bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mathematically and biologically consistent framework for presence-absence pairwise indices

A.L. Šizling[1*], P. Keil[2], E. Tjørve[3], K.M.C. Tjørve[3], J.D. Žárský[4], and D. Storch[1,4]

[1]Center for Theoretical Study, Jilská 1, 110 00, Praha 1, Czech Republic
[2] Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha – Suchdol, 16000, Czech Republic
[3]Inland Norway University of Applied Sciences, PO Box 952, NO-2604 Lillehammer, Norway;
[4]Department of ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2

sizling@cts.cuni.cz, storch@cts.cuni.cz.
kathy.tjorve@inn.no, even.tjorve@inn.no, jakub.dan.zarsky@gmail.com, pkeil@seznam.cz.
*corresponding author

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Aim Large number of indices for presence-absence data that compare two assemblages have been proposed or reinvented. Interpretation of these indices varies across the literature, despite efforts for clarification and unification. Most effort has focused on the mathematics behind the indices, their relationships with diversity, and between each other. At the same time, following issues have been largely overlooked: (i) inter-dependence of indices based on their informational value, (ii) overlap of the ecological phenomena that the indices aim to capture, (iii) requirement that a small re-arrangement of assemblages should only cause a small change in an index, and (iv) inferences from the indices about diversity patterns. Underappreciation of these issues has led to invention or reinvention of indices without increasing their information value. We offer a framework for pairwise diversity indices that accounts for these issues.

Methods We present a framework that links different ecological phenomena to pairwise indices, and show mathematical links between the indices. We distinguish statistical dependence of indices from their informational (mathematical) dependence. Using linear algebra, we found the minimum number of indices needed to detect different patterns.

Results We (1) classified existing indices into three major and four minor mutually independent families, (2) demonstrated how assemblage interrelation confuses conceptually different patterns, (3) showed what can be inferred about diversity phenomena from different indices, (4) demonstrated problems with most of the indices of nestedness, (5) provided formula linking mathematically (informationally) dependent indices, and (6) showed which combinations of indices can be used for meaningful ecological inference. Additionally, (7) we showed how to calculate any index from two presence-absence indices, which can be used to standardize and compare different indices across the literature.

Main conclusions It is impossible to purify an index of a single biodiversity phenomenon from the effects of other phenomena, because phenomena inevitably bound each other (e.g. a species richness gradient bounds possible values of Jaccard index of community similarity). Consequently, inventing indices which seemingly purify these effects (e.g. pure turnover or nestedness) leads to misleading inference. In contrast, a proper inference is obtained by using a combination of classical indices from different, information-independent families.

Key words: beta diversity, biodiversity, co-occurrence, partitioning, dependence, Simpson, dissimilarity

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Considerable effort has been made to capture variation in biological diversity using a single variable (e.g., Brualdi & Sanderson, 1999; Gotelli, 2000; Vellend, 2001; Koleff & Gaston, 2002; Gaston *et al*., 2007; Jost, 2007; Juranski *et al*., 2009; Ulrich *et al*., 2009; Baselga, 2010a; Tuomisto, 2010, Podani & Schmera, 2011; Almeida-Neto *et al*., 2012; Carvalho *et al*., 2012; Chao *et al*., 2012; Newbold *et al*., 2016; Schmera *et al*., 2020). Differences in species composition between assemblages in space and time have been assessed by various indices, with a heated debate and a little consensus about which are better for different purposes (Baselga, Podani & Schmera and Ulrich & Almeida-Neto, Baselga, 2010a, 2012; Podani & Schmera, 2011; Schmera *et al*., 2020; Almeida-Neto *et al.* ,2012; Ulrich *et al.*, 2009, 2017). In the debate, a particularly notable disagreement concerns indices of species nestedness, and how to purify them from the effects of species turnover and richness (and vice versa).

The debate over the proper index to compare species assemblages can be traced back to Simpson (1943), whose invention of the Simpson index (β_{Sim}; Tab. 1) resolved the undesirable interdependence between Jaccard index (J; Jaccard, 1912; Tab. 1) and the difference in species richness of assemblages. Since then, a number of indices have been introduced that carry equal information as Jaccard and Simpson indices (e.g. $\beta_{Sør}$; Sørensen, 1948; Tab. 1), or which are mathematically equivalent to earlier indices (e.g., $\beta_t = \beta_{SR} = \beta_{HK}$, defined by Wilson and Shmida 1984, Schluter and Ricklefs 1993, Harte and Kinzig 1997, respectively; Tab. 1). At the same time, we lack a general framework for relationships between existing indices, which would also apply for any index invented or introduced in the future. All this allows for generation of new, often redundant indices; for several indices new to the last decade see, e.g., Baselga, (2012) or Schmera *et al*. (2020).

Despite disagreements and redundancies, there has been progress. It is to Baselga's (2010a,b) credit that the field has refocused from classification of indices to their meaning. Lennon *et al*. (2001) and Newbold *et al*. (2016) showed that Simpson index is in fact Sørensen similarity relativized to the contrast in species richness. Koleff *et al*. (2003) and Legendre & De Caceres (2013) have shown independence of the Jaccard index from mean and total species richness. These authors also pointed out that Simpson and Jaccard represent two different groups of indices, which Koleff *et al*. (2003) labelled as broad-sense and narrow-sense turnover. This classification resonates with Almeida-Neto *et al.* (2008) and Šizling *et al*. (2009) who realized that reversed Simpson index (β'_{Sim}; Tab. 1) quantifies the phenomenon of nestedness, while the Jaccard index quantifies reversed species turnover. Lastly and importantly, Ulrich *et al.* (2017, 2018) pioneered null models that are able to disentangle mathematical and biological drivers of the dependence between various indices and species richness. All this abovementioned effort has concerned incidence-based indices (i.e. those based on binary presence-absence data). In parallel, a considerable effort concerned abundance-based indices, and their inclusion in the framework of Hill numbers (Hill, 1973; Jost, 2007; Tuomisto, 2010; Chao *et al*., 2012, 2014a,b; Chiu *et al*., 2014). However, accurate data on abundances are still more difficult to obtain and less common than simple presences and absences (Keil *et al.* 2021), particularly at large "macroecological" spatial scales. Because of this, pairwise indices based on presences and absences, rather than abundances, have been used in major global analyses of biodiversity (e.g. Dornelas *et al*., 2014; Blowes *et al*., 2019).

A particular attention has been paid to the dependence of the indices on species richness (Koleff *et al*. 2003, Jost, 2006,2007; Baselga, 2010b; Chao *et al.*, 2012; Schmera *et*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

al. 2020) and on each other (Simpson, 1943; Koleff *et al*. 2003; Baselga, 2010a; Baselga & Leprieur, 2015; Schmera *et al*. 2020), and there has been an effort to purify one effect from another by means of additive partitioning (Baselga, 2010a, 2012; Podani & Schmera, 2011). Jost (2006, 2007, 2010) and Chao *et al.* (2012) emphasized that while statistical dependence refers to the dependence between variables (two indices are statistically dependent if observing an index value modifies the probability to observe particular values of another index) the definititon of independence between indices should utilize their mathematical formulas. In order to replace the statistical dependence, Chao *et al.* (2012) defined unrelatedness as the situation where minimum or maximum of one index is affected by the variation of another index. Importantly, the relatedness says nothing about equal or different information carried by the variables in focus, but instead it induces statistical dependence as defined above. The reason is that changing constraints on minimum or maximum values necessarily modify probabilities to observe particular index values, which is the basis of statistical dependence.

Baselga (2010a,b, 2012), Podani & Schmera (2011) and Schmera *et al*. (2020) presented two new approaches to the problem of dependence among indices; both are based on mathematical partitioning of the indices, but differ in the way they standardize them. Baselga (2012) based his reasoning on the observation that some arrangements of assemblages affect both turnover and nestedness simultaneously, and the respective indices are thus 'related' (*sensu* Chao *et al*., 2012), which means that their random samples are statistically dependent. To address this dependence, Baselga (2012) defined "nestedness-resultant component" (β_{sne}), which is $\beta_{sne} \stackrel{\text{def}}{=} \beta_t - \beta_{Sim}$, where β_t is Sørensen dissimilarity and β_{Sim} is Simpson dissimilarity (see Tab. 1 for definitions). The subtraction supposedly removes the effect of turnover component represented by β_{Sim} from the overall dissimilarity represented by β_t, and what remains is considered to be 'nestedness resultant component'. In contrast, Podani & Schmera (2011) see nestedness as antithetic to replacement (a synonymum for turnover) and so indices of nestedness (β_{nps}; Podani & Schmera, 2011) and replacement (R_{ps}; Podani & Schmera, 2011) (see Tab. 1 for definitions) together sum to one. Their index of nestedness is thus $\beta_{nps} \stackrel{\text{def}}{=} 1 - R_{ps}$.

We argue that the abovementioned approaches have several problems (outlined in the chapter below) which have led to a confusing situation in which researchers are typically not sure which indices can be used in different situations and what the indices actually measure. To sort out the confusion, we take inspiration in physics, basic linear algebra, conditional entropy theory, and economics, and present a unified framework for incidence-based indices that addresses these issues. In our framework, the indices are tools to capture phenomena observed during fieldwork or data processing, and we show that these phenomena are already mutually dependent by their very nature, regardless of the indices which quantify them. This makes it impossible to find neither unrelated nor statistically independent indices. We further show that the interdependence between the phenomena makes it impossible to 'purify' any index from the effects of other phenomena. For instance, species spatial turnover (first phenomenon) is affected by the difference in species richness (second phenomenon). Partitioning them statistically on the level of indices does not lead to a proper inference, since the two phenomena affect each other irrespectively of any index.

Instead of statistical s-independence (or unrelatedness sensu Chao *et al.* 2012) of indices, we advocate for accounting for the independence of the indices in terms of their information content (which we call i-independence). We then show that only two i-independent indices carry enough information to compute any (even not yet introduced)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dimensionless index, and provide a user friendly tool to perform this conversion. We develop guidelines to attribute the indices to different diversty phenomena. Finally, we demonstrate that the interrelationship of assemblages introduces mutual i-dependence between otherwise i-independent indices, and that this effect can obscure differences between otherwise clearly distinguishable patterns. Our framework addresses the problems that have repeatedly acted as incentives for generating new indices. In doing so, we demonstrate that most of the work needed has already been achieved by Jaccard (1912), Simpson (1943) and others in early 20th century (see also Tjørve *et al*., 2018).

Problems with previously suggested indices

We distinguish five fundamental problems with the indices proposed in literature during last few decades. Our basic assumption is that different indices should measure different – albeit not neccesarily entirely independent – phenomena, and that the value of an index should reflect the strength of given phenomenon (Ulrich *et al*. 2017,2018). The problems listed below represent the cases when this general assumption is not fulfilled.

Problem 1: Some commonly used indices do not satisfy the *requirement of continuity* (Neumann & Morgenstern, 1953). The requirement, presented for ecologists by Podani & Schmera (2012), states that a *small* re-arrangement of species assemblages should lead to only a *small* change in the index value and conversly, a *small* change in index value should indicate only a small re-arrangement of the assemblages. This is important for two reasons: an index that changes considerably with a negligible re-arrangement of assemblages is (i) sensitive to errors in observed data, and (ii) leads to misleading inference about the pattern and/or ecological phenomenon. For example, common indices of nestedness (N_0, Patterson & Atmar, 1986; N_{NODF}, Almeida-Neto *et al.*, 2008) have been defined as zero when $S_X = S_Y$. This violates the the requirement of continuity (Fig. 1). To satisfy the requirement when $S_X = S_Y$, a value of nestedness should instead lie between nestedness for $S_X = S_Y - 1$ and nestedness for $S_X = S_Y + 1$. The requirement of continuity is also violated in β_{nps} (Podani and Shmera, 2011), which is discontinuously defined as $\beta_{nps} \stackrel{\text{def}}{=} 0$ when $S_{X\cap Y} = 0$, but $\beta_{nps} = (1 + |S_X - S_Y|)/S_{X\cup Y}$ when one species instead of zero is shared between the assemblages ($S_{X\cap Y} = 1$, Fig. 1). This discontinuity introduces sensitivity of β_{nps} to errors in data collection where $S_{X\cap Y}$ is small. Finally, the problematic assumption of zero nestedness for $S_X = S_Y$ also affects the "nestedness-resultant components" by Baselga (2010a, 2012) (Fig. 1). Baselga (2010a,2012) explicitly stated that in the absence of nestedness an index of dissimilarity equals to its turnover component. However, this statement would hold only if the equality of S_X and S_Y implied zero nestedness, even if the asemblages were preciselly identical and thus (in accord with the requirement of continuity) they should be similarily nested as two almost identical and absolutely nested assemblages.

Problem 2: The relationships between indices and ecological phenomena that the indices aim to capture (e.g. nestendess) have been typically derived using an ambiguous *from-pattern-to-index* implication: a pattern (particular arrangement of assemblages) affects the value of an index, so it is assumed that the index measures the phenomenon asociated with the pattern. However, the correct reasoning should follow a *from-index-to-pattern* implication: a value of an index should *always* correspond to a unique pattern that reflects

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

an ecological phenomenon (e.g. nestedness), making the index an unambiguous measure of the phenomenon. An example of the ambigous from-pattern-to-index implication is used by Podani and Shmera (2011), who argue that "*because species replacement and nestedness reflect contrasting ecological phenomena […] it is meaningful to express nestedness (β_{nps}) with the effect of species replacement (R_{ps}) completely removed from [its] maximum*". However, in Fig. 1 we show that cases with different levels of nestedness can have equal $\beta_{nps}(= 13/27)$, violating the from-index-to-pattern implication. Thus, β_{nps} is not an unambiguous measure of nestedness. Similarly, Baselga's reasoning described above is an example of a wrong implication: The fact that absence of nestedness leads to equal indices of Sørensen and Simpson does not at the same time mean that the equality of the indices implies zero nestedness. This problem of the incorrect implication was mentioned by several authors (e.g., Ulrich *et al*. 2017, 2018; Schmera *et al*. 2020) but has not been appreciated in practice.

Problem 3: So far, attempts to partition an effect of one index from another considered statistical dependence (we label it *s-dependence*) between indices (e.g., Simpson, 1943; Koleff *et al*. 2003; Baselga, 2010a, 2012, 2020; Lyasevska & Farnsworth, 2012). However, in terms of the exact meaning of the indices and their links to ecological phenomena, the dependence of indices in terms of their information content (hereafter *i-dependence*) is more fundamental. Information content of two indices is identical when each value of the first index uniquely determines the value of the second index, and each value of the second index uniquely determines the value of the first index (Orlitski, 2003) (i.e. if there is a bijection, which means that there is a function *F* so that $I_2 = F(I_1)$ and $I_1 = F^{-1}(I_2)$, where '*I*'s are the values of the first and second indices) (Fig. 2). When two indices have equal information content, they are i-dependent, which means that they cannot capture different phenomena. I-independent indices may be at the same time s-dependent, but this dependence stems from the dependence of the phenomena themselves and, as we show below, partitioning of s-dependent indices thus leads to a distorted link between the phenomena and the indices that are intended to measure them.

Problem 4: The mathematical operation of *subtraction* of indices does not remove one effect from another. According to Baselga (2010a, 2012) the subtraction results to nestedness-resultant component regardless of the index. Baselga (2012) applies the reasoning about nestedness-resultant components equally to β_t (see reasoning above) and Jaccard dissimilarity (β_G; Tab. 1). In the second case, the nestedness resultant component of Jaccard dissimilarity is $\beta_{jne} \stackrel{\text{def}}{=} \beta_G - \beta_{jtu}$, where β_{jtu} is 'turnover component of Jaccard dissimilarity' (Tab. 1). However, the subtraction can remove the effects only for β_t or β_G (or none of them), but not for both, and it is unknown which of β_{sne} and β_{jne} is the 'true' nesteness resultant component. The reason is that β_t is a unique transformation of β_G (Baselga 2012, SI.1), and that β_{Sim} is a unique transformation of β_{jtu} ($\beta_{jtu} = 2\beta_{Sim}/(1 + \beta_{Sim})$, SI.1). Thus β_t and β_G carry the same information and are i-dependent, and the same is true for β_{jtu} and β_{Sim}. If the subtraction produced nestedness-resultant components in both cases (Baselga 2012), then the subtraction would have to account for the same effect in both cases, and thus β_{sne} and β_{jne} would also have to be i-dependent, which is not the case (Baselga 2012). This proves that the subtraction does not work. This is analogous to the logarithmic transformation where a variable x carries equal information as $\log(x)$, and y carries the same information as $\log(y)$, but $x - y$ does not scale with $\log(x) - \log(y)$ -

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

there is bijection between $\log(x) - \log(y)$ and x/y (as $\log(x) - \log(y) = log(x/y)$) , but not between $\log(x) - \log(y)$ and $x - y$. In both cases the problem is that any nonlinear transformation also transforms the meaning of the operation between transformed variables. For the same reason the minus operator in $\beta_t - \beta_{Sim}$ ($\stackrel{\text{def}}{=} \beta_{sne}$) has a different meaning than the minus operator in $\beta_G - \beta_{jtu}$ ($\stackrel{\text{def}}{=} \beta_{jne}$). β_{sne} and β_{jne} thus cannot both capture nestedness-resultant components. Since the meaning of any mathematical operation between two indices depends on prior transformation of the indices, inferring the meaning of indices from the operations between them should be avoided.

Problem 5: More generally, although pairwise indices are dimensionless quantities and can take values within the same range (e.g. 0 and 1), they are not universally comparable (Schmera & Podani 2011). For example, the framework introduced in Baselga (2010a) subtracts and compares Sørensen dissimilarity, β_t, and Simpson dissimilarity β_{Sim} (Tab. 1); however, these all are not commensurable, as pointed out by Schmera & Podani (2011), who proposed to sum (or subtract) only indices with the same denominator.

Theory

In this part, we first describe five distinct phenomena that represent different aspects of the structure of species communities, and thus should be quantified by different indices. Second, we define the constraints necessary to attribute each of the indices to one spatial phenomenon. We argue that these indices (mathematical expressions) should capture the intuitive understanding of their respective phenomena. Third, we define the independence of biodiversity indices (including alpha diversity) in terms of their information value. We show that, to make correct and complete inference, we need to combine three indices that are i-independent. We contrast this i-independence with s-independence (including relatedness; Fig. 2), which was the major topic of large ecological literature (Simpson, 1943; Koleff & Gaston, 2002; Jost, 2007; Baselga, 2010a; Chao *et al.* 2012). We show that that s-independence is not achievable in most cases, and even if it was, it may not guarantee correct inference.

Distinguishing different diversity phenomena

We propose that most of the commonly used indices have originated as proxies for *five phenomena*: (1) Nestedness, (2) co-occurrence, (3) beta-diversity, (4) community turnover, and (5) species-richness gradient. Imagine ecologists roaming the landscape, leaving one site and approaching another. They see changes in species richness, changes in species composition, and overlap in species composition between species assemblages. These correspond, in the same order, to different (but not mutually exclusive) *patterns*, where: (i) species of the species-poor site are all present in the species-rich site; (ii) some species often co-occur at the same site, (iii) the checklist of observed species grows rapidly or slowly with increasing number of visited sites, (iv) new species replace the species of the site that the ecologist has left, but the species richness may or may not change, and (v) species richness either differs or remains the same between sites. Importantly, patterns (i-v) correspond to phenomena (1-5). A roaming ecologist observes a combination of these patterns, as there are species-poor sites that host species missing in species-rich sites, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

equally rich sites do not necessarily share the same set of species. Nevertheless, these five patterns delimit the whole range of possibilities. So, the ecologist can intuitively perceive the phenomena in the field even without the knowledge of the mathematically defined indices. This is like comparing temperature, speed, or mass without measuring them. Detailed descriptions of the five phenomena are:

Nestedness: The idea of nestedness was originally intended to capture the patterns in assemblages of archipelagos and/or of inland islands (e.g., mountain ridges). Patterson (1984), and Patterson & Atmar (1986) noticed that lists of species on mountain ridges varied, and that species of the species-poor sites were almost universally found at the species-rich sites. They suggested that this pattern was driven by selective extinction and called it a "nested pattern". Their index of nestedness, N_0, was designed to measure the deviation from a perfectly nested assemblage, in which species-poor sites have no unique species. This concept of nestedness fully agrees with the mathematical definition of nestedness where a subset is nested within the set. Although the original indices of nestedness (N_0, N_c, and Discrepancy, Patterson & Atmar, 1986; Wright & Reeves, 1992 and Brualdi & Sanderson, 1999) were designed for multisite assemblages, here we explore them using only two assemblages (Koleff *et al*., 2003; Gaston *et al*., 2007), provided that the indices for multi-site assemblages are extensions of two-site indices.

Co-occurrence: Co-occurrence has been originally tightly linked with the idea that competition structures communities (Connor & Simberloff, 1979; Diamond & Gilpin, 1982), and with assembly rules (Diamond, 1975; Gotelli & McCabe, 2002). Species co-occur if their spatial ranges (distributions) overlap (Gotelli, 2000). If a species occurs at several sites, and another species occurs at only one of the sites, co-occurrence of these two is apparently low, although their ranges are nested within each other. This is the main difference between co-occurrence and nestedness. In the literature, however, co-occurrence typically refers to the overlap of species ranges and nestedness mostly refers to the overlap in assemblages. A closer look at several measures of nestedness, however, reveals that indices of nestedness are often used to indicate the overlap of species ranges (Wright & Reeves, 1992; Brualdi & Sanderson, 1999; Šizling *et al*., 2009), which makes nestedness and co-occurrence seemingly inseparable from each other. Here, we argue that nestedness and co-occurrence may both refer to overlaps in species ranges and assemblages, but that they differ from each other where overlap is small due to a large contrast between range areas or species richness. Then nestedness can be high, but co-occurrence low.

Species spatial turnover and species-richness gradient: Species spatial turnover explicitly addresses the gain and the loss of species in space (Cody, 1975; Wilson & Shmida, 1984) and it is sometimes referred as species replacement (Podani & Schmera, 2011). This covers both a change in species richness (Harrison *et al*., 1992) and a change in species composition (Cody, 1975; Lennon *et al*., 2001). Although these two changes are bound by each other (change in species richness is always accompanied by a change in species composition to some extent), they are also mutually independent to some extent (two equally species-rich assemblages can have all or no species in common). It is therefore useful to split turnover, in its broad sense, into these two components (Lennon *et al*., 2001). Here we adopt the idea that the contrast between the species richness of two different assemblages (called species-richness gradient) and species spatial turnover are two different spatial phenomena (Lennon *et al*., 2001).

Beta diversity: The original idea behind β-diversity is that different regions have different relationship between local (alpha) and regional (gamma) diversity (Whittaker,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1960). Beta-diversity thus quantifies the contrast between gamma-diversity and average alpha-diversity (Whittaker, 1960). Beta-diversity thus does not primarily compare two assemblages with different locations, but instead a set of sub-assemblages with a merged assemblage of the whole region. The Whittaker (1960) formula that defines beta-diversity was later included in the indices of similarity between two different assemblages (Koleff *et al*., 2003) and several mathematical links between similarity indices and Whitaker beta-diversity were introduced (Koleff *et al.* 2003; Tuomisto, 2010; Chao *et al.* 2012). However, in accord with with Koleff *et al*. (2003), Jost (2007) and Tuomisto (2010) we contend that the two forms of comparison (i.e., between species richness of an assemblage and its sub-assemblages, and between two separate assemblages) should not be confused, and we will not deal with the issue of comparisons of multiple assemblages in this text, which is focused on pairwise indices.

These phenomena are interconnected and constrain each other. Apparently, if no species co-occur, then the turnover is higher than where all species co-occur. If all species co-occur, all sites are occupied by equal set of species, and so there is no richness gradient and no turnover. Removing species from species poor assemblages increases both the richness gradient and turnover, and perfectly nested assemblages show lower turnover than non-overlapping assemblages of the same richness. The interdependence and mutual constraints between the phenomena are their inevitable properties, and are consequently reflected by the indices.

Constraints needed to specify the phenomena

Here, we show the strict constraints that delimit the extreme cases of the phenomena described above. This will serve to attribute each index to a particular phenomenon, which is a prerequisite for any meaningful index. To describe the constraints mathematically, we will use six extreme re-arrangements of two sets as illustrated in the Venn diagram in Fig. 3. Each set can be considered either as a list of species, or as a species range (i.e. distribution in space). The constraints are:

- *Nestedness* has its maximum either where all the species from the species-poor assemblage are found in the species-rich assemblage (e.g., Patterson & Atmar, 1986; Baselga, 2010; Almeida-Neto *et al*., 2008), or where one range falls completely within the second range for all species in the assemblage (e.g., Wright & Reeves, 1992; Brualdi & Sanderson, 1999; Šizling *et al*., 2009). Therefore, two assemblages or ranges are maximally nested if one of them is completely contained within the other regardless of their size (as in re-arrangements r1, r2, r5, r6 in Fig. 3). The nestedness is at its minimum when none of the species is shared between two lists, or when there is no range overlap between two species (r3 and r4 in Fig. 3). No index should oscillate widely if re-arrangement changes only slightly, which is the abovementioned requirement of continuity *sensu* Neumann & Morgenstern (1953). Therefore, two *nearly* identical lists of species or two *nearly* identical ranges (r6 in Fig. 3) should have nestedness *near* to that of two *strictly* identical lists of species or ranges (r5 in Fig. 3). Since in the case of strictly identical lists one set of species is completely comprised within the other set, it also constitutes the maximum possible nestedness. This understanding is supported not only by the mathematical usage of the term "nestedness", which refers to a superset and its subsets, but it is also the basis of the conceptual (pre)definition by Patterson and Atmar (1986). Therefore, we do not attribute zero nestedness to two equally rich assemblages, or equally large ranges, as did the creators of N_0 (Patterson & Atmar, 1986), N_{NODH}

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Almeida-Neto *et al*., 2008), β_{nps} (Podani & Shmera, 2011), β_{sne} (Baselga, 2010a) and β_{jne} (Baselga, 2012); all of these indices are discontinuous, and as such fail to discriminate assemblages/ranges of equal size and zero overlap (r4 in Fig. 3) from assemblages/ranges of equal size and complete overlap (r5 in Fig. 3). This ambivalence renders inference from discontinuous indices spurious.

- Maximum *co-occurrence* is where two species occupy the same sites (r5 Fig. 3), whereas minimum co-occurrence is where the ranges are completely non-overlapping (r3, r4 in Fig. 3). Species that share only part of their ranges (r1, r2, r6 in Fig. 3) co-occur less than species with identical ranges (r5 in Fig. 3) and more than species with no shared occurrences (r3, r4 in Fig. 3). Importantly, the distinction between nestedness and co-occurrence therefore lies in re-arrangements r1 and r2 (Fig. 3), with equal nestedness ($\mathrm{Nest}[r1] = \mathrm{Nest}[r2]$ in Fig. 3) but different co-occurrence ($\mathrm{Co}[r1] < \mathrm{Co}[r2]$ in Fig. 3). We hold that species whose ranges are perfectly nested, but differ in their size, have a lower co-occurrence than species with two identical ranges ($\mathrm{Co}[r2] < \mathrm{Co}[r5]$ in Fig. 3).
- *Turnover* captures the contrast in species composition between two or more assemblages. Consequently, in Fig. 3, the turnover found in re-arrangement r3 must be larger than that of r1, which in turn represents a larger turnover than r2. In r5, the species lists or ranges are identical, hence there is no turnover at all, which means that the index of turnover is at its minimum. However, it is not intuitively evident whether the case r3 represents a larger turnover than that of r4. In our framework, we distinguish between turnover and species richness gradient (SRG, Lennon *et al*., 2001) and so we solve this dilemma by stating arbitrarily that $\mathrm{Turn}[\mathrm{r3}] = \mathrm{Turn}[\mathrm{r4}]$ but $\mathrm{SRG}[r3] > \mathrm{SRG}[r4]$ (Fig. 3).
- Because we define the *species-richness gradient* (SRG) as simply the contrast in species richness between two sites, the re-arrangements r1 and r3 in Fig. 3 represent the same SRG, as do r4 and r5 in Fig. 3. The latter also represents a zero SRG and therefore the minimum. Consequently, the value of the SRG in r2 Fig. 3 falls between the above.

Importantly, these constraints do not define the indices uniquely and two i-independent indices (see Fig. 2) may have the same constraints. In such case we say that these two indices capture two different aspects of the same phenomenon.

Definition of independence in terms of information content

To capture different phenomena by different indices, we need an idea of an independence of indices in terms of their information content – only the indices that do not carry the same information can distinguish different phenomena. To find out the information provided by an index, we need to see the definitions of the indices as equations to solve (Box 1). When we add an equation to the set of *n* equations, and if these *n*+1 equations provide equal solution as the set of n equations, then the new equation is said to be *i-dependent* on the others and carries no extra information. If the solution of the *n*+1 equations is a subset of (but not equal to) the solution of the *n* equations, then the new equation is *i-independent* of the others and carries extra information. Note that this is not a new term, but a widely accepted mathematical definition of independence in a system of equations (see also Box 1) and that it follows the theorem from information theory that only variables that are uniquely mapped to each other have equal informational content (Orlitski, 2003). In extreme circumstances, the set of equations with one unique solution carries complete information on assemblages. In the case of two assemblages, it all means that the complete

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

set of equations uniquely determines richness of both assemblages (S_X, S_Y) as well as species overlap ($S_{X\cap Y}$; for details see Box 1).

We utilize this mathematical definition of independent equations, and define the indices that are mutually i-independent if the equations of the indices are mutually independent (Box 1). The β'_{Sim} (Simpson nestedness) and J (Jaccard similarity) (Tab. 1) are mutually independent because there is no way to convert β'_{Sim} to J. This is easy to see from Fig. 4a and b, where more than one value of β'_{Sim} is attributed to each value of J ($J \neq 0$). This is an analogy to a case from physics, where the kinetic energy (E_k) and momentum (p) of a moving object cannot be calculated from each other although they are computed using the same variables ($E_k = 0.5mv^2$, $p = mv$; m is mass and v is speed). It induces relatedness (according to Chao *et al.* 2012), but *not one-to-one* correspondence (Fig. 2) between energy and momentum (Fig. 4c, d). However, because E_k and p are independent in terms of the lack of *one-to-one correspondence*, the equations of these variables can be combined to determine the mass and speed of an object. All this means that J and β'_{Sim} carry different information about assemblages, similarly to E_k and p that carry different information about moving objects, although statistical tests would reveal their mutual statistical dependence, and constraints reveal their relatedness.

In contrast to this definition of i-independence, the independence that is used in ecological literature is almost always of statistical nature, and we call it s-dependence. While the i-independence refers to a mathematical formula, the s-independence refers to data values (i.e., a conditional probability of measuring a particular value of one variable, given a value of another variable). Some indices are i-independent, while their values are s-dependent because their definitions are related sensu Chao *et al.* (2012). For example, J and β'_{Sim} are i-independent but related, and thus their values generated across multiple assemblages are correlated (Fig. 4a, b).

Results

Families of indices

In the previous section, we have developed a framework to assign the indices to different phenomena. Now, we will group them to families within which indices share equal information, and thus capture the same phenomena. Then we proceed by assigning a phenomenon to the index which is most often used or most convenient for some reason. Finally, we will show how to convert indices within their respective family to each other (Box 2). Mathematical details of these steps are in appendix SI.2.T1-19, and here we summarize the results.

There is a number of diversity indices and each has some information value. Some of them, however, are i-dependent, and they are thus equivalent in terms of any inference. To distinguish families of i-dependent indices, we first examined their mutual bivariate relationships, i.e. we simply plotted the indices against each other, using simulated and empirical data (Figs. 5,6). Where we did not see a one-to-one correspondence (Fig. 2), i.e. the line joining the points in the plot was neither strictly increasing nor strictly decreasing, the two plotted indices were considered i-independent. Where we found a one-to-one correspondence along a strictly increasing or decreasing curve, we had to prove the i-dependence by mathematical analysis (SI.2). This way we found three major and five minor families of mutually equivalent indices (Fig. 5), and our mathematical analysis (SI.2)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

confirmed the one-to-one correspondences (i-dependence) between the indices within these families. The three major families of more than one index can be attributed to turnover, nestedness, and species richness gradient.

1. Jaccard index family. The largest major family of indices is grouped around the Jaccard index (J, Jaccard, 1912). It has 9 indices of (dis)similarlity, including Sørensen ($\beta_{Sør}$, as defined in Gaston *et al*., 2007), β_{HK} (Harte & Kinzig, 1997), β_G (Gaston *et al*., 2001), β_{SR} (Schluter & Ricklefs, 1993), β_t (Wilson & Shmida, 1984) and β_{DICE} (Raup & Crick, 1979). The Bray-Curtis index of dissimilarity (β_{BC}, Bray & Curtis, 1957), belongs to the Jaccard index family when computed from incidences. A classical index of nestedness, N_c (Wright & Reeves, 1992) also belongs to the Jaccard-index family (Fig. 5, SI.2.T8) when standardized by $S_X + S_Y$ (Gotelli & McCabe, 2002). The indices of the Jaccard-index family capture the phenomenon of species turnover (Fig. 5, SI.2.T15). To be precise, J measures species co-occurrence (proportion of species that co-occur in both assemblages), which is a reversed measure of species turnover (the relationship between an index and the reversed index is strictly decreasing), and all indices from the Jaccard-index family that decrease with increasing J (e.g., β_t or $\beta_{Sør}$) are measures of turnover (Fig. 5). Indices from Jaccard family are measures of co-occurrence of two species if we replace S_X, S_Y and $S_{X\cap Y}$ by the numbers of occupied sites by species X and Y.

2. Simpson-beta family. The second major family consists of Simpson's beta (β_{Sim}, Simpson, 1943), r-Nestedness (rN, applied to species lists; Šizling *et al*., 2009), and Simpson's nestedness (β'_{Sim}, Simpson, 1943). As β'_{Sim} is identical to rN when applied to species lists, we do not plot these indices separately in Fig. 5. Two classical indices of nestedness, the Discrepancy (D, Brualdi & Sanderson, 1999), and the standardized N_C by Wright & Reeves (1992) also belong to the Simpson-beta family under certain circumstances. Specifically, D (standardized as in Greve *et al*., 2005) converges to the mean across all β_{Sim} (i.e. $\langle\beta_{Sim}\rangle$) as number of sites increases (SI.2.T13). In practice, n>10 guarantees

$$D \cong \langle\beta_{Sim}\rangle \qquad (1)$$

(Fig. SI.3); $\langle x\rangle$ stands for a simple mean accros all x's. N_C belongs to Simpson beta family if standardized as suggested by its authors (Wright & Reeves, 1992; see Tab. SI.1) and if species richness is high (SI.2.T9). The indices from the Simpson-beta family capture the phenomenon of nestedness (SI.2.T4). β_{Sim} and D are measures of reversed nestedness, whereas β'_{Sim} and rN measure nestedness, applied at the level of assemblages and ranges, respectively.

*3. Species richness gradient/uniformity***.** The third (minor) family so far consists of just two indices: β_{gl}, (Tab. 1, Lennon *et al*., 2001) and R (Tab. 1, Newbold *et al*., 2016). The β_{gl} captures the contrast in species richness between two sites, which is the phenomenon of species richness gradient (SI.2.T14). R scales negatively with β_{gl} and so we call R an index of species-richness uniformity. Although β_{gl} is older, the R is simpler and thus more tractable than β_{gl}, and so we will use R in equations describing the relationships between the families (Eqs. 2,4,5,B8, SI.1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The other minor families are each made up of only a single index (Figs. 5). Neither of these indices capture any phenomenon under consideration (SI.2.T7,10,12). Of these only Ruggiero index of beta-diversity (β_{Rib}, Ruggiero *et al*., 1998) satisfies the requirement of continuity. The others are: Baselga nestedness-resultant components of both Sørensen dissimilarity (β_{sne}, Baselga 2012, labelled as β_{nes} in Baselga, 2010) and Jaccard dissimilarity (β_{jne}, Baselga, 2012; not plotted); the classic index of nestedness, N_0 (computed for two sites; Patterson & Atmar, 1986); and the N_{NODF} index of nestedness (computed for two sites; Almeida-Neto *et al.*, 2008). The N_{NODF} would have belonged to the Simpson-beta family if it did not violate the requirement of continuity, which splits the scaling line in the bivariate plot into two different trajectories (Fig. 5), which makes any inference difficult.

i-independent combinations of indices

Although we have delineated four families of indices that satisfy the requirement of continuity (hereafter continuous indices), it does not mean that there are four i-independent indices. The reason is that three indices from any three families are i-dependent even if any single pair from this triplet is i-independent. Only two families are then i-independent with certainty, and the indices of the other families can be determined from these two (Box 1; this does not hold for Whittaker beta-diversity due to the fact it is not a pairwise index, and for this reason we avoid discussing this index). A clear case of mutual i-dependence of a triplet of indices is the combination of the Jaccard-index family, the Simpson-index family and the species-richness-gradient family. A simple rearrangement of Eq. B5 (where $R = I$ and $k_1, k_2, l_1, l_3 = 0,\ k_3, l_2 = 1$, see Box 1), gives

$$\frac{J}{1+J} = \beta'_{Sim}\frac{R}{1+R} \quad \text{and} \quad \langle\frac{J}{1+J}\rangle = \langle\beta'_{Sim}\rangle\langle\frac{R}{1+R}\rangle + cov(\beta'_{Sim}, \frac{R}{1+R}) \qquad (2)$$

which says that reversed turnover increases with increasing nestedness and species richness uniformity (reversed gradient). The covariance corresponds to the effect described by Diserud & Ødegaard (2007) and Baselga (2013). The result (Eq. 2) can be generalized (Box 2) to all the considered indices that can be defined by Eq. B1, which includes all dimensionless and continuous indices listed in Tab. 1 and SI.1 (Box 1), and even indices that have not yet been introduced.

The Eq. B5 unifies indices across ecological studies, assuming that authors calculated and published at least two i-independent presence-absence indices for their data.

Special cases: families of indices in the case of assemblage interrelation

The mutual i-independence of the indices implicitly assumes that species assemblages are completely 'independent' of each other. However, this may not be the case, as assemblages are often interrelated due to similarity in habitats or due to dispersal. This leads to two special cases of behavior of the indices. First, let us assume an effect that limits variation of species richness between sites. In extreme, $S_X = S_Y$. In this case, there is no species richness gradient, $R = 1$ and Eq. 2 turns into

$$\beta'_{Sim} = 2\frac{J}{1+J} \quad \text{and} \quad \langle\beta'_{Sim}\rangle = 2\langle\frac{J}{1+J}\rangle. \qquad (3)$$

This means that the Jaccard-index and the Simpson-beta families become i-dependent (Fig. SI.4). Second, assemblages can be perfectly nested, i.e. the species-poor assemblages have

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

no unique species. In this case, $S_{X\cap Y} = \min(S_X, S_Y)$ and so $\beta'_{Sim} = 1$. It follows from Eq. 2 that

$$\frac{J}{1+J} = \frac{R}{1+R} \Longrightarrow J = R, \text{ and } \langle\frac{J}{1+J}\rangle = \langle\frac{R}{1+R}\rangle. \quad (4)$$

It reads that the families of Jaccard-index and the species richness gradient are no longer i-independent (Fig. SI.4). Third, assemblages can have equal proportion of shared species which gives

$$R^{-1} = C \cdot \beta'_{Sim} - 1, \text{ and } \langle R^{-1}\rangle = C\langle\beta'_{Sim}\rangle - 1, \quad (5)$$

where C is a constant. These special cases can easily be recognized in any data. It follows that $\langle\beta'_{Sim}\rangle$ approaches one in the case of a fully nested pattern, and $\langle 2J/(1+J)\rangle$ (see Eq. 3) in the case of a low spatial variation in species richness. In the latter case R approaches one.

The loss of i-independences in these special cases is trivial, but it illustrates an important point: Even if indices are i-independent, they appear more dependent on each other as species richness values across sites become similar, and/or when assemblages approach complete nestedness (compare Figs. 6 and SI.4).

Statistical non-equivalence of the indices

We have argued that the indices within a given family are equivalent, since they are i-dependent and provide the same information. However, they are not equivalent statistically. The reason is that they often scale non-linearly with each other. For instance, within the Jaccard-index family there are three groups of indices that scale in a linear manner to each other (group i: J, β_G; group ii: N_c, t, β_{sr}, β_{HK}, β_{BC}, $\beta_{Sør}$; group iii: β_{DICE}; Fig. 5), and are statistically equivalent. Between these groups, however, the indices scale nonlinearly (see, e.g., J and $\beta_{Sør}$ in Fig. 5), and this decreases the correlation coefficient between otherwise i-dependent indices, which decreases s-dependence between the indices (e.g., Koleff & Gaston, 2002; Lyashevska & Farnsworth, 2012).

The values of an index within one family that scales in a non-linear manner (Fig. 5) with another index can also be seen as transformed values of the other index. For example, the $\beta_{Sør}$ and β_{HK} are two indices that originated as equal transformations of J (Fig. 5, Tab. SI.1). However, this means that these values have different frequency distributions, which can affect parametric statistical tests and their sensitivity. This could lead to the conclusion that it is best to use an index with the most symmetric/regular frequency distribution in a given system. However, the commonly used indices represent a rather poor spectrum of transformations (Fig. 5). We argue that it's more logical to first pick the index that best describes the phenomenon of interest, and transform/normalize it using an appropriate transformation (e.g., logit-transformation), rather than pick and choose from the range of existing indices with "good" statistical properties. Alternatively, it is possible to test the index values against a null model, and then the precise distribution of the values is not an issue (Ulrich & Gotelli, 2007, 2013; Chase *et al*., 2011).

Consequences of "purifying" an index from the effect of other phenomena

We have shown that it makes little sense to develop indices that purify one aspect of diversity patterns, controlling for the other aspects; an example is the attempt to partition

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the nestedness-resultant and turnover components (Baselga, 2010a, 2012). The reason is that s-dependence represents an inherent property of the phenomena, where different aspects bound each other (they are related sensu Chao *et al.* 2012). For instance, inequality in species richness (diversity gradient) bounds possible values of species turnover, imposing its bottom limit. In other words, s-dependence and relatedness of indices mirror dependence between phenomena, which makes summing and subtracting of indices problematic.

In all situations when we compare two regions or periods, or test for (non)randomness in diversity pattern with respect to a particular phenomenon, a single index that captures the focal phenomenon is sufficient, regardless of its s-dependence on other indices. We will demonstrate the irrelevance of s-dependence of indices on the triplet J, β'_{Sim}, R (as any pair of i-independent indices is convertible to these three indices; Box 1,2). S-dependence (including relatedness; Chao *et al*. 2012) between these indices is driven by inequalities:

$$0 \leq J =_1 R\frac{\beta'_{Sim}}{R+1-R\beta'_{Sim}} \leq_2 \frac{\beta'_{Sim}}{2-\beta'_{Sim}}, \text{ and } 0 \leq J \leq_3 R. \qquad (6)$$

The first equality ($=_1$) results from Eq. 2, the next inequality ($\leq_2$) involves β'_{Sim} and is a consequence of $0 \leq R \leq 1$, and the last inequality ($\leq_3$) is a consequence of both $0 \leq \beta'_{Sim} \leq 1$ and the first equality ($=_1$). The $\leq_3$ is well known relatedness between J and the contrast between alpha-diversities, which has led to the definition of β_{Sim} (Tab. 1). Both bounds ($\leq_{2,3}$) have been taken as an argument that J and β_{Sim} have to be corrected for the effect of the other indices (e.g., Simpson, 1943; Baselga, 2010a). We argue that, when we are interested in turnover, indices of the Jaccard family need no correction for variation in species-richness gradient. The rationale is that (i) the J is a measure of reversed turnover or co-occurrence regardless of its relatedness with R, and that (ii) the bounds only show the limits of maximum possible reversed turnover given the level of nestedness or uniformity (*R*) in species richness across sites. Obviously, it would be possible to use the value of *J* relative to its maximum ($\leq_1$), but this value reflects just the level of uniformity *R* (Fig. 7), so using this relative *J* is equivalent to direct measurement of *R*.

A special case is the widely discussed s-dependence (which includes relatedness) between pairwise indices and species richness (Koleff et al. 2003, Jost, 2006, 2007, 2010, Baselga, 2010b, Chao et al., 2012, Legendre & De Cáceres, 2013). All dimensionless indices are, by definition, i-independent of species richness (SI.5), and species richness can be used (together with, e.g., J and β_{Sim}) to make correct inferences about assemblages (Boxes 1,2). However, dimensionless indices are also necessarily s-dependent on species richness, especially when species richness is low (see SI.5 for the mechanism of s-dependence between dimensionless indices and species richness).

A case study: How to make inference from the indices

To provide an example of an inference based on similarity indices we use indices from Šizling *et al*. (2016) who studied temporal change of Central European plants during Holocene. They reported an increase in mean species richness $\langle S \rangle$ from 23 to 36, almost no change of β'_{Sim} ($\beta'_{Sim} \cong 0.7$), and an increase of average J from app. 0.38 at the end of the last Ice Age to the recent value of app. 0.48. J and β'_{Sim} determine *R* (eq. 2), and we can additionally calculate the contribution of species-poor ($S_{poor} = \langle S \rangle\, 2R/(R+1)$) and

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

species-rich ($S_{rich} = R \cdot S_{poor}$) sites to the average alpha diversity. These contributions result from the solution of Eqs. B3 and B4 (Box 1, see SI.2.T17 for the derivation). We thus find that R increased from 0.65 to 0.86, S_{poor} increased from app. 18 to 33, and S_{rich} increased from app. 28 to 39. We conclude that recent plant assemblages in Central Europe show lower spatial turnover (higher similarity), and lower contrasts in species-richness than those 14,000 years ago. Nestedness has not changed during this period, however, the contribution of species-poor sites to the mean alpha diversity increased more ($\Delta S_{poor} = 15$) than contribution of species rich sites ($\Delta S_{rich} = 11$). The spatial pattern in plant diversity therefore became more uniform with a proportionally higher increase of richness at species-poor sites, and temporally increasing average richness of all sites. Regardless of the higher recent uniformity, the species poor sites used to have similar proportion of unique to common species as today (an inference from the constant nestedness).

The results of this empirical analysis directly contradict Baselga's (2010a,2012) interpretation, problematic for all the reasons demonstrated in previous sections. When we used Baselga's (2012) partitioning method on our Holocene plant data, we got $\beta_{sne} = 0.15$, $\beta_{Sim} = 0.3$ for the past and $\beta_{sne} = 0.05$, $\beta_{Sim} = 0.3$ for the recent. This reads in the Baselga's (2010a,2012) framework as no temporal change of turnover component, and a decrease of nestedness-resultant component, which is the exact opposite to the trends described above. This shows that using the mathematically consistent framework is not just a cosmetic improvement of the otherwise intuitive and robust trends; rather, it can entirely invert the interpretation of the analysis. Furthermore, Baselga's (2010a,2012) interpretation suggests stronger contribution of turnover than nestedness in the past ($\beta_{Bas} < \beta_{Sim}$), but the patterns reported above, based on direct interpretation of i-independent indices, do not suggest anything like that. We have already shown that nestedness and turnover are incomparable (Box 3), and that it is impossible to say if nestedness contributes to species pattern more or less than turnover.

Practical guidelines

Practical inference from pairwise indices can have different purposes (Anderson *et al*., 2010), for instance: (i) exploration of diversity (e.g. Qian *et al*., 2009) or (dis)similarity (e.g., Simpson, 1943) between biotas of two sites or regions, (ii) revealing non-random origin of a spatial or temporal biodiversity patterns (e.g. Patterson & Atmar, 1986; Ulrich & Gotelli, 2013), (iii) meta-analysis based on indices extracted from the literature, and (iv) exploration of the behavior of the indices along a temporal, spatial or environmental gradients. Based on previous considerations, here are practical recommendations for using the presence-absence indices:

1. Avoid using 'component' indices that supposedly remove an effect of one phenomenon from an effect of another phenomenon (e.g. purifying turnover from the effect of richness gradient, or purifying nestedness from the effect of turnover). These indices are either mathematically flawed (e.g. they violate the requirement of continuity), and their meaning is thus unclear, or they measure a different phenomenon from what they were originally claimed to measure. An example of the latter is the standardization of an index relative to its maximum value, e.g. standardization of Jaccard index by its maximum given by the nestedness, which in fact directly measures the species richness gradient (Fig. 7). The reason is that nestedness, together with richness gradient *R*, uniquely determine *J*, so that measuring *J* relative to *Jmax* (bounded by nestedness) is equivalent to measuring *R*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. When choosing an index, first consider which phenomenon it should capture, then select a corresponding family of indices (see Fig. 5, Tab. SI.1, user-friendly calculator SI.6). It does not matter which index within the family is selected, as all indices within any family can be converted to each other (Tab. SI.1) and are thus practically equivalent. If there is a need for an appropriate statistical distribution of the index, a proper transformation can be used.

3. When comparing already published indices, use the equations in Tab. SI.1 or calculator, SI.6 to convert them to a common index which is the most desirable for a given purpose. If a desirable common index is not available in the publication, but two indices from two different families are, they can be used to calculate any other index (using equations from Boxes 1,2 or calculator SI.6), as two indices from different families uniquely determine an index from any other family.

4. For evaluating the importance of different phenomena (e.g. nestedness vs. turnover), use null models that randomize species incidences (Gotelli & Ulrich, 2012) to eliminate the undesirable phenomenon, and produces a distribution of an index (expected under the absence of the phenomenon) that can then be compared to the observed value of the index. Without a null model, the indices are not commensurable (see above) and comparison of their values does not make a sense.

6. When publishing indices from your research, ensure that these indices conserve complete information, i.e., publish at least two i-independent indices (J, and β_{Sim} or R are recommended) plus regional and mean local species richness, if available.

Discussion

We have proposed a new and mathematically consistent framework for presence-absence biodiversity indices. We have identified the sources of the proliferation of new indices in ecological literature as: (1) the search for indices that are statistically independent of each other and of species richness, (2) the habit of designing indices without considering all constraints necessary for delineating the phenomena that the indices aim to capture, and (3) the lack of tools to make inferences about spatial and temporal biodiversity patterns. Our framework avoids these pitfalls.

We demonstrated that the phenomena that are measured by different indices constrain each other, and respective indices are thus necessary related sensu Chao *et al.* (2012). This leads to their statistical dependence. It is pointless to search for s-independent indices, because (i) s-independence concerns index values instead of index formulas and thus s-dependence is a matter of the structure of data rather than index definitions; and (ii) s-dependent values emerge even from the desired i-independent indices. In other words, there are no s-independent indices that capture different, but necessarily mutually dependent phenomena.

On the other hand, i-independent indices are desiderable, as they provide (by definition) different information about the system. However, a sufficient number of i-independent indices have already been invented (and any not yet invented index is i-dependent on J, and β_{Sim}), and there is thus no need for new indices. We have

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mathematically proven that two i-independent indices, in combination with a species richness value (which acts as a scaling parameter), provide complete information about any system (SI.2.T1). Moreover, when variation of species richness across sites is negligible, or when there are no species that are unique to the species-poorer assemblage (i.e. perfect nestedness), only one index is sufficient. The interrelation between assemblages which leads to these effects thus leads to higher correlations between the observed index values than what would follow from the theory (Ulrich *et al.* 2017), which makes some families almost indistinguishable (Figs. 6, SI.4). This is apparently the reason why it has been so difficult to achieve agreement regarding which indices characterize different phenomena (Lennon *et al*., 2001; Gaston *et al*., 2007; Ulrich & Gotelli, 2007; Baselga, 2012; Podani & Schmera, 2011; Ulrich *et al.* 2017; Schmera *et al.* 2020).

In terms of practical utility of the indices, the finding that the five phenomena distinguished by ecologists are mutually dependent and that two i-independent indices fully characterize the system implies that it is reasonable to calculate just two indices belonging to different families (see Results section) to make a proper inference. These indices (for example J and β'_{Sim}) can be, if necessary, converted to any other index within respective families (and vice versa). However, we have shown that indices from different families cannot be directly compared to each other, as their absolute values characterize different phenomena. Different indices within a family of i-dependent indices are just different transformations of each other and their particular values thus reflect just different ways (formulas) how they are calculated. Therefore, if we need to evaluate the index values in terms to what is expected, a proper null model is needed (Ulrich & Gotelli, 2013). This does not exclude the possibility to compare the values of the indices, e.g., among different regions or taxa, but it is crucial to keep in mind that the values are inherently simultaneously affected by different phenomena, and any attempts to purify them leads to spurious results.

We have focused only on pairwise indices, and avoided speaking about beta-diversity in its strict, original sense (Whittaker, 1960), which comprises the relationship between local and regional species richness, or, more specifically, between alpha and gamma diversity. This relationship is related to the turnover, since when alpha is considerably lower than gamma, there must be high turnover among communities. However, exact mathematical links between pairwise community turnover and Whittaker's beta is a separate issue (Koleff *et al*., 2003; Tuomisto, 2010; Chao *et al.*, 2012).

Our findings have important consequences for understanding to the forces shaping distance decay in assemblage similarity (Nekola & White, 1999). A direct consequence of Eq. 2 is that distance decays of nestedness (β'_{Sim}) and turnover (measured by $J/(1+J)$) follow the same mathematical function only when there is no distance decay in the species richness gradient. This happens only in environments without geographical gradients that affect the species richness, for example, gradient in productivity (e.g., Willig *et al*., 2003; Currie *et al*., 2004). In these cases spatial structuring of the assemblages is anisotropic, and there are different rates of distance decay in different directions. Even more importantly, it is meaningless to assume a universal functional form of the distance decay of community similarity, e.g. exponential (Nekola & McGill, 2014). This follows from the mutual non-linear scaling between different indices, even between the i-dependent ones. Consequently, if one index reveals, say, exponential distance decay, another index can reveal a non-exponential distance decay.

Unfortunately, our results question the interpretation of hundreds of empirical assessments that used the subtractive partitioning of pairwise indices. To date, the original

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

publication describing the partitioning to nestedness and turnover components (Baselga, 2010a) has 2,493 citations, publication announcing the R package betapart which does the partitioning (Baselga & Orme, 2012) has 1,317 citations, and the R package on CRAN (Baselga, *et al*. 2018) has 510 citations (Google Scholar accessed on 19 May 2022). Most of these are empirical studies, including some high-profile ones in top journals (e.g., Molinos *et al*. 2016; Gotelli *et al*., 2017; Rocha *et al*., 2018; Blowes *et al.*, 2019; Chase *et al*., 2020). The approach has also been gaining momentum in young fields, for example in microbial ecology (Shade *et al*., 2013). Furthermore, the approach has been extended to partition functional (Villéger *et al*., 2013) and phylogenetic (Nowakowski *et al*., 2018) nestedness and turnover. To start with, our recommendation is to simply stop. The next step would then be to assess the extent of the damage that has been done, potentially re-analyzing and re-interpreting some of the studies; in this effort, our equations from Boxes 1,2 (or Tab. SI.1 or user friendly calculator SI.6) can be helpful.

In conclusion, our framework systematically deals with problems that unnecessarily generated new indices, and it resolves old issues concerning the mutual dependence of the indices. Based on the distinction between two types of dependence (i- vs s-dependence), we offer a tool for making inference using classical indices, a tool that can be further developed when new spatial or temporal phenomena are identified. Our framework is mathematically consistent as a whole, which is an advantage over earlier works that have accumulated new indices and phenomena without a unifying concept.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Almeida-Neto, M., Frensel, D.M.B. & Ulrich, W. (2012) Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). *Global Ecology and Biogeography,* **21**, 772–777.

Almeida-Neto, M., Guimaraes, P., Guimarães, P.R., Loyola, R.D., & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227-1239.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., & Harrison, S.P. (2010) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology letters*, **14**, 19-28.

Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity , *Methods in Ecology and Evolution*, **3**, 808–812.

Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur., F., Logez, M., Henriques-Silva, R., Martinez-Santalla, S., Martin-Devasa, R. Gomez-Rodriguez, C. & Crujeiras R. (2018) betapart: Partitioning beta diversity into turnover and nestedness components. R package version 1.5.6 http://CRAN.R-project.org/package=betapart

Baselga, A. (2010a). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134-143.

Baselga, A. (2010b) Multiplicative partition of true diversity yields independent alpha and beta components; additive partition does not. *Ecology*, **9**, 1974-1981.

Baselga, A. (2013) Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography*, **36**, 124–128.

Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223-1232.

Baselga, A. and Leprieur, F. 2015. Comparingmethods to separate components of beta Diversity. *Methods in Ecology and Evolution*, **6**, 1069–1079.

Blowes, S.A., Supp, S.R., Antão, L.H., Bates, A.,Bruelheide, H., Chase, J.M.Moyes, F., Magurran, A., McGill, B., Myers-Smith, I.H., Winter, M., Bjorkman, A.D., Bowler, D.E., Byrnes, J.E.K., Gonzalez, A., Hines, J., Isbell, F., Jones, H.P., Navarro, L.M., Thompson, P.L., Vellend, M., Waldock, C. and Dornelas, M.2019. The geography of biodiversity change in marine and terrestrial assemblages. *Science*, **336**, 339-345.

Bray, J.R. & Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs*, **27**, 325-349.

Brualdi, R.A. & Sanderson, J.G. (1999) Nested species subsets, gaps, and discrepancy. *Oecologia*, **119**, 256-264.

Carvalho, J.C., Cardoso, P. & Gomes, P. (2012) Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, **21**, 760-771.

Chao, A., Chiu, C.H. & Hsieh, T.C. (2012) Proposing a resolution to debates on diversity partitioning. *Ecology*, **93**, 2037-2051.

Chao, A., Chiu, C.H. & Jost, L. (2014a) Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 297-324.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014b) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45-67.

Chase, J.M., Blowes, S.A., Knight, T.M., Gerstner, K. & May, F. (2020) Ecosystem decay exacerbates biodiversity loss with habitat loss. *Nature* **584**, 238–243.

Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M. & Inouye, B.D. (2011) Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere*, **2**, 24

Chiu, C.H., Jost, L. & Chao, A. (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs*, **84**, 21-44.

Cody, M.L. (1975) Towards a theory of continental species diversities. In: *Ecology and Evolution of Communities* (ed. by M.L. Cody, and J.M. Diamond), pp. 214–257. Harvard University Press, Cambridge, Massachusetts, USA.

Connor, E.F. & Simberloff, D. (1979) The assembly of species communities: chance or competition? *Ecology*, **60**, 1132–1140.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.F., Hawkins, B.A., ... & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology letters*, **7**, 1121-1134.

Diamond, J.M. & Gilpin, M.E. (1982) Examination of the ''null'' model of Connor and Simberloff for species cooccurrences on islands. *Oecologia*, **52**, 64–74.

Diamond, J.M. (1975) Assembly of species communities. In: *Ecology and Evolution of Communities* (ed. by M.L. Cody, and J.M. Diamond), pp. 342-444. Harvard University Press, Cambridge, Massachusetts, USA.

Diserud, O.H. & Ødegaard, F. (2007) A multiple-site similarity measure. *Biology Letters,* **3**, 20–22.

Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. and Magurran, A.E. 2014. Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. *Science*, **344**, 296-299.

Gaston, K. J. and Blackburn, T. M. 2000. Pattern and process in macroecology. – Blackwell.

Gaston, K.J., Evans, K.L. & Lennon, J.J. (2007) The scaling of spatial turnover: pruning the thicket. In: *Scaling Biodiversity* (ed. by D. Storch, P.A. Marquet, and J.H. Brown), pp. 181–214. Cambridge University Press, Cambridge.

Gaston, K.J., Rodrigues, A.S.L., Van Rensburg, B.J., Koleff, P. & Chown, S.L. (2001) Complementary representation and zones of ecological transition. *Ecology Letters*, **4**, 4-9.

Gotelli, N. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, **121**, 171-180.

Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: A meta-analysis of JM Diamond's assembly rules model. *Ecology*, **83**, 2091-2096.

Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606-2621.

Gotelli, N.J., Shimadzu, H., Dornelas, M., McGill, B., Moyes, F. & Magurran, A.E. (2017) Community-level regulation of temporal trends in biodiversity. Science Advances, **3**, e1700315.

Greve, M., Gremmen, N.J., Gaston, K.J. & Chown, S.L. (2005) Nestedness of Southern Ocean island biotas: ecological perspectives on a biogeographical conundrum. *Journal of Biogeography*, **32**, 155-168.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Harrison, S., Ross, S.J. & Lawton, J.H. (1992) Beta diversity on geographic gradients in Britain. *Journal of Animal ecology*, **61**, 151-158.

Harte, J. & Kinzig, A.P. (1997) On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. *Oikos*, **80**, 417-427.

Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427-432.

Jaccard, P. (1912) The distribution of the flora in the alpine zone. *New phytologist*, **11**, 37-50.

Jost, L. (2006) Entropy and diversity. *Oikos*, **113**, 363-375.

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427-2439.

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427-2439.

Jost, L. 2010. Independence of alpha and beta diversities. *Ecology*, **9**, 1969-1974.

Jurasinski, G., Retzer, V. & Beierkuhnlein, C. (2009) Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia*, **159**, 15-26.

Keil, P., Wiegand, T., Tóth, A.B., McGlinn, D.J. and Chase, J.M.2021. Measurement and analysis of interspecific spatial associations as a facet of biodiversity. *Ecological Monographs*, **91**, e01452

Koleff, P. & Gaston, K.J. (2002) The relationships between local and regional species richness and spatial turnover. *Global Ecology and Biogeography*, **11**, 363-375.

Koleff, P., Gaston, K.J. & Lennon, J.J. (2003) Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, **72**, 367-382.

Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966-979.

Lyashevska, O. & Farnsworth, K.D. (2012) How many dimensions of biodiversity do we need? *Ecological Indicators*, **18**, 485-492.

Molinos, J.G., Halpern, B.S., Schoeman, D.S., Brown, C.J., Kiessling, W., Moore, P.J., Pandolfi, J.M., Poloczanska, E.S., Richardson, A.J. & Burrows, M.T. (2016) Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, **6**,83–88.

Nekola, J.C. & McGill, B.J. (2014) Scale dependency in the functional form of the distance decay relationship. *Ecography*, **37**, 309–320.

Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Neumann, J. von & Morgenstern, O. (1953) *Theory of Games and Economic Behavior*. Princeton, NJ. Princeton University Press.

Newbold, T., Hudson, L.N., Hill, S.L.L., Gray, S.C.C.L., Scharlemann, J.P.W., Börger, L., Phillips, H.R.P., Douglas Sheil, D., Lysenko, I. & Purvis, A. (2016) Global patterns of terrestrial assemblage turnover within and among land uses. *Ecography*, **39**, 1151–1163.

Nowakowski, A.J., Frishkoff, L.O., Thompson, M.E, Smith, T.M. & Todd, B.D. (2018) Phylogenetic homogenization of amphibian assemblages in human-altered habitats across the globe. *Proceedings of the National Academy of Sciences*, **115**, 3454-3462.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Orlitsky, A. (2003) Information Theory. In: *Encyclopedia of Physical Science and Technology* (ed. by R.A. Meyers), pp. 751-769. Elsevier Science Ltd., Academic Press.

Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society*, **28**, 65-82.

Patterson, B.D. (1984) Mammalian extinction and biogeography in the Southern Rocky Mountains. In: *Extinctions* (ed. By M.H. Nitecki), pp. 247-293. University of Chicago Press, Chicago.

Podani, J. and Schmera, D. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence-absence data. *Oikos*, **120**, 1625-1638.

Qian, H., Badgley, C. & Fox, D.L. (2009) The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. *Global Ecology and Biogeography*, **18**, 111–122.

Raup, D.M. & Crick, R.E. (1979) Measurement of faunal similarity in paleontology. *Journal of Paleontology*, **53**, 1213-1227.

Rocha, L.A., Pinheiro, H.T, Shepherd, B., Papastamatiou, Y.P., Luiz, O.J., Pyle, R.L. & Bongaerts, P. (2018) Mesophotic coral ecosystems are threatened and ecologically distinct from shallow water reefs. *Science*, **361**, 281–284.

Ruggiero, A., Lawton, J.H. & Blackburn, T.M. (1998) The geographic ranges of mammalian species in South America: spatial patterns in environmental resistance and anisotropy. *Journal of Biogeography*, **25**, 1093-1103.

Schluter, D. & Ricklefs, R.E. (1993) Species diversity: an introduction to the problem. In: *Species diversity in ecological communities: Historical and Geographical Perspectives* (ed. by R.E. Ricklefs and D. Schluter), pp. 1–12. Chicago University Press, Chicago.

Schmera, D. & Podani, J. (2011) Comments on separating components of beta-diversity. Community Ecology, **12**, 153-160.

Schmera, D., Podani, J. and Legendre, P. 2020. What do beta diversity components reveal from presence-absence community data? Let us connect every indicator to an indicandum! Ecological Indicators 117, 106540.

Shade, A., McManus, P.S. & Handelsman, J. (2013) Unexpected diversity during community succession in the apple flower microbiome. *mBio* **4**(2):e00602-12. doi:10.1128/mBio.00602-12.

Simpson, G.G. (1943) Mammals and the nature of continents. *American Journal of Science*, **241**, 1-31.

Šizling, A.L., Pokorný, P., Juřičková, L., Horáčková, J., Abraham, V., Šizlingová, E., Ložek, V., Tjørve, E., Tjørve, K.M.C. & Kunin, W. (2016) Can people change the ecological rules that appear general across space? *Global Ecology and Biogeography*, **25**, 1072-1084.

Šizling, A.L., Šizlingová, E., Storch, D., Reif, J. & Gaston, K.J. (2009) Rarity, commonness, and the contribution of individual species to species richness patterns. *The American Naturalist*, **174**, 82-93.

Sørensen, T.A. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab Biologiske Skrifter,* **5**, 1–34.

Tjørve, E., Tjørve, K.M.C., Šizlingová, E. & Šizling, A. L. (2018) Great theories of species diversity in space and why they were forgotten: The beginnings of a spatial ecology and the Nordic early 20th-century botanists. *Journal of Biogeography*, **45**, 530-540.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tuomisto, H. (2010) A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, **33**, 23-45.

Ulrich, W. & Gotelli, N.J. (2007) Disentangling community patterns of nestedness and species co-occurrence. *Oikos*, **116**, 2053-2061.

Ulrich, W. & Gotelli, N.J. (2013) Pattern detection in null model analysis. *Oikos*, **122**, 2-18.

Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3-17.

Ulrich, W., Baselga, A., Kusumoto, B., Shiono, T., Tuomisto, H., & Kubota, Y. (2017) The tangled link between β- and ϒ-diversity: a Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. *Global Ecology and Biogeography*, **26**, 1–5.

Ulrich, W., Kubota, Y., Kusumoto, B., Baselga, A., Tuomisto, H. & Gotelli, N.J. (2018) Species richness correlates of raw and standardized co-occurrence metrics. *Global Ecology and Biogeography*, **27**, 395–399.

Vellend, M. (2001) Do commonly used indices of β-diversity measure species turnover? *Journal of Vegetation Science*, **12**, 545-552.

Villéger, S., Grenouillet, G. & Brosse. S. (2013) Decomposing functional β-diversity reveals that low functional β-diversity is driven by low functional turnover in European fish assemblages. *Global Ecology and Biogeography,* **22**, 671–681.

Whittaker, R.H. (1960) Vegetation of the Siskiyou mountains, Oregon and California. *Ecological monographs*, **30**, 279-338.

Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273-309.

Wilson, M.V. & Shmida, A. (1984) Measuring beta diversity with presence-absence data. *The Journal of Ecology*, **72**, 1055-1064.

Wright, D.H. & Reeves, J.H. (1992) On the meaning and measurement of nestedness of species assemblages. *Oecologia*, **92**, 416-428.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: The frequently used pairwise indices. S_X and S_Y are species richness of the assemblages to compare , and $S_{X\cap Y}$ is the number of shared species between them. The labelling follows Gaston *et al*. (2007) and the symbol '$\stackrel{\text{def}}{=}$' stays for definitions. For complete list of indices and their mutual relationships see SI.1. Asterix (*) labels indices that capture none of the defined phenomena.

Definition	Name and Source	Phenomenon
$J \stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	Jaccard index, Jaccard similarity, Jaccard (1912)	reversed turnover, co-occurrence
$\beta_G = 1 - J$	Jaccard dissimilarity, Gaston *et al*. (2001), labeled as in Gaston *et al*. (2007)	turnover, reversed co-occurrence
$\beta_{Sim} \stackrel{\text{def}}{=} \frac{\min(S_X, S_Y) - S_{X\cap Y}}{\min(S_X, S_Y)}$	Simpson index, Simpson dissimilarity, Simpson (1943)	reversed nestedness
$\beta'_{Sim} \stackrel{\text{def}}{=} 1 - \beta_{Sim} = \frac{S_{X\cap Y}}{\min(S_X, S_Y)}$	Simpson similarity, Simpson nestedness, Lennon *et al.* (2001)	Nestedness
$R_{ps} \stackrel{\text{def}}{=} 2\frac{\min(S_X, S_Y) - S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	‚Replacement' (turnover) by Podani & Shmera (2011)	(*)
$\beta_{nps} \stackrel{\text{def}}{=} \frac{S_{X\cap Y} + \lvert S_X - S_Y \rvert}{S_X + S_Y - S_{X\cap Y}}$ if $S_{X\cap Y} \neq 0$ else $\beta_{nps} \stackrel{\text{def}}{=} 0$.	‚Nestedness' by Podani & Shmera (2011)	(*)
$\beta_{Sør} \stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y}$	Sørensen similarity, Sørensen (1948)	reversed turnover, Co-occurrence
$\beta_t \stackrel{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y} = 1 - \beta_{Sør}$	Wilson & Shmida (1984), labelled as in Gaston *et al*. (2007), Baselga (2010) calls this index Sørensen dissimilarity	Turnover
$\beta_{gl} \stackrel{\text{def}}{=} 2\frac{\lvert S_X - S_Y \rvert}{S_X + S_Y}$	Gaston & Lennon index of Species-Richness gradient, Lennon *et al.* (2001)	species-richness gradient reversed uniformity
$\beta_{sne} \stackrel{\text{def}}{=} \beta_{NES} \stackrel{\text{def}}{=} \beta_t - \beta_{Sim}$	'Nestedness-resultant component' of Sørensen dissimilarity Baselga (2010a,2012)	(*)
$\beta_{jtu} \stackrel{\text{def}}{=} \frac{2\min(S_X, S_Y) - 2S_{X\cap Y}}{2\min(S_X, S_Y) - S_{X\cap Y}}$	Turnover component of Jaccard dissimilarity Baselga (2012)	(*)
$\beta_{jne} \stackrel{\text{def}}{=} \beta_G - \beta_{jtu}$	Nestedness resultant component of Jaccard dissimilarity, Baselga (2012)	(*)
$N_{NODF} = \beta'_{Sim}$ if $S_X \neq S_Y$, else $N_{NODF} = 0$	Nestedness by Almeida-Neto *et al*. (2008)	(*)
$R \stackrel{\text{def}}{=} \frac{\min(S_X, S_Y)}{\max(S_X, S_Y)} = \frac{2 - \beta_{gl}}{2 + \beta_{gl}}$	Newbold *et al.* 2016	species-richness uniformity reversed gradient
$N_0 \stackrel{\text{def}}{=} \frac{\min(S_X, S_Y) - S_{X\cap Y}}{S_X + S_Y}$ if $S_X \neq S_Y$ else $N_0 \stackrel{\text{def}}{=} 0$.	Nestedness by Patterson & Atmar (1986)	(*)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2: Table of possible inference from variation of pairwise indices along temporal, geographic, or environmental gradients. Inference can be obtained from the comparison of the properties that vary (significant) with the properties which are constant (insignificant) along a given gradient. Some inference can be based on variation of a single index, regardless of other indices (first three rows in the table), but more specific inference follows from a situation when one index significantly changes, while another index reveals only small variation (all remaining rows).

Variation along a gradient		**Inference**
Yes (Significant)	**No (Insignificant)**	
J		Change in turnover between regions/periods, change in co-occurrence, change in proportion of common species relative to alpha diversity
β'_{Sim}		Change in nestedness compared between regions/periods (i.e. nestedness for region 1 vs nestedness for region 2), change in proportion of common species relative to richness of species poorer sites (this change may or may not be induced only by change in the list of species of species-richer sites)
R		Change in species richness gradient (uniformity) between regions/periods; change in spatial variation of alpha diversity
β'_{Sim}	S_{rich}	Change in nestedess between regions/periods is driven by exchange of species at species poor sites
$\langle S \rangle$	S_{rich}	Difference of alpha diversity between regions/periods is driven by sites with fewer species
$\langle S \rangle$	S_{poor}	Difference of alpha diversity between regions/periods is driven by species-richer sites
R	$\langle S \rangle$	Regions/periods differ in spatial variation of alpha diversity without change of its average value
β'_{Sim}	J	Change in nestedness but not turnover/co-occurrence; Simultaneous change in lists of species that are unique to species-richer and species-poorer sites; variation in β'_{Sim}, no change in J, and Eq. 2 induce necessarily a change in R
J	β'_{Sim}, S_{poor}	Change in turnover/co-occurrence but not nestedness; change in list of species that are unique to species-richer sites; necessarily a change in R

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

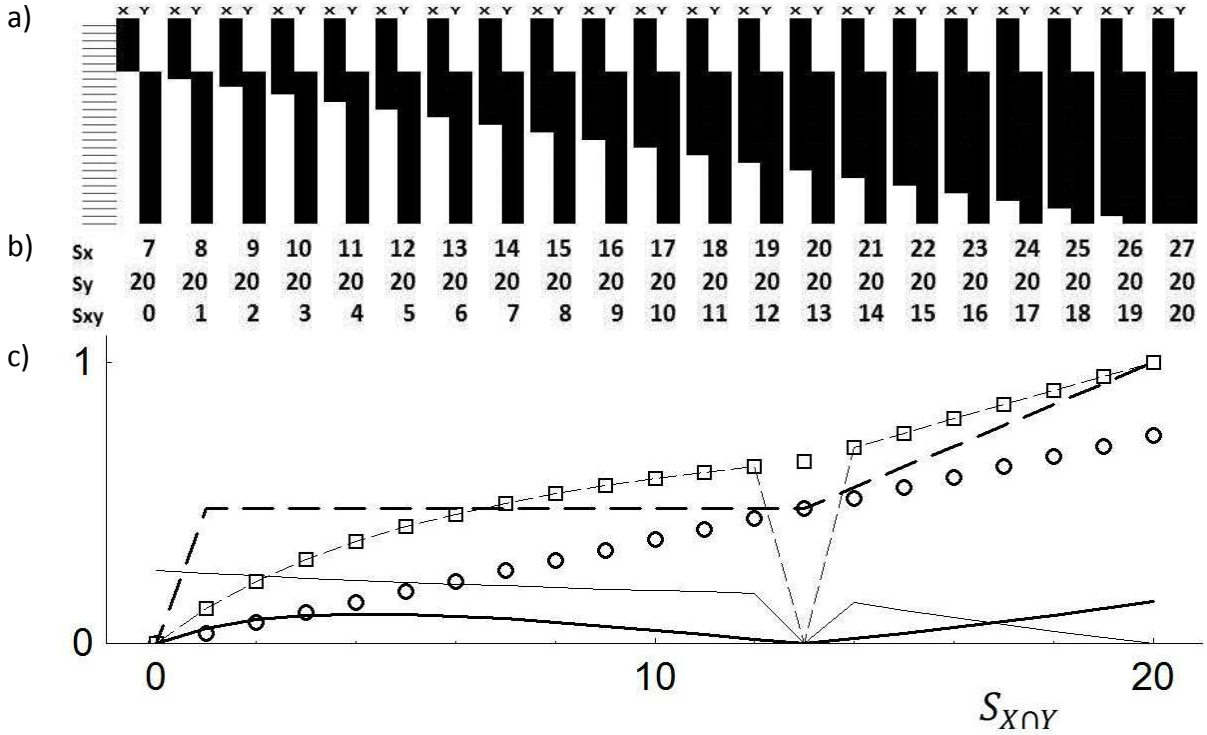


Fig. 1: The requirement of continuity (see Section Problems), and consequences of its violation. The top plot (a) shows a sequence of 20 pairs of assemblages X and Y, with a continuous change of the arrangement of assemblages, increasing similarity and nestedness from left to right, and constant total species richness. Columns show assemblages, lines species and blackened cells presences of species within assemblages. First two lines in (b) show numbers of species in the left (Sx) and right (Sy) assembages for each situation along the gradient. The third line in (b) shows $S_{X\cap Y}$ labeled as Sxy. The most bottom plot (c) shows values of β'_{Sim} (squares), J (circles), β_{sne} (solid bold line), N_{NODF} (dashed thin line), β_{nps} (dashed bold line), and No (solid thin line).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

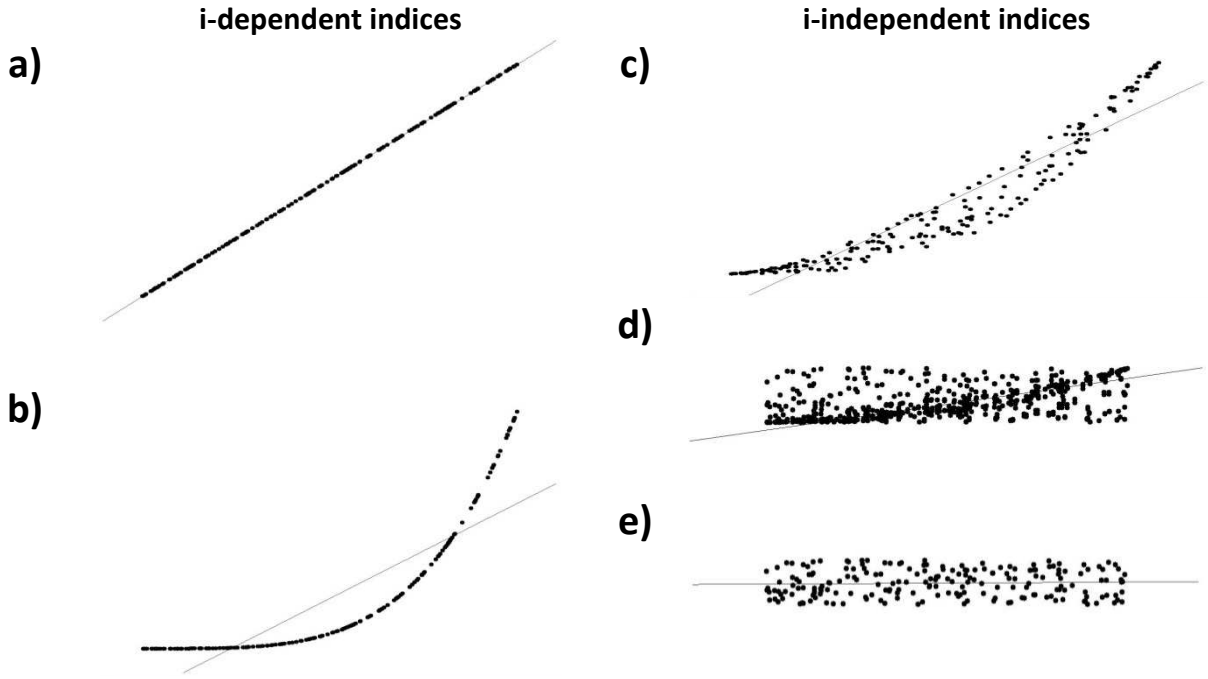


Fig. 2: An example of relationships between two i-dependent (a,b), i-independent (c-e), s-dependent (a-d), s-independent (e), related sensu Chao *et al*. (2012) (a-c) and unrelated (d,e) indices. Dots show 200 samples from the underlying relationships and lines show linear regressions. I-dependent indices (a,b) carry equal information because the strictly increasing or decreasing relationship between them uniquely transforms one index to another. The relatedness (Chao *et al.*, 2012) is one of the mechanisms that drive s-dependence, and i-dependence is a special case of relatedness where both upper and bottom constraints are identical (a,b). The existence of identical constraints must be proven by means of mathematical analysis using the formulae that define the indices for each particular case separately (SI.2.T19). Anyway, there are apparently no one-to-one (strictly increasing nor decreasing) continuous lines beyond the samples of i-independent indices (c-e) (which does not require analytical proof). Only the indices in panel a) scale in a linear manner. The indices in panel b) are continuous (non-linear) transformations of each other. This transformation is called homeomorphism. Indices in panels c-e can also be viewed as transformations of each other, but this transformation is not a homeomorphism and thus each index may or may not measure different phenomena, depending on the constraints (Fig. 3). Degree of s-dependence measured by Pearson correlation coefficient decreases from a) to e) (1; 0.95; 0.86; 0.73; 0; N=200). The i-dependences and i-independences are equal for a-b, and c-e, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

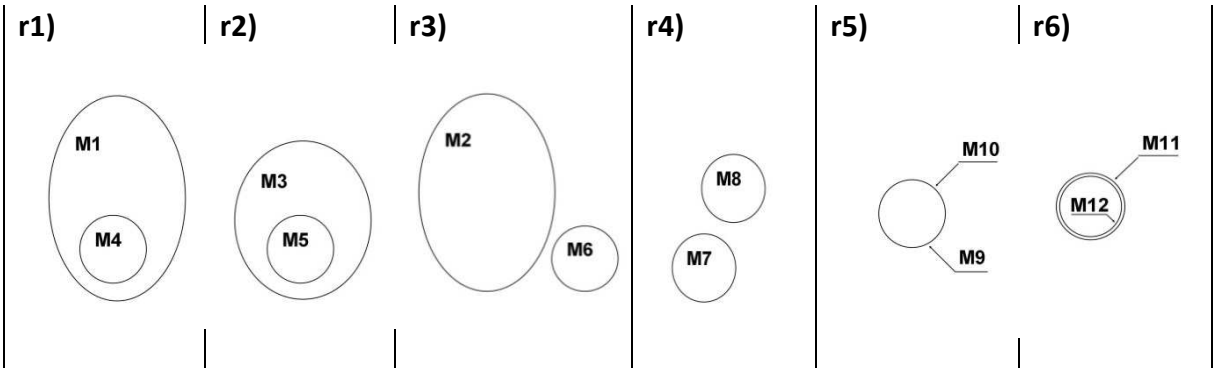


Fig. 3: Constraints by the five spatial phenomena on the indices. Various re-arrangements of two sets in the Venn diagrams define differences between the phenomena. The scheme equally applies to re-arrangements of two lists of species and two species spatial ranges, for the sets that would represent either lists of species or species spatial ranges, respectively. The sizes of the sets follow M1 = M2 > M3 > M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11>M12. The sets M9 and M10 are identical. The constraints that define the five spatial phenomena are as follows:

(Nestedness) $\mathrm{Min} = \mathrm{Nest[r4]} = \mathrm{Nest[r3]} < \mathrm{Nest[r1]} = \mathrm{Nest[r2]} = \mathrm{Nest[r6]} = \mathrm{Max} =_{(1)} \mathrm{Nest[r5]}$;
(Co-occurrence) $\mathrm{Min} = \mathrm{Co[r4]} = \mathrm{Co[r3]} < \mathrm{Co[r1]} < \mathrm{Co[r2]} < \mathrm{Co[r6]} < \mathrm{Co[r5]} = \mathrm{Max}$;
(β-diversity) $\mathrm{Min} = \mathrm{BD[r5]} < \mathrm{BD[r6]} < \mathrm{BD[r2]} < \mathrm{BD[r1]}$;
(Turnover) $\mathrm{Min} = \mathrm{Turn[r5]} < \mathrm{Turn[r6]} < \mathrm{Turn[r2]} < \mathrm{Turn[r1]} < \mathrm{Turn[r4]} \leq \mathrm{Turn[r3]}$; and
(sr-gradient) $\mathrm{Min} = \mathrm{SRG[r5]} = \mathrm{SRG[r4]} < \mathrm{SRG[r6]} < \mathrm{SRG[r2]} < \mathrm{SRG[r1]} = \mathrm{SRG[r3]}$.

Where '*Min*' and '*Max*' label the extreme values of the indices (usually *Min*=0 and *Max*=1) and the letters in brackets refer to the re-arrangements. The equation $=_{(1)}$ is often violated, and nestedness of r5 is in some indices defined as zero, which is the violation of continuity that results in problems.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

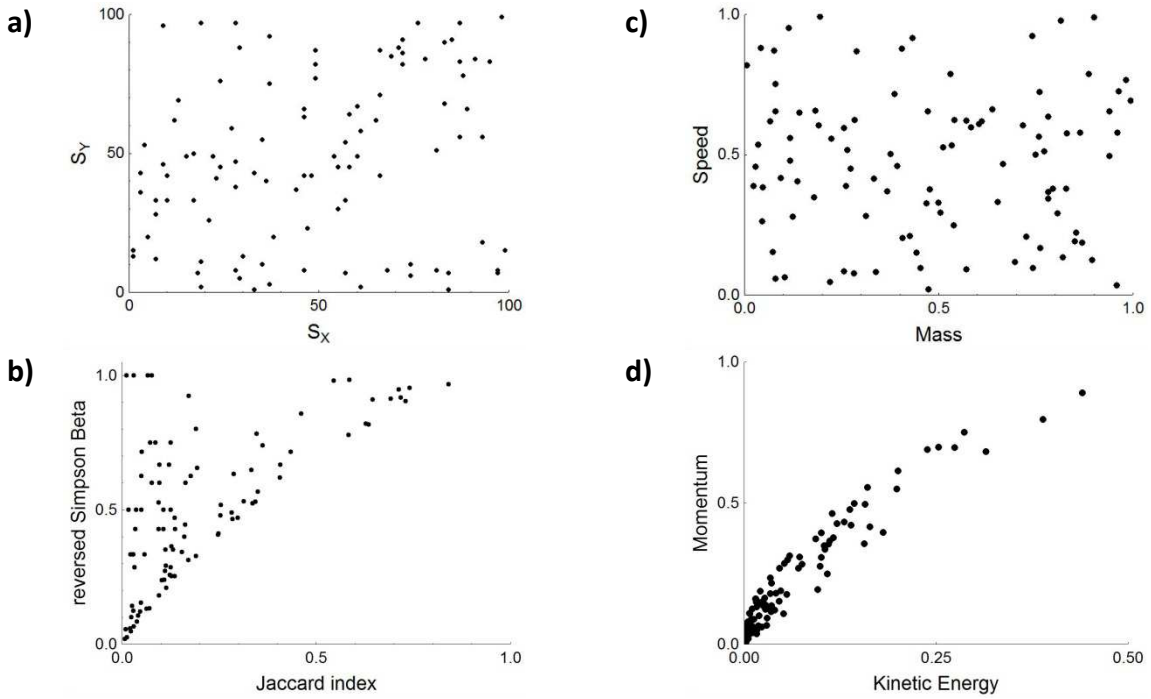


Fig. 4: Statistical and information (in)dependence in ecology (a, b) and physics (c, d). Mutually statistically independent (s-independent) values for species assemblages in ecology (S_X and S_Y; **a**), and moving objects in physics (mass and speed; **c**) do not necessarily result in statistically independent indices (J and β'_{Sim}; **b**) and variables (p and E_k; **d**). This is despite J, β'_{Sim}, p, and E_k carry a specific information and therefore are independent (i-independent) regarding their information value. Here S_X, S_Y, ($\in \{1, \cdots, 100\}$) and $S_{X \cap Y}$ ($\leq \min(S_X, S_Y)$) are drawn at random from a uniform distribution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

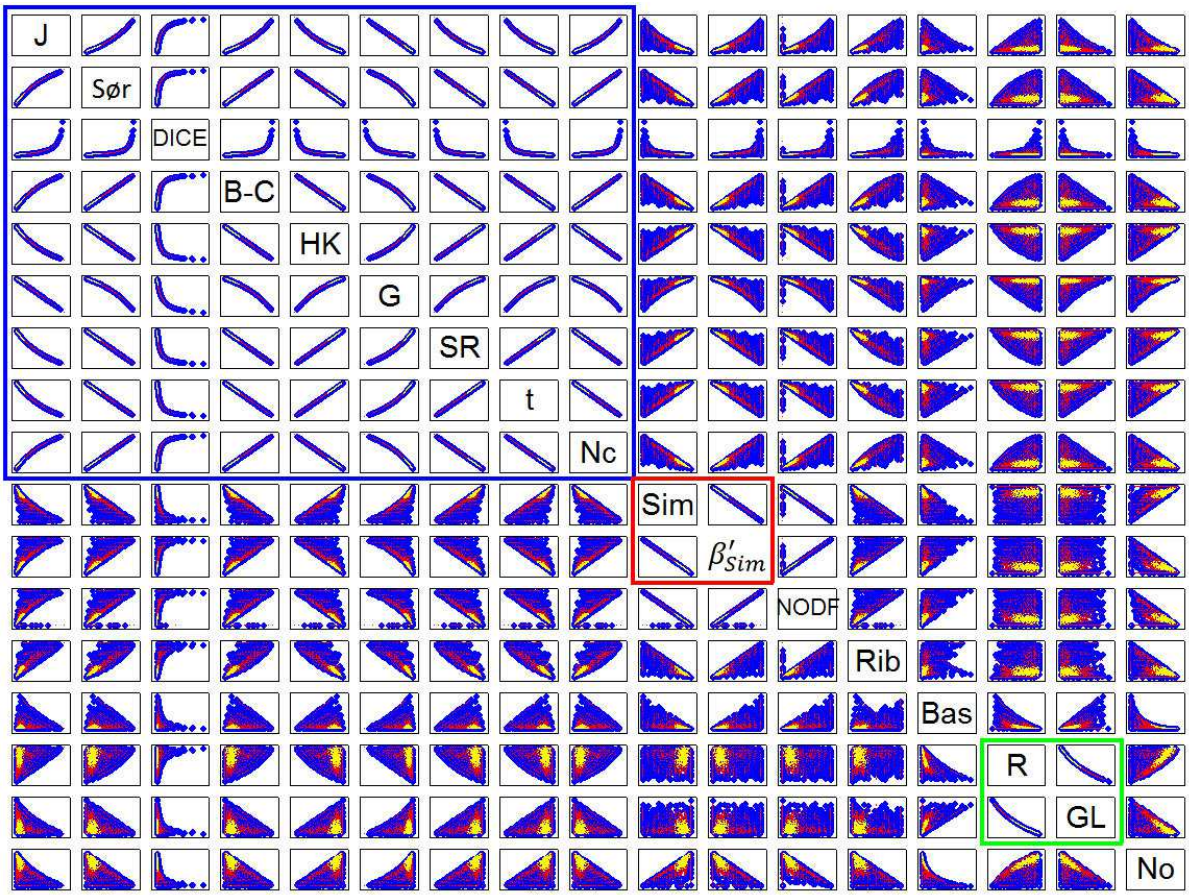


Fig. 5: Relationships between the indices for pairs of simulated random (blue dots, N=1352) and observed (red – plants, N=946; yellow – Ice Shield microbes, N=406) assemblages (see SI.7 data sources). The species richness values of assemblages in each pair are uncorrelated, i.e., numbers $S_{X\cap Y}$, $b = S_X - S_{X\cap Y}$ and $c = S_Y - S_{X\cap Y}$ vary between 0 and 1, they are drawn from a uniform distribution and are mutually s-independent. In N_0 and N_{NODF} (i.e., where the equality between S_X and S_Y affects the result) species richness is a random integer between 1 and 20 species. Where possible the notation is adapted from Gaston *et al*. (2007). The notation in the plot is simplified. From up to down: J – Tab. 1, $\beta_{Sør}$ – Tab. 1, β_{DICE} - Raup & Crick (1979), β_{BC} – Bray & Curtis (1957) (computed for incidences), β_{HK} – Harte & Kinzig (1997), β_G – Tab. 1, β_{SR} – Schluter & Ricklefs (1993), β_t – Tab. 1, N_C - Wright & Reeves (1992) standardized as in Gotelli & McCabe (2002), β_{Sim} – Tab. 1, β'_{Sim} - Tab. 1, it equals *rN* (range nestedness, Šizling *et al*., 2009) if applied to ranges, β_{rib} – Ruggiero *et al*. (1998), $\beta_{sne} \stackrel{\text{def}}{=} \beta_{BAS}$, Tab. 1, *R* – Tab. 1; β_{gl} – Tab. 1, N_0 - Patterson & Atmar (1986), N_{NODF} – Tab. 1. Blue, red, yellow and green rectangles delimit the families of Jaccard similarity, Simpson nestedness, and species-richness gradient, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

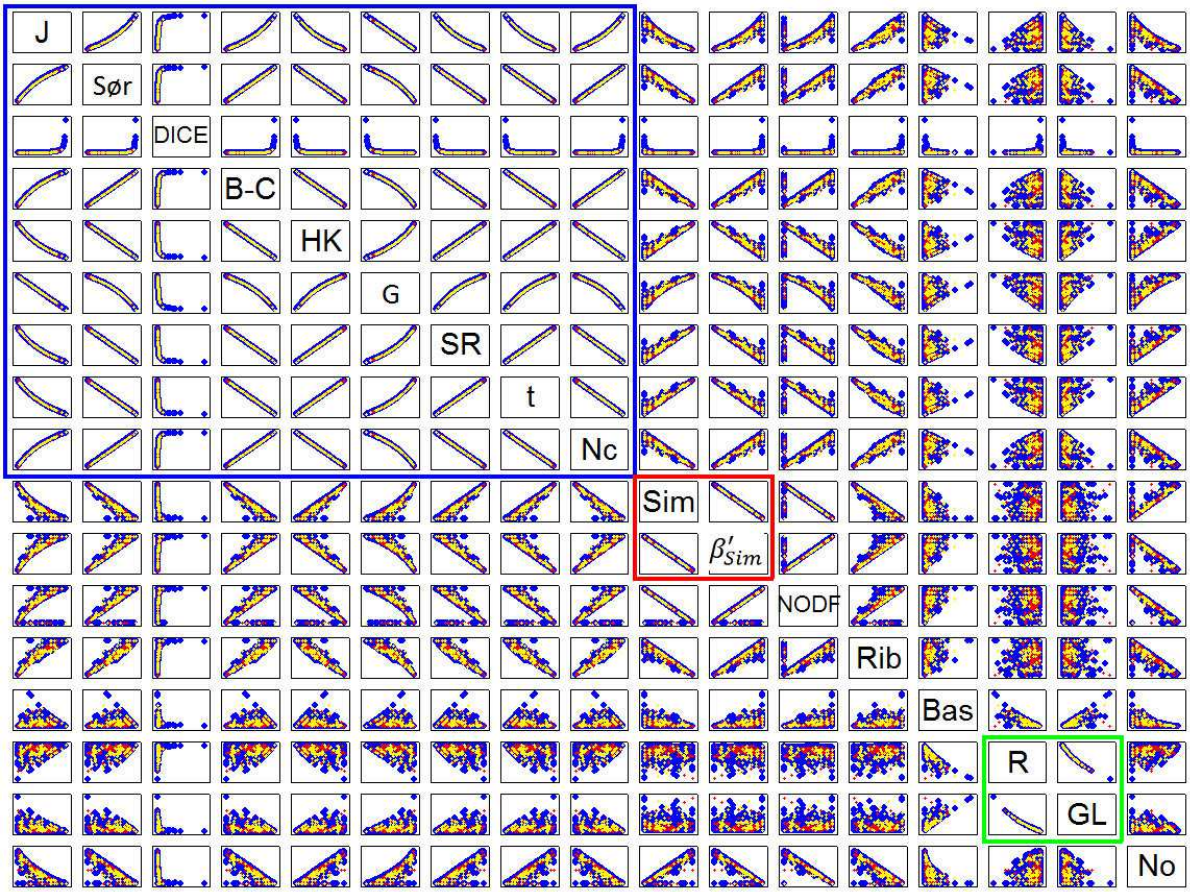


Fig. 6: Relationships between pairwise indices of assemblages which have low (20%) species richness variation (simulations) or which are spatially close to each other (observed data). Blue (N=1254) and yellow (N=98) dots show data generated by the simulations of random assemblages, and red dots show data for the plant and Ice-Shield microbe assemblages that are closer than 0.5 km from each other (N=98) (for details on data see SI.7). The cluster of yellow dots is a random subsample of the blue dots, with N equal to the observation to allow comparison. In accord with the theory, when there is low variation in species richness or when the assemblages are close to each other, the Jaccard-index, Simpson-beta, Ruggiero-beta, and N_0 families collapse together into one big family of almost i-dependent (and definitely s-dependent) indices. For the meaning of abbreviations see the legend to Fig. 5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

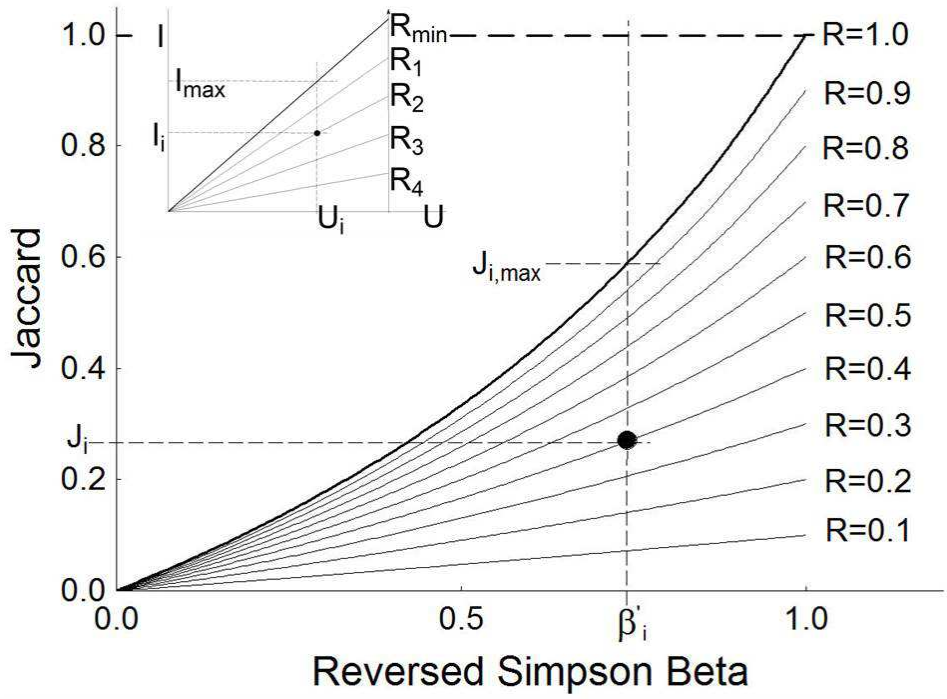


Fig. 7: The reason why mutual dependence of indices cannot be eliminated to purify individual effects. The relationship between Jaccard similarity (J) and Simpson nestedness (β'_{Sim}) is determined by species richness uniformity R (Eq. 6), each line shows points with equal R (the dot has R=0.4). J is limited by 1 (dashed bold line) and each particular J_i is limited by $J_{i,max}$, which is a function of β'_{Sim} (bold line, Eq. 6). The value of J_i relative to the limit of one should be read as similarity (or reversed turnover) without a need for any partitioning of J or β'_{Sim} into components. We say that J is a similarity *per se* (or reversed turnover *per se*). The value J_i rescaled by $J_{i,max}$ shows R_i, and should be read as evenness *per se* (or reversed species richness gradient *per se*). The fact that *R* affects *J* is a matter of i-dependence between triplet J,β'_{Sim},*R*, and thus the interplay between *J* and *R* should not invalidate their unique meaning determined by the constraints (section 'Constraints needed to define the phenomena'). This is somehow similar (inset) to the Ohm's law in physics where the triplet of voltage (*U*), current (*I*) and resistance (*R*) are mutually dependent (*U=RI*; see inset). Each *I* is therefore affected by *R* (R_{1-4}, for the dot $R = R_2$) and has its upper limit given by technological limits to produce a component of low resistance (R_{min}, bold line). Most importantly, the understanding in physics is that I_i is the current even though it also scales with resistance. No one would say that partitioning U minus I is the 'voltage resultant component'. In the same way, J_i is reversed turnover per se even though it also scales with reversed richness gradient *per se*. For the same reason β'_{Sim} is a measure of nestedness per se, without a need for partitioning into components of turnover and/or nestedness.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Box 1: *How many indices do we need?*

For a given system, multitude of indices can be calculated. What is the minimum number of indices that characterize a given system fully? Here we show that three i-independent indices (including species richness) are sufficient for full characterization of the system, so that any other index can be reconstructed using this information. We adopt the term of independence from mathematics, which is independence of equations in terms of their solutions (hereafter i-independence). For this purpose, we need to see definitions of the indices as equations to solve. This allows for two statements: (i) all the focal indices listed in SI.1 are i-dependent on the pair of Jaccard index, *J* (Tab. 1) and Simpson nestedness, β'_{Sim} (Tab. 1); and (ii) all information concerning the difference between two assemblages is captured by *J*, β'_{Sim}, and an index that is i-dependent on species richness (e.g. species richness itself). Here we show why these statements hold.

All the focal indices of diversity (SI.1) can be defined as a ratio of two linear functions

$$I \stackrel{\text{def}}{=} \frac{k_1 S_{X\cap Y}+k_2 S_X+k_3 S_Y}{l_1 S_{X\cap Y}+l_2 S_X+l_3 S_Y}, \quad (k_{1-3},\ l_{1-3} \in \boldsymbol{R}),\ \ |l_2|+|l_3|>0\,, S_X \geq S_Y > 0. \qquad \text{(Eq. B1)}$$

For *J* and β'_{Sim}, $\{k_{1-3}, l_{1-3}\}$={1,0,0,-1,1,1} and {1,0,0,0,0,1}, respectively. Following the logic: $I=\frac{a}{b} \ \rightarrow bI-a=0$, Eq. B1 can be converted into linear equation

$$S_X(Il_2-k_2)+S_Y(Il_3-k_3)+S_{X\cap Y}(Il_1-k_1)=0, \qquad \text{(Eq. B2)}$$

where S_X, S_Y, $S_{X\cap Y}$ are the unknowns and the *I* is a particular value of an index. Eq. B2 turns into

$$S_X J+S_Y J-S_{X\cap Y}(J+1)=0 \text{ and } S_Y\beta'_{Sim}-S_{X\cap Y}=0 \text{ or } S_X R-S_Y=0 \qquad \text{(Eq. B3)}$$

for *J* and β'_{Sim}, respectively. The two equations for two indices (Eq. B3), and the three unknown variables (S_X, S_Y, $S_{X\cap Y}$), do not provide a unique solution, and have zero at the right side. If the third equation had zero on the right side, the system would either provide multiple solutions or the only solution would be zero ($S_X, S_Y, S_{X\cap Y}=0,0,0$). We therefore need an independent equation with nonzero right side to get unique information on S_X, S_Y, $S_{X\cap Y}$. This equation is

$$S_X+S_Y=2\langle S\rangle, \qquad \text{(Eq. B4)}$$

where $\langle S\rangle \neq 0$ is the expected species richness of one assemblage, that is, the mean richness (for the proof that Eq. B4 is independent of equations B3, see SI.2.T1). The three equations (Eqs. B3,B4) determine $S_X, S_Y, S_{X\cap Y}$ uniquely, and thus no other index (even if its definition does not follow B1) can further specify the solution. Moreover, if we restrict our statements on the existing indices which can be expressed using Eq. B1, we can compute the value of any other index from the values of *J* and β'_{Sim} as

$$I=\frac{J(\beta'_{Sim}(k_1+k_2)-k_2+k_3)+\beta'_{Sim}k_2}{J(\beta'_{Sim}(l_1+l_2)-l_2+l_3)+\beta'_{Sim}l_2} \text{ for } J\neq 0 \text{ else } I=\frac{k_2+k_3R}{l_2+l_3R} \qquad \text{(Eq. B5)}$$

where k_i, l_i define the new index *I* (for the proof see SI3.T16)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Box 2: *How to unify indices from across the literature*

A large number of different indices is spread over the literature (e.g., Gaston *et al.* 2007; Baselga, 2012; Podani & Schmera, 2020). However, if one wishes to do a meta-analysis and compare indices from different published sources, one needs to make the indices comparable, and to convert them to only one or two reference indices. Here we show how this can be done, based on the idea that an index is defined by an equation. Firstly, we need a publication using at least two i-independent indices, I_1, I_2, that are expressed as Eq. B1 (see Box 1). Then we write a system of three independent equations: two equations B2, each for one of the indices, and one scaling equation Eq. B3. If $\langle S \rangle$ is missing from the publication, we can put $\langle S \rangle = 1$ without loss of generality. The reason is that the indices to be converted (Eq. B1) are i-independent of species richness and therefore the exact value of $\langle S \rangle$ does not matter in the case. The third equation then is

$$S_X + S_Y = 2. \qquad \text{(B7)}$$

The solution of the three equations (See SI.2.T18 for the guide how to solve them) is S_{Xcomp}, S_{Ycomp} and $S_{X \cap Ycomp}$. This solution is different from the solution based on the original data behind the published source, but we can get the original average values S_{Xorig}, S_{Yorig}, $S_{X \cap Yorig}$ by simple rescaling of the computed values ($S_{Xorig} = \langle S \rangle S_{Xcomp}$ and so on), in case we know $\langle S \rangle$. From S_{Xcomp}, S_{Ycomp}, $S_{X \cap Ycomp}$ we can compute any index that is i-independent of species richness, even an index that cannot be expressed as Eq. B1 (e.g.,

$I = S_X S_{X \cap Y} / S_Y^2$).

A special case are the indices that originated from additive partitioning ($I_1 = I_a - I_b$) such as β_{sne}, β_{jne}, of β_{nps} (Tab. 1). These indices cannot be universally expressed as Eq. B1, and the equations to solve are no longer linear. In the case of additive partitioning, we can compute I_b from two i-independent indices I_1 and I_2 using equation B8

$$I_b^2 D_{LL} - I_b (I_1 D_{LL} - D_{KL} + D_{LK}) + I_1 D_{LK} - D_{KK} = 0 \qquad \text{(B8)}$$

where D_{XY} are determinants listed in SI.2.T20 (see also SI.2.T20 for derivation). I_a then follows $I_1 + I_b$. If I_a, I_2, or I_b, I_2, or I_a, I_b are i-independent then we can follow the above algorithm that uses linear equations, and compute any even not yet invented index. However, as Eq. B8 may have two realistic solutions (both solutions are within miminum and maximum possible range of I_b), the partitioning often leads to the loss of information. An example is $\beta_{sne} = 0.15$, (Baselga, 2012; Tab. 1) and $\beta_{nps} \cong 0.714$ (Podani & Schmera, 2011; Tab. 1) ($\langle S \rangle = 25$) which produce $J \cong 0.43$, ($S_{X \cap Y} = 15$, $S_Y = 20$, $S_X = 30$), or $J \cong 0.04$ ($S_{X \cap Y} \cong 2.08$, $S_Y \cong 8.93$, $S_X \cong 41.07$) (for details See SI.6, last panel). As difference between $J = 0.43$ and $J = 0.04$ is considerable, β_{sne} and β_{nps} provide less information than, e.g. J and β_{Sim}, and any inference from them is ambiguous.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Box. 3: Turnover and nestedness are incomparable

There are three main reasons for incomparability of nestedness and turnover. First, dimensionless indices are universally incomparable, which can be demonstrated as follows: let's assume an index of turnover $I_T = 0.5$ and an index of nestedness $I_N = 0.4$. The temptation to claim that the nestedness is lower than turnover collides with this simple logic: I_T^2 is a strictly increasing transformation of I_T, and thus also measures turnover; however, $I_T^2 = 0.25$ is also a dimensionless index and suggests smaller turnover than nestedness, which contradicts the initial conclusion. Comparison and subtraction of dimensionless indices is thus mathematically problematic, and is only possible under some circumstances. Second, Eq. 2 implies that we cannot put an equal (nor unequality) sign between the indices of nestedness and turnover without taking into account the effect of the richness gradient (Schmera & Podani, 2012). Similarly, no one would directly compare U with I in Ohm's law, $U = RI$ (Fig. 7). Third, nestedness and turnover differ in their constraints (Fig. 3), and are thus distinct phenomena. In the exact sciences, we cannot compare measures of different phenomena even if they had equal units. The example from phisics is 'work' and 'torque' that are both computed as a product of force (F) and distance (r or d) ($W = F \cdot d$ and $\tau = F \cdot r$, in scalar form) and have equal units ($\mathrm{N} \cdot \mathrm{m}$), but differ in the direction of the acting force (they are different phenomena) and cannot thus be compared. Different phenomena can only be compared using their consequences, if these consequences are comparable.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information

Mathematically and biologically consistent framework for presence-absence indices of diversity

Content:

SI.1 Scaling properties **2-5**
SI.2 Theses and proofs **6-12**
SI.3 (Relationships between indices of nestedness) **13-14**
SI.4 Relationship between pairwise indices of interrelated assemblages **15**
SI.5 i-independence of and s-dependence on species richness **16**
SI.6 Software Calculator (manual) **17-19**
SI.7 Datasets **20**

SI.1: Scaling properties: The table shows the definitions of the diversity indices and their mutual scaling. The indices by the equation marks refer to the derivations of the relationships (see SI.2.T19 for derivations). The equation marks without affiliation are derived by simple rearranging the equations listed in the table. All the scaling properties listed are unique except those marked '≅'. Where possible, indices are defined as in Koleff *et al*. (2003) and Gaston *et al*. (2007); see the column 'Notes' for exceptions. In agreement with Koleff *et al*. (2003), and Gaston *et al*. (2007), *a* refers to the number of species shared by the focal assemblages ($a \stackrel{\text{def}}{=} S_{X\cap Y}$), and *b* and *c* refer to the number of species that belong only to the first and second assemblages ($b \stackrel{\text{def}}{=} S_X - S_{X\cap Y}$ and $c \stackrel{\text{def}}{=} S_Y - S_{X\cap Y}$). See the column 'Notes' for further references.

Notation	Definition	scaling properties	Notes
β_{Wh0}	$\stackrel{\text{def}}{=} \frac{S_T}{\langle S_1 \rangle}$	Gaston *et al*. (2007) call β_{Wh0} as β_{Wh2}, for the case of two assemblages; β_{Wh2} scales with indices of Jaccard-index family (SI.3.T19.4), but it has fundamentally different inferences	Overall Whittaker (1960) beta diversity For two assemblages (in this study), $S_T = S_{X\cup Y} = S_X + S_Y - S_{X\cap Y}$ and $\langle S_1 \rangle = (S_X + S_Y)/2$ Beta-diversity (SI.2.T1)
J	$\stackrel{\text{def}}{=} \frac{a}{a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	$J =_{(1)} \frac{\beta'_{Sim}}{1 + R^{-1} - \beta'_{Sim}}$	Jaccard (1912) reversed turnover or co-occurrence applied on between assemblage level (SI.2.T15)
β_{SR}	$\stackrel{\text{def}}{=} \frac{a+b+c}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X + S_Y - S_{X\cap Y}}{S_X + S_Y}$	$\beta_{SR} =_{(2)} \frac{1}{1+J}$	Schluter & Ricklefs (1993) turnover or reversed co-occurrence applied on between assemblage level (SI.2.T15)
$\beta_{Sør}$	$\stackrel{\text{def}}{=} \frac{2a}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y}$	$\beta_{Sør} =_{(3)} 2\frac{J}{1+J}$	Sørensen (1948) similarity, reversed turnover or co-occurrence applied on between assemblage level (SI.2.T15)
β_{DICE}	$\stackrel{\text{def}}{=} \frac{2a}{b+c}$ $\stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y - 2S_{X\cap Y}}$	$\beta_{DICE} =_{(17)} 2\frac{J}{1-J}$	Raup & Crick (1979) reversed turnover or co-occurrence applied on between assemblage level (SI.2.T15)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

β_{BC}	$\stackrel{\text{def}}{=} 1-\frac{a}{2a+b+c}$ $\stackrel{\text{def}}{=} 1-\frac{S_{X\cap Y}}{S_X+S_Y}$	$\beta_{BC} =_{(16)} \frac{1}{1+J}$	Bray & Curtis (1957) simplified for incidence data Turnover or reversed co-occurrence applied on between assemblage level (SI.2.T15)
β_{HK}	$\stackrel{\text{def}}{=} 1-\frac{2a}{2a+b+c}$ $\stackrel{\text{def}}{=} 1-\frac{2S_{X\cap Y}}{S_X+S_Y}$	$\beta_{HK} =_{(5)} \frac{1-J}{1+J}$	Harte & Kinzig (1997) Turnover or reversed co-occurrence applied on between assemblage level (SI.2.T15)
β_t	$\stackrel{\text{def}}{=} \frac{b+c}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X+S_Y-2S_{X\cap Y}}{S_X+S_Y}$	$\beta_t =_{(6)} \frac{1-J}{1+J}$	Wilson & Shmida (1984) in Gaston *et al.* (2007) Baselga (2010a,2012) calls this index Sørensen dissimilarity Turnover or reversed co-occurrence applied on between assemblage level (SI.2.T15)
β_G	$\stackrel{\text{def}}{=} \frac{b+c}{a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X+S_Y-2S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}}$	$\beta_G =_{(7)} 1-J$ $\beta_G =_{(6,7)} 2\frac{\beta_t}{1+\beta_t}$	Gaston *et al.* (2001) Turnover or reversed co-occurrence applied on between assemblage level (SI.2.T15)
β_{Sim}	$\stackrel{\text{def}}{=} \frac{\min(b,c)}{a+\min(b,c)}$ $\stackrel{\text{def}}{=} \frac{\min(S_X,S_Y)-S_{X\cap Y}}{\min(S_X,S_Y)}$	$\beta_{Sim} =_{(8)} 1-\beta'_{Sim}$	Simpson beta Simpson (1943); Lennon *et al.* (2001) reversed nestedness (SI.2.T4)
β'_{Sim} (rN)	$\stackrel{\text{def}}{=} \frac{a}{a+\min(b,c)}$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{\min(S_X,S_Y)}$	$\beta'_{Sim} =_{(8)} 1-\beta_{Sim}$, $\beta'_{Sim} =_{(9)} \frac{1+R}{R}\frac{J}{1+J}$	Šizling *et al.* (2009,2016) β'_{Sim} equals rN computed for assemblages Nestedness
β_{gl}	$\stackrel{\text{def}}{=} 2\frac{\lvert b-c\rvert}{2a+b+c}$ $\stackrel{\text{def}}{=} 2\frac{\lvert S_X-S_Y\rvert}{S_X+S_Y}$	$\beta_{gl} =_{(10)} 2\frac{1-R}{1+R}$ $\frac{R}{1+R} =_{(10)} \frac{2-\beta_{gl}}{4}$	Lennon *et al.* (2001) Species-richness gradient (SI.2.T14)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

R	$\stackrel{\text{def}}{=} \frac{a+\min(b,c)}{a+\max(b,c)}$ $\stackrel{\text{def}}{=} \frac{\min(S_X,S_Y)}{\max(S_X,S_Y)}$	$R =_{(10)} \frac{2-\beta_{gl}}{2+\beta_{gl}}$ $R =_{(9)} \frac{J}{\beta'_{Sim}(1+J)-J}$	Newbold *et al*. (2016)scales in one-to-one manner with β_{gl} by Lennon *et al*. (2001); reversed sr-gradient or species-richness evenness
	$\beta_{jtu} \stackrel{\text{def}}{=} 2\frac{\min(b,c)}{a-\min(b,c)}$ $\beta_{jtu} \stackrel{\text{def}}{=} 2\frac{\min(S_X,S_Y)-S_{X\cap Y}}{2\min(S_X,S_Y)-S_{X\cap Y}}$	$\beta_{jtu} =_{(21)} 2\frac{\beta_{Sim}}{1+\beta_{Sim}}$	Turnover component of Jaccard dissimilarity Baselga (2012)
$\beta_{Sne} \stackrel{\text{def}}{=} \beta_{NES}$	$\stackrel{\text{def}}{=} \beta_t - \beta_{Sim}$	does not scale with other indices	Baselga (2012) No family under examination (SI.2.T7)
β_{rib}	$\stackrel{\text{def}}{=} \frac{a}{a+c}$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_Y}$	$\beta_{rib} = \beta'_{Sim}$ where $S_Y = \min(S_X,S_Y)$ else it does not scale with other indices	Ruggiero *et al*. (1998) Nestedness or co-occurrence depending on the direction; S_Y stands for species richness of the approached assemblage if we leave assemblage X (SI.2.T11)
$\langle\beta_{Sim}\rangle$	$\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\beta_{Sim,i}$	$\langle\beta_{Sim}\rangle =_{(11)} 1-\langle\beta'_{Sim}\rangle$; $\langle\beta_{Sim}\rangle \cong_{(12)} D$	arithmetic mean across n Simpson indices as introduced in Lennon *et al*. (2001) (a pair wise index); it is labeled simply Sim in literature (e.g., Gaston *et al*. 2007) The scaling (12) works for only assemblages with an high number of sites (≥ 10) reversed Nestedness (SI.2.T17)
$\langle\beta'_{Sim}\rangle$	$\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\beta'_{Sim,i}$	$\langle\beta'_{Sim}\rangle =_{(11)} 1-\langle\beta_{Sim}\rangle$; $\langle\beta'_{Sim}\rangle \cong_{(11+12)} 1-D$; $\langle\frac{J}{1+J}\rangle \cong_{(13)} \langle\beta'_{Sim}\rangle\langle\frac{R}{1+R}\rangle$	arithmetic mean across n rN_2 indices, as introduced in Šizling *et al* (2009) (a pairwise index) The scaling between $\langle\beta_{Sim}\rangle$ and discrepancy works only for assemblages with an high number of sites ((≥ 30) Nestedness (SI.2.T13)
D	Discrepancy	$D \cong_{(12)} \langle\beta_{Sim}\rangle =_{(11)} 1-\langle\beta'_{Sim}\rangle$	Brualdi & Sanderson (1999)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	(defined by an algorithm)		Scaling between either of these works only for assemblages with a high number of sites (≥ 10) Reversed Nestedness (SI.2.T13)
N_{NODF}	$\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} N_{NODF,i}$ $N_{NODF,i} \stackrel{\text{def}}{=} 1-\beta_{Sim,i}$ if the two assemblages vary in their species richness; otherwise $N_{NODF,i} \stackrel{\text{def}}{=} 0$.	does not scale with other indices	Almeida-Neto *et al*. (2008) No family under examination (SI.2.T8)
N_0	Number of gaps (defined by an algorithm)	does not scale with other indices	Patterson & Atmar (1986) For two assemblages (in this study), $N_0 \stackrel{\text{def}}{=} \frac{\min(b,c)}{2a+b+c}$ if the two assemblages vary in their species richness; otherwise $N_0 \stackrel{\text{def}}{=} 0$. No family under examination (SI.2.T10)
N_C	(defined by an algorithm)	$N_{C1} =_{(19)} 1-\beta_{SR} =_{(20)} \frac{\beta_{S\emptyset r}}{2} =_{(3)} \frac{J}{1+J}$ $N_{C2} \xrightarrow{\langle S\rangle\to\infty}_{(T9)} 2\beta'_{Sim}-1$	Wright & Reeves (1992), For two assemblages (in this study), $N_{C1} \stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X+S_Y}$ (standardized as in Gotelli and McCabe 2002); $N_{C2} \stackrel{\text{def}}{=} \frac{S_{X\cap Y}-E(S_{X\cap Y})}{max(S_{X\cap Y})-E(S_{X\cap Y})}$ (standardized as in Wright & Reeves 1992) N_{C1}, Jaccard index family (SI.2.T8) N_{C2} Simpson beta family for high species richness (SI.2.T9)

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SI.2: Theses and proofs

Theses T1-15 state whether or not each index satisfies the constraints of the five spatial phenomena (Fig. 3). The evidences for the theses uses three parameters a, b and c (the number of shared species, $S_{X\cap Y}$, the number of species exclusive to the first assemblage, S_X, and the number of species exclusive to the second assemblage, S_Y, respectively). Individual arrangements in Fig. 3, are characterized by $a = 0$ in arrangements r3 and r4 in Fig. 3; by $b = 0$ in arrangements r1,r2,r5 and r6 in Fig. 3; by $c = 0$ in arrangement r5 in Fig. 3 and by $c_{r6} < c_{r2} < c_{r1}$ in arrangements r1,r2,r6,.

T1: The system of Eqs. B3,B4 is mutually i-independent.

Evidence: $\begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| \; 0 \\ 1 & 1 & 0 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| \; 0 \\ 0 & 0 & J+1 & 2J\langle S\rangle \end{bmatrix}$, this matrix provides unique solution of the system. Hence $\mathbf{det}\begin{bmatrix} J & J & -J-1 \\ 0 & \beta'_{Sim} & -1 \\ 0 & 0 & 1+J \end{bmatrix} = J(1+J)\cdot\beta'_{Sim} \neq 0$, if $J, \beta'_{Sim} \neq 0$.The condition $J \neq 0$ is equivalent to $\beta'_{Sim} \neq 0$ by definition. Then also $2J\langle S\rangle \neq 0$ and the solution is not trivial (it is non zero). As an alternative, the last, scaling equation can be replaced with $S_X + S_Y - S_{X\cap Y} = S_{TOT}$ where S_{TOT} is species richness of both sites.

T2: J, β'_{Sim} and $\langle S\rangle$ became i-dependent under conditions $S_X = S_Y$ and $S_Y = S_{X\cap Y}$.

Evidence: Firstly, let $S_X = S_Y = S$, then the system turns into (i) $2JS - (J+1)S_{X\cap Y} = 0$, (ii) $\beta'_{Sim}S - S_{X\cap Y} = 0$, and (iii) $2S = 2\langle S\rangle$. Then

$$\begin{bmatrix} 2J & -J-1 & 0 \\ \beta'_{Sim} & -1 & \| \; 0 \\ 2 & 0 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} 2J & -J-1 & 0 \\ -\beta'_{Sim}(J+1)+2J & 0 & \| \; 0 \\ 2 & 0 & 2\langle S\rangle \end{bmatrix}$$

$\sim \begin{bmatrix} 2J & -J-1 & \| \; 0 \\ 2 & 0 & 2\langle S\rangle \end{bmatrix}$. Hence β'_{Sim} is redundant.

Secondly, let $S_Y = S_{X\cap Y}$ then the system turns into (i) $JS_X - S_{X\cap Y} = 0$, (ii) $(\beta'_{Sim} - 1)S_{X\cap Y} = 0$ ($\beta'_{Sim} = 1$ in the case), and (iii) $S_X + S_{X\cap Y} = 2\langle S\rangle$. Then

$\begin{bmatrix} J & -1 & 0 \\ 0 & \beta'_{Sim}-1 & \| \; 0 \\ 1 & 1 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} J & -1 & 0 \\ 0 & \beta'_{Sim}-1 & \| \; 0 \\ J & 0 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} J & -1 & \| \; 0 \\ J & 0 & 2\langle S\rangle \end{bmatrix}$. Again, β'_{Sim} is redundant and thus carries no extra information.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

T3: Evidence for the constraints Eq. 1 is as follows - $J(S_X + S_Y) - JS_{X\cap Y} \stackrel{\text{def}}{=} S_{X\cap Y} \Rightarrow J(S_X + S_Y) = S_{X\cap Y}(1 + J) \Rightarrow \frac{J}{J+1}(S_X + S_Y) = S_{X\cap Y} =_{(\text{Eq. } T3)} \beta'_{Sim}\min(S_X, S_Y)$. It follows that $\frac{J}{J+1}\frac{S_X+S_Y}{\min(S_X,S_Y)} = \beta'_{Sim}$. Hence Eq. 1.

T4: *Simpson beta* (β_{Sim}, here *Sim*) *is a reversed index of nestedness*. **Evidence**: $\beta_{Sim} = \frac{\min[b,c]}{a+\min[b,c]}$. $\beta_{Sim}[r3] = \beta_{Sim}[r4] = 1$ because $a = 0$ in these cases; $\beta_{Sim}[r1] = \beta_{Sim}[r2] = \beta_{Sim}[r5] = \beta_{Sim}[r6] = 0$ because one of the variables b, c equal zero and $a \neq 0$ in these cases.

T5: *Whittaker index* (β_{Wh}) *is an index of beta diversity.* **Evidence**: $\beta_{Wh0} = \frac{S_T}{\langle S_1\rangle}$. S_T is the size of the large set and $\langle S_1\rangle$ is the size of the small set in re-arrangements r1,r2,r5 and r6 in Fig. 3. Hence, $Wh[r1] > Wh[r2] > Wh[r6] > Wh[r5] = 1$. These constraints are also met by the species-richness gradient (Tab. 1). Nevertheless, following Gaston *et al*. (2007), the difference between *R* and β_{Wh0} is given by the fact that β_{Wh0} is computed for two nested sites whilst *R* is computed for two non-overlapping sites.

T6: *C-score* (*CS*) *captures none of the five phenomena.* **Evidence**: This evidence is based on *C-S* indices computed for various pairs of sets with mutual re-arrangement as in Fig. 3 and sizes M1 = M2 = 15, M3=10, M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11=5, and M12=4. The computed order is $0 = CS[r1] = CS[r2] = CS[r5] = CS[r6] < CS[r3] < CS[r4] = 0.6$ for C-Scores that were standardized (in accord with Gotelli & McCabe, 2002) by numbers of pairs. For rough *C-Score* the order is as follows: $0 = CS[r1] = CS[r2] = CS[r5] = CS[r6] < CS[r4] < CS[r3] = 0.75$. This matches no experience of the five spatial phenomena.

T7: *Nestedness by Baselga* (β_{Bas}) *captures none of the five phenomena.* **Evidence**: $\beta_{Sne} = \frac{b+c}{2a+b+c} - \beta_{Sim}$. In re-arrangements r3 and r4 in Fig. 3 $\beta_{Sne} = 0$, for $a = 0$, which excludes turnover and beta diversity. In re-arrangements r1 and r2 in Fig. 3 $\beta_{Sne}[r1] < \beta_{Sne}[r2]$, for $\beta_{Sim}[r1] = \beta_{Sim}[r2] = 0$, $b = 0$, $c_{r2} < c_{r1}$, and a does not vary between these two cases. This excludes nestedness. Finally, $\beta_{Sne}[r5] = \beta_{Sne}[r3] = 0$, for $b = c$ and $\beta_{Sim}[r5] = 0$, which excludes cooccurrence.

T8: N_C, *standardized as in Gotelli & McCabe(2002) belongs to Jaccard index family and captures reversed turnover or co-occurence*. **Evidence**: In accord with T19.20 it scales positively with $\beta_{Sør}$, which in accord with T19.3 scales positively with J.

T9: : N_C, *standardized as in* Wright & Reeves (1992) *(labeled as C in* Wright & Reeves (1992) *) scales with Simpson beta and captures nestedness if Species Richness approaches infinity. N_C is i-dependent on species richness.*
Evidence: $N_{C2} \stackrel{\text{def}}{=} \frac{S_{X\cap Y} - E(S_{X\cap Y})}{\max(S_{X\cap Y}) - E(S_{X\cap Y})}$, where $\max(S_{X\cap Y}) = S_Y$ if we put $S_X \geq S_Y$, and E() stands for expectance. So, $\mathrm{E}(S_{X\cap Y}) = \frac{1+\cdots+S_Y}{S_Y} = \frac{S_Y+1}{2}$. Hence $N_{C2} = \frac{2S_{X\cap Y} - S_Y - 1}{S_Y - 1}$. In general, $N_{C2} = \frac{2rS_{X\cap Y} - rS_Y - 1}{rS_Y - 1}$, where $r > 0$ emulates variation in species richness. Apparently, N_{C2} is I dependent on r, and $\lim_{r\to\infty} N_C = 2\frac{S_{X\cap Y}}{S_Y} - 1 = 2\beta'_{Sim} - 1$.

T10: *Number of gaps* (N_0) *captures none of the five phenomena.* **Evidence**: This evidence is based on N_0 indices computed for various pairs of sets (see T6 above). The computed order is $0 = N_0[r1] = N_0[r2] = N_0[r4] = N_0[r5] = N_0[r6] < N_0[r3] = 0.3$ for the index that were standardized by number of incidences within the focal matrix. This matches no experience of the five spatial phenomena.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

T11: *Ruggiero index of beta-diversity* (β_{Rib}) *is co-occurrence or nestedness depending on the direction.* **Evidence**: The index *rib* depends on the order of the focal assemblages. It is defined as $\beta_{Rib} = \frac{a}{a+c}$, where *c* captures either the first or second assemblage. In our case, $0 = c_{r5} < c_{r6} < c_{r2} < c_{r1}$ and $a_{r5} = a_{r6} = a_{r2} = a_{r1}$. Hence $0 < \beta_{Rib}[r1] < \beta_{Rib}[r5]$. At the same time $0 = \beta_{Rib}[r3] < \beta_{Rib}[r4]$, for $a = 0$ in these cases. The index β_{Rib} thus captures co-occurrence. If we replace *b* with *c*, then $\beta_{Rib}[r1] = \beta_{Rib}[r2] = \beta_{Rib}[r6] = \beta_{Rib}[r5] = 1$, for *c*=0 in these cases. The index *rib* therefore captures nestedness in the case.

T12: *Nestedness by Almeida-Neto et al. (2008)* (N_{NODF}) *would be an index of nestedness if we ignored the condition that* $S_X = S_Y \implies N_{NODF} = 0$. **Evidence**: Where assemblages differ in their species richness, N_{NODF} equals β'_{Sim} from the Simpson-beta family, and which is an index of nestedness. Where assemblages have equal species richness $N_{NODF} = 0$. If we accepted that two equal-sized assemblages cannot be mutually nested, then N_{NODF} is an index of nestedness; but our framework excludes this possibility.

T13: *Discrepancy* (*D*) *for a high number of sites approaches the reversed constraints on nestedness.* **Evidence**: For a small number of sets to compare *D* does not follow any of the focal phenomena. The evidence is again based on *D* computed for various pairs of sets (see T6 above). The computed order is $0 = D[r1] = D[r2] = D[r5] = D[r6] < D[r3] < D[r4] = 0.5$ for *D* standardized by number of incidences within the focal matrix. For rough *D*, the order is as follows: $0 = D[r1] = D[r2] = D[r5] = D[r6] < D[r3] = D[r4] = 5$. This matches no experience of the five spatial phenomena. However, if the number of sets is large enough (simulations suggest more than 30 simulations), the standardization by number of incidences begins to work properly and the *D* will scale with the indices of nestedness (Fig. SI.3). The reason is that: Discrepancy is defined as the minimum number of incidences that must be shifted along rows of incidence matrix (rows – sites; columns - species) to get absolutely nested assemblages. Discrepancy is standardized by the number of incidences within the matrix. Discrepancy of absolutely nested matrix is thus by definition zero. Discrepancy of an absolutely non-nested matrix (where each site has its unique set of species) is computed as $\sum_{i=1}^{Sites} S_i - max_{i=1}^{Sites}[S_i]$, where S_i is the species richness of the i-th site. It is standardized by number of incidencies, i.e., $\sum_{i=1}^{Sites} S_i$. Hence, $D = \frac{\sum_{i=1}^{Sites} S_i - max_{i=1}^{Sites}[S_i]}{\sum_{i=1}^{Sites} S_i} = 1 - \frac{max_{i=1}^{Sites}[S_i]}{\sum_{i=1}^{Sites} S_i}$, which approaches one if $max_{i=1}^{Sites}[S_i] \ll \sum_{i=1}^{Sites} S_i$. This is the case for practically all datasets with large numbers of sites. It might only be broken if maximum species richness was high and species richness of the other sites was extremely small. Apparently, this cannot happen for almost regular (i.e., square-shaped matrices), as each site must be occupied by at least one species. Simulations show a one to one scaling of *D* with Simpson beta (non nestedness) for matrices of 30 and 100 sites (Fig. SI.3).

T14: *Indices of species-richness gradient* (β_{gl} and R) *capture the experience of species-richness gradient.* **Evidence**: R is the ratio between minimum to maximum species richness. Hence $R[r1] = R[r3] < R[r2] < R[r6] < R[r4] = R[r5] = 1$. *R* is therefore a strictly reversed value to the species-richness gradient and it thereby belongs to the species-richness-gradient family.

T15: *Jaccard index* (J) *is a reversed index of species turnover or co-occurrence*. **Evidence:** $J = \frac{a}{a+b+c}$. $a = 0$ in re-arrangements r3 and r4 in Fig. 3. Hence $J[r3] = J[r4] = 0$, which excludes the sr-gradient but support turnover as defined in our framework. $0 = c_{r5} < c_{r6} < c_{r2} < c_{r1}$, $b = 0$ and a does not vary between re-arrangements r1,r2,r5 and r6 in Fig. 3. Hence, $0 < J[r1] < J[r5] = 1$, which supports only reversed constraints of turnover or the constraints of co-occurrence.

T16 (i-dependence of three indices): The value of any index defined by Eq. B1 can b computed from values of Jaccard index (J, Eq. T1, SI2) and Simpson nestedness (β'_{Sim}, Eq. T3, SI2) using Eq. B5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Evidence: Put arbitrary $S_Y \leq S_X$, then

$$I =_{(Eq.B5)} \frac{J(\beta'_{Sim}(k_1+k_2)-k_2+k_3)+\beta'_{Sim}k_2}{J(\beta'_{Sim}(l_1+l_2)-l_2+l_3)+\beta'_{Sim}l_2} =_{(Eq.T1,T3)} \frac{\frac{S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}}(\frac{S_{X\cap Y}}{S_Y}(k_1+k_2)-k_2+k_3)+\frac{S_{X\cap Y}}{S_Y}k_2}{\frac{S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}}(\frac{S_{X\cap Y}}{S_Y}(l_1+l_2)-l_2+l_3)+\frac{S_{X\cap Y}}{S_Y}l_2}$$

$=\frac{k_1S_{X\cap Y}+k_2S_X+k_3S_Y}{l_1S_{X\cap Y}+l_2S_X+l_3S_Y}$, which is the definition Eq. B1. The Eq. B5 was gain as a solution of the equation $det\begin{vmatrix} -J & -J & J+1 \\ 0 & \beta'_{Sim} & -1 \\ Il_2-k_2 & Il_3-k_3 & Il_1-k_1 \end{vmatrix} = 0.$

T17 (*species rich and poor assemb lages*): The variation of $S_{poor} = 2\langle S\rangle R/(1+R)$ indicates inevitable change in richness of the species poorer assemblage, and the variation of $S_{rich} = 2\langle S\rangle/(1+R)$ indicates inevitable change in richness of the species richer assemblage.
Evidence: The evidence is based on the solution of the system of three equations (see Box 1 for the equations and T1 for the matrix of these equations). Then

$\begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| & 0 \\ 0 & 0 & J+1 & 2J\langle S\rangle \end{bmatrix} \Rightarrow S_{X\cap Y} = 2\langle S\rangle\frac{J}{J+1}$ & $S_Y = \frac{S_{X\cap Y}}{\beta'_{Sim}}$ & $S_X = \frac{(J+1)S_{X\cap Y}-JS_Y}{J}$. It follows that $S_Y = 2\langle S\rangle\frac{J}{(J+1)\beta'_{Sim}} =_{(Eq.7)} 2\langle S\rangle\frac{R}{1+R}$. The last equality follows from Eq. 7. Finally, $S_X = \frac{2\langle S\rangle J-2\langle S\rangle J\frac{R}{R+1}}{J} = 2\langle S\rangle\left(1-\frac{R}{R+1}\right) = 2\langle S\rangle/(R+1)$. Because $0 < R \leq 1$ then $S_Y \leq S_X$, and we relabel $S_{poor} := S_Y$ and $S_{rich} := S_X$.

T18. (*Guide* to *solve a system of three equations).* System can be written as $\begin{bmatrix} Ax+By+Cz & = & 0 \\ Dx+Ey+Fz & = & 0 \\ Gx+Hy+Kz & = & W \end{bmatrix}$ now we multiply the second equations by minus A and the first equation by D and add up them. Then we replace the second equation with the resulting equation. So

$= \begin{bmatrix} ADx+BDy+CDz & = & 0 \\ -ADx-AEy-AFz & = & 0 \\ Gx+Hy+Kz & = & W \end{bmatrix} = \begin{bmatrix} Ax+By+Cz & = & 0 \\ (BD-AE)y+(CD-AF)z & = & 0 \\ Gx+Hy+Kz & = & W \end{bmatrix}$. If D=0 we skip this step. Similarly we multiply the first equation with G and the third equation with minus A and add up them

$= \begin{bmatrix} AGx+BGy+CGz & = & 0 \\ (BD-AE)y+(CD-AF)z & = & 0 \\ -AGx-AHy-AKz & = & -AW \end{bmatrix} = \begin{bmatrix} Ax+By+Cz & = & 0 \\ (BD-AE)y+(CD-AF)z & = & 0 \\ (BG-AH)y+(CG-AK)z & = & -AW \end{bmatrix}$. If G=0 we skip this step. Now we do the same with the second and third equations to eliminate the term with y from the last equation. Thus

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$= \begin{bmatrix} Ax + By + Cz & = & 0 \\ (BD - AE)y + (CD - AF)z & = & 0 \\ ((CD - AF)(BG - AH) - (BD - AE)(CG - AK))z & = & AW(BD - AE) \end{bmatrix}.$$ If $BG - AH = 0$ we skip this step.

The ultimate goal is to get diagonal matrix $\begin{pmatrix} L_{1x}x & +L_{1y}y & +L_{1z}z & = & M_1 \\ 0 & +L_{2y}y & +L_{2z}z & = & M_2 \\ 0 & +0 & +L_3 z & = & M_3 \end{pmatrix}$ where $L_{1x} \neq 0$, $L_{2y} \neq 0$ and $L_{3z} \neq 0$, thus we stop the procedure whenever we gain this goal. Replacing lines with each other is allowed.

From the last equation in the above matrix we compute $z = \frac{AW(BD-AE)}{(CD-AF)(BG-AH)-(BD-AE)(CG-AK)}$. If the equations are independent (that is, the indices are i-independent) the denominator differs from zero. They can be computed from the second equation providing that we know z ($y = \frac{-(CD-AF)z}{BD-AE}$), and x can be computed from the first equation providing that we know z and y ($x = \frac{-Cz-By}{A}$). In the matrix notation (as in T1 and T17), we only did not write 'x','y','z' and '=' to hold the notation as simple as possible (this is a standard habit in mathematics).

T19 (scaling properties): *Relationships between the focal indices obey the equations as listed in Tab. SI.1*. Evidences for the relationships are as follows (the numbering '1-21' refers to the labels above the equation marks in Tab. SI.1; brackets $\langle . \rangle$ label a mean value):

1. $J = \frac{a}{c+b+a} = \frac{a}{(c+a)+(b+a)-a} = \frac{\frac{a}{\min[b,c]+a}}{\frac{c+a}{\min[b,c]+a} + \frac{b+a}{\min[b,c]+a} - \frac{a}{\min[b,c]+a}} = \frac{rN}{1+R^{-1}-rN} = \frac{\beta'_{Sim}}{1+R^{-1}-\beta'_{Sim}}$ To be continued at T15.8.

2. $\beta_{sr} = \frac{a+b+c}{2a+b+c} = \left(1 + \frac{a}{a+b+c}\right)^{-1} = \frac{1}{1+J}$.

3. $\beta_{Sor} = \frac{2a}{2a+b+c} = 2\left(1 + \frac{a+b+c}{a}\right)^{-1} = 2(1+J^{-1})^{-1} = 2\frac{J}{J+1}$.

4. $\beta_{Wh2} = \frac{a+b+c}{(2a+b+c)/2} =_{(2)} 2\beta_{sr} = 2\frac{1}{1+J}$.

5. $\beta_{HK} = 1 - \frac{2a}{2a+b+c} =_{(3)} 1 - \beta_{Sor} = 1 - 2\frac{J}{J+1} = \frac{1-J}{1+J}$.

6. $\beta_t = \frac{b+c}{2a+b+c} = \frac{2a+b+c-2a}{2a+b+c} = 1 - \frac{2a}{2a+b+c} =_{(5)} \frac{1-J}{1+J}$.

7. $\beta_G = \frac{b+c}{a+b+c} = \frac{a+b+c-a}{a+b+c} = 1 - \frac{a}{a+b+c} = 1 - J$.

8. $\beta_{Sim} = \frac{\min[b,c]}{a+\min[b,c]} = \frac{a+\min[b,c]-a}{a+\min[b,c]} = 1 - \frac{a}{a+\min[b,c]} = 1 - \beta'_{Sim}$. Consequently, $J =_{(1)} \frac{1-\beta'_{Sim}}{R^{-1}+\beta'_{Sim}}$.

9. $J =_{(8)} \frac{1-\beta'_{Sim}}{R^{-1}+\beta'_{Sim}} \Rightarrow J(R + 1 - \beta'_{Sim} \cdot R) = \beta'_{Sim} \cdot R \Rightarrow \frac{J}{1+J} = \beta'_{Sim}\frac{R}{1+R}$.

10. Let $b \leq c$ then $R = \frac{b+a}{c+a}$ and thus $\beta_{gl} = 2\frac{c-b}{2a+b+c} = 2\frac{(c+a)-(b+a)}{2a+b+c} = 2\frac{(c+a)(1-R)}{2a+b+c} =$
$= 2(1-R)\left(\frac{2a+b+c}{a+c}\right)^{-1} = 2(1-R)\left(\frac{a+b+a+c}{a+c}\right)^{-1} = 2(1-R)(1+R)^{-1} = 2\frac{1-R}{R+1} \Longrightarrow R = \frac{2-\beta_{gl}}{2+\beta_{gl}}$.

11. $\beta_{Sim} =_{(8)} 1 - \beta'_{Sim} \Longrightarrow \langle\beta_{Sim}\rangle = \langle 1 - \beta'_{Sim}\rangle \Longrightarrow \langle\beta_{Sim}\rangle = \langle 1\rangle - \langle\beta'_{Sim}\rangle$

12. Simulation based evidence (see Fig. SI.3).

13. $\frac{J}{J+1} =_{(9)} \beta'_{Sim}\frac{R}{1+R} \Rightarrow \langle\frac{J}{1+J}\rangle = \langle\beta'_{Sim}\frac{R}{1+R}\rangle \Rightarrow \langle\frac{J}{1+J}\rangle = \langle\beta'_{Sim}\rangle\langle\frac{R}{1+R}\rangle + \mathrm{cov}(\beta'_{Sim}, \frac{R}{1+R})$.

Where the i-th values $\beta'_{Sim,i}$ and $\frac{R_i}{1+R_i}$ are s-independent across all i, $\mathrm{cov}\left(\beta'_{Sim}, \frac{R}{1+R}\right) = 0$. Where the covariance is large small, the Eq. 8 is approximative. Nestedness, $\beta'_{Sim,i}$ and species richness gradient $\frac{R_i}{1+R_i}$ does not constraint each other (share only one sr component of the S_X, S_Y, $S_{X\cap Y}$) therefore does

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

not constraint each other and are likely s-independent (implying zero covariance), else an ecological driver makes covariance nonzero. For data verification see Fig. SI.8.

14. $0 \le R \le 1 \;\Rightarrow\; 0 \le \frac{R}{1+R} \le 0.5 \Rightarrow \frac{J}{1+J} =_{(\mathrm{Eq.12})} \beta'_{Sim}\frac{R}{1+R} \le 0.5\beta'_{Sim} \Rightarrow J \le \frac{\beta'_{Sim}}{2-\beta'_{Sim}}$.

15. $J =_{(6)} \frac{1-\beta_t}{1+\beta_t} =_{(\mathrm{Def}\,\beta Bas)} \frac{1-\beta_{Sne}-\beta_{Sim}}{1+\beta_{Sne}+\beta_{Sim}}$.

16. $\beta_{BC} \stackrel{\text{def}}{=} 1-\frac{S_{X\cap Y}}{S_X+S_Y} = 1-J\frac{S_X+S_Y-S_{X\cap Y}}{S_X+S_Y} = 1-J\beta_{BC} \Rightarrow \beta_{BC} = 1-J\beta_{BC} \;\Rightarrow \beta_{BC} = \frac{1}{1+J}$.

17. $\beta_{DICE} \stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X+S_Y-2S_{X\cap Y}} = 2J\frac{S_X+S_Y-S_{X\cap Y}}{S_X+S_Y-2S_{X\cap Y}} = 2J\left(1-\frac{S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}}\right)^{-1} = 2\frac{J}{1-J}$.

18. $\beta_{Sne} \stackrel{\text{def}}{=} \beta_t - \beta_{Sim} = \frac{S_X+S_Y-2S_{X\cap Y}}{S_X+S_Y} - \frac{S_{X\cap Y}}{S_Y} = \frac{S_XS_{X\cap Y}-S_YS_{X\cap Y}}{S_Y(S_X+S_Y)} = \frac{\frac{S_Y}{R}S_{X\cap Y}-S_YS_{X\cap Y}}{S_Y(S_X+S_Y)} = \frac{(R^{-1}-1)S_{X\cap Y}}{S_X+S_Y}$.

$S_Y \le S_X$ is an arbitrary choice without losing generality.

19. $N_{C1} \stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X+S_Y} = \frac{S_X+S_Y+S_{X\cap Y}-S_X-S_Y}{S_X+S_Y} = 1-\frac{S_X+S_Y+S_{X\cap Y}}{S_X+S_Y} = 1-\beta_{SR}$

20. $N_{C1} \stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X+S_Y} = \frac{1}{2}\frac{2S_{X\cap Y}}{S_X+S_Y} = \frac{\beta_{Sør}}{2}$

21. $\beta_{jtu} \stackrel{\text{def}}{=} 2\frac{\min(S_X,S_Y)-S_{X\cap Y}}{2\min(S_X,S_Y)-S_{X\cap Y}} = 2\frac{1-\beta'_{Sim}}{2-\beta'_{Sim}} = 2\frac{\beta_{Sim}}{1+\beta_{Sim}}$

T20. How to get Jaccard similarity and Simpson nestedness if one of the indices is partitioned.

Let

$$I_1 = \frac{K_{1a}S_{X\cap Y}+K_{2a}S_X+K_{3a}S_Y}{L_{1a}S_{X\cap Y}+L_{2a}S_X+L_{3a}S_Y} + \frac{K_{1b}S_{X\cap Y}+K_{2b}S_X+K_{3b}S_Y}{L_{1b}S_{X\cap Y}+L_{2b}S_X+L_{3b}S_Y} = I_a + I_b$$

(*if indices are subtracted the* k_{ib} *coefficients are multiplied with minus one*)

and

$$I_2 = \frac{K_{12}S_{X\cap Y}+K_{22}S_X+K_{32}S_Y}{L_{12}S_{X\cap Y}+L_{22}S_X+L_{32}S_Y}.$$

Then the triple I_a, I_b, and I_2 are necessarily i-dependent. Thus

$$\mathbf{det}\begin{bmatrix} I_aL_{1a}-K_{1a} & I_aL_{2a}-K_{2a} & I_aL_{3a}-K_{3a} \\ I_bL_{1b}-K_{1b} & I_bL_{2b}-K_{2b} & I_bL_{3b}-K_{3b} \\ I_2L_{12}-K_{12} & I_2L_{22}-K_{22} & I_2L_{32}-K_{32} \end{bmatrix} \equiv 0.$$

I_2, and I_1 are known ($I_1 = I_a + I_b$), so

$$\mathbf{det}\begin{bmatrix} (I_1-I_b)L_{1a}-K_{1a} & (I_1-I_b)L_{2a}-K_{2a} & (I_1-I_b)L_{3a}-K_{3a} \\ I_bL_{1b}-K_{1b} & I_bL_{2b}-K_{2b} & I_bL_{3b}-K_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix} \equiv 0$$

where $M_i = I_2L_{i2} - K_{i2}$.

After expansion

$$I_b^2D_{LL} - I_b(I_1D_{LL} - D_{KL} + D_{LK}) + I_1D_{LK} - D_{KK} = 0,$$

where $D_{LL} = \mathbf{det}\begin{bmatrix} L_{1a} & L_{2a} & L_{3a} \\ L_{1b} & L_{2b} & L_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$; $D_{KK} = \mathbf{det}\begin{bmatrix} K_{1a} & K_{2a} & K_{3a} \\ K_{1b} & K_{2b} & K_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$; $D_{LK} = \mathbf{det}\begin{bmatrix} L_{1a} & L_{2a} & L_{3a} \\ K_{1b} & K_{2b} & K_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$;

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and $D_{KL} = \mathbf{det}\begin{bmatrix} K_{1a} & K_{2a} & K_{3a} \\ L_{1b} & L_{2b} & L_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$.

Having I_b, we can pick up two i-independent indices and then use the algorithm from Box 2 to compute any dimensionless index.

The above solution works only if the pair I_a, I_b is i-independent, the pair I_a, I_1 is i-independent, and the pair I_1, I_b is i-independent (together they are always i-dependent). If either pair of indices is i-dependent, the solution is simpler. We just convert the i-dependent indices to an index from their family (Tab. SI.1) and then follow the algorithm from Box2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

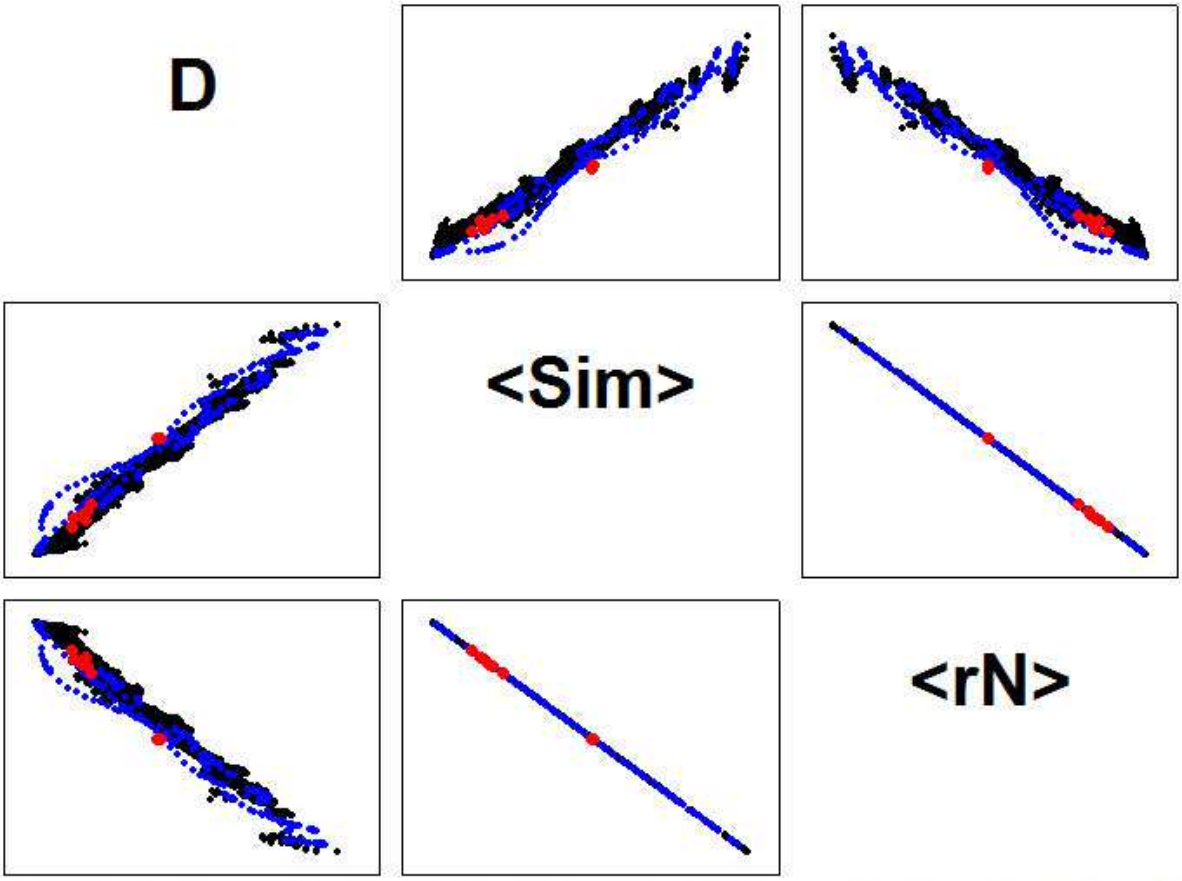


Fig. SI.3: Relationships between indices of nestedness computed for more assemblages as simulated (30 sites x 30 species – black dots; 100 sites x 100 species – blue dots) and observed (red dots, SI.7). Matrices were generated to cover as wide spectrum of assemblages as possible (see details on simulations below). *D* – discrepancy by Brualdi & Sanderson (1999); $\langle Sim \rangle$– a simple mean of the Simpson index across all pairs of adjacent plots (Lennon *et al*. 2001); $\langle rN \rangle \equiv \langle \beta'_{Sim} \rangle$ – simple mean of nestedness as defined in Šizling *et al.* 2009,2016; *D,* was standardized (i.e., divided) by total number of incidences within the focal matrix (Greve *et al*. 2005). The indices scale in one-to-one manner belonging to the same family and measuring the extent of nestedness. For exact evidence see SI.2.T12.

The incidence matrices for testing the relationships between various indices of nestedness of multiple assemblages

The 30x30 and 100x100 matrices of artificial assemblages were generated to cover the whole range of possible values for each index of nestedness as follows:

1. Matrices with various distributions of incidences in rows and columns were generated using the algorithm published by Šizling *et al*. (2009). This algorithm places each species' incidence randomly within a matrix to keep the required distributions. We created matrices with the distributions of incidences for each row or column randomly picked up from the regular distribution of the numbers (a) between 1 and 30, (b) between 10 and 20, (c) between 15 and 16 and (d) between 15 and 30, and used all the combinations of these possibilities for rows and columns. One hundred matrices were generated using each combination, which gave altogether 600 matrices.
2. Matrices with column and row counts of approximately 5, 10, 15, 20 and 25 (the column and row distributions were regular) were generated using the same algorithm. One hundred matrices were generated for each setting.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3. One hundred matrices with various combinations of bimodal (variously chosen bimodal distributions were used to cover the full range of indices values; the set of the matrices is available see the item 7 below) and unimodal distributions of incidences for rows and columns were generated. These make altogether 400 matrices.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

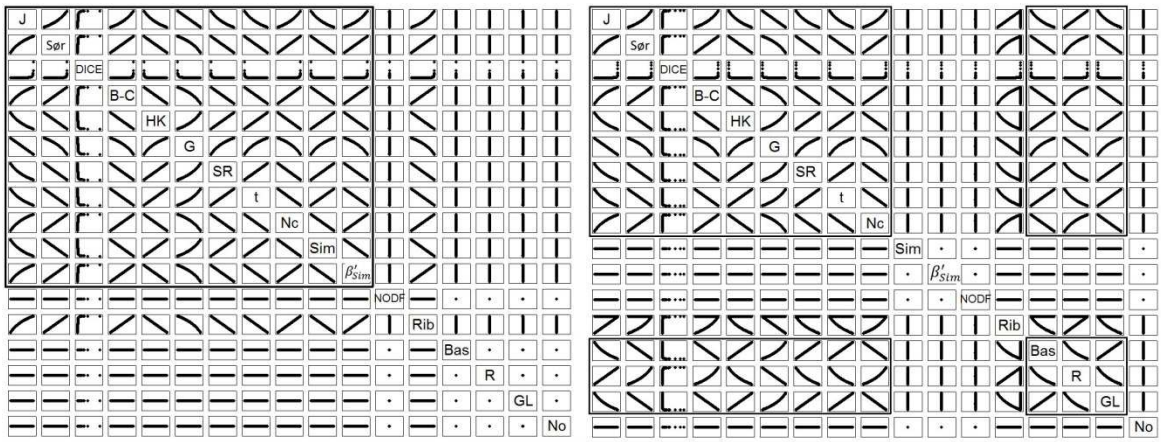


Fig. SI.4: Relationship between pairwise indices where there is no variation in species richness (left) and where assemblages are perfectly nested (right). As predicted, no variation in species richness merges families of J, β_{Sim} and N_0 together, and perfect nestedness merges families of J, β_{rb}, β_{sne}, and β_{gl}. The evidence that β_{sne} shows variability where nestedness is perfect disqualifies this measure from being a proxy for nestedness. Black rectangles delimit the merged families. For detailed legend see capture to Fig. 5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SI.5 i-independence of and s-dependence on species richness

The Jaccard index (*J*), and thus all the indices that scale one-to-one with *J* (Fig. 2), has been referred to as 'dependent' on species richness (Simpson, 1943; Lennon *et al*., 2001; Koleff & Gaston, 2002; Baselga, 2010a) and on the contrast between species richness of two assemblages (Simpson, 1943). This has led to a search for an index that is species richness 'independent', and to attempts to modify *J* , so that ecologists could compare assemblages that varied in species richness. This s-dependence between *J* and species richness was based on empirical experience (Koleff & Gaston, 2002), and on arguments that there are bounds on the *J* imposed by contrast in species richness (Simpson, 1943).

We found that all dimensionless (unitless) indices, including *J*, are i-independent of species richness. For all indices that can be expressed by the universal definition (Eq. B1) it holds that

$$I \stackrel{\text{def}}{=} \frac{k_1 S_{X\cap Y}+k_2 S_X+k_3 S_Y}{l_1 S_{X\cap Y}+l_2 S_X+l_3 S_Y} = \frac{k_1 \rho S_{X\cap Y}+k_2 \rho S_X+k_3 \rho S_Y}{l_1 \rho S_{X\cap Y}+l_2 \rho S_X+l_3 \rho S_Y}, (\rho \neq 0) \qquad \text{(Eq. SI.6)}$$

and thus the index does not change when S_X, S_Y, and $S_{X\cap Y}$ scale proportionally to each other. In this case, the index has the same value regardless of species-richness, and thus it is not uniquely determined by species richness (see SI.2.T1 for a proof using Cramer's rule). Any observed s-dependence between the dimensionless index and species richness is therefore caused by disproportional scaling between S_X, S_Y or $S_{X\cap Y}$.

Disproportional scaling can, however, appear at sites with small species richness because species richness is an integer. In this case, the frequency distribution of possible *J* values is affected by total species richness. This in turn affects the most likely value of *J*, imposing its s-dependence on species richness. The reason is that the Jaccard index can only have a finite number of values. For example, if $S_X = 1$ then $J = 1, 1/2, 1/3, 1/4, \ldots, 0$, accumulating possible values below 1/2. $S_X = 2$ then allows for 2/3, which is above 1/2, $S_X = 3$ allows for $3/4 > 2/3$ and so on. Further computation of possible *J* values for increasing S_X (Fig. SI.5) shows an increasingly even distribution of *J*-values. This mechanismus works for any index that can be expressed by Eq. B1, and the effect cannot be eliminated by inventing a new dimensionless index.

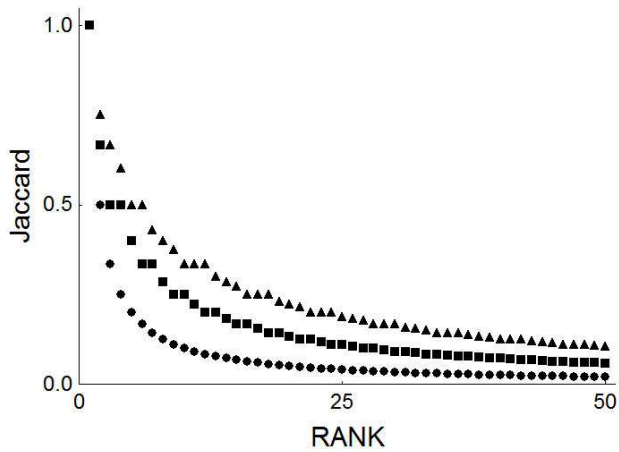
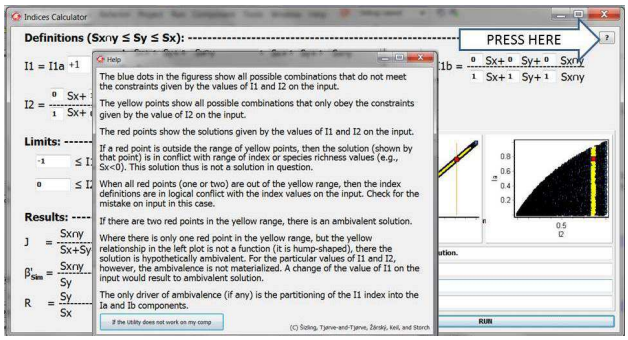


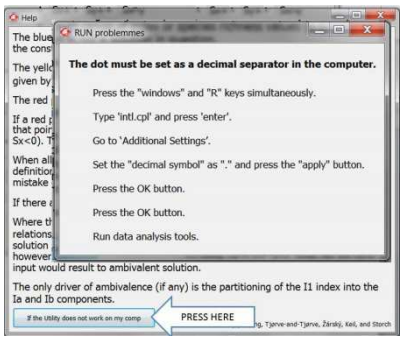
Fig. SI.5: Rank plot of the first fifty values that can reach Jaccard index if S_X is fixed and S_Y and $S_{X\cap Y}$ vary within their limits ($1 \leq S_Y < \infty$ and $0 \leq S_{X\cap Y} \leq \min(S_X, S_Y)$); the three distributions on display correspond to $S_X = 1$ (circles), $S_X = 2$ (squares) and $S_X = 3$ (triangles).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

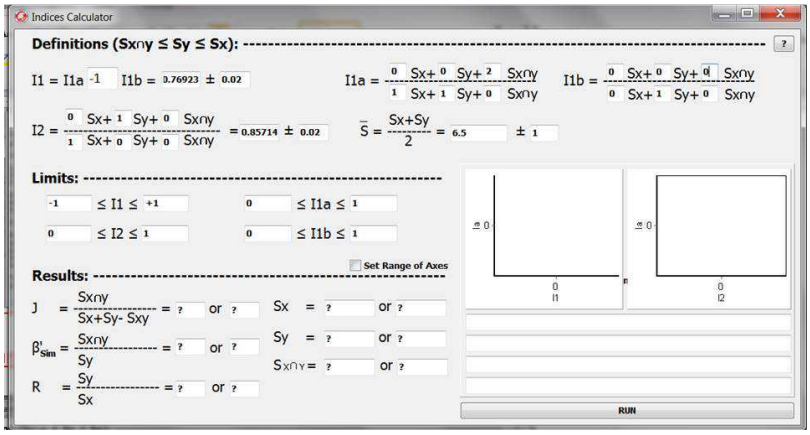
SI.6 Software Calculator (manual)

The calculator works properly only with mashines that utilize 'dot' as a decimal separator. See the help for detailed instructions.





After running the calculator (click your mouse) a pannel appears.

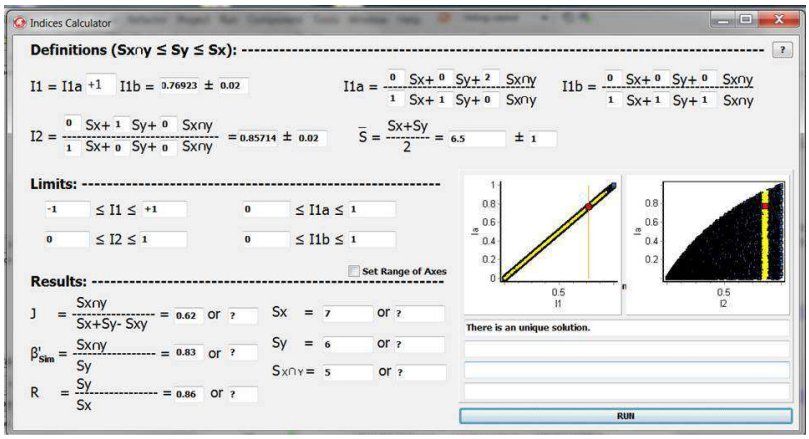


Fill the sections Definitions and Limits. At the pannel above 'Ia1' is defined as Sørensen simillarity (see Box 1) and I1b is always zero, thus the calculator will ignore the component I1b. I2 is species richness uniformity R. This is because $S_Y \leq S_X$ by definition. Mean species richness is known in this example and it is 6.5. Limits are theoretical limits of the indices.

The errors have to be reasonably defined (here 0.02 for indices and 1 for species richness). Too small errors may lead to wrong conclussion that there is no solution because index definitions and values in logical conflict.

Press the RUN button.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

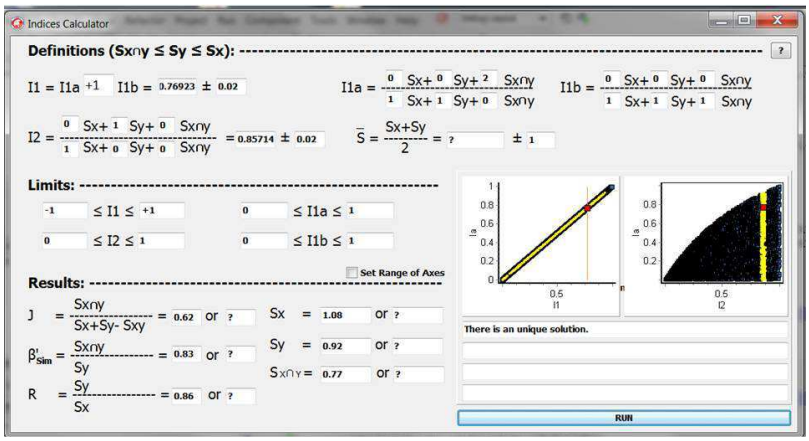


The Results section and the figures are displayed. The resulting values of Jaccard similarity, Simpson nestedness, species-richness uniformity, and three species richnesses are computed as described in Box 2 if there is no partitioning. If there is a partitioning (both I1a and I1b are defined), the equation SI.2.T20 is employed. This equation is quadratic, therefore two different solutions may exist for partitioned indices indicating lost of information.

Figures help to see whether there is a unique solution. The number of solutions is always none, one or two. Two solutions may appear only when index I1 is partitioned. The reason is that additive partitioning may decrease information.

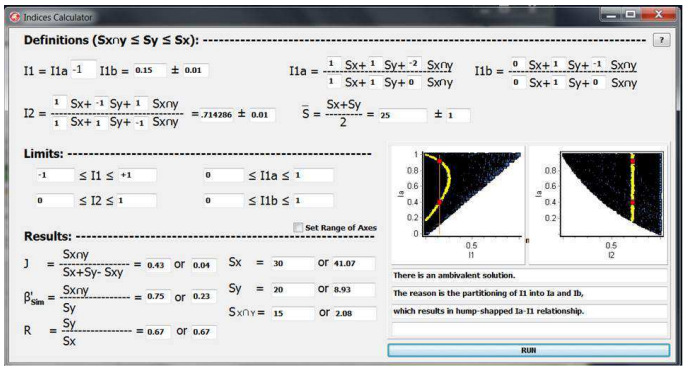
Blue area in figs shows all possible combinations, the yellow line shows all possible combinations constrained by the value of I2, and the red point shows the solution constrained by both indices I1, I2.

Thus, the red point that is out of the yellow line shows mutuall logical conflict of definitions and values of indices. Two red points within the yellow area indicate umbiguous solution.



There is no need for precise species richness. The indices are dimensionless, which means that they are i-independent of species richness. Just write the question mark '?' or number of one in the edit lane for average species richness. The resulting Sx,Sy and Sxy are then only proportional numbers.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



If I1 is partitioned (e.g., $I_1 = \beta_{sne}$, Baselga, 2012; $I_2 = \beta_{nps}$; Podani & Schmera, 2012), then two solutions may exist. Both solutions are displayed and computed in the case.

Enjoy the work and, please, report any error with the calculator (sizling@cts.cuni.cz).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.14.452244; this version posted June 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SI.7 Datasets

For comparison with the results from the artificial (simulated) data, we also plotted observed values of the indices. This allows us to identify relationships that are mathematically possible but may be rare or absent in nature. These observations consisted of three different datasets: a set of 29 microbial assemblages extracted from cryoconite on the Greenland Ice Shield, a set of 24 arctic plant assemblages (4 from Greenland, and 20 from Svalbard), and a set of 20 temperate zone plant assemblages (10 from the Czech Republic, and 10 from Southern Norway). The microbial assemblages were sampled by J.Ž. and A.Š, and processed by J.Ž. Plant assemblages were sampled by A.L.Š., Eva Šizlingová and E.T (see LINK TO DATA ARCHIVE after acceptance).

A list of plant species found in a 10x10 m area was recorded. The data are nested in the sense that the assemblages are grouped so that each group of five assemblages is located within a 1km diameter circle. For the purposes of this analysis, only Genera were used.

Microbial assemblages were sampled at 300 m intervals along two lines on the western margin of the Greenland Ice Sheet in the vicinity of Kangerlussuaq. Sampling and sample processing procedures followed Cameron *et al.* 2016. Here we use data inferred from environmental RNA using Illumina amplicon sequencing to detect the active part of the microbial assemblage. Processing of the sequencing output was performed using the QIIME2 environment (Bolyen *et al*., 2019), filtering for sequences present at least three times in the whole dataset, and rarefaction to the sampling depth of 5000 features per sample. This resulted in the exclusion of 7 samples out of 36 that had fewer features than the sampling depth. The remaining samples were used for the diversity analysis.

Cameron1, K.A., Stibal, M., Zarsky, J.D., Gozdereliler, E., Schostag, M. & Jacobsen, C.S.. (2016) Supraglacial bacterial community structures vary across the Greenland ice sheet. *FEMS Microbiology Ecology*, **92**, doi: 10.1093/femsec/fiv164

Bolyen, E., Rideout, J.R., Dillon, M.R., *et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, **37**, 852–857.