Mathematically and biologically consistent framework for presence-absence pairwise indices

Arnošt L. Šizling[1*], Petr Keil[2], Even Tjørve[3], Kathleen M.C. Tjørve[3], Jakub D. Žárský[4], and David 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: sizling@cts.cuni.cz

Open research statement: Data, index definitions, theses and proofs, calculator and its code are available in public Repository at https://doi.org/10.5281/zenodo.8084396.

The work was supported from public Foundations, see Acknowledgement for details. All data were collected by the authors by non-destructive methods and in accord with law.

Key words: beta diversity, biodiversity, nestedness, partitioning, dependence, Simpson, Jaccard, Baselga, dissimilarity

Abstract

A 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 with each other. At the same time, the following issues have been largely overlooked: (i) requirement that a small re-arrangement of assemblages should only cause a small change in an index, (ii) inferences from the indices about diversity patterns, (iii) inter-dependence of indices based on their information value, (iv) overlap of the ecological phenomena that the indices aim to capture, and (v) incomparability of measures of different phenomena. Neglecting these issues has resulted in the invention or reinvention of indices without increasing their information value, although this value is crucial for correct interpretation of the indices. We offer a framework for pairwise diversity indices that accounts for these issues. We differentiate between statistical and information dependence of indices and show mathematical links between all indices, even those that have not yet been developed. Using linear algebra, we show (1) which set of indices carries complete information on assemblage arrangement, (2) how to calculate any index from two presence-absence indices, which can be used to standardize and compare different indices across the literature, and (3) what can be inferred about diversity phenomena from different informationally independent indices. It is impossible to purify an index of a single biodiversity phenomenon from the effects of other phenomena, because these phenomena inevitably constrain each other. Consequently, many recently proposed indices do not measure the phenomena that they were intended to measure. In contrast, a proper inference can be made by combining classical indices from different, information independent families.

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; Jurasinski et al., 2009; Tuomisto, 2010b; Carvalho et al., 2012; Chao et al., 2012; Newbold et al., 2016; Matthews et al., 2019). Differences in species composition between assemblages in space and time have been assessed by various indices, with a heated debate about which are better for different purposes (Baselga, 2010, 2012; Podani & Schmera, 2011; Schmera et al., 2020; Almeida-Neto et al., 2012; Ulrich et al., 2009, 2017). Substantial work has been done on abundance-based indices, and their inclusion in the framework of Hill numbers (Hill, 1973; Jost, 2007; Tuomisto, 2010b; Chao et al., 2012, 2014a, 2014b; Chiu et al., 2014; Barwell et al., 2015). 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. Furthermore, abundance-based indices downweight rare species, which are often of conservation interest. Because of this, pairwise indices based on presences and absences, rather than abundances, continue to dominate most global analyses of biodiversity (e.g., Dornelas et al., 2014; Blowes et al., 2019; Xu et al., 2023). The aim of this study is to provide a consistent framework for using incidence-based indices.

A particularly intensive debate has concerned the links between indices of species turnover and nestedness, and how to purify them from each other and from species richness differences. This debate can be traced back to Simpson (1943), whose invention of the Simpson similarity (β'_{Sim}; Table 1) resolved the undesirable interdependence between the Jaccard index (J; Jaccard, 1912; Table 1) and the difference in species richness of assemblages. Since then, a number of indices have been introduced that carry equal information to Jaccard and Simpson indices (e.g., $\beta_{Sør}$; Sørensen, 1948; and β_{jtu}; Baselga, 2012; Table 1), or which are mathematically equivalent to earlier indices (e.g., $\beta_{t2} = \beta_{SR} = \beta_{HK}$; defined by Wilson & Shmida, 1984; Schluter & Ricklefs, 1993; Harte & Kinzig, 1997; respectively; Table 1; Table R1 in Šizling et al., 2023). Yet, we lack a

general framework for the relationships between indices, which would apply for any index already invented, or introduced in the future.

Despite disagreements and redundancies, there has been progress. It is to Baselga's (2010) credit that the field has refocused from the classification of indices to their meaning. Lennon et al. (2001), Tuomisto 2010b and Newbold et al. (2016) showed that the Simpson index is in fact Sørensen similarity relativized to the contrast in species richness. Koleff et al. (2003) and Legendre and De Cáceres (2013) pointed out that Simpson and Jaccard represent two different groups of indices, which were 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 the Simpson similarity (β'_{Sim}; Table 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 disentangle mathematical and biological drivers of the dependence between various indices and species richness.

The realization that pairwise indices are simultaneously affected by different phenomena (e.g., by proportion of shared species, as well as, species richness difference) has recently led to an effort to purify one effect from another by means of additive partitioning. Baselga (2010, 2012), Podani and Schmera (2011) and Schmera et al. (2020) presented two ways to mathematically partition the indices. 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 (Box 1). To address this dependence, Baselga (2012) defined the "nestedness-resultant component" (β_{sne}), which is $\beta_{sne} \stackrel{\text{def}}{=} \beta_{t2} - \beta_{Sim}$, where β_{t2} is Sørensen dissimilarity and β_{Sim} is Simpson dissimilarity (see Table 1 for definitions). The subtraction supposedly removes the effect of the turnover component represented by β_{Sim} (Baselga's 2012 interpretation) from the overall dissimilarity represented by β_{t2} (Baselga's 2012 interpretation), and what remains is considered to be the 'nestedness resultant component'. In contrast, Podani and Schmera (2011) see nestedness as

antithetic to replacement (their synonym for turnover) and so indices of nestedness (β_{nps}; Podani & Schmera, 2011) and replacement (β_{rps}; Podani & Schmera, 2011) (see Table 1 for definitions) together sum to one. Their index of nestedness is thus $\beta_{nps} \stackrel{\text{def}}{=} 1 - \beta_{rps}$.

We argue that the abovementioned approaches have several problems (outlined in the section below) which has led to a situation where researchers are typically not sure which indices can be used in different occasions and what the indices actually measure. To sort out the confusion, we present a unified framework for incidence-based indices that addresses these issues (see also Tuomisto 2010a, 2010b). 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 that quantify them. This makes statistically independent or unrelated indices impossible. We further show that the interdependence between the phenomena makes it impossible to 'purify' any index from the effects of other phenomena.

Instead of statistical independence (hereafter *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 here *i-independence* (Box 1; note that in mathematics this is called simply independence). We then show that only two i-independent indices, together with species richness, carry enough information to compute any (even not yet introduced) index, and we provide a user friendly tool to perform this conversion. We also develop guidelines to attribute the indices to different diversity phenomena. We demonstrate that particular types of interrelation of assemblages introduce mutual i-dependence between otherwise i-independent indices, and that this effect can obscure differences between otherwise clearly distinguishable patterns. Finally, we show that all the mathematically and ecologically meaningful indices have already been described by Jaccard (1912), Simpson (1943) and others in the early 20th century (see also Tjørve et al., 2018).

Problems with previously suggested indices

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

Problem 1: Some commonly used indices do not satisfy the *requirement of continuity* (Neumann & Morgenstern, 1953). This requirement, presented for ecologists by Podani and 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 an 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 as N_0 (Patterson & Atmar, 1986) and N_{NODF} (Almeida-Neto et al., 2008) have been defined as zero when species richness of one site (S_X) is equal to species richness of the other site (S_Y). This violates the requirement of continuity (Figure 1). To satisfy the requirement when $S_X = S_Y$, a value of nestedness should instead fall as close as possible to the nestedness values for $S_X = S_Y - 1$ and $S_X = S_Y + 1$. Similarly, the problematic assumption of zero nestedness for $S_X = S_Y$ also affects the "nestedness-resultant components" by Baselga (2010, 2012) (Figure 1). Baselga (2010, 2012) explicitly stated that, in the absence of nestedness, an index of dissimilarity equals its turnover component. However, this would hold only if the equality of S_X and S_Y implied zero nestedness (Figure 1). The problem is that two precisely identical assemblages should be, according to the requirement of continuity, similarily nested as two almost identical and absolutely nested assemblages, i.e. they must not have zero nestedness.

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-phenomenon-to-index* implication: a particular state of the phenomenon (e.g., high or low nestedness) affects the value of an index, so it is assumed that the index measures the phenomenon. However, if an index is to be useful, the appropriate reasoning should follow a *from-index-to-phenomenon* implication: a value of the index should *always* correspond to a unique state of the phenomenon (e.g., only to high or only to low nestedness, not both). In other words, one index value should not be assigned to multiple different states of the phenomenon (see the problem of *converse logic*; e.g., Audi, 1999). An example of the ambigous from-phenomenon-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 (β_{rps}) completely removed from [its] maximum*". However, in Figure 1 we show that cases with 1-13 shared species have equal β_{nps}, although they apparently have different nestedness, conceptually predefined as the extent in which species of smaller community occur also in the larger community (Podani and Schmera, 2011; Patterson, 1984; Baselga, 2010). The equality between their β_{nps} thus violates the from-index-to-phenomenon implication. Similarly, the fact that the absence of nestedness leads to $\beta_{Sør} = \beta_{Sim}$ (Baselga 2010, 2012) does not mean that $\beta_{Sør} = \beta_{Sim}$ implies no 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 remove an effect of one index from another have accounted for *s-dependence* (statistical dependence) between indices (e.g., Simpson, 1943; Koleff et al., 2003; Baselga, 2010, 2012) but failed to account for *i-dependence*, i.e. the dependence of indices in terms of their information content (Box 1). However, in terms of the exact meaning of the indices and their links to ecological phenomena, i-dependence is more fundamental, as we will show later. In short, two indices are

considered i-dependent if one can be uniquely determined by the other and vice versa (Box 1).

Problem 4: The mathematical operation of *subtraction* of indices does not remove one effect from another. Baselga (2010, 2012) has suggested that subtraction of an index of turnover component from an index of dissimilarity leads to the nestedness-resultant component, irrespective of the indices used. Baselga (2012) applies the reasoning about nestedness-resultant components equally to β_t and β_G, so that the nestedness-resultant component of Sørensen dissimilarity (β_{sne}; Table 1) is $\beta_t - \beta_{Sim}$, and the nestedness-resultant component of Jaccard dissimilarity (β_{jne}; Table 1) is $\beta_G - \beta_{jtu}$ (β_{Sim}, and β_{jtu} are the 'turnover components' of Sørensen and Jaccard dissimilarities, respectively (see Table 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} would be the 'true' nesteness resultant component. The reasoning is as follows: β_t is i-dependent to β_G (one can be calculated from the other) and, similarly, β_{jtu} is i-dependent to β_{Sim} (see Table 1, and Thesis T01.21 in Šizling et al., 2023). If the subtraction produced nestedness-resultant components in both cases, then the subtraction would have to account for the same effect in both cases, and thus β_{sne} and β_{jne} should also be i-dependent. However, it is not the case (Baselga, 2012). The problem is that the minus operator in $\beta_t - \beta_{Sim}$ ($\stackrel{\text{def}}{=} \beta_{sne}$) has a different meaning than the minus operator in $\beta_G - \beta_{jtu}$ ($\stackrel{\text{def}}{=} \beta_{jne}$); similarly the subtraction in $x - y$ does not have the same meaning as the subtraction in $\log x - \log y$ (since $\log(x) - \log(y) = log(x/y)$), even though x and $\log x$ are i-dependent. As a result, the logic that subtraction eliminates an effect cannot be applied to both β_{sne} and β_{jne}, and only one of them (if any) can capture the nestedness-resultant component, and we don't know which one.

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 (ISO, 2009; Schmera & Podani, 2011). For example, the framework introduced in Baselga (2010) subtracts and compares Sørensen dissimilarity, β_t, and Simpson dissimilarity β_{Sim} (Table 1); however, these all are not commensurable, as

pointed out by Schmera and Podani (2011), who proposed to sum (or subtract) only indices with the same denominator.

Theory

Here we first describe four distinct phenomena that represent different aspects of the structure of species communities, and can be quantified by different indices. Second, we define the constraints necessary to attribute each of the indices to one phenomenon. Third, we define the independence of the indices (including alpha diversity) in terms of their information content. We contrast this i-independence with s-independence (including relatedness sensu Chao et al., 2012; Figure 2; Box 1), which was a major topic of a large ecological literature (Simpson, 1943; Koleff & Gaston, 2002; Jost, 2007; Baselga, 2010).

Predefinitions of four diversity phenomena

We propose that most of the commonly used indices have originated as proxies for *four phenomena*: (1) nestedness, (2) species turnover, (3) species richness contrast, and (4) beta-diversity sensu stricto (i.e. the contrast between small-scale and large-scale richness). These four phenomena correspond to different (but strongly interrelated) *observations*, where: (i) species of the species-poor site are more or less present in the species-rich site; (ii) species lists vary between the visited sites (but the species richness may or may not change); (iii) species richness varies negligibly or considerably between visited sites, and (iv) species richness of a site represents lower or higher portion of species richness of the whole region.

Imagine an ecologist roaming the landscape, leaving one site, and approaching another, detecting changes in species richness, changes in species composition, and changes in overlap in species composition between species assemblages. They always observe a combination of these four patterns, as the phenomena are interconnected and constrain each other. Apparently, if there is no turnover, then all assemblages are identical and there is no richness contrast. The loss of a species from species poor

assemblages increases both the richness contrast 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. Nevertheless, these four 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 (similarly, we can compare the speed of a moving object without measuring it). The four 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 and Atmar (1986) noticed that species on species-poor mountain ridges 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.

Species turnover and replacement: Generally, species turnover addresses the gain and the loss of species in space or time (Cody, 1975; Wilson & Shmida, 1984; Tuomisto, 2010a). Turnover is sometimes narrowed down to *replacement*, defined by Podani and Schmera (2011) and Schmera et al. (2020) as a situation where one species replaces exactly one other species, irrespective of the ecological role or habitat requirements of the species. This one-for-one replacement is equivalent to Balselga's (2010) idea of turnover component (Schmera et al., 2020), because Baselga's (2010) aim was to separate turnover from species richness contrast. Here we follow the most

common approach (Lennon et al. 2001, Koleff et al. 2003, Gaston et al. 2007) and we understand turnover as a more general phenomenon referring to a change of species composition between assemblages.

Species richness contrast: Species turnover between sites, i.e. the difference in species composition (Cody, 1975; Lennon et al., 2001), can be determined by the difference in species richness (Harrison et al., 1992). Although these two differences are bound by each other (a difference in species richness between two assemblages is always accompanied by difference in their species composition), they are also mutually independent to some extent (two equally species rich assemblages can have all or no species in common). Here we adopt the idea that the contrast between the species richness of two different assemblages (called species richness gradient in Lennon et al., 2001) and species turnover are two distinct phenomena (Lennon et al., 2001) although they are not mutually exclusive - these two phenomena overlap each other and their measures must therefore correlate with each other.

Beta-diversity: The original idea behind beta-diversity is that different regions have different relationships between local (alpha) and regional (gamma) diversity (Whittaker, 1960). Beta-diversity thus quantifies the contrast between gamma-diversity and average alpha-diversity (Whittaker, 1960). Therefore, beta-diversity 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 (hereafter beta-diversity *sensu stricto*) 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 *sensu stricto* were highlighted (Koleff et al., 2003; Tuomisto, 2010b; Chao et al., 2012). However, in accord with Koleff et al. (2003), Jost (2007) and Tuomisto (2010b) 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.

Formal definitions of the phenomena based on their constraints

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 Figure 3. Each set can be considered as a list of species. The constraints are:

- *Nestedness* has its maximum where all the species from the species-poor assemblage are found also in the species-rich assemblage, and minimum where the two assemblages have no species in common (e.g., Patterson & Atmar, 1986; Wright & Reeves, 1992; Brualdi & Sanderson, 1999; Baselga, 2010). Therefore, the two assemblages 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 Figure 3).
- *Turnover* captures the contrast in species composition between two or more assemblages. Consequently, in Figure 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 are identical, hence there is no turnover at all, which means that the index of turnover is at its minimum. In the section above, we distinguish turnover (T) from species richness contrast (Lennon et al., 2001) and so we state, in accord with Tuomisto (2010a), that $\mathrm{T[r3]} = \mathrm{T[r4]}$ but $\mathrm{C}[r3] > \mathrm{C}[r4]$ (Figure 3), i.e. if no species are shared, turnover is maximal regardless of the species richness contrast.
- Because we define the *species richness contrast* as simply the contrast in species richness between two sites, the re-arrangements r1 and r3 in Figure 3 represent the same species richness contrast, as do r4 and r5 in Figure 3. The latter also represents a minimum species richness contrast.

Independence of indices 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 those indices that carry different 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 2). When we add an equation to the set of *n* equations, and if this new set of equations provides an identical solution to the previous set, 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 a widely accepted mathematical definition of independence in a system of equations (see also Box 2) and that it follows the theorem from information theory that only the variables that are uniquely mapped to each other have equal information content (Orlitsky, 2003).

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 2). In extreme circumstances, the set of equations with one unique solution carries complete information on both assemblages. It all means that the complete 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 2).

For example, β'_{Sim} (Simpson nestedness) and J (Jaccard similarity) (Table 1) are i-independent because there is no way to convert β'_{Sim} to J. This is easy to see from Figure 4a,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; see Box 1), but *not one-to-one* correspondence (Figure 2) between energy and momentum (Figure 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. The indices 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 the constraints reveal their relatedness.

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 used most often or is more convenient for some reason. Finally, we will show how to convert indices within their respective family to each other (Box 3). Mathematical details of these steps are in repository (Theses T01-T20 in Šizling et al., 2023), and here we summarize the results.

There is a number of pairwise 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 (Figure 5). Where we did not see a one-to-one correspondence (Figure 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 (Thesis T01 in Šizling et al., 2023). In this way, we found several families of mutually equivalent indices (Figure 5), and our mathematical analysis confirmed the one-to-one correspondences (i-dependence) between the indices within these families. The three families of more than one index can be attributed to turnover, nestedness, and species richness contrast.

1. Jaccard index family. The largest family of indices is grouped around the Jaccard index (J; Jaccard, 1912). It has 9 indices of (dis)similarity, 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_{t2} (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 (Figure 5; Thesis T01.20 in Šizling et al., 2023) when standardized by $S_X + S_Y$ (Gotelli & McCabe, 2002). The indices of the Jaccard index family capture the phenomenon of species turnover (Thesis T02 in Šizling et al., 2023). All indices from the Jaccard index family that decrease with increasing J (e.g., β_{t2} or $\beta_{Sør}$) are measures of turnover (Figure 5). Whittaker beta-diversity index, when applied to pairwise comparison, also belongs to this family (Tuomisto, 2010b; Table R1 in Šizling et al., 2023).

2. Simpson beta family. The second large family consists of Simpson's beta (β_{Sim}; Simpson, 1943), Simpson's nestedness (β'_{Sim}; Lennon et al., 2001) and β_{jtu} which was previously referred to as the turnover component (Baselga, 2012). 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. The pairwise index *D* does not belong to any family if we compare just two assemblages. However, if we increase the number of sites being compared, then D (standardized as in Greve et al. 2005) belongs to the Simpson beta family, because it converges to the mean value across all β_{Sim} (Thesis T03 in Šizling et al., 2023). In practice, n of sites >10 guarantees

$$D \cong \overline{\beta_{Sim}}. \quad (1)$$

The pairwise N_C also belongs to the Simpson beta family if standardized as suggested by its authors (Wright & Reeves, 1992; Table R1 in Šizling et al., 2023), and if species richness is high (Thesis T04 in Šizling et al., 2023). All the indices from the Simpson beta family capture either the phenomenon of nestedness (Thesis T05 in Šizling et al., 2023), or reversed nestedness (β_{Sim} and D).

*3. Richness contrast family***.** The third family so far consists of just two indices: β_{gl} (Table 1; Lennon et al., 2001) and R (Table 1; Newbold et al., 2016). The β_{gl} captures the contrast in species richness between two sites (Thesis T06 in Šizling et al., 2023). 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 rather than β_{gl} in our equations(e.g., Equation 2).

4. Other families. There are several other families of indices that do not capture any phenomenon in focus (Thesis T07; Thesis T08; Thesis T09; Thesis T10; Thesis T11; Thesis T12, and Thesis T13 in Šizling et al., 2023). The largest of them consists of β_{rps} and β_{nps} (Table 1). Both the indices were introduced by Podani and Schmera (2011) and were called replacement (meaning the one-for-one replacement) and nestedness, respectively. In addition, there are families, each consisting of a single index (Figure 5). These are: Ruggiero index of beta-diversity (β_{Rib}; Ruggiero et al., 1998), 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 (Figure 5), which makes any inference difficult.

i-independent combinations of indices

Although we have delineated nine families of indices, it does not mean that there are nine 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 calculated from these two (Box 2). A clear case of mutual i-dependence of a triplet of indices is the combination of the Jaccard index family, the Simpson beta family and the

richness contrast family. A simple rearrangement of Equation 12 (where $R = I$ and $k_1, k_2, l_1, l_3 = 0,\ k_3, l_2 = 1$, see Box 2), gives

$$\frac{J}{1+J} = \beta'_{Sim}\frac{R}{1+R} \tag{2}$$

which says that the reversed turnover increases with increasing nestedness and species richness uniformity (reversed contrast). Similarly, for the one-for-one replacement (β_{rps}; Schmera et al., 2020), Equation 12 gives

$$2J(1 - \beta'_{Sim}) = \beta_{rps}\beta'_{Sim}. \tag{3}$$

The results (Equation 2, and Equation 3) can be generalized to all the considered indices that can be defined by Equation 8 (Box 2), which includes all dimensionless and continuous indices listed in Table 1 and Table R1 in Šizling et al., 2023, and even indices that have not yet been introduced. This does not include indices that have been produced using partitioning (e.g., β_{sne}), because the partitioning often produces i-dependency (for the algorithm see Box 3, and Thesis T14 in Šizling et al., 2023).

The algorithm in Box 3 provides a tool for converting indices from multiple studies into common terms, assuming that the authors calculated and published at least two i-independent presence-absence indices for their data (e.g., as in Xu et al., 2003).

Special cases: families of indices when assemblages are interrelated

Species assemblages are often interrelated due to similarity in habitats or due to dispersal. This may lead to two special cases in which the indices that would normally belong to different families appear as i-dependent. 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 contrast, i.e. $R = 1$ and Equation 2 turns into

$$\beta'_{Sim} = 2\frac{J}{1+J}\,. \tag{4}$$

This means that the Jaccard index and the Simpson beta families become apparently i-dependent (Figure R3 in Šizling et al., 2023). Second, assemblages can be perfectly nested, i.e., the species-poor assemblages have no unique species. In this case, $S_{X\cap Y} = \min(S_X, S_Y)$ and so $\beta'_{Sim} = 1$. It follows from Equation 2 that

$$\frac{J}{1+J} = \frac{R}{1+R} \Longrightarrow J = R. \quad (5)$$

This means that in such special cases the families of the Jaccard index and the richness contrast appear no longer i-independent (Figure R3 in Šizling et al., 2023). Third, pairs of assemblages can have equal proportion of shared species which gives

$$R^{-1} = C \cdot \beta'_{Sim} - 1, \quad (6)$$

where C is a constant. These special cases can easily be recognized in any data.

The apparent 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 i-dependent on each other (i.e., their values begin to cluster along a strictly monotonic relationship and thus become more s-dependent) as species richness values across sites become similar, and/or when assemblages approach complete nestedness (Figure 6).

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 with each other (group i: J, β_G; group ii: N_c, t, β_{sr}, β_{HK}, β_{BC}, $\beta_{Sør}$; group iii: β_{DICE}; Figure 5), and are statistically equivalent. Between these groups, however, the indices scale nonlinearly (see, e.g., J and $\beta_{Sør}$ in Figure 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 (Figure 5) with another index can also be seen as transformed values of the other index. For example, the β_t and β_{HK} represent identical transformations of J (Figure 5; Thesis T01.6 in Šizling et al., 2023, and Thesis T01.5 in Šizling et al., 2023). This means that the values of an index and its nonlinear transformation have different frequency distributions, which can affect parametric statistical tests and their sensitivity. This may

crucially affect conclusions based on real data with measurement bias (Chao et al., 2005, 2006; Chao & Colwell, 2017**).**

One option to deal with these issues would be 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 (Figure 5). We argue that it is better to first pick any index (within a given family) that well 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 subtracting the indices

Schmera and Podani (2011) showed that only indices with equal denominators are comparable, and thus can be summed or subtracted. The reason is that the denominator scales the nominator and thus it acts as 'a unit' in dimensionless indices. However, if we subtract (or sum) the indices with the same denominator, and if these indices are symmetric (sensu Chao et al., 2006) then we necessarily subtract or sum two i-dependent indices, i.e., indices that measure the same phenomenon (see Thesis T15 in Šizling et al., 2023 for details). Since subtracting or summing i-dependent indices leads to another i-dependent index, the operation does not provide any new information. For fundamental reasons why indices of different phenomena cannot be compared and thus summed together nor subtracted from each other see Box 4.

Inadequacy of partitioning the indices

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 "overall beta-diversity" into the nestedness-resultant and turnover components (Baselga, 2010, 2012). The reason is that interdependence between phenomena is an intrinsic

characteristic of the phenomena, where different aspects bound each other and removing the interdependence necessarily alters the phenomena itself. For instance, inequality in species richness (species richness contrast) bounds possible values of species turnover.

More specifically, in all situations when we compare two regions or periods, or test for (non)randomness in a 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. The ideas demonstrated here are then universally applicable, since any pair of i-independent indices is convertible to a pair of indices J, β'_{Sim}, or R (Box 3). S-dependence (including relatedness; Chao et al., 2012) between these indices is driven by inequalities:

$$0 \leq J =_1 R\frac{\beta'_{Sim}}{R+1-R\beta'_{Sim}} \leq_2 \frac{\beta'_{Sim}}{2-\beta'_{Sim}}, \text{ and } 0 \leq J \leq_3 R. \quad (7)$$

The first equality ($=_1$) results from Equation 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 inequality $\leq_3$ is well known relatedness between J and the contrast between alpha-diversities, which has led to the definition of β_{Sim} (Table 1). Both bounds ($\leq_{2,3}$) have been taken as an argument that J and β_{Sim} must be corrected for the effect of the other indices (e.g., Simpson, 1943; Baselga, 2010). We argue that, when we are interested in turnover (not nestedness nor other phenomena), then the indices of the Jaccard family must not be corrected for variation in species richness contrast, regardless of any interdependence between these phenomena. The rationale is that (i) the J is a measure of reversed turnover regardless of its relatedness with R (i.e., *J* meets the reversed constraints of turnover in Figure 3; Thesis T15 in Šizling et al., 2023), and that (ii) the bounds ($\leq_{2,3}$) only show the limits of 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* (Figure 7), so using this relative *J* is equivalent to direct measurement of *R*.

A general intuition illuminating the inadequacy of the partitioning of indices (defined by Equation 8) stems from the realization that any arrangement of assemblages is fully described by two i-independent indices (plus species richness as a scaling parameter). This implies that if we have two i-independent indices, there are no degrees of freedom for residual information about the assemblage arrangement which would follow from the partitioning. A partitioned index that combines two i-independent indices thus does not measure a phenomenon purified from the effect of the other phenomenon, but a combination of the phenomena, which is a different phenomenon. A good example is the attempt of Simpson (1943) to purify turnover from species richness contrast, which has led to β'_{Sim}, i.e. the index which directly measures nestedness (Almeida-Neto et al., 2008). Any partitioning thus leads either (1) to an index that measures one of the phenomena we already measure by another established index (as in the above case) or (2) an index that is not strictly monotonically related to any of the phenomena that we want to measure (as in Baselga´s approach in Figure 1).

Empirical examples: How to make inference from the indices?

To provide real world examples of an inference based on similarity indices, we use data from Šizling et al. (2016) who studied temporal changes of Central European plants during Holocene, and Xu et al. (2023) who reported global patterns of angiosperm trees. The main purpose of these examples is to show that the increasingly popular partitioning of these indices (Baselga, 2010, 2012) may lead to interpretations that do not correspond to any of the four phenomena intuitively distinguished by ecologists (section *Predefinitions of four diversity phenomena*).

Example 1: Temporal change in Holocene plants. Šizling et al. (2016) reported a temporal increase in mean species richness $\langle S \rangle$ from 27 to 32, almost no change of β'_{Sim} ($\beta'_{Sim} \cong 0.67$), and a steep increase of average J from ca. 0.4 to ca. 0.5 during the short period (5,900-5,700 BP) when agriculture was introduced in the region. J and β'_{Sim} uniquely determine *R* (Equation 2), which increased from 0.74 to 0.99 (Figure 8). In addition, *R* uniquely determines the contribution of species-poor

($S_{poor} = \langle S \rangle\, 2R/(R+1)$) and species-rich ($S_{rich} = R^{-1} \cdot S_{poor}$) sites to the average alpha diversity (Thesis T16 in Šizling et al., 2023). S_{poor} increased from ca. 23 to 32, and S_{rich} increased from ca. 31 to 32. We conclude that recent plant assemblages in Central Europe show lower spatial turnover (Figure 8c), and higher uniformity (Figure 8c) than those before the introduction of agriculture. Since the average nestedness (Figure 8c) hasn't changed from before agriculture until today, the species poor sites have also kept a constant ratio of unique to common species. The introduction of agriculture was therefore accompanied by homogenization of plant assemblages, with proportionally higher increase of richness at species-poor sites, and increasing average species richness of all sites. This homogenization of assemblages as a consequence of agriculture is in agreement with other analyses (e.g., Kolář et. al., 2022) and corresponds to the visual inspection of Figure 8a.

When we apply Baselga's (2012) partitioning method to the data of Šizling et al. (2016), we get $\beta_{sne} = 0.1$, $\beta_{Sim} = 0.33$ for the pre-agricultural landscape and $\beta_{sne} = 0.00$, $\beta_{Sim} = 0.33$ for early agricultural landscape. This reads in the Baselga's (2010, 2012) framework as no temporal change of "turnover component", and decreases in both the "nestedness-resultant component" and total beta-diversity (Figure 8b). This interpretation based on Baselga's (2010, 2012) partitioning thus directly contradicts our framework (Figure 8c). Baselga's framework misses the increasing uniformity of species richness (due to the increase of species richness of species-poor assemblages) which seems to be driving the other observed patterns.

Example 2: Global patterns of trees. Xu et al. (2023) used the Baselga (2010) framework and reported an increase in the regression line of the nestedness-resultant component (β_{sne}) with an increasing temperature anomaly from ca. 0.11 to 0.32, and a decrease in the regression line of the turnover component (β_{Sim}) from ca. 0.68 to 0.48 (Figure 8e). The total beta-diversity ($\beta_{sne} + \beta_{Sim}$) did not show a significant trend along the temperature anomaly gradient.

To reanalyze Xu et al.'s (2023) data, we used our framework (Box 2; Box 3, and Calculator RI4 in Šizling et al., 2023) and computed the average S_X, S_Y, and $S_{X\cap Y}$ that correspond to the indices reported by Xu et al. (2023; Figure 8e). Since Xu et al. (2023) did not report richness, we standardized it to $S_X + S_Y - S_{X\cap Y} \cong 100$ without loss of generality. Our reanalysis of the data from Xu et al. (2023) suggests no significant trend in turnover (in contrast to a significant decrease from Baselga's framework), increasing nestedness (in line with Baselga's framework), and decreasing richness uniformity with an increasing temperature anomaly (Figure 8f). This indicates a heterogenization of assemblages as the temperature anomaly increases. Visual inspection of the arrangement of species assemblages (Figure 8d) indicates the pattern of increasing contrast in species richness of both assemblages (i.e. decreasing uniformity) associated with increasing nestedness, and no pattern of species turnover (as the overlap is constant). This agrees with our interpretation of the indices, but is in striking contrast to the interpretation within Baselga's framework.

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 temporal, spatial or environmental gradients. Based on previous considerations, here are practical recommendations for using the presence-absence indices:

1. Avoid partitioning of indices that supposedly removes an effect of one phenomenon from an effect of another phenomenon (e.g., purifying turnover from the effect of richness contrast 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. For instance, standardization of Jaccard index by its maximum determined by the nestedness (bold line in Figure 7) is in fact a measure of species richness contrast (Figure 7). The reason is that nestedness, together with richness contrast *R*, uniquely determine *J*, so that measuring *J* relative to *Jmax* (bounded by nestedness) is equivalent to measuring *R*.

2. When choosing an index, first consider which phenomenon it should capture (Table 1; Table 2), then select a corresponding family of indices (see Figure 5; Table R1 in Šizling et al., 2023; Manual RI3 in Šizling et al., 2023, and Calculator RI4 in Šizling et al., 2023). It does not matter which index within the family is selected, as all indices within any family can be converted to each other and are thus practically equivalent. If there is a need for an appropriate statistical distribution of the index, a proper transformation is preferable to an invention of, or search for, a new index.
3. When comparing already published indices, use the equations in Table R1 in Šizling et al. (2023) or the Calculator RI4 in Šizling et al. (2023) to convert them to a common index that is the most desirable for a given purpose. If a desirable common index is not available in the publication, the desirable index can be calculated from any i-dependent index, or any two i-independent indices (Box 2 and Box 3). The indices of two different phenomena (two i-independent indices) are never commensurable.
4. Use null models that randomize species incidences to assess the statistical significance of the phenomenon in focus (Gotelli & Ulrich, 2012). Null models are useful for distinguishing inherent correlations (Chao et al., 2012) from ecologically meaningful correlations and relationships. Note, however, that relevant null models are still under development (e.g. Chase et al., 2011; Legendre, 2019), and it may not always be clear what exactly to randomize in the site-by-species matrix.

5. 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. Use our Calculator RI4 in Šizling et al. (2023) when in doubt about whether the employed indices carry complete information.

Discussion

We demonstrated that the phenomena that are measured by different indices constrain each other at the conceptual level, 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 any effort to eliminate s-dependence (or relativize the index in focus) results in a measure of a different phenomenon. 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 desirable, 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}, or R; Box 2 and Box 3), and there is thus no need for new indices. We have mathematically proven that two i-independent indices, in combination with a species richness value (which acts as a scaling parameter), provide complete information about the assemblages. This is not true for the partitioned indices because the partitioning often reduces the amount of information. 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), i-independent indices appear as if they were i-dependent, and can be derived from each other. 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 of indices almost indistinguishable (Figure 6; Figure R3 in Šizling et al., 2023). 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; Tuomisto, 2010a; Tuomisto, 2010b; 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 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 section *Results*) to make a proper inference. These indices (for example J and R) can be, if necessary, converted to any other index within their respective family (and vice versa). However, we have shown that indices from different families cannot be directly compared to each other, as their values characterize different phenomena. Different indices within a family of i-dependent indices are different transformations of each other and their particular values thus reflect just different rescaling.

We have focused only on pairwise indices that compare spatially or temporarily non-overlapping assemblages and have avoided the topic of beta-diversity *sensu stricto* (Whittaker, 1960), which comprises the relationship between local and regional species richness, or, more specifically, between alpha and gamma diversity. This relationship is to some extent related to turnover, since when alpha is considerably lower than gamma, there must be high turnover among communities. However, the precise mathematical links between pairwise community turnover and Whittaker's beta is a separate issue (Koleff et al., 2003; Tuomisto, 2010b; Chao et al., 2012) and there is no straightforward way to derive turnover from beta-diversity *sensu stricto* for more than two assemblages (Šizling et al., 2011).

It is fundamental that two i-dependent indices cannot quantify two different phenomena. An index that is i-dependent on another index can be viewed as its rescaled version, and no rescaling can affect the phenomenon, since rescaling can be done at any time after the observation. This is in contrast to Tuomisto (2012) who defines 'conceptual independence' of indices (Box 1) to delimit her phenomena. In our

framework, the phenomena are defined more broadly using the constraints in Figure 3. If two indices obey the particular set of constraints, they quantify the same phenomenon. Mutually i-dependent indices necessarily satisfy the same constraints and thus they quantify the same phenomenon.

Our findings regarding the pairwise comparison of two assemblages can also be applied to pairwise comparisons of two species ranges, and generally to all pairwise indices of interspecific spatial association derived from binary data (Keil et al. 2021). In particular, S_X, S_Y, and $S_{X \cap Y}$ can represent the number of sites occupied by the first species, second species, and both species together, respectively. For instance, the co-occurrence of two species over multiple sites (Diamond, 1975; Connor & Simberloff, 1979; Diamond & Gilpin, 1982; Gotelli & McCabe, 2002) can be measured by the Jaccard index (Thesis T17 in Šizling et al., 2023). Analogically, indices from the Simpson beta family and richness contrast family can be used for characterizing nestedness of species' geographic ranges (Šizling et al., 2009) and the contrast in species extent of occurrence.

The relationships between indices and families have important implications for understanding the forces shaping distance decay (a strict increase or decrease of an index with increasing distance) in assemblage similarity (Nekola & White, 1999). A direct consequence of Equation 2 is that distance decays of nestedness (β'_{Sim}) and reversed turnover (measured by $J/(1+J)$) follow the same mathematical function when there is no distance decay in the species richness contrast. This happens only in environments without geographical gradients that affect the species richness, for example, the gradient in productivity (e.g., Willig et al., 2003; Currie et al., 2004; Bohdalková et al., 2021). If there are richness gradients, then 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.

Our results are relevant for the interpretation of hundreds of empirical assessments that used the subtractive partitioning of pairwise indices. To date, the original publication describing the partitioning to nestedness and turnover components (Baselga, 2010) has 3,365 citations, the publication announcing the R package betapart which does the partitioning (Baselga & Orme, 2012) has 1,884 citations, and the R package on CRAN (Baselga et al., 2018) has 852 citations (Google Scholar accessed on 3rd of April 2024). 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; Xu et al., 2023). The approach has also been gaining momentum in young fields, for example in microbial ecology (Shade et al., 2013). Based on the conceptual problems that we have described, we call for a critical re-evaluation of the practice. A need may also arise for re-analyzing and re-interpreting some of the studies; in this effort, our equations from Box 2, and Box 3 (or Calculator RI4 in Šizling et al., 2023) 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.

Acknowledgements: Special thanks to A. Bryson and G. Ridder for help with English, F. He and W. E. Kunin for their comments on the first versions, Z. Sokolíčková, E. Šizlingová for her help in the field, J. Klimešová and A. Bernard for lessons in Arctic plants, and R.K. Colwell and anonymous reviewers for insightful comments. The project was supported by the Czech Science Foundation (EXPRO 20-29554X and GA ČR 19-21341S). We used the Czech Scientific Infrastructure of the University of South Bohemia, J. Svoboda house on Svalbard – CzechPolar2 LM2015078 supported by MŠMT. PK was funded by the European Union (ERC, BEAST, 101044740). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Author Contributions:

AS, DS, and ET conceived the ideas. AS and DS worked out the definitions, concepts, and proofs. ET, JDŽ and ALŠ collected and provided the empirical data. ALŠ did the analyses, simulations, and coded the software tool. ALŠ, DS and PK led the writing with contributions from ET and KMCT.

References

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

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

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

Audi, R. 1999. *The Cambridge Dictionary of Philosophy*. 2nd ed., Cambridge University Press, Cambridge, UK.

Barwell, L. J., N. J. B. Isaac and W. E. Kunin. 2015. "Measuring ß-diversity with species abundance data." *Journal of Animal Ecology* 84: 1112-1122.

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

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

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

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

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

Bohdalková, E., A. Tószögyová, I. Šímová and D. Storch. 2021. “Universality in biodiversity patterns: variation in species–temperature and species–productivity relationships reveals a prominent role of productivity in diversity gradients.” *Ecography* 44: 1366–1378.

Bray, J. R. and J. T. Curtis. 1957. “An ordination of the upland forest communities of southern Wisconsin.” *Ecological monographs* 27: 325-349.

Brualdi, R. A. and J. G. Sanderson. 1999. “Nested species subsets, gaps, and discrepancy.” *Oecologia* 119: 256-264.

Carvalho, J. C., P. Cardoso and P. Gomes. 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., R. L. Chazdon, R. K. Colwell and T.-J. Shen. 2005. “A new statistical approach for assessing compositional similarity based on incidence and abundance data.” *Ecology Letters* 8: 148-159.

Chao, A., R. L. Chazdon, R. K. Colwell and T.-J. Shen. 2006. “Abundance-based similarity indices and their estimation when there are unseen species in samples.” *Biometrics* 62: 361-371.

Chao, A., C. H. Chiu and T. C. Hsieh. 2012. “Proposing a resolution to debates on diversity partitioning.” *Ecology* 93: 2037-2051.

Chao, A., C. H. Chiu and L. Jost. 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.

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

Chao, A. and R. K. Colwell. 2017. “Thirty years of progeny from Chao’s inequality: Estimating and comparing richness with incidence data and incomplete sampling.” *Statistics and Operations Research Transactions* 41: 1–52.

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

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

Chiu, C.H., L. Jost and A. Chao. 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*, edited by M.L. Cody, and J.M. Diamond, 214–257. Harvard University Press, Cambridge, Massachusetts, USA.

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

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

Diamond, J. M. 1975. “Assembly of species communities.” In *Ecology and Evolution of Communities* edited by M.L. Cody, and J.M. Diamond, 342-444. Harvard University Press, Cambridge, Massachusetts, USA.

Diamond, J. M. and M. E. Gilpin. 1982. “Examination of the ‘‘null’’ model of Connor and Simberloff for species cooccurrences on islands.” *Oecologia* 52: 64–74.

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

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

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

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

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

Gotelli, N. and W. Ulrich. 2012. “Statistical challenges in null model analysis.” *Oikos* 121: 171-180.

Gotelli, N. J., H. Shimadzu, M. Dornelas, B. McGill, F. Moyes and A. E. Magurran. 2017. “Community-level regulation of temporal trends in biodiversity.” *Science Advance*s 3:e1700315.

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

Harrison, S., S. J. Ross and J. H. Lawton. 1992. “Beta diversity on geographic gradients in Britain.” *Journal of Animal ecology* 61: 151-158.

Harte, J. and A. P. Kinzig. 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.

ISO. 2009. ‘*ISO 80000-1:2009(en) Quantities and units — Part 1: General*’ International Organization for Standardization. (https://www.iso.org)

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

Jost, L. 2007. “Partitioning diversity into independent alpha and beta components.” *Ecology* 88: 2427-2439.

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

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

Kolář, J., M. Macek, P. Tkáč, D. Novák and V. Abraham. 2022. “Long-term demographic trends and spatio-temporal distribution of past human activity in Central Europe: Comparison of archaeological and palaeoecological proxies.” *Quaternary Science Reviews* 297: 107834

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

Koleff, P., K. J. Gaston and J. J. Lennon. 2003. “Measuring beta diversity for presence–absence data.” *Journal of Animal Ecology* 72: 367-382.

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

Legendre, P. 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. *Ecology and Evolution* doi: 10.1002/ece3.4984.

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

Lyashevska, O. and K. D. Farnsworth. 2012. “How many dimensions of biodiversity do we need?” *Ecological Indicators* 18: 485-492.

Mari, L., M. Wilson, and A. Maul. 2023. *Measurement across the sciences*. 2nd ed., Springer Nature, Switzerland AG.

Matthews, T., T. Aspin, W. Ulrich, A. Baselga, Y. Kubota, K. Proios, K. A. Triantis, R.J. Whittaker, and G. Strona. 2019. “Can additive beta diversity be reliably partitioned into nestedness and turnover components?” *Global Ecology and Biogeography* doi: 10.1111/geb.12921.

Molinos, J. G., B. S. Halpern, D. S. Schoeman, C. J. Brown, W. Kiessling, P. J. Moore, J. M. Pandolfi, E. S. Poloczanska, A. J. Richardson and M. T. Burrows. 2016. “Climate

velocity and the future global redistribution of marine biodiversity." *Nature Climate Change* 6: 83–88.

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

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

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

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

Orlitsky, A. 2003. "Information Theory." In *Encyclopedia of Physical Science and Technology* edited by R.A. Meyers, 751-769. Elsevier Science Ltd., Academic Press.

Patterson, B. D. 1984. "Mammalian extinction and biogeography in the Southern Rocky Mountains." In *Extinctions* edited by M.H. Nitecki, 247-293. University of Chicago Press, Chicago.

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

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

Podani, J., and D. Schmera. 2012. "A comparative evaluation of pairwise nestedness measures." *Ecography* 35: 889–900.

Qian, H., C. Badgley and D. L. Fox. 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. and R. E. Crick. 1979. "Measurement of faunal similarity in paleontology." *Journal of Paleontology* 53: 1213-1227.

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

Ruggiero, A., J. H. Lawton and T. M. Blackburn. 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. and R. E. Ricklefs. 1993. “Species diversity: an introduction to the problem.” In *Species diversity in ecological communities: Historical and Geographical Perspectives*, edited. by R.E. Ricklefs and D. Schluter, 1–12. Chicago University Press, Chicago.

Schmera, D. and J. Podani. 2011. “Comments on separating components of beta-diversity.” *Community Ecology* 12: 153-160.

Schmera, D., J. Podani, and P. Legendre. 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., P. S. McManus and J. Handelsman. 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., E. Šizlingová, D. Storch, J. Reif and K. J. Gaston. 2009. “Rarity, commonness, and the contribution of individual species to species richness patterns.” *The American Naturalist* 174: 82-93.

Šizling, A.L., Kunin, W., Šizlingová, E., Reif, J. & Storch, D. 2011. “Between geometry and biology: the problem of universality of the species-area relationship.” *American Naturalist* 175: 602-609.

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

Šizling, A. L., P. Keil, E. Tjørve, K. M. C. Tjørve, J. D. Žárský, and D. Storch. 2023. "Mathematically and biologically consistent framework for presence-absence pairwise indices – Repository Information." *Zenodo* https://doi.org/10.5281/zenodo.8084396

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., K. M. C. Tjørve, E. Šizlingová and A. L. Šizling. 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.

Tuomisto, H. 2010a. "A diversity of beta diversities: straightening up a concept gone awry. Part 1. Quantifying beta diversity and related phenomena." *Ecography* 33: 2-22.

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

Tuomisto, H. 2012. "An updated consumer's guide to evenness and related indices." *Oikos* 121: 1203–1218.

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

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

Ulrich, W. and N. J. Gotelli. 2013. "Pattern detection in null model analysis." *Oikos* 122: 2-18.

Ulrich, W., A. Baselga, B. Kusumoto, T. Shiono, H. Tuomisto and Y. Kubota. 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., Y. Kubota, B. Kusumoto, A. Baselga, H. Tuomisto and N. J. Gotelli. 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.

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

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

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

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

Xu. W., W. Guo, J. M. Serra-Diaz, F. Schrodt, W. L. Eiserhardt, B. J. Enquist, B. S. Maitner, C. Merow, C. Violle, M. Anand, M. Belluau, H. H. Bruun, C. Byun, J. A. Catford, B. E. L. Cerabolini, E. Chacón-Madrigal, D. Ciccarelli, J. Hans, C. Cornelissen, A. T. Dang-Le, A. de Frutos, A. S. Dias, A. B. Giroldo, A. G. Gutiérrez, W. Hattingh, T. He, P. Hietz, N. Hough-Snee, S. Jansen, J. Kattge, B. Komac, N. J. B. Kraft, K. Kramer, S. Lavorel, C. H. Lusk, A. R. Martin, K. Ma, M. Mencuccini, S. T. Michaletz, V. Minden, A. S. Mori, Ü. Niinemets, Y. Onoda, R. E. Onstein, J. Peñuelas, V. D. Pillar, J. Pisek, M. J. Pound, B. J. M. Robroek, B. Schamp, M. Slot, M. Sun, Ê. E. Sosinski Jr., N. A. Soudzilovskaia, N. Thiffault, P. M. van Bodegom, F. van der Plas, J. Zheng, J-C. Svenning, A. Ordonez. 2023. "Global beta-diversity of angiosperm trees is shaped by Quaternary climate change." *Science Advances* eadd8553.

BOX 1 Key concept: The informational dependence of indices

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 (Orlitsky, 2003). We thus define informational dependence (i-dependence) of two variables as a situation when these variables are strictly monotonic transformations of each other (Figure 2). A strict monotony of two variables implies equality of their informational values, and when two indices have equal information content, they cannot capture different phenomena.

In contrast, the kind of independence that is used in ecological literature is almost always of statistical nature, and we call it *s-dependence*. While i-independence refers to a mathematical formula, 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).

Chao et al. (2012) distinguished statistical dependence from a dependence of mathematical formulas of indices; two indices are statistically dependent when one index value affects probability density of values of another index. Chao et al. (2012) use the term *relatedness* for dependence of the formulas, and defined it as a situation where the minimum or maximum of one index is affected by the variation of another index. Relatedness, however, is one of the mechanisms that induces s-dependence, since a change in the minimum and maximum possible value of an index has an effect on its probability density, which is the definition of statistical dependence (Figure 2). Importantly, relatedness says nothing about equal or different information carried by the variables in focus. For example, J and β'_{Sim} are i-independent, while their values are s-dependent (their values are correlated) because their definitions are related sensu Chao et al. (2012) (Figure 4a,b).

Tuomisto (2012) defines the independence of indices based on different defining formulas. According to Tuomisto (2012), two different formulas , e.g. $I_1 = (A - B)/B$ and $I_2 = A/B$, are conceptually independent of each other and quantify two different phenomena. The reason is that I_1 reads as 'relative difference' but I_2 reads as 'proportion'. However, in our terminology I_1 and I_2 are i-dependent and define different measures of the same phenomenon, because $I_1 = I_2 - 1$, and whatever can be inferred from I_1 can also be inferred from I_2.

BOX 2 *How many indices do we need?*

For a given system, a 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 (the i-independence from Box 1). 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 Table R1 in Šizling et al. (2023) are i-dependent on the pair of Jaccard index, *J* (Table 1) and Simpson nestedness, β'_{Sim} (Table 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 can be defined as a ratio of two linear functions (see also Table R1 in Šizling et al., 2023):

$$I \stackrel{\text{def}}{=} \frac{k_1 S_{X\cap Y}+k_2 S_X+k_3 S_Y}{l_1 S_{X\cap Y}+l_2 S_X+l_3 S_Y}, \quad (k_{1-3},\ l_{1-3} \in \boldsymbol{R}), \qquad |l_2|+|l_3|>0 \ \ , S_X \geq S_Y > 0. \tag{8}$$

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: $I=\frac{a}{b} \ \rightarrow bI-a=0$, Equation 8 can be converted into the linear equation

$$S_X(Il_2-k_2)+S_Y(Il_3-k_3)+S_{X\cap Y}(Il_1-k_1)=0, \tag{9}$$

where S_X, S_Y, $S_{X\cap Y}$ are the unknowns and the *I* is a particular value of an index. Equation 9 turns into

$$S_X J+S_Y J-S_{X\cap Y}(J+1)=0,\ S_Y\beta'_{Sim}-S_{X\cap Y}=0 \text{ and } S_X R-S_Y=0 \tag{10}$$

for *J*, β'_{Sim} and R, respectively. The two equations for two indices (Equation 10), 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 a nonzero right side to get unique information on S_X, S_Y, $S_{X\cap Y}$. This equation is

$$S_X+S_Y=2\bar{S}, \tag{11}$$

where $\bar{S}\neq 0$ is the expected alpha diversity (for the proof that Equation 11 is i-independent of Equation 10; see Thesis T18 in Šizling et al., 2023). The three equations (any two equations from Equation 10, and Equation 11) determine $S_X, S_Y, S_{X\cap Y}$ uniquely, and thus no other index (even if

its definition does not follow Equation 8) can further specify the solution. Moreover, if we focus on the indices which can be expressed using Equation 8, we can compute the value of any other index from the values of *J* and β'_{Sim} or R as

$$I = \frac{J(\beta'_{Sim}(k_1+k_2)-k_2+k_3)+\beta'_{Sim}k_2}{J(\beta'_{Sim}(l_1+l_2)-l_2+l_3)+\beta'_{Sim}l_2} \text{ for } J \neq 0 \text{ else } I = \frac{k_2+k_3R}{l_2+l_3R} \quad (12)$$

where k_i, l_i define the new index *I* (for the proof see Thesis T19 in Šizling et al., 2023).

BOX 3 *How to unify indices from across the literature*

A large number of different indices are spread over the literature (e.g., Gaston et al., 2007; Baselga, 2012; Schmera et al., 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 that utilizes at least two i-independent indices, I_1, I_2, that are expressed as in Equation 8 (see Box 2). Then we write a system of three independent equations: two equations Equation 9, each for one of the indices, and one scaling equation Equation 11. If $\bar{S}$ is missing from the publication, we can put $\bar{S} = 1$ without loss of generality. The reason is that the indices to be converted (Equation 8) are i-independent of species richness and therefore the exact value of $\bar{S}$ does not matter in the case. The third equation then is

$$S_X + S_Y = 2. \tag{13}$$

The solution of the three equations (see *Gauss Elimination Method*) 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 ($S_{Xorig} = \bar{S}\, S_{Xcomp}$ 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 for an index that cannot be expressed as Equation 8 (e.g., $I = S_X S_{X\cap Y}/S_Y^2$). Some special cases are the indices that originated from additive partitioning ($I_1 = I_a - I_b$) such as β_{sne}, β_{jne}, of β_{nps} (Table 1). These indices cannot be universally expressed as Equation 8, and their equations 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 14

$$I_b^2 D_{LL} - I_b(I_1 D_{LL} - D_{KL} + D_{LK}) + I_1 D_{LK} - D_{KK} = 0 \tag{14}$$

where D_{XY} are determinants listed in Thesis T14 in Šizling et al. (2023). 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 Equation 14 may have two realistic solutions (both solutions are within minimum and maximum possible range of I_b), the partitioning often leads to the loss of information. For example, if Šizling et al. (2016) had only reported two partitioned indices (e.g., $\beta_{sne} = 0.1$ and $\beta_{nps} = 0.61$ for pre-agricultural landscape, and $\beta_{sne} = 0.0$ and $\beta_{nps} = 0.51$ for agricultural landscape, it would not be sufficient to infer the homogenization of assemblages. This is because these two indices do not distinguish

between an increase or decrease in uniformity (R), as R could have changed from 0.74 to either 1 or 0.32, respectively (see Thesis T20 in Šizling et al., 2023 for further details).

BOX 4 *Different phenomena are incomparable*

There are several reasons why indices of different phenomena are incomparable. One of them is that the relationship between two phenomena (e.g., turnover and nestedness) is modulated by an index of a third phenomenon (e.g., species richness contrast), which is necessarily i-dependent on the two phenomena in comparison. This interdependence means that the indices cannot be viewed in isolation (Table 2). An example of the relationship of three indices is Equation 2, which can be transformed to involve any triplet of similarity indices (see Box 2 for the algorithm). We cannot put the equation sign nor inequality sign between the indices of nestedness and turnover without taking into account the effect of species richness contrast (Podani & Schmera, 2012). It is not β'_{Sim} but the product of β'_{Sim} and $R/(R+1)$ that is comparable with $J/(J+1)$. Similarly, no one would directly compare U with I in Ohm's law, because the relationship between U and I is modulated by R, $U = RI$ (Figure 7).

The most fundamental reason, however, stems from the theory of measurement (ISO, 2009; Mari et al., 2023), which states that: (i) there are different concepts of measurement (kinds of quantities) that are incomparable even though they may have the same units (e.g., torque and work are incomparable but have the same unit, the Newton meter), and that (ii) the dimensionless quantities are not commensurable if they capture different phenomena.

The phenomena addressed here capture different concepts of similarity between assemblages (they are different kinds of quantities) and are therefore incomparable. The reason is that they are defined using different sets of constraints (Figure 3). In particular, turnover is conceptually defined as species overlap ($S_{X\cap Y}$) relative to the richness of both assemblages (S_X, S_Y) (e.g., relative to total richness in the case of Jaccard, and mean richness in the case of Sørensen index). Unlike turnover, nestedness is conceptually defined only by the property of the poorer assemblage ($\min(S_X, S_Y)$, $S_{X\cap Y}$), which makes it i-independent of the species richness of the richer assemblage ($\max(S_X, S_Y)$)). This means that the index of turnover must include both S_X and S_Y in the denominator (Thesis T21 in Šizling et al., 2023), whereas the denominator of the index of nestedness can include either S_X or S_Y, but not both. Different denominators make the indices of turnover and nestedness incomparable. This incomparability of different phenomena implies that their quantities cannot be summed nor subtracted.

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. For index definitions using the a,b,c notation ($a = S_{X\cap Y}$, $b = S_X - a$; $c = S_Y - a$) see Table R1 in Šizling et al. (2023). The labelling follows Gaston et al. (2007) and the symbol '$\stackrel{\text{def}}{=}$' represents definitions. For a complete list of indices and their mutual relationships see Table R1 in Šizling et al. (2023). 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,
$\beta_G \stackrel{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}} = 1 - J$	Jaccard dissimilarity, Gaston et al. (2001), labeled as in Gaston et al. (2007)	turnover
$\beta_{Sør} \stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y} = 2\frac{J}{1+J}$	Sørensen similarity, Sørensen (1948)	reversed turnover
$\beta_{t2} \stackrel{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y} = 1 - \beta_{Sør}$	Wilson & Shmida (1984), for two sites, labelled as in Gaston et al. (2007), Baselga (2010) and Tuomisto (2010) call this index Sørensen dissimilarity and Whittaker's species turnover, respectively	turnover
$\beta_{Sim} \stackrel{\text{def}}{=} \frac{\min(S_X, S_Y) - S_{X\cap Y}}{\min(S_X, S_Y)}$	Simpson beta, Simpson dissimilarity, Simpson (1943), Koleff et. al. (2003) turnover component of β_{t2}, Baselga (2012)	reversed nestedness
$\beta'_{Sim} \stackrel{\text{def}}{=} \frac{S_{X\cap}}{\min(S_X, S_Y)} = 1 - \beta_{Sim}$	Simpson similarity, Simpson nestedness, Simpson (1943), Lennon et al. (2001)	nestedness
$\beta_{jtu} \stackrel{\text{def}}{=} \frac{2\min(S_X, S_Y) - 2S_{X\cap Y}}{2\min(S_X, S_Y) - S_{X\cap Y}} = = 2\beta_{Sim}/(1 + \beta_{Sim});$	turnover component of Jaccard dissimilarity, Baselga (2012)	reversed nestedness
$\beta_{gl} \stackrel{\text{def}}{=} 2\frac{\lvert S_X - S_Y \rvert}{S_X + S_Y}$	Gaston and Lennon index of species richness gradient, Lennon et al. (2001)	species richness contrast, reversed uniformity

$R \stackrel{\text{def}}{=} \frac{\min(S_X, S_Y)}{\max(S_X, S_Y)} = \frac{2-\beta_{gl}}{2+\beta_{gl}}$	Newbold et al. (2016)	species richness uniformity, reversed contrast
$\beta_{sne} \stackrel{\text{def}}{=} \beta_{t2} - \beta_{Sim}$	nestedness resultant component of β_{t2}, Baselga (2012), labeled β_{NES} in Baselga (2010)	(*)
$\beta_{jne} \stackrel{\text{def}}{=} \beta_G - \beta_{jtu}$	nestedness resultant component of Jaccard dissimilarity, Baselga (2012)	(*)
$\beta_{rps} \stackrel{\text{def}}{=} 2\frac{\min(S_X, S_Y) - S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	(one-for-one) replacement, labeled as R_{rel} in 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, labeled as N_{rel} in Podani & Shmera (2011)	(*)
$N_{NODF} \stackrel{\text{def}}{=} \beta'_{Si}$ if $S_X \neq S_Y$, else $N_{NODF} = 0$	nestedness by Almeida-Neto et al. (2008)	(*)
$N_0 \stackrel{\text{def}}{=} \frac{\min(S_X,S_Y)-S_{X\cap Y}}{S_X+S_Y}$ if $S_X \neq S_Y$ else $N_0 \stackrel{\text{def}}{=} 0$.	nestedness by Patterson & Atmar (1986)	(*)

TABLE 2 Table of possible inference from the variation of pairwise indices along temporal, geographic, or environmental gradients. Some inference can be based on variation of a single index, regardless of other indices (first three rows). However, more specific inference follows from situations when one index changes along the gradient, while the other remains constant or shows only negligible variation (all remaining rows).

Variation along a gradient		**Inference**
Yes	**No**	
J		Change in turnover between regions/periods, change in proportion of shared species relative to species richness of both assemblages
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.
R		Change in species richness contrast (uniformity) between regions/periods; change in 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; 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 Equation 2 induce necessarily a change in R
J	β'_{Sim}, S_{poor}	Change in turnover but not nestedness; change in list of species that are unique to species-richer sites; necessarily a change in R

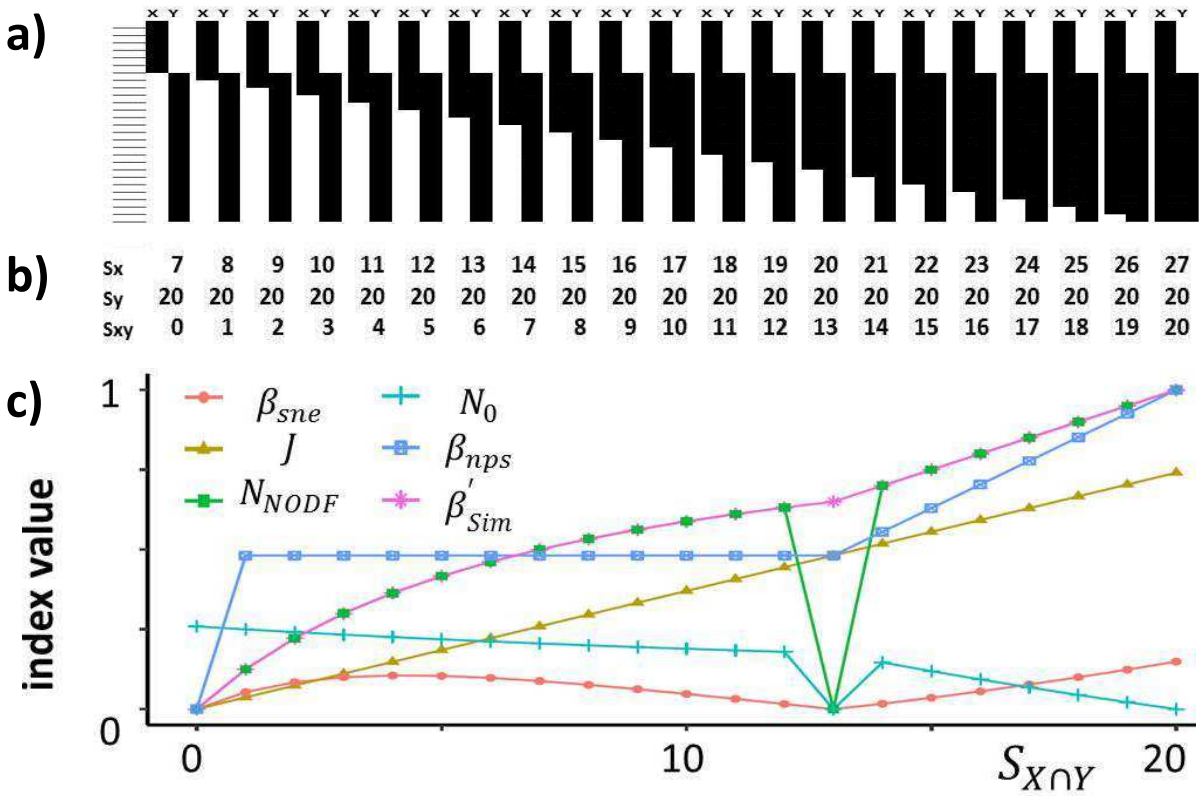

FIGURE 1 The requirement of continuity (see section *Problems with previously suggested indices*), and consequences of its violation. The top plot (a) shows a sequence of 20 pairs of assemblages X and Y, with a continuous change of the arrangement of assemblages, increasing similarity in species composition and nestedness from left to right, and constant total species richness. Columns show assemblages, lines species and blackened cells presences of species within assemblages. The 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 the values of β_{sne} (red circles), J (olive triangles), N_{NODF} (green squares), N_0 (green-blue crosses), β_{nps} (blue squares), and β'_{Sim} (purple asterisks).

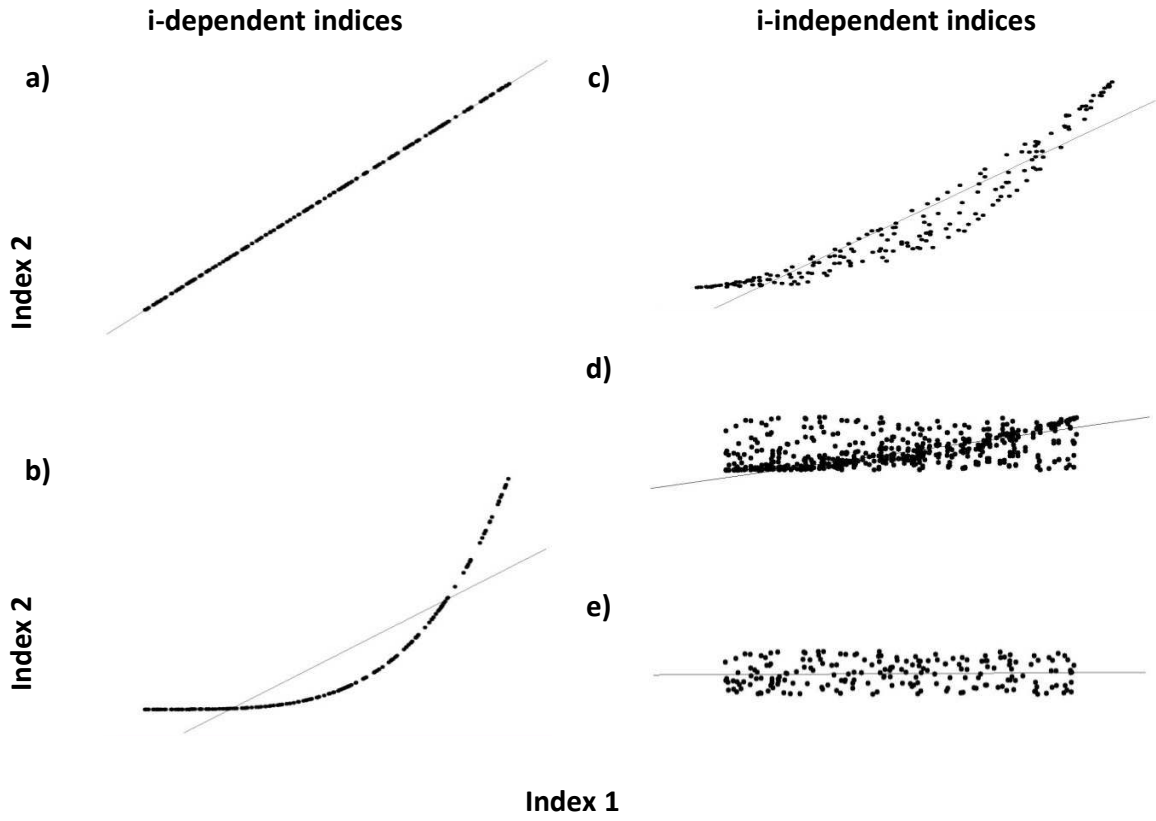

FIGURE 2 An example of relationships between two indices that are i-dependent (a,b), i-independent (c,d,e), s-dependent (a,b,c,d), s-independent (e), related sensu Chao et al. (2012) (a,b,c) and unrelated (d,e) indices. The dots show 200 samples from the underlying relationships and the lines show linear regressions. I-dependent indices (a,b) carry equal information because the strictly increasing or decreasing relationship between them uniquely transforms one index to another. 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 lower 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 (Thesis T01 in Šizling et al., 2023). Anyway, there are apparently no one-to-one (strictly increasing nor decreasing) continuous lines beyond the samples of i-independent indices (c,d,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. Indices in panels (c,d,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 (Figure 3). The degree of s-dependence measured by Pearson correlation coefficient decreases from (a) to (e) (1; 0.95; 0.86; 0.73; 0; N=200). Panels a and b show i-dependent relationships, panels c,d, and e show i-independent relationships.

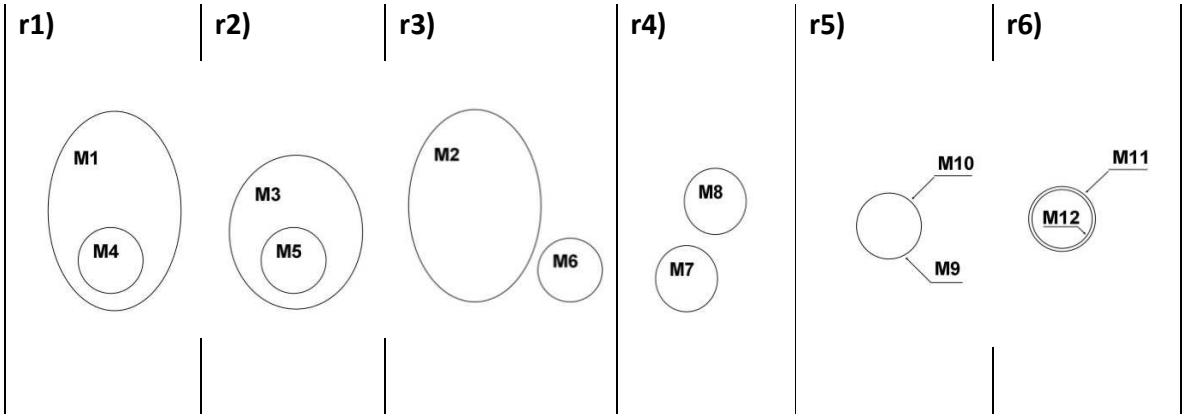

FIGURE 3 Constraints by the three (excluding beta-diversity *sensu stricto*) diversity phenomena on the indices. Various re-arrangements of two sets (lists of species) in the Venn diagrams define differences between the phenomena. The sizes of the sets follow M1 = M2 > M3 > M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11>M12. The sets M9 and M10 are identical. The constraints that define the spatial phenomena are as follows:
Nestedness: $\mathrm{Min} = \mathrm{N[r4]} = \mathrm{N[r3]} < \mathrm{N[r1]} = \mathrm{N[r2]} = \mathrm{N[r6]} = \mathrm{N[r5]} = \mathrm{Max}$;
Turnover: $\mathrm{Min} = \mathrm{T[r5]} < \mathrm{T[r6]} < \mathrm{T[r2]} < \mathrm{T[r1]} < \mathrm{T[r4]} =_{(1)} \mathrm{T[r3]}$; and
Richness Contrast: $\mathrm{Min} = \mathrm{Ct[r5]} = \mathrm{Ct[r4]} < \mathrm{Ct[r6]} < \mathrm{Ct[r2]} < \mathrm{Ct[r1]} = \mathrm{Ct[r3]}$.
The '*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 an arbitrary choice to separate the species richness contrast from species turnover.

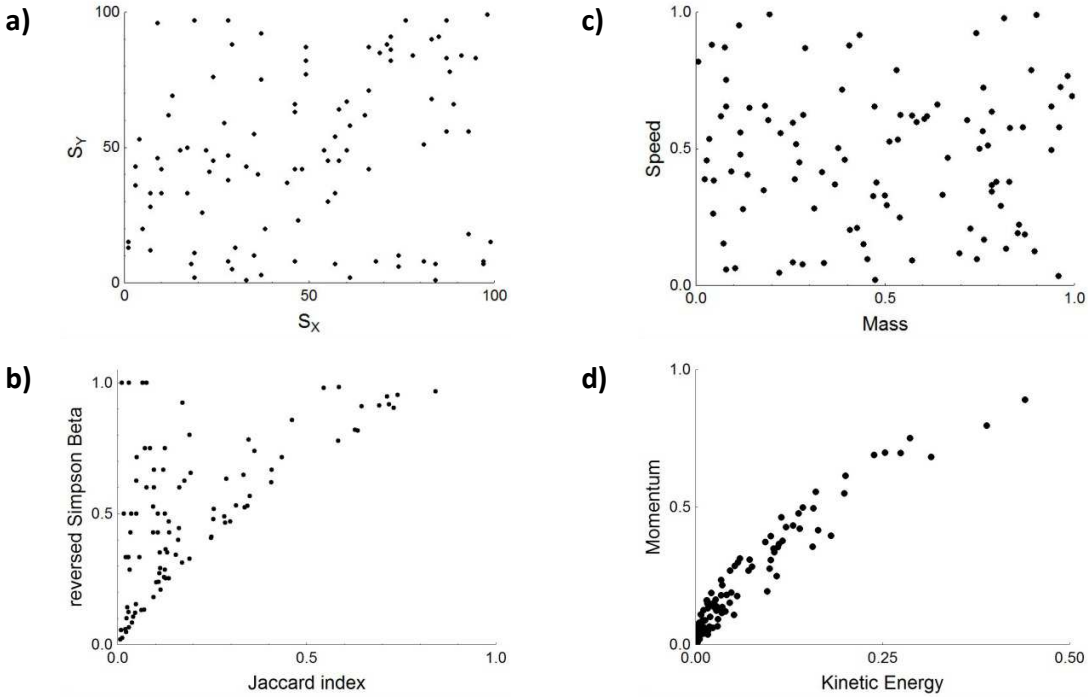

FIGURE 4 Statistical and informational (in)dependence in ecology (a,b) and physics (c,d). Mutually statistically independent (s-independent) values for species assemblages in ecology (S_X and S_Y) (a), and moving objects in physics (mass and speed) (c) do not necessarily result in statistically independent indices (J and β'_{Sim}) (b) and variables (p and E_k) (d). This is despite the fact that J, β'_{Sim}, p, and E_k carry a specific information and therefore are 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.

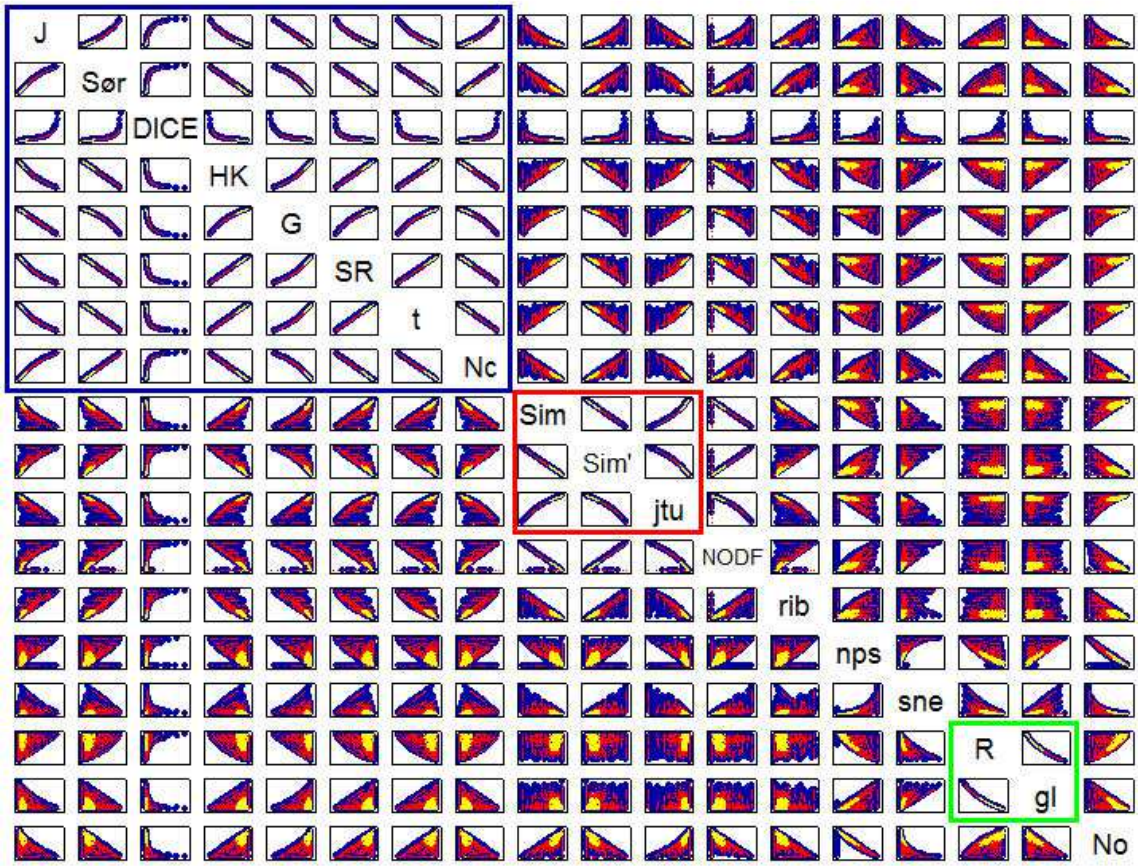

FIGURE 5 Relationships between the indices for pairs of simulated random (blue dots, N=1352; see Figure R3 in Šizling et al., 2023 for details) and observed (red – plants, N=946; yellow – Ice Shield microbes, N=406) assemblages (Data description R6 in Šizling et al., 2023). The species richness values 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 adopted from Gaston et al. (2007). The notation in the plot is simplified. From up to down: J – Table 1, $\beta_{Sør}$ – Table 1, β_{DICE} - Raup & Crick (1979), β_{HK} – Harte & Kinzig (1997), β_G – Table 1, β_{SR} – Schluter & Ricklefs (1993), β_{t2} – Table 1, N_C - Wright & Reeves (1992) standardized as in Gotelli & McCabe (2002), β_{Sim} – Table 1, β'_{Sim} - Table 1, β_{jtu} – Table 1, N_{NODF} – Table 1, β_{rib} – Ruggiero et al. (1998), , β_{nps} – Table 1 $\beta_{sne} \stackrel{\text{def}}{=} \beta_{NES}$, Table 1, R – Table 1; β_{gl} – Table 1, N_0 - Patterson & Atmar (1986) (see Table R1 in Šizling et al., 2023 for index definitions). Blue, red, yellow, and green rectangles delimit the families of Jaccard index, Simpson beta, and richness contrast, respectively.

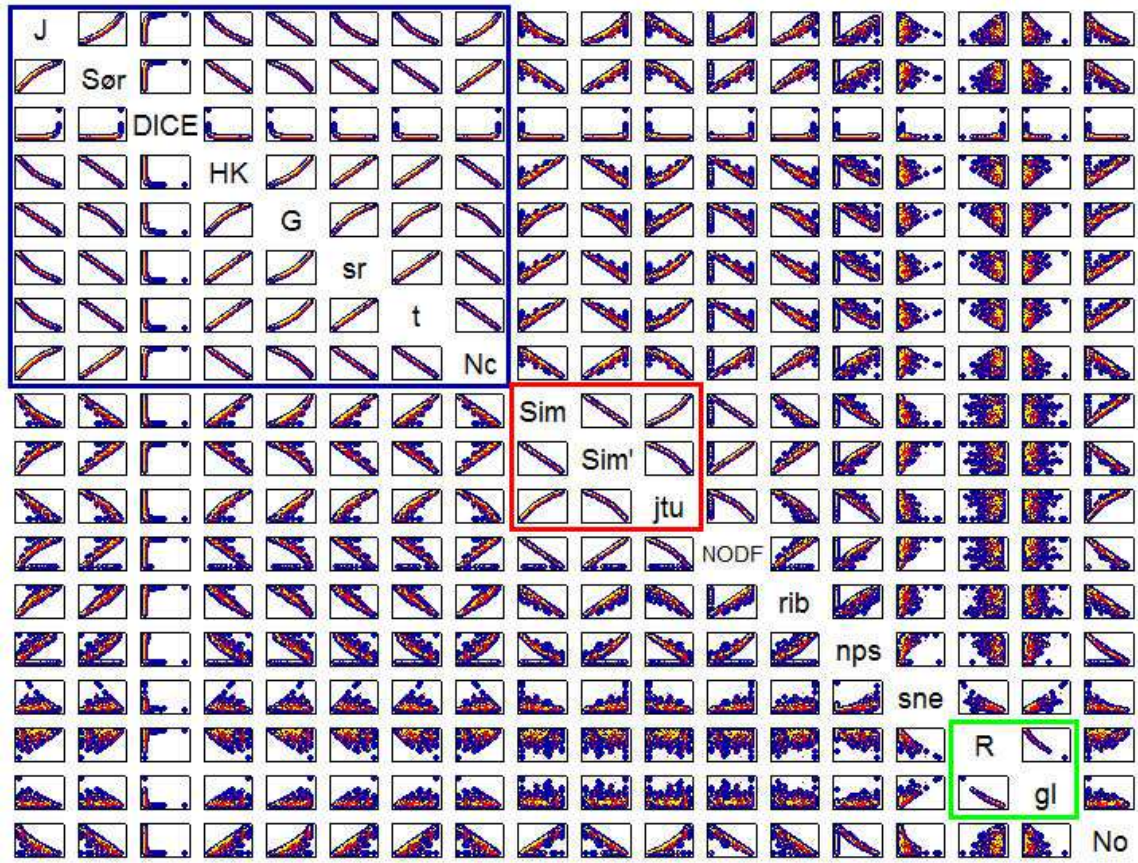

FIGURE 6 Relationships between pairwise indices of assemblages which have low (0%-20%) species richness contrast (simulations; see Figure 3 in Šizling et al., 2023 for details) or which are spatially close to each other (observed data; Data RI2 in Šizling et al., 2023). Blue (N=1254) and yellow (N=98) 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 Data description R6 in Šizling et al., 2023). The cluster of yellow dots is a random subsample of the blue dots, and the N of each subsample (yellow dots) is the same as the N of respective observations (red dots), to allow for 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 Figure 5.

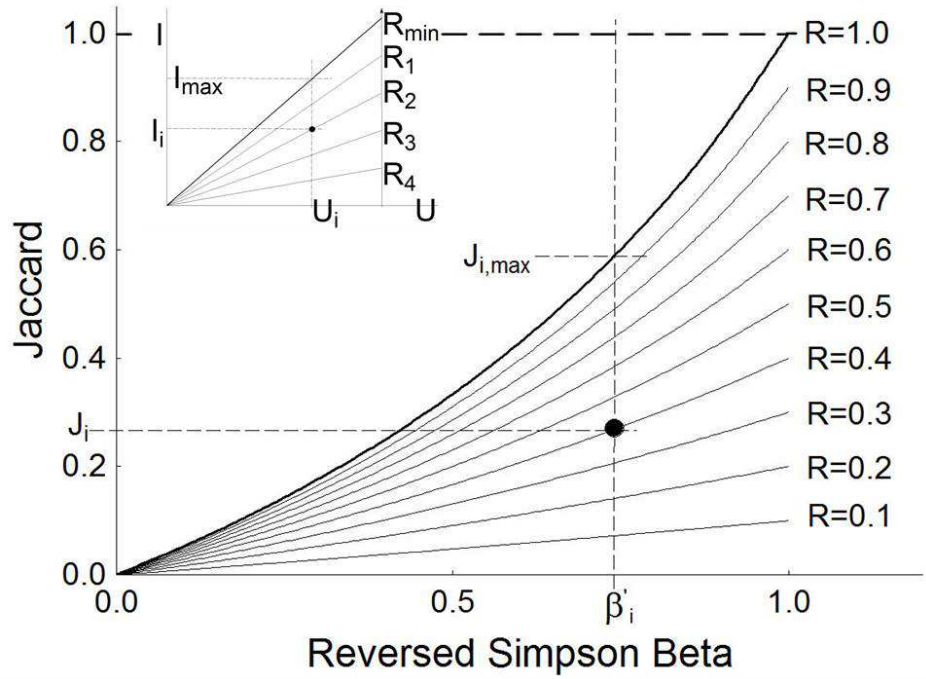

FIGURE 7 The reason why mutual dependence of indices cannot be eliminated to purify individual effects. The relationship between Jaccard similarity (J) and Simpson nestedness (β'_{Sim}) is determined by species richness uniformity R (Equation 7), each line shows points with equal R (the dot has R=0.4). 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, Equation 7). 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 uniformity *per se* (or reversed species richness contrast *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 *Formal definitions of the phenomena based on their constraints*). This is similar 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). The understanding in physics is that I_i is the current regardless of its dependence on *R* and *U*. No one would relativize I to R, nor say that in the absence of voltage U equals I, and thus U minus I measures the 'voltage resultant component'. In the same way, J_i is reversed turnover *per se* even though it also scales with reversed richness contrast *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.

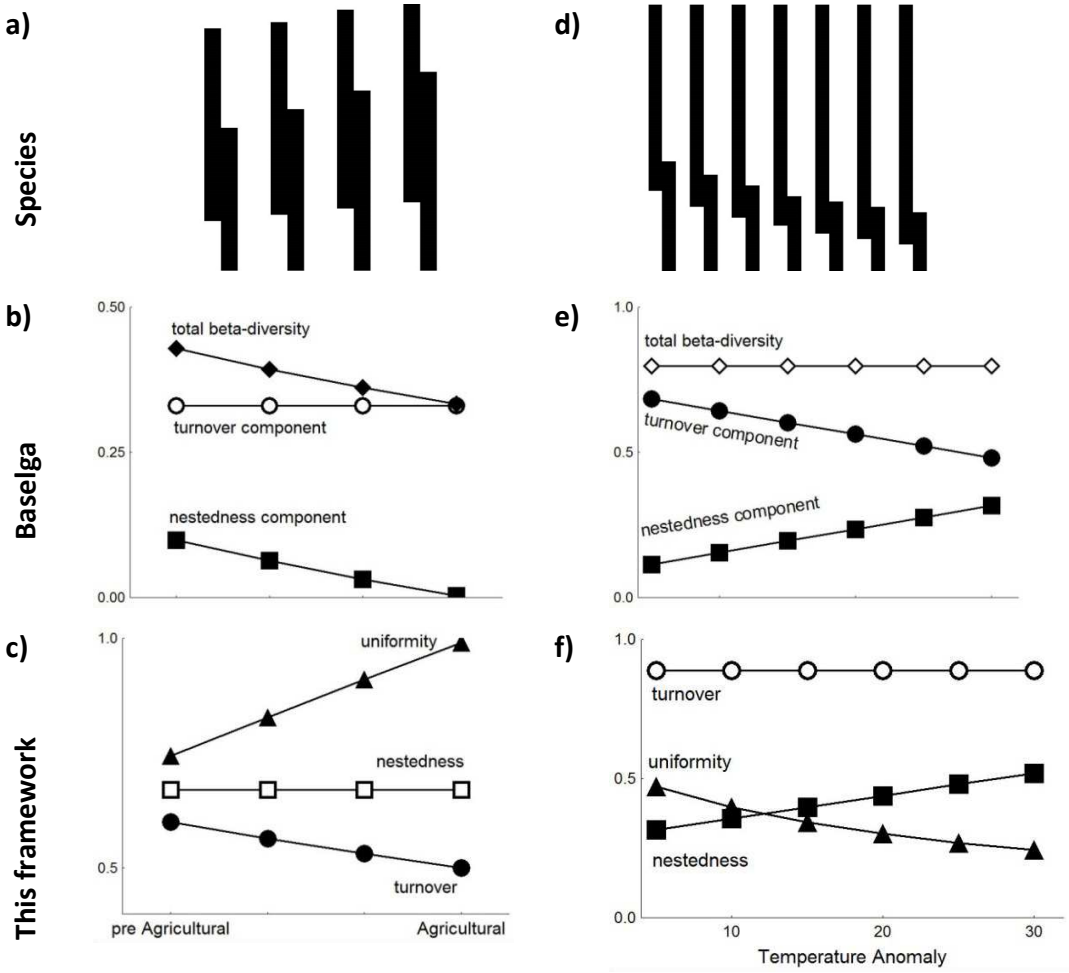

FIGURE 8 The illustration of the difference between the approach that uses Baselga's (2010) additive partitioning of pairwise indices (b,e) and the framework presented in this paper (c,f), using empirical data from Šizling et al. (2016, left) and Xu et al. (2023, right). Panels (a) and (d) show the species assemblages and respective overlaps in the lists of species (see caption to Figure 1). We used the algorithm in Box 2 and Box 3 to compute (a) and (b) from (c), and (f) and (d) from (e), as only data in (c) and (e) was published (the computed S_X ranges from 75.62 to 89.54; $S_Y \cong$ 35.66 to 21.74 and $S_{X \cap Y} \cong$ 11.28 to 11.28 in Xu. et al. 2023, S_X ranges from 30.97 to 32.16; $S_Y \cong$ 23.03 to 31.84 and $S_{X \cap Y} \cong$ 15.43 to 21.33 in Šizling et al. 2016). The figures (b) and (e) display trends for β_{Sim} (circles), β_{sne} (squares), and β_{t2} (diamonds), labeled according to Xu et al. (2023) as turnover component, nestedness component, and total beta-diversity, respectively ($\beta_{Sim} + \beta_{sne} = \beta_{t2}$). The figures (c) and (f) show trends for $1 - J$ (circles), β'_{Sim} (squares), and R (triangles), labeled according to this report as turnover, nestedness, and uniformity, respectively. Differences between (b) and (c), as well as (e) and (f), lead to divergent interpretations when applying our framework compared to Baselga's (2010, 2012). The lines show least square regressions; where the regression was not significant ($\alpha > 0.05$; open symbols) we plotted a constant line.

RI1: Repository Information for

Šizling, A.L., P. Keil, E. Tjørve, K.M.C. Tjørve, J.D. Žárský, and D. Storch. 202X. Mathematically and biologically consistent framework for presence-absence indices of diversity.

Content:

TABLE R1: Definitions and Scaling Properties	2-6
THESES T1-T21: Theses and Evidences	7-14
FIGURE R2: Average Values	15
FIGURE R3: Interrelated Assemblages	16
FIGURE R4: Multisite Version of Equation 2 in Šizling et al., (202X)	17
R5: i-Independence of and s-Dependence on Species Richness	18
R6: Data Description	19

For the list of the references see Šizling et al. (202X)

TABLE R1 (definitions and scaling properties): The table shows the definitions of diversity indices and their mutual scaling properties. The references (e.g., T01.4) attributed to the equation marks refer to the derivations of the relationships (see Theses and Proofs T01 for derivations). The equation marks without references are derived through simple rearrangements of the equations listed in the table. All listed scaling properties are exact except those marked as '≅'. Where possible, indices are defined according to Koleff *et al*. (2003) and Gaston *et al*. (2007); see the column 'Notes' for exceptions. Consistent 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 represent the numbers of species unique to the first and second assemblages ($b \stackrel{\text{def}}{=} S_X - S_{X\cap Y}$ and $c \stackrel{\text{def}}{=} S_Y - S_{X\cap Y}$). Further references can be found in the 'Notes' column.

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 only if the two sub-assemblages are adjacent, but it has fundamentally different inferences. $\beta_{Wh2} =_{(T01.4)} 2\frac{1}{1+J}$	Overall Whittaker (1960) beta-diversity of two sub-assemblages of one assemblage. (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 *sensu stricto* (T08)
J	$\stackrel{\text{def}}{=} \frac{a}{a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	$J =_{(T01.1)} \frac{\beta'_{Sim}}{1 + R^{-1} - \beta'_{Sim}}$	Jaccard index (1912) reversed turnover or co-occurrence if applied on between two species level (T02)
β_{SR}	$\stackrel{\text{def}}{=} \frac{a+b+c}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X + S_Y - S_{X\cap Y}}{S_X + S_Y}$	$\beta_{SR} =_{(T01.2)} \frac{1}{1+J}$	Schluter & Ricklefs (1993) turnover or reversed co-occurrence applied on between two species level (T02)
$\beta_{Sør}$	$\stackrel{\text{def}}{=} \frac{2a}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y}$	$\beta_{Sør} =_{(T01.3)} 2\frac{J}{1+J}$	Sørensen (1948) similarity, reversed turnover or co-occurrence applied on between two species level (T02)
β_{DICE}	$\stackrel{\text{def}}{=} \frac{2a}{b+c}$ $\stackrel{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y - 2S_{X\cap Y}}$	$\beta_{DICE} =_{(T01.17)} 2\frac{J}{1-J}$	Raup & Crick (1979) reversed turnover or co-occurrence applied on between two species level (T02)

β_{BC}	$\stackrel{\text{def}}{=} 1 - \frac{a}{2a+b+c}$ $\stackrel{\text{def}}{=} 1 - \frac{S_{X\cap Y}}{S_X + S_Y}$	$\beta_{BC} =_{(T01.16)} \frac{1}{1+J}$	Bray & Curtis (1957) simplified for incidence data Turnover or reversed co-occurrence if applied on between species level (T02)
β_{HK}	$\stackrel{\text{def}}{=} 1 - \frac{2a}{2a+b+c}$ $\stackrel{\text{def}}{=} 1 - \frac{2S_{X\cap Y}}{S_X + S_Y}$	$\beta_{HK} =_{(T01.5)} \frac{1-J}{1+J}$	Harte & Kinzig (1997) Turnover or reversed co-occurrence applied on between two species level (T02)
β_t	$\stackrel{\text{def}}{=} \frac{b+c}{2a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y}$	$\beta_t =_{(T01.6)} \frac{1-J}{1+J}$	Wilson & Shmida (1984) in Gaston *et al*. (2007) Baselga (2010a,2012) calls this index Sørensen dissimilarity Turnover or reversed co-occurrence applied on between two species level (T02)
β_G	$\stackrel{\text{def}}{=} \frac{b+c}{a+b+c}$ $\stackrel{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	$\beta_G =_{(T01.7)} 1 - J$ $\beta_G =_{(T01.6,T01.7)} 2\frac{\beta_t}{1+\beta_t}$	Gaston *et al*. (2001) Turnover or reversed co-occurrence applied on between two species level (T02)
β_{Sim}	$\stackrel{\text{def}}{=} \frac{\min(b,c)}{a+\min(b,c)}$ $\stackrel{\text{def}}{=} \frac{\min(S_X,S_Y) - S_{X\cap Y}}{\min(S_X,S_Y)}$	$\beta_{Sim} =_{(T01.8)} 1 - \beta'_{Sim}$	Simpson beta Simpson (1943); Lennon *et al*. (2001) reversed nestedness (T05)
β'_{Sim} (rN)	$\stackrel{\text{def}}{=} \frac{a}{a+\min(b,c)}$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}}{\min(S_X,S_Y)}$	$\beta'_{Sim} =_{(T01.8)} 1 - \beta_{Sim}$, $\beta'_{Sim} =_{(T01.9)} \frac{1+R}{R}\frac{J}{1+J}$	Lennon *et al*. (2001), Šizling *et al*. (2016) β'_{Sim} equals rN (range nestedness) Šizling *et al*. (2009) if computed between two species ranges Nestedness (T05)
β_{jtu}	$\stackrel{\text{def}}{=} 2\frac{\min(b,c)}{a+2\min(b,c)}$ $\stackrel{\text{def}}{=} 2\frac{\min(S_X,S_Y) - S_{X\cap Y}}{2\min(S_X,S_Y) - S_{X\cap Y}}$	$\beta_{jtu} =_{(T01.21)} 2\frac{\beta_{Sim}}{1+\beta_{Sim}}$	Introduced as turnover component of Jaccard dissimilarity in Baselga (2012) reversed nestedness (T05)

β_{gl}	$\stackrel{\text{def}}{=} 2\frac{\lvert b-c\rvert}{2a+b+c}$ $\stackrel{\text{def}}{=} 2\frac{\lvert S_X-S_Y\rvert}{S_X+S_Y}$	$\frac{R}{1+R} =_{(T01.10)} \frac{2-\beta_{gl}}{4}$ $\beta_{gl} =_{(*)} 2\frac{1-R}{1+R}$	Lennon *et al*. (2001) species richness contrast (T06) (* *follows from the above* $T19.10$ *equation*)
R	$\stackrel{\text{def}}{=} \frac{a+\min(b,c)}{a+\max(b,c)}$ $\stackrel{\text{def}}{=} \frac{\min(S_X,S_Y)}{\max(S_X,S_Y)}$	$R =_{(T01.10)} \frac{2-\beta_{gl}}{2+\beta_{gl}}$ $R =_{(T01.9)} \frac{J}{\beta'_{Sim}(1+J)-J}$	Newbold *et al*. (2016)scales in one-to-one manner with β_{gl} by Lennon *et al*. (2001); reversed species richness contrast or species richness uniformity (T06)
$\beta_{sne} =_{(*)} \beta_{NES}$	$\stackrel{\text{def}}{=} \beta_t - \beta_{Sim}$	does not scale with other indices	Baselga (2012) (* it is labeled as β_{NES} in Baselga 2010) No family under examination (T09)
β_{jne}	$\stackrel{\text{def}}{=} \beta_G - \beta_{jtu}$	does not scale with other indices	Baselga (2012) No family under examination (T13)
β_{rps}	$\stackrel{\text{def}}{=} 2\frac{\min(b,c)}{a+b+c}$ $\stackrel{\text{def}}{=} 2\frac{\min(S_X,S_Y)-S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}}$	does not scale with other indices	Introduced as (one-for-one) replacement and labeled R_{rel} in Podani & Schmera (2011) no phenomenon in consideration
β_{nps}	$\stackrel{\text{def}}{=} \frac{a+\lvert b-c\rvert}{a+b+c}$ if $a \neq 0$ & $b \neq c$ $\stackrel{\text{def}}{=} \frac{S_{X\cap Y}+\lvert S_X-S_Y\rvert}{S_X+S_Y-S_{X\cap Y}}$ if $S_{X\cap Y} \neq 0$ & $S_X \neq S_Y$ $\stackrel{\text{def}}{=} 0$ if $S_{X\cap Y} = 0$ or $S_X = S_Y$	$\beta_{nps} =_{(T01.22)} 1-\beta_{rps}$ if there were not for the extra conditions that make the index value zero	Introduced as nestedness and labeled N'_{rel} in Podani & Schmera (2011) no phenomenon in consideration
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 reversed turnover depending on the direction; S_Y stands for species richness of the approached assemblage if we leave assemblage X (T11)

$\langle\beta_{Sim}\rangle$	$\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\beta_{Sim,i}$	$\langle\beta_{Sim}\rangle =_{(T01.11)} 1-\langle\beta'_{Sim}\rangle$; $\langle\beta_{Sim}\rangle \cong_{(T01.12)} D$	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 a high number of sites (≥ 10) reversed Nestedness (T05,T03)
$\langle\beta'_{Sim}\rangle$	$\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\beta'_{Sim,i}$	$\langle\beta'_{Sim}\rangle =_{(T01.11)} 1-\langle\beta_{Sim}\rangle$; $\langle\beta'_{Sim}\rangle \cong_{(T01.11,T19.12)} 1-D$; $\langle\frac{J}{1+J}\rangle \cong_{(T01.13*)} \langle\beta'_{Sim}\rangle\langle\frac{R}{1+R}\rangle$	The scaling between $\langle\beta_{Sim}\rangle$ and D works only for assemblages with a high number of sites (≥ 10) Nestedness (T05,T03) (*) if $cov(\langle\beta'_{Sim}\rangle,\langle\frac{R}{1+R}\rangle) \cong 0$, which is likely
D	Discrepancy (defined with an algorithm)	$D \cong_{(T01.12)} \langle\beta_{Sim}\rangle =_{(T01.11)} 1-\langle\beta'_{Sim}\rangle$	Brualdi & Sanderson (1999) Scaling (T01.12) works only for assemblages with a high number of sites (≥ 10) Reversed Nestedness (T03)
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 it would scale with β'_{Sim} if there was not for condition $S_X = S_Y \Rightarrow N_{NODF} = 0$	Almeida-Neto *et al*. (2008) No family under examination (T12)
N_0	Number of gaps (defined with 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 (T10)

N_C	(defined with an algorithm)	$N_{C1} =_{(T01.19)} 1 - \beta_{SR} =_{(T01.20)} \frac{\beta_{Sør}}{2} =_{(T01.3)} \frac{J}{1+J}$ $N_{C2} \xrightarrow{\langle S \rangle \to \infty}_{(T04)} 2\beta'_{Sim} - 1$	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}, reversed turnover (T02) N_{C2} Nestedness (T04)

Theses (theses and proofs): Theses T1-21 state whether or not each index from Table R1 satisfies the constraints of the spatial phenomena (Figure 3 in Šizling et al. 202X), the conditions for i-independnce of the indices, and the scaling between indices within a family and between families. The evidences for the theses employ three parameters: a, (the number of shared species, $S_{X\cap Y}$,), b (the number of species exclusive to the first assemblage, $S_X - S_{X\cap Y}$), and c (the number of species exclusive to the second assemblage, $S_Y - S_{X\cap Y}$). The individual arrangements in Figure 3 in Šizling et al. (202X), are characterized as follows: $a = 0$ for the arrangements r3 and r4; $b = 0$ in arrangements r1,r2,r5 and r6; $b = 0$ and $c = 0$ in arrangement r5; and by $c_{r6} < c_{r2} < c_{r1}$ in arrangements r1,r2,r6.

Understanding the evidence for the theses requires a basic knowledge of linear algebra. Specifically, one should know: what a system of linear equations is and how it can be converted to a matrix (e.g., system: $a_{11}x_1 + a_{12}x_2 = b_1$; $a_{21}x_1 + a_{22}x_2 = b_2$; the matrix: $\begin{bmatrix} a_{11} & a_{12} & \| & b_1 \\ a_{21} & a_{22} & \| & b_2 \end{bmatrix}$); the Gauss elimination method; Cramer's rule and the theorem stating that a nonzero determinant indicates a set of mutually independent (in our work, i-independent) equations while a zero determinant indicates a set of mutually (i-)dependent equations. Formore details on the link between linear algebra and indices of diversity see Box 2 and Box 3 in Šizling et al. (202X).

T01 (scaling properties): *Relationships between the focal indices obey the equations listed in Table R1*. The evidence for the relationships is as follows (the numbering '1-22' refers to the labels associated with equation marks in Table *R1*; brackets $\langle . \rangle$ denote a mean value):

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

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

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

4. $\beta_{Wh2} = \frac{a+b+c}{(2a+b+c)/2} =_{(T19.2)} 2\beta_{sr} = 2\frac{1}{1+J}$.

5. $\beta_{HK} = 1 - \frac{2a}{2a+b+c} =_{(T19.3)} 1 - \beta_{Sor} = 1 - 2\frac{J}{J+1} = \frac{1-J}{1+J}$.

6. $\beta_t = \frac{b+c}{2a+b+c} = \frac{2a+b+c-2a}{2a+b+c} = 1 - \frac{2a}{2a+b+c} =_{(T19.5)} \frac{1-J}{1+J}$.

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. $\beta_{Sim} = \frac{\min[b,c]}{a+\min[b,c]} = \frac{a+\min[b,c]-a}{a+\min[b,c]} = 1 - \frac{a}{a+\min[b,c]} = 1 - \beta'_{Sim}$.

9. $JR(1+R^{-1}-\beta'_{Sim}) =_{(T01.1)} \beta'_{Sim} \Rightarrow J(R+1-\beta'_{Sim}\cdot R) = \beta'_{Sim}\cdot R \Rightarrow \frac{J}{1+J} = \beta'_{Sim}\frac{R}{1+R}$.

10. Let $b \le c$ 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. $\beta_{Sim} =_{(T01.8)} 1 - \beta'_{Sim} \Longrightarrow \langle\beta_{Sim}\rangle = \langle 1-\beta'_{Sim}\rangle \Longrightarrow \langle\beta_{Sim}\rangle = \langle 1\rangle - \langle\beta'_{Sim}\rangle$

12. Simulation based evidence (see Figure R2 below).

13. $\frac{J}{J+1} =_{(T01.9)} \beta'_{Sim}\frac{R}{1+R} \Rightarrow \langle\frac{J}{1+J}\rangle = \langle\beta'_{Sim}\frac{R}{1+R}\rangle \Rightarrow \langle\frac{J}{1+J}\rangle = \langle\beta'_{Sim}\rangle\langle\frac{R}{1+R}\rangle + \mathrm{cov}(\beta'_{Sim},\frac{R}{1+R})$.

Where the i-th values $\beta'_{Sim,i}$ and $\frac{R_i}{1+R_i}$ are s-independent across all i, $\mathrm{cov}\left(\beta'_{Sim},\frac{R}{1+R}\right) = 0$.

Nestedness, $\beta'_{Sim,i}$ and species richness uniformity $\frac{R_i}{1+R_i}$ does not constraint each other (share only one sr component of the S_X, S_Y, $S_{X\cap Y}$) and are likely s-independent (implying zero covariance). However, an ecological driver makes covariance nonzero and then the relationship in Table R1 does not work.

14. ($0 \leq R \leq 1 \;\Rightarrow\; 0 \leq \frac{R}{1+R} \leq 0.5$), it follows ($0 \leq \frac{R}{1+R} \leq 0.5$ & $\frac{J}{1+J} =_{(T01.9)} \beta'_{Sim}\frac{R}{1+R} \Longrightarrow \frac{J}{1+J} \leq 0.5\beta'_{Sim} \Rightarrow J \leq \frac{\beta'_{Sim}}{2-\beta'_{Sim}}$).

15. $J =_{(T01.6)} \frac{1-\beta_t}{1+\beta_t} =_{(\mathrm{Def}\,\beta sne)} \frac{1-\beta_{sne}-\beta_{Sim}}{1+\beta_{sne}+\beta_{Sim}}$.

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

22. $\beta_{nps} + \beta_{rps} \stackrel{\text{def}}{=}_{(*)} \frac{S_{X\cap Y}+|S_X-S_Y|}{S_X+S_Y-S_{X\cap Y}} + 2\frac{\min(S_X,S_Y)-S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}} =_{(Sy\leq Sx)} \frac{S_{X\cap Y}+S_X-S_Y}{S_X+S_Y-S_{X\cap Y}} + 2\frac{S_Y-S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}} = \frac{S_X+S_Y-S_{X\cap Y}}{S_X+S_Y-S_{X\cap Y}} = 1$. (*)=no discontinuity conditions

T02: *Jaccard index* (J) *is consistent with the constraints on reversed turnover*.
Note: All the indices that follow strictly increasing function of J are consistent with the constraints on reversed turnover, and all indices that follow a strictly decreasing function of J are consistent with the constraints on turnover. For the scaling, see Fig. 5 in Šizling et al. (202X), Thesis T01, and Table R1. For a new index, consult Repository RI4: Calculator.
Evidence: $J = \frac{a}{a+b+c}$. Here $a = 0$ in re-arrangements r3 and r4 in Figure 3 in Šizling et al. (202X). Hence $J[r3] = J[r4] = 0$, which excludes the species richness contrast but supports turnover as defined in our framework. Additionally, $J[r1] < J[r2] < J[r6] < J[r5]$ in Figure 3 in Šizling et al. (202X). Hence, $0 < J[r1] < J[r5] = 1$, which supports only the reversed constraints of turnover.

T03: *Discrepancy* (D) *for a high number of sites approaches a function that is consistent with the constraints on nestedness.*
Note: Figure R2 shows that more than 10 sites are sufficient to achieve consistency with the constraints.
Evidence: For a small number of sites to compare, index D does not follow any of the focal phenomena. This evidence is based on D computed for various pairs of sets (M1 = M2 = 15, M3=10, M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11=5, and M12=4.). The rough (non-standardized) Discrepancy is defined as the minimum number of incidences that must be shifted along rows of an incidence matrix (rows represent sites; columns represent species) to achieve absolutely nested assemblages (Brualdi and Sanderson 1999). This definition simplifies to '$\min(S_X, S_Y) - S_{X\cap Y}$' for two assemblages. Index D is standardized by the number of incidences within the focal matrix, thus $D = \min(S_X, S_Y) - S_{X\cap Y}/(S_X + S_Y)$ in the case. The computed order is therefore $0 = D[r1] = D[r2] = D[r5] = D[r6] < D[r3] < D[r4] = 0.5$. The equality $D[r1] = D[r5]$ excludes species richness contrast, nestedness, and turnover, and the inequality $D[r3] < D[r4]$ excludes nestedness, but not contrast. However, if the number of sets is large enough (simulations suggest more than 30 simulations), the standardization by the number of incidences begins to work properly and the index D will scale with the indices of nestedness (Figure R2 below). We demonstrate the reason using two extreme cases: with maximum and minimum possible D. The Discrepancy of an absolutely nested matrix is by definition zero (no shift of incidences is needed). Maximum Discrepancy occurs in a matrix where almost all incidences (except the incidences of one site) must be shifted to gain an absolutely nested matrix. The Discrepancy of an absolutely non-nested matrix (where each site

has its unique set of species) is then computed as the total species richness (sum across all sites) of species that has to be shifted (i.e., excluding one of the most species rich sites). This equals $\sum_{i=1}^{Sites} S_i - max_{i=1}^{Sites}[S_i]$, where S_i is the species richness of the i-th site. It is standardized by the 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 of practically all datasets with large numbers of sites, and therefore $D \cong 1$ can be attributed to absolutely non-nested multisite assemblages. It is apparent that all matrices between these two extremes have D values between 0 and 1. The condition $max_{i=1}^{Sites}[S_i] \ll \sum_{i=1}^{Sites} S_i$ might only be violated if the maximum species richness was high and species richness of the other sites was extremely small, which is unlikely. Simulations show a one-to-one scaling of *D* with Simpson beta (non nestedness) for matrices of 30 and 100 sites (Figure R2).

T04: N_C, *standardized as in* Wright & Reeves (1992) *(here labeled N_{C2}; labeled as C in Wright & Reeves 1992) approaches i-dependence on Simpson beta as Species Richness approaches infinity. N_{C2} is i-dependent on species richness.*
Note: N_{C2} is thus consistent with the constraints on Nestedness.
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 $\lim_{r\to\infty} N_C = 2\frac{S_{X\cap Y}}{S_Y} - 1 = 2\beta'_{Sim} - 1$.

T05: *Simpson beta β_{Sim}* and *β_{jtu} are consistent with the constraints on reversed nestedness and β'_{Sim} is consistent with the constraints on nestedness.*
Note: The evidence is done for β_{Sim}. For the strictly increasing function of β_{jtu} with β_{Sim}, and the strictly decreasing function of β'_{Sim} with β_{Sim} see Fig. 5 in Šizling et al. (202X), Thesis T01, and Table R1.
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 equals zero and $a \neq 0$ in these cases. All the indices are continuous and none of them depends on $\max(b, c)$. For scaling between β_{jtu}, β'_{Sim} and β_{Sim} see (T01.8, T01.21 and Figure 5 in Šizling et al. 202X).

T06: *β_{gl} and R are consistent with the constraints on species richness contrast and species richness uniformity (i.e., reversed species richness contrast), respectively.*
Evidence: R is the ratio of minimum to maximum species richness. Hence $R[r1] = R[r3] < R[r2] < R[r6] < R[r4] = R[r5] = 1$. *R* is therefore inversely related to species richness contrast, representing species richness uniformity. β_{gl} scales negatively with *R*, (see T01.10).

T07: *β_{nps} is inconsistent with the constraints on nestedness, turnover or species richness contrast.*
Evidence: If we ignore the extra conditions that make β_{nps} discontinuous (i.e., β_{nps} if $S_{X\cap Y} = 0$ or $S_X = S_Y$) then $0 = \beta_{nps}[r4] < \beta_{nps}[r3] < \beta_{nps}[r5] = \beta_{nps}[r6] = \beta_{nps}[r2] = \beta_{nps}[r1] = 1$, where r1,…,r6 are re-arrangements of two assemblages defined in Figure 3 in Šizling et al. (202X). The inequality $\beta_{nps}[r4] < \beta_{nps}[r3]$ excludes nestedness, the equality $\beta_{nps}[r2] = \beta_{nps}[r1]$ excludes the species richness contrast, and the equality $\beta_{nps}[r2] = \beta_{nps}[r1]$ excludes turnover.

If we add the condition: "$\beta_{nps} = 0$ if $S_{X\cap Y} = 0$" (the variant of the index under discussion in Šizling et al. 202X, which is labeled N_{rel} in Podani and Schmera 2011), then the relations become $0 = \beta_{nps}[r4] = \beta_{nps}[r3] < \beta_{nps}[r5] = \beta_{nps}[r6] = \beta_{nps}[r2] = \beta_{nps}[r1] = 1$. This again excludes species richness contrast and turnover, but it agrees with the constraints on nestedness (Figure 3 in Šizling et al. 202X). This is only due to the extra condition, so if we

consider that indices must follow the requirement of continuity, the β_{nps} does not capture the nestedness. However, there is another reason why β_{nps}does not capture the experience of nestedness. The constraints $N[r1] = N[r2]$ implies independence of nestedness from the species richness of the richer assemblage. If we violate this condition where there is an overlap between the assemblages (i.e., $S_{X\cap Y} > 0$), then we are inconsistent with the constraint $N[r1] = N[r2]$ (Figure 3 in Šizling et al. 202X). The intuition is that if we enter the sea, then the fraction of how much we are immersed does not depend on the size of the sea. It is easy to see that $\beta_{nps} \stackrel{\text{def}}{=} 2(S_Y - S_{X\cap Y})/(S_X + S_Y - S_{X\cap Y})$ changes with changing S_X $(= \max(S_X, S_Y))$, which excludes nestedness (arbitrarily without loss of universality $S_Y \leq S_X$).

If we add the condition: "$\beta_{nps} = 0$ if $S_{X\cap Y} = 0$ or $S_X = S_Y$" (this variant of the index that is labeled N'_{rel} in Podani and Schmera 2011) then the relations turn into $0 = \beta_{nps}[r4] = \beta_{nps}[r3] = \beta_{nps}[r5] = \beta_{nps}[r6] = \beta_{nps}[r2] = \beta_{nps}[r1] = 1$. Again, the equality $\beta_{nps}[r3] = \beta_{nps}[r5]$ excludes nestedness, the equality $\beta_{nps}[r2] = \beta_{nps}[r1]$ excludes the species richness contrast, and the equality $\beta_{nps}[r2] = \beta_{nps}[r1]$ excludes turnover.

T08: Index β_{rps} *is inconsistent with the constraints on nestedness, turnover and species richness contrast. The index* β_{rps} *is also inconsistent with the concept of One-for-One Replacement.*
Evidence: The relations for β_{rps} are reversed to the relations for β_{nps} where no extra conditions that make the index discontinuous are considered (i.e., $0 = \beta_{rps}[r4] < \beta_{rps}[r3] < \beta_{rps}[r5] = \beta_{rps}[r6] = \beta_{rps}[r2] = \beta_{rps}[r1] = 1$), for $\beta_{rps} = 1 - \beta_{nps}$ in this case (T01.22). Therefore, the reasoning for nestedness, species-richness contrast, and turnover follows the evidence T07 above. In the case of one-for-one replacement, there is a constraint $R[r4] = R[r3]$ on the one-for-one replacement that follows the point of Podani & Schmera (2011). The inequality $\beta_{rps}[r4] < \beta_{rps}[r3]$ therefore excludes one-for one replacement. Because the numerator of β_{rps} for r3 equals the numerator of β_{rps} for r4, the problem is introduced by denominator (i.e., standardization of the index with $S_X + S_Y - S_{X\cap Y}$). One may argue that the one-for-one replacement of r3 differs from one-for-one replacement of r4, and thus the inequality $\beta_{rps}[r4] < \beta_{rps}[r3]$ does not disqualify β_{rps} as a measure of the one-for-one replacement. In this case, however, there should be a difference between $\beta_{nps}(S_X = 8, S_Y = 20, S_{X\cap Y} = 1)$ and $\beta_{nps}(17,20,10)$, but $\beta_{nps}(8,20,1) = \beta_{nps}(17,20,10)$ (Figure 1 in Šizling et al. 202X).

T09: *Nestedness by Baselga* (β_{sne}) *is inconsistent with the constraints on nestedness, turnover and species richness contrast.*
Evidence: $\beta_{sne} = \frac{b+c}{2a+b+c} - \frac{\min(b,c)}{a+\min(b,c)}$**.** For re-arrangements r1 and r2 in Figure 3 in Šizling et al. (202X), $\beta_{sne}[r1] < \beta_{sne}[r2]$, because $\beta_{Sim}[r1] = \beta_{Sim}[r2] = 0$, $b = 0$, $c_{r2} < c_{r1}$, and a does not vary between these two cases. This excludes nestedness. $\beta_{sne}[r5] = \beta_{sne}[r3] = 0$, because $b = c$, which excludes turnover. $\beta_{sne}[r1] > \beta_{sne}[r3] = 0$, which excludes species richness contrast ($\beta_{sne}[r3] = 0$, because $a = 0$; and ($\beta_{sne}[r1] > 0$, because $min(b, c) = 0$ and $b \neq c$).

T10: *Number of gaps* (N_0) *is inconsistent with the constraints on nestedness, turnover and species richness contrast.*
Evidence: This evidence is based on N_0 indices computed for various pairs of sets (M1 = M2 = 15, M3=10, M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11=5, and M12=4.). 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 was standardized by number of incidences within the focal matrix. This matches no experience of the five spatial phenomena.

T11: *The Ruggiero index of beta-diversity* (β_{Rib}) *is consistent with the constraints on reversed turnover 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 $0 = \beta_{Rib}[r3] < \beta_{Rib}[r4]$, because $a = 0$ in these cases. The index β_{Rib} thus captures reversed turnover. If we replace *b* with *c*, then $\beta_{Rib}[r1] = \beta_{Rib}[r2] = \beta_{Rib}[r6] = \beta_{Rib}[r5] = 1$, because *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 consistent with the constraints on 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, which is an index of nestedness. Where assemblages have equal species richness $N_{NODF} = 0$. If we accepted that two equal sized assemblages are not mutually nested ($NODF(r5) = 0$, but $NODF(r6) = 1$), then N_{NODF} would be an index of nestedness. However, our framework excludes this possibility.

T13: *The Nestedness resultant component of Jaccard dissimilarity (β_{jne}) is inconsistent with the constraints on nestedness, turnover or species richness contrast.*
Evidence: This evidence is based on N_0 indices computed for various pairs of sets (M1 = M2 = 15, M3=10, M4 = M5 = M6 = M7 = M8 = M9 = M10 =M11=5, and M12=4.). The computed order is $\beta_{jne}[r3] = \beta_{jne}[r4] = \beta_{jne}[r5] = 0$, $\beta_{jne}[r6] = 0.2$, $\beta_{jne}[r2] = 0.5$, and $\beta_{jne}[r1] \cong 0.7$. The equality $\beta_{jne}[r4] = \beta_{jne}[r5]$ excludes nestedness and turnover, and tha equality $\beta_{jne}[r3] = \beta_{jne}[r4]$ excludes species richness contrast.

T14: *An algorithm to compute Jaccard similarity and Simpson nestedness from two i-independent indices if one of the indices is partitioned.*
The algorithm: 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 index I_b is subtracted the k_{ib} coefficients are multiplied by '-1'*)
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 $D_{LL} = \mathbf{det}\begin{bmatrix} L_{1a} & L_{2a} & L_{3a} \\ L_{1b} & L_{2b} & L_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$; $D_{KK} = \mathbf{det}\begin{bmatrix} K_{1a} & K_{2a} & K_{3a} \\ K_{1b} & K_{2b} & K_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$; $D_{LK} = \mathbf{det}\begin{bmatrix} L_{1a} & L_{2a} & L_{3a} \\ K_{1b} & K_{2b} & K_{3b} \\ M_1 & M_2 & M_3 \end{bmatrix}$;

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 3 in Šizling et al. XXXX to compute any dimensionless index.
This solution works only if the indices I_a, I_b are mutually i-independent, the indices I_a, I_1 are mutually i-independent, and the indices I_1, I_b are mutually i-independent (together they are always i-dependent). If any pair of indices is mutually i-dependent, the solution is simpler: convert the i-dependent indices to an index from their family (Table R1) and then follow the algorithm from Box 3 in Šizling et al. XXXX.

T15: Symmetric and linear indices with equal denominators are i-dependent.
That is: Let us have two indices that are defined as

$$I_\alpha \stackrel{\text{def}}{=} \frac{k_{1\alpha}f_1(S_{X\cap Y})+k_{2\alpha}f_2(S_X)+k_{3\alpha}f_3(S_Y)}{l_{1\alpha}f_1(S_{X\cap Y})+l_{2\alpha}f_2(S_X)+l_{3\alpha}f_3(S_Y)}, (k_{1\alpha-3\alpha},\ l_{1-3} \in \boldsymbol{R}), |l_{2\alpha}| + |l_{3\alpha}| > 0 \qquad \text{(T15.1)}$$

$$I_\beta \stackrel{\text{def}}{=} \frac{k_{1\beta}f_1(S_{X\cap Y})+k_{2\beta}f_2(S_X)+k_{3\beta}f_3(S_Y)}{l_{1\beta}f_1(S_{X\cap Y})+l_{2\beta}f_2(S_X)+l_{3\beta}f_3(S_Y)}, (k_{1\beta-3\beta},\ l_{1-3} \in \boldsymbol{R}), |l_{2\beta}| + |l_{3\beta}| > 0 \qquad \text{(T15.2)}$$

where $f_i(X)$ are strictly increasing or strictly decreasing functions from $\boldsymbol{R}^+$ to $\boldsymbol{R}^+$, and $k_{2\alpha} = k_{3\alpha}$, $k_{2\beta} = k_{3\beta}$, and $l_{1\alpha} = l_{2\alpha}$, $l_{1\beta} = l_{2\beta}$,
then $(l_{1\alpha} = l_{1\beta}\ \&\ l_{2\alpha} = l_{2\beta}) \Rightarrow I_\alpha$ is i-dependent on I_β.
Note 1: if $f_i(X) = X, \forall X \in \mathbf{R}$, then Eqs. T15.1 and T15.2 turn into Eq. 8 in Šizling et al. 202X.
Note 2: It follows that two asymmetric indices with equal denominators are i-independent.
Note 3: The symmetry required here is a stronger form of symmetry as it necessitates symmetry in both the numerator and denominator separately. For example, the index R is not a symmetric index according to this theorem.
Evidence:

$$\begin{aligned} det\begin{bmatrix} I_\alpha l_1 - k_{1\alpha} & I_\alpha l_2 - k_{2\alpha} & I_\alpha l_3 - k_{3\alpha} \\ I_\beta l_1 - k_{1\beta} & I_\beta l_2 - k_{2\beta} & I_\beta l_3 - k_{3\beta} \\ 0 & 1 & 1 \end{bmatrix} \\ = (I_\alpha l_1 - k_{1\alpha})(I_\beta l_2 - k_{2\beta}) + (I_\alpha l_3 - k_{3\alpha})(I_\beta l_1 - k_{1\beta}) \\ - (I_\alpha l_1 - k_{1\alpha})(I_\beta l_3 - k_{3\beta}) - (I_\alpha l_2 - k_{2\alpha})(I_\beta l_1 - k_{1\beta}) \\ = I_\alpha\left(k_{1\beta}(l_2 - l_3) + l_1(k_{3\beta} - k_{2\beta})\right) + I_\beta(l_1(k_{2\alpha} - k_{3\alpha}) + k_{1\alpha}(l_3 - l_2)) \\ + k_{1\alpha}(k_{2\beta} - k_{3\beta}) + k_{1\beta}(k_{3\alpha} - k_{2\alpha}) \end{aligned}$$

Therefore determinant is zero (i.e., indices are i-dependent) If $l_2 - l_3 = 0\ \&\ k_{3\alpha} - k_{2\alpha} = 0\ \&\ k_{3\beta} - k_{2\beta} = 0$.

T16: (*species rich and poor assemblages*): 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 the richness of the species richer assemblage.
Evidence: The evidence is based on the solution of the system of three equations for J, β'_{Sim} and $\langle S\rangle$. See Box 2 in Šizling et al. 202X. If S_X, S_Y, and $S_{X\cap Y}$ are mutually i-independent, the system follows $\begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| & 0 \\ 1 & 1 & 0 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| & 0 \\ 0 & 0 & J+1 & 2J\langle S\rangle \end{bmatrix} \Rightarrow S_{X\cap Y} = 2\langle S\rangle\frac{J}{J+1}\ \&\ S_Y = \frac{S_{X\cap Y}}{\beta'_{Sim}}\ \&$ $S_X = \frac{(J+1)S_{X\cap Y} - JS_Y}{J}$. It follows that $S_Y = 2\langle S\rangle\frac{J}{(J+1)\beta'_{Sim}} =_{(Eq.2)} 2\langle S\rangle\frac{R}{1+R}$. The last equality follows from Eq. 2. Finally, $S_X = \frac{2\langle S\rangle J - 2\langle S\rangle J\frac{R}{R+1}}{J} = 2\langle S\rangle\left(1 - \frac{R}{R+1}\right) = \frac{2\langle S\rangle}{R+1} = \frac{S_Y}{R}$. Because $0 < R \le 1$, then $S_Y \le S_X$, and we relabel $S_{poor} := S_Y$ and $S_{rich} := S_X$. If S_X, S_Y, and $S_{X\cap Y}$ are mutually i-dependent, then $\beta'_{Sim} = 1$ or $R = 1$ and the derived equations for S_{poor} and S_{rich} remain valid.

T17: *Jaccard index (J) is consistent with conceptual predefinition of co-occurrence if applied to more sites occupied by two species instead of more species that belong to two assemblages.*

Evidence: If we consider the Venn diagrams in Figure 3 (Šizling et al. 202X) to show two species ranges (sets of occupied sites), then the constraints on Co-occurrence follow:
$\mathrm{Min} = \mathrm{Co}[r4] = \mathrm{Co}[r3] < \mathrm{Co}[r1] < \mathrm{Co}[r2] < \mathrm{Co}[r6] < \mathrm{Co}[r5] = \mathrm{Max}$.
Because $0 = J[r4] = J[r3] < J[r1] < J[r2] < J[r6] < J[r5] = 1$, the evidence is completed.

T18: The system of Eq. 10 and Eq. 11 in Šizling et al. (202X) is mutually i-independent if $S_{X\cap Y} \neq 0$ and $S_X \not\equiv S_Y \not\equiv S_{X\cap Y} \not\equiv S_X$.
Evidence: If S_X, S_Y, and $S_{X\cap Y}$ are mutually i-independent, then the system follows
$\begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| \; 0 \\ 1 & 1 & 0 & 2\langle S\rangle \end{bmatrix} \sim \begin{bmatrix} J & J & -J-1 & 0 \\ 0 & \beta'_{Sim} & -1 & \| \; 0 \\ 0 & 0 & J+1 & 2J\langle S\rangle \end{bmatrix}$. If $J \neq -1$ and $J \neq 0$ and $\langle S\rangle \neq 0$ (mean S), this matrix provides a unique and non-trivial (i.e., non-zero) solution of the system.
Alternatively, $\mathbf{det}\begin{bmatrix} J & J & -J-1 \\ 0 & \beta'_{Sim} & -1 \\ 0 & 0 & 1+J \end{bmatrix} = J(1+J)\cdot\beta'_{Sim} \neq 0$, if $J, \beta'_{Sim} \neq 0$. Since $S_{X\cap Y} \neq 0$ implies $J \neq 0$, $\beta'_{Sim} \neq 0$, and $\langle S\rangle \neq 0$, it follows that $2J\langle S\rangle \neq 0$ 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. If the condition '$S_{X\cap Y} \neq 0$ and $S_X \not\equiv S_Y \not\equiv S_{X\cap Y} \not\equiv S_X$' is violated, then each index definition is a linear function of two variables and thus two equations are enough to get a unique solution, making one of the equations i-dependent on the others.

T19: *(i-dependence of three indices): If $S_{X\cap Y} \neq 0$ (i.e., $J \neq 0$), then the value of any index defined by Eq. 8 in Šizling et al. (202X) can be computed from values of Jaccard index (J) and Simpson nestedness (β'_{Sim}) using*

$$I = \frac{J(\beta'_{Sim}(k_1+k_2)-k_2+k_3)+\beta'_{Sim}k_2}{J(\beta'_{Sim}(l_1+l_2)-l_2+l_3)+\beta'_{Sim}l_2}. \qquad \text{(Equation 12 in Šizling et al. 202X)}$$

Evidence: Consider arbitrary $S_Y \leq S_X$. Then $S_{X\cap Y} \neq 0 \;\Rightarrow\; S_Y \neq 0 \;\Rightarrow\; S_X \neq 0$. The new index is defined as $I = \frac{k_1 S_{X\cap Y}+k_2 S_X+k_3 S_Y}{l_1 S_{X\cap Y}+l_2 S_X+l_3 S_Y}$ ($|l_2| + |l_3| > 0$ & $|k_2| + |k_3| > 0$), which is the definition Eq. 8 in Šizling et al. (202X). The Eq. 12 results as a solution of the determinant equation
$det\begin{vmatrix} -J & -J & J+1 \\ 0 & \beta'_{Sim} & -1 \\ Il_2 - k_2 & Il_3 - k_3 & Il_1 - k_1 \end{vmatrix} = 0$. The proof of the i-independence follows the reverse logic. (i) substitute definitions of J and β'_{Sim} into Eq. 12, (ii) convert Eq. 12 into the third equation of the system of linear equations and (iii) show that the determinant equals zero. In the same way is proved the equation for triplets I, J, R and I, β'_{Sim}, R. In these cases the restriction $S_{X\cap Y} \neq 0$ is replaced with $S_Y \neq 0$.

T20: *If Šizling et al. (2016) published only two partitioned indices (β_{sne}, β_{nps}) then the referred information would be unequivocal.*
Evidence: When we convert the indices reported by Šizling et al. (2016) into two partitioned indices, we got $\beta_{sne} = 0.1$, $\beta_{nps} = 0.61$ for the pre-agricultural landscape, and $\beta_{sne} = 0.00$, $\beta_{nps} = 0.51$ for early agricultural landscape. These indices suggest: $S_X \cong 31.01$, $S_Y \cong 22.99$, $S_{X\cap Y} \cong 15.48$ or $S_X \cong 41.63$, $S_Y \cong 12.37$, $S_{X\cap Y} \cong 2.28$ for pre-agricultural landscape; and $S_X \cong 32.00$, $S_Y \cong 32.00$, $S_{X\cap Y} \cong 21.60$ or $S_X \cong 48.32$, $S_Y \cong 15.68$, $S_{X\cap Y} \cong 0.00$ for early agricultural landscape (thesis T14 or Repository R4: calculator). This can be interpreted as either increasing species richness uniformity (R increases from 0.74 to 1), or decreasing species richness uniformity (R decreases from 0.74 to 0.32) species richness uniformity. This dual interpretation precludes any correct conclusion.

T21: *Any dimensionless index of turnover (defined by the constraints in Fig. 3) must have both S_X and S_Y in its denominator.*
Evidence: Dimensionless indices are ratios with numerators and denominators. Accurate measurement of turnover requires that information on both assemblages S_X, and S_Y be represented to capture their interaction fully. Including only one of S_X, or S_Y in the denominator while having both in the numerator creates an index without an upper limit. This contradicts the required bounded nature of the index, as specified by the constraints $\mathrm{T[r4]} = \mathrm{T[r3]} = \mathrm{Max}$ in Fig. 3.

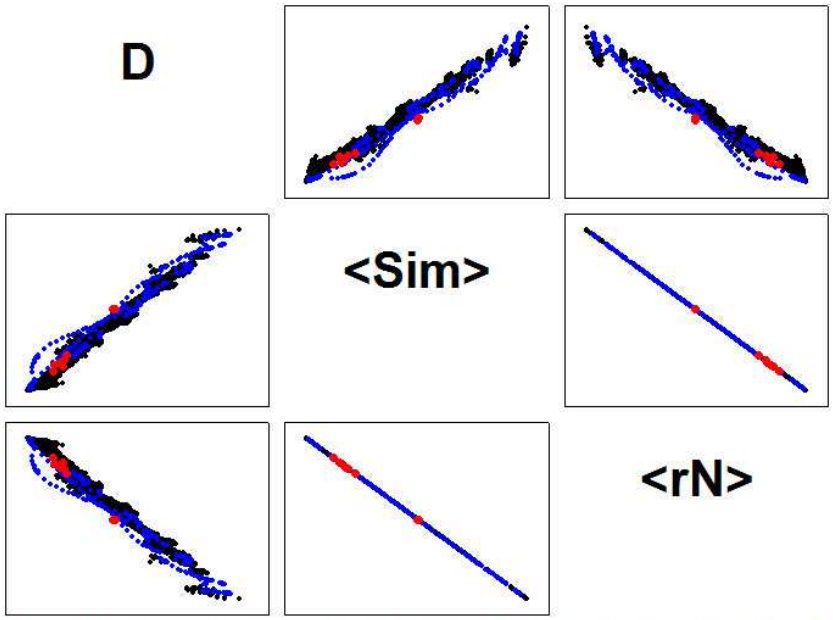

FIGURE R2 (average values): Relationships between indices of nestedness computed for more assemblages as simulated (30 sites x 30 species – black dots; 100 sites x 100 species – blue dots) and observed (red dots, R6 - data description). Matrices were generated to cover a wide spectrum of assemblages (see simulation details below). Indices include D (discrepancy by Brualdi & Sanderson, 1999); $\langle Sim \rangle$ (simple mean of the Simpson index across all pairs of adjacent plots; Lennon et al., 2001); and $\langle rN \rangle \equiv \langle \beta'_{Sim} \rangle$ (simple mean of nestedness as defined in Šizling et al., 2009); *D,* was standardized by dividing with the total number of incidences within the focal matrix (Greve et al., 2005). The indices scale with each other in a one-to-one manner, belonging to the same family and measuring the extent of nestedness. For exact evidence, see theses T03, and T05.

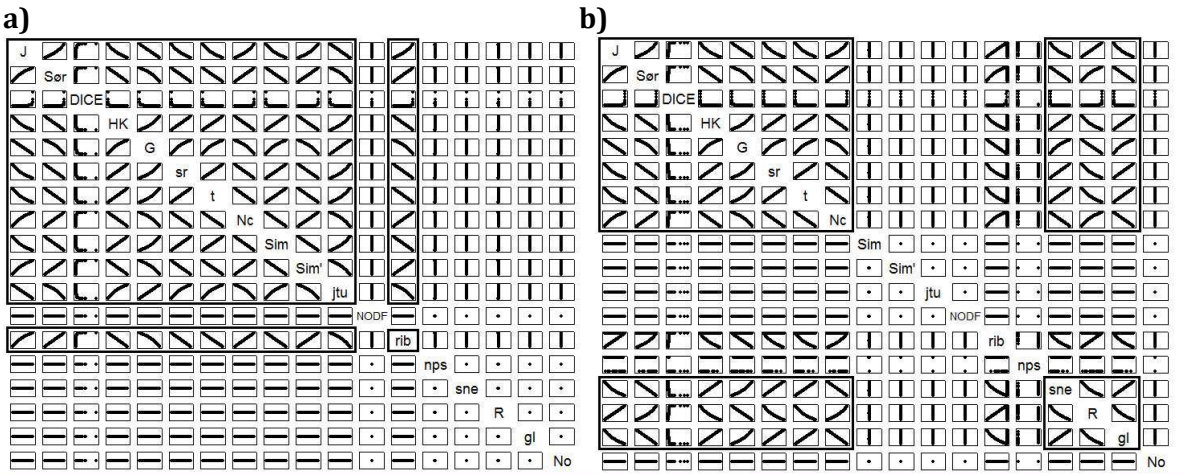

FIGURE R3 (interrelated assemblages): Relationship between pairwise indices where there is no contrast in species richness (a) and where assemblages are perfectly nested (b). As predicted, no contrast in species richness merges families of J, and β_{Sim} together, and perfect nestedness merges families of J, β_{sne}, and β_{gl} together. The evidence that β_{sne} shows variability where nestedness is perfect disqualifies this measure from being a proxy for nestedness component. Black rectangles delimit the merged families. For a detailed legend, see capture to Figure 5 in Šizling et al. 202X.

The random points for Fig. R3 and Figs 5,6 in Šizling et al. (202X) follow $S_{X\cap Y} = Rnd(1) * (1 - T), S_X = S_{X\cap Y} + trunc(100 * (1 - N) * Rnd(1))/100, S_Y = Trunc(100 * (Rnd(1) * (\min(1, S_X + C) - \max(0, S_X - C)) + \max(0, S_X - C))/100. T, N, C \in \langle 0,1 \rangle$ where T=1 indicates no shared species richness; T=0 indicates all possible overlaps in species richness; N=1 indicates absolute nestedness; N=0 indicates that nestednes is not constrained; C=1 indicates unlimited species richness contrast; C=0 indicates no species richness contrast; G=0.2 indicates a maximum 20% variation in the difference between species richness of the sites. In particular, T=0, N=0, C=1 (Figure 5 in Šizling *et al.* 202X); T=0, N=0, C=0.2 (Figure 6); T=0, N=0, C=0 (Figure R3A); T=0, N=1, C=1 (Figure R3B in Šizling *et al.* 202X).

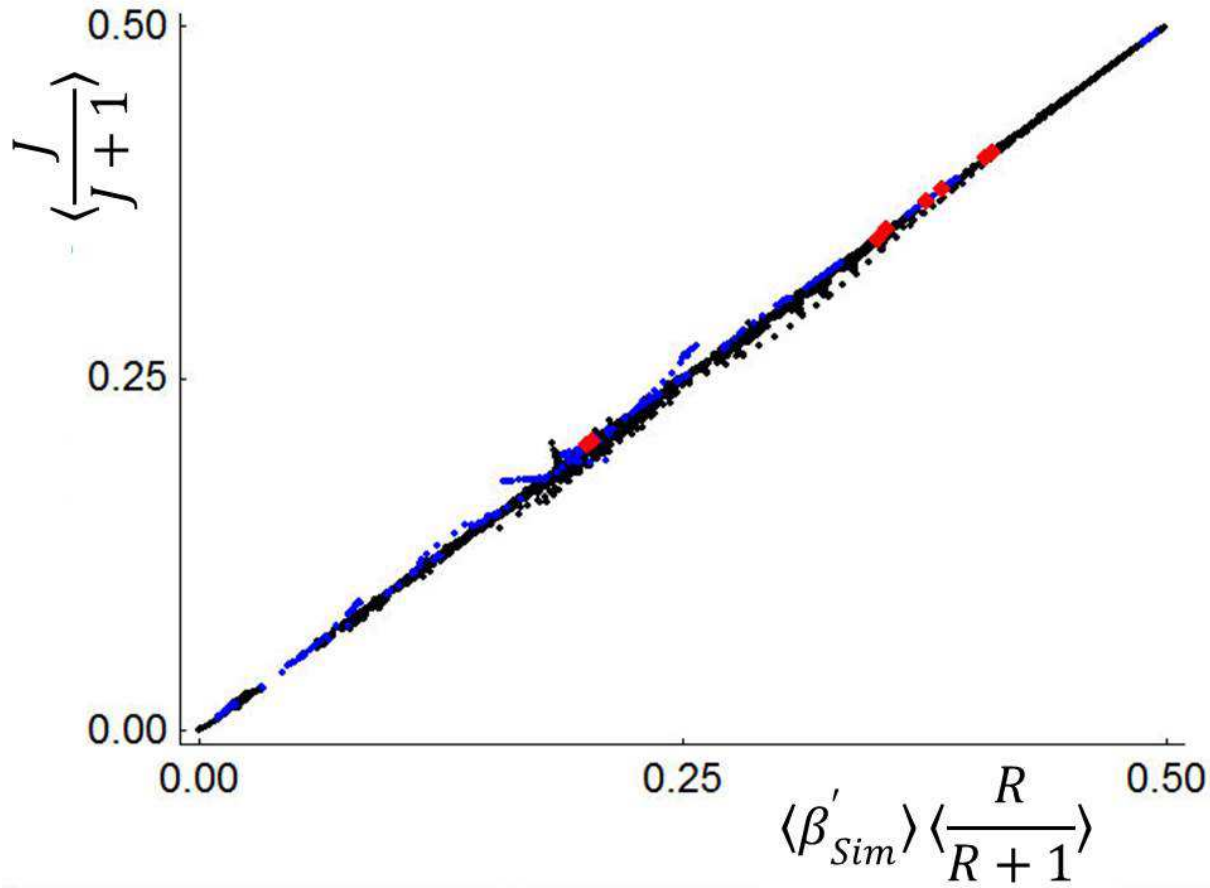

FIGURE R4 (Multisite Version of Eq. 2 in Šizling et al., 202X): The accuracy of the approximate Eq. 2 applied to mean values across multiple pairs of sites follows thesis T01.13. The exact equality between the X-values and Y values in the plot is affected by the covariance between β'_{Sim} and $R/(1+R)$, which is expected to be nearly zero. The figure shows the extent to which this expectation is met for simulated (30 sites x 30 species – black dots; 100 sites x 100 species – blue dots) and observed (red dots) multiple-site assemblages (R6 data description and RI2: Data). The y-axis shows the left side of the equation (i.e., the mean reversed turnover; $\langle J/(J+1) \rangle$), and the x-axis shows the right side of the equation (i.e., the product of the mean values of nestedness, $\langle \beta'_{Sim} \rangle$ and species richness uniformity $\langle R/R+1 \rangle$). The points approach an identity line, showing that the covariance approaches zero as expected. The 30x30 and 100x100 assemblages were produced by the algorithm adopted from Šizling et al. (2009).

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

R5 (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* (Figure R5), 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. 8 in Šizling et al. 202X) it holds that

$$I \stackrel{\text{def}}{=} \frac{k_1 S_{X\cap Y}+k_2 S_X+k_3 S_Y}{l_1 S_{X\cap Y}+l_2 S_X+l_3 S_Y} = \frac{k_1 \rho S_{X\cap Y}+k_2 \rho S_X+k_3 \rho S_Y}{l_1 \rho S_{X\cap Y}+l_2 \rho S_X+l_3 \rho S_Y}, (\rho \neq 0) \qquad \text{(Eq. R5)}$$

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. 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 (Figure R5) shows an increasingly even distribution of *J*-values. This mechanisms works for any index that can be expressed by Eq. 8 in Šizling et al. (202X), and the effect cannot be eliminated by inventing a new dimensionless index.

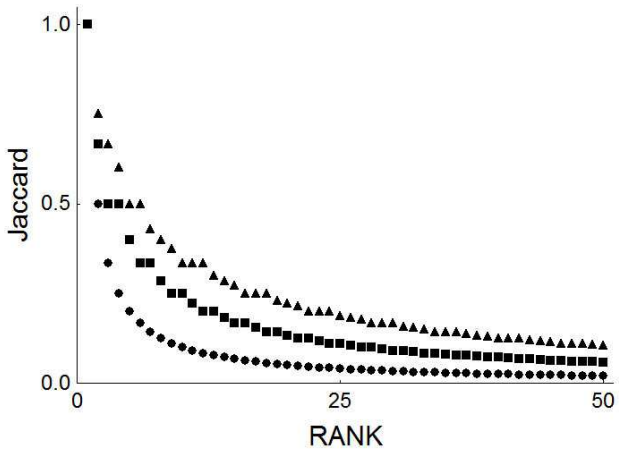

FIGURE R5: Rank plot of the first fifty values that can reach Jaccard index if S_X is fixed and S_Y and $S_{X\cap Y}$ vary within their limits ($1 \leq S_Y < \infty$ and $0 \leq S_{X\cap Y} \leq \min(S_X, S_Y)$); the three distributions on display correspond to $S_X = 1$ (circles), $S_X = 2$ (squares) and $S_X = 3$ (triangles).

R6 (data description):
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 Dataset RI.2).

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.

Cameron, K.A., M. Stibal, J. D. Žárský, E. Gozdereliler, M. Schostag and C. S. Jacobsen. 2016. Supraglacial bacterial community structures vary across the Greenland ice sheet. FEMS Microbiology Ecology 92, doi: 10.1093/femsec/fiv164

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