

## RESEARCH ARTICLE

# Decomposing diversity into measures of evenness, similarity, and richness

Bingzhang Chen  | Michael Grinfeld

Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

**Correspondence**

Bingzhang Chen, Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK.

Email: [bingzhang.chen@strath.ac.uk](mailto:bingzhang.chen@strath.ac.uk)

**Funding information**

Leverhulme Trust, Grant/Award Number: RPG-2020-389

**Abstract**

It has long been recognized that diversity has many measurable aspects, such as richness, evenness, and similarity among species. However, given a diversity index, it is unclear whether it necessarily can be decomposed into components that reflect these different aspects. Here, we present a scheme to decompose the Leinster and Cobbold diversity index, which subsumes and generalizes many other indices, into the components of richness, evenness and taxonomic similarity. Our approach addresses the problem that in general a vector of equal relative abundances does not maximize diversity. Furthermore, our approach uses all available information to give unbiased estimates of both evenness and similarity.

**KEYWORDS**

decomposition, diversity, evenness, Leinster–Cobbold diversity index, richness, similarity, taxonomic trees

**TAXONOMY CLASSIFICATION**

Biodiversity ecology

## 1 | INTRODUCTION

It has long been recognized that diversity has multiple aspects such as “richness” or “variety” reflecting the number of species present, dominance or rarity relations among the constituent species (“evenness” or “balance”), and “similarity” (or “disparity”) among the species (measures of distance between species based on taxonomic, phylogenetic, or functional traits relationships) (Daly et al., 2018; Leinster & Cobbold, 2012; Purvis & Hector, 2000; Stirling, 2007). Stirling (2007) even claims that “there seems no other obvious candidate for a fourth important general property of diversity beyond these three.”

In ecology, claimed biogeographical patterns of diversity often depend on the particular aspect of diversity being investigated. For example, Stuart-Smith et al. (2013) show that biodiversity hot spots can shift from the tropics to higher latitudes if in addition to

considering abundances, one takes account of functional traits similarity of species.

Several promising diversity indexes have been defined in the literature (Chao et al., 2014; Leinster & Cobbold, 2012; Rao, 1982; Stirling, 2007). However, a priori it is not clear whether from any given index one can extract information about richness, evenness, and species similarity. There is no argument why such a decomposition should exist and whether it has to be unique. In cases of non-uniqueness, one has to define what is the optimal way to perform such a decomposition. The situation here is similar to the decomposition of  $\gamma$  diversity into  $\alpha$  and  $\beta$  diversity indices either additively or multiplicatively (Anderson et al., 2011; Jost, 2007).

Being able to decompose a diversity index into biologically significant components would help us understand which aspect of diversity contributes to the changes in the relationship between

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

diversity and environmental variables (e.g., latitude or productivity). Such a decomposition is also helpful in understanding how different aspects of biodiversity influence the functioning of ecosystems (Hillebrand et al., 2018; Maureaud et al., 2019).

In an important study, van Dam (2019) proposed a straightforward decomposition of the Leinster–Cobbold (2012) diversity index that contains information about the richness, evenness, and similarity aspects of diversity. However, as will be detailed below, van Dam's approach fails to consider the fact that a homogenous community (i.e., in which all constituent species have equal relative abundances) may not maximize diversity if similarity among species is taken into account. Leinster and Meckes (2016) showed that in general, there exists a unique abundance vector  $\mathbf{p}^*$  that maximizes diversity that can differ from the equal abundance vector which we denote by  $\mathbf{p}_h$ . In other words, the maximal evenness of some communities is not achieved when all species have equal relative abundances. The criticism of van Dam's (2019) assumption that the component of evenness is maximal for an equal abundance vector can also be applied to the work of Daly et al. (2018) who stated that “the diversity measure is maximal for a fixed number of species  $S$  when all species' abundances are equal.”

In this study, we develop a decomposition approach that addresses the above problem. We focus on the Leinster–Cobbold (LC) index (Leinster & Cobbold, 2012) because it subsumes many other diversity indices that are widely used in ecology, such as the Hill numbers (Hill, 1973) and Rao's index (Rao, 1982; Ricotta & Moretti, 2011; Stirling, 2007), and has many other merits to be discussed later. For recent applications of this index, please see Mugabushaka et al. (2016) and Veresoglou et al. (2014). We take advantage of intrinsic properties of the LC index (e.g., the existence of the vector  $\mathbf{p}^*$ ) to achieve an unbiased (in a sense to be explained below) decomposition of the index.

The structure of this article is as follows. In Section 2, we discuss the notions of richness, evenness, and similarity using the material in Chao et al. (2014), Chiu et al. (2014), Daly et al. (2018), and Gregorius and Gillet (2022). Then, in Section 3, we comment biodiversity indices in general and present the required information about the LC index following Leinster and Cobbold (2012) and Leinster and Meckes (2016). In Section 4, we present first van Dam's, and then our decomposition and its consequences. Finally, in Section 5, we discuss the merits and limitations of our approach and suggest future work.

We also note that even if the quotation from Stirling (2007) reflects a deep property of biodiversity indices, it is not clear that a decomposition of a biodiversity index cannot contain additional terms. In fact, one interpretation of the decomposition we suggest contains an additional term, related to Hill's numbers and to the maximal diversity of a biocommunity (defined below). That is, the additional term does not carry any new information but is needed “to balance the books.” The quotation from Stirling also does not prescribe how richness, evenness, and similarity should be defined.

## 2 | DIVERSITY COMPONENTS

### 2.1 | Notation

For clarity, we first establish notation. Everywhere below we assume that the number of species (i.e., richness) in a community is fixed at  $n > 1$ .

We use  $\mathbf{p} = (p_1, \dots, p_n)$  to denote the unit vector of relative abundances, that is,

$$p_i > 0 \text{ for all } i = 1, \dots, n \text{ and } \sum_{i=1}^n p_i = 1.$$

We use the vector

$$\mathbf{p}_h = \left( \frac{1}{n}, \dots, \frac{1}{n} \right), \quad (1)$$

where the subscript  $h$  stands for “homogeneous” to represent the equal relative abundance of species in a community. Below  $\mathbf{1}$  will stand for an  $n$ -vector with all components equal to 1.

Next, we need to discuss  $n \times n$  matrices to record the information of species similarity. We denote the  $n \times n$  identity matrix by  $I_n$  which indicates that all species in the community are totally dissimilar with each other. We also use the notation  $J_n$  for the  $n \times n$  matrix of ones which represents an extreme case that all species are the same in the community (i.e., effectively we have only one species).

In this article, for simplicity, we will work with *ultrametric* matrices; this choice is motivated by the fact that similarity matrices (see Section 3.2) constructed using taxonomic trees are necessarily ultrametric and using ultrametric matrices simplifies the theory of Leinster and Mecke (2016). For more information on ultrametric matrices, please see Dellacheria et al. (2014), Leinster (2013), and Leinster and Meckes (2016).

### 2.2 | Evenness

Classically, the “most even” population of  $n$  species is one for which the vector of relative abundances is the homogeneous vector  $\mathbf{p}_h$ , that is, one where every species is equally represented (Daly et al., 2018). For thorough discussions of evenness see, for example, Chao et al. (2014), Chao and Ricotta (2019), Gregorius and Gillet (2022), and references therein. As the case of richness discussed below, the term “evenness” itself seems to be precluding discussion. As rightly pointed in Gregorius and Gillet (2022), a definition of maximal evenness in terms of  $\mathbf{p}_h$  leaves open the discussion of what would constitute “maximum unevenness.” Instead of “evenness,” van Dam (2019) uses “balance,” which seems to us a better term; this is the concept which, after defining it properly (see (12)) we will use below.

## 2.3 | Richness

Richness is usually defined as the number of species (Daly et al., 2018). Such a definition is open to the same criticism as the notion of maximal evenness defined by  $p_h$  discussed above. It is “species-centric” without considering species similarity, and takes into account only the last level of taxonomic classification. We prefer to call the number of species  $n$  “species richness.”

## 2.4 | Similarity

In this section, we discuss the construction of taxonomic similarity matrices  $Z$  for a community with  $n$  species.

The usual way of constructing similarity matrices  $Z$  (which are automatically ultrametric) is to assign distances between different levels of a taxonomic tree. Then the taxonomic distance  $d(i, j)$  between two species is the sum of distances from the nodes corresponding to these species to the first common node. Once the distance between nodes is defined, we could put  $Z_{ij} = e^{-d(i, j)}$  or, if the maximal distance in the tree has been normalized to 1, we could put  $Z_{ij} = 1 - d(i, j)$ . As an example, consider the tree in Figure 1 and set the species–genus and the genus–family distance to be 0.3. If we use the additive recipe for defining  $Z$ , that is, put  $Z_{ij} = 1 - d(i, j)$ , we obtain the similarity matrix

$$Z_1 = \begin{pmatrix} 1 & 0.7 & 0.4 \\ 0.7 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix}. \quad (2)$$

## 3 | THE LEINSTER–COBBOLD DIVERSITY INDEX

The Leinster and Cobbold (LC) diversity index was introduced in Leinster and Cobbold (2012). The LC index is a far-reaching generalization of Hill numbers, for discussions of which, see Chiu et al. (2014). For details on its properties, see Leinster and Cobbold (2012) and Leinster and Meckes (2016); here, we just collect the bare minimum in the framework of taxonomic (ultrametric) similarity matrices;

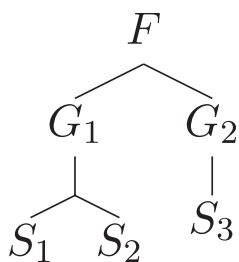


FIGURE 1 An example taxonomic tree with the root at the family level.

ultrametricity is the blanket assumption on similarity matrices we make from now on.

Note that compared to the diversity indices proposed by Chao et al. (2014), the LC index has the flexibility to take into account taxonomic, phylogenetic, and functional diversity simultaneously, although the resulting similarity matrices will in general not be ultrametric. We leave this as a more general case for future study.

Before we start discussing the LC index, let us review the requirements that have to be satisfied by any diversity index.

### 3.1 | Requirements for an diversity index

Leinster and Cobbold (2012, pp. 482–483) formulate conditions that any biodiversity index should satisfy; they number nine such conditions. The LC index satisfies all of them, that is, in brief, the LC index provides effective numbers, is modular, has the property of replication, is symmetric, is not changed if a species identical to one of those already present is added or by adding a new species of zero abundance, decreases if similarity between species increases, is larger if similarities between species are ignored, and always takes a value between 1 and the number of species  $n$ . Hence, the LC index is a legitimate starting point for the decomposition work we are interested in. We also note that while Leinster and Cobbold (2012) write that any index with sensitivity parameter  $q$  should be a monotone decreasing function of  $q$ , they do not add this criterion to the requirements on an index, as we do below.

### 3.2 | Definition of the LC index

As in the definition of Hill numbers, below  $q \in [0, \infty)$  is the sensitivity parameter, measuring the (un)importance given to rare species. Then for a community of  $n$  species with relative abundance vector  $\mathbf{p}$  and similarity matrix  $Z$ , we have

**Definition 1.** The LC diversity of order  $q$  is

$$F(Z, \mathbf{p}, q) := \left( \sum_{i=1}^n p_i (Z\mathbf{p}^T)_i^{q-1} \right)^{1/(1-q)}. \quad (3)$$

Leinster and Cobbold (2012) show that the LC index indeed satisfies all the requirements imposed on a diversity index in Section 3.1.

Note that Leinster and Cobbold (2012) use a different notation, similar to the Hill number notation in the literature; they denote the right-hand side of (3) by  ${}^q D^Z(\mathbf{p})$ . We prefer the notation used here as it clearly shows functional dependencies and allows easy generalization, which we discuss briefly in Section 5. We collected the required properties of the LC index below and in Section 3.3.

The LC index has the following properties:

- $F(\mathbf{p}, Z, q)$  is a monotone decreasing function of  $q$ ;
- $F(\mathbf{p}, Z, q) < F(\mathbf{p}, I_n, q)$  for all  $q$  if  $Z \neq I_n$ ;
- $F(\mathbf{p}_h, I_n, q) = n$  for all  $q$ ;
- $F(\mathbf{p}, J_n, q) = 1$  for all  $q$ .

For proofs of (a) and (b), see Leinster and Cobbold (2012); the other statements are immediate.

Following Leinster and Meckes (2016), we now discuss the concept of a *maximally balanced abundance vector* for a community of  $n$  species with an ultrametric similarity matrix  $Z$  which is the key starting point of our decomposition approach.

### 3.3 | $\mathbf{p}^*$ : A crucial property of the LC index

Leinster and Meckes (2016) prove that for each similarity matrix  $Z$ , there exists a unique abundance vector  $\mathbf{p}^*$  that maximizes the diversity index  $F(Z, \mathbf{p}, q)$  for every value of  $q \in [0, \infty)$ . We call  $\mathbf{p}^*$  the *maximally balanced abundance vector* (which replaces  $\mathbf{p}_h$  if taxonomic similarity is taken into account) and call  $F(Z, \mathbf{p}^*, q)$  (which is independent of  $q$ ) the maximal diversity of a biocommunity with similarity matrix  $Z$ .

If  $Z$  is ultrametric, computing  $\mathbf{p}^*$  is a simple matter of solving a system of linear equations and normalizing:

$$p_i^* = \frac{w_i}{\sum_{j=1}^n w_j} \quad (4)$$

where  $\mathbf{w}$  solves the system of equations  $Z\mathbf{w} = \mathbf{1}$ , with  $\mathbf{1}$  is a column vector of ones.

For the similarity matrix  $Z_1$  in (2), we have

$$\mathbf{p}^* \approx (0.286, 0.286, 0.429).$$

which differs from  $\mathbf{p}_h = (0.333, 0.333, 0.333)$ . We can check that when  $q = 0$ ,  $F(Z, \mathbf{p}^*, q) = 1.522$  which is greater than  $F(Z, \mathbf{p}_h, q) = 1.508$ . For  $q = 1, 2$ ,  $F(Z, \mathbf{p}^*, q)$  remains unchanged, but  $F(Z, \mathbf{p}_h, 1) = 1.504$  and  $F(Z, \mathbf{p}_h, 2) = 1.500$ .

Note that an alternative way of computing  $\mathbf{p}^*$  is provided in Leinster and Meckes (2016, lemma 6).

There are similarity matrices  $Z \neq I_n$  for which  $\mathbf{p}^* = \mathbf{p}_h$ . However, it is beyond the scope of this study to characterize exhaustively the set of such matrices. For now, we simply define a taxonomic tree to be *taxonomically equilibrated* if  $\mathbf{p}^* = \mathbf{p}_h$ . Of course if at each level of the tree all the nodes have the same degree, the taxonomic tree is equilibrated. However, the converse statement is not true. An example is provided by the tree in Figure 2.

It is not hard to show that assigning of species–genus, genus–family, and family–order distances to be 0.25, and using the additive recipe to compute  $Z_{ij}$ , results in a similarity matrix for which  $\mathbf{p}^* = \mathbf{p}_h$ . Thus, there is a trichotomy of taxonomic trees: those for which  $\mathbf{p}^* = \mathbf{p}_h$  holds for every assignment of distances; those where such assignments can be chosen, as in Figure 2, and such

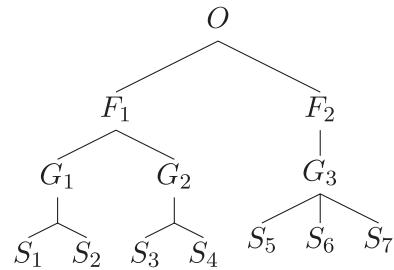


FIGURE 2 A taxonomic tree that does not have the same degree at all nodes but allows  $\mathbf{p}^* = \mathbf{p}_h$ .

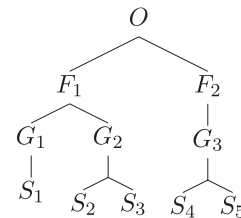


FIGURE 3 A taxonomic tree for which  $\mathbf{p}^* \neq \mathbf{p}_h$  is guaranteed.

that no assignment of distances results in a homogeneous maximally balanced abundance vector; an example of such a tree is shown in Figure 3.

## 4 | AN UNBIASED DECOMPOSITION SCHEME

We are now ready to propose a decomposition scheme for the LC index. It is best to start with the decomposition scheme suggested by van Dam (2019) and see why it has to be modified. van Dam writes

$$F(\mathbf{p}, Z, q) = \frac{F(\mathbf{p}, I_n, q)}{F(\mathbf{p}_h, I_n, q)} \cdot \frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}, I_n, q)} \cdot F(\mathbf{p}_h, I_n, q). \quad (5)$$

The second fraction in the right-hand side of (5),

$$D_{VD}(\mathbf{p}, Z, q) := \frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}, I_n, q)}, \quad (6)$$

is clearly a measure of dissimilarity, while the first fraction should therefore be measure of balance,

$$B_{VD}(\mathbf{p}, q) := \frac{F(\mathbf{p}, I_n, q)}{F(\mathbf{p}_h, I_n, q)}; \quad (7)$$

$F(\mathbf{p}_h, I_n, q) = n$ , so the last term on the right-hand side is the standard measure of richness. Note that, as it should be, the two fractions always lie in the interval  $(1/n, 1]$  (property (b) of the LC index ensures that the value  $1/n$  cannot be reached).

The difficulty with (5) is in accepting (7) as a measure of balance. First of all, it does not use the information contained in the taxonomic similarity matrix  $Z$ . Secondly, it introduces  $\mathbf{p}_h$  which a priori has no meaning in theories that take similarity into account.  $B_{VD}(\mathbf{p}, q)$  will

take the maximal value 1 when  $\mathbf{p} = \mathbf{p}_h$  for all values of  $q$ , while in the framework of the LC index, the measure of balance should be maximal when  $\mathbf{p} = \mathbf{p}^*$ . This argument makes the van Dam decomposition unusable in the framework of the LC index for which it is intended.

The dissimilarity measure  $\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}_h, q)}$  does use information from both  $\mathbf{p}$  and  $Z$ . Hence, the decomposition (5) can be said to be *asymmetrically biased*.

Note that an alternative decomposition

$$F(\mathbf{p}, Z, q) = \frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}_h, Z, q)} \cdot \frac{F(\mathbf{p}_h, Z, q)}{F(\mathbf{p}_h, I_n, q)} \cdot F(\mathbf{p}_h, I_n, q), \quad (8)$$

in addition to having the drawbacks of (5) of using the often irrelevant abundance vector  $\mathbf{p}_h$  and having an asymmetrically biased dissimilarity measure, also has the problem that its measure of balance, the first term in the right-hand side, can take values larger than 1 if  $Z$  is not a similarity matrix of an taxonomically equilibrated tree and  $\mathbf{p} = \mathbf{p}^*$ .

To ensure normalization and to be consistent with Leinster and Meckes (2016), consider the following decomposition:

$$F(\mathbf{p}, Z, q) = \frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}^*, Z, q)} \cdot \frac{F(\mathbf{p}^*, Z, q)}{F(\mathbf{p}^*, I_n, q)} \cdot F(\mathbf{p}^*, I_n, q). \quad (9)$$

It is asymmetrically biased as the second factor, a dissimilarity measure, does not use the information from the actual abundance vector  $\mathbf{p}$ . We will discuss the interpretation of  $F(\mathbf{p}^*, I_n, q)$  later.

To obtain an **unbiased** decomposition, we multiply (5) and (9) and take a square root. The result is

$$F(\mathbf{p}, Z, q) = \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}^*, Z, q)} \frac{F(\mathbf{p}, I_n, q)}{F(\mathbf{p}_h, I_n, q)}} \cdot \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}_h, I_n, q)} \frac{F(\mathbf{p}^*, Z, q)}{F(\mathbf{p}^*, I_n, q)}} \cdot \sqrt{nF(\mathbf{p}^*, I_n, q)}, \quad (10)$$

since  $F(\mathbf{p}_h, I_n, q) = n$ .

Note that the last term in the right-hand side of (10) can be rewritten as

$$\sqrt{nF(\mathbf{p}^*, I_n, q)} = \sqrt{\frac{F(\mathbf{p}^*, I_n, q)}{n}} n := E(Z, q)n, \quad (11)$$

which is appropriate notation as  $E(Z, q)$  is determined by  $Z$ . The term  $E(Z, q)$  expresses the lack of equilibration in the taxonomic tree and always lies in the interval  $(0, 1]$ . In the case of similarity matrices of taxonomically equilibrated trees, for which we have  $\mathbf{p}^* = \mathbf{p}_h$  and hence  $F(\mathbf{p}^*, I_n, q) = n$ , we have that  $E(Z, q) = 1$ . If  $Z$  does not correspond to a taxonomically equilibrated taxonomic tree,  $F(\mathbf{p}^*, I_n, q)$  is dependent on  $q$ ;  $E(Z, q)$  is monotone decreasing in  $q$ ; the limit of  $E(Z, q)$  as  $q \rightarrow \infty$  exists and may reflect properties of  $Z$ ; we leave analysis of  $E(Z, q)$  to future work.

The difficulty with the decomposition (10) as written is in interpreting

$$B_0(\mathbf{p}, Z, q) = \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}^*, Z, q)} \frac{F(\mathbf{p}, I_n, q)}{F(\mathbf{p}_h, I_n, q)}}. \quad (12)$$

as a measure of balance as it is not necessarily maximized by  $\mathbf{p} = \mathbf{p}^*$  for all values of  $q$  different from zero.

However, note that (10) can also be written as

$$F(\mathbf{p}, Z, q) = \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}^*, Z, q)}} \cdot \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}, I_n, q)}} \cdot \sqrt{F(\mathbf{p}^*, Z, q)F(\mathbf{p}, I_n, q)}, \quad (13)$$

In both (13), we have an unbiased measure of balance,

$$B(\mathbf{p}, Z, q) := \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}^*, Z, q)}}, \quad (14)$$

which correctly takes its maximal value 1 at  $\mathbf{p} = \mathbf{p}^*$ , and an unbiased measure of dissimilarity

$$D(\mathbf{p}, Z, q) := \sqrt{\frac{F(\mathbf{p}, Z, q)}{F(\mathbf{p}, I_n, q)}}, \quad (15)$$

which for any given  $\mathbf{p}$  is maximized by choosing  $I_n$ .

We suggest to define

$$R(\mathbf{p}, Z, q) := \sqrt{F(\mathbf{p}^*, Z, q)F(\mathbf{p}, I_n, q)}, \quad (16)$$

the harmonic mean of Hill's number of order  $q$ ,  $F(\mathbf{p}, I_n, q)$ , and  $F(\mathbf{p}^*, Z, q)$ , the maximal diversity for all  $q$  for a given similarity matrix  $Z$  (see Section 3.3), to be our notion of *effective richness* of order  $q$ . Note that it arises naturally in the process of unique unbiased decomposition of the LC index. This is not an inconsistent definition: if  $Z = I_n$  and  $\mathbf{p} = \mathbf{p}_h = \mathbf{p}^*$ , that is, for a maximally balanced population of totally unrelated species, we have that  $R(\mathbf{p}_h, I_n, q) = n$ , the species richness, for all  $q \geq 0$ . So our notion of richness truly generalizes species richness and takes into account similarity among species, the abundance vector  $\mathbf{p}$  and the sensitivity parameter  $q$ .

Hence, with this definition of richness, (13) becomes

$$F(\mathbf{p}, Z, q) = B(\mathbf{p}, Z, q)D(\mathbf{p}, Z, q)R(\mathbf{p}, Z, q). \quad (17)$$

An advantage of our non-“species-centric” definition of richness is that it allows to compare two communities with different number of species.

## 5 | DISCUSSION

We have proposed a decomposition of the LC index into components that include well-defined measures of balance (evenness) and (dis)similarity. Compared to a previous version of decomposition (van Dam, 2019), our approach delivers measures of balance and of dissimilarity of the community that employ all the available information and do not use the homogeneous abundance vector  $\mathbf{p}_h$  that does not have any biological significance in realistic communities.

The price for having three components in the decomposition was having to redefine richness as in (16). We could alternatively write (13) as

$$F(\mathbf{p}, Z, q) = B(\mathbf{p}, Z, q)D(\mathbf{p}, Z, q)n \cdot \frac{R(\mathbf{p}, Z, q)}{n}, \quad (18)$$

in which case we must retain in our decomposition an in general  $q$ -dependent fourth term.

However, if the choice is between (17) and (18), a decomposition into four terms, we argue that (17) is preferable: the last term in the right-hand side of (18) can only be interpreted as the ratio between a measure of “effective richness” and the species richness and hence provides the same information as (17).

One could ask whether one could not start with the components (14), (15),  $n$ , and define an index  $l(\mathbf{p}, Z, q)$  by

$$l(\mathbf{p}, Z, q) = B(\mathbf{p}, Z, q)D(\mathbf{p}, Z, q)n.$$

However, this “index” is not decreasing in  $q$ , so does not define a bona fide diversity index.

## 5.1 | Incorporating more information streams

Although Stirling (2007) argues that there are only three aspects (i.e., richness, balance, and disparity) that should be considered in diversity and though LC index also uses only three “information streams”: the number of species  $n$ , the relative abundance vector  $\mathbf{p}$ , and the similarity matrix  $Z$ , we could in theory consider a diversity index  $F(c_1, \dots, c_m; q)$  with  $m > 3$ , where  $c_1, \dots, c_m$  are various information streams. For example, one may wish to consider both the phylogenetic and functional diversity in the diversity index and we can have more than one  $Z$ . We could then follow the decomposition process of Section 4: find  $m!$  biased decompositions, multiply them together, take the  $m!$ -th root, and simplify. However, this is already unwieldy in the case of  $m = 3$ .

One approach to circumvent this problem is to take advantage of the flexibility of the LC index by incorporating different information streams into one unified similarity matrix. As explained in Section 2.4, one can define a similarity matrix by setting  $Z_{ij} = e^{-d(i,j)}$ , where  $d(i,j)$  is some suitably defined distance between species  $i$  and  $j$ . Hence, incorporating more information streams can be thought about as changing the distance function  $d(\cdot, \cdot)$ . In the process of incorporating such information, such as functional similarity, the ultrametricity of the similarity matrix may be lost; it is possible that the resulting function  $d(\cdot, \cdot)$  will no longer be a metric, becoming more generally a divergence measure. The point is that  $\mathbf{p}$  and (a suitably redefined)  $Z$  dependence of a diversity index is sufficient to incorporate all relevant information.

## 5.2 | Limitations and future work

Our approach relies on the computation of  $\mathbf{p}^*$  which is not always possible if the similarity matrix  $Z$  is not ultrametric. Leinster and Meckes (2016) also note that if the number of species becomes high enough, the computation of  $\mathbf{p}^*$  in the non-ultrametric case can be quite demanding. Extension of our work to the non-ultrametric case

relevant to phylogenetic trees and physiological traits similarity, and exploration of the ensuing computational issues is therefore a priority.

Sensitivity of similarity matrices, and hence the LC index and our decomposition, to the abundance of a species that is significantly distinct from other species in a community, is an area that needs to be explored.

Another good (linear algebra) research question is an exhaustive characterization of the trichotomy in taxonomic trees with respect to taxonomic tree equilibration, that is, understanding for which taxonomic trees any choice of similarity matrix leads to  $\mathbf{p}^* = \mathbf{p}_H$ , as opposed to ones only possible for particular choices of a similarity matrix, or impossible for any similarity matrix.

## AUTHOR CONTRIBUTIONS

**Bingzhang Chen:** Conceptualization (supporting); data curation (lead); funding acquisition (lead); investigation (supporting); methodology (supporting); software (equal); visualization (equal); writing – review and editing (supporting). **Michael Grinfeld:** Conceptualization (lead); formal analysis (lead); methodology (lead); writing – original draft (lead); writing – review and editing (lead).

## ACKNOWLEDGEMENTS

The authors sincerely thank the editors, and two anonymous reviewers for thorough and insightful comments that substantially improved the article.

## FUNDING INFORMATION

B. Chen was supported by a Leverhulme Trust Research Project Grant (RPG-2020-389).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

This article does not involve any dataset.

## ORCID

Bingzhang Chen  <https://orcid.org/0000-0002-1573-7473>

## REFERENCES

- Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N., Cornell, H., Comita, L., Davies, K., & Harrison, S. (2011). Navigating the multiple meanings of  $\beta$  diversity: A roadmap for the practicing ecologist. *Ecology Letters*, 14, 19–28.
- Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45, 297–324.
- Chao, A., & Ricotta, C. (2019). Quantifying evenness and linking it to diversity, beta diversity and similarity. *Ecology*, 100, e02852.
- Chiu, C.-H., Jost, L., & Chao, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs*, 81, 21–44.
- Daly, A. J., Baetens, J. M., & De Baets, B. (2018). Ecological diversity: measuring the unmeasurable. *Mathematics*, 6, 119.

- Dellacheria, C., Martinez, S., & San Martin, J. (2014). *Inverse M-matrices and ultrametric matrices*. Springer.
- Gregorius, H.-R., & Gillet, E. M. (2022). The concept of evenness/unevenness: Less evenness or more unevenness? *Acta Biotheoretica*, 70, 1–28.
- Hill, M. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54, 427–432.
- Hillebrand, H., Blasius, B., Borer, E. T., Chase, J. M., Downing, J. A., Eriksson, B. K., Filstrup, C. T., Harpole, W., Hodapp, D., Larsen, S., & Lewandowska, A. M. (2018). Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology*, 55(1), 169–184.
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, 88, 2427–2439.
- Leinster, T. (2013). The magnitude of metric spaces. *Documenta Mathematica*, 18, 857–905.
- Leinster, T., & Cobbold, C. (2012). Measuring diversity: The importance of species similarity. *Ecology*, 93, 477–489.
- Leinster, T., & Meckes, M. W. (2016). Maximizing diversity in biology and beyond. *Entropy*, 18, 88.
- Maureaud, A., Hodapp, D., van Denderen, P. D., Hillebrand, H., Gislason, H., Spaanheden Dencker, T., Beukhof, E., & Lindegren, M. (2019). Biodiversity–ecosystem functioning relationships in fish communities: Biomass is related to evenness and the environment, not to species richness. *Proceedings of the Royal Society B: Biological Sciences*, 286(1906), 20191189.
- Mugabushaka, A.-M., Kyriakou, A., & Papazoglou, T. (2016). Bibliometric indicators of interdisciplinarity: The potential of the Leinster–Cobbold diversity indices to study disciplinary diversity. *Scientometrics*, 107, 593–607.
- Purvis, A., & Hector, A. (2000). Getting the measure of biodiversity. *Nature*, 405(6783), 212–219.
- Rao, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, 21(1), 24–43.
- Ricotta, C., & Moretti, M. (2011). CWM and Rao's quadratic diversity: A unified framework for functional ecology. *Oecologia*, 167(1), 181–188.
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707–719.
- Stuart-Smith, R. D., Bates, A. E., Lefcheck, J. S., Duffy, J. E., Baker, S. C., Thomson, R. J., Stuart-Smith, J. F., Hill, N. A., Kininmonth, S. J., Airolidi, L., & Becerro, M. A. (2013). Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, 501(7468), 539–542.
- van Dam, A. (2019). Diversity and its decomposition into variety, balance and disparity. *Royal Society Open Science*, 6, 90452.
- Veresoglou, S. D., Powell, J. R., Davison, J., Lekberg, Y., & Rillig, M. C. (2014). The Leinster and Cobbold indices improve inferences about microbial diversity. *Fungal Ecology*, 11, 1–7.

**How to cite this article:** Chen, B., & Grinfeld, M. (2024). Decomposing diversity into measures of evenness, similarity, and richness. *Ecology and Evolution*, 14, e10952. <https://doi.org/10.1002/ece3.10952>