

# The partitioning of diversity: showing Theseus a way out of the labyrinth

Francesco de Bello, Sébastien Lavergne, Christine N. Meynard, Jan Lepš & Wilfried Thuiller

## Abstract

A methodology for partitioning of biodiversity into  $\alpha$ ,  $\beta$  and  $\gamma$  components has long been debated, resulting in different mathematical frameworks. Recently, use of the Rao quadratic entropy index has been advocated since it allows comparison of various facets of diversity (e.g. taxonomic, phylogenetic and functional) within the same mathematical framework. However, if not well implemented, the Rao index can easily yield biologically meaningless results and lead into a mathematical labyrinth. As a practical guideline for ecologists, we present a critical synthesis of diverging implementations of the index in the recent literature and a new extension of the index for measuring  $\beta$ -diversity. First, we detail correct computation of the index that needs to be applied in order not to obtain negative  $\beta$ -diversity values, which are ecologically unacceptable, and elucidate the main approaches to calculate the Rao quadratic entropy at different spatial scales. Then, we emphasize that, similar to other entropy measures, the Rao index often produces lower-than-expected  $\beta$ -diversity values. To solve this, we extend a correction based on equivalent numbers, as proposed by Jost (2007), to the Rao index. We further show that this correction can be applied to additive partitioning of diversity and not only its multiplicative form. These developments around the Rao index open up an exciting avenue to develop an estimator of turnover diversity across

different environmental and temporal scales, allowing meaningful comparisons of partitioning across species, phylogenetic and functional diversities within the same mathematical framework. We also propose a set of R functions, based on existing developments, which perform different key computations to apply this framework in biodiversity science.

**Keywords:** Biodiversity turnover; Biogeography; Community assembly; Functional diversity; Phylogenetic diversity; Simpson index.

## Introduction

The partitioning of biodiversity into different spatial components is critical to understand processes underlying species distributions and diversity turnover (Magurran 2004; Ackerly & Cornwell 2007; Prinzing et al. 2008; de Bello et al. 2009). In particular, proper management of ecosystems requires that we understand the processes by which  $\beta$ -diversity (i.e. the diversity across habitats or communities) is generated and maintained (Legendre et al. 2005). Several indices and mathematical frameworks have been developed for these purposes (Lande 1996; Veech et al. 2002; Crist & Veech 2006), making it possible to answer different ecological questions and unavoidably producing little consensus among methods (Koleff et al. 2003). Overall, it is widely accepted that the total diversity of a region ( $\gamma$ -diversity) can be partitioned into within-community ( $\alpha$ -diversity) and among-communities ( $\beta$ -diversity; Whittaker 1975; Magurran 2004 and references therein) components. Partitioning of diversity could then be additive (e.g.  $\gamma = \alpha + \beta$ ) or multiplicative ( $\gamma = \alpha * \beta$ ), depending on the models and mathematical indices used (Veech et al. 2002; Ricotta 2005a; Jost 2007; Jost et al. 2010).

Among the different existing mathematical frameworks (Magurran 2004), Rao's quadratic entropy index (1982) can provide a general approach for partitioning biodiversity into  $\alpha$ ,  $\beta$  and  $\gamma$  components. Indeed, the Rao entropy is currently the only existing estimator of diversity that formally combines different measures of species dissimilarity (e.g. phylogenetic or functional) with relative species abundances, providing a standardized methodology

**de Bello, F.** (corresponding author, fradebello@ctfc.es), **Lavergne, S.** (sebastien.lavergne@ujf-grenoble.fr) & **Thuiller, W.** (wilfried.thuiller@ujf-grenoble.fr): Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, 38041, Grenoble Cedex 9, France.

**Meynard, C.N.** (cnmeynard@gmail.com): UMR 5554-ISEM, Université Montpellier II, Place Eugène Bataillon, CC 065, 34095 Montpellier Cedex 5, France.

**Lepš, J.** (suspa@prf.jcu.cz): Faculty of Science, University of South Bohemia and Institute of Entomology, Biology Centre of ASCR, Branišovská 31, CZ-37005 České Budějovice, Czech Republic.

Present address: Department of Functional Ecology, Institute of Botany, Czech Academy of Sciences, Dukelská 135, CZ-379 82 Třeboň, Czech Republic.

|   |              |                |          |                   |                 |                 |
|---|--------------|----------------|----------|-------------------|-----------------|-----------------|
|  | <b>J V S</b> | <b>1195</b>    | <b>B</b> | Dispatch: 28.4.10 | Journal: JVS    | CE: Mahendra    |
|   | Journal Name | Manuscript No. |          | Author Received:  | No. of pages: 9 | Te: Aswani/Mini |

1 applicable to compare  $\alpha$ ,  $\beta$  and  $\gamma$  components  
 2 between different facets of diversity (e.g. taxonomic,  
 3 phylogenetic and functional diversity; Pavoine et al.  
 4 2004; Ricotta 2005a; Hardy & Senterre 2007). Fur-  
 5 thermore, the index provides one of the few direct  
 6 measures of species redundancy within and among  
 7 biological communities (de Bello et al. 2007, 2009).  
 8 These unique properties of the Rao index could  
 9 open new perspectives to understand mechanisms  
 10 driving the turnover of diversity along environ-  
 11 mental and temporal scales.

12 However, some key methodological issues re-  
 13 garding the spatial partitioning of diversity with the  
 14 Rao index have been hotly debated in the recent lit-  
 15 erature (Ricotta 2005a, b; Hardy & Jost 2008;  
 16 Villegger & Moullot 2008; de Bello et al. 2009).  
 17 Whether the index could lead to negative  $\beta$  values  
 18 has been discussed (Hardy & Jost 2008; Villegger &  
 19 Moullot 2008), with no clear agreement yet on how  
 20 Rao's index should be computed. Moreover, the in-  
 21 dex has not been able to offer a robust estimation of  
 22  $\beta$ -diversity (de Bello et al. 2009), giving system-  
 23 atically low estimates of  $\beta$ -diversity, even for  
 24 complete species replacement between communities.  
 25 To help ecologists find a way out of this mathemat-  
 26 ical labyrinth, we provide clarification and techni-  
 27 cal guideline to derive a more realistic partition  
 28 of diversity using the Rao index (i.e. producing a  
 29  $\beta$ -diversity that behaves as ecologists would expect).  
 30 We discuss these issues with numerical examples and  
 31 a case study, which demonstrate how promising a  
 32 corrected version of the Rao index for the partition-  
 33 ing of  $\alpha$ ,  $\beta$  and  $\gamma$  components of diversity can be. As  
 34 the idea for the study was conceived during the IAVS  
 35 meeting held in Crete (2009), we hope this study will  
 36 constitute a thread to follow for ecologists, as in the  
 37 Theseus myth, and revive classic ecological questions  
 38 using a reliable mathematical framework.

## 39 Different calculations of Rao $\alpha$ , $\beta$ 40 and $\gamma$ entropy

41 With the Rao index, within-community di-  
 42 versity ( $\alpha$ ) can be defined as the extent of  
 43 dissimilarity between species in a community. In  
 44 particular, the Rao index for  $\alpha$ -diversity represents  
 45 the expected dissimilarity between two randomly  
 46 chosen individuals from a sampled community. If  $p_{ic}$   
 47 is the proportion of species  $i$  in community  $c$  (i.e. the  
 48 relative abundance of the  $i$ th species in the  $c$ th sam-  
 49 pling unit or site),  $s$  is the number of species (species  
 50 richness) in the community, and  $d_{ij}$  is the dissim-  
 51 ilarity (or "distance") between each pair of species

52  $i$  and  $j$ , the Rao  $\alpha$ -diversity can be defined as follows:

$$53 \alpha \text{ Rao} = \sum_{i=1}^s \sum_{j=1}^s d_{ij} p_{ic} p_{jc} \quad (1) \quad 54$$

55 where  $p_{ic}$  can be calculated as the number of in-  
 56 dividuals of a species ( $A_{ic}$ ) over the sum of  
 57 individuals of all species in the community, i.e.  
 58  $p_{ic} = A_{ic} / \sum_{i=1}^s A_{ic}$ , which means that  $p_{ic} = 1/s$  if all  
 59 species are equally abundant. More generally,  $\alpha$  Rao  
 60 is the sum of the dissimilarity between all possible  
 61 pairs of species, weighted by the product of species  
 62 proportions (which can be based on any measure of  
 63 species abundance, e.g. biomass, cover, etc.).

64 There are several possible ways to calculate  $d_{ij}$   
 65 depending on the type of data and facet of biodi-  
 66 versity considered (Ricotta 2005a; Lepš et al. 2006;  
 67 Hardy & Senterre 2007). For taxonomic diversity,  
 68  $d_{ij} = 1$  for every  $i \neq j$  and  $d_{ij} = 0$  otherwise (i.e. a unity  
 69 matrix with null diagonal),  $\alpha$  Rao equals the Simpson  
 70 diversity index (Pavoine et al. 2004; Botta-Dukát  
 71 2005; Ricotta 2005a). More generally,  $d_{ij}$  can be ex-  
 72 pressed using various measures of species  
 73 differences, e.g. functional or phylogenetic dissim-  
 74 ilarity between species (e.g. Shimatani 2001; Pavoine  
 75 et al. 2004). It is recommended to constrain  $d_{ij}$  to vary  
 76 from 0 (species  $i$  and  $j$  are identical) to 1 (as maximum  
 77 distance; Botta-Dukát 2005). This way, Simpson's in-  
 78 dex represents the potential maximum value that the  
 79 Rao index can reach if species are completely differ-  
 80 ent, and biological redundancy can be calculated as  
 81 the difference or ratio between Simpson's index and  
 82 the Rao index (de Bello et al. 2007).

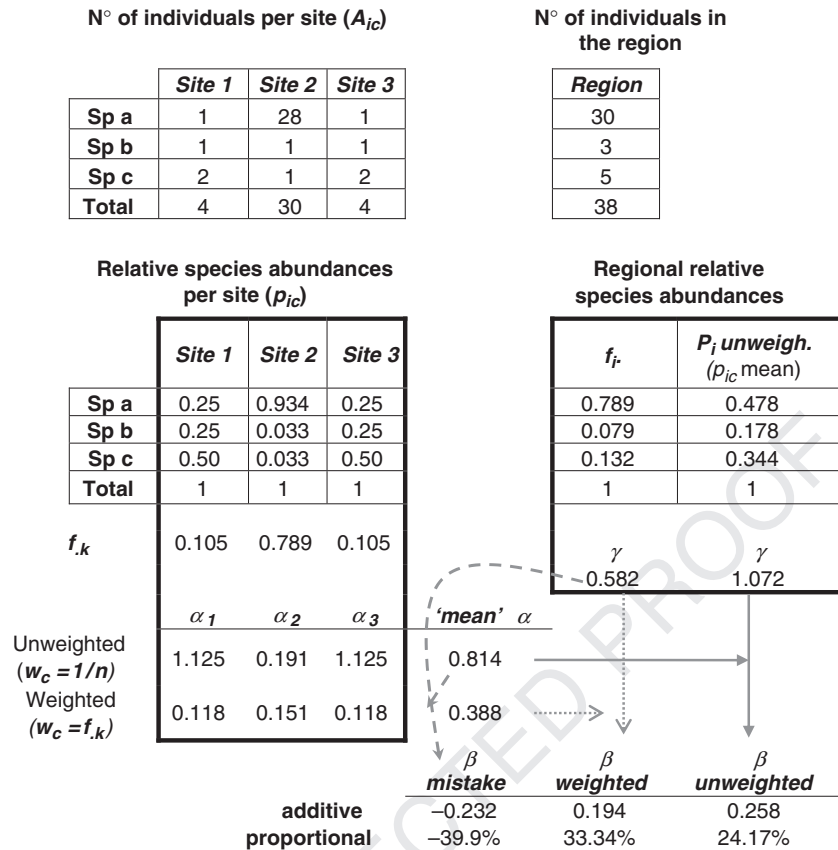
83 Application of the Rao index to calculate di-  
 84 versity at different spatial scales is bound to several  
 85 crucial choices. Overall, to calculate  $\gamma$  Rao it is neces-  
 86 sary to treat the study region as a single sampling unit  
 87 by pooling local communities together. Let  $S$  be the  
 88 total number of species in the region and  $P_i$  be the re-  
 89 gional species relative abundance for species  $I$ , then:

$$90 \gamma \text{ Rao} = \sum_{i=1}^S \sum_{j=1}^S d_{ij} P_i P_j \quad (2) \quad 91$$

92 The question is then what to consider as  $P_i$ . This  
 93 issue, we believe, resulted in a long-lasting confusion  
 94 in the literature. Ricotta (2005a), on the basis of  
 95 Rao's work (1982), originally stated that the re-  
 96 gional relative species abundances had to be  
 97 computed from the *average* (weighted or not, see  
 98 below) *of the local relative abundances of each species*  
 99 (Fig. 1). This equals:

$$100 P_i = \sum_{c=1}^n w_c p_{ic} \quad (3) \quad 101$$

55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108



**Fig. 1.** In the example, three hypothetical sampling units (or “Sites”) compose the species pool of a region (Region). From the local species abundances ( $A_{ic}$ ) the relative species abundance per site ( $p_{ic}$ ) can be calculated (as  $p_{ic} = A_{ic}/\sum A_{ic}$ ). The regional relative species abundances ( $P_i$ ) can be computed with two approaches. The first is using eqn 4, i.e. the average of  $p_{ic}$  (e.g. for species a:  $0.25/3+0.934/3+0.25/3$ ). This further implies calculating the mean regional  $\alpha$  as the average of all  $\alpha_c$  (i.e. “unweighted” with  $w_c = 1/n$  in eqn 5) and obtaining ever-positive  $\beta$  values (with  $\beta$  “unweighted” =  $\gamma -$  “unweighted” mean  $\alpha$ ). The second approach is using  $f_i$  of Vileger & Mouillot (2008), i.e. the number of individuals of a given species in the entire region divided by the total number of individuals of all species (e.g. for species a:  $30/38 = 0.789$ ). Using  $f_i$  implies a special use of  $w_c$  in both eqns 3 and 5, with the corresponding use of “weighted”  $\alpha_c$  values (otherwise negative  $\beta$ -diversity values, with  $\beta$  “mistake” =  $\gamma -$  “unweighted” mean  $\alpha$ , may be obtained). The weighted  $\alpha_c$  is obtained by multiplying  $\alpha_c$  by a factor  $f_{i,k}$ , which corresponds to the total number of individuals in a sampling unit divided by the total number of individuals in the region (e.g. for Site 1:  $4/38 = 0.105$ ). It should be noted that  $f_i$  can also be obtained from the sum of  $p_{ic}f_{i,k}$  (e.g. for the species a:  $0.25*0.105+0.934*0.789+0.25*0.105 = 0.789$ ). Note finally that, in the example, the original distances between species (either phylogenetic or functional) of Vileger and Mouillot are used ( $d_{ab} = 1, d_{bc} = 2, d_{ac} = 2$ ). Scaling distances between zero and one (which is recommended to allow comparisons with the Simpson species diversity) renders half of  $\alpha, \beta$  and  $\gamma$  values but the same proportion of  $\alpha$  and  $\beta$  over  $\gamma$ .

with  $n$  being the number of sites (1 to  $n$ ) in the region and  $w_c$  being a weighting parameter for the  $c$ th sampling unit or community. Most often  $w_c = 1/n$  (Pavoine et al. 2004; Ricotta 2005a), which corresponds to the “unweighted” form of  $P_i$ , simply calculated as the average of  $p_{ic}$ :

$$P_i \text{ unweighted} = \frac{\sum_{c=1}^n p_{ic}}{n} \quad (4)$$

Ricotta also showed that for whatever  $w_c$ ,  $\beta$ -diversity cannot be negative as long as  $\sum_{c=1}^n w_c = 1$  (see e.g. equation 3 in Ricotta 2005a) and  $\beta$  is expressed

as:

$$\beta = \gamma - \sum_{c=1}^n w_c \alpha_c \text{ with } \gamma \geq \sum_{c=1}^n w_c \alpha_c \quad (5)$$

where  $\sum w_c \alpha_c$  represent the weighted contribution of  $\alpha$ -diversities to the regional diversity  $\gamma$ . In this way, the same weight applied to calculate  $p_{ic}$  needs to be applied to weight  $\alpha_c$ .

However, certain ambiguity remains regarding how to calculate regional species relative abundances. An example can be found in Ricotta (2005b, Table 1, p. 369) where the used formula is not that

originally proposed by Ricotta (2005a). Along the same lines, Vileger & Mouillot (2008) showed an example of how to calculate regional species relative abundance as *the sum of the number of individuals of a given species in the whole region divided by the total number of individuals of all species*, i.e.

$$f_i = \frac{\sum_{c=1}^n A_{ic}}{\sum_{c=1}^n \sum_{i=1}^s A_{ic}} \text{ (see example in Fig. 1; note$$

that  $f_i$  of Vileger and Mouillot replaces  $P_i$  in equation 2). Here we show that  $f_i$  is only a special case of  $P_i$ , which occurs when  $w_c$  equals *the total number of individuals in a plot divided by the total number of individuals in a region* (called  $f_k$  by Vileger and Mouillot). In this case,  $f_i$  corresponds exactly to the sum of  $w_c P_{ic}$  (see Fig. 1). Note that  $f_i$  is equal to  $P_i$  *unweighted* only if the number of individuals is the same at all sites (e.g. Hardy & Senterre 2007, which in some sampling designs corresponds to having the same sampling effort in all sampling units).

When using  $f_i$  as a special case of  $P_i$ , one should not forget to apply the same value used for  $w_c$  to weight all  $\alpha$ -diversities (as in the original formula in equation 5). Otherwise  $\beta$  values can indeed be negative (Fig. 1). As a matter of fact, Vileger and Mouillot, to avoid negative  $\beta$  values when using  $f_i$ , further suggested to correct (as a general rule)  $\alpha$ -diversities by a factor  $f_k$ , which in fact corresponds exactly to  $w_c$  in Fig. 1 (i.e. if expressed, as mentioned above, as the total number of individuals in a plot divided by the total number of individuals in a region). In doing this, they did nothing more than applying the original formula of weighted  $P_i$  and  $\alpha_c$  (i.e. using the same value of  $w_c$  in both equations 3 and 5, rather than only in equation 3).

To summarize, the problem of negative  $\beta$  Rao values is due to a specific case of estimating  $P_i$  while not applying the same weighting to the mean  $\alpha$ -diversities of a region. Hence, the correction  $f_k$  proposed by Vileger and Mouillot for  $w_c$  should be applied only when  $f_i$  is used to estimate  $P_i$  and not as a general rule. Indeed, Hardy & Jost (2008) noticed that the cause of negative values obtained by Vileger & Mouillot (2008) probably resulted “from an inadequate mixing of parameter definitions and parameter estimators” and that the correction should be applied only in certain circumstances. We believe that a clear elucidation of the specific reasons of such inadequate use of the Rao index and mixing of parameters is needed in order to make the index a more practical and useful tool for ecologists.

In this sense, as convincingly discussed by Hardy & Jost (2008), the weighting parameter  $w_c$  in the calculation of the mean  $\alpha$  Rao should be used in the original context proposed by Pavoine et al.

(2004) and Ricotta (2005a), i.e. it should be most often equal to  $1/n$  (thus resulting in equation 4 for “unweighted”  $P_i$ ). Correcting for an uneven total number of individuals at sites is possible but should be performed only in very special contexts (Hardy & Jost 2008). As a guideline,  $w_c = f_k$  should be preferred if all sub-communities within a given habitat and community type have been exhaustively sampled, with the aim of estimating within-community heterogeneity and the overall community diversity. Most often, however, field sampling units are chosen as a representative selection within a given habitat and community type, with the main question being the extent of diversity among habitats ( $\beta$ ). Then  $w_c = 1/n$  should be used (i.e. “unweighted”  $P_i$ ). The development and application of other possible corrections for  $w_c$  is, in this sense, open to further research.

## Biased $\beta$ -diversity values and necessary transformations

Jost (2007) demonstrated for various species diversity indices, including the Simpson index of species diversity, that  $\beta$ -diversity approaches zero as  $\alpha$ -diversity becomes larger, even if the sampling units share no similar species. Overall, this means that the  $\beta$ -diversity will be low regardless of the actual species overlap and the change in diversity across sampling units (Jost 2007; de Bello et al. 2009). Therefore  $\beta$ -diversity estimated using Simpson’s formulation could lead to meaningless ecological results (Ricotta & Szeidl 2009; Jost et al. 2010). This was shown to also be the case for indices commonly used in population genetics (Jost 2008). This limitation of the Simpson index in partitioning the spatial components of taxonomic diversity can be resolved by applying the correction proposed by Jost (2007) derived from equivalent numbers:

$$\alpha_{\text{Eqv}} = \frac{1}{(1 - \alpha)}; \quad (6)$$

$$\gamma_{\text{Eqv}} = \frac{1}{(1 - \gamma)} \quad (7)$$

According to Jost (2007), the  $\beta$ -diversity in a region in terms of equivalent numbers can then be expressed as:

$$\beta_{\text{Eqv}} = \frac{\gamma_{\text{Eqv}}}{\alpha_{\text{Eqv}}} \quad (8)$$

Here  $\beta_{\text{Eqv}}$  represents the number of communities that have no diversity overlap (i.e. having no

species in common). Therefore, the lower limit of the index is 1 (all communities have the same composition) and the upper limit is the number of sampling units (if communities share no species). If we define a more intuitive  $\beta$ , called  $\beta_{\text{prop}}$ , which represents the proportion of diversity accounted for by the differentiation between communities (or sampling units) in a given region, we can rewrite equation 8 as follows:

$$\beta_{\text{Eqv}} = \frac{1}{(1 - \beta_{\text{prop}})} \quad (9)$$

By resolving equation 9 with equations 6–8, Jost (2007) shows that  $\beta_{\text{prop}}$  can be expressed as:

$$\beta_{\text{prop}} = \frac{(\gamma - \alpha)}{(1 - \alpha)} \quad (10)$$

Here we warn ecologists that, in this notation, the  $\alpha$ -diversity used to calculate  $\alpha_{\text{Eqv}}$  (equation 6),  $\beta_{\text{Eqv}}$  (equation 8) and  $\beta_{\text{prop}}$  (equation 10) should equate to the average  $\alpha$  in a region. This approach should be preferred to using the averages of  $\alpha_{\text{Eqv}}$  calculated from single sampling unit  $\alpha$ 's. The two approaches produce slightly different results (see Appendix S1, second section, for details) but, more importantly, equations 8 and 10 are only valid with the first approach (e.g.  $\alpha_{\text{Eqv}}$  calculated applying the Jost correction on the mean regional  $\alpha$ ).

The logic of the original correction of Jost (2006, 2007) is based on the concept of “equivalent communities”. If  $\alpha$  and  $\beta$  are to be independent of each other, so that one is not constrained in any way by the other, then this is the unique correct partitioning (Jost et al. 2010). This corresponds to calculating diversity for the case of  $s$  equally common species in a sampling unit (each species therefore with a proportion of  $1/s$ ), with a resulting  $\alpha$ -diversity expression that should equal the actual number of species in a community, i.e. species richness (Jost 2006). It should be noted that the correction proposed by Jost (2006, 2007) corresponds to the original formulation of the Simpson diversity, i.e. expressed as  $1/\text{dominance}$  instead of the other possible notation as  $1$  minus dominance (with dominance,  $D = \sum_{i=1}^s p_{ic}^2$ ; Magurran 2004). This notation of  $1/D$  also corresponds to the index of diversity  $N_2$  proposed by Hill (1973). Given the particular way to derive equivalent communities, the Jost-corrected Simpson diversity equals the number of species if all species have the same relative abundance in a sampling unit ( $p_i$ ) or in a region ( $P_i$ ; Fig. 2), which is intuitive and biologically interpretable (i.e. the maximum value of the re-

ciprocal of the Simpson index corresponds to the number of species in a community; Hill 1973).

Although Jost’s correction was originally proposed only for multiplicative partitioning of diversity (i.e.  $\beta = \gamma/\alpha$ ), we show here that the correction could be equivalently applied to additive partitioning (see below and Fig. 2). By resolving equation 10,  $\beta_{\text{prop}}$  can actually be expressed as a percentage of the diversity of a whole region:

$$\beta_{\text{prop}} = \frac{\gamma_{\text{Eqv}} - \alpha_{\text{Eqv}}}{\gamma_{\text{Eqv}}} \quad (11)$$

This last formula clearly shows that the difference between  $\gamma$  and  $\alpha$  in terms of equivalent numbers is a meaningful measure of  $\beta$ -diversity. If  $\alpha_{\text{Eqv}}$  expresses the average number of equivalent species at the sample scale, and  $\gamma_{\text{Eqv}}$  the number of equivalent species at the regional scale, their difference expresses how many equivalent species are found across sampling units. We call this the “ $\beta$ -equivalent-additive”, i.e.  $\beta_{\text{EqvAdd}}$ , with:

$$\beta_{\text{prop}} = \frac{\beta_{\text{EqvAdd}}}{\gamma_{\text{Eqv}}} \quad (12)$$

where

$$\beta_{\text{EqvAdd}} = \gamma_{\text{Eqv}} - \alpha_{\text{Eqv}} \quad (13)$$

With this extension of Jost’s correction,  $\beta$ -diversity can be expressed as a proportion of the total regional diversity, which can be very useful when comparing different facets of diversity together (e.g. taxonomic, functional, phylogenetic).

Indeed, equation 12 has an upper limit that depends on the number of sampling units (it equals  $1 - 1/n$  if the  $n$  samples are all completely distinct) and produce a  $\beta_{\text{prop}}$  value always lower than 1. For example, a maximum differentiation between two plots results into  $\beta_{\text{prop}} = 0.5$  (or, which is the same, 50%; Fig. 2 and Jost 2007). Thus, it should be used carefully to compare results from data sets with very different  $n$ . In this case, we propose to normalize this equation to the interval  $[0, 1]$ . For example:

$$\beta_{\text{Norm-prop}} = \frac{\beta_{\text{prop}}}{1 - 1/n} \quad (14)$$

It should be noted, however, that for sufficiently large data sets,  $1/n$  will tend to zero so that  $\beta_{\text{Norm-prop}}$  will not be significantly higher than  $\beta_{\text{prop}}$ .

As a next step, we show that these extensions of the Jost correction can be similarly applied for the Rao index, whenever it is used for functional or phylogenetic diversity. In fact, since the Rao

Relative species abundances

|      | Site1 | Site2 | Total<br>( $P_i$ ) | Case1 | Case2  |
|------|-------|-------|--------------------|-------|--------|
| Sp a | 0.25  | 0     | 0.125              | grass | grass  |
| Sp b | 0.25  | 0     | 0.125              | grass | grass  |
| Sp c | 0.25  | 0     | 0.125              | forb  | legume |
| Sp d | 0.25  | 0     | 0.125              | forb  | forb   |
| Sp e | 0     | 0.25  | 0.125              | tree  | tree   |
| Sp f | 0     | 0.25  | 0.125              | tree  | tree   |
| Sp g | 0     | 0.25  | 0.125              | fern  | fern   |
| Sp h | 0     | 0.25  | 0.125              | fern  | moss   |

|              | $\alpha_1$ | $\alpha_2$ | $\gamma$ | $\beta_{additive}$ | $\beta_{prop}(\%)$ |
|--------------|------------|------------|----------|--------------------|--------------------|
| N° species   | 4          | 4          | 8        | 4                  | 50                 |
| Simpson div. | 0.75       | 0.75       | 0.875    | 0.125              | 14.3               |
| Simp. - JOST | 4          | 4          | 8        | 4                  | 50                 |

Case 1: two couples of identical species within a site, complete dissimilarity between sites

|            | $\alpha_1$ | $\alpha_2$ | $\gamma$ | $\beta_{additive}$ | $\beta_{prop}(\%)$ |
|------------|------------|------------|----------|--------------------|--------------------|
| Rao        | 0.5        | 0.5        | 0.75     | 0.25               | 33.3               |
| Rao - JOST | 2          | 2          | 4        | 2                  | 50                 |

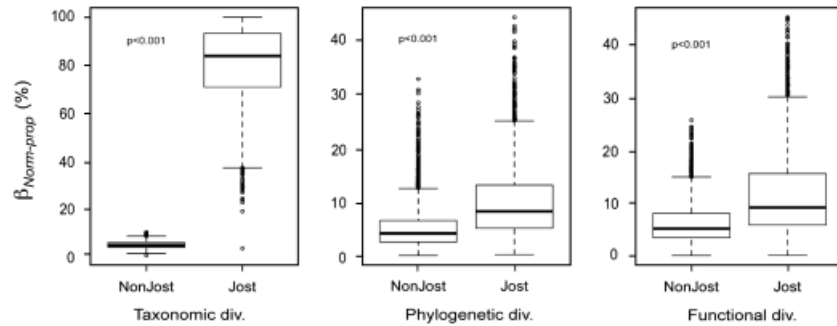
Case 2: one couple of identical species within a site, complete dissimilarity between sites

|            | $\alpha_1$ | $\alpha_2$ | $\gamma$ | $\beta_{additive}$ | $\beta_{prop}(\%)$ |
|------------|------------|------------|----------|--------------------|--------------------|
| Rao        | 0.625      | 0.625      | 0.813    | 0.19               | 23.1               |
| Rao - JOST | 2.67       | 2.67       | 5.33     | 2.67               | 50                 |

**Fig. 2.** Hypothetical example with two communities, with no species shared between them, having four species, each with the same relative abundance. Species are characterized by different functional forms in two hypothetical combinations (one where there are two couples of species that are identical within each site, and one where there is only one couple of identical species). For the example, we assume the distance ( $d$ ) between identical species to be equal to zero and, for different species, to be equal to 1 (in the case where  $d$  equals 1 for every couple of species, the Rao index equates to the Simpson index of diversity). Without the corrections proposed by Jost (2007), the Simpson index produces lower-than-expected  $\beta$  taxonomic diversity. The same can occur for uncorrected  $\beta$  Rao, which does not show a complete dissimilarity among sites (the first case corresponds to the fact that there are only two functionally or phylogenetically distinct species in a site; the second corresponds to three distinct species).  $\beta_{additive}$  expresses the differences between  $\gamma$  and mean  $\alpha$ ; for the Jost-corrected indices, this corresponds to equation 13 (i.e.  $\beta_{EqvAdd}$ ).  $\beta_{prop}$  expresses the proportion accounted by  $\beta_{additive}$  over  $\gamma$ ; for the Jost-corrected indices, this corresponds to equation 12 (here expressed in per cent).

decomposition of diversity is a generalization of the Simpson index (see above), we can imagine that  $\beta$ -diversity expressed by the Rao index does not behave as ecologists would expect (de Bello et al. 2009). This is shown in Fig. 2, where the  $\beta$  Rao index does not adequately depict a complete change in functional diversity across sites (in Case 1, with no functionally similar species shared by two “equivalent communities” with similar  $\alpha$ -diversity,  $\beta$ -diversity should equal 50% of  $\gamma$ -diversity). This underestimation of  $\beta$ -diversity could have devastating consequences on the interpretation of ecological patterns (Jost 2007; de Bello et al. 2009; Ricotta & Szeidl 2009).

Based on the concept of equivalent numbers, the example in Fig. 2 (and its extension in Appendix S1) shows also that the same correction proposed by Jost for the Simpson index can also be applied to the Rao index. This is not surprising since the Simpson index is only a specific case of the Rao index (Ricotta & Szeidl 2009). The  $\alpha$  and  $\gamma$  Jost-corrected indices should equate the number of species (i.e. species richness) in the case of equivalent communities. This is true in Case 1 in Fig. 2, with two equivalent communities sharing no functionally similar species, which shows that the Jost-corrected Rao index correctly returns the number of distinct species within local communities ( $\alpha_{Eqv}$ ), within



**Fig. 3.**  $\beta$ -diversity in the Guisane Valley, French Alps. The  $\beta$ -diversity was expressed as a normalized proportion of dissimilarity (i.e.  $\beta_{\text{Norm-prop}}$ , equation 14) between each pair of sampling units in the data set. The  $\beta$ -diversity was computed for taxonomic (Simpson diversity), phylogenetic (PD) and functional (FD) diversities with the Rao index, with and without the Jost-derived correction (i.e. Jost versus NonJost). The  $P$ -values within each panel indicate significantly higher  $\beta$ -diversity obtained when using the Jost-derived indices. Note that the panels have different scales.

an entire region ( $\gamma_{\text{Eqv}}$ ) and between communities ( $\beta_{\text{EqvAdd}}$ ). In addition, as convincingly demonstrated by Jost (2006), this correction does not produce negative  $\beta$  values. Since  $\gamma$  Rao is always greater than the average  $\alpha$  Rao (equation 5; Ricotta 2005a), the Jost-corrected  $\alpha$  Rao will always be lower than, or equal to, the corrected  $\gamma$  Rao. In this sense, it is important to emphasize that the distance matrix used to calculate this index must be based on ultrametric distances (i.e.  $d_{ij} \leq \max(d_{ik}, d_{kj})$ , Pavoine et al. 2004; Ricotta 2005a) and that  $d_{ij}$  should vary from 0 to 1. If  $d_{ij}$  is not scaled between 0 and 1, the results of the partitioning of diversity will be exactly the same, but the functional or phylogenetic diversity cannot be related, in absolute values, to taxonomic diversity.

## Case study

To show the importance of the Jost correction for the Rao index with real data, we compared diversity partitioning for taxonomic, functional and phylogenetic diversity obtained with and without the Jost-derived corrections. A case study from the Guisane Valley, in the French Alps, was used. The data set comprises 82 plant communities (of  $10\text{m} \times 10\text{m}$  in size with visual estimates of species composition and cover) sampled along multiple environmental gradients (altitude, soil characteristics, slope). For the 212 species in the data set, we used a phylogenetic supertree derived from the phylomatic web tool (Webb & Donoghue 2005). For functional diversity, we calculated species dissimilarity in terms of two plant traits: Specific Leaf Area (SLA), a quantitative trait (i.e. leaf area divided by the leaf dry weight), and Raunkiaer's classical life form, a

categorical trait. To calculate species dissimilarities with these two traits we applied the Gower distance, a standardized approach proposed by Botta-Dukát (2005) as appropriate for the computation of the Rao index.

The results of this case study show various important patterns illustrating the necessity of using the Jost correction for the partitioning of diversity. First, as shown by Jost (2007), the  $\beta$  taxonomic diversity is clearly strongly underestimated without the correction (4% TD-NonJost versus 80% for TD-Jost; Fig. 3). Second, without the Jost correction the  $\beta$  functional and phylogenetic diversity (FD- and PD-NonJost, respectively) can be higher than the taxonomic diversity (the values of FD- and PD-NonJost were sometimes higher than TD-NonJost; Fig. 3). This is a biological nonsense result, as  $\beta$  taxonomic diversity should represent the upper limit of functional and phylogenetic  $\beta$ -diversity (i.e. in the case that all species are different). This clearly indicates that comparing TD, FD and PD without the Jost correction can lead to biologically meaningless interpretations. Third, the FD- and PD-Jost indices showed higher  $\beta$ -diversity than the FD- and PD-NonJost indices ( $P < 0.001$ , Fig. 3), indicating that the Jost correction reduces the risks of a possible underestimation of  $\beta$ -diversity. It should be noted that, potentially, the Jost-corrected  $\beta$  Rao could be even lower than the "uncorrected"  $\beta$  (see Appendix S1). This is, however, a particular case that occurs, for example, between those sampling units where the FD- and PD-NonJost are higher than the TD-NonJost (see above).

At the same time, it should be noted that the Jost-corrected and non-corrected indices are, logically, strongly correlated (the Pearson correlations between Jost versus NonJost indices were 0.78 for

TD, 0.97 for PD and 0.96 for FD). Therefore, it is not surprising that using Jost versus NonJost indices does not markedly alter the relationship between  $\beta$ -diversity and environment (i.e. diversity turnover; Vellend 2001). For example, in a Mantel test between  $\beta$ -diversity (in pairs of communities) and environmental dissimilarity, we found a similar effect of environment on  $\beta$ -diversity with or without the Jost correction (for TD, FD and PD; not shown). Similarly, when using null models to compare the observed versus expected partitioning of diversity (see e.g. de Bello et al. 2009), we found no markedly different results comparing Jost versus NonJost indices (not shown). Therefore, the conclusions regarding possible community assembly mechanisms should remain rather unchanged with or without the Jost correction against random expectations.

Overall, these results highlight the idea that the Jost correction is especially important when comparing taxonomic diversity against functional and phylogenetic diversities. Such corrections may, therefore, prove to be fundamental if we intend to jointly use and compare biodiversity indices to understand mechanisms driving the assembly and functioning of natural communities across space and time (Ackerly & Cornwell 2007; Prinzing et al. 2008).

## Conclusions

The examples shown here (Figs 1 and 2) are intended to offer a simple guideline to ecologists aimed at comparing spatial partitioning of different facets of diversity with the same mathematical framework. When applying the Rao index framework, however, ecologists should be aware of different crucial choices that need to be made to correctly compute the index. These are synthesized here, as a guideline:

- Different views may lead to calculating  $P_i$ , relative abundance of a species in the study region, by modifying the parameter  $w_c$  within the general formula expressed in equation 3 (Fig. 1). In general,  $w_c$  should be equal to  $1/n$  (thus resulting in “unweighted”  $P_i$ ; equation 4), especially if sampling is not equally exhaustive in different habitats.
- Corrections for uneven numbers of individuals in sampling units (i.e. with  $w_c = f_{.k}$ , i.e. the total number of individuals in a plot divided by the total number of individuals in a region) should be

used only together with the corresponding weighting of average  $\alpha$ -diversities, i.e. in order to avoid potentially negative  $\beta$  values (equation 5; Fig. 1).

- Regardless of how  $P_i$  is calculated, the Rao index can produce systematic lower-than-expected  $\beta$ -diversity values (Fig. 2).
- This may be solved by applying simple corrections based on equivalent numbers, as proposed by Jost (2007) for the Simpson species diversity index. This expansion of the correction will certainly provide a helpful tool to compare spatial partitioning of taxonomic, phylogenetic and functional diversities.

The examples presented are further complemented by a function that applies these calculations within the statistical package R (Appendices S1 and S2). This new R function, called “Rao”, can be a useful tool for ecologists in partitioning diversity with the Rao index. The function further returns  $\alpha$ ,  $\beta$  and  $\gamma$ , both with and without the correction proposed here and by Jost (2006, 2007), for the Simpson and Rao indices of diversity.

**Acknowledgements.** We are grateful for the constructive inputs of Zoltán Botta-Dukát, Norman Mason and Valerio Pillar, which considerably improved this study. We also thank P. Choler for the initial discussions that stimulated this work and David Mouillot for providing insightful comments on an earlier draft of the manuscript. Marco Moretti and Guillaume Lentendu kindly helped to improve the clarity of the manuscript. Jan Jongepier improved the English language. The Station Alpine Joseph Fourier provided support during field sampling. FdB wishes to personally thank Lou Jost and Carlo Ricotta for their sincere and thoughtful suggestions on the final manuscript version. This research was funded through the IFB-ANR DIVERSITALP project (ANR 2008-2011, contract No. ANR 07 BDIV 014), the EU funded Eco-Change project (FP6 European Integrated project 2007–2011, contract No 066866 GOCE), the CNRS APIC RT PICs project 4876 and the project LC06073 (Czech Ministry of Education).

## References

- Ackerly, D.D. & Cornwell, W.K. 2007. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecology Letters* 10: 135–145.
- Botta-Dukát, Z. 2005. Rao’s quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science* 16: 533–540.



- 1 Crist, T.O. & Veech, J.A. 2006. Additive partitioning of  
2 rarefaction curves and species–area relationships:  
3 unifying alpha-, beta- and gamma-diversity with  
4 sample size and habitat area. *Ecology Letters* 9: 923–  
5 932.
- 6 de Bello, F., Lepš, J., Lavorel, S. & Moretti, M. 2007.  
7 Importance of species abundance for assessment of  
8 trait composition: an example based on pollinator  
9 communities. *Community Ecology* 8: 163–170.
- 10 de Bello, F., Thuiller, W., Lepš, J., Choler, P., Clement,  
11 J.C., Macek, P., Sebastia, M.T. & Lavorel, S. 2009.  
12 Partitioning of functional diversity reveals the scale  
13 and extent of trait convergence and divergence.  
14 *Journal of Vegetation Science* 20: 475–486.
- 15 Hardy, O.J. & Jost, L. 2008. Interpreting and estimating  
16 measures of community phylogenetic structuring.  
17 *Journal of Ecology* 96: 849–852.
- 18 Hardy, O.J. & Senterre, B. 2007. Characterizing the  
19 phylogenetic structure of communities by an additive  
20 partitioning of phylogenetic diversity. *Journal of  
21 Ecology* 95: 493–506.
- 22 Hill, M.O. 1973. Diversity and evenness – unifying  
23 notation and its consequences. *Ecology* 54: 427–432.
- 24 Jost, L. 2006. Entropy and diversity. *Oikos* 113: 363–375.
- 25 Jost, L. 2007. Partitioning diversity into independent  
26 alpha and beta components. *Ecology* 88: 2427–2439.
- 27 Jost, L. 2008. G(ST) and its relatives do not measure  
28 differentiation. *Molecular Ecology* 17: 4015–4026.
- 29 Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A. &  
30 Ricotta, C. 2010. Partitioning diversity for conserva-  
31 tion analyses. *Diversity and Distribution* 16: 65–76.
- 32 Koleff, P., Gaston, K.J. & Lennon, J.J. 2003. Measuring  
33 beta diversity for presence–absence data. *Journal of  
34 Animal Ecology* 72: 367–382.
- 35 Lande, R. 1996. Statistics and partitioning of species  
36 diversity, and similarity among multiple communities.  
37 *Oikos* 76: 5–13.
- 38 Legendre, P., Borcard, D. & Peres-Neto, P.R. 2005. Ana-  
39 lyzing beta diversity: partitioning the spatial variation  
40 of community composition data. *Ecological Mono-  
41 graphs* 75: 435–450.
- 42 Lepš, J., de Bello, F., Lavorel, S. & Berman, S. 2006.  
43 Quantifying and interpreting functional diversity of  
44 natural communities: practical considerations matter.  
45 *Preslia* 78: 481–501.
- 46 Magurran, A.E. 2004. *Measuring biological diversity*.  
47 Blackwell, Oxford, UK.
- 48 Pavoine, S., Ollier, S. & Pontier, D. 2004. Measuring,  
49 diversity from dissimilarities with Rao's quadratic  
50 entropy: are any dissimilarities suitable? *Theoretical  
51 Population Biology* 67: 231–239.
- 52 Prinzing, A., Reiffers, R., Braakhekke, W.G., Hennekens,  
53 S.M., Tackenberg, O., Ozinga, W.A., Schaminee,  
54 J.H.J. & van Groenendael, J.M. 2008. Less lineages –  
55 more trait variation: phylogenetically clustered plant  
56 communities are functionally more diverse. *Ecology  
57 Letters* 11: 809–819.
- 58 Rao, C.R. 1982. Diversity and dissimilarity coefficients – a  
59 unified approach. *Theoretical Population Biology* 21:  
60 24–43.
- 61 Ricotta, C. 2005a. A note on functional diversity measu-  
62 res. *Basic and Applied Ecology* 6: 479–486.
- 63 Ricotta, C. 2005b. Additive partitioning of Rao's quadra-  
64 tic diversity: a hierarchical approach. *Ecological  
65 Modeling* 183: 365–371.
- 66 Ricotta, C. & Szeidl, L. 2009. Diversity partitioning of  
67 Rao's quadratic entropy. *Theoretical Population Bio-  
68 logic* 76: 299–302.
- 69 Shimatani, K. 2001. On the measurement of species diver-  
70 sity incorporating species differences. *Oikos* 93:  
71 135–147.
- 72 Veech, J.A., Summerville, K.S., Crist, T.O. & Gering, J.C.  
73 2002. The additive partitioning of species diversity:  
74 recent revival of an old idea. *Oikos* 99: 3–9.
- 75 Vellend, M. 2001. Do commonly used indices of beta-  
76 diversity measure species turnover? *Journal of Vege-  
77 tation Science* 12: 545–552.
- 78 Villeger, S. & Moullot, D. 2008. Additive partitioning of  
79 diversity including species differences: a comment on  
80 Hardy and Senterre (2007). *Journal of Ecology* 96:  
81 845–848.
- 82 Webb, C.O. & Donoghue, M.J. 2005. Phylomatic: tree as-  
83 sembly for applied phylogenetics. *Molecular Ecology  
84 Notes* 5: 181.
- 85 Whittaker, R.J. 1975. *Communities and ecosystems*.  
86 Macmillan, New York, NY, US.

## Supporting Information

Additional supporting information may be found in the online version of this article:

### Appendix S1.

### Appendix S2.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 29 June 2009;

Accepted 4 April 2010.

Co-ordinating Editor: Dr. Valerio Pillar.

