A Journal of Conservation Biogeography

Diversity and Distributions

BIODIVERSITY RESEARCH

Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling

Pedro Cardoso[1,2,3]*, Paulo A. V. Borges[1] and Joseph A. Veech[4]

[1]*CITA-A (Azorean Biodiversity Group), Dep. Ciências Agrárias, Universidade dos Açores, Angra do Heroísmo, Portugal,* [2]*Natural History Museum of Denmark, Zoological Museum and Centre for Macroecology, University of Copenhagen, Copenhagen, Denmark,* [3]*Smithsonian Institution, National Museum of Natural History, Washington, DC, USA,* [4]*Department of Biology, Texas State University, San Marcos, TX, USA*

ABSTRACT

Aim Researchers measuring beta diversity have rarely concerned themselves with the problems of how complete the species lists of studied communities are, and of how the varying degrees of completeness can actually change estimates of beta diversity. No comprehensive assessment has been made regarding the behaviour of most beta diversity indices when applied to incomplete samples, a situation which is more common than usually recognized. Our objective was to assess the behaviour and robustness of a number of beta diversity measures for incidence data from undersampled communities.

Location Mainland Portugal and the Azorean archipelago (North Atlantic).

Methods Data from intensive sampling of spiders in mainland Portugal and arthropods in Azores were collected. We examined the properties of 15 beta diversity measures developed for incidence data. We simulated varying degrees of completeness, whereas computing beta diversity for selected pairs of samples. The robustness of these beta diversity accumulation curves was assessed for the purpose of finding the best measures for undersampled communities.

Results The Harrison *et al.* β_{-2} and the Williams β_{-3} are particularly robust to undersampling. These measures are also insensitive to differences of alpha diversity (species richness) between communities, and therefore to nestedness. Colwell & Coddington β_{cc} and the related Jaccard β_j and Gaston *et al.* β_g performed best of the measures sensitive to alpha diversity. They performed poorly, however, when compared communities exhibited very low values of beta diversity. In such cases, the Routledge β_r performed the best.

Main conclusions No index was found to perform without bias in all circumstances. Overall, β_{-2}, β_{-3} and β_{cc} (or related measures β_j and β_g) are recommended as they seem to be the most robust to undersampling.

Keywords
Accumulation curves, arthropods, beta diversity, completeness, sampling, similarity indices.

*Correspondence: Pedro Cardoso, Smithsonian Institution, National Museum of Natural History, PO Box 37012, MRC 105, Room E-509, Washington, DC 20013-7012, USA. E-mail: pcardoso@ennor.org

INTRODUCTION

Biological diversity has long been separated into different components according to the phenomena that interest ecologists (Whittaker, 1960, 1972). The components most often studied are: (i) alpha (local) diversity; (ii) gamma (regional) diversity, which can be considered as an equivalent to alpha diversity on a larger scale, but reflects the allopatric distribution of related taxa, and; (iii) beta diversity that measures turnover of species between communities, but for which there

is no universally accepted measure (Whittaker *et al.*, 2001; Koleff *et al.*, 2003).

Much attention has been given to the increase in estimates of alpha diversity resulting from increased sampling effort, especially when comparing alpha diversity between communities sampled with different levels of effort and completeness (i.e., the ratio of observed to actual richness) such that comparisons are biased and even misleading (see Soberón & Llorente, 1993; Gotelli & Colwell, 2001). Without complete inventories of communities, comparisons of species richness

cannot be reliably made if one does not consider the sampling effort or completeness attained. Many researchers address this concern by estimating alpha diversity using techniques that adjust for sampling effort or completeness, especially when dealing with hyperdiverse taxa, for which it is difficult to obtain complete inventories (Keating *et al.*, 1998; Longino *et al.*, 2002). However, researchers measuring beta diversity have rarely concerned themselves with problems relating to the potential effect that sampling effort and completeness within compared communities may have on estimates of beta diversity (e.g. Fisher, 1999; Chao *et al.*, 2005). Perhaps this is a consequence of the general lack of agreement about which index of beta diversity should be used in the first place (e.g. Koleff *et al.*, 2003).

Colwell & Coddington (1994) and Chao *et al.* (2005) revealed that, for different indices, beta diversity decreases with increasing sampling effort, resulting from the fact that with low effort, many species which are shared by two communities may not be sampled from one or both by pure chance, and differences between them are therefore artificially inflated (see also Fisher, 1999). This happens even when comparing samples that come from the same community, which should ideally exhibit no differences. On the other hand, the opposite problem is also possible. If the compared communities share a few very abundant species, and if there are many rare species that are exclusive to each community, estimated beta diversity may increase with increased sampling effort. In such cases, undersampling will tend to reveal only the shared, more abundant species and beta diversity will therefore be underestimated.

The problem of undersampling in measuring beta diversity can be addressed in one of the three following ways: (i) by ensuring sufficient sampling effort for the chosen beta diversity measure to reach a stable, and therefore reliable, value; (ii) by statistically correcting undersampling bias in indices (as in Chao *et al.*, 2000, 2005); or (iii) by explicitly choosing an index that is robust to relatively low levels of sampling effort.

Chao *et al.* (2000) have devised ways of implementing the second of these strategies by estimating the number of shared species between pairs of samples (or communities) based on the principles of the ACE estimator (Abundance-based Coverage Estimator, Chao & Lee, 1992). Chao *et al.* (2005) have also created estimators for the Jaccard and Sørensen similarity indices (complements of beta diversity indices for incidence data) for quantitative data. Their approach deserves consideration given that it recognizes, and attempts to statistically correct for undersampling. However, as with species richness estimators, the approach does not eliminate all sampling issues. A considerable amount of bias remains with underestimation still occurring, especially with respect to severe undersampling and for highly dissimilar samples. Additionally, the method requires that 'adjustment terms' be estimated to account for the shared undetected species (Chao *et al.*, 2005). These adjustment terms are estimated from abundance data (of recorded species) that may themselves be affected by undersampling or underestimation of the true abundances.

Our focus in this paper is on the third of the above strategies – i.e. choosing an index that is robust to relatively low levels of sampling effort. We examine the effects of undersampling on well-known indices measuring the similarity, or conversely dissimilarity (equivalent to beta diversity), between pairs of communities. Some new indices (e.g., Baselga *et al.*, 2007; Chao *et al.*, 2008) allow simultaneous measurement of similarity for three or more samples, but these new indices are also potentially affected by undersampling given that they are based on incidence or abundance data, as are the pairwise indices. These multi-sample indices are also new enough that they have not yet been extensively applied to empirical data.

Several authors (Wilson & Shmida, 1984; Gray, 2000 ; Koleff *et al.*, 2003; Magurran, 2004) have reviewed a number of pairwise indices based on incidence data and, in some cases, have assessed these for properties that are deemed desirable in an ideal beta diversity measure. But, to our knowledge, no comprehensive assessment has yet been made regarding the behaviour of such measures when applied to incomplete samples, a situation which is more common than usually recognized. While the effect of undersampling has been tested for a few indices – e.g. for complementarity (Fisher, 1999), Jaccard (Chao *et al.*, 2005) and Sørensen's (Plotkin & Muller-Landau, 2002; Chao *et al.*, 2005) – these effects have never been compared between indices, and most indices proposed to date have never been tested at all. Therefore, we have three objectives in this paper: (i) to demonstrate the utility of beta diversity accumulation curves in the assessment of the robustness to undersampling in pairwise beta diversity measures based on incidence data; (ii) to assess how different measures behave with regard to varying degrees of undersampling and; (iii) to recommend the measures that are relatively robust to undersampling.

METHODS

Indices assessed

Our evaluation focused on measures that assess beta diversity in terms of compositional dissimilarity between pairs of communities, rather than in terms of the overall beta diversity of larger sets of three or more communities. The vast majority of studies of pairwise beta diversity have focused on measures that employ incidence data, consisting of lists of species occurring in the compared communities (Koleff *et al.*, 2003). We therefore evaluated only measures that use such data and not indices that require abundance data. We also limited the study to those measures whose maximum values are finite. The values for these measures typically vary between 0, indicating that the two communities have identical species composition and 1, indicating that the communities share no species. Although one of the measures we evaluated did not originally present values in this range, it was modified to fulfil this condition (Williams β_{-3}; Table 1).

Many indices are basically equivalent, or they present values that are directly convertible to other indices by adding or

Table 1 Beta diversity measures compared in this study, re-expressed in terms of their pairwise matching/mismatching components (a = species shared by both communities; b, c = species exclusive to each of the two compared communities; notation following Koleff et al., 2003). Indices that are equivalent or that present similar properties for paired comparisons of communities are indicated. *The original formulation of Williams (1996) did not include the multiplication by 2, but we propose to use it to ensure that all indices change between 0, indicating when two communities have identical species composition, and 1, indicating when two communities do not share any species.

Formula	References	Indices with the same properties
$\beta_w = \frac{2(a+b+c)}{2a+b+c} - 1$	Whittaker (1960, 1972)	β_{-1} (Harrison et al., 1992) β_t (Wilson & Shmida, 1984) β_{me} (Mourelle & Ezcurra, 1997) β_{sor} (Sørensen, 1948) β_{hk} (Harte & Kinzig, 1997)
$\beta_r = \frac{(a+b+c)^2}{(a+b+c)^2 - 2bc} - 1$	Routledge (1977)	
$\beta_{-2} = \frac{\min(b,c)}{\max(b,c)+a}$	Harrison et al. (1992)	
$\beta_{co} = 1 - \frac{a(2a+b+c)}{2(a+b)(a+c)}$	Cody (1993)	
$\beta_{cc} = \frac{b+c}{a+b+c}$	Colwell & Coddington (1994)	β_g (Gaston et al., 2001) β_j (Jaccard, 1912)
$\beta_{-3} = 2\frac{\min(b,c)}{a+b+c}$	Williams (1996)*	
$\beta_{sim} = 1 - \frac{a}{\min(b,c)+a}$	Lennon et al. (2001)	
$\beta_z = 1 - \left[\frac{\log\left(\frac{2a+b+c}{a+b+c}\right)}{\log 2}\right]$	Lennon et al. (2001)	

multiplying fixed numbers (see Table 1). In such cases, we assessed only one of these equivalent indices in detail, although basic evaluation results are presented for all.

Datasets

To evaluate the behaviour of the different indices in relation to sampling effort, we used three different datasets: one, theoretical that simulated species' abundances of 11 fictitious communities and two, empirical datasets generated from the application of very comprehensive spider and arthropod sampling protocols (see details below).

Ten of the communities in the theoretical dataset each included 1,000,000 individuals and 1000 species, with the number of individuals for each species assigned according to a lognormal distribution. These were simulations of entire communities rather than of samples or inventories, as very few inventories sample this many individuals. However, arthropod communities typically contain such numbers. Also, typical communities follow a lognormal distribution of species abundance (May, 1975; Longino et al., 2002), although this is usually truncated by low sampling effort (Magurran, 2004; Coddington et al., 2009). Datasets were constructed in such a way that 10 species were represented by singletons (only one individual recorded). The first fictitious community was compared (pairwise) with the other nine communities. These nine communities shared 1%, 5%, 10%, 20%, 50%, 80%, 90%, 95% and 99% of species with the first community, thus representing varying degrees of sample similarity. The shared species were chosen randomly such that they represented species of varying abundances. To examine the performance of the pairwise indices on nested data, one of the fictitious communities with the 1,000,000/1000 lognormal distribution was compared with the 11th community that presented 500,000 individuals and 500 species also following a lognormal

distribution. All these 500 species were shared by both communities, in a fully nested structure.

One of the empirical datasets resulted from intensive fieldwork in mainland Portugal, which concentrated on spiders in three different habitat types, one in each of three protected areas: a mixed oak forest in the Peneda-Gerês National Park (hereafter simply referred to as Gerês), north-western Portugal (see Cardoso et al., 2008b); a cork oak forest in the Arrábida Nature Park (hereafter Arrábida), central Portugal (see Cardoso et al., 2008a); and a scrubland in the Vale do Guadiana Nature Park (hereafter Guadiana), south-eastern Portugal (see Cardoso et al., 2009). At each site, a delimited area of 1 ha was sampled with three different methods that covered all vertical layers (except the high canopy) with some degree of overlap. These methods were tree beating, sweeping and pitfall trapping. Every method had 64 samples, each equivalent to one-person-hour of work, which was set up as a semi-quantitative protocol (see Cardoso et al., 2008a,b, 2009; for details). In total, 336 species and 15,736 individuals were included in the analyses. Sampling completeness, calculated as the ratio between observed and estimated richness with the Chao1 estimator (Colwell, 2006), reached c. 90% at the first two sites, and c. 70% at the third site.

The second empirical dataset resulted from a large collecting effort of arthropods undertaken in the Azores archipelago, in the North Atlantic, in relatively large natural forest fragments on three different islands: Mistério da Prainha forest in the island of Pico; Serra de Santa Bárbara in Terceira; and Morro da Sé e Pico Alto in Flores. At each site, eight transects of 150×5 m were scattered in the forest to quantify arthropod diversity (see Borges et al., 2005, 2006; Ribeiro et al., 2005; Cardoso et al., 2007 for details). Each transect consisted of 30 pitfall traps and 30 beaten tree samples. Half of the pitfall traps contained Turquin solution and the other half, an ethylene glycol solution. In total, 233 species and 30,359 individuals

were collected and included in the analyses. The sampling completeness in each of the three forests was around 70%.

Statistical procedures

With the theoretical dataset, we first tested how the different measures change according to varying proportions of shared species. This provided confirmation of the sensitivity of measures to changes in composition, and therefore that any high robustness to undersampling observed in subsequent analyses would not be because of an inherent lack of sensitivity of any measure.

We evaluated the robustness of measures to incomplete samples using accumulation curves. Beta diversity accumulation curves can be constructed (in a similar way to species richness accumulation curves; see Fisher, 1999; Chao *et al.*, 2005) by calculating the mean beta diversity of paired samples simulating different amounts of undersampling. A computer program (Java code) was created to randomize the accumulation of individuals (without replacement), one at a time within each of the two samples that were compared (code available from the first author by request). This accumulation was performed at the same rate for both the samples. When datasets had a different maximum number of individuals, the accumulation process for the smaller dataset would reach the end before the larger dataset, which would then continue its accumulation process until all individuals were selected. Ten thousand curves were created and averaged for each pairwise comparison (see below), resulting in smooth accumulation curves for beta diversity values. Thus, the accumulation curves allowed us to quantify the beta diversity between the two datasets (series of paired samples) for the complete range of undersampling.

Three analyses were made with this approach. First, we assessed beta diversity indices (Table 1) with the theoretical datasets, resulting in 10 accumulation curves per measure. Secondly, we assessed how the indices behaved when applied to the two empirical datasets at the large scale. A total of six comparisons were made, with the three sites sampled for spiders in Portugal paired together (three pairwise comparisons) and then the same for the three sites sampled for arthropods in Azores (another three pairwise comparisons). Again, each pairwise comparison resulted in the generation of a beta diversity accumulation curve. Thirdly, we used the empirical datasets to assess the performance of indices at a more local scale by deriving these indices from results obtained using different sampling methods within a single site selected from each dataset – i.e. three methods compared for Gerês (three pairwise comparisons) and three methods compared for Pico (another three pairwise comparisons). These sites were chosen because they contained the richest fauna in mainland Portugal and the Azores, respectively. Although beta diversity is not commonly calculated for this type of scenario (i.e., beta diversity as produced by different sampling methods), each method was directed towards a different subset of organisms, with varying degrees of overlap between them. Moreover, these

comparisons enabled us to test the expectation that beta diversity would be higher between pitfall methods and vegetation displacement methods than among methods that sampled similar parts of the community.

To quantitatively evaluate the performance of the different indices for each pairwise comparison, we measured the robustness to undersampling of the entire curve, from the first to the next to last individual, using the following equation:

$$\text{Robustness} = 1 - \frac{\sum_{i=1}^{n-1} |\beta_i - \beta_n|}{n-1}$$

where n is the total number of individuals within each dataset; β_i is the average value (over all 10,000 curves) of the index when i individuals are included in the sample; β_n is the final value of the index with all individuals included in the samples. This index, derived from the Mean Absolute Error measure (Walther & Moore, 2005), reflects the average difference between the beta diversity values obtained with all possible degrees of undersampling of the two datasets, and the final value with all individuals of both datasets included (i.e., no undersampling). The robustness value varies between 0 when β_n is either 0 or 1 and all β_i values are 1 or 0, respectively, and 1 when all β_i values are equal to β_n.

Arthropod datasets very rarely reach the sampling completeness levels achieved for our empirical datasets (Cardoso, in press). Typical per-site sampling effort in ecology and conservation studies is as low as 1% of the effort we used (Cardoso, in press). We have therefore plotted the accumulation curves on a log10 scale and have calculated robustness for three of these curves, corresponding to sampling of 1%, 10% and 100% of individuals. The log10 scale also allowed for a clearer visual depiction of the robustness of the beta diversity measures when sampling was 10% or less.

RESULTS

All measures were found to be similarly sensitive to changes in compositional overlap between communities (Fig. 1). The values of most measures decrease linearly as the percentage of shared species is increased (holding richness equal between compared communities). This suggests that any high values of robustness to undersampling obtained in the results presented below are not because of an inherent lack of sensitivity for any of the evaluated measures.

The accumulation curves of beta diversity derived using the theoretical datasets (Fig. 2) revealed that the values of all measures change with varying sampling completeness. However, the nature of this change differs between measures. The β_{-2} and β_{-3} rapidly reach an asymptote when communities differ markedly in composition (Fig. 2c, d). On the contrary, β_r performs very badly under this scenario. This latter measure, however, is the best performer when communities are very similar, reaching the asymptote more rapidly than any other index (Fig. 2g–i). Four measures (β_r, β_{-2}, β_{-3}, and β_{sim}) are insensitive to differences in alpha diversity (species richness) and nestedness, although in this case, they reach the true value
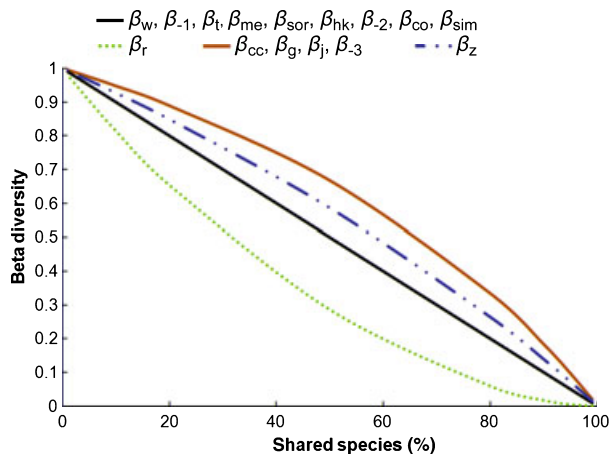
Figure 1 Value of the different pairwise beta diversity measures (see Table 1) according to the proportion of shared species, assuming similar species richness in both compared samples.

of 0 only with fully sampled datasets (Fig. 2j). This is because, although the two communities are actually completely nested, the accumulation process (representing undersampling) often produced samples that were not completely nested.

Both large and local-scale evaluations of the empirical datasets revealed relatively high robustness to undersampling for all beta diversity measures when at least 10% of the dataset was sampled (see Appendix S1). That is, many of the measures rapidly converged to the true beta diversity value as sampling effort increased (Figs 3 & 4). Two indices were found to be particularly robust to undersampling, β_{-2} (robustness at 1% = 0.855 averaged over all 12 large and local-scale pairwise comparisons; at 10% = 0.924; at 100% = 0.977) and β_{-3} (average robustness at 1% = 0.860; at 10% = 0.918; at 100% = 0.975). Their accumulation curves were usually very stable right from the beginning of the process, with a few exceptions (Figs 3 & 4). A relatively flat curve indicated a highly robust index, regardless of the beta diversity value. This was not the case for these two specific indices when compared communities exhibited low final beta diversity values (e.g. Fig. 3, Pico vs. Terceira or Fig. 4, beating vs. sweeping). In such cases, their behaviour was unreliable, with undersampling causing either overestimation or underestimation of beta diversity. The β_r index appeared to be the most robust measure when datasets had low dissimilarity (Fig. 2; see Appendix S1). However, β_r presented a mediocre performance when datasets had medium to high dissimilarity (average robustness at 1% = 0.781; at 10% = 0.884; at 100% = 0.960). The worst performers overall were β_w and related measures (average robustness at 1% = 0.747; at 10% = 0.868; at 100% = 0.960) and especially β_{sim} (average robustness at 1% = 0.739; at 10% = 0.859; at 100% = 0.957). In general, the robustness of the different measures, according to the large and local-scale empirical evaluations, may be ranked as: $\beta_{-2} \approx \beta_{-3} > \beta_{cc} = \beta_j = \beta_g > \beta_z > \beta_r > \beta_{co} > \beta_w = \beta_{-1} = \beta_t = \beta_{me} = \beta_{sor} = \beta_{hk} > \beta_{sim}$.

DISCUSSION

This study sought to explore the impact of undersampling on the performance of pairwise measures of beta diversity, or community dissimilarity. Although the use of pairwise measures is known to have its limitations when comparing a large number of communities simultaneously (Chao et al., 2008), many studies still use such measures. Using (dis)similarity or related coefficients as measures of beta diversity is a common practice (Johannsson & Minns, 1987; Philippi et al., 1998; Anderson et al., 2006; Ferrier et al., 2007; Ricotta & Marignani, 2007). As previously mentioned, the Whittaker index when calculated for pairwise comparisons is similar to the Sørensen coefficient of similarity. Likewise, the complementarity index is similar to the Jaccard coefficient. But not all similarity measures have been 'disguised' as beta diversity indices. Studies of distance decay of similarity mostly use the Sørensen incidence-based index, equivalent to β_w (Nekola & White, 1999; Morlon et al., 2008; Cardoso et al., in press). This continued and widespread usage of pairwise indices highlight the importance of studying the effects of undersampling on these beta diversity measures.

Previous studies examined the effect of alpha diversity (species richness) and sample size (number of individuals) on pairwise beta diversity indices (Ricklefs & Lau, 1980; Wolda, 1981). These studies were aimed at estimating the bias in each index where bias was defined as the difference between the theoretical maximum for the index and the observed value under different levels of species richness and sample size.

As several authors have previously pointed out for different datasets and measures (Fisher, 1999; Plotkin & Muller-Landau, 2002; Chao et al., 2005) theoretical or empirical, beta diversity typically decreases with increasing sampling effort. The decrease in beta diversity with an increasing number of sampled individuals is usually constant and diversity values often do not asymptote (Figs 3 & 4). The most commonly used indices of beta diversity for incidence data are the original formulation of Whittaker (1960, 1972) and the Wilson & Shmida (1984), which is similar to the first index with respect to pairwise comparisons (Table 1). Both are particularly prone to overestimation when in the presence of undersampling. In the case of comparisons of species lists, which are known to be incomplete or with an unknown level of completeness, these indices should be applied cautiously.

In this study, we have used beta diversity accumulation curves to examine robustness of the most widely-used indices to known levels of undersampling. These tests indicate that there are a few indices that are less sensitive to undersampling relative to other measures. Both β_{-2} and β_{-3} usually present very stable curves, retaining similar values throughout the simulated sampling process, which is confirmed by the high robustness that both show in most of the pairwise comparisons. These two indices are precisely the ones that incorporate the numerator min (b,c). The fact that only the smaller of the mismatching components used prevents the index from
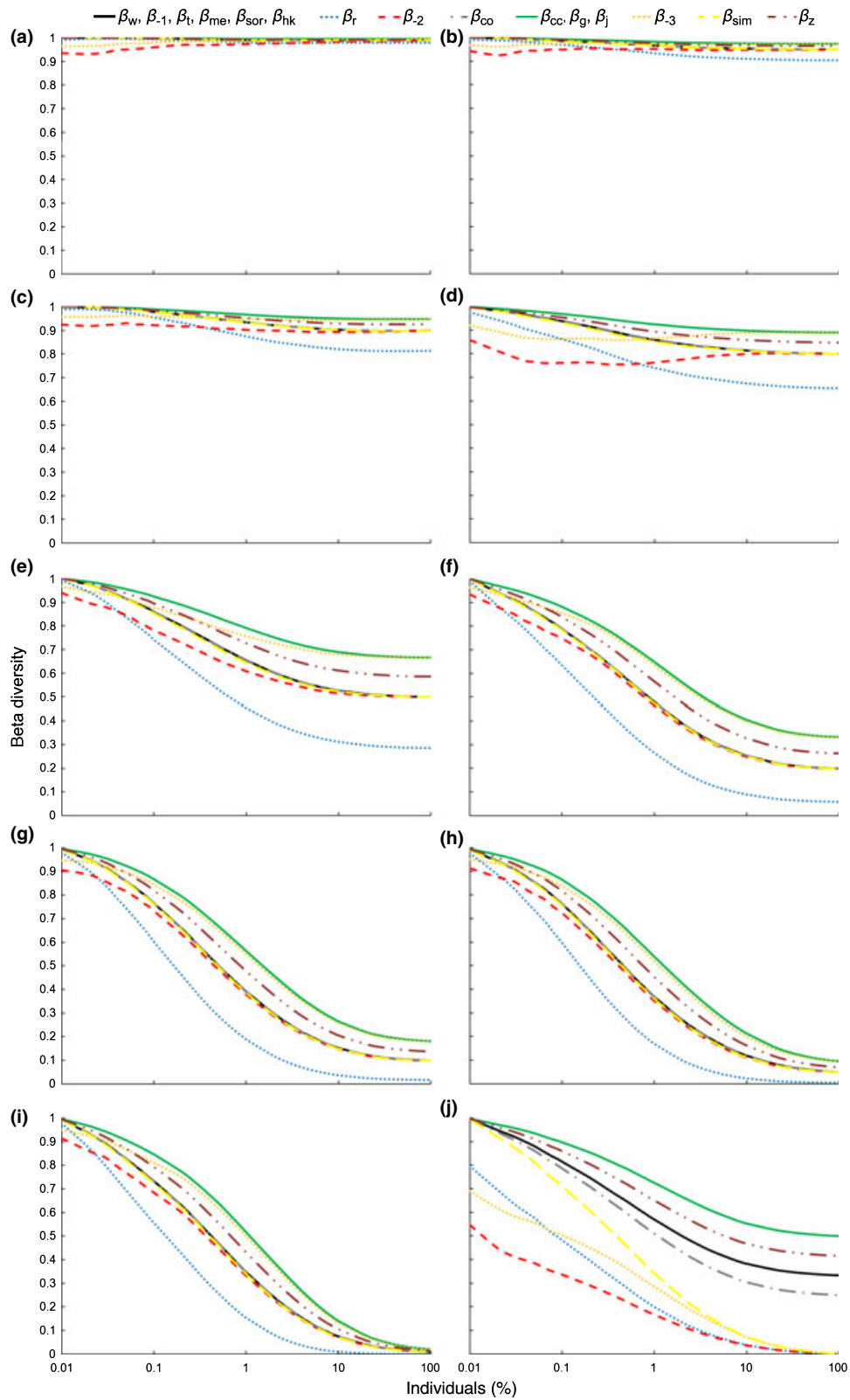
Figure 2 Beta diversity accumulation curves for the 15 measures evaluated when comparing theoretical communities with 1,000,000 individuals and 1000 species with a lognormal distribution. A single community was compared with communities sharing: (a) 1%; (b) 5%; (c) 10%; (d) 20%; (e) 50%; (f) 80%; (g) 90%; (h) 95%; (i) 99% of its species. The same community was compared with another presenting (j) 500,000 individuals and 500 species, all shared by both communities, in a fully nested design, with no species exclusive to the latter. The ordinate refers to the mean value of the given beta diversity index over 10,000 replicate simulations.
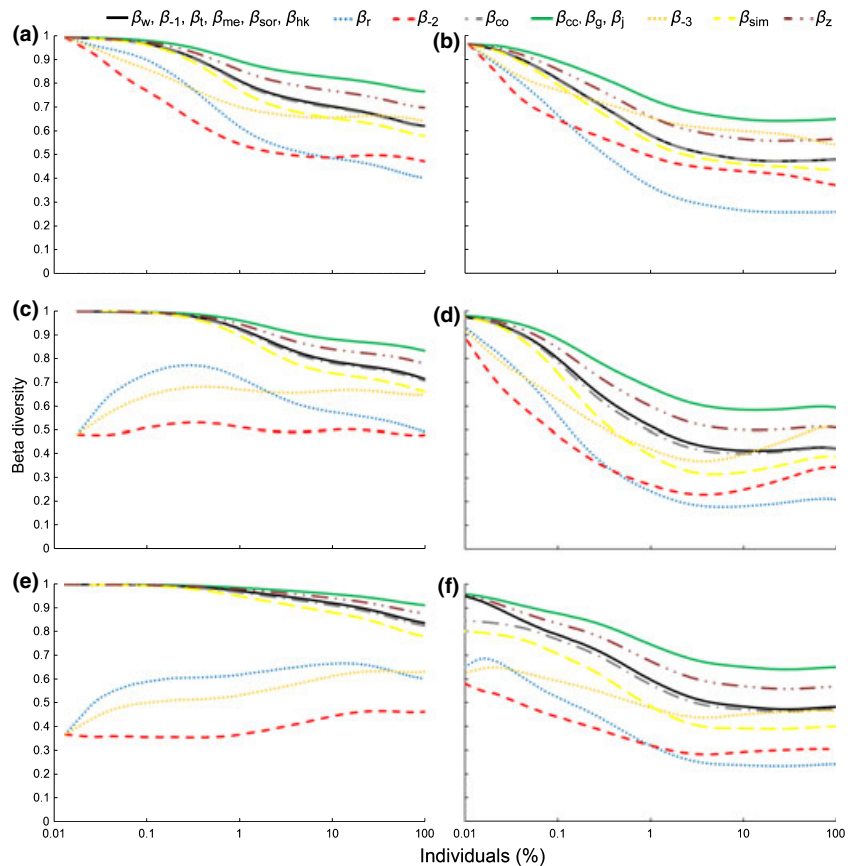
Figure 3 Beta diversity accumulation curves for the sampling at a large scale. Three sites sampled for spiders in mainland Portugal are compared: (a) Gerês × Arrábida; (c) Arrábida × Guadiana; (e) Gerês × Guadiana. As are three sites sampled for arthropods in the Azores: (b) Flores × Terceira; (d) Terceira × Pico; (f) Flores × Pico. The ordinate refers to the mean value of the given beta diversity index over 10,000 replicate simulations.

overestimating beta when there is undersampling, which occurs most of the time with the other indices that incorporate both components in their formulae. It may seem, on a first glance, that the β_{-2} and β_{-3} have performed well only because they are more insensitive to compositional differences than are other indices. However, the results presented in Fig.1 suggest that this is not the case.

The measures β_{-2} and β_{-3} are however, together with β_r and β_{sim}, insensitive to species richness differences between samples and consequently to nestedness (given that complete nestedness reflects only a difference in species richness, not species turnover). Complete nestedness (see Patterson & Atmar, 1986) occurs when any community has all the species that are present in all the less rich communities of the same region. Also, every species of that community will be present at all the communities in the region that are richer than itself. Nestedness patterns are usually because of impoverishment caused by selective extinction, for example, during fragmentation of large habitats, or as a result of enrichment driven by selective colonization (for example, in the Azores see Cardoso *et al.*, in press). Many authors do not perceive nestedness as beta diversity and the insensitivity of these indices to nestedness is therefore probably desirable in many studies involving community comparisons (e.g. Baselga *et al.*, 2007). In fact, this insensitivity is analogous to the requirement of independence between alpha and beta diversities (Wilson & Shmida, 1984; Jost, 2007; Ricotta, 2008).

On the other hand, many authors may not agree with this requirement of independence between diversity components. When comparing very different communities that differ both in richness and composition (e.g. polar and tropical biotas for most taxa), β_{-2}, β_{-3}, β_r, and β_{sim} will be limited by an upper bound lower than 1. This can cause low values of beta diversity even between communities with no shared species. The third best measures in terms of robustness, β_{cc}, β_j, and β_g, are sensitive to such differences. These may therefore be preferred in cases where differences in species richness between samples (or communities) need to be reflected in the measurement of beta diversity.

The measures β_{-2} and β_{-3} can be particularly prone to underestimation when very similar communities are undersampled. In this case, an alternative measure could be β_r, which exhibits the best behaviour for very similar communities. However, as shown by the lack of an asymptote for many indices in Figs 3 & 4 (see also Appendix S1), underestimation seems to be less common and less severe than overestimation. This bias was also reported by Ricklefs & Lau (1980) for some of these indices.

CONCLUSIONS

As a general recommendation, the most desirable measures of beta diversity based on our evaluation are those for which values obtained from varying levels of undersampling are as
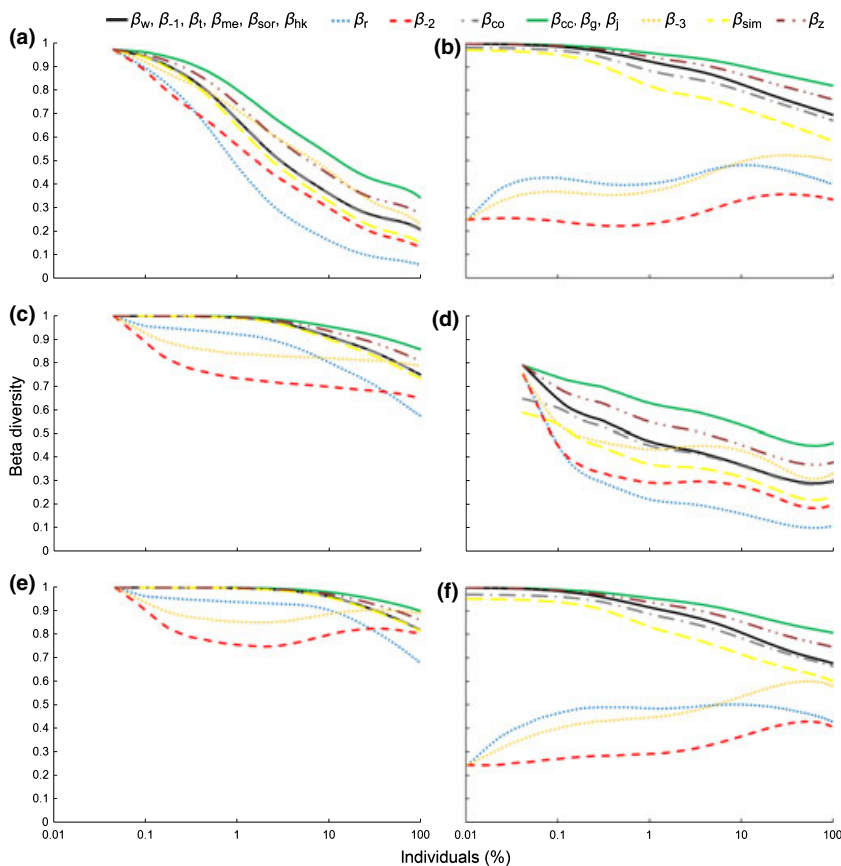
Figure 4 Beta diversity accumulation curves for the sampling at a local scale. The three methods employed at each of two studied sites are compared. For spiders in Gerês: (a) beating × sweeping; (c) sweeping × pitfall traps; (e) beating × pitfall traps. For arthropods in Pico: (b) beating × ethylene glycol traps; (d) ethylene glycol traps × Turquin traps; (f) beating × Turquin traps. The ordinate refers to the mean value of the given beta diversity index over 10,000 replicate simulations.

close as possible to the true value obtained from complete sampling of both communities. If the degree of sampling completeness of datasets is unknown or known to be low, we recommend the β_{-2} index of Harrison *et al.* (1992) as the most robust measure or, as an alternative, the β_{-3} of Williams (1996). Both of these indices were also amongst those recommended by Koleff *et al.* (2003) in their assessment of a number of other properties. If however, alpha diversity differences or nestedness need to be reflected in the measurement of beta diversity, then the best alternatives are β_{cc}, β_j, and β_g.

REFERENCES

Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**, 683–693.

Baselga, A., Jiménez-Valverde, A. & Niccolini, G. (2007) A multiple site similarity measure independent of richness. *Biology Letters*, **3**, 642–645.

Borges, P.A.V., Aguiar, C., Amaral, J., *et al.* (2005) Ranking protected areas in the Azores using standardized sampling of soil epigean arthropods. *Biodiversity and Conservation*, **14**, 2029–2060.

Borges, P.A.V., Lobo, J.M., Azevedo, E.B., Gaspar, C., Melo, C. & Nunes, L.V. (2006) Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species. *Journal of Biogeography*, **33**, 169–187.

Cardoso, P. (in press) Standardization and optimization of arthropod inventories – the case of Iberian spiders. *Biodiversity and Conservation*. doi: 10.1007/s10531-009-9690-7.

Cardoso, P., Borges, P.A.V. & Gaspar, C. (2007) Biotic integrity of the arthropod communities in the natural forests of Azores. *Biodiversity and Conservation*, **16**, 2883–2901.

Cardoso, P., Gaspar, C., Pereira, L.C., Silva, I., Henriques, S.S., Silva, R.R. & Sousa, P. (2008a) Assessing spider species richness and composition in Mediterranean cork oak forests. *Acta Oecologica*, **33**, 114–127.

Cardoso, P., Scharff, N., Gaspar, C., Henriques, S.S., Carvalho, R., Castro, P.H., Schmidt, J.B., Silva, I., Szüts, T., Castro, A. & Crespo, L.C. (2008b) Rapid biodiversity assessment of spiders (Araneae) using semi-quantitative sampling: a case study in a Mediterranean forest. *Insect Conservation and Diversity*, **1**, 71–84.

Cardoso, P., Henriques, S.S., Gaspar, C., Crespo, L.C., Carvalho, R., Schmidt, J.B., Sousa, P. & Szüts, T. (2009) Species richness and composition assessment of spiders in a Mediterranean scrubland. *Journal of Insect Conservation*, **13**, 45–55.

Cardoso, P., Gaspar, C., Dinis, F. & Borges, P.A.V. (in press) Patterns of alpha and beta diversity of arthropods at contrasting land-uses of an oceanic island. *Arthropods of Macaronesia* (ed. by A.R.M. Serrano, P.A.V. Borges, M. Boieiro and P. Oromí). Sociedade Portuguesa de Entomologia, Portugal.

Chao, A. & Lee, S.-M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Chao, A., Hwang, W.-H., Chen, Y.-C. & Kuo, C.-Y. (2000) Estimating the number of shared species in two communities. *Statistica Sinica*, **10**, 227–246.

Chao, A., Chazdon, R.L., Colwell, R.K. & Tsung-Jen, S. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, **8**, 148–159.

Chao, A., Jost, L., Chiang, S.C., Jiang, Y.-H. & Chazdon, R.L. (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, **64**, 1178–1186.

Coddington, J.A., Agnarsson, I., Miller., J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology*, **78**, 573–584.

Cody, M.L. (1993) Bird diversity components within and between habitats in Australia. *Species diversity in ecological communities: historical and geographical perspectives* (ed. by R.E. Ricklefs and D. Schluter), pp. 147–158, University of Chicago Press, Chicago, IL.

Colwell, R.K. (2006) EstimateS: statistical estimation of species richness and shared species from samples, Version 8.0. Available at http://viceroy.eeb.uconn.edu/estimates. Persistent URL http://purl.oclc.org/estimates (last accessed 1 January 2009).

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society, London B*, **345**, 101–118.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264.

Fisher, B.L. (1999) Improving inventory efficiency: a case study of leaf-litter ant diversity in Madagascar. *Ecological Applications*, **9**, 714–731.

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

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Gray, J.S. (2000) The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. *Journal of Experimental Marine Biology and Ecology*, **250**, 23–49.

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

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

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

Johannsson, O.E. & Minns, C.K. (1987) Examination of association indices and formulation of a composite seasonal dissimilarity index. *Hydrobiologia*, **150**, 109–121.

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

Keating, K.A., Quinn, J.F., Ivie, M.A. & Ivie, L.L. (1998) Estimating the effectiveness of further sampling in species inventories. *Ecological Applications*, **8**, 1239–1249.

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

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

Longino, J., Coddington, J.A. & Colwell, R.K. (2002) The ant fauna of a tropical rainforest: estimating species richness three different ways. *Ecology*, **83**, 689–702.

Magurran, A.E. (2004) *Measuring biological diversity*. Blackwell Science, Oxford, UK.

May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 81–120, Harvard University Press, Cambridge, MA.

Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R. & Green, J.L. (2008) A general framework for the distance-decay of similarity in ecological communities. *Ecology Letters*, **11**, 904–917.

Mourelle, C. & Ezcurra, E. (1997) Differentiation diversity of *Argentine cacti* and its relationship to environmental factors. *Journal of Vegetation Science*, **8**, 547–558.

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

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

Philippi, T.E., Dixon, P.M. & Taylor, B.E. (1998) Detecting trends in species composition. *Ecological Applications*, **8**, 300–308.

Plotkin, J.B. & Muller-Landau, H.C. (2002) Sampling the species composition of a landscape. *Ecology*, **83**, 3344–3356.

Ribeiro, S.P., Borges, P.A.V., Gaspar, C., Melo, C., Serrano, A.R.M., Amaral, J., Aguiar, C., André, G. & Quartau, J.A. (2005) Canopy insect herbivore diversity and distribution

in the native forests of the Azores: key host plant species in a highly generalist insect community. *Ecography*, **28**, 315–330.

Ricklefs, R.E. & Lau, M. (1980) Bias and dispersion of overlap indices: results of some Monte Carlo simulations. *Ecology*, **61**, 1019–1024.

Ricotta, C. (2008) Computing additive β-diversity from presence and absence scores: a critique and alternative parameters. *Theoretical Population Biology*, **73**, 244–249.

Ricotta, C. & Marignani, M. (2007) Computing β-diversity with Rao's quadratic entropy: a change of perspective. *Diversity and Distributions*, **13**, 237–241.

Routledge, R.D. (1977) On Whittaker's components of diversity. *Ecology*, **58**, 1120–1127.

Soberón, J. & Llorente, J. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7**, 480–488.

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 Selskabs Biologiske Skrifter*, **5**, 1–34.

Walther, B.A. & Moore, J.L. (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**, 815–829.

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

Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213–251.

Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.

Williams, P.H. (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proceedings of the Royal Society, London B*, **263**, 579–588.

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

Wolda, H. (1981) Similarity indices, sample size and diversity. *Oecologia*, **50**, 296–302.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Robustness values with 1%, 10% and 100% of individuals of the different beta diversity indices for all paired comparisons between communities.

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.

BIOSKETCHES

Pedro Cardoso is currently a postdoc at the Smithsonian Institution and the University of Azores. His research interests comprise biodiversity assessment and monitoring, prioritization of areas and species for conservation, ecological modeling, island biogeography, subterranean biology and spider taxonomy.

Paulo A. V. Borges is a professor in the Department of Agriculture of the University of the Azores and leads the Azorean Biodiversity Group of CITA-A. His ongoing research focuses on the study of the biodiversity of Azorean arthropods, with emphasis on taxonomy, ecology (macroecology), biospeleology and biogeography. He is also interested in the development of species-area models that could describe patterns of species richness and potential extinction on islands.

Joseph Veech is interested in landscape-level patterns of species diversity and the ecological factors that affect the patterns. He has also developed various statistical methods for analysing species diversity.

Editor: Hugh MacIsaac