

## treeNODF: nestedness to phylogenetic, functional and other tree-based diversity metrics

Adriano S. Melo\*, Marcus V. Cianciaruso and Mário Almeida-Neto

Departamento de Ecologia, Universidade Federal de Goiás, Caixa Postal 131, Goiânia, GO 74001-970, Brazil

### Summary

1. Available measures of nestedness consider descriptor variables (e.g. species) as equally associated, ignoring evolutionary or ecological dissimilarities. Here, we introduce treeNODF, a new class of nestedness that takes into account the resemblance of descriptor variables.
2. The method is an extension of the NODF index and can be applied to systems in which the resemblance of descriptor variables is described by a tree-like object. Computation of treeNODF is similar to NODF, but uses branch lengths instead of the sum of species occurrences. In this way, we can calculate a phyloNODF for metacommunities if a phylogeny is used to account for differences in phylogenetic diversity (PD) and traitNODF if a functional dendrogram constructed from species ecological traits is used to account for the functional diversity (FD) of communities. Similar to NODF, treeNODF can also be used to assess nestedness among species. In this case, envNODF uses a dendrogram describing the resemblance among the environmental conditions of different sites, to test whether rare species occur in a subset of the environmental conditions in the habitats occupied by frequent species.
3. treeNODF is a composite metric that can be additively partitioned into compositional (S.fraction) and tree-topology (topoNODF = treeNODF – S.fraction) components of the descriptor variables. Tests of treeNODF and its components can be carried out using null models and, if a hypothetical factor is used to order metacommunity data, permutation tests. We show that treeNODF is robust for matrix size and fill, as well as for tree topology.
4. Finally, we illustrate the use of treeNODF by analysing data on Caribbean bats using phyloNODF, traitNODF and envNODF, as well as their composition and tree-topology components.

**Key-words:** bats, community phylogenetics, ecophylogenetics, functional diversity, metacommunities, nested overlap based on decreasing fill, NODF, phylogenetic diversity

### Introduction

A nested pattern in species distribution is the tendency of the less ubiquitous species to occur solely in richer assemblages, whereas the common species occur in most assemblages. A nested pattern is thought to emerge as a result of coupled gradients in species traits and site characteristics (Ulrich, Almeida-Neto & Gotelli 2009; Ulrich & Almeida-Neto 2012), such as dispersal ability and isolation (Darlington 1957), and niche width and habitat heterogeneity (Wright & Reeves 1992).

Many community-wide metrics of nestedness are now available (see Ulrich, Almeida-Neto & Gotelli 2009 and references therein). For presence–absence data, the NODF index (nestedness based on overlap and decreasing Fill; Almeida-Neto *et al.* 2008) has gained popularity due to its straightforward interpretation and good statistical properties (independence on matrix size, shape and fill; low type I error rates; consistencies with the very definition of nestedness) (Almeida-Neto *et al.* 2008; Ulrich, Almeida-Neto & Gotelli 2009).

Interpretations of nested patterns in metacommunities are based on ecological and historical (i.e. biogeographical and evolutionary) processes. However, previous studies on nestedness and metacommunity structure have considered species as equivalent and independent units in ecological and evolutionary terms. Thus, traditional nested metrics ignore that, for example, species that share most of their functional traits should be mostly redundant for a particular ecological process or show a higher niche overlap. To solve this, we can use similarities in species traits to understand both how species affect ecosystem functioning and the degree to which they must compete with each other for resources. Similarly, communities composed of phylogenetically related species should, on average, show a lower potential for adaptation (see Forest *et al.* 2007) to environmental change than other communities that harbour the same species richness, but of less closely related species. This claim is supported by the relationship between phylogenetic diversity (PD) and features diversity (Faith 1992). Accordingly, PD-rich communities should harbour a high number of features that may allow most of its species to remain in place in face of environmental change. Furthermore, evidence of the decay in phylogenetic similarity with

\*Correspondence authors. E-mail: asm.adrimelo@gmail.com

geographical or environmental distance (Bryant *et al.* 2008) has been investigated and can shed light on the mechanisms underlying community assembly from an evolutionary point of view. Finally, ecological- and phylogeny-based approaches are needed to set appropriate conservation priorities (Faith 1992; Forest *et al.* 2007).

The current concept of nestedness applied to metacommunities is restricted to a single measure of diversity: species richness. Similarly, it is restricted to a single measure of niche width or habitat specialization: species incidence. To overcome these limitations, here, we generalize the nestedness concept to any measure of diversity which takes into account the resemblance among species or among sites. The basic approach is to replace species richness or species incidence with branch lengths of tree-like objects in NODF. Accordingly, we term the generalized index as treeNODF. For instance, if the relationship among species is expressed by a phylogenetic tree, phyloNODF can be calculated using phylogenetic diversity (PD: the total branch length linking all species present at a site; Faith 1992; Lozupone & Knight 2005). Similarly, traitNODF can be obtained from community data and a functional dendrogram constructed from the traits of individual species (Petchey & Gaston 2002). As well as NODF, treeNODF can also be applied to nestedness among species. In this case, envNODF measures nestedness among species, taking into account a dendrogram representing the environmental resemblance of sites, and assesses whether the restricted set of environmental conditions in which a habitat-specialist species occurs is a nested subset of the conditions exploited by the habitat-generalist species.

We propose that treeNODF is a composite metric that can be additively partitioned in compositional (S.fraction) and resemblance (topoNODF) components of descriptor variables. We present appropriate null models and permutation tests to be used with treeNODF and its composition and resemblance components. Additionally, we used simulations to assess how treeNODF behaves under a range of matrix and tree properties. Finally, we illustrate the use of treeNODF by analysing data on Caribbean bats, using phyloNODF, traitNODF and envNODF.

## NODF and treeNODF

### THE NODF INDEX

NODF can be calculated for communities, where species are the descriptor variables; for species, where communities are the descriptors; or both (Almeida-Neto *et al.* 2008). For purposes of brevity, the following description will focus on the nestedness of communities. If used in an exploratory way, the first step to calculate NODF is to order communities (or rows of the data set) by decreasing species richness. A better interpretation of the results is possible, however, if the ordering of communities is dictated by a gradient factor that is presumed to generate nestedness (e.g. a gradient of humidity for amphibian species richness; Silva *et al.* 2012). Next, all pairs of communities are obtained, and a  $NODF_{\text{paired}}$  is computed for each community pair. Then, a  $NODF_{\text{sites}}$  can

be calculated by averaging the  $NODF_{\text{paired}}$  values.  $NODF_{\text{sites}}$  quantifies, thus, how much of the diversity in the supposedly species-poor samples is shared with their supposedly species-rich counterparts. To compute  $NODF_{\text{paired}}$ , the first step is to evaluate whether the species richness in the supposed species-poor community is lower than the species richness present in the supposed species-rich community. If this condition is not met,  $NODF_{\text{paired}}$  is zero. If the condition is met,  $NODF_{\text{paired}}$  is simply the proportion of the species richness present in the species-poor community that also occurs in the species-rich community. This can be expressed as  $a/(a+b)$ , where  $a$  is the number of species shared by the two communities and  $b$  is the number of species exclusive to the species-poor community. This is the Simpson similarity formulae, but applied only to pairs of samples that conform to the first condition described above (Almeida-Neto, Frensel & Ulrich 2012).

### THE TREENODF INDEX

treeNODF is a straightforward extension of NODF, in which the resemblance of descriptors, represented by a tree-like object, is taken into account. This can be carried out simply by replacing the incidence metric (e.g. species richness of sites, frequency of occurrence of species) by the total branch length of descriptor variables. For instance, phyloNODF is obtained by replacing species richness with phylogenetic diversity (PD; Faith 1992). In this case, we first determine whether the PD of a supposedly PD-poor community is lower than the PD of the supposedly PD-rich community. If this first condition is not met, treeNODF is zero. If the condition is met, phyloNODF<sub>paired</sub> is simply the proportion of the branches present in the PD-poor community that is shared with the PD-rich community. If branch lengths in a phylogenetic tree are interpreted as proportional to number of features (Faith *et al.* 2009), phyloNODF<sub>paired</sub> is the proportion of the features present in the features-poor community that is shared with the features-rich community.

treeNODF can be generalized for all cases in which tree-like structures are used to represent the resemblance of the descriptor variables under study. Accordingly, traitNODF can be used to assess the nestedness of communities, taking into account the resemblance of species in terms of their traits. In this case, resemblance is represented by a dendrogram constructed from species traits (e.g. Petchey & Gaston 2002).

Previous studies usually assessed nestedness among sites or in the entire matrix (species and sites). NODF, and thus treeNODF, can also be used to quantify nestedness among species. For instance, envNODF can be computed using a dendrogram representing the environmental resemblance of sites. In this case, envNODF can be used to test whether species present in a restricted range of environmental conditions (env-poor or specialist species) occur in a subset of the conditions occupied by species tolerating a wide range of conditions (env-rich or generalist species). This does not mean that a species occurring in a few sites (rare species) will present a compulsory nested

distribution in relation to another species occurring in many sites (frequent species). For instance, a rare species may occur in environmental conditions not shared by the frequent species (a turnover-like phenomenon). Also, a rare species may occur in a few but environmentally distinct sites that include the range of environmental conditions occupied by a frequent species.

#### TREENODF PROPERTIES

A perceptive description of treeNODF can be made by comparing its behaviour in particular cases to that of NODF (Fig. 1). Using a phylogeny as an example of a tree-like object, we have that:

1 phyloNODF is maximum when species composition in a community with low phylogenetic diversity is a perfect subset of those present in a community with high phylogenetic

diversity. In this case, phyloNODF and NODF will produce identical values (i.e. 100). This is illustrated by community pairs A–B and C–E in Fig. 1.

2 phyloNODF and NODF will be identical when a polytomous (star) tree is used (Fig. 1). In this case, all species are similarly related and phylogenetic diversity is simply the multiplication of species richness by a constant branch length. Polytomies are usually present in trees used in many phylogenetic studies. For these cases, phyloNODF values will approach those produced by NODF as more polytomies are included. However, the effects of low tree resolution should be more severe as the polytomies appear towards the root of the tree than when they are more at the terminal nodes (Swenson 2009).

3 phyloNODF can produce values larger than zero even if no species is shared between the two communities, a situation in which NODF (or the phyloNODF with a star tree, Fig. 1) would be zero (see B–D and D–E in Fig. 1). This is because communities share some evolutionary history, represented in the phylogenetic tree as shared branch lengths.

4 phyloNODF can produce values larger than zero even when species richness in the hypothesized community with high phylogenetic diversity is identical to (B–E in Fig. 1) or smaller than (D–E in Fig. 1) that observed in the community with low phylogenetic diversity. In this situation, NODF (or the phyloNODF with a star tree) would be zero (see B–D and D–E in Fig. 1). This is because the diversity metric used in phyloNODF (total branch length) is distinct from the metric used in NODF (species richness).

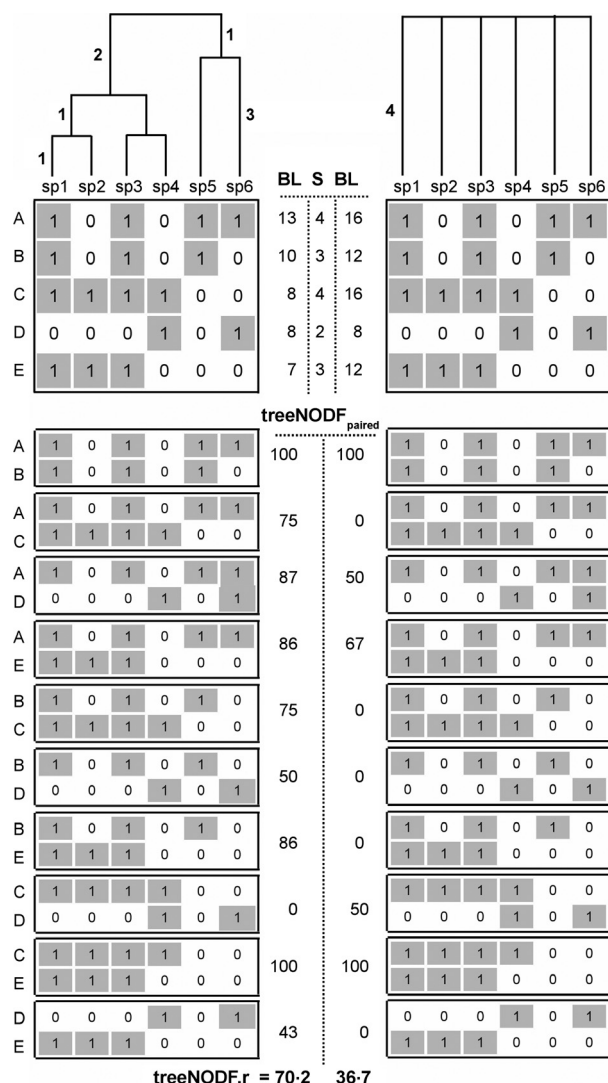
5 phyloNODF is affected by the phylogenetic originality of species, and thus, similar phyloNODF values can be produced when a community with low phylogenetic diversity shares half (A–D) or two-thirds (A–E) of its species richness (Fig. 1).

6 phyloNODF is affected by tree topology. Accordingly, the correlation of phyloNODF and NODF not necessarily will be high. For the example in Fig. 1, the correlation of paired communities for the two metrics is 0.37.

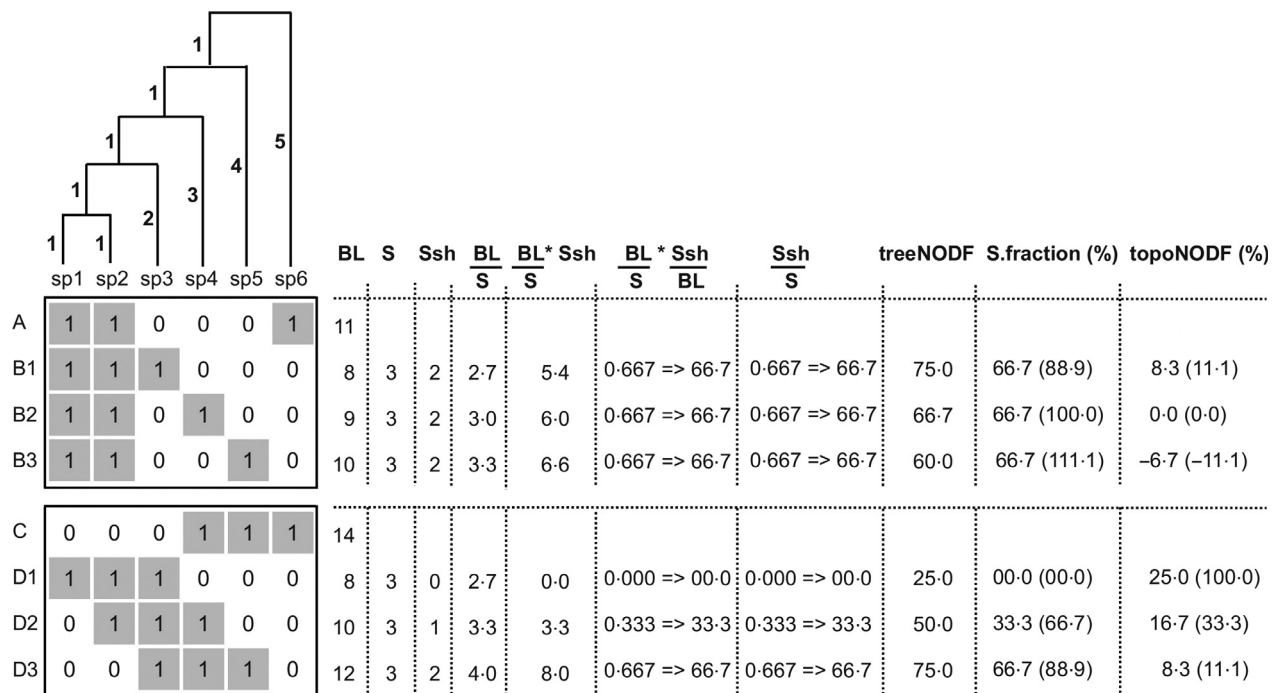
7 phyloNODF will be zero when phylogenetic diversity of communities is identical (C–D). This follows the basic rationale of the NODF, in which a fundamental condition to the existence of nestedness is that a supposedly diversity-poor community in fact presents lower diversity than a supposedly diversity-rich community.

#### Additive partitioning of treeNODF into compositional and resemblance components

treeNODF is a composite metric including components due to composition and resemblance of descriptor variables. The compositional component, termed here S.fraction, is the treeNODF fraction expected if the descriptor variables (e.g. species) are equally related (Fig. 2). S.fraction is the percentage of species richness in the BL-poor site that also occurs in the BL-rich site. For the pair A–B1 in Fig. 2, S.fraction is 66.7 because there are two shared species and the total species richness in BL-poor site (B1) is 3 ( $(2/3) \times 100 = 66.7$ ). An alternative understanding of the S.fraction can be obtained by examining



**Fig. 1.** treeNODF computed for rows of an artificial metacommunity data set and two trees representing resemblance among species. Values obtained using the polytomous tree are equal to that produced by NODF of species composition. BL = total branch length of the tree represented in a community site. S = species richness.



**Fig. 2.** treeNODF, S.fraction and topoNODF for rows of two sets of artificial metacommunity data sets. Values refer to comparisons of communities B1, B2 and B3 to A and D1, D2 and D3 to C. BL = total branch length of the tree represented in a community site. S = species richness. Ssh = number of species shared with the PD-rich (A or C) community. S.fraction = percentage of the species richness in the BL-poor community shared with the BL-rich community. topoNODF = treeNODF – S.fraction and quantifies effects of tree topology to treeNODF. See text for details.

the expected contribution of each species to the BL-poor site. For the B1 site, BL is 8 and species richness is 3. Thus, each species on average contributes 2.7 BL (8/3; Fig. 2). B1 shares two species with A, and thus, it could be expected that shared BL is 5.4 (2.7\*2). This expected shared BL represents 66.7% of the total BL of B1 (5.4/8). For the pair A–B2, treeNODF and S.fraction are identical because the expected shared BL assuming species are equally related  $[(2/3)*100 = 66.7]$  is equal to the observed BL shared  $[(6/9)*100 = 66.7]$ ; Fig. 2).

The second component is the fraction of the treeNODF due to tree topology and is obtained as  $\text{topoNODF} = \text{treeNODF} - \text{S.fraction}$ . That is, it is the treeNODF fraction not accounted by the shared species richness assuming species are equally related (S.fraction). Thus, it is the fraction resulting from particular tree topologies. For all the pairs in the first set of Fig. 2 (comparisons of A and each of B1, B2 and B3), species richness in the PD-poor site is 3 and shared species richness is 2 so that S.fraction is 66.7. It can be observed that the exclusive BL contribution of the species not shared with A in sites B1, B2 and B3 are 2, 3 and 4, respectively (Fig. 2). These non-shared BLs represent, respectively, 0.25 (2/8), 0.3 (3/9) and 0.4 (4/10) of the total BL in these BL-poor sites. Assuming species are equally related (a star tree), the expected contribution would be 0.3 in all three cases. Accordingly, topoNODF is positive (8.3) for A–B1 as the non-shared species (sp3) contributes a fraction of the total BL (0.25) that is lower than the expected if all species are equally related (0.3). For the pair A–B2, topoNODF is 0 because the observed contribution of the non-shared species (sp4) is equal to that expected if species were similarly related

(0.3). Finally, topoNODF is negative (–6.7) for the pair A–B3 because the non-shared species (sp5) contributes more (0.4) than the expected (0.3).

The second set of pairs of sites in Fig. 2 (C and each of D1, D2 and D3) present distinct numbers of shared species. For the pair C–D1, there is no shared species and thus S.fraction is 0 and treeNODF = topoNODF. For the other two pairs, the number of shared species is higher (2 and 3 for C–D2 and C–D3, respectively), thus decreasing the value of topoNODF (16.7 and 8.3, respectively).

### Effects of matrix properties and tree topology on treeNODF and topoNODF

Because some matrix properties affect the absolute values produced by nestedness metrics (Almeida-Neto *et al.* 2008), we investigated the effects of matrix size and matrix fill on treeNODF and topoNODF values. In all cases, treeNODF and topoNODF were calculated for sites (rows).

To investigate matrix fill, we created random incidence matrices of 30 rows  $\times$  30 columns. These matrices were filled with 10, 20, 30 ... 90% of presences. Fifteen matrices were generated for each level of matrix filling. Regarding matrix size, we generated 15 random matrices with 50% fill for each of the sizes 6 (rows)  $\times$  6 (columns), 10  $\times$  10, 20  $\times$  20 ... 100  $\times$  100. We generated random rooted coalescent trees of appropriate matrix sizes using the function *rcoal* of package *ape* (Paradis, Claude & Strimmer 2004) of the R statistical environment (R Core Team 2012).



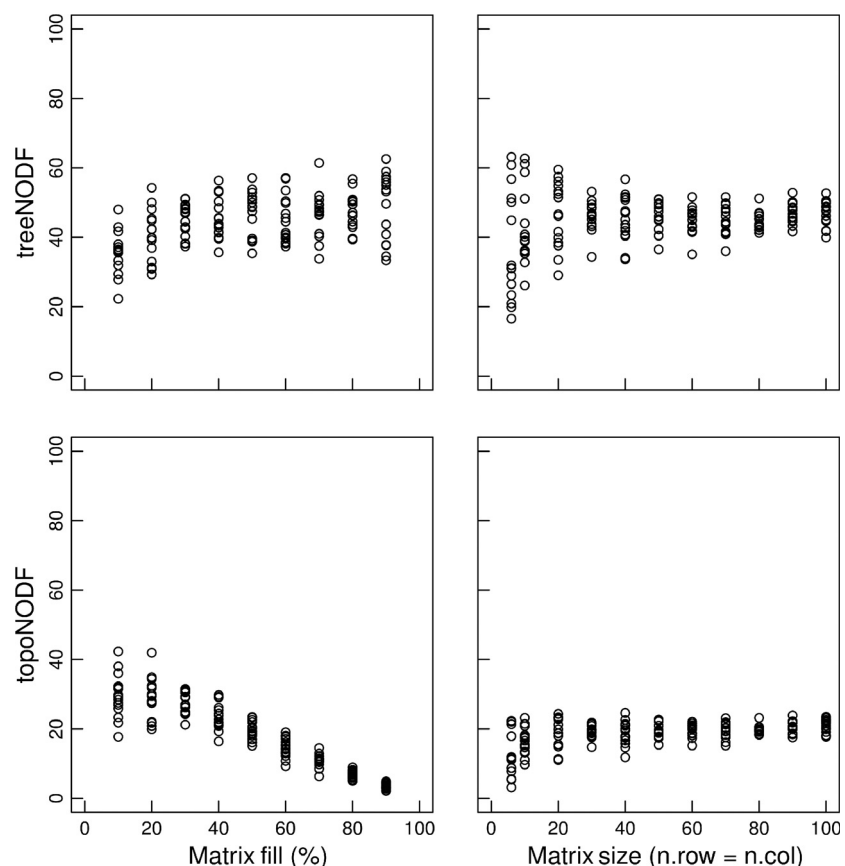
We also assessed the effects of tree shape properties, imbalance and relative position of internal nodes on absolute treeNODF and topoNODF values. For imbalance, we generated 500 random trees with the function *rreeshape* (model type 'biased' and random probabilities) and measured imbalance with the function *colless*, both included in the package *apTree-shape* (Bortolussi *et al.* 2011) of R. The random matrices were size  $30 \times 30$ , with 50% fill. To assess the effects of the relative position of internal nodes, we generated 200 random trees using the function *pbtree* and obtained the gamma statistics of Pybus & Harvey (2000) for these trees, using the function *litt*, both available in the package *phytools* (Revell 2012). Negative values of gamma indicate that most divisions in the tree-like object occur at its base (high level of differences), whereas positive values indicate more divisions at the tips (low level of differences). The random matrices were size  $30 \times 30$  with 50% fill.

When applied to simulated matrices that are not ordered by decreasing BL diversity, treeNODF was mostly insensitive to the matrix fill (Fig. 3). However, the topoNODF applied to the same simulated data showed a negative trend, reflecting the positive correlation of S.fraction with the matrix fill. As matrix fill increases, more species are shared among samples, and thus, S.fraction tends to increase. This behaviour has been previously noted for NODF (Almeida-Neto *et al.* 2008). treeNODF and topoNODF were almost completely insensitive to matrix size, although some variation was observed when small

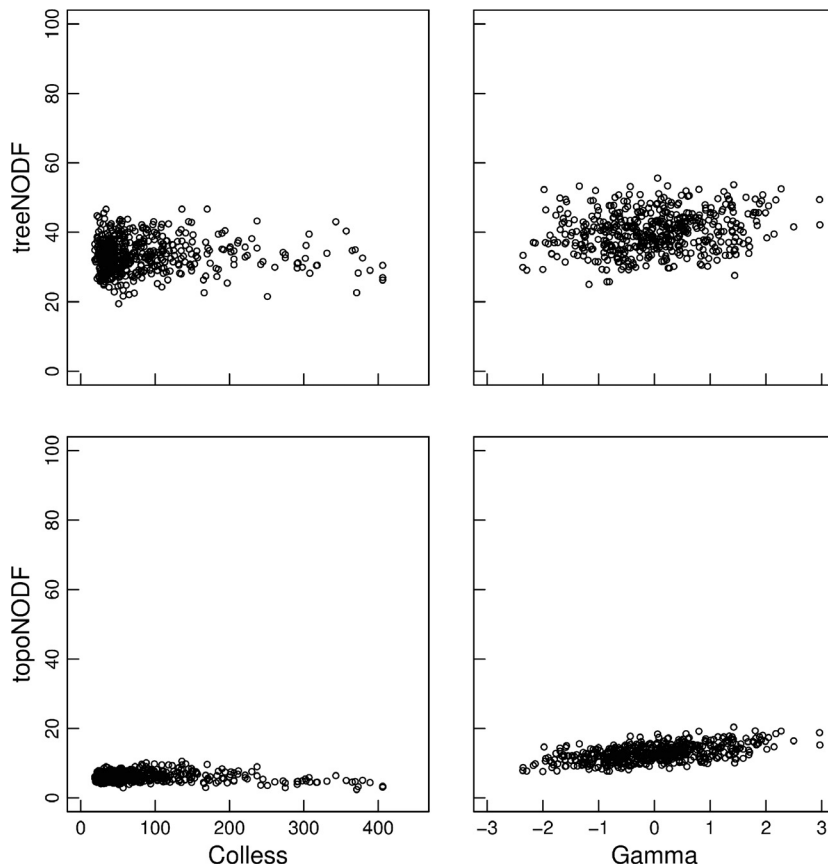
matrix sizes were used (Fig. 3). treeNODF and the topoNODF were not affected by tree imbalance, measured as the Colless shape statistic (Fig. 4). This indicates that summed differences between paired branch lengths at all nodes do not have an effect on values produced by treeNODF and topoNODF. Similarly, the relative position of nodes in the simulated trees, measured as Gamma statistic, presented no effect on treeNODF and topoNODF (Fig. 4), indicating that branching at the base or at the tips of random trees does not have a strong effect on indices.

When simulated matrices were ordered by decreasing BL diversity, high matrix fill and size caused treeNODF to converge to its maximum value of 100 (Fig. S1, Supporting information). topoNODF was affected by matrix fill, with low values observed in highly filled matrices. As highlighted above, this is a direct consequence of the positive relationship of S.fraction to matrix fill. topoNODF was mostly unaffected by matrix size (Fig. S1, Supporting information). treeNODF and topoNODF were unaffected by the tree imbalance and only slightly influenced by the relative position of tree branching (respectively measured by Colless and Gamma statistics; Fig. S2, Supporting information).

It should be made clear that results in this section refer to potential bias of matrix and tree properties on *absolute* values produced by the metrics. A proper interpretation of the importance of a particular value should be made using null models or permutation tests.



**Fig. 3.** Effects of matrix fill and size on treeNODF and topoNODF. Random matrices used to evaluate fill consisted of 30 rows and 30 columns. Random matrices used to evaluate size were filled with 50% presences. In all cases, nestedness measures were obtained over matrices not ordered by decreasing diversity.



**Fig. 4.** Effects of two tree properties on treeNODF and topoNODF. Colless measures the imbalance of the trees. Gamma indicates the relative position of internal nodes, with negative values indicating that most speciation occurs at the tree base, while positive values indicate recent speciation. Random matrices used consisted of 30 rows and 30 columns and were completed with 50% of presences. In all cases, nestedness measures were obtained over matrices that were not ordered by decreasing diversity.

### Null models and permutation tests

Values produced by treeNODF and its component metrics can be evaluated for significance using null models or permutation tests. A number of null models are available (see Gotelli 2000; Hardy 2008; Ulrich, Almeida-Neto & Gotelli 2009), among them:

**1 fixedFixed:** shuffles elements in the site by species matrix, while keeping the column and row totals preserved. This is a very conservative null model as most of the matrix structure is maintained. We propose here the use of the fixedFixed null model when sites and species are ordered simply by decreasing values of total branch length (e.g. PD in the case of phyloNODF).

**2 tipLabels:** shuffles the tip labels of the tree representation. The site-by-species matrix is maintained intact, but the relatedness among their constituent species is broken down. tipLabels is suitable for testing the phylogenetic-only (or resemblance to the other tree-like objects) component of the phyloNODF (or treeNODF). Accordingly, it should be very conservative to detect tree nestedness. It can be used with permutation tests (e.g. permRows or permColumns, see below) to evaluate whether tree nestedness is due to the ordering of sites (or species) or to the tree structure. According to Hardy (2008), this null model is not too liberal, as are other null models that randomize the site-by-species matrix.

Many metrics used to assess nestedness include an algorithm to order sites and/or species before computation (see Ulrich, Almeida-Neto & Gotelli 2009). This is also the case for NODF

and can be done simply by decreasing columns and rows sums. However, the automatic ordering of sites and/or species precludes the development of a solidly based inference regarding the mechanisms generating nestedness. An informed use of NODF can be accomplished by ordering sites and/or species according to a hypothetical factor generating nestedness (Ulrich, Almeida-Neto & Gotelli 2009; Ulrich & Almeida-Neto 2012). Both automatic and hypothesis-driven ordering can be used for treeNODF, although if ordering is done simply by decreasing diversity, the total branch length (e.g. PD for phylogenetic trees, FD for trait-based trees) must be used, rather than species richness. When a hypothetical factor is used to order rows (sites) or columns (species) in the metacommunity data set, a straightforward assessment of the importance of treeNODF, S.fraction and topoNODF can be made through a permutation test (Lomolino 1996):

**1 permRows:** randomly reorders rows of the matrix and calculates the treeNODF values for each rows-reordered matrix. Because entire rows are permuted, the original species and phylogenetic composition of sites, in the case of metacommunities, remain unchanged. The permRows test must only be used when rows (i.e. sites) are ordered according to some *a priori* hypothesis, such as area size, degree of isolation or time since the last disturbance event. In fact, automatic ordering of the matrix by decreasing diversity will produce simulated statistics in which the maximum is lower than or equal to the observed value, but never higher.

**2 permColumns:** randomly reorders columns of the matrix and calculates the treeNODF values for each columns-reordered

matrix. This test is similar to permRows, but is intended to test objects located in columns and with descriptor variables in rows. For a site-by-species matrix, it can be used to test whether the environmental conditions used by specialist species are a subset of the environmental conditions in which generalist species occur. This test must only be used when columns (i.e. species) are ordered according to some species-related *a priori* hypothesis, such as dispersal ability, body size or abundance.

## Case study

### DATA SETS

We used data available on the occurrence of bat species on Caribbean islands (Willig *et al.* 2010) to exemplify the empirical uses of the treeNODF metrics. The Caribbean is a hot spot of biodiversity, with high species richness and endemism, especially for bats (Willig *et al.* 2010). Such complex patterns of endemism and richness are the outcome of geological and evolutionary patterns (historical processes) as well as island size and distance from the mainland (ecological processes).

To calculate phyloNODF, we produced a phylogenetic tree according to the relationships proposed by Bininda-Emonds *et al.* (2008). For the traitNODF, we used data on body size and dietary guilds (Willig *et al.* 2010) to produce a dendrogram (Gower distance and UPGMA clustering algorithm) that represents species similarity with respect to these traits. These traits efficiently capture how bat species use and compete for resources and are commonly used in studies involving functional diversity, not only for bats (Willig, Kaufman & Stevens 2003) but also for other mammals (Safi *et al.* 2011). To calculate envNODF, we used island area and maximum altitude to produce a dendrogram (log data, Euclidean distance and UPGMA) representing island similarities. These features are thought to be good surrogates for habitat diversity and resource availability.

To evaluate whether the observed values differed from those expected by chance, we used permRows and permCols permutation tests (999 permutations) depending on the hypotheses being tested (Table 1), specifically:

**1** Phylogenetic and trait diversities on small islands are a subset of those present on large ones. Large islands usually have more habitat types and resources, thus allowing high levels of phylogenetic and trait diversity. We calculated phyloNODF and traitNODF and their components ordering the rows in the community matrix according to decreasing island area and used the simple permRows algorithm that randomizes rows (sites; islands in our case) to test whether the observed values of our metrics could be produced by chance.

**2** Body size is positively related to dispersal and colonization ability, and therefore, species with a smaller body size should occur nested in the environment occupied by species with large body size. Thus, envNODF was used to assess nestedness among species with respect to environmental conditions. We ordered species by decreasing body size and used the permCols test.

**Table 1.** treeNODF analyses of Caribbean bats. phyloNODF and traitNODF assess nestedness taking into account phylogenetic and trait relatedness, respectively. envNODF assesses nestedness among species taking into account the environmental resemblance of sites. The rows (sites) of the metacommunity data set used in phyloNODF and traitNODF were ordered by decreasing island area and assessed using the permRows test. The columns (species) of the metacommunity data set used in envNODF were ordered by decreasing bat body size and assessed using the permCols test. S.fraction measures nestedness due to composition, and topoNODF measures nestedness due to tree topology. In all cases, indices were obtained for a single matrix dimension (rows [sites] for phyloNODF and traitNODF and columns [species] for envNODF)

Metric	Observed	Mean	SD	Z	P
phyloNODF	53.89	36.14	3.23	5.50	0.001
S.fraction	38.33	24.99	2.41	5.54	0.001
topoNODF	15.56	11.15	0.96	4.57	0.001
traitNODF	53.86	36.60	3.44	5.02	0.001
S.fraction	38.42	24.83	2.52	5.39	0.001
topoNODF	15.44	11.77	1.15	3.20	0.002
envNODF	36.41	29.90	3.30	1.97	0.020
S.fraction	21.11	17.06	2.28	1.77	0.035
topoNODF	15.30	12.84	1.56	1.58	0.052

Mean = Mean of statistics obtained under a null model (999 randomizations). SD = Standard deviation of statistics obtained under a null model. Z = (Observed – Mean)/SD.

All analyses were carried out using the R environment. Functions to calculate and test treeNODF and its components are available in the R package CommEcol (Melo 2013).

### RESULTS

Bat assemblages were nested with respect to their phylogenetic and trait diversity (in all cases  $P = 0.001$ ; Table 1). Small islands tended to harbour not only assemblages that are a subset of larger islands, but also species that are phylogenetically and functionally similar to those present on large islands. Therefore, the phylogenetic and functional diversity present in small islands is not only reduced when compared to those present in large islands, but is a nested subset of those present in the latter.

envNODF indicated that the species were nested ( $P = 0.020$ ; Table 1). However, nestedness among species was mostly due to the species composition ( $P = 0.035$ ), indicating that small-bodied species tended to occur on subsets of the islands occupied by large-bodied species. The topoNODF component was slightly important ( $P = 0.052$ ), indicating a weak tendency of small-bodied species to occur in a restricted range of the environmental conditions (summarized in the dendrogram as island area and maximum altitude) occupied by large-bodied species.

## Discussion

The treeNODF approach is based directly on the NODF index (Almeida-Neto *et al.* 2008) and thus immune to some known problems and inconsistencies that are observed with other

species-based nestedness metrics. For instance, a common undesirable property of some metrics is that they count pairs of columns or rows positively towards the degree of nestedness, even when their information or diversity is the same (for further details, see Almeida-Neto *et al.* 2008; Ulrich, Almeida-Neto & Gotelli 2009).

Previous assessments of nestedness have been carried out mostly using traditional null models (Ulrich & Gotelli 2007), and this is also true for the use of NODF (e.g. Silva *et al.* 2012). In these cases, NODF is usually obtained after ordering rows and/or columns of the metacommunity matrix by decreasing species richness and/or decreasing species frequency. A major drawback of this procedure is that the finding of nestedness does not allow a solidly based inference regarding the mechanisms generating nestedness, as many potential mechanisms can be invoked (Ulrich, Almeida-Neto & Gotelli 2009). A much wiser use of NODF and treeNODF can be made, however, by ordering rows and/or columns according to a hypothetical factor generating nestedness. This approach was suggested early in the nestedness literature by Lomolino (1996) and later reinforced by Ulrich, Almeida-Neto & Gotelli (2009), but has been used only rarely (e.g. Silva *et al.* 2012). This hypothesis-driven approach not only allows for a stronger inference regarding causal mechanisms, but also avoids the difficult task of choosing a proper null model, as permutation tests can be used to assess the significance of the observed statistic.

#### CASE STUDY

On islands, nested species assemblages have usually been explained as an outcome of differential species extinctions and colonizations (e.g. Lomolino 1996). The former is suggested when nestedness is correlated with island area, whereas the latter should occur when it is correlated with the degree of island isolation (see references in Martínez-Morales 2005). We found a significant nested structure related to island area, not only for species composition (S.fraction) but also for phylogenetic and trait structures. This means that small islands are inhabited not only by a reduced number of bat species, but also by bat species that reflect the evolutionary history and diversity of functional traits present on large islands. Considering that area is often a good surrogate for habitat diversity, the high levels of phylogenetic and trait diversities on large islands would simply reflect different lineages or species with different traits that occupy different habitats. Studying the relative contribution of island area and habitat diversity to bat species richness in the Lesser Antilles, Ricklefs & Lovette (1999) found that whereas area explained about 30% of species richness, habitat diversity alone explained < 2%. Also, more than 40% of the variation remained unexplained. The inclusion of aspects of phylogeny and trait diversity in such an analysis could allow one to evaluate niche complementarity and thus to better understand the role of species interactions (e.g. competition) and environmental filters. Further analyses including habitat diversity on each island could shed some light on the mechanism behind this pattern.

#### APPLICATIONS IN ECOLOGY, BIOGEOGRAPHY AND CONSERVATION

The potential applications of the new nestedness approach proposed here are diverse. treeNODF is basically a way to express species or site resemblance through a tree-like representation, and thus, the approach can be further extended to other kinds of ecological information and also to networks of interacting species.

In metacommunities characterized by a gradient of environmental stress, for instance, phylogenetic nestedness may be present if the species present in communities at the extreme of the gradient are derived from a single or a few (adapted) clade(s). If this is the case, phylogenetic nestedness could be interpreted as the selection of features (Faith *et al.* 2009), clumped in the phylogeny, which allow species to survive in extreme portions of the gradient. This seems to be the case of hummingbird communities in Ecuador, where moist lowland communities tend to be phylogenetically overdispersed and include representatives of several lineages, in contrast to those in highlands where environmental constraints (e.g. low temperature) select phylogenetically clustered communities derived from fewer lineages (Graham *et al.* 2009). Similarly, a stress gradient may generate functional nestedness if conditions at increasingly stressed sites select species with a restricted set of traits that allow them to survive. In fragmented landscapes, phylogenetic nestedness is expected to occur if dispersal ability is a strongly conserved trait in the phylogeny.

With respect to island biogeography, islands in an archipelago can be partially nested in terms of species composition but not in terms of phylogeny if good dispersers are derived from many lineages. Additionally, islands may harbour a similar number of species and thus do not show nestedness in the composition of the biota, but may still show phylogenetic nestedness if species on large islands and/or close to the continent include more different lineages. phyloNODF may also be useful to test historical hypotheses in a geographical context at distinct hierarchical levels. For instance, one could hypothesize that under recent dispersal and migration of populations, alleles found in a recently occupied area are nested subsets of those in areas occupied by a species before dispersal (see a related example in Diniz-Filho *et al.* 2012). In a biogeographical context, a similar pattern of nested subsets would be expected under models explaining latitudinal diversity gradients, such as the tropical niche conservatism model (Wiens & Donoghue 2004) or the Out of Tropics (OTT) model (Jablonski, Roy & Valentine 2006). In such models, tropical areas harbour many (e.g. old and new) lineages (and thus high PD), but temperate regions are composed of a few derived lineages (and low PD) nested within tropical lineages.

Most studies on nestedness have been carried out using the entire matrix (for instance using the Temperature metric) (Ulrich, Almeida-Neto & Gotelli 2009). However, a desirable characteristic of the NODF metric, and thus of treeNODF, is that analyses can be restricted to each matrix margin separately. Accordingly, some recent studies have evaluated nestedness among communities (or rows; e.g. Schneck, Schwarzbold



& Melo 2011; Silva *et al.* 2012). However, ecologists have seldom assessed nestedness among species. treeNODF extends this mostly unexplored approach to assess nestedness among species in terms of the environmental conditions under which they are able to survive. For instance, for a monophyletic clade, species may be ordered by estimated time of appearance, to assess whether derived species inhabit specialized habitat conditions that are a subset of the conditions in the habitat exploited by the basal species.

The treeNODF allows one to extend the rationale applied to the nested structure in species composition to the protection of other important aspects of biodiversity such as phylogenetic and functional (trait) diversities. This expands the conservation biologist's toolkit, because finding a nested structure in species composition does not guarantee that phylogenetic or trait diversities are also nested. In such cases, species-poor communities would not be selected, even if they include species from different clades, for instance primitive or species-poor lineages, or species with unique ecological traits. That is, species-poor communities may include not only complementary features, but also share a portion of the features presented by species-rich communities as they share basal branch lengths. A framework where one applies both the NODF and treeNODF measures would improve our ability to conserve simultaneously these aspects of biodiversity.

## Acknowledgements

ASM, MVC and MAN received research fellowships (307479/2011-0, 306843/2012-9 and 306870/2012-6, respectively), and ASM received a research grant (558187/2009-9) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Janet Reid reviewed the English. José Alexandre F. Diniz-Filho suggested some applications of the treeNODF. Andrés Baselga, one anonymous referee and one associated editor provided useful suggestions.

## Data accessibility

Data on bats occurrence and feeding guilds, island area and maximum altitude were obtained from Willig *et al.* (2010), available at <http://digitalcommons.unl.edu/museummammalogy/130>. Data on bats body size were obtained from Jones *et al.* (2009), available at <http://esapubs.org/archive/ecol/e090/184/#data>. The bats phylogenetic tree was obtained from Bininda-Emonds *et al.* (2008), available at <http://www.nature.com/nature/journal/v446/n7135/extref/nature05634-s2-revised.txt>. Analyses were conducted using functions available in the R package CommEcol (Melo 2013), available at <http://r-forge.r-project.org/projects/commecol/>.

## References

- Almeida-Neto, M., Frensel, D.M.B. & Ulrich, W. (2012) Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). *Global Ecology and Biogeography*, **21**, 772–777.
- Almeida-Neto, M., Guimarães, P., Guimarães, P.R. Jr, Loyola, R.D. & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227–1239.
- Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., Macphee, R.D., Beck, R.M., Grenyer, R. *et al.* (2008) The delayed rise of present-day mammals. *Nature*, **456**, 274–274.
- Bortolussi, N., Durand, E., Blum, M. & Francois, O. (2011) *apTreeshape: Analyses of Phylogenetic Treeshape*. R package version 1.4-4. <http://CRAN.R-project.org/package=apTreeshape>.
- Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the USA*, **105**(Suppl. 1), 11505–11511.
- Darlington, P.J. Jr (1957) *Zoogeography: The Geographical Distribution of Animals*. John Wiley & Sons, New York.
- Diniz-Filho, J.A.F., Collevatti, R.G., Soares, T.N. & Telles, M.P.C. (2012) Geographical patterns of turnover and nestedness-resultant components of allelic diversity among populations. *Genetica*, **140**, 189–195.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.
- Faith, D.P., Lozupone, C.A., Nipperess, D. & Knight, R. (2009) The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *International Journal of Molecular Sciences*, **10**, 4723–4741.
- Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P. *et al.* (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, **445**, 757–760.
- Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.
- Graham, C.H., Parra, J.L., Rahbek, C. & McGuire, J.A. (2009) Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(Suppl. 2), 19673–19678.
- Hardy, O.J. (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.
- Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–106.
- Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L. *et al.* (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, **90**, 2648.
- Lomolino, M.V. (1996) Investigating causality of nestedness of insular communities: selective immigrations or extinctions? *Journal of Biogeography*, **23**, 699–703.
- Lozupone, C. & Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.
- Martínez-Morales, M.A. (2005) Landscape patterns influencing bird assemblages in a fragmented Neotropical cloud forest. *Biological Conservation*, **121**, 117–126.
- Melo, A.S. (2013) CommEcol: Community Ecology analyses. URL <http://commecol.r-forge.r-project.org/>
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Petchev, O.L. & Gaston, K.J. (2002) Functional Diversity (FD), species richness, and community composition. *Ecology Letters*, **5**, 402–411.
- Pybus, O.G. & Harvey, P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **267**, 2267–2272.
- R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- Ricklefs, R.E. & Lovette, I.J. (1999) The role of island area *per se* and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. *Journal of Animal Ecology*, **68**, 1142–1160.
- Safi, K., Cianciaruso, M.V., Loyola, R.D., Brito, D., Armour-Marshall, K. & Diniz-Filho, J.A.F. (2011) Understanding global patterns of mammalian functional and phylogenetic diversity. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **366**, 2536–2544.
- Schneck, F., Schwarzbald, A. & Melo, A.S. (2011) Substrate roughness affects stream benthic algal diversity, assemblage composition, and nestedness. *Journal of the North American Benthological Society*, **30**, 1049–1056.
- Silva, F.R., Almeida-Neto, M., Prado, V.H.M., Haddad, C.F.B. & Rossa-Feres, D.C. (2012) Humidity levels drive reproductive modes and phylogenetic diversity of amphibians in the Brazilian Atlantic Forest. *Journal of Biogeography*, **39**, 1720–1732.
- Swenson, N.G. (2009) Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLoS ONE*, **4**, e4390.
- Ulrich, W. & Almeida-Neto, M. (2012) On the meanings of nestedness: back to the basics. *Ecography*, **35**, 865–871.

- Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3–17.
- Ulrich, W. & Gotelli, N.J. (2007) Null model analysis of species nestedness patterns. *Ecology*, **88**, 1824–1831.
- Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.
- Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273–309.
- Willig, M.R., Presley, S.J., Bloch, C.P. & Genoways, H.H. (2010) Macroecology of Caribbean bats: effects of area, elevation, latitude, and hurricane-induced disturbance. *Island Bats: Evolution, Ecology, and Conservation* (eds T.H. Fleming & P.A. Racey), pp. 216–264. University of Chicago Press, Chicago, Illinois.
- Wright, D.H. & Reeves, J.H. (1992) On the meaning and measurement of nestedness of species assemblages. *Oecologia*, **92**, 416–428.

Received 12 May 2013; accepted 6 March 2014  
Handling Editor: Daniel Faith

## Supporting Information

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

**Figure S1.** Effects of matrix fill and size on treeNODF and topoNODF.

**Figure S2.** Effects of two tree properties on treeNODF and topoNODF.