

Article

MBI: an R package for calculating multiple-site beta diversity indices

YouHua Chen

Department of Zoology, University of British Columbia, Vancouver, V6T 1Z4, Canada

E-mail: yhchen@zoology.ubc.ca, haydi@126.com

Received 14 March 2013; Accepted 16 April 2013; Published online 1 June 2013

Abstract

Beta diversity is one of most important features in community ecology. Indices for pairwise comparison of beta diversity have been extensively developed, but the ones specifically designed for multiple-site comparison of beta diversity are still limited. Currently, by compiling all the available metrics based on the previous literature, plus some new metrics developed in the present report, we made the calculation of these multiple-site beta-diversity statistics become ready for ecologists using R computing environment. An empirical study was present using 290 real-world presence/absence matrices. The results showed that (1) mean pairwise indices could be good surrogates for multiple-site indices in principle, except the mean pairwise richness different index; (2) most of the indices were highly correlated, as indicated by Pearson correlation and significance test. The new R package “MBI” for calculating multiple-site diversity indices could be downloaded from the <http://cran.r-project.org/web/packages/MBI/>.

Keywords beta diversity; multiple-site comparison; community ecology; nestedness; turnover; richness.

<p>Computational Ecology and Software ISSN 2220-721X URL: http://www.iaees.org/publications/journals/ces/online-version.asp RSS: http://www.iaees.org/publications/journals/ces/rss.xml E-mail: ces@iaees.org Editor-in-Chief: WenJun Zhang Publisher: International Academy of Ecology and Environmental Sciences</p>
--

1 Introduction

Development of biodiversity indices for better quantifying beta diversity is one fundamental prospect of community ecology. In recent progresses, many endeavors have been devoted to design and develop new statistics to quantify different facets of beta diversity. For most of the available beta diversity indices, they are typically pairwise-site comparison, thus for communities with many sites, the calculated values must be present in the form of matrices. This pairwise comparison might thus limit the potential applications, for example, in the case of the comparison of different communities each of which is comprised of many sampling plots (Zhang, 2011a).

In recent years, there is a growing trend that people are now trying to quantify beta diversity at community (or multiple-site) level. For example, Baselga (Baselga et al., 2007; Baselga, 2010) partitioned beta diversity into turnover and nestedness components at community levels. Later, it was found that these two indices were too insensitive to species loss in the community (Almeida-Neto et al., 2012; Carvalho et al., 2012a,b). In a recent study, beta diversity was suggested to partition into the components of replacement and richness

differences (Almeida-Neto et al., 2012; Carvalho et al., 2012a,b) at the pairwise-site level. However, the multiple-site counterparts of these metrics were still lacking, which would be developed in the present study. Finally, WNODF index (Almeida-Neto and Ulrich, 2011) has been advocated as an effective indicator to quantify nestedness of the community.

Programs for calculating beta-diversity has been widely developed and there are a series of computer programs available for calculating pairwise indices, for example, function “*vegdist*” in the R packages “*vegan*” (Oksanen et al., 2012). However, there are just a few the programs specifically for multiple-site indices calculation, and they are typically only able to calculate one or two particular multiple-site indices. For example, WNODF index was available in a FORTRAN program developed by the authors (Almeida-Neto and Ulrich, 2011) and “*vegan*” package (Oksanen et al., 2012). Recently, Baselga and Orme (2012) developed an R package “*betapart*” for calculating the nestedness and turnover indices mentioned above. Thus, there were no a program that can put different multiple-site indices together and ecologists should feel hard to compare different multiple-site beta diversity indices simultaneously since some indices were not included in computer programs. Thus, it is very necessary to develop a general package specifically for multiple-site beta diversity calculations, thus allowing ecologists to compare various indices easily for their own data sets. Inspired from this issue, I tried to collect different multiple-site diversity indices together and make their calculation become available using R scientific computing platform (R Development Core Team, 2011).

In this short report, I demonstrated a new R package, called “*MBI*” (<http://cran.r-project.org/web/packages/MBI/>), for calculating different indices of beta diversity for the purpose of multiple-site comparison. All the previously proposed indices, with the addition of some more indices developed in the present study, were included (Appendix). Currently, there were 21 indices available, in which 10 novel ones were developed in the present study. Among these new indices, 7 were developed on the basis of pairwise indices which calculated the index values for only two sites each time iteratively and then average the values across all the pairs of sites in the community (Appendix). The other 3 ones, including multiple-site versions of replacement, richness difference and Lennon’s richness indices, were developed as the analogues of Baselga’s nestedness and turnover indices (Appendix). The deduction of these indices followed the same steps as Baselga’s indices (Baselga, 2010).

Thus, our multiple-site indices could be classified into two categories. The first group contained those deduced from mathematical formulation (called as “multiple-site indices without pairwise calculations”) (Baselga, 2010; Baselga et al., 2007), while the other group involved those with straightforward calculations by taking the averages over all the pairwise values (called as “mean pairwise indices”). The latter category included 7 new indices developed in the present study, while the former one was comprised of the other 14 indices (Appendix).

Because we introduced a series of new indices for measuring multiple-site diversity, in addition to the demonstration of MRI package, another objective of the study is to compare the effectiveness of these new indices on quantifying multiple-site diversity patterns with reference to the available ones previously published.

2 Materials and Methods

Here we tested the indices using the public 290 presence/absence real-world matrices, which were compiled by the previous work (Atmar and Patterson, 1995). These real-world matrices have varying species and site numbers. Thus, they can help quantify the relationships of these multiple-site indices from an empirical perspective.

Similar to the study in Zhang (2011b), a Pearson’s correlation analysis was performed to compare all the indices and the significance levels of the correlations were tested with Bonferroni correction when comparing

the values for different indices across the 290 matrices.

3 Results and Discussion

3.1 A comparison of all multiple-site indices

When observing the index values for the 290 real-world ecological matrices, most of the indices were tightly associated as indicated by Table 1. Typically each pair from the 21 available metrics has high correlation values and significance level of P values < 0.001. Values of indices for each of the 290 matrices were present in Supplementary Material.

One index, the mean pairwise richness difference index, had low consistent degrees of measurements when correlated it to other indices. Thus, except for this index, other indices shared a high congruence across the 290 matrices principally. Therefore, most of the diversity indices should be appropriate for measuring and comparing beta diversity among multiple communities.

Table 1 Pearson's correlation analysis and significance test among the multiple-site diversity indices across 290 real-world presence/absence matrices. Upper triangular zone indicated the P values, while lower triangular zone showed the correlation coefficients. The abbreviations of each index were adopted from the function names showed in Appendix.

	cn	ct	crep	crich	wnodfT	wnodfC	wnodfR	wbeta	harrison	ht	wt	do	cl	msorensen	mjaccard	mn	mt	mrep	mrich	ml	cfull	
cn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0
ct	-0.76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0
crep	-0.87	0.77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0
crich	0.69	-0.17	-0.72	0	0	0	0	0	0	0	0	0	0	0.42	0.54	0	0	0	0	0	0	0
wnodfT	0.56	-0.43	-0.56	0.46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.27	0	0.01
wnodfC	0.55	-0.43	-0.54	0.43	0.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.11	0	0
wnodfR	0.59	-0.56	-0.59	0.37	0.48	0.34	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0	0
wbeta	-0.31	0.61	0.22	0.23	-0.27	-0.29	-0.26	0	0	0.53	0	0	0	0	0	0.39	0	0	0	0	0	0
harrison	-0.24	0.26	0.39	-0.27	-0.42	-0.43	-0.27	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0
ht	-0.5	0.39	0.68	-0.59	-0.48	-0.46	-0.4	0.03	0.8	0	0	0	0	0	0	0	0	0	0	0.08	0	0.03
wt	-0.78	0.8	0.85	-0.48	-0.62	-0.62	-0.55	0.5	0.48	0.63	0	0	0	0	0	0	0	0	0	0	0	0
do	0.24	-0.26	-0.39	0.27	0.42	0.43	0.27	-0.18	-0.99	-0.8	-0.48	0	0	0	0	0	0	0	0	0	0.08	0
cl	0.69	-0.17	-0.72	1	0.46	0.43	0.37	0.23	-0.27	-0.59	-0.48	0.27	0	0.42	0.54	0	0	0	0	0	0	0
msorensen	-0.34	0.67	0.46	0.04	-0.35	-0.37	-0.4	0.61	0.68	0.5	0.62	-0.68	0.04	0	0	0.1	0	0	0	0	0	0
mjaccard	-0.35	0.68	0.48	0.03	-0.37	-0.38	-0.38	0.58	0.68	0.5	0.63	-0.68	0.03	0.98	0	0.16	0	0	0	0	0	0
mn	0.84	-0.48	-0.82	0.85	0.48	0.44	0.51	-0.05	-0.21	-0.54	-0.61	0.21	0.85	-0.09	-0.08	0	0	0	0	0	0	0.43
mt	-0.71	0.79	0.79	-0.39	-0.53	-0.52	-0.58	0.52	0.66	0.69	0.82	-0.66	-0.39	0.86	0.84	-0.59	0	0	0	0.1	0	0
mrep	-0.8	0.76	0.93	-0.63	-0.56	-0.54	-0.6	0.33	0.59	0.77	0.83	-0.59	-0.63	0.67	0.68	-0.73	0.92	0	0	0	0	0
mrich	0.12	0.33	-0.07	0.51	-0.06	-0.09	-0.08	0.57	0.48	0.1	0.21	-0.48	0.51	0.82	0.79	0.37	0.47	0.16	0	0	0	0
ml	0.55	-0.08	-0.57	0.84	0.24	0.2	0.26	0.33	0.1	-0.34	-0.24	-0.1	0.84	0.39	0.38	0.81	-0.09	-0.39	0.79	0	0	0
cfull	-0.24	0.8	0.35	0.37	-0.13	-0.15	-0.3	0.64	0.16	0.12	0.49	-0.16	0.37	0.71	0.71	0.04	0.55	0.41	0.62	0.37	0	0

3.2 Effectiveness of mean pairwise indices as the surrogates of other multiple-site indices without pairwise calculations

We found that many of the mean pairwise indices were tightly associated to the ones that don't take the average of pairwise calculations, as shown in Fig. 1.

All the four multiple-site indices without pairwise calculations could be well represented by their counterparts of mean pairwise indices, including nestedness, turnover, Lennon's richness and replacement indices. It is worth noting that the richness difference multiple-site index without pairwise calculation was not strongly correlated to its counterpart (Fig. 1 and Table 1). Thus, mean pairwise beta diversity indices proposed in the present work could be used to measure overall beta diversity for the whole community effectively in comparison to other multiple-site indices without iterative calculations of indices for each pair of sites.

3.3 Implications

MBI package should be the first R package available for ecologists to effectively calculate a comprehensive

suite of multiple-site diversity indices concurrently. In the next updated version of the package, it will be considered to integrate the calculation of multiple-site diversity indices at functional and phylogenetic levels.

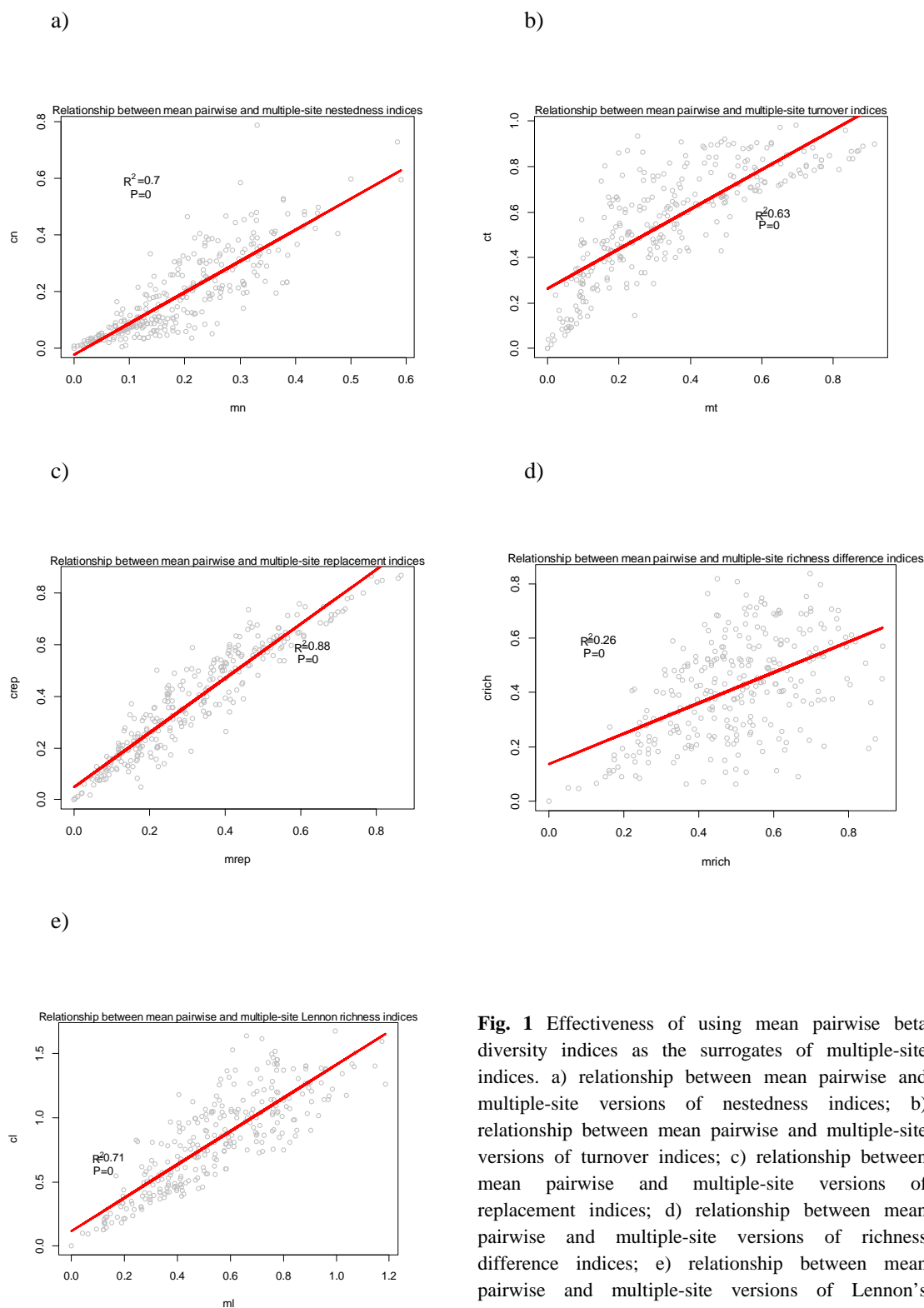


Fig. 1 Effectiveness of using mean pairwise beta diversity indices as the surrogates of multiple-site indices. a) relationship between mean pairwise and multiple-site versions of nestedness indices; b) relationship between mean pairwise and multiple-site versions of turnover indices; c) relationship between mean pairwise and multiple-site versions of replacement indices; d) relationship between mean pairwise and multiple-site versions of richness difference indices; e) relationship between mean pairwise and multiple-site versions of Lennon's richness indices.

MBI package is fully open-source, thus allowing users to modify and add functions to calculate new or self-defined indices *ad hoc*. Also, the power of data handling and statistical analysis in R environment allowed the researchers to effectively handle their own data and calculate the multiple-site beta diversity indices in a quick manner with the assistance of *MRI* package.

Finally, since there is a rapid growing trend to utilize R software for ecological researches, we expect that the program could have potential applications when one wants to perform multiple-site comparison of beta diversity at different spatial scales and across various taxa.

Appendix

Summary of 21 Multiple-site Diversity Indices Currently Available in *MBI* Package

S_i is the number of species found in site i , while S_T total number of species in the community. b_{ij} and b_{ji} are the number of species exclusive to sites i and j respectively when compared by pairs. T is the site number in the community. k_{ij} denotes the number of cells with lower values in column c_j compared to those in column c_i and N_j is the total number of non-empty cells in column c_j . k'_{ij} denotes the number of cells with lower values in row r_j compared to that for the row r_i and N'_j is the total number of non-empty cells in row r_j . N_{pairs} denotes the number of pairs of the sites.

Index	Formula	Function in the <i>MBI</i> package	Reference
Baselga's full index	$\frac{\sum_{i<j} \max(b_{ij}, b_{ji}) + \sum_{i<j} \min(b_{ij}, b_{ji})}{2(\sum_i S_i - S_T) + (\sum_{i<j} \min(b_{ij}, b_{ji})) + (\sum_{i<j} \max(b_{ij}, b_{ji}))}$	<i>cfull(data)</i>	(Baselga et al., 2007)
Baselga's nestedness	$\frac{\sum_{i<j} \max(b_{ij}, b_{ji}) - \sum_{i<j} \min(b_{ij}, b_{ji})}{2(\sum_i S_i - S_T) + \sum_{i<j} \max(b_{ij}, b_{ji}) + \sum_{i<j} \min(b_{ij}, b_{ji})} \times \frac{(\sum_i S_i - S_T)}{(\sum_i S_i - S_T) + \sum_{i<j} \min(b_{ij}, b_{ji})}$	<i>cn(data)</i>	(Baselga et al., 2007; Baselga, 2010)
Baselga's turnover	$\frac{\sum_{i<j} \min(b_{ij}, b_{ji})}{(\sum_i S_i - S_T) + (\sum_{i<j} \min(b_{ij}, b_{ji}))}$	<i>ct(data)</i>	(Baselga et al., 2007; Baselga, 2010)
Replacement	$2 \times \frac{\sum_{i<j} \min(b_{ij}, b_{ji})}{\left(\sum_i S_i - S_T \right) + \sum_{i<j} (b_{ij} + b_{ji})}$	<i>crep(data)</i>	This work; (Carvalho et al., 2012a)
Richness difference	$\frac{\sum_{i<j} \max(b_{ij}, b_{ji}) - \sum_{i<j} \min(b_{ij}, b_{ji})}{2(\sum_i S_i - S_T) + \sum_{i<j} \max(b_{ij}, b_{ji}) + \sum_{i<j} \min(b_{ij}, b_{ji})}$	<i>crich(data)</i>	This work; (Carvalho et al., 2012a)
WNODF index)	$WNODF = \frac{2(WNODFc + WNODFr)}{m(m-1) + n(n-1)}$ (overall index)	<i>wnodf(data)</i>	(Almeida-Neto et al., 2008, 2012; Almeida-Neto and Ulrich, 2011)

$$WNODFr = 100 \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{k'_{ij}}{N'_j} \quad (\text{index for rows})$$

$$WNODFc = 100 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{k_{ij}}{N_j} \quad (\text{index for columns})$$

Whittaker's beta	$\frac{S_T}{\sum_i S_i / T}$	<i>wbeta(data)</i>	(Whittaker, 1960)
Harrison's dissimilarity	$\left(\frac{S_T}{\sum_i S_i / T} - 1\right) / (T - 1)$	<i>harrison(data)</i>	(Harrison et al., 1992)
Diserud-Odegaard's index	$1 - \left(\frac{S_T}{\sum_i S_i / T} - 1\right) / (T - 1)$	<i>do(data)</i>	(Diserud and Ødegaard, 2007)
Harrison's turnover	$\left(\frac{S_T}{\max(S_i)} - 1\right) / (T - 1)$	<i>ht(data)</i>	(Harrison et al., 1992)
Williams's turnover	$1 - \frac{\max(S_i)}{S_T}$	<i>wt(data)</i>	(Williams, 1996)
Lennon richness	$\frac{2\left(\sum_{i<j} \max(b_{ij}, b_{ji}) - \sum_{i<j} \min(b_{ij}, b_{ji})\right)}{2\left(\sum_i S_i - S_T\right) + \sum_{i<j} \max(b_{ij}, b_{ji}) + \sum_{i<j} \min(b_{ij}, b_{ji})}$	<i>cl(data)</i>	This work; (Lennon et al., 2001)
Mean pairwise Jaccard distance	$\frac{1}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \frac{b_{ij} + b_{ji}}{a + b_{ij} + b_{ji}}$	<i>mjaccard(data)</i>	This work
Mean pairwise Sorensen distance	$\frac{1}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \frac{b_{ij} + b_{ji}}{2a + b_{ij} + b_{ji}}$	<i>msorensen(data)</i>	This work; (Sorensen, 1948)
Mean pairwise nestedness	$\frac{1}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \left(\frac{ b_{ij} - b_{ji} }{2a + b_{ij} + b_{ji}} \times \frac{a}{a + \min(b_{ij}, b_{ji})}\right)$	<i>mn(data)</i>	This work; (Baselga, 2010)
Mean pairwise turnover	$\frac{1}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \frac{\min(b_{ij}, b_{ji})}{a + \min(b_{ij}, b_{ji})}$	<i>mt(data)</i>	This work; (Baselga, 2010)
Mean pairwise replacement	$\frac{2}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \frac{\min(b_{ij}, b_{ji})}{a + b_{ij} + b_{ji}}$	<i>mrep(data)</i>	This work; (Carvalho et al., 2012a)
Mean pairwise richness difference	$\frac{1}{N_{pairs}} \sum_{i<j}^{N_{pairs}} \frac{ b_{ij} - b_{ji} }{a + b_{ij} + b_{ji}}$	<i>mrich(data)</i>	This work; (Carvalho et al., 2012a)

Mean pairwise Lennon richness	$\frac{2}{N_{pairs}} \sum_{i < j}^{N_{pairs}} \frac{\max(b_{ij}, b_{ji}) - \min(b_{ij}, b_{ji})}{2a + \min(b_{ij}, b_{ji}) + \max(b_{ij}, b_{ji})}$	<i>ml(data)</i>	This work; (Lennon et al., 2001)
----------------------------------	---	-----------------	-------------------------------------

References

- Almeida-Neto M, Frensel D, 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, Ulrich W. 2011. A straightforward computational approach for measuring nestedness using quantitative matrices *Environmental Modelling and Software*, 26: 173-178
- Atmar W, Patterson B. 1995. The Nestedness Temperature Calculator: A Visual Basic Program, Including 294 Presence/absence Matrices. AICS Research Inc., University Park, NM and The Field Museum, Chicago, IL, USA
- Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19: 134-143
- Baselga A, Orme C. 2012. betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, 3: 808-812
- Baselga, Jimenez-Valverde A, Niccolini G. 2007. A multiple-site similarity measure independent of richness. *Biology Letters*, 3: 642-645
- Carvalho J, Cardoso P, Borges P, et al. 2012a. Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos* (doi: 10.1111/j.1600-0706.2012.20980.x)
- Carvalho J, Cardoso P, Gomes P. 2012b. Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, 21: 760-771
- Oksanen J, Blanchet F, Kindt R, et al. 2012. *vegan: Community Ecology Package*. R package Version 2.0-4. <http://CRAN.R-project.org/package=vegan>.
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <http://www.R-project.org>.
- Zhang WJ. 2011a. A Java algorithm for non-parametric statistic comparison of network structure. *Network Biology*, 1(2): 130-133
- Zhang WJ. 2011b. Constructing ecological interaction networks by correlation analysis: hints from community sampling. *Network Biology*, 1(2): 81-98