APPLICATION

`betapart`: an R package for the study of beta diversity

Andrés Baselga[1]* and C. David L. Orme[2]

[1]*Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, c/Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain; and* [2]*Division of Biology, Department of Life Sciences, Silwood Park, Ascot, Berkshire SL5 7PY, UK*

Summary

1. Beta diversity, that is, the variation in species composition among sites, can be the result of species replacement between sites (turnover) and species loss from site to site (nestedness).

2. We present `betapart`, an R package for computing total dissimilarity as Sørensen or Jaccard indices, as well as their respective turnover and nestedness components.

3. `betapart` allows the assessment of spatial patterns of beta diversity using multiple-site dissimilarity measures accounting for compositional heterogeneity across several sites or pairwise measures providing distance matrices accounting for the multivariate structure of dissimilarity.

4. `betapart` also allows computing patterns of temporal difference in assemblage composition, and its turnover and nestedness components.

5. Several example analyses are shown, using the data included in the package, to illustrate the relevance of separating the turnover and nestedness components of beta diversity to infer different mechanisms behind biodiversity patterns.

Key-words: dissimilarity, distance matrices, multiple-site dissimilarity, nestedness, temporal change, turnover

Introduction

The term 'beta diversity' is applied in a broad sense to any measure of variation in species composition (Anderson *et al.* 2011). In the narrowest sense, it is the simple ratio between gamma and alpha diversities (Jost 2007; Tuomisto 2010; Jurasinski & Koch 2011), which only differs from 1 when local sites differ in species composition. A wide range of broader measures exist (see Anderson *et al.* 2011), including measures of differentiation and proportional diversity (Jurasinski, Retzer & Beierkuhnlein 2009; Jurasinski & Koch 2011), but all broadly aim at providing a measure of the difference between the assemblages present at each site, taking into account the identities of all species. This last characteristic makes beta diversity studies complementary of analyses of the variation in species richness, which ignores species identity. Therefore, compared to species richness, the analysis of beta diversity allows testing of different hypotheses about the processes driving species distributions and biodiversity.

The concept of 'change in species composition' or the question 'how different are two species assemblages' may apparently seem straightforward, but, as argued elsewhere (Baselga 2007, 2010, 2012; Baselga, Jiménez-Valverde & Niccolini 2007), there are two potential ways in which two species assemblages can be 'different'. One is species replacement (i.e. turnover), which consists in the substitution of species in one site by different species in the other site. The second way is species loss (or gain), which implies the elimination (or addition) of species in only one of the sites, and leads to the poorest assemblage being a strict subset of the richest one (a pattern called nestedness). Therefore, the selection of the dissimilarity measure used to quantify the differences between assemblages can be crucial, because different dissimilarity indices account for the two phenomena in different ways. For example, the strict sense definition of beta diversity (the ratio of gamma and alpha diversity: Whittaker 1960; Tuomisto 2010) yields a measure that accounts for turnover and nestedness as being equivalent, as both turnover and nested patterns make alpha diversity lower than gamma diversity. The same applies for the widely used Jaccard and Sørensen indices, which are monotonic transformations of gamma/alpha (Jost 2007; Chao, Chiu & Hsieh 2012). In contrast, the Simpson index of dissimilarity (Simpson 1943; Lennon *et al.* 2001) accounts only for turnover (species replacement), and building on this, Baselga (2010, 2012) proposed a method for partitioning total dissimilarity

*Correspondence author. E-mail: andres.baselga@usc.es

(i.e. Sørensen and Jaccard indices, both monotonic transformations of beta diversity) into two separate components accounting for the dissimilarity derived solely from turnover and the dissimilarity derived from nestedness. The two decompositions for a single pair of cells are shown below for the Sørensen (eqn 1) and Jaccard (eqn 2) indices, where a is the number of shared species between two cells, b the number of species unique to the poorest site and c the number of species unique to the richest site.

$$\beta_{sor} = \beta_{sim} + \beta_{sne} \equiv \frac{b+c}{2a+b+c} = \frac{b}{b+a} + \left(\frac{c-b}{2a+b+c}\right)\left(\frac{a}{b+a}\right) \qquad \text{eqn 1}$$

$$\beta_{jac} = \beta_{jtu} + \beta_{jne} \equiv \frac{b+c}{a+b+c} = \frac{2b}{2b+a} + \left(\frac{c-b}{a+b+c}\right)\left(\frac{a}{2b+a}\right) \qquad \text{eqn 2}$$

where β_{sor} is Sørensen dissimilarity, β_{sim} is Simpson dissimilarity (= turnover component of Sørensen dissimilarity), β_{sne} is the nestedness component of Sørensen dissimilarity, β_{jac} is Jaccard dissimilarity, β_{jtu} is the turnover component of Jaccard dissimilarity, and β_{jne} is the nestedness component of Jaccard dissimilarity. Pairwise dissimilarity between all pairs of sites can be used to investigate spatial patterning of turnover and nestedness-resultant dissimilarity. In addition, multiple-site measures of compositional dissimilarity across a set of sites can be calculated by substituting the multiple-site analogues of the values a, b and c into the two equations (see Baselga 2010, 2012). Note that these are the sum across all pairs of sites for b and c analogues, but not for the shared species a, whose multiple-site analogue is $\left[\sum_i S_i - S_T\right]$, where S_i is the number of species in site i, and S_T is the number of species in the total pool of sites. This component makes the indices real multiple-site measures and not averaged pairwise dissimilarities. We use capital letters to differentiate these multiple-site measures from pairwise measures (Baselga 2012): $\beta_{SOR} = \beta_{SIM} + \beta_{SNE}$ and $\beta_{JAC} = \beta_{JTU} + \beta_{JNE}$.

The betapart package: partitioning beta diversity

This application note introduces an R package, betapart, to compute these dissimilarity measures. The package provides two basic analytical functions (beta.multi and beta.pair), which calculate the multiple-site and pairwise partitions of beta diversity. It also provides a function (beta.sample), which uses randomly selected sites to generate a distribution of multiple-site dissimilarity measures for a given number of sites. Finally, the function beta.temp takes two presence–absence matrices of the same sites and species at two different time steps and computes pairwise turnover components over time within each site. In all functions, total

dissimilarity can be computed as Sørensen or Jaccard dissimilarity, as both are monotonic transformations of strict sense beta diversity (Jost 2007; Tuomisto 2010).

The betapart package is written entirely in the scientific computing language R (R Development Core Team, 2011) and can be installed from the Comprehensive R Archive Network (http://cran.r-project.org/web/packages/betapart/). The raw data accepted by all functions in the package is a matrix (x) codifying the presence (1) or absence (0) of m species (columns) in n sites (rows). The functions provided by the package are as follows:

1. betapart.core(x) computes the basic quantities needed for computing the multiple-site beta diversity measures and pairwise dissimilarity matrices from the presence–absence matrix (x), including pairwise matrices of shared and non-shared species between sites. As these matrices are used in several functions and can be time-consuming to calculate for large matrices, precalculating them using the betapart.core function can markedly improve the speed of subsequent analyses. The function returns a new object of class 'betapart' containing these quantities and which can be used as the input (x) to all remaining functions.

2. beta.multi(x, index.family). This function computes the total dissimilarity across all n sites, along with the turnover and nestedness components of that dissimilarity. The input x may be a presence–absence matrix or a betapart object. The argument index.family selects whether the Sørensen or Jaccard index is used as a measure of total dissimilarity (β_{SOR} or β_{JAC}) and the respective components of turnover (β_{SIM} or β_{JTU}) and nestedness (β_{SNE} and β_{JNE}). The function returns three values, which are the total multi-site dissimilarity across the sites, and its turnover and nestedness components.

3. beta.pair(x, index.family). This function computes the same three dissimilarity metrics as for the previous function and can again be set using the argument index.family to use the Sørensen or Jaccard index of total dissimilarity. Rather than returning three single values as in the previous function, beta.pair returns three matrices containing the pairwise between-site values of each component of beta diversity. The dissimilarity matrices yielded by beta.pair are objects of class dist and can be submitted to further analyses as, for example, Mantel tests, non-metric multidimensional scaling, cluster analysis using other R packages as vegan (Oksanen *et al.* 2011) or cluster (Maechler *et al.* 2005).

4. beta.sample(x, index.family, sites, samples) will resample the three multiple-site dissimilarities for a subset of sites of the original data frame. The number of sites in the subset can be specified along with the number of random samples used to calculate the distribution of dissimilarity measures. The function returns a data frame containing the individual sampled measures along with vectors of the means and standard deviations across samples of each measure.

5. beta.temp(x, y, index.family). This function computes dissimilarity values between matched sites from two data sets (x, y) describing the presence and absence of

species across the same set of sites at two separate times. Again, the index family may be set to use the Sørensen or Jaccard index of total dissimilarity. The function returns a data frame of the three values for the temporal dissimilarity within each site.

Although some of these computations can also be conducted in previously available R packages, that is, vegan (Oksanen *et al.* 2011) and simba (Jurasinski & Retzer 2011), betapart provides a unified framework for the partitioning of total dissimilarity into turnover and nestedness components. The package betapart implements a highly efficient workflow (Fig. 1) to analyse such beta diversity patterns, and the function betapart.core provides an underlying data structure (the 'betapart' object type) that permits efficient implementation for use with large matrices and for performing resampling.

Example analyses

The package contains two data sets containing the presence and absence of 634 species of longhorn beetles (Cerambycidae) in southern (ceram.s, 15 countries) and northern (ceram.n, 19 countries) European countries (see Danilevsky 2007; Baselga 2008 for details). The code below shows how the betapart package can be used to compare the multi-site and pairwise beta diversity components arising from turnover and nestedness, using the default choice of the Sørensen index.

```
library(betapart)
data(ceram.s)
data(ceram.n)
```

```
# get betapart objects
ceram.s.core<-betapart.core(ceram.s)
ceram.n.core<-betapart.core(ceram.n)
# multiple site measures
ceram.s.multi<-beta.multi(ceram.s.core)
ceram.n.multi<-beta.multi(ceram.n.core)
# sampling across equal sites
ceram.s.samp<-beta.sample(ceram.s.core,
sites=10, samples=100)
ceram.n.samp<-beta.sample(ceram.n.core,
sites=10, samples=100)
# plotting the distributions of components
dist.s<-ceram.s.samp$sampled.values
dist.n<-ceram.n.samp$sampled.values
plot(density(dist.s$beta.SOR),
xlim=c(0,0·8), ylim=c(0, 19), xlab='Beta
diversity', main='', lwd=3)
lines(density(dist.s$beta.SNE), lty=1, lwd=2)
lines(density(dist.s$beta.SIM), lty=2, lwd=2)
lines(density(dist.n$beta.SOR), col='grey60',
lwd=3)
lines(density(dist.n$beta.SNE), col='grey60',
lty=1, lwd=2)
lines(density(dist.n$beta.SIM), col='grey60',
lty=2, lwd=2)
# pairwise for south
pair.s<-beta.pair(ceram.s)
# plotting clusters
dist.s<-ceram.s.samp$sampled.values
dist.n<-ceram.n.samp$sampled.values
```
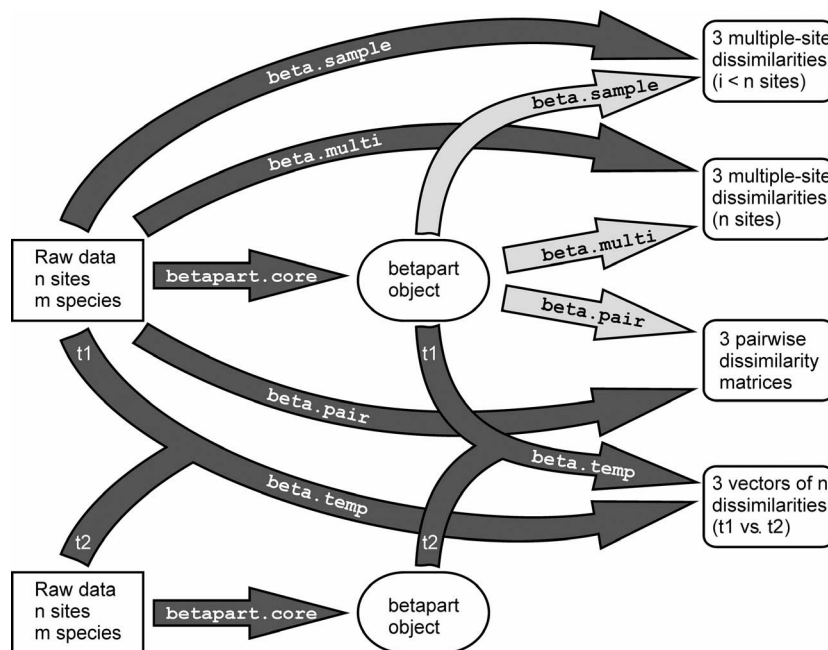

Fig. 1. Workflow of the betapart package. Functions are represented by arrows, and all functions appear twice as they can be applied to the raw data or 'betapart' objects. Arrow tone informs whether applying the function to a 'betapart' object implies saving computational effort (light grey) compared to applying the same function to a raw data table.

```
plot(hclust(pair.s$beta.sim,method="aver-
age"),hang=-1,main='',sub='',xlab='')
title(xlab=expression(beta[ sim]),line=0·3)
plot(hclust(pair.s$beta.sne,method="aver-
age"),hang=-1,main='',sub='',,xlab='')
title(xlab=expression(beta[ sne]),line=0·3)
```

These analyses provide the multiple-site dissimilarities across all sites and the estimated distribution of those values, controlling for the number of sites (Fig. 2a). Note that despite the similar values of total dissimilarity (β_{SOR}) in southern and northern Europe, turnover (β_{SIM}) is much higher in the south and nestedness-resultant dissimilarity (β_{SNE}) is much higher in the north. The pairwise analysis allows the identification of patterns of dissimilarity between Southern European countries derived from turnover and nestedness-resultant dissimilarity (Fig. 2b,c). In addition, the `betapart` package contains data on the presence and absence of 569 bird species across the United States. Raw data were taken from the North American Breeding Bird Survey (http://www.pwrc.usgs.gov/BBS) and have been aggregated to the state level for two time periods (1980–1985 and 2000–2005). This data are used below to explore the temporal differences in species composition of states and the relative contributions of nestedness and turnover (Fig. 2d).

```
data(bbsData)
bbs.t<-beta.temp(bbs1980,bbs2000,index.
family="sor")
# plotting root transformed components
with(bbs.t, plot(sqrt(beta.sim) ~ sqrt(beta.
sne),type='n',ylab=expression(sqrt(beta
[ sim])),xlab=expression(sqrt(beta[ sne]))))
with(bbs.t, text(y=sqrt(beta.sim),x=sqrt
(beta.sne),labels=rownames(bbs1980)))
```

Citation of betapart

Researchers using `betapart` in a published paper should cite this article. Users can also cite the `betapart` package directly. Citation information can be obtained by typing:
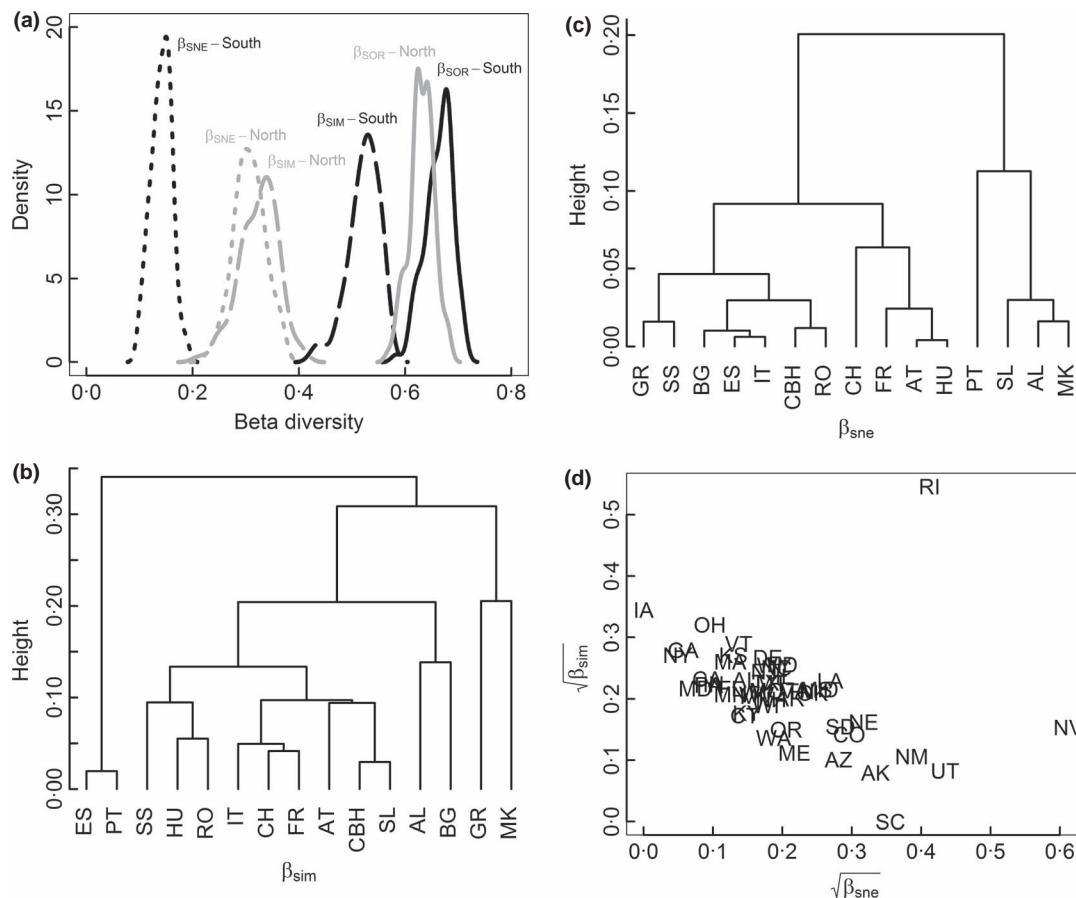
```
citation("betapart")
```


Fig. 2. Use of the *betapart* package. (a) The partition of β_{SOR} (solid line) into β_{SIM} (dashed line) and β_{SNE} (dotted line) for southern (black lines) and northern (grey lines) European countries, using 100 samples of 10 sites from each data set. (b, c) Clustering using average linkage of the β_{sim} and β_{sne} components of species dissimilarity between southern European countries. (d) Comparison of the square root transformed β_{sim} and β_{sne} components of β_{sor} between 1980 and 2000 for US birds by state, with points labelled by state abbreviation.

Acknowledgements

References

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C. & Swenson, N.G. (2011) Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.

Baselga, A. (2007) Disentangling distance decay of similarity from richness gradients: response to Soininen *et al.* 2007. *Ecography*, **30**, 838–841.

Baselga, A. (2008) Determinants of species richness, endemism and turnover in European longhorn beetles. *Ecography*, **31**, 263–271.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2011.00756.x. (in press).

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

Chao, A., Chiu, C.-H. & Hsieh, T.C. (2012) Proposing a resolution to debates on diversity partitioning. *Ecology*, (in press).

Danilevsky, M.L. (2007) A check-list of Longicorn Beetles (Coleoptera, Cerambycoidea) of Europe. http://www.cerambycidae.net/.

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

Jurasinski, G. & Koch, M. (2011) Commentary: do we have a consistent terminology for species diversity? We are on the way. *Oecologia*, **167**, 893–902.

Jurasinski, G. & Retzer, V. (2011) *simba: A Collection of Functions for Similarity Analysis of Vegetation Data. R Package Version 0.3-4*. http://CRAN.R-project.org/package = simba.

Jurasinski, G., Retzer, V. & Beierkuhnlein, C. (2009) Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia*, **159**, 15–26.

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.

Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. (2005) *Cluster Analysis Basics and Extensions. R package*. Available at: http://cran.r-project.org/.

Oksanen, J., Blanchet, G., Kindt, R., Minchin, P.R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2011) *vegan: Community Ecology Package. R package Version 2.0-2*. Available at: http://cran.r-project.org/.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing Version 2.13.1*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org.

Simpson, G.G. (1943) Mammals and the nature of continents. *American Journal of Science*, **241**, 1–31.

Tuomisto, H. (2010) A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, **33**, 2–22.

Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 280–338.