Beta diversity computed as Var(Y)

Description

Compute estimates of total beta diversity as the total variance in a community data matrix **Y**, as well as derived SCBD and LCBD statistics, for 21 dissimilarity coefficients or the raw data table. Computing beta diversity as Var(**Y**) for raw, untransformed community composition data is not recommended. Tests of significance of the LCDB indices are also produced.

Usage

```
beta.div(Y,method="hellinger",sqrt.D=FALSE,samp=TRUE,nperm=999,save.D=FALSE,
        clock=FALSE)
```

Arguments

`Y`	Community composition data. The file class can be either `data.frame` or `matrix`.
`method`	One of 21 dissimilarity coefficients available in the function, or "none". See Details. Names can be abbreviated to a non-ambiguous set of first letters. Default: method="hellinger".
`sqrt.D`	If `sqrt.D=TRUE`, the dissimilarities in matrix **D** are square-rooted before computation of SStotal, BDtotal and LCBD. This transformation may be useful for methods {"manhattan", "modmeanchardiff", "whittaker", "divergence", "canberra", "%difference", “ruzicka”, "wishart"} since square-root transformation of the dissimilarities makes these **D** matrices Euclidean. *Note 1* – Euclideanarity is useful for ordination by principal coordinate analysis; lack of this property does not adversely affect SStotal, BDtotal and LCBD. *Note 2* – The logical value given to parameter `sqrt.D` has no incidence on calculations through methods {"euclidean", "profiles", "hellinger", "chord", "chisquare", "none"} since no **D** matrix is computed in those cases. *Note 3* – For methods {"jaccard", "sorensen", "ochiai"}, the dissimilarity matrix is computed by function dist.binary() of package ade4. That function produces the dissimilarity matrix in the form sqrt(*D*), which is Euclidean.
`samp`	If `samp=TRUE`, the abundance-based distances (ab.jaccard, ab.sorensen, ab.ochiai, ab.simpson) are computed for sample data. If `samp=FALSE`, they are computed for true population data.
`nperm`	Number of permutations for the tests of significance of LCBD indices.
`save.D`	If `save.D=TRUE`, the distance matrix will appear in the output list.
`clock`	If `clock=TRUE`, the computation time is printed. Useful when `nperm` is large.

Details

Calculations may be carried out in two ways, depending on the selected method.

• For untransformed or transformed raw data, the total sum of squares (SStotal) is first computed, then the total beta diversity (BDtotal), which is SStotal divided by (n – 1), is calculated. This algorithm is used for methods {"euclidean", "profiles", "hellinger", "chord", "chisquare", "none"}. No transformation of the data is computed when the method is "euclidean" or "none" (no transformation). For methods "profiles", "hellinger", "chord" and "chisquare", the algorithm begins with computation of the same-name transformation of the community data (Legendre and Gallagher 2001; Legendre and Legendre 2012, Section 7.7); SStotal and BDtotal are then computed for the transformed data.

• Calculations of BDtotal can also be conducted from a dissimilarity matrix. SStotal is computed by summing the squared dissimilarities in the lower triangular dissimilarity matrix and dividing by n; then, total beta diversity (BDtotal) is computed as above. Whith option sqrt.D = TRUE, the computation of SStotal is equivalent to summing the distances instead of the squared distances. Choices are: method = {"manhattan", "modmeanchardiff", "whittaker", "divergence", "canberra", "%difference", “ruzicka”, "wishart", "kulczynski", "ab.jaccard", "ab.sorensen", "ab.ochiai", "ab.simpson", “jaccard”, “sorensen”, “ochiai”}. Equations for these dissimilarities are presented in Table 1 of Legendre and De Cáceres (2013). The Ružička index is described in Legendre (2014); this coefficient is suitable for beta diversity studies. See Chao et al. (2006) for details about the abundance-based (ab) coefficients.

Community composition data could be log-transformed prior to analysis. This transformation makes the distributions more symmetrical. Only the Euclidean distance option should be used with log-transformed data. It is meaningless to subject log-transformed data to the {"profiles", "hellinger", "chord", "chisquare"} transformations available in this function. One can use either the log(y+1 transformation (log1p() function of {base}), or Anderson et al. (2006) special log transformation available in {vegan}: decostand(mat, "log", logbase=10).

The Jaccard, Sørensen and Ochiai coefficients are the binary forms of 10 of the 12 dissimilarity coefficients (including the Ružička index) that are suitable for beta diversity assessment. The equivalences are described in Legendre and De Cáceres (2013, Table 1). These popular coefficients can be computed directly using function beta.div() without going to the trouble of applying the quantitative forms of these coefficients to data reduced to presence-absence form. The transformation to presence-absence is done directly by function dist.binary() of package ade4, which is used by beta.div() to compute these coefficients. That function produces the dissimilarity matrix in the form sqrt(D), which is Euclidean. Hence for these three coefficients, function beta.div() should be used with option sqrt.D=FALSE.

(1) *Species contributions to beta diversity* (SCBD indices for the species) are computed for the untransformed or transformed raw data, but not for dissimilarity matrices. (2) *Local contributions to beta diversity* (LCBD indices) represent the degree of uniqueness of the sites in terms of their species compositions. They can be computed in all cases: raw (not recommended) or transformed data, as well as dissimilarity matrices. See Legendre and De Cáceres (2013) for details.

LCBD indices are tested for significance by random, independent permutations within the columns of **Y**. This permutation method tests H_0 that the species are distributed at random,

independently of one another, among the sites, while preserving the species abundance distributions in the observed data. See Legendre and De Cáceres (2013) for discussion.

Value

Function beta.div returns a list containing the following results:

`SStotal_BDtotal`	Total sum of squares and total beta diversity [= Var(**Y**)] of the data matrix. `BDtotal` statistics computed with the same **D** index are comparable among data sets having the same or different numbers of sampling units (*n*), provided that they are of the same size or represent the same sampling effort.
`SCBD`	Vector of *Species contributions to beta diversity* (SCBD), if computed.
`LCBD`	Vector of *Local contributions to beta diversity* (LCBD) for the sites.
`p.LCBD`	P-values associated with the LCBD indices.
`method`	Method selected.
`note`	Notes indicate whether the selected coefficient is Euclidean or not.
`D`	The distance matrix if `save.D=TRUE`.

For two sites only, the LCBD results are not interesting. With all coefficients, the two LCBD indices are equal to 0.5. The two associated p-values are 1 because LCBD is 0.5 for all columnwise permutations of the data.

The calculation is aborted when the data file only contains two identical rows of data. In that case, SStotal and BDtotal are 0 and the LCBD indices cannot be computed (value NaN).

References

Chao, A., R. L. Chazdon, R. K. Colwell and T. J. Shen. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 62: 361–371.

Legendre, P. 2014. Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography* 23: 1324-1334.

Legendre, P. and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters* 16: 951-963.

Legendre, P. and E. D. Gallagher, E.D. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271–280.

Legendre, P. and Legendre, L. 2012. *Numerical Ecology. 3rd English edition*. Elsevier Science BV, Amsterdam.

Author

Pierre Legendre, Département de sciences biologiques, Université de Montréal.

Example

```
### Example: mite data available in the vegan package
require(vegan)
data(mite)
res = beta.div(mite, "hellinger", nperm=999)

# Plot a map of the LCDB indices
# First, load the file of Cartesian coordinates of the 70 mite sampling sites
data(mite.xy)
plot(mite.xy, asp=1, type="n", xlab="x coordinates (m)", ylab="y coordinates (m)", main="Map
of mite LCBD")
points(mite.xy, pch=21, col="white", bg="brown", cex=120*res$LCBD)

### Example using the mite abundance data and the percentage difference dissimilarity
res = beta.div(mite, "%diff", nperm=999, clock=TRUE)

# Plot a map of the LCDB indices
# First, load the file of cartesian coordinates of the 70 mite sampling sites
data(mite.xy)
signif = which(res$p.LCBD <= 0.05)          # Which are the significant LCDB indices?
nonsignif = which(res$p.LCBD > 0.05)        # Which are the non-significant LCDB indices?
plot(mite.xy, asp=1, type="n", xlab="x coordinates (m)", ylab="y coordinates (m)", main="Map
of mite LCBD (red = significant indices)")
points(mite.xy[nonsignif,], pch=21, col="white", bg="blue", cex=100*res$LCBD[nonsignif])
points(mite.xy[signif,], pch=21, col="white", bg="red", cex=100*res$LCBD[signif])
```