Difference between multivariate observations at T1 and T2

Description

Compute the differences between multivariate observations (frequency or presence-absence data) forming pairs observed at times T1 and T2. Temporal Beta-diversity Indices (TBI) are computed and tested. TBI are dissimilarity indices that measure beta differentiation through time. They are computed separately between T1 and T2 for each site.

Usage

```
TBI <- function(mat1, mat2, method="%difference", pa.tr=FALSE, nperm=99,
       permute.sp=1, BCD=TRUE, replace=FALSE, clock=FALSE)
```

Arguments

`mat1,mat2`	Two multivariate community composition or gene frequency data matrices (class `data.frame` or `matrix`) with the same number of rows and columns. The rows must correspond to the same objects (e.g. sites) and the columns to the same variables, e.g. species or alleles.
`method`	One of the following dissimilarity coefficients: {"%difference", "ruzicka", "hellinger", "chord", "sorensen", "jaccard", "ochiai", "euclidean"}. See Details. Names can be abbreviated to a non-ambiguous set of first letters. Default: method="%difference".
`pa.tr`	If `pa.tr=TRUE`, the data are transformed to binary (i.e. presence-absence, or *pa*) form. If `pa.tr=FALSE`, they are not.
`nperm`	Number of permutations for the tests of significance of the temporal beta indices.
`permute.sp`	The permutation method used in the permutation tests. permute.sp=1: permute the data separately in each column; the corresponding columns in the two matrices are permuted in the same way. permute.sp=2: permute the data separately in each column. Do not force the permutations to start with the same random seed in the two matrices. permute.sp=3: permute entire rows of data in each matrix separately.
`BCD`	If `BCD=TRUE`, the B and C components of the percentage difference and Ružička indices are computed and presented in an output table with three columns: B/*den*, C/*den*, D=(B+C)/*den*, where *den* is the denominator of the index, i.e. (2A+B+C) for the percentage difference index and (A+B+C) for the Ružička index. See **Details** and **Value**. If `pa.tr` is TRUE, the B and C components are the numbers of species lost or gained, and D is either the Sørensen or the Jaccard dissimilarity. In the BCD output table, column B contains B/*den*, C/*den*, D=(B+C)/*den*, as in the case of the percentage difference and Ružička indices.

If `BCD=FALSE`, that table is not produced. No table is (or can be) computed for indices other than the Ružička and percentage difference indices or their binary forms.

`replace` If `replace=FALSE` (default value), sampling is done without replacement, producing a regular permutation test. If `replace=TRUE`, sampling is done with replacement for the test of significance; the method is then bootstrapping.

`clock` If `clock=TRUE`, the computation time is printed. This option is useful to predict the calculation time when *n* and `nperm` are large.

Details

For each object, the function tests the hypothesis (H_0) that the *difference* between T1 and T2 for that object belongs to the same statistical population as the differences displayed by the other objects in the data files. If H_0 is rejected, the object is recognized as exceptionally different from the other objects for its difference between T1 and T2.

To fix ideas, an example in palaeoecology — A researcher is studying ancient and modern diatom communities in sediment cores. If a site displays an exceptional difference between T1 and T2, the researcher is justified to examine the reason for that difference. It could, for example, be caused by a change in land use at that site, which has caused the difference to be larger than at the other sites, compared to the differences caused by climatic change at all sites.

The temporal beta diversity indices available in this function belong to four groups, computed in different ways.

• Method "%difference" computes the percentage difference index, called the Bray-Curtis index in some software; it is the quantitative form of the Sørensen index. Method "ruzicka" computes the Ružička dissimilarity; this is one of the quantitative coefficients corresponding to the Jaccard dissimilarity for binary data. These indices are obtained by computing a dissimilarity function. When these indices are used to compute ordinations by principal coordinate analysis, it is recommended to take the square root of the dissimilarities before the ordination analysis because these indices do not have the Euclidean property. However, that precaution is not important here; the results of permutation tests will be the same for these dissimilarities square-rooted or not. If pa.tr=TRUE, either the Sørensen or the Jaccard coefficient are obtained by computing these two coefficients.

• Methods "hellinger" (Hellinger distance) and "chord" (chord distance) are obtained by transformation of the species data, as recommended by Legendre & Gallagher (2001), followed by calculation of the Euclidean distance. The transformations are carried out through function `decostand` of the vegan package. These two distances have the Euclidean property (Legendre & Legendre 2012, Legendre & De Cáceres 2013). If pa.tr=TRUE, the binary distance sqrt(2)*sqrt(1–Ochiai similarity) is obtained from these two coefficients.

• Methods {"jaccard", "sorensen", "ochiai"} implement the Jaccard, Sørensen and sqrt(2)*sqrt(1 – Ochiai similarity) dissimilarities. For these coefficients, the data are first transformed to presence-absence form (`pa.tr` is given the value `TRUE`), then the dissimilarities

are computed using the corresponding quantitative coefficients (Ružička, percentage difference and Hellinger).

• The Euclidean distance is also available in this function. It is not recommended for community composition or allele frequency data. One can compute it for log-transformed abundance data that do not contain zeros, or very few zeros (short gradients).

The temporal beta indices are tested for significance using permutation tests. The hypotheses are the following:
• H_0: the site under study (e.g. a species assemblage) is not exceptionally different between T1 and T2, compared to other sites that have been observed at the same two times. The differences between T1 and T2 al belong to the same statistical population of differences.
• H_1: the site under study is exceptionally different between times T1 and T2.

In the decomposition of the Ružička and percentage difference dissimilarities or their presence-absence forms (Jaccard, Sørensen), the components B and C are computed as follows:

• b_j is the part of the abundance of species j that is higher at time 1 than at time 2: $b_j = y_{1j} - y_{2j}$. B is the sum of the b_j values for all species in the group of species under study. It is the unscaled sum of **species losses** between time 1 and time 2. In the BCD output table, column B contains B/*den* where *den* is the denominator of the index, i.e. (2A+B+C) for the percentage difference index and (A+B+C) for the Ružička index.

• c_j is the part of the abundance of species j that is higher at time 2 than at time 1: $c_j = y_{2j} - y_{1j}$. C is the sum of the c_j values for all species in the group of species under study. It is the unscaled sum of **species gains** between time 1 and time 2. In the BCD output table, column C contains C/*den* where *den* is the denominator of the index, i.e. (2A+B+C) for the percentage difference index and (A+B+C) for the Ružička index.

Value

Function `TBI` returns a list containing the following results:

`TBI`	The vector of Temporal Beta-diversity Indices (TBI) between observations at times T1 and T2 for each object.
`p.TBI`	A corresponding vector of p-values. Significant p-values (e.g. p.dist ≤ 0.05) indicate exceptional objects for the difference of their species composition.
`p.adj`	The p-values are corrected for multiple testing using function `p.adjust` of {stats}. The adjustment is done using method="holm", which is the default option of the `p.adjust` function.
`BCD.mat`	An output table with three columns: B/*den*, C/*den*, D=(B+C)/*den*, where *den* is the denominator of the index, i.e. (2A+B+C) for the percentage difference index and (A+B+C) for the Ružička index. The decomposition is such that D = B/*den* + C/*den*. Columns B and C can help users determine which of the significant D values are associated with large B (species losses) or large C values (species gains), before proceeding to the analysis and interpretation of the D values, using environmental or spatial explanatory variables, through regression or classification tree analysis.

If `pa.tr` is TRUE, the B and C components are the numbers of losses and gains of species, and D is either the Sørensen or the Jaccard dissimilarity.
If `BCD=FALSE`, that table is not produced. No table is (or can be) computed for indices other than the Ružička and percentage difference indices or their binary forms.

References

Legendre, P. A temporal beta-diversity index to identify exceptional sites in space-time surveys. *Manuscript* (to be submitted).

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

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

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

van den Brink, P. J. & C. J. F. ter Braak. 1999. Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* 18: 138–148.

Author

Pierre Legendre, Département de sciences biologiques, Université de Montréal.
License: GPL(>=2).

Example

Invertebrate communities subjected to insecticide treatment.

As an example in their paper on Principal Response Curves (PRC), van den Brink & ter Braak (1999) used observations on the abundances of 178 invertebrate species (macroinvertebrates and zooplankton) subjected to treatments in 12 mesocosms by the insecticide chlorpyrifos. The mesocosms were sampled at 11 occasions. The data, available in the {vegan} package, are log-transformed species abundances, $y_{tranformed} = \log_e(10*y+1)$.

The data of survey #4 will be compared to those of survey #11 in this example. Survey #4 was carried out one week after the insecticide treatment, whereas the fauna of the mesocosms was considered by the authors to have fully recovered from the insecticide treatment at survey #11.

```
require(vegan)
data(pyrifos)
```

```
# The mesocosms had originally been attributed at random to the treatments. However, to facilitate presentation of the results, they will be listed here in order of increased insecticide doses: {0, 0, 0, 0, 0.1, 0.1, 0.9, 0.9, 6, 6, 44, 44} µg/L.

survey4.order = c(38,39,41,47,37,44,40,46,43,48,42,45)
survey11.order = c(122,123,125,131,121,128,124,130,127,132,126,129)

# Results using abundance data, percentage difference dissimilarity
res1 <- TBI(pyrifos[survey4.order,], pyrifos[survey11.order,], method="%diff", nperm=999,
permute.sp=1, BCD=TRUE, clock=TRUE)

# Results using presence-absence data, Sørensen dissimilarity
res2 <- TBI(pyrifos[survey4.order,], pyrifos[survey11.order,], method="sorensen", nperm=999,
permute.sp=1, BCD=TRUE, clock=TRUE)

# Identical results (Sørensen dissimilarity) are obtained with
res3 <- TBI(pyrifos[survey4.order,], pyrifos[survey11.order,], method="%diff", pa.tr=TRUE,
nperm=999, permute.sp=1, BCD=TRUE, clock=TRUE)
```