

# The betaper Package

June 9, 2008

**Type** Package

**Title** Distance decay of similarity among biological inventories in the face of taxonomic uncertainty

**Version** 1.0-0

**Date** 2008-06-08

**Author** Luis Cayuela and Marcelino de la Cruz

**Maintainer** Luis Cayuela <luis.cayuela@uah.es>

**Depends** vegan

**Description** Evaluates and quantifies distance decay of similarity among biological inventories in the face of taxonomic uncertainty

**License** GPL>=2

**LazyLoad** yes

## R topics documented:

Amazonia . . . . .	1
HCP . . . . .	2
betaper . . . . .	3
plot.betaper . . . . .	5
<b>Index</b>	<b>8</b>

---

Amazonia

*Tree abundance and soil data in Western Amazonia*

---

### Description

The `Amazonia` data frame has tree counts in nine 0.16-hectare inventory plots in Western Amazonia. `soils` contains data on soil cations at each location.

### Usage

```
data(Amazonia)
data(soils)
```

### Format

`Amazonia` is a data frame with 1188 observations (species) and 12 columns (taxonomic description and sites). The three first columns refer to family, genus and specific species Latin names. Columns 4 to 12 have tree abundance data for nine inventory plots.

`soils` is a data frame with 9 observations (inventory plots) and 4 columns (variables). Soil variables (Ca, K, Mg, Na) are given in cmol/kg.

### Details

Data from Western Amazonia includes tree inventories at nine lowland sites (approximately 100-150 m above sea level) near Iquitos, Peru. The sites were selected to represent regional variations in geology and were distributed along a soil nutrient gradient ranging from poor loamy soils to richer clayey soils. Each inventory consisted of 20 x 20 m plots (0.16 ha total area) distributed along 1.3-km transects. At each site, K. Ruokolainen and colleagues identified to species or morphospecies all woody, free-standing stems of > 2.5 cm dbh. The full inventories sampled 3980 individuals from 1188 species or morphospecies.

### References

Higgins, M.A. & Ruokolainen, K. 2004. Rapid tropical forest inventory: a comparison of techniques based on inventory data from western Amazonia. *Conservation Biology* 18(3): 799-811.

Ruokolainen, K., Tuomisto, H., Macia, M.J., Higgins, M.A. & Yli-Halla, M. 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology* 23: 13-25.

### Examples

```
data(Amazonia)
data(soils)
```

---

HCP

*Tree counts in tropical montane forest fragments*

---

### Description

HCP has tree abundance data from 16 forest fragments located in the Highlands of Chiapas, southern Mexico. `HCP.coords` contains the geographical UTM coordinates for the 16 forest fragments' centroids.

### Usage

```
data(HCP)
data(HCP.coords)
```

### Format

HCP is a data frame with 231 observations and 19 variables. The three first columns contain family, genus and specific species Latin names. Columns 4 to 19 have tree abundance data for the 16 forest fragments. `HCP.coords` is a data frame with two columns and 16 rows.

### References

Cayuela, L., Golicher, D.J., Rey Benayas, J.M., Gonzalez-Espinosa, M. & Ramirez-Marcial, N. 2006. Fragmentation, disturbance and tree diversity conservation in tropical montane forests. *Journal of Applied Ecology* 43: 1172-1181.

### Examples

```
data(HCP)
data(HCP.coords)
```

---

betaper

*Distance decay of similarity among floristic inventories in the face of taxonomic uncertainty*

---

### Description

Function `betaper` computes the uncertainty range of the estimated correlation between floristic and geographical or environmental distances in the face of taxonomic uncertainty.

### Usage

```
betaper(data, geodist, index = NULL, nsim = 100, vegdist.method = "bray", binary =
```

## Arguments

<code>data</code>	Community data matrix. The three first columns are factors referring to the family, genus and species specific names. The remaining columns are numeric vectors indicating species abundances at each site.
<code>geodist</code>	Dissimilarity matrix or a dist object of geographical or environmental distances.
<code>index</code>	List of additional parameters to determine the level at which species have been identified. Default values include 'Indet', 'indet', 'sp', 'sp1' to 'sp100', 'sp 1' to 'sp 100', ', ' and ' '.
<code>nsim</code>	Number of simulations of species' identities in assessing bounds to the estimated correlation.
<code>vegdist.method</code>	Dissimilarity index, as in <code>vegdist</code> . Partial match to "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "morisita", "horn", "mountford", "raup", "binomial" or "chao".
<code>binary</code>	Perform presence/absence standardization before calculation of vegetation distances using <code>vegdist</code> function.
<code>cor.method</code>	Correlation method, as accepted by <code>cor.test</code> : "pearson", "spearman" or "kendall".

## Details

The method describes the manner to incorporate taxonomic uncertainty to the calculation of the correlation coefficient that measures the degree of correspondence between floristic and geographical or environmental distances, by estimating some credible bounds under plausible scenarios of re-assigned species identities. The approach involves the following steps:

Step 1. Codification of the species according to their degree of identification. Identified species are codified as 0, morphospecies identified only to genus (e.g. *Eugenia* sp1) as 1, only to family (e.g. *Myrtaceae* sp1) as 2, and fully unidentified morphospecies (e.g. *Undetermined* sp1) are codified as 3.

Step 2. Morphospecies identified only to genus are then randomly re-assigned in the group of species and morphospecies that share the same genus provided they are not found in the same sites. In the re-assignment of the species identity, the species can also get its own identity.

Step 3. Once the re-assignment of morphospecies has been accomplished, we calculate a floristic distance matrix. A distance matrix is a square, symmetric matrix in which rows and columns are sites and cells are the floristic distance between pairs of sites. A single distance matrix thus represents the floristic pattern among sites for a single set of inventories. The Bray-Curtis' or Sorensen's coefficient is used by default to calculate floristic distance between sites. Note that although dissimilarity indexes are calculated with the `vegdist` function, function `betaper` calculates the complementary similarity index (i.e. 1 - dissimilarity index).

Step 4. The Pearson's correlation coefficient is computed between the floristic distance matrix and the geographical or environmental distance matrix.

Step 5. Steps 2 to 4 are iterated  $n$  times. The process can be time-consuming if the floristic matrix is large. The mean and standard deviation, and the minimum and maximum values of the Pearson correlation coefficient,  $r$ , are then estimated. These measures provide credible bounds to the true correlation coefficient in the face of taxonomic uncertainty.

**Value**

The function return a list of class `betaper` with the following components

<code>perm</code>	Permutated floristic distance matrices coerced to columns in a data frame. As many permutations (columns) as <code>nsim</code> .
<code>cordis</code>	Pearson's correlation values ( $r$ ) between simulated floristic distances and geographical or environmental distances.
<code>taxunc</code>	Summary of the number of species fully identified (0), identified to genus (1), identified to family (2), or fully undetermined (3).
<code>geodist</code>	Dissimilarity matrix of geographical or environmental distances.
<code>cond</code>	Numeric vector indicating the assignment of species to any of the <code>taxunc</code> categories.
...	

**Note**

The method uses floristic similarities instead of floristic dissimilarities, as calculated by `vegdist` function. The method simply computes floristic similarities by subtracting the value output by `vegdist` to one

**Author(s)**

Luis Cayuela (luis.cayuela@uah.es) and Marcelino de la Cruz (marcelino.delacruz@upm.es)

**See Also**

[mantel](#) for computation of the Mantel statistic, [vegdist](#) for methods to calculate vegetation distances. See also [bioenv](#) for methods to select environmental variables.

**Examples**

```
## Not run:
require(vegan)
#####
# Tree data from 16 0.1-ha plots from the Highlands of Chiapas, southern Mexico

data(HCP)
data(HCP.coords)

geodist.HCP <- dist(HCP.coords)/1000 # Geographical distances in kms

# Define a new index that includes the terms used in the HCP dataset to define undetermined
index.new <- c(paste("Genus", as.character(c(1:20))), sep=""), paste("Family", as.character(c(1:20))), sep="")

beta.HCP <- betaper(HCP, geodist = geodist.HCP, index = index.new)
beta.HCP # This results in a relatively low uncertainty range of the estimated correlation
beta.HCP$taxunc # Number of identified and unidentified species at different taxonomic levels

#####
```

```
# Tree data from nine lowland sites in Western Amazonia

data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the Amazonia dataset to define undetermined
index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

beta.Amazonia <- betaper(Amazonia, geodist = dist(log(soils)), index = index.Amazon, nsim = 100)
beta.Amazonia # The uncertainty range of the estimated correlation is much larger than in the
beta.Amazonia$taxunc # Number of identified and unidentified species at different taxonomic levels
## End(Not run)
```

---

plot.betaper	<i>Plot of distance decay of similarity among floristic inventories in the face of taxonomic uncertainty</i>
--------------	--

---

## Description

Plot of maximum and minimum correlation bounds between floristic and geographical or environmental distances under simulated scenarios of species identities in the face of taxonomic uncertainty

## Usage

```
plot.betaper(x, xlab = "Distances (km)", ylab = "Sorensen's similarity index", pch = 1)
```

## Arguments

x	a betaper object.
xlab	x-axis label.
ylab	y-axis label.
pch	A vector of the form c(x1, x2) specifying a symbol or a single character to represent the points that minimize and maximize the correlation between floristic and geographical or environmental distances.
...	any other graphical parameter.

## Details

This plot function produces a quick, standard plot with estimated minimum and maximum correlation bounds between floristic and geographical or environmental distances.

## Author(s)

Luis Cayuela <luis.cayuela@uah.es> and Marcelino de la Cruz <marcelino.delacruz@upm.es>

## See Also

[betaper](#)

**Examples**

```
## Not run:
require(vegan)
#####
# Tree data from 16 0.1-ha plots from the Highlands of Chiapas, southern Mexico

data(HCP)
data(HCP.coords)

geodist.HCP <- dist(HCP.coords)/1000 # Geographical distances in kms

# Define a new index that includes the terms used in the HCPdataset to define undetermined t
index.new <- c(paste("Genus", as.character(c(1:20))), sep=""), paste("Family", as.character(c

beta.HCP <- betaper(HCP, geodist = geodist.HCP, index = index.new)
beta.HCP # This results in a relatively low uncertainty range of the estimated correl
beta.HCP$taxunc # Number of identified and unidentified species at different taxonomic level

plot(beta.HCP)

#####
# Tree data from nine lowland sites in Western Amazonia

data(Amazonia)
data(soils)

# Define a new index that includes the terms used in the Amazonia dataset to define undeterm
index.Amazon <- c(paste("sp.", rep(1:20), sep=""), "Indet.", "indet.")

beta.Amazonia <- betaper(Amazonia, geodist = dist(log(soils)), index = index.Amazon, nsim =
beta.Amazonia # The uncertainty range of the estimated correlation is much larger than in
beta.Amazonia$taxunc # Number of identified and unidentified species at different taxonom

plot(beta.Amazonia)
## End(Not run)
```

# Index

## \*Topic **datasets**

Amazonia, [1](#)

HCP, [2](#)

## \*Topic **hplot**

plot.betaper, [5](#)

## \*Topic **htest**

betaper, [3](#)

## \*Topic **multivariate**

betaper, [3](#)

Amazonia, [1](#)

betaper, [3](#), [6](#)

bioenv, [5](#)

cor.test, [3](#)

HCP, [2](#)

mantel, [5](#)

plot.betaper, [5](#)

print.betaper(*betaper*), [3](#)

soils(*Amazonia*), [1](#)

vegdist, [3](#), [5](#)