# Package 'DarkDiv'

July 15, 2020

**Type** Package

**Title** Estimating Dark Diversity and Site-Specific Species Pools

**Version** 0.3.0

**Author** Carlos P. Carmona and Meelis Partel

**Maintainer** Carlos P. Carmona <perezcarmonacarlos@gmail.com>

**Description** Estimation of dark diversity and site-specific species pools using species co-occurrences.
It includes implementations of probabilistic dark diversity based on the
Hypergeometric distribution, as well as estimations based on the Beals index,
which can be transformed to binary predictions using different thresholds,
or transformed into a favorability index. All methods include the possibility of
using a calibration dataset that is used to estimate the indication matrix
between pairs of species, or to estimate dark diversity directly on a single
dataset. See De Caceres and Legendre (2008) <doi:10.1007/s00442-008-1017-y>,
Lewis et al. (2016) <doi:10.1111/2041-210X.12443>,
Partel et al. (2011) <doi:10.1016/j.tree.2010.12.004>, Real et al. (2017) <doi:10.1093/sysbio/syw072>
for further information.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.1.1

**Imports** vegan, stats

**Suggests** testthat

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-07-15 09:00:02 UTC

## R topics documented:

---

`DarkDiv`                              *Dark Diversity*

---

**Description**

Estimates dark diversity based on species co-occurrences using different methods.

**Usage**

```
DarkDiv(x, r = x, method = "Hypergeometric", limit = "min",
  const = 0.01, removeAbsent = T, wa = F, weights = NULL)
```

**Arguments**

| | |
|---|---|
| x | Study data with sites in rows and species in columns. Make sure that names of species are written in colnames(x) and that they coincide with the names given in 'r' (in case it is provided). |
| r | Dataset for reference, it can be either a matrix or data.frame with sites in rows and species in columns, or an indication matrix (a matrix containing the indication values between pairs of species estimated using some of the available methods). In case r is provided, the function will use it to estimate the indication matrix that will be later applied to predict dark diversity in 'x'. In case 'r' is not provided, the function will automatically use 'x' to estimate the indication matrix. |
| method | A character to choose between "Hypergeometric", "RawBeals", "Threshold-Beals" and "Favorability". This parameter determines which method is used to estimate dark diversity (see "Details" below). Defaults to "Hypergeometric". |
| limit | A character to choose between "quantile", "min", "const" and "outlier" indicating the method to choose which limit to apply to the thresholded Beals method (see "Details" below). Defaults to "min". |
| const | constant for limit (as quantile or as minimal) in the ThresholdBeals method. Defaults to 0.01. |
| removeAbsent | Logical indicating what to do with species with zero occurrences in the indication matrix (i.e. for which no indication values can be estimated). removeAbsent = TRUE indicates that these species should be removed from results (giving dark diversity and pool matrices whose dimensions might not coincide with x). removeAbsent = FALSE indicates that these columns will be kept in the results, but filled with NAs. Default to TRUE. |
| wa | Logical indicating whether abundance should be considered for estimations of dark diversity. Defaults to FALSE. If wa = T, abundance weighted values are given based on the values in 'weights', or in 'x' in case 'weights' is not provided. |
| weights | Matrix or data.frame with sites in rows and species in columns including the weights that will be used in case wa is set to TRUE. |

## Details

The species pool of a site encompasses all the species in the region that could potentially inhabit those particular ecological conditions (Zobel 2016). Generally, not all the species from the species pool are realised into local communities. A more or less large proportion of the species in the pool are not present; this unobserved part of the pool represents the "dark diversity" of a site (Partel et al. 2011). Dark diversity is, by definition, unobservable, at least with absolute certainty, and can only be estimated. Because of this, dark diversity can be defined as a fuzzy set, where the degree of certainty about a species membership is expressed as a probability. There are different methods to estimate this subset of species, including the use of species co-occurrence patterns (Lewis et al. 2016; de Bello et al. 2016).

The `DarkDiv` package implements estimations of dark diversity based on species co-occurrence patterns, including both probabilistic and binary estimations. The methods included in the package are:

*Hypergeometric*: the method is based in comparing, for each pair of species, their realised number of co-occurrences with the one that would be expected at random (i.e. if there was no association between species). The expected number of co-occurrences is given by the mean value of the hypergeometric distribution. The difference between observed and expected co-occurrences is then expressed as a probability, and an indication matrix is created showing the strength of the association between pairs of species. Subsequently, probabilities of belonging to local dark diversity are assigned to absent species based on the average indication values of the species that have been observed in each site.

*RawBeals*: Beals values reflect the probability that a species *occurs* at a site. Consequently Beals values are strongly and positively correlated with the frequency of the species in the region. Beals values have been used to estimate species that belong to species pools (Ewald 2002; Karger et al. 2016). However, although the Beals index is a good predictor of the probability of occurrence of the target species (De Caceres & Legendre 2008), it is not necessarily such a good predictor of their suitability in a given site, which is what a probabilistic indicator of dark diversity should return.

*Favorability*: The favorability correction (Real et al. 2017) can be applied to indices affected by species prevalence to turn them into pure indicators of the suitability of the local conditions for each particular species. Effectively, this informs on whether a species is more or less likely to be found in a site than random expectations (i.e. regardless its presence/absence ratio in the dataset). In `DarkDiv`, Favorability first estimates *RawBeals* values, and then applies the favorability correction.

*ThresholdBeals*: Thresholds can be applied to raw Beals values to transform the probabilities for each species in each site into a binary presence/absence indication. Species with Beals values below the threshold are given a 0 probability of belonging to local dark diversity, and 1 otherwise. The thresholds are based on different criteria, given by the argument `limit`: `limit = "quantile"` uses a percentile (given by `const`) of the distribution of Beals values of the sites in which the species is actually present into 0; `limit = "min"` estimates probabilities for the sites in which the species is actually present, and set the minimum of these values as the threshold; `limit = "const"` uses a constant (given by `const`) into 0; finally, `limit = "outlier"` sets the threshold as `quantile(x,probs = 0.25) -1.5 * IQR(x)`, where *x* is the distribution of Beals values of the sites where the species is present.

## Value

`DarkDiv` returns a list containing the following components:

*indication:* A square matrix (species x species) containing the indication values for all pairs of species for which there is at least one occurrence in the data (*r*, or *x* in case *r* is not given). The indication matrix contains the indicator value of each species –in columns– for all other (target) species –in rows– and it is estimated using whatever method is specified. However, if *r* is given already in the form of an indication matrix, then *indication* contains those values.

*AllProbs:* A matrix, with the same dimensions as *x*, with sites in rows and species in columns. Each cell contains the value given by each method (a probability for *Hypergeometric*, *RawBeals* and *Favorability*, or presence/absence for *ThresholdBeals* for all species, regardless whether they were present or absent in the site.

*Pool:* A matrix, with the same dimensions as *x*, with sites in rows and species in columns. Each cell contains the value given by each method (a probability for *Hypergeometric*, *RawBeals* and *Favorability*, or presence/absence for *ThresholdBeals*. Species that were present in each site (i.e. with positive abundance in *x*) are automatically assigned a 1 in this matrix, since present species are assumed to be part of the local species pool.

*Dark:* A matrix, with the same dimensions as *x*, with sites in rows and species in columns. Each cell contains the value given by each method (a probability for *Hypergeometric*, *RawBeals* and *Favorability*, or presence/absence for *ThresholdBeals*. Species that were present in each site (i.e. with positive abundance in *x*) are automatically assigned a NA in this matrix, since it does not make sense to estimate if a species actually observed in a site is part of its dark diversity.

## References

de Bello, F., et al. (2016). Measuring size and composition of species pools: a comparison of dark diversity estimates. Ecology and Evolution, 6(12), 4088-4101.

De Caceres, M., & Legendre, P. (2008). Beals smoothing revisited. Oecologia, 156(3), 657-669.

Ewald, J. (2002). A probabilistic approach to estimating species pools from large compositional matrices. Journal of Vegetation Science, 13(2), 191-198.

Karger, D. N., et al. (2016). Delineating probabilistic species pools in ecology and biogeography. Global Ecology and Biogeography, 25(4), 489-501.

Lewis, R. J. et al. (2017). Applying the dark diversity concept to nature conservation. Conservation Biology, 31(1), 40-47.

Partel, M. et al. (2011). Dark diversity: Shedding light on absent species. Trends in Ecology and Evolution, 26(3), 124-128.

Real, R., et al. (2017). Species distributions, quantum theory, and the enhancement of biodiversity measures. Systematic Biology, 66(3), 453-462.

Zobel, M. (2016). The species pool concept as a framework for studying patterns of plant diversity. Journal of Vegetation Science, 27(1), 8-18.

## Examples

```
#Compute dark diversity with the Hypergeometric method
require(vegan)
data(dune)
ddHyper <- DarkDiv(x = dune, method = "Hypergeometric")
#Compute dark diversity with the Beals method
```

```
ddBeals <- DarkDiv(x = dune, method = "RawBeals")
#Compute favorability using directly the indication matrix from 2.
ddFavor1 <- DarkDiv(x = dune, r = ddBeals$indication, method = "Favorability")
#Compute dark diversity with the Favorability method, and compare with 3a
ddFavor2 <- DarkDiv(x = dune, method = "Favorability")
identical(ddFavor1, ddFavor2)
```

# Index

DarkDiv,