

Package ‘biomapME’

September 2, 2020

Type Package

Title Biodiversity Mapping and Macroecology

Version 2.0

Date 2020-06-24

Author Greg R. Guerin

Maintainer Greg Guerin <ggueri01@gmail.com>

Depends R (>= 3.1.0), raster

Suggests

Imports geosphere,

simba,

SDMTools,

ape,

pracma,

maps,

vegan,

adehabitatHR,

dismo,

sp,

grDevices,

stats,

BAT,

PhyloMeasures

Description R functions for generating gridded biodiversity maps from point (incidence) data.

License GPL-3 + file LICENSE

LazyData TRUE

R topics documented:

biomapME	2
buffer.frag	3
convert.site.data	5
endemism.null.test	6
map.multi.metrics	8
map.pa.matrix	10
pe.null.test	12
phylogenetic.diversity.sites	14

phylogenetic.endemism	15
spp.range.metrics	19
weighted.endemism	21

Index	25
--------------	-----------

biomapME	<i>Biodiversity Mapping and Macroecology</i>
----------	--

Description

The biomapME package contains a range of R functions for manipulating, analysing and rasterising ecological data for use in macroecological analysis and mapping of biodiversity metrics. The functions are primarily geared towards generating (and analysing) gridded biodiversity maps based on various metrics from species incidence (and environmental) data.

Details

Package: biomapME
 Type: Package
 Version: 2.0
 Date: 2020-08-26
 License: GPL-3

biomapME is currently under active development. Additional metrics/functions/functionality are expected over time.

The main purpose of the package is to provide self-contained functions in the R environment for taking simple input data (i.e., georeferenced presence or plot-based species records) and calculating important biodiversity metrics, which are returned as automatically generated raster maps or in numeric formats.

Functions in the ausplotsR package:

At present, biomapME contains functions for calculating:

[buffer.frag](#) : Batch-processed habitat fragmentation/class statistics for circular buffers around focal sites.

[convert.site.data](#) : Converting plot based species occurrence matrices into individual species records.

[endemism.null.test](#) : Non-parametric tests as to whether observed endemism is higher than expected.

[map.multi.metrics](#) : Metrics based on sets of multiple presence/absence plots within map grid cells: 1. Nonparametric estimators of species richness. 2. Multi-site beta diversity. 3. Estimated phylogenetic diversity based on completeness of species sampling.

[map.pa.matrix](#) : Presence-absence matrices for species occurrences within map grid cells.

`pe.null.test` : Non-parametric tests as to whether observed phylogenetic diversity/endemism is higher than expected.

`phylogenetic.diversity.sites` : Phylogenetic diversity for a set of sample sites.

`phylogenetic.endemism` : Phylogenetic diversity, Phylogenetic range rarity and Georeferenced phylogenetic endemism.

`spp.range.metrics` : Various range metrics for batch-processed multiple species (i.e, number of occupied grid cells, extent of occurrence (range span, range area), longitudinal/latitudinal range).

`weighted.endemism` : Species richness, Range rarity richness and Georeferenced weighted endemism.

Author(s)

Greg R. Guerin

References

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

Guerin, G.R. & Lowe, A.J. (2015) Sum of inverse range-sizes ('SIR'), a biodiversity metric with many names and interpretations. *Biodiversity and Conservation* **24**, 2877-2882.

buffer.frag

Batch fragmentation/class stats for circular buffers

Description

Batch calculates habitat fragmentation indices ('fragstats') within a circular buffer zone around sites.

Usage

```
buffer.frag(XY, radius, vegetation.base.raster, plot=TRUE)
```

Arguments

XY	The x/longitude and y/latitude coordinates of a set of sites provided in either of the following formats. If the coordinates are in columns of a data.frame, then the subsetting data.frame should be given: e.g. XY = df[,c("Longitude", "Latitude")] if there are additional columns. If the coordinates are within a SpatialPoints-DataFrame, the SPDF can be given for XY (the function will locate the coordinates).
radius	The radius of the buffer zone around each site within which fragstats are calculated, in units of metres if the 'vegetation.base.raster' is in longlat.
vegetation.base.raster	A RasterLayer object containing integer classes for the presence of vegetation/habitat/land classes of 1 or more kind.
plot	Logical. If TRUE, the input sites and the requested buffers are overplotted on the raster.

Details

This function is a wrapper for the `ClassStat` function from package `SDMTools`, which calculates fragstats for a given raster representing a vegetation/habitat matrix. The `ClassStat` function is applied to a user defined circular buffer area surrounding a site/coordinate, and these are batch processed for a set of multiple sites.

Value

Returns a list of data frames, one for each focal location, with fragstat indices as columns and vegetation/habitat classes (read from the input raster) as rows. See example below for collapsing the list to a single data.frame object. By default, NA values within the buffer (generally representing non-habitat or non-vegetation) are returned with the value 100 for 'class'. 'Class' will otherwise be the same as the categories in the habitat raster (See `?ClassStat`)

Author(s)

Greg R. Guerin

References

Wrapper for `SDMTools` version of fragstats, which is based on statistics calculated by fragstats, see <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.

Jeremy VanDerWal, Lorena Falconi, Stephanie Januchowski, Luke Shoo and Collin Storlie (2014). `SDMTools`: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1-221. <http://CRAN.R-project.org/package=SDMTools>

See Also

[ClassStat](#)

Examples

```
eg.rast <- raster()
extent(eg.rast) <- c(0,10, 0, 10)
res(eg.rast) <- c(0.5,0.5)
eg.rast[] <- NA
eg.rast[sample(c(1:ncell(eg.rast)), 100)] <- 1
eg.coords <- data.frame(Longitude=c(3, 5, 7), Latitude=c(4, 6, 8))
row.names(eg.coords) <- c("SiteA", "SiteB", "SiteC")
eg.coords
frags <- suppressWarnings(buffer.fragstats(eg.coords, radius=150000, vegetation.base.raster=eg.rast))

#TO CONDENSE INTO SINGLE DATA.FRAME
frags <- do.call(rbind, frags)
frags
#NOTE:class = 100 is default output for NA input raster cells.
```

convert.site.data	<i>Convert species records from sites with spatial coordinates to individual records (i.e. long-format).</i>
-------------------	--

Description

This function converts species occurrence data from site- to individual species record-based. This conversion could be useful in any situation where individual species records are required, but was designed primarily to allow this data format as input to other biodiversity mapping functions, in which case this conversion makes plot -based data compatible with the format of herbarium or museum records, for example.

Usage

```
convert.site.data(species_records, site.coords)
```

Arguments

species_records	A data.frame with rows as sites and columns as species. Values greater than 0 are interpreted as indicating presence (abundances are not retained).
site.coords	A data.frame with the sites (/field plots) that match the column names of species_records as rows with longlat/XY coordinates.

Details

This function is partially a reverse operation to `simba::mama`, which converts long-format (individual) species records into matrix format.

Value

Returns a data.frame with three columns, "SPECIES", "LONGITUDE", "LATITUDE".

Author(s)

Greg R. Guerin & Lasse Ruokolainen

References

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

See Also

[mama](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)
mite_records <- convert.site.data(mite, mite.xy)
head(mite_records)
```

endemism.null.test	<i>Endemism Null Test</i>
--------------------	---------------------------

Description

Taking the outputs from the 'weighted.endemism' function, tests whether observed endemism is higher than expected, using non-parametric methods.

Usage

```
endemism.null.test(weighted.endemism.output, nrep = 100, outlier.range = 1.5)
```

Arguments

weighted.endemism.output	The returned object from a 'weighted.endemism' run (or equivalent constructed separately). Consists of a list of length 4: \$WE (named vector of endemism scores for grid cells), \$WE_raster (raster layer of endemism scores), \$weights (named vector of range weights for species used in the calculation of weighted endemism); \$grid.matrix (species against cells occurrence matrix - row names must match \$WE names, column names must match \$weights names). If 'corrected weighted endemism' was specified in the weighted.endemism run, this function will not accept the output. This is because correcting via dividing endemism by species richness is an alternative to this function, which uses more sophisticated methods for testing whether endemism is different than that expected for a given species richness.
nrep	The desired number of replicates when generating a null distribution from a random draw of species. Default is 100 for speed, but at least 1000 is recommended to ensure smooth null distributions and useful p-values.
outlier.range	Times outside the (null) interquartile range beyond which a record is considered an outlier.

Details

With the outputs from the 'weighted.endemism' function, performs the following tests:

- 1) Non-parametric significance test as to whether observed endemism is higher or lower than expected, given species richness (and observed species frequencies).
- 2) Identifies and maps outliers (i.e. in terms of map grid cells that have higher or lower endemism) based on quantiles. As categorical: whether endemism score lies more than 1.5 (or other user-defined amount) times outside the interquartile range; as continuous: the factor of the interquartile by which observed values differ from the median / 50 percent quantile). Returns vectors of values plus raster maps.

Raw weighted endemism scores are biased both by the completeness of species sampling and species richness itself. Correcting by dividing by the observed number of species ('corrected weighted endemism' of Crisp et al. 2001) is a proposed correction, but the relationship between endemism scores and species richness is not linear under a null model (random species draws), as increasingly infrequent species are drawn as richness increases, thereby increasing CWE. While correcting endemism scores in a more sophisticated way is possible, this function does not correct the scores per se, but compares them to a null distribution. This is achieved by making replicate random draws from the species pool based on the observed species richness (i.e. same number of

species) and the actual species frequencies (more frequent species more likely to be drawn). The distribution of the resulting set of null endemism scores is compared to observed endemism and subsequently grid cells can be mapped as higher or lower than expected (based on significance testing and comparison to null quantiles).

Value

Returns a list with following elements. Plots of observed and expected endemism against species richness are generated, as are plots of the generated rasters.

\$Quantile.25	Vector of expected lower interquartile range for a given species richness.
\$Quantile.75	Vector of expected upper interquartile range for a given species richness.
\$out.above.below	Vector assigning outliers categorically, with -1 for lower outlier, 0 non-outlier, 1 for upper outlier for each grid cell.
\$out.above.below.raster	Map of \$out.above.below.
\$out.continuous	Scores for each grid cell = by which factor of the interquartile range observed endemism differs from the median / 50 percent quantile.
\$out.continuous.raster	Map of \$out.continuous.
\$P.above	Vector of p-values for grid cells having higher than expected endemism. Very low scores = more likely to be higher than expected; very high scores = more likely to be lower than expected.
\$P.above.raster	Map of \$P.above.
\$richness	Simple observed species richness scores for cells.

Author(s)

Greg R. Guerin & Lasse Ruokolainen

References

Crisp, M.D., Laffan, S., Linder, H.P. & Monro, A.N.N.A. (2001) Endemism in the Australian flora. *Journal of Biogeography* **28**, 183-198.

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

See Also

[weighted.endemism](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)

endemism_mydata <- weighted.endemism(mite, site.coords=mite.xy, records="site")

endemism.test.example <- endemism.null.test(endemism_mydata)
```

map.multi.metrics	<i>Estimated alpha, multi-site beta and estimated phylogenetic diversity within grid cells or groups</i>
-------------------	--

Description

Given plot-based incidence data for species (i.e., presence/absence), calculates either alpha diversity (estimated using nonparametric methods), multi-site beta diversity (based on multiple plot samples), or phylogenetic diversity (estimated based on completeness of species sampling), associated with map grid cells or custom groupings.

Usage

```
map.multi.metrics(species_records, site.coords, alpha=TRUE, beta=TRUE,
  phylogenetic=FALSE, phylo.tree, frame.raster, deg.resolution=c(0.25,0.25),
  extent.vector, custom_grouping, plot.raster=TRUE)
```

Arguments

species_records	A data.frame of species occurrences in plots with rows as sites and columns as species. Cannot make use of presence-only records.
site.coords	A data.frame with rows as sites (/field plots) (names match the row names of species_records) and their geographic coordinates: first column must be x/longitude, second column y/latitude.
alpha	Logical with default TRUE. Whether to calculate alpha diversity for grid cells using nonparametric estimators.
beta	Logical with default TRUE. Whether to calculate multi-site beta diversity for grid cells.
phylogenetic	Logical with default FALSE. Whether to calculate phylogenetic diversity for grid cells with estimates based on completeness of species sampling.
phylo.tree	A phylogenetic tree of class 'phylo' (with branch lengths) containing species in the occurrence data. Must be supplied if phylogenetic=TRUE.
frame.raster	Optional. An existing rasterLayer object. User can elect to supply a raster. If not specified, the function generates a raster based on default or specified extent and resolution. Alpha and beta metrics are calculated based on plots in species_records that fall within grid cells of the frame.raster.
deg.resolution	Arguments specifying the map resolution (in degrees) the user wishes the calculations and mapping to use. If none are specified, default resolution (0.25) is used. If a frame.raster is specified, this argument is ignored.
extent.vector	Argument specifying the map extent (in degrees) the user wishes the calculations and mapping to use. If none are specified, default extent (data extent) is used. If a frame.raster is specified, this argument is ignored.
custom_grouping	Numeric constant identifying the column of site.coords which contains the desired custom groupings of sites (rows). If supplied, multi-site alpha, beta and phylogenetic diversity estimates are calculated by the groupings instead of by map grid cells.

`plot.raster` Logical with default TRUE. Whether to plot out generated rasters of alpha, beta and phylogenetic diversity. Either way, results tables and rasters are returned. Ignored if `custom_grouping` is supplied as no rasters are then produced.

Details

This function calculates diversity among plots located within the same grid cell, using metrics that are based on multiple community observations, rather than pooled observations over the entire grid cell. `map.multi.metrics` uses plot-based functions in packages BAT and vegan.

Alpha diversity estimation uses the [specpool](#) function, which provides observed species richness as well as estimates of richness based on chao, jackknife and bootstrap. All estimates are returned in the table but only chao is rasterised. This complements calculation of observed unweighted species richness in map grid cells (see [weighted.endemism](#) for plot-based datasets.)

For multi-site beta diversity, the [beta.multi](#) function is used to calculate diversity among the plots. Total beta diversity as well as its replacement and richness components (Cardoso et al. 2014a) are returned in the table but only total beta is rasterised.

Phylogenetic diversity estimation is based on the phylogenetic diversity of the entire species assemblage at grid cell level (i.e., all species in plots that fall in that cell) corrected by the completeness of species sampling in the relevant plots (Cardoso et al. 2014b). In this case, PD is divided by observed/estimated species richness (based on the above alpha diversity estimation), so that less complete sampling results in a larger boost to estimated PD. This complements calculation of observed PD in grid cells (see [phylogenetic.endemism](#)) for plot-based datasets and individual plot samples (see [phylogenetic.diversity.sites](#)).

The function is suitable for mapping grid cell level estimates of multi-site alpha, beta and phylogenetic diversity from distributed networks of vegetation plots. Alternatively, by supplying the `custom_grouping` argument, the user can calculate these metrics by predefined groups of plots, such as those resulting from classification or some other scheme.

Value

Returns a list of length 2, 4 or 6 depending on whether alpha, beta and phylogenetic are set to TRUE (length is halved if `custom_grouping` is supplied as no rasters are generated in that case):

<code>\$alpha_result</code>	A data.frame of alpha diversity metrics against rows which are named by grid cell numbers that match those in the raster.
<code>\$alpha_raster</code>	A RasterLayer object with values that represent chao estimates of species richness.
<code>\$beta_result</code>	A data.frame of beta diversity metrics against rows which are named by grid cell numbers that match those in the raster. Btotal is made up of the other two components, Brepl and Brich, which represent the richness difference and species replacement components of beta diversity.
<code>\$beta_raster</code>	A RasterLayer object with values that represent total multi-site beta diversity.
<code>\$phylo_result</code>	A data.frame of phylogenetic diversity (observed and estimated) against rows which are named by grid cell numbers that match those in the raster.
<code>\$phylo_raster</code>	A RasterLayer object with values that represent estimated phylogenetic diversity.

Author(s)

Greg R. Guerin

References

- Cardoso, P., Rigal, F., Carvalho, J.C., Fortelius, M., Borges, P.A.V., Podani, J. & Schmera, D. (2014a) Partitioning taxon, phylogenetic and functional beta diversity into replacement and richness difference components. *Journal of Biogeography* **41**, 749-761.
- Cardoso, P., Rigal, F., Borges, P.A. & Carvalho, J.C. (2014b) A new frontier in biodiversity inventory: a proposal for estimators of phylogenetic and functional diversity. *Methods in Ecology and Evolution* **5**, 452-461.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783-791.

See Also

[phylogenetic.diversity.sites](#)

[phylogenetic.endemism](#)

[weighted.endemism](#)

[specpool](#)

[beta.multi](#)

Examples

```
#####Preparation for this example:
library(vegan)
data(mite)
data(mite.xy)
#generate random phylogenetic tree
mite.tree <- ape::rtree(n=ncol(mite), tip.label=colnames(mite))
####Usage of the function:
my.multi <- map.multi.metrics(mite, mite.xy, phylogenetic=TRUE,
  phylo.tree=mite.tree, deg.resolution=c(1, 1), plot.raster=FALSE)
my.multi
plot(my.multi$phylogenetic_raster)

####Example with custom grouping instead of by map grid cells
#generate some random groupings for sites:
mite.xy$group <- sample(letters[1:5], nrow(mite.xy), replace=TRUE)
map.multi.metrics(mite, mite.xy, phylogenetic=TRUE, phylo.tree=mite.tree, custom_grouping=3)
```

map.pa.matrix

Map Presence-absence Matrix

Description

Given georeferenced incidence data for species, generates a binary presence/absence matrix associated with grid cells of a raster.

Usage

```
map.pa.matrix(species_records, records = "single", site.coords, species = "SPECIES",
longitude = "LONGITUDE", latitude = "LATITUDE", frame.raster,
deg.resolution = c(0.25, 0.25), extent.vector)
```

Arguments

<code>species_records</code>	A data.frame, either with: a) rows as individual species records, and columns that include fields for species name, longitude and latitude (see = species = , 'longitude', 'latitude' below); or b) rows as sites and columns as species, in which case 'site.coords' (below) must also be supplied.
<code>records</code>	Are the species_records in single/long format (the default, records="single") or in site-based/short format (records="site")?
<code>site.coords</code>	For site-based data (records="site"), a data.frame with rows as sites (/field plots) (names match the row names of species_records) and their geographic coordinates: first column must be x/longitude, second column y/latitude.
<code>species</code>	For records="single" (i.e. individual occurrence data); what colname in the supplied species_records contains species names?
<code>longitude</code>	For records="single"; what colname in the supplied species_records contains longitude values?
<code>latitude</code>	For records="single"; what colname in the supplied species_records contains latitude values?
<code>frame.raster</code>	An existing rasterLayer object. User can elect to supply a raster, in which case presences and absences are scored for grid cells in the raster. If not specified, the function generates a raster based on default or specified extent and resolution.
<code>deg.resolution</code>	Arguments specifying the map resolution (in degrees) the user wishes the calculations and mapping to use. If none are specified, default resolution (0.25) is used. If a frame.raster is specified, this argument is ignored.
<code>extent.vector</code>	Argument specifying the map extent (in degrees) the user wishes the calculations and mapping to use. If none are specified, default extent (data extent) is used. If a frame.raster is specified, this argument is ignored.

Details

This function generates a binary species presence/absence matrix associated with a raster layer based on georeferenced incidence data. This is a data processing step for mapping various biodiversity metrics onto raster layers. The outputs can be used as inputs into these functions, or if desired they can be used like site-based data (at the resolution of the raster) for various analysis such as ordination, or incidence/frequency data for particular species can be extracted.

Value

Returns a list of length 2:

<code>\$grid.matrix</code>	A binary data.frame of species occurrences against grid cell numbers that match those in \$pa.raster.
<code>\$pa.raster</code>	A RasterLayer object for which species presence/absence is scored in \$grid.matrix.

Author(s)

Greg R. Guerin & Lasse Ruokolainen

References

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

See Also

[lets.presab](#)

Examples

```
#####Preparation for this example:
library(vegan)
data(mite)
data(mite.xy)
####Usage of the function:
map.pa.matrix(mite, records="site", site.coords=mite.xy)
```

pe.null.test

Phylogenetic endemism null test

Description

Taking the outputs from the phylogenetic.endemism function, tests whether observed phylogenetic diversity/endemism is higher than expected, using non-parametric methods.

Usage

```
pe.null.test(phylogenetic.endemism.output, nrep = 100, outlier.range = 1.5,
pe.type = "weighted", phylo.tree)
```

Arguments

phylogenetic.endemism.output

The returned object from a [phylogenetic.endemism](#) run, a list of length 3 (for PD, pe.type="unweighted") or 7 (for PE: pe.type="weighted"): \$PD (numeric vector of phylogenetic diversity/endemism scores for grid cells); \$PD_raster (raster layer of PD/PE scores); \$grid.matrix (species against cells occurrence matrix - row names match \$PD names); \$ranges (numeric vector of range weights for phylogenetic branches); \$phyloMatrix (branches against species matrix); \$phylo.cell.matrix (branches against cells matrix); \$edge.lengths (numeric vector of edge.length in phylo.tree).

nrep

Desired number of replicates when generating a null distribution from a random draw of species. Default is 100 for speed (slow for large datasets), but at least 1000 is recommended to ensure smooth null distributions and useful p-values.

outlier.range

The multiplier of the interquartile range, outside of which scores are considered to be categorical outliers. Default is the typical value of 1.5.

pe.type	Refers back to how phylogenetic.endemism was run; alternatively set to pe.type="unweighted" for phylogenetic diversity with edge length uweighted by range size if this was the original setting (in which case 'phylo.tree' must be supplied).
phylo.tree	The phylogenetic tree that was used to run phylogenetic.endemism, only required for testing unweighted pd (pe.type="unweighted").

Details

With the outputs from the `phylogenetic.endemism` function, performs the following tests: **1)** non-parametric significance test as to whether observed phylogenetic diversity/endemism is higher or lower than expected, given species richness (and observed species frequencies); **2)** Identifies and maps outliers (i.e. in terms of map grid cells that have higher or lower PD/PE) based on quantiles. As categorical: whether score lies more than 1.5 (or other user-defined amount) times outside the interquartile range; as continuous: the factor of the interquartile by which observed values differ from the median / 50 percent quantile). Returns vectors of values plus raster maps.

Raw phylogenetic diversity/endemism scores are biased both by the completeness of species sampling and species richness itself. This function does not correct the scores per se, but compares them to a null distribution. This is achieved by making replicate random draws from the species pool based on the observed species richness (i.e. same number of species) and the actual species frequencies (more frequent species more likely to be drawn). The distribution of the resulting set of null scores is compared to the observed scores and subsequently grid cells can be mapped as higher or lower than expected (based on significance testing and comparison to null quantiles).

Value

Returns a list with following elements. Plots of observed and expected phylogenetic diversity/endemism against species richness are generated, as are plots of the generated rasters.

\$Quantile.25	Vector of expected lower interquartile range for a given species richness.
\$Quantile.75	Vector of expected upper interquartile range for a given species richness.
\$out.above.below	Numeric vector assigning outliers categorically, with -1 for lower outlier, 0 non-outlier, 1 for upper outlier, for each grid cell.
\$out.above.below.raster	Map of \$out.above.below
#\$out.continuous	Scores for each grid cell = by what factor of the interquartile range observed phylogenetic diversity/endemism differs from the median / 50 percent quantile.
\$out.continuous.raster	Map of \$out.continuous
\$P.above	Vector of p-values for grid cells having higher than expected phylogenetic diversity/endemism. i.e. very low score = likely to be higher than expected; very high scores = likely to be lower than expected.
\$P.above.raster	Map of \$P.above with default 2-tailed colour scheme
\$richness	Vector of simple observed species richness scores for cells.

Author(s)

Greg R. Guerin

References

Guerin, G.R. and Lowe, A.J. (submitted) Mapping phylogenetic endemism in R using georeferenced branch extents.

See Also

[phylogenetic.endemism](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)
#for example, generate phylogenetic tree of species with random relationships and branch lengths
mite.tree <- ape::rtree(n=ncol(mite), tip.label=colnames(mite))
mite.PE <- phylogenetic.endemism(mite, records="site", site.coords=mite.xy,
  sep.comm.spp="none", phylo.tree=mite.tree, sep.phylo.spp="none", weight.type="cell")
pe.mite.test <- pe.null.test(mite.PE)
pe.mite.test
```

phylogenetic.diversity.sites

Phylogenetic diversity for sites

Description

Calculates (unweighted and uncorrected) Faith's phylogenetic diversity for a set of community sample sites.

Usage

```
phylogenetic.diversity.sites(species_records, phylo.tree)
```

Arguments

species_records	A data.frame with rows as sites and columns as species (presence absence matrix).
phylo.tree	A phylogenetic tree of class 'phylo' containing species in the occurrence data. Must have branch lengths.

Details

Unweighted and uncorrected Faith's phylogenetic diversity for community samples organised into sites. Function takes a community matrix (could be species inventories from field plots or map grid cells) and an associated tree with branch lengths. Although coded differently, this is more or less equivalent to the 'pd' function in the 'picante' package. I have pulled it out as a stand-alone function for a community matrix from the unweighted (pd) case of the biomapME::phylogenetic.endemism function, which has point record inputs and gridded map outputs.

Value

Returns a vector of PD scores (list with \$PD as the vector) for sites in the community matrix.

Author(s)

Greg R. Guerin

References

Guerin, G.R. and Lowe, A.J. (2015) Mapping phylogenetic endemism in R using georeferenced branch extents. *SoftwareX* **3-4**, 22-26.

Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1-10.

See Also

[phylogenetic.endemism](#)

[pd](#)

[pd.query](#)

Examples

```
library(vegan)
data(mite)
#generate random phylogenetic tree
mite.tree <- ape::rtree(n=ncol(mite), tip.label=colnames(mite))
mite.PD <- phylogenetic.diversity.sites(mite, phylo.tree=mite.tree)
mite.PD
```

phylogenetic.endemism *Phylogenetic diversity, phylogenetic range rarity & georeferenced phylogenetic endemism*

Description

Calculates phylogenetic endemism (Faith's phylogenetic diversity inversely weighted by the spatial range of particular branch lengths) or alternatively (unweighted) phylogenetic diversity across gridded maps using individual or site-based point records.

Usage

```
phylogenetic.endemism(species_records, records = "single", site.coords,
  species = "SPECIES", longitude = "LONGITUDE", latitude = "LATITUDE",
  sep.comm.spp = " ", phylo.tree, sep.phylo.spp = "_", frame.raster,
  deg.resolution = c(0.25, 0.25), extent.vector, pe.type = "weighted",
  plot.raster = TRUE, weight.type = "cell", outlier_pct = 95, own.weights,
  own.grid.matrix, own.phylo.cell.matrix, own.phyloMatrix, pd.standard = FALSE)
```

Arguments

<code>species_records</code>	A data.frame, either with: a) rows as individual species records, and columns that include fields for species name, longitude and latitude (see 'species', 'longitude', 'latitude' below); or b) rows as sites and columns as species, in which case 'site.coords' (below) must also be supplied.
<code>records</code>	Are the species_records in single/long format (the default, records="single") or in site-based/short format (records="site").
<code>site.coords</code>	For site-based data (records="site"), a data.frame with rows as sites (/field plots) (names match the. row names of species_records) and their geographic (longlat) coordinates: first column must be x/longitude, second column y/latitude
<code>species</code>	For records="single" (i.e. individual occurrence data); what colname in the supplied species_records contains species names?
<code>longitude</code>	for records="single"; which colname in the supplied species_records contains longitude values?
<code>latitude</code>	For records="single"; which colname in the supplied species_records contains latitude values?
<code>sep.comm.spp</code>	The genus_species separating character in the community data. If there is none (i.e. taxon names represented by a single 'word'), argument should be set to sep.comm.spp="none". The purpose of this argument is that it is common for separators to differ between community data and phylogenetic trees due to the different processing functions applied, and to avoid having to reformat the data, this argument allows them to be matched within the function. Default is a space.
<code>phylo.tree</code>	A phylogenetic tree of class 'phylo' containing species in the occurrence data. Must have branch lengths.
<code>sep.phylo.spp</code>	The genus_species separating character in the phylo.tree (tip.labels). If there is none (i.e. taxon names represented by a single 'word'), argument should be set to sep.phylo.spp="none". See sep.comm.spp argument for purpose. Default for phylo.tree is an underscore, as this is a common format in community ecology using e.g. phylomatic trees.
<code>frame.raster</code>	An existing rasterLayer object. User can elect to supply a raster as the frame for calculations and mapping. If not specified, the function will generate a raster based on default or specified extent and resolution.
<code>deg.resolution</code>	Argument specifying the map resolution in degrees the user wishes the calculations and mapping to use. If none are specified, default resolution (0.25) is used. If a frame.raster is specified, these arguments are ignored (function bases mapping on the supplied raster).
<code>extent.vector</code>	Arguments specifying the map extent in degrees the user wishes the calculations and mapping to use. If none are specified, default extent (data extent) is used. If a frame.raster is specified, these arguments are ignored (function bases mapping on the supplied raster).
<code>pe.type</code>	Either "weighted" (default; = phylogenetic endemism), or "unweighted" (= phylogenetic diversity).
<code>plot.raster</code>	Logical. Whether or not to plot the output raster of endemism scores. Whether plotted or not, a raster object is stored in the returned output.
<code>weight.type</code>	Range weight type for calculation of phylogenetic endemism. Default is "cell" (grid cell-based range weights:: phylogenetic range rarity), while "geo" will

	calculate georeferenced range 'span' weights (georeferenced phylogenetic endemism). Argument is ignored if <code>pe.type="unweighted"</code> or if <code>own.weights</code> are provided.
<code>outlier_pct</code>	For the calculation of range 'span' via convex polygons, this argument can be used to remove outliers on a percentage basis via the 'mcp' function in package <code>adehabitatHR</code> , i.e. 95 means 5 percent most outlying points are removed. By default, 100, i.e., outliers are not excluded.
<code>own.weights</code>	An optional user-supplied numeric vector of branch length weights for calculating phylogenetic endemism. This argument is intended mainly so that time consuming calculation of georeferenced weights can be done once and the result stored and used for subsequent re-runs. Alternatively, user-calculated weights can be used in subsequent runs. The weights must be matched to a matrix representation of <code>phylo.tree</code> (see below).
<code>own.grid.matrix</code>	Optional user supplied binary matrix of species against grid cell numbers, rather than this being generated within the function. The purpose of this argument is to save time for repeat runs, given the step can be time consuming for large datasets. Must be provided if <code>own.phylo.cell.matrix</code> is supplied— although the function no longer requires the <code>grid.matrix</code> , it is still needed for inclusion in the outputs, so that they can be subsequently fed into the <code>pe.null.test</code> function (which does requires the <code>grid.matrix...</code>). If supplied, a <code>frame.raster</code> must also be supplied that has cell numbers which match the <code>row.names</code> of the <code>own.grid.matrix</code> . Assumes that genus/species words are separated by the <code>sep.phylo.spp</code> , as this is how the <code>grid.matrix</code> is returned from the function originally.
<code>own.phylo.cell.matrix</code>	User can supply the matrix representing occurrences of phylogenetic branches in map grid cells from previous runs to save computation time (i.e. <code>\$phylo.cell.matrix</code> of a previous <code>phylogenetic.endemism</code> output). A matching 'frame.raster' must also be supplied.
<code>own.phyloMatrix</code>	If user provides the above <code>own.phylo.cell.matrix</code> in a repeat run of the function, this must also be supplied, as the function still defaults to returning 'phylo.matrix' even though the function now skips its generation. e.g. <code>\$phyloMatrix</code> of an previous <code>phylogenetic.endemism.R</code> run).
<code>pd.standard</code>	Logical with default FALSE. Applies only to the calculation of unweighted phylogenetic diversity (i.e., <code>pe.type="unweighted"</code>). If set to TRUE, PD is standardised by species richness. See Details.

Details

This implementation of phylogenetic endemism for incidence data within map grid cells allows alternative calculation of weights for branch length geographic ranges. Weights can be calculated based on the frequency of occurrence in grid cells, or alternatively by the georeferenced span of the range.

Unweighted phylogenetic diversity can also be selected. In this case, if `pd.standard = TRUE`, the function becomes a rasterised wrapper for the [pd.query](#) function in package `PhyloMeasures`, which uses a null model to standardise PD for species richness (given PD is additive by species) by subtracting the mean and dividing by the standard deviation of expected PD.

Value

Returns a list of length 3 (for `pe.type="unweighted"` i.e. phylogenetic diversity/PD) or length 7 (for `pe.type="weighted"` i.e. phylogenetic endemism/PE). For both PD and PE:

`$PD` Vector of phylogenetic diversity/endemism scores.

`$PD_raster` Raster map with phylogenetic diversity/endemism scores.

`$grid.matrix` A binary data.frame of species against grid cell numbers used in the function which is returned so that it can be re-used to save computation time, and because it is required in the downstream `pe.null.test.R` function to calculate species sample probabilities. Note that this only includes species that are found in the `phylo.tree`.

For PE only:

`$ranges` A named numeric vector of weights used to calculate endemism (equivalent to range size in metres if `weight.type="geo"`, range size in cells if `weight.type="cell"` (default), or the user supplied weights if `own.weights` was supplied) (skipped if `pe.type="unweighted"`).

`$phyloMatrix` A binary matrix representation of the `phylo.tree`, returned mainly for convenience for downstream use (skipped if `pe.type="unweighted"`).

`$phylo.cell.matrix` A binary matrix recording the presence of particular phylogenetic branches in map grid cells used in the function, which is returned for re-use in subsequent runs for efficiency (e.g. with different weights), (skipped if `pe.type="unweighted"`).

`$edge.lengths` A numeric vector of edge lengths from the `phylo.tree`

Author(s)

Greg R. Guerin

References

- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1-10.
- Guerin, G.R. & Lowe, A.J. (2015) Mapping phylogenetic endemism in R using georeferenced branch extents. *SoftwareX* **3-4**, 22-26.
- Rosauer, D., Laffan, S.W., Crisp, M.D., Donnellan, S.C. & Cook, L.G. (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* **18**, 4061-4072.
- Tsirogiannis, C. & B. Sandel. (2016) PhyloMeasures: A package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography* **39**, 709-714,

See Also

[pe.null.test](#)

[phylogenetic.diversity.sites](#)

[map.multi.metrics](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)

#generate random phylogenetic tree of species in dataset
mite.tree <- ape::rtree(n=ncol(mite), tip.label=colnames(mite))

#calculate phylogenetic range rarity
mite.PE <- phylogenetic.endemism(mite, records="site", site.coords=mite.xy,
sep.comm.spp="none", phylo.tree=mite.tree, sep.phylo.spp="none", weight.type="cell")
mite.PE$PD

#calculate richness-standardised phylogenetic diversity
mite.PD <- phylogenetic.endemism(mite, records="site", site.coords=mite.xy,
sep.comm.spp="none", phylo.tree=mite.tree, sep.phylo.spp="none",
pe.type="unweighted", pd.standard=TRUE)
mite.PD$PD
```

spp.range.metrics	<i>Species range metrics</i>
-------------------	------------------------------

Description

Calculates selected species range metrics, given a set of georeferenced locations. Metrics include grid cell and point location based measures of Area of Occupancy and Extent of Occurrence.

Usage

```
spp.range.metrics(species_records, species = "SPECIES", longitude = "LONGITUDE",
latitude = "LATITUDE", coord.type = "longlat", weight.type = "cell",
geo.calc = "max.dist", outlier_pct = 100, verbose = TRUE, frame.raster,
deg.resolution = c(0.25, 0.25), extent.vector, plot.out = TRUE)
```

Arguments

species_records	A data.frame with rows as individual species records, and columns that include fields for species name, longitude/x and latitude/y (see species, longitude, latitude below).
species	What colname in the supplied species_records contains species names?
longitude	What colname in the supplied species_records contains longitude/Y values?
latitude	What colname in the supplied species_records contains latitude/Y values?
coord.type	Either the default of "longlat" or "custom" for any other coordinate systems. If the former, distances will be calculated in metres, if "custom", in the the same units as the XY coordinates.
weight.type	Default is "cell" (cell-based range metrics), while "geo" will calculate geographic range weights (no rasters).

<code>geo.calc</code>	Default is "max.dist". This argument is only considered when <code>weight.type</code> is set to "geo". If <code>geo.calc="max.dist"</code> the default maximum geographic distance ('span') between point locations is calculated for a species range. If <code>geo.calc</code> is set to "polygon", the function weights species ranges based on the area of a MCP (minimum convex polygon) that contains all points. Further arguments to this function can be included, such as changing the default <code>outlier_pct=95</code> (removes outlying locations). To calculate longitudinal/latitudinal range spans, set to: <code>geo.calc="LONG"</code> or <code>geo.calc="LAT"</code> .
<code>outlier_pct</code>	For the calculation of range span or area via convex polygons (at least 5 records of the species), this argument can be used to remove outliers on a percentage basis via the <code>mcp</code> function in package <code>adehabitat</code> , i.e. 95 means 5 percent most outlying points are removed (the default).
<code>verbose</code>	If TRUE (default) detail of progress by species is reported while the function is still running. Alternatively, set to FALSE.
<code>frame.raster</code>	An existing <code>rasterLayer</code> object the user can optionally elect to supply as the frame for calculations and mapping. If not specified, the function will generate a raster based on default or specified extent and resolution (if needed).
<code>deg.resolution</code>	Argument specifying the map resolution (in degrees for longlat) the user wishes the calculations and mapping to use. If no frame is specified, an arbitrary resolution is supplied. If a <code>frame.raster</code> is specified, argument is ignored (function bases mapping on the supplied raster).
<code>extent.vector</code>	Argument specifying the map extent (in degrees for longlat) the user wishes the calculations and mapping to use. If no frame is specified, an arbitrary extent is supplied. If a <code>frame.raster</code> is specified, argument is ignored (function bases mapping on the supplied raster).
<code>plot.out</code>	TRUE or FALSE. Whether or not to plot locations and rasters for cell count range scores (<code>weight.type="cell"</code>). Either way, the rasters are stored in the output.

Details

Given a list of georeferenced (longlat) species records, calculates range metrics according to user choice of number of occupied map grid cells (from a supplied raster or one automatically generated within the function), maximum span across the range, range area (area of polygon defined by occurrences) or the longitudinal/latitudinal extent.

Georeferenced distance and area range metrics can be calculated longlats (default and preferred) or any coordinate set. For non-longlat coordinates, distance and area calculations assume a flat surface (longlat calcs assume curved surface). This would be appropriate for smaller areas where distance measures won't be as biased by curvature of the Earth's surface, so that the function can be used directly on data sets where species coordinates are not recorded in longlat, e.g. fine-scale locations within plots, or UTM coordinates within one map zone etc.

Value

Returns a list. For cell-based frequency, contains a vector of range scores (number of cells) and a `RasterStack` containing a `RasterLayer` showing cell occupancy for each unique species. For georeferenced calculations of range, only a vector of range scores (in m or m² or degrees of longitude/latitude if longlat coordinates used)

Author(s)

Greg R. Guerin

References

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

See Also

[weighted.endemism](#)
[lets.rangesize](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)
mite_records <- convert.site.data(mite, mite.xy)
spp.range.metrics(mite_records, coord.type="custom", weight.type="geo")
```

weighted.endemism	<i>Range rarity richness, georeferenced weighted endemism and species richness</i>
-------------------	--

Description

Calculates (taxonomic / species) sum of inverse range-sizes ('SIR': species richness inversely weighted by species ranges) metrics across gridded maps using single or site-based point records.

Usage

```
weighted.endemism(species_records, records = "single", site.coords,
species = "SPECIES", longitude = "LONGITUDE", latitude = "LATITUDE",
frame.raster, deg.resolution = c(0.25, 0.25), extent.vector, type = "weighted",
plot.raster = TRUE, own.weights, weight.type = "cell", geo.type = "cell",
geo.calc = "max.dist", outlier_pct = 100, verbose = TRUE, own.grid.matrix)
```

Arguments

species_records	A data.frame, either with: a) rows as individual species records, and columns that include fields for species name, longitude and latitude (see species, longitude, latitude below); or b) rows as sites and columns as species, in which case site.coords (below) must also be supplied.
records	Are the species_records in single/long format (the default, "single") or in site-based/short format (records="site")?
site.coords	For site-based data (records="site"), a data.frame with the sites (/field plots) that match the column names of species_records and their longlat coordinates.
species	For records="single"; which colname in the supplied species_records contains species names?
longitude	For records="single"; which colname in the supplied species_records contains longitude values?

latitude	For records="single"; which colname in the supplied species_records contains latitude values?
frame.raster	An existing rasterLayer object the user can optionally elect to supply as the frame for calculations and mapping. If not specified, the function will generate a raster based on default or specified extent and resolution.
deg.resolution	Argument specifying the map resolution in degrees the user wishes the calculations and mapping to use. If no frame is specified, an arbitrary resolution is supplied. If a frame.raster is specified, these arguments are ignored (function bases mapping on the supplied raster)
extent.vector	Arguments specifying the map extent in degrees the user wishes the calculations and mapping to use. If no frame is specified, an arbitrary extent is supplied, that of the input point data. If a frame.raster is specified, these arguments are ignored (function bases mapping on the supplied raster)
type	Either "weighted" (default), or "corrected" (corrected weighted endemism - the 'per-species' weighted endemism as per Crisp et al. (2001). This is provided for convenience, but is not particularly recommended - instead, use the outputs of weighted.endemism in endemism.null.test function to compare to null expectations of endemism given the species richness)
plot.raster	Whether or not to plot the output raster with endemism scores. Either way, the raster object is stored in the output.
own.weights	An optional user-supplied numeric vector of species weights for calculating endemism. Values must have names that are a complete and exact match for the species names in species_records as each species must have a weight. This optional argument is intended mainly so that the more time consuming calculation of geographic weights can be done once and the result stored and used for subsequent re-runs of the endemism calculations.
weight.type	Default is "cell" (cell-based range weights), while "geo" will calculate geographic range weights. Weight.type "richness" sets the weights to 1, which is equivalent to calculating simple species richness (note setting type="corrected" in this case will give each cell a score of 1). Argument is ignored if own.weights is supplied.
geo.type	Default is "cell". This argument is only considered when weight.type is set to "geo" (above), in which case geo.type="cell" calculates geographic ranges based on map grid cell centroids. This can optionally be set to geo.type="point", in which case the geographic range weightings are calculated based on the point locations of each species.
geo.calc	Default is "max.dist". This argument is only considered when when weight.type is set to "geo" and applies to both geo.type="cell" and geo.type="point". If geo.calc="max.dist" the default maximum geographic distance ('span') between cell centroids/point locations is calculated for a species range. If geo.calc is set to "polygon", the function weights species ranges based on the area of a MCP (minimum convex polygon) that contains all points. Further arguments to this function can be included, such as changing the default outlier_pct=95 (removes outlying locations). Additionally, the geo.type="point" option is not recommended for the "polygon" method, as it is more likely to lead to errors where nearby point locations do not allow drawing of a spanning polygon. In this case, cell-centroid based calculations ensure that multiple records are spatially separated (in different cells) and that occurrences within a single cell are returned as the area of that cell.

`outlier_pct` For the calculation of range span or area via convex polygons (at least five records of the species), this argument can be used to remove outliers on a percentage basis via the `mcp` function in package `adehabitat`, i.e. 95 means 5 percent most outlying points are removed. Default is 100, i.e. no outlier exclusion.

`verbose`

`own.grid.matrix` User can supply a binary matrix of species against grid cell numbers, rather than this being generated within the function. The purpose of this argument is that the step can be time consuming for large datasets, so the user can return the matrix that is returned from the function in subsequent runs with different setting to improve speed. If this is supplied, a `frame.raster` must also be supplied that has cell numbers which match the `row.names` of the `own.grid.matrix`.

Details

This implementation of 'weighted endemism' allows alternative calculation of weights for species ranges as well as the option of user-supplied weights. Weights can be calculated based on the frequency of occurrence in grid cells, or alternatively by the geographical size of the species range, calculated in one (span) or two (area) dimensions.

Value

Returns a list of length 4:

`$WE` (`/$CWE`) Vector of 'weighted endemism' scores.

`$WE_Raster` (`/$CWE`) Raster map with endemism scores.

`$weights` A named numeric vector of weights used to calculate endemism (equivalent to range size in metres if `weight.type="geo"`, range size in cells if `weight.type="cell"` (default), or the user supplied weights if `own.weights` was supplied).

`$grid.matrix` A binary data.frame of species against grid cell numbers used in the function which is returned so that it can be re-used in subsequent runs to save time.

Author(s)

Greg R. Guerin & Lasse Ruokolainen

References

Guerin, G.R., Ruokolainen, L. & Lowe, A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution* **6**, 845-852.

See Also

[endemism.null.test](#)

Examples

```
library(vegan)
data(mite)
data(mite.xy)

#default, range rarity richness
```

```
endemism_mydata <- weighted.endemism(mite, site.coords = mite.xy, records="site")
endemism_mydata$WE
endemism_mydata$weights

#georeferenced weighted endemism; re-apply previous outputs
endemism_mydata2 <- weighted.endemism(mite, site.coords = mite.xy, records="site",
weight.type="geo", own.grid.matrix = endemism_mydata$grid.matrix,
frame.raster=endemism_mydata$WE_raster)
endemism_mydata$WE
endemism_mydata$weights

#re-apply weights from previous run
endemism_mydata3 <- weighted.endemism(mite, site.coords = mite.xy, records="site",
own.weights = endemism_mydata2$weights)

#unweighted species richness
endemism_mydata <- weighted.endemism(mite, site.coords = mite.xy,
records="site", weight.type="richness")
endemism_mydata$WE
endemism_mydata$weights
```


Index

beta.multi, [9](#), [10](#)
biomapME, [2](#)
buffer.frag, [2](#), [3](#)

ClassStat, [4](#)
convert.site.data, [2](#), [5](#)

endemism.null.test, [2](#), [6](#), [23](#)

lets.presab, [12](#)
lets.rangesize, [21](#)

mama, [5](#)
map.multi.metrics, [2](#), [8](#), [18](#)
map.pa.matrix, [2](#), [10](#)

pd, [15](#)
pd.query, [15](#), [17](#)
pe.null.test, [3](#), [12](#), [18](#)
phylogenetic.diversity.sites, [3](#), [9](#), [10](#),
[14](#), [18](#)
phylogenetic.endemism, [3](#), [9](#), [10](#), [12](#), [14](#), [15](#),
[15](#)

specpool, [9](#), [10](#)
spp.range.metrics, [3](#), [19](#)

weighted.endemism, [3](#), [7](#), [9](#), [10](#), [21](#), [21](#)