Package: BoskR (via r-universe)

August 30, 2024

Type Package

Title Assess Adequacy of Diversification Models Using Tree Shapes

Version 1.1.2

Date 2024-04-15

Description Given a phylogeny (or several) and a diversification model
(or several), the package makes use of graph Laplacians (as
implemented in RPANDA) and other tree shape metrics to infer
summary statistics. The overlap of those metrics will be
compared to the metrics from a set of trees simulated under the
model in question and based on the same parameters as the
initial tree. If the metrics indicate that the shapes of the
simulated trees significantly differ from the initial one, the
model should be deemed not adequate for this tree.

License GPL-3

Depends R (>= 3.5.1), ape

Imports RPANDA (>= 1.5), diversitree, phyloTop, DDD, TESS, TreeSim,
phangorn, picante, scales, scatterplot3d

Suggests knitr, testthat, spacodiR, rmarkdown, markdown

VignetteBuilder knitr

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Encoding UTF-8

Repository https://phylotastic.r-universe.dev

RemoteUrl https://github.com/oschwery/BoskR

RemoteRef HEAD

RemoteSha 47659baea551bff2c1971c87c197ea83bccb234d

Contents

BoskR-package   *BoskR - Assess Adequacy of Diversification Models Using Tree Shapes*

Description

Given a phylogeny (or several) and a diversification model (or several), the package makes use of
graph Laplacians (as implemented in RPANDA) to infer tree shape metrics. The overlap of those
metrics will be compared to the metrics from a set of trees simulated under the model in question
and based on the same parameters as the initial tree. If the metrics indicate that the shapes of the
simulated trees significantly differ from the initial one, the model should be deemed not adequate
for this tree.

Details

Index: This package was not yet installed at build time.

Author(s)

Maintainer: Orlando Schwery <oschwery@vols.utk.edu> (ORCID)

References

In prep.

See Also

CombineTrees, GetMetricTreeSets, GetTreeMetrics, GetTreeParams, PvalMetrics, ScatterMetrics,
TreeCorr, plotPvalMetricsCDF, plotPvalMetricsPDF

RPANDA

CombineTrees	*Rearrange Input Trees*

Description

Rearranges different input trees to correct format for downstream analyses.

Usage

```
CombineTrees(trees, sims = FALSE)
```

Arguments

trees
: Vector of tree objects: individual trees, lists of trees, and/or objects of class multiPhylo. Also accepts the output of `GetMetricTrees` from specified parameters.

sims
: Logical, `FALSE` if combining empirical trees, `TRUE` if combining simulations based on an empirical tree. This setting is mainly for the case where trees were simulated under a model that is not implemented, so they can be supplied to `GetTreeMetrics`. Default is `FALSE`.

Details

This function accepts different kinds of input phylogenies (see below) and rearranges them into a format that will work for the remaining functions of the package.

Value

List of trees in correct format to be used by downstream functions.

---

emptesttrees	*Set of Test Trees for BoskR*

Description

ADD DESCRIPTION HERE

Usage

```
data(epmtesttrees)
```

Format

A list of phylogenies (of class phy)

Source

ADD THEM HERE

References

ADD THEM HERE

GetMetricTreeSets   *Simulate trees based on empirical estimations or set parameters*

Description

Uses `GetMetricTrees` to simulate trees under a given model based on either parameter estimates from empirical trees or pre-set parameters.

Usage

```
GetMetricTreeSets(
  empirical_start = FALSE,
  empParams = empParams,
  current_method,
  N = NULL,
  Numbsim1,
  Lambda,
  Mu,
  l = NULL,
  a = NULL,
  LambdaFun = NULL,
  MuFun = NULL,
  TreeAge = NULL,
  BiSSEpars = NULL,
  tree = NULL
)
```

Arguments

empirical_start

TRUE to use parameters estimated from empirical trees, FALSE to use user-specified ones

empParams   Nested list object with tree parameters as inferred through `GetParams` from one or several empirical trees

current_method   Method to be used for simulation, either `"Yule"`, `"BD"`, `"TimeD-BD"`, `"DD"`, `"CD"`, `"TraitD"` for birth-death, time-dependent birth-death, diversity dependent, clade dependent, or trait dependent diversification respectively.

N   Number of taxa

Numbsim1   Number of trees to simulate per each

Lambda   Speciation rate

Mu   Extinction rate

l   Speciation rate

a	Extinction fracion (Mu/Lambda)
LambdaFun	Function for speciation rate
MuFun	Function for extinction rate
TreeAge	Stem age of tree
BiSSEpars	Parameters from BiSSE
tree	Phylogeny

Details

The function will simulate a number of trees based on either the parameters inferred from one or several empirical trees (given through empParams if empirical_start=TRUE), or user-specified parameters (if empirical_start=FALSE)

Value

A list of trees of class multiPhylo

GetTreeMetrics *Get metrics describing tree shape*

Description

GetTreeMetrics calculates a number of metrics describing tree shape for a tree or a set of trees.

Usage

GetTreeMetrics(trees, empirical_start = FALSE)

Arguments

trees	Tree or set of trees, list or multiPhylo-object, or list of tree sets
empirical_start	
	TRUE if started out from empirical trees, FALSE if started from user-specified parameters

Details

The function wraps around the internal 'GetMetrics', which will calculate five 'traditional' tree metrics (Colless, Sackin, number of cherries, number of pitchforks, ladder sizes), as well as standard and normalised graph Laplacian spectra and the associated summary metrics (principal eigenvalue, asymmetry, peakedness, eigengap), as implemented in RPANDA.

Value

A list with two elements: metrics: a matrix with the values for all tree metrics for each tree, and spectra: a list of raw values for the standard and normalised graph Laplacian spectra for each tree. If applied to the simulated trees based on a tree set, it will be one such two-element list for each tree set provided in a nested list.

GetTreeParams	*Get diversification parameters from trees*

Description

GetTreeParams estimates parameters from a supplied tree or tree set, which can subsequently be used as input for tree simulations using GetMetricTrees.

Usage

```
GetTreeParams(trees, current_method_est)
```

Arguments

trees   Tree or set of trees, list or multiPhylo-object, or list of tree sets

current_method_est

    String specifying the method to be used to estimate the parameters. For possible values see details section.

Details

The function wraps around the internal GetParams, and uses either ...

The parameter current_method_est can be '"Yule", "BD", "Time_lambda_mu", "DD_lambda_mu", for birth-death, time-dependent birth-death, or diversity dependent, respectively. For the time- and diversity-dependent models, "lambda" and "mu" in the name should be replaced with the kind of time dependence intended for the respective parameter, being "const", "lin", or "exp" for constant, linear or exponential respectively. For a pure-birth model (only time-dependent), mu can be set to "PB".

For diversity-dependent models, only five combinations are available: linear lambda, exponential lambda, linear mu, exponential mu, and both linear.

Example: a time dependent model with exponential speciation rate and constant extinction rate would be specified by "Time_exp_const".

Value

A nested list of parameter estimates for every tree in trees, or every tree in each tree set therein respectively.

---

plotPvalMetricsCDF   *Plot p-values on CDF for sets of tree metrics*

Description

Creates plots of p-values on their corresponding cumulative distribution function, based on sets of simulated and empirical distributions of tree metrics,

Usage

```
plotPvalMetricsCDF(pmetrics, set = NULL)
```

Arguments

pmetrics	Object with ECDs and p-values of empirical and simulated tree shapes, output of `PvalMetrics` or formatted the same way.
set	Numerical index for which of the sets of pairs of empirical and simulated metrics to be plotted; default NULL will plot all sets.

Value

An array of plots.

---

plotPvalMetricsPDF   *Plot p-values on PDF for sets of tree metrics*

Description

Creates plots of p-values on their corresponding probability density function, based on sets of simulated and empirical distributions of tree metrics,

Usage

```
plotPvalMetricsPDF(empMetrics, simMetrics, set = NULL, metricset = "spectR")
```

Arguments

empMetrics	Metrics of empirical tree or set of trees; output of `GetTreeMetrics` or formatted the same way.
simMetrics	Metrics of sets of simulated trees; output of `GetTreeMetrics` or formatted the same way.
set	Numerical index for which of the sets of pairs of empirical and simulated metrics to be plotted; default NULL will plot all sets.
metricset	String specifying which tree metrics to use; default is "spectR", other options are "spectRnorm", "classic", and "nodibranch"; for more information on the options see Details of `PvalMetrics()`.

Value

An array of plots.

PvalMetrics *Get p-values for tree metrics*

Description

Estimates p-values based on simulated and empirical distributions of tree metrics

Usage

```
PvalMetrics(
  empMetrics,
  simMetrics,
  empirical_start = TRUE,
  methodnr,
  metricset = "spectR"
)
```

Arguments

empMetrics	Metrics of empirical tree or set of trees; output of `GetTreeMetrics` or formatted the same way.
simMetrics	Metrics of sets of simulated trees; output of `GetTreeMetrics` or formatted the same way.
empirical_start	
	Indicator whether empMetrics is based on empirical or simulated initial trees, default is `TRUE` (=empirical); mainly important for data format reasons.
methodnr	Integral specifying which method is used: 1: BD, 2: TimeD-BD, 3: DD; is only used if `empirical_start` is `TRUE`
metricset	String specifying which tree metrics to use; default is "spectR", other options are "spectrRnorm", "classic", and "nodibranch"; for more information on the options see Details.

Details

The function uses an Empirical Cumulative Distribution function to determine the area under the curve of the metric values of the simulated trees, to get to a p-value for the position of the metrics of the empirical tree on that distribution. The argument `metricset` allows to chose between: ″spectR″- the standard (i.e. unnormalised) spectral densities, ″spectRnorm″ - the normalised spectral densities, ″classic″ - a couple of more 'conventional' measures of tree shape, being Colless index, Sackin index, number of cherries, number of pitchforks, average ladder size, and gamma statistic; finally '"nodibranch" - includes minimum, maximum, and median for both node ages and branch lengths respectively. For more information on the spectral densities, i.e. the Eigenvalues of the tree's modified graph Laplacian, see R package RPANDA and associated papers.

Value

A list with two entries: `ECDs` is a list of Empirical Cumulative Distributions; `pValues` is a matrix with p-values for the targeted metrics

---

ScatterMetrics *3D Metrics Scatterplot*

Description

Plots empirical trees and their simulations in tree metric space using a 3D scatterplot.

Usage

```
ScatterMetrics(
  empMetrics,
  simMetrics,
  pair = 1,
  skim = FALSE,
  combine = FALSE,
  colours = c("black", "red"),
  transparencyEmp = 0.8,
  transparencySim = 0.2,
  pch = 16,
  cex.symbols = 1.5,
  main = paste("Empirical vs. Simulated Metrics Set", pair, sep = " "),
  angle = -230
)
```

Arguments

empMetrics	Metrics of empirical tree or set of trees; output of `GetTreeMetrics` or formatted the same way.
simMetrics	Metrics of sets of simulated trees; output of `GetTreeMetrics` or formatted the same way.
pair	Numerical index for which of the sets of pairs of empirical and simulated metrics to be plotted. Value is ignored if `skim` or `combine` are TRUE.
skim	Logical, creates interactive plot of all pairs of empirical trees and their simulations if TRUE; one can advance through the plots by hitting enter.
combine	Logical, combines all empirical and simulated trees into one plot if TRUE.
colours	Vector of length two, indicating the desired colours for empirical trees and simulated treees, in that order (defaults are "black" and "red", respectively).
transparencyEmp	
	Value determining the transparency of the empirical tree plot points (0: completely transparent, 1: completely opaque; corresponding to `alpha` from package `scales`).

transparencySim

> Value determining the transparency of the simulated tree plot points (0: completely transparent, 1: completely opaque; corresponding to `alpha` from package `scales`).

`pch`	Shape of plot symbols; default 16.
`cex.symbols`	Size of plot symbols; default 1.5.
`main`	String for plot title; default "Empirical vs. Simulated Metrics Set", followed by pair number plotted, or "Combined".
`angle`	Rotation of the plot, determined by angle between x and y axis (corresponding to `scatterplot3d`); default -230.

Details

The function uses the internals `ScatterMetricsPair` and `ScatterMetricsCombo` and plots the empirical input-trees and their corresponding simulations in the metric space (asymmetry x peakedness x principal Eigenvalue) as a 3D scatterplot. It allows to either plot them all combined, or pairwise. The latter meaning each empirical tree is plotted with its corresponding simulations only, either one at a time or all together interactively (one advances through the plots by pressing enter). The basic function used is `scatterplot3d`, from the package with the same name.

Value

3D scatterplot of trees in metric space, or a series of such plots to skip through.

TreeCorr   *Run Tests and Corrections for Trees*

Description

Tests input treeset for branch length rounding errors, zero length branches, and order.

Usage

```
TreeCorr(emptrees)
```

Arguments

`emptrees`	Tree or list of trees.

Details

The function is a wrapper around the internals `CorrUltramet`, `CorrZerobranch`, and `ReorderCladewise`. Trees which are not ultrametric due to rounding errors are being corrected using `nnls.tree` as [discribed on the phytools blog](), polytomies are randomly resolved and all trees are reordered to 'cladewise' using the ape functions `multi2di` and `reorder.phylo` respectively.

Value

Same tree set as input, but corrected if necessary.

Index