Package 'iteRates'

February 20, 2015

Type Package

Title Parametric rate comparison

Version 3.1

Date 2012-12-03

Author Premal Shah, Benjamin Fitzpatrick, James Fordyce

Maintainer Ben Fitzpatrick <benfitz@utk.edu>

Description Iterates through a phylogenetic tree to identify regions of rate variation using the parametric rate comparison test.

License GPL (>= 3)

LazyLoad yes

Depends partitions, stats, VGAM, MASS, ape, apTreeshape, geiger, gtools

NeedsCompilation no

Repository CRAN

Date/Publication 2013-05-03 21:40:36

R topics documented:

`iteRates-package` *iteRates*

Description

Iterates through a phylogenetic tree to identify regions of rate variation using the parametric rate comparison test.

Details

Package:	iteRates
Type:	Package
Version:	3.0
Date:	2011-05-24
License:	GPL 3.0
LazyLoad:	yes

The user provides a phylogenetic tree of object class phylo. The package will iterate through all useable subtrees and identify regions of the tree with different rates of diversification using the parametric rate comparison test.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

Maintainer: Ben Fitzpatrick <benfitz@utk.edu>

`color.tree.plot` *color.tree.plot*

Description

This function plots phylogenetic trees on the current graphical device and indicates potential regions of the tree that might have undergone a shift in diversification rate.

Usage

```
color.tree.plot(out, tree, p.thres = 1, evid.thres=0, PorE=1, show.node.label = FALSE,
NODE = TRUE, PADJ = NULL, scale = 1, col.rank = TRUE, breaks = 50, ...)
```

Arguments

out	the output object from comp.subs.
tree	an object of class "phylo" used in the comp.subs analysis.
p.thres	a numeric between 0 and 1 setting the threshold to plot rate-shifts with p-value<=p.thres. Default is 1.0.
evid.thres	a numeric setting the threshold to plot rate-shifts with evidence ratio >=evid.thres. Default is 0.
PorE	a switch to indicate whether rate-shifts are indicated based on the p-value (PorE=1) or the evidence ratio (PorE=1).
show.node.label	a logical indicating whether the node labels need to be plotted with the tree. Default is FALSE.
NODE	a logical switch between identifying rate-shifts on trees by coloring "nodes" or "branches". Default is TRUE.
PADJ	a character vector to adjust p-values from comp.subs for multiple comparison. Options are identical to the ones in p.adjust in the stats package including "holm","hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none". Default is NULL.
scale	a numeric that controls the size of the colored nodes or thickness of colored branch lengths used to indicate rate-shifts. Default is 1.
col.rank	a logical indicating whether various instances of potential rate-shifts should be colored based on the rank of the p-value or the absolute magnitude of the rate-shift. Default is TRUE indicating use of ranks instead of magnitude.
breaks	a numeric indicating the range of colors to be used for plotting. Choosing a smaller value will lead to big differences in colors while a bigger value will lead to finer variations in colors.
...	additional arguments to be passed to plot.phylo in the ape package.

Details

When passing an object of class "phylo" (tree) follow the guidelines in plot.phylo in the ape package. Also make sure that the tree passed to color.tree.plot is the same as the one used to generate out from comp.subs.

Value

color.tree.plot returns only a graphical device output.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James A. Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

comp.subs, plot.phylo

Examples

```
data(geospiza)
attach(geospiza)

output.geospiza <- comp.subs(geospiza.tree)

color.tree.plot(out=output.geospiza, tree= geospiza.tree)
color.tree.plot(out=output.geospiza, tree= geospiza.tree, NODE=FALSE)
color.tree.plot(out=output.geospiza, tree= geospiza.tree, p.thres=1)
color.tree.plot(out=output.geospiza, tree= geospiza.tree, scale=2)
```

---

comp.fit.subs *comp.fit.subs*

Description

The function implements the K-clades parametric rate comparison test. This function compares rate estimates among defined subtrees and evaluates various groupings from 1 to k groups for these subtrees.

Usage

```
comp.fit.subs(trees, focal, k, mod.id = c(1, 0, 0, 0), min.val = 0.01)
```

Arguments

trees	A list from from function id.subtrees.
focal	A vector indicating the subtrees to compare
k	A value indicting the maximum number of groupings of subtrees to examine
mod.id	A vector with four elements of 0 or 1 indicating which models to consider. 1 indicates that the model should be considered. 0 indicates the model is not considered. These for elements refer to an exponential, Weibel, lognormal, and rate variable, respectively.
min.val	A value for determining the minimum edge length for a tree scaled against the longest edge length. A value of 0.01 (the default) rescales the minimum edge length to 1

Details

The list of possible subtrees is provided by the function id.subtrees. The function will explore all possible groupings of subtrees into k defined groups choosing the best fit model for each partition from among the models identified by mod.id.

Value

A dataframe that consists of the following:

k	The number of groups
Groups	The groupings for each subtree numbered as 1 to the number of subtrees indicated. The numbering corresponds to the order in which subtrees are identified by focal. Groups are separated with vs.
gi_Pj	The jth parameter value for the ith group in the analysis
gi_mod.id	The best model chosen for the ith group
gi_n.param	The number of parameters in the best model for the ith group
AIC	Akaike information criterion score for the entire model for a grouping scheme
AICc	Akaike information criterion corrected for sample size
dAICc	The delta AIC across all grouping schemes and k values relative to the best fit model

Note

The output can get very large as k increases. Function tab.summary is useful for reducing the size of the result table.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

`tab.summary` `id.subtrees`

Examples

```
data(hivtree.newick)
cat(hivtree.newick, file = "hivtree.phy", sep = "\n")
tree.hiv <- read.tree("hivtree.phy") # load tree
unlink("hivtree.phy") # delete the file "hivtree.phy"

idHIV<-id.subtrees(tree.hiv)
plot(idHIV$tree,show.node.label=TRUE)
cfsHIV<-comp.fit.subs(idHIV$subtree,focal=c(153,119,96,5),k=4)
```

comp.subs *comp.subs*

Description

The function implements the parametric rate comparison test. The function iterates through all subtrees of a phylogenetic tree and compares the distribution of branch lengths in the subtree to the "remainder" tree. It is intended to be used with a chronogram in order to test whether diversification rates differ among clades within a broader phylogeny. A variety of truncated distributions can be used and compared via likelihood.

Usage

```
comp.subs(tree, thr = 6, srt = "drop", min.val = 0.01,
mod.id = c(1, 0, 0, 0),verbose=TRUE)
```

Arguments

tree	An object of class phylo. To test variation in diversification rates, this should be a chronogram.
thr	Threshold subtree or remainder tree size below which comparisons should not be performed. thr is the minimum number of edges (in either the subtree or remainder tree) for a comparison to be made.
srt	Treatment of subtree root edge. Default is "drop" meaning the edge subtending each subtree will be left out of the comparison for that subtree. Alternatives "in" or "out" classify the subtree root edge as part of the subtree or part of the remainder tree, respectively.
min.val	Replacement of zero-length branches with a small positive number to avoid spurious zeros in likelihood calculations. This value is treated as a fraction of the maximum branch (it is multiplied by the maximum edge length and that resultant is substituted for zero-length branches in tree
mod.id	Indicator vector specifying statistical distributions to be fit to the data. In order, the distributions are exponential, Weibull, lognormal, and variable rates *Venditti et al. 2010*. Default is exponential only.
verbose	A logical indicating whether progress is updated on the screen

Details

All distributions are fit using the likelihood for the truncated form

Value

A data frame containing up to 15 variables for each subtree of tree. Each row corresponds to a subtree and the order is that returned by the function subtrees. Subtrees that are not tested (owing to failure to meet the thr threshold) have NA's for all variables:

Par1.tot	First estimated parameter of the best fit model for the pooled edge lengths of the subtree and remainder tree. For exponential, this is the rate. For Weibull it is the "shape" parameter. For lognormal it is mu. For the variable rates distribution it is alpha.
Par2.tot	Second estimated parameter of the best fit model for the pooled edge lengths. For exponential, it is NA. For Weibull it is the "scale" parameter. For lognormal, it is sigma. For variable rates, it is beta.
Par1.tr1	First estimated parameter for the best fit model for the subtree
Par2.tr1	Second estimated parameter for the best fit model for the subtree
Par1.tr2	First estimated parameter for the best fit model for the remainder tree
Par2.tr2	Second estimated parameter for the best fit model for the remainder tree
llk.1r	log likelihood of the best fit model for the pooled set of edges: the one-rate model.
llk.2r	log likelihood for the best two-rate model
mod.1r.tot	Best fit distribution for the one-rate model: 1=exponential, 2=Weibull, 3=log-normal, 4=variable rates
mod.2r.tr1	Best fit distribution for the subtree under the two-rate model
mod.2r.tr2	Best fit distribution for the remainder tree under the two-rate model
node1	Identifies the node corresponding to the most recent common ancestor of the subtree and its sister clade. That is, the node ancestral to the branch along which a rate change might have occured.
node2	Identifies the most recent common ancestor of all taxa in the subtree. That is, the descendant node of the branch along which a rate chage might have occurred.
p.val	P-value from the likelihood ratio test of the two-rate vs. one-rate model for the subtree defined by node2
EvidRatio	The evidence ratio from the AICc scores of the two-rate vs. one-rate model for the subtree defined by node2

Author(s)

Premal Shah, James A. Fordyce, Benjamin M. Fitzpatrick

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013 A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377. Venditti, C., A. Meade, and M. Pagel, 2010. Phylogenies reveal new interpretation of speciation and the red queen. Nature 463:349-352.

Examples

```
data(geospiza)
attach(geospiza)
comp.subs(geospiza.tree)
```

FP.comp.subs *FP.comp.subs*

Description

This function simulates pure birth trees with a given number of taxa and NA subtrees and calculates the null expectation for the number of significant rate differences.

Usage

```
FP.comp.subs(tree.size, na.present, sims = 100, missing = 0,
alpha = 0.05, verbose = FALSE, ...)
```

Arguments

tree.size	A value for the number of terminal taxa in the tree to simulate.
na.present	A value for the number of NA subtrees in the simulated trees.
sims	A value for the number of trees to simulate.
missing	A value indicating the number of missing taxa from the tree.
alpha	A value indicating the threshold for statistical significance.
verbose	A boolean indicating whether a summary of the simulations is printed to the screen.
...	Arguments passed on to comp.subs function

Details

This function is useful if the user wants to know the expected number of significant rate differences for a tree of a given size and number of NA subtrees. This function calls on comp.subs, and arguments can be passed on.

Value

A list that consists of the following:

tree.size	The number of terminal taxa provided by the user.
missing	The number of missing taxa from the tree.
sims	The number of simulated trees.
FPRthres	The number of significant rate difference detections expected based upon the alpha value provided by the user.

Note

comp.subs is an exploratory data analysis tool and concerns of false positives should be considered accordingly. The argument "missing" can be used for trees with incomplete taxon sampling. Thus, if a group should have 100 taxa included, but only 90 are present in the tree, tree.size=100 and missing=10.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

`comp.subs`

Examples

```
## Not run:
data(geospiza)
tree<-geospiza$geospiza.tree
na.count<-tree.na.Count(tree)
FP.comp.subs(tree.size=14,na.present=na.count,verbose=TRUE)

## End(Not run)
```

`id.subtrees` *id.subtrees*

Description

This function identifies and numbers all subtrees within a tree of object class phylo. It creates the object required for function `comp.fit.subs`.

Usage

```
id.subtrees(tree)
```

Arguments

`tree`	A tree of object class phylo.

Details

This function identifies all the subtrees in a tree. These identifiers are used to identify the focal subtrees used in the `comp.fit.subs` function.

Value

A list that consists of the following:

`tree`	The original tree as object class phylo with nodes labeled identifying the identification number for all subtrees.
`subtree`	A list of all possible subtrees as object class phylo.

Note

This function will rename all node labels.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

`comp.fit.subs`

Examples

```
## Not run:
data(hivtree.newick)
cat(hivtree.newick, file = "hivtree.phy", sep = "\n")
tree.hiv <- read.tree("hivtree.phy") # load tree
unlink("hivtree.phy") # delete the file "hivtree.phy"

idHIV<-id.subtrees(tree.hiv)
plot(idHIV$tree,show.node.label=TRUE)

## End(Not run)
```

`tab.summary` *tab.summary*

Description

This function provides an abridged output of results obtained from the `comp.fit.subs` function by restricting the output to a user provided delta AIC threshold.

Usage

```
tab.summary(res, daic = 2, show.rate = FALSE)
```

Arguments

`res`	A dataframe obtained from `comp.fit.subs` function.
`daic`	A value indicating a threshold of delta AIC relative to the best fit model for each k to be included in the output.
`show.rate`	A boolean indicting whether the rate parameters are included in the output.

Details

This function will provide a reduced output of the results provided by the `comp.fit.subs` function by allowing the user to choose a critical delta AIC for each value of k that determines which comparisons are included in the output. The best fit model for each k is included in the output regardless of delta AIC. The show.rate argument indicates whether the rate estimate for each of the subtrees is included in the output.

Value

A dataframe that consists of the following:

`k`	The number of groups
`Groups`	the groupings for each subtree numbered as 1 to the number of subtrees indicated. The numbering corresponds to the order in which subtrees are identified by focal. Groups are separated with 'vs.'.
`gi_rate`	The rate for the ith group in the analysis.
`LL`	The log likelihood for the entire model for a grouping scheme.
`AIC`	Akaike information criterion score for the entire model for a grouping scheme.
`AICc`	Akaike information criterion corrected for sample size.
`dAICc`	The delta AIC across all grouping schemes and k values relative to the best fit model.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

`tab.summary` `id.subtrees`

Examples

```
## Not run:
data(hivtree.newick)
cat(hivtree.newick, file = "hivtree.phy", sep = "\n")
tree.hiv <- read.tree("hivtree.phy") # load tree
unlink("hivtree.phy") # delete the file "hivtree.phy"

idHIV<-id.subtrees(tree.hiv)
plot(idHIV$tree,show.node.label=TRUE)
cfsHIV<-comp.fit.subs(idHIV$subtree,focal=c(153,119,96,5),k=4)
tab.summary(cfsHIV)
tab.summary(cfsHIV,daic=1)
```

```
tab.summary(cfsHIV,daic=0.01)

## End(Not run)
```

`tree.na.Count` *tree.na.Count*

Description

This function will identify the number of NA subtrees present in a given phylogenetic tree.

Usage

```
tree.na.Count(tree, thr = 6, srt = "drop", min.val = 0.01,
mod.id = c(1, 0, 0, 0))
```

Arguments

`tree`	A tree of object class phylo.
`thr`	The threshold for the minimum number of edges to be used for calculating the rate of a subtree.
`srt`	Determines how the edge leading to a subtree is dealt with when calculating rates. The default, "drop", excludes the edge leading to the subtree from the analysis. "in" will include the edge as part of the subtree and "out" will include the edge as part of the remaining tree.
`min.val`	A value for determining the minimum edge length for a tree scaled against the longest edge length. A value of 0.01 (the default) rescales the minimum edge length to 1
`mod.id`	A vector with four elements of 0 or 1 indicating which models to consider. 1 indicates that the model should be considered. 0 indicates the model is not considered. These for elements refer to an exponential, Weibel, lognormal, and rate variable, respectively.

Details

This function identifies the number of NA subtrees present in a given phylogenetic tree. This information might be useful if the user is interested in simulating trees with the same amount of information (i.e., useable edges) for calculating rates.

Value

A number indicating the number of NAs in the given tree.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

See Also

FP.comp.subs

Examples

```
## Not run:
data(geospiza)
tree<-geospiza$geospiza.tree
tree.na.Count(tree)

## End(Not run)
```

`tree.rand.test` *tree.rand.test*

Description

This function performs a randomization test for rate variation among clades.

Usage

```
tree.rand.test(tree, reps=1000, mod.id=c(1,0,0,0), trace=TRUE)
```

Arguments

`tree`	An ultrametric tree of object class phylo.
`reps`	Desired number of randomizations
`mod.id`	Indicator vector specifying statistical distributions to be fit to the data. In order, the distributions are exponential, Weibull, lognormal, and variable rates *Venditti et al. 2010.* Default is exponential only.
`trace`	If true, progress will be indicated by printing to the screen.

Details

This function addresses the potential for spurious inference of diversification rate variation when a phylogeny deviates from the pure birth model. Deviation from pure birth (e.g., when extinction is important or speciation probabilities change over time) distorts the distribution of branching times such that internode lengths do not satisfy the independent and identical distribution (iid) assumption of the PRC test. This function distinguishes among-clade rate variation from rate variation through time by holding the set of branching times constant and randomizing tree topologies. That is, it simulates the null hypothesis that speciation and extinction probabilities are constant across lineages at any given time. The function provides a null distribution for the false detection rate - the fraction of subtrees appearing to have deviant diversification rates when there is no true among-clade rate variation.

Value

A list that consists of the following:

tree	The original tree as object class phylo.
obs.p	Observed set of p-values from comp.subs.
ncs	A (potentially large) list of output (p-values and evidence ratios) from each randomization.
obs.detection	Detection rate for the observed tree. This is the fraction of qualified subtrees with rate variation according to a p-value less than 0.05
p.detection	The fraction of null trees that have more detections than the observed.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

Examples

```
## Not run:
data(geospiza)
tree <- geospiza$geospiza.tree
tree.rand.test(tree,reps=50) # few reps used to illustrate without taking too much time

## End(Not run)
```

---

trimTree *trimTree*

Description

This function will trim a specified amount of time, or branch length, from the tips of an ultrametric tree.

Usage

```
trimTree(phy, Time)
```

Arguments

phy	An ultrametric tree of object class phylo.
Time	A value indicating the amount of branch length (time) to be removed from the tips of the tree

Details

This function is useful if there is some ambiguity regarding the resolution of the tips. This might include possible over-splitting of taxa, or incomplete taxon sampling. For example, it might be desirable to analyze a tree where the most recent 1 million years is excluded to account for the possibility of incomplete sampling. It is important to note that analyses conducted on the trimmed tree is based on lineages that are still extant and cannot account for lineages that might have been present at the time of the trimming but has subsequently gone extinct.

Value

A list that consists of the following:

`o.tree`	The original tree as object class phylo.
`t.tree`	The tree after the designated amount of branch length has been trimmed from the tips as object class phylo.
`new.tip.clades`	A vector in the t.tree phylo object that gives the tip names following trimming that identifies the original tip names in the newly defined clades.

Author(s)

Premal Shah, Benjamin Fitzpatrick and James Fordyce.

References

Shah, P., B. M. Fitzpatrick, and J. A. Fordyce. 2013. A parametric method for assessing diversification rate variation in phylogenetic trees. Evolution 67:368-377.

Examples

```
## Not run:
data(hivtree.newick)
cat(hivtree.newick, file = "hivtree.phy", sep = "\n")
tree.hiv <- read.tree("hivtree.phy") # load tree
unlink("hivtree.phy") # delete the file "hivtree.phy"

trim.hiv<-trimTree(phy=tree.hiv,Time=0.1)#trims 0.1 branchlength units from the tree
par(mfrow=c(1,2))
plot.phylo(trim.hiv$o.tree);plot.phylo(trim.hiv$t.tree)

# Identify the names of the original terminal taxa
# that correspond to the newly defined, numbered tips.
trim.hiv$t.tree$new.tip.clades

## End(Not run)
```

Index