

Package ‘RERconverge’

October 22, 2018

Type Package

Title Analysis of Convergence Between Organismal Traits and DNA/Protein Sequences

Version 0.1.0

Author Maria Chikina, Nathan Clark, Amanda Kowalczyk, Weiguang (Wayne) Mao, Wynn Meyer, and Raghavendran Partha

Maintainer Maria Chikina <mchikina@pitt.edu> / Nathan Clark <nclark@pitt.edu>

Description RERConverge is a set of software written in R that estimates the correlation between relative evolutionary rates of genes and the evolution of a convergent binary or continuous trait across a phylogeny.

License GNU General Public License v3.0

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Depends RColorBrewer, gplots, phytools, geiger, knitr, Rcpp, RcppArmadillo, weights, phangorn, ggplot2, dplyr

Imports Rcpp (>= 0.11.0), ape

LinkingTo Rcpp, RcppArmadillo, ape

R topics documented:

char2Paths	2
click_select_foreground_branches	3
correlateWithBinaryPhenotype	3
correlateWithContinuousPhenotype	4
edgeVars	5
fastwilcoxGMTall	5
foreground2Paths	6
foreground2Tree	7
getAllCor	7
getAllResiduals	8
getStat	9
makeBinaryPaths	10

namePathsWSpecies	10
plotRers	11
plotTreeHighlightBranches	11
pruneTree	12
read.gmt	13
readTrees	13
RERconverge	14
tree2Paths	14

Index	15
--------------	-----------

char2Paths	<i>turns a named vector of characters into a paths vector to be used with getAllCor</i>
------------	---

Description

turns a named vector of characters into a paths vector to be used with [getAllCor](#)

Usage

```
char2Paths(tip.vals, treesObj, altMasterTree = NULL, metric = "diff",
           se.filter = -1, ...)
```

Arguments

tip.vals	the trait/phenotype/character value at the tip, names(tip.vals) should match some of the mastertree\$tip.label, though a perfect match is not required
treesObj	A treesObj created by readTrees
metric	The metric used to translate node values into branch values. "Diff" takes the difference and makes the result phylogenetically independent. Other possible values are "mean" (the mean of the two values) and "last" the value of the most recent species on the branch. The last two options are not phylogenetically independent and downstream computations for those are not yet implemented
se.filter	Will remove branch values that are not at least se.filter*edge.se away from 0 (where edge.se is the standard error in the estimate for the edge value). Only implemented for metric="diff". Set se.filter to a positive value to filter. By default no filtering is done.

Value

A vector of length equal to the number of paths in treesObj

```
click_select_foreground_branches
```

Interactive click-based function to select foreground branches showing convergent binary trait

Description

Interactive click-based function to select foreground branches showing convergent binary trait

Usage

```
click_select_foreground_branches(mastertree)
```

Arguments

`mastertree.` A phylo tree object defining the topology of all species

Value

A binary trait tree with branch lengths of 1 for selected foreground species and lengths of 0 for the rest

```
correlateWithBinaryPhenotype
```

Computes the association statistics between RER from [getAllResiduals](#) and a phenotype paths vector for a binary phenotype made with [tree2Paths](#)

Description

Computes the association statistics between RER from [getAllResiduals](#) and a phenotype paths vector for a binary phenotype made with [tree2Paths](#)

Usage

```
correlateWithBinaryPhenotype(RERmat, charP, min.sp = 10, min.pos = 2,
  weighted = "auto")
```

Arguments

<code>RERmat</code>	RER matrix returned by getAllResiduals
<code>charP</code>	phenotype vector returned by tree2Paths or char2Paths
<code>min.sp</code>	Minimum number of species that must be present for a gene
<code>min.pos</code>	Minimum number of species that must be present in the foreground (non-zero phenotype values)

weighted	perform weighted correlation. This option turns on weighted correlation that uses the weights computed by <code>foreground2Tree(wholeClade=T)</code> . This setting will treat each clade a single observation for the purpose of p-value estimation. The function will guess automatically if the <code>charP</code> vector is of "weighted" type and there should be not need to set this parameter.
method	Method used to compute correlations. Accepts the same arguments as <code>cor</code> . Set to "auto" to select automatically based on the number of unique values in <code>charP</code> . This will also auto set the winsorization for Pearson correlation. Set <code>winsorize=some number</code> to override

```
correlateWithContinuousPhenotype
```

Computes the association statistics between RER from `getAllResiduals` and a phenotype paths vector for a continuous phenotype made with `char2Paths`

Description

Computes the association statistics between RER from `getAllResiduals` and a phenotype paths vector for a continuous phenotype made with `char2Paths`

Usage

```
correlateWithContinuousPhenotype(RERmat, charP, min.sp = 10,
  winsorize = 3)
```

Arguments

RERmat	RER matrix returned by <code>getAllResiduals</code>
charP	phenotype vector returned by <code>tree2Paths</code> or <code>char2Paths</code>
min.sp	Minimum number of species that must be present for a gene
winsorize	Winsorize values before computing Pearson correlation. Winsorize=3, will set the 3 most extreme values at each end to the the value closest to 0.
method	Method used to compute correlations. Accepts the same arguments as <code>cor</code> . Set to "auto" to select automatically based on the number of unique values in <code>charP</code> . This will also auto set the winsorization for Pearson correlation. Set <code>winsorize=some number</code> to override

edgeVars	<i>maps a vector of traits onto a reference tree</i>
----------	--

Description

maps a vector of traits onto a reference tree

Usage

```
edgeVars(mastertree, tip.vals, metric = "diff", se.filter = -1,
         return.var = F)
```

Arguments

mastertree	the tree species the topology of the output tree and the branch lengths are used to infer ancestral states
tip.vals	the trait/phenotype/character value at the tip, <code>names(tip.vals)</code> should match some of the <code>mastertree\$tip.label</code> , though a perfect match is not required
metric	The metric used to translate node values into branch values. "Diff" takes the difference and makes the result phylogenetically independent. Other possible values are "mean" (the mean of the two values) and "last" the value of the most recent species on the branch. The last two options are not phylogenetically independent and downstream computations for those are not yet implemented
se.filter	Will remove branch values that are not at least <code>se.filter*edge.se</code> away from 0 (where <code>edge.se</code> is the standard error in the estimate for the edge value). Only implemented for <code>metric="diff"</code> . Set <code>se.filter</code> to a positive value to filter. By default no filtering is done.
return.var	Returns the variance instead of the mean. Useful for seeing which estimates have high confidence.

Value

A phylo tree with branch values computed from the input tip.values

fastwilcoxGMTall	<i>Performs Wilcoxon Rank-Sum pathway enrichment on pathway genesets</i>
------------------	--

Description

Performs Wilcoxon Rank-Sum pathway enrichment on pathway genesets

Usage

```
fastwilcoxGMTall(vals, annotList, ...)
```

Arguments

vals	The output from getStat run on RERconverge correlation functions (correlateWithContinuousPhenotype, correlateWithBinaryPhenotype, getAllCor)
annotList	Pathway annotations formatted as a list of lists of annotation information

Value

A list of pathways with enrichment statistics

foreground2Paths	<i>Creates paths from a set of foreground species</i>
------------------	---

Description

Creates paths from a set of foreground species

Usage

```
foreground2Paths(foreground, treesObj, plotTree = F,
  clade = c("ancestral", "terminal", "all", "weighted"),
  useSpecies = NULL)
```

Arguments

treesObj	A treesObj created by readTrees
plotTree	Plot a tree representation of the result
clade	A character string indicating which branches within the clade containing the foreground species should be set to foreground. Must be one of the strings "ancestral", "terminal", "all", or "weighted".
useSpecies	Give only a subset of the species to use for ancestral state reconstruction (e.g., only those species for which the trait can be reliably determined).
foreground.	A character vector containing the foreground species

Value

A vector of length equal to the number of paths in treesObj

foreground2Tree	<i>Creates a binary trait tree from a set of foreground species.</i>
-----------------	--

Description

Creates a binary trait tree from a set of foreground species.

Usage

```
foreground2Tree(foreground, treesObj, collapse2anc = T, plotTree = T,
  wholeClade = F, clade = c("ancestral", "terminal", "all",
    "weighted"), useSpecies = NULL)
```

Arguments

treesObj	A treesObj created by readTrees
collapse2anc	Put all the weight on the ancestral branch when the trait appears on a whole clade (redundant to "clade", kept for backwards compatibility)
plotTree	Plot a tree representation of the result
wholeClade	Whether to implement the weighted edge option across all members of a foreground clade (redundant to "clade", kept for backwards compatibility)
clade	A character string indicating which branches within the clade containing the foreground species should be set to foreground. Must be one of the strings "ancestral", "terminal", "all", or "weighted".
useSpecies	Give only a subset of the species to use for ancestral state reconstruction (e.g., only those species for which the trait can be reliably determined).
foreground.	A character vector containing the foreground species

Value

A tree with edge.lengths representing phenotypic states

getAllCor	<i>Computes the association statistics between RER from getAllResiduals and a phenotype paths vector made with tree2Paths or char2Paths</i>
-----------	---

Description

Computes the association statistics between RER from [getAllResiduals](#) and a phenotype paths vector made with [tree2Paths](#) or [char2Paths](#)

Usage

```
getAllCor(RERmat, charP, method = "auto", min.sp = 10, min.pos = 2,
  winsorize = NULL, weighted = F)
```

Arguments

RERmat	RER matrix returned by getAllResiduals
charP	phenotype vector returned by tree2Paths or char2Paths
method	Method used to compute correlations. Accepts the same arguments as cor . Set to "auto" to select automatically based on the number of unique values in charP. This will also auto set the winsorization for Pearson correlation. Set winsorize=some number to override
min.sp	Minimum number of species that must be present for a gene
min.pos	Minimum number of species that must be present in the foreground (non-zero phenotype values)
winsorize	Winsorize values before computing Pearson correlation. Winsorize=3, will set the 3 most extreme values at each end to the the value closest to 0.
weighted	perform weighted correlation. This option needs to be set if the clade weights computed in foreground2Tree (wholeClade=T) are to be used. This setting will treat the clade a single observation for the purpose of p-value estimation.

Value

A list object with correlation values, p-values, and the number of data points used for each tree

Note

winsorize is in terms of number of observations at each end, NOT quantiles

getAllResiduals	<i>main RER computation function</i>
-----------------	--------------------------------------

Description

main RER computation function

Usage

```
getAllResiduals(treesObj, cutoff = NULL, transform = "sqrt",
  weighted = T, useSpecies = NULL, min.sp = 10, scale = T,
  doOnly = NULL, maxT = NULL, scaleForPproj = F, mean.trim = 0.05)
```


Arguments

treesObj	A treesObj created by readTrees
transform	The transformation to apply to the trees branch values before computing relative rates. Available options are sqrt and log, sqrt is recommended.
weighted	Use weighted regression to compute relative rates, meant to correct for the non-constant mean-variance relationship in evolutionary rate data.
useSpecies	Give only a subset of the species to use for RER calculation. Some times excluding unusually long branches can provide more stable results
min.sp	The minimum number of species needed to compute RER
scale	Scale relative rates internally for each species subset. Increases computation time with little apparent benefit. Better to scale the final matrix.
doOnly	The index of a specific tree in the treesObj to calculate RER for. Useful if a single result is needed quickly.
maxT	The maximum number of trees to compute results for. Since this function takes some time this is useful for debugging.
a	cutoff value for branch lengths below which the branch lengths will be discarded, very data dependent but should roughly correspond to 0 or 1 sequence change on that branch. If left NULL this will be set to the bottom 0.05 quantile. Set to 0 for no cutoff.

Value

A number of trees by number of paths matrix of relative evolutionary rates. Only an independent set of paths has non-NA values for each tree.

getStat	<i>Calculates Rho-signed negative log-base-ten p-value for use in enrichment functions</i>
---------	--

Description

Calculates Rho-signed negative log-base-ten p-value for use in enrichment functions

Usage

```
getStat(res)
```

Arguments

res	The output from RERconverge correlation functions (correlateWithContinuousPhenotype, correlateWithBinaryPhenotype, getAllCor)
-----	---

Value

A dataframe of Rho-signed negative log-base-ten p-values for all genes, NAs removed

makeBinaryPaths	<i>Makes a binary path vector from either a tree of class "phylo" or a foreground species set supplied as a character vector</i>
-----------------	--

Description

Makes a binary path vector from either a tree of class "phylo" or a foreground species set supplied as a character vector

Usage

```
makeBinaryPaths(input, treesObj)
```

Arguments

input	Either a phenotype tree of class "phylo" (with branch length encoding a phenotype) or a character vector of foreground branches
treesObj	A treesObj created by readTrees

Value

A vector of length equal to the number of paths in treesObj

namePathsWSpecies	<i>Provides names for paths/RERs representing terminal branches for plotting Originally an internal function but necessary for the vignette/walk-through</i>
-------------------	--

Description

Provides names for paths/RERs representing terminal branches for plotting Originally an internal function but necessary for the vignette/walk-through

Usage

```
namePathsWSpecies(masterTree)
```

Arguments

masterTree	The master tree used for analysis
------------	-----------------------------------

Value

Names corresponding to the paths/RERs for terminal branches

plotRers	<i>Plot the residuals reflecting the relative evolutionary rates (RERs) of a gene across species present in the gene tree</i>
----------	---

Description

Plot the residuals reflecting the relative evolutionary rates (RERs) of a gene across species present in the gene tree

Usage

```
plotRers(rermat = NULL, index = NULL, phenv = NULL, rers = NULL,
         method = "k", xlims = NULL, plot = 1, xextend = 0.2,
         sortrers = F)
```

Arguments

rermat.	A residual matrix, output of the <code>getAllResiduals()</code> function
index.	A character denoting the name of gene, or a numeric value corresponding to the gene's row index in the residuals matrix
phenv.	A phenotype vector returned by tree2Paths or foreground2Paths

Value

A plot of the RERs with foreground species labelled in red, and the rest in blue

plotTreeHighlightBranches	<i>Plot the provided tree, (optionally) rerooted, with specified branches highlighted</i>
---------------------------	---

Description

Plot the provided tree, (optionally) rerooted, with specified branches highlighted

Usage

```
plotTreeHighlightBranches(tree, outgroup = NULL, hlspecies,
                          hlcols = NULL, main = "")
```

Arguments

<code>tree.</code>	A tree object.
<code>outgroup.</code>	A vector of species to use to root the tree. If not provided, the tree remains unrooted.
<code>hlspecies.</code>	A vector of species whose terminal branches to highlight, or a vector of branch numbers within the tree.
<code>hlcols.</code>	Colors to use in highlighting the branches. If not specified, will use default R colors.
<code>main.</code>	Main text for plot.

Value

A plot of the the (optionally rerooted) tree, with branches highlighted.

<code>pruneTree</code>	<i>wrapper around</i> <code>drop.tip</code>
------------------------	---

Description

wrapper around `drop.tip`

Usage

```
pruneTree(tree, tip.names)
```

Arguments

<code>tree</code>	a "phylo" tree
<code>tip.names</code>	The tip names to keep in the tree

Value

A new pruned tree

read.gmt	<i>Reads pathway data from a gmt-formatted file from GSEA-MSigDB</i>
----------	--

Description

Reads pathway data from a gmt-formatted file from GSEA-MSigDB

Usage

```
read.gmt(filename)
```

Arguments

filename	The gmt-formatted pathway file
----------	--------------------------------

Value

A list of pathway information with pathway names and genesets

readTrees	<i>reads trees from a 2 column , tab seperated, file The first columns is the gene name and the second column is the corresponding tree in parenthetic format known as the Newick or New Hampshire format</i>
-----------	---

Description

reads trees from a 2 column , tab seperated, file The first columns is the gene name and the second column is the corresponding tree in parenthetic format known as the Newick or New Hampshire format

Usage

```
readTrees(file, max.read = NA)
```

Arguments

file	The path to the tree file
max.read	this function takes a while for a whole genome so max.read is useful for testing

Value

A trees object of class "treeObj"

RERconverge	<i>RERconverge</i>
-------------	--------------------

Description

RERconverge

tree2Paths	<i>Generate a phenotype paths vector from a phenotype tree</i>
------------	--

Description

tree2Paths generates a phenotype paths vector matching the treesObject from a tree where branches specify phenotypes.

Usage

```
tree2Paths(tree, treesObj, binarize = NULL, useSpecies = NULL)
```

Arguments

tree	A phenotype tree, with branch length encoding a phenotype.
treesObj	A treesObject created by readTrees
binarize	Force binary path representation. Default action depends upon the type of data within the phenotype tree (binary or continuous). <ul style="list-style-type: none"> • If binary (all branch lengths == 0 or 1): Sets all positive path values to 1. Useful if the tree has non-zero branch lengths for an internal branch or branches; otherwise, values are simply added along branches when calculating paths. Default behavior: binarize = TRUE. • If continuous (not all branch lengths == 0 or 1): Sets all path values > the mean to 1 and all those <= the mean to 0. Converts a continuous phenotype to a binary phenotype, with state determined by comparison to the mean across all paths. Default behavior: binarize = FALSE.
useSpecies	Give only a subset of the species to use for ancestral state reconstruction (e.g., only those species for which the trait can be reliably determined).

Details

The tree topology of the phenotype tree must match that of the master tree within the treesObject.

Value

A vector of length equal to the number of paths in treesObj

Index

char2Paths, [2](#), [3](#), [4](#), [7](#), [8](#)
click_select_foreground_branches,
 [3](#)
cor, [4](#), [8](#)
correlateWithBinaryPhenotype, [3](#)
correlateWithContinuousPhenotype,
 [4](#)

drop.tip, [12](#)

edgeVars, [5](#)

fastwilcoxGMTall, [5](#)
foreground2Paths, [6](#), [11](#)
foreground2Tree, [4](#), [7](#), [8](#)

getAllCor, [2](#), [7](#)
getAllResiduals, [3](#), [4](#), [7](#), [8](#), [8](#)
getStat, [9](#)

makeBinaryPaths, [10](#)

namePathsWSpecies, [10](#)

plotRers, [11](#)
plotTreeHighlightBranches, [11](#)
pruneTree, [12](#)

read.gmt, [13](#)
readTrees, [2](#), [6](#), [7](#), [9](#), [10](#), [13](#), [14](#)
RERconverge, [14](#)
RERconverge-package
 (*RERconverge*), [14](#)

tree2Paths, [3](#), [4](#), [7](#), [8](#), [11](#), [14](#)