# Package 'RERconverge'

August 14, 2021

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 7.1.1

**Depends** R (>= 3.5.0), RColorBrewer, gplots, phytools, geiger, knitr, Rcpp, RcppArmadillo, weights, phangorn, ggplot2, dplyr

**Imports** Rcpp (>= 0.11.0), ape

LinkingTo Rcpp, RcppArmadillo, ape

RemoteType github

RemoteHost api.github.com

RemoteRepo RERconverge

RemoteUsername nclark-lab

RemoteRef master

RemoteSha db53e0799070b30239768edfd8d596454dfb78d6

**GithubRepo** RERconverge **GithubUsername** nclark-lab

GithubRef master

GithubSHA1 db53e0799070b30239768edfd8d596454dfb78d6

NeedsCompilation yes

## $\mathsf{R}$ topics documented:

calculateCorPermuted	3
calculatePermulatedPaths_apply	4
char2Paths	4
click_select_foreground_branches	5
combinePermData	6
correlateWithBinaryPhenotype	6
correlateWithContinuousPhenotype	7
edgeVars	8
estimatePhangornTree	9
estimatePhangornTreeAll	10
fastwilcoxGMTall	11
fixPseudoroot	11
foreground2Paths	12
foreground2Tree	13
foreground2TreeClades	14
generatePermulatedBinPhen	14
generatePermulatedBinPhenSSMBatched	15
getAllCor	16
getAllResiduals	17
getEnrichPerms	18
getForegroundInfoClades	19
getPermsBinary	20
getPermsContinuous	21
getStat	22
makeBinaryPaths	23
matchAllNodesClades	23
namePathsWSpecies	24
permpvalcor	24
permpvalenrich	25
permutevec	25
plotPositivesFromPermulations	26
plotRers	26
plotTreeHighlightBranches	27
pruneTree	28
read.gmt	29
readTrees	29
RERconverge	30
returnRersAsNewickStrings	30
returnRersAsTree	31
returnRersAsTreesAll	32
simBinPheno	32
simBinPhenoCC	33
simBinPhenoSSM	34
simBinPhenoVec	34
simpermvec	35
simulatevec	36

calculateCorPermuted 3

Index					41
	treePlotNew	 	 	 	 38

calculateCorPermuted

Calculate permulation correlation statistics

## **Description**

Calculate permulation correlation statistics

#### Usage

```
calculateCorPermuted(
  permulated.paths,
  RERmat,
  min.sp = 10,
  min.pos = 2,
  method = "k"
)
```

## Arguments

method

permulated.paths

 $A \ nested \ list \ of \ permulated \ paths \ (e.g., output \ of \ \texttt{calculatePermulatedPaths\_apply}$ 

An RER matrix calculated using getAllResiduals.

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)

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 winsorizetrait=some number and winsorizeRER=some number to override

#### Value

A nested list containing the correlation statistics for the permulations

4 char2Paths

```
calculatePermulatedPaths_apply
```

Calculates the paths for all permulated trees for a list of genes

#### **Description**

Calculates the paths for all permulated trees for a list of genes

#### Usage

```
calculatePermulatedPaths_apply(permulated.trees.list, map.list, treesObj)
```

## **Arguments**

```
A nested list of permulated phenotype trees for all the genes. The top layer lists the genes, and the nested layer lists the permulated binary trees (phylo objects) for each gene

map.list A list of maps corresponding to the genes listed in permulated.trees.list, the output of matchAllNodesClades

treesObj treesObj from readTrees
```

#### Value

permulated.paths.list A nested list of permulated paths corresponding to the nested list of permulated phenotype trees

char2Paths

turns a named vector of characters into a paths vector to be used with getAllCor

#### **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 val	lue at the tip, names	(tip.vals)	should match
----------	-----------------------------------	-----------------------	------------	--------------

some of the mastertree\$tip.label, though a perfect match is not re-

quired

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 (mastertre)
```

#### **Arguments**

mastertre. 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

combinePermData

Combines batches of permulations

## Description

Combines batches of permulations

## Usage

```
combinePermData(permdat1, permdat2, enrich = T)
```

## Arguments

permdat1 Batch of permulations output from getPermsContinuous

permdat2 Batch of permulations output from getPermsContinuous

enrich Default T. Specifies if both 'permdat1' and 'permdat2' contain enrichment per-

mulations

#### Value

Combined permulations

```
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**

RERmat	RER matrix returned by getAllResiduals
charP	phenotype vector returned by tree2Paths or char2Paths
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)
weighted	perform weighted correlation. This option turns on weighted correlation that uses the weights computed by <pre>foreground2Tree</pre> (wholeClade=T). This setting will treat each clade a single observation for the purpose of p-value estimation. The function will guess automatically if the charP 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 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 winsorizetrait=some number and winsorizeRER=some number 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,
  winsorizeRER = 3,
  winsorizetrait = 3
)
```

## **Arguments**

RERmat RER matrix returned by getAllResiduals

charP phenotype vector returned by tree2Paths or char2Paths

min.sp Minimum number of species that must be present for a gene

winsorizeRER Winsorize RER values before computing Pearson correlation. winsorizeRER=3 will set the 3 most extreme values at each end of each RER row to the the value

closest to 0.

8 edge Vars

winsorizetrait

Winsorize trait values before computing Pearson correlation. winsorizetrait=3 will set the 3 most extreme values of the trait values to the value closest to 0.

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 winsorizetrait=some number and winsorizeRER=some number to override

edgeVars

maps a vector of traits onto a reference tree

## 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, names (tip.vals) should match some of the mastertree\$tip.label, 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 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.
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

estimatePhangornTree 9

```
estimatePhangornTree
```

Estimate a ML tree from a given topology Uses 'pml' and 'optim.pml' from the 'phangorn' package to estimate the tree.

## Description

Estimate a ML tree from a given topology Uses 'pml' and 'optim.pml' from the 'phangorn' package to estimate the tree.

## Usage

```
estimatePhangornTree(
  alnfile,
  treefile,
  submodel = "LG",
  type = "AA",
  format = "fasta",
  k = 4,
  ...
)
```

#### **Arguments**

alnfile	The path to the sequence alignment file
treefile	The path to the master tree file (whose topology will be used to generate the tree)
submodel	Substitution model to be used to estimate the tree
type	"AA" for amino acid or "DNA" for DNA
format	Format of the alignment file (commonly used formats include fasta and phylip)
k	Number of intervals in the discrete gamma distribution for 'pml'
	Further arguments passed to 'pml' or 'optim.pml'

## Value

A list: 'tree.opt' is the tree from the optimized output of 'optim.pml'; 'results.opt' is the optimized output of 'optim.pml'; 'results.init' is the initial results estimated by 'pml'

## See Also

```
phyDat for alignment formats, pml and optim.pml for tree estimation
```

```
estimatePhangornTreeAll
```

Estimates a ML trees from a given topology for a list of alignments. Wirtes a text file compatble with readTrees(). Uses 'pml' and 'optim.pml' from the 'phangorn' package to estimate the tree.

## Description

Estimates a ML trees from a given topology for a list of alignments. Wirtes a text file compatble with readTrees(). Uses 'pml' and 'optim.pml' from the 'phangorn' package to estimate the tree.

#### Usage

```
estimatePhangornTreeAll(
  alnfiles = NULL,
  alndir = NULL,
  pattern = NULL,
  treefile,
  output.file = NULL,
  submodel = "LG",
  type = "AA",
  format = "fasta",
  k = 4,
  ...
)
```

## Arguments

alnfiles	A character vector of paths to alignment files. Such as one produced by list.files. Either alnfiles or alndir must be supplied.
alndir	The path to directories with alignment files. Trees will be named with the alignment file name after stripping the extension. Either alnfiles or alndir must be supplied.
pattern	An optional regular expression for files in the alndir director. As in ".*fasta".
treefile	The path to the master tree file (whose topology will be used to generate the tree)
output.file	The file where the output trees will be written. This file can be read with read-Trees().
submodel	Substitution model to be used to estimate the tree
type	"AA" for amino acid or "DNA" for DNA
format	Format of the alignment file (commonly used formats include fasta and phylip)
k	Number of intervals in the discrete gamma distribution for 'pml'
	Further arguments passed to 'pml' or 'optim.pml'

fastwilcoxGMTall 11

#### Value

A list: 'tree.opt' is the tree from the optimized output of 'optim.pml'; 'results.opt' is the optimized output of 'optim.pml'; 'results.init' is the initial results estimated by 'pml'

#### See Also

phyDat for alignment formats, pml and optim.pml for tree estimation

## 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 (correlate-

WithContinuousPhenotype, correlateWithBinaryPhenotype, getAllCor)

annotList Pathway annotations formatted as a list of lists of annotation information

#### Value

A list of pathways with enrichment statistics

fixPseudoroot pseudoroot trait tree to match the psuedoroot of trees from readTrees if the trees are reconcilable

## Description

pseudoroot trait tree to match the pseudoroot of trees from readTrees if the trees are reconcilable

#### Usage

```
fixPseudoroot(tree, treesObj)
```

#### Arguments

tree A trait tree with branch lengths representing trait values

treesObject created by readTrees

12 foreground2Paths

## Value

A trait tree with the correct topology

foreground2Paths Cre

Creates paths from a set of foreground species

## 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 13

foreground2Tree

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

## Description

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

## Usage

```
foreground2Tree(
  foreground,
  treesObj,
  plotTree = T,
  clade = c("ancestral", "terminal", "all"),
  weighted = F,
  transition = "unidirectional",
  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".
weighted	if set to TRUE weights foreground edges belonging to the same clade such that their branch lengths sum up to 1 (only done for clade options "all" and "terminal").
transition	A character string indicating whether transitions between background and foreground branches are "bidirectional" or "unidirectional" (no foreground to background transitions, the default)
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
collapse2anc	Put all the weight on the ancestral branch when the trait appears on a whole clade (redundant to "clade", kept for backwards compatibility)
wholeClade	Whether to implement the weighted edge option across all members of a foreground clade (redundant to "clade", kept for backwards compatibility)

#### Value

A tree with edge.lengths representing phenotypic states

foreground2TreeClades

Generates a binary phenotype tree using the list of tip foreground animals, foreground common ancestors, and their phylogenetic relationships

## Description

Generates a binary phenotype tree using the list of tip foreground animals, foreground common ancestors, and their phylogenetic relationships

## Usage

```
foreground2TreeClades(
  fg_vec,
  sisters_list = NULL,
  trees,
  plotTree = T,
  useSpecies = NULL
)
```

#### **Arguments**

```
fg_vec A vector containing the tip foreground species

sisters_list A list containing pairs of "sister species" in the foreground set (put NULL if empty)

trees treesObj from readTrees

plotTree A boolean indicator for plotting the output tree (default=FALSE)

useSpecies An array containing the tip labels in the output tree
```

#### Value

fg.tree A binary phenotype tree corresponding to the input information

```
generatePermulatedBinPhen
```

Produces binary permulations for a gene

## **Description**

Produces binary permulations for a gene

#### Usage

```
generatePermulatedBinPhen(
   tree,
   numperms,
   trees,
   root_sp,
   fg_vec,
   sisters_list,
   pathvec,
   permmode = "cc"
)
```

#### **Arguments**

tree	Tree of the gene of interest (if permmode="cc", set this as the masterTree in trees (i.e., the output from readTrees))
numperms	An integer number of permulations
trees	treesObj from readTrees
root_sp	The species to root the tree on
fg_vec	A vector containing the foreground species
sisters_list	A list containing pairs of "sister species" in the foreground set (put NULL if empty)
pathvec	A path vector generated from the real set of foreground animals

#### Value

output.list a list containing the set of binary permulated trees

```
generatePermulatedBinPhenSSMBatched

Produces binary SSM permulations for a list of genes
```

## Description

Produces binary SSM permulations for a list of genes

#### Usage

```
generatePermulatedBinPhenSSMBatched(
  trees_list,
  numperms,
  trees,
  root_sp,
  fg_vec,
  sisters_list,
  pathvec
)
```

16 getAllCor

## **Arguments**

trees\_list A list containing the trees of all genes of interest (formatted like trees in treesObj from readTrees)

numperms An integer number of permulations

trees treesObj from readTrees

root\_sp The species to root the tree on

fg\_vec A vector containing the foreground species

sisters\_list A list containing pairs of "sister species" in the foreground set (put NULL if empty)

pathvec A path vector generated from the real set of foreground animals

#### Value

simPhenoList A list containing binary permulated trees for each gene

getAllCor Computes the association statistics between RER from getAllResiduals and a phenotype paths vector made with tree2Paths or char2Paths

#### **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,
  winsorizeRER = NULL,
  winsorizetrait = 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 winsorizetrait=some number and winsorizeRER=some number to override

getAllResiduals 17

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)		
winsorizeRER	Winsorize RER values before computing Pearson correlation. winsorize RER=3 will set the 3 most extreme RER values at each end of each row to the value closest to 0.		
winsorizetrait			
	Winsorize trait values before computing Pearson correlation. winsorizetrait=3 will set the 3 most extreme trait values at each end to the value closest to 0.		
weighted	perform weighted correlation. This option needs to be set if the clade weights computed in <pre>foreground2Tree</pre> (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 main RER computation function

#### **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,
  plot = T
)
```

18 getEnrichPerms

## 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.
plot	Whether to plot the output of the correction for mean-variance relationship.
a	cutoff value for branch lengths bellow 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 whill be set to the bottom 0.05 quantile. Set to 0 for no cutoff.

## Value

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

lation permulations	getEnrichPerms	Performs enrichment statistic permulations using existing gene correlation permulations
---------------------	----------------	---

## Description

Performs enrichment statistic permulations using existing gene correlation permulations

## Usage

```
getEnrichPerms(corperms, realenrich, annotlist)
```

## Arguments

corperms	Gene correlation permulations from getPermsContinuous
realenrich	Pathway enrichment results using observed phenotype obtained from 'correlate-WithContinuousPhenotype' or 'correlate-WithBinaryPhenotype'
annotlist	Pathway annotations

#### Value

Full null permulation statistics and p-values for gene correlations and pathway enrichment

```
getForegroundInfoClades
```

Generates a binary phenotype tree and foreground clades information using the list of tip foreground animals, the presence of foreground common ancestors, and their phylogenetic relationships

## Description

Generates a binary phenotype tree and foreground clades information using the list of tip foreground animals, the presence of foreground common ancestors, and their phylogenetic relationships

#### Usage

```
getForegroundInfoClades(
  fg_vec,
  sisters_list = NULL,
  trees,
  plotTree = T,
  useSpecies = NULL
)
```

#### **Arguments**

```
fg_vec A vector containing the tip foreground species

sisters_list A list containing pairs of "sister species" in the foreground set (put NULL if empty)

trees treesObj from readTrees

plotTree A boolean indicator for plotting the output tree (default=FALSE)

useSpecies An array containing the tip labels in the output tree
```

#### Value

output.list A list containing 1) "tree" = a binary phenotype tree corresponding to the input information, 2) "fg.sisters.table" = a table containing all sister species in the foreground set

20 getPermsBinary

getPermsBinary	Calculates permuted correlation and enrichment statistics for binary phenotype

## Description

Calculates permuted correlation and enrichment statistics for binary phenotype

## Usage

```
getPermsBinary(
  numperms,
  fg_vec,
  sisters_list,
  root_sp,
  RERmat,
  trees,
  mastertree,
  permmode = "cc",
  method = "k",
  min.pos = 2,
  trees_list = NULL,
  calculateenrich = F,
  annotlist = NULL
```

## Arguments

numperms	An integer number of permulations
fg_vec	A vector containing the tip foreground species
sisters_list	A list containing pairs of "sister species" in the foreground set (put NULL if empty)
root_sp	The species to root the tree on
RERmat	An RER matrix calculated using getAllResiduals.
trees	treesObj from readTrees
mastertree	A rooted, fully dichotomous tree derived from the treesObj master tree from readTrees. Must not contain species not in traitvec
permmode	Mode of binary permulation ("cc" for Complete Cases (default), "ssm" for Species Subset Match)
method	statistical method to use for correlations (set to "k" (default) for Kendall Tau test)
min.pos	minimum number of foreground species (default 2)
trees_list	A list containing the trees of all genes of interest (formatted like trees in treesObj from readTrees)

getPermsContinuous 21

```
calculateenrich
```

A boolean variable indicating if null permulation p-values for enrichment statis-

tics

annotlist Pathway annotations

#### Value

A list object with enrichment statistics, correlation p-val, rho, and correlation effect size

getPermsContinuous Calculates permuted correlation and enrichment statistics

## Description

Calculates permuted correlation and enrichment statistics

#### Usage

```
getPermsContinuous(
  numperms,
  traitvec,
  RERmat,
  annotlist,
  trees,
  mastertree,
  calculateenrich = T,
  type = "simperm",
  winR = 3,
  winT = 3,
  method = "p",
  min.pos = 0
)
```

## **Arguments**

numperms An integer number of permulations

traitvec A named phenotype vector

RERmat An RER matrix calculated using getAllResiduals

annotlist Pathway annotations

trees treesObj from readTrees

mastertree A rooted, fully dichotomous tree derived from the treesObj master tree from

readTrees. Must not contain species not in traitvec

calculateenrich

A boolean variable indicating if null permulation p-values for enrichment statistics

22 getStat

type	One of "simperm", "sim", or "perm" for permulations, simulations, or permutations, respectively
winR	Integer winzorization value for RERmat
winT	Integer winzorization value for trait
method	statistical method to use for correlations
min.pos	minimum foreground species - should be set to 0

#### Value

A list object with enrichment statistics, correlation p-val, rho, and correlation effect size

#### Note

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

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

## Description

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

## Usage

```
getStat(res)
```

#### **Arguments**

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 23

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

#### **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 pheno-

type) or a character vector of foreground branches

treesObj created by readTrees

#### Value

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

```
matchAllNodesClades
```

Calculates the clade mappings between the gene tree and the master tree (with the complete topology)

## Description

Calculates the clade mappings between the gene tree and the master tree (with the complete topology)

#### Usage

```
matchAllNodesClades(gene_tree, treesObj)
```

## **Arguments**

```
gene_tree A binary phenotype tree of a gene
treesObj treesObj from readTrees
```

#### Value

output.map A list containing a dataframe of clades mapping

24 permpvalcor

namePathsWSpecies

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

## **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

permpvalcor Calculates correlation permutation pvals from output of getPermsContinuous

#### Description

Calculates correlation permutation pvals from output of getPermsContinuous

#### Usage

```
permpvalcor(realcor, permvals)
```

## **Arguments**

realcor Real enrichment statistics from fastwilcoxGMTall

permvals output from getPermsContinuous

#### Value

A vector with permulation p-values

permpvalenrich 25

#### **Description**

Calculates enrichment permutation pvals from output of getPermsContinuous

#### Usage

```
permpvalenrich(realenrich, permvals)
```

## Arguments

realenrich Real enrichment statistics from fastwilcoxGMTall permvals output from getPermsContinuous

#### Value

A list object with vectors of permulation p-values

permutevec Generates a permuted continuous phenotype given an observed continuous phenotype

## Description

Generates a permuted continuous phenotype given an observed continuous phenotype

#### Usage

```
permutevec(namedvec)
```

#### **Arguments**

namedvec A named numeric vector with phenotype values for each speices

#### Value

A vector with permuted phenotype values

26 plotRers

```
\verb"plotPositivesFromPermulations"
```

Plots changes in number of statistically significant rate acceleration or deceleration versus the number of permulations

## Description

Plots changes in number of statistically significant rate acceleration or deceleration versus the number of permulations

#### Usage

```
plotPositivesFromPermulations(
   res,
   perm.out,
   interval,
   pvalthres,
   output.res = FALSE
)
```

#### **Arguments**

res	$correlation\ statistic\ output\ from\ \texttt{correlateWithBinaryPhenotype}\ or\ \texttt{correlateWithConting}$
perm.out	output from getPermsBinary or getPermsContinuous
interval	interval of number of permulations (e.g., interval = $10$ means that number of positives with be calculated for number of permulations = $10$ , $20$ , $30$ ,)
pvalthres	p-value threshold for identifying statistically significant rate acceleration or de- celeration
output.res	Boolean defining whether to output a list object containing changes in the number of identified elements with changing number of permulations (default=FALSE)

#### Value

A list containing changes in the number of identified elements with changing number of permulations (default=NULL)

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

#### **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 getAllResiduals() 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
```

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

## Description

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

#### Usage

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

28 pruneTree

## **Arguments**

useGG option to creat plot with ggtree. Improves plot readability for most data sets.

tree. A tree object.

outgroup. A vector of species to use to root the tree. If not provided, the tree remains

unrooted.

hlspecies. A vector of species whose terminal branches to highlight, or a vector of branch

numbers within the tree.

hlcols. Colors to use in highlighting the branches. If not specified, will use default R

colors.

main. Main text for plot.

#### Value

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

pruneTree

wrapper around drop.tip

## **Description**

```
wrapper around drop.tip
```

#### Usage

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

#### **Arguments**

tree a "phylo" tree

tip.names The tip names to keep in the tree

#### Value

A new pruned tree

read.gmt 29

read.gmt

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

#### **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

 ${\tt readTrees}$ 

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

## 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,
   masterTree = NULL,
   minTreesAll = 20,
   reestimateBranches = F,
   minSpecs = NULL
)
```

#### **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 masterTree (optional) User can specify a master tree. Recommended only when the number

of available gene trees with all species is small.

minTreesAll The minimum number of trees with all species present in order to estimate mas-

ter tree edge lengths (default 20).

reestimateBranches

Boolean indicating whether to re-estimate branch lengths if master tree topology

is included (default FALSE)

minSpecs the minimum number of species that needs to be present in a gene tree to be

included in calculating master tree

#### Value

A trees object of class "treeObj"

A trees object of class "treeObj"

RERconverge

RERconverge

## Description

RERconverge

RERconverge

returnRersAsNewickStrings

Produce a vector of newick strings representing gene trees where the edge lengths correspond to RERs

#### **Description**

Produce a vector of newick strings representing gene trees where the edge lengths correspond to RERs

#### Usage

returnRersAsNewickStrings(treesObj, rermat)

#### **Arguments**

treesObj created by readTrees

rermat. A residual matrix, output of the getAllResiduals() function

returnRersAsTree 31

#### Value

A named character vector of newick strings, one per gene, representing RERs as edge lengths'

returnRersAsTree

Produce a gene tree with branch lengths representing RERs and (optionally) display these RERs as branch labels

#### **Description**

Produce a gene tree with branch lengths representing RERs and (optionally) display these RERs as branch labels

## Usage

```
returnRersAsTree(
  treesObj,
  rermat,
  index,
  phenv = NULL,
  rer.cex = 0.7,
  tip.cex = 0.7,
  nalab = "NA",
  plot = T
)
```

## **Arguments**

treesObj.	A treesObj created by readTrees
rermat.	A residual matrix, output of the getAllResiduals() 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
rer.cex.	Numeric expansion for RER labels
tip.cex.	Numeric expansion for tip labels
nalab.	Label given to any NA RERs
plot.	Whether to produce a plot displaying the RERs on the gene tree

## Value

An object of class "phylo" with edge lengths representing RERs for the given gene If plot = TRUE, also displays a plot of the gene tree with edges labeled with RERs

32 simBinPheno

returnRersAsTreesAll

Produce a multiPhylo object of all gene trees with branch lengths representing RERs for each gene

## Description

Produce a multiPhylo object of all gene trees with branch lengths representing RERs for each gene

#### Usage

```
returnRersAsTreesAll(treesObj, rermat)
```

## Arguments

treesObj. A treesObj created by readTrees

rermat. A residual matrix, output of the getAllResiduals() function

#### Value

An object of class "multiPhylo" of named gene trees with edge lengths representing RERs for the given gene

simBinPheno	Generates a permulated phylogenetic tree with specified number of foreground branches. User may specify the number of foreground
	branches that are internal branches.

## Description

Generates a permulated phylogenetic tree with specified number of foreground branches. User may specify the number of foreground branches that are internal branches.

## Usage

```
simBinPheno(trees, root, phenvec, fgnum = NULL, internal = 0, drop = NULL)
```

## Arguments

trees	treesObj output from readTrees
root	Species on which to root the master tree
phenvec	Named vector of 1's and 0's representing phenotype values for each species
fgnum	Total number of foreground species - only required if internal foreground branches are required

simBinPhenoCC 33

internal	Number of foreground species that should be internal branches - only required if internal foreground branches are required
drop	Character vector (or single character variable) of species names to be removed from the master tree (such as species in trees but not in phenotype vector)

#### Value

A tree with permulated phenotype values

simBinPhenoCC

Produces one CC binary permulation for a gene

## **Description**

Produces one CC binary permulation for a gene

## Usage

```
simBinPhenoCC(
   trees,
   mastertree,
   root_sp,
   fg_vec,
   sisters_list = NULL,
   pathvec,
   plotTreeBool = F
)
```

## Arguments

```
trees treesObj from readTrees

Mastertree A rooted, fully dichotomous tree derived from the treesObj master tree from readTrees.

The species to root the tree on

fg_vec a vector containing the foreground species

sisters_list A list containing pairs of "sister species" in the foreground set (put NULL if empty)

pathvec A path vector generated from the real set of foreground animals

plotTreeBool Boolean indicator for plotting the output tree (default=FALSE)
```

## Value

A CC binary permulated tree

34 simBinPhenoVec

simBinPhenoSSM

Produces one SSM binary permulation for a gene

## **Description**

Produces one SSM binary permulation for a gene

#### Usage

```
simBinPhenoSSM(
   tree,
   trees,
   root_sp,
   fg_vec,
   sisters_list = NULL,
   pathvec,
   plotTreeBool = F
)
```

#### **Arguments**

```
trees Tree of the gene of interest

trees treesObj from readTrees

root_sp The species to root the tree on

fg_vec A vector containing the foreground species

sisters_list A list containing pairs of "sister species" in the foreground set (put NULL if empty)

pathvec A path vector generated from the real set of foreground animals

plotTreeBool Boolean indicator for plotting the output tree (default=FALSE)
```

#### Value

A SSM binary permulated tree

1	lated phenotype vector whose phylogeny matches e. User may specify the number of foreground nternal branches.
***************************************	

#### **Description**

Generates a permulated phenotype vector whose phylogeny matches a desired structure. User may specify the number of foreground branches that are internal branches.

simpermvec 35

#### Usage

```
simBinPhenoVec(trees, root, phenvec, fgnum = NULL, internal = 0, drop = NULL)
```

#### **Arguments**

trees	treesObj output from readTrees
root	Species on which to root the master tree
phenvec	Named vector of 1's and 0's representing phenotype values for each species
fgnum	Total number of foreground species - only required if internal foreground branches are required
internal	Number of foreground species that should be internal branches - only required if internal foreground branches are required
drop	Character vector (or single character variable) of species names to be removed from the master tree (such as species in trees but not in phenotype vector)

#### Value

A vector of permulated foreground species

simpermvec	Generates a permulated continuous phenotype given an observed continuous phenotype and a phylogeny

## **Description**

Generates a permulated continuous phenotype given an observed continuous phenotype and a phylogeny

## Usage

```
simpermvec(namedvec, treewithbranchlengths)
```

#### **Arguments**

namedvec A named numeric vector with phenotype values for each speices treewithbranchlengths

A rooted phylogenetic tree with the same species as namedvec and branch lengths representing average evolutionary rate. The master tree from readTrees may be rooted and used for this parameter.

#### Value

A vector with permulated phenotype values

36 tree2Paths

simulatevec	Generates a simulated continuous phenotype given an observed continuous phenotype and a phylogeny

## Description

Generates a simulated continuous phenotype given an observed continuous phenotype and a phylogeny

#### Usage

```
simulatevec(namedvec, treewithbranchlengths)
```

## Arguments

 ${\tt namedvec} \qquad \text{$A$ named numeric vector with phenotype values for each speices} \\ {\tt treewithbranchlengths}$ 

A rooted phylogenetic tree with the same species as namedvec and branch lengths representing average evolutionary rate. The master tree from readTrees may be rooted and used for this parameter.

#### Value

A vector with simulated phenotype values

ŧ	tree2Paths	Generate a phenotype paths vector from a phenotype tree

## 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).

tree2PathsClades 37

• 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

tree2PathsClades

A modification of the tree2Paths function that takes in pre-calculated mappings

#### **Description**

A modification of the tree2Paths function that takes in pre-calculated mappings

## Usage

```
tree2PathsClades(tree, trees)
```

#### **Arguments**

tree the input tree to be converted into paths

trees treesObj from readTrees

38 treePlotNew

treePlotNew

Plot 'tree' with branch labels colored in a heatmap based on values in 'vals'

#### **Description**

Plot 'tree' with branch labels colored in a heatmap based on values in 'vals'

#### **Usage**

```
treePlotNew(
 tree,
  vals = NULL
  rank = F,
  nlevels = 9,
  type = "c",
  col = NULL,
  maintitle = NULL,
  useedge = F,
  doreroot = F,
  rerootby = NULL,
  useSpecies = NULL,
  species.names = NULL,
  speclist1 = NULL,
  speclist2 = NULL,
  aligntip = F,
  colpan1 = rgb(0, 119, 187, maxColorValue = 255),
  colpan2 = rgb(204, 51, 17, maxColorValue = 255),
  colpanmid = rgb(187, 187, 187, maxColorValue = 255),
  plotspecies = NULL,
  edgetype = NULL,
  textsize = 0.6,
  colbarlab = "",
  splist2sym = "psi",
  dolegend = T_{i}
  nacol = rgb(0, 0, 0),
  figwid = 10,
)
```

#### **Arguments**

tree. A phylo object, used for topology
vals. Values to use for heatmap to color branches, in same order as edges of tree
rank. Whether to plot ranks instead of values
nlevels. How many colors to use in the heatmap

treePlotNew 39

Type for plot.phylo type. Vector of user-defined colors col. maintitle. Main title label Whether to use edge lengths in 'tree' for plotting useedge. doreroot. Whether to re-root the tree before plotting If re-rooting, what to use to root the tree (verify by checking against unrooted rerootby. useSpecies. A vector of species to include in the plot (verify by checking against full set) species.names. Data fram for converting names in 'tree' to names to be plotted (row names should be tip labels of tree, and first column should contain the corresponding desired tip labels) speclist1. A vector of tip labels to highlight in bold, blue text speclist2. A vector of tip labels to which to add an asterisk aligntip. Whether to align tip labels (default FALSE) colpan1. Color for lowest value in heatmap (default blue) Color for highest value in heatmap (default rose) colpan2. Color for middle value in heatmap (default gray) colpanmid. plot species. A vector of tip labels to display on the tree (the remainder will be masked, but the corresponding tips will be plotted on the tree) Vector of line type for edges. edgetype. cex value for tip labels (default 0.6) textsize. colbarlab. Label for the color bar legend A value within species names to display as a symbol splist2sym. dolegend. Whether to display the heatmap legend. nacol. Color to display for any edges with length NA figwid. Adjust x limits of plot.phylo by 1/figwid. May be related to figure width (requires some optimization)

further arguments to be passed to 'plot' or to 'plot.phylo'

## Value

Plots a cladogram of 'tree' with branch colors determined by 'vals'

40 treePlotRers

treePlotRers	Plot a cladogram with RERs shown as either labels (type="label") or a color heatmap along the branches (type="color") Wraps around
	returnRersAsTree or treePlotNew, respectively

## Description

Plot a cladogram with RERs shown as either labels (type="label") or a color heatmap along the branches (type="color") Wraps around returnRersAsTree or treePlotNew, respectively

## Usage

```
treePlotRers(
  treesObj,
  rermat = NULL,
  index = NULL,
  type = c("label", "color"),
  phenv = NULL,
  figwid = 10,
  ...
)
```

## Arguments

treesObj.	A treesObj created by readTrees
rermat.	A residual matrix, output of the getAllResiduals() function
index.	A character denoting the name of gene, or a numeric value corresponding to the gene's row index in the residuals matrix
type.	Whether to display RERs as branch labels ('label') or a heatmap ('color')
phenv.	A phenotype vector returned by tree2Paths or foreground2Paths
• • • • • • • • • • • • • • • • • • • •	Additional parameters to be passed to returnRersAsTree or treePlotNew #' @param figwid. Adjust x limits of plot.phylo by 1/figwid. May be related to figure width (requires some optimization)

#### Value

Plots a cladogram of the master tree with RERs displayed as branch labels or colors

# **Index**

calculateCorPermuted, 3	namePathsWSpecies, 24
calculatePermulatedPaths_apply,	
3, 4	optim.pml, $9$ , $11$
char2Paths, 4, 7, 16	- 24
<pre>click_select_foreground_branches,</pre>	permpvalcor, 24
5	permpvalenrich, 25
combinePermData, 6	permutevec, 25
cor, 3, 7, 8, 16	phyDat, 9, 11
correlateWithBinaryPhenotype, 6,	plotPositivesFromPermulations, 26
26	plotRers, 26
correlateWithContinuousPhenotype,	plotTreeHighlightBranches, 27
7, 26	pml, 9, 11
7, 20	pruneTree, 28
drop.tip, 28	
1 1/	read.gmt, 29
edgeVars, 8	readTrees, 4, 5, 11-16, 18-21, 23, 29,
estimatePhangornTree,9	30–37, 40
estimatePhangornTreeAll, 10	RERconverge, 30
5 , -	returnRersAsNewickStrings, 30
fastwilcoxGMTall, 11, 24, 25	returnRersAsTree, 31, 40
fixPseudoroot, 11	returnRersAsTreesAll, 32
foreground2Paths, 12, 27, 31, 40	
foreground2Tree, 7, 13, 17	simBinPheno, 32
foreground2TreeClades, 14	simBinPhenoCC, 33
1010g10ana2110001aacb, 11	simBinPhenoSSM, 34
generatePermulatedBinPhen, 14	simBinPhenoVec, 34
generatePermulatedBinPhenSSMBatched,	simpermvec, 35
15	simulatevec, 36
getAllCor, 4, 16	
getAllResiduals, 3, 6, 7, 16, 17, 20, 21	tree2Paths, 6, 7, 16, 27, 31, 36, 40
getEnrichPerms, 18	tree2PathsClades, 37
getForegroundInfoClades, 19	treePlotNew, 38, 40
getPermsBinary, 20, 26	treePlotRers,40
=	
getPermsContinuous, $6$ , $18$ , $21$ , $24$ – $26$ getStat, $22$	
list.files, 10	
makeBinaryPaths, 23	
matchAllNodesClades 23	