Package ‘datelife’

August 22, 2017

Title Underlying code for DateLife service

Version 0.2.13

Author Brian O'Meara <omeara.brian@gmail.com>, Luna L. Sanchez Reyes, Jonathan Eastman, Tracy Heath,
April Wright, Klaus Schliep, Scott Chamberlain, Peter Midford, Luke Harmon,
Joseph Brown, Matt Pennell, Mike Alfaro.

Maintainer Brian O'Meara <omeara.brian@gmail.com>

Description This contains the core code for going from a tree or list of taxa to a chronogram.

Depends R (>= 3.2.3)

Imports ape, abind, bold, phangorn, phytools, ips, compare, geiger, stringr, rotl, knitcitations, phylobase, treebase

Suggests testthat

SystemRequirements mafft, PATHd8

URL https://github.com/phylotastic/datelife, http://www.datelife.org

License GPL (>=2)

LazyData true

RoxygenNote 6.0.1

R topics documented:

`AllMatching` *Are all desired taxa in the patristic.matrix?*

Description

Are all desired taxa in the patristic.matrix?

Usage

```
AllMatching(patristic.matrix, taxa)
```

Arguments

`patristic.matrix`	A patristic matrix, rownames and colnames must be taxa
`taxa`	A vector of taxon names

Value

A Boolean

`BindMatrices` *Convert list of patristic matrices to a 3D array*

Description

Convert list of patristic matrices to a 3D array

Usage

```
BindMatrices(patristic.matrix.list, pad = TRUE)
```

Arguments

`patristic.matrix.list`	List of patristic matrices
`pad`	If TRUE, pad missing entries

Value

A 3d array of patristic matrices

`CalculateOverlapTable` *Create an overlap table*

Description

Create an overlap table

Usage

```
CalculateOverlapTable(results.table)
```

Arguments

`results.table`	An author.results or curator.results data.frame

Value

Data.frame with info on people and what clades they've worked on

`CleanChronogram` *Clean up some issues with OToL chronograms*

Description

Clean up some issues with OToL chronograms

Usage

```
CleanChronogram(phy)
```

Arguments

`phy`	Input phylo object

Value

A cleaned up phylo object

`ConvertSpacesToUnderscores` *Convert spaces to underscores in trees*

Description

Convert spaces to underscores in trees

Usage

```
ConvertSpacesToUnderscores(phy)
```

Arguments

`phy`	A phylo object

Value

A phylo object

`ConvertUnderscoresToSpaces` *Convert underscores to spaces in trees*

Description

Convert underscores to spaces in trees

Usage

```
ConvertUnderscoresToSpaces(phy)
```

Arguments

`phy`	A phylo object

Value

A phylo object

`CreateContributorCache` *Create a cache from Open Tree of Life*

Description

Create a cache from Open Tree of Life

Usage

```
CreateContributorCache(outputfile = "contributorcache.rda", verbose = TRUE)
```

Arguments

`outputfile`	Path including file name
`verbose`	If TRUE, give status updates to the user

Value

List containing author and curator results

CreateTreeBaseCache *Create a cache from TreeBase*

Description

Create a cache from TreeBase

Usage

```
CreateTreeBaseCache(outputfile = "treebasecache.rda", verbose = TRUE)
```

Arguments

outputfile	Path including file name
verbose	If TRUE, give status updates to the user

Value

List containing author and curator results

EstimateDates *Core function to go from a vector of species, newick string, or phylo object get a chronogram or dates back*

Description

Core function to go from a vector of species, newick string, or phylo object get a chronogram or dates back

Usage

```
EstimateDates(input = c("Rhea americana", "Pterocnemia pennata",
  "Struthio camelus"), output.format = "phylo.sdm", partial = TRUE,
  usetnrs = FALSE, approximatematch = TRUE,
  cache = get("opentree_chronograms"), method = "PATHd8")
```

Arguments

input	A vector of names, a newick string, or a phylo object
output.format	The desired output format. See details.
partial	If TRUE, use source trees even if they only match some of the desired taxa
usetnrs	If TRUE, use OpenTree's services to resolve names. This can dramatically improve the chance of matches, but also take much longer
approximatematch	If TRUE, use a slower TNRS to correct mispellings, increasing the chance of matches (including false matches)
cache	The cached set of chronograms and other info from data(opentree_chronograms)
method	The method used for congruification. PATHd8 only right now, r8s and treePL later.

Details

The output formats are citations, mrca, newick.all, newick.sdm, newick.median, phylo.sdm, phylo.median, phylo.all, html

Value

Varies depending on the chosen format

Examples

```
ages <- EstimateDates(c("Rhea americana", "Pterocnemia pennata", "Struthio camelus", "Mus musculus"), output.f
```

FindMatchingStudyIndex
Find the index of relevant studies in a opentree_chronograms object

---

Description

Find the index of relevant studies in a opentree_chronograms object

Usage

```
FindMatchingStudyIndex(filtered.results, cache = get("opentree_chronograms"))
```

Arguments

`filtered.results`
The patristic.matrices that will be used

`cache` The cache of studies

Value

A vector with the indices of studies that have relevant info

GetAge
Get time of MRCA from patristic matrix

Description

Get time of MRCA from patristic matrix

Usage

```
GetAge(patristic.matrix, partial = TRUE)
```

Arguments

`patristic.matrix`
A patristic matrix

`partial` If TRUE, drop NA from the patristic matrix; if FALSE, will return NA if there are missing entries

Value

The depth of the MRCA

`GetAges` *Get vector of MRCA from list of patristic matrices*

Description

Get vector of MRCA from list of patristic matrices

Usage

```
GetAges(filtered.results, partial = TRUE)
```

Arguments

`filtered.results`
List of patristic matrices

`partial` If TRUE, drop NA from the patristic matrix; if FALSE, will return NA if there are missing entries

Value

Vector of MRCA ages with names same as in filtered.results

`GetAllCalibrations` *Get all calibrations given trees in database*

Description

Get all calibrations given trees in database

Usage

```
GetAllCalibrations(input = c("Rhea americana", "Pterocnemia pennata",
  "Struthio camelus"), partial = TRUE, usetnrs = FALSE,
  approximatematch = TRUE, cache = get("opentree_chronograms"))
```

Arguments

`input` A vector of names, a newick string, or a phylo object

`partial` If TRUE, use source trees even if they only match some of the desired taxa

`usetnrs` If TRUE, use OpenTree's services to resolve names. This can dramatically improve the chance of matches, but also take much longer

`approximatematch`
If TRUE, use a slower TNRS to correct mispellings, increasing the chance of matches (including false matches)

`cache` The cached set of chronograms and other info from data(opentree_chronograms)

Value

data.frame of calibrations

GetBoldOToLTree *Use Barcode of Life data to get branch lengths on the OToL tree*

Description

Use Barcode of Life data to get branch lengths on the OToL tree

Usage

```
GetBoldOToLTree(input = c("Rhea americana", "Struthio camelus",
  "Gallus gallus"), marker = "COI", otol_version = "v2", doML = FALSE)
```

Arguments

input	A vector of names
marker	Gene name to select
otol_version	Version of OToL to use
doML	Boolean; if TRUE, does ML brlen optimization

Value

A phylogeny with ML branch lengths

GetFilteredResults *Go from a vector of species, newick string, or phylo object to a list of patristic matrices*

Description

Go from a vector of species, newick string, or phylo object to a list of patristic matrices

Usage

```
GetFilteredResults(input = c("Rhea americana", "Pterocnemia pennata",
  "Struthio camelus"), partial = TRUE, usetnrs = FALSE,
  approximatematch = TRUE, cache = get("opentree_chronograms"),
  method = "PATHd8")
```

Arguments

input	A vector of names, a newick string, or a phylo object
partial	If TRUE, use source trees even if they only match some of the desired taxa
usetnrs	If TRUE, use OpenTree's services to resolve names. This can dramatically improve the chance of matches, but also take much longer
approximatematch	If TRUE, use a slower TNRS to correct mispellings, increasing the chance of matches (including false matches)
cache	The cached set of chronograms and other info from data(opentree_chronograms)
method	The method used for congruification. PATHd8 only right now, r8s and treePL later.

Value

List of patristic matrices

GetOToLChronograms *Get all chronograms from Open Tree of Life*

Description

Get all chronograms from Open Tree of Life

Usage

```
GetOToLChronograms(verbose = FALSE)
```

Arguments

verbose	If TRUE, give updates to the user

Value

A list with elements for the trees, authors, curators, and study ids

GetSubsetArrayDispatch *Figure out which subset function to use*

Description

Figure out which subset function to use

Usage

```
GetSubsetArrayDispatch(study.element, taxa, phy = NULL, phy4 = NULL,
  method = "PATHd8")
```

Arguments

study.element	The thing being passed in: an array or a phylo to serve as reference
taxa	Vector of taxon names to get a subset for
phy	A user tree to congruify in phylo format (ape)
phy4	A user tree to congruify in phylo4 format (phylobase)
method	Which method to use for congruification

Value

A patristic matrix with for the taxa.

HasBrlen *Check for branch lengths in a tree*

Description

Check for branch lengths in a tree

Usage

```
HasBrlen(x)
```

Arguments

`x`	A phylo object

Value

A TRUE or FALSE

IsGoodChronogram *Check to see that a chronogram is valid*

Description

Check to see that a chronogram is valid

Usage

```
IsGoodChronogram(phy)
```

Arguments

`phy`	Input phylo object

Value

Boolean: TRUE if good tree

PadMatrix *Function to fill in empty cells in a patristic matrix for missing taxa*

Description

Function to fill in empty cells in a patristic matrix for missing taxa

Usage

```
PadMatrix(patristic.matrix, all.taxa)
```

Arguments

`patristic.matrix`	A patristic matrix with row and column names for taxa
`all.taxa`	A vector of the names of all taxa you want, including ones not in the patristic matrix

Value

Patristic.matrix for all.taxa, with NA for entries between taxa where at least one was not in the original patristic.matrix

PatristicMatrixToNewick *Convert patristic matrix to a newick string*

Description

Convert patristic matrix to a newick string

Usage

```
PatristicMatrixToNewick(patristic.matrix)
```

Arguments

`patristic.matrix`	A patristic matrix

Value

A newick string

PatristicMatrixToTree *Convert patristic matrix to a phylo object*

Description

Convert patristic matrix to a phylo object

Usage

```
PatristicMatrixToTree(patristic.matrix)
```

Arguments

`patristic.matrix`
A patristic matrix

Value

A rooted phylo object

ProcessInput *Take input string, figure out if it's newick or list of species*

Description

Take input string, figure out if it's newick or list of species

Usage

```
ProcessInput(input = c("Rhea americana", "Pterocnemia pennata",
  "Struthio camelus"), usetnrs = FALSE, approximatematch = TRUE)
```

Arguments

`input`	A newick string or vector of taxa, or a phylo object
`usetnrs`	Whether to use OpenTree's TNRS for the input
`approximatematch`	If using TNRS, use approximate matching

Value

A list with the phy (or NA, if no tree) and cleaned vector of taxa

ProcessResultsList *Take results.list and process it*

Description

Take results.list and process it

Usage

```
ProcessResultsList(results.list, taxa = NULL, partial = FALSE)
```

Arguments

`results.list`	A list returned from using GetSubsetArrayDispatch on opentree_chronograms$trees
`taxa`	A vector of taxa to match
`partial`	If TRUE, return matrices that have only partial matches

Value

A list with the patristic.matrices that are not NA

ReorderMatrix *Function to reorder a matrix so that row and column labels are in alphabetical order*

Description

Function to reorder a matrix so that row and column labels are in alphabetical order

Usage

```
ReorderMatrix(patristic.matrix)
```

Arguments

`patristic.matrix`	A patristic matrix with row and column names for taxa

Value

patristic.matrix A patristic matrix with row and column names for taxa in alphabetial order

run — *Core function to generate results*

Description

Core function to generate results

Usage

```
run(input = c("Rhea americana", "Pterocnemia pennata", "Struthio camelus"),
  format = "citations", partial = "yes", plot.width = 600,
  plot.height = 600, usetnrs = "no", approximatematch = "yes",
  opentree_chronograms = NULL)
```

Arguments

`input`	A newick string or vector of taxa
`format`	The output format
`partial`	How to deal with trees that have a subset of taxa in the query
`plot.width`	Width in pixels for output plot
`plot.height`	Height in pixels for output plot
`usetnrs`	Whether to use OpenTree's TNRS for the input
`approximatematch`	If using TNRS, use approximate matching
`opentree_chronograms`	The list of lists containing the input trees and other info

Value

results in the desired format

RunSDM — *Function to compute the SDM supertree Criscuolo et al. 2006*

Description

Function to compute the SDM supertree Criscuolo et al. 2006

Usage

```
RunSDM(filtered.results, weighting = "flat")
```

Arguments

`filtered.results`	List of patristic matrices
`weighting`	flat, taxa, inverse

Details

Weighting is how much weight to give each input tree. flat = all trees have equal weighting taxa = weight is proportional to number of taxa inverse = weight is proportional to 1 / number of taxa Criscuolo A, Berry V, Douzery EJ, Gascuel O. SDM: a fast distance-based approach for (super) tree building in phylogenomics. Syst Biol. 2006;55(5):740–55. doi: 10.1080/10635150600969872.

Value

A list containing phy (a chronogram), and filtered.results that were actually used

SaveOToLChronograms *Save all chronograms from Open Tree of Life*

Description

Save all chronograms from Open Tree of Life

Usage

```
SaveOToLChronograms(file = "opentree_chronograms.rda", verbose = FALSE)
```

Arguments

file	Path including file name
verbose	If TRUE, give status updates to the user

Value

None

SummarizeResults *Summarize a filtered results list in various ways*

Description

Summarize a filtered results list in various ways

Usage

```
SummarizeResults(filtered.results, output.format, partial = TRUE,
  cache = get("opentree_chronograms"), suppress.citations = FALSE)
```

Arguments

filtered.results	A list of patristic matrices; labels correspond to citations
output.format	The desired output format
partial	If TRUE, use source trees even if they only match some of the desired taxa
cache	The cached trees and other information
suppress.citations	If using a format that would normally print() citations, turn this off

Value

Depends on output format

`SummaryPatristicMatrixArray` *Summarize patristic matrix array (by default, median)*

Description

Summarize patristic matrix array (by default, median)

Usage

```
SummaryPatristicMatrixArray(patristic.matrix.array, fn = stats::median)
```

Arguments

`patristic.matrix.array`	3D array of patristic matrices
`fn`	The function to use ot summarize

Value

A 2d array with the median (or max, or mean, etc) of the input array

---

`TabulateRelevantAuthors` *Return the relevant authors for a set of studies*

Description

Return the relevant authors for a set of studies

Usage

```
TabulateRelevantAuthors(results.index, cache = get("opentree_chronograms"))
```

Arguments

`results.index`	A vector from FindMatchingStudyIndex() with the indices of the relevant studies
`cache`	The cache

Value

A vector with counts of each author, with names equal to author names

TabulateRelevantCurators *Return the relevant curators for a set of studies*

Description

Return the relevant curators for a set of studies

Usage

```
TabulateRelevantCurators(results.index, cache = get("opentree_chronograms"))
```

Arguments

`results.index` A vector from FindMatchingStudyIndex() with the indices of the relevant studies

`cache` The cache

Value

A vector with counts of each curator, with names equal to curator names

UnpadMatrix *Function to remove missing taxa*

Description

Function to remove missing taxa

Usage

```
UnpadMatrix(patristic.matrix)
```

Arguments

`patristic.matrix` A patristic matrix with row and column names for taxa

Value

Patristic.matrix for all.taxa

UseAllCalibrations *Use all calibrations given trees in database to date a tree*

Description

Use all calibrations given trees in database to date a tree

Usage

```
UseAllCalibrations(phy = GetBoldOToLTree(c("Rhea americana",
  "Struthio camelus", "Gallus gallus")), partial = TRUE, usetnrs = FALSE,
  approximatematch = TRUE, cache = get("opentree_chronograms"),
  expand = 0.1, giveup = 100)
```

Arguments

`phy`	A phylo object
`partial`	If TRUE, use source trees even if they only match some of the desired taxa
`usetnrs`	If TRUE, use OpenTree's services to resolve names. This can dramatically improve the chance of matches, but also take much longer
`approximatematch`	If TRUE, use a slower TNRS to correct mispellings, increasing the chance of matches (including false matches)
`cache`	The cached set of chronograms and other info from data(opentree_chronograms)
`expand`	How much to expand by each step to get consistent calibrations
`giveup`	How many expansion to try before giving up

Details

This will try to use the calibrations as fixed ages. If that fails (often due to conflict between calibrations), it will expand the range of the minage and maxage and try again. And repeat. expand sets the expansion value: should be between 0 and 1

Value

list with chronogram, original calibrations, and expanded calibrations

Index