

Package ‘strap’

February 20, 2015

Type Package

Title Stratigraphic Tree Analysis for Palaeontology

Version 1.4

Date 2014-10-31

Author Mark A. Bell <mark.bell521@gmail.com> Graeme T. Lloyd <graemetlloyd@gmail.com>.

Maintainer Mark A. Bell <mark.bell521@gmail.com>

Depends ape,geoscale

Description Functions for the stratigraphic analysis of phylogenetic trees.

License GPL (>= 2)

LazyData yes

ByteCompile yes

NeedsCompilation no

Repository CRAN

Date/Publication 2014-11-05 20:34:50

R topics documented:

strap-package	2
Asaphidae	2
DatePhylo	3
Dipnoi	5
FindDescendants	6
geoscalePhylo	6
StratPhyloCongruence	10
UKzones	15
Index	17

strap-package

Stratigraphic Tree Analysis for Palaeontology

Description

A series of functions for stratigraphic analysis of phylogenetic trees.

Details

Package: strap
Type: Package
Version: 1.4
Date: 2014-10-31
License: GPL (>=2)
LazyLoad: yes

Author(s)

Mark A. Bell <mark.bell521@gmail.com> and Graeme T. Lloyd <graemetlloyd@gmail.com>

References

Wills, M. A., 1999, Congruence between phylogeny and stratigraphy: randomization tests and the gap excess ratio. *Systematic Biology*, 48(3), 559-580.

Examples

```
# Calculate stratigraphic fit measures treating ages as ranges:  
fit.to.strat.1 <- StratPhyloCongruence(trees=Dipnoi$tree, ages=Dipnoi$ages, rlen=0,  
method="basic", samp.perm=5, rand.perm=5, hard=TRUE,  
randomly.sample.ages=FALSE, fix.topology=FALSE, fix.outgroup=TRUE)
```

```
fit.to.strat.1$input.tree.results # Show just the output for the input tree(s)
```

Asaphidae

Phylogeny and age data for the Asaphidae

Description

Phylogeny (162 most parsimonious trees) and age data for Asaphidae genera (Trilobita, Asaphida) taken from Bell and Braddy (2012).

Usage

```
data(Asaphidae)
```

Format

A list containing 162 trees (`$tree`) and a matrix of first and last appearances (`$ages`).

References

Bell, M. A. and Braddy, S. J., 2012, Cope's rule in the Ordovician trilobite Family Asaphidae (Order Asaphida): patterns across multiple most parsimonious trees: *Historical Biology*, v. 24, no. 3, p. 223-230.

DatePhylo	<i>Calculates branch lengths for a topology</i>
-----------	---

Description

Calculates branch lengths for a topology given a tree and age data for the tips.

Usage

```
DatePhylo(tree, ages, rlen=0, method="basic", add.terminal=FALSE)
```

Arguments

<code>tree</code>	Tree as a phylo object.
<code>ages</code>	A two-column matrix of taxa (rows) against First and Last Appearance Datums (FADs and LADs). Note that rownames should be the taxon names exactly as they appear in <code>tree\$tip.label</code> and colnames should be "FAD" and "LAD". All ages should be in time before present.
<code>rlen</code>	Root length. This must be greater than zero if using a method other than basic.
<code>method</code>	The dating method used. Either basic (Norell 1992; Smith 1994), <i>ruta</i> (Ruta et al. 2006; requires input tree to have branch lengths) or equal (Brusatte et al. 2008).
<code>add.terminal</code>	An optional to add the range of a taxon (FAD minus LAD) to terminal branch lengths.

Details

The basic method (Norell 1992; Smith 1994) of dating a phylogenetic tree of fossil occurrences in palaeontology has been to make each internal node the age of its oldest descendant. In practical terms this means at least half of the branches in a fully bifurcating tree will have a duration of zero million years, as a hypothetical ancestor and its immediate descendant will have the same age, creating a major problem for a variety of rate-based approaches that use branch durations as a divisor.

Early solutions to this problem relied on adding some arbitrary value to each branch in order to enforce non-zero durations. However, more recently Ruta et al. (2006) argued for an approach that first dated the tree using the basic approach then, working from tip-to-root, whenever a zero duration branch was encountered it was assigned a share of the time available from the first directly ancestral branch of positive length. The size of this share is decided by some measure of evolutionary change along that branch. Ruta et al. (2006) used patristic dissimilarity (Wagner 1997), but conceivably any measure could be used. This approach was modified slightly by Brusatte et al. (2008), who preferred equal sharing. This has a couple of benefits over Ruta et al. (2006). Firstly, it avoids zero-length branches entirely - these could still happen with the Ruta et al. 2006 approach, as if no change occurs along a branch it gets zero share of any time. Secondly, it opens up the dating approach to trees without meaningful branch lengths, such as supertrees.

An undiscussed problem with the Ruta et al. (2006), and by extension the Brusatte et al. (2008) approach, concerns the inevitable zero-length branch at the base of the tree that has no preceding ancestral branch with which to share time. Here the obvious practical solution to this problem is implemented - to allow the user to pick a root length that the lowest branch(es) of the tree can share time with (Lloyd et al. 2012). Although selection of this value is potentially arbitrary, in most cases it will only effect a very small number of branches (potentially only a single branch). A recommended method for choosing root length is to use the difference between the oldest taxon in the tree and the age of the first outgroup to the tree that is older (ensuring a positive value).

Note that all three methods implemented here are effectively minimal approaches, in that they assume as little missing or unsampled history as possible. This is because they have their roots in maximum parsimony as an optimality criterion. Consequently the user should be aware that this function will likely return trees with relatively very short internal branch lengths, which may be a source of bias in subsequent analyses.

These approaches (with the exception of the Ruta method) are also implemented, along with others, in the `timePaleoPhy` function of the `paleotree` package.

Value

A phylo object with branch lengths scaled to time and the root age stored as `$root.time`.

Author(s)

Graeme T. Lloyd <graemetlloyd@gmail.com>

References

- Brusatte, S. L., Benton, M. J., Ruta, M. and Lloyd, G. T., 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, 321, 1485-1488.
- Lloyd, G. T., Wang, S. C. and Brusatte, S. L., 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (*Sarcopterygii*; *Dipnoi*). *Evolution*, 66, 330-348.
- Norell, M. A., 1992. Taxic origin and temporal diversity: the effect of phylogeny. In: *Extinction and Phylogeny*, Novacek, M. J. and Wheeler, Q. D. (eds.). Columbia University Press, New York, p89-118.
- Ruta, M., Wagner, P. J. and Coates, M. I., 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proceedings of the Royal Society B*, 273, 2107-2111.

Smith, A. B., 1994. Systematics and the Fossil Record. Blackwell Scientific, London, 223pp.

Wagner, P. J., 1997. Patterns of morphologic diversification among the Rostroconchia. *Paleobiology*, 23, 115-150.

See Also

timePaleoPhy in paleotree package

Examples

```
# Time-scale the lungfish tree using the "equal" method and a root length of 1 Ma:  
time.tree <- DatePhylo(Dipnoi$tree, Dipnoi$ages, method="equal", rlen=1)  
plot(time.tree, cex=0.5)
```

Dipnoi

Phylogeny and age data for dipnoans (lungfish)

Description

Phylogeny (first most parsimonious tree) and age data for lungfish (Osteichthyes, Sarcopterygii, Dipnoi) taken from Lloyd et al. (2012).

Usage

```
data(Dipnoi)
```

Format

A list containing a tree (`$tree`) and a matrix of first and last appearances (`$ages`).

References

Lloyd, G. T., Wang, S. C. and Brusatte, S. L., 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). *Evolution*, 66, 330-348.

FindDescendants	<i>Finds the tip numbers descending from a specific node in a phylo object.</i>
-----------------	---

Description

Finds the tip numbers descending from a specific node in a phylo object.

Usage

```
FindDescendants(n, tree)
```

Arguments

n	The node number.
tree	Tree as a phylo object.

Author(s)

Graeme T. Lloyd <graemetlloyd@gmail.com>

Examples

```
# Find descendants of the root node in the lungfish tree:
FindDescendants(87, Dipnoi$tree)
```

geoscalePhylo	<i>Plots a phylogeny against the geological time scale.</i>
---------------	---

Description

Plots a time-scaled phylogeny against the international geological time scale.

Usage

```
geoscalePhylo(tree, ages, direction="rightwards", units=c("Period", "Epoch", "Age"),
  boxes="Age", tick.scale="myr", user.scale, cex.age=0.3, cex.ts=0.3, cex.tip=0.3,
  width=1, label.offset, ts.col=TRUE, vers="ICS2013", x.lim, quat.rm=FALSE, erotate,
  arotate, urotate,...)
```

Arguments

tree	A tree as a phylo object.
ages	A dataset containing the first and last appearance datums, "FAD" and "LAD" respectively, of all taxa in the phylogeny. See the object \$ages in data(Dipnoi) for an example.
direction	The direction the tree is to be plotted in, options include "rightwards" and "upwards", see help(plot.phylo).
units	The temporal unit(s) to be included in the timescale, options include: "Eon", "Era", "Period", "Epoch", "Age" and "User". The option "User" is required when including a user-defined timescale. This also requires an object to be assigned to user.scale (see Details).
boxes	Option for including grey boxes at a certain temporal resolution, options are the same as for units.
tick.scale	The resolution of the tick marks at the base of the timescale, the default is the same as units. The resolution of the scale can also be chosen by specifying a value or removed entirely by using "no".
user.scale	The data object to be used when including a user-defined time scale, requires the option "User" to be included in units. See data(UKzones) as an example of the required data format.
cex.age	Size of the text on the scale bar.
cex.ts	Size of the text on the geological time scale.
cex.tip	Size of the tip labels on the phylogeny
width	Width of the edges of the phylogeny.
label.offset	A value for the distance between the nodes and tip labels, see help(plot.phylo).
ts.col	Option for using standard ICS colours on the time scale.
vers	The version of the geological time scale to use. Options include: "ICS2013", "ICS2012", "ICS2010", "ICS2009" or "ICS2008".
x.lim	A two item statement for the geological range, in millions of years, of the plot i.e. (0,65). If only one value is used it will be used as the upper limit, see help(plot.phylo).
quat.rm	Option to remove the names from Quaternary time bins, useful when plotting clades with long durations that range through to the recent.
erotate	A numerical value for the rotation for the Epoch/Series temporal units, default values are 0 when direction = "upwards" and 90 when direction = "rightwards".
arotate	A numerical value for the rotation for the Age/Stage temporal units, default values are 0 when direction = "upwards" and 90 when direction = "rightwards".
urotate	A numerical value for the rotation for the User temporal units, default values are 0 when direction = "upwards" and 90 when direction = "rightwards".
...	All other arguments passed to plot.phylo

Details

Often palaeontologists wish to display phylogenetic trees against a geological time scale for easy visualization of the temporal position of the taxa the generation of which can be time-consuming. `geoscalePhylo` fills this need and works in the same way as `geoscale.plot` from the package `geoscale` and allows users to plot a time-scaled phylogeny against the International Chronostratigraphic Chart (Gradstein, 2014). This function accepts any tree time-scaled through either the function `DatePhylo` in this package or the `timePaleoPhy` function in the library `paleotree`.

Built-in options allow the user control over which direction the tree is plotted in (either horizontally or vertically) as well as deciding which temporal units are included in the time scale (see below for example).

Temporal units

The function `geoscalePhylo` allows for a time-scaled phylogeny to be plotted against geologic time using either the current geologic time scale of Gradstein et al., 2012 or previously published time scales by the International Commission on Stratigraphy. The time scale that is plotted is comprised of a number of temporal components representing the different units that the geological time scale is divided into. There are five main temporal units that can be included, each of which has two alternative names and are as follows: Eon (Eonothem), Era (Erathem), System (Period), Series (Epoch), and Stage (Age). These alternative names can be used interchangeably i.e. both Eon and Erathem are accepted, however should both these alternative names be included then that temporal unit will only be included once. In addition, the order in which they are included into units does not affect the order in which they appear in the chart so `units=c("Period", "Epoch", "Age")` will produce the same results as `units=c("Age", "Epoch", "Period")` with the default order as they were listed previously with Eons plotted at the base and Stages at the top.

Including a user-defined time scale

There is a sixth option that can be included into the `units` argument. "User" allows for an additional temporal unit to be plotted i.e. biozonal or terrane-specific time scales. This requires a matrix of three columns named "Start", "End" and "Name" representing the bottom, top of each temporal bin (in millions of years) and the name to be plotted respectively. An example dataset called `UKzones` representing Stages of the UK Ordovician System is included in the package. See below for an example of how to implement this option.

Stratigraphic ranges

`geoscalePhylo` allows for the stratigraphic ranges to be included in the plot. This requires a matrix with the first appearance and last appearance dates in millions of years (FAD and LAD respectively) with the row names containing all the tip labels of the taxa in the tree, exactly as they appear in `tree$tip.label` and the column names should be "FAD" and "LAD". In order to add the stratigraphic ranges to the plot this matrix should be attached to the argument `ages`. See below for an example of this option.

Apparent appearance of polytomies

It should be noted that using certain methods for time-scaling a tree, such as the "basic" method (the default), can create the appearance of polytomies in a tree that is otherwise fully resolved due to the presence of a large number of zero length branches. This can be solved by using another time-scaling method such as the "equal" method which will enforce all the branches to have a positive length.

Author(s)

Mark A. Bell <mark.bell521@gmail.com>

References

Gradstein, F. M., Ogg, J. M., and Schmitz, M. 2012. A Geologic Time Scale. Elsevier, Boston, USA

Examples

```
### Example lungfish data
data(Dipnoi)

tree_1 <- DatePhylo(Dipnoi$tree, Dipnoi$ages, method="equal", rlen=1)

geoscalePhylo(tree=tree_1, boxes="Age", cex.tip=0.4)

# Plotting the tree with the stratigraphical ranges included
geoscalePhylo(tree=tree_1, ages=Dipnoi$ages, boxes="Age", cex.tip=0.4)

# Including all temporal units into the stratigraphic column
geoscalePhylo(tree_1, Dipnoi$ages, units = c("Eon","Era","Period","Epoch","Age"),
  boxes="Age", cex.tip=0.4)

# Plotting the numerical values on the time scale at Age resolution
geoscalePhylo(tree_1, Dipnoi$ages, units = c("Eon","Era","Period","Epoch","Age"),
  boxes="Age", cex.tip=0.4,tick.scale="Age")

### Example trilobite data
data(Asaphidae)

tree_a <- DatePhylo(Asaphidae$trees[[1]], Asaphidae$ages, method="equal", rlen=1)

geoscalePhylo(ladderize(tree_a), Asaphidae$ages, boxes="Age", x.lim=c(504,435),
  cex.tip=0.5, cex.ts=0.5,vers="ICS2009")

# Plotting the tree vertically
geoscalePhylo(ladderize(tree_a), Asaphidae$ages, boxes="Age", x.lim=c(504,435),
  cex.tip=0.5, cex.ts=0.5,direction="upwards",vers="ICS2009")

# Including a user-defined time scale

data(UKzones)
data(Asaphidae)

tree_a <- DatePhylo(Asaphidae$trees[[1]], Asaphidae$ages, method="equal", rlen=1)
```

```

geoscalePhylo(ladderize(tree_a), Asaphidae$ages, units = c("Eon", "Era", "Period",
  "Epoch", "User"), boxes="Age", cex.tip=0.4, user.scale=UKzones,
  vers="ICS2009", cex.ts=0.5, x.lim=c(440), direction="upwards")

# Rotating the text on the time scale

tree_a <- DatePhylo(Asaphidae$trees[[1]], Asaphidae$ages, method="equal", rlen=1)

geoscalePhylo(ladderize(tree_a), Asaphidae$ages, units = c("Period",
  "Epoch", "Age", "User"), boxes="Age", cex.tip=0.4, user.scale=UKzones,
  vers="ICS2009", cex.ts=0.5, x.lim=c(440), arotate=0, erotate=0, urotate=0)

```

StratPhyloCongruence *Calculates fit to stratigraphy metrics for a set of tree(s).*

Description

Calculates SCI, RCI, MSM*, and GER for a number of topologies.

Usage

```

StratPhyloCongruence(trees, ages, rlen=0, method="basic", samp.perm=1000,
  rand.perm=1000, hard=TRUE, randomly.sample.ages=FALSE, fix.topology=TRUE,
  fix.outgroup=TRUE)

```

Arguments

trees	Input tree(s) as either a phylo or multiphylo object.
ages	A two-column matrix of taxa (rows) against First and Last Appearance Datums (FADs and LADs) to be passed to DatePhylo . Note that rownames should be the taxon names exactly as they appear in <code>tree\$tip.label</code> and colnames should be "FAD" and "LAD". All ages should be in time before present.
rlen	Root length, to be passed to DatePhylo .
method	Tree dating method, to be passed to DatePhylo .
samp.perm	Number of sampled trees to be produced by resolving polytomies and/or drawing random dates for the tips for the input trees.
rand.perm	Number of random trees to be produced in calculating probabilities for the input trees, and (if used) the sampled trees.
hard	Whether to treat polytomies as hard or soft. If FALSE polytomies are resolved randomly.
randomly.sample.ages	Whether to treat FAD and LAD as a range (<code>randomly.sample.ages=FALSE</code>) or an uncertainty (<code>randomly.sample.ages=TRUE</code>). If the latter then two ages are randomly sampled from the range and these are used as the FAD and LAD.

<code>fix.topology</code>	Whether to allow tree shape to be random (<code>fix.topology=FALSE</code>) or to reflect the tree shape of the input tree(s) (<code>fix.topology=TRUE</code>).
<code>fix.outgroup</code>	Whether to force the randomly generated trees to share the outgroup of the first input tree (<code>fix.outgroup=TRUE</code>) or not (<code>fix.outgroup=FALSE</code>).

Details

Cladograms of fossil taxa make explicit predictions about the successive appearance of taxa in the fossil record that can be compared with their observed stratigraphic ranges. Several methods have been developed to quantify this "fit to stratigraphy" of phylogenetic hypotheses, and these can be assessed in both a statistic (measuring the apparent strength of this congruence) and an associated significance test (p-value) based on generating random topologies for the same taxon set.

This function produces both values for all four main metrics: the Stratigraphic Consistency Index (SCI; Huelsenbeck 1994), the Relative Consistency Index (RCI; Benton and Storrs 1994), the Manhattan Stratigraphic Measure (MSM*; Siddall 1998; Pol and Norell 2001), and the Gap Excess Ratio (GER; Wills 1999).

SCI - Stratigraphic Consistency Index

The SCI works by assessing the "consistency" of nodes. A node is considered stratigraphically consistent if its oldest descendant is the same age or younger than the oldest descendant of the preceding node. The SCI is thus given simply as:

$$SCI = C/N$$

Where C is the sum of all the consistent nodes and N is the total number of nodes - 1. (As there is no node preceding the root there is no basis on which to estimate its consistency.) This value can range from zero (maximally inconsistent) to one (maximally consistent). However, a potential criticism of the SCI is that a high value may be returned when in fact a single inconsistent node may represent a very large amount of missing history (measured in unsampled units or millions of years), whereas a low SCI may represent relatively few unsampled units or millions of years.

RCI - Relative Completeness Index

The RCI was the first method to explicitly account for the absolute amount of missing data implied by the tree. This figure is usually expressed as the Minimum Implied Gap (MIG), a term also used by both the MSM and GER (see below), and corresponds to the sum of the branch lengths excluding the duration of the terminals (the observed ranges of the taxa). The RCI expresses the MIG as a proportion of the sum of the observed ranges (Simple Range Length; SRL) of the taxa converted to a percentage:

$$RCI = (1 - (MIG/SRL)) * 100percent$$

Importantly this value is not confined to a 0 to 100 percent scale, and can have both negative values and values greater than 100 percent, which can make it difficult to interpret.

MSM - Manhattan Stratigraphic Measure

The MSM was the first method to account for both the absolute MIG and range on a confined zero to one scale. It is expressed as:

$$MSM = L_m/L_0$$

Where L_0 is the length of the tree expressed by optimising times of first appearance on to the tree as a Sankoff character and taking the total length. L_m represents the same process, but for the optimal possible tree given the same set of first appearances. However, Pol and Norell (2001) noted a critical flaw in this approach, specifically that the Sankoff optimisation is reversible, meaning that nodes in the topology are allowed to be younger than their descendants, leading in some cases to a poor fit to stratigraphy being perceived as a good fit. Instead they suggest modifying the character step matrix to make the cost of reversals effectively infinite and hence impossible. Thus the values for L_0 and L_m are modified accordingly. This approach they termed MSM* and is the implementation of MSM used here. This statistic can be expressed as:

$$MSM^* = G_{min}/MIG$$

Where G_{min} represents the *MIG* for the tree with the optimal fit to stratigraphy. In effect this is a completely unbalanced tree where the youngest pair of taxa are the most deeply nested and successive outgroups represent the next oldest taxon. Theoretically MSM* ranges from one (the best fit as the observed tree is the maximally consistent tree) to zero (the least optimal tree). However, in effect no tree can have a value of zero as its *MIG* would have to be equal to infinity.

GER - Gap Excess Ratio

The GER represents a method that accounts for *MIG*, ranges from zero to one, and the best and worst fits to stratigraphy are both practically realisable. It can be expressed as:

$$GER = 1 - ((MIG - G_{min}) / (G_{max} - G_{min}))$$

Where G_{max} represents the *MIG* of the tree with the worst possible fit to stratigraphy. This is in effect any topology where the oldest taxon is the most deeply nested such that every clade in the tree contains it and hence must be minimally that old.

P-values

In isolation all four methods suffer from an inability to reject the null hypothesis that an apparent good fit to stratigraphy may be generated by chance alone. In practice this can be tested by generating a set of random topologies, calculating the fit to stratigraphy measure, and then either fitting a normal distribution to the resulting values (to get an estimated p-value) or assessing the relative position of the *MIG* of the observed (and sampled) tree(s) to get an absolute p-value. (Note that the assumption of normality may not always hold and the former approach should be used at the user's discretion. However, it should be noted that for the SCI, MSM*, and GER p-values are calculated after first transforming the data by taking the arcsine of the square root of each value.) The reason for having two sets of p-values is that if the observed trees fall completely outside the range of the random topologies they will be given an extreme p-value (0 or 1) that may be misleading. In such cases the estimated value may be more accurate.

P-values should be interpreted as the probability of the null: that the observed tree(s) have an equal or worse fit to stratigraphy than the sample of random trees. Thus if the p-values are very small the user can reject the null hypothesis in favour of the alternative: that the observed tree(s) have a better fit to stratigraphy than expected by chance alone.

Modifications of the GER

More recently Wills et al. (2008) introduced two new versions of the GER that take advantage of the distribution of *MIGs* from the set of randomly generated topologies. The first of these (GERT)

uses the extreme values of the random topologies as modified versions of G_{max} and G_{min} , termed Gt_{max} and Gt_{min} respectively. GERT is thus expressed as:

$$GERT = 1 - ((MIG - Gt_{min}) / (Gt_{max} - Gt_{min}))$$

In practice the MIG of the observed tree(s) may fall outside of these ranges so here a correction factor is employed so that any value below zero is corrected to zero, and any value above one is corrected to one. An additional stipulation for GERT is that the overall tree topology is fixed and only the taxa themselves are shuffled. This is to give a more realistic set of random topologies as there are known biases towards unbalanced trees in many palaeontological data sets. Here this is implemented by selecting the `fix.topology=TRUE` option. However, here GERT can also be calculated when `fix.topology=FALSE`.

A second modification of GER is to use the position of the observed tree(s) in the sample of randomly generated topologies, such that:

$$GER* = 1 - (Fraction\ of\ distribution\ \leq\ MIG)$$

Thus if the MIG of the observed tree(s) is less than any randomly generated topology $GER*$ will be one (maximally optimal fit) and if it worse than any of the randomly generated topologies it will be zero (maximally suboptimal fit).

Note: it is recommended that you use a large number of random topologies in order to get reliable values for GERT and $GER*$ using `rand.perm=N`. Wills et al. (2008) used 50000, but the user should note that for many real world examples such values will take many hours to run.

Wills et al. (2008) also introduced the notion of referring to intervals sampled rather than absolute time by recasting MIG as $MIGu$: the sum of ghost ranges for intervals of unit length. Although not directly implemented here this can be done manually by converting the time values (in Ma) used to simple unit counts such that FADs and LADs of taxa are given as numbered time bins (the youngest being 1 and the oldest N, where there are N time bins).

Polytomies and age uncertainties

Alongside the input trees the user can also create an additional set of sampled trees based on the input trees. This option is automatically implemented when choosing either `hard=FALSE` or `randomly.sample.ages=TRUE`, and the total number of permutations to perform dictated by `samp.perm=N`. This process works by first sampling from the set of input tree(s) and then randomly resolving any polytomies (if `hard=FALSE`) to ensure all sampled trees are fully dichotomous. (At present the function does not allow the various options laid out in Boyd et al. 2011, but the user can achieve this effect by modifying the input trees themselves.) Then if `randomly.sample.ages=TRUE` the FAD and LAD are treated as bounds of a uniform distribution which is sampled at random. This allows the user to get results for a set of trees that account for uncertainty in dating (as outlined in Pol and Norell 2006). (Note that two dates are picked for each taxon to avoid the problem of having an SRL of zero that would cause a divide by zero error for the RCI metric.)

All fit to stratigraphy measures calculated for the input trees are then repeated for the sampled trees. However, if both `hard=TRUE` and `randomly.sample.ages=FALSE` (the defaults) no set of sampled trees will be created.

In all cases when using the function users will see a progress bar that indicates the general progress through the combined set of trees (input, sampled, and randomly generated). This serves as a useful indicator of the time it will take for the function to finish. Here default values for `samp.perm` and

rand.perm are both set at 1000, but the user may wish to lower these (to decrease calculation time) or increase them (to enhance accuracy).

Additional options

Note that because this function uses [DatePhylo](#) the user has the option of using different tree dating algorithms than the basic method (equivalent to the basic method in the paleotree package) employed in all the published studies cited above (and the default option here). The dating method used will apply to all trees generated, including the input, sampled, and randomly generated topologies. In all cases the time-scaled trees are returned with the function output.

A final option (`fix.outgroup=TRUE`) allows the user to always use the same outgroup taxon (based on the first input tree) for all randomly generated topologies. Because the outgroup will often be the oldest taxon and its position in the input topologies is not allowed to vary letting it do so in the random topologies may lead to inferring a better fit to stratigraphy for the observed tree(s) than is fair. Fixing the outgroup thus ameliorates this potential bias and is the default option here.

Value

input.tree.results

A matrix with a row for each input tree and columns indicating the values for SCI, RCI, GER and MSM* and their estimated probabilities assuming a normal distribution (`est.p.SCI`, `est.p.RCI`, `est.p.GER`, and `est.p.MSM*`) as well as `GERT`, `GER*`, `MIG`, and `p.Wills` (their probability as position within the MIGs of the random topologies).

samp.permutation.results

If used, a matrix with a row for each sampled tree (up to `samp.perm`) and columns indicating the values for SCI, RCI, GER and MSM* and their estimated probabilities assuming a normal distribution (`est.p.SCI`, `est.p.RCI`, `est.p.GER`, and `est.p.MSM*`) as well as `GERT`, `GER*`, `MIG`, and `p.Wills` (their probability as position within the MIGs of random topologies).

rand.permutations

A matrix with a row for each randomly generated tree (up to `rand.perm`) and columns indicating the values for SCI, RCI, GER, MSM*, and MIG.

input.trees

The input tree(s) as a phylo or multiphylo object, with branches scaled to time according to the input values passed to [DatePhylo](#).

samp.trees

The sampled tree(s) as a phylo or multiphylo object, with branches scaled to time according to the input values passed to [DatePhylo](#).

rand.trees

The randomly generated tree(s) as a phylo or multiphylo object, with branches scaled to time according to the input values passed to [DatePhylo](#).

Author(s)

Mark A. Bell <mark.bell1521@gmail.com> and Graeme T. Lloyd <graemetlloyd@gmail.com>

References

Benton, M. J. and Storrs, G. W., 1994. Testing the quality of the fossil record: palaeontological knowledge is improving. *Geology*, 22, 111-114.

Boyd, C. A., Cleland, T. P., Marrero, N. L. and Clarke, J. A., 2011. Exploring the effects of phylogenetic uncertainty and consensus trees on stratigraphic consistency scores: a new program and a standardized method. *Cladistics*, 27, 52-60.

Huelsenbeck, J. P., 1994. Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology*, 20, 470-483.

Pol, D. and Norell, M. A., 2001. Comments on the Manhattan Stratigraphic Measure. *Cladistics*, 17, 285-289.

Pol, D. and Norell, M. A., 2006. Uncertainty in the age of fossils and the stratigraphic fit to phylogenies. *Systematic Biology*, 55, 512-521.

Siddall, M. E., 1998. Stratigraphic fit to phylogenies: a proposed solution. *Cladistics*, 14, 201-208.

Wills, M. A., 1999. Congruence between phylogeny and stratigraphy: randomization tests and the Gap Excess Ratio. *Systematic Biology*, 48, 559-580.

Wills, M. A., Barrett, P. M. and Heathcote, J. F., 2008. The modified Gap Excess Ratio (GER*) and the stratigraphic congruence of dinosaur phylogenies. *Systematic Biology*, 57, 891-904.

Examples

```
# Calculate stratigraphic fit measures treating ages as ranges
# (permutation numbers used are lower than recommended for standard use):
fit.to.strat.1 <- StratPhyloCongruence(trees=Dipnoi$tree, ages=Dipnoi$ages, rlen=0,
method="basic", samp.perm=5, rand.perm=5, hard=TRUE,
randomly.sample.ages=FALSE, fix.topology=TRUE, fix.outgroup=TRUE)
fit.to.strat.1 # View all output
names(fit.to.strat.1) # Show output options
fit.to.strat.1$input.tree.results # Show just the output for the input tree(s)

# Calculate stratigraphic fit measures treating ages as uncertainties
# (permutation numbers used are lower than recommended for standard use):
#fit.to.strat.2 <- StratPhyloCongruence(trees=Dipnoi$tree, ages=Dipnoi$ages, rlen=0,
# method="basic", samp.perm=10, rand.perm=10, hard=TRUE,
# randomly.sample.ages=TRUE, fix.topology=TRUE, fix.outgroup=TRUE)
```

UKzones

British regional stages for the Ordovician

Description

The stratigraphic ranges for the British stages of the Ordovician.

Usage

```
data(UKzones)
```

Format

A matrix containing the start and end ages for the British stage subdivisions of the Ordovician.

References

Webby, B.D., Paris, F., Droser, M.L., and Percival, I.G. (editors), 2004. The Great Ordovician Biodiversity Event. Columbia University Press, New York, 496 pp.

Index

- *Topic **congruence**
 - StratPhyloCongruence, [10](#)
 - *Topic **datasets**
 - Asaphidae, [2](#)
 - Dipnoi, [5](#)
 - UKzones, [15](#)
 - *Topic **phylogeny**
 - FindDescendants, [6](#)
 - strap-package, [2](#)
 - *Topic **time scale, geological, plot, phylogeny, tree**
 - geoscalePhylo, [6](#)
 - *Topic **time-scaling, phylogeny, fossil**
 - DatePhylo, [3](#)
- Asaphidae, [2](#)
- DatePhylo, [3](#), [10](#), [14](#)
- Dipnoi, [5](#)
- FindDescendants, [6](#)
- geoscalePhylo, [6](#)
- strap (strap-package), [2](#)
- strap-package, [2](#)
- StratPhyloCongruence, [10](#)
- UKzones, [15](#)