Package 'MEDUSA'

July 24, 2012

Type Package

Title MEDUSA: Modeling Evolutionary Diversification Using Stepwise AIC

Version 0.93-4-6

Date 2012-07-24

Author Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer Joseph W. Brown <josephwb@uidaho.edu>

Description Fits piecewise diversification models from a time-
calibrated phylogenetic tree and complete extant species richness.

License GPL (>= 2)

LazyLoad yes

Depends ape, geiger, colorspace, shape, stringr

Suggests multicore (>= 0.1-3)

Enhances geiger

R topics documented:

MEDUSA-package	*MEDUSA: Modeling Evolutionary Diversification Using Stepwise AIC*

Description

Fits piecewise birth-death models to an ultrametric phylogenetic tree according to phylogenetic (edge-length) and taxonomic (richness) likelihoods. A reimplementation of the original **geiger** MEDUSA method, meant especially for the analysis of large trees. Optimal model size is determined via a stepwise AIC approach. Run with `MEDUSA` and summarize results with `medusaSummary`. Publication-quality trees with coloured branches (reflecting piecewise model membership) can be plotted using `plotPrettyTree`.

Details

Package:	MEDUSA
Type:	Package
Version:	0.19
Date:	2011-08-03
License:	GPL version 2 or greater
LazyLoad:	yes

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

References

Alfaro, ME, F Santini, C Brock, H Alamillo, A Dornburg, DL Rabosky, G Carnevale, and LJ Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* **106**: 13410-13414.

fitSisters	*Fit piecewise diversification models to sister clades in a time-calibrated phylogeny*

Description

Fits individual diversification models to the two sister clades branching from a specified node.

Usage

```
fitSisters(phy, richness=NULL, node=NULL, tips=NULL, model="mixed",
criterion="aicc", epsilon=NULL, r=NULL, b=NULL, d=NULL, plotSurface=FALSE,
initialR=0.05, initialE=0.5, verbose=TRUE, mc=FALSE, numCores=NULL, ...)
```

Arguments

phy	Ultrametric phylogenetic tree. Assumed to be of class `phylo` or (if multiple trees) `multiPhylo`.
richness	Optional; only required if tree is NOT completely sampled. A dataframe with minimally two columns `taxon` and `n.taxa`. The identifier `taxon` must match exactly with a tip.label in the phylogeny phy; this is checked. May also include an `exemplar` column, used for renaming incompletely-sampled clades that require collapsing; in this case, `exemplar` must match exactly with a tip label and `taxon` will be the name for that clade (say, genus, family, etc.). If no richness information is provided then it is assumed tips represent single species with complete sampling.
node	The node where the two clades of interest share a common ancestor. Value may be `node="root"` if the root node, or a numeric value based on APE's node numbering system. If the node number is unknown, `tips` can be provided instead (see below).
tips	If the node number (using APE's numbering system) of a node of interest is not known, tip labels for taxa sharing a MRCA at that node can instead be passed in. Specified as: `tips=c("tax1", "tax2")`, where tip labels must be given in quotes.
model	The flavour of piecewise models to be fit to the phylogeny. Can be birth-death (`model="bd"`), pure-birth (`model="yule"`), or mixed (`model="mixed"`) (the default). For `mixed` model analyses, both BD and Yule models are considered for all potential breaks.
criterion	The statistical criterion used to determine whether larger models fit the data significantly better. Options are `criterion="aicc"` (default) or `criterion="aic"`. As AIC and AICc will select the same breaks for any given model size, this option is only useful if `stop="threshold"`. In this case, the default is strongly advised, as for some trees AIC can continue to improve well after AICc has deteriorated.
epsilon	A fixed value for the extinction fraction. Will be held constant while r (net diversification) is estimated. Only for use in the birth-death model.
r	A fixed value for the net diversification rate. Will be held constant while epsilon (extinction fraction) is estimated. Only for use in the birth-death model.
b	A fixed value for the per-lineage speciation rate. Only for use in the birth-death model.
d	A fixed value for the per-lineage extinction rate. Will be held constant while epsilon (extinction fraction) is estimated.
plotSurface	Whether parameter likelihood surfaces should be plotted. Default is FALSE.
initialR	Starting guess for net diversification rate. May need to be decreased for the program to work on large trees.
initialE	Starting guess for relative extinction rate.
verbose	Print out additional information to screen, such as which taxa are pruned from the original tree given richness information. Mostly useful for debugging.
mc	Indicate whether multiple cores should be used (mc=TRUE). Works only on unix-like machines (i.e. not Windows), and only in non-GUI instantiations of R.
numCores	The number of cores to use if `mc=TRUE`. By default all cores detected will be used. This option is only useful if one wishes to limit the number of cares devoted to `MEDUSA` analysis.
...	Potential additional arguments for plotting etc.

Details

A diversification model is first fit to the entire subtree rooted at node node. Separate diversification models are then fit to the two clades emanating from node. MLEs and associated precision intervals are reported for parameters in both the Base and Sisters models. If a single tree in passed in, confidence intervals are calculated using profile likelihoods. If multiple trees are analyzed, the following are reported for each parameter: mean, min, max, and standard deviation.

In addition, the relative fit of the Base and Sisters models is also reported using AIC weights. In the case of multiple trees, the mean and precision of model weights is presented, as well as the proportion of times each model is the better fit model.

Value

An object of class "sisterFit". Returned invisibly. This is a list with elements:

$parameterSummary

A dataframe listing parameter estimates and precision statistics.

$modelSummary   A dataframe summarizing the relative fit of a single model vs. the implemented 2-piece sister-lineages model.

$fixPar   The fixed parameter value used in a constrained model. If a constrained model is not implemented, this will be NULL.

$models   A list containing the base and sister-clade diversificaton models estimated for each tree. Includes tree(s) (which may have been pruned using richness information).

Author(s)

Joseph W. Brown.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

---

getBD   *Get per-lineage birth and death rates*

Description

Get per-lineage birth (speciation) and death (extinction) rates from inferred net diversification (r) and relative extinction (epsilon) rates.

Usage

```
getBD(r, epsilon)
```

Arguments

r   Net diversification rate (b-d)

epsilon   Relative extinction rate (d/b)

Details

Useful if one is in interested in b (lambda) & d (mu) rather than r and epsilon.

Value

Returns a list containing:

$b	Per-lineage birth (speciation) rate
$d	Per-lineage death (extinction) rate

MEDUSA	*Fit piecewise diversification models to a time-calibrated phylogeny*

Description

Fits piecewise birth-death models to an ultrametric phylogenetic tree according to phylogenetic (edge-length) and taxonomic (richness) likelihoods. Optimal model size is determined via a stepwise AIC approach.

Usage

```
MEDUSA(phy, richness=NULL, model="mixed", modelLimit=20, stop="threshold",
shiftCut="both", criterion="aicc", stepBack=TRUE, preserveModelFlavour=FALSE, epsilon=NULL,
r=NULL, b=NULL, d=NULL, fixThreshold=NULL, initialR=0.05, initialE=0.5,
verbose=TRUE, mc=FALSE, numCores=NULL, resolveTree=FALSE, ...)
```

Arguments

phy	Ultrametric phylogenetic tree. Assumed to be of class phylo or (if multiple trees) multiPhylo.
richness	Optional; only required if tree is NOT completely sampled. A dataframe with minimally two columns taxon and n.taxa. The identifier taxon must match exactly with a tip.label in the phylogeny phy; this is checked. May also include an exemplar column, used for renaming incompletely-sampled clades that require collapsing; in this case, exemplar must match exactly with a tip label and taxon will be the name for that clade (say, genus, family, etc.). If no richness information is provided then it is assumed tips represent single species with complete sampling.
model	The flavour of piecewise models to be fit to the phylogeny. Can be birth-death (model="bd"), pure-birth (model="yule"), or mixed (model="mixed") (the default). For mixed model analyses, both BD and Yule models are considered for all potential breaks.
modelLimit	The number of piecewise models to be fit to the phylogeny. The program may overrule this value and fit fewer models is the tree is small (such that the denominator in the AIC correction factor becomes undefined).
stop	The criterion for when the analysis should be terminated. stop="modelLimit" will terminate after modelLimit piecewise models have been fit to the the phylogeny (see above). The default criterion is stop="threshold". Here, threshold refers to the improvement in AIC score that should be considered significant given the number of tips in the phylogeny; analysis will terminate when the addition of a piecewise BD model does not improve AIC score by threshold.

shiftCut	Determines where rate shifts are placed on the tree. Three options are available. For the first option shiftCut="stem", when a subtree is selected for sub-model consideration, it is assumed that the shift in diversification occurred in the branch leading to the clade. The second option shiftCut="node" assumes the shift occurred at the node representing the most recent common ancestor of the clade (that is, omitting the leading branch). The final and default option shiftCut="both" considers both possibilities for shift placement, and carries the shift with the higher AICc score forward to piecewise model selection.
criterion	The statistical criterion used to determine whether larger models fit the data significantly better. Options are criterion="aicc" (default) or criterion="aic". As AIC and AICc will select the same breaks for any given model size, this option is only useful if stop="threshold". In this case, the default is strongly advised, as for some trees AIC can continue to improve well after AICc has deteriorated.
stepBack	Determines whether parameter removal should be considered. Default is TRUE.
preserveModelFlavour	
	Only for use when using model="mixed". When a clade has been fit with model i (either "yule" or "bd"), and a rate shift is introduced within the clade, this option determines whether the basal rate is reestimated under model i (preserveModelFlavour=TRUE) or alternative models flavours are considered (preserveModelFlavour=FALSE). The default is to not preserve model flavour.
epsilon	A fixed value for the extinction fraction. Will be held constant while r (net diversification) is estimated. Only for use in the birth-death model.
r	A fixed value for the net diversification rate. Will be held constant while epsilon (extinction fraction) is estimated. Only for use in the birth-death model.
b	A fixed value for the per-lineage speciation rate. Only for use in the birth-death model.
d	A fixed value for the per-lineage extinction rate. Will be held constant while epsilon (extinction fraction) is estimated.
fixThreshold	User-provided value for what improvement in AIC score is deemed significant. A value provided for fixThreshold will overrule the default value determined through simulations, whether the provided value is more liberal or conservative than the default. This value is used to determine when to terminate the analysis when stop="threshold" (see above).
initialR	Starting guess for net diversification rate. May need to be decreased for the program to work on large trees.
initialE	Starting guess for relative extinction rate.
verbose	Print out additional information to screen, such as which taxa are pruned from the original tree given richness information. Mostly useful for debugging.
mc	Indicate whether multiple cores should be used (mc=TRUE). Works only on unix-like machines (i.e. not Windows), and only in non-GUI instantiations of R.
numCores	The number of cores to use if mc=TRUE. By default all cores detected will be used. This option is only useful if one wishes to limit the number of cares devoted to MEDUSA analysis.
resolveTree	MEDUSA maths require a fully resolved (i.e. binary) tree. If your tree contains polytomies, MEDUSA will complain and refuse to continue with analysis. If the tree is not binary and resolveTree is set to TRUE, polytomies will be randomly resolved, with zero-length edges inserted into the tree.
...	Potential additional arguments for plotting etc.

Details

A series of diversification models is fit to a combination of phylogenetic and taxonomic data. The input is a phylogenetic tree with branch lengths proportional to time showing the relationship amongst clades, and the diversity for living species in all of those clades. All taxa missing from the tree must be assigned to one of the tip clades in the richness matrix.

The algorithm first fits a single diversification model to the entire dataset. Then, a series of breaks are added, so that different parts of the tree evolve with different parameter values (per-lineage net diversification and relative extinction rates). The program first compares all single-breakpoint models with the overall model, and selects the best one. Then all possible two-breakpoint models are compared with the best single-breakpoint model, and so on.

If phy is of class multiPhylo then the analysis will be carried out on each tree in succession.

Summarize results with medusaSummary.

Value

An object of class "medusa". Returned invisibly. This is a list with elements:

$desc	A list of all descendants for each node in the tree; for summarizing results only.
$phy	The phylogeny that was analyzed. May have been pruned (using richness information) from originally tree passed in. Mainly for summarizing results.
$modelSummary	A dataframe containing break and fit values for optimal models at each size considered.
$fixPar	The fixed parameter value used in a constrained model. If a constrained model is not implemented, this will be NULL.
$threshold	The correct AIC threshold given the size of the tree analyzed. Used in summarizing results and (possibly) terminating the analysis.
$models	A list containing the optimal models at each successive size. For each model contains:

- $z A matrix listing branch times and richnesses; for summarizing results only.
- $par A matrix containing estimates of r and epsilon for all piecewise models.
- $lnLik.part The likelihoods of each piecewise model.
- $lnLik The total likelihood for the piecewise model.
- $split.at The node(s) where individual BD models originate.
- $aic AIC for the piecewise model.
- $aicc AICc for the piecewise model.
- $num.par The number of parameters (total r + epsilon + breakpoints) in the piecewise model.

If phy is of class multiPhylo then the returned object is of class "multiMedusa", and is a list of class "medusa" objects.

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

References

Alfaro, ME, F Santini, C Brock, H Alamillo, A Dornburg, DL Rabosky, G Carnevale, and LJ Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* **106**: 13410-13414.

See Also

medusaSummary plotPrettyTree

---

medusaSummary	*Summarize the fit of a piecewise diversification model*

Description

Summarize results from a MEDUSA analysis.

Usage

```
medusaSummary(results, plotTree=TRUE, time=TRUE, node.labels=TRUE, colourOffset=0,
cex=0.5, plotSurface=FALSE, printTitle=TRUE, n.points=100, ...)
```

Arguments

results	An object of class "medusa" produced from a MEDUSA analysis.
plotTree	Boolean value that determines whether the fitted tree is plotted. If plotTree=TRUE (the default) the tree is plotted with edges coloured according to which piecewise model they belong to, and positions (nodes) of break points are potentially indicated (see node.labels below).
time	Boolean value that indicates whether a timescale is printed below the tree. Only has meaning if plotTree=TRUE. For large trees it might be desirable to set time=FALSE to make tip labels legible.
node.labels	Boolean value that indicates whether node breaks are plotted on the tree. Only has meaning if plotTree=TRUE.
colourOffset	A plotting option for the colouring of branches by piecewise model membership. An integer value > 1; sets the colour cycling from a different starting point. Only has meaning if plotTree=TRUE.
cex	Specifies the font size for printing tip labels. Default is cex=0.5; may need to decrease for larger trees. Only has meaning if plotTree=TRUE.
plotSurface	Boolean value that indicates whether likelihood surface(s) for the selected piecewise model is plotted. If plotSurface=TRUE the optimal model is evaluated at n.points (see below), and contours are plotted. Default is no plot.
printTitle	Whether to print titles on surface plots.
n.points	The number of values at which r and epsilon are evaluated for the selected piecewise model (hence, a total of n.points x n.points parameter combinations considered for model "bd"). A greater value of n.points gives better resolution of the surface, but comes with a computational burden. Default is n.points=100. Only has meaning if plotSurface=TRUE.
...	Potential additional arguments for plotting etc.

Value

An object of class "medusa.sum". This is a list with elements:

$z	A matrix listing branch times and richnesses, coded by model membership. The model used is either the optimal model (the default) or a user-selected model.
$mm	Colour-coding for edges.
$break.point	The position in the tree where shifts in models occur.
$phy	The tree (possibly pruned from the original tree using richness information).

This object can be passed to plotPrettyTree to produce publication-quality trees.

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

See Also

MEDUSA plotPrettyTree

---

multiMedusaSummary	*Summarize the fit of piecewise diversification models across a distribution of trees*

Description

Summarize results from MEDUSA analysis across a distribution of trees.

Usage

```
multiMedusaSummary(res, conTree, cutOff=0.05, plotModelSizes=FALSE, plotTree=TRUE, cex=0.5, ...)
```

Arguments

res	An object of class "multiMedusa" produced from a MEDUSA analysis across a distribution of trees.
conTree	Tree of class "phylo" on which to summarize results. Although in principle any tree can be used, ideally this tree should share as many edges as possible with those across the distribution of analyzed trees e.g. some sort of consensus or maximum clade credibility tree. Rate shift positons and average edge-specific rates are mapped onto this tree. May be pruned using richness information (a component of results) to correspond with tip sampling of trees that were analyzed.
cutOff	Determines the threshold (expressed as a proportion) for which splits are recorded on conTree. Default is to keep shifts that occur in > 5
plotModelSizes	Boolean value that determines whether a histogram of the number of piecewise diversification models per tree is plotted. Default is no plot.
plotTree	Boolean value that determines whether an annotated tree is plotted. Default is to plot. A possible reason to set plotTree=FALSE is that the plotMultiMedusa function gives greater flexibility.

| cex | Specifies the font size for printing tip labels. Default is cex=0.5; may need to decrease for larger trees. Only has meaning if plotTree=TRUE. |
| ... | Potential additional arguments for plotting etc. |

Value

An object of class "multiMedusaSummary". This is a list with elements:

$model.sizes	A numeric vector of model sizes (i.e. number of splits) across all trees.
$num.trees	The number of trees analyzed.
$shift.summary	The position in the tree where shifts in models occur, together with their frequecny across trees.
$summary.tree	The tree (possibly pruned from the original tree using richness information) on which the results were summarized. Of class "phylo". In addition to standard "phylo" elements, includes a matrix $rates. This matrix is ordered according to summary.tree$edge, and includes columns: r.mean, r.median, r.sd, eps.mean, eps.median, eps.sd, and freq. freq corresponds to the frequency with which an edge is present accross all trees.

This object can be passed to [plotMultiMedusa](#) to produce publication-quality trees.

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

See Also

[MEDUSA](#) [plotMultiMedusa](#)

| plotMultiMedusa | *Plot the summary tree from a multiMEDUSA analysis.* |

Description

Plot a consensus tree annotated with results from a [MEDUSA](#) analysis across a distribution of trees.

Usage

```
plotMultiMedusa(summary, treeRearrange="down", annotateShift=TRUE, annotateRate="r.mean",
plotRichnesses=TRUE, richPlot="log", time=TRUE, tip.cex=0.3, shiftScale=1, label.offset=0.5, fon
shift.leg.pos="left", power=1.5, ...)
```

Arguments

summary	An object of class "multiMedusaSummary" produced from a [multiMedusaSummary](#) summary across a distribution of trees.
treeRearrange	Ladderize the tree for nicer viewing. Default is treeRearrange="up". Possible values are "up", "down", or NULL (for no rearrangement of edges).
plotRichnesses	Boolean value that determines whether bar plot of extant tip richnesses is plotted. Default is true.

richPlot	Flavour of richness to plot. Only has meaning if `plotRichnesses=TRUE`. By default, richnesses are plotted as `"log"`. Alternative is `"raw"` i.e. straight richnesses are plotted.
annotateShift	Boolean value that determines whether shift positions are plotted on the tree. Default is `true`.
annotateRate	Specifies which rate to annotate the tree with. Possibilities are: `"r.mean"`, `"r.median"`, `"r.sd"`, `"eps.mean"`, `"eps.median"`, and `"eps.sd"`.
time	Boolean value that indicates whether a timescale is printed below the tree. For large trees it might be desirable to set `time=FALSE` to make tip labels legible.
tip.cex	Specifies the font size for printing tip labels. Default is `cex=0.5`; may need to decrease for larger trees. Only has meaning if `plotTree=TRUE`.
shiftScale	Alter the size of plotted shift proportions. Takes numeric value.
label.offset	Specifies how far tip labels are offset rightward from the default `plot.phylo()` settings.
font	Specifies the font to display for tip labels. The default `font=3` plots labels in italic. Options are `font=1` (plain text), `font=2` (bold), `font=3` (italic, the default), and `font=4` (bold italic).
shift.leg.pos	Specifies the position of the shift legend. Default is `shift.leg.pos="bottomleft"`.
power	Alters the colour scheme for annotating rates.
...	Potential additional arguments for plotting etc.

Value

None.

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

See Also

MEDUSA multiMedusaSummary

---

plotPrettyTree	*Produce a publication-quality tree figure indicating the position of rate shifts*

Description

Plot a publication-quality colour-coded tree, where edge colours indicate piecewise diversification model membership from a MEDUSA analysis. A wrapper for plot.phylo in the APE package.

Usage

```
plotPrettyTree(treeParameters, time=TRUE, node.labels=FALSE, margin=FALSE, cex=0.5,
label.offset=0, font=3, colourTipLabels=FALSE, ...)
```

Arguments

treeParameters	An object of class "medusa.sum" produced from a medusaSummary af a MEDUSA analysis. Contains the tree and piecewise diversification model membership.
time	Whether a timescale should be plotted. Possibly not wanted for large trees. Default=true;
node.labels	Whether node labels should be plotted on the tree to indicate the position and order of diversification rate shifts.
margin	Whether plot should include margins. time (above) overrules this option, as a margin is required for plotting a timescale.
cex	Specifies the font size for printing tip labels. Default is cex=0.5; may need to decrease for larger trees.
label.offset	Specifies how far tip labels are offset rightward from the default plot.phylo() settings.
font	Specifies the font to display for tip labels. The default font=3 plots labels in italic. Options are font=1 (plain text), font=2 (bold), font=3 (italic, the default), and font=4 (bold italic).
colourTipLabels	
	Determines whether tip labels should be coloured according to their piecewise diversification model membership. Default=FALSE.
...	Potential additional arguments to be passed to plot.phylo.

Value

None.

Author(s)

Joseph W. Brown, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon.

Maintainer: Joseph W. Brown <josephwb@uidaho.edu>

See Also

MEDUSA medusaSummary

---

whales	*Whale phylogeny*

Description

Tree, species richnesses.

Usage

```
data(whales)
```

Format

The data are stored in RData (binary) format.

Source

Graham Slater, PaleoBioDB.

Examples

```
data(whales)
str(whales)
```

Index