**Package ‘phylometric’**

Jan 6, 2015

**Version**  0.1

**Date** 2015-01-06

**Title** Estimating statistical errors of phylogenetic metrics

**Depends** R (>= 2.1.1)

**Imports** mvtnorm

**ZipData** no

**Description**

phylometrics provides functions to estimate statistical errors of phylogenetic metrics particularly to detect binary trait influence on diversification, conduct validation test on macroevolutionary models suggested by other approaches, as well as a function to simulate trees backwardly with fixed number of taxa, trait prevalence, and sampling fraction.

**License** GPL (>=2)

**Author** Xia Hua, Lindell Bromham

**Maintainer** Xia Hua <xia.hua@anu.edu.au>

**Repository** CRAN

**Date/Publication** 2015

**Citation**

Hua, X. & Bromham, L. (submitted) Phylometrics: An R package for detecting macroevolutionary patterns, including estimation of statistical errors of phylogenetic metrics and backward tree simulation under a binary state birth-death model. *Methods in Evolution and Ecology.*

R topics documented:

treesim3

treestat4

tars5

noto6

sscd6

fpd7

|  |  |
| --- | --- |
| treesim | *Simulate trees with fixed number of sampled taxa and trait prevalance* |

**Description**

This function generates a tree that contains a defined number of sampled taxa for each of the two trait states, under trait-dependent speciation and extinction.

**Usage**

treesim(pars, N0, N1, sampling.f, max.t)

**Arguments**

|  |  |
| --- | --- |
| pars | a vector of parameters that describes the macroevolutionary processes of the tree. Parameters are in order of speciation rate for trait state 0, speciation rate for trait state 1, extinction rate for trait state 0, extinction rate for trait state 1, transition rate from state 0 to state 1, transition rate from state 1 to state 0. |
| N0 | the number of sampled taxa with trait state 0. |
| N1 | the number of sampled taxa with trait state 1. |
| sampling.f | a vector of sampling fraction of taxa with trait state 0 and trait state 1. |
| max.t | the maximum amount of time, above which tree simulation stops and reports the tree as not being able to coalesce. |

**Value**

An object of class “phylo” with the following elements:

|  |  |
| --- | --- |
| edge | a two-colum matrix where each row represents the start and the end of an edge. Tips are numbered 1,2,…. and the nodes are numbered after tips, counting from the root. |
| edge.length | a numeric vector giving the lengths of each edge |
| Nnode | The number of internal nodes |
| tip.lable | a vector of tip names in the same order as the tips in edge |
| tip.state | a vector of ‘0’ and ‘1’, giving the trait state of each tips, in the order of tip.label |

**References**

Bromham L., Hua X. & Cardillo M. (submitted) Tippy distribution of traits on phylogenies as a measure of macroevolutionary self-destruction. *Systematic Biology,* USYB-2014–232.

**Examples**

phy <- treesim(pars=c(0.1,0.1,0.05,0.05,0.1,0.1), N0=50, N1=50, sampling.f=c(1,1), max.t=Inf)

|  |  |
| --- | --- |
| treestat | *Significance test of user defined phylogenetic metric* |

**Description**

This function conducts the significance test of a phylogenetic metric defined by users.

**Usage**

treestat(phy, stlist = NULL, state = NULL, func, par = NULL,

traitevol = NULL, a = NULL, alternative = "two.sided", simplify = T)

**Arguments**

|  |  |
| --- | --- |
| phy | an object of class 'phylo'. |
| stlist | a vector of tip labels with trait state '1'. |
| state | a vector of '0' and '1' for trait state of each tip, in the same order as the tip labels. |
| func | a name of the function that calculates the metric. The first two inputs of the function need to be the state vector and the phylo class. |
| par | values of additional parameters used in the above function. |
| traitevol | the null model of trait evolution used in the significance test of the metric. traitevol = "TBM" is the threshold brownian motion with the same trait prevalence as observed. traitevol = "random" is the random reshuffle of tip states. If users only want to calculate the value of the metric or if the null model is defined in the 'func' already, traitevol = NULL. |
| a | number of traits simulated under the null model of trait evolution. |
| alternative | the alternative hypothesis for the significant test of the metric. alternative = "greater", if users want to test if the observed trait has significantly larger metric value than expected by the null model of trait evolution. alternative = "less", if users want to test if the observed trait has significantly smaller metric value. alternative = "two.sided" (default), if users have no prior knowledge on how the observed trait may differ from the null model. |
| simplify | if TRUE, the output is simplified. |

**Value**

The full output includes the following elements:

|  |  |
| --- | --- |
| p.value | result of the significance test |
| obsstat | the metric value of the input phylogeny |
| simstat | a list of metric value of each simulated null trait |
| simstate | a matrix, each row is the state vector of a simulated null trait |

The simplified output gives the p.value or the obsstat when traitevol=NULL

**Examples**

#Here we conduct significance test on a metric defined by 'noto' function for a simulated phylogeny. The null hypothesis is that the trait is randomly distributed across tips. We don't know if the trait will have larger or smaller metric value than expected, so we apply a "two.sided" test. 1000 traits are simulated to generate the null distribution of the metric.

#The state vector is included in the phylo class as the output of treesim, so we don't need to input state here.

phy <- treesim(pars=c(0.1,0.1,0.05,0.05,0.1,0.1), N0=50, N1=50, sampling.f=c(1,1), max.t=Inf)

treestat(phy, func=tars, traitevol="random", a=1000, alternative="two.sided", simplify=T)

#Becasue trait simulation takes time, users can conduct significance test on the same simulated traits for multiple metrics with different options of alternative hypotheses

treestat(phy, func=list(tars,noto), traitevol="random", a=1000, alternative= ("two.sided","greater"), simplify=T)

|  |  |
| --- | --- |
| tars | *calculate the tip age rank sum (TARS) metric* |

**Description**

This function calculates the TARS metric. To conduct significance test on TARS, use 'tars' as input of 'func' in 'treestat'

**Usage**

tars(state, phy)

**Details**

If users choose random distribution as the null model for trait evolution, significance test is the Wilcoxon rank-sum test. The test can be done without simulation by setting traitevol=NULL in ‘treestat’. The output of ‘treestat’ is then the rank sum of tip lengths with state 1, upon which users can conduct the Wilcoxon rank-sum test analytically.

**References**

Bromham L., Hua X. & Cardillo M. (submitted) Tippy distribution of traits on phylogenies as a measure of macroevolutionary self-destruction. *Systematic Biology,* USYB-2014–232.

|  |  |
| --- | --- |
| noto | *calculate the number of tips per origin (NoTO) metric* |

**Description**

This function calculates the NoTO metric. To conduct significance test on NoTO, use 'noto' as input of 'func' in 'treestat'

**Usage**

noto(state, phy)

**References**

Agnarsson I., Avilés L., Coddington J.A. & Maddison W.P. (2006) Sociality in Theridiid spiders: repeated origins of an evolutionary dead end. *Evolution* 60:2342-2351

Bromham L., Hua X. & Cardillo M. (submitted) Tippy distribution of traits on phylogenies as a measure of macroevolutionary self-destruction. *Systematic Biology,* USYB-2014–232.

|  |  |
| --- | --- |
| sscd | *calculate the sum of sister clade differences (SSCD) metric* |

**Description**

This function calculates the SSCD metric. To conduct significance test on SSCD, use 'sscd' as input of 'func' in 'treestat'

**Usage**

sscd(state, phy)

**References**

Bromham L., Hua X. & Cardillo M. (submitted) Tippy distribution of traits on phylogenies as a measure of macroevolutionary self-destruction. *Systematic Biology,* USYB-2014–232.

Fritz S.A. & Purvis A. (2010) Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24:1042-1051.

|  |  |
| --- | --- |
| fpd | *calculate the Fritz & Purvis D (FPD) metric* |

**Description**

This function calculates the FPD metric. To conduct significance test on FPD, use 'fpd' as input of 'func' in 'treestat'

**Usage**

fpd(state, phy)

**Details**

The Fritz & Purvis D statistic (FPD) calculates the difference between observed SSCD and expected SSCD under Brownian motion, scaled by the difference between SSCD under random distributions of the trait across the tips of the phylogeny and SSCD under Brownian motion. Thus, the absolute value of FPD equals one if the observed trait is randomly distributed across tips, and equals zero if the trait evolves under Brownian motion. A meaningful significance test on FPD could be FPD>=0, or FPD>=0.5, as suggested by Bromham et al. (submitted).

**References**

Bromham L., Hua X. & Cardillo M. (submitted) Tippy distribution of traits on phylogenies as a measure of macroevolutionary self-destruction. *Systematic Biology,* USYB-2014–232.

Fritz S.A. & Purvis A. (2010) Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24:1042-1051.