

Package ‘PHYLOGR’

January 2, 2012

Version 1.0.7

Date 2010/08/09

Title Functions for phylogenetically based statistical analyses

Author Ramon Diaz-Uriarte <rdiaz@cnio.es> and Theodore Garland, Jr
<tgarland@facstaff.wisc.edu>, <tgarland@citrus.ucr.edu>

Maintainer Ramon Diaz-Uriarte <rdiaz02@gmail.com>

Depends R(>= 1.8.1)

Description Manipulation and analysis of phylogenetically simulated
data sets and phylogenetically based analyses using GLS.

License GPL (>= 2)

URL <http://ligarto.org/rdiaz>

Repository CRAN

Date/Publication 2010-08-09 14:21:09

R topics documented:

cancor.phylog	2
cor.origin	4
GarlandJanis.IC	5
GarlandJanis.Original	6
GarlandJanis.varcov	7
Helper.functions	8
Lacertid.IC	9
Lacertid.Original	10
Lacertid.varcov	11
lm.phylog	12
matrix.D	14
phylog.gls.fit	15
plot.phylog.cancor	17

plot.phylog.lm	18
plot.phylog.prcomp	19
prcomp.phylog	20
print.summary.phylog	21
read.inp.data	22
read.pdi.data	23
read.phylip.data	25
read.phylog.matrix	26
read.sim.data	27
SimulExample	30
summary.phylog.cancor	31
summary.phylog.lm	33
summary.phylog.prcomp	34

Index	37
--------------	-----------

cancor.phylog	<i>Canonical Correlation Analysis from Simulated Data Sets</i>
---------------	--

Description

Performs a canonical correlation analysis on two sets of simulated data and returns the canonical correlations.

Usage

```
cancor.phylog(data1, data2, max.num=0, exclude.tips=NULL, lapply.size=100,
              xcenter=TRUE, ycenter=TRUE)
```

Arguments

data1	the columns from a data set returned from read.sim.data that you want to use as the first set in the canonical correlation analysis. The first column MUST be sim.counter and the second Tips.
data2	the columns from a data set returned from read.sim.data that you want to use as the second set in the canonical correlation analysis. The first column MUST be sim.counter and the second Tips.
max.num	if different from 0, the maximum number of simulations to analyze.
exclude.tips	an optional vector giving the names of tips to exclude from the analyses.
lapply.size	a tuning parameter that can affect the speed of calculations; see Details in phylog.lm.
xcenter	should the x columns be centered? defaults to yes; see cancor.
ycenter	should the y columns be centered? defaults to yes; see cancor.

Value

A list (of class phylog.cancor) with components

call the call to the function

CanonicalCorrelations

the canonical correlations. The one with sim.counter=0 corresponds to the original ("real") data.

WARNING

It is necessary to be careful with the null hypothesis you are testing and how the null data set—the simulations—are generated. For instance, suppose you want to examine the canonical correlations between sets x and y ; you will probably want to generate x and y each with the observed correlations within each set so that the correlations within each set are maintained (but with no correlations among sets). You probably do not want to generate each of the x 's as if they were independent of each other x , and ditto for y , since that will destroy the correlations within each set; see some discussion in Manly, 1997.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Krzanowski, W. J. (1990) *Principles of multivariate analysis* Oxford University Press.

Manly, B. F. J. (1997) *Randomization, bootstrapping, and Monte Carlo methods in biology*, 2nd ed. Chapman & Hall.

Morrison, D. F. (1990) *Multivariate statistical methods*, 3rd ed. McGraw-Hill.

See Also

[read.sim.data](#), [summary.phylog.cancor](#)

Examples

```
data(SimulExample)
ex1.cancor <- cancor.phylog(SimulExample[,c(1,2,3,4,5)], SimulExample[,c(1,2,6,7,8)])
ex1.cancor
summary(ex1.cancor)
plot(ex1.cancor)
```

`cor.origin`*Correlation Through the Origin*

Description

Return the correlation through the origin of two vectors. Generally used for independent contrasts

Usage

```
cor.origin(x, y)
```

Arguments

<code>x</code>	A vector
<code>y</code>	A vector (of same size as <code>x</code>)

Value

The correlation of `x` and `y`, from a model without intercept (i.e., forcing the line through the origin).

Note

This is a very simple function, provided for convenience. You can obtain the p-value, if you wish, with the usual formula for the t-statistic: $2 * (1 - pt(\sqrt{df} * \rho / \sqrt{1 - \rho^2}, df))$ where ρ is the correlation through the origin and df are the appropriate degrees of freedom—generally $N-1$.

Author(s)

R. Diaz-Uriarte and T. Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Examples

```
x <- rnorm(100)
y <- rnorm(100)
rho <- cor.origin(x,y)
rho # the correlation
2 * (1 - pt(sqrt(99) * rho / sqrt(1 - rho^2), 99)) # the p-value
```

GarlandJanis.IC *Independent contrasts for Garland & Janis data set*

Description

Independent contrast for Garland & Janis (1993) data set; they are based on the data in the file `GarlandJanis.Original` using the phylogeny in Garland et al. (1993) — see `GarlandJanis.varcov`

Format

The data frame contains seven columns. The first four columns are the (standardized) independent contrasts for the respective variables —see detailed explanation in `GarlandJanis.Original`. The rest of the columns are:

branch.lengths Standard Deviation of Contrast = square root of sum of corrected branch lengths.

names.contr the names of the contrasts, as the names of the two nodes that form the contrast

clade.contr A factor with levels Carnivore, Herbivore, root, that indicates whether the contrast is a contrast within the carnivore lineage, within the ungulates, or the root contrast —between the carnivore and the ungulate clades.

Source

Garland, T. Jr., and Janis, C. M. (1993). Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? *J. zoology, London*, **229**, 133–151.

Garland, T. Jr., Dickerman, A. W., Janis, C. M., Jones, J. A. (1993) Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, **42**, 265–292.

See Also

[GarlandJanis.Original](#), [GarlandJanis.varcov](#)

Examples

```
# Multiple regression with independent contrasts
# excluding the polar bear - grizzly bear contrast
data(GarlandJanis.IC)
lm(running.speed ~ body.mass + hind.l.length - 1,
  subset=names.contr!="Tm-Ur", data= GarlandJanis.IC)

## Not run:
# This data set can be obtained from the original files as:

garland.janis.ic <- cbind(read.table("49ms.fic")[,c(3,4)],
  read.table("49hmt.fic")[,c(3,4)])
```

```

branch.lengths <- read.table("49ms.fic")[,5]
garland.janis.ic <- garland.janis.ic/branch.lengths
names(garland.janis.ic) <- c("body.mass", "running.speed",
                           "hind.l.length", "mtf.ratio")
garland.janis.ic[garland.janis.ic$body.mass<0,] <-
  -1 * garland.janis.ic[ garland.janis.ic$body.mass<0, ]
garland.janis.ic$branch.lengths <- branch.lengths
garland.janis.ic$names.contr <-
  as.factor(read.table("49ms.fic")[,1])
garland.janis.ic$clade.contr <-
  as.factor( c("root",rep("Carnivore",18), rep("Herbivore",29)))

## End(Not run)

```

GarlandJanis.Original *Garland & Janis's 1993 data on mammalian running speed and limb length*

Description

This data set was used by Garland & Janis in their analysis of metatarsal/femur ration and running speed in cursorial mammals. The data refer to several ecomorphological characteristics for a set of 49 mammals (18 carnivores and 29 ungulates).

Format

This data frame contains the following columns:

Tips the code for each species

body.mass log 10 of body mass in kilograms

running.speed log 10 running or sprint speed in km/h

hind.l.length log 10 hind limb length —sum of femur, tibia, and metatarsal lengths—in cm

mtf.ratio metatarsal/femur ratio

clade a factor with levels Carnivore or Herbivore

Source

Garland, T. Jr., and Janis, C. M. (1993). Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? *J. Zoology, London*, **229**, 133–151.

See Also

[GarlandJanis.IC](#), [GarlandJanis.varcov](#)

Examples

```
# an example of a GLS fit
data(GarlandJanis.Original)
data(GarlandJanis.varcov)
attach(GarlandJanis.Original)
fit.gls.GJ <- phylog.gls.fit(cbind(body.mass,hind.l.length),
running.speed, GarlandJanis.varcov)
summary(fit.gls.GJ) # summary of the gls model; same as with IC
detach(GarlandJanis.Original)

## Not run:
# This data set can be prepared from the original pdi files
# (in directory Examples) as:
GarlandJanis.Orig <- read.pdi.data(c("49ms.pdi", "49hmt.pdi"),
variable.names = c("body.mass", "running.speed",
"hind.l.length", "mtf.ratio"))
GarlandJanis.Orig$clade <- as.factor(c(rep("Carnivore",19),
rep("Herbivore",30)))

## End(Not run)
```

GarlandJanis.varcov *Phylogenetic variance-covariance matrix for Garland & Janis (1993).*

Description

Phylogenetic variance-covariance matrix for the species in Garland & Janis (1993). Note that the phylogeny is not exactly the same as in Garland & Janis (1993), but actually corresponds to the more recent phylogeny in Garland et al. (1993).

Format

A matrix with the phylogenetic distance between species; every entry d_{ij} is the sum of branch segment lengths that species i and j share in common.

Source

Garland, T. Jr., and Janis, C. M. (1993). Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? *J. Zoology, London*, **229**, 133–151.

Garland, T. Jr., Dickerman, A. W., Janis, C. M., Jones, J. A. (1993) Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, **42**, 265–292.

See Also

[GarlandJanis.Original](#), [GarlandJanis.IC](#)

Examples

```
# an example of a GLS fit
data(GarlandJanis.Original)
data(GarlandJanis.varcov)
attach(GarlandJanis.Original)
fit.gls.GJ <- phylog.gls.fit(cbind(body.mass,hind.l.length),
  running.speed, GarlandJanis.varcov)
summary(fit.gls.GJ) # summary of the gls model; same as with IC
detach(GarlandJanis.Original)
```

```
## Not run:
# This data set can be obtained from the original dsc file as:
```

```
GarlandJanis.varcov <- read.phylog.matrix("49ms.dsc")
```

```
## End(Not run)
```

Helper.functions

Helper.functions

Description

Several helper functions used by the PHYLOGR main functions (i.e., these functions are called from other functions). These are all one to three lines functions, which are used in lieu of calls to read.table, scan, etc. They are of no immediate use for the end user, but might be helpful for further programming.

Value

Depends on the helper function; here is a summary:

```
num.sim.tips    number of tips and number of simulations of a simulated data set
number.of.simulations
                 the number of simulations of a simulated data set
number.of.tips.inp
                 number of tips in inp data file
```

number.of.tips.pdi
 ditto for pdi
 number.of.tips.sim
 ditto for sim
 scan.inp.file the two columns with data from inp file
 scan.pdi.file ditto for pdi
 scan.simulation.file
 ditto for sim file
 tips.names.inp the names of tips from an inp file
 tips.names.pdi ditto for pdi

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Lacertid.IC	<i>Independent contrasts for Bauwens and Diaz-Uriarte (1997) lacertid data set</i>
-------------	--

Description

Independent contrast for Bauwens and Diaz-Uriarte (1997) data set; they are based on the data in the file Lacertid.Original using the phylogeny in Buwens and Diaz-Uriarte (1997), Tree A — see [Lacertid.varcov](#)

Format

The data frame contains eight columns. The first seven columns are the (standardized) independent contrasts for the respective variables —see detailed explanation in [Lacertid.Original](#). The final column is

contr the names of the contrasts, as the names of the two nodes that form the contrast

Source

Bauwens, D., and Diaz-Uriarte, R. (1997) Covariation of life-history traits in lacertid lizards: a comparative study. *The American Naturalist*, **149**, 91-11

See Also

[Lacertid.varcov](#), [Lacertid.Original](#)

Examples

```

# Obtaining correlations through the origin;
# compare with Table 3 in Bauwens and Diaz-Uriarte (1997).

data(Lacertid.IC)
cor.lacert <- matrix(nrow=7,ncol=7)
for (i in 1:7) for (j in 1:7)
cor.lacert[i,j] <- cor.origin(Lacertid.IC[[i]],Lacertid.IC[[j]])
cor.lacert

## Not run:
# This data frame can be obtained from the fic data files as:

LacertidIC <- cbind(read.table("ifsmi.fic")[,c(3,4)],
                   read.table("ihshw.fic")[,c(3,4)],
                   read.table("iclag.fic")[,c(3,4)],
                   read.table("icfxx.fic")[,3])
stand <- read.table("ifsmi.fic")[,5]
LacertidIC <- LacertidIC/stand
LacertidIC$contr <- read.table("ifsmi.fic")[,1]
names(LacertidIC) <- c("svl", "svl.matur", "hatsvl", "hatweight",
                    "clutch.size", "age.mat", "cl.freq", "contr")

## End(Not run)

```

Lacertid.Original

Bauwens and Diaz-Uriarte (1997) lacertid data

Description

This is part of the data set used by Bauwens and Diaz-Uriarte (1997) in their analysis of lacertid life histories. The data include several life history traits of 18 lacertid species.

Format

This data frame contains the following columns:

Tips the code for each species

svl log 10 of mean adult female Snout-to-Vent length in mm

svl.matur log 10 of SVL when sexual maturity (females) is reached

hatsvl log 10 of hatchling svl in mm

hatweight log10 of hatchling mass in grams

clutch.size log10 of clutch size
age.mat log10 of age at maturity in months
cl.freq log10 of clutch frequency —number of clutches per year

Source

Bauwens, D., and Diaz-Uriarte, R. (1997) Covariation of life-history traits in lacertid lizards: a comparative study. *The American Naturalist*, **149**, 91-11

See Also

[Lacertid.varcov](#), [Lacertid.IC](#)

Examples

```
# a GLS fit
data(Lacertid.varcov)
data(Lacertid.Original)
ex.gls.phylog <-
phylog.gls.fit(Lacertid.Original$svl,Lacertid.Original$clutch.size,Lacertid.varcov)
ex.gls.phylog

## Not run:
# This data set can also be obtained from the pdi files
# (see example in GarlandJanis.Original), or as:

LacertidSim <- read.sim.data(c("ifsmi.sim","ihshw.sim","iclag.sim","icfxx.sim"),
  pdi.files=c("ifsmi.pdi","ihshw.pdi","iclag.pdi", "icfxx.pdi"),
  variable.names = c("svl","svl.matur","hatsvl","hatweight",
    "clutch.size", "age.mat", "cl.freq", "xx"))

LacertidSim <- LacertidSim[,-10]
LacertidOriginal <- LacertidSim[LacertidSim$sim.counter==0,-1]

## End(Not run)
```

Lacertid.varcov	<i>Variance-covariance matrix for lacertids from Bauwens and Diaz-Uriarte (1997)</i>
-----------------	--

Description

Phylogenetic variance-covariance matrix for 18 species of lacertids. It is based on Tree A of Bauwens and Diaz-Uriarte (1997).

Format

The `Lacertid.varcov` data frame has 18 rows and 18 columns, corresponding to each one of the 18 lacertidspecies; the matrix is the phylogenetic variance-covariance matrix between all 18 species; thus, each entry d_{ij} is the the sum of branch segment lengths that species i and j share in common.

Source

Bauwens, D., and Diaz-Uriarte, R. (1997) Covariation of life-history traits in lacertid lizards: a comparative study. *The American Naturalist*, **149**, 91-111

See Also

[SimulExample, Lacertid.Original](#)

Examples

```
# a GLS fit
data(Lacertid.varcov)
data(Lacertid.Original)
ex.gls.phylog <-
phylog.gls.fit(Lacertid.Original$svl,Lacertid.Original$clutch.size,Lacertid.varcov)
ex.gls.phylog

## Not run:
# This data can be obtained from the original dsc file as:
Lacertid.varcov <- read.phylog.matrix("ifsmi.dsc")

## End(Not run)
```

lm.phylog

Linear Models from Simulated Data Sets

Description

Fit a linear model to the data from a `read.sim.data`.

Usage

```
lm.phylog(formula, data, max.num=0, weights=NULL, exclude.tips=NULL,
           lapply.size=100)
```

Arguments

formula	a formula with the same syntax as for any other linear model in R (see help for <code>lm: ?lm</code>).
data	a data frame —with a particular structure— with the observations; it will often be the name of a data frame created using <code>read.sim.data</code> .
max.num	if different from 0, maximum number of simulations to analyze
weights	an optional vector of weights to perform weighted least squares; can be a column from the data frame or a vector in the parent environment (of same length as any column of data before the data is reduced, if appropriate, with <code>exclude.tips</code> and <code>max.num</code> arguments).
exclude.tips	an optional vector giving the names of tips to exclude from the analyses.
lapply.size	a tuning parameter that can affect the speed of calculations; see Details

Details

This function uses a loop over `lapply` calls (I got the idea from Venables and Ripley (2000), "S programming"). By changing the `lapply.size` you can change the size of the block over which `lapply` is used. Changes can make a difference in speed; for instance, in my machine from about 1 sec per simulation for a set with 49 species to less than 0.5 sec. The default value worked well in my machine, but your mileage will vary.

Value

A list of class `phylog.lm` with components

call	the function call.
Fits	a data frame with the fitted coefficients for that model; the coefficients for the non-simulated ("real") data correspond to <code>sim.counter=0</code> .
MarginalTests	a data frame with the F-tests. The first column is the overall F-test for the model, the rest are the marginal F-tests, respecting the marginality principle or hierarchy of terms included. The coefficients for the non-simulated ("real") data correspond to <code>sim.counter=0</code> .

Note

The marginal F-tests returned are obtained from `drop1`, and thus respect the marginality principle. For instance, if your model is $y \sim x_1 + x_2 * x_3$ you will see an F for x_1 and an F for $x_2 : x_3$ but no F's for x_2 or x_3 . Discussion can be found, for example, in Venables & Ripley, (1999), ch. 6; see also Searle, (1987), ch. 6, for the ANCOVA case.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Searle, S. R. (1987) *Linear models for unbalanced data*. Wiley

Venables, W. N. and Ripley, B. D. (1999) *Modern applied statistics with S-Plus, 3rd ed.* Springer-Verlag.

Venables, W. N. and Ripley, B. D. (2000) *S programming*. Springer-Verlag.

See Also

[summary.phylog.lm](#), [plot.phylog.lm](#), [read.sim.data](#), [drop1](#)

Examples

```
data(SimulExample)
ex1.lm <- lm.phylog(y ~ x1+diet, weights=x2, max.num=20,
                  exclude.tips=c("La", "Ls"), data=SimulExample)
ex1.lm
summary(ex1.lm)
par(mfrow=c(2,2))
plot(ex1.lm)
```

matrix.D

Produce the Matrix D of Garland and Ives (2000) for GLS

Description

Produces the matrix D in Garland and Ives, 2000, p. 361, for further use in GLS procedures. You will rarely need to call this function directly. It is called by `phylog.gls.fit`.

Usage

```
matrix.D(x)
```

Arguments

x a phylogenetic variance-covariance matrix, such as a matrix returned from a call to `read.phylog.matrix`.

Value

A variance-covariance matrix.

Note

The file read is a dsc file generated from PDDIST under option 5 with the additional options : - in matrix form; - with header; - with scaled values.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Garland, T. Jr. and Ives, A. R. (2000) Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155, 346-364.

See Also

[read.phylog.matrix](#), [phylog.gls.fit](#)

Examples

```
#perform GLS using lm function after transforming the variables
data(Lacertid.varcov)
mD <- matrix.D(Lacertid.varcov)
data(Lacertid.Original)
# obtain transformed variables
Z <- mD %**% Lacertid.Original$clutch.size
U <- mD %**% cbind(rep(1,18),Lacertid.Original$sv1) # intercept included
lm1 <- lm(Z ~ U - 1) # eliminate intercept, since already included in U matrix
summary(lm1)
```

phylog.gls.fit

Phylogenetically-Based GLS Model Fitting

Description

Fits a GLS linear model, as in Garland and Ives (2000), using a phylogenetic variance-covariance function

Usage

```
phylog.gls.fit(x, y, cov.matrix, intercept = TRUE, exclude.tips = NULL)
```

Arguments

x	The predictor or "X" variables; they must be numeric variables. If you are using a factor, you must recode it numerically, with the appropriate type of contrasts—see function contrasts and Venables & Ripley (1999) ch. 6
y	The response
cov.matrix	The phylogenetic variance-covariance matrix, which can be obtained from read.phylog.matrix.
intercept	Include an intercept in the model? Defaults to TRUE
exclude.tips	The tips to be excluded from the analyses. Defaults to NULL

Value

a fitted linear model

WARNING

This is one possible implementation of GLS that uses the transformation of the Y and X as explained in Garland and Ives (2000). Ideally, we would directly call gls from the NLME package, passing it the var-cov matrix, but there are some printing problems of the fitted object in the R implementation when we use a fixed correlation structure. The advantage of using gls from NLME is that the function is called using the typical syntax for linear models, and we do not need to worry about making categorical factors into numerical variables. Once the problem with NLME is solved I'll add functions to incorporate GLS into the analysis of data sets.

In the meantime, when using this function, you should be aware that:

- 1) the overall F-test reported is wrong (it is like comparing to a model without an intercept);
- 2) you can apply the usual plot(fitted.model) to see diagnostic plots, or other diagnostic functions such as lm.influence, influence.measures, etc. But most of these will be wrong and meaningless.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

- Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.
- Garland, T. Jr. and Ives, A. R. (2000) Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155, 346-364.
- Venables, W. N. and Ripley, B. D. (1999) *Modern applied statistics with S-Plus, 3rd ed.* Springer-Verlag.

See Also

[read.phylog.matrix](#), [matrix.D](#)

Examples

```
data(Lacertid.varcov)
data(Lacertid.Original)
ex.gls.phylog <-
phylog.gls.fit(Lacertid.Original$svl,Lacertid.Original$clutch.size,Lacertid.varcov)
ex.gls.phylog
```

plot.phylog.cancor *Plot a phylog.cancor object*

Description

Plots histogram of the canonical correlations for simulated data as returned from a phylog.cancor object; with vertical bars indicates the values from the original ("real") data (the one with sim.counter=0), and in parenthesis their 'correlation-wise' p-value (see summary.phylog.cancor).

Usage

```
## S3 method for class 'phylog.cancor'
plot(x, ...)
```

Arguments

x an object of class phylog.cancor returned from a previous call to cancel.phylog.
... other parameters to be passed to through to plotting functions (currently not used).

WARNING

These histograms are in the spirit of the 'correlation-wise' p-values returned from summary.phylog.cancor; see Details for summary.phylog.cancor.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[cancel.phylog](#), [summary.phylog.cancor](#)

Examples

```
data(SimulExample)
ex1.cancor <- cancor.phylog(SimulExample[,c(1,2,3,4,5)],SimulExample[,c(1,2,6,7,8)])
ex1.cancor
summary(ex1.cancor)
par(mfrow=c(1,3))
plot(ex1.cancor)
```

plot.phylog.lm	<i>Plot a phylog.lm object</i>
----------------	--------------------------------

Description

Plots histogram of the the fitted coefficients and F-tests for simulated data as returned from a phylog.lm object; with vertical bars indicates the values from the original ("real") data (the one with sim.counter=0). For F-values, the number in parenthesis indicates the p-value (see summary.phylog.lm).

Usage

```
## S3 method for class 'phylog.lm'
plot(x, ...)
```

Arguments

x	an object of class phylog.lm returned from a previous call to lm.phylog.
...	other parameters to be passed to through to plotting functions (currently not used).

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[lm.phylog](#), [summary.phylog.lm](#)

Examples

```
data(SimulExample)
ex1.lm <- lm.phylog(y ~ x1 + diet, weights=x2, max.num=20, data=SimulExample)
options(digits=5)
plot(ex1.lm)
```

plot.phylog.prcomp *Plot a phylog.prcomp object*

Description

Plots histogram of the eigenvalues for simulated data as returned from a phylog.prcomp object; with vertical bars indicates the values from the original ("real") data (the one with sim.counter=0).

Usage

```
## S3 method for class 'phylog.prcomp'  
plot(x, ...)
```

Arguments

x an object of class phylog.prcomp such as can be obtained from a previous call to prcomp.phylog.
... other parameters to be passed to through to plotting functions (currently not used).

WARNING

These histograms are in the spirit of the 'naive' p-values returned from summary.phylog.prcomp; see Details for summary.phylog.prcomp

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[prcomp.phylog](#), [summary.phylog.prcomp](#)

Examples

```
data(SimulExample)  
ex1.prcomp <- prcomp.phylog(SimulExample[,-11]) # 11th column is a factor  
options(digits=5)  
plot(ex1.prcomp)
```

prcomp.phylog

*Principal Components Analysis from Simulated Data Sets***Description**

Performs a principal component analyses on a set of simulated data and return the eigenvalues.

Usage

```
## S3 method for class 'phylog'
prcomp(x, max.num=0, exclude.tips=NULL, lapply.size=100,
       center=TRUE, scale=TRUE, ...)
```

Arguments

x	the columns from a data set returned from read.sim.data that you want to use in the PCA. The first column MUST be sim.counter and the second Tips.
max.num	if different from 0, the maximum number of simulations to analyze.
exclude.tips	an optional vector giving the names of tips to exclude from the analyses.
lapply.size	a tuning parameter that can affect the speed of calculations; see Details in phylog.lm.
center	should the data be centered before analyses? defaults to yes; see prcomp.
scale	should the data be scaled before the analyses? defaults to yes; see prcomp.
...	Not used.

Value

A list (of class phylog.prcomp) with components

call	the call to the function
Eigenvalues	all the eigenvalues from the PCA. The one with sim.counter=0 corresponds to the original ("real") data.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Krzanowski, W. J. (1990) *Principles of multivariate analysis* Oxford University Press.

Morrison, D. F. (1990) *Multivariate statistical methods, 3rd ed.* McGraw-Hill.

See Also

[read.sim.data](#), [summary.phylog.prcomp](#)

Examples

```
data(SimulExample)
ex1.prcomp <- prcomp.phylog(SimulExample[,-11]) # 11th col. is a factor
options(digits=5)
ex1.prcomp
summary(ex1.prcomp)
plot(ex1.prcomp)
```

print.summary.phylog *Printing summaries of PHYLOGR statistical functions*

Description

These are specific 'methods' for print, that are used with objects of classes summary.phylog.lm, summary.phylog.cancor, and summary.phylog.prcomp, respectively.

Usage

```
## S3 method for class 'summary.phylog.lm'
print(x, ...)
## S3 method for class 'summary.phylog.cancor'
print(x, ...)
## S3 method for class 'summary.phylog.prcomp'
print(x, ...)
```

Arguments

x an object of the appropriate class.
... further parameters to be passed (currently not used).

Details

These functions are called automatically whenever you type 'summary(object.name)' or you type the name of a summary object; these functions simply provide nicer formatted output.

Value

See explanation of output in summary.phylog.lm, summary.phylog.cancor, and summary.phylog.prcomp.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[summary.phylog.lm](#), [summary.phylog.prcomp](#), [summary.phylog.cancor](#)

read.inp.data

Read Inp Data Files

Description

Reads one or more inp data files, such as used by PDTREE, of the PDAP program bundle, and returns an R data frame. Allows to combine several inp files and to change the name of variables.

Usage

```
read.inp.data(input.inp.files, variable.names=NULL)
```

Arguments

`input.inp.files`

the name (with path if necessary), of the inp file(s).

`variable.names` an optional vector with the new names for the variables.

Value

A data frame (with class `pdi.file` and data frame) with first column the names of tips and remaining columns the data columns from the inp file(s).

Author(s)

Diaz-Uriarte, R., and Garland, T., Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[read.pdi.data](#), [read.phylip.data](#), [read.sim.data](#),

Examples

```

# This works under both Unix and Windows.
# First need to find out where the 'Examples' directory is located.
path.to.example <- paste(.path.package(package="PHYLOGR"), "Examples/", sep="/")

# a simple case
p49a <- paste(path.to.example, "49lbr.inp", sep="")
data.49a <- read.inp.data(p49a)
data.49a

# two files and rename columns
p49b <- paste(path.to.example, "49hmt.inp", sep="")
data.49.2 <- read.inp.data(c(p49a, p49b), variable.names=c("y", "x1", "x2", "x3"))
data.49.2

# You could jump directly to the call to the function if you
# are willing to enter the path explicitly.
# For example in some Linux systems the following works
# read.inp.data("/usr/lib/R/library/PHYLOGR/Examples/49lbr.inp")
# In Windows, maybe do:
# read.inp.data("c:\\progra~1\\rw1001\\library\\PHYLOGR\\Examples\\49lbr.inp")

```

read.pdi.data

Read Pdi Data Files

Description

Reads one or more pdi data files, such as used by PDTREE, of the PDAP program bundle, and returns an R data frame. Allows to combine several pdi files and to change the name of variables.

Usage

```
read.pdi.data(input.pdi.files, variable.names=NULL)
```

Arguments

input.pdi.files

the name(s), with path if necessary, of the pdi file(s).

variable.names an optional vector with the new names for the variables.

Value

A data frame (with class `pdi.file` and data frame) with first column the names of tips and remaining columns the data columns from the `pdi` file(s).

Author(s)

Diaz-Uriarte, R., and Garland, T., Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[read.inp.data](#), [read.phylip.data](#), [read.sim.data](#),

Examples

```
# This works under both Unix and Windows.
# First need to find out where the 'Examples' directory is located.
path.to.example <- paste(.path.package(package="PHYLOGR"), "Examples/", sep="/")

# a simple case
p49a <- paste(path.to.example, "49lbr.pdi", sep="")
data.49a <- read.pdi.data(p49a)
data.49a

# two files and rename columns
p49b <- paste(path.to.example, "49hmt.pdi", sep="")
data.49.2 <- read.pdi.data(c(p49a, p49b), variable.names=c("y", "x1", "x2", "x3"))
data.49.2

# You could jump directly to the call to the function if you
# are willing to enter the path explicitly.
# For example in some Linux systems the following works
# read.pdi.data("/usr/lib/R/library/PHYLOGR/Examples/49lbr.pdi")
# In Windows, maybe do:
# read.pdi.data("c:\\progra~1\\rw1001\\library\\PHYLOGR\\Examples\\49lbr.pdi")
```

read.phylip.data	<i>Read Phylip Infile Data Files</i>
------------------	--------------------------------------

Description

Reads one file with tip data, such as a PHYLIP infile, and returns an R data frame. The files have to follow the PHYLIP standard of having a first line with the number of species and traits. There can be multiple traits, as allowed in PHYLIP. Data for a species can extend over several lines, as for PHYLIP's sequential data format for continuous traits. It is assumed that all traits are quantitative.

Usage

```
read.phylip.data(input.phylip.file, variable.names=NULL)
```

Arguments

`input.phylip.file`
the name, with path if necessary, of the infile.

`variable.names` an optional vector with the new names for the variables.

Value

A data frame (with class `phylip.file` and data frame) with first column the names of tips and remaining columns the data columns from the phylip file.

Note

The format of PHYLIP's infiles is not exactly the same as the TIP files produced from PDTREE. First, PHYLIP's infiles can contain an arbitrary number of traits, whereas PDTREE's TIP files only have two. Second, PHYLIP's infiles have a first line with the number of tips and the number of traits, separated by blanks. Third, the first field or column of PHYLIP's infiles must be ten characters wide; if the tips names are shorter, they must be padded with blanks (see PHYLIP's documentation). This limitation does not apply to `read.phylip.data`, but you might want to follow it if you plan to use PHYLIP.

Author(s)

Diaz-Uriarte, R., and Garland, T., Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[read.inp.data](#), [read.pdi.data](#), [read.sim.data](#),

Examples

```
# This works under both Unix and Windows.
# First need to find out where the 'Examples' directory is located.
path.to.example <- paste(.path.package(package="PHYLOGR"), "Examples/", sep="/")

lacertid.data.name <- paste(path.to.example, "LacertidData.PhylipInfile", sep="")
lacertid.data <- read.phylip.data(lacertid.data.name, variable.names=c("sv1", "sv1.matur",
    "hatsv1", "hatweight", "clutch.size",
    "age.mat", "cl.freq", "xx"))

lacertid.data

# You could jump directly to the call to the function if you
# are willing to enter the path explicitly.
# For example in some Linux systems the following works
# read.phylip.data("/usr/lib/R/library/PHYLOGR/Examples/LacertidData.PhylipInfile")
# In Windows, maybe do:
# read.pdi.data("c:\\progra~1\\rw1001\\library\\PHYLOGR\\Examples\\LacertidData.PhylipInfile")
```

read.phylog.matrix *Read a Phylogenetic Covariance Matrix*

Description

Reads a dsc matrix file —returned from the PDDIST program— and converts into an R matrix for subsequent use.

Usage

```
read.phylog.matrix(x)
```

Arguments

x An ASCII data file such as the *.dsc file generated by the PDDIST program

Value

a phylogenetic variance-covariance matrix that can be used in R functions, such as for GLS models.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

Garland, T. Jr. and Ives, A. R. (2000) Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155, 346-364.

See Also

[matrix.D](#), [phylog.gls.fit](#)

Examples

```
# First need to find where the example data sets are
path.to.example <- paste(.path.package(package="PHYLOGR"), "Examples/", sep="/")

example.dsc.file <- paste(path.to.example, "ifsmi.dsc", sep="")
phylog.matrix1 <- read.phylog.matrix(example.dsc.file)

# You could jump directly to the call to the function if you
# are willing to enter the path explicitly.
# For example in some Linux systems the following works
# read.phylog.matrix("/usr/lib/R/library/PHYLOGR/Examples/hb12n.dsc")
# In Windows, maybe do:
# read.phylog.matrix("c:\\progra~1\\rw1001\\library\\PHYLOGR\\Examples\\hb12n.dsc")
```

read.sim.data

Read Sim Files From PDSIMUL

Description

Reads output file(s) from PDSIMUL —simulation of phenotypic evolution along a phylogeny— and converts into an R data frame. Can add original data from inp or pdi file(s).

Usage

```
read.sim.data(sim.files, inp.files = NULL, pdi.files = NULL,
             phylip.file = NULL, variable.names = NULL,
             other.variables = NULL, max.num = 0)
```

Arguments

<code>sim.files</code>	the sim file(s), with path if needed.
<code>inp.files</code>	the inp file(s), with path if needed. Might have already been processed by <code>read.inp.data</code> . If already read, enter name unquoted.
<code>pdi.files</code>	the pdi file(s), with path if needed. Might have already been processed by <code>read.pdi.data</code> . If already read, enter name unquoted.
<code>phylip.file</code>	the phylip.file, with path if needed. Might have already been processed by <code>read.phylip.data</code> . If already read, enter name unquoted.
<code>variable.names</code>	a optional vector of variable names
<code>other.variables</code>	an optional set of other variables that you want to add to the data set. Can be a vector, a matrix, or a data frame. It must be the same length as the number of tips in you sim file(s).
<code>max.num</code>	if different from 0, the number of simulations to process.

Details

You will almost always want to provide `inp` or `pdi` or `phylip` file(s) since this is what will be used for the analyses of simulated data sets. The `sim` and `pdi` (or `inp` or `phylip`) files should match one to one; for example, you might have used `f1.pdi` to obtain `f1.sim` and `f2.pdi` to produce `f2.sim`. Then, the order of files ought to be

```
read.sim.data(c("f1.sim", "f2.sim"), pdi.files=c("f1.pdi", "f2.pdi"))
```

or

```
read.sim.data(c("f2.sim", "f1.sim"), pdi.files=c("f2.pdi", "f1.pdi"))
```

but NOT

```
read.sim.data(c("f1.sim", "f2.sim"), pdi.files=c("f2.pdi", "f1.pdi"))
```

and NOT

```
read.sim.data(c("f2.sim", "f1.sim"), pdi.files=c("f1.pdi", "f2.pdi"))
```

since the last two will yield meaningless results.

Remember that the number of `sim` file(s) and `pdi` (or `inp`) files must match (since that is the only way the number of columns will match).

This does not apply to PHYLIP infiles, since a PHYLIP infile can contain multiple columns.

`Inp` and `Pdi` files can not (yet) be mixed in a single call. If you need to, you should use `read.inp.data` and `read.pdi.data`, and change the class of the output data frame.

If you are entering `inp` files only, you don't need to provide the argument name. If you are using `pdi` files you need to provide the `pdi.files=`.

Value

A data frame (of class `simul.phylog` and `data.frame`) where the first column is called `sim.counter`, for simulation counter (with value 0 for the `pdi`, or `inp`, data set), second column is called `tips`, and the rest are data columns (including, if given, the `other.variables` column).

Author(s)

Ramon Diaz-Uriarte and Thodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[read.pdi.data](#), [read.inp.data](#), [read.phylip.data](#). There are generic functions `plot` and `summary`.

Examples

```
# First we need to find where the Examples directory is.  
# You could enter it directly (see read.pdi.data for an example).  
path.to.example <- paste(.path.package(package="PHYLOGR"), "Examples/", sep="/")
```

```
# simple example  
p49.i <- paste(path.to.example, "49lbr.pdi", sep="")  
p49.s <- paste(path.to.example, "49lbr.sim", sep="")  
data.49.s <- read.sim.data(p49.s, pdi.files=p49.i)  
data.49.s  
  
# several files, added variables, change column names,  
# and limit number of cases  
  
f491s <- paste(path.to.example, "49lbr.sim", sep="")  
f492s <- paste(path.to.example, "49hmt.sim", sep="")  
f493s <- paste(path.to.example, "49ms.sim", sep="")  
  
f491i <- paste(path.to.example, "49lbr.pdi", sep="")  
f492i <- paste(path.to.example, "49hmt.pdi", sep="")  
f493i <- paste(path.to.example, "49ms.pdi", sep="")
```

```

data.hb <-
  read.sim.data(c(f491s, f492s, f493s), pdi.files=c(f491i, f492i, f493i),
    variable.names=c("x1", "x2", "x3", "x4", "x5", "x6"),
    other.variables=data.frame(
      mood=c(rep("good", 15),
        rep("bad", 15),
        rep("terrible", 19)),
      color=c(rep("blue", 20),
        rep("white", 29))),
    max.num=20)

data.hb

```

 SimulExample

A simulated data set

Description

A simulated data set; the phylogeny is based in Bauwens and Diaz-Uriarte (1997), such as is included in the file ifsm.pdi (in the Examples directory). But the data are all completely fictitious and have nothing to do with lacertids (or, for that matter, with any other creatures).

Format

This data frame contains the following columns:

sim.counter the simulation counter

Tips the name of tips; it matches those for the lacertid examples but, again, is unrelated to those

y one numeric variable

x1 another numeric variable

x2 ditto

x3 ditto

x4 ditto

x5 guess what? same thing

x6 again

x7 once more

diet a factor with fictitious levels Carnivore Herbivore Omnivore

Source

Bauwens, D., and Diaz-Uriarte, R. (1997) Covariation of life-history traits in lacertid lizards: a comparative study. *The American Naturalist*, **149**, 91-11

Examples

```
# a canonical correlation example
data(SimulExample)
ex1.cancor <- cancor.phylog(SimulExample[,c(1,2,3,4,5)],SimulExample[,c(1,2,6,7,8)])
ex1.cancor
summary(ex1.cancor)
plot(ex1.cancor)
```

summary.phylog.cancor *Summarize a phylog.cancor object*

Description

A 'method' for objects of class phylog.cancor. Shows the original data, and provides p-values and quantiles of the canonical correlations based on the simulated data. There is a print 'method' for this summary.

Usage

```
## S3 method for class 'phylog.cancor'
summary(object, ...)
```

Arguments

object an object of class phylog.cancor returned from a previous call to cancor.phylog.
... further arguments passed to or from other methods (currently not used).

Details

To test the hypothesis that all population canonical correlations are zero we use the likelihood-ratio statistic from Krzanowski (pp. 447 and ff.); this statistic is computed for the original data set and for each of the simulated data sets, and we obtain the p-value as (number of simulated data sets with LR statistic larger than original ("real") data + 1) / (number of simulated data sets + 1). Note that a test of this same hypothesis using the Union-Intersection approach is equivalent to the test we implement below for the first canonical correlation.

The p-values for the individual canonical correlations are calculated in two different ways. For the 'component-wise' ones the p-value for a particular correlation is (number of simulated data sets with canonical correlation larger than original ("real") data + 1) / (number of simulated data sets + 1). With this approach, you can find that the p-value for, say, the second canonical correlation is smaller than the first, which is not sensible. It only makes sense to examine the second canonical correlation if the first one is "significant", etc. Thus, when considering the significance of the second canonical correlation we should account for the value of the first. In other words, there is only support against the null hypothesis (of no significant second canonical correlation) if both the first and the second canonical correlations from the observed data set are larger than most of the simulated data sets. We can account for what happens with the first canonical correlation by computing the p-value of the second canonical correlation as the number of simulations in which the second simulated canonical correlation is larger than the observed, or the first simulated canonical correlation is larger than the

observed one, or both (so that the only cases that count against the null are those where both the first and second canonical correlations are smaller than the observed ones); these we call 'Multiple' p-values.

Value

A list (of class `summary.phylog.cancor`) with elements

`call` the call to function `cancor.phylog`.
`original.LR.statistic` the likelihood ratio statistic for the test that all canonical correlations are zero
`original.canonicalcorrelations` the canonical correlations corresponding to the original ("real") data set.
`p.value.overall.test` the p-value for the test that all canonical correlations are zero
`p.value.corwise` the correlation-wise p-value —see Details
`p.value.mult` the multiple correlations p-value; see Details
`quant.canonicalcorrelations` the quantiles from the simulated canonical correlations; linear interpolation is used. Note that these quantiles are in the spirit of the "naive p.values".
`num.simul` the number of simulations used in the analyses

WARNING

It is necessary to be careful with the null hypothesis you are testing and how the null data set—the simulations—are generated. For instance, suppose you want to examine the canonical correlations between sets x and y ; you will probably want to generate x and y each with the observed correlations within each set so that the correlations within each set are maintained (but with no correlations among sets). You probably do not want to generate each of the x 's as if they were independent of each other x , and ditto for y , since that will destroy the correlations within each set; see some discussion in Manly, 1997.

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

- Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.
- Krzanowski, W. J. (1990) *Principles of multivariate analysis* Oxford University Press.
- Manly, B. F. J. (1997) *Randomization, bootstrapping, and Monte Carlo methods in biology*, 2nd ed. Chapman & Hall.
- Morrison, D. F. (1990) *Multivariate statistical methods*, 3rd ed. McGraw-Hill.

See Also

[read.sim.data](#), [summary.phylog.cancor](#)

Examples

```
data(SimulExample)
ex1.cancor <- cancor.phylog(SimulExample[,c(1,2,3,4,5,6)],SimulExample[,c(1,2,7,8)])
summary(ex1.cancor)
```

summary.phylog.lm	<i>Summarize a phylog.lm object</i>
-------------------	-------------------------------------

Description

A 'method' for objects of class phylog.lm. Summarizes the results of the fitted linear models for the simulated (and original) data and returns p-values. There is a print 'method' for this summary.

Usage

```
## S3 method for class 'phylog.lm'
summary(object, ...)
```

Arguments

object	an object of class phylog.lm returned from a previous call to lm.phylog.
...	further arguments passed to or from other methods (currently not used).

Details

The p-value is computed as the (number of simulated data sets with F-value larger than original ("real") data + 1) / (number of simulated data sets + 1). The quantiles are obtained with the function quantile, and thus linear interpolation is used.

Value

A list (of class summary.phylog.lm) with components

call	the call to lm.phylog
original.model	the model fitted to the original data.
original.Fvalue	the F-values from the original data.
p.value	the p-values for the marginal F-tests and overall F.
quant.F.value	the quantiles of the distribution of F-values from the simulated data.
num.simul	the number of simulations used in the analyses

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.

See Also

[lm.phylog](#), [plot.phylog.lm](#)

Examples

```
data(SimulExample)
ex1.lm <- lm.phylog(y ~ x1 + diet, weights=x2, max.num=20, data=SimulExample)
summary(ex1.lm)
```

summary.phylog.prcomp *Summarize a phylog.prcomp object*

Description

A 'method' for objects of class phylog.prcomp. Shows the original data, and provides p-values and quantiles of the eigenvalues based on the simulated data. There is a print 'method' for this summary.

Usage

```
## S3 method for class 'phylog.prcomp'
summary(object, ...)
```

Arguments

object	an object of class phylog.prcomp such as one returned from a previous call to prcomp.phylog.
...	further arguments passed to or from other methods (currently not used).

Details

The p-values are calculated in two different ways. The 'component-wise' ones, where the p-value for a particular eigenvalue is (number of simulated data sets with eigenvalue larger than original ("real") data + 1) / (number of simulated data sets + 1). With this approach, you can find that the p-value for, say, the second eigenvalue is smaller than the first, which is not sensible. It only makes sense to examine the second eigenvalue if the first one is "significant", etc. Thus, when considering the significance of the second eigenvalue we should account for the value of the first. In other words, there is only support against the null hypothesis (of no significant second component) if both the first and the second eigenvalue from the observed data set are larger than most of the simulated data

sets. We can account for what happens with the first eigenvalue by computing the p-value of the second eigenvalue as the number of simulations in which the second eigenvalue is larger than the observed, or the first simulated eigenvalue is larger than the observed one, or both (so that the only cases that count against the null are those where both the first and second simulated eigenvalues are smaller than the observed ones). Therefore, with the second set of p-values, the p-values are increasing.

We also provide data for parallel analysis as in Horn (1965; see also Zwick & Velicer 1986 and Lautenschlager 1989) where each eigenvalue is compared to the average eigenvalue (for that factor) obtained from a simulation. These can then be used to construct scree plots showing both the original and the simulated data.

Value

A list (of class `summary.phylog.prcomp`) with elements

`call` the call to function `prcomp.phylog`.
`original.eigenvalues` the eigenvalues corresponding to the original ("real") data set.
`p.value.component` the component-wise p-value —see Details
`p.value.multiple` the 'multiple-eigenvalue' p-value; see Details
`quant.eigenvalue` the quantiles from the simulated eigenvalues; linear interpolation is used. Note that these quantiles are in the spirit of the "component-wise p.values".
`num.simul` the number of simulations used in the analyses

Author(s)

Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

- Diaz-Uriarte, R., and Garland, T., Jr., in prep. PHYLOGR: an R package for the analysis of comparative data via Monte Carlo simulations and generalized least squares approaches.
- Horn, J. L. (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Krzanowski, W. J. (1990) *Principles of multivariate analysis* Oxford University Press.
- Lautenschlager, G. J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24, 365-395.
- Morrison, D. F. (1990) *Multivariate statistical methods, 3rd ed.* McGraw-Hill.
- Zwick, W. R., and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

See Also

[read.sim.data](#), [prcomp.phylog](#)

Examples

```
data(SimulExample)
ex1.prcomp <- prcomp.phylog(SimulExample[,-11]) #the 11th column is a factor
summary(ex1.prcomp)
```

Index

- *Topic **array**
 - matrix.D, [14](#)
- *Topic **datasets**
 - GarlandJanis.IC, [5](#)
 - GarlandJanis.Original, [6](#)
 - GarlandJanis.varcov, [7](#)
 - Lacertid.IC, [9](#)
 - Lacertid.Original, [10](#)
 - Lacertid.varcov, [11](#)
 - SimulExample, [30](#)
- *Topic **file**
 - read.inp.data, [22](#)
 - read.pdi.data, [23](#)
 - read.phylip.data, [25](#)
 - read.sim.data, [27](#)
- *Topic **manip**
 - read.inp.data, [22](#)
 - read.pdi.data, [23](#)
 - read.phylip.data, [25](#)
 - read.phylog.matrix, [26](#)
 - read.sim.data, [27](#)
- *Topic **methods**
 - plot.phylog.cancor, [17](#)
 - plot.phylog.lm, [18](#)
 - plot.phylog.prcomp, [19](#)
 - print.summary.phylog, [21](#)
 - summary.phylog.cancor, [31](#)
 - summary.phylog.lm, [33](#)
 - summary.phylog.prcomp, [34](#)
- *Topic **misc**
 - Helper.functions, [8](#)
- *Topic **models**
 - cor.origin, [4](#)
 - lm.phylog, [12](#)
 - phylog.gls.fit, [15](#)
- *Topic **multivariate**
 - cancor.phylog, [2](#)
 - prcomp.phylog, [20](#)
 - summary.phylog.cancor, [31](#)
 - summary.phylog.prcomp, [34](#)
 - cancor.phylog, [2, 17](#)
 - cor.origin, [4](#)
 - drop1, [14](#)
 - GarlandJanis.IC, [5, 6, 8](#)
 - GarlandJanis.Original, [5, 6, 8](#)
 - GarlandJanis.varcov, [5, 6, 7](#)
 - Helper.functions, [8](#)
 - Lacertid.IC, [9, 11](#)
 - Lacertid.Original, [9, 10, 12](#)
 - Lacertid.varcov, [9, 11, 11](#)
 - lm.phylog, [12, 18, 34](#)
 - matrix.D, [14, 16, 27](#)
 - num.sim.tips (Helper.functions), [8](#)
 - number.of.simulations (Helper.functions), [8](#)
 - number.of.tips.inp (Helper.functions), [8](#)
 - number.of.tips.pdi (Helper.functions), [8](#)
 - number.of.tips.sim (Helper.functions), [8](#)
 - phylog.gls.fit, [15, 15, 27](#)
 - plot.phylog.cancor, [17](#)
 - plot.phylog.lm, [14, 18, 34](#)
 - plot.phylog.prcomp, [19](#)
 - prcomp.phylog, [19, 20, 35](#)
 - print.summary.phylog, [21](#)
 - read.inp.data, [22, 24, 25, 29](#)
 - read.pdi.data, [22, 23, 25, 29](#)
 - read.phylip.data, [22, 24, 25, 29](#)
 - read.phylog.matrix, [15, 16, 26](#)
 - read.sim.data, [3, 14, 21, 22, 24, 25, 27, 33, 35](#)
 - scan.inp.file (Helper.functions), [8](#)

scan.pdi.file (Helper.functions), 8
scan.simulation.file
 (Helper.functions), 8
SimulExample, 12, 30
summary.phylog.cancor, 3, 17, 22, 31, 33
summary.phylog.lm, 14, 18, 22, 33
summary.phylog.prcomp, 19, 21, 22, 34

tips.names.inp (Helper.functions), 8
tips.names.pdi (Helper.functions), 8