Package ‘PHYLOGR’

April 14, 2020

Version 1.0.11
Date 2020-04-14
Title Functions for Phylogenetically Based Statistical Analyses
Author Ramon Diaz-Uriarte <rdiaz@cnio.es> and Theodore Garland, Jr
         <theodore.garland@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
LazyData true
NeedsCompilation no
Repository CRAN
Date/Publication 2020-04-14 15:10:07 UTC

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 .......................................................... 12
lm.phylog ................................................................. 13
matrix.D ................................................................. 14
phylog.gls.fit ............................................................ 16
plot.phylog.cancor ..................................................... 17
cancor.phylog

Canonical Correlation Analysis from Simulated Data Sets

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
- 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


See Also
read.sim.data, summary.phylog.cancor

Examples
```r
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)
```
Description

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

Usage

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

Arguments

- `x`: A vector
- `y`: A vector (of same size as `x`)

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: 

\[ 2 \times (1 - pt(\sqrt{df} \times \text{abs}(\rho) / \sqrt{1 - \rho^2}), df) \]

where \( \rho \) is the correlation through the origin and `df` are the appropriate degrees of freedom — generally `N-1` — by using the absolute value of the coefficient and finding \( 2 \times \) the probability of upper tail \((1 - pt)\) this works for both positive and negative correlation coefficients.

Author(s)

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

References


Examples

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


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


See Also

GarlandJanis.Original, GarlandJanis.varcov

Examples

```r
# 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)])
```
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 ratio 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


See Also

GarlandJanis.IC, GarlandJanis.varcov
Examples

```r
## What do the data look like
head(GarlandJanis.Original)
head(GarlandJanis.varcov)

## An example of a GLS fit
fit.gls.GJ <- with(GarlandJanis.Original,
  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

## 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"))
Garland.Janis.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**


See Also

GarlandJanis.Original, GarlandJanis.IC

Examples

## What do the data look like
head(GarlandJanis.Original)
head(GarlandJanis.varcov)

## An example of a GLS fit
fit.gls.GJ <- with(GarlandJanis.Original, 
                 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

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

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

## End(Not run)
Author(s)
Ramon Diaz-Uriarte and Theodore Garland, Jr.

References

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

\textbf{contr}  the names of the contrasts, as the names of the two nodes that form the contrast

Source
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


See Also

[Lacertid.varcov](https://example.com/Lacertid.varcov), [Lacertid.IC](https://example.com/Lacertid.IC)

Examples

```r
# 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
```

```r
## 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

Variance-covariance matrix for lacertids from Bauwens and Diaz-Uriarte (1997)

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 lacertid species; the matrix is the phylogenetic variance-covariance matrix between all 18 species; thus, each entry $d_{ij}$ is the sum of branch segment lengths that species $i$ and $j$ share in common.

Source


See Also

SimulExample, Lacertid.Original

Examples

```r
# 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("ifsni.dsc")

## End(Not run)
```
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 lm: ?lm).
    data              a data frame —with a particular structure— with the observations; it will often be the name of a data frame created using read.sim.data.
    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 exclude.tips and max.num 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 sim.counter=0.
    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 sim.counter=0.
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


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


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$clutch.size
U <- mD %*% cbind(rep(1,18),Lacertid.Original$svl) # 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 correation 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.
plot.phylog.cancor

References

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 cancor.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


See Also
cancor.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

Plot a phylog.lm object

Description

Plots histogram of 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

plot.phylog.prcomp

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


See Also

prcomp.phylog, summary.phylog.prcomp
Examples

```r
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**

```r
## 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


See Also

read.sim.data, summary.phylog.prcomp

Examples

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

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

```r
## 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 automagically 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**


**See Also**

`summary.phylog.lm`, `summary.phylog.prcomp`, `summary.phylog.cancor`

---

**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**

```r
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**

read.pdi.data

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


See Also

- `read.inp.data`
- `read.phylip.data`
- `read.sim.data`

Examples

```r
# 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.
```
read.phylip.data

# 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")

---

**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**

```r
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**

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

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("svl", "svl.matur",
"hatsvl", "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")
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


See Also

`matrix.D`, `phylog.gls.fit`

Examples

```r
# 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**

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

**Arguments**

- `sim.files`: the sim file(s), with path if needed.
- `inp.files`: the inp file(s), with path if needed. Might have already been processed by read.inp.data. If already read, enter name unquoted.
- `pdi.files`: the pdi file(s), with path if needed. Might have already been processed by read.pdi.data. If already read, enter name unquoted.
- `phylip.file`: the phylip.file, with path if needed. Might have already been processed by read.phylip.data. If already read, enter name unquoted.
- `variable.names`: a optional vector of variable names
- `other.variables`: 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 your sim file(s).
- `max.num`: 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

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

or

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

but NOT

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

and NOT

```r
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

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

```r
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="")
```

```r
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)
```

```r
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 Ommnivore
Source

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: 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**


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**

*Summarize a phylog.lm object*

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

```r
## 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.
Summary

**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

```r
## 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


**See Also**

*read.sim.data, prcomp.phylog*

**Examples**

```r
data(SimulExample)
ex1.pcomp <- prcomp.phylog(SimulExample[,-11]) # the 11th column is a factor
summary(ex1.pcomp)
```
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, 12
  SimulExample, 30

*Topic **file**
  read.inp.data, 22
  read.pdi.data, 23
  read.phylip.data, 25
  read.sim.data, 28

*Topic **manip**
  read.inp.data, 22
  read.pdi.data, 23
  read.phylip.data, 25
  read.phylog.matrix, 26
  read.sim.data, 28

*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, 13
  phylog.gls.fit, 16

*Topic **multivariate**
  cancor.phylog, 2
  prcomp.phylog, 20
  summary.phylog.cancor, 31
  summary.phylog.prcomp, 34
  cancor.phylog, 2, 18
  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, 10, 12
  Lacertid.varcov, 9–11, 12
  lm.phylog, 13, 19, 34
  matrix.D, 14, 17, 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, 16, 27
  plot.phylog.cancor, 17
  plot.phylog.lm, 14, 18, 34
  plot.phylog.prcomp, 19
  prcomp.phylog, 19, 20, 36
  print.summary.phylog, 21

read.inp.data, 22, 24, 26, 29
read.pdi.data, 23, 23, 26, 29
read.phylip.data, 23, 24, 25, 29
read.phylog.matrix, 15, 17, 26
read.sim.data, 3, 14, 21, 23, 24, 26, 28, 33, 36
scan.inp.file (Helper.functions), 8

37
scan.pdi.file (Helper.functions), 8
scan.simulation.file
    (Helper.functions), 8
SimulExample, 12, 30
summary.phylog.cancor, 3, 18, 22, 31, 33
summary.phylog.lm, 14, 19, 22, 33
summary.phylog.prcomp, 19, 21, 22, 34
tips.names.inp (Helper.functions), 8
tips.names.pdi (Helper.functions), 8