Package ‘Rphylopars’

October 12, 2022

Version 0.3.9

Title Phylogenetic Comparative Tools for Missing Data and Within-Species Variation

Description Tools for performing phylogenetic comparative methods for datasets with multiple observations per species (intraspecific variation or measurement error) and/or missing data (Goolsby et al. 2017). Performs ancestral state reconstruction and missing data imputation on the estimated evolutionary model, which can be specified as Brownian Motion, Ornstein-Uhlenbeck, Early-Burst, Pagel's lambda, kappa, or delta, or a star phylogeny.

Date 2022-05-08

Depends ape

License GPL (>= 2)

URL https://github.com/ericgoolsby/Rphylopars/wiki

Imports Rcpp, doBy, phylolm, phytools, Matrix, MASS, numDeriv

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.1.1

NeedsCompilation yes

Author Eric W. Goolsby [aut, cre], Jorn Bruggeman [aut], Cecile Ane [aut]

Maintainer Eric W. Goolsby <eric.goolsby.evolution@gmail.com>

Repository CRAN

Date/Publication 2022-05-09 02:20:02 UTC

R topics documented:

Rphylopars-package ........................................... 2
anc.recon .................................................. 3
anova.phylopars.lm ......................................... 4
fast.SSC ..................................................... 5
logLik.phylopars ............................................. 6
logLik.phylopars.lm ......................................... 6
Description

Tools for performing phylogenetic comparative methods for datasets with with multiple observations per species (intraspecific variation or measurement error) and/or missing data (Goolsby et al. 2017). Performs ancestral state reconstruction and missing data imputation on the estimated evolutionary model, which can be specified as Brownian Motion, Ornstein-Uhlenbeck, Early-Burst, Pagel's lambda, kappa, or delta, or a star phylogeny.

Details

Package: Rphylopars
Type: Package
Version: 0.3.9
Date: 2022-05-08
License: GPL (>= 2)

Author(s)

Eric W. Goolsby, Jorn Bruggeman, Cecile Ane
Maintainer: Eric W. Goolsby <eric.goolsby.evolution@gmail.com>

References


Examples

# simulate data
sim_data <- simtraits(ntaxa = 15,ntraits = 4,nreps = 3,nmissing = 10)

# estimate parameters under Brownian motion
# pheno_error = TRUE assumes intraspecific variation
# pheno_correlated = FALSE assumes intraspecific variation is not correlated
# phylo_correlated = TRUE assumed traits are correlated
PPE <- phylopars(trait_data = sim_data$trait_data,tree = sim_data$tree,
      pheno_error = TRUE,phylo_correlated = TRUE,pheno_correlated = TRUE)

PPE

PPE$anc_recon # Ancestral state reconstruction and species mean prediction
PPE$anc_var # Prediction variance

###NOT RUN
# estimate parameters under multivariate OU
# PPE_OU <- phylopars(trait_data = sim_data$trait_data,tree = sim_data$tree,
#  # model="mvOU",pheno_error = TRUE,phylo_correlated = TRUE,
#  # pheno_correlated = TRUE)
#
#
# PPE

anc.recon

Ultra-fast maximum likelihood ancestral state reconstruction

Description

This function performs ancestral state reconstruction using a fast algorithm based on Ho and Ane (2014).

Usage

anc.recon(trait_data, tree, vars = FALSE, CI = FALSE)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trait_data</td>
<td>A vector or matrix of trait values. Names or row names correspond to species names. Data cannot have any missing data or within-species variation (this type of data can be handled by the phylopars function).</td>
</tr>
<tr>
<td>tree</td>
<td>An object of class phylo.</td>
</tr>
</tbody>
</table>
anova.phylopars.lm

vars Whether to return the variances of the restricted maximum likelihood estimates
CI Whether to return 95% confidence intervals of the restricted maximum likelihood estimates

Value

A named vector of maximum likelihood ancestral states (with names corresponding to node names if available or node numbers from the tree rearranged in postorder, as obtained by the command `reorder(tree,"postorder")`). If vars or CI is set to TRUE, a list is returned with these values included.

Author(s)


See Also

fastAnc, ace, pic

Examples

```r
require(ape)
tree <- rtree(10000) # random tree with 10,000 taxa
x <- setNames(rnorm(1e4),tree$tip.label) # random trait data
recon <- anc.recon(trait_data=x,tree=tree)
```

---

anova.phylopars.lm  Phylpars regression ANOVA

Description

Generic S3 method for phylopars

Usage

```r
## S3 method for class 'phylopars.lm'
anova(object, ...)
```

Arguments

- `object` Fitted phylopars.lm object
- `...`
Description

This function uses a fast ancestral state reconstruction algorithm (anc.recon, Goolsby, In review) to calculate the sum of squared changes between ancestral and descendant nodes/tips, as described in Klingenberg and Gidaszewski (2010). Significance is assessed via phylogenetic permutation.

Usage

fast.SSC(trait_data, tree, niter = 1000)

Arguments

- **trait_data**: A vector or matrix of trait values. Names or row names correspond to species names. Data cannot have any missing data or within-species variation.
- **tree**: An object of class phylo.
- **niter**: Number of iterations for hypothesis testing (default=1000).

Value

- **pvalue**: Description of 'comp1'
- **scaled.SSC**: Scaled sum of squared changes. A value less than 1 indicates less phylogenetic signal as measured by SSC than expected under Brownian motion, and a value greater than 1 indicates greater phylogenetic signal as measured by SSC than expected under Brownian motion.
- **SSC**: Total sum of squared changes (SSC)

Author(s)

Eric W. Goolsby

References


Examples

```r
sim_dat <- simtraits(ntaxa = 100, ntraits = 4)
fast.SSC(trait_data = sim_dat$trait_data, tree = sim_dat$tree)
```

### logLik.phylopars

*Extract Log likelihood*

#### Description

Generic S3 method for phylopars

#### Usage

```r
## S3 method for class 'phylopars'
logLik(object, ...)
```

#### Arguments

- `object`: Fitted phylopars object
- `...`: Additional arguments

### logLik.phylopars.lm

*Extract Log likelihood*

#### Description

Generic S3 method for phylopars

#### Usage

```r
## S3 method for class 'phylopars.lm'
logLik(object, ...)
```

#### Arguments

- `object`: Fitted phylopars.lm object
- `...`: Additional arguments
Description

This function estimates parameters for the phylogenetic and phenotypic variance-covariance matrices for datasets with missing observations and multiple within-species observations. This function can also be used to fit alternative evolutionary models, including Ornstein-Uhlenbeck, Early-Burst, star phylogeny, or Pagel’s lambda, kappa, or delta. Reconstructed ancestral states and predicted species means (i.e., for missing data), along with prediction variances, are also provided.

Usage

phylopars(trait_data, tree, model = "BM", pheno_error, phylo_correlated = TRUE, pheno_correlated = TRUE, REML = TRUE, full_alpha = TRUE, phylocov_start, phenocov_start, model_par_start, phylocov_fixed, phenocov_fixed, model_par_fixed, skip_optim = FALSE, skip_EM = FALSE, EM_Fels_limit = 1000, repeat_optim_limit = 1, EM_missing_limit = 50, repeat_optim_tol = 0.01, model_par_evals = 10, max_delta = 10000, EM_verbose = FALSE, optim_verbose = FALSE, npd = FALSE, nested_optim = FALSE, usezscores = TRUE, phenocov_list = list(), ret_args = FALSE, ret_level = 1, get_cov_CIs = FALSE)

Arguments

trait_data A data frame with the first column labeled "species" (with species names matching tips on the phylogeny) and one column per trait. Each row corresponds to a single observation, and multiple observations for species are allowed. Missing data should be represented with NA.

tree An object of class phylo

model Model of evolution. Default is "BM". Alternative evolutionary models include "mvOU" (for the multivariate Ornstein-Uhlenbeck), or univariate tree transformations: "OU" "lambda", "kappa", "delta", "EB", "star".

pheno_error If TRUE (default, unless <=1 observation per species is provided), parameters are estimated assuming within-species variation.

phylo_correlated If TRUE (default), parameters are estimated assuming traits are correlated.

pheno_correlated If TRUE (default), parameters are estimated assuming within-species observations traits are correlated.

REML If TRUE (default), the algorithm will return REML estimates. If FALSE, maximum likelihood estimates will be returned.

full_alpha Only applicable for the multivariate OU (model="mvOU"). If TRUE (default), a fully parametrized alpha matrix is fit. If FALSE, a diagonal alpha matrix is fit.

phylocov_start Optional starting value for phylogenetic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
phenocov_start  Optional starting value for phenotypic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
model_par_start  Optional starting parameters for the evolutionary model. For model="mvOU", must be of dimension n_traits by n_traits. Otherwise, must be a single value.
phylocov_fixed  Optional fixed value for phylogenetic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
phenocov_fixed  Optional starting value for phenotypic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
model_par_fixed  Optional fixed parameter for the evolutionary model. For model="mvOU", must be of dimension n_traits by n_traits. Otherwise, must be a single value.
skip_optim  Whether to skip BFGS optimization (not recommended unless all parameters are fixed).
skip_EM  Whether to skip Expectation-Maximization prior to generating starting parameters for BFGS optimization (not recommended unless providing fixed parameters).
EM_Fels_limit  Whether to skip Expectation-Maximization prior to generating starting parameters for BFGS optimization (not recommended unless providing fixed parameters).
repeat_optim_limit  The number of times to repeat numerical optimization (default is 1).
EM_missing_limit  Maximum number of iterations for EM.
repeat_optim_tol  Maximum tolerance for repeated numerical optimization (only relevant if repeat_optim_limit>1).
model_par_evals  Number of times to evaluate univariate tree transformation models along the range of possible parameter values. Used to generate informed starting values for alternative evolutionary models if nested_optim=TRUE.
max_delta  Maximum allowed difference between the log-likelihood for EM-generated starting parameters and new parameters tried under numerical optimization. Extremely large deltas are likely to be numerical artifacts. Prevents artificial convergence.
EM_verbose  Whether to print the log-likelihood during Expectation-Maximization.
optim_verbose  Whether to print log-likelihoods during numerical optimization.
npd  Whether to find the nearest positive-definite matrix for all covariance matrices during numerical optimization (slow – only set to TRUE if converging to singular matrices).
nested_optim  Only relevant if fitting a univariate alternative evolutionary model. Tries multiple tree transformation parameter values along the range of possible values to make informed starting parameters. Slower than the default (nested_optim=FALSE), in which all parameters are estimated simultaneously.
phylopars

usezscores  Whether or not to use centered and standardized data during numerical optimization (recommended).

phenocov_list  An optional named list of species-specific within-species covariance matrices to be held fixed, as in Ives et al (2007). This option forces pheno_error and pheno_correlated to be FALSE, and uses mean species values instead of raw data. Raw variance should be divided by the number of observations per species (i.e., squared standard errors). See Ives et al (2007) for more details.

ret_args  For internal use only.

ret_level  For internal use only.

get_cov_CIs  Whether to return 95-percent confidence intervals of covariance parameters (default=FALSE).

Value

An object of class phylopars. For models with phenotypic (within-species) covariance, the estimated percentage of variance explained by the phylogeny is provided as 100*(1 - phenotypic_variance/raw_variance), where raw_variance is the variance of all observations for a given trait across species (\( \text{var(PPE$trait_data[,2:ncol(PPE$trait_data)]}, na.rm=TRUE) \)).

logLik  The log-likelihood of the model

pars  A list composed of phylogenetic trait covariance and phenotypic (within-species) trait covariance, if estimated

model  The model of evolution (e.g., BM, OU, lambda, etc.), and any additional evolutionary model parameters estimated. For OU models, stationary covariance is calculated from both phylogenetic covariance (Sigma) and alpha (see Supplement 1 of Clavel et al. 2015).

mu  The estimate ancestral state at the root of the tree.

npars  The total number of parameters estimated by optimization (used for AIC and BIC).

anc_recon  Reconstructed ancestral states and species means. Row names correspond to species names (for the first \( 1:\text{nspecies} \) rows), and the remaining row names correspond to node numbers on a tree with edges in postorder: \( \text{reorder(tree,"postorder"}) \). Or, if node labels were included on the original tree, row names correspond to node labels.

anc_var  Variance of reconstructed ancestral estimates and imputed species means.

anc_cov  Covariance of estimates among variables.

tree  The phylogenetic tree supplied to phylopars

trait_data  The trait data supplied to phylopars

REML  TRUE if REML, FALSE if ML.

Author(s)

Eric W. Goolsby <eric.goolsby.evolution@gmail.com>, Cecile Ane, Jorn Bruggeman
# simulate data
sim_data <- simtraits(ntaxa = 15,ntraits = 4,nreps = 3,nmissing = 10)

# estimate parameters under Brownian motion
# pheno_error = TRUE assumes intraspecific variation
# pheno_correlated = FALSE assumes intraspecific variation is not correlated
# phylo_correlated = TRUE assumed traits are correlated

PPE <- phylopars(trait_data = sim_data$trait_data,tree = sim_data$tree,
          pheno_error = TRUE,phylo_correlated = TRUE,pheno_correlated = TRUE)

PPE

PPE$anc_recon # Ancestral state reconstruction and species mean prediction
PPE$anc_var # Prediction variance

###NOT RUN
# estimate parameters under multivariate OU
# PPE_OU <- phylopars(trait_data = sim_data$trait_data,tree = sim_data$tree,
#          model="mvOU",pheno_error = TRUE,phylo_correlated = TRUE,
#          pheno_correlated = TRUE)
# # PPE

phylopars.lm

## References


## Description

Performs phylogenetic regression.

## Usage

phylopars.lm(formula, trait_data, tree, model = "BM", pheno_error, phylo_correlated = TRUE, pheno_correlated = TRUE, REML = TRUE, full_alpha = TRUE, phylocov_start, phenocov_start, model_par_start, phylocov_fixed, phenocov_fixed, model_par_fixed, skip_optim = FALSE,
skip_EM = FALSE, EM_Fels_limit = 1000, repeat_optim_limit = 1,
EM_missing_limit = 50, repeat_optim_tol = 0.01, model_par_evals = 10,
max_delta = 10000, EM_verbose = FALSE, optim_verbose = FALSE, npd = FALSE,
nested_optim = FALSE, usezscores = TRUE, phenocov_list = list(),
ret_args = FALSE, ret_level = 1, get_cov_CIs = FALSE)

Arguments

formula            Model formula – e.g. Y~X1+X2
trait_data         A data frame with the first column labeled "species" (with species names matching tips on the phylogeny) and one column per trait. Each row corresponds to a single observation, and multiple observations for species are allowed. Missing data should be represented with NA.
tree               An object of class phylo
model              Model of evolution. Default is "BM". Alternative evolutionary models include "mvOU" (for the multivariate Ornstein-Uhlenbeck), or univariate tree transformations: "OU" "lambda", "kappa", "delta", "EB", "star".
pheno_error        If TRUE (default, unless <=1 observation per species is provided), parameters are estimated assuming within-species variation.
phylo_correlated   If TRUE (default), parameters are estimated assuming traits are correlated.
pheno_correlated   If TRUE (default), parameters are estimated assuming within-species observations traits are correlated.
REML               If TRUE (default), the algorithm will return REML estimates. If FALSE, maximum likelihood estimates will be returned.
full_alpha         Only applicable for the multivariate OU (model="mvOU"). If TRUE (default), a fully parametrized alpha matrix is fit. If FALSE, a diagonal alpha matrix is fit.
phylocov_start     Optional starting value for phylogenetic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
phenocov_start     Optional starting value for phenotypic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
model_par_start    Optional starting parameters for the evolutionary model. For model="mvOU", must be of dimension n_traits by n_traits. Otherwise, must be a single value.
phylocov_fixed     Optional fixed value for phylogenetic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
phenocov_fixed     Optional starting value for phenotypic trait variance-covariance matrix. Must be of dimension n_traits by n_traits.
model_par_fixed    Optional fixed parameter for the evolutionary model. For model="mvOU", must be of dimension n_traits by n_traits. Otherwise, must be a single value.
skip_optim         Whether to skip BFGS optimization (not recommended unless all parameters are fixed).
Whether to skip Expectation-Maximization prior to generating starting parameters for BFGS optimization (not recommended unless providing fixed parameters).

**EM_Fels_limit**
Whether to skip Expectation-Maximization prior to generating starting parameters for BFGS optimization (not recommended unless providing fixed parameters).

**repeat_optim_limit**
The number of times to repeat numerical optimization (default is 1).

**EM_missing_limit**
Maximum number of iterations for EM.

**repeat_optim_tol**
Maximum tolerance for repeated numerical optimization (only relevant if repeat_optim_limit>1).

**model_par_evals**
Number of times to evaluate univariate tree transformation models along the range of possible parameter values. Used to generate informed starting values for alternative evolutionary models if nested_optim=TRUE.

**max_delta**
Maximum allowed difference between the log-likelihood for EM-generated starting parameters and new parameters tried under numerical optimization. Extremely large deltas are likely to be numerical artifacts. Prevents artificial convergence.

**EM_verbose**
Whether to print the log-likelihood during Expectation-Maximization.

**optim_verbose**
Whether to print log-likelihoods during numerical optimization.

**npd**
Whether to find the nearest positive-definite matrix for all covariance matrices during numerical optimization (slow – only set to TRUE if converging to singular matrices).

**nested_optim**
Only relevant if fitting a univariate alternative evolutionary model. Tries multiple tree transformation parameter values along the range of possible values to make informed starting parameters. Slower than the default (nested_optim=FALSE), in which all parameters are estimated simultaneously.

**usezscores**
Whether or not to use centered and standardized data during numerical optimization (recommended).

**phenocov_list**
An optional named list of species-specific within-species covariance matrices to be held fixed, as in Ives et al (2007). This option forces pheno_error and pheno_correlated to be FALSE, and uses mean species values instead of raw data. Raw variance should be divided by the number of observations per species (i.e., squared standard errors). See Ives et al (2007) for more details.

**ret_args**
For internal use only.

**ret_level**
For internal use only.

**get_cov_CIs**
Whether to return 95-percent confidence intervals of covariance parameters (default=FALSE).

**Value**

A fitted phylopars.lm object.
Examples

# simulate data
sim_data <- simtraits(ntaxa = 15, ntraits = 4)

phylopars.lm(V4 ~ V1 + V2 + V3, trait_data = sim_data$trait_data, tree = sim_data$tree)

print.phylopars
Print phylopars

Description
Generic S3 method for phylopars

Usage

## S3 method for class 'phylopars'
print(x, ...)

Arguments

x Fitted phylopars object

...

print.phylopars.lm
Print phylopars.lm

Description
Generic S3 method for phylopars.lm

Usage

## S3 method for class 'phylopars.lm'
print(x, ...)

Arguments

x Fitted phylopars.lm object

...
print.SSC  

*Print SSC*

**Description**

Generic S3 method for objects returned by the function `fast.SSC`

**Usage**

```r
## S3 method for class 'SSC'
print(x, ...)
```

**Arguments**

- `x`: Object returned by `fast.SSC`
- `...`: Additional arguments

---

simtraits  

*Simulate traits for phylopars estimation*

**Description**

Simulates traits for codephylopars estimation.

**Usage**

```r
simtraits(ntaxa = 15, ntraits = 4, nreps = 1, nmissing = 0, tree, v, anc,
intraspecific, model="BM", parameters, nsim, return.type="data.frame")
```

**Arguments**

- `ntaxa`: Either number of taxa (`ntaxa`) or a tree can be supplied.
- `ntraits`: Number of traits to be simulated.
- `nreps`: Number of replicates per trait per species to simulate.
- `nmissing`: Number of randomly missing trait values.
- `tree`: Either number of taxa (`ntaxa`) or a tree can be supplied.
- `v`: Trait covariance (`v`) can be optionally supplied; otherwise off-diagonal elements are set to 0.8.
- `anc`: Value for ancestral state at root node.
- `intraspecific`: Optional value for within-species variance.
- `model`: Model of evolution (default="BM"). Other options include "OUfixedRoot", "OUrandomRoot", "lambda", "kappa", "delta", "EB".
- `parameters`: List of parameters for the model. alpha for the selection strength in the OU model, lambda, kappa, delta, or rate for the EB model.
- `nsim`: Number of simulations to perform (default is 1)
- `return.type`: Default is "data.frame". Can also specify "matrix" if `nreps`=1.
summary.phylopars

Value

trait_data Data for phylopars()

Tree

The original phylogenetic tree (either provided to the function or generated internally)

sim_tree The transformed tree on which trait simulations were performed (identical to tree if model="BM")

original_X If within-species variation is simulated, original_X is the original species mean values before adding within-species variation.

Author(s)

Eric W. Goolsby <eric.goolsby.evolution@gmail.com>

References


Examples

# simulate data
sim_data <- simtraits(ntaxa = 15,ntraits = 4,nreps = 3,nmissing = 10)

# estimate parameters under Brownian motion
# pheno_error = TRUE assumes intraspecific variation
# pheno_correlated = FALSE assumes intraspecific variation is not correlated
# phylo_correlated = TRUE assumed traits are correlated

PPE <- phylopars(trait_data = sim_data$trait_data,tree = sim_data$tree,
                 pheno_error = TRUE,phylo_correlated = TRUE,pheno_correlated = FALSE)

PPE

summary.phylopars  Phylopars summary

Description

Summarizes phylopars

Usage

## S3 method for class 'phylopars'
summary(object, ...)
write.phylopars

Arguments

object  Fitted phylopars object

---

summary.phylopars.lm  phylopars.lm summary

---

Description

Summarizes phylopars.lm

Usage

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

Arguments

object  Fitted phylopars.lm object

---

write.phylopars  Write data and tree files for Python phylopars compatibility.

---

Description

Writes data and tree files for Python phylopars compatibility.

Usage

write.phylopars(trait_data, tree, data_file, tree_file, species_identifier = "species")

Arguments

trait_data  A data frame with one column per trait, as well as a column labeled "species" (with species names matching tips on the phylogeny). Each row corresponds to a single observation, and multiple observation for species are allowed. Missing data should be represented with NA.

tree  An object of class phylo
data_file  Desired path to write data file.
tree_file  Desired path to write tree file.
species_identifier  Title of species column in data file. Defaulted to "species"
write.phylopars

Author(s)

Eric W. Goolsby <eric.goolsby.evolution@gmail.com>

References


Examples

```r
## Not run:
# simulate data
sim_data <- simtraits(nntaxa = 15, ntraits = 4, nreps = 3, nmissing = 10)

write.phylopars(trait_data = sim_data$trait_data, tree = sim_data$tree, data_file = "data_path.txt", tree_file = "tree_path.new")

## End(Not run)
```
Index

* Ancestral state reconstruction
  Rphylopars-package, 2
* Intraspecific variation
  Rphylopars-package, 2
* Measurement error
  Rphylopars-package, 2
* Missing data
  Rphylopars-package, 2
* Phylogenetic imputation
  Rphylopars-package, 2
* Phylogenetic variance-covariance
  Rphylopars-package, 2
* Within-species variation
  Rphylopars-package, 2
* package
  Rphylopars-package, 2

ace, 4
anc.recon, 3, 5
anova.phylopars.lm, 4

fast.SSC, 5
fastAnc, 4

logLik.phylopars, 6
logLik.phylopars.lm, 6

phylopars, 3, 7
phylopars.lm, 10
pic, 4
print.phylopars, 13
print.phylopars.lm, 13
print.SSC, 14

Rphylopars-package, 2

simtraits, 14
summary.phylopars, 15
summary.phylopars.lm, 16

write.phylopars, 16