Package ‘phylosignalDB’

May 9, 2026

Title Explore Phylogenetic Signals Using Distance-Based Methods

Version 0.4.2

Description A unified method, called M statistic, is provided for detecting phylogenetic signals in continuous traits, discrete traits, and multi-trait combinations. Blomberg and Garland (2002) <doi:10.1046/j.1420-9101.2002.00472.x> provided a widely accepted statistical definition of the phylogenetic signal, which is the ``tendency for related species to resemble each other more than they resemble species drawn at random from the tree''. The M statistic strictly adheres to the definition of phylogenetic signal, formulating an index and developing a method of testing in strict accordance with the definition, instead of relying on correlation analysis or evolutionary models. The novel method equivalently expressed the textual definition of the phylogenetic signal as an inequality equation of the phylogenetic and trait distances and constructed the M statistic. The M statistic implemented in this package is based on the methodology described in Yao and Yuan (2025) <doi:10.1002/ece3.71106>. If you use this method in your research, please cite the paper.

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.0)

LazyData true

Imports ape (>= 5.7), castor (>= 1.8.3), cluster (>= 2.1.6), doParallel (>= 1.0.17), foreach (>= 1.5.2)

URL `https://github.com/anonymous-eco/phylosignalDB`

BugReports `https://github.com/anonymous-eco/phylosignalDB/issues`

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

License GPL (>= 3)

NeedsCompilation no

Author Liang Yao [aut, cre],
Ye Yuan [aut]

Maintainer Liang Yao <`yaoliang986@ahnu.edu.cn`>

Repository CRAN

Date/Publication 2026-01-29 07:50:02 UTC

Contents

`gower_dist` *Calculate Gower distance*

Description

`gower_dist()` calculates Gower distance among observations or species.

Usage

```
gower_dist(x, type = list(), dist_format = c("matrix", "dist"))
```

Arguments

`x`	A data frame. The columns usually represent trait data, and the row names are species names.
`type`	A list for specifying the variable types of the columns in `x`. Default is numeric type. More details in `cluster::daisy()`.
`dist_format`	The class of the return value. Default is "matrix".

Value

A matrix or dist object containing the Gower distance among the rows of `x`.

References

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. Biometrics: 857-871.

Kaufman, L. & Rousseeuw, P.J. (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

See Also

`cluster::daisy()` which this function wraps.

Examples

```
data("turtles")
# Continuous trait
trait_df <- data.frame(M1 = turtles$traits$M1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df)

# Nominal discrete trait
trait_df <- data.frame(B1 = turtles$traits$B1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = 1))

# Ordinal discrete trait
trait_df <- data.frame(CS1 = turtles$traits$CS1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(ordered = 1))

# Multi-trait Combinations
trait_df <- data.frame(turtles$traits[, c("M1", "M2", "M3", "M4", "M5")],
                       row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = c("M4", "M5")))
```

`M_rand_perm` *Calculate M statistics after random permutations*

Description

`M_rand_perm` calculates M statistic for trait(s) after randomly permuting the species names or tip labels in phylogeny. The M statistic is a unified method for detecting phylogenetic signals in continuous traits, discrete traits, and multi-trait combinations. Blomberg and Garland (2002) provided a widely accepted statistical definition of the phylogenetic signal, which is the "tendency for related species to resemble each other more than they resemble species drawn at random from the tree". The M statistic strictly adheres to the definition of phylogenetic signal, formulating an index and developing a method of testing in strict accordance with the definition, instead of relying on correlation analysis or evolutionary models. The novel method equivalently expressed the textual definition of the phylogenetic signal as an inequality equation of the phylogenetic and trait distances and constructed the M statistic. The M statistic implemented in this package is based on the methodology described in Yao and Yuan (2025) doi:10.1002/ece3.71106. If you use this method in your research, please cite the paper.

Usage

```
M_rand_perm(
  trait_dist = NULL,
  phy = NULL,
  reps = 999,
  auto_multi2di = TRUE,
  cores = 1
)
```

Arguments

`trait_dist`	A distance object of class `matrix` or `dist`. Its row and column names should match the tip labels of the phylogenetic tree (phy). The functions `gower_dist()` and `cluster::daisy()` can be used to calculate distances using trait data.
`phy`	A phylogenetic tree of class `phylo`.
`reps`	An integer. The number of random permutations.
`auto_multi2di`	A logical switch, `TRUE` or `FALSE`. Default is `TRUE`, then function `ape::multi2di()` in ape package will be called to make the phylogeney (tree) be dichotomous if the tree (phy) contains some polytomies.
`cores`	Number of cores to be used in parallel processing. Default is 1, indicating no parallel computation is performed. If set to 0, parallel computation is executed using `parallel::detectCores() - 1` number of cores.

Value

A list object containing two components. Component `$permuted` is the vector of M values obtained after random permutation for `reps` times; component `$observed` is the value of M statistic obtained from the original input data.

References

Blomberg, S.P. & Garland, T., Jr (2002) Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology, 15(6): 899-910.

Yao, L. and Yuan, Y. (2025), A Unified Method for Detecting Phylogenetic Signals in Continuous, Discrete, and Multiple Trait Combinations. Ecology and Evolution, 15: e71106.

See Also

`M_stat()` `phylosignal_M()`

Examples

```
data("turtles")
# Continuous trait
trait_df <- data.frame(M1 = turtles$traits$M1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df)
M_rand_perm(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Nominal discrete trait
trait_df <- data.frame(B1 = turtles$traits$B1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = 1))
M_rand_perm(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Ordinal discrete trait
trait_df <- data.frame(CS1 = turtles$traits$CS1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(ordered = 1))
M_rand_perm(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Multi-trait Combinations
```

```
trait_df <- data.frame(turtles$traits[, c("M1", "M2", "M3", "M4", "M5")],
                       row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = c("M4", "M5")))
M_rand_perm(trait_dist, turtles$phylo, reps = 99) # reps=999 better
```

`M_stat` *Calculate M statistic*

Description

`M_stat` calculates the value of M statistic as a measurement of the strength of the phylogenetic signal for the trait(s). The trait(s) could be continuous, discrete, or multi-variable. Blomberg and Garland (2002) provided a widely accepted statistical definition of the phylogenetic signal, which is the "tendency for related species to resemble each other more than they resemble species drawn at random from the tree". The M statistic strictly adheres to the definition of phylogenetic signal, formulating an index and developing a method of testing in strict accordance with the definition, instead of relying on correlation analysis or evolutionary models. The novel method equivalently expressed the textual definition of the phylogenetic signal as an inequality equation of the phylogenetic and trait distances and constructed the M statistic. The M statistic implemented in this package is based on the methodology described in Yao and Yuan (2025) `doi:10.1002/ece3.71106`. If you use this method in your research, please cite the paper.

Usage

```
M_stat(trait_dist = NULL, phy = NULL, auto_multi2di = TRUE)
```

Arguments

`trait_dist`	A distance object of class `matrix` or `dist`. Its row and column names should match the tip labels of the phylogenetic tree (phy). The functions `gower_dist()` and `cluster::daisy()` can be used to calculate distances using trait data.
`phy`	A phylogenetic tree of class `phylo`.
`auto_multi2di`	A logical switch, `TRUE` or `FALSE`. Default is `TRUE`, then function `ape::multi2di()` in ape package will be called to make the phylogeny (tree) be dichotomous if the tree (phy) contains some polytomies.

Value

A value that lies between 0 and 1, inclusive.

References

Blomberg, S.P. & Garland, T., Jr (2002) Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology, 15(6): 899-910.

Yao, L. and Yuan, Y. (2025), A Unified Method for Detecting Phylogenetic Signals in Continuous, Discrete, and Multiple Trait Combinations. Ecology and Evolution, 15: e71106.

See Also

`M_rand_perm()` `phylosignal_M()`

Examples

```
data("turtles")
# Continuous trait
trait_df <- data.frame(M1 = turtles$traits$M1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df)
M_stat(trait_dist, turtles$phylo)

# Nominal discrete trait
trait_df <- data.frame(B1 = turtles$traits$B1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = 1))
M_stat(trait_dist, turtles$phylo)

# Ordinal discrete trait
trait_df <- data.frame(CS1 = turtles$traits$CS1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(ordered = 1))
M_stat(trait_dist, turtles$phylo)

# Multi-trait Combinations
trait_df <- data.frame(turtles$traits[, c("M1", "M2", "M3", "M4", "M5")],
                       row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = c("M4", "M5")))
M_stat(trait_dist, turtles$phylo)
```

`phylosignal_M` *Measure and test phylogenetic signal with M statistic*

Description

`phylosignal_M` computes the M statistic for trait(s) and evaluates its statistical significance through a random permutation test. The M statistic is a unified method for detecting phylogenetic signals in continuous traits, discrete traits, and multi-trait combinations. Blomberg and Garland (2002) provided a widely accepted statistical definition of the phylogenetic signal, which is the "tendency for related species to resemble each other more than they resemble species drawn at random from the tree". The M statistic strictly adheres to the definition of phylogenetic signal, formulating an index and developing a method of testing in strict accordance with the definition, instead of relying on correlation analysis or evolutionary models. The novel method equivalently expressed the textual definition of the phylogenetic signal as an inequality equation of the phylogenetic and trait distances and constructed the M statistic. The M statistic implemented in this package is based on the methodology described in Yao and Yuan (2025) `doi:10.1002/ece3.71106`. If you use this method in your research, please cite the paper.

Usage

```
phylosignal_M(
  trait_dist = NULL,
  phy = NULL,
  reps = 999,
  auto_multi2di = TRUE,
  output_M_permuted = FALSE,
  cores = 1
)
```

Arguments

`trait_dist`	A distance object of class `matrix` or `dist`. Its row and column names should match the tip labels of the phylogenetic tree (phy). The functions `gower_dist()` and `cluster::daisy()` can be used to calculate distances using trait data.
`phy`	A phylogenetic tree of class `phylo`.
`reps`	An integer. The number of random permutations.
`auto_multi2di`	A logical switch, `TRUE` or `FALSE`. Default is `TRUE`, then function `ape::multi2di()` in ape package will be called to make the phylogeny (tree) be dichotomous if the tree (phy) contains some polytomies.
`output_M_permuted`	A logical switch, `TRUE` or `FALSE`. Default is `FALSE`. If this logical switch is set to `TRUE`, the returned list will include the vector of M values obtained after random permutations.
`cores`	Number of cores to be used in parallel processing. Default is 1, indicating no parallel computation is performed. If set to 0, parallel computation is executed using `parallel::detectCores() - 1` number of cores.

Value

A list object containing two components. Component `$permuted` is the vector of M values obtained after random permutation for `reps` times; component `$observed` is the value of M statistic obtained from the original input data.

References

Blomberg, S.P. & Garland, T., Jr (2002) Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology, 15(6): 899-910.

Yao, L. and Yuan, Y. (2025), A Unified Method for Detecting Phylogenetic Signals in Continuous, Discrete, and Multiple Trait Combinations. Ecology and Evolution, 15: e71106.

See Also

`M_stat()` `M_rand_perm()`

Examples

```
data("turtles")
# Continuous trait
trait_df <- data.frame(M1 = turtles$traits$M1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df)
phylosignal_M(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Nominal discrete trait
trait_df <- data.frame(B1 = turtles$traits$B1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = 1))
phylosignal_M(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Ordinal discrete trait
trait_df <- data.frame(CS1 = turtles$traits$CS1, row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(ordered = 1))
phylosignal_M(trait_dist, turtles$phylo, reps = 99) # reps=999 better

# Multi-trait Combinations
trait_df <- data.frame(turtles$traits[, c("M1", "M2", "M3", "M4", "M5")],
                       row.names = turtles$traits$specie)
trait_dist <- gower_dist(x = trait_df, type = list(factor = c("M4", "M5")))
phylosignal_M(trait_dist, turtles$phylo, reps = 99) # reps=999 better
```

`turtles` *Ecological Traits and Phylogeny of Turtles*

Description

An ecological trait dataset for turtles. The dataset was derived from the recently published ReptTraits dataset (Oskyrko et al., 2024), extracting species classified under the major group Testudines (comprising 361 species). Only ecological traits with more than 50% of the species having trait records were retained. The phylogeny of turtles was derived by pruning from the maximum clade credibility tree with 288 tips provided in Thomson et al. (2021). Only those species that are present in both the ReptTraits dataset and the turtle phylogenetic tree were selected. Ultimately, the dataset comprised 240 species, encompassing 5 morphology traits, 2 behaviour traits, 2 life history traits, 5 habitat variables, and 2 variables concerning species conservation status.

Usage

```
turtles
data("turtles")
```

Format

`turtles` is a list object with 3 components:

traits The ecological traits of turtles as an object of class `data.frame/tibble`.

phylo The phylogeny of turtles as an object of class `phylo`.

traits_info The full names and id of ecological traits. More details in Oskyrko et al. (2024) and Thomson et al. (2021).

References

Oskyrko, O., Mi, C., Meiri, S. & Du, W. (2024) ReptTraits: a comprehensive dataset of ecological traits in reptiles. Scientific Data, 11(1): 243.

Thomson, R.C., Spinks, P.Q. & Shaffer, H.B. (2021) A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. Proceedings of the National Academy of Sciences, 118(7): e2012215118.

Index