Package 'PNC'

November 7, 2025

Type Package

Title Phylogenetic Niche Conservatism Analysis for Ecological
Communities

Version 0.1.0

Date 2025-11-5

Maintainer Yan He <heyan@njfu.edu.cn>

Description Provides functions for testing phylogenetic niche conservatism, a key prerequisite in community assembly studies. The package integrates global functional trait data across major taxonomic groups and implements methods such as Pagel's Lambda and Blomberg's K to quantify phylogenetic signals in ecological communities. Methods are described in Münkemüller et al. (2012) <doi:10.1111/j.2041-210X.2012.00196.x>.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports ape, phytools, stats, utils, geiger

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Yan He [aut, cre],
Yu Xia [aut],
Rui Yang [aut],
Lingfeng Mao [aut]

Repository CRAN

Date/Publication 2025-11-07 13:40:13 UTC

Contents

AmphiBIO   *AmphiBIO: Global Amphibian Ecological Traits Database*

Description

A comprehensive global database of ecological traits for amphibian species, compiled to provide insights into the life history and ecological characteristics of amphibians worldwide.

Usage

AmphiBIO

Format

A data frame with multiple variables:

species Scientific name of the amphibian species

genus Taxonomic genus of the species

family Taxonomic family of the species

Body_mass_g Maximum adult body mass.

Age_at_maturity_min_y Minimum age at maturation or sexual maturity.

Age_at_maturity_max_y Maximum age at maturation or sexual maturity.

Body_size_mm Maximum adult body size. In Anura, body size is reported as snout to vent length. In Gymnophiona and Caudata, body size is reported as total length.

Size_at_maturity_min_mm Minimum size at maturation or sexual maturity.

Size_at_maturity_max_mm Maximum size at maturation or sexual maturity.

Longevity_max_y Maximum life span.

Litter_size_min_n Minimum no. of offspring or eggs per clutch.

Litter_size_max_n Maximum no. of offspring or eggs per clutch.

Reproductive_output_y Maximum no. reproduction events per year.

Offspring_size_min_mm Minimum offspring or egg size.

Offspring_size_max_mm Maximum offspring or egg size.

References

Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. Scientific data, 4(1), 1-7. doi:10.1038/sdata.2017.123

Examples

```
# Load the dataset
data(AmphiBIO)
head(AmphiBIO)
```

AVONET   *AVONET Bird Morphological Dataset*

Description

Comprehensive morphological dataset for bird species, including taxonomic information from BirdLife International and detailed morphological measurements.

Usage

```
AVONET
```

Format

A data frame with 11,009 rows and 14 columns, where each row represents a bird species:

species Species scientific name

genus Genus name

family Family name, according to BirdLife International taxonomy

Beak.Length_Culmen Length from beak tip to skull base, in millimeters

Beak.Length_Nares Length from nostril anterior edge to beak tip, in millimeters

Beak.Width Beak width at the anterior edge of nostrils, in millimeters

Beak.Depth Beak depth at the anterior edge of nostrils, in millimeters

Tarsus.Length Tarsus length from posterior notch between tibia and tarsus to the last scale end, in millimeters

Wing.Length Length from carpal joint to longest primary feather tip, in millimeters

Kipps.Distance Length from first secondary feather tip to longest primary feather tip, in millimeters

Secondary1 Length from carpal joint to first secondary feather tip, in millimeters

Hand-Wing.Index 100*DK/Lw, where DK is Kipp's distance and Lw is wing length

Tail.Length Distance from longest rectrix tip to point where central rectrices protrude from skin, in millimeters

Mass Species average body mass, including both male and female, in grams

Details

This dataset provides comprehensive morphological measurements of birds, including beak, wing, tarsus, and body weight indicators. Data originates from a comprehensive study of bird morphological, ecological, and geographical characteristics.

Note

- Taxonomic information based on BirdLife International - Measurements represent species averages - Hand-Wing Index reflects flight capability and ecological adaptation

References

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., Darski, B., et al. (2022). AVONET: morphological, ecological and geographical data for all birds. Ecology Letters, 25(3), 581-597. doi:10.1111/ele.13898

Examples

```
data(AVONET)
head(AVONET)
```

BCI   *Barro Colorado Island (BCI) dataset*

Description

The Barro Colorado Island (BCI) dataset contains comprehensive ecological data from the 50-hectare forest dynamics plot on Barro Colorado Island, Panama. This dataset includes phylogenetic information and community composition data for tropical forest species.

Usage

```
BCI
```

Format

A list containing four main components:

splist A data frame with species information including species names, genus, and family classifi-
cations.

phy_species A phylogenetic tree representing species-level evolutionary relationships, rooted and
including branch lengths.

phy_genus A phylogenetic tree with 183 tips and 174 internal nodes, rooted and including branch
lengths.

com A community matrix showing species abundance across different sampling plots, with species
counts for each location.

Source

Barro Colorado Island (BCI)

References

Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., & Hubbell, S. P. (2019). Complete data from
the Barro Colorado 50-ha plot: 423617 trees, 35 years, 2019 version. Dryad Digital Repository.
doi:10.15146/5xcp0d46

Examples

```
# Load the dataset
data(BCI)
head(BCI)
```

COMBINE   *COMBINE: Mammal Trait Database*

Description

A comprehensive dataset of mammalian traits compiled from multiple sources, providing detailed
ecological and biological information for various mammal species.

Usage

```
COMBINE
```

Format

A data frame with the following columns:

species Species name

genus Genus name

family Taxonomic family

adult_mass_g Body mass of an adult individual in grams

adult_brain_mass_g Weight of the brain of an adult individual in grams

adult_body_length_mm Total length from tip of the nose to anus or base of the tail of an adult individual in millimeters

adult_forearm_length_mm Total length from elbow to wrist of an adult individual in millimeters, specific to order Chiroptera

max_longevity_d Maximum reported age at death for the species in days

maturity_d The amount of time needed to reach sexual maturity in days

female_maturity_d The amount of time needed for a female to reach sexual maturity in days

male_maturity_d Age at which females give birth to their first litter or their young attach to teats in days

age_first_reproduction_d Age at first reproduction in days

gestation_length_d Length of time of fetal growth in days

teat_number_n Total number of teats present in an individual of the species

litter_size_n Number of offspring born per litter per female

litters_per_year_n Number of litters per female per year

interbirth_interval_d Time between reproduction events in days

neonate_mass_g Weight of an individual at birth in grams

weaning_age_d Age at which primary nutritional dependency on the mother ends and independent foraging begins in days

weaning_mass_g Weight at weaning in grams

generation_length_d Average age of parents of the current cohort in days

dispersal_km The distance an animal travels between its place of birth to the place where it reproduces in kilometers

density_n_km2 Number of individuals of the species per squared kilometer

home_range_km2 Size of the area within which everyday activities of individuals or groups of individuals are typically restricted in km2

social_group_n Number of individuals in a group that spends most of their daily time together

dphy_invertebrate Percentage of the diet composed of invertebrates

dphy_vertebrate Percentage of the diet composed of vertebrates

dphy_plant Percentage of the diet composed of plants and/or fungi

det_inv Percentage of the diet composed of invertebrates

det_vend Percentage of the diet composed of mammals, birds

det_vect Percentage of the diet composed of reptiles, snakes, amphibians, salamanders

det_vfish Percentage of the diet composed of fish

det_vunk Percentage of the diet composed of vertebrates – general or unknown

det_scav Percentage of the diet composed of scavenge, garbage, offal, carcasses, trawlers, carrion

det_fruit Percentage of the diet composed of fruit, drupes

det_nect Percentage of the diet composed of nectar, pollen, plant exudates, gums

det_seed Percentage of the diet composed of seed, maize, nuts, spores, wheat, grains

det_plantother Percentage of the diet composed of other plant elements

det_diet_breadth_n Number of prevalent EltonTraits dietary categories consumed at 20 percent or more

upper_elevation_m Upper elevation limit at which the species can be found in meters

lower_elevation_m Lower elevation limit at which the species can be found in meters

altitude_breadth_m Difference between the upper and lower elevation limits of a species in meters

habitat_breadth_n Number of distinct suitable level 1 IUCN habitats

References

Soria, C. D., M. Pacifici, M. Di Marco, S. M. Stephen, and C. Rondinini. (2021). COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. Ecology, 102(6):e03344. doi:10.1002/ecy.3344

Examples

```
data(COMBINE)
head(COMBINE)
```

---

compnc   *Calculate Phylogenetic Niche Conservatism Across Multiple Communities*

Description

This function conducts comprehensive phylogenetic niche conservatism analysis across multiple communities simultaneously. It evaluates phylogenetic signal for trait data across different community assemblages using various statistical methods, enabling comparative assessment of niche conservatism patterns among communities. The function processes community composition matrices, species trait information, and phylogenetic trees to determine whether closely related species consistently occupy similar ecological niches across different habitats or sampling locations.

Usage

```
compnc(
  com,
  trait_data,
  phylo_tree,
  methods = c("lambda", "K"),
  pca_axes = c("PC1", "PC2"),
  sig_levels = c(0.001, 0.01, 0.05),
  min_abundance = 0,
  nsim = 1000,
  verbose = TRUE
)
```

Arguments

`com`	A community matrix with sites as rows and species as columns
`trait_data`	A data frame or matrix containing trait data with species as rows
`phylo_tree`	A phylogenetic tree object of class "phylo"
`methods`	Character vector specifying methods to use. Options: "lambda", "K"
`pca_axes`	Character vector specifying which PCA axes to include (e.g., c("PC1", "PC2"))
`sig_levels`	Numeric vector of significance levels for marking results
`min_abundance`	Minimum abundance threshold for including species
`nsim`	Number of permutations for significance testing
`verbose`	Logical indicating whether to show progress and warnings

Value

A data frame containing phylogenetic signal results for all communities

Examples

```
#' # Load example data
data(BCI)
data(TRY)

# Extract trait data
sp <- colnames(BCI$com)
subtraits <- extract_traits(sp, TRY, rank = "species",
                  traits = c("LA", "LMA", "LeafN", "PlantHeight", "SeedMass", "SSD"))

compnc(com = BCI$com, subtraits, BCI$phy_species, methods = "lambda", pca_axes = NULL)
```

compnc_robustness	*Test Robustness of Phylogenetic Niche Conservatism Analysis Across Multiple Communities*

Description

This function evaluates the robustness of phylogenetic signal estimates across multiple communities by simulating trait data with the same phylogenetic signal strength as observed, applying the original missing data pattern, and testing how consistently the statistical significance is recovered across multiple simulations for each community.

Usage

```
compnc_robustness(
  com,
  trait_data,
  phylo_tree,
  methods = "lambda",
  pca_axes = c("PC1", "PC2"),
  sig_levels = c(0.001, 0.01, 0.05),
  min_abundance = 0,
  n_simulations = 100,
  alpha_level = 0.05,
  tolerance = 0.05,
  verbose = TRUE
)
```

Arguments

com	A community matrix with sites as rows and species as columns
trait_data	A data frame or matrix containing trait data with species as rows
phylo_tree	A phylogenetic tree object of class "phylo"
methods	Character string specifying method to use. Options: "lambda" or "K". Default is "lambda"
pca_axes	Character vector specifying which PCA axes to include (e.g., c("PC1", "PC2")). Default is c("PC1", "PC2")
sig_levels	Numeric vector of significance levels for marking results
min_abundance	Minimum abundance threshold for including species
n_simulations	Integer. Number of simulations to run for robustness testing. Default is 100
alpha_level	Numeric. Significance level for statistical testing. Default is 0.05
tolerance	Numeric. Acceptable difference between target and estimated signal values during trait simulation. Default is 0.05
verbose	Logical indicating whether to show progress and warnings

Value

A data frame containing the original phylogenetic signal results with additional columns:

- robustness: Percentage of simulations that maintain the same statistical significance conclusion as the original analysis
- signal_sd: Standard deviation of phylogenetic signal values across successful simulations

Examples

```
# Load example data
data("HimalayanBirds")
str(HimalayanBirds)
data("AVONET")
head(AVONET)

# species level
sp <- colnames(HimalayanBirds$com)
sp
subtraits <- extract_traits(sp, AVONET, rank = "species")
head(subtraits)
coverage(subtraits)
pnc(subtraits, HimalayanBirds$phy_species, methods = "lambda", pca_axes = c("PC1", "PC2"))

compnc(com = HimalayanBirds$com, subtraits, HimalayanBirds$phy_species,
       methods = "lambda", pca_axes = NULL)

# Test robustness of phylogenetic signal analysis
# This function's runtime is long
compnc_robustness(HimalayanBirds$com,
                  subtraits,
                  HimalayanBirds$phy_species,
                  methods = "lambda",
                  pca_axes = NULL,
                  n_simulations = 5)
```

---

coverage	*Calculate Trait Coverage Statistics*

Description

This function calculates comprehensive coverage statistics for trait data, including individual trait coverage rates, complete case coverage, and overall data coverage. It provides both summary statistics and detailed breakdowns of missing and available data.

Usage

```
coverage(data)
```

Arguments

data   A data frame containing trait data. Each column represents a trait and each row represents an observation (e.g., species, samples).

Details

The function performs the following calculations:

- **Individual trait coverage**: For each trait, calculates the number and percentage of available (non-NA) values

- **Complete case coverage**: Counts rows with no missing values across all traits and calculates the percentage

- **Overall coverage**: Calculates the percentage of all cells in the dataset that contain non-missing values

The function also prints the overall trait coverage rate to the console before returning the detailed summary table.

Value

A data frame with the following columns:

Trait Character. Names of traits plus an "All" row for complete cases

Available_count Integer. Number of non-missing values for each trait

Missing_count Integer. Number of missing (NA) values for each trait

Trait_coverage_rate Character. Percentage of available data for each trait

The "All" row shows statistics for complete cases (rows with no missing values).

Examples

```
# Create sample trait data
trait_data <- data.frame(
  PlantHeight = c(1.2, 1.5, NA, 2.1, 1.8),
  LDMC = c(0.5, NA, 0.8, 1.2, 0.9),
  LA = c(15.2, 18.5, 12.3, NA, 16.7)
)

# Calculate coverage statistics
coverage(trait_data)
```

---

extract_traits   *Extract Plant Traits from Trait Database*

Description

This function extracts plant trait data from the TRY database or similar datasets for a specified list of taxa at different taxonomic ranks (species, genus, or family). For numeric traits at genus and family levels, it calculates mean values across all available records.

Usage

```
extract_traits(sp.list, dataset, rank = "species", traits = NULL)
```

Arguments

sp.list	A character vector containing the names of taxa to extract traits for. The names should match the taxonomic rank specified in the 'rank' parameter.
dataset	A data frame containing trait data. Default is TRY database. Must contain columns named "species", "genus", and "family" for taxonomic information.
rank	A character string specifying the taxonomic rank to match against. Must be one of "species", "genus", or "family". Default is "species".
traits	A character vector specifying which traits to extract. If NULL (default), all available traits in the dataset will be extracted. Available traits are all columns except "species", "genus", and "family".

Details

The function performs the following operations:

- Validates input parameters
- Identifies available traits in the dataset
- Matches input taxa with dataset entries
- Reports missing taxa
- Extracts trait data based on the specified taxonomic rank
- For numeric traits at genus/family level, calculates mean values
- For non-numeric traits, uses the first available value
- Handles NaN values by converting them to NA

Value

A data frame with taxa names as row names and trait names as column names. For species-level extraction, returns the first occurrence of each species. For genus/family-level extraction, returns mean values for numeric traits and the first occurrence for non-numeric traits. Missing values are represented as NA.

Examples

```
# Load the dataset
data(TRY)

# Extract all traits for species
species_list <- c("Acaena novae-zelandiae", "Adiantum capillus-veneris", "Zuelania guidonia")
extract_traits(species_list, TRY, rank = "species")

# Extract specific traits for species
extract_traits(species_list, TRY, rank = "species",
                traits = c("LA", "LMA", "LeafN", "PlantHeight", "SeedMass", "SSD"))

# Extract specific traits at genus level
genus_list <- c("Acaena", "Adiantum")
extract_traits(genus_list, TRY, rank = "genus",
                traits = c("LDMC", "PlantHeight", "SeedMass"))
```

Fishlife   *Fishlife Dataset*

Description

A comprehensive dataset of fish life history traits across multiple species, compiled by Thorson et al. (2023). The dataset provides various morphological, ecological, and biological characteristics of fish species.

Usage

```
Fishlife
```

Format

A data frame with multiple variables:

species Scientific species name

genus Genus of the fish species

family Family classification

age_max Maximum age, years

trophic_level Trophic level, where 1 is primary producers, etc., dimensionless

aspect_ratio Caudal fin height and length divided by area, dimensionless

fecundity Annual eggs produced, number/year

growth_coefficient von Bertalannffy growth coefficient, year-1

temperature Average temperature from portion of population sampled, celcius

length_max maximum length, cm

length_infinity von Bertalanffy asymptotic maximum length, cm

length_maturity Length at 50% maturity, cm

age_maturity Age at 50% sexual maturity, years

natural_mortality Natural mortality rate M, year-1

weight_infinity Asymptotic maximum weight, g

max_body_depth Maximum body depth, cm

max_body_width Maximum body width, cm

lower_jaw_length Length of lower jaw, cm

min_caudal_pedoncule_depth Depth of caudal pedoncule, connecting caudal fin to body

offspring_size Size of offspring, kg

References

Thorson, J. T., Maureaud, A. A., Frelat, R., Mérigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. Methods in Ecology and Evolution, 14(5), 1243-1255. doi:10.1111/2041210X.14076

Examples

```
data(Fishlife)
head(Fishlife)
```

HimalayanBirds   *Himalayan Birds Dataset*

Description

The 'HimalayanBirds' dataset provides information on bird species in the Himalayas, including their species names, genera, families, phylogenetic relationships, and community composition across elevation bands. This dataset is used to explore elevational patterns of bird functional and phylogenetic diversity and the ecological processes that structure bird communities.

Usage

```
HimalayanBirds
```

Format

A list with three components:

splist A data frame with 151 rows and 3 variables:

   species Scientific name of the bird species.
   genus Genus of the bird species.
   family Family of the bird species.

phy_species A phylogenetic tree (object of class "phylo") representing the evolutionary relation-ships among the bird species. It contains edge, edge.length, Nnode, tip.label, and node.label.

com A community matrix representing the presence (1) or absence (0) of each bird species across 12 elevation bands (ele1 to ele12). The rows represent the elevation bands, and the columns represent the bird species.

References

Ding, Z., Hu, H., Cadotte, M.W., Liang, J., Hu, Y., & Si, X. (2021). Elevational patterns of bird functional and phylogenetic structure in the central Himalaya. Ecography, 44(9), 1403-1417. doi:10.1111/ecog.05660

Examples

```
# Load the dataset
data(HimalayanBirds)
head(HimalayanBirds)
```

---

merge_dataset   *Merge Two Datasets Based on Species Column*

Description

This function merges two data frames based on the 'species' column, handling missing values and column differences intelligently. It provides flexible options for resolving conflicts when the same species appears in both datasets.

Usage

```
merge_dataset(main_data, additional_data, priority = "main")
```

Arguments

main_data   A data frame containing the primary dataset. Must include a 'species' column.

additional_data
                   A data frame containing the secondary dataset. Must include a 'species' column.

priority   A character string specifying how to handle conflicts when both datasets contain non-missing values for the same species and column. Options are:

- "main" (default): Use values from main_data
- "additional": Use values from additional_data
- "mean": Calculate mean for numeric values, use main_data for non-numeric

Details

The function performs the following operations:

- Combines all unique species from both datasets

- Includes all columns from both datasets

- Handles missing values by using available non-missing values

- Resolves conflicts based on the specified priority

- For duplicate species within a dataset, only the first occurrence is used

Value

A data frame containing all unique species from both input datasets, with all columns from both datasets. The 'species' column is placed first, followed by all other columns in alphabetical order.

Note

- Both input datasets must contain a 'species' column

- If a species appears multiple times in a dataset, only the first occurrence is used

- When priority is "mean", non-numeric values default to main_data values

- The function preserves the original data types of columns

Examples

```
# Create sample datasets
main_data <- data.frame(
  species = c("Abies alba", "Coussapoa trinervia", "Crataegus monogyna"),
  genus = c("Abies", "Coussapoa", "Crataegus"),
  family = c("Pinaceae", "Urticaceae", "Rosaceae"),
  LA = c(NA, 2050.24, 449.15),
  LeafN = c(13.10, 14.52, 17.46),
  Seedmass = c(53.64, NA, 95.92),
  stringsAsFactors = FALSE
)

additional_data <- data.frame(
  species = c("Abies alba", "Corydalis solida"),
  genus = c("Abies", "Corydalis"),
  family = c("Pinaceae", "Papaveraceae"),
  LA = c(25.58, NA),
  LMA = c(0.19, 0.2),
  PlantHeight = c(53.66, 0.14),
  stringsAsFactors = FALSE
)

# Merge with main data priority (default)
merge_dataset(main_data, additional_data)
```

---

pnc   *Analyze Phylogenetic Niche Conservatism in Ecological Communities*

Description

This function performs in-depth phylogenetic niche conservatism analysis for communities by quantifying phylogenetic signal in trait data using multiple statistical methods. The function integrates trait data preprocessing, phylogenetic tree manipulation, optional principal component analysis, and robust statistical testing to provide detailed insights into evolutionary constraints on trait evolution.

Usage

```
pnc(
  trait_data,
  phylo_tree,
  methods = "lambda",
  pca_axes = c("PC1", "PC2"),
  sig_levels = c(0.001, 0.01, 0.05),
  nsim = 1000,
  verbose = TRUE
)
```

Arguments

trait_data	A data frame or matrix containing trait data with species as rows
phylo_tree	A phylogenetic tree object of class "phylo"
methods	Character vector specifying methods to use. Options: "lambda", "K"
pca_axes	Character vector specifying which PCA axes to include (e.g., c("PC1", "PC2"))
sig_levels	Numeric vector of significance levels for marking results
nsim	Number of permutations for significance testing
verbose	Logical indicating whether to show progress and warnings

Value

A data frame containing phylogenetic signal results

References

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K. and Thuiller, W. (2012). How to measure and test phylogenetic signal. Methods in Ecology and Evolution, 3(4), 743-756. doi:10.1111/j.2041210X.2012.00196.x

Examples

```
#' # Load example data
data(BCI)
data(TRY)

# Extract trait data
sp <- colnames(BCI$com)
subtraits <- extract_traits(sp, TRY, rank = "species",
                        traits = c("LA", "LMA", "LeafN", "PlantHeight", "SeedMass", "SSD"))

# Calculate phylogenetic signal using Lambda method
pnc(subtraits, BCI$phy_species, methods = "lambda")

# Calculate without PCA analysis
pnc(subtraits, BCI$phy_species, methods = "lambda", pca_axes = NULL)
```

---

pnc_robustness   *Test Robustness of Phylogenetic Niche Conservatism Analysis*

Description

This function evaluates the robustness of phylogenetic signal estimates by simulating trait data with
the same phylogenetic signal strength as observed, applying the original missing data pattern, and
testing how consistently the statistical significance is recovered across multiple simulations.

Usage

```
pnc_robustness(
  trait_data,
  phylo_tree,
  methods = "lambda",
  pca_axes = c("PC1", "PC2"),
  n_simulations = 100,
  alpha_level = 0.05,
  tolerance = 0.05
)
```

Arguments

trait_data	A data frame or matrix containing trait data with species as rows
phylo_tree	A phylogenetic tree object of class "phylo"
methods	Character string specifying method to use. Options: "lambda" or "K". Default is "lambda"
pca_axes	Character vector specifying which PCA axes to include (e.g., c("PC1", "PC2")). Default is c("PC1", "PC2")

n_simulations	Integer. Number of simulations to run for robustness testing. Default is 100
alpha_level	Numeric. Significance level for statistical testing. Default is 0.05
tolerance	Numeric. Acceptable difference between target and estimated signal values during trait simulation. Default is 0.05

Details

The robustness testing procedure involves:

1. Performing baseline phylogenetic signal analysis using pnc()

2. For each trait, simulating new trait data with the same phylogenetic signal strength as observed in the original data

3. Applying the exact missing data pattern from the original dataset to the simulated data

4. Re-testing phylogenetic signal on the simulated data and recording p-values

5. Calculating the percentage of simulations that maintain the same statistical significance conclusion (significant vs. non-significant)

The function uses simulate_lambda_trait() or simulate_K_trait() internally to generate trait data with target phylogenetic signal values.

For PCA axes, the missing data pattern corresponds to complete cases from the original trait matrix. For individual traits, the original missing pattern is preserved exactly.

Value

A data frame containing the original phylogenetic signal results with additional columns:

- robustness: Percentage of simulations that maintain the same statistical significance conclusion as the original analysis

- signal_sd: Standard deviation of phylogenetic signal values across successful simulations

Returns the enhanced results from the baseline pnc() analysis

Examples

```
# Load example data
data(BCI)
data(TRY)

# Extract trait data
sp <- colnames(BCI$com)
subtraits <- extract_traits(sp, TRY, rank = "species",
                            traits = c("LA", "LMA", "LeafN", "PlantHeight"))

# Test robustness of phylogenetic signal analysis
# This function's runtime is long
pnc_robustness(subtraits, BCI$phy_species, methods = "lambda", n_simulations = 5)
```

ReptTraits *ReptTraits: A Comprehensive Dataset of Ecological Traits in Reptiles*

Description

A comprehensive dataset containing ecological and morphological characteristics of reptiles. The dataset provides detailed information about reptile species, including elevation, seasonal precipitation, body mass, and reproductive features.

Usage

```
ReptTraits
```

Format

A data frame with the following columns:

species Scientific species name

genus Genus name

family Family name

Minimal_elevation Minimum elevation where the species was observed (meters above sea level)

Maximum_elevation Maximum elevation where the species was observed (meters above sea level)

Mean_Annual_Temperature Mean annual temperature, °C

Temperature_Seasonality Temperature seasonality, standard deviation × 100

Seasonality_Precipitation Seasonal precipitation information

Maximum_Longevity Longevity data are the maximum age reported for each species from the literature, years

Maximum_body_mass Maximum body mass of the species (grams)

Maximum_length Maximum length ("SVL", mm)/straight carapace length for turtles ("SCL", mm)

Mean_number_of_offspring Mean number of offspring or eggs per clutch

Smallest_clutch_size Minimum clutch/litter size

Largest_clutch_size Maximum clutch/litter size

Mean_Tb The mean reported mean body temperatures of animal, °C

References

Oskyrko, O., Mi, C., Meiri, S., & Du, W. (2024). ReptTraits: a comprehensive dataset of ecological traits in reptiles. Scientific Data, 11(1), 243. doi:10.1038/s41597024030795

Examples

```
data(ReptTraits)
head(ReptTraits)
```

simulate_K_trait   *Simulate Trait Data with Target Phylogenetic Signal (Blomberg's K)*

Description

This function generates trait data that matches a specified phylogenetic signal strength (Blomberg's K) through iterative simulation and testing.

Usage

```
simulate_K_trait(target_K, tree, max_attempts = 1e+05, tolerance = 0.02)
```

Arguments

target_K   Numeric. The desired phylogenetic signal strength (K value). - K = 0: No phylogenetic signal (star phylogeny) - K = 1: Expected signal under Brownian motion evolution - K > 1: Stronger phylogenetic signal than expected under Brownian motion - 0 < K < 1: Weaker phylogenetic signal than expected under Brownian motion

tree   An object of class "phylo". The phylogenetic tree for trait simulation.

max_attempts   Integer. Maximum number of simulation attempts before giving up. Default is 100000.

tolerance   Numeric. Acceptable difference between target and estimated K. Default is 0.02.

Details

The function works by:

1. Transforming the phylogenetic tree according to the target K value

2. Simulating trait data using phytools::fastBM() on the transformed tree

3. Estimating the phylogenetic signal using phytools::phylosig()

4. Repeating until the estimated K is within tolerance of the target

Tree transformation strategies: - When target_K = 0: Creates a star phylogeny using ape::stree() - When target_K = 1: Uses the original tree without transformation - When target_K > 1: Scales all branch lengths by the target K value - When 0 < target_K < 1: Interpolates between original tree and uniform branch lengths

Value

A data.frame with one column named 'trait' containing the simulated trait values. Row names correspond to tip labels from the phylogenetic tree. Returns NULL if the target K cannot be achieved within the specified tolerance and attempts.

Note

Blomberg's K measures the strength of phylogenetic signal relative to what would be expected under a Brownian motion model of evolution. Unlike Pagel's lambda, K can exceed 1, indicating stronger phylogenetic clustering than expected.

The function may take considerable time to converge for certain K values. Consider adjusting the tolerance parameter if convergence is slow.

Examples

```
# Generate a random tree
tree <- ape::rtree(50)

# Simulate trait with expected Brownian motion signal
trait_data <- simulate_K_trait(0.9, tree)

# Verify the phylogenetic signal
trait_vector <- setNames(trait_data$trait, rownames(trait_data))
phytools::phylosig(tree, trait_vector, method = "K", test = TRUE)
```

---

  simulate_lambda_trait   *Simulate Trait Data with Target Phylogenetic Signal (Lambda)*

Description

This function generates trait data that matches a specified phylogenetic signal strength (Pagel's lambda) through iterative simulation and testing.

Usage

```
simulate_lambda_trait(
  target_lambda,
  tree,
  max_attempts = 1e+05,
  tolerance = 0.02
)
```

Arguments

target_lambda	Numeric. The desired phylogenetic signal strength (lambda value). Should be between 0 and 1. - 0: No phylogenetic signal (star phylogeny) - 1: Full phylogenetic signal (Brownian motion)
tree	An object of class "phylo". The phylogenetic tree for trait simulation.
max_attempts	Integer. Maximum number of simulation attempts before giving up. Default is 100000.
tolerance	Numeric. Acceptable difference between target and estimated lambda. Default is 0.02.

Details

The function works by:

1. Transforming the phylogenetic tree according to the target lambda value using rescale()

2. Simulating trait data using fastBM() on the transformed tree

3. Estimating the phylogenetic signal using phylosig()

4. Repeating until the estimated lambda is within tolerance of the target

Special cases: - When target_lambda = 0: Sets internal branch lengths to 0, keeping only terminal branches - When target_lambda = 1: Uses the original tree without transformation

Value

A data.frame with one column named 'trait' containing the simulated trait values. Row names correspond to tip labels from the phylogenetic tree. Returns NULL if the target lambda cannot be achieved within the specified tolerance and attempts.

Note

The function may take considerable time to converge for certain lambda values, especially those close to intermediate values.

Consider adjusting the tolerance parameter if convergence is slow.

If 'target_lambda' is greater than 1, it will be automatically capped at 1, as lambda values typically range from 0 to 1.

Examples

```
# Generate a random tree
tree <- ape::rtree(50)

# Simulate trait with strong phylogenetic signal
trait_data <- simulate_lambda_trait(0.8, tree)

# Verify the phylogenetic signal
trait_vector <- setNames(trait_data$trait, rownames(trait_data))
phytools::phylosig(tree, trait_vector, method = "lambda", test = TRUE)
```

TRY   *TRY Plant Trait Database*

Description

A comprehensive global database of plant functional traits from the TRY initiative. This dataset contains standardized measurements of key plant functional traits across multiple species, genera, and families.

Usage

TRY

Format

A data frame with 58,964 rows and 23 variables:

species Character. Species name

genus Character. Genus name

family Character. Family name

DispersalUnitLength Numeric. Dispersal unit length, mm. (TraitID: 237)

LA Numeric. Leaf area (in case of compound leaves: leaflet, undefined if petiole is in- or excluded), mm2. (TraitID: 3113)

LDMC Numeric. Leaf dry mass per leaf fresh mass (leaf dry matter content, LDMC), g/g. (TraitID: 47)

LeafC Numeric. Leaf carbon (C) content per leaf dry mass, mg/g. (TraitID: 13)

LeafN Numeric. Leaf nitrogen (N) content per leaf dry mass, mg/g. (TraitID: 14)

LeafNPratio Numeric. Leaf nitrogen/phosphorus (N/P) ratio, g/g. (TraitID: 56)

LeafNperArea Numeric. Leaf nitrogen (N) content per leaf area, g m-2. (TraitID: 50)

LeafP Numeric. Leaf phosphorus (P) content per leaf dry mass, mg/g. (TraitID: 15)

Leafdelta15N Numeric. Leaf nitrogen (N) isotope signature (delta 15N), per mill. (TraitID: 78)

Leaffreshmass Numeric. Leaf fresh mass, g. (TraitID: 163)

LMA Numeric. Leaf mass per area. (1/SLA)

PlantHeight Numeric. Plant height vegetative, m. (TraitID: 3106)

RootingDepth Numeric. Root rooting depth, m. (TraitID: 6)

SeedLength Numeric. Seed length, mm. (TraitID: 27)

SeedMass Numeric. Seed dry mass, mg. (TraitID: 26)

SeedNumber Numeric. Seed number per reproduction unit, number. (TraitID: 138)

SLA Numeric. Leaf area per leaf dry mass (specific leaf area, SLA or 1/LMA): petiole excluded, mm2 mg-1. (TraitID: 3115)

SSD Numeric. Stem specific density (SSD, stem dry mass per stem fresh volume) or wood density, g/cm3. (TraitID: 4)

StemConduitDensity Numeric. Stem conduit density (vessels and tracheids), mm-2. (TraitID: 169)

WoodVesselLength Numeric. Wood vessel element length; stem conduit (vessel and tracheids) element length, micro m. (TraitID: 282)

Details

The TRY database represents a global effort to compile plant functional trait data from multiple sources and research groups. Plant functional traits are morphological, physiological, and phenological characteristics that influence fitness and ecosystem functioning. This dataset includes key traits related to:

- Leaf economics (SLA, LDMC, leaf nutrients)
- Plant architecture (height, rooting depth)
- Reproductive strategy (seed mass, seed number)
- Wood anatomy (vessel length, conduit density)
- Chemical composition (C, N, P content)

Missing values (NA) are common in trait databases due to the difficulty of measuring all traits for all species.

Source

TRY Plant Trait Database (https://www.try-db.org/)

References

Kattge, J., Bönisch, G., Díaz, S., et al. (2020). TRY plant trait database – enhanced coverage and open access. Global Change Biology, 26(1), 119-188. doi:10.1111/gcb.14904

Examples

```
# Load the dataset
data(TRY)
```

Index