

Package: PNC (via r-universe)

September 28, 2026

Type Package

Title Evaluating Phylogeny as a Proxy for Ecological Similarity

Version 0.2.0

Date 2026-8-29

Maintainer Yan He <heyaneco@163.com>

Description Provides a trait-based workflow for evaluating whether phylogenetic relatedness is informative about similarity in measured quantitative traits within focal species pools and across multiple communities. Functions support trait data integration, taxon-specific trait extraction, coverage assessment, optional principal component analysis, and estimation of phylogenetic signal using Pagel's lambda or Blomberg's K. Curated quantitative trait datasets are included for plants, birds, mammals, reptiles, amphibians, and fishes. Paired simulations assess how observed patterns of missing trait data affect Pagel's lambda estimates and significance classifications for individual traits. Methods for quantifying phylogenetic signal are based on Pagel (1999) <doi:10.1038/44766>, Blomberg et al. (2003) <doi:10.1111/j.0014-3820.2003.tb00285.x>, and 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, geiger, phytools, stats, utils

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 <https://cran.r-universe.dev>
Date/Publication 2026-08-29 14:34:21 UTC
RemoteUrl <https://github.com/cran/PNC>
RemoteRef HEAD
RemoteSha 0b9df002a3ada6f557b55ea338c0fd1ba91105c2

Contents

AmphiBIO	2
AVONET	4
BCI	5
COMBINE	6
compnc	9
compnc_robustness	12
coverage	15
extract_traits	16
FishLife	18
HimalayanBirds	20
merge_dataset	21
pnc	23
pnc_robustness	26
ReptTraits	29
TRY	30
Index	33

AmphiBIO	<i>AmphiBIO Amphibian Trait Data</i>
----------	--------------------------------------

Description

A processed subset of the AmphiBIO database containing 12 quantitative morphological, life-history, and reproductive traits. The dataset contains one record for each of 6,776 amphibian species.

Usage

AmphiBIO

Format

A data frame with 6,776 rows and 15 variables:

- species** Character. Full binomial species name.
- genus** Character. Genus name.
- family** Character. Taxonomic family.

Body_mass_g Numeric. Maximum adult body mass, g.

Age_at_maturity_min_y Numeric. Minimum age at maturation or sexual maturity, years.

Age_at_maturity_max_y Numeric. Maximum age at maturation or sexual maturity, years.

Body_size_mm Numeric. Maximum adult body size, mm.

Size_at_maturity_min_mm Numeric. Minimum body size at maturation or sexual maturity, mm.

Size_at_maturity_max_mm Numeric. Maximum body size at maturation or sexual maturity, mm.

Longevity_max_y Numeric. Maximum lifespan, years.

Litter_size_min_n Numeric. Minimum number of offspring or eggs per clutch.

Litter_size_max_n Numeric. Maximum number of offspring or eggs per clutch.

Reproductive_output_y Numeric. Maximum number of reproductive events per year.

Offspring_size_min_mm Numeric. Minimum offspring or egg size, mm.

Offspring_size_max_mm Numeric. Maximum offspring or egg size, mm.

Details

The source AmphiBIO database contains 17 amphibian ecological traits. This processed dataset retains 12 numeric variables selected for use with PNC. These variables include continuous measurements and count variables.

Species names and higher taxonomy were retained from AmphiBIO. The source taxonomy follows Amphibian Species of the World version 5.5. Trait values and units were retained as provided in the source database. No additional unit conversions, transformations, or imputations were performed. Missing values are represented by NA.

Maximum body size is generally reported as snout-vent length for Anura and as total length for Gymnophiona and Caudata. For some Caudata records, only snout-vent length was available in the source database.

Offspring size generally refers to egg-yolk diameter, although some source studies measured total egg diameter including the surrounding jelly capsule. The measurement convention cannot always be distinguished from the processed data.

Trait coverage varies substantially among variables. In this dataset, 57 species have complete values for all 12 selected traits, whereas 796 species have no observed values for any of the selected traits. Missing-data coverage and the dependence between paired minimum and maximum variables should therefore be considered when selecting variables for multivariate analyses.

Source

AmphiBIO data repository: [doi:10.6084/m9.figshare.4644424](https://doi.org/10.6084/m9.figshare.4644424)

References

Oliveira, B. F., Sao-Pedro, V. A., Santos-Barrera, G., Penone, C., and Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*, 4, 170123. [doi:10.1038/sdata.2017.123](https://doi.org/10.1038/sdata.2017.123)

Examples

```
data(AmphiBIO)

dim(AmphiBIO)
head(AmphiBIO[, c(
  "species",
  "Body_size_mm",
  "Reproductive_output_y"
)])
```

 AVONET

Bird Morphological Traits from AVONET

Description

A species-level compilation of 11 quantitative morphological traits obtained from AVONET. The dataset also includes genus and family information based on BirdLife International taxonomy. Each row represents one bird species.

Usage

AVONET

Format

A data frame with 11,009 rows and 14 variables:

species Character. Scientific species name.

genus Character. Genus name.

family Character. Family name based on BirdLife International taxonomy.

Beak.Length_Culmen Numeric. Beak length measured along the culmen from the beak tip to the base of the skull, mm.

Beak.Length_Nares Numeric. Beak length from the anterior edge of the nares to the beak tip, mm.

Beak.Width Numeric. Beak width at the anterior edge of the nares, mm.

Beak.Depth Numeric. Beak depth at the anterior edge of the nares, mm.

Tarsus.Length Numeric. Tarsus length from the posterior notch between the tibia and tarsus to the end of the last scale of the acrotarsium, mm.

Wing.Length Numeric. Length from the carpal joint to the tip of the longest primary feather on the unflattened wing, mm.

Kipps.Distance Numeric. Distance between the tip of the first outermost secondary feather and the tip of the longest primary feather, mm. This value was measured directly or calculated as $\text{Wing.Length} - \text{Secondary1}$.

Secondary1 Numeric. Length from the carpal joint to the tip of the first outermost secondary feather, mm.

Hand-Wing.Index Numeric. Hand-wing index, calculated as $100 * \text{Kipps.Distance} / \text{Wing.Length}$; dimensionless.

Tail.Length Numeric. Distance from the tip of the longest rectrix to the point where the two central rectrices emerge from the skin, mm.

Mass Numeric. Species-level mean body mass across available individuals, including both sexes, g.

Details

Values are species-level summaries provided by AVONET. The four beak measurements describe feeding morphology, whereas measurements of the tarsus, wing, secondary feathers, and tail describe locomotor morphology. The hand-wing index measures wing elongation and is commonly used as an indicator of flight efficiency and dispersal ability. Body mass represents overall body size.

This PNC data object contains the 11 morphological traits and associated taxonomic information. The ecological and geographical variables included in the complete AVONET release are not included. Missing values are represented by NA.

Source

AVONET data repository (<https://figshare.com/s/b990722d72a26b5bf9ad>)

References

Tobias, J. A., Sheard, C., Pigot, A. L., et al. (2022). AVONET: morphological, ecological and geographical data for all birds. *Ecology Letters*, 25, 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 classifications.

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](https://doi.org/10.15146/5xcp0d46)

Examples

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

COMBINE

COMBINE Mammal Trait Data

Description

A processed subset of the COMBINE database containing 40 quantitative mammal traits and ecological attributes. The dataset contains one record for each of 6,234 extant or recently extinct mammal species.

Usage

COMBINE

Format

A data frame with 6,234 rows and 43 variables:

species Character. Full binomial species name.

genus Character. Genus name.

family Character. Taxonomic family.

adult_mass_g Numeric. Adult body mass, g.

adult_brain_mass_g Numeric. Adult brain mass, g.

adult_body_length_mm Numeric. Adult body length from the tip of the nose to the anus or base of the tail, mm.

adult_forearm_length_mm Numeric. Adult forearm length from the elbow to the wrist, mm. This variable applies to bats.

max_longevity_d Numeric. Maximum reported lifespan, days.

maturity_d Numeric. Age at sexual maturity, days.

female_maturity_d Numeric. Age at which a female reaches sexual maturity, days.

male_maturity_d Numeric. Age at which a male reaches sexual maturity, days.

age_first_reproduction_d Numeric. Age at which a female first gives birth or the young first attach to the teats, days.

gestation_length_d Numeric. Duration of gestation, days.

teat_number_n Numeric. Number of teats.

litter_size_n Numeric. Number of offspring per litter per female.

litters_per_year_n Numeric. Number of litters per female per year.

interbirth_interval_d Numeric. Time between successive reproductive events, days.

neonate_mass_g Numeric. Body mass at birth, g.

weaning_age_d Numeric. Age at which primary nutritional dependence on the mother ends, days.

weaning_mass_g Numeric. Body mass at weaning, g.

generation_length_d Numeric. Mean age of the parents of the current cohort, days.

dispersal_km Numeric. Distance between the natal location and the location of reproduction, km.

density_n_km2 Numeric. Population density, individuals per square kilometre.

home_range_km2 Numeric. Area used for the routine activities of an individual or social group, square kilometres.

social_group_n Numeric. Number of individuals in a social group.

dphy_invertebrate Numeric. Broad diet fraction composed of invertebrates, percent.

dphy Vertebrate Numeric. Broad diet fraction composed of vertebrates, percent.

dphy_plant Numeric. Broad diet fraction composed of plants or fungi, percent.

det_inv Numeric. Detailed diet fraction composed of invertebrates, percent.

det_vend Numeric. Detailed diet fraction composed of endothermic vertebrates, including mammals and birds, percent.

det_vect Numeric. Detailed diet fraction composed of ectothermic vertebrates, including reptiles and amphibians, percent.

det_vfish Numeric. Detailed diet fraction composed of fish, percent.

det_vunk Numeric. Detailed diet fraction composed of unspecified vertebrates, percent.

det_scav Numeric. Detailed diet fraction composed of scavenged material or carrion, percent.

det_fruit Numeric. Detailed diet fraction composed of fruits or drupes, percent.

det_nect Numeric. Detailed diet fraction composed of nectar, pollen, or plant exudates, percent.

det_seed Numeric. Detailed diet fraction composed of seeds, grains, nuts, or spores, percent.

det_plantother Numeric. Detailed diet fraction composed of other plant material, percent.

det_diet_breadth_n Numeric. Number of detailed EltonTraits diet categories that each contribute at least 20 percent of the diet.

upper_elevation_m Numeric. Upper known elevational limit, m.

lower_elevation_m Numeric. Lower known elevational limit, m.

altitude_breadth_m Numeric. Difference between the upper and lower elevational limits, m.

habitat_breadth_n Numeric. Number of distinct suitable level 1 IUCN habitat categories.

Details

Species names, trait values, and units follow the COMBINE database. This processed dataset contains one record per species. No additional unit conversions, trait transformations, or imputations were applied.

Source

COMBINE data repository: [doi:10.6084/m9.figshare.13028255.v4](https://doi.org/10.6084/m9.figshare.13028255.v4)

References

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

Users of individual traits should also consult and cite the relevant underlying data sources listed in the COMBINE data repository.

Examples

```
data(COMBINE)

dim(COMBINE)
head(COMBINE[, c("species", "adult_mass_g", "maturity_d")])
```

compnc

*Estimate Phylogenetic Signal Across Multiple Communities***Description**

‘compnc()’ estimates phylogenetic signal in quantitative traits across multiple ecological communities. The function identifies the species present in each community, matches them to a trait matrix and phylogenetic tree, and estimates Pagel’s lambda, Blomberg’s K, or both within each resulting local species pool.

Usage

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

Arguments

com	A numeric community matrix or data frame with communities as rows and species as columns. Cell values represent species abundance or occurrence and must be non-negative. Species names must be stored as unique column names. Row names are used as community names; when they are absent, names are generated automatically. Missing values are treated as absence.
trait_data	A data frame or matrix containing quantitative trait values, with species as rows and traits as columns. Species names must be stored as unique row names and are matched exactly to the species names in ‘com’ and the tip labels of ‘phylo_tree’. Trait columns must be numeric. Missing values may be ‘NA’; all other values must be finite.
phylo_tree	A phylogenetic tree object of class “phylo” with unique tip labels.
methods	Character vector specifying the phylogenetic signal metrics to calculate. Available options are “lambda” and “K”. The default is “lambda”.
pca_axes	Character vector naming the PC axes to include in the phylogenetic signal analysis. The default is ‘c(“PC1”, “PC2”)’. Set to ‘NULL’ to skip PCA.
sig_levels	Numeric vector of three increasing P-value thresholds used to assign significance symbols. The default is ‘c(0.001, 0.01, 0.05)’.
min_abundance	Minimum abundance required for a species to be treated as present in a community. A species is considered present when its value is greater than ‘min_abundance’. The default is 0.

<code>nsim</code>	Number of randomizations used to test Blomberg's K. This argument is ignored when only Pagel's lambda is requested. The default is 1000.
<code>verbose</code>	Logical. If 'TRUE', warnings about skipped or failed PCA and phylogenetic signal analyses are reported.

Details

When PC axes are requested, one principal component analysis (PCA) is fitted using the pooled set of eligible species across all communities. The resulting species scores are then used in every community, so that a given PC axis represents the same dimension of trait variation across communities.

Phylogenetic signal describes the tendency for related species to resemble one another in measured trait values. The community-specific estimates can be used to evaluate whether phylogenetic relatedness is informative about trait similarity within each local species pool and to compare this relationship among communities.

Individual traits are analyzed separately. Missing values in one trait therefore do not alter the species included in the analysis of another trait.

When PC axes are requested, the pooled PCA includes species that occur in at least one community, are represented in both 'trait_data' and 'phylo_tree', and have complete observations across all supplied traits. Trait values are centered and scaled before PCA. The PCA is fitted once, and the same species scores are used in every community.

The output reports both 'n_sp_in_plot' and trait-specific 'n_sp', allowing users to compare the original local pool with the species retained after trait and phylogenetic matching.

Phylogenetic signal is estimated only when at least four matched species have an observed value and the retained values show variation. Otherwise, 'signal' and 'p' are returned as 'NA'. The four-species threshold is a computational minimum used by the function, not a general recommendation for adequate sampling.

Community-specific estimates may be sensitive to the number and phylogenetic composition of the retained species. Results from small or phylogenetically restricted local pools should therefore be interpreted together with the reported 'n_sp' and the species retained in each analysis.

Value

A data frame with one row for each community-trait-method combination:

- 'plot': community name.
- 'trait': name of the individual trait or PC axis.
- 'coverage': formatted percentage of species present in the community with an available value. For an individual trait, this is trait availability before phylogenetic matching. For a PC axis, it is the percentage of species with a score from the pooled PCA.
- 'n_sp': number of species included after matching the trait or PC score to the phylogeny.
- 'signal': estimated Pagel's lambda or Blomberg's K.
- 'p': P-value for the phylogenetic signal estimate.
- 'significance': significance symbol assigned using 'sig_levels'.
- 'method': phylogenetic signal metric used.

- ‘n_sp_in_plot’: number of species present in the original local community according to ‘min_abundance’, before trait and phylogenetic matching.

The returned data frame also stores ‘methods’, ‘pca_axes’, ‘pca_failed’, ‘sig_levels’, ‘min_abundance’, ‘nsim’, ‘total_plots’, and ‘analyzed_plots’ as attributes. When PCA is successfully fitted, the pooled species scores and fitted model are stored in the ‘pca_results’ and ‘pca_model’ attributes, respectively.

References

- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877-884. doi:10.1038/44766
- Blomberg, S. P., Garland, T., Jr., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, 57, 717-745. doi:10.1111/j.0014-3820.2003.tb00285.x
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743-756. doi:10.1111/j.2041210X.2012.00196.x

Examples

```
data(HimalayanBirds)
data(AVONET)

sp <- colnames(HimalayanBirds$com)

subtraits <- extract_traits(
  sp,
  AVONET,
  rank = "species"
)

# Pagel's lambda using one pooled PCA across all communities
compnc(
  HimalayanBirds$com,
  subtraits,
  HimalayanBirds$phy_species,
  methods = "lambda"
)

# Pagel's lambda for individual traits only
compnc(
  HimalayanBirds$com,
  subtraits,
  HimalayanBirds$phy_species,
  methods = "lambda",
  pca_axes = NULL
)
```

compnc_robustness	<i>Assess Sensitivity of Pagel's Lambda Inference to Missing Trait Data Across Multiple Communities</i>
-------------------	---

Description

'compnc_robustness()' uses paired simulations to assess how observed patterns of missing data for individual quantitative traits affect Pagel's lambda estimates and their significance classifications across multiple ecological communities.

Usage

```
compnc_robustness(
  com,
  trait_data,
  phylo_tree,
  min_abundance = 0,
  n_simulations = 100,
  alpha_level = 0.05,
  verbose = TRUE
)
```

Arguments

com	A numeric community matrix or data frame with communities as rows and species as columns. Cell values represent species abundance or occurrence and must be non-negative. Species names must be stored as unique column names. Row names are used as community names; when they are absent, names are generated automatically. Missing values are treated as absence.
trait_data	A data frame or matrix containing quantitative trait values, with species as rows and traits as columns. Species names must be stored as unique row names and are matched exactly to the species names in 'com' and the tip labels of 'phylo_tree'. Trait columns must be numeric. Missing values may be 'NA'; all other values must be finite.
phylo_tree	A phylogenetic tree object of class '"phylo"' with finite, non-negative branch lengths and unique tip labels.
min_abundance	Minimum abundance required for a species to be treated as present in a community. A species is considered present when its value is greater than 'min_abundance'. The default is 0.
n_simulations	Number of paired simulations performed for each combination of community and individual trait. The default is 100.
alpha_level	P-value threshold used to classify the complete and incomplete analyses as statistically significant or non-significant. The default is 0.05.
verbose	Logical. If 'TRUE', a progress bar and warnings are shown. The default is 'TRUE'.

Details

For each combination of community and trait, the observed lambda is first estimated using `'compnc()'`. A local reference phylogeny is then constructed from species present in the community and represented in both the trait data and phylogeny. Complete trait values are simulated under Brownian motion on this phylogeny after it has been transformed according to the observed lambda.

Each simulated realization is analyzed twice: first using the complete local reference species pool and then after applying the observed missing data pattern of the focal trait. Because both analyses use the same simulated realization, their paired difference measures the change associated with applying that missing data pattern within the simulation.

The analysis is performed separately for each community and individual trait. For a given community, the local reference species pool consists of species that are present according to `'min_abundance'` and represented in both `'trait_data'` and `'phylo_tree'`. Species without an observed value for the focal trait remain in the complete simulated dataset and are excluded only when the observed missing data pattern is applied.

The output includes `'n_sp_in_plot'` and trait-specific `'n_sp'`, allowing users to compare the original local pool with the species retained after trait and phylogenetic matching.

For each combination of community and trait, the simulation is conditional on the Pagel's lambda estimated from the observed data. The local reference phylogeny is transformed according to this lambda, and complete trait values are simulated under Brownian motion on the transformed tree. Simulated lambda estimates are not forced to equal the observed value, so stochastic variation among simulated realizations is retained.

Within each simulation, the complete and incomplete analyses use identical trait values for species retained in both datasets. Their paired difference therefore compares the two data conditions within the same simulated realization rather than between independently simulated datasets.

The maximum valid lambda depends on the branch length structure of the local reference phylogeny. An observed lambda estimated from a smaller, trait-specific tree may occasionally exceed the maximum value that can be applied to the larger local reference phylogeny. In this case, the largest valid lambda for the local reference phylogeny is used for simulation, while the observed lambda reported in the output remains unchanged.

Sensitivity metrics are returned as `'NA'` with `'n_successful = 0'` when the observed lambda cannot be estimated, fewer than four matched species have observed values for the focal trait, the retained trait values show no variation, or a valid transformed local reference phylogeny cannot be constructed. The four-species threshold is a computational minimum, not a general recommendation for adequate sampling.

The current robustness assessment is applied to individual traits only; PC axes are not included because their paired comparison would require a separate multivariate simulation framework.

For repeated lambda estimation, the function uses cached likelihood calculations for the local phylogenies and falls back to `'phytools::phylosig()'` when necessary.

The resulting assessment is conditional on the observed lambda and the Brownian motion simulation model.

Value

A data frame with one row for each combination of community and individual trait. It contains the observed Pagel's lambda results from `'compnc()'` and the simulation based sensitivity metrics:

- ‘plot’: community name.
- ‘trait’: trait name.
- ‘coverage’: formatted percentage of species present in the community with an observed value for the focal trait, before phylogenetic matching.
- ‘n_sp’: number of species included in the observed Pagel’s lambda analysis after trait and phylogenetic matching.
- ‘signal’: Pagel’s lambda estimated from the observed data.
- ‘p’: P-value for the observed lambda estimate.
- ‘significance’: significance symbol for the observed estimate.
- ‘method’: phylogenetic signal metric, always “lambda”.
- ‘n_sp_in_plot’: number of species present in the original local community according to ‘min_abundance’, before trait and phylogenetic matching.
- ‘simulation_lambda’: lambda used to transform the local reference phylogeny for simulation. This normally equals the observed lambda but may be reduced when the observed value exceeds the maximum supported by the local reference phylogeny.
- ‘consistency’: formatted percentage of successful paired simulations in which the complete and incomplete analyses produced the same significance classification at ‘alpha_level’.
- ‘signal_bias’: mean paired difference ‘lambda_missing - lambda_complete’. Positive values indicate that applying the missing data pattern produced a higher lambda estimate on average, whereas negative values indicate a lower estimate.
- ‘signal_sd’: standard deviation of the paired lambda differences. This is ‘NA’ when fewer than two paired simulations are successful.
- ‘n_successful’: number of simulations in which both the complete and incomplete analyses produced finite lambda estimates and P-values.

The returned data frame stores ‘min_abundance’, ‘n_simulations’, and ‘alpha_level’ as attributes.

References

- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877-884. doi:10.1038/44766
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743-756. doi:10.1111/j.2041210X.2012.00196.x

Examples

```
data(HimalayanBirds)
data(AVONET)

sp <- colnames(HimalayanBirds$com)

subtraits <- extract_traits(
  sp,
  AVONET,
  rank = "species"
```

```

)

set.seed(123)

compnc_robustness(
  HimalayanBirds$com,
  subtraits,
  HimalayanBirds$phy_species,
  n_simulations = 10
)

```

coverage

Summarize Trait Coverage

Description

‘coverage()’ summarizes missingness in a trait dataset for individual traits, complete cases across all supplied traits, and all trait values combined.

Usage

```
coverage(data)
```

Arguments

data	A data frame or matrix with observations, typically species or other taxa, in rows and traits in columns. All supplied columns are treated as traits. Species or taxon identifiers should therefore be stored as row names or removed before calling the function.
------	--

Details

Coverage is summarized in three ways:

- Trait-specific coverage is the percentage of observations with a non-missing value for each trait.
- Coverage of complete cases is the percentage of observations with non-missing values for every supplied trait and is reported in the “All” row.
- Overall coverage is the percentage of all cells in the supplied dataset that contain non-missing values and is displayed when the function runs.

Missingness is defined using ‘is.na()’. The function summarizes data availability but does not evaluate whether non-missing values are biologically valid for a particular trait.

Value

A data frame with one row for each supplied trait and an additional "All" row. It contains the following columns:

Trait Trait name. The "All" row summarizes complete cases across all supplied traits.

Available_count For an individual trait, the number of observations with a non-missing value. For the "All" row, the number of observations with complete data for all traits.

Missing_count For an individual trait, the number of observations with a missing value. For the "All" row, the number of observations missing at least one trait.

Trait_coverage_rate The percentage of available observations, returned as a character string containing a percent sign.

Overall coverage across all cells is displayed as a message but is not included in the returned data frame.

Examples

```
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)
)

coverage(trait_data)
```

extract_traits

Extract Quantitative Traits for Specified Taxa

Description

'extract_traits()' retrieves selected quantitative traits for specified species, genera, or families.

Usage

```
extract_traits(
  sp.list,
  dataset,
  rank = "species",
  traits = NULL,
  within_species = c("error", "mean", "median"),
  among_species = c("mean", "median")
)
```

Arguments

<code>sp.list</code>	A non-empty character vector containing unique taxon names to extract. Names must correspond to the taxonomic rank specified by <code>'rank'</code> .
<code>dataset</code>	A non-empty data frame or tibble containing taxonomic information and quantitative trait data. A <code>"species"</code> column is required for all analyses. A <code>"genus"</code> or <code>"family"</code> column is additionally required when extracting traits at the corresponding rank.
<code>rank</code>	Taxonomic rank used to match the requested taxa. Must be one of <code>"species"</code> , <code>"genus"</code> , or <code>"family"</code> . The default is <code>"species"</code> .
<code>traits</code>	Character vector specifying the traits to extract. If <code>'NULL'</code> , all numeric columns other than <code>"species"</code> , <code>"genus"</code> , and <code>"family"</code> are extracted.
<code>within_species</code>	Rule used when a species has multiple different finite records for the same trait. Must be <code>"error"</code> , <code>"mean"</code> , or <code>"median"</code> . The default, <code>"error"</code> , requires the user to resolve the records or explicitly select a summary rule.
<code>among_species</code>	Rule used to aggregate species-level values within a genus or family. Must be <code>"mean"</code> or <code>"median"</code> . The default is <code>"mean"</code> . This argument is ignored when <code>'rank = "species"'</code> .

Details

Repeated records are first resolved separately for each species and trait. For genus and family extraction, the resulting species-level values are then aggregated so that each contributing species receives equal weight, regardless of its number of records in the original dataset.

Taxon names are matched exactly against the column corresponding to `'rank'`. The function does not standardize taxonomic names, spelling, capitalization, or synonyms.

Only numeric traits are extracted. Requested traits that are absent or non-numeric are ignored with a warning, and the function stops if no requested numeric traits are available.

For each requested trait, only records belonging to the requested taxa are examined. Non-finite numeric values (`'Inf'`, `'-Inf'`, and `'NaN'`) in these records are treated as missing and reported in a warning. Non-finite values in unrequested taxa or traits do not affect the extraction.

Within each species and trait, missing and non-finite records are excluded. If no finite values remain, `'NA'` is returned. A single finite value, or multiple identical finite values, is returned directly. Multiple different finite values either trigger an error or are summarized using the mean or median, according to `'within_species'`.

For genus and family extraction, the resolved species-level values are aggregated separately for each trait using `'among_species'`. Species without a finite value for the focal trait are excluded from that trait's aggregation. Each contributing species receives equal weight.

The resulting genus or family value represents the central tendency of the contributing species. When interpreting this aggregate, users should also consider the number of contributing species and the variation among their trait values.

Each species must be assigned to only one requested genus or family. Conflicting taxonomic assignments cause the function to stop. Values are not inferred from higher taxonomic levels.

Trait-specific validity constraints are not evaluated automatically. For example, the function does not assume that zero is invalid because zero or negative values may be valid for some quantitative traits.

Value

A data frame with one row for each requested taxon, in the order supplied in ‘sp.list’, and one column for each retained numeric trait. Taxon names are stored as row names. Taxa not found in ‘dataset’ are retained and returned with missing values for all traits.

Examples

```
trait_data <- data.frame(
  species = c("Species a", "Species a", "Species b", "Species c"),
  genus = c("Genus1", "Genus1", "Genus1", "Genus2"),
  family = c("Family1", "Family1", "Family1", "Family2"),
  Trait1 = c(1, 3, 5, 2),
  Trait2 = c(4, 4, 6, NA)
)

# Resolve repeated records using their mean
extract_traits(
  c("Species a", "Species c"),
  trait_data,
  traits = c("Trait1", "Trait2"),
  within_species = "mean"
)

# Calculate genus-level means from resolved species-level values
extract_traits(
  c("Genus1", "Genus2"),
  trait_data,
  rank = "genus",
  traits = c("Trait1", "Trait2"),
  within_species = "mean",
  among_species = "mean"
)
```

FishLife	<i>Observed Fish Functional Traits from FishLife</i>
----------	--

Description

A processed species-level dataset containing 17 quantitative fish functional traits extracted from the observed-data component of FishBase_and_Morphometrics in FishLife. The dataset contains one taxonomic record for each of 26,622 fish species.

Usage

FishLife

Format

A data frame with 26,622 rows and 20 variables:

species Character. Full binomial species name.

genus Character. Genus name.

family Character. Taxonomic family.

age_max Numeric. Maximum reported age, years.

trophic_level Numeric. Trophic level, dimensionless.

aspect_ratio Numeric. Caudal-fin aspect ratio, calculated as squared caudal-fin height divided by caudal-fin surface area, dimensionless.

fecundity Numeric. Absolute fecundity, expressed as the number of eggs or offspring produced per female per year or per spawning batch, depending on the source record.

growth_coefficient Numeric. von Bertalanffy growth coefficient K , year^{-1} .

temperature Numeric. Mean environmental temperature associated with the sampled population or stock, degrees Celsius.

length_max Numeric. Maximum recorded body length, cm.

length_infinity Numeric. von Bertalanffy asymptotic body length, cm.

length_maturity Numeric. Body length at sexual maturity, generally corresponding to the length at which 50 percent of individuals are mature, cm.

age_maturity Numeric. Age at sexual maturity, generally corresponding to the age at which 50 percent of individuals are mature, years.

natural_mortality Numeric. Instantaneous natural mortality rate M , year^{-1} .

weight_infinity Numeric. von Bertalanffy asymptotic body mass, g.

max_body_depth Numeric. Size-standardized maximum body depth, dimensionless.

max_body_width Numeric. Size-standardized maximum body width, dimensionless.

lower_jaw_length Numeric. Size-standardized lower-jaw length, dimensionless.

min_caudal_peduncle_depth Numeric. Size-standardized minimum caudal-peduncle depth, dimensionless. The variable name retains the spelling used in the source dataset.

offspring_size Numeric. Offspring size, measured as egg diameter, egg-case length, or newborn body length depending on taxon, mm.

Details

This dataset was derived from the Y_{ij} component of `FishBase_and_Morphometrics` in FishLife version 3.0.1. Y_{ij} contains observed trait records in transformed space and is distinct from the model-predicted and imputed trait values available elsewhere in the FishLife object.

Variables stored on the natural-log scale were back-transformed using `exp()`. The source data contained 41,546 observation-level records representing 26,622 unique species. When multiple finite observations were available for the same species and trait, they were summarized using the median. No model-based imputation was applied.

Of the 26,622 species, 16,237 have at least one observed value among the 17 selected traits, whereas 10,385 contain taxonomic information but no observed value for any of these traits. Only 46 species

have complete observations for all 17 traits. Users should therefore examine trait coverage and select biologically appropriate trait subsets before multivariate analyses.

The four specimen-based morphometric traits (max_body_depth, max_body_width, lower_jaw_length, and min_caudal_peduncle_depth) describe relative body shape. In the source analysis, each linear measurement was standardized by the geometric mean of specimen length, width, and height. These variables are therefore dimensionless.

Maximum body length is generally reported as total length, although some source records use standard length, fork length, disc width, or another reported measure of body size.

Fecundity is not consistently standardized to annual reproductive output. Depending on the source record, values may represent annual production for single-spawning species or production per spawning batch for species that reproduce multiple times within a year.

Offspring size represents egg diameter for most fishes, egg-case length for skates and rays, or body length of a newborn pup for sharks, expressed in millimetres.

Source

FishLife version 3.0.1: [doi:10.5281/zenodo.7590994](https://doi.org/10.5281/zenodo.7590994)

References

Thorson, J. T., Maureaud, A. A., Frelat, R., Merigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., and 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, 1259–1275. [doi:10.1111/2041210X.14076](https://doi.org/10.1111/2041210X.14076)

Examples

```
data(FishLife)

dim(FishLife)
head(FishLife[, c(
  "species",
  "length_max",
  "offspring_size"
)])
```

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 relationships 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](https://doi.org/10.1111/ecog.05660)

Examples

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

merge_dataset

Merge Two Trait Datasets by Species

Description

‘merge_dataset()’ combines two species-level trait datasets using the “species” column as the matching key. Species and variables occurring in either dataset are retained, and overlapping values are resolved using a user-selected priority rule.

Usage

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

Arguments

<code>main_data</code>	A data frame containing the primary dataset. It must include a <code>"species"</code> column containing non-missing, non-empty species names.
<code>additional_data</code>	A data frame containing the additional dataset. It must include a <code>"species"</code> column containing non-missing, non-empty species names.
<code>priority</code>	Character string specifying how overlapping non-missing values are resolved. Available options are: <code>"main"</code> retains the value from <code>'main_data'</code>. This is the default. <code>"additional"</code> retains the value from <code>'additional_data'</code>. <code>"mean"</code> uses the arithmetic mean of the two values when the corresponding columns in both datasets are numeric. For non-numeric columns, it retains the value from <code>'main_data'</code>.

Details

Species names are matched exactly. The function does not standardize taxonomic names, spelling, capitalization, or synonyms.

Each input dataset is expected to contain no more than one row per species. If duplicate species are present, the function issues a warning and uses only the first occurrence of each species in that dataset.

All columns other than `"species"` are merged using the selected priority rule, including trait variables and taxonomic metadata such as genus or family.

For a given species and variable, a non-missing value is retained when the corresponding value in the other dataset is missing. The selected priority rule is used only when both datasets contain non-missing values.

With `priority = "mean"`, averaging is performed only when the corresponding columns in both datasets are numeric. The two source values receive equal weight. When either value is missing, the available value is retained. For non-numeric columns, `priority = "mean"` has the same effect as `priority = "main"`.

Factor columns are converted to character columns before merging. Variable names should be unique within each input dataset because duplicated column names are not resolved by the function.

Value

A data frame containing all unique species and all variables present in either input dataset. The `"species"` column is placed first. Species retain their order from `'main_data'`, followed by species occurring only in `'additional_data'`. Other columns retain their order from `'main_data'`, followed by columns occurring only in `'additional_data'`.

Examples

```
main_data <- data.frame(
  species = c("Species a", "Species b"),
  genus = c("Genus1", "Genus2"),
  Trait1 = c(1, NA),
```

```

    Trait2 = c(4, 6)
  )

  additional_data <- data.frame(
    species = c("Species a", "Species c"),
    genus = c("Genus1", "Genus3"),
    Trait1 = c(3, 5),
    Trait3 = c(8, 9)
  )

  # Prioritize values from the main dataset
  merge_dataset(
    main_data,
    additional_data
  )

  # Average overlapping numeric values
  merge_dataset(
    main_data,
    additional_data,
    priority = "mean"
  )

```

pnc

*Estimate Phylogenetic Signal in Quantitative Traits***Description**

‘pnc()’ estimates phylogenetic signal in quantitative traits for a focal set of taxa. The function matches trait data to a phylogenetic tree and estimates Pagel’s lambda, Blomberg’s K, or both for individual traits and, when requested, principal component (PC) axes.

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 quantitative trait values, with taxa as rows and traits as columns. Taxon names must be stored as unique row names and

	are matched exactly to the tip labels of 'phylo_tree'. Trait columns must be numeric. Missing values may be 'NA'; all other values must be finite.
phylo_tree	A phylogenetic tree object of class "phylo" with unique tip labels.
methods	Character vector specifying the phylogenetic signal metrics to calculate. Available options are "lambda" and "K". The default is "lambda".
pca_axes	Character vector naming the PC axes to include in the phylogenetic signal analysis. The default is 'c("PC1", "PC2")'. Set to 'NULL' to skip PCA.
sig_levels	Numeric vector of three increasing P-value thresholds used to assign significance symbols. The default is 'c(0.001, 0.01, 0.05)'.
nsim	Number of randomizations used to test Blomberg's K. This argument is ignored when only Pagel's lambda is requested. The default is 1000.
verbose	Logical. If 'TRUE', warnings about skipped or failed PCA and phylogenetic signal analyses are reported.

Details

Phylogenetic signal describes the tendency for related taxa to resemble one another in measured trait values. The resulting estimates can be used to evaluate whether phylogenetic relatedness is informative about trait similarity in the analyzed taxon pool. Their ecological interpretation depends on the traits supplied and the taxa retained in the analysis.

Individual traits are analyzed separately. Missing values in one trait therefore do not alter the taxa included in the analysis of another trait.

When PC axes are requested, PCA is performed once using taxa represented in both 'trait_data' and 'phylo_tree' and having complete observations across all supplied traits. Trait values are centered and scaled before PCA. The requested PC scores are then analyzed in the same way as individual traits.

PCA is skipped if fewer than four matched taxa have complete trait data or if the PCA cannot be fitted. This does not prevent the individual traits from being analyzed.

For each individual trait or PC axis, phylogenetic signal is estimated only when at least four taxa have both an observed value and a matching phylogenetic tip, and when the retained values show variation. Otherwise, 'signal' and 'p' are returned as 'NA'. The four-taxon threshold is a computational minimum used by the function, not a general recommendation for adequate sampling.

Value

A data frame with one row for each trait-method combination:

- 'trait': name of the individual trait or PC axis.
- 'coverage': formatted percentage of input taxa with an available value. For an individual trait, this is trait availability before phylogenetic matching. For a PC axis, it represents complete-case coverage among taxa matched to the phylogeny.
- 'n_sp': number of taxa included after matching non-missing trait values to the phylogeny.
- 'signal': estimated Pagel's lambda or Blomberg's K.
- 'p': P-value for the phylogenetic signal estimate.
- 'significance': significance symbol assigned using 'sig_levels'.

- ‘method’: phylogenetic signal metric used.

The returned data frame also stores ‘methods’, ‘pca_axes’, ‘pca_failed’, ‘sig_levels’, and ‘nsim’ as attributes. When PCA is successfully fitted, the PCA scores and fitted model are stored in the ‘pca_results’ and ‘pca_model’ attributes, respectively.

References

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877-884. doi:[10.1038/44766](https://doi.org/10.1038/44766)

Blomberg, S. P., Garland, T., Jr., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, 57, 717-745. doi:[10.1111/j.0014-3820.2003.tb00285.x](https://doi.org/10.1111/j.0014-3820.2003.tb00285.x)

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

Examples

```
data(BCI)
data(TRY)

sp <- colnames(BCI$com)

subtraits <- extract_traits(
  sp,
  TRY,
  rank = "species",
  traits = c(
    "LA",
    "LMA",
    "LeafN",
    "PlantHeight",
    "SeedMass",
    "SSD"
  )
)

# Pagel's lambda for individual traits and PC1 and PC2
pnc(
  subtraits,
  BCI$phy_species,
  methods = "lambda"
)

# Pagel's lambda for individual traits only
pnc(
  subtraits,
  BCI$phy_species,
  methods = "lambda",
  pca_axes = NULL
)
```

)

pnc_robustness

*Assess Sensitivity of Pagel's Lambda Inference to Missing Trait Data***Description**

'pnc_robustness()' uses paired simulations to assess how the observed pattern of missing data for each individual quantitative trait affects Pagel's lambda estimates and their significance classifications within a focal species pool.

Usage

```
pnc_robustness(
  trait_data,
  phylo_tree,
  n_simulations = 100,
  alpha_level = 0.05,
  verbose = TRUE
)
```

Arguments

trait_data	A data frame or matrix containing quantitative trait values, with species as rows and traits as columns. Species names must be stored as unique row names and are matched exactly to the tip labels of 'phylo_tree'. Trait columns must be numeric. Missing values may be 'NA'; all other values must be finite.
phylo_tree	A phylogenetic tree object of class "phylo" with finite, non-negative branch lengths and unique tip labels.
n_simulations	Number of paired simulations performed for each individual trait. The default is 100.
alpha_level	P-value threshold used to classify the complete and incomplete analyses as statistically significant or non-significant. The default is 0.05.
verbose	Logical. If 'TRUE', a progress bar and warnings are shown. The default is 'TRUE'.

Details

For each trait, the observed lambda is first estimated using 'pnc()'. The corresponding reference phylogeny is then transformed according to this estimate, and a complete set of trait values is simulated under Brownian motion. Each simulated realization is analyzed twice: first using all species in the reference pool and then after applying the observed missing data pattern of the focal trait.

Because the complete and incomplete analyses use the same simulated realization, their paired difference measures the change associated with applying the observed missing data pattern within that simulation.

The analysis is performed separately for each individual trait. The reference species pool consists of species represented in both 'trait_data' and 'phylo_tree'. Species without an observed value for the focal trait are retained in the complete simulated dataset and excluded only when the observed missing data pattern is applied.

For each trait, the simulation is conditional on the Pagel's lambda estimated from the observed data. The reference phylogeny is transformed according to this lambda, and complete trait values are simulated under Brownian motion on the transformed tree. Simulated lambda estimates are not forced to equal the observed value, so stochastic variation among simulated realizations is retained.

Within each simulation, the complete and incomplete analyses use identical trait values for the species retained in both datasets. Their paired difference therefore compares the two data conditions within the same simulated realization rather than between independently simulated datasets.

The maximum valid lambda depends on the branch length structure of the reference phylogeny. An observed lambda estimated from a smaller, trait-specific tree may occasionally exceed the maximum value that can be applied to the larger reference phylogeny. In this case, the largest valid lambda for the reference phylogeny is used for simulation, while the observed lambda estimate reported in the output remains unchanged.

Sensitivity metrics are returned as 'NA' with 'n_successful = 0' when the observed lambda cannot be estimated, fewer than four species have observed values for the focal trait, or a valid transformed reference phylogeny cannot be constructed.

The current robustness assessment is applied to individual traits only; PC axes are not included because their paired comparison would require a separate multivariate simulation framework.

For repeated lambda estimation, the function uses a cached likelihood calculation and falls back to 'phytools::phylosig()' when necessary.

The resulting assessment is conditional on the observed lambda and the Brownian motion simulation model.

Value

A data frame with one row for each individual trait. It contains the observed Pagel's lambda results from 'pnc()' and the simulation based sensitivity metrics:

- 'trait': trait name.
- 'coverage': formatted percentage of input species with an observed value for the focal trait.
- 'n_sp': number of species included in the observed Pagel's lambda analysis after trait and phylogenetic matching.
- 'signal': Pagel's lambda estimated from the observed data.
- 'p': P-value for the observed lambda estimate.
- 'significance': significance symbol for the observed estimate.
- 'method': phylogenetic signal metric, always "lambda".
- 'simulation_lambda': lambda used to transform the reference phylogeny for simulation. This normally equals the observed lambda but may be reduced when the observed value exceeds the maximum supported by the reference phylogeny.

- ‘consistency’: formatted percentage of successful paired simulations in which the complete and incomplete analyses produced the same significance classification at ‘alpha_level’.
- ‘signal_bias’: mean paired difference ‘lambda_missing - lambda_complete’. Positive values indicate that applying the missing data pattern produced a higher lambda estimate on average, whereas negative values indicate a lower estimate.
- ‘signal_sd’: standard deviation of the paired lambda differences. This is ‘NA’ when fewer than two paired simulations are successful.
- ‘n_successful’: number of simulations in which both the complete and incomplete analyses produced finite lambda estimates and P-values.

The returned data frame stores ‘n_simulations’ and ‘alpha_level’ as attributes.

References

- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877-884. doi:[10.1038/44766](https://doi.org/10.1038/44766)
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743-756. doi:[10.1111/j.2041210X.2012.00196.x](https://doi.org/10.1111/j.2041210X.2012.00196.x)

Examples

```
data(BCI)
data(TRY)

sp <- colnames(BCI$com)

subtraits <- extract_traits(
  sp,
  TRY,
  rank = "species",
  traits = c("LA", "LMA", "LeafN", "PlantHeight")
)

set.seed(123)

pnc_robustness(
  subtraits,
  BCI$phy_species,
  n_simulations = 10
)
```

ReptTraits

Reptile Traits and Environmental Attributes

Description

A processed subset of the ReptTraits database containing 12 quantitative traits and environmental attributes for 12,060 extant reptile species. The dataset contains one record per species.

Usage

ReptTraits

Format

A data frame with 12,060 rows and 15 variables:

species Character. Full binomial species name.

genus Character. Genus name.

family Character. Taxonomic family.

Minimal_elevation Numeric. Minimum elevation at which the species was observed, metres above sea level.

Maximum_elevation Numeric. Maximum elevation at which the species was observed, metres above sea level.

Mean_Annual_Temperature Numeric. Mean annual temperature, degrees Celsius.

Temperature_Seasonality Numeric. Temperature seasonality, expressed as the standard deviation multiplied by 100.

Seasonality_Precipitation Numeric. Precipitation seasonality, expressed as the coefficient of variation multiplied by 100.

Maximum_Longevity Numeric. Maximum age reported for the species in the literature, years.

Maximum_body_mass Numeric. Maximum body mass, g.

Maximum_length Numeric. Maximum snout-vent length, or straight carapace length for turtles, mm.

Mean_number_of_offspring Numeric. Mean number of offspring per litter or eggs per clutch.

Smallest_clutch_size Numeric. Lowest reported mean clutch or litter size.

Largest_clutch_size Numeric. Highest reported mean clutch or litter size.

Mean_Tb Numeric. Mean of the reported mean body temperatures, degrees Celsius.

Details

The source ReptTraits database contains 40 variables describing reptile taxonomy, ecology, morphology, life history, physiology, behaviour, and conservation status. This processed dataset retains five elevational or climatic attributes and seven morphological, life-history, or physiological variables for use with PNC.

Species names and higher taxonomy follow the Reptile Database taxonomy adopted by ReptTraits. Trait values and units were retained as provided in the source database. No additional unit conversions, transformations, or imputations were applied when preparing this PNC subset. Missing values are represented by NA, and coverage varies among variables.

The Maximum_length variable represents snout-vent length for most reptile groups but straight carapace length for turtles. The three clutch or litter-size variables provide related summaries of the same reproductive dimension and should not be treated as independent traits in multivariate analyses.

Source

ReptTraits data repository: [doi:10.6084/m9.figshare.24572683](https://doi.org/10.6084/m9.figshare.24572683)

References

Oskyrko, O., Mi, C., Meiri, S., and Du, W. (2024). ReptTraits: a comprehensive dataset of ecological traits in reptiles. *Scientific Data*, 11, 243. [doi:10.1038/s41597024030795](https://doi.org/10.1038/s41597024030795)

Examples

```
data(ReptTraits)

dim(ReptTraits)
head(
  ReptTraits[
    ,
    c("species", "Maximum_body_mass", "Mean_Tb")
  ]
)
```

TRY	<i>Plant Functional Traits from the TRY Database</i>
-----	--

Description

A species-level compilation of 20 quantitative plant functional traits obtained from the TRY Plant Trait Database. Taxonomic names were standardized using the **U.Taxonstand** package. Each row represents one species, while the genus and family columns provide its higher-taxonomic classification.

Usage

TRY

Format

A data frame with 58,975 rows and 23 variables:

species Character. Standardized species name.

genus Character. Standardized genus name.

family Character. Standardized family name.

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

LA Numeric. Leaf area, or leaflet area for compound leaves, mm². Inclusion of the petiole may vary among source records. (TraitID: 3113)

LDMC Numeric. Leaf dry mass per leaf fresh mass, or leaf dry matter content, g g⁻¹. (TraitID: 47)

LeafC Numeric. Leaf carbon content per unit leaf dry mass, mg g⁻¹. (TraitID: 13)

LeafN Numeric. Leaf nitrogen content per unit leaf dry mass, mg g⁻¹. (TraitID: 14)

LeafNPratio Numeric. Leaf nitrogen-to-phosphorus mass ratio, g g⁻¹ (dimensionless). (TraitID: 56)

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

LeafP Numeric. Leaf phosphorus content per unit leaf dry mass, mg g⁻¹. (TraitID: 15)

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

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

LMA Numeric. Leaf mass per area, calculated as the reciprocal of SLA, mg mm⁻².

PlantHeight Numeric. Vegetative plant height, m. (TraitID: 3106)

RootingDepth Numeric. Rooting depth, m. (TraitID: 6)

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

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

SeedNumber Numeric. Number of seeds per reproductive unit. (TraitID: 138)

SLA Numeric. Leaf area per unit leaf dry mass, or specific leaf area, with the petiole excluded, mm² mg⁻¹. (TraitID: 3115)

SSD Numeric. Stem specific density, defined as stem dry mass per unit fresh volume, or wood density, g cm⁻³. (TraitID: 4)

StemConduitDensity Numeric. Density of stem conduits, including vessels and tracheids, mm⁻². (TraitID: 169)

WoodVesselLength Numeric. Length of stem conduit elements, including vessel elements and tracheids, μm. (TraitID: 282)

Details

The dataset contains 19 traits obtained directly from TRY and one derived variable, LMA. The latter was calculated as the reciprocal of SLA and therefore has no separate TRY TraitID. Because SLA is expressed as $\text{mm}^2 \text{mg}^{-1}$, LMA is expressed as mg mm^{-2} in this dataset.

The traits describe major dimensions of leaf structure and chemistry, plant size and rooting strategy, seed and dispersal strategy, and stem and wood anatomy. Missing trait values are represented by NA.

Source

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

References

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

Zhang, J., and Qian, H. (2023). U.Taxonstand: An R package for standardizing scientific names of plants and animals. *Plant Diversity*, 45, 1–5. doi:10.1016/j.pld.2022.09.001

Examples

```
data("TRY")  
head(TRY)
```

Index

* datasets

- AmphiBIO, [2](#)
- AVONET, [4](#)
- BCI, [5](#)
- COMBINE, [6](#)
- FishLife, [18](#)
- HimalayanBirds, [20](#)
- ReptTraits, [29](#)
- TRY, [30](#)

AmphiBIO, [2](#)

AVONET, [4](#)

BCI, [5](#)

COMBINE, [6](#)

compnc, [9](#)

compnc_robustness, [12](#)

coverage, [15](#)

extract_traits, [16](#)

FishLife, [18](#)

HimalayanBirds, [20](#)

merge_dataset, [21](#)

pnc, [23](#)

pnc_robustness, [26](#)

ReptTraits, [29](#)

TRY, [30](#)