

Package ‘RRgeo’

April 1, 2025

Type Package

Title Species Distribution Modelling for Rare Species

Version 0.0.3

Description Performs species distribution modeling for rare species with unprecedented accuracy (Mondanaro et al., 2023 <[doi:10.1111/2041-210X.14066](https://doi.org/10.1111/2041-210X.14066)>) and finds the area of origin of species and past contact between them taking climatic variability in full consideration (Mondanaro et al., 2025 <[doi:10.1111/2041-210X.14478](https://doi.org/10.1111/2041-210X.14478)>).

License GPL-2

Encoding UTF-8

Depends R (>= 3.6.0)

Imports ape, methods,pbapply,Rphylopars,RRphylo,dismo (>= 1.3),gtools, terra, adehabitatMA, ecospat (>= 3.2.1), foreach, doParallel, PresenceAbsence, parallel, ade4, sp, sf,scales, ks, leastcostpath, doSNOW, biomod2 (>= 4.2),utils

Suggests ggplot2, cowplot, openxlsx, Bchron, curl, rnaturalearth (>= 1.0.1), rnaturalearthhires, spatstat.geom, spatstat.explore, knitr, httr, jsonlite

Config/Needs/check ropensci/rnaturalearthhires

Additional_repositories <http://packages.ropensci.org>

VignetteBuilder knitr

RoxygenNote 7.3.2

NeedsCompilation no

Author Alessandro Mondanaro [aut],
Mirko Di Febbraro [aut],
Silvia Castiglione [aut, cre],
Carmela Serio [aut],
Marina Melchionna [aut],
Giorgia Girardi [aut],
Pasquale Raia [aut]

Maintainer Silvia Castiglione <silvia.castiglione@unina.it>

Repository CRAN

Date/Publication 2025-04-01 16:20:10 UTC

Contents

cal14C	2
ENphylo_modeling	4
ENphylo_prediction	8
eucop_data_preparation	10
getENphylo_results	13
RRphylogeography	14
Index	17

cal14C	<i>Radiocarbon Calibration of Occurrences</i>
--------	---

Description

The function is meant to automatically apply the calibration process to conventional radiocarbon ages, relying on the package Bchron (Haslett & Parnell 2008). Specifically, the function internally selects the appropriate calibration curve based on the latitude associated with each occurrence and the nature of the sample (i.e marine or terrestrial samples).

Usage

```
cal14C(dataset,age=NULL,uncertainty=NULL,latitude=NULL,domain=NULL,
        bounds=c(0.025,0.975), clust=0.5, save=TRUE, output.dir=NULL)
```

Arguments

dataset	a data.frame containing all the occurrences to be calibrated.
age	character. Name of the column in dataset containing the conventional radiocarbon dates.
uncertainty	character. Name of the column in dataset containing the uncertainty associated to conventional radiocarbon dates.
latitude	character. Name of the column in dataset containing the latitude (in decimal degrees) of each occurrence.
domain	character. Name of the column in dataset indicating which occurrences are marine radiocarbon samples. If NULL, all the occurrences are assumed as "terrestrial" radiocarbon samples.
bounds	numeric. An upper and lower bound (in quantiles) to define the limits of the density probability created for each radiocarbon age (default: 95%).
clust	numeric. The proportion of cores used to train cal14C. If NULL, parallel computing is disabled.
save, output.dir	if save = TRUE, cal14C outputs are saved in output.dir.

Details

If dataset includes marine samples, the user should indicate it in the domain column by indicating "marine" for the corresponding occurrences. In this case, the function uses the marine20 curve to calibrate the related radiocarbon ages accordingly.

Value

The initial dataset with additional columns providing detailed calibration information for each occurrence. The new columns indicate the calibration curve used for each occurrence ("curve"), the calibrated radiocarbon ages ("cal_age"), and the values corresponding to the specified confidence limits derived from the density estimate of the calibrated radiocarbon ages. If `save=TRUE`, the dataframe is saved as xlsx file in `output.dir`.

Author(s)

Alessandro Mondanaro, Silvia Castiglione, Pasquale Raia

See Also

[BchronCalibrate](#)

Examples

```
library(RRgeo)

## Create an example dataset with 100 random radiocarbon ages and errors
set.seed(2025)
data.frame(age=round(runif(100,20000,50000),0),
           uncertain=round(runif(100,20,300),0),
           latitude=round(runif(100,-90,90),2))->data
data$domain<-"domain"
rep("marine",5)-> data[sample(nrow(data),5),"domain"]

cal14C(dataset=data,
       age="age",
       uncertainty = "uncertain",
       latitude = "latitude",
       domain<-"domain",
       clust=NULL,
       save= FALSE)->res
```

ENphylo_modeling	<i>Calculating species marginality and specialization via ENFA and phylogenetic imputation</i>
------------------	--

Description

The function computes vectors of marginality and specialization according to *Rinnan & Lawler (2019)* via Environmental Niche Factor Analysis (ENFA) and phylogenetic imputation (*Garland & Ives, 2000*). It takes a list of Simple Features (or `sf`) objects and a phylogenetic tree to train ENFA and/or ENphylo models. Both model techniques are calibrated and evaluated while accounting for phylogenetic uncertainty. Calibrations are made on a random subset of the data under the bootstrap cross-validation scheme. The predictive power of the different models is estimated using five different evaluation metrics.

Usage

```
ENphylo_modeling(input_data, tree, input_mask, obs_col, time_col=NULL,
  min_occ_enfa=30, boot_test_perc=20, boot_reps=10, swap.args= list(nsim=10,
  si=0.2, si2=0.2), eval.args=list(eval_metric_for_imputation="AUC",
  eval_threshold=0.7,output_options="best"),clust=0.5,output.dir)
```

Arguments

<code>input_data</code>	a list of <code>sf::data.frame</code> objects containing species occurrence data in binary format (ones for presence, zero for background points) along with the explanatory continuous variables to be used in ENFA or ENphylo. Each element of the list must be named using the names of the target species. Alternatively, ENFA model outputs generated through ENphylo_modeling can be supplied as named elements of <code>input_data</code> list.
<code>tree</code>	an object of class <code>phylo</code> including all the species listed in <code>input_data</code> . The tree needs not to be ultrametric or fully dichotomous. Any species in the tree that do not match species in <code>input_data</code> are automatically dropped from the tree.
<code>input_mask</code>	a <code>SpatRaster</code> object. It represents the geographical mask defining the spatial domain encompassing the background area enclosing all the species in <code>input_data</code> .
<code>obs_col</code>	character. Name of the <code>input_data</code> column containing the vector of species occurrence data in binary format.
<code>time_col</code>	character. Name of the <code>input_data</code> column containing the time intervals associated to each species presence and background point (optional).
<code>min_occ_enfa</code>	numeric. The minimum number of occurrence data required for a species to be modeled with ENFA.
<code>boot_test_perc</code>	numeric. Percentage of data (ranging between 0 and 100) used to calibrate ENFA and/or ENphylo models within a bootstrap cross-validation scheme. The remaining percentage ($100 - \text{boot_test_perc}$) will be used to evaluate model performances.

<code>boot_reps</code>	numeric. Number of evaluation runs performed within the bootstrap cross-validation scheme to evaluate ENFA and/or ENphylo models. If set to 0, models evaluation is skipped and the internal evaluation element returns NULL.
<code>swap.args</code>	list of ENphylo parameters. It includes: <ul style="list-style-type: none"> • <code>nsim</code> = number of alternative phylogenies generated by altering topology and branch lengths of the reference tree by means of <code>swapONE</code>. <code>nsim</code> must be greater than or equal to 1 (see details); • <code>si,si2</code> = arguments passed to <code>RRphylo::swapONE</code>.
<code>eval.args</code>	list of evaluation model parameters. It includes: <ul style="list-style-type: none"> • <code>eval_metric_for_imputation</code> = evaluation metric used to select the most accurate ENphylo models. The viable options are: "AUC", "TSS", "CBI", "SORENSEN", or "OMR"; • <code>eval_threshold</code> = the minimum evaluation score required to assess ENFA and ENphylo performance. ENFA models having <code>eval_metric_for_imputation</code> lower than <code>eval_threshold</code> are compared to ENphylo models to keep the one fitting best. Additionally, within ENphylo, models derived from the swapped trees having <code>eval_metric_for_imputation</code> lower than <code>eval_threshold</code> are excluded from the output; • <code>output_options</code> = the strategy adopted to return ENphylo models results (see details). The viable options are: "full", "weighted.mean", and "best".
<code>clust</code>	numeric. The proportion of cores used to train ENFA and ENphylo models. If NULL, parallel computing is disabled. It is set at 0.5 by default.
<code>output.dir</code>	the file path wherein ENphylo_modeling creates "ENphylo_enfa_models" and "ENphylo_imputed_models" folders to store modeling outputs (see details).

Details

ENphylo_modeling automatically arranges `input_data` in a suitable format to run ENFA or ENphylo. The internal call of the function is "calibrated_enfa" for ENFA and "calibrated_imputed" for ENphylo, respectively.

Phylogenetic uncertainty

The function does not work with `nsim < 1` since one of the strongest points of ENphylo_modeling is to test alternative phylogenies to provide the most accurate reconstruction of species environmental preferences. Similarly, setting `nsim = 1` limits the power of the function, as it will use the original tree without generating alternative phylogenies.

Phylogenetic Imputation

ENphylo_modeling automatically switches from ENFA to ENphylo algorithm for any species having less than `min_occ_enfa` occurrences or ENFA model accuracy below `eval_threshold`. In this latter case, the function performs both models and retains the one performing best according to `eval_metric_for_imputation`. Phylogenetic imputation is allowed for up to 30% of the species on the tree. If the number of species to impute exceeds 30%, ENphylo_modeling automatically splits the original tree into smaller subtrees, so that the maximum percentage of imputation is observed. Each subtree is designed to impute phylogenetically distant species and to retain species phylogenetically close to the taxa to be imputed (so that imputation is robust). In this case, the function prints the number of phylogenies used.

Outputs

If ENphylo_modeling runs the ENphylo algorithm, the outputs depend on the strategy adopted by the user through the output_options argument. If output_options="full", all CO matrices and evaluation metrics for all the swapped trees tested are returned. Under output_options="weighted.mean", the output consists of a subset of CO matrices and evaluation metrics for those tree swapping iterations achieving a predictive accuracy in terms of eval_metric_for_imputation above eval_threshold. Finally, if output_options="best", a single CO matrix and evaluation scores list corresponding to the most accurate swapped tree is returned. If any tree swapping iterations under either "best" or "weighted.mean" results in accuracy below the threshold, the function automatically switches to "full" strategy.

Eventually, the function creates two new folders, "ENphylo_enfa_models" and "ENphylo_imputed_models", in output.dir. In each of these folders, a number of new named subfolders equal to the number of modeled species are created. Therein, model outputs and background area are saved as model_outputs.RData and study_area.tif, respectively. model_outputs.RData includes a list of three elements, regardless of whether ENFA or ENphylo is used:

1. **\$call** a character specifying the algorithm used to model the species (i.e. ENFA or ENphylo).
2. **\$formatted_data** a list of input data formatted to run either ENFA or ENphylo algorithms. Specifically, the list reports: the presence data points (\$input_ones), the background points (\$input_back), the name of the columns associated to the arguments OBS_col and time_col (if specified), the name of the column containing the cell numbers (geoID_col), and the coordinates of presence data only (\$one_coords).
3. **\$calibrated_model** a list. The output objects are different depending on whether ENFA or ENphylo is used to model the species:

ENFA

- **\$call**: a character specifying the algorithm used.
- **\$full_model**: a list containing marginality and specialization factors, the 'co' matrix, the number of significant axes, and all the other objects generated by applying ENFA on the entire occurrence dataset (see *Rinnan et al. 2019* for additional details).
- **\$evaluation**: a matrix containing the evaluation scores of the ENFA model assessed by all possible evaluation metrics (i.e. Area Under the Curve (AUC), True Skill Statistic (TSS), Boyce Index (CBI), Sorensen Index, and Omission Rate (OMR)) for each model evaluations run.

ENphylo

- **\$call**: a character specifying the algorithm used.
- **\$co**: a list of the 'co' matrices of length equal to the number of alternative phylogenies tested (i.e. nsim argument). The number of 'co' matrices also reflects the selected output_option strategy.
- **\$evaluation**: a data.frame containing the evaluation scores of ENphylo model assessed by all possible evaluation metrics for each alternative phylogeny. The output of this object depends on the strategy adopted by the user through the output_options argument. Specifically, the function internally selects the model (or models) with the highest evaluation score according to the specified evaluation metric.
- **\$output_options**: a character vector including the argument output_options and eval_metric_for_imputation set to run the of ENphylo model.

Value

The function does not return the output into `.GlobalEnv`. Use the function [getENphylo_results](#) to collect results from local folders.

Author(s)

Alessandro Mondanaro, Mirko Di Febbraro, Silvia Castiglione, Carmela Serio, Marina Melchionna, Pasquale Raia

References

- Rinnan, D. S., & Lawler, J. (2019). Climate-niche factor analysis: a spatial approach to quantifying species vulnerability to climate change. *Ecography*, 42(9), 1494–1503. doi/full/10.1111/ecog.03937
- Garland, T., & Ives, A. R. (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist*, 155(3), 346–364. doi.org/10.1086/303327
- Mondanaro, A., Di Febbraro, M., Castiglione, S., Melchionna, M., Serio, C., Girardi, G., Blefiore, A.M., & Raia, P. (2023). ENphylo: A new method to model the distribution of extremely rare species. *Methods in Ecology and Evolution*, 14: 911-922. doi:10.1111/2041-210X.14066

See Also

[getENphylo_results](#); [ENphylo vignette](#)

Examples

```
library(ape)
library(terra)
library(sf)
library(RRgeo)

newwd<-tempdir()
# newwd<-"YOUR_DIRECTORY"
latesturl<-RRgeo::get_latest_version("12734585")
curl::curl_download(url = paste0(latesturl,"/files/dat.Rda?download=1"),
                    destfile = file.path(newwd,"dat.Rda"), quiet = FALSE)
load(file.path(newwd,"dat.Rda"))
read.tree(system.file("exdata/Eucopdata_tree.txt", package="RRgeo"))->tree
tree$tip.label<-gsub("_", " ", tree$tip.label)
curl::curl_download(paste0(latesturl,"/files/X35kya.tif?download=1"),
                    destfile = file.path(newwd,"X35kya.tif"), quiet = FALSE)
rast(file.path(newwd,"X35kya.tif"))->map35
project(map35,st_crs(dat[[1]])$proj4string,res = 50000)->map

ENphylo_modeling(input_data=dat[c(1,11)],
                 tree=tree,
                 input_mask=map[[1]],
                 obs_col="OBS",
                 time_col="age",
                 min_occ_enfa=15,
```

```

boot_test_perc=20,
boot_reps=10,
swap.args=list(nsim=5,si=0.2,si2=0.2),
eval.args=list(eval_metric_for_imputation="AUC",
               eval_threshold=0.7,
               output_options="best"),
clust=NULL,
output.dir=newwd)

```

ENphylo_prediction *Project the ENFA and ENphylo models into new geographical space and time interval*

Description

The function projects species marginality and specialization factors in different geographical areas and timescales. The function is able to convert marginality and specialization factors in habitat suitability values by using the Mahalanobis distances method.

Usage

```

ENphylo_prediction(object, newdata,
                  convert.to.suitability=FALSE,output.dir,proj_name="outputs")

```

Arguments

object	a list of ENFA and ENphylo models. Each element of the list must be named using the names of the modelled species.
newdata	a SpatRaster object including explanatory variables onto which ENFA or ENphylo models are to be projected. The list of variables must match the list used to model the species.
convert.to.suitability	logical. If TRUE, ENphylo_prediction projects ENFA or ENphylo model predictions in different geographical areas and timescales.
output.dir	the file path wherein ENphylo_prediction creates an "ENphylo_prediction" folder to store prediction outputs for each species.
proj_name	name of the subfolder created within the individual species folders to contain the ENphylo_prediction outputs.

Details

If `convert.to.suitability` is set as TRUE, ENphylo_prediction uses the function [mahasuhab](#) from the **adehabitatHS** R package (*Calenge, 2006*) to compute the habitat suitability map of the species over a given area. The conversion of Mahalanobis distances into probabilities follows the

chi-squared distribution. Specifically, we set the degree of freedom equal to n rather than $n-1$ following Etherington (2019). To convert habitat suitability values into binary presence/absence values, ENphylo_prediction relies on three different thresholding methods available in the function `optimal.thresholds` (Freeman & Moisen, 2008).

Value

The function stores all the results in a number of nested subfolders all contained in the "ENphylo_prediction" folder created in `output.dir`. This contains a subfolder for each individual species in object, in which a subfolder named according to `proj_name` contains all the outputs. Specifically, the function saves the predictions for marginality and specificity (more than one depending on the number of significant axes selected by `ENphylo_modeling`) in the new geographical areas along with the suitability and binarized maps. The latter are calculated by using the three different predefined thresholds: MaxSensSpec (i.e. maximize TSS), SensSpec (i.e. equalize sensitivity and specificity) and 10th percentile of predicted probability.

Author(s)

Alessandro Mondanaro, Mirko Di Febbraro, Silvia Castiglione, Carmela Serio, Marina Melchionna, Pasquale Raia

References

- Calenge, C. (2006) The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197, 516-519.
- Etherington, T. R. (2019). Mahalanobis distances and ecological niche modelling: correcting a chi-squared probability error. *PeerJ*, 7, e6678.
- Freeman, E. A. & Moisen, G. (2008). PresenceAbsence: An R Package for Presence-Absence Model Analysis. *Journal of Statistical Software*, 23(11):1-31.
- Mondanaro, A., Di Febbraro, M., Castiglione, S., Melchionna, M., Serio, C., Girardi, G., Blefiore, A.M., & Raia, P. (2023). ENphylo: A new method to model the distribution of extremely rare species. *Methods in Ecology and Evolution*, 14: 911-922. doi:10.1111/2041-210X.14066

See Also

`vignette("ENphylo", package = "RRgeo")`
[getENphylo_results](#); [ENphylo vignette](#)

Examples

```
library(ape)
library(terra)
library(sf)
library(RRgeo)

newwd<-tempdir()
# newwd<-"YOUR_DIRECTORY"

latesturl<-RRgeo:::get_latest_version("12734585")
```

```

curl::curl_download(url = paste0(latesturl,"/files/dat.Rda?download=1"),
                    destfile = file.path(newwd,"dat.Rda"), quiet = FALSE)
load(file.path(newwd,"dat.Rda"))
read.tree(system.file("exdata/Eucopdata_tree.txt", package="RRgeo"))->tree
tree$tip.label<-gsub("_", " ", tree$tip.label)
curl::curl_download(paste0(latesturl,"/files/X35kya.tif?download=1"),
                    destfile = file.path(newwd,"X35kya.tif"), quiet = FALSE)
rast(file.path(newwd,"X35kya.tif"))->map35
project(map35,st_crs(dat[[1]])$proj4string,res = 50000)->map

ENphylo_modeling(input_data=dat[c(1,11)],
                 tree=tree,
                 input_mask=map[[1]],
                 obs_col="OBS",
                 time_col="age",
                 min_occ_enfa=15,
                 boot_test_perc=20,
                 boot_reps=10,
                 swap.args=list(nsim=5,si=0.2,si2=0.2),
                 eval.args=list(eval_metric_for_imputation="AUC",
                                eval_threshold=0.7,
                                output_options="best"),
                 clust=NULL,
                 output.dir=newwd)

getENphylo_results(input.dir =newwd,
                   mods="all",
                   species_name=names(dat)[c(1,11)])->mod

library(rnaturalearth)
ne_countries(returnclass = "sf")->globalmap
subset(globalmap,continent=="North America")->ame_map

map35[[c("bio1","bio4","bio11","bio19")]]->newmap
crop(newmap,ext(ame_map))->newmap
project(newmap,st_crs(dat[[1]])$proj4string,res = 50000)->newmap

ENphylo_prediction(object = mod,
                   newdata = newmap,
                   convert.to.suitability = TRUE,
                   output.dir=newwd,
                   proj_name="proj_example")

```

Description

The function is meant to automatically import and preprocess fossil mammal occurrences and paleoclimatic/vegetational data available in EutherianCop dataset (Mondanaro et al., 2025). It also provides two distinct approaches, both implemented within a user-defined study area, for sampling a specified number of pseudoabsences or alternatively defining the background points. This flexibility enables users to assemble a list of `sf` objects that can be easily used to train ENFA, ENphylo or any other SDM algorithms of their choice.

Usage

```
eucop_data_preparation(input.dir, species_name, variables="all", which.vars=NULL,
  calibration=FALSE, add.modern.occs=FALSE,
  combine.ages=NULL, remove.duplicates=TRUE, bk_points=NULL, output.dir)
```

Arguments

<code>input.dir</code>	the file path wherein EutherianCop mammal occurrences and paleoclimatic data are to be stored.
<code>species_name</code>	character. The name of the single (or multiple) species used by <code>eucop_data_preparation</code> .
<code>variables</code>	character. The name of paleoclimatic simulations to be used. The viable options are "climveg", "bio", or "all".
<code>which.vars</code>	character vector indicating the name of the variables to be downloaded. The list of accepted names can be found [here](https://www.nature.com/articles/s41597-024-04181-4/tables/1).
<code>calibration</code>	logical. If TRUE, <code>eucop_data_preparation</code> performs the 14C calibration process to convert the conventional radiocarbon age estimates included in EutherianCop raw data file.
<code>add.modern.occs</code>	logical. If TRUE, <code>eucop_data_preparation</code> adds the modern records (if present) related to species in <code>species_name</code> .
<code>combine.ages</code>	one of "mean" or "median". The method to be used to aggregate multiple ages for each site or layer within the site.
<code>remove.duplicates</code>	logical. If TRUE, <code>eucop_data_preparation</code> removes duplicated record for each grid cell within a given time bin.
<code>bk_points</code>	a list including parameters to add background/pseudoabsence (i.e. absence) points (following the procedure described in Mondanaro et al. 2024). The list includes: <ul style="list-style-type: none"> • <code>buff</code>: the proportional distance to set a buffer around the minimum convex polygon that encompasses all occurrences of the target species. • <code>bk_strategy</code>: the strategy to add the absence points. It can be one of "background" or "pseudoabsence". • <code>bk_n</code>: number of absence points. If provided as an empty <code>list()</code> , the function automatically sets <code>buff = 0.1</code> , <code>bk_strategy="background"</code> , <code>bk_n=10000</code> .
<code>output.dir</code>	the file path wherein <code>eucop_data_preparation</code> stores the results.

Details

The `variables` argument allows the selection of climatic and environmental variables ("climveg"), bioclimatic variables ("bio"), or both sets of variables.

Through the `bk_strategy` argument, `eucop_data_preparation` offers two different approaches to generate absence points. The definition of the study area is the same for both methods. Under `bk_strategy = "background"`, the `bk_n` argument defines the maximum number of background points sampled from the study area within each time bin. Under `bk_strategy = "pseudoabsence"`, the `bk_n` argument represents the maximum number of pseudoabsence points across all time bins. This flexibility allows users to accommodate the different requirements for training the traditional envelope models (i.e. ENFA, ENphylo) and the common correlative or machine learning models (i.e. generalized linear model, MaxEnt, Random Forest).

Additionally, if `bk_points` is not NULL, the ages of presences and pseudoabsences or background points are forced to 1 kyr resolution according to the temporal resolution of the paleoclimatic/vegetational or bioclimatic data.

Value

`eucop_data_preparation` does not store any results in the global environment. Instead, a list of GeoPackage files, one per selected species, is saved in the directory specified by `output.dir`. The names of these files depend on the combination of arguments chosen by users: they include the suffix "cal/uncal" and "combined/multi" depending on whether calibration (`calibration`) and age aggregation (`combine.ages`) steps are performed. In any case, output files include information about ages, a column called "OBS" including species occurrence data in binary format, spatial geometry, and all the data information derived from EutherianCop dataset.

Author(s)

Alessandro Mondanaro, Silvia Castiglione, Pasquale Raia

References

Mondanaro, A., Di Febbraro, M., Castiglione, S., Belfiore, A. M., Girardi, G., Melchionna, M., Serio, C., Esposito, A., & Raia, P. (2024). Modelling reveals the effect of climate and land use change on Madagascar's chameleons fauna. *Communications Biology*, 7: 889. doi:10.1038/s42003-024-06597-5.

Mondanaro, A., Girardi, G., Castiglione, S., Timmermann, A., Zeller, E., Venugopal, T., Serio, C., Melchionna, M., Esposito, A., Di Febbraro, M., & Raia, P. (2025). EutherianCoP. An integrated biotic and climate database for conservation paleobiology based on eutherian mammals. *Scientific Data*, 12: 6. doi:10.1038/s41597-024-04181-4.

See Also

[eucop_data_preparation vignette](#)

Examples

```
newwd<-tempdir()
```

```
# newwd<-"YOUR_DIRECTORY"  
  
eucop_data_preparation(input.dir=newwd, species_name="Ursus ingressus",  
                        variables="bio",which.vars = "bio1", calibration=FALSE, combine.ages="mean",  
                        bk_points=NULL,output.dir=newwd)
```

getENphylo_results *Import ENphylo_modeling results into global environment*

Description

The function retrieves the ENFA/ENphylo models generated by [ENphylo_modeling](#) and arranges them into a named list to be used in [ENphylo_prediction](#) function. It also offers the option to choose between retrieving ENFA or ENphylo models and to select the model produced for one or more focal species.

Usage

```
getENphylo_results(input.dir,mods="all",species_name=NULL,only_evaluations=FALSE)
```

Arguments

input.dir	the file path wherein the folders "ENphylo_enfa_models" and "ENphylo_imputed_models" generated by ENphylo_modeling are stored.
mods	character. Name of the models to be retrieved. Viable options are: "enfa" (enfa models), "enphylo" (ENphylo models), "all" (default, enfa plus ENphylo models).
species_name	character. The name of the single (or multiple) species for which model results must be imported.
only_evaluations	logical. If TRUE, getENphylo_results returns model performances only.

Value

A named list of outputs as described in [ENphylo_modeling](#).

Author(s)

Alessandro Mondanaro, Silvia Castiglione, Pasquale Raia

See Also

[ENphylo_modeling](#); [getENphylo_results vignette](#)

Examples

```

library(ape)
library(terra)
library(sf)
library(RRgeo)

newwd<-tempdir()
# newwd<-"YOUR_DIRECTORY"
latesturl<-RRgeo:::get_latest_version("12734585")
curl::curl_download(url = paste0(latesturl,"/files/dat.Rda?download=1"),
  destfile = file.path(newwd,"dat.Rda"), quiet = FALSE)
load(file.path(newwd,"dat.Rda"))
read.tree(system.file("exdata/Eucopdata_tree.txt", package="RRgeo"))->tree
tree$tip.label<-gsub("_", " ", tree$tip.label)
curl::curl_download(paste0(latesturl,"/files/X35kya.tif?download=1"),
  destfile = file.path(newwd,"X35kya.tif"), quiet = FALSE)
rast(file.path(newwd,"X35kya.tif"))->map35
project(map35,st_crs(dat[[1]])$proj4string,res = 50000)->map

ENphylo_modeling(input_data=dat[c(1,11)],
  tree=tree,
  input_mask=map[[1]],
  obs_col="OBS",
  time_col="age",
  min_occ_enfa=15,
  boot_test_perc=20,
  boot_reps=10,
  swap.args=list(nsim=5,si=0.2,si2=0.2),
  eval.args=list(eval_metric_for_imputation="AUC",
    eval_threshold=0.7,
    output_options="best"),
  clust=NULL,
  output.dir=newwd)

getENphylo_results(input.dir =newwd,
  mods="all",
  species_name=names(dat)[c(1,11)])->mod

```

Description

The function integrates phylogenetic and geographical data (i.e., habitat suitability maps), along with tools to model species' movements across landscapes. This integration allows for inference of the most probable area of origin (i.e. speciation) or regions of historical contact across time and space between a pair of target species.

Usage

```
RRphylogeography(spec1, spec2, pred, occs, aggr=NULL, time_col=NULL,
  kde_inversion=FALSE, resistance_map=NULL, th=0.5, clust=0.5, plot=FALSE,
  mask_for_pred=NULL, standardize=TRUE, output.dir)
```

Arguments

spec1, spec2	character. The names of the sister species whose area of origin should be inferred.
pred	a list of two <code>SpatRaster</code> objects containing the prediction maps in logistic output generated through any Species Distribution Model technique. List names must correspond to spec1 and spec2.
occs	a list of two <code>sf::data.frame</code> objects containing species occurrence data only in binary format (exclusively ones for presence). List names must correspond to spec1 and spec2.
aggr	positive integer. Aggregation factor expressed as number of cells in each direction to be aggregated by averaging cell values (optional).
time_col	character. Name of the occs column containing the time intervals associated to each species occurrence (optional).
kde_inversion	logical. If TRUE and time_col is provided, kernel density is estimated by inverting the weights associated to the occurrences of the oldest species.
resistance_map	a <code>SpatRaster</code> object to be optionally provided as a cost surface map to account for the difficulty of moving in each direction across the study area.
th	numeric. The threshold value to define most suitable cells of the species pair as predicted via SDMs.
clust	numeric. The proportion of the proportion of clusters to be used in parallel computing. Default is 0.5. If NULL, parallel computing is disabled.
plot	logical. If TRUE, the area of the origin (RPO) maps with and without the kernel estimation factors are plotted.
mask_for_pred	a <code>SpatRaster</code> object representing the geographical extent of the study area. This is used to crop the prediction maps define the area of interest for the analysis (optional).
standardize	logical. If TRUE, the Relative Probability (RPO) values are standardized between 0 and 1.
output.dir	character. The file path wherein RRphylogeography creates a new folder to store the outputs. This new folder is renamed by concatenating the names of the species pair.

Details

RRphylogeography identifies the most suitable cells for both target species by relying on the `th` value. This threshold represents a numeric quantile so that any habitat suitability value greater than 0 that exceeds the value is considered to belong to the most suitable cells for the species. Conversely, the cells having habitat suitability values lower than `th` are excluded from distance calculation.

Value

A list of `SpatRaster` objects which includes the area of the origin (RPO) and both relative probability (RPO) maps for the species pair calculated for each layer in the prediction maps.

Author(s)

Alessandro Mondanaro, Mirko Di Febbraro, Silvia Castiglione, Carmela Serio, Marina Melchionna, Pasquale Raia

References

Mondanaro, A., Castiglione, S., Di Febbraro, M., Timmermann, A., Girardi, G., Melchionna, M., Serio, C., Belfiore, A.M., & Raia, P. (2025). RRphylogeography: A new method to find the area of origin of species and the history of past contacts between species. *Methods in Ecology and Evolution*, 16: 546-557. 10.1111/2041-210X.14478

See Also

[RRphylogeography vignette](#)

Examples

```
library(RRgeo)
library(terra)
library(sf)

newwd<-tempdir()
# newwd<-"YOUR_DIRECTORY"

rast(system.file("exdata/U.arctos_suitability.tif", package="RRgeo"))->map1
rast(system.file("exdata/U.maritimus_suitability.tif", package="RRgeo"))->map2
load(system.file("exdata/Ursus_occurrences.Rda", package="RRgeo"))
list(Ursus_arctos=map1,Ursus_maritimus=map2)->pred
list(Ursus_arctos=occs_arctos,Ursus_maritimus=occs_marit)->occs

RRphylogeography(spec1="Ursus_arctos",
                  spec2="Ursus_maritimus",
                  pred=pred,
                  occs=occs,
                  aggr=5,
                  time_col="TIME_factor",
                  kde_inversion=FALSE,
                  resistance_map=NULL,
                  clust=NULL,
                  plot=FALSE,
                  mask_for_pred=NULL,
                  th=0.7,
                  standardize=TRUE,
                  output.dir=newwd)
```


Index

BchronCalibrate, [3](#)

cal14C, [2](#)

ENphylo_modeling, [4](#), [9](#), [13](#)

ENphylo_prediction, [8](#), [13](#)

eucop_data_preparation, [10](#)

getENphylo_results, [7](#), [9](#), [13](#)

mahasuhab, [8](#)

optimal.thresholds, [9](#)

RRphylogeography, [14](#)

swapONE, [5](#)