Package ‘DAISIE’

July 21, 2025

Type Package

Title Dynamical Assembly of Islands by Speciation, Immigration and Extinction

Version 4.6.0

Date 2025-05-30

Depends R (>= 4.2.0)

Imports DDD (>= 5.0), deSolve, doParallel, foreach, graphics, magrittr, Matrix, parallel, pracma, Rcpp (>= 1.0.10), stats, subplex, tensor, testit, utils

LinkingTo Rcpp (>= 1.0.10), RcppEigen, BH (>= 1.81.0-1)

Suggests ape, covr, dplyr, future, future.apply, ggplot2, ggtree (>= 3.0.0), gridExtra, knitr, phytools, purrr, rmarkdown, testthat (>= 2.1.0), tibble, tidytree, tidyr

Enhances doMC

License GPL (>= 3) | file LICENSE

Copyright See the file COPYRIGHTS for various DAISIE copyright details

Description Simulates and computes the (maximum) likelihood of a dynamical model of island biota assembly through speciation, immigration and extinction. See Valente et al. (2015) <doi:10.1111/ele.12461>.

NeedsCompilation yes

SystemRequirements C++17

Encoding UTF-8

VignetteBuilder knitr

URL `https://github.com/rsetienne/DAISIE`,
`https://rsetienne.github.io/DAISIE/`

BugReports `https://github.com/rsetienne/DAISIE/issues`

RoxygenNote 7.3.2

Author Rampal S. Etienne [aut, cre] (ORCID: <https://orcid.org/0000-0003-2142-7612>),
Luis Valente [aut] (ORCID: <https://orcid.org/0000-0003-4247-8785>),
Albert B. Phillimore [aut] (ORCID: <https://orcid.org/0000-0002-6553-1553>),
Bart Haegeman [aut] (ORCID: <https://orcid.org/0000-0003-2325-4727>),
Joshua W. Lambert [aut] (ORCID: <https://orcid.org/0000-0001-5218-3046>),
Pedro Santos Neves [aut] (ORCID: <https://orcid.org/0000-0003-2561-4677>),
Shu Xie [aut] (ORCID: <https://orcid.org/0000-0001-9594-946X>),
Richèl J.C. Bilderbeek [aut] (ORCID: <https://orcid.org/0000-0003-1107-7049>),
Hanno Hildenbrandt [aut] (ORCID: <https://orcid.org/0000-0002-6784-1037>),
Torsten Hauffe [ctb] (ORCID: <https://orcid.org/0000-0001-5711-9457>),
Giovanni Laudanno [ctb] (ORCID: <https://orcid.org/0000-0002-2952-3345>),
Nadiah Kristensen [ctb] (ORCID: <https://orcid.org/0000-0002-9720-4581>),
Ornela Dehayem Nanwou [ctb] (ORCID: <https://orcid.org/0009-0009-0507-1215>),
Raphael Scherrer [ctb] (ORCID: <https://orcid.org/0000-0002-1447-7630>)

Maintainer Rampal S. Etienne <`r.s.etienne@rug.nl`>

Repository CRAN

Date/Publication 2025-06-01 21:20:02 UTC

Contents

`add_column_to_dataframe` *Add a column to a data frame*

Description

Add a column to a data frame

Usage

```
add_column_to_dataframe(df, position, column_to_insert)
```

Arguments

`df`	data frame to add the column to
`position`	location in data frame where to insert the column. Position can also be a name of a column
`column_to_insert`	the elements of the column to insert. If the column has a name, this name will be copied into the data frame. Id is does not have a name, it will get the name "nc".

Value

A data frame with the column inserted

`archipelagos41` *DAISIE datalist object including bird phylogenetic data and physical data for 41 archipelagos*

Description

A datalist with 41 items representing the 41 archipelagos. Each archipelago can be called separately using archipelagos41[[x]] with x being a number between 1 and 41. Using archipelagos41[[x]][[1]] will show just the top part of the archipelago item where the archipelago name and physical features are displayed. The structure of each of the archipelagos is the same as regular DAISIE datalist generated using DAISIE_dataprep.

Format

A datalist containing data on the 41 archipelagos studied in Valente et al 2020 (Main Dataset D1). Contains colonisation and branching times for bird species in each of the archipelagos. It also contains information on archipelago name, area, age and distance from the nearest mainland.

Source

Valente L, Phillimore AB, Melo M, Warren BH, Clegg SM, Havenstein K, Tiedemann R, Illera JC, Thébaud C, Aschenbach T, Etienne RS. A simple dynamic model explains island bird diversity worldwide (2020) Nature, 579, 92-96

`archipelago_data` *Physical data on 41 archipelagos*

Description

A dataframe with in subsequent columns the name of the archipelago (Archipelago) the area of the archipelago (Area), the age (Age) and the distance from the mainland (Distance)

Format

A dataframe containing information on archipelago name, area, age and distance from the mainland

Source

Valente L, Phillimore AB, Melo M, Warren BH, Clegg SM, Havenstein K, Tiedemann R, Illera JC, Thébaud C, Aschenbach T, Etienne RS. A simple dynamic model explains island bird diversity worldwide (2020) Nature, 579, 92-96

`Bats_GreaterAntilles` *Colonization and branching times of Noctilionoid bats from the Greater Antilles.*

Description

A list containing the colonization and branching times of the noctilionoid bats of the Greater Antilles. Main dataset used in Valente, Etienne and Dávalos (2017) Nature Ecology and Evolution. Island age 20 Myr and mainland pool size of 100 species.
The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contain information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island.

For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
* Endemic_MaxAge: 5 or 6

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 17 elements, the first of which contains 2 elements and the following 16 containing 5 components.

Source

Valente L, Etienne RS, Dávalos (2017) Recent extinctions disturb path to equilibrium diversity in Caribbean bats. Nature Ecology and Evolution, 1, 26.

See Also

`DAISIE_dataprep`, `DAISIE_ML`, `DAISIE_SR_ML`

`Biwa_datalist` *Colonization and branching times of 68 fish clades.*

Description

A list containing the colonization and branching times of the fishes of Lake Biwa (Japan). Main dataset used in Hauffe et al (2020). This list can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.

The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
* Endemic_MaxAge: 5 or 6

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 69 elements, the first of which contains 2 elements and the following 68 containing 5 components.

Source

Hauffe, T., D. Delicado, R.S. Etienne and L. Valente. Lake expansion elevates equilibrium diversity via increasing colonisation. (2020) Journal of Biogeography

See Also

`DAISIE_dataprep`, `DAISIE_ML`, `DAISIE_SR_ML`

`create_area_pars` *Create named list of area parameters*

Description

Create named list of area parameters

Usage

```
create_area_pars(
  max_area,
  current_area,
  proportional_peak_t,
  total_island_age,
  sea_level_amplitude,
  sea_level_frequency,
  island_gradient_angle
)
```

Arguments

`max_area`	Numeric defining maximum area.
`current_area`	A numeric with the current island area at present (i.e., at the end of the simulation).
`proportional_peak_t`	Numeric value from 0 to 1 indicating where in the island's history the peak area is achieved.
`total_island_age`	Numeric defining total island age.
`sea_level_amplitude`	Numeric defining amplitude of area fluctuation from sea level.
`sea_level_frequency`	Numeric defining frequency of sine wave of area change from sea level.
`island_gradient_angle`	Numeric defining the angle in degrees specifying the slope of the island.

Value

list of numerical values containing area and sea level parameters for island ontogeny simulation

Author(s)

Richel J.C Bilderbeek, Joshua Lambert, Pedro Neves

Examples

```
area_pars <- create_area_pars(
  max_area = 10,
  current_area = 1,
  proportional_peak_t = 0.5,
  total_island_age = 5,
  sea_level_amplitude = 5,
  sea_level_frequency = 10,
  island_gradient_angle = 0


)
```

`create_CS_version` *Creates the list object for CS_version argument in DAISIE_ML_CS*

Description

Creates the list object for CS_version argument in DAISIE_ML_CS

Usage

```
create_CS_version(
  model = 1,
  function_to_optimize = "DAISIE",
  relaxed_par = NULL,
  par_sd = 0,
  par_upper_bound = Inf,
  integration_method = "standard",
  seed = 42,
  sample_size = 100,
  parallel = FALSE,
  n_cores = 1
)
```

Arguments

`model`	the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model
`function_to_optimize`	likelihood function that must be optimized in ML, either 'DAISIE', 'DAISIE_approx', or 'DAISIE_DE'
`relaxed_par`	the parameter to relax (integrate over). Options are `"cladogenesis"`, `"extinction"`, `"carrying_capacity"`, `"immigration"`, `"anagenesis"`
`par_sd`	standard deviation of the parameter to relax
`par_upper_bound`	upper bound of the parameter to relax
`integration_method`	method of integration, either 'standard','stratified' or 'MC'
`seed`	seed of the random number generator in case of 'MC'
`sample_size`	size of sample in case of 'MC' or 'stratified'
`parallel`	use parallel computing or not in case of 'MC' or 'stratified'
`n_cores`	number of cores to use when run in parallel

Value

A list of four elements

- model: the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model
- fumction_to_optimize likelihood function that must be optimized in ML, either 'DAISIE', 'DAISIE_approx', or 'DAISIE_DE'
- relaxed_par: the parameter to relax (integrate over), for model = 2.
- par_sd: the standard deviation of the parameter to relax
- par_upperbound: upper bound of the parameter to relax.
- integration_method: method of integration, either 'standard', 'stratified' or 'MC'
- seed: random seed in case of integration_method = 'MC'
- sample_size: size of sample in case of integration_method = 'MC' or 'stratified'
- parallel: use parallel computing or not in case of integration_method = 'MC' or 'stratified'
- n_cores: number of cores to use when run in parallel

`create_hyper_pars` *Create list of hyperparameters*

Description

Create list of hyperparameters

Usage

```
create_hyper_pars(d, x)
```

Arguments

`d`	Numeric defining the scaling parameter for exponent for calculating cladogenesis rate.
`x`	Numeric defining the exponent for calculating extinction rate.

Value

Named list with hyperparameters

Author(s)

Pedro Neves, Joshua Lambert

Examples

```
hyper_pars <- create_hyper_pars(d = 0.027, x = 0.15)
```

`create_pars` *Create vector of model parameters*

Description

Create vector of model parameters

Usage

```
create_pars(clado_rate, ext_rate, carr_cap, immig_rate, ana_rate)
```

Arguments

`clado_rate`	Numeric rate of cladogenesis
`ext_rate`	Numeric rate of extinction
`carr_cap`	Numeric carrying capacity
`immig_rate`	Numeric rate of immigration
`ana_rate`	Numeric rate of anagenesis

Value

Numeric vector with 5 elements

Examples

```
create_pars(clado_rate = 1,
            ext_rate = 1,
            carr_cap = 10,
            immig_rate = 0.1,
            ana_rate = 1)
```

`create_trait_pars` *Create named list of trait state parameters*

Description

Create named list of trait state parameters

Usage

```
create_trait_pars(
  trans_rate,
  immig_rate2,
  ext_rate2,
  ana_rate2,
  clado_rate2,
  trans_rate2,
  M2
)
```

Arguments

`trans_rate`	A numeric with the per capita transition rate with state1
`immig_rate2`	A numeric with the per capita immigration rate with state2
`ext_rate2`	A numeric with the per capita extinction rate with state2
`ana_rate2`	A numeric with the per capita anagenesis rate with state2
`clado_rate2`	A numeric with the per capita cladogenesis rate with state2
`trans_rate2`	A numeric with the per capita transition rate with state2
`M2`	A numeric with the number of species with trait state 2 on mainland

Value

list of numerical values containing trait state parameters

Examples

```
trait_pars <- create_trait_pars(
   trans_rate = 0.5,
   immig_rate2 = 0.1,
   ext_rate2 = 0.2,
   ana_rate2 = 0.3,
   clado_rate2 = 0.4,
   trans_rate2 = 0.5,
   M2 = 1000
 )
```

`create_trait_pars_2K` *Create named list of trait state parameters*

Description

Create named list of trait state parameters

Usage

```
create_trait_pars_2K(
  trans_rate,
  immig_rate2,
  ext_rate2,
  ana_rate2,
  clado_rate2,
  trans_rate2,
  M2,
  K2
)
```

Arguments

`trans_rate`	A numeric with the per capita transition rate with state1
`immig_rate2`	A numeric with the per capita immigration rate with state2
`ext_rate2`	A numeric with the per capita extinction rate with state2
`ana_rate2`	A numeric with the per capita anagenesis rate with state2
`clado_rate2`	A numeric with the per capita cladogenesis rate with state2
`trans_rate2`	A numeric with the per capita transition rate with state2
`M2`	A numeric with the number of species with trait state 2 on mainland
`K2`	A numeric with the carrying capacity for state 2

Value

list of numerical values containing trait state parameters

`DAISIE_abm_factor` *Sets or retrieves the factor to calculate the step-size used by the odeint::adams_bashforth[_moulton] solvers.*

Description

Sets or retrieves the factor to calculate the step-size used by the odeint::adams_bashforth[_moulton] solvers.

Usage

```
DAISIE_abm_factor(factor)
```

Arguments

`factor`	sets step-size to `factor * (t1 - t0)`.

Value

current factor

DAISIE_convertprobdist

Converts the joint distribution of endemics and non-endemics under the DAISIE model to list format

---

Description

This function converts the joint distribution of the number of endemics and non-endemics from the matrix format of DAISIE_probdist to a list format

Usage

```
DAISIE_convertprobdist(pb)
```

Arguments

pb	Probability distribution in matrix format as output by `DAISIE_probdist()`.

Value

A list of length nrow(pb) containing matrices of square dimensions of size sqrt(ncol - 1) containing the joint probabilities with endemics in the rows and non-endemics in the columns. The last element of the list is a vector a times at which the joint probability distribution is evaluated.

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Compute the probability distribution at t = 4 and t = 8, for a mainland pool
# size of 250 potential colonists and a vector of 5 parameters (cladogenesis, extinction,
# clade-level carrying capacity, immigration, anagenesis) starting from an empty
# island; store in list format

pb <- DAISIE_probdist(
   pars1 = c(0.3,0.35,Inf,0.75,0.012),
   pars2 = c(100,250),
   tvec = c(4,8),
   initEI = c(0,0),
```

```
    initprobs = NULL
    )
prob_dists <- DAISIE_convertprobdist(pb)
```

`DAISIE_count_species` *Count number of species in DAISIE datalist or simulated data.*

Description

Calculates various island diversity metrics from island datasets.

Usage

```
DAISIE_count_species(islands, sort_clade_sizes = TRUE)
```

Arguments

`islands`	Island datalist or simulated data in DAISIE datalist format. Can be a single island (empirical data) generated with DAISIE_dataprep or DAISIEprep. Can also be simulated data generated with DAISIE_sim function.
`sort_clade_sizes`	Default sort_clade_sizes = T outputs clade sizes sorted in ascending order of number of species. sort_clade_sizes=F outputs clade sizes in the same order as they appear in the input datalist.

Value

The output is a list containing the following items:

`clade_sizes_sorted`	List showing the total number of species in each island clade (including missing species). Each item [[i]] on the list gives the sizes of all clades for a single island. If option sort_clade_sizes = T, the clade sizes for are sorted by increasing number of species. If option sort_clade_sizes = F the clade sizes are given in the same order as in the input datalist.
`size_largest_clade`	The total number of species in the largest island clade for each island.
`mean_clade_size`	Mean clade size (average of all island clades)
`number_colonisations`	The total number of colonisations (clades) on each island.
`total_number_species`	The total number of species on each island. These are the extant species at present, including missing species; in case of simulations, this is the number of species present on the island at the end of the simulation.

Author(s)

Luis Valente

See Also

`DAISIE_dataprep`, `DAISIE_plot_island`

Examples

```
# Run function with clade sizes in the order they appear in the input data
data("NewZealand_birds_datalist")
species_count <- DAISIE_count_species(NewZealand_birds_datalist)

# Run function with clade sizes in ascending order
species_count_sorted <- DAISIE_count_species(
  NewZealand_birds_datalist,
  sort_clade_sizes = TRUE
)
```

`DAISIE_CS_max_steps` *CS iteration control*

Description

Sets or retrieves the max. number of iterations used by the odeint solver.

Usage

```
DAISIE_CS_max_steps(max_steps)
```

Arguments

`max_steps`	`max_steps`: sets max. iterations to `max_steps`.

Value

current max. iterations

`DAISIE_dataprep` *Prepare colonisation and branching time data to run in DAISIE.*

Description

This function produces a data object that can be run in DAISIE likelihood computation/optimization functions. The function converts a user-specified table to a DAISIE-compatible format. See Galapagos_datatable.Rdata for a template of an input table.)

Usage

```
DAISIE_dataprep(
  datatable,
  island_age,
  M,
  number_clade_types = 1,
  list_type2_clades = NA,
  prop_type2_pool = "proportional",
  epss = 1e-05,
  verbose = TRUE
)
```

Arguments

`datatable`	Data frame (table) with user-specified data. See file Galapagos_datatable.Rdata for a template of an input table. Each row on the table represents and independent colonisation event. Table has the following four columns. `$Clade_name` - name of independent colonization event `$Status` - One of the following categories: * "Non_endemic": applies to non-endemic species when an approximate colonisation time is known * "Non_endemic_MaxAge": applies to non-endemic species for cases where colonisation time is unknown * "Endemic": applies to endemic species or endemic clades when an approximate colonisation time is known * "Endemic_MaxAge": applies to endemic species or endemic clades for cases where the colonisation time is unknown, or when the user wants to specify an upper bound for colonisation. This could for example apply to endemic species that have recently gone extinct because of anthropogenic causes, and which are not included in the phylogeny ("NA" should be given in the branching times column). It could also apply to insular radiations with long stem branches, for which the time of the first cladogenetic event is known, but the precise time of colonisation is not. * "Endemic_MaxAge_MinAge": same as Endemic_MaxAge but also includes a minimum age for colonisation. * "Non_endemic_MaxAge_MinAge": same as Non_endemic_MaxAge but also

	includes a minimum age for colonisation.#' * "Endemic&Non_Endemic": when endemic clade is present and its mainland ancestor has re-colonized `$Missing_species` - Number of island species that were not sampled for particular clade (only applicable for "Endemic" clades). If NA is given in branching times column, this should be equal to the number of species in the clade minus 1 `$Branching_times` - Stem age of the population/species in the case of "Non_endemic", "Non_endemic_MaxAge" and "Endemic" species with no extant close relatives on the island. Set "NA" if colonisation time unknown and no upper bound is known. For "Endemic" cladogenetic species these should be branching times of the radiation, including the stem age of the radiation (colonisation time estimate).
`island_age`	Age of island in appropriate units
`M`	The size of the mainland pool, i.e the number of species that can potentially colonize the island
`number_clade_types`	Number of clade types. Default: number_clade_types = 1 all species are considered to belong to same macroevolutionary process. If number_clade_types = 2, there are two types of clades with distinct macroevolutionary processes.
`list_type2_clades`	If number_clade_types = 2, list_type2_clades specifies the names of the clades that have a distinct macroevolutionary process. The names must match those in the $Clade_name column of the source data table (e.g. list_type2_clades = "Finches"). If number_clade_types = 1, then list_type2_clades = NA should be specified (default)
`prop_type2_pool`	Specifies the fraction of potential mainland colonists that have a distinct macroevolutionary process. Applies only if number_clade_types = 2. Default "proportional" sets the fraction to be proportional to the number of clades of distinct macroevolutionary process that have colonised the island. Alternatively, the user can specify a value between 0 and 1 (e.g. if mainland pool size is 1000 and prop_type2_pool = 0.02 then number of type2 species is 20).
`epss`	Default= 1E-5 should be appropriate in most cases. This value is used to set the maximum age of colonisation of "Non_endemic_MaxAge" and "Endemic_MaxAge" species to an age that is slightly younger than the island for cases when the age provided for that species is older than the island. The new maximum age is then used as an upper bound to integrate over all possible colonisation times.
`verbose`	Boolean. States if intermediate results should be printed to console. Defaults to `TRUE`.

Details

The output is an R list containing the data formatted to be run on other DAISIE functions.

Value

`datalist`	R list object containing data: The first element of the list has two or three components: `$island_age` - the island age Then, depending on whether a distinction between species types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island or: `$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island `$not_present_type2` - the number of mainland lineages of type 2 that are not present on the island The following elements of the list each contains information on a single colonist lineage on the island and has 5 components: `$colonist_name` - the name of the species or clade that colonized the island `$branching_times` - island age and stem age of the population/species in the case of "Non-endemic", "Non-endemic_MaxAge" and "Endemic" anagenetic species. For "Endemic" cladogenetic species these are island age and branching times of the radiation including the stem age of the radiation. `$stac` - the status of the colonist * Non_endemic_MaxAge: 1 * Endemic: 2 * Endemic&Non_Endemic: 3 * Non_endemic: 4 * Endemic_MaxAge: 5 (if only colonisation time was given) * Endemic_MaxAge: 6 (if colonisation time and cladogenesis times were given) `$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades) `$type_1or2` - whether the colonist belongs to type 1 or type 2

Author(s)

Luis M Valente

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Create Galapagos data object where all taxa have the same macroevolutionary process

utils::data(Galapagos_datatable, package = "DAISIE")
```

```
DAISIE_dataprep(
   datatable = Galapagos_datatable,
   island_age = 4,
   M = 1000
   )

### Create Galapagos data object with a distinct macroevolutionary processes
# for the Darwin's finches. One process applies to type 1 species (all species
# except for Darwin's finches) and the other applies only to type 2 species
# (Darwin's finches). Set fraction of potential colonists of type 2 to be
# proportional to the number of type2 clades present on the island.

utils::data(Galapagos_datatable, package = "DAISIE")
DAISIE_dataprep(
   datatable = Galapagos_datatable,
   island_age = 4,
   M = 1000,
   number_clade_types = 2,
   list_type2_clades = "Finches"
   )

### Create Galapagos data object with a distinct macroevolutionary processes
# for the Darwin's finches. One process applies to type 1 species (all species
# except for Darwin's finches) and the other applies only to type 2 species
# (Darwin's finches). Set fraction of potential colonists of type 2 to be 0.163.

utils::data(Galapagos_datatable, package = "DAISIE")
DAISIE_dataprep(
   datatable = Galapagos_datatable,
   island_age = 4,
   M = 1000,
   number_clade_types = 2,
   list_type2_clades = "Finches",
   prop_type2_pool = 0.163
   )
```

`DAISIE_ExpEIN` *The expected number of endemics and non-endemics under the DAISIE model with no diversity-dependence*

Description

This function calculates the expected number of endemics, non-endemics and the sum of these for a given set of parameter values, a given mainland species pool size and a given time, assuming no diversity-dependence

Usage

```
DAISIE_ExpEIN(tvec, pars, M, initEI = c(0, 0))
```

Arguments

`tvec`	The times at which the probabilities need to be computed.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`initEI`	The initial values for the number of endemics and non-endemics. In `DAISIE_probdist()` or `DAISIE_margprobdist()` either this or initprobs must be NULL. In `DAISIE_numcol()` when it is NULL, it is assumed that the island is empty.

Value

The output is a list with three elements:

`ExpE` The number of endemic species
`ExpI` The number of non-endemic species
`ExpN` The sum of the number of endemics and non-endemics

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Compute the expected values at t = 4, for a mainland pool size of 1000 potential
# colonists and a vector of 5 parameters (cladogenesis, extinction, clade-level carrying
# capacity, immigration, anagenesis)

DAISIE_ExpEIN(
   tvec = c(2,4),
   pars = c(0.5,0.1,Inf,0.01,0.4),
   M = 1000
   )
```

DAISIE_ExpEIN2 *The expected number of endemics and non-endemics under the DAISIE model*

Description

This function calculates the expected number of endemics, non-endemics and the sum of these for a given set of parameter values, a given mainland species pool size and a given time, where there can be diversity-dependence

Usage

```
DAISIE_ExpEIN2(
  tvec,
  pars,
  M,
  initEI = NULL,
  res = 1000,
  ddmodel = 11,
  methode = "ode45",
  reltolint = 1e-16,
  abstolint = 1e-16
)
```

Arguments

`tvec`	The times at which the probabilities need to be computed.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model

	• `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`initEI`	The initial values for the number of endemics and non-endemics. In `DAISIE_probdist()` or `DAISIE_margprobdist()` either this or initprobs must be NULL. In `DAISIE_numcol()` when it is NULL, it is assumed that the island is empty.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade.
`ddmodel`	Sets the model of diversity-dependence: • ddmodel = 0 : no diversity dependence • ddmodel = 1 : linear dependence in speciation rate • ddmodel = 11: linear dependence in speciation rate and in immigration rate • ddmodel = 2 : exponential dependence in speciation rate • ddmodel = 21: exponential dependence in speciation rate and in immigration rate
`methode`	Method of the ODE-solver. Supported Boost `ODEINT` solvers (steppers) are: `"odeint::runge_kutta_cash_karp54"` `"odeint::runge_kutta_fehlberg78"` `"odeint::runge_kutta_dopri5"` `"odeint::bulirsch_stoer"` without `odeint::`-prefix, `ode` method is assumed. The default method overall is `"lsodes"` for `DAISIE_ML_CS()` and `"ode45"` from `ode()` for `DAISIE_ML_IW()`.
`reltolint`	Numeric relative tolerance of the integration
`abstolint`	Numeric absolute tolerance of the integration

Value

`tot_expEIN`	The output is a list with three elements: `ExpE` The number of endemic species at the times in tvec `ExpI` The number of non-endemic species at the times in tvec `ExpN` The sum of the number of endemics and non-endemics at the times in tvec

Author(s)

Rampal S. Etienne

Examples

```
DAISIE_ExpEIN2(tvec = c(0.000001,0.5,0.75,1),
                         pars = c(0.3,0.1,10,1,0.1),
                         M = 1000,
                         initEI = rbind(c(1,0),c(2,0),c(0,1)))
```

`DAISIE_format_CS_full_stt` *Formats clade-specific simulation output into standard DAISIE list output with complete STT table*

Description

Formats clade-specific simulation output into standard DAISIE list output with complete STT table

Usage

```
DAISIE_format_CS_full_stt(
  island_replicates,
  time,
  M,
  verbose = TRUE,
  trait_pars = NULL
)
```

Arguments

`island_replicates`	List output from `DAISIE_sim_core_cr()`, `DAISIE_sim_core_time_dep()`, `DAISIE_sim_core_cr_shif` or `DAISIE_sim_min_type2()` functions. Minimally, this must be a list that has as many elements as replicates. Each element must be a list with the elements `island_age`, `not_present` and `stt_all`. `stt_all` must be a data frame with the column names `Time`, `nI`, `nA`, `nC` and `present`.
`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.

`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`trait_pars`	A named list containing diversification rates considering two trait states created by `create_trait_pars`: • [1]:A numeric with the per capita transition rate with state 1 • [2]:A numeric with the per capita immigration rate with state 2 • [3]:A numeric with the per capita extinction rate with state 2 • [4]:A numeric with the per capita anagenesis rate with state 2 • [5]:A numeric with the per capita cladogenesis rate with state 2 • [6]:A numeric with the per capita transition rate with state 2 • [7]:A numeric with the number of species with trait state 2 on mainland

Value

List with CS DAISIE simulation output

`DAISIE_format_IW_full_stt`

Formats clade-specific simulation output into standard DAISIE list output with complete STT table

---

Description

Formats clade-specific simulation output into standard DAISIE list output with complete STT table

Usage

```
DAISIE_format_IW_full_stt(island_replicates, total_time, M, verbose)
```

Arguments

`island_replicates`	List output from `DAISIE_sim_core_cr()`, `DAISIE_sim_core_time_dep()`, `DAISIE_sim_core_cr_shif` or `DAISIE_sim_min_type2()` functions. Minimally, this must be a list that has as many elements as replicates. Each element must be a list with the elements `island_age`, `not_present` and `stt_all`. `stt_all` must be a data frame with the column names `Time`, `nI`, `nA`, `nC` and `present`.
`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.

`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

List with IW DAISIE simulation output

`DAISIE_IC` *Calculates information criterion from DAISIE ML estimates?*

Description

Calculates information criterion from DAISIE ML estimates?

Usage

```
DAISIE_IC(
  datalist,
  initparsopt,
  idparsopt,
  parsfix,
  idparsfix,
  endmc = 1000,
  res = 100,
  cond = 0,
  ddmodel = 0
)
```

Arguments

`datalist`	Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements. The first element of the list has two or three components: `$island_age` - the island age Then, depending on whether a distinction between types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island or:

	`$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island `$not_present_type2` - the number of mainland lineages of type 2 that are not present on the island The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components: `$colonist_name` - the name of the species or clade that colonized the island `$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade `$stac` - the status of the colonist - Non_endemic_MaxAge: 1 - Endemic: 2 - Endemic&Non_Endemic: 3 - Non_Endemic: 4 - Endemic_Singleton_MaxAge: 5 - Endemic_Clade_MaxAge: 6 - Endemic&Non_Endemic_Clade_MaxAge: 7 - Non_endemic_MaxAge_MinAge: 8 - Endemic_Singleton_MaxAge_MinAge: 9 `$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades) `$type1or2` - whether the colonist belongs to type 1 or type 2
`initparsopt`	The initial values of the parameters that must be optimized, they are all positive.
`idparsopt`	The ids of the parameters that must be optimized. The ids are defined as follows: id = 1 corresponds to lambda^c (cladogenesis rate) id = 2 corresponds to mu (extinction rate) id = 3 corresponds to K (clade-level carrying capacity) id = 4 corresponds to gamma (immigration rate) id = 5 corresponds to lambda^a (anagenesis rate) id = 6 corresponds to lambda^c (cladogenesis rate) for an optional subset of the species id = 7 corresponds to mu (extinction rate) for an optional subset of the species id = 8 corresponds to K (clade-level carrying capacity) for an optional subset of the species id = 9 corresponds to gamma (immigration rate) for an optional subset of the species id = 10 corresponds to lambda^a (anagenesis rate) for an optional subset of the species

	id = 11 corresponds to p_f (fraction of mainland species that belongs to the second subset of species.
`parsfix`	The values of the parameters that should not be optimized.
`idparsfix`	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda^c and K should not be optimized.
`endmc`	Numeric for how many simulations should run.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`ddmodel`	Sets the model of diversity-dependence: • ddmodel = 0 : no diversity dependence • ddmodel = 1 : linear dependence in speciation rate • ddmodel = 11: linear dependence in speciation rate and in immigration rate • ddmodel = 2 : exponential dependence in speciation rate • ddmodel = 21: exponential dependence in speciation rate and in immigration rate

Value

List of two numerics with WIC and AICb

`DAISIE_IW_num_threads` *IW concurrency control*

Description

Sets or retrieves the number of threads used by the odeint solver.

Usage

```
DAISIE_IW_num_threads(num_threads)
```

Arguments

`num_threads`	`num_threads` < `0` or `omitted`: retrieves number of threads. `num_threads` = `0`: sets the number of threads to the number of available cores. `num_threads` = `1`: single-threaded execution. `num_threads` > `1`: sets the number of threads to `num_threads`.

Value

number of threads

Note

The maximum number of threads is limited to the value of the C++ standard library function `std::thread::hardware_concurrency()`. This is also the default number of threads upon library load. Multithreading incurs some overhead. Therefore, single-threaded execution might be faster for small systems.

`DAISIE_loglik_IW` *Computes the loglikelihood of the DAISIE model with island-wide diversity-dependence given data and a set of model parameters*

Description

Computes the loglikelihood of the DAISIE model given colonization and branching times for lineages on an island, and a set of model parameters for the DAISIE model with island-wide diversity-dependence

Usage

```
DAISIE_loglik_IW(
  pars1,
  pars2,
  datalist,
  methode = "lsodes",
  abstolint = 1e-12,
  reltolint = 1e-10,
  verbose = FALSE
)
```

Arguments

`pars1`	Contains the model parameters: `pars1[1]` corresponds to lambda^c (cladogenesis rate) `pars1[2]` corresponds to mu (extinction rate) `pars1[3]` corresponds to K (clade-level carrying capacity) `pars1[4]` corresponds to gamma (immigration rate) `pars1[5]` corresponds to lambda^a (anagenesis rate) `pars1[6]` is optional; it may contain M, the total number of species on the mainland

`pars2`	Contains the model settings `pars2[1]` corresponds to lx = length of ODE variable x `pars2[2]` corresponds to ddmodel = diversity-dependent model, model of diversity-dependence, which can be one of ddmodel = 0 : no diversity dependence ddmodel = 1 : linear dependence in speciation rate ddmodel = 11: linear dependence in speciation rate and in immigration rate ddmodel = 2 : exponential dependence in speciation rate ddmodel = 21: exponential dependence in speciation rate and in immigration rate Only ddmodel = 11 is currently implemented `pars2[3]` corresponds to cond = setting of conditioning cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota `pars2[4]` Specifies whether intermediate output should be provided, because computation may take long. Default is 0, no output. A value of 1 means the parameters and loglikelihood are printed. A value of 2 means also intermediate progress during loglikelihood computation is shown.
`datalist`	Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements. The first element of the list has two or three components: `$island_age` - the island age Then, depending on whether a distinction between types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components: `$colonist_name` - the name of the species or clade that colonized the island `$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation. `$stac` - the status of the colonist * Non_endemic_MaxAge: 1 * Endemic: 2 * Endemic&Non_Endemic: 3 * Non_endemic: 4 * Endemic_MaxAge: 5

	`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`methode`	Method of the ODE-solver. Supported Boost `ODEINT` solvers (steppers) are: `'odeint::runge_kutta_cash_karp54'` `'odeint::runge_kutta_fehlberg78'` [default] `'odeint::runge_kutta_dopri5'` `'odeint::bulirsch_stoer'` `'odeint::adams_bashforth` `\code{'odeint::adams_bashforth_moulton_[1|2|3|4|5|6|7|8]` without `odeint::`-prefix, `ode` method is assumed.
`abstolint`	Absolute tolerance of the integration
`reltolint`	Relative tolerance of the integration
`verbose`	Logical controling if progress is printed to console.

Details

The output is a loglikelihood value

Value

The loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_ML_IW`, `DAISIE_loglik_CS`, `DAISIE_sim_cr`

`DAISIE_margprobdist` *The marginal distribution of endemics and non-endemics under the DAISIE model*

Description

This function calculates the marginal distribution of the number of endemics and non-endemics and their sum for a given set of parameter values, a given mainland species pool size and a given set of times

Usage

```
DAISIE_margprobdist(
  pars1,
  pars2,
  tvec,
  initEI = c(0, 0),
  initprobs = NULL,
  pb = NULL
)
```

Arguments

`pars1`	Vector of model parameters: `pars1[1]` corresponds to lambda^c (cladogenesis rate) `pars1[2]` corresponds to mu (extinction rate) `pars1[3]` corresponds to K (clade-level carrying capacity) `pars1[4]` corresponds to gamma (immigration rate) `pars1[5]` corresponds to lambda^a (anagenesis rate).
`pars2`	Vector of settings: `pars2[1]` corresponds to res, the maximum number of endemics or non-endemics for which the ODE system is solved; this must be much larger than the actual number for which the probability needs to be calculated.) `pars2[2]` corresponds to M, size of the mainland pool, i.e the number of species that can potentially colonize the island.
`tvec`	The times at which the probabilities need to be computed.
`initEI`	The initial values for the number of endemics and non-endemics. In `DAISIE_probdist()` or `DAISIE_margprobdist()` either this or initprobs must be NULL. In `DAISIE_numcol()` when it is NULL, it is assumed that the island is empty.
`initprobs`	The initial probability distribution for the number of endemics and non-endemics; either this or initEI must be NULL.
`pb`	Probability distribution in matrix format as output by `DAISIE_probdist()`.

Value

`out`	A list of three vectors: `pE` The probability distribution of the number of endemic species `pI` The probability distribution of the number of non-endemic species `pN` The probability distribution of the sum of the number of endemics and non-endemics

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Compute the marginal probability distributions at t = 4 and t = 8, for a mainland
# pool size of 250 potential colonists and a vector of 5 parameters (cladogenesis,
# extinction, clade-level carrying capacity, immigration, anagenesis) starting from
# an empty island

marg_prob_dists <- DAISIE_margprobdist(
   pars1 = c(0.3,0.35,Inf,0.75,0.012),
   pars2 = c(100,250),
   tvec = c(4,8),
   initEI = c(5,1),
   initprobs = NULL
   )
```

DAISIE_margprobdist2 *The probability distribution of the number of endemics and non-endemics under the DAISIE model*

Description

This function calculates the probability distribution of the number of endemics, non-endemics and the sum of these for a given set of parameter values, a given mainland species pool size and a given time, where there can be diversity-dependence

Usage

```
DAISIE_margprobdist2(
  tvec,
  pars,
  M,
  initEI_mat = NULL,
  res = 1000,
  ddmodel = 11,
  methode = "ode45",
  reltolint = 1e-16,
  abstolint = 1e-16
)
```

Arguments

`tvec`	The times at which the probabilities need to be computed.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`initEI_mat`	matrix where each row represents the initial number of endemic and non-endemic species per colonizing lineage.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade.
`ddmodel`	Sets the model of diversity-dependence: • ddmodel = 0 : no diversity dependence • ddmodel = 1 : linear dependence in speciation rate • ddmodel = 11: linear dependence in speciation rate and in immigration rate • ddmodel = 2 : exponential dependence in speciation rate • ddmodel = 21: exponential dependence in speciation rate and in immigration rate
`methode`	Method of the ODE-solver. Supported Boost `ODEINT` solvers (steppers) are: `"odeint::runge_kutta_cash_karp54"` `"odeint::runge_kutta_fehlberg78"` `"odeint::runge_kutta_dopri5"` `"odeint::bulirsch_stoer"` without `odeint::`-prefix, `ode` method is assumed. The default method overall is `"lsodes"` for `DAISIE_ML_CS()` and `"ode45"` from `ode()` for `DAISIE_ML_IW()`.
`reltolint`	Numeric relative tolerance of the integration
`abstolint`	Numeric absolute tolerance of the integration

Value

`probsEIN`	The output is a list with three elements: `probsE` The number of endemic species at the times in tvec `probsI` The number of non-endemic species at the times in tvec `probsN` The sum of the number of endemics and non-endemics at the times in tvec

Author(s)

Rampal S. Etienne

Examples

```
DAISIE_margprobdist2(tvec = c(0.000001,0.5,0.75,1),
                               pars = c(0.3,0.1,10,1,0.1),
                               M = 1000,
                               initEI_mat = rbind(c(1,0),c(2,0),c(0,1)))
```

`DAISIE_ML` *Maximization of the loglikelihood under the DAISIE model with clade-specific diversity-dependence*

Description

This function computes the maximum likelihood estimates of the parameters of the DAISIE model with clade-specific diversity-dependence for data from lineages colonizing an island. It also outputs the corresponding loglikelihood that can be used in model comparisons. The result of sort(c(idparsopt, idparsfix, idparsnoshift)) should be identical to c(1:10). If not, an error is reported that the input is incoherent. The same happens when the length of initparsopt is different from the length of idparsopt, and the length of parsfix is different from the length of idparsfix.
Including the 11th parameter (p_f) in either idparsopt or idparsfix (and therefore initparsopt or parsfix) is optional. If this parameter is not specified, then the information in the data is used, otherwise the information in the data is overruled.

Usage

```
DAISIE_ML_CS(
  datalist,
  datatype = "single",
  initparsopt,
  idparsopt,
  parsfix,
  idparsfix,
  idparsnoshift = 6:10,
  idparsmat = NULL,
  res = 100,
```

```
    ddmodel = 0,
    cond = 0,
    island_ontogeny = NA,
    eqmodel = 0,
    x_E = 0.95,
    x_I = 0.98,
    tol = c(1e-04, 1e-05, 1e-07),
    maxiter = 1000 * round((1.25)^length(idparsopt)),
    methode = "odeint::runge_kutta_cash_karp54",
    optimmethod = "simplex",
    CS_version = list(model = 1, function_to_optimize = "DAISIE"),
    verbose = 0,
    tolint = c(1e-16, 1e-10),
    jitter = 0,
    num_cycles = 1,
    equal_extinction = TRUE
  )
```

Arguments

`datalist` Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.
The first element of the list has two or three components:

`$island_age` - the island age
Then, depending on whether a distinction between types is made, we have:
`$not_present` - the number of mainland lineages that are not present on the island
or:
`$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island
`$not_present_type2` - the number of mainland lineages of type 2 that are not present on the island

The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade
`$stac` - the status of the colonist

- Non_endemic_MaxAge: 1

- Endemic: 2
- Endemic&Non_Endemic: 3
- Non_Endemic: 4
- Endemic_Singleton_MaxAge: 5
- Endemic_Clade_MaxAge: 6
- Endemic&Non_Endemic_Clade_MaxAge: 7
- Non_endemic_MaxAge_MinAge: 8
- Endemic_Singleton_MaxAge_MinAge: 9

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2

`datatype`	Sets the type of data: 'single' for a single island or archipelago treated as one, and 'multiple' for multiple archipelagoes potentially sharing the same parameters.
`initparsopt`	The initial values of the parameters that must be optimized, they are all positive.
`idparsopt`	The ids of the parameters that must be optimized. The ids are defined as follows: id = 1 corresponds to lambda^c (cladogenesis rate) id = 2 corresponds to mu (extinction rate) id = 3 corresponds to K (clade-level carrying capacity) id = 4 corresponds to gamma (immigration rate) id = 5 corresponds to lambda^a (anagenesis rate) id = 6 corresponds to lambda^c (cladogenesis rate) for an optional subset of the species id = 7 corresponds to mu (extinction rate) for an optional subset of the species id = 8 corresponds to K (clade-level carrying capacity) for an optional subset of the species id = 9 corresponds to gamma (immigration rate) for an optional subset of the species id = 10 corresponds to lambda^a (anagenesis rate) for an optional subset of the species id = 11 corresponds to p_f (fraction of mainland species that belongs to the second subset of species.
`parsfix`	The values of the parameters that should not be optimized.
`idparsfix`	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda^c and K should not be optimized.
`idparsnoshift`	For datatype = 'single' only: The ids of the parameters that should not be different between two groups of species; This can only apply to ids 6:10, e.g. idparsnoshift = c(6,7) means that lambda^c and mu have the same values for both groups.
`idparsmat`	For datatype = 'multiple' only: Matrix containing the ids of the parameters, linking them to initparsopt and parsfix. Per island system we use the following order: * lac = (initial) cladogenesis rate

	* mu = extinction rate * K = maximum number of species possible in the clade * gam = (initial) immigration rate * laa = (initial) anagenesis rate Example: `idparsmat = rbind(c(1, 2, 3, 4, 5), c(1, 2, 3, 6, 7))` has different rates of immigration and anagenesis for the two islands.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade.
`ddmodel`	Sets the model of diversity-dependence: • ddmodel = 0 : no diversity dependence • ddmodel = 1 : linear dependence in speciation rate • ddmodel = 11: linear dependence in speciation rate and in immigration rate • ddmodel = 2 : exponential dependence in speciation rate • ddmodel = 21: exponential dependence in speciation rate and in immigration rate
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`eqmodel`	Sets the equilibrium constraint that can be used during the likelihood optimization. Only available for datatype = 'single'. eqmodel = 0 : no equilibrium is assumed eqmodel = 13 : near-equilibrium is assumed on endemics using deterministic equation for endemics and immigrants. Endemics must be within x_E of the equilibrium value eqmodel = 15 : near-equilibrium is assumed on endemics and immigrants using deterministic equation for endemics and immigrants. Endemics must be within x_E of the equilibrium value, while non-endemics must be within x_I of the equilibrium value.
`x_E`	Sets the fraction of the equlibrium endemic diversity above which the endemics are assumed to be in equilibrium; only active for eqmodel = 13 or 15.
`x_I`	Sets the fraction of the equlibrium non-endemic diversity above which the system is assumed to be in equilibrium; only active for eqmodel = 15.

`tol`	Sets the tolerances in the optimization. Consists of: reltolx = relative tolerance of parameter values in optimization reltolf = relative tolerance of function value in optimization abstolx = absolute tolerance of parameter values in optimization.
`maxiter`	Sets the maximum number of iterations in the optimization.
`methode`	Method of the ODE-solver. Supported Boost `ODEINT` solvers (steppers) are: `"odeint::runge_kutta_cash_karp54"` `"odeint::runge_kutta_fehlberg78"` `"odeint::runge_kutta_dopri5"` `"odeint::bulirsch_stoer"` without `odeint::`-prefix, `ode` method is assumed. The default method overall is `"lsodes"` for `DAISIE_ML_CS()` and `"ode45"` from `ode()` for `DAISIE_ML_IW()`.
`optimmethod`	Method used in likelihood optimization. Default is ‘simplex‘ in the standard Clade Specific scenario. Alternative is ‘subplex‘ (see ‘subplex()‘ for full details) which was the default method in previous versions. In the Island Wide, two type scenarios, and split rate scenarios the default remains ‘subplex‘.
`CS_version`	a numeric or list. Default is CS_version = list(model = 1, function_to_optimize = ’DAISIE’), but for a relaxed-rate model the list can contain more elements: • model: the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model • function_to_optimize: the DAISIE loglikelihood function that will be optimized. Options are: `"DAISIE"`, default, the full DAISIE loglikelihood `"DAISIE_approx"`, an approximate loglikelihood `"DAISIE_DE"`, an exact loglikelkhood for K = Inf based on the D-E approach • integration_method: the method used to do integraion in the relaxed rate model. Options are: `'standard'` the default numerical integration `'MC'` Monte Carlo integration `'stratified'` using quantiles of the gamma distribution • relaxed_par: the parameter to relax (integrate over) in the relaxed rate model. Options are `"cladogenesis"`, `"extinction"`, `"carrying_capacity"`, `"immigration"`, or `"anagenesis"` • par_sd: standard deviation of the parameter to relax • par_upper_bound upper bound of the parameter to relax • seed: seed of the random number generator in case of ’MC’ • sample_size: size of sample in case of ’MC’ or ’stratified’ • parallel: use parallel computing or not in case of ’MC’ or ’stratified’ • n_cores: number of cores to use when run in parallel
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`tolint`	Vector of two elements containing the absolute and relative tolerance of the integration.

`jitter`	Numeric for `optimizer()`. Jitters the parameters being optimized by the specified amount which should be very small, e.g. 1e-5. Jitter when `link{subplex}{subplex}()` produces incorrect output due to parameter transformation.
`num_cycles`	The number of cycles the optimizer will go through. Default is 1.
`equal_extinction`	If FALSE the extinction rates of endemic and non-endemic species are different, otherwise they are set equal in optimization

Value

The output is a dataframe containing estimated parameters and maximum loglikelihood.

`lambda_c`	gives the maximum likelihood estimate of lambda^c, the rate of cladogenesis
`mu`	gives the maximum likelihood estimate of mu, the extinction rate
`K`	gives the maximum likelihood estimate of K, the carrying-capacity
`gamma`	gives the maximum likelihood estimate of gamma, the immigration rate
`lambda_a`	gives the maximum likelihood estimate of lambda^a, the rate of anagenesis
`lambda_c2`	gives the maximum likelihood estimate of lambda^c2, the rate of cladogenesis for the optional second group of species
`mu2`	gives the maximum likelihood estimate of mu2, the extinction rate for the optional second group of species
`K2`	gives the maximum likelihood estimate of K2, the carrying-capacity for the optional second group of species
`gamma2`	gives the maximum likelihood estimate of gamma2, the immigration rate for the optional second group of species
`lambda_a2`	gives the maximum likelihood estimate of lambda^a2, the rate of anagenesis for the optional second group of species
`loglik`	gives the maximum loglikelihood
`df`	gives the number of estimated parameters, i.e. degrees of feedom
`conv`	gives a message on convergence of optimization; conv = 0 means convergence

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852. <doi:10.1111/ele.12461>.

See Also

`DAISIE_loglik_all`, `DAISIE_sim_cr`, `DAISIE_sim_time_dep`, `DAISIE_sim_cr_shift`

Examples

```
cat("
### When all species have the same rates, and we want to optimize all 5 parameters,
# we use:

utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = Galapagos_datalist,
   initparsopt = c(2.5,2.7,20,0.009,1.01),
   ddmodel = 11,
   idparsopt = 1:5,
   parsfix = NULL,
   idparsfix = NULL
)

### When all species have the same rates, and we want to optimize all parameters
# except K (which we set equal to Inf), we use:

utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = Galapagos_datalist,
   initparsopt = c(2.5,2.7,0.009,1.01),
   idparsopt = c(1,2,4,5),
   parsfix = Inf,
   idparsfix = 3
   )

### When all species have the same rates except that the finches have a different
# rate of cladogenesis, and we want to optimize all parameters except K (which we
# set equal to Inf), fixing the proportion of finch-type species at 0.163, we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   initparsopt = c(0.38,0.55,0.004,1.1,2.28),
   idparsopt = c(1,2,4,5,6),
   parsfix = c(Inf,Inf,0.163),
   idparsfix = c(3,8,11),
   idparsnoshift = c(7,9,10)
   )

### When all species have the same rates except that the finches have a different
# rate of cladogenesis, extinction and a different K, and we want to optimize all
# parameters, fixing the proportion of finch-type species at 0.163, we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   ddmodel = 11,
   initparsopt = c(0.19,0.09,0.002,0.87,20,8.9,15),
   idparsopt = c(1,2,4,5,6,7,8),
   parsfix = c(Inf,0.163),
```

```
   idparsfix = c(3,11),
   idparsnoshift = c(9,10)
   )


### When all species have the same rates except that the finches have a different
# rate of extinction, and we want to optimize all parameters except K (which we
# set equal to Inf), and we also# want to estimate the fraction of finch species
# in the mainland pool. we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   initparsopt = c(2.48,2.7,0.009,1.01,2.25,0.163),
   idparsopt = c(1,2,4,5,7,11),
   parsfix = c(Inf,Inf),
   idparsfix = c(3,8),
   idparsnoshift = c(6,9,10)
   )

### When we have two islands with the same rates except for immigration and anagenesis rate,
# and we want to optimize all parameters, we use:

utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = list(Galapagos_datalist,Galapagos_datalist),
   datatype = 'multiple',
   initparsopt = c(2.5,2.7,20,0.009,1.01,0.009,1.01),
   idparsmat = rbind(1:5,c(1:3,6,7)),
   idparsopt = 1:7,
   parsfix = NULL,
   idparsfix = NULL
)

### When we consider the four Macaronesia archipelagoes and set all parameters the same
# except for rates of cladogenesis, extinction and immigration for Canary Islands,
# rate of cladogenesis is fixed to 0 for the other archipelagoes,
# diversity-dependence is assumed to be absent
# and we want to optimize all parameters, we use:

utils::data(Macaronesia_datalist)
DAISIE_ML(
   datalist = Macaronesia_datalist,
   datatype = 'multiple',
   initparsopt = c(1.053151832,0.052148979,0.512939011,0.133766934,0.152763179),
   idparsmat = rbind(1:5,c(6,2,3,7,5),1:5,1:5),
   idparsopt = c(2,4,5,6,7),
   parsfix = c(0,Inf),
   idparsfix = c(1,3)
)

")
```

`DAISIE_ML_IW` *Maximization of the loglikelihood under the DAISIE model with island-wide diversity-dependence*

Description

This function computes the maximum likelihood estimates of the parameters of the DAISIE model with island-wide diversity-dependence for data from lineages colonizing an island. It also outputs the corresponding loglikelihood that can be used in model comparisons.

Usage

```
DAISIE_ML_IW(
  datalist,
  initparsopt,
  idparsopt,
  parsfix,
  idparsfix,
  res = 100,
  ddmodel = 11,
  cond = 0,
  tol = c(1e-04, 1e-05, 1e-07),
  maxiter = 1000 * round((1.25)^length(idparsopt)),
  methode = "ode45",
  optimmethod = "subplex",
  verbose = 0,
  tolint = c(1e-16, 1e-14),
  jitter = 0,
  num_cycles = 1
)
```

Arguments

`datalist`	Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements. The first element of the list has two or three components: `$island_age` - the island age Then, depending on whether a distinction between types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island or: `$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island `$not_present_type2` - the number of mainland lineages of type 2 that are not

present on the island

The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade
`$stac` - the status of the colonist

- Non_endemic_MaxAge: 1
- Endemic: 2
- Endemic&Non_Endemic: 3
- Non_Endemic: 4
- Endemic_Singleton_MaxAge: 5
- Endemic_Clade_MaxAge: 6
- Endemic&Non_Endemic_Clade_MaxAge: 7
- Non_endemic_MaxAge_MinAge: 8
- Endemic_Singleton_MaxAge_MinAge: 9

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2

`initparsopt`	The initial values of the parameters that must be optimized, they are all positive.
`idparsopt`	The ids of the parameters that must be optimized. The ids are defined as follows: id = 1 corresponds to lambda^c (cladogenesis rate) id = 2 corresponds to mu (extinction rate) id = 3 corresponds to K (clade-level carrying capacity) id = 4 corresponds to gamma (immigration rate) id = 5 corresponds to lambda^a (anagenesis rate) id = 6 corresponds to lambda^c (cladogenesis rate) for an optional subset of the species id = 7 corresponds to mu (extinction rate) for an optional subset of the species id = 8 corresponds to K (clade-level carrying capacity) for an optional subset of the species id = 9 corresponds to gamma (immigration rate) for an optional subset of the species id = 10 corresponds to lambda^a (anagenesis rate) for an optional subset of the species id = 11 corresponds to p_f (fraction of mainland species that belongs to the second subset of species.
`parsfix`	The values of the parameters that should not be optimized.

`idparsfix`	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda^c and K should not be optimized.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade.
`ddmodel`	Sets the model of diversity-dependence: • ddmodel = 0 : no diversity dependence • ddmodel = 1 : linear dependence in speciation rate • ddmodel = 11: linear dependence in speciation rate and in immigration rate • ddmodel = 2 : exponential dependence in speciation rate • ddmodel = 21: exponential dependence in speciation rate and in immigration rate
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`tol`	Sets the tolerances in the optimization. Consists of: reltolx = relative tolerance of parameter values in optimization reltolf = relative tolerance of function value in optimization abstolx = absolute tolerance of parameter values in optimization.
`maxiter`	Sets the maximum number of iterations in the optimization.
`methode`	Method of the ODE-solver. Supported Boost `ODEINT` solvers (steppers) are: `"odeint::runge_kutta_cash_karp54"` `"odeint::runge_kutta_fehlberg78"` `"odeint::runge_kutta_dopri5"` `"odeint::bulirsch_stoer"` without `odeint::`-prefix, `ode` method is assumed. The default method overall is `"lsodes"` for `DAISIE_ML_CS()` and `"ode45"` from `ode()` for `DAISIE_ML_IW()`.
`optimmethod`	Method used in likelihood optimization. Default is ‘simplex‘ in the standard Clade Specific scenario. Alternative is ‘subplex‘ (see ‘subplex()‘ for full details) which was the default method in previous versions. In the Island Wide, two type scenarios, and split rate scenarios the default remains ‘subplex‘.
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`tolint`	Vector of two elements containing the absolute and relative tolerance of the integration.
`jitter`	Numeric for `optimizer()`. Jitters the parameters being optimized by the specified amount which should be very small, e.g. 1e-5. Jitter when `link{subplex}{subplex}()` produces incorrect output due to parameter transformation.
`num_cycles`	The number of cycles the optimizer will go through. Default is 1.

Details

The result of sort(c(idparsopt, idparsfix)) should be identical to c(1:5). If not, an error is reported that the input is incoherent. The same happens when the length of initparsopt is different from the length of idparsopt, and the length of parsfix is different from the length of idparsfix.

Value

The output is a dataframe containing estimated parameters and maximum loglikelihood.

`lambda_c`	gives the maximum likelihood estimate of lambda^c, the rate of cladogenesis
`mu`	gives the maximum likelihood estimate of mu, the extinction rate
`K`	gives the maximum likelihood estimate of K, the carrying-capacity
`gamma`	gives the maximum likelihood estimate of gamma, the immigration rate
`lambda_a`	gives the maximum likelihood estimate of lambda^a, the rate of anagenesis
`loglik`	gives the maximum loglikelihood
`df`	gives the number of estimated parameters, i.e. degrees of feedom
`conv`	gives a message on convergence of optimization; conv = 0 means convergence

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852. <DOI:10.1111/ele.12461>.

See Also

`DAISIE_loglik_IW`, `DAISIE_ML_CS` `DAISIE_sim_cr`

`DAISIE_MW_ML` *Maximization of the loglikelihood under the DAISIE model with clade-specific diversity-dependence and explicit dependencies on island area and isolation as hypothesized by MacArthur & Wilson*

Description

This function computes the maximum likelihood estimates of the parameters of the relationships between parameters of the DAISIE model (with clade-specific diversity-dependence) and island area and distance of the island to the mainland for data from lineages colonizing several islands/archipelagos. It also outputs the corresponding loglikelihood that can be used in model comparisons.

A note on the sigmoidal functions used in distance_dep: For anagenesis and cladogenesis, the functional relationship is k * (d/d0)^x/(1 + (d/d0)^x); for colonization the relationship is: k - k * (d/d0)^x/(1 + (d/d0)^x). The d0 parameter is the 11th parameter entered. In 'sigmoidal_col_ana', the 11th parameter is the d0 for colonization and the 12th is the d0 for anagenesis.

Usage

```
DAISIE_MW_ML(
  datalist,
  initparsopt,
  idparsopt,
  parsfix,
  idparsfix,
  res = 100,
  ddmodel = 11,
  cond = 0,
  island_ontogeny = NA,
  tol = c(1e-04, 1e-05, 1e-07),
  maxiter = 1000 * round((1.25)^length(idparsopt)),
  methode = "odeint::runge_kutta_cash_karp54",
  optimmethod = "simplex",
  CS_version = list(model = 1, function_to_optimize = "DAISIE"),
  verbose = 0,
  tolint = c(1e-16, 1e-10),
  distance_type = "continent",
  distance_dep = "power",
  parallel = "local",
  cpus = 3,
  num_cycles = 1
)
```

Arguments

`datalist`	Data object containing information on colonisation and branching times of species for several islands or archipelagos, as well as the area, isolation and age of each of the islands/archipelagos. See data(archipelagos41) for an example.
`initparsopt`	The initial values of the parameters that must be optimized; they are all positive
`idparsopt`	The ids of the parameters that must be optimized. The ids are defined as follows (see Valente et al 2020 Supplementary Tables 1 and 2 a better explanation of the models and parameters): id = 1 corresponds to lambda^c0 (cladogenesis rate for unit area) id = 2 corresponds to y (exponent of area for cladogenesis rate) id = 3 corresponds to mu0 (extinction rate for unit area) id = 4 corresponds to x (exponent of 1/area for extinction rate) id = 5 corresponds to K0 (clade-level carrying capacity for unit area) id = 6 corresponds to z (exponent of area for clade-level carrying capacity) id = 7 corresponds to gamma0 (immigration rate for unit distance) id = 8 corresponds to alpha (exponent of 1/distance for immigration rate) id = 9 corresponds to lambda^a0 (anagenesis rate for unit distance) id = 10 corresponds to beta (exponent of 1/distance for anagenesis rate) id = 11 corresponds to d0 in models M15 to M19, and models with distance_dep = 'sigmoidal_col', 'sigmoidal_ana' or 'sigmoidal_clado'; or d0 for colonisation (when specifying distance_dep = 'sigmoidal_col_ana'

	id = 12 corresponds to d0 for anagenesis when specifying distance_dep = 'sigmoidal_col_ana'
`parsfix`	The values of the parameters that should not be optimized
`idparsfix`	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda^c and K should not be optimized.
`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade
`ddmodel`	Sets the model of diversity-dependence: ddmodel = 0 : no diversity dependence ddmodel = 1 : linear dependence in speciation rate ddmodel = 11: linear dependence in speciation rate and in immigration rate ddmodel = 2 : exponential dependence in speciation rate ddmodel = 21: exponential dependence in speciation rate and in immigration rate
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota
`island_ontogeny`	type of island ontonogeny. If NA, then constant ontogeny is assumed
`tol`	Sets the tolerances in the optimization. Consists of: reltolx = relative tolerance of parameter values in optimization reltolf = relative tolerance of function value in optimization abstolx = absolute tolerance of parameter values in optimization
`maxiter`	Sets the maximum number of iterations in the optimization
`methode`	Method of the ODE-solver. See package deSolve for details. Default is "lsodes"
`optimmethod`	Method used in likelihood optimization. Default is "subplex" (see subplex package). Alternative is 'simplex' which was the method in previous versions.
`CS_version`	a numeric or list. Default is CS_version = list(model = 1, function_to_optimize = 'DAISIE'), but for a relaxed-rate model the list can contain more elements: • model: the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model • function_to_optimize: the DAISIE loglikelihood function that will be optimized. Options are: `"DAISIE"`, default, the full DAISIE loglikelihood `"DAISIE_approx"`, an approximate loglikelihood `"DAISIE_DE"`, an exact loglikelkhood for K = Inf based on the D-E approach • integration_method: the method used to do integraion in the relaxed rate model. Options are: `'standard'` the default numerical integration `'MC'` Monte Carlo integration `'stratified'` using quantiles of the gamma distribution • relaxed_par: the parameter to relax (integrate over) in the relaxed rate model. Options are `"cladogenesis"`, `"extinction"`, `"carrying_capacity"`, `"immigration"`, or `"anagenesis"`

	• par_sd: standard deviation of the parameter to relax • par_upper_bound upper bound of the parameter to relax • seed: seed of the random number generator in case of 'MC' • sample_size: size of sample in case of 'MC' or 'stratified' • parallel: use parallel computing or not in case of 'MC' or 'stratified' • n_cores: number of cores to use when run in parallel
`verbose`	sets whether parameters and likelihood should be printed (1) or not (0)
`tolint`	Vector of two elements containing the absolute and relative tolerance of the integration
`distance_type`	Use 'continent' if the distance to the continent should be used, use 'nearest_big' if the distance to the nearest big landmass should be used, and use 'biologically_realistic' if the distance should take into account some biologically realism, e.g. an average of the previous two if both are thought to contribute.
`distance_dep`	Sets what type of distance dependence should be used. Default is a power law, denoted as 'power' (models M1-14 in Valente et al 2020). Alternatives are additive or interactive contributions of distance and area to the rate of cladogenesis ("area_additive_clado"; "area_interactive_clado", "area_interactive_clado1" and "area_interactive_clado2"). Other alternatives are exponential relationship denoted by 'exp'; or sigmoids, either 'sigmoidal_col' for a sigmoid in the colonization, 'sigmoidal_ana' for sigmoidal anagenesis, 'sigmoidal_clado' for sigmoidal cladogenesis, and 'sigmoidal_col_ana' for sigmoids in both colonization and anagenesis. A key for the different options of distance_dep that should be specified to run the models from Valente et al 2020 (Supplementary Data Table 1 and 2) is given below: * M1 to M14 - 'power' * M15 -'area_additive_clado' * M16 and M19 -'area_interactive_clado' * M17 -'area_interactive_clado1' * M18 - 'area_interactive_clado2' * M20 and M24 - sigmoidal_col' * M21, M25 and M28 - sigmoidal_ana' * M22 and M26 - 'sigmoidal_clado' * M23 and M27 - 'sigmoidal_col_ana'
`parallel`	Sets whether parallel computation should be used. Use 'no' if no parallel computing should be used, 'cluster' for parallel computing on a unix/linux cluster, and 'local' for parallel computation on a local machine.
`cpus`	Number of cpus used in parallel computing. Default is 3. Will not have an effect if parallel = 'no'.
`num_cycles`	The number of cycles the optimizer will go through. Default is 1.

Value

The output is a dataframe containing estimated parameters and maximum loglikelihood.

`lambda_c0`	gives the maximum likelihood estimate of lambda^c, the rate of cladogenesis for unit area
`y`	gives the maximum likelihood estimate of y, the exponent of area for the rate of cladogenesis
`mu0`	gives the maximum likelihood estimate of mu0, the extinction rate
`x`	gives the maximum likelihood estimate of x, the exponent of 1/area for the extinction rate
`K0`	gives the maximum likelihood estimate of K0, the carrying-capacity for unit area
`z`	gives the maximum likelihood estimate of z, the exponent of area for the carrying capacity
`gamma0`	gives the maximum likelihood estimate of gamma0, the immigration rate for unit distance
`y`	gives the maximum likelihood estimate of alpha, the exponent of 1/distance for the rate of colonization
`lambda_a0`	gives the maximum likelihood estimate of lambda^a0, the rate of anagenesis for unit distance
`beta`	gives the maximum likelihood estimate of beta, the exponent of 1/distance for the rate of anagenesis
`loglik`	gives the maximum loglikelihood
`df`	gives the number of estimated parameters, i.e. degrees of feedom
`conv`	gives a message on convergence of optimization; conv = 0 means convergence

Author(s)

Rampal S. Etienne & Luis Valente

References

Valente L, Phillimore AB, Melo M, Warren BH, Clegg SM, Havenstein K, Tiedemann R, Illera JC, Thébaud C, Aschenbach T, Etienne RS. A simple dynamic model explains island bird diversity worldwide (2020) Nature, 579, 92-96

See Also

DAISIE_ML_CS,

Examples

```
cat("
### Fit the M19 model as in Valente et al 2020, using the ML
parameters as starting values (see Supplementary Tables 1 and 2).

utils::data(archipelagos41)

DAISIE_MW_ML(
```

```
datalist= archipelagos41,
initparsopt =
c(0.040073803,1.945656546,0.150429656,
67.25643672,0.293635061,0.059096872,0.382688527,
0.026510781),
idparsopt = c(1,3,4,7,8,9,10,11),
parsfix = c(0,Inf,0) ,
idparsfix = c(2,5,6),
res = 100,
ddmodel = 0,
methode = 'lsodes',
cpus = 4,
parallel = 'local',
optimmethod = 'subplex',
tol = c(1E-4, 1E-5, 1E-7),
distance_type = 'continent',
distance_dep = 'area_interactive_clado'
)
")
```

`DAISIE_numcol` *The expectation and marginal distribution of the number of colonizations (lineages) under the DAISIE model*

Description

This function calculates expectation and marginal distribution of the number of colonizations (lineages) for a given set of parameter values, a given mainland species pool size and a given set of times

Usage

```
DAISIE_numcol(pars1, pars2, tvec, initEI = NULL)
```

Arguments

`pars1`	Vector of model parameters: `pars1[1]` corresponds to lambda^c (cladogenesis rate) `pars1[2]` corresponds to mu (extinction rate) `pars1[3]` corresponds to K (clade-level carrying capacity) `pars1[4]` corresponds to gamma (immigration rate) `pars1[5]` corresponds to lambda^a (anagenesis rate).
`pars2`	Vector of settings: `pars2[1]` corresponds to res, the maximum number of endemics or non-endemics for which the ODE system is solved; this must be much larger than the actual number for which the probability needs to be calculated.)

	`pars2[2]` corresponds to M, size of the mainland pool, i.e the number of species that can potentially colonize the island.
`tvec`	The times at which the probabilities need to be computed.
`initEI`	The initial values for the number of endemics and non-endemics. In `DAISIE_probdist()` or `DAISIE_margprobdist()` either this or initprobs must be NULL. In `DAISIE_numcol()` when it is NULL, it is assumed that the island is empty.

Value

`out`	A list of three vectors: `expC` The expectation of the number of colonizations/lineages at the given times `pC` The probability distribution of the number of colonizations (lineages) at the given times

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Compute the marginal probability distributions at t = 4 and t = 8, for a mainland
# pool size of 250 potential colonists and a vector of 5 parameters (cladogenesis,
# extinction, clade-level carrying capacity, immigration, anagenesis) starting from
# an empty island

numcol <- DAISIE_numcol(
   pars1 = c(0.3,0.35,Inf,0.75,0.012),
   pars2 = c(100,250),
   tvec = c(4,8),
   initEI = list(c(0,1),c(0,2),c(3,1))
   )
```

`daisie_odeint_cs` *Driver for the boost::odeint solver for the CS model*

Description

Driver for the boost::odeint solver for the CS model

Value

Object of type ‘state_type‘, which itself is ‘vector_t‘<double>, with the result of the integration depending on the runmod chosen.

`daisie_odeint_iw` *Driver for the boost::odeint solver for the IW model*

Description

Driver for the boost::odeint solver for the IW model

`DAISIE_plot_age_diversity` *Plot clade age against clade diversity.*

Description

Plots clade age against clade diversity for all clades for which colonisation time is known.

Usage

```
DAISIE_plot_age_diversity(
  island,
  title = "Clade age vs clade diversity",
  island_age = NA
)
```

Arguments

`island`	Island data object. Can be in DAISIE list format (see Galapagos_datalist and DAISIE_data_prep for examples) or in table format (see Galapagos_datatable for an example).
`title`	Title of the plot
`island_age`	Age of island in appropriate units. In `DAISIE_plot_age_diversity()` and `DAISIE_plot_island()` if island input is in table format, the age of the island must be specified. If island input is in DAISIE list format, this option will override the island age specified in the island list.

Details

R plot showing for each clade in the island object the time of colonisation and the diversity of the clade. Only clades for which colonisation time is known are plotted. Blue - endemic; black - non-endemic. C = number of independent colonisations on island, N = number of species on the island

Value

R plot.

Author(s)

Luis Valente

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr`, `DAISIE_sim_time_dep`, `DAISIE_sim_cr_shift`, `DAISIE_plot_island`

Examples

```
### Plot Galapagos age-diversity for Galapagos dataset
utils::data(Galapagos_datalist)
DAISIE_plot_age_diversity(Galapagos_datalist)
```

`DAISIE_plot_input` *DAISIE tree plot*

Description

Shows the phylogenies of the multiple clades present on the island

Usage

```
DAISIE_plot_input(
  trees,
  age = NULL,
  tcols = NULL,
  metadata = NULL,
  mapping = NULL,
  xlen = 0.001,
  pargs = NULL,
  bckgd = "white"
)
```

Arguments

`trees`	A list of trees, in phylo format, named after their respective clades
`age`	The age of the island, on the same scale as the branch lengths of the trees. If unspecified, the depth of the deepest stem across trees.
`tcols`	A vector of colonization times, with one value per clade. Order must be the same as in ‘trees‘. If unspecified, all clades are assigned the age of the island as colonization time.
`metadata`	Optional data frame with clade-level metadata. One column must be named "clade".
`mapping`	Optional aesthetic mapping to apply to the trees, as returned by the ‘ggplot2::aes‘ function. Mapped variables can be anything in the columns of the ‘data‘ node-wise data frame associated to the ‘ggtree‘ plot being created (e.g. node, label, clade, mrca) or anything in the columns of the clade-wise ‘metadata‘, if provided (in this case the aesthetics is mapped to all nodes within each clade).
`xlen`	Length of the extra tips grafted to each tree at the island age. These are a hack for scaling the plot. Keep this value small.
`pargs`	Optional arguments to be passed to ‘geom_point‘ when plotting points at colonization events (e.g. size, shape...).
`bckgd`	Optional background color of the figure. This is because we use rectangles as a hack to hide tree branches prior to island colonization. Default to white background.

Value

A ‘ggtree‘ plot, which is also a ‘ggplot‘ object. The output is fully customizable, as any ‘ggplot‘ object.

Author(s)

Raphael Scherrer (github.com/rscherrer)

Examples

```
# check whether package tibble has been installed
if (!requireNamespace('tibble', quietly = TRUE)) {
   cat("Package tibble needed for this function to work. Please install it.")
} else {
set.seed(42)

# Random trees
t1 <- ape::rtree(10)
t1$tip.label <- gsub("t", "t1.", t1$tip.label)
t2 <- ape::rtree(3)
t2$tip.label <- gsub("t", "t2.", t2$tip.label)
t3 <- ape::rtree(2)
trees <- list(t1, t2, t3)
names(trees) <- c("A", "B", "C")
```

```
  # Toy colonization events for each clade
  tcols <- c(4.5, 5, 4.6)

  # Toy metadata
  metadata <- tibble::tibble(
    clade = names(trees),
    endemic = TRUE,  # whether each clade is endemic
    uncertain = FALSE  # whether colonization time is known for sure
  )
  metadata$endemic[3] <- FALSE
  metadata$uncertain[2] <- TRUE

  # Island age
  age <- 5

  # Make a plot
  p <- DAISIE_plot_input(
    trees,
    age,
    tcols,
    metadata,
    mapping = ggplot2::aes(color = endemic, linetype = uncertain),
    pargs = list(size = 3)
  )
 p
 }
```

DAISIE_plot_island *Plot colonisation and branching time of species found on an island dataset.*

Description

Produces an image with the times of colonisation, branching times, clade name and clade status for all clades found in a given dataset.

Usage

```
DAISIE_plot_island(island, island_age = NA)
```

Arguments

`island`	Island data object. Can be in DAISIE list format (see Galapagos_datalist and DAISIE_data_prep for examples) or in table format (see Galapagos_datatable for an example).
`island_age`	Age of island in appropriate units. In `DAISIE_plot_age_diversity()` and `DAISIE_plot_island()` if island input is in table format, the age of the island must be specified. If island input is in DAISIE list format, this option will override the island age specified in the island list.

Details

R plot showing for each clade in the island object: time of colonisation, branching times, species status (endemic or non-endemic), total number of species in clade (n), number of species not sampled (m).

Value

R plot.

Author(s)

Luis Valente

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr`, `DAISIE_sim_time_dep`, `DAISIE_sim_cr_shift`, `DAISIE_plot_age_diversity`

Examples

```
### Plot Galapagos islands dataset from data table (specify island age)

utils::data(Galapagos_datatable)
DAISIE_plot_island(Galapagos_datatable, island_age=4)

### Plot Galapagos islands dataset from datalist (no need to specify
## island age)
utils::data(Galapagos_datalist)
DAISIE_plot_island(Galapagos_datalist)
```

`DAISIE_plot_sims` *Plot island species-through-time (STT) plots*

Description

Produces STT plots. If only one type of species is present in the simulated islands, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.

R plots with number of total, endemic and non-endemic STTs for different types of species for the entire time span the islands were simulated. 2.5-97.5th percentiles are plotted in light grey, 25-75th percentiles plotted in dark grey.

Usage

```
DAISIE_plot_sims(
  island_replicates,
  plot_plus_one = TRUE,
  type = "all_species",
  sample_freq = 25,
  trait_pars = NULL
)
```

Arguments

`island_replicates`	List output from `DAISIE_sim_core_cr()`, `DAISIE_sim_core_time_dep()`, `DAISIE_sim_core_cr_shif` or `DAISIE_sim_min_type2()` functions. Minimally, this must be a list that has as many elements as replicates. Each element must be a list with the elements `island_age`, `not_present` and `stt_all`. `stt_all` must be a data frame with the column names `Time`, `nI`, `nA`, `nC` and `present`.
`plot_plus_one`	Boolean to indicate to plot all values plus one. Set to `TRUE` for default behavior. Set to `FALSE` to plot all values without adding one. Only works when there is one type of species.
`type`	String to indicate if stt of all species or all possible stt should be plotted. Default is `"all_species"`, `"type1_species"` or `"type2_species"` should be plotted.
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`trait_pars`	A named list containing diversification rates considering two trait states created by `create_trait_pars`: • [1]:A numeric with the per capita transition rate with state 1 • [2]:A numeric with the per capita immigration rate with state 2 • [3]:A numeric with the per capita extinction rate with state 2 • [4]:A numeric with the per capita anagenesis rate with state 2 • [5]:A numeric with the per capita cladogenesis rate with state 2 • [6]:A numeric with the per capita transition rate with state 2 • [7]:A numeric with the number of species with trait state 2 on mainland

Value

R plot.

Author(s)

Luis Valente

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr`, `DAISIE_sim_time_dep`, `DAISIE_sim_cr_shift`, `DAISIE_format_CS`

Examples

```
### Plot islands with single process (only one type of species)
utils::data(islands_1type_1000reps)
DAISIE_plot_sims(
  island_replicates = islands_1type_1000reps
  )


### Plot island with type 1 and type 2
utils::data(islands_2types_1000reps)
DAISIE_plot_sims(
  island_replicates = islands_2types_1000reps
  )
```

`DAISIE_probdist` *The joint distribution of endemics and non-endemics under the DAISIE model*

Description

This function calculates the joint distribution of the number of endemics and non-endemics for a given set of parameter values, a given mainland species pool size and a given set of times

Usage

```
DAISIE_probdist(pars1, pars2, tvec, initEI = c(0, 0), initprobs = NULL)
```

Arguments

`pars1`	Vector of model parameters: `pars1[1]` corresponds to lambda^c (cladogenesis rate) `pars1[2]` corresponds to mu (extinction rate) `pars1[3]` corresponds to K (clade-level carrying capacity) `pars1[4]` corresponds to gamma (immigration rate) `pars1[5]` corresponds to lambda^a (anagenesis rate).
`pars2`	Vector of settings: `pars2[1]` corresponds to res, the maximum number of endemics or non-endemics for which the ODE system is solved; this must be much larger than the actual

	number for which the probability needs to be calculated.) `pars2[2]` corresponds to M, size of the mainland pool, i.e the number of species that can potentially colonize the island.
`tvec`	The times at which the probabilities need to be computed.
`initEI`	The initial values for the number of endemics and non-endemics. In `DAISIE_probdist()` or `DAISIE_margprobdist()` either this or initprobs must be NULL. In `DAISIE_numcol()` when it is NULL, it is assumed that the island is empty.
`initprobs`	The initial probability distribution for the number of endemics and non-endemics; either this or initEI must be NULL.

Details

To obtain a matrix of probabilities with endemics in rows and non-endemics in columns for a certain time, one can run DAISIE_convertprobdist

Value

A matrix of dimensions 1 + length(tvec) and pars[2]^2 + 1] where the first column contains the times at which the probabilities are evaluated and the other columns contain the joint probabilities.

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Examples

```
### Compute the probability distribution at t = 4 and t = 8, for a mainland pool
# size of 250 potential colonists and a vector of 5 parameters (cladogenesis,
# extinction, clade-level carrying capacity, immigration, anagenesis) starting
# from an empty island

prob_dists <- DAISIE_probdist(
   pars1 = c(0.3,0.35,Inf,0.75,0.012),
   pars2 = c(100,250),
   tvec = c(4,8),
   initEI = c(0,0),
   initprobs = NULL
   )
```

DAISIE_sim *Simulate (non-)oceanic islands with given parameters under time-constant rates*

Description

This function simulates islands with given cladogenesis, extinction, Kprime, immigration and anagenesis parameters, all of which modelled as time-constant parameters. If a single parameter set is provided (5 parameters) it simulates islands where all species have the same macro-evolutionary process. If two paramater sets (10 parameters) are provided, it simulates islands where two different macro-evolutionary processes operate, one applying to type 1 species and other to type 2 species. Further, it allows for the simulation of non-oceanic islands, generating islands for which the starting condition includes potential endemic and non-endemic species.

Usage

```
DAISIE_sim_cr(
  time,
  M,
  pars,
  replicates,
  divdepmodel = "CS",
  nonoceanic_pars = c(0, 0),
  num_guilds = NULL,
  prop_type2_pool = NA,
  replicates_apply_type2 = TRUE,
  sample_freq = 25,
  plot_sims = TRUE,
  hyper_pars = create_hyper_pars(d = 0, x = 0),
 area_pars = create_area_pars(max_area = 1, current_area = 1, proportional_peak_t = 0,
    total_island_age = 0, sea_level_amplitude = 0, sea_level_frequency = 0,
    island_gradient_angle = 0),
  cond = 0,
  verbose = TRUE,
  files_to_write = FALSE,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters:

- `pars[1]`: lambda^c (cladogenesis rate)
- `pars[2]`: mu (extinction rate)
- `pars[3]`: K (carrying capacity), set K=Inf for diversity independence.
- `pars[4]`: gamma (immigration rate)
- `pars[5]`: lambda^a (anagenesis rate)
- `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model
- `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model
- `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence.
- `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model
- `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.

`replicates`	Integer specifying number of island replicates to be simulated.
`divdepmodel`	Option divdepmodel = 'CS' runs a model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist. Option divdepmodel = 'IW' runs a model with island-wide carrying capacity, where diversity-dependence operates within and among clades. Option divdepmodel = 'GW' runs a model with diversity-dependence operates within a guild.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`num_guilds`	The number of guilds on the mainland. The number of mainland species is divided by the number of guilds when `divdepmodel = "GW"`
`prop_type2_pool`	Fraction of mainland species that belongs to the second subset of species (type 2). Applies only when two types of species are simulated (length(pars) = 10). For `DAISIE_dataprep()` applies only if number_clade_types = 2. In `DAISIE_dataprep()` the default `"proportional"` sets the fraction to be proportional to the number of clades of distinct macroevolutionary process that have colonised the island.
`replicates_apply_type2`	Applies only when two types of species are being simulated. Default replicates_apply_type2 = TRUE runs simulations until the number of islands where a type 2 species has colonised is equal to the specified number of replicates. This is recommended if `prop_type2_pool` is small or if the rate of immigration

	of type two species (`pars[9]`) is low, meaning that more replicates are needed to achieved an adequate sample size of islands with type 2 species. Setting `replicates_apply_type2 = FALSE` simulates islands up to the specified number of replicates regardless of whether type 2 species have colonised or not.
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`plot_sims`	`Default = TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and 'DAISIEdataprep()' can be '1' or '0', where '1' gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: '0' does not print, '1' prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), '2' prints the same as '1 and also the intermediate output of the parameters and loglikelihood, while '3' the same as '2' and prints intermediate progress during likelihood computation.
`files_to_write`	number of files to write simulations to file
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replciate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: the number of mainland lineages that are not present on the island. It is only present if only 1 type of species is simulated. Becomes `$not_present_type1`: the number of mainland lineages of type 1 that are not present on the island and `$not_present_type2`: the number of mainland lineages of type 2 that are not present on the island, if two types are simulated.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$stt_stt_type1`: STT table for type 1 species on the island - only if 2 types of species were simulated (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present).
- `$stt_stt_type2`: STT table for type 2 species on the island - only if 2 types of species were simulated (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present).
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species.

  For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:
  1. Non_endemic_MaxAge
  2. Endemic
  3. Endemic&Non_Endemic
  4. Non_endemic_MaxAge
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente, Albert Phillimore, Joshua Lambert, Shu Xie, Pedro Neves, Richèl J. C. Bilderbeek, Rampal Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_plot_sims()` for plotting STT of simulation outputs.

Other simulation models: `DAISIE_sim_cr_shift()`, `DAISIE_sim_relaxed_rate()`, `DAISIE_sim_time_dep()`

Examples

```
## Simulate 2 islands for 1 million years, where all species have equal
## rates. Pool size 100.

clado_rate <- 0.5
ext_rate <- 0.2
carr_cap <- Inf
immig_rate <- 0.05
ana_rate <- 1
sim_pars <- c(clado_rate, ext_rate, carr_cap, immig_rate, ana_rate)
set.seed(1)
island_replicates <- DAISIE_sim_cr(
  time = 1,
  M = 100,
  pars = sim_pars,
  replicates = 2,
  plot_sims = FALSE,
  verbose = FALSE
)

## Simulate 2 islands for 1 million years with two types of species (type1
## and type 2). Pool size 100
## Fraction of type 2 species in source pool is 0.15. Function will
## simulate until number of islands where type 2 species has colonised is
## equal to number specified in replicates.

clado_rate <- 0.5
ext_rate <- 0.2
carr_cap <- Inf
immig_rate <- 0.005
ana_rate <- 1
sim_pars_type1 <- c(clado_rate, ext_rate, carr_cap, immig_rate, ana_rate)
sim_pars_type2 <- sim_pars_type1 * 2
set.seed(1)
island_replicates_2types <- DAISIE_sim_cr(
  time = 1,
  M = 100,
  pars = c(sim_pars_type1, sim_pars_type2),
  replicates = 2,
  prop_type2_pool = 0.15,
  plot_sims = FALSE,
  verbose = FALSE
)
## Simulate two non-oceanic island for 1 million years.
## Pool size 500. Island area as a proportion
## of mainland is 0.1, proportion of native species is 0.9.
clado_rate <- 0.5
ext_rate <- 0.2
carr_cap <- Inf
immig_rate <- 0.005
ana_rate <- 1
sim_pars <- c(clado_rate, ext_rate, carr_cap, immig_rate, ana_rate)
```

```
set.seed(1)
island_replicates <- DAISIE_sim_cr(
  time = 1,
  M = 500,
  pars = sim_pars,
  replicates = 2,
  nonoceanic_pars = c(0.1, 0.9),
  plot_sims = FALSE,
  verbose = FALSE
)

## Simulate 2 islands for 1 million years with a shift in immigration rate
## at 0.195 Ma, and plot the species-through-time plot. Pool size 296.

pars_before_shift <- c(0.079, 0.973, Inf, 0.136, 0.413)
pars_after_shift <- c(0.079, 0.973, Inf, 0.652, 0.413)
tshift <- 0.195
set.seed(1)
island_shift_replicates <- DAISIE_sim_cr_shift(
  time = 1,
  M = 296,
  pars = c(pars_before_shift, pars_after_shift),
  replicates = 2,
  shift_times = tshift,
  plot_sims = FALSE,
  verbose = FALSE
)
```

`DAISIE_sim_cr_cs` *Simulates island replicates with an clade-specific (CS) diversity-dependent constant-rate process*

Description

Simulates island replicates with an clade-specific (CS) diversity-dependent constant-rate process

Usage

```
DAISIE_sim_cr_cs(
  total_time,
  M,
  pars,
  replicates,
  nonoceanic_pars,
  prop_type2_pool,
  replicates_apply_type2,
  sample_freq,
  hyper_pars,
  area_pars,
```

```
    cond,
    verbose,
    files_to_write = 0
  )
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`prop_type2_pool`	Fraction of mainland species that belongs to the second subset of species (type 2). Applies only when two types of species are simulated (length(pars) = 10). For `DAISIE_dataprep()` applies only if number_clade_types = 2. In `DAISIE_dataprep()` the default `"proportional"` sets the fraction to be proportional to the number of clades of distinct macroevolutionary process that have colonised the island.

`replicates_apply_type2`	Applies only when two types of species are being simulated. Default replicates_apply_type2 = TRUE runs simulations until the number of islands where a type 2 species has colonised is equal to the specified number of replicates. This is recommended if `prop_type2_pool` is small or if the rate of immigration of type two species (`pars[9]`) is low, meaning that more replicates are needed to achieved an adequate sample size of islands with type 2 species. Setting `replicates_apply_type2 = FALSE` simulates islands up to the specified number of replicates regardless of whether type 2 species have colonised or not.
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`files_to_write`	number of files to write simulations to file

Value

A list. The highest level of the least corresponds to each individual replicate. See return for ‘DAISIE_sim_cr()‘ for details.

`DAISIE_sim_cr_gw` *Simulates island replicates with an guild-wide (GW) diversity-dependent process*

Description

Simulates island replicates with an guild-wide (GW) diversity-dependent process

Usage

```
DAISIE_sim_cr_gw(
  total_time,
  M,
  pars,
  replicates,
  nonoceanic_pars,
  num_guilds,
  sample_freq,
  hyper_pars,
  area_pars,
  verbose
)
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

	Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`num_guilds`	The number of guilds on the mainland. The number of mainland species is divided by the number of guilds when `divdepmodel = "GW"`
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

A list. The highest level of the least corresponds to each individual replciate. See return for ‘DAISIE_sim_cr()‘ for details.

`DAISIE_sim_cr_iw` *Simulates island replicates with an island-wide (IW) diversity-dependent constant-rate process*

Description

Simulates island replicates with an island-wide (IW) diversity-dependent constant-rate process

Usage

```
DAISIE_sim_cr_iw(
  total_time,
  M,
  pars,
  replicates,
  nonoceanic_pars,
  sample_freq,
  hyper_pars,
  area_pars,
  cond,
  verbose
)
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

	Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

A list. The highest level of the least corresponds to each individual replicate. See return for ‘DAISIE_sim_cr()‘ for details.

`DAISIE_sim_cr_shift` *Simulate (non-)oceanic islands with given parameters under a rate-shift regime*

Description

This function simulates islands with given cladogenesis, extinction, Kprime, immigration and anagenesis parameters, all of which modelled as time-constant parameters, which can be switched to a different diversification regime (i.e., different set of parameters) at one or more set times before the present. Further, it allows for the simulation of non-oceanic islands, generating islands for which the starting condition includes potential endemic and non-endemic species.

Usage

```
DAISIE_sim_cr_shift(
  time,
  M,
  pars,
  replicates,
  shift_times,
  divdepmodel = "CS",
  nonoceanic_pars = c(0, 0),
  num_guilds = NULL,
  sample_freq = 25,
  plot_sims = TRUE,
  hyper_pars = create_hyper_pars(d = 0, x = 0),
  area_pars = DAISIE::create_area_pars(max_area = 1, current_area = 1,
    proportional_peak_t = 0, total_island_age = 0, sea_level_amplitude = 0,
    sea_level_frequency = 0, island_gradient_angle = 0),
  cond = 0,
  verbose = TRUE,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence.

	• `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`shift_times`	a numeric vector specifying when the rate shifts occur before the present.
`divdepmodel`	Option divdepmodel = 'CS' runs a model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist. Option divdepmodel = 'IW' runs a model with island-wide carrying capacity, where diversity-dependence operates within and among clades. Option divdepmodel = 'GW' runs a model with diversity-dependence operates within a guild.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`num_guilds`	The number of guilds on the mainland. The number of mainland species is divided by the number of guilds when `divdepmodel = "GW"`
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`plot_sims`	`Default = TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate

`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replciate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: A numeric with the number of mainland lineages that are not present on the island.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:

    1. Non_endemic_MaxAge
    2. Endemic
    3. Endemic&Non_Endemic
    4. Non_endemic_MaxAge
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente, Albert Phillimore, Torsten Hauffe

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Hauffe, T., D. Delicado, R.S. Etienne and L. Valente (2020). Lake expansion elevates equilibrium diversity via increasing colonization.

See Also

`DAISIE_plot_sims()` for plotting STT of simulation outputs.

Other simulation models: `DAISIE_sim()`, `DAISIE_sim_relaxed_rate()`, `DAISIE_sim_time_dep()`

`DAISIE_sim_MW` *Simulate multiple islands based on hyperparameters that describe relationships between area, isolation and local parameters*

Description

This function simulates islands of given age, area and isolation. A list of islands/archipelagos with this information is provided as a data frame object (in the example, the archipelago_data from Valente et al 2020 Nature). For each island, local parameters are first calculated based on hyperparameters describing the dependency of rates on island area and isolation. Simulations are then run for each island, from island birth until the given island age based on the local parameters for each island (using the DAISIE_sim function).

Returns R list object that contains the simulated islands.

Usage

```
DAISIE_sim_MW(
  archipelago_data,
  M,
  pars,
  replicates,
  divdepmodel = "CS",
```

```
  distance_dep = "power",
  cladogenesis_dep = "NULL",
  sigmoidal_par = "NULL"
)
```

Arguments

`archipelago_data`	Object (data.frame) containing a table with columns with the following names: Archipelago: names of the archipelagos or islands to be simulated; Area - the area of the archipelago/island; Age: Age of the oldest island in the archipelago, or age of the island if single island. Distance - distance to the mainland (or other applicable isolation metric). As an example, the dataset from Valente et al 2020 for 41 archipelagos is provided in data(archipelago_data).
`M`	The size of the mainland pool, i.e the number of species that can potentially colonize the island(s).
`pars`	Contains the model hyperparameters: `pars[1]` corresponds to lambda^c0 (initial cladogenesis rate) `pars[2]` corresponds to y (dependency of cladogenesis on area) 0 - no dependency, positive value - cladogenesis increases with area, negative value - cladogenesis decreases with area `pars[3]` corresponds to mu_0 (initial extinction rate) `pars[4]` corresponds to x (dependency of extinction on area) 0 - no dependency, positive value - extinction increases with area, negative value - extinction decreases with area. `pars[5]` corresponds to K_0 (initial carrying capacity). Set K=Inf for non-diversity dependence. `pars[6]` corresponds to z (dependency of K on area). 0 - no dependency, positive value - K increases with area, negative value - K decreases with area. `pars[7]` corresponds to gamma_0 (initial immigration rate) `pars[8]` corresponds to alpha (dependency of immigration on island isolation). 0 - no dependency, positive value - immigration increases with distance, negative value - immigration decreases with distance `pars[9]` corresponds to lambda^a_0 (initial anagenesis rate) `pars[10]` corresponds to beta (dependency of anagenesis on island isolation). 0 - no dependency,positive value - anagenesis increases with distance, negative value - anagenesis decreases with distance. `pars[11]` corresponds to d0, parameter that describes interactive/additive effect of isolation and area on cladogenesis (power models); or that describes the shape of the sigmoidal relationship between isolation and a parameter (sigmoidal models).
`replicates`	Number of island replicates to be simulated per island in the table archipelago_data. If there are 5 islands in archipelago_data, and number of replicates is set to 10, 50 islands will be simulated.
`divdepmodel`	Option divdepmodel = 'CS' runs model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist.

	Option divdepmodel = 'IW' runs model with island-wide carrying capacity, where diversity-dependence operates within and among clades.
`distance_dep`	Shape of dependency of parameter with distance. Can be of type 'power' or type 'sigmoidal'.
`cladogenesis_dep`	When distance_dep = 'power' is specified, cladogenesis_dep specifies the type of effect of area and isolation on cladogenesis. The possible options are: 'NULL' (no additive or interactive effect of area on isolation, same as models M1-M14 in Valente et al 2020) 'additive' (additive effect of area and isolation on cladogenesis, same as model M15 in Valente et al 2020) 'interactive' (interactive effect of area and isolation on cladogenesis, same as models M16 and M19 in Valente et al 2020) 'interactive1' (interactive effect of area and isolation on cladogenesis, same as model M17 in Valente et al 2020) 'interactive2' (interactive effect of area and isolation on cladogenesis, same as model M18 in Valente et al 2020)
`sigmoidal_par`	When distance_dep = 'sigmoidal', sigmoidal_par specifies to which parameter the sigmoidal relationship with distance is applied. Options: 'cladogenesis' 'anagenesis' 'colonisation'.

Value

For each island listed in the rows of archipelago_data, a given number of islands is simulated depending on the number of replicates specified. The simulations for each island in archipelago_data are an element of the list, which can be called using [[x]]. Individual replicates for an island can be called with [[x]][[x]] For example if the object is called global_sims, the 1st replicate for the first island can be called using global_sims[[1]][[1]]. The 3rd replicate of the 4th island is called with global_sims[[4]][[3]] Each of the island replicates is a list in itself. The first (e.g. global_sims[[x]][[x]][[1]]) element of that list has the following components:
`$island_age` - the island or archipelago age `$not_present` - the number of mainland lineages that are not present on the island
`$stt_all` - STT table for all species on the archipelago/island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
`$area` Area of the island or archipelago
`$distance` Distance to the mainland
`$name` Name of the island or archipelago

The subsequent elements of the list each contain information on a single colonist lineage on the island and has 4 components:
`$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic with no close extant relatives on the islands species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.

`$stac` - the status of the colonist
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)

Author(s)

Luis Valente, Albert Phillimore, Rampal Etienne

References

Valente, LM, Phillimore AB, Melo M, Warren B, Clegg S, Havenstein K, Tiedemann R, Illera JC, Thebaud C, Aschenbach T and Etienne RS (2020). A simple dynamic model explain island bird diversity worldwide. Nature, 579, 92-96.

See Also

`DAISIE_sim_cr()`.

Examples

```
## Simulate 10 replicates for the each of the 41 archipelagos in the archipelago_data table,
##based on the hyperparameters of the M19 model (preferred model in Valente et al 2020)
## and the age, area and isolation values given in the archipelago_data table.
## Mainland pool size of 1000, clade-specific carrying capacity. The M19 model is a
## power model with an interactive effect of area and isolation on cladogenesis.
##
## data(archipelago_data)
## result <- DAISIE_sim_MW(
## archipelago_data = archipelago_data,
## M = 1000,
## pars = c(0.040073803,0,1.945656546,0.150429656,Inf,0,67.25643672,
## 0.293635061,0.059096872,0.382688527,0.026510781),
## replicates = 10,
## distance_dep = 'power',
## cladogenesis_dep = 'interactive',
## sigmoidal_par = 'NULL',
## divdepmodel = 'CS')
```

`DAISIE_sim_relaxed_rate`

Simulate (non-)oceanic islands with given parameters under a relaxed-rate model

---

Description

This function simulates islands with given cladogenesis, extinction, Kprime, immigration and anagenesis parameters, all of which can be modelled as time-constant parameters with variation between clades in one or multiple parameters. Further, it allows for the simulation of non-oceanic islands, generating islands for which the starting condition includes potential endemic and non-endemic species.

Usage

```
DAISIE_sim_relaxed_rate(
  time,
  M,
  pars,
  replicates,
  relaxed_par,
  nonoceanic_pars = c(0, 0),
  sample_freq = 25,
  plot_sims = TRUE,
  hyper_pars = create_hyper_pars(d = 0, x = 0),
 area_pars = create_area_pars(max_area = 1, current_area = 1, proportional_peak_t = 0,
    total_island_age = 0, sea_level_amplitude = 0, sea_level_frequency = 0,
    island_gradient_angle = 0),
  cond = 0,
  verbose = TRUE,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence.

	- `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model - `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`relaxed_par`	A string determining which parameter is relaxed in a relaxed rate model.
`nonoceanic_pars`	A vector of length two with: - [1]: the probability of sampling a species from the mainland - [2]: the probability of the species sampled from the mainland being nonendemic
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`plot_sims`	`Default` = `TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: - [1]: is d the scaling parameter for exponent for calculating cladogenesis rate - [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: - [1]: maximum area - [2]: current area - [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved - [4]: total island age - [5]: amplitude of area fluctuation from sea level - [6]: frequency of sine wave of area change from sea level - [7]: angle of the slope of the island
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model

`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replciate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: the number of mainland lineages that are not present on the island. It is only present if only 1 type of species is simulated. Becomes `$not_present_type1`: the number of mainland lineages of type 1 that are not present on the island and `$not_present_type2`: the number of mainland lineages of type 2 that are not present on the island, if two types are simulated.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$stt_stt_type1`: STT table for type 1 species on the island - only if 2 types of species were simulated (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present).
- `$stt_stt_type2`: STT table for type 2 species on the island - only if 2 types of species were simulated (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present).
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species.

  For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:
  1. Non_endemic_MaxAge
  2. Endemic
  3. Endemic&Non_Endemic
  4. Non_endemic_MaxAge
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente, Albert Phillimore, Joshua Lambert, Shu Xie, Pedro Neves, Richèl J. C. Bilderbeek, Rampal Etienne

See Also

`DAISIE_plot_sims()` for plotting STT of simulation outputs.

Other simulation models: `DAISIE_sim()`, `DAISIE_sim_cr_shift()`, `DAISIE_sim_time_dep()`

Examples

```
## Simulate an island for 1 million years, with a relaxed the rate of
## cladogenesis between clades. Pool size 500.

clado_rate <- 0.5
ext_rate <- 0.2
carr_cap <- Inf
immig_rate <- 0.005
ana_rate <- 1
par_sd <- 1
sim_pars <- c(clado_rate, ext_rate, carr_cap, immig_rate, ana_rate, par_sd)
set.seed(1)
island_replicates <- DAISIE_sim_relaxed_rate(
  time = 1,
  M = 500,
  pars = sim_pars,
  replicates = 2,
  relaxed_par = "cladogenesis",
  plot_sims = FALSE,
  verbose = FALSE
)
```

`DAISIE_sim_time_dep` *Simulate (non-)oceanic islands with given parameters under a time-dependent regime*

Description

This function simulates islands with given cladogenesis, extinction, Kprime, immigration and anagenesis parameters, all of which modelled as time-dependent parameters.

Time dependency aims to capture the effect of area changes islands undego from their emergence until subsidence. Thus, oceanic, volcanic island ontogeny scenarios can be modelled (by a beta function), as well as the effect of sea level fluctuations (modelled through a sine function). See paramter entry `area_pars` for details. Both island ontogeny and sea level fluctuations are allowed to operate simultaneuosly.

This function also allows for the simulation of non-oceanic islands, generating islands for which the starting condition includes potential endemic and non-endemic species.

Usage

```
DAISIE_sim_time_dep(
  time,
  M,
  pars,
  replicates,
  area_pars,
  hyper_pars,
  divdepmodel = "CS",
  nonoceanic_pars = c(0, 0),
  num_guilds = NULL,
  sample_freq = 25,
  plot_sims = TRUE,
  island_ontogeny = "const",
  sea_level = "const",
  extcutoff = 1000,
  cond = 0,
  verbose = TRUE,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

	Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`divdepmodel`	Option divdepmodel = 'CS' runs a model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist. Option divdepmodel = 'IW' runs a model with island-wide carrying capacity, where diversity-dependence operates within and among clades. Option divdepmodel = 'GW' runs a model with diversity-dependence operates within a guild.
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`num_guilds`	The number of guilds on the mainland. The number of mainland species is divided by the number of guilds when `divdepmodel = "GW"`
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`plot_sims`	`Default = TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time.

	In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny` = `NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replciate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: A numeric with the number of mainland lineages that are not present on the island.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.

- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:
    1. Non_endemic_MaxAge
    2. Endemic
    3. Endemic&Non_Endemic
    4. Non_endemic_MaxAge
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente and Albert Phillimore

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

Valente, L.M., Etienne, R.S. and Phillimore, A.B. (2014). The effects of island ontogeny on species diversity and phylogeny. Proceedings of the Royal Society B: Biological Sciences 281(1784), p.20133227.

See Also

`DAISIE_plot_sims()` for plotting STT of simulation outputs.

Other simulation models: `DAISIE_sim()`, `DAISIE_sim_cr_shift()`, `DAISIE_sim_relaxed_rate()`

`DAISIE_sim_time_dep_cs` *Simulates island replicates with an clade-specific (CS) diversity-dependent time-dependent process*

Description

Simulates island replicates with an clade-specific (CS) diversity-dependent time-dependent process

Usage

```
DAISIE_sim_time_dep_cs(
  total_time,
  M,
  pars,
  replicates,
  area_pars,
  hyper_pars,
  nonoceanic_pars,
```

```
    sample_freq,
    island_ontogeny,
    sea_level,
    peak,
    Amax,
    Amin,
    extcutoff,
    cond,
    verbose
  )
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level

	• [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`peak`	A numeric value specifying the peakiness (or shaprness) of the ontogeny curve. Higher values imply peakier ontogeny. This value is internally calculated by `calc_peak()` given the area at the present and the `area_pars`.
`Amax`	A numeric value for maximum island area during the simulation.
`Amin`	A numeric value for minimum island area during the simulation.
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed.

	The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

A list. The highest level of the least corresponds to each individual replicate. See return for ‘DAISIE_sim_time_dep()‘ for details.

`DAISIE_sim_time_dep_gw`

Simulates island replicates with an guild-wide (GW) diversity-dependent time-dependent process

Description

Simulates island replicates with an guild-wide (GW) diversity-dependent time-dependent process

Usage

```
DAISIE_sim_time_dep_gw(
  total_time,
  M,
  pars,
  replicates,
  area_pars,
  hyper_pars,
  nonoceanic_pars,
  num_guilds,
  sample_freq,
  island_ontogeny,
  sea_level,
  peak,
  Amax,
  Amin,
  extcutoff,
  verbose
)
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters:

- `pars[1]`: lambda^c (cladogenesis rate)
- `pars[2]`: mu (extinction rate)
- `pars[3]`: K (carrying capacity), set K=Inf for diversity independence.
- `pars[4]`: gamma (immigration rate)
- `pars[5]`: lambda^a (anagenesis rate)
- `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model
- `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model
- `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence.
- `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model
- `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.

`replicates`	Integer specifying number of island replicates to be simulated.
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`num_guilds`	The number of guilds on the mainland. The number of mainland species is divided by the number of guilds when `divdepmodel = "GW"`

`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`peak`	A numeric value specifying the peakiness (or shaprness) of the ontogeny curve. Higher values imply peakier ontogeny. This value is internally calculated by `calc_peak()` given the area at the present and the `area_pars`.
`Amax`	A numeric value for maximum island area during the simulation.
`Amin`	A numeric value for minimum island area during the simulation.
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

A list. The highest level of the least corresponds to each individual replicate. See return for ‘DAISIE_sim_time_dep()‘ for details.

`DAISIE_sim_time_dep_iw` *Simulates island replicates with an island-wide (IW) diversity-dependent time-dependent process*

Description

Simulates island replicates with an island-wide (IW) diversity-dependent time-dependent process

Usage

```
DAISIE_sim_time_dep_iw(
  total_time,
  M,
  pars,
  replicates,
  area_pars,
  hyper_pars,
  nonoceanic_pars,
  sample_freq,
  island_ontogeny,
  sea_level,
  peak,
  Amax,
  Amin,
  extcutoff,
  cond,
  verbose
)
```

Arguments

`total_time`	Numeric defining the length of the simulation in time units.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.

`replicates`	Integer specifying number of island replicates to be simulated.
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`nonoceanic_pars`	A vector of length two with: • [1]: the probability of sampling a species from the mainland • [2]: the probability of the species sampled from the mainland being nonendemic
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`peak`	A numeric value specifying the peakiness (or shaprness) of the ontogeny curve. Higher values imply peakier ontogeny. This value is internally calculated by `calc_peak()` given the area at the present and the `area_pars`.
`Amax`	A numeric value for maximum island area during the simulation.
`Amin`	A numeric value for minimum island area during the simulation.
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.

`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.

Value

A list. The highest level of the least corresponds to each individual replicate. See return for ‘DAISIE_sim_time_dep()‘ for details.

`DAISIE_sim_trait_dep` *Simulate islands with given trait-dependent parameters.*

Description

This function simulates islands with given cladogenesis, extinction, K, immigration and anagenesis parameters for binary trait states.

Returns R list object that contains the simulated islands

Usage

```
DAISIE_sim_trait_dep(
  time,
  M,
  pars,
  replicates,
  divdepmodel = "CS",
  sample_freq = 25,
  plot_sims = TRUE,
  island_ontogeny = "const",
  sea_level = "const",
  hyper_pars = create_hyper_pars(d = 0, x = 0),
  area_pars = DAISIE::create_area_pars(max_area = 1, current_area = 1,
    proportional_peak_t = 0, total_island_age = 0, sea_level_amplitude = 0,
    sea_level_frequency = 0, island_gradient_angle = 0),
  extcutoff = 1000,
  cond = 0,
```

```
  verbose = TRUE,
  trait_pars = NULL,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters: • `pars[1]`: lambda^c (cladogenesis rate) • `pars[2]`: mu (extinction rate) • `pars[3]`: K (carrying capacity), set K=Inf for diversity independence. • `pars[4]`: gamma (immigration rate) • `pars[5]`: lambda^a (anagenesis rate) • `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model • `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model • `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence. • `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model • `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.
`replicates`	Integer specifying number of island replicates to be simulated.
`divdepmodel`	Option divdepmodel = 'CS' runs a model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist. Option divdepmodel = 'IW' runs a model with island-wide carrying capacity, where diversity-dependence operates within and among clades. Option divdepmodel = 'GW' runs a model with diversity-dependence operates within a guild.
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.

`plot_sims`	`Default = TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and 'DAISIEdataprep()' can be '1' or '0', where '1' gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: '0' does not print, '1' prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), '2' prints the same as '1 and also the intermediate output of the parameters and

	loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`trait_pars`	A named list containing diversification rates considering two trait states created by `create_trait_pars`: • [1]:A numeric with the per capita transition rate with state 1 • [2]:A numeric with the per capita immigration rate with state 2 • [3]:A numeric with the per capita extinction rate with state 2 • [4]:A numeric with the per capita anagenesis rate with state 2 • [5]:A numeric with the per capita cladogenesis rate with state 2 • [6]:A numeric with the per capita transition rate with state 2 • [7]:A numeric with the number of species with trait state 2 on mainland
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replicate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: A numeric with the number of mainland lineages that are not present on the island.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente and Albert Phillimore

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852. Hauffe, T., D. Delicado, R.S. Etienne and L. Valente (submitted). Lake expansion increases equilibrium diversity via the target effect of island biogeography.

See Also

`DAISIE_format_CS` `DAISIE_plot_sims`

DAISIE_sim_trait_dep_2K

Simulate islands with given trait-dependent parameters.

Description

This function simulates islands with given cladogenesis, extinction, K, immigration and anagenesis parameters. In this version, rates and K are both trait-dependent.

Returns R list object that contains the simulated islands

Usage

```
DAISIE_sim_trait_dep_2K(
  time,
  M,
  pars,
  replicates,
  divdepmodel = "CS",
  sample_freq = 25,
  plot_sims = TRUE,
  island_ontogeny = "const",
  sea_level = "const",
  hyper_pars = create_hyper_pars(d = 0, x = 0),
  area_pars = DAISIE::create_area_pars(max_area = 1, current_area = 1,
    proportional_peak_t = 0, total_island_age = 0, sea_level_amplitude = 0,
    sea_level_frequency = 0, island_gradient_angle = 0),
  extcutoff = 1000,
  cond = 0,
  verbose = TRUE,
  trait_pars = NULL,
  ...
)
```

Arguments

`time`	Numeric defining the length of the simulation in time units. For example, if an island is known to be 4 million years old, setting time = 4 will simulate the entire life span of the island; setting time = 2 will stop the simulation at the mid-life of the island.
`M`	Numeric defining the size of mainland pool, i.e. the number of species that can potentially colonize the island.
`pars`	A numeric vector containing the model parameters:

- `pars[1]`: lambda^c (cladogenesis rate)
- `pars[2]`: mu (extinction rate)
- `pars[3]`: K (carrying capacity), set K=Inf for diversity independence.
- `pars[4]`: gamma (immigration rate)
- `pars[5]`: lambda^a (anagenesis rate)
- `pars[6]`: lambda^c (cladogenesis rate) for either type 2 species or rate set 2 in rate shift model
- `pars[7]`: mu (extinction rate) for either type 2 species or rate set 2 in rate shift model
- `pars[8]`: K (carrying capacity) for either type 2 species or rate set 2 in rate shift model, set K=Inf for diversity independence.
- `pars[9]`: gamma (immigration rate) for either type 2 species or rate set 2 in rate shift model
- `pars[10]`: lambda^a (anagenesis rate) for either type 2 species or rate set 2 in rate shift model

Elements 6:10 are required only when type 2 species are included or in the rate shift model. For `DAISIE_sim_relaxed_rate()` `pars[6]` is the standard deviation of the gamma distribution for the relaxed parameter and the parameter chosen by the `relaxed_par` argument is the mean of the gamma distribution for the relaxed parameter.

`replicates`	Integer specifying number of island replicates to be simulated.
`divdepmodel`	Option divdepmodel = 'CS' runs a model with clade-specific carrying capacity, where diversity-dependence operates only within single clades, i.e. only among species originating from the same mainland colonist. Option divdepmodel = 'IW' runs a model with island-wide carrying capacity, where diversity-dependence operates within and among clades. Option divdepmodel = 'GW' runs a model with diversity-dependence operates within a guild.
`sample_freq`	Numeric specifing the number of units times should be divided by for plotting purposes. Larger values will lead to plots with higher resolution, but will also run slower.
`plot_sims`	`Default = TRUE` plots species-through-time (STT) plots. It detects how many types of species are present. If only one type of species is present, STT is plotted for all species. If two types are present, three plots are produced: STT for all, STT for type 1 and STT for type 2.
`island_ontogeny`	In `DAISIE_sim_time_dep()`, `DAISIE_ML_CS` and plotting a string describing the type of island ontogeny. Can be `"const"`, `"beta"` for a beta function describing area through time. In all other functions a numeric describing the type of island ontogeny. Can be `0` for constant, `1` for a beta function describing area through time. In ML functions `island_ontogeny = NA` assumes constant ontogeny. Time dependent estimation is not yet available as development is still ongoing. Will return an error if called in that case.
`sea_level`	In `DAISIE_sim_time_dep()` and plotting a string describing the type of sea level. Can be `"const"` or `"sine"` for a sine function describing area through

	time. In all other functions a numeric describing the type of sea level. Can be `0` for constant, `1` for a sine function describing area through time.
`hyper_pars`	A named list of numeric hyperparameters for the rate calculations as returned by `create_hyper_pars()`: • [1]: is d the scaling parameter for exponent for calculating cladogenesis rate • [2]: is x the exponent for calculating extinction rate
`area_pars`	A named list containing area and sea level parameters as created by `create_area_pars()`: • [1]: maximum area • [2]: current area • [3]: value from 0 to 1 indicating where in the island's history the peak area is achieved • [4]: total island age • [5]: amplitude of area fluctuation from sea level • [6]: frequency of sine wave of area change from sea level • [7]: angle of the slope of the island
`extcutoff`	A numeric with the cutoff for the the maximum extinction rate preventing it from being too large and slowing down simulation.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota . cond > 1 : conditioning on island age and having at least cond colonizations on the island. This last option is not yet available for the IW model
`verbose`	A numeric vector of length 1, which in simulations and ‘DAISIEdataprep()‘ can be ‘1‘ or ‘0‘, where ‘1‘ gives intermediate output should be printed. For ML functions a numeric determining if intermediate output should be printed. The default: ‘0‘ does not print, ‘1‘ prints the initial likelihood and the settings that were selected (which parameters are to be optimised, fixed or shifted), ‘2‘ prints the same as ‘1 and also the intermediate output of the parameters and loglikelihood, while ‘3‘ the same as ‘2‘ and prints intermediate progress during likelihood computation.
`trait_pars`	A named list containing diversification rates considering two trait states created by `create_trait_pars`: • [1]:A numeric with the per capita transition rate with state 1 • [2]:A numeric with the per capita immigration rate with state 2 • [3]:A numeric with the per capita extinction rate with state 2 • [4]:A numeric with the per capita anagenesis rate with state 2 • [5]:A numeric with the per capita cladogenesis rate with state 2 • [6]:A numeric with the per capita transition rate with state 2 • [7]:A numeric with the number of species with trait state 2 on mainland
`...`	Any arguments to pass on to plotting functions.

Value

A list. The highest level of the least corresponds to each individual replciate. The first element of each replicate is composed of island information containing:

- `$island_age`: A numeric with the island age.
- `$not_present`: A numeric with the number of mainland lineages that are not present on the island.
- `$stt_all`: STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
- `$brts_table`: Only for simulations under `"IW"`. Table containing information on order of events in the data, for use in maximum likelihood optimization.).

The subsequent elements of the list pertaining to each replcate contain information on a single colonist lineage on the island and have 4 components:

- `$branching_times`: island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
- `$stac`: An integer ranging from 1 to 4 indicating the status of the colonist:
- `$missing_species`: number of island species that were not sampled for particular clade (only applicable for endemic clades)
- `$type_1or2`: whether the colonist belongs to type 1 or type 2

Author(s)

Luis Valente and Albert Phillimore

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852. Hauffe, T., D. Delicado, R.S. Etienne and L. Valente (submitted). Lake expansion increases equilibrium diversity via the target effect of island biogeography.

See Also

`DAISIE_format_CS` `DAISIE_plot_sims`

`DAISIE_SR_loglik_CS` *Computes the loglikelihood of the DAISIE model with clade-specific diversity-dependence given data and a set of model parameters that may shift at some time*

Description

Computes the loglikelihood of the DAISIE model with clade-specific diversity-dependence given colonization and branching times for lineages on an island, and a set of model parameters that may shift at some time

Usage

```
DAISIE_SR_loglik_CS(
  pars1,
  pars2,
  datalist,
  methode = "odeint::runge_kutta_cash_karp54",
  CS_version = list(model = 1, function_to_optimize = "DAISIE"),
  abstolint = 1e-16,
  reltolint = 1e-10
)
```

Arguments

`pars1` Contains the model parameters:

`pars1[1]` corresponds to lambda^c (cladogenesis rate)
`pars1[2]` corresponds to mu (extinction rate)
`pars1[3]` corresponds to K (clade-level carrying capacity)
`pars1[4]` corresponds to gamma (immigration rate)
`pars1[5]` corresponds to lambda^a (anagenesis rate)
`pars1[6]` corresponds to lambda^c (cladogenesis rate) after the shift
`pars1[7]` corresponds to mu (extinction rate) after the shift
`pars1[8]` corresponds to K (clade-level carrying capacity) after the shift
`pars1[9]` corresponds to gamma (immigration rate) after the shift
`pars1[10]` corresponds to lambda^a (anagenesis rate) after the shift
`pars1[11]` corresponds to the time of shift

`pars2` Contains the model settings

`pars2[1]` corresponds to lx = length of ODE variable x
`pars2[2]` corresponds to ddmodel = diversity-dependent model, model of diversity-dependence, which can be one of

ddmodel = 0 : no diversity dependence
ddmodel = 1 : linear dependence in speciation rate

	ddmodel = 11: linear dependence in speciation rate and in immigration rate ddmodel = 2 : exponential dependence in speciation rate ddmodel = 21: exponential dependence in speciation rate and in immigration rate `pars2[3]` corresponds to cond = setting of conditioning cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota cond > 1 : conditioning on island age and having at least cond colonizations on the island `pars2[4]` sets whether parameters and likelihood should be printed (1) or not (0)
`datalist`	Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements. The first element of the list has two or three components: `$island_age` - the island age Then, depending on whether a distinction between types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components: `$colonist_name` - the name of the species or clade that colonized the island `$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation. `$stac` - the status of the colonist - Non_endemic_MaxAge: 1 * Endemic: 2 - Endemic&Non_Endemic: 3 - Non_endemic: 4 - Endemic_MaxAge: 5 `$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`methode`	Method of the ODE-solver. See package deSolve for details. Default is "lsodes"
`CS_version`	a numeric or list. Default is CS_version = list(model = 1, function_to_optimize = 'DAISIE'), but for a relaxed-rate model the list can contain more elements: • model: the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model

- function_to_optimize: the DAISIE loglikelihood function that will be optimized. Options are: `"DAISIE"`, default, the full DAISIE loglikelihood `"DAISIE_approx"`, an approximate loglikelihood `"DAISIE_DE"`, an exact loglikelkhood for K = Inf based on the D-E approach
- integration_method: the method used to do integraion in the relaxed rate model. Options are: `'standard'` the default numerical integration `'MC'` Monte Carlo integration `'stratified'` using quantiles of the gamma distribution
- relaxed_par: the parameter to relax (integrate over) in the relaxed rate model. Options are `"cladogenesis"`, `"extinction"`, `"carrying_capacity"`, `"immigration"`, or `"anagenesis"`
- par_sd: standard deviation of the parameter to relax
- par_upper_bound upper bound of the parameter to relax
- seed: seed of the random number generator in case of 'MC'
- sample_size: size of sample in case of 'MC' or 'stratified'
- parallel: use parallel computing or not in case of 'MC' or 'stratified'
- n_cores: number of cores to use when run in parallel

`abstolint`	Absolute tolerance of the integration
`reltolint`	Relative tolerance of the integration

Details

The output is a loglikelihood value

Value

The loglikelihood

Author(s)

Rampal S. Etienne & Bart Haegeman

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_ML`, `DAISIE_sim_cr`

Examples

```
utils::data(Galapagos_datalist_2types)
pars1 = c(0.195442017,0.087959583,Inf,0.002247364,0.873605049,
          3755.202241,8.909285094,14.99999923,0.002247364,0.873605049,0.163)
pars2 = c(100,11,0,1)
DAISIE_loglik_all(pars1,pars2,Galapagos_datalist_2types)
```

`DAISIE_SR_ML` *Maximization of the loglikelihood under the DAISIE model with clade-specific diversity-dependence*

Description

This function computes the maximum likelihood estimates of the parameters of the DAISIE model with clade-specific diversity-dependence and a shift in parameters for data from lineages colonizing an island. It also outputs the corresponding loglikelihood that can be used in model comparisons.

The result of sort(c(idparsopt, idparsfix, idparsnoshift)) should be identical to c(1:10). If not, an error is reported that the input is incoherent. The same happens when the length of initparsopt is different from the length of idparsopt, and the length of parsfix is different from the length of idparsfix.
Including the 11th parameter (p_f) in either idparsopt or idparsfix (and therefore initparsopt or parsfix) is optional. If this parameter is not specified, then the information in the data is used, otherwise the information in the data is overruled.

Usage

```
DAISIE_SR_ML_CS(
  datalist,
  initparsopt,
  idparsopt,
  parsfix,
  idparsfix,
  idparsnoshift = 6:10,
  res = 100,
  ddmodel = 0,
  cond = 0,
  island_ontogeny = NA,
  tol = c(1e-04, 1e-05, 1e-07),
  maxiter = 1000 * round((1.25)^length(idparsopt)),
  methode = "odeint::runge_kutta_cash_karp54",
  optimmethod = "simplex",
  CS_version = list(model = 1, function_to_optimize = "DAISIE"),
  verbose = 0,
  tolint = c(1e-16, 1e-10),
  jitter = 0,
  num_cycles = 1
)
```

Arguments

`datalist`	Data object containing information on colonisation and branching times. This object can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.

	The first element of the list has two three components: `$island_age` - the island age Then, depending on whether a distinction between types is made, we have: `$not_present` - the number of mainland lineages that are not present on the island The remaining elements of the list each contains information on a single colonist lineage on the island and has 5 components: `$colonist_name` - the name of the species or clade that colonized the island `$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation. `$stac` - the status of the colonist - Non_endemic_MaxAge: 1 - Endemic: 2 - Endemic&Non_Endemic: 3 - Non_Endemic: 4 - Endemic_Singleton_MaxAge: 5 - Endemic_Clade_MaxAge: 6 - Endemic&Non_Endemic_Clade_MaxAge: 7 - Non_endemic_MaxAge_MinAge: 8 - Endemic_Singleton_MaxAge_MinAge: 9 `$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`initparsopt`	The initial values of the parameters that must be optimized
`idparsopt`	The ids of the parameters that must be optimized. The ids are defined as follows: id = 1 corresponds to lambda^c (cladogenesis rate) id = 2 corresponds to mu (extinction rate) id = 3 corresponds to K (clade-level carrying capacity) id = 4 corresponds to gamma (immigration rate) id = 5 corresponds to lambda^a (anagenesis rate) id = 6 corresponds to lambda^c (cladogenesis rate) after the shift id = 7 corresponds to mu (extinction rate) after the shift id = 8 corresponds to K (clade-level carrying capacity) after the shift id = 9 corresponds to gamma (immigration rate) after the shift id = 10 corresponds to lambda^a (anagenesis rate) after the shift id = 11 corresponds to the time of shift
`parsfix`	The values of the parameters that should not be optimized
`idparsfix`	The ids of the parameters that should not be optimized, e.g. c(1,3) if lambda^c and K should not be optimized.
`idparsnoshift`	The ids of the parameters that should not be different before and after the shift.

`res`	Sets the maximum number of species for which a probability must be computed, must be larger than the size of the largest clade
`ddmodel`	Sets the model of diversity-dependence: ddmodel = 0 : no diversity dependence ddmodel = 1 : linear dependence in speciation rate ddmodel = 11: linear dependence in speciation rate and in immigration rate ddmodel = 2 : exponential dependence in speciation rate ddmodel = 21: exponential dependence in speciation rate and in immigration rate.
`cond`	cond = 0 : conditioning on island age cond = 1 : conditioning on island age and non-extinction of the island biota
`island_ontogeny`	type of island ontonogeny. If NA, then constant ontogeny is assumed.
`tol`	Sets the tolerances in the optimization. Consists of: reltolx = relative tolerance of parameter values in optimization reltolf = relative tolerance of function value in optimization abstolx = absolute tolerance of parameter values in optimization.
`maxiter`	Sets the maximum number of iterations in the optimization.
`methode`	Method of the ODE-solver. See package deSolve for details. Default is "lsodes"
`optimmethod`	Method used in likelihood optimization. Default is "subplex" (see subplex package). Alternative is 'simplex' which was the method in previous versions.
`CS_version`	a numeric or list. Default is CS_version = list(model = 1, function_to_optimize = 'DAISIE'), but for a relaxed-rate model the list can contain more elements: • model: the CS model to run, options are `1` for single rate DAISIE model, `2` for multi-rate DAISIE, or `0` for IW test model • function_to_optimize: the DAISIE loglikelihood function that will be optimized. Options are: `"DAISIE"`, default, the full DAISIE loglikelihood `"DAISIE_approx"`, an approximate loglikelihood `"DAISIE_DE"`, an exact loglikelkhood for K = Inf based on the D-E approach • integration_method: the method used to do integraion in the relaxed rate model. Options are: `'standard'` the default numerical integration `'MC'` Monte Carlo integration `'stratified'` using quantiles of the gamma distribution • relaxed_par: the parameter to relax (integrate over) in the relaxed rate model. Options are `"cladogenesis"`, `"extinction"`, `"carrying_capacity"`, `"immigration"`, or `"anagenesis"` • par_sd: standard deviation of the parameter to relax • par_upper_bound upper bound of the parameter to relax • seed: seed of the random number generator in case of 'MC' • sample_size: size of sample in case of 'MC' or 'stratified' • parallel: use parallel computing or not in case of 'MC' or 'stratified' • n_cores: number of cores to use when run in parallel

`verbose`	sets whether parameters and likelihood should be printed (1) or not (0).
`tolint`	Vector of two elements containing the absolute and relative tolerance of the integration.
`jitter`	Numeric for `optimizer()`. Jitters the parameters being optimized by the specified amount which should be very small, e.g. 1e-5. Jitter when `link[subplex]{subplex}()` produces incorrect output due to parameter transformation.
`num_cycles`	The number of cycles the optimizer will go through. Default is 1.

Value

The output is a dataframe containing estimated parameters and maximum loglikelihood.

`lambda_c`	gives the maximum likelihood estimate of lambda^c, the rate of cladogenesis
`mu`	gives the maximum likelihood estimate of mu, the extinction rate
`K`	gives the maximum likelihood estimate of K, the carrying-capacity
`gamma`	gives the maximum likelihood estimate of gamma, the immigration rate
`lambda_a`	gives the maximum likelihood estimate of lambda^a, the rate of anagenesis
`lambda_c2`	gives the maximum likelihood estimate of lambda^c2, the rate of cladogenesis for the optional second group of species
`mu2`	gives the maximum likelihood estimate of mu2, the extinction rate for the optional second group of species
`K2`	gives the maximum likelihood estimate of K2, the carrying-capacity for the optional second group of species
`gamma2`	gives the maximum likelihood estimate of gamma2, the immigration rate for the optional second group of species
`lambda_a2`	gives the maximum likelihood estimate of lambda^a2, the rate of anagenesis for the optional second group of species
`loglik`	gives the maximum loglikelihood
`df`	gives the number of estimated parameters, i.e. degrees of feedom
`conv`	gives a message on convergence of optimization; conv = 0 means convergence

Author(s)

Rampal S. Etienne

References

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852. <DOI:10.1111/ele.12461>.

See Also

`DAISIE_loglik_all()`, `DAISIE_sim_cr`

Examples

```
## In all following DAISIE_ML calls very high tolerances and low system size
## are used  for fast computation for this example. Use default or better
## tol, tolint an res values in actual analyses.
##################
### When all species have the same rates, and we want to optimize all 5
### parameters, we use:
utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = Galapagos_datalist,
   initparsopt = c(2.5,2.7,20,0.009,1.01),
   ddmodel = 11,
   idparsopt = 1:5,
   parsfix = NULL,
   idparsfix = NULL,
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
)

### When all species have the same rates, and we want to optimize all parameters
# except K (which we set equal to Inf), we use:

utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = Galapagos_datalist,
   initparsopt = c(2.5,2.7,0.009,1.01),
   idparsopt = c(1,2,4,5),
   parsfix = Inf,
   idparsfix = 3,
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
   )

### When all species have the same rates except that the finches have a different
# rate of cladogenesis, and we want to optimize all parameters except K (which we
# set equal to Inf), fixing the proportion of finch-type species at 0.163, we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   initparsopt = c(0.38,0.55,0.004,1.1,2.28),
   idparsopt = c(1,2,4,5,6),
   parsfix = c(Inf,Inf,0.163),
   idparsfix = c(3,8,11),
   idparsnoshift = c(7,9,10),
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
   )
```

```
### When all species have the same rates except that the finches have a different
# rate of cladogenesis, extinction and a different K, and we want to optimize all
# parameters, fixing the proportion of finch-type species at 0.163, we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   ddmodel = 11,
   initparsopt = c(0.19,0.09,0.002,0.87,20,8.9,15),
   idparsopt = c(1,2,4,5,6,7,8),
   parsfix = c(Inf,0.163),
   idparsfix = c(3,11),
   idparsnoshift = c(9,10),
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
   )


### When all species have the same rates except that the finches have a different
# rate of extinction, and we want to optimize all parameters except K (which we
# set equal to Inf), and we also# want to estimate the fraction of finch species
# in the mainland pool. we use:

utils::data(Galapagos_datalist_2types)
DAISIE_ML(
   datalist = Galapagos_datalist_2types,
   initparsopt = c(2.48,2.7,0.009,1.01,2.25,0.163),
   idparsopt = c(1,2,4,5,7,11),
   parsfix = c(Inf,Inf),
   idparsfix = c(3,8),
   idparsnoshift = c(6,9,10),
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
   )

### When we have two islands with the same rates except for immigration and anagenesis rate,
# and we want to optimize all parameters, we use:

utils::data(Galapagos_datalist)
DAISIE_ML(
   datalist = list(Galapagos_datalist,Galapagos_datalist),
   datatype = 'multiple',
   initparsopt = c(2.5,2.7,20,0.009,1.01,0.009,1.01),
   idparsmat = rbind(1:5,c(1:3,6,7)),
   idparsopt = 1:7,
   parsfix = NULL,
   idparsfix = NULL,
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
```

```
)

### When we consider the four Macaronesia archipelagoes and set all parameters the same
# except for rates of cladogenesis, extinction and immigration for Canary Islands,
# rate of cladogenesis is fixed to 0 for the other archipelagoes,
# diversity-dependence is assumed to be absent
# and we want to optimize all parameters, we use:

utils::data(Macaronesia_datalist)
DAISIE_ML(
   datalist = Macaronesia_datalist,
   datatype = 'multiple',
   initparsopt = c(1.053151832,0.052148979,0.512939011,0.133766934,0.152763179),
   idparsmat = rbind(1:5,c(6,2,3,7,5),1:5,1:5),
   idparsopt = c(2,4,5,6,7),
   parsfix = c(0,Inf),
   idparsfix = c(1,3),
   tol = c(0.1, 0.02, 0.01),
   tolint = c(1e-4, 1e-2),
   res = 50
)
```

`frogs_datalist` *Colonization and branching times of 5 Eleutherodactylus clades from Hispaniola island.*

Description

A list containing the colonization and branching times of the Eleutherodactylus frogs of Hispaniola. Main dataset used in Etienne et al. This list can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.

The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.

`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
* Endemic_MaxAge: 5 or 6

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 6 elements, the first of which contains 2 elements and the following 5 elements containing 5 components.

Source

Etienne RS, Haegeman B, Dugo-Cota A, Vila C, Gonzalez-Voyer A & Valente L. The limits to ecological limits to diversification.

See Also

`DAISIE_dataprep`, `DAISIE_ML`, `DAISIE_SR_ML`

`frogs_datatable` *Colonization and branching times of 5 Eleutherodactylus (frogs) clades from the island of Hispaniola.*

Description

A table containing the colonization and branching times of the Eleutherodactylus frogs of the island of Hispaniola (Greater Antilles). Each row on the table represents and independent colonisation event. The table has four columns.

`$Clade_name` - name of independent colonization event
`$Status` - One of the following categories:
* Non_endemic: for non-endemic island species when an approximate time of colonisation is known
* Non_endemic_MaxAge: for non-endemic island species when colonisation time is unknown
* Endemic: for endemic species when an approximate colonisation time is known
* "Endemic_MaxAge": applies to endemic species or endemic clades for cases where the colonisation time is unknown, or when the user wants to specify an upper bound for colonisation. This could for example apply to endemic species that have recently gone extinct because of anthropogenic

causes, and which are not included in the phylogeny ("NA" should be given in the branching times column). It could also apply to insular radiations with long stem branches, for which the time of the first cladogenetic event is known, but the precise time of colonisation is not.
* Endemic&Non_Endemic: when endemic clade and mainland ancestor has re-colonized

Format

A table with 5 rows and 4 columns.

Source

Etienne RS, Haegeman B, Dugo-Cota A, Vila C, Gonzalez-Voyer A & Valente L. The limits to ecological limits to diversification.

`frogs_sim_datalist` *Colonization and branching times of a data set simulated with the MLE parameters of frogs_datalist*

Description

A list containing the colonization and branching times of a simulated data set using the MLE parameters of IW model for the frogs_datalist. It is an R list object with the following elements.

The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
* Endemic_MaxAge: 5 or 6

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with the first element containing 2 elements and the following elements containing 5 components.

Source

Etienne RS, Haegeman B, Dugo-Cota A, Vila C, Gonzalez-Voyer A & Valente L. The limits to ecological limits to diversification.

See Also

`DAISIE_ML`, `DAISIE_SR_ML`

`Galapagos_datalist` *Colonization and branching times of 8 terrestrial avifaunal clades in list format, accepted by DAISIE_ML and DAISIE_loglik_all*

Description

A list containing the colonization and branching times of the terrestrial avifauna in the Galapagos where no distinction is made between types of colonists. This list can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.

The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following 8 elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island cladeincluding the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 9 elements the first of which contains 2 elements and the following 8 containing 5 components.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_dataprep`, `DAISIE_ML`

`Galapagos_datalist_2types`

Colonization and branching times of 8 terrestrial avifaunal clades in list format, accepted by DAISIE_ML and DAISIE_loglik_all

Description

A list containing the colonization and branching times of the terrestrial avifauna in the Galapagos. This list can be generated using the DAISIE_dataprep function, which converts a user-specified data table into a data object, but the object can of course also be entered directly. It is an R list object with the following elements.

The first element of the list has three components:

`$island_age` - the island age
`$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island
`$not_present_type2` - the number of mainland lineages of type 2 that are not present on the island

The following 8 elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island.

For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island cladeincluding the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset only the finches are type 2

Format

A list with 9 elements the first of which contains 3 elements and the following 8 containing 5 components.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_dataprep`, `DAISIE_ML`

`Galapagos_datatable` *Colonization and branching times of 8 terrestrial avifaunal Galápagos clades in table format.*

Description

A table containing the colonization and branching times of the terrestrial avifauna in the Galápagos. Each row on the table represents and independent colonisation event. The table has four columns.

`$Clade_name` - name of independent colonization event
`$Status` - One of the following categories:
* Non_endemic: for non-endemic island species when an approximate time of colonisation is known
* Non_endemic_MaxAge: for non-endemic island species when colonisation time is unknown
* Endemic: for endemic species when an approximate colonisation time is known
* "Endemic_MaxAge": applies to endemic species or endemic clades for cases where the colonisation time is unknown, or when the user wants to specify an upper bound for colonisation. This could

for example apply to endemic species that have recently gone extinct because of anthropogenic causes, and which are not included in the phylogeny ("NA" should be given in the branching times column). It could also apply to insular radiations with long stem branches, for which the time of the first cladogenetic event is known, but the precise time of colonisation is not.
* Endemic&Non_Endemic: when endemic clade and mainland ancestor has re-colonized

`$Missing_species` - Number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$Branching_times` - Stem age of the population/species in the case of "Non_endemic", "Non_endemic_MaxAge" and "Endemic" species with no extant close relatives on the island. Set "NA" if colonisation time unknown and no upper bound is known. For "Endemic" cladogenetic species these should be branching times of the radiation, including the stem age of the radiation (colonisation time estimate).

Format

A table with 8 rows and 4 columns.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_dataprep`, `DAISIE_ML`

`islands_10reps_RAW` *1000 islands in RAW format simulated with the ML parameters of the CR model for the Galapagos data.*

Description

Each simulated dataset is an element of the list, which can be called using e.g. islands_10reps_RAW[[1]]
Each of the island replicates is a list in itself. The first (e.g. islands_10reps_RAW[[x]][[1]]) element of that list has the following components:
The following elements of the RAW list each contain information on a single colonist lineage on the island and has 5 components:
`$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
`$stac` - the status of the colonist
* Not_present: 0
* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4

`$stt_table` - Species-through-time table for the descendants of the mainland species (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species)
`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)

Format

A list with 10 items.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr()`, `DAISIE_plot_sims`

`islands_1type_1000reps`

1000 islands in DAISIE format simulated with the ML parameters of the CR model for the Galapagos data

---

Description

Each simulated dataset is an element of the list, which can be called using e.g. islands_1type_1000reps[[1]]
Each of the island replicates is a list in itself. The first (e.g. islands_1type_1000reps[[x]][[1]]) element of that list has the following components:
`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island
`$stt_all` - STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
The subsequent elements of the list each contain information on a single colonist lineage on the island and has 3 components:
`$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
`$stac` - the status of the colonist
* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)

Format

A list with 1000 items.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr()`, `DAISIE_plot_sims`

`islands_2types_1000reps`

1000 islands in DAISIE format simulated with the ML parameters of the CR_lamc_mu_K model for the Galapagos data (2 types of species)

Description

Each simulated dataset is an element of the list, which can be called using e.g. islands_2types_1000reps[[1]] Each of the island replicates is a list in itself. The first (e.g. islands_2types_1000reps[[x]][[1]]) element of that list has the following components:
`$island_age` - the island age
`$not_present_type1` - the number of mainland lineages of type 1 that are not present on the island
`$not_present_type2` - the number of mainland lineages of type 2 that are not present on the island
`$stt_all` - STT table for all species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
`$stt_stt_type1` - STT table for type 1 species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
`$stt_stt_type2` - STT table for type 2 species on the island (nI - number of non-endemic species; nA - number of anagenetic species, nC - number of cladogenetic species, present - number of independent colonisations present)
The subsequent elements of the list each contain information on a single colonist lineage on the island and has 4 components:
`$branching_times` - island age and stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge and Endemic anagenetic species. For cladogenetic species these should be island age and branching times of the radiation including the stem age of the radiation.
`$stac` - the status of the colonist
* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type_1or2` - whether the colonist belongs to type 1 or type 2

Format

A list with 1000 items.

Source

Valente, L.M., A.B. Phillimore and R.S. Etienne (2015). Equilibrium and non-equilibrium dynamics simultaneously operate in the Galapagos islands. Ecology Letters 18: 844-852.

See Also

`DAISIE_sim_cr()`, `DAISIE_plot_sims`

`Macaronesia_datalist` *Colonization and branching times of terrestrial avifaunal clades from Azores, Canary Islands, Cape Verde and Madeira in list format, accepted by DAISIE_ML and DAISIE_loglik_all*

Description

A list containing the colonization and branching times of the terrestrial avifauna in 4 archipelagos: Azores, Canary Islands, Cape Verde and Madeira. It is an R list object with the 4 main elements corresponding to each of the archipelagos (e.g. Macaronesia_datalist[[1]] calls the Azores data). Each of the four elements is then made of several elemants:

The first element of the list for an archipelago has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contains information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4

* Endemic_MaxAge: 5

`$missing_species` - number of island species that were not sampled for particular clade (only

applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 4 main elements for each archipelago. Each element has several sub-elements.

Source

Valente L., Illera J.C, Havenstein K., Pallien T., Etienne R.S., Tiedemann R. Equilibrium bird species diversity in Atlantic islands. 2017 Current Biology, 27, 1660-1666.

See Also

`DAISIE_dataprep`, `DAISIE_ML`

`NewZealand_birds_datalist`

Colonization and branching times of New Zealand birds.

Description

A list containing the colonization and branching times of the birds of New Zealand. Main dataset used in Valente, Etienne, Garcia-R (2019) Current Biology. Island age 52 Myr and mainland pool size of 1000 species.
The first element of the list has two components:

`$island_age` - the island age
`$not_present` - the number of mainland lineages that are not present on the island

The following elements of the list each contain information on a single colonist lineage on the island and has 5 components:

`$colonist_name` - the name of the species or clade that colonized the island
`$branching_times` - island age followed by stem age of the population/species in the case of Non-endemic, Non-endemic_MaxAge species and Endemic species with no close relatives on the island. For endemic clades with more than one species on the island (cladogenetic clades/ radiations) these should be island age followed by the branching times of the island clade including the stem age of the clade.
`$stac` - the status of the colonist

* Non_endemic_MaxAge: 1
* Endemic: 2
* Endemic&Non_Endemic: 3
* Non_endemic: 4
* Endemic_MaxAge: 5 or 6

`$missing_species` - number of island species that were not sampled for particular clade (only applicable for endemic clades)
`$type1or2` - whether the colonist belongs to type 1 or type 2. In this dataset all are equal to 1.

Format

A list with 40 elements, the first of which contains 2 elements and the following 39 containing 5 components.

Source

Valente L, Etienne RS, Garcia-R JC (2019) Deep Macroevolutionary Impact of Humans on New Zealand's Unique Avifauna. Current Biology, 29, 2563–2569.

See Also

`DAISIE_dataprep`, `DAISIE_ML`, `DAISIE_SR_ML`

`stac_table` *Explanatory table on meaning and use of* `stac` *settings*

Description

A table containing the information regarding the meaning of the `stac` codes utilised by DAISIE's ML functions. It is used to render the "DAISIE `stac` values" vignette.

`stac` stands for "status of the clade" formed by the immigrant. It is an important part of DAISIE objects that informs the likelihood functions about the endemicity status and type of data available for each insular clade. Each colonisation event that has extant species on the island needs to have a `stac` value specified. This also has implications in what is included in the DAISIE object `"branching_times"` vector, which also described in this table. The table is composed of the following columns:

- $stac: A numeric with each stac code from 1 until 9.
- $Input for DAISIE_dataprep table "Clade_Name": A character with the possibilities for input in DAISIE_dataprep() of each clade, which are then translated to a numeric `stac` code as in `$stac`.
- $Type of species or clade": A character with the plain English explanation of the different possible kinds of island lineages that can be considered by DAISIE.
- $Input for DAISIE_dataprep table "Branching_times": A character with the plain English explanation of what the branching times vector in the DAISIE object should contain.
- $Colonisation times: A character with the plain English explanation of what the colonisation time in the branching times vector of the DAISIE object (first element of the vector) means.

Format

A table with 9 rows and 5 columns.

Index