

# Package ‘**phyloland**’

February 20, 2015

**Type** Package

**Title** Modelling Competitive Exclusion and Limited Dispersal in a Statistical Phylogeographic Framework

**Version** 1.3

**Imports** ape

**Date** 2014-09-05

**Author** Louis Ranjard, Marie Paturel

**Maintainer** Louis Ranjard <l.ranjard@auckland.ac.nz>

## **Description**

Phyloland package models a space colonization process mapped onto a phylogeny, it aims at estimating limited dispersal and ecological competitive exclusion in a Bayesian MCMC statistical phylogeographic framework (please refer to phyloland-package help for details.)

**License** GPL-2

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2014-09-13 07:55:48

## **R topics documented:**

|                             |   |
|-----------------------------|---|
| phyloland-package . . . . . | 2 |
| Banza . . . . .             | 3 |
| PLD_interface . . . . .     | 4 |
| PLD_loc_mrca . . . . .      | 6 |
| PLD_plot_stat_mig . . . . . | 6 |
| PLD_plot_trees . . . . .    | 7 |
| PLD_stat_mig . . . . .      | 8 |
| tree_Banza . . . . .        | 8 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>10</b> |
|--------------|-----------|

---

phyloland-package      *Modelling competition and dispersal in a statistical phylogeographic framework*

---

## Description

The R package "phyloland" implements a landscape colonization model in a statistical phylogeographic framework, explicitly modelling dispersal and competition. The model aims at testing whether these two factors have had a significant impact on the process of landscape colonization that explains the current geographical distribution of the evolutionary units under study. Given (i) their phylogeny and (ii) their geographic location, the distributions of these two parameters is estimated.

The landscape is defined in a 2-dimensional space (latitude and longitude) as a set of locations that can be colonized through a sequence of dated dispersal events. According to this approach, dispersals are mapped on the genealogy, each internal node corresponding to a specific dispersal event.

This model takes into account two constraints on dispersals :

- The geographical distance separating the source and the post-dispersal locations: does the dispersal probability decrease with the geographic distance separating the locations of the landscape? A normal distribution centered on the source location is used to model dispersal. The standard deviation  $\sigma$  determines the geographical dispersal bias. Low values for  $\sigma$  indicate that the units of interest tend to disperse locally. Conversely, with highly mobile species or limited space, the probability to migrate to any locations in the landscape may be uniform, corresponding to larger values of  $\sigma$ .
- The occupancy state of the post-dispersal location: does the probability of colonizing a location vary with its occupancy state, i.e. whether it has previously been colonized or not? An occupied location is colonized with probability  $\lambda$ . If competition for space takes place, dispersals to occupied locations are prevented and the estimates of the parameter  $\lambda$  will fall below 1.

The 3 model parameters are :

- $\sigma$ : the standard deviation of the normal distribution for dispersal kernel.
- $\lambda$ : the probability of migrating to an occupied location.
- $\tau$ : the overall dispersal rate, the number of dispersals per unit of time.

Bayesian Markov chain Monte Carlo (MCMC) algorithms are used to sample the model parameters and the geographic locations at the internal nodes of the phylogeny. Uniform distributions are used as prior distribution for all parameters.

## Details

The following package functions allow the user to implement the model and investigate the results :

- The function `PLD_interface` performs the Bayesian estimations from a phylogeny and a location file.
- `PLD_plot_trees` plots a number of sampled trees with their internal and tips locations.
- `PLD_loc_mrca` returns the posterior distribution of each location of the Most Recent Common Ancestor of a set of tips. A barplot can be displayed.
- `PLD_stat_mig` lists the dispersal events from the genealogies sampled by the MCMC.
- `PLD_plot_stat_mig` displays the density plots of the dispersal time points for each location of the space.

---

Banza

*Banza MCMC output*

---

### Description

This data set was generated by the function `PLD_interface.Rd` using a phylogeny constructed from the data of Shapiro et al. (2006). The command was: `names_locations = c("Maui_Nui","Maui_Nui","Maui_Nui","Maui_Nui")`  
`Banza = PLD_interface(fileTREES=system.file("extdata","tree_Banza.nex.txt",package="phyloland"),fileDATA=system.file("extdata","Banza.nex.txt",package="phyloland"),num_step=10000,freq=10,ess_lim=500,id_filena="Banza",names_locations=names_locations)`

### Usage

`banza`

### Format

An object of class "phyloland" comprising 7 attributes:

1. trees 1000 multiPhylo list
2. locations 41 data.frame list
3. tips 21 -none- character
4. space 10 -none- numeric
5. index 1 -none- list
6. mcmc 8 -none- list
7. sigma\_limit 2 -none- numeric

### Source

Output generated by `phyloland:PLD_interface.Rd`

### References

Shapiro LH, Strazanac JS, Roderick GK. Molecular phylogeny of Banza (Orthoptera: Tettigoniidae), the endemic katydid of the Hawaiian Archipelago. *Mol. Phylogenet. Evol.* 41(1):53-63, 2006.

---

 PLD\_interface

*Model parameters and internal locations estimations*


---

### Description

Function that estimates model parameters, genealogies and internal locations through Bayesian Markov chain Monte Carlo (MCMC) algorithm.

### Usage

```
PLD_interface(fileTREES, fileDATA, num_step = 1e+05, freq = 100,
burnin = 0, ess_lim = 100, sigma = NA, lambda = NA, tau = NA,
num_step_sigma = 1, num_step_lambda = 1, num_step_tau = 1,
id_filena = NA, pattern_trees_likelihood = "treeLikelihood",
names_locations = NA)
```

### Arguments

|           |  |
|-----------|--|
| fileTREES | a character string that specifies the name of the file containing the phylogenetic trees. A Nexus file (e.g. BEAST output) with tips names (no space characters allowed), the phylogenetic trees and the trees likelihood. If only one tree is contained in the file, the genealogies are not sampled in the MCMC.   |
| fileDATA  | a character string that specifies the name of the file containing the tips locations : a text file with 3 columns (separated by tabs). The first one with the tips names (the same as in fileTrees), the second one with the location latitudes and the last one with the location longitudes (both in decimal degrees). No header.                        |
| num_step  | a strictly positive integer specifying the length of the Markov Chain (the number of MCMC steps). Suggested values : 50 000 - 100 000.   |
| freq      | a strictly positive integer specifying how often the sampled model parameters and the locations are saved. Suggested values : 50 - 100.  |
| burnin    | a strictly positive integer specifying the number of trees dropped from the file fileTREES. Suggested value : 10% of the total number of trees, 0 in case of only one tree.  |
| ess_lim   | a strictly positive integer specifying the Effective Sample Size. The MCMC stops when the ess of all parameters reach this value.  |
| sigma     | a vector of two elements, strictly positive values specifying the standard deviation of the normal distribution for dispersal steps in each dimension. Default is NA: the 2 sigma parameters are estimated in the MCMC. Alternatively, if values are specified the sigma parameters are not estimated by the MCMC.   |
| lambda    | a strictly positive value specifying the probability of migrating to an occupied location. Default is NA: the lambda parameter is estimated in the MCMC. Alternatively, if a value is specified the lambda parameter is not estimated by the MCMC. A lambda value equal to 1 indicates that competitive exclusion did not impact the colonization process. |

|                                       |   |
|---------------------------------------|---|
| <code>tau</code>                      | a strictly positive value specifying the overall dispersal rate. Default is NA: the tau parameter is estimated in the MCMC. Alternatively, if a value specified the tau parameter is not estimated by the MCMC. |
| <code>num_step_sigma</code>           | a strictly positive integer specifying how many times the sigma parameters are sampled at each MCMC step.   |
| <code>num_step_lambda</code>          | a strictly positive integer specifying how many times the lambda parameter is sampled at each MCMC step.  |
| <code>num_step_tau</code>             | a strictly positive integer specifying how many times the tau parameter is sampled at each MCMC step.   |
| <code>id_filena</code>                | an optional character string specifying the ID for the output files.  |
| <code>pattern_trees_likelihood</code> | a character string specifying the text pattern to retrieve the trees likelihood in fileTREES.   |
| <code>names_locations</code>          | a vector of character strings that specifies the location names, in the same order as in fileDATA.  |

### Details

MCMC runs stop when the `num_step` is reached or when the ess of all estimated parameters equal `ess_lim`.

The respective ess for estimating convergence of the chains regarding each parameter is calculated by monitoring the autocorrelation between successive sampled states in the MCMC.

### Value

an object of class "phyloland" with the following components:

|                          |  |
|--------------------------|--|
| <code>trees</code>       | an object of class "multiPhylo" or "phylo" (package ape) containing the trees sampled in the MCMC. If the <code>ess_lim</code> is not reached, the number of trees is <code>num_step/freq</code> . |
| <code>locations</code>   | a matrix containing the internal locations sampled in the MCMC and the tips locations (columns) for each sampled tree(rows).   |
| <code>tips</code>        | a vector of character strings containing the tips names (from fileTREES and fileDATA).   |
| <code>space</code>       | a matrix containing the unique locations from fileDATA.  |
| <code>mcmc</code>        | a list containing the posterior distributions of each parameters.  |
| <code>sigma_limit</code> | a vector indicating the threshold for sigma below which limited dispersal occurs.  |

---

PLD\_loc\_mrca                      *Location of Most Recent Common Ancestor*

---

### Description

Function that returns the posterior distribution of the geographic locations of the Most Recent Common Ancestor of a set of tips. A barplot can be generated.

### Usage

```
PLD_loc_mrca(x, tips, sub_sample = 0, plot_distrib = FALSE, col = NA)
```

### Arguments

x                      an object of class "phyloland" generated by a call to [PLD\\_interface](#).

tips                    a vector of character strings specifying the set of tips.

sub\_sample            a vector of integers that specifies the indexes of the trees and locations to take into account. Default is 0, all the trees and locations are considered.

plot\_distrib          a logical value, if TRUE a barplot is plotted. The default is FALSE.

col                    a set of colors for the barplot.

### Value

frequencies          matrix with 3 columns containing location latitudes, longitudes and frequencies.

locationsMRCA        vector containing all the sub\_sample sampled locations for the tips Most Recent Common Ancestor.

### Examples

```
data(Banza)
locations = PLD_loc_mrca(x = Banza, tips = Banza$tips, sub_sample = 100:200, plot_distrib = TRUE)
locations$frequencies
```

---

PLD\_plot\_stat\_mig                *Plot dispersal statistics*

---

### Description

Function that displays the density plots of the dispersal times for each location.

### Usage

```
PLD_plot_stat_mig(timemat, color = NA, xy_legend = NA, group = 0)
```

**Arguments**

|           |  |
|-----------|--|
| timemat   | a 3-dimensional (a*b*c) array with (a) the departure locations, (b) the destination locations and (c) the time of the dispersal from (a) to (b), generated by a call to <a href="#">PLD_stat_mig</a> . |
| color     | a set of colors for the density plot. If NA, rainbow (grDevices) is used.  |
| xy_legend | a vector of 2 values specifying the legend coordinates.  |
| group     | locations are grouped according to timemat rownames.   |

**Examples**

```
data(Banza)
stat = PLD_stat_mig(x = Banza, sub_sample = 100:200, first = TRUE)
PLD_plot_stat_mig(timemat = stat$timemat)
```

---

|                |                           |
|----------------|---------------------------|
| PLD_plot_trees | <i>Plot sampled trees</i> |
|----------------|---------------------------|

---

**Description**

Function that plots sampled trees with their locations.

**Usage**

```
PLD_plot_trees(x, sub_sample = 0, one_plot = FALSE)
```

**Arguments**

|            |   |
|------------|---|
| x          | an object of class "phyloLand" generated by a call to <a href="#">PLD_interface</a> .   |
| sub_sample | a vector of integers that specifies the indexes of the tree(s) to plot. Default is 0, all the trees are plotted.                |
| one_plot   | a logical value. Default is FALSE, the trees are plotted one after the other. If TRUE the trees are plotted on the same device. |

**Examples**

```
data(Banza)
PLD_plot_trees(x = Banza, sub_sample = 1)
PLD_plot_trees(x = Banza, sub_sample = 100:105)
```

---

PLD\_stat\_mig                      *Dispersal statistics*

---

### Description

Function that lists the dispersal events from the genealogies sampled by the MCMC.

### Usage

```
PLD_stat_mig(x, sub_sample = 0, first = FALSE)
```

### Arguments

**x**                      an object of class "phyloLand" generated by a call to [PLD\\_interface](#).

**sub\_sample**            a vector of integers that specifies the indexes of the trees and locations to take into account to list the dispersals events. Default is 0, all trees and locations are considered.

**first**                 a logical value. If TRUE only the oldest dispersal to a location is taken into account. If FALSE all dispersals are taken into account.

### Value

**migmat**                a matrix containing the number of occurrences of each dispersal event (departures in rows and destinations in columns).

**timemat**                a 3-dimensional (a\*b\*c) array with (a) the departure locations, (b) the destination locations and (c) the time (node height) of the dispersal from (a) to (b).

### Examples

```
data(Banza)
stat = PLD_stat_mig(x = Banza, sub_sample = 100:200, first = TRUE)
stat$migmat
stat$timemat[, , 100]
```

---

tree\_Banza                      *Phylogenetic tree of Banza*

---

### Description

This data contains the tree used to generate the file banza.RData.

### Usage

```
banza
```



**Format**

A phylogenetic tree in nexus format. Taxa names match the sample names in *locations\_Banza.txt*.

**Source**

Generated in Beast using data from Shapiro et al. (2006)

**References**

Shapiro LH, Strazanac JS, Roderick GK. Molecular phylogeny of *Banza* (Orthoptera: Tettigoniidae), the endemic katydids of the Hawaiian Archipelago. *Mol Phylogenet Evol.* 2006 Oct;41(1):53-63.

# Index

## \*Topic **datasets**

Banza, [3](#)

tree\_Banza, [8](#)

Banza, [3](#)

phyloland (phyloland-package), [2](#)

phyloland-package, [2](#)

PLD\_interface, [3](#), [4](#), [6–8](#)

PLD\_loc\_mrca, [3](#), [6](#)

PLD\_plot\_stat\_mig, [3](#), [6](#)

PLD\_plot\_trees, [3](#), [7](#)

PLD\_stat\_mig, [3](#), [7](#), [8](#)

tree\_Banza, [8](#)