Package 'pgHMAtools'

November 2, 2020

Title Phylogenetic and population genetic heteroduplex mobility analyses

Version 0.0.1

Description A package containing several R-
functions for phylogenetic and population genetic analyses based on heteroduplex mobility data

Depends R (>= 2.10)

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Imports HDInterval,
ape,
ggplot2,
phangorn,
zoo,
stats,
scales

R topics documented:

---

get_DH   *Get the demographic history from each genealogy*

Description

This function was used to obtain the demographic history from the genealogy Uses classic skyline implemented in ape package to find the time at the end of each coalescent interval and effective population size of each interval.

Usage

```
get_DH(i, tree, lower_HDI, epsilon = 0, n = 1000)
```

Arguments

i	which number of the phylogenetic tree need to be analyzed.
tree	either an ultra-metric phylogenetic tree (i.e. an object of class "phylo") or a vector of interval lengths.
lower_HDI	Lower HDI value of the phylogenetic tree.
epsilon	collapsing parameter that controls the amount of smoothing.
n	A number of intervals used to build the matrix.

Value

Normalized coalescent interval and effective population size for the phylogenetic tree.

---

get_HDI_matrix   *Get a HDI matrix for the phylogenetic tree*

Description

This function was used to obtain a HDI matrix of the phylogenetic tree. Uses hdi (Highest Density Interval) algorithm as implemented in HDInterval package.

Usage

```
get_HDI_matrix(dat, margin = 2, ci = 0.95)
```

Arguments

dat	Data matrix used for extracting highest density intervals, and assume that each row represents a sample and each column show the coalescent interval.
margin	Default by column.
ci	The value of credible interval.

Value

A HDI matrix with each column is the value of lower, median, and upper bound.

---

get_lower_HDI *Get Lower HDI value of the phylogenetic tree*

Description

This function was used to calculate the lower HDI value of the phylogenetic tree. Uses hdi (Highest Density Interval) algorithm as implemented in HDInterval package. Uses coalescent.intervals as implemented in ape package.

Usage

```
get_lower_HDI(tree, ci = 0.95)
```

Arguments

tree	either an ultra-metric phylogenetic tree (i.e. an object of class "phylo") or a vector of interval lengths.
ci	The value of credible interval.

Value

lower limits of the HDI for all trees in the tree file.

---

get_slide_matrix *Get a smooth HDI matrix by sliding window method*

Description

This function was used to obtain a smooth HDI matrix by sliding window. Uses zoo algorithm as implemented in zoo package.

Usage

```
get_slide_matrix(x, w = 100, s = 20)
```

Arguments

x	Data matrix of each row represents a sample and each column show the coalescent interval.
w	window size of the sliding window.
s	step size of the sliding window.

Value

A smooth HDI matrix with each column is the value of lower, median, and upper bound.

---

pair_data	*Distance data matrix for 16 samples including the heteroduplex mobility data and nucleotide sequencing data.*

Description

Pairwise distance data matrix for 16 samples

Usage

```
pair_data
```

Format

A data frame with 120 rows and 2 variables:

ober Heteroduplex mobility data

real Nucleotide sequencing data ...

---

para_boot	*Parametric bootstrap sampling using the predict interval*

Description

This function was used to bootstrap sampling for HMA data using the predict interval. Uses neighbour-joining (nj) algorithm as implemented in ape package. Uses upgma method as implemented in phangorn package.

Usage

```
para_boot(pair_data, dframe, method)
```

Arguments

pair_data	Assume by default that the data matrix is the pairwise distances matrix, which contain two column named ober (HMA data) and real (DNA data).
dframe	A data frame for the samples.
method	only allow "nj" or "upgma" method to build the phylogeny.

Value

the phylogenetic tree of the parametric bootstrap.

| plot_lm | *Plot the linear regression with prediction interval* |

Description

This function was used to perform the correlation analysis between HMA and nucleotide sequencing data. Uses ggplot2 package.

Usage

```
plot_lm(pair_data)
```

Arguments

pair_data   Assume by default that the data matrix is the pairwise distances matrix, which contain two column named ober (HMA data) and real (DNA data).

Examples

```
data(pair_data)
plot_lm(pair_data)
```

| Skyline_plot | *Draw the Skyline plot with or without Y log axis.* |

Description

This function was used to draw the Skyline plot from the HDI matrix with or without Y log axis. Use ggplot2 package. Assuming each column is a different time point, and the dimension between times and skyline_matrix is match.

Usage

```
Skyline_plot(
  times,
  skyline_matrix,
  log_axis = FALSE,
  col = "blue",
  fill = "gray85",
  xlims = NULL,
  ylims = NULL,
  xlab = "Time",
  ylab = "Effective population size",
  main = "",
  base_size = 12
)
```

Arguments

`times`	The time points or mutational units (X-axis) used to draw the Skyline plot.
`skyline_matrix`	The HDI matrix (including three rows of lower, median and upper bound) used to draw the Skyline plot.
`log_axis`	Use the Y log axis or not. Default is false.
`col`	Set the color for the median trend.
`fill`	Set the color for the credible interval.
`xlims`	scale limits for X-axis.
`ylims`	scale limits for Y-axis.
`xlab`	name for X-axis.
`ylab`	name for Y-axis.
`main`	name for the plot title.
`base_size`	Set the base font size.

Index