Package ‘harrietr’

December 1, 2017

Type Package

Title Wrangle Phylogenetic Distance Matrices and Other Utilities

Version 0.2.3

Description Harriet was Charles Darwin's pet tortoise (possibly). 'harrietr' implements some function to manipulate distance matrices and phylogenetic trees to make it easier to plot with 'ggplot2' and to manipulate using 'tidyverse' tools.

Depends R (>= 3.3.0)

License GPL-3 | file LICENSE

URL `https://github.com/andersgs/harrietr`

BugReports `https://github.com/andersgs/harrietr/issues`

Encoding UTF-8

LazyData true

Imports ape (>= 4.1), ggtree (>= 1.8.1), magrittr (>= 1.5), lazyeval (>= 0.2.0), dplyr (>= 0.7.2), tidyr (>= 0.7.0), rlang (>= 0.1.2)

Suggests ggplot2 (>= 2.2.1)

RoxygenNote 6.0.1

NeedsCompilation no

Author Anders Gonçalves da Silva [aut, cre]

Maintainer Anders Gonçalves da Silva <`andersgs@gmail.com`>

Repository CRAN

Date/Publication 2017-12-01 04:30:18 UTC

R topics documented:

`dist_long` *Return evolutionary distance in long format*

Description

This will take an alignment, will calculate the evolutionary distance between all pairs of sequence, and will transform the distance matrix to long format. It will remove upper triangle, and diagonal elements, so you end with only (n)*(n-1)/2 rows, where n are the total number of rows in the distance matrix.

Usage

```
dist_long(aln, order = NULL, dist = "N", tree = NULL)
```

Arguments

`aln`	An object of class matrix, it must be square
`order`	A character vector of size n with the order of the columns and rows (default: NULL)
`dist`	A string naming the model to calculate distances (accepted values are those in ape::dist.dna)
`tree`	An object of class phylo

Details

If a tree is optionally given, a fourth column is returned with the cophenetic distance across all elements of tree. It assumes the tree was generated from the alignment.

Value

A data.frame with three or four columns: (1) iso1; (2) iso2: (3) dist. If a tree is given then a fourth column (evol_dist) containig the distances from the tree is also supplied.

Examples

```
## Not run:
data(woodmouse)
dist_df <- dist_long(woodmouse)

## End(Not run)
```

`get_node_support` *Get node support from a tree produced with IQTREE*

Description

In IQTREE it is possible to obtain node support values by SH aproximate likelihood ratio tests (SH-aLRT), and ultrafast bootstraps (uBS). Often, we do both, which IQTREE encodes as two numbers separated by a '/' as the internal node label. This function will return a data.frame with the number of the internal nodes, and the support values for each.

Usage

```
get_node_support(tree)
```

Arguments

`tree`	An object of type phylo generated using IQTREE

Value

A data.frame with internal node information, plus two columns: (1) SH-aLRT; and (2) uBS

Examples

```
## Not run:
data(woodmouse_iqtree)
get_node_support(woodmouse_iqtree)

## End(Not run)
```

`harrietr` `harrietr` *package*

Description

harrietr: Wrangle Phylogenetic Distance Matrices and Other Utilities

Details

See the README on CRAN or GitHub

`join_metadata` *Add metadata to long distance matrix*

Description

This functions takes the output from `dist_long`, plus a data.frame with metadata, and attaches it to the data.frame output from `dist_long`. It uses a column in the metadata data.frame as a key to join the two data.frames. So, it requires a column of data in the metadata data.frame to have same ID labels as those in the pairwise comparison table.

Usage

```
join_metadata(dist, meta, isolate = "ISOLATES", group = "CLUSTER",
  remove_ind = TRUE)
```

Arguments

`dist`	A data.frame produced by dist_long function
`meta`	A data.frame with one column of IDs that match the IDs in `dist_long`
`isolate`	A character string with the name of the column in the meta data.frame with the ID data
`group`	A character string with the name of column containing the grouping variable
`remove_ind`	A boolean whether to remove all non-essential columns

Details

The output from `dist_long` with an additional column containing a factor, with levels composed of joining the categories in the `group` colum of the metadata data.frame for each pairwise comparison. For example, if one row has distance between samples id1 and id2, and in the grouping column of the metadata id1 is identified as part of group 'A', and id2 is identified as part of group 'B', then the output from that row will 'AB'. If they were from the same group, say 'A', the output would be just 'A'. In this way it is easy to identify pairs of isolates that are from the same group, and pairs of isolates that are from different groups.

Examples

```
## Not run:
data(woodmouse)
data(woodmouse_meta)
dist_df <- dist_long(woodmouse)
join_metadata(dist_df, woodmouse_meta, isolate = 'SAMPLE_ID', group = 'CLUSTER', remove_ind = TRUE)

## End(Not run)
```

`melt_dist` *Melt a square distance matrix into long format*

Description

This will take a square distance matrix, and will transform in to long format. It will remove upper triangle, and diagonal elements, so you end with only (n)*(n-1)/2 rows, where n are the total number of rows in the distance matrix.

Usage

```
melt_dist(dist, order = NULL, dist_name = "dist")
```

Arguments

`dist`	An object of class matrix, it must be square
`order`	A character vector of size n with the order of the columns and rows (default: NULL)
`dist_name`	A string to name the distance column in the output (default: dist)

Value

A data.frame with three columns: (1) iso1; (2) iso2; (3) dist. iso1 and iso2 indicate the pair being compared, and dist indicates the distance between that pair.

Examples

```
## Not run:
data(woodmouse)
dist <- ape::dist(woodmouse, model = 'N', as.matrix = TRUE)
dist_df <- melt_dist(dist)

## End(Not run)
```

`woodmouse` *Woodmouse dataset*

Description

Woodmouse dataset

Usage

```
woodmouse
```

Format

An object of class `DNAbin` with 15 rows and 965 columns.

Source

"ape" package woodmouse

`woodmouse_iqtree` *Woodmouse IQTREE newick tree*

Description

Generated a multiFASTA, and used IQTREE to generate a tree with the following command:

Usage

```
woodmouse_iqtree
```

Format

An object of class `phylo` of length 5.

Details

```
iqtree -s woodmouse.fasta -m TEST -nt 4 -bb 1000 -alrt 1000
```

The tree was loaded into ‘R‘ using ‘ape::read.tree‘, and saved to a dataset.

Source

"ape" package woodmouse

`woodmouse_meta` *Woodmouse metadata*

Description

A dummy metadata table generated to demonstrate the use of `add_metadata`.

Usage

```
woodmouse_meta
```

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 15 rows and 2 columns.

Index