

Package ‘scistreer’

January 6, 2023

Title Maximum-Likelihood Perfect Phylogeny Inference at Scale

URL <https://github.com/kharchenkolab/scistreer>,
<https://kharchenkolab.github.io/scistreer/>

Version 1.1.0

Description

Fast maximum-likelihood phylogeny inference from noisy single-cell data using the 'ScisTree' algorithm by Yufeng Wu (2019) <[doi:10.1093/bioinformatics/btz676](https://doi.org/10.1093/bioinformatics/btz676)>. 'scistreer' provides an 'R' interface and improves speed via 'Rcpp' and 'RcppParallel', making the method applicable to massive single-cell datasets (>10,000 cells).

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 4.1.0)

biocViews

Imports ape, dplyr, ggplot2, ggtree, igraph, parallelDist, patchwork,
phangorn, Rcpp, reshape2, RcppParallel, RhpcBLASctl, stringr,
tidygraph

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

LinkingTo Rcpp, RcppArmadillo, RcppParallel

NeedsCompilation yes

SystemRequirements C++11, GNU make

Author Teng Gao [cre, aut],
Evan Biederstedt [aut],
Peter Kharchenko [aut],
Yufeng Wu [aut]

Maintainer Teng Gao <tgaoteng@gmail.com>

RoxygenNote 7.2.2

Repository CRAN

Date/Publication 2023-01-06 22:30:04 UTC

R topics documented:

annotate_tree	2
get_mut_graph	3
gtree_small	3
ladderize	4
mut_nodes_small	4
mut_to_tree	5
perform_nni	5
plot_phylo_heatmap	6
P_example	7
P_small	7
run_scistree	7
score_tree	8
to_phylo	9
tree_small	9
tree_upgma	10
Index	11

annotate_tree	<i>Find maximum likelihood assignment of mutations on a tree</i>
---------------	--

Description

Find maximum likelihood assignment of mutations on a tree

Usage

```
annotate_tree(tree, P)
```

Arguments

tree	phylo Single-cell phylogenetic tree
P	matrix Genotype probability matrix

Value

tbl_graph A single-cell phylogeny with mutation placements

Examples

```
gtree_small = annotate_tree(tree_small, P_small)
```

get_mut_graph	<i>Convert a single-cell phylogeny with mutation placements into a mutation graph</i>
---------------	---

Description

Convert a single-cell phylogeny with mutation placements into a mutation graph

Usage

```
get_mut_graph(gtree)
```

Arguments

gtree	tbl_graph	The single-cell phylogeny
-------	-----------	---------------------------

Value

igraph Mutation graph

Examples

```
mut_graph = get_mut_graph(gtree_small)
```

gtree_small	<i>Smaller example annotated tree built from P_small</i>
-------------	--

Description

Smaller example annotated tree built from P_small

Usage

```
gtree_small
```

Format

An object of class tbl_graph (inherits from igraph) of length 199.

ladderize	<i>From ape; will remove once new ape version is released https://github.com/emmanuelparadis/ape/issues/54</i>
-----------	---

Description

From ape; will remove once new ape version is released <https://github.com/emmanuelparadis/ape/issues/54>

Usage

```
ladderize(phy, right = TRUE)
```

Arguments

phy	phylo	The phylogeny
right	logical	Whether ladderize to the right

Examples

```
tree_small = ladderize(tree_small)
```

mut_nodes_small	<i>Mutation placements calculated from tree_small and P_small</i>
-----------------	---

Description

Mutation placements calculated from tree_small and P_small

Usage

```
mut_nodes_small
```

Format

An object of class `data.frame` with 9 rows and 2 columns.

mut_to_tree	<i>Transfer mutation assignment onto a single-cell phylogeny</i>
-------------	--

Description

Transfer mutation assignment onto a single-cell phylogeny

Usage

```
mut_to_tree(gtree, mut_nodes)
```

Arguments

gtree	tbl_graph	The single-cell phylogeny
mut_nodes	dataframe	Mutation placements

Value

tbl_graph A single-cell phylogeny with mutation placements

Examples

```
gtree_small = mut_to_tree(gtree_small, mut_nodes_small)
```

perform_nni	<i>Maximum likelihood tree search via NNI</i>
-------------	---

Description

Maximum likelihood tree search via NNI

Usage

```
perform_nni(  
  tree_init,  
  P,  
  max_iter = 100,  
  eps = 0.01,  
  ncores = 1,  
  verbose = TRUE  
)
```

Arguments

tree_init	phylo	Initial tree
P	matrix	Genotype probability matrix
max_iter	integer	Maximum number of iterations
eps	numeric	Tolerance threshold in likelihood difference for stopping
ncores	integer	Number of cores to use
verbose	logical	Verbosity

Value

multiPhylo List of trees corresponding to the rearrangement steps

Examples

```
tree_list = perform_nni(tree_upgma, P_small)
```

plot_phylo_heatmap *Plot phylogeny and mutation heatmap*

Description

Plot phylogeny and mutation heatmap

Usage

```
plot_phylo_heatmap(tree, P, branch_width = 0.5, root_edge = TRUE)
```

Arguments

tree	phylo or tbl_graph	Phylogeny
P	matrix	Genotype probability matrix
branch_width	numeric	Branch width
root_edge	logical	Whether to plot root edge

Value

ggplot Plot visualizing the single-cell phylogeny and mutation probability heatmap

Examples

```
p = plot_phylo_heatmap(tree_small, P_small)
```

P_example	<i>Example genotype probability matrix</i>
-----------	--

Description

Example genotype probability matrix

Usage

P_example

Format

An object of class `matrix` (inherits from `array`) with 1000 rows and 25 columns.

P_small	<i>Smaller example genotype probability matrix</i>
---------	--

Description

Smaller example genotype probability matrix

Usage

P_small

Format

An object of class `matrix` (inherits from `array`) with 100 rows and 25 columns.

run_scistree	<i>Run the scistree workflow</i>
--------------	----------------------------------

Description

Run the scistree workflow

Usage

```
run_scistree(  
  P,  
  init = "UPGMA",  
  ncores = 1,  
  max_iter = 100,  
  eps = 0.01,  
  verbose = TRUE  
)
```

Arguments

P	matrix	Genotype probability matrix (cell x mutation). Each entry is a probability (0-1) that the cell harbors the mutation
init	character	Initialization strategy; UPGMA or NJ
ncores	integer	Number of cores to use
max_iter	integer	Maximum number of iterations
eps	numeric	Tolerance threshold in likelihood difference for stopping
verbose	logical	Verbosity

Value

phylo A maximum-likelihood phylogeny

Examples

```
tree_small = run_scistree(P_small)
```

score_tree

Score a tree based on maximum likelihood

Description

Score a tree based on maximum likelihood

Usage

```
score_tree(tree, P, get_l_matrix = FALSE)
```

Arguments

tree	phylo object
P	genotype probability matrix
get_l_matrix	whether to compute the whole likelihood matrix

Value

list Likelihood scores of a tree

Examples

```
tree_likelihood = score_tree(tree_upgma, P_small)$l_tree
```

to_phylo	<i>Convert the phylogeny from tidygraph to phylo object modified from R package alakazam, converts a tbl_graph to a phylo object</i>
----------	--

Description

Convert the phylogeny from tidygraph to phylo object modified from R package alakazam, converts a tbl_graph to a phylo object

Usage

```
to_phylo(graph)
```

Arguments

graph tbl_graph The single-cell phylogeny

Value

phylo The single-cell phylogeny

Examples

```
tree_small = to_phylo(annotate_tree(tree_small, P_small))
```

tree_small	<i>Smaller example tree built from P_small</i>
------------	--

Description

Smaller example tree built from P_small

Usage

```
tree_small
```

Format

An object of class phylo of length 5.

tree_upgma

Example tree built using UPGMA from P_small

Description

Example tree built using UPGMA from P_small

Usage

tree_upgma

Format

An object of class phylo of length 4.

Index

* datasets

- gtree_small, 3
- mut_nodes_small, 4
- P_example, 7
- P_small, 7
- tree_small, 9
- tree_upgma, 10

annotate_tree, 2

get_mut_graph, 3

gtree_small, 3

ladderize, 4

mut_nodes_small, 4

mut_to_tree, 5

P_example, 7

P_small, 7

perform_nni, 5

plot_phylo_heatmap, 6

run_scistree, 7

score_tree, 8

to_phylo, 9

tree_small, 9

tree_upgma, 10