

Package ‘DCLEAR’

February 16, 2023

Version 1.0.12

Date 2023-2-16

Title Distance Based Cell Lineage Reconstruction

Author Il-Youp Kwak [aut, cre],
Wuming Gong [aut]

Maintainer Il-Youp Kwak <ikwak2@cau.ac.kr>

License GPL-3

VignetteBuilder knitr

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.1.2

Depends R (>= 4.1.0), tensorflow(>= 2.2.0)

biocViews

Encoding UTF-8

Imports BiocParallel, dplyr, Matrix, matrixStats, ape, phangorn, Rcpp,
igraph, methods, purrr, stringr, tidyr, rBayesianOptimization,
rlang, BiocGenerics

Suggests knitr, rmarkdown, markdown

Description R codes for distance based cell lineage reconstruction. Our methods won both sub-challenges 2 and 3 of the Allen Institute Cell Lineage Reconstruction DREAM Challenge in 2020. References: Gong et al. (2021) <[doi:10.1016/j.cels.2021.05.008](https://doi.org/10.1016/j.cels.2021.05.008)>, Gong et al. (2022) <[doi:10.1186/s12859-022-04633-x](https://doi.org/10.1186/s12859-022-04633-x)>.

URL <https://github.com/ikwak2/DCLEAR>

NeedsCompilation yes

Repository CRAN

Date/Publication 2023-02-16 16:00:07 UTC

R topics documented:

add_deletion	3
add_dropout	4
as_igraph	4
as_igraph,data.frame-method	5
as_igraph,phylo-method	5
as_lineage_tree	6
as_lineage_tree,phyDat,phylo,lineage_tree_config-method	6
as_phylo	7
as_phylo,igraph-method	7
DCLEAR	7
dist_kmer_replacement_inference	8
dist_replacement	8
dist_replacement,phyDat,kmer_summary,integer-method	9
dist_replacement,phyDat,missing,integer-method	9
dist_weighted_hamming	10
dist_weighted_hamming,phyDat,numeric-method	10
downsample	12
downsample,igraph-method	12
downsample,lineage_tree-method	13
get_distance_prior	13
get_leaves	14
get_leaves,lineage_tree-method	14
get_node_names	15
get_replacement_probability	15
get_sequence	16
get_transition_probability	16
lineages	17
positional_mutation_prob	17
process_sequence	18
process_sequence,phyDat-method	18
prune	19
prune,igraph-method	19
prune,lineage_tree-method	20
random_tree	20
rbind,phyDat-method	21
sample_mutation_outcome	21
sample_mutation_site	22
sample_outcome_prob	22
score_simulation	23
simulate	23
simulate,lineage_tree_config,missing-method	24
simulate,lineage_tree_config,phyDat-method	24
simulate_core	25
sim_seqdata	26
substr_kmer	27
substr_kmer,kmer_summary-method	28

<i>add_deletion</i>	3
subtract	28
subtract,lineage_tree,lineage_tree-method	29
subtree	29
subtree,lineage_tree-method	30
subtree,phylo-method	30
summarize_kmer	31
summarize_kmer,phyDat-method	31
summarize_kmer_core	32
WH	33
WH_train	34
WH_train_fit	35
Index	36

<code>add_deletion</code>	<i>add_deletion</i>
---------------------------	---------------------

Description

Add deletion

Usage

```
add_deletion(x, tree, mutation_site, config)
```

Arguments

<code>x</code>	a character matrix
<code>tree</code>	a matrix representing the lineage tree
<code>mutation_site</code>	a binary matrix for mutation site
<code>config</code>	a <code>lineage_tree_config</code> object

Value

a character matrix with deletions

add_dropout *add_dropout*

Description

Add dropout events

Usage

```
add_dropout(x, config)
```

Arguments

x a character matrix
config a lineage_tree_config object

Value

a character matrix with dropout events

as_igraph *Generic function for as_igraph*

Description

Generic function for as_igraph

Usage

```
as_igraph(x, ...)
```

Arguments

x a phylo object
... additional parameters

as_igraph,data.frame-method
as_igraph

Description

Convert an phylo object to an igraph object, while keeping the weight (in contrast to igraph::as.igraph)

Usage

```
## S4 method for signature 'data.frame'  
as_igraph(x, config)
```

Arguments

x a phylo object
config a 'lineage_tree_config' object

Value

an igraph object

as_igraph,phylo-method
as_igraph

Description

Convert an phylo object to an igraph object, while keeping the weight (in contrast to igraph::as.igraph)

Usage

```
## S4 method for signature 'phylo'  
as_igraph(x)
```

Arguments

x a phylo object

Value

an igraph object

as_lineage_tree *Generic function for as_lineage_tree*

Description

Generic function for as_lineage_tree

Usage

```
as_lineage_tree(x, y, config, ...)
```

Arguments

x	a phyDat object
y	a phylo object
config	a lineage_tree_config object
...	additional parameters

as_lineage_tree,phyDat,phylo,lineage_tree_config-method
as_lineage_tree

Description

Convert a phylo object and a phyDat object to a lineage_tree object

Usage

```
## S4 method for signature 'phyDat,phylo,lineage_tree_config'  
as_lineage_tree(x, y, config, ...)
```

Arguments

x	a phyDat object
y	a phylo object
config	a lineage_tree_config object
...	additional parameters

Value

a lineage_tree object

as_phylo	<i>Generic function for as_phylo</i>
----------	--------------------------------------

Description

Generic function for as_phylo

Usage

```
as_phylo(x, ...)
```

Arguments

x	a graph object
...	additional parameters

as_phylo, igraph-method	<i>as_phylo</i>
-------------------------	-----------------

Description

Convert an igraph object to a phylo object

Usage

```
## S4 method for signature 'igraph'  
as_phylo(x)
```

Arguments

x	an igraph object
---	------------------

Value

a phylo object or a igraph object

DCLEAR	<i>DCLEAR: A package for DCLEAR: Distance based Cell LinEAge Re- construction</i>
--------	---

Description

Distance based methods for inferring lineage trees from single cell data

dist_kmer_replacement_inference

Core function of computing kmer replacement distance

Description

Compute the sequence distance matrix using inferred kmer replacement matrix

Usage

```
dist_kmer_replacement_inference(x, kmer_summary, k = 2)
```

Arguments

x	input data in phyDat format
kmer_summary	a kmer_summary object
k	k-mers (default k=2)

Value

a dist object

Author(s)

Wuming Gong (gongx030@umn.edu)

dist_replacement

Generic function for dist_replacement

Description

Generic function for dist_replacement

Usage

```
dist_replacement(x, kmer_summary, k, ...)
```

Arguments

x	a sequence object
kmer_summary	a kmer_summary object
k	k-mer length
...	additional parameters

dist_replacement, phyDat, kmer_summary, integer-method
Compute the kmer replacement distance

Description

Compute the kmer replacement distance between sequences

Usage

```
## S4 method for signature 'phyDat,kmer_summary,integer'  
dist_replacement(x, kmer_summary, k = 2, ...)
```

Arguments

x	input data in phyDat format
kmer_summary	a kmer_summary object
k	k-mer length
...	other arguments passed to substr_kmer

Value

a dist object

Author(s)

Wuming Gong (gongx030@umn.edu)

dist_replacement, phyDat, missing, integer-method
Compute the kmer replacement distance

Description

Compute the kmer replacement distance between sequences

Usage

```
## S4 method for signature 'phyDat,missing,integer'  
dist_replacement(x, kmer_summary, k = 2L, ...)
```

Arguments

x	input data in phyDat format
kmer_summary	a kmer_summary object
k	k-mer length
...	other arguments passed to substr_kmer

Value

a dist object

Author(s)

Wuming Gong (gongx030@umn.edu)

`dist_weighted_hamming` *Generic function for dist_weighted_hamming*

Description

Generic function for `dist_weighted_hamming`

Usage

```
dist_weighted_hamming(x, wVec, ...)
```

Arguments

x	a sequence object
wVec	weight vector
...	additional parameters

`dist_weighted_hamming,phyDat,numeric-method`
dist_weighted_hamming

Description

implementation of weighted hamming algorithm

Usage

```
## S4 method for signature 'phyDat,numeric'
dist_weighted_hamming(x, wVec, dropout = FALSE)
```

Arguments

x	Sequence object of 'phyDat' type.
wVec	Weight vector for the calculation of weighted hamming distance
dropout	Different weighting strategy is taken to consider interval dropout with dropout = 'TRUE'. Default is, dropout = 'FALSE'.

Value

Calculated distance matrix of input sequences. The result is a 'dist' class object.

Author(s)

Il-Youp Kwak

Examples

```

library(DCLEAR)
library(phangorn)
library(ape)

set.seed(1)
mu_d1 = c( 30, 20, 10, 5, 5, 1, 0.01, 0.001)
mu_d1 = mu_d1/sum(mu_d1)
simn = 10 # number of cell samples
m = 10 ## number of targets
sD = sim_seqdata(sim_n = simn, m = m, mu_d = 0.03,
  d = 12, n_s = length(mu_d1), outcome_prob = mu_d1, p_d = 0.005 )
## RF score with hamming distance
D_hm = dist.hamming(sD$seqs)
tree_hm = NJ(D_hm)
RF.dist(tree_hm, sD$tree, normalize = TRUE)

## RF score with weighted hamming
InfoW = -log(mu_d1)
InfoW[1:2] = 1
InfoW[3:7] = 4.5
D_wh = dist_weighted_hamming(sD$seqs, InfoW, dropout = FALSE)
tree_wh = NJ(D_wh)
RF.dist(tree_wh, sD$tree, normalize = TRUE)

## RF score with weighted hamming, considering dropout situation
nfoW = -log(mu_d1)
InfoW[1] = 1
InfoW[2] = 12
InfoW[3:7] = 3
D_wh2 = dist_weighted_hamming(sD$seqs, InfoW, dropout = TRUE)
tree_wh2 = NJ(D_wh2)
RF.dist(tree_wh2, sD$tree, normalize = TRUE)

```

downsample *Generic function for downsample*

Description

Generic function for downsample

Usage

```
downsample(x, ...)
```

Arguments

x a data object
... additional parameters

downsample,igraph-method
downsample

Description

Sample a lineage tree

Usage

```
## S4 method for signature 'igraph'  
downsample(x, n = 10L, ...)
```

Arguments

x a igraph object
n number of leaves (tips) in the down-sampled tree
... additional parameters

Value

a phylo object

downsample,lineage_tree-method
downsample

Description

Sample a lineage tree

Usage

```
## S4 method for signature 'lineage_tree'  
downsample(x, n = 10L, ...)
```

Arguments

x	a lineage_tree object
n	number of leaves (tips) in the down-sampled tree
...	additional parameters

Value

a lineage_tree object

get_distance_prior *get_distance_prior*

Description

prior distribution of distance

Usage

```
get_distance_prior(x)
```

Arguments

x	a kmer_summary object
---	-----------------------

Value

a probabilistic vector of the distribution of nodal distances

Author(s)

Wuming Gong (gongx030@umn.edu)

<code>get_leaves</code>	<i>Generic function for <code>get_leaves</code></i>
-------------------------	---

Description

Generic function for `get_leaves`

Usage

```
get_leaves(x, ...)
```

Arguments

<code>x</code>	a <code>lineage_tree</code> object
<code>...</code>	additional parameters

<code>get_leaves,lineage_tree-method</code>
<i>get_leaves</i>

Description

Get the leaf sequences

Usage

```
## S4 method for signature 'lineage_tree'  
get_leaves(x, ...)
```

Arguments

<code>x</code>	a <code>lineage_tree</code> object
<code>...</code>	additional parameters

Value

a `phyDat` object

get_node_names *get_node_names*

Description

Convenient function for get node names

Usage

```
get_node_names(x)
```

Arguments

x node id

Value

node names

Author(s)

Wuming Gong (gongx030@umn.edu)

get_replacement_probability
 get_replacement_probability

Description

Compute $p(A, B|d)$, the conditional probability of seeing a replacement of from kmer A to B or vice versa

Usage

```
get_replacement_probability(x)
```

Arguments

x a kmer_summary object

Value

an 3D probabilistic array (kmers by kmers by distances)

Author(s)

Wuming Gong (gongx030@umn.edu)

get_sequence	<i>get_sequence</i>
--------------	---------------------

Description

Get sequencees

Usage

```
get_sequence(x, tree, outcome, config)
```

Arguments

x	a character matrix
tree	a matrix representing the lineage tree
outcome	a character matrix
config	a lineage_tree_config object

Value

a character matrix

get_transition_probability	<i>get_transition_probability</i>
----------------------------	-----------------------------------

Description

Compute $p(A, X|B, Y, d)$, the conditional probability of seeing a replacement from A to B given the previous replacement B from Y at nodal distance d

Usage

```
get_transition_probability(x)
```

Arguments

x	a kmer_summary object
---	-----------------------

Value

an 3D probabilistic array (kmers by kmers by distances)

Author(s)

Wuming Gong (gongx030@umn.edu)

lineages	<i>Lineage data</i>
----------	---------------------

Description

Lineage data

Usage

```
data(lineages)
```

Format

An object of class `list` of length 100.

Examples

```
data(lineages)
```

positional_mutation_prob	<i>positional_mutation_prob</i>
--------------------------	---------------------------------

Description

Convenient function for get node names

Usage

```
positional_mutation_prob(x, config)
```

Arguments

x	a phyDat object
config	a lineage_tree_config object

Value

a positional mutation probability matrix

process_sequence *Generic function for process_sequence*

Description

Generic function for process_sequence

Usage

```
process_sequence(x, ...)
```

Arguments

x	a sequence object
...	additional parameters

process_sequence, phyDat-method
Process sequences

Description

Process sequences

Usage

```
## S4 method for signature 'phyDat'
process_sequence(
  x,
  division = 16L,
  dropout_character = "*",
  default_character = "0",
  deletion_character = "-"
)
```

Arguments

x	input data in phyDat format
division	cell division
dropout_character	Dropout character (default: '*')
default_character	Default character (default: '0')
deletion_character	Deletion character (default: '-')

Value

a 'lineage_tree_config' object

Author(s)

Wuming Gong (gongx030@umn.edu)

prune	<i>Generic function for prune</i>
-------	-----------------------------------

Description

Generic function for prune

Usage

```
prune(x, ...)
```

Arguments

x	a lineage_tree object
...	additional parameters

prune, igraph-method	<i>prune</i>
----------------------	--------------

Description

Trim a full lineage tree into phylogenetic tree

Usage

```
## S4 method for signature 'igraph'
prune(x, weighted = TRUE, ...)
```

Arguments

x	an igraph object
weighted	whether or not keep the edge weight (default: TRUE)
...	additional parameters

Value

an igraph object

prune, lineage_tree-method
prune

Description

Trim a full lineage tree into phylogenetic tree

Usage

```
## S4 method for signature 'lineage_tree'  
prune(x, ...)
```

Arguments

x	a lineage_tree object
...	additional parameters passed to as_phylo()

Value

a lineage_tree object

random_tree *random_tree*

Description

Simulate a random lineage tree

Usage

```
random_tree(n_samples, division = 16L)
```

Arguments

n_samples	number of samples to simulate
division	number of cell division

Value

a data frame

Author(s)

Wuming Gong (gongx030@umn.edu)

rbind,phyDat-method *rbind*

Description

Concatenate multiple phyDat objects

Usage

```
## S4 method for signature 'phyDat'
rbind(..., deparse.level = 1)
```

Arguments

... a list of phyDat objects
deparse.level see definition in generic rbind

Value

a phyDat object

sample_mutation_outcome
sample_mutation_outcome

Description

Sample mutation outcome

Usage

```
sample_mutation_outcome(x, mp = NULL, config)
```

Arguments

x an igraph object
mp a mutation site matrix
config a lineage_tree_config object

Value

a outcome matrix

sample_mutation_site *sample_mutation_site*

Description

Sample mutation site

Usage

```
sample_mutation_site(tree, config)
```

Arguments

tree	a data frame
config	a lineage_tree_config object

Value

a mutation site matrix

sample_outcome_prob *sample_outcome_prob*

Description

Sampling outcome probability based on a gamma distribution

Usage

```
sample_outcome_prob(config, num_states = 20L, shape = 0.1, scale = 2)
```

Arguments

config	a lineage_tree_config object
num_states	number of states used in simulation.
shape	shape parameter in gamma distribution
scale	scale parameter in gamma distribution

Value

a probability vector for each alphabet

Author(s)

Wuming Gong (gongx030@umn.edu)

score_simulation	<i>score_simulation</i>
------------------	-------------------------

Description

Compare two sets of sequences

Usage

```
score_simulation(x, y, config)
```

Arguments

x	a character matrix
y	a character matrix
config	a lineage_tree_config object

Value

numeric scores

simulate	<i>Generic function for simulate</i>
----------	--------------------------------------

Description

Generic function for simulate

Usage

```
simulate(config, x, ...)
```

Arguments

config	a lineage_tree_config object
x	a sequence object
...	additional parameters

simulate,lineage_tree_config,missing-method

simulate

Description

Simulate a cell lineage tree Adoped from https://github.com/elifesciences-publications/CRISPR_recorders_sims/blob/master

Usage

```
## S4 method for signature 'lineage_tree_config,missing'  
simulate(config, x, n_samples = 200, ...)
```

Arguments

config	simulation configuration; a lineage_tree_config object
x	missing
n_samples	number of samples to simulate
...	additional parameters

Value

a lineage_tree object

Author(s)

Wuming Gong (gongx030@umn.edu)

simulate,lineage_tree_config,phyDat-method

simulate

Description

Simulate a cell lineage tree based on a set of sequences

Usage

```
## S4 method for signature 'lineage_tree_config,phyDat'  
simulate(config, x, n_samples = 200L, k = 50, greedy = TRUE, ...)
```


Arguments

config	simulation configuration; a lineage_tree_config object
x	a sequence object
n_samples	number of samples to simulate
k	Number of trials
greedy	Whether or not use a greedy search
...	additional parameters

Value

a lineage_tree object

Author(s)

Wuming Gong (gongx030@umn.edu)

simulate_core	<i>simulate_core</i>
---------------	----------------------

Description

Simulate a cell lineage tree Adoped from https://github.com/elifesciences-publications/CRISPR_recorders_sims/blob/master

Usage

```
simulate_core(config, tree, mutation_site, outcome)
```

Arguments

config	simulation configuration; a lineage_tree_config object
tree	a matrix representing the lineage tree
mutation_site	a binary matrix indicating the mutation sites
outcome	a character matrix

Value

a 'lineage_tree' object

`sim_seqdata`*sim_seqdata*

Description

Generate single cell barcode data set with tree shaped lineage information

Usage

```
sim_seqdata(  
  sim_n = 200,  
  m = 200,  
  mu_d = 0.03,  
  d = 15,  
  n_s = 23,  
  outcome_prob = NULL,  
  p_d = 0.003  
)
```

Arguments

<code>sim_n</code>	Number of cell samples to simulate.
<code>m</code>	Number of targets.
<code>mu_d</code>	Mutation rate. (a scalar or a vector)
<code>d</code>	Number of cell divisions.
<code>n_s</code>	Number of possible outcome states
<code>outcome_prob</code>	Outcome probability vector (default is NULL)
<code>p_d</code>	Dropout probability

Value

The result is a list containing two objects, 'seqs' and 'tree'. The 'seqs' is 'phyDat' object of 'sim_n' number of simulated barcodes corresponding to each cell, and The 'tree' is a 'phylo' object, a ground truth tree structure for the simulated data.

Author(s)

Il-Youp Kwak

Examples

```
library(DCLEAR)  
library(phangorn)  
library(ape)
```

```

set.seed(1)
mu_d1 = c( 30, 20, 10, 5, 5, 1, 0.01, 0.001)
mu_d1 = mu_d1/sum(mu_d1)
simn = 10 # number of cell samples
m = 10 ## number of targets
sD = sim_seqdata(sim_n = simn, m = m, mu_d = 0.03,
                 d = 12, n_s = length(mu_d1), outcome_prob = mu_d1, p_d = 0.005 )
## RF score with hamming distance
D_hm = dist.hamming(sD$seqs)
tree_hm = NJ(D_hm)
RF.dist(tree_hm, sD$tree, normalize = TRUE)

## RF score with weighted hamming
InfoW = -log(mu_d1)
InfoW[1:2] = 1
InfoW[3:7] = 4.5
D_wh = dist_weighted_hamming(sD$seqs, InfoW, dropout=FALSE)
tree_wh = NJ(D_wh)
RF.dist(tree_wh, sD$tree, normalize = TRUE)

## RF score with weighted hamming, cosidering dropout situation
nfoW = -log(mu_d1)
InfoW[1] = 1
InfoW[2] = 12
InfoW[3:7] = 3
D_wh2 = dist_weighted_hamming(sD$seqs, InfoW, dropout = TRUE)
tree_wh2= NJ(D_wh2)
RF.dist(tree_wh2, sD$tree, normalize = TRUE)

```

substr_kmer

Generic function for substr_kmer

Description

Generic function for substr_kmer

Usage

```
substr_kmer(x, ...)
```

Arguments

x	a kmer object
...	additional parameters

substr_kmer, kmer_summary-method
Subsetting a kmer_summary object

Description

Summarize the short k-mer summary from the long k-mer summary

Usage

```
## S4 method for signature 'kmer_summary'  
substr_kmer(x, k = 2)
```

Arguments

x	a kmer_summary object
k	k-mer length(default: 2)

Value

a new kmer_summary object

Author(s)

Wuming Gong (gongx030@umn.edu)

subtract *Generic function for subtract*

Description

Generic function for subtract

Usage

```
subtract(x, y, ...)
```

Arguments

x	a lineage_tree object
y	a lineage_tree object
...	additional parameters

`subtract,lineage_tree,lineage_tree-method`
subtract

Description

Subtract a subtree from a large tree

Usage

```
## S4 method for signature 'lineage_tree,lineage_tree'  
subtract(x, y, ...)
```

Arguments

<code>x</code>	a <code>lineage_tree</code> object
<code>y</code>	a <code>lineage_tree</code> object
<code>...</code>	additional parameters

Value

a `lineage_tree` object

`subtree` *Generic function for subtree*

Description

Generic function for subtree

Usage

```
subtree(x, ...)
```

Arguments

<code>x</code>	a <code>lineage_tree</code> object
<code>...</code>	additional parameters

subtree,lineage_tree-method
subtree

Description

Extract a subtree with specific leaves

Usage

```
## S4 method for signature 'lineage_tree'
subtree(x, leaves = NULL, ...)
```

Arguments

x	a lineage_tree object
leaves	leaves of the extracted tree
...	additional parameters

Value

a lineage_tree object

subtree,phylo-method *subtree*

Description

Extract a subtree with specific leaves

Usage

```
## S4 method for signature 'phylo'
subtree(x, leaves = NULL, ...)
```

Arguments

x	a phylo object
leaves	leaves of the extracted tree
...	additional parameters

Value

a phylo object

summarize_kmer	<i>Generic function for summarize_kmer</i>
----------------	--

Description

Generic function for summarize_kmer

Usage

```
summarize_kmer(x, ...)
```

Arguments

x	a sequence object
...	additional parameters

summarize_kmer, phyDat-method	<i>summarize_kmer</i>
-------------------------------	-----------------------

Description

Summarize kmer distributions with input sequences

Usage

```
## S4 method for signature 'phyDat'
summarize_kmer(
  x,
  division = 16L,
  k = 2,
  reps = 20L,
  n_samples = 200L,
  n_nodes = 100L,
  n_targets
)
```

Arguments

x	input data as a phyDat object
division	number of cell division
k	k-mer (default = 2)
reps	number of simulated trees
n_samples	number of samples to simulate

n_nodes	number of nodes to sample (including both leaves and interval nodes)
n_targets	sequence length. If this argument is missing, the length of the input sequences will be used.

Value

a kmer_summary object

Author(s)

Wuming Gong (gongx030@umn.edu)

summarize_kmer_core	<i>summarize_kmer_core</i>
---------------------	----------------------------

Description

Summarize kmer distributions (core function)

Usage

```
summarize_kmer_core(
  k = 2,
  reps = 20L,
  n_samples = 200L,
  n_nodes = 100L,
  config = NULL
)
```

Arguments

k	k-mer (default = 2)
reps	number of simulated trees
n_samples	number of samples to simulate
n_nodes	number of nodes to sample (including both leaves and interval nodes)
config	lineage tree configuration (a lineage_tree_config object)

Value

a kmer_summary object

Author(s)

Wuming Gong (gongx030@umn.edu)

 WH

 WH

Description

implementation of weighted hamming algorithm

Usage

```
WH(x, InfoW, dropout = FALSE)
```

Arguments

x	Sequence object of 'phyDat' type.
InfoW	Weight vector for the calculation of weighted hamming distance
dropout	Different weighting strategy is taken to consider interval dropout with dropout = 'TRUE'. Default is, dropout = 'FALSE'.

Value

Calculated distance matrix of input sequences. The result is a 'dist' class object.

Author(s)

Il-Youp Kwak

Examples

```
set.seed(1)
library(phangorn)
mu_d1 = c( 30, 20, 10, 5, 5, 1, 0.01, 0.001)
mu_d1 = mu_d1/sum(mu_d1)
simn = 10 # number of cell samples
m = 10 ## number of targets
sD = sim_seqdata(sim_n = simn, m = m, mu_d = 0.03,
                 d = 12, n_s = length(mu_d1), outcome_prob = mu_d1, p_d = 0.005 )

## RF score with hamming distance
D_h = dist.hamming(sD$seqs)
tree_h= NJ(D_h)
RF.dist(tree_h, sD$tree, normalize = TRUE)

## RF score with weighted hamming
InfoW = -log(mu_d1)
InfoW[1:2] = 1
InfoW[3:7] = 4.5

D_wh = WH(sD$seqs, InfoW)
```

```

tree_wh= NJ(D_wh)
RF.dist(tree_wh, sD$tree, normalize = TRUE)

## RF score with weighted hamming, considering dropout situation
nfoW = -log(mu_d1)
InfoW[1] = 1
InfoW[2] = 12
InfoW[3:7] = 3

D_wh2 = WH(sD$seqs, InfoW, dropout=TRUE)
tree_wh2= NJ(D_wh2)
RF.dist(tree_wh2, sD$tree, normalize = TRUE)

```

WH_train

Train weights for WH

Description

Train weights for WH and output weight vector

Usage

```
WH_train(X, loc0 = 2, locDropout = 1, locMissing = FALSE)
```

Arguments

X	a list of k number of input data, X[[1]] ... X[[k]]. The ith data have sequence information as phyDat format in X[[i]][[1]], and tree information in X[[i]][[2]] as phylo format.
loc0	weight location of initial state
locDropout	weight location of dropout state
locMissing	weight location of missing state, FALSE if there is no missing values

Value

a weight vector

Author(s)

Il-Youp Kwak (ikwak2@cau.ac.kr)

WH_train_fit	<i>Train weights for WH, and output distance object</i>
--------------	---

Description

Train weights for WH using the given data, and fit the distance matrix for a input sequence.

Usage

```
WH_train_fit(x, X)
```

Arguments

x	input data in phyDat format
X	a list of k number of input data, X[[1]] ... X[[k]]. The ith data have sequence information as phyDat format in X[[i]][[1]], and tree information in X[[i]][[2]] as phylo format.

Value

a dist object

Author(s)

Il-Youp Kwak (ikwak2@cau.ac.kr)

Index

- * **datasets**
 - lineages, [17](#)
- add_deletion, [3](#)
- add_dropout, [4](#)
- as_igraph, [4](#)
- as_igraph, data.frame-method, [5](#)
- as_igraph, phylo-method, [5](#)
- as_lineage_tree, [6](#)
- as_lineage_tree, phyDat, phylo, lineage_tree_config-method, [6](#)
- as_phylo, [7](#)
- as_phylo, igraph-method, [7](#)
- DCLEAR, [7](#)
- dist_kmer_replacement_inference, [8](#)
- dist_replacement, [8](#)
- dist_replacement, phyDat, kmer_summary, integer-method, [9](#)
- dist_replacement, phyDat, missing, integer-method, [9](#)
- dist_weighted_hamming, [10](#)
- dist_weighted_hamming, phyDat, numeric-method, [10](#)
- downsample, [12](#)
- downsample, igraph-method, [12](#)
- downsample, lineage_tree-method, [13](#)
- get_distance_prior, [13](#)
- get_leaves, [14](#)
- get_leaves, lineage_tree-method, [14](#)
- get_node_names, [15](#)
- get_replacement_probability, [15](#)
- get_sequence, [16](#)
- get_transition_probability, [16](#)
- lineages, [17](#)
- positional_mutation_prob, [17](#)
- process_sequence, [18](#)
- process_sequence, phyDat-method, [18](#)
- prune, [19](#)
- prune, igraph-method, [19](#)
- prune, lineage_tree-method, [20](#)
- random_tree, [20](#)
- rbind, phyDat-method, [21](#)
- sample_mutation_outcome, [21](#)
- sample_mutation_site, [22](#)
- sample_mutation_outcome_prob, [22](#)
- score_simulation, [23](#)
- sim_seqdata, [26](#)
- simulate, [23](#)
- simulate, lineage_tree_config, missing-method, [24](#)
- simulate, lineage_tree_config, phyDat-method, [24](#)
- simulate_core, [25](#)
- substr_kmer, [27](#)
- substr_kmer, kmer_summary-method, [28](#)
- subtract, [28](#)
- subtract, lineage_tree, lineage_tree-method, [29](#)
- subtree, [29](#)
- subtree, lineage_tree-method, [30](#)
- subtree, phylo-method, [30](#)
- summarize_kmer, [31](#)
- summarize_kmer, phyDat-method, [31](#)
- summarize_kmer_core, [32](#)
- WH, [33](#)
- WH_train, [34](#)
- WH_train_fit, [35](#)