R topics documented:  

- as_lineage_tree ........................................... 4  
- as_lineage_tree.phyDat.phylo,lineage_tree_config-method 4  
- as_phylo .................................................. 5  
- as_phylo.igraph-method .................................... 5  
- DCLEAR ..................................................... 5  
- dist_kmer_replacement_inference ............................ 6  
- dist_replacement ........................................... 6  
- dist_replacement.phyDat.kmer_summary.integer-method 7  
- dist_replacement.phyDat.missing.integer-method ........ 7  
- dist_weighted_hamming .................................... 8  
- dist_weighted_hamming.phyDat.numeric-method ........... 8  
- downsample ............................................... 10  
- downsample.igraph-method ................................ 10  
- downsample.lineage_tree-method .......................... 11  
- get_distance_prior ........................................ 11  
- get_leaves ............................................... 12  
- get_leaves.lineage_tree-method ......................... 12  
- get_node_names ............................................ 13  
- get_replacement_probability ............................... 13  
- get_transition_probability ................................ 14  
- lineages ................................................... 14  
- positional_mutation_prob .................................. 15  
- process_sequence .......................................... 15  
- process_sequence.phyDat-method .......................... 16  
- prune ...................................................... 16  
- prune.igraph-method ..................................... 17  
- prune.lineage_tree-method ................................ 17  
- random_tree .............................................. 18  
- rbind.phyDat-method ..................................... 18  
- sample_outcome_prob ...................................... 19  
- simulate .................................................... 19  
- simulate.lineage_tree_config.missing-method .......... 20  
- simulate.lineage_tree_config.phyDat-method .......... 20  
- simulate_core ............................................. 21  
- sim_seqdata .............................................. 22  
- substr_kmer ............................................... 23  
- substr_kmer.kmer_summary-method ....................... 24  
- subtract ................................................... 24  
- subtract.lineage_tree.lineage_tree-method .............. 25  
- subtree .................................................... 25  
- subtree.lineage_tree-method ................................ 26  
- subtree.phylo-method ..................................... 26  
- summarize_kmer ............................................ 27  
- summarize_kmer.phyDat-method ............................ 27  
- summarize_kmer_core ..................................... 28  
- WH ......................................................... 29  
- WH_train .................................................. 30  
- WH_train_fit ............................................. 31
as_igraph

Description

Generic function for as_igraph

Usage

as_igraph(x, ...)

Arguments

x a phylo object
... additional parameters

Value

an igraph object
as_lineage_tree

Generic function for as_lineage_tree

Description

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

Usage

```r
## 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

Generic function for as_phylo

Description

Generic function for as_phylo

Usage

as_phylo(x, ...)

Arguments

x

a graph object

...  

additional parameters

as_phylo,igraph-method

as_phylo

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

DCLEAR: A package for DCLEAR: Distance based Cell LinEAge Re-construction

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
Compute the kmer replacement distance

Description

Compute the kmer replacement distance between sequences

Usage

## S4 method for signature 'phyDat,kmer_summary,integer'

```r
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)
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)

Description

Generic function for dist_weighted_hamming

Usage

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

Arguments

- **x**: a sequence object
- **wVec**: weight vector
- **...**: additional parameters

Description

implementation of weighted hamming algorithm

Usage

```r
## 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

```r
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
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
InfoW = -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**
```r
## 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**

```r
## 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**

**Description**

prior distribution of distance

**Usage**

```r
generate_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)
get_leaves

Generic function for get_leaves

Description

Generic function for get_leaves

Usage

get_leaves(x, ...)

Arguments

x  a lineage_tree object
...
  additional parameters

get_leaves,lineage_tree-method

Description

Get the leaf sequences

Usage

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

Arguments

x  a lineage_tree object
...
  additional parameters

Value

a phyDat object
**get_node_names**

**Description**
Convenient function for get node names

**Usage**
get_node_names(x)

**Arguments**
x

**Value**
node names

**Author(s)**
Wuming Gong (gongx030@umn.edu)

---

**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

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

**Author(s)**
Wuming Gong (gongx030@umn.edu)
get_transition_probability

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 (kmer by kmers by distances)

Author(s)
Wuming Gong (gongx030@umn.edu)

lineages
Lineage data

Description
Lineage data

Usage
data(lineages)

Format
An object of class list of length 100.

Examples
data(lineages)
**positional_mutation_prob**

Description

Convenient function for get node names

Usage

```r
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

```r
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)

Arguments

x input data in phyDat format
division cell divisions (default: 16L)

Value

a ‘lineage_tree_config’ object

Author(s)

Wuming Gong (gongx030@umn.edu)

---

prune

Generic function for prune

Description

Generic function for prune

Usage

prune(x, ...)

Arguments

x a lineage_tree object
... additional parameters
prune.igraph-method

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

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

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

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_outcome_prob**

**Description**
Sampling outcome probability based on a gamma distribution

**Usage**
```r
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)

---

**simulate**

*Generic function for simulate*

**Description**
Generic function for simulate

**Usage**
```r
simulate(config, x, ...)
```

**Arguments**
- `config`: a lineage_tree_config object
- `x`: a sequence object
- `...`: additional parameters
Description

Simulate a cell lineage tree. Adoped from https://github.com/elifesciences-publications/CRISPR_recorders_sims/blob/master/MATLAB_sims/GESTALT_30hr_1x_simulation.m

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)

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, ...)

...
**Arguments**

- `config`: simulation configuration; a lineage_tree_config object
- `x`: a sequence object
- `n_samples`: number of samples to simulate
- `...`: additional parameters

**Value**

- a lineage_tree object

**Author(s)**

Wuming Gong (gongx030@umn.edu)

---

**Description**

Simulate a cell lineage tree Adopted from https://github.com/elifesciences-publications/CRISPR_recorders_sims/blob/master/MATLAB_sims/GESTALT_30hr_1x_simulation.m

**Usage**

```r
simulate_core(config, mp = NULL, n_samples = 200L, ...)
```

**Arguments**

- `config`: simulation configuration; a lineage_tree_config object
- `mp`: site specific mutation probability
- `n_samples`: number of samples to simulate
- `...`: additional parameters
Description

Generate singe cell barcode data set with tree shaped lineage information

Usage

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

Arguments

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

```r
library(DCLEAR)
library(phangorn)
library(ape)
```
set.seed(1)
mu_d1 = c(30, 20, 10, 5, 5, 1, 0.01, 0.001)/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
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)

substr_kmer  

Generic function for substr_kmer

Description

Generic function for substr_kmer

Usage

substr_kmer(x, ...)

Arguments

x a kmer object
... additional parameters
Description

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

Usage

```r
## 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

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

Arguments

- `x`: a `lineage_tree` object
- `y`: a `lineage_tree` object
- `...`: additional parameters
Description

Subtract a subtree from a large tree

Usage

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

Arguments

- `x`: a `lineage_tree` object
- `y`: a `lineage_tree` object
- `...`: additional parameters

Value

a `lineage_tree` object

Description

Generic function for `subtree`

Usage

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

Arguments

- `x`: a `lineage_tree` object
- `...`: additional parameters
subtree, lineage_tree-method

subtree

Description

Extract a subtree with specific leaves

Usage

```r
# 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

```r
# 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

Generic function for summarize_kmer

Description

Generic function for summarize_kmer

Usage

summarize_kmer(x, ...)

Arguments

x a sequence object
...
additional parameters

summarize_kmer,phyDat-method

summarize_kmer

Description

Summarize kmers 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
summarize_kmer_core

\begin{verbatim}
  n_nodes number of nodes to sample (including both leaves and internal 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)
\end{verbatim}

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 internal nodes)
config  lineage tree configuration (a lineage_tree_config object)

Value

a kmer_summary object

Author(s)

Wuming Gong (gongx030@umn.edu)
Description

implementation of weighted hamming algorithm

Usage

\texttt{WH(x, InfoW, dropout = FALSE)}

Arguments

\begin{itemize}
  \item \texttt{x} \hspace{1cm} Sequence object of 'phyDat' type.
  \item \texttt{InfoW} \hspace{1cm} Weight vector for the calculation of weighted hamming distance
  \item \texttt{dropout} Different weighting strategy is taken to consider interval dropout with dropout = 'TRUE'. Default is, dropout = 'FALSE'.
\end{itemize}

Value

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

Author(s)

Il-Youp Kwak

Examples

\begin{verbatim}
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
D_wh = WH(sD$seqs, InfoW)
\end{verbatim}
tree_wh = NJ(D_wh)
RF.dist(tree_wh, sD$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

Train weights for WH, and output distance object

Description

Train weights for WH using the given data, and fit the distance matrix for an 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, 14
  as_igraph, 3
  as_igraph, phylo-method, 3
  as_lineage_tree, 4
  as_lineage_tree, phyDat, phylo, lineage_tree_config-method, 4
  as_phylo, 5
  as Phylo, igraph-method, 5
  DCLEAR, 5
  dist_kmer_replacement_inference, 6
  dist_replacement, 6
  dist_replacement, phyDat, kmer_summary, integer-method, 7
  dist_replacement, phyDat, missing, integer-method, 7
  dist_weighted_hamming, 8
  dist_weighted_hamming, phyDat, numeric-method, 8
  downsample, 10
  downsample, igraph-method, 10
  downsample, lineage_tree-method, 11
  get_distance_prior, 11
  get_leaves, 12
  get_leaves, lineage_tree-method, 12
  get_node_names, 13
  get_replacement_probability, 13
  get_transition_probability, 14
  lineages, 14
  positional_mutation_prob, 15
  process_sequence, 15
  process_sequence, phyDat-method, 16
  prune, 16
  prune, igraph-method, 17
  prune, lineage_tree-method, 17
  random_tree, 18
  rbind, phyDat-method, 18
  sample_outcome_prob, 19
  sim_seqdata, 22
  simulate, 19
  simulate, lineage_tree_config, missing-method, 20
  simulate, lineage_tree_config, phyDat-method, 20
  simulate_core, 21
  substr_kmer, 23
  substr_kmer, kmer_summary-method, 24
  subtract, 24
  subtree, lineage_tree-method, 25
  subtree, phylo-method, 26
  summarize_kmer, 27
  summarize_kmer, phyDat-method, 27
  summarize_kmer_core, 28
  WH, 29
  WH_train, 30
  WH_train_fit, 31