

Package ‘GeRnika’

March 27, 2025

Type Package

Title Simulation, Visualization and Comparison of Tumor Evolution Data

Version 1.1.0

Description Simulating, visualizing and comparing tumor clonal data by using simple commands. This aims at providing a tool to help researchers to easily simulate tumor data and analyze the results of their approaches for studying the composition and the evolutionary history of tumors.

Encoding UTF-8

Imports data.tree, purrr, reshape2, DiagrammeR, colorspace, dplyr, methods, vctrs, magrittr, stats

Depends R (>= 4.00)

RoxygenNote 7.3.2

VignetteBuilder knitr, rmarkdown, knitcitations, ggpubr, ggplot2

Suggests knitr, rmarkdown, knitcitations, ggpubr, ggplot2

NeedsCompilation no

LazyData TRUE

License GPL (>= 3)

Author Aitor Sánchez-Ferrera [cre, aut]
(<https://orcid.org/0000-0001-6127-0686>),
Maitena Tellaetxe-Abete [aut] (<https://orcid.org/0000-0003-1894-4547>),
Borja Calvo [aut] (<https://orcid.org/0000-0001-9969-9664>)

Maintainer Aitor Sánchez-Ferrera <aitor.sanchezf@ehu.eus>

Repository CRAN

Date/Publication 2025-03-27 19:10:05 UTC

Contents

add_noise	2
B_mats	3
B_to_phylo tree	4

combine_trees	5
create_B	6
create_F	7
create_instance	8
create_phylotree	10
create_U	11
equals	12
find_common_subtrees	13
palettes	14
Phylotree-class	15
plot	15
plot_proportions	16

Index	18
--------------	-----------

add_noise	<i>Add noise to the VAF values in an F matrix</i>
-----------	---------------------------------------------------

Description

This function adds noise to the variant allele frequency (VAF) values in an F matrix, simulating the effect of sequencing errors. The noise is modeled as a negative binomial distribution for the depth of the reads and a binomial distribution for both the variant allele counts and the mismatch counts.

Usage

```
add_noise(F_matrix, depth, overdispersion)
```

Arguments

F_matrix	A matrix representing the true VAF values of a series of mutations in a set of samples (F matrix).
depth	A numeric value representing the mean depth of sequencing.
overdispersion	A numeric value representing the overdispersion parameter for the negative binomial distribution used to simulate the depth of sequencing.

Value

A matrix containing noisy VAF values of a series of mutations in a set of samples.

Examples

```
# Calculate the noisy VAF values of a series of mutations in a set of samples, given the true
# VAF values in the F matrix F_true, a depth of 30 and an overdispersion of 5

# Simulate the noise-free F matrix of a tumor with 50 clones,
# 10 samples, k = 5, following a positive selection model
F_true <- create_instance(
  n = 50,
```

```

m = 10,
k = 5,
selection = "positive",
noisy = FALSE)$F_true

# Then we add the noise using a depth of 30 and an overdispersion of 5.
noisy_F <- add_noise(F_true, 30, 5)

```

B_mats	<i>A set of 10 trios of B matrices for experimenting with the methods of GeRnika</i>
--------	--------------------------------------------------------------------------------------

Description

A list of lists composed by 10 trios of B matrices; a real B matrix, a B matrix got by using the one algorithm (alg1) method and another one as a result of another algorithm (alg2). These matrices can be used as examples for the methods of GeRnika.

Usage

B_mats

Format

A list of lists composed by 10 trios of B matrices.

Trio 1 B_real, B_alg1 and B_alg2 (matrices composed by 5 clones)

Trio 2 B_real, B_alg1 and B_alg2 (matrices composed by 5 clones)

Trio 3 B_real, B_alg1 and B_alg2 (matrices composed by 5 clones)

Trio 4 B_real, B_alg1 and B_alg2 (matrices composed by 5 clones)

Trio 5 B_real, B_alg1 and B_alg2 (matrices composed by 5 clones)

Trio 6 B_real, B_alg1 and B_alg2 (matrices composed by 10 clones)

Trio 7 B_real, B_alg1 and B_alg2 (matrices composed by 10 clones)

Trio 8 B_real, B_alg1 and B_alg2 (matrices composed by 10 clones)

Trio 9 B_real, B_alg1 and B_alg2 (matrices composed by 10 clones)

Trio 10 B_real, B_alg1 and B_alg2 (matrices composed by 10 clones)

Source

Local source; as a result of the Grasp and the ILS methods used for solving the Clonal Deconvolution and Evolution Problem (CDEP).

B_to_phyloree	<i>Create a Phylotree object from a B matrix.</i>
---------------	---------------------------------------------------

Description

This function creates a Phylotree class object from a B matrix.

Usage

```
B_to_phyloree(B, labels = NA)
```

Arguments

B	A square matrix that represents the phylogenetic tree.
labels	An optional vector containing the tags of the genes in the phylogenetic tree. NA by default.

Value

A Phylotree class object.

Examples

```
# Create a B matrix instance
# composed by 10 subpopulations of
# clones
B <- create_instance(
  n = 10,
  m = 4,
  k = 1,
  selection = "neutral")$B

# Create a new 'Phylotree' object
# on the basis of the B matrix
phylotree <- B_to_phyloree(B = B)

# Generate the tags for the genes of
# the phylogenetic tree
tags <- LETTERS[1:nrow(B)]

# Create a new 'Phylotree' object
# on the basis of the B matrix and
# the list of tags
phylotree_tags <- B_to_phyloree(
  B = B,
  labels = tags)
```

combine_trees	<i>Get consensus tree between two phylogenetic trees</i>
---------------	----------------------------------------------------------

Description

Returns a graph representing the consensus tree between two phylogenetic trees.

Usage

```
combine_trees(
  phylotree_1,
  phylotree_2,
  palette = GeRnika::palettes$Simpsons,
  labels = FALSE
)
```

Arguments

phylotree_1	A Phylotree class object.
phylotree_2	A Phylotree class object.
palette	A vector composed by the hexadecimal code of three colors. "The Simpsons" palette used as default.
labels	A boolean, if TRUE the resulting graph will be plotted with the tags of the genes in the phylogenetic trees instead of their mutation index. FALSE by default.

Value

a `dgr_graph` object representing the consensus graph between `phylotree_1` `phylotree_2`.

Examples

```
# Load the predefined B matrices of the package
B_mats <- GeRnika::B_mats

B_real <- B_mats[[2]]$B_real
B_alg1 <- B_mats[[2]]$B_alg1

# Generate the tags for the genes of
# the phylogenetic tree
tags <- LETTERS[1:nrow(B_real)]

# Instantiate two \code{Phylotree} class objects on
# the basis of the B matrices
phylotree_real <- B_to_phylotree(
  B = B_real,
```

```

        labels = tags)

phylotree_alg1 <- B_to_phylotree(
  B = B_alg1,
  labels = tags)

# Create the consensus tree between phylotree_real
# and phylotree_alg1
consensus <- combine_trees(
  phylotree_1 = phylotree_real,
  phylotree_2 = phylotree_alg1)

# Render the consensus tree
DiagrammeR::render_graph(consensus)

# Load another palette
palette_1 <- GeRnika::palettes$Lancet

# Create the consensus tree between phylotree_real
# and phylotree_alg1 using tags and another palette
consensus_tag <- combine_trees(
  phylotree_1 = phylotree_real,
  phylotree_2 = phylotree_alg1,
  palette = palette_1,
  labels = TRUE)

# Render the consensus tree using tags and the
# selected palette
DiagrammeR::render_graph(consensus_tag)

```

create_B

Create tumor phylogenetic tree topology

Description

This function generates a mutation matrix (B matrix) for a tumor phylogenetic tree with a given number of nodes. This matrix represents the topology and it is created randomly, with the probability of a node to be chosen as a parent of a new node being proportional to the number of its ascendants raised to the power of a constant 'k'.

Usage

```
create_B(n, k)
```

Arguments

- n An integer representing the number of nodes in the phylogenetic tree.
- k A numeric value representing the constant used to calculate the probability of a node to be chosen as a parent of a new node.

Value

A square matrix representing the mutation relationships between the nodes in the phylogenetic tree. Each row corresponds to a node, and each column corresponds to a mutation. The value at the i-th row and j-th column is 1 if the i-th node has the j-th mutation, and 0 otherwise.

Examples

```
# Create a mutation matrix for a phylogenetic tree with 10 nodes and k = 2
B <- create_B(10, 2)
```

create_F *Calculate the variant allele frequency (VAF) values in a set of samples*

Description

This method generates the F matrix that contains the mutation frequency values of a series of mutations in a collection of tumor biopsies or samples.

Usage

```
create_F(U, B, heterozygous = TRUE)
```

Arguments

- U A matrix where each row corresponds to a sample, and each column corresponds to a clone. The value at the i-th row and j-th column is the frequency of the j-th clone in the i-th sample.
- B A matrix representing the mutation relationships between the nodes in the phylogenetic tree.
- heterozygous A logical value indicating whether to adjust the clone proportions for heterozygous states. If 'TRUE', the clone proportions are halved. If 'FALSE', the clone proportions are not adjusted. Default is 'TRUE'.

Value

A matrix containing the VAF values of a series of mutations in a set of samples.

Examples

```
# Create random topology with 10 nodes and k = 2
B <- create_B(10, 2)

# Create U matrix with parameter m=4 and "positive" selection
U <- create_U(B = B, m = 4, selection = "positive")

# Then we compute the F matrix for a heterozygous tumor
F <- create_F(U = U, B = B, heterozygous = TRUE)
```

<code>create_instance</code>	<i>Create a tumor phylogenetic tree instance</i>
------------------------------	--------------------------------------------------

Description

This function generates a tumor phylogenetic tree instance, composed by a mutation matrix (B matrix), a matrix of true variant allele frequencies (F_true), a matrix of noisy variant allele frequencies (F), and a matrix of clone frequencies in samples (U).

Usage

```
create_instance(
  n,
  m,
  k,
  selection,
  noisy = TRUE,
  depth = 30,
  seed = Sys.time()
)
```

Arguments

<code>n</code>	An integer representing the number of clones.
<code>m</code>	An integer representing the number of samples.
<code>k</code>	A numeric value that determines the linearity of the tree topology. Also referred to as the topology parameter. Increasing values of this parameter increase the linearity of the topology. When 'k' is set to 1, all nodes have equal probabilities of being chosen as parents, resulting in a completely random topology.
<code>selection</code>	A character string representing the evolutionary mode the tumor follows. This should be either "positive" or "neutral".
<code>noisy</code>	A logical value indicating whether to add noise to the frequency matrix. If 'TRUE', noise is added to the frequency matrix. If 'FALSE', no noise is added. 'TRUE' by default.
<code>depth</code>	A numeric value representing the mean depth of sequencing. 30 by default.
<code>seed</code>	A numeric value used to set the seed for the random number generator. Sys.time() by default.

Details

The B matrix is a square matrix representing the mutation relationships between the clones in the tumor, or, in other words, it represents the topology of the phylogenetic tree. The F_true matrix represents the true variant allele frequencies of the mutations present in the tumor in a set of samples. The F matrix represents the noisy variant allele frequencies of the mutations in the same set of samples. The U matrix represents the frequencies of the clones in the tumor in the set of samples.

Value

A list containing four elements: 'F', a matrix representing the noisy frequencies of each mutation in each sample; 'B', a matrix representing the mutation relationships between the clones in the tumor; 'U', a matrix that represents the frequencies of the clones in the tumor in the set of samples; and 'F_true', a matrix representing the true frequencies of each mutation in each sample.

Examples

```
# Create an instance of a tumor with 10 clones,  
# 4 samples, k = 1, neutral evolution and  
# added noise with depth = 500  
I1 <- create_instance(  
  n = 10,  
  m = 4,  
  k = 1,  
  selection = "neutral",  
  depth = 500)  
  
# Create an instance of a tumor with 50 clones,  
# 10 samples, k = 5, positive selection and  
# added noise with depth = 500  
I2 <- create_instance(  
  n = 50,  
  m = 10,  
  k = 5,  
  selection = "positive",  
  noisy = TRUE,  
  depth = 500)  
  
# Create an instance of a tumor with 100 clones,  
# 25 samples, k = 0, positive selection without  
# noise  
I3 <- create_instance(  
  n = 100,  
  m = 25,  
  k = 0,  
  selection = "positive",  
  noisy = FALSE)
```

create_phyloree *Create a Phylotree object*

Description

This is the general constructor of the Phylotree S4 class.

Usage

```
create_phyloree(B, clones, genes, parents, tree, labels = NA)
```

Arguments

B	A square matrix that represents the phylogenetic tree.
clones	A numeric vector representing the clones in the phylogenetic tree.
genes	A numeric vector representing the genes in the phylogenetic tree.
parents	A numeric vector representing the parents of the clones in the phylogenetic tree.
tree	A data.tree object containing the tree structure of the phylogenetic tree.
labels	An optional vector containing the tags of the genes in the phylogenetic tree. NA by default.

Value

A Phylotree class object.

Examples

```
# Create a B matrix instance
# composed by 10 subpopulations of
# clones
B <- create_instance(
  n = 10,
  m = 4,
  k = 1,
  selection = "neutral")$B

# Create a new 'Phylotree' object
# on the basis of the B matrix
phylotree1 <- B_to_phyloree(B = B)

# Create a new 'Phylotree' object
# with the general constructor of
# the class
phylotree2 <- create_phyloree(
  B = B,
```

```

        clones = phylotree1@clones,
        genes = phylotree1@genes,
        parents = phylotree1@parents,
        tree = phylotree1@tree)

# Generate the tags for the genes of
# the phylogenetic tree
tags <- LETTERS[1:nrow(B)]

# Create a new 'Phylotree' object
# with the general constructor of
# the class using tags
phylotree_tags <- create_phylotree(
    B = B,
    clones = phylotree1@clones,
    genes = phylotree1@genes,
    parents = phylotree1@parents,
    tree = phylotree1@tree,
    labels = tags)

```

create_U

Calculate tumor clone frequencies in samples

Description

This function calculates the frequencies of each clone in a set of samples, given the global clone proportions in the tumor and their spatial distribution.

Usage

```
create_U(B, m, selection, n_cells = 100)
```

Arguments

B	A matrix representing the mutation relationships between the nodes in the phylogenetic tree (B matrix).
m	An integer representing the number of samples taken from the tumor.
selection	A character string representing the evolutionary mode the tumor follows. This should be either "positive" or "neutral".
n_cells	An integer representing the number of cells sampled from the multinomial distribution. Default is 100.

Value

A matrix where each row corresponds to a sample, and each column corresponds to a clone. The value at the i-th row and j-th column is the frequency of the j-th clone in the i-th sample.

Examples

```
# Create random topology with 20 nodes and k = 3
B <- create_B(20, 3)

# Create U matrix with parameter m=4 and "positive" selection
U <- create_U(B = B, m = 4, selection = "positive")
```

equals

Check if two phylogenetic trees are equal

Description

Checks whether two phylogenetic trees are equivalent or not.

Usage

```
equals(phyloree_1, phyloree_2)
```

Arguments

phyloree_1 A Phylotree class object.
phyloree_2 A Phylotree class object.

Value

A boolean, TRUE if they are equal and FALSE if not.

Examples

```
# Load the predefined B matrices of the package
B_mats <- GeRnika::B_mats

B_real <- B_mats[[2]]$B_real
B_alg1 <- B_mats[[2]]$B_alg1

# Instantiate two \code{Phylotree} class objects on
# the basis of the B matrices
phyloree_real <- B_to_phylotree(
  B = B_real)

phyloree_alg1 <- B_to_phylotree(
  B = B_alg1)

equals(phyloree_real, phyloree_alg1)
```

find_common_subtrees *Find the set of common subtrees between two phylogenetic trees*

Description

Plots the common subtrees between two phylogenetic trees and prints the information about their similarities and their differences.

Usage

```
find_common_subtrees(phyloree_1, phyloree_2, labels = FALSE)
```

Arguments

phyloree_1 A Phylotree class object.
phyloree_2 A Phylotree class object.
labels A boolean, if TRUE the rendered graph will be plotted with the tags of the genes in the phylogenetic trees instead of their gene index. FALSE by default.

Value

A plot of the common subtrees between two phylogenetic trees and the information about the distance between them based on their independent and common edges.

Examples

```
# Load the predefined B matrices of the package
B_mats <- GeRnika::B_mats

B_real <- B_mats[[2]]$B_real
B_alg1 <- B_mats[[2]]$B_alg1

# Generate the tags for the genes of
# the phylogenetic tree
tags <- LETTERS[1:nrow(B_real)]

# Instantiate two Phylotree class objects on
# the basis of the B matrices using tags
phyloree_real <- B_to_phylotree(
  B = B_real,
  labels = tags)

phyloree_alg1 <- B_to_phylotree(
  B = B_alg1,
  labels = tags)
```

```
# find the set of common subtrees between both
# phylogenetic trees
find_common_subtrees(
  phylotree_1 = phylotree_real,
  phylotree_2 = phylotree_alg1)

# find the set of common subtrees between both
# phylogenetic trees using tags
find_common_subtrees(
  phylotree_1 = phylotree_real,
  phylotree_2 = phylotree_alg1,
  labels = TRUE)
```

palettes

Palettes for the methods of GeRnika

Description

A data.frame containing 3 default palettes for the parameters used in the methods of GeRnika.

Usage

```
palettes
```

Format

A data.frame containing 3 palettes.

Lancet #0099B444, #AD002A77, #42B540FF

NEJM #FFDC9177, #7876B188, #EE4C97FF

Simpsons #FED43966, #FD744688, #197EC0FF

Source

Lancet, NEJM and The Simpsons palettes; inspired by the plots in Lancet journals, the plots in the New England Journal of Medicine and the colors used in the TV show The Simpsons, respectively.

Phylotree-class	<i>S4 class to represent phylogenetic trees.</i>
-----------------	--------------------------------------------------

Description

S4 class to represent phylogenetic trees.

Slots

B A data.frame containing the square matrix that represents the ancestral relations among the clones of the phylogenetic tree.

clones A vector representing the equivalence table of the clones in the phylogenetic tree.

genes A vector representing the equivalence table of the genes in the phylogenetic tree.

parents A vector representing the parents of the clones in the phylogenetic tree.

tree A Node class object representing the phylogenetic tree.

labels A vector representing the tags of the genes in the phylogenetic tree.

plot	<i>Plot a Phylotree object.</i>
------	---------------------------------

Description

Plot a Phylotree object.

Usage

```
plot(object, labels = FALSE)
```

```
## S4 method for signature 'Phylotree'
```

```
plot(object, labels = FALSE)
```

Arguments

object A Phylotree object.

labels A label vector.

plot_proportions *Plot a phylogenetic tree with proportional node sizes and colors*

Description

This function plots a phylogenetic tree with nodes sized and colored according to the proportions of each clone. If a matrix of proportions is provided, multiple phylogenetic trees will be plotted, each corresponding to a row of proportions.

Usage

```
plot_proportions(phytotree, proportions, labels = FALSE)
```

Arguments

phytotree	A Phylotree class object representing the phylogenetic tree to be plotted.
proportions	A numeric vector or matrix representing the proportions of each clone in the phylogenetic tree. If a matrix is provided, each row should represent the proportions for a separate tree.
labels	A logical value indicating whether to label the nodes with gene tags (if TRUE) or gene indices (if FALSE). Default is FALSE.

Value

A graph representing the phylogenetic tree, with node sizes and colors reflecting clone proportions.

Examples

```
# Create an instance
# composed by 5 subpopulations of clones
# and 4 samples
instance <- create_instance(
  n = 5,
  m = 4,
  k = 1,
  selection = "neutral")

# Extract its associated B matrix
B <- instance$B

# Create a new 'Phylotree' object
# on the basis of the B matrix
phytotree <- B_to_phylotree(B = B)

# Generate the tags for the genes of
# the phylogenetic tree
tags <- LETTERS[1:nrow(B)]
```



```
# Plot the phylogenetic tree taking  
# into account the proportions of the  
# previously generated instance  
plot_proportions(phyloree, instance$U, labels=TRUE)
```

Index

* datasets

B_mats, 3

palettes, 14

add_noise, 2

B_mats, 3

B_to_phylotree, 4

combine_trees, 5

create_B, 6

create_F, 7

create_instance, 8

create_phylotree, 10

create_U, 11

equals, 12

find_common_subtrees, 13

palettes, 14

Phylotree-class, 15

plot, 15

plot, Phylotree-method (plot), 15

plot_proportions, 16