Package 'HiMC'

June 26, 2018

Version	0.1	.1	٠.

Title High-Throughput Mitochondrial Haplogroup Classification

Description Assign high-level mitochondrial haplogroups given SNP information from standard PLINK *.map* and *.ped* files. A reduced phylotree[1] is used as a classification tree to determine samples' haplogroup assignment. Please see the included package vignette ``Hi-MC: Overview Guide" for a detailed algorithm description and usage examples. [1] Mitchell SL, Goodloe R, Brown-Gentry K, Pendergrass SA, Murdock DG, Crawford DC (2014) <doi:10.1007/s00439-014-1421-9>.

Depends R (>= 3.1.2)

Imports stringr, methods

VignetteBuilder knitr

Suggests knitr

License GPL-3

LazyData false

RoxygenNote 6.0.1

NeedsCompilation no

Author Eric Farber-Eger [aut, cre],

Dana Crawford [aut],

Nicholas R. Wheeler [aut] (0000-0003-2248-8919),

Sandra Smieszek [aut]

Maintainer Eric Farber-Eger <eric.h.farber-eger@vumc.org>

R topics documented:

generate_snp_data	
getAllPaths	3
getClassifications	4
getFinalPathList	4
getGroupFromPath	5
getPathList	5
naybeNode-class	5
missingSnps	6
node-class	6
node_a	6
node_a2	7
node_b2	7

25

Index

node b4bde	. 7
node b5	. 8
10de_c	
node d	
node_d1	
10de_d2	
10de_d2	
10de_d4	. 9 . 10
	. 10
node_h2	
node_h2a	. 10
node_h2a2a	. 11
node_hv	. 11
node_{-} i	. 11
$node_{-\mathbf{j}}$. 12
node_jt	. 12
node_jt_alt	. 12
node_k	. 13
node_k1	. 13
node_k_alt	. 13
node_10	. 14
node_11	. 14
node_l1b	. 14
node_12	. 15
node_123456	. 15
node_12346	. 15
node_13	. 16
node_134	. 16
node_m	. 16
	. 17
node nlal	
node nlalb	
node_n_alt	
node_r	
node_r_alt	
node_t	
node t1	
node_u	
node_u8b	. 20
node_u8b_alt	
node_u_alt	. 20
node v	
	. 21
-	. 21
node_x	
node_x2	. 22
numChildren	. 22
numReqs	. 22
numSnps	. 23
root	
validData	. 24
validPath	. 24

generate_snp_data 3

```
generate_snp_data SNP Data Generator
```

Description

Takes in a plink map file and a plink ped file and creates a dataframe with headers

Usage

```
generate_snp_data(map_file, ped_file)
```

Arguments

map_file The name of the .map file. Should be tab delimited with no header.

ped_file The name of the .ped file. Should be space delimited with no header.

Examples

```
mapfile <- system.file("extdata","HapMap_Mito_Seq_QC_2.map",package="HiMC")
pedfile <- system.file("extdata","HapMap_Mito_Seq_QC_2.ped",package="HiMC")
generate_snp_data(mapfile,pedfile)</pre>
```

getAllPaths

Path Generator

Description

Internal function. Takes a SNP dataframe, a node, and current path, and returns all available paths.

Usage

```
getAllPaths(df, node, path)
```

Arguments

df SNP dataframe
node Node to be checked
path Current path checked

Examples

```
#internal function, not meant to be called externally
mapfile <- system.file("extdata", "HapMap_Mito_Seq_QC_2.map", package="HiMC")
pedfile <- system.file("extdata", "HapMap_Mito_Seq_QC_2.ped", package="HiMC")
snpfile <- generate_snp_data(mapfile, pedfile)
full_path_list <- getAllPaths(snpfile, HiMC::root, "")</pre>
```

4 getFinalPathList

```
getClassifications getClassifications output generator
```

Description

Takes in a dataframe generated by "generate_snp_data" and returns each subject classified with full classification paths

Usage

```
getClassifications(source_df)
```

Arguments

```
source_df The snp_data generated dataframe
```

Examples

```
mapfile <- system.file("extdata", "HapMap_Mito_Seq_QC_2.map", package="HiMC")
pedfile <- system.file("extdata", "HapMap_Mito_Seq_QC_2.ped", package="HiMC")
snpfile <- generate_snp_data(mapfile, pedfile)
classifications <- getClassifications(snpfile)</pre>
```

```
getFinalPathList getFinalPathList internal function
```

Description

Takes in a nested list of prettified paths and returns those of the greatest length

Usage

```
getFinalPathList(plist)
```

Arguments

```
plist A nested list of prettified paths
```

Examples

```
#internal function, not meant to be called externally
filtered_paths <- getFinalPathList(c("A -> B","A -> B -> C"))
```

getGroupFromPath 5

```
getGroupFromPath getFinalPathList internal function
```

Description

Takes in a list of paths and returns the final classification of that path

Usage

```
getGroupFromPath(string)
```

Arguments

string

The path in question

Examples

```
#internal function, not meant to be called externally group <- getGroupFromPath("A \rightarrow B \rightarrow C")
```

getPathList

getPathList internal function

Description

Takes in a nested list of paths in dataframe format and returns a top-level path assignment

Usage

```
getPathList(df)
```

Arguments

df

The FASTMAP dataframe row

Examples

```
#internal function, not meant to be called externally
paths <- getPathList(data.frame())</pre>
```

maybeNode-class

A class to represent a maybe-node.

Description

Union between classes NULL and node

6 node_a

missingSnps

validData boolean check

Description

Internal function. Takes in a dataframe and a node and returns true if the dataframe row is missing any SNPs for the node in question

Usage

```
missingSnps(df, node)
```

Arguments

df The FASTMAP dataframe row

node The node in question

Examples

```
#internal function, not meant to be called externally
is_dataframe_row_missing_snps_from_node <- missingSnps(data.frame(),HiMC::root)</pre>
```

node-class

A class to represent a node.

Description

A class to represent a node.

Slots

```
name A character vector snps A list req A list children A list
```

node_a

A node node_a.

Description

An instance of the node class.

Usage

node_a

Format

node_a2 7

node_a2

A node node_a2.

Description

An instance of the node class.

Usage

node_a2

Format

An object of class node of length 1.

node_b2

A node node_b2.

Description

An instance of the node class.

Usage

node_b2

Format

An object of class node of length 1.

node_b4bde

A node node_b4bde.

Description

An instance of the node class.

Usage

node_b4bde

Format

node_d

node_b5

A node node_b5.

Description

An instance of the node class.

Usage

node_b5

Format

An object of class node of length 1.

node_c

A node node_c.

Description

An instance of the node class.

Usage

node_c

Format

An object of class node of length 1.

node_d

A node node_d.

Description

An instance of the node class.

Usage

node_d

Format

node_d1

node_d1

 $A \ node \ node_d1.$

Description

An instance of the node class.

Usage

node_d1

Format

An object of class node of length 1.

node_d2

A node node_d2.

Description

An instance of the node class.

Usage

node_d2

Format

An object of class node of length 1.

node_d4

A node node_d4.

Description

An instance of the node class.

Usage

node_d4

Format

node_h2a

node_h

 $A \ node \ node_h.$

Description

An instance of the node class.

Usage

node_h

Format

An object of class node of length 1.

node_h2

A node node_h2.

Description

An instance of the node class.

Usage

node_h2

Format

An object of class node of length 1.

node_h2a

A node node_h2a.

Description

An instance of the node class.

Usage

node_h2a

Format

node_h2a2a 11

node_h2a2a

A node node_h2a2a.

Description

An instance of the node class.

Usage

```
node_h2a2a
```

Format

An object of class node of length 1.

node_hv

A node node_hv.

Description

An instance of the node class.

Usage

```
node_hv
```

Format

An object of class node of length 1.

node_i

A node node_i.

Description

An instance of the node class.

Usage

node_i

Format

node_jt_alt

node_j

A node node_j.

Description

An instance of the node class.

Usage

node_j

Format

An object of class node of length 1.

node_jt

A node node_jt.

Description

An instance of the node class.

Usage

node_jt

Format

An object of class node of length 1.

node_jt_alt

A node node_jt_alt.

Description

An instance of the node class.

Usage

node_jt_alt

Format

node_k

node_k

A node node_k.

Description

An instance of the node class.

Usage

node_k

Format

An object of class node of length 1.

node_k1

A node node_k1.

Description

An instance of the node class.

Usage

node_k1

Format

An object of class node of length 1.

node_k_alt

A node node_k_alt.

Description

An instance of the node class.

Usage

node_k_alt

Format

14 node_11b

node_10

 $A\ node\ node_l0.$

Description

An instance of the node class.

Usage

```
node_10
```

Format

An object of class node of length 1.

node_11

A node node_l1.

Description

An instance of the node class.

Usage

```
node_l1
```

Format

An object of class node of length 1.

node_l1b

A node node_l1b.

Description

An instance of the node class.

Usage

```
node_l1b
```

Format

node_12

node_12

A node node_l2.

Description

An instance of the node class.

Usage

node_12

Format

An object of class node of length 1.

node_123456

A node node_l23456.

Description

An instance of the node class.

Usage

node_123456

Format

An object of class node of length 1.

node_12346

A node node_l2346.

Description

An instance of the node class.

Usage

node_12346

Format

node_m

node_13

A node node_l3.

Description

An instance of the node class.

Usage

node_13

Format

An object of class node of length 1.

node_134

A node node_l34.

Description

An instance of the node class.

Usage

node_134

Format

An object of class node of length 1.

node_m

A node node_m.

Description

An instance of the node class.

Usage

node_m

Format

node_n

node_n

 $A \ node \ node_n.$

Description

An instance of the node class.

Usage

node_n

Format

An object of class node of length 1.

node_n1a1

A node node_n1a1.

Description

An instance of the node class.

Usage

node_n1a1

Format

An object of class node of length 1.

node_n1a1b

 $A node node_n1a1b.$

Description

An instance of the node class.

Usage

node_n1a1b

Format

node_r_alt

node_n_alt

 $A \ node \ node_n_alt.$

Description

An instance of the node class.

Usage

```
node_n_alt
```

Format

An object of class node of length 1.

node_r

A node node_r.

Description

An instance of the node class.

Usage

```
node_r
```

Format

An object of class node of length 1.

node_r_alt

A node node_r_alt.

Description

An instance of the node class.

Usage

```
node_r_alt
```

Format

node_t

node_t

 $A \ node \ node_t.$

Description

An instance of the node class.

Usage

node_t

Format

An object of class node of length 1.

node_t1

A node node_t1.

Description

An instance of the node class.

Usage

node_t1

Format

An object of class node of length 1.

node_u

A node node_u.

Description

An instance of the node class.

Usage

node_u

Format

node_u_alt

node_u8b

 $A\ node\ node_u8b.$

Description

An instance of the node class.

Usage

```
node_u8b
```

Format

An object of class node of length 1.

node_u8b_alt

A node node_u8b_alt.

Description

An instance of the node class.

Usage

```
node_u8b_alt
```

Format

An object of class node of length 1.

node_u_alt

A node node_u_alt.

Description

An instance of the node class.

Usage

Format

node_v 21

node_v

A node node_v.

Description

An instance of the node class.

Usage

node_v

Format

An object of class node of length 1.

node_w

A node node_w.

Description

An instance of the node class.

Usage

node_w

Format

An object of class node of length 1.

node_x

A node node_x.

Description

An instance of the node class.

Usage

node_x

Format

22 numReqs

node_x2

A node node_x2.

Description

An instance of the node class.

Usage

```
node_x2
```

Format

An object of class node of length 1.

numChildren

NumChildren

Description

Internal function. Takes in a node object and returns the total number of that node's children

Usage

```
numChildren(node_object)
```

Arguments

Examples

```
#internal function, not meant to be called externally
node_tail_length = numChildren(HiMC::root)
```

numReqs

NumReqs

Description

Internal function. Takes in a node object and returns the number of SNPs that it requires for valida-

Usage

```
numReqs(node_object)
```

numSnps 23

Arguments

```
node_object The FASTMAP node in question
```

Examples

```
#internal function, not meant to be called externally
number_of_required_snps <- numReqs(HiMC::root)</pre>
```

numSnps

NumSnps

Description

Internal function. Takes in a node object and returns the number of SNPs that belong to it

Usage

```
numSnps(node_object)
```

Arguments

Examples

```
#internal function, not meant to be called externally
amount_of_snps_in_node <- numSnps(HiMC::root)</pre>
```

root

A node root.

Description

An instance of the node class.

Usage

root

Format

24 validPath

validData

validData boolean check

Description

Internal function. Takes in a dataframe and a node and returns true if the dataframe row has the required SNPs for the node

Usage

```
validData(df, node)
```

Arguments

df The FASTMAP dataframe row

node The node in question

Examples

```
#internal function, not meant to be called externally
df_row_meets_criteria_for_node_requirements <- validData(data.frame(), HiMC::root)</pre>
```

validPath

validPath boolean check

Description

Internal function. Takes in a dataframe and a node and returns true if the dataframe row represents a valid path for the node

Usage

```
validPath(df, node)
```

Arguments

df The FASTMAP dataframe row

node The node in question

Examples

```
#internal function, not meant to be called externally
df_row_meets_criteria_for_node_path <- validPath(data.frame(), HiMC::root)</pre>
```

Index

*Topic children	node_t, 19
numChildren, 22	node_t1, 19
*Topic classification	node_u, 19
getClassifications,4	node_u8b, 20
getGroupFromPath, 5	node_u8b_alt, 20
*Topic datasets	node_u_alt, 20
node_a, 6	node_v, 21
node_a2, 7	node_w, 21
node_b2,7	node_x, 21
node_b4bde, 7	node_x2, 22
node_b5,8	root, 23
node_c, 8	*Topic data
node_d, 8	validData, 24
node_d1,9	*Topic df
node_d2,9	missingSnps, 6
node_d4,9	validData,24
node_h, 10	validPath, 24
node_h2, 10	*Topic haplogroup
node_h2a, 10	getClassifications,4
node_h2a2a, 11	${ t getGroupFromPath, 5}$
node_hv, 11	*Topic list
node_i, 11	getClassifications,4
node_j, 12	getFinalPathList,4
node_jt, 12	${ t getGroupFromPath,5}$
node_jt_alt, <mark>12</mark>	getPathList, 5
node_k, 13	*Topic map
node_k1, 13	generate_snp_data,3
node_k_alt, 13	*Topic missing
node_10, 14	missingSnps, 6
node_11, 14	*Topic node
node_11b, <mark>14</mark>	getAllPaths, 3
node_12, 15	missingSnps, 6
node_123456, <mark>15</mark>	numChildren, 22
node_12346, 15	numReqs, 22
node_13, 16	numSnps, 23
node_134, <mark>16</mark>	validData,24
node_m, 16	validPath, 24
node_n, 17	*Topic path
node_n1a1,17	getAllPaths, 3
node_n1a1b, 17	getClassifications, 4
node_n_alt, 18	getFinalPathList,4
node_r, 18	getGroupFromPath,5
node_r_alt, 18	getPathList,5

26 INDEX

and Markle 24	
validPath, 24	node_11, 14
*Topic ped	node_11b, 14 node_12, 15
<pre>generate_snp_data, 3 *Topic prettify</pre>	node_123456, 15
getPathList, 5	node_12346, 15
*Topic snp	node_13, 16
generate_snp_data, 3	node_134, 16
missingSnps, 6	node_m, 16
numChildren, 22	node_n, 17
numRegs, 22	node_n1a1, 17
numSnps, 23	node_n1a1b, 17
validData, 24	node_n_alt, 18
validPath, 24	node_r, 18
*Topic valid	node_r_alt, 18
validData, 24	node_t, 19
validPath, 24	node_t1, 19
	node_u, 19
<pre>generate_snp_data, 3</pre>	node_u8b, 20
getAllPaths, 3	node_u8b_alt, 20
getClassifications, 4	node_u_alt, 20
<pre>getFinalPathList,4</pre>	node_v, 21
<pre>getGroupFromPath,5</pre>	node_w, 21
getPathList,5	node_x, 21
	node_x2, 22
<pre>maybeNode (maybeNode-class), 5</pre>	numChildren, 22
<pre>maybeNode-class,5</pre>	numReqs, 22
missingSnps,6	numSnps, 23
-	_
node (node-class), 6	numSnps, 23
node (node-class), 6 node-class, 6	root, 23
node (node-class), 6 node-class, 6 node_a, 6	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7	root, 23
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_d4, 9 node_h, 10	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_d4, 9 node_h, 10 node_h2, 10	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_d4, 9 node_h2, 10 node_h2, 10 node_h2a, 10	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_hv, 11 node_i, 11	root, 23 validData, 24
node (node-class), 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_h, 11 node_i, 11 node_j, 12	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_hy, 11 node_j, 12 node_jt, 12	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_h2, 10 node_h2, 10 node_h2a2a, 11 node_hy, 11 node_i, 11 node_j, 12 node_jt_alt, 12	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_i, 11 node_i, 11 node_j, 12 node_jt_alt, 12 node_k, 13	root, 23 validData, 24
node (node-class), 6 node_class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d, 9 node_d2, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_h, 11 node_j, 12 node_jt_alt, 12 node_jt_alt, 12 node_k, 13 node_k1, 13	root, 23 validData, 24
node (node-class), 6 node-class, 6 node_a, 6 node_a2, 7 node_b2, 7 node_b4bde, 7 node_b5, 8 node_c, 8 node_d, 8 node_d1, 9 node_d2, 9 node_d4, 9 node_h, 10 node_h2, 10 node_h2a, 10 node_h2a2a, 11 node_i, 11 node_i, 11 node_j, 12 node_jt_alt, 12 node_k, 13	root, 23 validData, 24