# Package ‘CovidMutations’

## September 18, 2020

**Type** Package  
**Title** Mutation Analysis Toolkit for COVID-19 (Coronavirus Disease 2019)  
**Version** 0.1.3  
**Date** 2020-08-29  
**Depends** R (>= 3.6)  
**License** Artistic-2.0  
**Encoding** UTF-8  
**LazyData** true  
**Imports** ggplot2, cowplot, seqinr, stringr, grDevices, graphics, utils, ggpubr, dplyr, VennDiagram  
**RoxygenNote** 7.1.1  
**NeedsCompilation** no  
**URL** https://github.com/MSQ-123/CovidMutations  
**BugReports** https://github.com/MSQ-123/CovidMutations/issues  
**Suggests** testthat  
**Author** Shaoqian Ma [aut, cre] (<https://orcid.org/0000-0001-8950-0711>), Yongyou Zhang [aut] (<https://orcid.org/0000-0003-2413-9106>)  
**Maintainer** Shaoqian Ma <shaoqianma@qq.com>  
**Repository** CRAN  
**Date/Publication** 2020-09-18 12:00:39 UTC
R topics documented:

AssayMutRatio . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
assays . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
chinalist . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
covid_annot . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
doubleAssay . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
gene_position . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
gff3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
globalProteinMut . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
globalSNPprofile . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
indelSNP . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
LastfiveNrMutation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
mergeEvents . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
MutByGene . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
mutStat . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
nucmer . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
nucmerr . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
nucmerRMD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
plotMutAnno . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
plotMutProteins . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
refseq . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

Index 18

AssayMutRatio Calculate the mutation detection rate using different assays

Description

This function is to use the well established assays information to detect mutations in different SARS-CoV-2 genomic sites. The output will be series of figures presenting the mutation profile using a specific assay and a figure for comparison between the mutation detection rate in each primers binding region.

Usage

AssayMutRatio(
  nucmerr = nucmerr,
  assays = assays,
  totalsample = totalsample,
  plotType = "barplot",
  outdir = NULL
)

Arguments

nucmerr  Mutation information containing group list (derived from "nucmer" object using "nucmerRMD" function).
assays   Assays dataframe including the detection ranges of mutations.
totalsample  Total sample number, total cleared GISAID fasta data.
plotType  Figure type for either "barplot" or "logtrans".
outdir  The output directory.

Value

Plot the selected figure type as output.

Examples

data("nucmerr")
data("assays")
Total <- 52  ## Total Cleared GISAID fasta data, sekitseq
#outdir <- tempdir()
#Output the results
AssayMutRatio(nucmerr = nucmerr,
assays = assays,
totalsample = Total,
plotType = "logtrans",
outdir = NULL)

assays  

Assays for mutation detection using different primers and probes

Description

These assays include the primer detection ranges in which mutations may occur.

Usage

data(assays)

Format

A dataframe with 10 rows and 7 columns.

References


Examples

data(assays)
chinalist  

_A list of places in China_

**Description**

The list is used for displacing some original cities’ names with “China” in order to make the downstream analysis easier.

**Usage**

data(chinalist)

**Format**

A dataframe with 31 rows and 1 column.

**Source**

This data is created by Zhanglab in Xiamen University.

**Examples**

data(chinalist)

covid_annot  

_Mutation annotation results produced by "indelSNP" function_

**Description**

A dataframe which could be used for downstream analysis like mutation statistics description.

**Usage**

data(covid_annot)

**Format**

A dataframe with 394 rows and 10 columns.

**Source**

[https://www.gisaid.org/](https://www.gisaid.org/)

**Examples**

data(covid_annot)
doubleAssay

Detection of co-occurring mutations using double-assay information

Description

The detection of SARS-CoV-2 is important for the prevention of the outbreak and management of patients. Real-time reverse-transcription polymerase chain reaction (RT-PCR) assay is one of the most effective molecular diagnosis strategies to detect virus in clinical laboratory. It will be more accurate and practical to use double assays to detect some samples with co-occurring mutations.

Usage

doubleAssay(nucmerr = nucmerr, assay1 = assay1, assay2 = assay2, outdir = NULL)

Arguments

nucmerr Mutation information containing group list(derived from "nucmer" object using "nucmerRMD" function). 
assay1 Information of the first assay(containing primers locations and probe location, see the format of assays provided as example data. e.g. data(assays); assay1< assays[1,])
assay2 Information of the second assay, the format is the same as the first assay.
outdir The output directory. If NULL print the plot in Rstudio.

Value

Plot three figures in a single panel, including two results of assays and a "venn" plot for co-occurring mutated samples.

Examples

data("nucmerr")
data("assays")
assay1 <- assays[1,]
assay2 <- assays[2,]
#outdir <- tempdir()
doubleAssay(nucmerr = nucmerr,
                        assay1 = assay1,
                        assay2 = assay2,
                        outdir = NULL)
**gene_position**  
"GFF3" format gene position data for SARS-Cov-2

**Description**  
This "GFF3" data is used for counting the mutations in each gene in virus sample.

**Usage**  
data(gene_position)

**Format**  
A dataframe with 26 rows and 10 columns.

**Source**  

**Examples**  
data(gene_position)

---

**gff3**  
"GFF3" format annotation data for SARS-Cov-2

**Description**  
This "GFF3" data is used for annotating the effects of mutations in virus sample.

**Usage**  
data(gff3)

**Format**  
A dataframe with 26 rows and 10 columns.

**Source**  
https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2697049

**Examples**  
data(gff3)
globalProteinMut

**Global mutational events profiling of proteins**

**Description**

This function is to visualize the global protein mutational pattern in the SARS-CoV-2 genome.

**Usage**

```r
globalProteinMut(
  covid_annot = covid_annot,
  outdir = NULL,
  figure_Type = "heatmap",
  top = 10,
  country = "global"
)
```

**Arguments**

- `covid_annot`: The mutation effects provided by "indelSNP" function.
- `outdir`: The output directory.
- `figure_Type`: Figure type for either "heatmap" or "count".
- `top`: The number of variants to plot.
- `country`: Choose a country to plot the mutational pattern or choose "global" to profile mutations across all countries. The default is "global".

**Value**

Plot the selected figure type as output.

**Examples**

```r
data("covid_annot")
outdir <- tempdir()
# make sure the covid_annot is a dataframe
 covid_annot <- as.data.frame(covid_annot)
globalProteinMut(covid_annot = covid_annot,
  outdir = outdir,
  figure_Type = "heatmap",
  top = 10,
  country = "USA")
```
**globalSNPprofile**

*Global single nucleotide polymorphism (SNP) profiling in virus genome*

**Description**

This function is to visualize the global SNP pattern in the SARS-CoV-2 genome.

**Usage**

```r
globalSNPprofile(
  nucmerr = nucmerr,
  outdir = NULL,
  figure_Type = "heatmap",
  country = "global",
  top = 5
)
```

**Arguments**

- **nucmerr**: Mutation information containing group list (derived from "nucmer" object using "nucmerRMD" function).
- **outdir**: The output directory.
- **figure_Type**: Figure type for either "heatmap" or "count".
- **country**: Choose a country to plot the mutational pattern or choose "global" to profile mutations across all countries. The default is "global".
- **top**: The number of mutational classes to plot.

**Value**

Plot the selected figure type as output.

**Examples**

```r
data("nucmerr")
outdir <- tempdir()
globalSNPprofile(nucmerr = nucmerr,
  outdir = outdir,
  figure_Type = "heatmap",
  country = "global",
  top = 5)
```
**Description**

This function is to annotate the mutational events and indicate their potential effects on the proteins. Mutational events include SNP, insertion and deletion.

**Usage**

```r
indelSNP(
  nucmer = nucmer,
  saveRda = FALSE,
  refseq = refseq,
  gff3 = gff3,
  annot = annot,
  outdir = NULL
)
```

**Arguments**

- `nucmer`: An object called "nucmer", mutation information derived from "nucmer.snp" variant file by "seqkit" software and "nucmer SNP-calling" scripts. To be processed by "indelSNP" function, the nucmer object should be first transformed by "mergeEvents" function.
- `saveRda`: Whether to save the results as ".rda" file.
- `refseq`: SARS-Cov-2 genomic reference sequence.
- `gff3`: "GFF3" format annotation data for SARS-Cov-2.
- `annot`: Annotation of genes(corresponding proteins) list from "GFF3" file by "setNames(gff3[,10],gff3[,9])".
- `outdir`: The output directory.

**Value**

Write the result as ".csv" file to the specified directory.

**Examples**

```r
data("nucmer")
# Fix IUPAC codes
nucmer<-nucmer[!nucmer$qvar%in%c("B","D","H","K","M","N","R","S","V","W","Y"),]
nucmer<- mergeEvents(nucmer = nucmer)## This will update the nucmer object
data("refseq")
data("gff3")
annot <- setNames(gff3[,10],gff3[,9])
#outdir <- tempdir()
```
nucmer<- indelSNP(nucmer = nucmer,
saveRda = FALSE,
refseq = refseq,
gff3 = gff3,
annot = annot,
outdir = NULL)

**LastfiveNrMutation**  
*Batch assay analysis for last five Nr of primers*

**Description**
Last five nucleotides of primer mutation count/type for any reverse transcription polymerase chain reaction (RT-PCR) primer.

**Usage**

```r
LastfiveNrMutation(
  nucmerr = nucmerr,
  assays = assays,
  totalsample = totalsample,
  figurelist = FALSE,
  outdir = NULL
)
```

**Arguments**

- `nucmerr`: Mutation information containing group list(derived from "nucmer" object using "nucmerRMD" function).
- `assays`: Assays dataframe including the detection ranges of mutations.
- `totalsample`: Total sample number, total cleared GISAID fasta data.
- `figurelist`: Whether to output the integrated plot list for each assay.
- `outdir`: The output directory. If `figurelist = TRUE`, output the figure in the R session.

**Value**
Plot the mutation counts(last five nucleotides for each primer) for each assay as output.

**Examples**

```r
data("nucmerr")
data("assays")
totalsample <- 434
#outdir <- tempdir()
LastfiveNrMutation(nucmerr = nucmerr,
  assays = assays,
  totalsample = totalsample,
  figurelist = FALSE,
  outdir = NULL)
```
mergeEvents

Merge neighboring events of single nucleotide polymorphism (SNP), insertion and deletion.

Description

The first step for handling the nucmer object, then effects of mutations can be analysed using "in-delSNP" function.

Usage

mergeEvents(nucmer = nucmer)

Arguments

nucmer

An object called "nucmer", mutation information derived from "nucmer.snp" variant file by "seqkit" software and "nucmer SNP-calling" scripts.

Value

An updated "nucmer" object.

Examples

#The example data:
data("nucmer")
#options(stringsAsFactors = FALSE)

#The input nucmer object can be made by the comment below:
#nucmer<-read.delim("nucmer.snps",as.is=TRUE,skip=4,header=FALSE)
#colnames(nucmer)<-c("rpos","rvar","qvar","qpos","","","","","rlength","qlength","","","rname","qname")
#rownames(nucmer)<-paste0("var",1:nrow(nucmer))

# Fix IUPAC codes
nucmer<-nucmer[!nucmer$qvar%in%c("B","D","H","K","M","N","R","S","V","W","Y"),]
nucmer<- mergeEvents(nucmer = nucmer)## This will update the nucmer object

MutByGene

Plot mutation counts for certain genes

Description

After annotating the mutations, this function is to plot the counts of mutational events for each gene in the SARS-CoV-2 genome.
Usage

MutByGene(nucmerr = nucmerr, gff3 = gff3, figurelist = FALSE, outdir = NULL)

Arguments

nucmerr  Mutation information containing group list(derived from "nucmer" object using "nucmerRMD" function).
gff3 "GFF3" format gene position data for SARS-Cov-2(the "GFF3" file should include columns named: "Gene", "Start", "Stop").
figurelist Whether to output the integrated plot list for each gene.
outdir The output directory, if the figurelist = TRUE, output the figure in the R session.

Value

Plot the mutation counts figure for each gene as output.

Examples

data("nucmerr")
data("gene_position")
outdir <- tempdir()  
MutByGene(nucmerr = nucmerr, gff3 = gene_position, figurelist = FALSE, outdir = NULL)  
#if figurelist = TRUE, the recommendation for figure display(in pixel) is: width=1650, height=1300

mutStat

Plot mutation statistics for nucleotide

Description

Visualization for the top mutated samples, average mutational counts, top mutated position in the genome, mutational density across the genome and distribution of mutations across countries.

Usage

mutStat(
  nucmerr = nucmerr,
  outdir = NULL,
  figure_Type = "TopMuSample",
  type_top = 10,
  country = FALSE,
  mutpos = NULL
)
**nucmer**

**Arguments**

- **nucmerr**: Mutation information containing group list (derived from "nucmer" object using "nucmerRMD" function).
- **outdir**: The output directory.
- **figure_Type**: Figure type for: "TopMuSample", "AverageMu", "TopMuPos", "MutDens", "CountryMutCount", "TopCountryMut".
- **type_top**: To plot the figure involving "top n"("TopMuSample", "TopMuPos", "TopCountryMut"), the "type_top" should specify the number of objects to display.
- **country**: To plot the figure using country as groups("CountryMutCount" and "TopCountryMut"), the "country" should be TRUE.
- **mutpos**: If the figure type is "TopCountryMut", "mutpos" can specify a range of genomic position (eg. 28831:28931) for plot

**Value**

Plot the selected figure type as output.

**Examples**

```r
data("nucmerr")
outdir <- tempdir()
mutStat(nucmerr = nucmerr,
    outdir = outdir,
    figure_Type = "TopCountryMut",
    type_top = 10,
    country = FALSE,
    mutpos = NULL)
```

---

**nucmer**

*Mutation information derived from "nucmer" SNP analysis*

**Description**

The "nucmer.snps" variant file is obtained by processing the SARS-Cov-2 sequence from Gisaid website (complete, high coverage only, low coverage exclusion, Host=human, Virus name = hCoV-19) with "seqkit" software and "nucmer" scripts. The example data is downsampled from complete data in 2020-07-28 (0.001 proportion, 52 samples).

**Usage**

```r
data(nucmer)
```

**Format**

A dataframe with 437 rows (mutation sites) and 14 columns.
Source
https://www.gisaid.org/

Examples
data(nucmer)

<table>
<thead>
<tr>
<th>nucmerr</th>
<th>Preprocessed &quot;nucmer.snps&quot; file using &quot;nucmerRMD&quot; function</th>
</tr>
</thead>
</table>

Description
A dataset contains some group information subtracted from the "nucmer" object by "nucmerRMD" function in order to best describe the results.

Usage
data(nucmerr)

Format
A dataframe with 437 rows (downsampled mutation sites) and 10 columns.

Source
https://www.gisaid.org/

Examples
data(nucmerr)

<table>
<thead>
<tr>
<th>nucmerRMD</th>
<th>Preprocess &quot;nucmer&quot; object to add group information</th>
</tr>
</thead>
</table>

Description
Manipulate the "nucmer" object to make the analysis easier.

Usage
nucmerRMD(nucmer = nucmer, outdir = NULL, chinalist = chinalist)
plotMutAnno

Arguments

nucmer An object called "nucmer", mutation information derived from "nucmer.snp" variant file by "seqkit" software and "nucmer SNP-calling" scripts.
outdir The output directory.
chinalist A list of places in China, for displacing some original cities with "China" in order to make the downstream analysis easier.

Value

Saving the updated "nucmer" object.

Examples

data("nucmer")
data("chinalist")
#outdir <- tempdir()
nucmerr<- nucmerRMD(nucmer = nucmer, outdir = NULL, chinalist = chinalist)

plotMutAnno

Plot the mutation statistics after annotating the "nucmer" object by "indelSNP" function

Description

Basic descriptions for the mutational events.

Usage

plotMutAnno(covid_annot = covid_annot, figureType = "MostMut", outdir = NULL)

Arguments

covid_annot The mutation effects provided by "indelSNP" function.
figureType Figure type for: "MostMut", "MutPerSample", "VarClasses", "VarType", "NucleoEvents", "ProEvents".
outdir The output directory.

Value

Plot the selected figure type as output.

Examples

data("covid_annot")
# make sure the covid_annot is a dataframe
covid_annot <- as.data.frame(covid_annot)
#outdir <- tempdir() specify your output directory
plotMutAnno(covid_annot = covid_annot, figureType = "MostMut", outdir = NULL)
`plotMutProteins`  
*Plot the most frequent mutational events for proteins in the SARS-CoV-2 genome*

**Description**

Plot the most frequent mutational events for proteins selected. The protein name should be specified correctly (only for SARS-CoV-2).

**Usage**

```r
plotMutProteins(
  covid_annot = covid_annot,
  proteinName = "NSP2",
  top = 20,
  outdir = NULL
)
```

**Arguments**

- `covid_annot`  
The mutation effects provided by "indelSNP" function.
- `proteinName`  
Proteins in the SARS-CoV-2 genome, available choices: 5'UTR, NSP1~NSP10, NSP12a, NSP12b, NSP13, NSP14, NSP15, NSP16, S, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8, N, ORF10.
- `top`  
The number of objects to display.
- `outdir`  
The output directory.

**Value**

Plot the mutational events for selected proteins as output.

**Examples**

```r
data("covid_annot")
# make sure the covid_annot is a dataframe
covid_annot <- as.data.frame(covid_annot)
#outdir <- tempdir() specify your output directory
plotMutProteins(covid_annot = covid_annot,proteinName = "NSP2", top = 20, outdir = NULL)
```
refseq

SARS-Cov-2 genomic reference sequence from NCBI

Description
This reference sequence is derived from "fasta" file, preprocessed by "read.fasta" function(refseq<-read.fasta("NC_045512.2.fa",forceDNAtolower=FALSE)[[1]]). It is used for annotating mutations in virus samples.

Usage
data(refseq)

Format
"SeqFastadna" characters.

Source

Examples
data(refseq)
Index

* datasets
  assays, 3
  chinalist, 4
  covid_annot, 4
  gene_position, 6
  gff3, 6
  nucmer, 13
  nucmerr, 14
  refseq, 17

AssayMutRatio, 2
assays, 3

chinalist, 4
covid_annot, 4

doubleAssay, 5
gene_position, 6
gff3, 6
globalProteinMut, 7
globalSNPprofile, 8

indelSNP, 9
LastfiveNrMutation, 10
mergeEvents, 11
MutByGene, 11
mutStat, 12

nucmer, 13
nucmerr, 14
nucmerRMD, 14

plotMutAnno, 15
plotMutProteins, 16

refseq, 17