

Package ‘ggmsa’

October 7, 2019

Title Plot Multiple Sequence Alignment using 'ggplot2'

Version 0.0.1

Description Supports visualizing multiple sequence alignment of DNA and protein sequences using 'ggplot2'. It supports a number of colour schemes, including Chemistry, Clustal, Shapely, Taylor and Zappo. Multiple sequence alignment can easily be combined with other 'ggplot2' plots, such as aligning a phylogenetic tree produced by 'ggtree' with multiple sequence alignment.

Depends R (>= 3.5.0)

Imports ggplot2, ggseqlogo, magrittr, tidyr, treeio, utils

Suggests Biostrings

License Artistic-2.0

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Author Guangchuang Yu [aut, cre] (<<https://orcid.org/0000-0002-6485-8781>>),
Lang Zhou [aut],
Huina Huang [ctb]

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Repository CRAN

Date/Publication 2019-10-07 17:50:02 UTC

R topics documented:

color_Clustal	2
geom_msa	2
ggmsa	3
msa_data	4

Index	5
--------------	----------

color_Clustal	<i>A color scheme of Clustal.This function is a algorithm to assign colors for Multiple Sequence.</i>
---------------	---

Description

A color scheme of Clustal.This function is a algorithm to assign colors for Multiple Sequence.

Usage

```
color_Clustal(y)
```

Arguments

y	A data frame, data of a tidy fasta,created by 'tidy_fasta()'.
---	---

geom_msa	<i>geom_msa</i>
----------	-----------------

Description

Multiple sequence alignment layer for ggplot2. It creates background frames with/without sequence labels.

Usage

```
geom_msa(data, ...)
```

Arguments

data	A data frame generated by mas_data()
...	additional parameter

Value

A list

Author(s)

guangchuang yu

`ggmsa`*ggmsa*

Description

Plot multiple sequence alignment using ggplot2 with multiple color schemes supported.

Usage

```
ggmsa(fasta, start = NULL, end = NULL, font = "helvetica_regular",  
      color = "Clustal")
```

Arguments

<code>fasta</code>	Aligned FASTA format file for representing either nucleotide sequences or peptide sequences.
<code>start</code>	Start position to plot, If font=NULL, only the background frame is drawn, and no character.
<code>end</code>	End position to plot, If font=NULL, only the background frame is drawn, and no character.
<code>font</code>	Character font, Defaults is 'helvetica_regular'.
<code>color</code>	A Color scheme. One of 'Clustal', 'Chemistry_AA', 'Shapely_AA', 'Zappo_AA', 'Taylor_AA', 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Clustal'.

Value

ggplot object

Author(s)

guangchuang yu

Examples

```
#plot multiple sequence alignment  
f <- system.file("extdata/sample.fasta", package="ggmsa")  
ggmsa(f, 164, 213, color="Chemistry_AA")
```

`msa_data`*msa_data*

Description

This function parses FASTA file and convert it to a tidy data frame. The function will also assign color to each molecule (amino acid or nucleotide) according to the selected color scheme. Sequence logo data for drawing alignment label will also be added if font != NULL. The output of `msa_data()` is the input of `geom_msa()`.

Usage

```
msa_data(fasta, start = NULL, end = NULL, font = "helvetica_regular",
         color = "Clustal")
```

Arguments

<code>fasta</code>	Aligned fasta file.
<code>start</code>	Start position to plot, Defaults = NULL.
<code>end</code>	End position to plot, Defaults = NULL.
<code>font</code>	Character font, Defaults is 'helvetica_regular'.
<code>color</code>	A Color scheme. One of 'Clustal', 'Chemistry_AA', 'Shapely_AA', 'Zappo_AA', 'Taylor_AA', 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Clustal'.

Value

A data frame

Author(s)

guangchuang yu

Examples

```
fasta <- system.file("extdata/sample.fasta", package="ggmsa")
data <- msa_data(fasta, 20, 120, font = 'helvetica_regular', color = 'Chemistry_AA' )
```

Index

*Topic **clustal**

color_Clustal, 2

color_Clustal, 2

geom_msa, 2

ggmsa, 3

msa_data, 4