Package 'ggmsa'

May 15, 2020

Title Plot Multiple Sequence Alignment using 'ggplot2'

Version 0.0.4

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 Biostrings, ggplot2, magrittr, tidyr, utils, stats, stringr

Suggests ape, cowplot, ggtree (>= 1.17.1), knitr, methods, seqmagick

License Artistic-2.0

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

VignetteBuilder knitr

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>

R topics documented:

---

available_colors   *List Color Schemes currently available*

Description

This function lists color schemes currently available that can be used by 'ggmsa'

Usage

```
available_colors()
```

Value

A character vector of available color schemes

Author(s)

Lang Zhou

Examples

```
available_colors()
```

---

available_fonts   *List Font Families currently available*

Description

This function lists font families currently available that can be used by 'ggmsa'

Usage

```
available_fonts()
```

Value

A character vector of available font family names

Author(s)

Lang Zhou

Examples

```
available_fonts()
```

---

available_msa	*List MSA objects currently available*

Description

This function lists MSA objects currently available that can be used by 'ggmsa'

Usage

```
available_msa()
```

Value

A character vector of available objects

Author(s)

Lang Zhou

Examples

```
available_msa()
```

---

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

Description

A color scheme of Culstal.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()'.

---

facet_msa   *segment MSA*

Description

The MSA would be plot in a field that you set.

Usage

```
facet_msa(field)
```

Arguments

field   a numeric vector of the field size

Author(s)

Lang Zhou

Examples

```
library(ggplot2)
f <- system.file("extdata/sample.fasta", package="ggmsa")
# 2 fields
ggmsa(f, end = 120, font = NULL, color="Chemistry_AA") + facet_msa(field = 60)
# 3 fields
ggmsa(f, end = 120, font = NULL,  color="Chemistry_AA") + facet_msa(field = 40)
```

---

geom_asterisk   *a ggplot2 layer of star as a polygon*

Description

a ggplot2 layer of star as a polygon

Usage

```
geom_asterisk(
  mapping = NULL,
  data = NULL,
  stat = "identity",
  position = "identity",
  na.rm = FALSE,
  show.legend = NA,
  inherit.aes = TRUE,
  ...
)
```

Arguments

`mapping`	aes mapping
`data`	a data frame
`stat`	the statistical transformation to use on the data for this layer, as a string.
`position`	position adjustment, either as a string, or the result of a call to a position adjustment function.
`na.rm`	a logical value
`show.legend`	a logical value
`inherit.aes`	a logical value
`...`	additional parameters

Value

ggplot2 layer

Author(s)

Lang Zhou

Examples

```
library(ggplot2)
ggplot(mtcars, aes(mpg, disp)) + geom_asterisk()
```

---

geom_GC	*geom_GC*

Description

Multiple sequence alignment layer for ggplot2. It plot points of GC content.

Usage

```
geom_GC()
```

Author(s)

Lang Zhou

Examples

```
library(ggplot2)
#plot GC content
f <- system.file("extdata/LeaderRepeat_All.fa", package="ggmsa")
ggmsa(f, font = NULL, color="Chemistry_NT") + geom_GC()
```

---

geom_msa   *geom_msa*

Description

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

Usage

```
geom_msa(
  data,
  font = "helvetical",
  color = "Clustal",
  char_width = 0.9,
  none_bg = FALSE,
  posHighligthed = NULL,
  seq_name = NULL,
  ...
)
```

Arguments

data	multiple sequence alignment in tidy data frame, generated by tidy_fasta()
font	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'helvetical'. If font = NULL, only plot the background tile
color	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'.
char_width	characters width. Defaults is 0.9.
none_bg	a logical value indicating whether backgroud should be produced. Defaults is FALSE.
posHighligthed	A numeric vector of the position that need to be highlighted.
seq_name	a logical value indicating whether seqence names should be displayed. Defaults is 'NULL' which indicates that the sequence name is displayed when 'font = null', but 'font = char' will not be displayed. If 'seq_name = TRUE' the sequence name will be displayed in any case. If 'seq_name = FALSE' the sequence name will not be displayed under any circumstances.
...	additional parameter

Value

A list

Author(s)

Guangchuang Yu

---

geom_seed	*geom_seed*

Description

Highlighting the seed in miRNA sequences

Usage

```
geom_seed(seed, star = FALSE)
```

Arguments

seed   character, like 'GAGGUAG' a miRNA seed sequence

star   a logical value indicating whether symbol of stars should be produced

Value

a ggplot layer

Author(s)

Lang Zhou

Examples

```
miRNA_sequences <- system.file("extdata/seedSample.fa", package="ggmsa")
ggmsa(miRNA_sequences, font = 'DroidSansMono', color = "Chemistry_NT", none_bg = TRUE) +
geom_seed(seed = "GAGGUAG")
ggmsa(miRNA_sequences, font = 'DroidSansMono', color = "Chemistry_NT") +
geom_seed(seed = "GAGGUAG", star = TRUE)
```

---

geom_seqlogo   *geom_seqlogo*

Description

Multiple sequence alignment layer for ggplot2. It plot sequence motifs.

Usage

```
geom_seqlogo(font = "helvetical", color = "Chemistry_NT")
```

Arguments

font   font families, possible values are 'helvetical', 'mono', and 'DroidSansMono',
                'TimesNewRoman'. Defaults is 'helvetical'.

color   A Color scheme. One of 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Tay-
                lor_NT'. Defaults is 'Chemistry_NT'.

Value

A list

Author(s)

Lang Zhou

Examples

```
#plot multiple sequence alignment and sequence motifs
f <- system.file("extdata/LeaderRepeat_All.fa", package="ggmsa")
ggmsa(f,font = NULL,color = "Chemistry_NT") + geom_seqlogo()
```

---

ggmotif   *ggmotif*

Description

plot sequence motif for nucleotide sequences based 'ggolot2'

Usage

```
ggmotif(
  msa,
  start = NULL,
  end = NULL,
  font = "helvetical",
  color = "Chemistry_NT"
)
```

Arguments

msa	Multiple aligned sequence file or object for representing either nucleotide sequences or peptide sequences.
start	Start position to plot, If font=NULL, only the background frame is drawn, and no character.
end	End position to plot,If font=NULL, only the background frame is drawn, and no character.
font	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'helvetical'.
color	A Color scheme. One of 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_NT'.

Value

ggplot object

Author(s)

Lang Zhou

Examples

```
#plot sequence motif independently
nt_sequence <- system.file("extdata", "LeaderRepeat_All.fa", package = "ggmsa")
ggmotif(nt_sequence, color = "Chemistry_NT")
```

---

ggmsa	*ggmsa*

Description

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

Usage

```
ggmsa(
  msa,
  start = NULL,
  end = NULL,
  font = "helvetical",
  color = "Clustal",
  char_width = 0.9,
  none_bg = FALSE,
  posHighligthed = NULL,
  seq_name = NULL
)
```

Arguments

`msa`	Multiple aligned sequence file or object for representing either nucleotide sequences or peptide sequences.
`start`	a numeric.Start position to plot.
`end`	a numeric. End position to plot.
`font`	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'helvetical'. If font = NULL, only plot the background tile
`color`	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'.
`char_width`	characters width. Defaults is 0.9.
`none_bg`	a logical value indicating whether backgroud should be produced. Defaults is FALSE.
`posHighligthed`	A numeric vector of the position that need to be highlighted.
`seq_name`	a logical value indicating whether seqence names should be displayed. Defaults is 'NULL' which indicates that the sequence name is displayed when 'font = null', but 'font = char' will not be displayed. If 'seq_name = TRUE' the sequence name will be displayed in any case. If 'seq_name = FALSE' the sequence name will not be displayed under any circumstances.

Value

ggplot object

Author(s)

Guangchuang Yu

Examples

```
#plot multiple sequences by loading fasta format
fasta <- system.file("extdata", "sample.fasta", package = "ggmsa")
ggmsa(fasta, 164, 213, color="Chemistry_AA")

#XMultipleAlignment objects can be used as input in the 'ggmsa'
#AAMultipleAlignment <- Biostrings::readAAMultipleAlignment(fasta)
#ggmsa(AAMultipleAlignment, 164, 213, color="Chemistry_AA")

#XStringSet objects can be used as input in the 'ggmsa'
#AAStringSet <- Biostrings::readAAStringSet(fasta)
#ggmsa(AAStringSet, 164, 213, color="Chemistry_AA")

#Xbin objects from 'seqmagick' can be used as input in the 'ggmsa'
#AAbin <- seqmagick::fa_read(fasta)
#ggmsa(AAbin, 164, 213, color="Chemistry_AA")
```

---

tidy_msa   *tidy_msa*

Description

Convert msa file/object to tidy data frame

Usage

```
tidy_msa(msa, start = NULL, end = NULL)
```

Arguments

msa	multiple sequence alignment file or sequence object in DNAStringSet, RNAStringSet, AAStringSet, BStringSet, DNAMultipleAlignment, RNAMultipleAlignment, AAMultipleAlignment, DNAbin or AAbin
start	start position to extract subset of alignment
end	end position to extract subset of alignemnt

Value

tibble data frame

Author(s)

Guangchuang Yu

Index