Package 'ggmsa'

May 13, 2024

Title Plot Multiple Sequence Alignment using 'ggplot2'

Version 1.11.0

Description A visual exploration tool for multiple sequence alignment and associated data. Supports MSA of DNA, RNA, and protein sequences using 'ggplot2'. Multiple sequence alignment can easily be combined with other 'ggplot2' plots, such as phylogenetic tree Visualized by 'ggtree', boxplot, genome map and so on. More features: visualization of sequence logos, sequence bundles, RNA secondary structures and detection of sequence recombinations.

Depends R (>= 4.1.0)

Imports Biostrings, ggplot2, magrittr, tidyr, utils, stats, aplot, RColorBrewer, ggalt, ggforce, dplyr, R4RNA, grDevices, seqmagick, grid, methods, statebins, ggtree (>= 1.17.1)

Suggests ggtreeExtra, ape, cowplot, knitr, BiocStyle, rmarkdown, readxl, ggnewscale, kableExtra, gggenes, testthat (>= 3.0.0)

License Artistic-2.0

Encoding UTF-8

URL https://doi.org/10.1093/bib/bbac222(paper),
https://www.amazon.com/Integration-Manipulation-Visualization-Phylogenetic-Computational-ebook/dp/B0B5NLZR1Z/
(book)

BugReports https://github.com/YuLab-SMU/ggmsa/issues

biocViews Software, Visualization, Alignment, Annotation, MultipleSequenceAlignment

RoxygenNote 7.1.2

VignetteBuilder knitr

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/ggmsa

git_branch devel

git_last_commit 6f21768

git_last_commit_date 2024-04-30

Repository Bioconductor 3.20

Date/Publication 2024-05-13

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

Maintainer Lang Zhou <nyzhoulang@gmail.com>

Contents

`adjust_ally` *adjust_ally*

Description

adjust the tree branch position after assigning ancestor node

Usage

```
adjust_ally(tree, node, sub = FALSE, seq_colname = "mol_seq")
```

Arguments

`tree`	ggtree object
`node`	internal node in tree
`sub`	logical value.
`seq_colname`	the colname of MSA on tree$data

Value

tree

Author(s)

Lang Zhou

assign_dms *assign_dms*

Description

assign dms value to alignments.

Usage

```
assign_dms(x, dms)
```

Arguments

x	data frame from tidy_msa()
dms	dms data frame

Value

tree

Author(s)

Lang Zhou

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()
```

extract_seq *extract_seq*

Description

extract ancestor sequence from tree data

Usage

```
extract_seq(tree_adjust, seq_colname = "mol_seq")
```

Arguments

tree_adjust	ggtree object
seq_colname	the colname of MSA on tree$data

Value

character

Author(s)

Lang Zhou

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.

Value

ggplot layers

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_GC *geom_GC*

Description

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

Usage

```
geom_GC(show.legend = FALSE)
```

Arguments

show.legend	logical. Should this layer be included in the legends?

Value

a ggplot layer

Author(s)

Lang Zhou

Examples

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

geom_helix *geom_helix*

Description

The layer of helix plot

Usage

```
geom_helix(helix_data, color_by = "length", overlap = FALSE, ...)
```

Arguments

helix_data	a data frame. The file of nucleltide secondary structure and then read by readSSfile().
color_by	generate colors for helices by various rules, including integer counts and value ranges one of "length" and "value"
overlap	Logicals. If TRUE, two structures data called predict and known must be given(eg:heilx_data = list(known = data1, predicted = data2)), plots the predicted helices that are known on top, predicted helices that are not known on the bottom, and finally plots unpredicted helices on top in black.
...	additional parameter

Value

ggplot2 layers

Author(s)

Lang Zhou

Examples

```
RF03120 <- system.file("extdata/Rfam/RF03120_SS.txt", package="ggmsa")
RF03120_fas <- system.file("extdata/Rfam/RF03120.fasta", package="ggmsa")
SS <- readSSfile(RF03120, type = "Vienna")
ggmsa(RF03120_fas, font = NULL,border = NA,
    color = "Chemistry_NT", seq_name = FALSE) +
geom_helix(SS)
```

geom_msa *geom_msa*

Description

Multiple sequence alignment layer for ggplot2. It creates background tiles with/without sequence characters.

Usage

```
geom_msa(
  data,
  font = "helvetical",
  mapping = NULL,
  color = "Chemistry_AA",
  custom_color = NULL,
  char_width = 0.9,
  none_bg = FALSE,
  by_conservation = FALSE,
  position_highlight = NULL,
  seq_name = NULL,
  border = NULL,
  consensus_views = FALSE,
  use_dot = FALSE,
  disagreement = TRUE,
  ignore_gaps = FALSE,
  ref = NULL,
  position = "identity",
  show.legend = FALSE,
  dms = FALSE,
  position_color = FALSE,
  ...
)
```

Arguments

data	sequence alignment with data frame, generated by tidy_msa().
font	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'helvetical'.
mapping	aes mapping If font = NULL, only plot the background tile.
color	A Color scheme. One of 'Clustal', 'Chemistry_AA', 'Shapely_AA', 'Zappo_AA', 'Taylor_AA', 'LETTER','CN6',, 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_AA'.
custom_color	A data frame with two column called "names" and "color".Customize the color scheme.

char_width	a numeric vector. Specifying the character width in the range of 0 to 1. Defaults is 0.9.
none_bg	a logical value indicating whether background should be displayed. Defaults is FALSE.
by_conservation	a logical value. The most conserved regions have the brightest colors.
position_highlight	A numeric vector of the position that need to be highlighted.
seq_name	a logical value indicating whether sequence 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.
border	a character string. The border color.
consensus_views	a logical value that opening consensus views.
use_dot	a logical value. Displays characters as dots instead of fading their color in the consensus view.
disagreement	a logical value. Displays characters that disagreement to consensus(excludes ambiguous disagreements).
ignore_gaps	a logical value. When selected TRUE, gaps in column are treated as if that row didn't exist.
ref	a character string. Specifying the reference sequence which should be one of input sequences when 'consensus_views' is TRUE.
position	Position adjustment, either as a string, or the result of a call to a position adjustment function, default is 'identity' meaning 'position_identity()'.
show.legend	logical. Should this layer be included in the legends?
dms	logical.
position_color	logical.
...	additional parameter

Value

A list

Author(s)

Guangchuang Yu, Lang Zhou seq_name' work position_highlight' work border' work none_bg' work

Examples

```
library(ggplot2)
aln <- system.file("extdata", "sample.fasta", package = "ggmsa")
tidy_aln <- tidy_msa(aln, start = 150, end = 170)
ggplot() + geom_msa(data = tidy_aln, font = NULL) + coord_fixed()
```

geom_msaBar *geom_msaBar*

Description

Multiple sequence alignment layer for ggplot2. It plot sequence conservation bar.

Usage

```
geom_msaBar()
```

Value

A list

Author(s)

Lang Zhou

Examples

```
#plot multiple sequence alignment and conservation bar.
f <- system.file("extdata/sample.fasta", package="ggmsa")
ggmsa(f, 221, 280, font = NULL, seq_name = TRUE) + geom_msaBar()
```

geom_seed *geom_seed*

Description

Highlighting the seed in miRNA sequences

Usage

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

Arguments

seed	a character string.Specifying the miRNA seed sequence like 'GAGGUAG'.
star	a logical value indicating whether asterisks should be displayed.

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", star = FALSE)
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 = "DroidSansMono",
  color = "Chemistry_AA",
  adaptive = TRUE,
  top = TRUE,
  custom_color = NULL,
  show.legend = FALSE,
  ...
)
```

Arguments

font	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'DroidSansMono'.
color	A Color scheme. One of 'Clustal', 'Chemistry_AA', 'Shapely_AA', 'Zappo_AA', 'Taylor_AA', 'LETTER', 'CN6', 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_AA'.
adaptive	A logical value indicating whether the overall height of seqlogo corresponds to the number of sequences.If is FALSE, seqlogo overall height = 4,fixedly.
top	A logical value. If TRUE, seqlogo is aligned to the top of MSA.
custom_color	A data frame with two cloumn called "names" and "color".Customize the color scheme.
show.legend	logical. Should this layer be included in the legends?
...	additional parameter

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()
```

---

gghelix *gghelix*

Description

Plots nucleltide secondary structure as helices in arc diagram

Usage

```
gghelix(helix_data, color_by = "length", overlap = FALSE)
```

Arguments

helix_data	a data frame. The file of nucleltide secondary structure and then read by readSSfile().
color_by	generate colors for helices by various rules, including integer counts and value ranges one of "length" and "value"
overlap	Logicals. If TRUE, two structures data called predict and known must be given(eg:heilx_data = list(known = data1, predicted = data2)), plots the predicted helices that are known on top, predicted helices that are not known on the bottom, and finally plots unpredicted helices on top in black.

Value

ggplot object

Author(s)

Lang Zhou

Examples

```
RF03120 <- system.file("extdata/Rfam/RF03120_SS.txt", package="ggmsa")
helix_data <- readSSfile(RF03120, type = "Vienna")
gghelix(helix_data)
```

ggmaf *ggmaf*

Description

plot MAF

Usage

```
ggmaf(
  data,
  ref,
  block_start = NULL,
  block_end = NULL,
  facet_field = NULL,
  heights = c(0.4, 0.6),
  facet_heights = NULL
)
```

Arguments

data	a tidy MAF data frame.You can get it by tidy_maf_df()
ref	character, the name of reference genome. eg:"hg38.chr1_KI270707v1_random"
block_start	a numeric vector(>0). The start block to plot.
block_end	a numeric vector(< max block). The end block to plot.
facet_field	a numeric vector. The field in a facet panel.
heights	two numeric vector.The plot proportion between "Genomic location" panel(upon) and "Alignment" panel(down). Default:c(0.4,0.6)
facet_heights	Numeric vectors.The facet proportion.

Value

ggplot object

Author(s)

Lang Zhou

ggmsa *ggmsa*

Description

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

Usage

```
ggmsa(
  msa,
  start = NULL,
  end = NULL,
  font = "helvetical",
  color = "Chemistry_AA",
  custom_color = NULL,
  char_width = 0.9,
  none_bg = FALSE,
  by_conservation = FALSE,
  position_highlight = NULL,
  seq_name = NULL,
  border = NULL,
  consensus_views = FALSE,
  use_dot = FALSE,
  disagreement = TRUE,
  ignore_gaps = FALSE,
  ref = NULL,
  show.legend = FALSE
)
```

Arguments

msa	Multiple aligned sequence files or objects representing either nucleotide sequences or AA sequences.
start	a numeric vector. Start position to plot.
end	a numeric vector. 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', 'LETTER', 'CN6', 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_AA'.
custom_color	A data frame with two column called "names" and "color".Customize the color scheme.
char_width	a numeric vector. Specifying the character width in the range of 0 to 1. Defaults is 0.9.

none_bg	a logical value indicating whether background should be displayed. Defaults is FALSE.
by_conservation	a logical value. The most conserved regions have the brightest colors.
position_highlight	A numeric vector of the position that need to be highlighted.
seq_name	a logical value indicating whether sequence 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.
border	a character string. The border color.
consensus_views	a logical value that opening consensus views.
use_dot	a logical value. Displays characters as dots instead of fading their color in the consensus view.
disagreement	a logical value. Displays characters that disagreememt to consensus(excludes ambiguous disagreements).
ignore_gaps	a logical value. When selected TRUE, gaps in column are treated as if that row didn't exist.
ref	a character string. Specifying the reference sequence which should be one of input sequences when 'consensus_views' is TRUE.
show.legend	logical. Should this layer be included in the legends?

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")

## Not run:
#XMultipleAlignment objects can be used as input in the 'ggmsa'
AAMultipleAlignment <- readAAMultipleAlignment(fasta)
ggmsa(AAMultipleAlignment, 164, 213, color="Chemistry_AA")

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

#Xbin objects from 'seqmagick' can be used as input in the 'ggmsa'
```

```
AAbin <- fa_read(fasta)
ggmsa(AAbin, 164, 213, color="Chemistry_AA")

## End(Not run)
```

---

ggSeqBundle *ggSeqBundle*

Description

plot Sequence Bundles for MSA based 'ggolot2'

Usage

```
ggSeqBundle(
  msa,
  line_widch = 0.3,
  line_thickness = 0.3,
  line_high = 0,
  spline_shape = 0.3,
  size = 0.5,
  alpha = 0.2,
  bundle_color = c("#2ba0f5", "#424242"),
 lev_molecule = c("-", "A", "V", "L", "I", "P", "F", "W", "M", "G", "S", "T", "C",
    "Y", "N", "Q", "D", "E", "K", "R", "H")
)
```

Arguments

msa	Multiple sequence alignment file(FASTA) or object for representing either nucleotide sequences or peptide sequences.Also receives multiple MSA files. eg:msa = c("Gram-negative_AKL.fasta", "Gram-positive_AKL.fasta").
line_widch	The width of bundles at each site, default is 0.3.
line_thickness	The thickness of bundles at each site, default is 0.3.
line_high	The high of bundles at each site, default is 0.
spline_shape	A numeric vector of values between -1 and 1, which control the shape of the spline relative to the control points. From geom_xspline() in ggalt package.
size	A numeric vector of values between o and 1, which control the size of each lines.
alpha	A numeric vector of values between o and 1, which control the alpha of each lines.
bundle_color	The colors of each sequence bundles. eg: bundle_color = c("#2ba0f5","#424242").
lev_molecule	Reassigning the Y-axis and displaying letter-coded amino acids/nucleotides arranged by physiochemical properties or others.eg:amino acids hydrophobicity lev_molecule = c("-","A", "V", "L", "I", "P", "F", "W", "M", "G", "S","T", "C", "Y", "N", "Q", "D", "E", "K","R", "H").

Value

ggplot object

Author(s)

Lang Zhou

Examples

```
aln <- system.file("extdata", "Gram-negative_AKL.fasta", package = "ggmsa")
ggSeqBundle(aln)
```

Gram-negative_AKL.fasta

Gram-negative_AKL

Description

Amino acids in the adenylate kinase lid (AKL) domain from Gram-negative bacteria.

Format

A MSA fasta with 100 sequences and 36 positions.

Source

http://biovis.net/year/2013/info/redesign-contest

Gram-positive_AKL.fasta

Gram-positive_AKL

Description

Amino acids in the adenylate kinase lid (AKL) domain from Gram-positive bacteria.

Format

A MSA fasta with 100 sequences and 36 positions.

Source

http://biovis.net/year/2013/info/redesign-contest

GVariation *GVariation*

Description

A folder containing 4 MAS files as a sample data set to identify the sequence recombination event.

Format

a folder

Details

- A.Mont.fas MSA with sequences of 'Mont' and 'CF_YL21'
- B.Oz.fas MSA with sequences of 'Oz' and 'CF_YL21'
- C.Wilga5.fas MSA with sequences of 'Wilga5' and 'CF_YL21'
- sample_alignment.fa MSA with sequences of 'Mont', 'CF_YL21', 'Oz', and 'Wilga5'

Source

https://link.springer.com/article/10.1007/s11540-015-9307-3

LeaderRepeat_All.fa *A sample DNA alignment sequences*

Description

DNA alignment sequences with 24 sequences and 56 positions.

Format

A MSA fasta

merge_seq *merge_seq*

Description

merge two MSA

Usage

```
merge_seq(previous_seq, gap, subsequent_seq, adjust_name = TRUE)
```

Arguments

previous_seq	previous MSA
gap	gap length
subsequent_seq	subsequent MSA
adjust_name	logical value. merge seq name or not

Value

tidy MSA data frame

Author(s)

Lang Zhou

plot *plot method for SeqDiff object*

Description

plot method for SeqDiff object

Usage

```
## S4 method for signature 'SeqDiff,ANY'
plot(
  x,
  width = 50,
  title = "auto",
  xlab = "Nucleotide Position",
  by = "bar",
  fill = "firebrick",
  colors = c(A = "#ff6d6d", C = "#769dcc", G = "#f2be3c", T = "#74ce98"),
  xlim = NULL
)
```

Arguments

x	SeqDiff object
width	bin width
title	plot title
xlab	xlab
by	one of 'bar' and 'area'
fill	fill color of upper part of the plot
colors	color of lower part of the plot
xlim	limits of x-axis

Value

plot

Author(s)

guangchuang yu

Examples

```
fas <- list.files(system.file("extdata", "GVariation", package="ggmsa"),
                  pattern="fas", full.names=TRUE)
x1 <- seqdiff(fas[1], reference=1)
plot(x1)
```

---

readSSfile *readSSfile*

Description

Read secondary structure file

Usage

```
readSSfile(file, type = NULL)
```

Arguments

file	A text file in connect format
type	file type. one of "Helix, "Connect", "Vienna" and "Bpseq"

Value

data frame

Author(s)

Lang Zhou

Examples

```
RF03120 <- system.file("extdata/Rfam/RF03120_SS.txt", package="ggmsa")
helix_data <- readSSfile(RF03120, type = "Vienna")
```

---

read_maf *read_maf*

Description

read 'multiple alignment format'(MAF) file

Usage

```
read_maf(multiple_alignment_format)
```

Arguments

multiple_alignment_format	a multiple alignment format(MAF) file

Value

data frame

Author(s)

Lang Zhou

---

reset_pos *reset_pos*

Description

reset MSA position

Usage

```
reset_pos(seq_df)
```

Arguments

seq_df	MSA data

Value

data frame

Author(s)

Lang Zhou

`Rfam` *Rfam*

Description

A folder containing seed alignment sequences and corresponding consensus RNA secondary structure.

Format

a folder

Details

- RF00458.fasta seed alignment sequences of Cripavirus internal ribosome entry site (IRES)
- RF03120.fasta seed alignment sequences of Sarbecovirus 5'UTR
- RF03120_SS.txt consensus RNA secondary structure of Sarbecovirus 5'UTR

Source

https://rfam.xfam.org/

`sample.fasta` *A sample data used in ggmsa*

Description

A dataset containing the alignment sequences of the phenylalanine hydroxylase protein (PH4H) within nine species

Format

A MSA fasta with 9 sequences and 456 positions.

seedSample.fa *microRNA data used in ggmsa*

Description

Fasta format sequences of mature miRNA sequences from miRBase

Format

A MSA fasta with 6 sequences and 22 positions.

Source

https://www.mirbase.org/ftp.shtml

seqdiff *seqdiff*

Description

calculate difference of two aligned sequences

Usage

```
seqdiff(fasta, reference = 1)
```

Arguments

fasta	fasta file
reference	which sequence serve as reference, 1 or 2

Value

SeqDiff object

Author(s)

guangchuang yu

Examples

```
fas <- list.files(system.file("extdata", "GVariation", package="ggmsa"),
                  pattern="fas", full.names=TRUE)
seqdiff(fas[1], reference=1)
```

seqlogo *seqlogo*

Description

plot sequence logo for MSA based 'ggolot2'

Usage

```
seqlogo(
  msa,
  start = NULL,
  end = NULL,
  font = "DroidSansMono",
  color = "Chemistry_AA",
  adaptive = FALSE,
  top = FALSE,
  custom_color = NULL
)
```

Arguments

msa	Multiple sequence alignment file or object for representing either nucleotide sequences or peptide sequences.
start	Start position to plot.
end	End position to plot.
font	font families, possible values are 'helvetical', 'mono', and 'DroidSansMono', 'TimesNewRoman'. Defaults is 'DroidSansMono'. If font=NULL, only the background tiles is drawn.
color	A Color scheme. One of 'Clustal', 'Chemistry_AA', 'Shapely_AA', 'Zappo_AA', 'Taylor_AA', 'LETTER', 'CN6','Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_AA'.
adaptive	A logical value indicating whether the overall height of seqlogo corresponds to the number of sequences. If FALSE, seqlogo overall height = 4,fixedly.
top	A logical value. If TRUE, seqlogo is aligned to the top of MSA.
custom_color	A data frame with two cloumn called "names" and "color".Customize the color scheme.

Value

ggplot object

Author(s)

Lang Zhou

Examples

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

`sequence-link-tree.fasta` *sequence-link-tree*

Description

Alignment sequences used to demonstrate circular MSA layout

Format

A MSA fasta with 28 sequences and 480 positions.

`show` *show method*

Description

show method

Usage

```
show(object)
```

Arguments

`object`	SeqDiff object

Value

message

Examples

```
fas <- list.files(system.file("extdata", "GVariation", package="ggmsa"),
                  pattern="fas", full.names=TRUE)
x1 <- seqdiff(fas[1], reference=1)
x1
```

simplify_hdata *simplify_hdata*

Description

reset hdata data position

Usage

```
simplify_hdata(hdata, sim_msa)
```

Arguments

hdata	data from tidy_hdata()
sim_msa	MSA data frame

Value

data frame

Author(s)

Lang Zhou

simplot *simplot*

Description

Sequence similarity plot

Usage

```
simplot(
  file,
  query,
  window = 200,
  step = 20,
  group = FALSE,
  id,
  sep,
  sd = FALSE,
  smooth = FALSE,
  smooth_params = list(method = "loess", se = FALSE)
)
```

Arguments

file	alignment fast file
query	query sequence
window	sliding window size (bp)
step	step size to slide the window (bp)
group	whether grouping sequence.(eg. For "A-seq1,A-seq-2,B-seq1 and B-seq2", using sep = "-" and id = 1 to divide sequences into groups A and B)
id	position to extract id for grouping; only works if group = TRUE
sep	separator to split sequence name; only works if group = TRUE
sd	whether display standard deviation of similarity among each group; only works if group=TRUE
smooth	FALSE(default)or TRUE; whether display smoothed spline.
smooth_params	a list that add params for geom_smooth, (default: smooth_params = list(method = "loess", se = FALSE))

Value

ggplot object

Author(s)

guangchuang yu

Examples

```
fas <- system.file("extdata/GVariation/sample_alignment.fa",
                   package="ggmsa")
simplot(fas, 'CF_YL21')
```

---

theme_msa *theme_msa*

Description

Theme for ggmsa.

Usage

```
theme_msa()
```

Author(s)

Lang Zhou

tidy_hdata *tidy_hdata*

Description

tidy protein-protein interactive position data

Usage

```
tidy_hdata(gap, inter, previous_seq, subsequent_seq)
```

Arguments

gap	gap length
inter	protein-protein interactive position data
previous_seq	previous MSA
subsequent_seq	subsequent MSA

Value

helix data

Author(s)

Lang Zhou

tidy_maf_df *tidy_maf_df*

Description

tidy MAF data frame

Usage

```
tidy_maf_df(maf_df, ref)
```

Arguments

maf_df	a MAF data frame.You can get it by read_maf()
ref	character, the name of reference genome. eg:"hg38.chr1_KI270707v1_random"

Value

data frame

Author(s)

Lang Zhou

`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

Examples

```
fasta <- system.file("extdata", "sample.fasta", package = "ggmsa")
aln <- tidy_msa(msa = fasta, start = 10, end = 100)
```

`tp53.fa` *TP53 MSA*

Description

Alignment sequences of used to show graphical combination

Format

A MSA fasta with 5 sequences and 404 positions.

TP53_genes.xlsx *genome locus*

Description

The local genome map shows the 30000 sites around the TP53 gene.

Format

xlsx

treeMSA_plot *treeMSA_plot*

Description

plot Tree-MSA plot

Usage

```
treeMSA_plot(
  p_tree,
  tidymsa_df,
  ancestral_node = "none",
  sub = FALSE,
  panel = "MSA",
  font = NULL,
  color = "Chemistry_AA",
  seq_colname = NULL,
  ...
)
```

Arguments

p_tree	tree view
tidymsa_df	tidy MSA data
ancestral_node	vector, internal node in tree. Assigning a internal node to display "ancestral sequences",If ancestral_node = "none" hides all ancestral sequences, if ancestral_node = "all" shows all ancestral sequences.
sub	logical value. Displaying a subset of ancestral sequences or not.
panel	panel name for plot of MSA data
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', 'LETTER', 'CN6', 'Chemistry_NT', 'Shapely_NT', 'Zappo_NT', 'Taylor_NT'. Defaults is 'Chemistry_AA'.
seq_colname	the colname of MSA on tree$data
...	additional parameters for 'geom_msa'

Details

'treeMSA_plot()' automatically re-arranges the MSA data according to the tree structure,

Value

ggplot object

Author(s)

Lang Zhou

Index