Package: mzinspectr (via r-universe)

November 5, 2024

Type Package

Title Read and Analyze Mass Spectrometry Alignment Files

Version 0.4.2

Maintainer Ethan Bass <ethanbass@gmail.com>

Description A few functions for analyzing MS-DIAL alignments in R. Includes functions for feature normalization, subtraction of blanks, and mass library (msp) search.

URL https://github.com/ethanbass/mzinspectr

BugReports https://github.com/ethanbass/mzinspectr/issues

Imports dplyr, Formula, fs, OrgMassSpecR, pbapply, purrr, scales, stats, stringr, tidyr

Suggests DT, plotly, rcdk, shiny

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.0

Config/pak/sysreqs make libicu-dev

Repository https://ethanbass.r-universe.dev

RemoteUrl https://github.com/ethanbass/mzinspectr

RemoteRef HEAD

RemoteSha 436fc7511d42855c250844543ca4fd3bc8323fe2

Contents

`boxplot.ms_alignment` *Make boxplot from MS peak table.*

Description

The function can take multiple response variables on the left hand side of the formula (separated by +). In this case, a separate boxplot will be produced for each response variable.

Usage

```
## S3 method for class 'ms_alignment'
boxplot(x, formula, ...)
```

Arguments

`x`	A peak_table object
`formula`	A formula object
`...`	Additional arguments to `boxplot`

ms_attach_metadata *Attach experimental metadata*

Description

Attaches experimental metadata to 'ms_alignment' object. One of the columns in the supplied metadata must match exactly the row names of the peak table.

Usage

```
ms_attach_metadata(x, metadata, col)
```

Arguments

x	A ms_alignment object.
metadata	A 'data.frame' containing the sample metadata.
col	The name of the column containing the sample names.

Value

A ms_alignment object with attached metadata in the $sample_meta slot.

Author(s)

Ethan Bass

ms_calculate_RIs *Convert retention times to retention indices in alignment object.*

Description

Convert retention times to retention indices in alignment object.

Usage

```
ms_calculate_RIs(x, Ris)
```

Arguments

x	An ms_alignment object.
Ris	A matrix or data.frame containing retention times in column one and retention indices in column two.

ms_call_msdial	*Call MS-DIAL console app For help configuring the MS-DIAL console app on mac OSX or linux, please see the* R*hrefhttps://github.com/Jiung-Wen/msdialinstructions helpfully compiled by* R*hrefhttps://github.com/Jiung-Wen/Jiung-Wen Chen.*

Description

Call MS-DIAL console app For help configuring the MSDIAL console app on mac OSX or linux, please see the instructions helpfully compiled by Jiung-Wen Chen.

Usage

```
ms_call_msdial(
  system,
  path_in,
  path_out,
  method,
  settings,
  p = FALSE,
  mce = FALSE
)
```

Arguments

system	Either gcms, lcmsdda, lcsdia, lcmsdia, lcimmsdda, or lcimmsdia.
path_in	Path to files.
path_out	Path to output directory
method	A method file.
settings	Settings in lieu of a method file.
p	Logical.
mce	Logical

Value

Returns MSDIAL alignment.

`ms_filter_alignment` *Filter alignment by provided indices.*

Description

Filter alignment by provided indices.

Usage

```
ms_filter_alignment(x, idx, what = c("rows", "cols"), inverse = FALSE)
```

Arguments

`x`	An `ms_alignment` object or matrix with rows as samples and features as columns.
`idx`	Indices to be retained or excluded according to the value of `inverse`.
`what`	Which dimension to filter on. Either (`rows`) or columns (`cols`).
`inverse`	Whether to retain (default) or remove the specified columns.

Author(s)

Ethan Bass

`ms_find_peak` *Find peak based on retention time and/or mass*

Description

Find peak based on retention time and/or mass

Usage

```
ms_find_peak(x, rt, mz, rt.tol = 0.01, mz.tol = 0.05, plot_it = TRUE)
```

Arguments

`x`	An `ms_alignment` object
`rt`	Retention time
`mz`	Quant.mass
`rt.tol`	Tolerance for matching retention time
`mz.tol`	Tolerance for matching Quant.mass
`plot_it`	Logical. Whether to plot the spectra.

Value

Returns EI spectrum as a `data.frame`.

Author(s)

Ethan Bass

`ms_get_spectrum` *Get spectrum from MSDIAL alignment object*

Description

Get spectrum from MSDIAL alignment object

Usage

```
ms_get_spectrum(x, col)
```

Arguments

`x`	An `ms_alignment` object or matrix with rows as samples and features as columns.
`col`	Index of the feature (column).

Value

Returns spectrum as a data.frame with two columns: "mz" and "intensity".

Author(s)

Ethan Bass

`ms_mirror_plot` *Mirror plot function*

Description

Mirror plot function

Plot two spectra as a mirror plot.

Plot two spectra as a mirror plot.

Usage

```
ms_mirror_plot(x, ...)

## S3 method for class 'data.frame'
ms_mirror_plot(
  x,
  y,
  plot_labels = TRUE,
  type = c("plotly", "base"),
  scale = TRUE,
  lab_int = 0.2,
  digits = 1,
  bar_width = 1,
  match_score = TRUE,
  ...
)

## S3 method for class 'ms_alignment'
ms_mirror_plot(
  x,
  cols,
  ref,
  type = c("plotly", "base"),
  scale = TRUE,
  plot_labels = TRUE,
  lab_int = 0.2,
  digits = 1,
  bar_width = 1,
  match_score = TRUE,
  ...
)
```

Arguments

x	A ms_alignment object.
...	Additional arguments
y	Mass spectrum as data.frame with m/z values in column one and ionization intensity in column two.
plot_labels	Logical. Whether to label m/z values on plot.
type	What kind of plot to produce. Either base R (base) or plotly (plotly).
scale	Logical. Whether to scale mass spectrum. Defaults to TRUE.
lab_int	Labels will be plotted above the specified proportion of the largest ion.
digits	How many figures to include on m/z labels.
bar_width	Width of bars.
match_score	Logical. Whether to plot match score or not.

cols	One or more columns in the peak table tab to plot.
ref	A row in the matches slot corresponding to the provided column.

`ms_normalize_itsd` *Internal standard normalization*

Description

Normalize by internal standard.

Usage

```
ms_normalize_itsd(x, idx, plot_it = FALSE)
```

Arguments

x	An ms_alignment object or matrix with rows as samples and features as columns.
idx	Column index of internal standard.
plot_it	Logical. Whether to plot ITSD against total peak area.

Value

A normalized ms_alignment object or matrix, according to the input.

Author(s)

Ethan Bass

`ms_normalize_pqn` *Probabilistic Quotient Normalization*

Description

Performs Probabilistic Quotient Normalization on peak table.

Usage

```
ms_normalize_pqn(x, ref = c("median", "mean"), QC = NULL)
```

Arguments

x	A ms_alignment object or matrix with rows as samples and features as columns.
ref	Reference for normalization: either median (default) to use the overall median of variables as the reference, or mean to use the overall average of variables as the reference.
QC	vector of number(s) to specify samples which average to use as reference (e.g. QC samples)

Value

A normalized `ms_alignment` object or `matrix`, according to the input.

Note

Adapted from the Rcpm package by Rico Derks (licensed under GPL3).

Author(s)

E. Nevedomskaya

Rico Derks

Ethan Bass

References

Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in H1 NMR Metabonomics. Anal. Chem. 78, 4281-4290 (2006).

`ms_normalize_tsn` *Total sum normalization*

Description

Divides each row by the sum of the features in that row.

Usage

```
ms_normalize_tsn(x)
```

Arguments

`x`	An `ms_alignment` object or matrix with rows as samples and features as columns.

Value

A normalized `ms_alignment` object or `matrix`, according to the input.

`ms_plot_spectrum` *Plot mass spectrum of peak given by* `col`.

Description

Plot mass spectrum of peak given by `col`.

Usage

```
ms_plot_spectrum(
  x,
  col,
  plot_labels = TRUE,
  lab_int = 0.2,
  title = TRUE,
  type = c("plotly", "base"),
  scale = FALSE,
  bar_width = 1,
  digits = 1,
  ...
)
```

Arguments

`x`	An alignment object.
`col`	Spectrum to plot.
`plot_labels`	Logical. Whether to plot labels or not.
`lab_int`	Labels will be plotted above the specified proportion of the largest ion.
`title`	Logical. Whether to plot title. Defaults to TRUE.
`type`	What kind of plot to produce. Either base R (`base`) or plotly (`plotly`)
`scale`	Logical. Whether to scale mass spectrum. Defaults to FALSE.
`bar_width`	Width of bars.
`digits`	How many figures to include on mz labels
`...`	Additional arguments.

Value

If `export` is `TRUE`, returns spectrum as `data.frame`. Otherwise, no return value.

Author(s)

Ethan Bass

ms_read_alignment *Read MSDIAL alignment file*

Description

Read MSDIAL alignment file

Usage

```
ms_read_alignment(path, format = c("msdial"))
```

Arguments

path	Path to mass spectrometry alignment file.
format	The format of the provided alignment file. Currently, only MS-DIAL '.txt' files are supported (msdial).

Value

Returns ms_alignment object. A list of 3 data.frames, containing peak data (tab), peak metadata (peak_meta) and sample metadata (sample_meta).

Author(s)

Ethan Bass

ms_reshape_peaktable *Reshape peak table*

Description

Convert peak table to tidy format for plotting.

Usage

```
ms_reshape_peaktable(
  x,
  peaks,
  metadata,
  treatments = NULL,
  fixed_levels = TRUE
)
```

Arguments

x	An MS dial alignment object.
peaks	A character vector specifying the peaks to include in tidy output. If the character vector is named, the names of the vector elements will be used in place of the original peak names.
metadata	A character vector specifying the metadata to include in the tidy output.
treatments	This argument is deprecated as of version 0.3.2. It is synonymous with the new metadata argument which should be used instead.
fixed_levels	Logical. Whether to fix factor levels of features in the order provided. Defaults to TRUE.

Value

If export is TRUE, returns spectrum as data.frame. Otherwise, no return value.

Author(s)

Ethan Bass

---

ms_rt_to_ri *Convert retention times to retention indices.*

Description

Convert retention times to retention indices.

Usage

```
ms_rt_to_ri(rts, RIs)
```

Arguments

rts	A vector of retention times.
RIs	A matrix or data.frame containing retention times in column one and retention indices in column two.

ms_search_gadget *Launch MS search gadget for interactive viewing of spectral matches.*

Description

Launch MS search gadget for interactive viewing of spectral matches.

Usage

```
ms_search_gadget(data)
```

Arguments

data	An ms_alignment object.

ms_search_spectra *Search spectra in MS alignment against database*

Description

This function can be used to identify peaks in a peak table by matching them to a spectral database (db). It takes several arguments that can be used to customize the matching algorithm, including ri_thresh, spectral weight, n_results. The retention index threshold (ri_thresh) is used to subset the provided database, which greatly improves the search speed. Only database entries with a retention index falling within the specified threshold will be considered. The spectral weight affects the weight given to spectral similarity (versus retention index similarity) when calculating the the total similarity score, which is used to rank matches.

Usage

```
ms_search_spectra(
  x,
  db,
  cols,
  ...,
  ri_thresh = 100,
  spectral_weight = 0.6,
  n_results = 10,
  parallel,
  mc.cores = 2,
  print = FALSE,
  progress_bar = TRUE
)
```

Arguments

x	An ms_alignment object.
db	MSP database. The provided object should be a nested list, where the sublists contain the following elements: retention indices in an element named RI and mass spectra in an element called Spectra. All other elements are optional.
cols	Index or indices of feature(s) to be identified.
...	Additional arguments to SpectrumSimilarity.
ri_thresh	Maximum difference between retention indices for a match. to be considered. Defaults to 100. Use NULL to search database without filtering by retention time (this will take a long time for large databases).
spectral_weight	A number between 0 and 1 specifying the weight given. to spectral similarity versus retention index similarity. Defaults to 0.6.
n_results	How many results to return. Defaults to 10.
parallel	Logical. Whether to use parallel processing. (This feature does not work on Windows).
mc.cores	How many cores to use for parallel processing? Defaults to 2.
print	Logical. Whether to print the results after each search. Defaults to FALSE.
progress_bar	Logical. Whether to display progress bar or not.

Value

Returns a modified ms_alignment object with database matches in the matches slot as a list of data frames. Each data.frame will contain the database matches as rows and columns corresponding to the elements of the database entry (e.g. "Name", "InChIKey", etc.) as well as match scores for spectral similarity (spectral_match), retention index similarity (ri_match) and the total similarity score (total_score).

Note

See mspcompiler for help compiling an msp database.

Author(s)

Ethan Bass

---

ms_subtract_blanks *Subtract blanks*

Description

Subtract blanks

Usage

```
ms_subtract_blanks(
  x,
  blanks.idx,
  blanks.pattern,
  what = c("mean", "median"),
  drop = TRUE
)
```

Arguments

x	A ms_alignment object.
blanks.idx	Indices of blank samples
blanks.pattern	A string that uniquely identifies blank samples by name
what	Whether to subtract the mean or median value
drop	Logical. Whether to drop columns containing only zeros. Defaults to TRUE.

Value

A ms_alignment object with the mean or median of the blanks subtracted from each peak.

---

ms_tidy_msdial *Reshape peak table*

Description

Converts peak table to tidy format for plotting. This function is deprecated as of version 0.3.3. Please use ms_reshape_peaktable instead.

Usage

```
ms_tidy_msdial(x, peaks, metadata, treatments = NULL)
```

Arguments

x	An MS dial alignment object.
peaks	A character vector specifying the peaks to include in tidy output. If the character vector is named, the names of the vector elements will be used in place of the original peak names.
metadata	A character vector specifying the metadata to include in the tidy output.
treatments	This argument is deprecated as of version 0.3.2. It is synonymous with the new metadata argument which should be used instead.

Value

If export is TRUE, returns spectrum as data.frame. Otherwise, no return value.

Author(s)

Ethan Bass

`plot.ms_alignment` *Plot spectrum from peak table*

Description

Plots the trace and/or spectrum for a given peak in peak table.

Usage

```
## S3 method for class 'ms_alignment'
plot(
  x,
  col,
  plot_spectrum = TRUE,
  box_plot = FALSE,
  vars = NULL,
  spectrum_labels = TRUE,
  engine = c("base", "plotly"),
  ...
)
```

Arguments

`x`	A `ms.alignment` object.
`col`	A vector specifying the peak(s) that you wish to plot.
`plot_spectrum`	Logical. If TRUE, plots the mass spectrum of the chosen peak. Defaults to TRUE.
`box_plot`	Logical. If TRUE, plots box plot using factors defined by `vars`.
`vars`	Independent variables for boxplot. Righthand side of formula.
`spectrum_labels`	Logical. If TRUE, plots labels on maxima in spectral plot. Defaults to TRUE.
`engine`	Which plotting engine to use: either `base` or `plotly`.
`...`	Additional arguments to `boxplot`.

Details

Can be used to confirm the identity of a peak or check that a particular column in the peak table represents a single compound. Can also be used to create simple box-plots to examine the distribution of a peak with respect to variables defined in sample metadata.

Value

No return value.

Side effects

If `plot_spectrum` is TRUE, plots the spectrum for the specified chromatogram at the specified retention time. The spectrum is a single row from the chromatographic matrix.

If `box_plot` is TRUE, produces a `boxplot` from the specified peak with groups provided by `vars`.

Author(s)

Ethan Bass

`spectral_similarity` *Calculate spectral similarity between two peaks*

Description

This function is slightly adapted from the `SpectrumSimilarity` function in [OrgMassSpecR](https://orgmassspec.github.io/) where it is licensed under BSD-2 (© 2011-2017, Nathan Dodder). The function was re-factored here for increased speed.

Usage

```
spectral_similarity(
  spec.top,
  spec.bottom,
  tol = 0.25,
  b = 10,
  xlim = c(50, 1200),
  x.threshold = 0
)
```

Arguments

`spec.top`	data frame containing the experimental spectrum's peak list with the m/z values in the first column and corresponding intensities in the second.
`spec.bottom`	data frame containing the reference spectrum's peak list with the m/z values in the first column and corresponding intensities in the second.
`tol`	numeric value specifying the tolerance used to align the m/z values of the two spectra.
`b`	numeric value specifying the baseline threshold for peak identification. Expressed as a percent of the maximum intensity.
`xlim`	numeric vector of length 2, defining the beginning and ending values of the x-axis.
`x.threshold`	numeric value of length 1 specifying the m/z threshold used for the similarity score calculation. Only peaks with m/z values above the threshold are used in the calculation. This can be used to exclude noise and/or non-specific ions at the low end of the spectrum. By default all ions are used.

Author(s)

Nathan G. Dodder

Ethan Bass

Index