Package 'RAMClustR'

September 21, 2019

Type Package

Title Mass Spectrometry Metabolomics Feature Clustering and Interpretation

Version 1.0.9

Date 2019-09-17

Author Corey D. Broeckling, Fayyaz Afsar, Steffan Neumann, Asa Ben-Hur, Jessica Prenni.

Maintainer ``Broeckling,Corey" <Corey.Broeckling@ColoState.EDU>

Imports dynamicTreeCut, fastcluster, ff, InterpretMSSpectrum, BiocManager, httr, jsonlite, preprocessCore, e1071, gplots, pcaMethods, stringr, xml2, utils, webchem, stringi, RCurl

License GPL (>= 2)

Description A feature clustering algorithm for non-targeted mass spectrometric metabolomics data. This method is compatible with gas and liquid chromatography coupled mass spectrometry, including indiscriminant tandem mass spectrometry <DOI: 10.1021/ac501530d> data.

URL https://github.com/cbroeckl/RAMClustR

Encoding UTF-8

biocViews MassSpectrometry, Metabolomics

RoxygenNote 6.1.1

Suggests knitr, rmarkdown, xcms

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2019-09-20 23:10:02 UTC

R topics documented:

`annotate` *evalute ramSearch, MSFinder mssearch, MSFinder Structure, MSFinder Formula, and findmain output to annotate spectra of ramclustR object*

Description

After running RAMSearch (msp) and MSFinder on .mat or .msp files, import the spectral search results

Usage

```
annotate(ramclustObj = NULL,
  msfinder.dir = "C:/MSFinder/MSFINDER ver 3.24",
  standardize.names = FALSE, min.msms.score = 3.5,
  database.priority = "all", any.database.priority = TRUE,
  reset = TRUE)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`msfinder.dir`	full path to MSFinder directory - used for naming refinement.
`standardize.names`	logical: if TRUE, use inchikey for standardized chemical name lookup (http://cts.fiehnlab.ucdavis.edu/)

`min.msms.score` numerical: what is the minimum MSFinder similarity score acceptable. default = 3.5

`database.priority`
character. Formula assignment prioritization based on presence in one or more databases. Can be set to a single or multiple database names. must match database names as they are listed in MSFinder precisily. Can also be set to 'all' (note that MSFinder reports all databases matched, not just selected databases). If any database is set, the best formula match to that (those) database(s) is selected, rather than the best formula match overall.

`any.database.priority`
logical. First priority in formula assignment is based on any of the 'database.priority' values. Secondary priority from all other databases (determined in original MSFinder search) if TRUE. If false, formula assignment score from MSFinder used independent of structure search results.

`reset` logical. If TRUE, removes any previously assigned annotations.

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with the at $msfinder.formula, $msfinder.formula.score, $ann, and $ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

http://cts.fiehnlab.ucdavis.edu/static/download/CTS2-MS2015.pdf

annotation.summary *annotation.summary()*

Description

Write a .csv file containing a summary of the annotations in the ramclustR object.

Usage

```
annotation.summary(ramclustObj = NULL, outfile = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
outfile	file path/name of output csv summary file. if NULL (default) will be exported to spectra/annotaionSummary.csv

Details

this function exports a csv file summarizing annotation evidence for each compound

Value

nothing

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

assign.z *assign.z*

Description

infer charge state of features in ramclustR object.

Usage

```
assign.z(ramclustObj = NULL, chargestate = c(1:5), mzError = 0.02,
  nEvents = 2, minPercentSignal = 10, assume1 = TRUE)
```

Arguments

ramclustObj	ramclustR object to annotate
chargestate	integer vector. vector of integers of charge states to look for. default = c(1:5)
mzError	numeric. the error allowed in charge state m/z filtering. absolute mass units
nEvents	integer. the number of isotopes necessary to assign a charnge state > 1. default = 2.
minPercentSignal	numeric. the ratio of isotope signal (all isotopes) divided by total spectrum signal * 100 much be greated than minPercentSignal to evaluate charge state. Value should be between 0 and 100.
assume1	logical. when TRUE, m/z values for which no isotopes are found are assumed to be at z = 1.

Details

Annotation of ramclustR spectra. looks at isotope spacing for clustered features to infer charge state for each feature and a max charge state for each compound

Value

returns a ramclustR object. new slots holding:

zmax. vector with length equal to number of compounds. max charge state detected for that compound

fm. vector of inferred 'm', m/z value * z value

fz. vector of inferred 'z' values based on analysis of isotopes in spectrum.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

change.annotation *evalute ramSearch, MSFinder mssearch, MSFinder Structure, MSFinder Formula, and findmain output to annotate spectra of ramclustR object*

Description

After running RAMSearch (msp) and MSFinder on .mat or .msp files, import the spectral search results

Usage

```
change.annotation(ramclustObj = NULL,
  msfinder.dir = "C:/MSFinder/MSFINDER ver 3.22",
  standardize.names = FALSE, min.msms.score = 3.5,
  database.priority = "all", any.database.priority = TRUE,
  reset = TRUE)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
msfinder.dir	full path to MSFinder directory - used for naming refinement
standardize.names	logical: if TRUE, use inchikey for standardized chemical name lookup (http://cts.fiehnlab.ucdavis.edu/)
min.msms.score	numerical: what is the minimum MSFinder similarity score acceptable. default = 3.5
database.priority	character. Formula assignment prioritization based on presence in one or more databases. Can be set to a single or multiple database names. must match database names as they are listed in MSFinder precisily. Can also be set to 'all' (note that MSFinder reports all databases matched, not just selected databases). If any database is set, the best formula match to that (those) database(s) is selected, rather than the best formula match overall.
any.database.priority	logical. First priority in formula assignment is based on any of the 'database.priority' values. Secondary priority from all other databases (determined in original MSFinder search) if TRUE. If false, formula assignment score from MSFinder used independent of structure search results.
reset	logical. If TRUE, removes any previously assigned annotations.

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with the at $msfinder.formula, $msfinder.formula.score, $ann, and $ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

http://cts.fiehnlab.ucdavis.edu/static/download/CTS2-MS2015.pdf

---

`defineExperiment` *defineExperiment*

Description

Create an Experimental Design R object for record-keeping and msp output

Usage

```
defineExperiment(csv = FALSE, force.skip = FALSE)
```

Arguments

`csv`	logical or filepath. If csv = TRUE , csv template called "ExpDes.csv" will be written to your working directory. you will fill this in manually, ensuring that when you save you retain csv format. ramclustR will then read this file in and and format appropriately. If csv = FALSE, a pop up window will appear (in windows, at leaset) asking for input. If a character string with full path (and file name) to a csv file is given, this will allow you to read in a previously edited csv file.
`force.skip`	logical. If TRUE, ramclustR creates a pseudo-filled ExpDes object to enable testing of functionality. Not recommended for real data, as your exported spectra will be improperly labelled.

Value

an Exp Des R object which will be used for record keeping and writing spectra data.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

`do.findmain` *do.findmain*

Description

Cluster annotation function: inference of 'M' - molecular weight of the compound giving rise to each spectrum - using the InterpretMSSpectrum::findMain function

Usage

```
do.findmain(ramclustObj = NULL, cmpd = NULL, mode = "positive",
  mzabs.error = 0.01, ppm.error = 10, ads = NULL, nls = NULL,
  scoring = "auto", plot.findmain = TRUE, writeMat = TRUE,
  writeMS = TRUE, use.z = TRUE)
```

Arguments

`ramclustObj`	ramclustR object to annotate.
`cmpd`	integer: vector defining compound numbers to annotated. if NULL (default), all compounds
`mode`	character: "positive" or "negative"
`mzabs.error`	numeric: absolute mass deviation allowd, default = 0.01
`ppm.error`	numeric: ppm mass error _added_ to mzabs.error, default = 10
`ads`	character: vector of allowed adducts, i.e. c("[M+H]+"). if NULL, default positive mode values of H+, Na+, K+, and NH4+, as monomer, dimer, and trimer, are assigned. Negative mode include "[M-H]-", "[M+Na-2H]-", "[M+K-2H]-", "[M+CH2O2-H]-" as monomer, dimer, and trimer.

nls	character: vector of allowed neutral losses, i.e. c("[M+H-H2O]+"). if NULL, an extensive list derived from CAMERA's will be used.
scoring	character: one of 'imss' , 'ramclustr', or 'auto'. default = 'auto'. see details.
plot.findmain	logical: should pdf polts be generated for evaluation? detfault = TRUE. PDF saved to working.directory/spectra
writeMat	logical: should indidual .mat files (for MSFinder) be generated in a 'mat' subdirectory in the 'spectra' folder? default = TRUE.
writeMS	logical: should indidual .ms files (for Sirius) be generated in a 'ms' subdirectory in the 'spectra' folder? default = TRUE. Note that no import functions are yet written for Sirius output.
use.z	logical: if you have previously run the 'assign.z' function from ramclustR, there will be a slot reflecting the feature mass after accounting for charge (fm) - if TRUE this is used instead of feature m/z (fmz) in interpreting MS data and exporing spectra for annotation.

Details

a partially annotated ramclustR object. base structure is that of a standard R heirarchical clustering output, with additional slots described in ramclustR documentation (?ramclustR). New slots added after using the interpretMSSpectrum functionality include those described below.

Value

$M: The inferred molecular weight of the compound giving rise to the each spectrum

$M.ppm: The ppm error of all the MS signals annotated, high error values should be considered 'red flags'.

$M.ann: The annotated spectrum supporting the intepretation of M

$use.findmain: Logical vector indicating whether findmain scoring (TRUE) or ramclustR scoring (FALSE) was used to support inference of M. By default, findmain scoring is used. When ramclustR scoring differs from findmain scoring, the scoring metric which predicts higher M is selected.

$M.ramclustr: M selected using ramclustR scoring

$M.ppm.ramclustr: ppm error of M selected using ramclustR scoring. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.ann.ramclustr: annotated spectrum supporing M using ramclustR scoring

$M.nann.ramclustr: number of masses annotated using ramclustR scoring. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.space.ramclustr: the 'space' of scores between the best and second best ramclustR scores. Calculated as a ratio. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.findmain: M selected using findmain scoring

$M.ppm.findmain: ppm error of M selected using findmain scoring. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.ann.findmain: annotated spectrum supporing M using findmain scoring

$M.nann.findmain: number of masses annotated using findmain scoring. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

$M.space.findmain: the 'space' of scores between the best and second best findmain scores. Calculated as a ratio. Used to resolve concflicts between ramclustR and findmain M assignment when scoring = auto.

Author(s)

Corey Broeckling

References

Jaeger C, ... Lisec J. Compound annotation in liquid chromatography/high-resolution mass spectrometry based metabolomics: robust adduct ion determination as a prerequisite to structure prediction in electrospray ionization mass spectra. Rapid Commun Mass Spectrom. 2017 Aug 15;31(15):1261-1266. doi: 10.1002/rcm.7905. PubMed PMID: 28499062.

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

---

export.msfinder.formulas

export MSFinder formula prediction results in tabular format.

Description

After running MSFinder, results have been imported to the ramclustR object. This function exports as a .csv file for ease of viewing.

Usage

```
export.msfinder.formulas(ramclustObj = NULL, export.all = FALSE,
  output.directory = NULL)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`export.all`	logical: default = FALSE. If TRUE, export all columns, if FALSE, only columns 1: "exactmass"
`output.directory`	valid path: default = NULL. If NULL, results are exported to spectra/mat directory.

Details

this function exports a .csv file containing all returned MSFinder molecular formula hypotheses. this file is saved (by default) to the working directory spectra/mat/ directory

Value

an updated ramclustR object, with the RC$ann and RC$ann.conf slots updated to annotated based on output from 1. ramsearch output, 2. msfinder mssearch, 3. msfinder predicted structure, 4. msfinder predicted formula, and 5. interpretMSSpectrum inferred molecular weight, with listed order as priority.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

---

exportDataset *exportDataset*

Description

export one of 'SpecAbund', 'SpecAbundAve', 'MSdata' or 'MSMSdata' from an RC object to csv

Usage

```
exportDataset(ramclustObj = NULL, which.data = "SpecAbund",
  label.by = "ann")
```

Arguments

ramclustObj	ramclustR object to export from
which.data	name of dataset to export. SpecAbund, SpecAbundAve, MSdata, or MSMSdata
label.by	either 'ann' or 'cmpd', generally. name of ramclustObj slot used as csv header for each column (compound)

Details

Useful for exporting the processed signal intensity matrix to csv for analysis elsewhere.

Value

nothing is returned. file exported as csf to 'datasets/*.csv'

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`findmass` *findmass*

Description

see if any features match a given mass, and whether they are plausibly M0

Usage

```
findmass(ramclustObj = NULL, mz = NULL, mztol = 0.02, rttol = 2,
  zmax = 6, m.check = TRUE)
```

Arguments

`ramclustObj`	R object: the ramclustR object to explore
`mz`	numeric: mz value to search for
`mztol`	numeric: absolute mass tolerance around mz
`rttol`	numeric: when examining isotope patterns, feaure retention time tolerance around features matching mz +- mztol
`zmax`	integer: maximum charge state to consider. default is 6.
`m.check`	logical: check whether the matching masses are plausibly M0. That is, we look for ions 1 proton mass (from charge state 1:zmax) below the target m/z at the same time that have intensities consistent with target ion being a non-M0 isotope.

Details

a convenience function to perform a targeted search of all feaures for a mass of interest. Also performs a crude plausibility check as to whether the matched feature could be M0, based on the assumption of approximately 1 carbon per 17 m/z units and natural isottopic abundance of 1.1

Value

returns a table to the console listing masses which match, their retention time and intensity, and whether it appears to be plausible as M0

Author(s)

Corey Broeckling

`fooddb2msfinder` *foodb2msfinder*

Description

convenience function for converting FoodDB database export format to MSFinder custom database import format. Before running this, please have downloaded .csv files from FoodDB with the appropriate Display Field Headers (see details)

Usage

```
fooddb2msfinder(foodb.files = NULL, out.dir = NULL,
  out.name = "FoodDB_for_MSFinder.txt")
```

Arguments

`foodb.files`	default = NULL, if path is set, will read automatically. If NULL, direcory selection by user.
`out.dir`	default = NULL. Can set to exiseting directory with full path name. If NULL, direcory selection by user.
`out.name`	default = "FoodDB_for_MSFinder.txt".

Details

Input file(s) should be csv formatted, with required headers of 'Name', 'Smiles', 'Inchikey', 'Chemical formula', and 'Mono mass' - case sensitive. Ouput will be in tab delimited text format in directory of choice.

Value

Nothing is returned - output file written to directory set by 'out.dir' and name set by 'out.name'

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

`get.synonyms` *get.synonyms()*

Description

Use chemical translation service to retreive synonyms for ramclustR inchikeys

Usage

```
get.synonyms(ramclustObj = NULL, get.db = TRUE, update.names = TRUE,
  lipid.short.hand = TRUE)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`get.db`	logical - should the HMDB, LipidMaps, CheBI, and Pubchem names be retreived when available?
`update.names`	logical - should the ramclustObj$ann slot (annotation) be updated based on the above? selection of which synonym to chose is difficult to automate well - new name may not be the most commonly used.
`lipid.short.hand`	- logical - should the stringr pacakge be used to look for lipid short hand nomenclature in the synonyms? Only used when update.names = TRUE

Details

this function uses the chemical translation service (http://cts.fiehnlab.ucdavis.edu/), HMDB, LipidMaps, and PubChem databases to retreive synonymns and compound names ror a given inchikey). Lipid shorthand (i.e. PC(36:6)) can be identified and used when available. Precendence for naming is lipid.short.hand > HMDB > LipidMaps > Pubchem > original assignment.

Value

an updated ramclustR object, with the RC$synonyms slot containing a list with length equal to the number of compounds in the dataset. Each list element represents a character vector of all the synonyms returns (or NA, for compounds with no inchikey)

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

http://cts.fiehnlab.ucdavis.edu/static/download/CTS2-MS2015.pdf

---

`getClassyFire` *getClassyFire*

Description

use classyfire web API to look up full ClassyFire heirarchy for each inchikey

Usage

```
getClassyFire(ramclustObj = NULL, get.all = TRUE, max.wait = 10,
  posts.per.minute = 5)
```

Arguments

`ramclustObj`	ramclustR object to ClassyFy
`get.all`	logical; if TRUE, when inchikey classyfire lookup fails, submits for classyfication. Can be slow. max.wait (below) sets max time to spend on each compound before moving on. default = FALSE.
`max.wait`	numeric; maximum time to wait per compound when 'get.all' = TRUE.
`posts.per.minute`	integer; a limit set when 'get.all' is true. ClassyFire server accepts no more than 5 posts per minute when calculating new ClassyFire results. Slows down submission process to keep server from denying access.

Details

The $inchikey slot is used to look up the

Value

returns a ramclustR object. new dataframe in $classyfire slot with rows equal to number of compounds.

Author(s)

Corey Broeckling

References

Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, and Wishart DS. ClassyFire: Automated Chemical Classification With A Comprehensive, Computable Taxonomy. Journal of Cheminformatics, 2016, 8:61. DOI: 10.1186/s13321-016-0174-y

---

`getSmilesInchi` *getSmilesInchi*

Description

use PubChem API to look up full smiles and inchi notation for each inchikey

Usage

```
getSmilesInchi(ramclustObj = NULL)
```

Arguments

`ramclustObj`	ramclustR object to ClassyFy

Details

The $inchikey slot is used to look up parameters from pubchem. PubChem CID, a pubchem URL, smiles (canonical) and inchi are returned. if smiles and inchi slots are alread present (from MS-Finder, for example) pubchem smiles and inchi are used to fill in missing values only, not replace.

Value

returns a ramclustR object. new dataframe in $classyfire slot with rows equal to number of compounds.

Author(s)

Corey Broeckling

References

Kim S, Thiessen PA, Bolton EE, Bryant SH. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. Nucleic Acids Res. 2015;43(W1):W605-11.

`import.msfinder.formulas` *import.msfinder.formulas*

Description

After running MSFinder on .mat or .msp files, import the formulas that were predicted and their scores

Usage

```
import.msfinder.formulas(ramclustObj = NULL, mat.dir = NULL,
  msp.dir = NULL)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`mat.dir`	optional path to .mat directory
`msp.dir`	optional path to .msp directory

Details

this function imports the output from the MSFinder program to support annotation of the ramclustR object

Value

new slot at $msfinder.formula.details

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

`import.msfinder.mssearch` *import.MSFinder.mssearch*

Description

After running MSFinder on .mat or .msp files, import the spectral search results

Usage

```
import.msfinder.mssearch(ramclustObj = NULL, mat.dir = NULL,
  msp.dir = NULL)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`mat.dir`	optional path to .mat directory
`msp.dir`	optional path to .msp directory

Details

this function imports the output from the MSFinder program to annotate the ramclustR object

Value

an updated ramclustR object, with new slots at $msfinder.mssearch.details and $msfinder.mssearch.scores

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

`import.msfinder.structures` *write.methods*

Description

write RAMClustR processing methods and citations to text file

Usage

```
import.msfinder.structures(ramclustObj = NULL, mat.dir = NULL,
  msp.dir = NULL)
```

Arguments

`ramclustObj`	R object - the ramclustR object which was used to write the .mat or .msp files
`mat.dir`	directory in which to look for mat file MSFinder output - by default the /spectra/mat in the working directory
`msp.dir`	directory in which to look for msp file MSFinder output - by default the /spectra/msp in the working directory

Details

this function exports a file called ramclustr_methods.txt which contains the processing history, parameters used, and relevent citations.

Value

an annotated ramclustR object

nothing - new file written to working director

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

impRamSearch *impRamSearch*

Description

import ramsearch output for annotating an RC object

Usage

```
impRamSearch(ramclustObj = NULL, ramsearchout = "spectra/results.rse")
```

Arguments

ramclustObj	ramclustR object to annotate
ramsearchout	path to .rse file to import

Details

Annotation of ramclustR exported .msp spectra is accomplished using RAMSearch. Exported ramsearch annotations (.rse) can be imported with this function

Value

returns a ramclustR object. new slots holding .rse data

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

manual.annotation.template
manual.annotation.template

Description

export a .csv formatted template for manually editing MSFinder annotations

Usage

```
manual.annotation.template(ramclustObj = NULL,
  outfile = "manual.annotation.template.csv")
```

Arguments

ramclustObj	ramclustR object to annotate
outfile	output file directory and name. default = 'manual.annotation.template.csv'

Details

While unsupervised annotation is rapid and objective, subjective knowledge can be used to improve annotations. This function writes a template file containing compound name, computationally assigned inchikey, and an empty column for your manually inferred inchikey. Upon completion of manual annotation, you can reimport this file and update your ramclustR object to reflect your manual input.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem. 2016 Aug 16;88(16):7946-58. doi: 10.1021/acs.analchem.6b00770. Epub 2016 Aug 4. PubMed PMID: 27419259.

mergeRCobjects *mergeRCobjects*

Description

merge two ramclustR objects

Usage

```
mergeRCobjects(ramclustObj.1 = NULL, ramclustObj.2 = NULL,
  mztol = 0.02, rttol = 30, mzwt = 2, rtwt = 1)
```

Arguments

ramclustObj.1	ramclustR object 1: this object will be the base for the new object. That is all the features from ramclustObj.1 will be retained.
ramclustObj.2	ramclustR object 2: this object will mapped and appended to racmlustObj1. That is only features which appear consistent with those from ramclustObj.1 will be retained.
mztol	numeric: absolute mass tolerance around mz
rttol	numeric: feaure retention time tolerance. Value set by this option will be used during the initial anchor mapping phase. Two times the standard error of the rt loess correction will be used for the full mapping.
mzwt	numeric: when mapping features, weighting value used for similarities between feature mass values (see rtwt)
rtwt	numeric: when mapping features, weighting value used for similarities between feature retention time values (see mzwt)

Details

Two ramclustR objects are merged with this function, mapping features between them. The first (ramclustObj.1) object use used as the template - all data in it is retained. ramclustObj.2 is mapped to ramclustObj.1 feature by feature - only mapped features are retained. A new ramlcustObj is returned, with a new SpecAbund dataset with the same column number as the ramclustObj.1$SpecAbund set.

Value

returns a ramclustR object. All values from ramclustObj.1 are retained. SpecAbund dataset from ramclustObj.1 is moved to RC$SpecAbund.1, where RC is the new ramclustObj.

Author(s)

Corey Broeckling

ramclustR *ramclustR*

Description

Main clustering function for grouping features based on their analytical behavior.

Usage

```
ramclustR(xcmsObj = NULL, ms = NULL, idmsms = NULL,
  taglocation = "filepaths", MStag = NULL, idMSMStag = NULL,
  featdelim = "_", timepos = 2, st = NULL, sr = NULL,
  maxt = NULL, deepSplit = FALSE, blocksize = 2000, mult = 5,
  hmax = NULL, sampNameCol = 1, collapse = TRUE, usePheno = TRUE,
  mspout = TRUE, ExpDes = NULL, normalize = "TIC",
  qc.inj.range = 20, order = NULL, batch = NULL, qc = NULL,
  minModuleSize = 2, linkage = "average", mzdec = 3,
  cor.method = "pearson", rt.only.low.n = TRUE, fftempdir = NULL)
```

Arguments

xcmsObj	xcmsObject: containing grouped feature data for clustering by ramclustR
ms	filepath: optional csv input. Features as columns, rows as samples. Column header mz_rt
idmsms	filepath: optional idMSMS / MSe csv data. same dim and names as ms required
taglocation	character: "filepaths" by default, "phenoData[,1]" is another option. referse to xcms slot
MStag	character: character string in 'taglocation' to designat MS / MSe files e.g. "01.cdf"
idMSMStag	character: character string in 'taglocation' to designat idMSMS / MSe files e.g. "02.cdf"
featdelim	character: how feature mz and rt are delimited in csv import column header e.g. ="-"
timepos	integer: which position in delimited column header represents the retention time (csv only)
st	numeric: sigma t - time similarity decay value
sr	numeric: sigma r - correlational similarity decay value
maxt	numeric: maximum time difference to calculate retention similarity for - all values beyond this are assigned similarity of zero
deepSplit	logical: controls how agressively the HCA tree is cut - see ?cutreeDynamicTree
blocksize	integer: number of features (scans?) processed in one block =1000,
mult	numeric: internal value, can be used to influence processing speed/ram usage
hmax	numeric: precut the tree at this height, default 0.3 - see ?cutreeDynamicTree

sampNameCol	integer: which column from the csv file contains sample names?
collapse	logical: reduce feature intensities to spectrum intensities?
usePheno	logical: tranfer phenotype data from XCMS object to SpecAbund dataset?
mspout	logical: write msp formatted specta to file?
ExpDes	either an R object created by R ExpDes object: data used for record keeping and labelling msp spectral output
normalize	character: either "none", "TIC", "quantile", or "batch.qc" normalization of feature intensities. see batch.qc overview in details.
qc.inj.range	integer: how many injections around each injection are to be scanned for presence of QC samples when using batch.qc normalization? A good rule of thumb is between 1 and 3 times the typical injection span between QC injections. i.e. if you inject QC ever 7 samples, set this to between 7 and 21. smaller values provide more local precision but make normalization sensitive to individual poor outliers (though these are first removed using the boxplot function outlier detection), while wider values provide less local precision in normalization but better stability to individual peak areas.
order	integer vector with length equal to number of injections in xset or csv file
batch	integer vector with length equal to number of injections in xset or csv file
qc	logical vector with length equal to number of injections in xset or csv file.
minModuleSize	integer: how many features must be part of a cluster to be returned? default = 2
linkage	character: heirarchical clustering linkage method - see ?hclust
mzdec	integer: number of decimal places used in printing m/z values
cor.method	character: which correlational method used to calculate 'r' - see ?cor
rt.only.low.n	logical: default = TRUE At low injection numbers, correlational relationships of peak intensities may be unreliable. by defualt ramclustR will simply ignore the correlational r value and cluster on retention time alone. if you wish to use correlation with at n < 5, set this value to FALSE.
fftempdir	valid path: if there are file size limitations on the default ff pacakge temp directory - getOptions('fftempdir') - you can change the directory used as the fftempdir with this option.

Details

Main clustering function output - see citation for algorithm description or vignette('RAMClustR') for a walk through. batch.qc. normalization requires input of three vectors (1) batch (2) order (3) qc. This is a feature centric normalization approach which adjusts signal intensities first by comparing batch median intensity of each feature (one feature at a time) QC signal intensity to full dataset median to correct for systematic batch effects and then secondly to apply a local QC median vs global median sample correction to correct for run order effects.

Value

$featclus: integer vector of cluster membership for each feature

$frt: feature retention time, in whatever units were fed in (xcms uses seconds, by default)

$fmz: feature retention time, reported in number of decimal points selected in ramclustR function

$xcmsOrd: the original XCMS (or csv) feature order for cross referencing, if need be

$clrt: cluster retention time

$clrtsd: retention time standard deviation of all the features that comprise that cluster

$nfeat: number of features in the cluster

$nsing: number of 'singletons' - that is the number of features which clustered with no other feature

$ExpDes: the experimental design object used when running ramclustR. List of two dataframes.

$cmpd: compound name. C#### are assigned in order of output by dynamicTreeCut. Compound with the most features is classified as C0001...

$ann: annotation. By default, annotation names are identical to 'cmpd' names. This slot is a placeholder for when annotations are provided

$MSdata: the MSdataset provided by either xcms or csv input

$MSMSdata: the (optional) MSe/idMSMS dataset provided be either xcms or csv input

$SpecAbund: the cluster intensities after collapsing features to clusters

$SpecAbundAve: the cluster intensities after averaging all samples with identical sample names

- 'spectra' directory is created in the working directory. In this directory a .msp is (optionally) created, which contains the spectra for all compounds in the dataset following clustering. if MSe/idMSMS data are provided, they are listed witht he same compound name as the MS spectrum, with the collision energy provided in the ExpDes object provided to distinguish low from high CE spectra.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

RCQC *RCQC*

Description

filter RC object and summarize quality control sample varation

Usage

```
RCQC(ramclustObj = NULL, qctag = "QC", npc = 4, scale = "pareto",
  which.data = "SpecAbund", outfile = "ramclustQC.pdf")
```

Arguments

ramclustObj	ramclustR object to analyze
qctag	"QC" by default - rowname tag to identify QC samples
npc	number of Principle components to calcuate and plot
scale	"pareto" by default: PCA scaling method used
which.data	which dataset to use. "SpecAbund" by default
outfile	name of output pdf file.

Details

plots a ramclustR summary plot. first page represents the correlation of each cluster to all other clusters, sorted by retention time. large blocks of yellow along the diaganol indicate either poor clustering or a group of coregulated metabolites with similar retention time. It is an imperfect diagnostic, partuclarly with lipids on reverse phase LC or sugars on HILIC LC systems. Page 2: histogram of r values from page 1 - only r values one position from the diagonal are used. Pages 3:5 - PCA results, with QC samples colored red. relative standard deviation calculated as sd(QC PC scores) / sd(all PC scores). Page 6: histogram of CV values for each compound int he dataset, QC samples only.

Value

new RC object, with QC samples moved to new slot. prints output summary plots to pdf.

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Broeckling CD, Ganna A, Layer M, Brown K, Sutton B, Ingelsson E, Peers G, Prenni JE. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. Anal Chem. 2016 Sep 20;88(18):9226-34. doi: 10.1021/acs.analchem.6b02479. Epub 2016 Sep 8. PubMed PMID: 7560453.

write.methods *write.methods*

Description

write RAMClustR processing methods and citations to text file

Usage

```
write.methods(ramclustObj = NULL, filename = NULL)
```

Arguments

ramclustObj	R object - the ramclustR object which was used to write the .mat or .msp files
filename	define filename/path to write. uses 'ramclustr_methods.txt' and the working directory by default.

Details

this function exports a file called ramclustr_methods.txt which contains the processing history, parameters used, and relevent citations.

Value

an annotated ramclustR object

nothing - new file written to working director

Author(s)

Corey Broeckling

References

Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014 Jul 15;86(14):6812-7. doi: 10.1021/ac501530d. Epub 2014 Jun 26. PubMed PMID: 24927477.

Index