Package 'R2DGC'

August 22, 2017

Type Package

Title Multiple Peak Alignment for 2D Gas Chromatography Mass Spectrometry Metabolomics Analysis

Version 1.0.3

Author Ryne Ramaker <ryneramaker@gmail.com> Emily Gordon <egordon@hudsonalpha.org> Sara Cooper <sjcooper@hudsonalpha.org>

Maintainer Ryne Ramaker <ryneramaker@gmail.com>

Description Provides functions for aligning 2D gas chromatography mass spectrometry derived metabolite peaks obtained from primary processing and generates an alignment table that allows for a comparison of common peaks across samples and metabolite identification. Publication describing the package in detail is available at the following citation: Ryne C. Ramaker, Emily Gordon, Sara J. Cooper (2017) <doi:10.1101/179168>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.0.0)

RoxygenNote 6.0.1

Imports parallel

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2017-08-22 17:26:21 UTC

R topics documented:

`ConsensusAlign` *Takes a vector of paths to input files and aligns common metabolites into a final table. Will also identify metabolites if a reference library is provided*

Description

Takes a vector of paths to input files and aligns common metabolites into a final table. Will also identify metabolites if a reference library is provided

Usage

```
ConsensusAlign(inputFileList, RT1_Standards = NULL, RT2_Standards = NULL,
  seedFile = 1, RT1Penalty = 1, RT2Penalty = 10,
  autoTuneMatchStringency = TRUE, similarityCutoff = 90,
  disimilarityCutoff = similarityCutoff - 90, numCores = 1,
  commonIons = c(), missingValueLimit = 0.75,
  missingPeakFinderSimilarityLax = 0.85, quantMethod = "T",
  standardLibrary = NULL)
```

Arguments

`inputFileList`	Vector of file paths to align
`RT1_Standards`	Vector of standard names used to adjust first retention time. All names must be found in input files. Defaults to NULL.
`RT2_Standards`	Vector of standard names used to adjust second retention time. All names must be found in input files. Defaults to NULL.
`seedFile`	File number in inputFileList to initialize alignment. Can also input a vector of different seed files (3 is usually sufficient) to prevent bias from seed file. Defaults to 1.
`RT1Penalty`	Penalty used for first retention time errors. Defaults to 1.
`RT2Penalty`	Penalty used for first retention time errors. Defaults to 10.
`autoTuneMatchStringency`	Will automatically find optimal match threshold. If TRUE, will ignore similarityCutoff. Defaults to TRUE.
`similarityCutoff`	Adjusts peak similarity threshold required for alignment. Adjust in concordance with RT1 and RT2 penalties. Will be ignored if autoTuneMatchStrigency is TRUE. Defaults to 90.
`disimilarityCutoff`	Defaults to similarityCutoff-90. Sets the threshold for including a new peak in the alignment table to ensure new metabolites aren't just below alignment thresholds

`numCores` Number of cores used to parallelize alignment. See parallel package. Defaults to 4.

`commonIons` Provide a vector of ions to ignore from the FindProblemIons function. Defaults to empty vector.

`missingValueLimit`
Maximum fraction (Numeric between 0 and 1) of missing values acceptable for retaining a metabolite in the final alignment table. Defaults to 0.25.

`missingPeakFinderSimilarityLax`
Fraction of Similarity Cutoff to use to find missing alignments just below threshold. Set to 1 to prevent searching for missing peaks. Defaults to 0.85.

`quantMethod` Set to U, A, or T to indicate if unique mass (U), appexing masses (A), or total ion chormatograph (T) was used to quantify peak areas. Defaults to T. If "T" or "A", peaks meeting similarity thresholds will simply be summed. If "U", peaks with the same unique mass with be summed and a proportional conversion will be used before combining peaks with different unique masses.

`standardLibrary`
Defaults to NULL. Provide standard library generated from MakeReference function to ID metabolites with retention index.

Value

A list with three items: AlignmentMatix - A dataframe with peak areas for all metabolites matched in sufficient number of samples. MetaboliteInfo - An info file with RT, spectra, and metabolite ID info for each metabolite in the AlignmentMatrix. UnmatchedQuantMasses- Info on metabolites combined that had different unique masses (if quantMethod="U") or greater than 50

Examples

```
ConsensusAlign(c(system.file("extdata", "SampleA.txt", package="R2DGC"),
    system.file("extdata", "SampleB.txt", package="R2DGC")), RT1_Standards= c())
```

`FindProblemIons` *This function scans over a range of potential ions and looks for ions that are not present in any peak or are common enough to decrease alignment quality. Can use as input to the consensus align function to avoid including these ions during alignment to speed up processing time and improve alignments*

Description

This function scans over a range of potential ions and looks for ions that are not present in any peak or are common enough to decrease alignment quality. Can use as input to the consensus align function to avoid including these ions during alignment to speed up processing time and improve alignments

Usage

```
FindProblemIons(inputFile, possibleIons = c(70:600), numCores = 1,
  absentIonThreshold = 0.01, commonIonThreshold = 2, plotData = T)
```

Arguments

inputFile	The file path of a representative chromatof file to use in searching for ions to filter
possibleIons	A numeric vector of possible ions to search. Make sure each ion listed is present in the input file. Defaults to 70 through 600.
numCores	The number of cores to use for parallel processing. Defaults to 1
absentIonThreshold	Numeric indicating the fraction of total ion intensity an ion has to greater than in at least one peak to not be filtered as an absent ion. Defaults to 0.01.
commonIonThreshold	Numeric indicating the number of standard deviations below the mean an ion has to decrease the global metabolite similarity score to be filtered as a common ion. Defaults to 2.
plotData	Boolean. If true, relative ion impact scores will be plotted.

Value

Two column data frame identifying filtered ions and the reason they were filtered (absent or common). If plotData is TRUE, plots common ion scores. Y-axis is the z-scored number of pairwise metabolite comparisons with a similarity score greater than 50. X-axis is the ion with the filtered ions labeled in red.

Examples

```
FindProblemIons(inputFile=system.file("extdata", "SampleA.txt", package="R2DGC"),
    possibleIons = 70:78,plotData=FALSE)
```

MakeReference *This function takes input chromatof files from metabolite standards and parses them into a dataframe of retention time indexed standards that can be used as an input for ConsensusAlign function*

Description

This function takes input chromatof files from metabolite standards and parses them into a dataframe of retention time indexed standards that can be used as an input for ConsensusAlign function

Usage

```
MakeReference(inputFileList, RT1_Standards = NULL, RT2_Standards = NULL)
```

Arguments

inputFileList	A character vector of full file paths to the metabolite standard chromatof files to include in library.
RT1_Standards	A character vector with the name of all first retention time standards to use to index metabolites. Defaults to NULL.
RT2_Standards	A character vector with the name of all second retention time standards to use to index metabolites. Defaults to NULL.

Value

Returns a list with a retention standard indexed metabolite library that can be used in the standard library argument of the ConsensusAlign function

Examples

```
MakeReference(c(system.file("extdata", "Alanine_150226_1.txt", package="R2DGC"),
    system.file("extdata", "Serine_022715_1.txt", package="R2DGC")))
```

PrecompressFiles *This function is an optional pre-prossessing step before running consensus align to identify peaks that likely need to be combined prior to running consensus align and will perform a rough combine of these peaks depending on the quant method as an output.*

Description

This function is an optional pre-prossessing step before running consensus align to identify peaks that likely need to be combined prior to running consensus align and will perform a rough combine of these peaks depending on the quant method as an output.

Usage

```
PrecompressFiles(inputFileList, RT1Penalty = 1, RT2Penalty = 10,
  similarityCutoff = 95, numCores = 1, commonIons = c(),
  quantMethod = "T", outputFiles = F)
```

Arguments

inputFileList	A character vector with full file paths to chromatof files for processing.
RT1Penalty	A numeric indicating penalty used for first retention time differences. Defaults to 1
RT2Penalty	A numeric indicating penalty used for second retention time differences. Defaults to 100
similarityCutoff	A numeric indicating the similarity threshold (max=100) to use for declaring peaks to combine. Defaults to 95

numCores	Number of cores used to parallelize alignment. See parallel package for details. Defaults to 1
commonIons	A numeric vector of ions to exclude from alignment scores. Can provide first column of output from FindProblemIons function.
quantMethod	Character indicating the quant method used in computing peak areas on chromatof. Accepts "U", "T", or "A" for unique mass, total ion chromatograph or apexing mass. Defaults to "T". If "T" or "A", peaks meeting similarity thresholds will simply be summed. If "U", peaks with the same unique mass with be summed and a proportional conversion will be used before combining peaks with different unique masses.
outputFiles	A boolean indicating if putative peak combinations should be outputted. Will be present at the same path as the input file with _Processed.txt appended to the end.

Value

Returns a data frame with peaks recommended to be combined. If outputFiles is TRUE, peaks returned will be combined and new sample files will be written to the original directory with "_Processed.txt" added to the file name.

Examples

```
PrecompressFiles(inputFileList=system.file("extdata", "SampleA.txt", package="R2DGC"))
```

StandardLibrary_030117

Example standard library

Description

Example standard library for use in identifying metabolites. Can be used as argument for standardLibrary flag in ConsensusAlign function

Format

A dataframe with 298 rows and 14 columns

Author(s)

Ryne Ramaker <ryneramaker@gmail.com>

Index