

Survey DeNovo

Daniel Vogel

2022-10-07

This will produce a data report for use in the documentation

List the files in the PEAKS data directory

```
peaks_path <- "../data-raw/PEAKS"
print( paste0("Directory for PEAKS output: ", peaks_path ))
```

```
## [1] "Directory for PEAKS output: ../data-raw/PEAKS"
```

```
data_files<-list.files(path = peaks_path, pattern = "*.csv", full.names=TRUE)
cat(c("There are", length(data_files),"data files"))
```

```
## There are 9 data files
```

```
basename( data_files )
```

```
## [1] "denovo.milk.csv"           "denovo.neg_control.csv"
## [3] "denovo.orange-juice.csv"   "denovo.oyster.csv"
## [5] "denovo.ricin.555510.csv"   "denovo.ricin.555558.csv"
## [7] "denovo.salvenlinus_namayacush.csv" "multi_file_PEAKS_test.csv"
## [9] "Rcom_GCH4_1_2_PEAKS.csv"
```

```
# check the stats for the first file
#file1 <- check_peaks_data( data_files[1] )
```

```
# check the stats for the entire directory of .csv
myfiles <- sapply(data_files, check_peaks_columns)
myfiles
```

```
## $'../data-raw/PEAKS/denovo.milk.csv'
## [1] "Fraction"           "Source.File"         "Feature"
## [4] "Peptide"            "Scan"                "Tag.Length"
## [7] "Denovo.Score"       "ALC..."            "length"
## [10] "m.z"                "z"                   "RT"
## [13] "Predict.RT"         "Area"                "Mass"
```

```

## [16] "ppm"          "PTM"          "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.neg_control.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"
## [16] "ppm"           "PTM"           "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.orange-juice.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"
## [16] "ppm"           "PTM"           "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.oyster.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"
## [16] "ppm"           "PTM"           "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.ricin.555510.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"
## [16] "ppm"           "PTM"           "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.ricin.555558.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"
## [16] "ppm"           "PTM"           "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-raw/PEAKS/denovo.salvenlinus_namayacush.csv'
## [1] "Fraction"      "Source.File"   "Feature"
## [4] "Peptide"       "Scan"          "Tag.Length"
## [7] "Denovo.Score"  "ALC...."       "length"
## [10] "m.z"           "z"             "RT"
## [13] "Predict.RT"    "Area"          "Mass"

```

```
## [16] "ppm"          "PTM"          "local.confidence...."
## [19] "tag....0.."    "mode"
##
## $'../data-row/PEAKS/multi_file_PEAKS_test.csv'
## [1] "Fraction"      "Scan"         "Source.File"
## [4] "Peptide"       "Tag.Length"   "ALC...."
## [7] "length"        "m.z"          "z"
## [10] "RT"            "Area"         "Mass"
## [13] "ppm"          "PTM"          "local.confidence...."
## [16] "tag....0.."    "mode"
##
## $'../data-row/PEAKS/Rcom_GCH4_1_2_PEAKS.csv'
## [1] "Fraction"      "Scan"         "Source.File"
## [4] "Peptide"       "Tag.Length"   "ALC...."
## [7] "length"        "m.z"          "z"
## [10] "RT"            "Area"         "Mass"
## [13] "ppm"          "PTM"          "local.confidence...."
## [16] "tag....0.."    "mode"
```

```
# show the columns as translated by R which removed the special character like ()
myfiles2 <- sapply(data_files, check_peaks_columns_literal)
str(myfiles2 )
```

```
## List of 9
## $ ../data-row/PEAKS/denovo.milk.csv          :'data.frame': 1 obs. of  20 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Source File"
## ..$ V3 : chr "Feature"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Scan"
## ..$ V6 : chr "Tag Length"
## ..$ V7 : chr "Denovo Score"
## ..$ V8 : chr "ALC (%)"
## ..$ V9 : chr "length"
## ..$ V10: chr "m/z"
## ..$ V11: chr "z"
## ..$ V12: chr "RT"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"
## $ ../data-row/PEAKS/denovo.neg_control.csv    :'data.frame': 1 obs. of  20 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Source File"
## ..$ V3 : chr "Feature"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Scan"
## ..$ V6 : chr "Tag Length"
## ..$ V7 : chr "Denovo Score"
## ..$ V8 : chr "ALC (%)"
```

```

## ..$ V9 : chr "length"
## ..$ V10: chr "m/z"
## ..$ V11: chr "z"
## ..$ V12: chr "RT"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"
## $ ../data-raw/PEAKS/denovo.orange-juice.csv      :'data.frame': 1 obs. of  20 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Source File"
## ..$ V3 : chr "Feature"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Scan"
## ..$ V6 : chr "Tag Length"
## ..$ V7 : chr "Denovo Score"
## ..$ V8 : chr "ALC (%)"
## ..$ V9 : chr "length"
## ..$ V10: chr "m/z"
## ..$ V11: chr "z"
## ..$ V12: chr "RT"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"
## $ ../data-raw/PEAKS/denovo.oyster.csv          :'data.frame': 1 obs. of  20 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Source File"
## ..$ V3 : chr "Feature"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Scan"
## ..$ V6 : chr "Tag Length"
## ..$ V7 : chr "Denovo Score"
## ..$ V8 : chr "ALC (%)"
## ..$ V9 : chr "length"
## ..$ V10: chr "m/z"
## ..$ V11: chr "z"
## ..$ V12: chr "RT"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"

```

```

## $ ../data-raw/PEAKS/denovo.ricin.555510.csv      : 'data.frame': 1 obs. of  20 variables:
##   ..$ V1 : chr "Fraction"
##   ..$ V2 : chr "Source File"
##   ..$ V3 : chr "Feature"
##   ..$ V4 : chr "Peptide"
##   ..$ V5 : chr "Scan"
##   ..$ V6 : chr "Tag Length"
##   ..$ V7 : chr "Denovo Score"
##   ..$ V8 : chr "ALC (%)"
##   ..$ V9 : chr "length"
##   ..$ V10: chr "m/z"
##   ..$ V11: chr "z"
##   ..$ V12: chr "RT"
##   ..$ V13: chr "Predict RT"
##   ..$ V14: chr "Area"
##   ..$ V15: chr "Mass"
##   ..$ V16: chr "ppm"
##   ..$ V17: chr "PTM"
##   ..$ V18: chr "local confidence (%)"
##   ..$ V19: chr "tag (>=0%)"
##   ..$ V20: chr "mode"
## $ ../data-raw/PEAKS/denovo.ricin.555558.csv      : 'data.frame': 1 obs. of  20 variables:
##   ..$ V1 : chr "Fraction"
##   ..$ V2 : chr "Source File"
##   ..$ V3 : chr "Feature"
##   ..$ V4 : chr "Peptide"
##   ..$ V5 : chr "Scan"
##   ..$ V6 : chr "Tag Length"
##   ..$ V7 : chr "Denovo Score"
##   ..$ V8 : chr "ALC (%)"
##   ..$ V9 : chr "length"
##   ..$ V10: chr "m/z"
##   ..$ V11: chr "z"
##   ..$ V12: chr "RT"
##   ..$ V13: chr "Predict RT"
##   ..$ V14: chr "Area"
##   ..$ V15: chr "Mass"
##   ..$ V16: chr "ppm"
##   ..$ V17: chr "PTM"
##   ..$ V18: chr "local confidence (%)"
##   ..$ V19: chr "tag (>=0%)"
##   ..$ V20: chr "mode"
## $ ../data-raw/PEAKS/denovo.salvenlinus_namayacush.csv: 'data.frame': 1 obs. of  20 variables:
##   ..$ V1 : chr "Fraction"
##   ..$ V2 : chr "Source File"
##   ..$ V3 : chr "Feature"
##   ..$ V4 : chr "Peptide"
##   ..$ V5 : chr "Scan"
##   ..$ V6 : chr "Tag Length"
##   ..$ V7 : chr "Denovo Score"
##   ..$ V8 : chr "ALC (%)"
##   ..$ V9 : chr "length"
##   ..$ V10: chr "m/z"
##   ..$ V11: chr "z"

```

```

## ..$ V12: chr "RT"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"
## $ ../data-raw/PEAKS/multi_file_PEAKS_test.csv      :'data.frame': 1 obs. of 17 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Scan"
## ..$ V3 : chr "Source.File"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Tag.Length"
## ..$ V6 : chr "ALC...."
## ..$ V7 : chr "length"
## ..$ V8 : chr "m.z"
## ..$ V9 : chr "z"
## ..$ V10: chr "RT"
## ..$ V11: chr "Area"
## ..$ V12: chr "Mass"
## ..$ V13: chr "ppm"
## ..$ V14: chr "PTM"
## ..$ V15: chr "local.confidence...."
## ..$ V16: chr "tag....0.."
## ..$ V17: chr "mode"
## $ ../data-raw/PEAKS/Rcom_GCH4_1_2_PEAKS.csv      :'data.frame': 1 obs. of 17 variables:
## ..$ V1 : chr "Fraction"
## ..$ V2 : chr "Scan"
## ..$ V3 : chr "Source File"
## ..$ V4 : chr "Peptide"
## ..$ V5 : chr "Tag Length"
## ..$ V6 : chr "ALC (%)"
## ..$ V7 : chr "length"
## ..$ V8 : chr "m/z"
## ..$ V9 : chr "z"
## ..$ V10: chr "RT"
## ..$ V11: chr "Area"
## ..$ V12: chr "Mass"
## ..$ V13: chr "ppm"
## ..$ V14: chr "PTM"
## ..$ V15: chr "local confidence (%)"
## ..$ V16: chr "tag (>=0%)"
## ..$ V17: chr "mode"

```

```

# show the sha-256 checksum hash values
myfiles3 <- sapply(data_files, hash_file_sha256 )
myfiles3

```

```

## ..$ V12: chr "RT"
## "bf66d00a43e2bfcee0d17456abecc59d6842167504189bf2a8338b8916f82a8e"
## ..$ V13: chr "Predict RT"
## ..$ V14: chr "Area"
## ..$ V15: chr "Mass"
## ..$ V16: chr "ppm"
## ..$ V17: chr "PTM"
## ..$ V18: chr "local confidence (%)"
## ..$ V19: chr "tag (>=0%)"
## ..$ V20: chr "mode"

```

```
##          ../data-raw/PEAKS/denovo.orange-juice.csv
## "871ad17a7f97db9d458886005c87806fe711bde488fd87af149a144990a50c73"
##          ../data-raw/PEAKS/denovo.oyster.csv
## "eb7f3f351e2a03ca090db4798a5c0c100125f346234b7365da20f9f92dd6198c"
##          ../data-raw/PEAKS/denovo.ricin.555510.csv
## "e1549bfc1b0a4a444381504a96acb695ca0ef3a710577f85a97877895a5b821b"
##          ../data-raw/PEAKS/denovo.ricin.555558.csv
## "fa8a081f0ec2b5ae1fd369607e3ff5aa75cdda2f31c6e97442e89c148231875b"
##          ../data-raw/PEAKS/denovo.salvenlinus_namayacush.csv
## "ff11cc0bdae0966443e372fea3db8ce2261fec9d3d471995df42808b79cf6643"
##          ../data-raw/PEAKS/multi_file_PEAKS_test.csv
## "01ddb03c1619c6f9df0487df6bd4c67af58f7920a774482f0d80af2abe08b888"
##          ../data-raw/PEAKS/Rcom_GCH4_1_2_PEAKS.csv
## "3920abeb30220c91e185da76b50d859f80325f388e3de596717c5c4c30071f2b"
```