

CommonContam User Documentation

Description

CommonContam is a handy tool that identifies and removes common contaminant masses in a related set of MALDI MS spectra. It generates a list of all the masses found and then removes duplicates that are within the user defined mass tolerance range. The user then selects a cut-off value. All masses that have been found in that many files or more will be removed. A new set of cleaned data files are generated ready for database searching.

Usage

1. Place a set of mass lists from your MALDI MS data, they should all come from the same gel or off line LC into your sample data, in a folder. The files should be in text format (not MSWord Doc format) with the masses only. For Bruker MALDI files you can use the UltraMassList utility to generate the mass lists. UltraMassList is available from the Ludwig Institute for Cancer Research - University College London branch at http://www.ludwig.ucl.ac.uk/bachem_html/software.htm.
2. Start the program and select the data directory with the `Starting Directory` button.
3. Select the tolerance type and value. Depending on the calibration across your data set you may need to use a larger mass tolerance than normal. Comparisons are made to the lowest mass of a set. if the tolerance is not set high enough you will observe the counts of common masses been split in two with the second mass just outside of the tolerance value used.
4. Enter a name for the results files then press the `Process Files` button. Two tab delimited results files are generated with the xls file extension. They can be opened with excel or a graphing program of your choice.
 - a. The file named `_graph.xls` has three columns of data. Column 1 contains the original file path and name, column 2 has the masses column 3 the file index number assigned by commonContam. Plot column 2 vs. column 3 in the XY (`Scatter`) chart type (figure1).

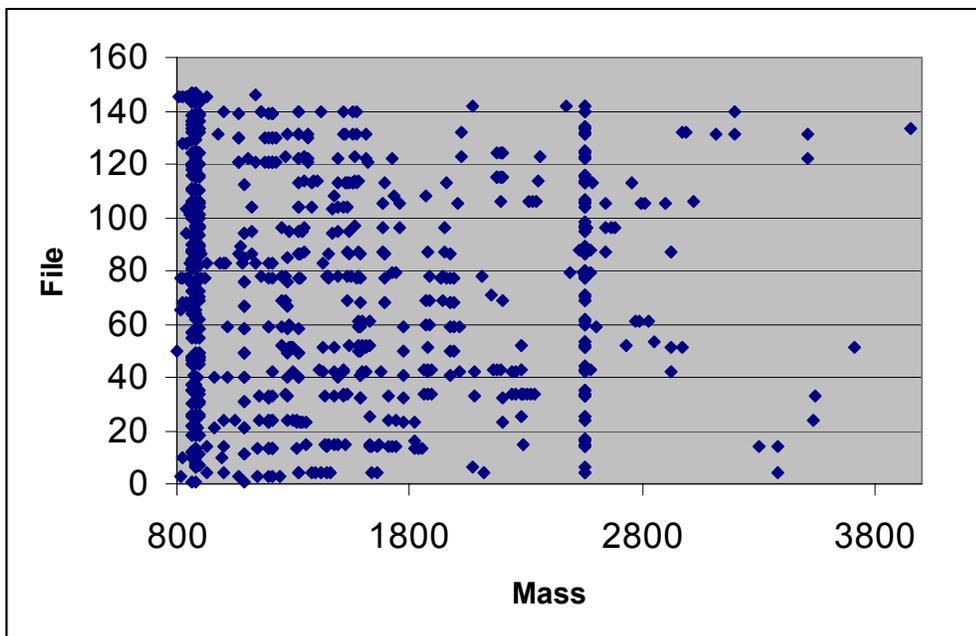


Figure 1. Mass vs file index number for fractions from a RPHPLC separation. Major contaminant peaks can be seen as vertical lines in the graph.

- b. The file named `_count.xls` can also be plotted. Either as a XY scatter plot or as a bar chart. Figure 2 shows a bar chart of the masses vs file count.

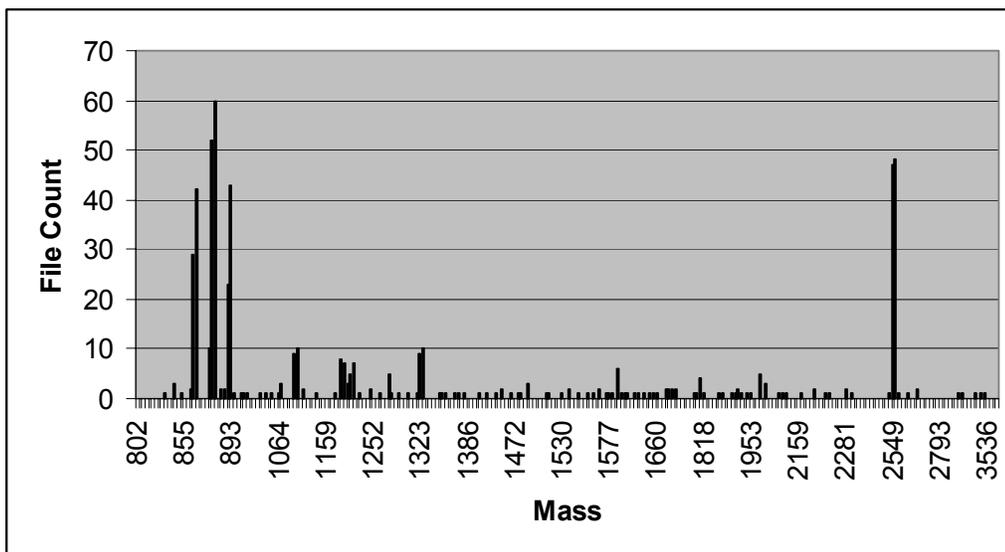


Figure 2. Mass vs File count. The mass axis is not to scale. There are four major contaminant peaks at the nominal masses 861, 877, 892, 2549 along with a number of masses that occur in around 10 of the 150 files.

5. After reviewing the two graphs enter a cut-off value. The maximum file count is displayed after a / see figure 3 ringed area. All masses that occur in more than the cut-off value number of files will be removed. Press the `remove` button to generate cleaned files. The cleaned files have the same root name but with the suffix `_cleaned.txt`.
6. The program will need to be restarted to process a new data set.



Figure 3. The commonContam GUI with the cut of value set to 5 and the highest count for a mass ringed in red.

Bugs

1. The browse for folder window has an array reference at the top. This is purely cosmetic and unfortunately there is no fix as yet.
2. The program should reset itself after a new data directory has been set so that multiple directories can be analysed without having to restart the program.
3. The mmu tolerance type has not been implemented.

Feature requests

1. Show the total number of files found as well as the maximum file count for the most frequent mass.
2. Display the two graphs in a window in the program.

If you find a bug or have a feature request write me an email r.jacob@ucl.ac.uk.