**Running Amalgamator user guide**

**Purpose**

This guide details how to use the Amalgamator module to combine files from positive and negative runs into a single file with duplicates removed ready for putative identification using the WebSearch module.

Duplicates are identified by comparing the negative file with the positive file within a small retention time tolerance and a corrected m/z tolerance (negative m/z + 2H+, followed by negative m/z + H+ +CH3+ for phosphotidylcholine and sphingomyelins with phosphocholine head group). Any hits are classed as a match.

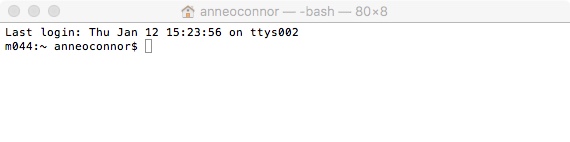
Where a match is found after correcting for charge the frame with the highest average intensity is retained and the other discarded. There is an option to add the intensity of the discarded matching ion to that of the retained ion.

**Prerequisites**

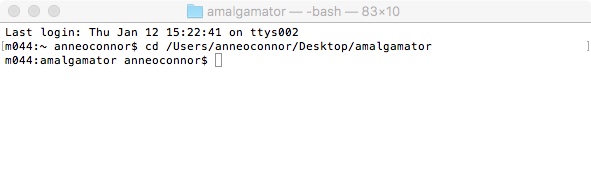
1. Python 2.7 minimum with specific libraries installed as per Python installation guide.
2. A positive and a negative mode output csv file that have been processed with PeakFilter. The original pre-processing for these files can either be SIEVE or XCMS.

**Running Amalgamator using Mac OSX/Linux**

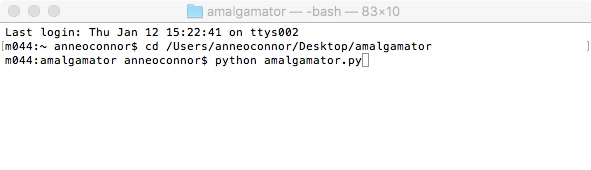
1. Open Terminal



1. Navigate to the folder where the Amalgamator code is.



1. Type in the console – “python amalgamator.py” (this is case sensitive) – This will start the Amalgamator program and import all dependant libraries.



1. Amalgamator will prompt the user to enter the number of replicate sample columns in the PeakFilter processed files. A default will be suggested from the current parameters.csv file. This number entered will be the same for both files.



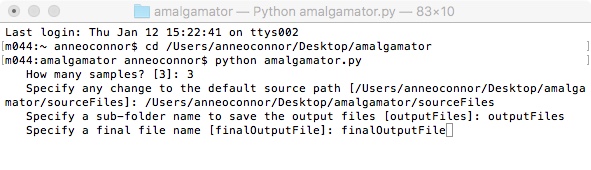
1. The default file location is ‘PeakFilter/sourceFiles’. If there are multiple input files they must reside in the same folder. To change the default path insert the full path to the input file folder here.



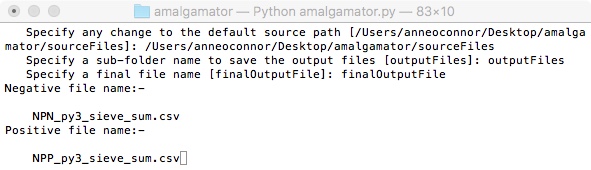
1. Two output files are created by Amalgamator. These files, by default are stored in a sub folder of the PeakFilter folder in a folder named ‘outputFiles’ (PeakFilter/outputFiles). Here the user can specify a different sub-folder to save the Amalgamator output, if it does not exist it will be created provided the naming is legal.



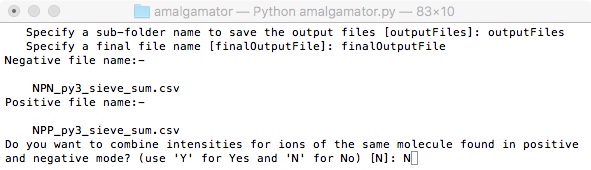
1. By default, the two files generated during the Amalgamator are prefixed with the name ‘finalOutputfFile’. At this point the user has the option to change this default prefix to something more meaningful.



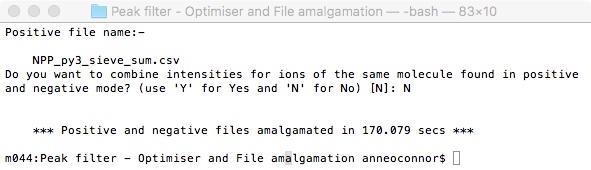
1. Amalgamator will now prompt for the two input files. These are the two PeakFilter summary files; negative first, positive second.



1. Amalgamator can add the intensity of the discarded matched ion to the retained, the user is now prompted to choose if this is required. Default is not to add.

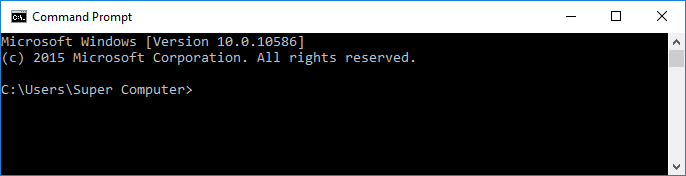


1. Amalgamator will now run to completion. Two output files will be created; one will be the combined positive and negative file with any duplicates removed, the other will list duplicates and provide the corresponding record from each file.

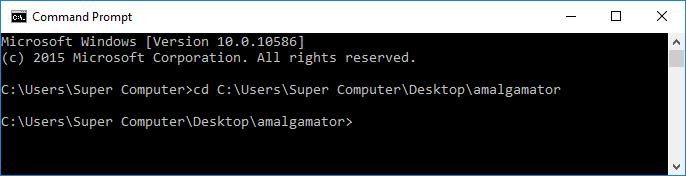


**Running Amalgamator using Windows**

1. Open Command prompt.



1. Navigate to the folder where the Amalgamator code is.



1. Type in the console – “python amalgamator.py” (this is case sensitive) – This will start the Amalgamator program and import all dependant libraries. Follow the instructions from section 4 under ‘Running Amalgamator using Mac OSX/Linux above.

**Output file naming**

The name of each output file is made up of 3 parts, a prefix, which is the file name as entered at stage 10 above, file type (results or matches) and a common suffix indicating the column type, polarity, year, month, date, hour, minute and second. For example:

my\_Data\_matches\_20151207-160538.csv

(file name) (stage) (year, month, date) (hour, minute, second)