

# **ADAP**

## **User Manual**

Version 3.1.0

Du-lab Team

Department of Bioinformatics and Genomics

University of North Carolina at Charlotte

dulab.binf@gmail.com

<http://www.du-lab.org>

February 20, 2017

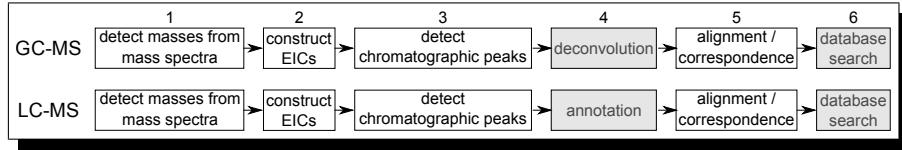
## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Download and Installation</b>	<b>3</b>
<b>3</b>	<b>ADAP-LC</b>	<b>4</b>
3.1	Detection of Masses from Mass Spectra . . . . .	4
3.2	Construction of Extracted Ion Chromatograms . . . . .	7
3.3	Detection of Peaks from EICs . . . . .	10
3.4	Annotation of EIC Peaks Using CAMERA . . . . .	12
3.5	Alignment . . . . .	16
3.6	Database Search . . . . .	19
3.7	Results Export . . . . .	23
<b>4</b>	<b>ADAP-GC</b>	<b>25</b>
4.1	Detection of Masses and Construction of EICs . . . . .	25
4.2	Detection of Peaks from EICs . . . . .	26
4.3	Deconvolution . . . . .	27
4.4	Alignment . . . . .	30
4.5	Spectra Export . . . . .	31
<b>5</b>	<b>List of Additions and Changes Du-lab Team Made to MZmine 2</b>	<b>34</b>

## 1 Introduction

ADAP (Automated Data Analysis Pipeline) was developed for pre-processing untargeted mass spectrometry-based metabolomics data. It consists of two components: ADAP-GC and ADAP-LC for pre-processing GC-MS and LC-MS data, respectively. Figure 1 depicts the workflows of the two pipelines. The two pipelines share modules 1, 2, and 5. The differences between the two pipelines lie in modules 3, 4, and 6. Deconvolution is unique to ADAP-GC while peak annotation is unique to ADAP-LC. In addition, the database search module for ADAP-GC searches against EI-GC-MS spectra while the database search module for ADAP-LC searches against the PubChem database. Compound identification in ADAP-GC is achieved by comparing spectral similarity while compound identification in ADAP-LC is achieved by comparing experimental masses and isotopic distributions against exact masses and theoretical isotopic distributions.

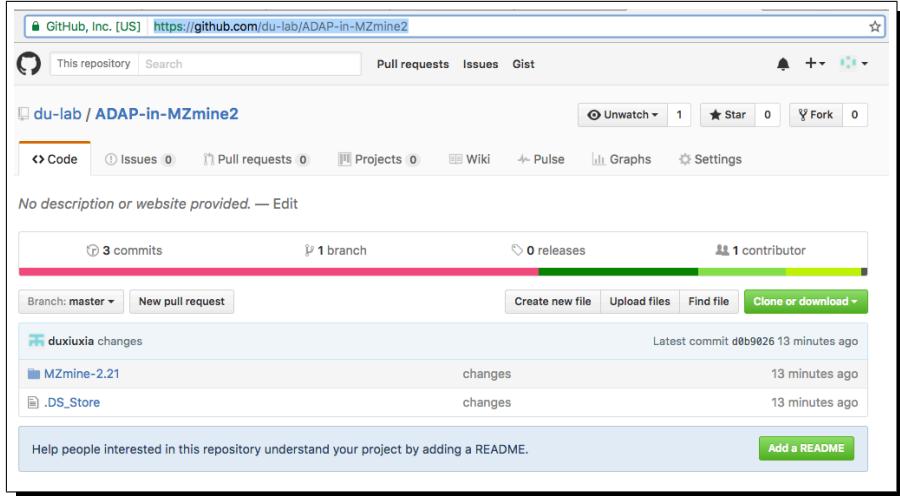
All of the computing modules have been implemented in Java or C++ and incorporated into the framework of MZmine 2. Next we describe how to use ADAP-GC and ADAP-LC. For other capabilities of MZmine 2, please refer to the MZmine 2 website [1].



**Figure 1:** Workflows for pre-processing GC- and LC-MS data.

## 2 Download and Installation

The software currently runs on Mac OS X only. The Du-lab team is working on making it available for Windows and Linux operating systems. The software can be downloaded from Github at <https://github.com/du-lab/ADAP-in-MZmine2>. Click the green *Clone or Download* button on the right as shown in Figure 2.



**Figure 2:** Download ADAP-in-MZmine2 from Github.

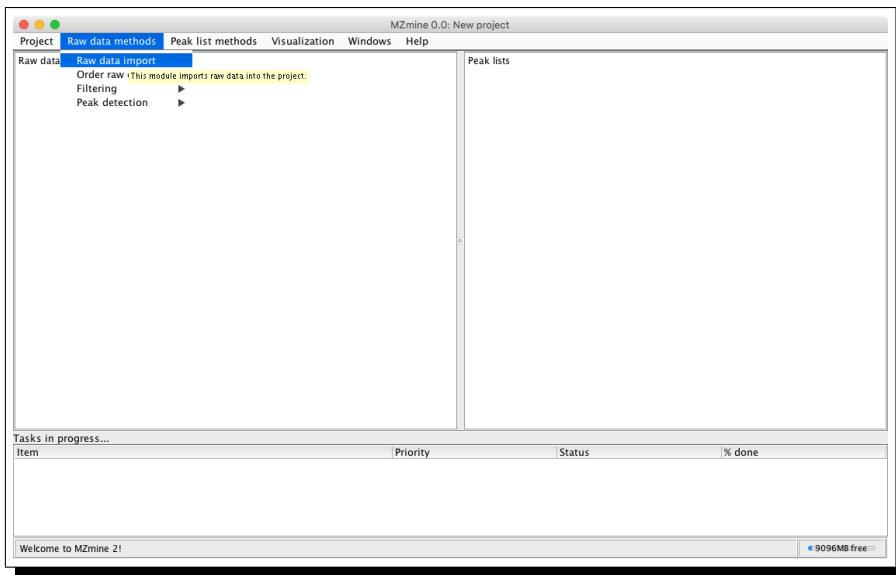
Unzip the downloaded file. Enter the folder *ADAP-in-MZmine2-master → MZmine-2.21* and then click on *startMZmine\_MacOSX.command* to start the program.

### 3 ADAP-LC

We will illustrate how to use the ADAP-LC workflow using three data files. The data is in profile mode and so we will start with detecting masses from the mass spectra, i.e. centroiding.

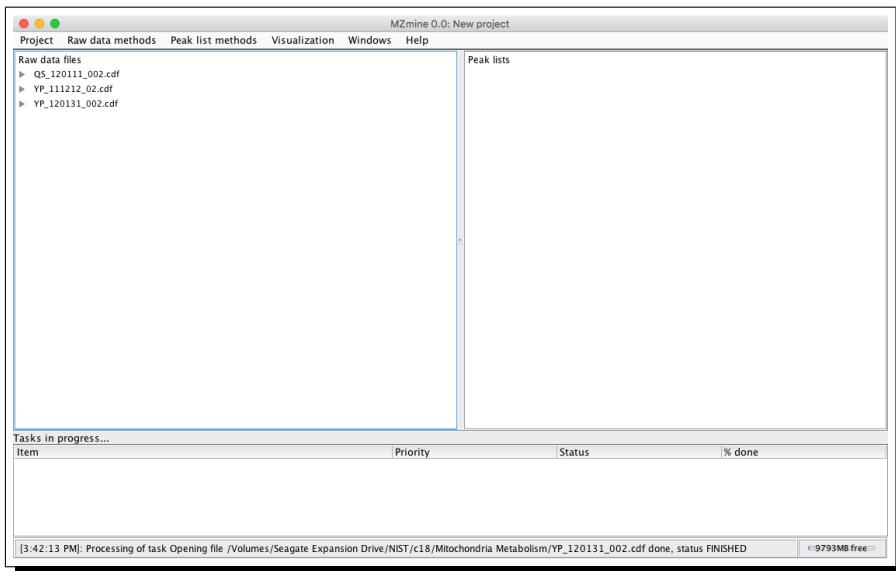
#### 3.1 Detection of Masses from Mass Spectra

Click on *Raw data methods* → *Raw data import*, shown in Figure 3.



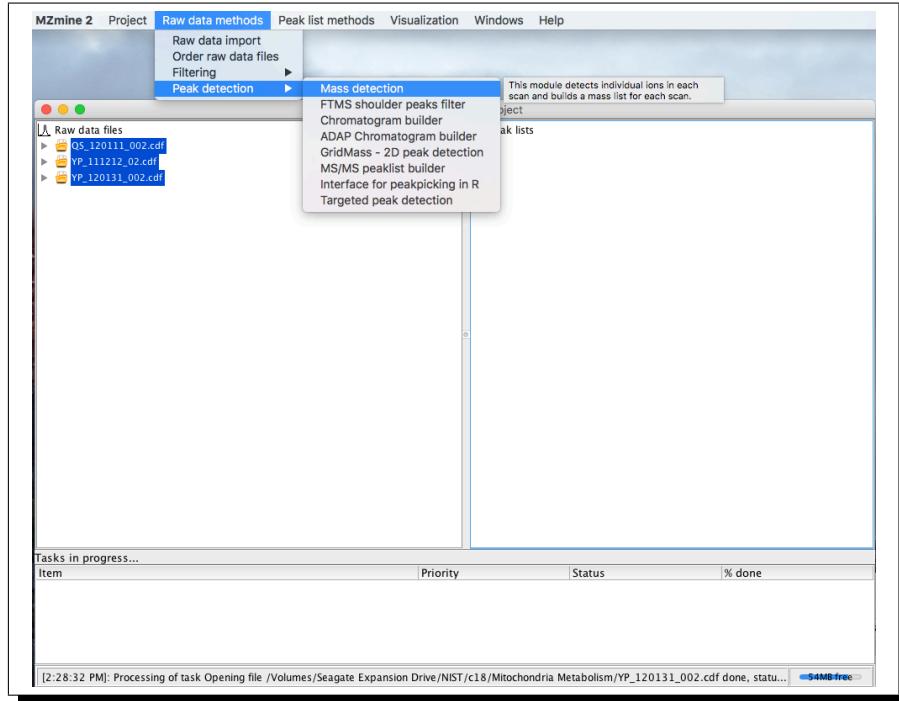
**Figure 3:** Import the raw data file.

This will open a window from which the desired data files may be chosen. The imported data files will appear in the left hand window of the GUI, labeled *Raw data files*, as shown in Figure 4.



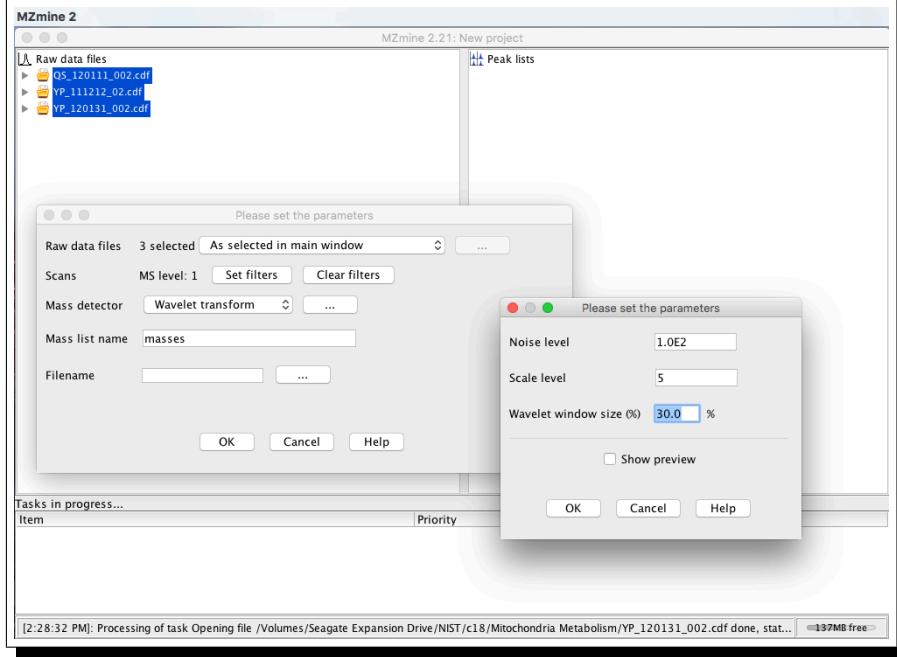
**Figure 4:** Imported data files.

To detect masses from the profile mass spectra, select the files that have been imported and then click *Raw data methods* → *Peak detection* → *Mass detection* as shown in Fig. 5.



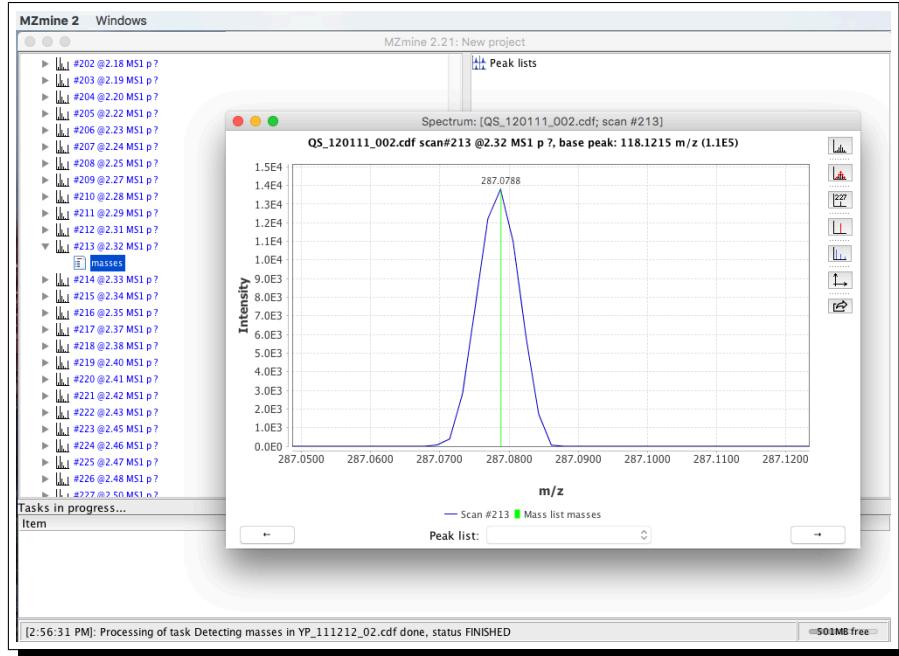
**Figure 5:** Mass detection from profile mass spectra

This will open a window with several options. From this window click on the *Mass detector* drop down box and choose *Wavelet transform*, then click on the ellipsis box directly to the right of the drop down box. The ellipsis button opens up a parameter selection window for the wavelet transform parameters. Both of these windows and the good parameters for these data files are shown in Figure 6.



**Figure 6:** Mass detection by continuous wavelet transform.

Click *OK* in both windows in Figure 6 and start the mass detection process. The process status will be shown in the bottom panel. After the process is finished, click on the triangle immediately to the left of each data file and you will see the list of the profile spectra. Then click on the triangle to the left of each profile spectrum and you will find that the centroid spectrum labelled as *masses* is shown immediately below the corresponding profile spectrum. Double click on the *masses* brings up a window displaying the profile spectra in blue and centroid masses that have been detected in green as shown in Figure 7. By stacking together the centroid spectrum and the profile spectrum, you can check how well the mass detection works.

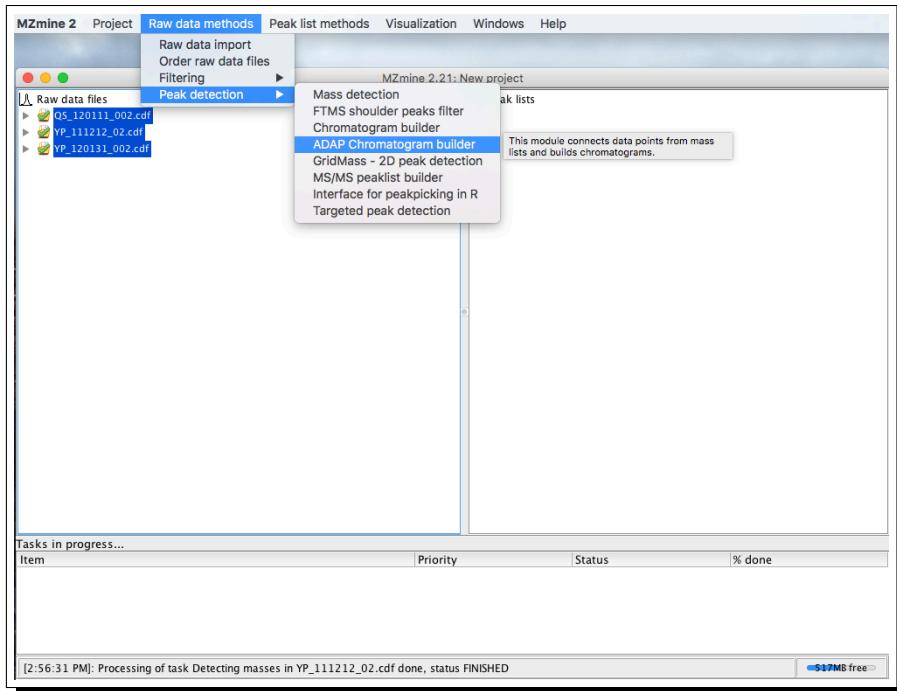


**Figure 7:** Mass detection result.

You can also use a third party software package, for example `msConvert`, for detecting masses and then import the centroid data into MZmine 2.

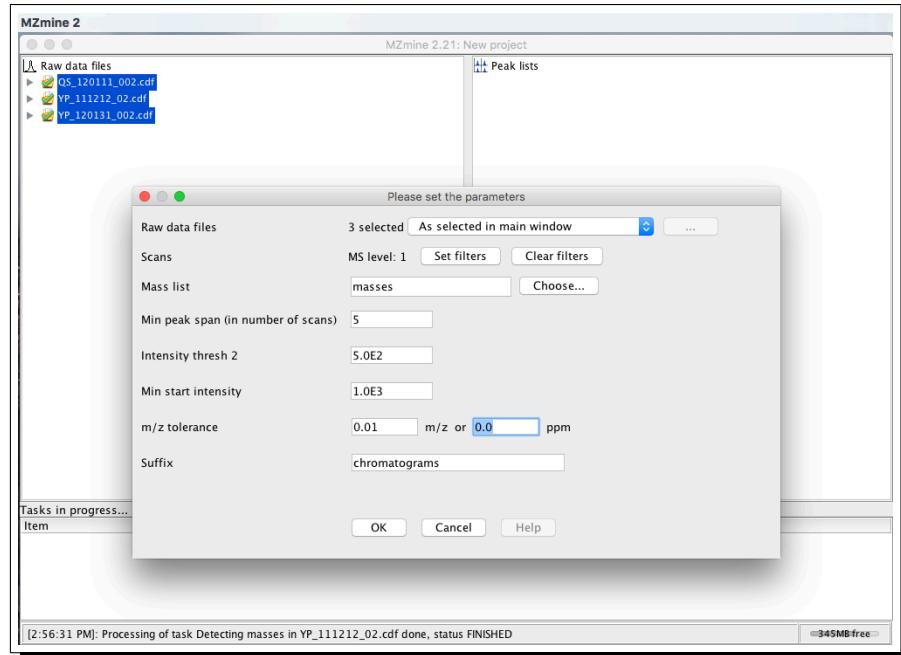
### 3.2 Construction of Extracted Ion Chromatograms

Chromatogram building builds extracted ion chromatograms (EIC) for masses that have been detected by the mass spectrometry continuously over a certain duration of time. To perform chromatogram building using the ADAP method, click *Raw data methods* → *Peak detection* → *ADAP chromatogram builder* as shown in Figure 8.



**Figure 8:** Selecting the ADAP chromatogram building.

This will pull up a window to set the parameters for the ADAP chromatogram building. The window and an example of the good parameters for the example file are shown in Figure 9.

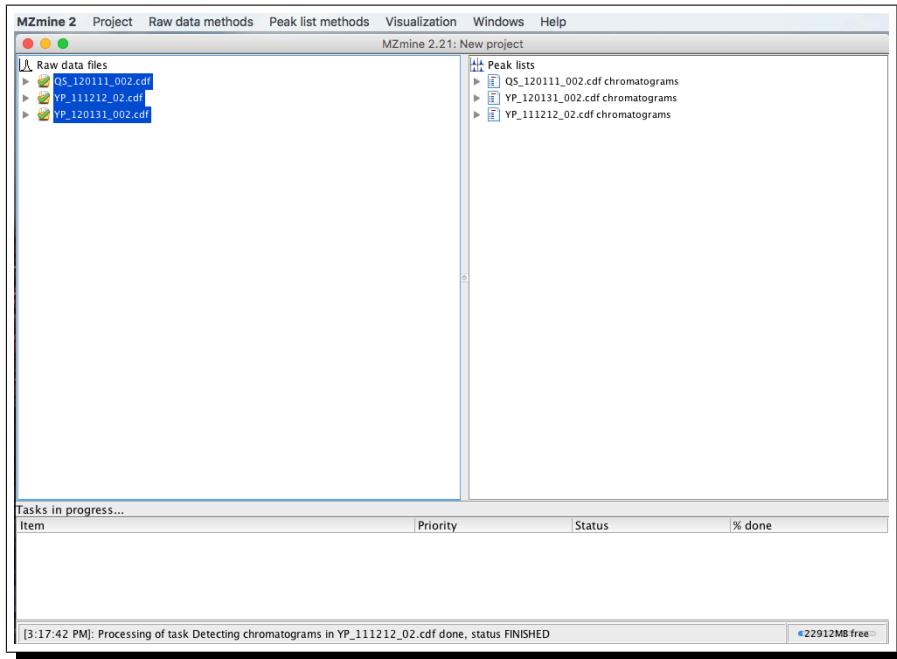


**Figure 9:** Example of ADAP chromatogram building parameters.

Description of parameters:

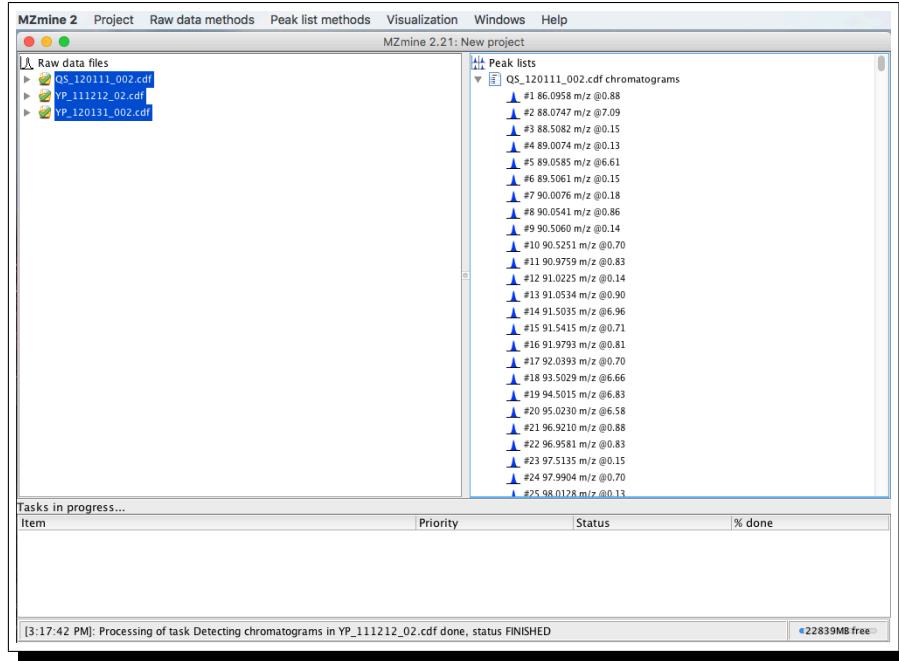
- *Min peak span*: In the entire chromatogram there must be at least this number of sequential scans having points above the *Intensity thresh 2* set by the user.
- *Intensity thresh 2*: See above.
- *Min start intensity*: There must be at least one point in the chromatogram that has an intensity greater than or equal to this value.
- *m/z tolerance*: The tolerance used in constructing the chromatogram to determine if points belong to the chromatogram or not. Twice the *m/z tolerance* set by the user is the maximum width of a mass trace. We strongly recommend setting the *m/z* value and **not** the ppm value. Whichever value is set to 0.0 will not be used.

Clock *OK* starts the chromatogram building process. After the process is complete, a list of chromatograms will appear in the right hand window of the GUI labeled *Peak Lists* as shown in Figure 10.



**Figure 10:** Results of chromatogram building.

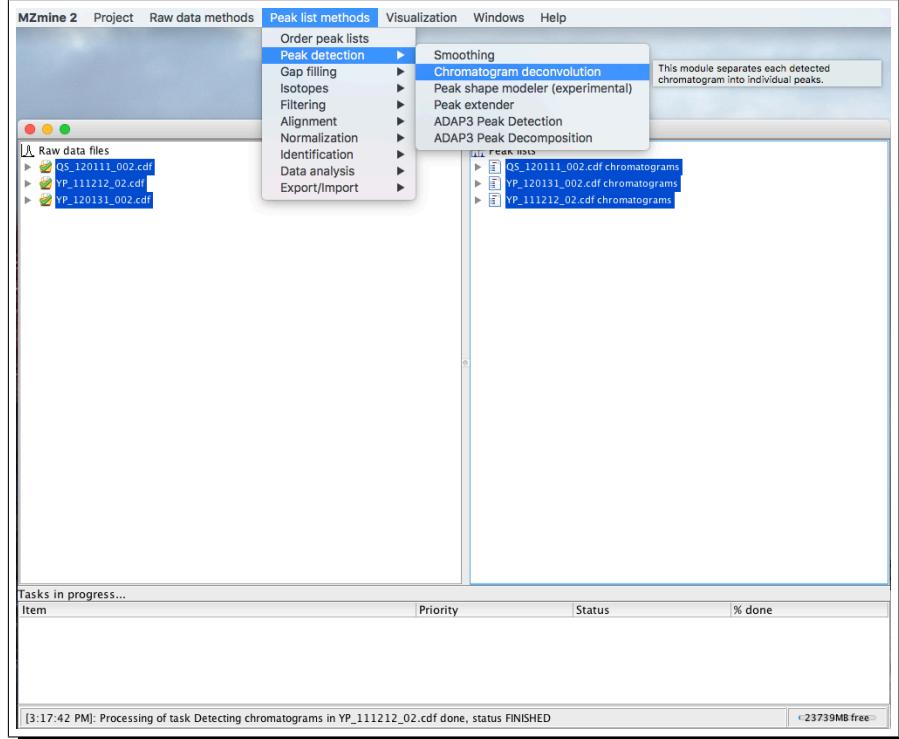
Click the triangle to the left of each data file expands the list of EICs as shown in Figure 11.



**Figure 11:** List of EICs that have been constructed.

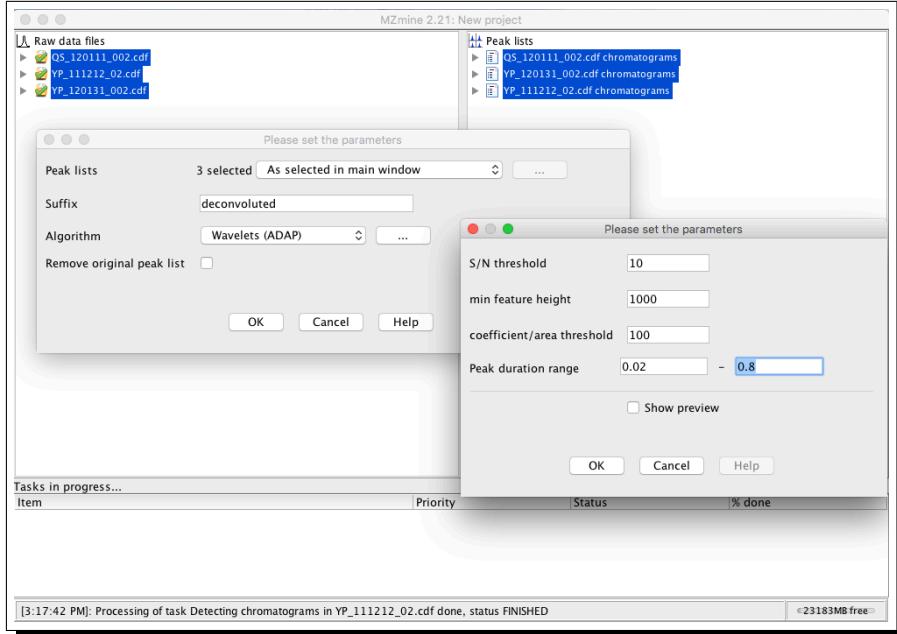
### 3.3 Detection of Peaks from EICs

Each EIC that has been constructed spans the entire duration of the chromatography. To detect the peaks from all of the EICs, select the EICs and click *Peak list methods* → *Peak detection* → *Chromatogram deconvolution* as shown in Figure 12.



**Figure 12:** Detect peaks from EICs.

This will open a window with a drop down box for selecting the peak detection method. From the drop down box choose the *Wavelets (ADAP)* option and then click on the ellipsis box/button next to the drop down box. The ellipsis button will open a window for setting the peak detection parameters. Both windows are shown in Fig. 13.

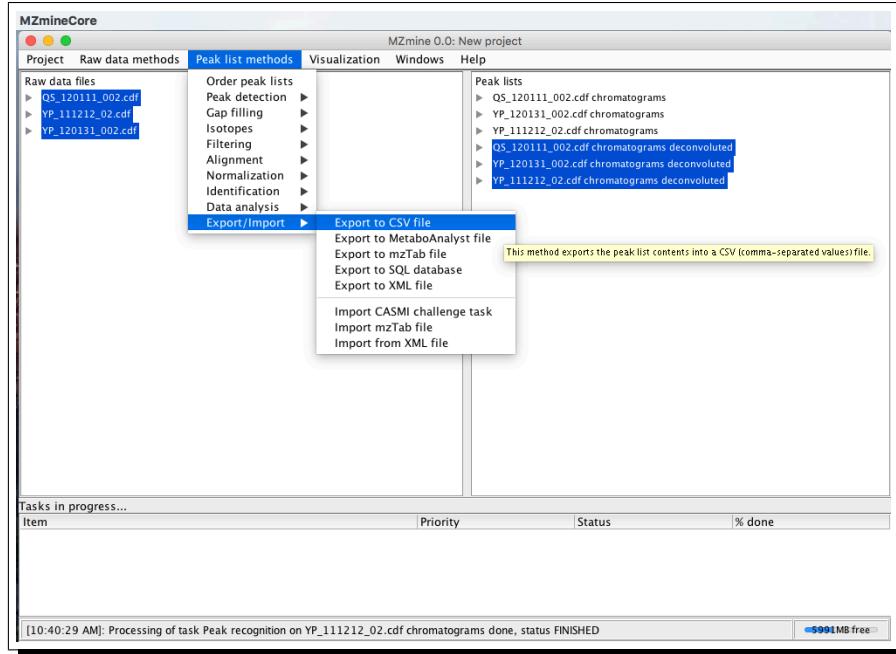


**Figure 13:** EIC peak detection parameters.

#### Description of parameters:

- *S/N threshold*: The minimum signal to noise ratio a peak must have to be considered a real feature. Values greater than or equal to 7 will work well and will only detect a very small number of false positive peaks.
- *min feature height*: The smallest intensity a peak can have and be considered a real feature.
- *coefficient/area threshold*: This number must be chosen by looking at examples using the *show preview button* at the bottom of the window. This is the best coefficient found by taking the inner product of the wavelet at the best scale and the peak, and then dividing by the area under the peak. Values around 100 work well for most data.

After the detection of chromatographic peaks is complete, a list of chromatographic peaks will appear below the list of chromatograms in the *Peak lists* window for each data file. Each list of peaks can be exported, separately, by selecting the peaks detected from one data file and clicking on *Peak list methods*, mousing over the *Export/Import* option and then selecting the desired export method (Figure 14). Figure 15 shows a sample export of the chromatographic peak detection results.



**Figure 14:** Export results from chromatographic peak detection.

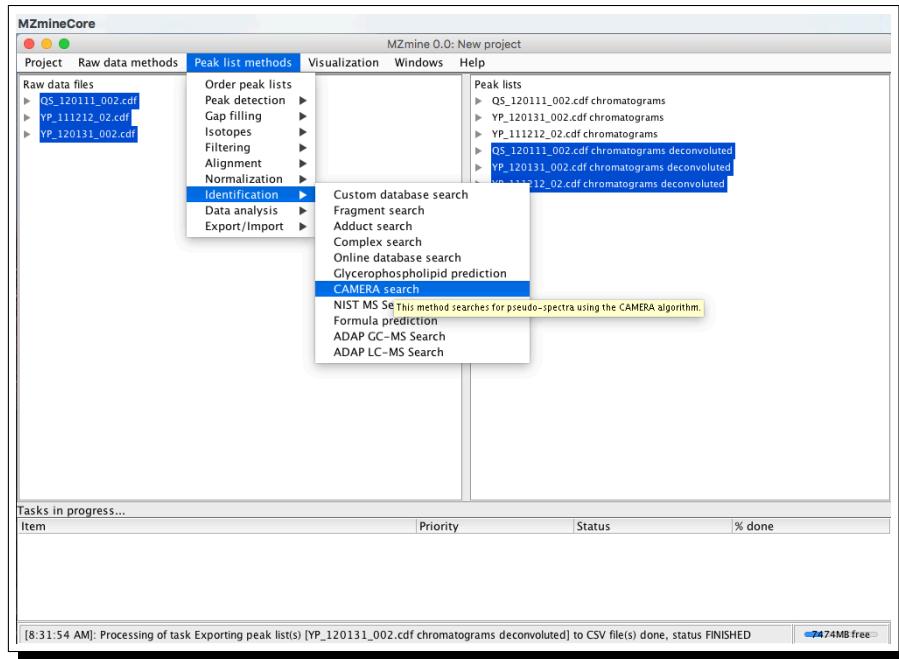
A	B	C	D	E	F	G	H	I	J	K	L
row ID	row m/z	row retention time	row comment	row number	All identity	Y_P_120131_1	Y_P_120131_1	Y_P_120131_1	Y_P_120131_1	Y_P_120131_1	Y_P_120131_1
1	86.0957031	0.90133833		1	DETECTED	86.0957031	0.90133833	0.84557	1.12000333	0.27441	
2	90.0542068	0.90133833		1	DETECTED	90.0542068	0.90133833	0.87351833	0.957405	0.08381	
3	90.5253906	0.718745		1	DETECTED	90.5253906	0.718745	0.63083	0.78977167	0.15891	
4	90.9760208	0.81773667		1	DETECTED	90.9760208	0.81773667	0.73311833	0.91526833	0.11	
5	91.0270233	0.73311833		1	DETECTED	91.0270233	0.73311833	0.66087167	0.74731167	0.01	
6	91.0270233	6.80784		1	DETECTED	91.0270233	6.80784	6.77885	6.80784	0.01	
7	91.0534515	0.92924667		1	DETECTED	91.0534515	0.92924667	0.91526833	1.02967	0.11441	
8	91.9795227	0.83164833		1	DETECTED	91.9795227	0.83164833	0.74731167	0.87351833	0.12621	
9	92.5217133	0.73311833		1	DETECTED	92.5217133	0.73311833	0.69002333	0.74731167	0.05721	
10	94.0445557	0.73311833		1	DETECTED	94.0445557	0.73311833	0.76550333	0.76158333	0.01	
11	94.0445557	4.37849833		1	DETECTED	94.0445557	4.37849833	4.37849833	4.40693167	0.02841	
12	96.9212341	0.87351833		1	DETECTED	96.9212341	0.87351833	0.84557	1.04465	0.11	
13	97.5137787	4.37849833		1	DETECTED	97.5137787	4.37849833	4.33529167	4.435335	0.10001	
14	97.9908905	0.718745		1	DETECTED	97.9908905	0.718745	0.66087167	0.77578	0.11491	
15	98.9181213	0.85954667		1	DETECTED	98.9181213	0.85954667	0.83164833	1.00039167	0.16871	
16	99.0545044	0.957405		1	DETECTED	99.0545044	0.957405	0.84557	1.00039167	0.15481	
17	99.0545044	0.18226333		1	DETECTED	99.0545044	0.18226333	0.096375	0.31042833	0.21401	
18	99.5306015	0.73311833		1	DETECTED	99.5306015	0.73311833	0.64588	0.77578	0.11	
19	99.5306015	5.59725667		1	DETECTED	99.5306015	5.59725667	5.582525	5.64120333	0.05861	
20	100.028282	0.70448833		1	DETECTED	100.028282	0.70448833	0.67550333	0.73311833	0.051	
21	100.057961	0.67550333		1	DETECTED	100.057961	0.67550333	0.64588	0.76158333	0.11571	
22	100.111687	0.18226333		1	DETECTED	100.111687	0.18226333	0.096375	0.28156	0.18	
23	100.111687	2.29399167		1	DFTFCFTFD	100.111687	2.29399167	2.26317833	2.33838333	0.071	

**Figure 15:** Sample export of results from chromatographic peak detection.

### 3.4 Annotation of EIC Peaks Using CAMERA

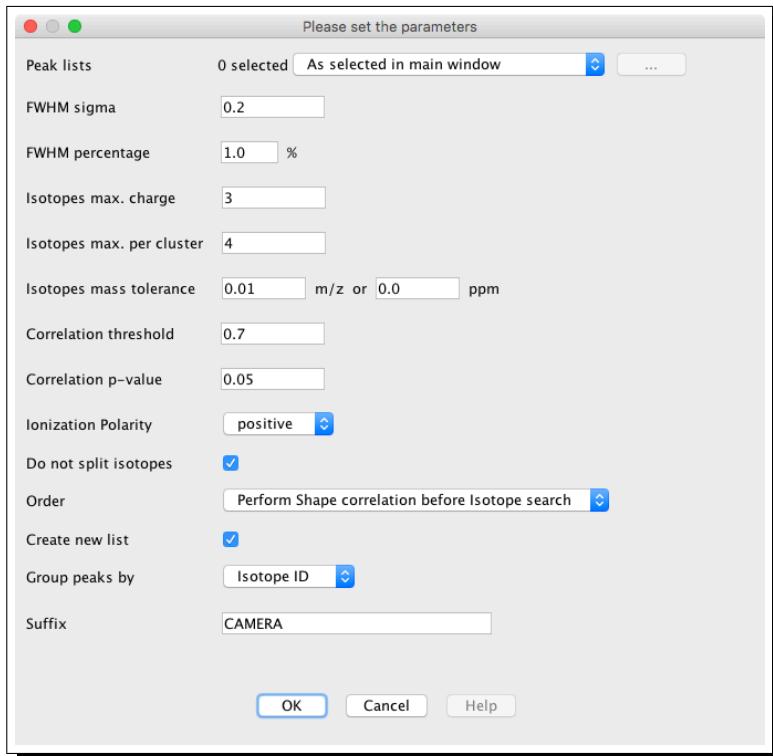
CAMERA is an R package that provides a strategy for compound spectra extraction and annotation of LC-MS datasets. It has been implemented by the MZmine 2 team into MZmine 2. The Du-lab team modified the CAMERA process slightly for extracting experimental isotopic patterns. The isotopic patterns will be used for identifying the analytes. For details about CAMERA, refer to [2, 3].

To do the annotation using CAMERA, click *Peak list methods* → *Identification* → *CAMERA search* (Figure 16).



**Figure 16:** Use CAMERA for annotation of EIC peaks.

A window will pull up as shown in Figure 17 allowing users to specify parameters.

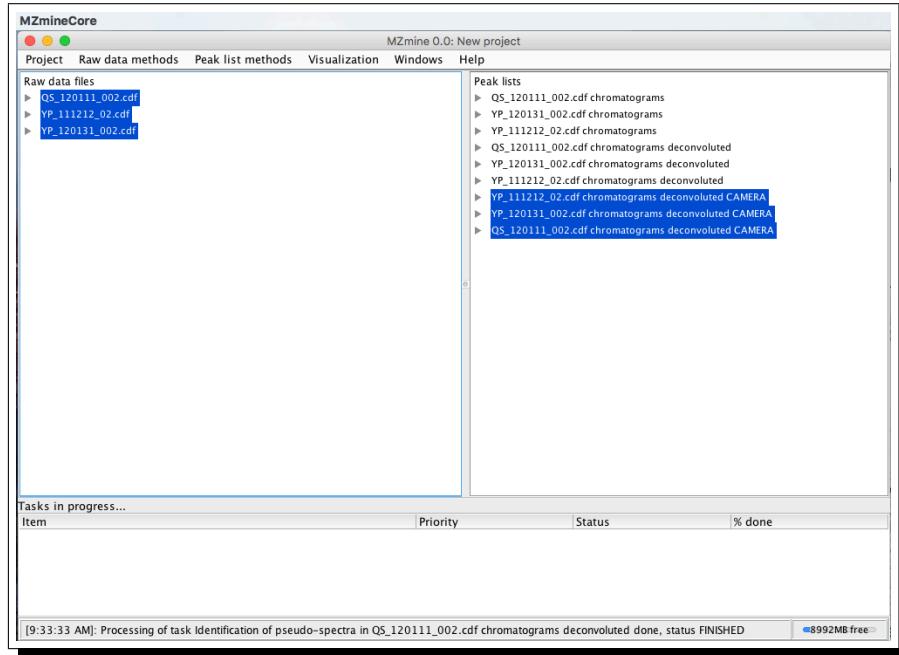


**Figure 17:** Specify parameters for CAMERA.

With the slight modification by the Du-lab team, an option (item *Order* in Figure 17) is provided to

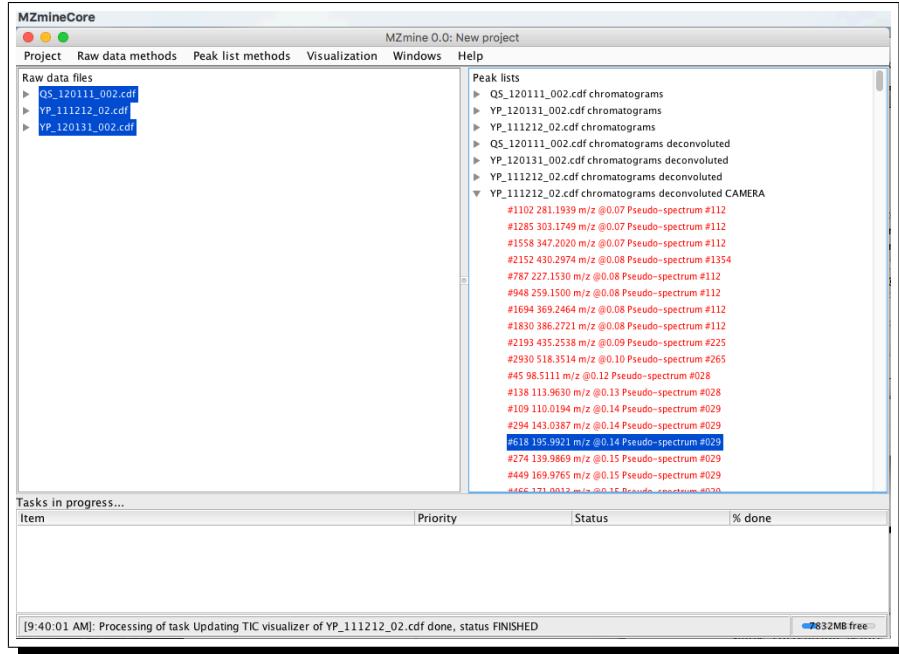
*perform shape correlation before isotope search* for stricter requirement of determining an isotopic pattern. With this stricter requirement, the mass peaks that form an isotopic pattern will have to meet not only the  $m/z$  requirement, but peak shape similarity as well. You can use the original CAMERA too by selecting *Perform Isotope search before Shape correlation*. Be aware that it could take a while for a CAMERA search to finish.

After CAMERA does finish the search, the results are displayed as shown in Figure 18.



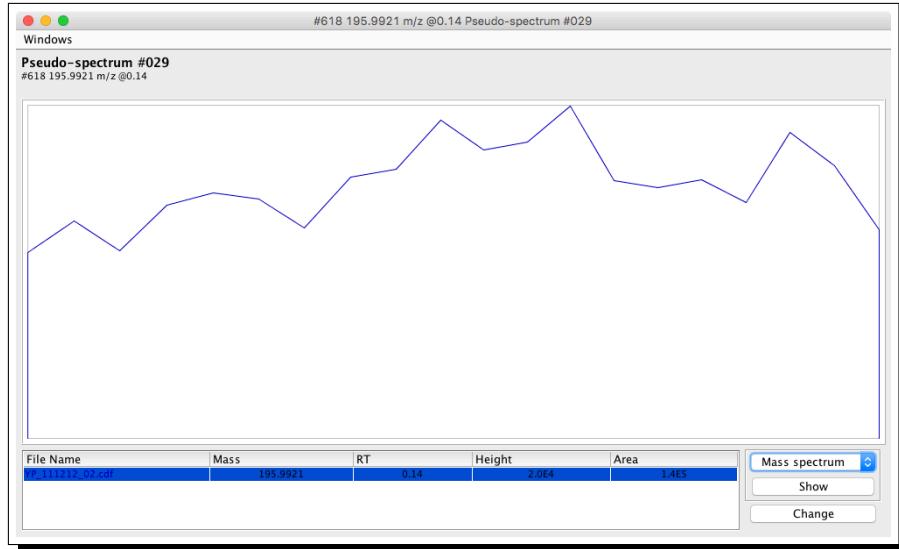
**Figure 18:** CAMERA finishes searches and results are displayed.

Click on the triangle immediately to the left of *YP\_111212\_02.cdf chromatograms deconvoluted CAMERA* will display the CAMERA search results (Figure 19) for data file *YP\_111212\_02.cdf*.



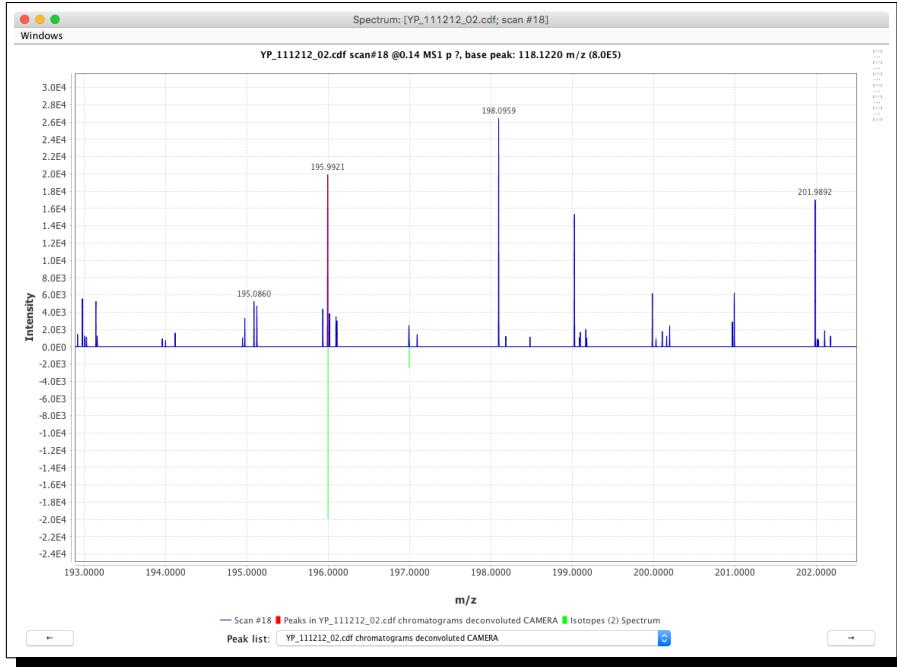
**Figure 19:** List of pseudo-spectra are displayed.

Each pseudo-spectrum can be displayed in the context of the raw spectrum. For example, to display pseudo-spectrum #029 in data file *YP\_111212\_02.cdf*, double click the pseudo-spectrum. A window will pull up as shown in Figure 20



**Figure 20:** First step of visualizing a pseudo spectrum.

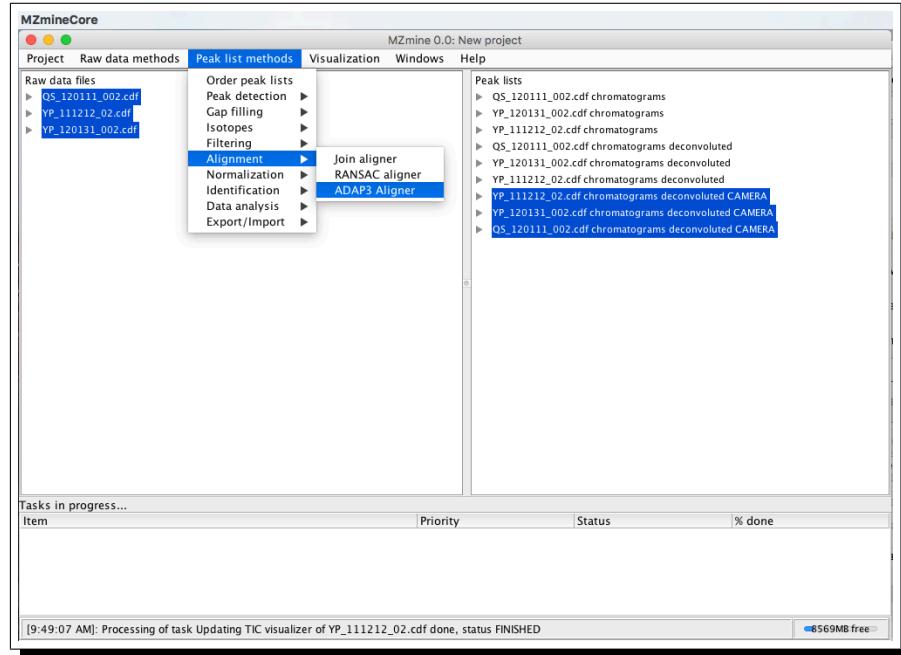
Select *Mass spectrum* in the bottom-right corner and then click on *Show* will pull up a window displaying the pseudo spectrum (green sticks) in the context of the raw spectrum (Figure 21).



**Figure 21:** Second step of visualizing a pseudo spectrum.

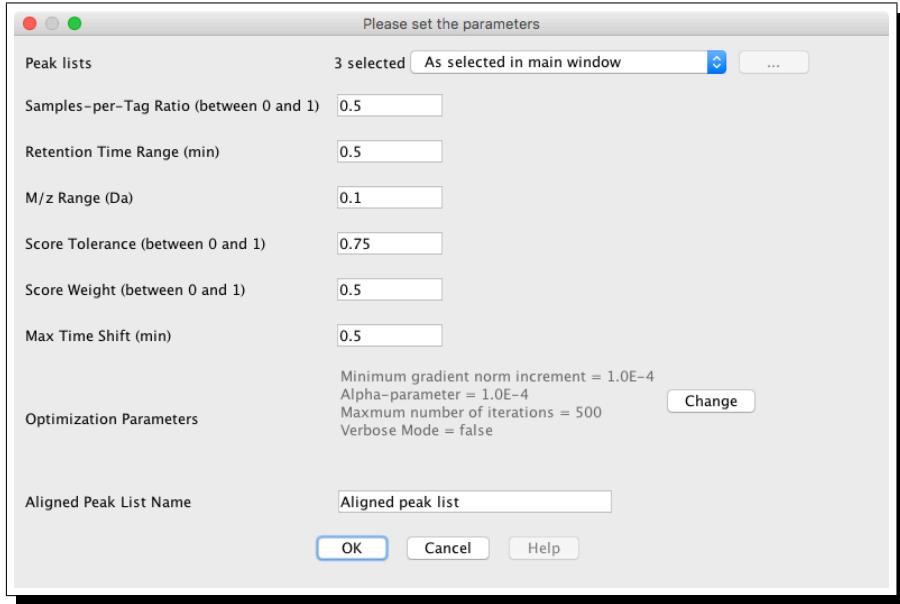
### 3.5 Alignment

Alignment corrects retention time shifts across data files. ADAP alignment algorithm is compound-based, i.e. it only aligns EIC peaks that it considers to be the same analyte. To perform alignment, click on *Peak list methods* → *Alignment* → *ADAP Aligner* as shown in Figure 22.



**Figure 22:** Alignment.

A window is pulled up as shown in Figure 23 to allow you to specify parameters for the alignment.

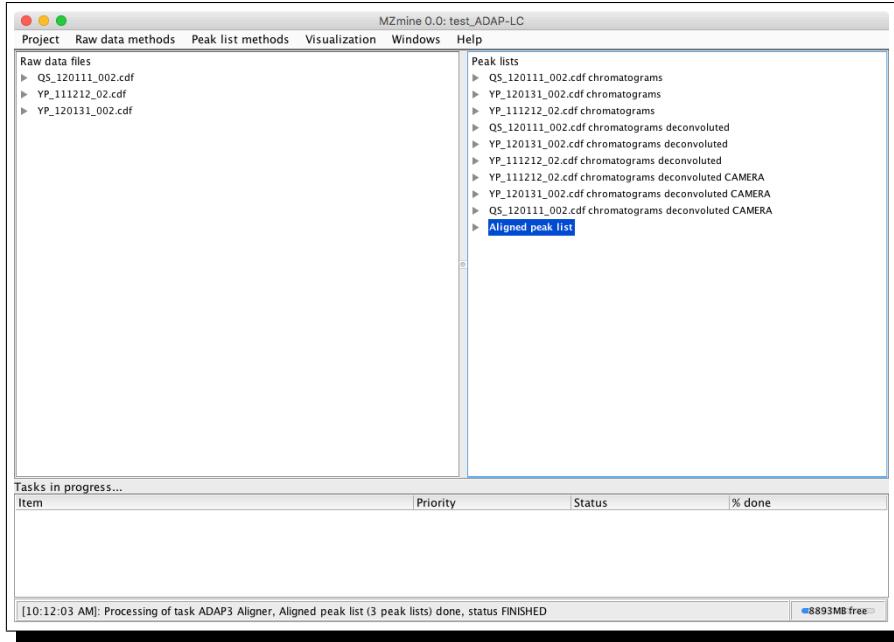


**Figure 23:** Alignment parameters.

Description of parameters:

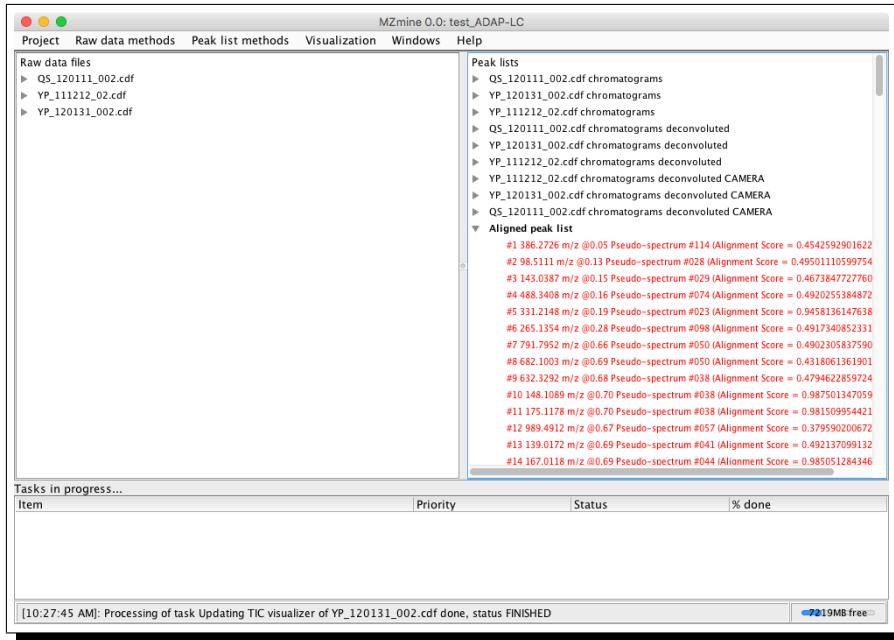
- *Samples-per-Tag Ratio*: Only components found in at least  $N$  samples will be aligned, where  $N$  is the total number of samples times the specified ratio.
- *Retention Time Range*: Only components within this range will be aligned.
- *m/z Range*: Only components within the specified *m/z* range will be aligned. Notice that, depending on the deconvolution and annotation algorithms, the same components might have completely different *m/z* values assigned to them in different samples. In this case, it is possible to set *m/z* range to some large value (e.g. 1000 Da).
- *Score Tolerance*: In order to perform alignment, a similarity score is calculated for each pair of components. The components with the similarity score above the tolerance will be aligned.
- *Score Weight*: The similarity score used in alignment is a linear combination of the EIC shape similarity and the spectral similarity with weights  $w$  and  $1 - w$  respectively, where  $w$  is the specified *Score Weight*.
- *Max Time Shift*: The maximum allowed time shift of one component relative to another.

After alignment is finished, the aligned peak list is shown as *Aligned peak list* in the *Peak lists* panel (Figure 24).



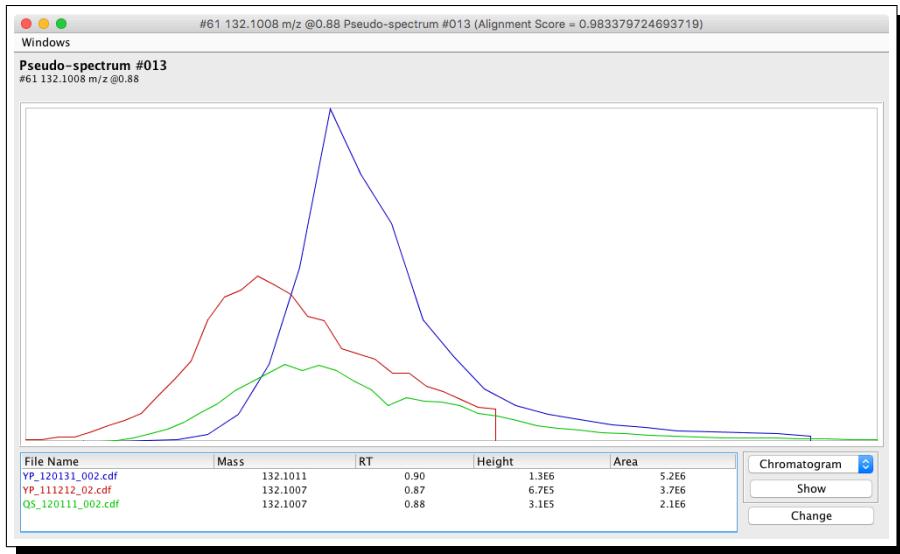
**Figure 24:** Aligned peak list.

Expand the *Aligned peak list* to see the list of peaks (Figure 25) that have been aligned across the three data files.



**Figure 25:** List of aligned peaks.

Double click on a particular peak pulls up a window (Figure 26) displaying the EICs for this peaks in the three data files.

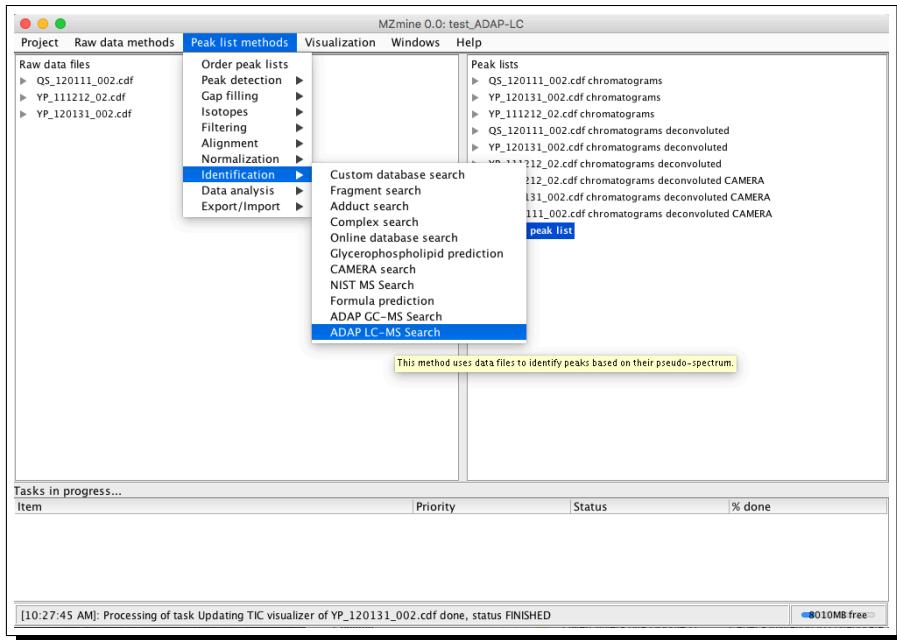


**Figure 26:** EICs of the peaks in three data files that are considered the same analyte by the alignment algorithm.

### 3.6 Database Search

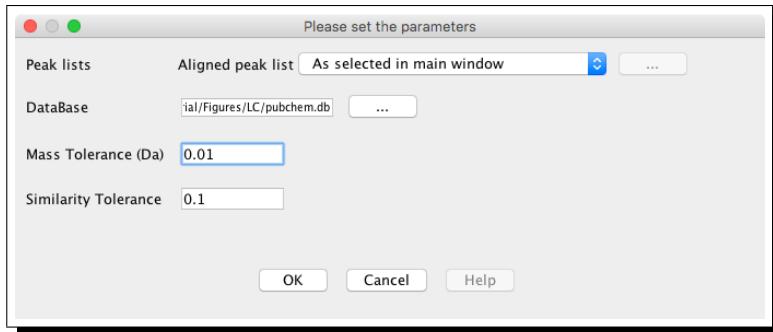
Analytes are identified by first searching the experimental monoisotopic mass against the masses of compounds in PubChem using a user-specified mass tolerance. All of the compound candidates are then scored based on the similarity between the experimental isotopic pattern and the theoretical isotopic pattern. To speed up the search, all of the compound entries (more than 9 million) in PubChem have been pre-downloaded and unique chemical formula have been extracted and stored as a database file. The total number of unique chemical formula is about 3 million.

To search against this database, click *Peak list methods* → *Identification* → *ADAP LC-MS Search* as shown in Figure 27.



**Figure 27:** Identification of analytes by searching compound databases.

A window pulls up as shown in Figure 28. Click on the ellipsis to select the database file and specify the mass tolerance and similarity threshold between experimental and theoretical isotopic pattern.

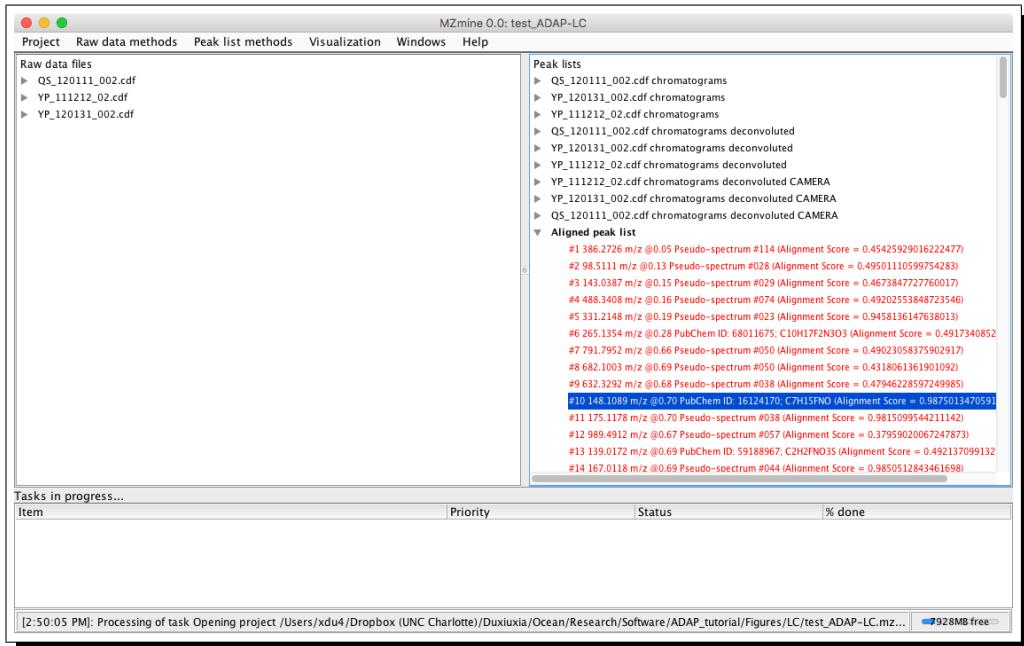


**Figure 28:** Parameters for identification of analytes by searching compound databases.

Description of parameters:

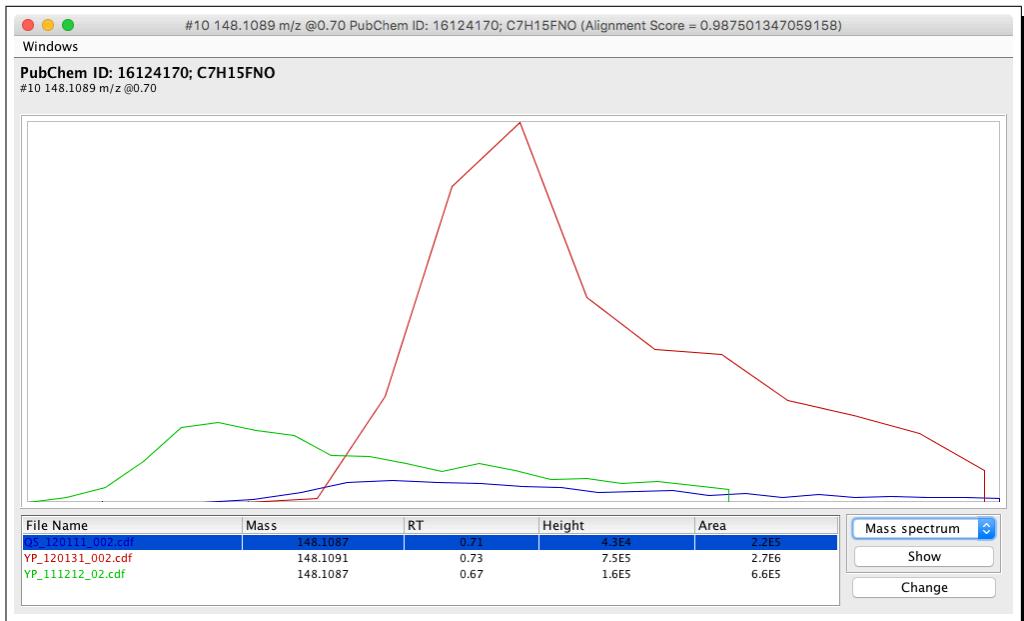
- *DataBase*: Path to a database file containing a table with chemical formulas and the corresponding molecular weights.
- *Mass Tolerance*: Experimental monoisotopic mass is matched to the library values within this tolerance.
- *Similarity Tolerance*: An experimental pseudo-spectrum or isotopic pattern is matched to a database of such by calculating the spectral similarity. Any match that is within this tolerance will be considered good matches. 0 corresponds to a high similarity and 1 corresponds to a poor similarity.

After the database search is complete, peaks that have matches in the database will have the corresponding PubChem IDs listed as shown in Figure 29.



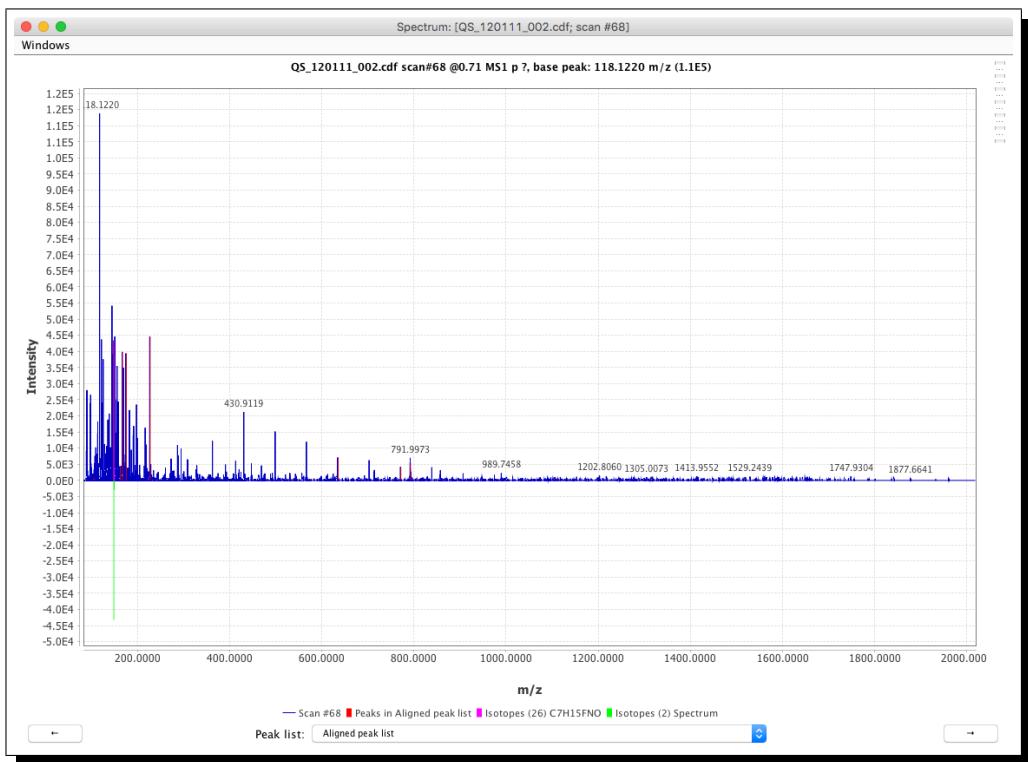
**Figure 29:** Results of searching PubChem.

Double click on a peak that have matching PubChem IDs pulls up a window as shown in Figure 30. This figure shows that this peak was observed in the three data files. Select any one of the three data files and then select *Mass spectrum* in the bottom-right drop-down menu.



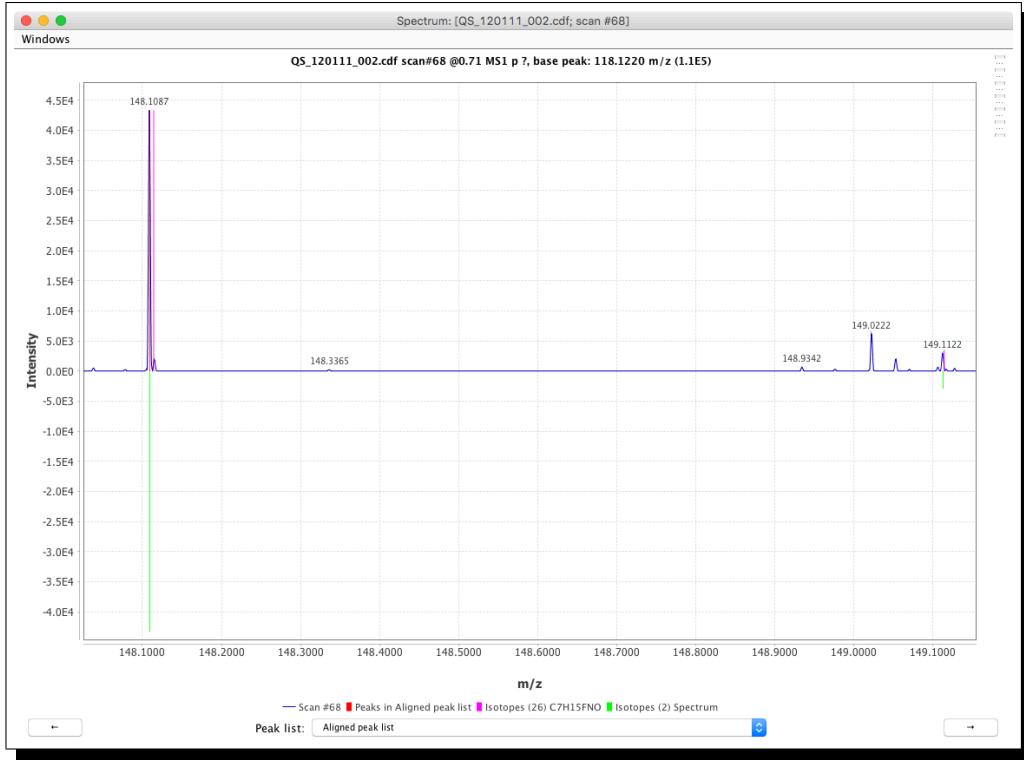
**Figure 30:** Results of searching PubChem.

Click *Show* will pull up a spectrum window as shown in Figure 31.



**Figure 31:** Results of searching PubChem.

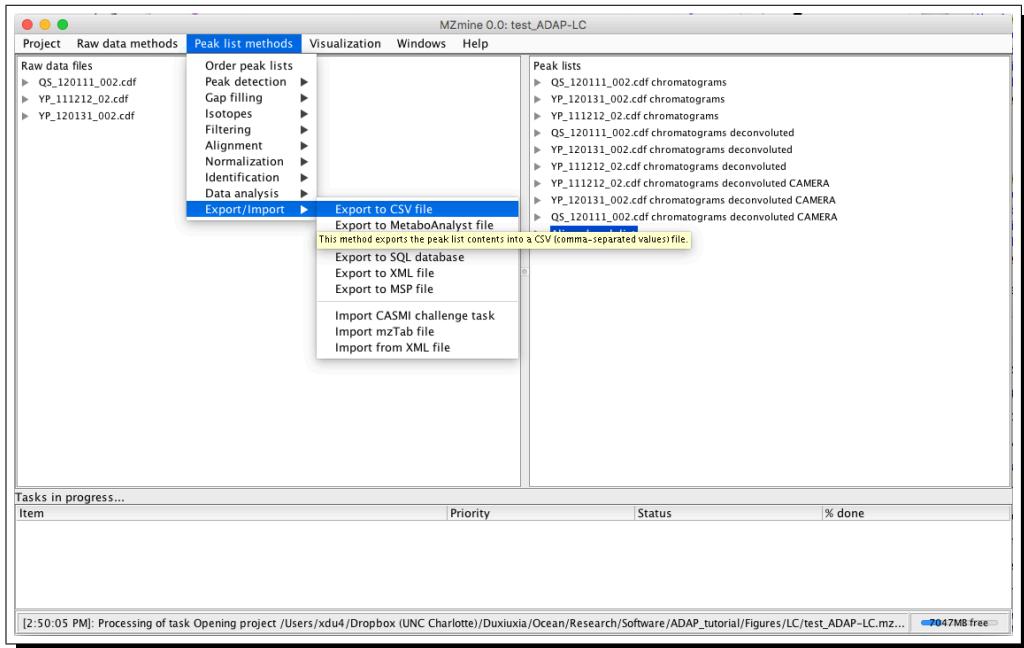
Zoom-in the  $m/z$  at 148.1089 will allow you to see the experimental centroid isotopic distribution of the peak in green and the theoretical isotopic distribution of the PubChem compound in magenta (Figure 32).



**Figure 32:** Experimental isotopic distribution in green vs. the theoretical isotopic distribution in magenta after searching PubChem compound database.

### 3.7 Results Export

The final results after detection of EIC peak detection, alignment, and database search can be exported. Click *Peak list methods* → *Export/Import* → *Export to CSV file* as shown in Figure 33.



**Figure 33:** Export results.

A window pulls up as shown in Figure 34 allowing to select what to export.

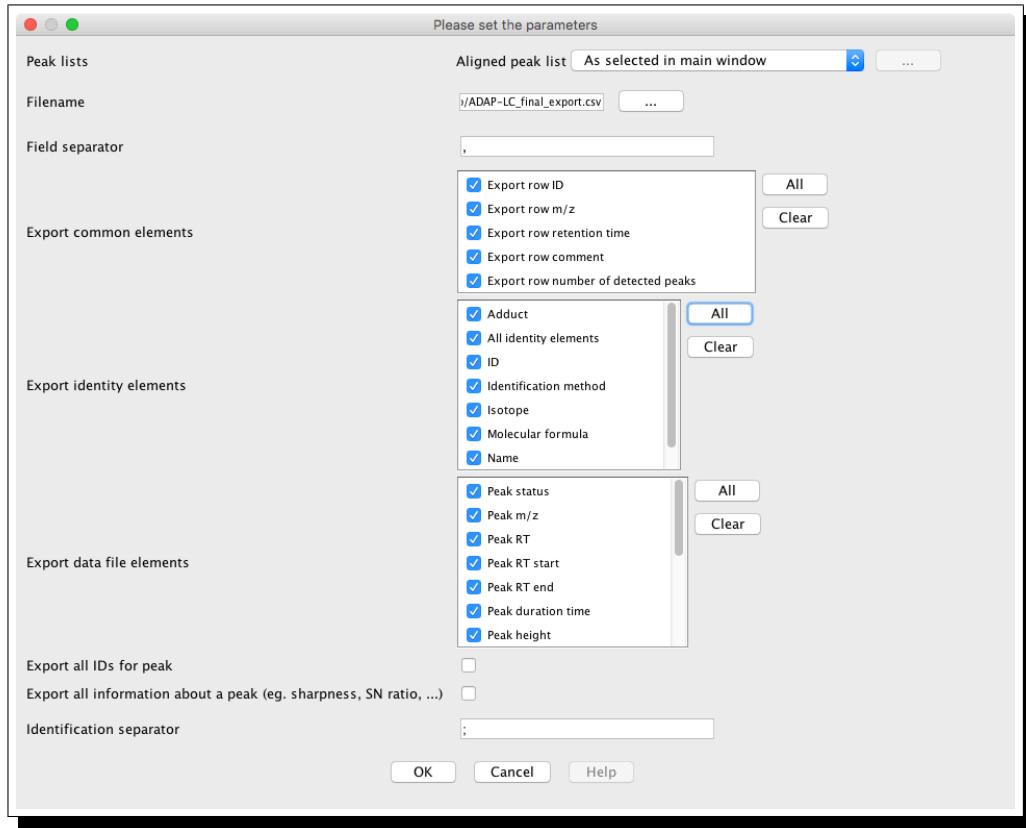


Figure 34: Select what to be exported to a CSV file.

Figure 35 shows part of the exported results.

row ID	row m/z	row retention time	row comment	ber of detect	Adduct	All identity elements	ID
1	386.2726135	0.051240833	Alignment Score = 0.45425929016222477	2			
2	98.51113892	0.1310555	Alignment Score = 0.49501110599754283	2			
3	143.0386505	0.148198333	Alignment Score = 0.4673847727760017	2			
4	488.3408051	0.161140833	Alignment Score = 0.49202553848723546	2			
5	331.2147827	0.192872778	Alignment Score = 0.458136147638013	3	[M+H]+ 330.207		
6	265.1353607	0.281286667	Alignment Score = 0.49173408523311857	2			PubChem ID: 68011675
7	791.7952271	0.6648	Alignment Score = 0.49023058375902917	2	[M+K]+ 752.828 [M+H]+ 790.786		
8	682.1002502	0.6931475	Alignment Score = 0.4318061361901092	2	[M+H]+ 681.096		
9	632.3291626	0.675619167	Alignment Score = 0.47946228597249985	2	[M+H]-CSH8O4]+ 763.364		
10	148.1088511	0.702736111	Alignment Score = 0.987501347059158	3			PubChem ID: 16124170
11	175.11728436	0.702736111	Alignment Score = 0.9815099544211142	3			
12	989.4912415	0.668359167	Alignment Score = 0.37959020067247873	2			
13	139.0171585	0.686798333	Alignment Score = 0.4921370991329182	2			PubChem ID: 59188967
14	167.0117544	0.694780556	Alignment Score = 0.9850512843461698	3	[M+K]+ 128.047		
15	151.0345383	0.693926667	Alignment Score = 0.49308078162242647	2	[M+Na]+ 128.047		
16	1314.740112	0.7716925	Alignment Score = 0.4911225491395063	2			
17	1348.740316	0.767528889	Alignment Score = 0.9617347059339171	3			
18	1220.751017	0.775256111	Alignment Score = 0.9568964656370008	3	[M+K]+ 1181.79		
19	1110.783407	0.779871111	Alignment Score = 0.976151979888738	3			
20	1246.7547476	0.779871111	Alignment Score = 0.9671774751707978	3			
21	1076.782837	0.7679225	Alignment Score = 0.484151015041907	2	[M+Na]+ 1053.79 [M+H-HCOOH]+ 1121.78		
22	1043.300537	0.768465	Alignment Score = 0.48785714356139953	2	[M-Na]+NaCOOH]+ 952.318		
23	1383.229777	0.770042778	Alignment Score = 0.9316060724098225	3			
24	906.8179118	0.784915	Alignment Score = 0.9677718177311489	3	[M+Na]+NaCOOH]+ R15.843		

Figure 35: Exported results.

## 4 ADAP-GC

The first three steps of pre-processing GC-MS data as well as alignment are the same as those for LC-MS data. The major differences between the two pipelines lie in the detection of EIC peaks and deconvolution. Therefore, we will only describe in detail these two steps. We will use three example high mass measurement accuracy data files that were acquired on ThermoFisher GC-Orbitrap to illustrate.

### 4.1 Detection of Masses and Construction of EICs

The three data files are in centroid mode already, so the *Centroid* method in MZmine 2 will be used for mass detection as shown in Figure 36.

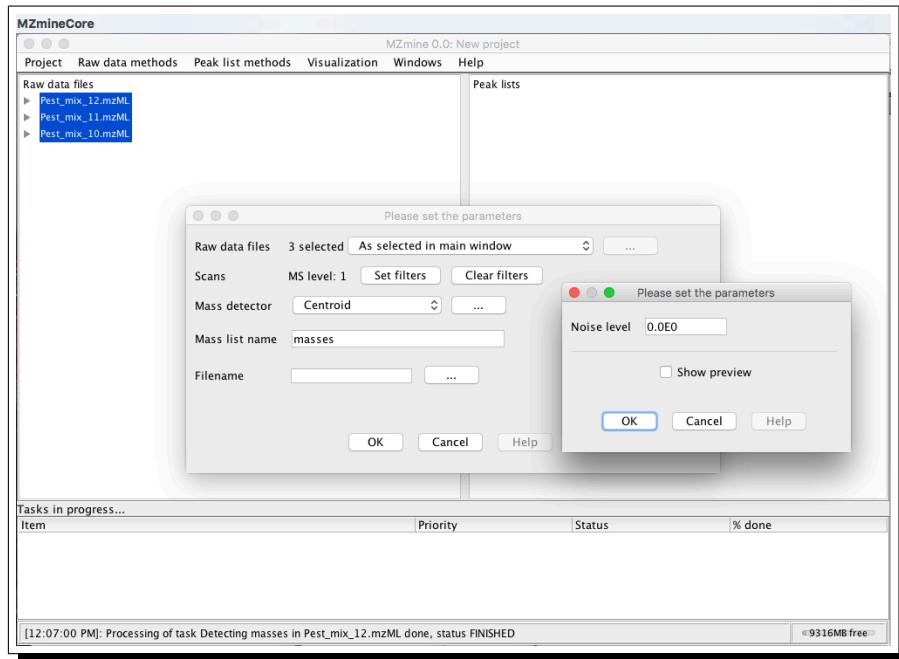
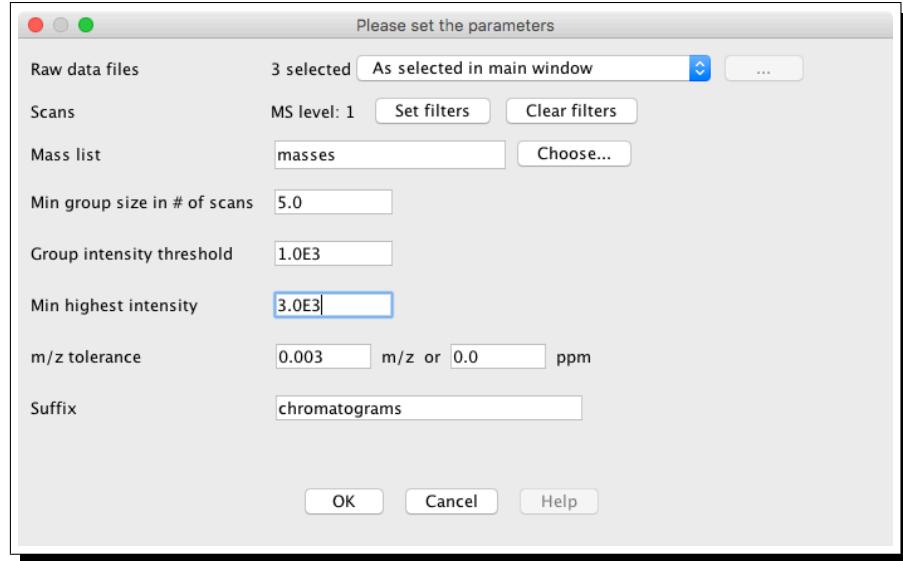


Figure 36: Mass detection of centroid data.

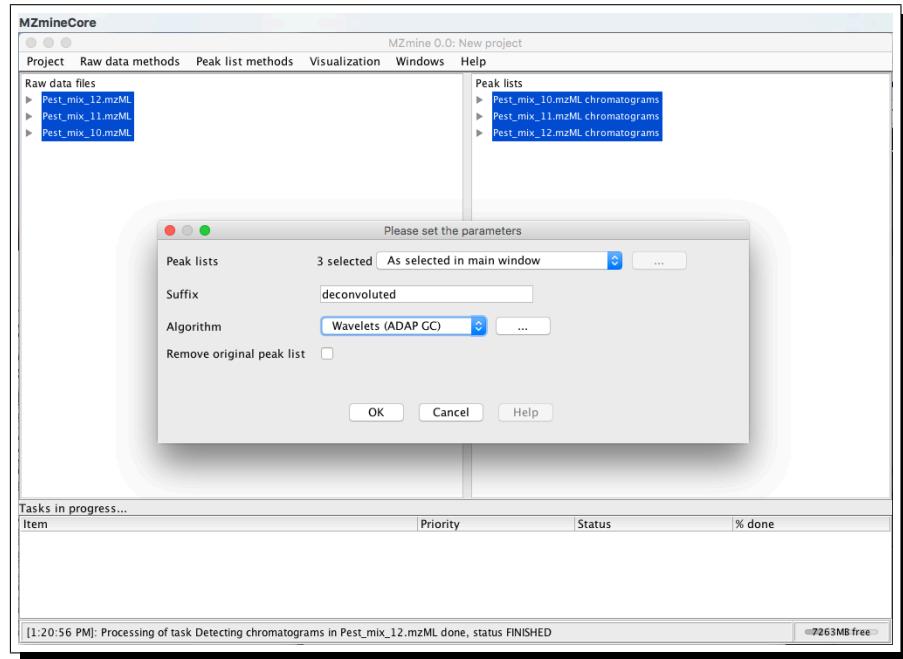
Parameters for constructing EICs are shown in Figure 37.



**Figure 37:** Example parameters for constructing EICs from GC-Orbitrap data.

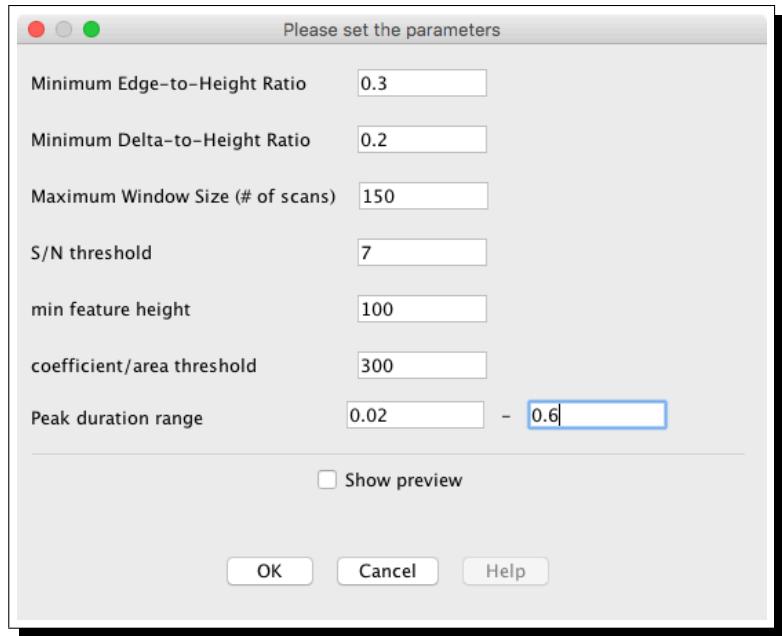
#### 4.2 Detection of Peaks from EICs

Detection of chromatographic peaks is invoked by clicking *Peak list methods* → *Peak detection* → *Chromatogram deconvolution*. A window will open. Select the *Wavelets (ADAP-GC)* algorithm as shown in Figure 38.



**Figure 38:** Select *Wavelet (ADAP-GC)* for detecting peaks from EICs for GC-MS data.

Click the ellipse to open the parameter window. Figure 39 shows example parameters.

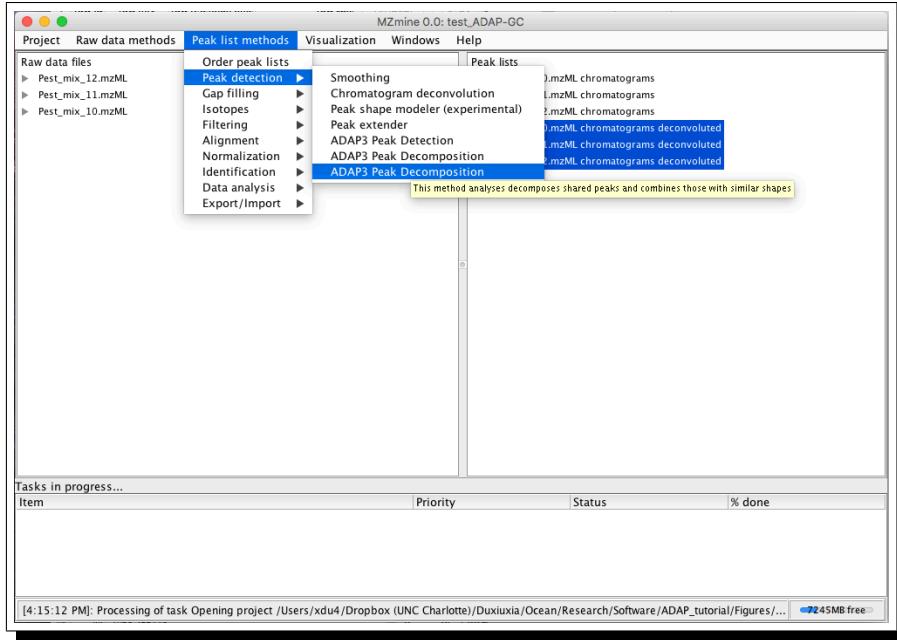


**Figure 39:** Example parameters for detecting peaks from EICs.

Click *OK* to start peak detection. If the duration of chromatography is long, this step could take a while.

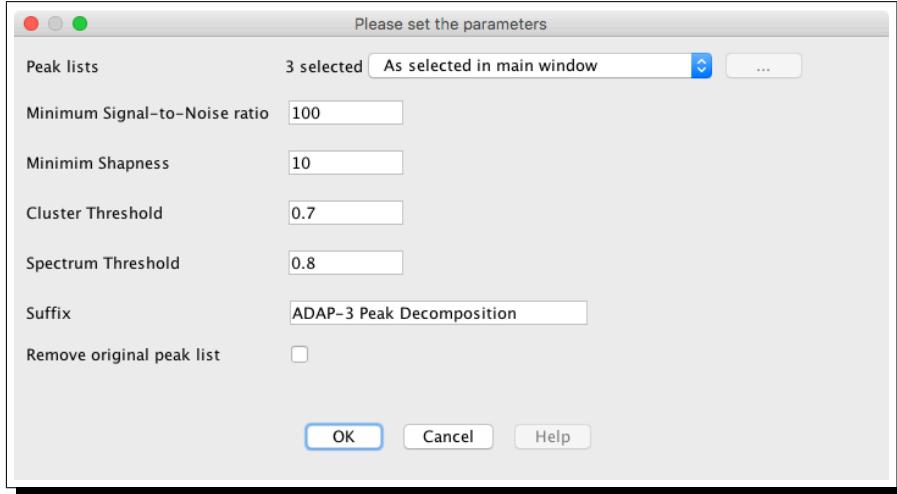
### 4.3 Deconvolution

The term *Deconvolution* here means to decompose a share EIC peak into constituent parts contributed by co-eluting analytes. For details about the underlying algorithm, please refer to [4, 5, 6]. To distinguish this deconvolution from the deconvolution used in LC-MS data pre-processing, we will sometimes use *decomposition* here instead. To perform decomposition, select all of the chromatographic peaks detected from one or more data files, then click *Peak list methods* → *Peak detection* → *ADAP3 peak decomposition* as shown in Figure 40.



**Figure 40:** Deconvolution of chromatographic peaks.

A window as shown in Figure 41 pulls up allowing you to specify parameters for deconvolution.

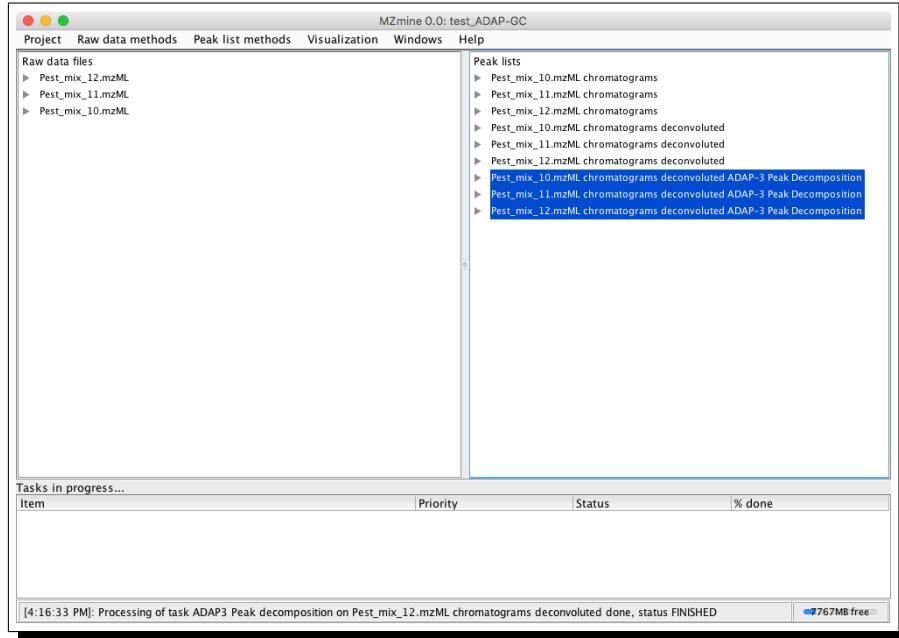


**Figure 41:** Specify parameters for decomposition of chromatographic peaks.

- *Minimum Signal-to-Noise ratio:* Minimum allowed signal-to-noise ratio for a component's mode peak
- *Minimum Sharpness:* Minimum allowed sharpness of a component's model peak
- *Cluster Threshold:* All peaks are clustered based on their shape-similarity. Cluster Threshold determines the “radius” of a cluster: only peaks with shape-similarity above the threshold are included in a cluster.
- *Spectrum Threshold:* While building spectra, each peak is decomposed into a linear combination of model peaks. Spectrum Threshold determines the quality of the decomposition: if

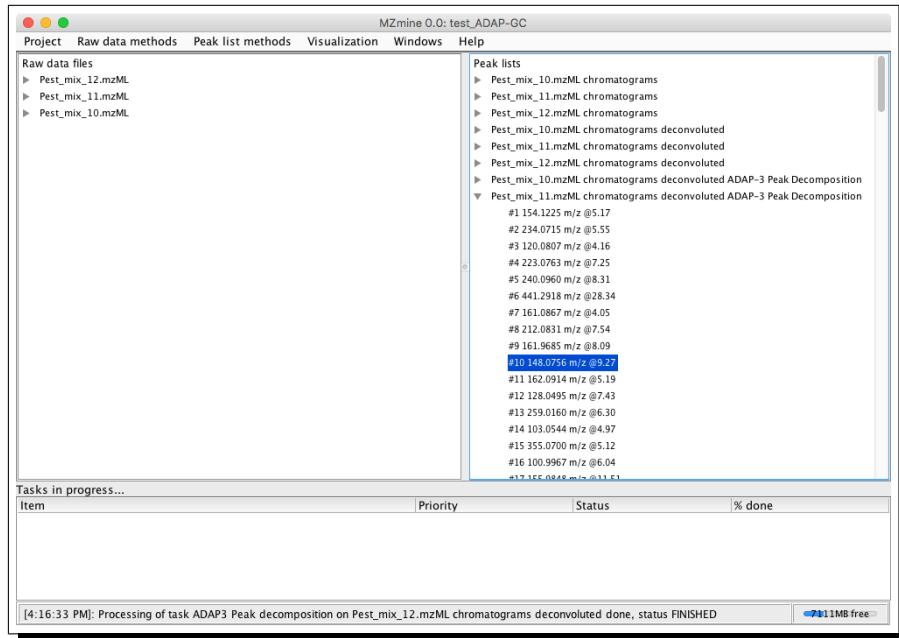
the quality is below Spectrum Threshold, the peak is not included in spectra.

After decomposition is finished, the results are displayed as shown in Figure 42.



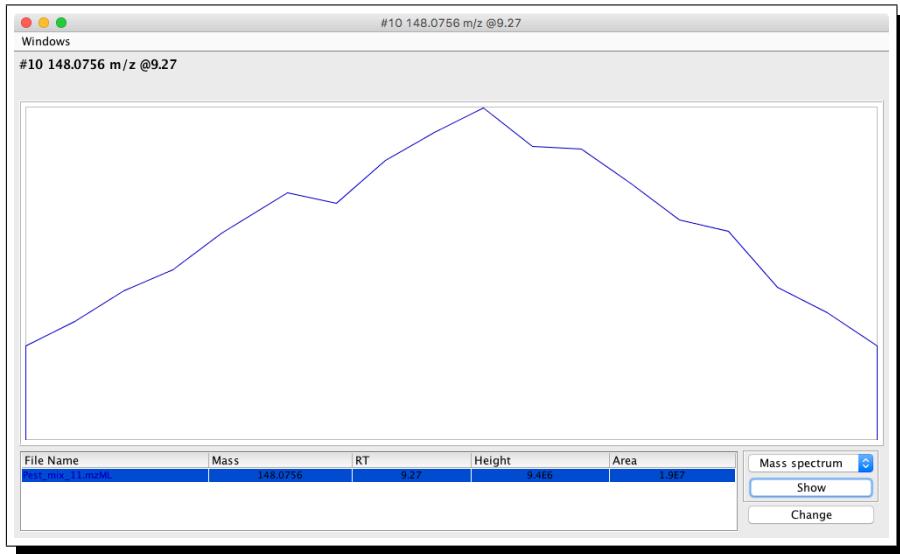
**Figure 42:** Decomposition results.

Expand the peak decomposition results for each data file by clicking on the left triangle, you will see a list of mass spectra that have been constructed by the deconvolution algorithm (Figure 43). The  $m/z$  for each entry is the  $m/z$  of the model peak for this spectrum.



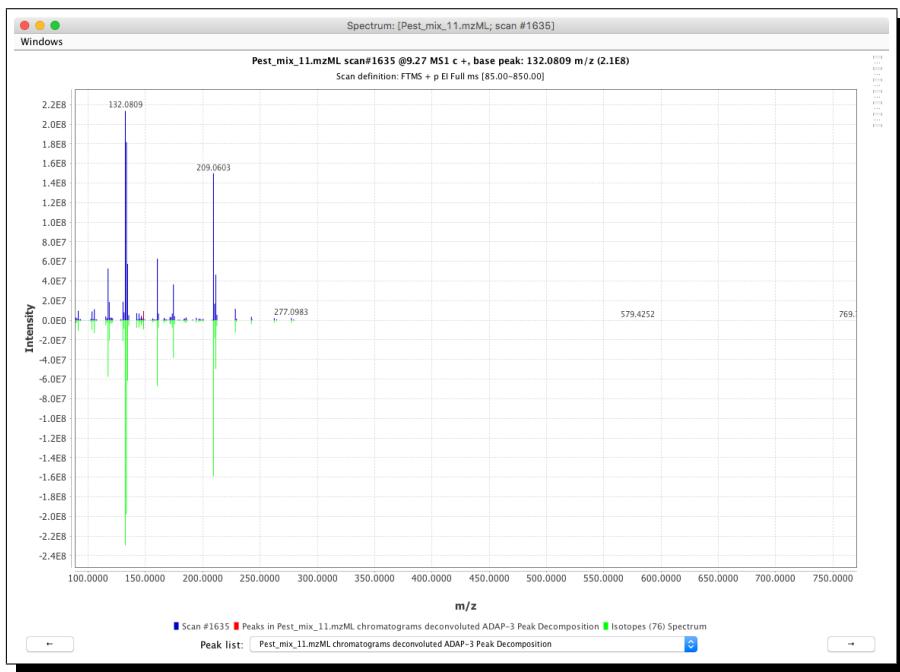
**Figure 43:** List of mass spectra constructed by the decomposition algorithm.

Double click on a particular mass spectrum will pull up a window as shown in Figure 44.



**Figure 44:** Elution profile of the analyte.

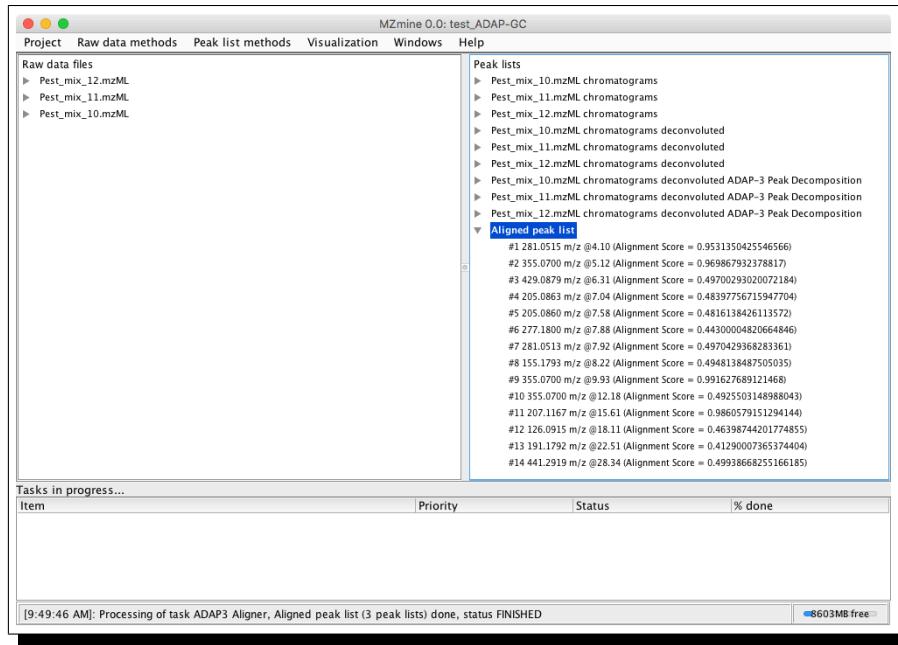
Click on the data file name and then select *Mass spectrum* in the drop-down menu on the right. The spectrum that has been constructed in the context of the raw spectrum is displayed (Figure 45).



**Figure 45:** Mass spectra constructed by the decomposition algorithm.

#### 4.4 Alignment

Perform alignment as described in LC-MS data pre-processing. The result is shown in Figure 46.

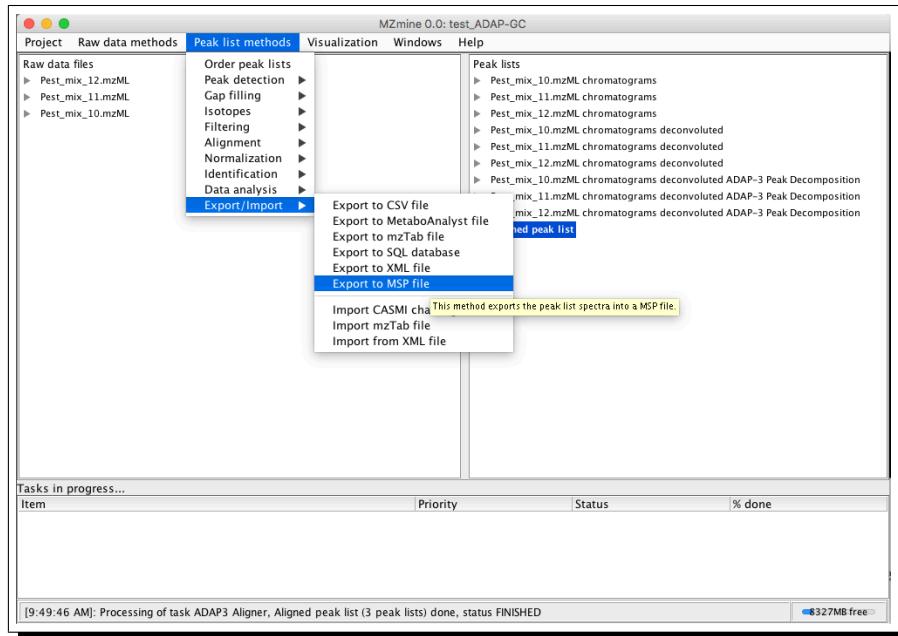


**Figure 46:** Alignment result.

You may double click each spectrum and visually examine the EICs and spectra as described previously.

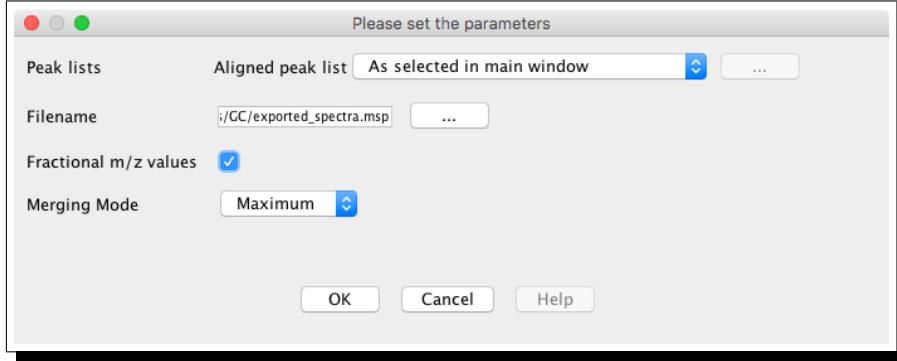
#### 4.5 Spectra Export

The mass spectra that have been constructed can be exported in .msp format and then imported to *NIST MS Search* for identification. To export the spectra, select the *Aligned peak list* and then click *Peak list methods* → *Export/Import* → *Export to MSP file* as shown in Figure 47.



**Figure 47:** Export mass spectra to a MSP file.

A window as shown in Figure 48 will pull up. You will need to choose a location and file name for the .msp file, check whether or not to round the  $m/z$  values for searching against unit-mass spectral libraries, and the merging mode when rounding is selected (i.e. two or more peaks exist within a 1 dalton window).



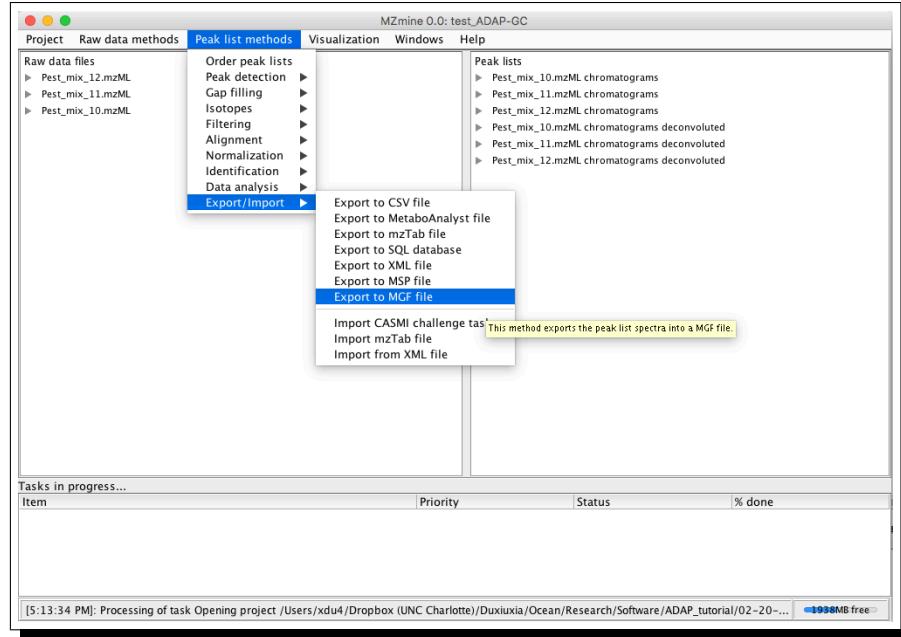
**Figure 48:** Export mass spectra to a MSP file.

Open the exported .msp file in a text editor. You will see that the mass spectra after alignment have been exported. Figure 49 shows a small portion of the .msp file.

```
Name: #1 281.0515 m/z @4.10 (Alignment Score = 0.9531350425546566)
DB#: 1
Num Peaks: 62
92.02568054199219 287278.71075470065
185.02923583984375 874378.1240155919
106.04122161865234 81468.73115179865
123.06768035888672 245995.74908396933
149.04736328125 2572571.3844742966
150.04652404785156 234995.55398492888
164.94801330566406 34306.16408879444
178.96493530273438 86455.19008466511
191.001220703125 278791.58316731517
192.98019409179688 266150.5970837946
285.01695251464844 50189.81177968083
287.03268432617188 475207.5692497589
289.01161193847656 241779.2090340591
221.08436584472656 1064530.8474890182
223.0816650390625 99404.0289137579
225.04273986816406 251728.76147294548
236.9520263671875 59029.37689508891
239.0953826904297 184586.40998280022
248.98899841308594 289881.40850885562
252.98326110839844 47241.12614709242
```

**Figure 49:** Example .msp file exported by ADAP-GC.

The constructed mass spectra can also be exported in .mgf format. To do so, select the *Aligned peak list* and then click *Peak list methods* → *Export/Import* → *Export to MGF file* as shown in Figure 50.



**Figure 50:** Export mass spectra to a MGF file.

## 5 List of Additions and Changes Du-lab Team Made to MZmine 2

For details about the following changes and addition, please refer to the main text of the tutorial.

- Category: *Raw data methods* → *Peak detection*
  - **Mass detection:** added *Filename* for choosing the directory and filename to output detected masses to.
  - **ADAP Chromatogram builder:** a new method of chromatogram building.
- Category: *Peak list methods* → *Peak Detection*
  - **Chromatogram Deconvolution: Wavelets (ADAP) and Wavelets (ADAP GC).**
  - **ADAP3 Peak Detection:** Interface to R-script of ADAP-GC 3.0 peak picking [6]. This is legacy code and will be replaced by the new **Wavelets (ADAP GC)** peak picking.
  - **ADAP3 Peak Decomposition:** Improved version of ADAP-GC peak decomposition compared to ADAP-GC 3.0 [6].
- Category: *Peak list methods* → *Alignment*
  - **ADAP3 Aligner:** Improved version of ADAP-GC alignment compared to ADAP-GC 1.0 [4].
- Category: *Peak list methods* → *Identification*
  - **CAMERA search:** Modified CAMERA search.
  - **ADAP GC-MS Search and ADAP LC-MS Search:** Local database search.
- Category: *Peak list methods* → *Export / Import*
  - **Export to MSP file:** exporting constructed spectra to a file in MSP format
- Category: *Visualization*
  - **Point 2D visualizer:** Heat map visualization of intensities in RT and  $m/z$  domain.

## References

- [1] **MZmine 2** [<http://mzmine.github.io/>]
- [2] Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S: **CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets.** *Anal Chem* 2012, 84(1):283-289.
- [3] **CAMERA** [<https://bioconductor.org/packages/release/bioc/html/CAMERA.html>]
- [4] Jiang W, Qiu Y, Ni Y, Su M, Jia W, Du X: **An automated data analysis pipeline for GC-TOF-MS metabonomics studies.** *J Proteome Res* 2010, 9(11):5974-5981.
- [5] Ni Y, Qiu Y, Jiang W, Suttlemyre K, Su M, Zhang W, Jia W, Du X: **ADAP-GC 2.0: deconvolution of coeluting metabolites from GC/TOF-MS data for metabolomics studies.** *Anal Chem* 2012, 84(15):6619-6629.
- [6] Ni Y, Su M, Qiu Y, Jia W, Du X: **ADAP-GC 3.0: Improved Peak Detection and Deconvolution of Co-eluting Metabolites from GC/TOF-MS Data for Metabolomics Studies.** *Anal Chem* 2016, 88(17):8802-8811.