

# **ADAP**

## **User Manual**

Version 3.3.0

Du-Lab Team

Department of Bioinformatics and Genomics

University of North Carolina at Charlotte

xiuxia.du@uncc.edu

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

January 25, 2018

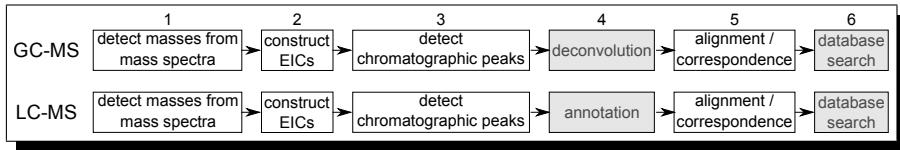
## Contents

<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>3</b>
3.1	Detection of Masses from Mass Spectra . . . . .	3
3.2	Construction of Extracted Ion Chromatograms . . . . .	6
3.3	Detection of Peaks from EICs . . . . .	9
3.4	Annotation of EIC Peaks Using CAMERA . . . . .	12
3.5	Results Export . . . . .	15
<b>4</b>	<b>ADAP-GC</b>	<b>18</b>
4.1	Detection of Masses and Construction of EICs . . . . .	18
4.2	Detection of Peaks from EICs . . . . .	19
4.3	Spectral Deconvolution . . . . .	20
4.4	Spectra Export . . . . .	25
<b>5</b>	<b>Batch Processing</b>	<b>28</b>
<b>6</b>	<b>List of Additions and Changes Du-lab Team Made to MZmine 2</b>	<b>29</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, 3, and 5. The differences between the two pipelines lie in modules 4 and 6. Deconvolution is unique to ADAP-GC while peak annotation is unique to ADAP-LC. 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.

The computing modules for construction of EICs, detecting peaks, and deconvolution have been developed by Du-lab team, implemented in Java, 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

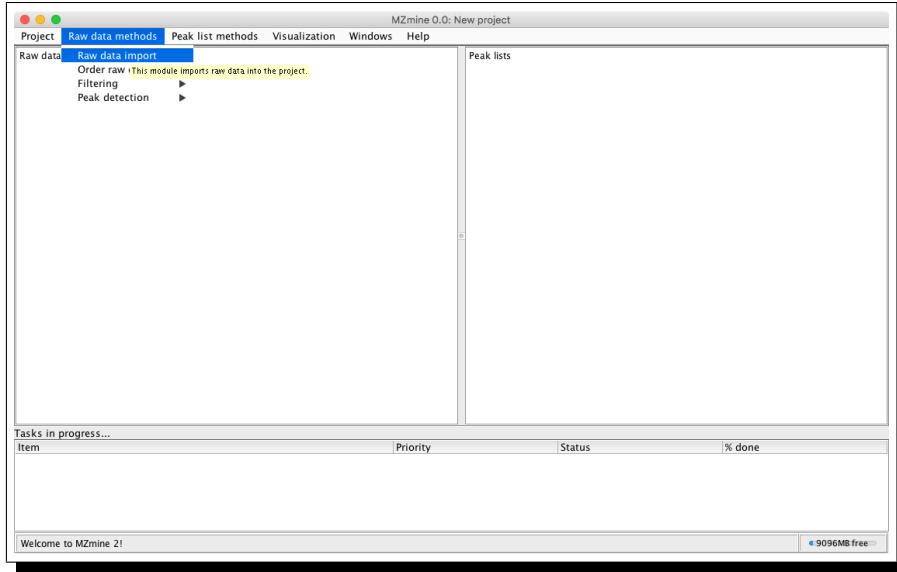
ADAP computational modules have been part of MZmine 2 since version MZmine 2.24. No installation of extra packages is required. For description on how to download and install MZmine 2, please refer to the MZmine 2 manual [1].

## 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.

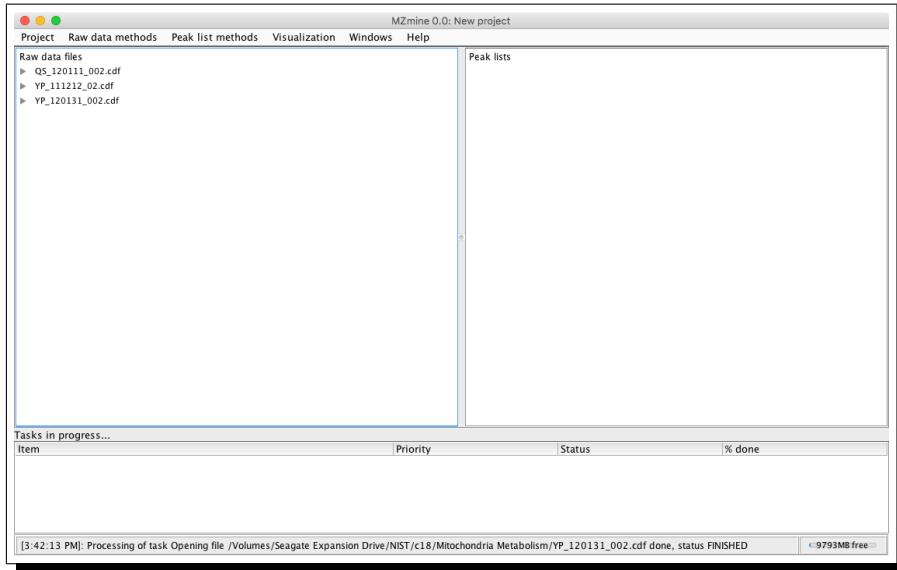
### 3.1 Detection of Masses from Mass Spectra

Click on *Raw data methods* → *Raw data import*, shown in Figure 2.



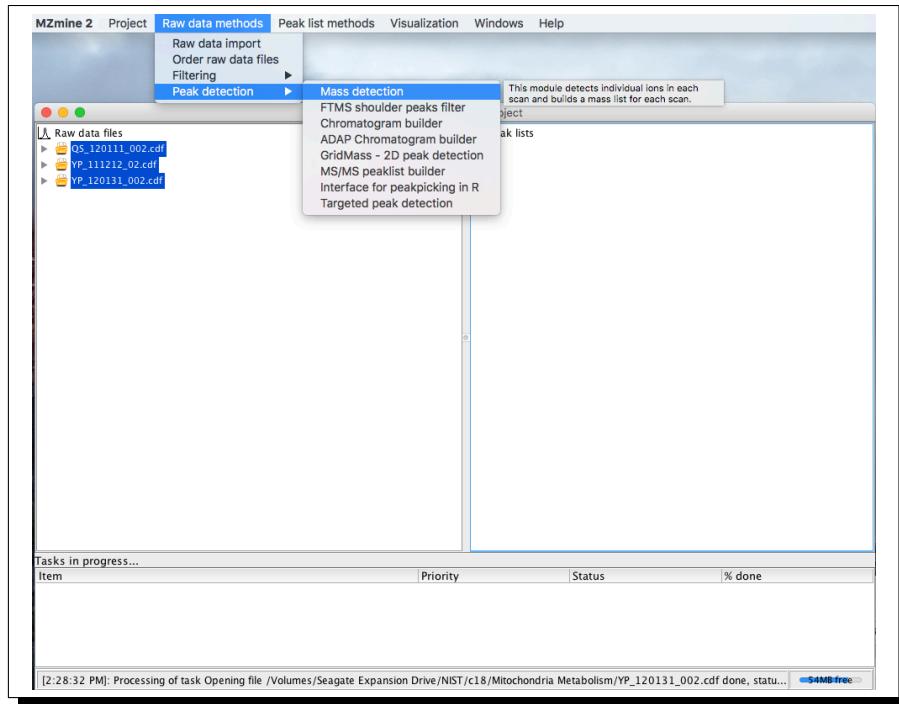
**Figure 2:** Import the raw data file.

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 3.



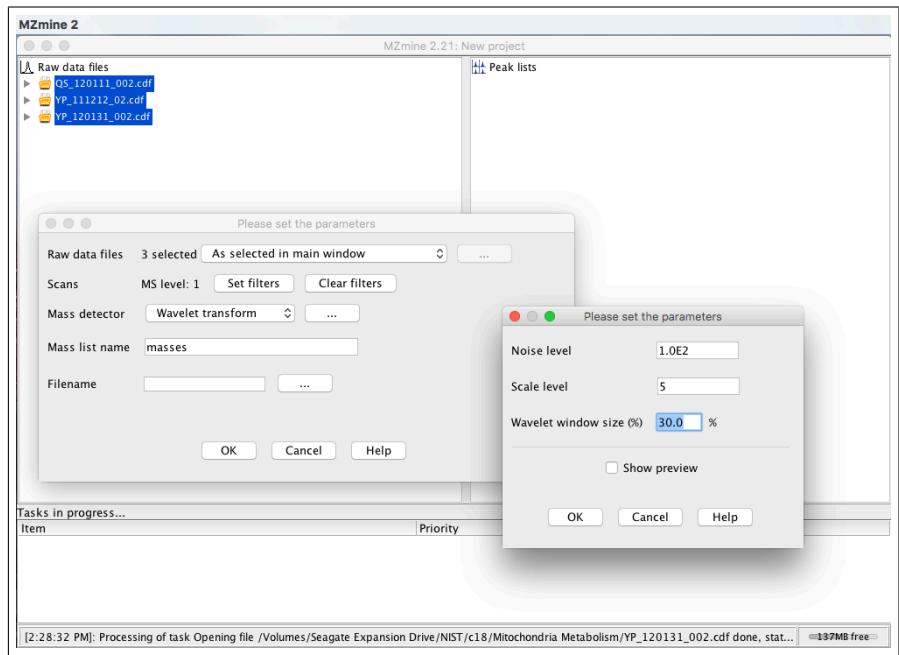
**Figure 3:** Imported data files.

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. 4.



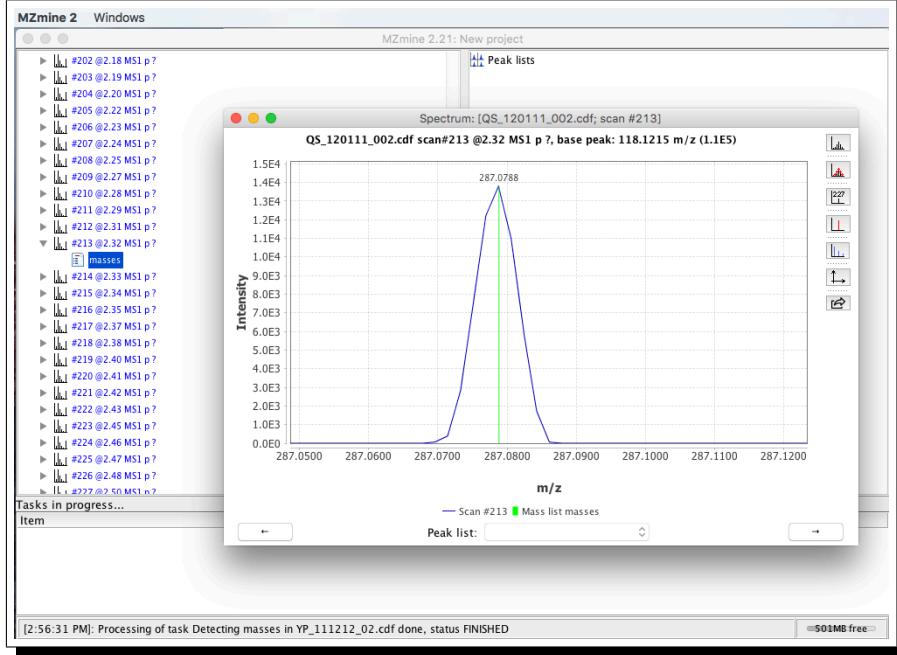
**Figure 4:** Mass detection from profile mass spectra

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 5.



**Figure 5:** Mass detection by continuous wavelet transform.

Click *OK* in both windows in Figure 5 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 6. By stacking together the centroid spectrum and the profile spectrum, you can check how well the mass detection works.

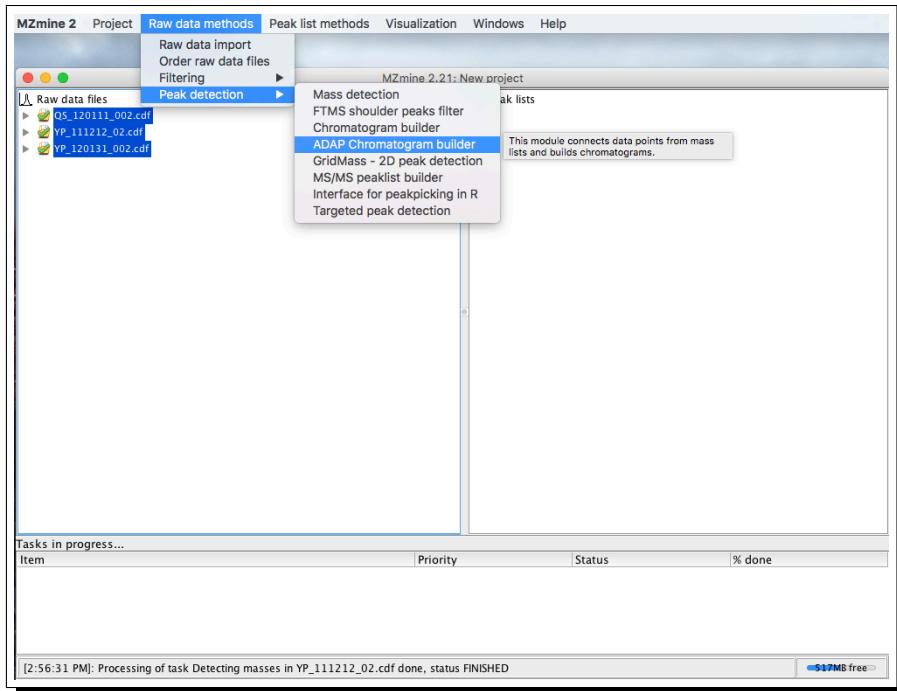


**Figure 6:** Mass detection result.

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

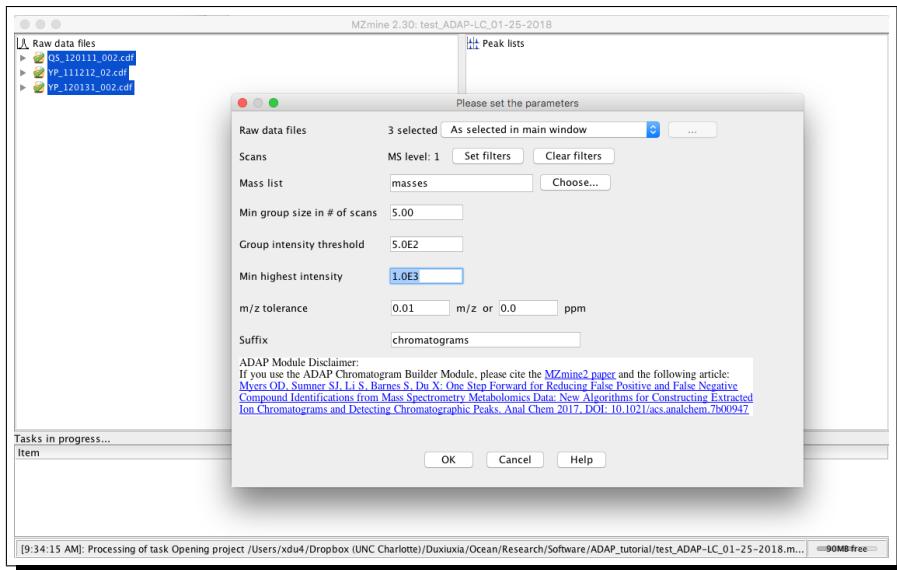
### 3.2 Construction of Extracted Ion Chromatograms

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 7.



**Figure 7:** 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 8.



**Figure 8:** Example of ADAP chromatogram building parameters.

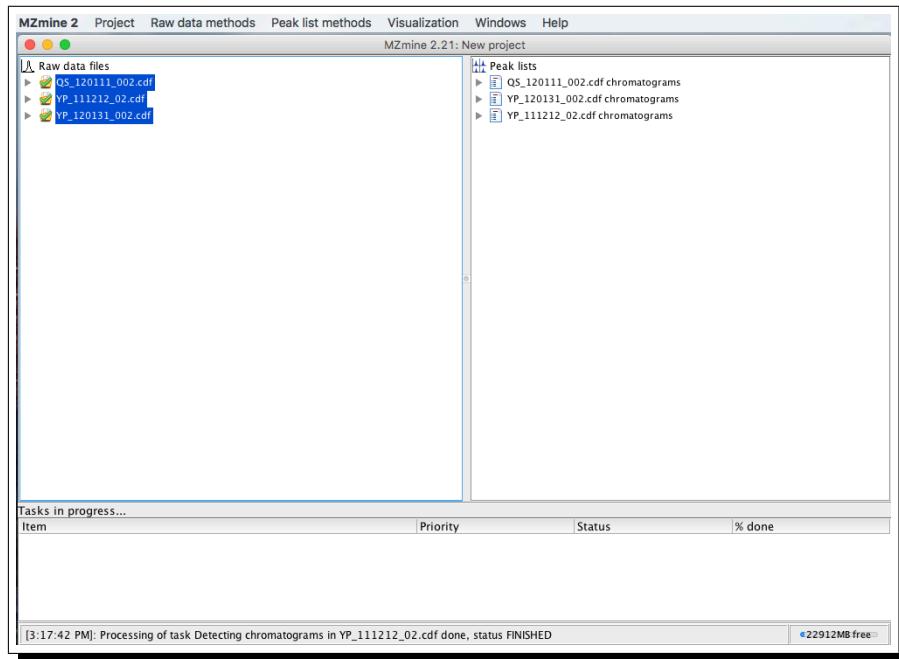
#### Description of parameters:

- *Min group size in # of scans*: In the entire chromatogram there must be at least this number of sequential scans having points above the *Group intensity threshold* set by the user.

The optimal value depends on the chromatography system setup. The best way to set this parameter is by studying the raw data and determining what is the typical time span of chromatographic peaks.

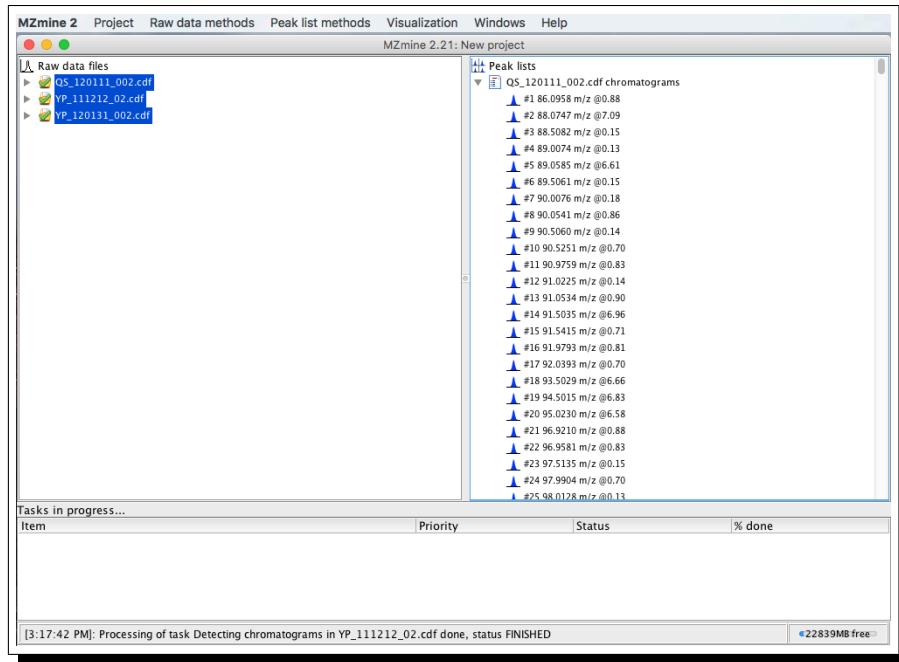
- *Group intensity threshold*: See above.
- *Min highest intensity*: There must be at least one point in the chromatogram that has an intensity greater than or equal to this value.
- *m/z tolerance*: Maximum m/z difference of data points in consecutive scans in order to be connected to the same chromatogram. 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.
- *Suffix*: The resulting chromatogram will be named *file name + suffix*.

Click *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 9.



**Figure 9:** Results of chromatogram building.

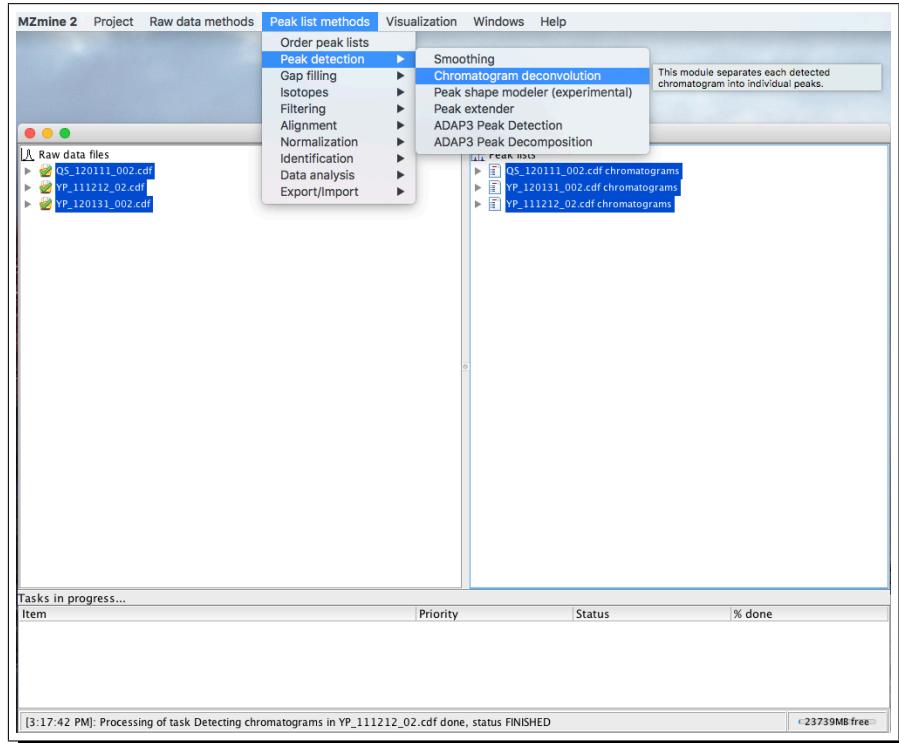
Click the triangle to the left of each data file expands the list of EICs as shown in Figure 10.



**Figure 10:** List of EICs that have been constructed.

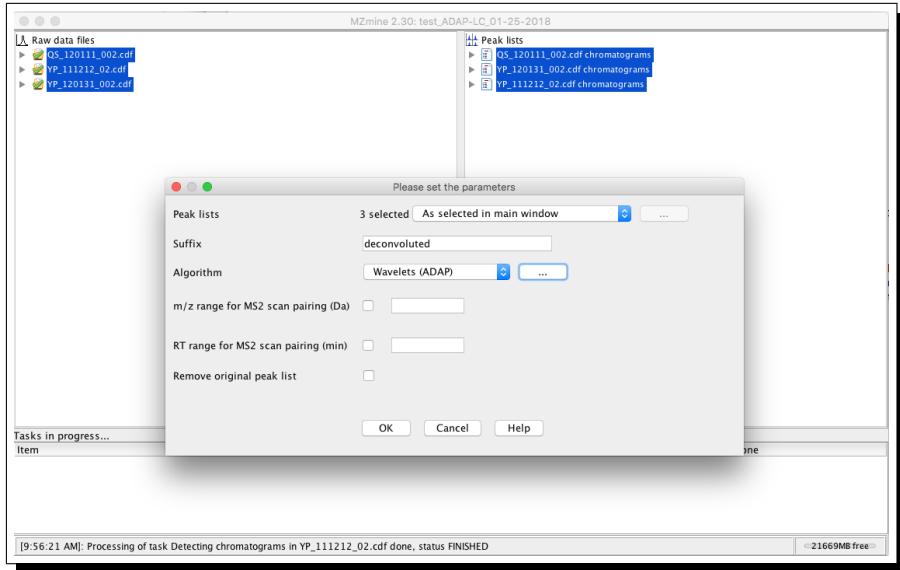
### 3.3 Detection of Peaks from EICs

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 11.



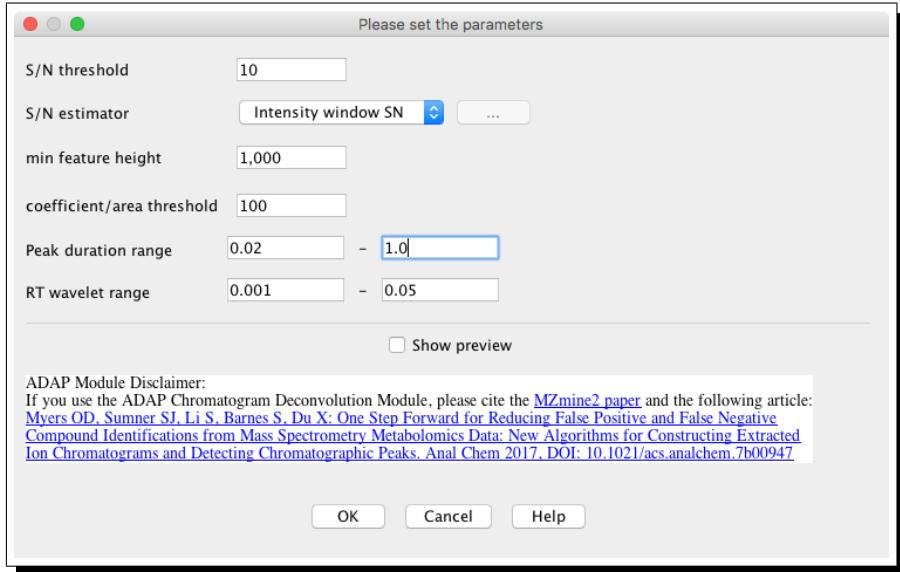
**Figure 11:** Detect peaks from EICs.

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 as shown in Fig. 12.



**Figure 12:** Select ADAP peak detection.

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.



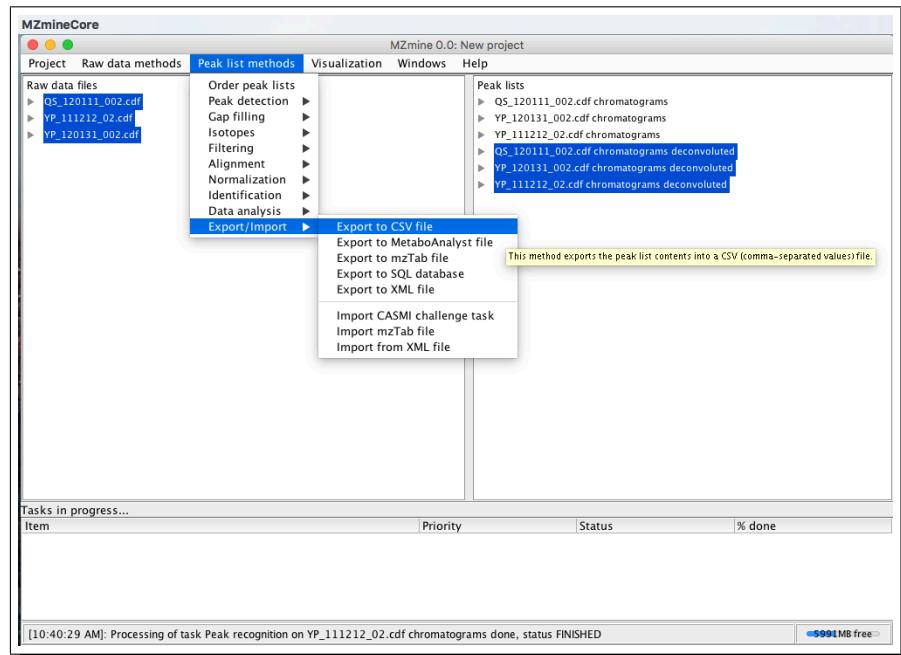
**Figure 13:** EIC peak detection parameters.

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.

- *S/N estimator*: User can choose one of two estimators of the signal-to-noise ratio
  - *Intensity window SN* (tested on LC-MS datasets) uses the peak height as the signal level and the standard deviation of intensities around the peak as the noise level;
  - *Wavelet Coeff. SN* (tested on GC-MS datasets) uses the continuous wavelet transform coefficients to estimate the signal and noise levels. Analogous approach is implemented in R-package *wmtsa*.
- *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.
- *Peak duration range*: Minimum and maximum widths allowed for a peak.
- *RT wavelet range*: Minimum and maximum widths of the wavelets used for detecting peaks.

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	e YP_120131_002.cdf chromatograms	YP_120131_002.cdf chromatograms				
1	86.0957031	0.90133833		1	DETECTED	86.0957031	0.90133833	0.84557	1.12000333	0.2744:	
2	90.0542068	0.90133833		1	DETECTED	90.0542068	0.90133833	0.87351833	0.957405	0.0838:	
3	90.5253906	0.718745		1	DETECTED	90.5253906	0.718745	0.63083	0.78977167	0.1589:	
4	90.9760208	0.81773667		1	DETECTED	90.9760208	0.81773667	0.73311833	0.91526833	0.1:	
5	91.0270233	0.73311833		1	DETECTED	91.0270233	0.73311833	0.66087167	0.74731167	0.08	
6	91.0270233	6.80784		1	DETECTED	91.0270233	6.80784	6.77885	6.80784	0.0:	
7	91.0534515	0.92924667		1	DETECTED	91.0534515	0.92924667	0.91526833	1.02967	0.1144:	
8	91.9795227	0.83164833		1	DETECTED	91.9795227	0.83164833	0.74731167	0.87351833	0.1262:	
9	92.5217133	0.73311833		1	DETECTED	92.5217133	0.73311833	0.69002333	0.74731167	0.0572:	
10	94.0445557	0.73311833		1	DETECTED	94.0445557	0.73311833	0.67550333	0.76158333	0.08	
11	94.0445557	4.37849833		1	DETECTED	94.0445557	4.37849833	4.37849833	4.40693167	0.0284:	
12	96.9212341	0.87351833		1	DETECTED	96.9212341	0.87351833	0.84557	1.04465	0.1:	
13	97.5137787	4.37849833		1	DETECTED	97.5137787	4.37849833	4.33529167	4.435335	0.1000:	
14	97.9908905	0.718745		1	DETECTED	97.9908905	0.718745	0.66087167	0.77578	0.1149:	
15	98.9181213	0.85954667		1	DETECTED	98.9181213	0.85954667	0.83164833	1.000039167	0.1687:	
16	99.0545044	0.957405		1	DETECTED	99.0545044	0.957405	0.84557	1.000039167	0.1548:	
17	99.0545044	0.18226333		1	DETECTED	99.0545044	0.18226333	0.096375	0.31042833	0.2140:	
18	99.5306015	0.73311833		1	DETECTED	99.5306015	0.73311833	0.64588	0.77578	0.:	
19	99.5306015	5.59725667		1	DETECTED	99.5306015	5.59725667	5.582525	5.64120333	0.0586:	
20	100.028282	0.70448833		1	DETECTED	100.028282	0.70448833	0.67550333	0.73311833	0.05:	
21	100.057961	0.67550333		1	DETECTED	100.057961	0.67550333	0.64588	0.76158333	0.1157:	
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	2.33838133	0.07:	

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).

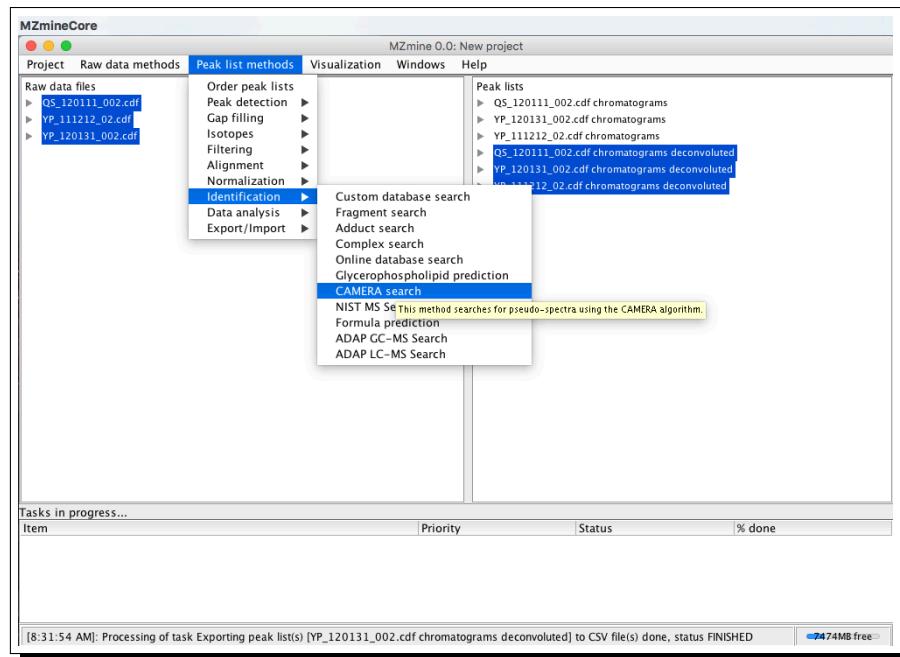
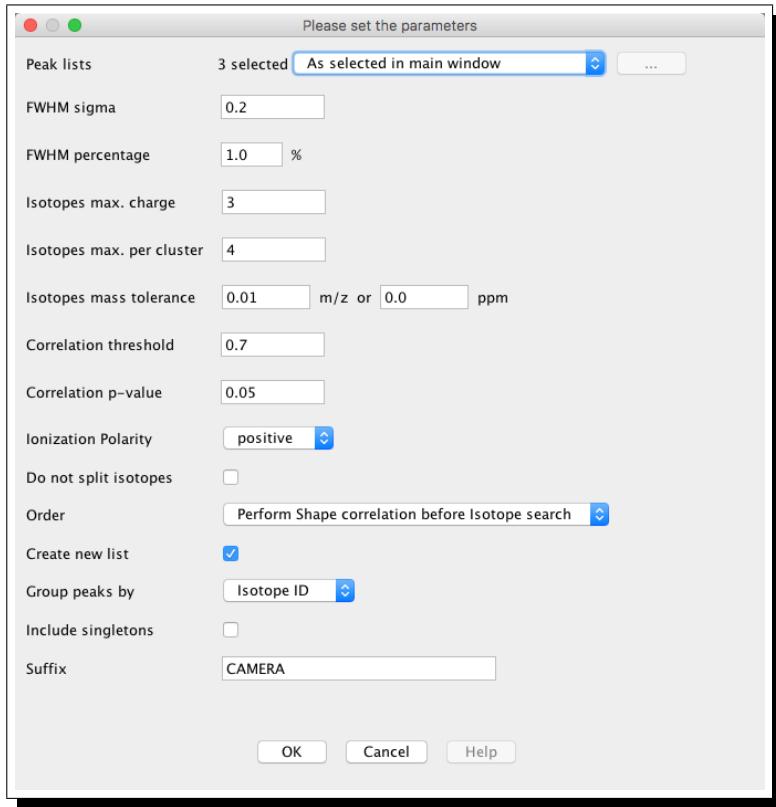


Figure 16: Use CAMERA for annotation of EIC peaks.

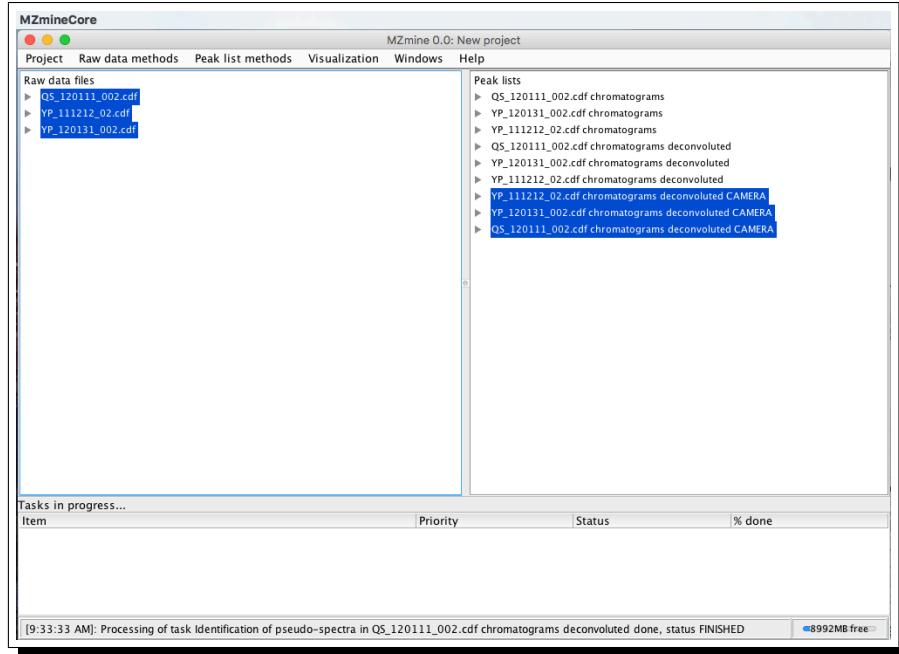
A window will pull up as shown in Figure 17 allowing users to specify parameters.



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

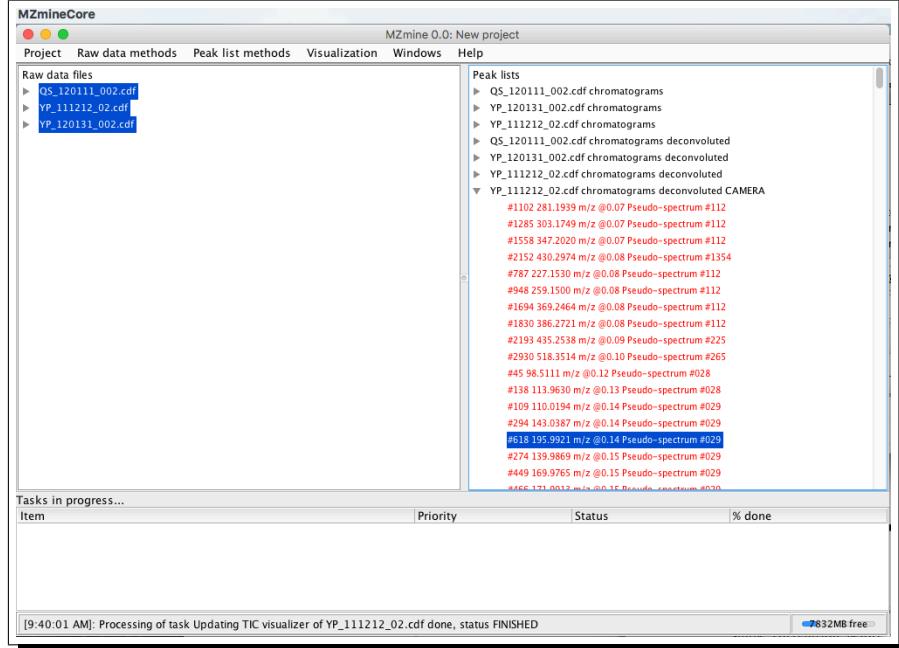
With the slight modification by the Du-lab team, an option (item *Order* in Figure 17) is provided to *perform shape correlation before isotope search* for stricter requirement of determining an isotopic pattern. With this stricter requirement, the mass peaks that form an isotopic pattern will have to 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.



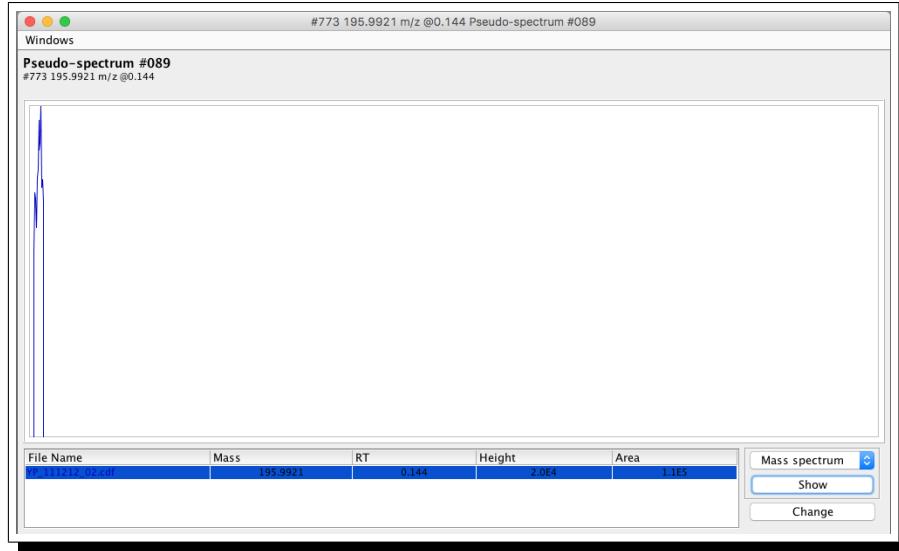
**Figure 18:** CAMERA finishes searches and results are displayed.

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*.



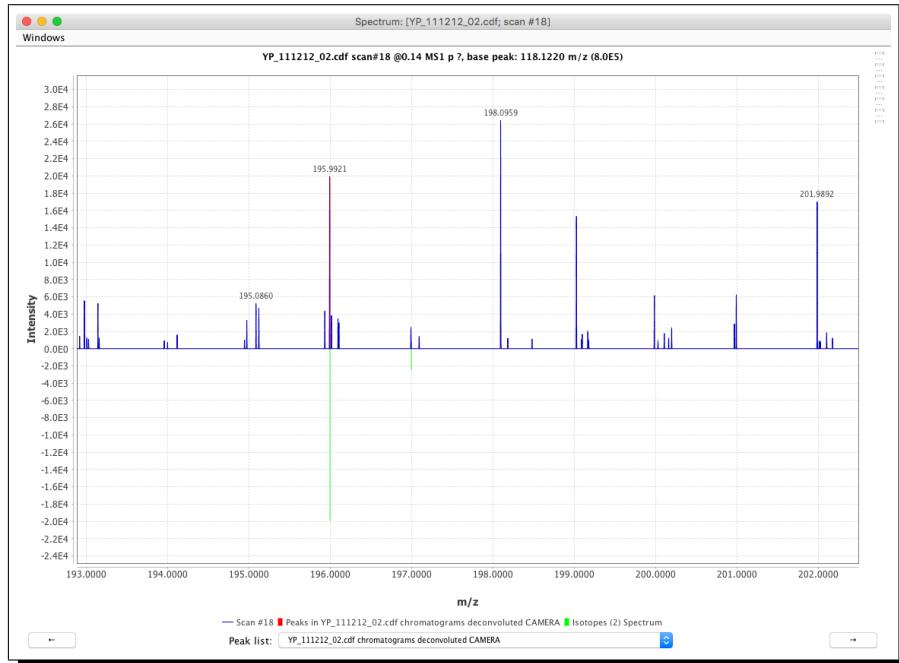
**Figure 19:** List of pseudo-spectra are displayed.

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.

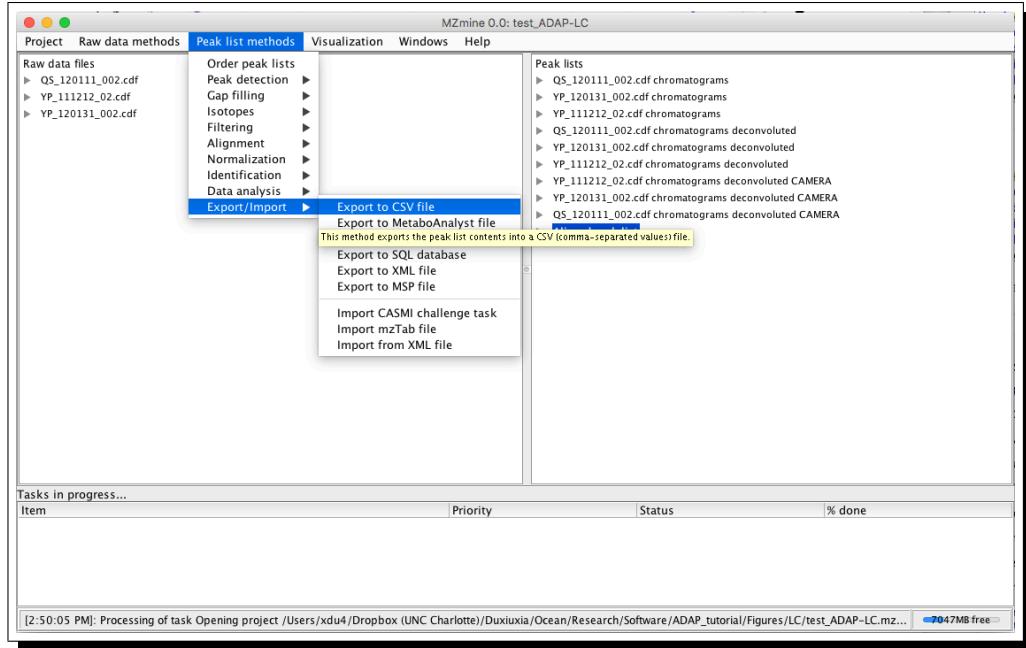
Select *Mass spectrum* in the bottom-right corner and then click on *Show* will pull up a window displaying the pseudo spectrum (green sticks) in the context of the raw spectrum (Figure 21).



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

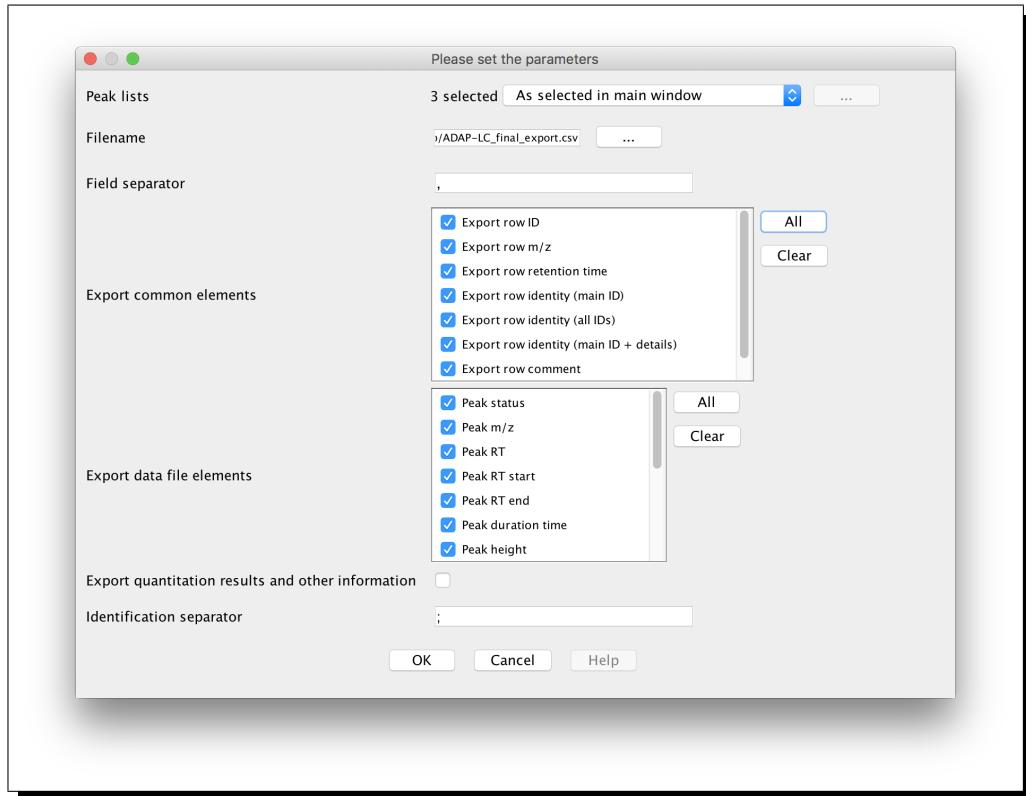
### 3.5 Results Export

The final results after detection of EIC peak detection can be exported. Click *Peak list methods* → *Export/Import* → *Export to CSV file* as shown in Figure 22.



**Figure 22:** Export results.

A window pulls up as shown in Figure 23 allowing to select what to export.



**Figure 23:** Select what to be exported to a CSV file.

Figure 24 shows part of the exported results. The CAMERA results can be found in the column “row identity (main ID + details)”.

row ID	row m/z	row retention time	row identity (main ID)	row identity (all IDs)	row identity (main ID + details)	row (YP_111212_0YP_111212_02)
1314	303.1749268	0.0706000	Pseudo-spectrum #131	Pseudo-spectrum #131	Name: Pseudo-spectrum #131;Isotope: [44]M+;Adduct: [M+Na]+;280.186;identification method: Bioconductor CAMERA	1 DETECTED 303.1749268
1352	342.2467797	0.078593333	Pseudo-spectrum #150	Pseudo-spectrum #150	Name: Pseudo-spectrum #150;Isotope: [57]M+;Adduct: [M+NH3]+;161.194;identification method: Bioconductor CAMERA	1 DETECTED 342.2467797
1393	342.2467797	0.078593333	Pseudo-spectrum #150	Pseudo-spectrum #150	Name: Pseudo-spectrum #150;Isotope: [57]M+;Adduct: [M+NH3]+;324.212;identification method: Bioconductor CAMERA	1 DETECTED 342.2467797
1780	369.2464294	0.078593333	Pseudo-spectrum #150	Pseudo-spectrum #150	Name: Pseudo-spectrum #150;Isotope: [68]M+;Adduct: [M+H]+;368.238;identification method: Bioconductor CAMERA	1 DETECTED 369.2464294
1930	386.2721252	0.078593333	Pseudo-spectrum #150	Pseudo-spectrum #150	Name: Pseudo-spectrum #150;Isotope: [79]M+;Adduct: [M+NH3]+;368.238;identification method: Bioconductor CAMERA	1 DETECTED 386.2721252
1970	391.2273865	0.078593333	Pseudo-spectrum #150	Pseudo-spectrum #150	Name: Pseudo-spectrum #150;Isotope: [81]M+;Adduct: [M+Na]+;368.238;identification method: Bioconductor CAMERA	1 DETECTED 391.2273865
3162	518.3513794	0.103911667	Pseudo-spectrum #136	Pseudo-spectrum #136	Name: Pseudo-spectrum #136;Isotope: [163]M+;Adduct: [M+Na]+;identification method: Bioconductor CAMERA	1 DETECTED 518.3513794
938	259.1502244	0.160346667	Pseudo-spectrum #135	Pseudo-spectrum #135	Name: Pseudo-spectrum #135;Isotope: [23]M+;Adduct: [M+H]+;236.161;identification method: Bioconductor CAMERA	1 DETECTED 259.1502244
2783	488.3432137	0.164888333	Pseudo-spectrum #147	Pseudo-spectrum #147	Name: Pseudo-spectrum #147;Isotope: [137]M+;Adduct: [M+H]+;470.305;identification method: Bioconductor CAMERA	1 DETECTED 488.3432137
2800	488.3432137	0.164888333	Pseudo-spectrum #147	Pseudo-spectrum #147	Name: Pseudo-spectrum #147;Isotope: [137]M+;Adduct: [M+H]+;470.305;identification method: Bioconductor CAMERA	1 DETECTED 488.3432137
2411	317.12934744	0.168388333	Pseudo-spectrum #176	Pseudo-spectrum #176	Name: Pseudo-spectrum #176;Isotope: [47]M+;Adduct: [M+Na]+;identification method: Bioconductor CAMERA	1 DETECTED 317.12934744
1731	361.216217	0.168388333	Pseudo-spectrum #147	Pseudo-spectrum #147	Name: Pseudo-spectrum #147;Isotope: [175]M+;Adduct: [M+Na]+;514.333;identification method: Bioconductor CAMERA	1 DETECTED 361.216217
3347	532.3635864	0.168388333	Pseudo-spectrum #147	Pseudo-spectrum #147	Name: Pseudo-spectrum #147;Isotope: [175]M+;Adduct: [M+H]+;514.333;identification method: Bioconductor CAMERA	1 DETECTED 532.3635864
4324	664.4437864	0.176581667	Pseudo-spectrum #366	Pseudo-spectrum #366	Name: Pseudo-spectrum #366;Isotope: [223]M+;Adduct: [M+H]+;330.207;identification method: Bioconductor CAMERA	1 DETECTED 664.4437864
401	166.1107025	0.209265	Pseudo-spectrum #024	Pseudo-spectrum #024	Name: Pseudo-spectrum #024;Isotope: [9]M+;Adduct: [M+H2]+;130.176;identification method: Bioconductor CAMERA	1 DETECTED 166.1107025
1514	331.2145081	0.209265	Pseudo-spectrum #024	Pseudo-spectrum #024	Name: Pseudo-spectrum #024;Isotope: [54]M+;Adduct: [M+H]+;278.184;identification method: Bioconductor CAMERA	1 DETECTED 331.2145081
1301	301.1683655	0.2174000	Pseudo-spectrum #024	Pseudo-spectrum #024	Name: Pseudo-spectrum #024;Isotope: [49]M+;Adduct: [M+H]+;203.166;identification method: Bioconductor CAMERA	1 DETECTED 301.1683655
2486	329.1959167	0.230191667	Pseudo-spectrum #027	Pseudo-spectrum #027	Name: Pseudo-spectrum #027;Isotope: [11]M+;Adduct: [M+H]+;338.228;identification method: Bioconductor CAMERA	1 DETECTED 329.1959167
1418	317.12934744	0.301091667	Pseudo-spectrum #205	Pseudo-spectrum #205	Name: Pseudo-spectrum #205;Isotope: [175]M+;Adduct: [M+Na]+;352.244;identification method: Bioconductor CAMERA	1 DETECTED 317.12934744
1846	275.1232449	0.301091667	Pseudo-spectrum #204	Pseudo-spectrum #204	Name: Pseudo-spectrum #204;Isotope: [175]M+;Adduct: [M+Na]+;352.244;identification method: Bioconductor CAMERA	1 DETECTED 275.1232449
964	265.1346113	0.309098333	Pseudo-spectrum #110	Pseudo-spectrum #110	Name: Pseudo-spectrum #110;Isotope: [25]M+;Adduct: [M+Na]+;176.176;identification method: Bioconductor CAMERA	1 DETECTED 265.1346113
1040	273.1654358	0.309098333	Pseudo-spectrum #110	Pseudo-spectrum #110	Name: Pseudo-spectrum #110;Isotope: [29]M+;Adduct: [M+Na]+;202.186;identification method: Bioconductor CAMERA	1 DETECTED 273.1654358
2949	502.3528137	0.328845	Pseudo-spectrum #93	Pseudo-spectrum #93	Name: Pseudo-spectrum #93;Isotope: [149]M+;Adduct: [M+Na]+;502.3528137	1 DETECTED 502.3528137
2545	458.3284302	0.446018333	Pseudo-spectrum #220	Pseudo-spectrum #220	Name: Pseudo-spectrum #220;Isotope: [126]M+;Adduct: [M+Na]+;458.3284302	1 DETECTED 458.3284302
1149	285.1659241	0.454478333	Pseudo-spectrum #254	Pseudo-spectrum #254	Name: Pseudo-spectrum #254;Isotope: [36]M+;Adduct: [M+Na]+;286.176;identification method: Bioconductor CAMERA	1 DETECTED 285.1659241
2988	500.3524760	0.4617085	Pseudo-spectrum #109	Pseudo-spectrum #109	Name: Pseudo-spectrum #109;Isotope: [11]M+;Adduct: [M+Na]+;352.244;identification method: Bioconductor CAMERA	1 DETECTED 500.3524760
1333	375.2323439	0.471685	Pseudo-spectrum #306	Pseudo-spectrum #306	Name: Pseudo-spectrum #306;Isotope: [73]M+;Adduct: [M+Na]+;352.244;identification method: Bioconductor CAMERA	1 DETECTED 375.2323439
506	182.9840951	0.602060667	Pseudo-spectrum #045	Pseudo-spectrum #045	Name: Pseudo-spectrum #045;Isotope: [31]M+;Adduct: [M+Na+COOH]+;92.005;identification method: Bioconductor CAMERA	1 DETECTED 182.9840951
3797	614.1139038	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [209]M+;Adduct: [M+Na]+;614.1139038	1 DETECTED 614.1139038
7173	1106.085693	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [334]M+;Adduct: [M+Na]+;1106.085693	1 DETECTED 1106.085693
4585	714.5251465	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [233]M+;Adduct: [M+Na]+;714.5251465	1 DETECTED 714.5251465
4618	717.0214844	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [237]M+;Adduct: [M+Na+K2+]++;1032.82;identification method: Bioconductor CAMERA	1 DETECTED 717.0214844
4661	722.1855469	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [237]M+;Adduct: [2M+Na+K2+]++;1032.82;identification method: Bioconductor CAMERA	1 DETECTED 722.1855469
3294	579.1139038	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [36]M+;Adduct: [M+Na]+;579.1139038	1 DETECTED 579.1139038
4111	630.3843694	0.645143333	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [24]M+;Adduct: [M+Na]+;630.3843694	1 DETECTED 630.3843694
4431	682.2994263	0.653176667	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [225]M+;Adduct: [M+H]+;681.096;identification method: Bioconductor CAMERA	1 DETECTED 682.2994263
4617	716.8546753	0.653176667	Pseudo-spectrum #049	Pseudo-spectrum #049	Name: Pseudo-spectrum #049;Isotope: [234]M+;Adduct: [2M+2N+K+]++;1032.82;identification method: Bioconductor CAMERA	1 DETECTED 716.8546753

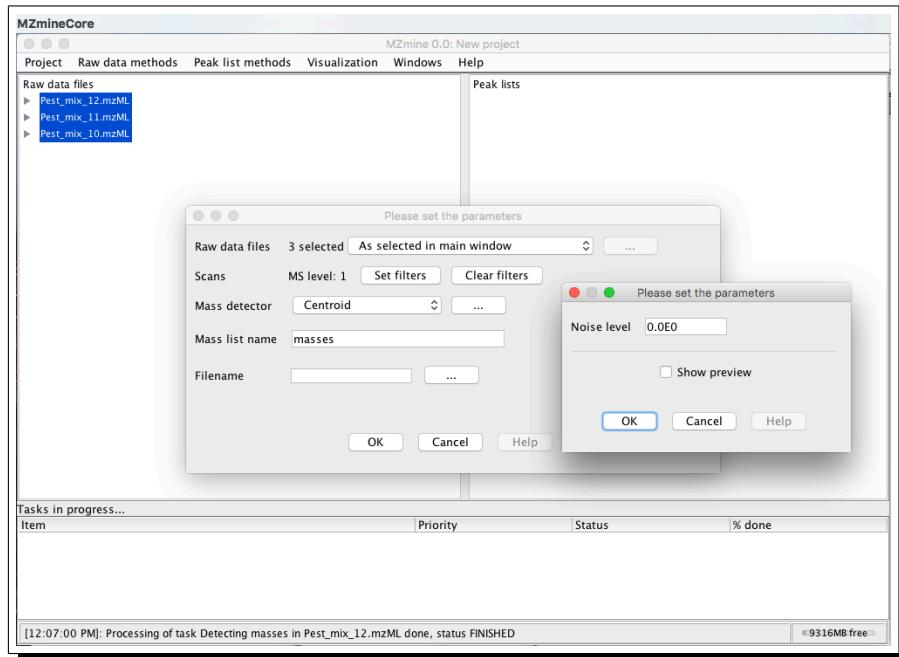
Figure 24: Exported results.

## 4 ADAP-GC

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

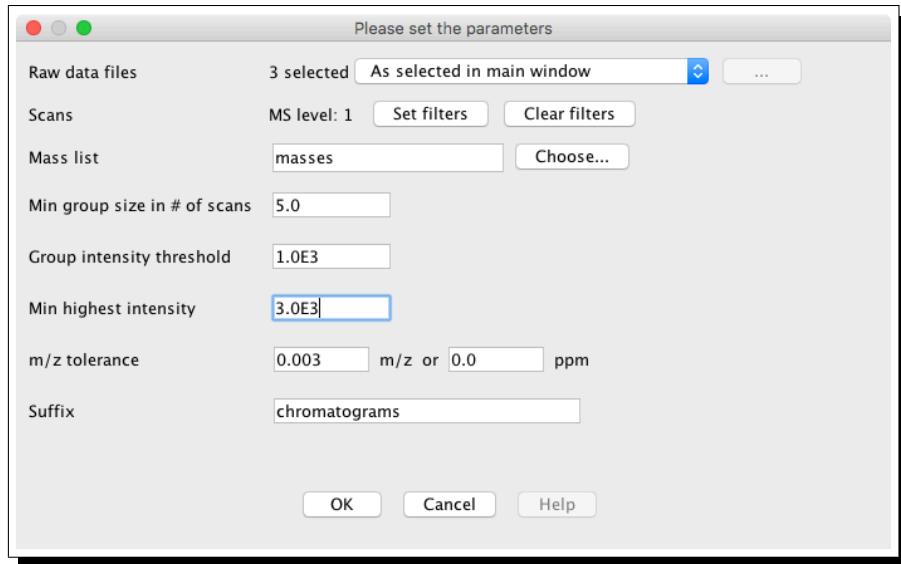
### 4.1 Detection of Masses and Construction of EICs

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 25.



**Figure 25:** Mass detection of centroid data.

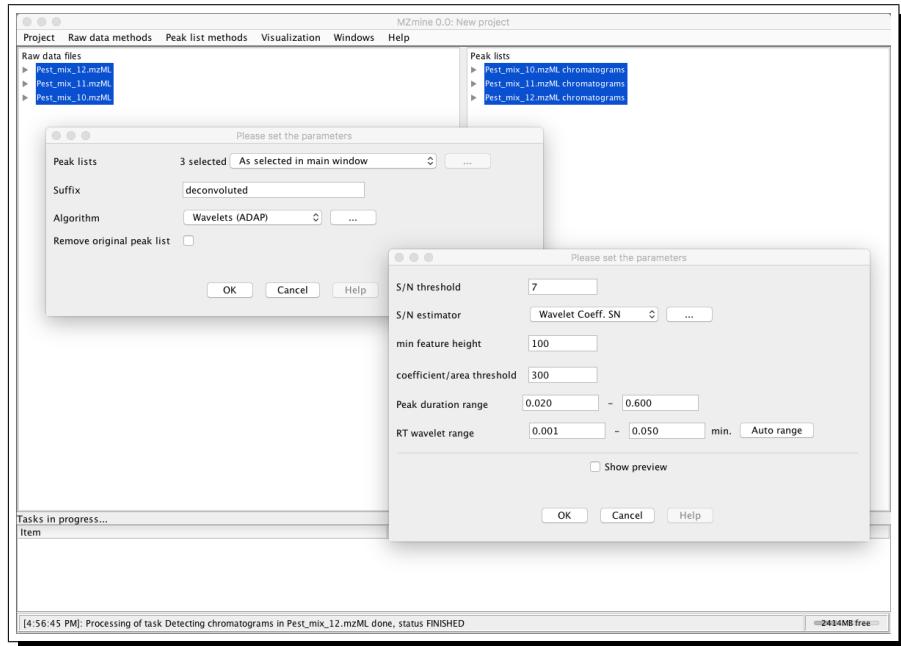
Parameters for constructing EICs are shown in Figure 26.



**Figure 26:** 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* algorithm as shown in Figure 27.



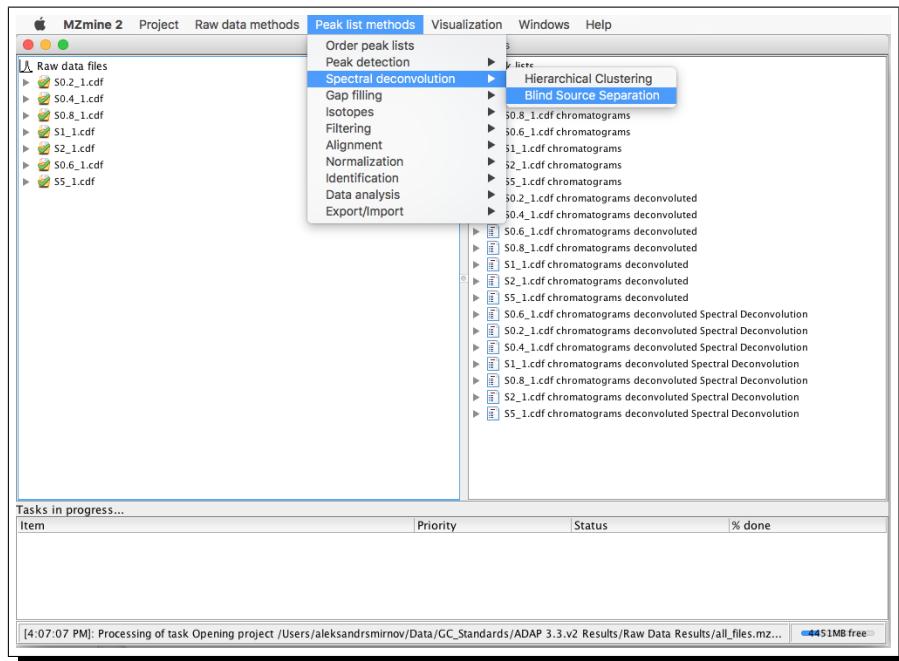
**Figure 27:** Select *Wavelet (ADAP)* for detecting peaks from EICs for GC-MS data.

Click the ellipse to open the parameter window. Figure 27 shows example parameters. Click *OK* to start peak detection. If the duration of chromatography is long, this step could take a while.

### 4.3 Spectral Deconvolution

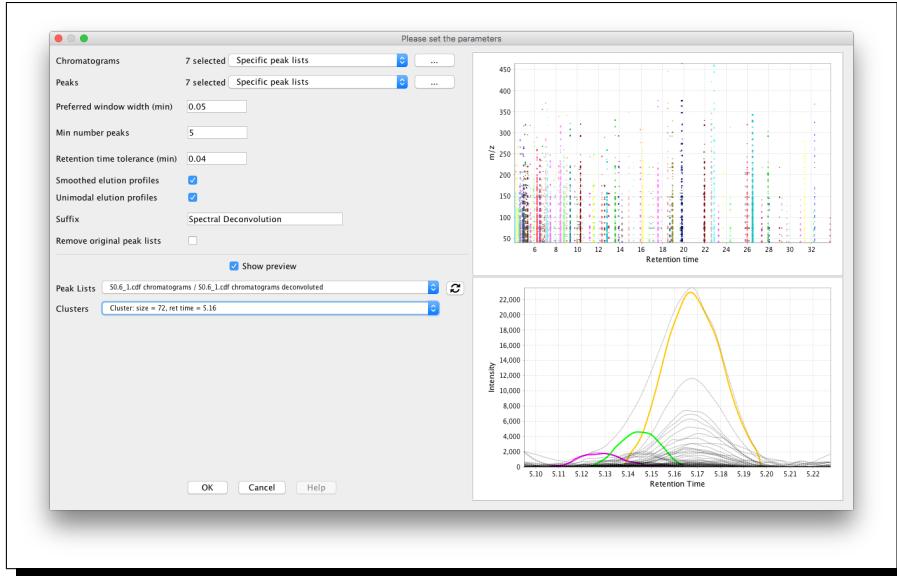
The term *Spectral Deconvolution* refers to detecting analytes by combining similar peaks into clusters and using their intensities to construct fragmentation mass spectra. Detection of analytes is performed by two clustering steps and one filtering step in between. Correspondingly, first all peak are combined into clusters based on their retention times. Then, model peaks are determined, that best describe the peaks in a cluster. Finally, all peaks in a cluster are decomposed into linear combination of the model peaks and their fragmentation mass spectra are constructed. Choice of the model peaks may affect the quality of the constructed fragmentation spectra.

To perform deconvolution, click *Peak list methods* → *Spectral Deconvolution* → *Blind Source Separation* as shown in Figure 28. To see preview of the deconvolution results, select *Show preview* option at the bottom of the parameter window.



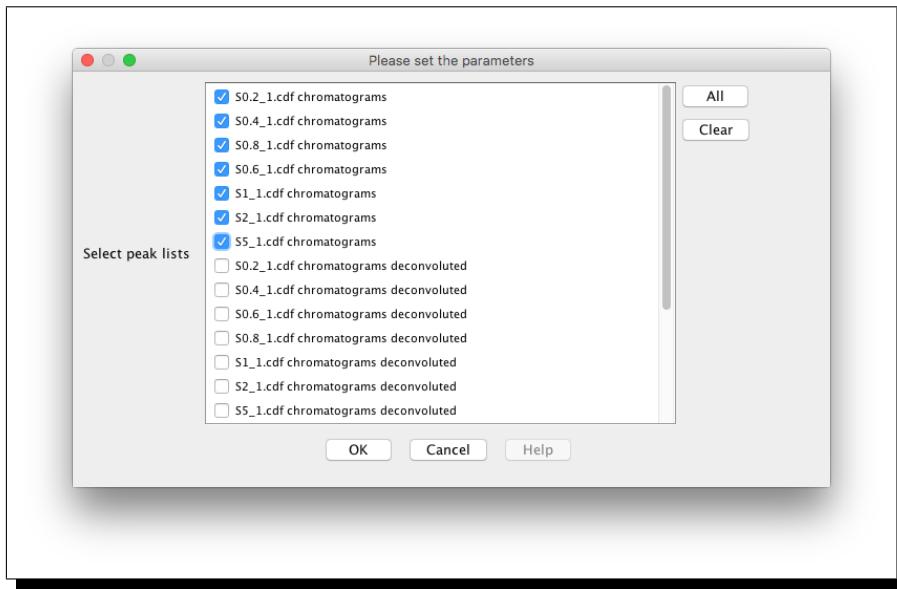
**Figure 28:** Deconvolution of chromatographic peaks.

A window as shown in Figure 29 pulls up allowing you to specify parameters for deconvolution.

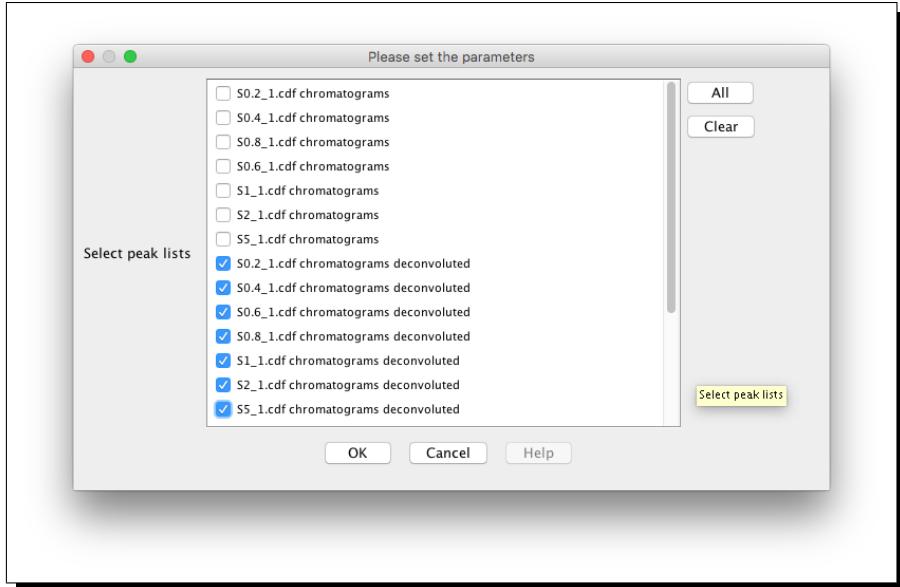


**Figure 29:** Specify parameters for decomposition of chromatographic peaks.

The spectral deconvolution uses both constructed chromatograms and detected peaks. The list of constructed chromatograms is specified by selecting *Specific peak lists* for parameter *Chromatograms*, clicking on the ellipsis button, and choosing one or more lists with chromatograms in the popup window (Figure 30). The list of detected peaks is specified by selecting *Specific peak lists* for parameter *Peaks*, clicking on the ellipsis button, and choosing one or more lists with detected peaks in the popup window (Figure 31).



**Figure 30:** Choosing chromatograms for Spectral Deconvolution.



**Figure 31:** Choosing peaks for Spectral Deconvolution.

The Spectral Deconvolution consists of two steps:

1. Entire retention time interval is split into deconvolution windows so that
  - Peaks produced by the same component or by coeluting components belong to the same deconvolution window,
  - Number of peaks in deconvolution window is significantly smaller than the total number of peaks.

The deconvolution windows are displayed in the top plot of the preview (see Figure 29), where small dash lines denote peaks in the (retention time, m/z)-plane, and peaks belonging to one deconvolution window have the same color. The vertical sequences of peaks usually mark the presence of one or several compounds, so it is important that those peaks are assigned to the same deconvolution window, i.e. they have the same color on the plot. On the other hand, if deconvolution windows contain too many peaks, it will significantly slow down the spectral deconvolution computations, so the deconvolution windows should be as short (in the retention time domain) as possible.

The following two parameters control the deconvolution window selection:

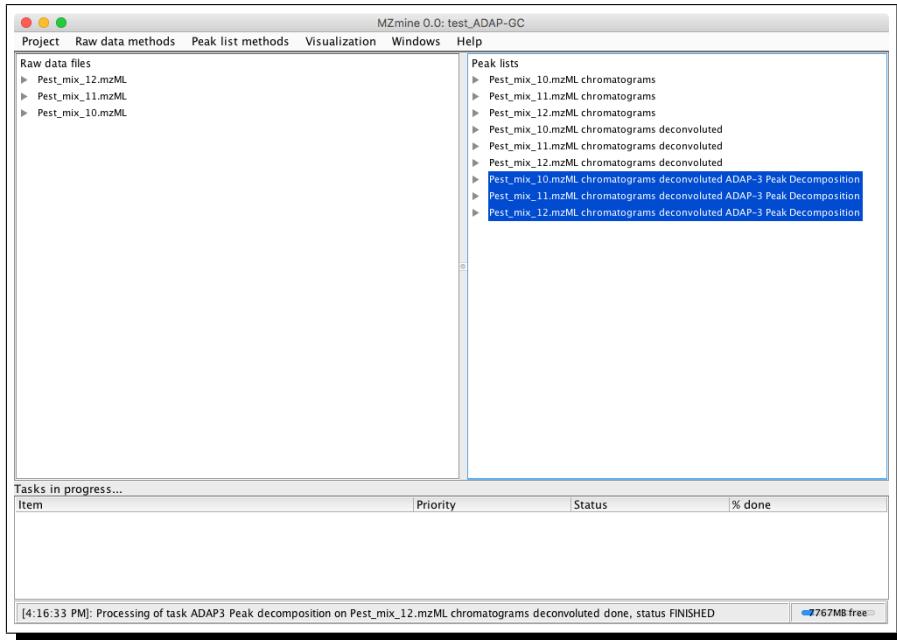
- *Preferred window width (min).* The algorithm will try to produce deconvolution windows so that their width (in retention time domain) does not exceed the value of this parameter. However, the algorithm can fail to produce deconvolution windows of the specified width if there are too many peaks. The recommended value of this parameter is about two peak widths. In our tests, we used value 0.5 min.
- *Min number of peaks.* Each deconvolution window will contain at least the specified number of peaks. The value of this parameter depends on the total number of detected peaks. It can be as small as 5 or as big as 50. Higher values of this parameter produce smaller deconvolution windows.

2. The algorithm infers the number of components in each deconvolution window and construct their elution profiles and fragmentation spectra. The inferred number of components is controlled by parameter *Retention time tolerance (min)*, which is the smallest time-gap between any two components. The value of this parameter should be a fraction of the average peak width. In our tests, we use 0.04 min.

The shape of constructed elution profiles is constrained by the following two parameters:

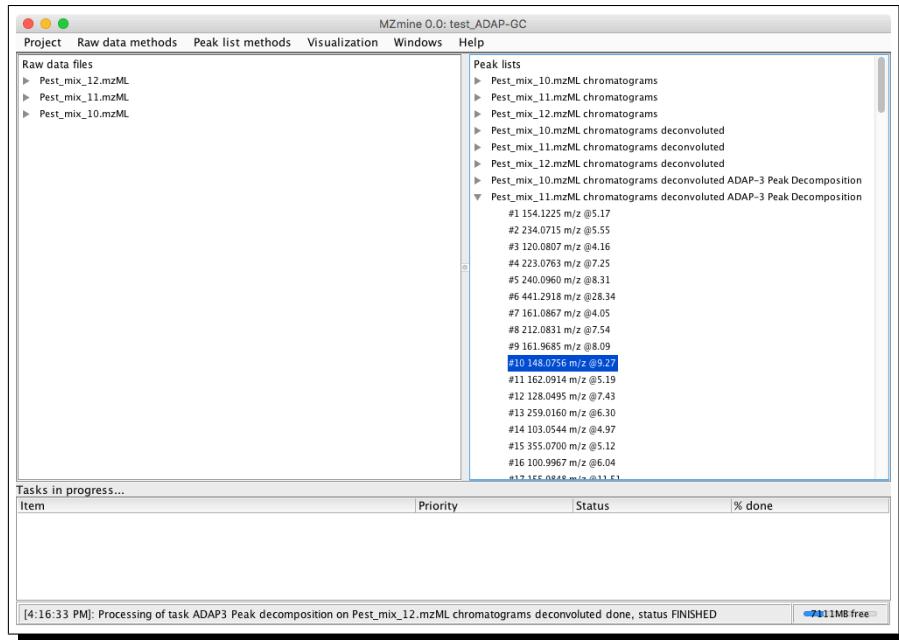
- *Smoothed elution profiles*. If checked, elution profiles of components are smoothed using the moving average algorithm,
- *Unimodal elution profiles*. If checked, each elution profile has only one local maximum.

After spectral deconvolution is finished, the results are displayed as shown in Figure 32.



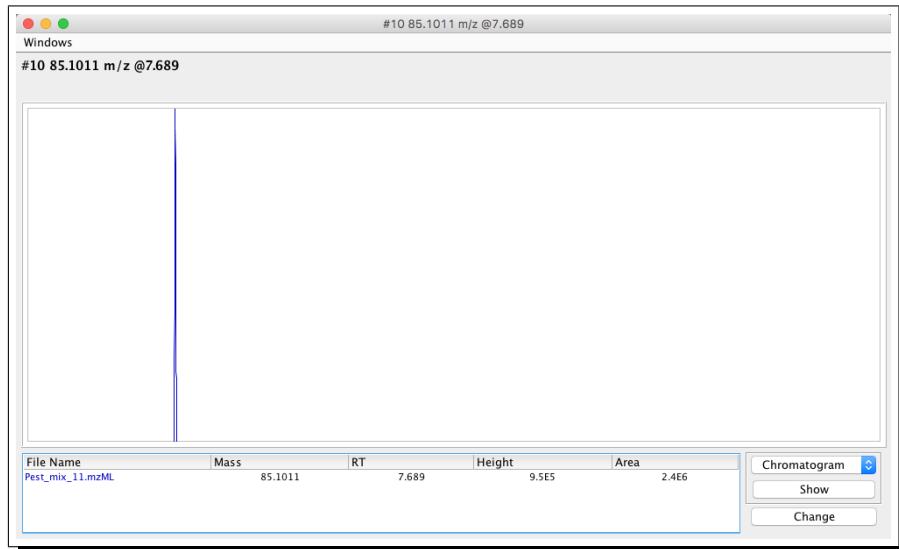
**Figure 32:** Decomposition results.

Expand the 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 33). The  $m/z$  for each entry is the  $m/z$  of the model peak for this spectrum.



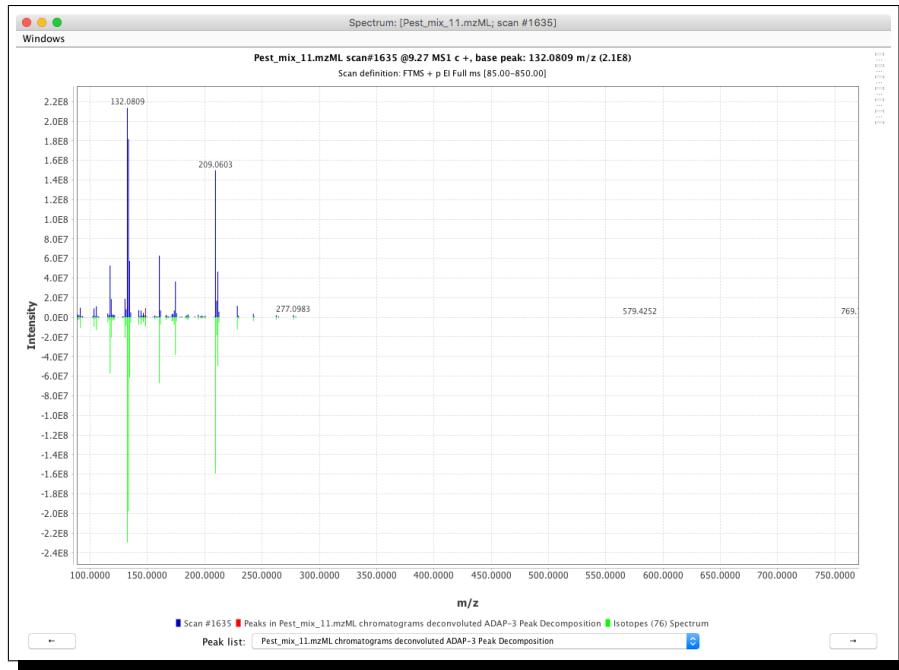
**Figure 33:** List of mass spectra constructed by the decomposition algorithm.

Double click on a particular mass spectrum will pull up a window as shown in Figure 34.



**Figure 34:** Peak information window.

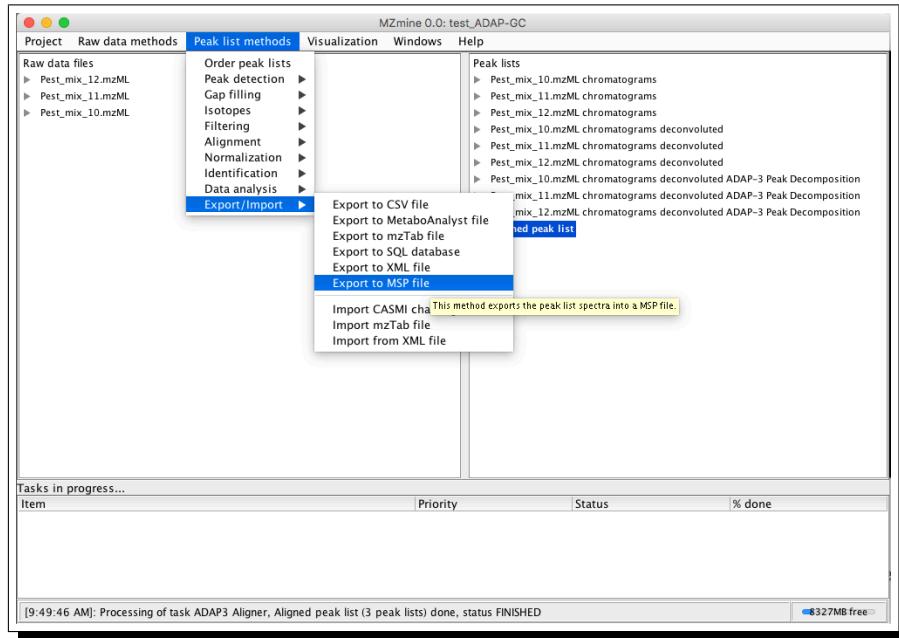
Click on the data file name and then select *Mass spectrum* in the drop-down menu on the right. The spectrum that has been constructed (green) in the context of the raw spectrum (blue) is displayed (Figure 35).



**Figure 35:** Mass spectra constructed by the decomposition algorithm.

#### 4.4 Spectra Export

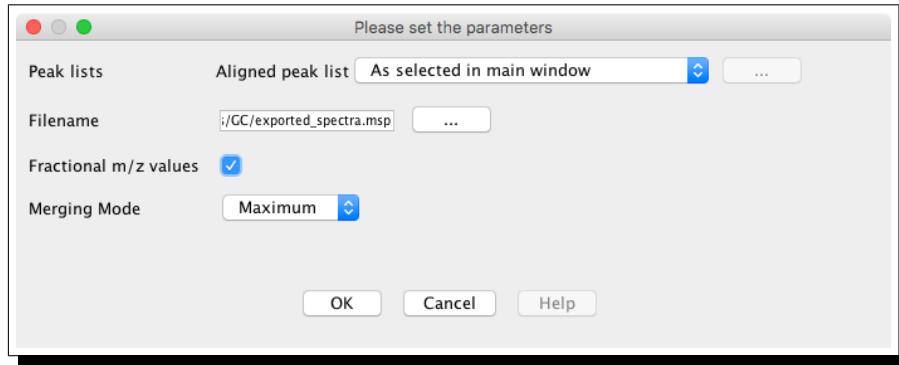
The mass spectra that have been constructed can be exported in .msp format and then imported to *NIST MS Search* for identification. To export the spectra, select the *Aligned peak list* and then click *Peak list methods* → *Export/Import* → *Export to MSP file* as shown in Figure 36.



**Figure 36:** Export mass spectra to a MSP file.

A window as shown in Figure 37 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 37:** 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 38 shows a small portion of the .msp file.

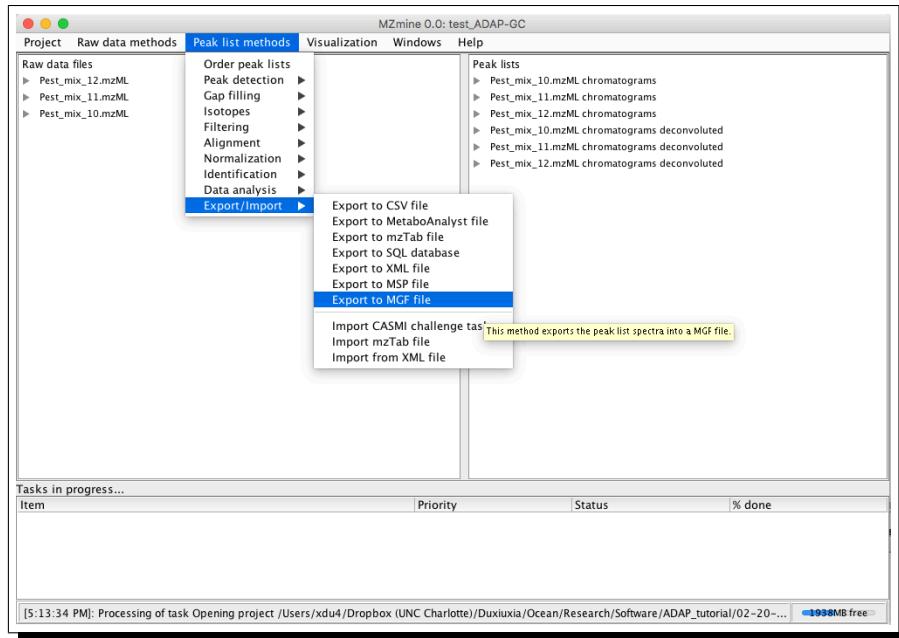
```

Name: #1 281.0515 m/z @4.10 (Alignment Score = 0.9531350425546566)
DB#: 1
Num Peaks: 62
92.02568054199219 287278.71075470065
105.02923583984375 874378.1240155919
106.04122161865234 81468.73115179865
123.06768035888672 245995.74908396933
149.04736328125 2572571.3844742966
150.04652404785156 234995.55398492888
164.94801338566486 34306.16408879444
178.96493538273438 86455.19008466511
191.001220703125 278791.58316731517
192.988019409179688 266150.5970837946
205.01695251464844 50189.81177968083
207.03268432617188 475207.5692497589
209.01161193847656 241779.2090340591
221.08436584472656 1064530.8474890182
223.0816650390625 99494.0289137579
225.04273986016406 251728.76147294548
236.9520263671875 59029.37689508891
239.0953826904297 184586.40998280022
248.98899841308594 289881.40850885626
252.98326110839844 47241.12614709242

```

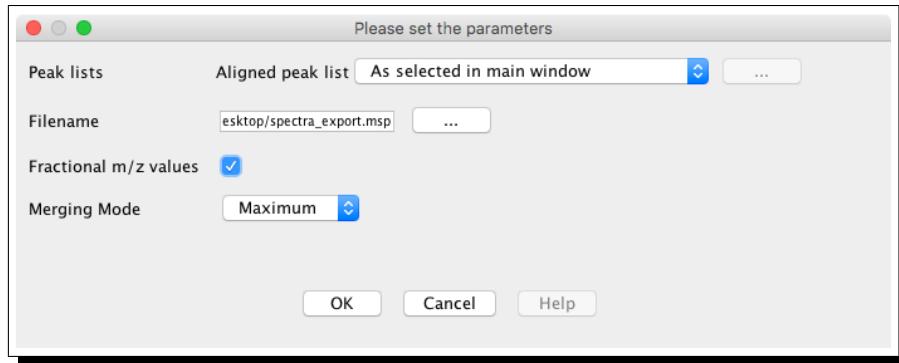
**Figure 38:** Example .msp file exported by ADAP-GC.

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 39.



**Figure 39:** Export mass spectra to a MGF file.

A window as shown in Figure 40 is open allowing you to name the export file.



**Figure 40:** Export mass spectra to a MGF file.

Figure 41 shows part of a .mgf file exported from MZmine 2.

```

spectra_export.mgf
BEGIN IONS
FEATURE_ID=1
PEPMASS=252.32110850016275
RTINSECONDS=181.004754
SCANS=1
MSLEVEL=2
CHARGE=1+
85.1011734008789 236595.53642372743
91.05420684814453 1592361.6816624396
92.06222534179688 193380.26696339928
93.07002258300781 1017302.1767548586
94.0777359008789 247600.75240741248
95.0857162475586 1377785.6425238154
96.0935287475586 249896.40878734837
97.10134887695312 573715.6438461392
98.07260131835938 499856.7958652754
102.04652404785156 46205.2682039636

```

**Figure 41:** Example .mgf file exported by ADAP.

## **5 Batch Processing**

Create the parameter file for batch processing.

On Mac, run `./startMZmine_MacOSX.command` ‘‘path to and name of the batch processing file’’ in the terminal.

## 6 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. The checkbox allows the user to choose if they would like to output this file or not.
  - **ADAP Chromatogram builder:** a new method of chromatogram building.
- Category: *Peak list methods* → *Peak Detection*
  - **Chromatogram Deconvolution: Wavelets (ADAP):**
  - **Spectral deconvolution:** a new method for pre-processing GC-MS data by detecting analytes and constructing their fragmentation spectra.
- Category: *Peak list methods* → *Identification*
  - **CAMERA search:** Modified CAMERA search.
- Category: *Peak list methods* → *Export / Import*
  - **Export to MSP file:** exporting constructed spectra to a file in MSP format
  - **Export to MGF file:** exporting constructed spectra to a file in MGF 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.