

Oct 30, 2019

Version 2

# METABOLIC PATHWAY ANALYSIS BY LIQUID CHROMATOGRAPHY (UHPLC) COUPLED TO HIGH RESOLUTION MASS SPECTROMETRY (LC/MS) V.2

DOI

[dx.doi.org/10.17504/protocols.io.8vjhw4n](https://dx.doi.org/10.17504/protocols.io.8vjhw4n)

Angelica S. Gouveia<sup>1</sup>, Lucas L. Lima<sup>1</sup>, Flaviane S. Coutinho<sup>1</sup>, Juliano M. Rodrigues<sup>1</sup>, Valquiria J. M. Pinheiro<sup>1</sup>, Maria Eduarda S. Ramos<sup>1</sup>, Camilo E. Vital<sup>1</sup>, Claudia S.L. Pontes<sup>2</sup>, Daniella P. Pinheiro<sup>2</sup>, Pedro M. Vidigal<sup>2</sup>, Edvaldo Barros<sup>2</sup>, Humberto J O Ramos<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology - Universidade Federal de Viçosa - UFV, BIOAGRO/INCT-IPP - Viçosa-MG Brazil;

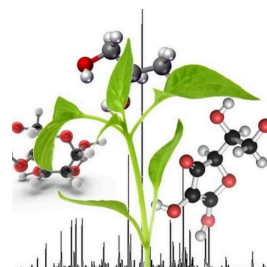
<sup>2</sup>Center for Biomolecules Analysis (NuBioMol) Universidade Federal de Viçosa - UFV Viçosa-MG Brazil

Metabolomics Protocols & Workflows  
Tech. support email: [bbmisraccb@gmail.com](mailto:bbmisraccb@gmail.com)



**Humberto J O Ramos**

Universidade Federal de Viçosa



## Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

[Create free account](#)

OPEN  ACCESS



**DOI:** <https://dx.doi.org/10.17504/protocols.io.8vjhw4n>

**Protocol Citation:** Angelica S. Gouveia, Lucas L. Lima, Flaviane S. Coutinho, Juliano M. Rodrigues, Valquiria J. M. Pinheiro, Maria Eduarda S. Ramos, Camilo E. Vital, Claudia S.L. Pontes, Daniella P. Pinheiro, Pedro M. Vidigal, Edvaldo Barros, Humberto J O Ramos 2019. METABOLIC PATHWAY ANALYSIS BY LIQUID CHROMATOGRAPHY (UHPLC) COUPLED TO HIGH RESOLUTION MASS SPECTROMETRY (LC/MS). **protocols.io** <https://dx.doi.org/10.17504/protocols.io.8vjhw4n>

**License:** This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

**We use this protocol and it's working**

**Created:** October 30, 2019

**Last Modified:** October 30, 2019

**Protocol Integer ID:** 29323

**Keywords:** Metabolomic, Plant, metabolite extraction, workflow, XCMS, MetaboAnalyst , metabolic pathway analysis by liquid chromatography, comparison of the metabolite profile, metabolomic, global analysis of the metabolome, metabolite profile, metabolic pathway analysis, metabolome, mass spectrometry, metabolite, high resolution mass spectrometry, coupled to high resolution mass spectrometry, ms data for determination, liquid chromatography, several samples by lc, functional enrichment analyze, small molecules present in biological system, automatic processing for both spectra, ms data

## Abstract

Plants produce a variety of compounds with diverse structure and abundance that play an important role in the development, growth and response to the environment. Metabolomics involves studies of a great number of metabolites, which are small molecules present in biological systems. Metabolite profiles can be obtained from several samples by LC/MS (liquid chromatography coupled to mass spectrometry). However, the large number of ions detected for each run from different treatments and replicates require automatic processing for both spectra (metabolite identification and quantification) and global analysis of the metabolome. Thus, this protocol describes in detail, step-by-step, the comparison of the metabolite profiles from LC/MS data for determination of dysregulated pathways, as well as for Exploratory Statistical and Functional Enrichment Analyzes.

## Materials

### REAGENTS

- Methanol (LiChrosolv, Merck, Order n° 1.06001 )
- Acetonitrile (LiChrosolv, Merck , Order n° )
- Acetic Acid (Sigma)
- High pure water (18.2M  $\Omega\text{cm}^{-1}$ ) provided by a Milli-Q system (Burlington, Massachusetts, USA)
- Liquid Nitrogen

### EQUIPMENTS AND SUPPLIES

- Liquid Chromatography System coupled to mass spectrometry (LC/MS) with the following specifications:  
NanoAcquity UHPLC (Waters) coupled to Mass Spectrometric Micro-Tof QII (Bruker)
- Capillary column ProteCol GHQ303 C18 3,0  $\mu\text{m}$  – 300  $\mu\text{m}$   $\times$  150 mm
- Thermomixer (benchtop laboratory incubator)
- Benchtop centrifuge
- Sppedvac (vacuum centrifugation drier)
- Ultra-freezer
- Benchtop balance
- Mortar and pestle
- Vials, caps and septa.
- Microtubes and tips

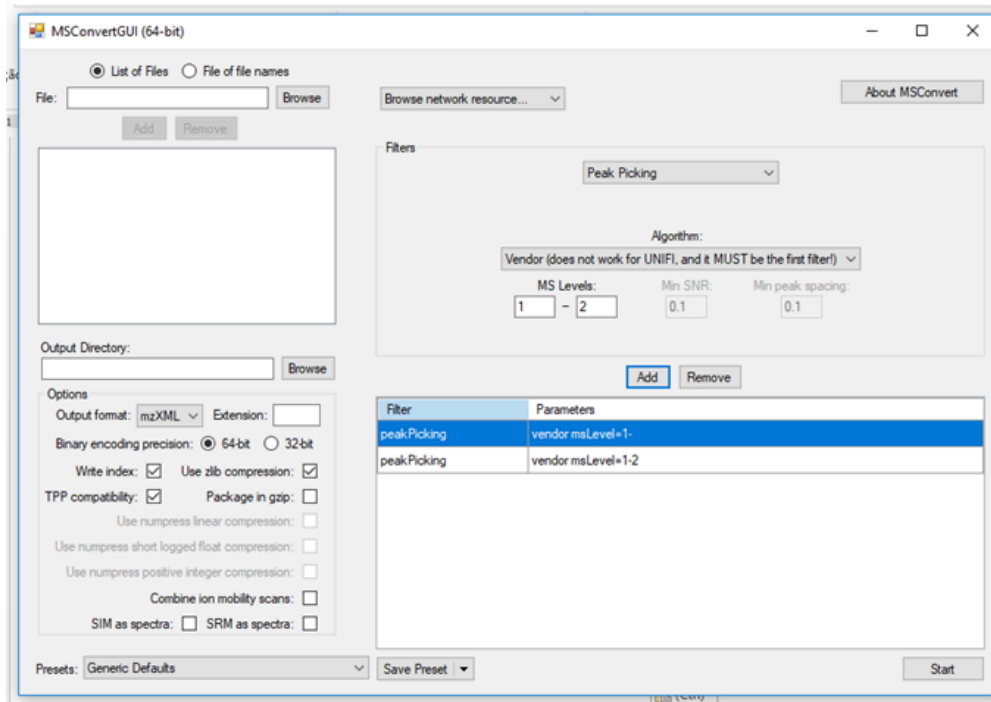
## Troubleshooting

## METABOLITE EXTRACTION AND ANALYSIS BY LC/MS

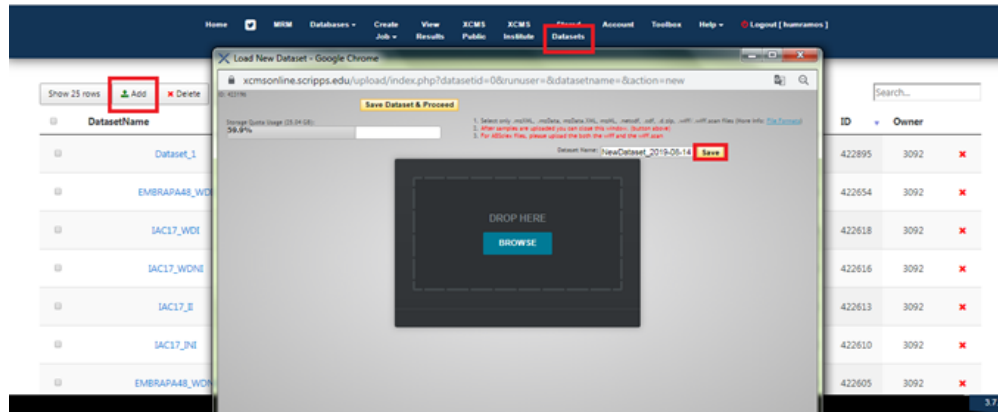
- 1 **1)** Collect samples of plant tissues, immediately freeze in liquid nitrogen and store them in freezer -80°C until use.  
  
**2)** Macerate the samples in liquid nitrogen using mortar and pestle. Do not allow to thaw. Weigh approximately 150mg of each sample into microtubes (2ml) and annotate the weight (used for normalization ).  
Note: Always use microtubes and tips of good quality.  
  
**3)** Add 500 µl of extractive solution (80% methanol/ 0,1% formic acid) and vortex for 20 seconds and centrifuged by 14,000 g for 20 min.  
  
**4)** After centrifugation, filter the supernatant using a membrane of PVDF 0.22 µm and transfer 80 µL of the solution to a vial.  
  
**5)** Inject an aliquot of 10 µL for analysis through Nano Liquid Chromatography – Mass Spectrometry (nanoLC/MS) using the nanoACQUITY UPLC system (Waters, Milford, MA, USA), containing a trap column and a capillary column ProteCol GHQ303 C18 3,0 µm – 300 µm × 150 mm, operating at a flow rate of 5.0µL.min<sup>-1</sup>, online mode with a microESI ionization needle. This step consisted of mobile phase solutions used for the gradient program, as follows: **(A)** water and 0.02% acetic acid (v/v) and **(B)** acetonitrile and 0.02% acetic acid (v/v). The following gradient program was used: a linear rising ramp starting at 5%, increasing to 50% (B) for 30 min, 50% (B) for 5 min; linear rising ramp starting at 50%, increasing to 90% of (B) for 3 min; 90% (B) for 2 min, linear gradient descent starting at 90%, decreasing to 10% (B) for 3 min, followed by a steady condition at 10% (B) for 3 min.  
  
**6)** Ion scanning for MS1 spectra in positive or negative mode can be carried out for masses ranging between 100 and 1000 m/z, and between 50 and 1000 m/z for the MS2 spectra. The data were acquired over 57.0 min in each LC-MS/MS analysis, using the Hystar software program, version 3.2 (Bruker Daltonics, Bremen, Germany) and the spectra were processed through the DataAnalysis software program, version 4.0 (Bruker Daltonics, Bremen, Germany), using the default settings for metabolomics. The mass spectrometer was operated in Auto-MSn mode, which collected MS2 spectra for the most intense ions in each whole scan spectrum.

## DATA PROCESSING AND LC/MS ALIGNMENT

- 2
- 1) Install the **Proteowizard package** for your operational system (**32 bits or 64 bits**).
  - 2) Convert the data to mzXML format using ProteoWizard by executing MSConvertGUI.exe.
  - 3) Locate the directories to input all spectra and output; choose filters **Peak Picking** and MS level 1, click **add** and repeat this step for **MS levels 1-2**; Add samples to **Browse** and click **Start**. Wait for all jobs to be completed



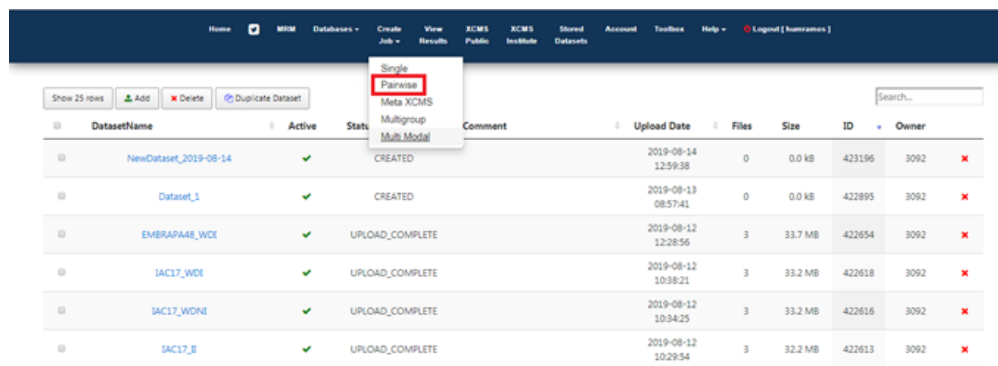
- 4) Go to the output directory and rename the spectrums **\*\*\*.mzXML** and group the replicates in subdirectories according to the treatments.
- 5) Go to the XCMS online website <https://xcmsonline.scripps.edu> and create an account.
- 6) To enter your data click **Dataset >>> Add** and click **upload (Browse)** the **\*\*\*.mzXML** spectra from biological replicates for each treatment one at a time, this guarantees that all the replicates will be grouped in a directory for each treatment



7) Rename the **Dataset Name** according to the treatment (use an abbreviation in English related to the treatment). Wait for the upload to finish and click **Save**.

8) Repeat this procedure for all the treatments.

9) Run the *pairwise* or *multigroup* method: Click **Create job >>> Pairwise**.



10) Click **Select dataset**: click in the directory for the first treatment (example "IAC17 NI – non infested").



**11)** Click **Next** >>> select the second treatment >>> **Next**.

**12)** Select the Parameters >>> click **HPLC/Q-TOF**>>> click **Customized** >>> click **View and Edit** >>>.

**Note:** estimate the m/z accuracy and the maximum peak width for your LC/MS system. In the module “identification”, you need search for the presence of the metabolite library of the organism in study.



## View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General **Feature Detection** Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous

Method: centWave

Option	Value	Note:
ppm	25	maximal tolerated m/z deviation in consecutive scans, in ppm (parts per million)
minimum peak width	10	minimum chromatographic peak width in seconds note: must be less than max peak width. See also <a href="#">here</a> .
maximum peak width	60	maximum chromatographic peak width in seconds note: must be greater than min peak width. See also <a href="#">here</a> .
View Advanced Options		
mzdiff	0.01	minimum difference in m/z for peaks with overlapping retention times, can be negative to allow overlap
Signal/Noise threshold	6	Signal/Noise threshold
Integration method	2	Integration method. If =1 peak limits are found through descent on the mexican hat filtered data, if =2 the descent is done on the real data. Method 2 is very accurate but prone to noise, while method 1 is more robust to noise but less exact.
prefilter peaks	3	Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
prefilter intensity	500	Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]
Noise Filter	0	optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity < noise are omitted from ROI detection

**Save** **Create New** **Delete** **Cancel**

## View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection **Retention Time Correction** Alignment Statistics Annotation Identification Visualization Miscellaneous

Method: obiwarp

Option	Value	Note:
profStep	0.5	step size (in m/z) to use for profile generation from the raw data files

**Save** **Create New** **Delete** **Cancel**

## View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.

The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction **Alignment** Statistics Annotation Identification Visualization Miscellaneous

Option	Value	Note:
bw	5	Allowable retention time deviations, in seconds. In more detail: bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram
minfrac	0.5	minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group
mzwid	0.025	width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples
View Advanced Options		
minsamp	1	minimum number of samples necessary in at least one of the sample groups for it to be a valid group
max	100	maximum number of groups to identify in a single m/z slice

**Save** **Create New** **Delete** **Cancel**



### View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.  
The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment **Statistics** Annotation Identification Visualization Miscellaneous

Option	Value	Note:
Statistical test	Unpaired non-parametric (Mann-Whitney)	Statistical test method. Welch t-test (unequal variances) or Wilcoxon Rank Sum test. The selected statistical test is performed as a paired test. The sample pairs need to be specified.
Perform paired test		
Perform post-hoc analysis	True	Perform post-hoc analysis [multigroup only].
p-value threshold (highly significant features)	0.01	Features with a p-value less than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
fold change threshold (highly significant features)	1.5	Features with a fold change greater than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
p-value threshold (significant features)	0.05	Features with a p-value less than this threshold are not considered significant and are omitted from some calculations to save time and space. EICs, annotations and database ID's are not generated for features with p-values above this threshold.
View Advanced Options		
value	into	intensity values to be used for the diffreport. If value="into", integrated peak intensities are used. If value="maxo", maximum peak intensities are used.
Normalization	Median fold change	Normalize the intensity values by either probabilistic quotient or cyclic loess normalization.

**Save** **Create New** **Delete** **Cancel**

### View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.  
The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment **Statistics** **Annotation** Identification Visualization Miscellaneous

Option	Value	Note:
ppm	20	ppm error
m/z absolute error	0.015	m/z absolute error
Search for	isotopes	Search for 1.) just isotopic features or 2.) isotopic features and adducts formations, dimers, trimers, neutral losses, etc. WARNING: searching for all adducts can increase the total processing time by approximately 50 %

**Save** **Create New** **Delete** **Cancel**

### View/Edit Parameters for Job

Polarity is defined on the General tab and will affect values on the Annotation and Identification (adducts) tabs. Job results will be misleading if this value is not correctly defined.  
The current parameter set is read-only. Use **Create New** button below to modify parameters to suit your job.

General Feature Detection Retention Time Correction Alignment **Statistics** **Annotation** **Identification** Visualization Miscellaneous

Option	Value	Note:
ppm	20	tolerance for database search
adducts	[M+H] <sup>+</sup> [M+H+4] <sup>+</sup> [M+Na] <sup>+</sup> [M+H+H <sub>2</sub> O] <sup>+</sup> [M+H-2H <sub>2</sub> O] <sup>+</sup> [M+K] <sup>+</sup> [M+ACN+H] <sup>+</sup> [M+ACN+Na] <sup>+</sup> [M+2Na+H] <sup>+</sup> [M+2H] <sub>2</sub> <sup>+</sup>	adducts to be considered for database search
sample biosource	SELECTED: ARA	Select your species/cell line, etc. that correspond to your samples. Default human.
pathway ppm deviation	20	metabolite pathway lookup
input intensity threshold		minimum intensity cut-off for pathway analysis
significant list p-value cutoff	AUTO	significant list p-value cut-off

**Save** **Create New** **Delete** **Cancel**

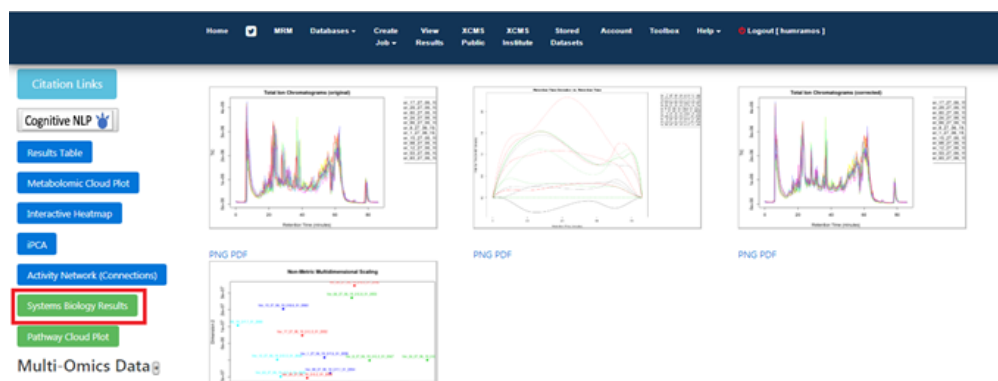
**13) Click Save.** Back to the module **"identification"** and verified if **the metabolite library of the organism in study was selected!! Click and save using a name "myXCMS parameter" for your customized method.**

Click **Next>>>>** click **Submit**. The job will **start**

**14) Go to View Results >>>> You will see the job progress. If you see the percentage values, the job is running without error. Wait and click **View result** to see the results of the completed job**

EXPTYPE	Status	JobID	Progress	JobName	Datasets / Sources	Created	Parameters (ID#)	Group	Share	Delete
PAIR	<a href="#">VIEW</a>	1358056	job complete	P_2019-08-13_09:25	IAC7_I1 (#422613) IAC7_WD1 (#422618)	2019-08-13 09:25:12	56806			
MULTI	<a href="#">VIEW</a>	1357827	job complete	MC_2019-08-12_13:42	IAC7_I1 (#422610) IAC7_I2 (#422613) IAC7_WDNI (#422616)	2019-08-12 13:42:41	56806			
MULTI	<a href="#">VIEW</a>	1357814	job complete	MC_2019-08-12_12:50	EMBRAPAAB (#422603) EMBRAPAAB (#422604) EMBRAPAAB (#422605)	2019-08-12 12:50:22	56806			
MULTI	<a href="#">VIEW</a>	1357784	job complete	MC_2019-08-12_11:07	BR16_I1 (#422588) BR16_I2 (#422589) BR16_WDNI (#422599)	2019-08-12 11:07:37	56806			
PAIR	<a href="#">VIEW</a>	1357593	job complete	P_2019-08-11_16:16	IAC7_I1 (#384773) IAC7_I2 (#384778)	2019-08-11 16:15:55	56806			
PAIR	<a href="#">VIEW</a>	1357066	job complete	P_2019-08-08_15:58	C_confirm (#421663) fungo_conf (#421673)	2019-08-08 15:58:55	56806			

**15) Download your full results using the link in the upper of the page.**



Click **System Biology Results >>>> click the column **Overlapping putative metabolites** to classify in descending order.**

**>>>> click the column **Overlapping putative metabolites** to classify in descending order. You could click in the **link** to see the **metabolite information**for each **pathway** enriched.**

Or **select and copy** the all table and **edit** in **Excel** and **Word Editor** for include the more significant and relevant informations such as:

Pathway	Overlapping putative metabolites <sup>1</sup>	All metabolites <sup>2*</sup>	p-values
flavonoid biosynthesis (in equisetum)	9	10	3.4e-4
kaempferol glycoside biosynthesis (Arabidopsis)	9	9	3.3e-4
sanguinarine and macarpine biosynthesis	8	13	9.9e-4
brassinosteroid biosynthesis I	7	17	2.2e-2
cytokinins 7-N-glucoside biosynthesis	6	10	2.2e-3
flavonoid biosynthesis	6	8	7.6e-4
2,3-cis-flavanols biosynthesis	5	5	5.0e-4
cytokinins 9-N-glucoside biosynthesis	5	8	2.8e-3
gibberellin inactivation I (2 $\beta$ -hydroxylation)	5	19	5.5e-1
leucopelargonidin and leucocyanidin biosynthesis	5	6	8.0e-4
matairesinol biosynthesis	5	9	1.2e-1
gibberellin biosynthesis I (non C-3, non C-13 hydroxylation)	4	9	3.0e-2
glucosinolate biosynthesis from hexahomomethionine	4	4	8.1e-4
leucodelphinidin biosynthesis	4	6	3.8e-3
luteolin biosynthesis	4	5	1.7e-3

<b>luteolin glycosides biosynthesis</b>	<b>4</b>	<b>4</b>	<b>8.1e-4</b>
<b>quercetin glycoside biosynthesis (Arabidopsis)</b>	<b>4</b>	<b>6</b>	<b>3.8e-3</b>
<b>steviol glucoside biosynthesis (rebaudioside A biosynthesis)</b>	<b>4</b>	<b>7</b>	<b>5.4e-2</b>
<b>aliphatic glucosinolate biosynthesis, side chain elongation cycle</b>	<b>3</b>	<b>4</b>	<b>6.0e-3</b>
<b>anthocyanin biosynthesis (cyanidin 3-O-glucoside)</b>	<b>3</b>	<b>4</b>	<b>6.0e-3</b>
<b>arginine biosynthesis II (acetyl cycle)</b>	<b>3</b>	<b>5</b>	<b>1.5e-2</b>
<b>brassinosteroid biosynthesis II</b>	<b>3</b>	<b>14</b>	<b>3.5e-1</b>
<b>indole glucosinolate breakdown (active in intact plant cell)</b>	<b>3</b>	<b>3</b>	<b>5.4e-2</b>
<b>flavonol biosynthesis</b>	<b>3</b>	<b>4</b>	<b>6.0e-3</b>
<b>gibberellin biosynthesis II (early C-3 hydroxylation)</b>	<b>3</b>	<b>7</b>	<b>5.4e-2</b>
<b>glucosinolate biosynthesis from dihomomethionine</b>	<b>3</b>	<b>4</b>	<b>6.0e-3</b>

**16 )Unzip** the "**result.rar**" that you downloaded and open the file "**XCMS-diffreport-MultiClass.xlsx**" using the **OpenOffice**.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	featurename	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax	npeaks	IAC17_T0	IAC17_T1	UFV105T0	UFV105T1	CV	pvalue	qvalue	IAC17_T0.mean	IAC17_T1.mean
2	1M339739	339.125	339.1243	339.1257	38.38842	38.754	38.92283	2	0	2	0	0	0	0.015564	0.025372	795616.437961155	625304.22208
3	2M487729	487.1689	487.1674	487.1693	28.76725	28.7145	28.82	2	2	0	0	0	0	0.015564	0.025372	349063.754727065	250266.5351
4	3M522738	522.2981	522.2977	522.2984	38.04617	37.96583	38.06717	3	0	0	1	2	0	0.015564	0.025372	81586.828125	116977.4915
5	4M527738	527.2501	527.2496	527.2517	38.04617	38.04617	38.11939	3	0	0	2	1	0	0.015564	0.025372	208346.302300347	652459.0203
6	5M221738	221.1869	221.1867	221.1874	38.05658	37.96583	38.11939	4	0	0	2	2	0	0.018785	0.025372	252225.459853431	404461.1465
7	6M271741	271.06	271.0594	271.0607	40.71567	40.69317	40.813	5	0	0	2	3	0	0.018785	0.025372	739558.67997054	919886.566
8	7M519740	519.1206	519.1204	519.1217	40.02883	39.94367	40.07917	4	0	0	1	3	0	0.018785	0.025372	152521.526855469	1814529.242
9	8M347711	347.1008	347.1006	347.1009	11.01763	11.01795	11.0179	2	0	0	2	0	0	0.018785	0.025372	117989.145898438	74227.18133
10	9M487738	487.2607	487.2602	487.2612	38.046	37.96583	38.06717	3	0	0	1	2	0	0.020056	0.025372	0	0
11	10M345742	345.1483	345.1475	345.1491	42.4725	42.441	42.504	2	0	2	0	0	0	0.02129	0.025372	19217.017956058	1754462.581
12	11M540738	543.2274	543.2272	543.2275	38.0165	37.96583	38.06717	2	0	0	0	2	0	0.02129	0.025372	14272.5016276042	106081.1765
13	12M341731	341.1359	341.1355	341.1367	31.04017	31.02933	31.04017	4	1	0	1	2	0	0.021629	0.025372	490155.695384928	357525.6073
14	13M377734	377.1615	377.1598	377.1619	33.542	33.3629	33.73933	3	2	0	1	0	0	0.021629	0.025372	747959.304043511	660737.2679

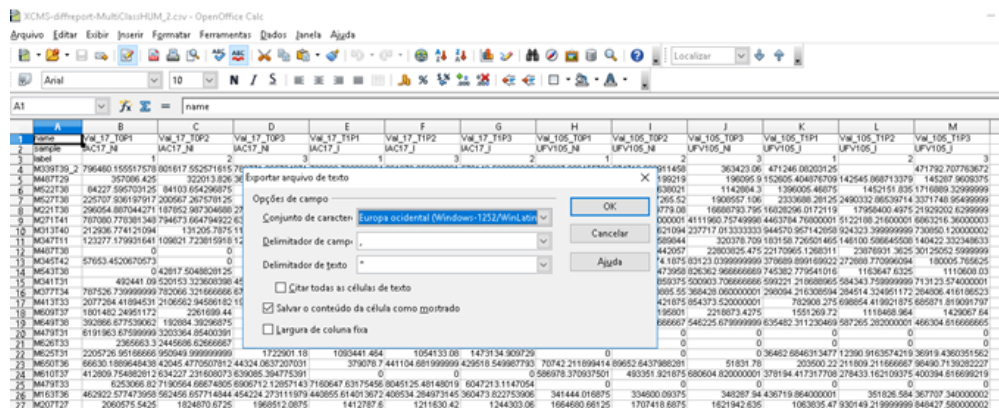
**Delete** the all column **"A"** and maintain **"B"**. Also, **delete** the columns from **"C"** to **"Z"**. Now you will maintain the information of the **XIC area** from the **all ions detected** and **aligned by XCMS**.

	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	UFV105T0	UFV105T1	maxint	postHoc	Val_17_T0	Val_17_T1	Val_17_T2	Val_17_T3	Val_17_T4	Val_17_T5	Val_17_T6	Val_17_T7	Val_17_T8	Val_17_T9	Val_17_T10	Val_17_T11
2	17210.42	27526.34	11185	0	796460.2	801617.6	788771.6	703886.7	570149.6	601876.4	397228.2	374712.6	363423.1	471246.1	423844.7	471792.7
3	15377.71	5200.294	5313	0	357086.4	322013.8	3680891	245431	247524.8	257843.7	165943.1	180373.6	196095.9	152605.4	142545.9	145288
4	178947.3	171369.4	29610	0	84227.6	84103.65	76429.23	115338.3	130517.2	105077	1145510	834259.8	1142884	1396005	1452152	1716889
5	184304.1	559613.2	49299	0	225707.9	200567.3	198763.7	649598.8	627427.9	680350.3	2114953	1747266	1908557	2333688	2490333	3371749
6	3746467	2679044	436548	0	296054.9	187853	272768.5	557625.9	356579.5	299175	17610370	10709779	16688794	16828296	17958400	21929203
7	839837.6	1239754	134206	0	787080.8	794673.7	636897.6	880975.1	894646.4	984038.2	4627425	2985242	4111961	4463785	5122188	6863216
8	738839.2	117992	27906	0	212936.8	131205.8	113422	1709437	2411864	1322269	913124.1	297686.3	237717	944571	924323.4	730805.1
9	76539.68	23209.01	3073	0	123277.2	109821.7	120868.5	77660.64	74371.34	70649.57	172167.7	213105.5	320378.7	183158.7	146100.6	140422.3
10	6539035	4187618	509476	0	0	0	0	0	0	0	23954502	11763268	22803825	22170965	23876931	30125053
11	83427.13	99412.04	102628	0	57653.45	0	0	794624.3	3673790	794973.3	242762.8	204974.2	83123.04	378689.9	272888.8	180005.8
12	78991.5	227723.6	18526	0	42817.5	0	0	137825.2	44685.56	137892.7	785806.2	673852.5	826363	745382.8	1163648	1110608
13	24092.39	70450.13	8418	0	492441.1	520153.3	457884.7	278273.6	410668.9	368634.3	516095.6	548100.6	500903.7	599221.2	584343.8	713123.6

**17) Edit** the table for the use in the **MetaboAnalyst** platform. You need inset the lines **"name"**, **"sample"** and **"label"** to indicate the treatments and replicates. **Use** **appropriated codifications for the treatments because it will be used in the exported figures for publication!!!**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	name	Val_17_T0	Val_17_T1	Val_17_T2	Val_17_T3	Val_17_T4	Val_17_T5	Val_17_T6	Val_17_T7	Val_17_T8	Val_17_T9	Val_17_T10	Val_17_T11
2	label	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N	IAC17_N
3		1	2	3	4	5	6	7	8	9	10	11	12
4	3M3738_2	796460.2	801617.6	788771.6	703886.7	570149.6	601876.4	397228.2	374712.6	363423.1	471246.1	423844.7	471792.7
5	4M47729	27526.34	11185	0	796460.2	801617.6	788771.6	703886.7	570149.6	601876.4	397228.2	374712.6	363423.1
6	5M52738	522.2981	522.2977	522.2984	38.04617	37.96583	38.06717	3	0	0	1	2	0
7	6M52738	527.2501	527.2496	527.2517	38.04617	38.04617	38.11939	3	0	0	2	1	0
8	7M221738	221.1869	221.1867	221.1874	38.05658	37.96583	38.11939	4	0	0	2	2	0
9	8M271741	271.06	271.0594	271.0607	40.71567	40.69317	40.813	5	0	0	2	3	0
10	9M519740	519.1206	519.1204	519.1217	40.02883	39.94367	40.07917	4	0	0	1	3	0
11	10M347711	347.1008	347.1006	347.1009	11.01763	11.01795	11.0179	2	0	0	2	0	0
12	11M487738	487.2607	487.2602	487.2612	38.046	37.96583	38.06717	3	0	0	1	2	0
13	12M345742	345.1483	345.1475	345.1491	42.4725	42.441	42.504	2	0	2	0	0	0
14	13M540738	543.2274	543.2272	543.2275	38.0165	37.96583	38.06717	2	0	0	0	2	0
15	14M341731	341.1359	341.1355	341.1367	31.04017	31.02933	31.04017	4	1	0	1	2	0
16	15M377734	377.1615	377.1598	377.1619	33.542	33.3629	33.73933	3	2	0	1	0	0

**18) Save as a txt (csv) file (click in "edit filter configuration and maintaining the format), such as "XCMS-diffreport-MultiClass\_yourprofile.csv", using the following parameters:**



## METABOLITE DATA ANALYSIS

3 **1)** Go to <https://www.metaboanalyst.ca/faces/ModuleView.xhtml>

Click in **Statistical Analysis**>>> "Select **"Peak Intensity Table"** and format **"sample in columns (unpaired)"**.

Select the file **"XCMS-diffreport-MultiClass\_yourprofile.csv"**

Click **in submit** >>> **click in skip**.

In the next window choice the method **"Standard Deviation"** for remove low quality results. **Click** in **"process"**

**2)** The next module is **"Data Normalization"**.

**Note:** Now, the aligned **raw data** can be normalized using different statistical methods. During normalization process the "Data transformation" and "data scaling" have been recommended when some requirement of the parametric statistic, such as normality of

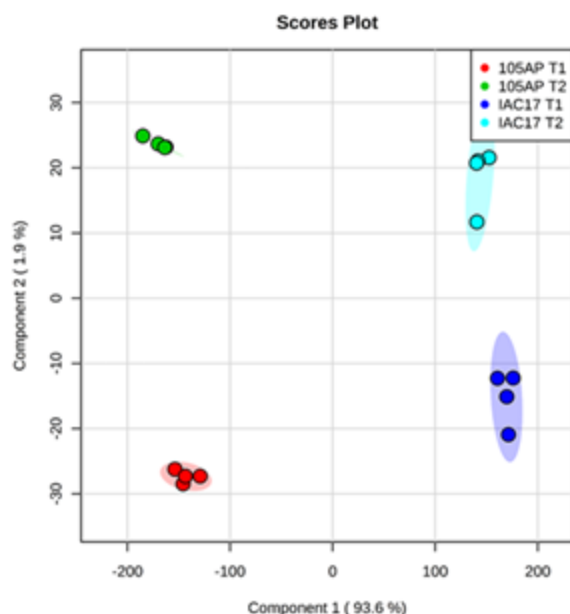
the error distribution, homogeneity of variance and etc., are not met. The **MetaboAnalyst pipeline** enable to **choice the better normalization options**, indicating the result after and before the procedure. Data normalization by media or median also show satisfactory results and could be used. Use data transformation by **"ParetoScaling"**. See the results before and after normalization to verify if the data is under normal distribution and if the variability were reduced.

Now, you can choice the analysis according with your **experimental design and hypothesis**. It convenient start with **"Cluster Analysis"** such as **PCA** enable view the general behaviors of the treatments and replicates, identifying the effect of specific treatments or genotype relative to the metabolite abundances. Adjustment for differences among your sample, data transformation and scaling are two different approaches to make individual features more comparable. You can use one or combine them to achieve better results. For details, see documentation on the **MetaboAnalyst** site (<https://www.metaboanalyst.ca/faces/docs/Tutorial.xhtml>).

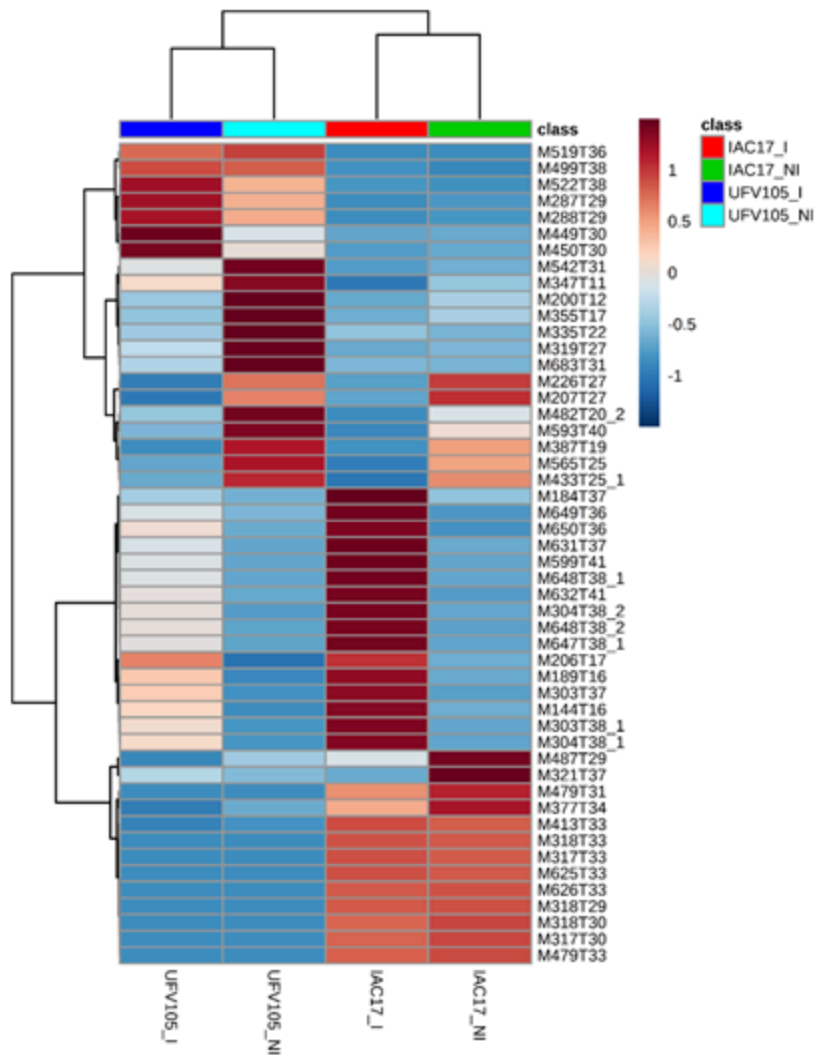
**Click** in Normalization by **median** and **Pareto Scaling>>>>submit** and **see** the **results.....**

### 3) Analysis by MetaboAnalyst plataform

**Note:** Now, you can choice the analysis according with your experimental design and hypothesis. It convenient start with **"Cluster Analysis"** such as **PCA** enable view the general behaviors of the treatments and replicates, identifying the effect of specific treatments or genotype relative to the metabolite abundances.



**"Cluster Analysis"** by **PSDLA** and **heatmap** enable also identify which **metabolites** showed **higher variations** for each group (the output indicate the relative concentrations of the corresponding metabolite in each group under study).



## ANOVA

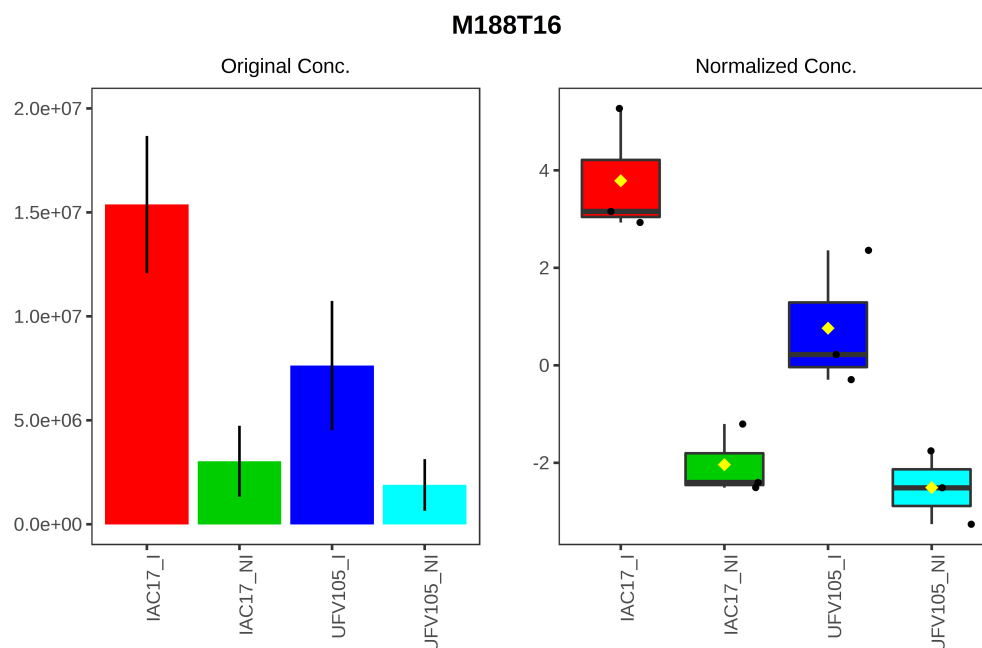
**Click in ANOVA one-way**

**Click in "red filled circles >>> you will see the dysregulated compounds**

Or **click** in the **link** to see the **table** containing **all ANOVA results**

You could also search the most significant ions listed from the **PSDLA and heatmap above**.

To **export**: click in **picture icon** and select **TIF** and **600dpi** and full page.



## METABOLITE IDENTIFICATION FROM MS/MS SPECTRUM

- 4 **Note:** Metabolite identification based in MS and MS/MS spectra is a challenging task. To obtain high confidence putative identifications require the use of multiple approaches. Currently, this process relies on tandem mass spectrometry (MS/MS) spectra generated separately for peaks of interest identified from previous MS runs. Frequently, the start point is compare the query spectrum against a database of MS/MS spectra of reference compounds and rank the candidates based on their similarity or identity to the query spectrum. This approach allow also select putative compounds sharing chemical groups that could be useful for structure elucidation. Complementary approaches can be combined to confirm the identifications, such as the use of isotope pattern analysis for detecting the molecular formula and further analyses the fragmentation pattern of a compound using fragmentation trees (Sirius Package) and/or of Competitive Fragmentation Modeling-ID (CFM-ID) to accurately predict electrospray ionization-MS/MS (ESI-MS/MS) spectra from chemical structures under investigation.

The followed steps describe the **use of the NIST package for searches of the MS/MS spectrum** of ions selected as dysregulated using the XCMS and Metaboanalyst

platforms. The fragmentation spectra (ESI MS/MS and LC/MS) were download from MassBank of North America (MoNA) repository (<https://mona.fiehnlab.ucdavis.edu/downloads>), imported and formatted for the use by the NISTdemo.exe. Use the trial version or purchase a commercial license for the NIST packages and libraries.

**Download** the "**Tutorial\_LCMS.rar**" from <https://figshare.com/s/952a2e51cc79592deb9d> containing the trial versions, videos and .mgf files as example.

**1)** Unzip the **DataAnalysis.rar** and install ....use the evaluation license for 30days or purchase a definitive license from Bruker Daltonics.

**Note:** other packages could be used to export the MS/MS peak list. Verify the input formats for the NIST.exe.

**2)** Run the **DataAnalysis.exe**.

**3)** Go to "**file**" >>>> "**open**" your LC/MS spectrum file. You will see the **MS and MS/MS** profiles.

**4)** Go to "**Method**" >>>> "**open**" >>> in the directory "**processing**" choice the appropriated **method**, such as "**DataAnalysis microTOF Default\_HUM\_altaIntensidade100.m**" to detect and process for **100 higher intensity ions**.

**5)** Use the mouse **right-click** over the "**TIC +all MSn**" Windows and select "**find compounds Auto MSn**" to generate the **compound list**

Use the mouse **right-click** over the spectrum window >>>> select "**Display Parameters**" configure "**mass precision**" for **4**. High accuracy mass spectrum will be displayed.

**6)** Click over the LC/MS spectrum file on the upper left side to select all **compounds** and **chromatograms**>>>>> go to "**File**" >>> **export**>>>> **mass spectrum**>>> select **mgf format** >>> maintain the file name "**myfilename.mgf**" click in **OK** to **export**.....  
**Note:** You could export a file .mgf containing a spectrum of interest or a file containing multiple MS/MS spectrum.

**7)** Unzip the **NIST.rar** and install in your PC. Click in **cancel** when was required to **search** for the library in your PC.

**8)** Unzip the "**libraries.rar**" and **paste** the directory library in "**D:\NIST14\MSSEARCH\**"

## 9) Run the NIST "*nist ms search.exe*".

## 10) NIST configuration:

Go to "options" >>> "library search options"

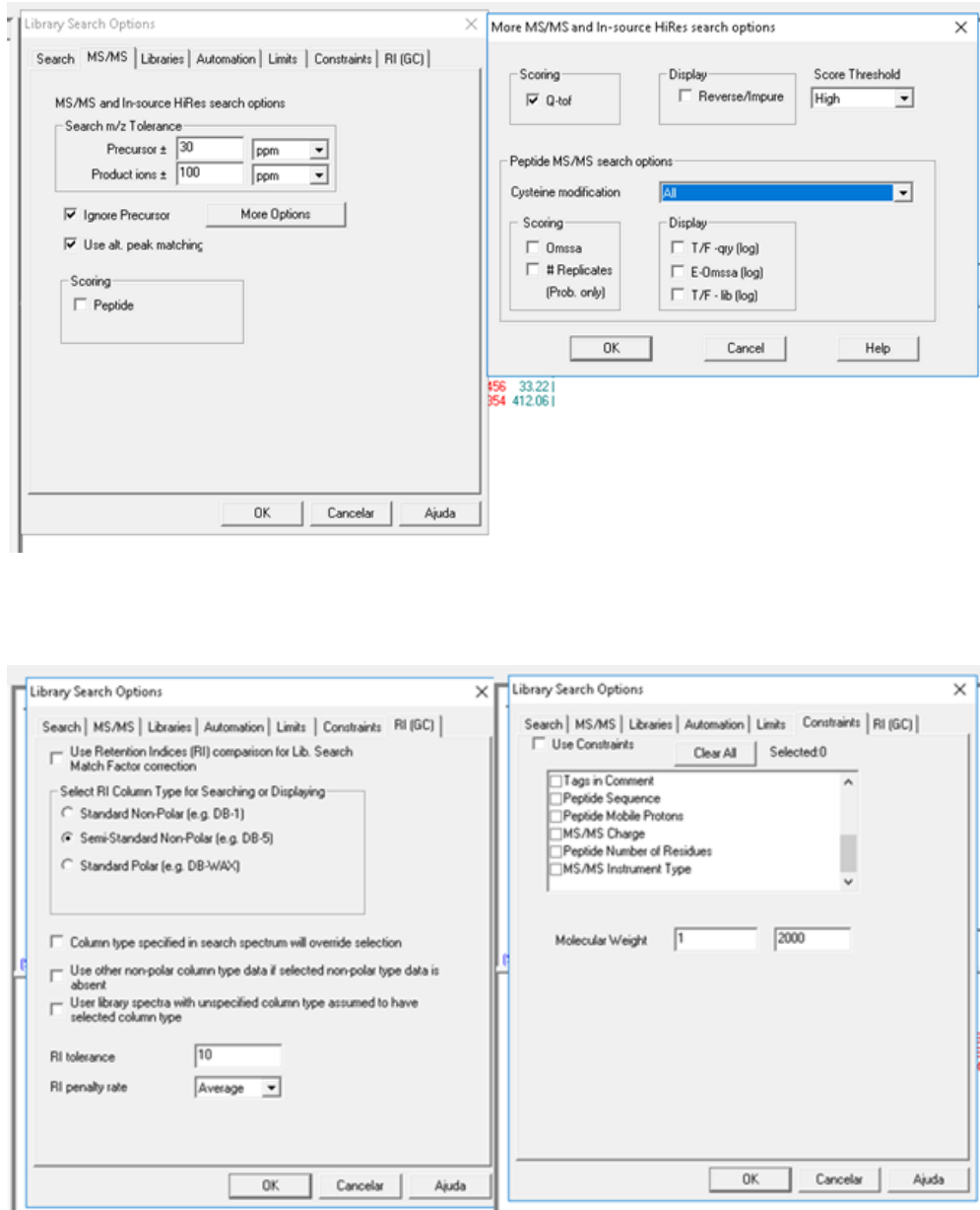
**Note:** The NIST package enable search your **MS/MS**spectra against the **MSM/MS spectra** present in the **libraries**. There are three types: GC M/MS library (main lib), GC RI (retention index) and ESI - LC MS/MS.

Thus, **select all libraries except "mainlib, raplib and nist\_ri**

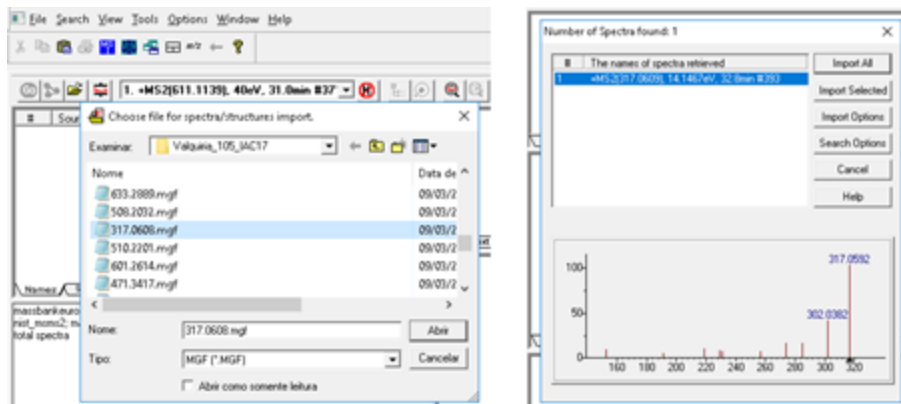


## Go to "MS/MS" and configure for Q-TOF and low mass molecules.

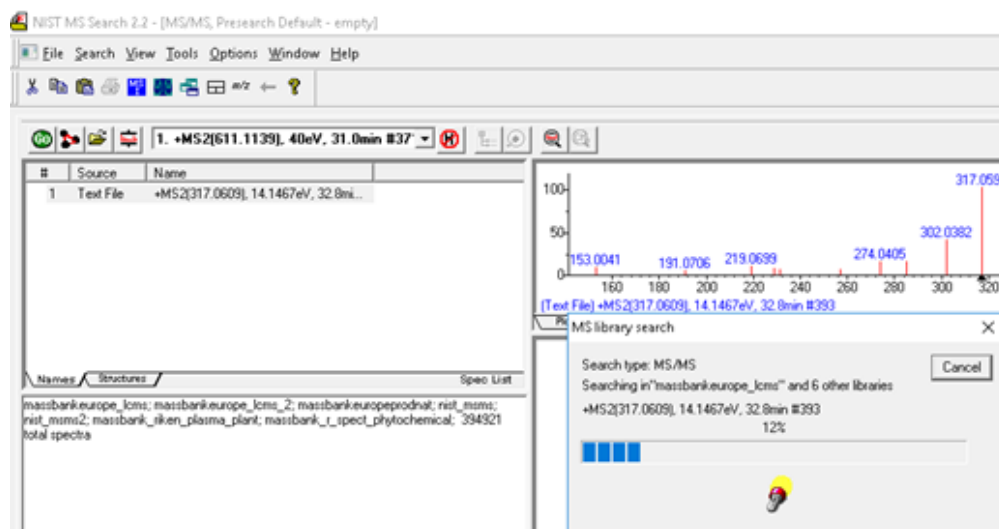
When you are performing the a search the option "**Ignore the Precursor**" could be **unselect** for enable the spectrum match with **related compounds** containing similar chemical groups. For example, flavonoids core with different glicoconjugates. The "**Score Threshold**" could also be modified to **enable** matching with spectra generated from different mass spectrometers using lower and higher collision energies;



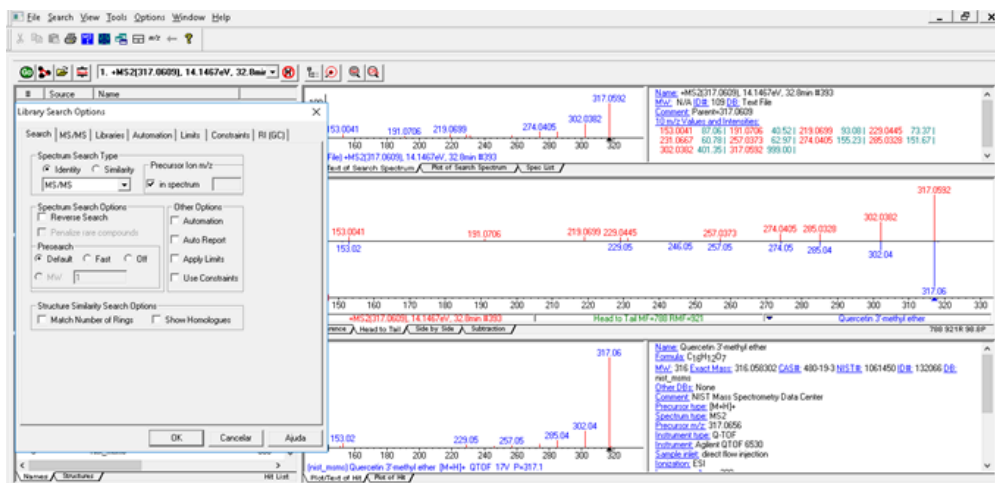
**11) Import and open the mass spectrum exported as .mgf**



12) Process a **double-click** on the **MS2file** and the search will be **start....**

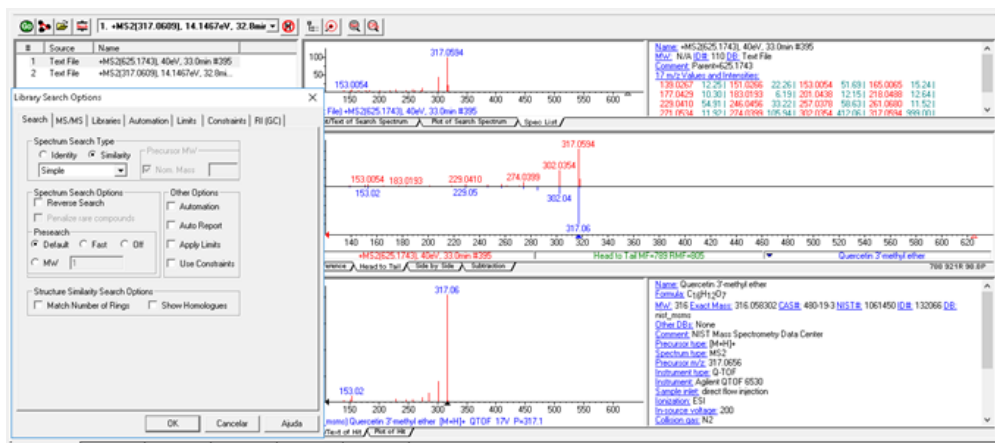


**Note:** this search was performed using the options "**Identity**" and MS/MS and the **option** "precursor ion m/z" "in **spectrum**" as selected. Thus, enable hits with higher similarities for the fragments and precursor ions over the mass accuracy and scores thresholds!!!



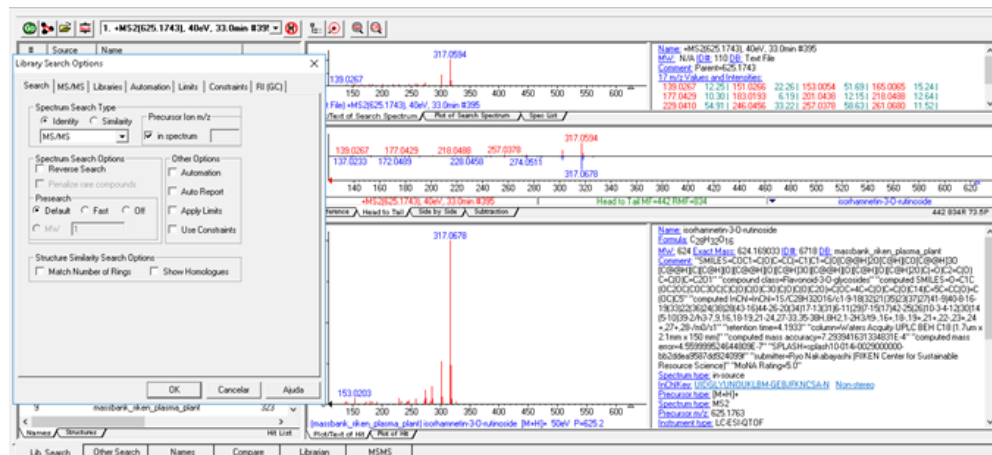
If this options **fail** to find compounds sharing the same accuracy mass, try the **"Similarity" "simple" search option!**

In this example, the ion 625.1743 before the search using these option return with a matching for Quercetin 3-methyl ester (Isorhamnetin) (317.0656), thus could be a Methyl quercetin glycosilated.



The search was performed **again** using the **high constrains options** such as options **"Identity"** and MS/MS and the option **"precursor ion m/z"** **"in spectrum"** as selected and **not ignore the precursor** (MS/MS options).

Observe that the identify compound sharing the same high accuracy mass as a Rutin methylated or Isoharmnetin 3-rutinoside



Thus, if the libraries contain the MS/MS spectrum for your compound you will see a match for your spectrum (**Remind** that libraries contain **repeated spectra** for the same compound, however generated with different instruments and energy from different labs!!!).

Thus, the similarity and identity searches will generate information about a putative chemical group and will be useful when applying other methods for putative compound identification from MS data.

## REFERENCES

- 5 Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem.* 84:5035-9. <https://doi.org/10.1021/ac300698c>. Epub 2012 May 10.

Chong, J., Wishart, D.S. and Xia, J. (2019) **Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis**. *Current Protocols in Bioinformatics* 68, e86

## ACKNOWLEDGMENTS

- 6 The authors would like to thank the Núcleo de Análises de Biomoléculas (NuBioMol) of the Universidade Federal de Viçosa for providing the facilities for the data analysis. The authors also acknowledge the financial support provided by the following Brazilian



agencies: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – Finance code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (Finep), and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).