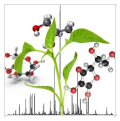


Oct 30, 2019

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

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

[1]Department of Biochemistry and Molecular Biology - Universidade Federal de Viçosa - UFV, BIOAGRO/INCT-IPP - Viçosa-MG Brazil, [2]Center for Biomolecules Analysis (NuBioMol) Universidade Federal de Viçosa - UFV Viçosa-MG Brazil

1 *Works for me* dx.doi.org/10.17504/protocols.io.8vjhw4n

OpenPlant Project Metabolomics Protocols & Workflows

Humberto J O Ramos
Universidade Federal de Viçosa

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 TEXT

REAGENTS

- Methanol (LiChrosolv, Merck, Order n° 1.06001)
- Acetonitrile (LiChrosolv, Merck , Order n°)
- Acetic Acid (Sigma)
- High pure water (18.2M Ωcm-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 µm – 300 µm × 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

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 um 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^{-1}, 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.

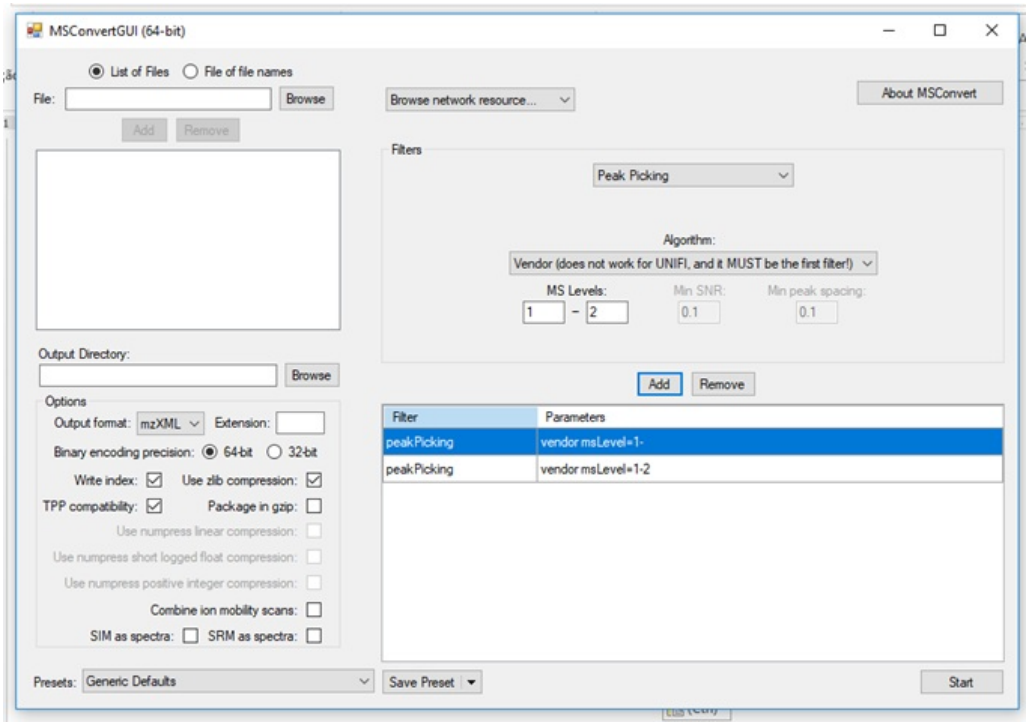
6) Ion scanning for MS1 spectra in positive or negative mode can be carried out for masses ranging between 100 and 1000 m/z, and between 50 and 1000 m/z for the MS2 spectra. The data were acquired over 57.0 min in each LC-MS/MS analysis, using the Hystar software program, version 3.2 (Bruker Daltonics, 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.

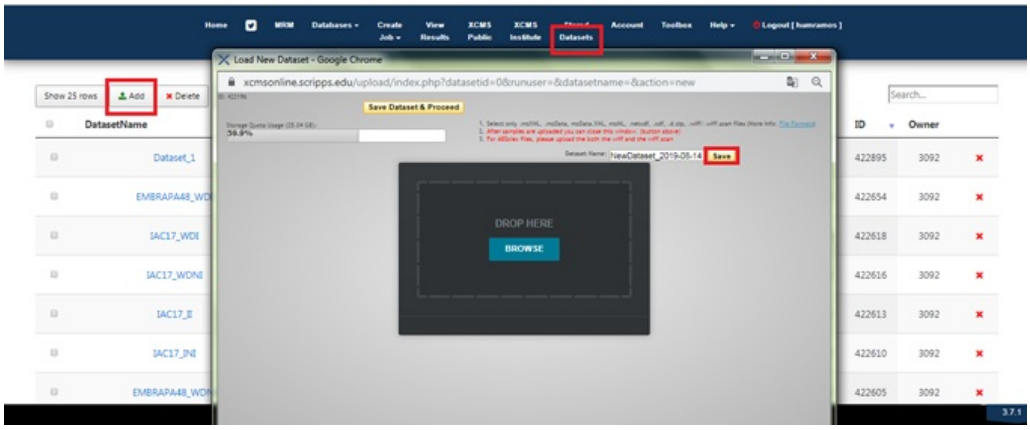
3) Locate the directories to input all spectra and output; choose filters **Peak Picking** and MS level 1, click **add** and repeat this step for **MS levels 1-2**; Add samples to **Browse** and click **Start**. Wait for all jobs to be completed



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.eduand create an account.

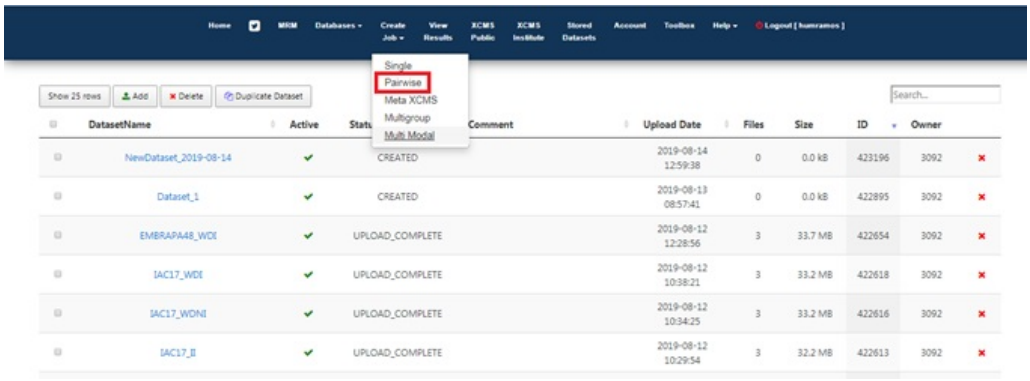
6) To enter your data click **Dataset** >>>> **Add** and click **upload** (**Browse**) the ***.**mzXML** spectra from biological replicates for each treatment one at a time, this guarantees that all the replicates will be grouped in a directory for each treatment



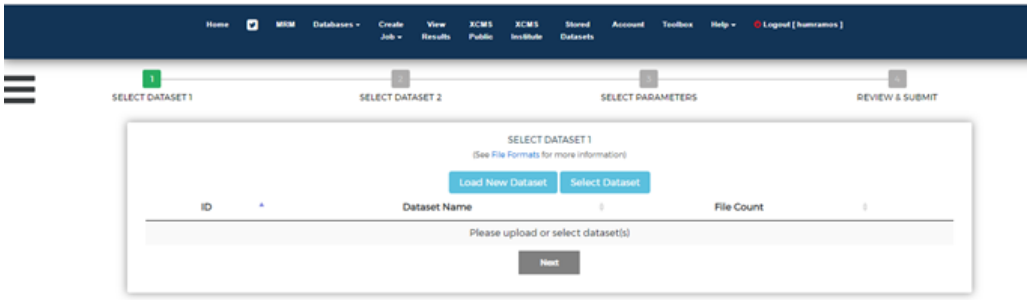
7) Rename the **Dataset Name** according to the treatment (use an abbreviation in English related to the treatment). Wait for the upload to finish and click **Save.**

8) Repeat this procedure for all the treatments.

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



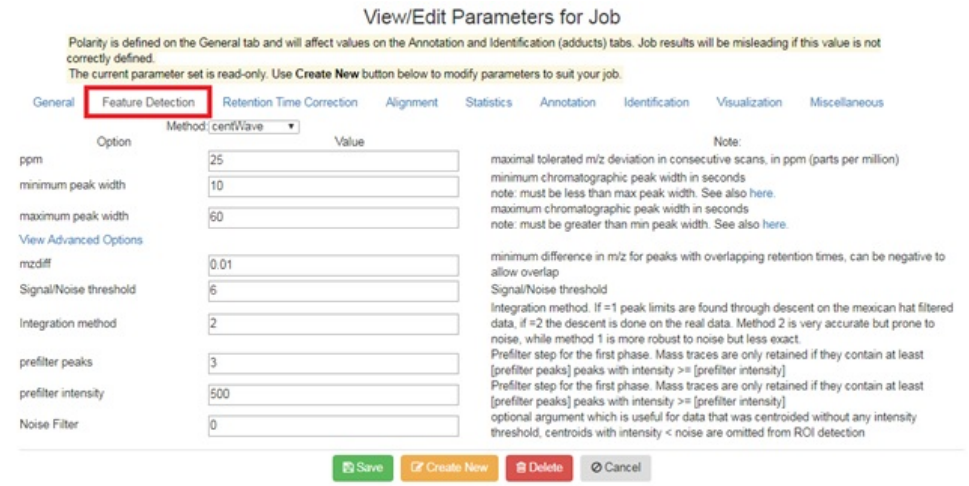
10) Click **Select dataset**: click in the directory for the first treatment (example "IAC17 NI – non infested).

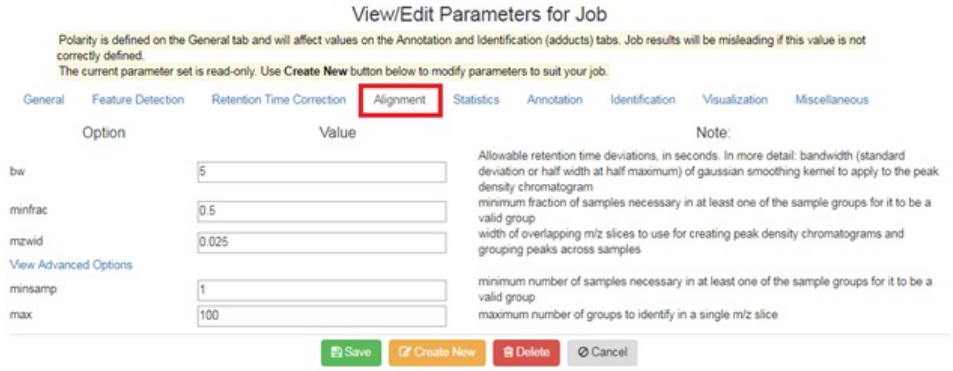


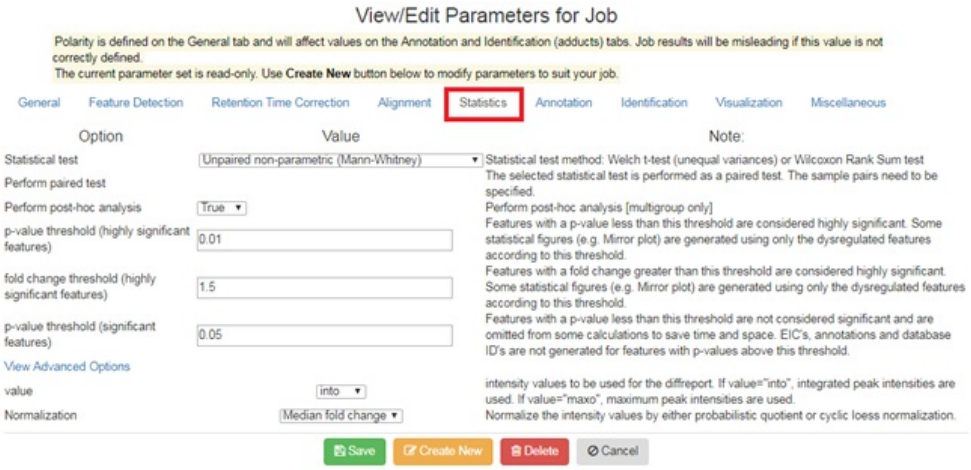
11) Click **Next** >>>> select the second treatment >>>> **Next.**

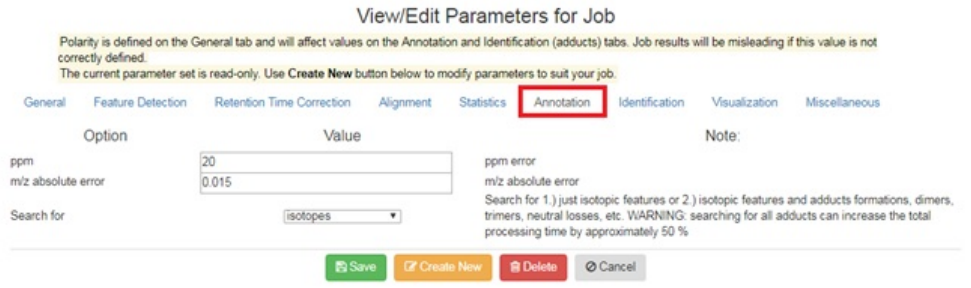
12) Select the Parameters >>>> click **HPLC/Q-TOF** >>>> click **Customized** >>>> click **View and Edit** >>>>.

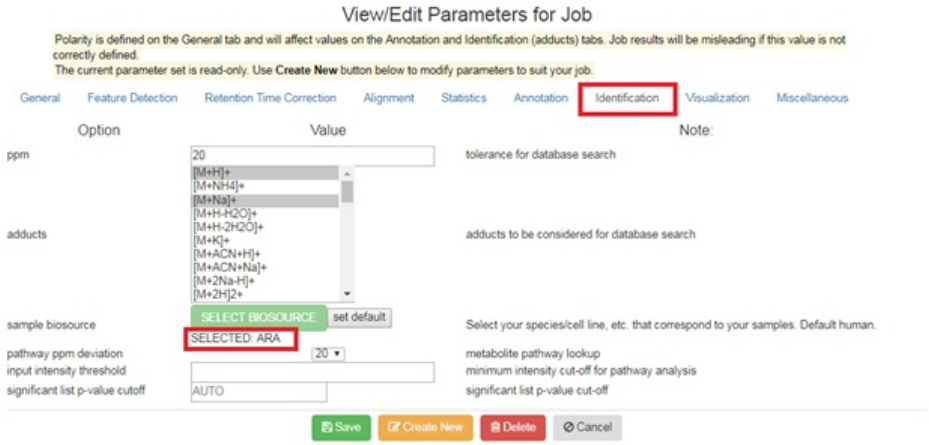
Note: estimate the m/z accuracy and the maximum peak width for your LC/MS system. In the module "identification", you need search for the presence of the metabolite library of the organism in study.





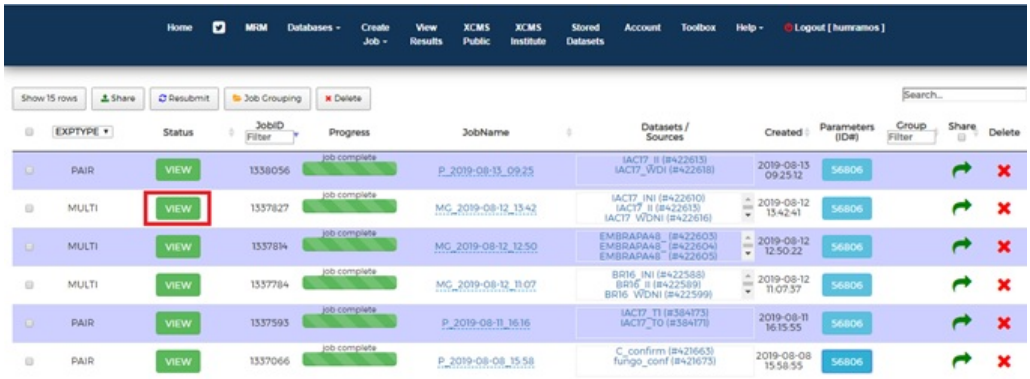




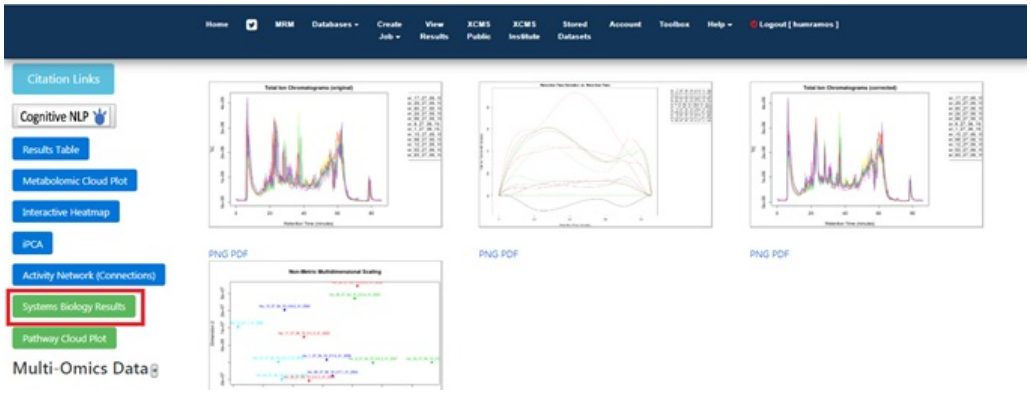


13) Click **Save**. Back to the module **"identification"** and verified if **the metabolite library of the organism in study was selected!! Click and save using a name "myXCMS parameter" for your customized method.**
Click N**ext**>>>> click **Submit**. The job will **start**

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



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.

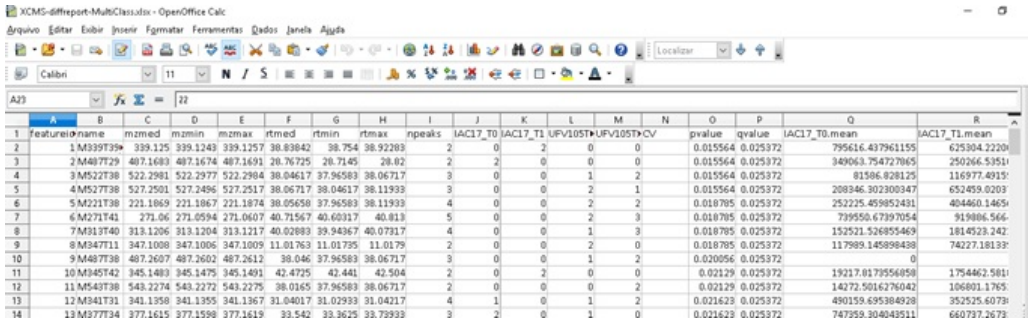
>>>> click the column ***Overlapping putative metabolites*** to classify in descending order. You could click in the **link** to see the **metabolite information** for each **pathway** enriched.

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

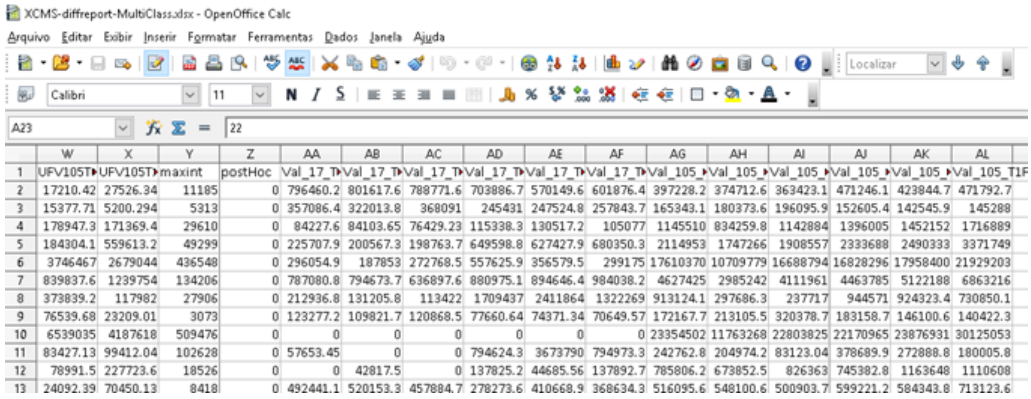
Pathway	Overlapping putative metabolites1	All metabolites2*	p-values
flavonoid biosynthesis (in equisetum)	9	10	3.4e-4
kaempferol glycoside biosynthesis (Arabidopsis)	9	9	3.3e-4
sanguinarine and macarpine biosynthesis	8	13	9.9e-4
brassinosteroid biosynthesis I	7	17	2.2e-2
cytokinins 7-N-glucoside biosynthesis	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 (2β-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**
luteolin glycosides biosynthesis	**4**	**4**	**8.1e-4**
quercetin glycoside biosynthesis (Arabidopsis)	**4**	**6**	**3.8e-3**
steviol glucoside biosynthesis (rebaudioside A biosynthesis)	**4**	**7**	**5.4e-2**
aliphatic glucosinolate biosynthesis, side chain elongation cycle	**3**	**4**	**6.0e-3**
anthocyanin biosynthesis (cyanidin 3-O-glucoside)	**3**	**4**	**6.0e-3**
arginine biosynthesis II (acetyl cycle)	**3**	**5**	**1.5e-2**
brassinosteroid biosynthesis II	**3**	**14**	**3.5e-1**
indole glucosinolate breakdown (active in intact plant cell)	**3**	**3**	**5.4e-2**
flavonol biosynthesis	**3**	**4**	**6.0e-3**
gibberellin biosynthesis II (early C-3 hydroxylation)	**3**	**7**	5.4e-2
glucosinolate biosynthesis from dihomomethionine	**3**	**4**	**6.0e-3**

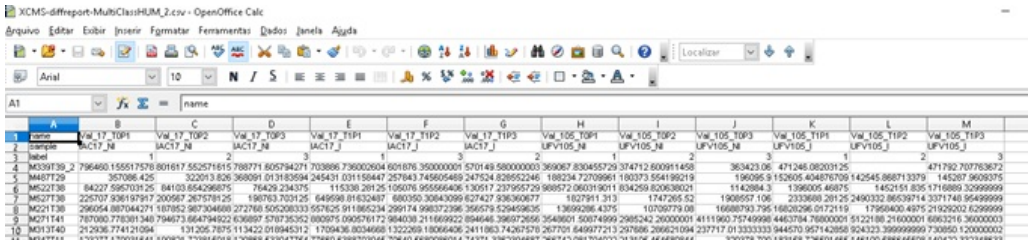
16)Unzip the "**result.rar**" that you downloaded and open the file "**XCMS-diffreport-MultiClass.xlsx**" using the **OpenOffice**.



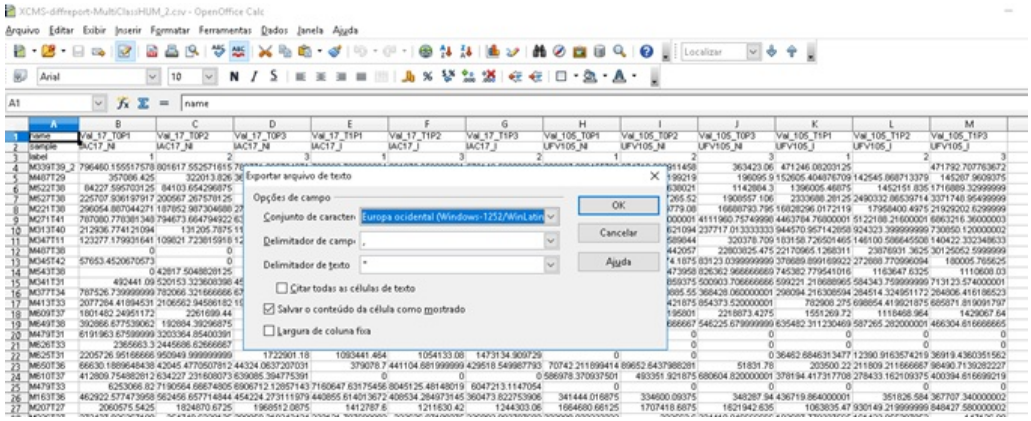
Delete the all column **"A" and maintain "B".** Also, **delete** the columns from **"C" to "Z".** Now you will maintain the information of the **XIC area** from the **all ions detected** and **aligned** by **XCMS**.



17) Edit the table for the use in the **MetaboAnalyst plataform**. 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!!!**



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**"

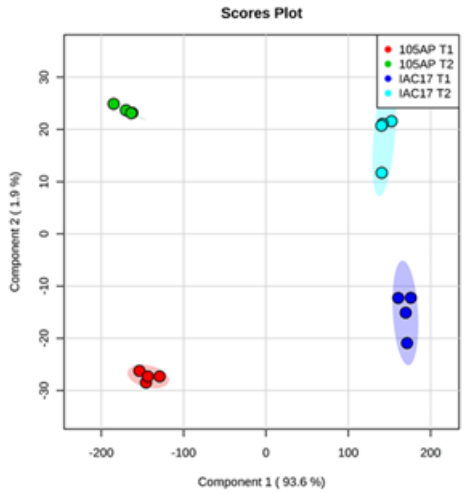
2) The next module is **"Data Normalization".**

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).

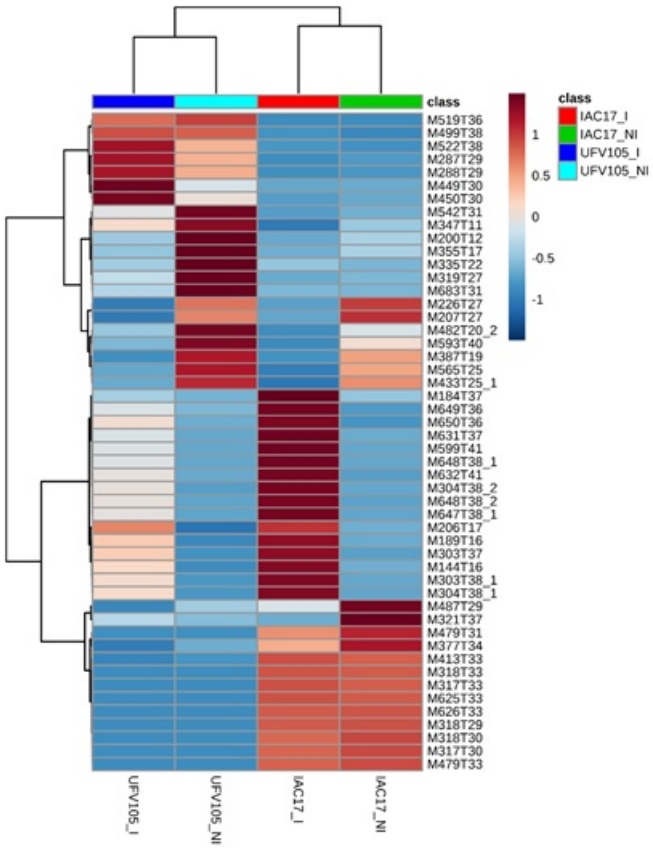
Click in Normalization by **median** and **Pareto Scaling>>>>>submit** and **see** the **results**.....

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.



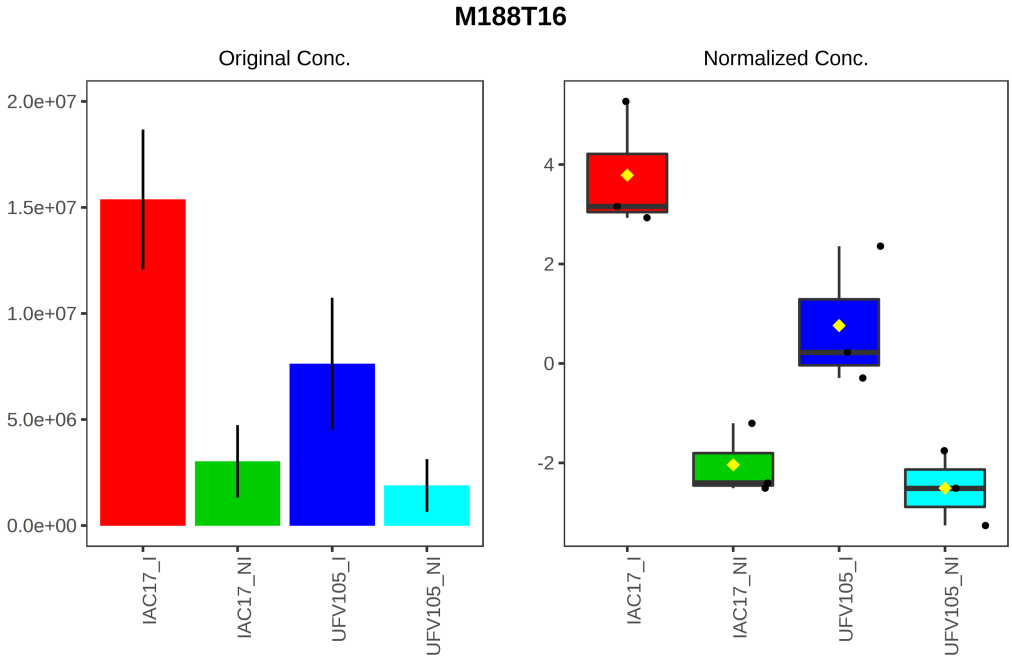
"Cluster Analysis" by **PSDLA and heatmap** enable also identify which **metabolites** showed **higher variations** for each group (the output indicate the relative concentrations of the corresponding metabolite in each group under study).



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 obtained 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 confirmate 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 plataforms. 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.

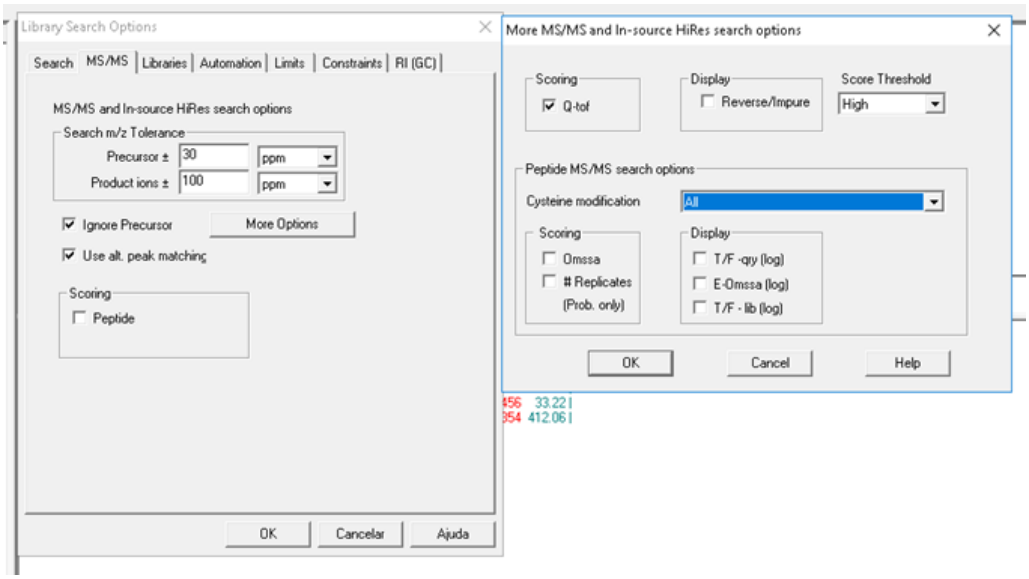
8) Unzip the "**libraries**.rar" and **paste** the directory library in "**D:\NIST14\MSSEARCH\"**

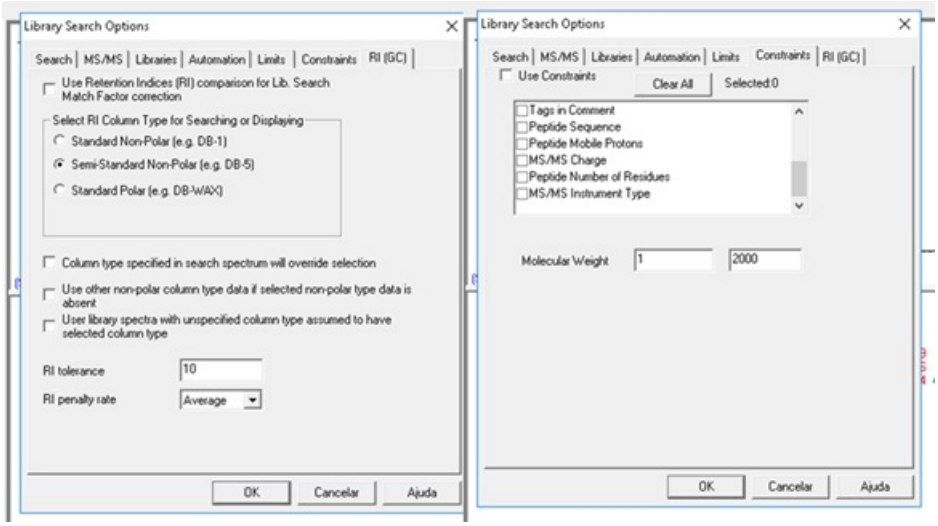
9) Run the NIST "*nist ms search.exe*".

10) NIST configuration:
Go to "options" >>> "library search options"

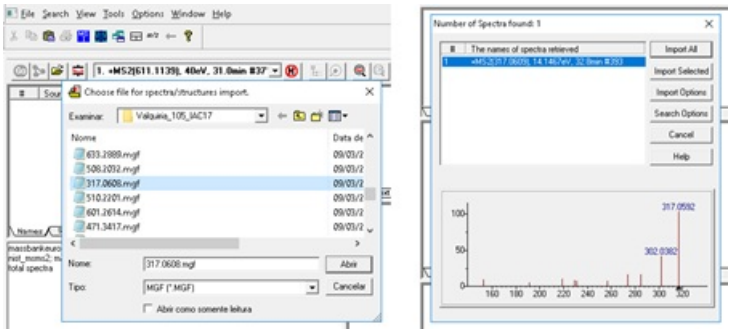
Note: The NIST package enable search your **MS/MS**spectra against the **MSM/MS spectra** present in the **libraries**. There are three types: GC M/MS library (main lib), GC RI (retention index) and ESI - LC MS/MS.
Thus, **select all libraries except "mainlib, raplib and nist_ri**

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;

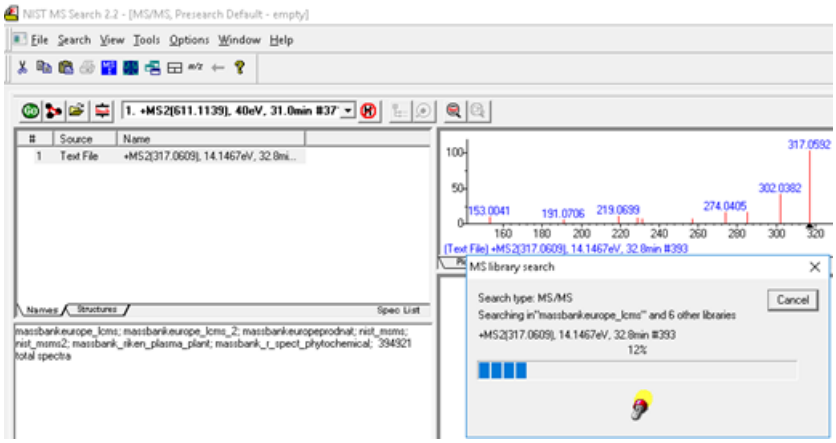




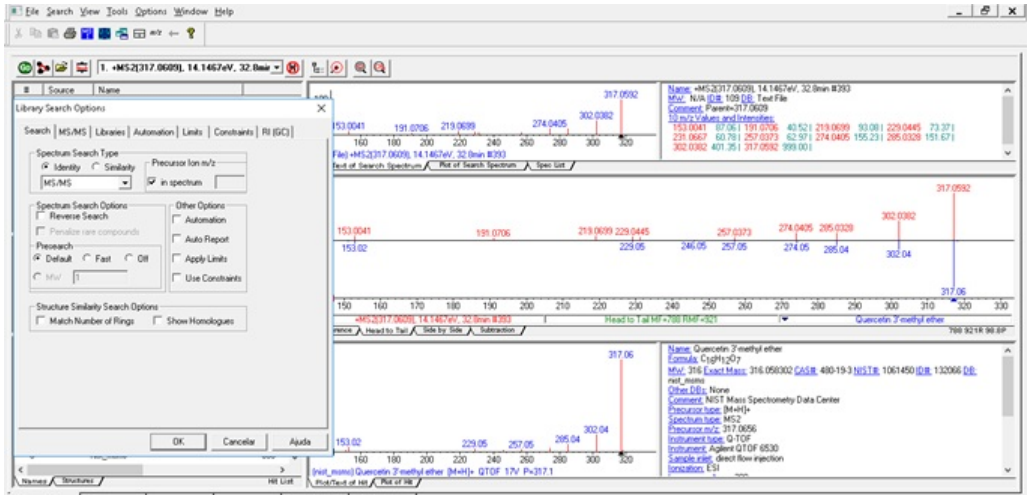
11) Import and open the mass **spectrum** exported as **.mgf**



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

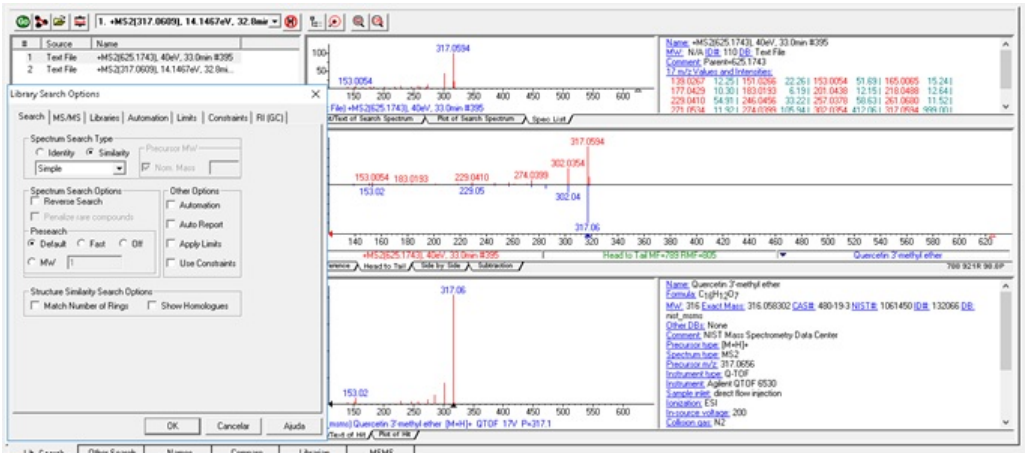


Note: this search was performed using the options "**Identity**" and MS/MS and the **option** "precursor ion m/z" "in **spectrum**" as selected. Thus, enable hits with higher similarities for the fragments and precursor ions over the mass accuracy and scores thresholds!!!



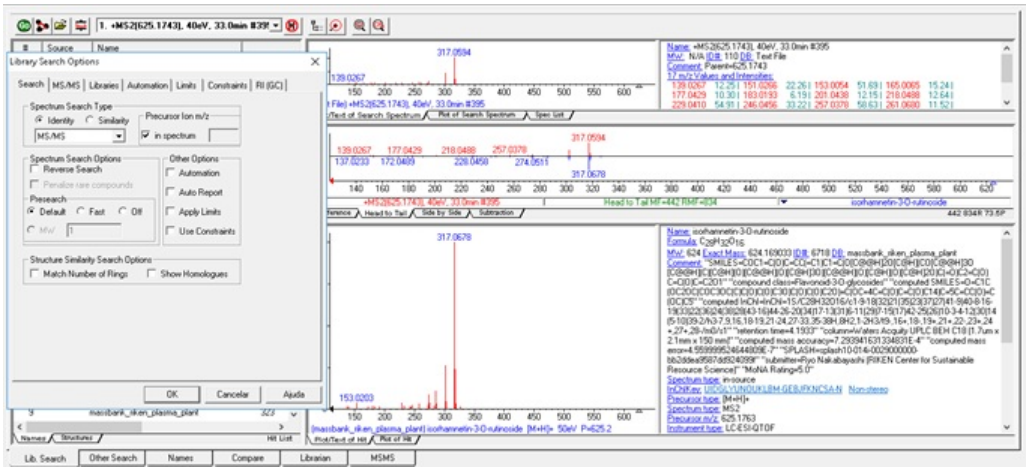
If this options **fail** to find compounds sharing the same accuracy mass, try the "**Similarity**" "**simple**" **search option!**

In this example, the ion 625.1743 before the search using these option return with a matching for Quercetin 3-methyl ester (Isorharmnetin) (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 methyled or Isoharmnetin 3-rutinoside



Thus, if the libraries contain the MS/MS spectrum for your compound will see a matches 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 informations about a putative chemical groups and will be useful when applying other methods for putative compounds 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: 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).