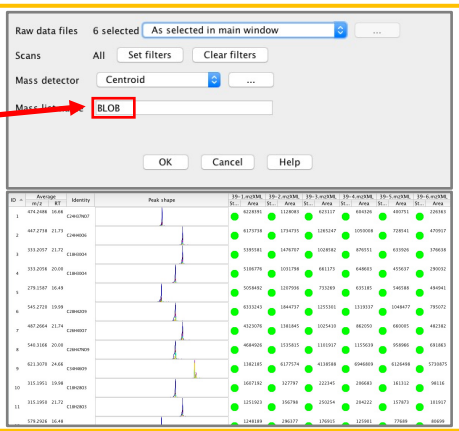
S1. Detailed procedure for .mgf and .xls data exportation

1. LC-MS² analysis and data conversion: conversion of proprietary files into universal .mzXML format using MSConvert (ProteoWizard).

2. MZmine 2 Preprocessing (user dependent!):



A. Feature detection:

- Mass detection: **All spectra must be designated as "BLOB" in the mass list name and not "mass"!**
- Chromatogram builder
- Deconvolution
- Isotopic Peak Grouper

B. Semi-quantitative info:

- Alignment
- Gap-filling

] For peak lists comparison

C. Identification (molecular formula, *m/z* databases…)

→ **Final peak list, ready to be exported.**

3. Exportation:

python™ + or

- Download **Python 3.5** (or upper version) and **Mamp 4.1.1** (Mac users) or **WampServer 3.0.6** (Windows users).
- Download the python scripts afforded at https://github.com/ElMouth/MZM2-MN
- Open **Mamp** and start the servers or simply open **WampServers**.

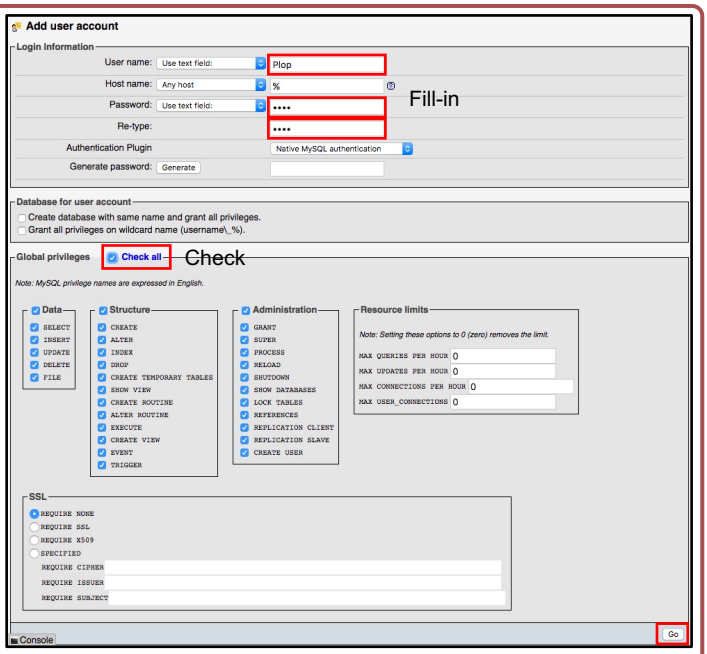
Only for first use: - Update the *pip* package for Python by launching the *get-pip.py* script via **Python Launcher**.
- Import the *Pandas*, *XLWT*, and *PyMySQL* libraries for Python by opening a new Command Promt (or Terminal) window and entering "pip3 install pandas", "pip3 install xlwt" and "pip3 install pymysql".

For Mac users, only for the first exportation:

- Open your web browser and connect to your **phpMyAdmin** interface using the following address : http://localhost:8888/phpMyAdmin/?lang=english

- From the home phpMyAdmin page, go to the "User accounts" tab and add a user by clicking "Add user account".

- Fill the "User name", "Password" and "Re-type" fields and check the "Global privileges" tab. Complete the user account creation by clicking on "Go" at the bottom of the page.



- Open the *bootstrap.py* script using the integrated **Python IDLE**.

- Edit the user_account_name = "**root**" line with your phpMyAdmin account previously created, and save the script.

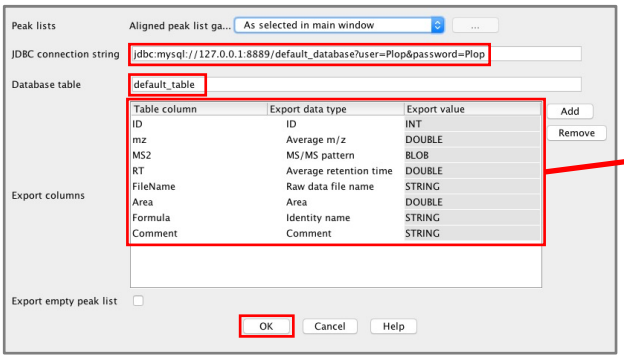
```
####################################
## MacOS part
user_account_name="root" # Value to change

mamp_ip_adress = "127.0.0.1" # Default value
mamp_port = 8889 # Default value
```

- Make sure your *bootstrap.py* and *mysql_processing.py* (Mac or Windows) are in the same folder.
- With your MZmine 2 processed data in hand and the local server started, **launch the *mysql_processing.py* (Mac or Windows) script using the Python launcher** (Mac users may have to use **Python IDLE** (then press F5) and are asked to enter their MySQL database password).

- A new mySQL database (default_database) and table (default_table) are automatically generated. Then, you are asked to export your MZmine 2 data using the given *JDBC connection string* and *Database table* parameters:

```
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "copyright", "credits" or "license()" for more information.
>>> WARNING: The version of Tcl/Tk (8.5.9) in use may be unstable.
Visit http://www.python.org/download/mac/tcltk/ for current information.

 RESTART: /Users/Florent/Desktop/Flo/mysql_processing1.2_MacOS/mysql_processing_MacOS.py
Connection to Mamp : Successfull
Connection to the Database : Successful
##########################################################

Please export your prossessed data from MzMine2
Use these parameters
JDBC connection string : jdbc:mysql://127.0.0.1:8889/default_database?user=Plop&password=Plop
Database table : default_table
Press Enter to continue
```

- Go back to MZmine 2 and click in the “peak list methods” tab, “Import/Export” and “Export to SQL database”.
- Fill the “JDBC connection string” and “Database table” fields using the parameters afforded in the script and select the data you want to export. Click “OK”:



Note that the script can proceed only those limited options. The selected options must absolutely be named in the “Table column” field as following:

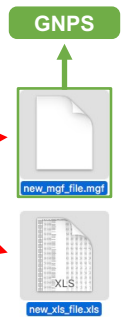
- ID as “ID”
- Average m/z as “mz”
- MS/MS pattern as “MS2”
- average retention time as “RT”
- Raw data file name as “FileName”
- Area as “Area”
- Identity name as “Formula”
- Comment as “Comment”.

- Go back to your Python process and press “Enter”. The script will successively write the .mgf then the .xls files in the script directory, and will clear the database.

- Press “y” and “Enter” if you want to use a group mapping file. Make sure that your file is located in the script directory and press “Enter” to continue.

- Select the Group mapping file you want to run.

```
Database table : default_table
Press Enter to continue
Exporting MGF file, please wait.
##########################################################

MGF file creation : Successful
Exporting XLS file, please wait.
XLS file creation : successful

##########################################################

Clearing database : successful

##########################################################
Do you want to add a Group Mapping file ? y/n
Your answer : y
Please be sure that a group mapping file is present in
the current directory and press any key to continue.
Press Enter to continue

Please select one of the following files :
0 = Group mapping 1.txt
1 = Group mapping 2.txt
File number : 0

Group mapping column have been added to the CSV file.
End of the script, thanks for using it!
Press Enter to exit
```



4. GNPS:

- Launch a GNPS job with the exported .mgf file. Make sure the MS-Clustering tab is **deactivated** and the Minimum cluster size set to 1**.**
- Download the clustered data from GNPS and open the *clusterinfosummarygroup_attributes_withIDs* file with Excel. Simply order this table by increasing RT (*RTMean* column), that corresponds in fact to you Mzmine *ID*. This column can be renamed *ID* and all other useless columns can be delete (all except *cluster index*, the newly renamed *ID*, *precursor mass*, *ProteoSAFeClusterLink*, *LibraryID*, *NumberOrganismIDs* and *AllOrganisms*).
- Open the .xls file from MZmine and order it by increasing *ID* value too. Copy this table and just append it to the *clusterinfosummarygroup_attributes_withIDs* file:

Final *clusterinfosummarygroup_attributes_withIDs* file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	cluster index	precursor mass	mz	RT	Formula	39-1.mzXML	39-2.mzXML	39-3.mzXML	39-4.mzXML	39-5.mzXML	39-6.mzXML	ProteoSAFeC	LibraryID	NumberOrganismIDs	AllOrganisms	ID
2	1	621.307	621,307048	24,6641	C34H46O9	1382185	6177574	4138588	6946809	6126498	5730875	http://gnps.u	N/A	0	N/A	1
3	2	593.276	593,275642	21,3625	C32H42O9	31635	2478138	1445363	2673782	2309745	1421278	http://gnps.u	Pheophorbid	1	GNPS-LIBRARY	3
4	18	365.105	365,105248	0,7636	-	2231	3436	73600	453722	629532	985795	http://gnps.u	N/A	0	N/A	4
5	10	637.302	637,30189	23,9876	C34H46O10	174974	954452	1429181	718801	765444	687581	http://gnps.u	N/A	0	N/A	5
6	3	609.27	609,270448	20,7384	C32H42O10	1707	1434761	4090401	1496745	1242896	717509	http://gnps.u	N/A	0	N/A	7
7	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

GNPS Info **MZmine 2 Preprocessing stage** **GNPS Info** **Both!**

- Open Cytoscape and import a new network from the *networkedges_selfloop* file. Load the modified data table *clusterinfosummarygroup_attributes_withIDs* for accurate semi-quantitative pie chart drawing based on relative areas and molecular formula implementation.