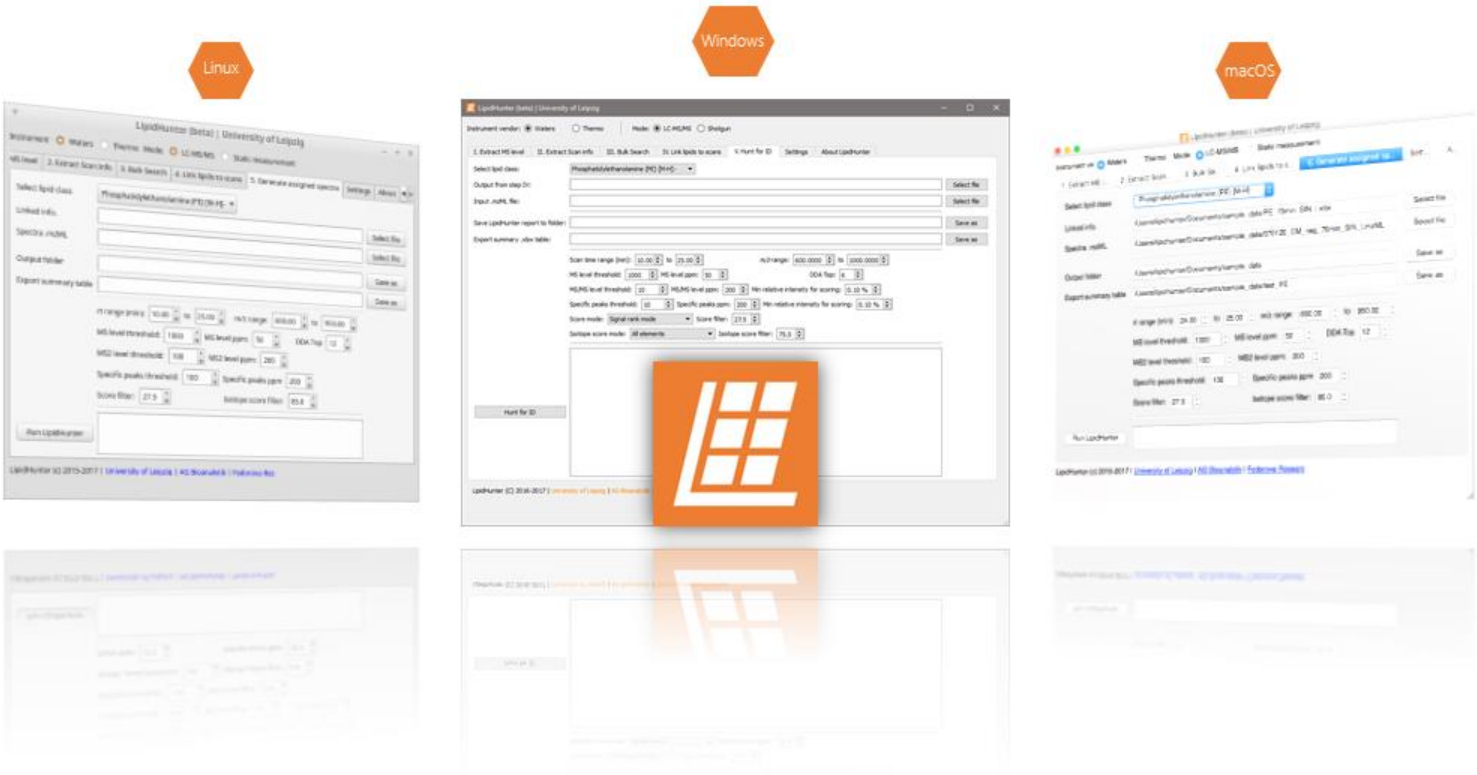
User Guide To LipidHunter



For LipidHunter Beta version 03, March, 2017

User guide version 08, March, 2017

UNIVERSITÄT LEIPZIG

Table of Contents

1. Introduction

LipidHunter is an open source software for high-throughput identification of phospholipids from LC-MS/MS and shotgun data-dependent lipidomics experiments. LipidHunter is designed to resemble manual spectra annotation process and provide full control and transparency for each identification and verification step.

This user guide will lead you through LipidHunter workflow based on the example of PL identification from LC-MS/MS DDA experiment performed on ESI-Q-TOF instrument using negative ion mode.

2. License

LipidHunter is Dual-licensed

- For academic and non-commercial use: GPLv2 License
    - https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html
- For commercial use: please contact the develop team by email.

Please cite our publication in an appropriate form.

3. Downloads

There are two versions of LipidHunter. The Windows executable version and the source code version. General information and installation instructions can be found here:
https://bitbucket.org/SysMedOs/lipidhunter/wiki/Home

- Windows executable version
    - Executable version is provided for Windows 7, 8 and 10 64bit system only.
    - Please read the instructions of LipidHunter windows version:
    - https://bitbucket.org/SysMedOs/lipidhunter_exe
    - The download link:
    - https://bitbucket.org/SysMedOs/lipidhunter_exe/downloads/
- Source code version
    - For developers or other platform users (Linux, macOS), LipidHunter source code is available.
    - Please read the instructions of LipidHunter source code version:
    - https://bitbucket.org/SysMedOs/lipidhunter
- Sample spectra
    - The sample spectra in .mzML format can be downloaded from:
    - https://bitbucket.org/SysMedOs/lipidhunter_exe/downloads/
- Updates of LipidHunter user guide
    - Please check following link for the latest version of LipidHunter user guide
    - https://bitbucket.org/SysMedOs/lipidhunter_exe/downloads/

4. Data conversion to .mzML

LipidHunter is designed to work with mzML files obtained from LC-MS/MS and shotgun data-dependent acquisition experiments. Original data should be converted to .mzML files using ProteoWizard MSconvert tool.

- Go to http://proteowizard.sourceforge.net/downloads.shtml
- Download the version suitable for your system
- Install ProteoWizard
- Open MSConvert from downloaded Folder

!!! General Notes

- The conversion of .mzML is the most critical step for LipidHunter workflow. The .mzML file provided in the LipidHunter test file were converted by ProteoWizard version 3.0.9134.
- The ProteoWizard version you download from the ProteoWizard website might be different.
- The .mzML file converted from specific instruments by different ProteoWizard version might be different, thus they might not compatible with LipidHunter. Please find the suitable ProteoWizard version to convert your raw files and keep using the same ProteoWizard version for further analysis if there are not critical updates of ProteoWizard.
- If you failed to convert or the .mzML cannot be processed by LipidHunter, please read our wiki to understand the LipidHunter specific requirement of .mzML file. Please contact ProteoWizard team to get proper version to convert your files.
- Please read LipidHunter wiki about essential data sections in the .mzML files at: https://bitbucket.org/SysMedOs/lipidhunter/wiki/Home

To accelerate the processing speed, for **LC-MS mode**, two different .mzML files (MS1 and MS2) have to be generated from the original dataset.
MS1mzML file is a reduced version of the dataset which contain only MS survey scan information and will be used for Bulk Search step.
MS2mzML file is a full data conversion which contains both MS1 and MS2 information.
For **shotgun mode**, use the **MS2mzML file** for step I is also supported.

4.1 Convert MS1 to .mzML

To generate **MS1mzML file** containing only MS survey scan information:

- Input your data in .raw format
- Click “Add”
- Choose binary coding precision “32 bit” to minimize the file size
- Choose “MS Level” and type “1-1”
- Click “Add"
- Other parameters (e.g. scan time (in seconds), mzWindow, threshold(absolute) should be specified properly

Options	Required parameters
binary coding precision	32-bit
Write index	True
Use zlib compression	True
TPP compatibility	True
Package in gzip	False
Use numpress options	False to all
MS level	1-1

Options	Suggested parameters
Threshold peak filter	Absolute intensity, Most intense, 1000
Subset --> Scan Time	Scan time range in seconds
Subset --> mz Window	*m/z* range of interest

4.2 Convert MS2 to .mzML

To generate **MS2mzML file** containing both MS survey scan and MS/MS information:

- Please remove all parameters from the MS1 file conversion
- Please create a new folder and save MS2mzML file to the new folder. (Save MS2mzML file to the previous location might over write the MS1mzML file.)
- Input your data in .raw format
- Click “Add”
- Choose binary coding precision “32 bit” to minimize the file size
- Other parameters (e.g. scanTime, mzWindow, threshold) can be specified

Options	Required parameters
binary coding precision	32-bit
Write index	True
Use zlib compression	True
TPP compatibility	True
Package in gzip	False
Use numpress options	False to all
MS level	Do not apply filter here (Important for Waters .raw files)

Options	Suggested parameters
Threshold peak filter	Absolute intensity, Most intense, 10
Subset --> Scan Time	Scan time range in seconds
Subset --> mz Window	Do not apply filter here

4.3 Check converted .mzML file by ProteoWizard SeeMS tool

It is very important to check your converted mzML file before start to work with LipidHunter. The SeeMS tool in the ProteoWizard program folder provide excellent way to view .mzML files.

Please check following parameters of converted mzML files, especially MS2mzML files.

- The MS level range.
    - for Waters files, the MS-Level 1 corresponding to MS survey scan, MS-Level 2 corresponding to DDA rank 1, MS-Level 3 corresponding to DDA rank 2, thus DDA rank N experiment should have MS-Level N+1 converted.
- The MS2 spectra *m/z* range
    - Please make sure that the MS2 spectra contains *m/z* range of specific fragments and neutral losses for identification
- The spectra intensity
    - Please check the intensity of structure representative peaks. Please adjust the "Absolute intensity" filter accordingly and convert again. Absolute intensity filter above 1000 for MS2 is NOT recommended.
    - If MS survey scan is always in good quality, Absolute intensity filter in MS1mzML can be set to 2500 or even higher based on spectra quality. Please select few m/z of interest to check if the isotope distribution of [M+2] ions are visible.
    - Please start your own data with lower threshold to finish the workflow of LipidHunter. You can adjust these conversion settings carefully based on the spectra quality and from preliminary LipidHunter results.

5 LipidHunter workflow overview

LipidHunter workflow is subdivided into five main steps.

I. **Extract MS level:** extracts MS1 Data (*m/z* values and intensities) from .mzML files of different experiments and merges this data together into one file

II. **Extract Scan info:** index all precursors selected for the fragmentation to the corresponding DDA rank, MS/MS scan number and scan time

III. **Bulk search:** MS information extracted in step I is used for bulk lipid identification using LIPIDMAPS database at lipidmaps.org

IV. **Link lipids to scans:** Bulk matches are further linked to the corresponding MS and MS/MS scans

V. **Hunt for ID:** linked MS/MS scans are used to assign distinct lipid species identification based on the combination of product ions. Final results are exported as .xlsx tables and HTML files containing six-panel graphical images allowing to evaluate and verify each identification step.

!!! General Notes

- Several .xlsx and .csv tables will be generated and used by LipidHunter. To avoid "comma vs dot" derived errors, please make sure that language on the computer is set to "US English", while working with LipidHunter.
- Note that during computational steps LipidHunter will display *"Not responding"* message on the window title. Step IV and V in LC-MS mode might take very long time to run, during this time, LipidHunter GUI might freeze and do not respond to user

actions, please be patient and wait until processing finishes. LipidHunter source code version users can monitor detailed information during the run.

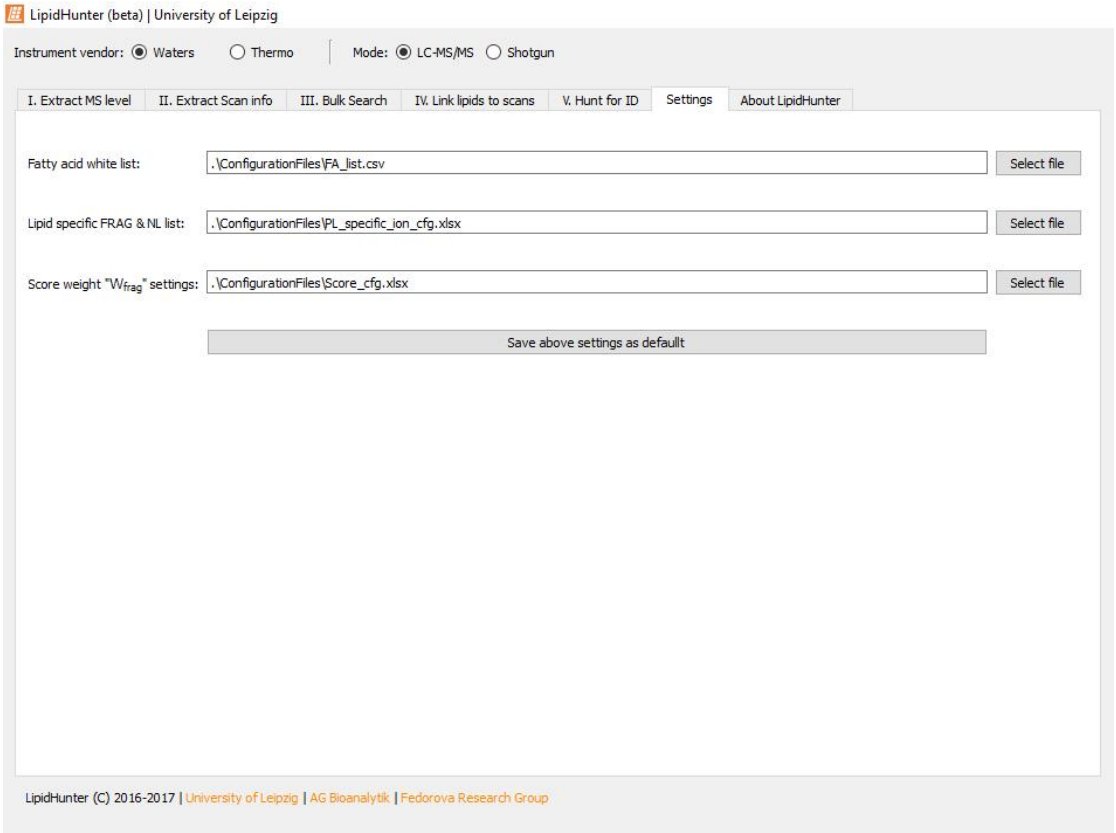
5.1 Before you start

Before you start LipidHunter workflow, please specify which Instrument vendor (Currently LipidHunter supports .mzML files generated from .raw/.RAW files from ***Waters*** and ***Thermo Fisher Scientific*** instruments; working with other files might require adjustment of file format) and experimental mode (***LC-MS/MS*** or ***shotgun***).

Configuration files containing fatty acid white list, PL specific product/neutral loss signals, and weight factor tables should be upload to the **Settings** tab. Instructions how to modify/create user specific configuration files are provided in section 6.

The default configuration files are provided in the "ConfigurationFiles" folder under LipidHunter folder. Please select the files according to the following instructions and press "save above settings as default" button to save the settings.

Please restart LipidHunter, and check the **Settings** tab to see if the configure files are loaded correctly before you start.



5.2 Step I Extract MS level

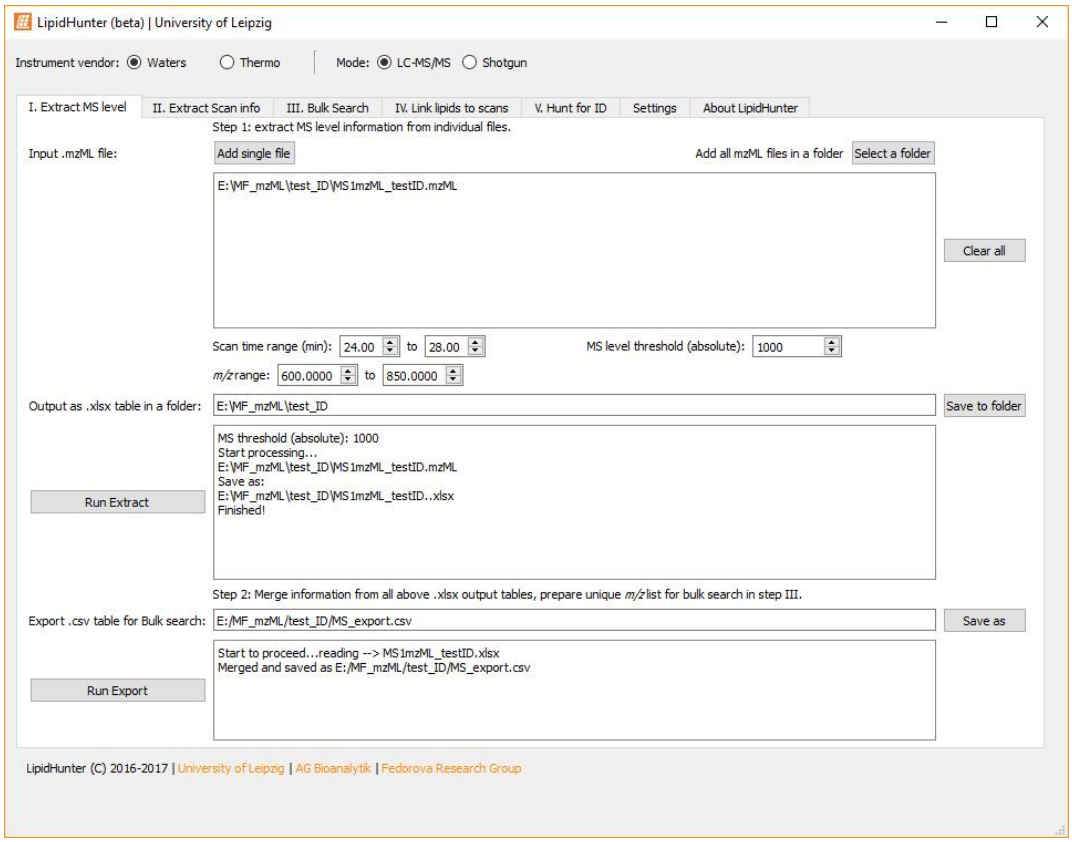
This step will use **MS1mzML** files to extract MS survey scan information into .csv table

- Click *"**Add single file**"* to use a single MS1mzML file for the extraction or click *"**Select a Folder**"* to select multiple files.
  ***!!! Note**: Combination of multiple files is possible only at this step of the workflow and will allow you to generate a merged output .csv table which will be used in a single*

Bulk Search reducing analysis time. All other steps of LipidHunter workflow will consider individual files.

- Set *"**Scan time range**"* to specify retention time range to be considered for the identification (e.g. at experimental conditions used to run test_ID sample, PE lipids elute between 24 and 28 min of LC gradient).
- Set *"**MS level threshold (absolute)**"* above which MS signals will be extracted (this values is data dependent and need to be determined from raw LC-MS data for each experiment for optimal performance; e.g. here we used 1000 counts).
- Set *"**m/z range**"* to choose a range of *m/z* values to be extracted (e.g. here we expect PE lipids to be detected between 600 and 850 *m/z*)
- Select an output folder in which the generated data will be saved
- The output .xlsx file will be saved in the selected folder under the same named of corresponding mzML file
- Click *"**Run Extract**"*
- When extraction is complete you will get the message "*Finished!*"
- Select an output folder to save merged results

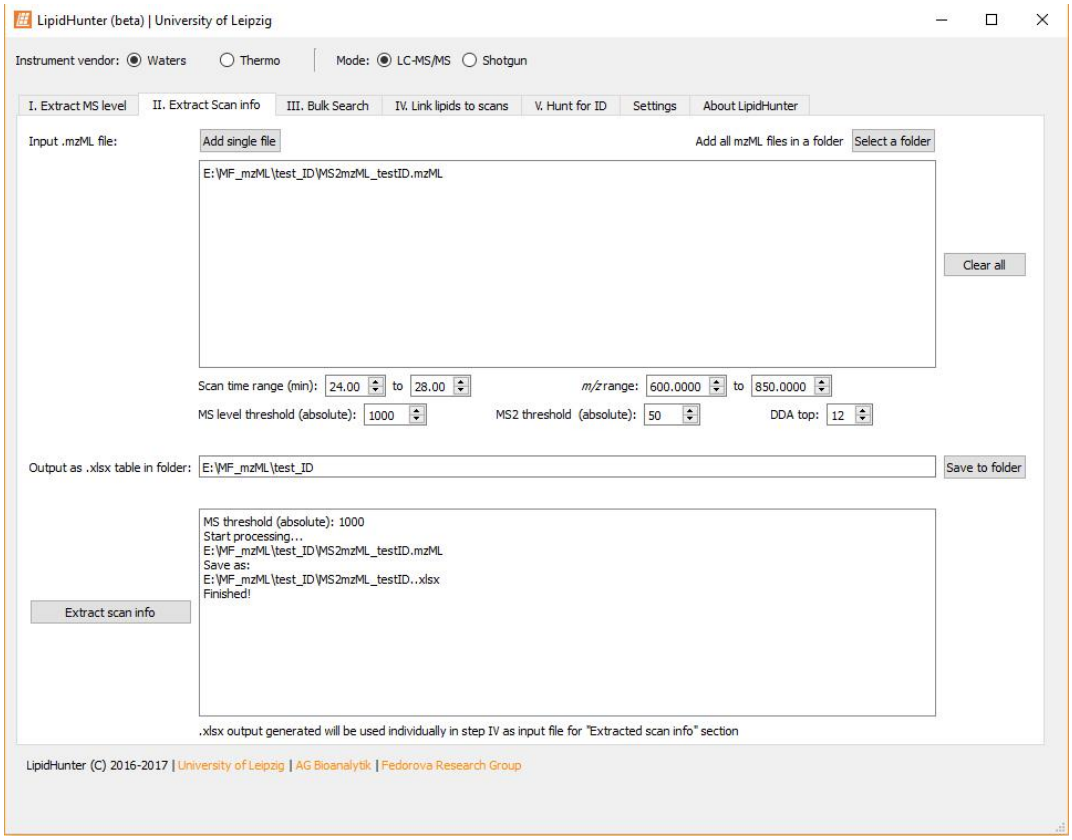
  ***!!! Note**: it is necessary to run merging step even if you have a single file*
- Click "***Run Export***" to generate .csv file which contain information on *m/z* values and corresponding intensities.



5.3 Step II Extract Scan info

This step will use **MS2mzML** files to extract MS/MS scans information into .xslx table

- Click *"**Add single file**"* to use a single MS2mzML file for the extraction or click *"**Select a Folder**"* to select multiple files.
  ***!!! Note**: Selecting multiple files for extraction you can reduce time but each individual files will result in separate output; results will not be merged.*
- Set *"**Scan time range**"* to specify retention time range to be considered for the identification (e.g. at experimental conditions used to run test_ID sample, PE lipids elute between 24 and 28 min of LC gradient).
- Set *"**m/z range**"* to choose a range of *m/z* values to be extracted (e.g. here we expect PE lipids to be detected between 600 and 850 *m/z*)



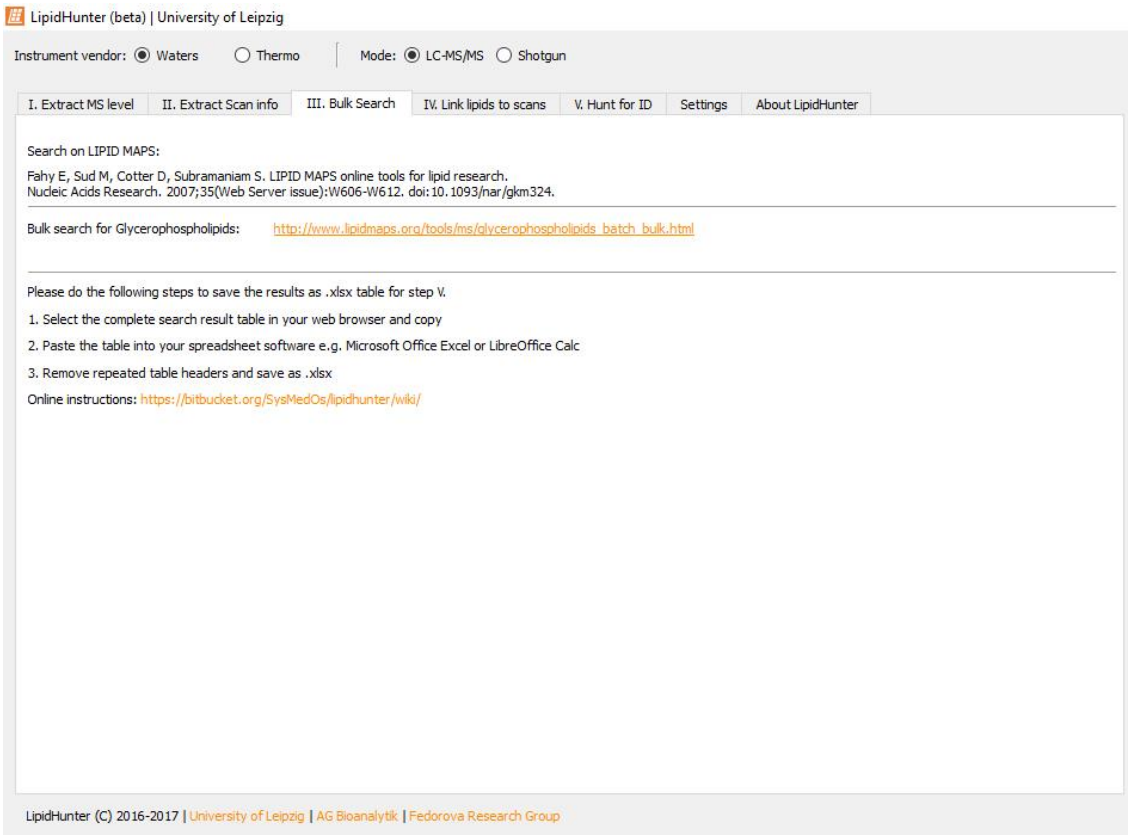
- Set *"**MS level threshold (absolute)**"* above which MS signals will be extracted (this values is data dependent and need to be determined from raw data for each experiment for optimal performance; e.g. here we used 1000 counts).
- Set "***MS2 threshold (absolute)***" above which signals from MS/MS scans will be extracted (this values is data dependent and need to be determined from raw data for each experiment for optimal performance; e.g. here we used 50 counts)
- Set "***DDA top***" to define number of MS/MS evens in each DDA duty cycle (e.g. here DDA top 12 experiment was used)
- Select an output folder in which the generated data will be saved

- Click *"**Extract Scan info**"*
- When extraction is complete you will get the message "*Finished!*" and the file *your_name*_ms2_info.xslx will be generated (e.g. here MS2mzML_testID_ms2_info.xslx).
- This file contains information on *m/z* value for each precursor (MS2_PR_mz), scan time, scan rank and corresponding DDA function (DDA_rank).

	A	B	C	D	E
1	DDA_rank	scan_number	scan_time	MS2_PR_mz	
2	1	328	24.0058498	800.9472046	
3	6	328	24.0304832	782.9272461	
4	7	328	24.0354004	778.911438	
5	1	329	24.0779171	790.9162598	
6	2	329	24.0828667	740.8892822	
7	3	329	24.0877991	840.9554443	
8	5	329	24.0976505	803.9539795	
9	6	329	24.1025829	793.9164429	
10	8	329	24.1124325	810.979126	
11	1	330	24.1499672	747.944458	
12	2	330	24.1548996	812.9581299	
13	3	330	24.1598167	762.9150391	
14	5	330	24.1696491	712.8512573	
15	7	330	24.1794834	766.9257813	
16	8	330	24.1844006	843.9602051	
17	9	330	24.1893501	744.8775635	
18	10	330	24.1942673	832.9691162	
19	1	331	24.2221508	826.979248	
20	2	331	24.2271004	737.90802	
21	4	331	24.2497005	788.9397583	

5.4 Step III_Bulk Search

Bulk identification of lipids is performed using Lipid Maps database and online tool "bulk glycerophospholipids search" (Fahy E. et al. LIPID MAPS online tools for lipid research. Nucleic Acid Research. 2007. 35 (Web Server issue): W606-W612).



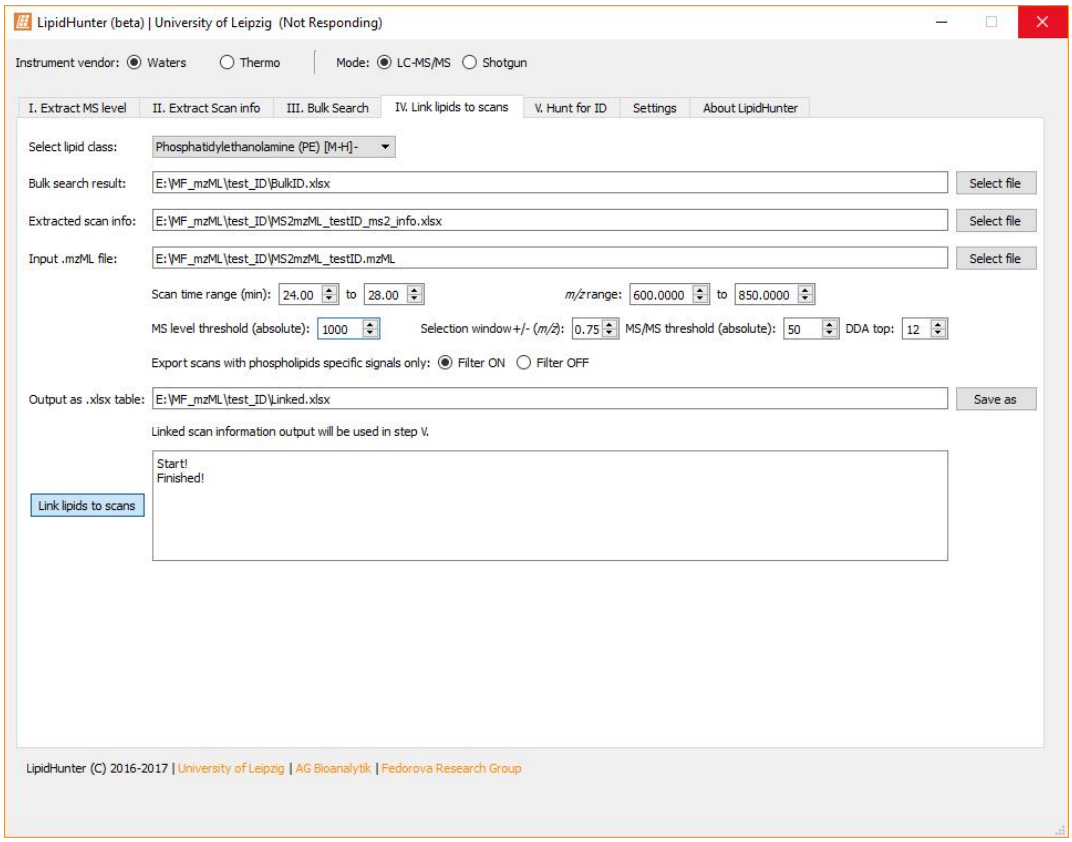
- Follow the link provided at the LipidHunter step III tab (http://www.lipidmaps.org/tools/ms/glycerophospholipids_batch_bulk.html)
- Open the MS Export .csv table obtained at Step I, copy to the input window of bulk search tool

  !!! Note: *if necessary .csv file containing comma separated values should be transformed into an excel workbook as following: highlight column A, Go to header "data", Select "Text to Columns", Select "Delimiters: Comma".*
- In the Bulk Search Lipid Maps tool interface window select MS threshold, PL class, mass tolerance, ion type and FA compositions you wish to consider for your bulk identifications (e.g. here we selected 1000 MS threshold, mass tolerance of 0.01, PE head group, and $[M-H]^-$ ion).

  !!! Note: *FA composition should correspond to the one provided in LipidHunter configuration file.* ***Please search each class of phospholipid individually.***
- Select "***Limit search to molecules containing ONLY the checked chains***"
- Submit the Bulk Search
- When Bulk Search is finished, select and copy all results in a new excel sheet
- Sort them according to "*matched mass*" (Excel "Data"→"Sort")
- Delete duplicated headers at the end of the list
- Save the table (e.g. *BulkID.xlsx*)

5.5 Step IV Link Lipids to Scans

In this step the results of the bulk search are linked to the corresponding MS/MS scans.



- Select PL class with a corresponding ion type
  !!! Note: *take care that it is the same one you used for the Bulk Search*
- "***Bulk search result***": Select output file from the Bulk Search step (e.g here BulkID.xlsx)
- "***Extracted scan info***": select output file from step II (e.g. here MS2mzML_testID_ms2_info.xlsx)
- "***Input .mzML file***": select corresponding MS2mzML file
- Define parameters as described above for previous steps
- Set "***Selection window +/- m/z***" to define the mass range used to fit MS1 and MS2 reported precursors
  !!! Note: *precursor m/z values reported for MS/MS scans might differe from MS1 reported precursor in .RAW Waters files or when LTQ offset is used on Thermo instruments.*
- You can choose to activate filter "***Export scans with phospholipids specific signals only***" – it will allow to filter out all MS/MS scan which do not contain fragment ions specific for certain PL class (loss of 60 amu from precursor for PC formate adducts, presence of deprotonated precursor ion for PE, and loss of 87 amu from precursor for deprotonated PS ions). This preliminary filtering step allows to increase specificity and reduce processing time during step V.

!!! Note: *filter available only for [PE-H]$^-$, [PS-H]$^-$ and [PC+HCOO]$^-$ ions.*

- Select an output folder in which the generated data will be saved (e.g. here Linked.xlsx)
- Click *"**Link Lipids to scans**"*

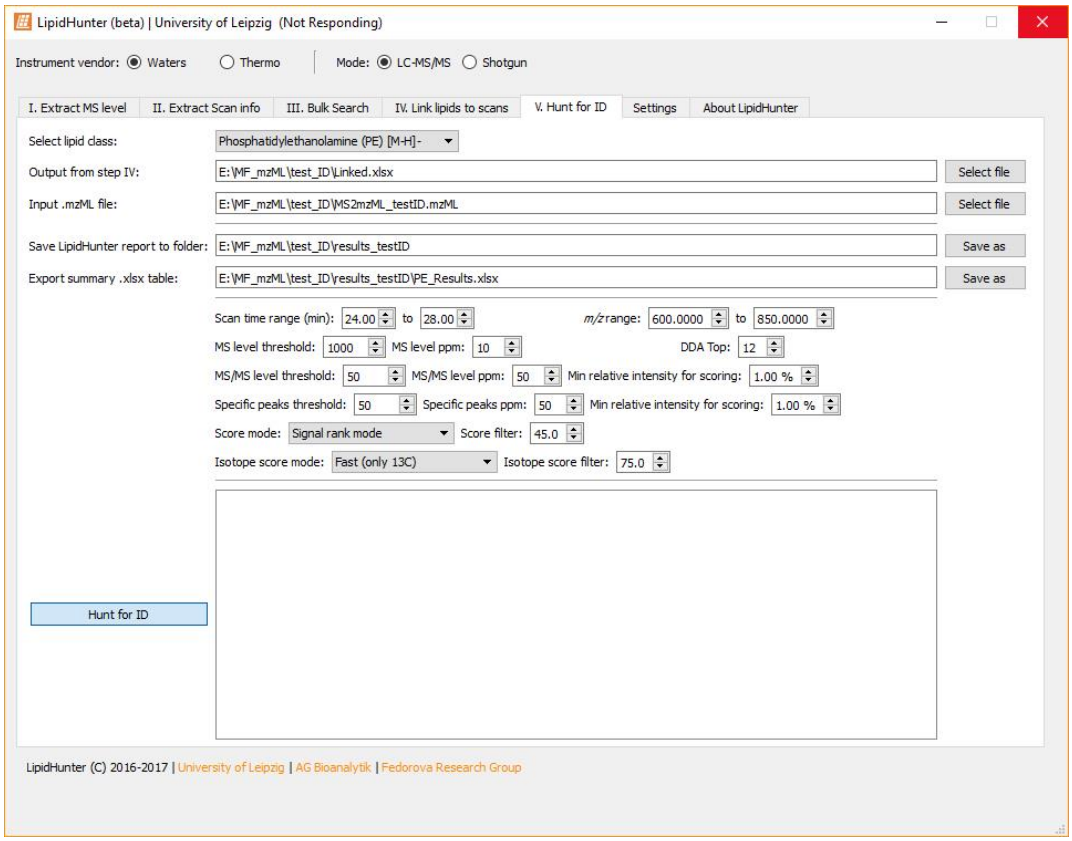
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DDA_rank	scan_number	MS2_PR_mz	scan_time	MS1_obs_mz	Lib_mz	Abbreviation	Formula	Ion	Class	DDA_even	MZ_PR	I_PR
2	1	329	790.91626	24.0779171	790.4931641	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
3	1	329	790.91626	24.0779171	790.5054321	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
4	1	329	790.91626	24.0779171	790.5177002	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
5	1	329	790.91626	24.0779171	790.5299683	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
6	1	329	790.91626	24.0779171	790.5422974	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
7	1	329	790.91626	24.0779171	790.5545654	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
8	1	329	790.91626	24.0779171	790.5668335	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
9	1	329	790.91626	24.0779171	790.5791016	790.5392	PE(40:6)	C45H77NO8P	[M-H]-	PE	1	790.53	699
10	1	330	747.944458	24.1499672	748.4890747	748.5286	PE(P-38:5)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
11	1	330	747.944458	24.1499672	748.4890747	748.5286	PE(O-38:6)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
12	1	330	747.944458	24.1499672	748.5010376	748.5286	PE(P-38:5)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
13	1	330	747.944458	24.1499672	748.5010376	748.5286	PE(O-38:6)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
14	1	330	747.944458	24.1499672	748.5130005	748.5286	PE(P-38:5)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
15	1	330	747.944458	24.1499672	748.5130005	748.5286	PE(O-38:6)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
16	1	330	747.944458	24.1499672	748.5249634	748.5286	PE(P-38:5)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
17	1	330	747.944458	24.1499672	748.5249634	748.5286	PE(O-38:6)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
18	1	330	747.944458	24.1499672	748.5368652	748.5286	PE(P-38:5)	C43H75NO7P	[M-H]-	PE	2	748.1427	51
19	1	330	747.944458	24.1499672	748.5368652	748.5286	PE(O-38:6)	C43H75NO7P	[M-H]-	PE	2	748.1427	51

The output excel file contains information about MS1 precursor masses, assosiated bulk identifications, and matching MS2 scans.

5.6 Step V Hunt for ID

This is the core function of LipidHunter which rely on the linked bulk identifications and scan information from step IV to perform assignment of MS/MS spectra.

Since Step V might take relative long time to run under LC-MS/MS mode. We recommend to restart LipidHunter before you run Step V.



- Select PL class with a corresponding ion type
  !!! Note: *take care that it is the same .mzML file you used for the Bulk Search and step IV*
- "***Output from step IV***": select file generetaed at step IV (e.g. here Linked.xlsx)
- "***Input .mzML file***": select corresponding MS2mzML file
- "***Save LipidHunter report to folder***": select folder where HTML report file and folder with images will be saved
- "***Export summary .xlsx table***": select folder where .xlsx summury will be saved
- Define parameters as described above for previous steps
- Set "**MS level ppm**" to define mass accuracy range for MS1 signals (! Instrument dependent; e.g. here we used 10 ppm)
- Set "**MS/MS level ppm**" to define mass accuracy range for MS2 signals (! Instrument dependent; e.g. here we used 50 ppm)
- Set "***Specific peaks threshold***" above which PL class specific signals from MS/MS scans will be considered (this values is data dependent and need to be determined from raw data for each experiment for optimal performance; e.g. here we used 50 counts)
- Set "***Min relative intensity for scoring***" above which FA and PL class specific fragment ions will be considered (this values is data dependent and need to be determined from raw data for each experiment for optimal performance; e.g. here we used 1 %)

Select "***Score mode***": Scoring mode refelects how LipidHunter will calcualte *Rank factor* R_{frag}. In "***Relative intensity mode***" R_{frag} directly corresponds to the relative intensity of the signal.

Whereas in the "***Signal rank mode***", the R_{frag} is calculated based on the *RankIndex*. "Signal rank mode" is more suitable for identification of as many lipids as possible in the complex samples with partially co-eluting (and thus co-fragmenting) structural isomers. Whereas "Relative intensity mode" will provide more strict identification of the most abundant isomer.

- Select "***Isotope score mode***": using "***All elements***" all elements (exept phosphorus) are used to calculate theoretical isotopic distribution, whereas in "***Fast*** (***only 13C)***" mode only carbon is used for the calculations. For PL species, ^{13}C isotope has a dominant contribution in isotopic pattern and thus "*only 13C*" mode is usually sufficient.
- "***Score filter***" and "***Isotope score filter***" for results output can be applied (e.g. here 45 and 75, respectively)
- Click "***Hunt for ID***"!
- LipidHunter create a log file named "LipidHunter_Params-Log_YYYY-MM-DD_HH-min.txt" and a HTML file with associated folder of image and "LipidHunter_Results_YYYY-MM-DD_HH-min.html".
- You can open the HTML report in your web browser and refresh from time to time to check the latest results. It may take few minutes until the first image result generated. We recommend Mozilla Firefox, Chrome or Chromium to open the HTML report.

6 LipidHunter results output

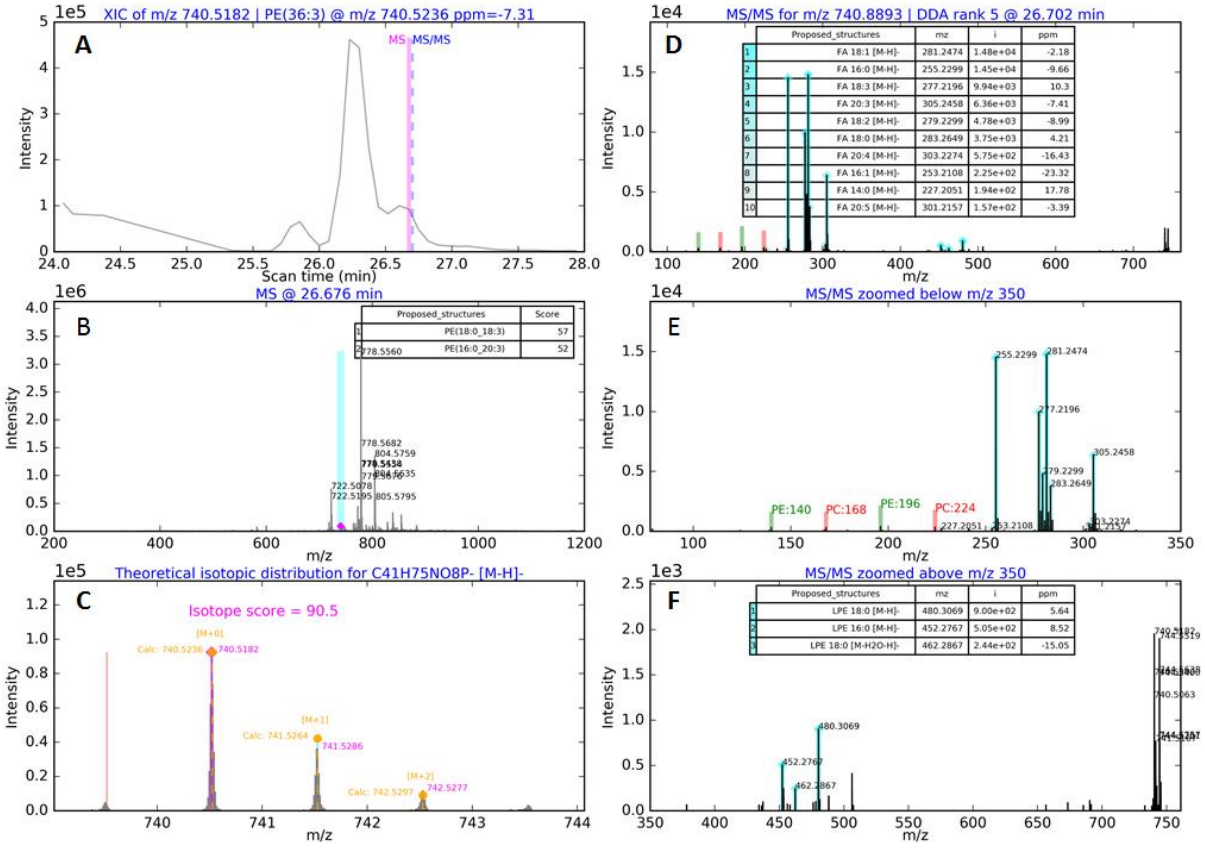
For each submitted dataset LipidHunter provides the **output .xlsx table** which summarize lipid identities (bulk identification, proposed discrete structure, elemental composition, theoretical and observed *m/z* values, mass accuracy, retention time), identification metrics (LipidHunter and isotope scores, relative intensities of matched fragments, PL specific and unspecific signals) and data specific details (DDA rank, scan number).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Bulk_identification	Proposed_structures	Formula_neutral	Formula_ion	Charge	Lib_mz	ppm	LipidHunter_Score	MS1_obs_mz	MS1_obs_i	Isotope_score	MS2_PR_mz	MS2_scan_time	DDA#	Scan#
2	PE(40:6)	PE(18:0_22:6)	C45H78NO8P	C45H77NO8P-	[M-H]-	790.5392	3.92	72	790.5423	2.94e+05	88.36	790.928528	27.347	1	373
3	PE(40:7)	PE(18:1_22:6)	C45H76NO8P	C45H75NO8P-	[M-H]-	788.5236	-7.4	67	788.5178	8.40e+04	98.08	788.915222	25.649	2	350
4	PE(38:4)	PE(18:1_20:3)	C43H78NO8P	C43H77NO8P-	[M-H]-	766.5392	1.57	62	766.5404	4.92e+04	77.20	766.925781	26.833	2	366
5	PE(38:5)	PE(18:1_20:4)	C43H76NO8P	C43H75NO8P-	[M-H]-	764.5236	-1.09	66	764.5228	3.26e+04	98.90	764.907043	27.077	5	369
6	PE(38:5)	PE(16:0_22:5)	C43H76NO8P	C43H75NO8P-	[M-H]-	764.5236	-1.09	50	764.5228	3.26e+04	98.90	764.907043	27.077	5	369
7	PE(38:6)	PE(18:2_20:4)	C43H74NO8P	C43H73NO8P-	[M-H]-	762.5079	-0.2	72	762.5078	7.04e+04	89.84	762.89093	24.683	4	337
8	PE(38:6)	PE(18:1_20:5)	C43H74NO8P	C43H73NO8P-	[M-H]-	762.5079	-0.2	52	762.5078	7.04e+04	89.84	762.89093	24.683	4	337
9	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43H75NO7P-	[M-H]-	748.5286	-4.86	50	748.525	7.08e+04	90.35	748.900452	26.717	8	364
10	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43H75NO7P-	[M-H]-	748.5286	-4.86	47	748.525	6.61e+05	85.39	748.900452	27.202	1	371
11	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43H75NO7P-	[M-H]-	748.5286	-4.86	45	748.525	7.04e+04	96.07	748.900452	27.722	1	378
12	PE(P-38:6)	PE(P-16:0_22:6)	C43H74NO7P	C43H73NO7P-	[M-H]-	746.513	-7.69	45	746.5073	5.46e+04	84.19	746.881653	26.111	5	356
13	PE(36:3)	PE(18:0_18:3)	C41H76NO8P	C41H75NO8P-	[M-H]-	740.5236	-7.31	57	740.5182	9.24e+04	90.48	740.889282	26.702	5	364
14	PE(36:3)	PE(16:0_20:3)	C41H76NO8P	C41H75NO8P-	[M-H]-	740.5236	-7.31	52	740.5182	9.24e+04	90.48	740.889282	26.702	5	364
15	PE(P-36:5)	PE(P-16:0_20:5)	C41H72NO7P	C41H71NO7P-	[M-H]-	720.4973	6.63	50	720.5021	8.51e+04	96.38	720.862244	25.492	2	348
16	PE(38:5)	PE(18:0_20:5)	C43H76NO8P	C43H75NO8P-	[M-H]-	764.5236	-1.09	67	764.5228	9.74e+04	91.89	764.907043	26.53	2	362
17	PE(36:2)	PE(18:0_18:2)	C41H78NO8P	C41H77NO8P-	[M-H]-	742.5392	0.55	62	742.5396	6.73e+05	93.19	742.911804	27.956	2	381
18	PE(36:3)	PE(18:1_18:2)	C41H76NO8P	C41H75NO8P-	[M-H]-	740.5236	-7.31	61	740.5182	3.68e+04	92.34	740.87738	24.71	7	337
19	PE(36:3)	PE(16:0_20:3)	C41H76NO8P	C41H75NO8P-	[M-H]-	740.5236	-7.31	56	740.5182	3.68e+04	92.34	740.87738	24.71	7	337
20	PE(36:4)	PE(18:2_18:2)	C41H74NO8P	C41H73NO8P-	[M-H]-	738.5079	4.68	100	738.5114	3.71e+04	78.10	738.881348	24.767	4	338
21	PE(P-36:3)	PE(P-16:0_20:3)	C41H76NO7P	C41H75NO7P-	[M-H]-	724.5286	-0.81	50	724.528	1.08e+05	94.69	724.890381	27.51	2	375
22	PE(P-36:4)	PE(P-16:0_20:4)	C41H74NO7P	C41H73NO7P-	[M-H]-	722.513	-7.26	50	722.5078	5.29e+05	93.89	722.869019	26.972	1	368
23	PE(34:1)	PE(16:0_18:1)	C39H76NO8P	C39H75NO8P-	[M-H]-	716.5236	-1.67	60	716.5224	2.44e+04	91.03	716.880371	27.742	5	378
24	PE(34:3)	PE(16:0_18:3)	C39H72NO8P	C39H71NO8P-	[M-H]-	712.4923	4.47	81	712.4955	4.27e+04	91.33	712.851257	24.715	8	337
25	PE(40:5)	PE(18:0_22:5)	C45H80NO8P	C45H79NO8P-	[M-H]-	792.5549	2.74	49	792.5571	3.84e+04	87.48	792.969116	27.13	1	370
26	PE(38:4)	PE(18:0_20:4)	C43H78NO8P	C43H77NO8P-	[M-H]-	766.5392	1.57	47	766.5404	1.32e+05	86.36	766.913696	25.203	2	344
27	PE(36:1)	PE(18:0_18:1)	C41H80NO8P	C41H79NO8P-	[M-H]-	744.5549	-3.97	62	744.5519	7.29e+04	76.09	744.925232	26.712	7	364
28	PE(34:1)	PE(16:0_18:1)	C39H76NO8P	C39H75NO8P-	[M-H]-	716.5236	-1.67	60	716.5224	1.71e+05	94.08	716.868652	25.113	1	343
29	PE(34:1)	PE(18:0_16:1)	C39H76NO8P	C39H75NO8P-	[M-H]-	716.5236	-1.67	49	716.5224	1.71e+05	94.08	716.868652	25.113	1	343

PL class specific (marked here green) and unspecific (red) fragment ions are not considered for LipidHunter score calculations but can be used to provide additional confidence criteria:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Bulk_identification	Proposed_structures	Formula_neutral	hula	harg	b_ppm	nte	obs	olpe	_PR	can	DDA	can	i_sn1	i_sn2	i_[M-H]-sn1	i_[M-H]-sn2	M-H]-sn1-H	i_[M-H]-sn2-H2O	#Specific_peaks	PE:196	PE:140	#Unspecific_peaks
2	PE(40:6)	PE(18:0_22:6)	C45H78NO8P	C45 [M-	##	4	72	##	2.94	88.1	##	27	1	##	100	60.13		12.92		2.96	2	5.59	2.99	1
3	PE(40:7)	PE(18:1_22:6)	C45H76NO8P	C45 [M-	##	-7	67	##	8.40	98.1	##	26	2	##	100	37.61		5.08		2.36	2	4.32	1.79	2
4	PE(38:4)	PE(18:1_20:3)	C43H78NO8P	C43 [M-	##	2	62	##	4.92	77.2	##	27	2	##	52.3	100		4.39			2	2.24	1.03	0
5	PE(38:5)	PE(18:1_20:4)	C43H76NO8P	C43 [M-	##	-1	66	##	3.26	98.1	##	27	5	##	98.17	86.74		13.11		1.63	2	9.85	2.97	3
6	PE(38:5)	PE(16:0_22:5)	C43H76NO8P	C43 [M-	##	-1	50	##	3.26	98.1	##	27	5	##	100	15.26		5.78			2	9.85	2.97	3
7	PE(38:6)	PE(18:2_20:4)	C43H74NO8P	C43 [M-	##	-0	72	##	7.04	89.1	##	25	4	##	66.43	100		4.12		2.22	2	2.95	1.84	3
8	PE(38:6)	PE(18:1_20:5)	C43H74NO8P	C43 [M-	##	-0	52	##	7.04	89.1	##	25	4	##	57.29	11.51		3.33		1.72	2	2.95	1.84	3
9	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43 [M-	##	-5	50	##	7.01	90.1	##	27	8	##		100		19.22		9.54	2	3.56	2.71	0
10	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43 [M-	##	-5	47	##	6.61	85.1	##	27	1	##		50.16		6.1		2.63	2	2.77	1.1	0
11	PE(P-38:5)	PE(P-16:0_22:5)	C43H76NO7P	C43 [M-	##	-5	45	##	7.04	96.1	##	28	1	##		31.99		5.92		3.31	2	2.36	1.86	1
12	PE(P-38:6)	PE(P-16:0_22:6)	C43H74NO7P	C43 [M-	##	-8	45	##	5.40	84.1	##	26	5	##		39.6		21.04		11.16	2	6.44	4	2
13	PE(36:3)	PE(18:0_18:3)	C41H76NO8P	C41 [M-	##	-7	57	##	9.24	90.4	##	27	5	##	25.41	67.34		6.1		1.65	2	2.68	1.99	2
14	PE(36:3)	PE(16:0_20:3)	C41H76NO8P	C41 [M-	##	-7	52	##	9.24	90.4	##	27	5	##	98.4	43.1		3.42			2	2.68	1.99	2
15	PE(P-36:5)	PE(P-16:0_20:5)	C41H72NO7P	C41 [M-	##	7	50	##	8.51	96.1	##	25	2	##		100		31.27		14.5	2	4.14	2.02	0
16	PE(38:5)	PE(18:0_20:5)	C43H76NO8P	C43 [M-	##	-1	67	##	9.74	91.1	##	27	2	##	16.78	17.38		5.72		2.01	1	1.46		2
17	PE(36:2)	PE(18:0_18:2)	C41H78NO8P	C41 [M-	##	1	62	##	6.73	93.1	##	28	2	##	31.47	100		3.19			1	1.84		0
18	PE(36:3)	PE(18:1_18:2)	C41H76NO8P	C41 [M-	##	-7	61	##	3.66	92.1	##	25	7	##	7.3	64.77	1.32		4.5	2.72	1	3.43		2
19	PE(36:3)	PE(16:0_20:3)	C41H76NO8P	C41 [M-	##	-7	56	##	3.66	92.1	##	25	7	##	100	4.18		3.43	1.09	4.62	1	3.43		2
20	PE(36:4)	PE(18:2_18:2)	C41H74NO8P	C41 [M-	##	5	##	##	3.71	78.1	##	25	4	##	93.66	93.66	3.2	3.2	1.08	1.08	1	1.25		1
21	PE(P-36:3)	PE(P-16:0_20:3)	C41H76NO7P	C41 [M-	##	-1	50	##	1.08	94.1	##	28	2	##		100		10.62		5.13	1	1.68		0
22	PE(P-36:4)	PE(P-16:0_20:4)	C41H74NO7P	C41 [M-	##	-7	50	##	5.25	93.1	##	27	1	##		100		16.83		6.7	1	1.95		0
23	PE(34:1)	PE(16:0_18:1)	C39H76NO8P	C39 [M-	##	-2	60	##	2.44	91.1	##	28	5	##	86.23	100		3.58			1	1.94		0
24	PE(34:3)	PE(16:0_18:3)	C39H72NO8P	C39 [M-	##	4	81	##	4.21	91.1	##	25	8	##	100	87.12		4.59	1.2	1.33	1	3.94		2
25	PE(40:5)	PE(18:0_22:5)	C45H80NO8P	C45 [M-	##	3	49	##	3.84	87.1	##	27	1	##	16.97	3.24		6.05		2.12	0			4
26	PE(38:4)	PE(18:0_20:4)	C43H78NO8P	C43 [M-	##	2	47	##	1.32	86.1	##	25	2	##	1.01	66.52		7.64		3	0			3
27	PE(36:1)	PE(18:0_18:1)	C41H80NO8P	C41 [M-	##	-4	62	##	7.25	76.0	##	27	7	##	3.17	100		2.16		1.02	0			2
28	PE(34:1)	PE(16:0_18:1)	C39H76NO8P	C39 [M-	##	-2	60	##	1.71	94.0	##	25	1	##	76.58	2	1.73			2.59	0			2
29	PE(34:1)	PE(18:0_16:1)	C39H76NO8P	C39 [M-	##	-2	49	##	1.71	94.0	##	25	1	##		100		2.95		1.35	0			2

Furthermore, LipidHunter generate a separate **six-panel image for each identified lipid**:



Output image generated by LipidHunter for precursor at *m/z* 740.518 identified as PE(16:0_20:3) lipid.

Panel A: Extracted ion chromatogram (XIC) for ion observed at *m/z* 740.5182. Results of bulk identification [PE(36:3)], theoretical (740.5236) and observed (740.5182) *m/z* values, mass accuracy error (-7.31 ppm) are listed at the header. Positions of MS (magenta line) and MS/MS (blue dashed line) scans used for the identification are indicated on XIC.

Panel B: MS spectrum with the corresponding precursor marked (magenta diamond indicating apex of the monoisotopic signal) within defined mass accuracy range (cyan bar). Corresponding MS scan time (26.676 min) is listed at the header.

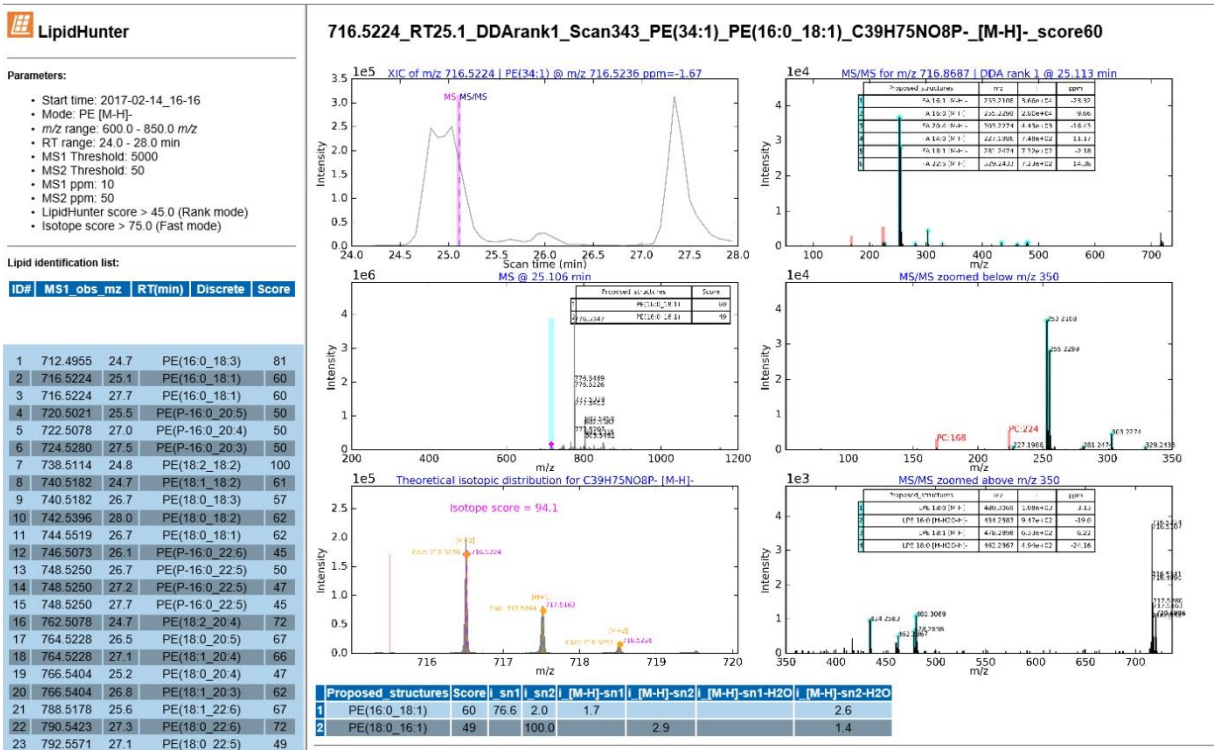
Panel C: Zoomed MS spectrum illustrating precursor isotopic distribution. Elemental composition used for calculations is listed at the header. Pseudo [M+0] signal marked with a magenta bar (width of the bar corresponds to a defined mass accuracy range; height of the bar corresponds to the precursor signal intensity). Theoretically calculated *m/z* values for the precursor isotopic signals are listed (orange text) and marked (dashed orange line with apex of the signal indicated by orange circle). Defined mass accuracy range shown by the width of a cyan bar. Height of the bar corresponds to the calculated intensities of theoretical isotopic signals. *M/z* values for the observed precursor isotopic signals are listed (magenta text) and marked (magenta line with the apex of the signal indicated by magenta diamond). Calculated isotope score shown at the upper right corner (magenta text).

Panel D: MS/MS spectrum used for the assignment. *m/z* value of the precursor listed for the corresponding MS/MS scan (740.8893), DDA rank (5) and MS/MS scan time (26.702 min) are listed at the header. Product ions assigned to the specific signals are marked with cyan (FA product and neutral loss ions), green (PL class specific) and magenta (PL class unspecific product or neutral loss ions) lines. Top 10 FA specific product ions ranked according to their intensities (used for *Rank factor* R_{frag} calculations) are listed in the table (proposed structure, *m/z*, intensity and mass accuracy error).

Panel E: MS/MS spectrum zoomed below *m/z* 350 for better illustration of FA (cyan), PL class specific (green) and unspecific (magenta lines) product ions.

Panel F: MS/MS spectrum zoomed above *m*/z 350 for better illustration of FA (cyan), PL class specific (green) and unspecific (magenta lines) neutral loss ions. Top 10 FA specific neutral loss ions ranked according to their intensities (used for *Rank factor* R_{frag} calculations) are listed in the table (proposed structure, *m/z*, intensity and mass accuracy error).

All output images are integrated and indexed in an informative **HTML report file** for manual reviewing. Availability of the graphical data representation and organized report files allows fast evaluation of identification results. The report HTML file with the built-in identification table and a log of the corresponding LipidHunter parameters provides a simple solution for data tracking and storage.

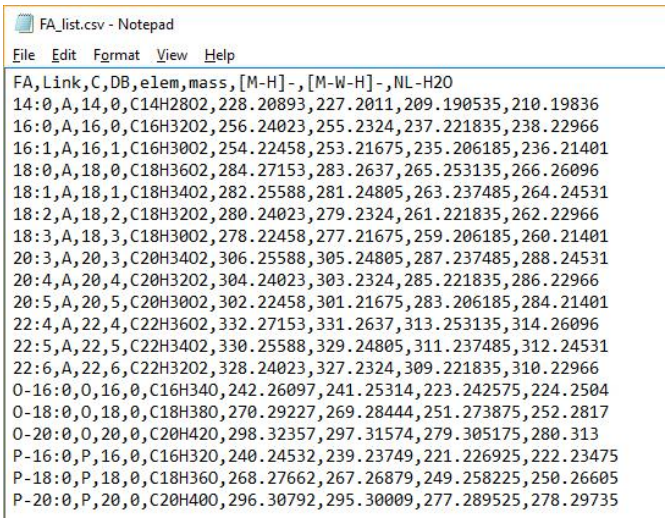


ID#	MS1_obs_mz	RT(min)	Discrete	Score
1	712.4955	24.7	PE(16:0_18:3)	81
2	716.5224	25.1	PE(16:0_18:1)	60
3	716.5224	27.7	PE(16:0_18:1)	60
4	720.5021	25.5	PE(P-16:0_20:5)	50
5	722.5078	27.0	PE(P-16:0_20:4)	50
6	724.5280	27.5	PE(P-16:0_20:3)	50
7	738.5114	24.8	PE(18:2_18:2)	100
8	740.5182	24.7	PE(18:1_18:2)	61
9	740.5182	26.7	PE(18:0_18:3)	57
10	742.5396	28.0	PE(18:0_18:2)	62
11	744.5519	26.7	PE(18:0_18:1)	62
12	746.5073	26.1	PE(P-16:0_22:6)	45
13	748.5250	26.7	PE(P-16:0_22:5)	50
14	748.5250	27.2	PE(P-16:0_22:5)	47
15	748.5250	27.7	PE(P-16:0_22:5)	45
16	762.5078	24.7	PE(18:2_20:4)	72
17	764.5228	26.5	PE(18:0_20:5)	67
18	764.5228	27.1	PE(18:1_20:4)	66
19	766.5404	25.2	PE(18:0_20:4)	47
20	766.5404	26.8	PE(18:1_20:3)	62
21	788.5178	25.6	PE(18:1_22:6)	67
22	790.5423	27.3	PE(18:0_22:6)	72
23	792.5571	27.1	PE(18:0_22:5)	49

Proposed_structures	Score	i_sn1	i_sn2	i_[M-H]-sn1	i_[M-H]-sn2	i_[M-H]-sn1-H2O	i_[M-H]-sn2-H2O
1 PE(16:0_18:1)	60	76.6	2.0	1.7			2.6
2 PE(18:0_16:1)	49		100.0		2.9		1.4

7 Configuration files

You can easily modify/create configuration files to define your own FA preferences using template files provided with LipidHunter.



!!! Note: *If you edit this file in Excel, take care that FA abbreviation e.g. 14:0 are saved correctly and not converted in the time format (e.g. 14:00:00). Check modified .csv file by opening it with Notepad.*

You can also adapt PL specific fragments list and weight factors depending on the MS instruments, collision energies and ion adducts used in the study by modifying simple .xlsx tables provided with LipidHunter.

	A	B	C	D	E	F	G	H	I	J
1	CLASS	TYPE	EXACTMASS	FORMULA	CHARGE_MODE	PR_CHARGE	LABEL	REMARKS		
2	PA	NL	97.9769	H3O4P	NEG	[M-H]-	PA:-98	-PA Head Group		
3										
4	PC	FRAG	168.0458	C4H11O4NP-	NEG	[M+HCOO]-	PC:168	demethylated PC [M-H]-		
5	PC	FRAG	224.0688	C7H15O5NP-	NEG	[M+HCOO]-	PC:224	demethylated PC dehydrated glycerol ester [M-H]-		
6	PC	FRAG	242.0794	C7H17O6NP-	NEG	[M+HCOO]-	PC:242	demethylated PC glycerol ester [M-H]-		
7	PC	NL	60.0211	C2H4O2	NEG	[M+HCOO]-	PC:-60	-methyl formate (-CH3COOH)		
8	PC	NL	183.0660	C5H14NO4P	NEG	[M+HCOO]-	PC:-183	-PC Head Group		
9										
10	PE	FRAG	140.0113	C2H7O4NP-	NEG	[M-H]-	PE:140	PE Head Group [M-H]-		
11	PE	FRAG	196.0375	C5H11O5NP-	NEG	[M-H]-	PE:196	Deprotonated doubly dehydrated glycerol phosphocholine		
12	PE	NL	141.0191	C2H8NO4P	NEG	[M-H]-	PE:-141	-PE Head Group		
13	PE	NL	43.0422	C2H5N	NEG	[M-H]-	PE:-43	-PE Head Group part		
14										
15	PG	FRAG	171.0059	C3H8O6P-	NEG	[M-H]-	PG:171	PG Head Group [M-H]-		
16	PG	FRAG	152.9953	C3H6O5P-	NEG	[M-H]-	PG:153	PG Head Group [M-H2O-H]-		
17	PG	NL	172.0137	C3H9O6P	NEG	[M-H]-	PG:-172	-PG Head Group		
18										
19	PI	FRAG	241.0113	C6H10O8P -	NEG	[M-H]-	PI:241	PI Head Group [M-H]-		
20	PI	NL	162.0528	C6H10O5	NEG	[M-H]-	PI:-162	-inositol		
21										
22	PS	FRAG	184.0011	C3H7NO6P-	NEG	[M-H]-	PS:184	PS Head Group [M-H]-		
23	PS	NL	87.0320	C3H5NO2	NEG	[M-H]-	PS:-87	-serine		
24										

	A	B	C	D	E	F
1	Type	Weight	[M-H]-	[M+HCOO]-	FA	H2O
2	sn1	25	0	0	1	0
3	sn2	25	0	0	1	0
4	[M-H]-sn1	15	1	0	-1	1
5	[M-H]-sn2	15	1	0	-1	1
6	[M-H]-sn1-H2O	10	1	0	-1	0
7	[M-H]-sn2-H2O	10	1	0	-1	0