**Release Note: LipidXplorer 1.2.8, October 14th, 2019**

**Improved Import Performance and Implementation of Frequency Filtering**

Fadi Al Machot, Nils Hoffmann, Jacobo Miranda Ackerman and Dominik Schwudke

**Aim:**

This release is focused on improving the import module and implementing the recently introduced filtering approach based on counting reoccurrences of peaks in MS-scans and MS/MS-scans conceptually following Schuhmann et al. [1]. All other general functionalities of the former releases should not be affected [2-5]. Further changes were made for better documentation and simpler user guidance. Some experimental and/or unused options were deactivated (grouping samples, heuristic hierarchical alignment, dta/csv import). Details are listed below.

**Usage:**

The frequency based filtering is active for the \*.mzXML and \*.mzML import module. For this release we have tested both \*.mzXML and \*.mzML files after conversion with msConvert of \*.raw files of a Q Exactive instrument (Thermo Fisher Scientific, Bremen, Germany)[6]. The LipidXplorer (LX) software and also this improved import module were developed for shotgun lipidomics experiments.

Figure 1) Import settings for LipidXplorer 1.2.8

An additional setting *frequency filter* can be applied. Values between *0* and *1* can be chosen representing the fraction of scans in which a certain peak has to be present. The filter is applied for both MS and MS/MS according to the given import setting for each scan type (please refer to [4]). For each precursor ion all associated MS/MS scans are collected according to *selection window*. Counting of MS/MS scans and association of precursor is done automatically. Accordingly, number of MS/MS scan can vary for each precursor ion in the shotgun acquisition, as might occur when using a DDA strategy.

**Benchmark:**

An extract of commercially available human serum (P9523, Sigma-Aldrich) was analysed with direct infusion shotgun MS/MS on a Thermo Q Exactive (in Thermo RAW format) using Thermo Xcalibur software (version 2.8-280502/2.8.1.2806), in positive ion mode comprising 31 high resolution MS scans (*m/z* 350 -1200) and 2560 MS/MS scans. An IDA approach for selecting precursors was chosen using a comprehensive inclusion list for the *m/z* range 350 – 1000. Each MS/MS was repeated four times. Ten independent technical replicates were acquired. mzML and mzXML data were generated using msConvert (ProteoWizard version 3.0.9706) with 32 bit encoding precision and vendor peak picking for centroiding of the RAW (Thermo) data and removal of extra zero samples. Conversion was performed *with* and *without* zlib scan compression to assess the effects of the compression on LX results.

In all instances, LX was evaluated using the same set of parameters, except for the *MSFilter* and *MSMSFilter* parameters that were added to the LX 1.2.8 version to implement vendor-independent frequency/repetition filtering of m/z signals.

**Benchmark Results:**

Summary:

1. We found only minor qualitative differences between LipidXplorer 1.2.7 and 1.2.8 for mzML and mzXML files when the frequency filter was not applied (LX 1.2.8 – Filter set to *ZERO*).

2. Small differences in the imports were observed depending on the usage of the zlib scan compression feature of msConvert.

3. Loading times for mzML could be decreased 23.5 fold from 3511 seconds to 149 seconds by eager loading and subsequent caching of mzML scans during the first load of scan info data. This comes at the expense of higher memory requirements, which are comparable to those of imports with mzXML files.

4. The resulting Masterscan for LX 1.2.7 using the \*.mzML format was with 68,85 MB more than two-fold larger than the 1.2.8 Masterscan (with filtering) with 30,65 MB (Table 1, Figure 2).

**Table 1) Statistics for import of mzXML and mzML between LX 1.2.7 and 1.2.8 and filter usage**

|  |  |  |  |
| --- | --- | --- | --- |
| Import format | mzXML – nozlib | mzML – nozlib | mzML |
| LX Version\* | **1.2.7** | **1.2.8** | **1.2.8** | **1.2.7** | **1.2.8** |
| Filter | - | 0.0 | 0.6 | 0.0 | 0.6 | - | 0.0 | 0.6 |
| Masterscan \*.sc size (MB) | 68.85 | 69.37 | 30.68 | 69.37 | 30.68 | 68.85 | 68.85 | 30.65 |
| MS-entries after alignment | 4315 | 4315 | 1724 | 4315 | 1724 | 4301 | 4301 | 1723 |
| MS/MS-entries after alignment | 622,328 | 622,328 | 127,915 | 622,328 | 127,915 | 616,880 | 616,880 | 127,569 |
| Total import time (s) | 153 | 214 | 87 | 272 | 155 | 3511 | 149 | 90 |

\* test was performed on Dell Latitude 7490 with Intel(R) Core(TM) CPU i7-8650U @ 1.9 GHz and 16 GB of RAM running Windows 10 Professional, with a Micron 2200 NVMe 512GB SSD. LX 1.2.7 does not support the MS filter setting.

Lipid identification was performed using MFQL scripts (provided in the MFQL folder) covering PC/PC-O, SM and TAG in MS and LPC, PE/PE-p and CE based on MS/MS signature ions. The application of frequency filter should lead to improved consistency of the Masterscan at the cost of mostly low intensity MS signals. In this perspective, one has to expect a reduction in number of lipid identifications without losing abundant lipids. Thus, we compared lipid identification and the recorded intensities after the respective import (Figure 2, Table 2). For PC/PC-O, SM and TAG we identified 129 lipid species with LX 1.2.7 (without filter) and 106 with LX 1.2.8 (with filter, see details: 190930\_mzML\_compressed\_127\_vs\_128\_filter.xlsx). As expected, we found that in most cases IDs were filtered out for lower abundant signals. We found a very good correlation between both import settings (Figure 2B). In cases where reduced intensities were reported for LX 1.2.8 (with filter), one could recognize that there were overlapping signals with low frequency that subsequently were filtered out (Figure 2C, D).

For lipid identification based on MS/MS quantifier ions, we found for LPC, PE/PE-p and CE 60 species for LX 1.2.7 (without filter) and 44 for LX 1.2.8 (with filter). For all lipids identified in both versions, we found very good correlations for the reported intensities. However, lipid species which were clearly still detectable by their signature ions in MS/MS were in a number of cases not reported because the required precursor signal in the MS survey scan was filtered out (Figure 3).

**Table 2) Lipid identification for mzML import between LX 1.2.7 and 1.2.8 and filter usage**

|  |  |  |
| --- | --- | --- |
| **LX Version** | **1.2.7** | **1.2.8 - Filtered** |
| **Lipid IDs - MS 1** | **129** | **106** |
| SM | 17 | 15 |
| PC / PC-O | 51 | 43 |
| TAG | 61 | 48 |
| **Lipid IDs - MS/MS** | **60** | **44** |
| CE | 18 | 15 |
| PEp | 19 | 13 |
| PE | 13 | 9 |
| LPC | 10 | 7 |
| **Lipid Ids** | **189** | **150** |



Figure 2) Correlation between LipidXplorer 1.2.7 and 1.2.8 lipid identifications.

A) Identified PC/PC-O species sorted according to abundance from top to bottom. Heat map representing ten technical replicates for the same extract and its pairwise change in intensity (log(2)). B) Correlation for PC/PC-O species intensities identified in both LX versions. C,D) Examples for overlapping peaks with overall low frequency of reoccurrence in MS survey scans (indicated in blue %). Indicated bin sizes were approximated as two times FWHM at the *m/z* according to LX import settings. Positioning of bins is based on spectral definition while starting points in LX starts from lowest *m/z* of centroid masses for recognized peak (see for details [4]).

All data, settings and MFQL can be found in folder *LipidXplorer-1.2.8-benchmark-data*.

For lipid identification and quantitation based in MS/MS signals, either relaxed filter settings for MS should be chosen or the mass accuracy for precursor matching could be relaxed. These settings might have to be customized for different studies and depend on the structure of the acquisition files.



Figure 3) Examples for LipidXplorer lipid identifications based on MS/MS.

A) Averaged MS spectrum section for the precursor *m/z* of LPC 16:1. In LX 1.2.7, a separate bin is retained with the *m/z* of 494.3240, which is filtered out in LX 1.2.8. B) MS spectrum section for PE p16:0/18:2, which is filtered out by LX 1.2.8 according to the frequency MS filter set to =0.6. Frequency of MS signal is indicated as blue %. All data, settings and MFQL can be found in folder *LipidXplorer-1.2.8-benchmark-data*. Positioning of bins is based on spectral definition while starting points in LX start from lowest *m/z* of centroid masses for the recognized peak (see for details [4]).

**Detailed Changelog**

*New Functionality*

1. Added frequency MS-Filter and MS/MS-Filter

The MS-Filter and MS/MS-Filter each take a value between 0 and 1. It represents the fraction of scans in which a certain peak has to be present.

The filter is applied for both MS and MS/MS according to the given import setting for each scan type.

For each precursor ion, all associated MS/MS scans are collected according to the selection window. Counting of MS/MS scans and association of precursor is done automatically.

Accordingly, the number of MS/MS scan can vary for each precursor ion in the shotgun acquisition, as might occur when using a DDA strategy.

1. Improved mzML file loading speed from O(n^2) to O(n).

This improves loading speed of mzML files significantly and brings them on par with mzXML files considering import of data.

1. Added application and taskbar icons.
2. (Developers) Anaconda package management.

LX now comes with a dependency definition for an Anaconda environment for development under Windows. Please check the [GitLab README.md](https://gitlab.isas.de/lifs-public/lipidxplorer/tree/release-1.2.8) file for further information.

1. LX executable distributions (convenience binaries).

LX is now distributed as a Windows and Linux archive containing executables that come with all dependencies included. These can be downloaded from [here](https://lifs.isas.de/lipidxplorer.html).

*Bug Fixes*

- Updated Wiki locations to https://lifs.isas.de/wiki

- Fixed MFQL editor dialog closing error on Windows 10 and Python 2.7.

- Fixed application not exiting properly due to non-terminated threads.

- Improved file handle closing by switching to auto closing behaviour with 'with'.

- Fixed output file creation for user accounts with '.' characters in path.

- Fixed UI label truncations and spacing issues on Windows 10.

- Updated wxPython dependency from 3.0 to 4.0.4.

*Removed Functionality[[1]](#footnote-1)*

- Removed sample grouping

- Removed DTA/CSV + PIS support for input spectra

*Deprecated Functionality*

- We plan to remove the mzXML format in one of the next LipidXplorer releases. mzXML is not supported anymore by the mass spectrometry community and deprecated. If you need this feature, please contact us for help with migration.

**Acknowledgements:**

The current maintainers, developers and contributors for LX would like to thank Ronny Herzog – initial author of the software for his continued support.

**References:**

1. Schuhmann, K., et al., *Intensity-Independent Noise Filtering in FT MS and FT MS/MS Spectra for Shotgun Lipidomics.* Anal Chem, 2017. **89**(13): p. 7046-7052.

2. Eggers, L.F. and D. Schwudke, *Shotgun Lipidomics Approach for Clinical Samples.* Methods Mol Biol, 2018. **1730**: p. 163-174.

3. Herzog, R., et al., *LipidXplorer: a software for consensual cross-platform lipidomics.* PLoS One, 2012. **7**(1): p. e29851.

4. Herzog, R., et al., *A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language.* Genome Biol, 2011. **12**(1): p. R8.

5. Herzog, R., D. Schwudke, and A. Shevchenko, *LipidXplorer: Software for Quantitative Shotgun Lipidomics Compatible with Multiple Mass Spectrometry Platforms.* Curr Protoc Bioinformatics, 2013. **43**: p. 14 12 1-30.

6. Chambers, M.C., et al., *A cross-platform toolkit for mass spectrometry and proteomics.* Nat Biotechnol, 2012. **30**(10): p. 918-20.

1. If you rely on this functionality, please contact us at lifs-support@isas.de [↑](#footnote-ref-1)