Automated Lipid Profiling of Malaria Samples Using Orbitrap Velos Pro Mass Spectrometer with LipidSearch and SimLipid® Software

PREMIER Biosoft
Accelerating Research in Life Sciences

Ningombam Sanjib Meitei[1,2], Arun Apte[1], Fatima Rahlouni[3]; David J. Sullivan[4]; Vladimir Shulaev[3]
[1]PREMIER Biosoft, Palo Alto, CA; [2]PREMIER Biosoft, Indore, M.P.; [3]University of North Texas, Denton, Texas; [4]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD
Corresponding author e-mail: sanjib@premierbiosoft.com

1

INTRODUCTION

Liquid chromatography—Mass spectrometry (LC-MS) provides one of the most popular platforms for lipidomic analysis. Comparative studies of the complex lipid mixtures found in cells and tissues could potentially reveal lipid biomarkers. We have developed a lipidomics analysis method – LC coupled with high resolution Orbitrap Velos Pro MS instrument methods; lipid identification using database search of MS/MS data from both the positive, and negative modes; and multivariate statistical analysis – to identify the significant lipid species that are responsible for classifying multiple biological groups of malaria samples. Commercially available software tools namely, SimLipid, and LipidSearch were used for lipid identification using MS/MS data, and SIMCA-P software tools were used for downstream data analysis that includes multivariate statistical techniques.

METHODS

Lipids were extracted from *Plasmodium berghei wild type strains or hemolysin III* knockout ANKA (PbHlyIII KO) isolate using Folch method. Two independent experiments using different KO isolates were performed with 3 biological replicates per wild type or KO strain each resulting in total of 12 samples.

LC-MS: LC-MS analysis was carried out on Accela 1250 quaternary pump with Accela Open autosampler on-line coupled to an LTQ Orbitrap Velos Pro Hybrid MS (Thermo Fisher Scientific, USA). Plasmodium lipid extracts were separated on an Accucore C18 2.1x150 mm 2.6 μm column using 30 min gradient (Hu et al., 2008).

Data Analysis: Lipid profiling: SimLipid v. 6.00 (PREMIER Biosoft, USA), and LipidSearch v. 4.1.30 (Thermo Scientific, USA) software.
Statistical analysis: SIMCA-P software.

OVERVIEW OF THE DATA ANALYSIS

The lipidomics analysis constitutes the following steps: (1) raw data processing, (2) lipid identification using MS/MS data, and (3) statistical analysis. The statistical analysis was performed on the following experimental design: group 1: KO1+KO2 vs group 2: WT1+WT2. Multivariate techniques namely, Principal Component Analysis (PCA), and Partial Least Square-Discriminant Analysis (PLS-DA) were also conducted to identify the lipid species that significantly contributed in classifying the groups.

The raw data were processed for peak detection, peak picking, and molecular feature finding with the two software tools – SimLipid (SL), and LipidSearch (LS)–using their respective parameter settings. The outputted peak lists were then subjected to their respective MS/MS database searches with a 5 ppm error tolerance for both the precursor, and product ions. Stringent filter criteria were applied in SL software to remove false positives in the results. For LS, all the identified lipids with Grades A-D were considered. Figure 1 (A) shows typical graphical user interfaces of SL software with optimized parameter settings for peak detection from LC-MS data generated by LTQ Orbitrap Pro Velos system. The MS/MS database search parameters shown in Figure 1(B) were used for lipid identification by SL software. Figure 2 shows some of the parameters of the LS software.

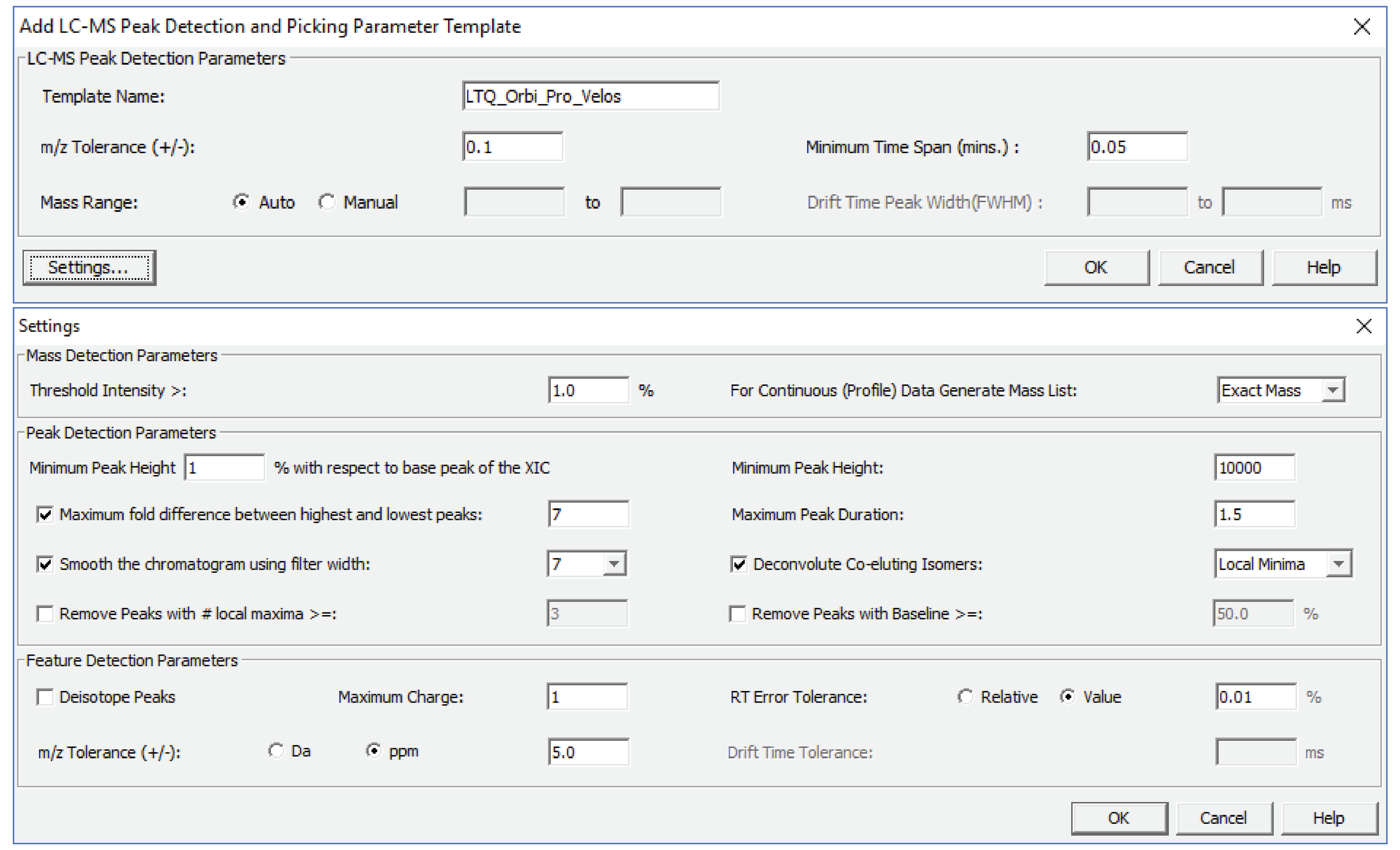

Add LC-MS Peak Detection and Picking Parameter Template
LC-MS Peak Detection Parameters
Template Name:
LTQ_Orbi_Pro_Velos
m/z Tolerance (+/-):
0.1
Minimum Time Span (mins.) :
0.05
Mass Range:
Auto
Manual
to
Drift Time Peak Width(FWHM) :
ms
Settings...
OK
Cancel
Help
Settings
Mass Detection Parameters
Threshold Intensity >:
1.0
%
For Continuous (Profile) Data Generate Mass List:
Exact Mass
Peak Detection Parameters
Minimum Peak Height
% with respect to base peak of the XIC
Minimum Peak Height:
10000
Maximum fold difference between highest and lowest peaks:
Maximum Peak Duration:
1.5
Smooth the chromatogram using filter width:
Deconvolute Co-eluting Isomers:
Local Minima
Remove Peaks with # local maxima >=:
Remove Peaks with Baseline >=:
50.0
Feature Detection Parameters
Deisotope Peaks
Maximum Charge:
RT Error Tolerance:
Relative
Value
0.01
m/z Tolerance (+/-):
Da
ppm
5.0
Drift Time Tolerance:

Figure 1(A): Typical GUIs of SimLipid software showing the LC-MS data processing parameters for LTQ Orbitrap Velos Pro Hybrid MS.

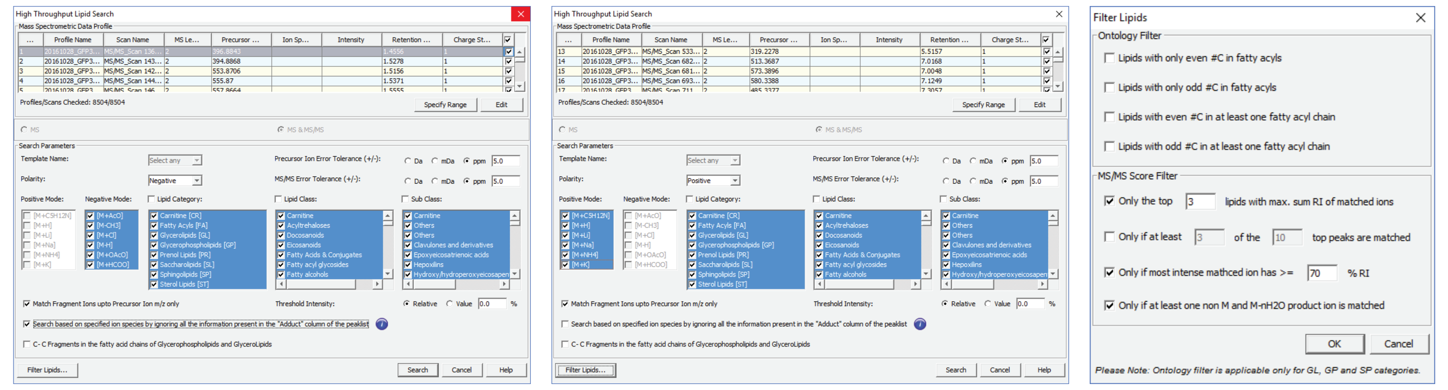

High Throughput Lipid Search
Filter Lipids
Ontology Filter
Lipids with only even #C in fatty acyls
Lipids with only odd #C in fatty acyls
Lipids with even #C in at least one fatty acyl chain
Lipids with odd #C in at least one fatty acyl chain
MS/MS Score Filter
Only if at least one non M and M-nH2O product ion is matched
Please Note: Ontology filter is applicable only for GL, GP and SP categories.

Figure 1(B): Typical GUIs of SimLipid software showing the MS/MS database search parameters for structural identification of lipids using product ions data.

2

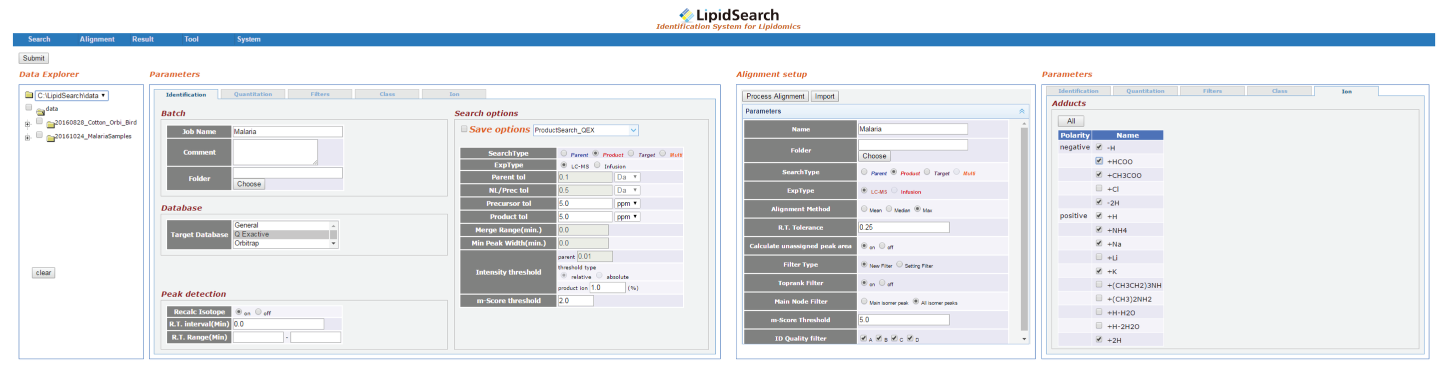

LipidSearch
Identification System for Lipidomics
Data Explorer
Parameters
Batch
Database
Peak detection
Search options
Alignment setup
Adducts

Figure 2: Typical GUIs of LipidSearch software showing LC-MS peak detection, and search parameters.

RESULTS AND DISCUSSION

The following numbers of unique lipids were identified by SL/LS database searches from the samples:

Software →	SL	LS	SL	LS	SL	LS	SL	LS	SL	LS	SL	LS	SL	LS	SL	LS
Mode →	-ve	-ve	+ve	+ve	-ve	-ve	+ve	+ve	-ve	-ve	+ve	+ve	-ve	-ve	+ve	+ve
Main Class ↓	KO1	KO1	KO1	KO1	KO2	KO2	KO2	KO2	WT1	WT1	WT1	WT1	WT2	WT2	WT2	WT2
PC	87	80	128	91	87	80	138	91	80	73	125	29	65	63	129	92
PE	77	94	46	50	73	97	49	46	75	88	49	52	71	88	46	41
PS	38	39	8	14	39	43	7	11	35	39	9	12	37	34	7	14
Phosphosphingolipids	22	21	26	38	21	21	25	34	20	20	25	4	19	21	25	31
TAG	0	0	39	57	0	0	32	53	0	0	29	38	0	0	42	67
PG	42	26	1	0	39	24	1	0	43	26	1	0	25	20	1	0
Ceramides	3	0	5	18	4	0	8	14	2	0	7	8	1	0	5	22
DAG	1	0	10	8	0	0	15	7	1	0	16	10	0	0	13	7
PI	18	0	2	0	20	0	2	0	21	0	1	0	17	1	1	0
PA	7	5	7	1	4	5	3	0	5	5	5	0	5	4	3	0
Cardiolipin	5	2	1	0	9	7	0	0	3	3	0	0	5	3	1	0
Fatty esters	0	0	6	0	0	0	7	0	0	0	11	0	0	0	6	0
Oxidized GP	0	0	5	0	0	0	4	0	0	0	2	0	0	0	4	0
AcCa	0	0	0	3	0	0	0	3	0	0	0	4	0	0	0	4
So	0	0	0	2	0	0	0	4	0	0	0	2	0	0	0	2
Neutral glycosphingolipids	2	0	1	0	0	0	1	0	2	0	1	0	0	0	2	0
Fatty Acids & Conjugates	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	0
GlyceroPE	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
CerG1	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0
Eicosanoids	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
Fatty amides	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0
Sterols	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
Co	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Acidic glycosphingolipids	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Other Glycerolipids	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Secosteroids	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Steroids	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Total	306	268	288	283	299	278	293	266	289	255	283	159	247	234	289	281

Table 1: Results generated by SL, and LS software tools. PC shows the sum of number of lipid species belonging to both the Diacylglycerophosphocholines (PC), and Monoacylglycerophosphocholines (LPC); similarly, PE is LPE+PE.

#	Main Class	SL	LS	#	Main Class	SL	LS
1	PC	210	150	15	Sterols	2	0
2	PE	106	116	16	Fatty amides	2	0
3	TAG	73	104	17	Eicosanoids	2	0
4	PG	63	31	18	Steroids	1	0
5	PS	59	51	19	Secosteroids	1	0
6	Phosphosphingolipids	37	46	20	Other Glycerolipids	1	0
7	PI	32	1	21	Glycerophosphonoethanolamines	1	0
8	PA	25	8	22	Fatty Acids & Conjugates	1	1
9	DAG	23	11	23	Acidic glycosphingolipids	1	0
10	Cardiolipin	15	8	24	AcCa	0	4
11	Fatty esters	13	0	25	Sphingosine (So)	0	4
12	Ceramides	10	24	26	CerG1	0	2
13	Oxidized glycerophospholipids	6	0	27	Co	0	1
14	Neutral glycosphingolipids	5	0		Total	689	562

Table 2: Summary of the lipid species identified from both the positive and negative modes data by SL, and LS software tools.

Parent Ion Molecules				
Main Class	SL (-ve Data)	LS (-ve Data)	SL (+ve Data)	LS (+ve Data)
PC	$[M+HCOO]^-$, $[M-H]^-$, $[M+CH_3]^-$, $[M+Cl]^-$	$[M+HCOO]^-$	$[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$	$[M+H]^+$, $[M+NH_4]^+$
PE	$[M-H]^-$	$[M-H]^-$	$[M+H]^+$, $[M+Na]^+$	$[M+H]^+$, $[M+NH_4]^+$
PS	$[M-H]^-$, $[M+CH_3]^-$	$[M-H]^-$	$[M+H]^+$, $[M+Naa]^+$, $[M+K]^+$	$[M+H]^+$, $[M+NH_4]^+$
SM	$[M+HCOO]^-$	$[M+HCOO]^-$	$[M+H]^+$, $[M+Na]^+$	$[M+H]^+$
TAG	-		$[M+NH_4]^+$, $[M+H]^+$	$[M+NH_4]^+$
PG	$[M-H]^-$, $[M+OAcO]^-$	$[M-H]^-$	$[M+H]^+$	-
Cer	$[M-H]^-$	-	$[M+H]^+$	$[M+H]^+$
DAG	-	-	$[M+H]^+$, $[M+NH_4]^+$	$[M+H]^+$, $[M+NH_4]^+$
PI	$[M-H]^-$	-	$[M+H]^+$	-
PA	$[M-H]^-$	$[M-H]^-$	$[M+Piperidine]^+$	$[M+H]^+$
CL	$[M-H]^-$	$[M-H]^-$	$M+H]^+$	-

Table 3: Parent Ion Molecules of the lipids identified by SL, and LS software tools. – denotes not reported.

Table 3 shows the parent molecular ions of the lipid species belonging to the top 11 lipid classes from which maximum number of lipids were identified by SL, and LS from the +ve, and -ve mode data.

SL VS LS: OVERVIEW OF THE CHARACTERISTIC IONS

Example 1: PC(16:0/18:2)

PC(16:0/18:2) at around 19 minutes was identified by both the software tools using the reported characteristic ions (Table 4(A)). SimLipid identification is based on the nine characteristic ions (Figure 3(A)) while LS did it on based on four ions (Figure 3(B)). Description of the characteristic ions of the lipid is described in Table 4(B):

	$[M-HCOO]^-$ (SL)	$[M-HCOO]^-$ (LS)
Data ID	808	1485
Retention Time	18.9134	18.922757
Precursor m/z	802.5599	802.560266
Delta mass	0.5385 (in ppm)	-.000095 (in Da)
Best Score/Grade	0.2202 (very good)	A
Group (replicates)	WT1 (replicate 2)	WT1 (replicate 2)
Characteristic Ions	15:0 COO-(255.2327), 17:2 COO-(279.2326), M-18:2-H2O(462.2989), M-18:2(480.3089), M-16:0-H2O(486.2975), M-16:0(504.3083), M-C5H13N(671.4639), M-CH3(742.5374), M(802.558)	FA(16:0)-H(255.232681), FA(18:2)-H(279.232574), M-H3(742.537385), M+HCOO(802.557995)

Table 4(A): Table showing the characteristic ions based on which SL, and LS identify the lipid species using MS/MS data.

Fragment Type	Description	SL nomenclature	LS Nomenclature
A	Sn1 chain	15:0 COO -	FA(16:0)-H
B	Sn2 chain	17:0 COO -	FA(18:2)-H
A*	M- Sn1 chain + Adduct	M-16:0 + HCOO-	Not reported
A**	A* - H2O	M-16:0 -H2O + HCOO-	Not reported
B	M- Sn2 chain + Adduct	M-18:2+HCOO-	Not reported
B*	B* - H2O	M-18:2-H2O+HCOO-	Not reported
C	M- Choline	M-C5H13N	Not reported
D	M-CH3	M-CH3	M-CH3
E	M+Adduct	M	M+HCOO-

Table 4(B): Description of the characteristic ions of the lipid PC(16:0/18:2).

3

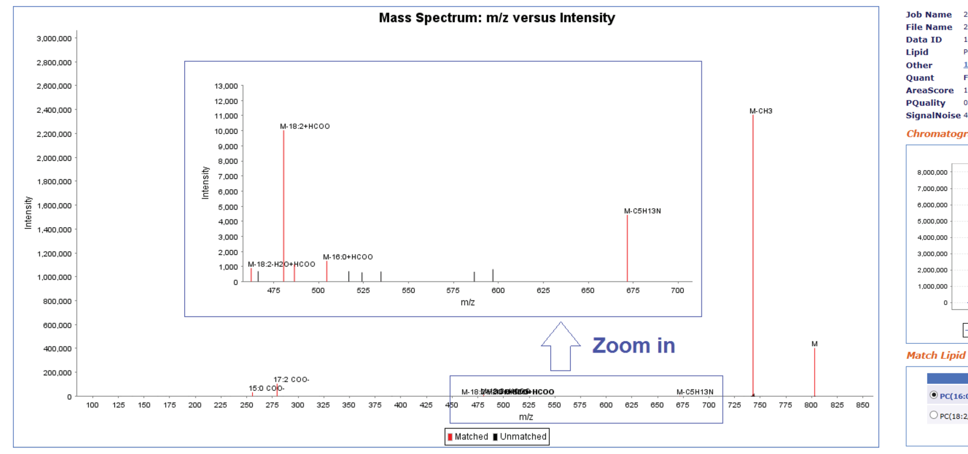

Mass Spectrum: m/z versus Intensity
Zoom in
Matched
Unmatched

Figure 3(A): Typical GUIs of SL software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species PC(16:0/18:2). The zoom in spectrum has been manually inserted for clear visualization of the fragment ions.

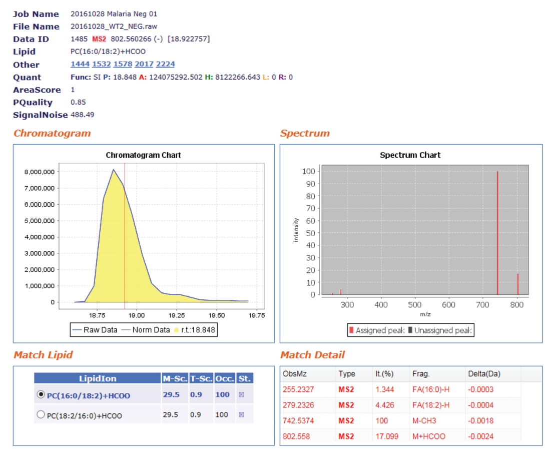

Chromatogram
Spectrum
Match Lipid
Match Detail

Figure 3(B): Typical GUIs of LS software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species PC(16:0/18:2).

Example 2: PE(16:0/18:2)

For this lipid, both SL, and LS software utilize the characteristic ions 313.274, 337.274, 575.5042. In addition to these common characteristic ions, SimLipid software make use of the characteristic ions 124.0159 (observed in the reported standard spectrum: http://www.hmdb.ca/spectra/ms_ms/29673), and the characteristic ions corresponding to the fatty acids' acylium ions 239.2372, and 263.2372 (Figure 4(A)). These main characteristic ions were not annotated by LS software (Figure 4(B)).

The ions a-j reported by LS software could be noise signals.

	$[M+H]^+$ (SL)	$[M+H]^+$ (LS)
Data ID	364	1797
Retention Time:	19.26	19.259955
Precursor m/z	716.5235	716.523855
Delta mass	2.1007 (in ppm)	-0.001372 (in Da)
Best Score/Grade	0.1586 (good)	A
Group (replicates)	WT2 (replicate 2)	WT2 (replicate 2)
Characteristic Ions	i. HG-H2O (124.0159), ii. 15:0C=O+(239.2372), iii. 17:2C=O+(263.2372), iv. 16:0 PE (313.274), v. 18:2 PE (337.274), vi. M- NL+H (575.5042)	a. C8H11(107.085472), b. C9H13(121.101234), c. C9H15(123.116913), d. C9H17(125.132568), e. C10H15(135.116989), f. C10H17(137.132614), g. C11H17(149.132614), h. C11H19(151.14827), i. C12H17(161.132721), j. C13H19(175.148315), k. MG(16:0)-OH(313.273987), l. MG(18:2)-OH(337.274017), m. NL[PE](575.50415)

Table 5: Table showing the characteristic ions based on which SL, and LS identify the lipid species using MS/MS data.

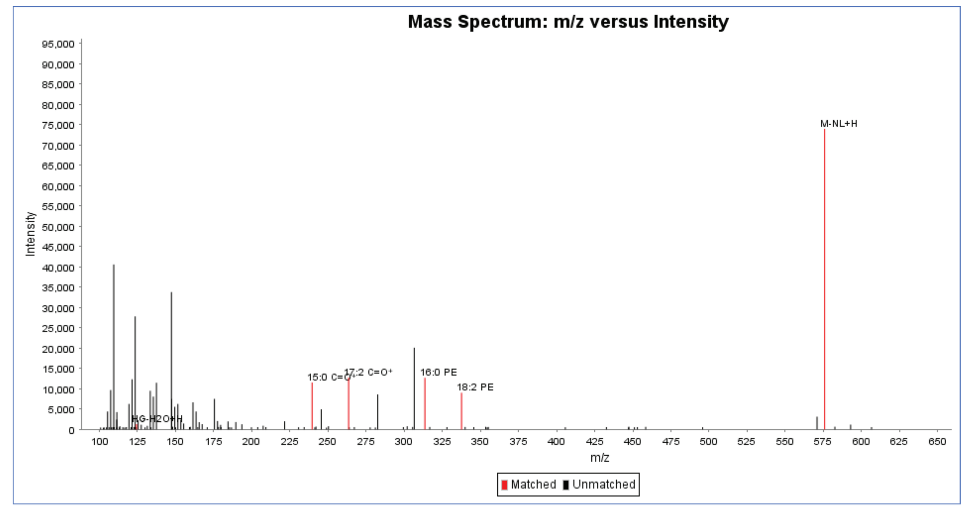

Mass Spectrum: m/z versus Intensity
M-NL+H
16:0 PE
18:2 PE
Matched
Unmatched

Figure 4(A): Typical GUIs of SL software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species PE(16:0/18:2). The MS/MS spectra from positive mode data feature a noisy region (m/z values < 200).

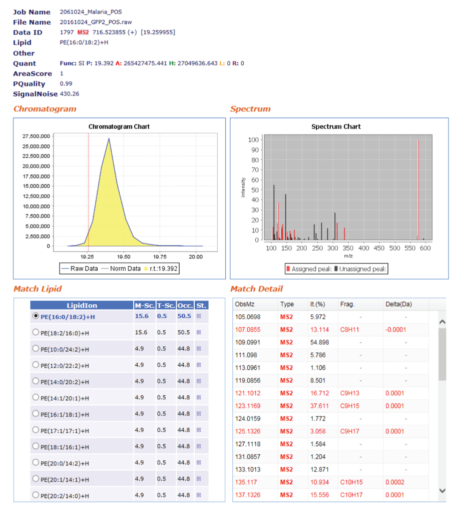

Chromatogram
Spectrum
Match Lipid
Match Detail

Figure 4(B): Typical GUIs of LS software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species PE(16:0/18:2).

ANNOTATION OF DIFFERENT LIPID SPECIES FOR LC-DETECTED COMPOUND BY SL, AND LS

We have observed that some of the LC-detected peaks (reported by both SL, and LS in their respective peaklists) have been assigned different lipid species by the software tools.

CARDIOLIPIN LIPIDS

Sample: WT2 Technical Replicate: 3LS software identified 8 lipid species from Cardiolipin class namely, (1) CL(18:2/16:0/16:0/20:4), (2) CL(18:2/16:0/18:2/20:4), (3) CL(18:2/20:3/16:0/20:4), (4) CL(20:4/16:0/16:0/20:3), (5) CL(20:4/16:0/16:0/20:4), (6) CL(20:4/16:0/18:0/20:4), (7) CL(20:4/16:0/20:3/20:4), and (8) CL(22:6/16:0/16:0/20:4). However, except for the lipid species (2) CL(18:2/16:0/18:2/20:4) identified in the sample KO2 (-ve), LS software annotate all the remaining species with grades either C or D. For example, the identification of isobaric lipids CL (72:4) by SL, and LS software tools.

	SimLipid	LipidSearch
Data ID	2544	1927
Retention Time:	23.55	23.554
Precursor m/z	1447.9648	1447.964188
Lipid (Short Name)	CL (72:8)	CL (72:8)
Lipid (Common Name)	CL(18:2/18:2/18:2/18:2)	CL(20:4/16:0/16:0/20:4)
Delta mass	0.0745 (in ppm)	-0.00076653 (in Da)
Score/Grade	0.1821 (good)	C
Group (Replicate)	WT2 (Replicate 3)	WT2 (Replicate 3)

LS software miss assigned the compound with CL(20:4/16:0/16:0/20:4) with the following annotations:
PA(20:4/16:0)-H(695.463989):MS2, PA(20:4/16:0)-H [isotope](696.46759):MS2, PA(16:0/20:4)-H(695.463989):MS2, PA(16:0/20:4)-H [isotope](696.46759):MS2, M-H(1447.963204):MS2.

SL assigned this compound with the lipid species CL(18:2/18:2/18:2/18:2) with the following annotations (Also shown in Figure 5):
M-C42H74O9P(695.464), M-C39H68O7P-H2O(751.4904), HG-H2O(831.4563), M-18:2-H2O(1167.7212), M(1447.9606)

LIPIDS NOT IDENTIFIED BY LS SOFTWARE

SL identified 32 lipid species belonging to the lipid main class phosphoinositol(PI), 29 out of the identified lipid species from the negative data. However, no PI lipid was reported by LS software. From the positive mode data, LS reported 1 PI lipid species. Figure 6 shows the annotated MS/MS spectrum of the lipid species PI(18:0/22:6) by SL software.

4

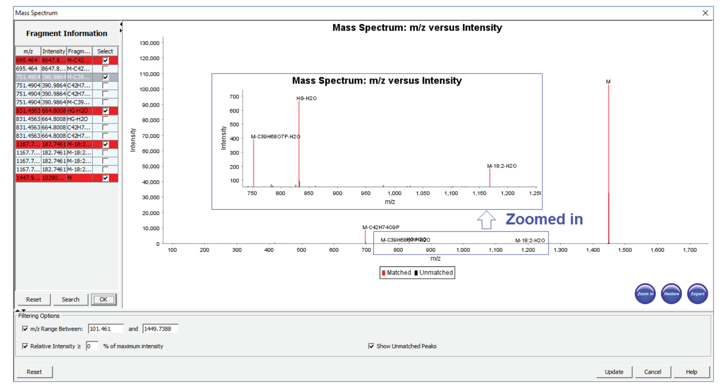

Mass Spectrum
Fragment Information
Mass Spectrum: m/z versus Intensity
Zoomed in
Matched
Unmatched

Figure 5: Typical GUIs of SL software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species CL(18:2/18:2/18:2/18:2). The zoom in spectrum has been manually inserted for clear visualization of the fragment ions.

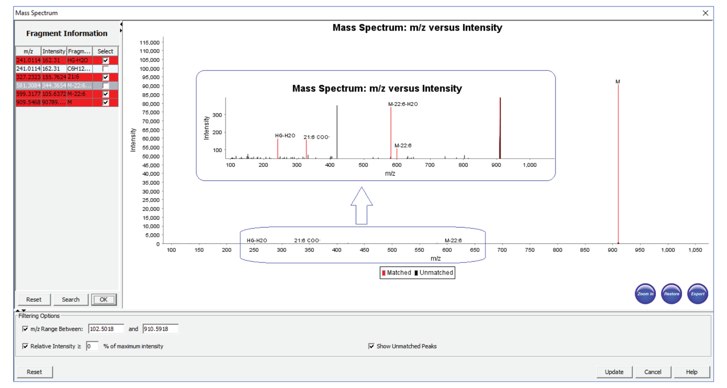

Mass Spectrum
Fragment Information
Mass Spectrum: m/z versus Intensity
Matched
Unmatched

Figure 6: Typical GUIs of SL software showing the MS/MS spectrum annotated with fragment ions of the identified lipid species PI(18:0/22:6). The zoom in spectrum has been manually inserted for clear visualization of the fragment ions. LS software did not identify any PI lipids in negative mode data.

STATISTICAL ANALYSIS

Unique lipids identified from all the biological groups, and their technical replicates were exported into MS excel file along with their peak areas.

DATA NORMALIZATION

This is an important step to ensure that we get correct and reliable results from the statistical analyses that we may subject our data to further hypothesis testing. For a biological group, the exported peak area of every IDed lipid was normalized with respect to the sum of peak areas of all the identified IDed lipids from all the technical replicates of the biological group. This data normalization is done to fulfil the assumption that if a lipid is identified, its abundance is observed consistently throughout the technical replicates.

The normalized data was subjected into the following multivariate statistical techniques namely, Principal component analysis (PCA), Partial least squares Discriminant Analysis (PLS-DA), Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA). From every technique, we display the (a) Score plot: showing the classification of the biological groups while their corresponding (b) Correlation loadings plot: showing the lipid species responsible for classification of the biological groups.

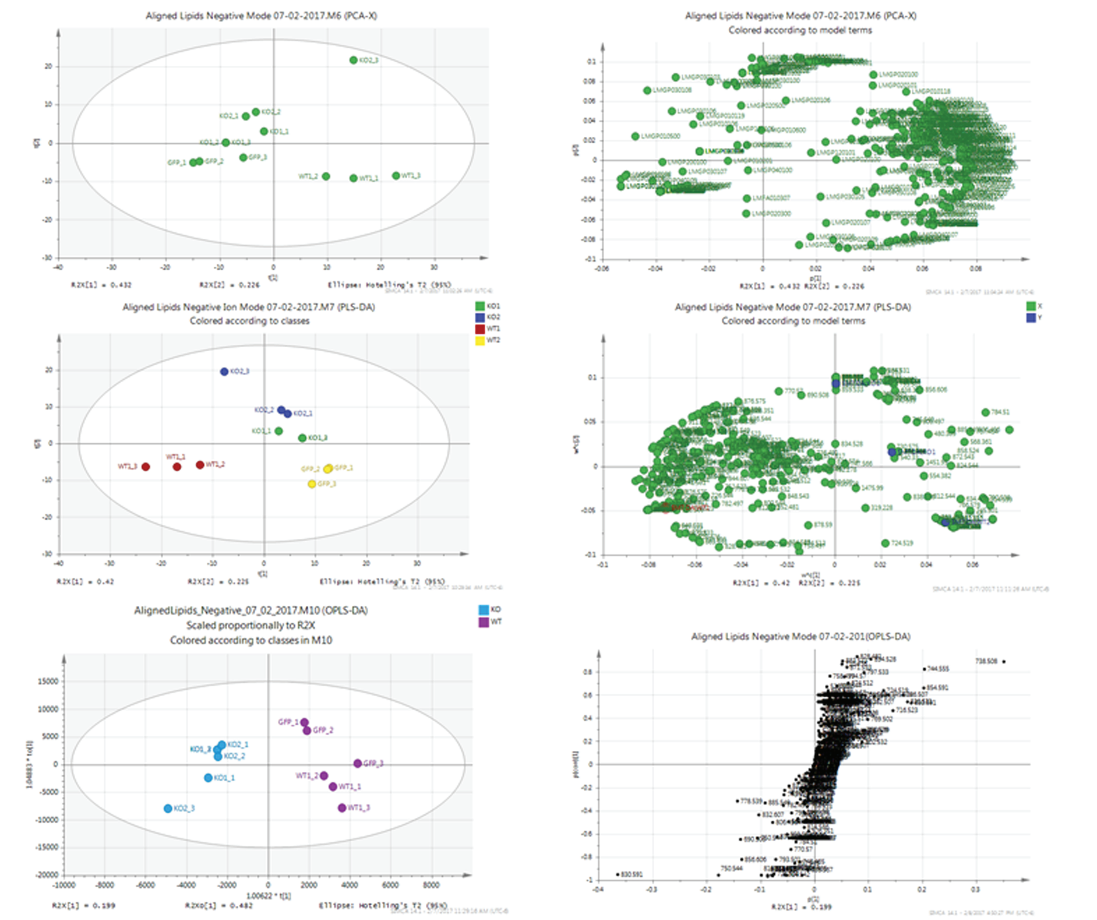

Aligned Lipids Negative Mode 07-02-2017.M6 (PCA-X)
Aligned Lipids Negative Mode 07-02-2017.M6 (PCA-X)
Colored according to model terms
Aligned Lipids Negative Ion Mode 07-02-2017.M7 (PLS-DA)
Colored according to classes
Aligned Lipids Negative Mode 07-02-2017.M7 (PLS-DA)
Colored according to model terms
AlignedLipids_Negative_07_02_2017.M10 (OPLS-DA)
Scaled proportionally to R2X
Colored according to classes in M10
Aligned Lipids Negative Mode 07-02-201(OPLS-DA)

Figure 7: Typical GUIs of the SIMCA-P software showing Scores plots, and their corresponding loadings plots from PCA, PLS-DA, and OPLS-DA observed in the score plot (only data from negative mode used).

CONCLUSION

1. We have developed a lipidomics analysis method –LC coupled with high resolution Thermo's Orbitrap Velos Pro MS instrument methods; lipid identification using database search of MS/MS data from both the positive, and negative modes.

2. SimLipid software identified 689 unique lipid species across the 12 samples while LipidSearch identified 562. The number of unique lipids identified by LS software further goes down when we removed those IDs with grades C, or D. It may be noted that, out of the 2024 lipids identified by LipidSearch software between the 12 samples, 590 lipids were identified with Grades C, and D.

3. The MS/MS spectra in positive mode feature a noisy region (peaks with m/z <200) which lead to false identification of fatty ester lipids by SimLipid software instead of TAG lipids. This could be controlled by removing fatty esters class in the SimLipid software database search filters.

4. The identified lipid species belonging to the main classes 16-27 (Table 2) may be false positives that were annotated due to matched of characteristic ions in the noise signals.

5. The method needs to be worked on to produce better MS/MS spectra for the positive mode.

REFERENCES

1. Hu, C. et al. (2008) *Proteome Res.* 7: 4982–4991.