

Lipid annotation with MS2Analyzer

Yan Ma

10/24/2013

Checklist before you start...

You need to have:

- 1.A computer with Java environment and Office(2003 or higher)
- 2.MS/MS spectra in MGF files
- 3.Latest version of MS2Analyzer jar file
- 4.Query text files

For additional validation with lipidblast, download the library and NIST search software in the following link(download the 87 MB full version):

<http://fiehnlab.ucdavis.edu/projects/LipidBlast>

Instruction outline

Step1: Using MS2Analyzer to search the spectra feature

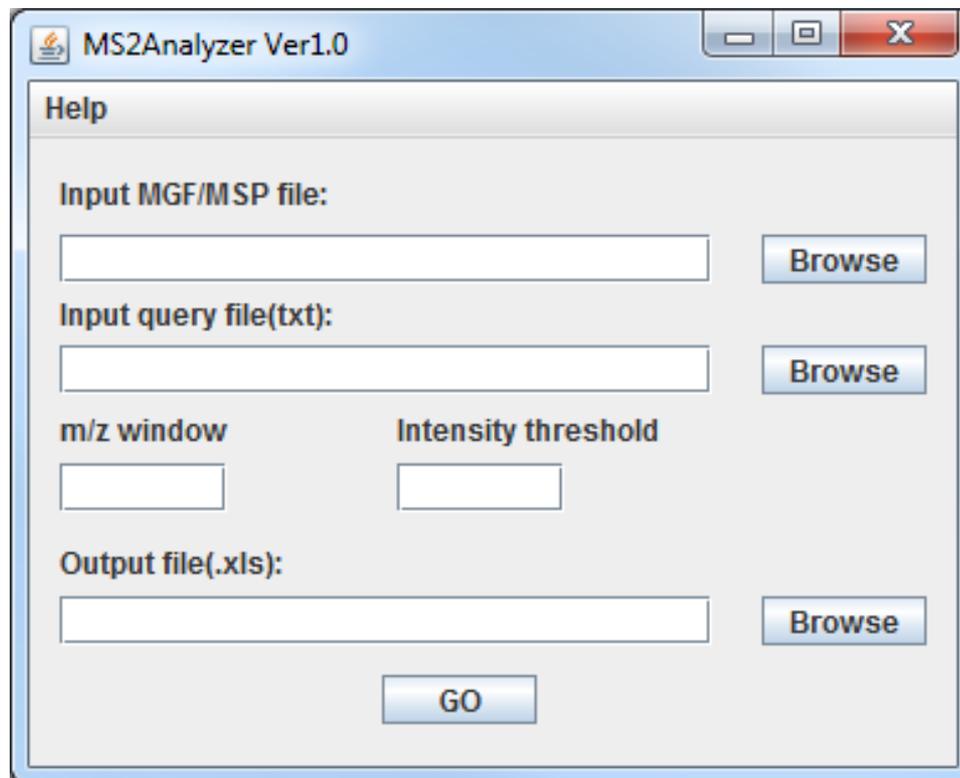
Step2: Using excel filters to annotate lipids

Step3: Confirmation by checking retention time
(optional)

Step 4:Using NIST search for library validation(optional)

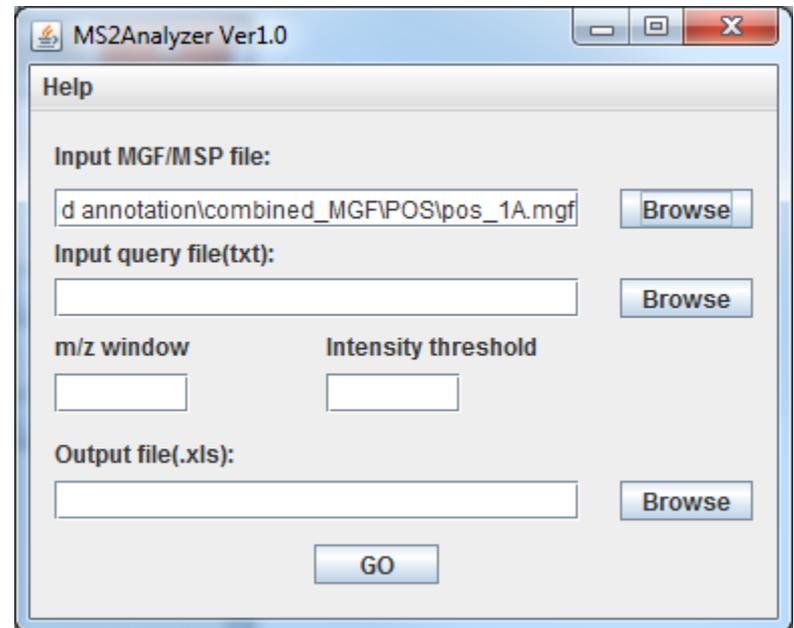
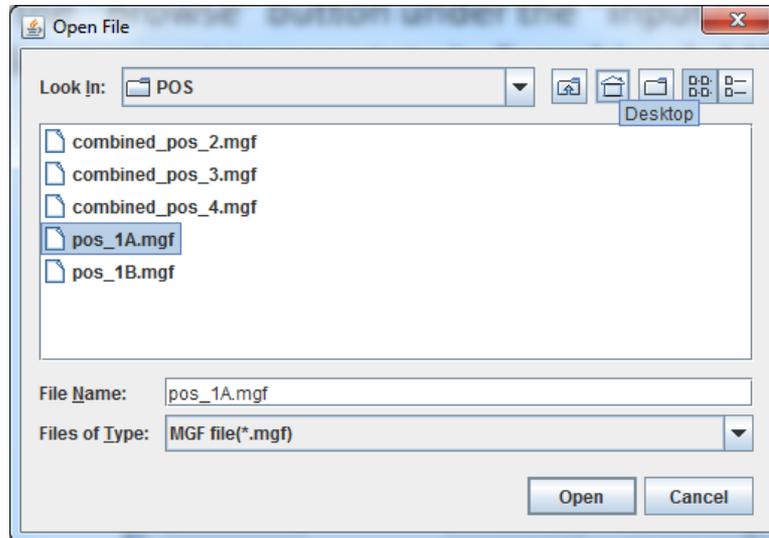
Step 1: Using MS2Analyzer to search the spectra feature

1. Double click on MS2Analyzer-ver1.jar to start the program



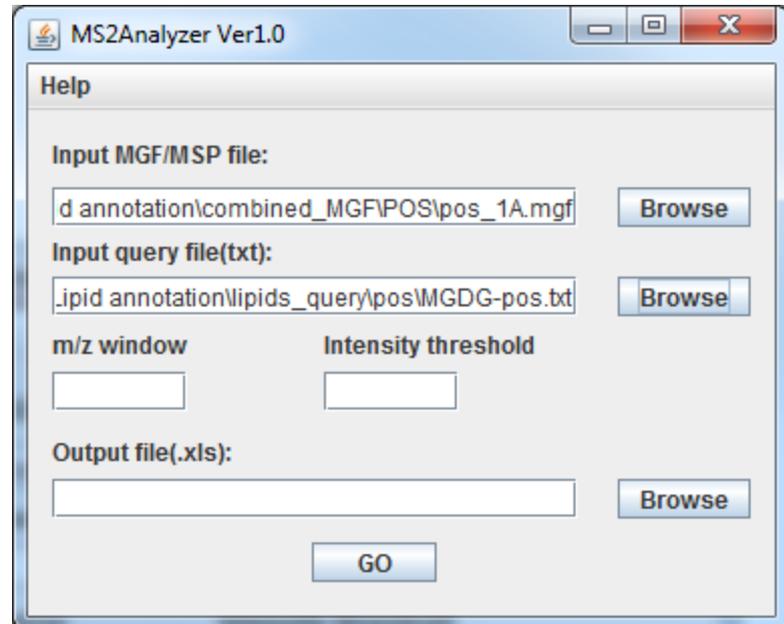
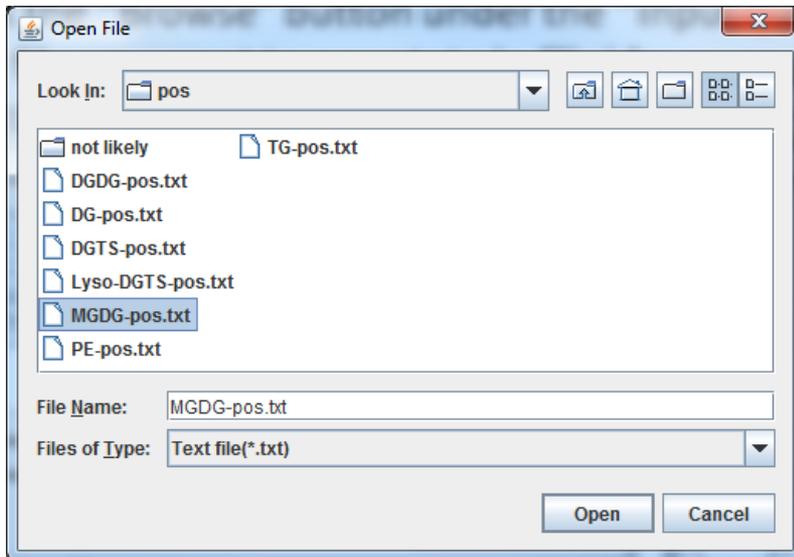
Step 1: Using MS2Analyzer to search the spectra feature

2. Click the “Browse” button under the “Input MGF/MSP file” and select the MGF file you want to annotate.



Step 1: Using MS2Analyzer to search the spectra feature

3. Click the “Browse” button under the “Input query file(txt)” and select the query file you want.



Step 1: Using MS2Analyzer to search the spectra feature

4.Type your desired parameters into “m/z window” and “intensity threshold”.

The first one depends on the mass accuracy of the instrument and the second one depends how noisy the spectra are. Intensity threshold are relative intensity to the base peak, from 0 to 1.0.

MS2Analyzer Ver1.0

Help

Input MGF/MSP file:
d annotation\combined_MGF\POS\pos_1A.mgf

Input query file(txt):
.lipid annotation\lipids_query\pos\MGDG-pos.txt

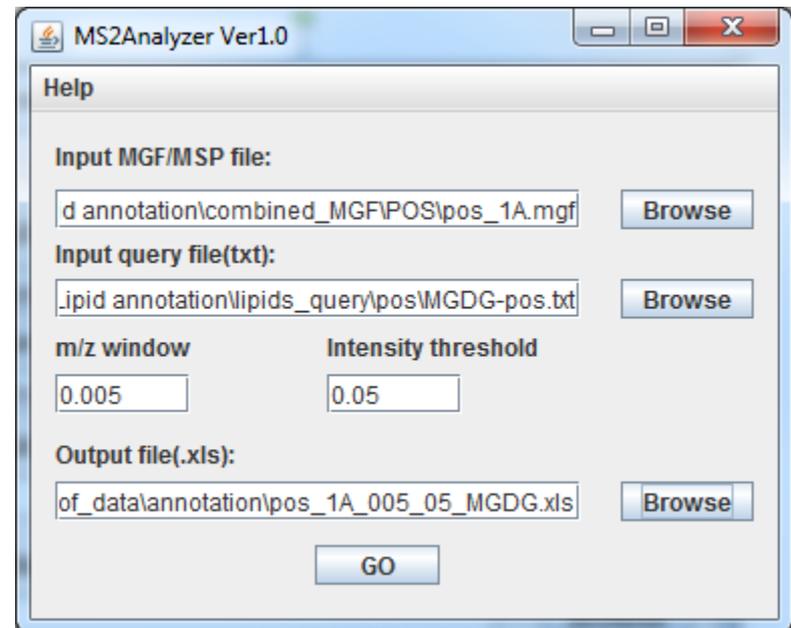
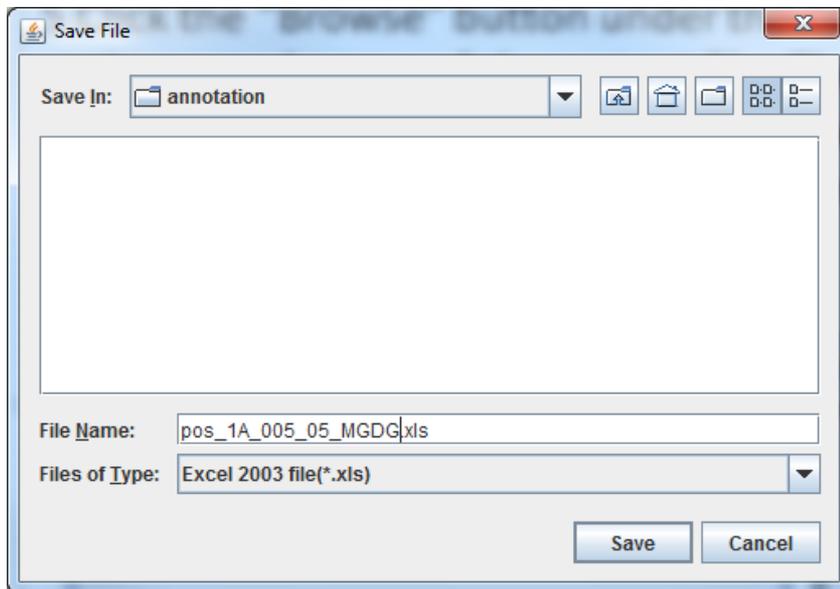
m/z window Intensity threshold
0.005 0.05

Output file(.xls):

Step 1: Using MS2Analyzer to search the spectra feature

5. Click the “Browse” button under the “Output files (.xls)” and choose the pathway and name of the output file.

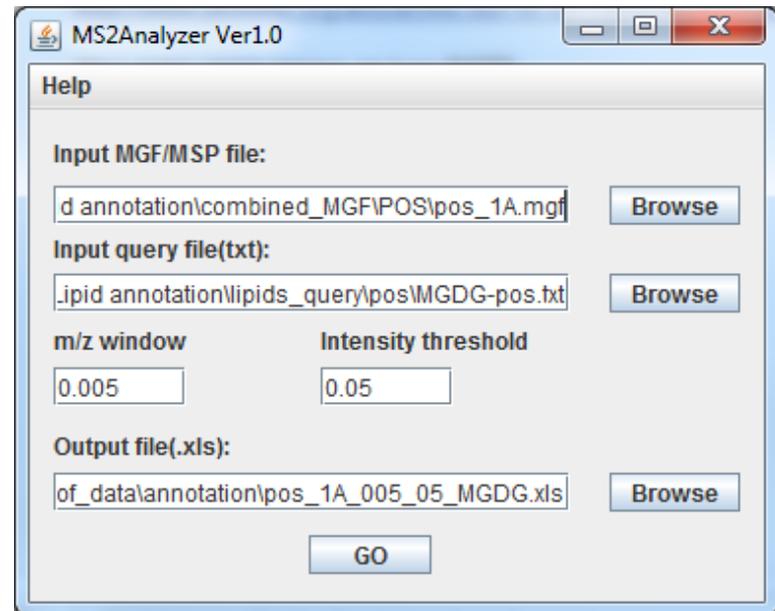
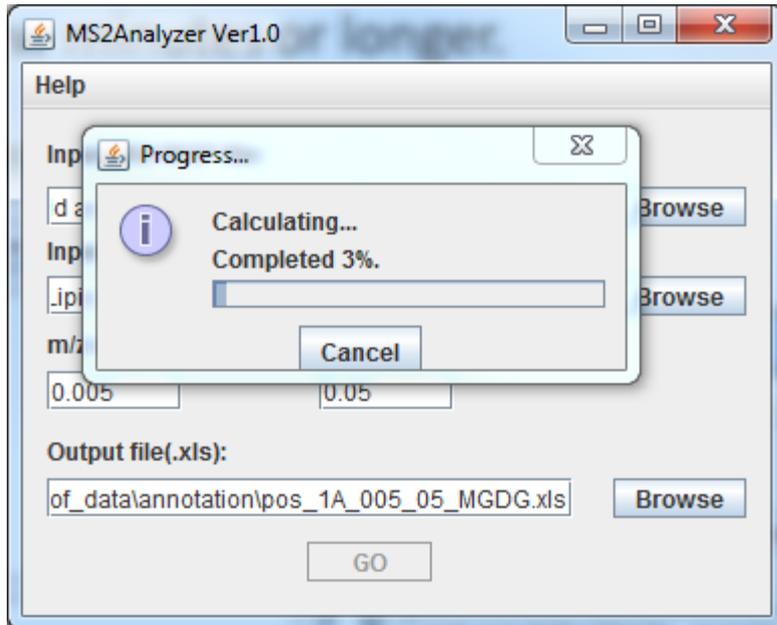
The default name of the output file is same as the name of the MGF files, but you may want to add more information about the lipid class and parameters. Here’s an example:



Step 1: Using MS2Analyzer to search the spectra feature

6. Click on “GO” button. The program will start searching mass spectra features in the MGF file and generate a Excel 2003 file.

Depending on the speed of the computer, it will take a few minutes or longer. A progress window will show up. When it's done, the progress window will disappear, and you are ready to move to the next Step!



Step 2: Using excel filters to annotate lipids

4. Click on each filter under the name of lipids (in red), and check if there's 1 value there. If yes, filter out the 0 and show only the 1 value.

The screenshot shows an Excel spreadsheet with a filter dropdown menu open over column M. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		MGDG	16:0+179	16:1+179	16:2+179	16:3+179	16:4+179	18:0+179	18:1+179	18:2+179	18:3+179	18:4+179	MGDG34 :7[M+NH 4]+	MGDG32 :0[M+NH 4]+	MGD :1[M+ 4]+
2	Title	179.0793	435.3195	433.3039	431.2882	429.2726	427.2569	463.3508	461.3352	459.3195	457.3039	455.2882	762.5155	748.5938	746.5
90	5h6kMSMSp.d, MS/MS of 279.1546631 1+ at 0.42085 mins	69	862	37	878	386	894	846	354	862	37	878	952	412	92
221	5h6kMSMSp.d, MS/MS of 504.3187866 0 at 0.859133333333333 mins	↑	0	0	0	0	0	0	0	0	0	0	0	0	0
1075	5h6kMSMSp.d, MS/MS of 762.5173035 1+ at 3.659283333333333 mins	↑	0	0	0	0	0	↑	0	0	0	0	0	0	0
1147	5h6kMSMSp.d, MS/MS of 764.5323079 0 at 3.821683333333333 mins	↑	0	0	0	0	↑	↑	0	0	0	0	0	0	0
1149	5h6kMSMSp.d, MS/MS of 765.5352173 1+ at 3.823516666666667 mins	↑	0	0	0	↑	↑	0	0	0	0	0	0	0	0
1150	5h6kMSMSp.d, MS/MS of 765.5352173 1+ at 3.824416666666667 mins	↑	0	0	0	0	↑	0	0	0	0	0	0	0	0
1197	5h6kMSMSp.d, MS/MS of 764.5323079 1+ at 3.942 mins	↑	0	0	0	0	↑	↑	0	0	0	0	0	0	0
1237	5h6kMSMSp.d, MS/MS of 764.5323079 1+ at 4.033666666666667 mins	↑	0	0	0	0	↑	↑	0	0	0	0	0	0	0
1277	5h6kMSMSp.d, MS/MS of 766.5473429 0 at 4.128183333333333 mins	↑	0	0	0	↑	↑	0	0	0	0	0	0	0	0
1387	5h6kMSMSp.d, MS/MS of 766.5473429 1+ at 4.38465 mins	↑	0	0	0	0	0	↑	0	0	0	0	0	0	0
1427	5h6kMSMSp.d, MS/MS of 768.5623474 1+ at 4.483683333333333 mins	↑	0	↑	↑	0	0	0	0	0	0	0	0	0	0

The filter dropdown menu is open over cell M1, showing the following options:

- Sort Smallest to Largest
- Sort Largest to Smallest
- Sort by Color
- Clear Filter From "762.5155952"
- Filter by Color
- Number Filters
- Search
- (Select All)
- 0
- 1

Buttons: OK, Cancel

Step 2: Using excel filters to annotate lipids

5. Check the 1 values in the acyl chain loss columns(16:0 to 18:4 in this case) and see if the sum of them matches the number in lipids name. If the sum matches, this is a correct hit!

Example 1: 16:4+18:3=34:7 which means this lipid is

MGDG 34:7(16:4/18:3) [M+NH₄]⁺

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		MGDG	16:0+179	16:1+179	16:2+179	16:3+179	16:4+179	18:0+179	18:1+179	18:2+179	18:3+179	18:4+179	MGDG34 :7[M+NH 4] ⁺	MGDG32 :0[M+NH 4] ⁺	MGDG32 :1[M+NH 4] ⁺
2	Title	179.0793	435.3195	433.3039	431.2882	429.2726	427.2569	463.3508	461.3352	459.3195	457.3039	455.2882	762.5155	748.5938	746.5781
1075	5h6kMSMSp.d, MS/MS of 762.5173035 1+ at 3.65928333333333 mins	69	862	37	878	386	894	846	354	862	37	878	952	412	92
4970			↑	0	0	0	0	↑	0	0	0	↑	0	1	0

Example 2: Sometimes there's only one acyl chain loss found, but maybe that's also a potential hit. In this case, the lipid name is written as

MGDG 34:7(16:4/18:3) [M+NH₄]⁺

While 16:4 is found by program, and 18:3 is back calculated: 34:7-16:4=18:3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		MGDG	16:0+179	16:1+179	16:2+179	16:3+179	16:4+179	18:0+179	18:1+179	18:2+179	18:3+179	18:4+179	MGDG34 :7[M+NH 4] ⁺	MGDG32 :0[M+NH 4] ⁺
2	Title	179.0793	435.3195	433.3039	431.2882	429.2726	427.2569	463.3508	461.3352	459.3195	457.3039	455.2882	762.5155	748.5938
1075	5h6kMSMSp.d, MS/MS of 762.5173035 1+ at 3.65928333333333 mins	69	862	37	878	386	894	846	354	862	37	878	952	412
4970			↑	0	0	0	0	↑	0	0	0	0	0	1

Step 2: Using excel filters to annotate lipids

5. Example 3: Sometimes, there are multiple possibilities...

First spectra:

16:2+18:3=34:5 MGDG 34:5(16:2/18:3) [M+NH₄]⁺

Second spectra:

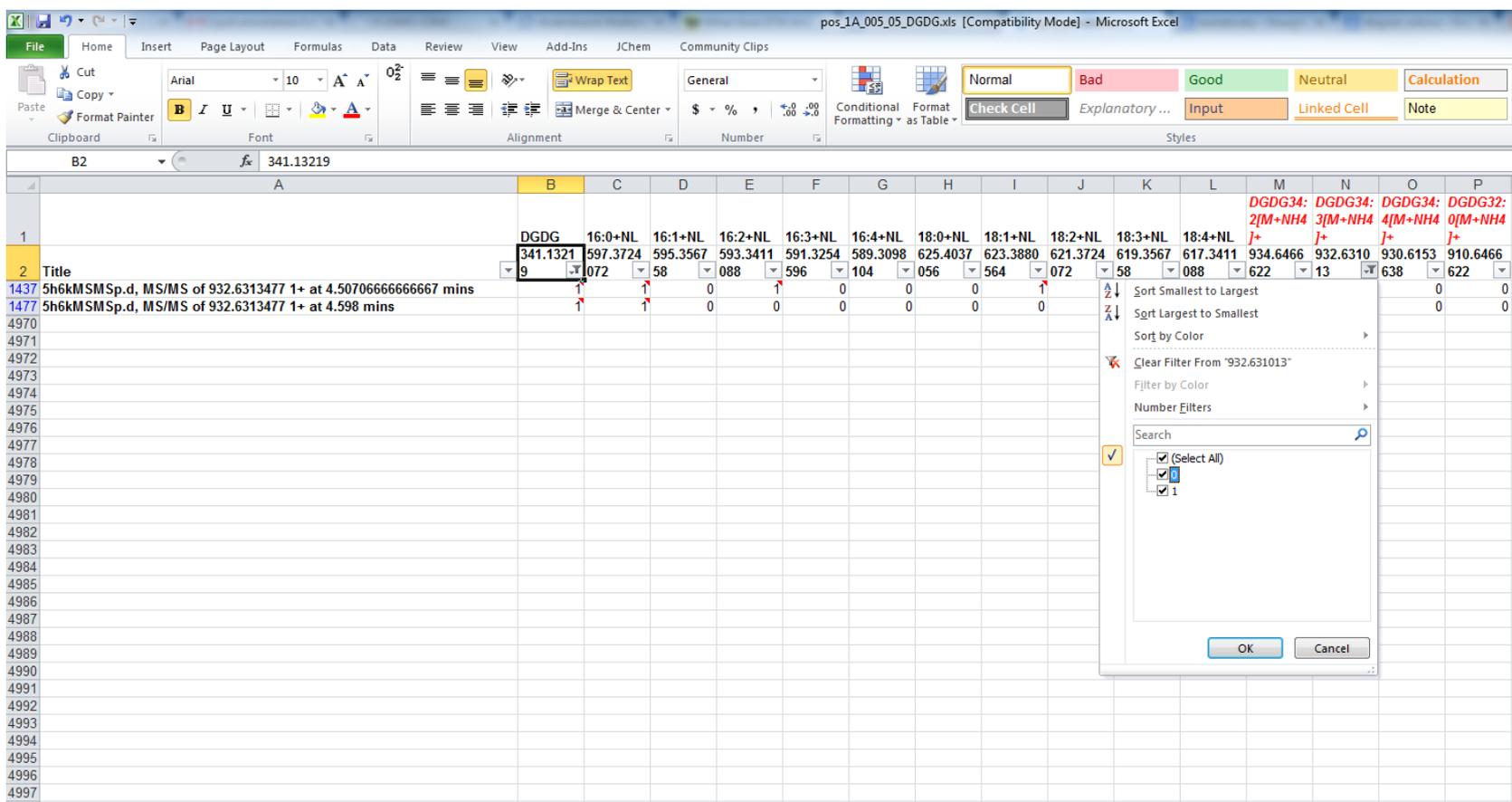
16:4+18:1=34:5 MGDG 34:5(16:4/18:1) [M+NH₄]⁺

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		MGDG	16:0+179	16:1+179	16:2+179	16:3+179	16:4+179	18:0+179	18:1+179	18:2+179	18:3+179	18:4+179	MGDG34 :5[M+NH 4] ⁺	MGDG34 :7[M+NH 4] ⁺
2	Title	179.0793	435.3195	433.3039	431.2882	429.2726	427.2569	463.3508	461.3352	459.3195	457.3039	455.2882	766.5468	762.5155
1277	5h6kMSMSp.d, MS/MS of 766.5473429 0 at 4.128183333333333 mins	69	862	37	878	386	894	846	354	862	37	878	936	952
1387	5h6kMSMSp.d, MS/MS of 766.5473429 1+ at 4.38465 mins	69	862	37	878	386	894	846	354	862	37	878	936	952
4970														
4971														

Both are correct and should be reported separately.

Step 2: Using excel filters to annotate lipids

5. (Continue) After you are done with one lipid precursor m/z , clear the filter by checking the 0 value under its filter again, and move to the next precursor m/z.



The screenshot shows the Microsoft Excel interface with a spreadsheet titled "pos_1A_005_05_DGDG.xls". The spreadsheet has columns A through P. Column A contains lipid precursor information, and columns B through P contain numerical values. A filter dropdown menu is open over column M, showing options to sort and filter data. The table contains lipid precursor information and associated m/z values.

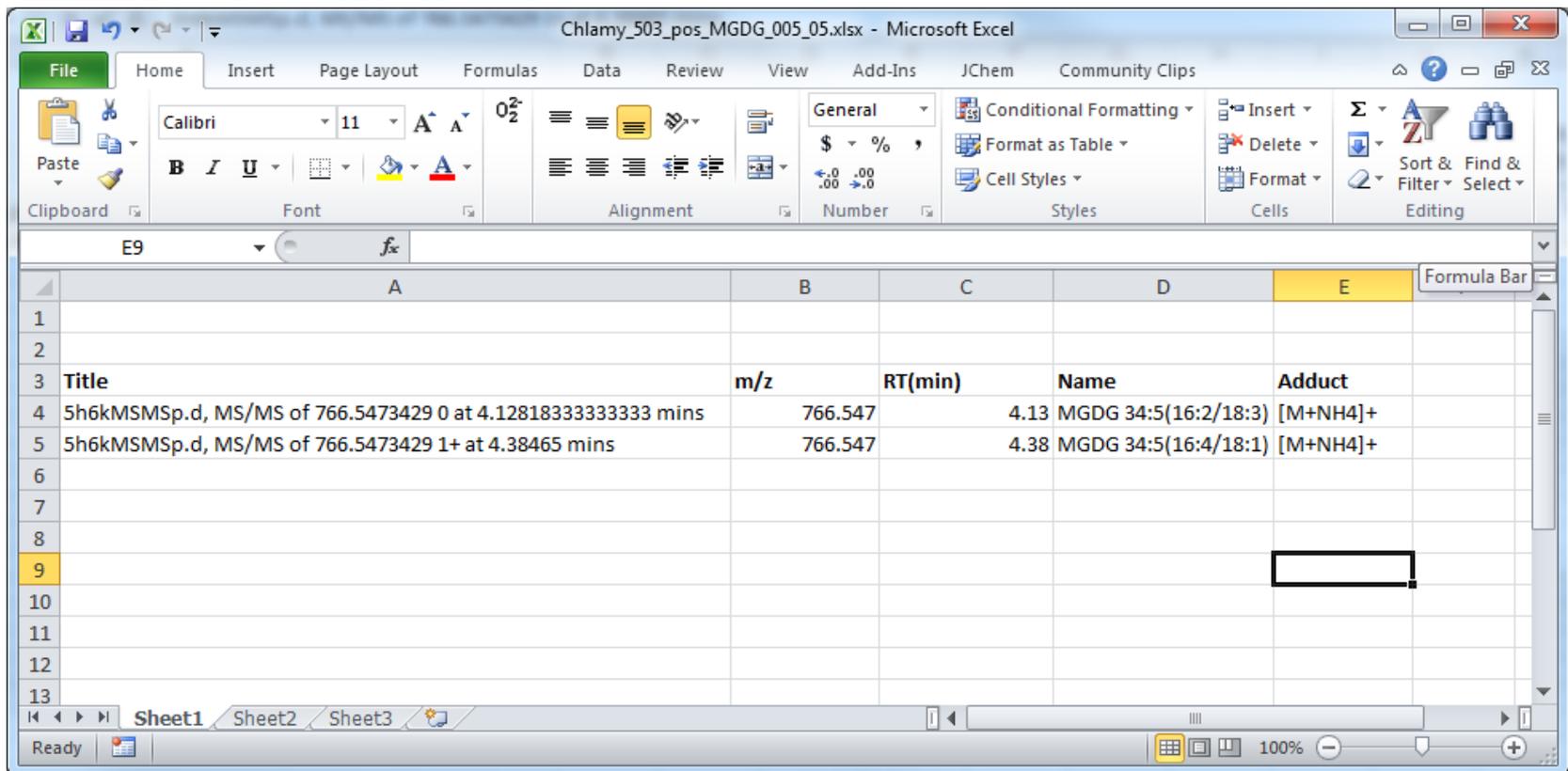
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		DGDG	16:0+NL	16:1+NL	16:2+NL	16:3+NL	16:4+NL	18:0+NL	18:1+NL	18:2+NL	18:3+NL	18:4+NL	DGDG34: 2[M+NH4]	DGDG34: 3[M+NH4]	DGDG34: 4[M+NH4]	DGDG32: 0[M+NH4]
2	Title	341.1321	597.3724	595.3567	593.3411	591.3254	589.3098	625.4037	623.3880	621.3724	619.3567	617.3411	934.6466	932.6310	930.6153	910.6466
1437	5h6kMSMSp.d, MS/MS of 932.6313477 1+ at 4.50706666666667 mins	9	072	58	088	596	104	056	564	072	58	088	622	13	638	622
1477	5h6kMSMSp.d, MS/MS of 932.6313477 1+ at 4.598 mins				0	0	0	0	0	0					0	0

Step 2: Using excel filters to annotate lipids

6. Write the result into a Excel file.

The following is an example of the report:

Report: Title, m/z(with 3 decimals, e.g. 766.547), RT(retention time, with 2 decimals, e.g. 4.13), Name of lipid, and Adduct



Chlamy_503_pos_MGDG_005_05.xlsx - Microsoft Excel

	A	B	C	D	E
1					
2					
3	Title	m/z	RT(min)	Name	Adduct
4	5h6kMSMSp.d, MS/MS of 766.5473429 0 at 4.12818333333333 mins	766.547	4.13	MGDG 34:5(16:2/18:3)	[M+NH4] ⁺
5	5h6kMSMSp.d, MS/MS of 766.5473429 1+ at 4.38465 mins	766.547	4.38	MGDG 34:5(16:4/18:1)	[M+NH4] ⁺
6					
7					
8					
9					
10					
11					
12					
13					

Step 2: Using excel filters to annotate lipids

TIPS:

MGDG, DGDG, DGTS, PE all have head groups and two acyl chains (same as the example).

Lyso-DGTS only has head group and one acyl chain, in this case, the acyl chain loss must be found in order to report.

DG and TG don't have head groups. Therefore,

DG: only look for the two acyl chain loss (both must be found)

TG: only look for the three acyl chain loss (at least two of them must be found)

Step3: Confirmation by checking retention time

If a RP column is used for the separation, among lipids in each lipid class, the retention times usually follow the following rule:

Retention times increase with the number of carbon, while decrease with the number of double bond.

This can be used as a further confirmation of the annotation.

Step3: Confirmation by checking retention time

For example

RT MGDG 34:0 < RT MGDG 36:0

RT DGDG 32:2 > RT DGDG 32:3

Quiz:

RT PE 34:1 ? RT PE 32:4

RT TG 50:4? RT DG 36:4

Step 4: Using NIST search for library validation

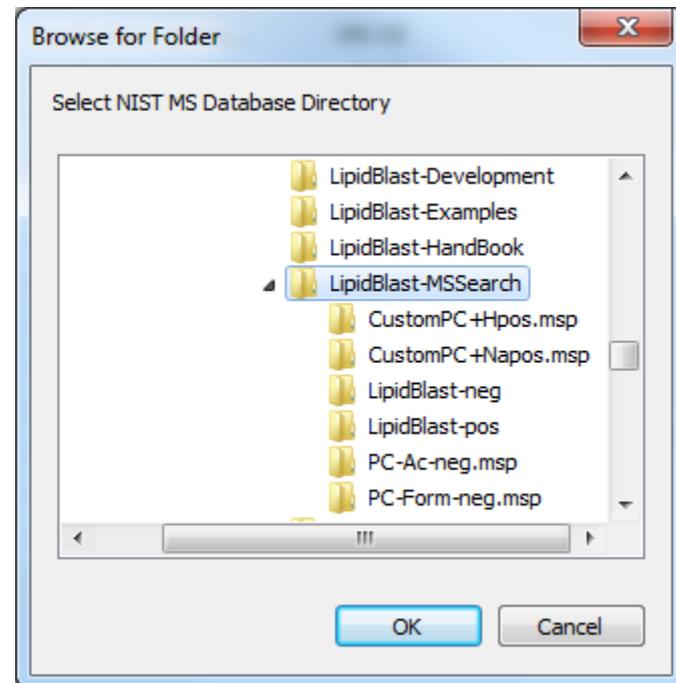
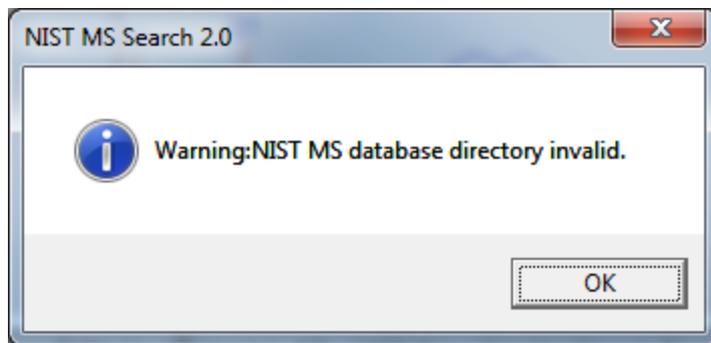
Now you have a list of potential lipid with its title and annotation, you can visualize the spectra in NIST search and search against available libraries to see if the annotation results match.

Here we use lipidblast library as an example.

Step 4: Using NIST search for library validation

1. Extract the LipidBlast-Full.zip file, and double click on the nistms.exe file in the LipidBlast-MSSearch folder.

When you open Nist search for the first time, it will give you a warning message. After clicking OK, a window pops up to ask you to select library directory. Here we just keep it highlighting LipidBlast-MSSearch and click OK, then the software will open.



Step 4: Using NIST search for library validation

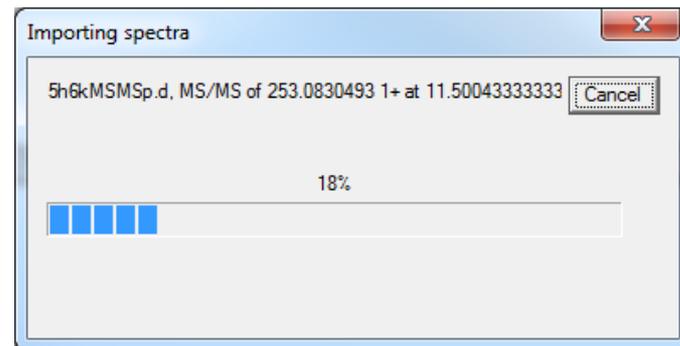
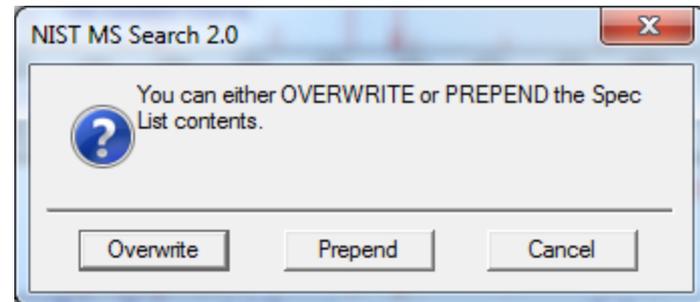
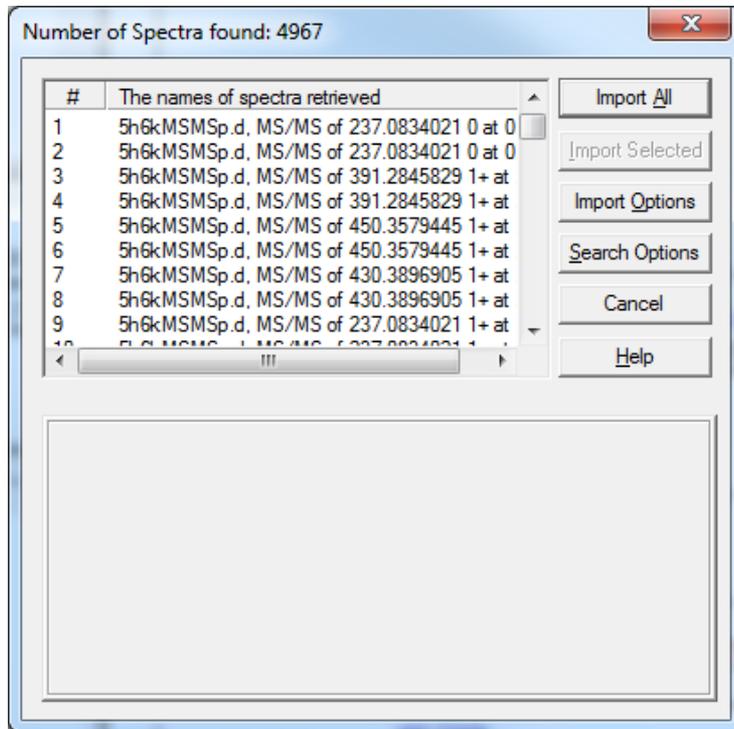
2. Go to the import button, and select the MGF file where you have your lipids in.

The screenshot shows the NIST MS Search 2.0 interface. The 'import' button is circled in red. A file selection dialog box is open, showing a list of MGF files in the 'POS' directory. The file 'pos_1A.mgf' is selected.

Name	Date modified	Type	Size
combined_pos_2.mgf	10/22/2013 3:15 PM	MGF File	131,312 KB
combined_pos_3.mgf	10/22/2013 3:27 PM	MGF File	87,084 KB
combined_pos_4.mgf	10/22/2013 3:28 PM	MGF File	98,017 KB
pos_1A.mgf	10/7/2013 12:36 AM	MGF File	93,153 KB
pos_1B.mgf	10/7/2013 12:35 AM	MGF File	96,332 KB

Step 4: Using NIST search for library validation

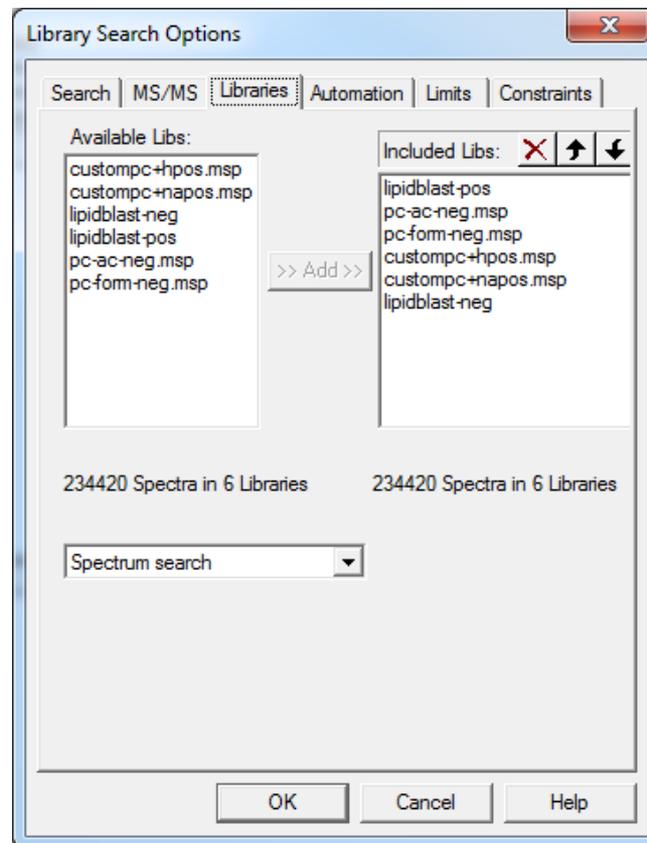
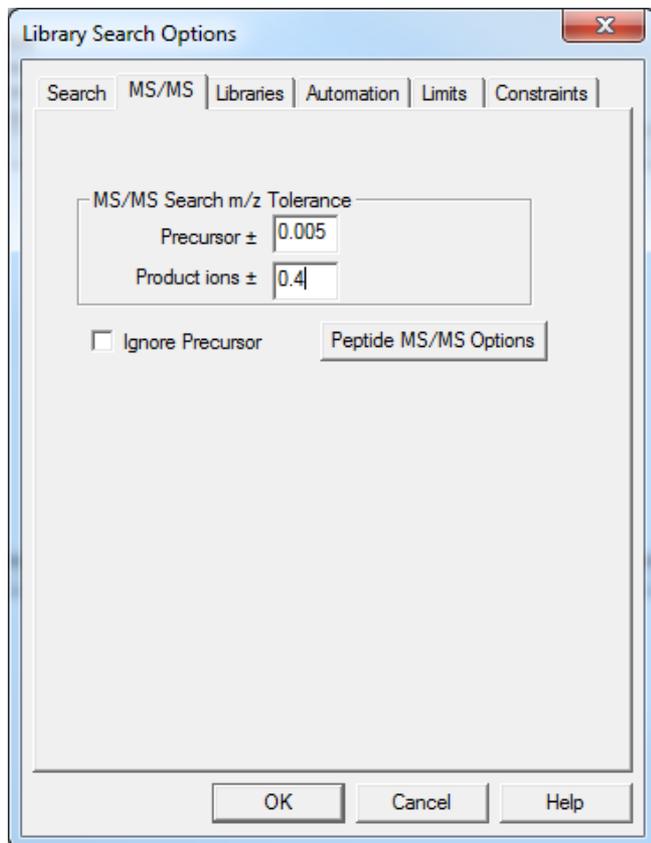
Click on "Import All" and in the next window, choose Overwrite or Prepend. It'll take some time to load all the spectra.



Step 4: Using NIST search for library validation

3. When the importing is done, go to the Library Search Options button, and set appropriate parameters.

For details, please refer to the user manual of lipidblast.

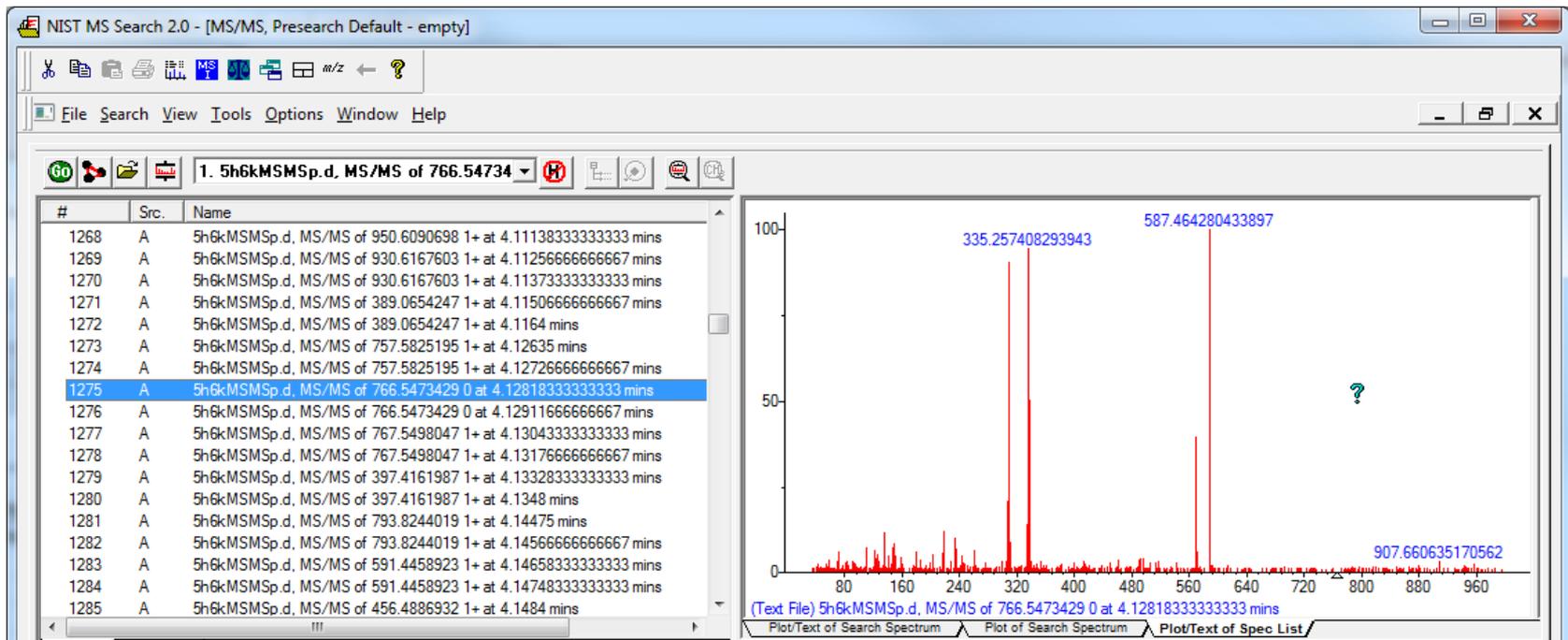


Step 4: Using NIST search for library validation

4. Scroll down in the upper-left window and find your potential lipids by its title (mainly RT, since RT is the increasing order).

Click on its title to show its spectra in the upper-right window.

Title	m/z	RT(min)	Name	Adduct
5h6kMSMSp.d, MS/MS of 766.5473429 0 at 4.128183333333333 mins	766.547	4.13	MGDG 34:5(16:2/18:3)	[M+NH4] ⁺



Step 4: Using NIST search for library validation

Double click on the title to search against libraries. Sometimes there's no good match in lipidblast because the specific lipid is not covered in lipidblast.

The screenshot displays the NIST MS Search 2.0 interface. The main window shows a search for a peptide with a mass of 766.54734. The search results are displayed in a table with columns for #, Src, Name, and Score. The top result is a peptide with a score of 186. The mass spectrum plot shows the relative intensity of ions versus their mass-to-charge ratio (m/z). The x-axis ranges from 0 to 1000, and the y-axis ranges from 0 to 100. The base peak is at m/z 335.257408293943. Other significant peaks are labeled at m/z 587.464280433897, 907.660635170562, 135.117528648017, 412.365935741688, 474.26225, and 667.278227157134. The library hit list shows a match with a score of 186. The hit list table is as follows:

#	Lib	Score	Dot Pro...	Prob. (%)	Rev-Dot	Name
1	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(6Z,9Z,12Z)/20:1(11E))
2	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(6Z,9Z,12Z)/20:1(11Z))
3	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(6Z,9Z,12Z)/20:1(13E))
4	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(6Z,9Z,12Z)/20:1(13Z))
5	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(9Z,12Z,15Z)/20:1(11E))
6	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(9Z,12Z,15Z)/20:1(11Z))
7	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(9Z,12Z,15Z)/20:1(13E))
8	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(18:3(9Z,12Z,15Z)/20:1(13Z))
9	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(11E)/18:3(6Z,9Z,12Z))
10	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(11E)/18:3(9Z,12Z,15Z))
11	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(11Z)/18:3(6Z,9Z,12Z))
12	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(11Z)/18:3(9Z,12Z,15Z))
13	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(13E)/18:3(6Z,9Z,12Z))
14	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(13E)/18:3(9Z,12Z,15Z))
15	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(13Z)/18:3(6Z,9Z,12Z))
16	li	153	186	4.06	714	PE 38.4; [M-H] ⁻ ; GPEn(20:1(13Z)/18:3(9Z,12Z,15Z))
17	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(6E,4E)/20:1(11Z,14Z))
18	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(6E,4E)/20:2(5Z,8Z))
19	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(6Z,9Z)/20:2(11Z,14Z))
20	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(6Z,9Z)/20:2(5Z,8Z))
21	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9E,10E)/20:2(11Z,14Z))
22	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9E,10E)/20:2(5Z,8Z))
23	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9E,12E)/20:2(11Z,14Z))
24	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9E,12E)/20:2(5Z,8Z))
25	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9Z,11Z)/20:2(11Z,14Z))
26	li	95	117	0.74	844	PE 38.4; [M-H] ⁻ ; GPEn(18:2(9Z,11Z)/20:2(5Z,8Z))

Step 4: Using NIST search for library validation

If the library search result match the MS2Analyzer result, then the annotation is further confirmed. Here's an example of good match.

Title	m/z	RT(min)	Name	Adduct
5h6kMSMSp.d, MS/MS of 586.5369873 1+ at 6.739 mins	586.537	6.74	DG 32:0(16:0/16:0)	[M+NH4] ⁺

NIST MS Search 2.0 - [Peptide, Research Default - 17 spectra]

File Search View Tools Options Window Help

1. 5h6kMSMSp.d, MS/MS of 586.5369873

#	Src	Name
2287	A	5h6kMSMSp.d, MS/MS of 552.434595 0 at 6.716583333333333 mins
2288	A	5h6kMSMSp.d, MS/MS of 552.434595 0 at 6.718133333333333 mins
2289	A	5h6kMSMSp.d, MS/MS of 553.4381104 1+ at 6.728666666666667 mins
2290	A	5h6kMSMSp.d, MS/MS of 553.4381104 1+ at 6.730166666666667 mins
2291	A	5h6kMSMSp.d, MS/MS of 628.5865479 1+ at 6.731916666666667 mins
2292	A	5h6kMSMSp.d, MS/MS of 628.5865479 1+ at 6.73365 mins
2293	A	5h6kMSMSp.d, MS/MS of 612.5553284 1+ at 6.73545 mins
2294	A	5h6kMSMSp.d, MS/MS of 612.5553284 1+ at 6.737233333333333 mins
2295	A	5h6kMSMSp.d, MS/MS of 586.5369873 1+ at 6.739 mins
2296	A	5h6kMSMSp.d, MS/MS of 586.5369873 1+ at 6.740766666666667 mins
2297	A	5h6kMSMSp.d, MS/MS of 654.6063232 1+ at 6.750716666666667 mins
2298	A	5h6kMSMSp.d, MS/MS of 654.6063232 1+ at 6.751633333333333 mins
2299	A	5h6kMSMSp.d, MS/MS of 371.0549708 1+ at 6.75235 mins

lipidblast-pos: pc-ac-neg.msp, pc-form-neg.msp, custompc-ppos.msp, custompc-nrpos.msp, lipidblast-neg. 234420 total spectra

Plot of Search Spectrum

#	Lib	Score	Dot Pro...	Prob. (%)	Rev-Dot	Name
1	li	319	489	94.8	889	DG 32:0, [M+NH4] ⁺ , DG(16:0/16:0/0)
2	li	76	144	0.94	687	DG 32:0, [M+NH4] ⁺ , DG(14:0/18:0/0)
3	li	76	144	0.94	687	DG 32:0, [M+NH4] ⁺ , DG(18:0/14:0/0)
4	li	66	127	0.66	631	DG 32:0, [M+NH4] ⁺ , DG(15:0/17:0/0)
5	li	66	127	0.66	631	DG 32:0, [M+NH4] ⁺ , DG(17:0/15:0/0)
6	li	33	67	0.17	365	DG 32:0, [M+NH4] ⁺ , DG(12:0/20:0/0)
7	li	33	67	0.17	365	DG 32:0, [M+NH4] ⁺ , DG(20:0/12:0/0)
8	li	33	65	0.17	356	DG 32:0, [M+NH4] ⁺ , DG(8:0/24:0/0)
9	li	33	65	0.17	356	DG 32:0, [M+NH4] ⁺ , DG(13:0/19:0/0)
10	li	33	65	0.17	356	DG 32:0, [M+NH4] ⁺ , DG(19:0/13:0/0)
11	li	33	65	0.17	356	DG 32:0, [M+NH4] ⁺ , DG(24:0/8:0/0)
12	li	32	65	0.16	356	DG 32:0, [M+NH4] ⁺ , DG(9:0/23:0/0)
13	li	32	65	0.16	356	DG 32:0, [M+NH4] ⁺ , DG(23:0/9:0/0)
14	li	32	64	0.16	352	DG 32:0, [M+NH4] ⁺ , DG(10:0/22:0/0)
15	li	32	64	0.16	352	DG 32:0, [M+NH4] ⁺ , DG(22:0/10:0/0)
16	li	21	42	0.11	235	DG 32:0, [M+NH4] ⁺ , DG(11:0/21:0/0)
17	li	21	42	0.11	235	DG 32:0, [M+NH4] ⁺ , DG(21:0/11:0/0)

Difference | Head to Tail | Side by Side | Subtraction

Head to Tail MF=319 RMF=489 DG 32:0, [M+NH4]⁺, DG(16:0/16:0/0) 319.489R 84.8P

Plot of Hit

lipidblast-pos: DG 32:0, [M+NH4]⁺, DG(16:0/16:0/0)

Plot of Hit

Peptide

That's it!

Thanks for using MS2Analyzer.

Feel free to email yanma@ucdavis.edu if you have any question and comment.