Lipid Identification with Greazy and LipidLama

David L. Tabb, Ph.D.

dtabb1973@gmail.com

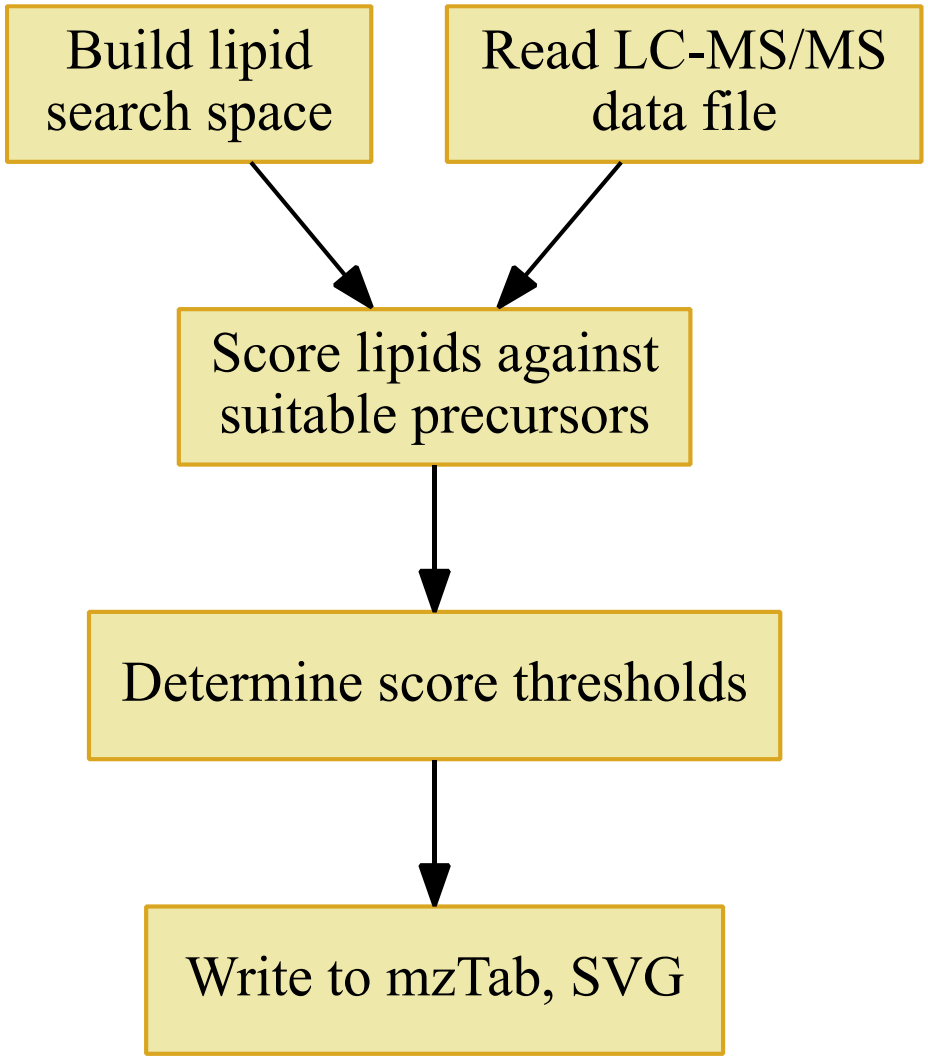
Overview

- Introduction to Greazy software for phospholipid identification
- Evaluation of identification performance using NIST LC-MS/MS library

Goals of Greazy Project

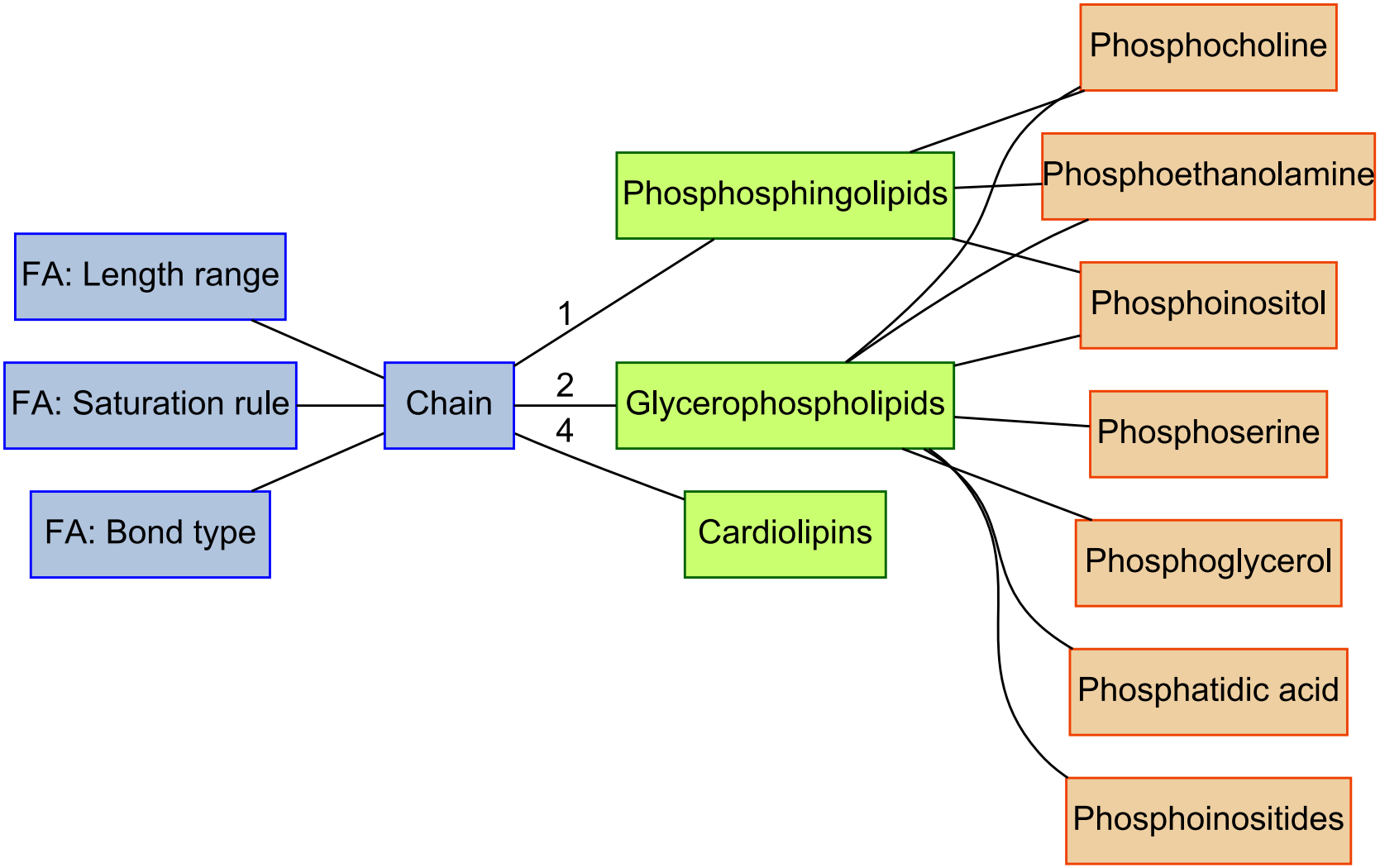
- Apply lessons from shotgun proteomics to phospholipid tandem mass spectra
  - Define search spaces for lipids to be analogous to protein sequence database.
  - Develop models to predict fragments observed from phospholipids under CID conditions.
  - Adapt scoring systems to reward matching of small peaks.
  - Determine thresholds for a given FDR.

Greazy and LipidLama Workflow



Michael Kochen, *in preparation*

Extrapolating a search space



Modeling a positive MS/MS

GP nonmetal fragments	PC	PE	PS	PG	PA	PI	PIP	PIP2	PIP3
HG	X	X	X						
loss of HG	X	X	X	X	X	X	X	X	X
loss of FA1 as ketene	X	X	X	X	X	X	X	X	X
loss of FA1 as carboxylic acid	X	X	X	X	X	X	X	X	X
loss of HG and FA1 as ketene	X	X	X	X	X	X	X	X	X
R1CO	X	X	X	X	X	X	X	X	X
loss of FA2 as ketene	X	X	X	X	X	X	X	X	X
loss of FA2 as carboxylic acid	X	X	X	X	X	X	X	X	X
loss of HG and FA2 as ketene	X	X	X	X	X	X	X	X	X
R2CO	X	X	X	X	X	X	X	X	X
Choline	X								
loss of NC3H9	X								

Phosphatidylcholines produce a few distinctive fragments, but but most fragments are universal across types.

Modeling a negative MS/MS

GP negative mode	PC	PE	PS	PG	PA	PI	PIP	PIP (-2)	PIP2	PIP2 (-2)	PIP3
G3P-H2O	X	X	X	X	X	X	X	X	X	X	X
G3P	X	X	X	X	X	X					
H2PO4	X	X	X	X	X	X	X	X	X	X	X
PO3	X	X	X	X	X	X	X	X	X	X	X
FA1 carboxylate anion	X	X	X	X	X	X	X	X	X	X	X
loss of FA1 as ketene	X	X	X	X	X	X	X				
loss of FA1 as carboxylic acid	X	X	X	X	X	X	X				
FA2 carboxylate anion	X	X	X	X	X	X	X	X	X	X	X
loss of FA2 as ketene	X	X	X	X	X	X	X	X			
loss of FA2 as carboxylic acid	X	X	X	X	X	X	X				
loss of head		X	X	X		X					
loss of head and FA1 as carboxylic acid		X	X	X		X					
loss of head and FA1 as ketene		X	X	X		X					
loss of head and FA2 as carboxylic acid		X	X	X		X					
loss of head and FA2 as ketene		X	X	X		X					
loss of FA1 as carboxylate anion								X		X	
loss of FA2 as carboxylate anion								X		X	
loss of FA2 as ketene and PO3								X			
loss of FA2 as ketene and FA1 carboxylate anion								X			
head group		X	X			X					
head group - H2O		X				X					
head group - 2*H2O						X					

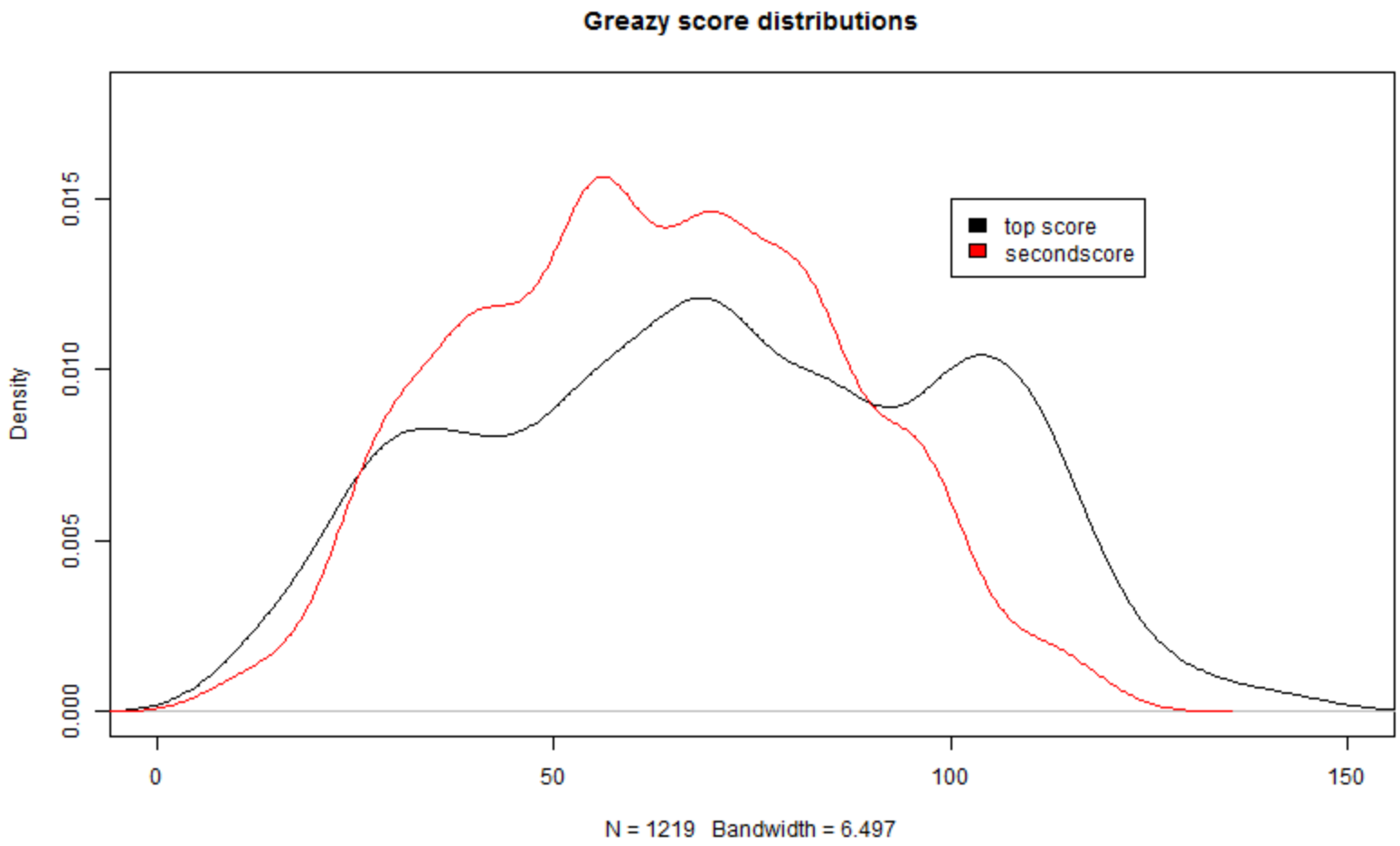
more

Negative mode models vary more by class of lipid, with many fragments found in just one type.

Scoring a lipid-spectrum match

- Hypergeometric score: what is the probability that we would match *at least* as many peaks as we did by random chance alone?
- Intensity score: what is the probability of matching *at least* as much of the MS/MS total intensity as we did by random chance alone?
- Scores are expressed as p-values and combined through Fisher's method.

Setting thresholds



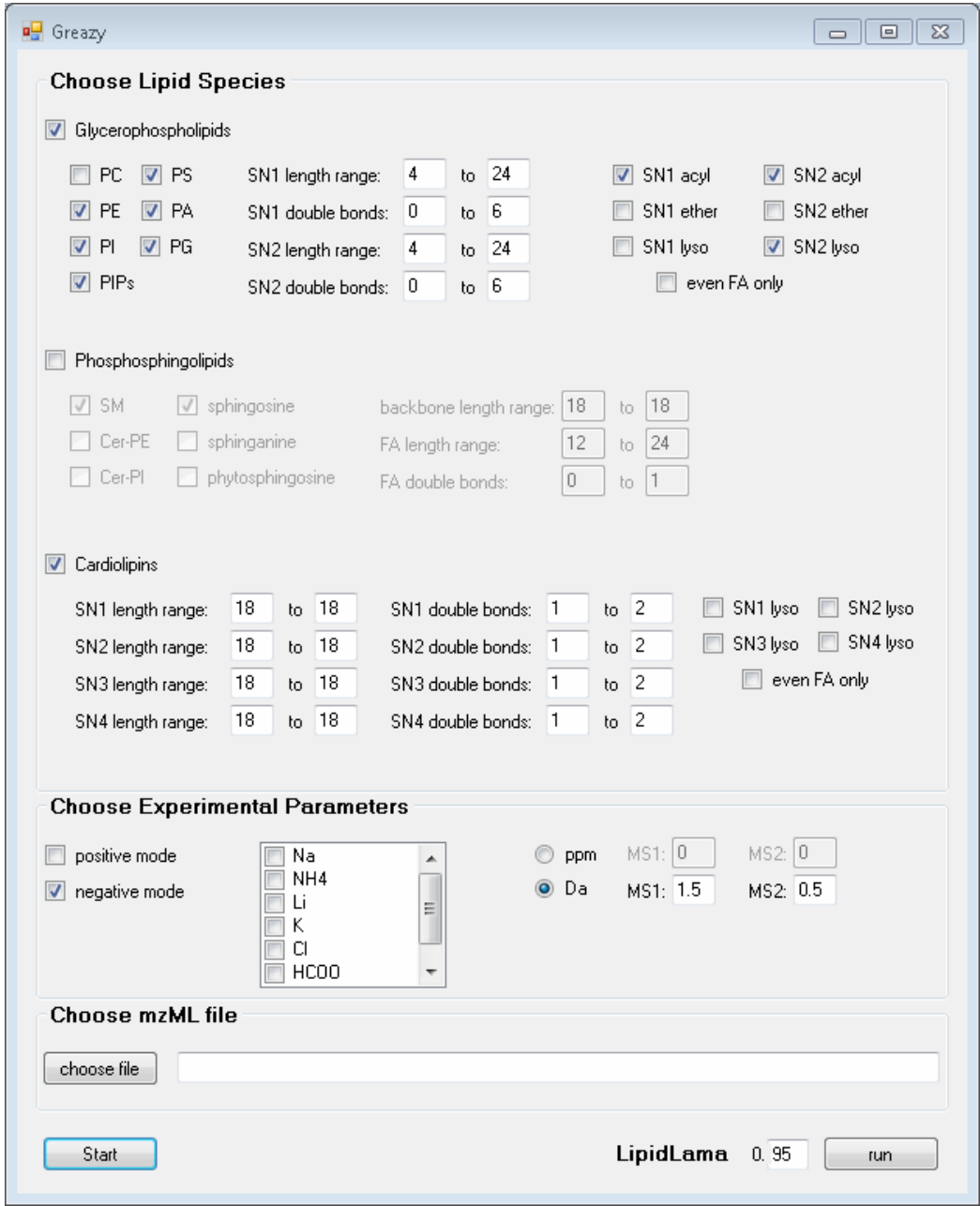
Second-ranked match scores are used to model the distribution of false scores.

NIST Data Set

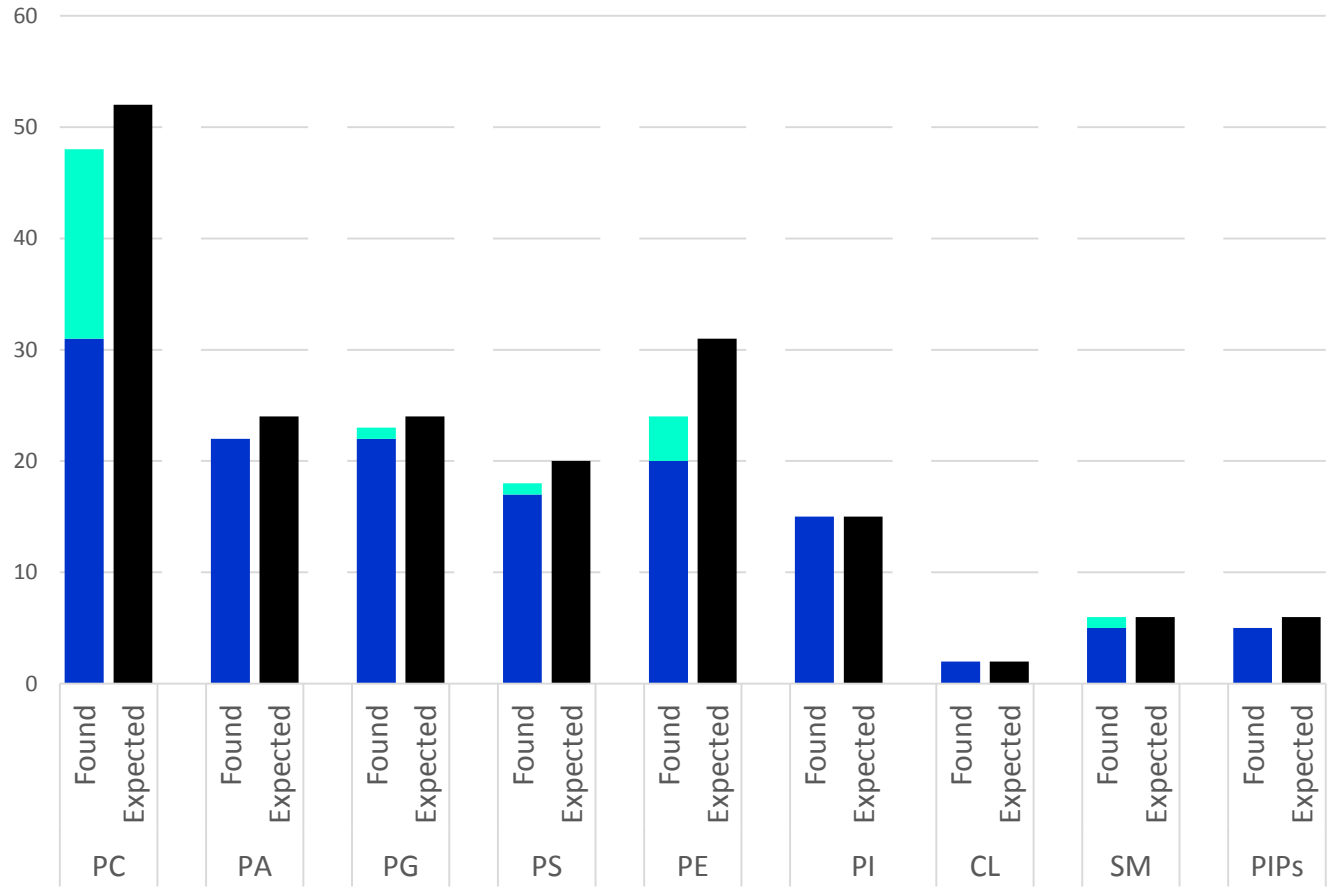
- SDF file was split in two based on ion mode
  - Positive mode: 161,355 spectra
  - Negative mode: 31,764 spectra
- Phospholipids were extracted based on the strings "`glycero-3-phos`" & "`sphing`"
  - Positive mode: 3590 spectra, 198 distinct lipids
  - Negative mode: 1295 spectra, 206 distinct lipids
- Data were converted to MGF format

Lipid Space Configuration

- Glycerophospholipids
  - Chain lengths: 4-24
  - Double bonds: 0-6
  - Types: acyl & lyso
  - Head groups: PC, PE, PG, PS, PI, & PIPs
- Sphingomyelin
  - Backbone: d18:1
  - Chain length: 12-24
  - Double bonds: 0-1
- Cardiolipins
  - Chain length: 18
  - Double bonds: 1-2



Identifications by Class



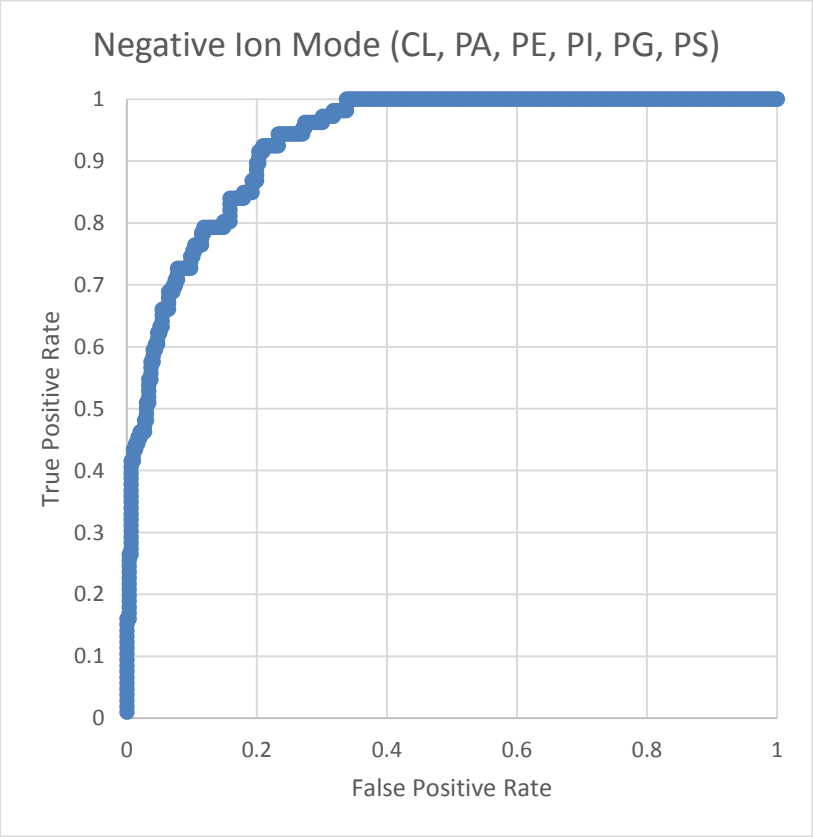
-180 of 259 distinct lipids were within search space.

-160 lipids were best matches for at least one MS/MS.

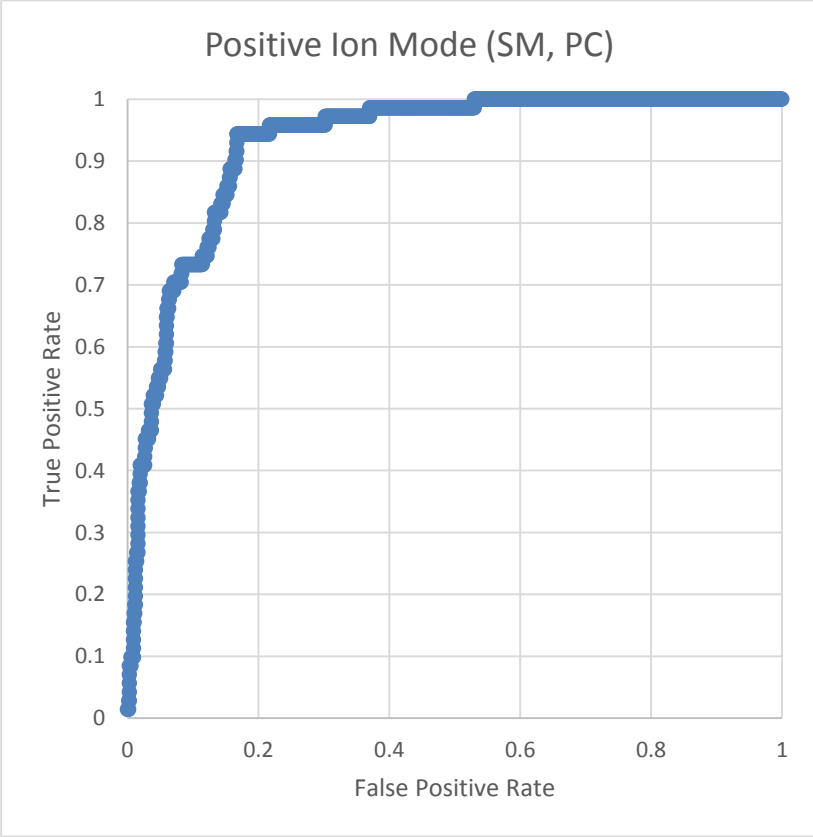
-136 lipids exceeded 95% threshold.

ROC Curves

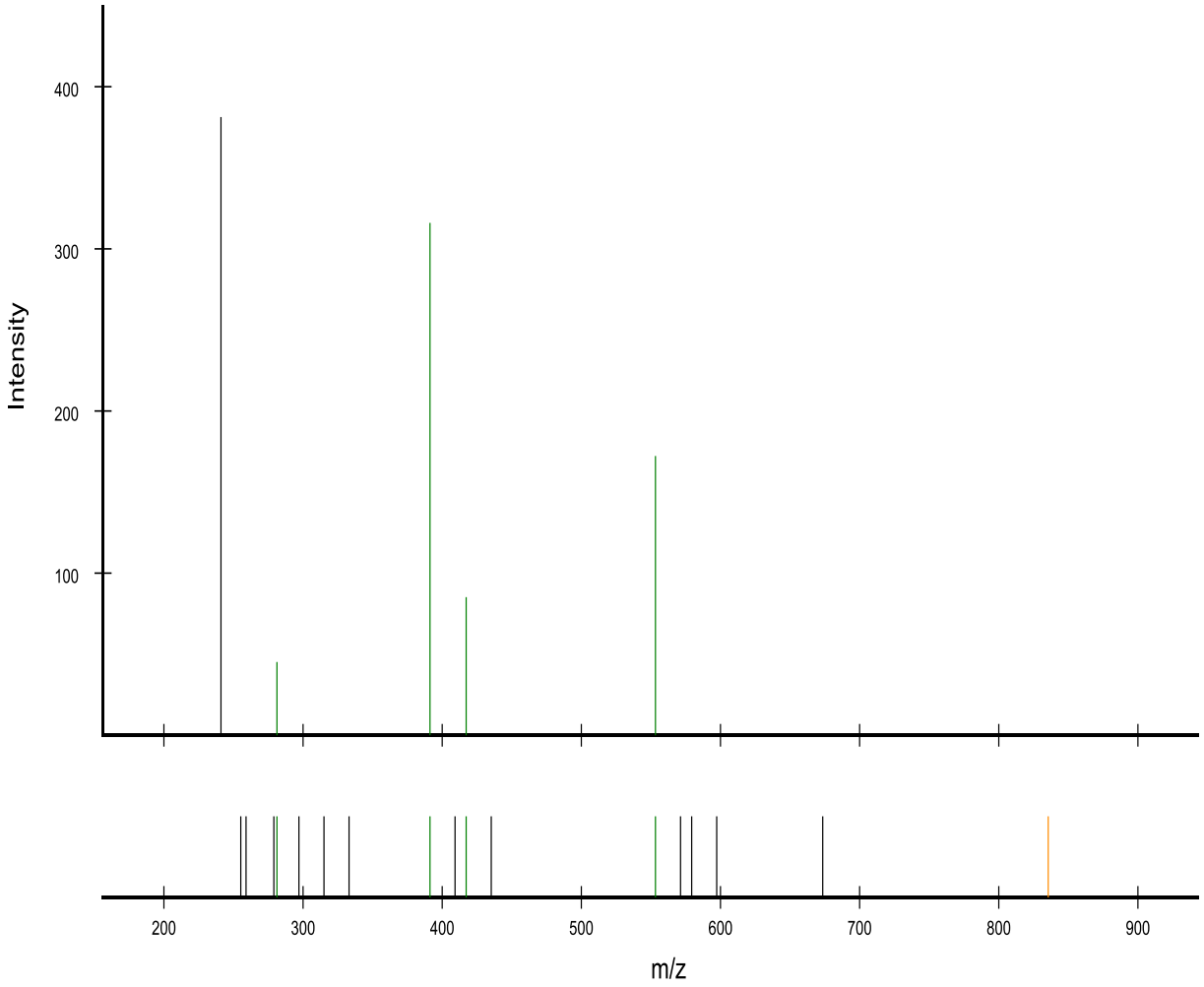
AUC: 0.933



AUC: 0.928



Collision Energy: 33 eV

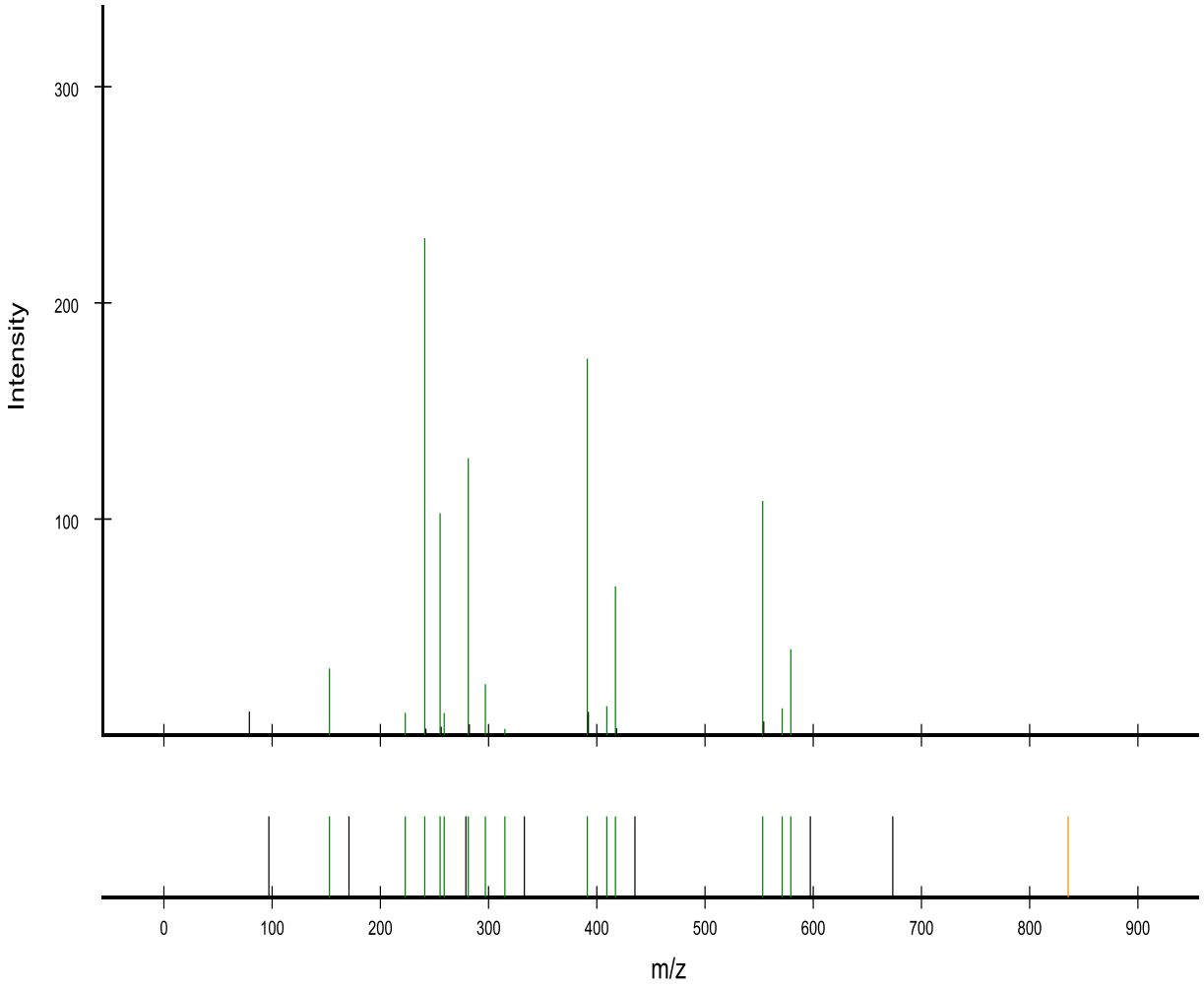


Lipid: PI(18:1/16:0)
Score: 32.5247

Precursor: 835.534 (deprotonated)

m/z	Intensity	Fragment
255.232	0	FA2 carboxylate anion
259.022	0	phosphoinositol
279.027	0	glycerophosphoinositol - 3*H2O
281.248	45.1958	FA1 carboxylate anion
297.038	0	glycerophosphoinositol - 2*H2O
315.048	0	glycerophosphoinositol - H2O
333.059	0	glycerophosphoinositol
391.225	316.036	loss of inositol; loss of FA1 as carboxylic acid
409.236	0	loss of inositol; loss of FA1 as ketene
417.241	85.2025	loss of inositol; loss of FA2 as carboxylic acid
435.251	0	loss of inositol; loss of FA2 as ketene
553.278	172.246	loss of FA1 as carboxylic acid
571.288	0	loss of FA1 as ketene
579.293	0	loss of FA2 as carboxylic acid
597.304	0	loss of FA2 as ketene
673.481	0	loss of inositol

Collision Energy: 41 eV

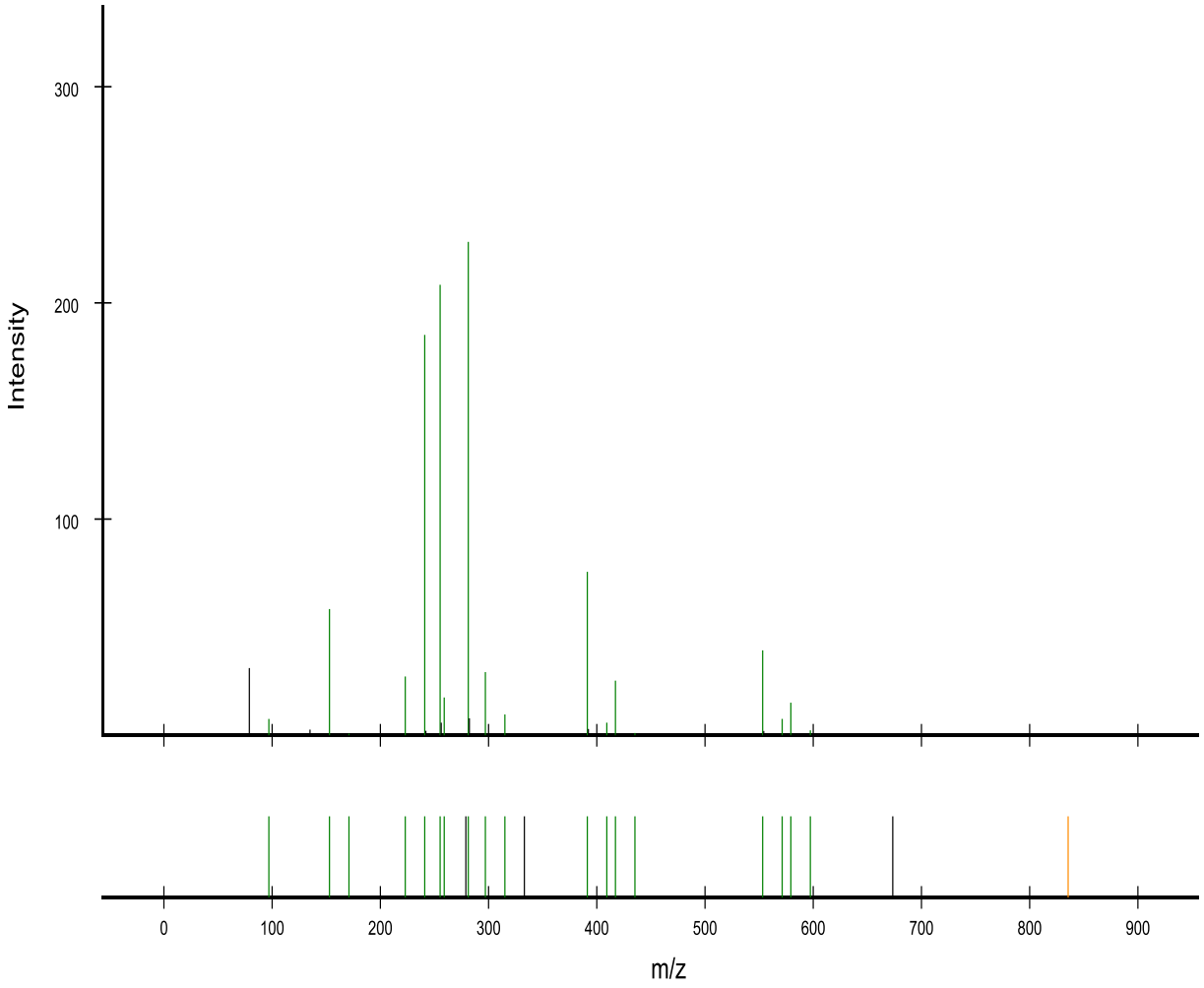


Lipid: PI(18:1/16:0)
Score: 125.19

Precursor: 835.534 (deprotonated)

m/z	Intensity	Fragment
96.9691	0	H2PO4
152.995	30.8947	G3P-H2O
171.006	0	G3P
223.001	10.3974	phosphoinositol - 2*H2O
241.011	229.921	phosphoinositol - H2O
255.232	102.685	FA2 carboxylate anion
259.022	10.2982	phosphoinositol
279.027	0	glycerophosphoinositol - 3*H2O
281.248	128.128	FA1 carboxylate anion
297.038	23.6691	glycerophosphoinositol - 2*H2O
315.048	2.97351	glycerophosphoinositol - H2O
333.059	0	glycerophosphoinositol
391.225	174.267	loss of inositol; loss of FA1 as carboxylic acid
409.236	13.47	loss of inositol; loss of FA1 as ketene
417.241	68.916	loss of inositol; loss of FA2 as carboxylic acid
435.251	0	loss of inositol; loss of FA2 as ketene
553.278	108.424	loss of FA1 as carboxylic acid
571.288	12.4788	loss of FA1 as ketene
579.293	39.8053	loss of FA2 as carboxylic acid
597.304	0	loss of FA2 as ketene
673.481	0	loss of inositol

Collision Energy: 54 eV



Lipid: PI(18:1/16:0)
Score: 160.818

Precursor: 835.534 (deprotonated)

m/z	Intensity	Fragment
96.9691	7.55552	H2PO4
152.995	58.3834	G3P-H2O
171.006	0.753953	G3P
223.001	27.2063	phosphoinositol - 2*H2O
241.011	185.196	phosphoinositol - H2O
255.232	208.386	FA2 carboxylate anion
259.022	17.5054	phosphoinositol
279.027	0	glycerophosphoinositol - 3*H2O
281.248	228.242	FA1 carboxylate anion
297.038	29.2374	glycerophosphoinositol - 2*H2O
315.048	9.63232	glycerophosphoinositol - H2O
333.059	0	glycerophosphoinositol
391.225	75.6169	loss of inositol; loss of FA1 as carboxylic acid
409.236	5.84199	loss of inositol; loss of FA1 as ketene
417.241	25.2894	loss of inositol; loss of FA2 as carboxylic acid
435.251	1.07381	loss of inositol; loss of FA2 as ketene
553.278	39.281	loss of FA1 as carboxylic acid
571.288	7.57837	loss of FA1 as ketene
579.293	15.0859	loss of FA2 as carboxylic acid
597.304	2.30527	loss of FA2 as ketene
673.481	0	loss of inositol

Takeaway messages

- Identification can employ synthetic spectra, not just those that have been experimentally created.
- Lessons learned from proteomics can be adapted to metabolomics with good effect.
- Lipids produce relatively few fragments for structural confirmation.