# Automated lipid classification by machine learning

Ryan Taylor, Ryan H. Miller, Ryan D. Miller, Michael Porter, James Dalgleish and John T. Prince

# Supplementary Materials

## 1 Master Table of Misclassified Lipids

## 1.1 Table 1: Overview of identified misclassifications

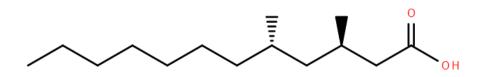
Lipid(s)	<b>Current Classification</b>	Suggested Classification	Details
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] > Branched	
LMFA01010053	Straight chain fatty acids [FA0101]	fatty acids [FA0102]	2.1
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01020363, LMFA01020364	Branched fatty acids [FA0102]	Unsaturated fatty acids [FA0103]	2.2
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
LMFA01030188, LMFA01030189,	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01030191	Unsaturated fatty acids [FA0103]	Carbocyclic fatty acids [FA0114]	2.3
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01020274, LMFA01020276	Branched fatty acids [FA0102]	Oxo fatty acids [FA0106]	2.4
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01030579	Unsaturated fatty acids [FA0103]	Oxo fatty acids [FA0106]	2.5
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01030675, LMFA01030676	Unsaturated fatty acids [FA0103]	Halogenated fatty acids [FA0109]	2.6
	Fatty Acyls [FA] >	Fatty Acyls [FA] >	
	Fatty Acids and Conjugates [FA01] >	Fatty Acids and Conjugates [FA01] >	
LMFA01030714	Unsaturated fatty acids [FA0103]	Hydroxy fatty acids [FA0105]	2.7
		Fatty Acyls [FA] >	
	Fatty Acyls [FA] >	Eicosanoids [FA03] >	
	Fatty Acids and Conjugates [FA01] >	Hydroxy/hydroperoxyeicosapentae-noic acids	
LMFA01030717	Unsaturated fatty acids [FA0103]	[FA0307]	2.8
		Fatty Acyls [FA] >	
	Fatty Acyls [FA] >	Eicosanoids [FA03] >	
	Fatty Acids and Conjugates [FA01] >	Hydroxy/hydroperoxyeicosatetrae-noic acids	
LMFA01030719	Unsaturated fatty acids [FA0103]	[FA0306]	2.9
LMFA01030750, LMFA01030790,			
LMFA01030792, LMFA01030796,	Fatty Acyls [FA] >		
LMFA01030797, LMFA01030798,	Fatty Acids and Conjugates [FA01] >	Fatty Acyls [FA] >	
LMFA01030799, LMFA01030893,	Unsaturated fatty acids [FA0103]	Fatty Acids and Conjugates [FA01] >	
LMFA01030895, LMFA01030905		Branched fatty acids [FA0102]	2.10
	Glycerophospholipids [GP] >		
	Glycerophosphocholines [GP01] >	Glycerophospholipids [GP] >	
	1-(1Z-alkenyl),2-	Glycerophosphoethanolamines [GP02] >	
LMGP01030016	acylglycerophosphocholines [GP0103]	Diacylglycerophosphoethanolamines [GP0201]	2.11
	Glycerophospholipids [GP] >	Glycerophospholipids [GP] >	
	Glycerophosphoglycerols [GP04] >	Glycerophosphoglycerols [GP04] >	
	Dialkylglycerophosphoglycerols	1-acyl,2-alkylglycero-phosphoglycerols	
LMGP04040006	[GP0404]	[GP0411]	2.12
	Sphingolipids [SP] >	Sphingolipids [SP] >	
	Neutral glycosphingolipids [SP05] >	Neutral glycosphingolipids [SP05] >	
	Galβ1-4GlcNAcβ1-3Galβ1-4Glc-	GalNAcβ1-4Galβ1-4Glc- (Ganglio series)	
LMSP0505DO01	(Neolacto series) [SP0505]	[SP0503]	2.13

LMSP0505DP01-			
LMSP0505DP08,			
LMSP0505DQ01-			
LMSP0505DQ08,	Sphingolipids [SP] >	Sphingolipids [SP] >	
LMSP0505DR01-	Neutral glycosphingolipids [SP05] >	Neutral glycosphingolipids [SP05] >	
LMSP0505DR08,	Galβ1-4GlcNAcβ1-3Galβ1-4Glc-	GalNAcβ1-4Galβ1-4Glc- (Ganglio series)	
LMSP0505DS01-LMSP0505DS08	(Neolacto series) [SP0505]	[SP0503]	2.14
LMSP0505DA01-			
LMSP0505DA08,			
LMSP0505DB01-			
LMSP0505DB08,			
LMSP0505DJ01-LMSP0505DJ08,	Sphingolipids [SP] >	Sphingolipids [SP] >	
LMSP0505DK01-	Neutral glycosphingolipids [SP05] >	Neutral glycosphingolipids [SP05] >	
LMSP0505DK08,	Galβ1-4GlcNAcβ1-3Galβ1-4Glc-	GalNAcβ1-3Galα1-3Galβ1-4Glc- (Isoglobo	
LMSP0505DL01-LMSP0505DL08	(Neolacto series) [SP0505]	series) [SP0506]	2.15

Table 1 shows the lipids that were identified as misclassified by LIPID MAPS. The current classification and suggested classification are shown.

## 2 Ontology Reasoning

2.1



Lipid(s): LMFA01010053

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Straight chain fatty

acids [FA0101]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Branched fatty acids

[FA0102]

Discussion: This lipid exhibits two branching events characteristic of lipids in the "Branched

fatty acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01010001	LMFA01020001

ОН	ОН
LMFA01010002	LMFA01020002
ОН	ОН
LMFA01010003	LMFA01020003
ОН	ОН
LMFA01010004	LMFA01020004
ОН	ОН ОН

Lipid(s): LMFA01020363, LMFA01020364

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Branched fatty acids

[FA0102]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Discussion: These lipids have unsaturated bonds characteristic of lipids in the "Unsaturated

fatty acids" subclass, but do not exhibit branching.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01020001	LMFA01030001
ОН	ОН
LMFA01020002	LMFA01030002

ОН	ОН
LMFA01020003	LMFA01030004
ОН	ОН
LMFA01020004	LMFA01030005
ОН	ОН

Lipid(s): LMFA01030188, LMFA01030189, LMFA01030191

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Carbocyclic fatty acids

[FA0114]

Discussion: These lipids have only have unsaturated bonds in a carbocyclic ring characteristic

of some lipids in the "Carbocyclic fatty acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01030001	LMFA01140025
ОН	OH OH

LMFA01030002	LMFA01140028
ОН	OH
LMFA01030004	LMFA01140018
ОН	OH OH
LMFA01030005	LMFA01140023
ОН	ОН

Lipid(s): LMFA01020274, LMFA01020276

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Branched fatty acids

[FA0102]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Oxo fatty acids

[FA0106]

Discussion: These lipids exhibit an additional carbonyl group characteristic of lipids in the

"Oxo fatty acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01020001	LMFA01060002

ОН	
LMFA01020002	LMFA01060111
ОН	
LMFA01020003	LMFA01060157
ОН	ОН
LMFA01020004	LMFA01060178
ОН	→ No Septiment (Septiment)

Lipid(s): LMFA01030579

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Oxo fatty acids

[FA0106]

Discussion: This lipid exhibits an additional carbonyl group characteristic of lipids in the

"Oxo fatty acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:

LMFA01030006	LMFA01060148
ОН	OH OH
LMFA01030023	LMFA01060111
ОН	
LMFA01030043	LMFA01060093
ОН	
LMFA01030048	LMFA01060095
ОН	Н

Lipid(s): LMFA01030675, LMFA01030676

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

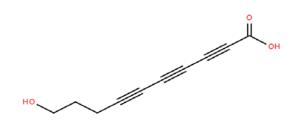
Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Halogenated fatty acids

[FA0109]

Discussion: These lipids are halogenated which is a characteristic of lipids in the

"Halogenated fatty acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01030006	LMFA01090031
ОН	ОМ
LMFA01030023	LMFA01090073
ОН	СІ
LMFA01030043	LMFA01090088
ОН	в он
LMFA01030048	LMFA01090100
ОН	Br OH



Lipid(s): LMFA01030714

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Hydroxy fatty acids

[FA0105]

Discussion: This lipid has a hydroxy group characteristic of lipids in the "Hydroxy fatty

acids" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
--	--

LMFA01030006	LMFA01050232
ОН	но
LMFA01030023	LMFA01050258
ОН	
LMFA01030043	LMFA01050272
ОН	
LMFA01030048	LMFA01050324
ОН	<u>=</u> он

Lipid(s): LMFA01030717

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Suggested classification: Fatty Acyls [FA] > Eicosanoids [FA03] > Hydroxy/hydroperoxyeicosapentaenoic

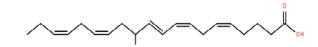
acids [FA0307]

Discussion: This lipid has a hydroxy group as well as five double bonds and 20 carbons, all

characteristics of lipids in the "Hydroxy/hydroperoxyeicosatetraenoic acids"

subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01030006	LMFA03070031



ОН	
LMFA01030023	LMFA03070028
ОН	
LMFA01030043	LMFA03070041
ОН	
LMFA01030048	LMFA03070049
ОН	HO, II, OH

Lipid(s): LMFA01030719

Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids Current classification:

[FA0103]

Fatty Acyls [FA] > Eicosanoids [FA03] > Hydroxy/hydroperoxyeicosatetraenoic Suggested classification:

acids [FA0306]

This lipid has a hydroxy group as well as four double bonds and 20 carbons, all characteristics of lipids in the "Hydroxy/hydroperoxyeicosatetraenoic acids" Discussion:

subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01030006	LMFA03060030

ОН	
LMFA01030023	LMFA03060012
ОН	
LMFA01030043	LMFA03060018
ОН	
LMFA01030048	LMFA03060044
ОН	ОН

Lipid(s): LMFA01030750, LMFA01030790, LMFA01030792, LMFA01030796,

LMFA01030797, LMFA01030798, LMFA01030799, LMFA01030893,

LMFA01030895, LMFA01030905

Current classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Unsaturated fatty acids

[FA0103]

Suggested classification: Fatty Acyls [FA] > Fatty Acids and Conjugates [FA01] > Branched fatty acids

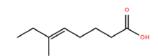
[FA0102]

Discussion: These lipids exhibit branching events characteristic of lipids in the "Branched

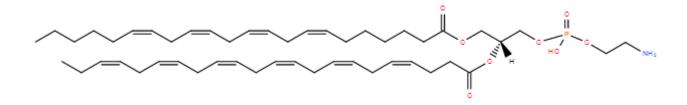
fatty acids" subclass. Examination of lipids classified by LIPID MAPS shows

branching takes precedence over unsaturation.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMFA01030006	LMFA01020103



ОН	
LMFA01030023	LMFA01020207
ОН	
LMFA01030043	LMFA01020045
ОН	
LMFA01030048	LMFA01020209
ОН	



Lipid(s): LMGP01030016

Current classification: Glycerophospholipids [GP] > Glycerophosphocholines [GP01] > 1-(1Z-

alkenyl),2-acylglycerophosphocholines [GP0103]

Suggested classification: Glycerophospholipids [GP] > Glycerophosphoethanolamines [GP02] >

Diacylglycerophosphoethanolamines [GP0201]

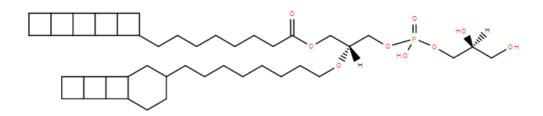
Discussion: The structure consists of two fatty acyls joined to the glycerol backbone at the 1

and 2 positions. Thus, it is inaccurate to describe it as a alkenyl containing

group. The right classification is the diacylglycero-phosphocholine.

Representative lipids from current subclass:	Representative lipids from suggested subclass:	
--	--	--

LMGP01030004	LMGP02010003
	NMs
LMGP01030006	LMGP02010004
LMGP01030008	LMGP02010005
LMGP01030008	LMGP02010005



Lipid(s): LMGP04040006

Current classification: Glycerophospholipids [GP] > Glycerophosphoglycerols [GP04] >

Dialkylglycerophosphoglycerols [GP0404]

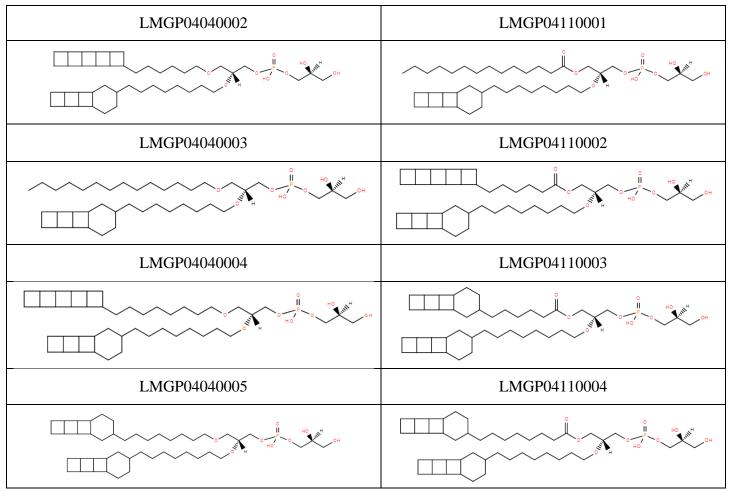
Suggested classification: Glycerophospholipids [GP] > Glycerophosphoglycerols [GP04] > 1-acyl,2-

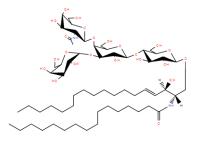
alkylglycerophosphoglycerols [GP0411]

Discussion: This lipid has an ester group characteristic of lipids in the "1-acyl,2-

alkylglycerophosphoglycerols" subclass.

Representative lipids from current subclass:	Representative lipids from suggested subclass:





Lipid(s): LMSP0505DO01,

Current classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > Galβ1-4GlcNAcβ1-

3Galβ1-4Glc- (Neolacto series) [SP0505]

Suggested classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > GalNAcβ1-4Galβ1-

4Glc- (Ganglio series) [SP0503]

Discussion: The sugar chain starting from Ceramide fits the Ganglio series root exactly

(GalNAc-Gal-Gal-Glc-Cer.). Discussion continued on the next page.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMSP0505AA01	LMSP0503AA01

LMSP0505AA02	LMSP0503AN01
LMSP0505AA03	LMSP0503AO01
LMSP0505AA04	LMSP0503AP01

-Continued Discussion of Lipids similar to LMSP0505DO01-08

Lipid(s): LMSP0505DP01-LMSP0505DP08, LMSP0505DQ01-LMSP0505DQ08,

LMSP0505DR01-LMSP0505DR08, LMSP0505DS01-LMSP0505DS08 (1-8 in each sub-sub section because only the Ceramide chain changes)

Current classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] >  $Gal\beta1-4GlcNAc\beta1-$ 

3Galβ1-4Glc- (Neolacto series) [SP0505]

Suggested classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > GalNAcβ1-4Galβ1-

4Glc- (Ganglio series) [SP0503]

Discussion: The structure is branched but the structure fits one root better than the other and

based on the 1997 IUPAC guidelines for naming glycolipids and the LIPID

MAPS own grouping rules, the root structure determines the group.

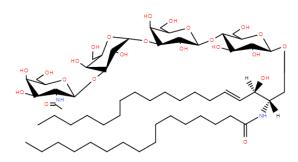
Representative lipids from current subclass:	Representative lipids from suggested subclass:

# LMSP0505DP01 Each of the sub-st DP-DS are the exa group's structure, one sugar to the s group. This allows part of the Gangli of these glycolipid root. LMSP0505DR01 LMSP0505DS01

## LMSP0503 Ganglio series

Each of the sub-sub groups after DO from DP-DS are the exact same as DO sub-sub group's structure, but with the addition of one sugar to the side chain for each new group. This allows them to be grouped as part of the Ganglio series as well. Also none of these glycolipids have the neolacto series root

2.15



Lipid(s): LMSP0505DA01-LMSP0505DB08, LMSP0505DB01-LMSP0505DB08,

LMSP0505DJ01-LMSP0505DJ08, LMSP0505DK01-LMSP0505DK08,

LMSP0505DL01-LMSP0505DL08 (1-8 in each sub-sub section because only the

Ceramide chain changes)

Current classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > Galβ1-4GlcNAcβ1-

3Galβ1-4Glc- (Neolacto series) [SP0505]

Suggested classification: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > GalNAcβ1-3Galα1-

3Galβ1-4Glc- (Isoglobo series) [SP0506]

Discussion: Based on the 1997 IUPAC guidelines for naming glycolipids and the sub-sub

groupings of LIPID MAPS themselves, these lipids fit the glyco-root of the Isoglobo series because these lipids don't have an N-acetyl-glucosamine in the third position from the Ceramide and have all the correct linkage for the Isoglobo

series.

Representative lipids from current subclass:	Representative lipids from suggested subclass:
LMSP0505AA01	LMSP0506AD01
LMSP0505 (Neolacto series)	LMSP0506 (Isoglobo series)
Galβ1-4GlcNAcβ1-3Galβ1-4Glc-Cer	GalNAcβ1-3Galα1-3Galβ1-4Glc-Cer

## 3 New Ontology

## 3.1 Table 2: Overview of suggested new ontology additions

Current Ontology	New Suggested Ontology:	Details
LMSP0505DC01-08 LMSP0505DD01-08 LMSP0505DE01-08 LMSP0505DF01-08	LMSP05 <u>10</u> (gluco-globo series) [10]	4.1
LMSP0505DM01-08 LMSP0505DN01-08	LMSP05 <u>11</u> (galacto-gluco series) [11]	4.2

Table 2 shows the lipids that would fit more appropriately in a new subclass of lipids. The new subclass LMSP0510 (gluco-globo) is suggested for the LMSP0505DC-DF lipids and the new subclass LMSP0511 is suggested for the LMSP0505DM- LMSP0505DN lipids.

## 4 New Ontology Reasoning

4.1

Lipid(s): LMSP0505DC01-08, LMSP0505DD01-08, LMSP0505DE01-08,

LMSP0505DF01-08

Current ontology: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > Galβ1-4GlcNAcβ1-

3Galβ1-4Glc- (Neolacto series) [SP0505]

Suggested ontology: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > GlcNAcβ1-3Galα1-

3Galβ1-4Glcβ-Cer (Gluco-globo series) [FA0106]

Discussion: The new subclass LMSP0510 (gluco-globo) is suggested for the LMSP0505DC-

DF lipids because the root of these lipids is similar to the isoglobo series and

contains a terminal N-acetyl glucosamine.

Current Root for Neolacto series [SP0505]	Root for New Suggested Ontology
Galβ1-4GlcNAcβ1-3Galβ1-4Glc-Cer	GlcNAcβ1-3Galα1-3Galβ1-4Glcβ-Cer (LMSP0510)

4.2

Lipid(s): LMSP0505DM01-08, LMSP0505DN01-08

Current ontology: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] > Galβ1-4GlcNAcβ1-

3Galβ1-4Glc- (Neolacto series) [SP0505]

Suggested ontology: Sphingolipids [SP] > Neutral glycosphingolipids [SP05] >  $Gal\alpha 1-3Gal\alpha 1-3Gal\beta 1-$ 

4Glcβ-Cer (Galacto-gluco series) [LMSP0511]

Discussion: The new subclass LMSP0511 (Galacto-lacto) is suggested for the

LMSP0505DM-DN lipids because Galacto-gluco illustrates the relationship to the

Gala series (LMSP0509) in that it contains repeated galactose monomers and

highlights the terminal glucose monomer.

Current Root for Neolacto series [SP0505]	Root for New Suggested Ontology
Galβ1-4GlcNAcβ1-3Galβ1-4Glc-Cer	Galα1-3Galα1-3Galβ1-4Glcβ-Cer (LMSP0511)

# **5 Lipid Category Decision Trees**



Figure 1: Root decision tree

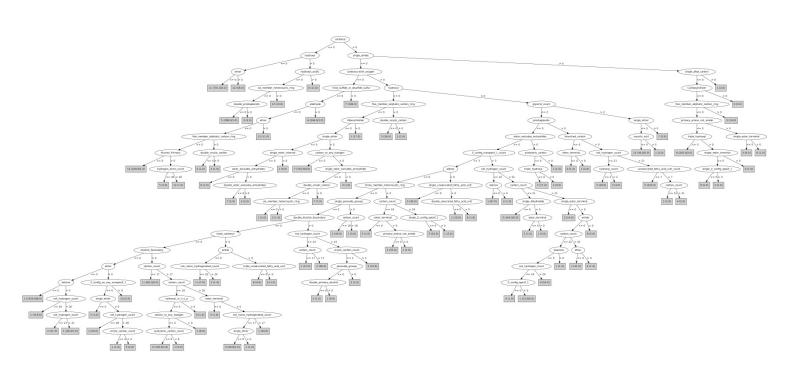


Figure 2: Fatty Acyls [FA] decision tree

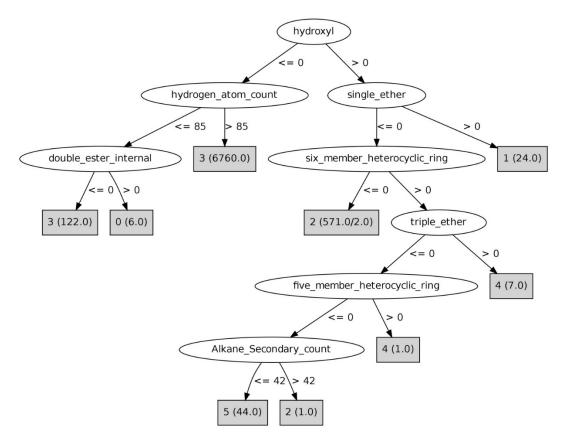


Figure 3: Glycerolipids [GL] decision tree

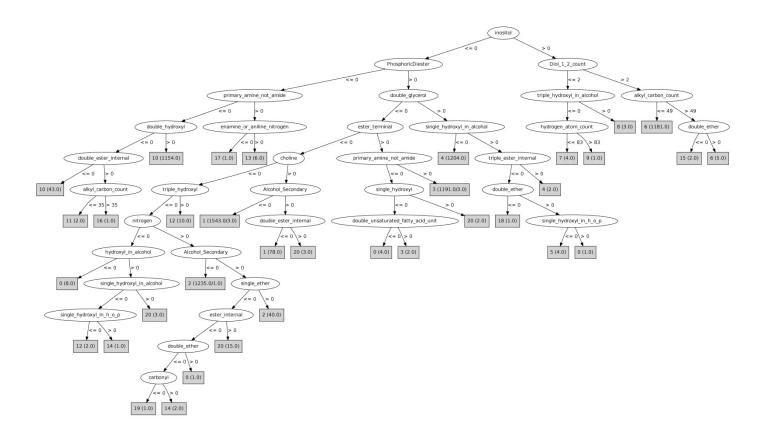


Figure 4: Glycerophospholipids [GP] decision tree

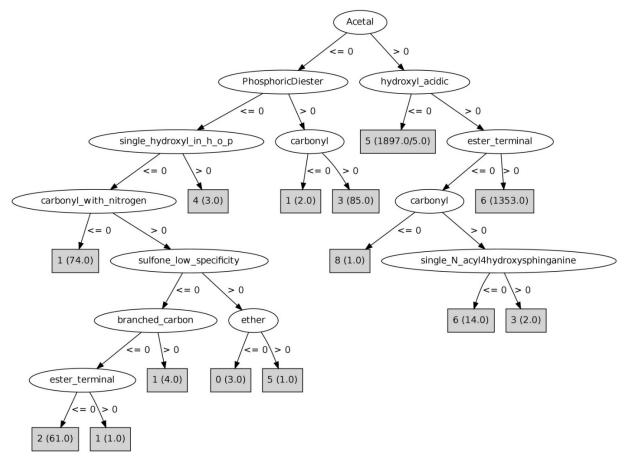


Figure 5: Sphingolipids [SP] decision tree

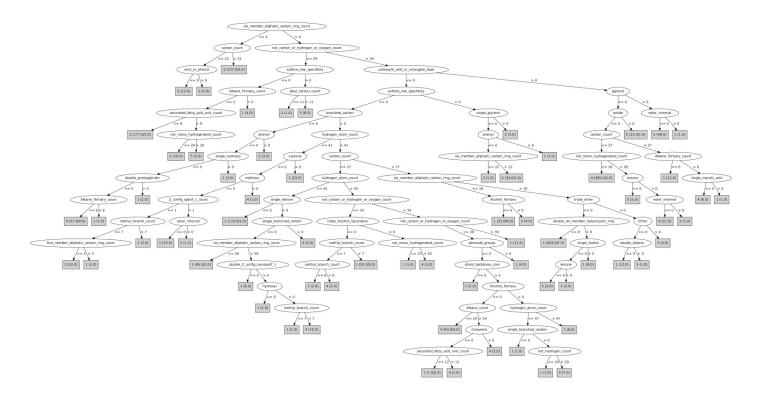


Figure 6: Sterol lipids [ST] decision tree

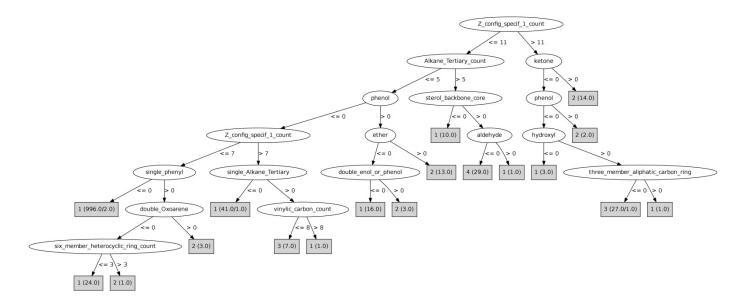


Figure 7: Prenol lipids [PR] decision tree

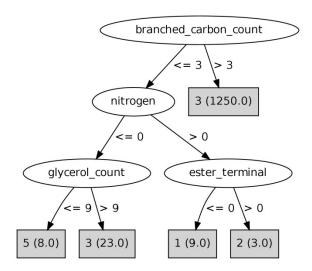


Figure 8: Saccharolipids [SL] decision tree

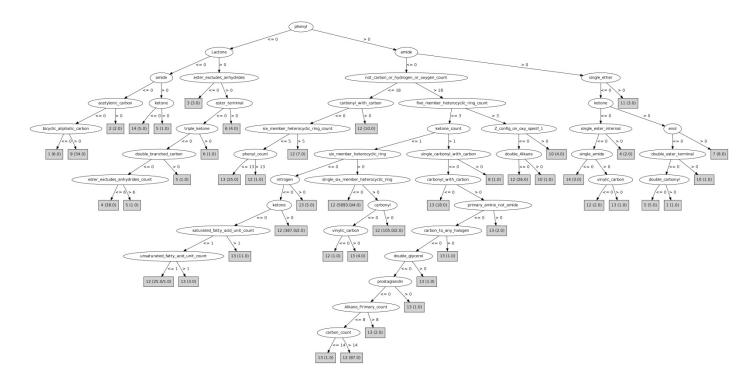


Figure 9: Polyketides [PK] decision tree