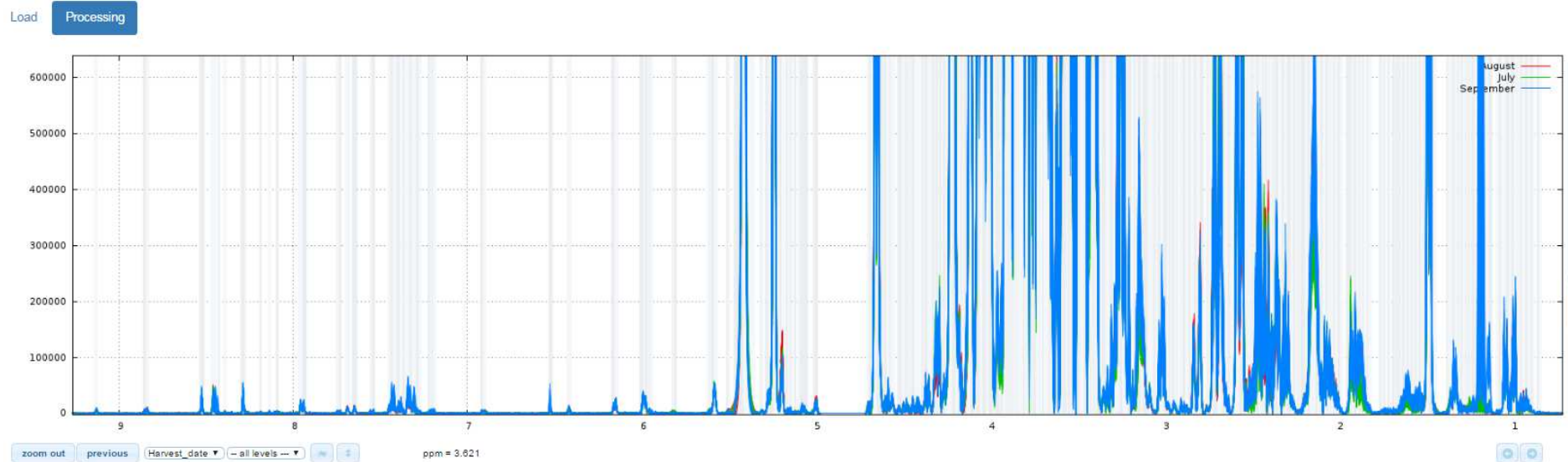


Help in the identification of compounds

Daniel Jacob



Processing Bucketing Data Export

Data Type to Export:

- ☒ Data matrix
- ☐ Buckets table
- ☐ SNR matrix

Export Format:

Semicolon Separator Value (CSV)

SNR threshold:

4

noisy PPM range:

10.5 10.2

Normalization Method:

None

PPM range of the Reference:

10.15 10.07

Export Data Matrix



<http://biostatflow.org/>



BioStatFlow
v.2.7.7 (C) INRA 2015

Dataset Workflow Results **New Session** **Save Session**

Authentication
Logged as **admin**
Logout

Session
S27974

Description
none

Upload your data table (CSV format with either a semicolon or a comma as separator) ?

Study Design - Number of Factors: 2 ▼

Data format: samples in rows ▼

Separator: semicolon (;) ▼

Imported data file (.csv): data_NM...bc7.csv ✓

Association file (.csv) (optional): Aucun fichier choisi

Validate **RESET**

Information about the uploaded dataset

Filename: data_NMRFRIM3-4_bc7.csv (43 rows x 360 columns)
Dataset: 2 Factors, 43 samples, 357 variables
Samples: Samplecode : F3-001 ... F4-093
Factor1: Condition : Control Shadow
Factor2: Stage : J08 J15 J28 J55
Variables: B9_1272 ... B0_8271

Next Step ➡

<http://biostatflow.org/?session=S27974>

A clustering-based workflow applied on the NMR spectra buckets (NMRProcFlow)

Authentication

Register

Log in

Session

S27974

Description

Tomato - FRIM 3 and 4 - FingerPrinting - Cluster Analysis : Reveal of latent variables

Selection

Default Workflow | Proceed from the Dataset : Scaling Reset ?

Clustering of Variables

☒ Launch

Distance Method: Euclidean | Agglomeration Method: Complete

relative distance threshold for variable clustering: ☒ Auto ☐ User

Save Associations: ☒ Yes ☐ No

ANOVA-PCA

☒ Launch

Scale: ☐ Yes ☒ No | Residuals QQ-Plot: ☒ Yes ☐ No

Labels on Score graphs: ☒ Yes ☐ No | Scores within ellipses: ☒ Yes ☐ No | Label size: 2

Use Associations Labels: ☒ Yes ☐ No | Highlighted: ☒ Yes ☐ No

OPLS-DA

☒ Launch

Nb. Comp. Cutoff: auto | Nb. Comp: 2 | permutations for validation: 100

Labels on Score graphs: ☒ Yes ☐ No | Scores within ellipses: ☒ Yes ☐ No | Label size: 2

Use Associations Labels: ☒ Yes ☐ No | Highlighted: ☒ Yes ☐ No

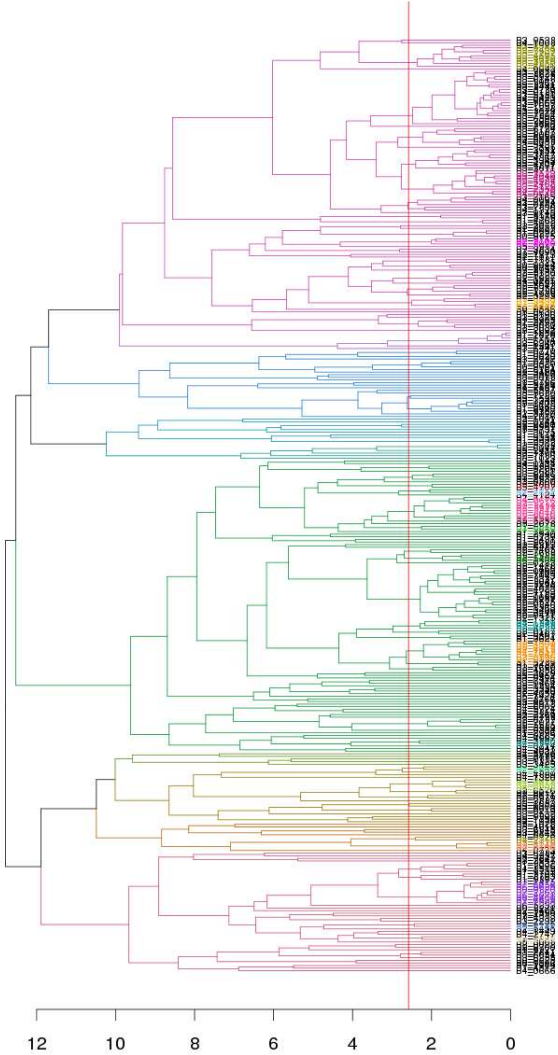
-- Add an analysis --

Launch Analysis

<http://biostatflow.org/?session=S27974>

Cluster Analysis : Highlighting of latent variables

HCA

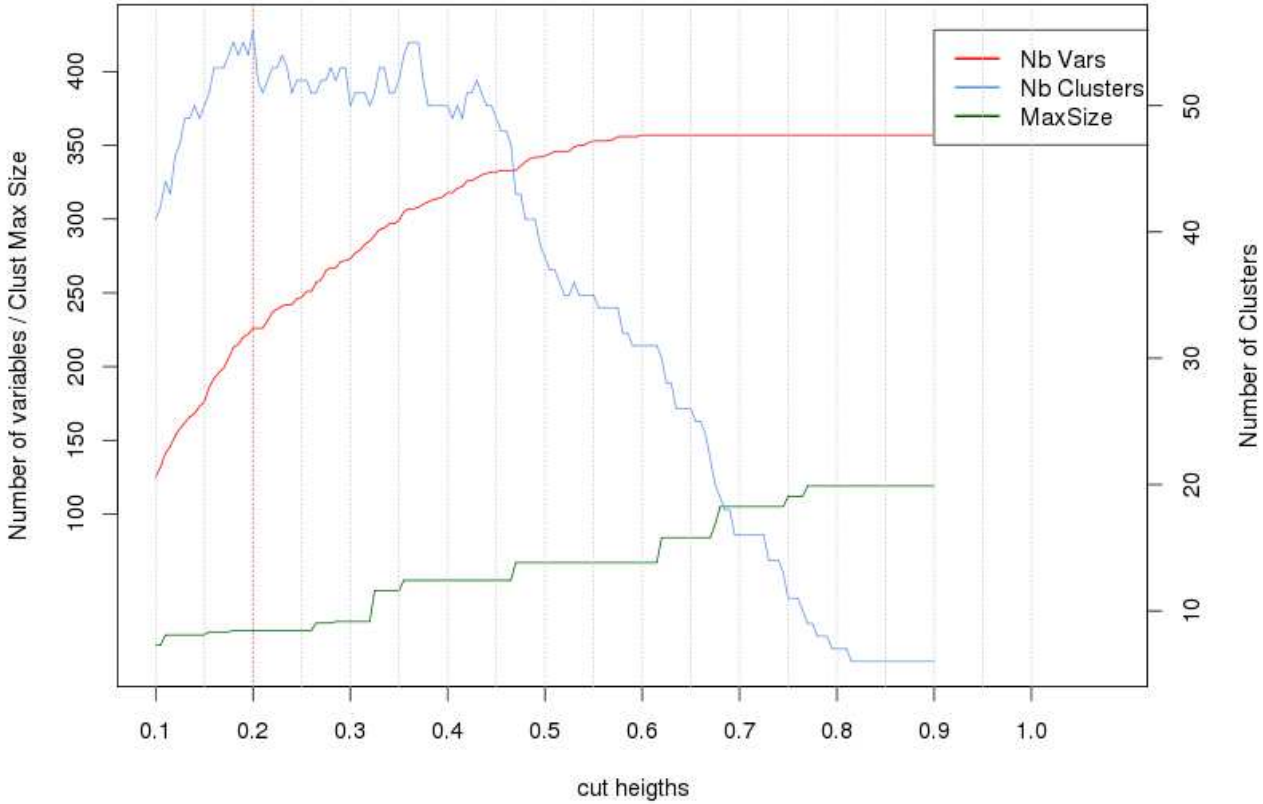


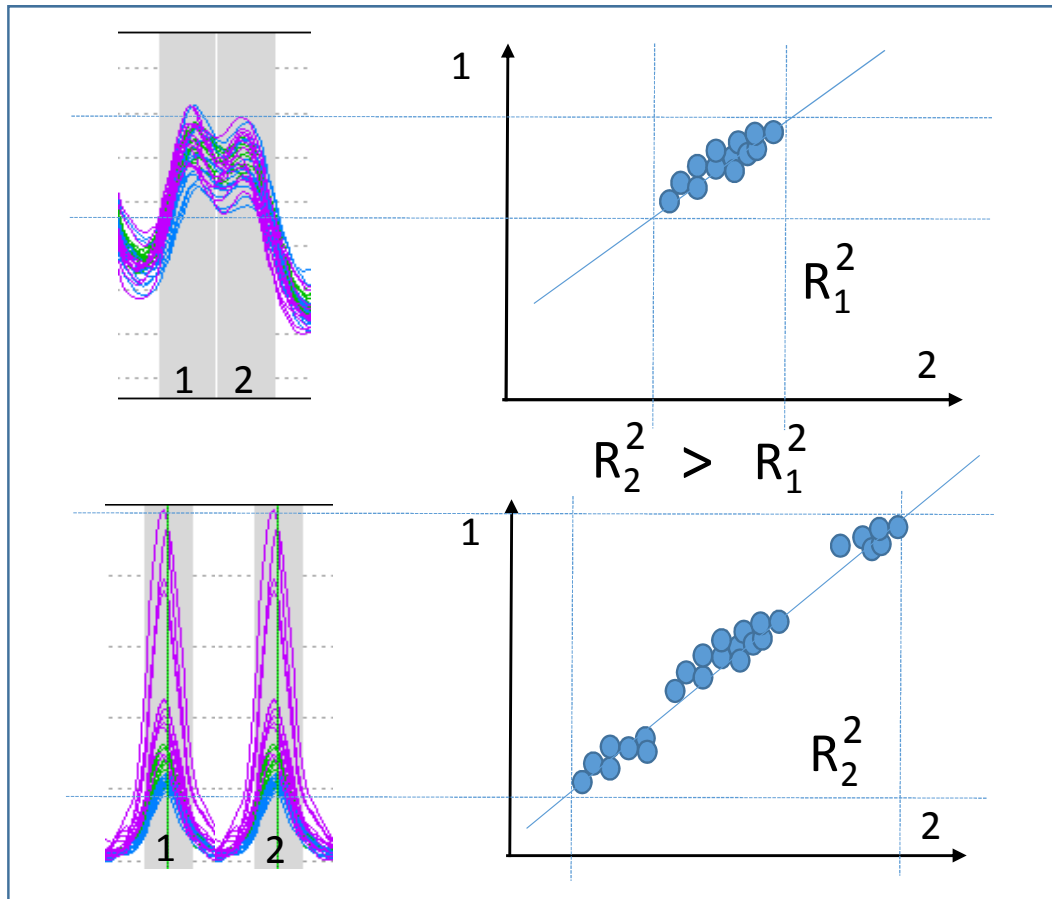
Optimized in order to

Maximize

Minimize

Cuttree = 0.200, Nb Clust = 56, Nb Vars = 226, Clust Max Size = 21





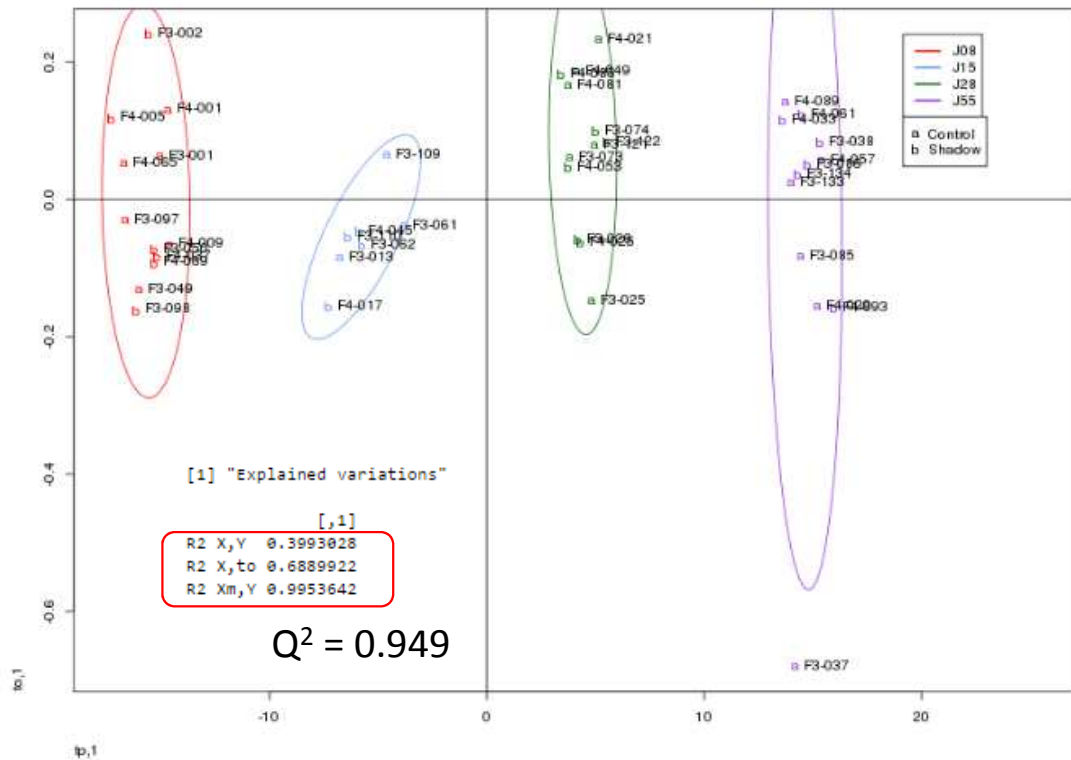
Highlighting biomarkers

- By choosing a good “cut threshold”, clusters link mainly the buckets that have a “**between-groups**” variance,
- Hoping that these “groups” corresponds to factor levels.

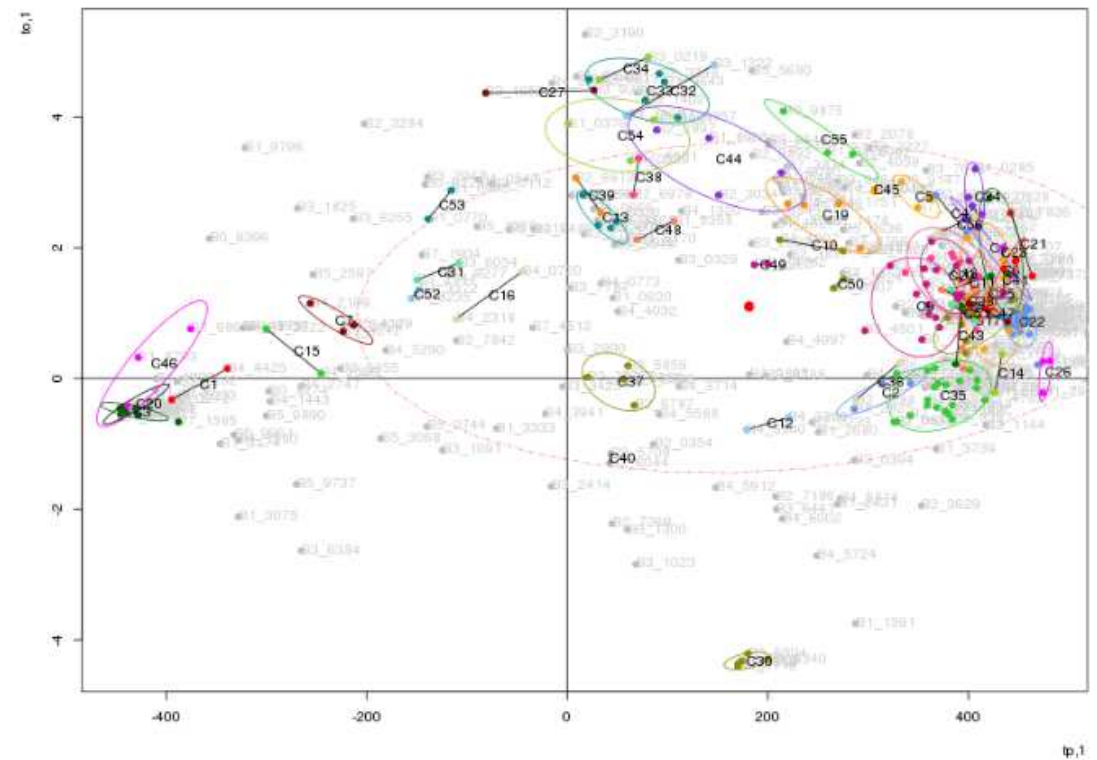


O-PLS-DA

Score Plot : Orthogonal vs. Between groups



Loadings Plot : Orthogonal vs. Between groups



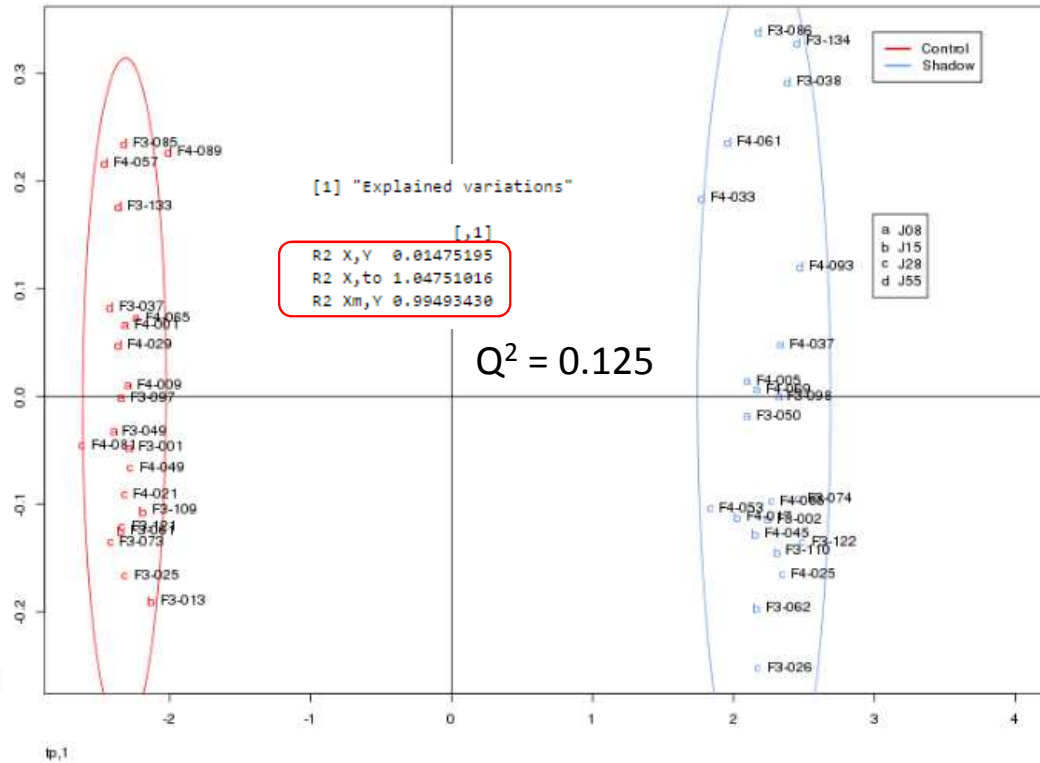
- The "between-groups" variance within clusters seems mostly correspond to this factor



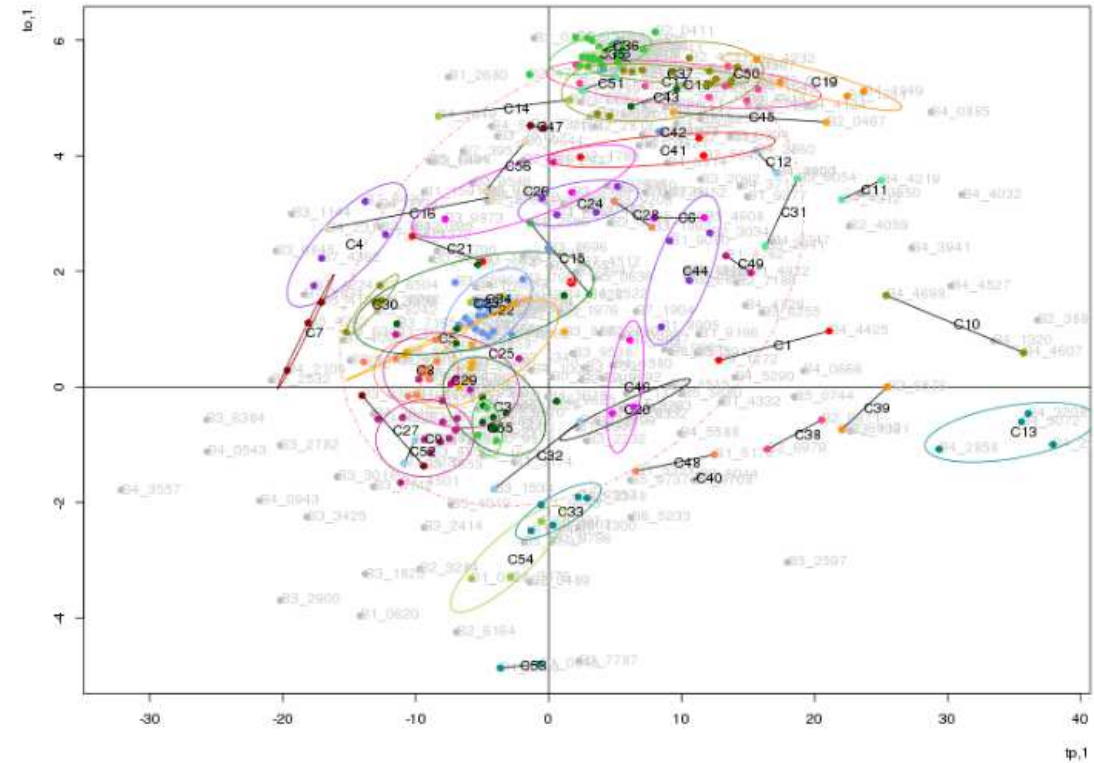
Clusters mainly located at the periphery of a circle => **biomarkers are highlighted**

O-PLS-DA

Score Plot : Orthogonal vs. Between groups



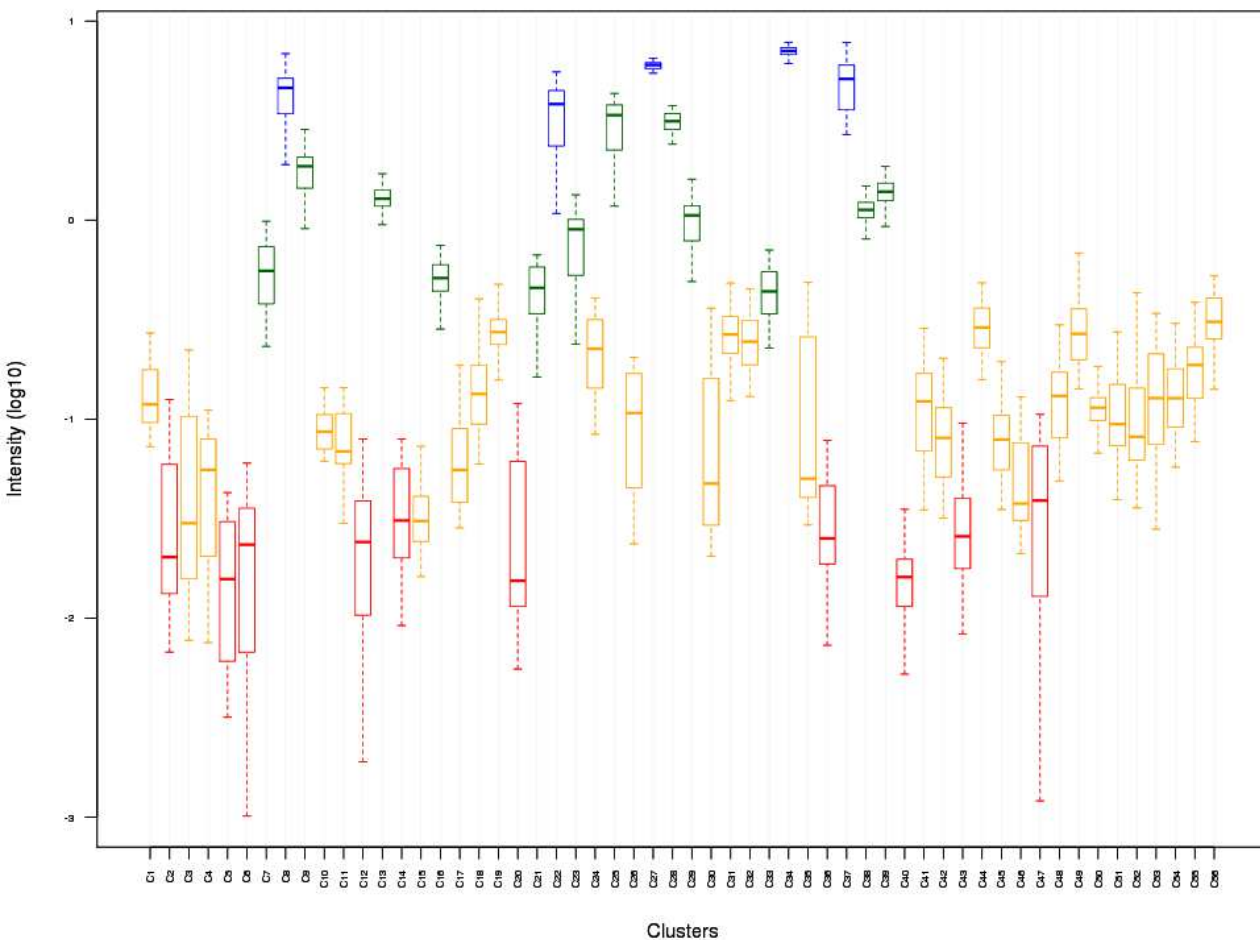
Loadings Plot : Orthogonal vs. Between groups



- The "**between-groups**" variance within clusters seems **NOT** (or very slightly) correspond to this factor

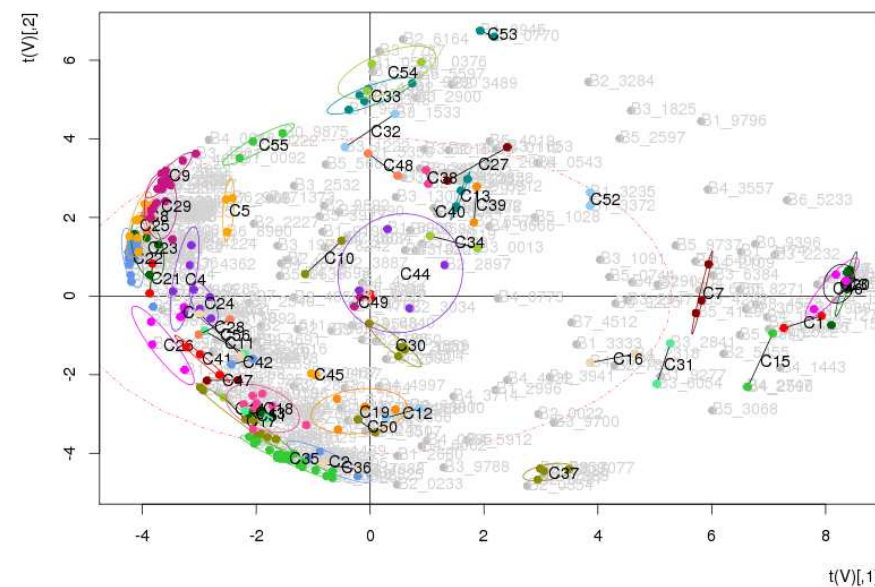


Boxplot by clusters (log10 transformed)



Clusters cover the whole range of intensities
(here we have 4 levels of magnitudes)

MDS - Variables; cutree = 0.200000



Export the Cluster file

BioStatFlow Dataset Workflow Results New Session Save Session

← previous step

Exports : CSV Association 0_dataset 0_factors 5_scaling

Scaling

Clustering of Variables

HCACLUSTR

HCA_Samples

HCA_Variables

MDS_Clusters

MDS_Samples

MDS_Variables

Stats_Clustering

boxclusters

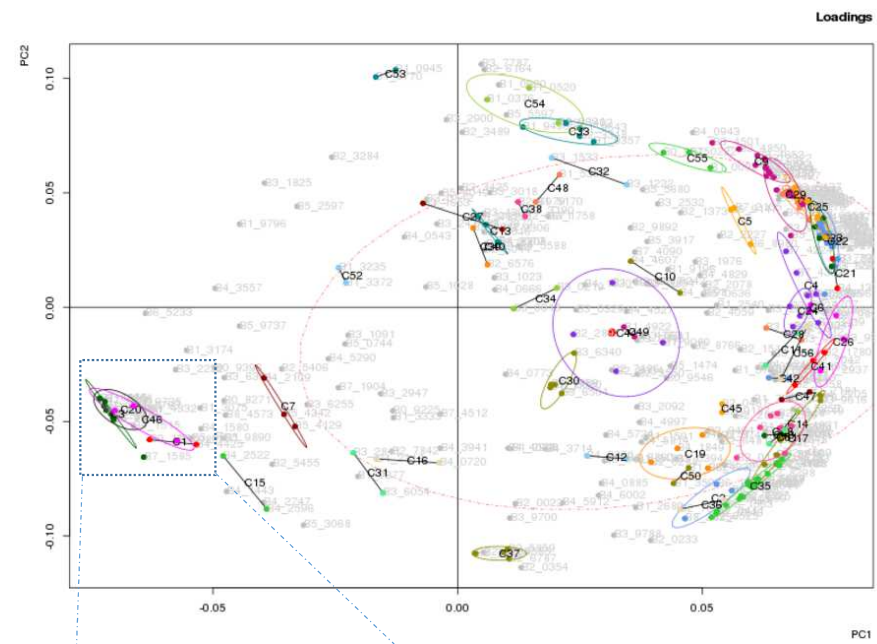
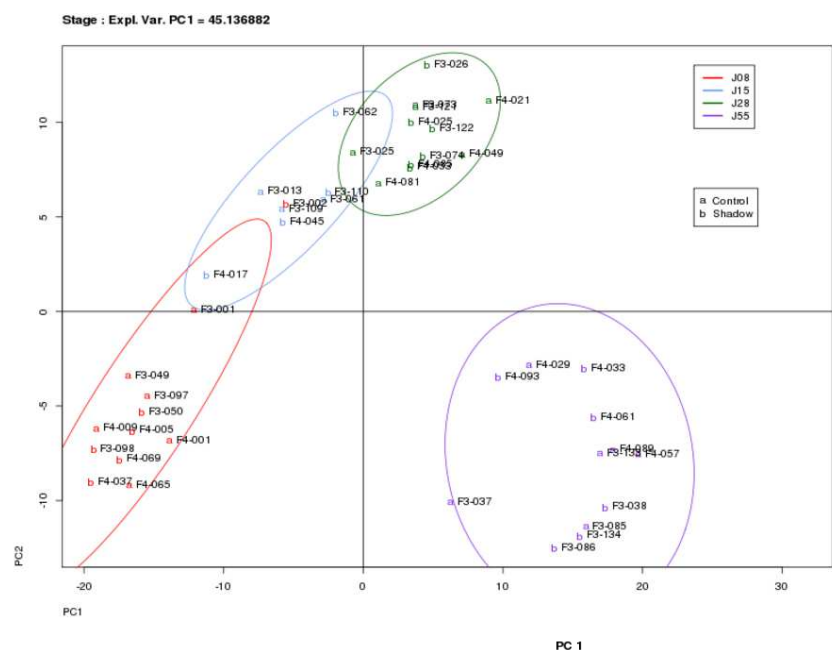
clusters

p6/MDS_Clusters.txt

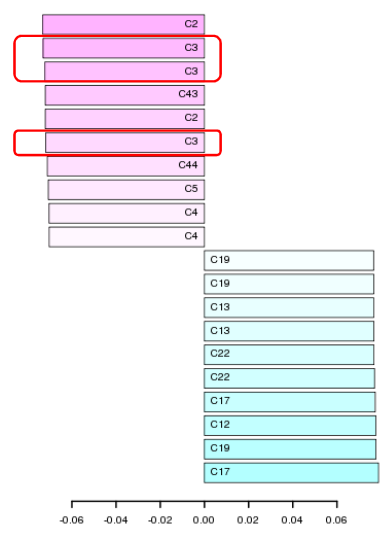
VAR	CLID	PPM
B9_1272	C1	9.1272
B4_4425	C1	4.4425
B8_5408	C2	8.5408
B8_2825	C2	8.2825
B6_1439	C2	6.1439
B7_6956	C3	7.6956
B7_6635	C3	7.6635
B7_2263	C3	7.2263
B7_1585	C3	7.1585
B7_1415	C3	7.1415
B6_9820	C3	6.9820
B6_9625	C3	6.9625
B6_4400	C3	6.4400
B6_4071	C3	6.4071
B7_4362	C4	7.4362
B7_4224	C4	7.4224

Download

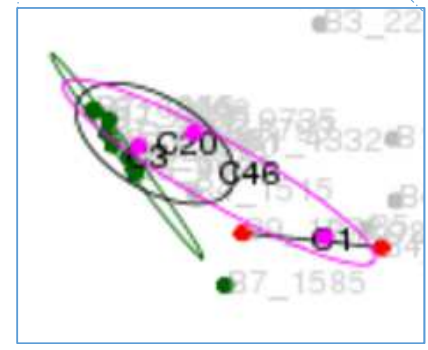
<

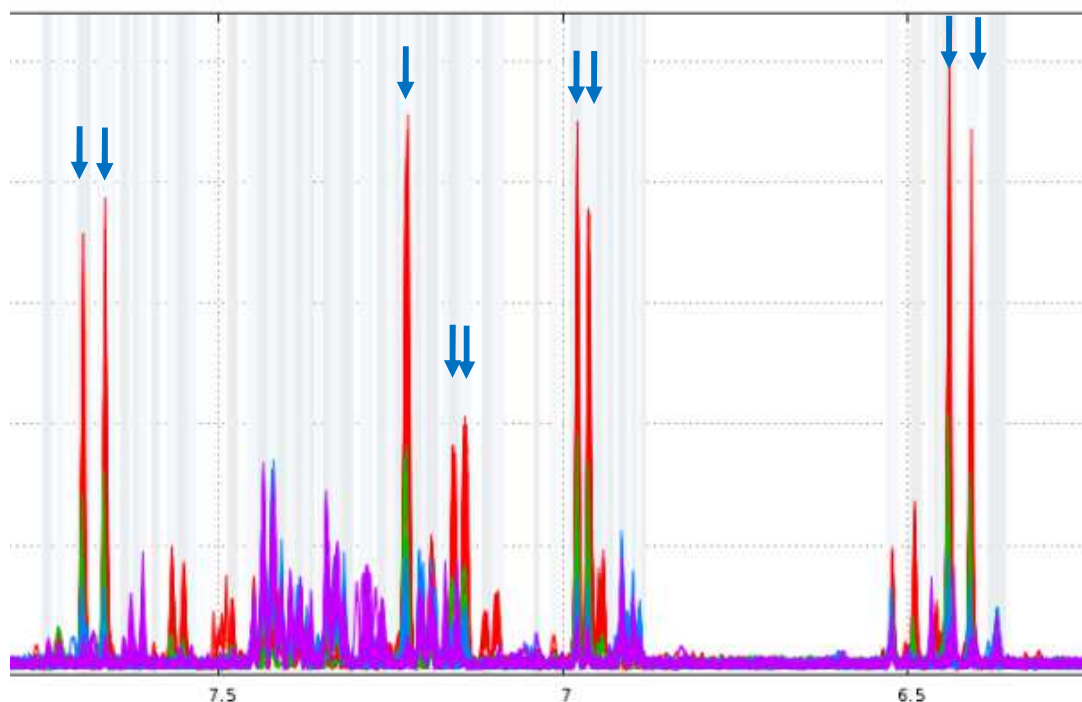


ANOVA-PCA

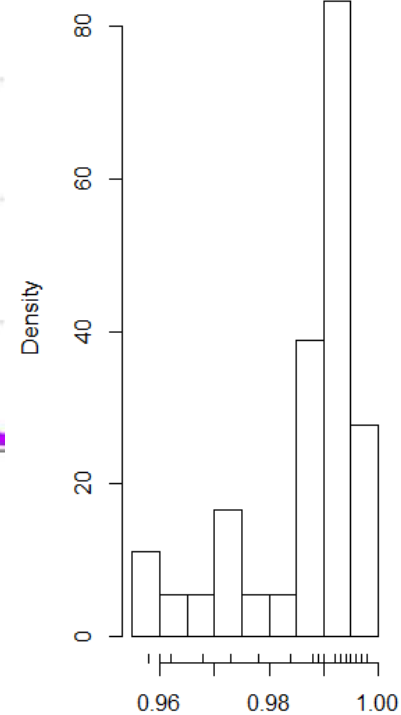


C3



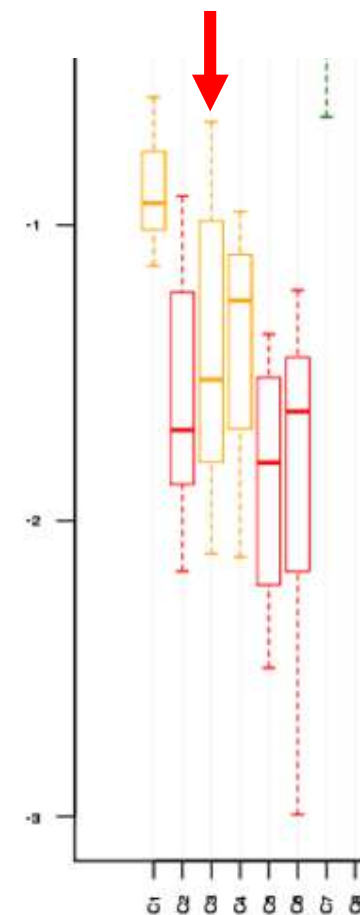


C3



Corr

Intensity (log10)



Cluster C3

	B7_6956	B7_6635	B7_2263	B7_1585	B7_1415	B6_9820	B6_9625	B6_4400	B6_4071
B7_6956	.	0.989	0.994	0.958	0.994	0.992	0.988	0.984	0.99
B7_6635	.	.	0.996	0.96	0.993	0.995	0.988	0.992	0.993
B7_2263	.	.	.	0.968	0.997	0.998	0.993	0.994	0.996
B7_1585	0.962	0.973	0.973	0.978	0.973
B7_1415	0.995	0.989	0.99	0.992
B6_9820	0.994	0.994	0.996
B6_9625	0.988	0.993
B6_4400	0.994
B6_4071



<https://peakforest.org/>



B7_6635	C3	7.6635
B7_2263	C3	7.2263
B7_1585	C3	7.1585
B7_1415	C3	7.1415
B6_9820	C3	6.982
B6_9625	C3	6.9625
B6_4400	C3	6.44
B6_4071	C3	6.4071

From the cluster file, copy-paste the corresponding PPM values to C3 cluster

Peak Matching - NMR

pH 6

Peak list

6.9625
6.44
6.4071

Matching method: ☐ All peaks ☒ At least one peak

¹H tolerance (ppm) 0.02

powered by NMR PeakMatching 1.0 - © INRA UMR 1332 - MetaboHUB

[reset](#) [load demo](#)

[Close](#) [Search](#)

[LC-MS](#)

[NMR](#)

NMR:{"pH":6,"d":0.02,"mm":"one","pl":7.6956,7.6635,7.2263,7.

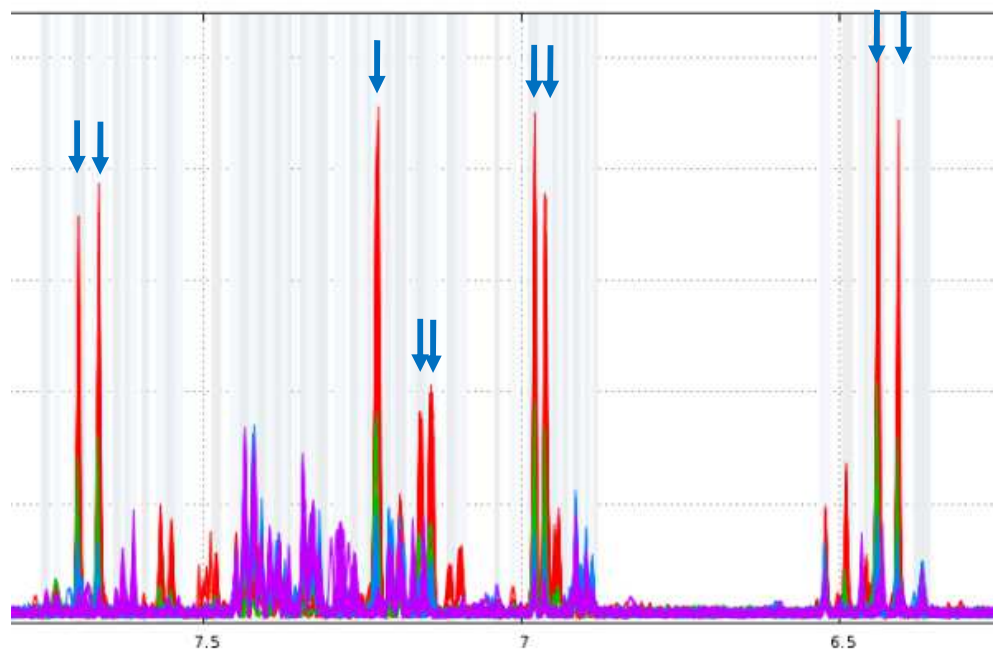
Peak Matching



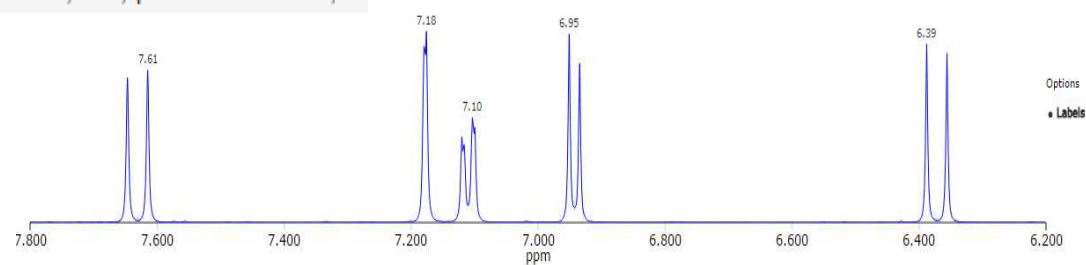
[add filters](#)

#	Name	Score	Preview
PFs000478	Chicoric acid; 6.0; Proton-1D (ZG) -500MHz;	0.4738	
PFs000577	chlorogenic acid; 6.0; Proton-1D (ZG) -500MHz;	0.3189	
PFs000564	Shikimic acid; 6.0; Proton-1D (ZG) -500MHz;	0.0751	
PFs000610	L-Tryptophan; 6.0; Proton-1D (ZG) -500MHz;	0.0648	

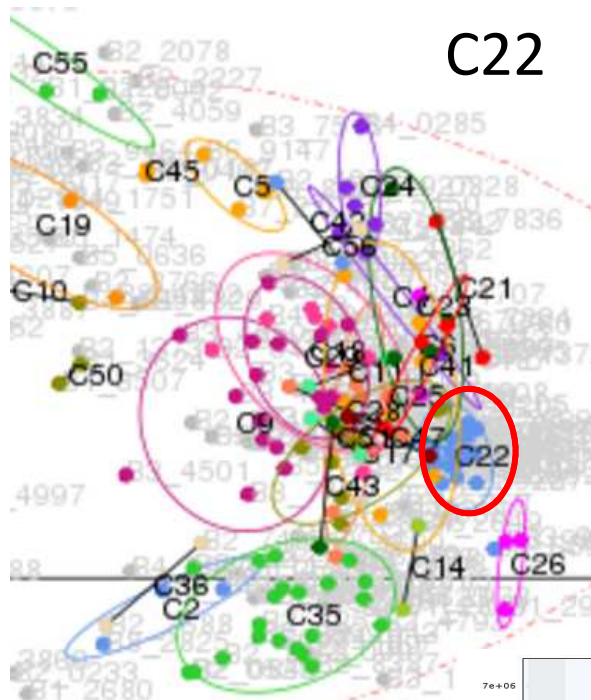
« 1 »



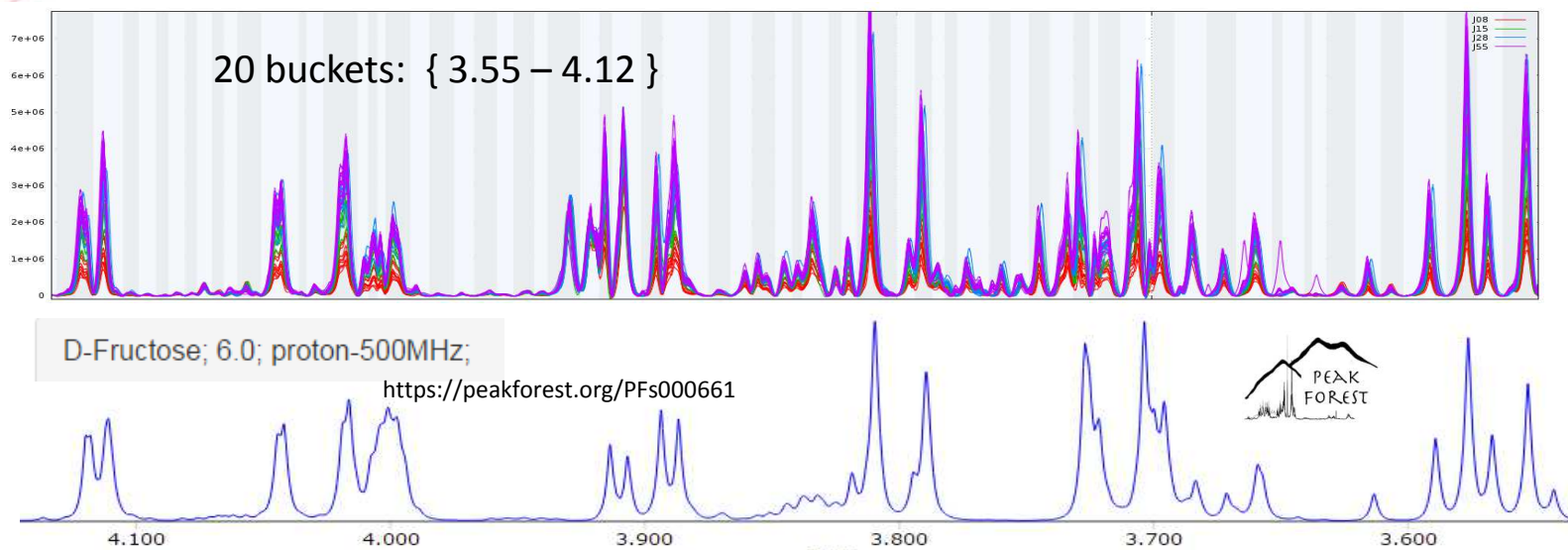
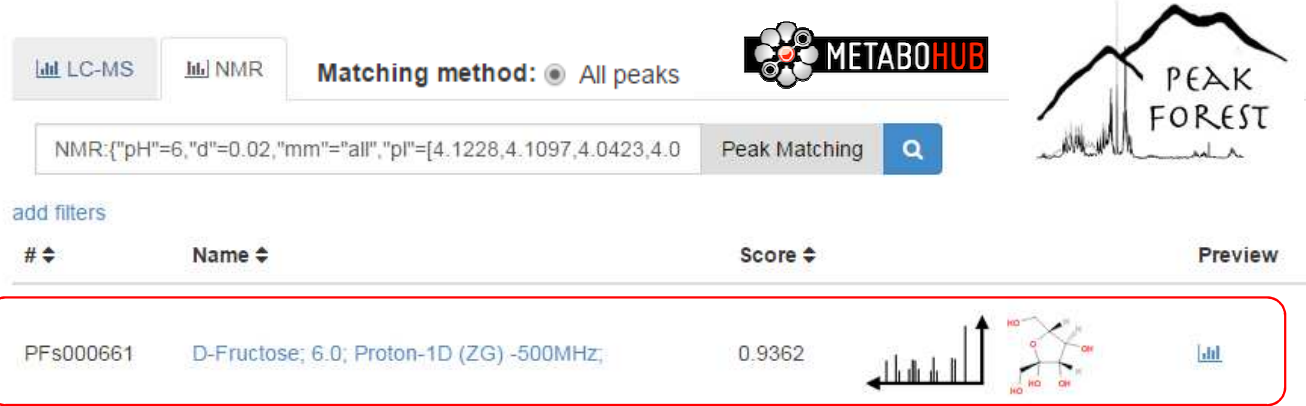
chlorogenic acid; 6.0; proton-500MHz;



<https://peakforest.org/PFs000478>

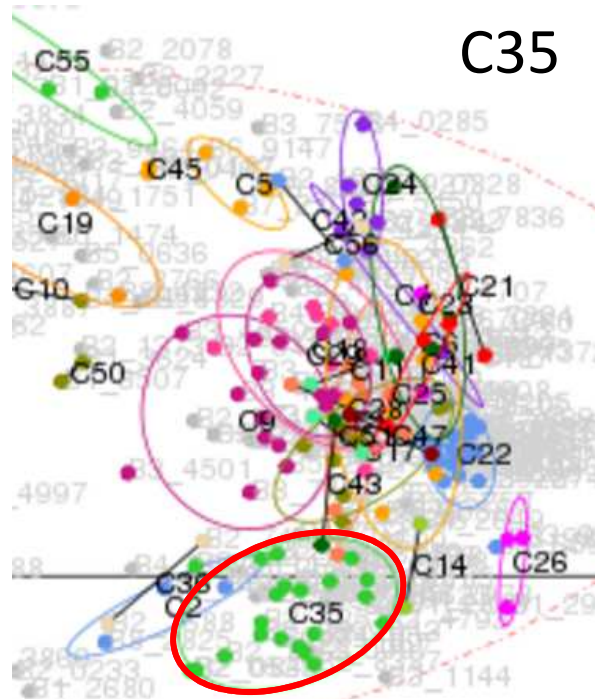


B4_1228	C22	4.1228
B4_1097	C22	4.1097
B4_0423	C22	4.0423
B4_0185	C22	4.0185
B4_0023	C22	4.0023
B3_8962	C22	3.8962
B3_8828	C22	3.8828
B3_8696	C22	3.8696
B3_8444	C22	3.8444
B3_8391	C22	3.8391
B3_8079	C22	3.8079
B3_7965	C22	3.7965
B3_7894	C22	3.7894
B3_7274	C22	3.7274
B3_7072	C22	3.7072
B3_6596	C22	3.6596
B3_6148	C22	3.6148
B3_5925	C22	3.5925
B3_5780	C22	3.578
B3_5669	C22	3.5669
B3_5551	C22	3.5551

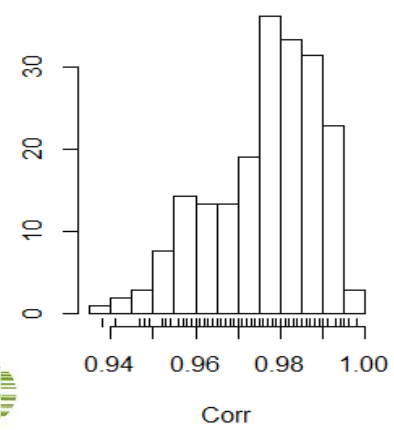


C35

B2_8387	C35	2.8387
B2_8298	C35	2.8298
B2_8041	C35	2.8041
B2_7943	C35	2.7943
B2_7579	C35	2.7579
B2_4793	C35	2.4793
B2_4623	C35	2.4623
B2_3449	C35	2.3449
B2_3409	C35	2.3409
B2_1563	C35	2.1563
B2_1466	C35	2.1466
B2_1418	C35	2.1418
B2_1325	C35	2.1325
B2_1257	C35	2.1257
B2_1185	C35	2.1185
B2_1109	C35	2.1109
B2_0997	C35	2.0997
B2_0846	C35	2.0846
B2_0699	C35	2.0699
B2_0525	C35	2.0525
B2_0411	C35	2.0411



C35

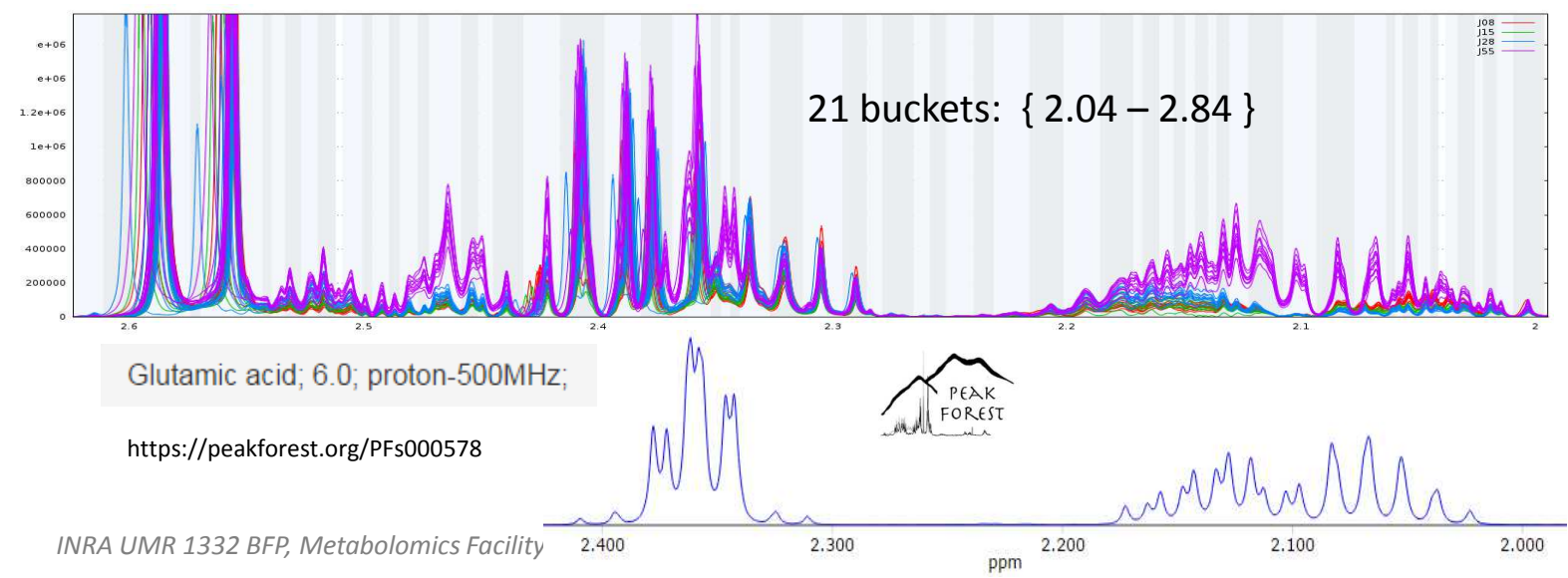


Matching method: ☒ At least one peak

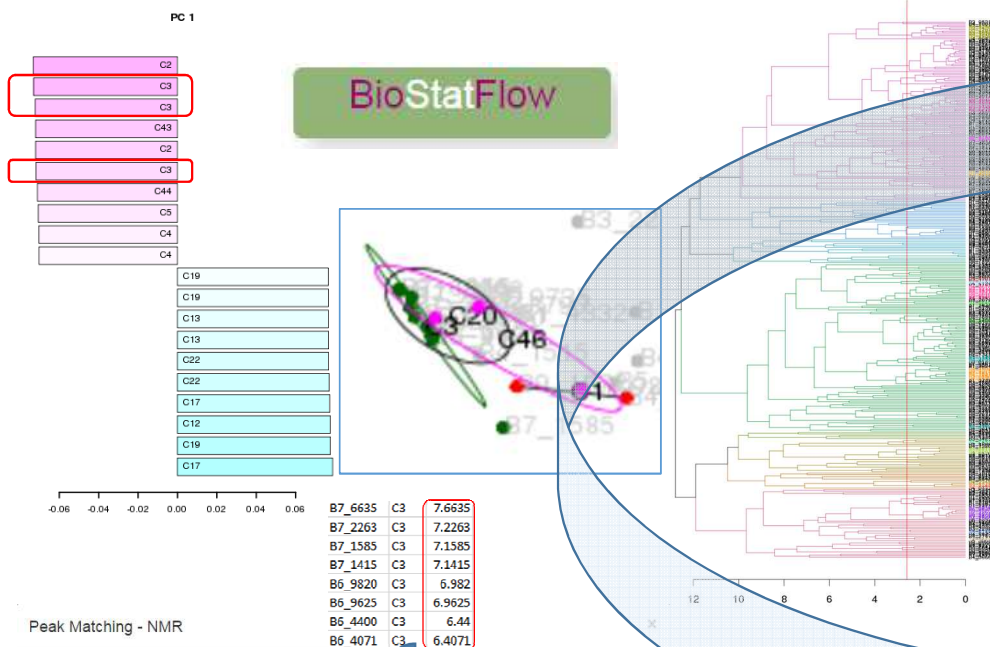
NMR:{"pH":6,"d":0.02,"mm":"one","pl":2.8387,2.8298,2.8041,2.7943,2.7579,2.4793,2.4623,2.3449,2.3409,2.1563,2.1466,2.1418,2.1325,2.1257,2.1185,2.1109,2.0997,2.0846,2.0699,2.0525,2.0411}

Peak Matching

#	Name	Score	Preview
PFs000578	Glutamic acid; 6.0; Proton-1D (ZG) -500MHz;	0.6134	
PFs000664	L-Glutamine; 6.0; Proton-1D (ZG) -500MHz;	0.478	
PFs000579	DL-Proline; 6.0; Proton-1D (ZG) -500MHz;	0.4321	



To summarize ...



Peak Matching - NMR

pH 6

Peak list

6.9625
6.44
6.4071

Matching method: ☐ All peaks ☒ At least one peak

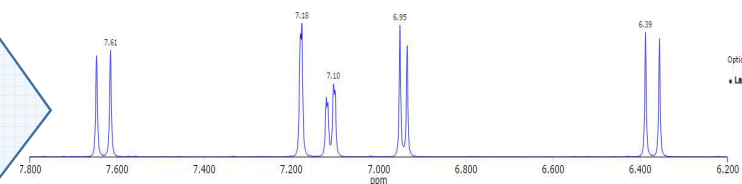
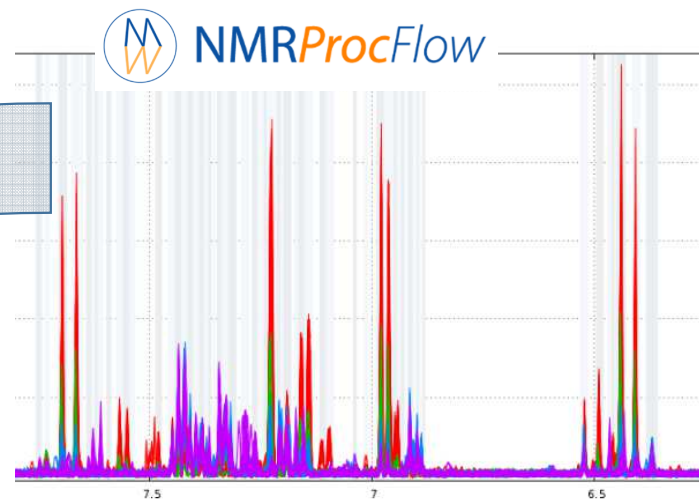
¹H tolerance (ppm) 0.02

powered by NMR PeakMatching 1.0 - © INRA UMR 1332 - MetaboHUB

add filters

#	Name	Score	Preview
PFs000478	Chloric acid, 6.0, Proton-1D (ZG) -500MHz;	0.4738	
PFs000577	chlorogenic acid, 6.0, Proton-1D (ZG) -500MHz;	0.3189	
PFs000564	Shikimic acid, 6.0, Proton-1D (ZG) -500MHz;	0.0751	
PFs000610	L-Tryptophan; 6.0, Proton-1D (ZG) -500MHz;	0.0648	

Manual Peak Matching



<https://peakforest.org/>

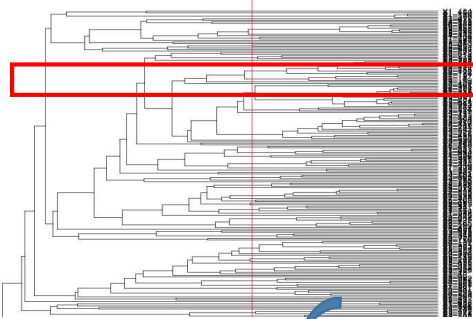


Towards Automatic Peak Matching

Automatic Peak Matching

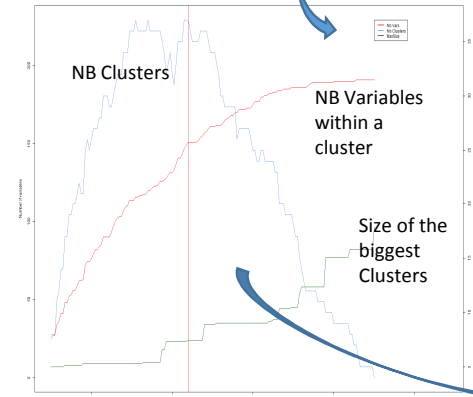
```
#-- Dataset --
# Samples( 54 ): F8_1_1 ... S45_3_2
# Variables( 191 ): X0_932 ... X9_126
# Factor1: tissue - 2 levels: fruit seed
# Factor2: developmental_stage - 5 levels: FF_01 FF_02 FF_03 FF_04 FR_04
```

Hierarchical Cluster Analysis (HCA)



Apply a "cut-tree"

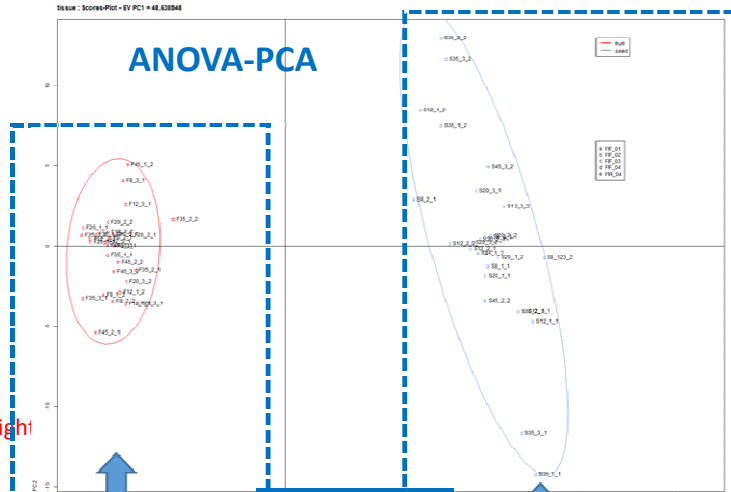
Optimal Cut Height



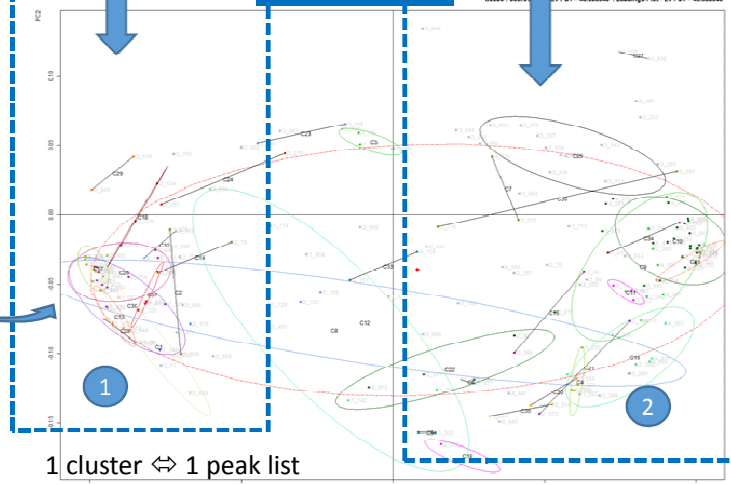
Cut-Height

Size of the biggest Clusters

ANOVA-PCA



Clustering



1 cluster ⇔ 1 peak list

1

Peak Matching

C17(10)	2.436, 2.502, 3.704, 3.733, 3.889, 3.903, 3.921, 4.018, 4.046, 4.118
DBREF0011(8):	Fructose (HMDB00660)(C10906) - Scores=0.6887 - Mean Corr=0.947
C26(9)	3.423, 3.432, 3.444, 3.461, 3.483, 3.754, 4.633, 5.243, 5.272
DBREF0014(8):	Glucose (HMDB00122)(C00267) - Scores=0.7518 - Mean Corr=0.16
C17(3)	5.833, 8.855, 9.126
DBREF0024(5):	Trigonelline (HMDB00878)(C01004) - Scores=0.6391 - Mean Corr=0.978
C19(9)	7.325, 7.344, 7.392, 7.424, 7.432
DBREF0020(10):	Phenylalanine (HMDB00159)(C02077) - Scores=0.7539 - Mean Corr=0.812
C12(5)	2.132, 2.444, 2.455, 2.471, 2.486
DBREF0047(5):	Glutamine (HMDB00641)(C00303) - Scores=0.7508 - Mean Corr=0.934
C1(4)	0.932, 0.993, 1.042, 1.06
DBREF0027(3):	Valine (HMDB00002)(C16426) - Scores=0.5103 - Mean Corr=0.697
C10(24)	1.875, 1.894, 1.91, 1.926, 2.283, 2.311, 2.327, 3, 3.015, 3.036, 4.235, 4.283, 4.283, 5.595, 6.403, 6.411, 6.526, 6.561, 6.586, 7.132, 7.225, 7.656, 7.658, 7.964
DBREF0013(10):	GABA (HMDB00112)(C00334) - Scores=0.3806 - Mean Corr=0.948
DBREF0040(4):	Chlorogenic acid (HMDB003164)(C00502) - Scores=0.5125 - Mean Corr=0.092
C31(5)	4.06, 4.216, 4.232, 5.416, 5.424
DBREF0019(8):	Sucrose (HMDB00288)(C00089) - Scores=0.7737 - Mean Corr=0.91
C9(8)	1.85, 2.005, 2.017, 2.029, 2.042, 2.057, 3.633, 4.086
DBREF0000(5):	Chlorogenic acid (HMDB003164)(C00502) - Scores=0.5125 - Mean Corr=0.092
C11(3)	2.072, 2.087, 6.14
DBREF0021(2):	Proline (HMDB00162)(C16435) - Scores=0.3771 - Mean Corr=0.984
C16(7)	2.388, 2.395, 2.418, 4.292, 4.301, 4.311, 4.33
DBREF0017(6):	Malic acid (HMDB00744)(C00711) - Scores=0.6095 - Mean Corr=0.910
C4(5)	1.255, 1.272, 1.287, 1.304, 1.321
DBREF0015(5):	Isoleucine (HMDB00172)(C16434) - Scores=0.7831 - Mean Corr=0.876

Clustering

Distance Method: euclidean
Agglomeration Method: average
Cutting Tree threshold: 0.44

Peak Matching

Nb CMPD in DBREF: 82
Total Cluster: 37
Nb Cluster with Size >= 2 : 37
Clustered variables: 151 / 191
Nb Cluster with matching: 36
Chem. Shift Tol.: 0.02
Limit Result size: 3



<https://peakforest.org/>



May 20

