Untargeted LC-MS analysis of the raspberry transnational trial

Sean Conner

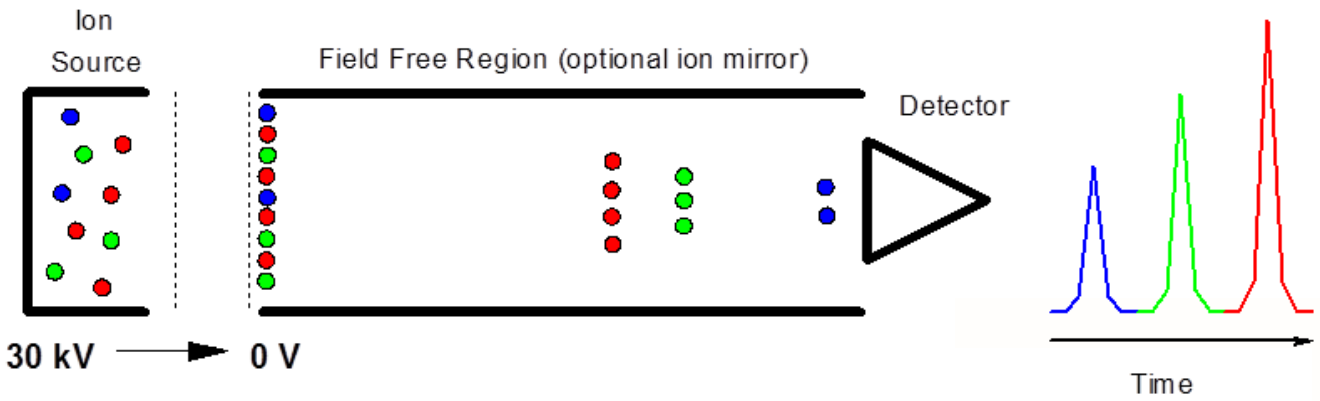
The James Hutton Institute

Future-proofing berryfruit CLIMAFRUIT

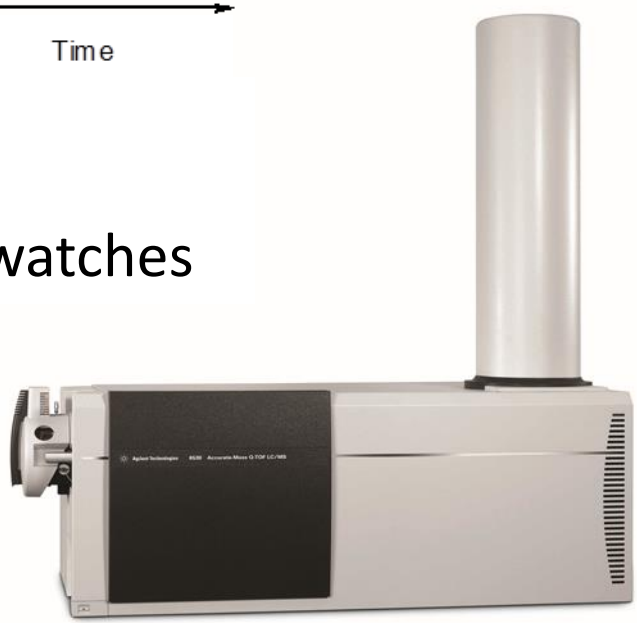
The Interreg IVB North Sea Region Programme

funding by Interreg IVB as part of the NSR programme: (Climafruit)Project No. 35-2-05-09)

Time-of-Flight Mass Spectrometry

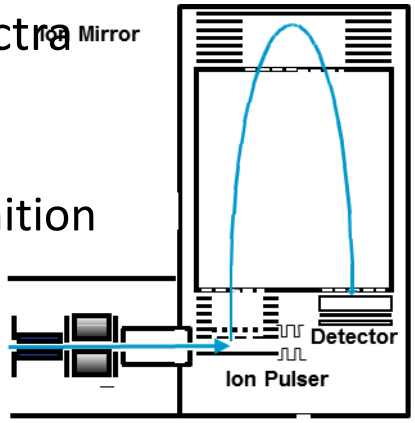


TOF Analysers are VERY accurate stopwatches

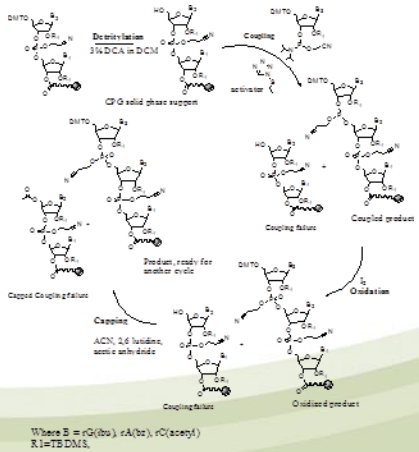


Time-of-Flight Mass Spectrometry

- Non Scanning Instrument collecting full spectra
  - 20,000 m/z mass range,
  - Up to 20 spectra/sec for accurate LC definition



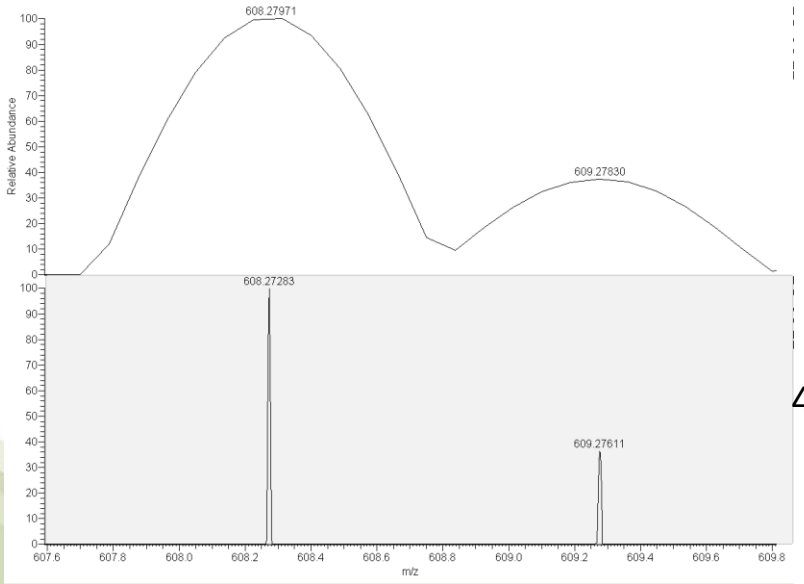
- High Resolution MS data (40,000FWHM)
  - Provides accurate mass measurements (1ppm)
- High Sensitivity
  - Provides ability to detect low level compounds
- Wide Detection Dynamic Range
  - 4-5 orders



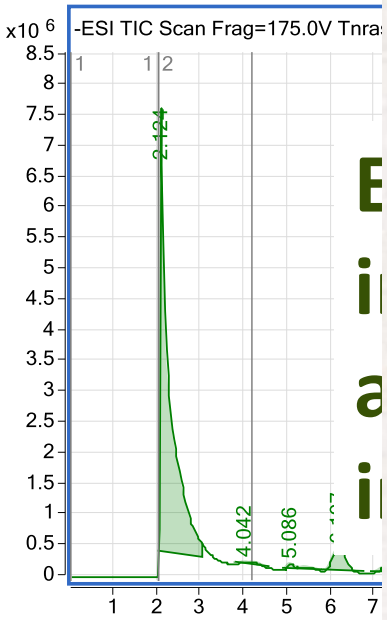
The Advantage of Accurate Mass Measurement - *increased specificity*

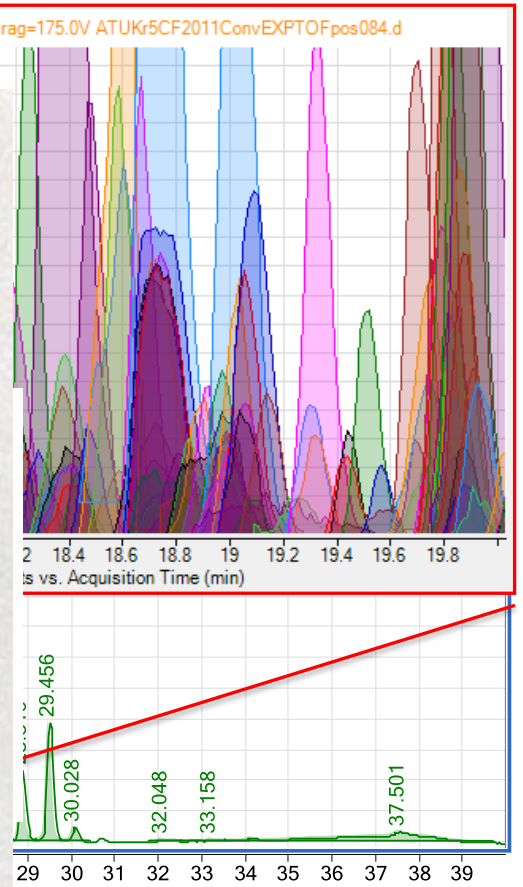
- $C_{33}H_{40}N_2O_9$ has a protonated ion at 609.28066
- Quadrupole MS reports mass to +/- 0.1Da = 165 ppm
- High Resolution MS reports to <2ppm

- Possible Formulas (C,H,N,O)
  - 165 ppm 209
  - 10 ppm 13
  - 5 ppm 7
  - 3 ppm 4
  - 2 ppm 2



Finding the needle in a haystack
Dealing





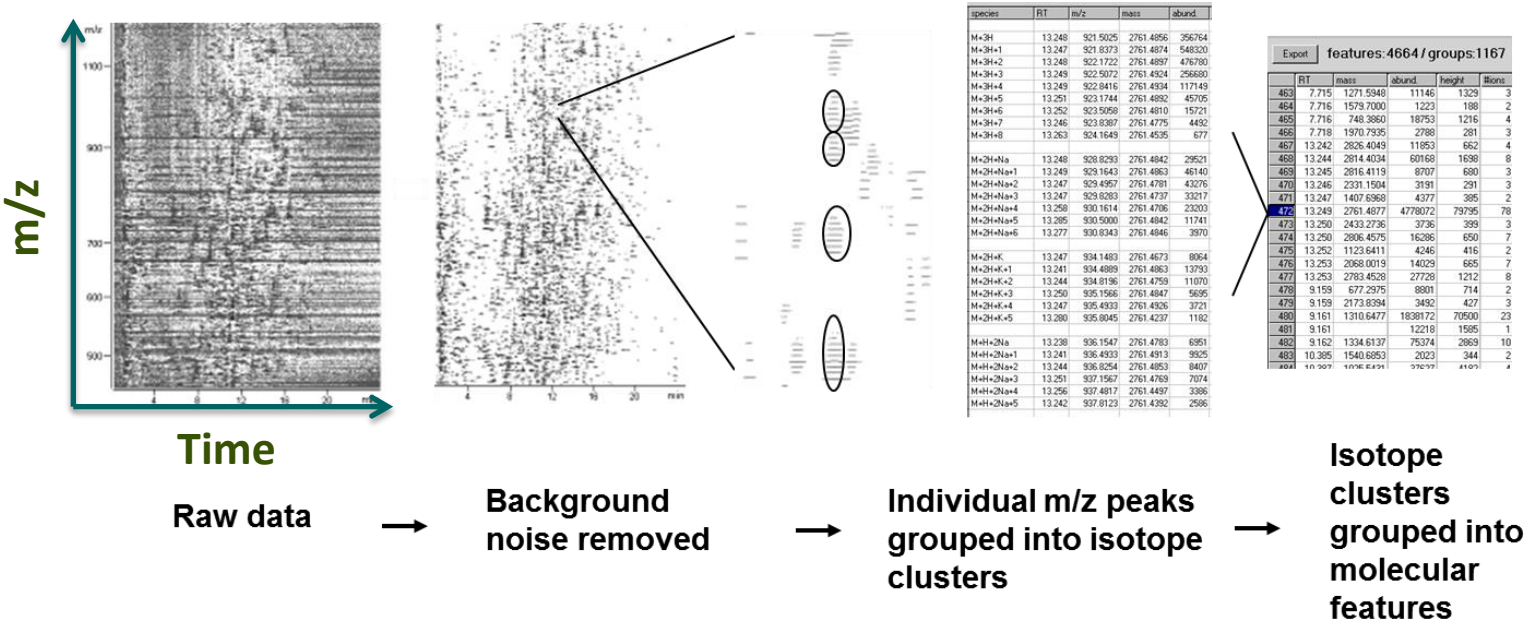
Hunts Needle in a Haystack

HOW LONG does it take to find a needle in a haystack? Jim Moran, Washington, D. C., publicity man, recently dropped a needle into a convenient pile of hay, hopped in after it, and began an intensive search for (a) some publicity and (b) the needle. Having found the former, Moran abandoned the needle hunt.

Time consuming

- 300000 peaks in total –
- Assuming 15 seconds per peak (average) to check re-integrate if necessary 4500000s, 75000min, 125h, 52.1days or assuming 8 hour working day 156.3 working days.
- Clearly faster way to mine data is required

Mass Profiler – Find by molecular feature



Raw data → **Background noise removed** → **Individual m/z peaks grouped into isotope clusters** → **Isotope clusters grouped into molecular features**

Identification, Quantification, Differential Analysis are performed on *chemically qualified compound data* - automatically

Raspberry TN - Experimental design

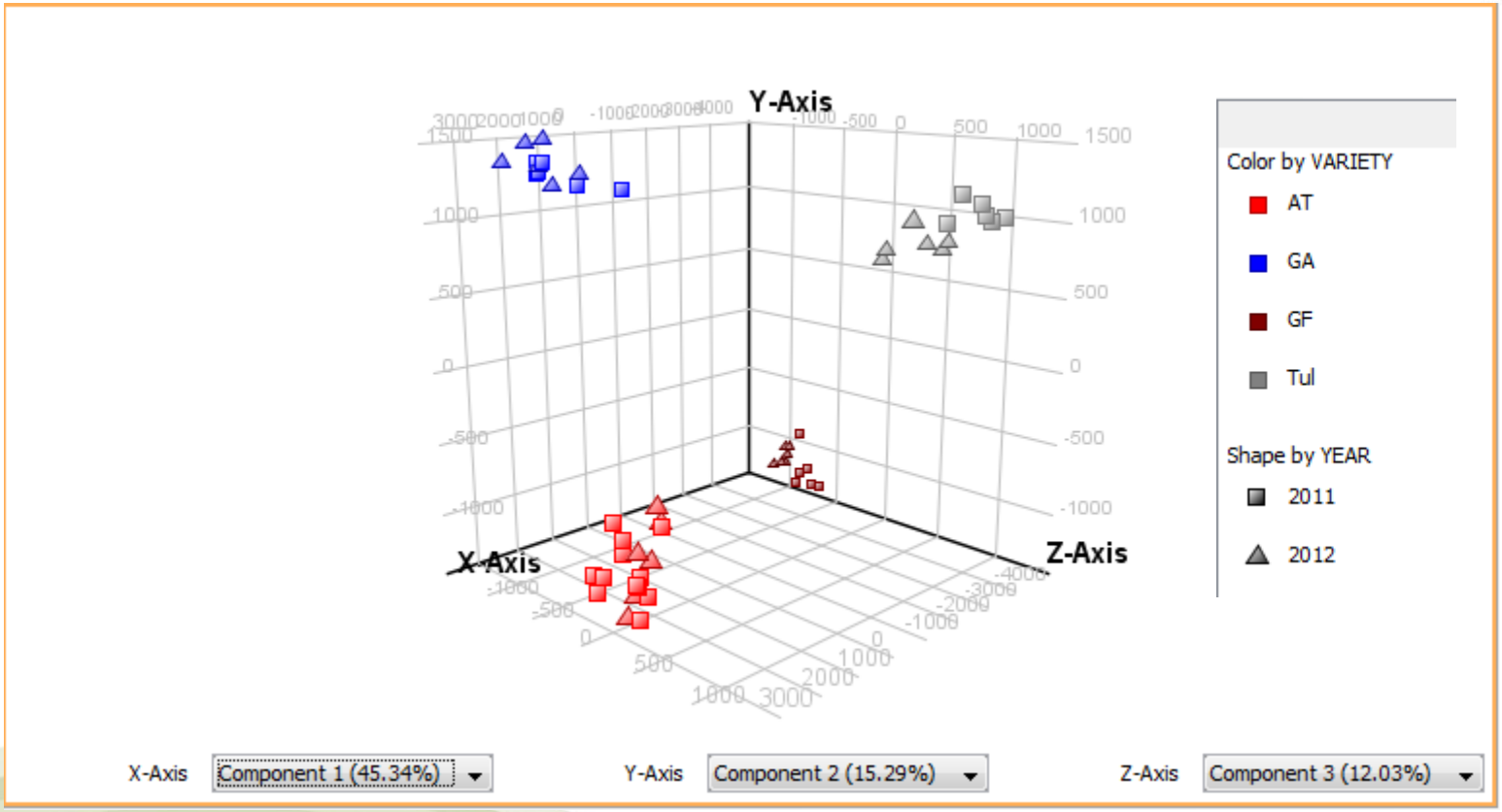
- Randomised block design
- Years 2011, 2012
- Conventional /(Organic)
- UK, Norway, Sweden, Germany, Denmark
- Autumn Treasure, Glen Ample, Erika, Fall Gold, Sugana, Glen Doll, Cascade Delight, Cowichane, Malling Hestia, Glen Magna, Tulameen, Polka, Glen Rosa

All Cultivars UK – Positive mode

The James Hutton Institute

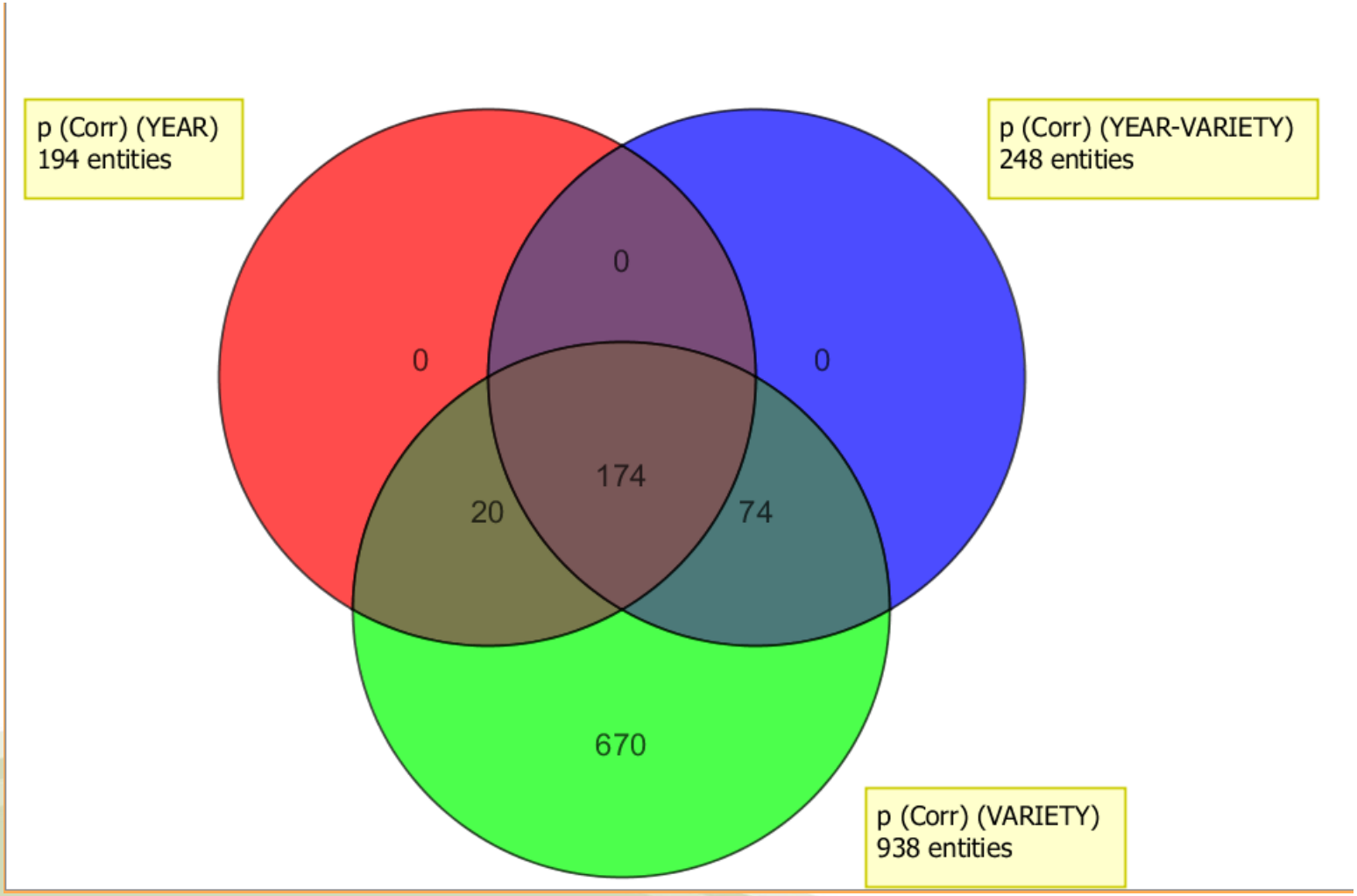
PCA Analysis showing variability across raspberry varieties

PCA analysis indicates that there is more difference between varieties than there is between years. Therefore, within the UK, harvest year is not significant for these compounds.



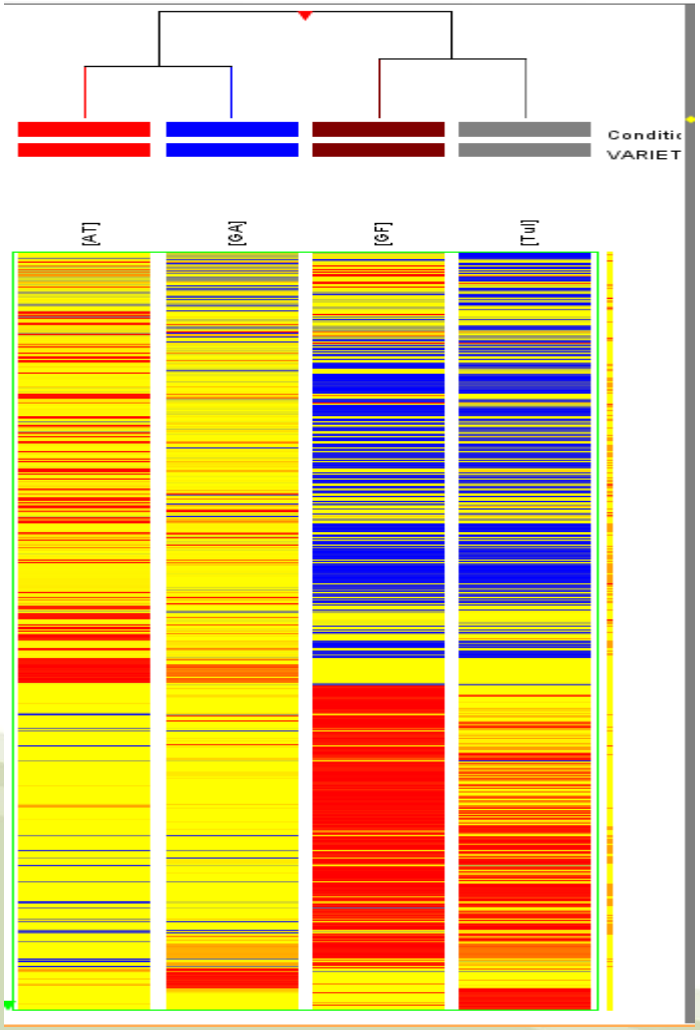
Significance Analysis – 2-way ANOVA (p <0.05)

This Venn Diagram supports the PCA in that differences between the samples were due to the raspberry variety and not the year of harvest.



Hierarchical Clustering

In this heat-map each line represents an entity while the colour indicates abundance of that compound. This dendrogram indicates similarities between AT and GA and GF and Tul while highlighting that AT and GF are more different from GF and Tul than each pair is from each other.

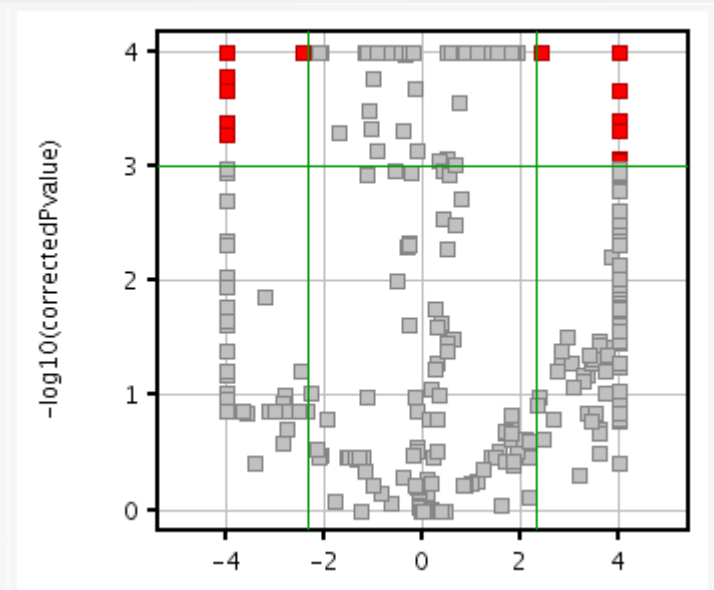


Volcano Plot

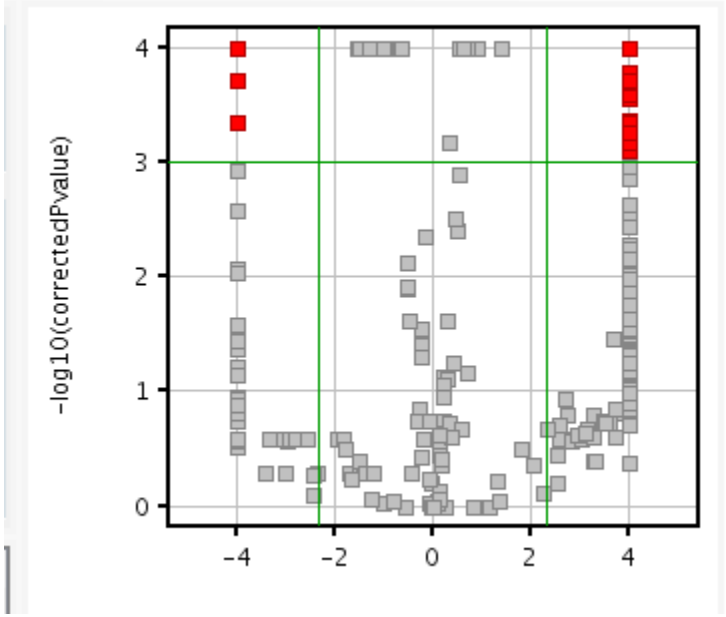
Within the more similar pairs a volcano plot can be used to show those entities (red) which vary by a specified fold-change (5 x here) and a specified statistical significance (p<0.001 here). Therefore the red compounds are 5x more or less abundant in one variety vs the other with a significant of p<0.001.

The James Hutton Institute

Volcano Plot AT vs GA



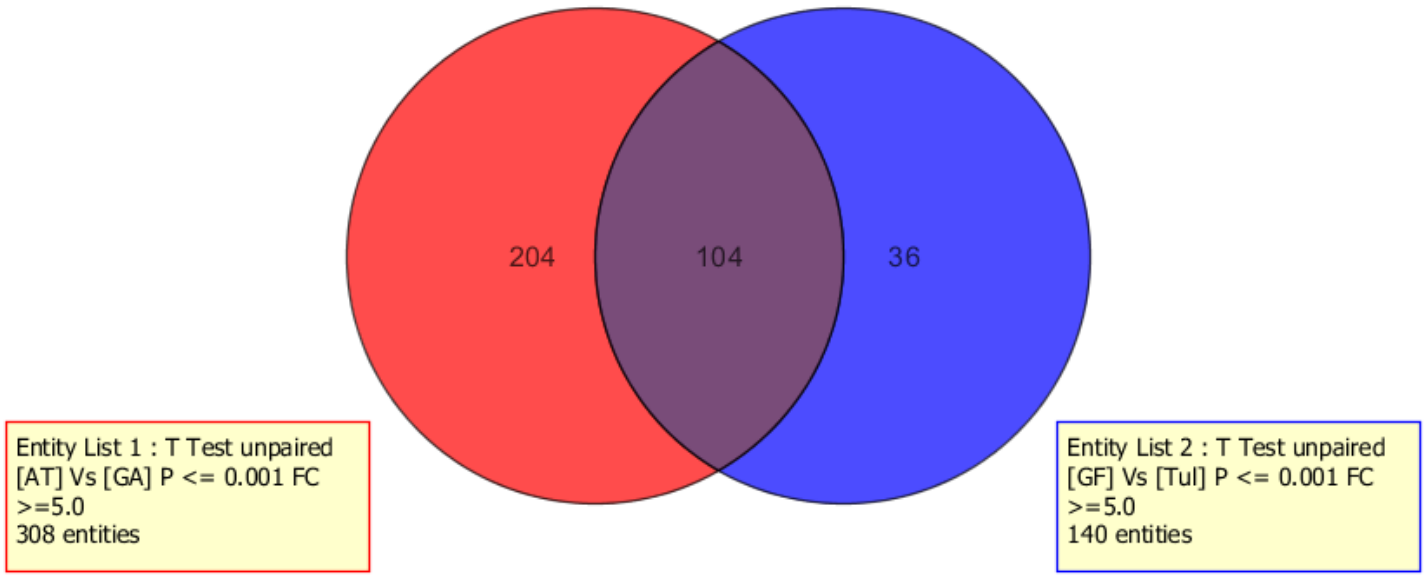
Volcano Plot GF vs Tul



Venn Diagram – Comparison of entities from Volcano plot (t-test) of two pairs of varieties.

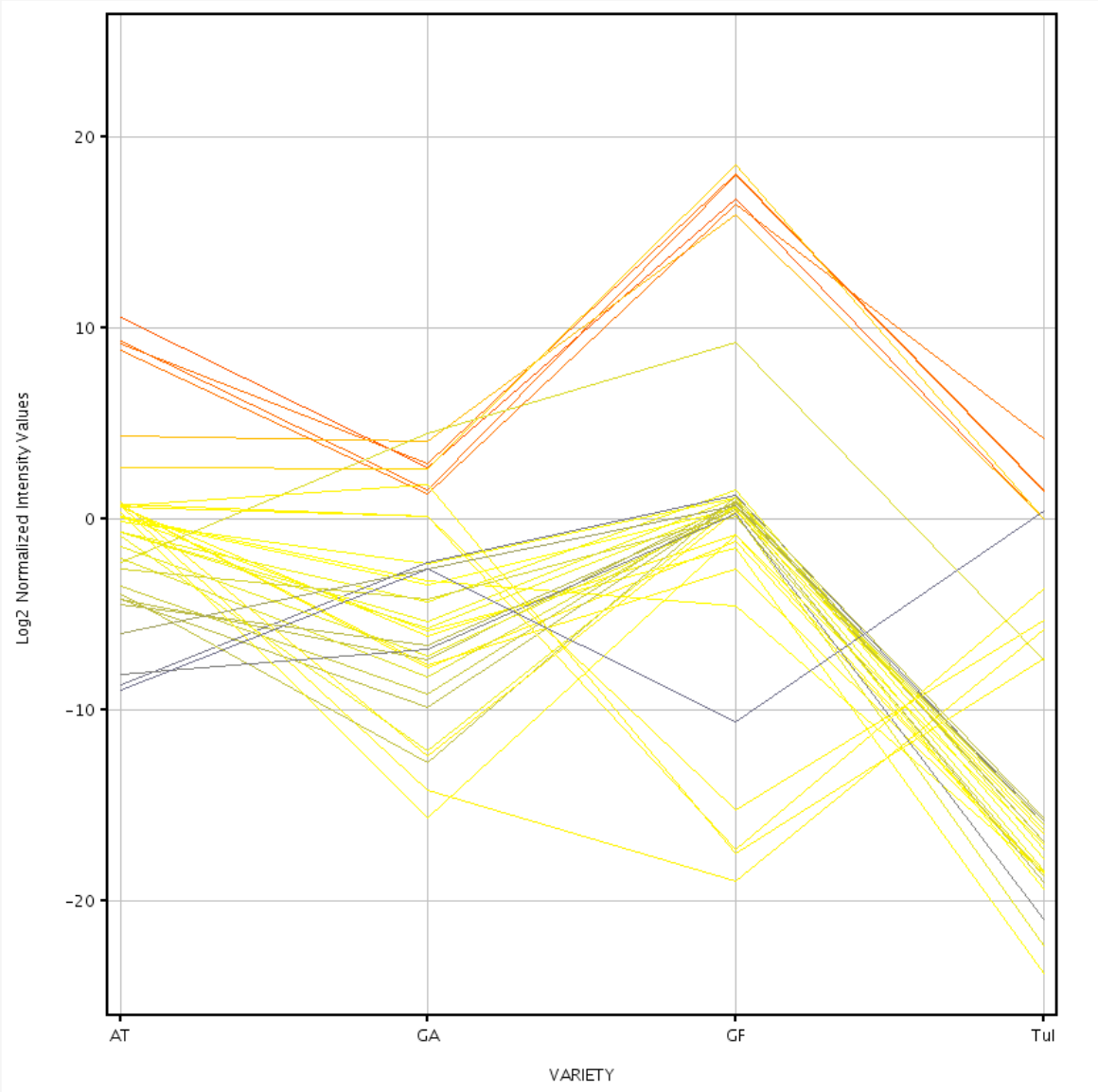
The James Hutton Institute

The entities identified from the volcano plots of t-testing of two pairs of varieties are examined using a Venn Diagram. This indicates that the same 104 compounds are 5x different in each pair of varieties. 36 compounds are 5x different only in the GF and Tul varieties.



Profile Plot of Entities

The 36 entities are shown here on a profile plot across all four investigated varieties. .



Identification

The entity list can be searched against a database (METLIN) and also used for formula generation

The James Hutton

Compound List

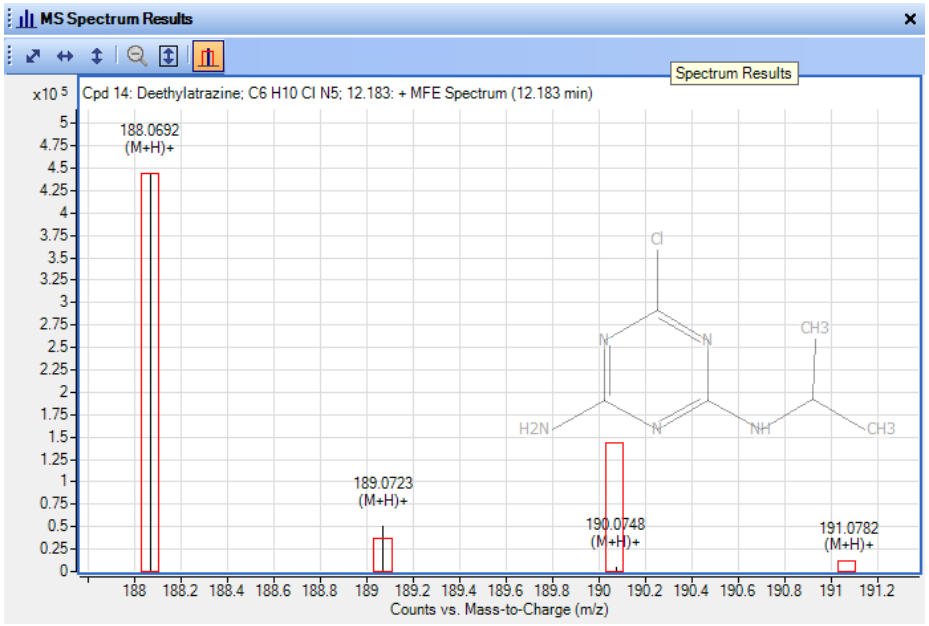
Cpd	Name	Notes	Diff (DB, ppm)	Hits (DB)	LMP	Formula	Score	Mass	RT	Mass (DB)	Mass (MFG)
1	Vitexin 2"-p-hydroxybenzoate		-3.13	3	LMPK12110225	C28 H24 O12	60.91	552.1285	2.675	552.1268	
25	Luteolin 7-galactoside-4'-glucoside		1.04	5	LMPK12110670	C27 H30 O16	98.92	610.1528	15.564	610.1534	
28	Iridodial glucoside tetraacetate		1.76	2		C24 H34 O11	75.61	498.2092	16.793	498.2101	
17	Indole	Group A 07/27/04	-3.07	3		C8 H7 N	82.49	117.0582	12.165	117.0578	
36	His His Arg		-1.31	3		C18 H28 N10 O4	84.55	448.2301	25.195	448.2295	
34	Glucosylgalactosyl hydroxylysine		0.64	2		C18 H34 N2 O13	86.09	486.2058	18.492	486.2061	
23	Disperse Blue 1		2.26	2		C14 H12 N4 O2	85.14	268.0954	14.976	268.096	
14	Deethylatrazine		1.65	1		C6 H10 Cl N5	68.39	187.0622	12.183	187.0625	
26	Asn-Trp-OH		0.51	5		C20 H18 N4 O7	98.46	426.1173	16.162	426.1175	
33	Asiatic acid		1.47	5		C30 H48 O5	97.12	488.3495	28.203	488.3502	
32	Argiotoxin 659		-0.42	3		C31 H53 N11 O5	68.59	659.4234	25.291	659.4231	
27	6-Chloropurine riboside		0.99	1		C10 H11 Cl N4...	65.32	286.0466	16.17	286.0469	
29	4-Hydroxybenzaldehyde		0.22	5		C7 H6 O2	87.34	122.0368	16.987	122.0368	
21	2-Hydroxy-7,8-dehydrograndiflorone		4.81	5	LMPK12120415	C19 H20 O5	85.67	328.1295	14.063	328.1311	
13	1-Benzylimidazole		12.86	3		C10 H10 N2	58.92	158.0824	12.183	158.0844	

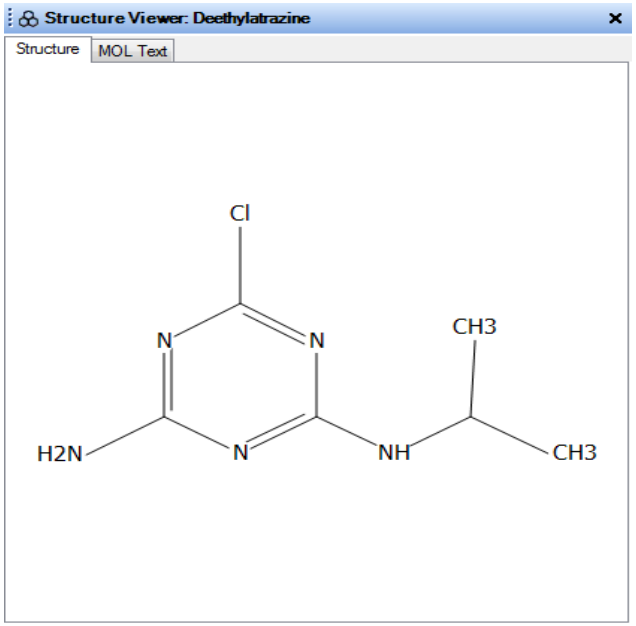
Compound List

Cpd	Label	Formula	Score	Diff (MFG, ppm	Mass	RT	Mass (MFG)	Diff (MFG, mDa)
31	Cpd 31: C11 H6 N6 O4; 19.700	C11 H6 N6 O4	83.29	1.85	286.0445	19.7	286.0451	0.53
24	Cpd 24: C13 H20 N2 O11; 15.018	C13 H20 N2 O11	92.45	1.2	380.1063	15.018	380.1067	0.46
11	Cpd 11: C9 H10 N5 O; 12.179	C9 H10 N5 O	99.06	1.01	204.0883	12.179	204.0885	0.21
8	Cpd 8: C24 H37 N O7 S; 3.035	C24 H37 N O7 S	90.19	0.9	483.2286	3.035	483.2291	0.44
20	Cpd 20: C32 H40 N O15; 14.068	C32 H40 N O15	92.76	0.71	678.2393	14.068	678.2398	0.48
30	Cpd 30: C12 H4 N3 O7; 19.713	C12 H4 N3 O7	84.98	0.57	302.0048	19.713	302.0049	0.17
10	Cpd 10: C14 H13 N7 O7; 8.221	C14 H13 N7 O7	93.97	0.55	391.0874	8.221	391.0876	0.22
9	Cpd 9: C20 H29 N4 O4; 3.019	C20 H29 N4 O4	83.84	0.45	389.2187	3.019	389.2189	0.18
2	Cpd 2: C3 H11 N5 O4; 2.816	C3 H11 N5 O4	87	-0.28	181.0812	2.816	181.0811	-0.05
18	Cpd 18: C7 H5 N4; 12.167	C7 H5 N4	95.86	-0.55	145.0515	12.167	145.0514	-0.08
35	Cpd 35: C11 H16 N3; 21.017	C11 H16 N3	84.77	-2.49	190.1349	21.017	190.1344	-0.47

Identification – Deethylatrazine ??

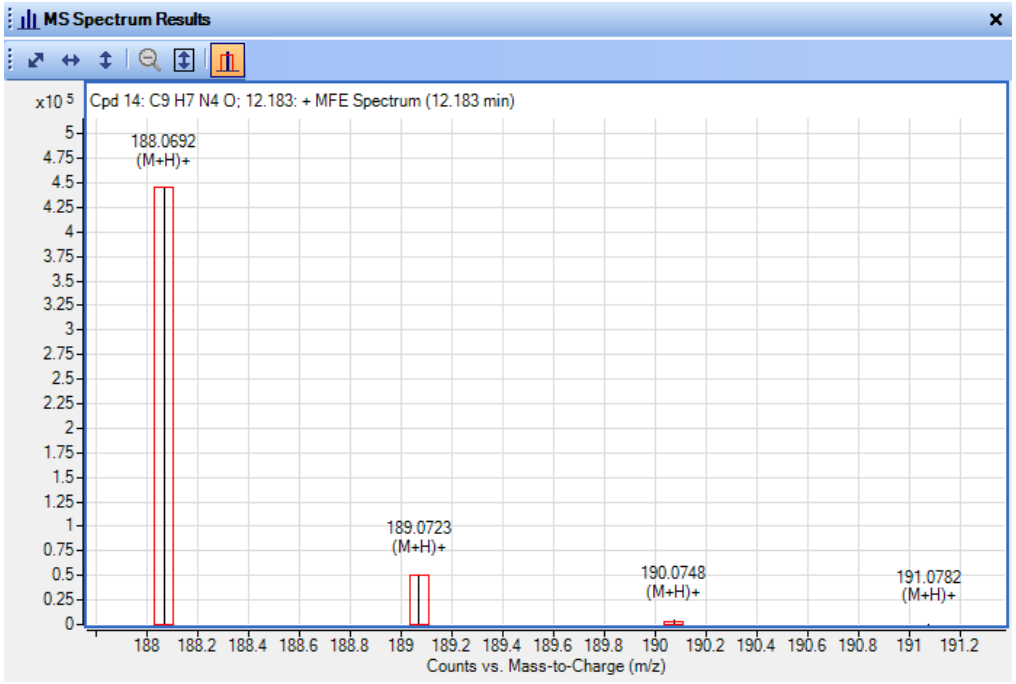
Database proposal of Deethylatrazine for compound 14 is not supported by the MS isotope pattern.





Identification – Formula Generation

Formula generation proposes a formula of C9H7N4O which has a much larger score and better isotope pattern fit.

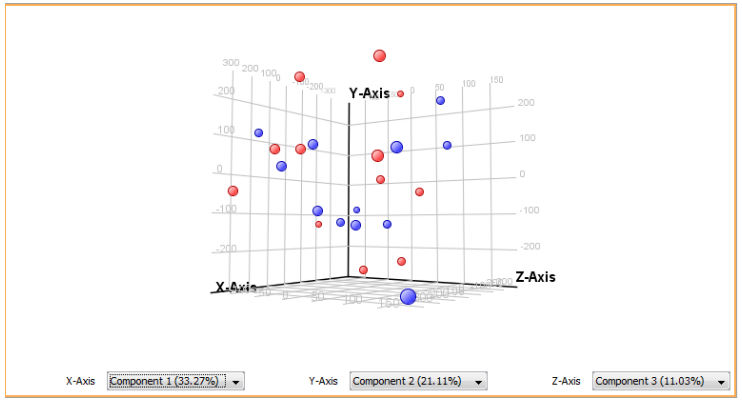


14	Cpd 14: C9 H7 N4 O; 12.183				1		C9 H7 N4 O	49.75	-0.19	187.062		187.062	-0.04	1	
Best	Name	Formula	Score	Mass	Mass (DB)	Mass (MFG)	Diff (ppm)	Diff (abs. ppm)	Diff (mDa)	RT	ID Source	Score (DB)	Score (MFG)	DBE	
◉		C9 H7 N4 O	49.75	187.062		187.062	-0.19	0.19	-0.04	12.183	MFG		99.51	8.5	
○		C9 H15 S2	40.52	187.0621		187.0615	-3.05	3.05	-0.57	12.183	MFG		81.05	2.5	
○	Deethylatrazine	C6 H10 Cl N5	34.2	187.0622	187.0625		1.65	1.65	0.31	12.183	DBSearch	68.39			

Does Organic Cultivation Make A Difference ??

A single variety (AT) was grown under organic and conventional cultivation conditions.

The James Hutton Institute



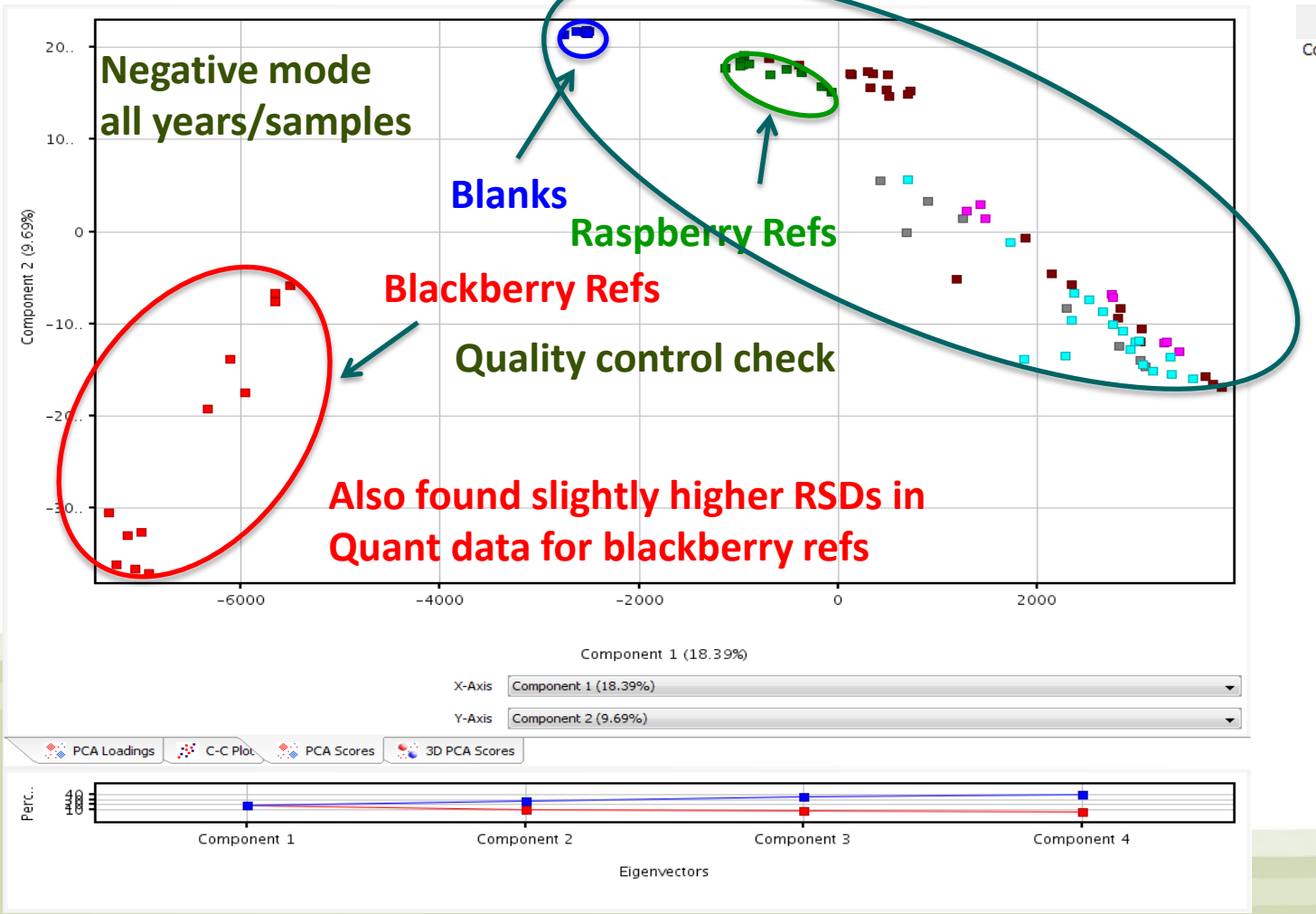
Result Summary

	P all	P < 0.05	P < 0.02	P < 0.01	P < 0.0050	P < 0.0010
FC all	151	0	0	0	0	0
FC > 1.1	81	0	0	0	0	0
FC > 1.5	50	0	0	0	0	0
FC > 2.0	50	0	0	0	0	0
FC > 3.0	21	0	0	0	0	0
Expected by chance		0	0	0	0	0

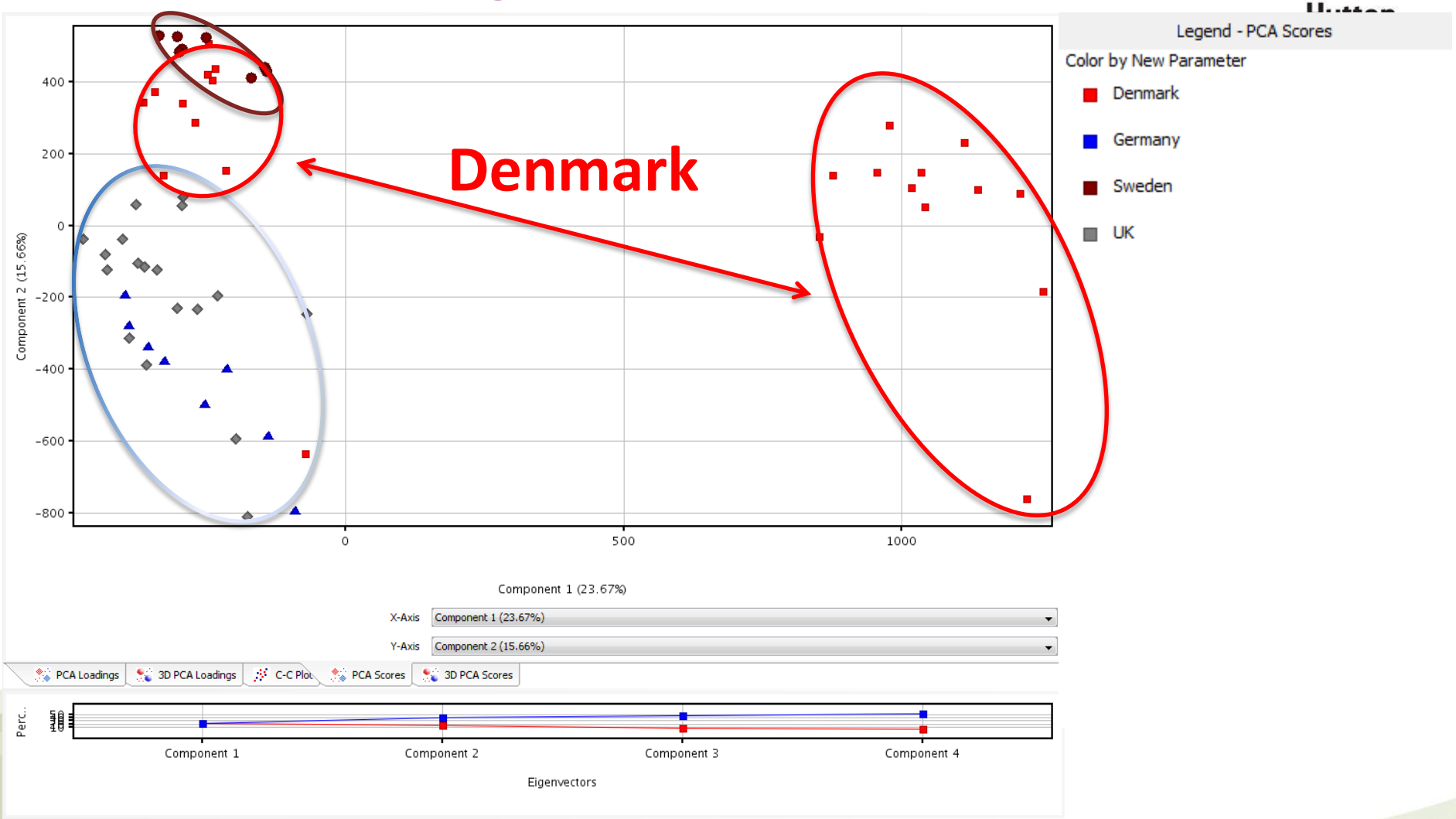
Use of Mass Profiler Professional for Raspberry TN Project

The James Hutton Institute

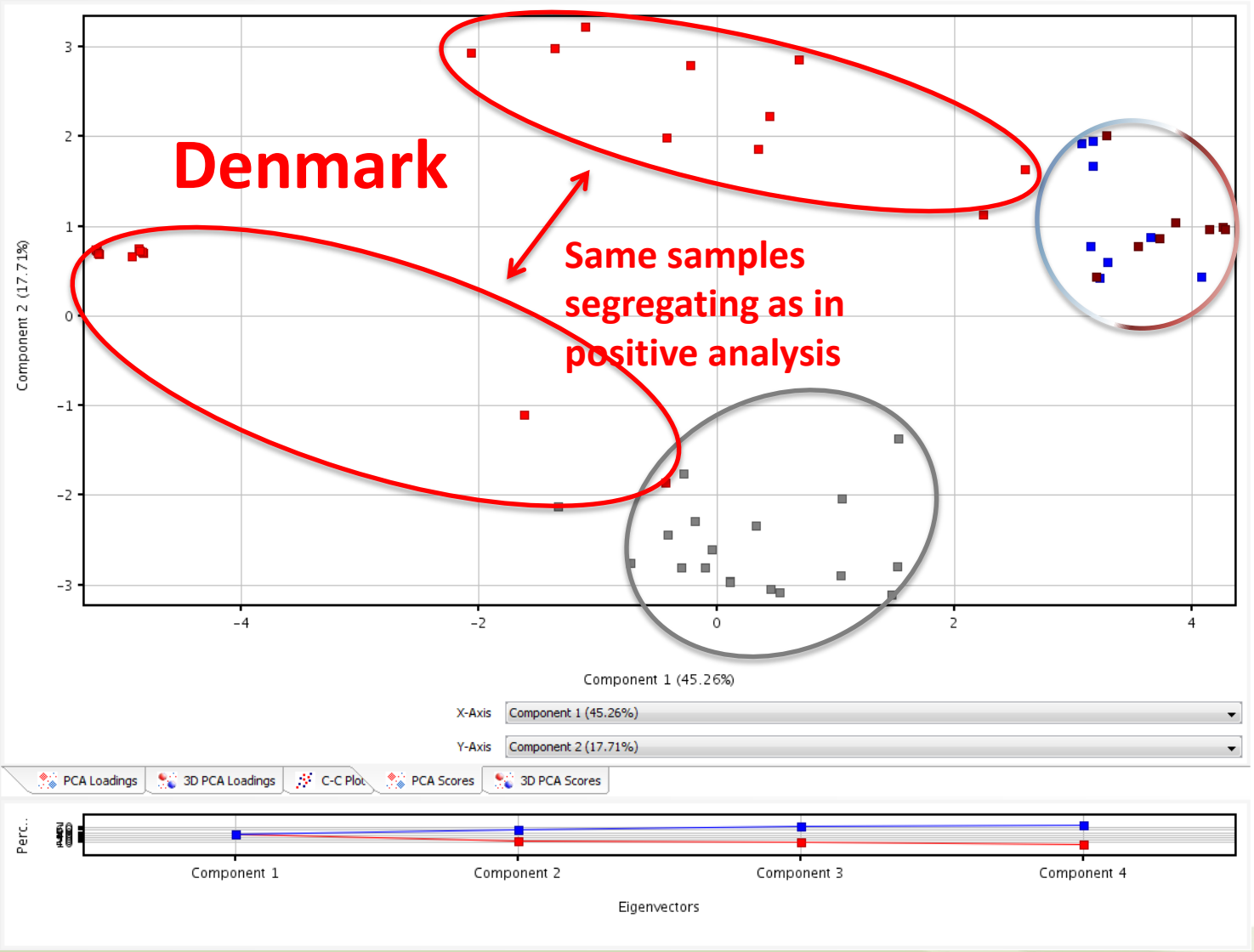
Focus on Autumn Treasure

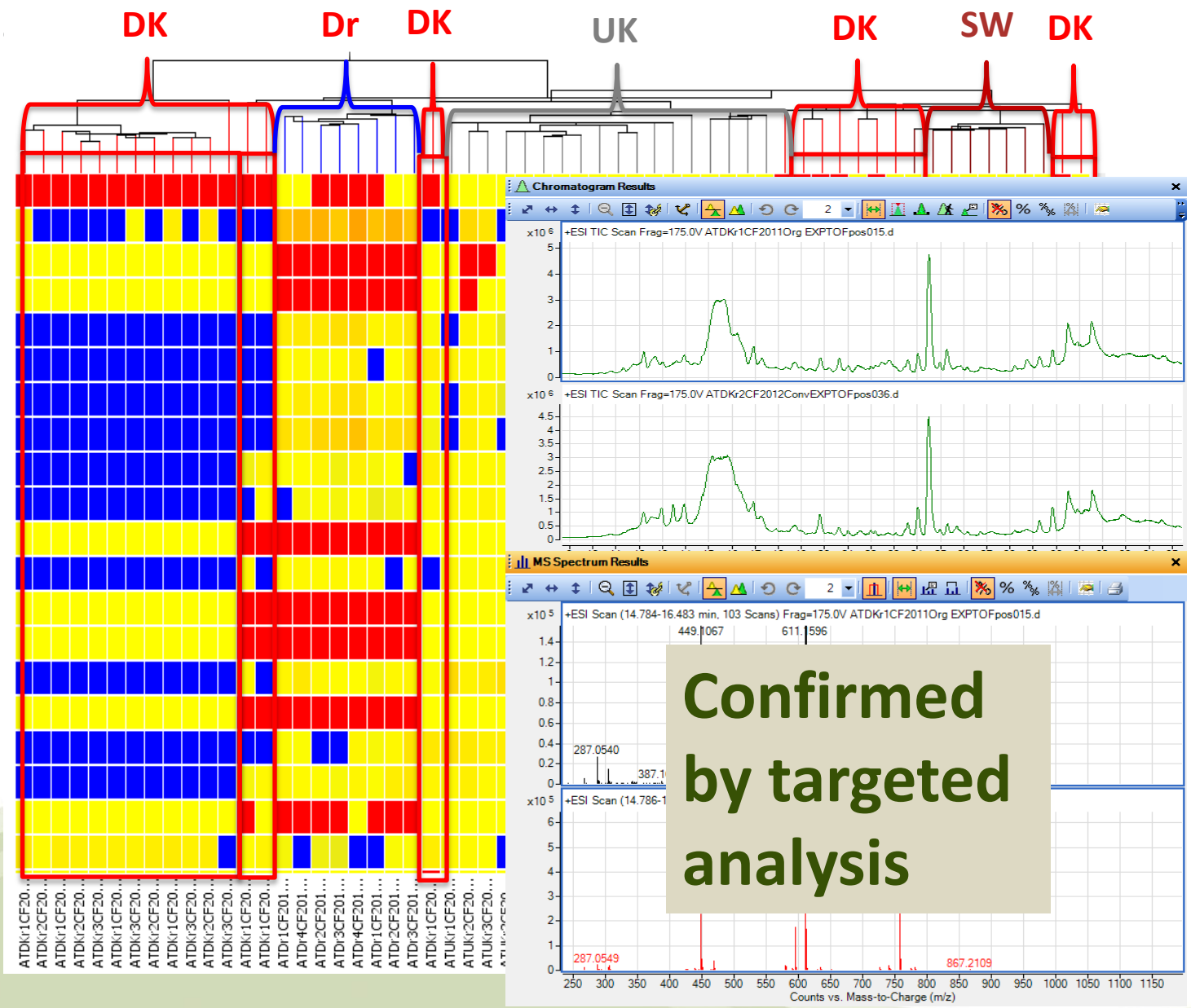


PCA 1V2 Autumn Treasure Blanks & Refs Removed - negative



PCA 1V2 Autumn Treasure Blanks & Refs Removed - positive




DK
Dr
DK
UK
DK
SW
DK
Chromatogram Results
+ESI TIC Scan Frag=175.0V ATDKr1CF2011Org EXPTOFpos015.d
+ESI TIC Scan Frag=175.0V ATDKr2CF2012ConvEXPTOFpos036.d
MS Spectrum Results
+ESI Scan (14.784-16.483 min, 103 Scans) Frag=175.0V ATDKr1CF2011Org EXPTOFpos015.d
449.1067
611.1596
287.0540
287.0549
867.2109
Counts vs. Mass-to-Charge (m/z)
Confirmed by targeted analysis

Data Mining – Germany V UK

- Generate comparisons between groups

Compound	Regulation ([Germany] Vs [UK]) : Normalized	[UK] avg : Raw Scaled	[Germany] avg : Raw Scaled
331.1254@11.294491	down	544620.44	74760.03
934.062@12.121386	down	150638.33	308.43866
146.0365@15.242655	down	36365.863	568.7739
578.1404@15.94045	down	1800159.9	689792.75
288.0618@16.075474	down	491891.94	480.73743
594.1558@16.361885	down	4037242	8669.79
316.1289@16.39073	down	49595.95	361.40717
666.1374@16.700771	down	18886.486	279.36105
344.1434@16.773407	down	124209.56	499.5067
290.0793@16.93731	down	3749013.2	1439901.4
138.0318@16.937773	down	844450.06	348932.4
448.0979@17.111551	down	196229.69	90.58586
562.1467@17.28109	down	1804926	627998.1
548.1849@17.863102	down	12757.754	66.47821
354.1646@18.170513	down	97231.336	4.24261
412.1722@18.951117	down	148233.42	6500.221
574.2218@18.966572	down	13294.727	285.72235
300.1178@19.010069	down	555576.4	63141.94
616.3293@19.015388	down	139080.8	22777.861
502.182@19.841	down	159154.61	74.02454
302.041@19.863255	down	168513.53	80.861275
452.1636@20.588594	down	84964.96	33669.996
412.1716@20.605858	down	75233.58	27738.545
414.1842@20.609915	down	83095.33	7292.14
208.1445@20.661974	down	113777.1	33436.777
458.2113@21.649847	down	144178.92	1297.5588
538.236@22.262154	down	475333.2	166300.77
538.2348@22.59311	down	134007.28	67.54607
372.1751@22.783226	down	205339.23	49811.453
314.1716@22.790148	down	39418.4	94.46004
392.1798@22.812517	down	205553.27	20.123981
1351.6868@23.456982	down	436162.62	12479.019
430.2172@31.072823	down	248587.66	18.42234
400.2064@31.176596	down	237898.92	6759.6934
313.2618@35.263504	down	749947.44	302638.38
898.1965@13.248199	down	25566.326	60.536125
426.094@15.933382	down	752.3607	1
866.202@16.533382	down	247834.55	1142.058

Data Mining – Germany V UK

The James Hutton Institute

794.2197@16.543953	down	134230.7	22661.197
342.126@16.5533	down	30233.844	4.500508
560.1854@16.9795	down	15190.528	4.41167
340.1464@17.059498	down	90000.266	21.748276
594.1388@17.343193	down	3918.1968	1
850.2087@17.520666	down	13429.745	15.9429035
1154.2625@17.642384	down	19.512495	1
578.1402@17.679657	down	13137.5	1
348.0812@17.891365	down	256389.81	7658.992
850.2063@18.092688	down	29861.643	4.1919293
578.2525@19.016273	down	20709.883	17.062534
834.2122@19.076162	down	68738.85	17.969711
610.1508@20.078224	down	15319.686	1
610.1495@20.630749	down	143338.3	1
190.1349@22.258749	down	278159.75	20.015024
436.2963@24.044271	down	3283.9822	263.42566
490.2016@24.84116	down	21189.645	4.5057735
498.2969@25.703676	down	22161.68	1
700.3378@25.712849	down	10324.328	1
516.3065@25.723032	down	57048.75	1
1368.6866@29.844738	down	2810.5662	775.38354
482.3017@29.84846	down	14930.771	17.362873
578.1404@14.607002	down	346.26913	4.161607
122.0367@17.265379	down	3.5282555	1
410.0994@17.276167	down	54.302586	1
562.1457@19.306913	down	11.747843	1
300.1339@23.516916	down	94.241646	4.155276
482.3011@24.034658	down	72.98053	4.0415845
322.1294@24.41619	down	43.46028	4.142021
296.1046@27.056261	down	7.029359	1
498.2967@29.834234	down	18.70755	1
970.6127@29.841938	down	62.024723	1
962.5887@29.843134	down	3.1465805	1
454.3078@29.84497	down	536.23584	4.010313
436.2974@29.849314	down	7681.9814	16.961914
982.6143@30.423685	down	5670.6426	52.242374
938.6239@30.479555	down	3.1482642	1
801.1008@10.977267	down	44.17492	1

- This data can then be imported back into chromatographic software for compound identification by database or formula generation as previously described.

Conclusions

The James Hutton Institute

Acknowledgements

- A. Foito, S. Freitag, S. Devlin, R. Faby, B, Svensson, K. Rumpunen, A. Wold, A. Nes, L. Andersen, H. Pedersen, M. Williams and D. Stewart