MZmine Tutorial

First presented at EMBL-EBI Industry Programme & Metabolights Project Workshop

22nd May 2012

Mark Earll - Syngenta UK



About MZmine

MZmine was developed at Okina wa Institute of Science and Technology, Japan and VTT Finland.More recently some development has been sponsored by Syngenta. It is a Java based program and is therefore platform independent. You may download it from the following website however it should be pre-installed on the EBI machines for this tutorial.

http://MZmine.sourceforge.net/

MZmine will import the following filetypes: Net CDF, mzData, mzML, mzXML, Xcalibur Raw files, Agilent CSV files. (For the Thermo Xcalibur files it is necessary either to have the Thermo Xcalibur software installed on the same machine or to have downloaded and installed the free ThermoMSFilereader software to be found at

http://sjsupport.thermofinnigan.com/public/detail.asp?id=586.

The version of MZmine used in the following examples was 2.8

Processing a simple Metabolomics dataset in MZmine

In the example dataset we have an excerpt of a metabolomic study on the ripening of fruits. We have nine samples of two different varieties, Wild-type and non-ripening Mutant plus ten control samples which consist of a large batch of identical fruit extract that are run at every fifth sample. In addition the fruit are sampled everyday from the onset of ripening between 47 and 54 days. (This example datasets is only a small excerpt of a larger replicated study). The data were collected on a Thermo Velos Orbitrap running in ESI+ mode with a UPLC column.

Loading the data

One of the great advantages of MZmine is its interactivity. Firstly we begin by importing the data. *Raw Data Methods/Import*

Look in:	MetData-E	BI	- 🔊 🕫	E) 📰 🔛
e.	MU47_917_1	0.mzData 🗿 QC6_917_24.mzData		
	MU48_917_3	19.mzData 🌆 QC7_917_30.mzData		
Recent		9.mzData 🎧 QC8_917_41.mzData		
Items	MU49_917_5	8.mzData 🍘 QC9_917_46.mzData		
1000	MU50_917_3	3.mzData 🌆 WT47_917_13.mzData		
	MU51_917_4	15.mzData 🌆 WT48_917_40.mzData		
Desktop	MU52_917_2	1.mzData 🗿 WT48_921_17.mzData		
	MU53_917_0	6.mzData 🍘 WT49_917_59.mzData		
T.	MU54_917_2	6.mzData 🕋 WT50_917_32.mzData		
	@ QC10_917_5	1.mzData 🍘 WT51_917_43.mzData		
My	QC1_917_01	.mzData m WT52_917_20.mzData		
Documents	G QC2_917_02	mzData G WT53_917_06.mzData		
1.	GC3_917_08	.mzData m WT54_917_29.mzData		
1000	GT QC4_917_14	mzData		
Computer	@ QC5_917_19	e.mzData		
C.				(-
-	File name:			Open
Network	Files of type:	All raw data files	Ť.	Cancel

Once the data is imported we can right click on the data file to reveal several display options

MZmine 2.8: New proj	ect	A Please set	the parame	lers		
	ect Peak list methods Visualization Windows Help Show TIC Show mass spectrum Show 2D visualizer Show 3D visualizer Remove	Data files MS level Pict type Ratention time es/2	wass, and same peak of	el acties (10 acties) 20 acties) 20 acties) 20 acties) 20 acties) 20 acties) 20 acties -	Al Cear Auto sange Auto range	Fran formale.
CC3_917_08.mzData		Peaks		Clear OK Cancel	Help.	
B GC2 917 02.mzData		-				

The TIC option offers the option of Base peak or TIC and allows you to set various ranges. Clicking OK leads to a high quality spectrum plot. The plot is fully zoomable and interactive, and double clicking a peak leads to its mass spectrum. Clicking and dragging upward or to the left is a gesture which results in zooming back out to maximum zoom. Clicking and dragging downwards or to the right zooms in.



NB: Make sure you take a note of the height of the baseline and the height of the smallest peaks. This will be useful later.

The mass spectrum plot also enables you to see associated ms-ms data. (In this dataset the MS-MS information has been removed).

					rData) ecan #				l
			MS1, RT 2	.0, hase pea	< 136.062 mit; (1	1.507)			
1.3E7	100.002								- 1
1.067									
9.0E0 -									
8.000									
2/101									- 1
6.001									-
7.001 6.005 5.066	1								
4.000									_
3.066									-
2.066									
1.066	1.	204303			407.175				-
0.080			a			827.106		794.428	878
10	0.000 3	00.000 30	00.000	405.000	500.000	600,000	700.001	800.000	
					m/z				
				1 Sea					
-			Peak list:			*			-

Peak detection is a three step process:

- (1) Mass detection
- (2) Chromatogram building
- (3) Peak Deconvolution

Mass Detection

Click on Raw data methods/Peak Detection/Mass detection



In this case we have imported mzData files which are centroided during the conversion from .RAW so the only option is Centroid mode. (If you have imported Thermo .RAW files then the data is continuous and you can use the exact mass, local maxima, recursive threshold or wavelet methods).

The 'Show preview' option allows you to interactively set the threshold for peak detection in the mass dimension. The aim is to detect peaks but not too many noisy features.

lass detector	Centroid	Noise level 1.0E4
IS level	1 •	Show preview
lass list name	masses MS1	
_		OK Cancel Help

[MU51_917_45.mzData] scan #1
MS1, RT 0.0, base peak: 127.039 m/z (1.555)
1.5E5 127 039
1.265
1.065
Atjene 17.5E4
§ 7.564
E
5.064
2.5E4
893.107 794.429
0.0E0 100.000 200,000 300,000 400,000 500,000 600,000 700,000 800,000
4000000 40000000 0000000 1000000 000000 000000 000000
m/z

After clicking OK MZmine will build the mass list. Depending on the speed of your computer this may take some time. When the mass list is built the icon will show a green tick mark.



Chromatogram Builder

The Chromatogram builder is found under Raw data methods/Peak detection/chromatogram builder



Click on the *Choose…* button to select the mass list just generated and fill in the parameters as shown (for your own data these will vary)

Aass list	masses M	IS1		Choose
Ain time span (min)	0.017			
Min height	1.0E4			
m/z tolerance	0.0050	m/z or	5.0	ppm
Suffix	chromatog	rams		

You will then see a number of chromatograms listed in the left hand pane



Double clicking on a chromatogram will bring up the results:

New project	* E NK8	53_917_06 mzData chromatograms				
↓ Rave data Nes ± ± ± ± ± ± ± ± ± ± ± ± ± ±	D	Average Identi	ty Comment	Peak shape	MU53_917 Height	06 mrData
1 2 1048_917_39 rudbala 2 2 1048_917_31 rudbala	an	101.509 10.0		1	. 1364	1.365
H 4 1048_301_19 m05ala	146	102 834 9.3			. 5.754	5.728
E 2 10.40 917 55 mcDate	37	102.005 1.5		Anna Ma		1.167
12 2 H054_917_28 Hutbera		100000 //5				
El 💕 QC10_917_91 ecCata El 💕 QC3_917_95 ecCata	-	102.091 10.4	tra M	THIT WITH TA M	+ 1 284	1.067
8 2 0C4_917_14.msData		142.128 5.5		1	· 2.065	1.462
E GC5_917_19.xuDete GC2_917_92.ruDete	390	102.970 1.1		100 mm - 100 mm	0.2.905	2.157
🗄 🛫 GC1_917_91.md5ala	at .	101.010 1.6	-		* 4.124	5.455
H @ 006,817,34,mcDate H @ 007,817,30,mcDate	92	103.054 7.0			· toes	1.008
GC0_917_45Au20ata GC0_917_41 mcData	211	102.807 11.8			# 1.0E4	3.455
- WT40_017_40.mzData		A CONTRACT OF A			and the second	
E WT49_547_58.nuData	34	163.131 6.5			± 1.364	1.364
3 WT40_521_17.mpDate	96	103.825 1.5			1.254	1.304
1 WT50_517_32.rccData	90	103.891 1.4			# 1.964	+764
H WTS2_517_22 milliota WTS2_517_05.milliota	- 17	104.806 1.5			.2464	2.054
1 WT54_817_39.mzData	14	104.571 1.5			* 6.228	1.627
the Peak lists					-	
E AUA7_017_10 mcData chromatograma	39	104.877 1.5	-		# 2.1E4	3.364
WU48_921_19 mcGate chrometograme WU48_921_19 mcGate chrometograme WU51_917_45 hgCate chrometograme	A DB	104.102 1.4			# 2.064	5.584
12 10 00,48_917_56 mcCate chromelograme	101	104.107 1.4			* 1.187	8.987
MO48_917_39 mcData chromatograms MO12_917_31 mcData chromatograms MO12_917_31 mcData chromatograms MO12_917_31 mcData chromatograms	hog	104.110 1.4			# 3.1E#	5.004

The peak list is comprised of a series of ion chromatograms taken at each point that was detected in the chromatogram builder. Some ion chromatograms may contain more than one peak so a second peak deconvolution stage is required.

Peak Deconvolution

Make sure you highlight the chromatograms in the left hand pane. Click *Peak list methods/Chromatogram deconvolution*. You have a choice between Baseline Cut-off, Noise Amplitude, Savitsky-Golay, Local minimum search [and Wavelets (XCMS) in a future version]

↓ MZmine 2.2: New project					
Project Raw data methods	Peak list methods Vis	sualiz	ation Windows Help		
New project	Peak detection	ŀ	Smoothing		
	Gap filling		Chromatogram deconvolution Ctrl+D	le atitus	
Peak lists	Isotopes	•	Peak shape modeler (experimental)	lentity Resolving individual peaks within each c	
⊞ QC2_917_02.raw peak lis	Filtering	•	Peak extender	Resolving Individual peaks within each ci	nromatogram
	Alignment	•		1	
	Normalization	140.0	0817 12:34.185		
	Identification	F	12.34.103		
	Data analysis	► 300.2	2020 08:57.304		
	Export/Import	•			
	14	207.1	1371 09:09.475		

Here we use *Noise Amplitude* [My personal preference is Wavelets XCMS but it may not be implemented in this demonstration version]

Auffix	deconvoluted_NA	Min peak height	1.0E4	
Jgorithm	nithm Noise amplitude	Peak duration range (min)	0.0	- 0.8
Remove original peak I	st 🔟	Amplitude of noise	5.0E4	
C	K Cancel Help		Show pro	eview
		OK	Cancel	Help

Fill in the boxes with the appropriate values. One of the recent key improvements to MZmine is the ability to specify a maximum for peak duration. This is very useful for removing some of the artefact peaks caused by column bleed. Here we use 0.8 mins. Another trick used here is to set the amplitude of noise slightly higher than the min peak height. This means that peaks with raised baselines as in the example get detected. The downside is a slight loss in integration accuracy. (The alternative is to baseline correct first - see later, or use a more robust peak picker such as the wavelet option). Peak picking is always a compromise and requires a lot of experimentation for optimal results.

n peak height	1E4																	-
ak dutation range (min)	0.0	- 0.8		4.066	10		-	-	-	-		-	-	-	-			
nplitude of noise	5E4			3.8E6 3.5E6														
R	Show preview			3.366														1 1 1
ak list ANUSD_917	.00.msData strome	eloigiante .																
romatogram #1 85 029	niz		- 17	2.066 2.566 2.266														
				# 2.0E6														
				35 1.5E6														
				1.0E6														
				7.5E5														
				5.0E5														
				2.5E5														
				0.0E0	10	20	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12/0	15.0	14.0

After the peak deconvolution step MZmine produces a resolved peak list with one peak per row:

WTail 517, 45 mpData chromatograma	* []	63 917 06 mgData chrs	and solutions. As can edit	Get RDK		0
W147_517_13.4z3ata chromatograma W148_517_55.ez3ata chromatograma W148_521_17.ez3ata chromatograma	0	Average m/z Ret.time	Identity	Comment	Peak shape	MU53_917 2
GC7_917_30 mptate chrometograms W750_917_32 mptate chrometograms	1	R1429 15		1		+ 1905 *
WTS2_SIT_30.mcData chromatograma WTS1_SIT_43.mcData chromatograma		85.040 1.5				+ 1.254
WTS3_917_06.mpDate obromatograme	-				1	
 W754_817_29 mcDate chromatograms GC0_517_40 mbDate chromatograms 	2	85.576 13.8				+ 1.4ES
d QC6_917_41.msDate chromelograme	4	01.032 1.5				+ 2.985
BUJ47 917 10 mcDate chromelograme deconvoluted INA	5	89,090 1.5				+ 8.055
80.01, 917 40 m20ata chromatograme deconvoluted (k. 80.46, 821, 15 m20ata chromatograme deconvoluted, h.k		86.097 5.3			1	w 2.125
UU49_917_58 nzData chromatograma deconvoluted_bA	7	MUNT 5.6			1	+ 2.85
HU48_917_38 md2ata chromatograma deconvolved_NA HU52_917_38 md2ata chromatograma deconvolved_NA		84.945 11			1	a. 6.525
80,50_917_33 mcGala chromatograme seconvolved_NA						
8054,917,28 muCata chromatograma deconvoluted NA WT48 917,42 mcData chromatograms deconvoluted NA		87.544 1.5				# 1.450
🚺 GC1_017_01 redbas chromatograms deconvelded_RA	14	87.058 1.4				+ 1.165
 GC2_917_82.m2Data chromatograma deconvalided_NA GC5_917_19.m2Data chromatograma deconvalided_NA 	11	87.964 1.5				# 5.924
QC5_917_95 m25als chromatograms deconvoluted_NA QC4_917_14 m25als chromatograms deconvoluted_NA	12	87394 1.1				+ 8.254
GOL_\$17_34.mcDate chromatograme deconvelided_NA		80.040 14				* \$ 664
 GC18_917_S1 m2bata chromatograme deconvoluted_NA WT47_917_11 m2bata chromatograme deconvoluted_NA 						and the second second
W152_817_20.mzData chromatograma deconvoluted_NA	E 14	81.546 1.5				# 0.104
WT53_817_06 mcDate chrometograme deconvoluted_SA WT80_917_02 mcDate chrometograme deconvoluted_SA	18	88.676 8.1				+ 1.264
WT48_521_17 mcDate chromatograms deconvoluted_NA	16	80.076 52				# 1414
WT49_917_28.rs20ate chromatograme deconvoluted_SA E_WT49_917_59.rs20ate chromatograme deconvoluted_bik	12	88.576 9.2				+ 5.164
WTSt_617_43.mcData chromatograme deconvoluted_NA		1 W 12 52				
GCT_91T_30 m2Data chromatograms deconvelided_3A E_GC0_91T_40.m2Data chromatiograms deconvoluted_8A	18	80.070 8.3			2	# £363 #

We can visualise the peaks using the 3D visualiser plot on the raw data. This is a useful check of the accuracy of peak picking.

Project Raw data	methods Peak list methods Visu	dization Windows H	A Please set the paran	1 -					×
A Raw data tie	and the second se	î.		0.0		14.0	Auto range		
E Z ND4	Show TIC Show mass spectrum		Retention time resolution	2000					
10. 10. 00.	Show 2D visualizer		m/z	85.001	-	899.989	Auto range	From formula	1
10- 00 MILA	Show 3D visualizer Remove		m/z resolution	2000					
0C10_917	51.mzData			6	OK .	Cancel	Help		





There is also a 2D "gel view" of the data - click Show 2D visualiser.

Right clicking a peak in the peak list From the peak list and selecting (*show... chromatogram quick*) shows the peak and the peak integration in pink. (There is also an option to see the peak in 3D but this appears to be broken in 2.8)



Peak Alignment

In MZmine peak alignment is done after the peaks are picked. To adjust for any slight variation in retention time a retention time normaliser is provided. Click on *Peak list methods /Normalisation /Retention time normaliser*.

ject Raw data methods Pa	ait list methods Vi	scalars	tion Windows Help	
1 2 WT54_917_28 mzDate	Peak detection		*	
E E NUS3_917_06 w2Date	Gap filing			
10 10 10 10 MD47_917_10 mpDate	Isotopes			
E I HU45_521_19 mcDate I HU51_917_45 mcDate	Filtering	•		
£ 4049,917,58.mzDete	Alignment			
HU40_017_38 mcOate MU52_017_21 mcOate	Normalization		Retention time normalizer	
1 WUS0_017_33 Actives	Identification	The	isleiður kva turnalítar aðarspís fu ruðura D	deviation of relevation times between peak lists, by exercising for comman peaks in these peak lists and using them as normalization dat
0C10_917_20 mpDate	Data analysis		Standard compound normalizer	
E GC5_017_18 mgDeta	Expert/Import			

Name suffix	RT_norma	lized	
m/z tolerance	0.0050	m/z or 5.0	ppm
Retention time tolerance	5.0	relative (%)	
Minimum standard intensity	1.0E6		
Remove original peak list	171		

Next we will combine the peaks using the Peak list methods/Alignment/Join aligner.



The alignment is based upon RT and m/z tolerance. There are options to only merge ions with the same charge state, the same ID or by isotope pattern. We will not use "Require same ID" because we have not identified any compounds yet.

Peak list name	Aligned pe	ak list		Isotope m/z tolerance	0.0050	m/z or 5.0	ppm
m/z tolerance	0.0050	m/z or 5.0	ppm	Minimum absolute intensity	1E3		
Weight for m/z	20			Minimum score	65 %		
Retention time tolerance	3.0	relative (%) 🔹		winingin score	00 100		
Weight for RT	10			OK	Cancel	Help	
Require same charge state		Score for perfectly matching	ng RT values	-	- Circumstering	J (ministran)	
Require same ID	10						

You should now have an aligned peak list. Green dots indicate the presence of that peak in the scan. A red dot indicates the peak was not detected. After the identification process we will return to fill in these gaps with baseline levels from the other scans. (Gap-filling may alter the accuracy of the m/z value due to averaging)

unit, art, ricesters morenegens in *	1,744	pined paints	141												-CR	
10702_017_30.esData chromatograma der 10702_017_30.esData chromatograma der	0		renage -		1.			1000		MARE 917	N-milles	MJA7, 917	Non-Dea	MART SHT	AL MICHAE	N
with art, to refuse changing and de	- 9	- 10	Batting	Manapy	Connect			wak shape		Thought	Ana	(Theogen	Aine	- Haught	Alter	
1752_621_17 active chonespece de		101.000	1.06							4 1.800	4.882	4188	1.477	++386	2.147	-
1/154_311_35 rs/Sets chronelogues de 1/166_311_56 rs/Sets chronelogues de	- E	-									12112					
WTD1_BTT_AD excluse strengtypers de	12	201.046	102							4.6.355	1.085			0.5188	1.94	. **
UCT_Sr1_32 edites prometypane less	1	80,076	1679						1	4.1882	1.865			+1.965	1.008	
223, 917, MinuSee Universidate Inc. 223, 917, 91 Audies Universidate Dio		and sold	196						-	+ 2.89	1.001	+ 120	4.701	+2185	140	
10/01, 917 dd-rollana chronaltograma dar	F										1.41					
Wolf_311_15.mcDate chronialograme de-	D	10.001	348							# 8.401	3.001	+ 1 (12)	1.888	a 1988	2.388	
MURT, BTT, AL registry constationants in MURE, BT1, 10-registry constation are	R		4.00							+ 2 495	4.708	+ 1 10 1	8.00	+185	2.084	
Work, PT, Municipale concessing area in-	F	-					-			_						
Most, PT, Northeas monetagene in	11	86.047	1.88							+2.80	7.881	4.0.000	1.100	+ 1185	0.000	
MUSE_W1_31 exclusis connexinginitie do MUSE_W1_22 exclusio connexinginitie der		100,000	1.16							+ 6.925	1.84	+ + 101	1.000	+ 4.185	1.969	
mills \$17.38 will be chromotoppere des		87.844	1.00			-				4 1 460	1.155	-0.1.000	1.948	+1.888	1.005	
10160_517_40 collars riverenting with the	r	11.000	144								1.100	4100	1.00	0.1466	1.00	. *
GC1_R1 ² _31 reliate chronologiane deci GC2_R1 ² _31 reliate chronologiane deci		10.000	100							a 1.000	1.803	-0.010	1.000	+ 1100	1489	
\$22,877,18 Addee strending are dear	44	int may	106			1					1.004	+ 1 711	1.200	+ 4.864	6164	
GCI, N.T., M. H.Calle chromologiame deci	-					_										-
GOLUT, 11 reflets investigants ten GOLUT, 31 reflets investigants ten	4	10.000	10							4.8.234	1.45	4.6394	1401	+191	1.000	
GC10, 817, 51 million recomming arts the		10.040	1.00			1				a 1.00x	2.80				1.011	
19797_011_13.rsDate chromatiquares der	1	-								-						
1752, 977, 20 million chromologium de 1753, 977, 20 million chromologium de	14	20.040	130							68.04	2.424	*		+176+	1.265	
1750, 217, 32 rollats chronalogians de	1	painter.	10							4.5388	8.404	+1.04	1.331	10.00	1.880	
WHAT ST, 17 will are chonen prove the		10.075	**					_	-	+1.04	1.04	+ 140+	1.201			
10754,817,35 estilata conceptione de 10748,917,35 estilata conceptione de		-	***													-
settin and an entropy and an		10.075	827						1	4.2.154	4.034	48.891	1.05			٠
DIT_NT_30 rulling drovening are inco		100.010	4.54						0	4.2.894	2.90	1.0.294	1.800			
COX, NT, N milles dromaturare and		-							<u>+</u>	-						
Contemporary -	E*	10.175	947							1.1.21	4.451					٠

Identification

In order to identify peaks a database of m/z or m/z and Retention times are required. There are two options. One is a custom database compiled on your own instrument based upon the measured masses of the molecular ion and any adducts or fragments. The other option is an online database search based purely on the accurate mass and isotopic pattern matching.

Custom database search

Peak list methods/Custom database search

Project Raw data methods P	wark has matched any via	URICAL	ion ynno	ows rielp				
🗄 🚺 WT47_917_13.mzDati	Peak detection	Alg	ned peak i	int				
H WTS2_917_20 mcDate WTS3_917_06 mcDate	Gap Sling	in l	A	verage	Identity	Comment		14
1 WT10_917_32.mtDet	Isotopes 1		m'z	Rettime	Identity	Comment		
H WT48_921_17 HUD48 H WT54_917_39 rs204t	Filtering	1	85.029	1.56				
WTV-347,40.mtDet VTV-347,40.mtDet Ord,917,30.mtDet Ord,917,31.mtDet Ord,917,41.mtDet Ord,917,41.mtDet Ord,917,41.mtDet Data anal Ord,917,41.mtDet	Alignment Normalization		85.048	1.52			1	
	Identification 4	land.	Custom da	tabase seat	a			
	Data analysis	The	mettod per	eshes a cadota	database (CS	V tite) using m/p a	and retarition time rate	446
	nromatograme dei		Adduct se Complex s Online dat	10111			1	1
Model 17, 30 models of Model 17, 30 models of Model 17, 31 models of Model 17, 31 models of Model 17, 33 models of Model 17, 33 models of Model 17, 35 models of Model 17, 35 models of WT48, NT7, 40 models of	rromatograms de- tromatograms de- tromatograms de- tromatograms de-	1		ospholipid pr search			4	1
E C OC1_BIT_D1 suDate ctv	onetoprame deco	10	87.056	1.38			1	

In the dialog box select POS_mzRT_database.csv (which will be with the demo datasets).

This is a database we have compiled for our library using our UPLC-MS system in positive ESI+ mode. (NB: Your own data will differ in RT and possibly the ionisation profiles - you will need to compile your own library database relevant to your own system!) Here is an excerpt of our custom database. The first column is KEGG ID, then accurate m/z, Retention time, Identity and Formula. We have also included common adducts and dimers in the list such as [M+K], [M+Na], [2M+H]

1	ID	m/z	Retention	Identity	Formula
28	C00041	90.05496	1.45	alanine [M+H]	C3H7NO2
29	na	161.0921	1.71	alanine-alanine [M+H]	C6H12N2O3
30	C01551	159.0513	1.62	allantoin [M+H]	C4H6N4O3
31	C06464	181.0707	1.52	altrose [M+H]	C6H12O6
32	C00216 C00259	151.0601	1.62	arabinose [M+H]	C5H10O5
33	C01112	231.0264	1.56	arabinose 5 phosphate [M+H]	C5H11O8P
34	C00532	153.0758	1.51	arabitol [M+H]	C5H12O5
35	C00792	175.119	1.37	arginine [M+H]	C6H14N4O2
36	C00049	134.0448	1.45	aspartic acid [M+H]	C4H7NO4
37	C00099	90.05496	1.38	beta-alanine [M+H]	C3H7NO2
38	C02512	115.0502	1.48	beta-cyano-l-alanine [M+H]	C4H6N2O2
39	C00719	118.0863	1.57	trimethylglycine [M+H]	C5H11NO2
40	C00308	177.0982	1.34	canavanine [M+H]	C5H12N4O3
41	C09773	363.1286	5.98	catalpol [M+H]	C15H22O10
42	C00185	343.1235	1.77	cellobiose [M+H]	C12H22O11
43	C01484	209.0961	10.94	chalcone [M+H]	C15H12O
44	C00852	355.1024	9.1	chlorogenic acid [M+H]	C16H18O9

Field separator			
mit.			
identity Formula	n Sime (xnin)		
ignore first line 🗵			
m/z tolerance 0.0010	0 m/z or 5.0	ppm	
Retention time tolerance 5	relative (%)	1-	

Look in	MetData-	EBI			100	
Rocert Bems Desktop My Documents Computer	Costs FileCosts F	35 nuOate 19 nuOate 55 nuOate 33 nuOate 45 nuOate 36 nuOate 36 nuOate 36 nuOate 36 nuOate 37 nuOate 31 nuOate	QC5_917_15 ncDes QC5_917_15 ncDes QC5_917_25 ncDes QC5_917_25 ncDes QC5_917_15 ncDes QC5_917_11 ncDes QC5_917_11 ncDes QC5_917_11 ncDes QC1_917_11 ncDes QC1_917_11 ncDes QC1_917_11 ncDes QC1_917_11 ncDes QC1_917_11 ncDes QC1_917_12 ncDes	🖀 9774,917,28 mcData		
	File name:	POS_m	zRT_database csv		1 6	Select fi
Network	Files of type	ALFie	Sector The sector sector sector		5 6	Cano

Clicking the Identity tab twice should bring all the identified peaks to the top of the list. Notice that some of the identified component are isotopic internal standards which we use to normalise the data (outside MZmine).

IS3_817_05.muData chromatograms der "	Align	et peak 1	ist .							101	ā
147_917_12 mcData chromatograms der 191_917_45 mcData chromatograms der 146_921_13 mcData chromatograms der	0		Ret. time	identity +	Comment		1	Peak shape	MU53_917_06 mg		7_9
H9_917_50 nzDala chronatograme der		195/2					_			100 1 1 1	tes,
140_917_39 m2Data chromatograms dei 152_917_21 m2Data chromatograms dei	4021	124.039	1.39	unnigent went (merid					•		
	4030	125.015	3.37	recolinamize (MAN)							
254_917_28.mcDala chromatograms and						_			• 1363 1365		2
car to far and the second first of the	1185	275.076	9.00	natigenais (Mrid)				11	• 1369 1365	• 3.65	2
31_917_01 ncData chromatograma deci 22_917_02 ncData chromatograma deci	0078	273.076	3.34	namperain (MHZ)							
A DAY AND											-
13_917_98.npData chromatograms daci	4100	192.098	3.31	mattornina (stret)							
D4_917_14.mgData chromatograms deci	756	147.113	1.22	fairs (block)					• 1.925 3.125		
05,917_24 mcDate chromatograms deci 10, 917,51 mcDate chromatograms deci					-						
T47 \$17 13 rollala chromatograma dei	2118-	627.158	1.79	teettee (N=fka)							
	610	132 102	5.33	inclausing (Minis)		1		. 6.405 5.005	• 2.95	4	
153_917_86 mcDela chrometograme de-			13211	acteurs (Anne)							-
	526	122,102	5.55	factive (street)			1		 1.328 3.768 	• 1.92	£.,
748_901_17 mcData chromatograma dei 154_917_29 mcData chromatograma dei	4078	132.103	6.99	leasure (Alrea)			-				
T48, 917, 50 mgCala utrumatograms der	Tere	1000.000		and the state of t	-				-		_
Tht \$17.43 multiple chrometograms dei	442	171.917	4.85	in_dd_d5-phenyleterine[kini]					# 3.128 2.008	# 1.12	8
27_317_30 moDela streamstograme deci	a handa d	113.050	1.44	to a figure of the Local Day		1		_			-
a transfer and the constraints and the	4196	111.000	1.49	14.04 (0.00 globaries and (0.00)		1			•		
35_917_41 mcData chromatograms deci. ¹⁰ 153_917_05 mcData chromatograms dec	1789	218.041	4.32	INCOME AND A DESIGNATION OF THE OWNER.							
of not on							-				-
151_917_85.mcData chromatograms.der	\$\$47	219.041	4.25	juiling?an exact acys. (grade)							
148, 921, 18 mcData christiatograms des	4100	197.059	4.30	ini ani an anno ann 2014							
145_917_58 mcDate chromatograms der						-	_				-
140_917_38 mcData chromatograms der 152_917_21 mcData chromatograms der	100	94.000	1.4Z	(rt_std_sHaterone (0144)					+ 3.469 3.268	* 2.H	6
	4227	160.029	2.26	int, stat, all-matrix active (Minhad)							
154 917 28 mcData chromatograms der						_				-	-
	5676	268.000	6.29	coming (tabob)							
1_917_91 nuOata chromalograma deco	MI	198.017	1.35	Topolitica (Mintel)		1			+2.65	+ 1.25	1
12_817_02 maDete chrometogrees deci +	00.1	rowany		untering formal					* 2 861 * 4 663	* 1.45	

Adduct search

Under peak list methods/identification there is an adduct search option. Note you can now load or save a customised list of adducts. Set the RT and m/z tolerance and the relative adduct peak height.

RT tolerance	3		relative	(%) *				
	(i) ge-	-	(71.962 m			•	- A8	
			37.866 mil				Clear	
			24121.969			9	Add	
Adducts			17.826 etc			13	Import	3
			04197.967				Export	3
	12 (M-	H20	03) 62 800	92		.]	Resot	
m/z tolerance	0.005	3	miz e	5.0	ppr	í.		
Max relative adduct peak height	100	1	6					
(OK.	í G	Cancel	Help	1			

10	A	verage	Identity -	Comment	Peak shape
	m/z	Rat time	iositzty -	Comment	Preak shape
\$378	378.123	10.16	(M+24-H) 78.918 mit active of 301.301 mit		
40.06	228.014	1.60	(d+24/m) 70.018 mit adduct of 101.082 mit		
4427	207.088	1.32	(51+26-ref) 70.010 (min animum of 100.000 res)		
963	100.067	1.58	(3142A2944)(33.389 mit adduit of 97.229 mit		
5128	833.268	8.22	(M+2+D++H) 83 060 mit watuut of 460 136 mit		
\$121	833.298	8.25	(NH-DACHHH) \$3,060 mid adduct of 460 YM mid		
5122	533.248	9.25	(51+24/24-14 83 080 mit adduct of 452 198 mit		
5028	S18 172	1.36	DI+24CH+ Herne: 30+24CH++(83.060 w/c extract of 4 Identification method: Adduct search	60 106 mir	1.1
5106	218.229	8.23	(34+24/2014) \$3,000 mix assuut of 433,071 mix		
1089	407.167	1.77	(NI+2ACH++)(03.000 m/s adduct of 404.138 m/s		
1799	471.228	8.55	(N+2A2N+++) 83.069 m/z edoko) of 388.199 m/z		
1010	w37 237	0.14	(NFCADOriel \$3.580 mic adduct of 584 178 m/s		

There is a similar option for fragment search (based upon MS/MS data). This cannot be used on the current dataset as the MS2 information has been removed in the conversion from RAW to mzData. There is also a method for removing isotopic Peaks *list methods/Isotopes/Isotopic peaks grouper* (not shown).

Online Search

Identification by online searching should be done <u>with caution</u>. We have found that this method often returns many research compounds, drugs and pharmaceuticals which are irrelevant to our plant based studies. For this reason we recommend searching <u>individual peaks</u> using the peak list.

Let's keep things simple and search a single peak:

Right click/Search/Search online database:

	-	ist deisotop	1			A CONTRACTOR OF		1	10.00 August	-
0	and the second second	verage	Identity	Comment	Peak shape	MU53_917_0	36.mzOsta	and the state of t	10 mzData	MU5
~	m/z	Ret time	(instance)		1.000	Height	Area -	Height	Area	1.1.9
	104.107	141	1		Show			*		•
781	148,091	1.52	glubernic acid (Minn)		Search			Search online	e database	
509	150.050	1.44			Export		1	NIST MS Set		
758	547.077	1.43	glutamina (bloot)		Identities Dist using im	lensity Plot mode	- 1	Predict mole	cular formula	10
1310	325.114	1.50			Manually def					•
230	104.871	1.49	pete (M+H)		Delete select	ted rows				•
804	136,055	1.79	trigonalizaa (kimi)		Add new row					
492	127.839	1.55				. 3.000	2.007			

You are presented with a number of online database options: Let's try a search of KEGG. Set the charge appropriate to the technique (here it is ESI+ so we set +H). Set the m/z tolerance and the isotope pattern filter (as before)

latabase	KEGG Compound Database		Database	KEGG Compound D	
	PubChem Compound Database Human Metabolome Database (HMDB)			m/z: 130.050	Charge: 1
Neutral mass	Yeast Metabolome Database (YMDB)		Neutral mass	Ionization type:	+Há * 🔫
	METLIN Database	=		Calculated mass	129.043
Number of results	LipidMaps Database MassBank Database		Number of results	100	
m/z tolerance	ChemSpider Database PlantCyc Database	-	m/z tolerance	0.0010 m/z or	r 5.0 ppm
Isotope pattern filter	r 📝 Setup.		isotope pattern filte	r 🔽 Setup	
radiape passent me	OK Cancel Help		ranne baren ane	OK Cancel	Help

MZmine will start the search and any hits are displayed in a new window. The isotope pattern and structure may be viewed. If you think the structure is an appropriate match then the identity may be added using the *Add identity* button.





The otheralternative is to search the whole list from the *Peak list methods/Identification/Online* database search menu. WARNING If you do this be prepared for many hours of deleting irrelevant peaks!

[PLEASE DO NOT DO THIS DURING THE DEMO FOR BANDWIDTH REASONS !!]

Gap Filling

We will now try to fill in the gaps where peaks were detected in some scans but not others. There are a number of occasions where a peak may be present but not detected well due to being close to the detection limit in some samples. Gap filling is done by searching the target window where a peak was detected and looking for appropriate peak features in that window. There are two options "Peak Finder" or "Same mz and RT range gap filler". Let's first use the "Peak Finder" option. The gap filled peak list will appear as a new item in the left hand pane.

	Aligned peak	list 2	filled	-	
(ш		-	*	
3	28:07 PM]: Finished	gap-fillin	g on Aligne	d peak lis	t
_					
ų.	Please set the paran	neters			
ta	me suffix	gap-filled	í.		
	me suffix ensity tolerance	Consideration of the	%		
nte		gap-filled	í	5.0	ppm
nte n/;	ensity tolerance z tolerance	gap-filled	%	manage 1	ppm
nte n/; Cel	ensity tolerance z tolerance	gap-filled 50 0.0010	% m/z or	manage 1	ppm

Here is the final gap filled spreadsheet peak list. Filled gaps are shown with a yellow icon. There may still be some gaps in which no evidence for the peak was found.

Al estivas ricorgio +	Aige	ed please it	of page block	D'										5	54
30 extinte photosals	0	A	ret lage	Martily -	Caintrant		Page shape		ML83 917	di mellula	MUH7, 917	til euclatu	ML01 917	45.AuGate	MU
38 estilata circolario 21 estilata circolario		112	Retore	- many -	Christian		That shape		. Huight	Avea	Hught	Ana	. Huight	Ans	
The State Investor 488	F	10.00	216	using (tree)					10.1000	5.889	+ 9.000	2,288	+ 1788	1.001	
25 NGORA (FOUND) 30 NGORA (FOUND)		107 (19)	512	unitera (hirita)					+ 3,015	1.00	+110	1965	+ 1088	1.00	
1. reflats months		241.071	8.12	unphie Shirik			-		* 4.025	1.000	+ 1 45	4.023	**002	1.005	+2
16 mgData chromating		MILTOT		andrea (Marriel			-		13.810	1.162	1 2 384	4.942	1.000	1.263	
A suffrage of the suffrage of							-								
25 millats chromatog	•	101.098	8.72	184					4.1.66	+160	+ 9.805	1.004	0.1400	3.49	= 1
11 milliola convenion (481	π	121.047	329	14.00		1					i dett	1.162	+ 402	2.81	= 4
Disches months and	11	101100	1.00	aniset Reside					+ 2014	130	+ 100	1.014	4.004	4 (62	
Ministrate deventer		Aug days	4.76	tennes Mari			1	-	+ 1.201	1.88	+ 1 885	1.468	A 1.006	2188	#1
U subdo drivenio		100.001	***	Same (1995)			-		+ 2.765	6.782	- 1.161	1400	a 1.000	1.400	
18 nuthate through							-								
All millala chromato 183 X. millala chromatogi	11	201.091	1.00	Aprilatian (894)					+2485	1.015	# 1.999	1881	+ 2.685	+289	+2
M-natiata dimension	1	100.100	1.05	tomprer (20ve)	1			1	1.100	2.401	18.003	1.901	5.482	3385	- 1
HEData (Investation)		1011107	110	Augustation (2014)				1	+ 0.35e	1.801	1.140	1.912	# 1.829	1.405	- 1
allada chronigin	14	ARC YOF	8.24	Party I	-		- 1	1	-2.95	8.852	- 2.853	4.162	- 1 102	1872	- 2
Cala strumate all						1		<u> </u>	+ 2 105	1.000	+139	1.001	1.40	viete.	-
a cheurage.				Energy plane (And		-									
da chronato 213	88	110.001	07	Spratta (Mile)	-	1			× 4.204	2.401	- 3.00+	140	a 1.053	10.004	-
Leichda throngin 204		100.000	176	Ingeneting (Ave)	1				a 1 188	3.97	*****	1887	+ 1.000	187	=1
Coulde shares Ing	10	276 162	1.79	mperains (Street					+ 3.665	4.600	+ 1485	3.965		1.885	
Lodes months +								1							

By right clicking on the peak list and *Show/Chromatogram* (*dialog*) and carefully selecting a scan with a detected peak and a scan with a filled peak you can see if the filling has been done in a sensible way. (Right clicking on the chromatogram and selecting *Show /Chromatogram* (*dialog*) brings up all

peaks overlaid but right clicking on an individual peak column and selecting *Show/Chromatogram* (quick) brings up just that peak). In the chromatogram the pink peak is the original and the yellow is the Gap filled by Peak finder. Note: in this case the earlier decision to set the baseline high may be causing a non-optimal integration of larger peaks. Peak picking is always a compromise between detection and accurate peak representation. (From the authors personal experience the XCMS wavelet method seems to be a more robust peak picker in practice).



However the peak finder option can often backfire because it may detect previously removed broad artefact peaks.



This means our carefully removed artefacts are now back with us! An alternative way of looking at this is that we can used this as a way to detect peaks that may be artefacts anyway!

	4	and and a second	tinth.	termet							1.	MART 947	10 reclass	Mint, 917	30 mi	1
1				Comment.				-				Itaget	Auto	D theight	1.8	
	10.000	10			1	10						1188	1.007	# 1.000	180	5
	****					-						4.000	1.000	8.2.990	1.00	
	(8-11)	10		-	1						4	1.000	4.92	4170	10	
		18			1							8788	1.88	1.1.421	1.00	
	10.14*	81					1					1995	4.10	1100	10	
	10.11			-			1					1.001	1.66	4.0.001	1.91	
	in tel	44		-	1							4160	1444	+1.01	185	
	10.044	10		-								1000	1.88	411108	140	
	10.00	14			1							1.111	184	+ 11/64	1181	
1	10.00	18										4794	1.80	44.04	100	
	17.944	-0.			1							1.00	140	a \$100 .	1.97	
	10.11	41								-		0.000	4.99	0.01	19	
	10.470	94		-					-	-		1.07	1.00	1.014	10	
	10.17	47		-					-	-			1.89	4.004	100	
	10.07	87							10	-	-	4.772	180	1.000	100	
	10.00	8.0		-					- 10	-	-	0.084	1.00	+1.01	+0	
	10.15	18.8							1		-	10.080	1,80	+144	10	
	18.219	111								-	-	1464	1.00	+1464	10	1
		-		1												

The other alternative for gap filling is the "Same RT and m/z range gap filler" this limits the gap fill to features within the original detected peak window. This results in much cleaner results.

oject Raw data methods	Peak list methods	Visualiza	ation Windows Help
HU51_917_45.mzDate	Peak detection	n + d_NA	
HU48_921_19.mzData			Peak finder
MU49_917_58.mzData	Isotopes		Same RT and m/z range gap fille
HU52_917_21.mzData		D_NA	
	neters gap-filled SameRTr	mz	
Name suffix	gap-filled SameRTr	mz or 5.0	ppm
Name suffix m/z tolerance	gap-filled SameRTr 0.0050 m/z		ppm
Please set the paran Name suffix m/z tolerance Remove original peak list	gap-filled SameRTr 0.0050 m/z		ppm

ID	As	verage	identity - Comment			Bard Shares	MU47_917	10.mzData	MU48_917_	39
	m/z	Ret.time	identity -	Comment		Peak shape	Height	Area	Height	T
5771	131,090	12		- 20	S				•	+
5772	138.087	11.6				1	<u>≻</u> 5.1E4	1.365	5.284	1
5772	136.087	11.6					* 5.1E4	9.5E4	6 5.364	
5774	136 087	11.7			14		0-5.7E4	1 865	<u>= 5.064</u>	
\$775	136 087	13.0			-		5.6E4	4.7E5	= 5.8E4	4
5776	138.055	6.9					•			
\$777	138.055	9.0					× 6.9E3	4.683	8.4E3	•
5778	139.968	10,1					•		2.165	
5779	141.959	3.7					= 4.2E4	0.0E4	4.164	4
5784	146.980	14.0					6.4E4	1.9E5	- 4.6E4	
5785	147.072	1.4					4.764	6.584	6.064	1
5791	158.961	0.3					5.5E4	2.565	= 5.4E4	2
5793	159.805	6.5			7		8 2.963	2.003	= 7.4E3	5
5795	177.007	4.8			1		≥ 6.9E3	4.0E4	1.263	é.
5798	154 996	12.2					9.0E4	8.7E5	6.464	1
5795	154.905	12.4					. 7.124	1.365	3.954	4
5000	154.986	12.6					7.7E4	3.6E5	4.664	-
5801	104 305	13.8					# 7.7E4	3.265	4.554	1
	ing ann	·			1				4	

If we look at the same example as above we can now see the detected peak is now cut rather than detected in full. In such cases the hope is that the peak cut-off is applied consistently across all peaks to preserve relative quantitation. Again a compromise is made.



Base peak plot, MS1, m/z: 160.0292 - 160.0298

net.sf.mzmine.modules.peaklistmethods.gapfilling.samerange.SameRangePeak@ff8e75c _ 160.0295 m/z @2.28 [QC1_917_01.mzData] 160.0295 m/z @2.28 [QC1_917_01.mzData] — QC1_917_01.mzData — VVT48_917_40.mzData

Looking again at one of the artefact peaks notice the gap filled peak is now defined by the mz tolerance and RT tolerance. The gap filled is broader than the original peak due to the RT window which is defined by all the peaks in the row. The variation in the retention time across the row is a function of the earlier tolerances used in both RT Normalisation and Join Align. This demonstrates the need to be careful when setting up the parameters from the very beginning.



Despite the limitations of gap filling it is far preferable to have some estimate of baseline levels than to report the value as missing for later statistical analysis.

Export of results



The final exported data is displayed in excel. At this stage it is probable that the data will be processed further in a commercial data analysis package but as we will see in later examples there are some possibilities to use open source tools to analyse metabolomics data.

		- 14		-	-	_				
A sea l'asse l'asse		-		1000						
Carthey Software 14	1 16-16		1 10 1 10 10 10 10 10 10 10 10 10 10 10	-	- N	1.22	100	-	11.4	7 B.
-	24					-	14	14400 ···	1412	in the h
1	-						part - 17	roug -		1.1.1.1.1.1
14 C			and a local	And in case		100 Aug		100	- 8	-101/0
AL	de ine	£3								
A DESCRIPTION OF THE OWNER	1		1	1.1.1	and shared and					6 A
No. of Concession, Name of Street, or other	an easily the side of	-	NAME AND ADDRESS OF TAXABLE	Allowide Links	alan di	महामगर	1000	108.57	Di Ilric	17451
THE INTERN LARDIN	1.8					100-1-1		101.000		
of which then	-					And in the		And inte		atta marite
35 254 9400						Linest eve		制化	100.00	約2 時間
TO WINE CARE	1					MARLIN BOA		집감		the second
10 x54 1100						MARCH STREET		812	1010.4	23.7 104.14
OR WING LINEY						Panel Contra		212		412 0000 Mail 100
DE 254 1781						Y00147 2668		101.00		1.14 100210
THE REAL PROPERTY.						(040) R. 7(0) (600) P. 1800		200.00		101 10.1
the maintain a result			224	and a state		form y new	Change 1	ALC: NO.	140-16 47	This is straight of
IL MORE LADO		204	TOD, NAPP, SERVICE	CHEMPSE.	A. 10, P 444	WHEN DO		101108		NUM OF THE
tan muccui to here						14764 W 2012	21 22214	46/62-17	1008.85 (7	Phy 10 101001 48
the Wolcow Street						2011 0 202		1008.08		Ing of Concerns
THE MOMENT TITLE	- F					ALC: 1010	10.00001	-0101	100.00	##4 1418 I
Col Miland College						TABLE & ADD		1000.0		418 750 B
MARKER COMP.	1					status the	11.03454	19854	100014-00	MAL COM
N BAR LM						1040.1 100		했다		CODE POINT
NU WORK 14278						22764 1776	TT BATTLE	96211	ments (Pag. 120714
WEITER THE	18					150-07 150- March 1000		128-07		8-17 19-17 80.1 10000
No. 2 Lot 1 Lots	- 2					TRACK LAND		10084		the arts.
W1 8'09 1525						447.62		200811		this graph
HE 8'200 7270						1002410.0000		1228.1		TIPE MALL
NE 1672 1540	18					200111 40-07		4000.4		white methics
RI BOAN TATA						ADD N NO.		4000.0		10.21 JANE
81 BUN 136						1604.254		10001		THE REPORT
No BOLY LAD						WHEN PER-		And a		10115 _ 17006
TE 0.001 1400	28 1	19	NuMER internet	10041	anani-take	intera Auto	64 Alter41	010817	DOM: N	YEAR 21000-
ST BOAL DOD MANUEL	- 6					TIME NO		Marrie I		1964 PL18
915 (E1441 TW94						202041 2224	14 219405	12756	26/58 22	0HR 20HC1
\$1.00 million (1920)						20841 728		1287		2379. GART?

Batch analysis

MZmine contains a Batch mode tool which allows a chain of processes to be set up which is a very useful feature with large datasets as the processing can be set to run overnight in unattended operation. The output of the previous operation is fed to the next operation.

Another useful feature is that once the parameters for a particular operation have been set up MZmine remembers the last used settings so we can apply the peak picking we developed above to every sample in a study.

The MZmine Batch command is to be found under the Project menu. The screenshot below shows a typical sequence.

	Load	Save		
Batch queue	Raw data imp Mass detection		-	Configure
	Chromatogram	222		Remove
	Chromatogran Retention time	m deconvolution	=	Clear
	Join aligner	e normanzer		
	Custom datab		-	
		▼ Add		
	Adduct search		·	▼ Add

A recently added feature is the ability to save or load batch sequences.

As an exercise try adding the steps you went through above to a new batch sequence.

Baseline Correction

Another new feature is baseline correction. This is applied to the raw data before peak detection. It is found under *Raw data methods/Baseline correction*.



The baseline correction dialog box has two main options, the Smoothing and Asymmetry factors. Try playing with different settings of these factors and comparing the TIC plots before and after (NB: Ensure the "Remove source file after baseline correction" is switched OFF

Filename suffix	baseline-corrected	1127 04:00 G400F
Chromatogram type	TIC 🔹	1007
WS level	1 -	A MANAGE AND AND AND AND AND
imoothing	1,000	2.005
symmetry	0.1	4.00 mm
lse m/z bins	1	A res
s/z bin width	1.000	3.00 (m.m. 190.00 (m.m.
Remove source file after baseline	conection 🔲	
OK	Cancel Help	1.00 2.00 3.00 4.00 5.00 9.00 1.00 1.00 1.00 1.00 1.00 1.00 1
Towns of the second sec		ACH JIT IN ACCOUNT AND ACHT JIT IN ACCOUNT AND A ACTIVATION

Initial data analysis in MZmine

Limited data analysis tools are included in MZmine. In order to use them it is necessary to 'Set Sample Parameters' under the project menu. You can define as many parameters as you wish.

In the example data we have an excerpt of a metabolomic study on the ripening of fruits. We have nine samples of two different varieties, Wild-type and non-ripening Mutant plus ten control samples which consist of a large batch of identical fruit extract that are run at every fifth sample. In addition the fruit are sampled everyday from the onset of ripening between 47 and 54 days. (This example datasets is only a small excerpt of a larger replicated study).

We now set a new experimental parameter called Type with the values "Wildtype", "Mutant" and "Control"

And an and a second	
Add experimental parameter	
Name Type	
Numerical values	
Set of values	
Wildtype Hutant Control	
Add value Remove value	
Free text	

Raw data	Type	
wan.80, 510, 58, may	Edani	
West \$17, 45,mm	Euleri	
wan 80, 719, 54, 64, 64	Butard	
Wer, 12, 110, 28, M	Rutari	
Ward 917 58 me	Rutert	
WUNE 921 15/18/1	Rutert	
WUSE 917 33 rev	Futert	
wantin Tre Teute	Rubert	
W154 (FT) 28 mer	Rubert	
ace in Fide rew	Control	
QC2 917 02.raw	Control	
QC4 917 16/8W	Control	
QC10_917_81/ww	Control	
GC1_917_01.rew	Central	
GC8_917_18.rem	Central	
0C3_917_08 rew	Control	
GC7_917_38.rpw	Control	
008_917_A1.rew	Control	
UT50_017_32 rew	Withore	
uff47_017_15.mm	Wittype	
9745_917_40.mm	Wittype	
1749_917_59.mm	Wittype	
1740_821_17.mm	Wittype	
0C8_917_48.raw	Control	
151_917_43.sev	Wittype	
W152_817_06.4eV	Wildype	
wms2_917_20.4avi	Widge	
UT56_917_28.464	Walkype	
Add new parameter	Import parameters and values Remove selected parameter	

MZmine has a number of data analysis options, *Coefficient of variation (CV) analysis, Log Ratio Analysis, Principal Component Analysis, Curvilinear Distance Analysis, Sammon's projection,* and *clustering.* The most useful of these options are described below:

Coefficient of variation analysis	Select files for analysis	
calculates the coefficient of variation of each peak and displays the result as a	MU53_917_06.raw MU54_917_26.raw QC2_917_02.raw QC3_917_08.raw	
colour coded plot. <i>(Ensure you have</i> your final peak list highlighted)	W114_391_1538W > 0C4_917_1678W W174_997_91078W CC5_917_1978W CC5_917_1978W W174_997_997W E CC6_917_2478W W175_977_3278W CC7_917_3078W CC6_917_4478W CC6_917_4478W W175_977_3378W W175_977_3078W CC6_917_4478W CC6_917_4478W W175_977_307W CC7_917_3078W CC9_917_4478W CC9_917_4478W W175_977_5078W CC9_917_3078W CC9_917_4478W CC9_917_4478W W175_977_6678W CC9_917_3078W CC9_917_4478W CC9_917_4678W	
In this case we have selected the 10 control samples. The graph shows that	WT54_917_29.raw Peak measuring approach height area OK Cancel	Help
some of the later peaks are showing some unexpectedly high variation.		





Other features in MZmine

There are many more features in MZmine, including some support for ms/ms data and formula prediction. More features are being added all the time, recent developments in this area include links to the NIST MS Search program to allow the use of MZmine for GC-MS data.

Get Involved !

Please join the community! Not just for programmers - testers and document authors always appreciated!

Developers Mailing List:

http://sourceforge.net/mailarchive/forum.php?forum_name=MZmine-devel

Acknowledgements:

I would like to acknowledge the following people for help with the current round of MZmine development.

Tomas Pluskal, Matej Orešič, Mikko Katajamaa, Carsten Kuhl, Ralf Tautenhahn, Steffen Neumann Christoph Steinbeck, Stephan Beisken, Chris Pudney, Mark Seymour, Mark Forster, Martin Cip, Dave Portwood, Aniko Kende, Madalina Oppermann, Erik Johansson, Johan Trygg.

I would particularly like to thank those in the XCMS community who were very gracious in allowing us to combine some of the features into MZmine and to Chris Pudney our hardworking programmer.

Mark Earll mark.earll@syngenta.com

Wednesday, 16 May 2012