Introduction to Metabolomics

Introduction to metabolomics

What is it and what are its applications?

mpd39@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Department of Plant Sciences

Focus of the two sessions:

The aim of this course is to provide an overview of the applications, laboratory equipment and online bioinformatic portals for metabolomics research.

Plant Bias – all techniques transferable to other organisms

Introduction to metabolomics– part 1:

09:30 – 10:00 – **Introduction to metabolomics** – what is it and what are its applications?

10:00 – 11:00 – **overview of techniques** – targeted and non-targeted metabolomics (metabolite extraction procedures, equipment – GC-MS, HPLC-PDA-MS)

11:00 – 11:20 – **coffee break**

11:20-12:30 – **how to identify a metabolite** – online practical

12:30 - 13:30 – lunch

Introduction to metabolomics – part 2:

13:30 – 14:00 – **Introduction to chemometrics** – how to analyse many metabolites

14:00 – 15:00 – **online practical involving multivariate statistics**

15:00-15:15 – **tea break**

15:15-15:45 – **metabolic mapping** – identifying proteins and genes associated with your metabolites

15:45-17:00 – **online practical** - **metabolic mapping** – identifying proteins and genes associated with your metabolites

What is the metabolome?

- Total quantitative collection of chemical compounds (metabolites) present in an organism *eg. sugars, amino acids, phenolics, lipids*
- Not proteins or peptides
- Highly complex
- Thorough and unbiased assessment of all metabolites within an organism

Complexity

- **Physically and chemically complex**
  Large range of molecular weights 10's to 100's MW (and size)
  Polar and non-polar metabolites
  Volatiles

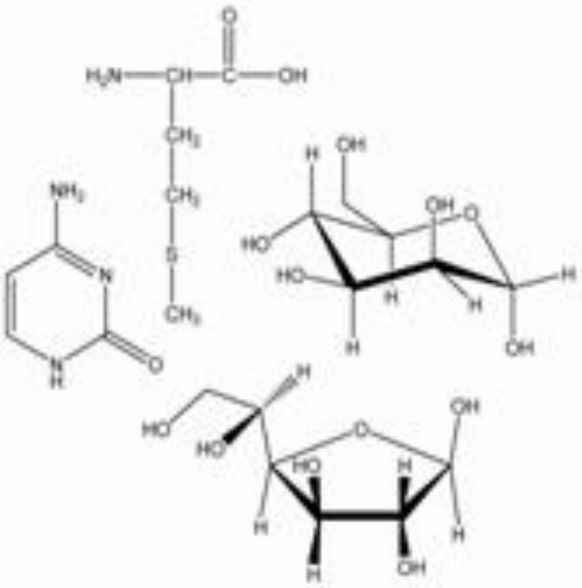
- **Variation in number of known metabolites per species**
  Yeast *Saccharomyces cerevisiae* (584)
  E. coli (436)
  Plant kingdom (up to 200 000)
  Human (2900)

- **Very wide concentration range**
  mM to sub-pM

- **Temporal changes** - flux

What is the metabolome?

http://www.biomics.se/



GENES/mRNA

PROTEINS

METABOLITES

TRANSCRIPTOMICS

PROTEOMICS

METABOLOMICS

"Metabolomics" first appeared in the literature in 1998 (Fiehn et al. 2007. *Metabolomics*)

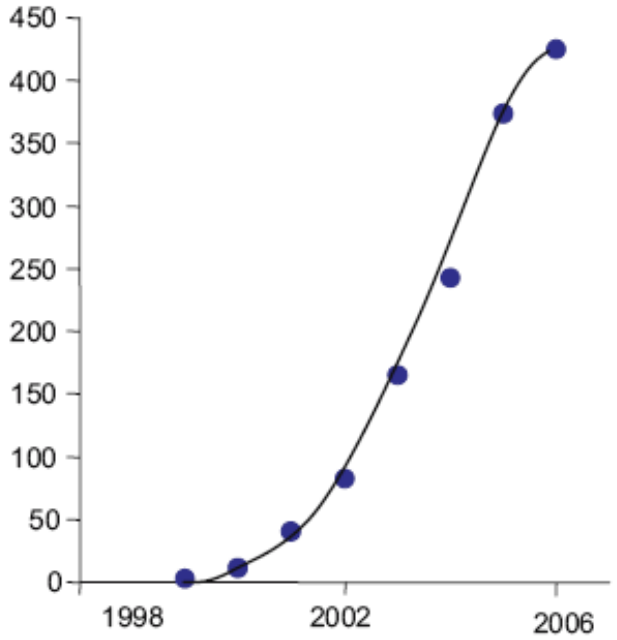


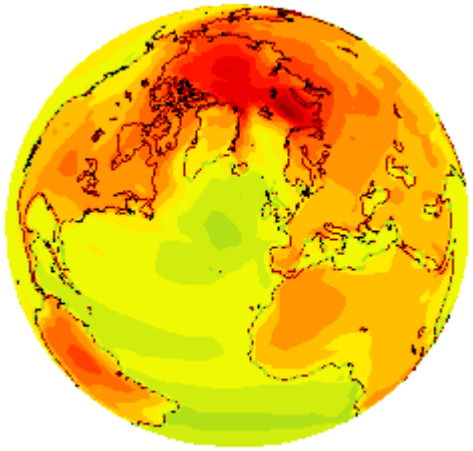
Fig. 1 Number of publications per year based on ISI Web-of-Science database query using the search phrases *metabolom** OR *metabonom**

Measuring many metabolites is nothing new, but…the ***scale*** of the analysis is

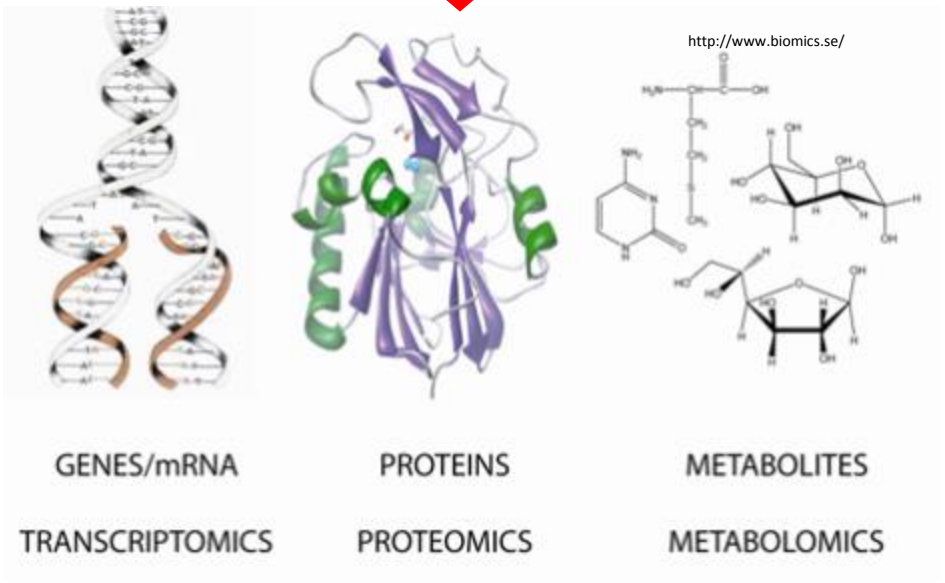
Rather than look at individual reactions to understand an organism (***reductionist theory***) an attempt is made to measure the whole system (***systems biology***)

Why study the metabolome? – direct link between genetic and environmental signals

Environment



Expression



Assign
Function
Response

Why study the metabolome?

- Need to understand metabolites and metabolic pathways before we can exploit them
- Metabolic status of cells provides a clearer indication of health than mRNA or proteins
- Advance systems biology

Applications

Trait development in crops
eg, salt and drought tolerance;
defence; photoprotection

Genetic engineering
safety – substantial equivalences

High value products – cosmetics; medicine

Plant disease biomarkers

Plant **population** / **evolutionary** studies

Biofuels

Biomarkers for:
Disease
drug intervention
environmental stress

Nutrigenomics

Personal health assessments

Personalised medicine

Metabolic engineering

Why study the metabolome?

Examples of early applications:

Diagnosis of coronary heart disease using metabonomics
Brindle, J.T. et al.(2002). Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nature Medicine, 8, 1439-1444.

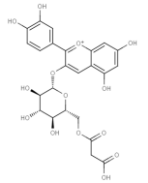
Metabolite profiling for plant functional genomics
Arabidopsis thaliana – model species. Quantified and identified many metabolites and related different genotypes to their metabolic profiles (by GC-MS)
Fiehn et al.(2000). Metabolite profiling for plant functional genomics. Nature Biotechnology, 18, 1157-1161.

Why is this application important in natural systems?

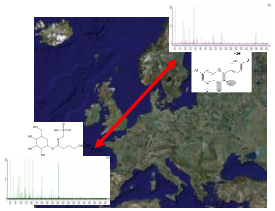
Plants
Abiotic
Biotic

Grow
(morphological traits)

Function
(biochemical traits)



Time
Population
spread



Natural selection (traits related to fitness eg, survival, reproduction)
Local adaptation (variation in traits)

Very little is known about variation in such metabolic traits – how do we measure them?

Techniques used in measuring the metabolome

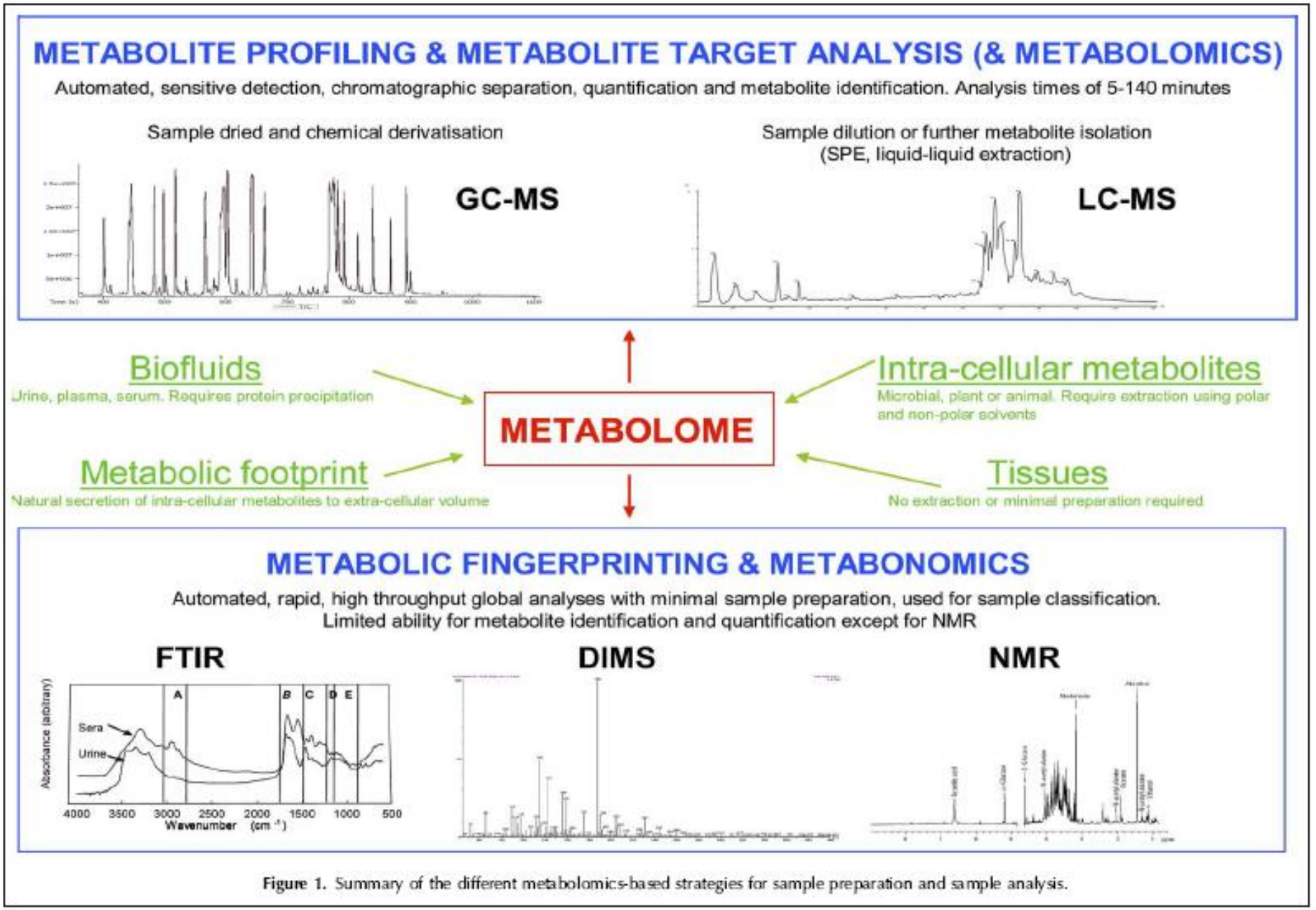


Figure 1. Summary of the different metabolomics-based strategies for sample preparation and sample analysis.

Dunn et al. 2005

Sample collection
STOP metabolism

- Cold methanol
- Hot Ethanol or Methanol
- Freeze clamping for plants
- Liquid nitrogen (-196’C)

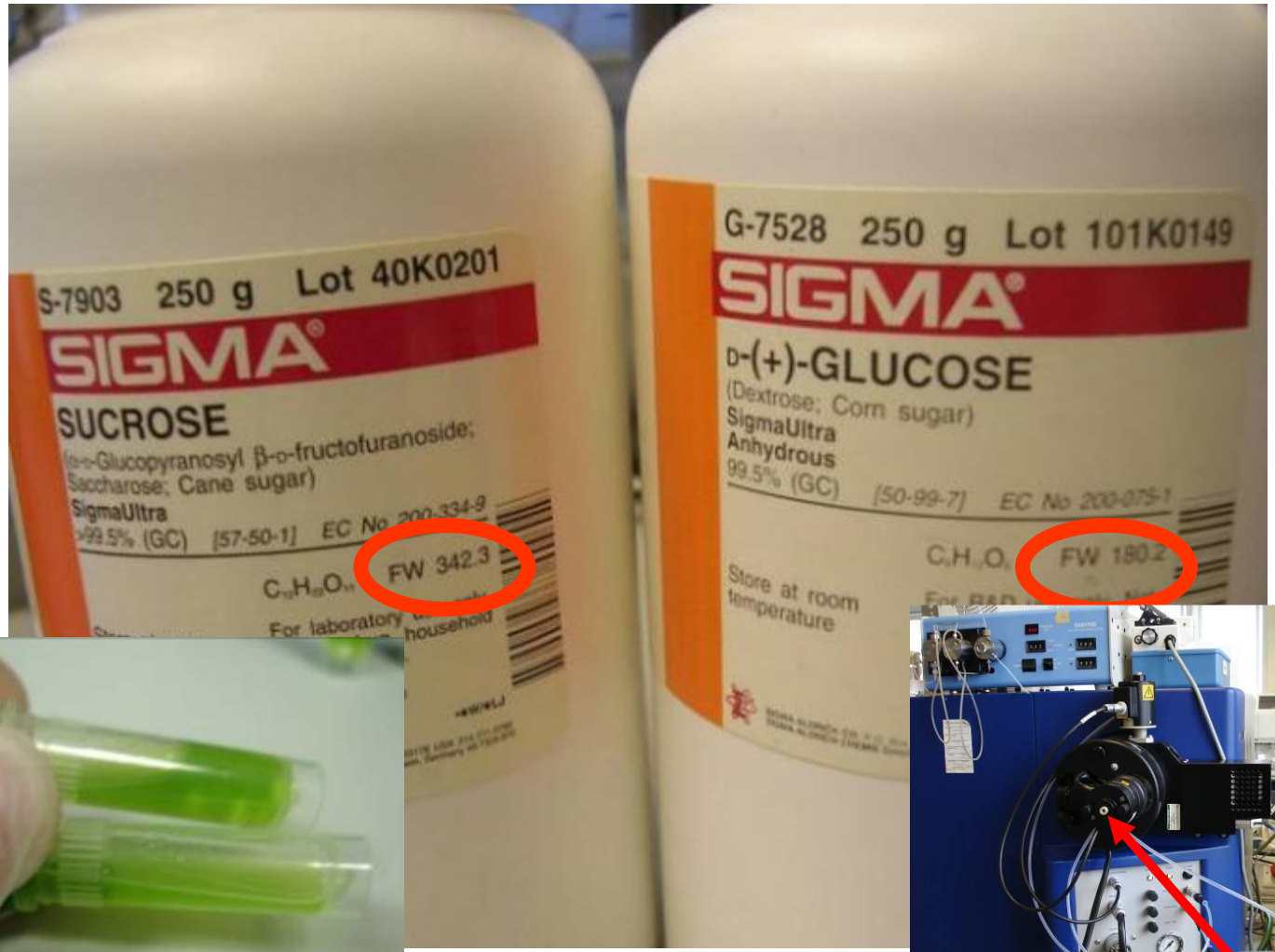
How to get metabolites out of the cell
Solvent extraction and storage

- Methanol (hot or cold)
- Methanol/chloroform/water
- Hot ethanol
- Ball milling or grinding with mortar/pestle
- Store at -80’C

Detecting metabolites – Metabolic Fingerprinting

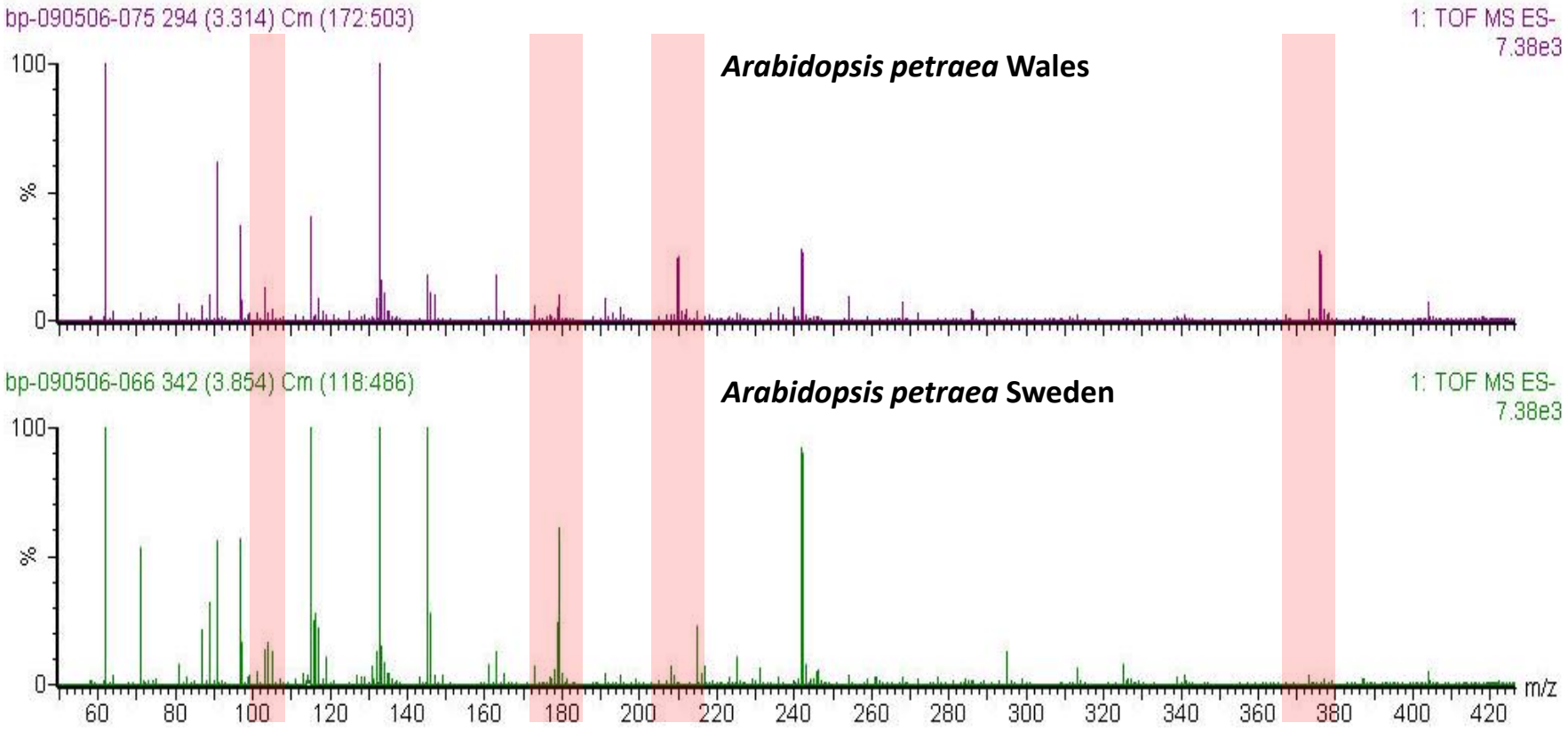
High throughput screening
for metabolic phenotypes

Mass Spectrometry (MS)
Nuclear Magnetic Resonance (NMR)
Fourier Transform Infra Red (FT-IR)



Direct injection of crude extract

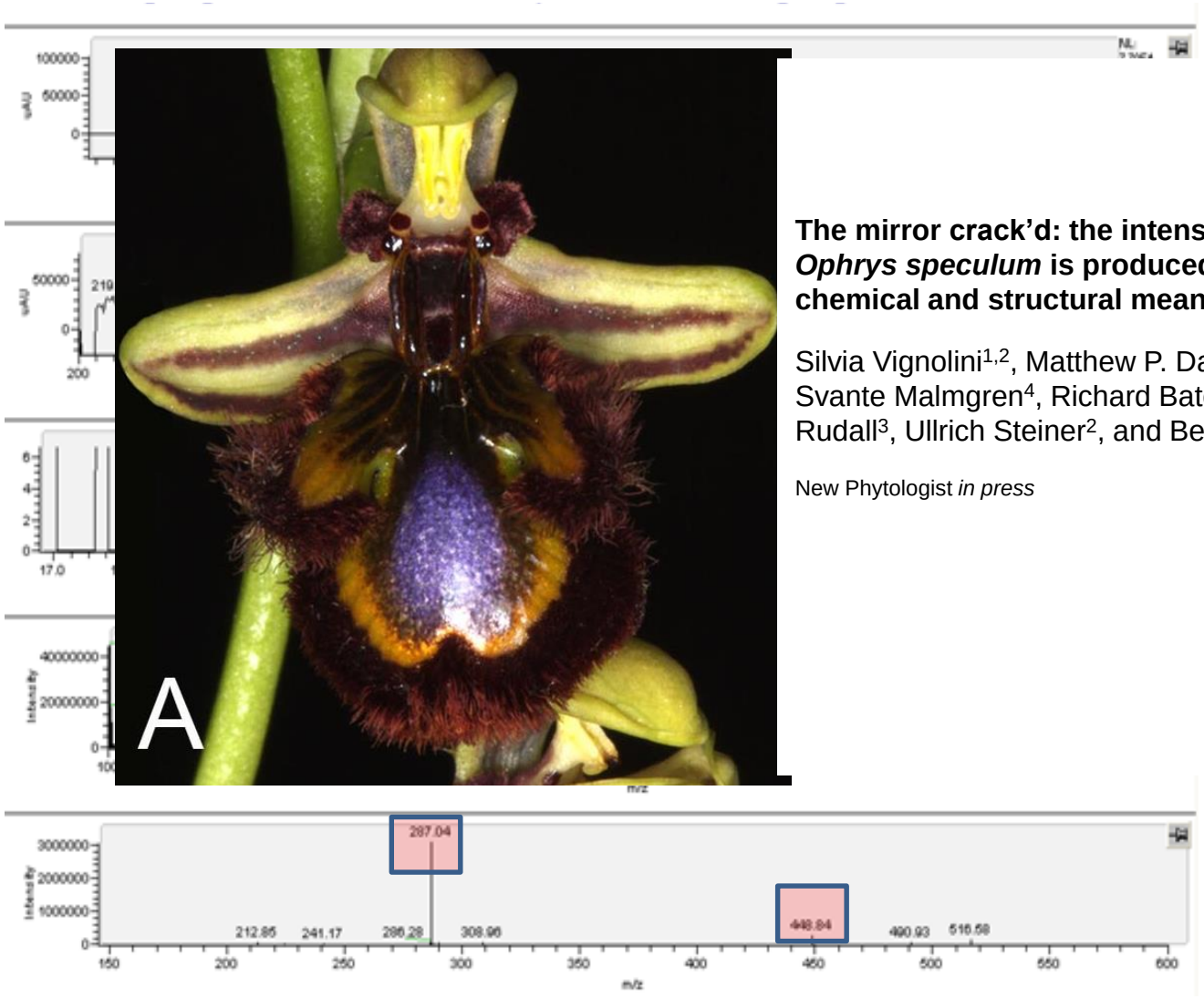
Metabolite fingerprinting



Davey et al. 2008

Detecting metabolites – Metabolic Profiling

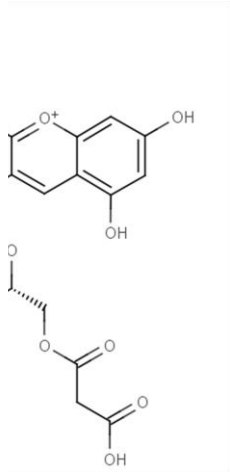
High Performance Liquid Chromatography (HPLC) – Photodiode array (PDA) – Mass spectrometry (MS)



The mirror crack'd: the intense blue colour of *Ophrys speculum* is produced by both chemical and structural means

Silvia Vignolini[1,2], Matthew P. Davey[1], Julia Tratt[3], Svante Malmgren[4], Richard Bateman[3], Paula Rudall[3], Ullrich Steiner[2], and Beverley J. Glover[1]

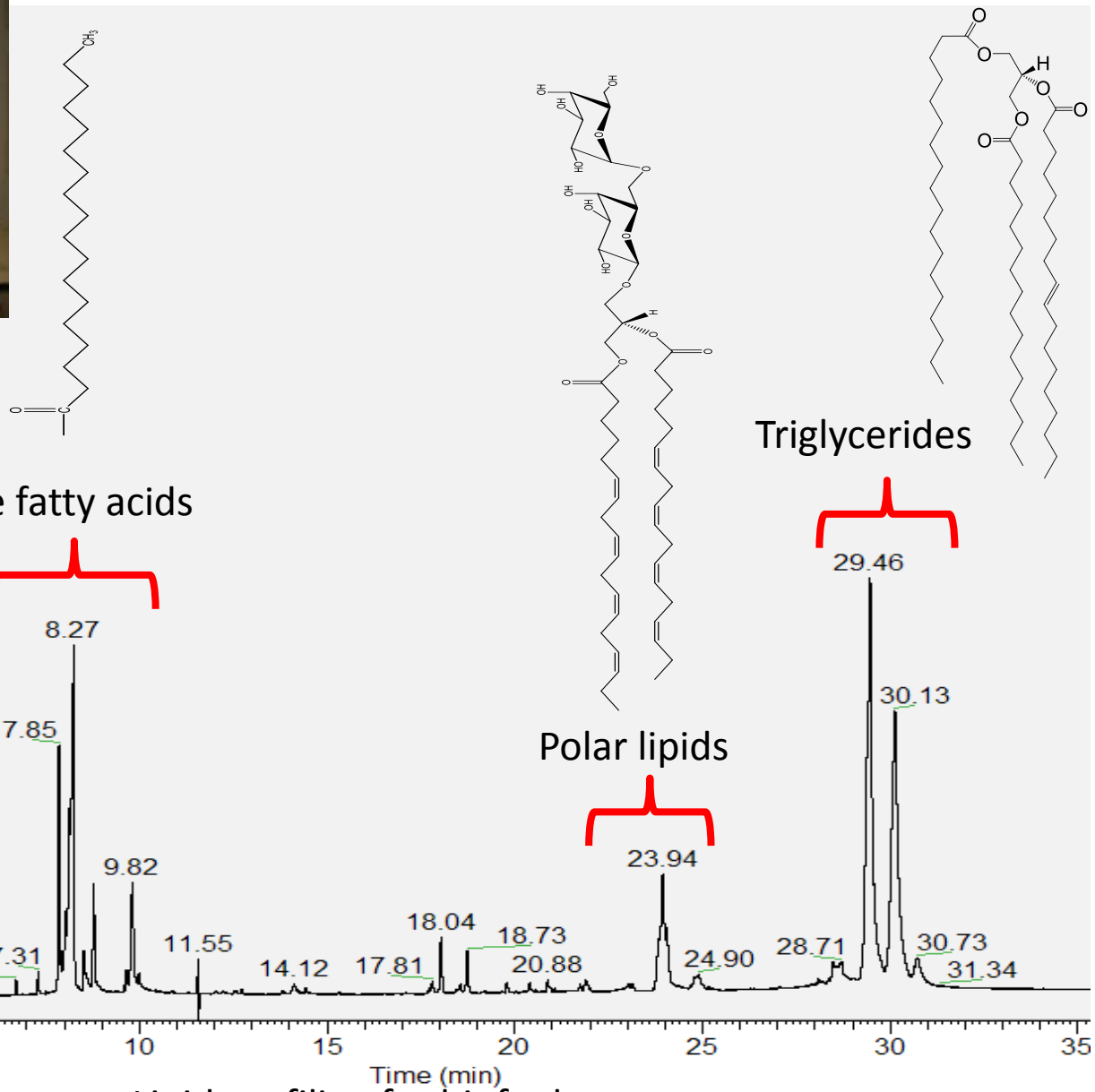
New Phytologist *in press*



MS/MS

Cyanidin 3-(3-malonyl glucoside)
534.90
448.9 (-malonyl)
287.04 (-hexose and malonyl)

Metabolic Profiling - Gas Chromatography (GC) – Flame Ionisation Detector (FID)



Lipid profiling for biofuels

Mass
(Bin)
(1000's)

Output files – can you see the difference?

Need multivariate statistics

codes	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day accl	2 day acc
pop	Helin	Helin	Helin	Helin	Helin	Helin	Helin	Helin	Helin	Helin	Helin
59	0.008	0.0225	0.0093	0.0137	0.0128	0.0221	0.0223	0.0075	0.0176	0.0073	0.0108
60	0	0.0046	0	0.003	0	0.0067	0.0052	0	0	0	0.0027
60.5	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0
61.1	0	0	0	0	0	0	0	0	0	0	0
61.2	0	0	0	0	0	0	0	0.0011	0.0008	0	0
61.3	0	0	0.0016	0	0	0.0012	0	0.0025	0.0034	0	0.0009
61.4	0	0	0.0014	0	0	0.0013	0	0	0.0023	0	0
61.5	0	0	0	0	0.0518	0.026	0.0514	0.0789	0.135	0	0.0922
61.6	0.1223	0	0.0483	0	0.0564	0.1332	0	0	0.0529	0.0727	0.0617
61.7	0.0542	0.0629	0.03	0.0427	0	0.0704	0	0.0914	0.1323	0.0437	0
61.8	0.1164	0.023	0.0419	0.063	0.0456	0.0964	0	0	0.185	0	0.1583
61.9	0	0.0406	0.1193	0.1481	0.1525	0.0937	0.0777	0	0.0427	0	0
62	12.0269	5.7907	8.0946	8.7906	9.8656	9.3396	5.5138	17.1029	14.9041	14.5248	18.8468
62.1	0.0102	0	0.0359	0.0153	0.0083	0.0116	0.0112	0.022	0.0094	0.0234	0.0136
62.2	0.025	0	0	0.0112	0	0.0276	0	0	0.0579	0.0365	0.052
62.3	0	0.0118	0.0356	0	0.0237	0.0471	0.01	0	0.0123	0.0231	0
62.4	0.0198	0.0132	0.0418	0.0425	0	0.0479	0.0139	0.0727	0.0238	0.0495	0.0186
62.5	0	0.0066	0.03	0.0122	0.0321	0.036	0	0	0	0.0317	0.0467
62.6	0	0	0	0.0298	0.0134	0.0125	0.0138	0.0606	0.0367	0	0
62.7	0.0139	0.0033	0.0055	0.0114	0	0.0136	0.0094	0.0649	0.0094	0	0.0445
62.8	0.0104	0.009	0.0105	0.0139	0.025	0.0138	0.0041	0.0315	0.0088	0	0.0294
62.9	0	0.0091	0.0048	0.013	0	0.0229	0.003	0.0118	0	0	0.0391
63	0.0803	0.0337	0.0665	0.067	0.0718	0.0661	0.0386	0.1276	0.0973	0.106	0.1092
63.1	0.006	0.0047	0.0042	0.0042	0.0059	0.0104	0	0.0209	0.0077	0.0063	0.0145
63.2	0.0038	0	0	0.0031	0.004	0.0036	0.0019	0.0171	0.0128	0	0.0244
63.3	0.0058	0.0025	0.0043	0.003	0	0.0027	0.0035	0	0.0098	0	0.0093
63.4	0.0105	0.0026	0.0024	0.0055	0.0109	0.0079	0.0008	0	0.0235	0	0
63.5	0	0.001	0	0	0	0.0104	0.0012	0.0083	0.0077	0.007	0.0056
63.6	0	0.0033	0.0041	0	0	0.0059	0.0014	0.006	0.0111	0.0058	0.0071
63.7	0	0	0.0024	0	0.0017	0.011	0	0.007	0.0049	0	0.0058
63.8	0.0059	0	0	0.0031	0.004	0	0.0012	0	0.0104	0.0094	0.0062
63.9	0.0021	0.001	0.0028	0.0027	0.0016	0.0016	0	0	0.0033	0.0044	0
64	0.1049	0.0463	0.0792	0.0779	0.0873	0.0831	0.048	0.1639	0.1225	0.1366	0.1583

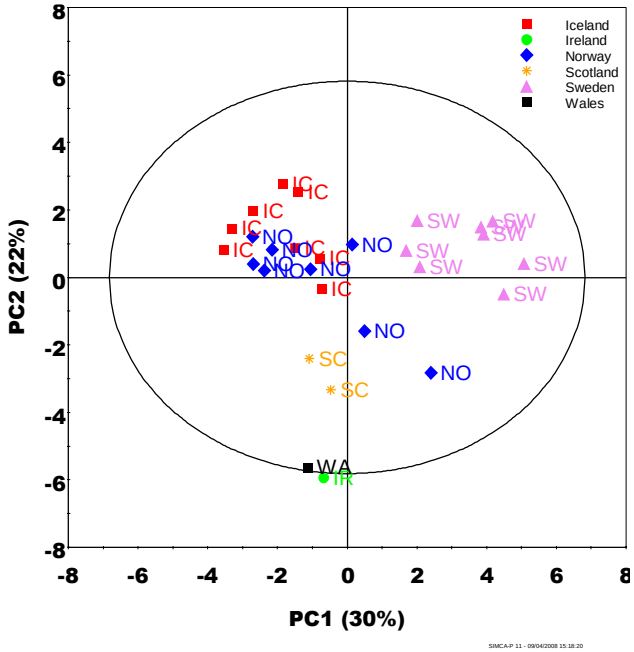
← Sample name (100's)

← Intensity
Total Ion Counts
(1000's)

Multivariate Data Analysis

Unsupervised

Principal Component Analysis (PCA)

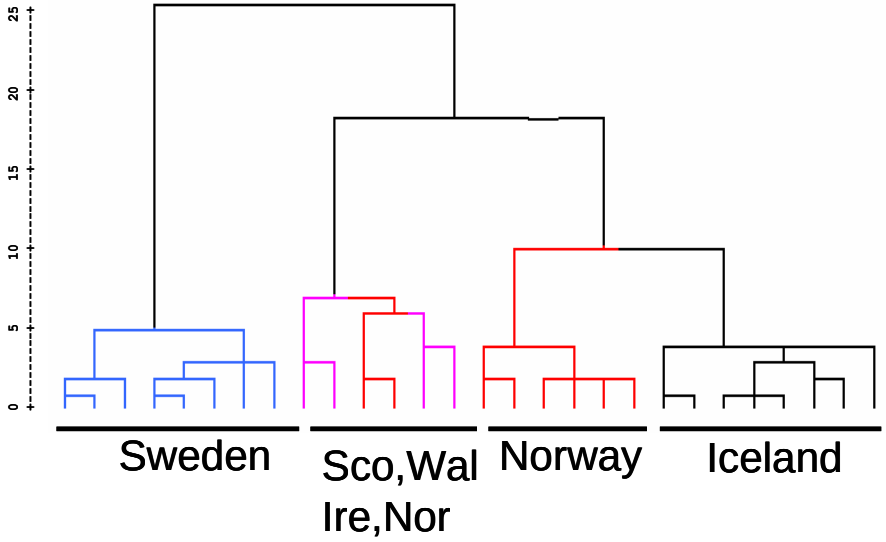


Supervised

Partial Least Squares

-Discriminant Analysis (PLS-DA)

Hierarchical Cluster Analysis (HCA)



Trygg et al. 2007

Metabolite matches and mapping based on mass matching

Large number of online databases for metabolite and network mapping

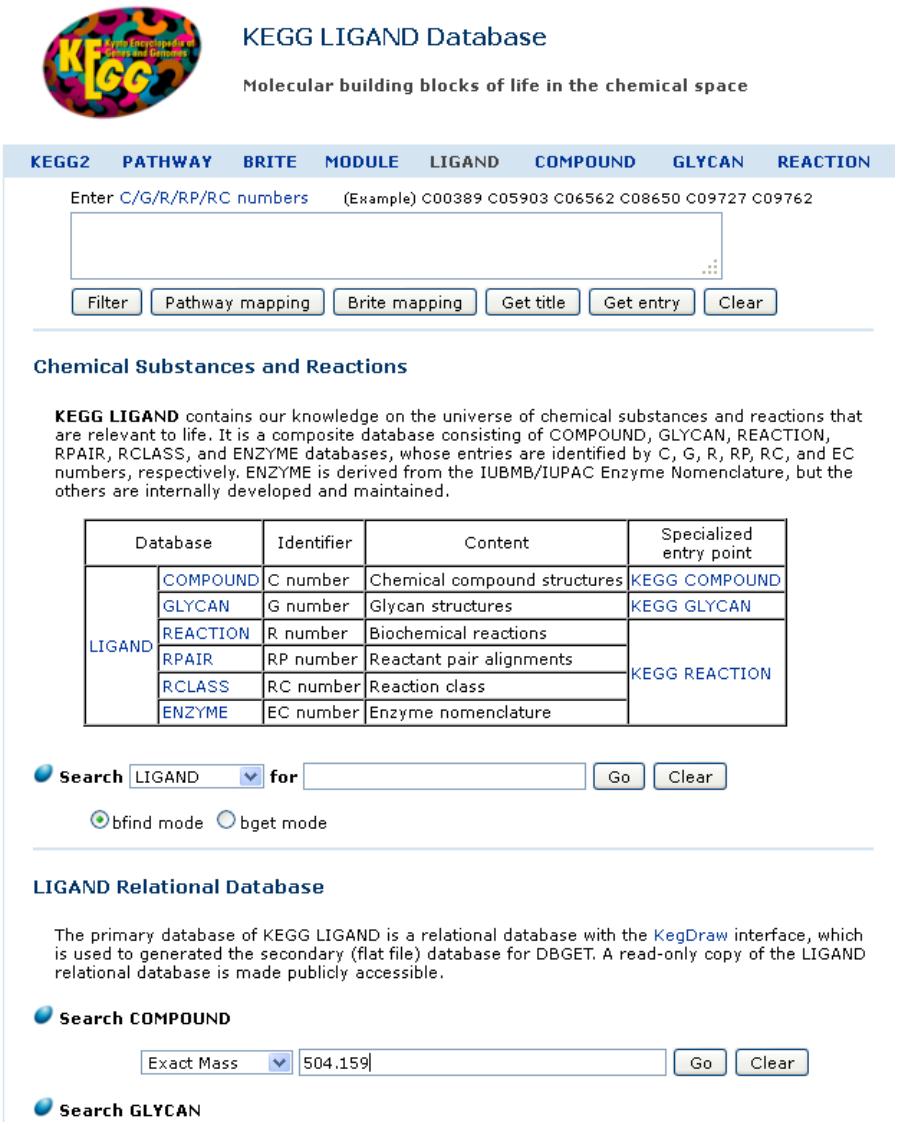
Updated yearly in Nucleic Acids Research
Galperin and Fernández-Suárez 2012

Brown et al. 2009

KEGG is the most widely used/known site for metabolic mapping
– there are errors but getting better!

Another common site is MAPMAN

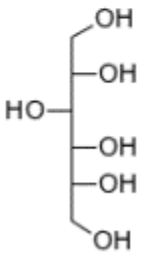
Eg. Search mass 504.159 →

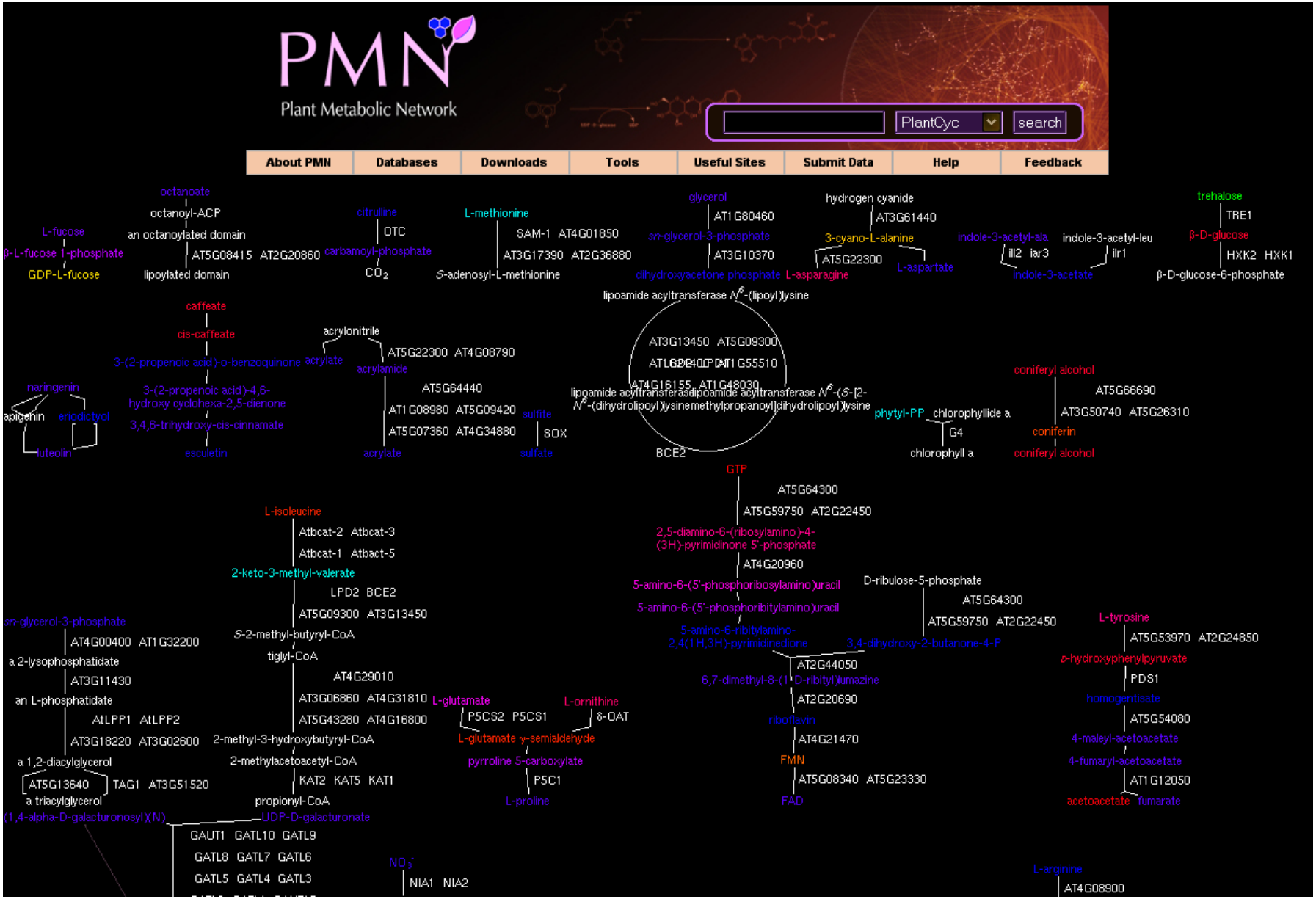


KEGG

COMPOUND: C00794

Help

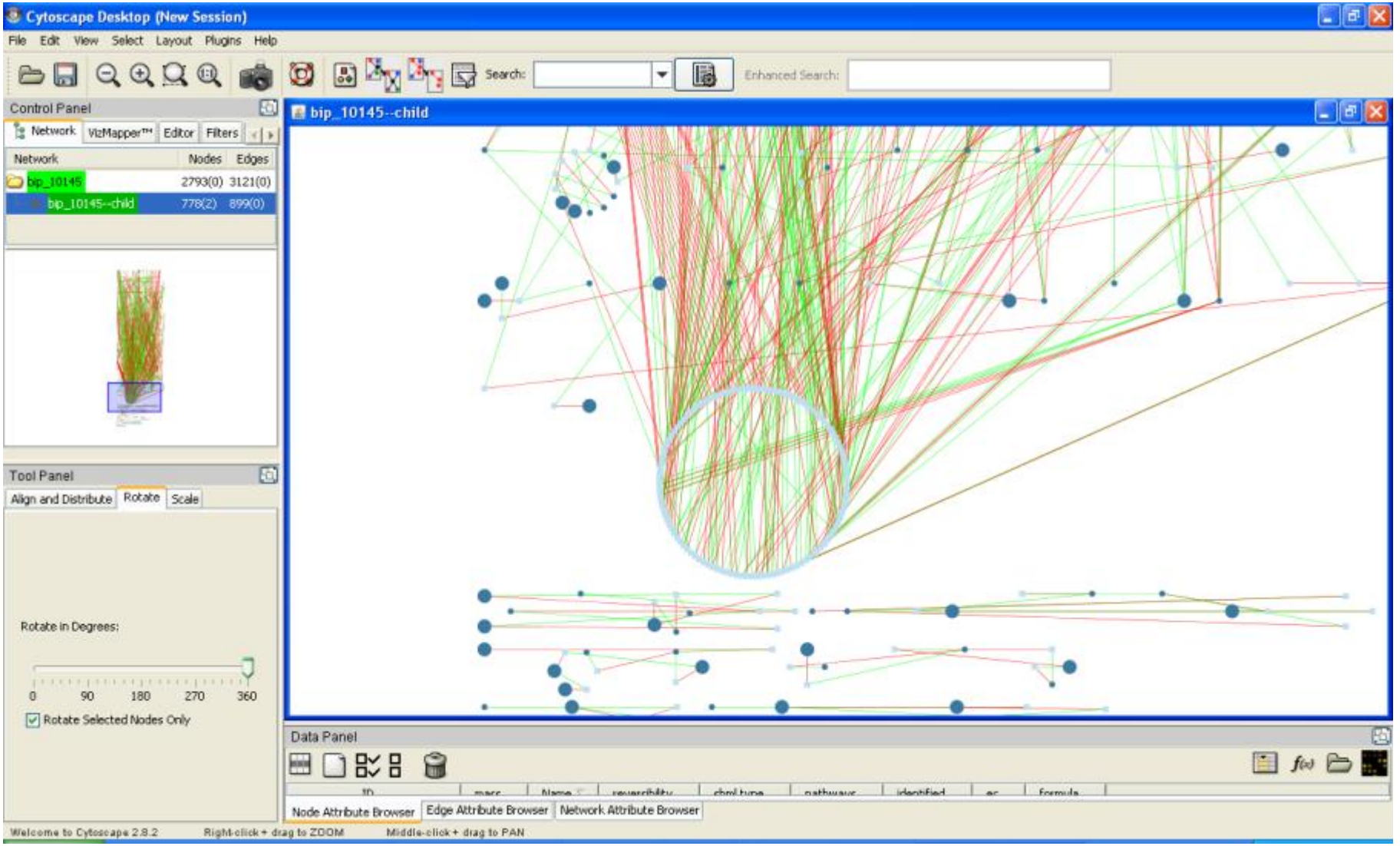
Entry	C00794 Compound
Name	D-Sorbitol; D-Glucitol; L-Gulitol; Sorbitol
Formula	C6H14O6
Mass	182.079
Structure	 C00794 Mol file KCF file DB search Jmol KegDraw
Remark	Same as: D00096 BRITE hierarchy
Reaction	R00874 R00875 R01697 R01787 R02865 R02866 R02867 R02868 R02925 R02926 R05820 R07346
Pathway	PATH: ko00051 Fructose and mannose metabolism PATH: ko00052 Galactose metabolism PATH: ko01100 Metabolic pathways PATH: ko02010 ABC transporters PATH: ko02060 Phosphotransferase system (PTS)
Enzyme	1.1.1.14 1.1.1.15 1.1.1.21 1.1.1.289 1.1.99.21 1.1.99.28 2.7.1.1 2.7.1.69 3.1.3.50 3.2.1.22
Other DBs	CAS: 50-70-4 PubChem: 4052 ChEBI: 17924 KNApSAcK: C00001173 3DMET: B04724 NIKKAJI: J2.299C
KCF data	Show


PMN
Plant Metabolic Network
PlantCyc
search
About PMN
Databases
Downloads
Tools
Useful Sites
Submit Data
Help
Feedback
octanoate
octanoyl-ACP
L-fucose
an octanoylated domain
β-L-fucose 1-phosphate
AT5G08415 AT2G20860
GDP-L-fucose
lipoylated domain
citrulline
OTC
carbamoyl-phosphate
CO2
L-methionine
SAM-1 AT4G01850
AT3G17390 AT2G36880
S-adenosyl-L-methionine
glycerol
AT1G80460
sn-glycerol-3-phosphate
AT3G10370
dihydroxyacetone phosphate
hydrogen cyanide
AT3G61440
3-cyano-L-alanine
AT5G22300
L-asparagine
L-aspartate
indole-3-acetyl-ala
indole-3-acetyl-leu
ill2 iar3
ilr1
indole-3-acetate
trehalose
TRE1
β-D-glucose
HXK2 HXK1
β-D-glucose-6-phosphate
caffeate
cis-caffeate
3-(2-propenoic acid)-o-benzoquinone
acrylate
3-(2-propenoic acid)-4,6-hydroxy cyclohexa-2,5-dienone
3,4,6-trihydroxy-cis-cinnamate
esculetin
acrylonitrile
AT5G22300 AT4G08790
acrylamide
AT5G64440
AT1G08980 AT5G09420
AT5G07360 AT4G34880
acrylate
sulfite
SOX
sulfate
naringenin
apigenin
eriodictyol
luteolin
lipoamide acyltransferase N6-(lipoyl)lysine
AT3G13450 AT5G09300
AT4G16155 AT1G48030
lipoamide acyltransferase N6-(dihydrolipoyl)lysine
lipoamide acyltransferase N6-(S-[2-methylpropanoyl]dihydrolipoyl)lysine
BCE2
phytyl-PP
chlorophyllide a
G4
chlorophyll a
coniferyl alcohol
AT5G66690
AT3G50740 AT5G26310
coniferin
coniferyl alcohol
GTP
AT5G64300
AT5G59750 AT2G22450
2,5-diamino-6-(ribosylamino)-4-(3H)-pyrimidinone 5'-phosphate
AT4G20960
5-amino-6-(5'-phosphoribosylamino)uracil
5-amino-6-(5'-phosphoribitylamino)uracil
5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione
D-ribulose-5-phosphate
AT5G64300
AT5G59750 AT2G22450
3,4-dihydroxy-2-butanone-4-P
AT2G44050
6,7-dimethyl-8-(1-D-ribityl)lumazine
AT2G20690
riboflavin
AT4G21470
FMN
AT5G08340 AT5G23330
FAD
L-isoleucine
Atbcat-2 Atbcat-3
Atbcat-1 Atbact-5
2-keto-3-methyl-valerate
LPD2 BCE2
AT5G09300 AT3G13450
S-2-methyl-butyryl-CoA
tiglyl-CoA
AT4G29010
AT3G06860 AT4G31810
AT5G43280 AT4G16800
2-methyl-3-hydroxybutyryl-CoA
2-methylacetoacetyl-CoA
KAT2 KAT5 KAT1
propionyl-CoA
L-glutamate
P5CS2 P5CS1
L-glutamate γ-semialdehyde
L-ornithine
δ-OAT
pyrroline 5-carboxylate
P5C1
L-proline
sn-glycerol-3-phosphate
AT4G00400 AT1G32200
a 2-lysophosphatidate
AT3G11430
an L-phosphatidate
AtLPP1 AtLPP2
AT3G18220 AT3G02600
a 1,2-diacylglycerol
AT5G13640
TAG1 AT3G51520
a triacylglycerol
L-tyrosine
AT5G53970 AT2G24850
p-hydroxyphenylpyruvate
PDS1
homogentisate
AT5G54080
4-maleyl-acetoacetate
4-fumaryl-acetoacetate
AT1G12050
acetoacetate fumarate
(1,4-alpha-D-galacturonosyl)(N)
UDP-D-galacturonate
GAUT1 GATL10 GATL9
GATL8 GATL7 GATL6
GATL5 GATL4 GATL3
NO3-
NIA1 NIA2
L-arginine
AT4G08900

PMN
Plant Metabolic Network

Can overlay transcriptomic and proteomic data – very complex!

MetExplore and CytoScape



Summary

- Metabolomics – logical progression of genomic and post-genomic science
- Diverse range of applications – especially trait identification
- Range of fingerprinting and profiling techniques
- Large datasets require multivariate statistics
- Large number of online databases for metabolite identification and mapping

Overview of techniques

Targeted and non-targeted metabolomics
(metabolite extraction procedures, equipment
GC-MS, HPLC-PDA-MS)