Natural variation of root exudates in Arabidopsis thailiana linking metabolomic and genomic data

Susann Mönchgesang*, Nadine Strehmel*, Stephan Schmidt*, Lore Westphal, Franziska Taruttis†, Erik Müller, Siska Herklotz, Steffen Neumann & Dierk Scheel

Begelson Lab Meeting
Justin Rigby
12/04/17

General Overview

- 19 separate lab accessions were used in the paper [e.g. Col-0, Pol-0, ect]
- MAGIC lines [Multiparent advanced generation inter-cross] were constructed from the accessions (Huang et al. 2012)
- A generic workflow including metabolomics and sequencing was used
- Fold change in metabolite presence and mutated abrupt stop codons were measured and correlated to determine a relationship

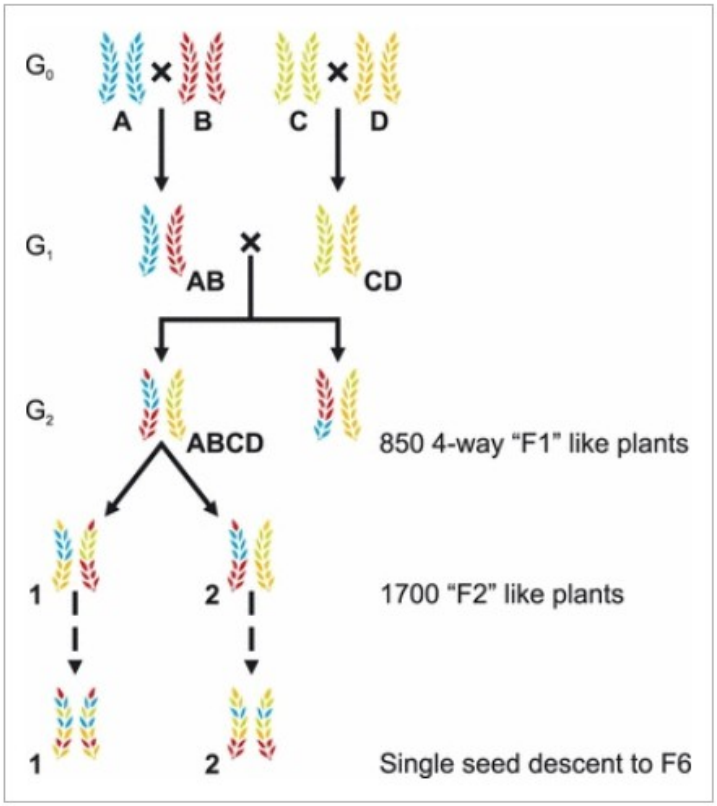


Fig 1. Definition of a MAGIC line
Huang et al. 2012

Hypotheses

- To determine if the root exudates are genetically determined, and if they can be found by a mixture of identifying dis-regulated metabolites and presence of SNP’s in the genome which lead to premature stop codons

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon

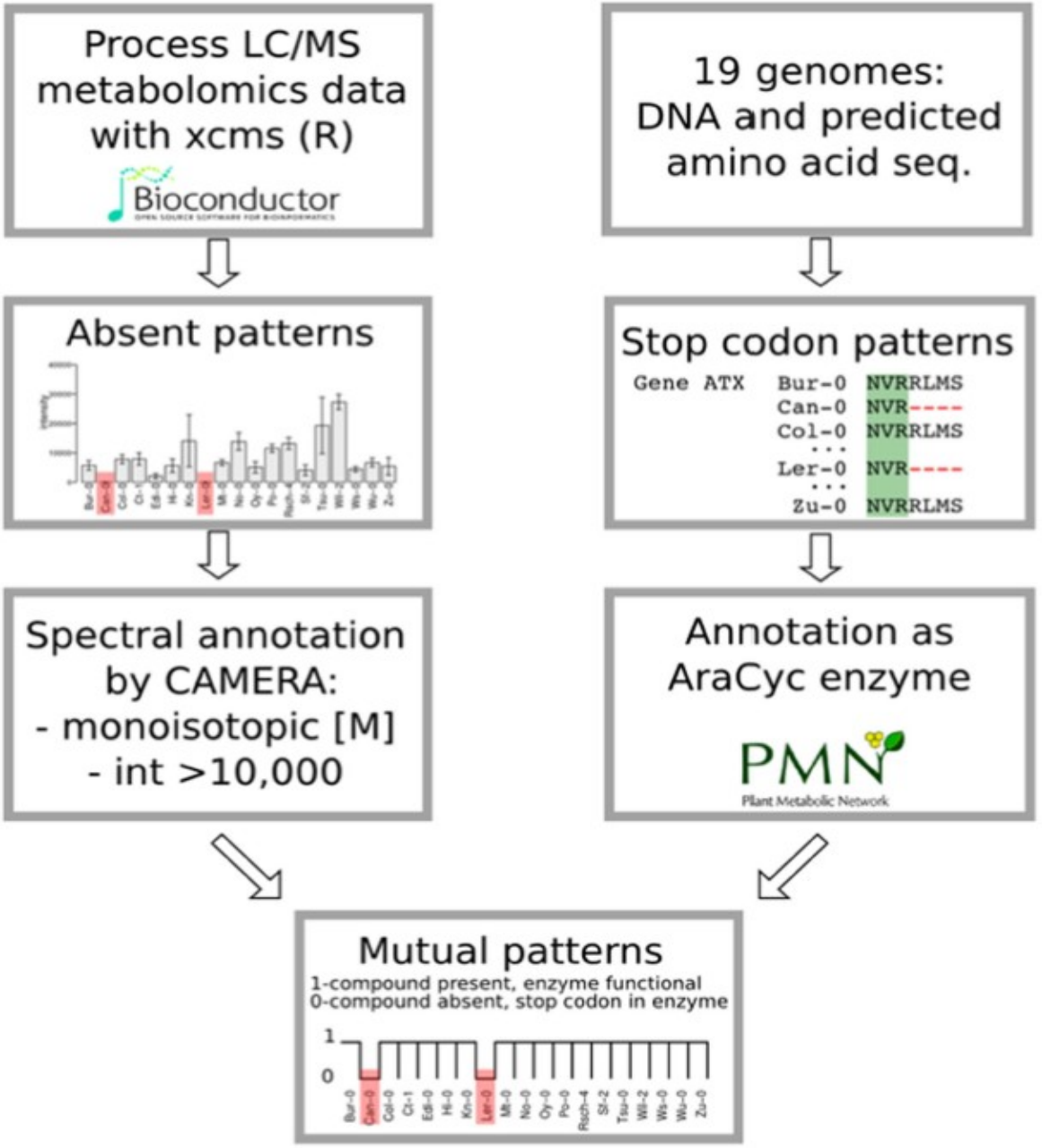
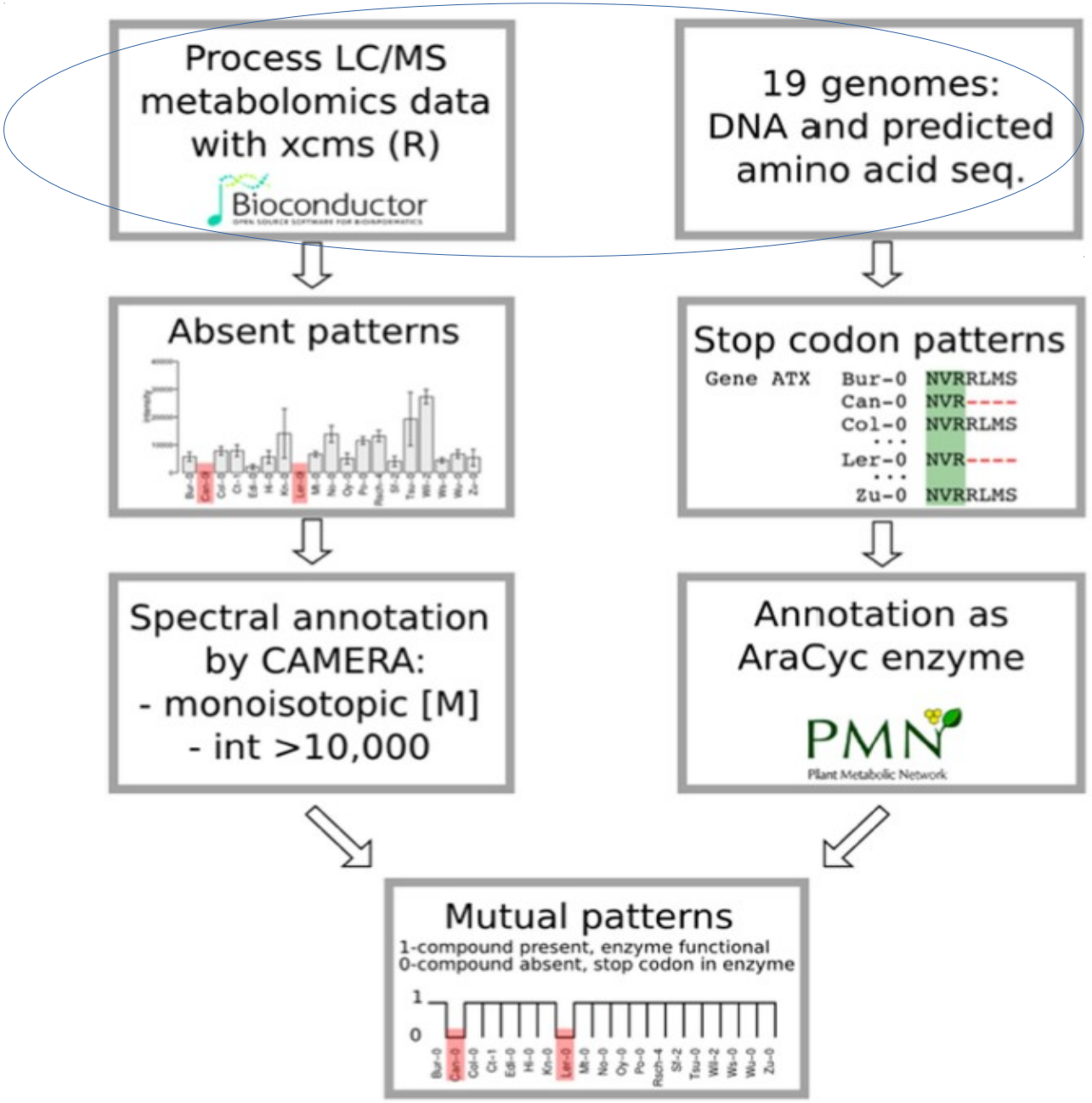


Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon



Plants wished to be sampled must have both their metabolome and their genome sequenced to a degree.

Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon

De convolution of data and creation of a peak table

Looking the absence of chemicals generally seen / assumed to be present

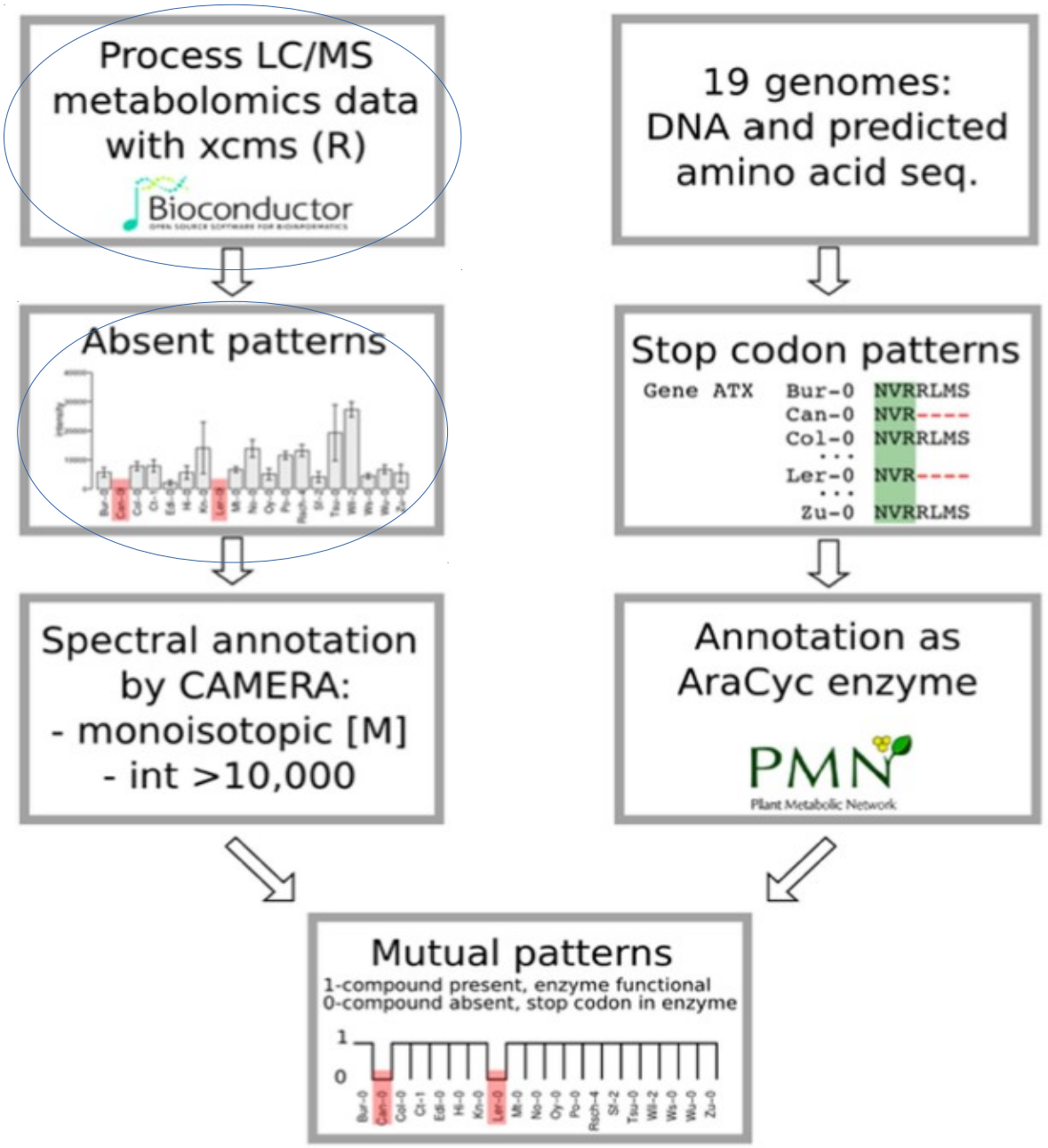
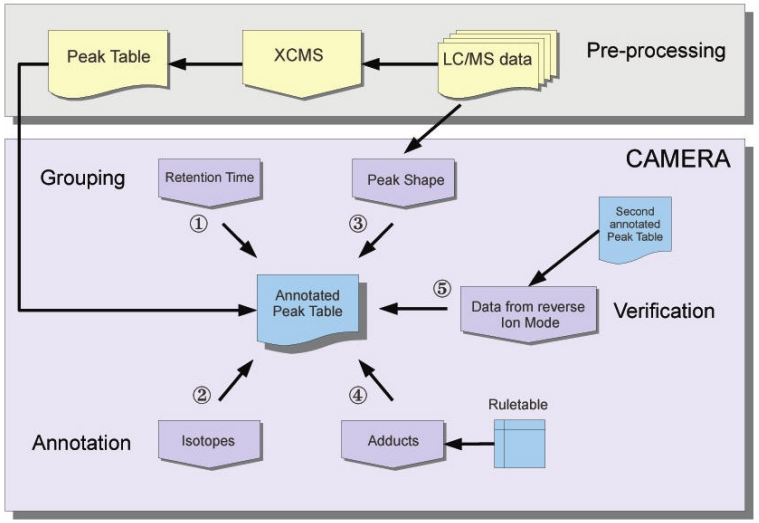


Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon

R package designed to analyse LC-MS peaks in order to determine the isotopes and accurate compound mass. (Kuhl et al. 2011)



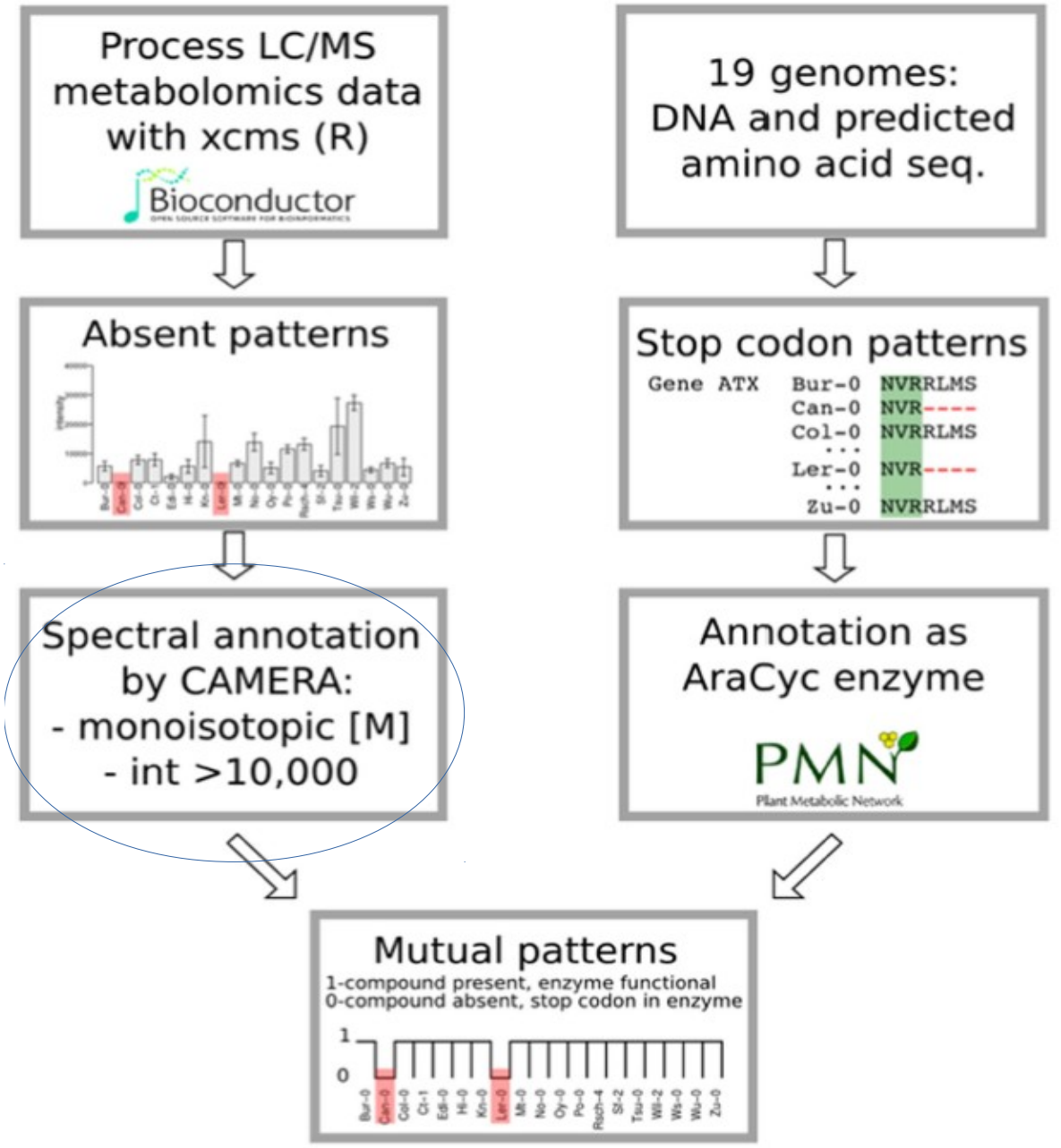
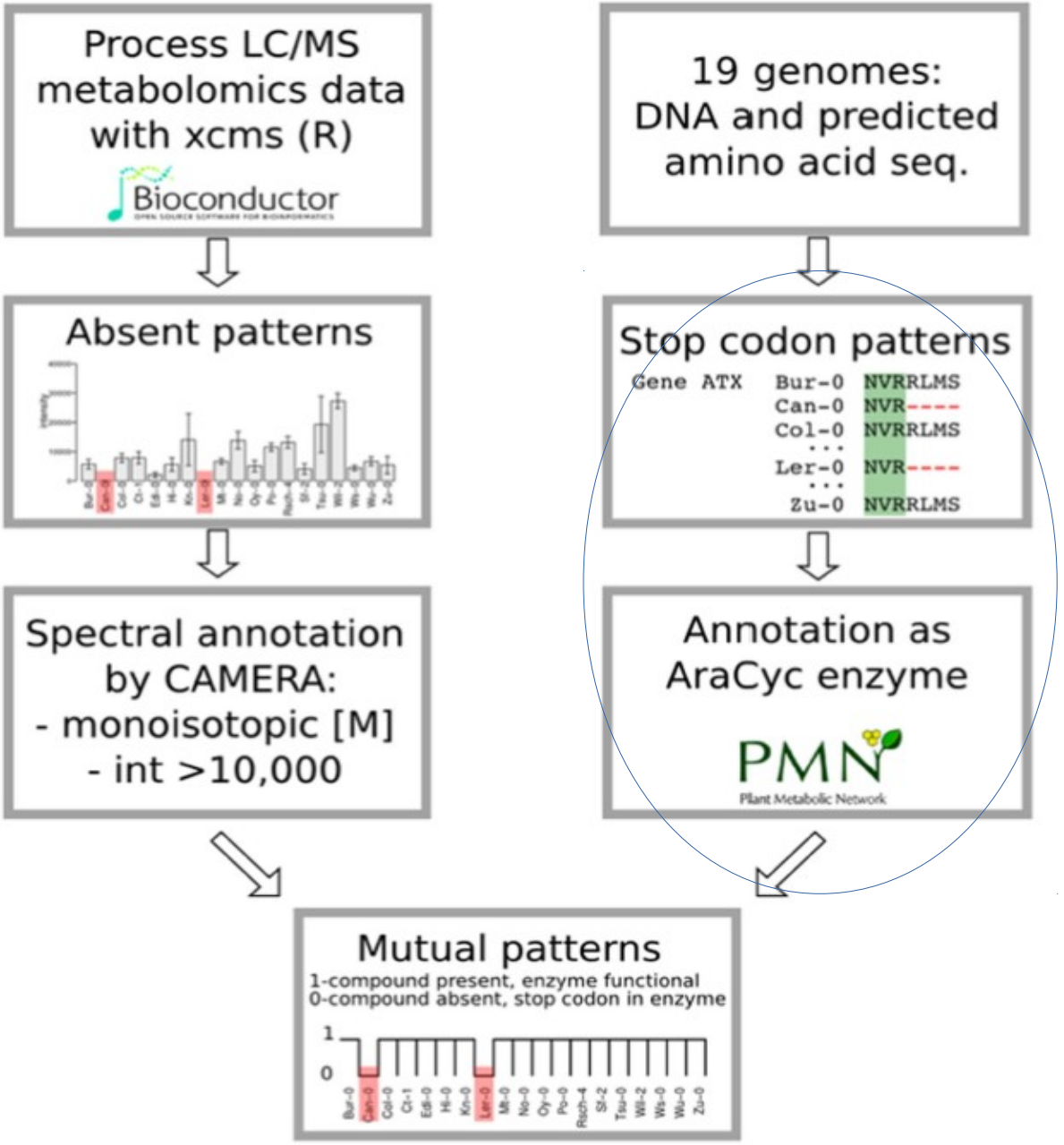


Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon



Large assumption that mutations in genes that encode enzymes are always negatively correlated with chemical abundances

Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Workflow for matching metabolic patterns of absence with mutation of a premature stop codon

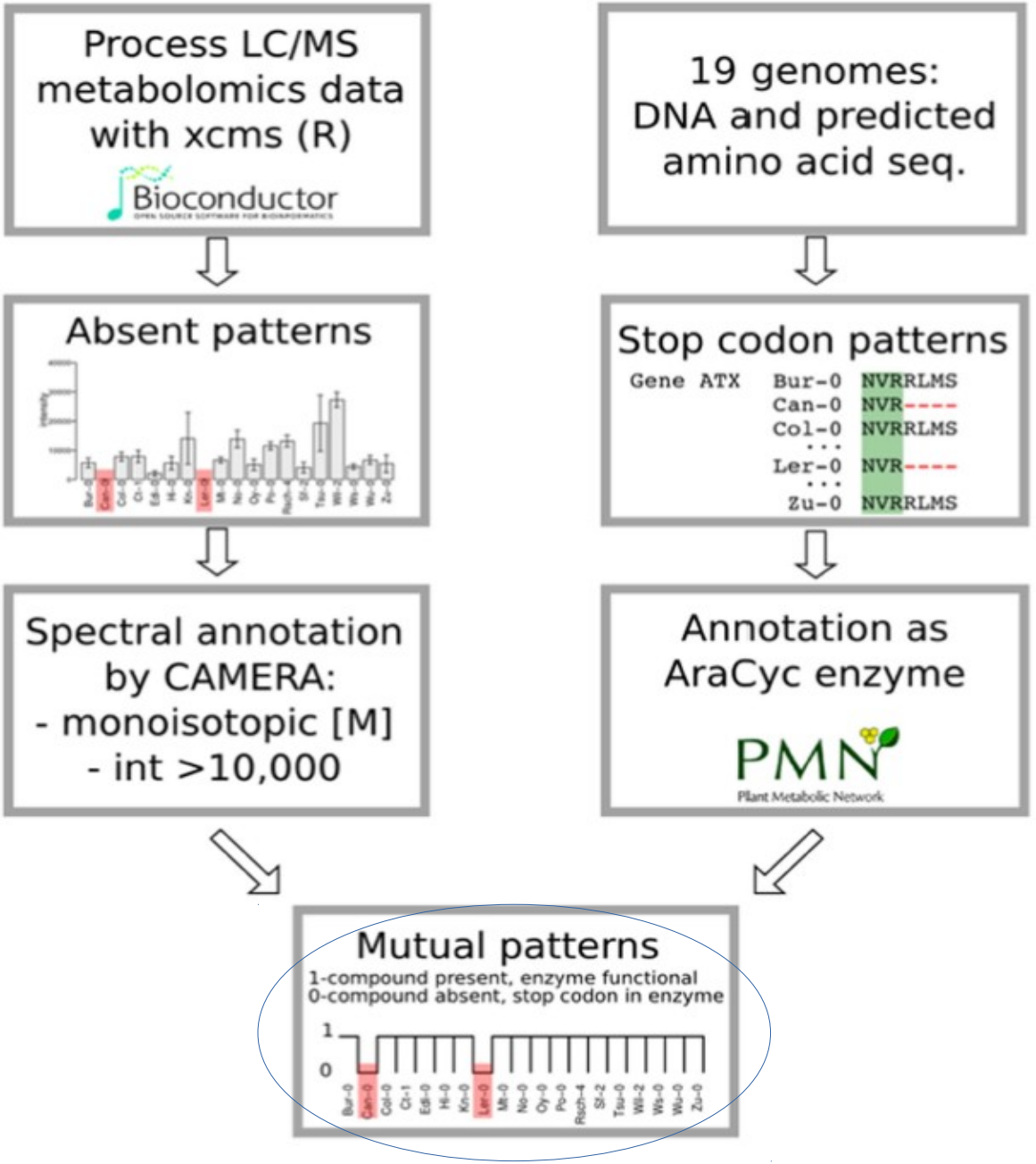
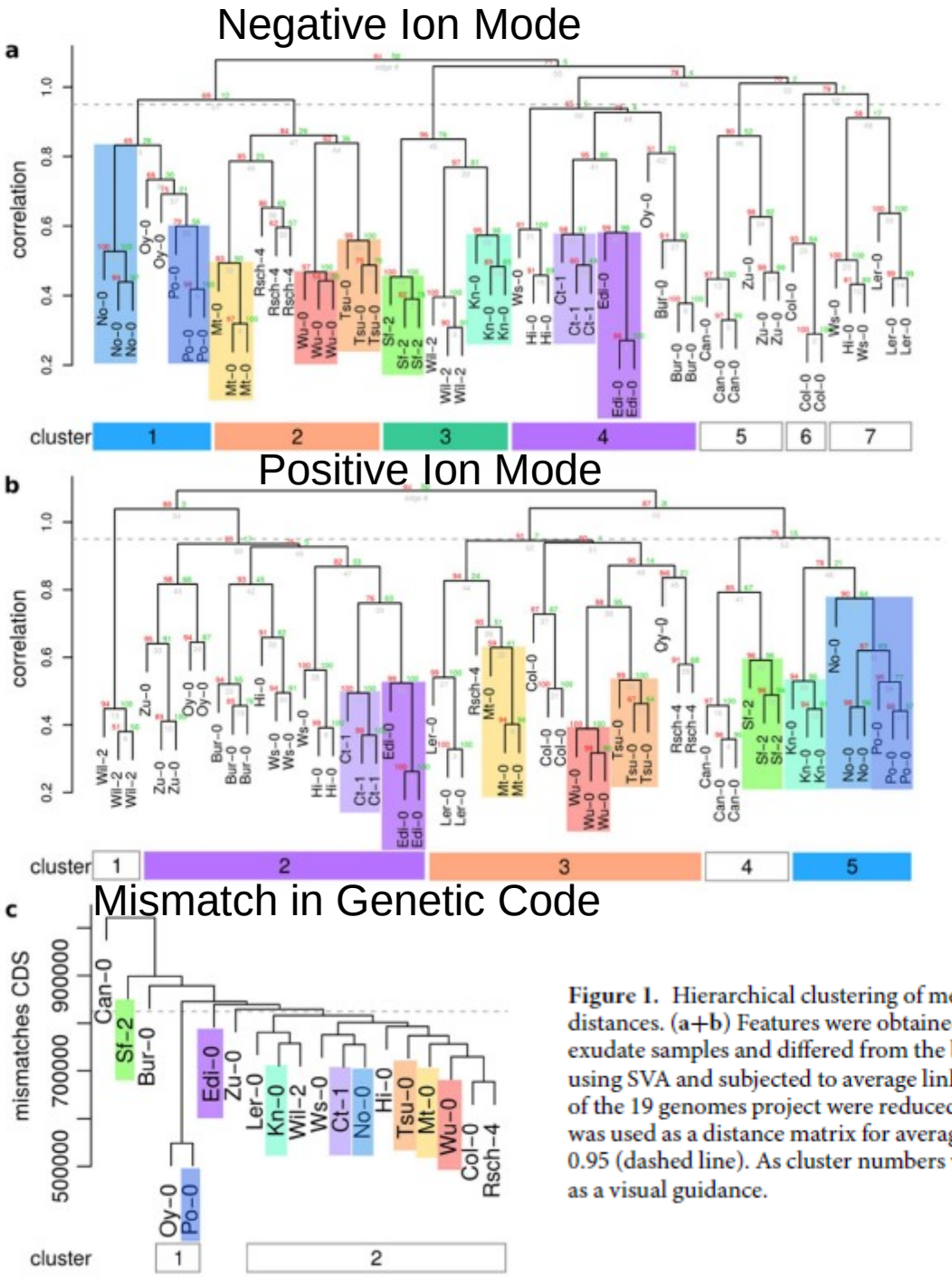


Figure 3. Workflow for matching metabolic patterns of absence with stop codons in genes annotated as AraCyc enzymes. For the metabolic data, 384 out of 455 metabolic features from the ESI(−) data set were absent in at least one accession. 38 of them were annotated as monoisotopic peak [M] by CAMERA. Approximately 32,000 stop codons were detected. 1,588 of AraCyc enzyme-encoding genes displayed a prematurely ended amino acid sequence possibly representing non-functional enzymes that can be causative for metabolite absence.

Results

Looking for clustering of metabolome compositions in the 19 accessions.



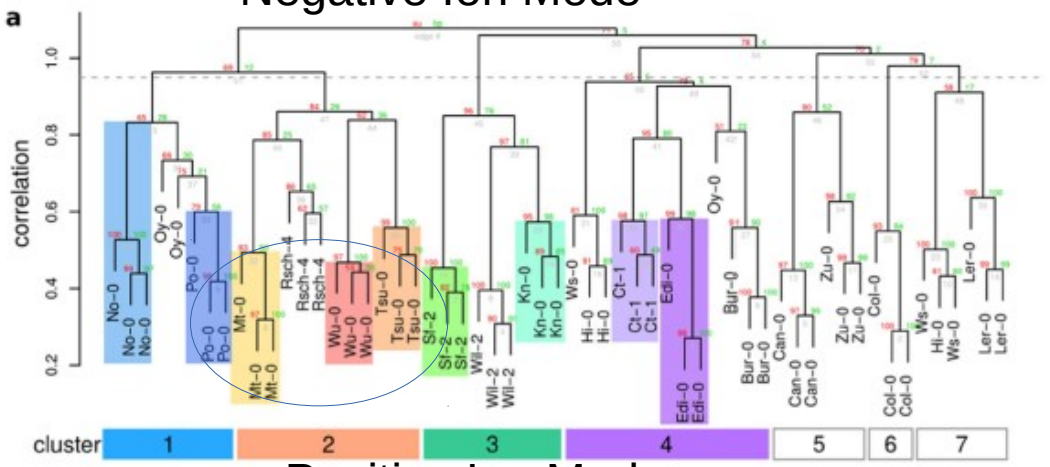
Clusters of similar metabolomes in the 19 accessions

Clusters of genotypes which typically resides next to one another are labelled with colour highlights

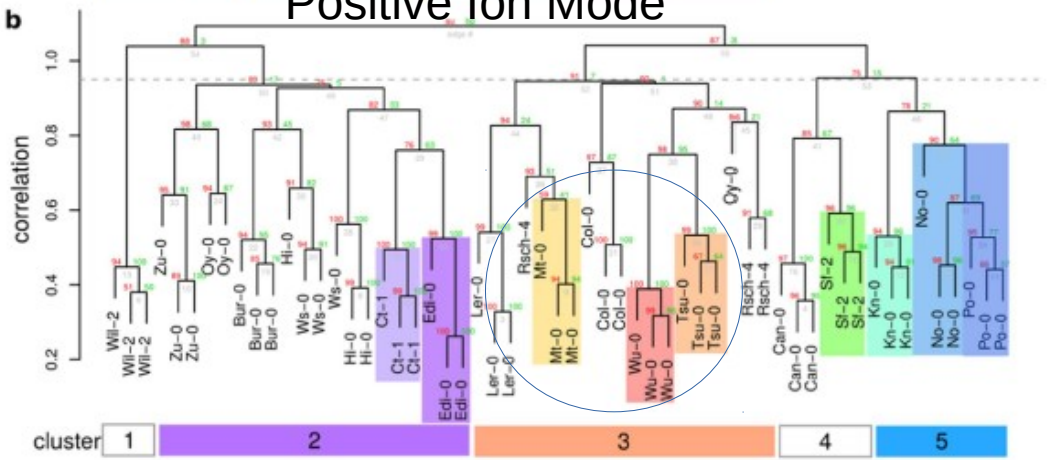
The clustering of some colours is more apparent than others, and does not guarantee that the clustering is significant across all figures

Figure 1. Hierarchical clustering of metabolic features from (**a**) exudates ESI(−), (**b**) ESI(+) and of (**c**) genetic distances. (**a**+**b**) Features were obtained by UPLC/ESI(−)-QTOF-MS (**a**) or UPLC/ESI(+)-QTOF-MS (**b**) from exudate samples and differed from the blank (Welch test, $p < 0.05$). Intensities were corrected for batch effects using SVA and subjected to average linkage clustering with correlation as a distance measure. (**c**) Variant tables of the 19 genomes project were reduced to coding regions, as annotated by TAIR. The sum of all mismatches was used as a distance matrix for average linkage clustering. Dendrograms were cut at a correlation threshold of 0.95 (dashed line). As cluster numbers were not comparable, consistent clusters were coloured across ion modes as a visual guidance.

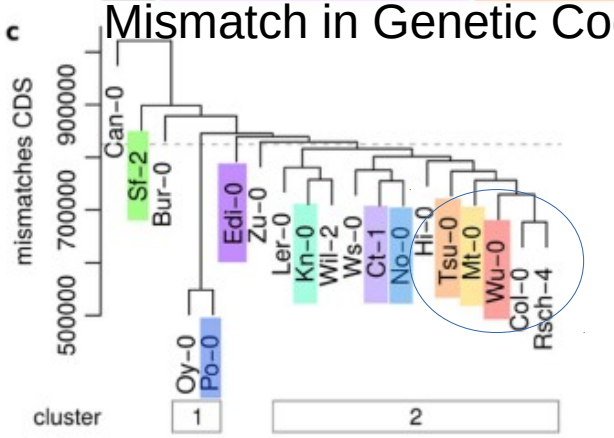
Negative Ion Mode



Positive Ion Mode



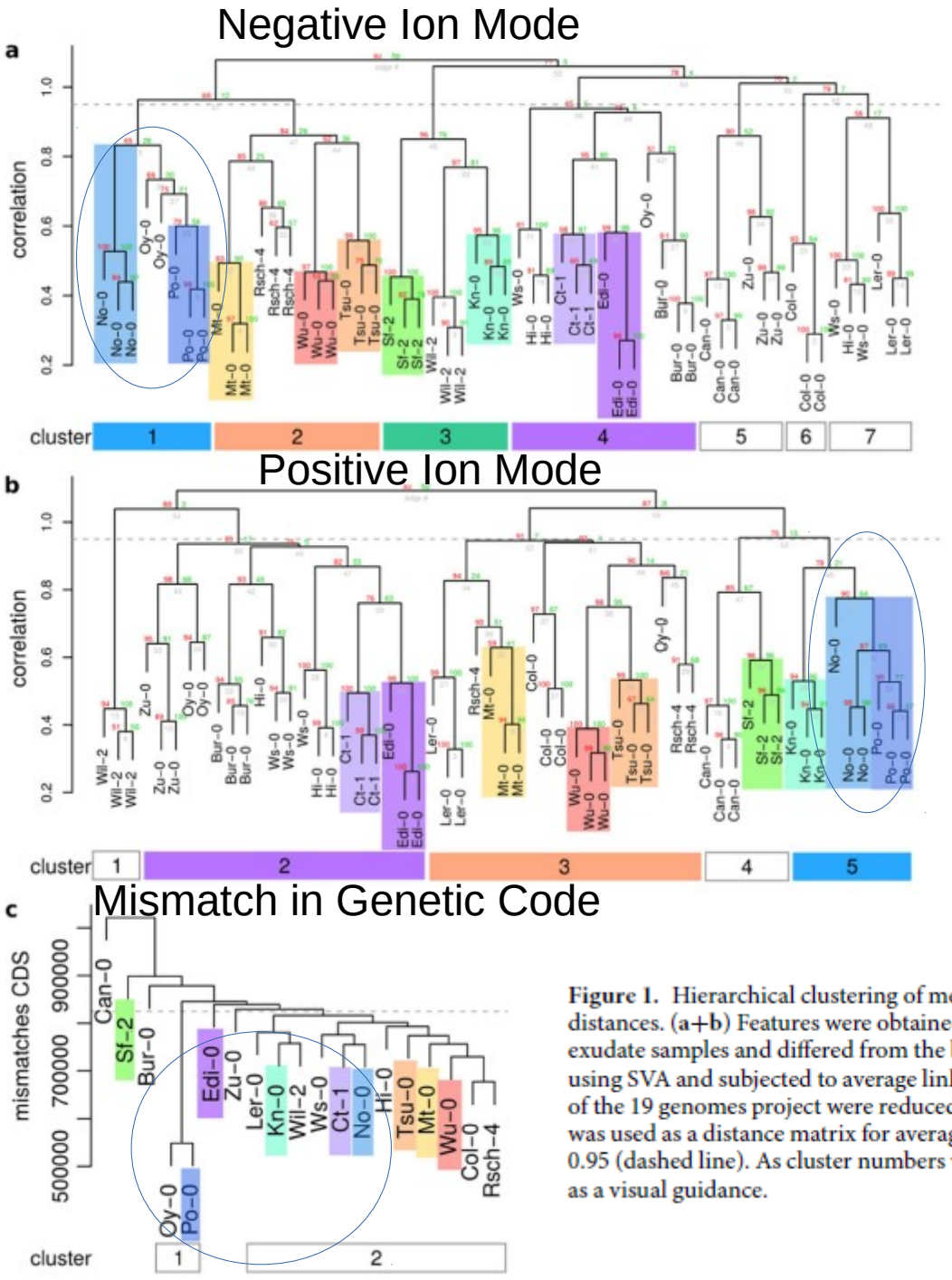
Mismatch in Genetic Code



Here three clusters of accessions tend to cluster next to one another in similar fashion

Possible indication of genetic abnormalities causing differential expression of metabolites

Figure 1. Hierarchical clustering of metabolic features from (**a**) exudates ESI(−), (**b**) ESI(+) and of (**c**) genetic distances. (**a**+**b**) Features were obtained by UPLC/ESI(−)-QTOF-MS (**a**) or UPLC/ESI(+)-QTOF-MS (**b**) from exudate samples and differed from the blank (Welch test, $p < 0.05$). Intensities were corrected for batch effects using SVA and subjected to average linkage clustering with correlation as a distance measure. (**c**) Variant tables of the 19 genomes project were reduced to coding regions, as annotated by TAIR. The sum of all mismatches was used as a distance matrix for average linkage clustering. Dendrograms were cut at a correlation threshold of 0.95 (dashed line). As cluster numbers were not comparable, consistent clusters were coloured across ion modes as a visual guidance.



Here two clusters of accessions tend to cluster next to one another in dissimilar fashion

The Po-0 genotype (along with Oy-0) tend to have the greatest amount of mismatches in their coding genome than the rest of the accessions

Interstingly, the level of mismatch CDS (Sequence polymorhpisms in coding sequences) does not seem to be similar between the Po-0 and Ct-1/No-0 genotypes

Figure 1. Hierarchical clustering of metabolic features from (a) exudates ESI(−), (b) ESI(+) and of (c) genetic distances. (a+b) Features were obtained by UPLC/ESI(−)-QTOF-MS (a) or UPLC/ESI(+)-QTOF-MS (b) from exudate samples and differed from the blank (Welch test, $p < 0.05$). Intensities were corrected for batch effects using SVA and subjected to average linkage clustering with correlation as a distance measure. (c) Variant tables of the 19 genomes project were reduced to coding regions, as annotated by TAIR. The sum of all mismatches was used as a distance matrix for average linkage clustering. Dendrograms were cut at a correlation threshold of 0.95 (dashed line). As cluster numbers were not comparable, consistent clusters were coloured across ion modes as a visual guidance.

Results

Looking at the fold changes of metabolites across the 19 accessions

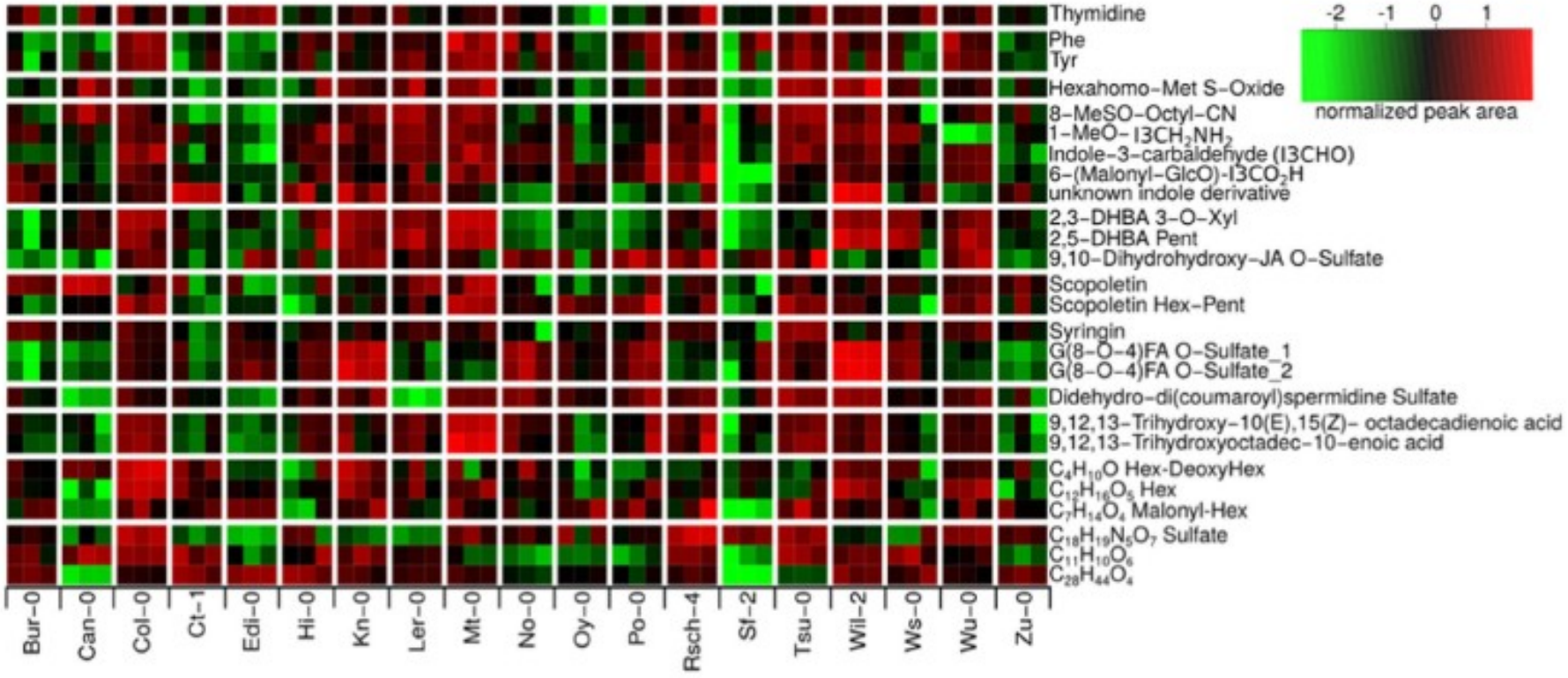


Figure 2. Colour-coded intensity matrix of differential metabolites occurring in exudates. Integrated peak areas were log-transformed and scaled to zero mean and standard variance. A Welch-test was used to find differentially abundant metabolites between the 19 accessions.

- Black: No difference
- Green: Down-regulation
- Red: Up-regulation

Three biological replicates for each genotype

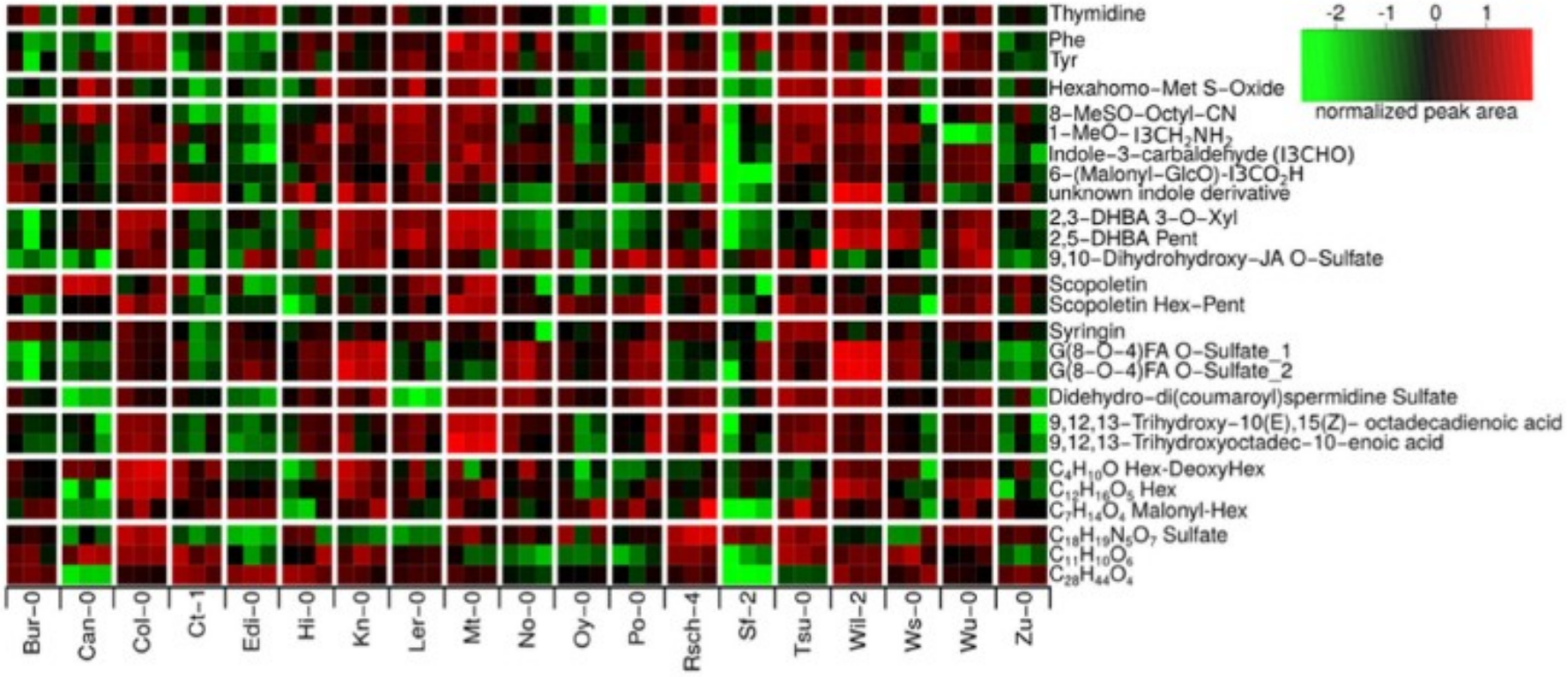


Figure 2. Colour-coded intensity matrix of differential metabolites occurring in exudates. Integrated peak areas were log-transformed and scaled to zero mean and standard variance. A Welch-test was used to find differentially abundant metabolites between the 19 accessions.

Significant negative fold changes in metabolite presence dependant on the genotype

Significant positive fold changes is hard to determine since the ‘2’ for the normalize peak area is not present

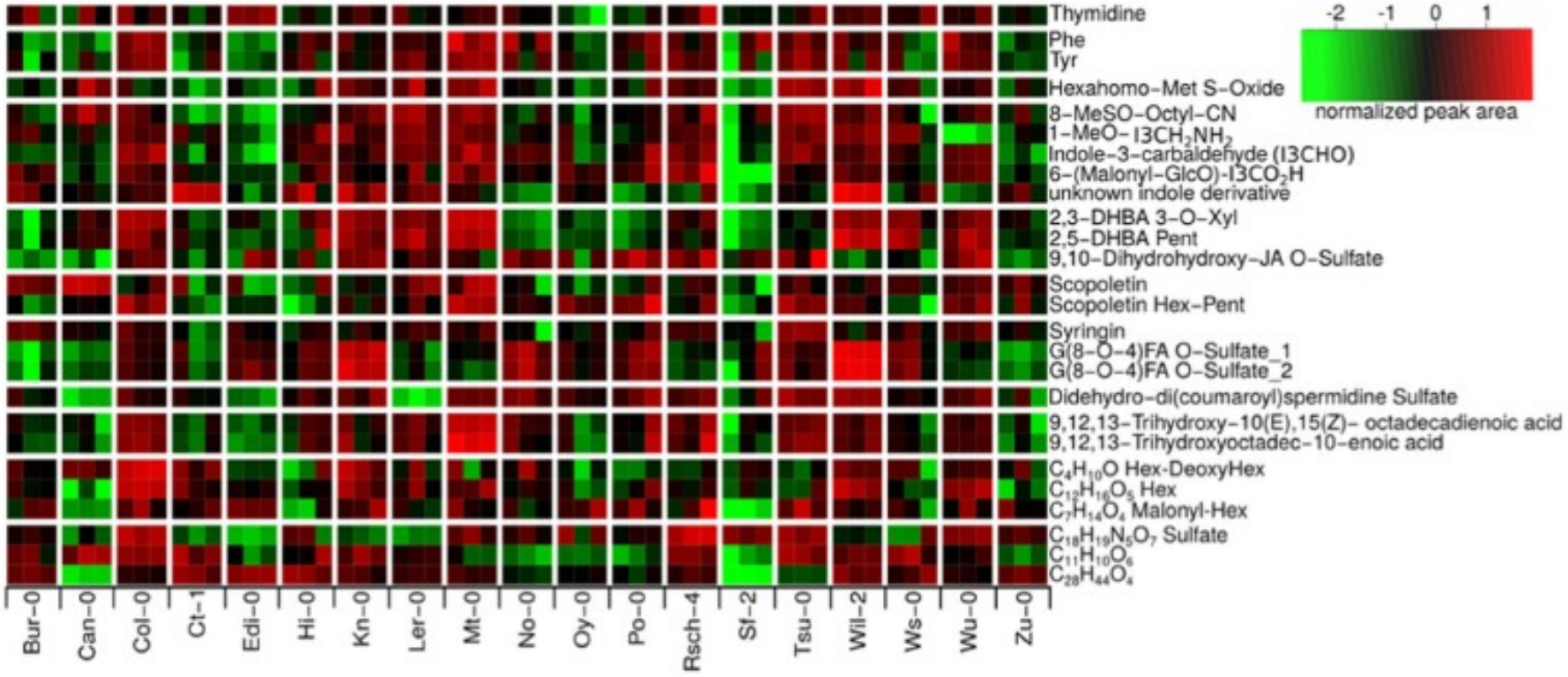


Figure 2. Colour-coded intensity matrix of differential metabolites occurring in exudates. Integrated peak areas were log-transformed and scaled to zero mean and standard variance. A Welch-test was used to find differentially abundant metabolites between the 19 accessions.

Raises question as to why this is;
Could be an artefact finding influenced by accidental relaxed selection in a lab setting
Or maybe
Difference in population metabolite composition

Results

Identifying a possible
cyclic didehydro-di(coumaroyl)spermidine sulfate
gene

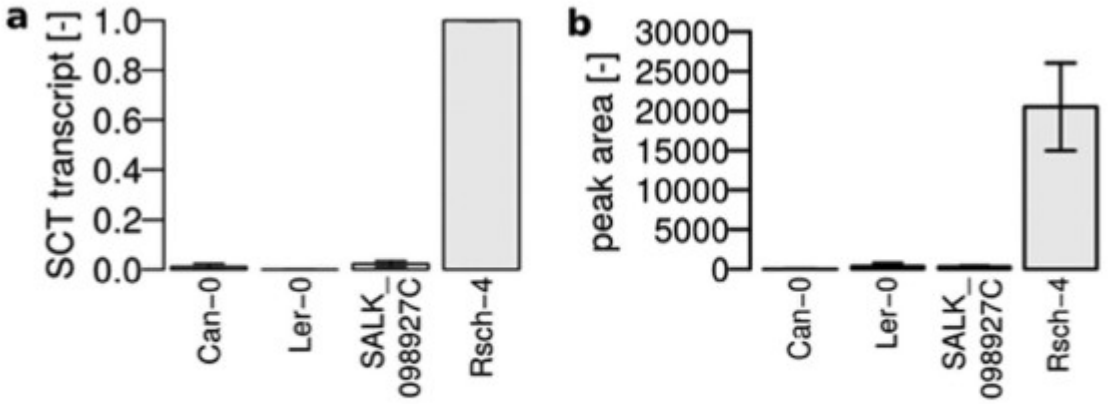


Figure 4. Natural and T-DNA insertion knockouts of SCT. (a) Relative transcript levels of SCT in root tissue as determined by qPCR, PP2A as reference, normalized to Rsch-4, mean ± s.e.m., n = 3. (b) Peak area counts of cyclic didehydro-di(coumaroyl)spermidine sulfate in exudates, mean ± s.e.m., n = 3.

Looking at the level of transcription (a) and the level of metabolites in question (b) in this figure

- Just because a single gene knock-out causes the same phenotype is not grounds for assumption of a relationship and the biological relevance of the chemical is still unknown

Results

Identifying a possible Robinin gene

Results

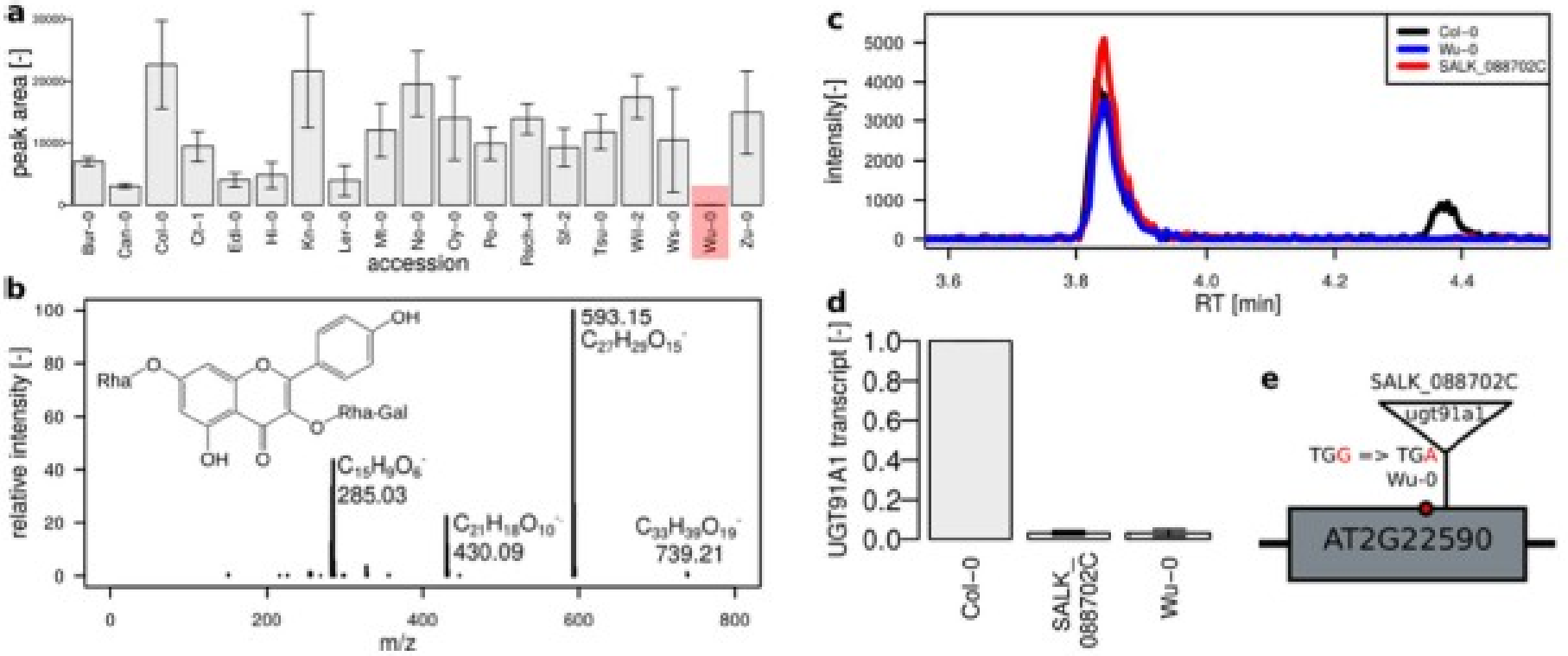


Figure 5. Robinin absence is linked to a stop codon in the UGT91A1 encoding gene. (a) Peak area counts, mean ± s.e.m. (n = 3) with absence in Wu-0 (highlighted in red) (b) MS/MS spectrum of robinin, 30 eV, (c) extracted ion chromatogram at *m/z* 739.21 with kaempferol 3-*O*-Rha(1→2)Glc 7-*O*-Rha eluting at 3.9 min and the galactose-conjugated robinin eluting at 4.3 min not detected in the natural knockout Wu-0 and T-DNA insertion line SALK_088702C, (d) relative transcript levels of UGT91A1 in roots as determined by qPCR, PP2A as reference, normalized to Col-0, mean ± s.e.m., n = 4, (e) schematic representation of the UGT91A1 gene (one exon) and the loss-of-function mutations in Wu-0 and SALK_088702C.

Conclusion

- The new protocols introduced in this paper pave the path for new high-throughput techniques and analysis between genes and metabolites

- The chance of a type I error is high due to the methods implored (no empirical data on the gene transcription or purpose is shown or accounted for)

Supplemental Images

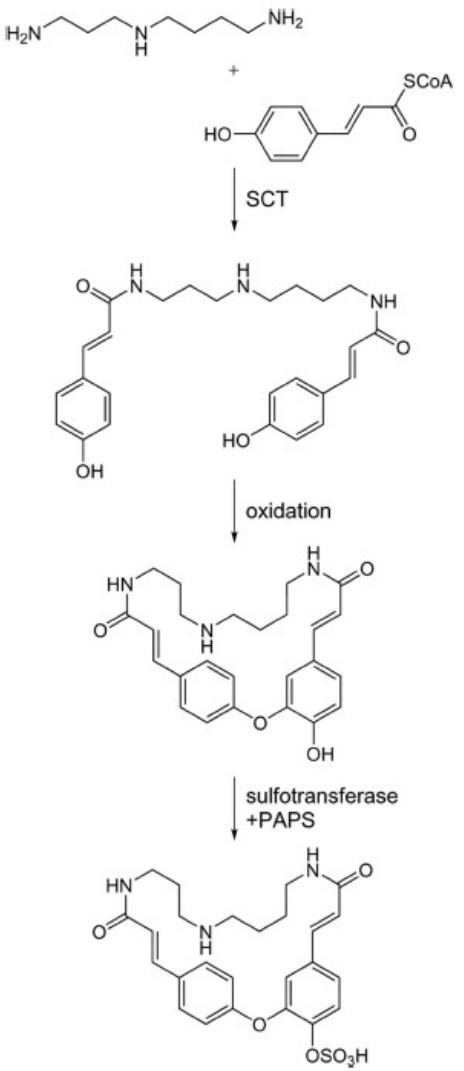
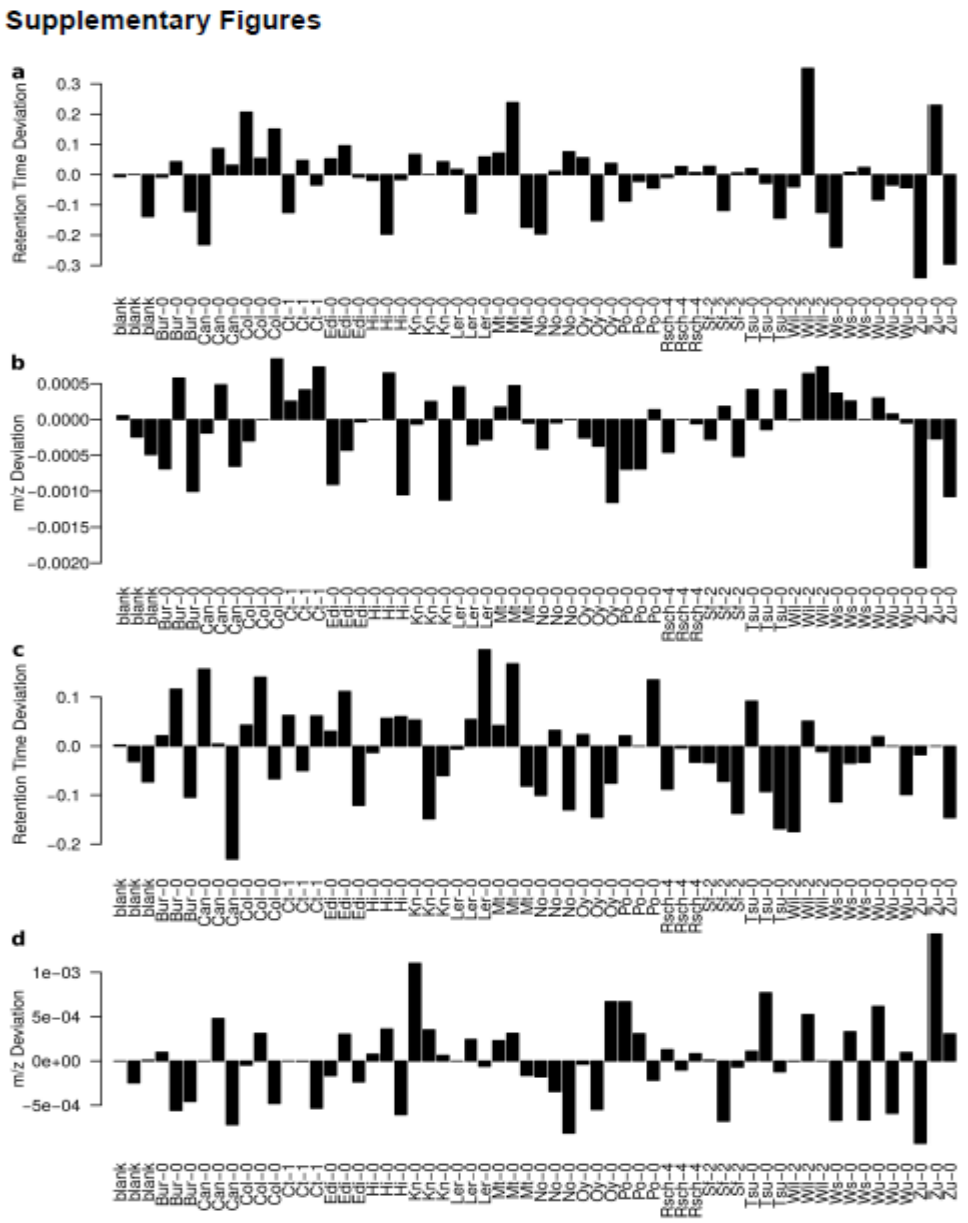


Figure 6. Biosynthetic pathway of cyclic didehydro-di(coumaroyl) spermidine sulfate. Di(coumaroyl) spermidine is synthesized by SCT[47] and subsequent oxidative ring closure and sulfonylation leads to cyclic didehydro-di(coumaroyl) spermidine sulfate, PAPS = 3′-phosphoadenosine-5′-phosphosulfate.

Supplemental Images



Supplemental Images

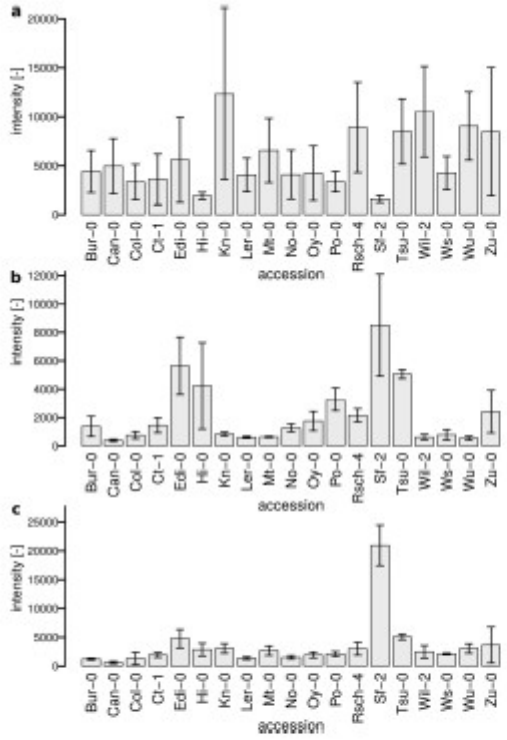


Fig. S2: Intensity distribution of salicylic acid and DHBA hexosides in exudates.

a) Salicylic acid, b) 2,5-DHBA hexose, c) 2,3-DHBA hexose, mean ± s.e.m. (n=3).

Supplemental Images

Supplementary Tables

Table S1: List of *Arabidopsis thaliana* accessions used for the analysis.

Native Name	Short Name	Region
Burren	Bur-0	Ireland
Canary Islands	Can-0	Spain
Catania	Ct-1	Italy
Columbia	Col-0	USA
Edinburgh	Edi-0	UK
Hilversum	Hi-0	Netherlands
Kaunas	Kn-0	Lithuania
Landsberg	*Ler*-0	Poland
Martuba	Mt-0	Libya
Nossen	No-0	Germany
Oystese	Oy-0	Norway
Poppelsdorf	Po-0	Germany
Rschew	Rsch-4	Russia
San Feliu	Sf-2	Spain
Tsushima	Tsu-0	Japan
Wilna	Wil-2	Lithuania
Wassilewskija	Ws-0	Russia
Würzburg	Wu-0	Germany
Zürich	Zu-0	Switzerland

Supplemental Images

Supplementary Tables

Table S1: List of *Arabidopsis thaliana* accessions used for the analysis.

Native Name	Short Name	Region
Burren	Bur-0	Ireland
Canary Islands	Can-0	Spain
Catania	Ct-1	Italy
Columbia	Col-0	USA
Edinburgh	Edi-0	UK
Hilversum	Hi-0	Netherlands
Kaunas	Kn-0	Lithuania
Landsberg	*Ler*-0	Poland
Martuba	Mt-0	Libya
Nossen	No-0	Germany
Oystese	Oy-0	Norway
Poppelsdorf	Po-0	Germany
Rschew	Rsch-4	Russia
San Feliu	Sf-2	Spain
Tsushima	Tsu-0	Japan
Wilna	Wil-2	Lithuania
Wassilewskija	Ws-0	Russia
Würzburg	Wu-0	Germany
Zürich	Zu-0	Switzerland