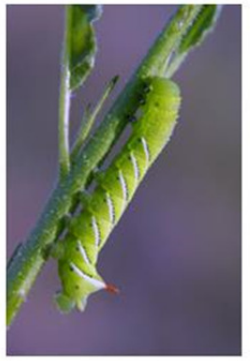
METABOSCAPE – A METABOLITE PROFILING PIPELINE DRIVEN BY AUTOMATIC COMPOUND IDENTIFICATION



OR

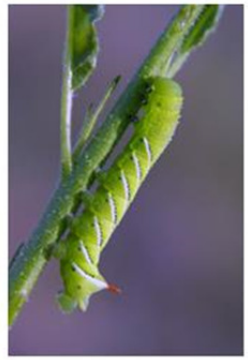
HOW TO LINK HRAM QTOF PLANT METABOLOMICS DATA TO BIOLOGY

Aiko Barsch, Bruker Daltonics, Bremen, Germany

Outline

- If you want to automate data processing workflows in Metabolomics you have to be able to rely on the raw data.

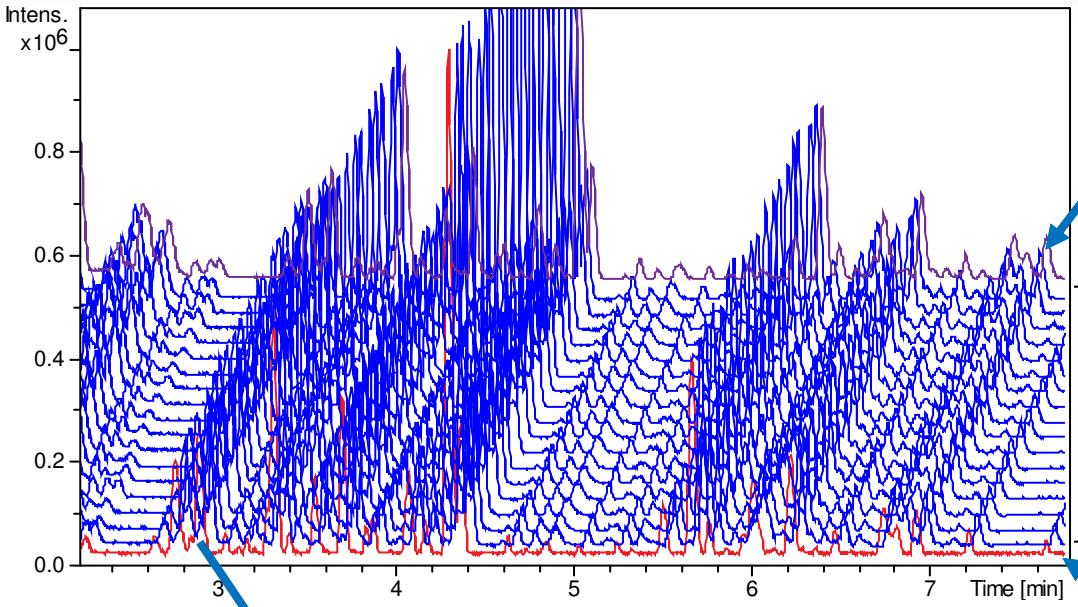
- Introducing MetaboScape. Turning MS data into knowledge!
  - An example studying the “war” between plants and insects!



Automatic data processing workflows in Metabolomics rely on the raw data quality

- The metabolome is an extremely complex mixture of chemically diverse small molecules covering a large dynamic range of concentrations.
- Robustness **and** Sensitivity **and** Dynamic range **and** High resolution, mass accuracy and isotopic fidelity in **one scan mode** at **LC speed** is absolute must in metabolomics.
- So, lets have a look at the capabilities of a QTOF instrument. Can this technology fulfill these needs?

Instrument robustness:
>100 injections of complex human Urine sample

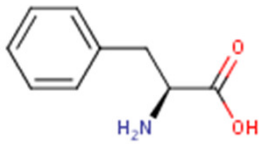


Sample Injection #100

selected BPCs from sequence of human urine sample injected 100 times

Sample Injection #1

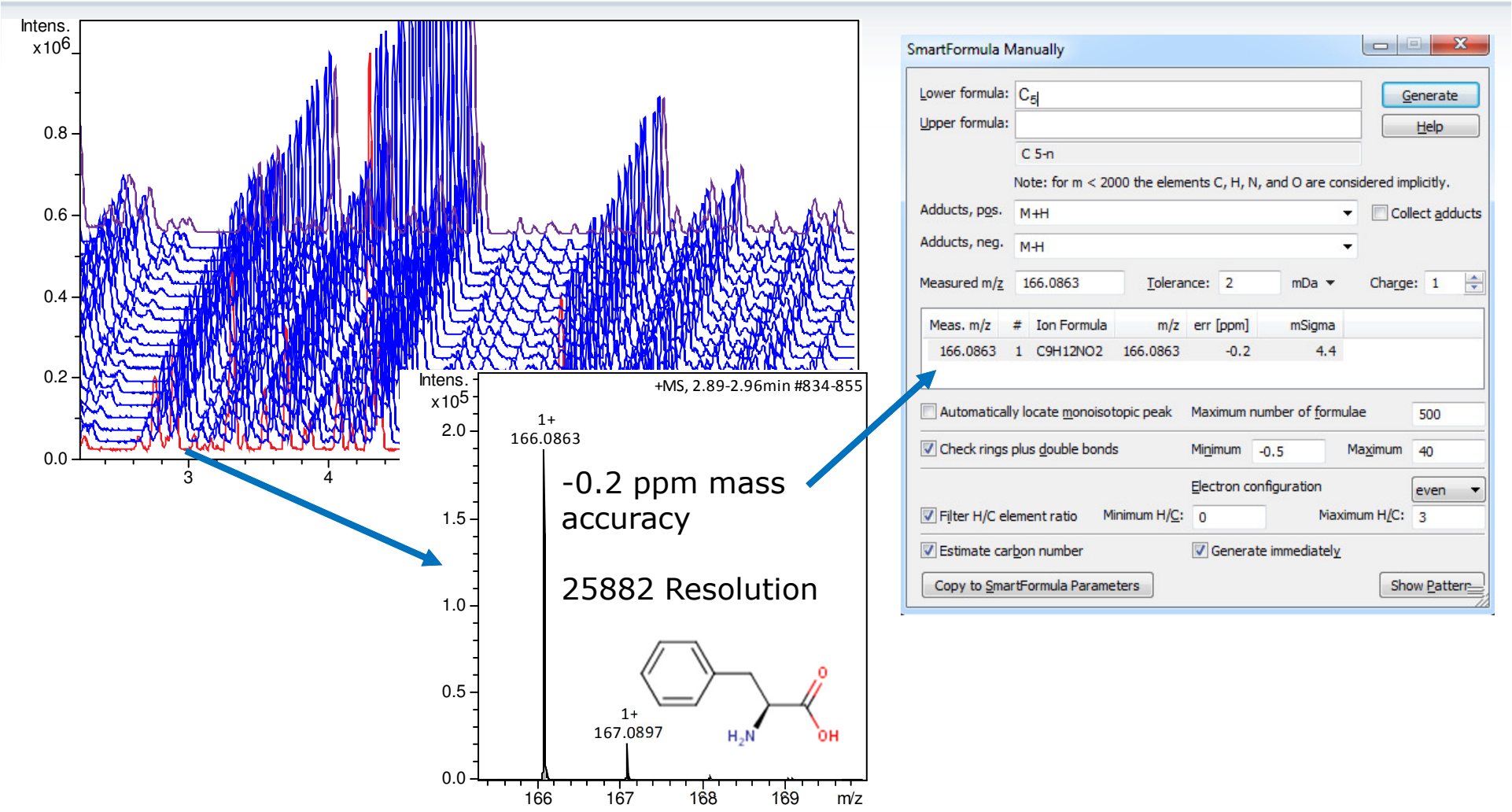
Phenylalanine peak



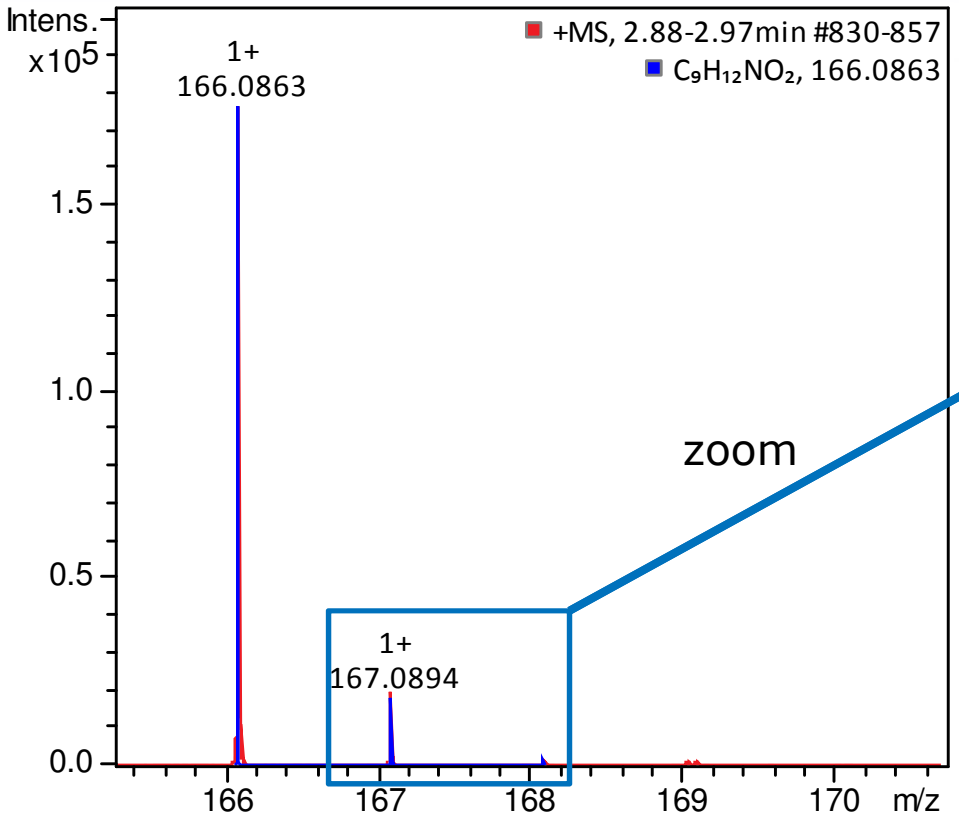
	Peak Area
Average	730594
StDEV	19720
CV	2.70

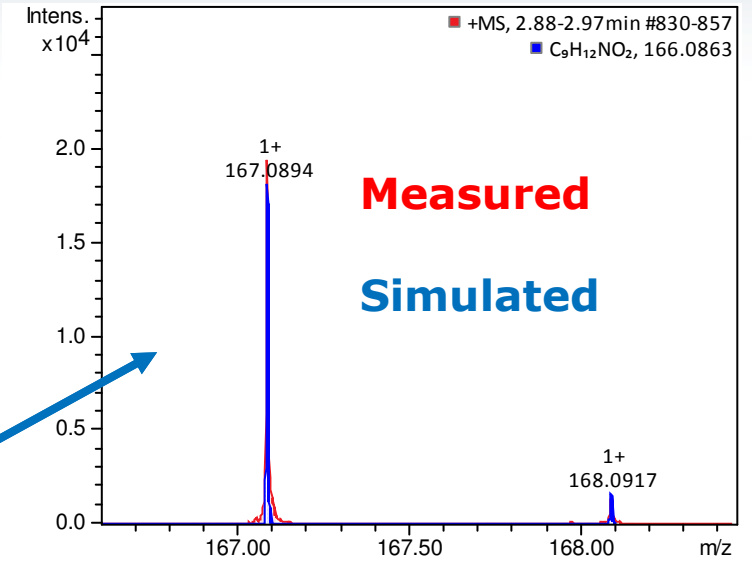
Peak area for phenylalanine deviates less than 3% (n=100 injections).

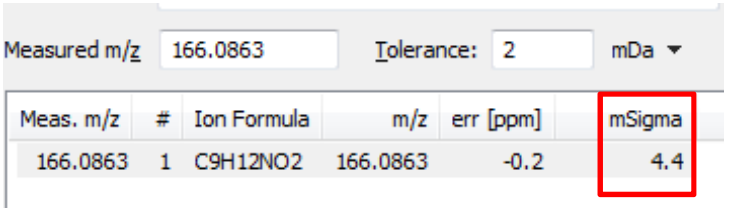
Example - Phenylalanine: **SmartFormula** provides the **correct molecular formula** based on accurate mass and isotopic pattern fit: $C_9H_{12}O_2$



SmartFormula makes use of the perfect fit between measured and simulated spectrum



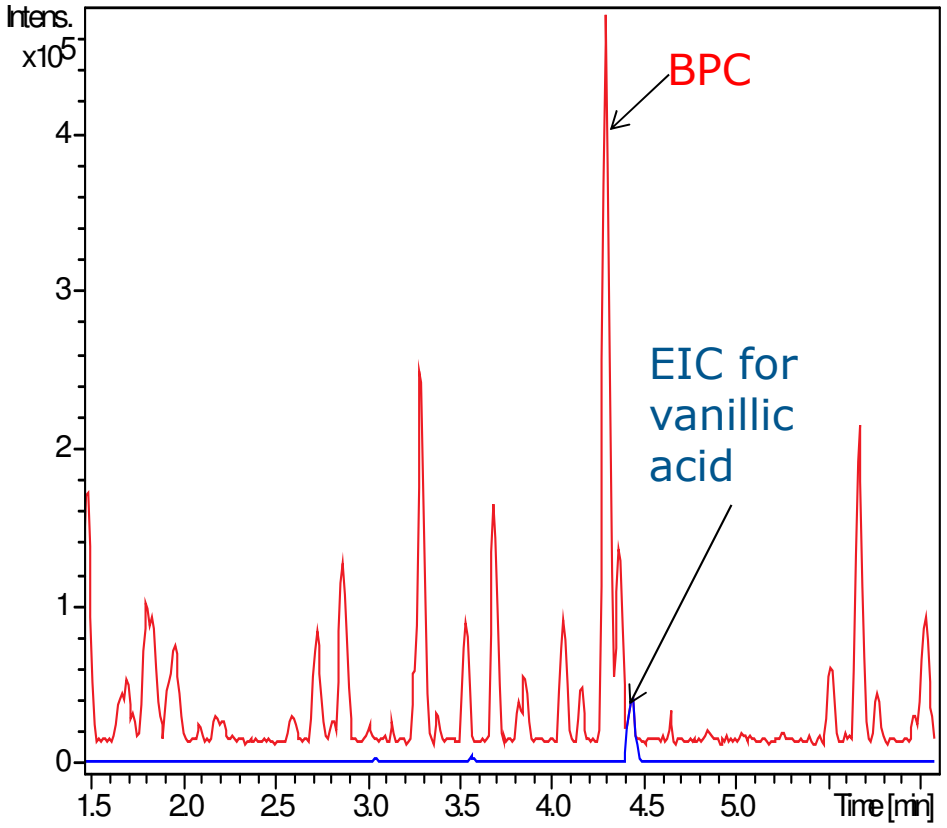




Mass accuracy & isotopic fidelity
= **confidence** in ID

expressed in a low mSigma value

Reliable ID of target compounds vanillic acid spiked into the urine sample

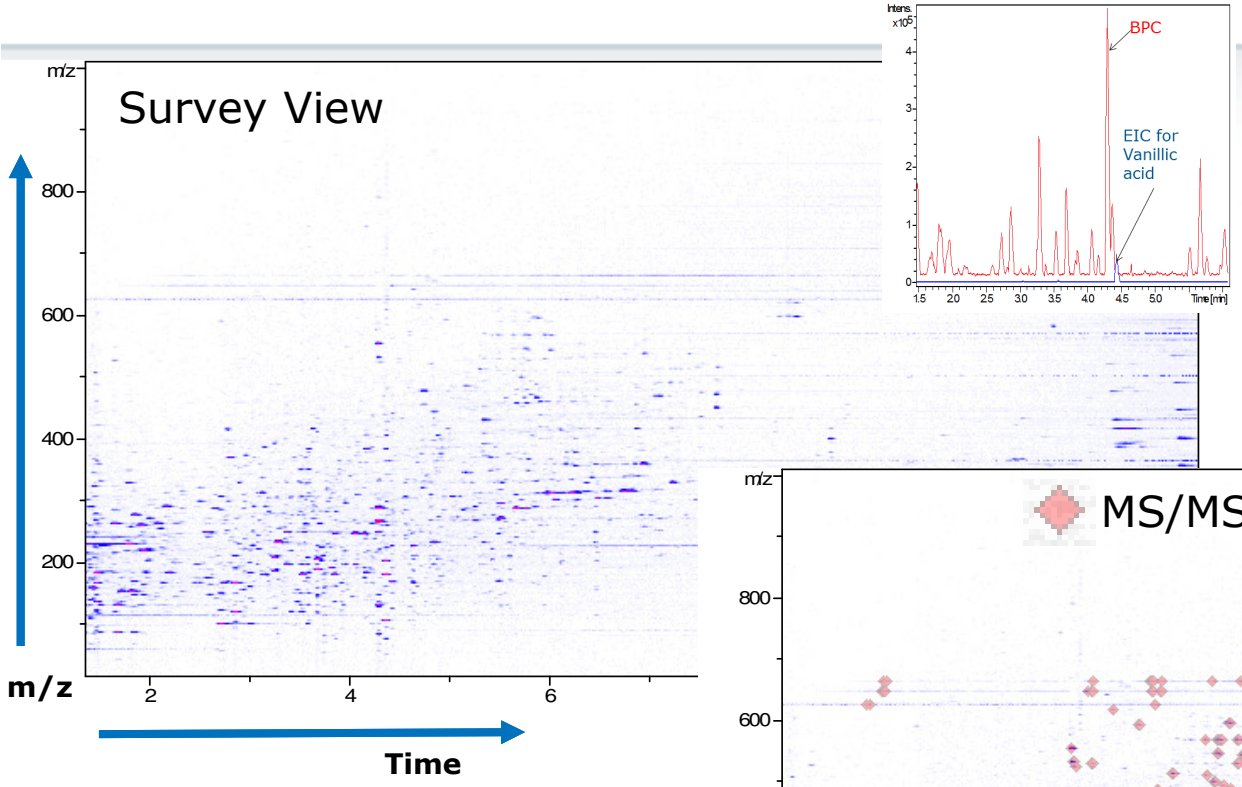


Target compound spiked into the human urine sample can be detected by creating a high resolution EIC trace for the target mass 169.0495m/z.

Detection by **h**igh **r**esolution EIC

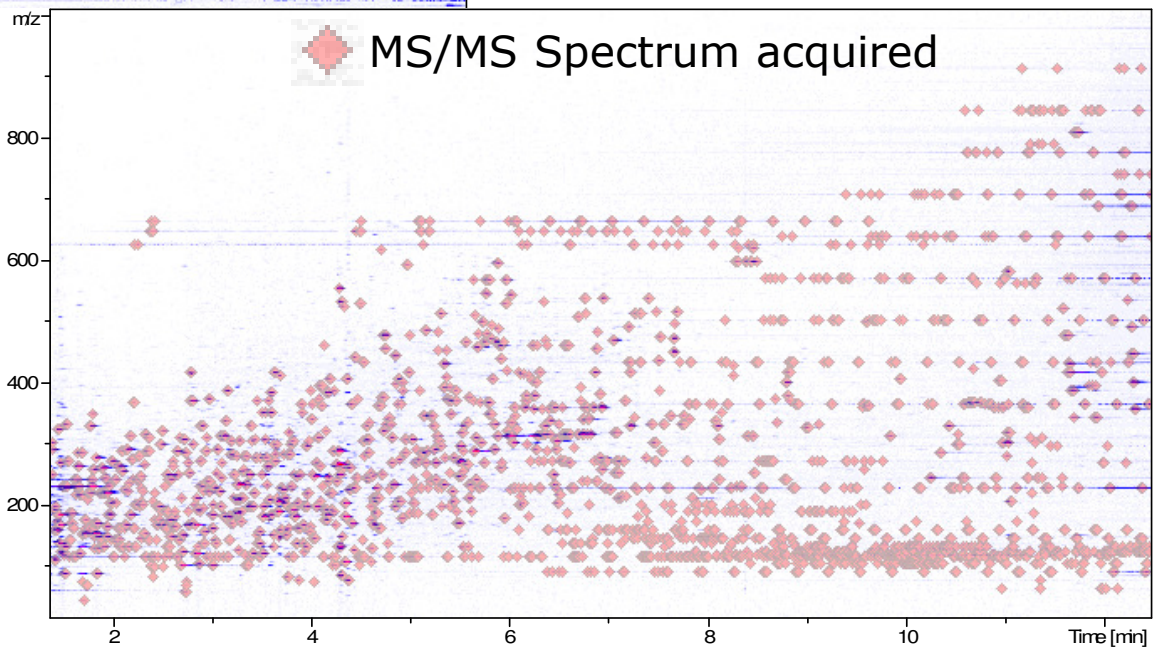
Basis for hrEIC
=
mass stability across entire dynamic concentration range

Impact II - High dynamic range in MS/MS:
up to 50Hz instrument scan speed combined with easy to use Instant Expertise software



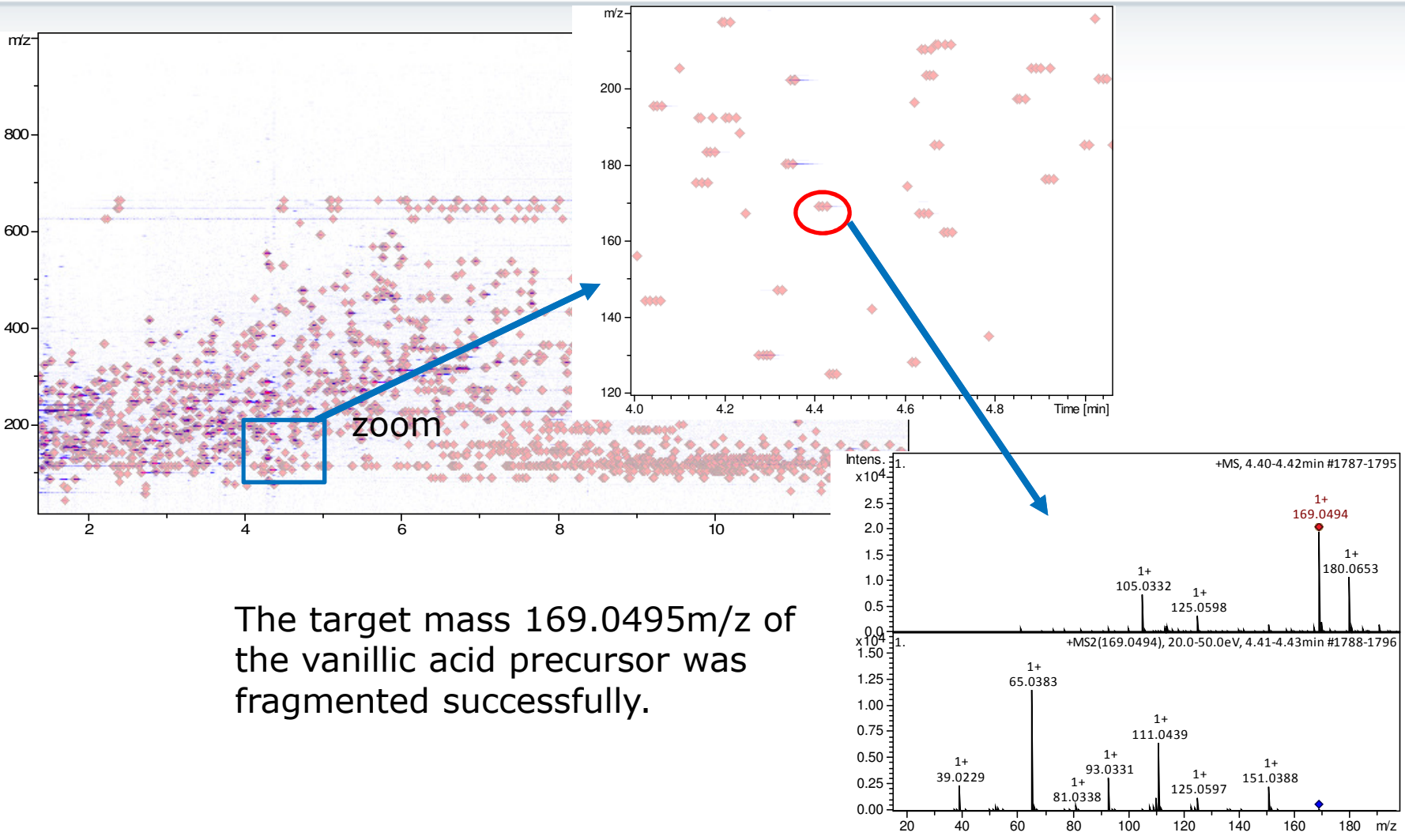
Human **urine** sample spiked with vanillic acid

- Impact II
- 4296 MS/MS spectra acquired at this sample complexity level



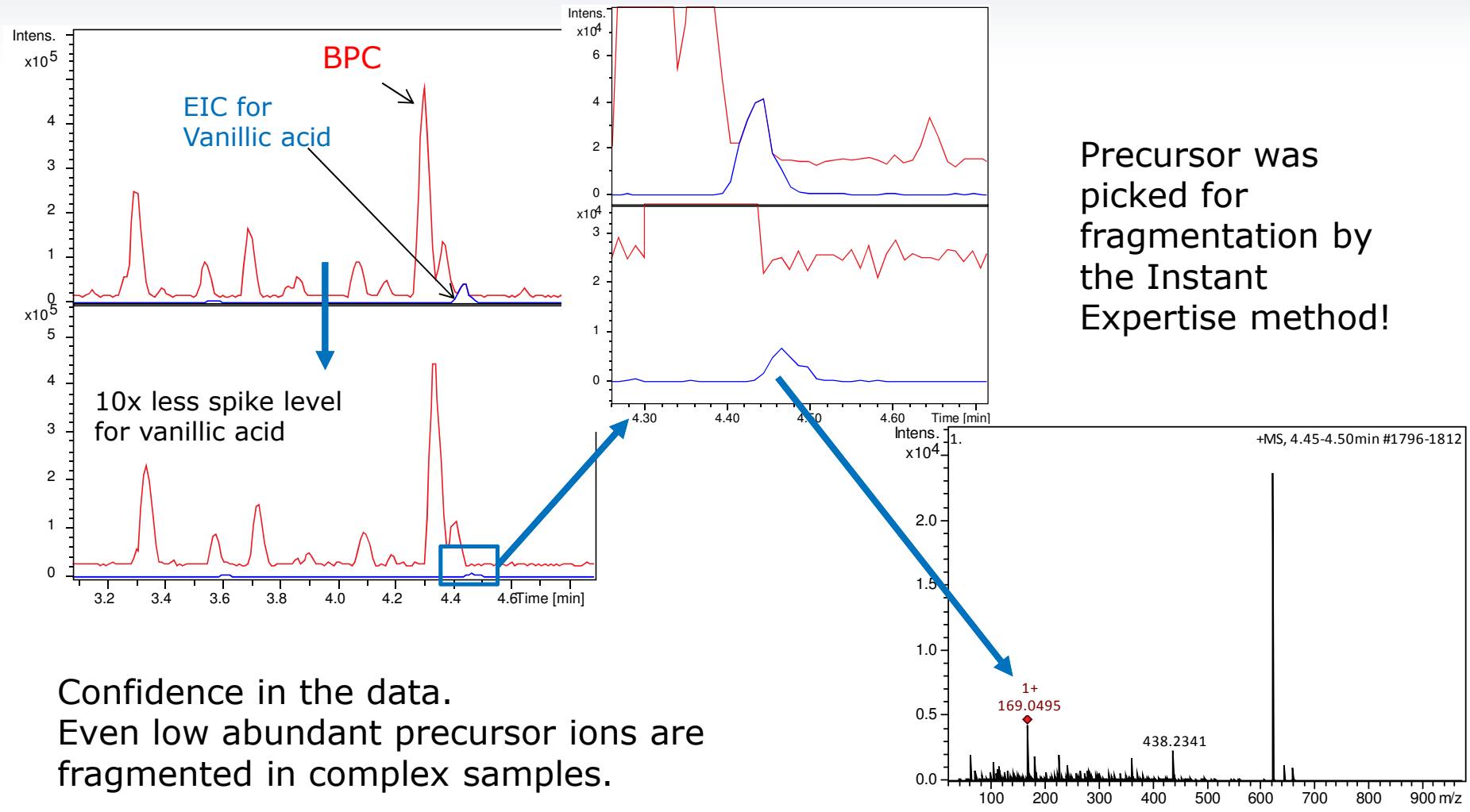
All precursor ions fragmented in **"one shot" acquisition**.

Instant Expertise selected vanillic acid spiked into the sample for MS/MS fragmentation



The target mass 169.0495m/z of the vanillic acid precursor was fragmented successfully.

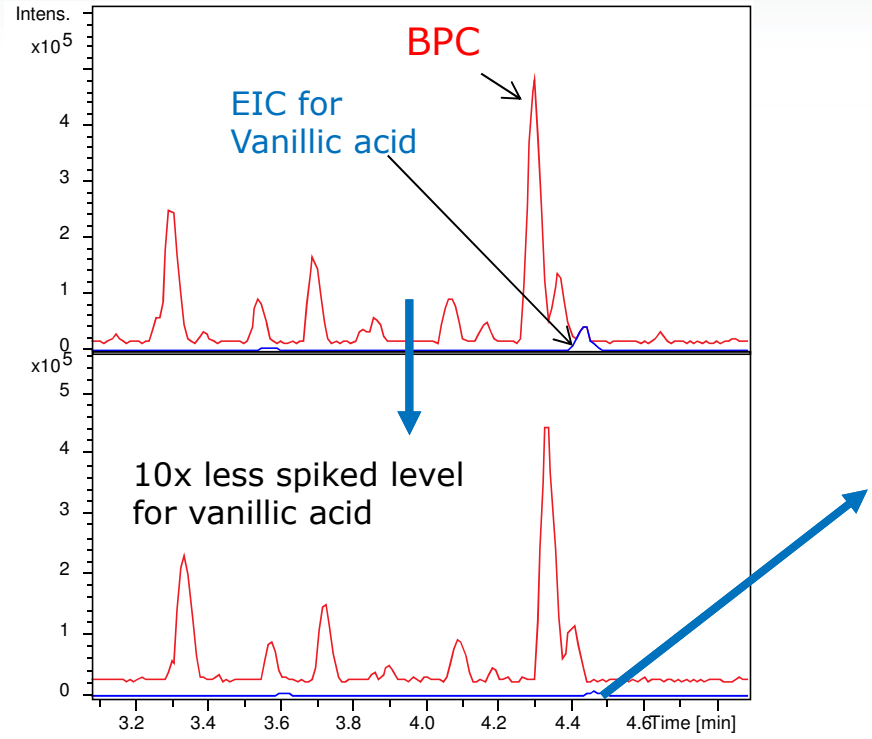
10x lower concentration is still fragmented
-> high dynamic range in MS & MS/MS data acquisition

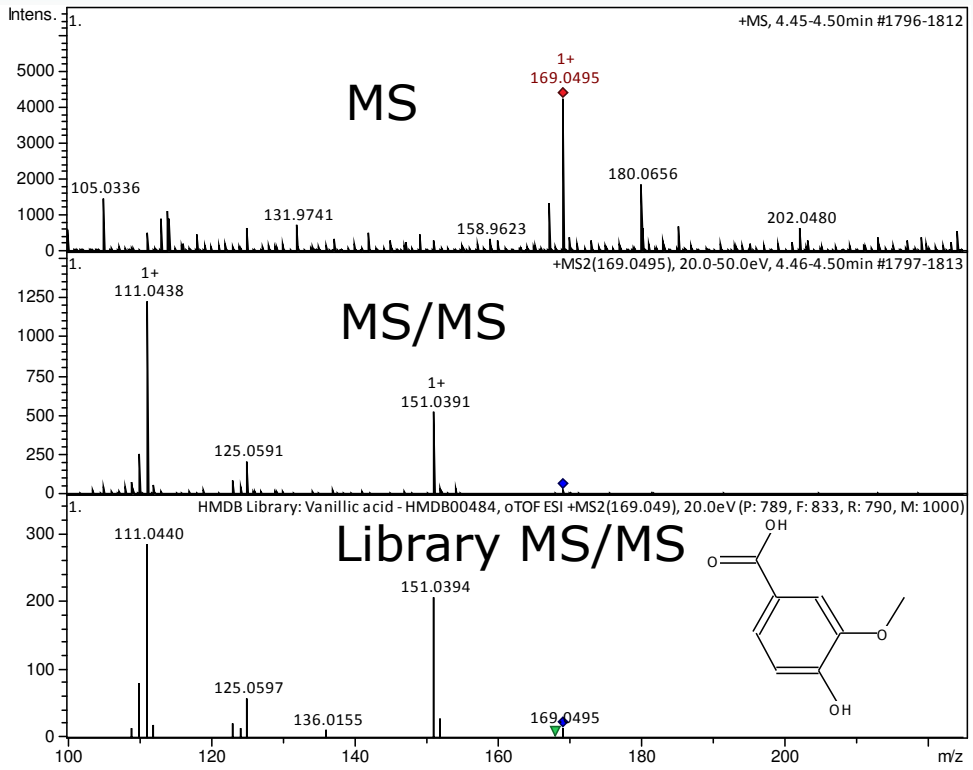


Precursor was picked for fragmentation by the Instant Expertise method!

Confidence in the data.
Even low abundant precursor ions are fragmented in complex samples.

MS/MS spectrum enables identification via **Bruker HMDB Metabolite Library** query





MS/MS data acquired permits confident ID via spectral library query.
In this case via the **Bruker HMDB Metabolite Library**.

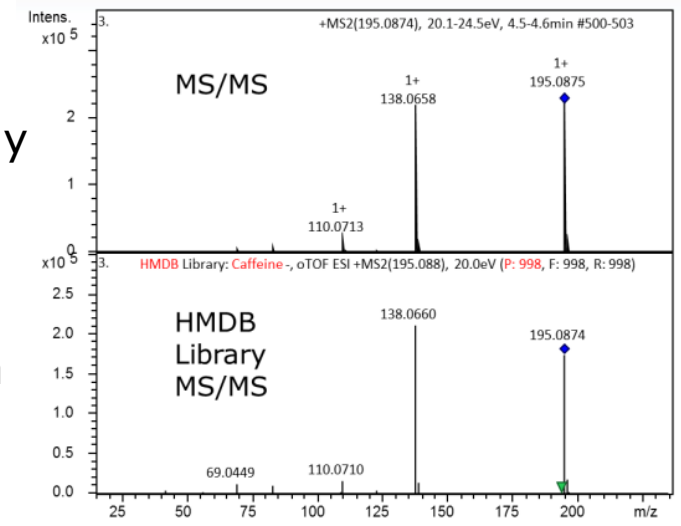
Deeper insights with the
Bruker HMDB Metabolite Library
"Human Metabolome Data Base" (HMDB)

Acquired in collaboration with the "**HMDB** team" at the University of Alberta. Lead by **Profs. Liang Li and David Wishart**.

- **~800** pure reference standards , measured on an **impact** HD **QTOF**
- up to **5** different **collision energies,** each mass spectrum **manually curated**

Germany

Canada



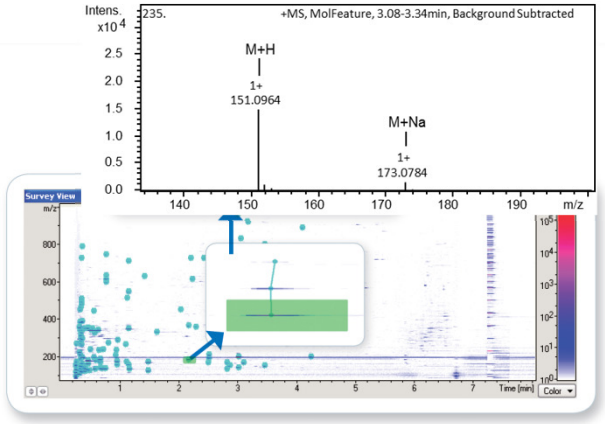
Perfect matching between Library and experimental spectrum.

Non- targeted & targeted Metabolomics

Both can be addressed using one ESI-TOF-MS data set

Non-targeted Metabolomics:

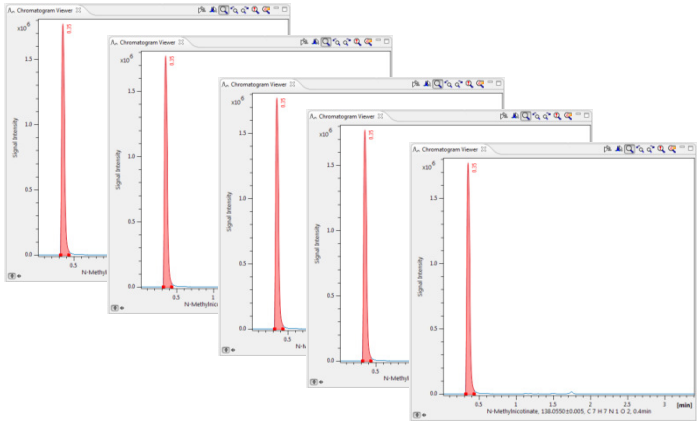
- "Think" extract all **Features** first



MetaboScape

Targeted Metabolomics:

- "Think" **hrEICs –** if you know what you are looking for



PathwayScreener

Acknowledgement




The study presented was conducted in collaboration with

Sven Heiling, Klaus Gase, Ian T. Baldwin
Max Planck Institute for Chemical Ecology, Molecular Ecology Department, Jena, Germany

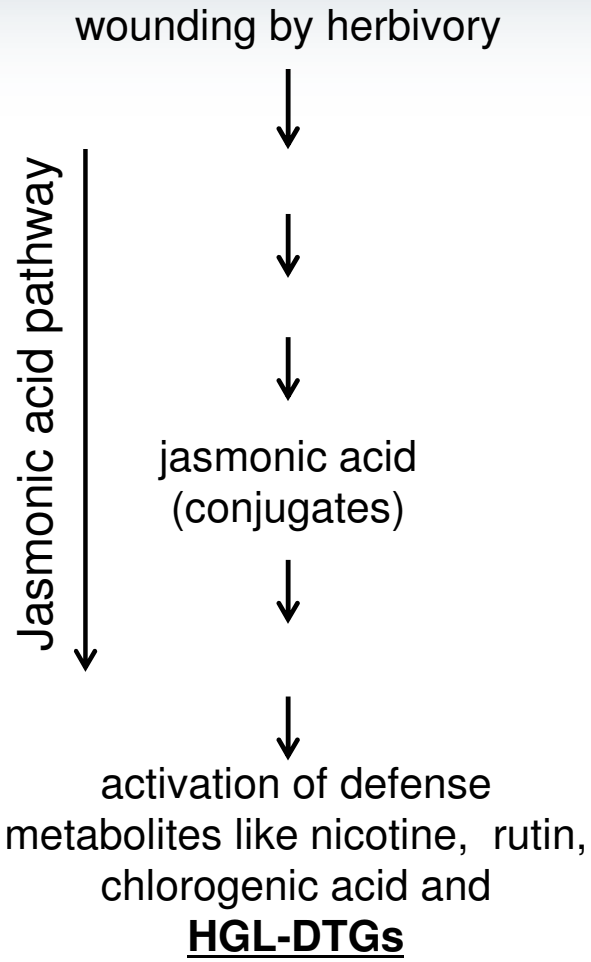
Emmanuel Gaquerel
Centre for Organismal Studies Heidelberg,
Plant Defense Metabolism Research Group, Heidelberg, Germany

Max Planck Society

Introduction – tobacco plants have enemies…

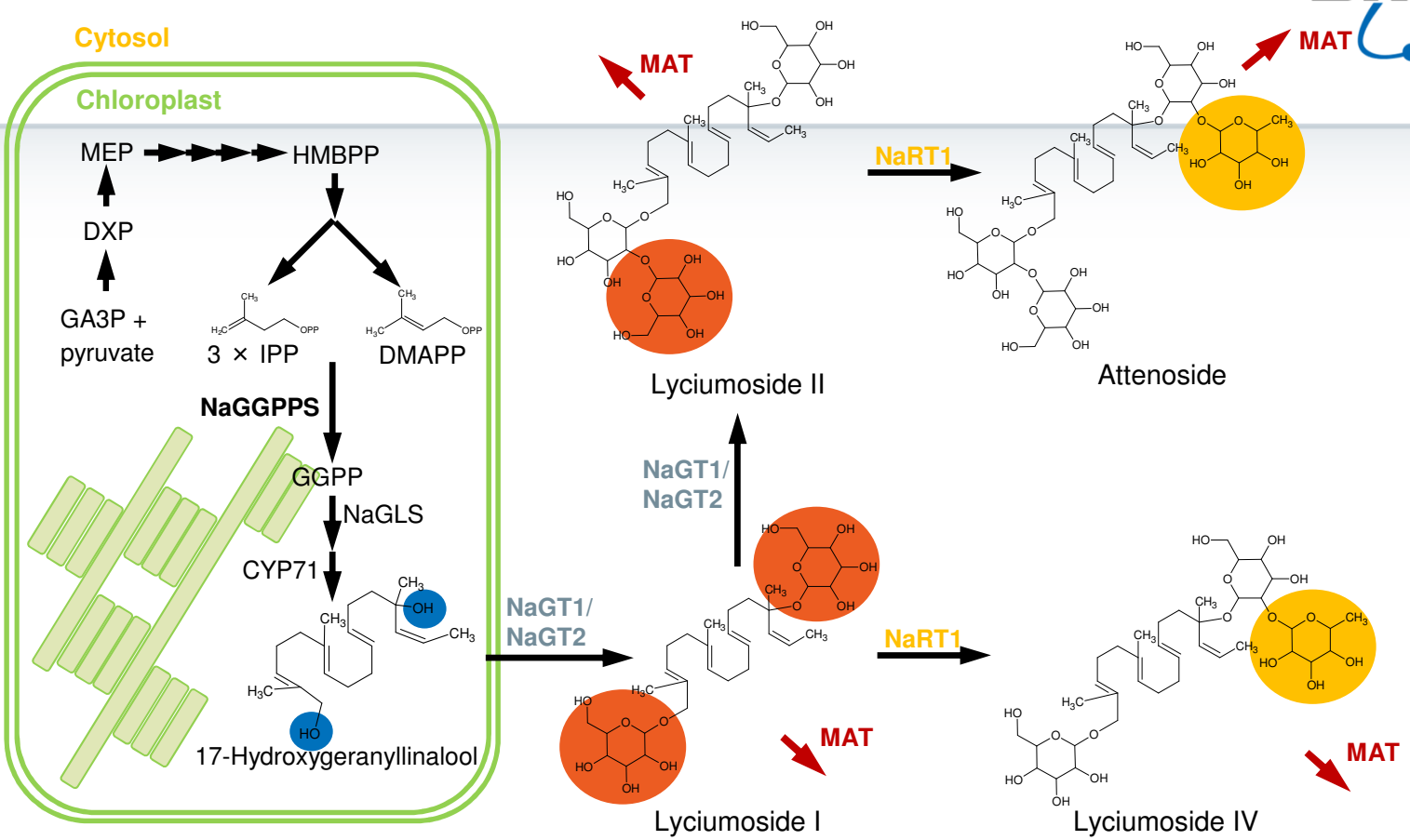
***Manduca sexta* induces the Jasmonic acid pathway which activates plant defense mechanisms**

- Specialist *Manduca sexta* – mayor pest on various solanaceae species
- feeds on green leaf material



Introduction

17-HGL-DTG biosynthesis



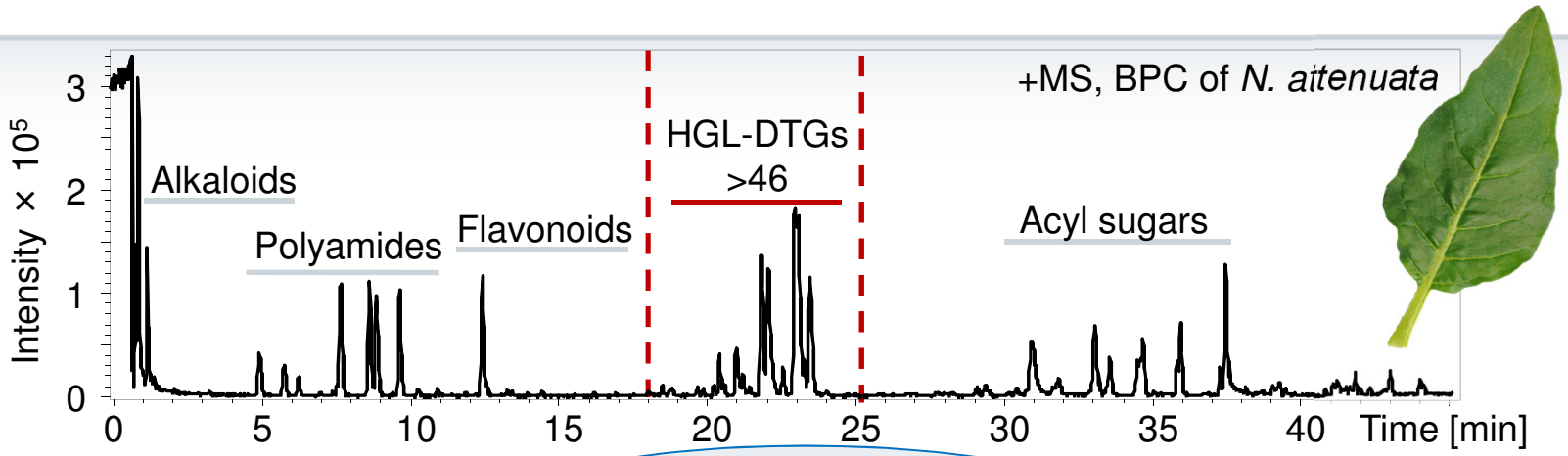
- 17-HGL is glucosylated at C-3 and C-17 hydroxy groups
- HGL-DTGs can be further glycosylated by either glucose or rhamnose
- HGL-DTGs can be malonylated up to three times
- Until now we identified **46 HGL-DTGs** in *N. attenuata*

These 2 genes will be important later in this talk

NaRT1

NaGT1

Databases of HRAM MS features



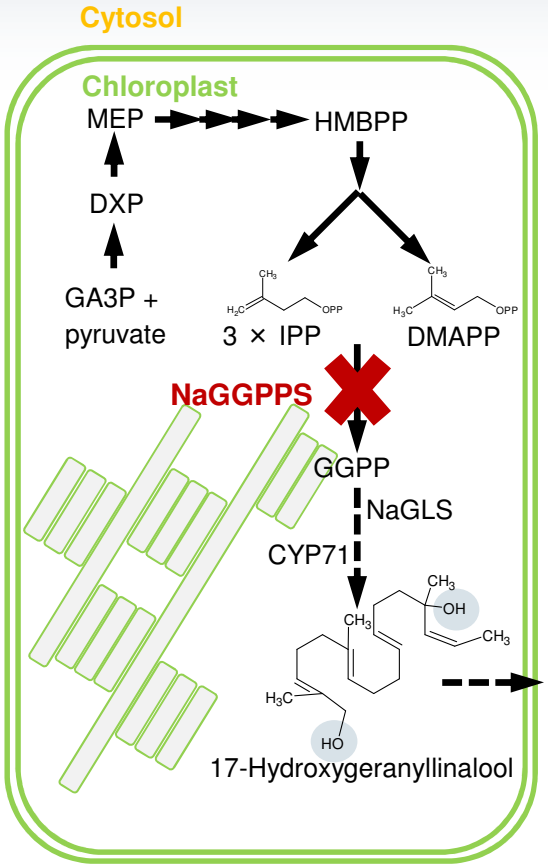
In-house database

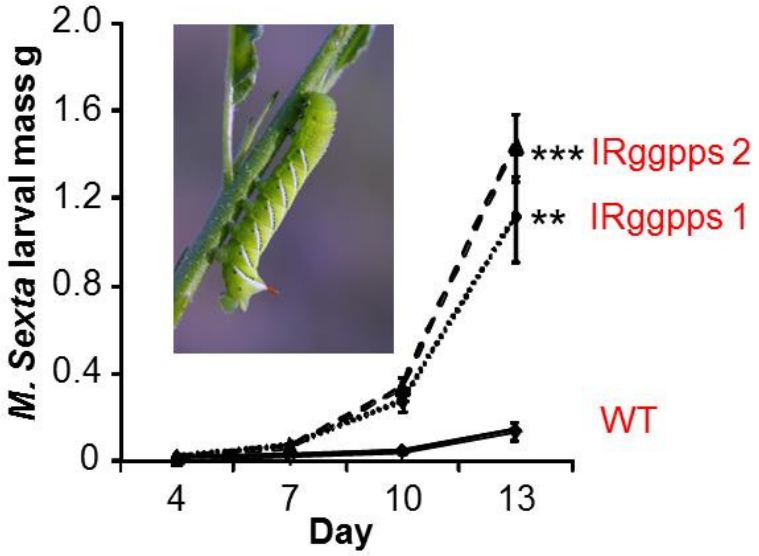
- high resolution mass spectrometry database
- 137 HGL-DTGs in 35 different solanaceous species
- 46 HGL-DTGs in *N. attenuata*
- detailed fragmentation information (IS-CID, MS, MS/MS) leading to 955 non-redundant explained fragments

In-house database

- 30 polyamides
- 15 acyl sugars
- 15 flavonoids
- 7 amino acids
- 4 alkaloids
- others

Ecological function of HGL-DTGs

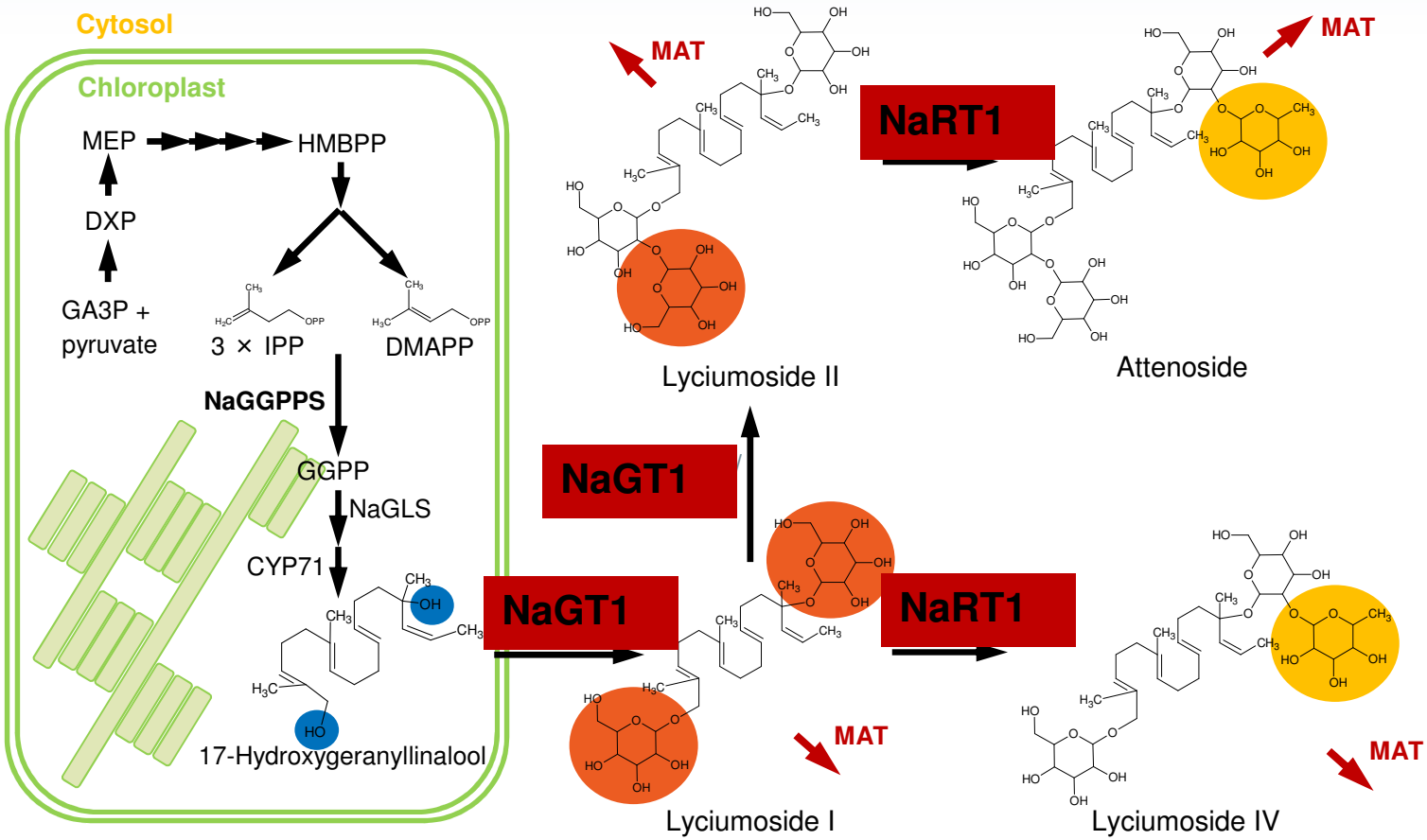




- Silencing NaGGPPS decreases the total HGL-DTG concentration
- Total HGL-DTGs have a significant effect on the growth of the specialist herbivore *M. sexta*

Heiling *et al.* 2010, *The Plant Cell*, 22, 273-292

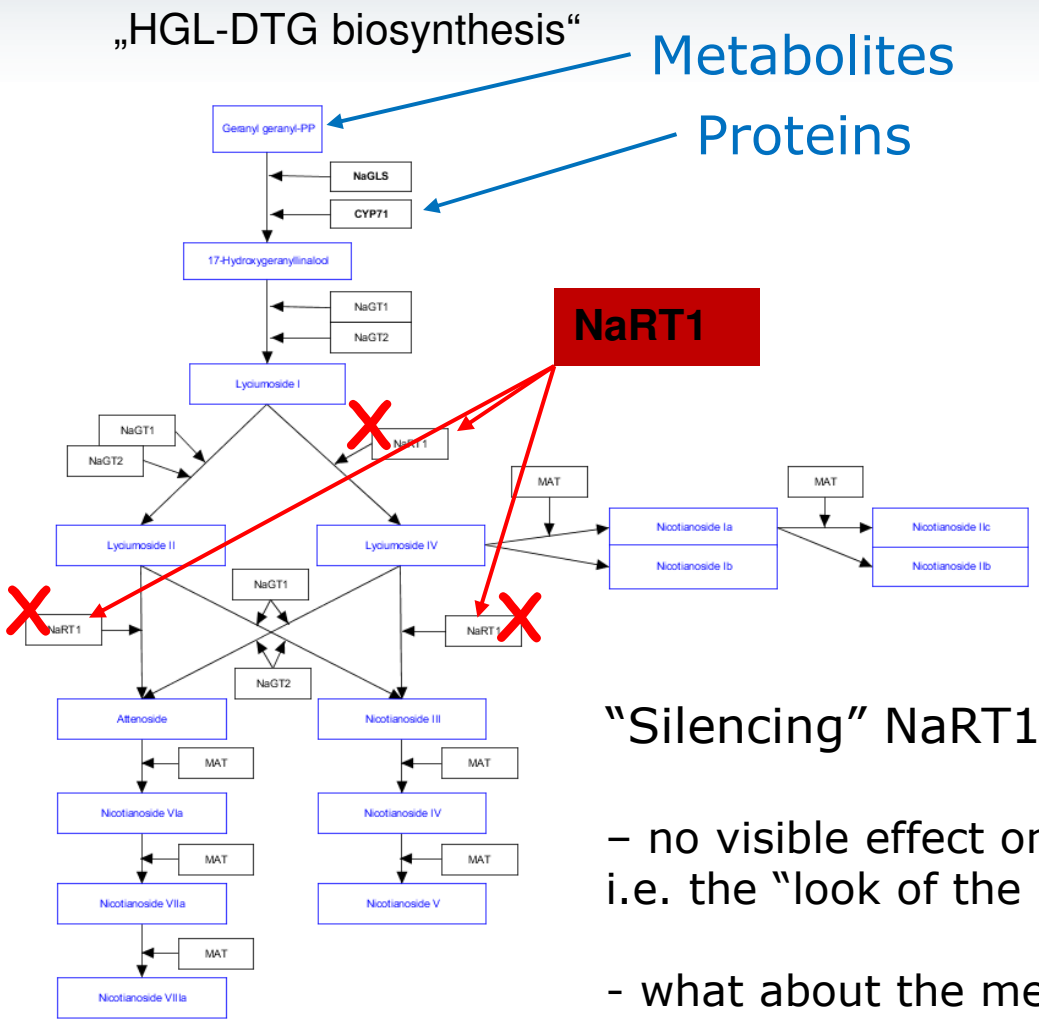
What is the biological effect if NaGT1 or NaRT1 are silenced?



"Silencing" NaRT1 has no visible effect on the plant phenotype

WT

NaRT1



"Silencing" NaRT1

– no visible effect on Phenotype, i.e. the "look of the plant"

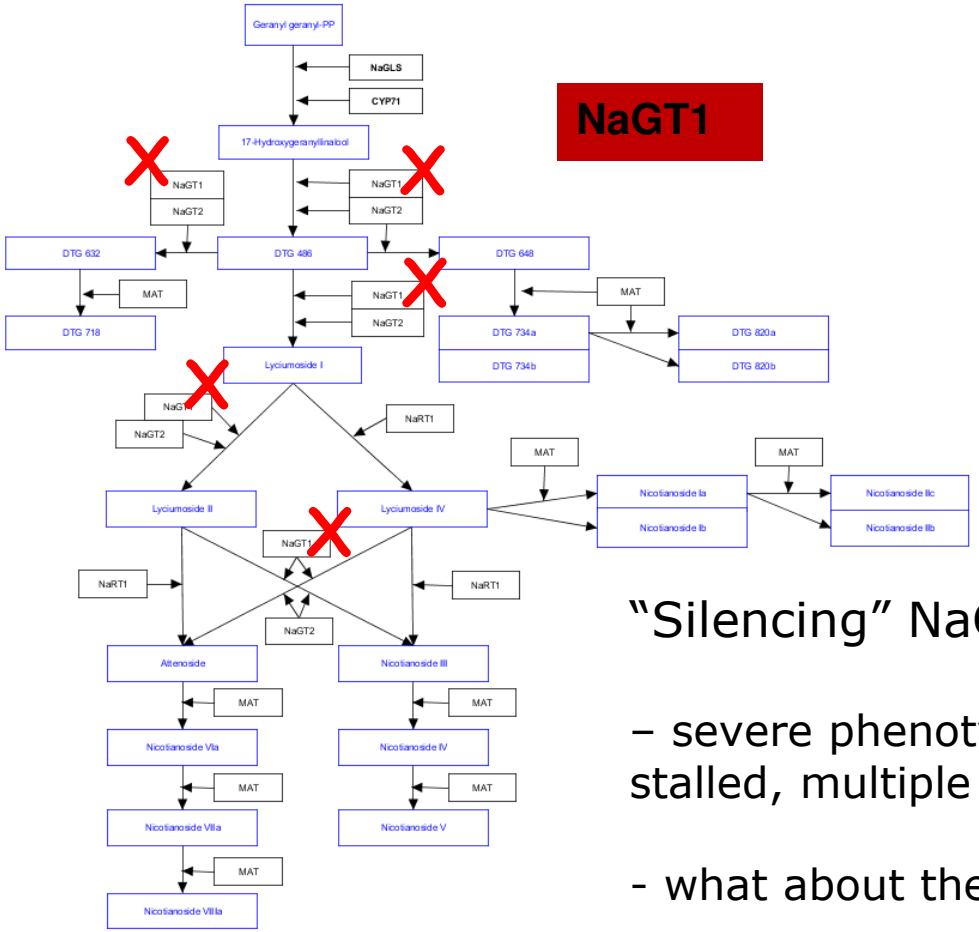
- what about the metabolome?

"Silencing" NaGT1 causes a severe plant phenotype, buds mostly stalled, multiple branching

WT

NaGT1

„HGL-DTG biosynthesis"



NaGT1

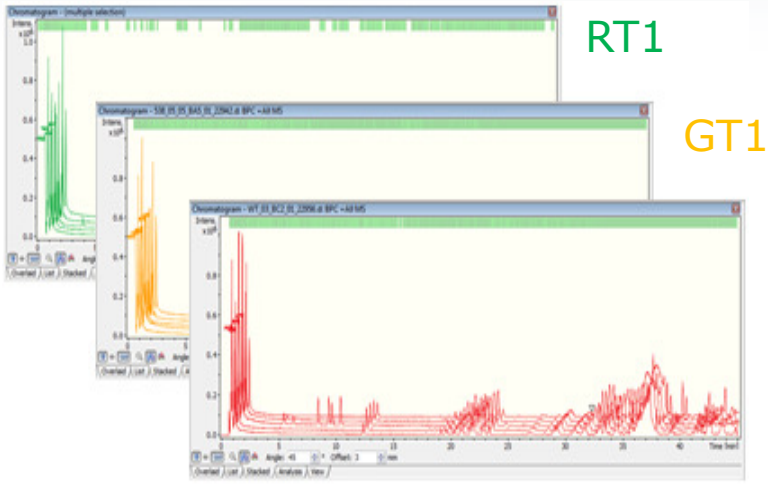
"Silencing" NaGT1

– severe phenotype, buds mostly stalled, multiple branching

- what about the metabolome?

Pathway designed by Sven Heiling with PathVisio

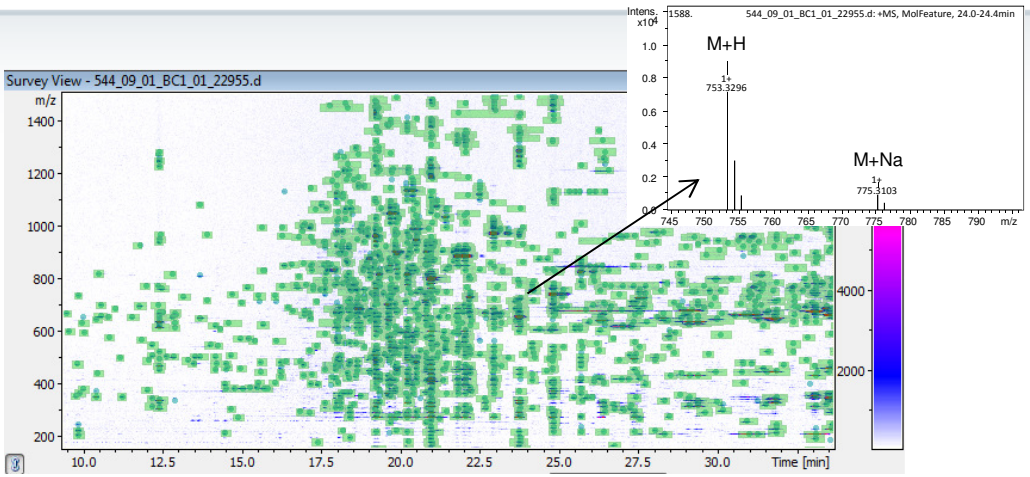
Metabolic profiling of WT, GT1 & RT1 tobacco plants



Seamless data acquisition of LC-MS/MS measurements for the three *N. attenuata* lines via **impact II** & **Instant Expertise™**

Metabolic profiling of WT, GT1 & RT1 tobacco plants

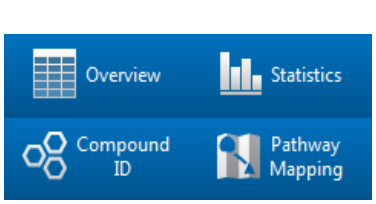
Seamless data evaluation by MetaboScape

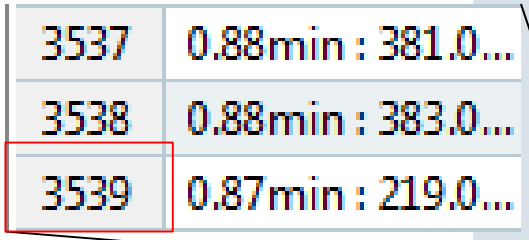


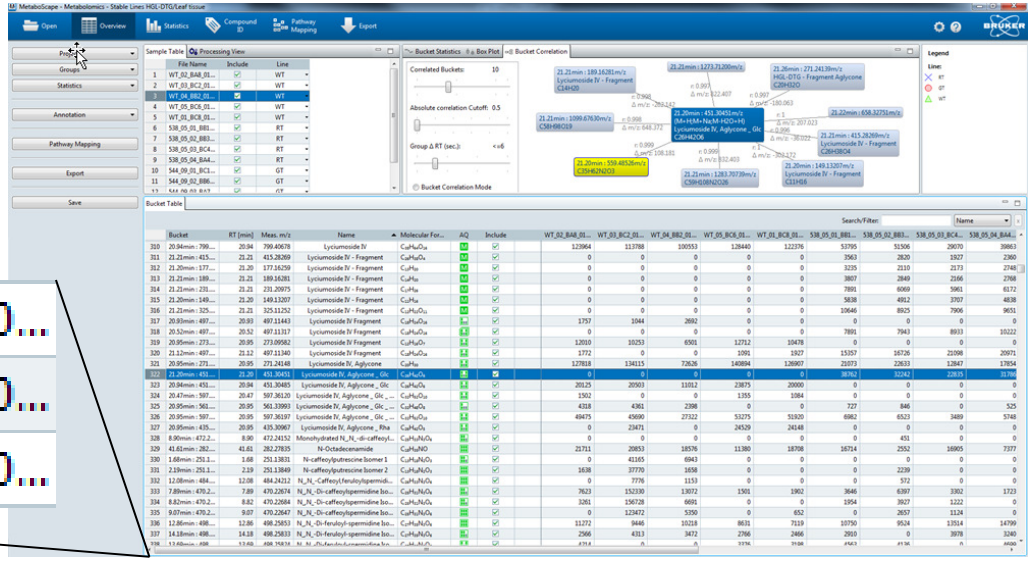
Comprehensive feature extraction by "Find Molecular Features" algorithm

- *Combines adducts, charge states and isotopes* belonging to the same compound detected within a sample
- ~3300 FMF compounds extracted per sample

- **Combining** extracted FMF features resulted in **3539 buckets** for further analysis in novel **MetaboScape** software



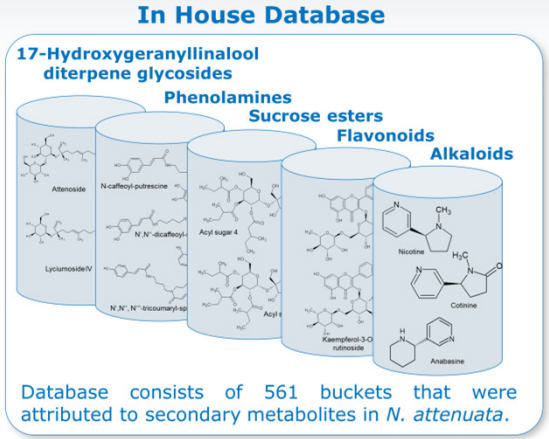


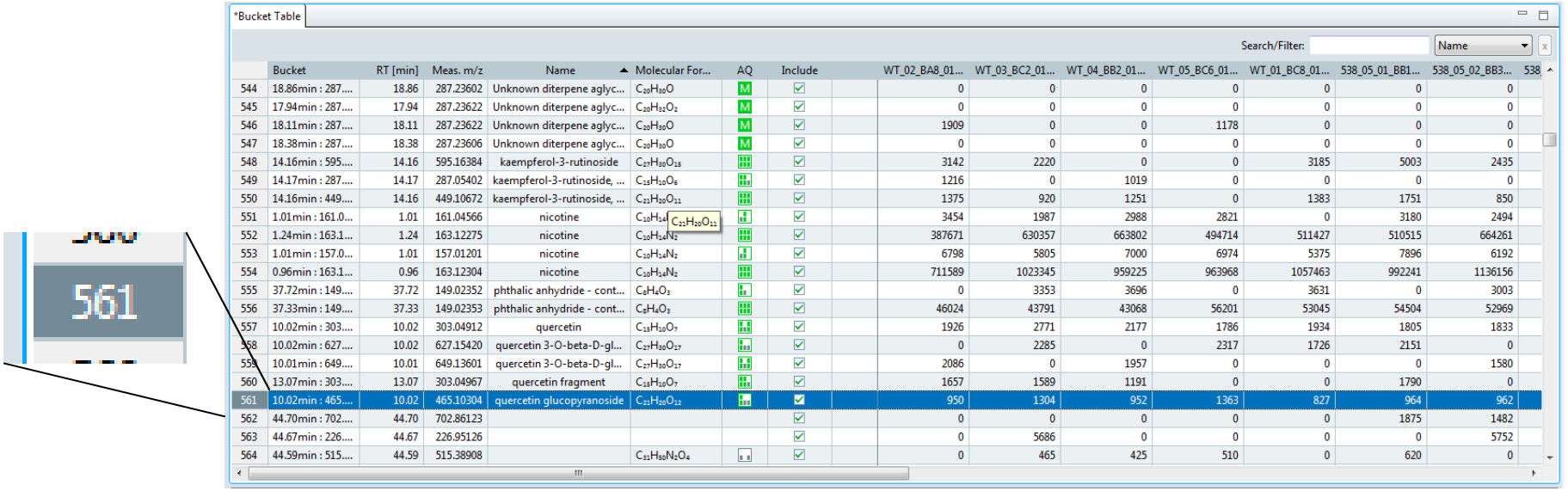


Automatic identification of known compound >550 buckets

Fast and confident de-replication using in-house DB

- **3539 buckets** from extracted FMF features in HRAM LC-MS data of samples from WT, GT1 and RT
- **Database** in **simple CSV format** of 561 fragments consists of molecular formula, name and retention time
- MetaboScape – with one push of a button: Automatically & confidently annotate **561 buckets**



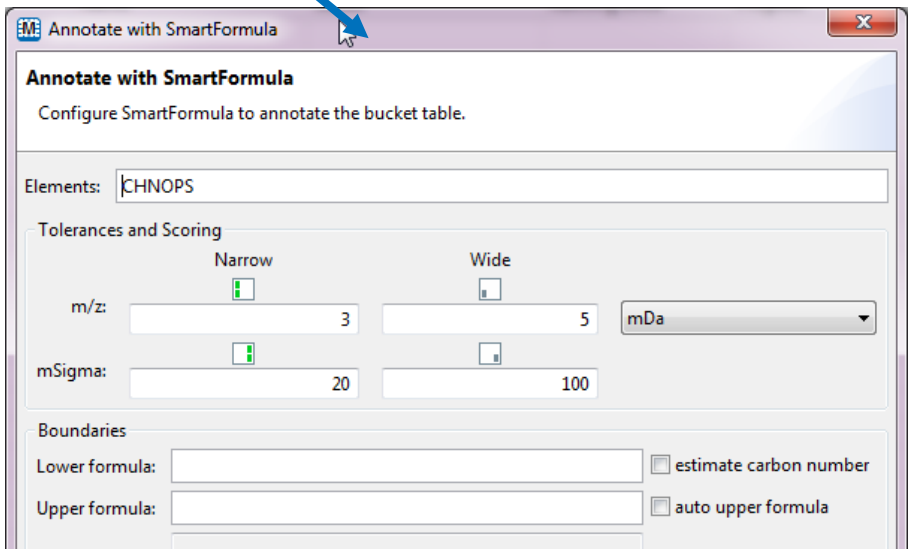


	Bucket	RT [min]	Meas. m/z	Name	Molecular For...	AQ	Include	WT_02_BA8_01...	WT_03_BC2_01...	WT_04_BB2_01...	WT_05_BC6_01...	WT_01_BC8_01...	538_05_01_BB1...	538_05_02_BB3...	538
544	18.86min : 287....	18.86	287.23602	Unknown diterpene aglyc...	$C_{20}H_{30}O$	M	☑	0	0	0	0	0	0	0	
545	17.94min : 287....	17.94	287.23622	Unknown diterpene aglyc...	$C_{20}H_{32}O_2$	M	☑	0	0	0	0	0	0	0	
546	18.11min : 287....	18.11	287.23622	Unknown diterpene aglyc...	$C_{20}H_{30}O$	M	☑	1909	0	0	1178	0	0	0	
547	18.38min : 287....	18.38	287.23606	Unknown diterpene aglyc...	$C_{20}H_{30}O$	M	☑	0	0	0	0	0	0	0	
548	14.16min : 595....	14.16	595.16384	kaempferol-3-rutinoside	$C_{27}H_{30}O_{15}$		☑	3142	2220	0	0	3185	5003	2435	
549	14.17min : 287....	14.17	287.05402	kaempferol-3-rutinoside, ...	$C_{15}H_{10}O_6$		☑	1216	0	1019	0	0	0	0	
550	14.16min : 449....	14.16	449.10672	kaempferol-3-rutinoside, ...	$C_{21}H_{20}O_{11}$		☑	1375	920	1251	0	1383	1751	850	
551	1.01min : 161.0...	1.01	161.04566	nicotine	$C_{10}H_{14}$		☑	3454	1987	2988	2821	0	3180	2494	
552	1.24min : 163.1...	1.24	163.12275	nicotine	$C_{10}H_{14}N_2$		☑	387671	630357	663802	494714	511427	510515	664261	
553	1.01min : 157.0...	1.01	157.01201	nicotine	$C_{10}H_{14}N_2$		☑	6798	5805	7000	6974	5375	7896	6192	
554	0.96min : 163.1...	0.96	163.12304	nicotine	$C_{10}H_{14}N_2$		☑	711589	1023345	959225	963968	1057463	992241	1136156	
555	37.72min : 149....	37.72	149.02352	phthalic anhydride - cont...	$C_8H_4O_3$		☑	0	3353	3696	0	3631	0	3003	
556	37.33min : 149....	37.33	149.02353	phthalic anhydride - cont...	$C_8H_4O_3$		☑	46024	43791	43068	56201	53045	54504	52969	
557	10.02min : 303....	10.02	303.04912	quercetin	$C_{15}H_{10}O_7$		☑	1926	2771	2177	1786	1934	1805	1833	
558	10.02min : 627....	10.02	627.15420	quercetin 3-O-beta-D-gl...	$C_{27}H_{30}O_{17}$		☑	0	2285	0	2317	1726	2151	0	
559	10.01min : 649....	10.01	649.13601	quercetin 3-O-beta-D-gl...	$C_{27}H_{30}O_{17}$		☑	2086	0	1957	0	0	0	1580	
560	13.07min : 303....	13.07	303.04967	quercetin fragment	$C_{15}H_{10}O_7$		☑	1657	1589	1191	0	0	1790	0	
561	10.02min : 465....	10.02	465.10304	quercetin glucopyranoside	$C_{21}H_{20}O_{12}$		☑	950	1304	952	1363	827	964	962	
562	44.70min : 702....	44.70	702.86123				☑	0	0	0	0	0	1875	1482	
563	44.67min : 226....	44.67	226.95126				☑	0	5686	0	0	0	0	5752	
564	44.59min : 515....	44.59	515.38908		$C_{31}H_{50}N_2O_4$		☑	0	465	425	510	0	620	0	

Seamless identification of unknowns

Automated molecular formula generation using SmartFormula

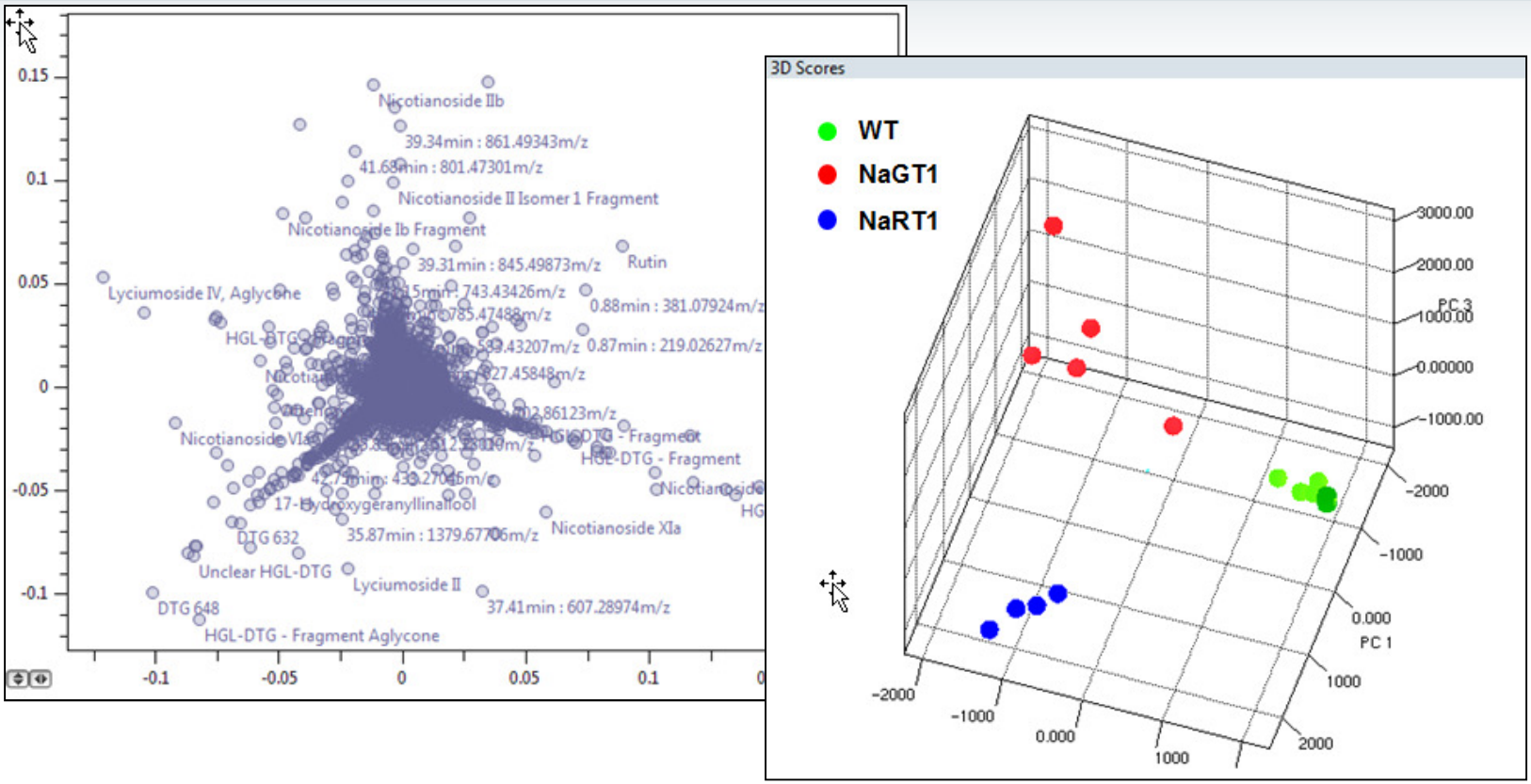
	Bucket	RT [min]	Meas. m/z	Name	Molecular Formula	AQ
2424	21.05min : 831.344m/z	21.05	831.344		$C_{49}H_{42}N_{12}S$	
2425	17.92min : 831.394m/z	17.92	831.394		$C_{38}H_{64}O_{18}$	
2426	17.68min : 831.394m/z	17.68	831.394		$C_{32}H_{61}N_{10}O_{10}PS$	
2427	37.15min : 831.450m/z	37.15	831.450		$C_{32}H_{55}N_{20}O_5P$	
2428	38.11min : 831.475m/z	38.11	831.475		$C_{37}H_{50}N_{24}$	



- Remaining ~ **3000 features automatically assigned** with candidate molecular formula with SmartFormula™

Quickly identify relevant information in complex data sets by statistical evaluation

PCA analysis separates WT, GT1, and RT1

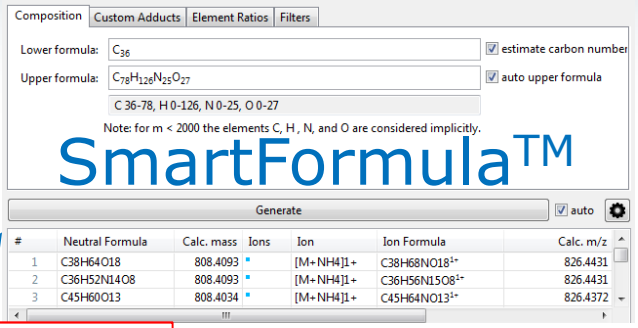


- PCA scores plot separates metabolites extracts from wild tobacco plants and the plants transformed to alter the expression of rhamnosyltransferase (RT1) and glycosyltransferase (GT1) genes.
- Loadings plot points to relevant compounds for further investigation.

Seamless identification of relevant compounds

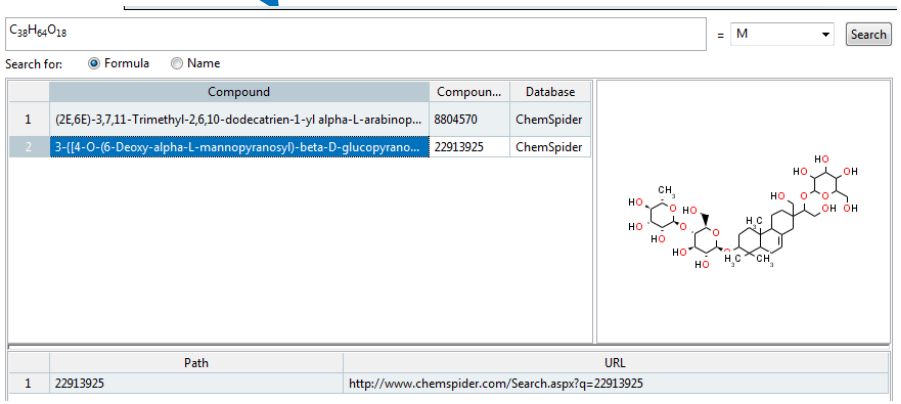
Automated molecular formula generation using SmartFormula & structural assignment through DB query

- ~ **3000 features automatically assigned** with candidate molecular formula with SmartFormula™



	Bucket	RT [min]	Meas. m/z	Name	Molecular Formula	AQ
2424	21.05min : 831.344m/z	21.05	831.344		$C_{49}H_{42}N_{12}S$	
2425	17.92min : 831.394m/z	17.92	831.394		$C_{38}H_{64}O_{18}$	
2426	17.68min : 831.394m/z	17.68	831.394		$C_{32}H_{61}N_{10}O_{10}PS$	
2427	37.15min : 831.450m/z	37.15	831.450		$C_{32}H_{55}N_{20}O_5P$	
2428	38.11min : 831.475m/z	38.11	831.475		$C_{37}H_{50}N_{24}$	

- Followed by structural assignment through public database (KEGG*, CheBI, PubChem) queries.



* KEGG Authorized Service Provider

MetaboScape 1.0

Pathway mapping of identified target compounds links data to the underlying biology

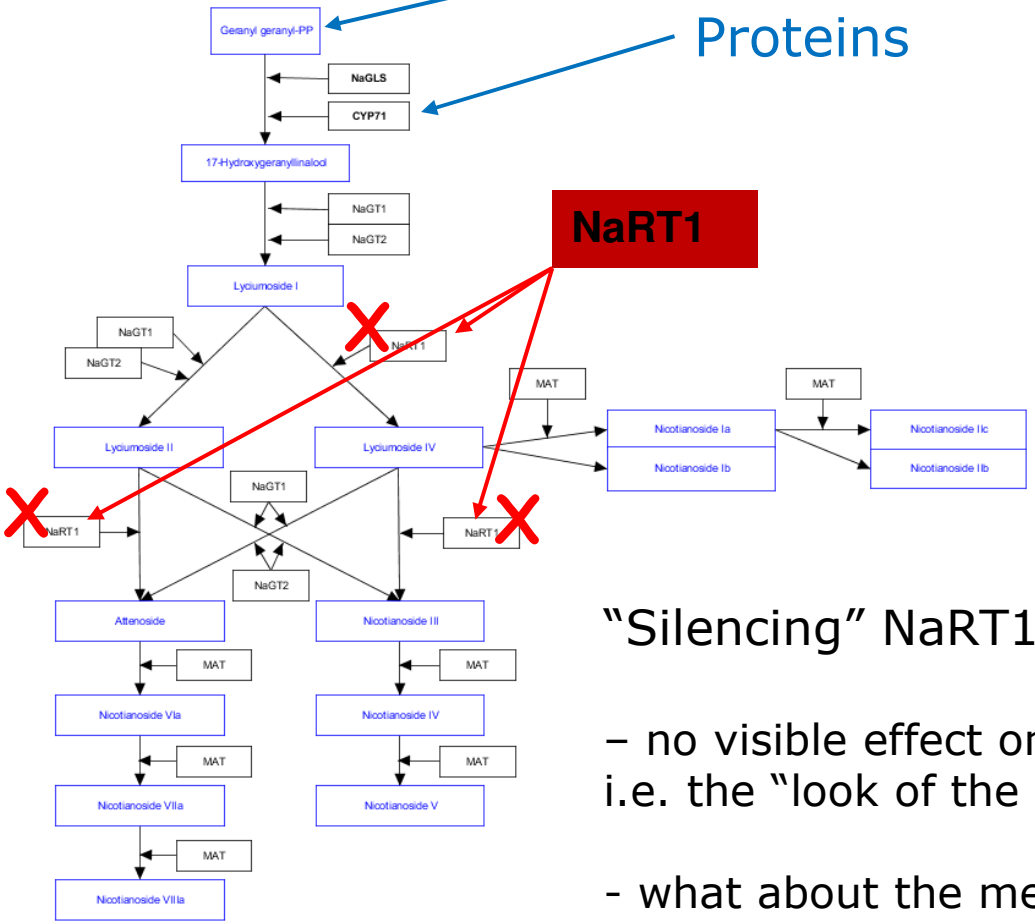
WT

NaRT1

„HGL-DTG biosynthesis“

Metabolites

Proteins



NaRT1

“Silencing” NaRT1

– no visible effect on Phenotype, i.e. the “look of the plant”

- what about the metabolome?

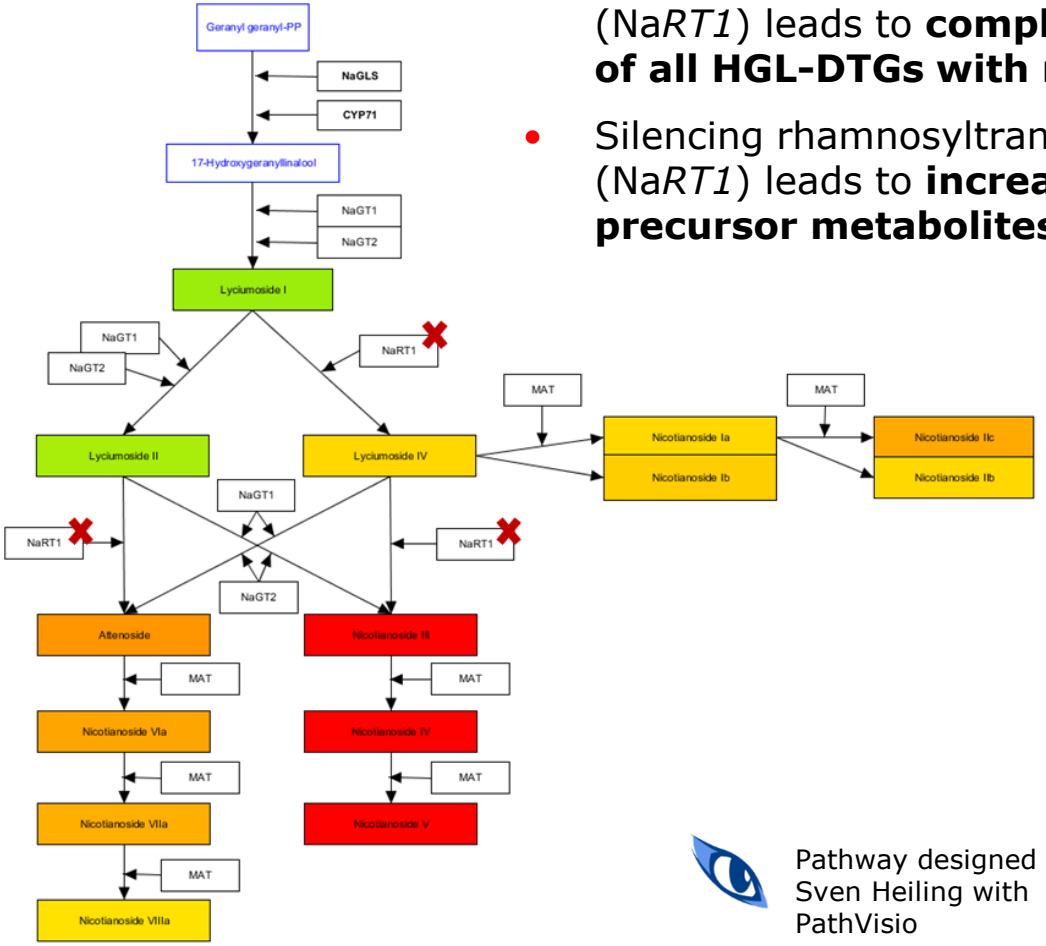
MetaboScape 1.0

Pathway mapping of identified target compounds links data to the underlying biology

„HGL-DTG biosynthesis“

WT

NaRT1



- Silencing rhamnosyltransferase (Na*RT1*) leads to **complete reduction of all HGL-DTGs with rhamnose**
- Silencing rhamnosyltransferase (Na*RT1*) leads to **increase in precursor metabolites**

Pathway designed by Sven Heiling with PathVisio

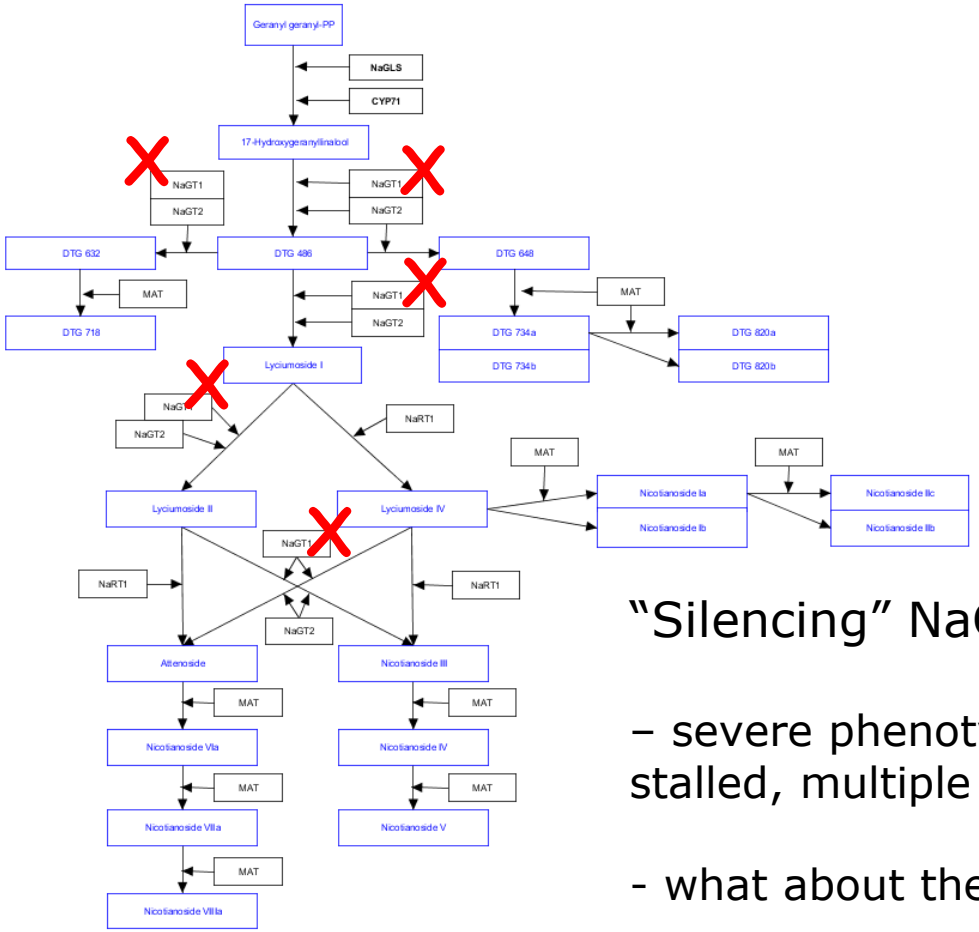
MetaboScape 1.0

Pathway mapping of identified target compounds links data to the underlying biology

„HGL-DTG biosynthesis“

WT

NaGT1



“Silencing” NaGT1

– severe phenotype, buds mostly stalled, multiple branching

- what about the metabolome?

Pathway designed by Sven Heiling with PathVisio

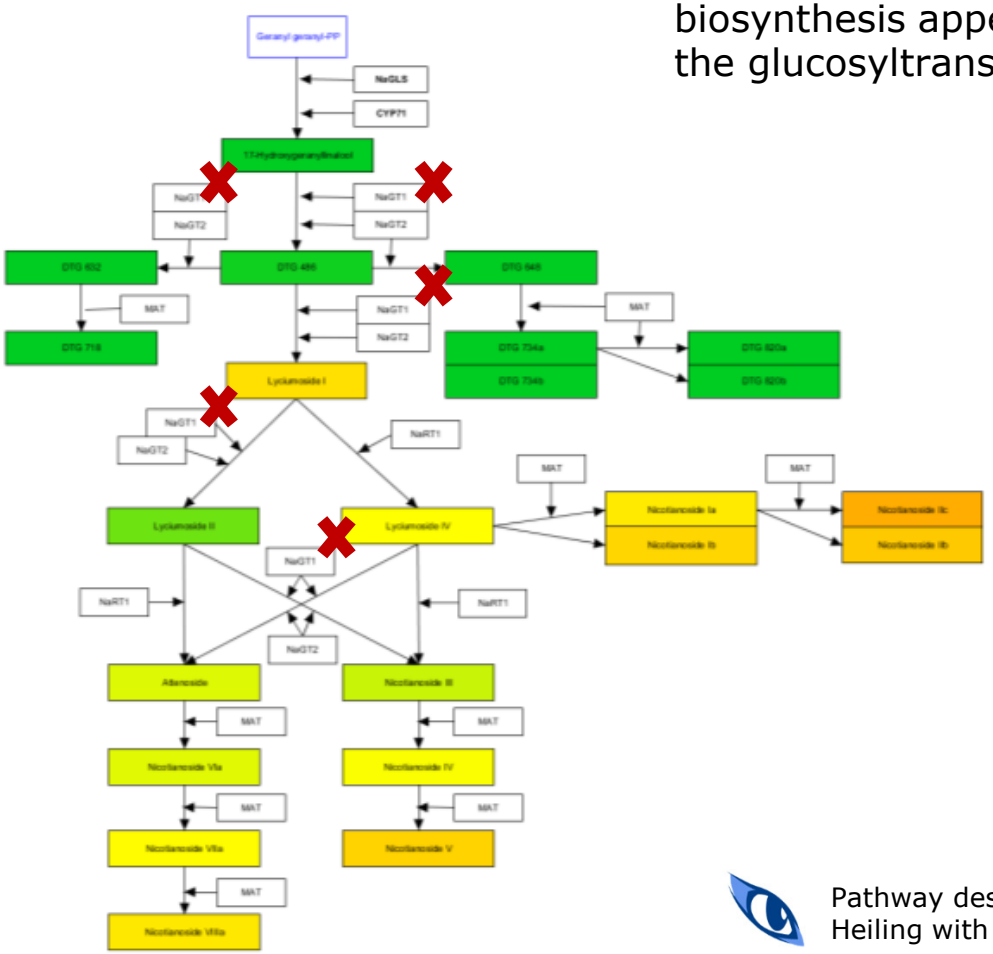
MetaboScape 1.0

Pathway mapping of identified target compounds links data to the underlying biology

WT

NaGT1

„HGL-DTG biosynthesis“

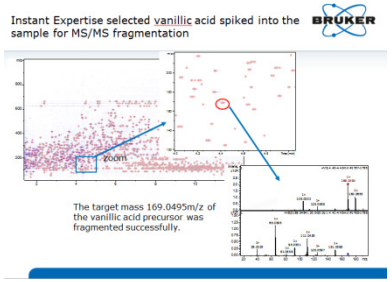


- Intermediate products of HGL-DTGs biosynthesis appear after silencing the glucosyltransferase NaGT1

Pathway designed by Sven Heiling with PathVisio

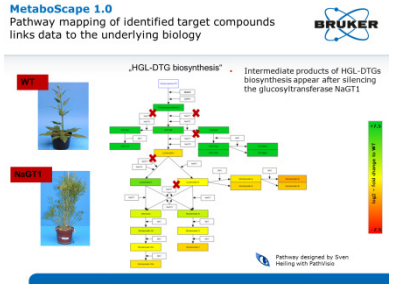
Outline -> Summary

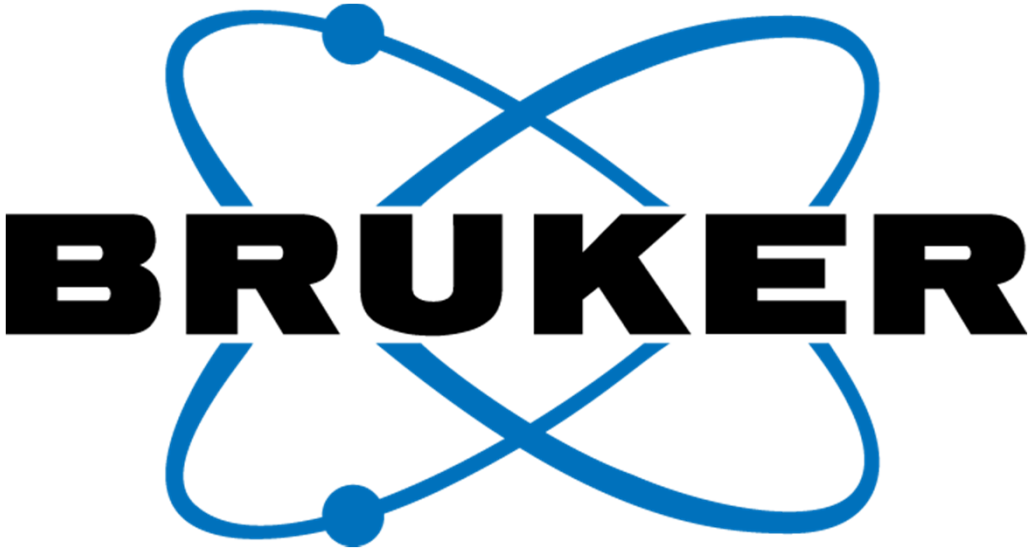
- If you want to automate data processing workflows in Metabolomics you have to be able to rely on the raw data.
  - Robustness and Sensitivity and Dynamic range and High resolution, mass accuracy and isotopic fidelity in one scan mode at LC speed is absolute must in Metabolomics: Impact II



- Introducing MetaboScape. Turning MS data into knowledge!
  - MetaboScape combines statistical data mining, automatic compound identification routines, and pathway mapping functionalities for linking MS data to biology in discovery metabolomics.

  - Here, enabling to detected and link diverse biological effects in genetically silenced lines of *N. attenuata* impaired in the biosynthetic steps of HGL-DTGs.





www.bruker.com