Supporting Information

Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory

Dapeng Li, Sven Heiling, Ian T. Baldwin and Emmanuel Gaquerel

Materials and Methods

Conditions for UHPLC-ESI/qTOF-MS analysis

An Acclaim column (150×2.1 mm, particle size 2.2 µm) with a 4 mm×4 mm guard column of the same material was used for the analysis. The following binary gradient was used with a Dionex Ultimate 3000 UHPLC system: 0 to 1 min, isocratic 90% A (de-ionized water, 0.1% [vol/vol] acetonitrile and 0.05% formic acid), 10% B (acetonitrile and 0.05% formic acid); 1 to 40 min, gradient phase to 15% A, 85% B; 40 to 45 min, isocratic 15% A, 85% B. Flow rate was 300 µL/min. Eluted compounds were detected by a high-resolution micrOTOF-Q II mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization source operating in positive ionization mode. Typical instrument settings were as follows: capillary voltage 4500 V, capillary exit 130 V, dry gas temperature 180°C, dry gas flow of 8 L/min. Ions were detected from *m/z* 50 to 1400 at a repetition rate of 1 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution of NaOH in 50/50% vol/vol isopropanol/water containing 0.2% formic acid). Raw data files were converted to netCDF format using the export function of the Data Analysis v4.0 software (Bruker Daltonics, Bremen, Germany).

Additional rules for the assembly of compound-specific idMS/MS

To reduce false positive errors resulting from spurious correlations from background noise due to the fact that some *m/z* features are only detected in a few samples, we compared data processing results obtained with and without the “fill peaks” function of XCMS (use for

background noise correction) and calculated a background noise value from the average correction estimate used by this function to replace “NA” not detected peak intensities. When the “fill peaks” function is used, there still were many “0” intensity values in the dataset which affect the calculation of correlations, and these were replaced with the calculated background value. We also only considered features with intensities that were more than 3 times the background value and considered these as “true peaks”. Only *m/z* signals with at least six “true peaks” for the 28 samples precursors (MS1) and fragments datasets were considered for PCC calculation.
A precursor mass feature is further defined if its intensities across samples significantly correlate with the decreased intensities of the same mass feature subjected to low or high collision energies and that this feature is not annotated as an isotope peak by CAMERA. The correlation analysis was then conducted by calculating all possible precursor-to-product pairs within 9s – estimated maximum retention time window for peak deviation. Logically, *m/z* values for fragments should be lower than that of the precursor and MS/MS fragmentation should occur in the same sample position within the 28 sample dataset as the precursor from which it is derived.

Based on these two simple rules, we excluded assigned fragments at *m/z* values larger than that of the identified precursor as well as fragments mismatched with precursor sample position. Many in-source-fragmentation-generated mass features produced in the MS1 mode can also be selected as candidate precursors resulting in redundant compound idMS/MSs. To reduce such data redundancy, we merged spectra if their NDP similarity exceeded 0.9 and they belong to the chromatographic “pcgroup” annotated by CAMERA. Finally we merged all the 4 collision energy results for precursor-to-fragment associations into a final deconvoluted spectrum by choosing the highest intensity peak among all candidate peaks of the same *m/z* value at the different collision energies. This latter processing step is based on the composite spectrum concept and accounts for the different collision energy conditions required to maximize fragmentation possibilities since certain fragments are detected only at specific collision energies. After applying the entire pipeline and set of rules, 895 deconvoluted non-redundant spectra were reconstructed from the tissue-wide analysis.

Information theory framework for defining tissue metabolic diversity and specialization

Tissue metabolic diversity was calculated using Shannon entropy of idMS/MS tissue-level frequency distribution by the following equation as described in Martinez et al., (2008) (1):

$$Hj = -\sum_{i=1}^{m} P_{ij} \log_2(P_{ij})$$

where P_{ij} correspond to relative frequency of the ith idMS/MS (i = 1, 2, …, m) in the jth tissue (j = 1, 2, …, t).

The average frequency of the ith idMS/MS among tissues was calculated as:

$$P_i = \frac{1}{t}\sum_{j=1}^{t} P_{ij}$$

idMS/MS specificity was calculated as:

$$S_i = \frac{1}{t}\left(\sum_{j=1}^{t} \frac{P_{ij}}{P_i}\log_2 \frac{P_{ij}}{P_i}\right)$$

The tissue specialization δj index was measured for each jth tissue, the average of the idMS/MS specificities using the following formula:

$$\delta_j = \sum_{i=1}^{m} P_{ij}\, S_i$$

idMS/MS molecular networking by bi-clustering

To perform this clustering, we used the R package DiffCoEx which is based an extension of the Weighted Gene Coexpression Analysis (WGCNA). Using NDP and NL-scoring matrices for 895 idMS/MS spectra, we computed a comparative correlation matrix using DiffCoEx with the parameters of “cutreeDynamic” set to method="hybrid", cutHeight = 0.999, deepSplit = T, minClusterSize = 10. The R source code of DiffCoEx is downloaded from additional file 1 in Tesson et al. (2010) (2), the required R WGCNA package can be found at http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA.

Gene-to-metabolite tissue-association similarities

Cross-tissue gene-to-metabolite associations provide valuable clues in formulating functional hypothesis about metabolic genes. To do so, we used as data input the idMS/MS and RNAseq binary data-sets computed separately for the following Z-scores: a Z-score of 1 indicating tissue-specificity and 0 indicating no tissue-specificity for a given feature. Similarities in gene and metabolite tissue-specificity were calculated using Ochiai coefficient calculated as following:

$$Ochiai = \frac{a}{\sqrt{(a+b)}\sqrt{(a+c)}}$$

where a is the number of tissue-associations for which both a metabolite and a gene exhibit a Z-score of 1, b is the number of tissue-associations where a metabolite is 1 and the gene is 0, c is the number of tissue-associations where the metabolite is 0 and the gene is 1.

Virus-induced gene silencing (VIGS)

Vector construction, plant growth, and inoculation conditions were as described by Saedler and Baldwin (2004) (3). Briefly, 200- to 300-bp fragments of *N. attenuata* target genes were amplified by PCR using specific primer pairs as listed in ***SI Appendix,* Table S1**. Amplified fragments were cloned into pTV00 vector, and plasmids were transformed by electroporation into *Agrobacterium tumefaciens* strain GV3101. A pTV00 plasmid without insert (EV) was used as a negative control in all experiments. Three leaves of 24- to 25-d-old *N. attenuata* plants were infiltrated with a 1:1 mixture of *A. tumefaciens* transformed with pBINTRA and one of the gene fragment-containing construct or the pTV00 construct. *Phytoene desaturase* (pTVPDS) causing bleaching of tobacco leaves due to the depletion of carotenoids was used as a positive control to monitor the progression of VIGS in a separate set of inoculated plants. VIGS-silenced plants were used for treatment after PDS-VIGS leaves developed a strong bleaching phenotype. Silencing efficiency was verified by RT-qPCR of target gene transcripts after RNA extraction and cDNA synthesis.

RT-qPCR analysis of gene silencing efficiency

Total RNA was extracted by adding Trizol reagent (Invitrogen; http://www.invitrogen.com) to approximately 150 mg of powdered leaf material ground in liquid nitrogen following the manufacturer's protocol. A total of 500 ng of DNA-free RNA samples was reverse transcribed using oligo(dT)18 primers and SuperScript II enzyme (Invitrogen) following the manufacturer's recommendations. All RT-qPCR assays were performed with a Stratagene MX3005P instrument (http://www.stratagene.com) as recommended by the manufacturer. To normalize transcript levels, primers specific for the elongation factor-1α gene from *Nicotiana tabacum* (EF1-α; accession no. D63396) were used. Specific primers in the 5′ to 3′ direction used for SYBR Green-based analyses are listed in ***SI Appendix,* Table S1**.

Accession numbers

For construction of the UDP-Glycosyltransferase tree in different species, we used the following Genbank accessions for *Allium cepa* (UGT73G1, AAP88406.1; UGT73J1, AAP88407.1), *Antirrhinum majus* (AmC4GT, BAE48239; UGT73E2 (Amugt36), BAG16513.1;

UGT73N1 (Amugt38), BAG16514.1; UGT88D4, BAG31945), *Arabidopsis thaliana* (AtF3G7GT, Q9ZQ95; AtF3GT, AAM91339; AtF5GT, AAM91686; AtF7GT, AAL90934; AtUGT89C1, AAP31923; DOGT1, NP_181218; UGT75D1, AAB58497.1), *Aralia cordata* (AcGAT, BAD06514), *Avena sativa* (AvUGT80A1, CAB06081), *Bellis perennis* (UGT94B1 (BpUGAT), BAD77944), *Beta vulgaris* (BvUGT71F1, AAS94330; BvUGT73A4, AAS94329.1), *Brassica napus* (UGT84A9 (BnSGT1), AF287143_1), *Catharanthus roseus* (CaUGT3, BAH80312; CaUGT1, BAD29721; CaUGT2, BAD29722; UGT85A2a (CrUGT6), BAK55749; UGT709C2 (CrUGT8),BAO01109), *Celosia cristata* (CcCDOPA5GT, BAD91804), *Citrus maxima* (CmF7G12RT, AAL06646), *Citrus sinensis* (CsUFGT, AAS00612), *Citrus unshiu* (CuLGT, BAA93039), *Crocus sativus* (CsGT45, ACM66950.1; CsUGT707B1, CCG85331; Glt2 (UGTCs2), AAP94878.1), *Dianthus caryophyllus* (DcF3GT, BAD52004; DicGT1, BAD52003; DicGT2, BAD52005; DicGT4 (DcC2GT), BAD52006; DicGT5, BAD52007), *Dorotheanthus bellidiformis* (DbB5GT, CAB56231; DbB6GT, AAL57240), *Forsythia x intermedia* (FiF3GT, AAD21086), *Fragaria x ananassa* (FaFGT, AAU12367; FaGT2, AAU09443), *Gentiana triflora* (Gt5GT7, BAG32255; GtF3GT, BAA12737; GtGTX, BAC54092), *Glycine max* (GmF3G6R, BAN91401; GmIF7GT, BAF64416), *Glycyrrhiza echinata* (GeIF7GT, BAC78438), *Hordeum vulgare subsp. vulgar* (HvF3GT, CAA33729), *Ipomoea nil* (In3GGT(InA32GT), BAD95885; InGTase1, BAF75917), *Ipomoea purpurea* (Ip3GGT(IpA32GT), BAD95882), *Iris x hollandica* (Ih3GT, BAD83701; Ih5GT, BAD06874), *Lamium galeobdolon* (LgF3GT, AEB61487), *Linaria vulgaris* (LvC4GT, BAE48240), *Lycium barbarum* (Ugt73a10, BAG80536), *Maclura pomifera* (MpUGT75L4, ABL85474; MpUGT88A4, ABL85471), *Medicago trunculata* (MtUGT73C8, ABI94020; MtUGT73K1, AAW56091; MtUGT73P1, ABI94026; MtUGT78G1, ABI94025; MtUGT84F1, ABI94023; MtUGT85H2, ABI94024.1; MtUGT88E1, ABI94021; MtUGT88E2, ABI94025; UGT71G1, AAW56092), *Mirabilis jalapa* (CDOPA5GT, BAD91804), *Nicotiana tabacum* (NTGT1A, BAB60720.1; NTGT1b, BAB60721.1; NtGT2, BAB88935; NtGT3, BAB88934; NtSAGT, AAF61647; TOGT 1, AAK28303; TOGT 2, AAK28304), *Perilla frutescens* (PfA5GT, BAA36421; PfF3GT, BAA19659; PfUGT88D7 (F7GAT), BAG31948), *Petunia x hybrida* (PhA3ART, CAA50376; PhA3GT, BAA89008; PhA5GT, BAA89009; PhF3GalTase, AF165148_1), *Phaseolus lunatus* (PlZOG1, AAD04166); *Phaseolus vulgaris* (PvZOX1, AF116858_1), *Phytolacca americana* (PaGT2, BAG71125; PaGT3, BAG71127), *Pilosella officinarum* (PoUGT95A1, ACB56927), *Prunus dulcis* (PdUGT85A19, ABV68925), *Pyrus communis* (PcF7GT, AAY27090), *Quercus robur* (QrUGT84A13, AHA54051), *Rauwolfia serpentine* (RsAS, CAC35167), *Rhodiola sachalinensis* (RsUGT73B6, AAS55083; RsUGT74R1, ABP49574; RsUGT72B14, ACD87062), *Rosa hybrida* (RhA53GT, BAD99560), *Scutellaria baicalensis* (SbB7GAT, BAD99560; SbF7GT, BAA83484), *Scutellaria laeteviolacea var. yakusimensis* (SlUGT88D5, BAG31946), *Sesamum indicum* (SiUGT88D6, BAG31947),

Solanum aculeatissimum (SaGT4A, BAD89042.1), Solanum berthaultii (SbGT, AAB62270.1), Solanum lycopersicum (Gtsatom, CAI62049.1), *Solanum melongena* (SmUGT76, CAA54558.1), *Solanum tuberosum* (Sgt2.1, ABB29873.1; Sgt2.2, ABB29874.1; StSgt1, AAB48444.1; StSgt3, ABB84472.1), *Stevia rebaudiana* (SrUGT74G1, AY345982; SrUGT76G1, AY345974; SrUGT85C2, AY345978), *Torenia hybrida* (ThA5GT, BAC54093), *Verbena hybrida* (VhA5GT, BAA36423), *Vigna angularis* (VaABAGT, BAB83692), Vigna mungo (VmUF3GaT, BAA36972; VmUFGlyT, BAA36410), *Vitis labrusca* (VlGT, ABR24135), *Vitis vinifera* (VvGT1, AAB81682), *Withania somnifera* (WsFGT, FJ560880; WsUGT73A16, FJ654696/ACO44747.1), *Zea mays* (Zm-BX8, AF331854_1; Zm-BX9, CAX02221; ZmcisZog1, AAK53551; ZmcisZog2, AAL92460; ZmIaglu, AAA59054; ZmUFGT, CAA30760; ZmUGT71A1, CAA31856).

Tables

Table S1. List of primers used for qRT-PCR and gene fragment cloning for VIGS

Name	For VIGS cloning (5' to 3')	For qRT-PCR (5' to 3')
UGT-A forward	GCGGCGCTCGAGGTTGAGCATTATACTAAGGTGC	GACGCTAGAAGGAGTTTCAGG
UGT-A reverse	GCGGCGGGATCCCAGGCAACCAATCTTCGTTGTC	CCACTGAATCGAACCAACAC
UGT-B forward	GCGGCGCTCGAGGTGGTCCTACTGTATATGACCG	GGGAATTATTCATTCAGGTTGGG
UGT-B reverse	GCGGCGGGATCCGGTAGCCCAGTTTGCTCCAGAC	GGCAACATAACCACTTGACAG
Elongation factor forward		TGGTATGGTTAAGATGCTTCCC
Elongation factor reverse		TGTCAACGCTCTTGATAACAC

Table S2. Overview of RNAseq transcriptome data

NCBI accession number	Tissue type	Treatment/development stage	Additional note on sampling procedure
NA1498ROT	Root (ROT)	Rosette stage plants, treated with 5 µL 1:1 diluted *M. sexta* oral secretion three times in leaves	Roots of rosette stage plants that were treated three times on leaves were collected for RNA extraction. The treatments were performed at 10 am and 6pm on the day before sampling and 10 am on the day of sampling. Samples were collected at 11 am.
NA1500LET	Leaf (LET)	Rosette stage plants, treated with 5 µL 1:1 diluted *M. sexta* oral secretion three times in leaves	Local leaves of rosette stage plants that were treated three times on leaves were collected for RNA extraction. The treatments were performed at10 am, 6pm on the day before sampling and 10 am on the day of sampling. Samples were collected at 11 am.
NA1717LEC	Leaf (LEC)	Rosette stage plants, no treatment	Rosette stage leaves were collected for RNA extraction. Samples were collected at 11 am.

NA1504STT	Stem (STT)	Rosette stage plants, treated with 5 µL 1:1 diluted *M. sexta* oral secretion three times in leaves	Stems of rosette stage plants that were treated three times on leaves were collected for RNA extraction. The treatments were performed at 10 am, 6pm on the day before sampling and 10 am on the day of sampling. Samples were collected at 11 am.
NA1505COE	Corolla (COE)	Early developmental stage, no treatment	Samples were collected in the afternoon, 60 samples were pooled.
NA1515COL	Corolla (COL)	Late developmental stage, no treatment	Samples were collected at 6 pm (open flowers) and 9am (closed flower after opening), 4-10 samples were pooled.
NA1506STI	Stigma (STI)	Mature stigma, no treatment	Stigma samples were collected in the afternoon, 40 samples were pooled.
NA1507POL	Pollen tube (POL)	No treatment	Pollen tubes were pooled.
NA1508SNP	Style (SNP)	Mature style without pollination	Styles were collected at 7 am, anthers were removed one day before, and 50 samples were pooled.
NA1509STO	Style (STO)	Mature style, pollinated with pollens from different genotype	Styles were collected at two hours after pollination, at 7 am. Anthers were removed one day before, and 30 samples were pooled.
NA1510STS	Style (STS)	Mature style, self-pollinated	Styles were collected at two hours after pollination, at 7 am. Anthers were removed one day before, and 30 samples were pooled.
NA1511NEC	Nectary (NEC)	Mature nectary, no treatment	Samples were collected in the afternoon, 60 samples were pooled.
NA1512ANT	Anther (ANT)	Mature anther no treatment	Samples were collected in the afternoon, 60 samples were pooled.
NA1513OVA	Ovary (OVA)	Mature ovary, no treatment	Samples were collected in the afternoon, 60 samples were pooled.
NA1514PED	Pedicel (PED)	Mature pedicel, no treatment	Samples collected at 9 am (heading down) and 4 pm (heading up) were pooled.
NA1516OFL	Flower (OFL)	Fully opened flowers, no treatment	Both morning (7 am) and evening (6 pm) flowers were collected, 1 sample of each were pooled.
NA1517FLB	Flower bud (FLB)	Two early developmental stages of flowers, no treatment	Samples were collected at 6pm, 1 bud and 1 middle stage flower were collected. Sepals were removed from the samples.
NA1501SES	Seed (SES)	Treated with liquid smoke	100 mg seeds treated with 1:50 diluted liquid smoke solution for 9-15 min were used for RNA extraction.
NA1502SEW	Seed (SEW)	Treated with water	100 mg seeds treated with water for 9-15 min were used for total RNA extraction.
NA1503SED	Seed (SED)	Dry seeds	100 mg dried seeds directly used for total RNA extraction.

Table S3. Performance of the Ochiai distance-based integration of genes and idMS/MS data for known gene-metabolite associations

Gene name	Function	Reference		Metabolites	PCC		Ochiai	
					coeff.	*P*-value	coeff.	*P*-value
		1,2		N-caffeoylputrescine (mz 251.14 @127s)	0.00	1.0	0.63	0.0
MYB8	transcription factor	1,2		N-feruloylputrescine (mz 265.15 @280s)	-0.21	0.5	0.35	0.3
		1,2		N-Caffeoylspermidine (mz 308.20 @418s)	-0.01	1.0	0.45	0.1
		1,2		N-caffeoylputrescine (mz 251.14 @127s)	0.07	0.8	0.71	0.0
AT1	hydroxycinnamoyl-CoA: putrescine transferase	1,2		N-feruloylputrescine (mz 265.15 @280s)	-0.14	0.7	0.47	0.1
		1,2		Unidentified putrescine-based phenolamide (mz 350.21 @221s)	-0.17	0.6	0.57	0.1
		1,2		N-Caffeoylspermidine (mz 308.20 @418s)	0.16	0.6	0.68	0.0
		1,2		N',N''-Di-caffeoylspermidine (mz 470.23 @396s)	0.55	0.1	0.72	0.0
DH29	hydroxycinnamoyl-CoA: spermidine transferase	1,2		N,N'-Coumaroyl,caffeoylspermidine (mz 454.23 @480s)	-0.03	0.9	0.72	0.0
		1,2		N',N''-Caffeoyl,feruloylspermidine (mz 484.24 @506s)	0.03	0.9	0.72	0.0
		1,2		N',N''-Di-feruloyl-spermidine (mz 498.26 @591s)	0.23	0.5	0.76	0.0
		1,2		Unidentified spermidine-based phenolamide (mz 468.21 @228s)	0.12	0.7	0.71	0.0
		1,2		N',N''-Di-caffeoylspermidine (mz 470.23 @396s)	0.44	0.2	0.63	0.0
CV86	hydroxycinnamoyl CoA:hydroxycinnamoylspermidine-conjugating activity	1,2		N,N'-Coumaroyl,caffeoylspermidine (mz 454.23 @480s)	0.43	0.2	0.63	0.0
		1,2		N',N''-Caffeoyl,feruloylspermidine (mz 484.24 @506s)	0.28	0.4	0.63	0.0
		1,2		N',N''-Di-feruloyl-spermidine (mz 498.26 @591s)	0.45	0.1	0.67	0.0
PAL	Phenylalanine ammonia lyase	1,2	Other phenylpropanoids	Phenylalanine (mz 166.09 @132s)	-0.17	0.6	0.35	0.3
THT	Tyramine N-hydroxycinnamoyltransferase	3		N-coumaroyltyramine (mz 284.10 @95s)	-0.01	1.0	0.71	0.0
HQT	hydroxycinnamoyl CoA quinate transferase	4		Chlorogenic acid (mz 355.10 @285s)	0.36	0.2	0.44	0.2
		This study		Rutin (mz 611.16 @513s)	0.75	0.0	0.52	0.1
		This study		Quercetin-3-O-sophoroside (mz 627.16 @430s)	-0.13	0.7	0.26	0.4
		This study		Quercetin-3-O-glucose (mz 465.10 @546s)	0.09	0.8	0.71	0.0
UGT-A	Flavonoid UDP-Glucosyltransferase	This study		Kaempferol-3-O-sophoroside (mz 611.16 @486s)	0.01	1.0	0.26	0.4
		This study		Kaempferol-3-O-glucose-rhamnose (mz 595.17 @580s)	0.07	0.8	0.52	0.1
		This study		Kaempferol-3-O-glucose (mz 449.11 @616s)	0.08	0.8	0.52	0.1
		This study		Kaempferol-3-O-rutinoside-glucoside (mz 757.22 @342s)	-0.16	0.6	0.58	0.0
		This study		Rutin (mz 633.14 @513s)	0.46	0.1	0.71	0.0
		This study		Quercetin-3-O-sophoroside (mz 627.16 @430s)	-0.44	0.1	0.45	0.1
UGT-B	Flavonoid UDP-Rhamnosyltransferase	This study		Quercetin-3-O-glucose (mz 465.10 @546s)	-0.22	0.5	0.61	0.0
		This study		Kaempferol-3-O-sophoroside (mz 611.16 @486s)	-0.38	0.2	0.45	0.1
		This study		Kaempferol-3-O-glucose-rhamnose (mz 595.17 @580s)	-0.27	0.4	0.45	0.1
		This study		Kaempferol-3-O-glucose (mz 449.11 @616s)	-0.26	0.4	0.45	0.1
		5,6		Lyciumoside I (mz 653.35 @1099s)	0.91	0.0	0.75	0.0
GGPPS	geranylgeranyl pyrophosphate synthase	5,6		Lyciumoside IV (mz 799. 41 @1087s)	0.52	0.1	0.53	0.1
		5,6		Lyciumoside II (mz 815.10 @999s)	0.63	0.0	0.76	0.0
		7		Lyciumoside I (mz 653.35 @1099s)	0.89	0.0	0.89	0.0
GLS	Geranyllinalool Synthase	7		Lyciumoside IV (mz 799. 41 @1087s)	0.51	0.1	0.79	0.0
		7		Lyciumoside II (mz 815.10 @999s)	0.55	0.1	0.85	0.0
		in preparation		Lyciumoside IV (mz 799. 41 @1087s)	0.50	0.1	0.72	0.0
UGT74P3	HGL-DTG UDP-Glycosyltransferase	in preparation		Attenoside (mz 939.48 @990s)	0.07	0.8	0.72	0.0
		in preparation		Nicotianoside III (mz 940.49 @992s)	-0.01	1.0	0.72	0.0
		in preparation		Lyciumoside II (mz 815.10 @999s)	0.58	0.0	0.77	0.0
		in preparation		Lyciumoside IV (mz 799. 41 @1087s)	0.52	0.1	0.79	0.0
UGT91T1	HGL-DTG UDP-Glycosyltransferase	in preparation		Attenoside (mz 939.48 @990s)	0.25	0.4	0.79	0.0
		in preparation		Nicotianoside III (mz 940.49 @992s)	0.15	0.7	0.79	0.0
		in preparation		Lyciumoside II (mz 815.10 @999s)	0.63	0.0	0.68	0.0
		8		O-AS#9 Class II (mz 631.29 @1471s)	-0.22	0.5	0.63	0.0
		8		O-AS#2 Class III (mz 687.32 @1771s)	-0.22	0.5	0.89	0.0
ASAT1	acylsucrose acyltransferase	8		O-AS#8 Class III (mz 654.33 @1550s)	-0.13	0.7	0.77	0.0
		8		O-AS#19 Class IV (mz 715.315 @1856s)	-0.20	0.5	0.77	0.0
		8		O-AS#45 Class IV (mz 701.30 @1738s)	-0.18	0.6	0.63	0.0
		8		O-AS#9 Class II (mz 631.29 @1471s)	0.77	0.0	0.53	0.1
ASAT2	acylsucrose acyltransferase	8		O-AS#2 Class III (mz 687.32 @1771s)	0.92	0.0	0.76	0.0
		8		O-AS#8 Class III (mz 654.33 @1550s)	0.55	0.1	0.65	0.0
		8		O-AS#4 Class IV (mz 679.35 @1892s)	0.81	0.0	0.85	0.0
ASH1	acylsugar acylhydrolase	9		O-AS#2 Class III (mz 687.32 @1771s)	0.45	0.1	0.50	0.1
		9		O-AS#4 Class IV (mz 679.35 @1892s)	0.40	0.2	0.67	0.0
		9		O-AS#9 Class II (mz 631.29 @1471s)	-0.05	0.9	0.35	0.3
ASH2	acylsugar acylhydrolase	9		O-AS#2 Class III (mz 687.32 @1771s)	0.06	0.9	0.50	0.1
		9		O-AS#4 Class IV (mz 679.35 @1892s)	-0.07	0.8	0.45	0.1

References for Table S3: 1, Onkokesung et al. Plant Physiology 158.1 (2012): 389-407; **2,** Jillian et al. Planta 221.6 (2005): 904-914; **3,** Gaquerel et al. Journal of Agricultural and Food Chemistry 58.17 (2010): 9418-9427; **4,** Ricarda et al. Nature Biotechnology 22.6 (2004): 746-754; **5,** Heiling et al. The Plant Cell 22.1 (2010): 273-292; **6,** Heiling, et al. The Plant Journal 85.4 (2016): 561-577; **7,** Falara, et al. Plant Physiology 166.1 (2014): 428-441; **8,** Fan, et al. Proc. Natl. Acad. Sci. U. S. A. 113.2 (2015): E239-E248; **9,** Schilmiller, et al. Plant Physiology 170.3 (2016): 1331-1344

Datasets

Dataset S1. idMS/MS data analysis.

Spreadsheet S1.1, XCMS-processed data-set; **Spreadsheet S1.2,** cross-tissue ZMAD scaled data and results of the Kurtosis analysis; **Spreadsheet S1.3,** deconvoluted idMS/MS spectra; **Spreadsheet S1.4,** structural and tissue annotation of the idMS/MS bi-clustering.

Figures

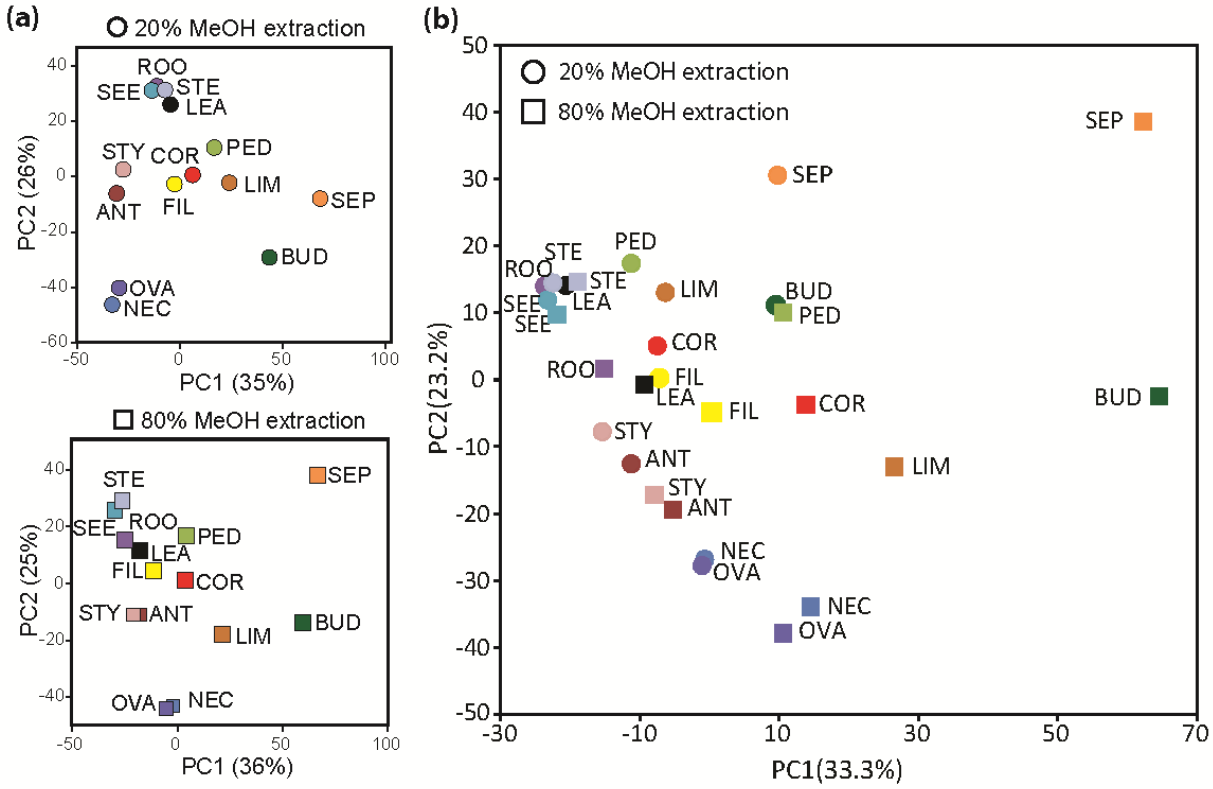

Figure S1. Principal component analysis (PCA) of the UHPLC-ESI/qTOF-MS metabolic profiles for all tissue and extraction types. The PCA score plot was conducted on the auto-scaled complete mass feature matrix resulting from XCMS processing of the samples extracted by 20% and 80% methanol (vol/vol) which are represented as circle and rectangular shapes respectively. Detailed explanations of the tissue collection procedure are provided in the Method section. **(a)** Separate PCA score plot of 20% and 80% methanolic (vol/vol) extracts. **(b)** PCA of the combined data-set.

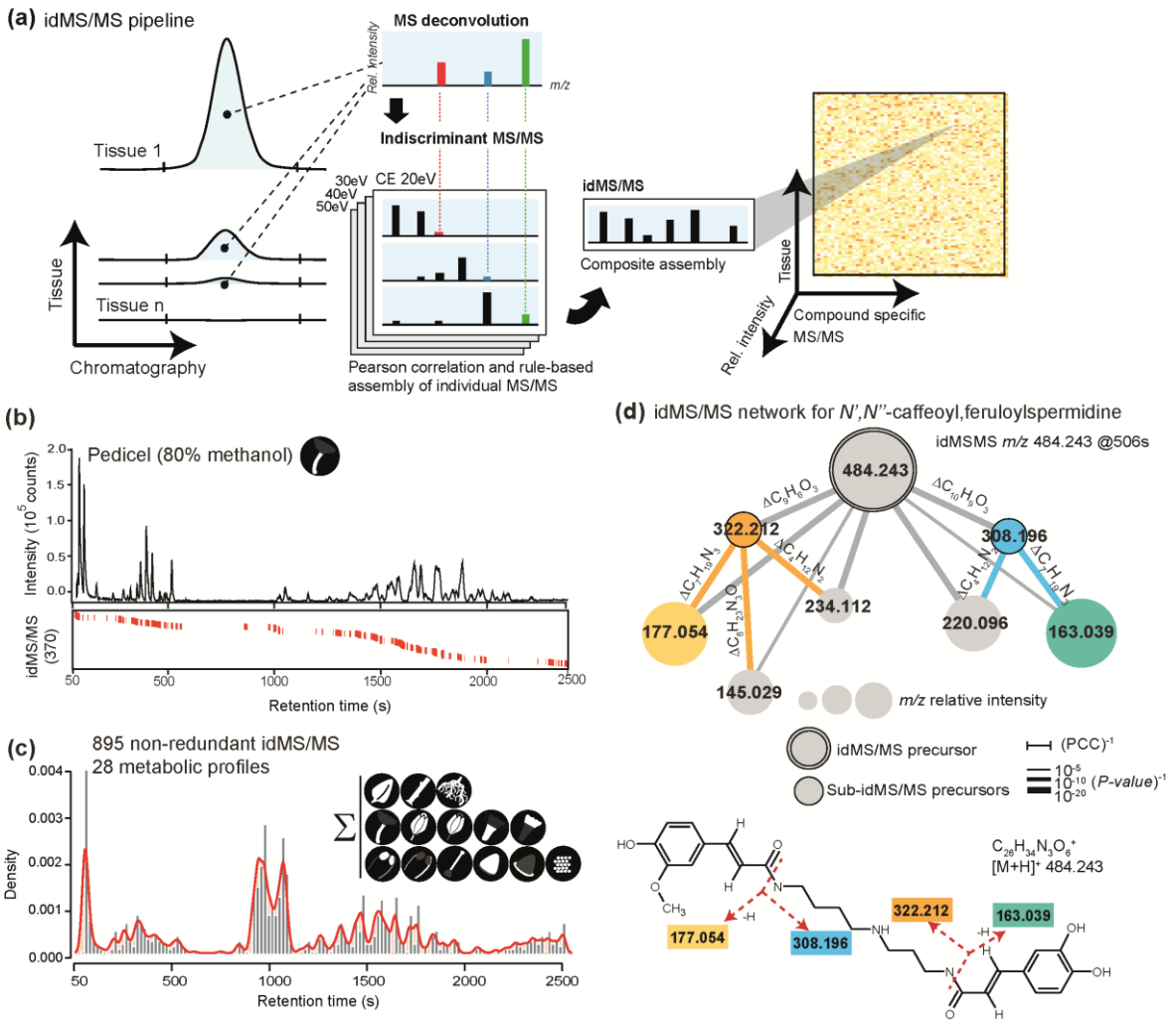

Figure S2. Tissue-wide indiscriminant acquisition and assembly of metabolite MS/MS spectral information. (a) Optimized pipeline to achieve indiscriminant acquisition of MS/MS data from tissue-wide metabolic variations. Indiscriminant MS/MS targets every mass signal within a range of 50 to 1400 *m/z* for fragmentation using 4 increasing collision energies (CE). Each tissue sample is processed using idMS/MS via individual analyses performed at different CID voltages in order to maximize fragment diversity. Information regarding a fragment's assignment to a given precursor mass is lost during indiscriminant MS/MS (idMS/MS) but can be computationally-retrieved for each CID voltage run, based on mathematical and chromatographic correlation analyses (see Method section). The pipeline harnesses the important chemical diversity among tissue samples. Reciprocally, cross-tissue quantitative variation provides the statistical power required for computing high confidence Pearson correlation coefficients (PCC) among the variation in intensities of precursors and candidate fragments. idMS/MSs assembled at each CID voltage for a given precursor *m/z* occurring at a

given retention time are then merged together into a composite MS/MS which displays the complete fragment diversity. The resulting output is a three-dimensional entry matrix with non-redundant idMS/MS spectra across tissues and their relative intensities. **(b)** idMS/MS coverage for a representative 80% methanol (vol/vol) pedicel sample. The lower heatmap displays the retention time position, aligned to the total ion current chromatogram, of the 370 idMS/MS spectra recorded for this sample. **(c)** Density plot summarizing the idMS/MS coverage across all 14 analyzed tissues denoted by their symbols. Bars represent the relative density of collected idMS/MSs for a 9 s retention time window. Red lines represent smoothed density curves. **(d)** Example of the idMS/MS obtained for *m/z* 484.243 @ 506 s corresponds to the $[M+H]^+$ of an *N',N''*-caffeoyl,feruloylspermidine isomer – ' and '' denote that the exact position of the caffeoyl and feruloyl moieties on the spermidine backbone cannot be assigned by MS analysis alone. idMS/MSs were also assembled for *m/z* 322.212 and *m/z* 308.196 as these *m/z* signals are already present as in-source ionization derived fragments in the MS mode analysis.

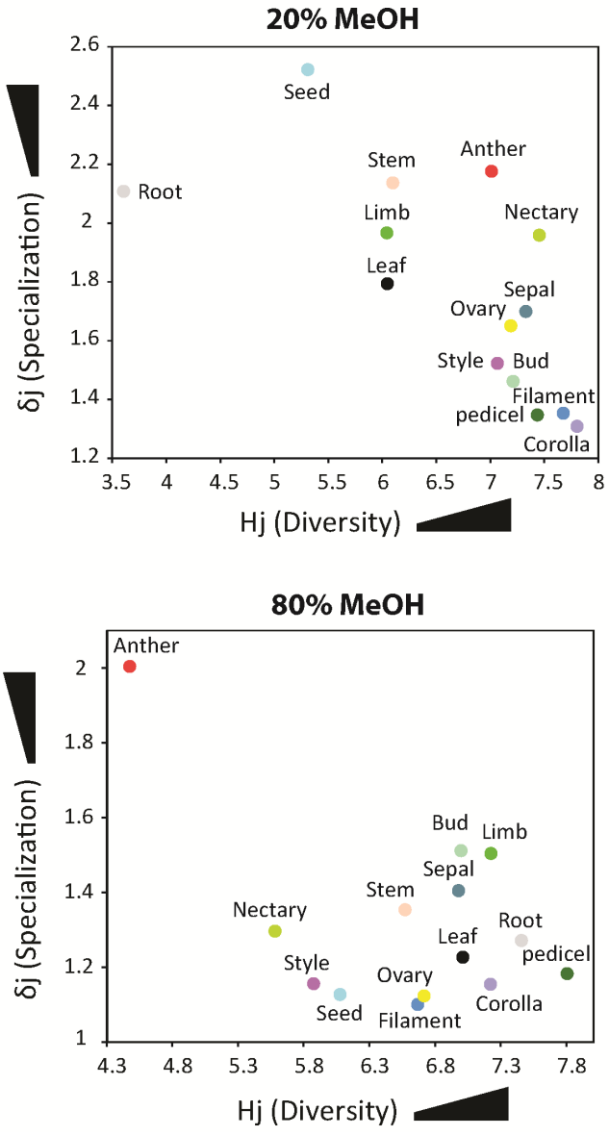

Figure S3. Information theory-based analysis of the degree of specialization and diversity in the idMS/MS composition across tissues. Tissue specialization (δj) and diversity (Hj) are mapped in a two-dimensional space using two indexes, where Hj, tissue-level idMS/MS diversity is calculated by Shannon entropy of idMS/MS frequency distribution of each tissue and δj, tissue-level idMS/MS specialization is measured by the average specificity of each idMS/MS component by taking into consideration its frequency among tissues. Tissue metabolomes extracted for two different extraction conditions (20% or 80% methanol (vol/vol)) were separately analyzed and visualized into two panels and tissue types are differentiated by different colors.

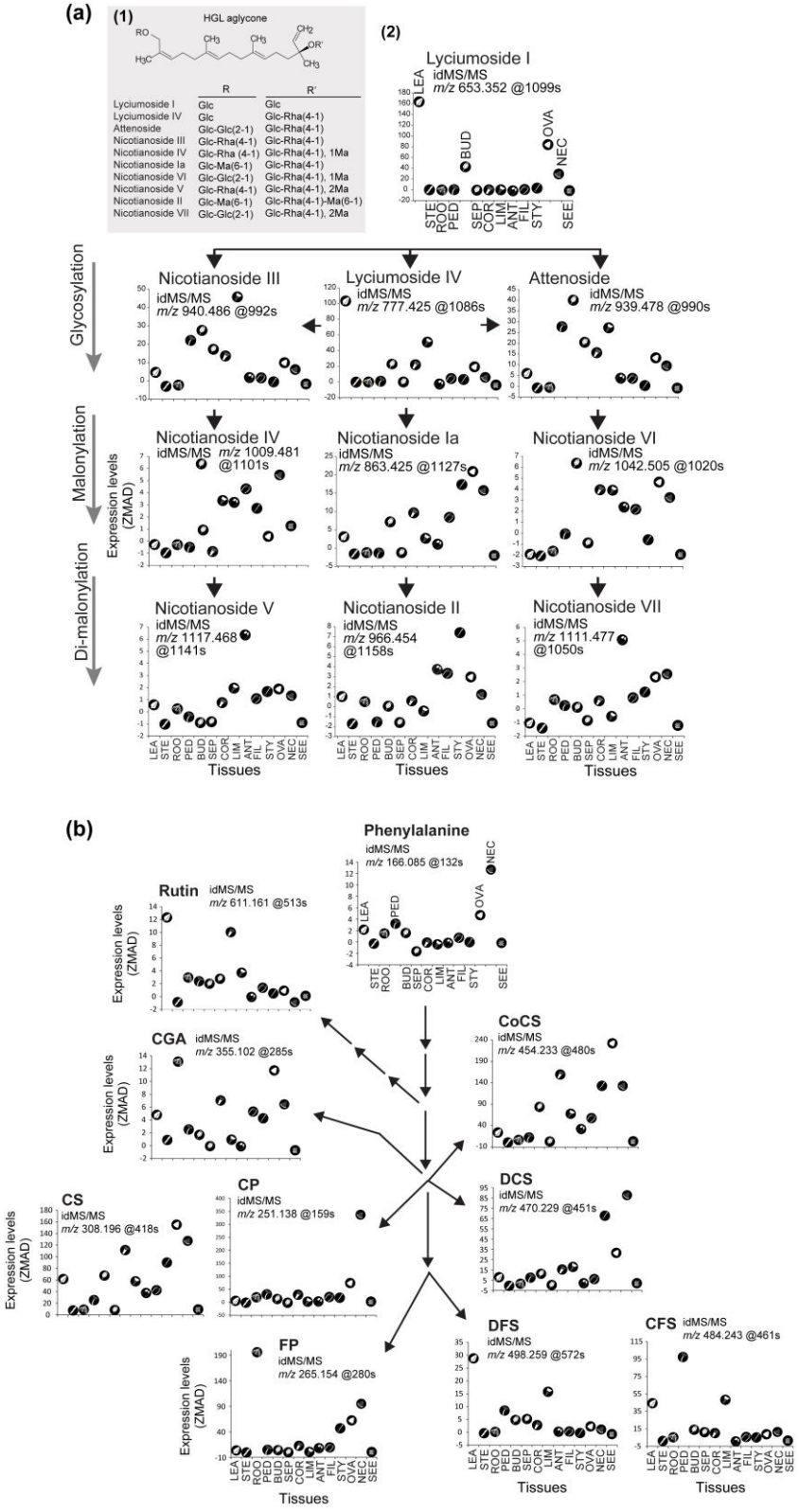

Figure S4. Tissue-based variations in the accumulation of HGL-DTGs and phenolic derivatives. 17-hydroxygeranyllinalool (HGL) diterpene glycosides are abundant secondary

metabolites in *N. attenuata* which differ in the number and types of sugar (Glc, glucose; Rha, Rhamnose; with or without malonyl, Ma, groups) decorations added to the acyclic HGL backbone which is characteristic of this compound family **(a.1)**. Z-score-normalized median absolute distances are employed to visualize cross-tissue variations in the idMS/MSs corresponding to the main **(a.2)** HGL-DTG intermediates as well as **(b)** major phenylpropanoid-quinate and -polyamine conjugates. For both metabolic groups, important changes in cross-tissue variations are detected. In the case of the HGL-DTG metabolic pathway, a progressive enrichment of certain metabolites is noticeable in reproductive floral tissues. CFS, *N',N''*-caffeoyl,feruloyl-spermidine; CGA, Chlorogenic acid; CoCS, *N',N''*-coumaraoyl,caffeoyl-spermidine; CP, *N*-caffeoylputrescine; CS, *N*-caffeoylspermine; FP, *N*-feruloylputrescine; DCS, *N', N''*-dicaffeoylspermidine; DFS, *N',N''*-diferuloyl-spermidine.

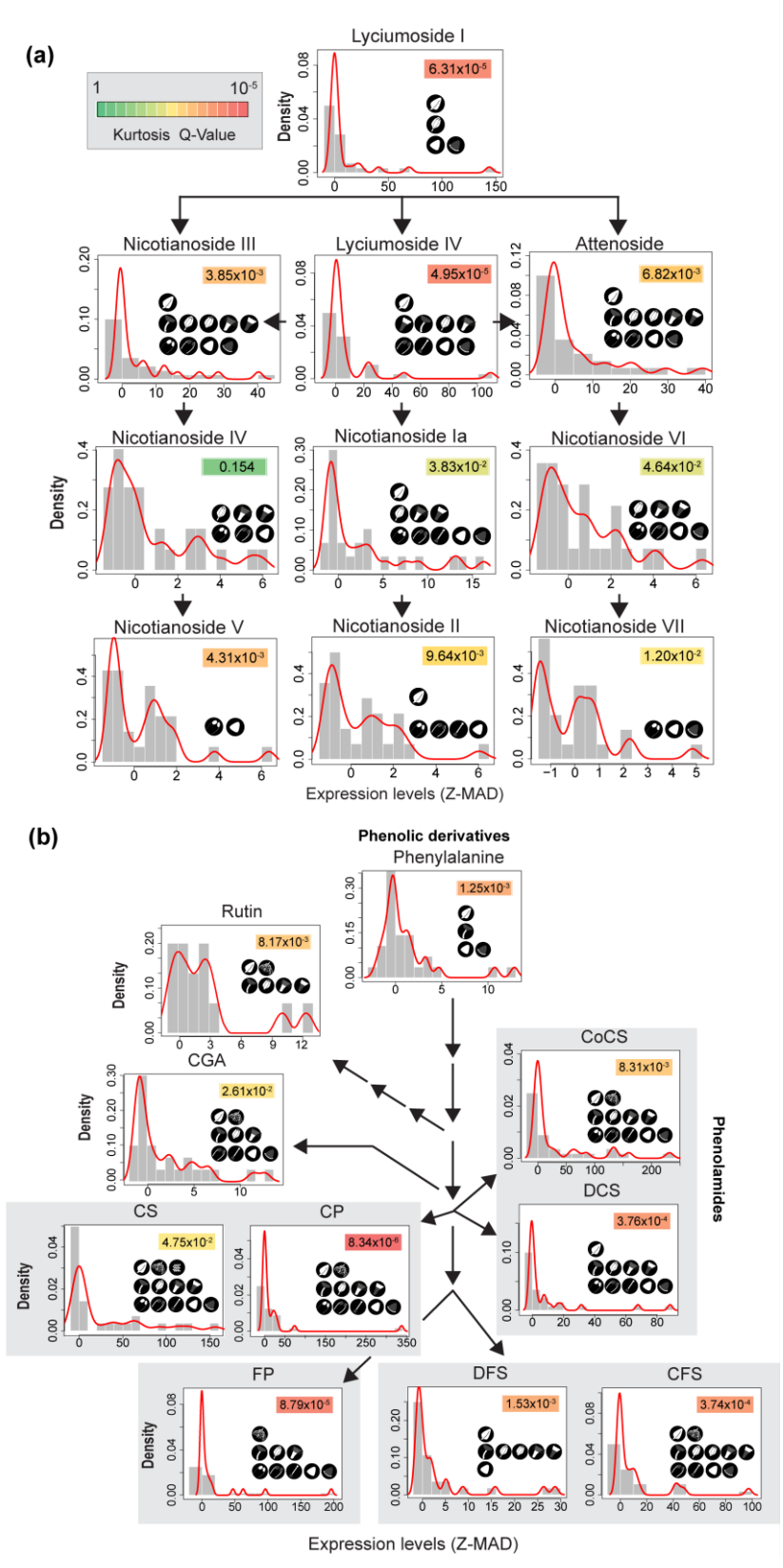

Figure S5. Significant levels of tissue-specificity derived from Kurtosis filtering for particular metabolic steps in the HGL-DTG and phenolic pathway. Q-values present in color

boxes derived from Kurtosis filtering analysis (see Method section) for the distribution of idMS/MSs corresponding to particular HGL-DTGs (17-hydrogeranyllinalool diterpene glycosides) **(a)** and phenolic derivatives **(b)**. Associated tissues for each idMS/MS are presented as tissue icons below the Q-value box. Low Q-values indicate strong tissue-specificity in the accumulation of particular HGL-DTGs and phenolic derivatives. Strongest tissue-associations are detected for upstream steps in the pathway indicating clear tissue-specificities in the biosynthesis of these upstream intermediates from which the complete HGL-DTG chemotype derives from. Complete results of the Kurtosis analysis are presented in ***SI Appendix,* Dataset S1**. CFS, *N',N''*-caffeoyl,feruloyl-spermidine; CGA, Chlorogenic acid; CoCS, *N',N''*-coumaraoyl,caffeoyl-spermidine; CP, *N*-caffeoylputrescine; CS, *N*-caffeoylspermine; FP, *N*-feruloylputrescine; DCS, *N', N''*-dicaffeoylspermidine; DFS, *N',N''*-diferuloyl-spermidine.

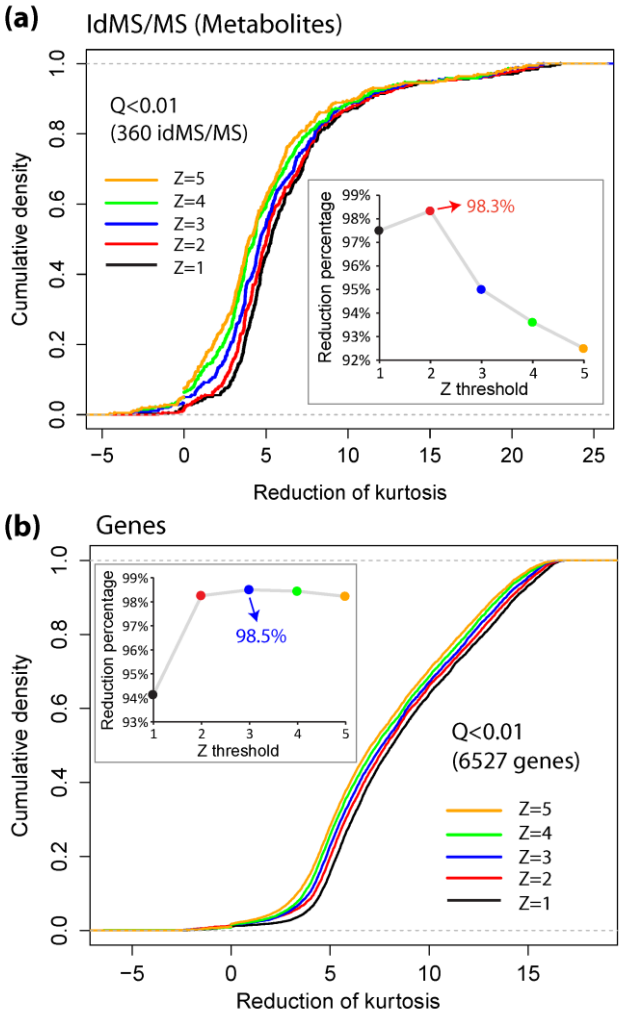

Figure S6. Implementing the reduction of Kurtosis analysis to infer idMS/MS and gene expression with high tissue-specificity. Cumulative distribution of idMS/MS **(a)** and gene **(b)** cross-tissue expression generated for q-value (Q) <0.01 obtained from R qvalue package (6). The analysis is based on the Anscombe test for kurtosis using the Anscombe.test function in the R "moments" package as described in Li et al. (7) and in the Method section. The x axis reports on the Kurtosis reduction when a certain Z threshold (ZMAD-normalized expression values) was applied, the y axis reports on the corresponding cumulative density. The insert panels correspond to the Kurtosis reduction percentage of idMS/MSs or genes when a given Z threshold was selected. Different colors denote for different Z thresholds and the corresponding cumulative curves. Z=2 was selected as the threshold to extract idMS/MSs with leptokurtic behaviors and Z=3 for genes.

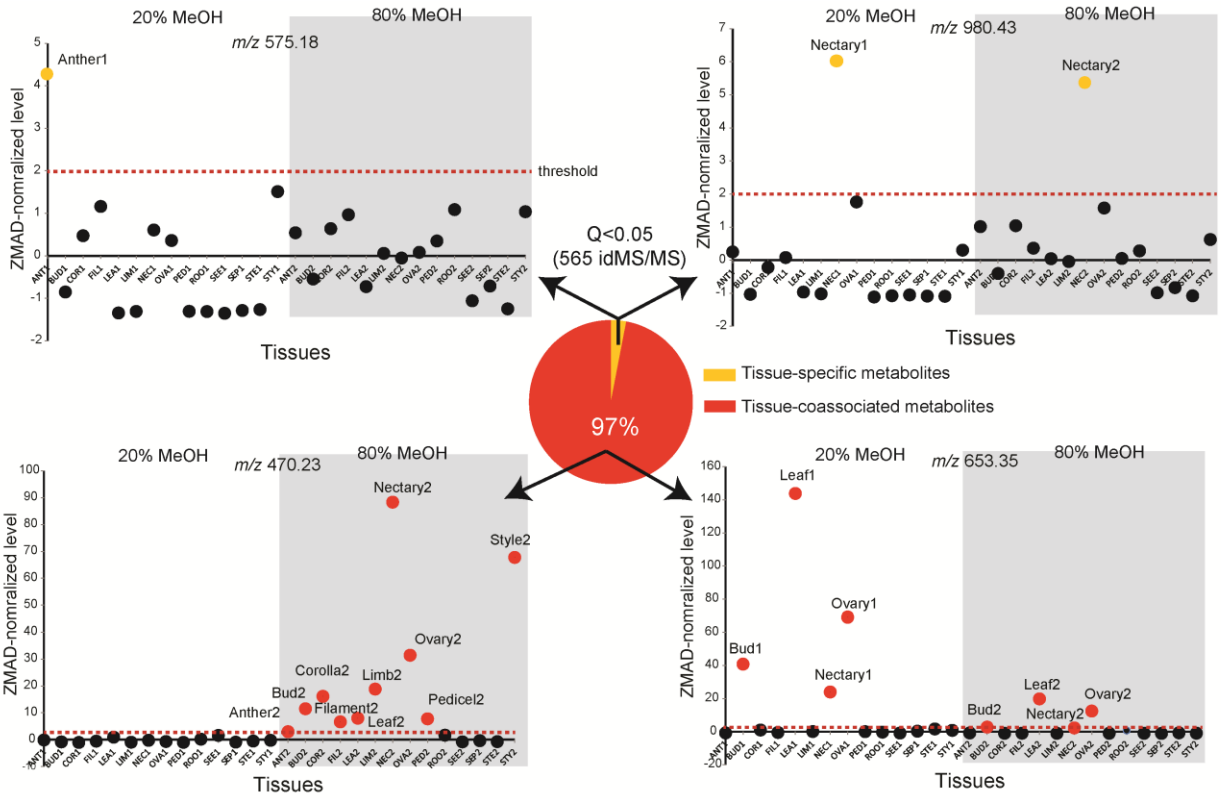

Figure S7. Different degrees of idMS/MS tissue-specificity inferred from the Kurtosis reduction analysis. The kurtosis analysis was used to discriminate leptokurtic idMS/MSs from those distributed in all tissues and the Z threshold of 2 that obtained from the reduction of Kurtosis analysis enabled the detection of single-tissue-specific and tissue-co-associated metabolites. The approach not only extracts single-tissue-specific idMS/MSs (3% of the 565 idMS/MSs with a Q<0.05 for the Kurtosis analysis; two upper panel as examples) but also those preferentially accumulating in several tissues (97% of idMS/MSs; two lower panels as examples). 1 and 2 indices after tissue names refer to the 20 % and 80 % methanol (vol/vol) extractions respectively.

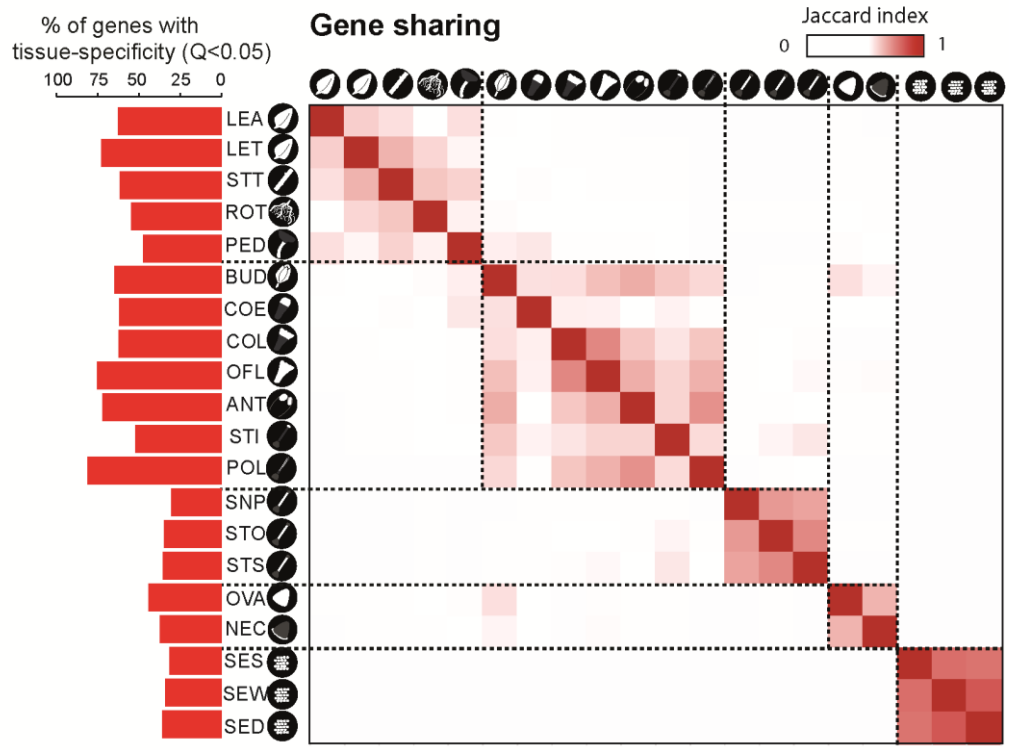

Figure S8. Gene-tissue specificity and gene-sharing among tissues of the transcriptome data-set. Left bar chart, % of genes showing tissue-specificity (Q<0.05) per tissue. Right Heatmap matrix visualizes genes sharing between tissues as measured using Jaccard index. LEA, rosette stage leaves; LET, rosette stage leaves treated with *Manduca sexta* oral secretion (OS); STT, stem from plants with leaves treated with OS; ROT, root from plants with leaves treated with OS; PED, pedicels; BUD, flower buds; COE, non-matured corolla collected 3 days after protrusion from the calyx; COL, matured corolla collected 5 days after protrusion from the corolla; OFL, open flowers; ANT, anthers; STI, stigma; POL, pollen tubes; SNP, style without pollination; STO, style outcross-pollinated (2h after pollination); STS, style self-pollinated (2h after pollination); OVA, ovary; NEC, nectary; SES, seeds treated with liquid smoke; SEW, watered seeds; SED, matured seeds. Detailed overview of RNAseq dataset is available in ***SI Appendix,* Table S2.**

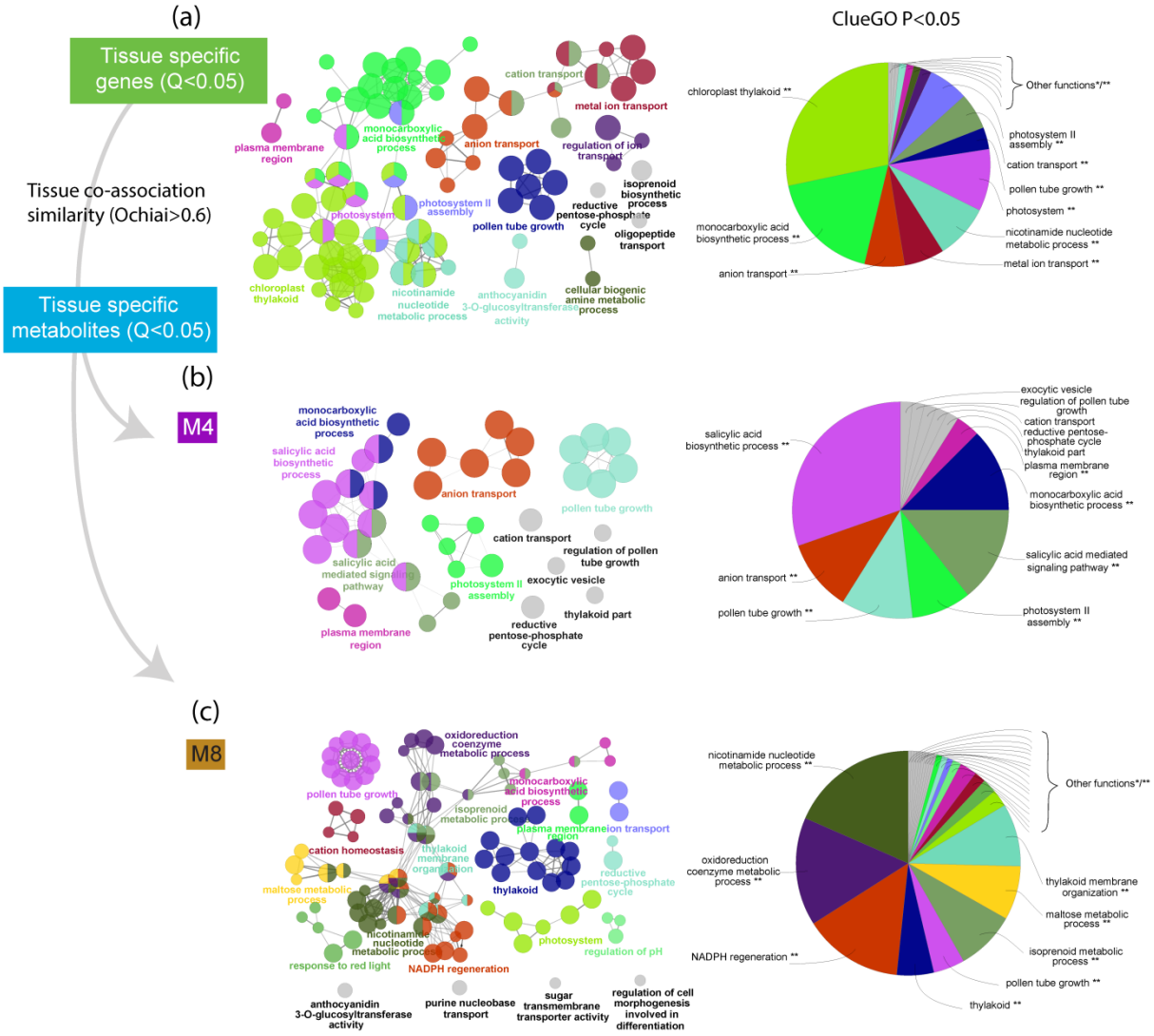

Figure S9. GO enrichment analysis for the pool of genes exhibiting tissue-specificity and co-tissue association with idMS/MSs classified as parts of modules M4 and M8. GO enrichment was generated via GlueGO (8). The significant GO enrichment was depicted in overview pie charts as well as functionally grouped networks with terms as nodes and edges linked based on their kappa score level (>0.4), where only the label of the most significant term per group is shown. Node size is in proportion to the term enrichment significance and functionally related groups partially overlap. **(a)** Significant GO enrichment obtained for all tissue-specific genes (Q<0.05) from transcriptomic data. **(b)** and **(c)** GO enrichment from transcriptomic data sharing tissue co-associations (Ochiai>0.6) with the tissue-specific metabolites (Q<0.05) in module 4 and 8.

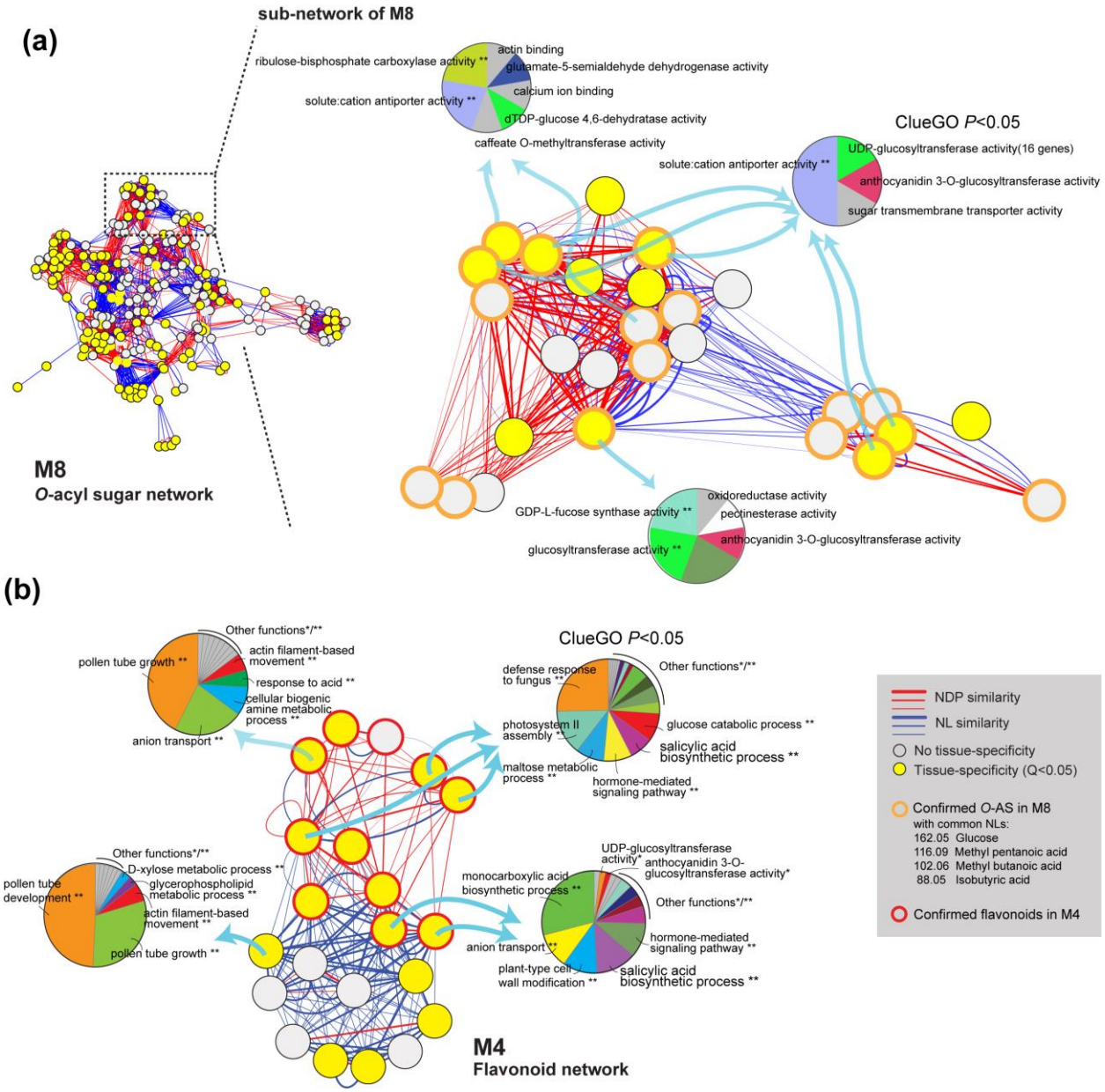

Figure S10. GO terms from transcriptomic data sharing significant tissue co-associations with targeted metabolites of modules 8 and 4. Molecular networks constructed from modules M8 **(a)** and M4 **(b)** are enriched in *O*-acyl sugars and flavonoids respectively. Nodes represent idMS/MS spectra and edges correspond to structural similarities based on the two score types (NDP similarity calculated based on shared fragments between spectra and NL similarity calculated based on shared common neutral losses between spectra). Tissue-specificity is mapped onto the molecular network with node colors. Nodes in the sub-network of module M8 that share typical *O*-acyl sugars neutral losses of glucose, methyl pentanoic acid, methyl butanoic acid and isobutyric acid are additionally circled in apricot. Identified flavonoids in module M4 are circled in red. A zoom-in of the network depicts metabolite-to-gene tissue co-

association calculated as Ochiai similarity. GO terms were generated by GlueGO (8) from transcriptomic data sharing significant tissue co-associations with targeted metabolites.

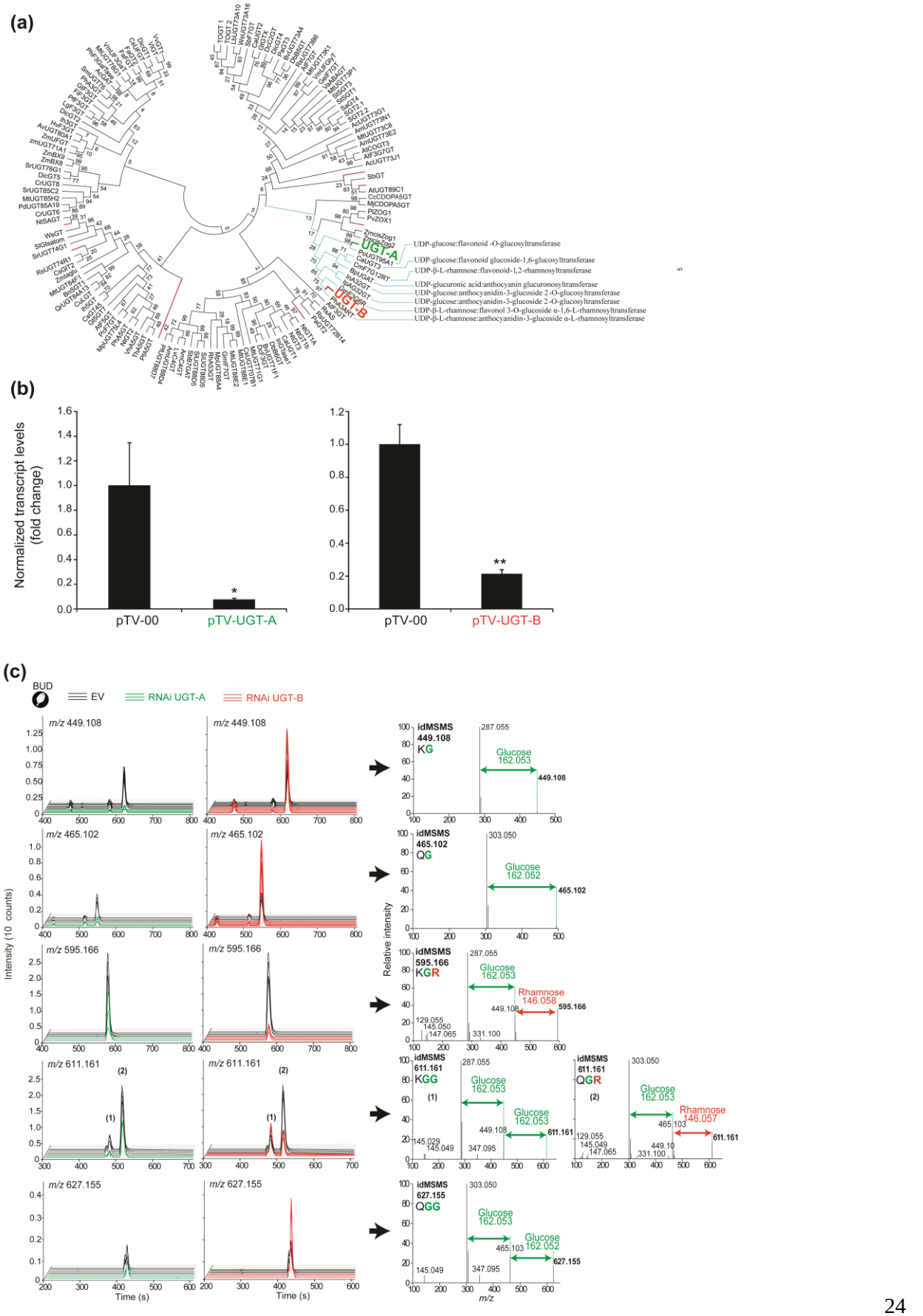
(a)
UGT-A
UGT-B
UDP-glucose:flavonoid -O-glucosyltransferase
UDP-glucose:flavonoid glucoside-1,6-glucosyltransferase
UDP-β-L-rhamnose:flavonoid-1,2-rhamnosyltransferase
UDP-glucuronic acid:anthocyanin glucuronosyltransferase
UDP-glucose:anthocyanidin-3-glucoside 2-O-glucosyltransferase
UDP-glucose:anthocyanidin-3-glucoside 2-O-glucosyltransferase
UDP-β-L-rhamnose:flavonol 3-O-glucoside α-1,6-L-rhamnosyltransferase
UDP-β-L-rhamnose:anthocyanidin-3-glucoside α-L-rhamnosyltransferase
(b)
Normalized transcript levels (fold change)
pTV-00
pTV-UGT-A
pTV-UGT-B
(c)
BUD
EV
RNAi UGT-A
RNAi UGT-B
Intensity (10 counts)
Relative intensity
Time (s)
m/z
idMSMS 449.108 KG
idMSMS 465.102 QG
idMSMS 595.166 KGR
idMSMS 611.161 KGG
idMSMS 611.161 QGR
idMSMS 627.155 QGG
Glucose 162.053
Rhamnose 146.058

Figure S11. Characterization of UGT-A and UGT-B association with flavonoid metabolism. **(a)** The tree was obtained by aligning characterized glycosyltransferases of the GT superfamily 1 (GT 1, 136 GT amino acid sequences) and inferring their phylogenetic relationship using the Maximum Likelihood method (bootstrap = 1000) based on the JTT matrix-based model(9). Evolutionary analyses were conducted in MEGA5 (10). UGT-A (UGT-02515) and UGT-B (UGT-02184) analyzed in the present study are highlighted with purple branches. The red marked branches represent 8 other putative GTs (no names reported) initially considered for virus-induced gene-silencing in *N. attenuata*. Accession numbers for construction of the UDP-Glycosyltransferase tree in different species are list in ***SI Appendix, Materials and Methods.*** **(b)** Gene silencing efficiency for the UDP-glycosyltransferases tested during VIGS experiments. Transcript levels (left panel, *pTV-UGT-A*; right panel, *pTV-UGT-B*) normalized to those of *ELONGATION FACTOR1* were determined in flower buds. Asterisks denote for significant differences between pTV-00 and UGT silenced lines (t-test, **P* < 0.05, ** < 0.01). pTV-00, empty vector; pTV-UGT-A, UDP-glycosyltransferase-A silencing VIGS construct; pTV-UGT-B, UDP-glycosyltransferase-B silencing VIGS construct. **(c)** UHPLC-MS analysis of flower buds of plants inoculated with empty vector and gene silencing constructs for *UGT-A* and *UGT-B*. UHPLC-MS analysis of flower buds of plants inoculated with empty vector and gene silencing constructs for *UGT-A* and *UGT-B.* Chromatograms are traces corresponding to idMS/MS signals with strong co-association with these two UGTs. As supported by the annotation of idMS/MS spectra, silencing *UGT-A* decreases the glucosylation of flavonols while silencing *UGT-B* decreases their additional rhamnosylation (**Figure 5c**). Asterisks denote significant differences between empty vector (EV) and UGT silenced lines (t-test, **P* < 0.05, ** < 0.01). KG, kaempferol-3-*O*-glucoside; KGG, kaempferol-3-*O*-sophoroside (glucosyl(1-2)glucoside); KGR, kaempferol-3-*O*-rutinoside (glucosyl(1-2)rhamnoside); QG, quercetin-3-*O*-glucoside; QGG, quercetin-3-*O*-sophoroside (glucosyl(1-2)glucoside); QGR (Rutin), kaempferol-3-*O*-rutinoside (glucosyl(1-2)rhamnoside).

References

1. Martinez O & Reyes-Valdes MH (2008) Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *P Natl Acad Sci USA* 105(28):9709-9714.
2. Tesson BM, Breitling R, & Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics* 11.
3. Saedler R & Baldwin IT (2004) Virus-induced gene silencing of jasmonate-induced direct defences, nicotine and trypsin proteinase-inhibitors in *Nicotiana attenuata*. *J Exp Bot* 55(395):151-157.
4. Gaquerel E, Heiling S, Schoettner M, Zurek G, & Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass

spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *Journal of agricultural and food chemistry* 58(17):9418-9427.
5. Heiling S, Khanal, S., Barsch, A., Zurek, G., Baldwin, I. T., Gaquerel, E. (2016) Using the knowns to discover the unknowns: MS-based dereplication uncovers structural diversity in 17-hydroxygeranyllinalool diterpene glycoside defense production in the Solanaceae. *The Plant Journal* 85(4):561-577.
6. Storey JD (2002) A direct approach to false discovery rates. *J Roy Stat Soc B* 64:479-498.
7. Li S*, et al.* (2012) Gene-sharing networks reveal organizing principles of transcriptomes in *Arabidopsis* and other multicellular organisms. *The Plant cell* 24(4):1362-1378.
8. Bindea G*, et al.* (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8):1091-1093.
9. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* 8(3):275-282.
10. Tamura K*, et al.* (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731-2739.