Supporting Information

Navigating natural variation in metabolism: structural analysis of MS/MS profiles from herbivory-induced coyote tobacco populations

Dapeng Li, Ian T. Baldwin, Emmanuel Gaquerel

Materials and Methods

Genetic manipulation of phenolamide-biosynthetic and lipoxygenase genes

The stably-transformed RNAi line silenced for *MYB8* as well virus-induced gene silencing (VIGS) constructs for *AT1* (*N*-acyltransferase producing putrescine-based phenolamides) and DH29 (*N*-acyltransferase producing spermidine-based phenolamides) are described in (1) and were used under the plant growth and inoculation conditions reported in this article. Briefly, 200- to 300-bp fragments of *N. attenuata AT1* and *DH29* genes were amplified by PCR. 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 pTVAT1, pTVDH29, or pTV00 constructs. 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. After PDS-VIGS leaves developed a strong bleaching phenotype, AT1-, DH29- and EV-VIGS-silenced plants were used for W+OS treatment as for in the case of the natural accession experiment.

Plants stably silenced for LOX2 expression have been notably been characterized by Allmann *et al.* (2). ir*LOX2* do not emit detectable levels of green leaf volatiles. Previous work has shown that this fatty oxidation branch and that of the LOX3-dependent jasmonate biosynthesis function independently and that manipulating one of the two LOX does not significantly alter the

flux through the other LOX-dependent branch (2). The cross between irLOX2 and irLOX3 is deficient in both GLV emissions and jasmonate accumulation.

Metabolite extraction for UHPL-ESI/qTOF-MS analysis

For metabolite profiling, approximately 100 mg of ground leaf tissue was weighed. 1 mL extraction buffer (50 mM acetate buffer, pH 4.8, containing 40 % methanol) per 100 mg tissue was added and samples were homogenized in a ball mill (Genogrinder 2000; SPEX CertiPrep) for 45 s at 1× rate and 250 strokes per min. Homogenized samples were centrifuged at 16,000g, 4°C for 30 min, and supernatants were transferred into 1.5 mL microcentrifuge tubes and re-centrifuged as before. Supernatants of 400 µL were transferred to 2 mL glass vials for mass spectrometry-based metabolomics.

Conditions for UHPLC-ESI/TOF-MS profile mode analysis

The column used was a Acclaim column (150×2.1 mm, particle size 2.2 µm) with a 4 mm×4 mm i.d. guard column of the same material. The following binary gradient was applied using Dionex Ultimate 3000 UHPLC system: 0 to 1 min, isocratic 90% A (de-ionized water, 0.1% [v/v] acetonitrile and 0.05% formic acid), 10% B (acetonitrile and 0.05% formic acid); 1 to 9 min, gradient phase to reach 20% A, 80% B; 9 to 15 min, isocratic 20% A, 80% B. Flow rate was 200 µL/min. Eluted compounds were detected by a high-resolution MicroToF 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 200°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% v/v 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

Below are the rules we implemented to improve the correlation calculation output. Rules and the PCC calculation were implemented as a C# script. An issue is that many false positives are produced during the correlation analysis due to the fact that some *m/z* features are only detected in a few samples and that background noise correction can spuriously generate high

correlations. To reduce such false positive errors, we filtered the results of the correlation analysis. To this end, we compared data processing 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 5 times the background value and considered these as "true peaks". Only *m/z* signals with at least three "true peaks" for the 10 samples precursors (MS1) and fragments datasets were considered for PCC calculation. 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 10 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 based the sample position for occurrence precursor and assigned fragments. 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 is based on the composite spectrum concept and takes into account the different collision energy conditions required to maximize fragmentation possibilities since some fragments can be detected only under certain collision energies. After applying the entire pipeline and set of rules, 360 deconvoluted non-redundant spectra were reconstructed.

Co-expression network construction and statistical analysis of the effects of natural variation

The intra-distance for each location was calculated by the average of Euclidean distance crosswise samples within each location and inter-distance for each location was calculated by first computing the average of samples at each location and then computing the average Euclidean distance of all other locations calculated with this location in **Fig. 1b**. Euclidean distance in **Fig. 1c** was calculated for each sample pair. The geographic distance was inferred from the GPS coordinates of samples using the Great-circle Distance algorithm. The degree of variation of each mass feature of the dataset was estimated using the relative median absolute distance calculated as follows:

$$relative\ MAD = \frac{median\ (|Xi - median(X)|)}{median\ (X)}$$

with *Xi* being *i*th value across the population for the m/z signal denoted as *X*.

Coexpression calculations for network construction were computed using the Cytoscape plugin MetaNetter (v2.1) (3).

idMS/MS molecular networking by biclustering

A comparative correlation heat map of 360 idMS/MS spectra was constructed using DiffCoEx (4). The parameters of "cutreeDynamic" were set to cutHeight = 0.999, deepSplit = 1, minClusterSize = 10. The R source code of DiffCoEx is downloaded from additional file 1 in (4), the required R WGCNA package can be found at: http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA.

Tables

Table S1. GPS coordinates of the 43 accessions.

Plant label	Accession	latitude	longitude
1-6	A1	37° 9'28.27"N	113°47'18.16"W
7-11	A2	37° 4'36.82"N	113°46'19.12"W
12-17	A3	37°13'15.83"N	113°48'20.86"W
18-20	A4	37°19'30.66"N	113°57'54.91"W
21-24	A5	37° 9'1.30"N	113°47'43.36"W
25-27	A6	37° 1'2.00"N	113°48'44.10"W
28-29	A7	37° 8'19.58"N	114° 1'35.10"W
30-34	A8	37°20'47.18"N	114° 2'59.84"W
35-36	A9	37° 7'22.15"N	114° 1'36.67"W
37-40	A10	37°19'39.83"N	113°56'27.96"W
41-42	A11	37°20'52.474"N	114° 05'57.799"W
43-47	A12	37°21'02.580"N	114° 05'53.661"W
48-51	A13	37°21'07.103"N	114° 05'50.298"W
52-55	A14	37°19'18.648"N	114°06'53.802"W
56-60	A15	37°19'35.48"N	113°57'38.28"W
61-62	A16	37°21'35.24"N	113°56'38.68"W
63-67	A17	37°13'52.699"N	113° 50'37.113"W
68-72	A18	37°15'28.5"N	114°10'13.3"W
73-77	A19	37°17'09.1"N	114°07'31.5"W
78-82	A20	37°21'12.0"N	114°04'16.7"W
83-88	A21	37°08'35.294"N	114°01'30.659"W
89-91	A22	37°07'03.674"N	113° 49'06.197"W
92-98	A23	37°06'12.5"N	113°49'36.6"W
99-101	A24	37° 8'34.19"N	114° 1'22.95"W
102-106	A25	37° 8'20.56"N	114° 1'29.48"W
107-112	A26	37° 8'16.83"N	114° 1'38.25"W
113-117	A27	37° 9'45.30"N	114° 0'58.52"W
118-120	A28	37°19'33.89"N	113°57'54.16"W
121-122	A29	37°13'5.50"N	113°48'24.25"W
123-135	U30	37°19'36.26"N	113°57'53.05"W
136-138	A31	37°20'22.52"N	114° 2'40.86"W
139-141	A32	37°21'11.21"N	113°57'6.03"W
142-148	A33	37°21'1.04"N	113°57'5.17"W
149-151	A34	37°16'38.65"N	113°53'35.18"W
152	A35	37°20'49.11"N	113°57'7.97"W
153-156	A36	37°21'24.29"N	113°57'0.96"W
157-160	A37	37°18'14.26"N	113°56'23.81"W
161-163	A38	37°19'52.45"N	113°57'31.64"W
164	A39	37°20'9.96"N	114° 2'12.62"W

165-168	A40	37°14'27.05"N	113°49'36.71"W
169-170	A41	35°13'7.04"N	111°27'46.10"W
171-172	A42	35°12'56.07"N	111°27'41.29"W
173-183	A43	37°45'19.61"N	118°35'41.82"W

Table S2. List of the 52 neutral losses used for spectral similarity calculations.

The neutral losses used in binary NL matrix generation for DiffCoEx biclustering. In addition to neutral losses that are commonly encountered during tandem MS fragmentation, which corresponded to *N. attenuata* secondary metabolite classes that had been previously annotated for MS/MS spectra, for instance, putrescine and spermidine NL in the case of phenolamides; isobutyric, methyl butanoic and methyl pentanoic acids as well as acetylated fructose losses for *O*-acyl sugars; and finally malonic anhydride and deoxy-hexoses for HGL-DTGs are also included in the NL table. If one of these NLs was detected in an idMS/MS, a score of 1 (otherwise 0) was assigned in order to create a binary vector.

No.	Loss name	Loss formula	Mass difference ±
1	Alkane chains, waxes, fatty acids, methylation	CH_2	14.0157
2	Methane	CH_4	16.0313
3	Ammonium adduct/neutral ammonium loss	NH_3	17.0265
4	Water addition/loss	H_2O	18.0106
5	Salt adduct	K^+ to NH_4^+	20.9293
6	Salt adduct	Na^+ to H^+	21.9819
7	Ethine	C_2H_2	26.0157
8	Carbon monoxide	CO	27.9949
9	Ethene	C_2H_4	28.0313
10	Methylimine	CH_3N	29.0266
11	Formaldehyde	CH_2O	30.0106
12	Methylamine	CH_5N	31.0422
13	Sulfur	S	31.9721
14	Hydrogen sulfide	H_2S	33.9877
15	Salt adduct	K^+ to H^+	37.9559
16	Ketene	C_2H_2O	42.0106
17	Propylation	C_3H_6	42.0470
18	Cyanic acid	CHNO	43.0058
19	Carbon dioxide	CO_2	43.9898
20	Formic acid	CH_2O_2	46.0055

21	Butylation	C_4H_8	56.0626
22	Trimethylamine	C_3H_9N	59.0735
23	Acetic acid	$C_2H_4O_2$	60.0211
24	Urea	CH_4N_2O	60.0324
25	Sulfur dioxide	SO_2	63.9619
26	Pentene	C_5H_8	68.0626
27	Propionic acid	$C_3H_6O_2$	74.0368
28	Benzene	C_6H_6	78.0470
29	Sulfur trioxide	SO_3	79.9568
30	Malonic anhydride	$C_3H_2O_3$	86.0004
31	Isobutyric acid	$C_4H_8O_2$	88.0517
32	Putrescine	$C_4H_{12}N_2$	88.1000
33	Sulfuric acid	H_2SO_4	97.9674
34	Phosphoric acid	H_3PO_4	97.9769
35	Methyl butanoic acid	$C_5H_{10}O_2$	102.0618
36	Malonic acid	$C_3H_4O_4$	104.0110
37	Methyl pentanoic acid	$C_6H_{12}O_2$	116.0861
38	Dihydrogen vinyl phosphate	$C_2H_5O_4P$	123.9926
39	Pentose equivalent	$C_5H_8O_4$	132.0423
40	Spermidine	$C_7H_{19}N_3$	145.1579
41	Deoxy-Hexose equivalent	$C_6H_{10}O_4$	146.0579
42	Hexose-H2O	$C_6H_{10}O_5$	162.0528
43	Deoxy-Hexose equivalent	$C_6H_{12}O_5$	164.0685
44	Glucuronic acid-H2O	$C_6H_8O_6$	176.0321
45	Hexose	$C_6H_{12}O_6$	180.0634
46	Glucuronic acid	$C_6H_{10}O_7$	194.0427
47	Acyl-fructose	$C_8H_{12}O_6$	204.0655
48	Sinapic acid - H2O	$C_{11}H_{10}O_4$	206.0579
49	Glutathione+O-H2O	$C_{10}H_{15}N_3O_6S$	305.0682
50	Glutathione	$C_{10}H_{17}N_3O_6S$	307.0838
51	Sucrose-H2O	$C_{12}H_{20}O_{10}$	324.1057
52	Sucrose	$C_{12}H_{22}O_{11}$	342.1162

Table S3. Comparative analysis of *N*-caffeoylputrescine and unknown at *m/z* 347.19 regulation in AT1 and DH29-VIGs silenced plants.

Data files (4 biological replicates per genotype) from the experiment presented in Fig. 5d were processed by XCMS (cf. script in Dataset S1) to extract peak area values for precursor ions of isomers of *N*-caffeoylputrescine and unknown at *m/z* 347.19. Fold changes between average peak intensity values of empty vector VIGs control samples and of *AT1*- or *DH29*-VIGs silenced samples. Statistics derived from unpaired are additionally reported. Mass features for these metabolites were not detected in ir*MYB8* samples.

m/z	RT (s)	Annot.	*P* value, unpaired t-test		Fold-change		Presence in irMYB8
			AT1-VIGs	DH29-VIGs	AT1-VIGs	DH29-VIGs	
251.134	244	*N*-caffeoyl-putrescine, [M+H]+	0.0000	0.4733	0.16	1.46	**Not detected**
251.134	319	*N*-caffeoyl-putrescine, [M+H]+	0.0000	0.0706	0.18	1.69	**Not detected**
347.193	690	Unidentified *N*-caffeoyl-putrescine metabolite, [M+H]+	0.0000	0.0518	0.20	1.28	**Not detected**
347.193	733	Unidentified *N*-caffeoyl-putrescine metabolite, [M+H]+	0.0000	0.0620	0.14	1.19	**Not detected**

Datasets

Dataset S1. Output of the XCMS and CAMERA processing and statistical analysis and XCMS and CAMERA scripts.

Dataset S2. 360 idMS/MS deconvoluted spectra and NDP and NL matrix used for biclustering. First column corresponds to the 360 idMS/MS precursors in the format of "pcgroup_mz_rt". First row corresponds to the 360 idMS/MS precursors and the 52 NL used for calculation.

Figures

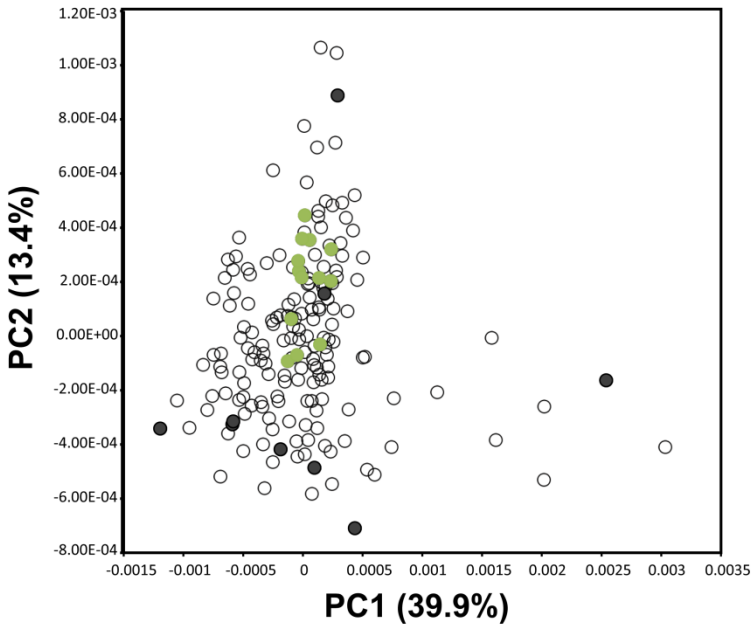

Figure S1. Principal component analysis of the UHPLC-ToFMS metabolic profiles elicited by herbivory for 183 glasshouse-grown plants from 43 natural accessions.

Principal component analysis score plot constructed from autoscaled data derived from the metabolic profiles of 183 herbivory-induced leaf samples of plants grown in the glasshouse from seeds collected in different locations of Southwest of the USA. The two first principal components (PCs) extracted during the modeling accounted for 39.9% and 13.4% of the total variance existing in the sample population. Sample coordinates on the extracted PCs are reported in Dataset S1. Black dots indicate the 10 samples according to PCA-based distant clustering (when considering the first six PCs), geographic origin of the corresponding seeds and manual inspection of chromatographic profiles for metabolic space analysis by indiscriminant MS/MS (see SI Appendix, Fig. S3). Green dots correspond to U30 (Utah accession inbred in the glasshouse for 30 generations) metabolic profiles on the PCA score plot.

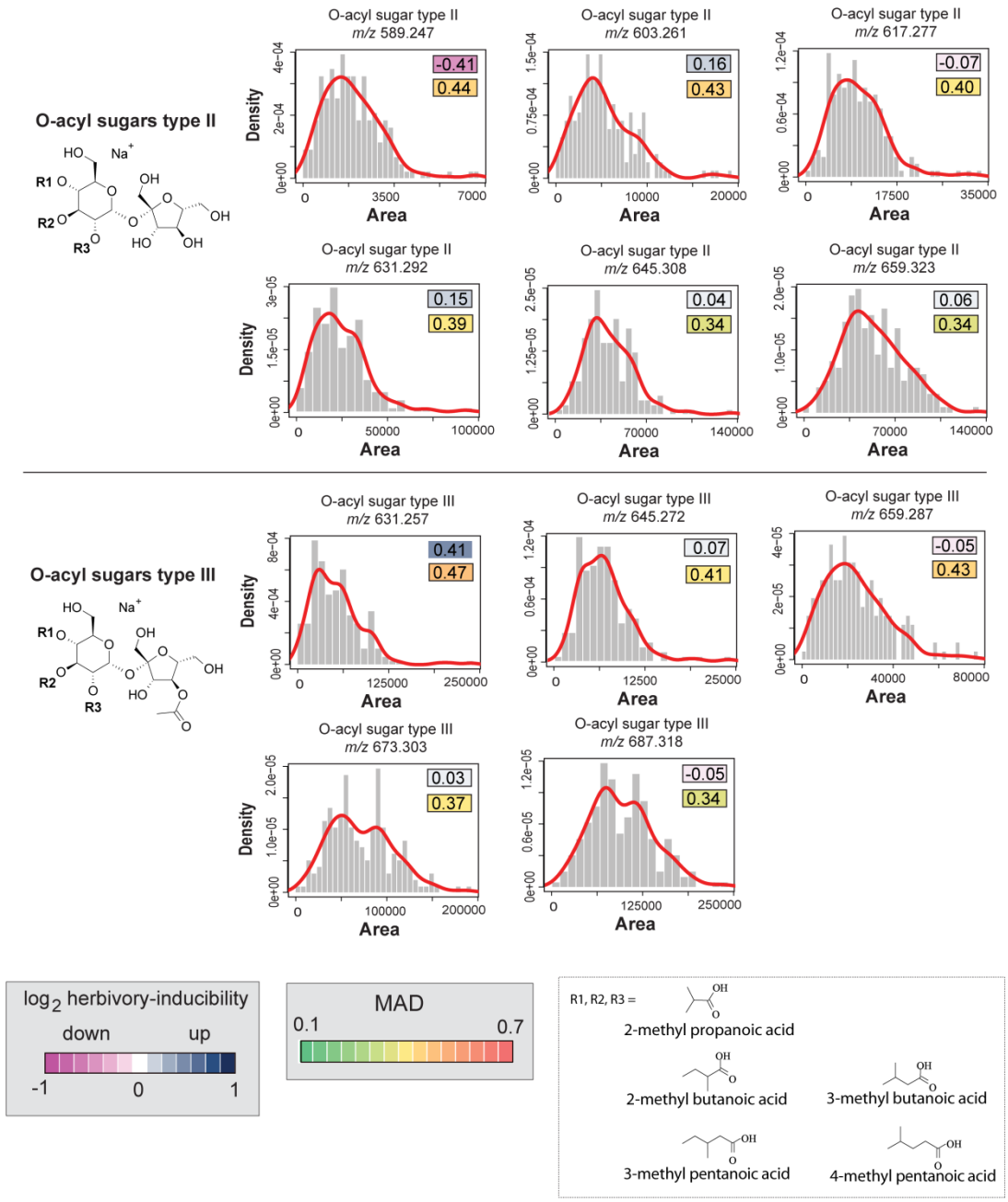

Figure S2. Natural variation analysis for type II and type III *O*-acyl sugars.

Distribution histogram plots are overlaid with a density line depicting distribution of intensities for type II and type III *O*-acyl sugars across 183 individual plants with the x axis corresponding to the area of intensities and the y axis corresponding to the fitted density curve for the histogram. The degree of inducibility for each individual *O*-acyl sugar is denoted by different color from pink (low) to blue (high) as log2-scaled fold change value between herbivory-induced and control samples of U30. The natural variation coefficients are calculated as relative median absolute

distance (relative MAD). The degree of MAD is indicated by the node color from green (low) to red (high). The structures of *O*-acyl sugars are given next to the distribution histogram plots. *O*-acyl sugars are not inducible by the W+OS treatment (5) but natural variations in their abundance were relatively important (MAD ranging from 0.34 to 0.47 for the corresponding $(M+H)^+$ *m/z* signals). Type II and type III *O*-acyl sugars differ in the presence of an acetyl group on their fructose moiety (6, 7, 8). Statistical analysis on fold-change regulation is summarized in Dataset S1.

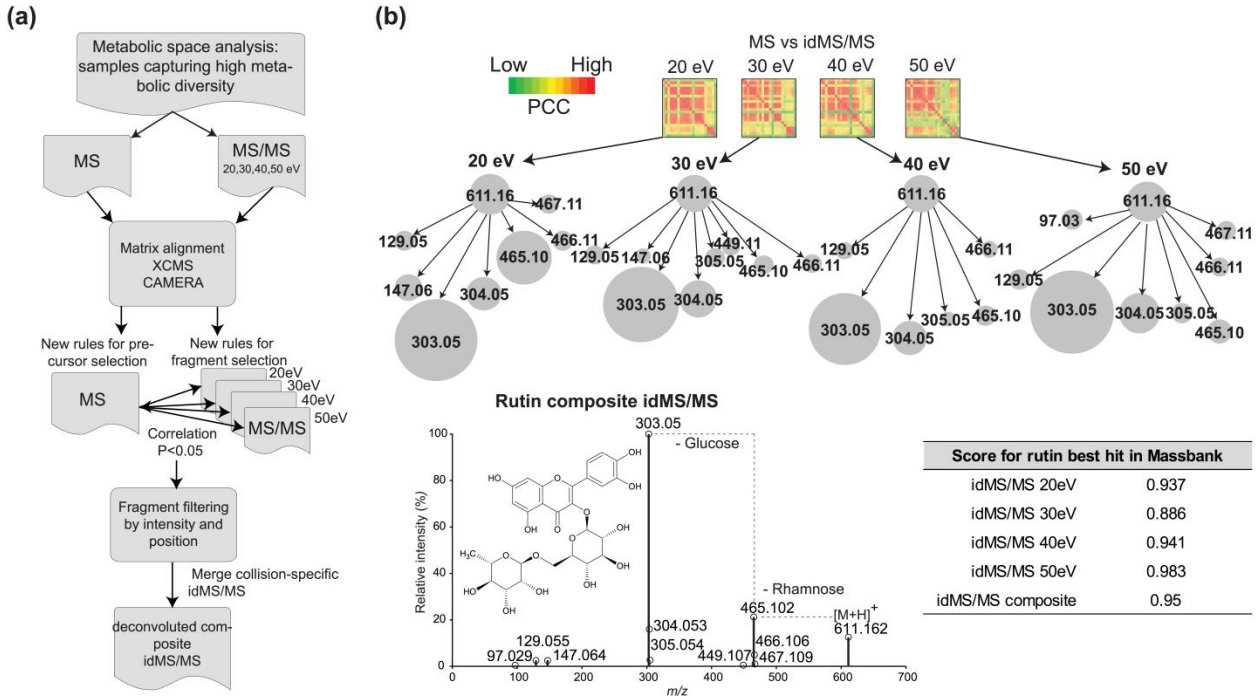

Figure S3. Workflow for metabolic space analysis by shotgun MS/MS and assembly of compound-specific idMS/MS. (a) idMS/MS workfow. To maximize the coverage by shotgun MS/MS of structurally diverse metabolites within a wide dynamic range of concentrations, 10 samples from the clustering analysis of 183 *N. attenuata* accessions were first selected based on the dissimilarity of their processed chemical profiles. These 10 samples were first analyzed by UHPL-qTOFMS using the single MS mode (low fragmentation condition) and then as independent measurements using idMS/MS at 4 different collision energies of 20, 30, 40 and 50ev without any precursor ion selection. Correlation analysis was applied to assign precursor and fragment relationships according to the resulting P-values and newly implemented peak intensity and position filtering functions (see Methods). *m/z* signals co-selected by this pipeline

were further assembled as an individual idMS/MS. **(b)** Correlation-based assembly of rutin (CHEBI:28527) idMS/MS. Heatmaps depict Pearson Correlation Coefficient (PCC) among *m/z* signals for the rutin chromatographic group and the progressive assignment of precursor-fragment relationships. CID voltage-specific (20 to 50ev) precursor-fragment assignments based on PCC are additionally visualized as partial networks with branch length indicative of the PCC significance calculated within the chromatographic group "PCGROUP" suggested by CAMERA and node size is indicative the *m/z* relative intensity. Edge length is inversely proportional to the *P* value: fragments with highly statistically significant precursor-to-fragment PC, the closer the fragments are pulled towards the precursor and the higher the intensity of the fragment, the larger of the space occupied by the circle. Smaller fragments like *m/z* 97.03, are only produced at high collision energy (50ev) and are therefore assigned with a significant correlation. Other fragments are detected only under certain collision energy, for example, *m/z* 449.11 is captured solely for the collision energy of 30ev whereas others, such as *m/z* 465.10, vary in intensity but are ubiquitously detected. To validate the quality idMS/MS assembly, we calculated matching scores with the different collision energy reconstructed spectra against the Massbank database that contains targeted MS/MS measurements on different instruments for rutin standards. All highest scores corresponded to rutin hits and the highest score of 0.983 retrieved from the search was obtained when using the idMS/MS at the 50ev collision energy. The composite idMS/MS for rutin returned a matching score of 0.950 in Massbank.

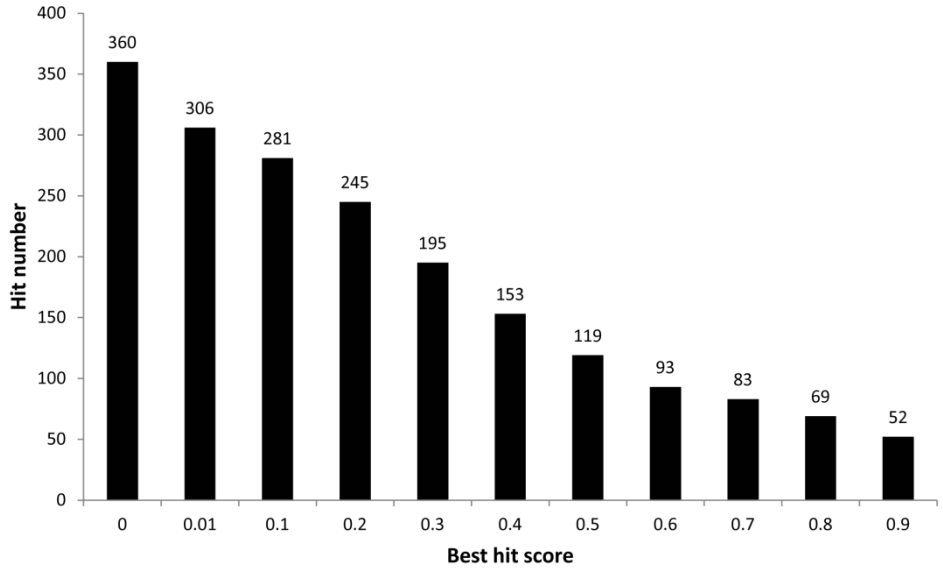

Figure S4. Distribution of idMS/MS score hits during alignment with the MS/MS database Massbank.

Deconvoluted idMS/MS spectra were searched against the MS2 database Massbank. The best hit score from Massbank of each idMS/MS are reported on the x axis. The number of hits that were above of a certain score value is plotted along the y axis.

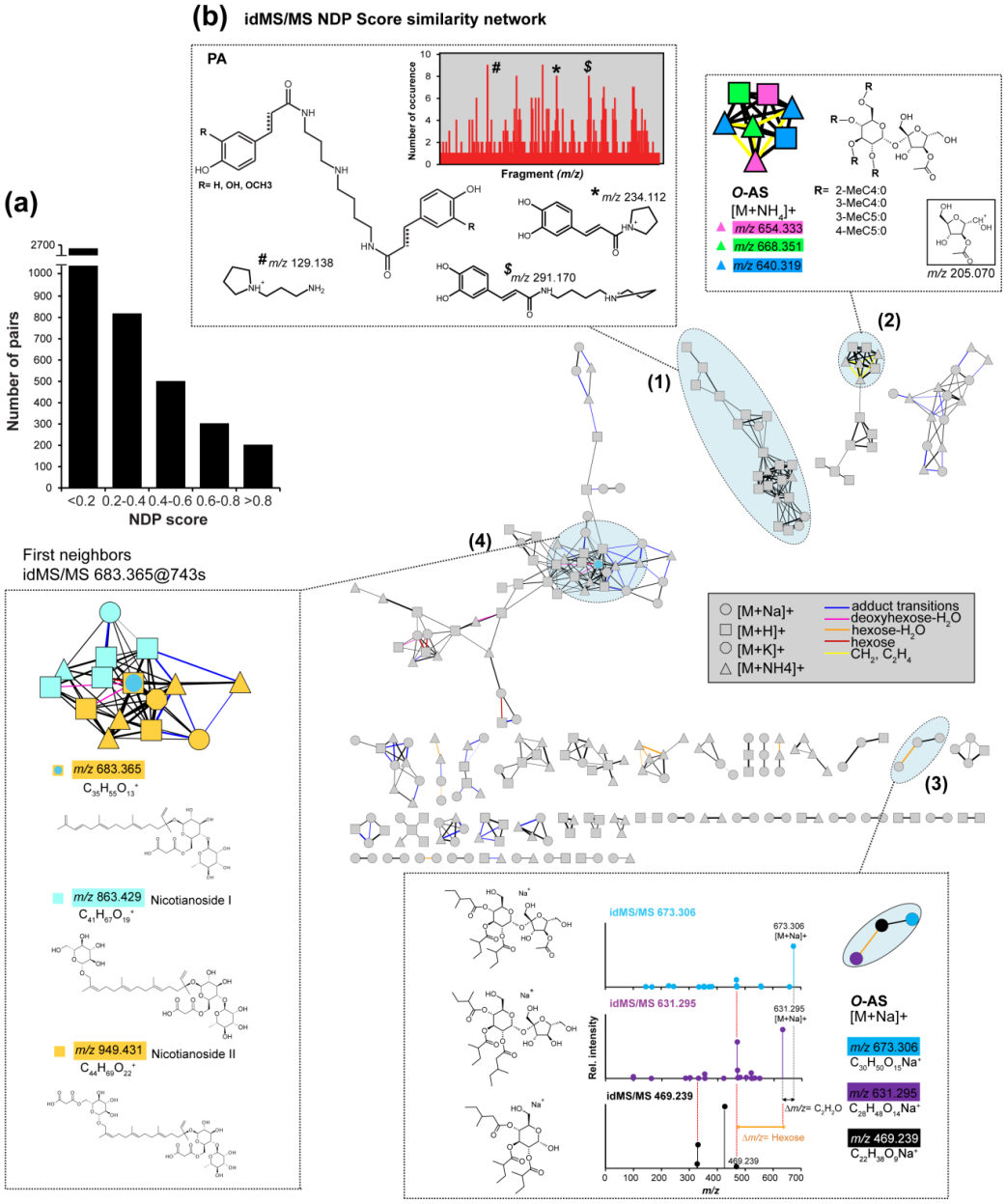

Figure S5. idMS/MS spectral similarity network constructed solely based on pairwise NDP fragment similarity calculations.

(a) Distribution histogram of pairwise NDP scores. **(b)** Molecular network constructed in Cytoscape using the Biolayout for NDP scores > 0.2. A first attempt for high-resolution idMS/MS classification was undertaken using only NDP scores based on fragment similarities. However, since many of the small molecules composing the metabolic profile analyzed produced a limited number of fragments, often < 5 fragments, many pairwise NDP scores are low which does not provide sufficient connectivity for the assembly of a sufficiently resolved molecular

network. The most prominent clusters formed in this analysis corresponded to the alignment of the multiple pseudo-MS3 idMS/MS collected for single metabolites being prone to in-source fragmentation during ionization such as HGL-DTGs but which limits their utility in mining known-to-unknown metabolite relationships. Nevertheless, selected clusters are examined. Cluster **(1)** contains several known phenolamides (PA). This cluster illustrates how groups of known compounds with spectral data classified by NDP fragment similarity can be mined for "overrepresented" m/z fragments. Here, *m/z* signals corresponding to the arrangement of phenolic acyl moieties were more frequently detected than average. As expected, idMS/MS corresponding to the different adduct forms for the same metabolite are also consistently clustered. This is the case of cluster **(2)** which derives from the fragmentation of one *O*-acyl sugar metabolite, in this example the higher than average NDP value results from the shared fragment *m/z* 205.07 (acetylated fructose residue). Cluster **(3)** associates two *O*-acyl sugar metabolites ($[M+H]^+$, *m/z* 673.306 and $[M+H]^+$, *m/z* 631.295). The third idMS/MS (with precursor *m/z* 469.239) corresponds to a pseudo-MS3 of idMS/MS at *m/z* 631.295. Using NDP alone, the clustering of HGL-DTGs is rather difficult to mine. We illustrate for cluster **(4)**, a reconstruction of a directed sub-network for the first neighbors of idMS/MS 683.365, a *m/z* fragment shared by nicotianoside I and II. We conclude that although interesting insights can be visualized from this network, sequences of fragmentation monitored by neutral loss analysis have also to be taken into account when evaluating spectral similarity in order to better assess known-to-unknown metabolite relationships (Fig. 5). Different node symbols denote the idMS/MS precursors for different adduct forms. Edge colors correspond to the most frequently annotated neutral losses.

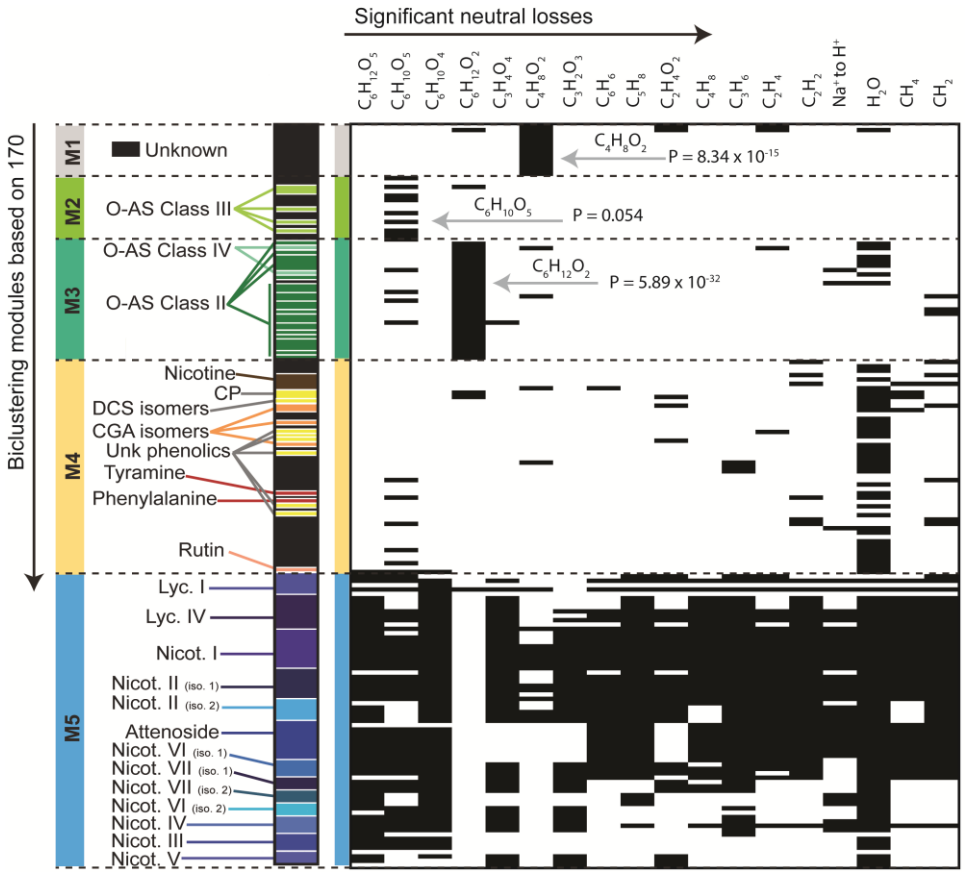

Figure S6. Binary neutral loss map. Black cells within the neutral loss (NL) map (see SI Appendix, Table S2) denote for shared and significantly overrepresented (Chi-Square score analysis) NLs within one module. Statistics for NL significantly overrepresented in modules 1 to 3 are reported for sugar-derived NLs from the backbone structure of *O*-acyl sugars classified into modules 2 and 3. An additional analysis of module 1 is presented in SI Appendix, Fig. S8. As predicted from previous knowledge on their structural characteristics, overrepresentation NLs derived from glucose ($P = 1.38\text{x}10^{-10}$), rhamnose ($P = 1.07\text{x}10^{-33}$) and malonic acid ($P = 4.10\text{x}10^{-22}$)-derived NLs – favored the clustering of HGL-DTGs into one module.

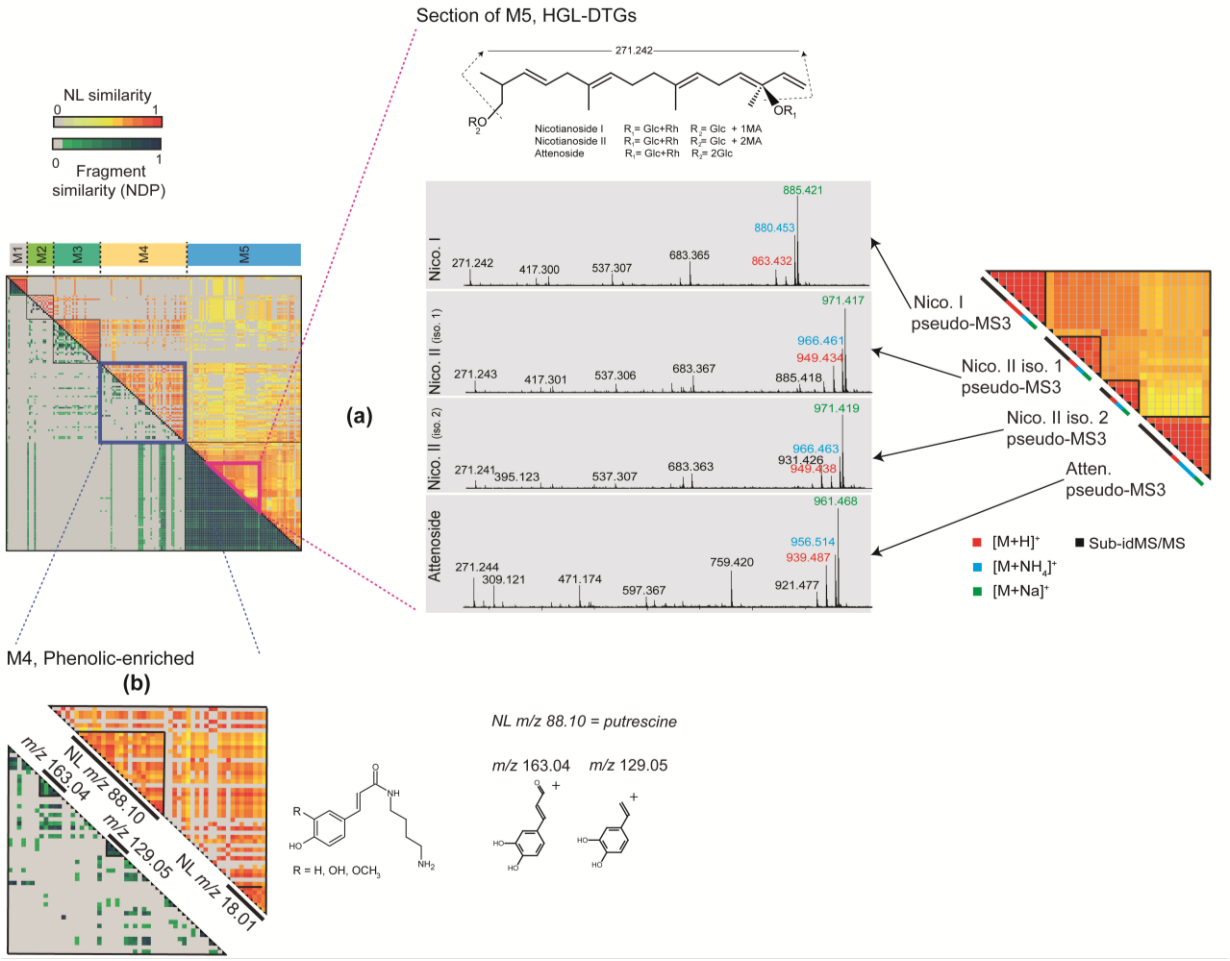

Figure S7. Close-up views on modules 4 and 5 from the biclustering of idMS/MS.

(a) Sub-section of module 5 highlighting the assembly of multiple pseudo-MS3 for different 17-hydroxygeranylinalool diterpene glycosides (HGL-DTGs). HGL-DTG fragmentation has been described by Heiling *et al.* (9). idMS/MS for different adduct forms of single type of HGL-DTGs and the corresponding sub-idMS/MS that derived from in-source fragmentation during ionization are tightly aligned due to high NL and NDP similarities explained by a common backbone-derived fragment at *m/z* 271.242 and common NLs for glucose, rhamnose and malonyl group. **(b)** Close-up view on module 4, previously shown to be enriched in phenolic derivatives. Among others, putrescine-based phenolamide derivatives are clustered due to the presence of a shared NL corresponding to the loss putrescine. High NDP similarities between some of these idMS/MS notably result from a shared fragment corresponding to a caffeoyl moiety. Nico., Nicotianoside; Atten., Attenoside.

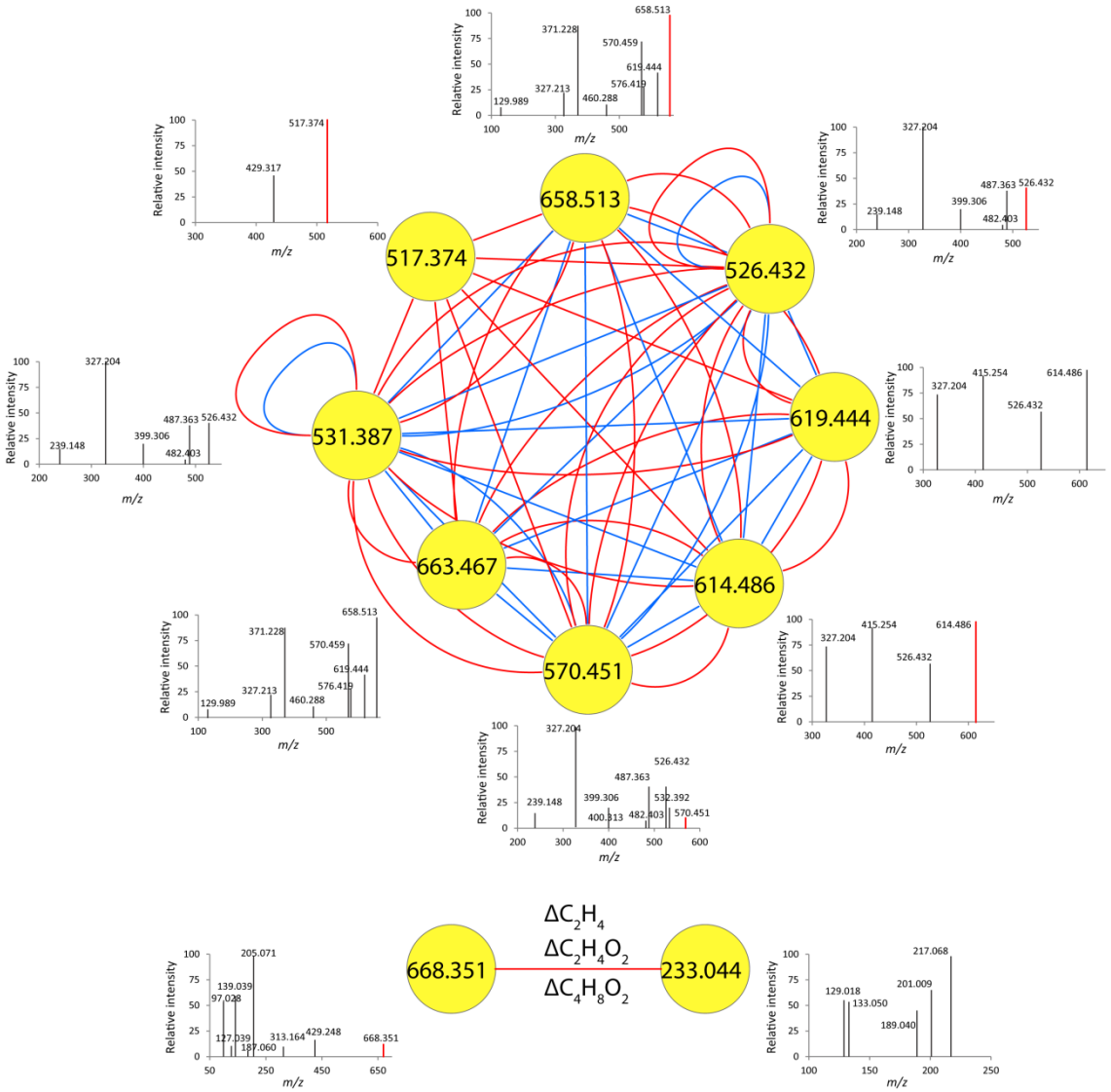

Figure S8. idMS/MS similarity network for unknown metabolites of Module 1 obtained from the biclustering analysis.

Each node represents one precursor of the corresponding idMS/MS spectrum. All possible NDP and NL similarity-based pairs with a score above 0.6 as network edges are assigned with different colors to distinguish NDP and NL connectivities: blue edges correspond to the NDP similarity and red edges to the NL similarity. The approach indicates that Module 1 corresponds likely to three metabolites: $[M+H]^+$, *m/z* 614.486, ($C_{37}H_{64}N_3O_4$); $[M+H]^+$, *m/z* 668.351, ($C_{32}H_{46}N_9O_7$); $[M+H]^+$, *m/z* 233.044, ($C_{12}H_9O_5$). Pseudo-MS3 idMS/MS for the metabolite at $[M+H]^+$, *m/z* 614.486, form a densely connected cluster

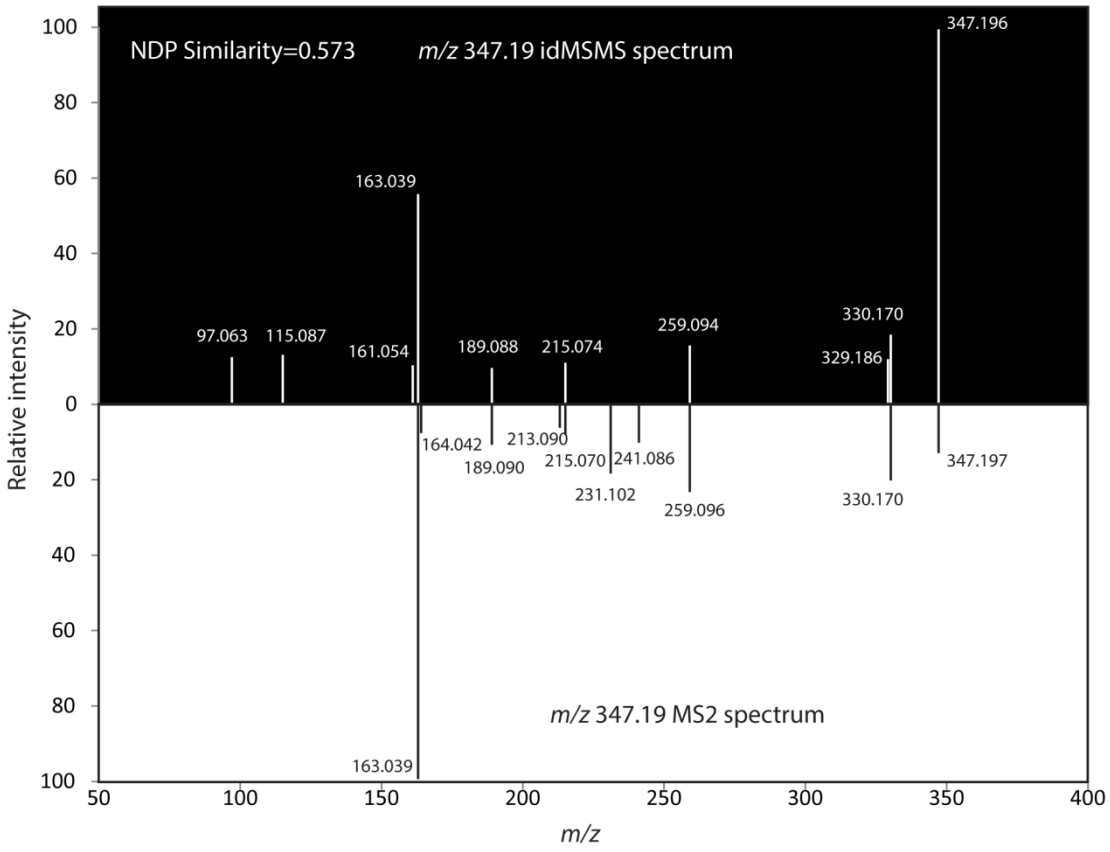

Figure S9. Comparison of the computationally-assembled idMS/MS for *m/z* 347.19 with that obtained by "regular" MS/MS data acquisition.

The experimental idMS/MS spectrum for *m/z* 347.19 is shown (top spectrum) with the corresponding MS2 spectrum obtained with "regular" MS/MS data acquisition using a 0.05 Da mass selection window for a collision energy of 35 eV (bottom spectrum). The similarity score is calculated as the normalized dot product similarity (NDP) score between the experimental idMS/MS spectrum and the MS2 spectrum.

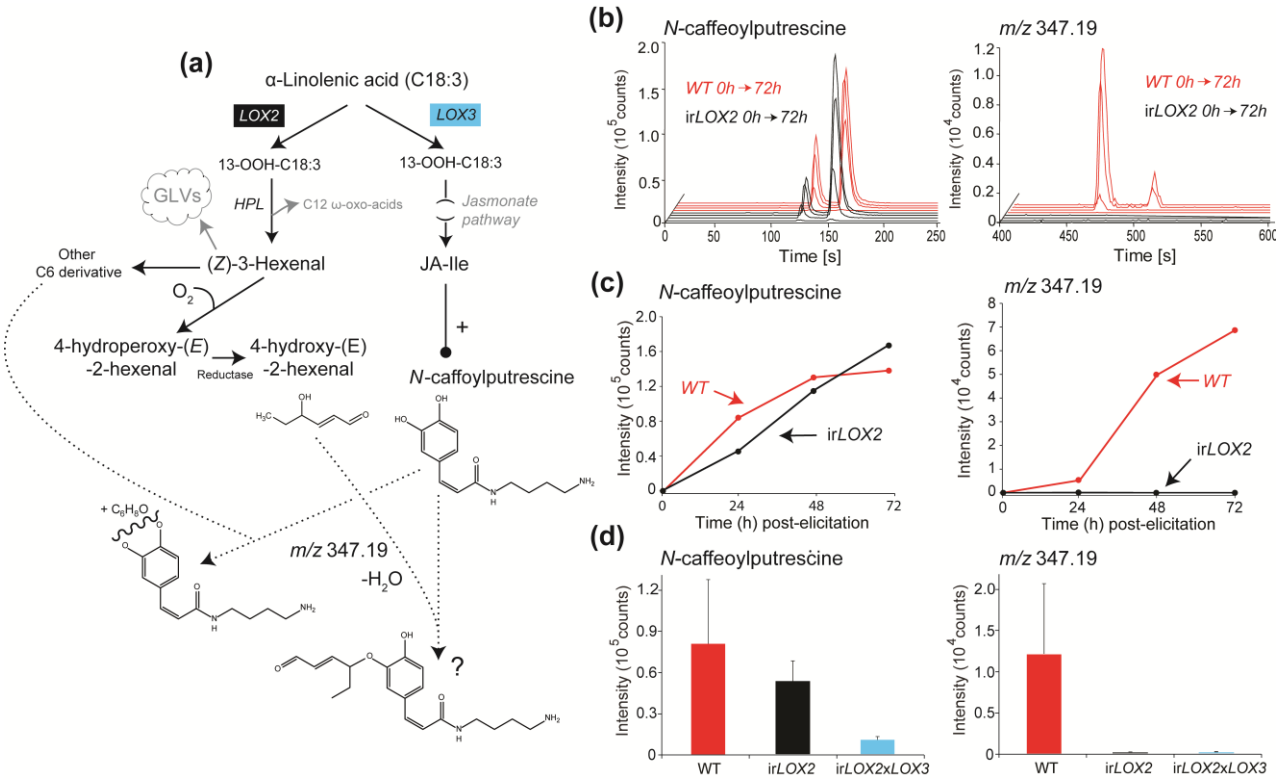

Figure S10. *LIPOXYGENASE2*-mediated fatty acid oxidation into reactive C6 derivatives is required for the metabolic conversion of *N*-caffeoylputrescine into the unknown at *m/z* 347.19.

(a) Hypothesis for the interaction between the LOX2-dependent C6 pathway and *N*-caffeoyputrescine metabolism resulting in the formation of the unknown at *m/z* 347.19. Transient *AT1* silencing and the idMS/MS networking approach indicated that the unknown at *m/z* 347.19 is derived from *N*-caffeoylputrescine by a biochemical transformation involving the attachment of a C_6H_8O residue to the caffeoyl moiety. Results in panels **(b-d)** support that this C6 residue derives from the LOX2-dependent fatty acid oxidation pathway. In this pathway, (*Z*)-3-hexenal is produced by HPL-dependent cleavage of linolenic acid 13-hydroperoxides (13-OOH-C18:3) resulting from the activity of LOX2. Previous work has shown that this fatty oxidation branch and that of the LOX3-dependent jasmonate biosynthesis function independently and that manipulating one of the two LOX does not significantly alter the flux through the other LOX-dependent branch (2). (*Z*)-3-hexenal serves for the production of the wide panel of GLVs, but is also abundantly oxidized into 4-hydroperoxy-(*E*)-2-hexenal which is reduced into 4-hydroxy-(*E*)-2-hexenal. This common metabolite, which has been previously detected in *Nicotiana* spp. in both controlled and stress conditions, is highly reactive and toxic for plant cells. Our current

hypothesis is that it reacts (loss of water involved) with *N*-caffeoylputrescine to form the unknown at *m/z* 347. An hypothesis for the product of this reaction is provided, in which 4-hydroxy-(*E*)-2-hexenal reacts with the C3 alcohol. We do not rule out that other LOX2-based C6 derivatives could be involved in the formation of the unknown at *m/z* 347. On the other hand, the LOX3 pathway serves for JA/JA-IIe signaling which regulates at the transcriptional level the biosynthesis of phenolamides including that of *N*-caffeoylputrescine. **(b)** Extracted ion current traces for *N*-caffeoylputrescine isomers and unknown at *m/z* 347.19 accumulation 0, 24, 48 and 72 h after W+OS treatment in WT (red) and ir*LOX2* (black) leaves. **(c)** Mean normalized levels depicting the temporal accumulation *N*-caffeoylputrescine and unknown at *m/z* 347.19 in WT and ir*LOX2*. **(d)** Mean normalized levels *N*-caffeoylputrescine and *m/z* 347.19 in WT, ir*LOX2* and a cross between ir*LOX2* and ir*LOX3* (no C6 and no jasmonate signaling), 72 h after W+OS elicitation. As expected disrupting jasmonate signaling impairs *N*-caffeoylputrescine accumulation and thereby that of unknown at *m/z* 347.19. But, *LOX2* silencing alone is sufficiently to abolish the conversion of *N*-caffeoylputrescine into the unknown *m/z* 347.19 by C_6 residue addition. LOX2, LIPOXYGENASE2; LOX3, LIPOXYGENASE3; HPL, HYDROPEROXIDE LYASE; GLVs, Green Leaf Volatiles.

Supporting References

1. Allmann, S., Halitschke, R., Schuurink, R. C., & Baldwin, I. T. (2010). Oxylipin channelling in *Nicotiana attenuata*: lipoxygenase 2 supplies substrates for green leaf volatile production. *Plant, cell & environment*, 33(12): 2028-2040.
2. Onkokesung N, Gaquerel E, Kotkar H, Kaur K, Baldwin IT, & Galis I. (2012). MYB8 Controls Inducible Phenolamide Levels by Activating Three Novel Hydroxycinnamoyl-Coenzyme A:Polyamine Transferases in *Nicotiana attenuata*. *Plant Physiology* 158: 389-407.
3. Jourdan F, Breitling R, Barrett MP, & Gilbert D (2008) MetaNetter: inference and visualization of high-resolution metabolomic networks. *Bioinformatics* 24(1):143-145.
4. Tesson BM, Breitling R, & Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics* 11:497.
5. 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.
6. Kim J, et al. (2012) Striking Natural Diversity in Glandular Trichome Acylsugar Composition Is Shaped by Variation at the Acyltransferase2 Locus in the Wild Tomato *Solanum habrochaites*. *Plant physiology* 160(4):1854-1870.
7. Arrendale RF, et al. (1990) Characterization of the Sucrose Ester Fraction from *Nicotiana-Glutinosa*. *Journal of agricultural and food chemistry* 38(1):75-85.
8. Weinhold A & Baldwin IT (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proceedings of the National Academy of Sciences of the United States of America* 108(19):7855-7859.
9. Heiling, S., Schuman, M. C., Schoettner, M., Mukerjee, P., Berger, B., Schneider, B., ... & Baldwin, I. T. (2010). Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *The Plant Cell Online*, *22*(1): 273-292.