Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/20/eaba0721/DC1

Supplementary Materials for

The evolutionary origins of the cat attractant nepetalactone in catnip

Benjamin R. Lichman*, Grant T. Godden, John P. Hamilton, Lira Palmer, Mohamed O. Kamileen, Dongyan Zhao, Brieanne Vaillancourt, Joshua C. Wood, Miao Sun, Taliesin J. Kinser, Laura K. Henry, Carlos Rodriguez-Lopez, Natalia Dudareva, Douglas E. Soltis, Pamela S. Soltis, C. Robin Buell*, Sarah E. O'Connor*

*Corresponding author. Email: benjamin.lichman@york.ac.uk (B.R.L.); buell@msu.edu (C.R.B.); oconnor@ice.mpg.de (S.E.O.)

Published 13 May 2020, *Sci. Adv.* **6**, eaba0721 (2020)
DOI: 10.1126/sciadv.aba0721

This PDF file includes:

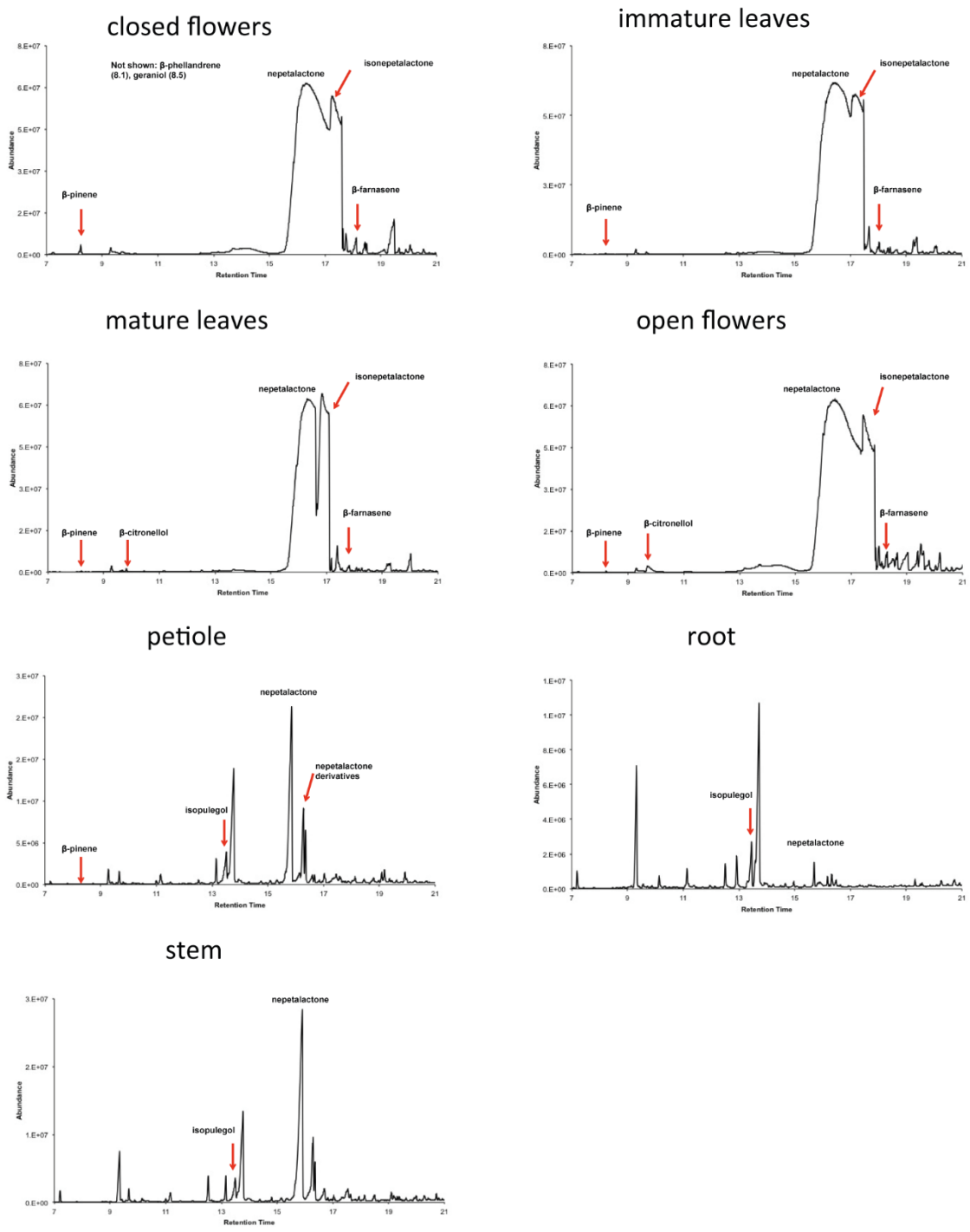


Fig. S1. GC-MS analysis of *Nepeta cataria* tissues.

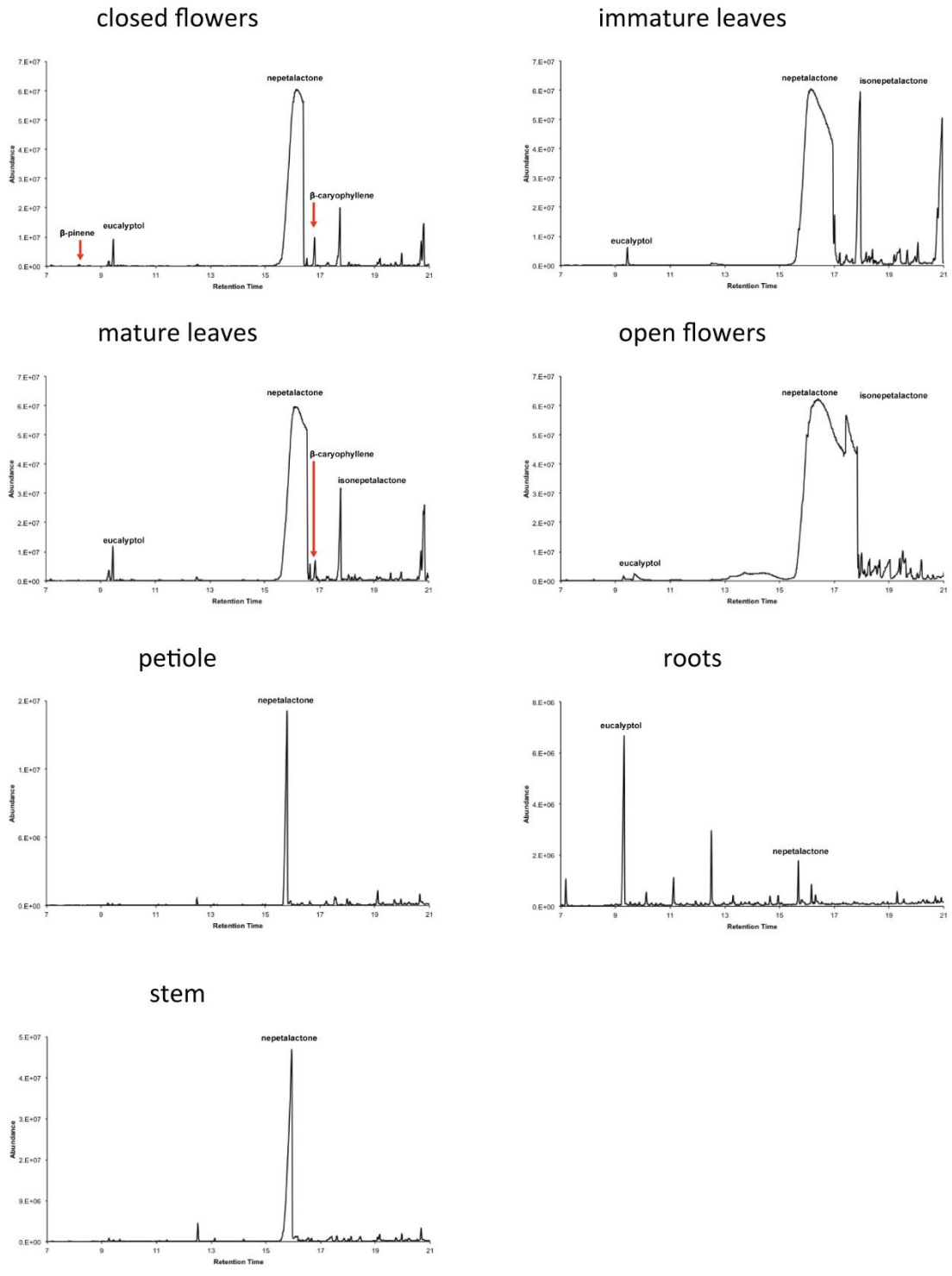


Fig. S2. GC-MS analysis of *Nepeta mussinii* tissues.

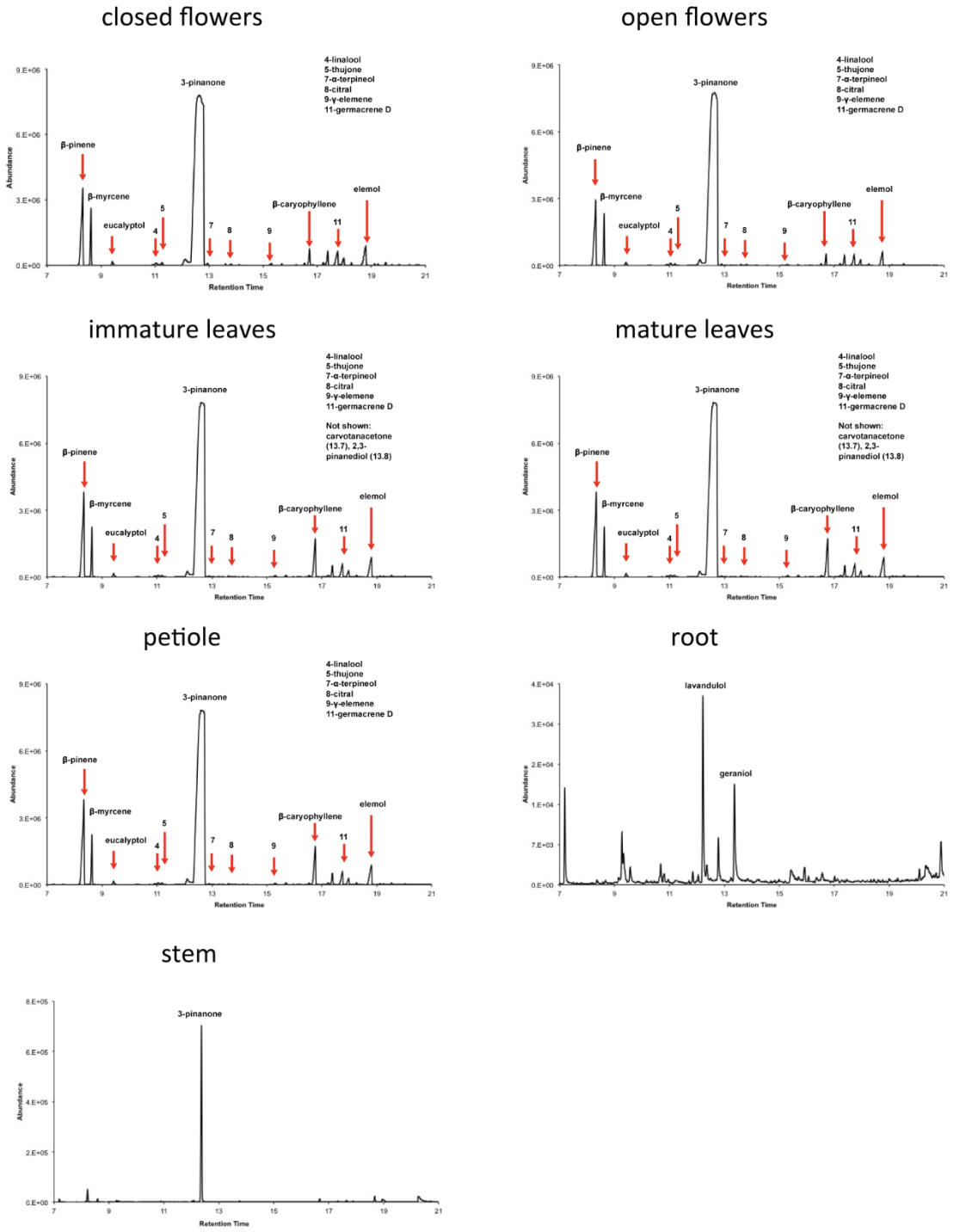


Fig. S3. GC-MS analysis of *Hyssopus officinalis* tissues.

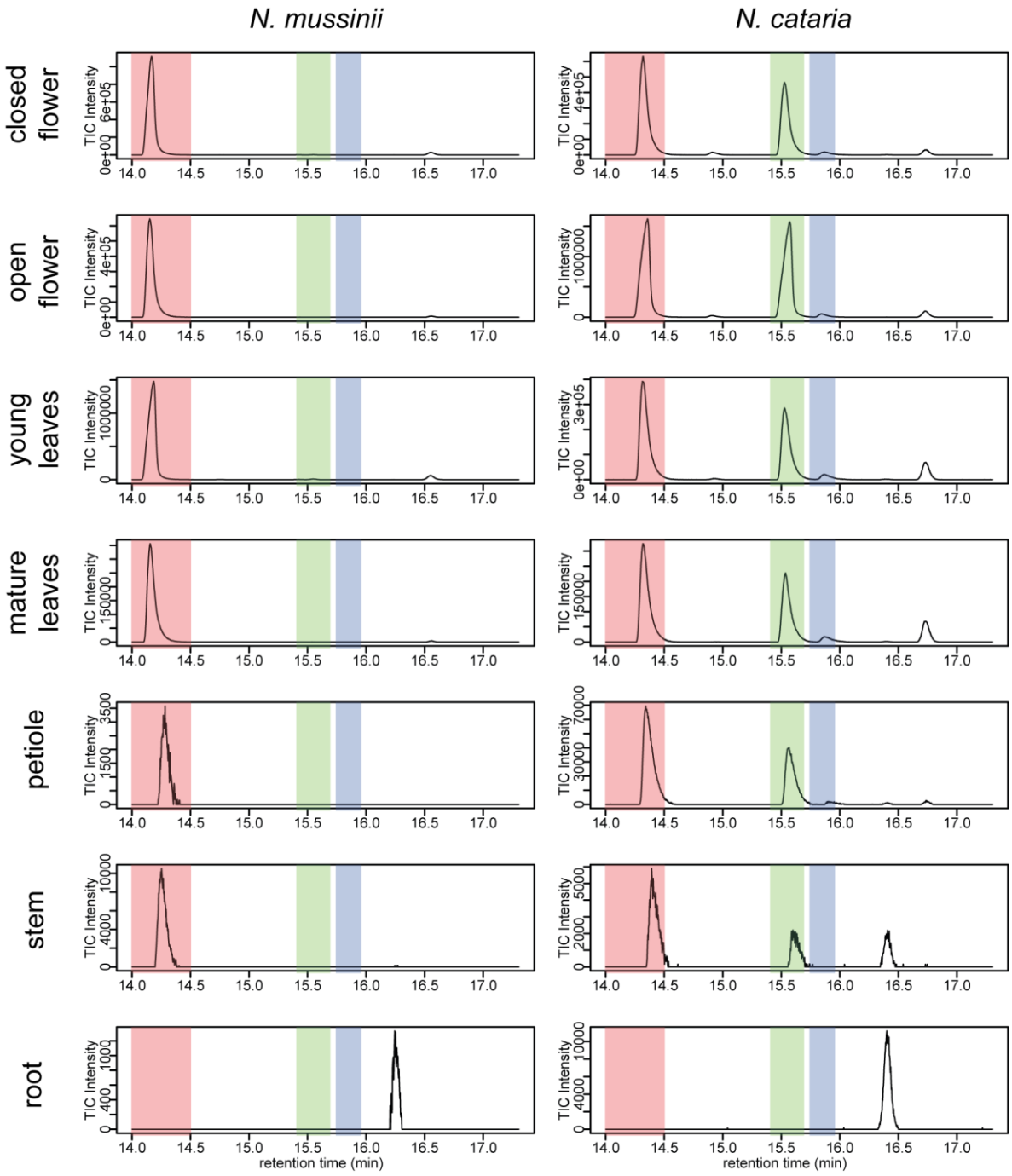
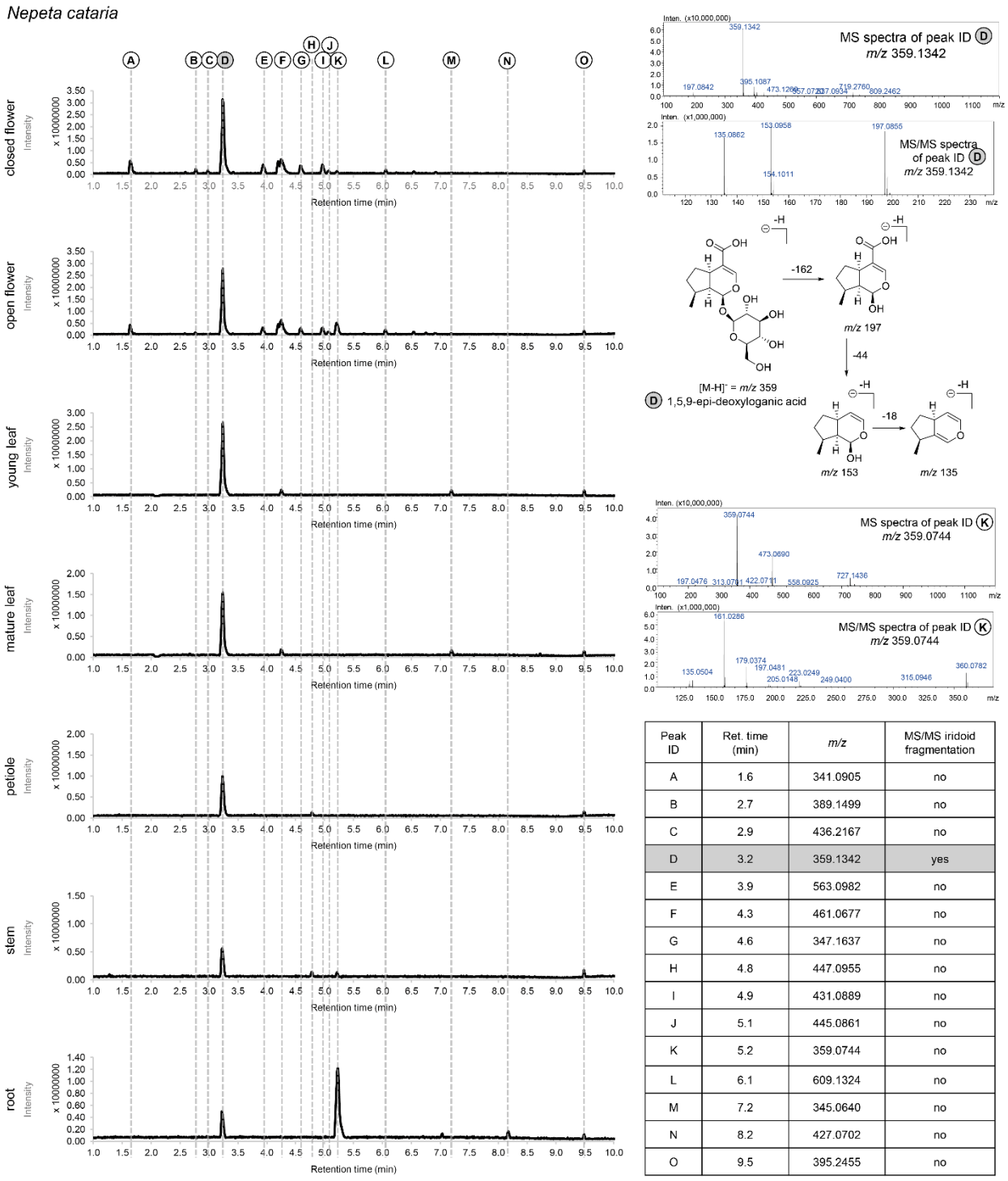
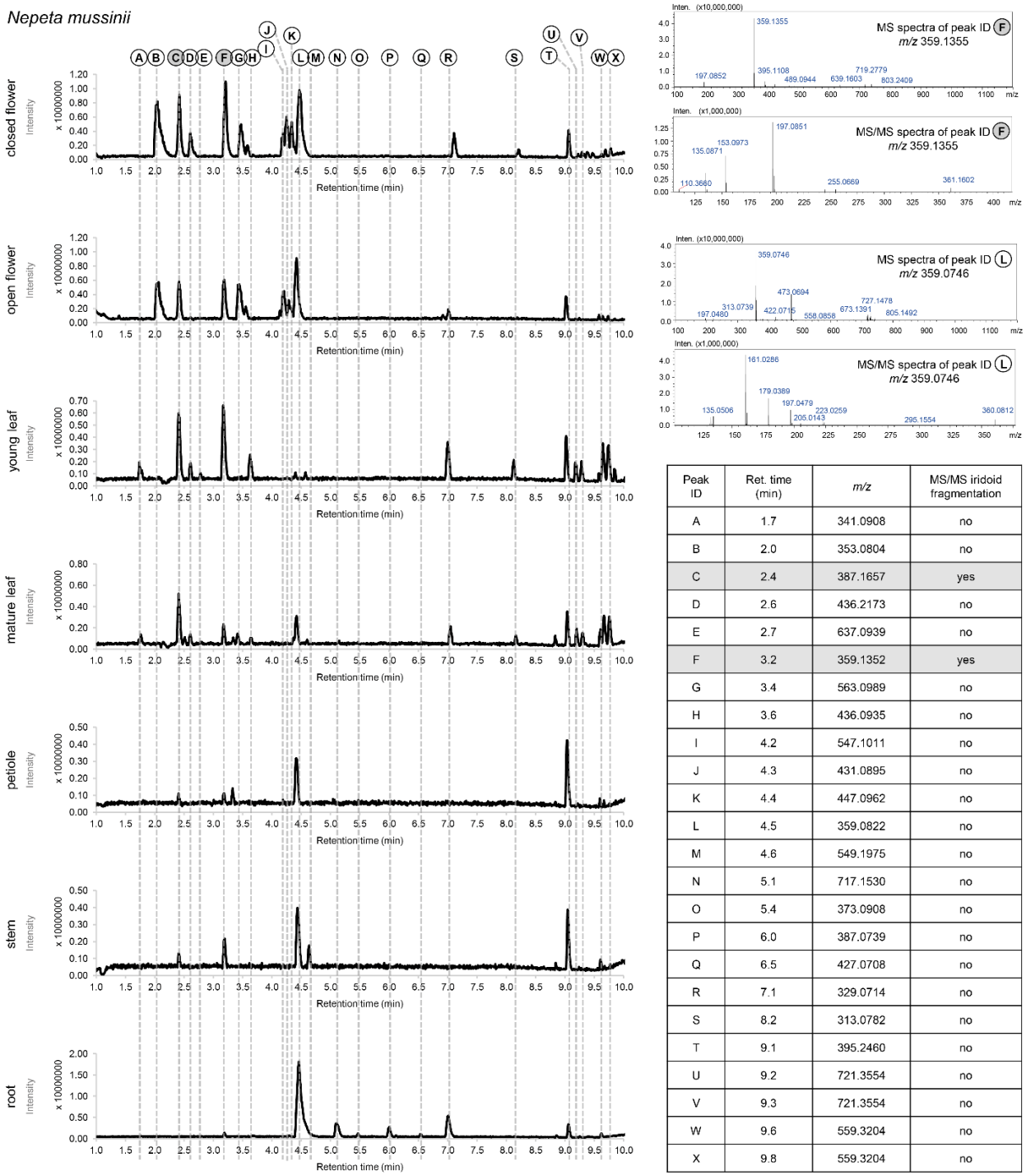


Fig. S4. Nepetalactone content of *Nepeta* tissues. Analysis conducted by GC-MS analysis, with the total ion chromatograms (TICs) depicted. Pink peak: cis-trans-nepetalactone; green peak: trans-cis-nepetalactone; lilac peak: cis-cis-nepetalactone.



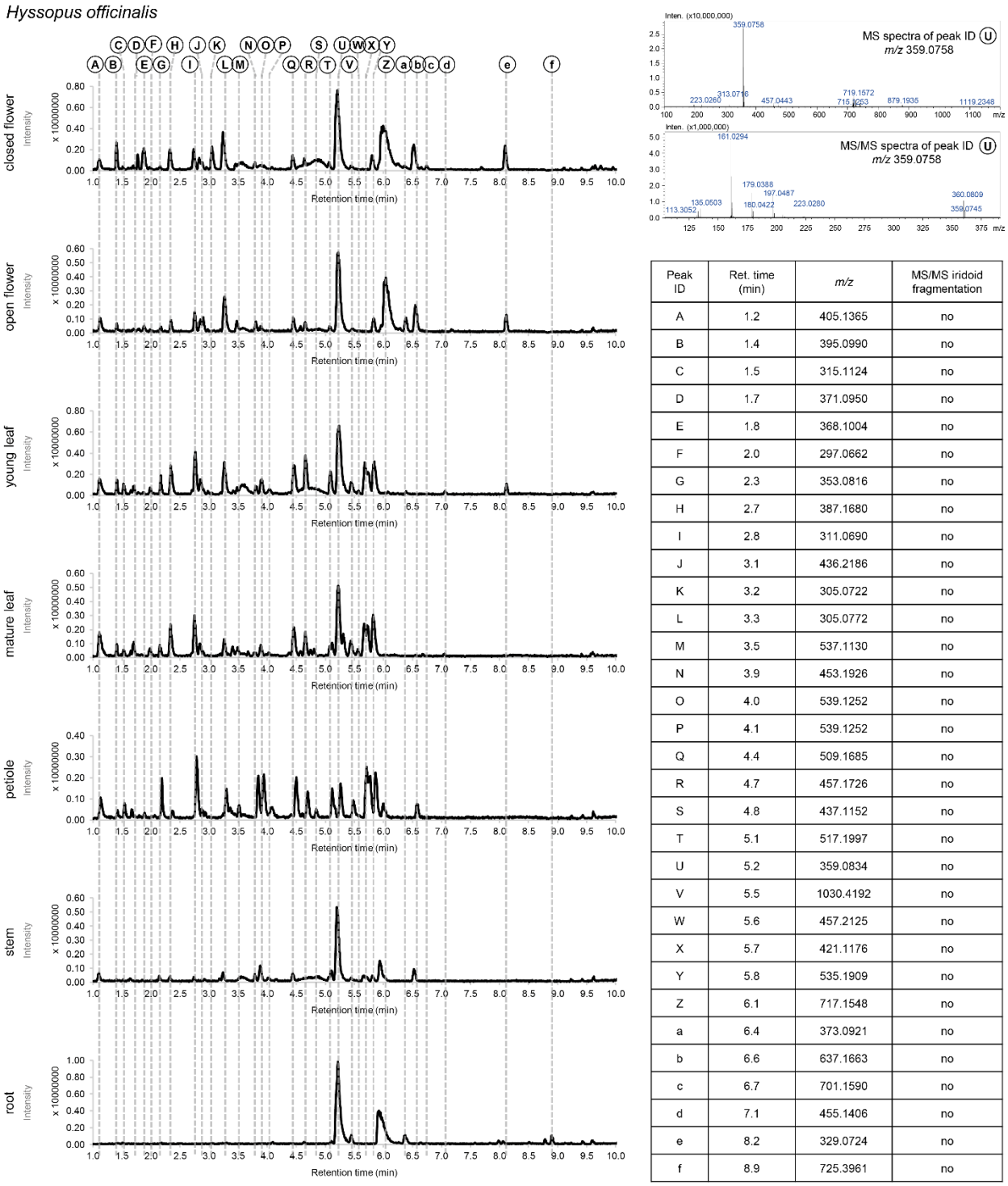
Peak ID	Ret. time (min)	*m/z*	MS/MS iridoid fragmentation
A	1.6	341.0905	no
B	2.7	389.1499	no
C	2.9	436.2167	no
D	3.2	359.1342	yes
E	3.9	563.0982	no
F	4.3	461.0677	no
G	4.6	347.1637	no
H	4.8	447.0955	no
I	4.9	431.0889	no
J	5.1	445.0861	no
K	5.2	359.0744	no
L	6.1	609.1324	no
M	7.2	345.0640	no
N	8.2	427.0702	no
O	9.5	395.2455	no

Fig. S5. LC-MS analysis of *N. cataria* tissues.Untargeted tandem mass LC-MS (-ESI) analysis of *N. cataria* tissues resulted in the tentative identification of an iridoid glucoside 1,5,9-epi-deoxylaoganic acid (D; highlighted in grey). The MS/MS spectra of peak K (*m/z* 359) was different from that of D and observed to be of a typical flavonoid glucoside. Chromatograms are total ion chromatograms (TICs).



Peak ID	Ret. time (min)	*m/z*	MS/MS iridoid fragmentation
A	1.7	341.0908	no
B	2.0	353.0804	no
C	2.4	387.1657	yes
D	2.6	436.2173	no
E	2.7	637.0939	no
F	3.2	359.1352	yes
G	3.4	563.0989	no
H	3.6	436.0935	no
I	4.2	547.1011	no
J	4.3	431.0895	no
K	4.4	447.0962	no
L	4.5	359.0822	no
M	4.6	549.1975	no
N	5.1	717.1530	no
O	5.4	373.0908	no
P	6.0	387.0739	no
Q	6.5	427.0708	no
R	7.1	329.0714	no
S	8.2	313.0782	no
T	9.1	395.2460	no
U	9.2	721.3554	no
V	9.3	721.3554	no
W	9.6	559.3204	no
X	9.8	559.3204	no

Fig. S6. LC-MS analysis of *N. mussinii* tissues.Untargeted tandem mass LC-MS (-ESI) analysis of *N. mussinii* tissues resulted in the tentative identification of two iridoid glucosides (highlighted in grey), 2-O-methyl-1,5,9-epi-deoxyloganic acid (C) and 1,5,9-epi-deoxyloganic acid (F). MS/MS spectra of C and F were similar to *N. cataria* compound D (fig. S5). An isomer of the *N. cataria* flavonoid glucoside K (Fig S5) was tentatively observed in peak L with similar MS/MS spectra. Chromatograms are total ion chromatograms (TICs).



Peak ID	Ret. time (min)	*m/z*	MS/MS iridoid fragmentation
A	1.2	405.1365	no
B	1.4	395.0990	no
C	1.5	315.1124	no
D	1.7	371.0950	no
E	1.8	368.1004	no
F	2.0	297.0662	no
G	2.3	353.0816	no
H	2.7	387.1680	no
I	2.8	311.0690	no
J	3.1	436.2186	no
K	3.2	305.0722	no
L	3.3	305.0772	no
M	3.5	537.1130	no
N	3.9	453.1926	no
O	4.0	539.1252	no
P	4.1	539.1252	no
Q	4.4	509.1685	no
R	4.7	457.1726	no
S	4.8	437.1152	no
T	5.1	517.1997	no
U	5.2	359.0834	no
V	5.5	1030.4192	no
W	5.6	457.2125	no
X	5.7	421.1176	no
Y	5.8	535.1909	no
Z	6.1	717.1548	no
a	6.4	373.0921	no
b	6.6	637.1663	no
c	6.7	701.1590	no
d	7.1	455.1406	no
e	8.2	329.0724	no
f	8.9	725.3961	no

Fig. S7. LC-MS analysis of *H. officinalis* tissues. Untargeted tandem mass LC-MS (-ESI) analysis of *H. officinalis* tissues failed to identify any iridoid glucosides. MS/MS spectra of peak U (*m/z* 359) was observed as a flavonoid glucoside with similar MS/MS to *N. cataria* peak K (Fig. S5), and *N. mussinii* peak L (Fig. S6). Chromatograms are total ion chromatograms (TICs).

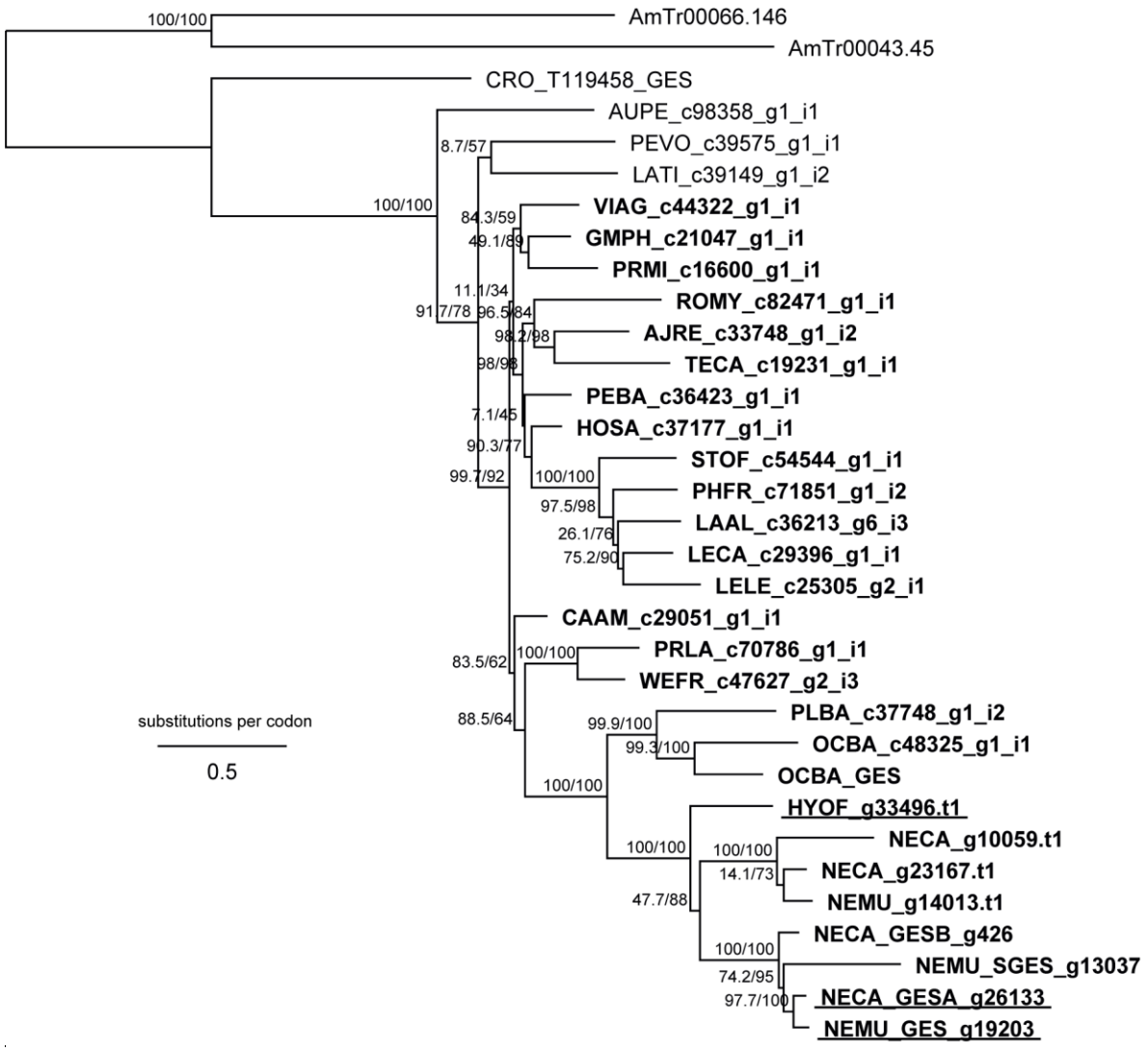


Fig. S8. Geraniol synthase (*GES*) phylogeny. Lamiaceae sequences in bold, assayed sequences underlined. Sequences obtained by orthogroup analysis (OG0011057) and supplemented with sequences from the genomes. *C. roseus* GES (CRO) and *Ocimum basilicum* GES (OCBA_GES) have previously been characterised. Tree inference with iQTree. Support values are SH-aLRT support (%) / ultrafast bootstrap support (%). Outgroups are sequences from *Amborella trichopoda*. Active *GES* genes in *Nepeta* are orthologous to the *C. roseus* gene (CRO_T119458). Species acronyms can be found in Table S5.

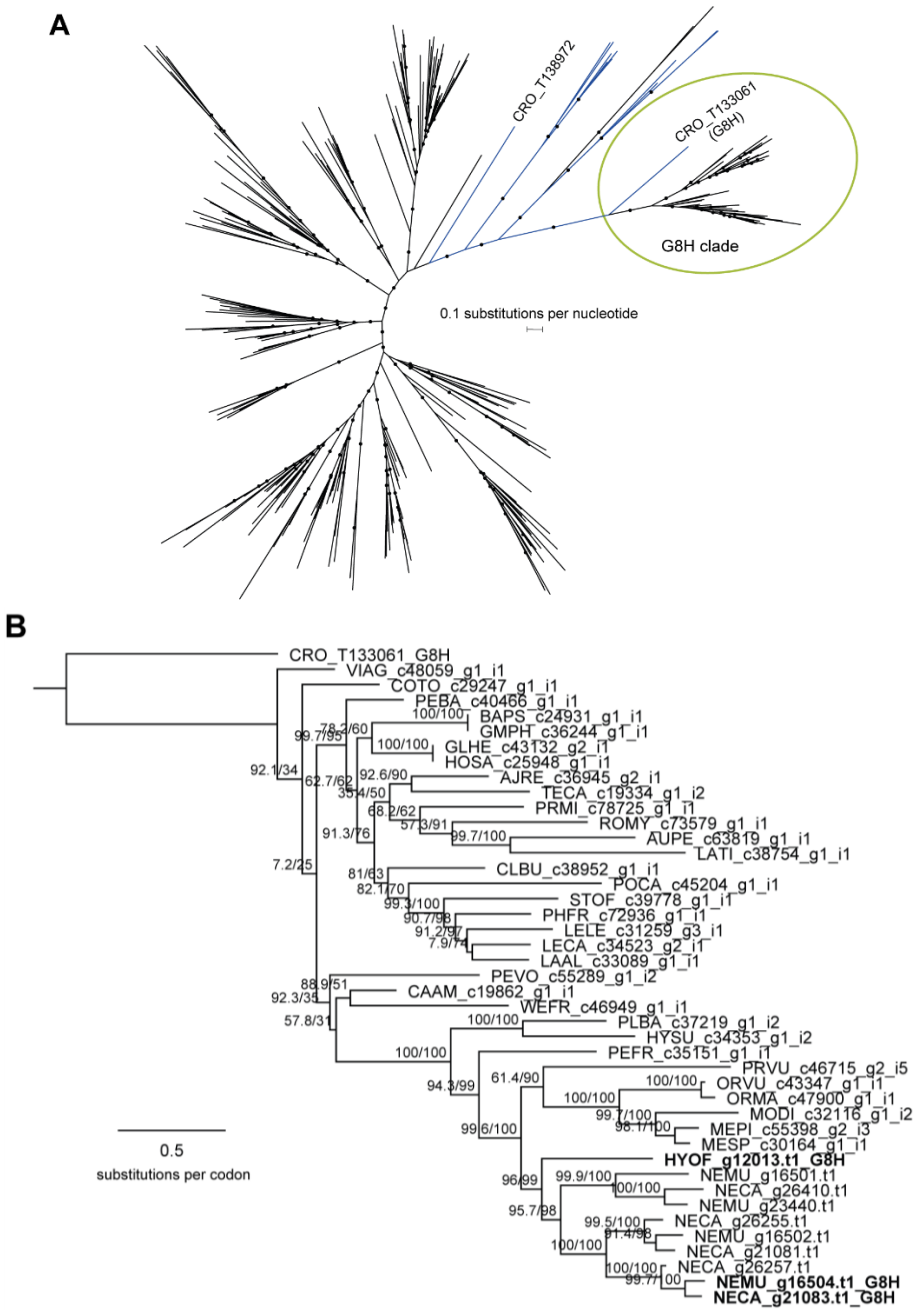


Fig. S9. Geraniol 8-hydroxylase (*G8H*) phylogeny. (**A**) Phylogeny of *G8H* (CYP76B6) orthogroup (OG0000115). Branches in blue represent sequences that are not in the Lamiales (the order containing Lamiaceae). Based on *Arabidopsis* genes, this family contains CYP76B and CYP76C sub-family sequences. Sequences orthologous to the functionally characterised *Catharanthus roseus G8H* (CRO_T133061) form a distinctive clade. Tree inference performed with FastTree and represented using iTOL. Circles represent support values greater than 0.9. (**B**) Phylogeny of *G8H* clade. All sequences, except *C. roseus G8H*, are from Lamiaceae. Assayed sequences in bold. Tree inference performed with iQTree; support values are SH-aLRT support (%) / ultrafast bootstrap support (%). Species acronyms can be found in Table S5.

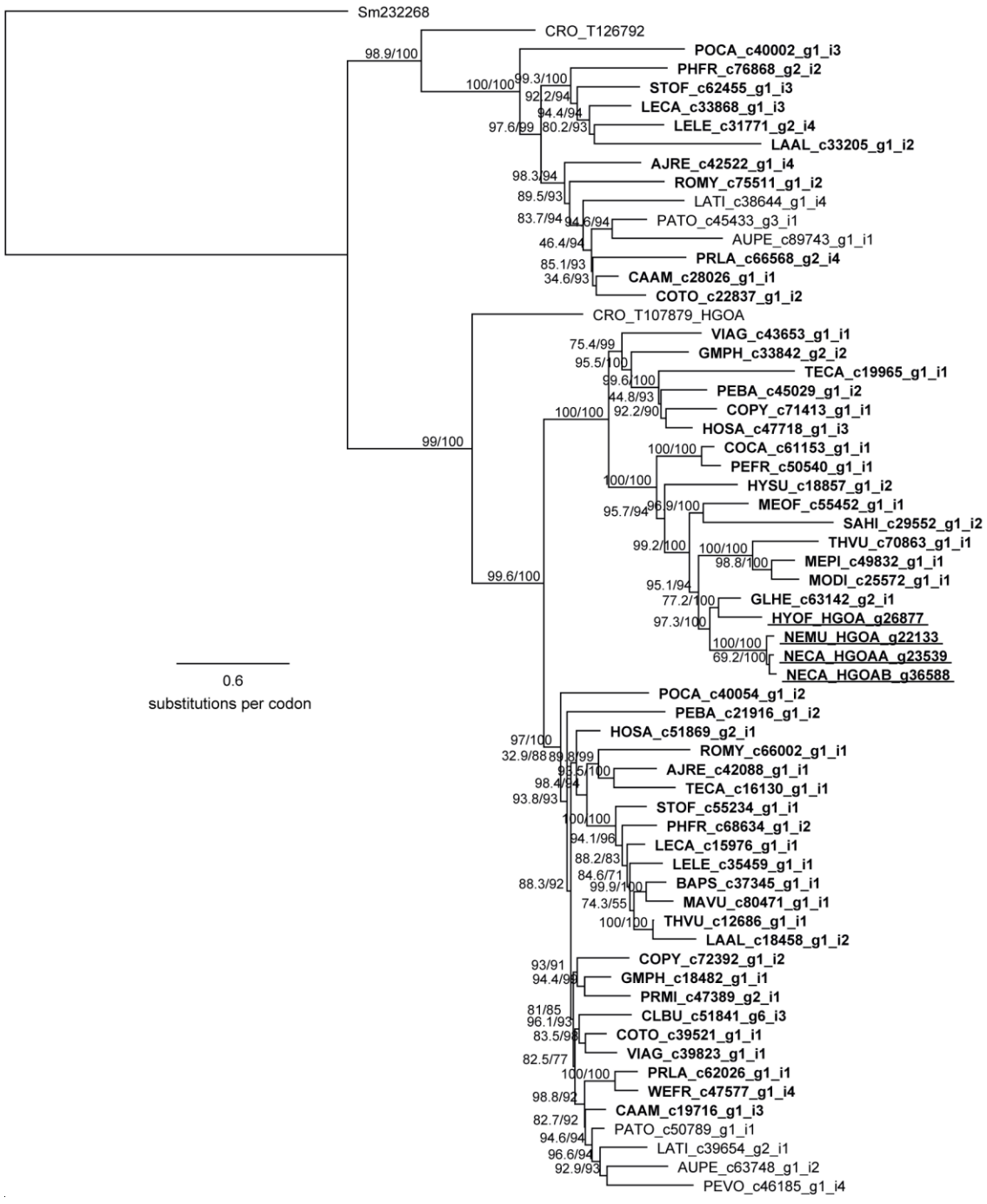


Fig. S10. 8-Hydroxygeraniol oxidoreductase A (*HGOA*) phylogeny. Lamiaceae sequences in bold; assayed sequences underlined. Sequences obtained by orthogroup analysis (OG0003405) and supplemented with sequences from the genomes. *Catharanthus roseus* HGOA has previously been functionally characterised. Tree inference with iQTree. Support values are SH-aLRT support (%) / ultrafast bootstrap support (%). Outgroups is a sequence from *Selaginella moellendorffii*. Active *HGOA* genes in *Nepeta* are orthologous to the *C. roseus* gene (CRO_T107879). Species acronyms can be found in Table S5.

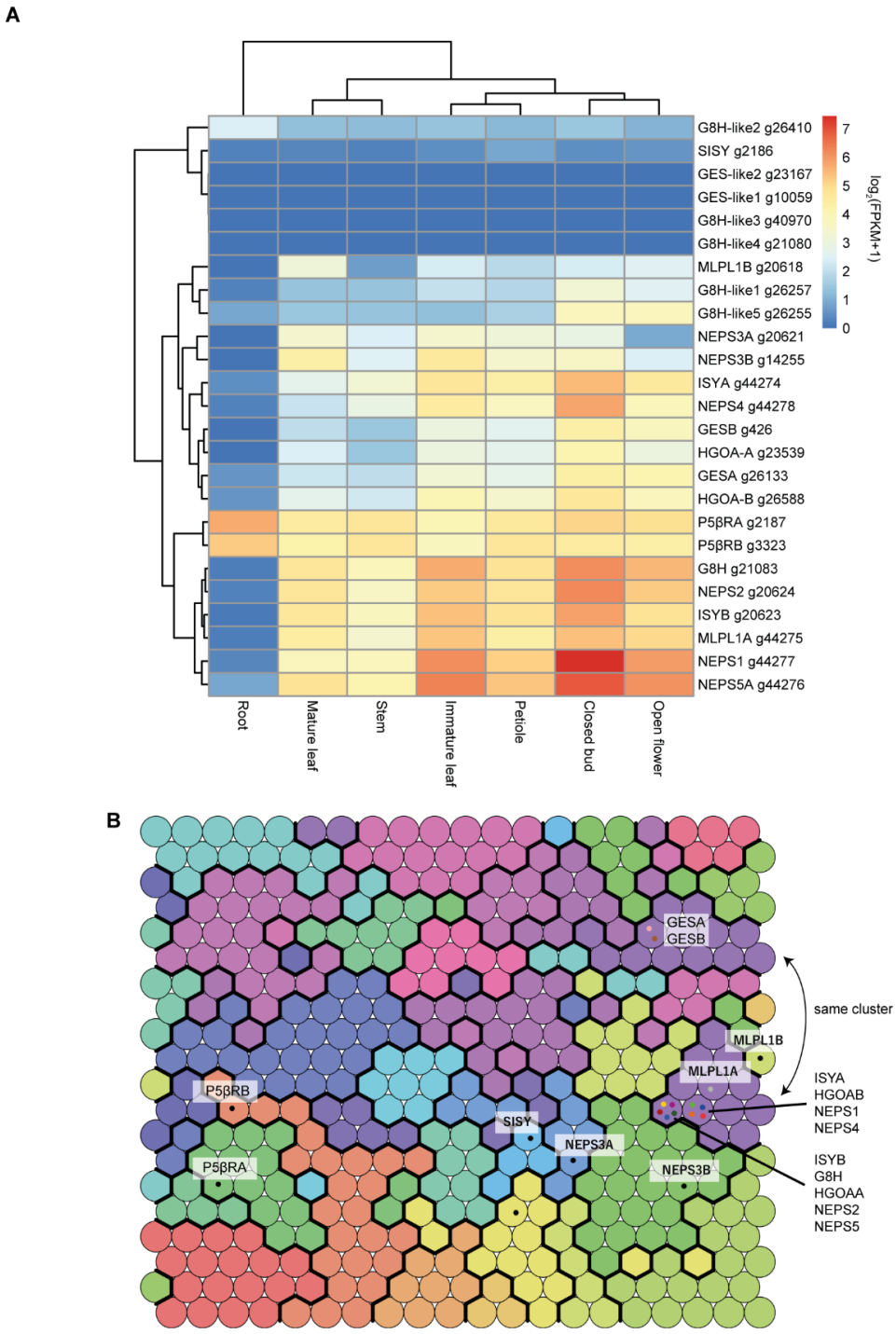


Fig. S11. Expression of iridoid related genes in *N. cataria*. (**A**) Tissue specific gene expression patterns of selected iridoid related genes and homologs, hierarchically clustered. (**B**) Self-organizing map of entire tissue atlas gene expression with iridoid related genes highlighted. Each circle contains a group of genes with very similar expression patterns; these circles are organised into clusters (separated by black lines) with somewhat similar expression patterns. Most of the selected biosynthetic genes are enriched within the same cluster ($p < 1.8 \times 10^{-9}$).

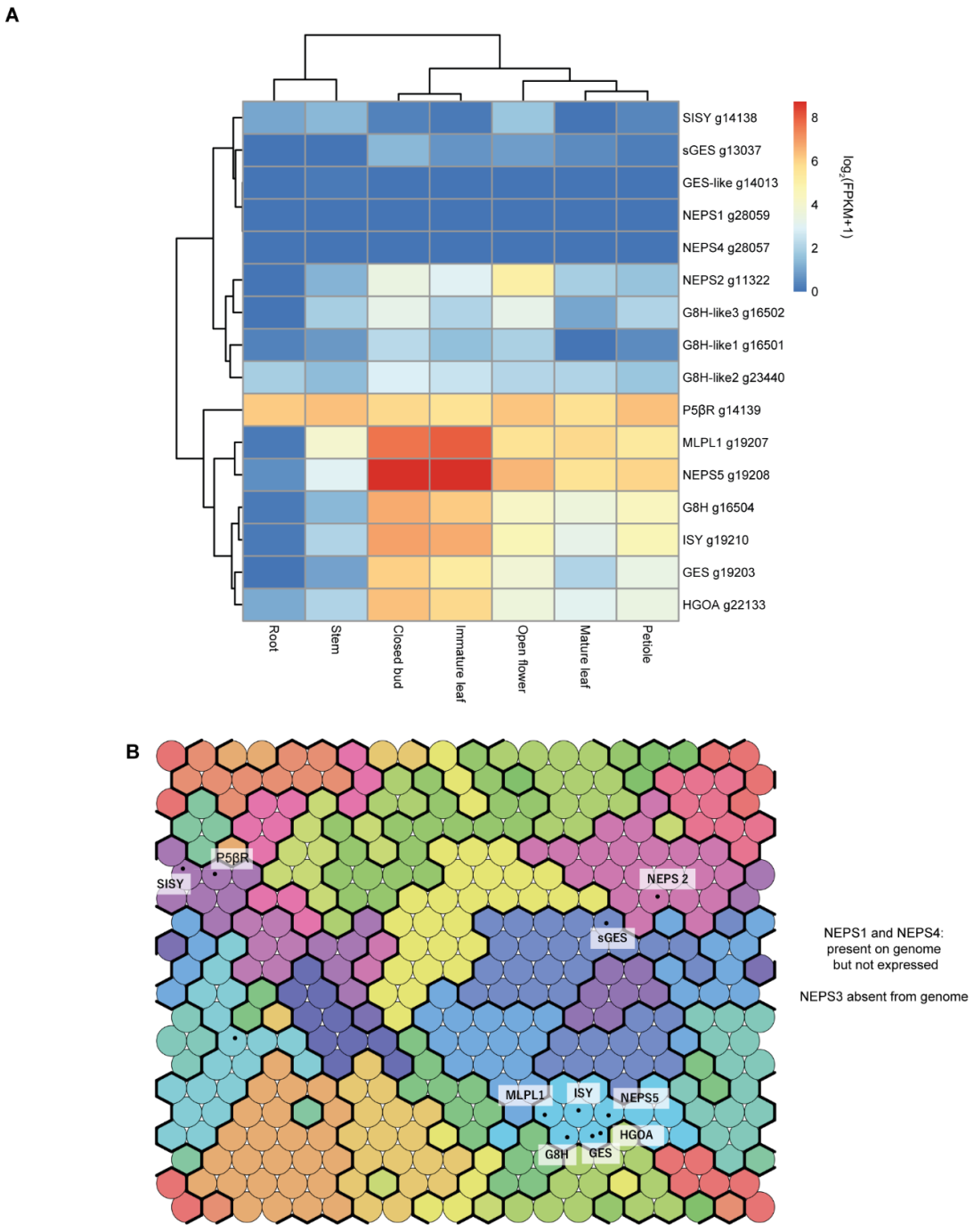


Fig. S12. Expression of iridoid related genes in *N. mussinii*. (**A**) Tissue specific gene expression patterns of selected iridoid related genes and homologs, hierarchically clustered (**B**) Self-organizing map of entire tissue atlas gene expression with iridoid related genes highlighted. Each circle contains a group of genes with very similar expression patterns; these circles are organised into clusters (separated by black lines) with somewhat similar expression patterns. Most of the selected biosynthetic genes are enriched within the same cluster ($p < 3.5 \times 10^{-7}$).

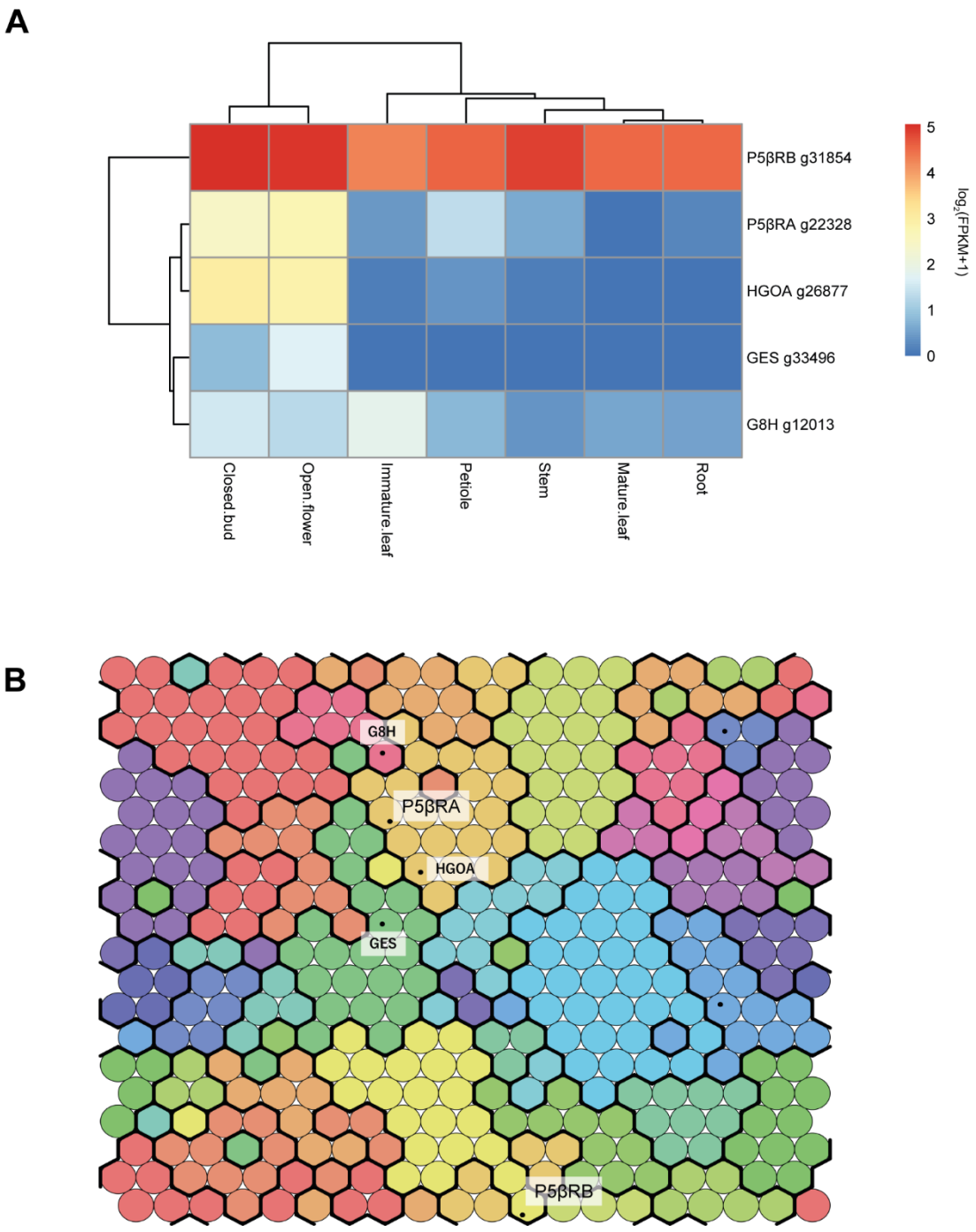


Fig. S13. Expression of iridoid related genes in *H. officinalis*. (**A**) Tissue specific gene expression patterns of selected *GES*, *G8H*, *HGOA* and *PRISE* homologs, hierarchically clustered. (**B**) Self-organizing map of entire tissue atlas gene expression with iridoid related genes highlighted. Each circle contains a group of genes with very similar expression patterns; these circles are organised into clusters (separated by black lines) with somewhat similar expression patterns. None of the selected genes are significantly enriched in a cluster ($p > 0.13$).

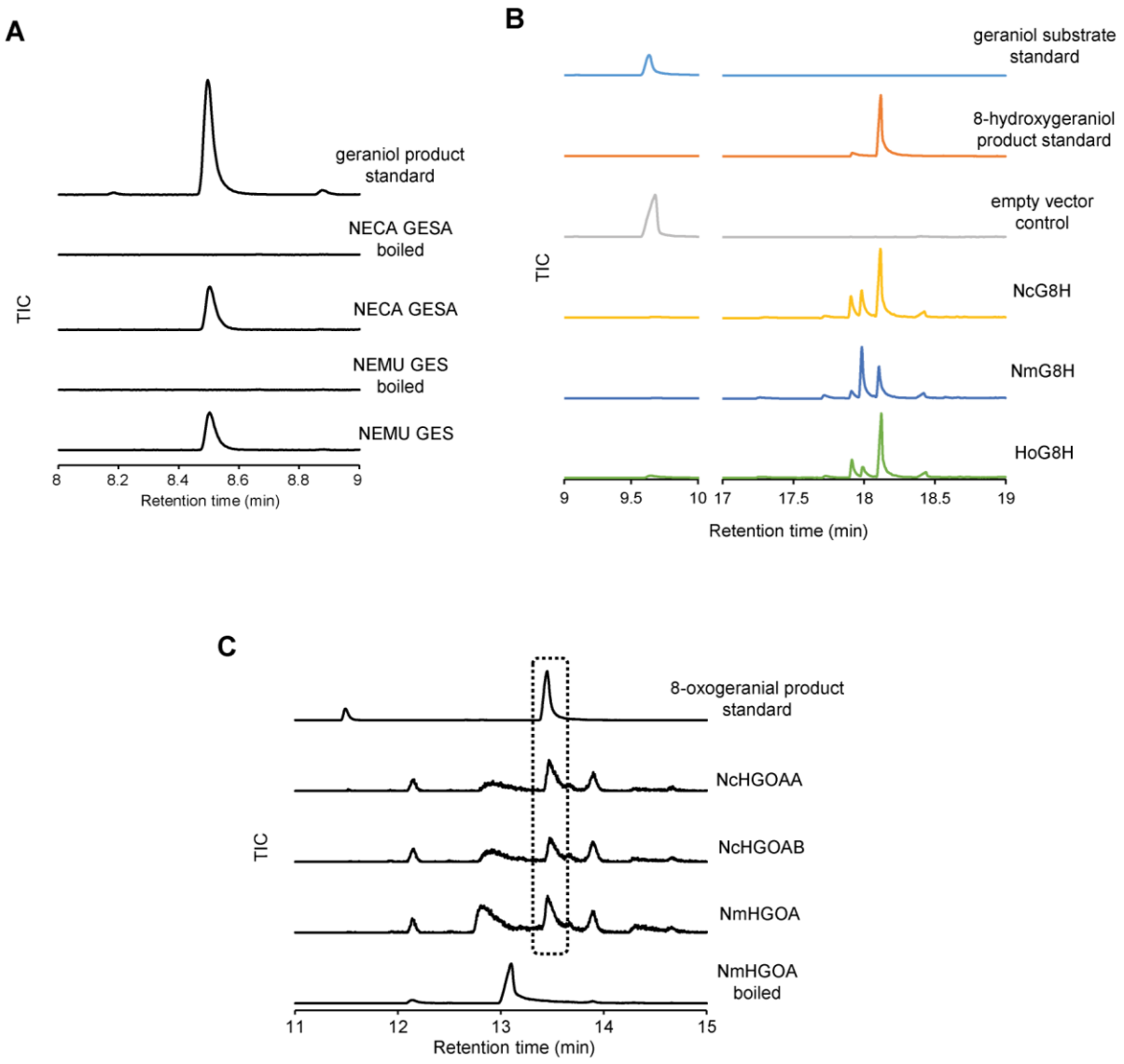


Fig. S14. *In vitro* activity of early iridoid enzymes. Analysis conducted by GC-MS analysis, with the total ion chromatograms (TICs) depicted. Sample intensity of standards and controls have been adjusted for presentation. (**A**) *In vitro* activity of GES candidates. GES from hyssop could not be expressed/purified, so its activity is unknown. (**B**) *In vitro* activity of G8H candidates. Extra product peaks are likely to be over-oxidized products (i.e. aldehydes). (**C**) *In vitro* activity of HGOA candidates. The HGOA reaction involves two NAD^+-controlled oxidation steps. Along with the double oxidised product, there are also two intermediates with single oxidations (8-hydroxygeranial and 8-oxogeraniol) formed in this reaction. Without a downstream step in the reactions, these compounds exist at equilibrium and multiple peaks are observed. HGO from hyssop could not be expressed/purified, so its activity is unknown.

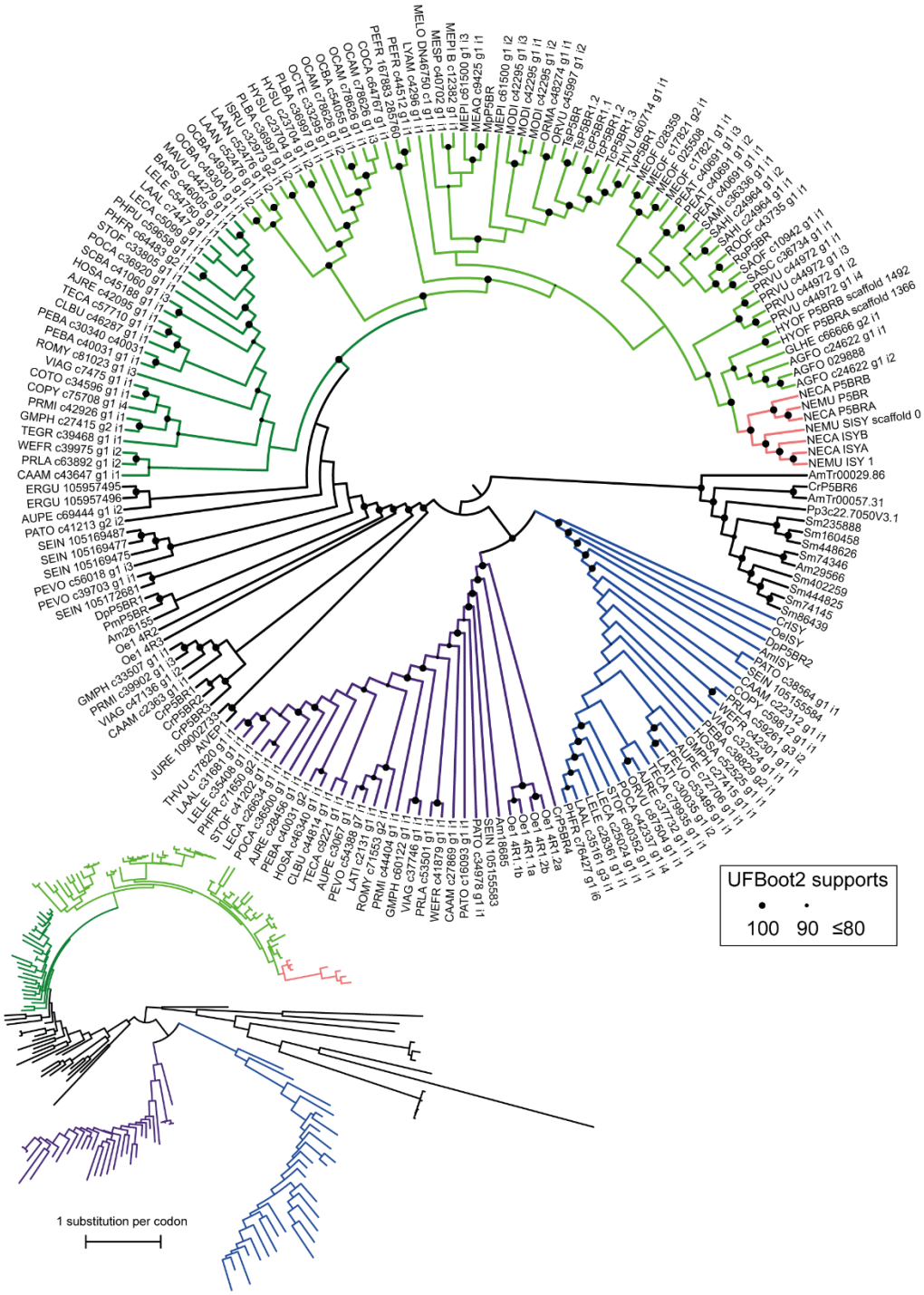


Fig. S15. Phylogeny of the *PRISE* enzyme family. Main cladogram with gene names. Blue branches show the *ISY*-clade; purple the *ISY-paralog* clade; dark green *P5βR* sequences from non-nepetoideae Lamiaceae; light green *P5βR* sequences from nepetoideae; red are sequences from *Nepeta*; all other branches are black. Ultra-fast bootstrap 2 values of range from 80 to 100 are depicted on nodes proportional to black circle size. Inset shows phylogram with branch lengths. Species acronyms can be found in Table S5.

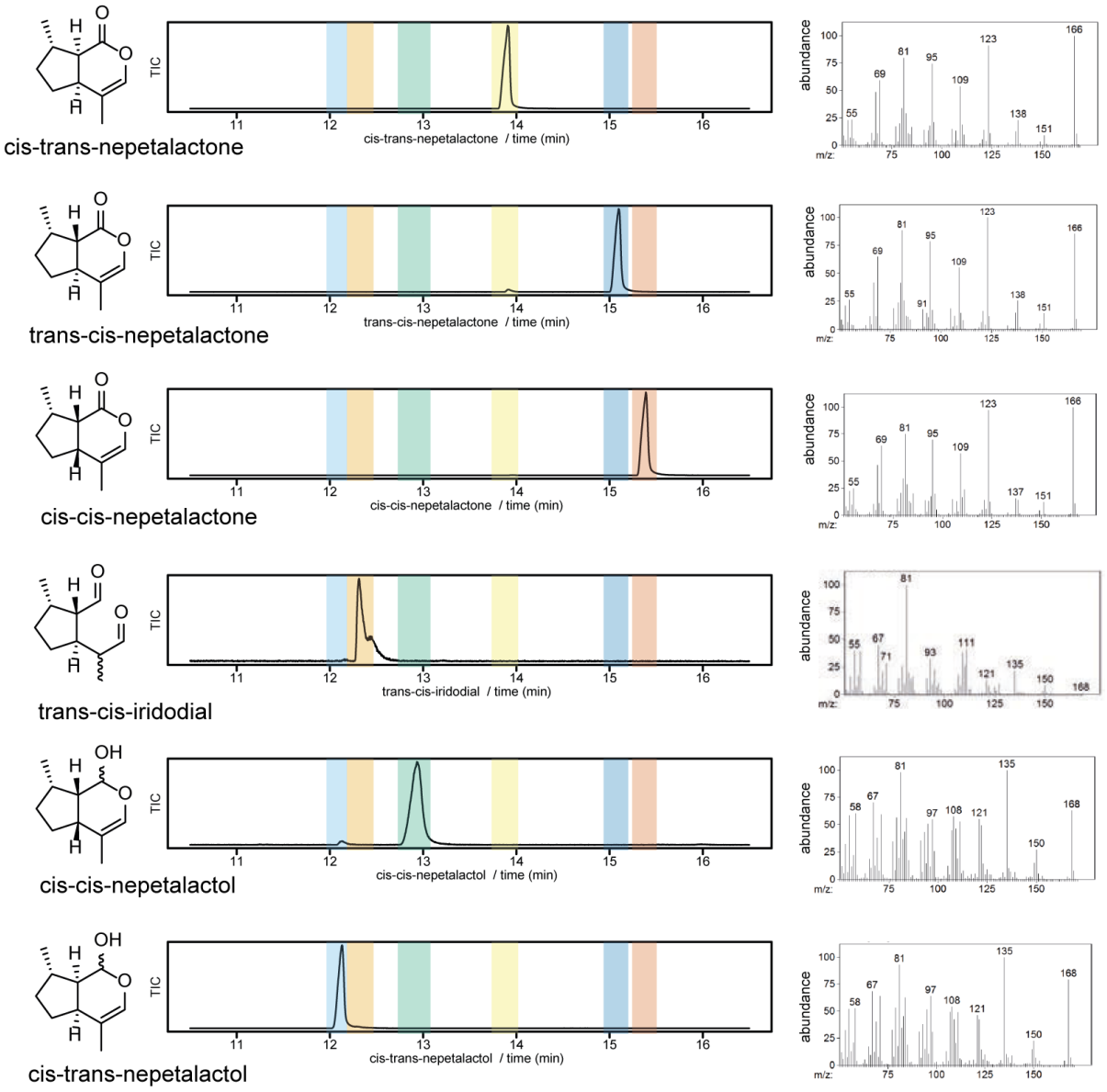


Fig. S16. Product standards for GC-MS assays. Total ion chromatograms (TICs) depicted, with corresponding electron impact (EI) spectra. Colored bars highlight retention times of product standards for assays presented in figs. S17-20. All enzyme product identities were also verified through manual assessment of EI-MS spectra compared to product standards.

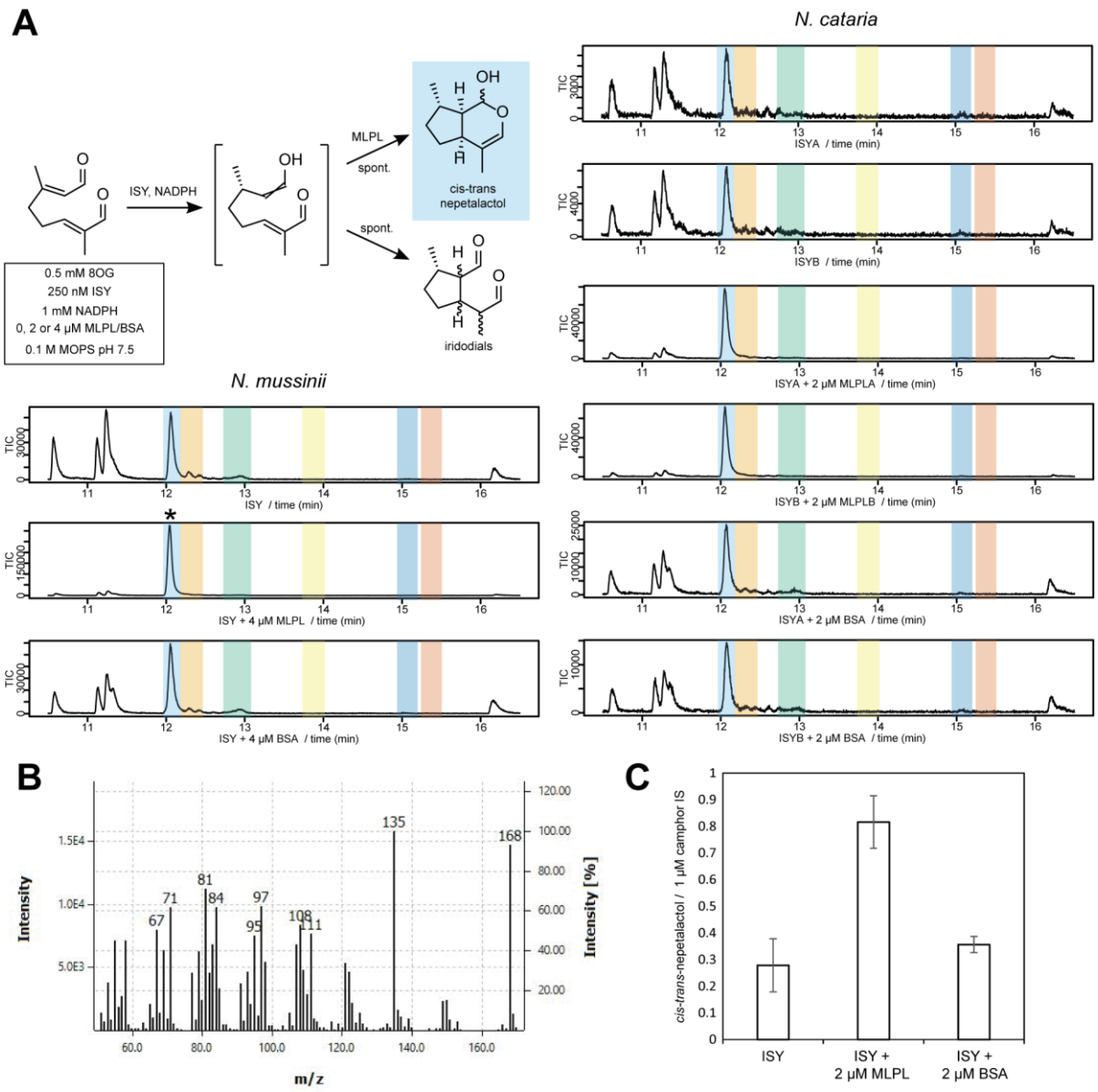


Fig. S17. Cyclisation activity of MLPL. (**A**) Coupled assays with 8-oxogeranial, ISY and MLPL from *Nepeta mussinii* and *N. cataria*. MLPL catalyzes the formation of cis-trans-nepetalactol. Bovine-serum-albumin (BSA) used as a control for MLPL activity. Standards can be found in fig. S16. Assays were performed independently three times with similar outcomes. Total ion chromatograms (TICs) depicted. (**B**) Electron impact (EI) spectrum for *cis-trans*-nepetalactol product from *N. mussinii* ISY + MLPL (4 µM) (starred peak). (**C**) Cis-trans-nepetalactol peak area / camphor internal standard (IS) for *N. mussinii* MLPL (2 µM) reactions. Error bars represent standard errors of three reactions. The ISY+ MLPL peak is significantly larger than ISY alone (t-test, p=0.02) and ISY + BSA (t-test, p=0.01) but ISY peak is not significantly different to ISY + BSA.

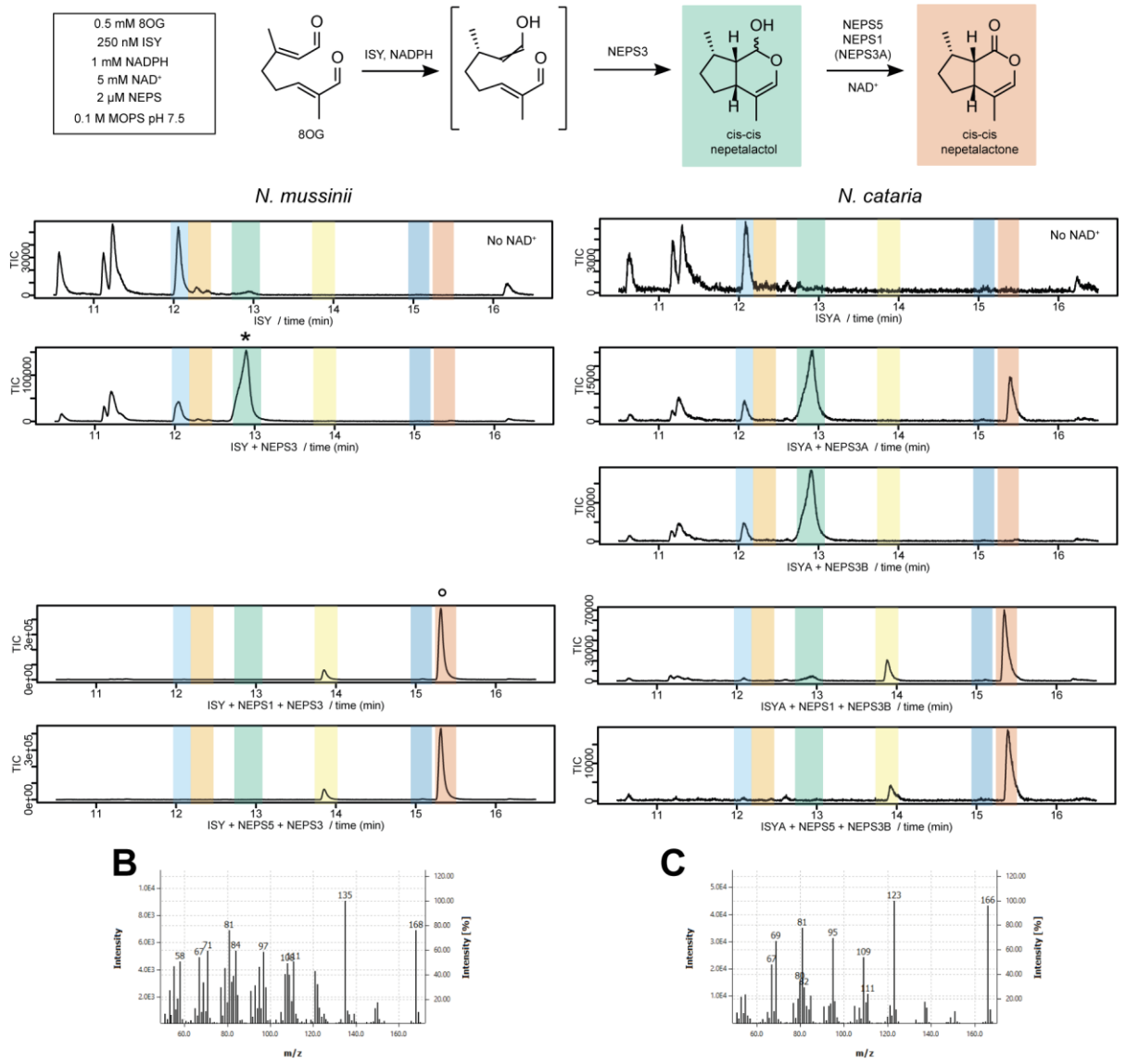


Fig. S18. Enzymatic formation of cis-cis-nepetalactone. (**A**) Coupled assays with 8-oxogeranial, ISY and NEPS from *Nepeta mussinii* and *N. cataria*. Standards can be found in fig. S16. Assays were performed independently three times with similar outcomes. Total ion chromatograms (TICs) depicted. (**B**) Electron impact (EI) spectrum for *cis-cis*-nepetalactol product from *N. mussinii* ISY + NEPS3 (4 µM) (starred peak). (**C**) EI spectrum for *cis-cis*-nepetalactone product from *N. mussinii* ISY + NEPS3 + NEPS1 (peak label ˚).

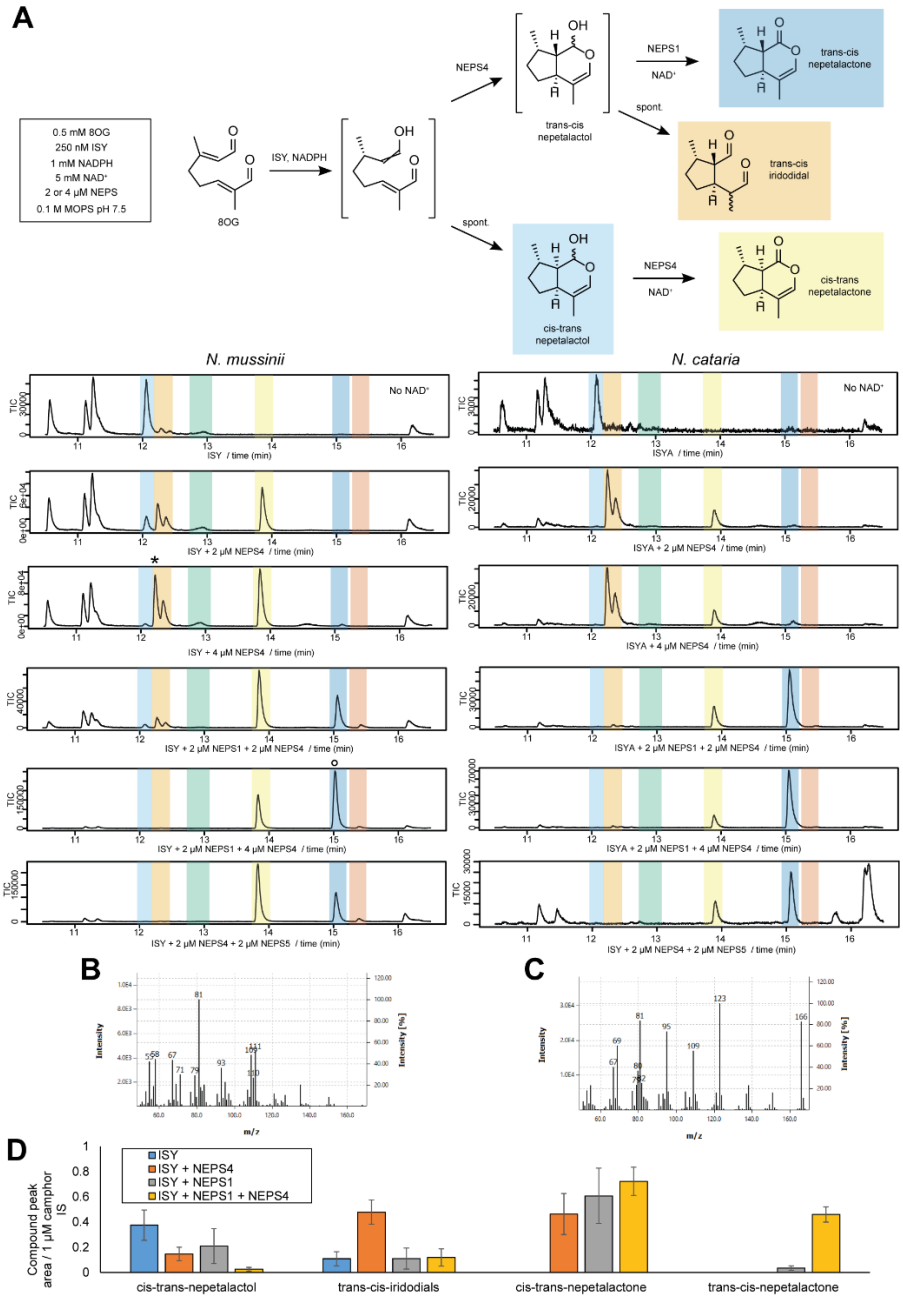


Fig. S19. Enzymatic formation of trans-cis-nepetalactone. (**A**) Coupled assays with 8-oxogeranial, ISY and NEPS from *Nepeta mussinii* and *N. cataria*. Standards can be found in fig. S16. Assays were performed independently three times with similar outcomes. (**B**) Electron impact (EI) spectrum for *trans-cis*-iridodial product from *N. mussinii* ISY + NEPS4 (4 µM) (starred peak). (**C**) EI spectrum for *trans-cis*-nepetalactone product from *N. mussinii* ISY + NEPS4 (4 µM) + NEPS1 (peak label ˚). (**D**) Compound peak areas relative to an internal standard for *N. mussinii* enzyme assays. Data points are means of 4 independent measurements and show standard errors.

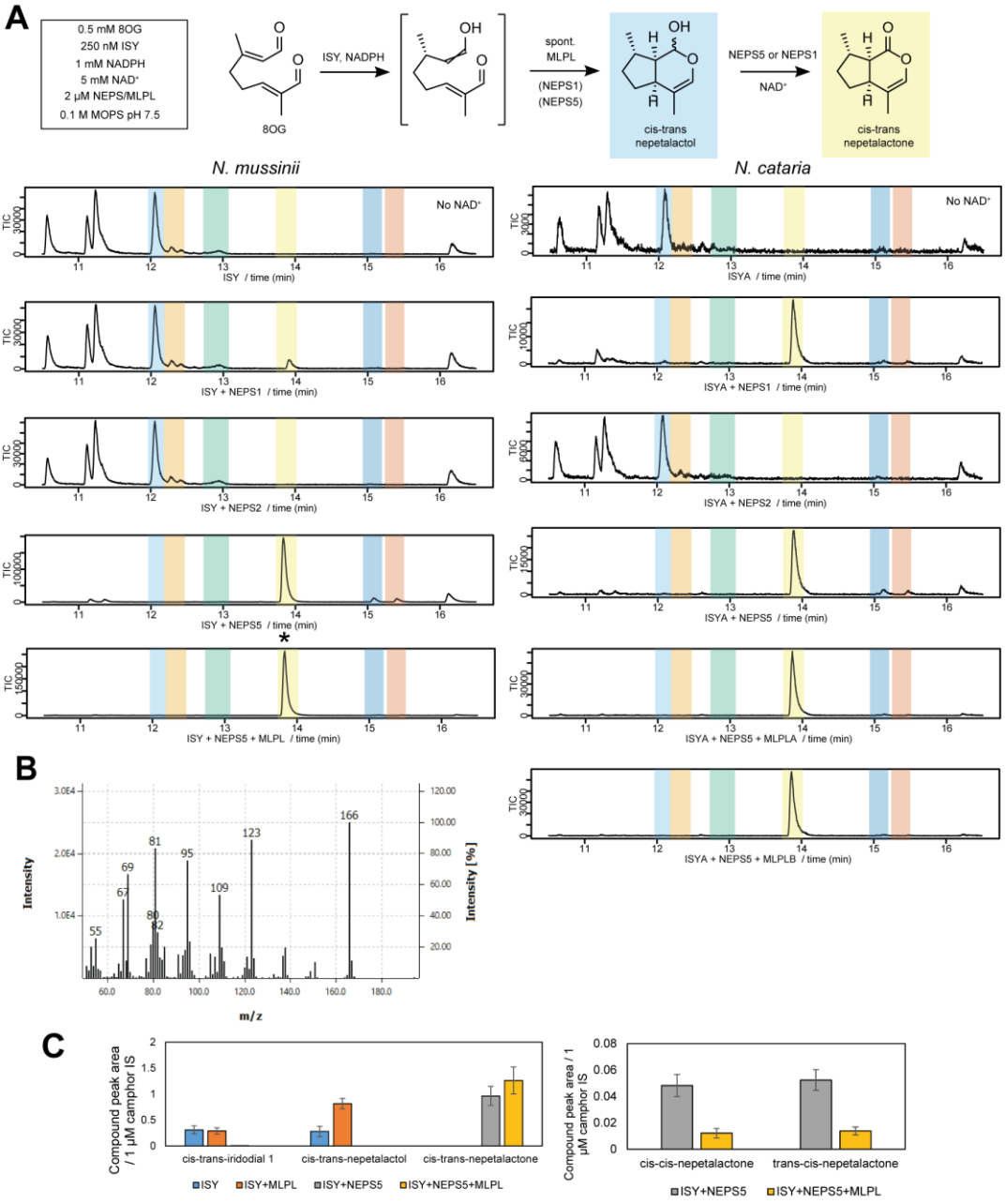


Fig. S20. Enzymatic formation of cis-trans-nepetalactone. (**A**) Coupled assays with 8-oxogeranial, ISY and NEPS from *Nepeta mussinii* and *N. cataria*. Standards can be found in fig. S16. Assays were performed independently three times with similar outcomes. Total ion chromatograms (TICs) depicted. (**B**) Electron impact (EI) spectrum for *cis-trans*-nepetalactone product from *N. mussinii* ISY + MLPL + NEPS5 (starred peak). (**C**) Compound peak areas relative to an internal standard for *N. mussinii* enzyme assays. Data points are means of three independent measurements and show standard errors. For ISY+NEPS5 reactions, addition of MLPL slightly increases cis-trans-nepetalactone, and

reduces other nepetalactone isomers, indicating it promotes cis-trans-nepetalactone at the expense of other isomers.

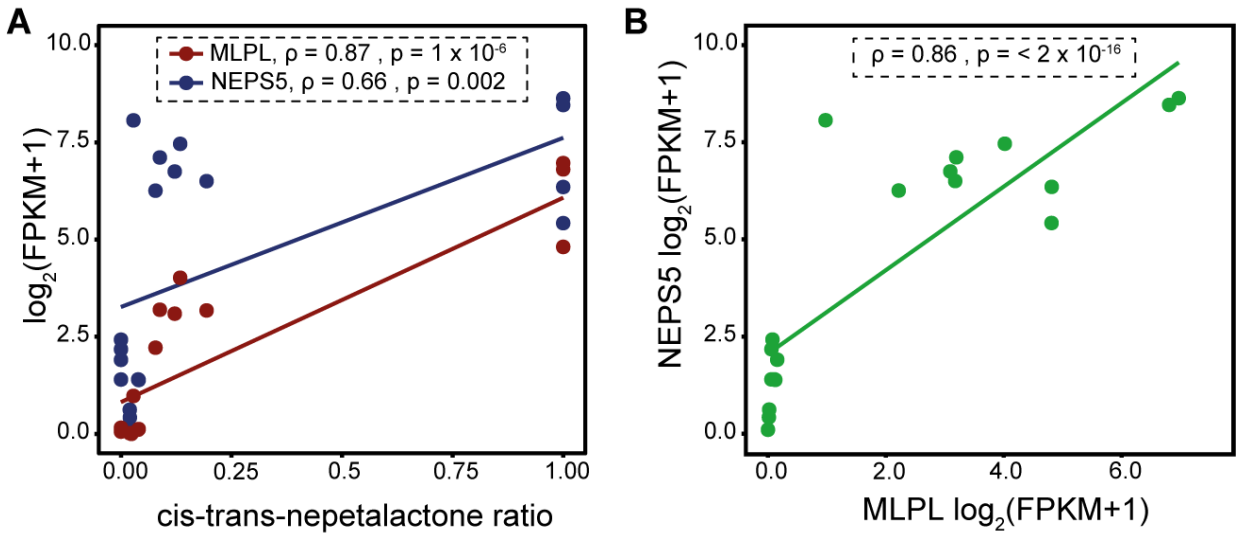


Fig. S21 *NEPS5, MLPL* and cis-trans-nepetalactone correlation in *N. mussinii*. (**A**) Expression of *NEPS5* and *MLPL* genes compared to cis-trans-nepetalactone ratios in plant. Straight lines are linear models. Spearman's rank correlation results are inset. *MLPL* and *NEPS5* expression levels are positively correlated with cis-trans-nepetalactone ratios, though the correlation is greater and more significant for *MLPL*. (**B**) Correlation between *NEPS5* and *MLPL* gene expression. Straight line is linear model. Spearman's rank correlation result is inset: the expression of the two genes are very significantly positively correlated. Overall this figure indicates that *MLPL*, *NEPS5* and cis-trans-nepetalactone accumulation are significantly correlated.

Fig. S22. Sequence alignment of *PRISE* genes from *Nepeta* and hyssop. Continued on next page.

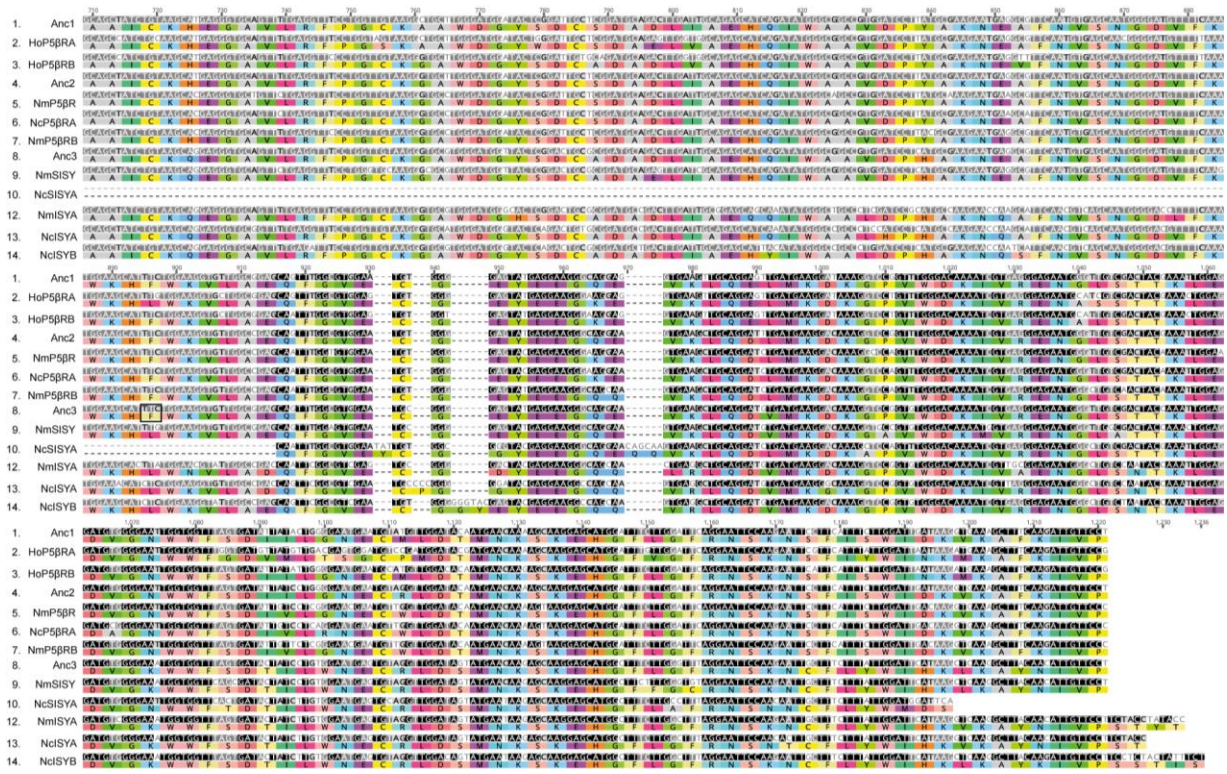


Fig. S22. Sequence alignment of *PRISE* genes from *Nepeta* and hyssop. Codon alignment of *PRISE* genes from *Nepeta* and hyssop. The delta (Δ) represents the N-terminal truncation for recombinant proteins. The *NcSISY* pseudogenes were absent from phylogenetic analyses as they destabilized the tree topology. With the exception of the ancestral genes, and *SISYs* which could not be cloned, all sequences here are the cloned sequence verified genes used in assays. The phylogenies (figs. S15 and S23) were constructed with genomic sequences from hyssop. The boxed codon in Anc3 shows the location of the equivocal TTG-TTC (see methods for details)

Fig. S23. Phylogeny of sub-sampled *PRISE* enzyme family. Tree phylogeny recalculated to focus on *P5βR* clade in Lamiaceae, with the Lamiaceae most recent common ancestor (MRCA) node labelled. Support values (UFBoot2) are labelled. This tree was used for positive selection analysis and ancestral reconstruction analysis. Branches tested for positive selection are labelled; nodes reconstructed are labelled. Species acronyms can be found in Table S5.

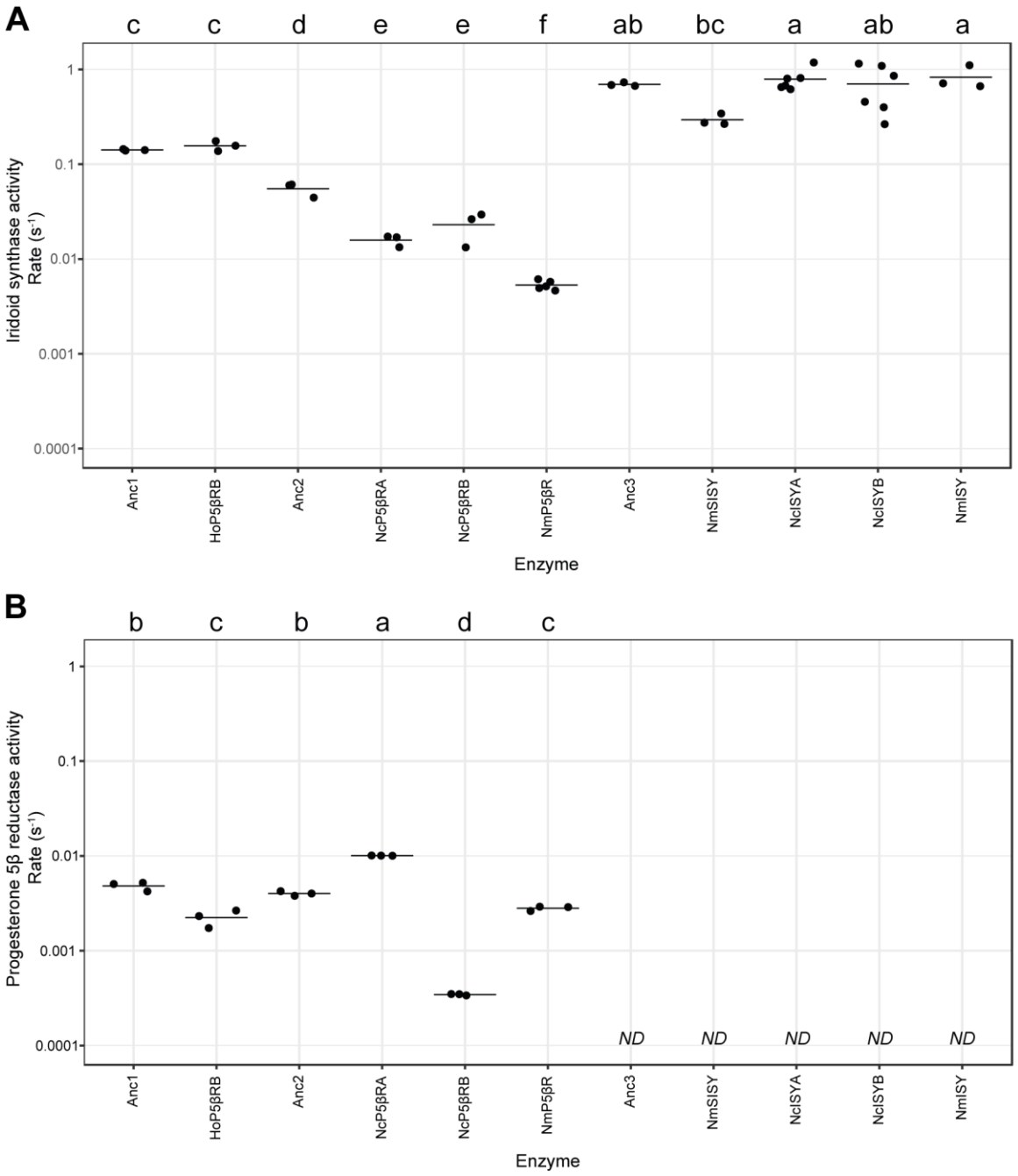


Fig. S24. Activities of PRISE enzymes from *Nepeta* and hyssop. In vitro enzyme activities of extant PRISEs from hyssop and *Nepeta* and reconstructed ancestral enzymes. Note logarithmic scale. Letters above plots represent groups of means with statistically significant (p<0.05) differences, as determined by analysis of variation and Tukey's post-hoc test. (**A**) ISY activity: reduction of 8-oxogeranial, measured spectrophotometrically by tracking depletion of NADPH (change in absorbance at 340 nm). (**B**) Progesterone 5β-reductase activity: reduction of progesterone, measured using a GC-MS timecourse.

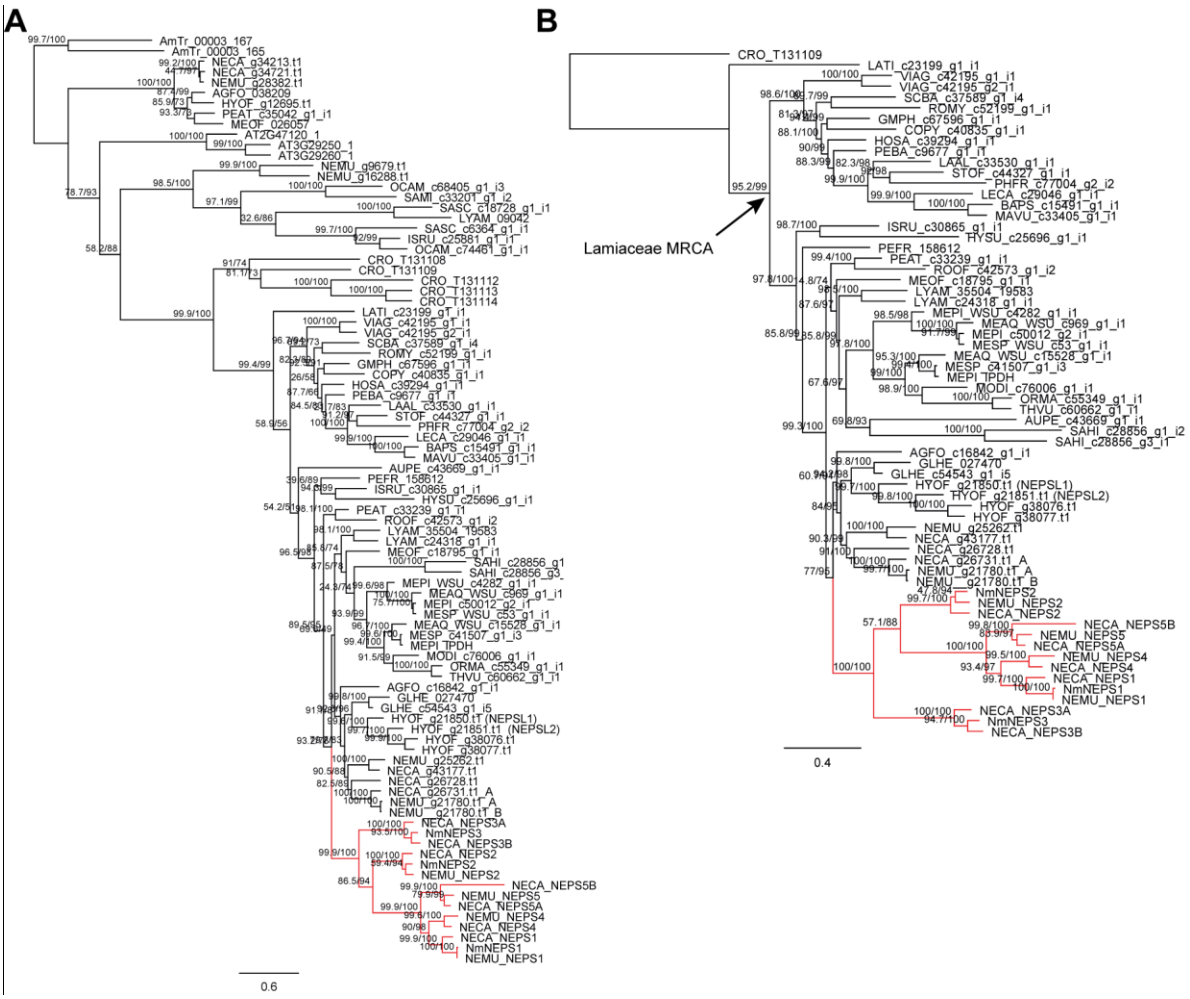


Fig. S25. *NEPS* phylogeny.(**A**) Phylogram of whole *NEPS-like* OG0004576 orthogroup. (**B**) Phylogram of selected sequences from OG0004576 showing a single Lamiaceae clade, with Lamiaceae most recent common ancestor (MRCA) node labelled. The *NEPS* clade is highlighted in red. Species acronyms can be found in Table S5.

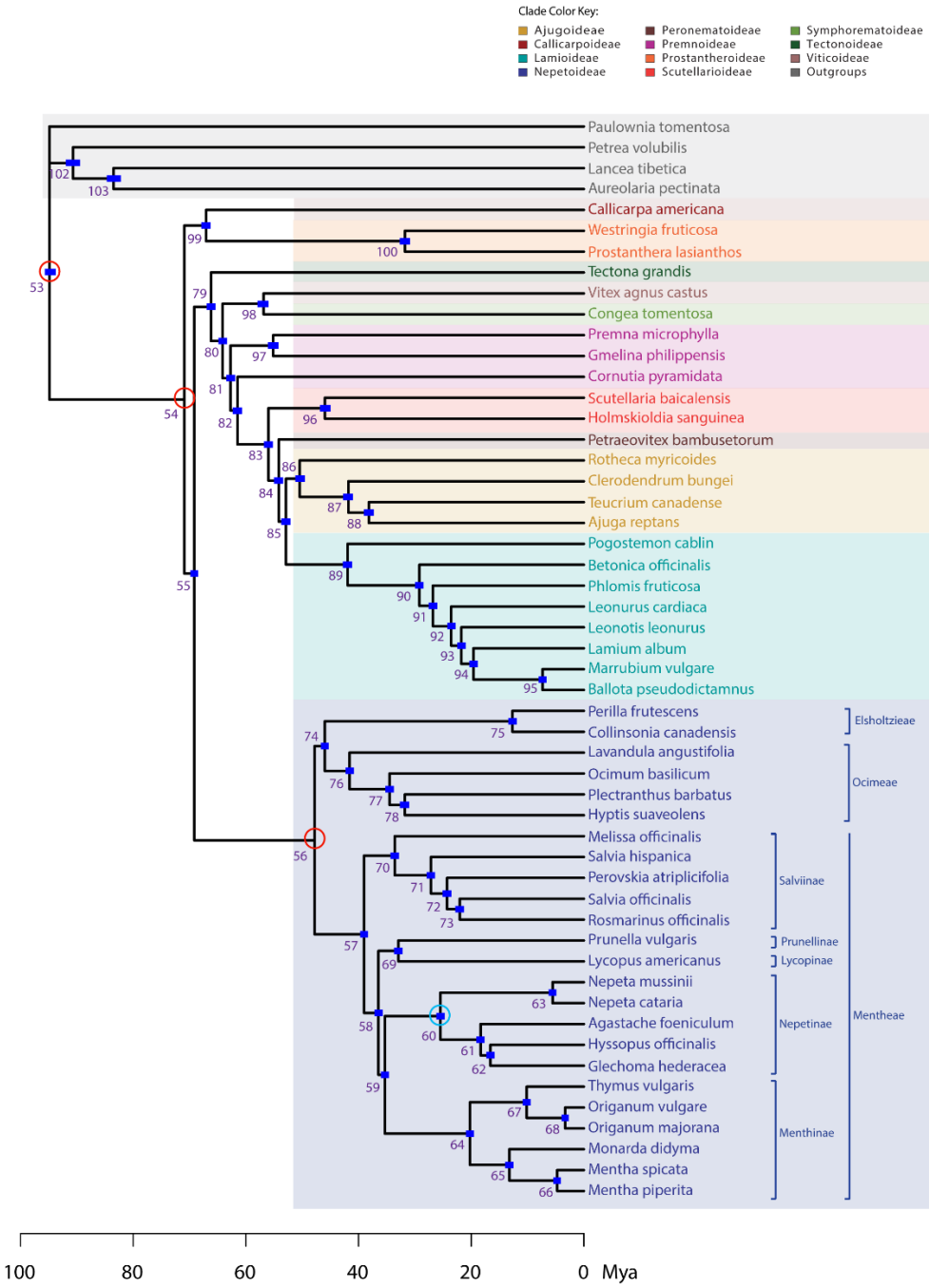


Fig. S26. Species chronogram. Divergence times (millions of years ago, Mya) were estimated assuming a relaxed molecular clock, as implemented in treePL with a penalized likelihood approach. Shown here is a summary of estimated ages from 100 dated bootstrap trees, with blue bars showing 95% confidence intervals. Open red circles denote nodes with age calibrations and an open blue circle indicates the confidence interval extracted for use as a secondary calibration in the PRISE and NEPS divergence times analyses. Subfamilial classifications for all species are color coded according to the key provided, and tribal and subtribal are indicated for Nepetoideae (*75*, *76*).

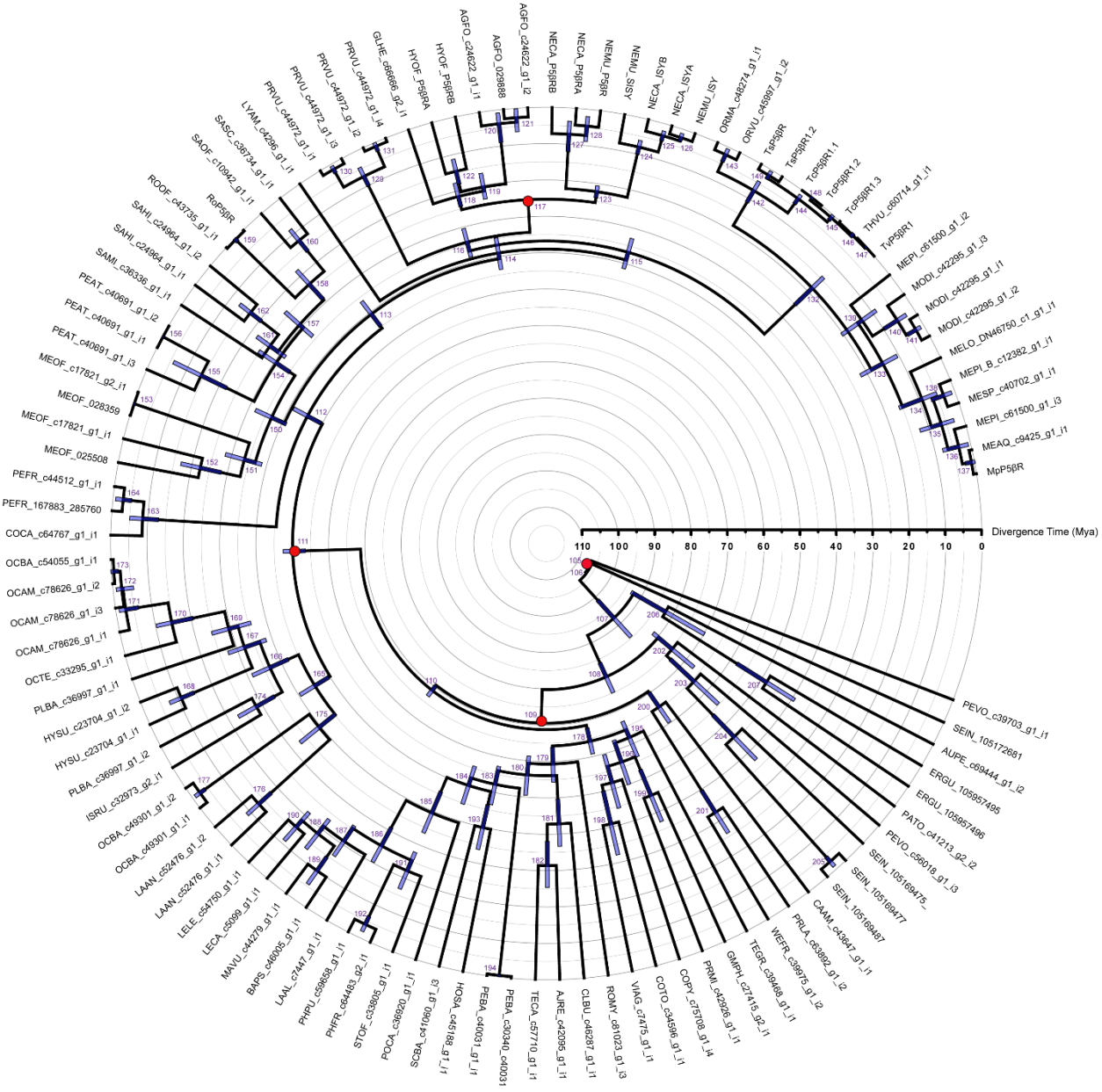


Fig. S27. Chronogram of Progesterone 5β-reductase/iridoid synthase (*PRISE*) evolution. Divergence times (millions of years ago, Mya) were estimated assuming a relaxed molecular clock, as implemented in treePL with a penalized likelihood approach. Shown here is a summary of estimated ages from 1000 dated bootstrap trees, with blue bars showing 95% confidence intervals. Red circles denote calibrated nodes. Species acronyms can be found in Table S5.

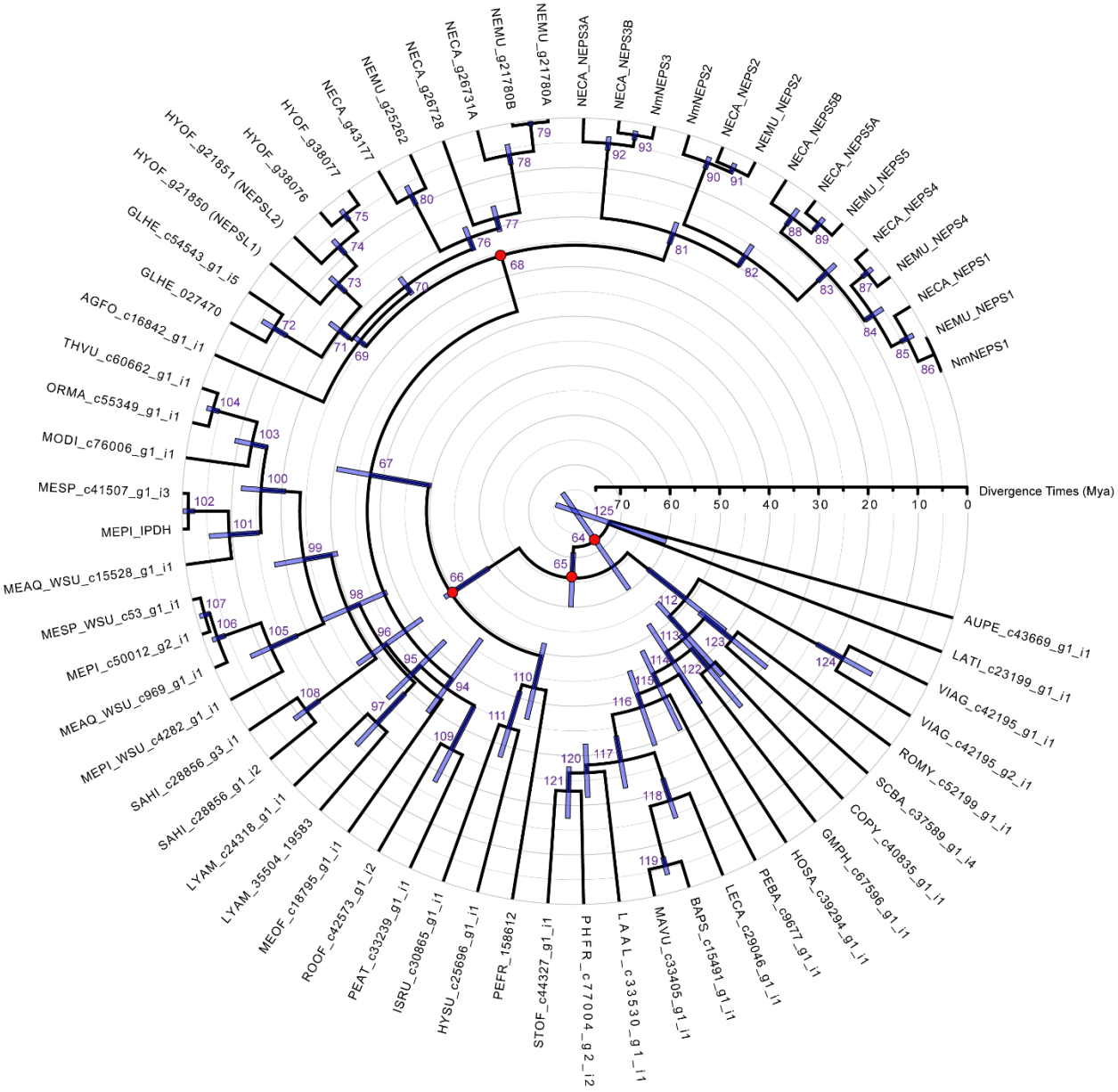


Fig. S28. Chronogram of *NEPS* and *NEPS-like* evolution. Divergence times (millions of years ago, Mya) were estimated assuming a relaxed molecular clock, as implemented in treePL with a penalized likelihood approach. Shown here is a summary of estimated ages from 1000 dated bootstrap trees, with blue bars showing 95% confidence intervals. Red circles denote calibrated nodes. Species acronyms can be found in Table S5.

.

Table S1. Libraries used in assembly and annotation of *H. officinalis* and *N. cataria*

Library ID	Library Type	Read Length (nt)	Type of Sequencing	Library Size (bp)	Total Reads	Bioproject	Biosample
Hyssopus officinalis							
HYS_AA	Whole genome sequencing	150	Paired end	362	92,937,185	PRJNA529676	SAMN11466421
HYS_AF	Mate pair 2kb	150	Paired end	778	39,475,350	PRJNA529676	SAMN11466421
HYS_AH	Mate pair 4kb	150	Paired end	594	50,761,467	PRJNA529676	SAMN11466421
HYS_AI	Mate pair 6kb	150	Paired end	548	50,176,401	PRJNA529676	SAMN11466421
HAY_AJ	Mate pair 10kb	150	Paired end	570	44,717,692	PRJNA529676	SAMN11466421
HYS_AB_Run1	Pacific Biosciences Selection >20kb	-	-	-	3.4Gb	PRJNA529676	SAMN11467189
HYS_AB_Run2	Pacific Biosciences Selection >15kb	-	-	-	6Gb	PRJNA529676	SAMN11467189
MINT_FM	RNAseq: closed flower buds	150	Paired end	252	40,282,518	PRJNA529676	SAMN11466428
MINT_FH	RNAseq: immature leaf	150	Paired end	259	42,372,404	PRJNA529676	SAMN11466423
MINT_FG	RNAseq: mature leaf	150	Paired end	264	43,763,710	PRJNA529676	SAMN11466422
MINT_FL	RNAseq: open flower buds	150	Paired end	250	42,485,284	PRJNA529676	SAMN11466427
MINT_FJ	RNAseq: petiole	150	Paired end	263	42,850,202	PRJNA529676	SAMN11466425
MINT_FK	RNAseq: root	150	Paired end	254	52,057,979	PRJNA529676	SAMN11466426
MINT_FI	RNAseq: stem	150	Paired end	257	44,804,481	PRJNA529676	SAMN11466424
Nepeta cataria							
MINT_EK	Whole genome sequencing	150	Paired end	385	78,581,091	PRJNA359989	SAMN06200307
MINT_EL	Whole genome sequencing	150	Paired end	388	104,966,807	PRJNA557218	SAMN12391618
MINT_EB	Mate pair 4kb	150	Paired end	607	39,971,767	PRJNA557218	SAMN12391618
MINT_EC	Mate pair 7kb	150	Paired end	595	44,574,956	PRJNA557218	SAMN12391618
MINT_ED	Mate pair 10kb	150	Paired end	632	43,738,395	PRJNA557218	SAMN12391618
MINT_EN	RNAseq: closed flower buds	150	Paired end	290	29,523,126	PRJNA529674	SAMN11475949
MINT_EO	RNAseq: immature leaf	150	Paired end	349	30,998,008	PRJNA529674	SAMN11475950
MINT_EP	RNAseq: mature leaf	150	Paired end	332	31,767,798	PRJNA529674	SAMN11475951
MINT_EQ	RNAseq: open flowers	150	Paired end	240	27,980,687	PRJNA529674	SAMN11475952
MINT_ER	RNAseq: petiole	150	Paired end	291	39,296,506	PRJNA529674	SAMN11475953
MINT_ES	RNAseq: root	150	Paired end	362	38,808,562	PRJNA529674	SAMN11475954
MINT_ET	RNAseq: stem	150	Paired end	302	39,126,095	PRJNA529674	SAMN11475955

Table S2. Libraries used in assembly and annotation of *N. mussinii*

Library ID	Library Type	Read Length (nt)	Type of Sequencing	Library Size (bp)	Total Reads	Bioproject	Biosample
Nepeta mussinii (reference genotype)							
MINT_EI	Whole genome sequencing	150	Paired end	381	42,485,828	PRJNA359989	SAMN06200308
MINT_EJ	Whole genome sequencing	150	Paired end	367	49,147,531	PRJNA557218	SAMN12391619
MINT_CI	Whole genome sequencing	150	Paired end	387	15,987,403	PRJNA359989	SAMN06200308
MINT_EE	Mate pair 2kb	150	Paired end	631	46,999,157	PRJNA557218	SAMN12391619
MINT_EF	Mate pair 4kb	150	Paired end	616	37,746,807	PRJNA557218	SAMN12391619
MINT_EG	Mate pair 7kb	150	Paired end	578	46,191,916	PRJNA557218	SAMN12391619
MINT_EH	Mate pair 10kb	150	Paired end	575	51,733,959	PRJNA557218	SAMN12391619
MINT_EU	RNAseq: closed flower buds	150	Paired end	274	28,507,931	PRJNA529674	SAMN11475956
MINT_EV	RNAseq: immature leaf	150	Paired end	311	32,035,293	PRJNA529674	SAMN11475957
MINT_EW	RNAseq: mature leaf	150	Paired end	325	35,098,168	PRJNA529674	SAMN11475958
MINT_EX	RNAseq: open flowers	150	Paired end	325	36,437,355	PRJNA529674	SAMN11475959
MINT_EY	RNAseq: petiole	150	Paired end	339	39,058,249	PRJNA529674	SAMN11475960
MINT_EZ	RNAseq: root	150	Paired end	326	41,021,013	PRJNA529674	SAMN11475961
MINT_FA	RNAseq: stem	150	Paired end	310	36,866,609	PRJNA529674	SAMN11475962

Table S3. Genome assembly and quality assessment metrics

	Hyssopus officinalis	*Nepeta cataria*	*Nepeta mussinii*
Assembly metrics			
N50 scaffold Size (bp)	29,954,448	296,091	503,797
Total Assembly Size (bp)	341,469,530	494,912,498	233,432,308
Number of Scaffolds	1,464	7,710	3,615
Maximum Scaffold Length (bp)	50,735,408	2,006,709	4,040,511
Minimum Scaffold Length (bp)	895	897	891
BUSCO metrics			
C - Complete BUSCOs	1316 (91.4%)	1316 (91.4%)	1299 (90.2%)
S - Complete and single-copy BUSCOs	1245 (86.5%)	408 (28.3%)	1199 (83.3%)
D - Complete and duplicated BUSCOs	71 (4.9%)	908 (63.1%)	100 (6.9%)
F - Fragmented BUSCOs	33 (2.3%)	25 (1.7%)	43 (3.0%)
M - Missing BUSCOs	91 (6.3%)	99 (6.9%)	98 (6.8%)
Total BUSCOs	1440	1440	1440

Table S4. Genome annotation metrics.

	Hyssopus officinalis		*Nepeta cataria*		*Nepeta mussinii*	
	Working Models	High Confidence Models	Working Models	High Confidence Models	Working Models	High Confidence Models
Number of Gene Models	61,330	50,669	87,501	78,876	47,924	42,887
Number of Loci	40,752	30,421	54,395	46,187	30,147	25,403
Maximum Transcript Length (bp)	17,511	17,511	16,296	15,707	16,657	15,706
Maximum CDS Length (bp)	17,511	17,511	16,158	15,210	16,233	15,210
Average Transcript Length (bp)	1,597.3	1,819.4	1,730.0	1,841.0	1,802.3	1,920.3
Average CDS Length (bp)	1,162.3	1,299.9	1,251.3	1,319.6	1,280.9	1,350.6
Average Exon Length (bp)	293.5	297.0	283.2	285.9	299.3	302.2
Average Intron Length (bp)	335.7	326.0	352.0	346.2	340.6	335.5
Single Exon Transcripts	22,650	14,268	20,292	16,162	12,370	9,696

Table S5. Species acronyms and sequence origins. Acronyms of species used in phylogenetic analyses, and origins of sequence data. Sequences labelled "Mint Genome Project" sequences are those reported in (Mint Evolutionary Genomics Consortium, 2019 (*7*)).

Acronym	Taxon	Origin
AGFO	*Agastache foeniculum* (Pursh) Kuntze	Mint Genome Project
AJRE	*Ajuga reptans* L.	Mint Genome Project
AmTr	*Amborella trichopoda* Baill.	Mint Genome Project (Phytozome)
Am	*Antirrhinum majus* L.	NCBI/EBI GenBank
At	*Arabidopsis thaliana*(L.) Heynh.	NCBI/EBI GenBank
AUPE	*Aureolaria pectinata* (Nutt.) Pennell	Mint Genome Project
BAPS	*Ballota pseudodictamnus* (L.) Benth.	Mint Genome Project
BEOF	*Betonica officinalis* L. [syn. *Stachys officinalis* (L.) Trevis.]	Mint Genome Project
CAAM	*Callicarpa americana* L.	Mint Genome Project
Cr	*Catharanthus roseus* (L.) G.Don	Mint Genome Project (Medicinal Plant Genomics Resource)
CLBU	*Clerodendrum bungei* Steud.	Mint Genome Project
COCA	*Collinsonia canadensis* L.	Mint Genome Project
COTO	*Congea tomentosa* Roxb.	Mint Genome Project
COPY	*Cornutia pyramidata* L.	Mint Genome Project
ERGU	*Erythranthe guttata (*DC.) G.L. Nesom [syn. *Mimulus guttatus* DC.]	NCBI/EBI GenBank
GLHE	*Glechoma hederacea* L.	Mint Genome Project
GMPH	*Gmelina philippensis* Cham.	Mint Genome Project
HOSA	*Holmskioldia sanguinea* Retz.	Mint Genome Project
HYSU	*Hyptis suaveolens* (L.) Poit.	Mint Genome Project
HYOF	*Hyssopus officinalis* L.	This work
ISRU	*Isodon rubescens* (Hemsl.) H.Hara	NCBI/EB SRA
JURE	*Juglans regia* L.	NCBI/EBI GenBank
LAAL	*Lamium album* L.	Mint Genome Project
LATI	*Lancea tibetica* Hook.f. & Thomson	Mint Genome Project
LAAN	*Lavandula angustifolia* Mill.	Mint Genome Project
LELE	*Leonotis leonurus* (L.) R.Br.	Mint Genome Project
LECA	*Leonurus cardiaca* L.	Mint Genome Project
LYAM	*Lycopus americanus* Muhl. ex W.P.C.Barton	Mint Genome Project
MAVU	*Marrubium vulgare* L.	Mint Genome Project
MEOF	*Melissa officinalis* L.	Mint Genome Project
MEPI	*Mentha* × *piperita* L.	Mint Genome Project
MEAQ	*Mentha aquatica* L.	Mint Genomics Resource (Lange lab)
MELO	*Mentha longifolia* (L.) L.	Mint Genomics Resource (Lange lab) (*77*)
MESP	*Mentha spicata* L.	Mint Genome Project
MEPI_B	*Mentha x piperita* L.	Mint Genomics Resource (Lange lab)
Mp	*Mentha x piperita* L.	NCBI/EBI GenBank
MODI	*Monarda didyma* L.	Mint Genome Project

Table S5. Continued.

Acronym	Taxon	Origin
NECA	*Nepeta cataria* L.	This work
NEMU	*Nepeta mussinii* Spreng. ex Henckel.	This work
OCAM	*Ocimum americanum* L.	NCBI/EB SRA
OCBA	*Ocimum basilicum* L.	Mint Genome Project
OCTE	*Ocimum tenuiflorum* L.	NCBI/EB SRA
Oe	*Olea europaea* L.	NCBI/EBI GenBank
ORMA	*Origanum majorana* L.	Mint Genome Project
ORVU	*Origanum vulgare* L.	Mint Genome Project
PATO	*Paulownia tomentosa* (Thunb.) Steud.	Mint Genome Project
PEFR	*Perilla frutescens* (L.) Britton	Mint Genome Project
PEAT	*Perovskia atriplicifolia* Benth.	Mint Genome Project
PEBA	*Petraeovitex bambusetorum* King & Gamble	Mint Genome Project
PEVO	*Petrea volubilis* L.	Mint Genome Project
PHFR	*Phlomis fruticosa* L.	Mint Genome Project
PHPU	*Phlomis purpurea* L.	NCBI/EB SRA
Pp	*Physcomitrella patens* (Hedw.) Bruch & Schimp.	Mint Genome Project (Phytozome)
Pm	*Plantago major* L.	NCBI/EBI GenBank
PLBA	*Plectranthus barbatus* Andrews	Mint Genome Project
POCA	*Pogostemon cablin* (Blanco) Benth.	Mint Genome Project
PRMI	*Premna microphylla* Turcz.	Mint Genome Project
PRLA	*Prostanthera lasianthos* Labill.	Mint Genome Project
PRVU	*Prunella vulgaris* L.	Mint Genome Project
Ro	*Rosmarinus officinalis* L.	NCBI/EBI GenBank
ROOF	*Rosmarinus officinalis* L.	Mint Genome Project
ROMY	*Rotheca myricoides* (Hochst.) Steane & Mabb.	Mint Genome Project
SAHI	*Salvia hispanica* L.	Mint Genome Project
SAMI	*Salvia miltiorrhiza* Bunge	NCBI/EB SRA (*78*)
SAOF	*Salvia officinalis* L.	Mint Genome Project
SASC	*Salvia sclarea* L.	NCBI/EB SRA
SCBA	*Scutellaria baicalensis* Georgi	Mint Genome Project
Sm	*Selaginella moellendorffii* Hieron.	Mint Genome Project (Phytozome)
SEIN	*Sesamum indicum* L.	NCBI/EBI GenBank
TEGR	*Tectona grandis* L.f.	Mint Genome Project (*41*)
TECA	*Teucrium canadense* L.	Mint Genome Project
Ts	*Thymus serpyllum* L.	NCBI/EBI GenBank
THVU	*Thymus vulgaris* L.	Mint Genome Project
Tv	*Thymus vulgaris* L.	NCBI/EBI GenBank
Tc	*Thymus x citriodorus* (Pers.) Schreb.	NCBI/EBI GenBank
VIAG	*Vitex agnus-castus* L.	Mint Genome Project
WEFR	*Westringia fruticosa* (Willd.) Druce	Mint Genome Project

Table S6. Libraries used in *Nepeta mussinii* diversity panel

Library ID	Library Type	Read Length (nt)	Type of Sequencing	Library Size (bp)	Total Reads	Bioproject	Biosample
Nepeta mussinii Diversity Panel							
NEP_AF	RNAseq: young leaf replicate A, plant B1B	50	Single end	301	30,482,538	PRJNA557218	SAMN12391686
NEP_AG	RNAseq: young leaf replicate B, plant B1B	50	Single end	339	36,076,107	PRJNA557218	SAMN12391686
NEP_AH	RNAseq: young leaf replicate C, plant B1B	50	Single end	247	35,574,798	PRJNA557218	SAMN12391686
NEP_AI	RNAseq: young leaf replicate A, plant B1C	50	Single end	282	45,257,016	PRJNA557218	SAMN12391687
NEP_AJ	RNAseq: young leaf replicate B, plant B1C	50	Single end	301	53,673,445	PRJNA557218	SAMN12391687
NEP_AK	RNAseq: young leaf replicate C, plant B1C	50	Single end	281	42,400,576	PRJNA557218	SAMN12391687
NEP_AL	RNAseq: young leaf replicate A, plant B1E	50	Single end	317	25,879,664	PRJNA557218	SAMN12391688
NEP_AM	RNAseq: young leaf replicate B, plant B1E	50	Single end	275	33,983,367	PRJNA557218	SAMN12391688
NEP_AN	RNAseq: young leaf replicate C, plant B1E	50	Single end	281	38,554,164	PRJNA557218	SAMN12391688
NEP_BA	RNAseq: young leaf replicate A, plant 5	50	Single end	278	31,832,160	PRJNA557218	SAMN12391689
NEP_BB	RNAseq: young leaf replicate B, plant 5	50	Single end	277	46,535,207	PRJNA557218	SAMN12391689
NEP_BC	RNAseq: young leaf replicate C, plant 5	50	Single end	283	29,930,382	PRJNA557218	SAMN12391689
NEP_BD	RNAseq: young leaf replicate A, plant 7	50	Single end	291	44,028,302	PRJNA557218	SAMN12391690
NEP_BE	RNAseq: young leaf replicate B, plant 7	50	Single end	270	33,053,357	PRJNA557218	SAMN12391690
NEP_BF	RNAseq: young leaf replicate C, plant 7	50	Single end	291	33,222,611	PRJNA557218	SAMN12391690
MINT_EU*	RNAseq: closed flower buds	150	Paired end	274	28,507,931	PRJNA529674	SAMN11475956
MINT_EV*	RNAseq: immature leaf	150	Paired end	311	32,035,293	PRJNA529674	SAMN11475957
MINT_EW*	RNAseq: mature leaf	150	Paired end	325	35,098,168	PRJNA529674	SAMN11475958
MINT_EX*	RNAseq: open flowers	150	Paired end	325	36,437,355	PRJNA529674	SAMN11475959
MINT_EY*	RNAseq: petiole	150	Paired end	339	39,058,249	PRJNA529674	SAMN11475960
MINT_EZ*	RNAseq: root	150	Paired end	326	41,021,013	PRJNA529674	SAMN11475961
MINT_FA*	RNAseq: stem	150	Paired end	310	36,866,609	PRJNA529674	SAMN11475962
*Read 1 was clipped to 50bp for use in comparison with diversity panel data							

Table S7. *P5βR* locus synteny. Synteny between *H. officinalis* (ho), *N. cataria* (nc), and *N. mussinii* (nm) at the *P5βR* locus. Syntenic region identified using MCScanX; only a portion of the syntenic block is shown. The synteny between *SISY*s and *P5βR*s was adjusted manually based on phylogenetic relationships. The *SISY* genes are in bold; the P5βR genes are underlined. The *N. cataria SISYB* pseudogene was identified using blast/homology but does not have a gene model name. Hyssop *P5βRB* may be syntenic to either Nepeta *SISY*s or *P5βR*s.

nc7	**nc12**	**nm64**	**ho8**
nc_g2176.t1	nc_g3337.t1	nm_g14127.t1_g14128.t1	
nc_g2177.t1	nc_g3336.t1	nm_g14129.t1	
nc_g2178.t1	nc_g3335.t1	nm_g14130.t1	ho_g31829.t1
nc_g2179.t1	| |	| |	| |
nc_g2180.t1	nc_g3334.t1	nm_g14131.t1	| |
nc_g2181.t1	nc_g3333.t1	nm_g14132.t1	| |
nc_g2182.t1	nc_g3330.t1	nm_g14133.t1	| |
nc_g2183.t1	nc_g3329.t1	nm_g14134.t1	| |
nc_g2184.t1	nc_g3327.t1	nm_g14136.t1	| |
nc_g2185.t1	nc_g3326.t1	nm_g14137.t1	| |
nc_g2186.t1 (*SISYA*)	**(*SISYB*)**	**nm_g14138.t1 (*SISY*)**	| |
nc_g2187.t1 (*P5βRA*)	nc_g3323.t1 (*P5βRA*)	nm_g14139.t1 (*P5βR*)	ho_g31854.t1 (*P5βRB*)
nc_g2188.t1	nc_g3322.t1	nm_g14140.t1	ho_g31860.t1
nc_g2189.t1	nc_g3319.t1	nm_g14142.t1	ho_g31862.t1
nc_g2190.t1	nc_g3318.t1	nm_g14143.t1	ho_g31871.t1
nc_g2191.t1	nc_g3317.t1	nm_g14144.t1	ho_g31873.t1
nc_g2192.t1	nc_g3315.t1	| |	ho_g31874.t1
nc_g2193.t1	| |	nm_g14145.t1	| |
nc_g2194.t1	| |	| |	| |
nc_g2196.t1	nc_g3314.t1	nm_g14148.t1	ho_g31877.t1
nc_g2197.t1	nc_g3312.t1	nm_g14150.t1_g14151.t1	ho_g31880.t1.
nc_g2198.t1	nc_g3311.t1.	nm_g14152.t1	ho_g31881.t1

Table S8. Kinetic analysis of iridoid synthase (ISY) reaction for selected PRISE enzymes.

Initial rates were obtained by measuring reduction of 8-oxogeranial through depletion of NADPH (absorbance at 340 nm). Rates at 1.5625, 3.125, 6.25, 12.5, 25, 50 or 100 µM 8-oxogeranial were measured. Michaelis-Menten or substrate inhibition models were fit to initial rate data; estimated parameters and associated significance (*p*-value) is presented.

Enzyme	k_{cat} (s^{-1})	*p*	K_m (µM)	*p*	K_i (µM)	*p*	k_{cat}/K_M (mM^{-1} s^{-1})	*n*	[E] (nM)
Anc1	0.155 ± 0.008	<0.001	2.5 ± 0.6	<0.001			62.0	22	25 or 100
Anc2	0.079 ± 0.004	<0.001	0.7 ± 0.3	0.032			113	24	200
NmP5BR	0.0075 ± 0.0004	<0.001	1.9 ± 0.5	0.0015			3.95	32	250 or 300
Anc3	1.00 ± 0.07	<0.001	0.9 ± 0.4	0.047	229 ± 86	0.016	1110	21	10
NmSISY	0.277 ± 0.015	<0.001	3.6 ± 0.8	<0.001			76.9	32	50
NmISY	3.50 ± 0.48	<0.001	2.1 ± 1.0	0.05	77 ± 28	0.01	1670	26	10

Table S9. Positive selection analysis of *PRISE* genes.

Selection analysis conducted with codeml software using subsampled *PRISE* tree. Tree and branch annotations can be found on fig. S23. The M0 model was used to select codon model via Bayesian information criterion (BIC), using number of informative codon patterns as *n* (376). The branch-site model MA was used to test for positive selection on selected branches. P-values were calculated using the χ^2 distribution and adjusted with Benjamini-Hochberg multiple test correction.

Model	Codon model	Branch	lnL	df	BIC	2ΔlnL	*p*	*p-corrected*
M0	equal	NA	-23324.59	167	47639.41			
M0	F1X4+F	NA	-23471.90	167	47934.05			
M0	F1X4	NA	-23302.31	170	47612.64			
M0	F3X4+F	NA	-23483.61	167	47957.45			
M0	F3X4	NA	-23192.69	176	47428.99			
M0	Fcodon+F	NA	-23696.63	167	48383.51			
M0	Fcodon	NA	-23059.02	227	47464.06			
M0	fMutSel0+F	NA	-23117.37	170	47242.77			
M0	fMutSel0	NA	-23094.45	189	47309.59			
M0	fMutSel+F	NA	-23005.95	170	47019.93			
M0	fMutSel	NA	-22903.77	230	47171.35			
MA_H0	fMutSel+F	c	-22455.28	172	45930.45			
MA_H1	fMutSel+F	c	-22451.25	173	45928.32	8.06	0.00453	0.023
MA_H0	fMutSel+F	a	-22460.71	172	45941.3			
MA_H1	fMutSel+F	a	-22460.71	173	45947.23	0.00	1	1
MA_H0	fMutSel+F	b	-22460.71	172	45941.3			
MA_H1	fMutSel+F	b	-22459.85	173	45945.52	1.71	0.1908	0.318
MA_H0	fMutSel+F	e	-22459.65	172	45939.2			
MA_H1	fMutSel+F	e	-22458.11	173	45942.04	3.09	0.0788	0.197
MA_H0	fMutSel+F	d	-22458.68	172	45937.25			
MA_H1	fMutSel+F	d	-22458.68	173	45943.18	0.00	1	1

Table S10. *GES* locus synteny. Synteny between *H. officinalis* (ho), *N. cataria* (nc), and *N. mussinii* (nm) at the geraniol synthase (*GES*) locus. Syntenic region identified using MCScanX; only a portion of the syntenic block is shown. The *GES* genes are in bold. The *N. mussinii SGES* (g13037) is a pseudogene, whilst the active *GES* is in the *NEPS* locus (Fig. 3A).

nc1	nc238	nm54	ho9
nc_g409.t1	nc_g26145.t1	nm_g13049.t1	ho_g33517.t1
nc_g410.t1	nc_g26144.t1	nm_g13048.t1	ho_g33510.t1
nc_g411.t1	nc_g26143.t1	nm_g13047.t1	ho_g33509.t1
nc_g412.t1_g413.t1	nc_g26141.t1	nm_g13046.t1	ho_g33508.t1
nc_g414.t1	nc_g26140.t1	||	ho_g33506.t1
nc_g415.t1	nc_g26139.t1	||	ho_g33505.t1
nc_g416.t1	||	nm_g13044.t1	ho_g33504.t1
nc_g417.t1	||	nm_g13043.t1	ho_g33503.t1
nc_g418.t1	||	nm_g13042.t1	ho_g33502.t1
nc_g419.t1	||	||	||
nc_g420.t1	||	nm_g13041.t1	ho_g33500.t1
nc_g421.t1	nc_g26134.t1	nm_g13038.t1	ho_g33497.t1
nc_g422.t1	||	||	||
nc_g424.t1	**||**	**||**	**||**
nc_g426.t1 (*GESB*)	**nc_g26133.t1 (*GESA*)**	**nm_g13037.t1 (*SGES*)**	**ho_g33496.t1 (*GES*)**
nc_g427.t1	nc_g26131.t1	nm_g13036.t1	ho_g33494.t1
nc_g428.t1	||	||	||
nc_g429.t1	nc_g26128.t1	||	||
nc_g430.t1	||	||	ho_g33491.t1
nc_g431.t1	nc_g26127.t1	||	ho_g33489.t1
nc_g432.t1_g433.t1	nc_g26126.t1	nm_g13034.t1	ho_g33488.t2
nc_g434.t1	nc_g26124.t1	nm_g13033.t1	ho_g33487.t1
nc_g435.t1	nc_g26123.t1	nm_g13032.t1	ho_g33486.t1
nc_g436.t1	nc_g26122.t1	nm_g13031.t1	ho_g33481.t1
nc_g437.t1	nc_temp_model_24.2.5b2 d6aea	nm_g13030.t1	ho_g33480.t1

REFERENCES AND NOTES

1. A. O. Tucker, S. S. Tucker, Catnip and the catnip response I. *Econ. Bot.* **42**, 214–231 (1988).

2. S. Bol, J. Caspers, L. Buckingham, G. D. Anderson-Shelton, C. Ridgway, C. A. Buffington, S. Schulz, E. M. Bunnik, Responsiveness of cats (Felidae) to silver vine (*Actinidia polygama*), Tatarian honeysuckle (*Lonicera tatarica*), valerian (*Valeriana officinalis*) and catnip (*Nepeta cataria*). *BMC Vet. Res.* **13**, 70 (2017).

3. T. Eisner, Catnip: Its raison d'etre. *Science* **146**, 1318–1320 (1964).

4. M. A. Birkett, A. Hassanali, S. Hoglund, J. Pettersson, J. A. Pickett, Repellent activity of catmint, *Nepeta cataria*, and iridoid nepetalactone isomers against Afro-tropical mosquitoes, ixodid ticks and red poultry mites. *Phytochemistry* **72**, 109–114 (2011).

5. K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. van der Krol, R. Lugan, T. Ilc, R. Verpoorte, Kirsi-Marja Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink, D. Werck-Reichhart, The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606 (2014).

6. F. Geu-Flores, N. H. Sherden, V. Courdavault, V. Burlat, W. S. Glenn, C. Wu, E. Nims, Y. Cui, Sarah E. O'Connor, An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138–142 (2012).

7. N. H. Sherden, B. Lichman, L. Caputi, D. Zhao, M. O. Kamileen, C. R. Buell, S. E. O'Connor, Identification of iridoid synthases from *Nepeta* species: Iridoid cyclization does not determine nepetalactone stereochemistry. *Phytochemistry* **145**, 48–56 (2018).

8. H. Kries, F. Kellner, M. O. Kamileen, S. E. O'Connor, Inverted stereocontrol of iridoid synthase in snapdragon. *J. Biol. Chem.* **292**, 14659–14667 (2017).

9. B. R. Lichman, M. O. Kamileen, G. R. Titchiner, G. Saalbach, CE. M. Stevenson, D. M. Lawson, S. E. O'Connor, Uncoupled activation and cyclisation in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2019).

10. Mint Evolutionary Genomics Consortium, Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* **11**, 1084–1096 (2018).

11. G. W. Stull, M. Schori, D. E. Soltis, P. S. Soltis, Character evolution and missing (morphological) data across Asteridae. *Am. J. Bot.* **105**, 470–479 (2018).

12. S. Dobler, G. Petschenka, H. Pankoke, Coping with toxic plant compounds - The insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* **72**, 1593–1604 (2011).

13. J. L. Magnard, A. Roccia, J. C. Caissard, P. Vergne, P. Sun, R. Hecquet, A. Dubois, L. Hibrand-Saint Oyant, F. Jullien, F. Nicolè, O. Raymond, S. Huguet, R. Baltenweck, S. Meyer, P. Claudel, J. Jeauffre, M. Rohmer, F. Foucher, P. Hugueney, M. Bendahmane, S. Baudino, Biosynthesis of monoterpene scent compounds in roses. *Science* **349**, 81–83 (2015).

14. J. Petersen, H. Lanig, J. Munkert, P. Bauer, F. Müller-Uri, W. Kreis, Progesterone 5β-reductases/iridoid synthases (PRISE): Gatekeeper role of highly conserved phenylalanines in substrate preference and trapping is supported by molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **34**, 1667–1680 (2016).

15. H. W. Nützmann, A. Huang, A. Osbourn, Plant metabolic clusters – From genetics to genomics. *New Phytol.* **211**, 771–789 (2016).

16. I. Liblikas, E. M. Santangelo, J. Sandell, P. Baeckström, M. Svensson, U. Jacobsson, C. R. Unelius, Simplified isolation procedure and interconversion of the diastereomers of nepetalactone and nepetalactol. *J. Nat. Prod.* **68**, 886–890 (2005).

17. I. Gabur, H. S. Chawla, R. J. Snowdon, I. A. P. Parkin, Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* **132**, 733–750 (2019).

18. S. Brown, M. Clastre, V. Courdavault, S. E. O'Connor, De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3205–3210 (2015).

19. G. D. Moghe, R. L. Last, Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* **169**, 1512–1523 (2015).

20. B. Christ, C. Xu, M. Xu, F.-S. Li, N. Wada, A. J. Mitchell, X.-L. Han, M.-L. Wen, M. Fujita, J.-K. Weng, Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* **10**, 3206 (2019).

21. R. Huang, A. J. O'Donnell, J. J. Barboline, T. J. Barkman, Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10613–10618 (2016).

22. M. Kaltenbach, J. R. Burke, M. Dindo, A. Pabis, F. S. Munsberg, A. Rabin, S. C. L. Kamerlin, J. P. Noel, D. S. Tawfik, Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555 (2018).

23. P. Bauer, J. Munkert, M. Brydziun, E. Burda, F. Müller-Uri, H. Gröger, Y. A. Muller, W. Kreis, Highly conserved progesterone 5β-reductase genes (P5βR) from 5β-cardenolide-free and 5β-cardenolide-producing angiosperms. *Phytochemistry* **71**, 1495–1505 (2010).

24. I. B. Rogozin, F. Belinky, V. Pavlenko, S. A. Shabalina, D. M. Kristensen, E. V. Koonin, Evolutionary switches between two serine codon sets are driven by selection. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13109–13113 (2016).

25. Y. Matsuba T. T. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schäfer, D. Kudrna, R. A. Wing, A. M. Bolger, B. Usadel, A. Tissier, A. R. Fernie, C. S. Barry, E. Pichersky, Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell* **25**, 2022–2036 (2013).

26. K. Miyamoto, M. Fujita, M. R. Shenton, S. Akashi, C. Sugawara, A. Sakai, K. Horie, M. Hasegawa, H. Kawaide, W. Mitsuhashi, H. Nojiri, H. Yamane, N. Kurata, K. Okada, T. Toyomasu, Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. *Plant J.* **87**, 293–304 (2016).

27. A. M. Pohlit, N. P. Lopes, R. A. Gama, W. P. Tadei, V. F. Andrade-Neto, Patent literature on mosquito repellent inventions which contain plant essential oils—A review. *Planta Med.* **77**, 598–617 (2011).

28. G. Dawson, D. C. Griffiths, N. F. Janes, A. Mudd, J. A. Pickett, L. J. Wadhams, C. M. Woodcock, D. C. Griffiths, N. F. Janes, A. Mudd, J. A. Pickett, L. J. Wadhams, C. M. Woodcock, Identification of an aphid sex pheromone. *Nature* **325**, 614–616 (1987).

29. S. Koczor, F. Szentkirályi, M. A. Birkett, J. A. Pickett, E. Voigt, M. Tóth, Attraction of *Chrysoperla carnea* complex and *Chrysopa* spp. lacewings (Neuroptera: Chrysopidae) to aphid sex pheromone components and a synthetic blend of floral compounds in Hungary. *Pest Manag. Sci.* **66**, 1374–1379 (2010).

30. C. Li, X. L. Zhang, X. Y. Xue, F. F. Zhang, Q. Xu, X. M. Liang, Structural characterization of iridoid glucosides by ultra-performance liquid chromatography/electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **22**, 1941–1954 (2008).

31. C. Formisano, D. Rigano, F. Senatore, Chemical constituents and biological activities of *Nepeta* species. *Chem. Biodivers.* **8**, 1783–1818 (2011).

32. M. A. Saghai-Maroof, K. M. Soliman, R. A. Jorgensen, R. W. Allard, Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population-dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 8014–8018 (1984).

33. D. Zhao, J. P. Hamilton, G. M. Pham, E. Crisovan, K. Wiegert-Rininger, B. Vaillancourt, D. Della Penna, C. R. Buell, De novo genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *Gigascience* **6**, 1–7 (2017).

34. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).

35. R. M. Leggett, B. J. Clavijo, L. Clissold, M. D. Clark, M. Caccamo, NextClip: An analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566–568 (2014).

36. S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, D. B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011).

37. D. M. Bickhart, B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. A. Ponce de León, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell, T. P. Smith, Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017).

38. A. C. English, S. Richards, Y. Han, M. Wang, V. Vee, J. Qu, X. Qin, D. M. Muzny, J. G. Reid, K. C. Worley, R. A. Gibbs, Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLOS ONE* **7**, e47768 (2012).

39. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).

40. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

41. D. Zhao, J. P. Hamilton, W. W. Bhat, S. R. Johnson, G. T. Godden, T. J. Kinser, B. Boachon, N. Dudareva, D. E. Soltis, P. S. Soltis, B. Hamberger, C. R. Buell, A chromosomal-scale genome assembly of *Tectona grandis* reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *Gigascience* **8**, giz005 (2019).

42. M. S. Campbell, M. Law, C. Holt, J. C. Stein, G. D. Moghe, D. E. Hufnagel, J. Lei, R. Achawanantakun, D. Jiao, C. J. Lawrence, D. Ware, S. H. Shiu, K. L. Childs, Y. Sun, N. Jiang, M. Yandell, MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014).

43. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

44. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

45. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

46. M. Stanke, B. Morgenstern, AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).

47. B. J. Haas, J. R. Wortman, C. M. Ronning, L. I. Hannick, R. K. Smith Jr, R. Maiti, A. P. Chan, C. Yu, M. Farzad, D. Wu, O. White, C. D. Town, Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release. *BMC Biol.* **3**, 7 (2005).

48. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

49. S. R. Eddy, HMMER user's guide (2010); ftp://selab.janelia.org/pub/software/hmmer3/.

50. Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

51. L. Meisel, B. Fonseca, S. González, R. Baeza-Yates, V. Cambiazo, R. Campos, M. Gonźalez, A. Orellana, J. Retamales, H. Silva, A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. *Biol. Res.* **38**, 83–88 (2005).

52. R Core Team, R: A language and environment for statistical computing. (2018); www.r-project.org/.

53. R. Wehrens, L. M. C. Buydens, Self- and super-organizing maps in R: The kohonen Package. *J. Stat. Softw.* **21** (2007).

54. R. Kolde, pheatmap: Pretty Heatmaps (2019).

55. T. T. T. Dang, J. Franke, E. Tatsis, S. E. O'Connor, Dual catalytic activity of a cytochrome p450 controls bifurcation at a metabolic branch point of alkaloid biosynthesis in *Rauwolfia serpentina*. *Angew. Chem. Int. Ed.* **56**, 9440–9444 (2017).

56. J. X. Chin, B. K. Chung, D. Lee, Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* **30**, 2210–2212 (2014).

57. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability article fast track. *Mol. Biol. Evol.* **30**, 772–780 (2013).

58. L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

59. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

60. D. T. Hoang, O. Chernomor, A. Von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).

61. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).

62. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

63. J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, I. Birol, ABySS: A parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).

64. A. Loytynoja, N. Goldman, Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).

65. H. Ashkenazy, O. Penn, A. Doron-Faigenboim, O. Cohen, G. Cannarozzi, O. Zomer, T. Pupko, FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584 (2012).

66. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

67. Z. Yang, R. Nielsen, Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.* **25**, 568–579 (2008).

68. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

69. S. A. Smith, B. C. O'Meara, TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).

70. M. J. Sanderson, Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109 (2002).

71. R. K. Kar, On the Indian origin of *Ocimum* (Lamiaceae): A palynological approach. *Palaeobotanist* **43**, 43–50 (1996).

72. G. Yao, B. T. Drew, T. S. Yi, H. F. Yan, Y. M. Yuan, X. J. Ge, Phylogenetic relationships, character evolution and biogeographic diversification of *Pogostemon* s.l. (Lamiaceae). *Mol. Phylogenet. Evol.* **98**, 184–200 (2016).

73. S. B. Janssens, E. B. Knox, S. Huysmans, E. F. Smets, V. S. F. T. Merckx, Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. *Mol. Phylogenet. Evol.* **52**, 806–824 (2009).

74. M. J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (1996).

75. B. Li, P. D. Cantino, R. G. Olmstead, G. L. C. Bramley, C.-L. Xiang, Z.-H. Ma, Y.-H. Tan, D.-X. Zhang, A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Sci. Rep.* **6**, 34343 (2016).

76. B. O. Li, R. G. Olmstead, Two new subfamilies in Lamiaceae. *Phytotaxa* **313**, 222–226 (2017).

77. K. J. Vining, S. R. Johnson, A. Ahkami, I. Lange, A. N. Parrish, S. C. Trapp, R. B. Croteau, S. C. K. Straub, I. Pandelova, B. M. Lange, Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Mol. Plant* **10**, 323–339 (2017).

78. H. Xu, J. Song, H. Luo, Y. Zhang, Q. Li, Y. Zhu, J. Xu, Y. Li, C. Song, B. Wang, W. Sun, G. Shen, X. Zhang, J. Qian, A. Ji, Z. Xu, X. Luo, L. He, C. Li, C. Sun, H. Yan, G. Cui, X. Li, X. Li, J. Wei, J. Liu, Y. Wang, A. Hayward, D. Nelson, Z. Ning, R. J. Peters, X. Qi, S. Chen, Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* **9**, 949–952 (2016).