Supplementary information

Identification and characterization of three strictosidine synthases in *Nauclea orientalis*

Chentao Wei,[#a,b] Pengzhen Bao,[#a,b] Jianyu Gao,[b] Zhiyin Yu,[c] Jing Yang,[b,c] Jian-Ping Huang,[b] Sheng-Xiong Huang,[a,b,c] Zhengjia Wang,*[a] Yin Chen,*[b] and Yong-Jiang Wang*[c]

a. State Key Laboratory for Development and Utilization of Forest Food Resources, Zhejiang A&F University, Hangzhou, 311300, China.
b. State Key Laboratory of Phytochemistry and Natural Medicines, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China.
c. Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China.
[#] These authors contributed equally. * Corresponding authors.
E-mail addresses: yjwang1993@163.com; chenyin@mail.kib.ac.cn; wzhj21@163.com.

Supplementary information

Experimental Procedures

General experimental procedures

Leaves, stems, and roots of *N*. *orientalis* were collected in November 2022 from Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (China). PCR amplifications were then performed using a Bio-Rad T100 thermal cycler with Phanta® Super Fidelity DNA Polymerase (P505-d3; Vazyme). Experimental reagents, solvents, and restriction enzymes were purchased from standard commercial sources and used as specified. All primers used in this study are listed in Table S1. Primer synthesis and DNA sequencing were performed by TsingKe Co. (China). Liquid chromatography–mass spectrometry (LC–MS) analyses were performed on an Agilent 1260 Infinity HPLC system coupled to an Agilent 6120 single quadrupole mass spectrometer, with data acquired in positive ion mode over a mass range of *m/z* 100–800. LC–MS/MS analyses were carried out on an Agilent 1200 series HPLC system coupled to an Agilent 6540 quadrupole time-of-flight (Q-TOF) mass spectrometer, scanning from *m/z* 50–800. For both LC–MS and LC–MS/MS, separations were conducted on a YMC-Triart C_{18} column (4.6 × 250 mm, 5 µm) at 25 °C with a flow rate of 1.0 mL/min. The mobile phases consisted of water containing 0.1% (v/v) formic acid (solvent A) and acetonitrile (solvent B), using the following gradient: 5% to 100% B from 0–20.0 min, 100% B from 20.0–24.0 min, 100% to 5% B from 24.0–24.1 min, and 5% B from 24.1–28.0 min. The injection volume was 20 µL. LC–MS and LC–MS/MS data were processed and visualized using Agilent MassHunter Qualitative Analysis software (version B.06.00; Agilent Technologies, USA) and OriginPro 2025, respectively. Chemical structures were drawn using ChemDraw Prime 23.1.2.

LC–MS/MS analysis was carried out using an Agilent 1290 series HPLC system coupled to an Agilent 6540 quadrupole time-of-flight (Q-TOF) mass spectrometer, with MS data acquired in positive ion mode over a mass range of 50–800 *m/z*. The electrospray ionization (ESI) source parameters were as follows: 8.0 L/min drying gas (N2) flow rate, 300 °C, nebulizer 1.2 bar, and capillary voltage (Vcap), 3500 V in positive mode. The cone voltage was 35 V, and the collision energy (CE) was 10, 20, 30, 40, and 50 eV in positive MS/MS mode. Chemical structures were drawn using ChemDraw Prime 23.1.2. LC–MS/MS data were visualized with Origin Pro 2025.

Content analysis of MIAs in *N. orientalis*

The fresh tissues (root, leaf, and stem) of *N.orientalis* were collected and dried at 50 ℃. 200 mg of finely ground tissue powder was macerated in 5 mL of 70% Ethanol and was sonicated at room temperature for 40 min. After centrifugation (12,000 rpm, 10 min) and filtration through a 0.22 µm membrane, 20 µL of the filtrate was injected into LC–MS/MS for analysis. All plant tissues were performed in three independent biologic replicates.

RNA extraction and transcriptome sequencing

Root, stem, and leaf tissues of *N. orientalis* were collected and sent to Beijing Biomarker Technologies Co., Ltd. (Beijing, China) for RNA sequencing, with three biological replicates per tissue type. For each sample, 2 µg of total RNA was used for cDNA library construction. Next-generation sequencing (NGS) transcriptome analysis was performed on the Illumina NovaSeq 6000 platform (paired-end, 150 bp reads). In parallel, full-length cDNA sequencing was carried out using the PacBio platform. Raw sequencing reads were processed to remove adapter and primer sequences, and low-quality reads were filtered to generate clean data for downstream analyses. NGS transcriptome sequencing yielded 59.88 Gb of high-quality clean data after quality control, with each sample contributing at least 5.86 Gb and achieving Q30 scores above 92.92%. For full-length transcriptome analysis, PacBio single-molecule real-time (SMRT) sequencing generated 38.74 Gb of raw data, which was processed via circular consensus sequencing (CCS) to obtain high-fidelity long reads. These reads were subsequently clustered and polished to assemble a non-redundant reference transcriptome. Functional annotation of the transcripts was conducted using DIAMOND,[1] InterProScan,[2] and HMMER[3] tools against the following databases: NCBI non-redundant protein sequences (NR),[4] Swiss-Prot,[5] Gene Ontology (GO),[6] Clusters of Orthologous Groups (COG),[7] eukaryotic Orthologous Groups (KOG),[8] Protein family (Pfam),[9] and the Kyoto Encyclopedia of Genes and Genomes (KEGG).[10] Using the non-redundant transcriptome reference generated from PacBio sequencing, clean reads were aligned with STAR,[11] and transcript expression was quantified using Kallisto.[12] Gene expression levels are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM).

Biochemical regents

Loganic acid (wkq-00367-100mg), and loganin (wkq-00366-1g) were purchased from Sichuan Weikeqi Biological Technology Co., Ltd (China). Tryptamine (T46250-5g) was purchased from ACMEC (China). Strictosidine (BBP05454) was obtained from BioBioPha (China). All compounds were dissolved in DMSO to a final concentration of 10 mM prior to use.

Analysis of candidate genes

The MIAs biosynthetic genes (*NoGES*, *geraniol synthase*; *NoG8H*, *geraniol 8-hydroxylase*; *No8HGO*, *8-hydroxygeraniol oxidoreductase*; *NoIS*, *iridoid synthase*; *NoIO*, *iridoid oxidase*; *No7DLGT*, *7-deoxyloganetic acid glycosyl transferase*; *No7DLH*, *7-deoxyloganic acid hydroxylase*; *NoLAMT*, *loganic acid O-methyltransferase*; *NoSLS*, *secologanin synthase*; and *NoSTR*, *strictosidine synthase*) were identified by performing a Blastp search against known functional reference proteins (*Cr*GES: AFD64744.1, *Cr*G8H: CAC80883.1, *Cr*8HGO: Q6V4H0.1, *Cr*IS: AFW98981.1, *Cr*IO: W8JIS5.1, *Cr*7DLGH: BAO01109.1, *Cr*7DLH: AGX93062.1, *Cr*LAMT: B2KPR3, *Cr*SLS: AAA33106.1, *Cr*STR2: CAA37671.1) with an E-value cut-off of 1×10^{-5} and a maximum of 20 hits per query.

Bioinformatics analysis

Amino acid sequences (Table S2) were downloaded from the National Center for Biotechnology Information (NCBI) database. Multiple sequence alignments were performed using the ClustalW algorithm implemented in MEGA 11 and visualized by ESPript 3.2.[13,14] The phylogenetic tree based on protein sequences was constructed using the neighbor-joining method with 1,000 bootstrap replications in MEGA 11.[15] The phylogenetic tree was visualized with the interactive tree of life (iTOL) webtool.[16] The expression patterns of *NoLAMTs*, *NoSLSs*, and *NoSTRs* candidates in root, stem, and leaf of *N*. *orientalis* were visualized by the HeatMap program of the TBtools package.[17]

Expression and purification of recombinant protein in *E*. *coli*

The genes encoding *NoLAMTs* and *NoSTRs* were amplified by PCR from *N*. *orientalis* root or stem cDNA. The resulting PCR products were cloned into the pET28a vector (digested with *BamH*I / *Hind*III) or pCold-TF vector (digested with *BamH*I / *Sal*I), and the resultant constructs were subsequently transformed into *E*. *coli* BL21 (DE3) for protein expression. Subsequently, positive transformants were first grown overnight at 37 ℃ in LB medium (10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.0) with the appropriate antibiotic (50 μg/mL kanamycin for pET28a or 100 μg/mL ampicillin for pCold-TF). The overnight cultures were then used to inoculate 1 L of fresh LB medium containing the same antibiotic. Cells grew at 37 ℃ with continuous shaking until OD_{600} value of 0.6–0.8. At this point, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. The induced cultures were subsequently incubated at 16 ℃ for 16–18 h with shaking. The cell culture was centrifuged at 4,000 rpm for 25 minutes at 4 °C. The resultant pellet was resuspended with 50 mL buffer A (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 8.0). Cell lysis was performed on ice using a sonicator to disrupt the suspended cells. After centrifugation at 24,000 rpm for 30 min, the cleared supernatants were subsequently filtered through a filter membrane (0.22 μm) and loaded into a HisTrap™ FF (GE Healthcare) to purify using the AKTA pure system. The samples were eluted by a linear imidazole gradient of buffer A and buffer B (50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, pH 8.0) at a flow rate of 2 mL/min. Finally, the purified protein fractions were concentrated using an Amicon® Ultra-4 centrifugal filter (10 kDa NMWL, Merck Millipore) and stored at -80 ℃ in storage buffer (50 mM Tris-HCl, 100 mM NaCl, 20% (v/v) glycerol, pH 8.0). Protein concentrations were determined by measuring UV absorbance at 280 nm using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA).

***In vitro* characterization of recombinant proteins**

LAMT: reactions were performed at a final volume of 100 μL with 50 mM Tris-HCl (pH 7.5) as the assay buffer, 0.5 mM loganic acid as the substrate, 1 mM *S*-adenosyl-*L*-methionine (SAM) as the methyl donor, and 20 μg NoLAMTs as the catalyst. After incubation at 40 ℃ for 30 minutes, the reactions were terminated by the addition of acetonitrile (200 μL).

STR: reactions were performed in a final volume of 100 μL assay buffer containing 50 mM Phosphate Buffered Saline (PBS) (pH 7.0), 0.2 mM secologanin, 0.1 mM tryptamine as substrates, and 50 μg TF-NoSTRs. After incubation at 35 ℃ for 1 h, the reactions were terminated by the addition of acetonitrile (200 μL).

All reaction mixtures were centrifuged at 12,000 rpm for 10 minutes and the supernatant was used for LC/MS analysis. All experiments were performed in triplicate.

Transient expression in *N*. *benthamiana*

The *N. benthamiana* plants were cultured with a 16 h : 8 h, light : dark cycle at ambient lab temperature for 4–5 weeks to transiently express candidate genes, as previously described.[18] The full-length sequence of *NoLAMTs*, *NoSLSs*, and *NoSTRs* were amplified using the primers listed in Supplementary Table 1 and then ligated to the pEAQ-HT vector digested by *Age*I and *Xho*I. The recombinant plasmids were transformed into chemically competent *E*. *coli* DH5α and grown overnight on LB agar plates (kanamycin 50 μg/mL) at 37 ℃. Single colonies were then inoculated into LB liquid medium containing kanamycin (50 μg/mL) and grown overnight at 37 °C with shaking at 200 rpm. The positive constructs were determined by PCR and plasmid sequencing. The verified plasmids were then mobilized into electrocompetent *Agrobacterium tumefaciens* LBA4404 cells for heterologous expression. Successful transformants were isolated by plating on LB medium containing a triple-antibiotic cocktail of rifampicin (25 μg/mL), streptomycin (50 μg/mL), and kanamycin (50 μg/mL) at 30 °C for 2–3 days. Bacterial cells were harvested and resuspended in MMA induction buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$, 150 μM acetosyringone) to a final OD_{600} of 0.4. The Agrobacterium suspensions were infiltrated into *N. benthamiana* leaves after incubation for 2 hours at room temperature using a 1 mL syringe. After 3 days, a mixture of potential substrates (0.2 mM loganic acid and 0.2 mM tryptamine, or 0.2 mM loganin and 0.2 mM tryptamine) were infiltrated into previously Agrobacterium-infiltrated leaves. 36 h later, the infiltrated leaves were harvested, powdered in liquid nitrogen, and then directly extracted with 25 × mL 95% (v/v) ethanol. Following centrifugation, the ethanolic layers were collected and extracted with 25 × mL petroleum ether for three times. After evaporating *in vacuo*, the residue was dissolved in 2 mL of methanol. After centrifugation at 12,000 rpm for 10 minutes, the supernatant was used for LC–MS analysis. All experiments were performed in triplicate.

Supplementary Figures

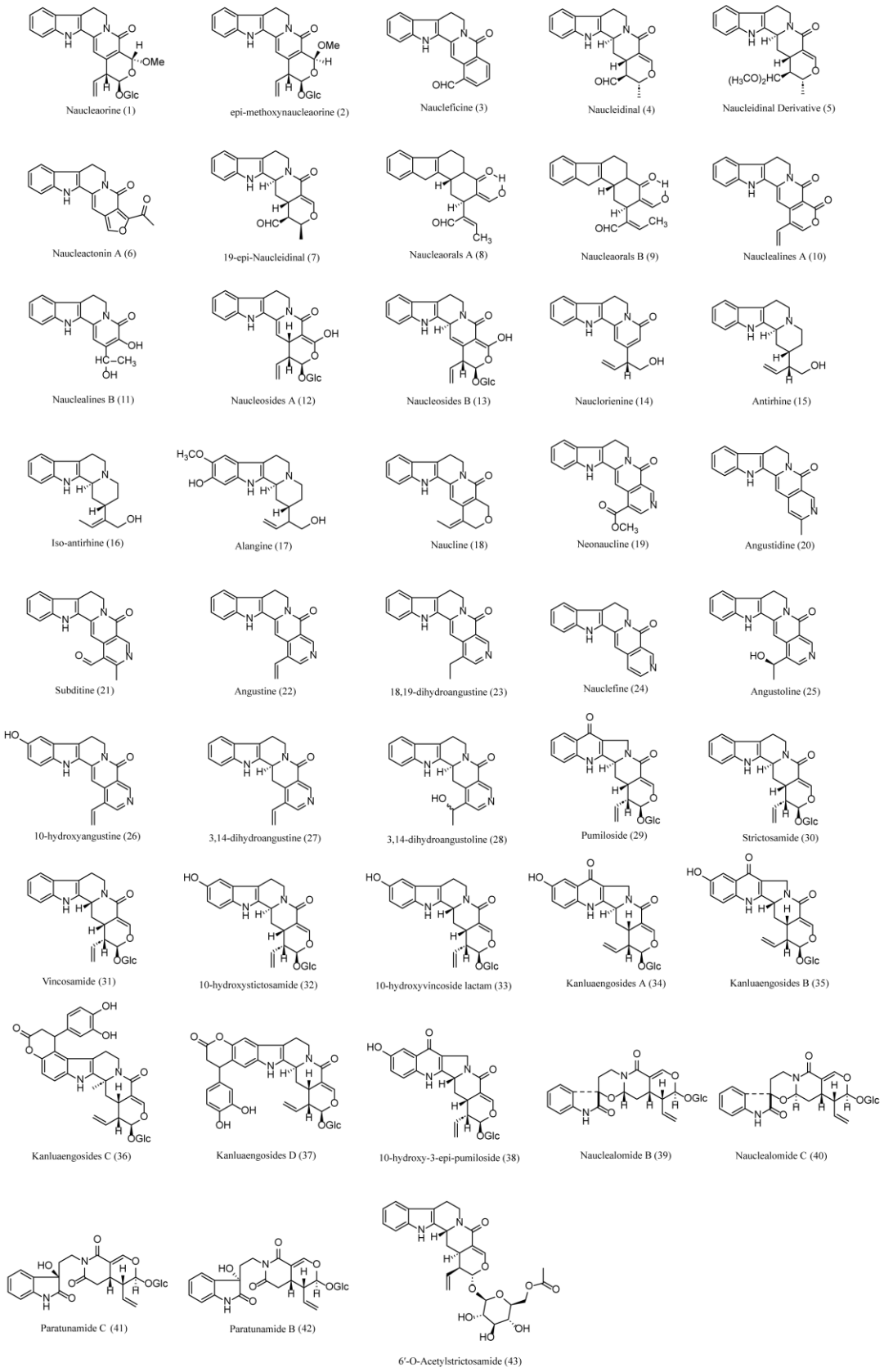

Figure S1. Chemical structures of monoterpene indole alkaloids 1-43 isolated from *N. orientalis*.[19-27]

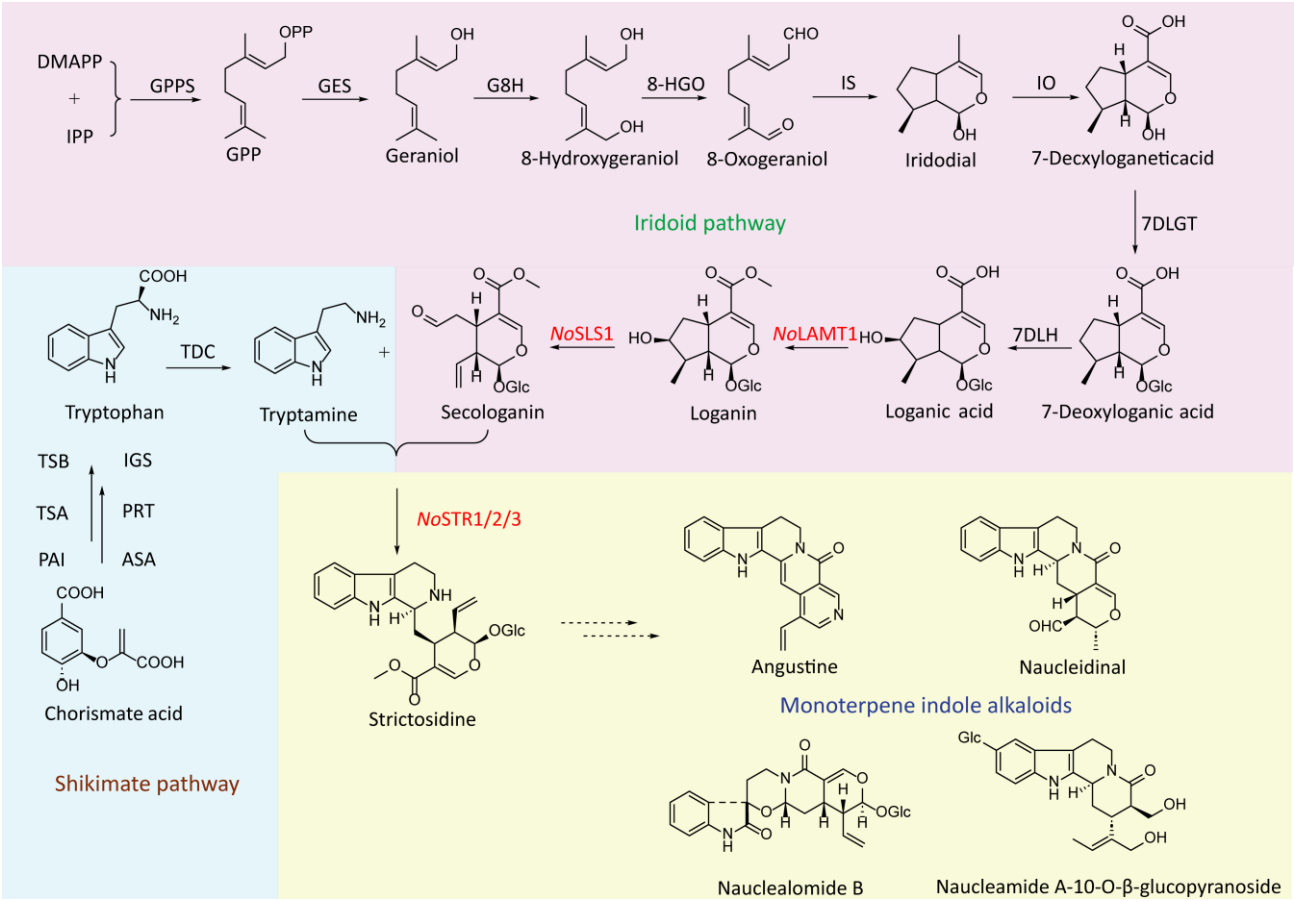

Figure S2. The proposed biosynthetic pathway of MIAs in *N. orientalis*[28]. The enzymes characterized in this study are highlighted in red. Abbreviations for biosynthetic intermediates and enzymes: IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl pyrophosphate; GPPS, GPP synthase; GES, geraniol synthase; G8H, geraniol 8-hydroxylase; 8HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glycosyl transferase; 7DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid *O*-methyltransferase; SLS, secologanin synthase; ASA, anthranilate synthase; PRT, phosphoribosyl diphosphate anthranilate transferase; PAI, PR-anthranilate isomerase; IGS, indole-3-glycerol phosphate synthase; TSA, tryptophan synthase *α*; TSB, tryptophan synthase *β*; TDC, tryptophan decarboxylase; STR, strictosidine synthase.

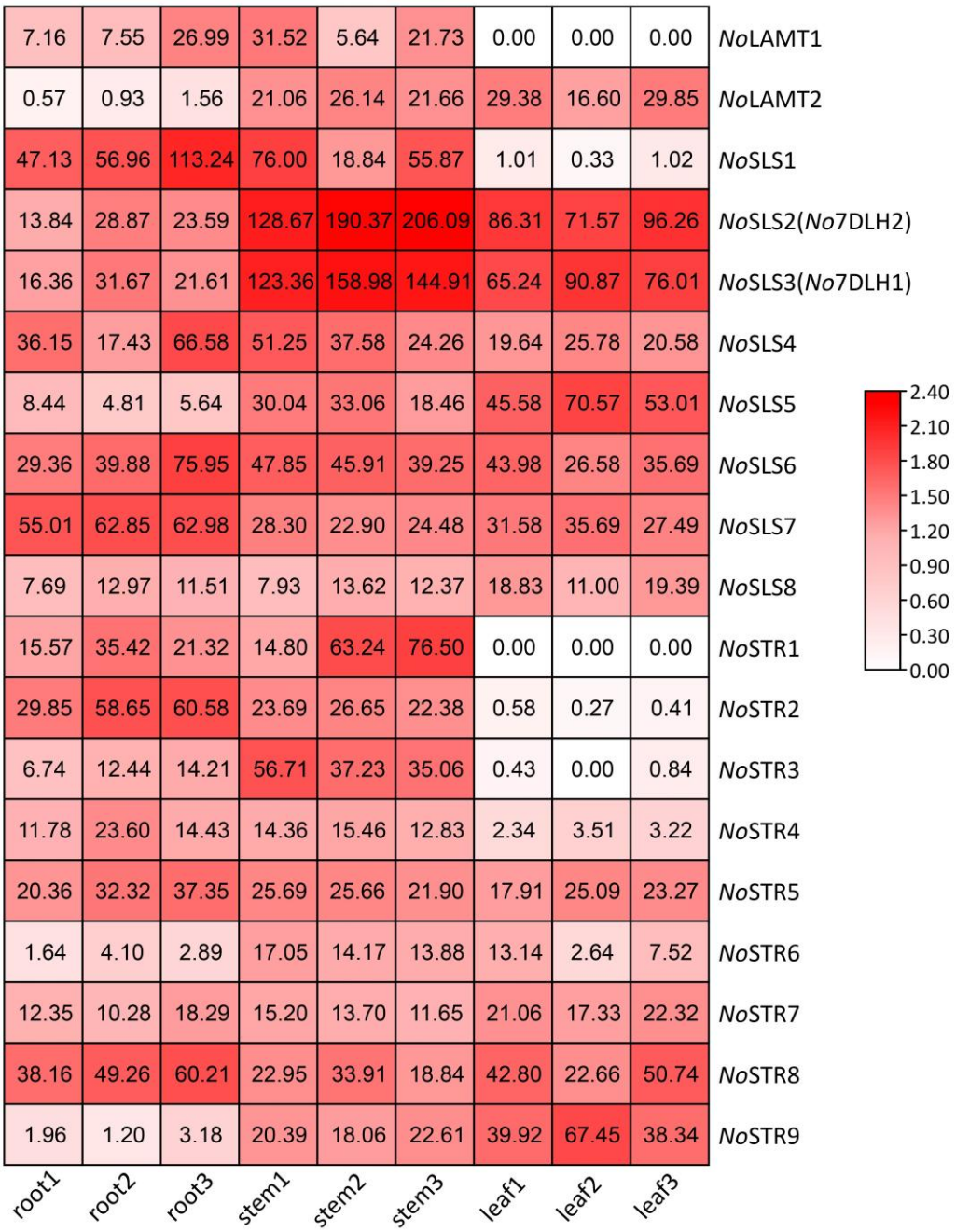

Figure S3. Expression patterns of *No*LAMT1-2, *No*SLS1-8, and *No*STR1-9. The expression levels of each gene were represented as the $\log_{10}$FPKM values. The heatmaps were generated by the HeatMap program of TBtools package.

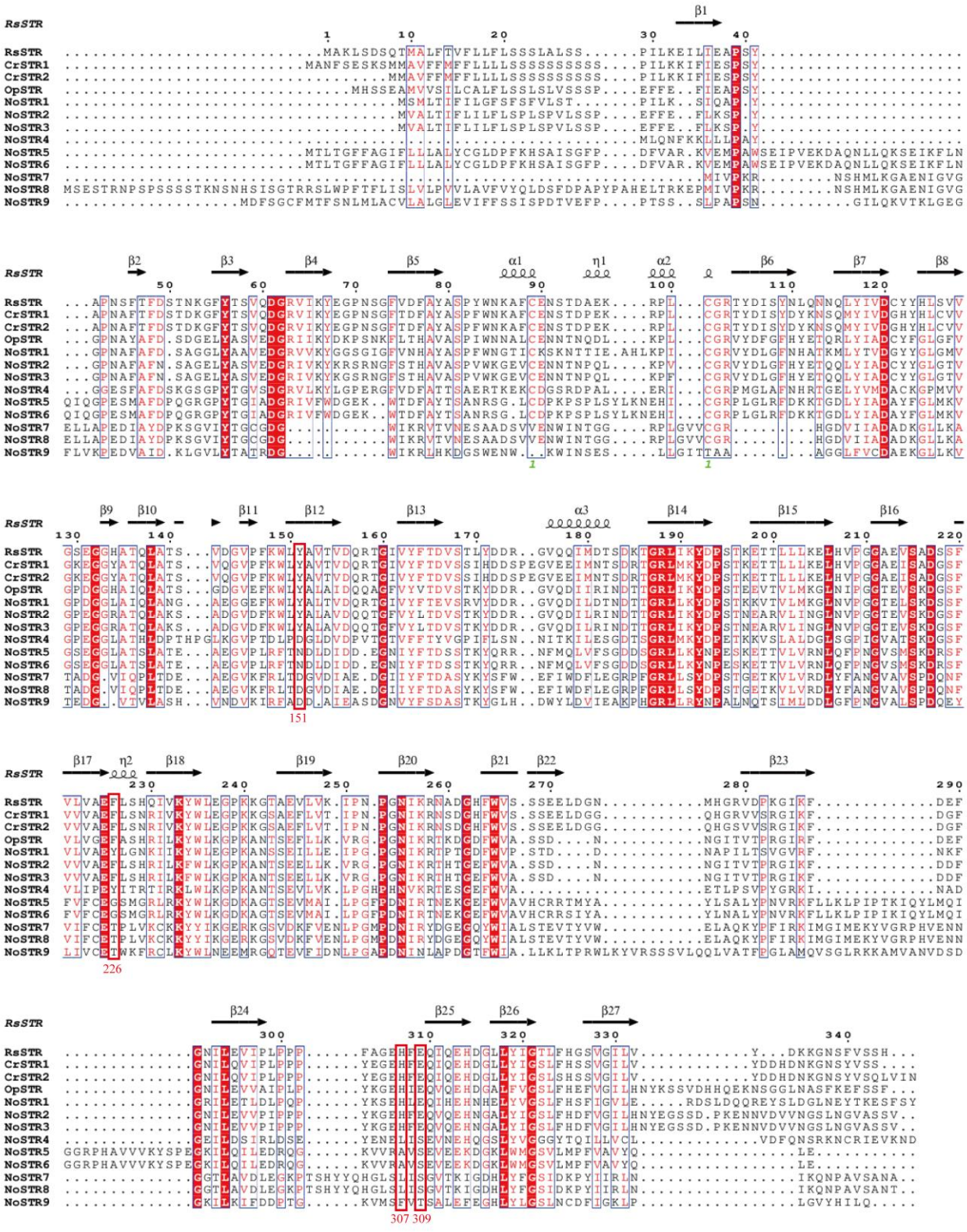

Figure S4. The amino acid conservative sequence analysis of *No*STRs. Amino acid sequence alignment of *No*STR1-9 and four known STRs from *Rauvolfia serpentina*,[29] *Catharanthus roseus*,[28] and *Ophiorrhiza pumila*.[30] Key amino acid residues are numbered and highlighted with red boxes.

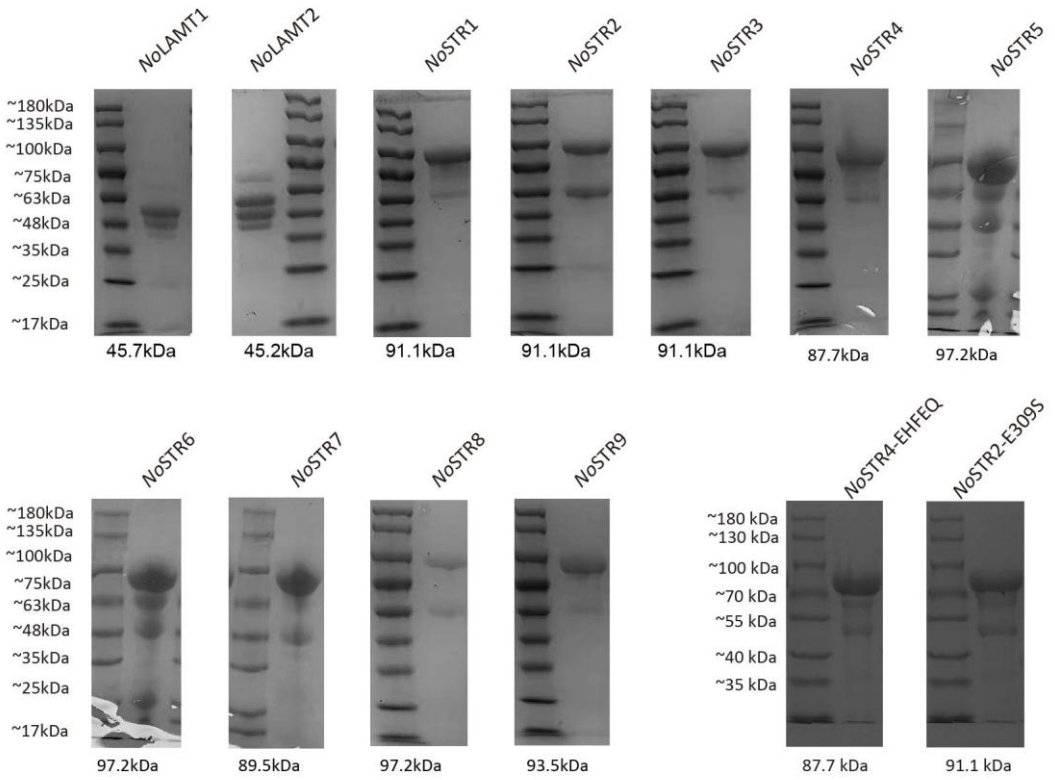

Figure S5. SDS-PAGE analyses of *No*LAMTs and *No*STRs expression in *E*. *coli*. Predicted molecular masses were TF-*No*STR1, 91.1 kDa; TF-*No*STR2, 91.1 kDa; TF-*No*STR3, 91.1 kDa; TF-*No*STR4, 87.7 kDa; TF-*No*STR5, 97.2 kDa; TF-*No*STR6, 97.2 kDa; TF-*No*STR7, 89.5 kDa; TF-*No*STR8, 97.2 kDa; TF-*No*STR9, 93.5 kDa; 6 x His-*No*LAMT1, 45.7 kDa; 6 x His-*No*LAMT2, 45.2 kDa; TF-*No*STR2-E309S, 91.1 kDa; TF-*No*STR4-EHFEQ, 87.7 kDa.

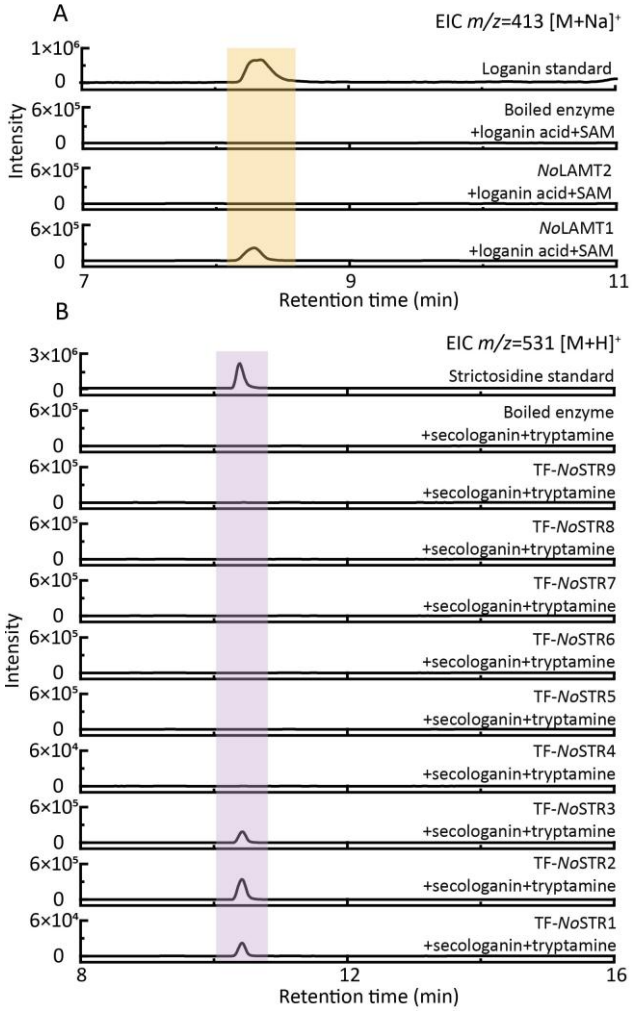

Figure S6. *In vitro* functional analysis of *No*LAMTs and *No*STRs. (A) LC–MS chromatograms of the enzymatic product of *No*LAMT1-2 recombinant proteins with *m/z* = 413 $[M+Na]^+$. (B) LC-MS analysis of the enzymatic product of TF-*No*STR1-9 with a mass corresponding to *m/z* = 531 $[M+H]^+$.

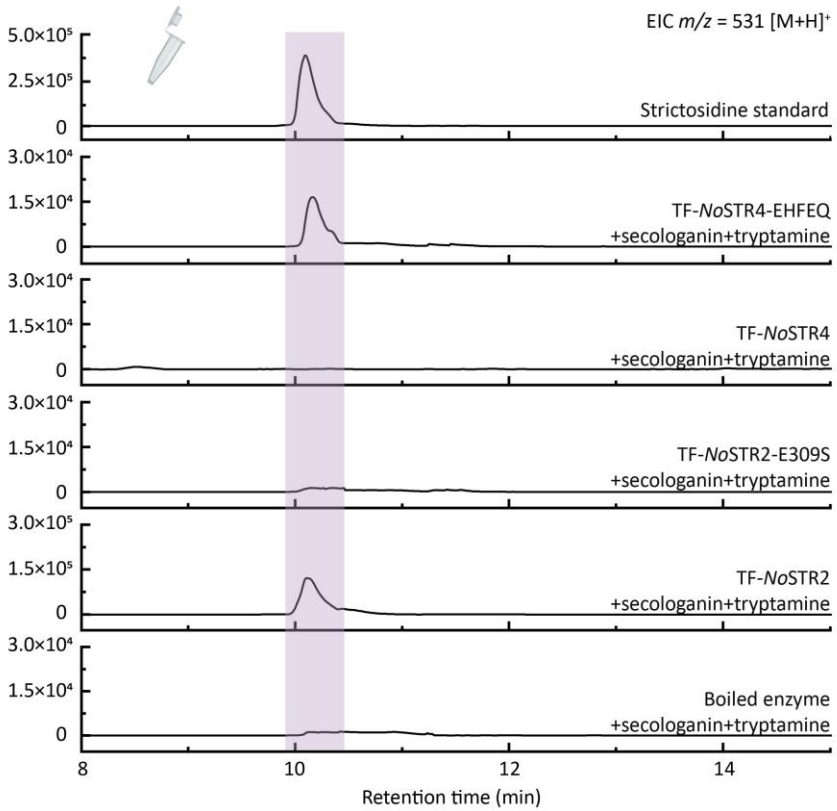

Figure S7. *In vitro* functional analysis of *No*STR2-E309S and *No*STR4-EHFEQ using secologanin and tryptamine as substrates. The products were monitored using EIC at *m/z* = 531 $[M+H]^+$.

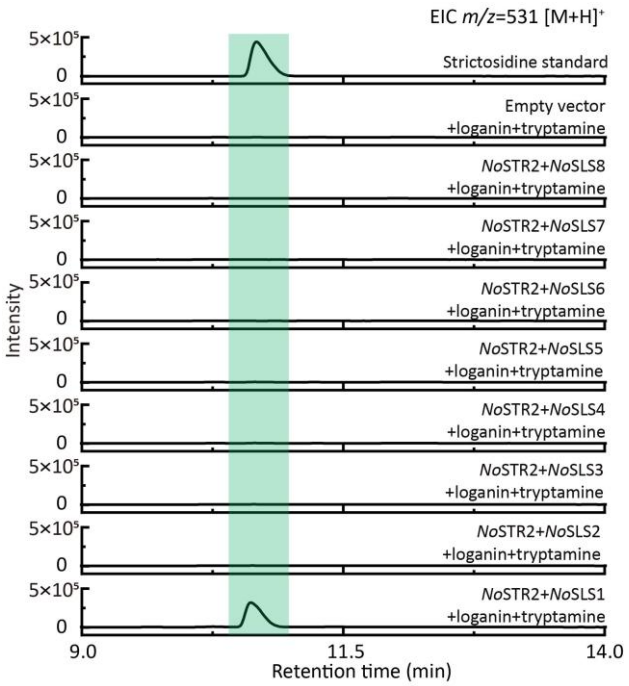

Figure S8. Functional characterization of *No*SLSs in *N. benthamiana*. EIC of strictosidine (*m/z* = 531 $[M+H]^+$) detected in *N. benthamiana* transiently expressing *No*SLS1-8, and *No*STR2 (following infiltration with loganin and tryptamine).

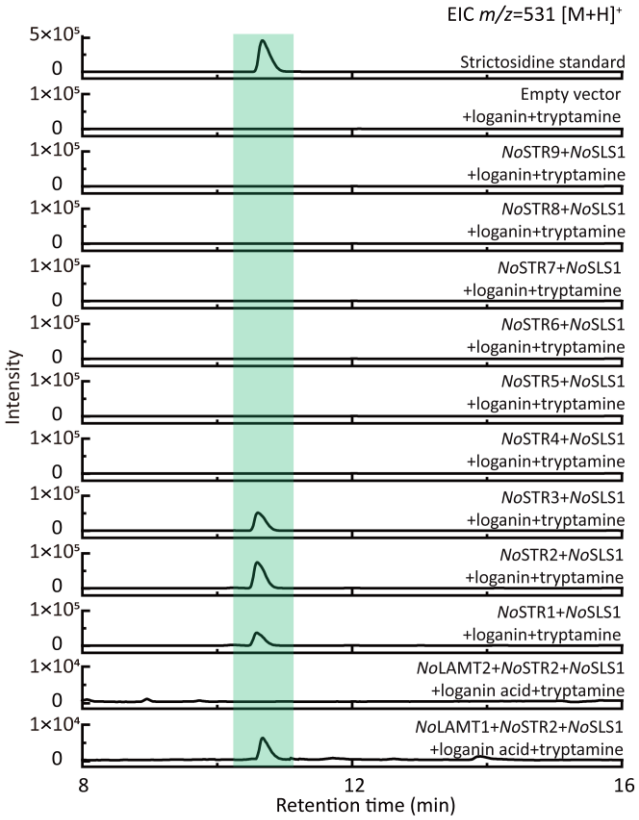

Figure S9. Functional characterization of *No*LAMTs and *No*STRs in *N. benthamiana*. EIC of strictosidine (*m/z* = 531 $[M+H]^+$) detected in *N. benthamiana* transiently expressing *No*SLS1, *No*LAMT1-2, and *No*STR1–9 (following infiltration with loganin and tryptamine or loganin acid and tryptamine).

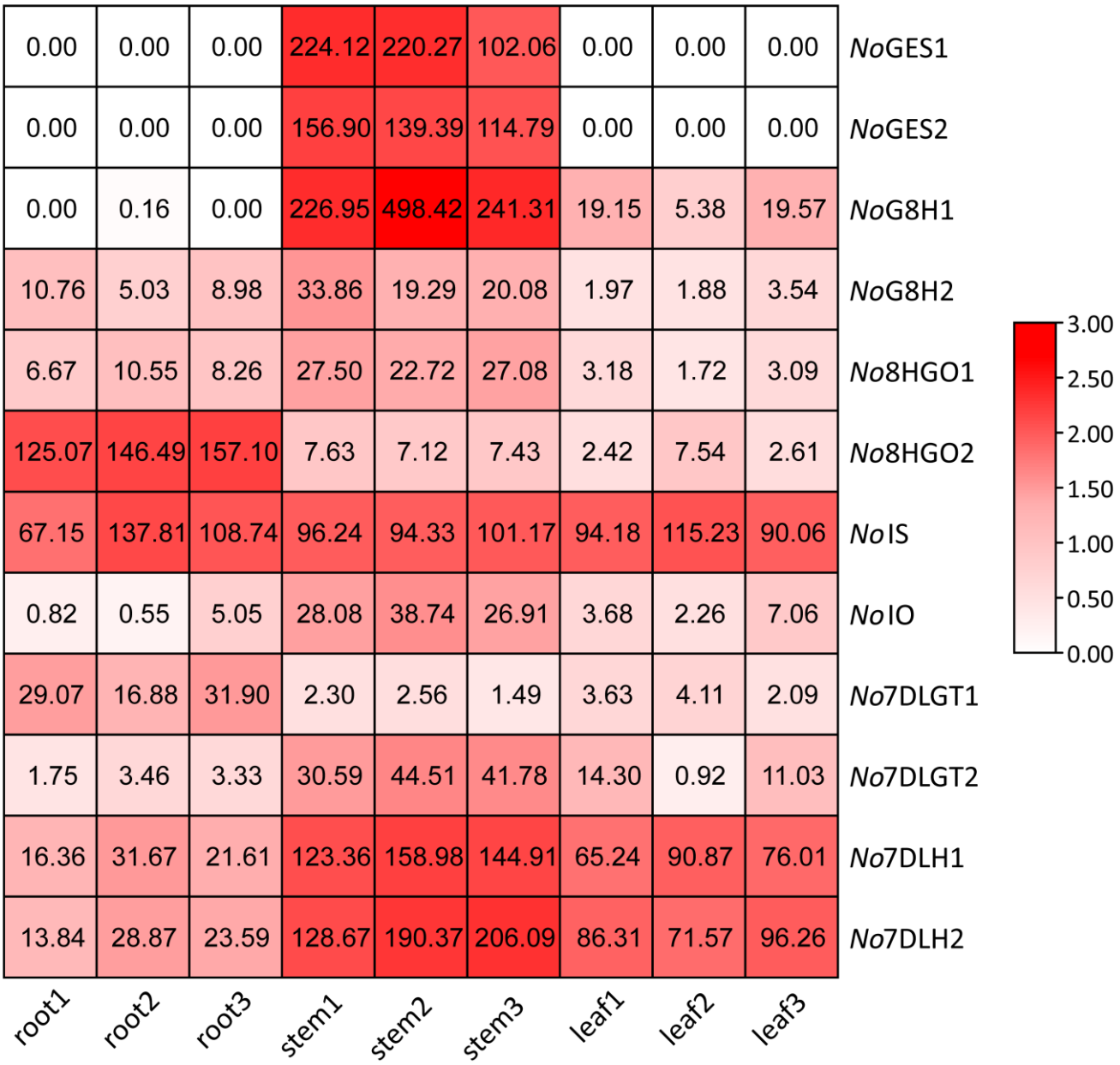

Figure S10. Expression patterns of *NoGES1-2*, *NoG8H1-2*, *No8HGO1-2*, *NoIS*, *NoIO*, *No7DLGT1-2*, and *No7DLH1-2*. The expression levels of each gene were represented as the $\log_{10}$FPKM values. The heatmaps were generated by the HeatMap program of TBtools package.

Supplementary Tables

Table S1. Targeted metabolic profiling analysis of metabolites in *N. orientalis*.

No	Identification compound	Molecular formula	biological activity	Reference
1	Naucleaorine	$C_{27}H_{30}N_2O_9$	Antimalarial	21
2	epi-methoxynaucleaorine	$C_{27}H_{30}N_2O_9$	Antimalarial	21
3	Naucleficine	$C_{20}H_{14}N_2O_2$	Anti-tumor	25
4	Naucleidinal	$C_{20}H_{20}N_2O_3$	Anti-tumor	25
5	Naucleidinal Derivative	$C_{22}H_{26}N_2O_4$	Anti-tumor	25
6	Naucleactonin A	$C_{19}H_{14}N_2O_3$	Anti-inflammatory, Antivirus	25
7	19-epi-Naucleidinal	$C_{20}H_{20}N_2O_3$	Anti-tumor	25
8	Naucleaorals A	$C_{20}H_{20}N_2O_3$	Anti-tumor	24
9	Naucleaorals B	$C_{20}H_{20}N_2O_3$	Anti-tumor, Antimalarial	24
10	Nauclealines A	$C_{20}H_{14}N_2O_3$	N/A	27
11	Nauclealines B	$C_{17}H_{16}N_2O_3$	N/A	27
12	Naucleosides A	$C_{26}H_{29}N_2O_9$	N/A	27
13	Naucleosides B	$C_{26}H_{28}N_2O_9$	N/A	27
14	Nauclorienine	$C_{19}H_{18}N_2O_2$	Anti-tumor	23
15	Antirhine	$C_{19}H_{24}N_2O$	Anti-tumor	23
16	Iso-antirhine	$C_{19}H_{24}N_2O$	Anti-tumor	23
17	Alangine	$C_{18}H_{25}NO_3$	Anti-tumor	23
18	Naucline	$C_{20}H_{18}N_2O_2$	Vasorelaxant, Anticholinesterase	23
19	Neonaucline	$C_{20}H_{15}N_3O_3$	Vasorelaxant	23
20	Angustidine	$C_{19}H_{15}N_3O$	Anticholinesterase	23
21	Subditine	$C_{20}H_{15}N_3O_2$	Anti-tumor	23
22	Angustine	$C_{20}H_{15}N_3O$	Antivirus, Vasorelaxant	19,26
23	18,19-dihydroangustine	$C_{20}H_{17}N_3O$	N/A	19,26
24	Nauclefine	$C_{18}H_{13}N_3O$	Anti-inflammatory, Vasorelaxant	19
25	Angustoline	$C_{20}H_{17}N_3O_2$	Anti-inflammatory, Anti-tumor	19,26
26	10-hydroxyangustine	$C_{20}H_{15}N_3O_2$	Anti-tumor	19
27	3,14-dihydroangustine	$C_{20}H_{18}N_3O$	Anti-inflammatory, Antivirus	19,26
28	3,14-dihydroangustoline	$C_{20}H_{19}N_3O_2$	Anticholinesterase, Anti-inflammatory	19,26
29	Pumiloside	$C_{26}H_{28}N_2O_9$	Anti-tumor	22,25,26
30	Strictosamide	$C_{26}H_{30}N_2O_8$	Anti-tumor, Analgesic, Antivirus, Antimalarial	20-22,25,26
31	Vincosamide	$C_{26}H_{30}N_2O_8$	Anti-tumor, Anti-microbial, Anti-inflammatory	20,22
32	10-hydroxystictosamide	$C_{26}H_{30}N_2O_9$	N/A	20,22,26
33	10-hydroxyvincoside lactam	$C_{26}H_{30}N_2O_9$	N/A	22,26
34	Kanluaengosides A	$C_{26}H_{28}N_2O_{10}$	N/A	22
35	Kanluaengosides B	$C_{26}H_{28}N_2O_{10}$	N/A	22
36	Kanluaengosides C	$C_{35}H_{36}N_2O_{12}$	N/A	22
37	Kanluaengosides D	$C_{35}H_{36}N_2O_{12}$	N/A	22
38	10-hydroxy-3-epi-pumiloside	$C_{26}H_{28}N_2O_{10}$	N/A	26
39	Nauclealomide B	$C_{26}H_{30}N_2O_{10}$	N/A	26
40	Nauclealomide C	$C_{26}H_{30}N_2O_{10}$	N/A	26
41	Paratunamide C	$C_{26}H_{30}N_2O_{11}$	N/A	22,26
42	Paratunamide B	$C_{26}H_{30}N_2O_{11}$	N/A	22
43	6’-O-Acetylstrictosamide	$C_{26}H_{30}N_2O_9$	N/A	20

Table S2. Targeted metabolic profiling analysis of metabolites in *N. orientalis*.[31]

Name	Molecular formula	Calculated for $[M+H]^+$	Error (ppm)	Experiment		
				Retention time (min)	$[M+H]^+$	MS/MS fragments
Naucleamide A-10-O-β-glucopyranoside	$C_{26}H_{34}N_2O_9$	519.234	-1.93	7.62	519.233	357.181, 339.170, 187.087
Pumiloside	$C_{26}H_{28}N_2O_9$	513.187	0.00	9.03	531.187	351.133, 281.092, 186.078
Paratunamide B	$C_{26}H_{30}N_2O_{11}$	547.192	0.00	9.15	547.192	367.128, 297.084, 158.063
Nauclealomide B	$C_{26}H_{30}N_2O_{10}$	531.197	1.88	10.24	531.198	369.145, 353.150, 282.096
Angustoline	$C_{20}H_{17}N_3O_2$	332.140	0.00	11.47	332.140	316.108
Strictosamide	$C_{26}H_{30}N_2O_8$	499.208	-2.00	11.67	499.207	337.155, 267.113, 171.091
Naucleidinal	$C_{20}H_{20}N_2O_3$	337.155	-2.97	12.07	337.154	267.113, 171.092, 144.081
Angustine	$C_{20}H_{15}N_3O$	314.129	0.00	13.89	314.129	299.105, 174.059, 117.063
Nauclefidine	$C_{16}H_{12}N_2O_2$	265.097	0.00	14.48	265.097	250.069, 182.082
3,14-dihydroangustine	$C_{20}H_{17}N_3O$	316.144	0.00	15.82	316.144	171.091, 146.060

Table S3. The candidate genes *No*LAMTs, *No*SLSs, and *No*STRs were obtained from the *N. orientalis* transcriptome based on a homology search with known genes from *Catharanthus roseus*.[32]

Name	Amino acid (base pairs)	Predicted Function	Amino acid similarity
NoLAMT1	374(1122)	Loganic acid *O*-methyltransferase	79.2%
NoLAMT2	367(1101)	Loganic acid *O*-methyltransferase	35.1%
NoSTR1	345(1035)	Strictosidine synthase	54.6%
NoSTR2	345(1035)	Strictosidine synthase	54.3%
NoSTR3	343(1029)	Strictosidine synthase	51.5%
NoSTR4	317(951)	Strictosidine synthase	37.9%
NoSTR5	391(1173)	Strictosidine synthase	31.6%
NoSTR6	391(1173)	Strictosidine synthase	31.6%
NoSTR7	328 (984)	Strictosidine synthase	29.5%
NoSTR8	397(1191)	Strictosidine synthase	29.5%
NoSTR9	365(1095	Strictosidine synthase	30.2%
NoSLS1	521(1563)	Secologanin synthase	86.0%
NoSLS2	519(1557)	Secologanin synthase	56.5%
NoSLS3	519(1557)	Secologanin synthase	56.5%
NoSLS4	523(1569)	Secologanin synthase	52.3%
NoSLS5	516(1548)	Secologanin synthase	54.0%
NoSLS6	518(1554)	Secologanin synthase	53.6%
NoSLS7	515(1545)	Secologanin synthase	54.7%
NoSLS8	513(1539)	Secologanin synthase	45.9%

Table S4. The expression profiling of candidate genes in *N. orientalis* (FPKM). Genes that have been functionally characterized in this study were distinctly marked in red.

Gene_ID	Root-1	Root-2	Root-3	Stem-1	Stem-2	Stem-3	Leaf-1	Leaf-2	Leaf-3
NoLAMT1	7.16	7.55	26.99	31.52	5.64	21.73	0	0	0
NoLAMT2	0.57	0.93	1.56	21.06	26.14	21.66	29.38	16.6	29.85
NoSLS1	47.13	56.96	113.24	76	18.84	55.87	1.01	0.33	1.02
NoSLS2(No7DLH2)	13.84	28.87	23.59	128.67	190.37	206.09	86.31	71.57	96.26
NoSLS3(No7DLH1)	16.36	31.67	21.61	123.36	158.98	144.91	65.24	90.87	76.01
NoSLS4	36.15	17.43	66.58	51.25	37.58	24.26	19.64	25.78	20.58
NoSLS5	8.44	4.81	5.64	30.04	33.06	18.46	45.58	70.57	53.01
NoSLS6	29.36	39.88	75.95	47.85	45.91	39.25	43.98	26.58	35.69
NoSLS7	55.01	62.85	62.98	28.3	22.9	24.48	31.58	35.69	27.49
NoSLS8	7.69	12.97	11.51	7.93	13.62	12.37	18.83	11	19.39
NoSTR1	15.57	35.42	21.32	14.8	63.24	76.5	0	0	0
NoSTR2	29.85	58.65	60.58	23.69	26.65	22.38	0.58	0.27	0.41
NoSTR3	6.74	12.44	14.21	56.71	37.23	35.06	0.43	0	0.84
NoSTR4	11.78	23.6	14.43	14.36	15.46	12.83	2.34	3.51	3.22
NoSTR5	20.36	32.32	37.35	25.69	25.66	21.9	17.91	25.09	23.27
NoSTR6	1.64	4.1	2.89	17.05	14.17	13.88	13.14	2.64	7.52
NoSTR7	12.35	10.28	18.29	15.2	13.7	11.65	21.06	17.33	22.32
NoSTR8	38.16	49.26	60.21	22.95	33.91	18.84	42.8	22.66	50.74
NoSTR9	1.96	1.2	3.18	20.39	18.06	22.61	39.92	67.45	38.34
NoGES1	0	0	0	224.12	220.27	102.06	0	0	0
NoGES2	0	0	0	156.9	139.39	114.79	0	0	0
NoG8H1	0	0.16	0	226.95	498.42	241.31	19.15	5.38	19.57
NoG8H2	10.76	5.03	8.98	33.86	19.29	20.08	1.97	1.88	3.54
No8HGO1	6.67	10.55	8.26	27.5	22.72	27.08	3.18	1.72	3.09
No8HGO2	125.07	146.49	157.1	7.63	7.12	7.43	2.42	7.54	2.61
NoIS	67.15	137.81	108.74	96.24	94.33	101.17	94.18	115.23	90.06
NoIO	0.82	0.55	5.05	8.08	38.74	26.91	3.68	2.26	7.06
No7DLGT1	29.07	16.88	31.9	2.3	2.56	1.49	3.63	4.11	2.09
No7DLGT2	1.75	3.46	3.33	30.59	44.51	41.78	14.3	0.92	11.03
No7DLH1	16.36	31.67	21.61	123.36	158.98	144.91	65.24	90.87	76.01
No7DLH2	13.84	28.87	23.59	128.67	190.37	206.09	86.31	71.57	96.26

Table S5. The candidate genes for MIAs biosynthesis were obtained from the *N. orientalis* transcriptome based on a homology search with known genes from *C. acuminata* and *C. roseus*.[32-37]

Name	Amino acid (base pairs)	Predicted Function	Amino acid similarity	References
NoGES1	571(1713)	Geraniol synthase	41.4%	32
NoGES2	543(1629)	Geraniol synthase	38.6%	32
NoG8H1	504(1512)	Geraniol 8-hydroxylase	48.5%	34
NoG8H2	497(1491)	Geraniol 8-hydroxylase	53.0%	34
No8HGO1	362(1086)	8-hydroxygeraniol oxidoreductase	68.4%	33
No8HGO2	359(1077)	8-hydroxygeraniol oxidoreductase	69.0%	33
NoIS	387(1161)	Iridoid synthase	61.0%	32
NoIO	500(1500)	Iridoid oxidase	35.9%	36
No7DLGT1	484(1452)	7-deoxyloganetic acid glucosyltransferase	34.4%	32, 35
No7DLGT2	459(1377)	7-deoxyloganetic acid glucosyltransferase	34.5%	32, 35
No7DLH1(NoSLS3)	519(1557)	7-deoxyloganic acid-7-hydroxylase	87.1%	32, 37
No7DLH2(NoSLS2)	519(1557)	7-deoxyloganic acid-7-hydroxylase	86.8%	32, 37

Table S6. LAMTs, SLSs, 7DLHs and STRs homologs from other species in phylogenetic tree.

LAMT (loganic acid *O*-methyltransferase)		
Gene	**Species**	**GenBank accession number**
CrLAMT	*Catharanthus roseus*	ABW38009.1
OpLAMT	*Ophiorrhiza pumila*	QWX38535.1
OeLAMT	*Olea europaea*	ATN39845.1
PtLAMT	*Populus trichocarpa*	AOW44494.1
UtLAMT	*Uncaria tomentosa*	QDD67562.1
PaLAMT	*Populus alba*	XP_034927111.1
HpLAMT	*Hamelia patens*	UYZ56985.1
MsLAMT	*Mitragyna speciosa*	WIE96233.1
NtLAMT	*Nothapodytes tomentosa*	WMS56963.1
NsLAMT	*Nicotiana sylvestris*	XP_009781450.1
ZjLAMT	*Ziziphus jujuba*	XP_048329770.1
JrLAMT	*Juglans regia*	XP_018843448.2
CasLAMT	*Cucurbita argyrosperma*	KAG6600862.1
SLS (secologanin synthase)		
Gene	**Species**	**GenBank accession number**
NtSLS	*Nothapodytes tomentosa*	WMS56962.1
CrSLS1	*Catharanthus roseus*	AAA33106.1
CrSLS2	*Catharanthus roseus*	AGX93064.1
CaCYP72A610	*Camptotheca acuminata*	QDC27813.1
CaCYP72A565	*Camptotheca acuminata*	AGX93062.1
TeSLS	*Tabernaemontana elegans*	AGX93048.1
UrSLS	*Uncaria rhynchophylla*	WRW51048.1
MsSLS	*Mitragyna speciosa*	WIE96234.1
OpSLS	*Ophiorrhiza pumila*	BAP90521.1
RsSLS	*Rauvolfia serpentina*	AGX93049.1
NnSLS	*Nothapodytes nimmoniana*	AQW38832.1
7DLH (7-deoxyloganic acid hydroxylase)		
Gene	**Species**	**GenBank accession number**
Cr7DLH	*Catharanthus roseus*	QDC27812.1
Cc7DLH	*Cinchona calisaya*	AGX93057.1
Rs7DLH	*Rauvolfia serpentina*	AGX93059.1
Ca7DLH	*Camptotheca acuminata*	WMD26767.1
STR (strictosidine synthase)		
Gene	**Species**	**GenBank accession number**
GsSTR	*Gelsemium sempervirens*	AXK92563.1
CpSTR	*Cinchona pubescens*	XKQ11353.1
RmSTR	*Rauvolfia mannii*	CAA45025.1
OjSTR	*Ophiorrhiza japonica*	ACF21007.1
UrSTR2	*Uncaria rhynchophylla*	USE06681.1
UrSTR1	*Uncaria rhynchophylla*	USZ80136.1
OpSTR	*Ophiorrhiza pumila*	BAB47180.1
RsSTR	*Rauvolfia serpentina*	CAA44208.1
RvSTR	*Rauvolfia verticillata*	AAY81922.1
CrSTR1	*Catharanthus roseus*	CAA71255.1
CrSTR2	*Catharanthus roseus*	CAA37671.1
NnSTR	*Nothapodytes nimmoniana*	AZZ89575.1
AtSSL4	*Arabidopsis thaliana*	At3g51420
AtSSL5	*Arabidopsis thaliana*	At3g51430
AtSSL6	*Arabidopsis thaliana*	At3g51440
AtSSL7	*Arabidopsis thaliana*	At3g51450

Table S7. Sequence identities between each of *No*STR1, *No*STR2, *No*STR3 and each of *No*STR4–9.

Identity (%)	***No*STR4**	***No*STR5**	***No*STR6**	***No*STR7**	***No*STR8**	***No*STR9**
***NoS*TR1**	40.71	33.04	36.27	28.88	28.70	26.57
***No*STR2**	41.18	33.13	33.43	29.86	29.86	27.20
***No*STR3**	41.52	33.44	33.74	29.86	29.86	27.20

Table S8. The primers used in this study.

Primer names	Primers	Target plasmid
28a-*No*LAMT1-F	agcaaatgggtcgcggatccATGGCCCCAACAATGGACAC	pET28a-*No*LAMT1
28a-*No*LAMT1-R	gcggccgcaagcttgtcgacTCAATTGCTCTTACGTTTTAGAACAAGGAATAAATCAG	
28a-*No*LAMT2-F	agcaaatgggtcgcggatccATGGAAGTGCACCAAGTTCTTC	pET28a-*No*LAMT2
28a-*No*LAMT2-R	gcggccgcaagcttgtcgacTCACTTTTTTGCAAGGGAAACAACCAAATTATC	
pCold-TF-*No*STR1-F	tcggtaccctcgagggatccATGTCAATGCTCACCATTTTCATTCTTG	pCold-TF-*No*STR1
pCold-TF-*No*STR1-R	tatctagactgcaggtcgacTCAGTAAGAGAAAGACTCCTTAGTATATTCGTTAAGTC	
pCold-TF-*No*STR2-F	tcggtaccctcgagggatccATGGTTGCGTTAACCATTTTTTTGATCC	pCold-TF-*No*STR2
pCold-TF-*No*STR2-R	tatctagactgcaggtcgacTCAGACAGAAGAAGCCACTCCATTCAAAG	
pCold-TF-*No*STR3-F	tcggtaccctcgagggatccATGGTTGCGTTAACCATTTTTTTGATCC	pCold-TF-*No*STR3
pCold-TF-*No*STR3-R	tatctagactgcaggtcgacTCAGACAGAAGAAGCCACTCCATTCAAAG	
pCold-TF-*No*STR4-F	tcggtaccctcgagggatccATGCTTCAAAACTTCAAGAAACTCCTACTTC	pCold-TF-*No*STR4
pCold-TF-*No*STR4-R	tatctagactgcaggtcgacTTAATCATTCTTCACCTCTATCCTACAATTCTTCCTTG	
pCold-TF-*No*STR5-F	tcggtaccctcgagggatccATGACGCTGACCGGATTCTTTG	pCold-TF-*No*STR5
pCold-TF-*No*STR5-R	tatctagactgcaggtcgacTCACTCTAATTGGTAGACTGCAACAAAAGG	
pCold-TF-*No*STR6-F	tcggtaccctcgagggatccATGACGCTGACCGGATTCTTTG	pCold-TF-*No*STR6
pCold-TF-*No*STR6-R	tatctagactgcaggtcgacTCACTCTAATTGGTAGACTGCAACAAAAGG	
pCold-TF-*No*STR7-F	tcggtaccctcgagggatccATGATAGTGCCGAAACGCAACTC	pCold-TF-*No*STR7
pCold-TF-*No*STR7-R	tatctagactgcaggtcgacTTATGCATTAGCGCTGACTGCAGG	
pCold-TF-*No*STR8-F	tcggtaccctcgagggatccATGTCCGAGTCGACTCGCAAC	pCold-TF-*No*STR8
pCold-TF-*No*STR8-R	tatctagactgcaggtcgacTTATGTATTAGCGCTGACTGCAGGATTTTG	
pCold-TF-*No*STR9-F	tcggtaccctcgagggatccATGGATTTTAGCGGTTGCTTCATGAC	pCold-TF-*No*STR9
pCold-TF-*No*STR9-R	tatctagactgcaggtcgacCTATTGCAGAATATGATATACTCCTAGCGGTAATTTTCC	
pCold-TF-*No*STR2-E309S-F	GGTGAACATTTCTCGCAAGTTCAAGAGCATAATGGTGCACTC	pCold-TF-*No*STR2-E309S
pCold-TF-*No*STR2-E309S-R	CTCTTGAACTTGCGAGAAATGTTCACCTTTATATGGTGGAGGG	
pCold-TF-*No*STR4-EHFEQ-F	TTCTGAGTATGAAAATGAACATTTCGAACAAGTGAATGAACATCAAGGTTCG	pCold-TF-*No*STR4-EHFEQ
pCold-TF-*No*STR4-EHFEQ-R	GATGTTCATTCACTTGTTCGAAATGTTCATTTTCATACTCAGAATCAAGCCTAAT	
pEAQ-*No*LAMT1-F	tgcccaaattcgcgaccggtATGGCCCCAACAATGGACAC	pEAQ-*No*LAMT1
pEAQ-*No*LAMT1-R	ccagagttaaaggcctcgagTCAATTGCTCTTACGTTTTAGAACAAGGAATAAATCAG	
pEAQ-*No*LAMT2-F	tgcccaaattcgcgaccggtATGGAAGTGCACCAAGTTCTTCG	pEAQ-*No*LAMT2
pEAQ-*No*LAMT2-R	ccagagttaaaggcctcgagTCACTTTTTTGCAAGGGAAACAACCAAATTATCG	
pEAQ-*No*STR1-F	tgcccaaattcgcgaccggtATGTCAATGCTCACCATTTTCATTCTTG	pEAQ-*No*STR1
pEAQ-*No*STR1-R	ccagagttaaaggcctcgagTCAGTAAGAGAAAGACTCCTTAGTATATTCGTTAAGTC	
pEAQ-*No*STR2-F	tgcccaaattcgcgaccggtATGGTTGCGTTAACCATTTTTTTGATCC	pEAQ-*No*STR2
pEAQ-*No*STR2-R	ccagagttaaaggcctcgagTCAGACAGAAGAAGCCACTCCATTCAAAG	
pEAQ-*No*STR3-F	tgcccaaattcgcgaccggtATGGTTGCGTTAACCATTTTTTTGATCC	pEAQ-*No*STR3
pEAQ-*No*STR3-R	ccagagttaaaggcctcgagTCAGACAGAAGAAGCCACTCCATTCAAAG	
pEAQ-*No*STR4-F	tgcccaaattcgcgaccggtATGCTTCAAAACTTCAAGAAACTCCTACTTC	pEAQ-*No*STR4
pEAQ-*No*STR4-R	cagagttaaaggcctcgagGATTAATCATTCTTCACCTCTATCCTACAATTCTTCCTTG	
pEAQ-*No*STR5-F	tgcccaaattcgcgaccggtATGACGCTGACCGGATTCTTTG	pEAQ-*No*STR5
pEAQ-*No*STR5-R	ccagagttaaaggcctcgagTCACTCTAATTGGTAGACTGCAACAAAAG	

pEAQ-*No*STR6-F	tgcccaaattcgcgaccggtATGACGCTGACCGGATTCTTTG	pEAQ-*No*STR6
pEAQ-*No*STR6-R	ccagagttaaaggcctcgagTCACTCTAATTGGTAGACTGCAACAAAAG	
pEAQ-*No*STR7-F	tgcccaaattcgcgaccggtATGATAGTGCCGAAACGCAACTC	pEAQ-*No*STR7
pEAQ-*No*STR7-R	ccagagttaaaggcctcgagTTATGCATTAGCGCTGACTGCAGG	
pEAQ-*No*STR8-F	tgcccaaattcgcgaccggtATGTCCGAGTCGACTCGCAAC	pEAQ-*No*STR8
pEAQ-*No*STR8-R	ccagagttaaaggcctcgagTTATGTATTAGCGCTGACTGCAGGATTTTG	
pEAQ-*No*STR9-F	tgcccaaattcgcgaccggtATGGATTTTAGCGGTTGCTTCATGACATTTTC	pEAQ-*No*STR9
pEAQ-*No*STR9-R	ccagagttaaaggcctcgagCTATTGCAGAATATGATATACTCCTAGCGGTAATTTTC	
pEAQ-*No*SLS1-F	tgcccaaattcgcgaccggtATGGAGATGGATATTGTCCATCAGG	pEAQ-*No*SLS1
pEAQ-*No*SLS1-R	ccagagttaaaggcctcgagTCACTCGAGCTTGCGGTAGATCAC	
pEAQ-*No*SLS2-F	tgcccaaattcgcgaccggtATGGAAATGAACTTCAGCTCAGTTGC	pEAQ-*No*SLS2
pEAQ-*No*SLS2-R	ccagagttaaaggcctcgagCTAGAGCTTGTGCAAGATCAAGTGAG	
pEAQ-*No*SLS3-F	tgcccaaattcgcgaccggtATGGAAATGAACTTCAGCTCAGTTGC	pEAQ-*No*SLS3
pEAQ-*No*SLS3-R	ccagagttaaaggcctcgagCTAGAGCTTGTGCAAGATCAAGTGAG	
pEAQ-*No*SLS4-F	tgcccaaattcgcgaccggtATGGAGATAGCCTACAACCCCATAGC	pEAQ-*No*SLS4
pEAQ-*No*SLS4-R	ccagagttaaaggcctcgagTCACACAAGCACATTCTGAAGATTATGCAAAATC	
pEAQ-*No*SLS5-F	tgcccaaattcgcgaccggtATGGATACGTTTAACAGCATAATTGGAGTTTTATG	pEAQ-*No*SLS5
pEAQ-*No*SLS5-R	ccagagttaaaggcctcgagCTACAGTTTCCTCAAAATCAACTGAGC	
pEAQ-*No*SLS6-F	tgcccaaattcgcgaccggtATGAGACTAGTGGAGAACATGTTAAGCATAG	pEAQ-*No*SLS6
pEAQ-*No*SLS6-R	ccagagttaaaggcctcgagCTACAGTTTGTGTAAAACTAAGTGAGCACCATG	
pEAQ-*No*SLS7-F	tgcccaaattcgcgaccggtATGGAAACTGCAGCCTATAGCTTGATTATTG	pEAQ-*No*SLS7
pEAQ-*No*SLS7-R	ccagagttaaaggcctcgagCTACAATTTGTGCAAAATTAAATGAGCTCCATACTG	
pEAQ-*No*SLS8-F	tgcccaaattcgcgaccggtATGGAAGGATCTATACTGAAGGGCTTGG	pEAQ-*No*SLS8
pEAQ-*No*SLS8-R	ccagagttaaaggcctcgagTTAAATTTGTTGCAATATGATCGGAGCTCC	

Table S9. Nucleotide sequences of MIAs biosynthetic genes in *N. orientalis*.

Gene	Sequence (5' → 3')
NoLAMT1	ATGGCCCCAACAATGGACACCATCCCATCTGTGGATATCAAAGAAATGCCTGAAGAAGCACACCCCATGAAAGGTGGTGATGACCTCAATAGCTATTCTCA AAATTCTAGCTATCAGAGGGGAGTTATAGATGCTGCAAAAACAGTCATTATTGAAGCAGTAACTGAAAAGCTGAATCTTGAGAATAATGTTAATTTTAATCC ATCTAAGCCTTTCCGCATTGCTGATTTTGGGTGCTCAACCGGCCCTAACACATATTTTGCCATGCAAAATGTTGTAGAAGCCGTGGGCCAAAAATACAAATC TCTGCAAGGAAAAAGCCAGCCAGAATTCCATGTTTTTTTCAACGATCATGTCAACAATGATTTCAACATTCTCTTCAGATCTCTCCCACAAGACAAGAATTAT TTTGCTGCCGGGGTCCCTGGATCATTTTACACAAGGGTTTTTCCCAAGGCCAGTTTGCATTTTGCTCATTGCTCATATGCACTTCACTGGCTATCCAAGGTTC CCAAGGAAATCCAGGACAAGAATTCTCCTGCCTTTAACAGAGGGAGAATTCACTACACTGGGACTGAAAAACATGTTGTGAAAGCTTACTTTTCTCAATTC CAAAAAGACATGGGCTCATTTCTGAAAGCTAGAGCTCAAGAAATTGTGGGAGGAGGGTTGATGGTGATTCAAATTCCTGGCCTTCCAAGTGGTGAGGTC CTCTTCTCTAGGACTGGTGCTGGCATGCTTCATGCTCTTTTGGGATCTTCCTTGATGGAACTCGTAAATTTGGGAGTAATAAATGAAGAAAGTGTAGATTCC TTCAACTTGCCTCAATACCATCCCTCAATTGAGGAATTGGAAATGGTGATAGAGATGAACAATAGTTTTACCCTTGAAAAATTTGGAGCATTGAATCACCCC ATGAAAAATTTACCATTTGATGTTCAAATGACTTCTCTACAAGTAAGAGCTGTCATGGAGGGTATTCTAGGTGATCATTTTGGAAAAAAAATCTTGGATCAAT TGTTCGAAATCTACACCAAGAAACTGCAAGAAAATTACAACGTCTTCGACAAGGAAATCCGAAGAGATGCTGATTTATTCCTTGTTCTAAAACGTAAGAGC AATTGA
NoSLS1	ATGGAGATGGATATTGTCCATCAGGCCATTGCTGCCACTTGCTTTCTTGCACTTTTTGCTTGGGCATGGAGAGTCCTGAACTGGGCTTGGTTTACACCTAAG AGCATCGAGAAGCGTCTCAGACAACAAGGTTTCAGAGGAAACCCTTATAGATGCTTGGTTGGAGATGTTAAAGAGAGCGGTATGATGCTCCAAGAAGCC ATGTCTAAGCCCATCCCACTCAACAACGATATCGTTCCTCGTATCATGCCTCACATTGACAAAACCATCAAGACCTATGGTAAGAATTCCTTTACCTGGATGG GACGTATTCCTAGGATTCACGTGATGGATCCCGATCTCATTCGGGAAGTCTTGACACACTCAAACAAATTCATGAAGAATTTTGATGTACACAATCCACTCGT TAAGTTCCTACTCACTGGAGTCGGAAGCTTTGAGGGTGAAAAATGGTCCAAACATAGAAGAATTATTTCCCCAGCCTTCACTCTTGAGAAGTTGAAGACA ATGTTGCCTGCTTTTGCGGTTTGCTATGACGATCTTTTGCGTAAATGGGAGCAAACTGCCACAAGAAAGGGATCCGTTGAGGTGGATATCTTTCCAACATTC GATGTCCTAACAAGTGATGTGATTTCCAAGGTTGCCTTTGGTAGCACCTACGAGGAAGGTGGCAAAATCTTCCAACTTCTTAAAGAACTTATGGAGCTGAC AATTCAGACCATGAGAGATGTATACATTCCAGGATGGAGTTTATTGCCAACAAAGAGGAACAAGAGGATGAAAGAAATTAACAACGAGATCACAGACATG TTGAGGAACATCATCAACAAGAGAGTTAAGGCAATGAAGGCTGGAGAACCAAGTGAGGATGATTTGCTTGGTGTTCTTTTGGAATCCAACTACCAAGAA ATTCAAAAGCAAGGAAACAACAAGAATGTTGGAATGACTATTGAGGATGTCATTGAGGAATGCAAACTTTTCTATTTTGCTGGACAAGAAACCACAGGTG TGCTTCTCACTTGGACAACCATTCTATTGAGCAAGCACCCCGAGTGGCAAGAGCGTGCAAGAGAAGAAGTCTTTCAAGCCTTCGGCAAGAACAAGCCCG AATTTGACAAATTGAATCACCTCAAATATGTTTCAATGATTTTGTTCGAGGTCCTAAGGTTGTATCCCCCGGTCATTGACCTGACCAAAATTGTTCACGAAGA TACCAAGCTAGGACCATACACCATCCCAGCTGGAACCCAAATTATGTTGCCAACAGTAATGCTTCACAGGGAGAAGAGCATTTGGGGAGAAGATGCAATG GAATTCAATCCGGAAAGGTTTGCAGATGGAGTTGCTAATGCAACCAAGAACCAAGTCACGTATCTGCCATTCAGCTGGGGACCTCGTGTCTGCCTGGGCC AGAATTTTGCACTATTGCAGGCCAAGTTGGGGCTGGCAATGATCTTACAGCGCTTCTCCTTCGATGTTTCGCCATCCTATGCTCATGCTCCTTACACCATCCT AACTCTTCAACCCCAACTTGGTTCTCATGTGATCTACCGCAAGCTCGAGTGA
NoSTR1	ATGTCAATGCTCACCATTTTCATTCTTGGATTCTCCTTTTCATTTGTTCTCTCTACTCCAATCTTAAAATCCATTCAAGCTCCCTACGGTCCCAACGCCTTCGCT TTCGACTCGGCCGGTGGACTTTACGCTGCCGTGGAAGATGGTAGAGTTGTCAAATATGGAGGATCAGGCATTGGTTTTGTGAACCACGCTTATGCCTCTCC TTTCTGGAATGGAACAATTTGTAAGAGCAAGAATACCACGATTGAAGCACACCTAAAGCCCATCTGTGGGAGGGTATACGACTTGGGATTCAACCATGCG ACAAAGATGCTTTATACAGTTGATGGCTATTATGGTCTTGGCATGGTTGGACCTGATGGAGGGCTTGCCATACAACTTGCCAATGGTGCAGAGGGAGGTGA ATTCAAGTGGCTTTATGCCTTAACTGTGGATCAGAGAACTGGGATTGTTTATTTCACGGAAGTAAGCAGAGTCTACGATGACAGAGGTGTGCAGGATATAC TAGATACAAACGACACAACCGGAAGATTATTGAAGTATGATCCCTCAACTAAGAAAGTGACAGTTTTGATGAAAGGGTTAAATGTACCAGGTGGTACGGA ACTCAGCAAAGACAGCTCCTTTGTTCTTGTTGCTGAATATCTGGGCAACAAAATTATCAAGTATTGGCTAGAGGGTCCTAAAGCAAATTCTTCTGAGATTTT GTTGGAAATTCCAGGTCCAGGGAACATAAAGAGGACTCCAACTGGAGATTTTTGGGTGCCATCCACTGATGATAATGCACCAATACTAACATCCGTAGGAG TAAGATTCAACAAATTTGGCAGAATATTGGAAACTTTGGATCTTCCACAGCCATATAAAAGCGAACATCTTGAGCAAATTCATGAGCATAATCATGAACTTTA TGTTGGATCATTGTTCCACAGCTTCATTGGTGTTCTGGAGAGGGATTCGCTTGATCAACAGAGGGAATACAGTCTAGATGGACTTAACGAATATACTAAGG AGTCTTTCTCTTACTGA
NoSTR2	ATGGTTGCGTTAACCATTTTTTTGATCCTGTTCTTGTCTCCTCTTTCACCTGTTCTATCTTCGCCGGAATTTTTCGAATTCCTCAAGTCACCCTACGGCCCCAA CGCCTTCGCTTTTAACTCAGCTGGTGAACTTTACGCTTCCGTCGAAGACGGCAGAATTGTTAAGTATAAACGATCACGCAACGGGTTCTCGACCCACGCTG TTGCCTCTCCAGTCTGGAAGGGAGAAGTTTGTGAGAATAATACTAACCCTCAGCTAAAACCCGTTTGTGGGAGGGTATATGACCTTGGATTTCACTATGAA ACTCAGCAGTTGTATATTGCTGATTGTTATTATGGTCTTGGTACGGTTGGCCCTGAAGGAGGCCGTGCAACTCAACTTGCTAAGAGTGCAGATGGAGTGGA CTTCAAGTGGCTTTATGCCTTAGCCGTGGACCAGCAAACTGGCTTTGTTTACCTCACTGATGTTAGCACAAAATATGATGACAGAGGTGTTCAAGACATCCT AAGGATAAATGATACAACAGGAAGATTAATCAAATATGATCCCTCAACTAATGAAGCTAGAGTTTTAATCAATGGGCTGAATGTACCAGGTGGTACCGAAGT TAGCAAAGATGGCTCGTTTGTTGTTGTTGCTGAATTCTTGAGCCACAGAATTCTCAAGTTTTGGTTAAAGGGTCCTAAGGCAAATACTTCTGAGGAATTATT GAAAGTGAGGGGCCCAGGTAACATAAAGCGGACCCATACTGGAGAATTTTGGGTAGCCTCTAGTGACAATAATGGAATTACGGTTACGCCTAGGGGTATA AAGTTTGATGATTTTGGCAATATTTTAGAAGTCGTGCCTATCCCTCCACCATATAAAGGTGAACATTTCGAACAAGTTCAAGAGCATAATGGTGCACTCTACA TCGGGTCTTTGTTCCATGACTTTGTGGGTATATTACACAATTACGAGGGTTCATCCGATCCCAAAGAAAATAATGTAGATGTGGTCAATGGATCTTTGAATGG AGTGGCTTCTTCTGTCTGA
NoSTR3	ATGGTTGCGTTAACCATTTTTTTGATCCTGTTCTTGTCTCCTCTTTCACCTGTTTTATCTTCGCCGGAATTTTTCGAATTCCTCAAGTCACCCTACGGCCCCAA CGCCTTCGCTTTTAACTCGGCTGGTGAACTTTACGCTTCCGTCGAAGACGGCAGAATTGTTAAGTATAAAGGATCACGCAACGGGTTCTCGACCCACGCTG TTGCCTCTCCAGTCTGGAAGGGAGAAGTTTGTGAGAATAATACTAACCCTCAGCTAAAACCCTTTTGTGGGAGGGTATATGACCTTGGATTTCACTATGAAA CTCAGCAGTTGTATATTGCTGATTGTTATTATGGTCTTGGTACGGTTGGACCTGAAGGAGGCCGTGCAACTCAACTTGCTAAGAGTGCAGATGGAGTGGAC TTCAAGTGGCTTTATGCCTTAGCCGTGGACCAGCAAACTGGCTTTGTTTACCTCACTGATGTTAGCACAAAATACGATGACAGAGGTGTTCAAGACATCCTA AGGATAAATGATACAACAGGAAGATTAATCAAATATGATCCCTCAACTAATGAAGCTAGAGTTTTAATCAATGGGCTGAATGTACCAGGTGGTACCGAAGTT AGCAAAGATGGCTCGTTTGTTGTTGTTGCTGAATTCTTGAGCCACAGAATTCTCAAGTTTTGGTTAAAGGGTCCTAAGGCAAATACTTCTGAGGAATTATT GAAAGTGAGGGGCCCAGGTAACATAAAGCGGACCCATACTGGAGAATTTTGGGTAGCCTCTAGTGACAATAATGGAATTACGGTTACGCCTAGGGGTATA AAGTTTGATGATTTTGGCAATATTTTAGAAGTCGTGCCTATCCCTCCACCATATAAAGGTGAACATTTCGAACAAGTTCAAGAGCATAATGGTGCACTCTACA TCGGGTCTTTGTTCCATGACTTTGTGGGTATATTACACAATTACGAGGGTTCATCCGATCCCAAAGAAAATAATGTAGATGTGGTCAATGGATCTTTGAATGG AGTGGCTTCTTCTGTCTGA

References

1 B. Buchfink, C. Xie and D. H. Huson, *Nat. Methods.,* **2015**, *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

2 P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez and S. Hunter, *Bioinformatics.,* **2014**, *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

3 L. S. Johnson, S. R. Eddy and E. Portugaly, *BMC Bioinf.,* **2010**, *11*, 431. https://doi.org/10.1186/1471-2105-11-431.

4 K. D. Pruitt, T. Tatusova and D. R. Maglott, *Nucleic Acids Res.,* **2007**, *35*, D61–D65. https://doi.org/10.1093/nar/gkl842.

5 The UniProt Consortium, *Nucleic Acids Res.,* **2023**, *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

6 The Gene Ontology Consortium, *Genetics,* 2023, *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.

7 M. Y. Galperin, Y. I. Wolf, K. S. Makarova, R. V. Alvarez, D. Landsman and E. V. Koonin, *Nucleic Acids Res.,* **2021**, *49*, D274–D281. https://doi.org/10.1093/nar/gkaa1018.

8 E. V. Koonin, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, D. M. Krylov, K. S. Makarova, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, I. B. Rogozin, S. Smirnov, A. V. Sorokin, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin and D. A. Natale, *Genome Biol.,* **2004**, *5*, R7. https://doi.org/10.1186/gb-2004-5-2-r7.

9 J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn and A. Bateman, *Nucleic Acids Res.,* **2021**, *49*, D412-D419. https://doi.org/10.1093/nar/gkaa913.

10 M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe, *Nucleic Acids Res.,* **2016**, *44*, D457–462. https://doi.org/10.1093/nar/gkv1070.

11 A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras, *Bioinformatics.,* **2013**, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

12 N. L. Bray, H. Pimentel, P. Melsted and L. Pachter, *Nat. Biotechnol.,* **2016**, *34*, 525–527. https://doi.org/10.1038/nbt.3519.

13 J. D. Thompson, T. J. Gibson and D. G. Higgins, *Curr. Protoc. Bioinf.,* **2002**, *00*, 2.3.1–2.3.22 https://doi.org/10.1002/0471250953.bi0203s00.

14 X. Robert and P. Gouet, *Nucleic Acids Res.,* **2014**, *42*, W320–W324. https://doi.org/10.1093/nar/gku316.

15 K. Tamura, G. Stecher and S. Kumar, *Mol. Biol. Evol.,* **2021**, *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120.

16 I. Letunic and P. Bork, *Nucleic Acids Res.,* **2021**, *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

17 C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He and R. Xia, *Mol. Plant.,* **2020**, *13*, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009.

18 Y.-J. Wang, J.-P. Huang, T. Tian, Y. Yan, Y. Chen, J. Yang, J. Chen, Y.-C. Gu and S.-X. Huang, *J. Am. Chem. Soc.,* **2022**, *144*, 22000–22007. https://doi.org/10.1021/jacs.2c09091.

19 C. A. Erdelmeier, U. Regenass, T. Rali and O. Sticher, *Planta Med.,* **1992**, *58*, 43–48. https://doi.org/10.1055/s-2006-961387.

20 C. A. J. Erdelmeier, A. D. Wright, J. Orjala, B. Baumgartner, T. Rali and O. Sticher, *Planta Med.,* **1991**, *57*, 149–152. https://doi.org/10.1055/s-2006-960052.

21 Z.-D. He, C.-Y. Ma, H.-J. Zhang, G. T. Tan, P. Tamez, K. Sydara, S. Bouamanivong, B. Southavong, D. D. Soejarto, J. M. Pezzuto and H. H. S. Fong, *Chem. Biodiversity,* **2005**, *2*, 1378–1386. https://doi.org/10.1002/cbdv.200590110.

22 T. Kanchanapoom, P. Sahakitpichan, N. Chimnoi, C. Petchthong, W. Thamniyom, P. Nangkoed and S. Ruchirawat, *Phytochem. Lett.,* **2021**, *41*, 83-87. https://doi.org/10.1016/j.phytol.2020.11.016.

23 Y.-P. Liu, P.-K. Ju, J.-T. Long, L. Lai, W.-H. Zhao, C. Zhang, Z.-J. Zhang and Y.-H. Fu, *Nat. Prod. Res.,* **2018**, *32*, 2922–2927. https://doi.org/10.1080/14786419.2017.1395429.

24 J. Sichaem, S. Surapinit, P. Siripong, S. Khumkratok, J. Jong-aramruang and S. Tip-pyang, *Fitoterapia,* **2010**, *81*, 830–833. https://doi.org/10.1016/j.fitote.2010.05.004.

25 J. Sichaem, W. Worawala and S. Tip-pyang, *Chem. Nat. Compd.,* **2012**, *48*, 827–830. https://doi.org/10.1007/s10600-012-0393-z.

26 W. Songoen, J. Brunmair, F. Traxler, V. C. Wieser, W. Phanchai, W. Pluempanupat, L. Brecker and J. Schinnerl, *Molecules,* **2022**, *27*, 5176. https://doi.org/10.3390/molecules27165176.

27 Z. Zhang, H. N. ElSohly, M. R. Jacob, D. S. Pasco, L. A. Walker and A. M. Clark, *J. Nat. Prod.,* **2001**, *64*, 1001–1005. https://doi.org/10.1021/np010042g.

28 Q. Pan, N. R. Mustafa, K. Tang, Y. H. Choi and R. Verpoorte, *Phytochem. Rev.,* **2015**, *15*, 221–250. https://doi.org/10.1007/s11101-015-9406-4.

29 X. Ma, S. Panjikar, J. Koepke, E. Loris and J. Stöckigt, *Plant Cell,* **2006**, *18*, 907-920. https://doi.org/10.1105/tpc.105.038018.

30 E. Eger, A. Simon, M. Sharma, S. Yang, W. B. Breukelaar, G. Grogan, K. N. Houk and W. Kroutil, *J. Am. Chem. Soc.*, 2020, 142, 792–800. https://doi.org/10.1021/jacs.9b08704.

31 G. Wang, H. Wang, Z. Lin, L. Hou, J. Y. Wang and L. Sun, *J. Ethnopharmacol.,* **2022**, *282*, 114560. https://doi.org/10.1016/j.jep.2021.114560.

32 K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. van der Krol, R. Lugan, T. Ilc, R. Verpoorte, K.-M. Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink and D. Werck-Reichhart, *Nat. Commun.,* **2014**, *5*, 3606. https://doi.org/10.1038/ncomms4606.

33 H. Liu, S. Liang, M. Zhu, W. Shi, C. Xu, W. Wei, R. Zhan and D. Ma, *Planta,* **2024**, *259*, 62. https://doi.org/10.1007/s00425-023-04287-w.

34 G. Collu, N. Unver, A. M. G. Peltenburg-Looman, R. van der Heijden, R. Verpoorte and J. Memelink, *FEBS Lett.,* **2001**, *508*, 215–220. https://doi.org/10.1016/S0014-5793(01)03045-9.

35 Y. Li, X. Wang, H. Jiang, S. Xu, Y. Xu, Z. Liu and Y. Luo, *Plant Physiol. Biochem.,* **2025**, *218*, 109305. https://doi.org/10.1016/j.plaphy.2024.109305.

36 V. Salim, B. Wiens, S. Masada-Atsumi, F. Yu and V. De Luca, *Phytochemistry,* **2014**, *101*, 23–31.

https://doi.org/10.1016/j.phytochem.2014.02.009.
37 V. Salim, F. Yu, J. Altarejos and V. De Luca, *Plant J.,* **2013**, *76*, 754–765. https://doi.org/10.1111/tpj.12330.