Originally posted 10 September 2015; corrected 16 September 2015

www.sciencemag.org/content/349/6253/1224/suppl/DC1

Supplementary Materials for

Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone

Warren Lau and Elizabeth S. Sattely*

*Corresponding author. E-mail: sattely@stanford.edu

Published 11 September 2015, *Science* **349**, 1224 (2015)
DOI: 10.1126/science.aac7202

This PDF file includes

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/349/6253/1224/suppl/DC1)

Correction (16 September 2005): In forming the PDF, the Table of Contents was garbled, and page 5 is impossible to read. These translation errors are corrected in the revised SM. The originally posted version can be seen here.

TABLE of CONTENTS

Materials and Methods

Transcriptomics data mining and analysis of the Medicinal Plants Consortium data sets

RNA-Sequencing data from *P. hexandrum* was downloaded from the Medicinal Plants Consortium database (http://www.medplantrnaseq.org/, Podophyllum_hexandrum_2.tgz) (16). The number of fragments per kilobase of contig per million mapped reads (FPKM) for each contig and condition (leaf-1, leaf-2, rhizome-1, rhizome-2, stem-1, stem-2, stem-3) was determined using the number of paired aligned reads provided in the .counts files. Contigs representing known (–)-podophyllotoxin biosynthetic genes were determined and annotated by tblastn with query sequences obtained from the NCBI database: (*DIR,* AIA24413.1; *PLR,* ABY75535.2; *SDH*, ABN14311.1; *CYP719A23*, AGC29953.1). Contigs were also annotated by blastx using an Arabidopsis proteome (The Arabidopsis Information Resource) database. Approximately 100 O-methyltransferase (*OMT)* and 300 cytochrome P450 (*CYP)* contigs were identified. Multiple contigs were found to represent the same gene. *OMT* gene candidates from this data set were chosen based on the availability of full coding sequence, similar expression profiles to known biosynthetic genes, and similarity (best blastx hit and % identity) to known plant genes involved in secondary metabolism (*OMT1*, *Ph13451* [Contig number, Ph refers to contigs from the Medicinal Plants Consortium dataset]; *OMT2*, *Ph14232*; *OMT3*, *Ph18546*; *OMT4*, *Ph6083*). *CYP* gene candidates (See Table S2 for primers) were chosen similarly. For biosynthetic steps that occur downstream of (–)-deoxypodophyllotoxin formation, CYP contigs with rhizome-specific expression were considered.

mRNA extraction and cDNA template preparation

RNA was extracted from *P. hexandrum* var. chinense ex MD97150 (Far Reaches Farm, Port Townsend, WA) and *P. hexandrum* (Dancing Oaks Nursery, Monmouth, OR) leaf and rhizome tissues using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions. cDNA was prepared from extracted mRNA using the SuperScript III First-Strand Synthesis System (Invitrogen).

qRT-PCR analysis of *OMT3* and known podophyllotoxin biosynthetic genes after *P. hexandrum* leaf wounding

A *P. hexandrum* (Dancing Oaks Nursery, Monmouth, OR) leaf was removed from the stem with scissors and 0 h time point (control) tissue samples were collected (one from each leaflet, three total from a single leaf), flash frozen with liquid nitrogen and stored at -80^{o}C for later manipulation. The leaf was then immediately wounded by piercing the leaf approximately a hundred times with fine tweezers throughout the entire surface of the leaf. The wounded leaf was placed in a petri dish with distilled water and moved to a growth chamber under a 16 h light cycle (photon flux of 100 µmol m^{-2} s^{-1}, 22°C, 50% relative humidity) approximately an hour after the start of the light cycle. Three tissue samples, one from each leaflet, were removed from the leaf 3, 6, 9, 12 and 24 h after wounding, flash frozen by liquid nitrogen and stored at -80^{o}C for later use in metabolomics analysis, gene profiling and RNA-Sequencing.

cDNA templates were prepared from the isolated tissues as described above. Quantitative RT-PCR was performed using cDNA templates and gene-specific primers for *DIR, PLR, SDH, CYP719A23, OMT3* and *ACTIN* (Table S2 for primers), (25) . Each cDNA was amplified by real-time PCR using SensiMix SYBR Hi-ROX Kit (Bioline) and the ABI StepOnePlus (Applied Biosystems). *ACTIN* expression was used to normalize the expression values in each sample, and relative expression values were determined against the 0 h time point (before wounding) using the comparative Ct method ($2^{-\Delta\Delta Ct}$), (26). Analysis was repeated once with another biological sample; similar results were obtained.

RNA-Seq library preparation, next-generation sequencing and co-expression analysis

A multiplexed RNA-Seq cDNA library was prepared from RNA isolated from the *P. hexandrum* wounding experiment (three biological replicates of 0, 3, 9 and 12 h after wounding, 12 samples total) using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's instructions. The quality and average length (insert size was approximately 200 bp) of cDNAs in the library were determined using a High Sensitivity DNA chip on a 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced (paired-end, 100 bp) on a single lane of HiSeq2000 (Illumina) at the Sequencing Service Center by the Stanford Center for Genomics and Personalized Medicine. The FASTX-toolkit (27) was used for quality assessment and cleaning of reads. Reads were trimmed at the 5'-end by 13 bp to remove biases associated with random priming. Reads with Phred quality scores of less than 20 were trimmed from the 3'-end; reads were discarded if resulting length was less than 50 bp. Low complexity sequences were removed. A transcriptome was assembled *de novo* using Velvet and Oases (28) with a k-mer of 71. The clustering tool CD-HIT-EST (29) was used to identify sequences with greater than 99% identity from the assembly; only the longest representative transcript was kept. The set of transcripts was further assembled with CAP3 (30) to combine contigs with significant overlaps (minimum 95% identity over at least 100 bp). Assembled transcripts were annotated by blastx using the Arabidopsis proteome (The Arabidopsis Information Resource) as a database and the NCBI non-redundant database, if necessary. eXpress (31) was used to quantify gene expression levels under each experimental condition by mapping paired reads to assembled transcripts. Effective counts were TMM (Trimmed Mean of M-values) normalized by edge (32). Contigs assembled using this data are annotated Phex followed by the contig number.

RNA-Seq co-expression analysis and candidate selection

On the basis of the predicted enzyme activities required for the missing pathway steps, the transcriptome data was mined for gene sequences annotated *O*-methyltransferases (*OMT*), cytochromes P450 (*CYP*), 2-oxoglutarate/Fe(II)-dependent dioxygenases (*2-ODD*), and polyphenol oxidases (*PPO*). For these 336 genes, linear regression analysis (33) of TMM-normalized reads was performed using *DIR* as a bait gene (fig. S10). Genes were ranked in order of increasing p-value. *OMT3* (rank 2 and 3) and *SDH* (rank 28) were among the genes whose expression profiles most closely matched that of *DIR*. Statistical tests were considered significant using the Benjamini–Hochberg (34) procedure with a false discovery rate of 0.05 (multiple testing corrections). For *DIR,* genes ranked less than 33 were significant. Known pathway genes were highly expressed compared to other genes (1-2 orders of magnitude difference in number of normalized read counts; median TMM normalized effective counts were greater than 500). An expression level cutoff was used (contigs with median TMM normalized effective counts < 500 were removed). After filtering by median expression level, 14 genes remained. From these candidates, four potential gene candidates were chosen: *Phex30848* (*2-ODD*), *Phex32688* (*CYP*), *Phex13114* (*OMT1*, previously tested), and *Phex359* (*PPO*). We chose to focus on the identification of an oxidative enzyme first based on the next predicted biosynthetic step and therefore, additional *OMT* candidates were not chosen. The qRT-PCR gene profiling after leaf wounding from above revealed two similar but distinct expression patterns (*DIR, SDH* and *OMT3* vs. *PLR* and *CYP719A23*), suggesting separate transcriptional regulons. Therefore, linear regression analysis was also performed using *CYP719A23* as a bait gene as above (fig. S11). Genes ranked less than 26 were considered significant using the Benjamini–Hochberg procedure with a false discovery rate of 0.05. After filtering by expression level cutoff (contigs with median TMM normalized effective counts < 500 were removed), 13 genes remained. Two additional candidates were chosen from this list: *Phex34339 (PPO)* and *Phex524 (CYP71CU1)*. As before, we chose to focus on the identification of an oxidative enzyme first based on the next predicted biosynthetic step and therefore, additional *OMT* candidates were not chosen.

Phex15199 (CYP) was also chosen because it showed an interesting elicitation pattern distinct from that observed from known pathway genes; this candidate did not satisfy any of the prior criteria.

To further validate candidate selection, average linkage, hierarchical clustering analysis was performed with CLUSTER (version 3.0) using expression data (TMM normalized, log_2-scaled, and median-centered effective counts) from transcript sequences annotated as cytochromes P450 (*CYP*), O-methyltransferases (*OMT*), Fe(II) and 2-oxoglutarate dependent dioxygenases (*2-ODD*), polyphenol oxidases (*PPO*) and known (–)-podophyllotoxin biosynthetic genes (*DIR, PLR, SDH, CYP719A23* and *OMT3*) (Fig. 2B and Table S1). Contigs with median TMM normalized effective read counts less than 500 were removed. Six of our seven candidates, three of four known pathway genes and *OMT3* were contained in a single clade of 22 genes. As before, multiple transcripts were found to represent the same gene. The heat map was visualized by Treeview. When computationally taxing, bioinformatics work was performed on a computational cluster operated by the Stanford Genetics Bioinformatics Service Center.

Cloning of podophyllotoxin biosynthetic genes and candidate genes

Phusion High-Fidelity DNA Polymerase (Thermo Scientific) was used for all PCR amplification steps according to the manufacturer's instructions. All other enzymes used for cloning were purchased from New England Biolabs. Oligonucleotide primers were purchased from Integrated DNA Technologies. DNA excised from agarose gels was purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research). *E. coli* TOP10 cells (Invitrogen) were used for plasmid isolation prior to transformation into other heterologous hosts. Plasmid DNA was isolated from *E. coli* cultures using the QIAprep Spin Miniprep Kit (Qiagen). For a list of primers used for cloning, see Table S2.

For *N. benthamiana* transient expression, all gene sequences were amplified from *P. hexandrum* cDNA template. Purified amplicons were inserted into pEAQ-HT (35) (Kan^R) plasmid digested with AgeI and XhoI in an isothermal DNA assembly reaction, as described by Gibson et al (36).

For *N. benthamiana* transient expression of multiple genes from the same *Agrobacterium* strain, a multi-gene construct was created using Golden Gate assembly. A Golden Gate dedicated pEAQ-HT plasmid (pEAQ-HT-GG) was constructed by PCR amplification and Gibson assembly to remove the region containing promoter, 5' UTR, multi-cloning site, 3' UTR and terminator sequences, and to insert two XbaI restriction sites linked to two distinct recognition sites separated by a linker sequence. Gene sequences were amplified from pEAQ-HT vectors carrying the gene of interest starting from promoter and ending with terminator and flanked with XbaI restriction sites linked with distinct recognition sites. Purified amplicons were inserted into pEAQ-HT-GG by Golden Gate assembly as described by Engler et al. (37) to construct pEAQ-HT-GG:*DIR-PLR-SDH* and pEAQ-HT-GG:*CYP719A23-OMT3* (Figure S29).

For the expression of CYPs in *S. cerevisiae*, the sequences were amplified from cloned pEAQ-HT vectors carrying the *CYP* sequence of interest and the purified amplicons were individually inserted into pYeDP60 (Amp^R) plasmid digested with BamHI and XhoI in a Gibson assembly reaction.

For the expression of OMT3 in *E. coli*, the sequence was amplified from pEAQ-HT:*OMT3*. The purified amplicon was inserted into pET28a (Kan^R) plasmid digested with NdeI and XhoI by Gibson assembly. The final construct was designed to express OMT3 with an N-terminal 6xHis tag.

For the expression of OMT1 and 2-ODD in *E. coli*, the sequences were amplified from their respective pEAQ-HT vectors. The purified amplicons were individually inserted into pET24b (Kan^R) plasmid digested with NdeI and XhoI by Gibson assembly. The final constructs were designed to express the enzymes with a C-terminal 6xHis tag.

The assembly reaction mixtures were used directly to transform *E. coli* TOP10 cells and the isolated plasmids harboring the desired insert were confirmed by Sanger DNA sequencing performed by Elim Biopharm.

Transient expression and candidate gene screening in *N. benthamiana*

pEAQ-HT constructs were transformed into *Agrobacterium tumafaciens* (GV3101) using the freeze-thaw method. Transformants were grown on LB plates containing 50 μg/mL kanamycin and 30 μg/mL gentamicin at 30°C. Cells were removed with a 1 mL pipette tip and resuspended in 1 mL of LB medium, centrifuged at 5000 g for 5 min and supernatant removed. Pellet was resuspended in 10 mM MES buffer, pH = 5.6, 10 mM $MgCl_2$, 150 μM acetosyringone and incubated at room temperature for 1 h. *Agrobacterium* suspensions (OD_{600} = 0.3 for each strain) were infiltrated into the underside of *N. benthamiana* leaves with a needleless 1 mL syringe. Plants were grown 4-5 weeks under a 16 h light cycle prior to infiltration. Leaves were harvested 5 days post-infiltration, flash frozen and stored at -80°C for later processing. Biological replicates consisted of several leaves all from different tobacco plants. Infiltrated leaf areas typically showed some signs of chlorosis or yellowing; leaves expressing GFP as a control also showed a similar phenotype.

For substrate infiltration studies, generally, 100 μM of (–)-matairesinol in 0.1% DMSO in water was infiltrated into the underside of previously *Agrobacterium*-infiltrated leaves with a needleless 1 mL syringe 4 days post-infiltration. Substrate infiltration was no more difficult than *Agrobacterium* infiltration. Leaves were harvested 1 day later, flash frozen and stored at -80°C for later processing.

Pathway reconstitution in *N. benthamiana*

Full reconstitution of the (–)-deoxypodophyllotoxin pathway starting from (+)-pinoresinol in *N. benthamiana* was achieved by transient expression of DIR (not required), PLR, SDH, CYP719A23, OMT3, CYP71CU1, OMT1 and 2-ODD (fig. S25) following the transient expression procedure above using an OD_{600} = 0.3 for each strain. Transient expression of GFP (OD_{600} = total OD_{600} for all pathway genes) was used as a control. For substrate infiltration, 100 μM (+)-pinoresinol (~10 μg per leaf) in 0.1% DMSO in water was infiltrated into the underside of previously *Agrobacterium*-infiltrated leaves with a needleless 1 mL syringe 4 days post-infiltration. Leaves were harvested 1 day later, flash frozen and stored at -80°C for later processing.

Full reconstitution of the (–)-epipodophyllotoxin (fig. S26) and (–)-4'-desmethyl-epipodophyllotoxin pathway (Fig. 3B) starting from (+)-pinoresinol in *N. benthamiana* was achieved by the additional transient expression of CYP82D61, and CYP82D61 and CYP71BE54, respectively.

Standard curves for (–)-deoxypodophyllotoxin and (–)-epipodophyllotoxin based on UV absorption at 280 nm, and (–)-4'-desmethyl-epipodophyllotoxin based on ion abundance were constructed and used to quantify amounts produced in *N. benthamiana*.

Metabolite extraction

Frozen leaf tissues were lyophilized to dryness. The samples were homogenized on a ball mill (Retsch MM 400) using 5 mm diameter stainless steel beads, shaking at 25 Hz for 2 min. 20 μL for tobacco or 250 μL for *P. hexandrum* of an 80:20 MeOH/H_2O solution was added per milligram of dry tissue, and the mixture was heated at 65°C for 10 min and filtered through 0.45 μm PTFE filters before liquid chromatography-mass spectrometry (LC-MS) analysis.

For initial studies in the reconstitution of the (–)-pluviatolide pathway in tobacco, β-glucosidase treatment was performed before metabolite analysis. The leaf extracts from above were briefly spun in a microcentrifuge and 200 μL of supernatant was removed and dried under a stream of nitrogen. Residues were resuspended with 200 μL of 0.1 M sodium acetate, pH = 5.0 with 2 mg/mL of β-glucosidase from almonds (Sigma-Aldrich) and incubated at 37°C for 16 h. The solutions were extracted with 2 volumes of ethyl acetate and the organic phase was dried under a stream of nitrogen. 200 μL of an 80:20 MeOH/H_2O

solution was added to the residue, and the mixture was heated at 65°C for 10 min and analyzed by LC-MS analysis.

LC-MS analysis of metabolite extracts

Metabolomics samples were analyzed by reversed-phase chromatography on an Agilent 1260 HPLC, using a 5 μm, 2 × 100 mm Gemini NX-C18 column (Phenomenex). Water with 0.05% acetic acid (A) and ACN with 0.05% acetic acid (B) [negative ion mode] or water with 0.1% FA (A) and ACN with 0.1% FA (B) [positive ion mode] were used as the mobile phase components at a flow rate of 0.4 mL/min with the following 41 min gradient: 0-30 min, 3-50% B; 30-31 min, 50-97% B; 31-36 min, 97% B; 36-37 min, 97-3% B; 37-41 min, 3% B. A coupled Agilent 6520 Accurate-Mass Q-TOF ESI mass spectrometer was used to collect MS data in either positive ion mode or negative ion mode (parameters: mass range: 100-1700 m/z; drying gas: 300°C, 1 L/min; nebulizer: 25 psig; capillary: 3500 V; fragmentor: 150 V; skimmer: 65 V; octupole 1 RF Vpp: 750 V; 1000 ms per spectrum). The first minute of each run was discarded to avoid salt contamination of the MS apparatus. For tandem mass spectrometry (MS/MS) analysis, 5, 10, 20 and 40 V collision energies were used with an m/z window of 1.3 centered on the m/z analyzed.

Metabolomics and MS data analysis

HRMS data were analyzed using MassHunter Qualitative Analysis software (Agilent) and XCMS (Scripps Center for Metabolomics) (38). For untargeted metabolomics, MassHunter (Agilent) data files were converted to mzXML format using trapper (Seattle Proteome Center). Grouped mzXML files were preprocessed and analyzed by XCMS, using the following sample R script:

```
library(xcms)
xset<-xcmsSet()
xset<-group(xset)
xset2<-retcor(xset,family="s",plottype="m")
xset2<-group(xset2)
xset3<-retcor(xset2,family="s",plottype="m")
xset3<-group(xset3)
xset4<-retcor(xset3,family="s",plottype="m")
xset4<-group(xset4,bw=10)
xset5<-fillPeaks(xset4)
reporttab<-diffreport(xset5,"cyp719a23+omt3","cyp719a23","cyp719a23+omt3 vs cyp719a23",2000)
```

The resulting report contains a mass peak list with m/z values, peak intensity fold change, statistical significance (p value, two-tailed unequal variance Student's t-test), retention times and extracted peak intensities. The list was filtered using cutoff criteria (generally, p value less than 0.05, t value greater than 0, fold change greater than 5, retention time less than 1800 s and average peak intensity greater than 5×10^4 in the experimental sample group [in the above case, cyp719a23+omt3]) to yield a narrow list for further analysis. A summary of XCMS results can be found in Figure S30.

Generally, ion abundances were determined by manual integration of EICs using the MassHunter Qualitative Analysis software. For Fig. S4, they were derived from XCMS analysis.

Heterologous expression and purification of OMT3, OMT1 and 2-ODD

pET28a:*OMT3* was transformed into BL21 (DE3) *E. coli*. Transformants were grown on LB plates containing 50 μg/mL kanamycin. A single colony was inoculated into 40 mL of LB medium containing 50 μg/mL kanamycin and grown for 16 h at 30°C. The overnight culture was inoculated into 2 L of LB medium containing 50 μg/mL kanamycin and grown at 37°C until OD_{600} = 0.6, at which point the culture

was cooled to 25°C, and induced with 500 μM IPTG. The culture was further incubated at 25°C for 6 h, cooled in ice water and centrifuged at 10,800 g for 10 min at 4°C. The supernatant was discarded and the pellet was resuspended in 40 mL of lysis buffer (50 mM sodium phosphate, pH = 8.0 and 300 mM NaCl) containing 10 mM imidazole and 20 mM of β-mercaptoethanol. The cell slurry was passed through a French press (Thermo) three times and centrifuged at 38,700 g for 30 min at 4°C. All subsequent manipulations were performed at 4°C. The supernatant was incubated with 1 mL of Ni-NTA agarose resin (Qiagen) pre-equilibrated with lysis buffer containing 10 mM imidazole for 1 h. The slurry was run through a fritted column and protein was eluted with lysis buffer containing increasing imidazole concentrations up to 400 mM. Fractions containing purified protein as determined by SDS-PAGE gel were combined, concentrated and buffer exchanged into 50 mM sodium phosphate, pH = 7.5, 100 mM NaCl, 2 mM DTT and 10% glycerol using Amicon Ultra-15 centrifugal filter units, 10kDa cutoff (Millipore). Total protein content was estimated by measuring UV absorbance at 280 nm on a NanoDrop 1000 spectrophotometer. Purified protein was flash-frozen as pellets in liquid nitrogen and kept at –80°C for long term storage.

The above protocol was also followed for the expression and purification of OMT1 from BL21 (DE3) *E. coli* harboring pET24b:*OMT1*.

The above protocol was also followed for the expression and purification of 2-ODD from BL21 (DE3) *E. coli* harboring pET24b:*2-ODD* except for the following changes: the 2 L culture at OD_{600} = 0.6 was cooled to 16°C and induced with 100 μM IPTG. The culture was further incubated at 16°C for 24 h and cooled in ice water before proceeding with purification.

Expression of cytochromes P450 and microsome isolation from *S. cerevesiae* WAT11

The pYeDP60 constructs were individually transformed into *Saccharomyces cerevisiae* WAT11 (carrying a chromosomal copy of the *A. thaliana* ATR1 cytochrome P450 reductase gene) using the lithium acetate method (39). Yeast growth, induction and microsome preparation were performed according to Pompon et al (40). Microsomes enriched with P450 were isolated in 50 mM Tris-Cl, pH = 7.4, 20% glycerol, 1 mM EDTA, flash-frozen as pellets in liquid nitrogen and kept at –80°C for long term storage. Total protein content was estimated by measuring UV absorbance at 280 nm on a NanoDrop 1000 spectrophotometer.

***In vitro* characterization of recombinant OMT3**

For steady state enzyme kinetics, enzyme assays (140 μl reaction volume) contained 10 μg/mL of purified OMT3, 1 mM SAM, (–)-pluviatolide at various concentrations (0.5, 1, 1.5, 2, 3, 5, 10 and 20 μM from a 100% DMSO stock; final concentration of DMSO was kept constant at 5%) and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4). Reactions were initiated by the addition of enzyme and incubated at 22°C. Assays lacking enzyme or SAM served as negative controls. 20 μL aliquots from the reaction were quenched by the addition of 1 volume of acetonitrile with 0.1 % TFA at various time points (0.5, 1, 1.5, 2, 3, 4, 6, 8 and 16 min; time points were chosen based on the starting substrate concentration), and clarified by centrifugation for 10 min at 17,000 g. Quenched samples were analyzed by reversed-phase chromatography on an Agilent 1100 HPLC with an Agilent 1260 diode array detector and a 2.7 μm, 3 × 50 mm Poroshell 120 EC-C18 column (Agilent). Water with 0.1% TFA (A) and ACN with 0.1% TFA (B) were used as the mobile phase components at a flow rate of 0.6 mL/min with the following gradient: 0-2 min, 3-20% B; 2-8 min, 20-60% B; 8-8.5 min, 60-97% B; 8.5-10 min, 97% B; 10-11 min, 97-3% B; 11-15 min, 3% B. The absorption at 280 nm was monitored to detect (–)-pluviatolide and (–)-5'-desmethoxy-yatein, and the peak area response observed was analyzed by Chemstation software (Agilent). A standard curve for (–)-pluviatolide was made in order to determine its extinction coefficient. Because the extinction coefficient for (–)-5'-desmethoxy-yatein was estimated to be similar to (–)-pluviatolide due to a similar change in absorbance during kinetic assays, kinetic rates were determined by measuring the rate

of (–)-pluviatolide consumption. Experiments were repeated in triplicate and the data points were fit to a linear line using linear regression with GraphPad Prism 6. Only the linear portion of the data, dependent on initial substrate concentration, was used. The initial velocity of the reaction was determined from the slope of the fitted line. The kinetic constants for OMT3 and (–)-pluviatolide, apparent K_m and v_{max}, were determined by non-linear regression.

For determining substrate specificity, (+)-pinoresinol (400 μM), (–)-matairesinol (40 μM) and (–)-arctigenin (40 μM) were also tested as substrates under the same assay conditions. Reactions were quenched after 2 h.

In vitro assays of CYP71CU1, OMT1, 2-ODD and CYP71BE54

CYP71CU1-microsome assays contained 1 mg/mL of CYP71CU1-microsomes, 1 mM NADPH, 30 or 50 μM of substrate and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4). Assays with microsomes from WAT11 harboring an empty vector or lacking NADPH served as negative controls. OMT1 enzyme assays contained 10 μg/mL of purified OMT1, 1 mM SAM, 30 or 50 μM of substrate and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4). Assay lacking SAM served as negative control. Enzyme assays with both CYP71CU1-microsomes and OMT1 contained 1 mg/mL of CYP71CU1-microsomes, 1 mM NADPH, 10 μg/mL of purified OMT1, 1 mM SAM, 30 or 50 μM of substrate and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4).

2-ODD enzyme assays contained 5 μg/mL of purified 2-ODD, 10 mM 2-oxoglutarate, 10 mM ascorbic acid, 0.5 mM $FeSO_4$, 30 μM of substrate and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4). Assays lacking enzyme or 2-oxoglutarate served as negative controls.

CYP71BE54-microsome assays contained 4 mg/mL of CYP71BE54-microsomes, 1 mM NADPH, 50 μM of substrate and 100 mM NaCl in sodium phosphate buffer (50 mM, pH 7.4). Assays with microsomes from WAT11 harboring an empty vector or lacking NADPH served as negative controls.

All substrates [(–)-matairesinol, (–)-arctigenin, (–)-pluviatolide, (–)-5'-desmethoxy-yatein, (–)-yatein, (–)-deoxypodophyllotoxin, and (–)-podophyllotoxin] were derived from a 100% DMSO stock; final concentration of DMSO in enzyme assays was kept constant at 5%. Total reaction volumes were 50 μL. Reactions were initiated by the addition of enzyme and incubated at 30°C. After 2 h of incubation, reactions were quenched by the addition of 1 volume of acetonitrile with 0.1 % TFA and clarified by centrifugation for 10 min at 17,000 g. Quenched samples were either analyzed by HPLC as above for the *in vitro* characterization of OMT3 or by LC-MS as above for the analysis of plant metabolite extracts.

Chemicals

Commercially available chemicals

(+)-pinoresinol was purchased from ArboNova (Turku, Finland). (–)-podophyllotoxin was purchased from Sigma-Aldrich. NMR solvents were purchased from Cambridge Isotope Laboratories.

General procedures

Compounds were assayed for purity by LC-MS and ^{1}H NMR. ^{1}H NMR spectra were acquired at room temperature on a Varian 400 MHz spectrometer. Shifts are reported in parts per million downfield from tetramethylsilane and referenced to the residual solvent peak. All NMR spectra can be found in Figures S31-S38.

Isolation of (–)-matairesinol and (–)-arctigenin from *Forsythia x intermedia*

Leaves from *Forysthia x intermedia* 'Northern Sun' (Forestfarm, Williams, OR) were harvested, flash frozen and lyophilized (5.6 g). Leaves were crushed into powder form with mortar and pestle in liquid nitrogen and heated at 65°C for 45 min in 300 mL of methanol. The extract was filtered and dried under reduced pressure. The residue was resuspended with 100 mL of 0.1 M sodium acetate, pH = 5.0 containing 1 mg/mL of β-glucosidase from almonds (Sigma-Aldrich) and incubated at 37°C for 24 h. The solution was extracted 3 times with 1 volume of ethyl acetate and the organic phase was filtered and dried under reduced pressure. The residue was further purified by preparative HPLC using an Agilent 1260 Infinity preparative-scale HPLC system with an Agilent 1100 diode array detector and a TARGA C18 10 µm 250 × 20 mm column (Higgins Analytical). Water with 0.1% TFA (A) and acetonitrile with 0.1% TFA (B) were used as the mobile phase components at a flow rate of 36 mL/min with the following method: 0-5 min, 3-25% B; 5-20 min, 25-40% B; 20-22 min, 40% B; 22-23 min, 40-97% B; 23-25 min, 97% B; 25-26 min, 97-3% B. 84 mg of (–)-matairesinol and 94 mg of (–)-arctigenin were isolated.

(–)-matairesinol: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.40 – 2.64 (m, 4H), 2.87 (dd, *J* = 14.1, 6.9 Hz, 1H), 2.95 (dd, *J* = 14.1, 5.2 Hz, 1H), 3.80 (s, 3H), 3.81 (s, 3H), 3.88 (dd, *J* = 9.1, 7.3 Hz, 1H), 4.07 – 4.19 (m, 1H), 5.57 (s, 1H), 5.59 (s, 1H), 6.40 (d, *J* = 1.9 Hz, 1H), 6.50 (dd, *J* = 8.0, 1.9 Hz, 1H), 6.58 – 6.60 (m, 1H), 6.61 (s, 1H), 6.79 (d, *J* = 7.8 Hz, 1H), 6.81 (d, *J* = 8.0 Hz, 1H); UV/Vis: Relative absorbance maximum at λ = 282 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{20}H_{23}O_6^+$, 359.1489; found, 359.1490; $[\alpha]_D$ = -32.3 (c = 0.62, acetone) {lit (41). $[\alpha]_D$ = -44 (c = 0.62, acetone)}.

(–)-arctigenin: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.40 – 2.68 (m, 4H), 2.89 (dd, *J* = 14.1, 6.6 Hz, 1H), 2.94 (dd, *J* = 14.1, 5.3 Hz, 3H), 3.80 (s, 6H), 3.84 (s, 3H), 3.88 (dd, *J* = 9.6, 1.8 Hz, 1H), 4.06 – 4.17 (m, 1H), 5.60 (s, 1H), 6.45 (d, *J* = 2.1 Hz, 1H), 6.54 (dd, *J* = 8.1, 2.1 Hz, 1H), 6.58 – 6.64 (m, 2H), 6.74 (d, *J* = 8.1 Hz, 1H), 6.81 (d, *J* = 7.9 Hz, 1H); UV/Vis: Relative absorbance maximum at λ = 280 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{21}H_{25}O_6^+$, 373.1646; found, 373.1641.

Synthesis of (–)-pluviatolide using *S. cerevisiae* WAT11 cells expressing *CYP719A23*

Following Pompon et al. [40] and Giddings et al. [42], a 1 L culture of *S. cerevisiae* WAT11 harboring pYeDP60:*CYP719A23* was grown and induced, and 6 h after induction, the culture was supplemented with (–)-matairesinol at a final concentration of 100 µM from a 100 mM stock in DMSO. 24 h after induction, the culture was centrifuged at 10,800 g for 15 min at 4°C. The supernatant was extracted 3 times with 1 volume of ethyl acetate and the organic phase was filtered and concentrated under reduced pressure. The desired product was further purified from the crude mixture by silica gel flash column chromatography with ethyl acetate:hexanes (1:1) and preparative HPLC using the same method used for (–)-matairesinol and (–)-arctigenin isolation. Compound was assayed for purity by LC-MS and ^{1}H NMR. 5 mg of (–)-pluviatolide was isolated.

(–)-pluviatolide: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.39 – 2.65 (m, 4H), 2.88 (dd, *J* = 14.1, 6.8 Hz, 1H), 2.95 (dd, *J* = 14.0, 5.0 Hz, 1H), 3.84 (s, 3H), 3.85 – 3.91 (m, 1H), 4.06 – 4.15 (m, 1H), 5.92 (s, 2H), 6.41 – 6.48 (m, 2H), 6.58 – 6.71 (m, 3H), 6.83 (d, *J* = 8.0 Hz, 1H); UV/Vis: Relative absorbance maximum at λ = 284 nm; HRMS (m/z): $[M + H]^+$ calcd. for $C_{20}H_{21}O_6^+$, 357.1333; found, 357.1326.

Isolation of OMT3 enzymatic product [(–)-5'-desmethoxy-yatein]

A scaled-up enzyme reaction was utilized to generate sufficient amounts of product for ^{1}H NMR analysis. The reaction mixture, containing 50 mM sodium phosphate (pH=7.4), 100 mM NaCl, 280 µM pluviatolide, 1 mM SAM and 50 µg/mL of OMT3, was incubated at room temperature for 3 h. The reaction was extracted 3 times with 1 volume of ethyl acetate and the organic phase was dried with sodium sulfate, filtered and dried under reduced pressure. The residue was further purified by preparative

HPLC using the same method used for (–)-matairesinol and (–)-arctigenin isolation. Compound was assayed for purity by LC-MS and ^{1}H NMR. 4.4 mg of (–)-5'-desmethoxy-yatein was isolated.

(–)-5'-desmethoxy-yatein: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.39 – 2.65 (m, 4H), 2.89 (dd, *J* = 14.1, 7.0 Hz, 1H), 2.97 (dd, *J* = 14.1, 5.1 Hz, 1H), 3.83 (s, 3H), 3.86 (s, 3H), 3.87 – 3.89 (m, 1H), 4.12 (dd, J = 9.1, 6.7 Hz, 1H), 5.92 (d, J = 1.5 Hz, 1H), 5.93 (d, J = 1.4 Hz, 1H), 6.42 – 6.48 (m, 2H), 6.63 – 6.71 (m, 3H), 6.76 – 6.81 (m, 1H); Relative absorbance maximum at λ = 284 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{21}H_{23}O_6^+$, 371.1489; found, 371.1480.

Isolation of (–)-deoxypodophyllotoxin and (–)-yatein from *Anthriscus sylvestris*

Shoots and rhizomes from *A. sylvestris* (Digging Dog Nursery, Albion, CA) were harvested, washed to remove soil, and lyophilized (~20 g dry weight). Dried plant material was soaked in ~10 mL of methanol per g of tissue for 16 h and then sonicated for 30 min. Methanolic extracts were removed and an additional 10 mL of methanol per g of tissue was added, sonicated for 30 min and repeated once more. The combined methanolic extracts were filtered and dried under reduced pressure. The residue was partially purified by preparative HPLC as above using the following method: 0-30 min, 3-50% B; 30-31 min, 50-97% B; 31-35 min, 97% B; 35-36 min, 97-3% B. The crude mixture was further purified by silica gel flash column chromatography with a mobile phase of ethyl acetate:hexanes (3:7 → 1:1). The remaining crude product was separated once more by preparative HPLC yielding 29.4 mg of (–)-deoxypodophyllotoxin and 3.5 mg of (–)-yatein. Compounds were assayed for purity by LC-MS and ^{1}H NMR.

(–)-yatein: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.41 – 2.66 (m, 4H), 2.83 – 2.97 (m, 2H), 3.83 (s, 9H), 3.88 (dd, *J* = 9.2, 7.5 Hz, 1H), 4.18 (dd, *J* = 9.1, 7.2 Hz, 1H), 5.93 (d, *J* = 1.5 Hz, 1H), 5.94 (d, *J* = 1.4 Hz, 1H), 6.35 (s, 2H), 6.43 – 6.50 (m, 2H), 6.69 (d, *J* = 7.7 Hz, 1H); Relative absorbance maximum at λ = 286 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{22}H_{25}O_7^+$, 401.1595; found, 401.1586; $[\alpha]_D$ = -16.6 (c = 0.32, chloroform) {lit (43). $[\alpha]_D$ = -28.4 (c = 0.32, chloroform)}.

(–)-deoxypodophyllotoxin: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.69 – 2.82 (m, 3H), 3.01 – 3.12 (m, 1H), 3.73 (s, 6H), 3.79 (s, 3H), 3.85 – 3.97 (m, 1H), 4.41 – 4.51 (m, 1H), 4.58 (d, *J* = 3.2 Hz, 1H), 5.92 (d, *J* = 1.3 Hz, 1H), 5.94 (d, *J* = 1.4 Hz, 2H), 6.33 (s, 2H), 6.50 (s, 1H), 6.66 (s, 1H); Relative absorbance maximum at λ = 292 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{22}H_{23}O_7^+$, 399.1438; found, 399.1435; $[\alpha]_D$ = -81.8 (c = 0.63, methanol) {lit (44). $[\alpha]_D$ = -66.6 (c = 0.63, methanol)}.

Synthesis of (–)-epipodophyllotoxin and (–)-4'-desmethyl-epipodophyllotoxin

Following Kamal et al. (45), (–)-epipodophyllotoxin and (–)-4'-desmethyl-epipodophyllotoxin were synthesized in two separate reactions starting with 10 mg (–)-podophyllotoxin. Crude reactions were purified by silica gel flash column chromatography with a mobile phase of ethyl acetate:hexanes (3:2), and further purified by preparative HPLC using the same method used for (–)-yatein and (–)-deoxypodophyllotoxin isolation, yielding 7.9 mg of (–)-epipodophyllotoxin and 1.3 mg of (–)-4'-desmethyl-epipodophyllotoxin. Compounds were assayed for purity by LC-MS and ^{1}H NMR.

(–)-epipodophyllotoxin: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.76 – 2.90 (m, 1H), 3.28 (dd, *J* = 14.1, 5.2 Hz, 1H), 3.73 (s, 6H), 3.79 (s, 3H), 4.31 – 4.43 (m, 2H), 4.61 (d, *J* = 5.2 Hz, 1H), 4.87 (d, *J* = 3.4 Hz, 1H), 5.97 (d, *J* = 1.3 Hz, 1H), 6.00 (d, *J* = 1.3 Hz, 1H), 6.27 (s, 2H), 6.54 (s, 1H), 6.88 (s, 1H); Relative absorbance maximum at λ = 288 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{22}H_{23}O_8^+$, 415.1387; found, 415.1388.

(–)-4'-desmethyl-epipodophyllotoxin: ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.73 – 2.86 (m, 1H), 3.25 (dd, *J* = 14.2, 5.1 Hz, 1H), 3.75 (s, 6H), 4.30 – 4.41 (m, 2H), 4.59 (d, *J* = 5.0 Hz, 1H), 4.82 – 4.87 (m,

1H), 5.38 (s, 1H), 5.95 (d, J = 1.7 Hz, 1H), 5.98 (d, J = 1.7 Hz, 4H), 6.27 (s, 2H), 6.53 (s, 1H), 6.85 (s, 1H); Relative absorbance maximum at λ = 284 nm ; HRMS (m/z): $[M + H]^+$ calcd. for $C_{21}H_{21}O_8^+$, 401.1231; found, 401.1235.

Supplementary Figures

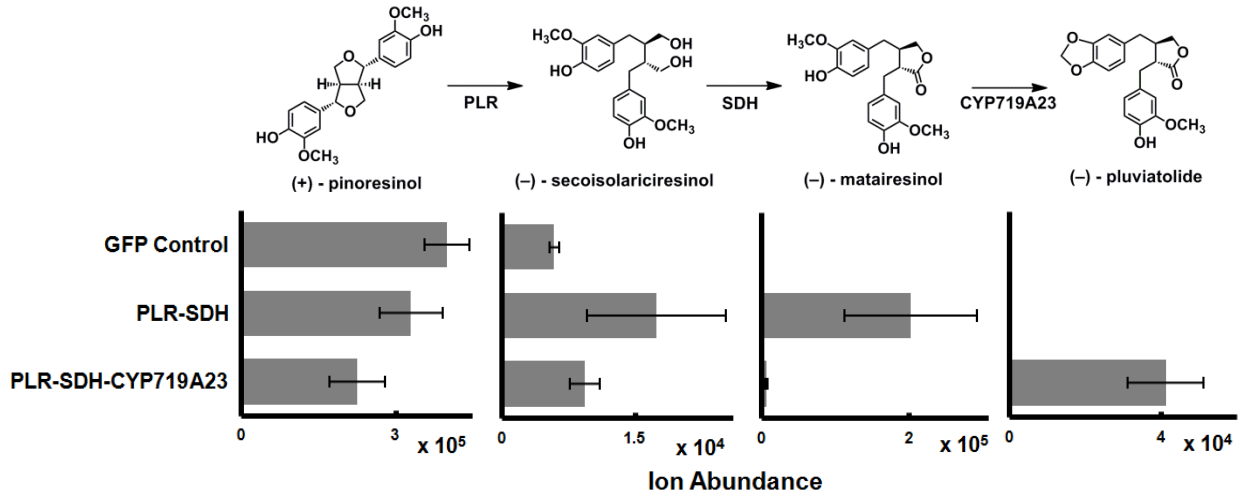


Fig. S1. Transient expression of early (–)-podophyllotoxin biosynthetic genes in *N. benthamiana*.

Transient expression of combinations of PLR, SDH and CYP719A23 in *N. benthamiana* leaves produces (–)-matairesinol (m/z = 357) and (–)-pluviatolide (m/z = 355), which are not normally observed in tobacco leaves (GFP only control). Data bars indicate the mean ion abundance [(–)-mode] ± 1 standard deviation, based on three biological replicates. (+)-pinoresinol (m/z = 357) and (–)-secoisolariciresinol (m/z = 361, no authentic standard) are natively produced in *N. benthamiana* leaves. Metabolite extracts were treated with β-glucosidase before LC-MS analysis.

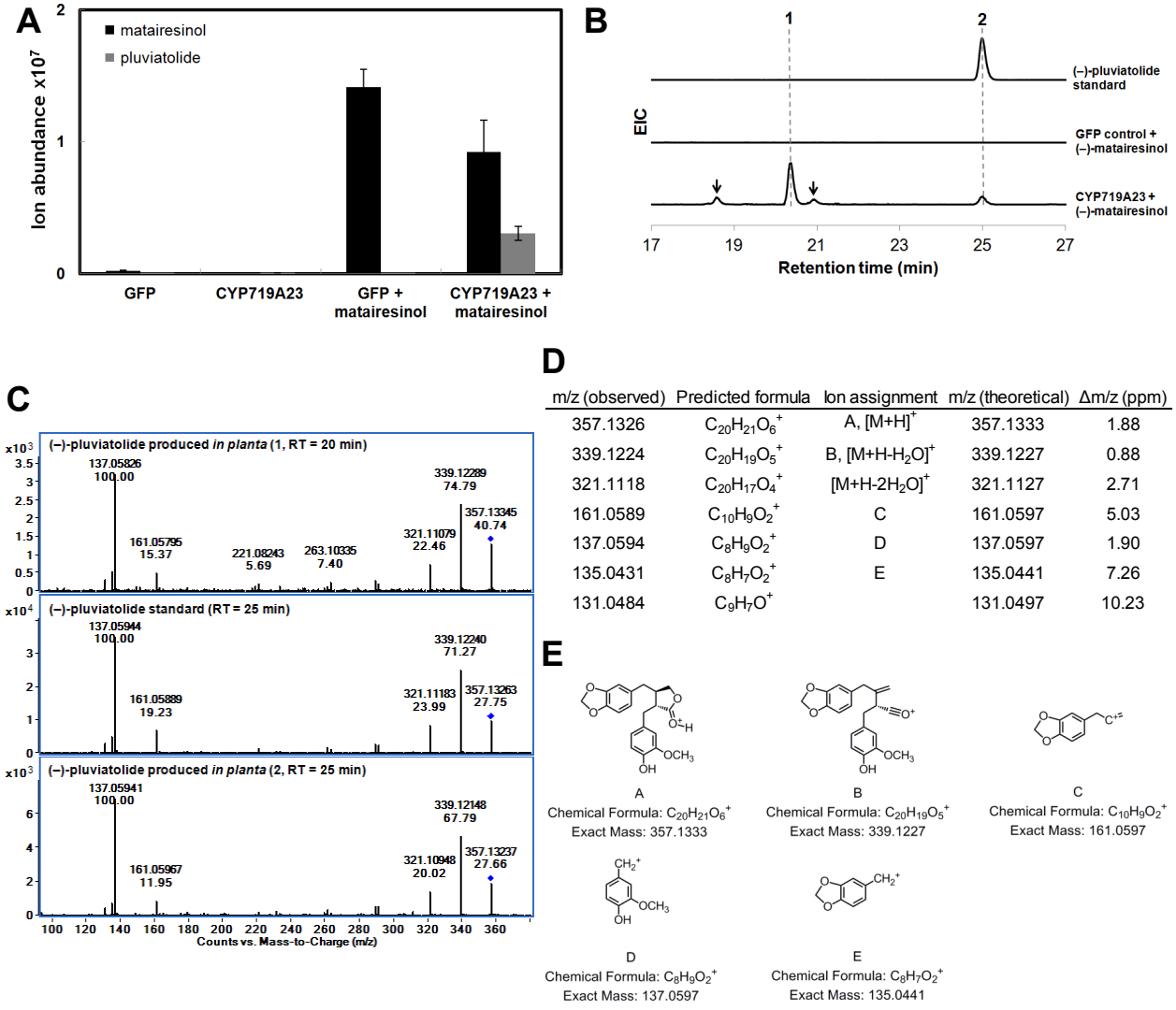
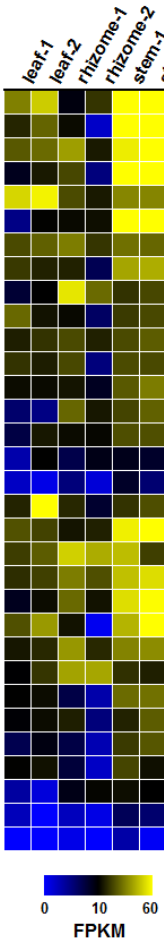


Fig. S2. (–)-Matairesinol infiltration enhances (–)-pluviatolide production in tobacco leaves expressing CYP719A23

(A) Infiltration of (–)-matairesinol (m/z = 357) into *N. benthamiana* leaves transiently expressing CYP719A23 alone drastically increases (–)-pluviatolide (m/z = 355) levels. No substrate infiltration and GFP only controls are shown for comparison. Data bars indicate the mean ion abundance [(–)-mode] ± 1 standard deviation, based on three biological replicates. **(B)** Extracted ion chromatograms (EIC) of (–)-pluviatolide (m/z = 357) from the infiltration of (–)-matairesinol into *N. benthamiana* leaves transiently expressing CYP719A23 are shown and compared to GFP control and authentic standard [(+)-mode]; peaks observed at earlier labeled retention times (1 and arrows) are in-source fragments of (–)-pluviatolide glucosylated derivatives produced by endogenous tobacco enzymes. Only chromatograms from *in planta* experiments are to scale. **(C)** MS/MS (10 V) spectra of (–)-pluviatolide produced *in planta* (retention times, 1 and 2) compared to authentic standard [(+)-mode] **(D)** Table of MS/MS peak assignments and **(E)** putative ion structures.



Contig	Gene	Best blastx hit as of March 2015	% Identity
14232	*OMT2*	norcoclaurine 6-O-methyltransferase [Podophyllum peltatum]	83
18546	*OMT3*	norcoclaurine 6-O-methyltransferase [Podophyllum peltatum]	65
13451	*OMT1*	caffeic acid O-3-methyltransferase [Pinus pinaster]	59
11943	*CYP*	PREDICTED: cytochrome P450 CYP736A12 [Vitis vinifera]	55
3089	*CYP*	PREDICTED: flavonoid 3'-monooxygenase-like [Nelumbo nucifera]	69
8214	*CYP*	PREDICTED: cytochrome P450 71A1-like [Nelumbo nucifera]	56
6083	*OMT4*	PREDICTED: caffeic acid 3-O-methyltransferase [Nelumbo nucifera]	79
3787	*CYP*	PREDICTED: geraniol 8-hydroxylase-like [Nelumbo nucifera]	83
2327	*CYP*	CYP81B57 [Sinopodophyllum hexandrum]	100
4178	*CYP*	PREDICTED: cytochrome P450 89A2-like [Nelumbo nucifera]	63
4466	*CYP*	cytochrome P450 [Coptis japonica var. dissecta]	73
5690	*2-ODD*	PREDICTED: 1-aminocyclopropane-1-carboxylate oxidase homolog 4-like [Nelumbo nucifera]	63
5422	*CYP*	PREDICTED: cytochrome P450 71A1-like [Nelumbo nucifera]	73
2456	*CYP*	PREDICTED: flavonoid 3'-monooxygenase-like [Fragaria vesca subsp. vesca]	54
2443	*CYP*	CYP81B56 [Sinopodophyllum hexandrum]	100
2002	*CYP*	(S)-N-methylcoclaurine 3'-hydroxylase isozyme 2 [Morus notabilis]	53
172938	*CYP*	putative N-methylcoclaurine 3-hydroxylase [Podophyllum peltatum]	84
8051	***DIR***		
225796	***CYP719A23***		
102408	***CYP719A23***		
17736	***CYP719A23***		
26599	***SDH***		
212985	***CYP719A23***		
6901	***PLR***		
140193	***PLR***		
12248	***SDH***		
31365	***CYP719A23***		
34018	***CYP719A23***		
38318	***CYP719A23***		
12826	***SDH***		
145347	***CYP719A23***		
193171	***CYP719A23***		

Fig. S3. Candidate genes selected for initial screening from the Medicinal Plants Consortium P. hexandrum transcriptome.

Candidates (grey and red) were chosen based on similar expression profiles to known (–)-podophyllotoxin biosynthetic genes (black), similarity to known enzymes involved in plant secondary metabolism based on best blastx hit, and percent identity and availability of full coding sequences. In red are (–)-podophyllotoxin biosynthetic genes that were discovered in this report. Multiple contigs were found to represent the same gene for known pathway enzymes. Contigs labeled *CYP* are distinct. Heat map shows FPKM (fragments per kilobase of transcript per million mapped reads) levels of transcripts.

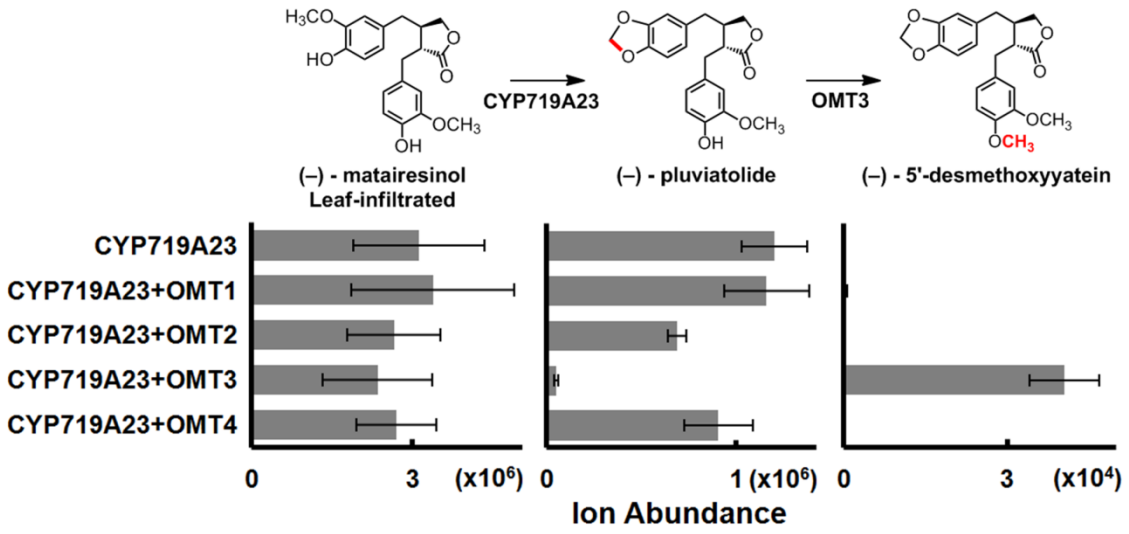
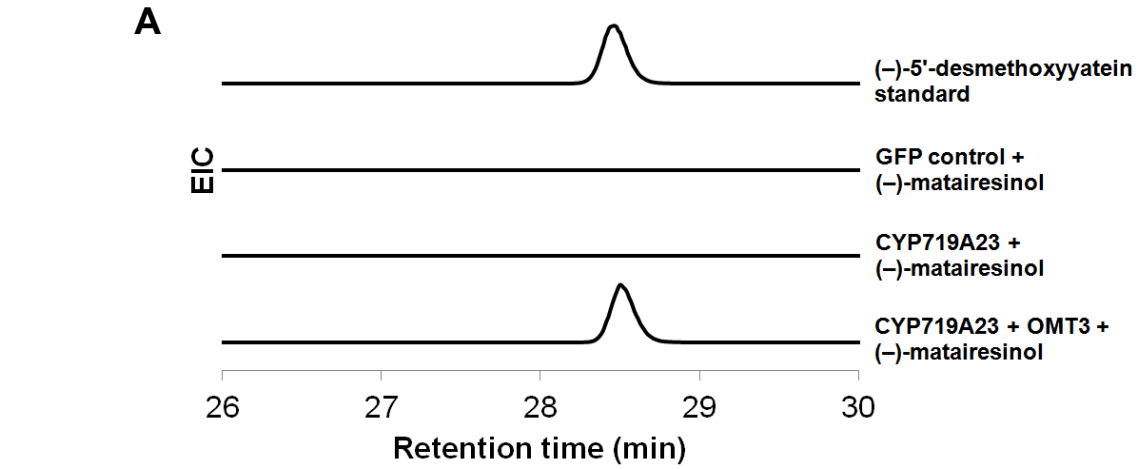
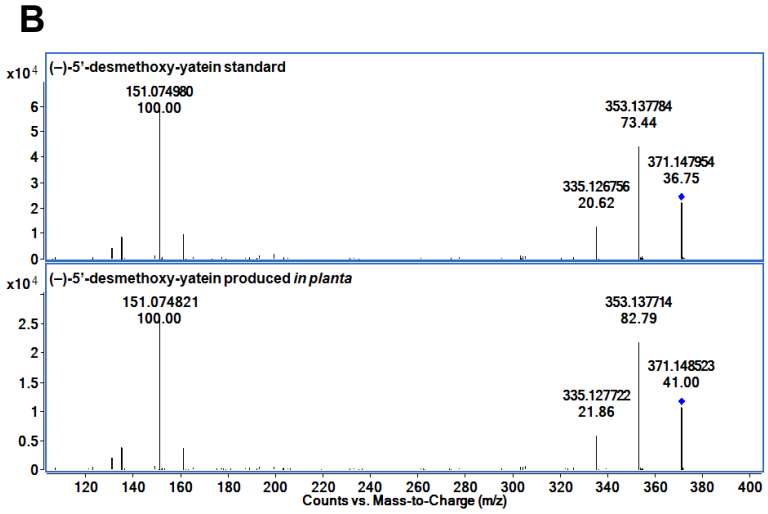


Fig. S4. Screening of OMT enzyme candidates in *N. benthamiana* by co-expression with CYP719A23 and (–)-matairesinol infiltration.

Tobacco leaves expressing CYP719A23 individually with OMT candidates were infiltrated with (–)-matairesinol. Average LC-MS [(–)-mode] ion abundance ± one S.D. from three biological replicates of (–)-matairesinol and reaction products, (–)-pluviatolide and (–)-5'-desmethoxy-yatein are shown.





C

m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
371.1480	$C_{21}H_{23}O_6^+$	A, $[M+H]^+$	371.1489	2.55
353.1378	$C_{21}H_{21}O_5^+$	B, $[M+H-H_2O]^+$	353.1384	1.74
335.1268	$C_{21}H_{19}O_4^+$	$[M+H-2H_2O]^+$	335.1283	4.61
161.0590	$C_{10}H_9O_2^+$	C	161.0597	4.60
151.0750	$C_9H_{11}O_2^+$	D	151.0754	2.78
135.0433	$C_8H_7O_2^+$	E	135.0441	6.15
131.0492	$C_9H_7O^+$		131.0497	3.85

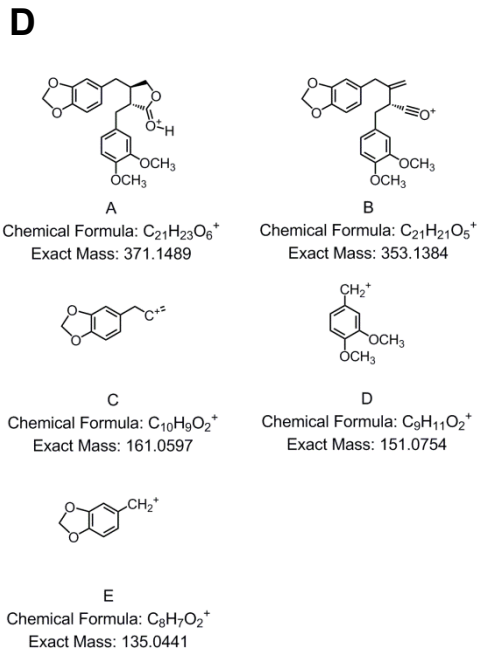


Fig. S5. ***N. benthamiana*** **transient expression of CYP719A23 and OMT3 and (–)-matairesinol infiltration produces (–)-5'-desmethoxy-yatein.**

(A) Aligned EICs of (–)-5'-desmethoxy-yatein [m/z = 371, (+)-mode] from the transient expression of CYP719A23 and OMT3, compared to CYP719A23 alone, GFP alone and an authentic standard. Only chromatograms from *in planta* experiments are to scale. **(B)** MS/MS (10 V) spectra of (–)-5'-desmethoxy-yatein produced *in planta* compared to authentic standard. **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.

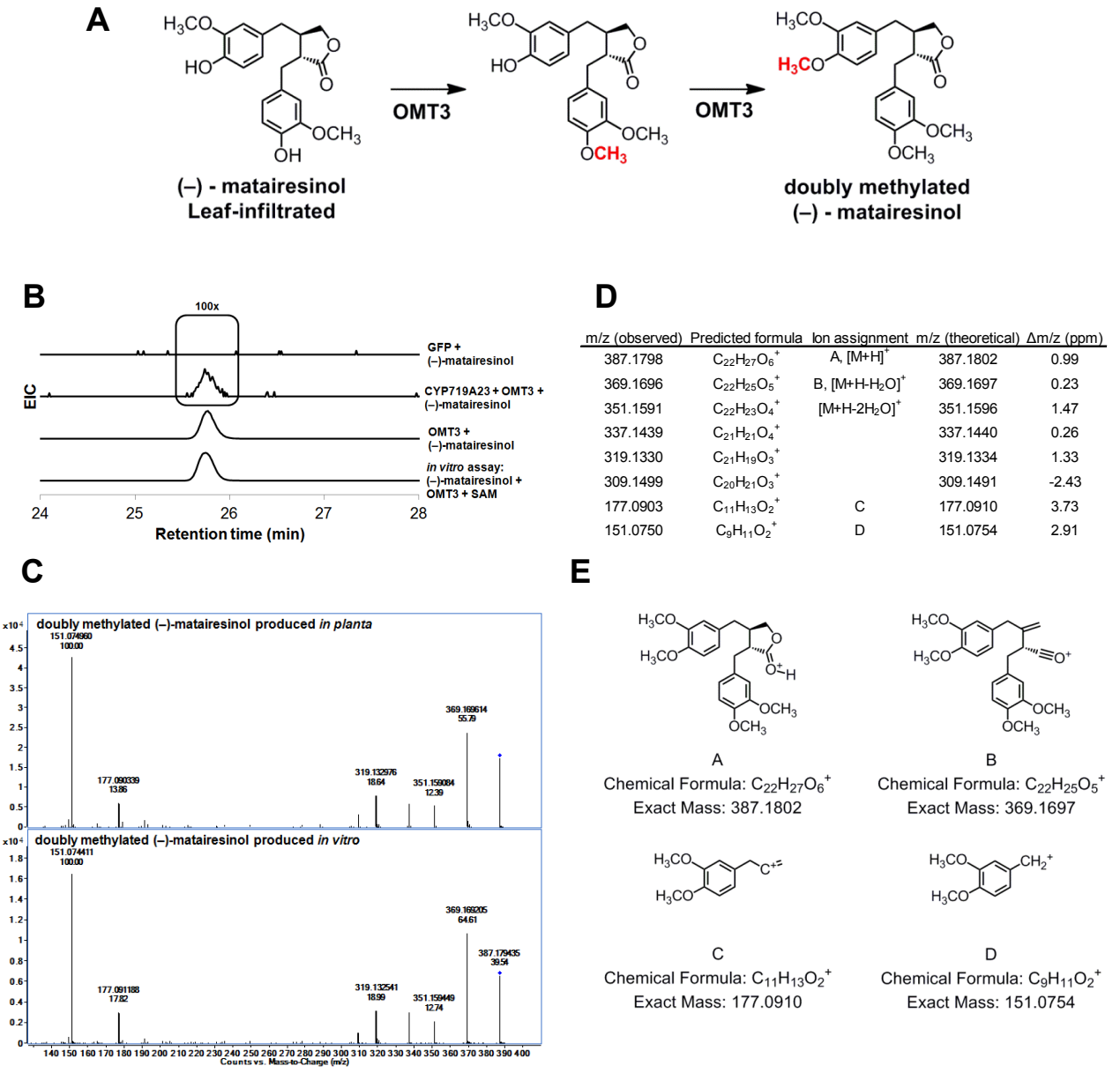


Fig. S6. *N. benthamiana* transient expression of CYP719A23 and OMT3 or OMT3 alone and (–)-matairesinol infiltration produces doubly methylated (–)-matairesinol.

(A) Proposed reactions occurring *in planta* from the expression of OMT3 **(B)** Aligned EICs of the product [m/z = 387, (+)-mode] from the transient expression of CYP719A23 and OMT3, compared to expression of OMT3 alone and GFP alone, and *in vitro* assay with purified OMT3 from *E. coli*, SAM and (–)-matairesinol. Only chromatograms from *in planta* expression of CYP719A23 + OMT3 and GFP alone are to scale (approximately 100x magnified with respect to chromatogram from *in planta* expression of OMT3). **(C)** MS/MS (10 V) spectra of the product produced *in planta* compared to *in vitro* product formed by incubation of purified OMT3 from *E. coli*, SAM and (–)-matairesinol. **(D)** Table of MS/MS peak assignments and **(E)** putative ion structures.

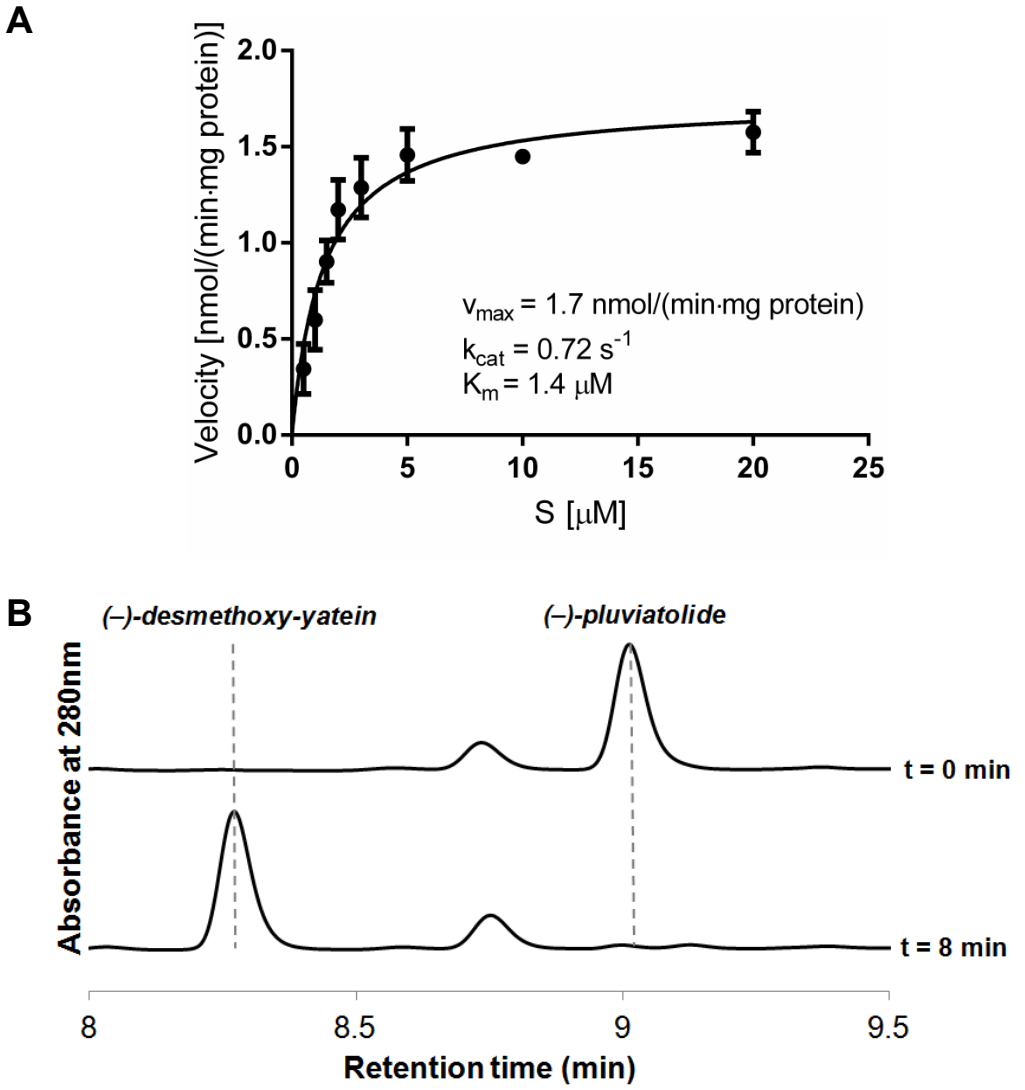


Fig. S7. In vitro characterization of (–)-pluviatolide O-methyltransferase (OMT3).

(A) Michaelis-Menten kinetic parameters were derived from initial rates of (–)-pluviatolide consumption by HPLC-DAD (280 nm) analysis with heterologously expressed and purified OMT3 from *E. coli*. Error bars represent the standard deviation from three independent assays. **(B)** HPLC-DAD (280 nm) profile of an *in vitro* reaction.

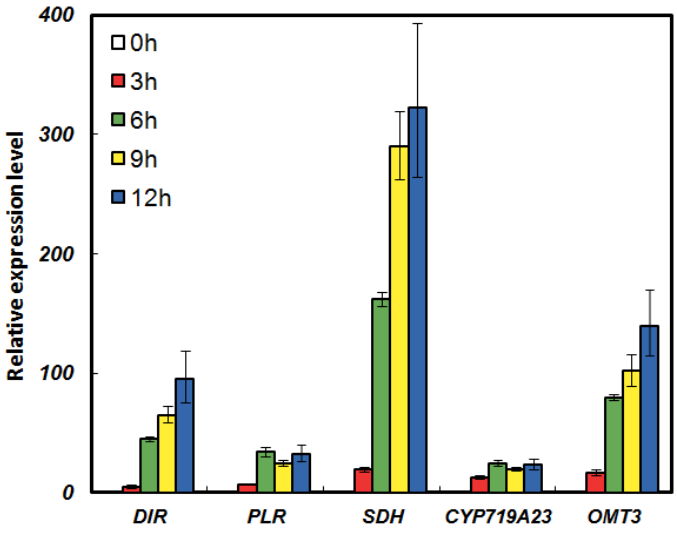


Fig. S8. Gene expression profiling of a *P. hexandrum* leaf after wounding.

qRT-PCR gene expression analysis of *OMT3* and other known podophyllotoxin biosynthetic genes after *P. hexandrum* leaf wounding (biological replicate distinct from sample used in Figure 3A)**.** Relative expression levels were determined with respect to 0 h time point (before wounding). Data show average values (n = 3 technical replicates) ± one S.D.

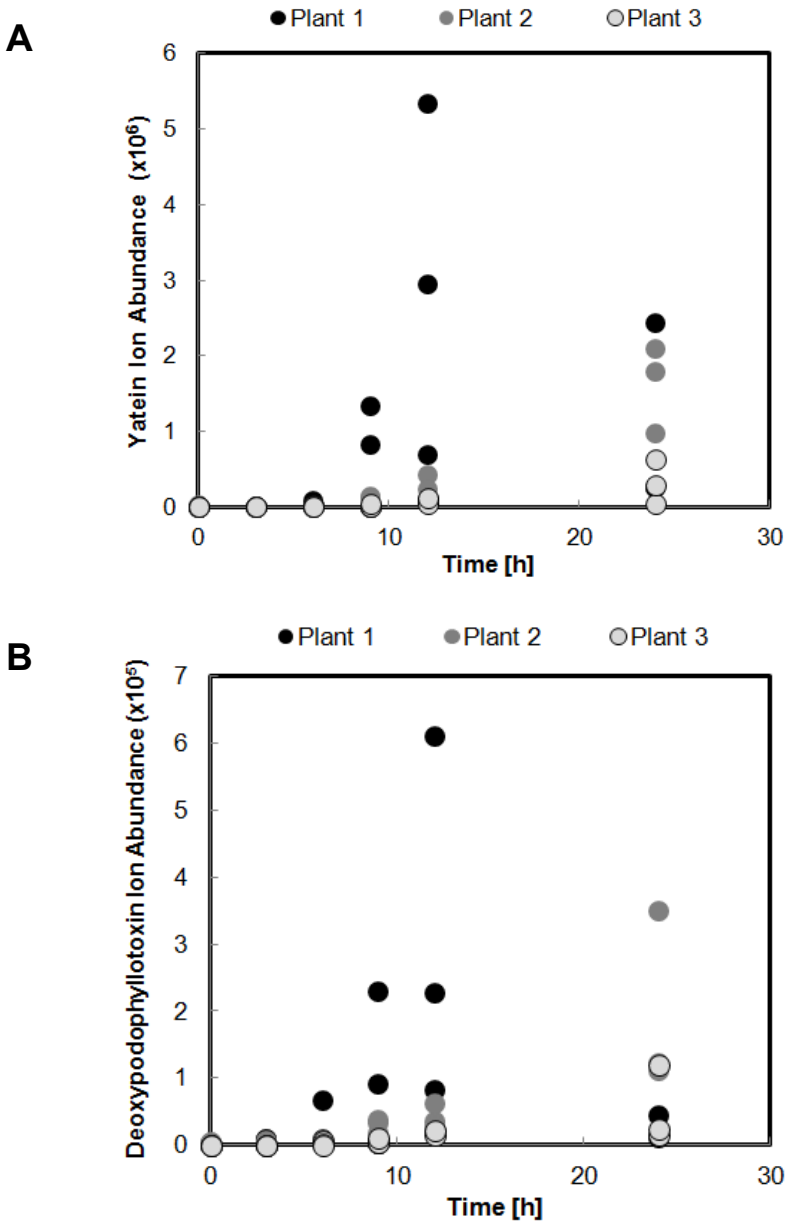


Fig. S9. Metabolite profiling of *P. hexandrum* leaves after wounding.

P. hexandrum leaf wounding induces changes in lignan levels over time**.** Ion abundances [(+)-mode] are shown for triplicate samples from three plants for **(A)** (–)-yatein (m/z = 401) and **(B)** (–)-deoxypodophyllotoxin (m/z = 399). Plant 1, which showed the strongest metabolite response, was used for qRT-PCR analysis (results in Figure 3A) and RNA Sequencing.

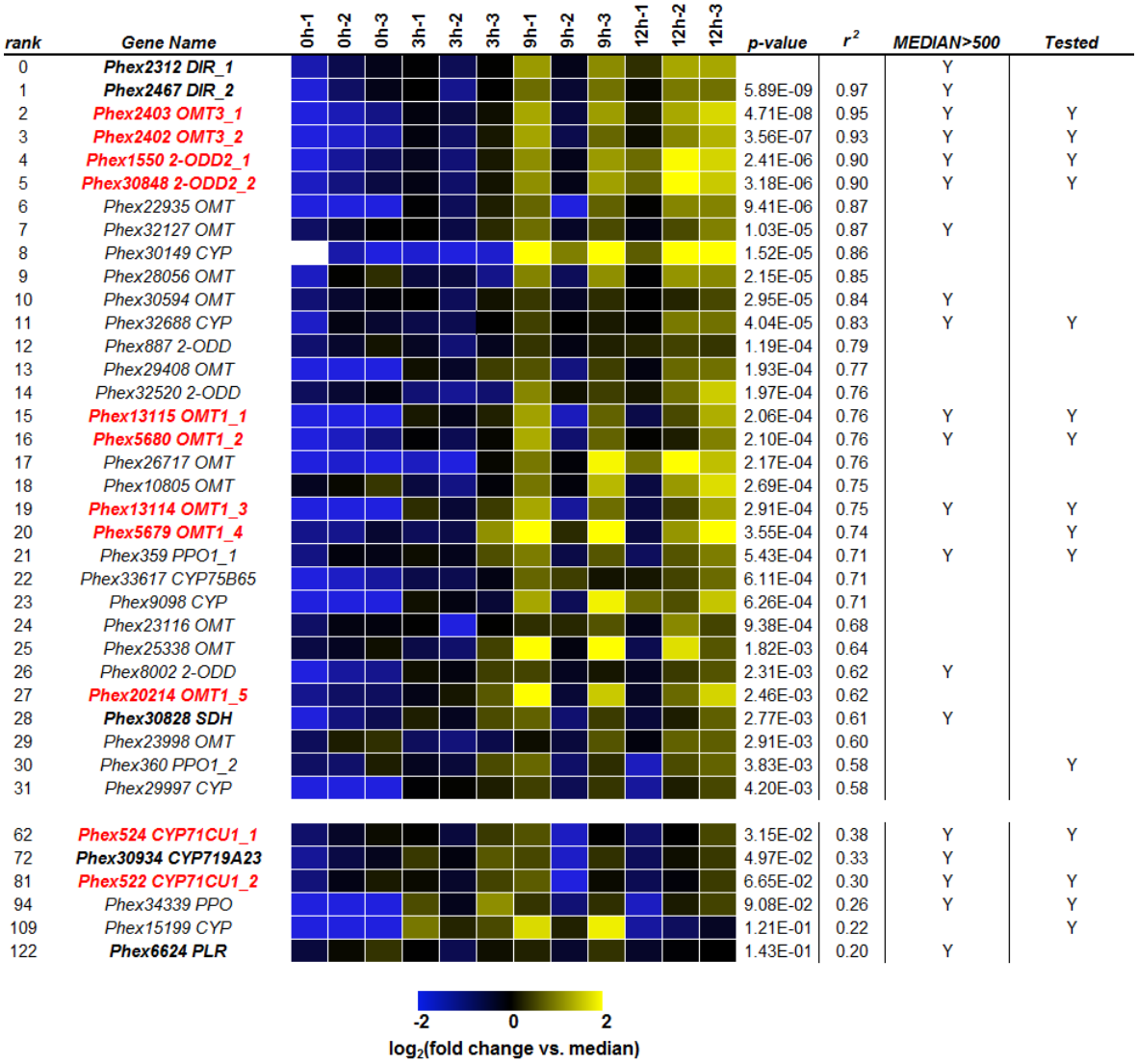
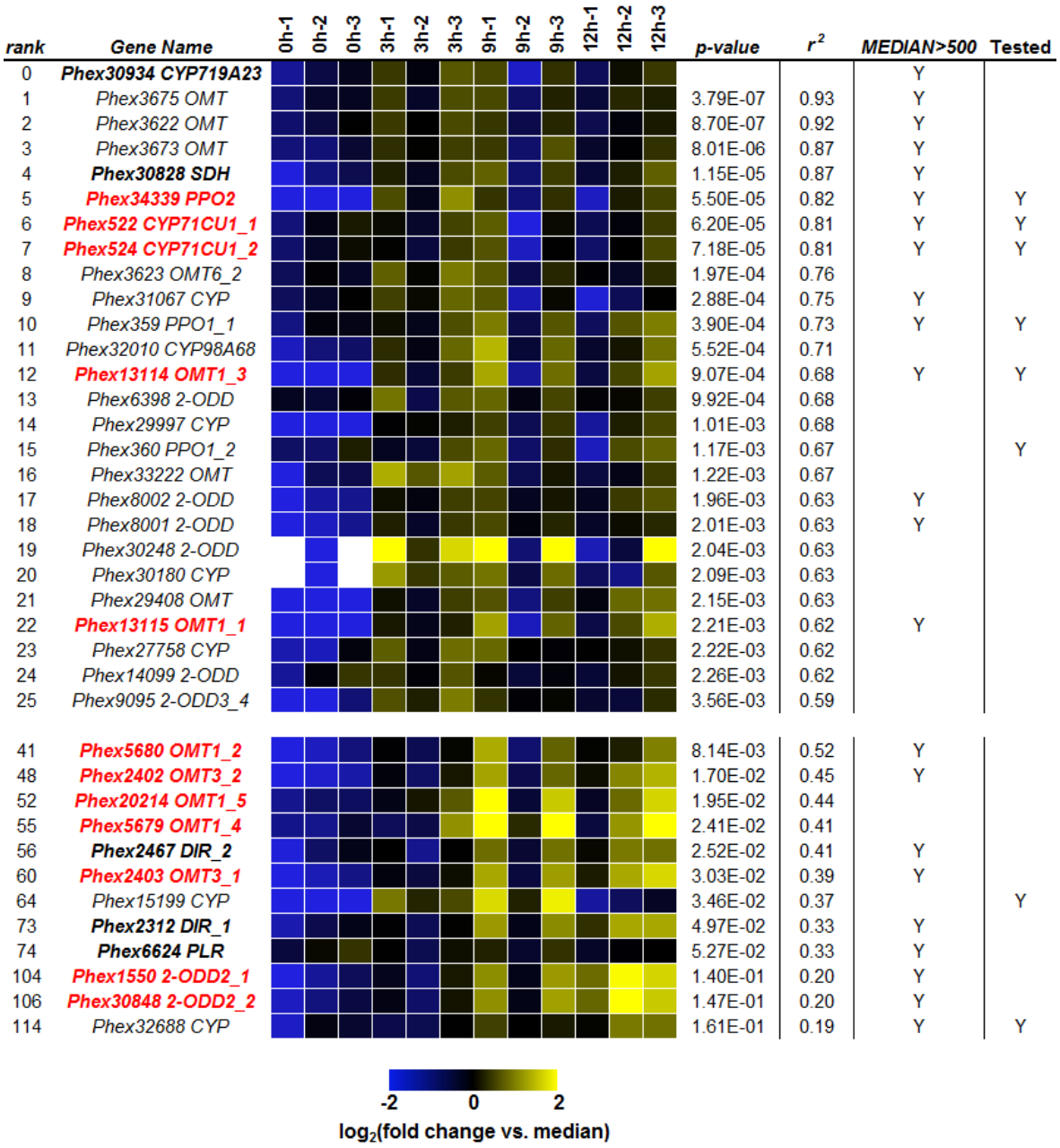


Fig. S10. Co-expression analysis of *P. hexandrum* leaf RNA-Seq expression data after wounding using *DIR* as a bait gene.

The top 31 ranked genes based on p-value from linear regression analysis were considered significant using the Benjamini–Hochberg procedure with a false discovery rate of 0.05 (multiple testing corrections). Other (–)-podophyllotoxin biosynthetic genes previously known (bold black) and discovered in this report (red), and other screened candidates based on co-expression analysis using *CYP719A23* as the bait gene (non-bold black) are also displayed. Heat map depicts the expression levels (effective counts, TMM normalized, log_2-scaled and median-centered) after *P. hexandrum* leaf wounding. Candidate genes that were screened from this data were chosen based on co-expression with *DIR* and *CYP719A23* (fig. S11), and level of expression (TMM-normalized effective counts > 500). Note that there are multiple contigs that represent the same gene.



rank	Gene Name	p-value	r^2	MEDIAN>500	Tested
0	**Phex30934 CYP719A23**			Y	
1	*Phex3675 OMT*	3.79E-07	0.93	Y	
2	*Phex3622 OMT*	8.70E-07	0.92	Y	
3	*Phex3673 OMT*	8.01E-06	0.87	Y	
4	**Phex30828 SDH**	1.15E-05	0.87	Y	
5	**Phex34339 PPO2**	5.50E-05	0.82	Y	Y
6	**Phex522 CYP71CU1_1**	6.20E-05	0.81	Y	Y
7	**Phex524 CYP71CU1_2**	7.18E-05	0.81	Y	Y
8	*Phex3623 OMT6_2*	1.97E-04	0.76		
9	*Phex31067 CYP*	2.88E-04	0.75	Y	
10	*Phex359 PPO1_1*	3.90E-04	0.73	Y	Y
11	*Phex32010 CYP98A68*	5.52E-04	0.71		
12	**Phex13114 OMT1_3**	9.07E-04	0.68	Y	Y
13	*Phex6398 2-ODD*	9.92E-04	0.68		
14	*Phex29997 CYP*	1.01E-03	0.68		
15	*Phex360 PPO1_2*	1.17E-03	0.67		Y
16	*Phex33222 OMT*	1.22E-03	0.67		
17	*Phex8002 2-ODD*	1.96E-03	0.63	Y	
18	*Phex8001 2-ODD*	2.01E-03	0.63	Y	
19	*Phex30248 2-ODD*	2.04E-03	0.63		
20	*Phex30180 CYP*	2.09E-03	0.63		
21	*Phex29408 OMT*	2.15E-03	0.63		
22	**Phex13115 OMT1_1**	2.21E-03	0.62	Y	
23	*Phex27758 CYP*	2.22E-03	0.62		
24	*Phex14099 2-ODD*	2.26E-03	0.62		
25	*Phex9095 2-ODD3_4*	3.56E-03	0.59		
41	**Phex5680 OMT1_2**	8.14E-03	0.52	Y	
48	**Phex2402 OMT3_2**	1.70E-02	0.45	Y	
52	**Phex20214 OMT1_5**	1.95E-02	0.44		
55	**Phex5679 OMT1_4**	2.41E-02	0.41		
56	**Phex2467 DIR_2**	2.52E-02	0.41	Y	
60	**Phex2403 OMT3_1**	3.03E-02	0.39	Y	
64	*Phex15199 CYP*	3.46E-02	0.37		Y
73	**Phex2312 DIR_1**	4.97E-02	0.33	Y	
74	**Phex6624 PLR**	5.27E-02	0.33	Y	
104	**Phex1550 2-ODD2_1**	1.40E-01	0.20	Y	
106	**Phex30848 2-ODD2_2**	1.47E-01	0.20	Y	
114	*Phex32688 CYP*	1.61E-01	0.19	Y	Y

Fig. S11. Co-expression analysis of *P. hexandrum* leaf RNA-Seq expression data after wounding using *CYP719A23* as a bait gene.

The top 25 ranked genes based on p-value from linear regression analysis were considered significant using the Benjamini–Hochberg procedure with a false discovery rate of 0.05 (multiple testing corrections). Other (–)-podophyllotoxin biosynthetic genes previously known (bold black) and discovered in this report (red), and other screened candidates based on co-expression analysis using *DIR* as the bait gene (non-bold black) are also displayed. Heat map depicts the expression levels (effective counts, TMM normalized, $\log_2$-scaled and median-centered) after *P. hexandrum* leaf wounding. Candidate genes that were screened from this data were chosen based on co-expression with *DIR* (fig. S10) and *CYP719A23*, and level of expression (TMM-normalized effective counts > 500). Note that there are multiple contigs that represent the same gene.

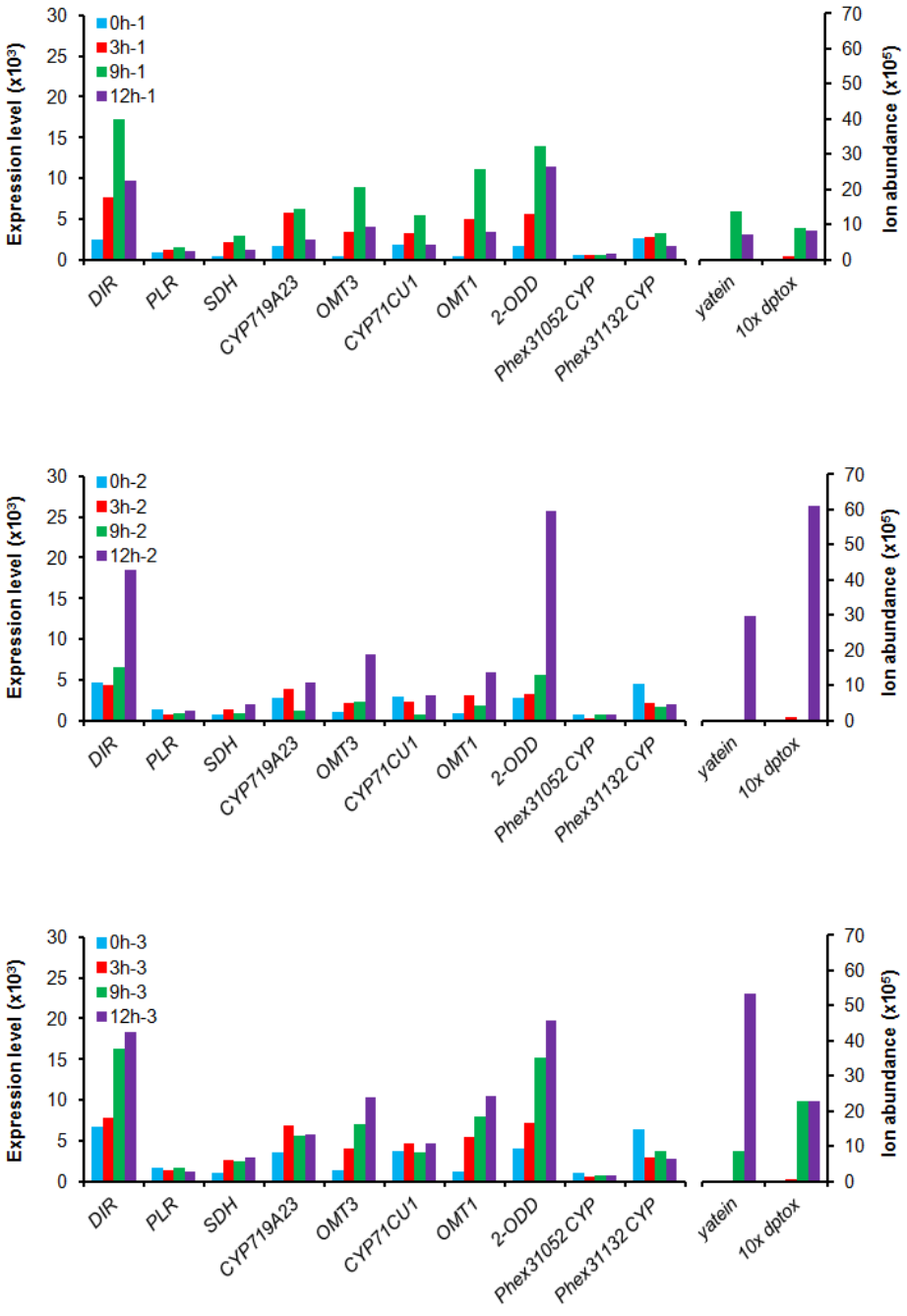
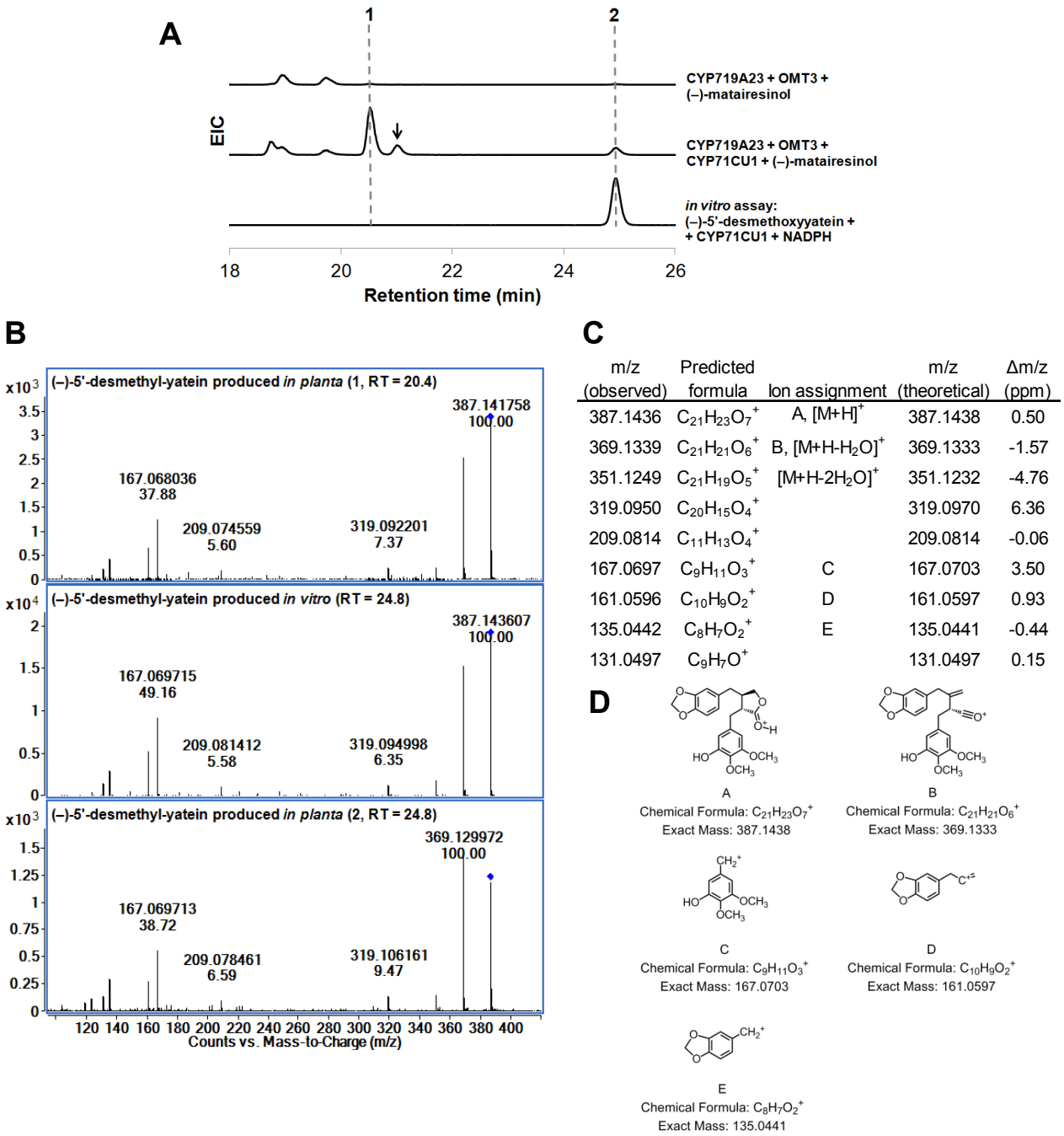


Fig. S12. RNA-Seq expression profiles and metabolite data after leaf wounding of all known (–)-podophyllotoxin genes and genes discovered in this report.

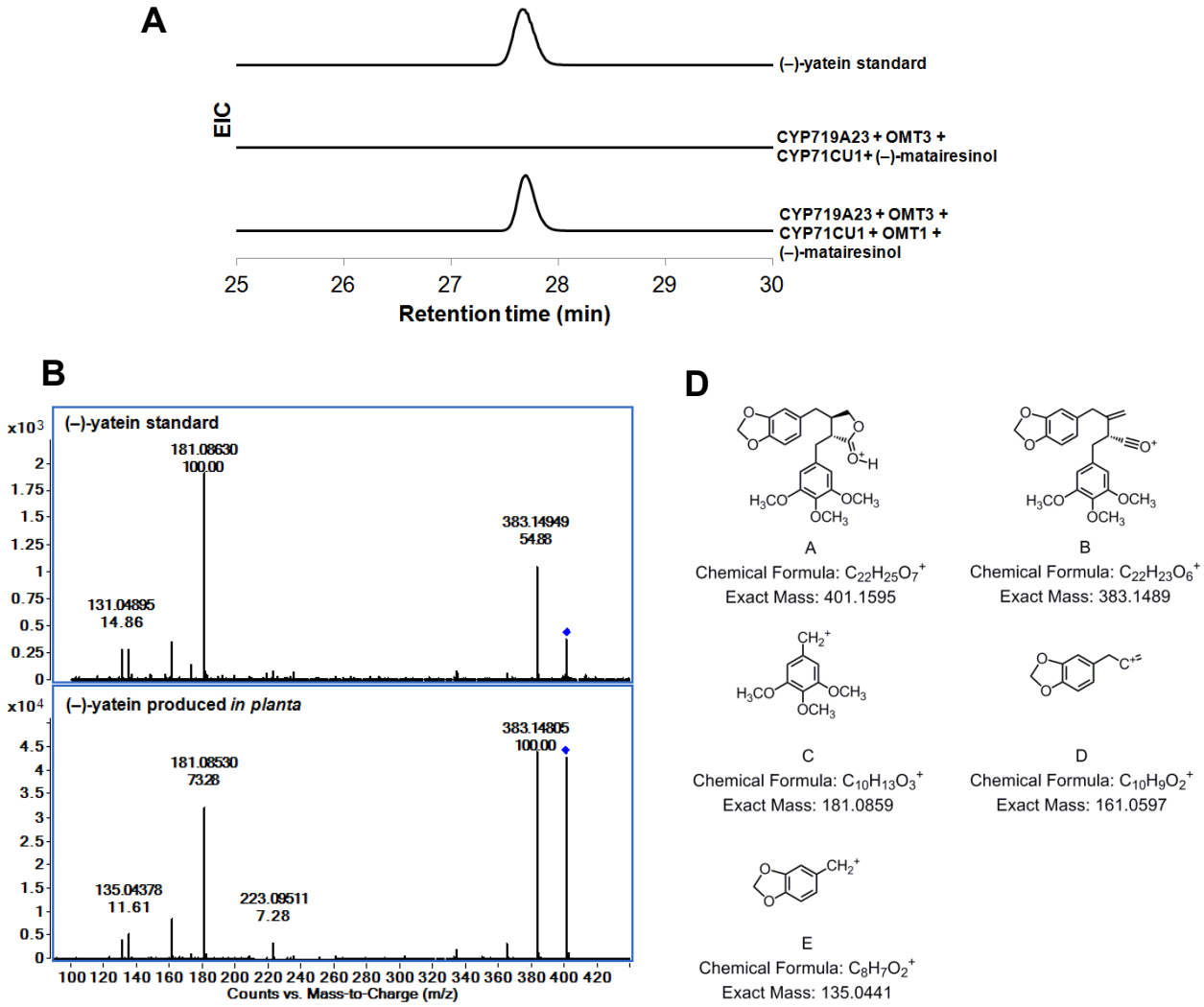
Each graph shows expression and metabolite data (dptox, deoxypodophyllotoxin) for individual leaflets from the same leaf after wounding. The expression and metabolite profiles are distinct for each set of data. Note that *PLR* was not significantly induced after leaf wounding. Expression profiles of *Phex31052 (CYP)* and *Phex31132 (CYP)* are also shown to demonstrate that not all genes are induced by wounding. Closest Arabidopsis homologs to *Phex31052* and *Phex31132* are *CYP704A2* and *CYP74B2* (hydroperoxide lyase), respectively.



m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
387.1436	$C_{21}H_{23}O_7^+$	A, $[M+H]^+$	387.1438	0.50
369.1339	$C_{21}H_{21}O_6^+$	B, $[M+H-H_2O]^+$	369.1333	-1.57
351.1249	$C_{21}H_{19}O_5^+$	$[M+H-2H_2O]^+$	351.1232	-4.76
319.0950	$C_{20}H_{15}O_4^+$		319.0970	6.36
209.0814	$C_{11}H_{13}O_4^+$		209.0814	-0.06
167.0697	$C_9H_{11}O_3^+$	C	167.0703	3.50
161.0596	$C_{10}H_9O_2^+$	D	161.0597	0.93
135.0442	$C_8H_7O_2^+$	E	135.0441	-0.44
131.0497	$C_9H_7O^+$		131.0497	0.15

Fig. S13. *N. benthamiana* transient expression of CYP719A23, OMT3, and CYP71CU1 and (–)-matairesinol leaf infiltration produces (–)-5'-desmethyl-yatein.

(A) Aligned EICs of (–)-desmethyl-yatein [m/z = 387, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1, compared to CYP719A23 + OMT3, and an *in vitro* assay with CYP71CU1–enriched microsomes, NADPH and (–)-5'-desmethoxy-yatein; peaks observed at earlier labeled retention times (1 and arrow) are derived from in-source fragmentation ions of (–)-5'-desmethyl-yatein glucosylated derivatives produced by endogenous tobacco enzymes. Only chromatograms from *in planta* experiments are to scale. **(B)** MS/MS (10 V) spectra of (–)-desmethyl-yatein produced *in planta* (from labeled retention times, 1 and 2) compared to *in vitro* assay. **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.

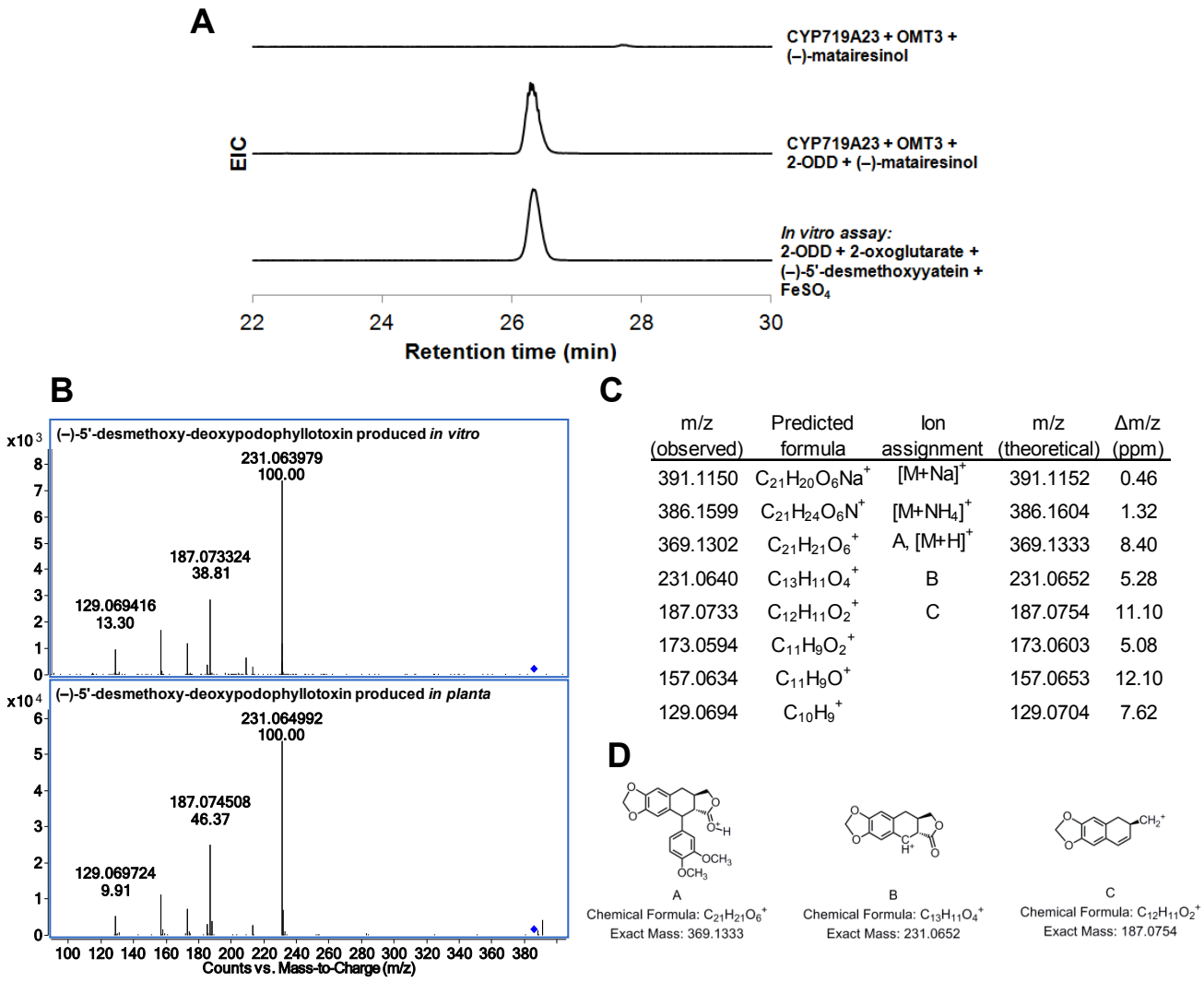


C

m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
401.1586	$C_{22}H_{25}O_7^+$	A, $[M+H]^+$	401.1595	2.19
383.1481	$C_{22}H_{23}O_6^+$	B, $[M+H-H_2O]^+$	383.1489	2.22
181.0853	$C_{10}H_{13}O_3^+$	C	181.0859	3.31
161.0586	$C_{10}H_9O_2^+$	D	161.0597	6.64
135.0438	$C_8H_7O_2^+$	E	135.0441	2.37
131.0485	$C_9H_7O^+$		131.0497	9.00

Fig. S14. *N. benthamiana* transient expression of CYP719A23, OMT3, CYP71CU1, and OMT1 and (–)-matairesinol leaf infiltration produces (–)-yatein.

(A) Aligned EICs of (–)-yatein [m/z = 401, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1 + OMT1, compared to CYP719A23 + OMT3 + CYP71CU1 and an authentic standard. Only chromatograms from *in planta* experiments are to scale. **(B)** MS/MS (10 V) spectra of (–)-yatein produced *in planta* compared to authentic standard. **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.



m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
391.1150	$C_{21}H_{20}O_6Na^+$	$[M+Na]^+$	391.1152	0.46
386.1599	$C_{21}H_{24}O_6N^+$	$[M+NH_4]^+$	386.1604	1.32
369.1302	$C_{21}H_{21}O_6^+$	A, $[M+H]^+$	369.1333	8.40
231.0640	$C_{13}H_{11}O_4^+$	B	231.0652	5.28
187.0733	$C_{12}H_{11}O_2^+$	C	187.0754	11.10
173.0594	$C_{11}H_9O_2^+$		173.0603	5.08
157.0634	$C_{11}H_9O^+$		157.0653	12.10
129.0694	$C_{10}H_9^+$		129.0704	7.62

Fig. S15. *N. benthamiana* transient expression of CYP719A23, OMT3 and 2-ODD and (–)-matairesinol leaf infiltration produces (–)-5'-desmethoxy-deoxypodophyllotoxin.

(A) Aligned EICs of 5'-desmethoxy-deoxypodophyllotoxin (m/z = 391, $[M+Na]^+$) from the transient expression of CYP719A23 + OMT3 + 2-ODD, compared to CYP719A23 + OMT3 and an *in vitro* assay with 2-ODD purified from *E. coli*, Fe_2SO_4, 2-oxoglutarate and (–)-5'-desmethoxy-yatein. Only chromatograms from *in planta* experiments are to scale. **(B)** MS/MS (20 V) spectra of (–)-5'-desmethoxy-deoxypodophyllotoxin (m/z = 386, $[M+NH_4]^+$) produced *in planta* compared to *in vitro* product **(C)** Table of MS and MS/MS peak assignments and **(D)** putative ion structures.

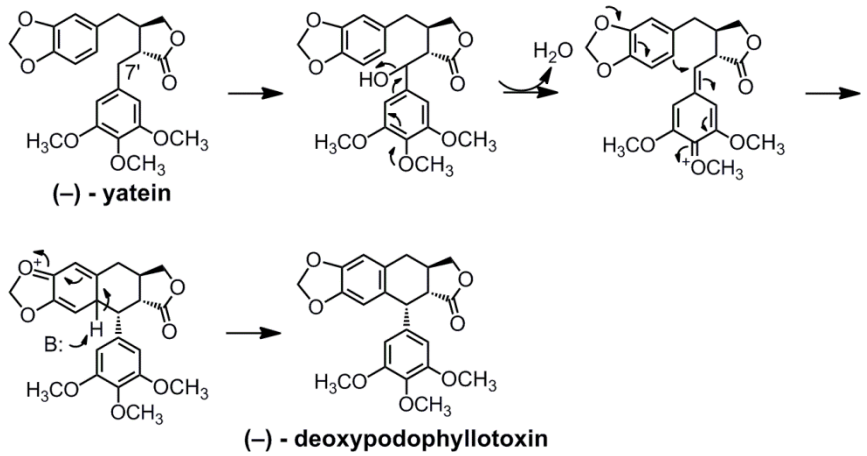


Fig. S16. Proposed reaction mechanism of the conversion of (–)-yatein to (–)-deoxypodophyllotoxin catalyzed by 2-ODD.

Oxygen is incorporated via hydroxylation of (–)-yatein at the 7' position, followed by its removal as water to form a quinone methide intermediate. Stereoselective cyclization then occurs, followed by removal of a proton to restore aromaticity, forming (–)-deoxypodophyllotoxin.

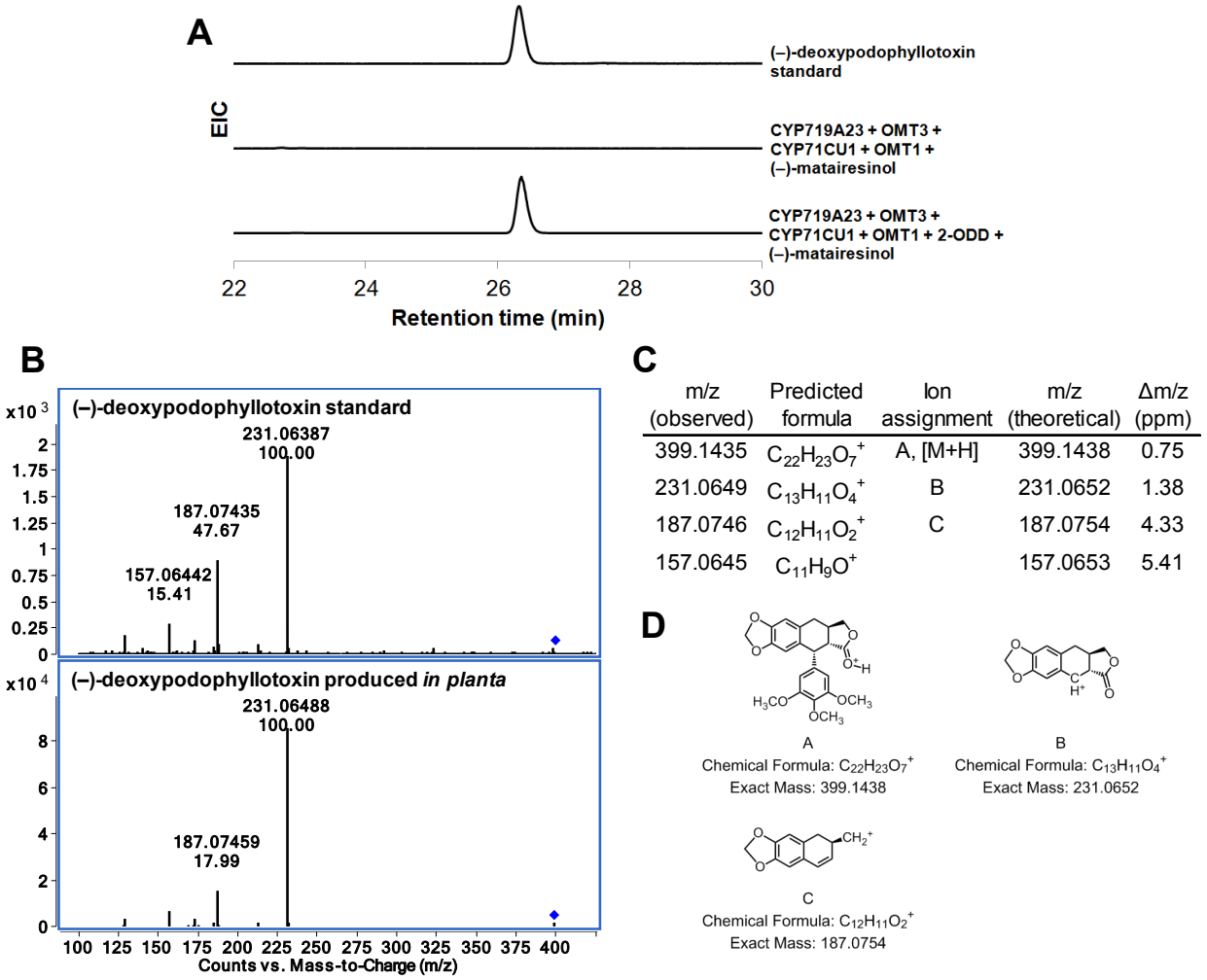


Fig. S17. *N. benthamiana* transient expression of CYP719A23, OMT3, CYP71CU1, OMT1, and 2-ODD and (–)-matairesinol leaf infiltration produces (–)-deoxypodophyllotoxin.

A) Aligned EICs of (–)-deoxypodophyllotoxin [m/z = 399, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD, compared to CYP719A23 + OMT3 + CYP71CU1 + OMT1 and an authentic standard. Only chromatograms from *in planta* experiments are to scale. **(B)** MS/MS (10 V) spectra of (–)-deoxypodophyllotoxin produced *in planta* compared to authentic standard **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.

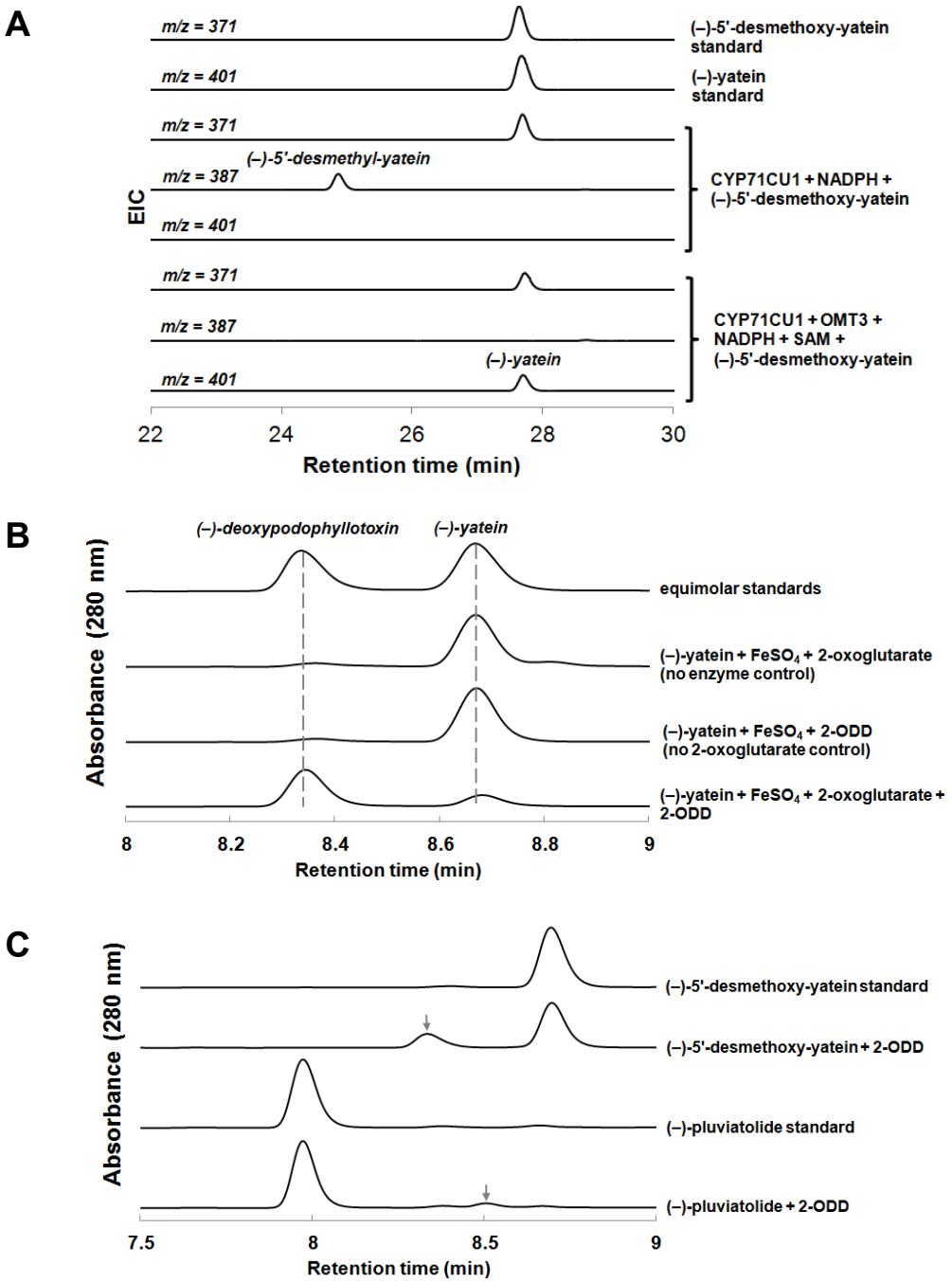


Fig. S18. Enzymatic assays of heterologously expressed CYP71CU1, OMT1 and 2-ODD.

(A) Microsomal fractions containing CYP71CU1 were incubated with (–)-5'-desmethoxy-yatein [m/z = 371, (+)-mode] and NADPH, forming (–)-5'-desmethyl-yatein [m/z = 387]. CYP71CU1-microsomes and OMT1, expressed and purified from *E. coli*, were incubated with (–)-5'-desmethoxy-yatein, NADPH and SAM, forming (–)-yatein [m/z = 401]. (–)-5'-desmethoxy-yatein co-elutes with (–)-yatein. **(B)** 2-ODD, expressed and purified from *E. coli*, was incubated with (–)-yatein, 2-oxoglutarate and $FeSO_4$ to form (–)-deoxypodophyllotoxin. **(C)** The enzyme also had minor activity on (–)-5'-desmethoxy-yatein and almost negligible activity on (–)-pluviatolide under the same assay conditions. All chromatograms are to scale. Arrows point to products.

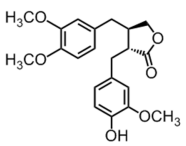
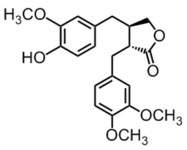
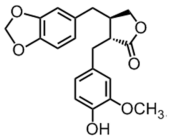
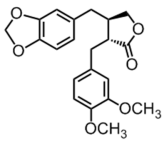
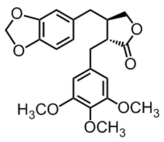
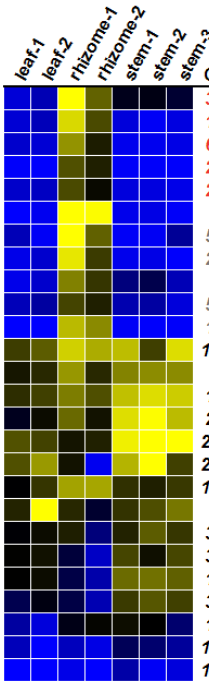
Enzyme	 (–)-arctigenin	 (–)-matairesinol	 (–)-pluviatolide	 (–)-5'-desmethoxy-yatein	 (–)-yatein
OMT3	<5%	<5%	100%	Not tested	Not tested
CYP71CU1	0%	0%	0%	100%	Not tested
OMT1	0%	0%	0%	Not tested	Not tested
2-ODD	0%	0%	<1%	<20%	100%

Fig. S19. Substrate specificity of OMT3, CYP71CU1, OMT1 and 2-ODD as determined by *in vitro* characterization.

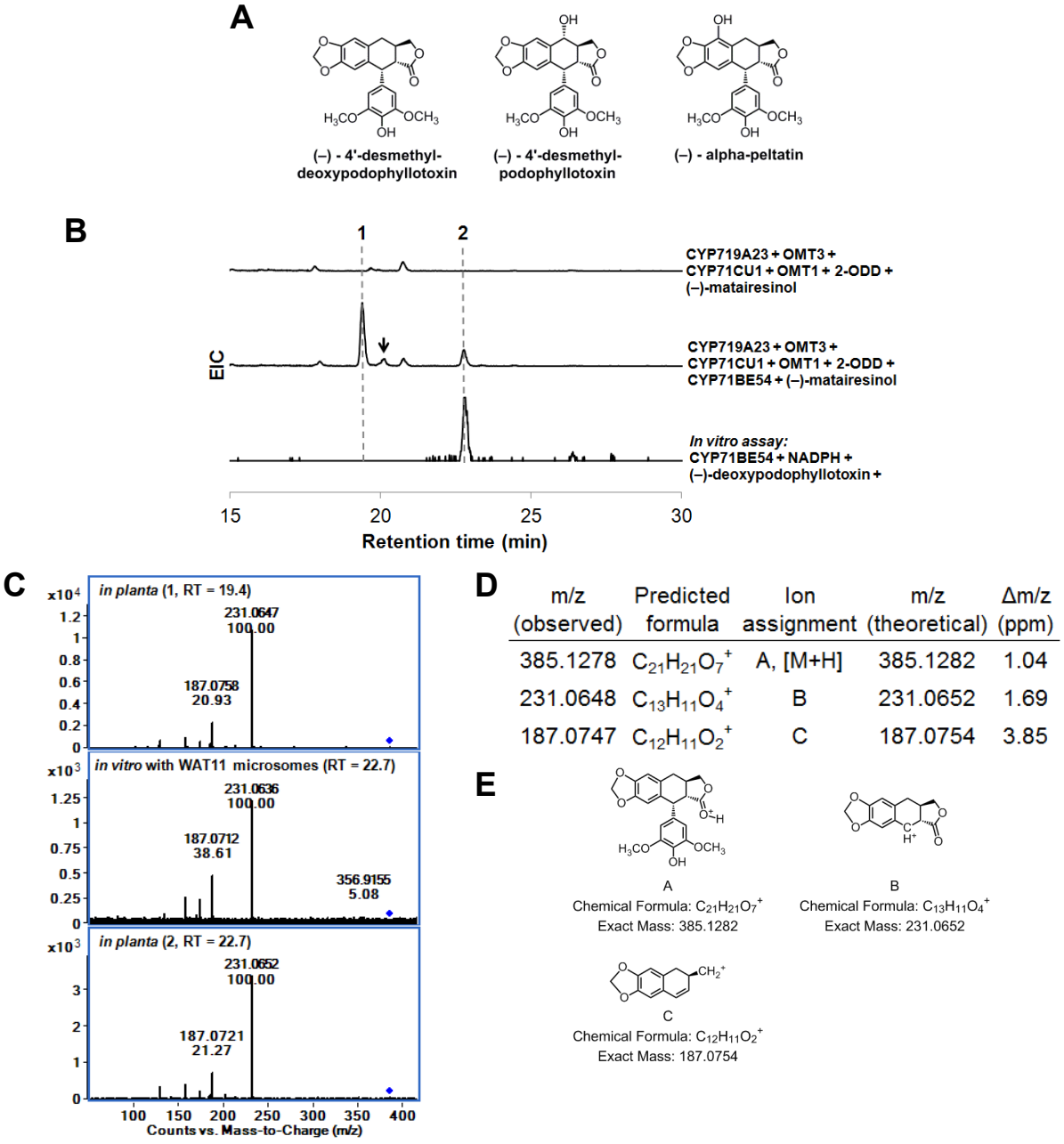
The table reports % activity of OMT3, CYP71CU1, OMT1 and 2-ODD as determined *in vitro* on a number of available lignan substrates. The % activity is determined with respect to the maximum conversion detected in all substrates tested. The results suggest that the order of events in the biosynthesis of (–)-4'-desmethyl-epipodophyllotoxin proceeds as depicted in Figure 4A. The predicted substrate for OMT1 is (–)-5'-desmethyl-yatein and was tested by incubation of OMT1 and SAM with (–)-5'-desmethoxy-yatein, CYP71CU1-microsomes isolated from *WAT11* and NADPH to generate the substrate *in vitro*.



Contig	Gene	Best blastx hit as of March 2015	% Identity
35407	CYP82D61	CYP82D61 [Sinopodophyllum hexandrum]	99
14372	CYP71BE54_1	PREDICTED: premnaspirodiene oxygenase-like isoform X1 [Solanum tuberosum]	59
62371	CYP71BE54_2		
22113	CYP71BE54_3		
20475	CYP71BE54_4		
5041	CYP	PREDICTED: cytochrome P450 CYP736A12-like [Pyrus x bretschneideri]	52
57270	CYP	CYP71BE30 [Sinopodophyllum hexandrum]	100
29054	CYP	PREDICTED: geraniol 8-hydroxylase-like [Nelumbo nucifera]	68
4618	CYP	PREDICTED: cytochrome P450 CYP736A12-like [Malus domestica]	52
55921	NADP	PREDICTED: cinnamoyl-CoA reductase 2-like [Elaeis guineensis]	58
18808	NADP	RecName: Full=Non-functional NADPH-dependent codeinone reductase 2 [Papaver somniferum]	63
102408	CYP719A23		
6901	PLR		
17736	CYP719A23		
26599	SDH		
225796	CYP719A23		
212985	CYP719A23		
140193	PLR		
8051	DIR		
31365	CYP719A23		
38318	CYP719A23		
12248	SDH		
34018	CYP719A23		
12826	SDH		
145347	CYP719A23		
193171	CYP719A23		

Fig. S20. Rhizome-specific candidate genes selected for screening from the Medicinal Plants Consortium *P. hexandrum* transcriptome.

Candidate genes (grey and red) for enzymes involved in biosynthetic steps downstream of (–)-deoxypodophyllotoxin production were chosen based on rhizome-specific expression and the availability of full coding sequences. CYPs were chosen for hydroxylating activity on (–)-deoxypodophyllotoxin and NADP(H)-dependent oxidoreductases (NADP) were chosen for oxidative and reductive activity to epimerize (–)-epipodophyllotoxin to (–)-podophyllotoxin. In red are lignan biosynthetic genes that were discovered and characterized in this report. Heat map shows FPKM (fragments per kilobase of transcript per million mapped reads) levels of transcripts.



m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
385.1278	$C_{21}H_{21}O_7^+$	A, [M+H]	385.1282	1.04
231.0648	$C_{13}H_{11}O_4^+$	B	231.0652	1.69
187.0747	$C_{12}H_{11}O_2^+$	C	187.0754	3.85

Fig. S21. N. benthamiana transient expression of CYP719A23, OMT3, CYP71CU1, OMT1, 2-ODD and CYP71BE54, and (–)-matairesinol infiltration produces (–)-4'-desmethyl-deoxypodophyllotoxin.

(A) Known 4'-desmethyl lignans present in *P. hexandrum*. **(B)** Aligned EICs of the product [m/z = 385, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + CYP71BE54, compared to CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD, and *in vitro* assay with CYP71BE54-enriched microsomes, NADPH and (–)-deoxypodophyllotoxin; peaks observed at earlier labeled retention times (1 and arrow) are in-source fragmentation ions from a (–)-4'-desmethyl-deoxypodophyllotoxin glucosylated derivative produced by endogenous tobacco enzymes. Only chromatograms from *in planta* expression are to scale. **(C)** MS/MS spectra (10 V) of **(–)**-4'-desmethyl-deoxypodophyllotoxin produced *in planta* (from labeled retention times, 1 and 2); ion abundance of *in vitro* product formed by incubation of (–)-deoxypodophyllotoxin with CYP71BE54-enriched microsomes was too low to obtain a MS/MS spectrum. **(D)** Table of MS/MS peak assignments and **(E)** and putative ion structures.

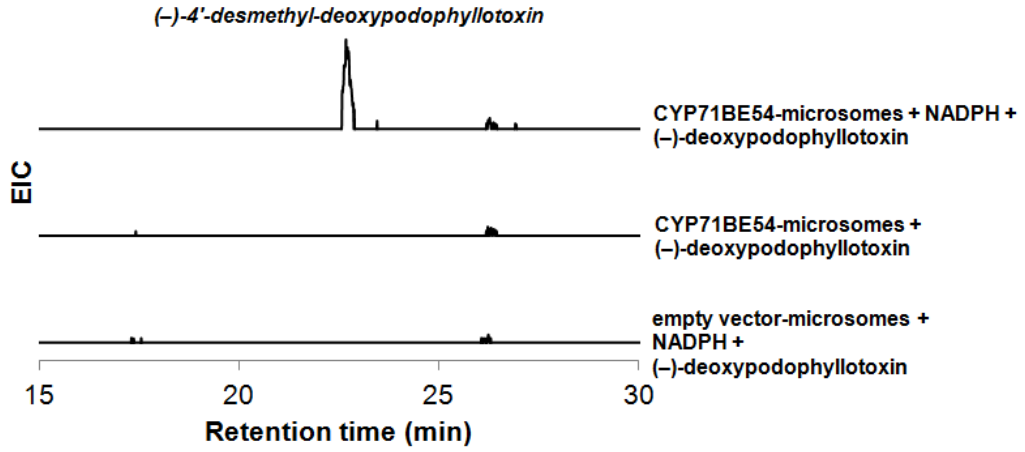


Fig. S22. Enzymatic assays of CYP71BE54-enriched microsomal fractions isolated from *WAT11*.

Microsomal fractions containing CYP71BE54 were incubated with (–)-deoxypodophyllotoxin and NADPH, forming (–)-4'-desmethyl-deoxypodophyllotoxin [m/z = 385, (+)-mode]. Aligned EICs are shown and are all to scale. Enzyme assays lacking NADPH or containing microsomes isolated from *WAT11* harboring empty vector served as negative controls. No activity was observed on (–)-podophyllotoxin, (–)-5'-desmethoxy-podophyllotoxin or (–)-yatein.

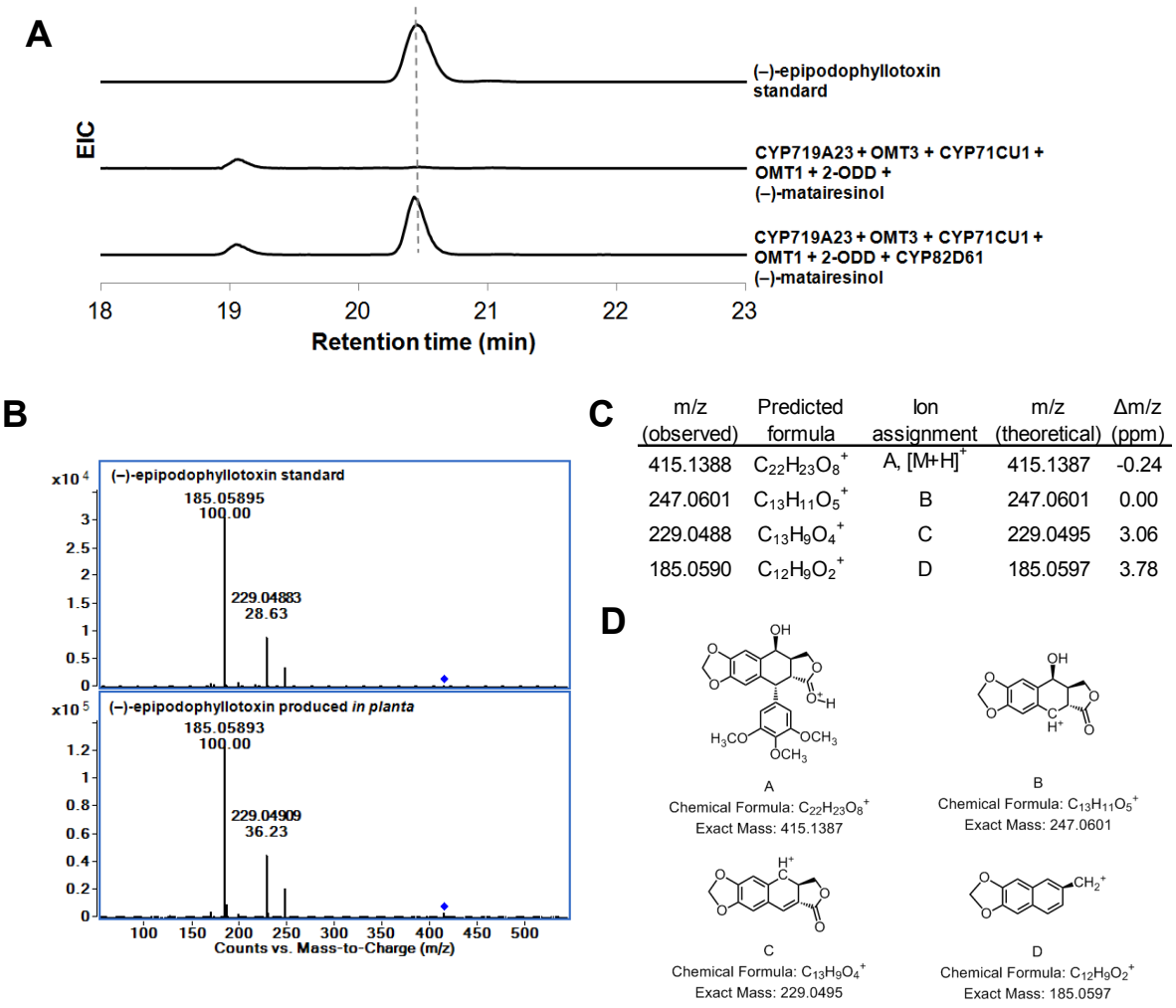


Fig. S23. *N. benthamiana* transient expression of CYP719A23, OMT3, CYP71CU1, OMT1, 2-ODD and CYP82D61, and (–)-matairesinol infiltration produces (–)-epipodophyllotoxin.

(A) Aligned EICs of (–)-epipodophyllotoxin [m/z = 415, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + CYP82D61, compared to CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD and an authentic standard. Only chromatograms from *in planta* expression are to scale. Interestingly, no production of glucosylated derivates by endogenous tobacco enzymes is observed, despite the introduction of a free hydroxyl. **(B)** MS/MS spectra (10 V) of (–)-epipodophyllotoxin produced *in planta* compared to authentic standard. **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.

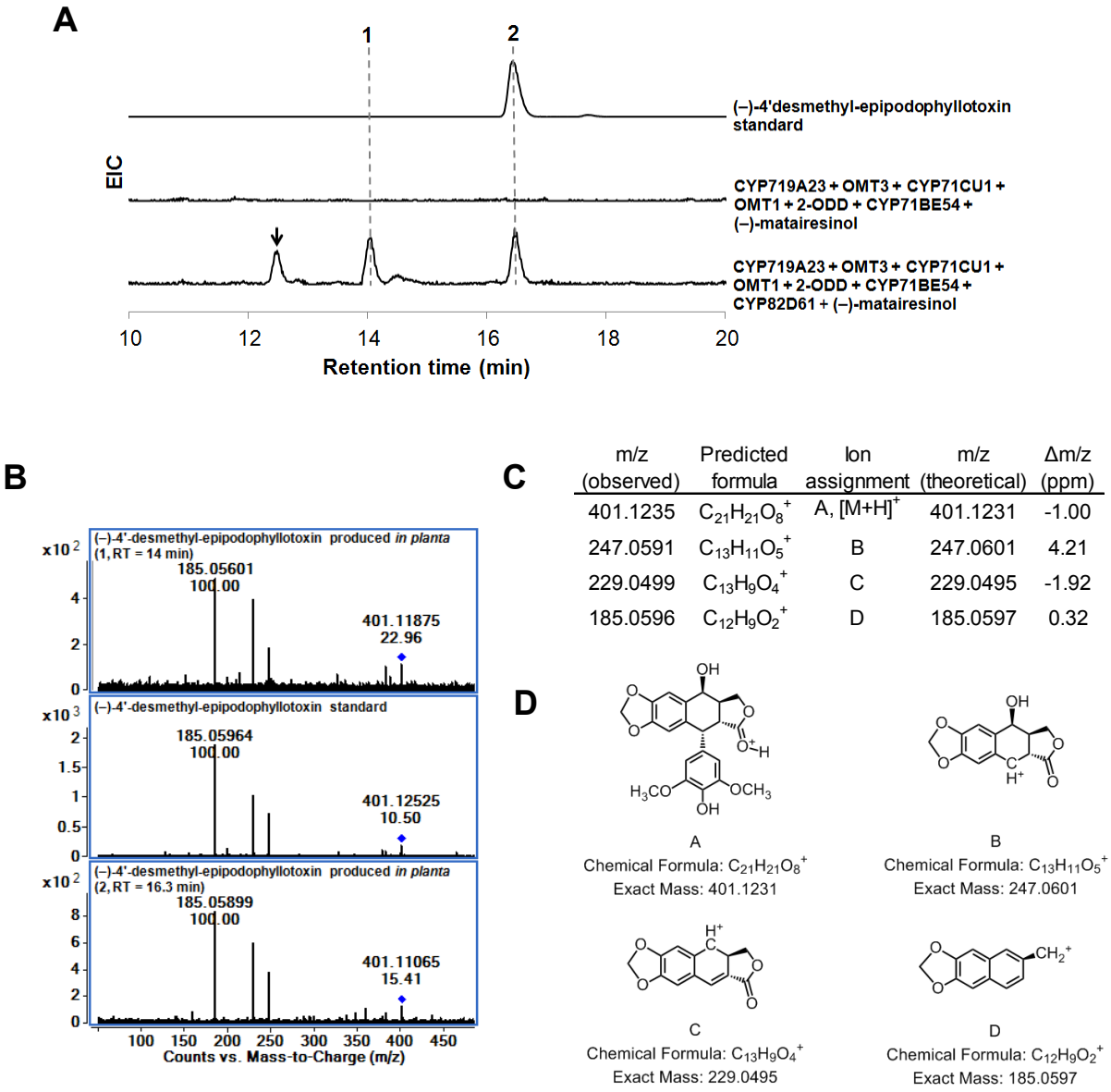
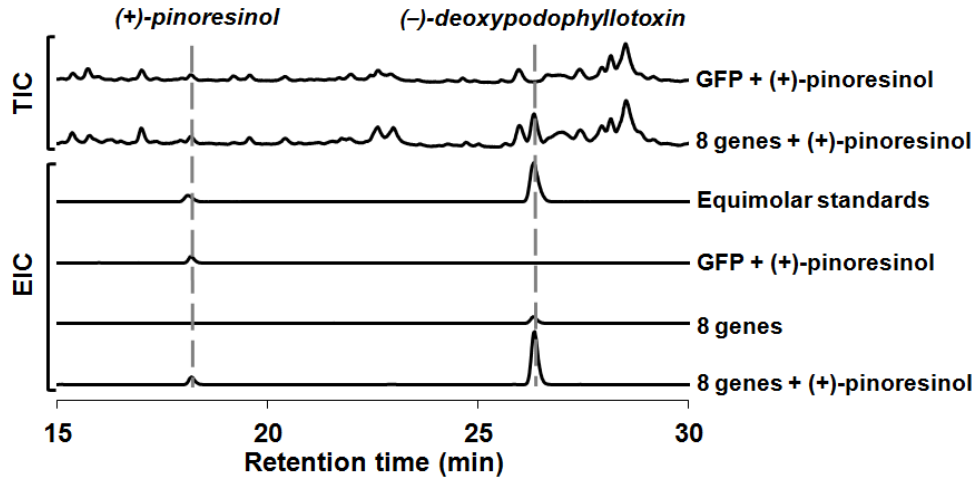


Fig. S24. ***N. benthamiana*** **transient expression of CYP719A23, OMT3, CYP71CU1, OMT1, 2-ODD, CYP71BE54 and CYP82D61, and (–)-matairesinol infiltration produces (–)-4'-desmethyl-epipodophyllotoxin, the etoposide alglycone.**

(A) Aligned EICs of 4'-desmethyl-epipodophyllotoxin [m/z = 401, (+)-mode] from the transient expression of CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + CYP71BE54 + CYP82D61, compared to CYP719A23 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + CYP71BE54 and an authentic standard; peaks observed at earlier labeled retention times (1 and arrow) are in-source fragments of (–)-pluviatolide glucosylated derivatives produced by endogenous tobacco enzymes.. Only chromatograms from *in planta* expression are to scale. **(B)** MS/MS spectra (5 V) of (–)-4'-desmethyl-epipodophyllotoxin produced *in planta* (from labeled retention times, 1 and 2) compared to authentic standard **(C)** Table of MS/MS peak assignments and **(D)** putative ion structures.

A

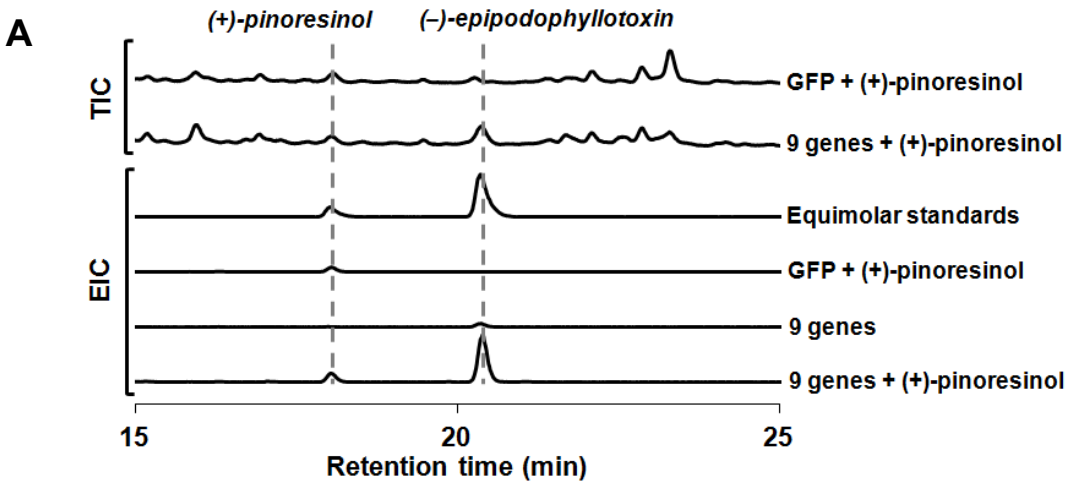


B

Genes	Substrate	(–)-deoxypodophyllotoxin [ng/mg of plant dry weight]
8 genes	(+)-pinoresinol	87.1 ± 9.7
8 genes	None	11.4 ± 3.8
GFP	(+)-pinoresinol	ND
GFP	None	ND

Fig. S25. *In planta* reconstitution of (–)-deoxypodophyllotoxin biosynthesis from (+)-pinoresinol.

(A) Total ion chromatogram (TIC) and combined extracted ion chromatograms (EIC) for (+)-pinoresinol (m/z = 341, $[M\text{-}H_2O]^+$) and (–)-deoxypodophyllotoxin (m/z = 399) are compared between the *in planta* expression of GFP with (+)-pinoresinol infiltration and the *in planta* expression of DIR, PLR, SDH, CYP719A23, OMT3, CYP71CU1, OMT1, and 2-ODD (8 genes) with and without (+)-pinoresinol infiltration. An equimolar standard (10 μM) of (+)-pinoresinol and (–)-deoxypodophyllotoxin is shown. **(B)** Amounts of (–)-deoxypodophyllotoxin produced *in planta* were quantified using a standard curve constructed from an authentic standard. Standard deviation is based on three biological replicates.



B

Genes	Substrate	(–)-epipodophyllotoxin [ng/mg of plant dry weight]
9 genes	(+)-pinoresinol	57.8 ± 6.4
9 genes	None	< 5
GFP	(+)-pinoresinol	ND
GFP	None	ND

Fig. S26. *In planta* reconstitution of (–)-epipodophyllotoxin biosynthesis from (+)-pinoresinol.

(A) Total ion chromatogram (TIC) and combined extracted ion chromatograms (EIC) for (+)-pinoresinol (m/z = 341, $[M\text{-}H_2O]^+$) and (–)-epipodophyllotoxin (m/z = 415) are compared between the *in planta* expression of GFP with (+)-pinoresinol infiltration and the *in planta* expression of DIR, PLR, SDH, CYP719A23, OMT3, CYP71CU1, OMT1, 2-ODD, and CYP82D61 (9 genes) with and without (+)-pinoresinol infiltration. An equimolar standard (10 µM) of (+)-pinoresinol and (–)-epipodophyllotoxin is shown.

(B) Amounts of (–)-epipodophyllotoxin produced *in planta* were quantified using a standard curve constructed from an authentic standard. Standard deviation is based on three biological replicates.

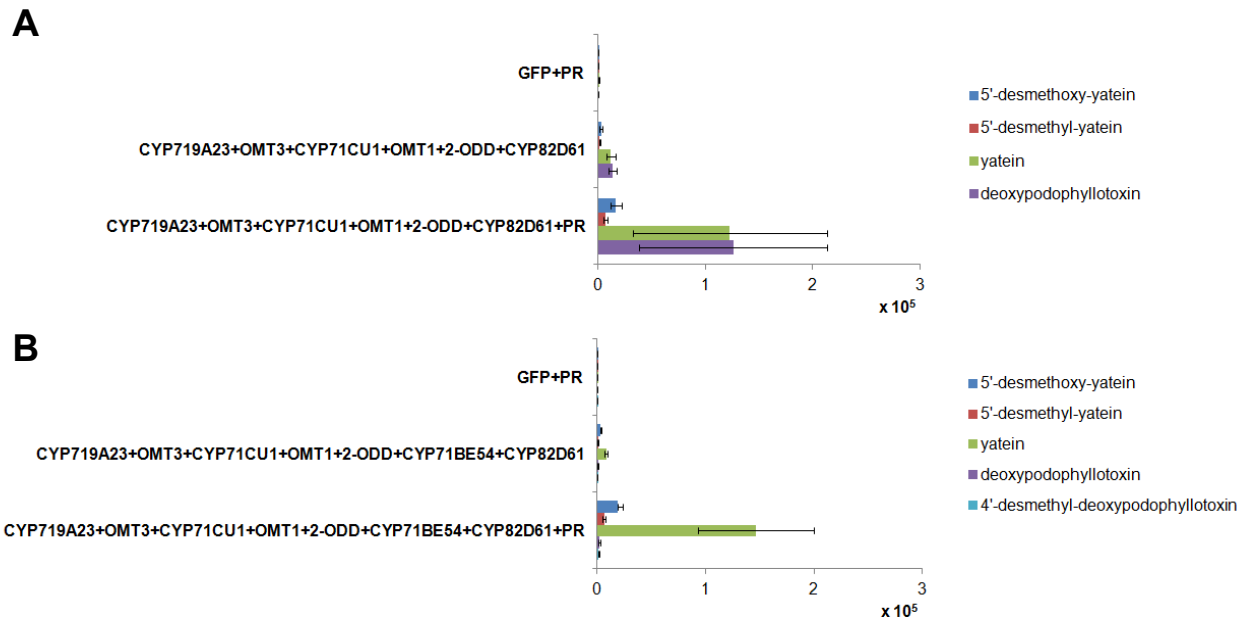


Fig. S27. Ion abundance of intermediates from *in planta* reconstitution of (–)-epipodophyllotoxin and (–)-4'-desmethyl-epipodophyllotoxin biosynthesis from (+)-pinoresinol.

(A) Ion abundance of intermediates from the reconstitution of (–)-epipodophyllotoxin biosynthesis from (+)-pinoresinol [PR] with and without PR infiltration compared to GFP with PR infiltration. **(B)** Ion abundance of intermediates from the reconstitution of (–)-4'-desmethyl-epipodophyllotoxin biosynthesis from PR with and without PR infiltration compared to GFP with PR infiltration. In both cases, ion abundances of intermediates are significantly less than the maximum amount observed to accumulate when the intermediates are produced as final products as in Fig. 4A.

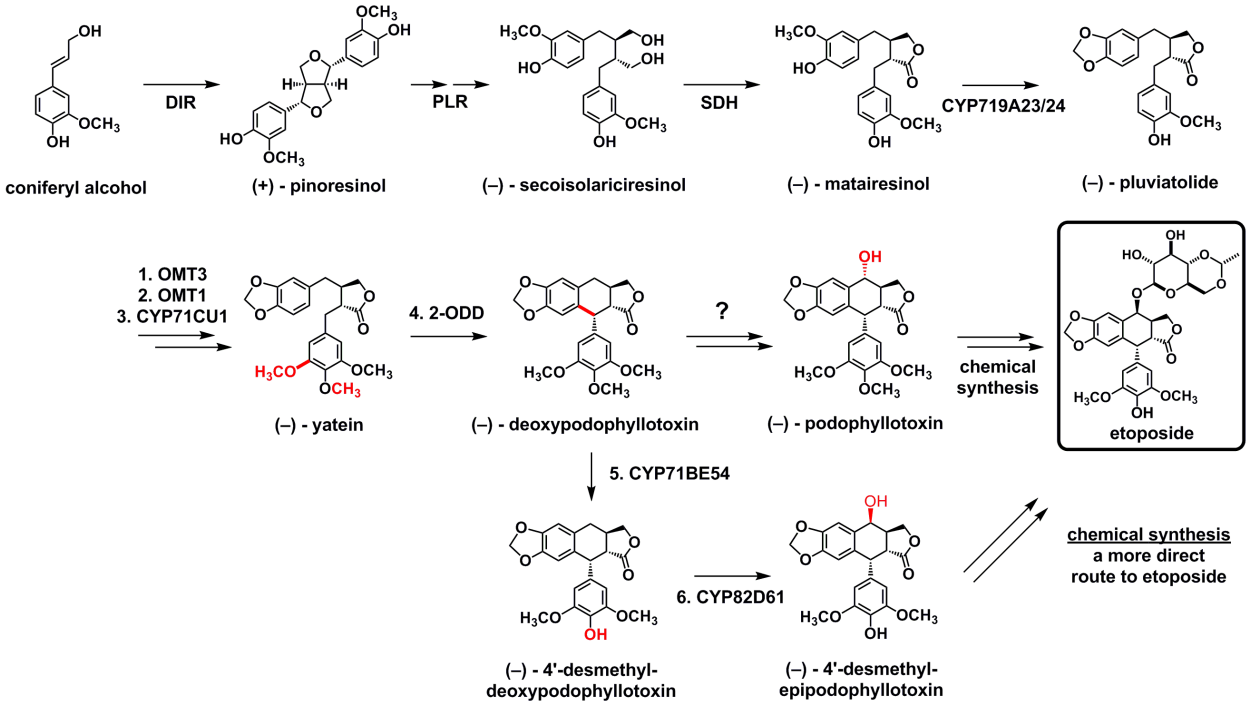


Figure S28. Biosynthetic route to (–)-4'-desmethyl-epipodophyllotoxin, the etoposide alygcone.

The discovery of the six enzymes in this report (OMT3, CYP71CU1, OMT1, 2-ODD, CYP71BE54 and CYP82D61) enable the production of (–)-4'-desmethyl-epipodophyllotoxin, the etoposide alygcone in a heterologous host such as *N. benthamiana*. This maps a simpler and more direct route to etoposide that circumvents the semisynthetic epimerization and demethylation currently required for etoposide production from (–)-podophyllotoxin isolated from Mayapple.

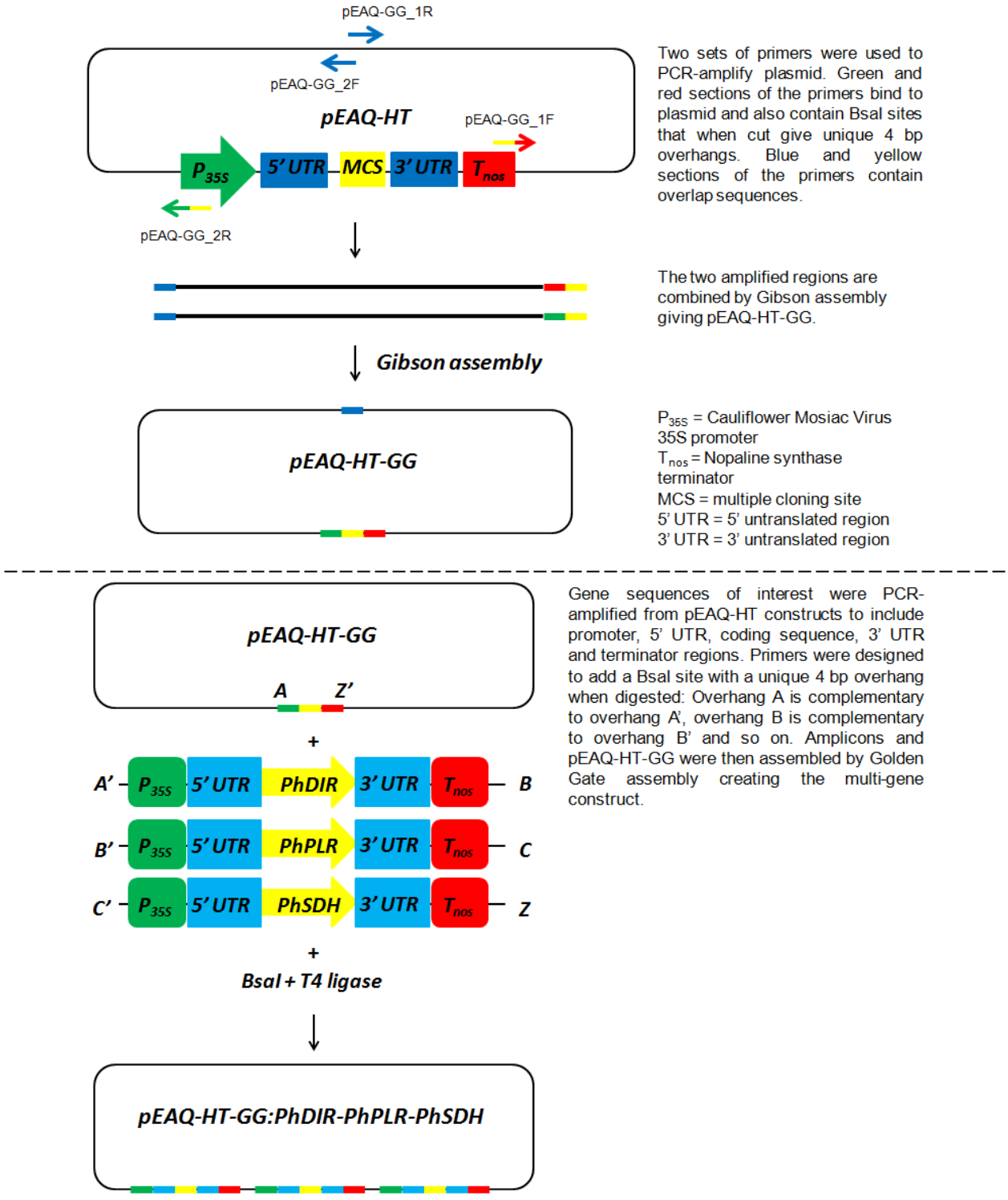


Figure S29. Golden Gate assembly strategy for multiple-gene constructs for transient expression in *N. benthamiana*.

A CYP719 + **OMT3** vs. CYP719 + OMT3 no mataireSinol

rank	mass and retention time	fold change	average ion abundance
1)	**M371T1711**	**487**	**3.21E+06**
2)	M393T1700	122	1.83E+06
3)	**M763T1711**	**10950**	**1.43E+06**
4)	**M388T1711**	**442**	**6.92E+05**
5)	**M372T1711**	**694**	**6.75E+05**
6)	**M764T1711**	**3450**	**6.44E+05**
7)	**M394T1711**	**1047**	**3.99E+05**
8)	**M151T1711**	**73**	**3.69E+05**
9)	M371T1701	11	3.06E+05
10)	**M353T1711**	**40**	**2.97E+05**

B CYP719 + OMT3 + **CYP71CU1** vs. CYP719 + OMT3

rank	mass and retention time	fold change	average ion abundance
1)	**M387T1233**	**36**	**1.31E+06**
2)	**M657T1233**	**42**	**1.12E+06**
3)	**M337T1233**	**23**	**6.12E+05**
4)	**M429T1233**	**32**	**5.24E+05**
5)	M571T1125	17	3.86E+05
6)	**M658T1233**	**31**	**3.55E+05**
7)	**M652T1233**	**34**	**3.38E+05**
8)	**M388T1233**	**28**	**2.84E+05**
9)	**M599T1233**	**28**	**2.84E+05**
10)	**M453T1233**	**27**	**2.56E+05**
11)	**M387T1496**	**13**	**2.37E+05**

C CYP719 + OMT3 + CYP71CU1 + **B38** vs. CYP719 + OMT3

rank	mass and retention time	fold change	average ion abundance
1)	**M401T1662**	**265**	**1.51E+07**
2)	**M823T1663**	**3146**	**4.46E+06**
3)	**M423T1663**	**194**	**4.40E+06**
4)	M195T718	68	3.46E+06
5)	**M402T1662**	**116**	**3.15E+06**
6)	**M824T1663**	**5518**	**2.02E+06**
7)	M195T805	18	1.65E+06
8)	M235T718	50	1.54E+06
9)	**M418T1662**	**101**	**1.07E+06**
10)	**M424T1663**	**160**	**9.51E+05**

D CYP719 + OMT3 + **2-ODD** vs. CYP719 + OMT3

rank	mass and retention time	fold change	average ion abundance
1)	**M231T1579**	**933**	**6.75E+06**
2)	**M759T1579**	**Inf**	**3.60E+06**
3)	**M391T1579**	**11924**	**3.50E+06**
4)	**M386T1579**	**14532**	**2.50E+06**
5)	M476T918	79	2.25E+06
6)	**M760T1579**	**Inf**	**1.56E+06**
7)	**M232T1579**	**1106**	**8.37E+05**
8)	**M225T1579**	**Inf**	**7.76E+05**
9)	**M392T1579**	**14857**	**7.29E+05**
10)	M477T919	131	5.51E+05

E CYP719 + OMT3 + CYP71CU1 + OMT1 + **2-ODD** vs. CYP719 + OMT3

rank	mass and retention time	fold change	average ion abundance
1)	**M399T1582**	**233**	**7.40E+06**
2)	**M421T1582**	**109**	**4.52E+06**
3)	**M819T1582**	**268**	**4.13E+06**
4)	**M231T1582**	**170**	**3.97E+06**
5)	**M820T1582**	**261**	**1.87E+06**
6)	**M400T1582**	**214**	**1.55E+06**
7)	**M416T1582**	**189**	**1.27E+06**
8)	**M422T1583**	**97**	**9.77E+05**
9)	**M821T1582**	**241**	**5.30E+05**
10)	**M232T1582**	**182**	**5.10E+05**

F CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + **CYP71BE54** vs. CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD

rank	mass and retention time	fold change	average ion abundance
1)	**M231T1163**	**42**	**8.32E+05**
2)	**M655T1163**	**304**	**6.32E+05**
3)	**M385T1163**	**160**	**4.05E+05**
4)	**M336T1163**	**161**	**3.69E+05**
5)	**M656T1163**	**297**	**2.07E+05**
6)	**M231T1366**	**8**	**1.72E+05**
7)	**M407T1366**	**56**	**1.71E+05**
8)	M650T1194	7	1.69E+05
9)	**M597T1163**	**112**	**1.66E+05**
10)	**M232T1163**	**13**	**1.16E+05**
14)	**M385T1366**	**13**	**7.67E+04**

G CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + **CYP82D61** vs. CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD

rank	mass and retention time	fold change	average ion abundance
1)	**M415T1226**	**53**	**3.81E+06**
2)	**M437T1226**	**21**	**2.83E+06**
3)	**M851T1226**	**89**	**2.34E+06**
4)	**M852T1226**	**80**	**1.08E+06**
5)	**M416T1226**	**46**	**8.35E+05**
6)	**M248T1226**	**39**	**7.68E+05**
7)	**M432T1226**	**43**	**7.35E+05**
8)	**M438T1226**	**19**	**6.27E+05**
9)	**M229T1226**	**36**	**6.09E+05**
10)	**M185T1226**	**23**	**5.37E+05**

H CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD + **CYP71BE54** + **CYP82D61** vs. CYP719 + OMT3 + CYP71CU1 + OMT1 + 2-ODD

rank	mass and retention time	fold change	average ion abundance
1)	**M585T745**	**139**	**2.86E+05**
2)	**M671T840**	**44**	**1.60E+05**
3)	**M344T839**	**12**	**1.54E+05**
4)	**M401T839**	**9**	**1.18E+05**
5)	M399T765	7	9.45E+04
6)	**M586T745**	**67**	**8.83E+04**
7)	M601T765	39	5.75E+04
8)	**M401T745**	**6**	**5.58E+04**
9)	**M229T839**	**7**	**5.52E+04**
10)	**M666T840**	**10**	**4.98E+04**

Figure S30. XCMS analysis summary of in planta candidate enzyme screening.

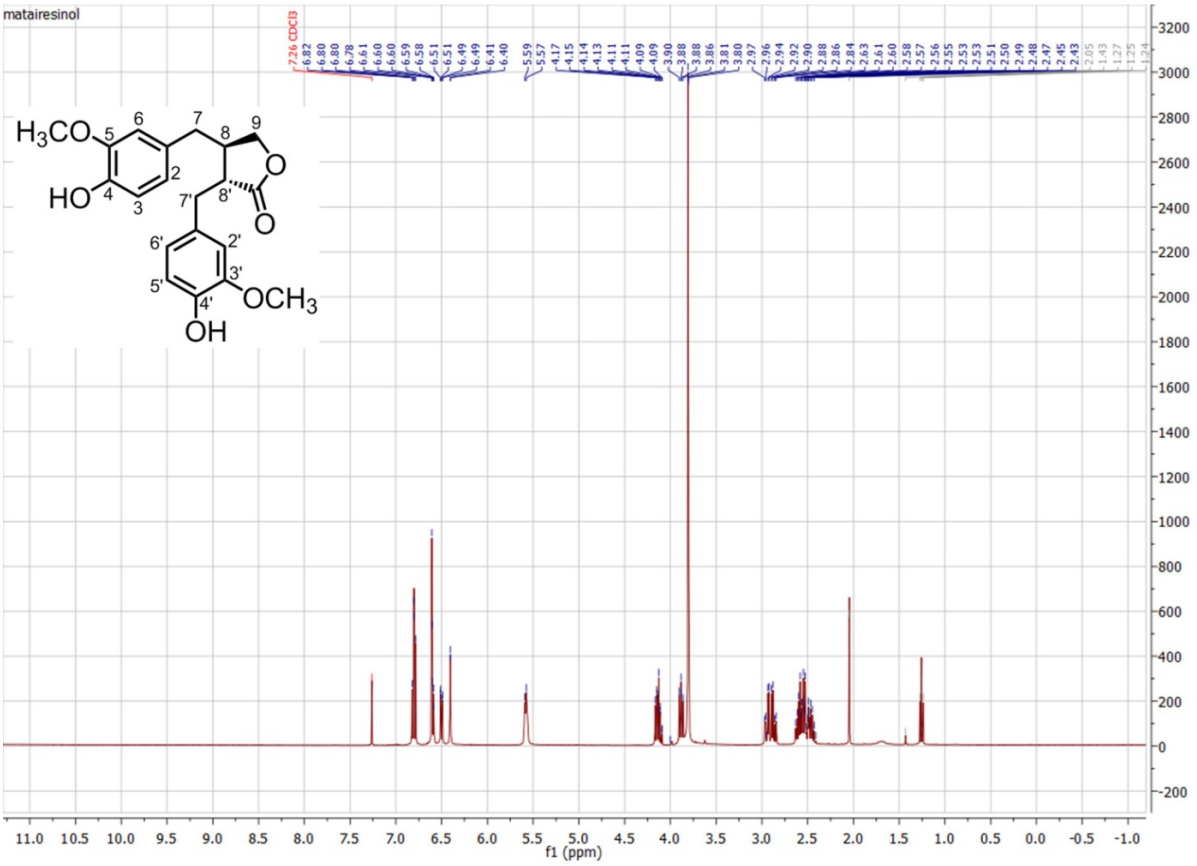
Results from the computational comparison of the metabolomes generated from the transient expression of different gene combinations in *N. benthamiana* are summarized here. The top ten most abundant mass signals that differ for a given co-infiltration relative to a control are reported, along with other mass signals of interest. Results were reproduced at least thrice, each time with three biological replicates for each sample condition. In bold are mass signals derived from identified products in this report (isotope peaks, adducts, in-source fragments, derivatives, etc.). Highlighted in grey are important parent, fragment or adduct ions that have been referred to throughout this report: **(A)** (–)-5'-desmethoxy-yatein $[M+H]^+$, m/z = 371 **(B)** (–)-5'-desmethyl-yatein $[M+H]^+$, m/z = 387 **(C)** (–)-yatein $[M+H]^+$, m/z = 401 **(D)** (–)-5'-desmethoxy-deoxypodophyllotoxin $[M+NH_4]^+$, m/z = 386; $[M+Na]^+$, m/z = 391 **(E)** (–)-deoxypodophyllotoxin $[M+H]^+$, m/z = 399 **(F)** (–)-4'-desmethyl-deoxypodophyllotoxin $[M+H]^+$, m/z = 385 **(G)** (–)-epipodophyllotoxin $[M+H]^+$, m/z = 415 **(H)** (–)-4'-desmethyl-epipodophyllotoxin $[M+H]^+$, m/z = 401

NMR Spectra of Authentic Standards

Minor impurities and residual solvent impurities are marked in grey. $CDCl_3$ solvent peak is marked in red.

Figure S31. (–)-matairesinol

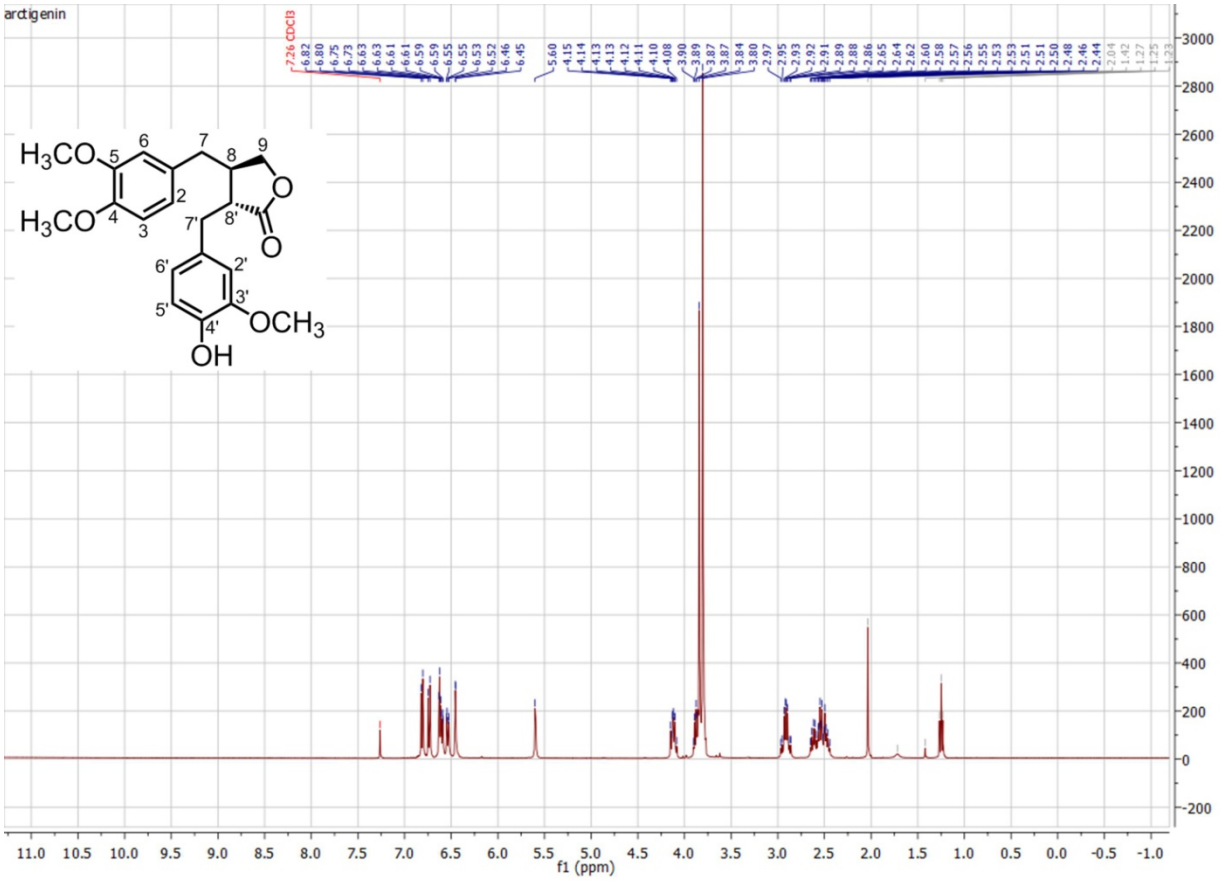
^{1}H NMR spectrum of (–)-matairesinol in $CDCl_3$ and comparison to literature (46)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
2, 3, 5', 6'	6.50 (dd, J = 8.0, 1.9 Hz, 1H), 6.58 – 6.60 (m, 1H), 6.79 (d, J = 7.8 Hz, 1H), 6.81 (d, J = 8.0 Hz, 1H)	6.51 (d, J = 7.8 Hz, 1H), 6.60 (d, J = 7.3 Hz, 1H), 6.80 (d, J = 8.3 Hz, 1H), 6.82 (d, J = 7.8 Hz, 1H),
4-OH, 4'-OH	5.57 (s, 1H), 5.59 (s, 1H)	5.49 (s, 1H), 5.52 (s, 1H)
5-OCH_3, 3'-OCH_3	3.80 (s, 3H), 3.81 (s, 3H)	3.81 (s, 3H), 3.82 (s, 3H)
6, 2'	6.40 (d, J = 1.9 Hz, 1H), 6.61 (s, 1H)	6.41 (s, 1H), 6.61 (s, 1H)
7, 8, 8'	2.41-2.64 (m, 4H)	2.42 - 2.64 (m, 4H)
7'	2.87 (dd, 1H, 14, 7 Hz), 2.95 (dd, 1H, J = 14, 5 Hz)	2.88 (dd, 1H, J = 14.2, 6.8 Hz), 2.95 (dd, 1H, J = 14.2, 5.9 Hz)
9	3.88 (dd, 1H, J = 9, 7 Hz), 4.09-4.18 (m, 1H)	3.89 (dd, 1H, J = 6.8, 6.8 Hz), 4.15 (dd, 1H, J = 9.3, 7.8 Hz)

Figure S32. (–)-arctigenin

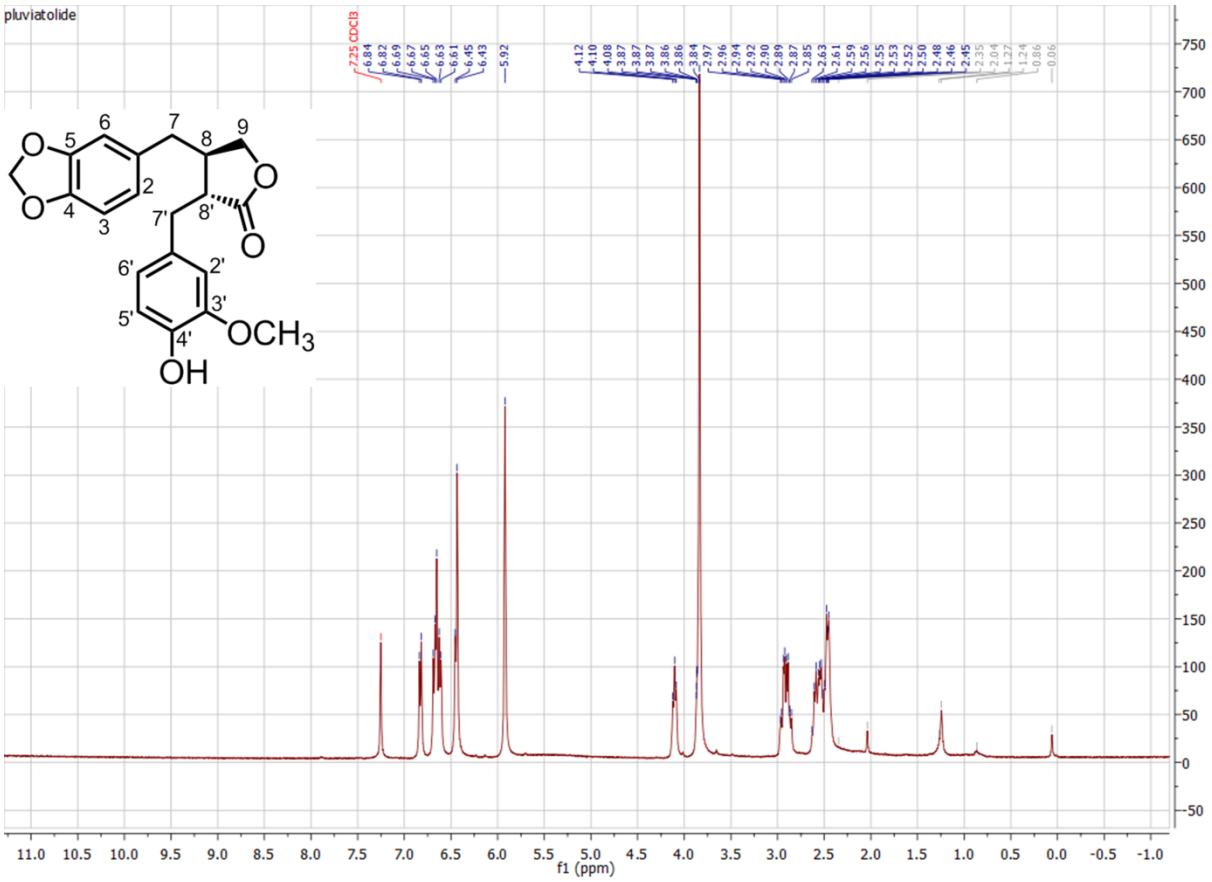
^{1}H NMR spectrum of (–)-arctigenin in CDCl3 and comparison to literature (46)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
2, 3, 6, 2', 5', 6'	6.45 (d, J = 2.1 Hz, 1H), 6.54 (dd, J = 8.1, 2.1 Hz, 1H), 6.58 – 6.64 (m, 2H), 6.74 (d, J = 8.1 Hz, 1H), 6.81 (d, J = 7.9 Hz, 1H)	6.46 (d, J = 2.0 Hz, 1H), 6.55 (dd, J = 7.8, 2.0 Hz, 1H), 6.61 (dd, J = 7.8, 2.0 Hz, 1H), 6.64 (d, J = 2.0 Hz, 1H), 6.75 (d, J = 7.8 Hz, 1H), 6.82 (d, J = 7.8 Hz, 1H)
4'-OH	5.60 (s, 1H)	5.55 (s, 1H)
5-OCH_3, 4-OCH_3, 3'-OCH3	3.80 (s, 6H), 3.84 (s, 3H)	3.81 (s, 6H), 3.85 (s, 3H)
7, 8, 8'	2.40 – 2.68 (m, 4H)	2.43 - 2.67 (m, 4H)
7'	2.89 (dd, J = 14.1, 6.6 Hz, 1H), 2.94 (dd, J = 14.1, 5.3 Hz, 3H)	2.85 - 3.00 (m, 2H)
9	3.88 (dd, J = 9.6, 1.8 Hz, 1H), 4.06 – 4.17 (m, 1H)	3.88 (dd, J = 8.8, 7.3 Hz, 1H), 4.13 (dd, J = 8.8, 6.8 Hz, 1H)

Figure S33. (–)-pluviatolide

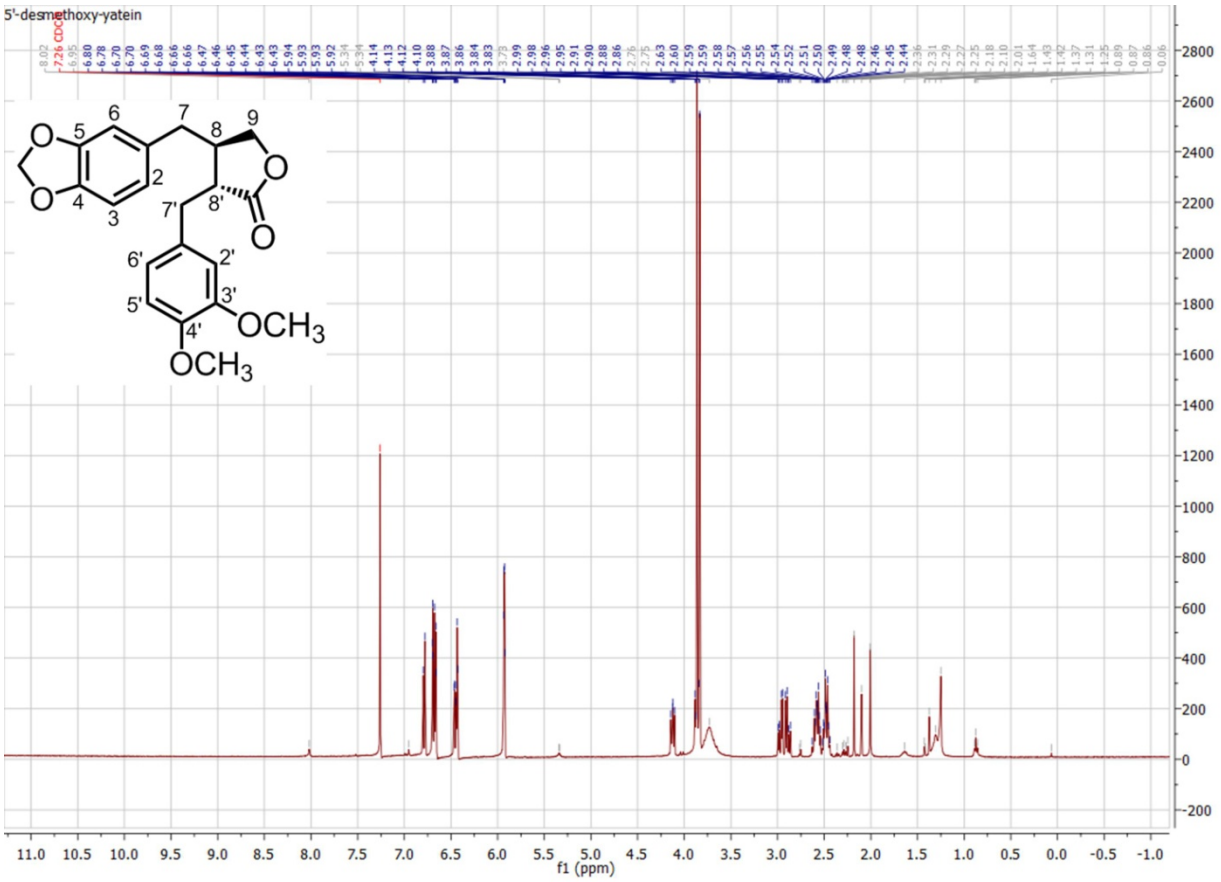
^{1}H NMR spectrum of (–)-pluviatolide in $CDCl_3$ and comparison to literature (47)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
2, 3, 6, 2', 5', 6'	6.41 – 6.48 (m, 2H), 6.58 – 6.71 (m, 3H), 6.83 (d, J = 8.0 Hz, 1H)	6.44-6.47 (m, 2H), 6.63 (dd, J=1.8, 7.9 Hz, 1H), 6.67 (J = 1.8 Hz, d, 1H), 6.69 (d, J=7.7 Hz, 1H), 6.84 (d, J=8 Hz, 1H)
4, 5 O-CH_2-O	5.92 (s, 2H)	5.93 (d, J=1.4 Hz, 1H), 5.94 (d, J=1.4 Hz, 1H)
3'-OCH_3	3.84 (s, 3H)	3.85 (s, 3H)
7, 8, 8'	2.39 – 2.65 (m, 4H)	2.45 - 2.62 (m, 4H)
7'	2.88 (dd, J = 14.1, 6.8 Hz, 1H), 2.95 (dd, J = 14.0, 5.0 Hz, 1H),	2.89 (dd, J=7.0, 14.1 Hz, 1H), 2.96 (dd, J=5.2, 14.0 Hz, 1H)
9	3.85 – 3.91 (m, 1H), 4.06 – 4.15 (m, 1H)	3.86 (dd, J=7.4, 9.1 Hz, 1H), 4.11 (dd, J=7.1, 9.2 Hz, 1H)

Figure S34. (–)-5'-desmethoxy-yatein

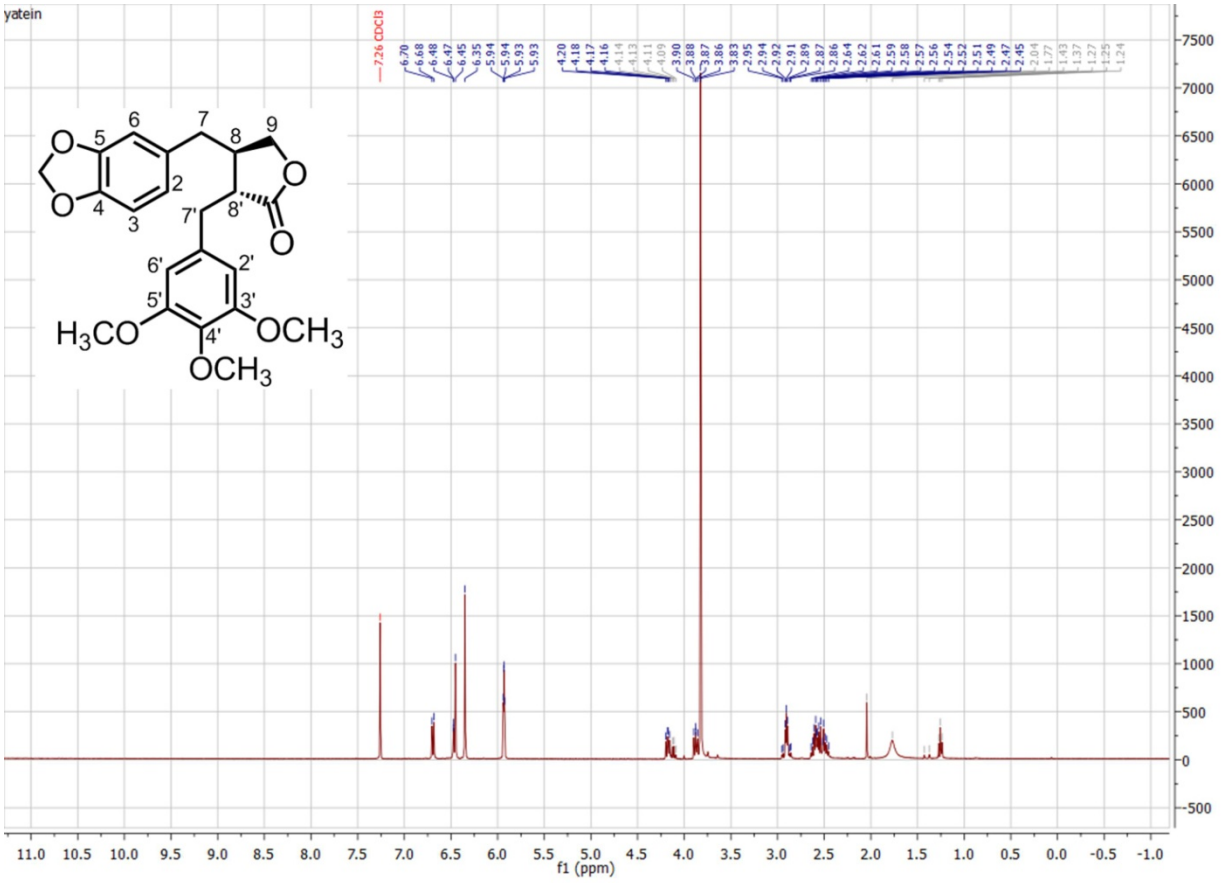
^{1}H NMR spectrum of (–)-5'-desmethoxy-yatein in $CDCl_3$ and comparison to literature (48)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
2, 3, 6, 2', 5', 6'	6.42 – 6.48 (m, 2H), 6.63 – 6.71 (m, 3H), 6.76 – 6.81 (m, 1H)	6.41-6.45 (m, 2H) 6.64-6.68 (m, 3H), 6.77 (d, J = 8.0, 1H)
4, 5 O-CH_2-O	5.92 (d, J = 1.5 Hz, 1H), 5.93 (d, J = 1.4 Hz, 1H)	5.90-5.91 (m, 2H)
3'-OCH_3, 4'-OCH_3	3.83 (s, 3H), 3.86 (s, 3H	3.82 (s, 3H), 3.84 (s, 3H)
7, 8, 8'	2.39 – 2.65 (m, 4H)	2.44-2.58 (m, 4H)
7'	2.89 (dd, J = 14.1, 7.0 Hz, 1H), 2.97 (dd, J = 14.1, 5.1 Hz, 1H)	2.87 (dd, J = 7.1, 14.2, 1H), 2.95 (dd, J = 5.1, 14.0, 1H)
9	3.87 – 3.89 (m, 1H), 4.12 (dd, J = 9.1, 6.7 Hz, 1H)	3.84 (m, 1H), 4.10 (dd, J = 6.8, 9.0 Hz, 1H)

Figure S35. (–)-yatein

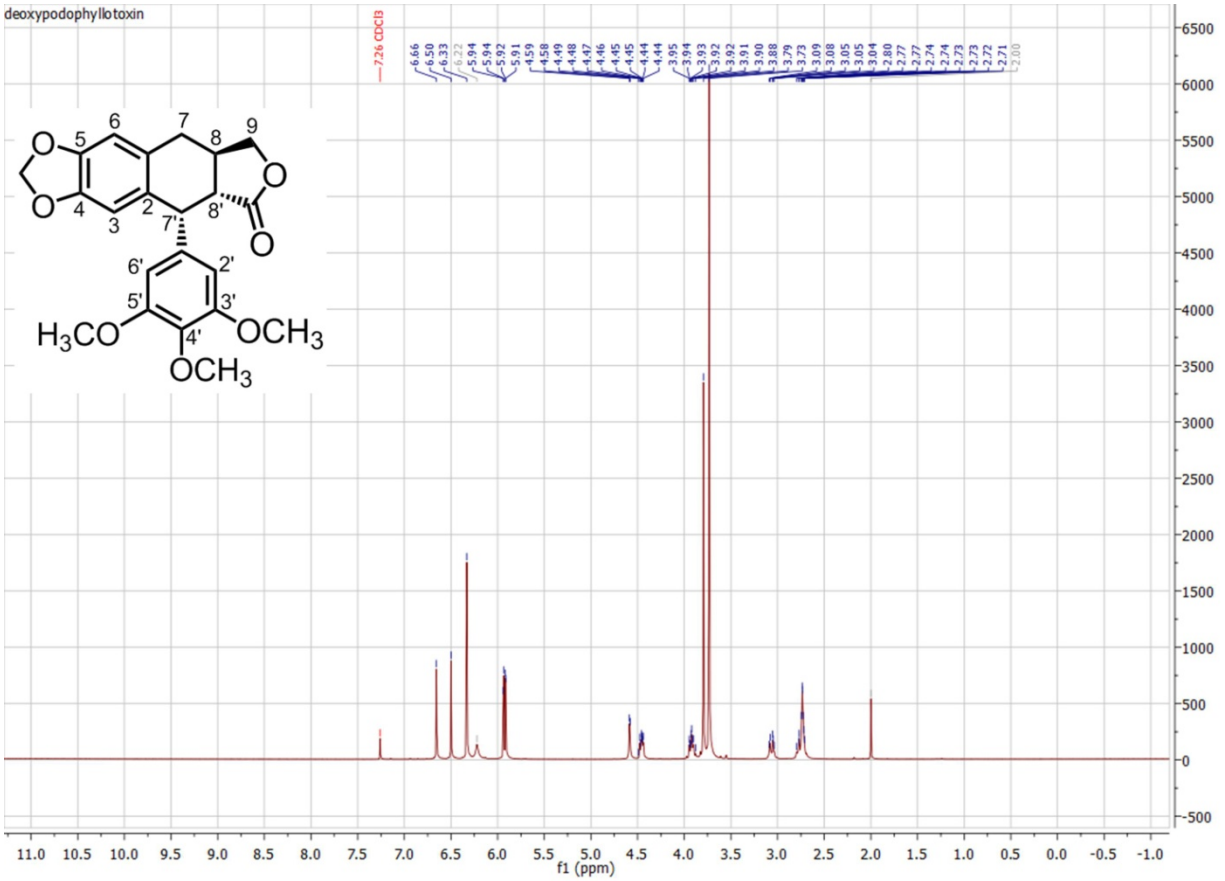
^{1}H NMR spectrum of (–)-yatein in $CDCl_3$ and comparison to literature (43)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
2, 3, 6, 2', 6'	6.35 (s, 2H), 6.43 – 6.50 (m, 2H), 6.69 (d, J = 7.7 Hz, 1H)	6.36 (s, 2H), 6.46 (d, J = 1.7 Hz, 1H), 6.47 (dd, J = 6.9, 1.7 Hz, 1H), 6.69 (d, J = 8.3 Hz, 1H)
4, 5 O-CH_2-O	5.93 (d, J = 1.5 Hz, 1H), 5.94 (d, J = 1.4 Hz, 1H)	5.93 (d, J = 1.5 Hz, 1H), 5.94 (d, J = 1.5 Hz, 1H)
3'-OCH_3, 4'-OCH_3, 5'-OCH_3	3.83 (s, 9H)	3.82 (s, 9H)
7, 8, 8'	2.41 – 2.66 (m, 4H)	2.48 - 2.63 (m, 4H)
7'	2.83 – 2.97 (m, 2H)	2.89 - 2.92 (m, 2H)
9	3.88 (dd, J = 9.2, 7.5 Hz, 1H), 4.18 (dd, J = 9.1, 7.2 Hz, 1H)	3.88 (dd, J=9.3, 7.3 Hz, 1H), 4.18 (dd, J = 9.3, 7.3 Hz, 1H)

Figure S36. (–)-deoxypodophyllotoxin

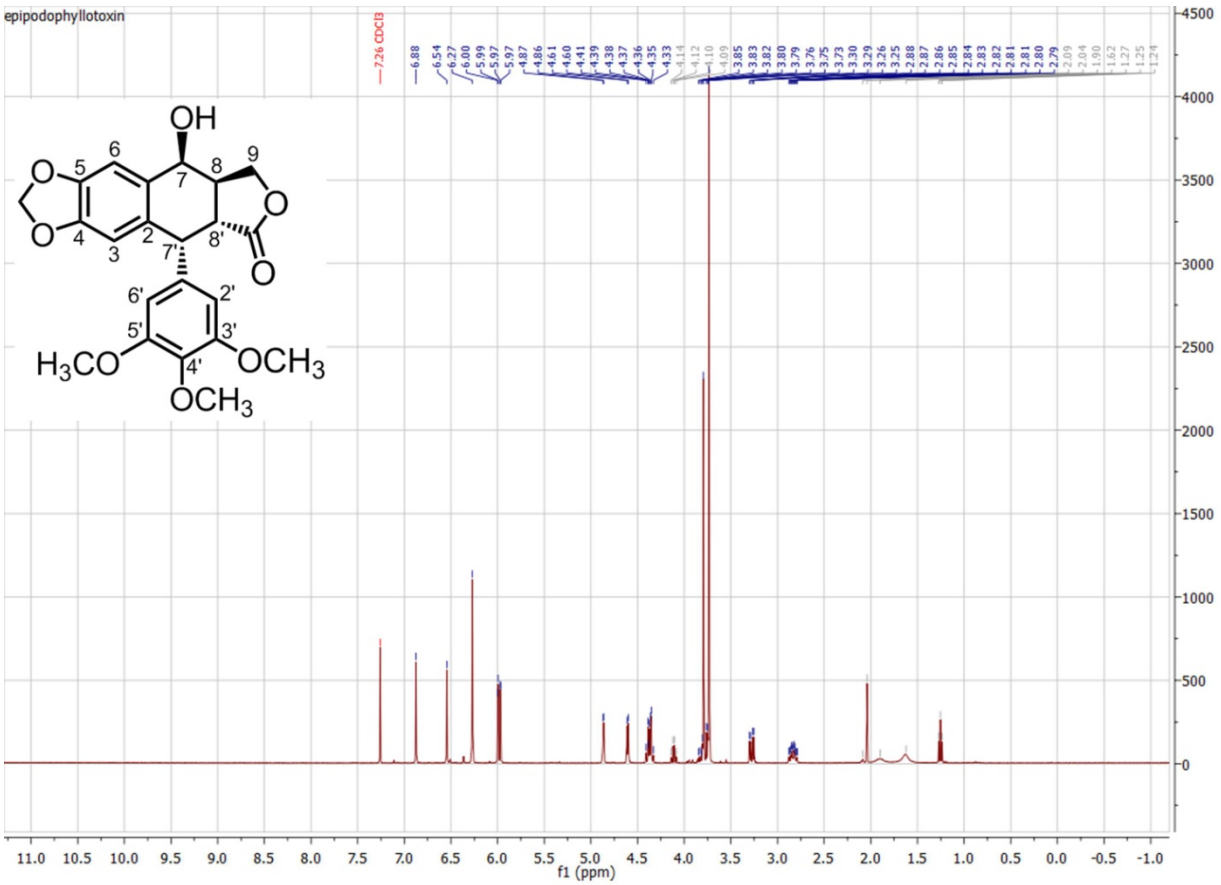
^{1}H NMR spectrum of (–)-deoxypodophyllotoxin in $CDCl_3$ and comparison to literature (49)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
3, 6, 2', 6'	6.33 (s, 2H), 6.50 (s, 1H), 6.66 (s, 1H)	6.34 (s, 2H), 6.52 (s, 1H), 6.67 (s, 1H)
4, 5 O-CH_2-O	5.92 (d, J = 1.3 Hz, 1H), 5.94 (d, J = 1.4 Hz, 2H)	5.93 (d, J = 1.0 Hz, 1H), 5.95 (d, J = 1.0 Hz, 1H)
3'-OCH_3, 4'-OCH_3, 5'-OCH_3	3.73 (s, 6H), 3.79 (s, 3H)	3.75 (s, 6H), 3.80 (s, 3H)
7, 8, 8'	2.69 – 2.82 (m, 3H), 3.01 – 3.12 (m, 1H)	2.74 (m, 2H), 2.77 (m, 1H), 3.08 (d, J = 11.2 Hz, 1H)
7'	4.58 (d, J = 3.2 Hz, 1H)	4.60 (d, J = 2.8 Hz, 1H)
9	3.85 – 3.97 (m, 1H), 4.41 – 4.51 (m, 1H)	3.91 (td, 1H), 4.45 (td, 1H)

Figure S37. (–)-epipodophyllotoxin

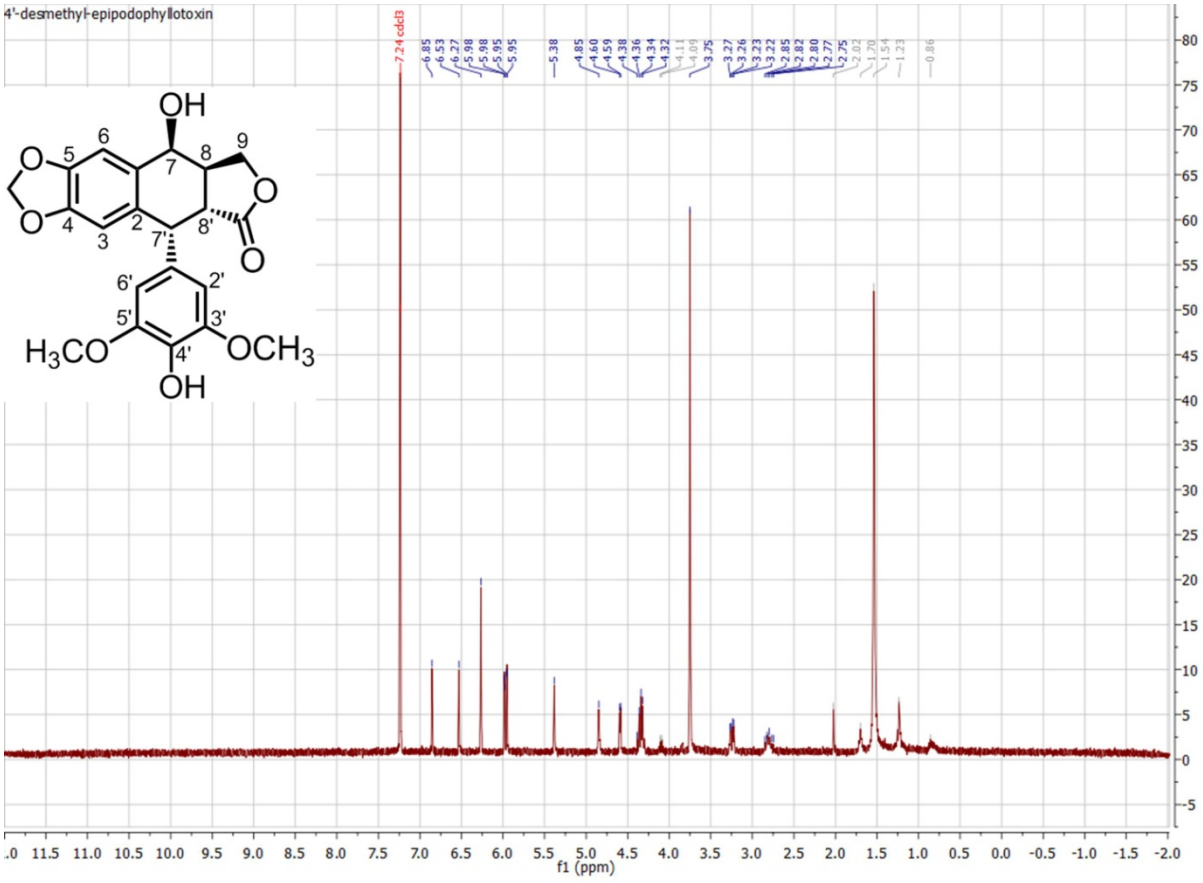
^{1}H NMR spectrum of (–)-epipodophyllotoxin in $CDCl_3$ and comparison to literature (50)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
3, 6, 2', 6'	6.27 (s, 2H), 6.54 (s, 1H), 6.88 (s, 1H)	6.28 (s, 2H), 6.56 (s, 1H), 6.88 (s, 1H)
4, 5 O-CH_2-O	5.97 (d, J = 1.3 Hz, 1H), 6.00 (d, J = 1.3 Hz, 1H)	5.98 (d, J = 1.3 Hz, 1H), 6.00 (d, J = 1.3 Hz, 1H)
3'-OCH_3, 4'-OCH_3, 5'-OCH_3	3.73 (s, 6H), 3.79 (s, 3H)	3.74 (s, 6H), 3.80 (s, 3H)
8, 8'	2.76 – 2.90 (m, 1H), 3.28 (dd, J = 14.1, 5.2 Hz, 1H)	2.84 (dddd, J = 14.1, 10.6, 7.9, 3.4 Hz, 1H), 3.29 (dd, J = 14.1, 5.1 Hz, 1H)
7, 7'	4.61 (d, J = 5.2 Hz, 1H), 4.87 (d, J = 3.4 Hz, 1H)	4.62 (d, J = 5.1 Hz, 1H), 4.88 (dd, J = 4.3, 3.4 Hz, 1H)
9	4.31 – 4.43 (m, 2H)	4.36 (dd, J = 10.6, 8.5 Hz, 1H), 4.39 (dd J = 8.5, 7.9, 1H)

Figure S38. (–)-4'-desmethyl-epipodophyllotoxin

^{1}H NMR spectrum of (–)-4'-desmethyl-epipodophyllotoxin in $CDCl_3$ and comparison to literature (51)



Carbon No.	1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$	Published 1H δ (ppm) [m, J_{HH} (Hz), area] in $CDCl_3$
3, 6, 2', 6'	6.27 (s, 2H), 6.53 (s, 1H), 6.85 (s, 1H)	6.22 (s, 2H), 6.48 (s, 1H), 6.81 (s, 1H)
4, 5 O-CH_2-O	5.95 (d, J = 1.7 Hz, 1H), 5.98 (d, J = 1.7 Hz, 4H)	5.92 (dd, J = 8.8, 1.3 Hz, 2H)
-OH	5.38 (s, 1H)	5.35 (s, 1H)
3'-OCH_3, 5'-OCH_3	3.75 (s, 6H)	3.70 (s, 6H)
8, 8'	2.73 – 2.86 (m, 1H), 3.25 (dd, J = 14.2, 5.1 Hz, 1H)	2.75 (dddd, J = 14.1, 6.3, 3.8, 3.5 Hz, 1H), 3.19 (dd, J = 14.1, 5.1 Hz, 1H)
7, 7'	4.59 (d, J = 5.0 Hz, 1H), 4.82 – 4.87 (m, 1H)	4.54 (d, J = 5.1 Hz, 1H), 4.79 (d, J = 3.5 Hz, 1H)
9	4.30 – 4.41 (m, 2H)	4.27 (d, J = 3.8 Hz, 1H), 4.31 (d, J = 6.3 Hz, 1H)

Table S1 and S2 Legends

Table S1 (Separate File). Hierarchical clustering of RNA-Seq expression data after filtering by enzyme family (CYPs, OMTs, 2-ODDs, PPOs). Heat map depicts gene expression levels; scale is below the heat map. Known biosynthetic genes (black), candidate genes (red), genes identified in this report (red with black arrows).

Table S2 (Separate File). Primers.

References

25. Bhattacharyya, Dipto, et al. "De novo transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, Podophyllum hexandrum." *BMC genomics* 14.1 (2013): 748.
26. Schmittgen, Thomas D., and Kenneth J. Livak. "Analyzing real-time PCR data by the comparative CT method." *Nature protocols* 3.6 (2008): 1101-1108.
27. Gordon, A., and G. J. Hannon. "Fastx-toolkit." *FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab. cshl. edu/fastx_toolkit* (2010).
28. Schulz, Marcel H., et al. "Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels." *Bioinformatics* 28.8 (2012): 1086-1092.
29. Li, Weizhong, and Adam Godzik. "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." *Bioinformatics* 22.13 (2006): 1658-1659.
30. Huang, Xiaoqiu, and Anup Madan. "CAP3: A DNA sequence assembly program."*Genome research* 9.9 (1999): 868-877.
31. Roberts, Adam, and Lior Pachter. "Streaming fragment assignment for real-time analysis of sequencing experiments." *Nature methods* 10.1 (2013): 71-73.
32. Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* 26.1 (2010): 139-140.
33. Wei, Hairong, et al. "Transcriptional coordination of the metabolic network in Arabidopsis." *Plant physiology* 142.2 (2006): 762-774.
34. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
35. Sainsbury, F., E. C. Thuenemann and G. P. Lomonossoff (2009). "pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants." Plant biotechnology journal **7**: 682-693.
36. Gibson, Daniel G., et al. "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nature methods* 6.5 (2009): 343-345.
37. Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. "A one pot, one step, precision cloning method with high throughput capability." *PloS one* 3.11 (2008): e3647.
38. Smith, Colin A., et al. "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification." *Analytical chemistry* 78.3 (2006): 779-787.
39. Gietz, R. Daniel, and Robert H. Schiestl. "High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method." *Nature protocols* 2.1 (2007): 31-34.
40. Pompon, Denis, et al. "Yeast expression of animal and plant P450s in optimized redox environments." *Methods in enzymology* 272.B (1996): 51-64.

41. Umezawa, Toshiaki, Laurence B. Davin, and Norman G. Lewis. "Formation of lignans (-)-secoisolariciresinol and (-)-matairesinol with Forsythia intermedia cell-free extracts." *Journal of Biological Chemistry* 266.16 (1991): 10210-10217
42. Giddings, Lesley-Ann, et al. "A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in Catharanthus roseus." *Journal of Biological Chemistry* 286.19 (2011): 16751-16757.
43. Miyata, Masaru, Kazutaka Itoh, and Sanro Tachibana. "Extractives of Juniperus chinensis L. I: Isolation of podophyllotoxin and yatein from the leaves of J. chinensis." *Journal of wood science* 44.5 (1998): 397-400.
44. Kawazu, K., et al. "Isolation of the cytotoxic constituent deoxypodophyllotoxin from the leaves of Juniperus chinensis." *Scientific Reports of the Faculty of Agriculture-Okayama University (Japan)* 86 (1997):1-5
45. Kamal, Ahmed, N. Laxman, and G. Ramesh. "Facile and efficient one-pot synthesis of 4β-arylaminopodophyllotoxins: synthesis of DNA topoisomerase II inhibitors (NPF and W-68)." *Bioorganic & medicinal chemistry letters* 10.18 (2000): 2059-2062
46. Fischer, J., et al., *Radical carboxyarylation approach to lignans. Total synthesis of (-)-arctigenin,(-)-matairesinol, and related natural products.* Organic letters, 2004. **6**(9): p. 1345-1348.
47. Okunishi, T., T. Umezawa, and M. Shimada, *Enantiomeric compositions and biosynthesis ofWikstroemia sikokiana lignans.* Journal of wood science, 2000. **46**(3): p. 234-242.
48. Rojas-Sepúlveda, A.M., et al., *Cytotoxic podophyllotoxin type-lignans from the steam bark of Bursera fagaroides var. fagaroides.* Molecules, 2012. **17**(8): p. 9506-9519.
43. Miyata, M., K. Itoh, and S. Tachibana, *Extractives of Juniperus chinensis L. I: Isolation of podophyllotoxin and yatein from the leaves of J. chinensis.* Journal of wood science, 1998. **44**(5): p. 397-400.
49. Cho, E.J., et al., *Antibacterial activity and protective effect against gastric cancer by Anthriscus sylvestris fractions.* Horticulture, Environment, and Biotechnology, 2013. **54**(4): p. 326-330.
50. Engelhardt, U., A. Sarkar, and T. Linker, *Efficient Enantioselective Total Synthesis of (−)-Epipodophyllotoxin.* Angewandte Chemie International Edition, 2003. **42**(22): p. 2487-2489.
51. Quirion, Jean-Charles, Deliencourt-Godefroy Geraldine Castelot, Christophe Audouard. “Novel gem-difluorinated c-glycoside compounds derived from podophyllotoxin, their preparation and their applications.” Patent CA 2650384 C. Aug 5, 2014.