PNAS
www.pnas.org

Supplementary Information for

Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants

Lingfeng Mao,[‡,1] Hiroshi Kawaide,[‡,2] Toshiya Higuchi,[‡,3] Meihong Chen,[1] Koji Miyamoto,[4] Yoshiki Hirata,[2] Honoka Kimura,[2] Sho Miyazaki,[2] Miyu Teruya,[3] Kaoru Fujiwara,[3] Keisuke Tomita,[3] Hisakazu Yamane,[4] Ken-ichiro Hayashi,[5] Hideaki Nojiri,[3] Lei Jia,[1] Jie Qiu,[1] Chuyu Ye,[1] Michael P. Timko[6], Longjiang Fan[*,1], and Kazunori Okada[*,3]

[1]Institute of Crop Sciences & Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China; [2]Graduate school of Agriculture, Tokyo University of Agriculture and Technology, Tokyo, Japan; [3]Biotechnology Research Center, The University of Tokyo, Tokyo, Japan; [4]Teikyo University, Department of Biosciences, Tochigi, Japan; [5]Okayama University of Science, Department of Biochemistry, Okayama, Japan; [6] Department of Biology, University of Virginia, Charlottesville, Virginia, USA

‡ These authors contributed equally to this work.

*Corresponding authors
* Kazunori Okada
Email: ukazokad@mail.ecc.u-tokyo.ac.jp
*Longjiang Fan.
Email: fanlj@zju.edu.cn

This PDF file includes:

SI Materials and Methods

De novo genome assembly

MaSuRCA assembler version 3.2.2 (1) was used for our hybrid assembly. Clean Illumina data filtered by NGSQC v2.3.3 (2) with default settings and raw Pacbio data were used to improve the raw assembly in two rounds with SSPACE v1.0 (3), PBjelly (4), and Gapfiller (5) (Table S3). The assembly was polished using Pilon v1.23 (4, 6) with all clean Illumina data at the last step. Meanwhile, we also constructed assembly-only based Illumina data using Soapdenovo2 (7) to improve our hybrid assembly. In the target region containing candidate momilactone gene cluster, Illumina-only assembly exhibited a scaffold, whereas it exhibited two scaffolds in the hybrid assembly. Finally, we merged the two scaffolds from the hybrid assembly according the synteny relationship between the two assemblies. Furthermore, we confirmed the link between the two scaffolds using PCR and Sanger sequencing (Table S5a).

Genome annotation

To annotate repeat elements in the assembly, we first built a *de novo* repeat library using RepeatModeler (8) and predicted repetitive elements of *C. plumiforme* genome using RepeatMasker 4.0.8 (8) with default settings. *ab initio* gene structure predictions were performed using AUGUSTUS 3.2.2 (9), GeneMark.hmm (10), and FGENESH 2.6 (11). Homology evidence for gene structure was searched against the coding sequences of moss *P. patens* v3.3 (reference 17 in the main text) and liverwort *Marchantia polymorpha* v3.1 (reference 23 in the main text) using gmap (12). RNA-seq data from our previous study (reference 5 in the main text) were aligned and the RNA-seq reads were assembled using Tophat 2.1.1 (13), Cufflinks 2.2.1 (14), and Trinity 2.4.0 (15) with default settings. The assembled transcripts from Cufflinks (reference-based) and Trinity (reference-free) were merged for the final transcriptome evidence using PASA 2.0.2 (16). Finally, all the evidence to support gene models were integrated to generate the combined gene set using EVM (15) with higher weight in the evidence of RNA-seq results.

Phylogenetic tree and divergence time estimation. Monophyletic constraints were imposed for the nodes, which were used to calibrate the evolutionary rates (BLOSUM62 and an uncorrelated exponential relaxed model). A Yule speciation process which specifies a constant rate of species divergence was used. Normal priors for land plants–vascular plants split time (mean: 540 mya, std dev: 6.0), gymnosperm–angiosperm split time (mean: 313 mya, std dev: 6.0), monocot–dicot split time (mean: 150.0 mya, std dev: 4.0), and *O. sativa–E. crus-galli* split time (mean: 40.0 mya, std dev: 3.0), were used. The MCMC chains in BEAST were run for 10,000,000 generations while sampling at every 1,000 steps. Convergence between the runs and the amount of burn-in (throwing away some iterations at the beginning of an MCMC run) was determined using Tracer v1.10.4, which was used to assess the effective sample size and to check the consistency of the results. The tree was drawn with FigTree v1.4 (17).

The primary coding sequences (CDS) and protein sequence of *P. patens* and *C. plumiforme* were used to BLAST against each other, the genes with mutual best BLAST hits between them were selected to calculate nonsynonymous (Ka) and synonymous (Ks) substitution rates using KaKs_calculator v2.0 with GMYN model (18). The K_s distribution between *P. patens* and *C. plumiforme* was plotted using ggplot2 (19) and K_s was applied to observe the historical genome duplication of *C. plumiforme*.

Identification of candidate momilactone gene cluster in the plant kingdom

Genome annotated genes of 107 plants were downloaded from Phytozome 12.0 (phytozome.jgi.doe.gov), Ensemble plants (www.plants.ensembl.org/), and other projects (reference 16 in the main text, 20). We first identified the clustered genes through Pfam domains (PF01397 and PF03936 for *TPS* genes; PF00067 for P450; PF13561, PF00106, and PF08659 for *SDR* genes) using pfamscan (*SI Appendix*, Fig. S7). Based on the three types of genes identified above, we scanned the genomes of the 107 plants with a 100 kb window size using in-house scripts to find the candidate regions where the three types of genes co-existed. Adjacent windows containing the three types of genes would be merged. As a control, all known momilactone gene clusters in *Oryza* species and barnyard grass (references 8 and 16 in the main text) must be successfully identified.

RNA extraction and quantitative RT-PCR

For the stress treatment, the *C. plumiforme* gametophores were incubated with BCDATG liquid media (reference 37 in the main text) containing 0.5 mg/mL chitosan (chitosan oligosaccharide lactate from Aldrich, Japan, was dissolved in water by the addition of acetic acid, and the final pH of the solution was adjusted to pH 6.0.) or 0.5 mM aqueous solution of copper (II) chloride dihydrate (Nacalai Tesque, Japan). Total RNA was extracted from the *C. plumiforme* gametophores treated with each of the elicitors at regular time intervals using Sepasol (Nacalai Tesque) and subjected to cDNA synthesis using a PrimeScript RT reagent Kit with gDNA Eraser (Takara Bio, Japan). Quantitative RT-PCR (qRT-PCR) was performed using a Power SYBR Green PCR Master Mix (Applied Biosystems, CA, USA) for the *CpDTC1*/*HpDTC1*, *CpMAS*, *CpCYP970A14*, *CpCYP964A1*, and *CpACT3* genes on an ABI 7500 Fast Real-Time PCR System (Applied Biosystems) with the standard mode according to a previous report (reference 9 in the main text). For each sample, the mean value from triplicate amplifications was used to calculate the transcript abundance. The $2^{-\Delta\Delta CT}$ method was used for a relative quantification of fold changes of the target genes at inductive condition compared to non-treated condition. Sequences of PCR primers used for qRT-PCR analysis are provided in *SI Appendix*, Table S5b.

Functional identification of *CpMAS*

The full‐length cDNA of *CpMAS* was amplified by end‐to‐end RT‐PCR using Ex Taq (Takara, Japan) with the following oligonucleotide as primers: CpMAS‐full‐F, 5′-ATGGCGTCAGGGAAAGAAGC-3′; CpMAS‐full‐R, 5′-CGGTCACCAAGTGAAATGGA-3′. The amplified CpMAS cDNA was cloned into pT7Blue T-vector (Invitrogen, http://www.invitrogen.com/) to generate pT7‐CpMAS, and the sequence was confirmed. For functional analysis, the cDNA fragment including *CpMAS* orf was prepared by restriction enzyme digestion in BamHI and SpeI sites, then the purified fragment was inserted into the same sites of pQE31 vector (Qiagen) to yield pQE-CpMAS. The construct was transformed into *E. coli* JM109. This strain was precultured for 18 h at 37°C in 2 mL of LB medium containing ampicillin (50 μg/mL), and then cultured at 37 °C in 800 mL of LB medium containing ampicillin until the exponential growth phase. Recombinant protein expression was induced by adding isopropyl β-D-thiogalactopyranoside (1 mM) and successively cultured for 18 h at 30 °C. Recombinant proteins were affinity-purified using HisTrap HP column according to the manufacturer's instruction (GE Healthcare Life Sciences). For functional analyses, 100 μL of reaction mixtures containing 0.4 μM of each recombinant protein, 0.2 M 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris)-HCl (pH 8.0), 10 mM 2-mercaptoethanol, 1 mM NAD^+, and 1 μg of 3β-hydroxy-9βH-pimara-7,15-dien-19,6β-olide were prepared. After incubation at 30 °C for 40 min, the mixtures were extracted with ethyl acetate, and the extract was evaporated to dryness *in vacuo*. The residue was dissolved in 100 μL of methanol and subjected to GC-MS and LC-MSMS analysis as described previously (reference 8 in the main text).

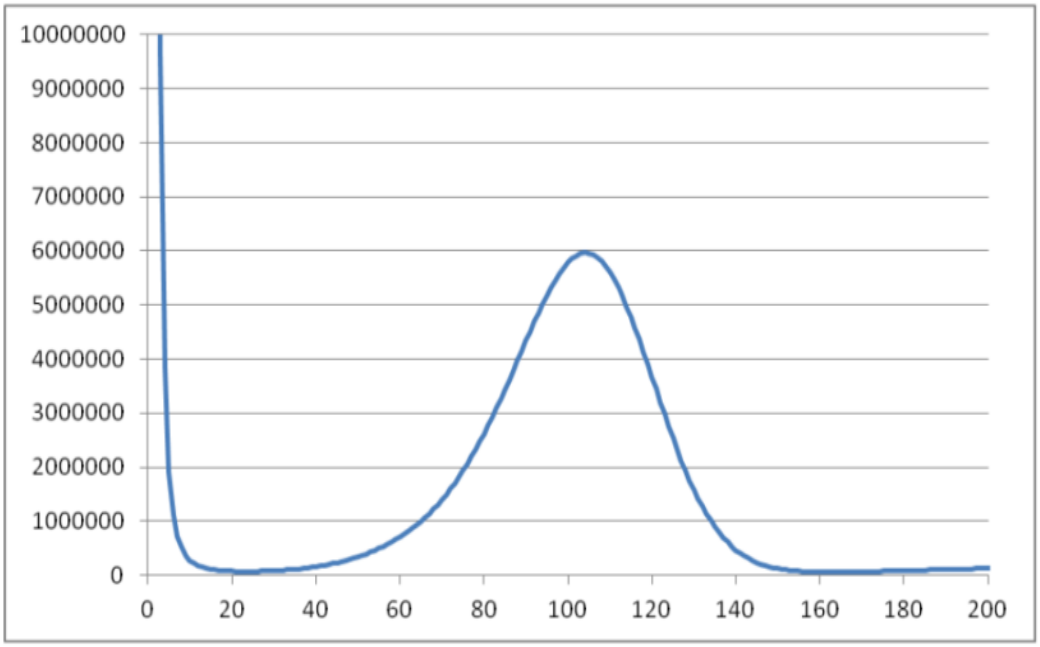
Biotransformation of 3OH-*syn*-pimaradienolide in fission yeast

Log-phase culture of *Schizosaccharomyces pombe* L972 was diluted in a fresh YES medium (0.5% yeast extract, 3% glucose, 225 mg/L adenine, 225 mg/L uracil, 225 mg/L leucine, 225 mg/L histidine, and 225 mg/L lysine) to the optical density at 600 nm of 0.1 and then incubated for 48 h at 30 °C with 5 μM 3OH-*syn*-pimaradienolide. The cell culture was extracted with ethyl acetate, and the extract was evaporated in vacuo. The residue was dissolved in methanol and subjected to LC-MSMS analysis as previously described (reference 8 in the main text).

Functional identification of *CpCYP970A14* and *CpCYP964A1* using both yeast and *N. benthamiana* systems.

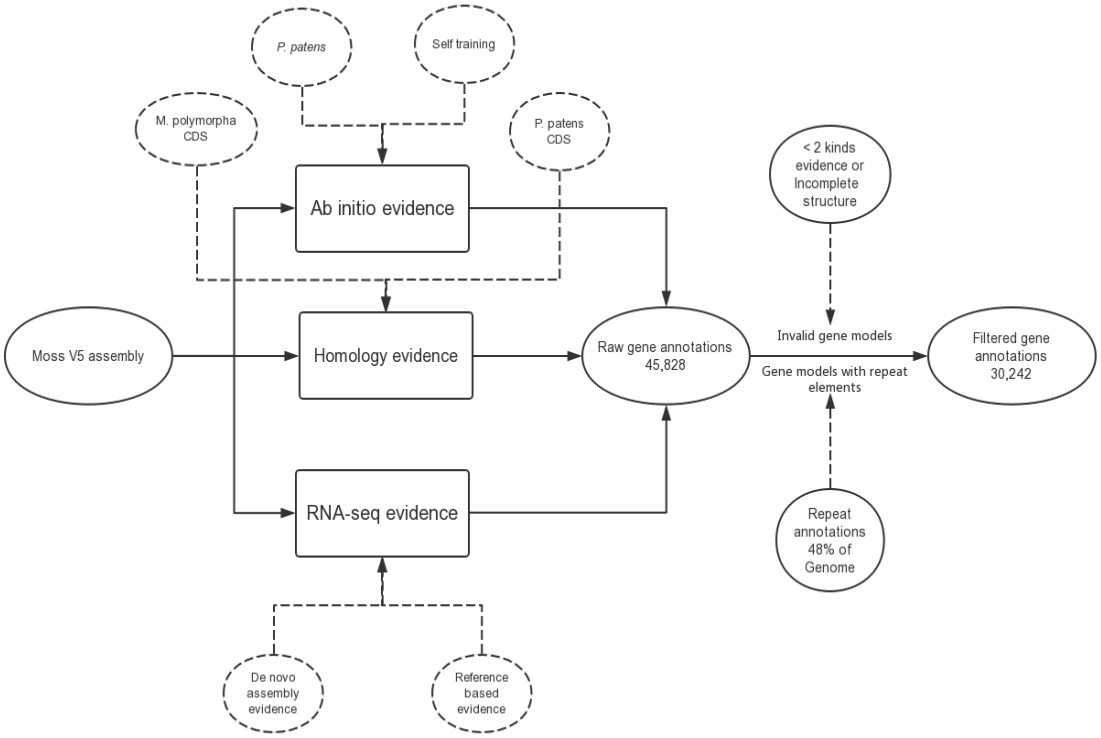
Based on the sequence data, the open reading frames of unigene12783 (*CpCYP964A1*) and unigene16484 (*CpCYP970A14*) were amplified from a *C. plumiforme* cDNA library reported previously (reference 9 in the text) using gene-specific primers which introduced restriction enzyme cleavage sites at the 3′- and 5′-ends: unigene12783-fwd (underlined as KpnI site): 5′-A<u>GGTACC</u>ATGGACCCGTTGCTGGG-3′, unigene12783-rev (underlined as NotI site): 5′-

AGCGGCCGCAATTGACATATGCTCCTCTTCTG-3′ , unigene16484-fwd (underlined as EcoRI site): 5′ -GAATTCATGGAGCTCTCTGTCTGG-3′ , unigene16484-rev (underlined as NotI site): 5′ -GCGGCCGCAAGCTGAAGTCTTGTATACATCC-3′ . The *CpCYP970A14* and *CpCYP964A1* amplicons were ligated into pPICZA (Thermo Fisher) and the recombinant constructs introduced into *Pichia* strain X-33 harboring the *ATR1* gene that enhances CYP reductase activity (21) according to manufacturer's protocols. Transformants were then selected on YPDS plates. The coding region of *CYP99A3* was also amplified by PCR with the gene-specific primers CYP99A3-F (underlined as EcoRI site): 5′ -GGAATTCATGTGGAGATAAACTCAGAAG-3′ , and CYP99A3-R (underlined as KpnI site): 5′ -GGGGTACCACTTTGCATGGAAATCG-3′ , then cloned into the pPICZA vector and to transform *Pichia* strain X-33 in the same manner as described above.
For *N. benthamiana* expression system, *CpCYP970A14* and *CpCYP964A1* cDNAs were amplified by PCR using gene-specific primers (*CpCYP970A14*: Forward primer 5′ -TTCTGCCCAAATTCGATGGAGCTCTCTGTCTGG-3′ , and Reverse primer 5′ -GTGATGGTGATGCCCAGCTGAAGTCTTGTATACATC-3′ . *CpCYP964A1*: Forward primer 5′ -TTCTGCCCAAATTCGATGGACCCGTTGCTGGGC-3′ , and Reverse primer 5′ -GTGATGGTGATGCCCATTGACATATGCTCCTCTTC-3′ ; underlines indicate 15 base pairs vector arm sequences either up and down stream at NruI or SmaI sites. In each gene, reverse primers were designed to truncate termination codon) and amplified fragments were cloned into NruI-SmaI double digested site of pEAQ-HT vectors by In-fusion cloning to express as C-terminal histidine-tagged enzymes. The resulting plasmids pEAQ-HT-CpCYP970A14 and pEAQ-HT-CpCYP964A1 were introduced into *Agrobacterium tumefaciens* LB4404 and used for agroinfiltration.
For functional analysis, the various *Pichia* transformants were cultured for 72 h in 50 mL of a minimal medium using methanol as a carbon source. The culture was started then methanol was added to a final concentration of 0.5% (v/v) every 24 h of the culture. The substrate, *syn*-pimara-7,15-diene (7 μg dissolved in methanol), was added to the culture medium 24 h after the start of culture and converted *in vivo*. After culturing for 72 h, 50 mL of ethyl acetate was added to the culture solution and stirred vigorously. The ethyl acetate layer was recovered by centrifugation (4000 rpm, 25°C, 5 min). This extraction procedure was performed twice. The ethyl acetate layers were combined and concentrated *in vacuo*. The ethyl acetate extracts were derivatized with diazomethane to form methyl ester derivatives and analyzed by GC-MS (GCMSD 5975 Series, Agilent Technologies, Santa Clara, CA, USA). GC and MS analytical conditions were previously described (21, 22).
In the *N. benthamiana* expression system, 3-week-old *N. benthamiana* grown in a plant incubator (23°C, 16 h light/8 h dark photoperiod) was used. After agroinfiltration, the plants were further grown for five days under same condition, following which 4~5 pieces of the leaves were frozen in liquid nitrogen and homogenized with multi-beads shocker (Yasui Kikai, Osaka, Japan). The homogenized samples were extracted with 80% (v/v) methanol containing 5% (v/v) formic acid or hexane as extraction solutions. Purification steps were as described by Miyazaki et al. (23). The methanol eluents were concentrated in vacuo and derivatized with diazomethane to form methyl ester derivatives then analyzed by GC-MS (GCQ1000-K9, JEOL, Tokyo, Japan). The conditions of the GC-MS analysis were previously described (21, 22).

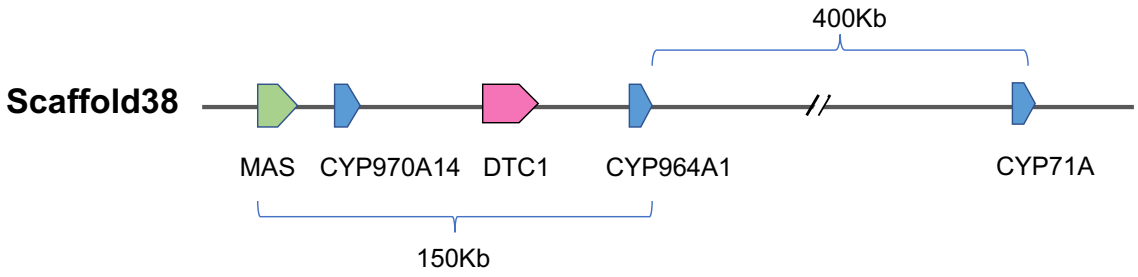


Estimate Genome size	434,417,786

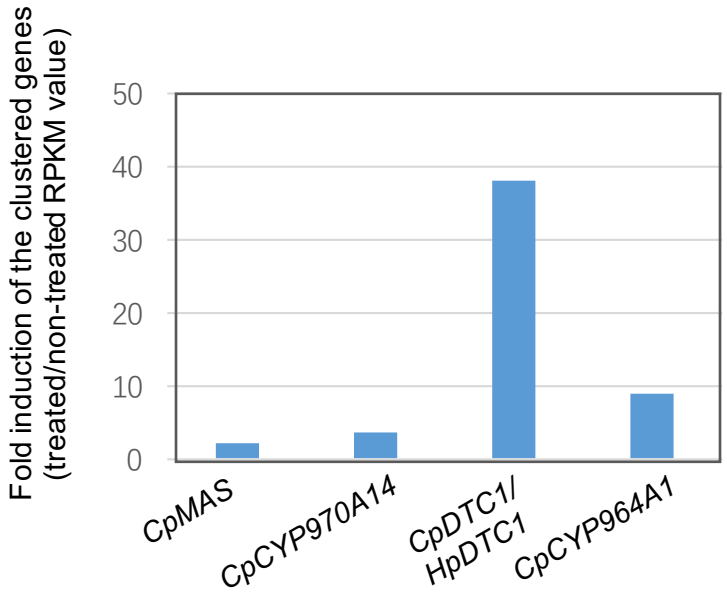
***SI Appendix* Fig. S1.** Genomic survey of *C. plumiforme*. K-mer distribution: X-depth of K-mer; Y-number of K-mer



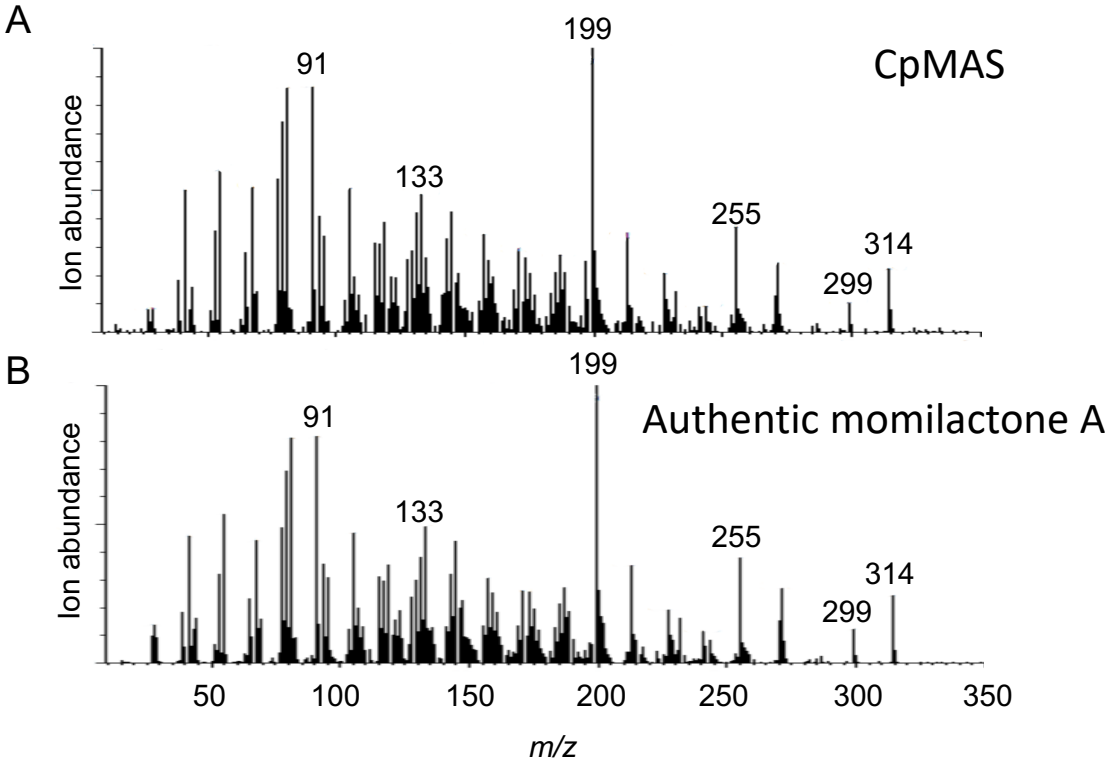
***SI Appendix* Fig. S2.** Gene annotation strategy and results of *C. plumiforme*



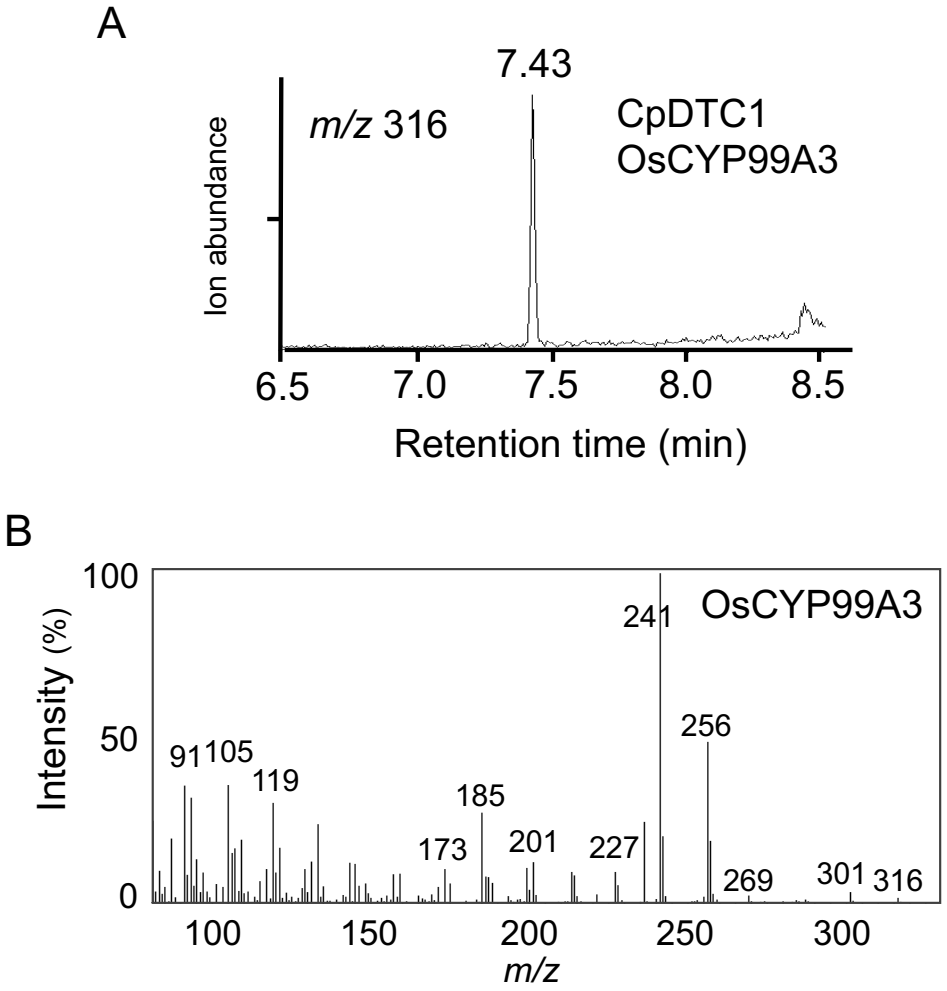
***SI Appendix* Fig. S3.** Illustration of Scaffold38



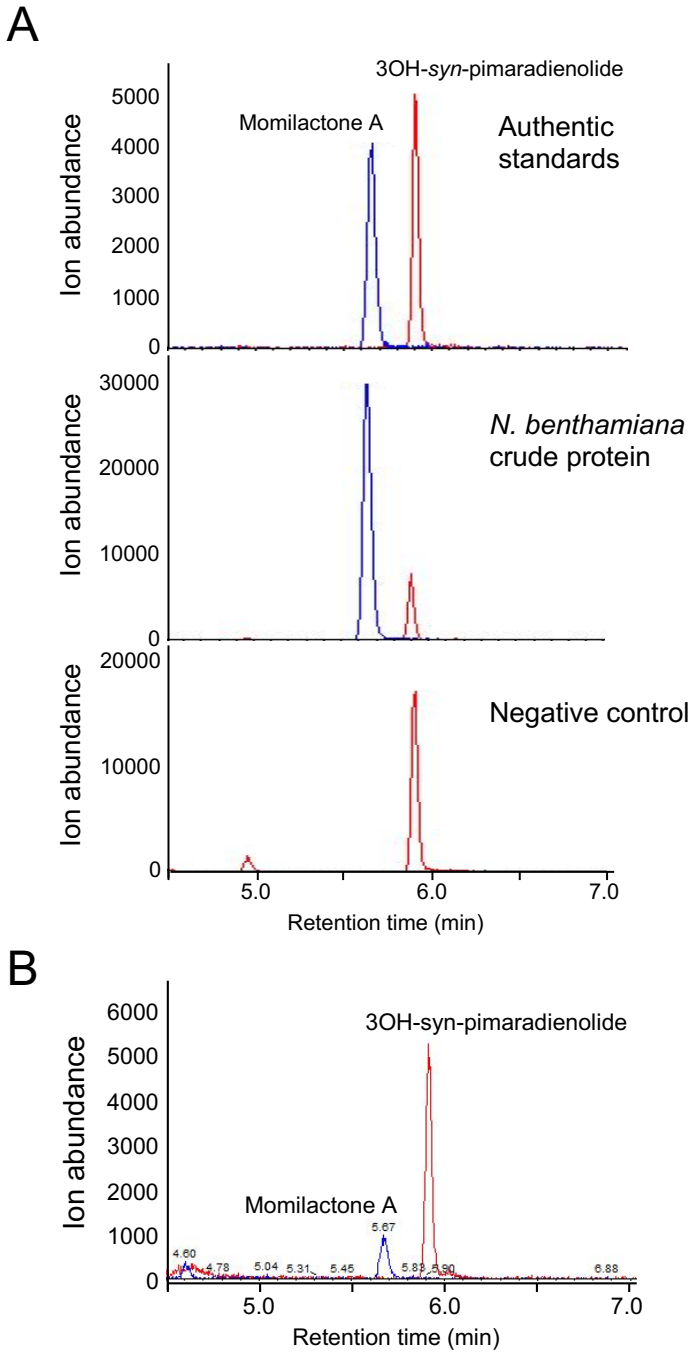
***SI Appendix* Fig. S4.** Inductive expression levels of the clustered genes after $CuCl_2$ treatment shown as calculated RPKM values from RNA-seq data. *CpMAS*: Unigene_5051, LC494432; CpCYP970A14: Unigene_16484, LC494433; CpDTC1/HpDTC1: LC128408; CpCYP964A1: Unigene_12783, LC494434.



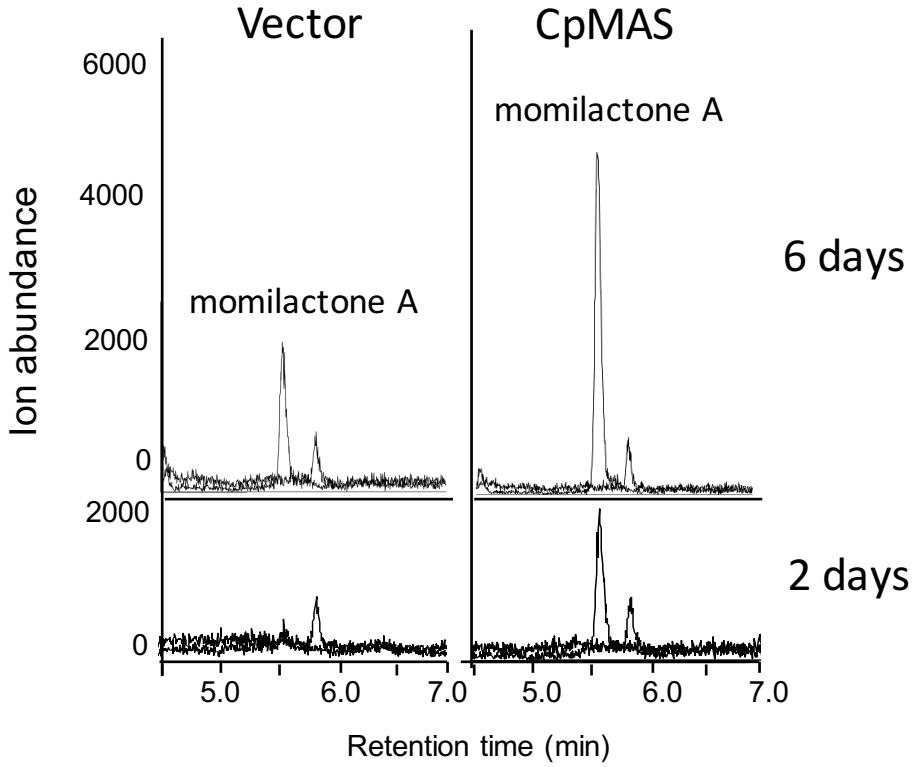
***SI Appendix* Fig. S5.** GC-MS analysis of momilactone A in the enzymatic reaction. Mass spectra of the CpMAS reaction products (A) and authentic momilactone A (B) shown in figure 3B.



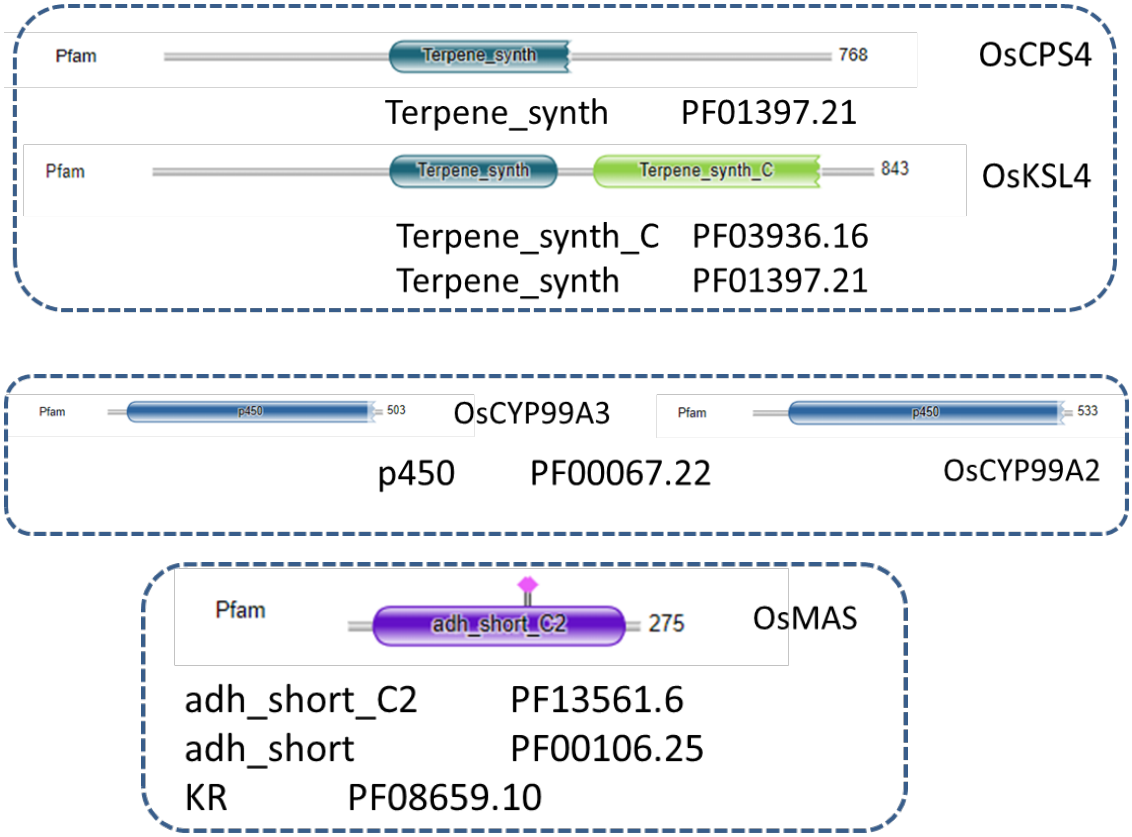
***SI Appendix* Fig. S6.** GC-MS analysis of product catalyzed by rice CYP99A3 in *N. benthamiana.* (A) Total ion chromatogram (scanned on *m/z* 316) of reaction product methyl ester derivative of methanol extract from *N. benthaminana* leaves expressing both CpDTC1 and rice CYP99A3. (B) Mass spectrum of the peak on GC at 7.43 min.



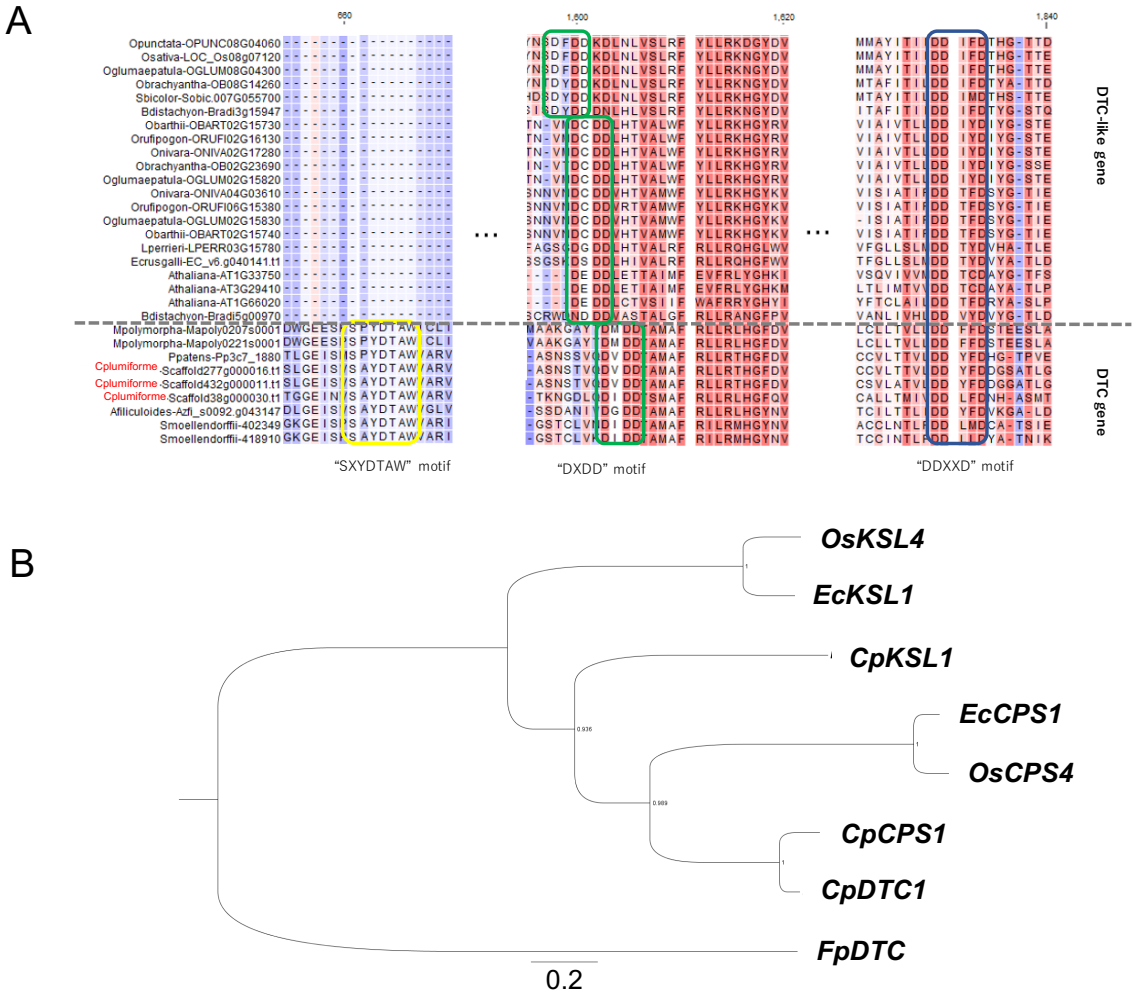
***SI Appendix* Fig. S7.** (A) Momilactone A [6] converted from the substrate 3β-hydroxy-9βH-pimara-7,15-dien-19,6β-olide [5] by crude protein from *N. benthamiana* leaves is analyzed by LC-MSMS. Upper panel, authentic standards; middle panel, product catalyzed by *N. benthamiana* crude protein (0.1 mg) at 33°C for 24h; lower panel, negative control without enzyme. (B) *In planta* conversion of 3OH-*syn*-pimaradienolide to momilactone A in *N. benthamiana* feeding assay. Approximately 1 µg of 3OH-*syn*-pimaradienolide was injected into one leaf of *N. benthamiana.*



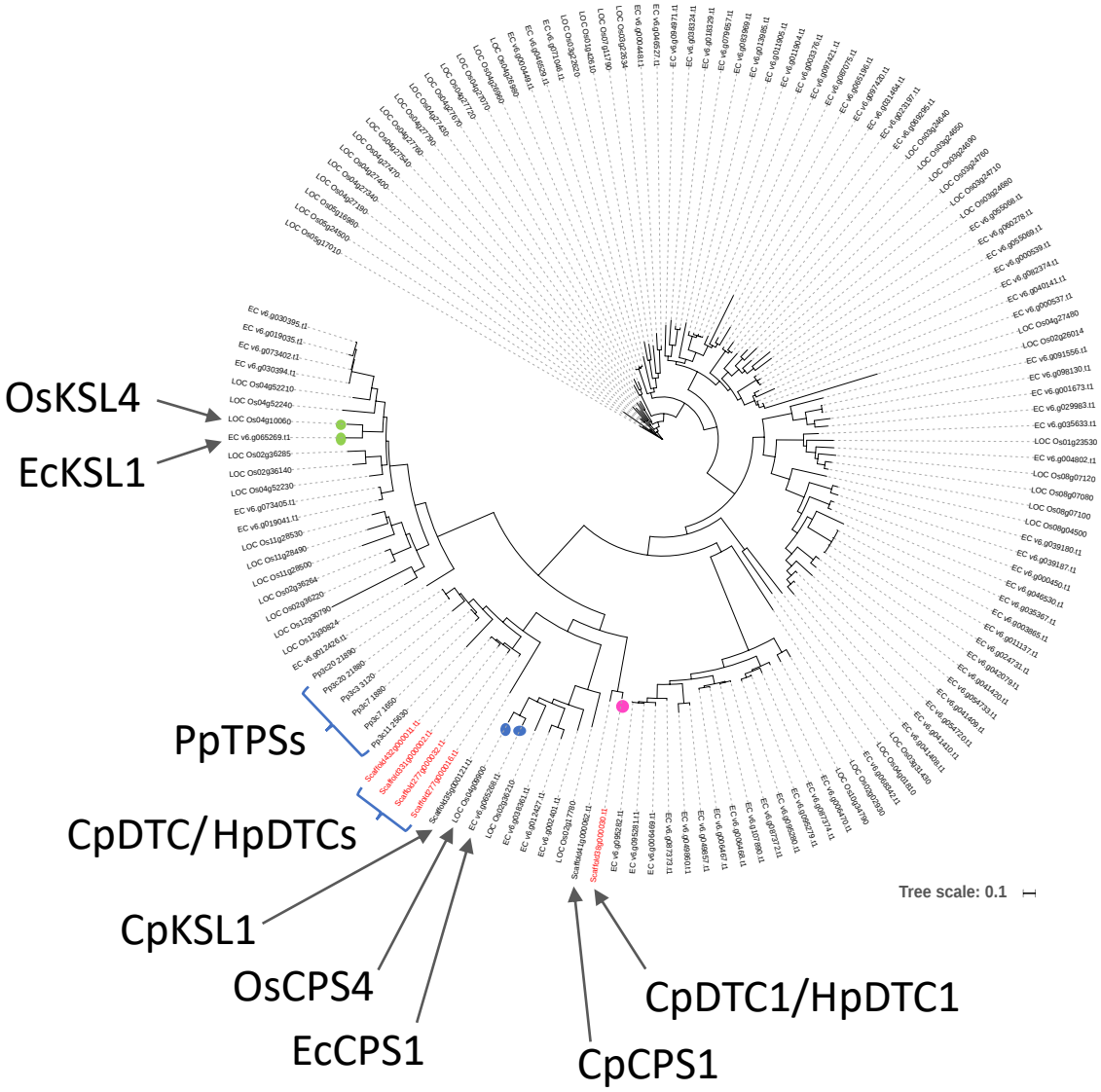
***SI Appendix* Fig. S8.** Enhanced production of momilactone A in the *N. benthamiana* leaves expressing CpMAS with feeding of 3OH-pimaradienolide. Reaction products from the leaves harvested 2 days and 6 days after the infiltration are analyzed by LC-MSMS. Approximately 1 µg of 3OH-*syn*-pimaradienolide was injected into one leaf of *N. benthamiana.*



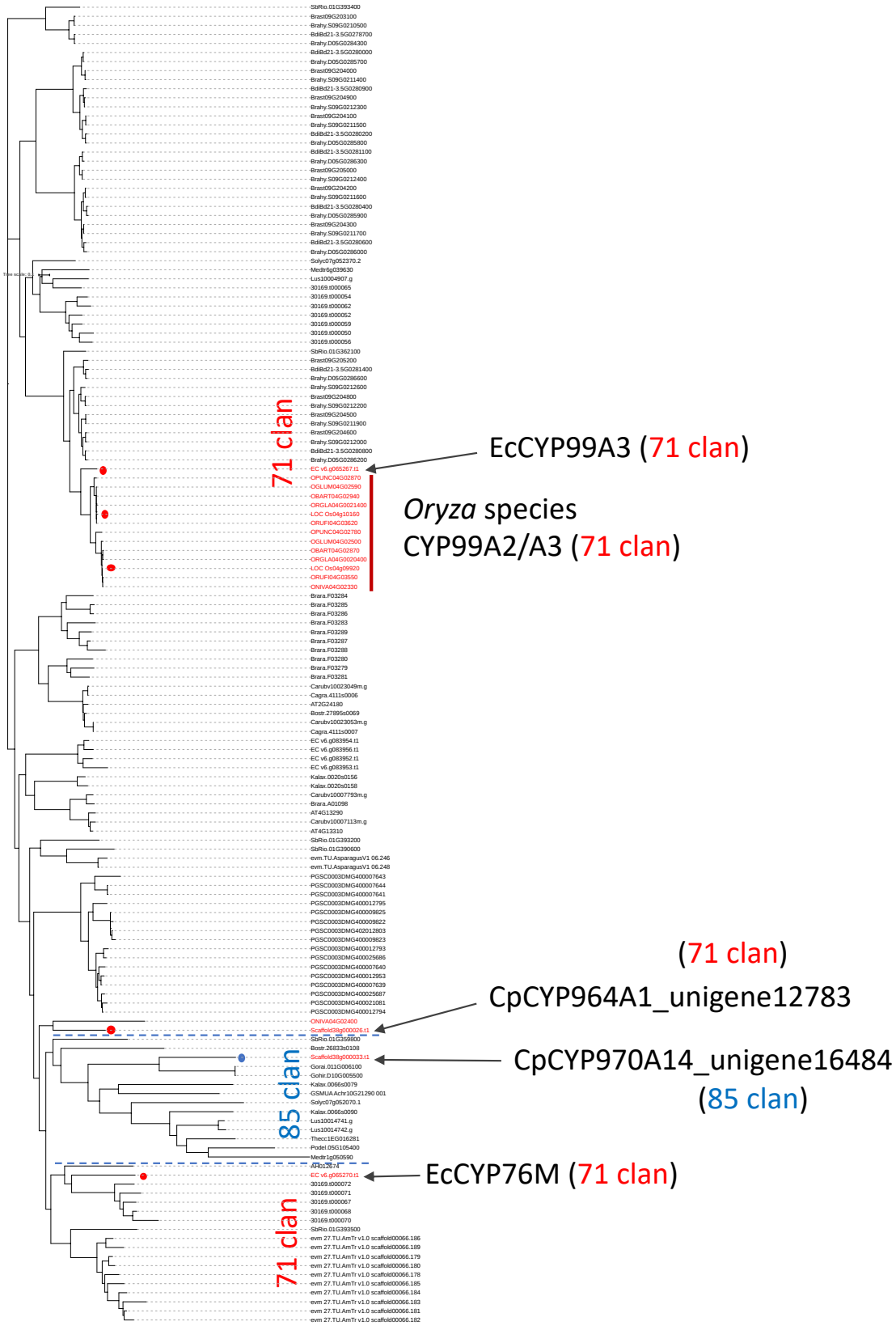
***SI Appendix* Fig. S9.** Domains of terpene synthase (TPS), cytochrome P450 monooxygenase (P450) and momilactone A synthase (MAS)



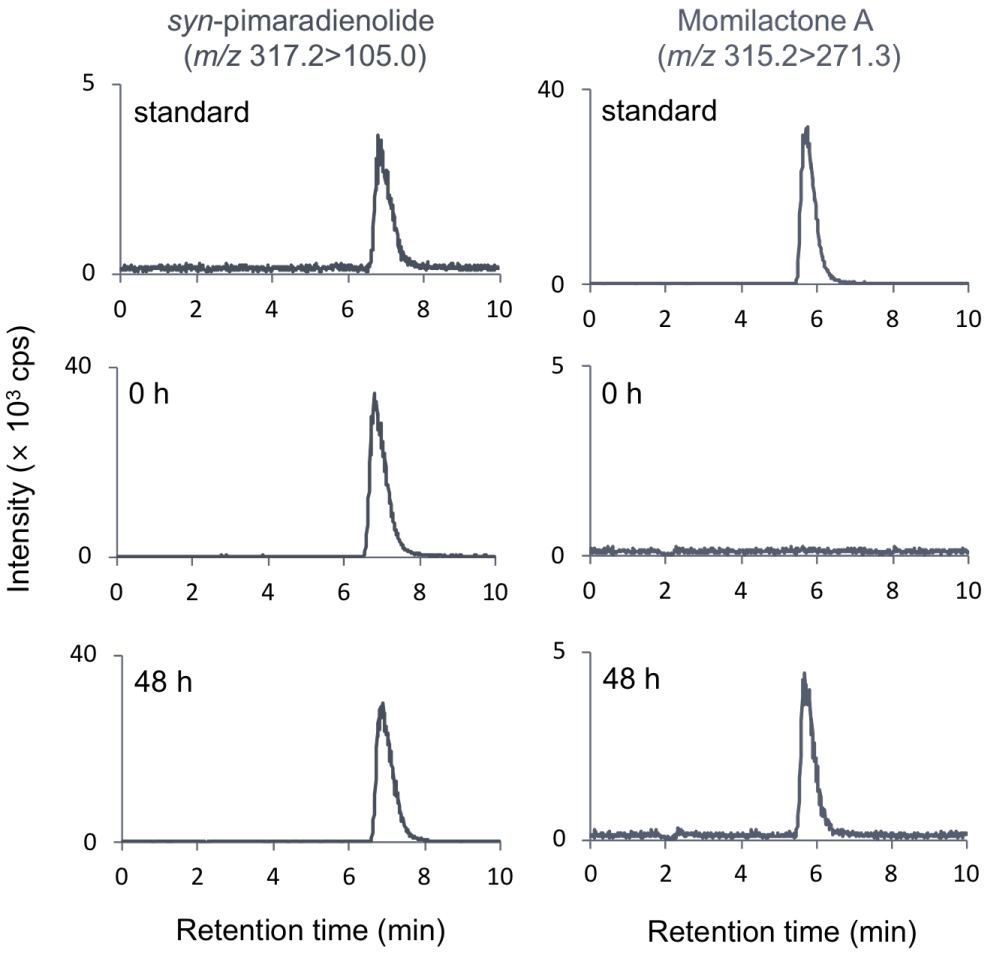
***SI Appendix* Fig. S10.** Evolution of momilactone biosynthesis-related terpene synthase (TPS) genes in plants. (A) Sequence alignment of two types of TPS genes (DTC and DTC-like) with the three functional motifs ("DXDD", "DDXXD" and "SXYDTAW") in plants. DTC-like TPS genes have "DXDD" motif but its "DXDD" motif do not locate at the typical position. "SXYDTAW" motif is not found in DTC-like TPS. (B) Phylogenetic tree of momilactone gene cluster-related TPS genes in the momilactone gene clusters in three plants species (Os, Ec and Cp). The ortholog (accession number: CAP74389.1) from the fungus *Fusarium proliferatum* was used as outgroup. In addition to the terpene gene *CpDTC1/HpDTC1* in the cluster, two other terpene genes (*CpCPS* and *CpKSL*) in the *Calohypnum* genome were also included.



***SI Appendix* Fig. S11.** Phylogenetic tree of terpene synthase genes from two mosses and two grasses. Four types of genes with different three functional motifs shown in figure S8 were included (e.g. DTC, DTC-like, CPS, KS have three, two and one motifs, respectively). CpDTC1/HpDTC1, OsCPS4, EcCPS4, OsKSL4, EcCPS4, CpDTC/HpDTC homologs, and PpTPS homologs are indicated.



***SI Appendix* Fig. S12.** Phylogenetic tree of CYP genes in 43 candidate gene cluster listed in supplementary table S7.



***SI Appendix* Fig. S13.** Biotransformation of momilactone A from 3OH-syn-pimaradienolide in fission yeast. Yeast fed with 3OH-syn-pimaradienolide was further grown for 48 h. Metabolized product was extracted by ethyl acetate and analyzed by LC-MSMS.

Table S1a Summary of the sequenced genomic data of *C. plumiforme*

Type	Libs	Read length	Raw bases	Coverage
Pair-end	360bp	90	29.4G	67.7
	350bp	90	33.4G	76.9
Total			~62.8G	~145x
Mate-pair	10k	140	14G	32.3
	20k	140	11.4G	26.3
			~25.4G	~59x
Total		75-150	~88.2G	~203x
RSII	10k	5.26k	~1.54G	
Total			~1.54G	~3.5x

Table S1b RNA-seq data information

Type	Treatments	Reads length	Total base
RNA-seq	$CuCl_2$ 0h	90	~2.4G
	$CuCl_2$ 8h	90	~2.3G
Total			~4.7G

***SI Appendix* Table S1. (a)** Summary of the sequenced genomic data of *C. plumiforme.* (b) RNA-seq data information

Table S2 Summary of genome assembly of *C. plumiforme*

	V1	V2	V3	V4	V5
Scaffold length(Mb)	332.3（76.1%）	336 (77.3%)	425.9(98.0%)	332.1（76.4%）	335.0(77.0%)
Estimated genome size (Mb)	434.4	434.4	434.4	434.4	434.4
Scaffold N50 (bp)	94,597	193,269	6,450,362	784,076	790,020
Contig length (Mb)	332.3	335.3	402.6	326.8	333.3
Gap size (Mb)	0.1	0.7	23.3	5.3	1.7
Contig N50 (bp)	92,913	173,015	24,050	110,463	224333

V1: Original assembly from Macrogen Japan
V2: Original assembly with gap filling of Pacbio data
V3: Pipeline1, NGS assembly with gap filling of Pacbio data
V4: Pipeline2, Hybrid assembly of NGS and Pacbio data
V5: Improved version based on V4

***SI Appendix* Table S2.** Summary of genome assembly of *C. plumiforme*

Table S3 V5 assembly results of all stages

Assembly stages	Genome size	Scaffold N50	Contig N50	Contig No.	Scaffold No.
1_MaSuRCA	332.06Mb	635.68Kb	110.46Kb	7,205	1,202
2_SSPACE	332.15Mb	784.08Kb	110.46Kb	7,201	1,043
3_Pbjelly	332.94Mb	784.08Kb	113.14Kb	7,025	1,042
4_GapFiller	334.36Mb	788.70Kb	162.13Kb	4,742	1,042
5_SSPACE	334.36Mb	788.70Kb	162.13Kb	4,742	1,040
6_Pbjelly	334.50Mb	788.70Kb	163.11Kb	4,712	1,040
7_GapFiller	335.23Mb	790.01Kb	197.03Kb	3,983	1,040
8_pilon	335.01Mb	790.02Kb	224.33Kb	3,474	1,040
9_merge	335.01Mb	790.02Kb	224.32Kb	3,474	1,040

***SI Appendix* Table S3.** V5 assembly results of all stages

Table S4 Summary of repeat annotation

Type of repeat elements	Length	Number	Percentage (%)
SINES	13,032	53	0.00
LINEs	3,215,244	7,253	0.96
LTR elements	54,985,261	137,958	16.41
DNA elements	13,342,737	25,481	3.98
Unclassified	90,821,706	280,791	27.11
total	162,377,980	451,536	48.46

***SI Appendix* Table S4.** Summary of repeat annotation

Table S5a Primer information

Primers	Primer sequences (5'-3')
Forward primer	TGCAAGGATGATCAATCACTATAAG
Reverse primer	CAACTGACCAATAGTGCTATGGAA

PCR_products and sequecing results

>Contig1

tgtaaaacatgccttaatgtcaaattgaacacatttttgtgtctaggttgatgtcctctaaaatttgtcatgattaacaagcaaaaaatgtccaaa
atttcaagtaacatgcttaattttggacttgtttttttttaacattacaaacaacatggtctttcaagcttcccaaaatttaccacataggaaattttt
catccaacttgatgatgtggcatctgacgtggtggtgggaaaaatgatgtggcaaccaaaaatgataacgtggcatgccatgtcaacaca
tttttgttgaaaatttttgatatttttttcaaatgttaaattagacattaattaaaacataccctgtccacaacacaaaattaattttaaaaatagtta
cttatttcaaaattaaattataaatttaaaaaaaaaacttcaaaaaaatcctaaaatcaacaaaaagtaaagaaaacttaccatctaccatc
aaaagaatttttgatggcaaaatgaaggaaatacttcattaaaagataatgaaaatgaaagaaatgagcttctatatatacccaatctcaac
ccatggtgagacaaaagatatgggataactagttttggttacttgatacttgtttatgttgacattatttttgatgtcacatgttctatctaatgtttgca
tgttctttctgacatcttgatgttctttttgacattctatactaaacacaattcattttgacaaatcccatatggattgtctcccttcctatcaacccatt
tgtataaagcctaactcttttcattcattgtgcacatcttatttcatccatccatttccttttttttacaaagtaagccacatggttgccctttcctaaa
attttcacaaatgggcagcattttactataaataatattaaattttattaattttcacaaaattggaaataaatctacttgttgacgtggtagcctatg
ttttttttggatattatcaaaaacacaaacttgtttttttgaacatcattaaaaacaaatgtgtccatagtgacatctacactaattttttttacgatttttg
caatgtcatatcaaattcacacttttgcaaccgtcaattgttgaacatagtgacgtttttagaaaacatgacacatttttttaataacataaccaa
aaacacagttttcttatagtgaatgcattactgagaattatgtgaaa

Table S5b Primers for RT-qPCR

Primers	Primer sequences (5'-3')
CpDTC-qPCR-F	TGCTGCTCAGCATGTATCGT
CpDTC-qPCR-R	GGACTCTGGAACGCAAGACT
CpMAS-qPCR-F	ACCGCAGACGAAGCAGAATA
CpMAS-qPCR-R	CCTTTACATTCAGCCGAAGC
CpCYP970-qPCR-F	CGGGAGATACTGAAGGTCCA
CpCYP970-qPCR-R	CTCCGTAAGGCGAGAAGATG
CpCYP964-qPCR-F	ATGTGCTAGATGCGCTGATG
CpCYP964-qPCR-R	AGAGGTCTCATGGCCTGCTA
CpACT3-qPCR-F	CGAGCAGCATGAAGATCAAG
CpACT3-qPCR-R	GTACTCGCTCTTCGCAATCC

***SI Appendix* Table S5.** (a) Primer information and gap sequence of two scaffolds. (b) Primer information for qPCR.

late momilactone gene clusters in plant kingdom					
Species	Version	CHR	START	END	GENES
Oryza sativa	Osativa_323_v7.0	Chr4	5300000	5500000	LOC_Os04g09900(TPS-c|type1),LOC_Os04g09920(71-99A3),LOC_Os04g10000(mas),LOC_Os04g10010(mas),LOC_Os04g10060(TPS-e/f|type2),LOC_Os04g10160(71-99A3)
Oryza glumaepatula	.Oryza_glumaepatula_v1.5	4	3500000	3800000	OGLUM04G02490(TPS-c|type1),OGLUM04G02500(71-99A3),OGLUM04G02520(mas),OGLUM04G02540(mas),OGLUM04G02550(mas),OGLUM04G02560(mas),OGLUM04G02570(TPS-e/f|type2),OGLUM04G02590(71-99A3)
Oryza rufipogon	Oryza_rufipogon.OR_W1943	4	4200000	4400000	ORUFI04G03540(TPS-c|type1),ORUFI04G03550(71-99A3),ORUFI04G03580(mas),ORUFI04G03590(mas),ORUFI04G03600(TPS-e/f|type2),ORUFI04G03620(71-99A3)
Oryza barthii	O.barthii_v1	4	2800000	3000000	OBART04G02860(TPS-c|type1),OBART04G02870(71-99A3),OBART04G02910(mas),OBART04G02920(TPS-e/f|type2),OBART04G02940(71-99A3)
Oryza punctata	Oryza_punctata_v1.2	4	4500000	4700000	OPUNC04G02770(TPS-c|type1),OPUNC04G02780(71-99A3),OPUNC04G02800(mas),OPUNC04G02820(mas),OPUNC04G02860(TPS-e/f|type2),OPUNC04G02870(71-99A3)
Oryza nivara	Oryza_nivara.Oryza_nivara_v1.0	4	2900000	3000000	ONIVA04G02320(TPS-c|type1),ONIVA04G02330(71-99A3),ONIVA04G02380(mas),ONIVA04G02400(71-79A42)
Echinochloa crus-galli	Ecrus-galli_v6.prot	scaffold303	1100000	1200000	EC_v6.g065267.t1(71-99A3),EC_v6.g065268.t1(TPS-c|type1),EC_v6.g065269.t1(TPS-e/f|type2),EC_v6.g065270.t1(71-76M6),EC_v6.g065271.t1(mas),EC_v6.g065273.t1(mas)
Calohypnum plumiforme	Cplumiforme_V5	Scaffold38	100000	300000	Scaffold38g000024.t1(mas),Scaffold38g000026.t1(71-736A104),Scaffold38g000030.t1(TPS-a|type0),Scaffold38g000033.t1(85-716A10)
Oryza glaberrima	Oryza_glaberrima.AGI1.1	4	2600000	2900000	ORGLA04G0020300(TPS-c|type1),ORGLA04G0020400(71-99A3),ORGLA04G0020700(mas),ORGLA04G0020800(mas),ORGLA04G0021100(TPS-e/f|type2),ORGLA04G0021400(71-99A3)
Amaranthus hypochondriacus	Ahypochondriacus_459_v2.1	Scaffold_8	2400000	2600000	AH012674(71-736A104),AH012690(mas),AH012694(TPS-c|type1)
Echinochloa crus-galli	Ecrus-galli_v6.prot	scaffold502	0	300000	EC_v6.g083952.t1(71-71B78),EC_v6.g083953.t1(71-71AN5),EC_v6.g083954.t1(71-71D84P),EC_v6.g083956.t1(71-71B78),EC_v6.g083969.t1(TPS-a|other),EC_v6.g083992.t1(mas),EC_v6.g083993.t1(mas)
Solanum lycopersicum	Slycopersicum_390_ITAG2.4	SL2.50ch07	60600000	60900000	Solyc07g052070.1(72-72A85),Solyc07g052120.2(TPS-a*|other),Solyc07g052140.2(TPS-a*|other),Solyc07g052270.2(mas),Solyc07g052370.2(71-71D204)
Populus deltoides	PdeltoidesWV94_445_v2.1	Chr05	7500000	7800000	Podel.05G105400(85-96D2),Podel.05G106400(TPS-a|type2),Podel.05G106500(TPS-a|type2),Podel.05G106600(TPS-a|type2),Podel.05G106700(TPS-a|type2),Podel.05G106900(mas),Podel.05G107000(TPS-a|type2)
Boechera stricta	Bstricta_278_v1.2	Scaffold26833	300000	600000	Bostr.26833s0052(mas),Bostr.26833s0091(TPS-a|type2),Bostr.26833s0108(71-89A94)
Boechera stricta	Bstricta_278_v1.2	Scaffold27895	500000	800000	Bostr.27895s0060(TPS-b|other),Bostr.27895s0061(TPS-b|other),Bostr.27895s0062(TPS-b|type2),Bostr.27895s0063(TPS-b|other),Bostr.27895s0064(TPS-b|type2),Bostr.27895s0067(mas),Bostr.27895s0069(71-71B6)
Brachypodium stacei	Bstacei_316_v1.1	Chr09	16800000	17200000	Brast09G203100(71-71C77),Brast09G203800(TPS-e/f|type2),Brast09G203900(TPS-e/f|type2),Brast09G204000(71-71F9),Brast09G204100(71-71F9),Brast09G204200(71-71AF5),Brast09G204300(71-71AF5),Brast09G204400(TPS-e/f|type2),Brast09G204500(71-99A20),Brast09G204600(71-99A13),Brast09G204800(71-99A13),Brast09G204900(71-71F9),Brast09G205000(71-71AF5),Brast09G205100(TPS-e/f|type2),Brast09G205200(71-99A6),Brast09G207100(mas)
Capsella rubella	Crubella_183_v1.0	scaffold_4	1500000	1600000	Carubv10023053m.g(71-71B6),Carubv10023049m.g(71-71B6),Carubv10023750m.g(mas),Carubv10022869m.g(TPS-b|type2),Carubv10022874m.g(TPS-b|type2)
Capsella rubella	Crubella_183_v1.0	scaffold_7	7700000	7900000	Carubv10005273m.g(mas),Carubv10007793m.g(71-71A20),Carubv10007667m.g(TPS-a*|type2),Carubv10007162m.g(TPS-a|type2)
Capsella rubella	Crubella_183_v1.0	scaffold_7	11300000	11500000	Carubv10007113m.g(71-71A20),Carubv10006722m.g(TPS-a|type2),Carubv10004643m.g(mas),Carubv10005531m.g(mas)
Kalanchoe laxiflora	Klaxiflora_309_v1.1	scaffold_20	1200000	1500000	Kalax.0020s0156(71-71D82),Kalax.0020s0157(TTC|type2),Kalax.0020s0158(71-71D82),Kalax.0020s0177(mas)
Kalanchoe laxiflora	Klaxiflora_309_v1.1	scaffold_66	700000	1000000	Kalax.0066s0074(mas),Kalax.0066s0075(mas),Kalax.0066s0076(mas),Kalax.0066s0077(mas),Kalax.0066s0079(72-72A85),Kalax.0066s0090(85-96D2),Kalax.0066s0098(TPS-a|type2)
Linum usitatissimum	Lusitatissimum_200_v1.0	scaffold584	100000	500000	Lus10014673.g(TPS-e/f|type2),Lus10014688.g(mas),Lus10014689.g(mas),Lus10014690.g(mas),Lus10014693.g(mas),Lus10014694.g(mas),Lus10014724.g(TPS-a|type2),Lus10014741.g(85-96D2),Lus10014742.g(86-96B17)
Linum usitatissimum	Lusitatissimum_200_v1.0	scaffold858	0	200000	Lus10004906.g(TPS-e/f|other),Lus10004907.g(71-71D343P),Lus10004908.g(mas),Lus10004925.g(mas)
Gossypium raimondii	Graimondii_221_v2.1	Chr11	300000	500000	Gorai.011G003700(mas),Gorai.011G005500(TPS-a|type2),Gorai.011G005600(TPS-a|type2),Gorai.011G005700(TPS-a|type2),Gorai.011G005800(TPS-a|other),Gorai.011G005900(TPS-a|type2),Gorai.011G006100(85-716A20)
Arabidopsis thaliana	Athaliana_167_TAIR10	Chr2	10200000	10300000	AT2G24180(71-71B6),AT2G24190(mas),AT2G24210(TPS-b|type0)
Arabidopsis thaliana	Athaliana_167_TAIR10	Chr4	7600000	7800000	AT4G13180(mas),AT4G13250(mas),AT4G13280(TPS-a|type2),AT4G13290(71-71A19),AT4G13300(TPS-a|type2),AT4G13310(71-71A20)
Ricinus communis	Rcommunis_119_v0.1	30169	200000	500000	30169.t000050(71-726A17),30169.t000051(mas),30169.t000052(71-726A17),30169.t000053(TPS-a|type2),30169.t000054(71-726A15),30169.t000056(71-726A17),30169.t000059(71-726A17),30169.t000060(TPS-a|type2),30169.t000062(71-726A18),30169.t000063(TPS-a|type2),30169.t000064(mas),30169.t000065(71-71D343P),30169.t000066(TPS-g|type2),30169.t000067(71-80T1),30169.t000068(71-80T1),30169.t000070(71-76C6),30169.t000071(71-80T1),30169.t000072(71-80T1),30169.t000079(mas),30169.t000091(TPS-a|type2)
Gossypium hirsutum	Ghirsutum_458_v1.1	D10	300000	500000	Gohir.D10G003300(mas),Gohir.D10G005100(TPS-a|type2),Gohir.D10G005200(TPS-a|other),Gohir.D10G005300(TPS-a|other),Gohir.D10G005500(85-716A20)
Amborella trichopoda	Atrichopoda_291_v1.0	AmTr_v1.0_scaffold00066	2400000	2700000	evm_27.TU.AmTr_v1.0_scaffold00066.167(mas),evm_27.TU.AmTr_v1.0_scaffold00066.178(71-81S16),evm_27.TU.AmTr_v1.0_scaffold00066.179(71-81Q59),evm_27.TU.AmTr_v1.0_scaffold00066.180(71-81Q59),evm_27.TU.AmTr_v1.0_scaffold00066.181(71-81S16),evm_27.TU.AmTr_v1.0_scaffold00066.182(71-81Q59),evm_27.TU.AmTr_v1.0_scaffold00066.183(71-81S16),evm_27.TU.AmTr_v1.0_scaffold00066.184(71-81Q59),evm_27.TU.AmTr_v1.0_scaffold00066.185(71-81Q59),evm_27.TU.AmTr_v1.0_scaffold00066.186(71-81S16),evm_27.TU.AmTr_v1.0_scaffold00066.189(71-92A63),evm_27.TU.AmTr_v1.0_scaffold00066.195(TPS-e/f|type2)
Brachypodium distachyon	BdistachyonBd21_3_460_v1.0	Bd5	24300000	24600000	BdiBd21-3.5G0278700(71-71C77),BdiBd21-3.5G0279800(TPS-e/f|type2),BdiBd21-3.5G0279900(TPS-e/f|type2),BdiBd21-3.5G0280000(71-71F9),BdiBd21-3.5G0280200(71-71F9),BdiBd21-3.5G0280400(71-71AF5),BdiBd21-3.5G0280600(71-71AF5),BdiBd21-3.5G0280700(TPS-e/f|type2),BdiBd21-3.5G0280800(71-99A13),BdiBd21-3.5G0280900(71-71F9),BdiBd21-3.5G0281100(71-71AF5),BdiBd21-3.5G0281400(71-99A15P),BdiBd21-3.5G0281900(TPS-e/f|type2),BdiBd21-3.5G0284500(mas)
Asparagus officinalis	Aofficinalis_498_V1.1	AsparagusV1_06	4000000	4200000	evm.TU.AsparagusV1_06.243(TTC|type2),evm.TU.AsparagusV1_06.244(TPS-a|type2),evm.TU.AsparagusV1_06.245(mas),evm.TU.AsparagusV1_06.246(71-71Z9),evm.TU.AsparagusV1_06.247(TPS-a|type2),evm.TU.AsparagusV1_06.248(71-705A2)
Theobroma cacao	Tcacao_233_v1.1	scaffold_3	32100000	32300000	Thecc1EG016274(mas),Thecc1EG016281(85-96D2),Thecc1EG016298(TTC|type2),Thecc1EG016299(TTC|type2)
Sorghum bicolor	SbicolorRio_468_v2.0	chromosome_1	61900000	62200000	SbRio.01G359800(71-78A150),SbRio.01G361300(mas),SbRio.01G362100(71-99A6),SbRio.01G362200(TPS-a|type2)
Sorghum bicolor	SbicolorRio_468_v2.0	chromosome_1	65400000	65700000	SbRio.01G390600(71-705A34P),SbRio.01G393100(TPS-a|other),SbRio.01G393200(71-92A63),SbRio.01G393400(71-71C77),SbRio.01G393500(71-81L6),SbRio.01G393900(mas)
Solanum tuberosum	Stuberosum_448_v4.03	ST4.03ch04	52800000	53600000	PGSC0003DMG400012797(mas),PGSC0003DMG400012790(TPS-a*|type2),PGSC0003DMG400012793(71-736A104),PGSC0003DMG400012794(71-736A104),PGSC0003DMG400012795(71-736A104),PGSC0003DMG402012803(71-736A104),PGSC0003DMG400009823(71-736A104),PGSC0003DMG400009825(71-736A104),PGSC0003DMG400009822(71-736A104),PGSC0003DMG400021081(71-736A104),PGSC0003DMG400025687(71-736A104),PGSC0003DMG400025686(71-736A104),PGSC0003DMG400007640(71-736A104),PGSC0003DMG400007644(71-736A104),PGSC0003DMG400007643(71-736A104),PGSC0003DMG400007639(71-736A104),PGSC0003DMG400007641(71-736A104),PGSC0003DMG400012953(71-736A104)
Brachypodium hybridum	Bhybridum_463_v1.1	BhD5	24100000	24500000	Brahy.D05G0284300(71-71C77),Brahy.D05G0285500(TPS-e/f|type2),Brahy.D05G0285600(TPS-e/f|type2),Brahy.D05G0285700(71-71F9),Brahy.D05G0285800(71-71F9),Brahy.D05G0285900(71-71AF5),Brahy.D05G0286000(71-71AF5),Brahy.D05G0286100(TPS-e/f|type2),Brahy.D05G0286200(71-99A13),Brahy.D05G0286300(71-71AF5),Brahy.D05G0286500(TPS-e/f|type2),Brahy.D05G0286600(71-99A15P),Brahy.D05G0286800(TPS-e/f|type2),Brahy.D05G0289500(mas)
Brachypodium hybridum	Bhybridum_463_v1.1	BhS9	14800000	15100000	Brahy.S09G0210500(71-71C77),Brahy.S09G0211200(TPS-e/f|type2),Brahy.S09G0211300(TPS-e/f|type2),Brahy.S09G0211400(71-71F9),Brahy.S09G0211500(71-71F9),Brahy.S09G0211600(71-71AF5),Brahy.S09G0211700(71-71AF5),Brahy.S09G0211800(TPS-e/f|type2),Brahy.S09G0211900(71-99A20),Brahy.S09G0212000(71-99A20),Brahy.S09G0212200(71-99A13),Brahy.S09G0212300(71-71F9),Brahy.S09G0212400(71-71AF5),Brahy.S09G0212600(71-99A13),Brahy.S09G0212700(TPS-e/f|type2),Brahy.S09G0214800(mas)
Musa acuminata	Macuminata_304_v1	chr10	26700000	27000000	GSMUA_Achr10G20900_001(TPS-e/f|other),GSMUA_Achr10G20910_001(TPS-e/f|type2),GSMUA_Achr10G20930_001(TPS-e/f|other),GSMUA_Achr10G21050_001(mas),GSMUA_Achr10G21290_001(71-71D345)
Capsella grandiflora	Cgrandiflora_266_v1.1	Scaffold4111	0	41871	Cagra.4111s0001(TPS-b|type2),Cagra.4111s0005(mas),Cagra.4111s0006(71-71B6),Cagra.4111s0007(71-71B6)
Medicago truncatula	Mtruncatula_285_Mt4.0v1	chr1	19700000	20000000	Medtr1g050590(85-96D2),Medtr1g050430(TPS-a|other),Medtr1g051165(mas)
Medicago truncatula	Mtruncatula_285_Mt4.0v1	chr6	14200000	14500000	Medtr6g039440(TPS-a|type2),Medtr6g039630(71-71D79),Medtr6g041980(mas)
Brassica rapa	BrapaFPsc_277_v1.3	A01	5800000	6000000	Brara.A01098(71-71A20),Brara.A01099(TPS-a*|type2),Brara.A01119(mas)
Brassica rapa	BrapaFPsc_277_v1.3	A06	25500000	26000000	Brara.F03253(mas),Brara.F03279(71-705A36P),Brara.F03280(71-705A36P),Brara.F03281(71-705A36P),Brara.F03283(71-705A36P),Brara.F03284(71-705A36P),Brara.F03285(71-705A36P),Brara.F03286(71-71B6),Brara.F03287(71-71F11),Brara.F03288(71-71B6),Brara.F03289(71-705A36P),Brara.F03314(TPS-b|type2),Brara.F03315(TPS-b|type2),Brara.F03317(TPS-b|type2)

SI Appendix **Table S6.** 43 candidate momilactone gene clusters in plant kingdom. Also supplied as excel data

Table S7. Numbers of homolog genes for terpene synthase, CYP, and MAS in 107 plant species

Speceis	Version	Number of terpene related gene	DTC and DTC-like	CPS	KSL	Other TPS	P450	Mas
Aquilegia coerulea	Acoerulea_322_v3.1	76	1	13	52	10	496	120
Ananas comosus	Acomosus_321_v3	21	0	3	10	8	181	76
Arabidopsis halleri	Ahalleri_264_v1.1	25	2	1	15	7	191	75
Amaranthus hypochondriacus	Ahypochondriacus_459_v2.1	34	1	1	14	18	200	68
Arabidopsis lyrata	Alyrata_384_v2.1	26	2	1	12	11	241	87
Anacardium occidentale	Aoccidentale_449_v0.9	86	10	4	57	15	428	158
Asparagus officinalis	Aofficinalis_498_V1.1	18	0	4	10	4	177	71
Arabidopsis thaliana	Athaliana_167_TAIR10	34	5	1	24	4	248	91
Amborella trichopoda	Atrichopoda_291_v1.0	38	1	2	16	19	233	64
Azolla_filiculoides	Azolla_filiculoides_v1.1	1	1	0	0	0	101	80
Botryococcus braunii	Bbraunii_502_v2.1	0	0	0	0	0	92	66
Brachypodium distachyon	Bdistachyon_314_v3.1	24	5	1	16	2	264	86
Beta vulgaris	Beta_vulgaris.RefBeet-1.2.2	20	1	2	15	2	199	70
Brachypodium hybridum	Bhybridum_463_v1.1	45	9	3	32	1	512	166
Brassica oleracea capitata	Boleraceacapitata_446_v1.0	32	0	2	20	10	274	101
Brassica rapa	BrapaFPsc_277_v1.3	35	0	2	29	4	343	128
Brachypodium stacei	Bstacei_316_v1.1	22	4	1	17	0	243	81
Boechera stricta	Bstricta_278_v1.2	26	0	1	18	7	209	88
Brachypodium sylvaticum	Bsylvaticum_490_v1.1	29	5	2	19	3	260	89
Cicer arietinum	Carietinum_492_v1.0	23	0	2	12	9	208	71
Citrus clementina	Cclementina_182_v1.0	11	0	2	8	1	339	92
Capsella grandiflora	Cgrandiflora_266_v1.1	26	0	1	19	6	208	73
Chondrus crispus	Chondrus_crispus.ASM35022v2	0	0	0	0	0	17	18
Carica papaya	Cpapaya_113_ASGPBv0.4	31	3	4	10	14	179	65
Chenopodium quinoa	Cquinoa_392_v1.0	62	3	2	43	14	480	155
Chlamydomonas reinhardtii	Creinhardtii_281_v5.5	0	0	0	0	0	41	44
Capsella rubella	Crubella_183_v1.0	34	1	1	28	4	237	84
Cucumis sativus	Csativus_122_v1.0	28	1	1	21	5	222	78
Citrus sinensis	Csinensis_154_v1.1	66	3	1	34	28	305	85
Coccomyxa subellipsoidea	CsubellipsoideaC_169_227_v2.0	0	0	0	0	0	27	48
Cyanidioschyzon merolae	Cyanidioschyzon_merolae.ASM9120v1	0	0	0	0	0	5	22
Chromochloris zofingiensis	Czofingiensis_461_v5.2.3.2	0	0	0	0	0	49	67
Daucus carota	Dcarota_388_v2.0	36	1	3	23	9	312	115
Dioscorea rotundata	Dioscorea_rotundata.TDr96_F1_Pseudo_Chromosome_v1.0	17	0	2	11	4	213	75
Dunaliella salina	Dsalina_325_v1.0	0	0	0	0	0	30	45
Echinochloa crus-galli	Ecrus-galli_v6.prot	75	2	2	42	29	900	295
Eucalyptus grandis	Egrandis_297_v2.0	99	8	1	73	17	572	147
Eutrema salsugineum	Esalsugineum_173_v1.0	18	1	1	13	3	213	80
Fragaria vesca	Fvesca_226_v1.1	56	8	6	28	14	292	77
Galdieria sulphuraria	Galdieria_sulphuraria.ASM34128v1	0	0	0	0	0	8	26
Gossypium hirsutum	Ghirsutum_458_v1.1	78	6	5	63	4	653	209
Glycine max	Gmax_275_Wm82.a2.v1	36	0	3	25	8	441	160
Gossypium raimondii	Graimondii_221_v2.1	69	2	3	51	13	373	122
Helianthus annuus	Hannuus_494_r1.2	98	0	6	76	16	604	177
Calohypnum plumiforme	Cplumiforme_V5	7	5	1	1	0	150	93
Hordeum vulgare	Hvulgare_462_r1	39	2	4	27	6	394	124
Kalanchoe fedtschenkoi	Kfedtschenkoi_382_v1.1	45	0	1	31	13	321	106
Kalanchoe laxiflora	Klaxiflora_309_v1.1	74	1	2	60	11	526	177
Leersia perrieri	Leersia_perrieri.Lperr_V1.4	31	2	3	22	4	291	86
Lactuca sativa	Lsativa_467_v5	54	0	1	42	11	426	148
Lupinus angustifolius	Lupinus_angustifolius.LupAngTanjil_v1.0	5	0	0	4	1	90	40
Linum usitatissimum	Lusitatissimum_200_v1.0	57	6	3	31	17	436	194
Musa acuminata	Macuminata_304_v1	59	3	1	32	23	238	94
Malus domestica	Mdomestica_196_v1.0	87	4	10	35	38	596	186
Manihot esculenta	Mesculenta_305_v6.1	50	0	2	40	8	342	121
Mimulus guttatus	Mguttatus_256_v2.0	57	2	11	38	6	345	78
Marchantia polymorpha	Mpolymorpha_320_v3.1	20	3	2	3	12	148	46
Micromonas pusilla	MpusillaCCMP1545_228_v3.0	0	0	0	0	0	12	33

***SI Appendix* Table S7**. Numbers of homolog genes for terpene synthase, CYP, and MAS in 107 plant species.
*Also supplied as excel data.

Table S7. Numbers of homolog genes for terpene synthase, CYP, and MAS in 107 plant species (*continue*)								
Speceis	Version	Number of terpene related gene	DTC and DTC-like	CPS	KSL	Other TPS	P450	Mas
Chara braunii	mRNA_Chbra_active_pep_20170414.tfa	0	0	0	0	0	19	29
Miscanthus sinensis	Msinensis_497_v7.1	62	2	2	45	13	600	211
Micromonas	MspRCC299_229_v3.0	0	0	0	0	0	12	41
Medicago truncatula	Mtruncatula_285_Mt4.0v1	48	1	4	35	8	424	128
Nicotiana attenuata	Nicotiana_attenuata.NIATTr2	50	0	3	34	13	279	107
Olea europaea	Oeuropaea_451_v1.0	75	0	5	52	18	346	101
Ostreococcus lucimarinus	Olucimarinus_231_v2.0	0	0	0	0	0	11	26
Oryza barthii	Oryza_barthii.O.barthii_v1	55	2	4	37	12	285	86
Oryza brachyantha	Oryza_brachyantha.Oryza_brachyantha.v1.4b	32	3	3	23	3	280	86
Oryza glaberrima	Oryza_glaberrima.AGI1.1	42	1	3	31	7	295	102
Oryza glumaepatula	Oryza_glumaepatula.ALNU02000000	48	5	3	31	9	321	86
Oryza indica	Oryza_indica.ASM465v1	57	3	3	35	16	399	144
Oryza meridionalis	Oryza_meridionalis.Oryza_meridionalis_v1.3	39	1	1	27	10	247	62
Oryza nivara	Oryza_nivara.AWHD00000000	48	4	3	33	8	309	91
Oryza punctata	Oryza_punctata.AVCL00000000	31	2	2	21	6	270	83
Oryza rufipogon	Oryza_rufipogon.OR_W1943	48	5	3	32	8	299	79
Oryza sativa	Osativa_323_v7.0	53	2	3	38	10	360	111
Oropetium thomaeum	Othomaeum_386_v1.0	13	1	2	4	6	100	60
Picea abies	Pabies1.01.0	56	1	2	35	18	382	87
Populus deltoides	PdeltoidesWV94_445_v2.1	68	1	2	50	15	432	137
Panicum hallii	Phallii_495_v3.1	32	2	4	24	2	285	95
Physcomitrella patens	Ppatens_318_v3.3	6	1	1	2	2	94	64
Prunus persica	Ppersica_298_v2.1	44	8	1	23	12	294	98
Populus trichocarpa	Ptrichocarpa_444_v3.1	60	1	2	42	15	415	130
Porphyra umbilicalis	Pumbilicalis_456_v1.5	0	0	0	0	0	14	33
Panicum virgatum	Pvirgatum_450_v4.1	125	3	17	73	32	691	231
Phaseolus vulgaris	Pvulgaris_442_v2.1	38	0	2	31	5	272	104
Ricinus communis	Rcommunis_119_v0.1	49	2	1	31	15	265	148
Sorghum bicolor	Sbicolor_454_v3.1.1	43	2	2	34	5	349	105
Sphagnum fallax	Sfallax_310_v0.5	6	3	1	1	1	205	103
Setaria italica	Sitalica_312_v2.2	43	3	5	29	6	363	109
Solanum lycopersicum	Slycopersicum_390_ITAG2.4	49	0	1	35	13	255	104
Selaginella moellendorffii	Smoellendorffii_91_v1.0	25	3	4	13	5	291	65
Spirodela polyrhiza	Spolyrhiza_290_v2	17	0	2	5	10	152	70
Salix purpurea	Spurpurea_289_v1.0	51	6	2	35	8	363	127
Solanum tuberosum	Stuberosum_448_v4.03	77	3	3	44	27	471	113
Setaria viridis	Sviridis_500_v2.1	47	3	4	32	8	356	108
Triticum aestivum	Taestivum_296_v2.2	149	2	18	84	45	1115	308
Theobroma cacao	Tcacao_233_v1.1	43	1	1	30	11	296	86
Trifolium pratense	Tpratense_385_v2	41	0	2	23	16	419	115
Triticum dicoccoides	Triticum_dicoccoides.WEWSeq_v.1.0	76	1	8	48	19	819	222
Volvox carteri	Vcarteri_317_v2.1	0	0	0	0	0	18	37
Vigna angularis	Vigna_angularis.Vigan1.1	25	0	2	21	2	256	109
Vigna radiata	Vigna_radiata.Vradiata_ver6	48	0	1	40	7	305	155
Vigna unguiculata	Vunguiculata_469_v1.1	29	0	2	26	1	301	127
Vitis vinifera	Vvinifera_145_Genoscope.12X	97	8	4	44	41	319	102
Zizania latifolia	Zlat_V1.protein	34	1	4	23	6	246	83
Zostera marina	Zmarina_324_v2.2	2	0	1	1	0	110	41
Zea mays	Zmays_284_Ensembl-18_2010-01-MaizeSequence	36	4	3	26	3	300	90

***SI Appendix* Table S7**. Numbers of homolog genes for terpene synthase, CYP, and MAS in 107 plant species (*continue*). *Also supplied as excel data

***SI Appendix* Referencesx**

1. A. V. Zimin *et al.*, Hybrid assembly of the large and highly repetitive genome of. *Genome Res* **27**, 787-792 (2017).
2. R. K. Patel, M. Jain, NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**, e30619 (2012).
3. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579 (2011).
4. A. C. English *et al.*, Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7**, e47768 (2012).
5. F. Nadalin, F. Vezzi, A. Policriti, GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics.***13** Suppl 14:S8 (2012).
6. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
7. R. Luo *et al.*, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
8. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4:Unit 4.10. (2009).
9. M. Stanke *et al.*, AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34:W435-439 (2006).
10. A. V. Lukashin, M. Borodovsky, GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26:1107-1115 (1998).
11. A. A. Salamov, V. V. Solovyev, Ab initio gene finding in Drosophila genomic DNA. *Genome Res* 10:516-522 (2000).
12. T. D. Wu, C. K. Watanabe, GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859-1875. (2005).
13. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
14. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. May;28(5):511-515 (2010).
15. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).
16. B. J. Haas *et al.*, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666 (2003).
17. A. Rambaut, FigTree v1. 4.2. A graphical viewer of phylogenetic trees. Institute of Evolutionary Biology University of Edinburgh. (2014).
18. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. **8**(1):77-80. (2010).

19. H. Wickham, ggplot2[J]. *Wiley Interdisciplinary Reviews Computational Statistics*, 3(2):180-185 (2015).

20. L. Guo, *et al.*, A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. *Plant J*. **83**(4):600-609 (2015).

21. S. Miyazaki, T. Katsumata, M. Natsume, H. Kawaide, The CYP701B1 of *Physcomitrella patens* is an *ent*-kaurene oxidase that resists inhibition by uniconazole-P. *FEBS Lett*. **585**(12):1879-1883 (2011)

22. Y. Sawada, *et al.*, Germination of photoblastic lettuce seeds is regulated via the control of endogenous physiologically active gibberellin content, rather than of gibberellin responsiveness. *J. Exp. Bot.* **59**, 3383–3393 (2008)

23. S. Miyazaki *et al.*, Analysis of *ent*-kaurenoic acid by ultra-performance liquid chromatography-tandem mass spectrometry. *Biochem Biophys Rep.* **2**, 103-107 (2015)