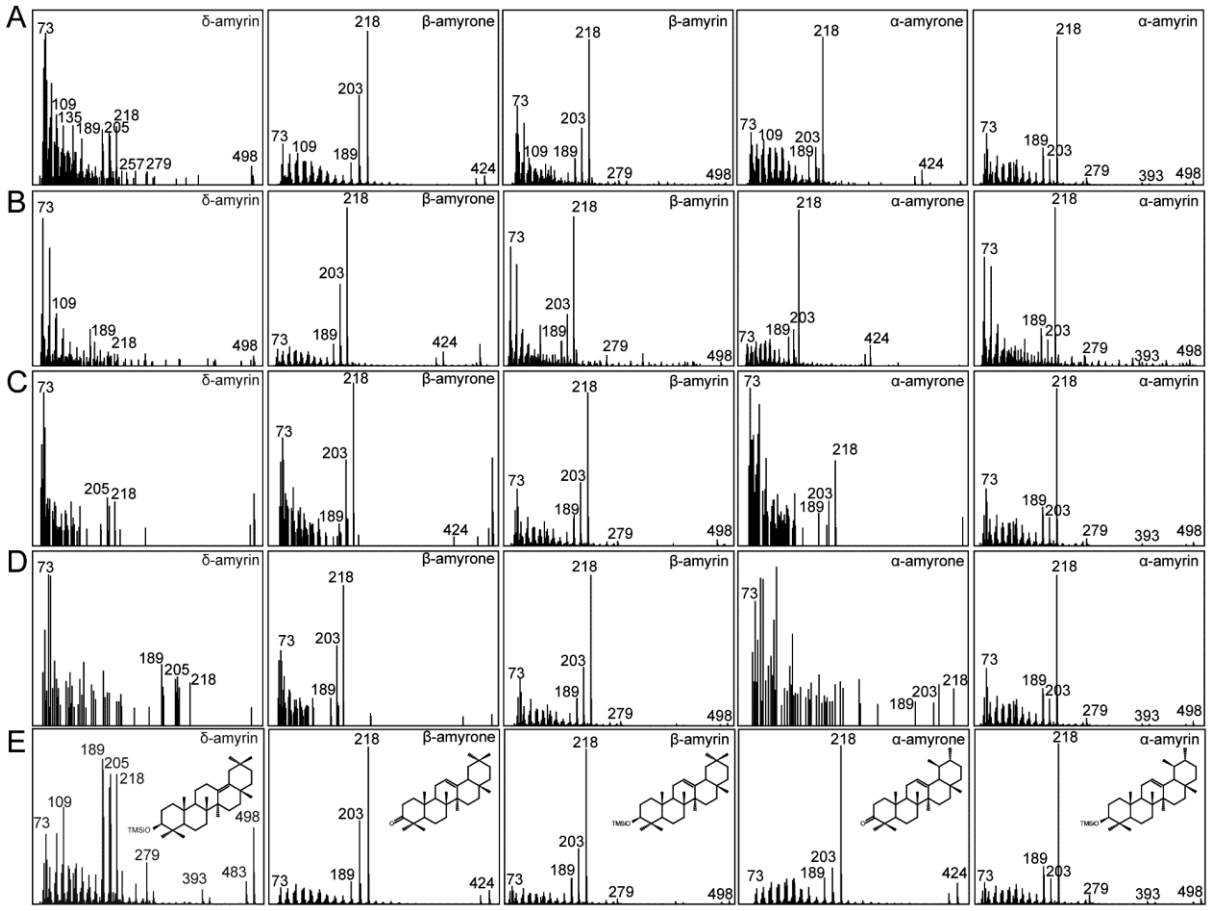
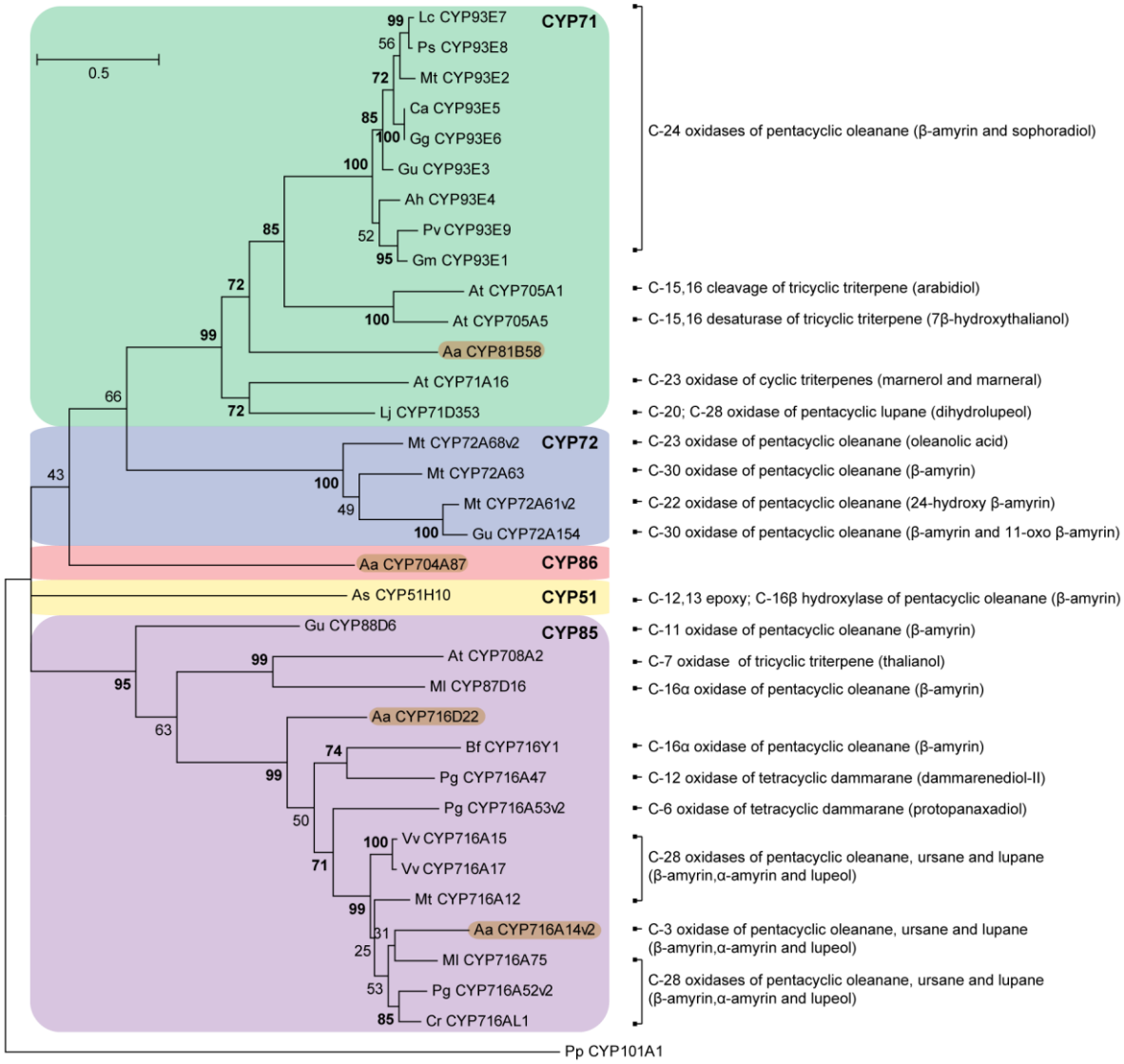
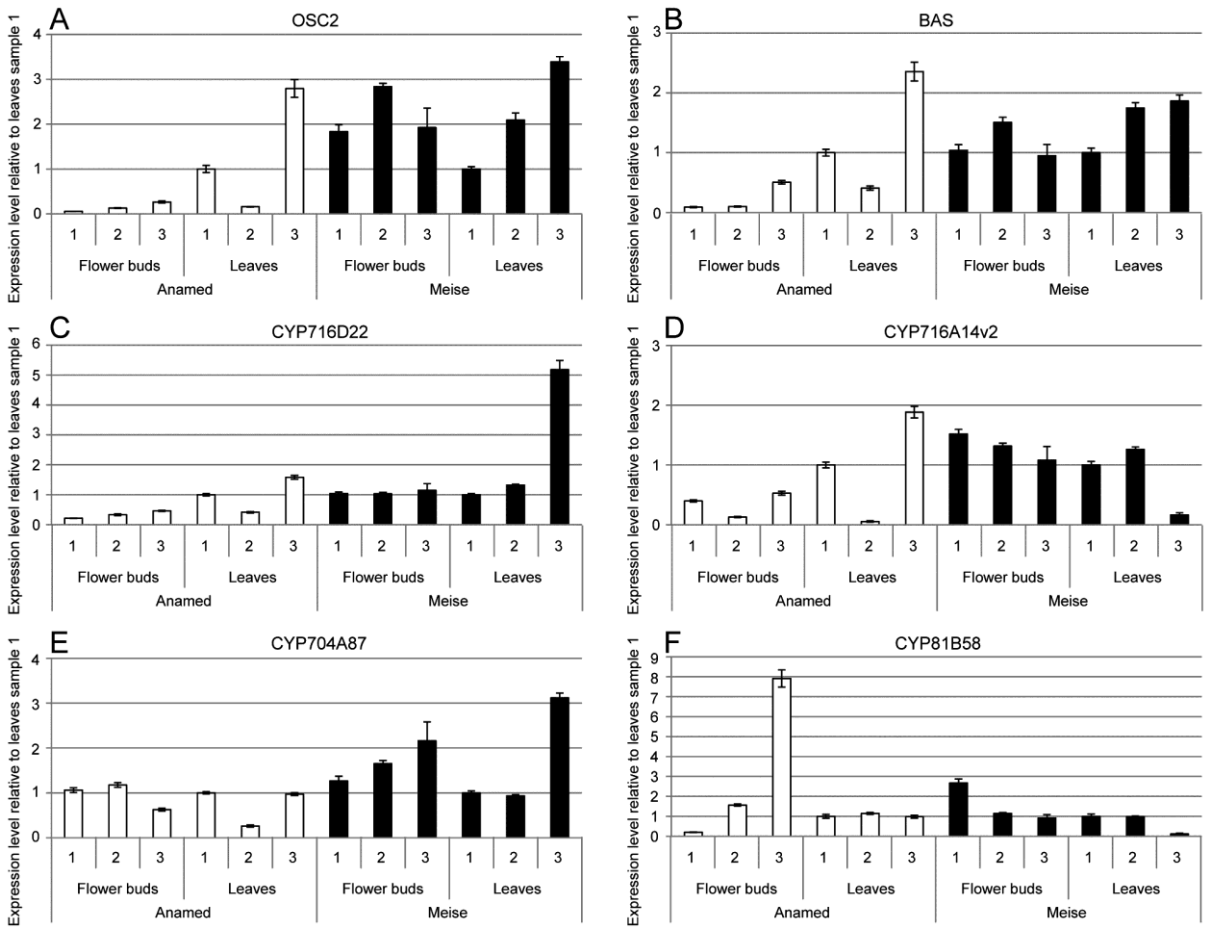
SUPPLEMENTAL DATA



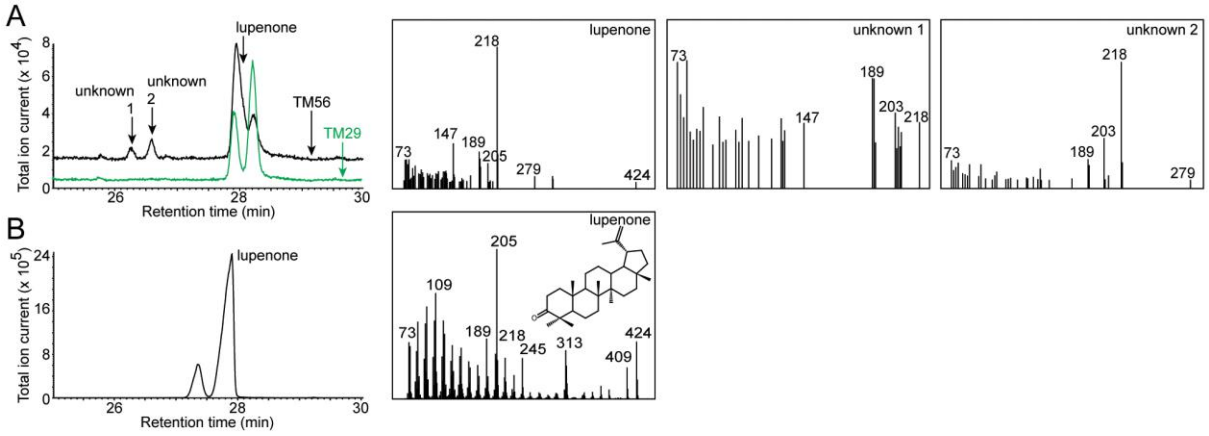
Supplemental Figure 1. Identity of *A. annua* triterpenoids. EI-MS data for the triterpenes and triterpenoids inherent to *A. annua* flower buds **(A,B)** and leaves **(C,D)** from the Anamed A3 **(A,C)** and Meise **(B,D)** cultivars. The peaks corresponding to these EI-MS are given in Figure 4. **(E)** EI-MS and structures of authentic trimethylsilylated δ-amyrin, β-amyrin and α-amyrin, and underivatized β-amyrone and α-amyrone.



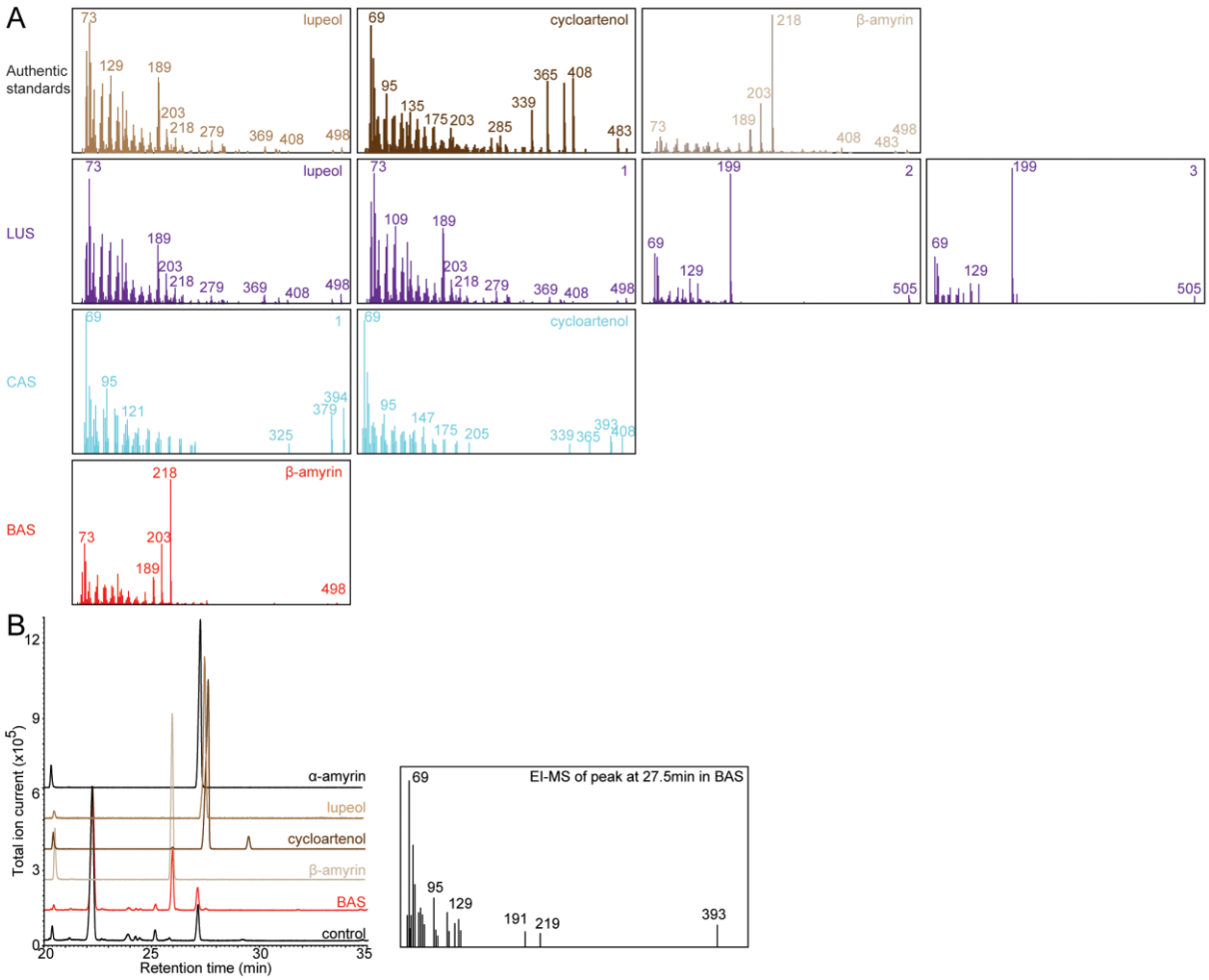
Supplemental Figure 2. Phylogenetic analysis of *A. annua* candidate P450s and other P450s involved in specialized triterpenoid biosynthesis. CYP716D22, CYP716A14v2, CYP704A87, CYP81B58 are highlighted in brown. The enzymatic activity of the P450s is indicated on the right. The percentage of replicate trees that clustered together in the bootstrap test is indicated on the branches. High scoring bootstrap values (≥70%) are highlighted in bold. The bacterial CYP101A1 is included as an outgroup. The P450 clan names are mentioned in bold and are color coded as green, CYP71 clan; blue, CYP72 clan; red, CYP86 clan; yellow, CYP51 clan; purple, CYP85 clan. The scale bar represents the number of amino acid substitutions per site. The amino acid sequences were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Aa, *Artemisia annua*; Ah, *Arachis hypogaea*; As, *Avena strigosa*; At, *Arabidopsis thaliana*; Bf, *Bupleurum falcatum*; Ca, *Cicer arietinum*; Cr, *Catharanthus roseus*; Gg, *Glycyrrhiza glabra*; Gm, *Glycine max*; Gu, *Glycyrrhiza uralensis*; Lc, *Lens culinaris*; Lj, *Lotus japonicus*; Ml, *Maesa lanceolata*; Mt, *Medicago truncatula*; Pg, *Panax ginseng*; Pp, *Pseudomonas putida*; Ps, *Pisum sativum*; Pv, *Phaseolus vulgaris*; Vv, *Vitis vinifera*.



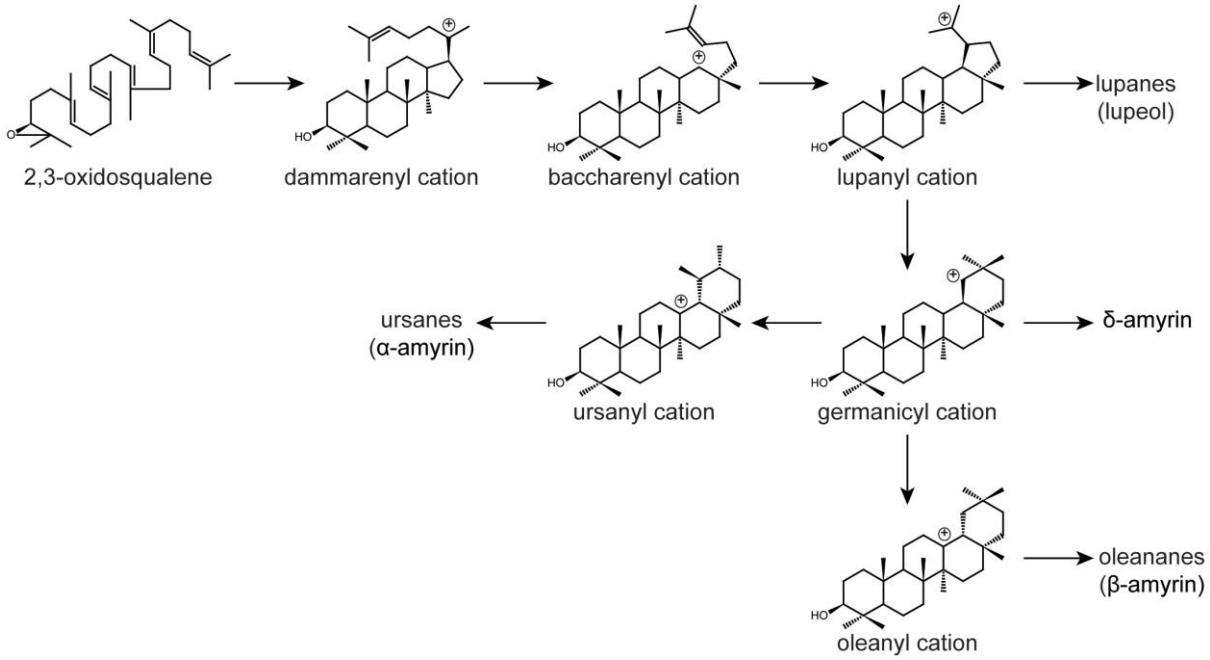
Supplemental Figure 3. Organ-specific expression of candidate OSC- and P450-family triterpene biosynthesis genes in *A. annua* cultivars. The organ-specific expression of the *OSC*s, *OSC2* **(A)** and *BAS* **(B)**, and candidate *P450*s, *CYP716D22* **(C)**, *CYP716A14v2* **(D)**, *CYP704A87* **(E)** and *CYP81B58* **(F)**, were verified by quantitative real-time polymerase chain reaction (qPCR) in flower buds and leaves of Anamed A3 and Meise cultivars. Y-axis, expression levels relative to those in the flower buds. The labels 1 to 3 in the X-axis refer to the three biological repeats. Error bars are standard error of the mean for n = 3.



Supplemental Figure 4. CYP716A14v2 is a multifunctional P450. **(A)** Overlay of GC-MS chromatograms showing the oxidation of lupeol to lupenone and other compounds (unknown 1 and unknown 2) by CYP716A14v2 in the yeast strain TM56 (black), but not in the control yeast strain TM29 (green). **(B)** GC-MS chromatogram of an authentic standard of lupenone, the structure of which is depicted in the inset. Insets at the right show the EI-MS spectra of unknown 1, unknown 2 and lupenone. The GC retention times of unknown 1, unknown 2 and lupenone are 26.2 min, 26.6 min and 27.9 min, respectively.



Supplemental Figure 5. Identification of the products of *A. annua* OSCs. **(A)** EI-MS spectra of new peaks and authentic standards in Figure 8. **(B)** Overlay of GC-MS chromatograms of authentic standards of α-amyrin, lupeol, cycloartenol, β-amyrin, yeast strain TM106 expressing *BAS* (red), and control strain TM111 (black) showing the cyclization of 2,3-oxidosqualene by BAS to β-amyrin and trace amounts of a compound eluting at 27.5 min. The EI-MS spectra of the peak eluting at 27.5 min is shown as an inset on the right.



Supplemental Figure S6. Schematic of the 2,3-oxidosqualene cyclization cascade mediated by oxidosqualene cyclases to generate oleanane, ursane and lupane triterpene backbones.

Supplemental Table 1. RNA-Seq triterpene contig counts for glandular and filamentous trichomes.

Gene name	Representative Contig	Counts FT[1]	Counts GT[1]	Ratio FT/GT	p-value
OSC2	*comp7642*	257.57	6.27	41.06	1.09 E-04
BAS	*comp33386*	147.15	2.04	72.16	1.92 E-01
CYP704A87	*comp8932*	62.45	0.00	∞	1.18 E-04
CYP81B58	*comp9232*	255.87	2.43	105.35	8.03 E-07
CYP716D22	*comp33076*	118.26	16.18	7.31	6.38 E-01
CYP716A14v2	*comp11264*	461.97	24.16	19.12	3.01 E-01

[1]Sum of normalized counts from 6 samples of filamentous (FT) and glandular (GT) trichomes.

Supplemental Table 2. Yeast strains generated in this study.

Strain	Genotype
S288c BY4742	MATa *his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0*
TM1	S288c BY4742; P_{erg7}::P_{MET3}-*ERG7*
TM5	TM1; pESC-URA[*GAL10/tHMG1*]
TM6	TM1; pESC-URA[*GAL10/tHMG1*; *GAL1*/At-*LUS1*]
TM29	TM6; pAG423, pAG415[*GAL1*/At-*ATR1*]
TM50	TM1; pESC-URA[*GAL10/tHMG1*; *GAL1/OSC2*]
TM51	TM50; pAG423[*GAL1/CYP716D22*], pAG415[*GAL1*/At-*ATR1*]
TM52	TM50; pAG423[*GAL1/CYP716A14v2*], pAG415[*GAL1*/At-*ATR1*]
TM53	TM50; pAG423[*GAL1/CYP704A87*], pAG415[*GAL1*/At-*ATR1*]
TM54	TM50; pAG423[*GAL1/CYP81B58*], pAG415[*GAL1*/At-*ATR1*]
TM55	TM50; pAG423, pAG415[*GAL1*/At-*ATR1*]
TM56	TM6; pAG423[*GAL1/CYP716A14v2*], pAG415[*GAL1*/At-*ATR1*]
TM106	TM5; pAG423[*GAL1/BAS*]
TM107	TM5; pAG423[*GAL1/CAS*]
TM108	TM5; pAG423[*GAL1/LUS*]
TM109	TM5; pAG423[*GAL1/OSC2*]
TM110	TM5; pAG423[*GAL1/OSC3*]
TM111	TM5; pAG423

Table S3. Primers used for cloning.

Primer No.	Sequence (5' to 3')
P1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGGAAGTTGAAGGTA
P2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATTTAGTTTTGACCCA
P3	GGGGACAAGTTTGTACAAAAAAGCAGGCTTActcgagATGTGGAAGTTGAAGGTAGC
P4	GGGGACCACTTTGTACAAGAAAGCTGGGTTgctagcTTATTTAGTTTTGACCCAAAC
P5	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGATACAAGTTTTAACATC
P6	GGGGACCACTTTGTACAAGAAAGCTGGGTATCATACTTGATGAGGATGAAG
P7	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGTTGATCTCATTCTC
P8	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGGCATAACGAGGATA
P9	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAAGACCCCTTTCTTC
P10	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGAAGCGATGAAAAGC
P11	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATACATTATATGTT
P12	GGGGACCACTTTGTACAAGAAAGCTGGGTATTACAGCTGAGATATTAA
P13	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCCCGGGCATGTGGAGATTGAAAA TAGC
P14	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTCGAGCTAGGTGCCTTTGAGCTGT GG
P15	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGGAAATTGAAGATCGCAG
P16	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAGTTTCCACTTTAAGTAC
P17	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGGAAGTTGAAGATAGCGG
P18	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGGTAGTGATCCCATCAC
P19	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGGAGATTGAAAATAGCAG
P20	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATGTATTGATTTGAACTCTTG
P21	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATAGCGAAGCAGTTCGTAAAGC
P22	GGGGACCACTTTGTACAAGAAAGCTGGGTAGATATTCTCCAAGTGCCCAAAG
P23	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCAAATGCTTTGTGATGAAGAG
P24	GGGGACCACTTTGTACAAGAAAGCTGGGTACCAACCTTTTGGAATGGAATAA

The sequences in lower case represent the restriction recognition sites used for restriction enzyme-mediated cloning.

Table S4. Primers used for qPCR analysis.

Name	Sequence (5' to 3')
OSC2	CGGTCGAGCGTCAAGAAGTA CGCATAAGCAAATCACCGCA
BAS	GGTGGGTGGGGATTTCACAT CACGACCTTCCCCCAAAAGA
CYP716D22	CTGCTGCGGCACATAAAGAC GAGTTGGTCCTGCACCTTCA
CYP716A14v2	AAGCTGACATCGCTGACGTA TGCACATGTTGAGCTTGCAT
CYP704A87	AGCGGCTCTCAAGGAAAACA AGCTGCTCCCTCTTGTTTCG
CYP81B58	AGGCACGACGGTTTAAGGAG CTCCGATCCACTTCCACCAC
Actin	TCTCACAATTTCCCGCTCTG CCACACGCCATCCTTCG
PP2AA3	GGGAAGTTGTTACAGCCCCA TACCATATACCGCACACGCC
PPR	GGGAAACTTAAGGTTGCGGT ACCATCCCGAGAGTACCCAT