Supporting Information

Field et al. 10.1073/pnas.1109273108

SI Materials and Methods

Plant Materials. *Arabidopsis thaliana* accession Columbia was used as wild type for this work. The transfer DNA (T-DNA) insertion mutants *mro1-1* (SALK_073803), *mro1-2* (SALK_045988), and SALK_017783 were provided by the Signal Insertion Mutant Library (http://signal.salk.edu/cgi-bin/tdnaexpress/) and were obtained via the Nottingham *Arabidopsis* Stock Centre (http://nasc.life.nott.ac.uk/). Insertion sites are indicated (Fig. 1*B*). Homozygous insertion mutants were confirmed by PCR genotyping. RT-PCR was used to determine the presence of full-length transcript in each mutant (Fig. S6).

Cloning and Transformation. The *MRN1* (*At5g42600*) coding sequence was amplified from *Arabidopsis* root cDNA with Phusion polymerase (Finnzymes) and introduced into the galactose inducible yeast expression vector pYES/DEST-52 (Invitrogen). Triterpene production was analyzed in the yeast strain GIL77 as described previously (1). The *MRN1* coding sequence also was introduced into the plant overexpression vector pGGW42 (2) (kindly provided by Alan Herr, Department of Pathology, University of Washington, Seattle, WA), which was used to transform wild-type plants by floral dipping to give *MRN1*-overexpressing plants (35S:MRN/At5g42600). The marneral oxidase (*MRO*) coding sequence (*CYP71A16*/*At5g42590*) was amplified from *Arabidopsis* root cDNA with Phusion polymerase and was introduced into the plant overexpression vector pGGW6 (2) (kindly provided by Alan Herr). The resulting vector then was used to transform wild-type plants to give *MRO*-overexpressing plants (35S:CYP71A16/At5g42590). Plants overexpressing *MRN1* and *MRO* were produced by crossing plants overexpressing the single genes.

Triterpene Analysis. For standard extractions, lyophilized plant tissue was saponified in 10% KOH (wt/vol) in 80% EtOH (vol/vol) with 0.5 mg/mL butylated hydroxytoluene (Sigma-Aldrich) at 70 °C for 2 h, extracted with hexane, derivatized with Tri-Sil Z (Pierce), and analyzed by GC-MS as described previously (2). The products of MRO could be identified only when the saponification step was replaced by incubation in 80% EtOH (vol/vol) with 0.5 mg/mL butylated hydroxytoluene. The alcohol marnerol was detected in saponified and nonsaponified plant extracts. The aldehyde marneral could not be detected in nonderivatized extracts. Therefore, marneral may be converted to marnerol *in planta* or during the extraction process. We did not observe changes in thalianol or thalianol derivatives in *MRN1* or *MRO* mutant plants.

Phylogenetic Analysis. Predicted protein sequences were retrieved from Phytozome (www.phytozome.net) or the National Center for Biotechnology Information (NCBI) and were aligned with the Multiple Alignment using Fast Fourier Transform (MAFFT) software package (3). Phylogenetic analyses were conducted in the Molecular Evolutionary Genetics Analysis (MEGA) v. 4 software package (4). Evolutionary relationships were inferred using multiple tree-building methods and distance matrices. All methods tested provided broadly similar phylogenetic trees that included Brassicaceae-specific clades.

Genome Analysis. Chromosome maps are derived from the Arabidopsis Information Resource Release 9 database for *A. thaliana* and Phytozome for *Arabidopsis lyrata*. Comparative genome analysis was performed online using the GeVO tool (http://synteny.cnr.berkeley.edu/CoGe/) (5). The promoters (3 kb upstream of the start codon) of the thalianol and marneral cluster genes were analyzed using the Blastn algorithm in GeVO. Several short matches were identified between promoter pairs in the thalianol and marneral clusters. The match with the lowest *E*-value occurred between the promoters of *THAS*/*At5g48000* and *At5g47980* and covered 21 nucleotides (90.4% identity, *E*-value = 0.04). The enrichment of promoters in known motifs was analyzed with the ATHENA (6) (3-kb promoter sequence) and AtCOEICiS (7) (1-kb promoter sequence) promoter-analysis tools.

1. Haralampidis K, et al. (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc Natl Acad Sci USA* 98:13431–13436.
2. Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* 320:543–547.
3. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298.
4. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
5. Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* 53:661–673.
6. O'Connor TR, Dyreson C, Wyrick JJ (2005) Athena: A resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. *Bioinformatics* 21:4411–4413.
7. Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol* 150:535–546.

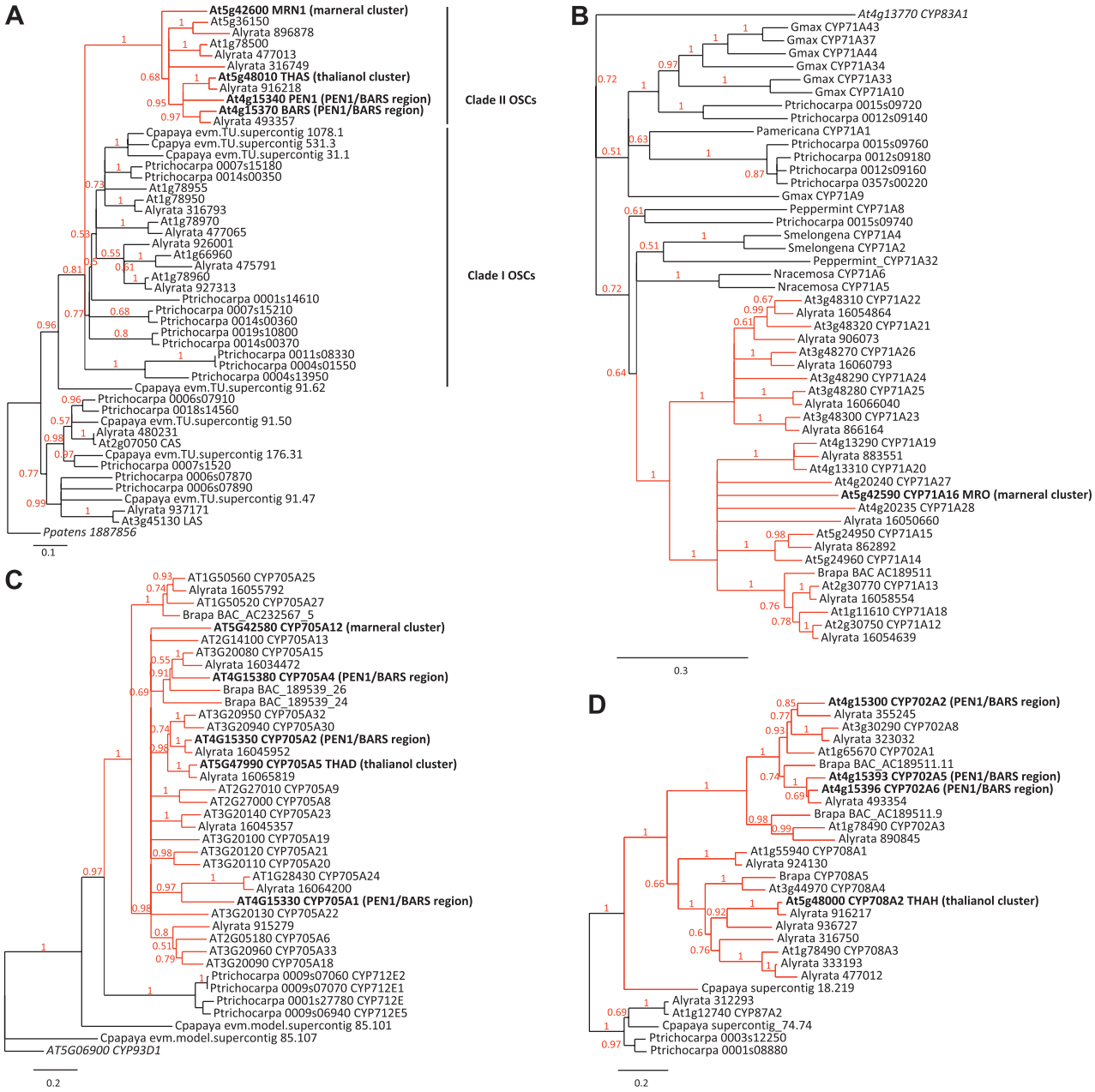

Fig. S1. Evolution of the marneral and thalianol cluster gene products. Phylogenetic trees for the (*A*) oxidosqualene cyclase (OSC), (*B*) CYP71, (*C*) CYP705, and (*D*) CYP702/708 enzyme families. Brassicaceae-specific and post-α duplication clades are shown by red branches. (*A*) OSCs were retrieved from the genomes of *Physcomitrella patens* (Ppatens), *Populus trichocarpa* (Ptrichocarpa), *Carica papaya* (Cpapaya), *Arabidopsis lyrata* (Alyrata), and *Arabidopsis thaliana* (At). Phytozome accession names are shown. We previously identified two distinct OSC clades in *Arabidopsis* (1). Clade II appears to have undergone accelerated evolution compared with clade I, and our phylogenetic analysis shows that it is restricted to the Brassicaceae. Accessions in bold text are encoded by genes within the marneral or thalianol functional gene clusters or the putative *pentacyclic triterpene synthase 1* (*PEN1*) and *baruol synthase 1* (*BARS1*) functional gene cluster. (*B–D*) Known CYP71, CYP705, and CYP702/8 subfamily members were retrieved from the cytochrome P450 homepage (http://drnelson.uthsc.edu/CytochromeP450.html). Known P450s are labeled with the species of origin and the P450 identifier. Other likely members of these P450 subfamilies were retrieved from the genomes of *Populus trichocarpa* (Ptrichocarpa), *Carica papaya* (Cpapaya), and *Arabidopsis lyrata* (Alyrata) using BLAST searches. Phytozome accession names are shown for these sequences. Other likely subfamily members also were identified in sequenced *Brassica rapa* BACs. In these cases the NCBI BAC accession number is shown. Cytochrome P450 subfamily trees are rooted with sister subfamilies. The evolutionary histories were inferred using the neighbor-joining method, and distances were calculated with the Poisson correction. One thousand bootstrap replicates were performed, and percentage support is shown at branches. Branches with less than 50% support were collapsed. Where trees are rooted, the roots are shown in italics.

1. Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* 320:543–547.

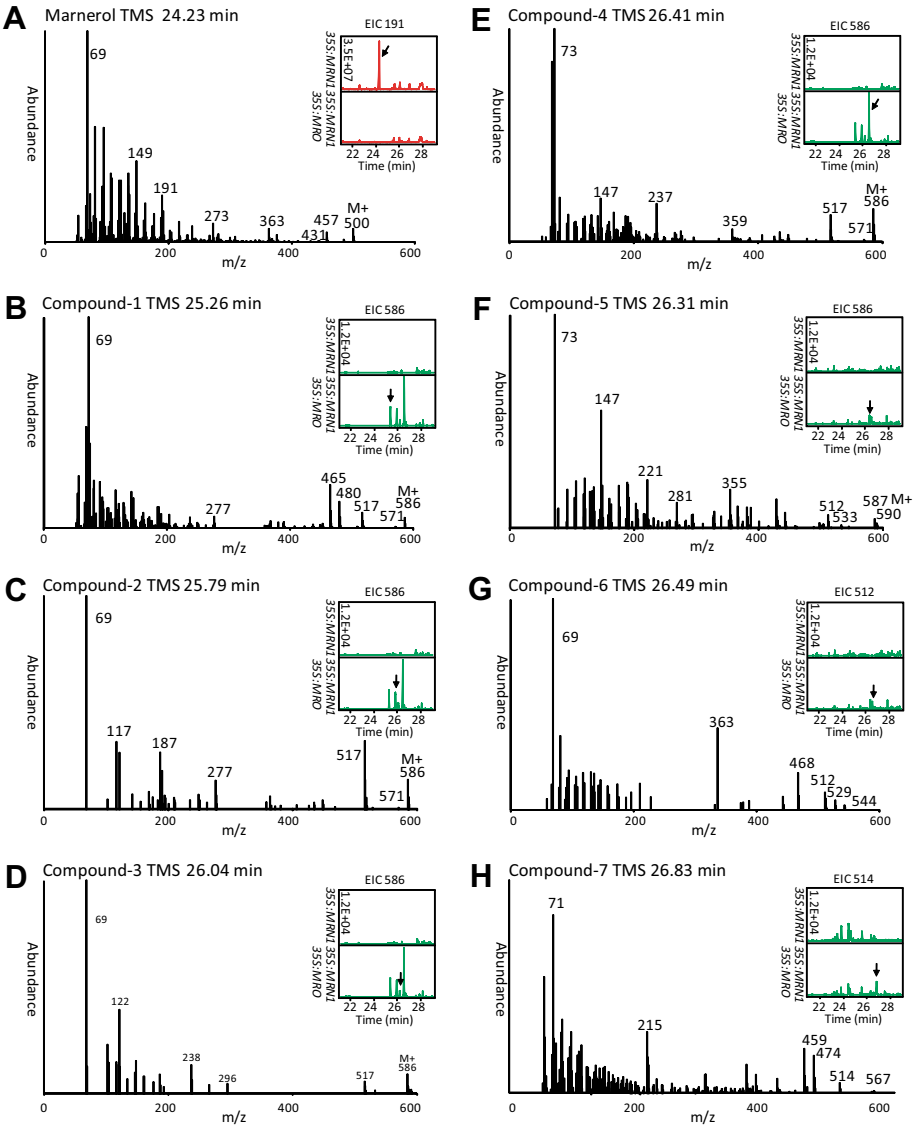

Fig. S2. Ion spectra for marnerol and the marnerol derivatives found in transgenic *Arabidopsis* plants. Ion component spectra extracted using the Automated Mass Spectral Deconvolution and Identification System (AMDIS, http://www.amdis.net) for (*A*) marnerol trimethylsilyl ether (TMS) from leaf extracts of *MRN1*-overexpressing plants and (*B–H*) seven marnerol TMS derivatives identified in plants overexpressing *MRN1* and *MRO*. No marnerol was detected in these plants, indicating complete conversion. The marnerol derivatives could not be detected in extracts from wild-type plants, plants overexpressing MRO alone, or plants overexpressing MRN alone. Compounds 1–4 were present at relatively high abundance and have a predicted molecular ion at *m/z* = 586 that is consistent with the molecular formula $C_{36}H_{66}O_2Si_2$. Therefore, compounds 1–4 are likely to represent different isomers of hydroxylated desaturated marnerol. Compounds 5–7 were present at lower levels and have ionization spectra that suggest that they are derived from marnerol. Compound 5 has a predicted molecular ion at *m/z* = 590 that is consistent with the molecular formula $C_{36}H_{70}O_2Si_2$. No molecular ion could be proposed for compounds 6 or 7, and because of their low abundance it is possible that the molecular ion is not present. M+ indicates the likely molecular ion.

Fig. S3. The phenotypes of plants overexpressing *MRN1* and/or *MRO*. (*A*) Plants that overexpress *MRN1* (*35S:MRN1*) and accumulate marneral develop differently from wild-type plants. (*B*) Plants that overexpress *MRN1* and *MRO* (*35S:MRN1 35S:MRO*) and accumulate marneral derivatives are dwarfed and develop slowly.

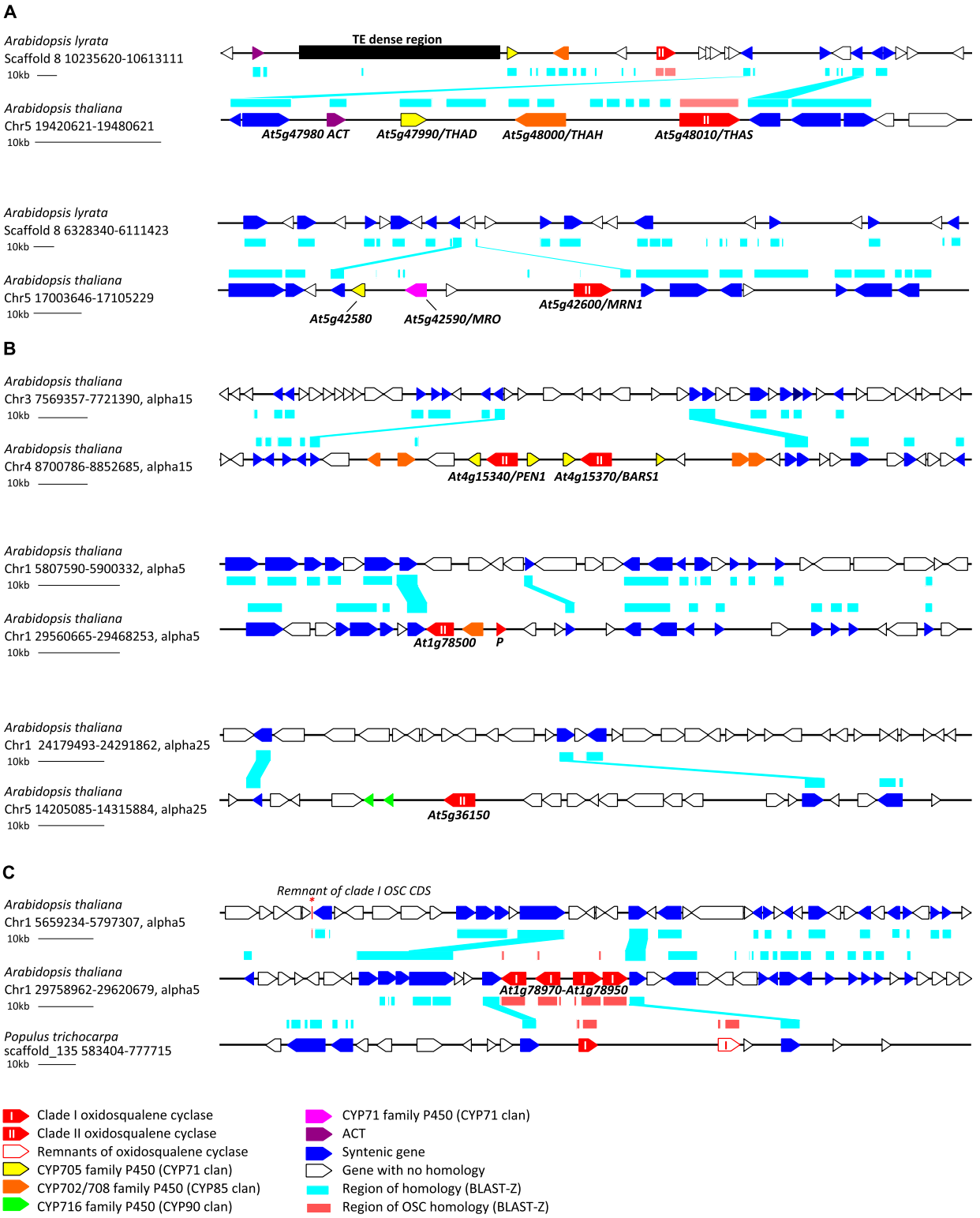

Fig. S4. Maps of chromosomal regions containing the gene *OSC* showing regions of synteny between α duplication segment pairs or different species. (*A*) Maps of the thalianol and marneral regions in *Arabidopsis lyrata* and *Arabidopsis thaliana*. (*B*) Maps of the regions surrounding the four remaining clade II *OSCs* compared with the equivalent region in the matching α duplication segment pair. In each case no trace of a clade II *OSC* or associated *CYP450* could be detected in the matching α duplication segment pair. These results suggest that the clade II *OSCs* arrived in these regions by SD after the α whole-genome duplication event. (*C*) A map of the region surrounding four clade I *OSCs* compared with the matching α duplication segment pair and with the corresponding region in *Populus trichocarpa* (the *Arabidopsis-Populus* species divergence predates the α duplication event). We found traces of a clade I *OSC* in the matching α duplication segment pair: a 244-nucleotide fragment with 66% homology to the coding sequence of *At1g78955* and *At1g78950*. In the corresponding region of *Populus trichocarpa* there are two clade I-like *OSCs*. Together, this evidence, which is supported further by an OSC phylogenetic tree (Fig. S1), indicates that the emergence of the clade I *OSCs* predates both the α duplication event and the *Arabidopsis–Populus* species divergence.

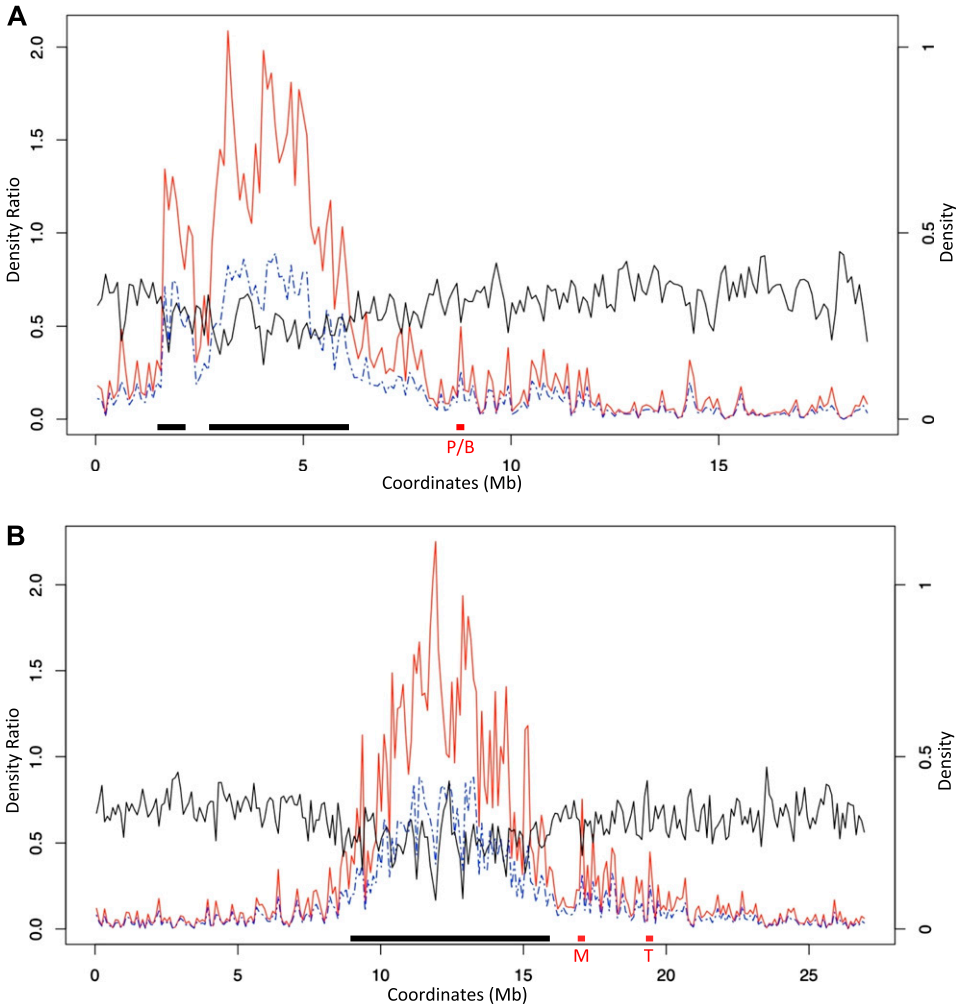

Fig. S5. The chromosomal context of *Arabidopsis* gene clusters. Distribution of gene density (black line), transposable element (TE) density (blue line), and the TE/gene density ratio (red line) across *Arabidopsis thaliana* chromosomes 4 (*A*) and 5 (*B*). The *PEN1*/*BARS* region (*P*/*B*) and the marneral (M) and thalianol (T) clusters are indicated. Centromeric and pericentromeric regions are marked by black bars. TE and gene densities were calculated for windows of 150 kb with an overlap of 5 kb.

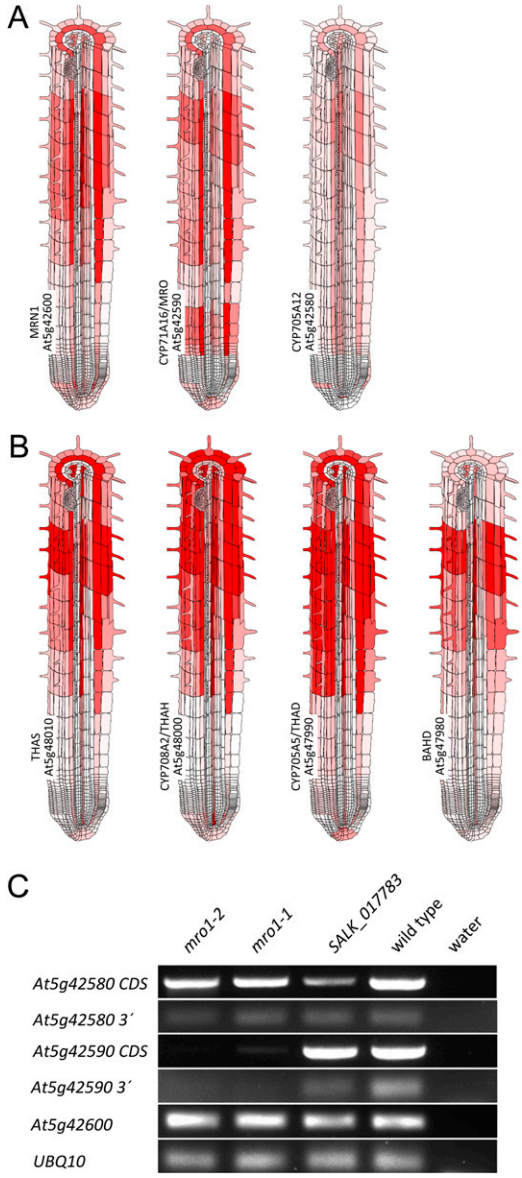

Fig. S6. Root-specific expression patterns of marneral and thalianol cluster genes and validation of T-DNA insertion mutants. Root expression patterns of (*A*) marneral and (*B*) thalianol cluster genes. Data and images are from ArexDB (http://www.arexdb.org/) (1). (*C*) Validation of TDNA insertion mutants. RT-PCR analysis of gene expression in *CYP71A16* T-DNA insertion mutants *mro1-1* and *mro1-2*, and the *At5g42580* TDNA insertion mutant SALK_017783. Primers were used to amplify either the full coding sequence (CDS) or the 3′ region of each gene from cDNA prepared from root extracts. RT-PCR of *MRN1* and *UBQ10* was used as a control. Insertion SALK_017783 is located in the promoter region of *At5g42580* and was the only publicly available, sequenced T-DNA insertion line likely to disrupt the transcription of this gene. We found that there was a small reduction in *At5g42580* transcript levels in the insertion line, but there were no detectable differences in triterpene content.

1. Cartwright DA, Brady SM, Orlando DA, Sturmfels B, Benfey PN (2009) Reconstructing spatiotemporal gene expression data from partial observations. *Bioinformatics* 25:2581−2587.

Dataset S1. TE composition and regions of homology between thalianol and marneral cluster regions

Dataset S1 (XLS)

Dataset S2. Analysis of the promoter regions of the marneral and thalianol cluster genes

Dataset S2 (XLS)