

Supporting Information

Lee et al. 10.1073/pnas.1009975107

SI Materials and Methods

Plant Material. Homozygous gene-knockout plants were screened according to the provider's instructions (<http://signal.salk.edu> and <http://www.gabi-kat.de/>). The *coi1-1* mutant was provided by J. Turner (University of East Anglia, Norwich, United Kingdom). The male sterile flowers of homozygous *coi1* plants were pollinated with pollen from *Pro*_{35S}:*GES* #3 (1). The resulting seeds were further propagated and seedlings of individual plants were characterized in the next generation for homogenous appearance of lesions on the cotyledons and male sterility, which are indicative of the *Pro*_{35S}:*GES* phenotype and homozygosity of the *coi1* allele, respectively. Seeds from these lines were used for further analysis.

Plant Treatment. For treatment with the elicitor alamethicin, six to eight fully expanded rosette leaves (0.5–0.8 g FW) were detached and incubated for 30 h with their petioles placed in a 10-mL solution of 5 µg/mL alamethicin (in 0.1% ethanol). Mock treatments were performed with 0.1% ethanol. Feeding of *Plutella xylostella* larvae on leaves of wild-type and *Pro*_{CYP82G1}:*GUS* plants was conducted by placing four to five larvae in the third-fourth instar on a single rosette. Insects were allowed to feed for 48 h under an 8-h light/16-h dark photoperiod before harvesting of damaged leaves for real-time RT-PCR analysis and GUS-staining assays.

Volatile Analysis. Volatiles collected from plant material via closed-loop stripping were analyzed using a Shimadzu GC-QP2010S. Separation was performed on an Rxi-XLB column (Restek) of 30-m length × 0.25-mm inside diameter × 0.25 µm thickness. Helium was the carrier gas (flow rate of 1.4 mL/min). A splitless injection (injection volume of 1 µL) was used, and a temperature gradient of 5 °C/min from 40 °C (1-min hold) to 220 °C was applied. Mass spectrometry was performed with a transfer line temperature of 280 °C, ion source temperature of 260 °C, 1-kV detector voltage, and a scan range of 50 to 300 atomic mass units. Volatiles collected by solid-phase microextraction (SPME) from P450 enzyme assays were adsorbed with a 100 µm polydimethylsiloxane fiber (Supelco) for 1 h at 30 °C from the headspace of a screw-capped vial (10-mL vials for in vitro enzyme assay and 20-mL vials for in vivo yeast culture assay). Compounds were desorbed from the fiber at 240 °C (4 min) with a 2:1 split injection and analyzed under the same conditions as described above but with a temperature gradient of 4 °C/min from 35 °C (5-min hold) to 60 °C followed by a gradient of 6 °C/min from 60 to 180 °C and 20 °C/min from 180 to 260 °C (3-min hold). The scan range was from 30 to 350 atomic mass units. The identities of all compounds were determined by comparison of retention times and mass spectra with those of authentic standards and with mass spectra in the National Institute of Standards and Technology and Wiley libraries (John Wiley & Sons, Inc., New York, NY). For quantification of geranylinalool and homoterpenes collected by closed-loop stripping, the primary ion peaks of each compound were integrated (single-ion method) and the amounts were calculated based on calibration curves established for each compound. Calibration of SPME for quantitative analysis of homoterpenes is described under *Yeast Expression and Enzyme Assay*.

Transcript Analysis of CYP82G1, GES, and Selected Candidate Genes. Semiquantitative RT-PCR of selected candidate genes (Table S1) was performed with 0.2 µM of each gene-specific primer (Table S3), 0.2 mM dNTP mix, and 0.5 units of Taq polymerase (New England Biolabs). PCR conditions were 95 °C for 3 min, followed by 26 cycles of 30 s at 95 °C, 30 s at 55 °C, and 40 s at 72 °C. *Actin 8*

was used as the endogenous control. The number of PCR cycles for each primer pair was optimized by determining the number of cycles at which the amount of *Actin 8* transcript was in an exponential amplification stage. The identity of all PCR products was verified by sequencing in both strands. Semiquantitative RT-PCR analysis of *CYP82G1* and *GES* transcripts in mock- or alamethicin-treated wild-type and transgenic plants was performed as described above but with an extension time of 1.5 min per cycle and a total of 30 cycles, using primers in Table S3. Quantitative PCR analysis of *CYP82G1* and *GES* transcript levels in treated *Arabidopsis* leaves was performed with an Applied Biosystems 7300 Real-Time PCR System. Concentrations of cDNAs and primers were first optimized according to the instrument user's manual (Applied Biosystems). Single-band amplification and the amplicon size for each of the primer pairs (sequence information in Table S3) were confirmed by agarose gel electrophoresis. Quantitative RT-PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems) with 1 µL of 30-fold diluted cDNA and 0.7-µM primers in a 25-µL reaction volume. PCR conditions were as follows: 10 min initial denaturation at 94 °C followed by 40 cycles of 94 °C for 15 s and 60 °C for 1 min. Protein phosphatase 2A (*PP2A*, At1g13320) was used to normalize transcript levels according to Czechowski et al. (see Table S3 for primer sequences) (2). The ΔC_T (Δ -threshold cycle) of *CYP82G1* was calculated relative to the *PP2A* gene C_T averaging three technical replicates. Fold differences were calculated with the $2^{-\Delta\Delta C_T}$ formula by comparing to the lowest value (3).

Genetic Complementation Analysis. The coding region of *CYP82G1* (1,548 bp) was amplified by RT-PCR from total RNA using primers A and B (Table S3). RT-PCR conditions were the same as those described for selected gene candidates (see *Transcript Analysis of CYP82G1, GES, and Selected Candidate Genes*) except for an extension time of 1.5 min per cycle and a total of 30 cycles. The *CYP82G1* amplicon was subcloned into the pENTR/D-TOPO vector, generating pENTR-*CYP82G1*, and transferred to the binary destination vector pK7WG2 (4) by an LR recombination reaction (Invitrogen). A heme-domain truncated *CYP82G1* gene fragment (1,121 bp) was amplified with the primer pair A and G (Table S3) and cloned into pK7WG2 using the same recombination method.

Yeast Expression and Enzyme Assay. The *CYP82G1* cDNA was transferred from pENTR-*CYP82G1* via an LR recombination reaction to the YEp352 destination vector. The construct was transformed into the WAT11 yeast strain (5) and transformed cells were selected on SGI selection medium (6). Recombinant *CYP82G1* was expressed by culturing transformed yeast for 24 h at 30 °C and 200 rpm in 5 mL SGI (+trp/+ade) liquid medium. To measure *CYP82G1* enzyme activity in vivo, 100 µL of the culture were then transferred to a sterile 20-mL PTFE/Silicon Septa screw cap glass vial containing 5 mL of YPAD medium (10 g/L yeast extract, 10 g/L peptone, 2% (wt/vol) glucose, and 200 mg/L adenine). After 24 h of further culturing, 10 µM of substrate (Table S2) was individually added and the culture incubated for another 4 h. The enzymatic reaction was terminated by adding HCl to a final concentration of 0.05 N with a syringe needle penetrating the septum of the cap. Volatiles in the culture headspace were then collected and analyzed by automated SPME-GC-MS as described under *Volatile Analysis*. Yeast cells carrying the empty YEp352 vector were used as a control.

For microsomal preparation, 5 mL of the transformed WAT11 starter culture were transferred into 500 mL of YPAD medium and *CYP82G1* protein was expressed for 24 h at 30 °C. Microsomes

were prepared as described previously except using a Bead beater HBB908 (BioSpec Products, Inc.) (7). The amount of properly folded P450 protein was estimated by CO-induced difference spectra ($A_{450} - A_{490}$) according to the method of Omura and Sato (8).

For in vitro enzyme assays, yeast microsomes were extracted from the same yeast line as described above (7). Enzyme assays were performed in a total reaction volume of 1 mL of 100 mM Tris-HCl (pH 7.0) in a 10-mL PTFE/Silicon Septa screw cap glass vial. Microsomal protein (110–140 pmol of CYP82G1 calculated by CO-difference spectroscopy) was first incubated for 5 min with 0.1 to 41.1 μ M of (*E,E*)-geranylinalool or 0.2 to 19.7 μ M of (*E*)-nerolidol (dissolved in DMSO). Reactions were initiated by the addition of 1 mM NADPH and then incubated for 20 min at 30 °C. Assays were terminated by acidification and immediately analyzed by automated SPME/GC/MS as described above. Quantification of homoterpene products was performed in the total ion mode.

Calibration of the SPME-based in vivo and in vitro assays for C_{16} -homoterpene 4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT) and C_{11} -homoterpene 4,8-dimethyl-1,3,7-nonatriene (DMNT) was achieved by simulating the assay without substrate in the presence of the yeast culture or with microsomes extracted from empty vector-expressing yeast cells. Incubation was carried out in the presence of homoterpene standards before acidification. Volatiles were then analyzed as described above. Linear calibration was performed between 1 and 100 ng (for TMTT, $R^2 = 0.9857$ and for DMNT, $R^2 = 0.9988$).

Comparative Modeling of CYP82G1. The translated peptide sequence of CYP82G1 was retrieved from the TAIR database (accession no: 2090275), and the first 25 amino acids, corresponding to the membrane binding peptide, were deleted from the original sequence (final sequence ranging from amino acid 26 to 515). The BLAST server (9) was used to search for potential templates for CYP82G1 in the RCSB Protein Data Bank. Because of the lack of an appropriate plant P450 structure in the database, four mammalian (human) P450 structures with highest resolution and sequence identity ($\sim 25\%$ identity, $\sim 42\%$ similarity) to CYP82G1 were selected for homology modeling: P450 2D6 (PDB: 2F9Q) (10), 2R1 (PDB: 3CZH) (11), 2E1 (PDB: 3E6I) (12), and 46A1 (PDB: 2Q9F) (13). These enzymes catalyze hydroxylation reactions of a variety of different substrates with narrow specificity (cholesterol: 2Q9F; vitamin D: 3CZH) and broad specificity (metabolism of xenobiotics, oxidation of fatty acids: 2F9Q, 3E6I). Similar to CYP82G1, several of the selected P450s bind hydrophobic substrates in hydrophobic active-site cavities. The selected P450s were used for single and multiple template comparative modeling using Modeler 9v7. Two scoring functions, DOPE (discrete optimized protein energy) score (14) and GA341 (a combined statistical potential z-score of the model) score (15) were applied to select the final model out of 500 independent modelings. Based on these scores, the model generated by a combination of four templates instead of single templates was selected. The loop from residues 270 to 279 was further refined by 100 independent simu-

lations of ab-initio refinement (16) to reduce the DOPE score in that region. To position the heme group into the active site of the model, a modified alignment was used corresponding to a model based on the heme-binding loop of 2F9Q_A and the equivalent loop 120 to 131 of CYP82G1, which share $\approx 70\%$ identity (17). The quality of the final model was verified using PROCHECK (18) and the internal energy of the model was checked using Anolea (19).

Determination of Charge Distribution of Heme. The atomic partial charges on the heme atoms were determined using Gaussian03cv2 (20). The geometrical parameters of heme were taken from the CYP82G1 homolog structures, where the heme group is totally planar and in its five-coordinate state. The Fe atom was modeled as a +3 ion in a high-spin state ($S = 5/2$), the experimentally observed state for the binary complex of most P450s. QM(B3LYP) single point energy at the DFT level of theory, in the ground state with the LANL2DZ basis set was used to calculate Mulliken charges for heme.

Docking of Molecules into the Active Site of CYP82G1. Autodock4 (21, 22) was used to find possible binding positions for ligands on CYP82G1. The 3D structures of the compounds are from the ZINC databases (23). The molecules that were docked are listed in Table S4. The selected ligands were fully flexible during docking and their Gasteiger partial charges (24) and torsion angles were determined by AutoDock Tools 1.5.4. The protonation state of amino acids in the 3D model of CYP82G1 was predicted by H^{++} (25), and hydrogen atoms were added to the enzyme and ligands using Chimera (26). The center of the grid box was placed in the center of active site, which was determined based on the structural alignment of modeled CYP82G1 to homologous structures in complex with their ligands. Grid maps with 0.375 Å spacing were created using AutoGrid in the AutoDock package. The Lamarckian genetic algorithm was used to determine the docked poses. At least 1,000 runs were performed for each docked ligand. The population size was 250 and the maximal number of energy evaluations and generations was 5×10^7 and 30,000, respectively. The rate of gene mutation was 0.02 and the rate of gene crossover was 0.8. The local search component involved 300 iterations per local Solis and Wets search (27). The most favored docked structure for each compound was selected based on the lowest binding energy, largest number of docked poses in a cluster, using a clustering tolerance of 2.0 Å root mean square deviation, and expected orientation of ligands within the active site for catalysis.

Sequence Alignment and Phylogenetic Analysis. Amino acid sequence alignment of plant CYP82 proteins was conducted with ClustalW (Lasergene 8) and exported as a Nexus file. Phylogenetic analysis of the data set was conducted using maximum parsimony in PAUP* (28). Support for the clades was obtained by performing bootstrap (29) searches with 1,000 replicates and 10 random sequence replicates. Trees were compiled using Tree-Graph2 (30).

- Herde M, et al. (2008) Identification and regulation of TPS04/GES, an *Arabidopsis* geranylinalool synthase catalyzing the first step in the formation of the insect-induced volatile C_{16} -homoterpene TMTT. *Plant Cell* 20:1152–1168.
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol* 139:5–17.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25:402–408.
- Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7:193–195.
- Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *J Biol Chem* 272:19176–19186.
- Takahashi S, et al. (2007) Metabolic engineering of sesquiterpene metabolism in yeast. *Biotechnol Bioeng* 97:170–181.
- Ralston L, et al. (2001) Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from tobacco (*Nicotiana tabacum*). *Arch Biochem Biophys* 393:222–235.
- Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J Biol Chem* 239:2370–2378.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Rowland P, et al. (2006) Crystal structure of human cytochrome P450 2D6. *J Biol Chem* 281:7614–7622.
- Strushkevich N, Usanov SA, Plotnikov AN, Jones G, Park H-W (2008) Structural analysis of CYP2R1 in complex with vitamin D3. *J Mol Biol* 380:95–106.
- Porubsky PR, Meneely KM, Scott EE (2008) Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. *J Biol Chem* 283:33698–33707.
- Mast N, et al. (2008) Crystal structures of substrate-bound and substrate-free cytochrome P450 46A1, the principal cholesterol hydroxylase in the brain. *Proc Natl Acad Sci USA* 105:9546–9551.
- Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524.
- Melo F, Sali A (2007) Fold assessment for comparative protein structure modeling. *Protein Sci* 16:2412–2426.

16. Fiser A, Do RKG, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753–1773.
17. Marti-Renom MA, et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291–325.
18. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12:345–364.
19. Melo FDD, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: A www server to assess protein structures. *Proc Int Conf Intell Syst Mol Biol* 5:187–190.
20. Frisch MJ, et al. (2004) *Gaussian 03, Revision C.02* (Gaussian Inc., Wallingford, CT).
21. Morris GM, et al. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662.
22. Morris GM, et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791.
23. Irwin JJ, Shoichet BK (2005) ZINC—A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182.
24. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36:3219–3228.
25. Gordon JC, et al. (2005) H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33(Web Server issue):W368–W371.
26. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
27. Solis FJ, Wets RJB (1981) Minimization by random search techniques. *Math Oper Res* 6: 19–30.
28. Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.* (Sinauer Associates, Sunderland, MA).
29. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evol* 39:783–791.
30. Stöver BC, Müller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11:7.

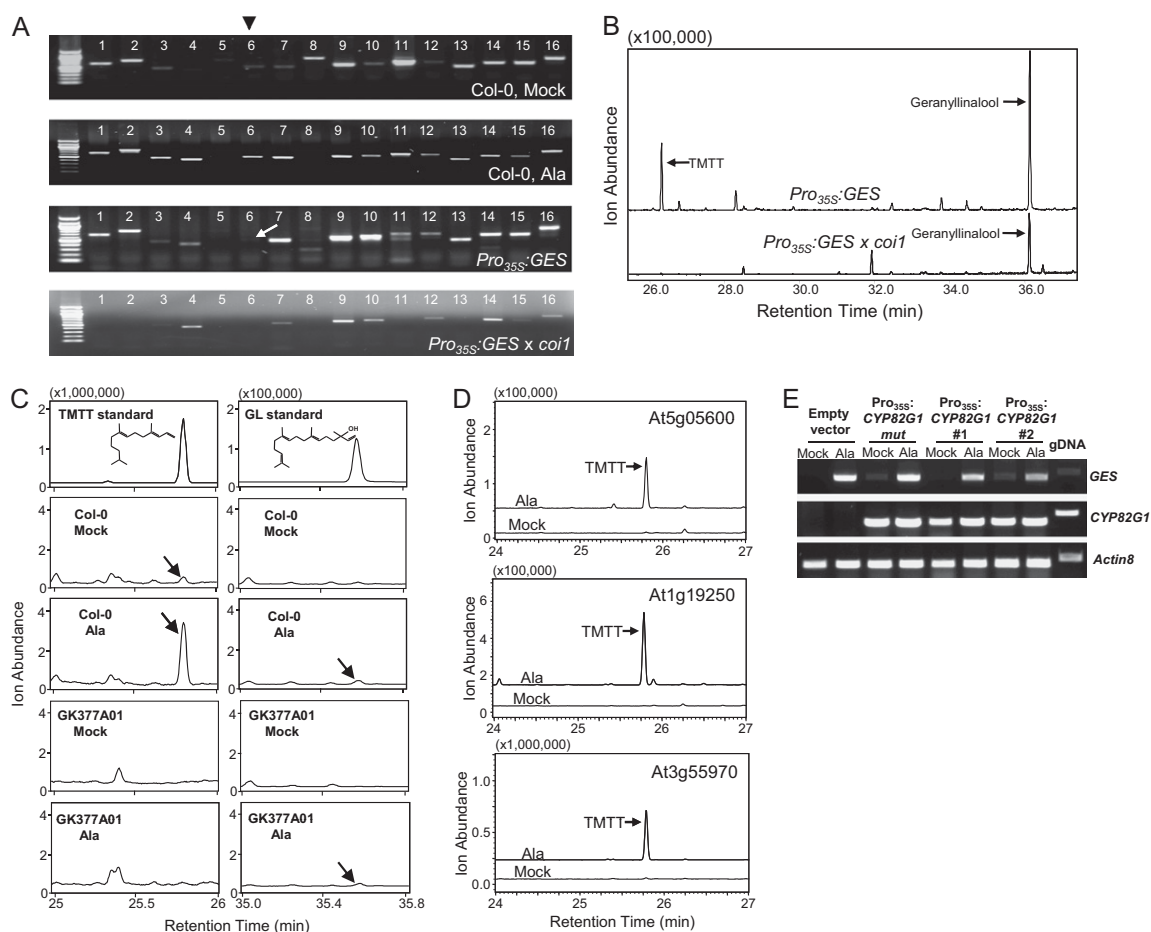
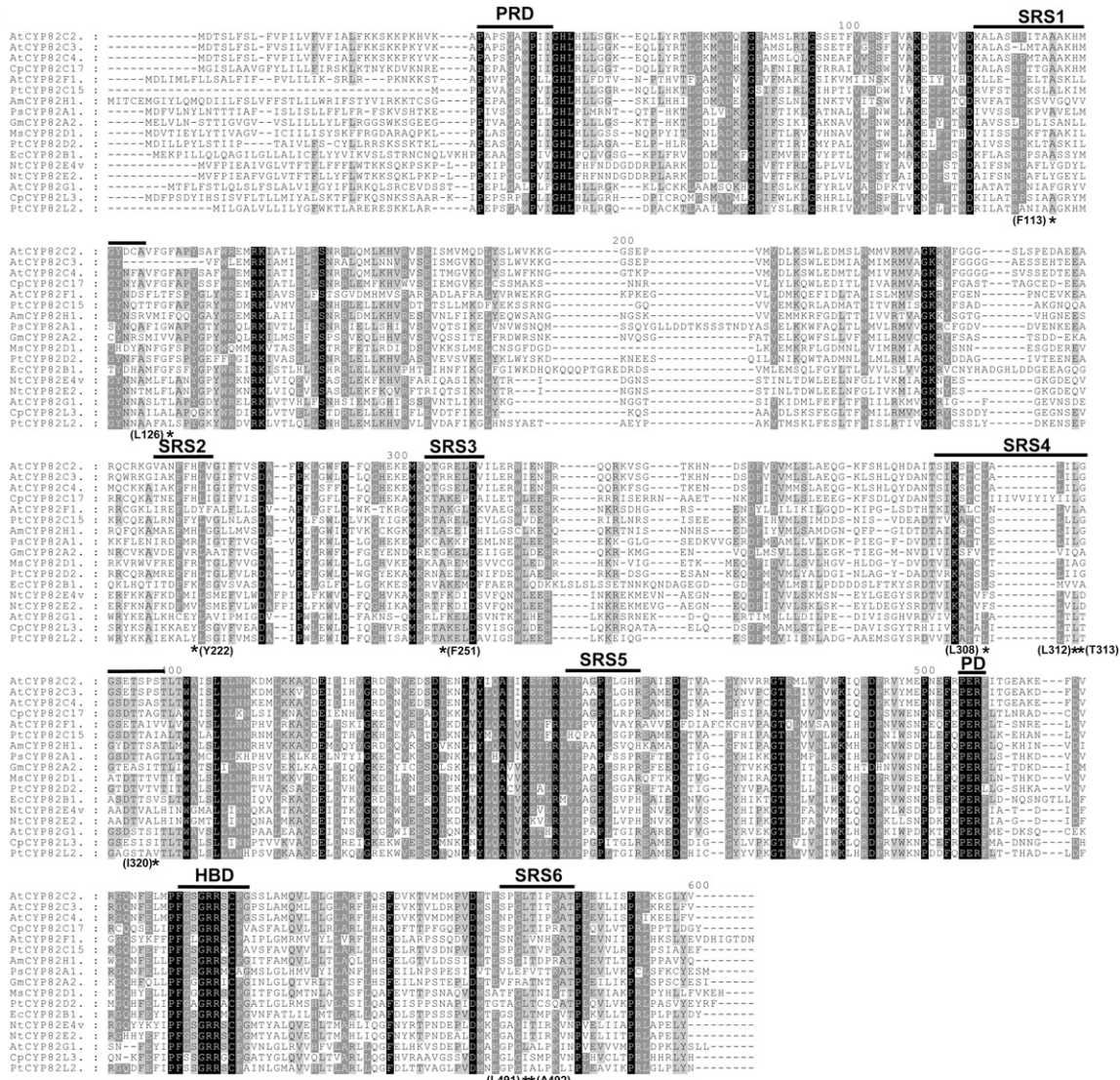


Fig. S1. Screening of gene candidates and analysis of the corresponding gene-knockout plant lines and the complemented *CYP82G1* gene-knockout line. (A) Semiquantitative RT-PCR analysis of the gene candidates shown in Table S1. Transcript analysis was performed in leaves of Col-0 wild-type in response to mock- and alamethicin (Ala)-treatment. Transcripts were also analyzed in untreated transgenic plants constitutively expressing *GES* in the background of the *GES* knockout line *salk_039864* (*Pro_{35S}:GES*) (1) and in *Pro_{35S}:GES* plants crossed into the *coi1* mutant background. Transcripts of gene *CYP82G1* (At3g25180) are shown in lane 6 (black arrowhead). The white arrow indicates a weak *CYP82G1* amplicon. (B) Dependency of TMTT formation on CO11. Selected ion GC-MS chromatograms (*m/z* = 69) of TMTT and its precursor (*E,E*)-geranylinalool in plants constitutively expressing *GES* in the background of the *GES* knockout line *salk_039864* (*Pro_{35S}:GES*) (Upper) and in *Pro_{35S}:GES* plants crossed into the *coi1* mutant background (Lower). Results are representative for at least three individual lines. (C) Total ion GC-MS chromatograms of TMTT and geranylinalool (GL) emitted from mock- and alamethicin-treated leaves of Col-0 wild type and GK377A01 in comparison with TMTT and GL standards (Upper). Arrows indicate peaks for TMTT or GL. (D) Alamethicin-induced emission of TMTT from knockout mutants of other selected candidate genes, *salk_026163* (for At1g19250), *salk_114795* (for At3g55970), and *salk_073705* (for At5g05600) (Table S1). (E) RT-PCR analysis of *GES* and *CYP82G1* gene transcripts in mock- and Ala-treated rosette leaves of GK377A01 plants expressing *CYP82G1* under the control of the CaMV 35S promoter. Results from two independent lines (#1, 2) are shown in comparison with GK377A01 lines carrying the empty expression vector or an expression construct with a truncated *CYP82G1* gene lacking both the PERF and the heme-binding region (*Pro_{35S}:CYP82G1mut*). *Actin8* was used as an endogenous gene expression control.

A



B

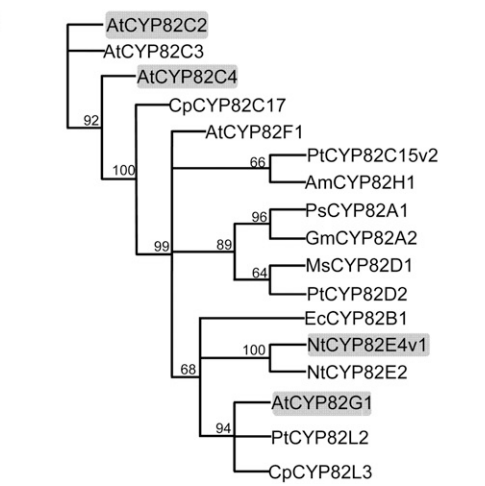


Fig. S2. (A) Alignment of deduced amino acid sequences of the CYP82 family in plant species using the ClustalW algorithm (<http://www.ch.embnet.org/>). The sequences were visualized and edited in GeneDoc 2.7 (<http://www.nrbc.org/gfx/genedoc>). Representative conserved domains among P450s were indicated as follows: PRD, proline-rich domain; PD, PERF domain; HBD, heme-binding domain. SRS1-6 indicate substrate recognition sites as predicted for *Arabidopsis* P450s (1). Residues with an asterisk represent the putative substrate-interacting residues identified in the AtCYP82G1 model. **(B)** Phylogenetic relationship of P450s of the plant CYP82 family. Enzymes with known biochemical function are highlighted in gray. Phylogenetic analysis was conducted using maximum parsimony in

Legend continued on following page

PAUP* 4.0 beta 10. Bootstrap values with 1,000 replications were calculated and only bootstrap values $\geq 60\%$ are shown. AtCYP82G1 showed highest sequence similarity to P450s from poplar and papaya (2, 3). At, *Arabidopsis thaliana*; Cp, *Carica papaya*; Pt, *Populus trichocarpa*; Am, *Ammi majus*; Ps, *Pisum sativum*; Gm, *Glycine max*; Ms, *Medicago sativa*; Ec, *Eschscholzia californica*; Nt, *Nicotiana tabacum*.

1. Rupasinghe S, Schuler M (2006) Homology modeling of plant cytochrome P450s. *Phytochem Rev* 5:473–505.
2. Ming R, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996.
3. Tuskan GA, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604.

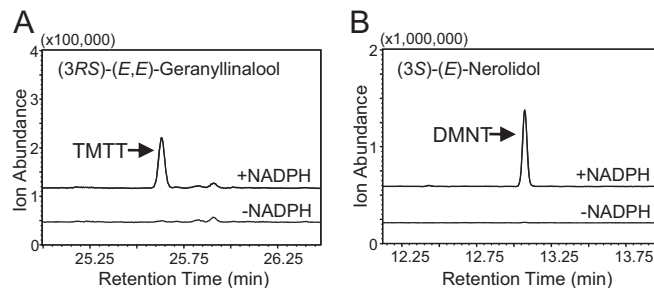


Fig. 53. In vitro microsomal enzyme assay of CYP82G1. GC-MS chromatograms of TMTT and DMNT produced by the recombinant CYP82G1 enzyme with (*E,E*)-geranylinalool (*A*) and (*E*)-nerolidol (*B*) as substrates in the presence and absence of NADPH.

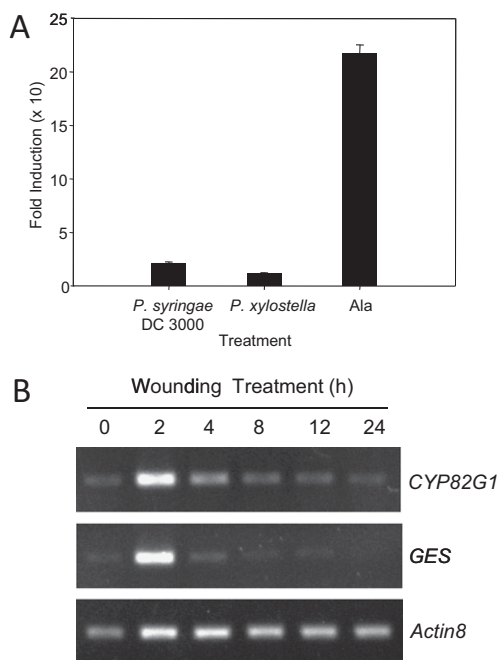
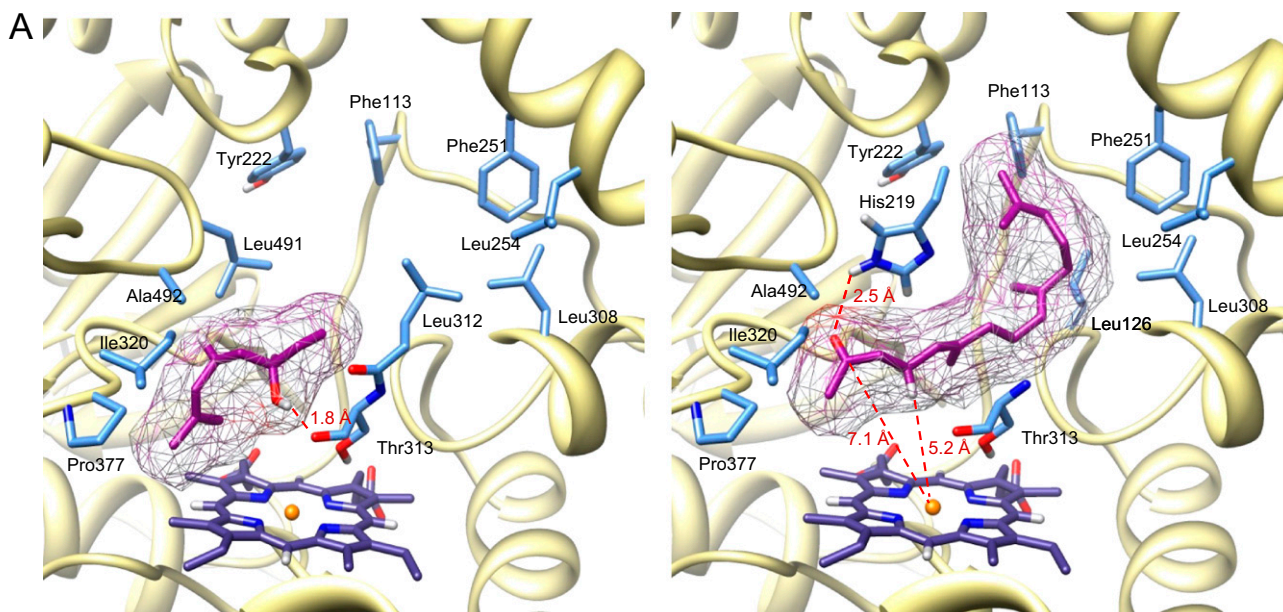


Fig. 54. (*A*) Real-time PCR of *CYP82G1* transcripts in response to biotic stress and elicitor-treatment. Induction of transcript levels is shown relative to those from mock-controls. Values were normalized to transcript levels of At1g13320 (2) and are means \pm SEM of three technical replicates. The experiments were repeated at least once with similar results. Treatments were as described under *Materials and Methods*. Ala, alamethicin. (*B*) *CYP82G1* and *GES* expression in mechanically wounded *Arabidopsis* leaves. Leaves were wounded as described under *Materials and Methods* and tissue from the wounded region was harvested at indicated time points. *CYP82G1* and *GES* transcripts were analyzed by semiquantitative RT-PCR. *Actin8* was used as an endogenous control.



B

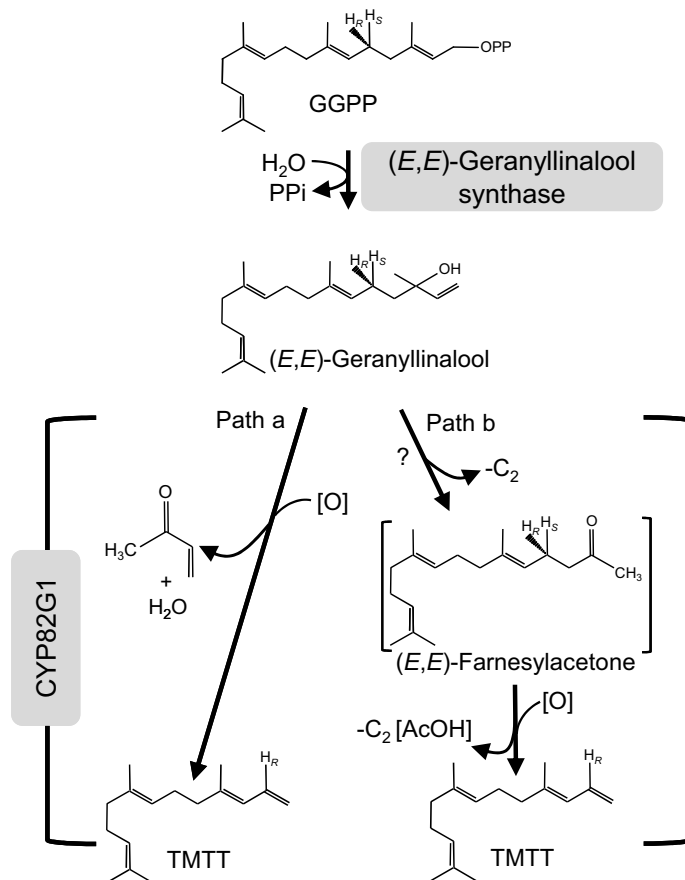


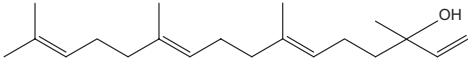
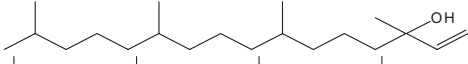
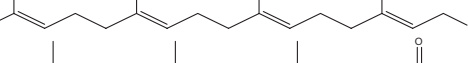
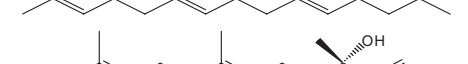
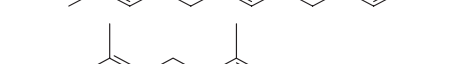

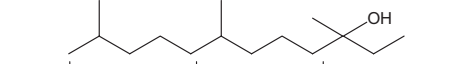
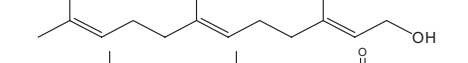
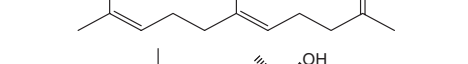
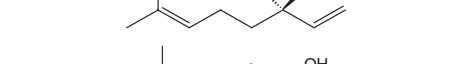
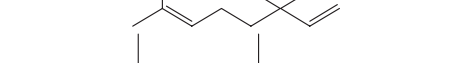
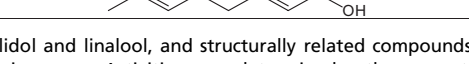
Fig. S5. (A) Predicted position of (*S*)-linalool (Left) and (*E,E*)-farnesylacetone (FA) (Right) in the active site of the CYP82G1 model. The main interacting residues are presented. The C-3 hydroxyl group of linalool forms a hydrogen bond with Thr313, although the alkyl moiety mainly interacts with Pro377 and Ala492 through hydrophobic interactions. The position of the alkyl chain of linalool is therefore not equivalent to that of (*E,E*)-geranylgeranyl alcohol and (*E*)-nerolidol (Fig. 5). Hydrogen atoms at C-5 are inaccessible to oxygen attack. FA was docked to CYP82G1 mainly through a hydrogen bond made by the carbonyl group to the protonated Ne of His219 causing a different position of the molecule in the binding site relative to (*E,E*)-geranylgeranyl alcohol. The carbonyl group of FA is in antiperiplanar orientation relative to the C-4 hydrogen and does not support a *syn*-periplanar orientation of the C4-hydrogen atom and the acyl moiety (1) required for rapid transfer of the Fe-bound hydroxyl group to the reactive acyl radical. (B) Scheme of putative reaction mechanisms for the conversion of (*E,E*)-geranylgeranyl alcohol to TMTT catalyzed by CYP82G1. Path a is a single-step fragmentation starting with the abstraction of the C-5 H_s atom followed by an immediate cleavage of the allylic radical intermediate to TMTT and but-1-en-3-one. Path b is a two-step fragmentation with a consecutive loss of two C₂ moieties and formation of a C₁₈ ketone intermediate. GGPP, geranylgeranyl diphosphate.

Table S1. Candidate genes coexpressed with *GES* and queried from publicly available coexpression databases

No.	AGI No.	Gene annotation	COR	No.	AGI No.	Gene annotation	COR
1	At1g06620	Putative 2-oxoglutarate-dependent dioxygenase	0.53	9	At4g03410	Peroxisomal membrane protein-related	0.61*
2	At3g55970	Oxidoreductase, 2OG-Fe(II) oxygenase family protein	0.62	10	At5g43450	Putative 2-oxoglutarate-dependent dioxygenase	0.63*
3	At1g19250	Flavin-dependent monooxygenase	0.42	11	At5g05600	Oxidoreductase, 2OG-Fe(II) oxygenase family protein	0.48
4	At5g05340	Putative peroxidase	0.39	12	At4g37370	CYP81D8	0.40
5	At2g30830	Putative 2-oxoglutarate-dependent dioxygenase	0.28	13	At2g38240	Oxidoreductase, 2OG-Fe(II) oxygenase family protein	0.50
6	At3g25180	CYP82G1	0.64	14	At4g02940	Oxidoreductase, 2OG-Fe(II) oxygenase family protein	0.37
7	At2g37770	Oxidoreductase	0.38	15	At1g14130	Putative 2-oxoglutarate-dependent dioxygenase	0.35
8	At5g63450	CYP94G1	0.53*	16	At5g36220	CYP81D1 or CYP91A1	0.50

Annotation and correlation coefficients (COR) were adopted from ATTED-II (<http://atted.jp/>) and the Expression Angler* (<http://bbc.botany.utoronto.ca/>). Genes are listed in the order as presented in Fig. S1A. *CYP82G1* showing the highest correlation coefficient is in boldface.

Table S2. Substrate specificity of the recombinant CYP82G1 enzyme

Substrate	Structure	Relative activity
C ₂₀ (3 <i>RS</i>)-(<i>E,E</i>)-geranylinalool		1
Isophytol		N.D.
(<i>E,E,E</i>)-geranylgeraniol		0.02 ± 0.0
C ₁₈ (<i>E,E</i>)-farnesylacetone		0.01 ± 0.0
C ₁₅ (3 <i>S</i>)-(<i>E</i>)-nerolidol		5.3 ± 0.4
(<i>Z</i>)-nerolidol		N.D.
3,7,11-Trimethyl-3-dodecanol		N.D.
(<i>E,E</i>)-farnesol		Trace
C ₁₃ (<i>E</i>)-geranylacetone		0.02 ± 0.0
C ₁₀ (<i>R</i>)-(-)-linalool		N.D.
(±)-Linalool		N.D.
(<i>E</i>)-geraniol		N.D.

(*E,E*)-geranylinalool, its C₁₅- and C₁₀-analogs, (*E*)-nerolidol and linalool, and structurally related compounds, as well as proposed ketone intermediates were tested. Results are shown for yeast *in vivo* assays. Activities were determined as the amount of homoterpene produced in 4 h from 5-mL culture in the presence of 10 μM substrate. Activities are shown relative to that with geranylinalool as substrate. Unless indicated otherwise, chiral compounds are assumed to be racemic mixtures. N.D., no enzyme activity was detected with 10 or 50 μM substrate.

Table S3. List of primers used in this study

Name	Sequence (5'-3')	Name	Sequence (5'-3')
Primers used in plasmid construction		Primers used in RT-PCR assays (cont'd)	
Primer A	<u>CACCATGACTTTTCTCTTAGTACTCTCCAG</u>	At4g03410-R	AGTAATAGTGAGACAGTGAACCATGG
Primer B	<u>TTAGAGTAAACTATACAACCTCGGGTCG</u>	At5g43450-F	TCTTCAAGAAACGTTGTCATTTCTGAAG
Primer C	<u>CACCAGCTTAGTGGGTTTTAACTTCAGC</u>	At5g43450-R	AATCACCAATGTTGACAACAAGAGC
Primer D	<u>GGTGATCACTTTGGGTTAATGTATGGGA</u>	At5g05600-F	AGGATCGATGGATGAGTGGCC
Primer E	<u>TTAAGTCGGTGCCTTCTTAACAATAG</u>	At5g05600-R	TGTGGAGAGAGACCAAGAGCCAG
Primers used in RT-PCR assays		At4g37370-F	AAGCCAAATCTACCTCCGAGTC
At3g25180-F	ATTGGTTTCGGTGAAGTGAATAGCG	At4g37370-F	TTGCAAGAACTCGTCGAGCC
At3g25180-R	TTAAGTCGGTGCCTTCTTAACAATAG	At2g38240-F	TAACAGAGACGTTGTCAGAGAGTTAGG
At1g06620-F	TCACAGAGATTCCTCGATATTCGTG	At2g38240-R	ACTTGGTTTTTCCACAAGGACC
At1g06620-R	TGGTCATGAAGAACTGAAGCCC	At4g02940-F	AAGAGAGAGCTGATTCAGCTTGGT
At3g55970-F	AAGACTGGCTGAGCCTATAGTCC	At4g02940-R	ATTGGTTGAACCGGTGGTGG
At3g55970-R	TAACGATGAAAGCATGAGGAGCTGG	At1g14130-F	TGATAAGATTCTGAATCAGAAAATCCGTG
At1g19250-F	TCGAGAGTAGCCATCATCGGTG	At1g14130-R	ACAATCTTCGGTTGCTCCATATCG
At1g19250-R	TTGGAACGTCGCCGATTTTC	At5g36220-F	ACTGTTGTGGAATTGAAACCAATGC
At5g05340-F	ACAACCTACGACTAACCTTCACTCAACCTC	At5g36220-R	ATTGTATCAATGACCCGAGAGCTAG
At5g05340-R	GCCGGAATGTTACTATTCGCTG	Actin8-F	ATGAAGATTAAGTCTGGTGGCAC
At2g30830-F	GCGGGCAACTATGATCGTG	Actin8-R	GTTTTATCCGAGTTGAAGAGGC
At2g30830-R	GATCTCCAAGTTAACGATTAGAGC	Primers used in real-time PCR	
At2g37770-F	TCATGACCCTCAAGATGTCCCG	CYP82G1-F	CTGATGAACCACTGGATATGGCT
At2g37770-R	GCACACTGTGACCCATTTGGAG	CYP82G1-R	GAGTAAACTATACAACCTCGGGTCCG
At5g63450-F	AACCCTAAAAGCCAAAACCCGA	GES-F	GATAGCGAACCAACGAGGAT
At5g63450-R	CAAAAACAGCCACGTCATCGC	GES-R	CTTGTGTTGTAGCACTTCAGAAA
At4g03410-F	ATGGCAGCTCTCTGTTGTTGTC	PP2A-F	TAACGTGGCCAAAATGATGC
		PP2A-R	GTTCTCCACAACCGCTTGGT

The CACC addition at the 5' end of primers for directional TOPO-cloning is underlined.

Table S4. Binding of true and selected nonfavorable substrates to the active site of CYP82G1 as analyzed by Autodock4.1

Molecule	Source	Torsion	Grid box size	Binding energy (kcal/mol)	K _d (μM)
(3S)-(6E)-(10E)-geranylinalool	ZINC: 4262175	11	50 × 50 × 50	-7.58	3.09
(3R)-(6E)-(10E)-geranylinalool	ZINC: 2522792	11	50 × 50 × 50	-7.48	3.39
(6E)-(10E)-farnesylacetone	ZINC: 1235887	9	48 × 48 × 48	-7.13	5.95
(3S)-(6E)-nerolidol	ZINC: 1351550	8	40 × 40 × 40	-6.09	34.09
(3R)-(6E)-nerolidol	ZINC: 2040970	8	40 × 40 × 40	-6.05	36.72
(3S)-linalool	ZINC: 1529819	5	30 × 30 × 30	-4.59	433.73
(3R)-linalool	ZINC: 1529820	5	30 × 30 × 30	-4.59	431.59

The 3D structures of the compounds are from the ZINC database (1). The estimated binding energies and binding constants refer to the selected docked structures. (*E,E*)-Geranylinalool is predicted to bind more favorably than (*E*)-nerolidol by about 1.5 kcal/mol because of additional hydrophobic interactions involving the alkyl tail (C12–C17) with Leu126 and Leu308 and formation of closer interactions to the side chains of Phe113 and Phe251 (Fig. 5). This difference in binding energy was not clearly supported by determination of apparent *K_m* values for both substrates (Table 1), reflecting the fact that factors other than substrate binding contribute to catalysis. The (3S)-enantiomers of (*E,E*)-geranylinalool and (*E*)-nerolidol displayed slightly more favorable binding energies relative to their *R*-configurations with the difference appearing to involve space restraints of C1 and C2 in the *R*-configuration in close proximity to the Ala378 side chain. Although the difference in docking energies in the model is not large enough to firmly support a preferred conversion of *S*-enantiomers, again reflecting the complexity of the catalytic process, the findings support previous observations for a primary conversion of the *S*-enantiomer of (*E*)-nerolidol in leaves and flowers of different angiosperms (2). Docking of (*R*)- and (*S*)-linalool occurred with the least favorable binding energy among the compounds studied.

1. Irwin JJ, Shoichet BK (2005) ZINC—A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182.

2. Gäbler A, Boland W, Preiss U, Simon H (1991) Stereochemical studies on homoterpene biosynthesis in higher plants; mechanistic, phylogenetic, and ecological aspects. *Helv Chim Acta* 74:1773–1789.