Supporting Information

Schlüter et al. 10.1073/pnas.1013313108

SI Methods

Details of Expression Analysis and RACE Experiments. RT-PCRs were set up as described in the main text, with a typical cycling program as follows: 95 °C 3 min, 37 × (95 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min 30 s), and 72 °C for 10 min. *Pfu* DNA polymerase (Promega) was added to all reactions to be sequenced or re-cloned at a final concentration of 0.015 unit/μL.

5′ and 3′ RACE were carried out by using GeneRacer kit (Invitrogen) and as described (1, 2), except for the use of a different RNA oligonucleotide (Table S6; a gift of S. Schauer, University of Zürich) for 5′ RACE. For RACE, PCRs were carried out with JumpStart REDAccuTaq LA DNA Polymerase (Sigma) or Advantage GC 2 Polymerase (Clontech) and a PCR extension step at 68 °C. RACE was complemented by PCR walking on genomic DNA extracted with Nucleon PhytoPure (GE Healthcare) by using the DNA Walking SpeedUp Kit (Seegene). Products were cloned into pDRIVE (Qiagen) and sequenced. In addition to the *SAD* genes, a full-length *Ophrys G3PDH* cDNA (putatively encoding a glycolytic glyceraldehyde 3-phosphate dehydrogenase) was isolated as a control.

Locus specificity of amplification was initially determined by gradient PCRs (as outlined above) by using 1.5 ng/μL *SAD* plasmid as a template and locus-specific full-length primers (Table S6). *SAD2*-specific primers did not result in bands visible on agarose gel from *SAD1* plasmid at annealing temperatures (T_A) of 60 °C or higher. Conversely, *SAD1*-specific primers did not amplify *SAD2* at $T_A \geq 63$ °C annealing temperature. Because using the same high stringency of PCR conditions reduced the amplification from cDNA, T_A was lowered to 58 °C in RT-PCR.

For analysis of tissue- and stage-specific expression levels, RT-PCR was carried out at $T_A = 58$ °C and *att*B-containing primers allowing recombination cloning of PCR products. This experiment was conducted for 80 tissue samples (different floral developmental stages from flowers and buds at different positions relative to the first open flower; labella, sepal and petal, and green leaves or bracts) from 6 *O. sphegodes* and 4 *O. exaltata* individuals, for which GC data from the same tissues were available. For every plant individual in the experiment, PCR products from open flowers were cloned into pDONR221 by BP recombination (Invitrogen), and 5–10 clones were sequenced completely. All clones were found to be from the correct *SAD* locus. To measure expression level, RT-PCR from 0.8 ng/μL template was carried out independently at two cycle numbers for each gene investigated (29 and 33, except for *G3PDH*, where 29 and 31 cycles were used). For *G3PDH*, the primer combination used (1f/2r; Table S6) amplifies a larger, intron-containing region from genomic DNA and, thus, served as a control for genomic DNA contamination in RT-PCR experiments (appropriate controls were included). Five microliters of each PCR product were loaded on 0.8% TAE (Tris-acetate-EDTA)-agarose gel containing 0.1 μg/mL ethidium bromide and electrophoresed for 40 min at 95 V. Loading ensured that PCR products of different genes of the same biological samples were loaded beside each other. Digital photographs of gels were taken under UV at three constant exposure settings for each experiment and band intensity was quantified by using Image J 1.43u software (3). *SAD* expression levels were normalized against *G3PDH* and measurements averaged for each sample.

Statistical Analysis of Hydrocarbon and Expression Data. Because hydrocarbons were analyzed from buds of different developmental stages, use of comparable amounts of tissue could not be ensured. Hydrocarbons were thus analyzed as relative alkane and alkene amounts (as percentages of the total amount of alkanes and alkenes per sample). Data analysis was performed in R 2.11.0 (4). Among-species hydrocarbon differences were tested in a Wilcoxon signed-rank test. The pairwise correlation of hydrocarbons was investigated by clustering based on a correlation distance (pvclust method) (5). For this cluster analysis only, additional hydrocarbon data were added ($n = 62+96$ additional data points from 7 *O. exaltata* and 12 *O. sphegodes* individuals, respectively, for which no RNA was available). The effect of *SAD* expression on relative alkane and alkene proportions (arcsine square-root transformed) was assessed in a generalized linear model (GLM) and a linear mixed-effect model (LME). The GLM used a Gaussian error distribution and modeled hydrocarbon levels with the explanatory variables species, tissue type, and *SAD1*, *SAD2,* and *SAD3* expression. The LME used the same explanatory variables and added biological individual as a random-effect factor in a maximum likelihood model. The models were simplified by using stepwise factor removal by AIC (Akaike information criterion; stepAIC method).

Characterization of *Arabidopsis* T-DNA Insertion Line SALK_036854. *Arabidopsis* line SALK_036854 (6) carries a T-DNA insertion in the *SSI2*/*FAB2* (*SUPPRESSOR OF SA-INSENSITIVITY2*/*FATTY ACID BIOSYNTHESIS 2*) locus (At2g43710). The position of the insertion was determined by sequencing a PCR product amplified with primers SSI2Dn1R and LB (Table S6) and was found to be between nucleotide positions 6 and 7 of the second intron of At2g43710. The insertion line displayed a recessive dwarf phenotype similar to other loss-of-function *ssi2* alleles (7–9). These phenotypic dwarves exhibited a poor transformation rate and a high mortality rate. Genotyping of plants was done by PCR using primer combinations SSI2Up1F/SSI2Dn1R (wild-type allele) and LB/SSI2Dn1R (T-DNA allele), or using the combination SSI2Up2F/LB/SSI2Dn1R for concomitant amplification of both alleles. Further, for confirmation of transgenic plants carrying orchid *SAD*, primer combinations B1-SAD1f/B2-SAD1r (*SAD1*) or B1-SAD2f/B2-SAD2r (*SAD2*) were used on genomic DNA (for genotyping) or cDNA (for confirmation of expression). For RT-PCR, an intron-containing region of *ACT1* (*ACTIN 1*; At2g37620; primers ACT1f/ACT1r) was used as a control.

Orchid Desaturase Constructs for Bacterial Expression. Three sets of desaturase constructs (pRSET-*SAD*, pET9d-MAX-*SAD*, and pET9d-RcX-*SAD* constructs) were made and evaluated for protein expression and feasibility of desaturase assays, as described below:

pRSET-GW and pRSET*-SAD *constructs. The plasmid pRSET-B (Invitrogen) for bacterial protein expression with an N-terminal His_6-tag was modified by replacing the multicloning site by a Gateway cassette. This replacement was done by digesting the plasmid with *Bam*HI and *Eco*RI, blunting by T4 DNA polymerase treatment (10), ligation of Gateway conversion cassette C.1 (Invitrogen), and transformation into *Escherichia coli* DB3.1 (Invitrogen), followed by selection on 25 μg/mL chloramphenicol and 50 μg/mL ampicillin. Plasmids were isolated and their sequence confirmed, selecting only plasmids in the correct orientation, and the resulting plasmids were called pRSET-GW. pRSET-GW was functionally validated by recombination with a pENTR-*GUS* control plasmid (Invitrogen). pRSET-*SAD* constructs were made by LR recombination among pRSET-GW and pENTR207-*SAD* constructs (see main text), followed by transformation into *E. coli* TOP10 (Invitrogen) and selection on 50 μg/mL ampicillin. All plasmids

were confirmed by sequencing. pRSET-*SS1* (*OsSAD1*), pRSET-*SS2* (*OsSAD2*), and pRSET-*ES2* (*OeSAD2*) constructs were transformed into *E. coli* BL21 (DE3) cells (Invitrogen), their sequences were validated, proteins were expressed as recommended by Invitrogen, and were purified by using His SpinTrap columns (GE Healthcare) under nondenaturing conditions as recommended by the manufacturer. Because no desaturase activity was detected in initial enzyme assays (as described in the main text, but using unlabeled fatty acids from Sigma-Aldrich, and *E. coli* ACP and acyl-ACP synthetase from Invitrogen), we reasoned that the presence of additional amino acids at the N terminus (tag and possibly transit peptide) might adversely affect enzyme activity. Therefore, additional constructs were made and evaluated.

pET9d-MAX*-SAD *constructs. Untagged bacterial expression constructs in pET9d (Novagen) were made deleting amino acid residues 2–5 (ELHL) to remove part of the putative chloroplast transit peptide from the *Ophrys* desaturases. *OsSAD1* (*SS1*), *OsSAD2* (*SS2*), and *OeSAD2* (*ES2*) were reamplified from pENTR207 clones by PCR, introducing an *Nco*I site before the translation start site at the 5′ end and a *Bgl*II site after the stop codon at the 3′ end of the gene, using PCR primers NcoI-MA-SS1f/xS2f (forward) and BglII-SS1r/SS2r/ES2r (reverse; Table S6). PCR products were purified, cloned into pJET1.2 (Fermentas), chemically transformed into *E. coli* DH5α-T1^R (Invitrogen), and transformants were selected by using 100 μg/mL ampicillin. Colonies were screened for positive clones by PCR using the primers recommended by the supplier, and plasmids were extracted by alkaline lysis and sequenced (10). Because the plasmids still contained an internal undesired *Nco*I site, this site was removed by site-directed mutagenesis (10), using oligonucleotides SAD1-XNcoI-f/r and SAD2-XNcoI-f/r, Jumpstart Red AccuTaq LA DNA polymerase mix (Simga-Aldrich), and *Dpn*I digestion to select against methylated source plasmids. Resulting plasmids were transformed into DH5α cells, and colonies were screened by PCR and digestion with *Nco*I. The desaturase sequences were transferred to pET9d by digestion with *Nco*I and *Bgl*II and ligation into *Nco*I/*Bam*HI-digested pET9d, followed by transformation into DH5α cells, selection on 50 μg/mL kanamycin, colony PCR using T7 primers, plasmid extraction, and sequence verification, to yield pET9d-MAX-*SS1*, pET9d-MAX-*SS2,* and pET9d-MAX-*ES2*.

pET9d-RcX*-SAD *constructs. Additional constructs were made for *Ophrys SAD2* to remove the putative chloroplast transit peptide completely and replace the N-terminal end of the *Ophrys* desaturases, up to half of the first predicted α-helix, with the corresponding sequence from the *Ricinus communis* Δ^9-SAD enzyme. The first amino acid residue of *SS2*/*ES2* sequence in these constructs was Trp52. The front end of the *Ricinus* enzyme was PCR amplified from a pET9d plasmid containing the mature *Ricinus* wild-type 18:0-ACP Δ^9 desaturase (11) by using primers NcoI-MastL-RcSADf and a chimeric RcSAD/*Ophrys* SAD2 primer NRcOSAD2r carrying a *Bsp*TI site. The purified PCR product was cloned into pDRIVE, transformed into DH5α cells, and selected on 50 μg/mL kanamycin. Resulting plasmids were mixed with pET9d-MAX-*SAD* constructs, digested with *Nco*I and *Bsp*TI, ligated and transformed into DH5α cells, selected on kanamycin, and transformants were screened for the correct insert by colony PCR. The resulting plasmids pET9d-RcX-*SS2* and pET9d-RcX-*ES2* were confirmed by sequencing.

pET9d-MAX-*SAD* and pET9d-RcX-*SAD* constructs were transformed into *E. coli* BL21 (DE3) cells and assayed for protein expression under different conditions. For OsSAD1 and OsSAD2 desaturase assays, pET9d-MAX-*SAD* cultures were induced with IPTG for 4 h at room temperature, and protein was extracted and enriched to >90% purity by 20CM cation exchange chromatography (Applied Biosystems) as described (12). Soluble expression of OeSAD2 could not be attained.

Identification of desaturase homologs in public sequence databases. *Arabidopsis* SSI2 protein sequence (NP_850400) was chosen as a seed for retrieval of homologous sequences from public sequence databases by using BLAST searches (13). SSI2 protein sequence was first compared against the National Center for Biotechnology Information (NCBI) NR database (www.ncbi.nlm.nih.gov) by BLASTP (BLAST+ executables from the NCBI C++ Toolkit; www.ncbi.nlm.nih.gov). All protein sequences matching at least 65% of SSI2 sequence in BLASTP were retrieved. From this set, all sequences with at least 98% pairwise amino acid sequence identity to any other sequence in the dataset were considered redundant and only one of the redundant sequences kept. The resulting sequences were used for tBLASTN comparison with NCBI genome and NCBI cDNA databases (both from www.ncbi.nlm.nih.gov), the *Populus trichocarpa* (http://genome.jgi-psf.org) and *Vitis vinifera* genomes (www.cns.fr/vitis), and the PlantGDB EST and genome database (www.plantgdb.org), in their most current versions as of 29 July 2009. *Ophrys* sequences from this study were also included at this step, which was carried out by using the FGF scripts of (14), kindly provided by Hongkun Zheng, using the default parameters (i.e., a minimum of 50% of sequence match at the amino acid level). Resulting nucleic acid sequences were aligned to the input protein sequences and ORFs were identified by using GeneWise 2.2 (15). Redundant sequences were removed from the resulting set of sequences if pairwise identity was at least 96% at the nucleic acid level (excepting *Ophrys* sequences), keeping only the best possible hit, to yield the final sequence set for phylogenetic analysis. Perl scripts and data for Table S1 are available upon request.

1. Scotto-Lavino E, Du G, Frohman MA (2006) 3′ end cDNA amplification using classic RACE. *Nat Protoc* 1:2742–2745.
2. Scotto-Lavino E, Du G, Frohman MA (2006) Amplification of 5′ end cDNA with 'new RACE'. *Nat Protoc* 1:3056–3061.
3. Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image processing with ImageJ. *Biophoton Int* 11:36–42.
4. R Development Core Team (2010) R: A language and environment for statistical computing (R Found for Stat Comput, Vienna) 2.11.0.
5. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22:1540–1542.
6. Alonso JM, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653–657.
7. Lightner J, James DW, Jr., Dooner HK, Browse J (1994) Altered body morphology is caused by increased stearate levels in a mutant of *Arabidopsis*. *Plant J* 6:401–412.
8. Kachroo A, et al. (2007) The *Arabidopsis* stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Mol Biol* 63:257–271.
9. Kachroo P, Shanklin J, Shah J, Whittle EJ, Klessig DF (2001) A fatty acid desaturase modulates the activation of defense signaling pathways in plants. *Proc Natl Acad Sci USA* 98:9448–9453.
10. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 3rd Ed.
11. Cahoon EB, Lindqvist Y, Schneider G, Shanklin J (1997) Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proc Natl Acad Sci USA* 94:4872–4877.
12. Whittle EJ, Tremblay AE, Buist PH, Shanklin J (2008) Revealing the catalytic potential of an acyl-ACP desaturase: Tandem selective oxidation of saturated fatty acids. *Proc Natl Acad Sci USA* 105:14738–14743.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
14. Zheng H, et al. (2007) FGF: a web tool for Fishing Gene Family in a whole genome database. *Nucleic Acids Res* 35 (Web Server issue):W121–W125.
15. Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. *Genome Res* 14:988–995.
16. Zhang J, Kumar S, Nei M (1997) Small-sample tests of episodic adaptive evolution: A case study of primate lysozymes. *Mol Biol Evol* 14:1335–1338.
17. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300:1005–1016.
18. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 10:1–6.
19. Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 8:978–984.

20. Bjellqvist B, et al. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14:1023–1031.
21. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, ed Walker JM (Humana Press, Totowa, NJ), pp 571–607.
22. Wiederstein M, Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35 (Web Server issue):W407–W410.
23. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291.
24. Schauer AML (2010) Protein dynamics of pollen development. PhD thesis (University of Zurich, Zurich).
25. Schmidt A, et al. (2004) UPS1 and UPS2 from *Arabidopsis* mediate high affinity transport of uracil and 5-fluorouracil. *J Biol Chem* 279:44817–44824.
26. Penninckx IAMA, et al. (1996) Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. *Plant Cell* 8:2309–2323.
27. Lindqvist Y, Huang W, Schneider G, Shanklin J (1996) Crystal structure of Δ^9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J* 15:4081–4092.
28. Schiestl FP, et al. (2000) Sex pheromone mimicry in the early spider orchid (*Ophrys sphegodes*): Patterns of hydrocarbons as the key mechanism for pollination by sexual deception. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 186: 567–574.
29. Mant JG, et al. (2005) Cuticular hydrocarbons as sex pheromone of the bee *Colletes cunicularius* and the key to its mimicry by the sexually deceptive orchid, *Ophrys exaltata*. *J Chem Ecol* 31:1765–1787.
30. Mant JG, Peakall R, Schiestl FP (2005) Does selection on floral odor promote differentiation among populations and species of the sexually deceptive orchid genus *Ophrys*? *Evolution* 59:1449–1463.
31. Vijayan P, Routaboul J-M, Browse J (1998) A genetic approach to investigating membrane lipid structure and biosynthetic function. *Lipids in Photosynthesis: Structure, Function and Genetics, Advances in Photosynthesis and Respiration*, eds Siegenthaler P-A, Murata N (Kluwer Academic, Dordrecht, The Netherlands), pp 263–285.

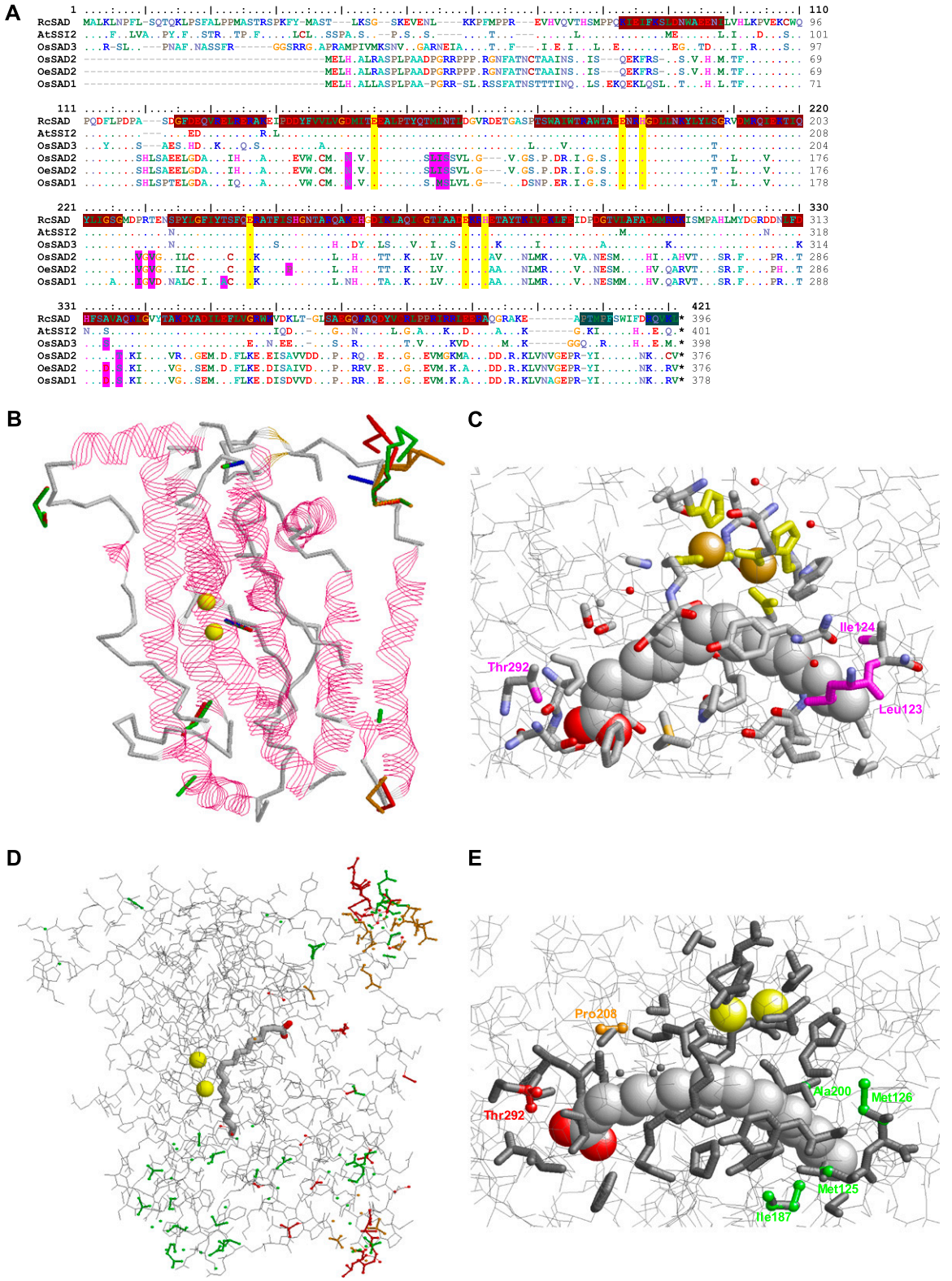

Fig. S1. Orchid SAD homologs and protein structures. (*A*) Alignment of amino acid sequences of *Ricinus communis* SAD (RcSAD) and *Arabidopsis thaliana* SSI2 (AtSSI2) with *Ophrys* SADs, where gaps and identities are denoted by "-" and ".", respectively. Background color indicates α-helix (red), β-sheet (dark green), conserved Fe-coordinating residues (yellow), and sites in orchid SADs within 6 Å of substrate position that differ from RcSAD (magenta). (*B*) Comparison of protein backbones among the RcSAD crystal structure (27) and the superimposed *Ophrys* SAD homology models. The RcSAD structure is shown, with magenta and orange strands indicating α-helices and β-sheets, respectively, and a thick gray line indicating the protein backbone elsewhere. Active site Fe atoms are shown as yellow spheres. The backbones of *Ophrys* SAD models are indicated by colored lines wherever their positions deviate from the RcSAD structure by at

Legend continued on following page

least 1.0 Å. (*C*) OsSAD2 homology model, showing sites within 6 Å of 18:0 substrate or Fe atoms, indicating conserved Fe-binding sites (yellow) and differences from RcSAD within 4 Å of substrate (magenta). (*D*) Differences among OsSAD1, OsSAD2, and OeSAD2 homology models, showing OsSAD2 as a gray wireframe and the active site Fe atoms (yellow) and modeled 18:0 substrate (from ref. 27). All atoms deviating by >1.0 Å in their positions among structures are highlighted in color. (*E*) Amino acid differences among OsSAD1, OsSAD2, and OeSAD2 homology models within 6.0 Å of active site Fe and modeled 18:0 substrate. Where there are no differences, the OsSAD2 structure is shown in gray; amino acid differences are superimposed in color. (*B, D,* and *E*) Parts specific to OsSAD1 are shown in green, OsSAD2 in red, OeSAD2 in orange, and OsSAD3 in dark blue.

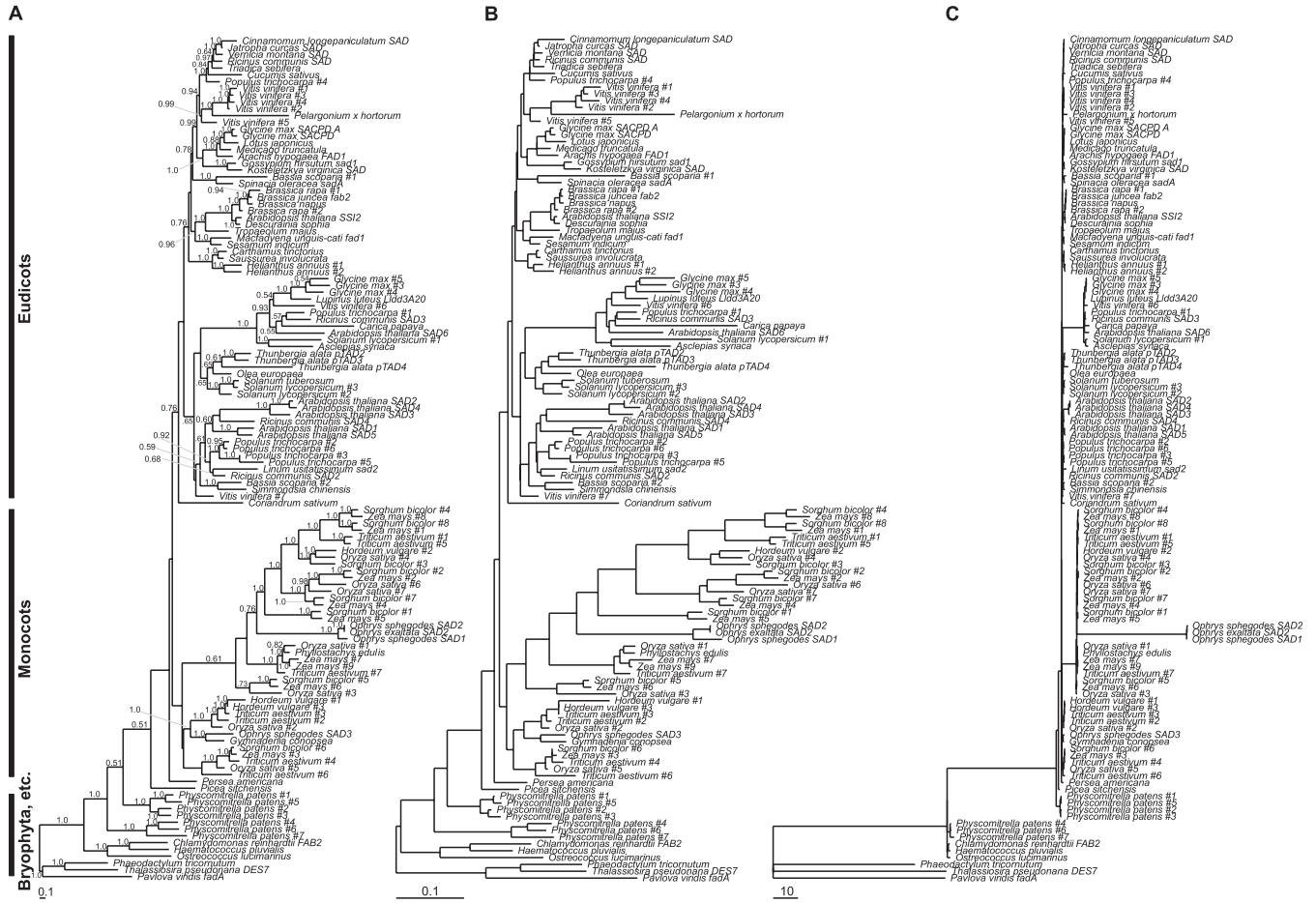

Fig. S2. Phylogenetic trees of plant stearoyl-ACP desaturase homologs. (*A*) Phylogenetic tree with topology from Bayesian inference (BI) reconstruction of sequences in Table S1, with BaseML-optimized branch lengths. Values near branches indicate statistical support as given by BI posterior probabilities (where >0.5). A subsection of this tree is presented and annotated in Fig. 1. (*B* and *C*) Same tree topology with branch lengths from CodeML's free-ratio model M1, indicating rates of nonsynonymous site change d_N (*B*) or synonymous site change d_S (*C*).

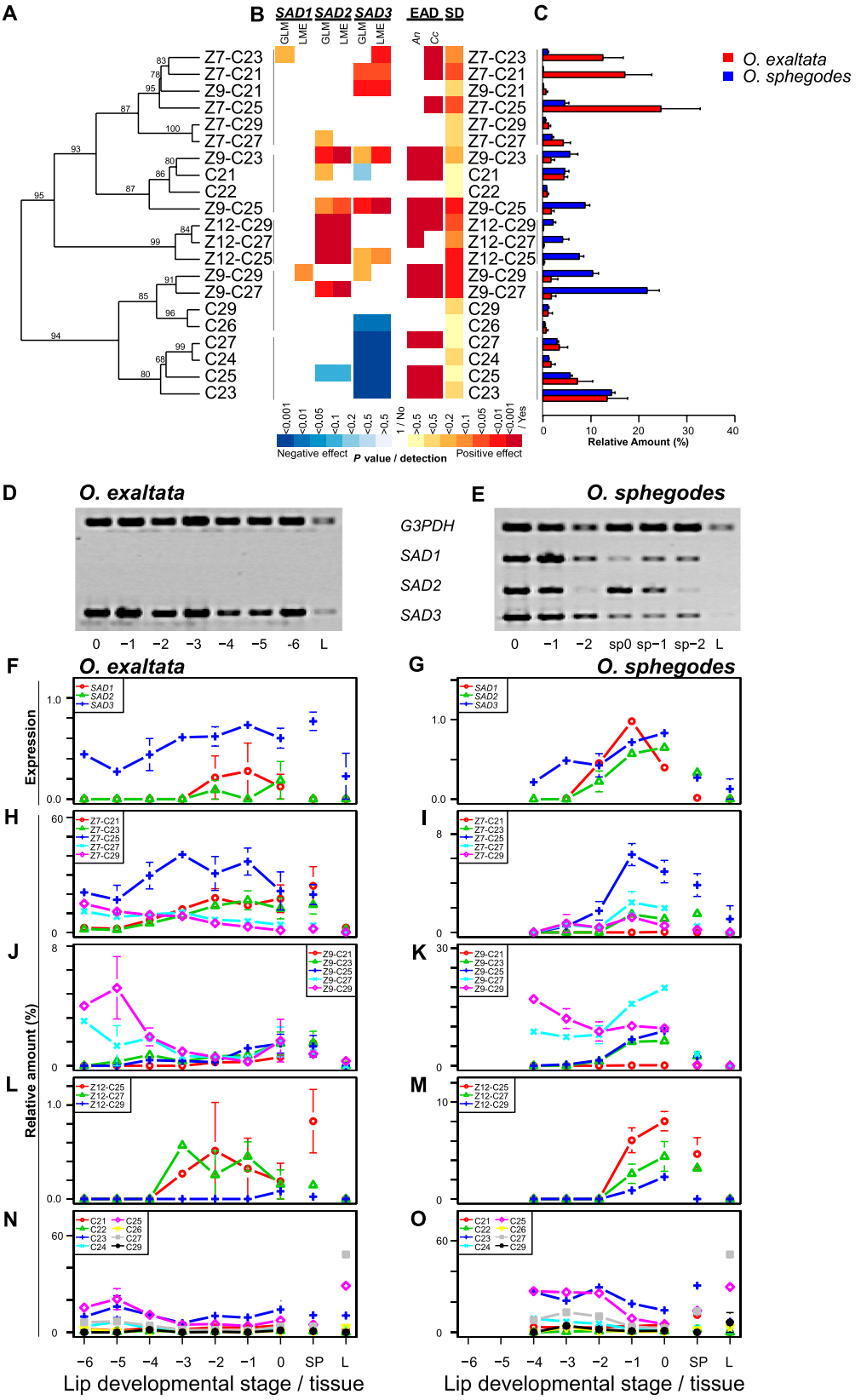

Fig. S3. Details of *SAD* expression and hydrocarbons. (*A*) Ward clustering of hydrocarbons by pairwise correlation across different tissues and flower stages of *O. sphegodes* and *O. exaltata*. Labels above branches represent approximately unbiased *P* value percentages. (*B*) Color-coded summary of statistical tests and compound detectability by pollinators, showing the significance of correlations of *SAD1*, *SAD2,* and *SAD3* expression with hydrocarbons in GLM and LME, the electro-antennographic detection (EAD) of compounds by pollinators *Andrena nigroaenea* (*An*) and *Colletes cunicularius* (*Cc*) (from refs. 28 and 29) and SD, the species difference in compounds in a Wilcoxon signed-rank test. (*C*) Relative amounts and SEM of hydrocarbons in mature labella of *O. sphegodes* and

Legend continued on following page

O. exaltata. This pattern is concordant with previous reports (30). (*D* and *E*) Example gel pictures of RT-PCRs from *O. exaltata* (*D*) and *O. sphegodes* (*E*) showing expression of *SAD* genes (33 cycles) and *G3PDH* (31 cycles). The *O. exaltata* individual shown, together with a second individual in this experiment, shows no detectable *SAD2* expression at 33 cycles, although *SAD2* expression for the same individuals was detectable at higher cycle numbers (≥38). (*F–O*) Gene expression and hydrocarbon abundance in different flower stages and tissues, bars showing SEM, in *O. exaltata* (*F, H, J, L,* and *N*) and *O. sphegodes* (*G, I, K, M,* and *O*). (*F* and *G*) Normalized expression of *SAD1*, *SAD2,* and *SAD3* relative to *G3PDH*. (*H–O*) Relative amounts (%) of 7-alkenes (*H* and *I*), 9-alkenes (*J* and *K*), 12-alkenes (*L* and *M*), and alkanes (*N* and *O*). Labella are labeled 0 (anthetic flower) to -X (bud, X positions before flower 0). Pooled sepal/petal samples are labeled sp followed by flower position, or SP for mature sepal/petal, and leaves are labeled L.

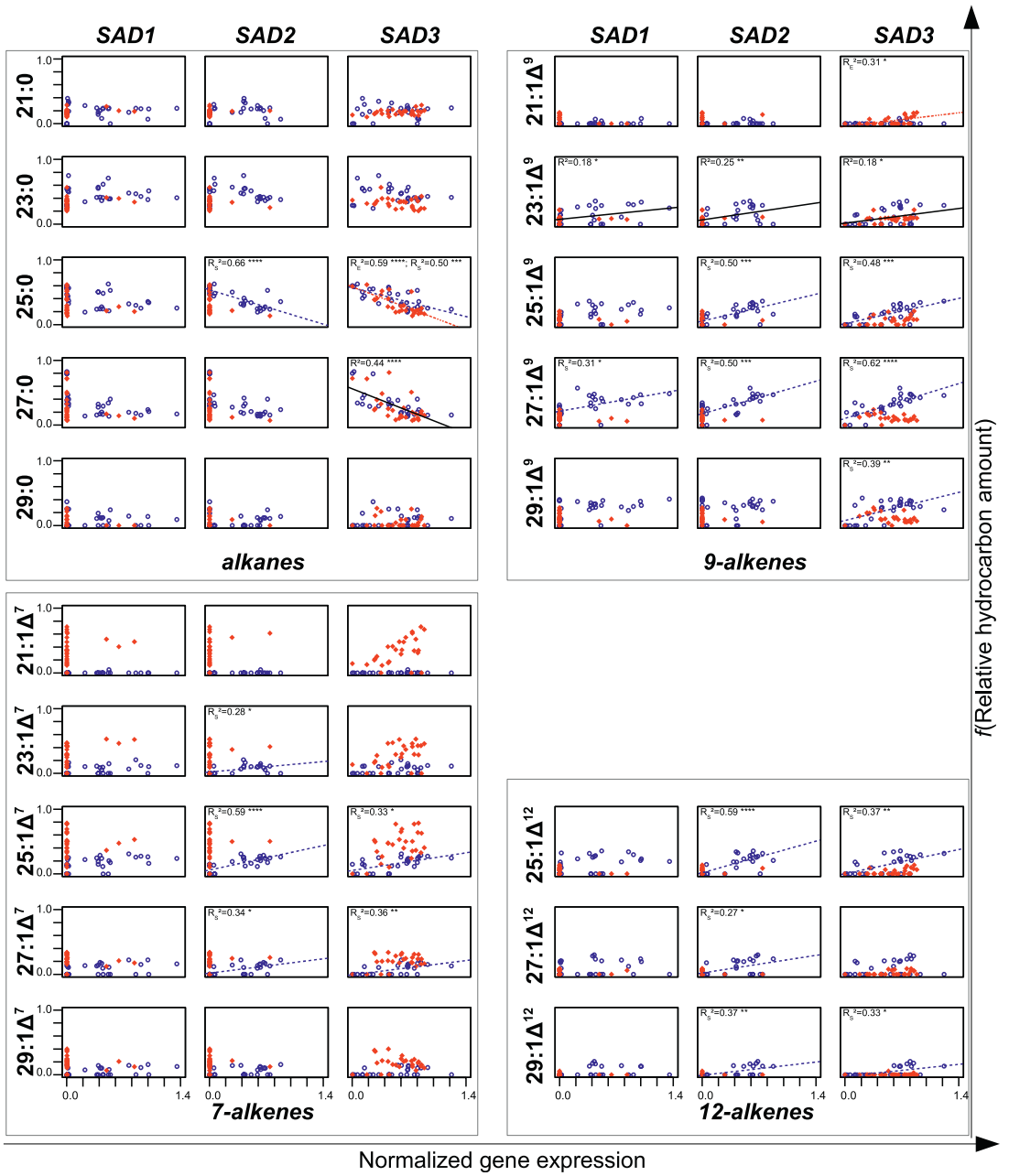

Fig. S4. Correlations of hydrocarbons with gene expression. Normalized gene expression of *SAD1*, *SAD2*, and *SAD3* versus $f(x) = \text{arcsin } x^{0.5}$ transformed relative amounts of alkanes and alkenes from all tissues and floral stages, plotted individually for each comparison. *O. exaltata* is shown in red diamonds and *O. sphegodes* in blue circles. Regression lines are plotted for significant correlations ($P < 0.05$, after correction of P values for multiple testing by using Holm's method). Where correlations were significantly different for the two species (one-way ANOVA; $P < 0.05$), separate regression lines are shown for them. R^2 values are shown, where E and S subscripts denote *O. exaltata* and *O. sphegodes*, respectively, and their significance indicated: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. Note that this correlation analysis is less powerful than the analyses summarized in Fig. S3*B*.

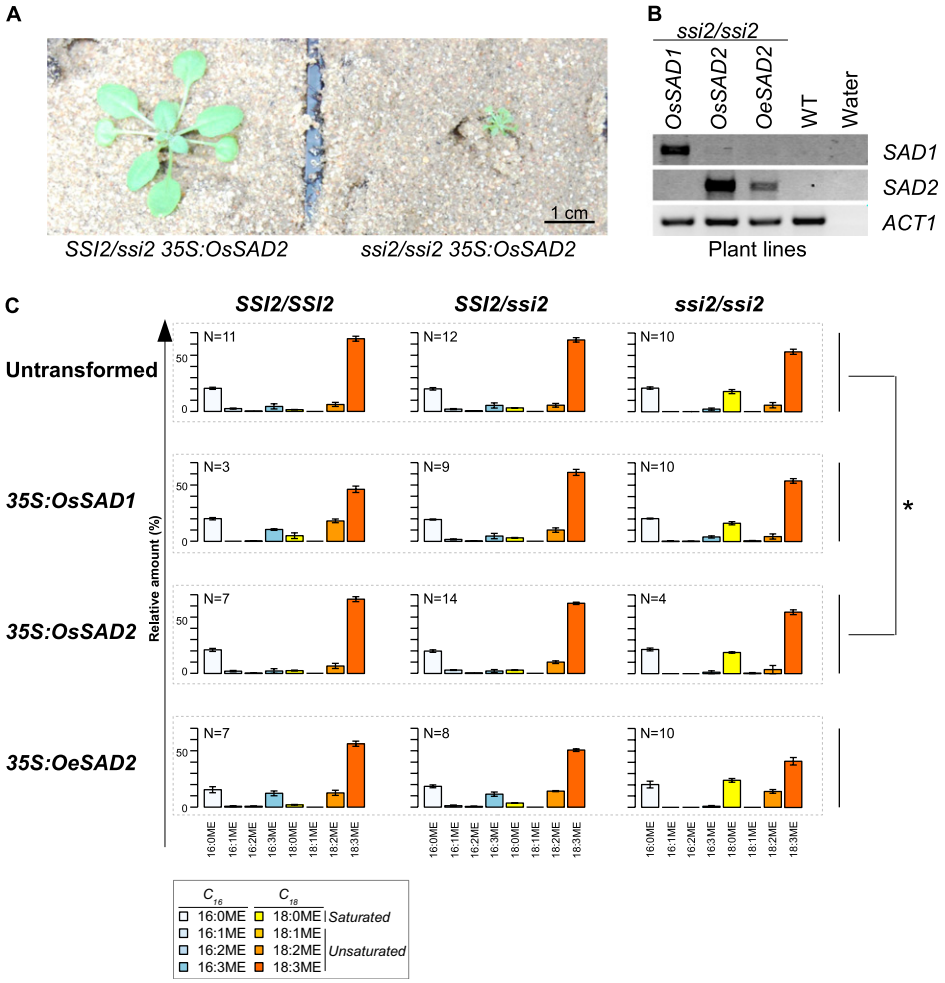

Fig. S5. Comparison of *Arabidopsis* lines. (*A*) Photograph of two *Arabidopsis* plants before bolting. Both plants carry the *2×35S:OsSAD2* transgene. Plants heterozygous for the *ssi2* T-DNA insertion (*Left*) have a wild-type (WT) phenotype, whereas *ssi2/ssi2* homozygous plants (*Right*) have a dwarf phenotype that is not complemented by *Ophrys SAD* transgenes. (*B*) Expression of *Ophrys SAD* transgenes (and *ACT1* control) in *Arabidopsis ssi2/ssi2* plants, as observed after RT-PCR (T_A = 56 °C), with untransformed WT plant and water as controls. (*C*) Relative amounts of C_{16} and C_{18} FAMEs from leaves of *Arabidopsis* plant lines (or whole plants, without roots, in the case of phenotypic dwarves), error bars indicating SEM. The sample number *N* shown refers to pooled samples (typically 5 individuals per sample) analyzed by GC/MS. The levels of fatty acids (FAs) in untransformed plants are similar to those listed in ref. 31. The effect of any transgene on C_{16} and C_{18} unsaturated FA levels was tested in a generalized linear model and found to be significant ($P < 0.05$) only for the *OsSAD2* transgene (indicated by an asterisk; $P = 0.012$ and $P = 0.004$ for unsaturated C_{16} and C_{18} FAs, respectively).

Table S1. List of stearoyl-ACP desaturase homologs identified

Species and gene symbol	Source	ID or accession no.
Arabidopsis thaliana SAD1	TAIR	At5g16240
Arabidopsis thaliana SAD2	TAIR	At3g02610
Arabidopsis thaliana SAD3	TAIR	At5g16230
Arabidopsis thaliana SAD4	TAIR	At3g02620
Arabidopsis thaliana SAD5	TAIR	At3g02630
Arabidopsis thaliana SAD6	TAIR	At3g02630
Arabidopsis thaliana SSI2	TAIR	At2g43710
Arachis hypogaea FAD1	NCBI	AF172728.1
Asclepias syriaca	NCBI	U60277.1
Bassia scoparia #1	NCBI	AF315599.1
Bassia scoparia #2	NCBI	AF315600.1
Brassica juncea fab2	NCBI	AF153420.1
Brassica napus	NCBI	AY642537.1
Brassica rapa #1	NCBI	DQ886528.1
Brassica rapa #2	NCBI	X60978.1
Carica papaya	NCBI	AY451852.1
Carthamus tinctorius	NCBI	M61109.1
Chlamydomonas reinhardtii FAB2	NCBI	XM_001691545.1
Cinnamomum longepaniculatum SAD	NCBI	EU131523.1
Coriandrum sativum	NCBI	AY770499.1
Cucumis sativus	NCBI	M59858.1
Descurainia sophia	NCBI	EF524186.1
Glycine max SACPD	NCBI	L34346.1
Glycine max SACPD A	NCBI	AY885234.2
Glycine max #3	NCBI	AK245255.1
Glycine max #4	PlantGDB	Gm13A
Glycine max #5	PlantGDB	Gm13B
Gossypium hirsutum sad1	NCBI	AJ132636.1
Gymnadenia conopsea	NCBI	EF051334.1
Haematococcus pluvialis	NCBI	EF586860.1
Helianthus annuus #1	NCBI	U91339.1
Helianthus annuus #2	NCBI	U91340.1
Hordeum vulgare #1	PlantGDB	PUT-157a-Hordeum_vulgare-46041
Hordeum vulgare #2	PlantGDB	PUT-157a-Hordeum_vulgare-65964
Hordeum vulgare #3	NCBI	151419245
Jatropha curcas SAD	NCBI	DQ084491.1
Kosteletzkya virginica SAD	NCBI	FJ750952.1
Linum usitatissimum sad2	NCBI	AJ006958.1
Lotus japonicus	NCBI	DQ020280.1
Lupinus luteus Lldd3A20	NCBI	AF139377.1
Macfadyena unguis-cati fad1	NCBI	AF051134.1
Medicago truncatula	NCBI	BT052760.1
Olea europaea	NCBI	U58141.1
Ophrys exaltata subsp. archipelagi SAD2	This study	FR688110
Ophrys sphegodes SAD1	This study	FR688108
Ophrys sphegodes SAD2	This study	FR688109
Ophrys sphegodes SAD3	This study	FR688106
Oryza sativa #1	TIGR	Os01g0880800
Oryza sativa #2	TIGR	Os01g0919900
Oryza sativa #3	TIGR	Os02g0504800
Oryza sativa #4	TIGR	Os03g0423300
Oryza sativa #5	TIGR	Os04g0379900
Oryza sativa #6	TIGR	Os08g0199400
Oryza sativa #7	TIGR	Os08g0200100
Ostreococcus lucimarinus	NCBI	XM_001417261.1
Pavlova viridis fadA	NCBI	EU000382.1
Pelargonium x hortorum	NCBI	U40344.1
Persea americana	NCBI	AF116861.1
Phaeodactylum tricornutum	NCBI	XM_002177381.1
Phyllostachys edulis	NCBI	FP091892.1
Physcomitrella patens #1	NCBI	168039273
Physcomitrella patens #2	NCBI	168040944
Physcomitrella patens #3	NCBI	168044588
Physcomitrella patens #4	NCBI	168065984

Table S1. Cont.

Species and gene symbol	Source	ID or accession no.
Physcomitrella patens #5	PlantGDB	scaffold
Physcomitrella patens #6	PlantGDB	scaffold
Physcomitrella patens #7	PlantGDB	scaffold
Picea sitchensis	NCBI	EF085288.1
Populus trichocarpa #1	NCBI	XM_002307442.1
Populus trichocarpa #2	NCBI	XM_002311231.1
Populus trichocarpa #3	NCBI	XM_002316135.1
Populus trichocarpa #4	NCBI	XM_002330664.1
Populus trichocarpa #5	NCBI	XM_002321290.1
Populus trichocarpa #6	*Populus* genome	LG_X
Ricinus communis SAD	NCBI	M59857.1
Ricinus communis SAD2	NCBI	EU878554.1
Ricinus communis SAD3	NCBI	EU878553.1
Ricinus communis SAD4	NCBI	FJ643486.1
Saussurea involucrata	NCBI	DQ516384.1
Sesamum indicum	NCBI	D49832.1
Simmondsia chinensis	NCBI	M83199.1
Solanum lycopersicum #1	NCBI	AK325946.1
Solanum lycopersicum #2	NCBI	BT014263.1
Solanum lycopersicum #3	NCBI	BT014387.1
Solanum tuberosum	NCBI	M91238.1
Sorghum bicolor #1	PlantGDB	*Sorghum* chromosome 1
Sorghum bicolor #2	NCBI	XM_002464685.1
Sorghum bicolor #3	NCBI	XM_002465071.1
Sorghum bicolor #4	NCBI	XM_002460450.1
Sorghum bicolor #5	NCBI	XM_002453832.1
Sorghum bicolor #6	NCBI	XM_002446233.1
Sorghum bicolor #7	NCBI	XM_002443958.1
Sorghum bicolor #8	NCBI	XM_002437426.1
Spinacia oleracea sadA	NCBI	FJ418167.1
Thalassiosira pseudonana DES7	NCBI	XM_002297328.1
Thunbergia alata pTAD2	NCBI	U07597.1
Thunbergia alata pTAD3	NCBI	U07605.1
Thunbergia alata pTAD4	NCBI	U09269.1
Triadica sebifera	NCBI	EF079655.1
Triticum aestivum #1	NCBI	AK331076.1
Triticum aestivum #2	PlantGDB	PUT-163b-Triticum_aestivum-014567
Triticum aestivum #3	NCBI	241986144
Triticum aestivum #4	PlantGDB	PUT-163b-Triticum_aestivum-019534
Triticum aestivum #5	NCBI	AK333552.1
Triticum aestivum #6	PlantGDB	PUT-163b-Triticum_aestivum-766165441
Triticum aestivum #7	PlantGDB	PUT-163b-Triticum_aestivum-157676
Tropaeolum majus	NCBI	AY188688.1
Vernicia montana SAD	NCBI	EU072353.1
Vitis vinifera #1	NCBI	225434117
Vitis vinifera #2	NCBI	225434123
Vitis vinifera #3	NCBI	XR_077944.1
Vitis vinifera #4	*Vitis* genome	chr5
Vitis vinifera #5	NCBI	XM_002274672.1
Vitis vinifera #6	NCBI	XM_002264759.1
Vitis vinifera #7	NCBI	XM_002265150.1
Zea mays #1	NCBI	EU946993.1
Zea mays #2	NCBI	EU966614.1
Zea mays #3	NCBI	NM_001139285.1
Zea mays #4	NCBI	BT063291.1
Zea mays #5	NCBI	NM_001157868.1
Zea mays #6	NCBI	226506507
Zea mays #7	NCBI	BT018751.1
Zea mays #8	NCBI	NM_001157610.1
Zea mays #9	PlantGDB	PUT-163a-Zea_mays-103945

Gene symbols are reported with the species, where such symbols were available in the source databases. In other cases, different homologs in a species are simply numbered (e.g., #1). The ID or accession number within a source database set refers to the best partial match of a given sequence identified by our script. TAIR, The Arabidopsis Information Resource; TIGR, The Institute for Genomic Research; NCBI, National Center for Biotechnology Information.

Table S2. Analysis of selection using PAML: Maximum likelihood models

Substitution model	*ln* L	Parameter MLEs
M0 (one ω ratio)	−60531.7	ω_0 = 0.101
Site models		
M1a (nearly neutral)	−60211.2	ω_0 = 0.097, ω_1 = 1.0 (p_0 = 0.896, p_1 = 0.104)
M2a (positive selection)	−60211.2	ω_0 = 0.097, ω_1 = 1.0, ω_2 = 1.0 (p_0 = 0.896, p_1 = 0.033, p_2 = 0.071)
M7 (β)	−58747.8	p = 0.794, q = 4.738
M8 (β+ω>1)	−58745.1	p_0 = 0.991, p = 0.813, q = 5.059 (p_1 = 0.009), ω=1.055
M8a (β+ω=1)	−58745.1	p_0 = 0.991, p = 0.813, q = 5.057 (p_1 = 0.009), ω=1.0
Branch models		
M1 (free-ratio)	−59625.5	ω_A = 0.004, ω_B = 85.215, ω_C= 0.206, ω_D = 0.071, ω_E = 0.060, etc.*
M2 (2-ratio), A: *SAD1*/*SAD2*	−60527.0	ω_0 = 0.102, ω_1=ω_A = 0.009
M2 (2-ratio), B: *SAD1*	−60506.9	ω_0 = 0.100, ω_1=ω_B = 999
M2 (2-ratio), C: *SAD2*	−60515.7	ω_0 = 0.101, ω_1=ω_C = 999
M2 (2-ratio), D: *SAD3*/ *Gymnadenia SAD*	−60531.0	ω_0 = 0.102, ω_1=ω_D = 0.063
M2 (2-ratio), E: *SAD3*	−60529.8	ω_0 = 0.102, ω_1=ω_E = 0.060
M2 (3-ratio), A,B	−60503.0	ω_0 = 0.101, ω_1=ω_A = 0.0098, ω_2=ω_B = 999
M2 (3-ratio), A,C	−60509.6	ω_0 = 0.101, ω_1=ω_A = 0.008, ω_2=ω_C = 999
M2 (3-ratio), B,C	−60505.1	ω_0 = 0.100, ω_1=ω_B = 999, ω_2=ω_C = 0.210
M2 (4-ratio), A,B,C	−60501.0	ω_0 = 0.101, ω_1=ω_A = 0.009, ω_2=ω_B = 999, ω_3=ω_C = 0.218

*This model has different ω ratios (where $\omega = d_N/d_S$) for every branch in the tree (not listed).

Likelihood values (*ln* L) and parameter maximum likelihood estimates (MLEs) for substitution models (names reflect naming in PAML output). Letters after models M2 (and as parameter subscripts) refer to branches as labeled in Fig. 1.

Table S3. Analysis of selection using PAML: Likelihood ratio tests

Model comparison	2δ	d.f.	*P*	Significance
M1a - M2a	$4.56{\cdot}10^{-4}$	2	>0.999	NS
M7 - M8	5.424	2	0.066	NS
M8 - M8a	0.040	1	0.841	NS
M0 - M1	1812.276	245	$7.58{\cdot}10^{-237}$	****
M0 - M2, A	9.451	1	0.002	**
M0 - M2, B	49.541	1	$1.942{\cdot}10^{-12}$	****
M0 - M2, C	31.998	1	$1.543{\cdot}10^{-8}$	****
M0 - M2, D	1.426	1	0.232	NS
M0 - M2, E	3.676	1	0.055	NS
M2, A - M2, A,B	47.987	1	$4.29{\cdot}10^{-12}$	****
M2, A - M2, A,C	34.758	1	$3.734{\cdot}10^{-9}$	****
M2, B - M2, A,B	7.897	1	0.005	**
M2, B - M2, B,C	3.612	1	0.057	NS
M2, C - M2, A,C	12.211	1	$4.752{\cdot}10^{-4}$	***
M2, C - M2, B,C	21.155	1	$4.235{\cdot}10^{-6}$	****
M2, A,B - M2, A,B,C	3.937	1	0.047	*
M2, A,C - M2 A,B,C	17.166	1	$3.425{\cdot}10^{-5}$	****
M2, B,C - M2 A,B,C	8.222	1	0.004	**
M2, B - M1	1762.735	244	$5.58{\cdot}10^{-228}$	****
M2, C - M1	1780.278	244	$2.86{\cdot}10^{-231}$	****
M2, A,B,C - M1	1750.901	242	$1.27{\cdot}10^{-226}$	****

Significance: NS, not significant ($P \geq 0.05$); *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. Likelihood ratio tests for differences among substitution models (from Table S2), where $2\delta = 2(ln\ L_1 - ln\ L_2)$ is compared against a χ^2 distribution with the degrees of freedom (d.f.) corresponding to the difference in the number of free model parameters.

Table S4. Fisher's exact tests

Branch/gene	Model	Test	N	S	$N_t - N$	$S_t - S$	P
B/*SAD1*	M2, B	Neutral	61	0	1,227	404	$4.18 \cdot 10^{-8}$
B/*SAD1*	M1	Neutral	55	0	1,233	404	$2.30 \cdot 10^{-7}$
C/*SAD2*	M2, C	Neutral	47	0	1,241	405	$2.14 \cdot 10^{-6}$
C/*SAD2*	M1	Neutral*	32	49	1,256	355	$5.40 \cdot 10^{-13}$
C/*SAD2*	M1	Relaxed	32	49	9,944	175,745	$<2.20 \cdot 10^{-16}$

Fisher's exact tests (one-sided) comparing the inferred number of nonsynonymous (N) and synonymous (S) changes at the branch tested to the expected total number of (non)synonymous sites (N_t and S_t), all obtained from PAML output (models from Table S2). This follows the "test of positive selection" (16) to test for deviations from neutral expectation ("Neutral"); to test for relaxation of purifying selection ("Relaxed"), N and S were compared with the total numbers of inferred (non)synonymous changes, which represent overall purifying selection. Letters B and C refer to branches as labeled in Fig. 1.

*In this case, the test outcome signifies purifying rather than positive selection. This branch/model combination was therefore subsequently tested for relaxation of purifying selection.

Table S5. Predicted protein characteristics of orchid desaturases

Species/protein	Length*	Average MW*	pI*	TP length	TP (score)	Z score
OsSAD1	378 (355)	42,704.7 (40,255.7)	5.71 (5.44)	23	Yes (0.517)	−8.52
OsSAD2	376	42,359.5	6.13	33	No (0.497)	−8.11
OeSAD2	376	42,464.5	6.13	33	No (0.497)	−7.73
Os/OeSAD3†	398 (369)	45,412.7 (42,325.1)	6.16 (5.64)	29	Yes (0.539)	−8.92

The species prefix used in protein names is Os for *O. sphegodes* and Oe for *O. exaltata*. MW denotes molecular weight, pI the isoelectric point, and TP a chloroplast transit peptide. Protein and TP length are given in amino acids. The column TP indicates whether a transit peptide was predicted, quoting the prediction score in parentheses. TP prediction was done by using TargetP1.1 (specificity threshold 0.9) and ChloroP1.1 (17–19), and pI/MW were predicted by using the ExPASy server (20, 21). The ProSA-web (22) Z score refers to the respective protein homology models, and is −9.06 for the template RcSAD crystal structure. Ramachandran plots generated by Procheck (23) showed 0.3% of angles in disallowed regions for RcSAD, OsSAD2, OeSAD2, and SAD3, and 1.0% for OsSAD1.

*Values in parentheses for proteins with putative TP removed.

†These proteins are identical on the amino acid level.

Table S6. Oligonucleotides used in this study

Name	Oligonucleotide sequence (5′→3′)	Purpose/comments
FAB2_dn_L_Forw	TGGGCTRTYTGGACTMGRGCDTG	Initial isolation of *SAD* homologs; first amplification.
FAB2_dn_L_Rev	CCATCRTACATWARRTGWGCWGGC	Initial isolation of *SAD* homologs; first amplification.
FAB2_dn2_Forw	GCDGARGARAAYMGDCAYGG	Initial isolation of *SAD* homologs; nested PCR.
FAB2_dn2_Rev	GAYYTTYTTCCTCACTCATDTCRGCAA	Initial isolation of *SAD* homologs; nested PCR.
5′RACE RNA oligo	GGGAGAUUGGUUGCCAUAAGCGGAUCAUCGGGAGG AGAAACGGAU	5′ RACE (24).
LB	CTTTGACGTTGGAGTCCAC	T-DNA left border primer (25) for *Arabidopsis* genotyping.
SSI2Up1F	TGAAACAGGTGCTAGTCCTACTTCA	*Arabidopsis* genotyping.
SSI2Up2F	GCATGTTCAAGTCTTGCACTCCATGC	*Arabidopsis* genotyping.
SSI2Dn1R	CACCTGAAAGCCCGGTTAAGTC	*Arabidopsis* genotyping.
ptB1fs-SAD1f (M)	AAAAAGCAGGCTTTATGGAACTTCACCTTGCACTACT	*Ophrys SAD1* (RT-)PCR, with partial *att*B1 site, allowing N-terminal protein fusion.
ptB2-SAD1r (*)	AGAAAGCTGGGTTTTACACGCGGACCTGCTTGTTGAA	*Ophrys SAD1* (RT-)PCR, with partial *att*B2 site, includes stop codon.
ptB1fs-SAD2f (M)	AAAAAGCAGGCTTTATGGAACTTCACCTTGCACTACG	*Ophrys SAD2* (RT-)PCR, with partial *att*B1 site, allowing N-terminal protein fusion.
ptB2-SAD2r (*)	AGAAAGCTGGGTACTACACGCRGACTTGCTTATTGAATATC	*Ophrys SAD2* (RT-)PCR, with partial *att*B2 site, includes stop codon.
ptB1-SAD3f	AAAAAGCAGGCTAACGAACAGTAGGAGAAGGAGAACAAC	*Ophrys SAD3* (RT-)PCR, with partial *att*B1 site, binds upstream of putative start codon.
ptB2-SAD3r	AGAAAGCTGGGTTTCAAAGCATCACTTCCCGGTC	*Ophrys SAD3* (RT-)PCR, with partial *att*B2 site, includes stop codon.
OsG3PDH-1f	ATGTTCAAGTATGACACTGTGCATGG	*G3PDH* (RT-)PCR.
OsG3PDH-2r	GTCGGCACACGGAAAGACATACCAGTCAA	*G3PDH* (RT-)PCR.
OsG3PDH3f	CGATCCCTACATCACCACCGAATAC	*G3PDH* (RT-)PCR (binds upstream of 1f).
OsG3PDH4r	GCACTTTCCCAACAGCCTTGGCAG	*G3PDH* RT-PCR (binds upstream of 2r). Spans intron/exon boundary.
ACT1f (OWB270)	GGCGATGAAGCTCAATCCAAACG	*Arabidopsis* RT-PCR control (26).
ACT1r (OWB271)	GGTCACGACCAGCAAGATCAAGACG	*Arabidopsis* RT-PCR control (26).
NcoI-MA-SS1f	AAAAAACCATGGCACTACTTGCATCCCCATTGC	PCR; construction of pET9d-MAX-*SS1*.
NcoI-MA-xS2f	AAAAAACCATGGCACTACGTGCATCCCCATTAC	PCR; construction of pET9d-MAX-*SS2*/*ES2*.
BglII-SS1r	AAAAAAAGATCTTTTACACGCGGACCTGCTTGTTGAA	PCR; construction of pET9d-MAX-*SS1*.
BglII-SS2r	AAAAAAAGATCTTCTACACGCAGACTTGCTTATTGAATATC	PCR; construction of pET9d-MAX-*SS2*.
BglII-ES2r	AAAAAAAGATCTTCTACACGCGGACTTGCTTATTGAATATC	PCR; construction of pET9d-MAX-*ES2*.
SAD1-XNcoI-f	GGAGGAGAACCGTCATGGCGATCTTCTGAAC	Site-directed mutagenesis; construction of pET9d-MAX-*SS1*.
SAD2-XNcoI-f	GGAGGAGAACCGTCATGGTGATCTTCTGAAC	Site-directed mutagenesis; construction of pET9d-MAX-*SS2*/*ES2*.
SAD1-XNcoI-r	GTTCAGAAGATCGCCATGACGGTTCTCCTCC	Site-directed mutagenesis; construction of pET9d-MAX-*SS1*.
SAD2-XNcoI-r	GTTCAGAAGATCACCATGACGGTTCTCCTCC	Site-directed mutagenesis; construction of pET9d-MAX-*SS2*/*ES2*.
NcoI-MastL-RcSADf	AAAAAACCATGGCCTCTACCCTCAAGTCTGGTTCT	PCR; construction of pET9d-RcX-*SS1*/*SS2*.
NRcOSAD2R	AACAGGCTTAAGGAAGGTGAGCATGTTATGCTCCACC CAATTGTCTAGGGATTTAAAGATCTCAATCTTT	PCR; construction of pET9d-RcX-*SS1*/*SS2*.

For primers compatible with Gateway (Invitrogen) cloning, only primers with partial *att*B sites are shown; these can be extended to bear full-length *att*B sites as described in Invitrogen's manuals.