nature plants

SUPPLEMENTARY INFORMATION

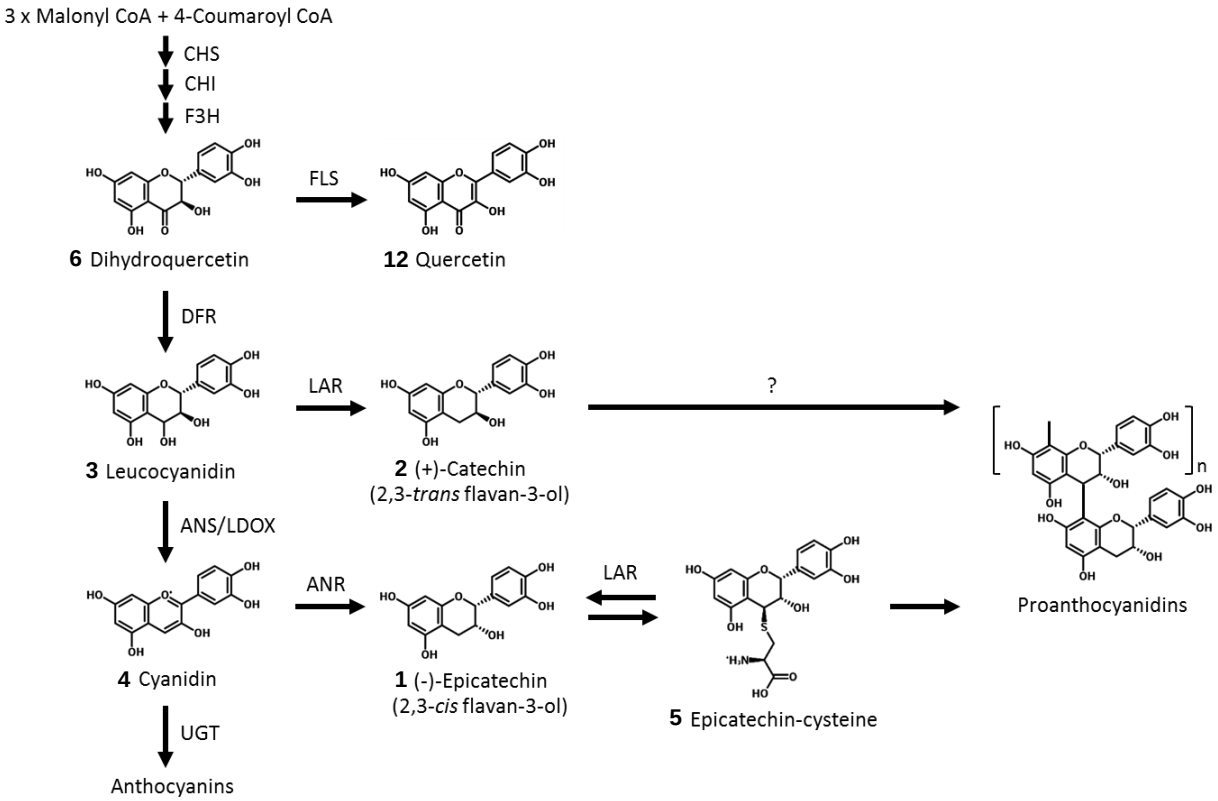
https://doi.org/10.1038/s41477-018-0292-9

In the format provided by the authors and unedited.

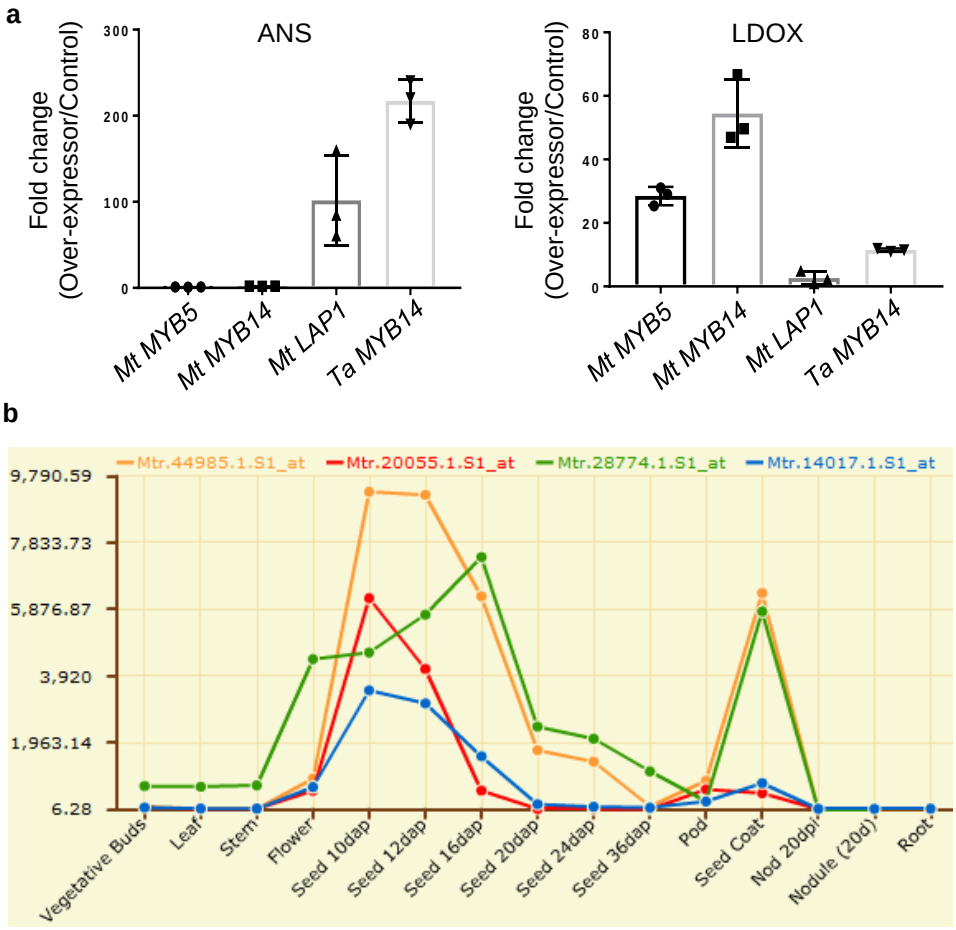
Proanthocyanidin subunit composition determined by functionally diverged dioxygenases

Ji Hyung Jun, Xirong Xiao, Xiaolan Rao and Richard A. Dixon*

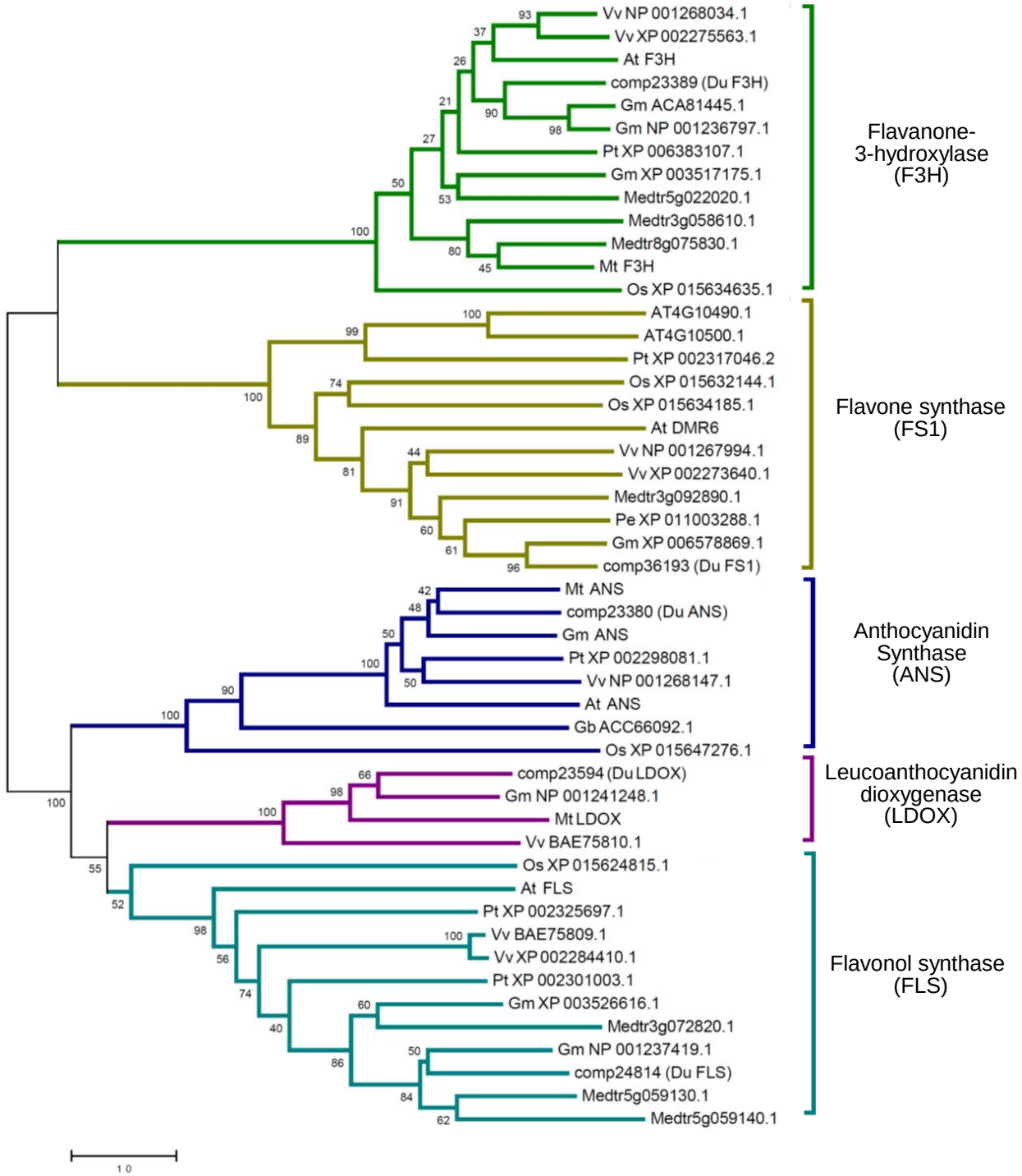
BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX, USA. *e-mail: Richard.Dixon@unt.edu



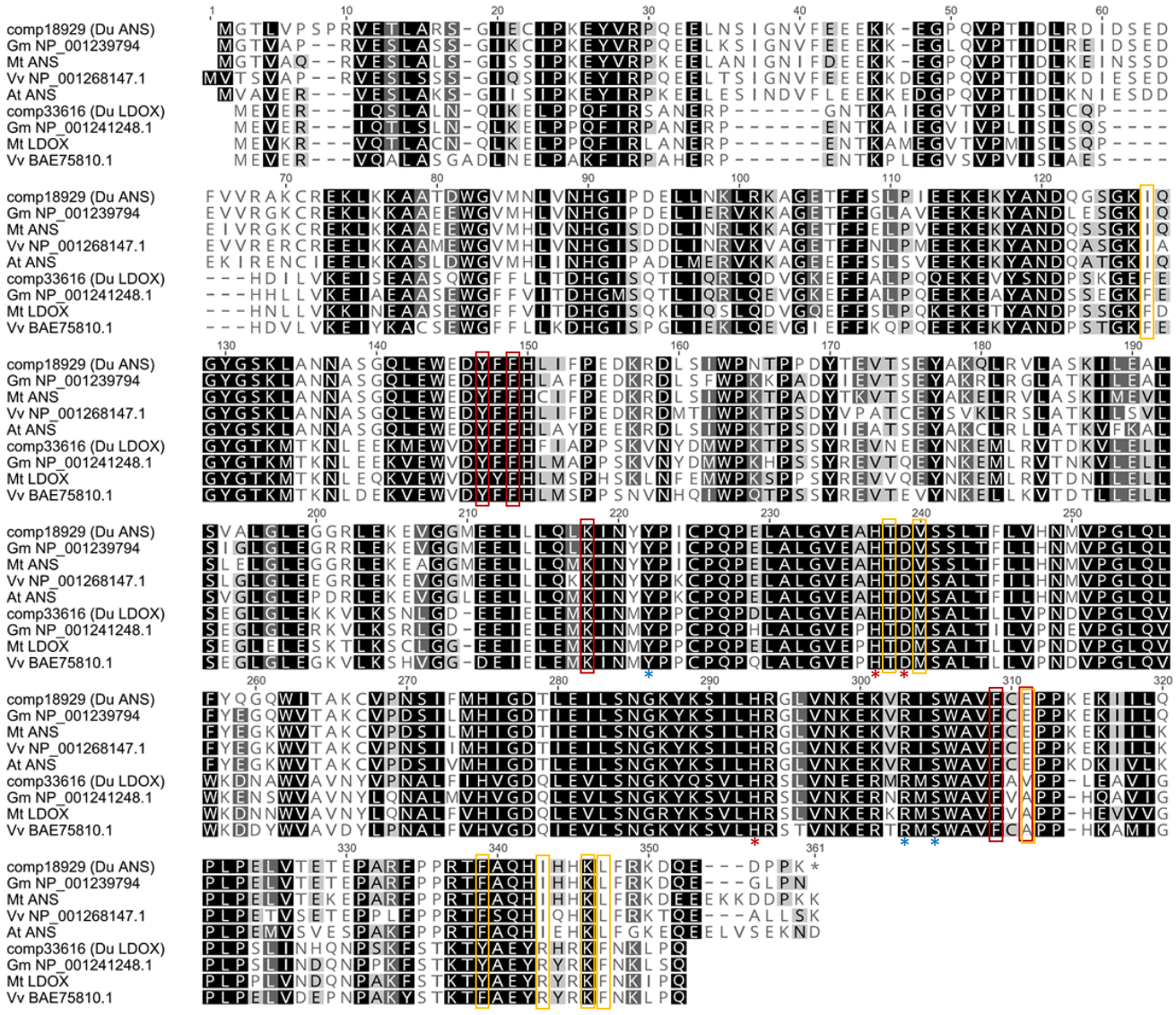
Supplementary Figure 1. Scheme of the flavonoid pathway leading to proanthocyandin(PA) production in *M. truncatula*. The classes of compounds are indicated, with structures illustrating the precursors of catechin and epicatechin. Abbreviations for enzymes are: CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; LDOX, leucoanthocyanidin dioxygenase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; UGT, UDP-glucosyltransferase.



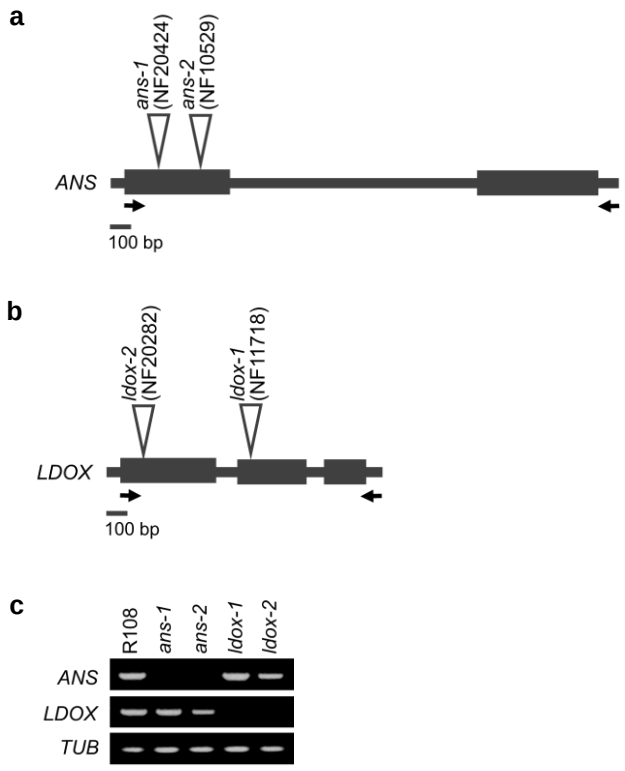
Supplementary Figure 2. Expression of *ANS* and *LDOX*. a, Microarray analysis of Mt *ANS* (probe set ID, Mtr.28774.1.S1_at) and Mt *LDOX* (probe set ID, Mtr.14017.1.S1_at) expression in *M. truncatula* hairy roots overexpressing Mt *MYB14* and Mt *MYB5* and in transgenic *Medicago sativa* (alfalfa) plants with overexpression of Mt *LAP1* and Ta *MYB14*. Microarray experiments were performed on three independent biological replicates, respectively. The relative expression was calculated with normalization to tubulin (probe set ID Mtr.40026.1.S1_at). The data are shown as the mean ± SD (*n* = 3 biologically independent samples). **b**, Tissue specific expression profile of *ANS* and *LDOX* in the microarray data of the *Medicago truncatula* Gene Expression Atlas (http://mtgea.noble.org/v3/probeset.php?id) obtained using the probe set Mtr.28774.1.S1_at (ANS) and Mtr.14017.1.S1_at (LDOX), Mtr.44985.1.S1_at (ANR) and Mtr.20055.1.S1_at (LAR) as queries. Both *ANS* and *LDOX* transcripts are strongly expressed in seed coat and developing seeds.



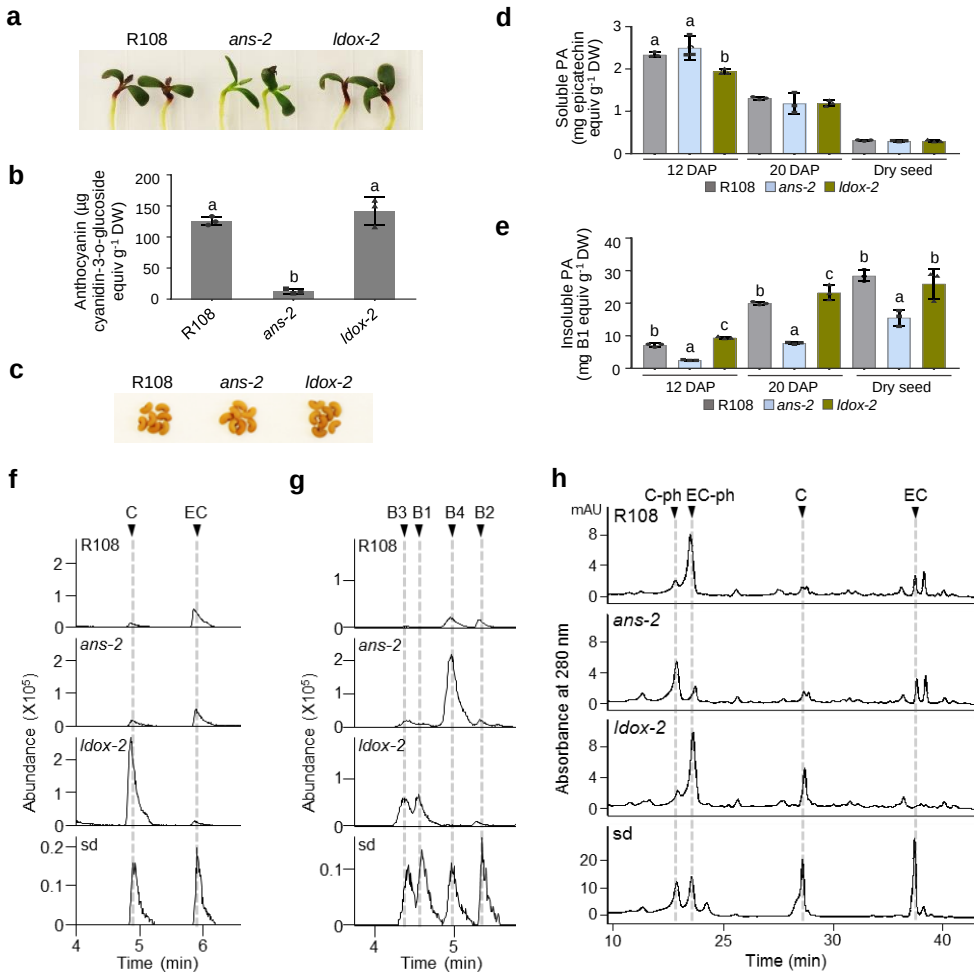
Supplementary Figure 3. Phylogeny of 2OG-Fe(II) oxygenase family oxidoreductase proteins in the flavonoid biosynthetic pathway. The phylogenetic tree was constructed using MEGA 6 [38] with 1000 bootstrap replicates. Numbers indicate the percentage of consensus support. The scale measures evolutionary distances in substitutions per amino acid. The homologous sequences for protein from *Medicago truncatula* were searched in *Glycine, Arabidopsis*, *Vitis*, *Populus*, *Oryza* and *Ginkgo* using BLAST, and Mt LDOX homologous proteins were found in limited taxa compared to other 2-ODDs of the flavonoid pathway. Protein sequences were obtained from GenBank with the following accession numbers: At F3H (BAD89980.1), Mt F3H (XP_003629323.1), At DMR6 (Q9FLV0.1), Mt ANS (XP_003611189.1), Gm ANS (NP_001239794), At ANS/LDOX (NP_194019.1), Mt LDOX (XP_003601080.1). Du F3H, Du FS1 (partial), Du ANS, Du LDOX and Du FLS homologous protein sequences for *Desmodium uncinatum* were identified from RNA sequencing analysis in this study. At, *Arabidopsis thaliana*; Gb, *Ginkgo biloba*; Gm, *Glycine max*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; Pe, *Populus euphratica*; Pt, *Populus trichocarpa*; Vv, *Vitis vinifera*.



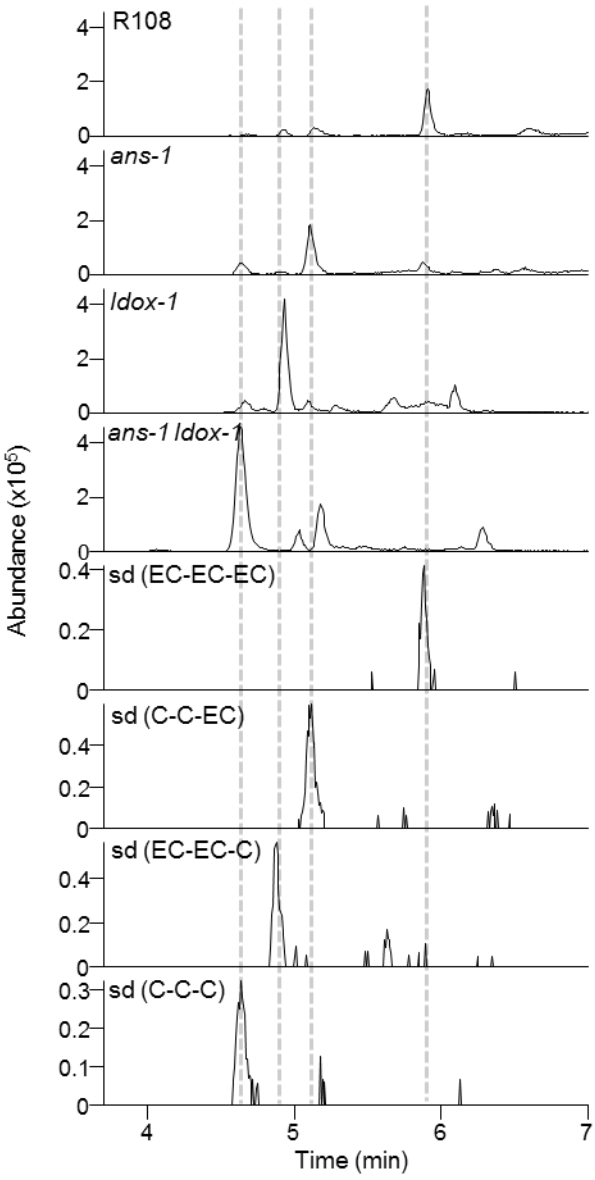
Supplementary Figure 4. Protein sequence alignment of ANS and LDOX proteins. Entire protein sequences were aligned by the ClustalW program with default settings integrated in Geneious Pro 10.1.3 software (Biomatters). * and * indicate Fe(II) and 2-oxoglutarate binding residues, respectively. The amino acids involved in binding to first *trans*-dihydroquercetin (DHQ-1) **6** and second *trans*-dihydroquercetin (DHQ-2) **6** are denoted in the red and yellow boxes, respectively (34). Protein sequences were obtained from GenBank with the following accession numbers: Mt LDOX (XP_003601080.1), At ANS (NP_194019.1), Mt ANS(XP_003611189.1). ANS and LDOX homologous protein sequences for *Desmodium uncinatum* were identified from RNA sequencing analysis in this study. At, *Arabidopsis thaliana*; Du, *Desmodium uncinatum*; Gm, *Glycine max*; Mt, *Medicago truncatula*; Vv, *Vitis vinifera*.



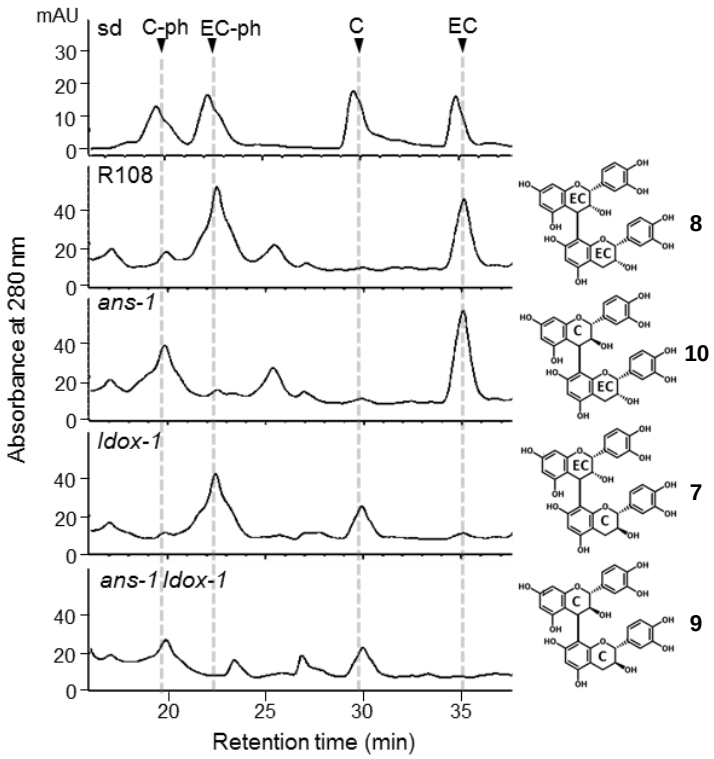
Supplementary Figure 5. Identification of *Tnt1* insertional mutants of ANS and LDOX. a, Schematic of the *ANS* gene depicting *Tnt1* insertion positions in *ans-1* (NF20424) and *ans-2* (NF10529). Gene direction is from left to right. Solid boxes indicate exons and thinner lines indicate introns. The positions of primer pairs used for the PCR in **c** are shown by arrows. **b**, As above, indicating *Tnt1* insertion positions in *ldox-1* (NF11718) and *ldox-2* (NF20282). **c**, RT-PCR for detecting full-length *ANS* and *LDOX* transcripts in wild-type (R108), *ans-1*, *ans-2*, *ldox-1* and *ldox-2*. The representative results of three independent analyses with similar results obtained for each are presented.



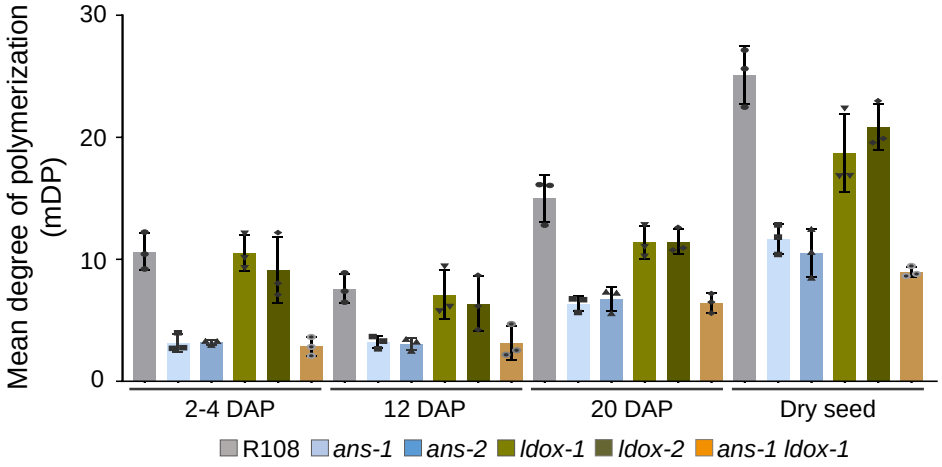
Supplementary Figure 6. Anthocyanin and PA accumulation in *ans-2* and *ldox-2* mutants of *M. truncatula*. **a**, Anthocyanin pigmentation in wild type (R108), *ans-2* and *ldox-2* seedlings at 7 d after germination. **b**, Anthocyanin levels in the same lines. **c**, Seed coat phenotypes of wild type (R108), *ans-2* and *ldox-2* dry mature seeds. **d**, Soluble PA levels in wild type (R108), *ans-2* and *ldox-2* seeds during maturation as quantified with DMACA reagent and expressed as epicatechin equivalents. **e**, Insoluble PA levels in seeds of the same genotypes quantified by the butanol-HCl method and expressed as procyanidin B1 equivalents. **f**, Extracted ion chromatograms (EIC) of epicatechin and catechin monomers (m/z 289.0718 ± 20 ppm) in soluble PAs extracted from seeds of R108, *ans-2* and *ldox-2* at 12 DAP. **g**, EIC of procyanidin dimers (m/z 577.1348 ± 20 ppm) in soluble PAs in the above extracts. **h**, Phloroglucinolysis of soluble PAs from R108, *ans-2* and *ldox-2* 12 DAP seeds. Epicatechin-phloroglucinol **11** and epicatechin **1** released from procyanidin B2 **8** and catechin-phloroglucinol and catechin released from procyanidin B3 **9** were analyzed for comparison. In **a**, **c**, **f**, **g** and **h**, representative result of three independent analyses with similar results obtained for each is presented. In **b**, **d** and **e**, data are shown as the mean ± SD (*n* = 3 biologically independent samples); the different letters above the bars represent statistically significant differences determined by one-way ANOVA (LSD, P = 0.01). C-ph, catechin-phloroglucinol; EC-ph, epicatechin-phloroglucinol; C, catechin; EC, epicatechin; B1, procyanidin B1; B2, procyanidin B2; B3, procyanidin B3; B4, procyanidin B4.



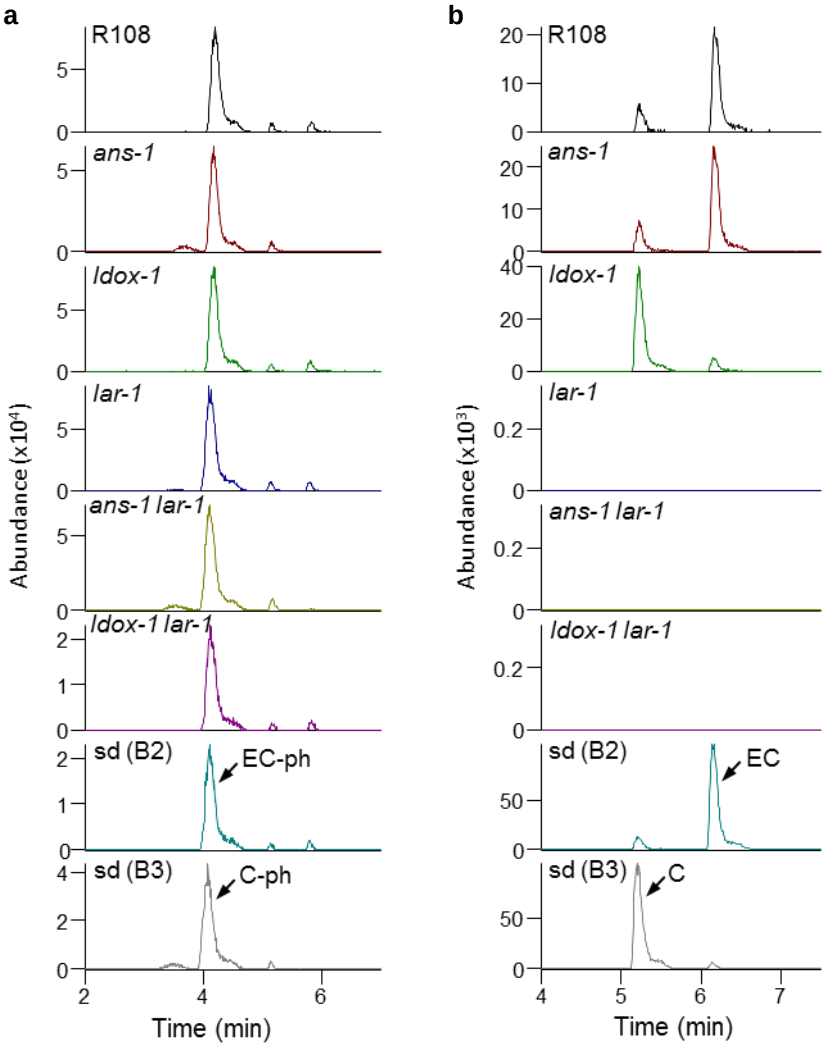
Supplementary Figure 7. Extracted ion chromatograms (EICs) of trimers (m/z 865.1985±20 ppm) in soluble PAs extracted from pods of R108, *ans-1*, *ldox-1* and *ans-1 ldox-1 at* 12 DAP. The representative results of three independent analyses with similar results obtained for each are presented. EC-EC-EC, epicatechin-epicatechin-epicatechin trimer (procyanidin C1); C-C-EC, catechin-catechin-epicatechin trimer; EC-EC-C, epicatechin-epicatechin-catechin trimer; C-C-C, catechin-catechin-catechin trimer (procyanidin C2).



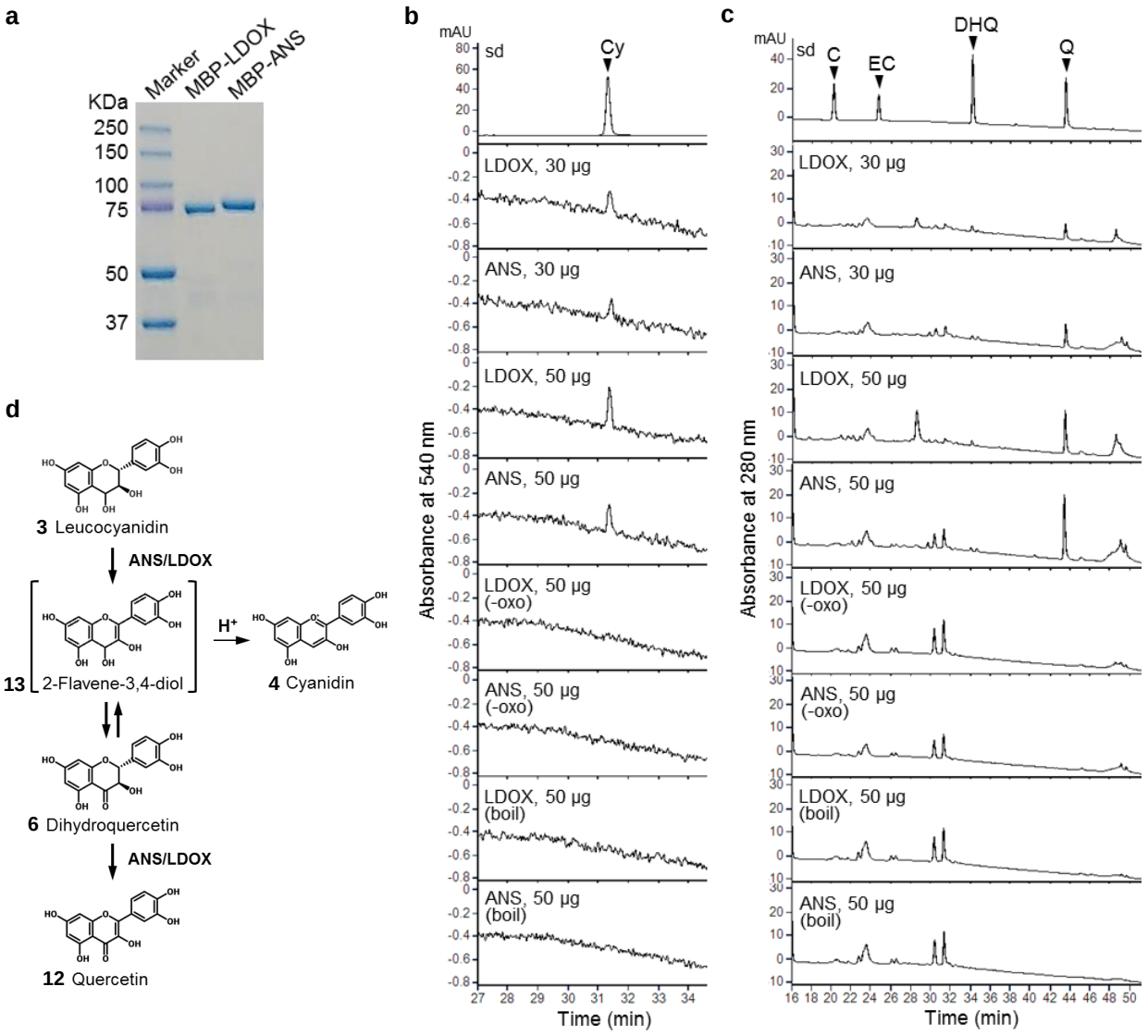
Supplementary Figure 8. Composition of the low molecular weight PA fraction (monomers and oligomers) from seeds of the *ans-1*, *ldox-1* and *ans-1 ldox-1* mutants. HPLC traces of phloroglucinolysis products from wild-type R108 and mutant lines were monitored at 280 nm. Top trace shows standards (sd) of catechin, epicatechin and their phloroglucinol (Ph) adducts. Structures of major dimer for each genotype are presented. The representative results of three independent analyses with similar results obtained for each are presented.



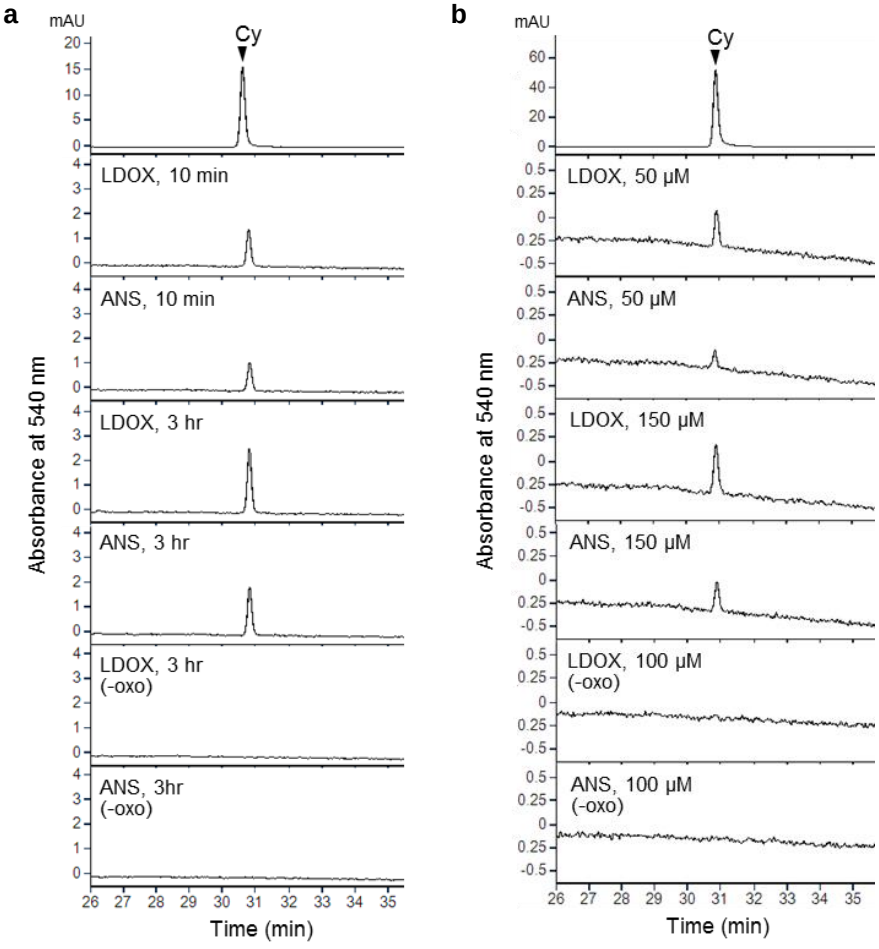
Supplementary Figure 9. Measurement of mDP of soluble PAs from *Mediacago truncatula* seeds. Mean degree of polymerization (mDP) was calculated based on relative proportions of extension to starter units released from the phloroglucinolysis of soluble PAs in wild-type and mutant seeds during maturation. The data are means ±SD of biological triplicates.



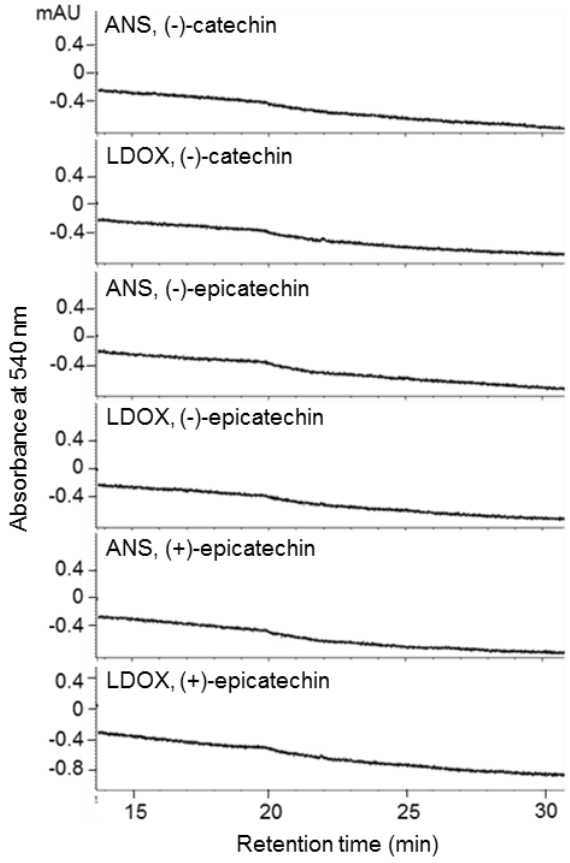
Supplementary Figure 10. LC/MS analysis of phlorogulucinolysis products from wild-type R108 and mutant lines. a, Extracted ion chromatograms (EICs) of epicatechin phloroglucinol (EC-ph) **11** and catechin phloroglucinol (C-ph) adducts (m/z 413.0877 ± 20 ppm) in phloroglucinolysis products using soluble PAs extracted from pods of R108, *ans-1*, *ldox-1, lar-1, ans-1 lar-1 and ldox-1 lar-1* at 2-4 DAP. **b**, EIC of epicatechin (EC) and catechin (C) monomers (m/z 289.0718 ± 20 ppm) in the above extracts. Arrows indicate epicatechin-phloroglucinol **11** and epicatechin **1** released from procyanidin B2 **8** and catechin-phloroglucinol and catechin released from procyanidin B3 **9** analyzed for comparison. In **a** and **b**, the representative results of three independent analyses with similar results obtained for each are presented.



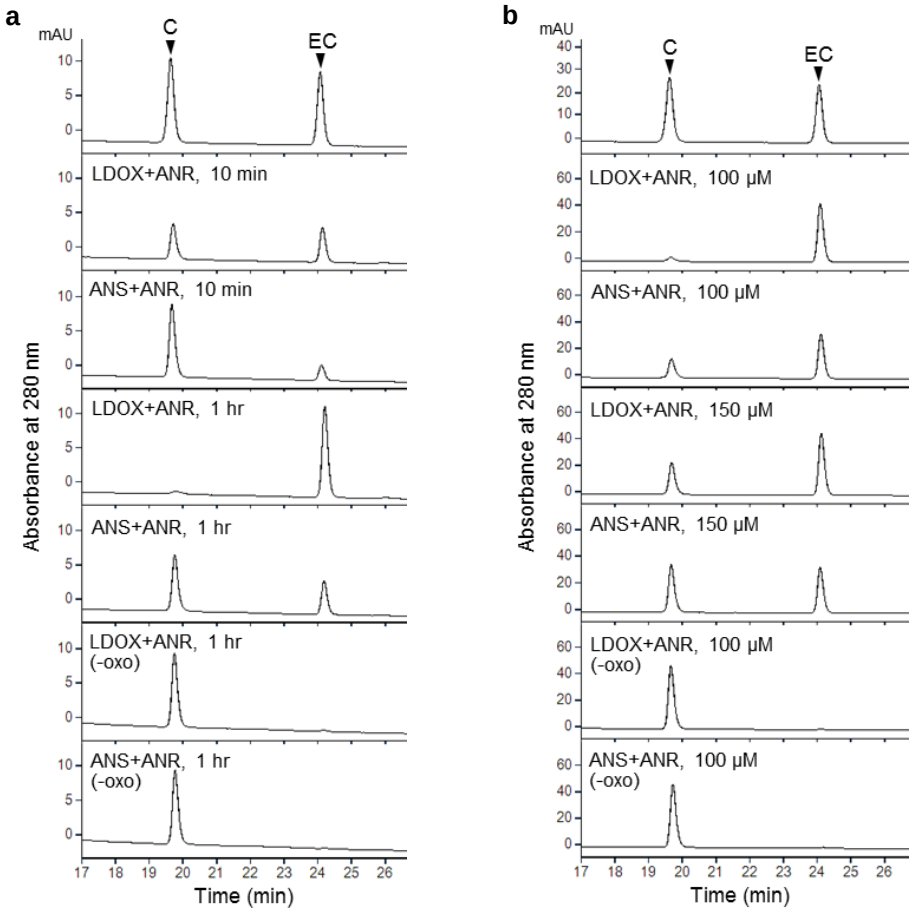
Supplementary Figure 11. In vitro enzymatic reactions of ANS and LDOX with leucocyanidin. a, Expression and purification of recombinant Mt ANS and Mt LDOX. SDS-PAGE gel of purified recombinant proteins fused with maltose binding protein (MBP) was stained with Coomassie blue. **b**, HPLC profile of reaction products with leucocyanidin **3** as substrate and recombinant ANS or LDOX proteins, monitored at 540 nm. Reactions without 2-oxoglutarate were run as negative controls and denoted as -oxo. **c**, As above, with HPLC profile monitored at 280 nm. **d**, Scheme of reactions of ANS and LDOX with leucocyanidin **3** *in vitro*. Arrows in **b** and **c** indicate the peaks corresponding to standards. In **b** and **c**, the representative results of three independent analyses with similar results obtained for each are presented. C, catechin; Cy, cyanidin; DHQ, dihydroquercetin; EC, epicatechin; Q, quercetin.



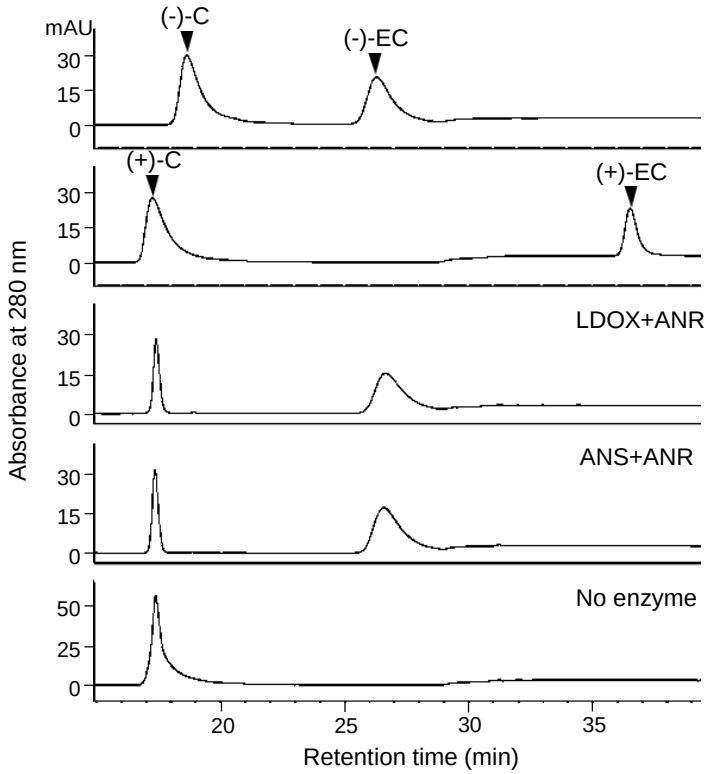
Supplementary Figure 12. Comparison of enzymatic reactions of ANS or LDOX with (+)-catechin as substrate. a, HPLC profiles at 540 nm for enzyme reaction performed with 200 µM (+)-catechin **2** as substrate and 3 µg recombinant ANS or LDOX protein for 10 min or 3 h. **b**, HPLC profiles at 540 nm for enzyme reactions performed with 50 µM or 150 µM (+)-catechin **2** as substrate and 500 ng recombinant ANS or LDOX protein for 1 h. Reactions without 2-oxoglutarate were run as negative controls and are denoted as -oxo. Arrows indicate the peak for cyanidin (Cy) **4** standard. In **a** and **b**, the representative results of three independent analyses with similar results obtained for each are presented.



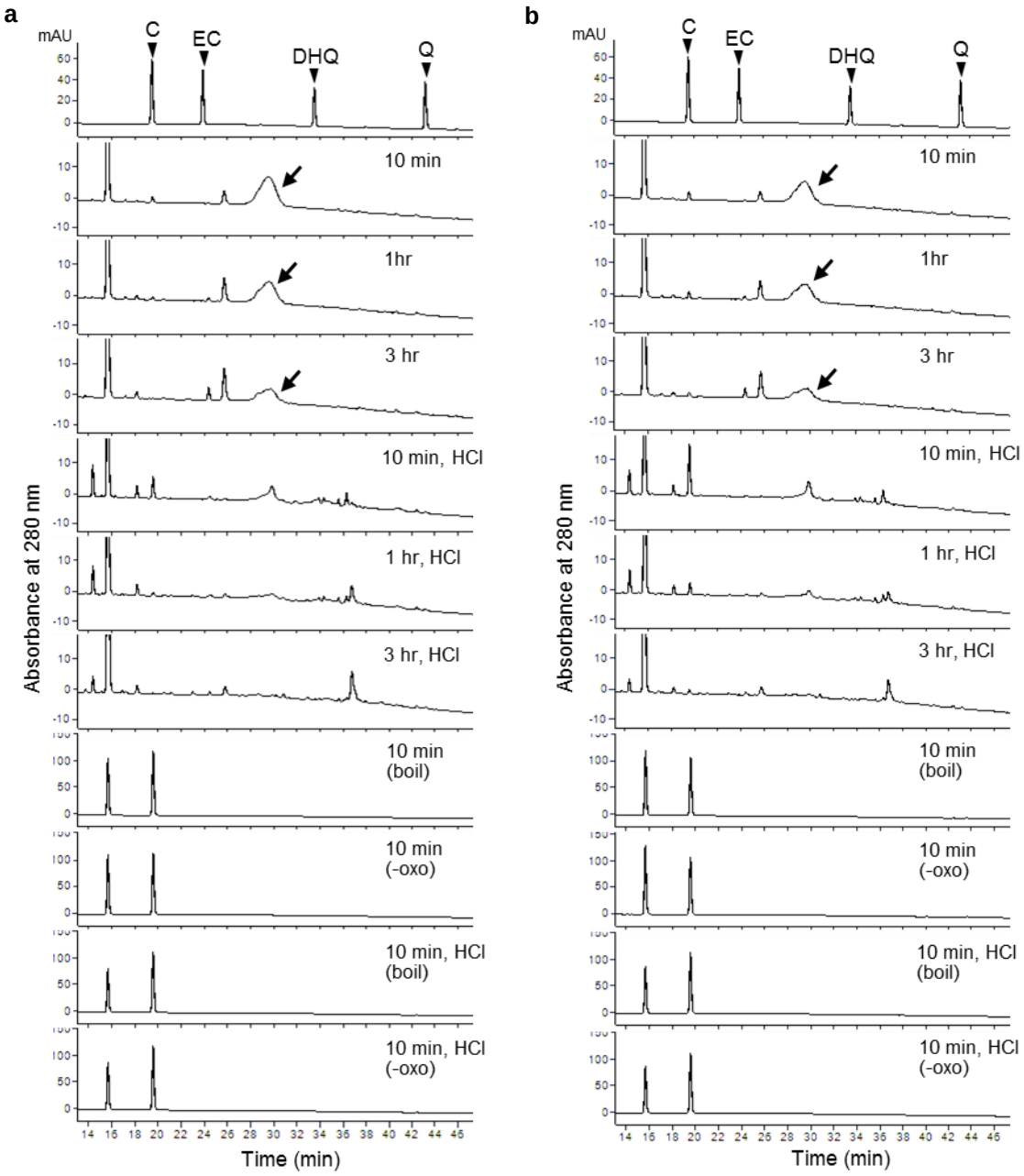
Supplementary Figure 13. Enzymatic reactions of ANS and LDOX with (-)-catechin, (+)-epicatechin and (-)-epicatechin. The enzyme reaction was performed with 200 μM (-)-catechin, (+)-epicatechin or (-)-epicatechin **1** as substrate and 3 μg recombinant ANS or LDOX protein for 3 h. HPLC profiles were monitored at 540 nm. The representative results of three independent analyses with similar results obtained for each are presented.



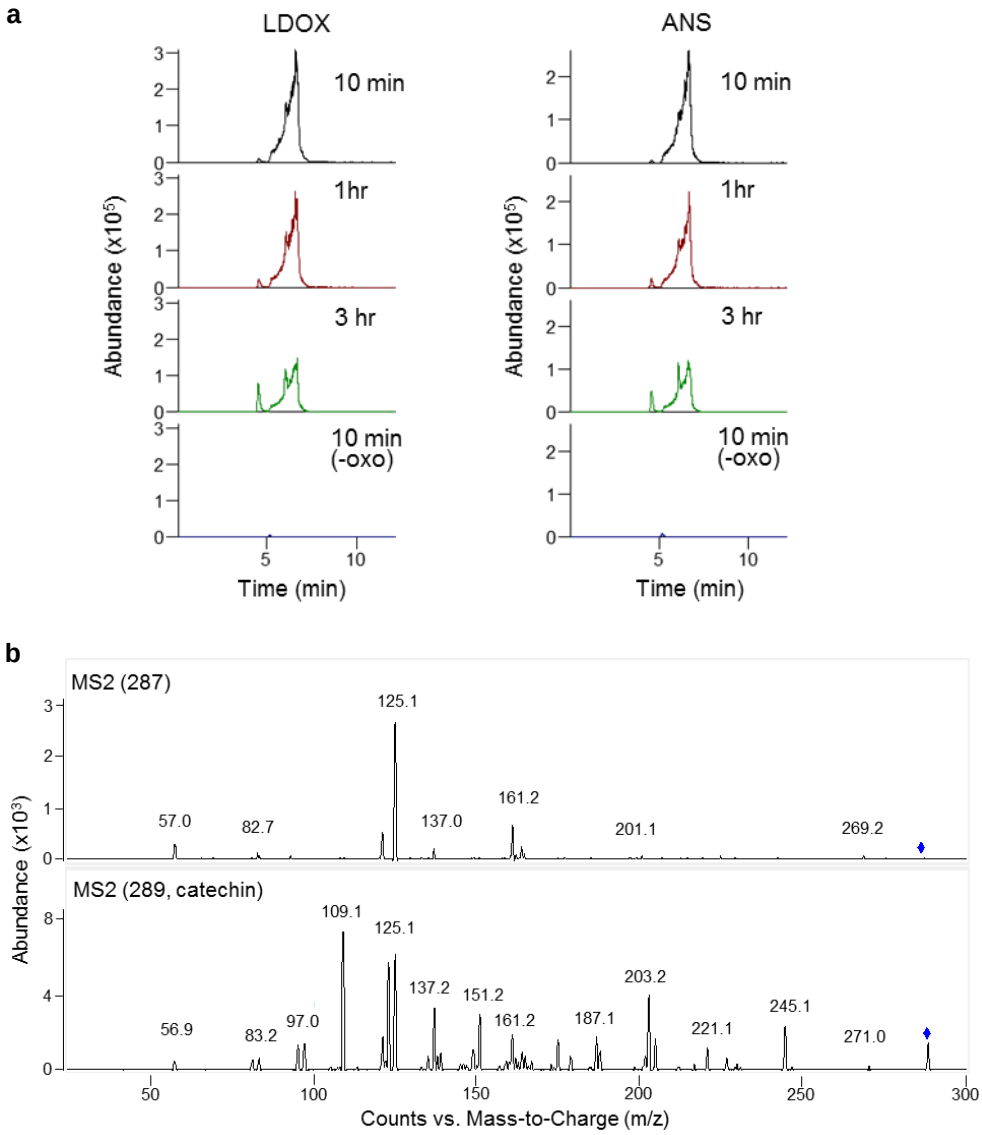
Supplementary Figure 14. Comparison of coupled enzymatic reactions of ANS or LDOX with ANR, with (+)-catechin as initial substrate. a, HPLC profiles at 280 nm for the enzyme reaction performed with 20 µM (+)-catechin **2** as substrate and 200 ng recombinant ANS or LDOX combined with 200 ng ANR protein for 10 min or 1 hr. **b**, HPLC profiles at 280 nm for the enzyme assay performed with 100 µM or 150 µM (+)-catechin **2** as substrate and 1 µg recombinant ANS or LDOX combined with 1 µg ANR protein for 10 min. Reactions without 2-oxoglutarate were run as negative controls and are denoted as -oxo. Arrows indicate the peak for (+)-catechin (C) **2** and (-)-epicatechin (EC) **1** standards. In **a** and **b**, the representative results of three independent analyses with similar results obtained for each are presented.



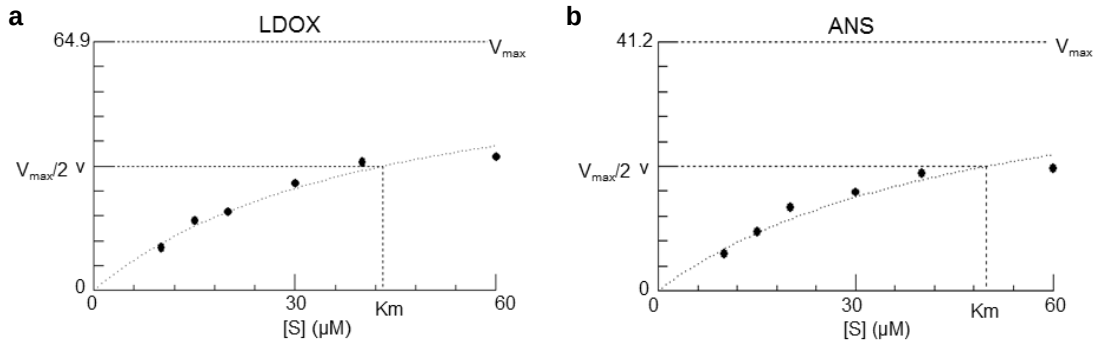
Supplementary Figure 15. Chiral chromatographic analysis of coupled in vitro enzymatic reactions of ANS or LDOX with ANR, with (+) catechin as initial substrate. HPLC profile from chromatography on a chiral column was monitored at 280 nm. The first and second chromatograms show standards of (-)-catechin, (-)-epicatechin, (+)-catechin and (+)-epicatechin. The representative results of three independent analyses with similar results obtained for each are presented.



Supplementary Figure 16. Analysis of the chemical intermediate generated in ANS or LDOX enzyme reactions, with (+)-catechin as initial substrate. a, HPLC profiles at 280 nm for the enzyme reaction performed with 200 µM (+)-catechin **2** as substrate and 3 µg recombinant LDOX protein for 10 min, 1 h or 3 h. **b**, HPLC profiles at 280 nm for the enzyme reaction performed with 200 µM (+)-catechin **2** as substrate and 3 µg recombinant ANS protein for 10 min, 1 h or 3 h. Negative controls were reactions without 2-oxoglutarate (-oxo) and using boiled enzyme (boil). Reactions were also terminated by addition of 1 µl 36% HCl and 10 µl 100% methanol (HCl). Arrows indicate transient accumulation of intermediate. Arrowheads indicate the peaks corresponding to standards. In **a** and **b**, the representative results of three independent analyses with similar results obtained for each are presented. C, catechin; EC, epicatechin; DHQ, dihydroquercetin; Q, quercetin.



Supplementary Figure 17. Detection of flav-2-en-3-ol as an intermediate in the generation of cyanidin from (+)-catechin catalyzed by ANS or LDOX in vitro. a, EIC of flav-2-en-3-ol (m/z 287.0561 ± 20 ppm) with the enzyme reaction performed with 200 µM (+)-catechin **2** as substrate and 3 µg recombinant LDOX or ANS protein for 10 min, 1 h or 3 h. Reactions without 2-oxoglutarate were run as negative controls and are denoted as -oxo. **b**, MS/MS spectra of the intermediate. The fragmentation of purified intermediate compound was analyzed by a triple quadrupole mass spectrometer. The Fragmentation pattern of catechin was analyzed for comparison. The peak from catechin displayed fragments at m/z 271, 245, 203, and 137 corresponding to the loss of H2O, CO2, C4H6O2, and C8H8O3 fragments, respectively. Fragmentation of the purified intermediate showed m/z 269 and 201 and the same smaller ions as seen in the fragmentation of catechin, indicating that the intermediate has the same structure as catechin but with loss of two hydrogens. In **a** and **b**, the representative results of three independent analyses with similar results obtained for each are presented.

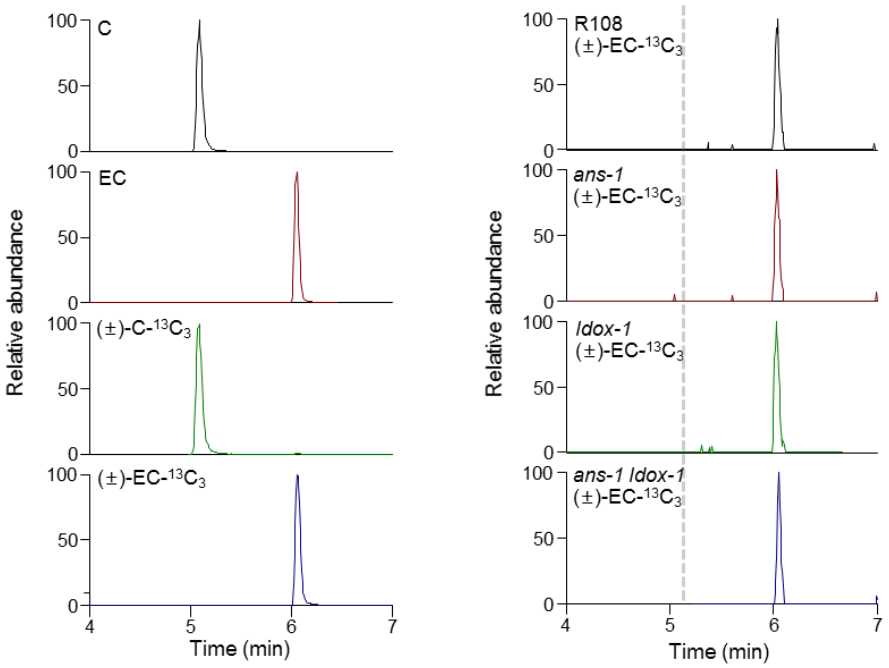


c

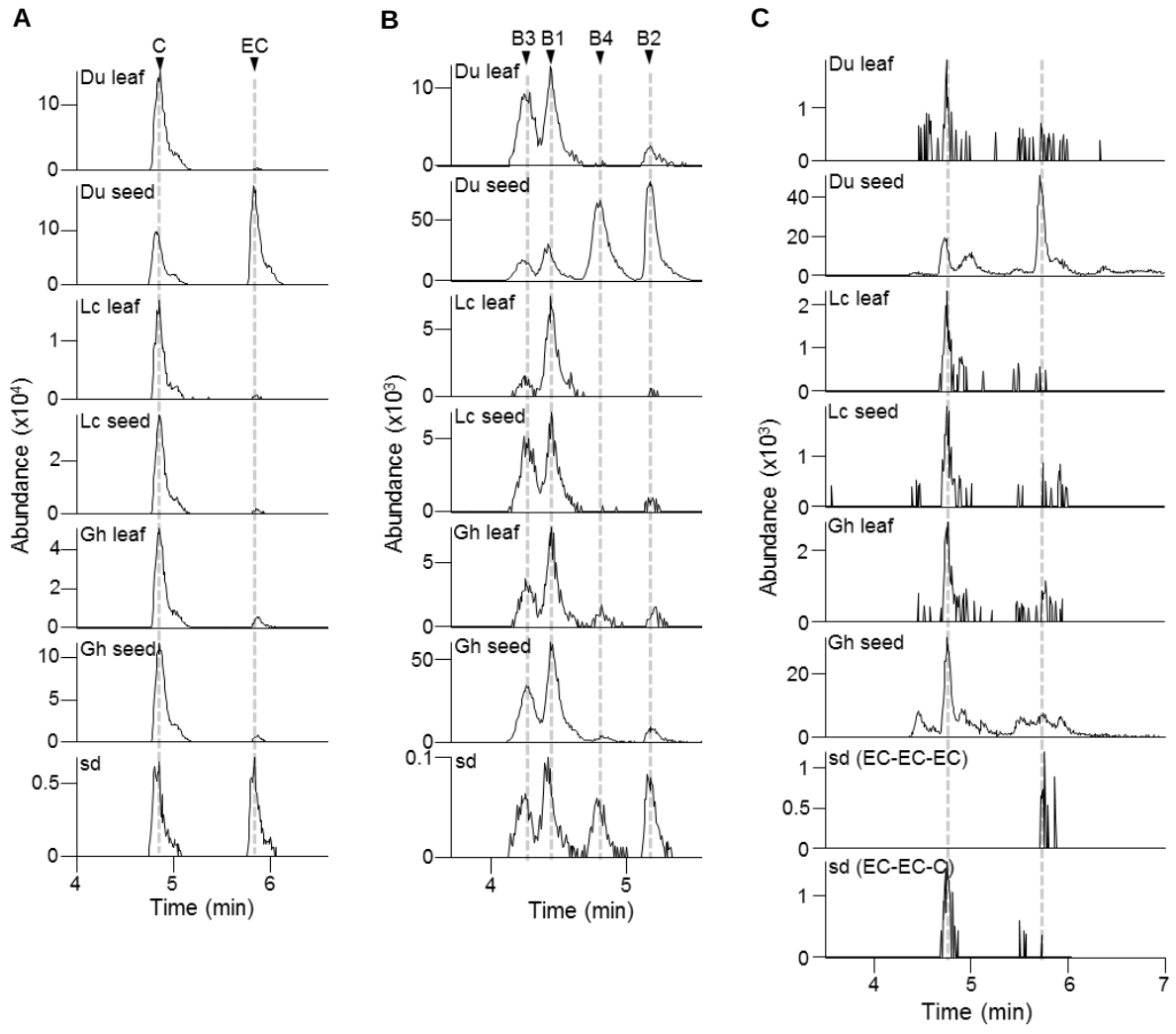
Kinetic parameter of recombinant ANS and LDOX enzyme

Enzyme	K_m (μM)	V_{max} (nkat mg protein^{-1})	K_{cat} (s^{-1})	K_{cat}/K_m ($s^{-1}\mu M^{-1}$)
LDOX	43.17 ± 6.43	67.89 ± 9.16	5.43	0.13
ANS	58.62 ± 7.75	40.94 ± 0.86	3.37	0.07

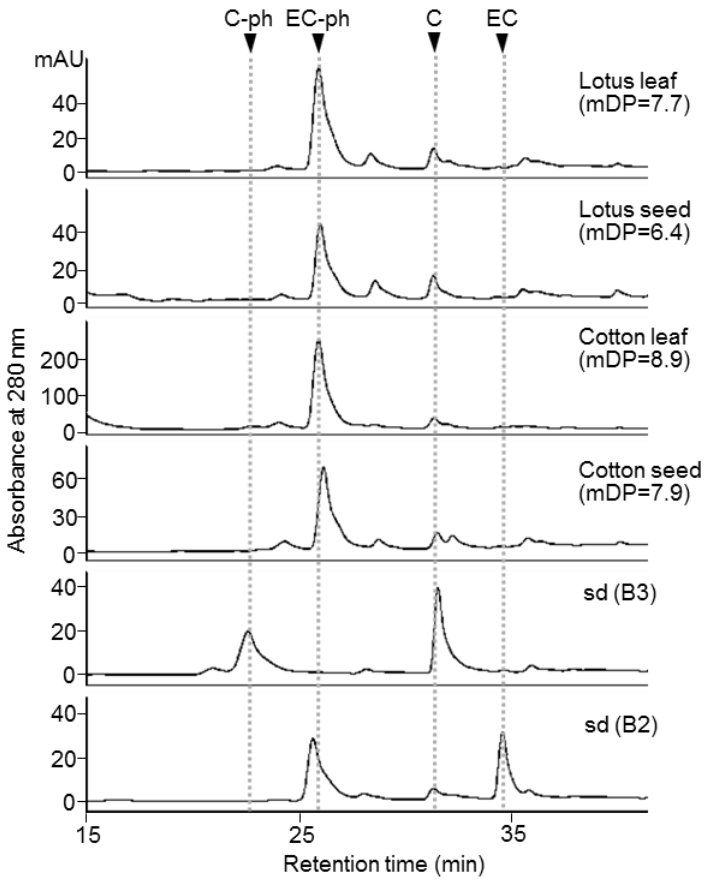
Supplementary Figure 18. Kinetic analysis of recombinant LDOX and ANS with (+)-catechin as substrate. a, Plot of initial velocity at different (+)-catechin **2** concentrations for LDOX. **b**, Plot of initial velocity at different (+)-catechin **2** concentrations for ANS. **c**, Kinetic parameters of recombinant LDOX and ANS for conversion of (+)-catechin **2** to flav-2-en-3-ol **14** intermediate. Kinetic constants were determined in a total volume of 100 μL containing 300 ng recombinant proteins and 0-60 μM (+)-catechin **2** substrate with 5 min incubation after termination by addition of 10 μl 100% methanol. Hyperbolic regression results was calculated and plotted by using Hyper32. **c**, Kinetic parameter of recombinant ANS and LDOX enzyme. The data are means ±SD of biological triplicates. Representative results of triple independent experiments are presented in **a** and **b**.



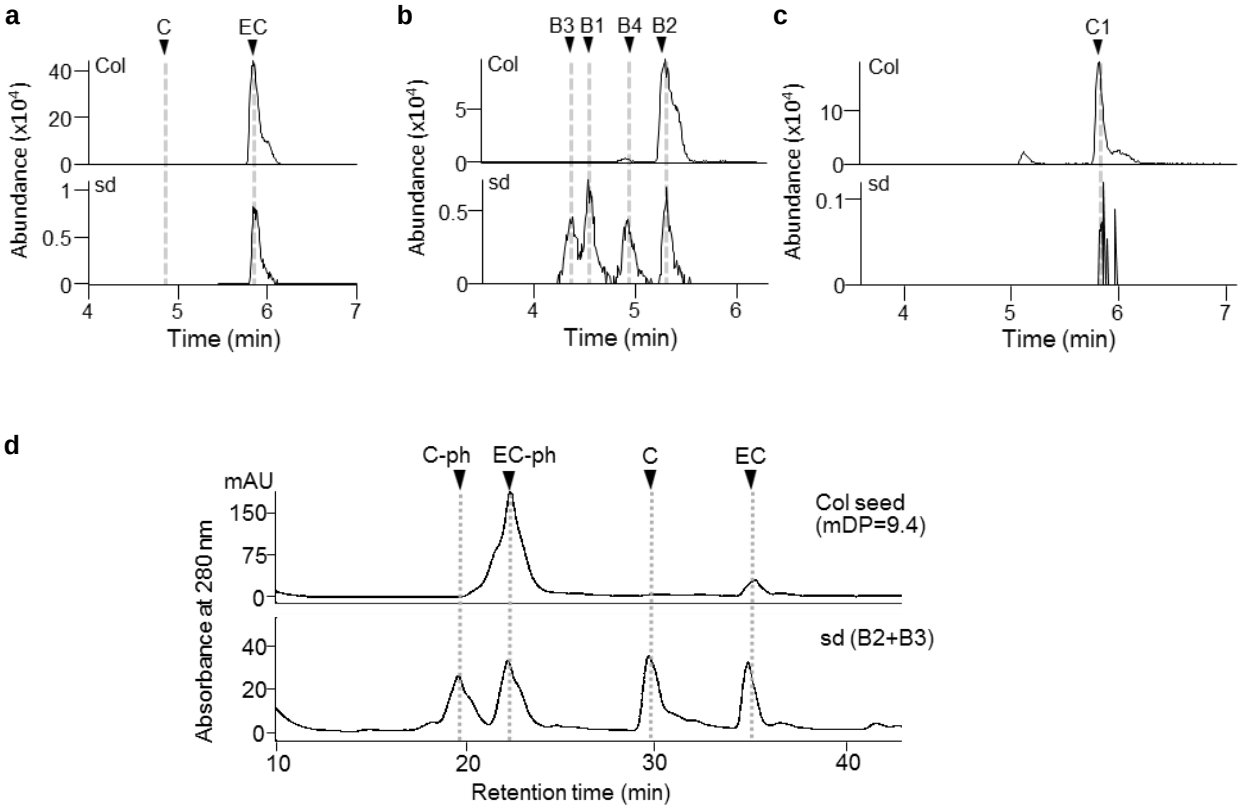
Supplementary Figure 19. Lack of conversion of (±)-epicatechin-2,3,4-$^{13}C_3$ to catechin in pods of wild-type and mutant *M. truncatula* at 2-4 DAP. The EICs of catechin or epicatechin (m/z 289.0718 ± 10 ppm) and catechin-2,3,4-$^{13}C_3$ or epicatechin-2,3,4-$^{13}C_3$ (m/z 292.0814 ± 10 ppm) are presented in the left panels. Incubation of pods with (±)-epicatechin-2,3,4-$^{13}C_3$ is denoted under the genotype in each chromatogram (right panel). The retention time of (+)-catechin **2** is denoted by the grey dashed line. The representative results of three independent analyses with similar results obtained for each are presented.



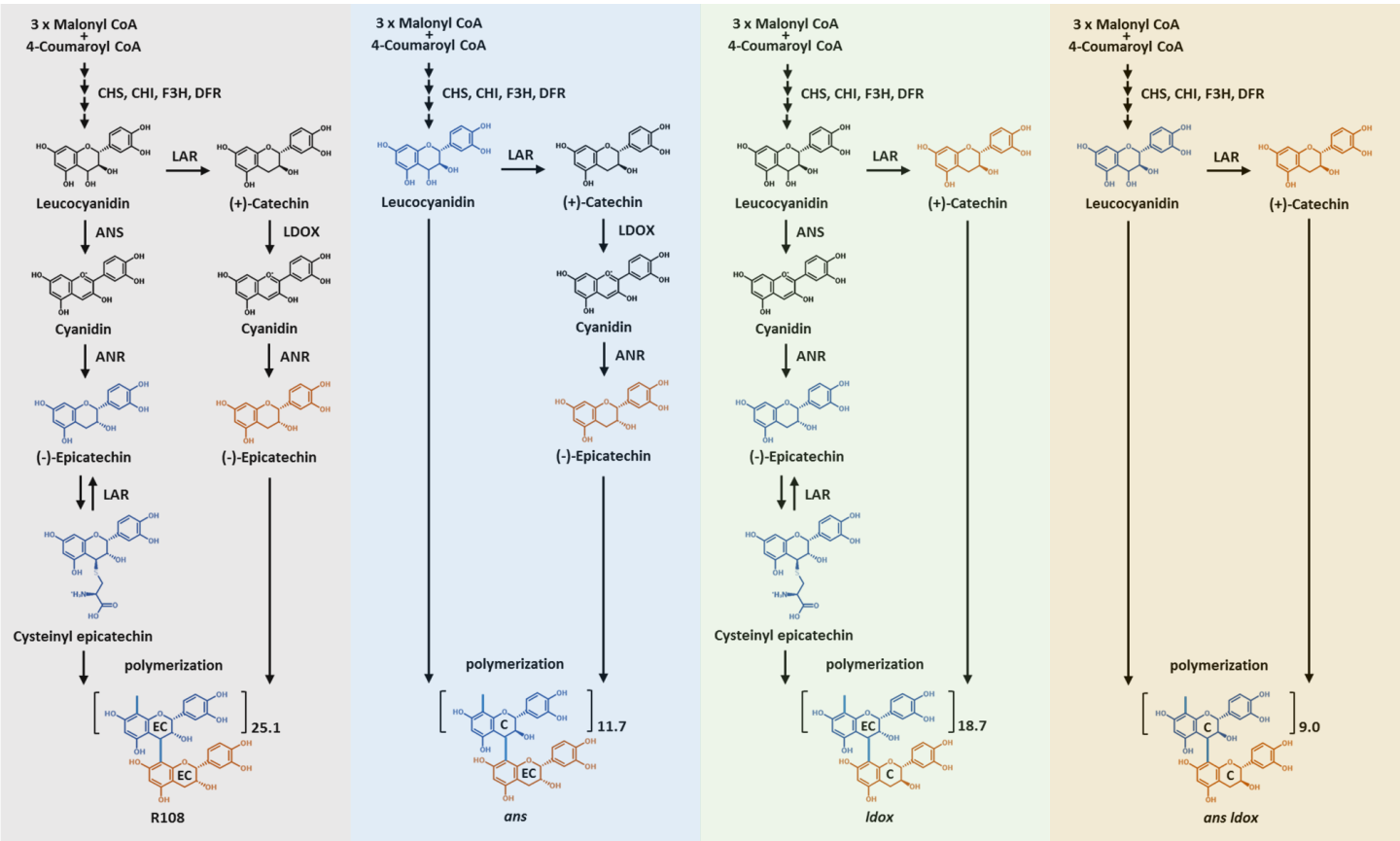
Supplementary Figure 20. Monomer, dimer and trimer composition of soluble PAs from *Desmodium uncinatum*, *Lotus corniculatus* and *Gossypium hirsutum.* **a**, Extracted ion chromatograms (EICs) of epicatechin and catechin monomers (m/z 289.0718 ± 20 ppm) in soluble PAs extracted from leaves or developing seeds. **b**, As above, showing EICs of procyanidin dimers (m/z 577.1348 ± 20 ppm). **c**, As above, showing EICs of trimers (m/z 865.1985±20 ppm). In **a**-**c**, the representative results of three independent analyses with similar results obtained for each are presented. C, catechin; EC, epicatechin; B1, procyanidin B1; B2, procyanidin B2; B3, procyanidin B3; B4, procyanidin B4; EC-EC-EC, epicatechin-epicatechin-epicatechin trimer; EC-EC-C, epicatechin-epicatechin-catechin trimer.



Supplementary Figure 21. Phloroglucinolysis of soluble PAs from *Lotus corniculatus* and *Gossypium hirsutum*. The bottom two panels show standards (sd) released from proanthocyanidins B2 **8** (epicatechin-epicatechin) and B3 **9** (catechin-catechin) respectively. Mean degree of polymerization (mDP) based on relative proportions of extension to starter units, are shown on the right. The representative results of three independent analyses with similar results obtained for each are presented. C-ph, catechin-phloroglucinol; EC-ph, epicatechin-phloroglucinol; C, catechin; EC, epicatechin.



Supplementary Figure 22. Monomer and dimer composition of soluble PAs from *Arabidosis thaliana* (Columbia). **a**, Extracted ion chromatograms (EICs) of epicatechin and catechin monomers (m/z 289.0718 ± 20 ppm) in soluble PAs extracted from young seeds. **b**, As above, showing EICs of procyanidin dimers (m/z 577.1348 ± 20 ppm). **c**, As above, showing EICs of trimers (m/z 865.1985±20 ppm). **d**, Phloroglucinolysis of soluble PAs in Arabidopsis seeds. The bottom panel shows standards released from procyanidins B2 **8** (epicatechin-epicatechin) and B3 **9** (catechin-catechin). Mean degree of polymerization (mDP) based on relative proportions of extension to starter units, is shown on the right. In **a-d**, the representative results of three independent analyses with similar results obtained for each are presented. Standards (sd) are: C-ph, catechin-phloroglucinol; EC-ph, epicatechin-phloroglucinol; C, catechin; EC, epicatechin; B1, procyanidin B1; B2, procyanidin B2; B3, procyanidin B3; B4, procyanidin B4; C1, procyanidin C1.



Supplementary Figure 23. Summary of PA biosynthesis in wild type (R108), *ans*, *ldox* and *ans ldox* double mutants. The separate contribution of ANS and LDOX proteins in the parallel pathways to provide starter and extension unit of PAs is proposed based on genetic and biochemical analysis of wild type (R108), *ans*, *ldox*, and *ans ldox* plants. The change of PA composition in each genotype is summarized and mDP of dry seeds is indicated. It is possible that leucocyanidin **3** and (+)-catechin **4** are incorporated as extension unit and starter unit in the absence of ANS and LDOX activity, respectively. The model explains how PAs are exclusively composed of epicatechin subunits in *Medicago truncatula* in spite of the expression of *LAR*. The *ans lar* double mutant produces PAs mainly composed of catechin extension units, indicating that leucocyanidin **3** can be directly conjugated into PA polymers in *ans* and *ans ldox* double mutant. Additionally, the increase of catechin monomer accumulation does not affect the composition of extension units in the *ldox* mutant, consistent with catechin production by LAR being mainly concerned with starter unit generation. Change of flux from leucocyanidin **4** through the ANS and LDOX pathways and/or availability of different substrates for conjugation (cysteinyl epicatechin **5**, leucocyanidin **4**, catechin and epicatechin) also affects polymer chain length (mDP), with catechin incorporation resulting in shorter chain length.

Supplementary Table 1. PA monomer and dimer accumulation in seeds of *ans-2* and *ldox-2 M. truncatula*, and in leaf and seeds of *Lotus corniculatus* (Lc) and *Gossypium hirsutum* (Gh).

Subunit	Flavan-3-ol monomer and dimer profile (% area)[a]					
	ans-2	*ldox-2*	Lc leaf	Lc seed	Gh leaf	Gh seed
Monomer						
Catechin (C)	23.74±1.32	94.11±0.33	97.81±0.32	96.36±0.92	91.97±0.26	94.98±0.76
Epicatechin (EC)	76.26±1.32	5.89±1.42	2.19±0.32	3.64±0.92	8.03±0.26	5.02±0.76
Dimer						
Procyanidin B1 (EC-C)	1.78±0.17	51.02±17.93	79±3.52	52.28±4.90	51.48±3.40	52.78±4.01
Procyanidin B2 (EC-EC)	5.31±3.52	5.35±3.25	3.18±0.81	6.07±1.38	7.53±0.21	6.57±0.41
Procyanidin B3 (C-C)	6.32±0.42	41.71±13.02	17.69±2.64	40.45±5.98	32.56±3.85	37.73±4.60
Procyanidin B4 (C-EC)	86.77±4.05	1.91±1.73	0.13±0.14	1.2±0.68	8.43±0.38	2.92±0.21

[a] The peak areas of EICs of epicatechin and catechin monomers (m/z 289.0718 ± 20 ppm) and procyanidin type B dimers (m/z 577.1348 ± 20 ppm) in soluble PAs extracted from young seeds of *M. truncatula*, *L. uncinatum* and *G. hirsutum* and leaf of *L. uncinatum* and *G. hirsutum* were measured by LC/MS and calculated as % value of total monomers and dimers. The data are means ±SD of biological triplicates.

Supplementary Table 2. RNA sequencing reads and de novo assembly information for *D. uncinatum* transcriptome

Sample	Seq_name	Raw_reads	Trimmed_reads	Assembly_transcripts
D. Uncinatum seed	index15_S3	58555332	58289358	68876

Next generation sequencing was performed using the Illumina NextSeq 500 platform with a mid-output 2x150 paired-end reagent cartridge. Approximately 58M raw reads were produced. De novo assembly of transcriptomes was performed using Trinity with the default parameters [41].

Supplementary Table 3. *D. uncinatum* cDNA sequences identified in this study.

Name	ID of assembly	cDNA sequence
Du F3H	comp23389_c0_seq1	ATGGCTCCAACCGCCAAGACTCTAACTTTCTTGGCTCAGGAAAAGACCCTCGAGTCCAGTTTCGTTCGAGACGAGACCGAGCGCCCTAAGGTTCCGTA CAATGAGTTCAGCAACGAGATTCCGGTGATTTCACTCGATGGAATCGACGAAGTGGATGGACGCAGAACCGAAATTCGCAGGAAGATCGTTGAGGC TTGTGAAAACTGGGGTATATTTCAGGTTGTGGATCACGGTGTGGACACTGAACTCGTTTCTCAGATGACTCGTCTTGCCAAAGAGTTCTTCGCTTTACC ACCCGATGAGAAGCTTCGCTTTGACATGTCTGGTGGTAAAAAGGGTGGCTTCATTGTCTCCAGCCATCTCCAAGGTGAATCGGTGCAGGACTGGAGA GAGATCGTGACATACTTCTCATACCCAATAAGAGAGAGAGACTATTCGCGGTGGCCAGACACACCAGAAGGGTGGAAAAAAGTGACTGAAGAATAC AGCGATAAACTCATGGATTTGGCGTGTAAACTGTTGGAGGTGTTGTCGGAGGCAATGGGTTTGGAGAAAGATGCTCTTACGAAAGCGTGTGTTGATA TGGACCAGAAGATTGTGGTGAATTATTACCCTAAGTGTCCTCAACCTGATCTAACTCTTGGGCTGAAGCGCCACACTGACCCCGGCACCATCACGCTC TTGCTTCAAGACCAAGTTGGTGGACTTCAGGCCACTAGGGATAATGGCAAAACTTGGATCACTGTTCAGCCTGTCGACGGTGCTTACGTTGTTAATCT TGGAGATCATGGTCACTTTCTGAGCAATGGAAGGTTCAAGAATGCTGATCACCAAGCAGTGGTGAACTCAAACCATAGTCGTTTGTCCATAGCCACAT TTCAGAACCCAGCACCAAATGCGACTGTGTACCCTTTGAAGATCAGAGAGGGTGAAAAACCTGTGCTTGAAGAACCAATCACTTTTGCTGAAATGTAC AGGAGAAAGATGAGCAAGGACCTTGAGCTTGCTAGGCTCAAGAAACTCGCCAAAGAAAAGCAATTGCAGGAACAACTTGACAAAGAAAAACAATTG CAGGCACAACTTGACAAAGAAAAAGAATTACAAGATCTTCCAAAGCCAAAACTTGAGACCAAGCCTTTGAATGAGATTCTTGCTTAG
Du FS1 (partial)	comp36193_c0_seq1	ATGGACACAAAGGTTCTTTCTTCTGGAATCCACTACTCAAATCTTCCGGAGAGTTACATCCGACCCGAATCGGAACGACCTCGACTCTCGGAAGTTTCC GAGTGTGAGGATGTTCCTGTCATCGACTTGAGTTCTCAAAACAGACCCCAGATTGTTCACCAAATCGGTGAAGCTTGCAGGTCCTATGGCTTTTTCCA GGTAATTAATCACGGGGTGGCTGTAGAAGCGGTTAAGGAAATGACAGAGGTGGCACATGGATTCTTCAAATTGCCAGTGGAGGAGAAGCTGAAGCT GTACTCTGAAGATCCATCAAAGACAATGAGATTGTCAACTAGTTTTAATGTGAAAAAAGAGAAGGTGCATAATTGGAGGGATTATCTTAGACTCCATT GCTATCCATTGGACAAGTATGTGCCTGAATGGCCTTCTAATCCTCCCTCTTTCAAGGAAAGCGTGAGAAGCTACTGCAAAGAAGTTCGTGAATTGGGA TTCAGAATACAAGAATACATATCAGAGAGTTTGGGCTAA
Du ANS	comp23380_c0_seq1	ATGGGAACGTTGGTTCCTAGTCCAAGAGTTGAAACCTTGGCGAGAAGTGGGATTGAGTGTATCCCTAAGGAGTATGTGAGGCCACAAGAAGAGTTG AATAGCATTGGGAATGTGTTTGAGGAGGAAAAGAAAGAGGGTCCTCAAGTTCCAACAATTGACTTGAGAGATATTGACTCTGAGGATTTTGTTGTGA GAGCTAAATGCCGAGAGAAGCTGAAGAAAGCAGCCACAGATTGGGGGGTCATGAACTTGGTTAACCATGGCATCCCTGATGAGCTATTGAACAAGC TTAGGAAAGCTGGTGAAACCTTTTTCAGTCTTCCAATTGAAGAAAAAGAGAAGTATGCGAATGATCAAGGTTCTGGAAAGATTCAAGGCTATGGAAG CAAACTAGCTAACAATGCCAGTGGTCAACTTGAATGGGAAGATTACTTCTTTCACCTCATTTTCCCTGAAGACAAACGTGACTTGTCCATCTGGCCAAA CACACCCCCAGATTACACTGAGGTTACAAGCGAGTATGCAAAGCAACTGAGGGTTCTGGCAAGCAAGATACTGGAGGCGCTATCTGTTGCATTGGGT TTAGAAGGAGGGAGGCTTGAGAAAGAAGTTGGTGGAATGGAAGAGCTTTTGCTTCAATTGAAGATCAACTATTACCCTATTTGTCCCCAGCCTGAAC TTGCTCTTGGAGTTGAAGCTCACACTGACGTAAGTTCACTCACTTTCCTTGTCCATAATATGGTCCCGGGTCTCCAACTTTTCTATCAAGGTCAATGGAT CACAGCAAAATGCGTTCCCAATTCCATTTTCATGCACATTGGTGATACCCTTGAGATTCTCAGCAATGGCAAGTACAAGAGTATTCTTCATAGGGGTTT GGTGAACAAGGAAAAGGTTAGAATATCATGGGCAGTGTTCTGTGAACCCCCCAAGGAGAAGATCATTCTGCAACCACTTCCTGAACTTGTCACTGAA ACAGAACCTGCACGGTTTCCTCCTCGCACTTTTGCTCAGCATATTCATCACAAACTCTTCAGGAAGGATCAGGAAGATCCCCCAAAGTGA
Du LDOX	comp23594_c0_seq1	ATGGAAGTGGAAAGAATACAAAGTTTGGCTTTGAACCAAATCAAGGAACTTCCACCCCAGTTCATCCGTTCAGCCAATGAAAGGCCAGGAAACACCA AAGCCATAGAAGGTGTGACTGTGCCTCTCATCTCATTGTGTCAGCCTCATGATATTTTGGTCAAGGAAATATCTGAAGCTGCTTCTCAATGGGGATTTT TCCTCCTAACTGATCATGGTATATCACAAACATTGATCCAACGTTTGCAAGATGTTGGTAAGGAGTTCTTTGCACTCCCTCAACAAGAGAAAGAGGTTT ATTCAAATGACCCTTCTAAGGGTGAGTTTGAGGGCTATGGCACAAAAATGACAAAAAACCTTGAAGAAAAGATGGAGTGGGTTGATTATTTTTTTCAT TTCATTGCTCCTCCTTCAAAGGTCAATTATGACATGTGGCCCAAAACCCCTTCTTCCTACAGGGAAGTGAATGAAGAATACAATAAAGAGATGTTAAG GGTAACAGACAAGGTCTTGGAGCTTTTGTCTGAAGGGTTGGGCCTAGAGAAGAAGGTTTTGAAGTCCAATTTGGGAGATGAAGAAATAGAACTAGA AATGAAGATAAATATGTATCCACCATGTCCACAACCAGATTTGGCTTTGGGAGTTGAGGCTCATACTGACATGAGTGCTCTTACCCTATTGGTCCCAAA CGATGTTCCTGGCCTACAAGTTTGGAAGGACAATGCTTGGGTCGCAGTCAATTATGTGCCAAATGCACTGTTCATACACGTTGGTGATCAACTTGAGG TGTTGAGTAATGGAAAGTACCAGAGTGTTCTACATCGAAGTTTGGTGAACGAGGAAAGAATGCGCATGTCATGGGCAGTGTTTGCTGTGCCACCACT TGAGGCAGTGATAGGGCCTCTTCCTTCTCTCATCAATCATCAAAATCCATCCAAGTTTTCAACAAAAACTTATGCTGAGTATCGCCATCGCAAATTCAA CAAGCTTCCCCAATAA
Du FLS	comp24814_c0_seq1	ATGGAGGTGCTAAGGGTGCAAACCATAGCTTCAGAGTCCAAAGATGCTACTATCCCTGCCATGTTTCTTAGGCCAGAGACAGAACAACCAGGCACCA CAACGGTTCATGGGGTGAATCTTGAGGTACCCATAATTGATTTTAGCAACCCTGACAAAGAGAAAGTGGTTCGAGAAATTGTGGAGGCAAGTAGGG ATTGGGGTATGTTCCAAATTGTGAACCATGACATACCAAATCATGTTATAAGGAAATTGCAAAGTGTGGGCAAAGAGTTCTTTGAGTTACCACAAGAA GAGAAAGAACTATATGCTAAGCTTCCAGGATCTAATTCTTTAGAAGGGTATGGCACAAAACTTCAAAAAGAGTTGAATGACAAGAAAGGGTGGGTG GATCATTTATTCCACATTATATGGCCTCCTTCTTCCATTAACTATCGTTTCTGGCCTAAGAATCCCTCTTCTTACAGGGAGGTGAATGAGGAATACTGCA AGTACCTACGTTGGGTGGTAGAGAAACTGTTGAAAAGCATGTCAATAGGGCTAGAACTTGAAGAAAACGAGTTAAAGGAAGCTGCAAATGAAGATG ACATGCAATACCTGTTAAAGATAAACTATTATCCACCTTGTCCATGTCCTGATTTGGTCCTGGGTGTGCCACCACACACAGACATGTCGTACCTCACAA TTCTGGTCCCCAATGAGGTTCAGGGGCTTCAAGCCTCTAGGGATGGTCACTGGTACGATGTTAAGTATGTCTCCAACGCCCTCGTCGTTCACATTGGC GATCAAATGGAGATACTGAGCAATGGGAAATATAAGGCAGTTTTTCACAGAACAACAGTGAACAAAGAAGAGACAAGAATGTCATGGCCAGTGTTC ATAGAGCCGAAAGCAGAGAATCTAGTTGGTCCTCACCCAAAATTGGTCAACCAAGACAATCCACCAAAATACAAGACCAAGAAATACAAGGATTATG CTTATTGTAAGCTCAATAAGATCCCTCAGTGA
Du ANR	comp23356_c0_seq1	ATGGGCAGCATCGAGCCAATCGGAAAGAAGGCATGTGTGATCGGTGGTAGTGGATTTGTGGCCTCTTTGTTGATCAAGCAGTTGCTTGAGAAGGGTT ATGCAGTTAATACAACTGTTAGAGACCCAGATAATGCTAAAAAAATATCTCACCTTCTGGCACTGCAAAGTTTGGGCGAATTGAACATATTTGGAGCA GATTTAACTGGTGAAAAAGATTTTGATGCCCCTGTAGCTGGCTGTGAACTTGTCTTCCAGCTCGCCACACCTGTGAACTTTGCTTCTGAAGATCCTGAG AATGACATGATCAAGCCTGCAATCTCAGGCGTATTGAATGTGTTGAAAGCATGTGCGCGGGCAAAAGATGTCAAACGAGTCATCTTAACATCTTCTGC AGCTGCAATAACCATAAGCGATGTCAAGGGGACTGGTCTGGTTATGGATGAAAGCAACTGGACTGATGTTGAGTTCTTGAGCACTGCAAAGCCACCC ACTTGGGGGTACCCCGTCTCCAAAACACTAGCTGAGAAGACAGCATGGAAATTTGCTGAAGAAAATCACATTGATCTCGTCACTGTAATTCCTTCTCT CACAACTGGTCCTTCTGTCACTCCAGACATCCCTTCAAGTGTTGGCCTCGCCACGTCCCTATAACAGGCAATGATTTCCTCATAAACGCTTTGAAAGG CATGCAGATGCTTTCTGGTTCAATATCCATCACTCATGTGGAGGATATTTGCCGAGCACATGTATTCGTAGCAGAGAAAGAATCCGCTTCTGGTAGAT ACATTTGCTGTGCTCACAATACCAGTGTTCCTGAGCTTGCAAAGTTTCTTAGCAAACGATACCCTCAGTATAAAATTCCAACTGAATTTGGTGATTTCCC CTCAAAGGCAAAGGTAATTGTCTCTTCTGAAAAGCTTGTCAAAGAAGGATTCAATTTCAAGTATGGGGTTGAAGAAATTTATGATCAGACTGTGGAGT ACTTGAAGAGCAAAGGGGCTCTAAAGAACTGA
Du LAR	comp20507_c0_seq1	ATGACGGTATCGGCTGCAATTCCTTCAATGACCAAGAACCGAACTTTGGTGATCGGAGGAACTGGGTTCATAGGTCAGTTCGTAACTGAGGCAAGTC TTGCCTTTGGGTACCCTACCTTTTTGCTCGTAAGGCCAGGACCTGTCTCACCTTCCAAGGCTGTCATGATCAAAACCTTTCAAGACAAAGGTGCTACGG TTATCTATGGCGTAATTAATGACAAGGAATGCATGGAGAAGATTTTGAAGGAGTACGAGATTGATGTCGTCATTTCTCTTGTAGGAGGCGCCCGACTA TTGGATCAGCTTACCTTGTTGGAGGCCATAAAATCTGTGAAGACTGTCCAGAGGTTTCTGCCATCAGAGTTTGGGCACGATGTGGATAGGACAGATC CTGTAGAGCCAGGATTGACAATGTACAAAGAGAAGCGTCTGGTTAGGCGTACTGTTGAGGAATATGGGATTCCTTTCACCAACATTTGCTGCAACTCC ATTGCTTCTTGGCCTTATTATGACAATTGTCACCCTTCCCAGGTCCCTCCACCCATGGATCAGTTTCAAATCTATGGTGATGGCAACACCAAAGCTTACT TCATTGATGGCAATGATATTGGAAAGTTCACAATGAAGGCCATTGATGATATCAGAACACTGAACAAAAATGTTCATTTTCGACCCTCGAGCAACTGT TATTCCATCAATGAACTTGCTTCTTTATGGGAAAAGAAAATTGGACGTACACTTCCCAGATTCACCGTAACAGCGGATAAACTTCTTGCTCATGCTGCA GAAAATATTATACCAGAAAGTATTGTATCATCGTTCACCCATGATATTTTCATCAACGGTTGCCAAGTTAACTTCAGCATAGATGAACATAGTGATGTT GAGATTGACACACTCTATCCAGATGAAAAATTTCGATCCTTGGAGGATTGCTATGAGGACTTTGTTCCCATGGTCCATGACAAGATTCATACAGGAAA AAGTGGAGAAATTAAAATTAAAGATAGAAAGCCCTTGGTAGAGACCGGAACAATTGAAGAAATTAATAAGGACATAAAGACTTTGGTAGAGGCAGA ACCAAATGAAGAAATTAACAAGGATATGAAGCCTTTGGTAGAGGCAGTGCCAATTTCAGCTATGGGCTAG

Supplementary Table 4. Sequences of primers used in this work.

Primer name	Sequence (5'-3')	Purpose
ANS-F	ATGGGAACGGTGGCTCAAAGAGTTGA	RT-PCR
ANS-R	TCATTTTTTGGGATCATCTTTCTTCTCCT	RT-PCR
LDOX_F	ATGGAAGTCAAAAGAGTACAAACTTTAGCT	RT-PCR
LDOX_R	CTACTGGGGAATCTTGTTGAATTTGC	RT-PCR
Tub-F	TTTGCTCCTCTTACATCCCGTG	RT-PCR
Tub-R	GCCATGGAGAGACCTCTAGG	RT-PCR
Tnt1-F	ACAGTGCTACCTCCTCTGGATG	genotyping
ANS-Tn-F	ATGGGAACGGTGGCTCAAAGA	genotyping
Tnt1-R	TGTAGCACCGAGATACGGTAATTAACAAGA	genotyping
LDOX-Tn-F	GGAAGTCAAAAGAGTACAAACT	genotyping
qDuANS_F	GGGTCCTCAAGTTCCAACAA	RT-qPCR
qDuANS_R	ATCTGTGGCTGCTTTCTTCA	RT-qPCR
qDuLDOX_F	TGTATCCACCATGTCCACAAC	RT-qPCR
qDuLDOX_R	TAGGCCAGGAACATCGTTTG	RT-qPCR
qDuANR-F	CCATCACTCATGTGGAGGATATT	RT-qPCR
qDuANR-R	AGCTCAGGAACACTGGTATTG	RT-qPCR
qDuLAR-F	GACCCTCGAGCAACTGTTATT	RT-qPCR
qDuLAR-R	GCAGCATGAGCAAGAAGTTTAT	RT-qPCR
qRPOC-F1	GGGATTCAATGCAGACTTTGATG	RT-qPCR
qRPOC-R1	CTGGAGACAAGAGGTTCGTATG	RT-qPCR