Supplementary Information

The *Euscaphis japonica* genome and the evolution of malvids

Wei-Hong Sun[1,2,3], Zhen Li[4,5], Shuang Xiang[1,2,3], Lin Ni[2], Diyang, Zhang[3], De-Qiang Chen[1,2,3], Meng-Yuan Qiu[1,2,3], Qi-Gong Zhang[1,2], Lin Xiao[1,2], Le Din[1,2,3], Yifan Li[1,2], Xing-Yu, Liao[3], Xue-Die, Liu[3], Yu-Ting Jiang[3], Pei-Lan Zhang[1,2,3], Hui Ni[1,2], Yifan Wang[1,2], Yi-Xun Yue[1,2], Xi Wu[1,2], Xiang-Qing Din[2,3], Wei Huang[1,2], Zhi-Wen Wang[6], Xiaokai Ma[1,3], Bobin, Liu[1,2], Xiao-Xing Zou[1,2], Yves Van de Peer[4,5,7,8†], Zhong-Jian Liu[3†], Shuang-Quan Zou[1,2,3†]

[1] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

[2] Fujian Colleges and Universities Engineering Research Institute of Conservation and Utilization of Natural Bioresources, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

[3] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

[4] Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium;

[5] VIB Center for Plant Systems Biology, 9052 Gent, Belgium;

[6] PubBio-Tech, Wuhan 430070, China;

[7] Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa;

[8] Academy for Advanced Interdisciplinary Studies and College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China.

†Correspondence and requests for materials should be addressed to Z-J. L. (zjliu@fafu.edu.cn), Y. V. d. P. (yves.vandepeer@psb.vib-ugent.be) or S-Q. Z. (zou@fafu.edu).

Content

Supplementary Notes

Supplementary Note 1. Chromosome number assessment

The material for chromosome karyotype analysis was obtained from the root tips of two-year-old *Euscaphis japonica* seedlings cultivated in the field laboratory of Fujian Agriculture and Forestry University. The root tips were placed in 8-hydroxyquinoline staining solution for 2 h; then, the root tips were fixed using a fixation liquid and stored in a refrigerator at 4 °C. The chromosome number of *E. japonica* was determined from photographs under an Olympus BX63 fluorescence microscope at the Fujian Agriculture and Forestry University in China.

Supplementary Note 2. Whole-genome duplication identification and dating

To determine the time of the whole genome duplication (WGD) event, the distribution of synonymous nucleotide substitution rates between paralogous genes were examined. We compared the anchor-pair *K*s distribution of *E. japonica* and the orthologous *K*s distributions between *E. japonica* and *Populus trichocarpa* and *Vitis vinifera* and found that the peak values of *K*s distribution of *E. japonica – V. vinifera*, *P. trichocarpa – V. vinifera*, and *E. japonica – P. trichocarpa* were 0.76, 0.96, and 0.9, respectively (**Figure S9**). To quantify the differences in substitution rates among these species, we performed a relative rated test using *V. vinifera* as an outgroup to calculate *K*s distance after the divergence between *E. japonica* and *P. trichocarpa*. We calculated that the evolution rate of the *P. trichocarpa* branch was 1.57 times that of the *E. japonica* branch; then, the *K*s value of *E. japonica – P. trichocarpa* was corrected based on the *K*s rate of *E. japonica*, and the peak value after correction was 0.7. According to the differentiation time of *E. japonica* and *P. trichocarpa*, which was found to be 108 Mya (95% CI: 101.57-120.957), the *K*s change rate of *E. japonica* could be calculated as: 6.407e-9 (95% CI: 5.787e-9, 6.892e-9), and the recent peak of WGD in *E. japonica* was found to be 0.34; therefore, we could infer that the time of the recent WGD of *E. japonica* was 53.06 Mya (95% CI: 49.33 Mya, 58.75 Mya).

Supplementary Note 3. Observation of *E. japonica* seed dispersal

The fruit abscission of *E. japonica* after maturity in September revealed mature black seeds. The seeds were tightly attached to the endocarp, and they fell with the fruits in the period until March of the following year. To explore why the fruits of *E. japonica* remain hanging on the plant for a long time, we investigated whether there are animals eating seeds or pericarps during this period. On 10 September 2018, three outdoor timing cameras (Forsafe® H501) were set up in the GYC population (27°54N, 117°10E; Quanzhou City, Fujian Province, China), JS population (27°3'N, 117°14'E; Nanping City, Fujian Province, China), and HX population (24°48'N, 117°11'E; Nanjing City, Fujian Province, China) to monitor the fruits on the tree. After four months of continuous observation, the cameras were taken back on 10 February 2019.

Birds are usually attracted to red fruits, which promote long-distance seed dispersal (EI-Kereamy et al., 2010; Holzwarth et al., 2012; Honda et al., 2014). We found that the great tit (*Parus major*) and masked finches (*Eophona personata*) took the pericarp from the *E. japonica* tree and flew to other branches to feed on the pericarp, whereas they discarded the seeds (**Figures S14a**, **b**). This process promotes the long-distance dispersal of *E. japonica* seeds and helps the population spread.

The scatter-hoarding behaviour of granivorous rodents plays an important role in seed dispersal and seedling regeneration (Gu et al., 2017; Xiao et al., 2020). After leaving the parent plant, seeds have to be dispersed away from maternal sources and ultimately reach microsites suitable for establishment before germinating. At this time, rodents that hoard seeds may place seeds nonrandomly in microsites where the chances of seeding establishment are enhanced (Urgoiti et al., 2018). We found that at night, wild mountain mice and squirrels sometimes climb along the trunk of *E. japonica* trees all the way to the top of the branches to pick the fruit, and then use their mouths to bring the fruit back to the ground (**Figures S14c**, **d**). In addition, the fruits stored by the mountain mouse, and while most of the seeds may be eaten by

rats, the ‘lucky’ ones survive and grow into seedlings (**Figures S14e, f**). To observe how CS7 mice and squirrels eat these fruits, we used a camera (Nikon/Nikon D 810) to record the process. We found that CS7 mice grasped the seeds with their two front paws, picked them from the endocarp, nibbled the hard seed coat outside the seed, and ate the white kernel inside the seed (**Figures S14c**, **f**). The squirrels grabbed the seeds with their front paws and ate them directly (**Figure S14d**). Although both mountain mice and grey squirrels ate the seeds, their food storage behaviour was conducive to the short-distance spread of seeds.

Supplementary Note 4. Determination of pentacyclic triterpene substances

The branches, leaves, and pericarps of *E. japonica* were ground and dried, and 10 kg of dried coarse powder was extracted with 80.0 L of 95% ethanol three times (3 h each time). The extract was concentrated under reduced pressure to obtain 670 g of ethanol extract. Then, 600 g of the ethanol extract were dissolved in methanol and mixed with 600 g of white diatomaceous soil; then, it was successively eluted with 6.0 L of petroleum ether, 6.0 L of ethyl acetate, and 6.0 L of ethanol, and the solvent was recovered under reduced pressure to obtain petroleum ether fractions (106 g), ethyl acetate (150 g), and ethanol (149 g). The ethyl acetate fraction was chromatographed on silica gel (48–75 μm), followed by gradient elution according to V (dichloromethane):V (methanol) = 100:0, 50:1, 30:1, 20:1, 10:1, 5:1, 2:1, 1:1, and 0:100. Finally, 16 fractions were obtained (Fr. 1–Fr. 16). Among them, Fr. 5 was eluted with PRP resin 70% ethanol to obtain compound 19a-hydroxyursolic acid, and Fr. 9 was eluted with PRP resin 70% to obtain euscaphic acid. Finally, it was identified as 19a-hydroxyursolic acid and euscaphic acid by hydrogen nuclear magnetic resonance spectroscopy and carbon nuclear magnetic resonance spectroscopy.

Supplementary Figure

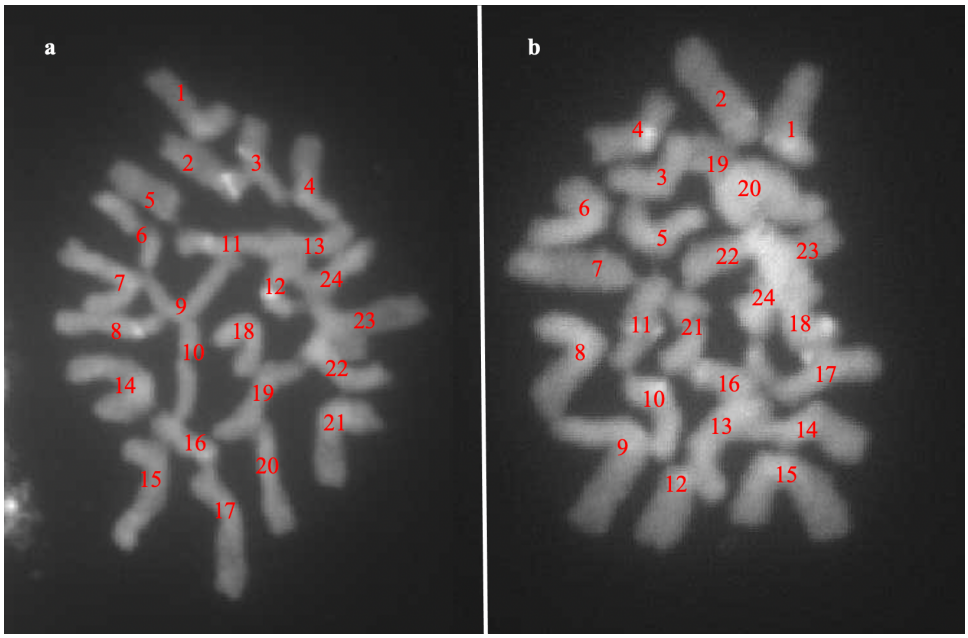


Figure S1. Cytogenetic analysis of *E. japonica.* **a** and **b** are the number of chromosomes of *E. japonica* under the Olympus BX63 fluorescence microscope. Karyotype analysis showed that *E. japonica* contains 24 chromosomes (2n = 2x = 24).

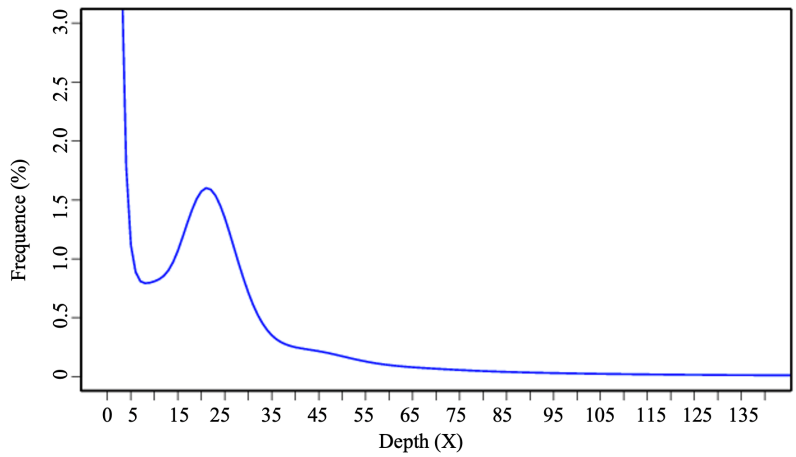


Figure S2. Genome size and heterozygosity of *E. japonica* estimation using 17 *K*-mer distribution. *K*-mer analysis shows that the genome size of *E. japonica* is about 1.39 Gb with heterozygosity of 0.5%.

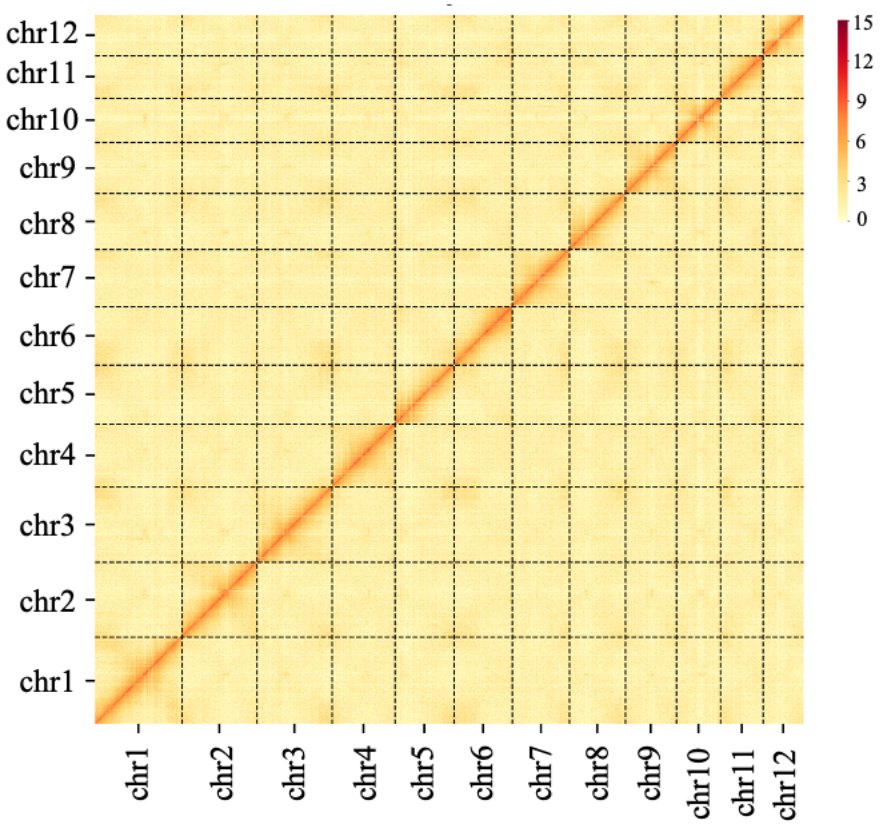


Figure S3. Interchromosomal Hi-C contact map of *E. japonica* genome. The intensity of each pixel represents the number of Hi-C links of 500 kb resolution in the chromosomes. Darker red pixels denote higher contact probabilities. Most interactions were observed within the chromosomes.

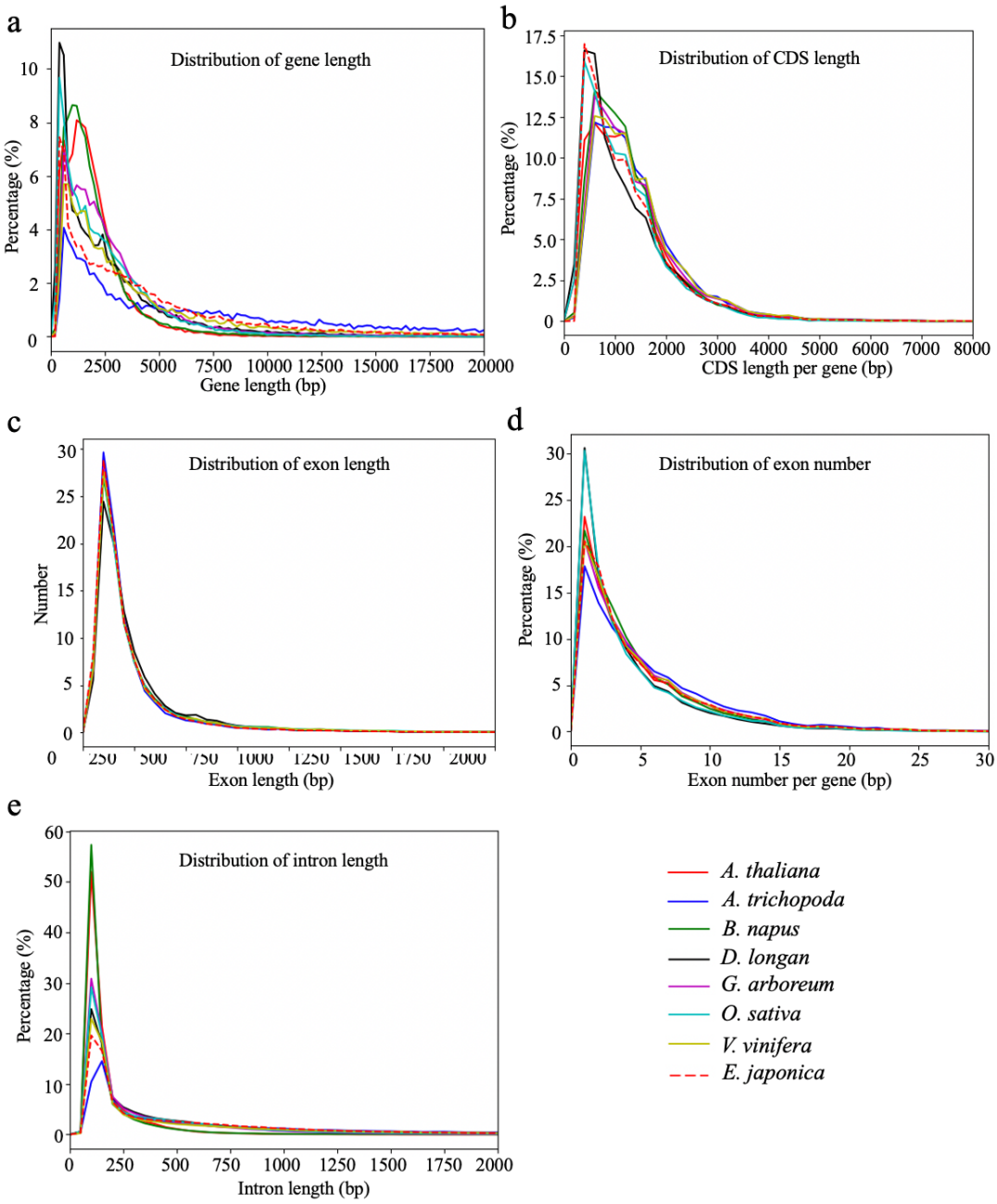


Figure S4. Gene structure prediction results of *E. japonica* and other species. a. Gene length; **b**. CDS length per gene; **c**. exon length; **d**. exon number per gene; **e**. intron length.

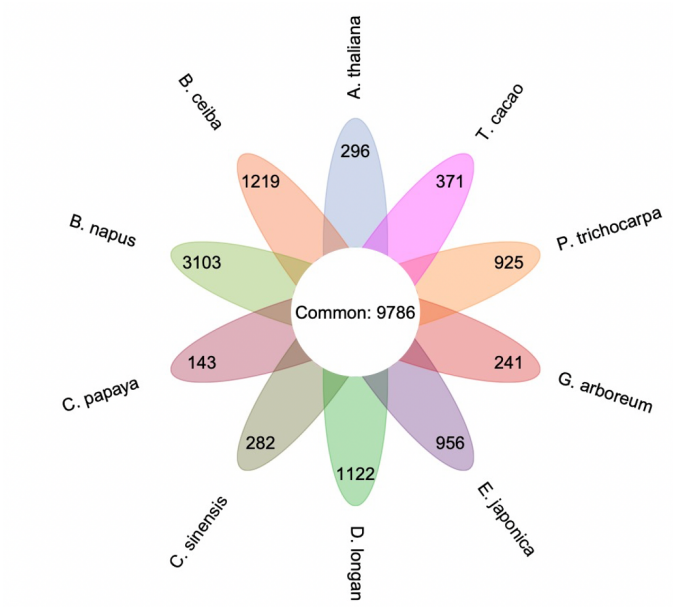


Figure S5. Venn diagram shows gene families of malvids.

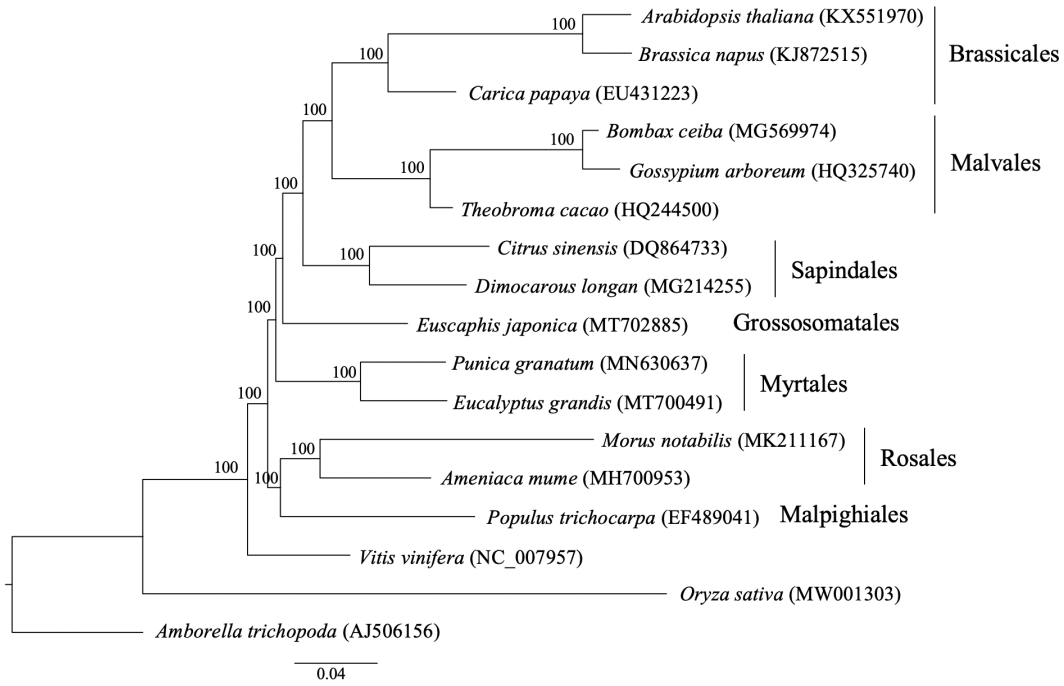


Figure S6. Phylogenetic tree constructed by chloroplast genomes from 17 species. The accession number of each species is in parentheses.

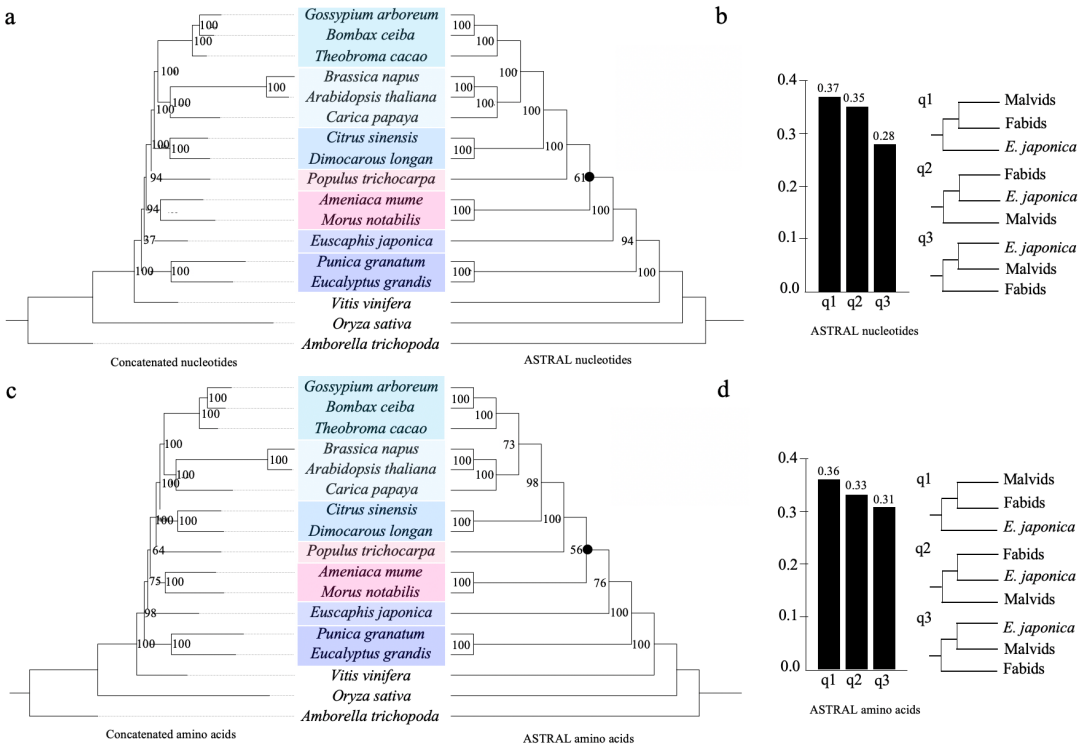


Figure S7. Concatenated- and ASTRAL-based phylogenetic trees.
a. Phylogenetic trees based on the concatenated (left) and multi-species coalescent (MSC) methods (right) using nucleotide sequences. **b.** Estimated proportions of the single-copy gene trees with different topologies based on nucleotide sequences alignments. The x-axis labels q1, q2, and q3 refer to the quartet support for the main topology, the first alternative, and the second alternative, respectively. **c.** Phylogenetic trees based on the concatenated (left) and multi-species coalescent (MSC) methods (right) using amino acid sequences. **d.** Estimated proportions of the single-copy gene trees with different topologies based on amino acid sequences alignments.

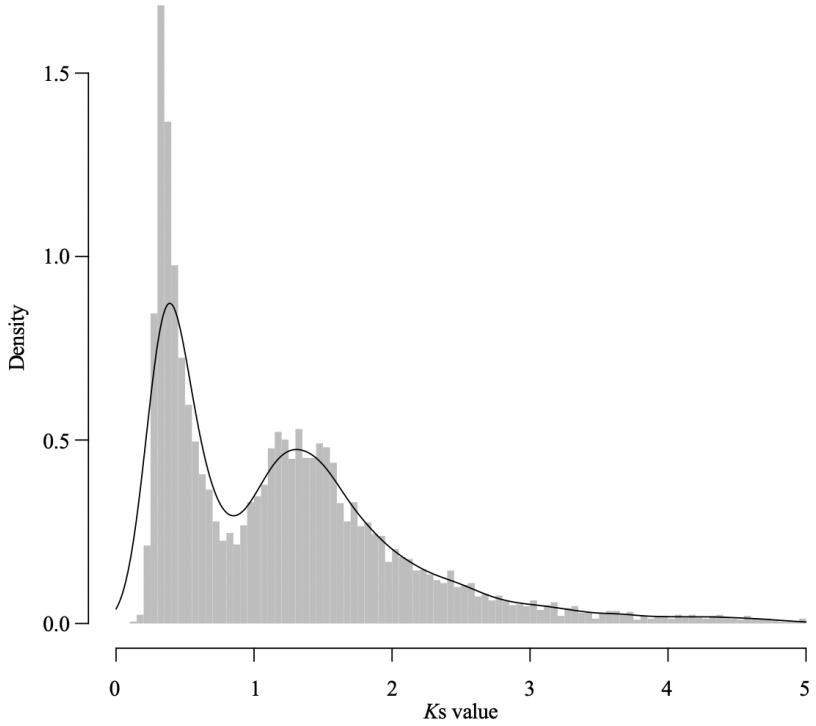


Figure S8. *K*s distribution in *E. japonica*. Distributions of *K*s for paralogs in the genome of *E. japonica* showed two clear peaks, one at *K*s1≈0.34 and the other at *K*s2≈1.29.

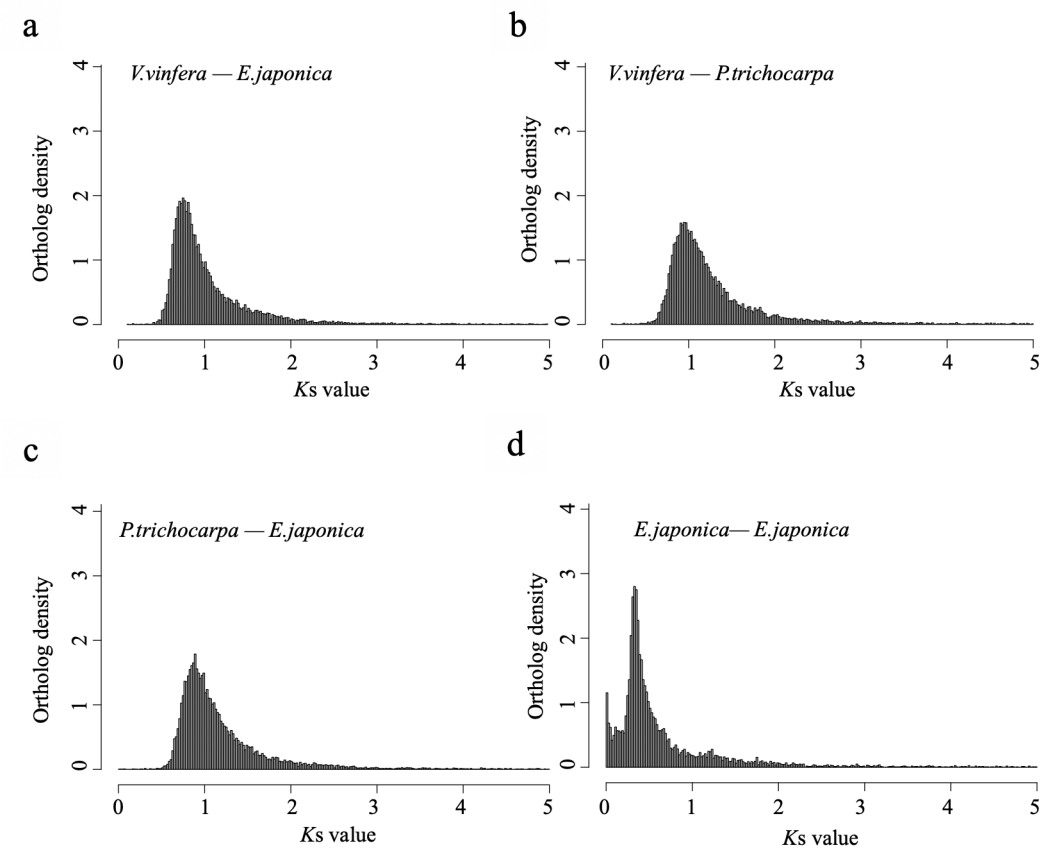


Figure S9. Distributions of synonymous substitutions per synonymous site (*Ks*) of one-to-one orthologs identified between *E. japonica* and *P. trichocarpa* and *V. vinifera*. a. *V. vinifera – E. japonica*. **b**. *V. vinifera – P. trichocarpa*. **c**. *P. trichocarpa – E. japonica*. **d**. *E. japonica – E. japonica*.

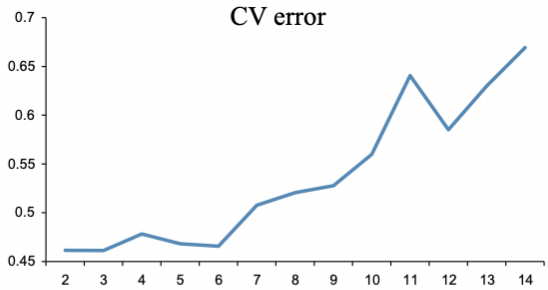


Figure S10. Population structure plot. The ordinate represents cross-validation error rate (CV-value), and the abscissa represents number of clusters (*K*).

	GYC	HX	JS	MHS	MZ	NJ	SF	SLS	WF	WYS	YA
GYC	-	0.2486	0.2108	0.7345	0.245	0.7326	0.7323	0.7234	0.7347	0.6996	0.7327
HX	0.2486	-	0.1468	0.7566	0.1538	0.7565	0.7551	0.7461	0.7579	0.7269	0.7553
JS	0.2108	0.1468	-	0.7413	0.0868	0.7404	0.7407	0.7312	0.7429	0.7104	0.74
MHS	0.7345	0.7566	0.7413	-	0.7503	0.077	0.1298	0.0574	0.1275	0.0583	0.0451
MZ	0.245	0.1538	0.0868	0.7503	-	0.75	0.7494	0.7401	0.7518	0.7207	0.7491
NJ	0.7326	0.7565	0.7404	0.077	0.75	-	0.1045	0.0761	0.1015	0.0482	0.0668
SF	0.7323	0.7551	0.7407	0.1298	0.7494	0.1045	-	0.131	0.0371	0.1045	0.1209
SLS	0.7234	0.7461	0.7312	0.0574	0.7401	0.0761	0.131	-	0.1286	0.0558	0.0542
WF	0.7347	0.7579	0.7429	0.1275	0.7518	0.1015	0.0371	0.1286	-	0.1012	0.1187
WYS	0.6996	0.7269	0.7104	0.0583	0.7207	0.0482	0.1045	0.0558	0.1012	-	0.0521
YA	0.7327	0.7553	0.74	0.0451	0.7491	0.0668	0.1209	0.0542	0.1187	0.0521	-

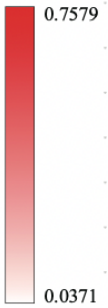


Figure S11. The fixation index (F_{ST}) heat map among *E. japonica* populations.

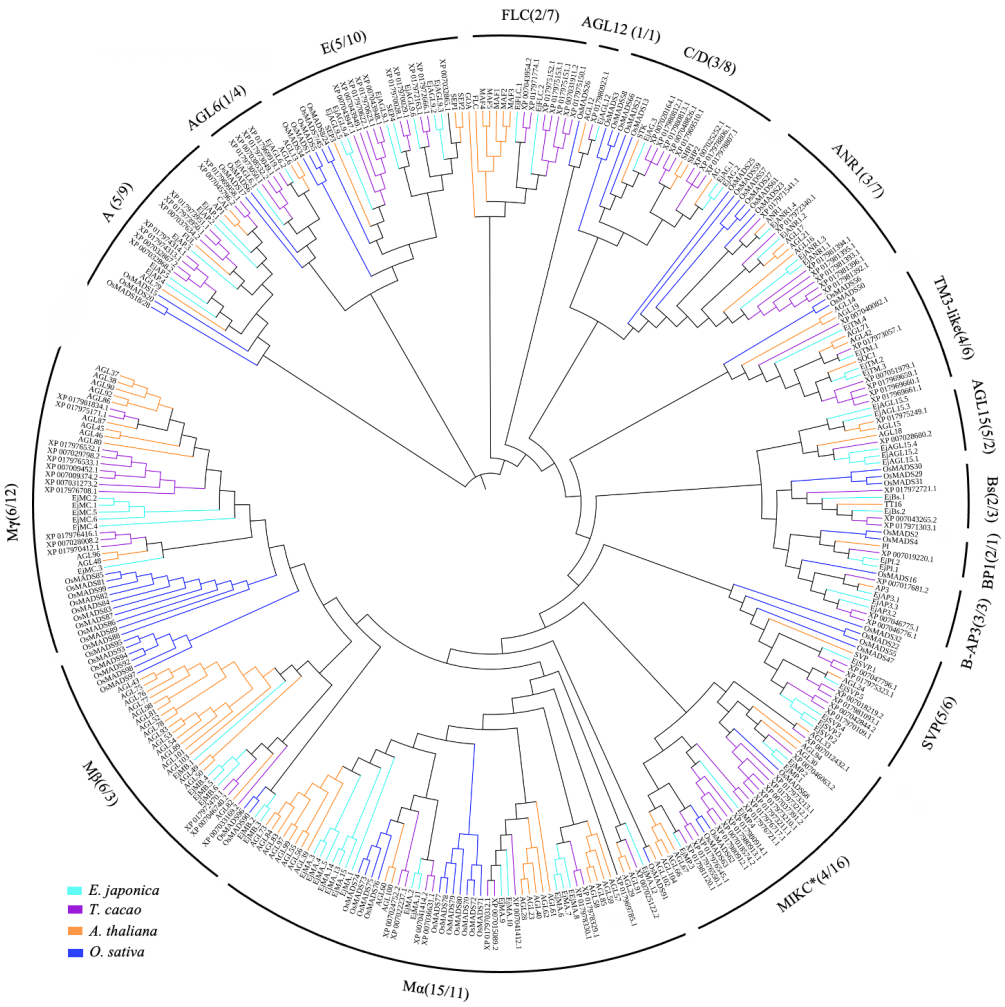


Figure S12. Phylogenetic analysis of MADS-box genes from *O. sativa*, *A. thaliana*, *E. japonica*, and *T. cacao*. The left side of the bracket is the number of *E. japonica* MADS-box subfamily, and the right side of the bracket is the number of *T. cacao* MADS-box subfamily.

Figure S13. Observation the fruit development. a. *E. japonica* blooming. **b**. Young fruit period. **c**. Pericarp discoloration period. **d**. Fruit mature, and the pericarp turns red and begins to dehiscence. **e**. The pericarp begins to fall off. Because the seeds are tightly attached to the endocarp, the seeds fall of as the pericarp falls off. **f**. There were still a few fruits hanging on the branches until March of the following years. We sequenced the transcriptome of the pericarp in four different periods, including the young fruit period (Fr_I), fruit discolouration period (Fr_II), fruit maturity (Fr_III), and fruit abscission period (Fr_IV).

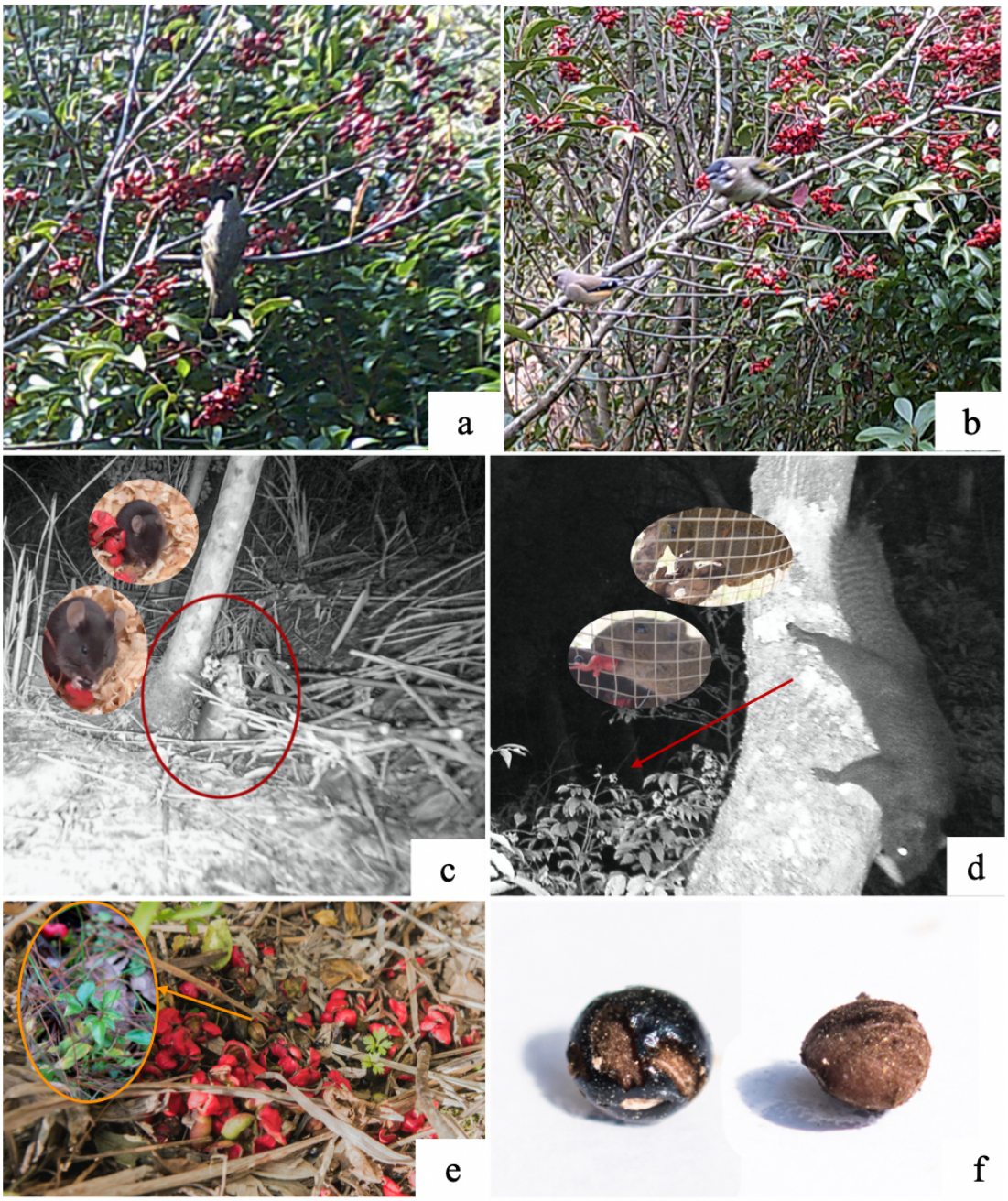


Figure S14. Animal seed dispersal. a and **b**, the masked hawfinch *(Eophona personata*) and big tit (*Parus major*) feed on the fruit of *E. japonica*. **c** and **d**, mountain mice and gray squirrel feed on the fruit of *E. japonica*. **e**. The fruit are stored by mountain mouse, most of the seeds have been eaten, but there are still lucky seeds to germinate after a year. **f**. The seeds of the coat were taken off by the mountain mouse.

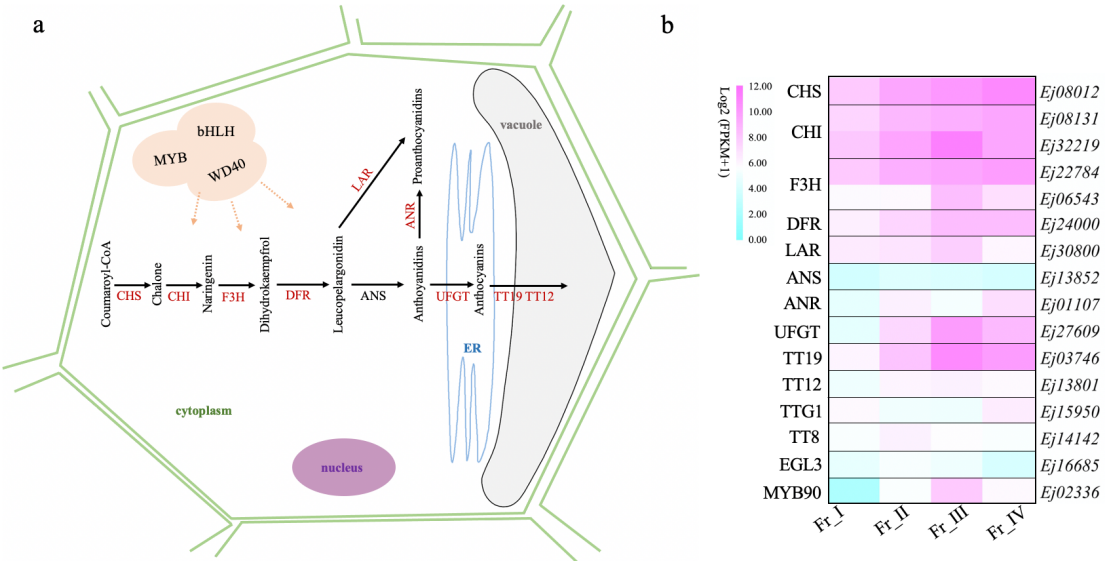


Figure S15. The anthocyanin biosynthesis in *E. japonica* fruits. a. The anthocyanin biosynthesis pathway. **b.** The heatmap of different expression levels of genes in the anthocyanin biosynthesis pathways of fruits at different development stages. Transparent testa glabra 1 (*TTG1*) belonging to WD40 gene family, transparent testa 8 (*TT8*) and enhancer of glabra 3 (*EGL3*) belonging to bHLH, and MYB domain protein 90 (*MYB90*) regulate anthocyanin biosynthesis through the formation of the MYB-bHLH-WD40 (MBW) protein complex (An et al., 2012). (see the full name of gene in **Table S21**)

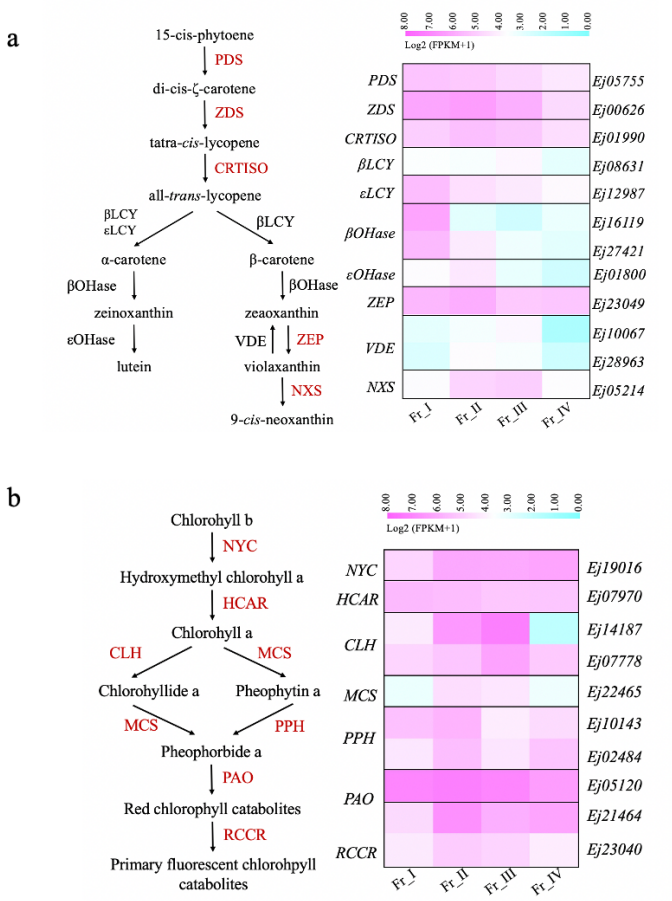


Figure S16. The carotenoid accumulation and the chlorophyll degradation in *E. japonica* fruits. a. The carotenoid biosynthesis pathway (left) and **t**he heatmap of different expression levels of genes in the carotenoid biosynthesis pathways in fruits at different development stages (right). The genes marked in red represent genes that are highly expressed during the green pericarp stage (Fr_I and Fr_II). **b**. The chlorophyll degradation pathway (left) and the heatmap of different expression levels of genes in the chlorophyll degradation pathways of fruits at different development stages (right). The genes marked in red represent genes that are highly expressed during the redness of the pericarp (Fr_II and Fr_III). (see the full name of genes in **Table S21**).

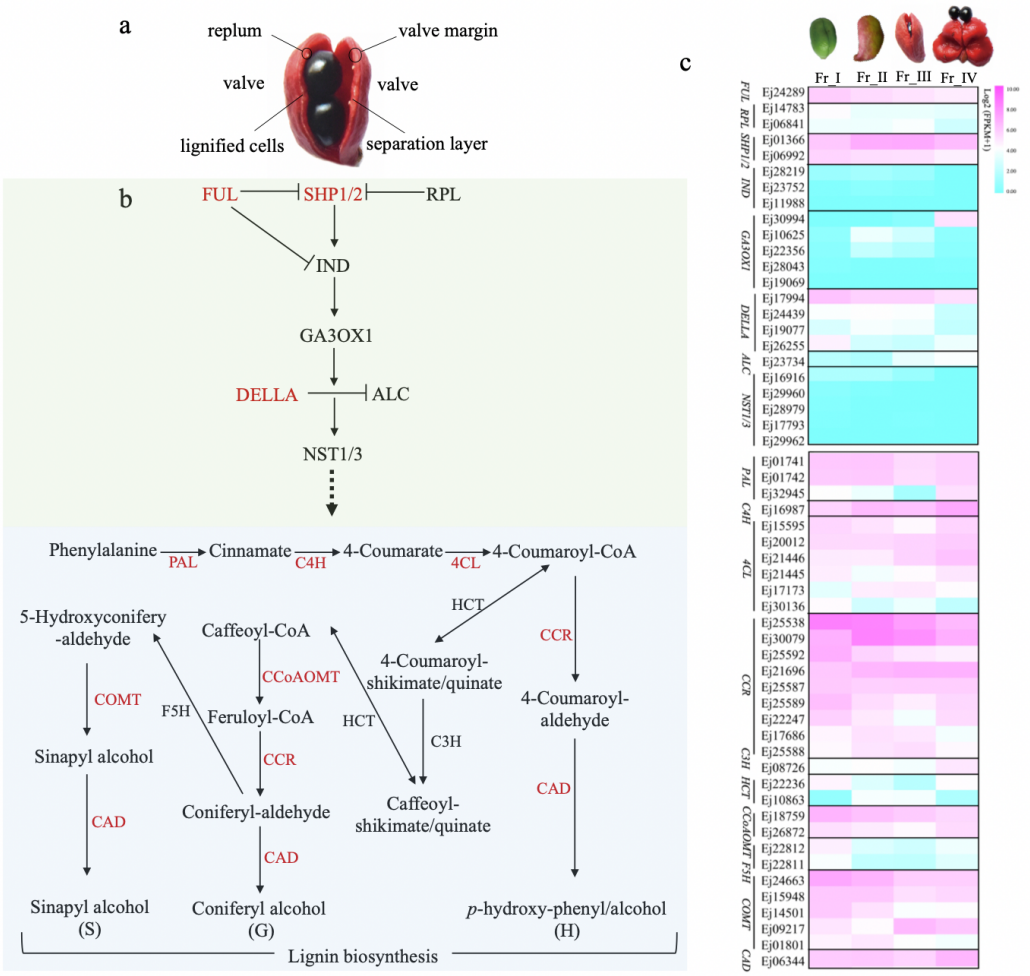


Figure S17. Expression pattern of fruit dehiscence-related genes. a. The structure of *E. japonica* fruit. **b**. Regulation of genes related to fruit dehiscence. The genes marked in red are highly expressed during fruit dehiscence (Fr_II and Fr_III). **c.** Expression profile of fruit dehiscence-related genes. (see the full name of genes in **Tables S22, 23**)

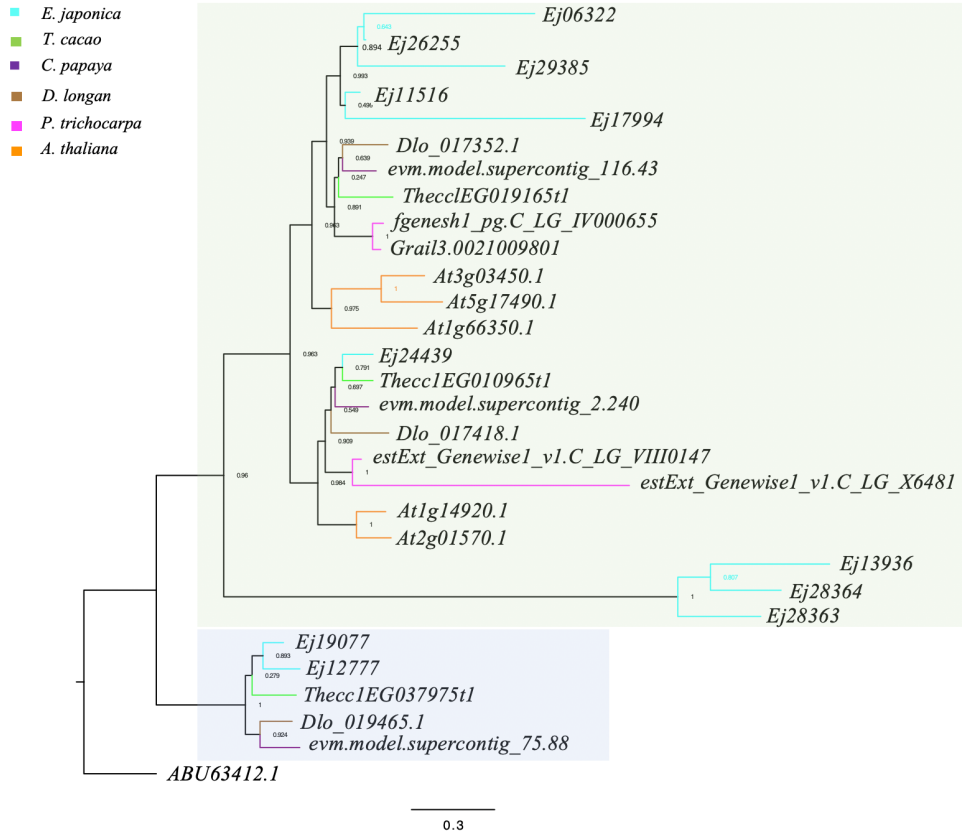


Figure S18. Phylogenetic tree of *DELLA* genes obtained from six malvids species. The *DELLA* gene from *Selaginella kraussiana* (*ABU63412.1*) as outer group. The phylogenetic tree shows that these *DELLA* genes are mainly divided into two group, except for *A. thaliana* and *P. trichocarpa*, the other species are distributed in both types.

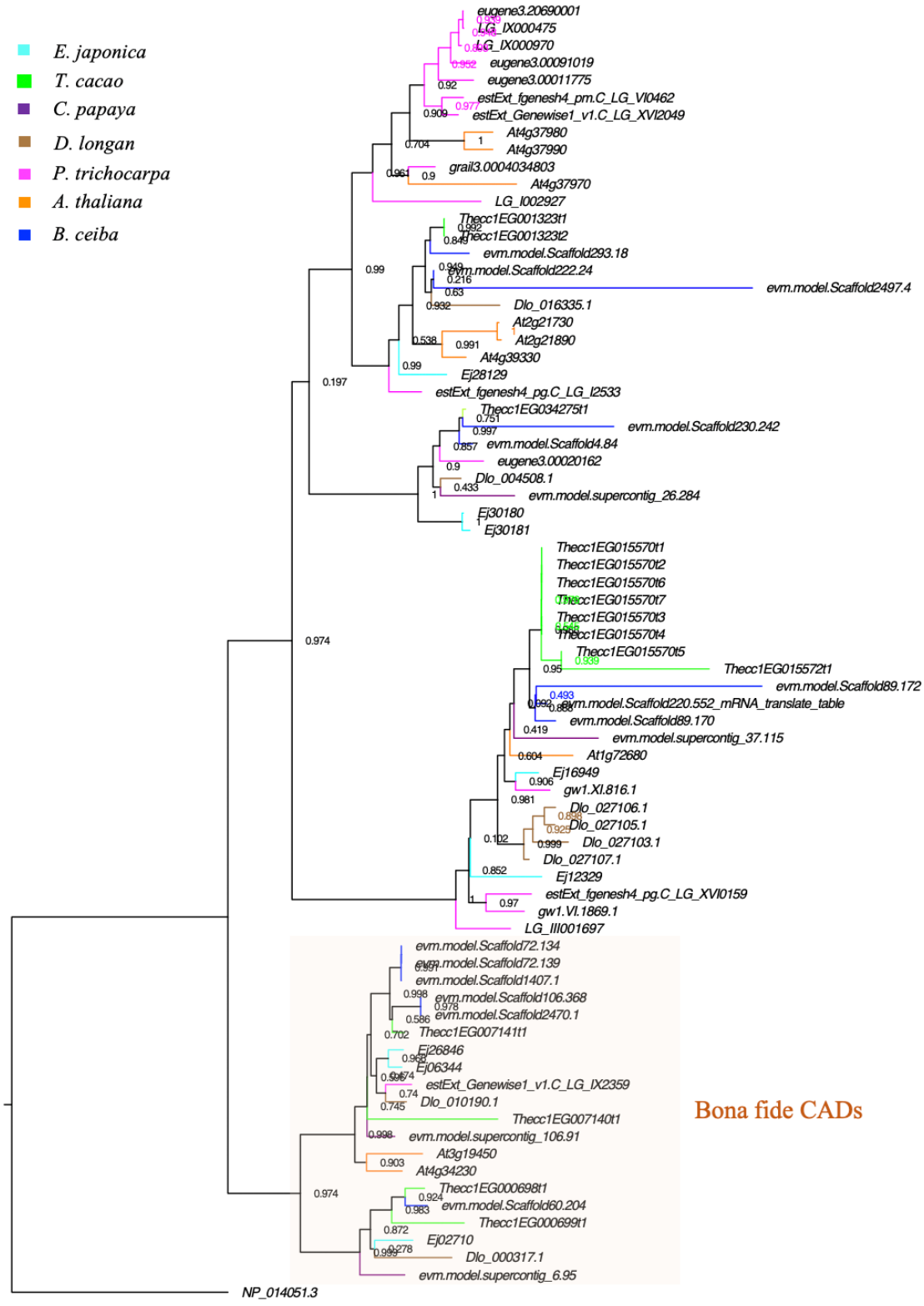


Figure S19. Phylogenetic tree of *CAD* genes obtained from seven malvids species. The *CAD* gene from *Saccharomyces cerevisiae* S288C (*NP_014051*) as the outer group.

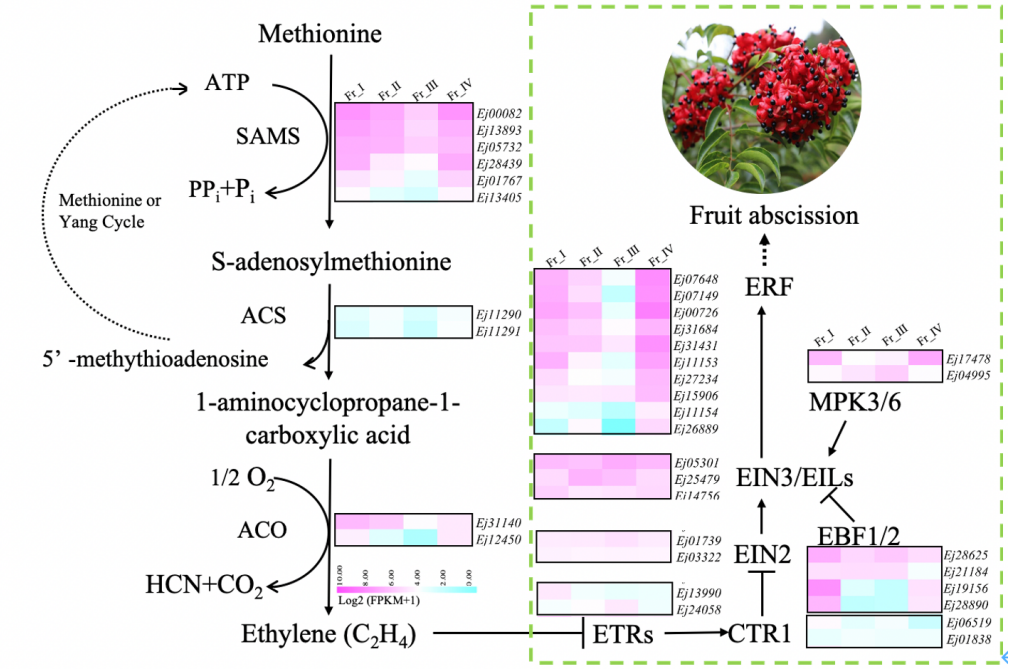
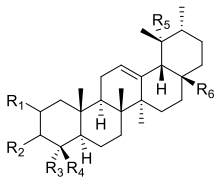
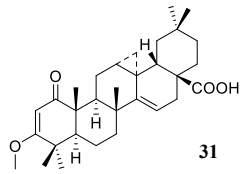


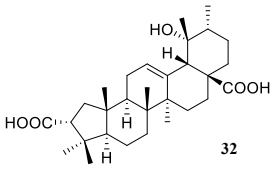
Figure S20. Expression pattern of fruit abscission-related genes. The ethylene insensitive 4 (EIN4) and ethylene response sensor 1 (ERS1) genes belong to ETRs family. (see the full name of genes in **Table S24**)



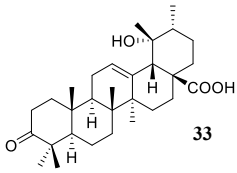
	R_1	R_2	R_3	R_4	R_5	R_6
1	αOH	αOH	CH_3	CH_3	αOH	COOH
2	βOH	βOH	CH_3	CH_3	αOH	COOH
3	αOH	βOH	CH_3	CH_3	αOH	COOH
4	H	βOH	CHO	CH_3	αOH	COOH
5	H	βOH	CH_2OH	CH_3	αOH	COOH
6	H	βOH	CH_3	CH_2OH	αOH	COOH
7	αOH	βOH	CH_3	CH_3	H	COOH
8	H	βOH	CH_3	CH_3	αOH	COOH
9	H	βOH	CH_3	CH_2OH	αOH	$COOCH_3$
10	αOH	αOH	$COOCH_3$	CH_3	αOH	COOH
11	αOH	βOH	$COOCH_3$	CH_3	αOH	COOH
12	H	βOH	CH_3	CH_3	H	COOH



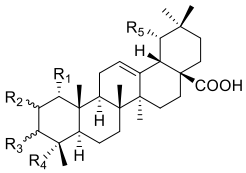
31



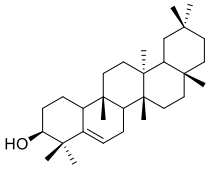
32



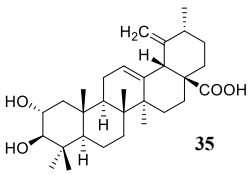
33



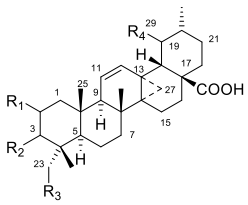
	R_1	R_2	R_3	R_4	R_5
13	H	αOH	βOH	CH_3	H
14	H	αOH	αOH	CH_2OH	H
15	H	H	βOH	CH_2OH	H
16	OH	H	βOH	CH_3	H
17	H	H	βOH	CH_2OH	OH
18	H	H	βOH	CH_3	H
19	H	αOH	βOH	CH_3	OH



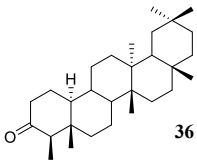
34



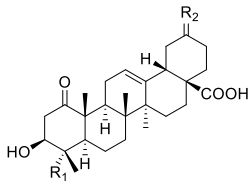
35



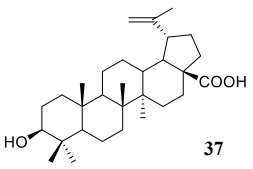
	R_1	R_2	R_3	R_4
20	H	βOH	H	βCH$_3$,αOH
21	H	βOH	OH	βCH$_3$,αOH
22	αOH	αOH	H	βCH$_3$,αOH
23	αOH	βOH	H	βCH$_3$,αOH
24	αOH	βOH	H	βCH$_3$
25	αOH	βOH	H	=CH_2



36



	R_1	R_2
26	CH_2OH	CH_2
27	CH_3	CH_2
28	CH_3	CH_3,CH_2OH
29	CH_3	CH_3, CH_3
30	CH_2OH	CH_3, CH_3



37

Figure S21. Structure of pentacyclic triterpene compounds separated from *Euscaphis.*

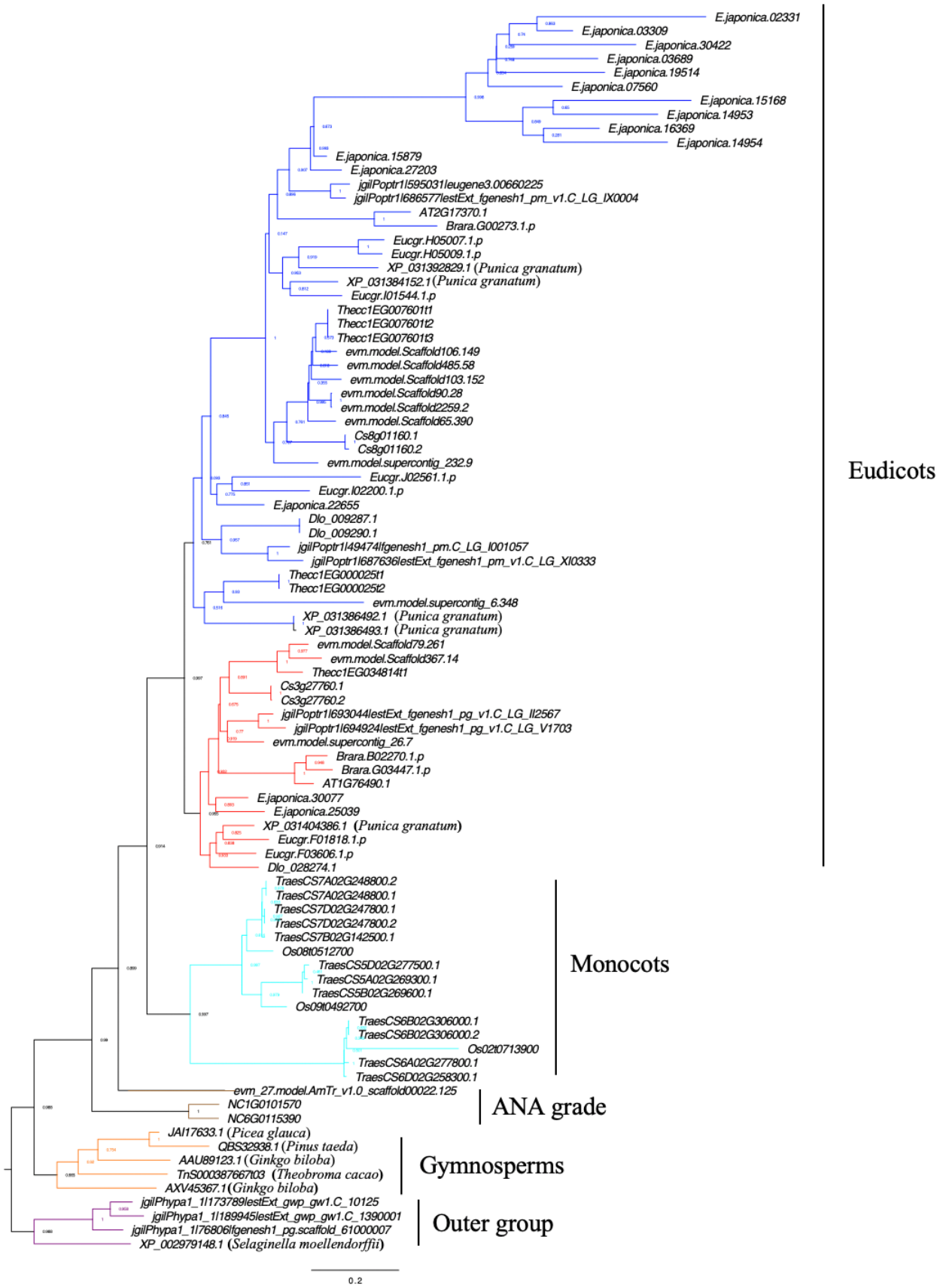


Figure S22. Phylogenetic tree of *HMGR* gene in plants. The *HMGR* gene from *Selaginella moellendorffii* (*XP_002979148.1*) and *Physcomitrella patens*

(*jgilPhypa1_1/173789lestExt_gwp_gw1.C_10125*, *jgilPhypa1_1/189945lestExt_gwp_gw1.C_1290001*, *jgilPhypa1_1/76806fgenesh 1_pg.scaffold_61000007*) as the outer group. The gene ID numbers begin with 'jgiPoptre' to represent the gene IDs of *Populus trichocarpa*, 'Eucgr' to represent the gene IDs of *Eucalyptus grandis*, 'Thecc' to represent the gene IDs of *Theobroma cacao*, 'evm.model.Scaffold' to represent the gene IDs of *Carica papaya*, 'Cc' to represent the gene IDs of *Citrus sinensis*, 'Dlo' to represent the gene IDs of *longyan*, 'Brara' to represent the gene IDs of *Brassica napus*, 'Os' to represent the gene IDs of *Oryza sativa*, 'NC' to represent the gene IDs of *Nymphaea colorata*, 'evm_27.model.AmTr' to represent the gene IDs of *Amborella trichopoda*, and 'AT' to represent the gene IDs of *A. thaliana*. The HMGR is one of the key enzymes contributed to terpene biosynthesis (Darabi et al., 2012). A total of 89 *HMGR* genes from lower land plant, gymnosperms, ANA grade, monocots and eudicots were identified, but no genes were found in all algal genomes. The phylogenetic analysis revealed that the plant *HMGR*s were derived from a common ancestor, and all eudicots *HMGR* genes are divided into two groups. There are 15 members of *HMGR* genes in *E. japonica* genome, which is more than the number of *HMGR* genes in all plants.

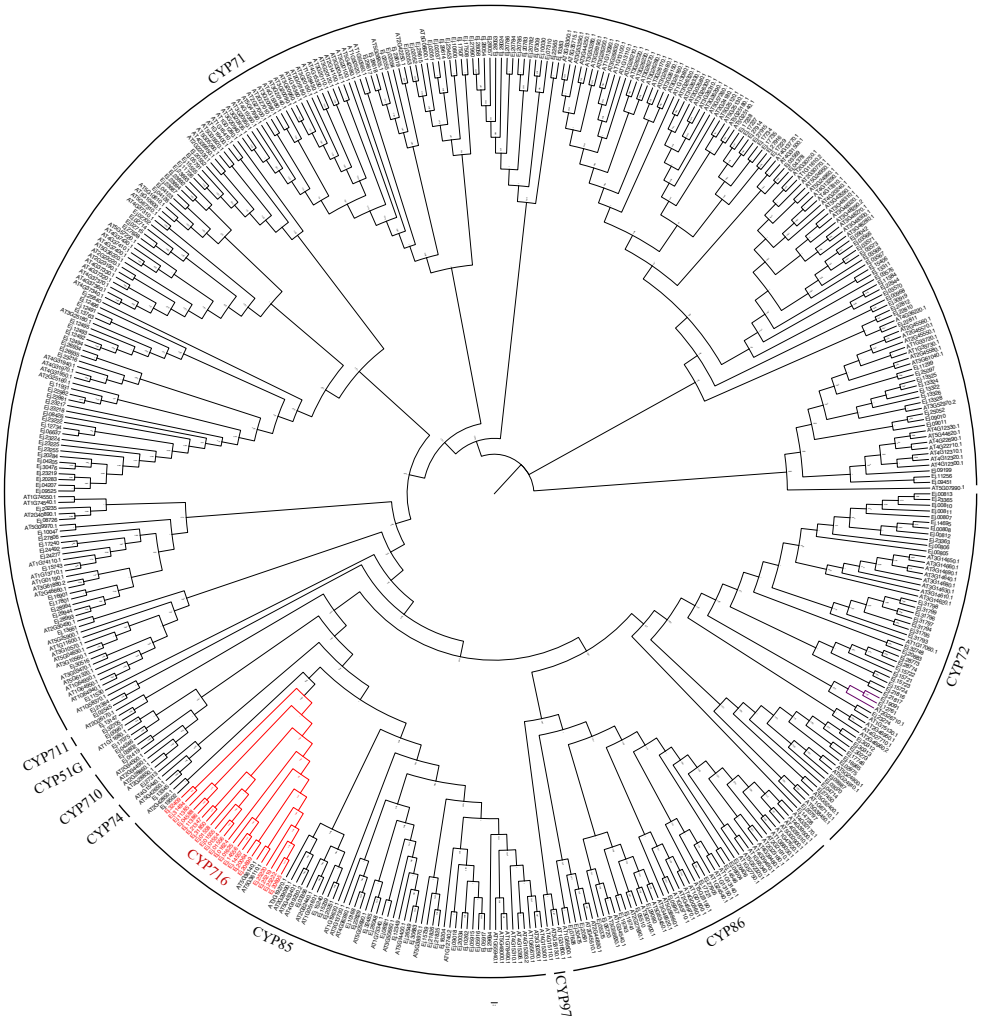


Figure S23. Phylogenetic tree of P450s gene family obtained from *A. thaliana* and *E. japonica.*

Supplementary Table

Table S1. Assembled statistics of *E. japonica* genome.

	Item	*E. japonica*
Illumina sequencing assembly	Clean data (bp)	56,321,483,400
	N50 scaffold (length/number)	(3,879 bp/81,747)
	N50 contig (length/number)	(227 bp/779,614)
	N90 scaffold (length/number)	(387 bp/542,741)
	N90 contig (length/number)	(75 bp/3,471,113)
	Longest scaffold (length/number)	(107,385 bp/1)
	Longest contig (length/number)	(31,653 bp/1)
	Total scaffold (length/number)	(1,389,358,896 bp/1,350,011)
	Total contig (length/number)	(716,596,440 bp/2,822,050)
PacBio sequencing assembly	Clean data (bp)	132,714,369,567
	N50 contig (length/number)	(11,647,650 bp/32)
	N90 contig (length/number)	(2,249,515 bp/121)
	Longest contig (length/number)	(41,602,068 bp/1)
	Total contig (length/number)	(1,199,971,138 bp/382)
Hi-C sequencing assembly	Clean data (bp)	188,758,230,011
	N50 scaffold (length/number)	(98,652,080 bp/5)
	N50 contig (length/number)	(11,099,599 bp/33)
	N90 scaffold (length/number)	(71,790,553 bp/11)
	N90 contig (length/number)	(2,110,939 bp/131)
	Longest scaffold (length/number)	(146,224,083 bp/11)
	Longest contig (length/number)	(40,600,000 bp/1)
	Total scaffold (length/number)	(1,199,999,138 bp/122)
	Total contig (length/number)	(1,199,971,138 bp/402)

Table S2. Evaluation of *E. japonica* genome assembly.

(a) BUSCO evaluation of *E. japonica* genome assembly

Type	Number	Percent (%)
Complete BUSCOs (C)	1,334	97.01
Complete and single-copy BUSCOs (S)	1,246	90.62
Complete and duplicated BUSCOs (D)	88	6.40
Fragmented BUSCOs (F)	15	0.01
Missing BUSCOs (M)	26	0.02
Total BUSCO groups searched	1,375	

(b) CEGMA evaluation of *E. japonica* genome assembly

Group	Proteins	% Completeness
Complete	244	98.39
Group 1	64	96.97
Group 2	54	96.43
Group 3	61	100
Group 4	65	100
Partial	245	98.79
Group 1	65	98.48
Group 2	54	96.43
Group 3	61	100
Group 4	65	100

A core gene database was constructed from 248 core eukaryotic genes of six model species, including *A. thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*.

Table S3. The chromosome length of *E. japonica*.

Chr	Size(bp)	Contig Num
LG01	146,221,383	28
LG02	126,012,888	25
LG03	125,905,229	34
LG04	105,580,116	30
LG05	98,649,780	24
LG06	98,561,950	10
LG07	95,388,557	32
LG08	93,599,311	20
LG09	85,591,010	25
LG10	73,869,617	19
LG11	71,789,253	14
LG12	67,405,898	31
Total	1,188,574,992	292

Table S4. Prediction of gene structures of the *E. japonica* genome.

Gene set		Total number of gene	Average transcript length(bp)	Average CDS length(bp)	Average exons number per gene	Average exon length(bp)	Average intron length(bp)
De novo	Augustus	42,071	4,735.21	1,063.7	5.07	209.76	901.86
Homology	*A. thaliana*	31,994	5,748.74	1,292.93	5.65	228.97	958.93
	A. trichopoda	29,673	6,374.36	1,291.43	5.59	231.08	1,107.73
	B. napus	51,446	5,696.06	1,217.41	5.26	231.63	1,052.33
	D. longan	39,789	5,778.33	1,183.4	4.95	238.88	1,162.09
	G. arboreum	40,450	6,098.89	1,239.63	5.31	233.56	1,128.08
	O. sativa	34,418	5,554.85	1,159.06	5.03	230.42	1,090.69
	V. vinifera	36,025	5,740.21	1,283.87	5.39	238.07	1,014.47
RNA-seq	Transdecoder	19,920	5,418.34	1,351.6	5.85	230.93	838.01
Final set	EVM	32,950	5,089.02	1,154.47	5.19	222.56	939.63

Table S5. Statistics on the function annotation of the *E. japonica* genome.

Database		Number	Percent (%)
Annotation	KOG	17,181	52.14
	KEGG	11,058	33.56
	NR	30,756	93.34
	SwissProt	24,740	75.08
	GO	14,857	45.09
All_annotation		30,873	93.70

Table S6. Non-coding RNA annotation results of *E. japonica* genome.

Type		Copy Number	Average Length(bp)	Total Length(bp)	Percentage (%) of Genome
rRNA	rRNA	759	1,541.11	1169704	0.081443
	18S	184	1,856.75	341642	0.023787
	28S	202	3,848.18	777333	0.054123
	5.8S	200	153.99	30798	0.002144
	5S	173	115.21	19931	0.001388
snRNA	snRNA	3,940	105.30	414,877	0.028887
	CD-box	3,783	104.23	394,289	0.027453
	HACA-box	63	125.32	7,895	0.000550
	splicing	94	135.03	12,693	0.000884
miRNA		349	117.34	40,953	0.002851
tRNA		626	74.34	46,538	0.003240

Table S7. BUSCO assessment of the *E. japonica* annotated genome.

Type	Number	Percent (%)
Complete BUSCOs	1,348	98.03
Complete and single-copy BUSCOs (S)	1,260	91.64
Complete and duplicated BUSCOs (D)	88	6.40
Fragmented BUSCOs (F)	13	0.95
Missing BUSCOs (M)	14	1.10
Total BUSCO groups searched	1,375	-

Table S8. Statistic of different repeat sequence types in *E. japonica* genome.

Type	Number	Length(bp)	% of genome
Retroelements	779,609	542,007,515	45.17
SINEs:	767	29,448	0.01
Penelope	99	8,331	0.01
LINEs:	46,025	22,667,855	1.89
CRE/SLACS	2	56	0
L2/CR1/Rex	17,984	9,622,745	0.80
RTE/Bov-B	229	28,239	0
L1/CIN4	27,600	13,002,128	1.09
LTR elements:	732,817	519,310,212	43.28
Copia	205,289	188,292,983	15.69
Gypsy	332,434	234,041,183	19.50
Retroviral	537	81,567	0.01
DNA transposons	349,986	122,164,143	10.18
hobo-Activator	110,662	37,888,791	3.16
Tc1-IS630-Pogo	2,832	147,891	0.01
Tourist/Harbinger	25,127	6,699,460	0.56
Unclassified:	827,464	242,532,149	20.21
Satellites:	3,135	651,011	0.05
Simple repeats:	3,249	454,524	0.04
Low complexity:	186	18,213	0
Total		872,815,690	72.74

LINE, long interspersed nuclear element; SINE, short interspersed element; LTR, long terminal repeat.

Table S9. Gene-clustering statistics for 17 species.

Species	Genes	Unclustered genes	Clustered genes	Familys	Unique families	Unique families genes	Common families	Common families genes	Single copy	Average genes per family
A. mume	23,340	491	22,849	13,296	214	1,036	6,830	12,087	28	1.718
A. thaliana	27,416	1,653	25,763	14,674	252	968	6,830	12,794	28	1.756
A. trichopoda	26,846	4,646	22,200	13,122	899	4,698	6,830	9,765	28	1.692
B. ceiba	52,705	7,755	44,950	15,810	902	3,374	6,830	21,964	28	2.843
B. napus	96,801	4,006	92,795	17,859	2,948	16,168	6,830	39,979	28	5.196
C. papaya	17,506	493	17,013	12,288	56	198	6,830	10,168	28	1.385
C. sinensis	24,287	443	23,844	13,324	198	939	6,830	11,835	28	1.79
D. longan	39,282	2,887	36,395	14,407	940	8,388	6,830	13,279	28	2.526
E. grandis	35,932	1,482	34,450	13,895	817	5,255	6,830	15,739	28	2.479
E. japonica	32,950	2,179	30,771	14,495	737	2,775	6,830	15,577	28	2.123
G. arboreum	34,411	713	33,698	13,656	192	1,051	6,830	17,936	28	2.468
M. notabilis	21,348	418	20,930	12,945	132	582	6,830	11,400	28	1.617
O. sativa	27,694	5,469	22,225	11,804	1,123	4,366	6,830	11,909	28	1.883
P. granatum	29,127	3,325	25,802	13,001	759	5,122	6,830	12,002	28	1.985
P. trichocarpa	41,335	4,742	36,593	14,564	756	3,199	6,830	18,346	28	2.513
T. cacao	29,452	3,521	25,931	14,377	293	1,400	6,830	11,836	28	1.804
V. vinifera	26,346	3,755	22,591	13,289	519	1,710	6,830	11,838	28	1.7

Unclustered genes refer to the number of genes endemic to the species. Unique families refer to the gene family that are Unique to a species.

Table S10. KEGG enrichment result of unique genes families of *E. japonica*. (see separate files)

Table S11. Gene Ontology (GO) and KEGG enrichment result of significant shared by malvids gene families. (see separate files)

Table S12. Gene Ontology (GO) and KEGG enrichment result of significant expansion of *E. japonica* gene families. (see separate files)

Table S13. Gene Ontology (GO) enrichment result of significant contraction of *E. japonica* gene families. (see separate files)

Table S14. Statistical sampling population information.

Population	Samples of Num.	Location	Longitude	Latitude	Altitude (m)
HX	10	Fujian, Nanjing	117°11′	24°48′	450
JS	9	Fujian, Shaowu	117°14’	27°3’	408
MZ	10	Guandong, Meizhou	116°15’	24°36’	320
GYC	8	Fujian, Quanzhou	117°10′	27°54′	447
MHS	9	Fujian, Longyan	116°52′	25°18′	1270
NJ	8	Jiangsu, Nanjing	118°34′	32°6′	60
SF	10	Yunnan, Guangyuan	104°14′	28°36′	680
SLS	10	Fujian, Quanzhou	118°29′	25°38′	1090
WF	9	Sichuan, Guangyuan	106°4′	31°57′	780
WYS	9	Fujian, Nanping	117°44′	27°45′	1250
YA	9	Fujian, Sanming	117°27’	26°00’	522

Table S15. Statistics population resequencing information.

Population	Bases (Gb)	Indels	Non-Synonymous	Synonymous	Intron	Intergenic	ncRNA	SNPs	Coverage (%)
HX	21.71	265,731	30,232	24,795	439,516	4,809,117	8,228	5,303,660	97.86
JS	19.22	270,899	35,429	29,268	510,445	5,576,663	8,537	6,151,805	90.46
MZ	19.99	272,364	35,721	29,533	512,200	5,520,043	8,401	6,097,497	97.53
GYC	20.46	254,301	23,539	18,869	370,702	4,281,299	7,941	4,694,409	98.38
MHS	20.72	422,695	104,597	81,636	1,468,707	15,837,634	14,188	17,492,574	74.56
NJ	18.57	418,207	103,385	80,665	1,449,277	15,552,539	13,536	17,185,866	74.30
SF	19.73	419,469	103,226	80,639	1,439,774	15,458,776	13,964	17,082,415	75.11
SLS	18.54	424,501	107,238	84,116	1,528,355	16,411,020	14,407	18,130,729	73.84
WF	19.77	416,079	99,226	76,475	1,407,830	15,213,464	13,432	16,796,995	74.86
WYS	15.77	424,585	122,174	97,796	1,676,945	17,488,148	14,910	19,385,063	72.87
YA	20.46	422,293	103,652	80,554	1,469,043	15,830,335	14,149	17,483,584	71.25
Total	214.94	4,011,124	868,419	684,346	12,272,794	131,979,038	131,693	145,804,597	/
Average	19.54	364,647	78,947	62,213	1,115,709	11,998,094	11,927	13,254,963	81.90

Table S16. Statistical nucleotide polymorphisms in the populations.

Population	Tajima'D	Pi ($\theta\pi$)	Theta (θw)
HX	0.985525	0.001411	0.0011
JS	0.891749	0.001622	0.001285
MZ	0.944524	0.001645	0.001299
GYC	1.021963	0.001385	0.001042
MHS	1.498027	0.00539	0.003964
NJ	1.386049	0.005377	0.004052
SF	1.389768	0.005098	0.003787
SLS	1.558582	0.005521	0.003997
WF	1.369312	0.005092	0.00382
WYS	1.406919	0.005945	0.004443
YA	1.507934	0.005427	0.003987

$\theta\pi$ average number of pairwise nucleotide differences per site, θw Watterson's estimator of θ per base pair.

Table S17. Candidate positive selection genes (PSGs) in the evergreen population. (see separate files)

Table S18. Candidate positive selection genes (PSGs) in the deciduous population. (see separate files)

Table S19. Gene Ontology (GO) enrichment result of significant PSGs in the evergreen population. (see separate files)

Table S20. List of MADS-box genes identified in *E. japonica*.

Gene ID	Name	Type	Subfamily
Ej15843	*EjMP.1*	MIKC*	
Ej17442	*EjMP.2*	MIKC*	
Ej23690	*EjMP.3*	MIKC*	
Ej28569	*EjMP.4*	MIKC*	
Ej18774	*EjAGL15.1*	MIKCc	AGL15
Ej08576	*EjAGL15.2*	MIKCc	AGL15
Ej18773	*EjAGL15.3*	MIKCc	AGL15
Ej02487	*EjAGL15.4*	MIKCc	AGL15
Ej25576	*EjAGL15.5*	MIKCc	AGL15
Ej04020	*EjANR1.1*	MIKCc	ANR1
Ej24158	*EjANR1.2*	MIKCc	ANR1
Ej09764	*EjANR1.3*	MIKCc	ANR1
Ej06477	*EjANR1.4*	MIKCc	ANR1
Ej06496	*EjSVP.1*	MIKCc	SVP
Ej06497	*EjSVP.2*	MIKCc	SVP
Ej26712	*EjSVP.3*	MIKCc	SVP
Ej16548	*EjSVP.4*	MIKCc	SVP
Ej25896	*EjSVP.5*	MIKCc	SVP
Ej14565	*EjAGL12*	MIKCc	AGL12
Ej30016	*EjBs.1*	MIKCc	Bs
Ej16881	*EjBs.2*	MIKCc	Bs
Ej22817	*EjAP3.1*	MIKCc	B-AP3
Ej24034	*EjAP3.2*	MIKCc	B-AP3
Ej10175	*EjAP3.3*	MIKCc	B-AP3
Ej24677	*EjPI.1*	MIKCc	B_PI
Ej14296	*EjPI.2*	MIKCc	B_PI
Ej30461	*EjAGL6.1*	MIKCc	AGL6
Ej29674	*EjAGL6.2*	MIKCc	AGL6
Ej18166	*EjAGL9.1*	MIKCc	E
Ej07442	*EjAGL9.2*	MIKCc	E
Ej24288	*EjAGL9.3*	MIKCc	E
Ej15732	*EjAGL9.4*	MIKCc	E
Ej12040	*EjAGL9.5*	MIKCc	E
Ej17129	*EjAGL9.6*	MIKCc	E
Ej26631	*EjFLC.1*		FLC
Ej17130	*EjFLC.2*		FLC
Ej18165	*EjAP.1*	MIKCc	A
Ej07441	*EjAP.2*	MIKCc	A
Ej24289	*EjAP.3*	MIKCc	A
Ej15731	*EjAP.4*	MIKCc	A
Ej21818	*EjAP.5*	MIKCc	A
Ej01366	*EjAG.1*	MIKCc	C/D
Ej06992	*EjAG.2*	MIKCc	C/D
Ej14561	*EjAG.3*	MIKCc	C/D
Ej11312	*EjTM.1*	MIKCc	SOC1

Ej25080	*EjTM.2*	MIKCc	SOC1
Ej02106	*EjTM.3*	MIKCc	SOC1
Ej30456	*EjTM.4*	MIKCc	SOC1
Ej27700	*EjMB.1*	Type I	Mβ
Ej05770	*EjMB.2*	Type I	Mβ
Ej05768	*EjMB.3*	Type I	Mβ
Ej15059	*EjMB.4*	Type I	Mβ
Ej15060	*EjMB.5*	Type I	Mβ
Ej28123	*EjMB.6*	Type I	Mβ
Ej29500	*EjMA.1*	Type I	Mα
Ej13838	*EjMA.2*	Type I	Mα
Ej30325	*EjMA.3*	Type I	Mα
Ej12459	*EjMA.4*	Type I	Mα
Ej00891	*EjMA.5*	Type I	Mα
Ej03792	*EjMA.6*	Type I	Mα
Ej03793	*EjMA.7*	Type I	Mα
Ej01409	*EjMA.8*	Type I	Mα
Ej26774	*EjMA.9*	Type I	Mα
Ej26775	*EjMA.10*	Type I	Mα
Ej06424	*EjMA.11*	Type I	Mα
Ej01410	*EjMA.12*	Type I	Mα
Ej22133	*EjMA.13*	Type I	Mα
Ej22095	*EjMA.14*	Type I	Mα
Ej20916	*EjMA.15*	Type I	Mα
Ej14337	*EjMC.1*	Type I	Mγ
Ej14340	*EjMC.2*	Type I	Mγ
Ej29491	*EjMC.3*	Type I	Mγ
Ej00012	*EjMC.4*	Type I	Mγ
Ej07559	*EjMC.5*	Type I	Mγ
Ej19161	*EjMC.6*	Type I	Mγ

Table S21. The genes involved in anthocyanin biosynthesis, carotenoid biosynthesis, and chlorophyll degradation.

Enzyme	Abbreviation	*Arabidopsis*	*E. japonica* gene ID			
genes related to anthocyanin synthesis						
Chalcone synthase	CHS	*AT5G13930*	*Ej04030* *Ej24661* *Ej24665* *Ej24840*	*Ej27478* *Ej27477* *Ej27475* *Ej08012*	*Ej02859* *Ej26474* *Ej10469* *Ej24398*	Ej27585 Ej16194 Ej24445
Chalcone-flavanone isomerase	CHI	*AT3g55120*	*Ej00802*	*Ej08131*	*Ej32219*	
Flavanone-3-hydroxylase	F3H	*A T3g51240*	*Ej11142* *Ej22783*	*Ej27227*	*Ej22784*	Ej06543
Dihydroflavonols 4-reductase	DFR	*AT5g42800*	*Ej24000*	*Ej23330*		
Leucoanthocyanidins reductase	LAR		*Ej30800*			
Anthocyanidin synthase	ANS	*AT4G22880*	*Ej13852*	*Ej02565*		
Anthocyanidin reductase	ANR	*AT1G61720*	*Ej01107*	*Ej26659*		
UDP (Uridine diphosphate) flavonoid Glucosyltransferase	UFGT	*AT5G17050*	*Ej23961*	*Ej23912*	*Ej27609*	Ej23268
Transparent testa 19	TT19	*AT5G17220*	*Ej03746*			
Transparent testa12	TT12	*AT3G59030*	*Ej13801*	*Ej13802*		
Transparent testa glabra 1	TTG1	*AT5G24520*	*Ej27273*	*Ej15950*		
Transparent testa 8	TT8	*AT4G09820*	*Ej14142*			
Enhancer of glabra 3	EGL3	*AT1G63650*	*Ej16685*			
MYB domain protein 90	MYB90	*AT1G66390*	*Ej02336*	*Ej29284*		
genes related to carotenoid synthesis						
Phytoene desaturase	PDS	*At4g14210*	*Ej05755*			
ζ-carotene desaturase	ZDS	*At3g04870*	*Ej00262*			

Carotenoid isomerase	CRTISO	*At1g06820*	*Ej27739*	*Ej01990*	*Ej20227*
β-carotene cyclase/ (Lycopersicum)	βLCY1/2	*At3g10230*	*Ej08631*		
ε-cyclase	εLCY	*At5g57030*	*Ej12987*		
β-carotene hydroxylase	βOHase1/2	*At4g25700/At5g52570*	*Ej27421*	*Ej16119*	*Ej16720*
ε-carotene hydroxylase	εOHase	*At3g53130*	*Ej01800*		
Zeaxanthin epoxidase	ZEP	*At5g67030*	*Ej23049*	*Ej28738*	
Violaxanthin deepoxidase	VDE	*At1g08550*	*Ej28963*	*Ej10067*	
Neoxanthin synthase	NXS		*Ej05214*		
genes related to chlorophyII degradation					
non-yellow coloring 1/non-yellow coloring 1 like	NYC	*AT4G13250/AT5G04900*	*Ej19016*	*Ej12822*	
7-Hydroxymethyl chlorophyll a reductase	HCAR	*AT1G04620*	*Ej07970*		
chlorophyllase 1/2	CLH	*AT1G19670/AT5G43860*	*Ej14187*	*Ej07778*	
Mg-chelatase	MCS		*Ej22465*		
pheophytinase	PPH	*AT5G13800*	*Ej02484*	*Ej10143*	*Ej04577*
pheophrbide a oxygenase	PAO	*AT3G44880*	*Ej21464*	*Ej05120*	
red chlorophyll cataabolite reductase	RCCR	*AT4G37000*	*Ej23040*		

Table S22. Identification fruit dehiscence-related genes in *E. japonica*.

Enzyme	Abbreviation	Arabidopsis	Gene ID
FRUITFULL	FUL	*AT5G60910*	*Ej24289*
SHATTERPROOF 1/2	SHP1/SHP2	*AT3G58780/AT2G42830*	*Ej01366*; *Ej06992*
REPLUMLESS	RPL	*AT5G02030*	*Ej14783*; *Ej06841*; *Ej05075*
INDEHISCENT	IND	*AT4G00120*	*Ej28219*; *Ej23752*; *Ej11988*
GIBBERELLIN 3-OXIDASE 1	GA3OX1	*AT1G15550*	*Ej10625*; *Ej22356*; *Ej30994*; *Ej28043*; *Ej19069*
ALCATRAZ	ALC	*AT5G67110*	*Ej23734*
NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1/3	NST1/3	*AT2G46770/AT1G32770*	*Ej16916*; *Ej29960*; *Ej28979*; *Ej17793*; *Ej29962*
DELLA			*Ej17994*; *Ej26255*; *Ej24439*; *Ej19077*; *Ej12777*; *Ej28364*; *Ej29385*; *Ej28363*; *Ej13936*

Table S23. Genes related to lignin synthesis that are highly expressed during pericarp dehiscence.

Gene name	Enzyme	*E. japonica ID*
PAL	Phenylalanone ammonia-lyase	*Ej01741*; *Ej01742*; *Ej32945*
C4H	Cinnamate 4-hydroxylase	*Ej16987*
4CL	Coumarate-4-CoA ligase	*Ej15595*; *Ej20012*; *Ej21446*; *Ej21445*; *Ej17173*; *Ej30136*
HCT	Hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoytransferase	*Ej22236*; *Ej10863*
C3H	4-Coumarate 3-hydroxylase	*Ej08726*
CCoAOMT	Caffeoyl-CoA O-methytransferase	*Ej18759*; *Ej26872*
CCR	Cinnamoyl CoA reductase	*Ej25538*; *Ej30079*; *Ej25592*; *Ej21696*; *Ej25587*; *Ej25589*; *Ej22247*; *Ej17686*; *Ej25588*
CAld5H/F5H	Ferulate 5-hydroxylase	*Ej22812*; *Ej22811*
COMT	Caffeare O-methytransferase	*Ej24663*; *Ej15948*; *Ej14501*; *Ej09217*; *Ej01801*
CAD	Cinnamyl alcohol dehydrogenase	*Ej06344*

Table S24. Gene expression levels (FPKMs) of fruit abscission-related genes in pericarp.

Enzyme	Abbreviation	Gene ID	Fr_I	Fr_II	Fr_III	Fr_IV
S-adenosy-L-methionine synthetase	SAMS	Ej00082	399.47	219.63	86.83	389.09
		Ej13893	279.39	185.14	67.55	182.85
		Ej05732	203.73	170.09	84.29	137.21
		Ej28439	186.9	42.04	27.66	209.45
		Ej01767	48.66	33.56	13.68	77.25
		Ej13405	26.69	12.08	8.4	31.12
1-aminocyclopropane-1-carboxylic acid synthase	ACS	Ej11290	11.99	19.67	9.23	19.37
		Ej11291	9.68	18.49	6	22.69
1-aminocyclopropane-1-carboxylic acid oxidase	ACO	Ej31140	148.73	93.29	26.37	42.76
		Ej12450	36.15	11.26	2.37	42.93
constitutive triple response 1	CTR1	Ej06519	27.88	12.84	18.84	5.93
		Ej01838	14.26	18.77	16.47	17.34
ethylene response sensor 1	ERS1	Ej13990	42.7	18.53	14.5	16.08
ethylene insensitive 4	EIN4	Ej24058	20.09	25.31	42.66	17.31
ethylene insensitive 2	EIN2	Ej01739	38.95	41.56	52.18	40.12
		Ej03322	33.91	30.82	32.52	30.54
ethylene insensitive 3	EIN3	Ej05301	139.1	112.06	206.83	144.67
		Ej25479	62.35	168.48	123.55	62.19
		Ej14756	80.18	45.06	50.05	65.18
EIN3-binding F-box protein 1/2	EBF1/2	Ej19156	413.07	10.61	4.98	51.36
		Ej28625	203.74	68.9	101.24	59.79
		Ej28890	153	4.28	4.83	46.67
		Ej21184	53.65	59.01	56.05	20.39
mitogen-activated protein kinase 3/6	MPK3/6	Ej17478	150.29	24.94	34.45	225.51
		Ej04995	28.17	49.3	86.55	25.94
ethylene responsive transcription factors	ERF	Ej07648	175.65	76.85	19.71	458.16
		Ej07149	204.94	57.46	4.98	403.41
		Ej00726	244.38	120.87	14.39	651.19
		Ej31684	126.8	74.5	27.1	163
		Ej31431	105.99	99.47	45.03	431.56
		Ej11153	189.27	31.16	12.84	271.92
		Ej27234	62.1	24.73	19.06	145.48
		Ej15906	47.07	42.48	45.39	135.37
		Ej11154	15.45	11.12	3.63	37.77
		Ej26889	5.05	28.39	0	64.78

Table S25. Triterpene compounds separated from *Euscaphis*

NO.	Name	Ref.	NO.	Name	Ref.
1	euscaphic acid	Cheng et al., 2010	16	1α,3β-dihydroxy-12-oleanen-28-oic acid	Ghosal et al., 1983
2	2*α*-hydroxypomolic acid	Cheng et al., 2010	17	ilexosapogenin A	Zhang et al., 2012
3	tormentic acid	Cheng et al., 2010; Abe et al., 2005	18	oleanic acid	Zhou et al., 2013
4	23-aldehydepomolic acid	Cheng et al., 2010	19	arjunic acid	Zhang et al., 2012
5	rotundic acid	Cheng et al., 2010	20～25	euscaphis acids A-F	Cheng et al., 2010
6	rotungenic acid	Cheng et al., 2010	26～28	euscaphis acids G-I	Zhang et al., 2012
7	2*α*-hydroxyursolic acid	Cheng et al., 2010	29	virgatic acid	Cheng et al., 2010
8	pomolic acid	Xiang et al., 2015	30	*3β*,23-dihydroxy-1-oxo-olean-12-en-28-oic aicd	Ghosal et al., 1983
9	methyl rotundate	Lee et al., 2009	31	(12*R*,13*S*)-3-methoxy-12,13-cyclo-taraxerene-2,14-diene-1-one-28-oic acid	Chang et al., 2006
10	euscaphic acid K	Zhang et al., 2012	32	euscaphic acid L	Zhang et al., 2012
11	2*α*,3*β*,19*α*-trihydroxyurs-12-ene-23,28-dioic acid-23-methyl ester	Zhang et al., 2012	33	pomonic acid	Xiang et al., 2015
12	ursolic acid	Xiang et al., 2015	34	glut-5-en-ol	Lee et al., 2009
13	maslinic acid	Xiang et al., 2015	35	euscaphic acid J	Zhang et al., 2012
14	2*α*,3*α*,23-trihydroxyolean-12-en-28-oic acid	Ghosal et al., 1983	36	friedline	Lee et al., 2009
15	hederagenin	Ghosal et al., 1983	37	betulinic acid	Cheng et al., 2010

Table S26. Number of putative pentacyclic triterpene-related genes in the malvids species.

Pathways	Enzyme	Acronym	Pfam Domains	*A. thaliana*	*E. japonica*	*T. cacao*	*C. papaya*	*D. longan*	*B. ceiba*	*P. trichocarpa*
MVA	Acetoacetyl-CoA thiolase	ACAT	PF02803/PF00108	2	5	6	2	6	3	3
	3-hydroxy-3-methylglutaryl CoA synthase	HMGS	PF01154/PF08540	3	2	6	1	2	5	3
	3-hydroxy-3-methylglutaryl CoA reductase	HMGR	PF00368	2	15	6	3	3	8	6
	Mevalonate kinase	MVK	PF00288	1	1	2	1	2	1	4
	5-phosphomevalonate kinase	PMK	PF00288/PF08544	1	2	1	1	1	2	3
	5-diphosphomevalonate decarboxylase	MDV	PF00288	1	1	4	1	1	2	2
MEP	1-deoxy-D-xylulose 5-phosphate synthase	DXS	PF13292/PF02780/PF02779	8	6	10	3	6	7	7
	1-deoxy-D-xylulose 5-phosphate reductoisomerase	DXR	PF13288/PF08436/PF02670	2	4	2	1	3	2	3
	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	MCT	PF01128	2	1	4	1	2	1	1
	4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase	CMK	PF08544/PF00288	1	1	1	0	1	1	2
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	MDS	PF02542	2	1	1	1	2	3	2

	1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase	HDS	PF04551	5	1	2	1	1	3	2
	1-hydroxy-2-methyl-2-butenyl 4-diphosphate reductase	HDR	PF02401	1	2	2	1	2	5	4
OSC groups	β-amyrin synthase	bAS	PF00432/PF13243	2	6	4	8	13	5	9
	lupeol synthase	LUS	PF00432	8	1	2	1	2	3	2
	Cycloartenol synthase	CAS	No	3	4	8	3	2	6	7
IPP isomerase	Isopentenyl diphosphate isomerase	IDI	PF00293	2	2	1	1	2	1	1
IPP-related downstream	Geranylgeranyl diphosphate synthase	GGPS	PF00348	7	7	4	1	6	13	5
	farnesyl diphosphate synthase	FPS	PF00348	2	2	4	1	2	3	2
	Squalene Synthase	SQS	PF00494	4	7	3	3	2	2	3
	Squalene epoxidase	SES	PF01266/PF08491/PF13450	5	7	4	1	3	8	1

Table S27. Identified pentacyclic triterpene synthesis-related genes in *E. japonica* genome.

Acronym	Gene ID	Name	Acronym	Gene ID	Name
ACAT	Ej.11320	EjAACT.1	HDR	Ej.06216	EjHDR.1
	Ej.19247	EjAACT.2		Ej.26952	EjHDR.2
	Ej.00954	EjAACT.3	bAS	Ej.10860	EjbAS.1
	Ej.25074	EjAACT.4		Ej.15179	EjbAS.2
	Ej.12375	EjAACT.5		Ej.22170	EjbAS.3
HMGS	Ej.11320	EjHMGS.1		Ej.22169	EjbAS.4
	Ej.25074	EjHMGS.2		Ej.22168	EjbAS.5
HMGR	Ej.15879	EjHMGR.1		Ej.03620	EjbAS.6
	Ej.27203	EjHMGR.2	LUS	Ej.13682	EjLAS
	Ej.30077	EjHMGR.3	CAS	Ej.25224	EjCAS.1
	Ej.25039	EjHMGR.4		Ej.13675	EjCAS.2
	Ej.22655	EjHMGR.5		Ej.13670	EjCAS.3
	Ej.03689	EjHMGR.6		Ej.13674	EjCAS.4
	Ej.02331	EjHMGR.7	IDI	Ej.00396	EjIDI.1
	Ej.03309	EjHMGR.8		Ej.22074	EjIDI.2
	Ej.14954	EjHMGR.9	GGPS	Ej.11901	EjGGPS.1
	Ej.30422	EjHMGR.10		Ej.22937	EjGGPS.2
	Ej.16369	EjHMGR.11		Ej.05959	EjGGPS.3
	Ej.15168	EjHMGR.12		Ej.11903	EjGGPS.4
	Ej.14953	EjHMGR.13		Ej.31474	EjGGPS.5
	Ej.19514	EjHMGR.14		Ej.17992	EjGGPS.6
	Ej.07560	EjHMGR.15		Ej.15313	EjGGPS.7
MVK	Ej.00182	EjMVK	FPS	Ej.31237	EjFPS.1
PMK	Ej.18316	EjPMK.1		Ej.00473	EjFPS.2
	Ej.03456	EjPMK.2	SQS	Ej.06100	EjSqs.1
MVD	Ej.30292	EjMVD		Ej.27034	EjSqs.2
DXS	Ej.11576	EjDXS.1		Ej.03747	EjSqs.3
	Ej.02279	EjDXS.2		Ej.23366	EjSqs.4
	Ej.31011	EjDXS.3		Ej.26320	EjSqs.5
	Ej.09770	EjDXS.4		Ej.04566	EjSqs.6
	Ej.15630	EjDXS.5		Ej.24490	EjSqs.7
	Ej.25913	EjDXS.6	SQE	Ej.26761	EjSqe.1
DXR	Ej.20141	EjDXR.1		Ej.09718	EjSqe.2
	Ej.15325	EjDXR.2		Ej.06445	EjSqe.3
	Ej.20142	EjDXR.3		Ej.05838	EjSqe.4
	Ej.20144	EjDXR.4		Ej.23630	EjSqe.5
MCT	Ej.27779	EjMCT		Ej.06741	EjSqe.6
CMK	Ej.21632	EjCMK		Ej.06444	EjSqe.7
MDS	Ej.21632	EjMDS			
HDS	Ej.21397	EjHDS			

Table S28. Statistical simple sequence repeat.

Type	Unit size (Repeat number)	number
P1	1(>=10)	624,989
P2	2(>=6)	183,636
P3	3(>=5)	51,795
P4	4(>=5)	11,424
P4	5(>=5)	2,478
P6	6(>=5)	1,407
Total identified SSR		875,729

Type: SSR Type, P1-single base repeat, P2-double base repeat, P3-triple base repeat, and so on.
Unit Size (Repeat Number): Number of SSR repeat bases (minimum number of repeats allowed).

Supplementary reference

Abe, I., Utsumi, Y., Oguro, S., Morita, H., Sano, Y., Noguchi, H. (2005) A plant type III polyketide synthase that produces pentaketide chromone. *J. Am. Chem. Soc*. **5**, 1362–1363, doi: 10.1021/ja0431206

An, X. H. Tian, Y., Chen, K. Q., Wang, X. F. & Hao, Y. J. (2012) The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. *J. Plant Physiol*. **169**, 710–717, doi: 10.1016/j.jplph.2012.01.015

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., Lepiniec, L. (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J*. **39**, 366–380, doi: 10.1111/j.1365-313X.2004.02138.x

Chang, X., Li, W., Koike, K., Wu, L., Nikaido, T. (2006) Phenolic constituents from the rhizomes of *Dryopteris crassirhizoma. Chem. Pharm. Bull.* **37**, 748–750, doi: 10.1248/cpb.54.748

Cheng, J. J., Zhang, L. J., Cheng, H. L., Chiou, C. T., Lee, I. J., Kuo, Y. H. (2010) Cytotoxic hexacyclic triterpene acids from *Euscaphis japonica*. *J. Nat. Prod*. **73**, 1655–8, doi: 10.1021/np1003593

El-Kereamy, A., Chervin, C., Roustan, J. P. Cheynier, V., Bouzayen, M. (2010) Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. *Physiol. Plantarum* **2**, 175–182, doi: 10.1034/j.1399-3054.2003.00165.x

Ghosal, S., Kumar, Y., Singh, S. (1983). Biflorin, a chromone-C-glucoside from Pancratium biflorum. *Phytochemistry* **22**, 2591–2593

Gu H., Zhao, Q., Zhang, Z.B. (2017) Does scatter-hoarding of seeds benefit cache owners or pilferers? *Integr. Zool*. **6**, 477–488, doi: 10.1111/1749-4877.12274

Holzwarth, M. Korhummel. S., Kammerer, D. R., Carle, R. (2012) Thermal inactivation of strawberry polyphenoloxidase and its impact on anthocyanin and color retention in strawberry (*Fragaria* x *ananassa* Duch.) purées. *Eur. Food Res. Technol* **6**, 1171–1180, doi: 10.1007/s00217-012-1852-2

Honda, C., Bessho, H., Murai, M., Iwanami, H., Tatsuki, M. (2014) Effect of temperature on anthocyanin synthesis and ethylene production in the fruit of early- and medium-maturing apple cultivars during ripening stages. *HortScience*. **12**, 1510–1517.

Lee, M. K., Lee, K. Y., Jeon, H. Y., Sung, S. H., Kim, Y. C. (2009) Antifibrotic activity of triterpenoids from the aerial parts of *Euscaphis japonica* on hepatic stellate cells. *J Enzyme. Inhib. Med Chem*. **24**, 1276–1279, doi: 10.3109/14756360902829709

Urgoiti, J., Muñoz, A., Espelta, J. M., Bonal, R. (2018) Distribution and space use of seed-dispersing rodents in central Pyrenees: implications for genetic diversity, conservation and plant recruitment. *Integr. Zool*. **3**, 307–318, doi: 10.1111/1749-4877.12301

Xiang, D.B., Qiao, H. U., Tan, Y., Meng, Y. C., Pei, G. (2015) Isolation and identification of triterpenoids from fruits of *Euscaphis fukienensis. Chin. Tradit. Pat. Med*. **37**, 793–796

Xiao, Z. S. & Huang, X. Q. (2020) How seed defense and seed abundance predict dispersal and survival patterns in Camellia. *Integr. Zool*. **2**, 1749–4877, doi: 10.1111/1749-4877.12408

Zhang, L. J., Cheng, J. J., Liao, C. C., Cheng, H. L., Huang, H. T. et al. (2012) Triterpene acids from *Euscaphis japonica* and assessment of their cytotoxic and anti-NO activities. *Planta Med*. **78,** 1584–1590, doi: 10.1055/s-0032-1315040

Zhou, W., Liu, Z., Wang, H. J., Li, Y. J., Liao, S. G. (2013) Study on chemical constituents of ethyl acetate fraction of *Euscaphis japonica*. *J. Chin. Exper. Tradit. Med. For*. **19**, 121–123