

Cultivated hawthorn (*Crataegus pinnatifida* var. major) genome sheds light on the evolution of Maleae (apple tribe)

Ticao Zhang^{1†*}, Qin Qiao^{2†*}, Xiao Du^{3,4}, Xiao Zhang³, Yali Hou³, Xin Wei³, Chao Sun³, Rengang Zhang⁵, Quanzheng Yun⁵, M. James C. Crabbe^{6,7,8}, Yves Van de Peer^{9,10,11}, Wenxuan Dong^{3*}

SUPPORTING INFORMATION



Figure S1. Information of collected sample

(A) Aerial panoramic view of National Hawthorn Germplasm Repository of China at Shenyang Agricultural University. (B) The sequenced individual of hawthorn cultivar 'Qiujiexing' (<https://www.cgris.net/query/croplist.php>, identification number: SZP016). (C) The fruits picture of hawthorn cultivar 'Qiujiexing'.

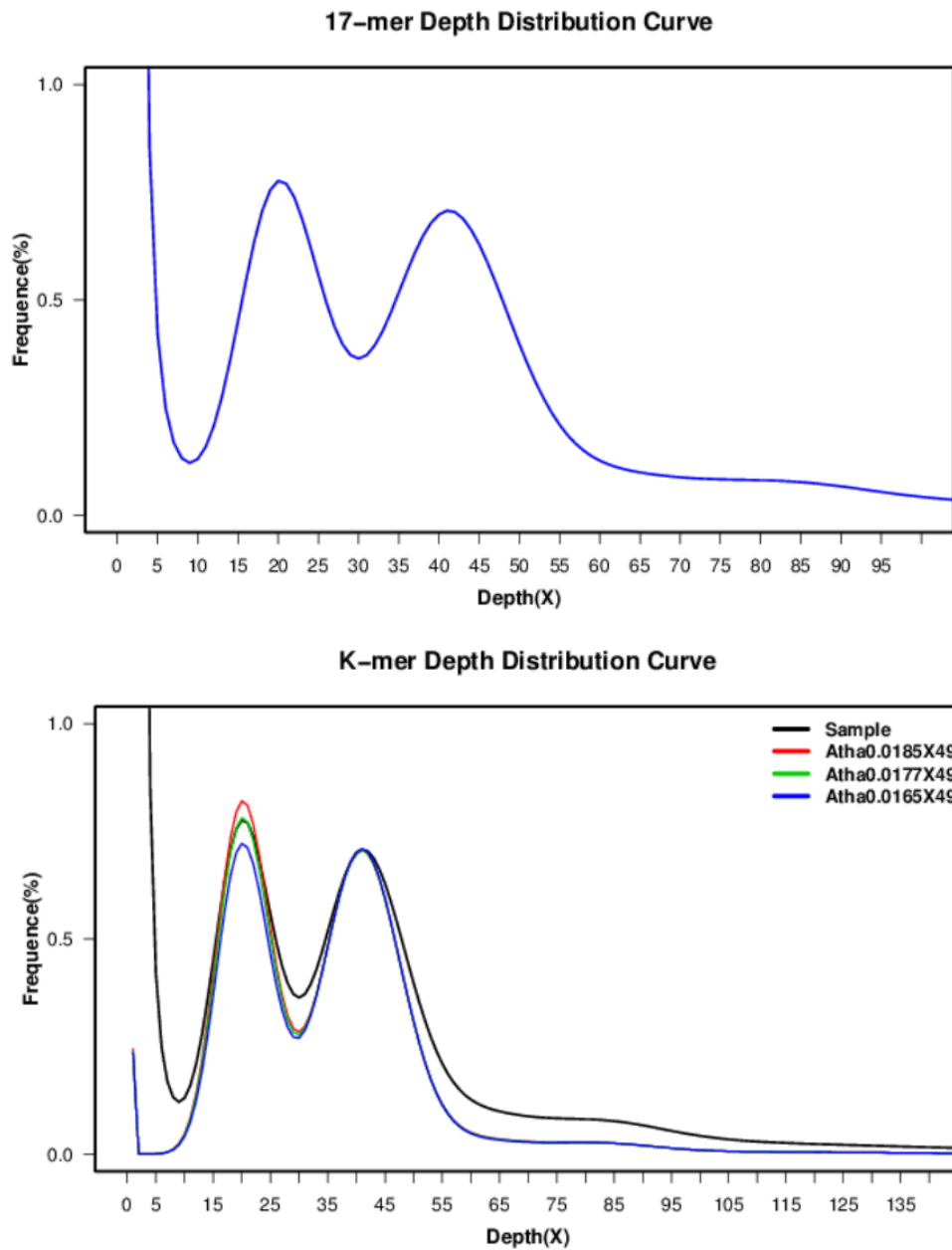


Figure S2. Frequency distribution of depth of 17-mer (upper) and *K*-mer (below) in genome survey of cultivated hawthorn

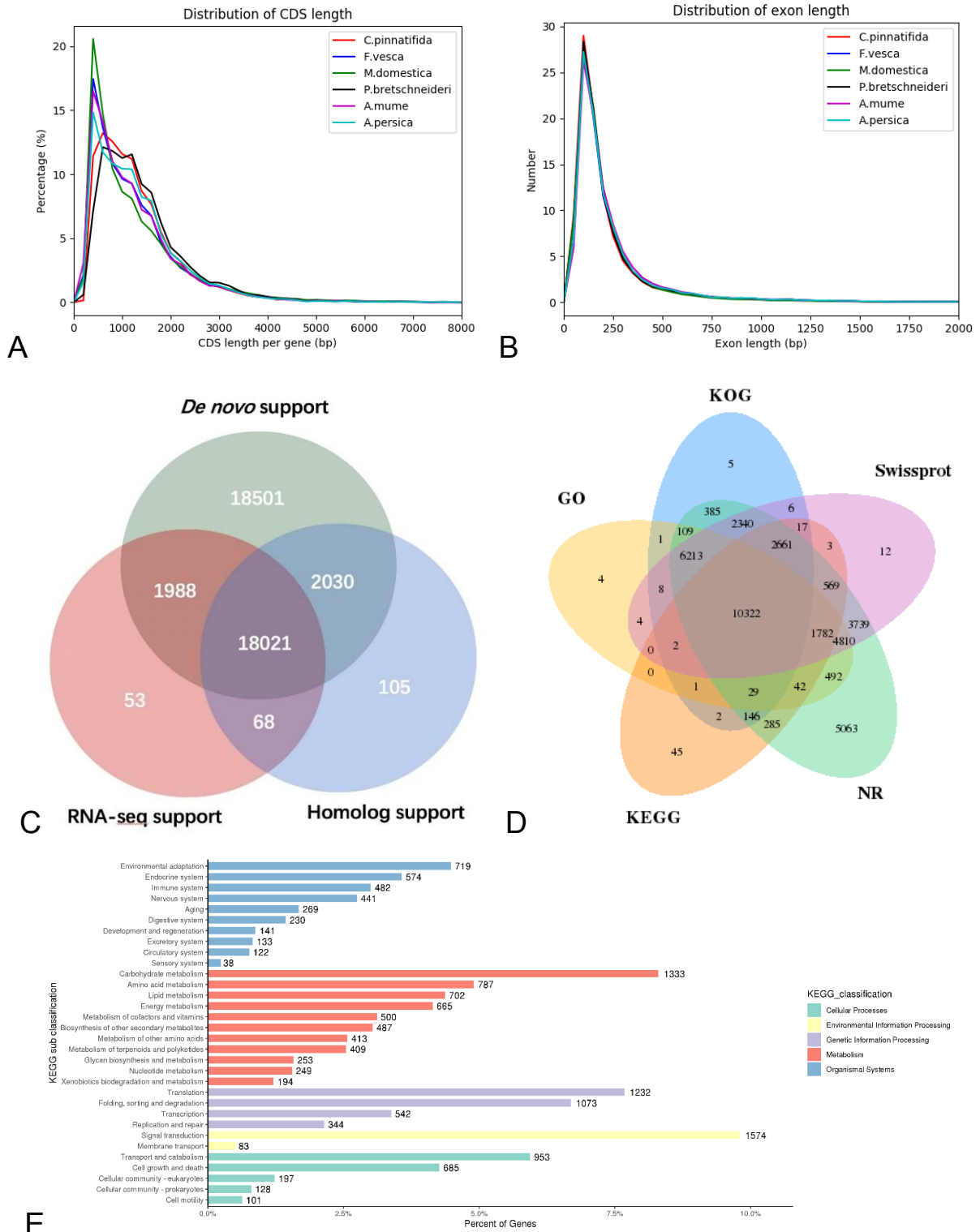


Figure S3 The genome annotation of hawthorn

(A) The distribution of CDS length. (B) The distribution of exon length. (C) Gene prediction using three methods. (D) Gene functional annotation in five databases. (E) KEGG enrichment analysis of predicted genes.

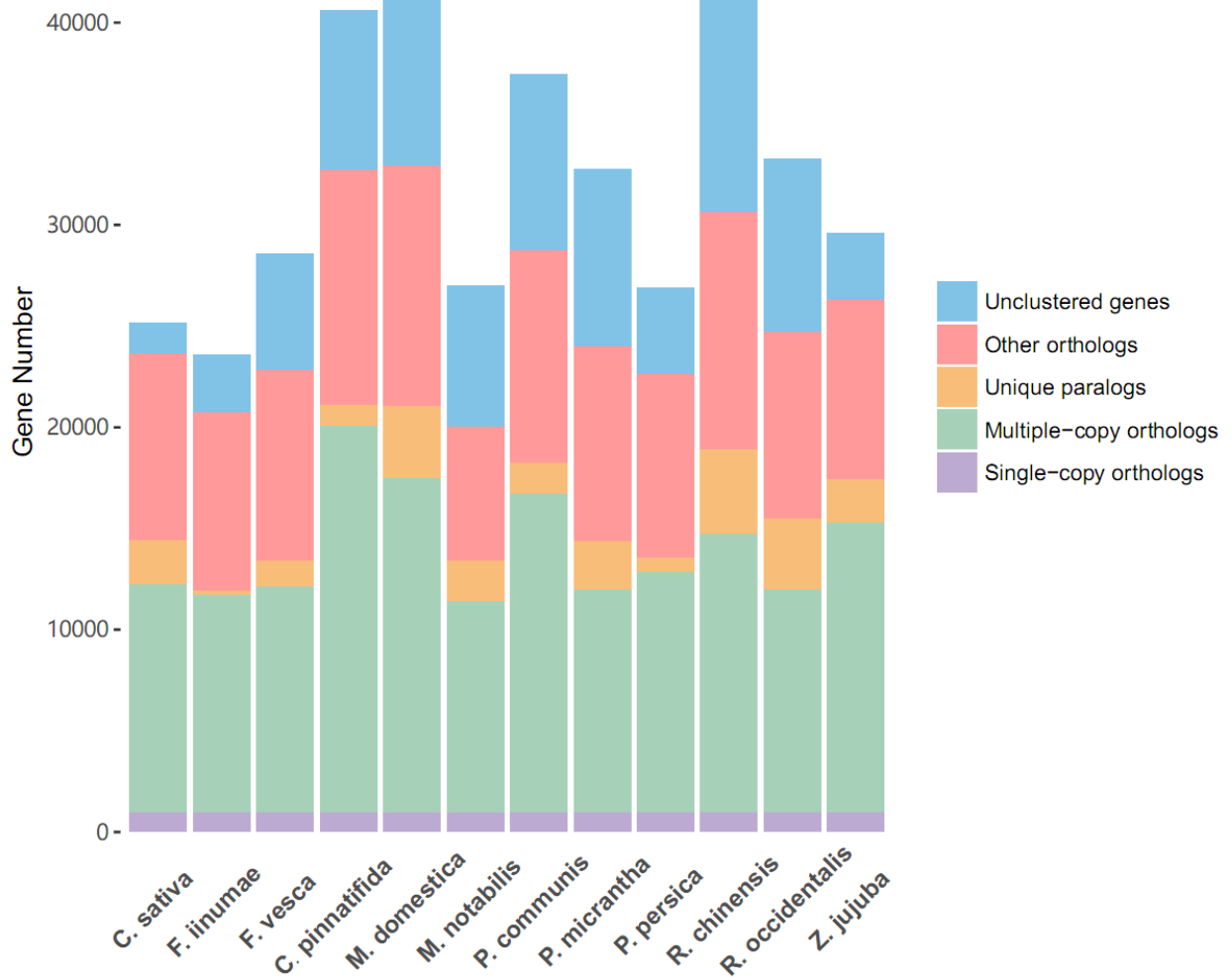


Figure S4. Classification statistics of cultivated hawthorn genes

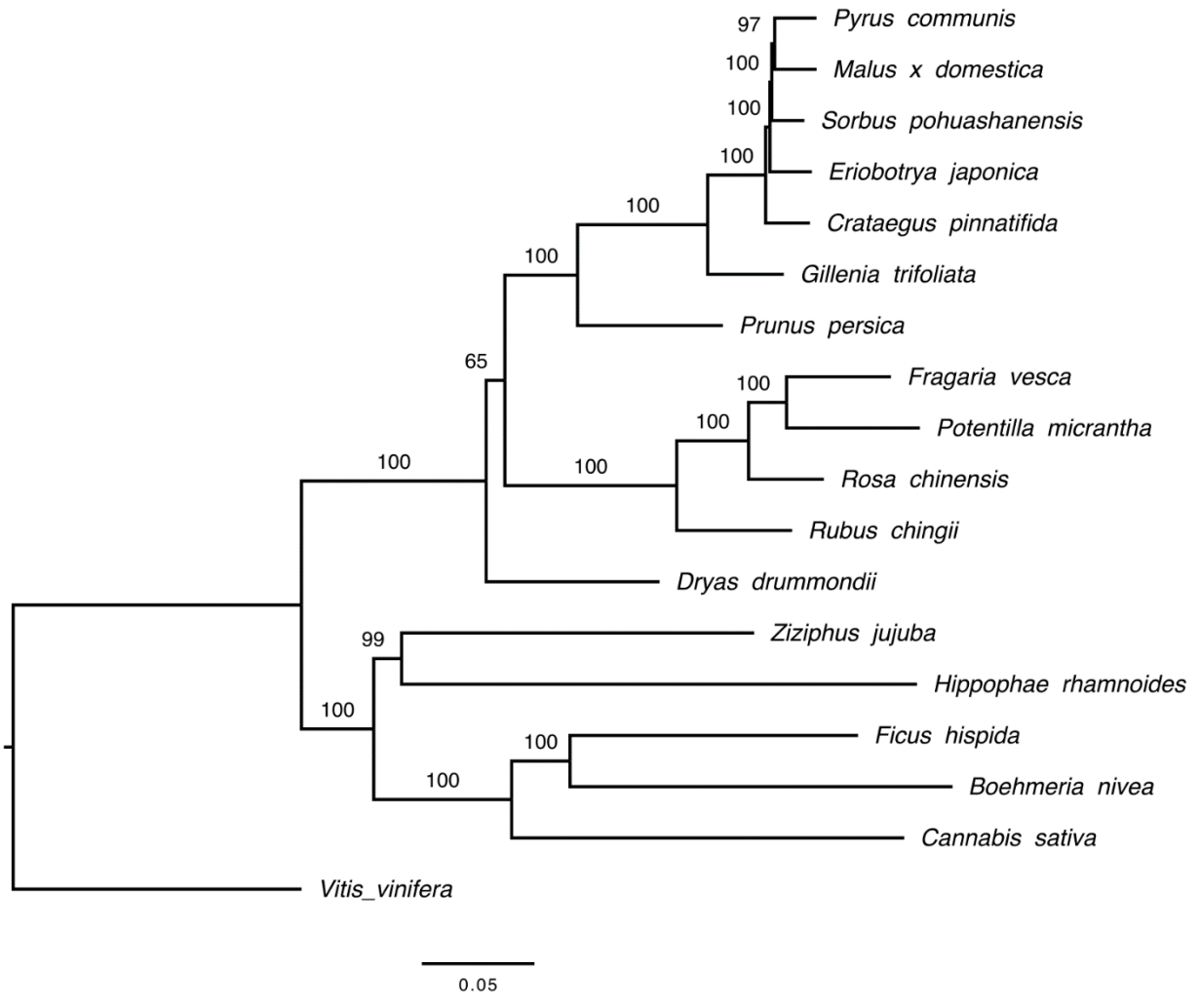


Figure S5. The Maximum likelihood phylogenetic tree of Rosaceae with bootstraps

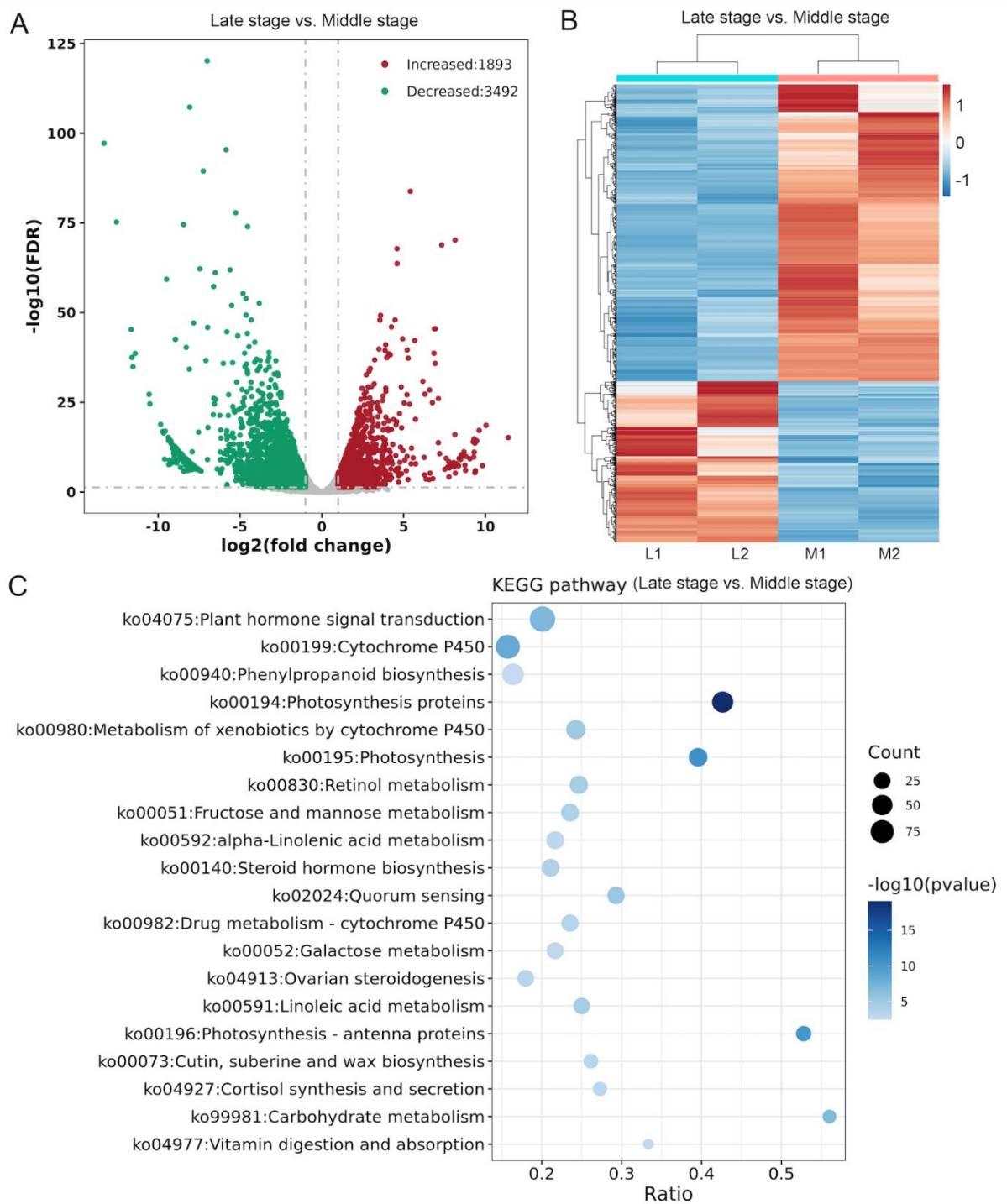


Figure S6. Changes in mRNA expression in hard-fleshed hawthorn ‘Qiujinxing’ (A) Volcano plots showing up-and down-regulated mRNAs of late and middle stages of fruit development, respectively. (B) Heatmap of different expressed genes between two stages. (C) KEGG enrichment analysis of different expressed genes between two stages.

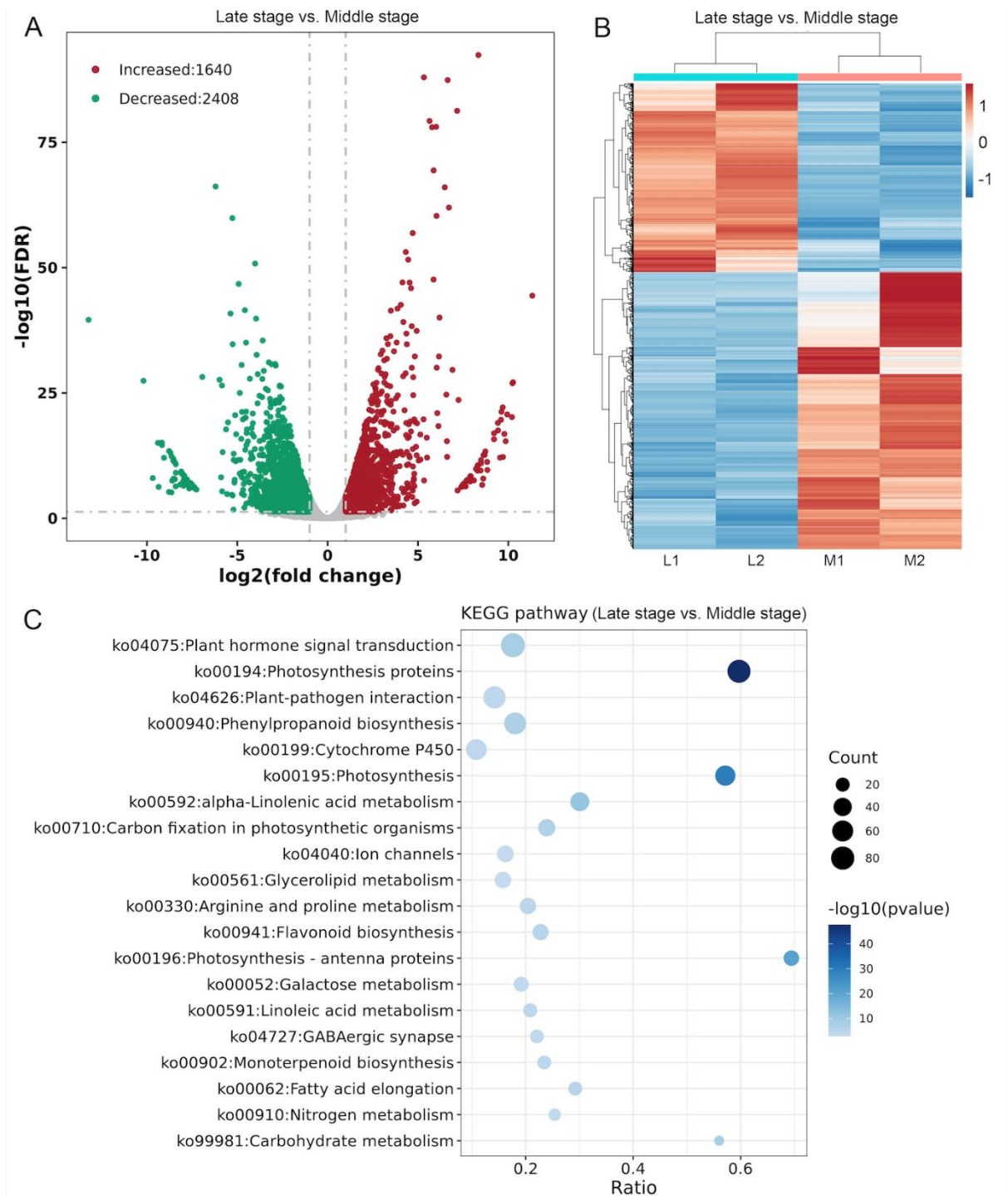


Figure S7. Changes in mRNA expression in soft-fleshed hawthorn 'Ruanrou Shanlihong #3'

(A) Volcano plots showing up-and down-regulated mRNAs of late and middle stages of fruit development, respectively. (B) Heatmap of different expressed genes between two stages. (C) KEGG enrichment analysis of different expressed genes between two stages.

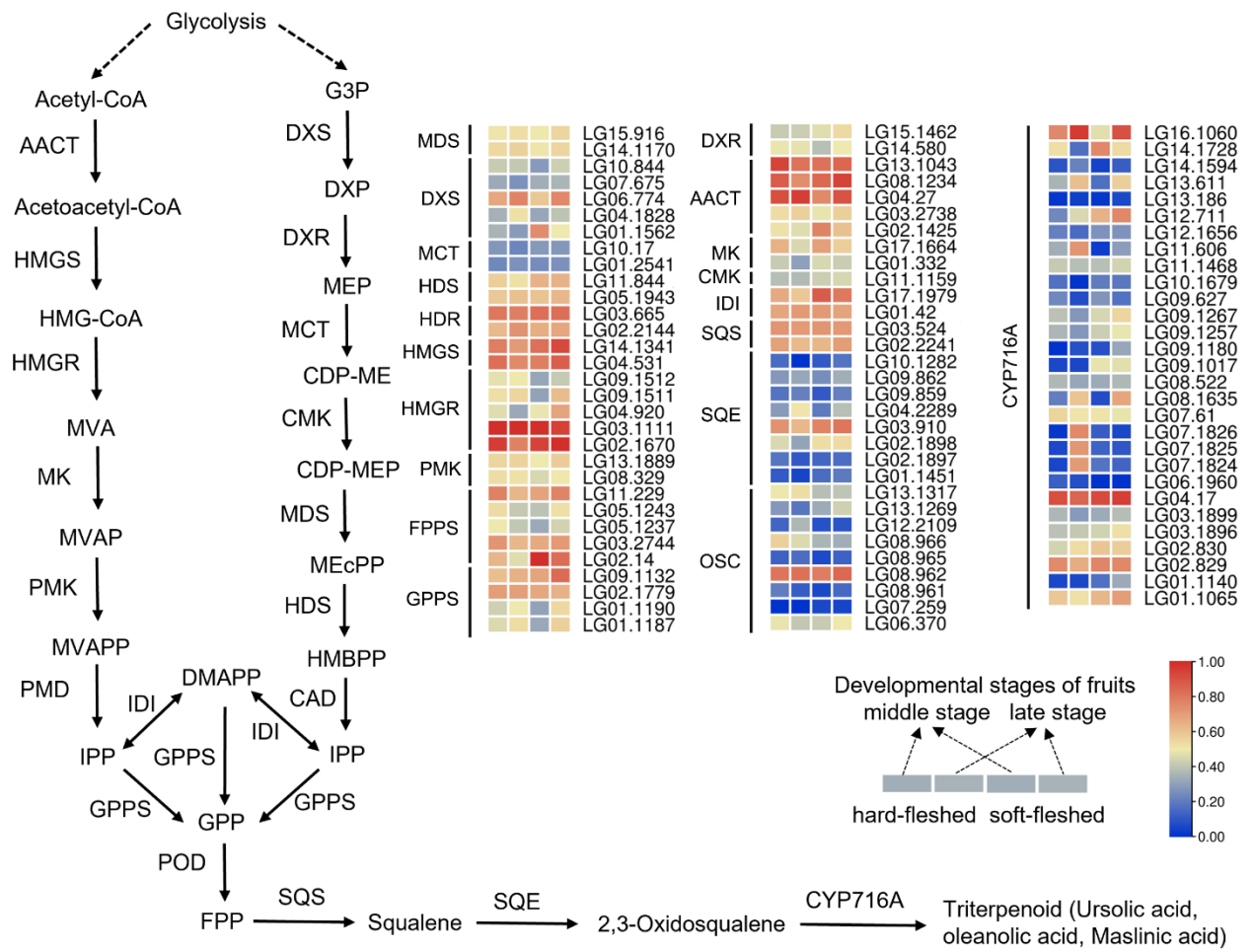


Figure S8. Triterpene biosynthesis pathway in *Crataegus pinnatifida*

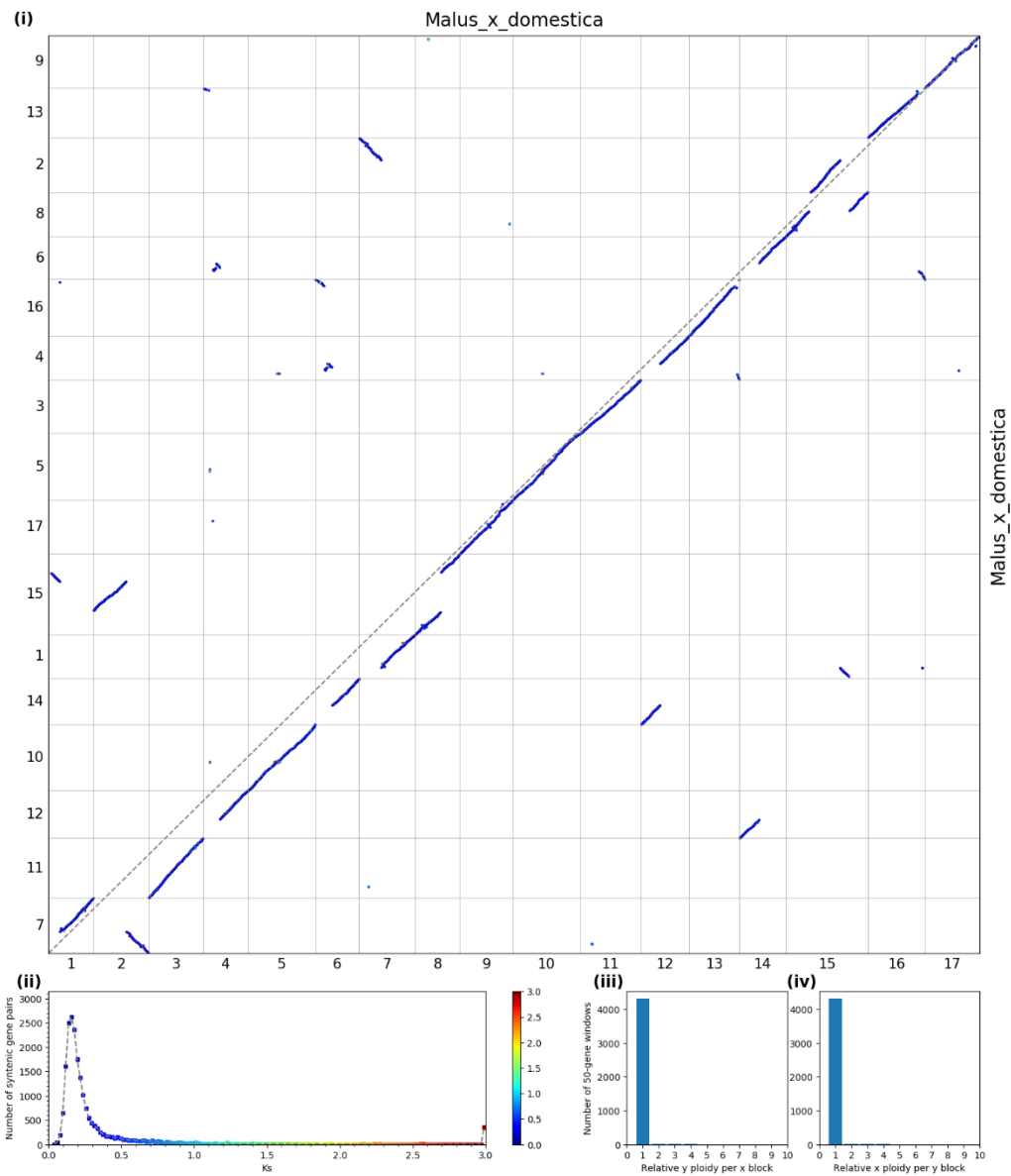


Figure S9. Syntenic dot plot and Ks distribution within the apple genome

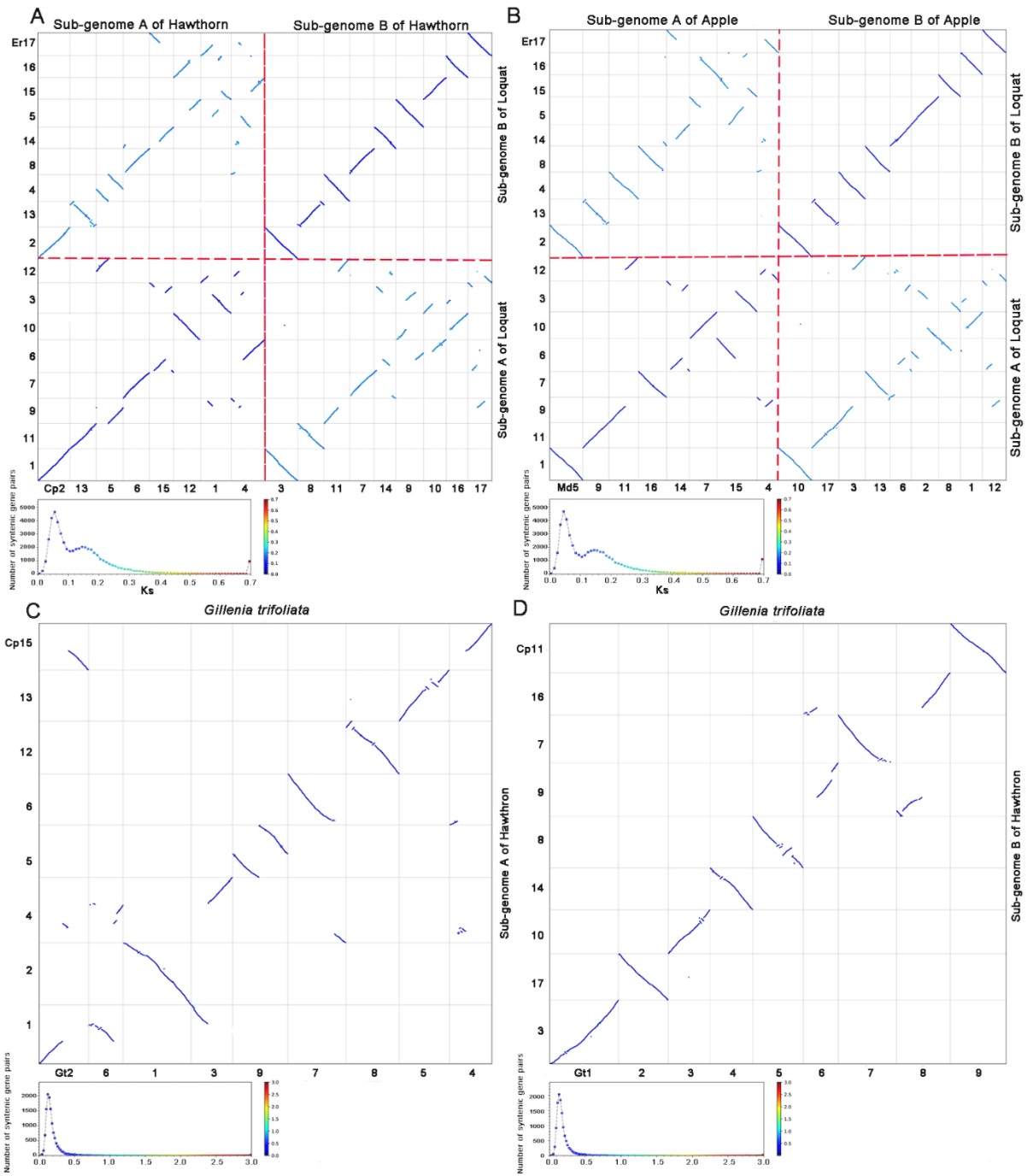


Figure S10. Syntenic dot plot and Ks distribution between two sub-genomes of hawthorn and Loquat (**A**), apple and Loquat (**B**). Syntenic dot plot and Ks distribution between *Gillenia trifoliata* and sub-genome A of hawthorn (**C**) and sub-genome B of hawthorn (**D**).

Table S1. Statistics of genome survey data

Parameters	Cultivated hawthorn
Raw Reads	815,627,604
Raw Bases (bp)	122,344,140,600
Clean Reads	778,206,444
Clean Bases (bp)	112,839,934,380
Insert Size (bp)	400
Read Length (bp)	150
Q30	0.92
GC content	38%
<i>K</i> -mer	17
<i>K</i> -mer number	35,132,166,024
<i>K</i> -mer depth	41
Estimated genome size (bp)	856,882,098
Estimated heterozygosity (%)	1.77

Table S2. Statistics of paired-end reads based on Hi-C technology

Parameters	Cultivated hawthorn
Clean Paired-end Reads	290,205,090
Unmapped Paired-end Reads	14,557,915
Unmapped Paired-end Reads Rate (%)	5.02
Paired-end Reads with Singleton	55,740,825
Paired-end Reads with Singleton Rate (%)	19.21
Multi Mapped Paired-end Reads	95,404,870
Multi Mapped Ratio (%)	32.87
Unique Mapped Paired-end Reads	124,501,480
Unique Mapped Ratio (%)	42.90
Dangling End Paired-end Reads	68,300,176
Self Circle Paired-end Reads	83,120
Dumped Paired-end Reads	17,588,100
Valid Paired-end Reads	31,256,516
Valid reads of unique mapping reads (%)	25.11
Valid reads of clean reads (%)	10.77

Table S3. Statistics of the lengths of 17 pseudo-chromosomes in the cultivated hawthorn genome

Chromosomes	Size (bp)	Scaffolds Num.
LG01	55,836,375	52
LG02	53,314,146	34
LG03	53,092,358	59
LG04	52,245,379	44
LG05	51,806,852	31
LG06	47,353,093	39
LG07	45,635,702	40
LG08	44,939,157	33
LG09	44,935,142	42
LG10	44,156,052	28
LG11	43,380,348	49
LG12	43,192,490	43
LG13	42,315,842	27
LG14	40,841,768	29
LG15	40,501,693	37
LG16	39,678,409	41
LG17	36,018,979	38
Total	779,243,785	666

Table S4. Predicted genes and gene features of the cultivated hawthorn

Gene set	Total number of genes	Average gene length(bp)	Average CDS length(bp)	Average exons number per gene	Average exon length(bp)	Average intron length(bp)
<i>De novo</i>	40,540	3,144.86	1,212.26	5.85	207.22	398.46
Homology	20,224	2,969.66	1,342.27	5.25	255.5	382.61
RNA-seq	20,130	3,627.87	1,384.39	6.27	220.85	425.82
Final set	40,571	3,352.32	1,227.45	5.6	219.2	461.97

Table S5. Gene functional annotation of the cultivated hawthorn

Type		Number	Percent (%)
Annotation	Swissprot	32,488	80.08
	KEGG	15,906	39.21
	KOG	22,247	54.83
	GO	23,819	58.71
	NR	38,987	96.10
Total	Annotated	39,097	96.37
	Gene	40,571	-

Table S6. Predicted RNA features of the cultivated hawthorn

Type		Copy Number	Average Length(bp)	Total Length(bp)	Percentage of sequence(%)
rRNA (5,294)	18S	13	1,707.92	22,203	0.0027
	28S	11	4,391.09	48,302	0.0059
	5.8S	2	148.00	296	0.0000
	5S	5,268	114.35	602,374	0.0732
Small RNA (1,417)	snRNA	289	99.35	28,713	0.0035
	miRNA	152	135.49	20,595	0.0025
	spliceosomal	315	159.58	50,269	0.0061
	other	661	117.20	77,466	0.0094
Regulatory	cis-regulatory elements	11	70.27	773	0.0001
tRNA	tRNA	782	75.12	58,744	0.0071

Table S7. Conserved genes using the BUSCO method

Type	Number	Percent (%)
Complete BUSCOs (C)	1,342	97.60
Complete and single-copy BUSCOs (S)	848	61.67
Complete and duplicated BUSCOs (D)	494	35.93
Fragmented BUSCOs (F)	8	0.58
Missing BUSCOs (M)	25	1.82
Total BUSCO groups searched	1,375	100

Table S8. The enriched KEGG functional categories of species-specific genes (*P* value < 0.05) in the cultivated hawthorn

Categories	<i>P</i> value	Gene Num.	Num In Categories	Term	Gene ID
ko00220	0.000712	15	68	Arginine biosynthesis	LG0.751, LG01.970, LG01.971, LG01.1804, LG05.316, LG06.2166, LG07.2025, LG07.2033, LG08.525, LG08.532, LG12.541, LG13.1766, LG15.24, LG16.1238
ko00250	0.00114	18	93	Alanine, aspartate and glutamate metabolism	LG0.751, LG01.1595, LG01.1722, LG02.1041, LG02.1042, LG05.316, LG05.376, LG06.2166, LG07.2025, LG07.2033, LG10.314, LG10.316, LG11.1050, LG12.541, LG13.1766, LG15.24, LG16.125, LG16.1238
ko04141	0.001434	57	434	Protein processing in endoplasmic reticulum	LG0.1039, LG0.1040, LG0.2225, LG0.2226, LG01.1295, LG01.1296, LG01.1661, LG02.390, LG02.565, LG02.2037, LG02.2378, LG03.980, LG03.2162, LG04.739, LG04.740, LG04.1392, LG04.1692, LG04.2556, LG04.2738, LG04.2739, LG05.1622, LG05.1628, LG06.172, LG07.820, LG07.1597, LG07.1598, LG07.1600, LG07.1714, LG08.476, LG09.476, LG09.1126, LG09.1324, LG09.1815, LG09.1939, LG10.255, LG10.256, LG10.376, LG10.1116, LG10.1550, LG11.618, LG11.1183, LG12.246, LG12.855, LG13.997, LG13.1618, LG14.1757, LG15.484, LG15.485, LG15.498, LG15.499, LG15.500, LG15.1920, LG16.1005, LG16.1007, LG16.1593, LG17.1076
ko04136	0.004056	16	88	Autophagy - other	LG03.887, LG04.285, LG05.239, LG05.1625, LG07.1238, LG08.856, LG08.1385, LG09.99, LG12.1106, LG13.316, LG13.1701, LG14.1262, LG14.1369, LG15.878, LG17.1106, LG17.1357
ko00340	0.00426	10	44	Histidine metabolism	LG0.1314, LG0.2081, LG0.2082, LG05.215, LG07.1300, LG07.1300, LG11.813, LG12.1492, LG14.584, LG14.590
ko03008	0.010979	25	175	Ribosome biogenesis in eukaryotes	LG0.291, LG01.1125, LG01.1503, LG01.1504, LG01.1505, LG01.1967, LG02.2228, LG02.2229, LG02.2564, LG03.1663, LG03.1663, LG04.2225, LG05.89, LG06.245, LG06.909, LG07.784, LG08.652, LG10.1208, LG10.1209, LG12.2037, LG14.1249, LG15.534, LG15.722, LG16.1801, LG17.666
ko01240	0.011315	62	527	Biosynthesis of cofactors	LG0.688, LG0.1346, LG0.1605, LG0.2054, LG0.2055, LG0.2082, LG0.2140, LG0.2191, LG01.195, LG01.510, LG01.1510, LG01.1511, LG01.1722, LG01.1735, LG01.1852, LG01.1852, LG01.2332, LG01.2392, LG02.769, LG02.812, LG02.1012, LG03.62, LG03.80, LG04.783, LG05.215

					LG05.406, LG05.493, LG06.175, LG06.264, LG07.1300, LG07.33, LG07.67, LG07.812, LG07.1300, LG08.1030, LG09.407, LG09.410, LG10.132, LG10.316, LG10.627, LG11.121, LG11.123, LG13.132, LG13.133, LG13.1147, LG13.1777, LG13.1925, LG14.584, LG14.590, LG15.930, LG15.930, LG16.475, LG16.479, LG16.480, LG16.1600, LG16.1600, LG17.373, LG17.467
ko03015	0.012916	37	288	mRNA surveillance pathway	LG0.389, LG0.391, LG0.393, LG0.394, LG0.395, LG0.396, LG0.1406, LG01.806, LG02.1288, LG02.1290, LG02.1291, LG03.471, LG03.473, LG03.476, LG03.477, LG03.478, LG04.2042, LG04.2043, LG05.716, LG06.1320, LG08.316, LG08.1282, LG09.99, LG12.771, LG12.2118, LG13.416, LG13.1561, LG13.1850, LG16.309, LG16.797, LG16.798, LG16.800, LG16.977, LG16.1253, LG16.1481, LG16.1482, LG17.743
ko00903	0.013404	5	17	Limonene and pinene degradation	LG0.2081, LG0.2082, LG07.1300, LG14.584, LG14.590
ko00020	0.014665	17	109	Citrate cycle (TCA cycle)	LG0.1346, LG01.1471, LG01.1472, LG03.1413, LG03.1796, LG03.1796, LG05.15, LG07.812, LG08.1687, LG09.2118, LG10.616, LG11.1386, LG13.265, LG13.1241, LG13.1710, LG13.1710, LG17.373
ko00910	0.016721	11	61	Nitrogen metabolism	LG0.751, LG05.316, LG05.2171, LG06.2166, LG07.2033, LG07.2033, LG09.928, LG10.596, LG12.541, LG15.260, LG16.1238
ko00053	0.016919	19	128	Ascorbate and aldarate metabolism	LG0.688, LG0.1017, LG0.2081, LG0.2082, LG01.1764, LG01.1764, LG02.769, LG03.963, LG04.776, LG04.1143, LG05.1017, LG05.1638, LG07.31, LG07.1300, LG09.410, LG09.410, LG13.1537, LG14.584, LG14.590
ko00770	0.017358	15	94	Pantothenate and CoA biosynthesis	LG0.2081, LG0.2082, LG01.1851, LG01.1852, LG05.716, LG06.780, LG07.58, LG07.1300, LG08.406, LG14.590, LG14.590, LG15.260, LG16.475, LG16.479, LG16.480
ko00500	0.018879	41	334	Starch and sucrose metabolism	LG01.1735, LG02.406, LG02.833, LG04.621, LG04.1143, LG04.2121, LG05.1597, LG06.1249, LG06.1333, LG07.1518, LG08.1427, LG09.455, LG09.462, LG09.1803, LG09.1804, LG10.509, LG10.515, LG10.785, LG10.860, LG10.1316, LG10.1676, LG11.515, LG11.1143, LG11.2082, LG12.1600, LG13.548, LG14.764, LG14.1704, LG15.1016, LG15.1017

					LG15.1068, LG16.61, LG16.403, LG16.406, LG16.409, LG16.410, LG16.621, LG16.1008, LG16.1311, LG17.1182
ko00071	0.019806	16	104	Fatty acid degradation	LG0.2081, LG0.2082, LG02.1429, LG02.2041, LG02.2582, LG03.90, LG03.93, LG03.97, LG03.99, LG03.399, LG03.1311, LG07.1300, LG14.584, LG14.590, LG17.1235
ko00010	0.023007	32	252	Glycolysis / Gluconeogenesis	LG0.688, LG0.1346, LG0.2081, LG0.2082, LG01.1619, LG01.1619, LG02.2041, LG02.2582, LG03.90, LG03.97, LG03.99, LG04.1169, LG07.657, LG07.812, LG07.1300, LG08.810, LG08.1687, LG09.407, LG09.407, LG09.1803, LG09.1804, LG11.438, LG11.439, LG11.584, LG12.1231, LG13.1740, LG14.584, LG14.590, LG14.716, LG14.1126, LG17.373
ko00280	0.026708	15	99	Valine, leucine and isoleucine degradation	LG0.1346, LG0.2081, LG0.2082, LG02.1429, LG03.399, LG03.1309, LG03.1311, LG07.812, LG07.1300, LG09.1803, LG12.1064, LG14.584, LG14.590, LG16.125, LG17.373
ko00780	0.028552	8	42	Biotin metabolism	LG0.2191, LG05.406, LG06.175, LG07.67, LG08.1064, LG10.655, LG11.121, LG11.123
ko00380	0.029454	22	163	Tryptophan metabolism	LG0.1346, LG0.2081, LG0.2082, LG01.1239, LG01.1454, LG03.397, LG03.939, LG03.943, LG05.2171, LG06.2171, LG07.295, LG07.299, LG07.812, LG07.1300, LG08.1584, LG11.1386, LG11.1984, LG11.1993, LG14.119, LG14.584, LG14.590, LG17.373
ko00730	0.032461	8	43	Thiamine metabolism	LG01.2392, LG04.783, LG07.675, LG07.676, LG10.1064, LG13.131, LG13.132, LG13.133
ko00562	0.033678	21	156	Inositol phosphate metabolism	LG01.1624, LG01.1625, LG02.47, LG02.2408, LG03.812, LG04.1141, LG04.2675, LG05.1520, LG05.1625, LG07.1235, LG08.856, LG10.180, LG11.463, LG11.464, LG13.399, LG13.1537, LG15.878, LG16.506, LG17.1238, LG17.1238, LG17.1697

Table S9. The enriched KEGG functional categories of significantly (P value < 0.05) expanded genes in the cultivated hawthorn

Categories	P value	Gene Num.	Num In Categories	Term	Gene ID
ko00944	1.01E-31	22	27	Flavone and flavonol biosynthesis	LG08.1578, LG08.1579, LG08.1580, LG08.1581, LG08.1582, LG08.1583, LG08.1585, LG08.1666, LG08.1667, LG08.1668, LG08.1669, LG08.1670, LG08.1716, LG08.1717, LG08.1719, LG13.531, LG13.584, LG13.585, LG13.586, LG13.587, LG13.588, LG13.669
ko04626	1.34E-25	55	421	Plant-pathogen interaction	chr0.1472, chr0.1482, LG01.741, LG02.170, LG02.1685, LG02.2376, LG03.379, LG03.1100, LG03.2593, LG04.2781, LG05.31, LG05.78, LG05.251, LG05.1624, LG05.1626, LG05.1629, LG05.1630, LG05.1632, LG05.1633, LG05.1635, LG05.1636, LG08.1018, LG08.1497, LG09.491, LG09.1315, LG10.65, LG10.66, LG10.68, LG10.70, LG10.72, LG10.1826, LG11.556, LG11.558, LG11.1172, LG11.1245, LG11.1409, LG12.1821, LG13.759, LG13.1304, LG13.1387, LG14.470, LG14.1144, LG14.1627, LG14.1631, LG14.1632, LG15.374, LG15.385, LG15.603, LG15.939, LG15.1568, LG17.346, LG17.360, LG17.538, LG17.541, LG17.1247
ko00943	4.63E-25	22	42	Isoflavonoid biosynthesis	LG08.1578, LG08.1579, LG08.1580, LG08.1581, LG08.1582, LG08.1583, LG08.1585, LG08.1666, LG08.1667, LG08.1668, LG08.1669, LG08.1670, LG08.1716, LG08.1717, LG08.1719, LG13.531, LG13.584, LG13.585, LG13.586, LG13.587, LG13.588, LG13.669
ko00330	1.76E-12	22	137	Arginine and proline metabolism	LG08.1578, LG08.1579, LG08.1580, LG08.1581, LG08.1582, LG08.1583, LG08.1585, LG08.1666, LG08.1667, LG08.1668, LG08.1669, LG08.1670, LG08.1716, LG08.1717, LG08.1719, LG13.531, LG13.584, LG13.585, LG13.586, LG13.587, LG13.588, LG13.669
ko00909	1.83E-10	15	73	Sesquiterpenoid and triterpenoid biosynthesis	chr0.1310, chr0.1613, chr0.1614, LG02.1897, LG02.1898, LG02.1901, LG03.910, LG04.2289, LG09.335, LG09.344, LG09.346, LG09.862, LG10.1282, LG12.1986, LG12.2018
ko00902	5.61E-10	16	74	Monoterpenoid biosynthesis	LG05.442, LG06.399, LG08.1187, LG08.1188, LG08.1190, LG08.1191, LG10.542, LG10.543, LG10.1054, LG13.1096, LG13.1097, LG13.1098, LG13.1130, LG13.2149, LG16.199, LG16.200
ko00100	7.19E-10	15	80	Steroid biosynthesis	chr0.1310, chr0.1613, chr0.1614, LG02.1897, LG02.1898, LG02.1901, LG03.910, LG04.2289, LG09.335, LG09.344, LG09.346, LG09.862, LG10.1282, LG12.1986, LG12.2018
ko00950	7.49E-07	9	43	Isoquinoline alkaloid biosynthesis	LG03.2011, LG03.2012, LG03.2013, LG03.2014, LG03.2015, LG03.2016, LG03.2017, LG03.2018, LG03.2019

ko00592	1.60E-06	15	140	alpha-Linolenic acid metabolism	LG01.2154, LG01.2155, LG01.2156, LG01.2157, LG01.2158, LG01.2159, LG01.2160, LG01.2161, LG01.2162, LG01.2163, LG01.2164, LG01.2165, LG01.2166, LG01.2167, LG10.302
ko00040	2.54E-06	19	225	Pentose and glucuronate interconversions	LG01.181, LG04.941, LG05.932, LG05.933, LG06.2019, LG06.2020, LG06.2021, LG06.2022, LG07.1888, LG07.1894, LG09.727, LG09.1541, LG12.452, LG12.1044, LG12.1045, LG16.728, LG16.1332, LG16.1333, LG16.1334
ko00460	2.81E-06	14	128	Cyanoamino acid metabolism	chr0.1361, chr0.1362, LG05.1731, LG05.1732, LG05.1733, LG05.1734, LG05.1735, LG05.1736, LG05.1740, LG05.1741, LG05.1742, LG11.639, LG11.640, LG11.641
ko00860	0.000243	10	105	Porphyrin and chlorophyll metabolism	LG01.2329, LG01.2330, LG01.2335, LG01.2337, LG01.2352, LG01.2360, LG01.2364, LG01.2369, LG01.2381, LG10.142
ko02010	0.000407	9	92	ABC transporters	chr0.1134, chr0.1135, LG08.1851, LG08.1852, LG08.1858, LG08.1859, LG13.370, LG13.371, LG13.373
ko04016	0.000412	18	300	MAPK signaling pathway - plant	chr0.522, chr0.523, chr0.1496, chr0.1499, LG01.1276, LG01.1498, LG01.1943, LG01.1944, LG01.2397, LG02.2134, LG03.713, LG03.716, LG04.423, LG09.1057, LG10.131, LG10.530, LG14.1480, LG16.212
ko04075	0.000471	24	467	Plant hormone signal transduction	chr0.522, chr0.523, chr0.1496, chr0.1499, LG01.1276, LG01.1498, LG01.1943, LG01.1944, LG01.2397, LG02.2134, LG03.713, LG03.716, LG04.423, LG09.1057, LG10.131, LG10.530, LG13.1394, LG14.1480, LG15.1508, LG15.1511, LG15.1512, LG15.1513, LG16.212, LG16.392
ko04145	0.017519	9	161	Phagosome	LG02.510, LG03.2214, LG03.2216, LG03.2217, LG03.2218, LG03.2219, LG03.2221, LG03.2223, LG13.1843
ko00480	0.025644	9	172	Glutathione metabolism	LG11.1999, LG11.2000, LG11.2002, LG11.2003, LG11.2004, LG11.2006, LG11.2007, LG11.2008, LG11.2010

Table S10. The enriched KEGG functional categories of significantly (P value < 0.05) differentially expressed genes between two fruit developmental stages in the hard-fleshed ('Qiu Jinxing') hawthorn cultivar

KEGG ID	Description	Gene Ratio	Background Ratio	P value
ko00194	Photosynthesis proteins	55/2415	122/20044	9.05E-20
ko00195	Photosynthesis	36/2415	90/20044	1.49E-11
ko00196	Photosynthesis - antenna proteins	19/2415	30/20044	4.7E-11
ko00199	Cytochrome P450	83/2415	364/20044	5.24E-09
ko04075	Plant hormone signal transduction	98/2415	482/20044	1.08E-07
ko99981	Carbohydrate metabolism	14/2415	25/20044	1.59E-07
ko02024	Quorum sensing	29/2415	99/20044	3.34E-06
ko00980	Metabolism of xenobiotics by cytochrome P450	41/2415	167/20044	5.62E-06
ko00591	Linoleic acid metabolism	24/2415	81/20044	1.84E-05
ko00830	Retinol metabolism	34/2415	135/20044	1.93E-05
ko00051	Fructose and mannose metabolism	32/2415	134/20044	0.000101
ko00140	Steroid hormone biosynthesis	31/2415	129/20044	0.000114
ko00982	Drug metabolism - cytochrome P450	28/2415	117/20044	0.000257
ko04913	Ovarian steroidogenesis	25/2415	102/20044	0.000364
ko00073	Cutin, suberine and wax biosynthesis	17/2415	60/20044	0.000528
ko04927	Cortisol synthesis and secretion	15/2415	51/20044	0.000719
ko00592	alpha-Linolenic acid metabolism	31/2415	143/20044	0.000779
ko00052	Galactose metabolism	26/2415	116/20044	0.001195
ko04977	Vitamin digestion and absorption	10/2415	30/20044	0.001937
ko00940	Phenylpropanoid biosynthesis	60/2415	350/20044	0.003003
ko00500	Starch and sucrose metabolism	62/2415	365/20044	0.003215
ko04976	Bile secretion	22/2415	100/20044	0.003496
ko00380	Tryptophan metabolism	37/2415	196/20044	0.003615
ko04712	Circadian rhythm - plant	22/2415	102/20044	0.004502
ko00680	Methane metabolism	23/2415	109/20044	0.004964
ko00941	Flavonoid biosynthesis	25/2415	123/20044	0.005821
ko04626	Plant-pathogen interaction	74/2415	461/20044	0.006006
ko04920	Adipocytokine signaling pathway	19/2415	86/20044	0.006068
ko04217	Necroptosis	29/2415	152/20044	0.007939
ko04621	NOD-like receptor signaling pathway	72/2415	452/20044	0.007966
ko00600	Sphingolipid metabolism	15/2415	64/20044	0.007972
ko00561	Glycerolipid metabolism	32/2415	172/20044	0.008036
ko00010	Glycolysis / Gluconeogenesis	45/2415	263/20044	0.009512
ko00053	Ascorbate and aldarate metabolism	27/2415	141/20044	0.00968
ko04612	Antigen processing and presentation	26/2415	135/20044	0.010199
ko00565	Ether lipid metabolism	13/2415	54/20044	0.010362
ko04917	Prolactin signaling pathway	17/2415	79/20044	0.011884

ko00860	Porphyrin and chlorophyll metabolism	24/2415	125/20044	0.013687
ko00232	Caffeine metabolism	5/2415	13/20044	0.014131
ko04978	Mineral absorption	11/2415	45/20044	0.015743
ko00910	Nitrogen metabolism	14/2415	63/20044	0.016195
ko00760	Nicotinate and nicotinamide metabolism	10/2415	41/20044	0.021131
ko04213	Longevity regulating pathway - multiple species	24/2415	130/20044	0.02149
ko02010	ABC transporters	19/2415	98/20044	0.02383
ko00983	Drug metabolism - other enzymes	26/2415	146/20044	0.026285
ko00071	Fatty acid degradation	20/2415	106/20044	0.027264
ko00350	Tyrosine metabolism	19/2415	100/20044	0.028889
ko04146	Peroxisome	31/2415	183/20044	0.031184
ko04922	Glucagon signaling pathway	31/2415	183/20044	0.031184
ko00564	Glycerophospholipid metabolism	33/2415	198/20044	0.033103
ko00790	Folate biosynthesis	13/2415	63/20044	0.03535
ko00960	Tropane, piperidine and pyridine alkaloid biosynthesis	10/2415	45/20044	0.038671
ko00904	Diterpenoid biosynthesis	17/2415	90/20044	0.03919
ko04216	Ferroptosis	12/2415	58/20044	0.041386
ko04068	FoxO signaling pathway	28/2415	167/20044	0.04369
ko00534	Glycosaminoglycan biosynthesis - heparan sulfate / heparin	5/2415	17/20044	0.045189

Table S11. The enriched KEGG functional categories of significantly (P value < 0.05) differentially expressed genes between two fruit developmental stages in the soft-fleshed ('Ruanrou Shanlihong #3') hawthorn cultivar

KEGG ID	Description	Gene Ratio	Background Ratio	P value
ko00194	Photosynthesis proteins	77/1889	122/20044	2.08E-48
ko00195	Photosynthesis	52/1889	90/20044	2.34E-30
ko00592	alpha-Linolenic acid metabolism	43/1889	143/20044	2.72E-12
ko04075	Plant hormone signal transduction	86/1889	482/20044	4.69E-09
ko99981	Carbohydrate metabolism	14/1889	25/20044	6.82E-09
ko00940	Phenylpropanoid biosynthesis	66/1889	350/20044	3.36E-08
ko00710	Carbon fixation in photosynthetic organisms	33/1889	134/20044	2.00E-07
ko00062	Fatty acid elongation	19/1889	64/20044	4.07E-06
ko00941	Flavonoid biosynthesis	28/1889	123/20044	8.40E-06
ko00330	Arginine and proline metabolism	29/1889	138/20044	2.95E-05
ko00902	Monoterpenoid biosynthesis	19/1889	74/20044	4.00E-05
ko00591	Linoleic acid metabolism	20/1889	81/20044	4.65E-05
ko04626	Plant-pathogen interaction	69/1889	461/20044	7.70E-05
ko00199	Cytochrome P450	57/1889	364/20044	9.28E-05
ko00910	Nitrogen metabolism	16/1889	63/20044	0.000183
ko04727	GABAergic synapse	19/1889	83/20044	0.00021
ko00052	Galactose metabolism	23/1889	116/20044	0.000459
ko04040	Ion channels	31/1889	183/20044	0.000949
ko00561	Glycerolipid metabolism	29/1889	172/20044	0.001465
ko00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	15/1889	71/20044	0.002223
ko00380	Tryptophan metabolism	31/1889	196/20044	0.002896
ko00053	Ascorbate and aldarate metabolism	24/1889	141/20044	0.003159
ko04020	Calcium signaling pathway	20/1889	112/20044	0.003836
ko00905	Brassinosteroid biosynthesis	10/1889	42/20044	0.004778
ko00450	Selenocompound metabolism	9/1889	36/20044	0.005176
ko04724	Glutamatergic synapse	14/1889	71/20044	0.005899
ko00944	Flavone and flavonol biosynthesis	8/1889	31/20044	0.006713
ko04976	Bile secretion	17/1889	100/20044	0.011894
ko04014	Ras signaling pathway	32/1889	225/20044	0.012249
ko00906	Carotenoid biosynthesis	13/1889	70/20044	0.012891
ko04621	NOD-like receptor signaling pathway	57/1889	452/20044	0.014307
ko00909	Sesquiterpenoid and triterpenoid biosynthesis	13/1889	73/20044	0.018026
ko04977	Vitamin digestion and absorption	7/1889	30/20044	0.01917
ko00564	Glycerophospholipid metabolism	28/1889	198/20044	0.019503
ko04975	Fat digestion and absorption	6/1889	24/20044	0.021196

ko00500	Starch and sucrose metabolism	46/1889	365/20044	0.025842
ko00051	Fructose and mannose metabolism	20/1889	134/20044	0.026184
ko00480	Glutathione metabolism	25/1889	180/20044	0.031761
ko00600	Sphingolipid metabolism	11/1889	64/20044	0.03566
ko00900	Terpenoid backbone biosynthesis	15/1889	97/20044	0.037716
ko04962	Vasopressin-regulated water reabsorption	12/1889	73/20044	0.039462
ko04016	MAPK signaling pathway - plant	39/1889	311/20044	0.040172
ko00402	Benzoxazinoid biosynthesis	5/1889	21/20044	0.0419
ko04010	MAPK signaling pathway	56/1889	472/20044	0.042832

Table S12. Statistics of repeat sequences, including transposable elements (TEs) in hawthorn, loquat, apple and pear genomes

Species	Genome Size (bp)	non-TE (bp)	non-specific LTR (bp)	non-specific non-LTR (bp)	specific LTR (bp)	specific non-LTR (bp)	Repeats proportion (%)
Hawthorn	823484868	308930973	156422968	80703069	269891955	7535903	62.49%
Loquat	760132407	286149766	99452657	68631146	291383321	14515517	62.36%
Apple	660495220	268895836	119111701	54387018	210598696	7501969	59.43%
Pear	498265991	253970737	63615352	40888819	127791564	11999519	49.17%

		Hawthorn			Loquat			Apple			Pear		
	Class	Count	bpMasked	%masked	Count	bpMasked	%masked	Count	bpMasked	%masked	Count	bpMasked	%masked
LTR	Copia	151434	98627890	11.98%	157444	97469795	12.82%	138528	80745570	12.25%	92291	47736157	9.61%
	Gypsy	191703	148373234	18.02%	177731	160297052	21.09%	146937	124001980	18.82%	82796	69341782	13.96%
	unknown	399570	179313799	21.78%	271626	133069131	17.51%	306153	124962847	18.97%	184264	74328977	14.96%
TIR	CACTA	3626	2420629	0.29%	4571	3328435	0.44%	3167	2196326	0.33%	1309	1045143	0.21%
	Mutator	5925	2443200	0.30%	7655	2848913	0.37%	6723	2591005	0.39%	490	451229	0.09%
	PIF_Harbinger	4283	2630029	0.32%	8892	4555419	0.60%	7985	3632512	0.55%	3100	2192851	0.44%
	Tc1_Mariner	128	81904	0.01%	88	79522	0.01%	140	95494	0.01%	--	--	--
	hAT	6814	3889050	0.47%	4347	2501422	0.33%	5066	3297032	0.50%	870	611665	0.12%
	polinton	810	194807	0.02%	107	21544	0.00%	--	--	--	44	16195	0.00%
nonLTR	LINE_element	6791	4845944	0.59%	5684	2775051	0.37%	4013	2398466	0.36%	5199	2322233	0.47%
	unknown	257	140999	0.02%	282	230494	0.03%	837	735624	0.11%	501	357450	0.07%
nonTIR	helitron	2826	1823740	0.22%	2044	1495620	0.20%	1706	1280734	0.19%	1977	1606708	0.32%
repeat_region		276339	69768670	8.47%	246451	65310243	8.59%	191848	45661794	6.93%	188964	44284864	8.91%
Total		1050506	514553895	62.49%	886922	473982641	62.36%	813103	391599384	59.43%	561805	244295254	49.17%

Table S13. Statistics of orthologs between hawthorn, apple and loquat genomes

	Chromosomes Sub-genome A	in Genes Num.	Chromosomes Sub-genome B	in Genes Num.
Loquat (Ej)	1	3088	2	3069
	11	2431	13	2479
	3	2936	4	2584
	6	3170	5	2669
	7	2479	8	2488
	9	2421	14	2057
	10	2532	15	2102
	12	2300	16	2105
			17	2198
Hawthorn (Cp)	2	2689	3	2763
	5	2269	11	2137
	13	2208	8	2228
	1	2552	7	2081
	4	2826	9	2312
	6	2207	10	1899
	12	2286	14	1820
	15	2015	16	1825
		17	2014	
Apple (Md)	11	2867	3	2571
	5	3191	10	3175
	9	2522	17	2591
	16	2703	13	2387
	4	2117	6	2067
	14	2202	12	2284
	7	2654	1	2114
	15	3874	2	2620
		8	2117	

The color of the table corresponds to the squares of chromosomes in Figure 4E in paper.