

Article

Characterization of the MADS-Box Gene Family in *Akebia trifoliata* and Their Evolutionary Events in Angiosperms

Shengfu Zhong^{1,†}, Huai Yang^{1,†}, Ju Guan¹, Jinliang Shen², Tianheng Ren¹ , Zhi Li¹, Feiquan Tan¹, Qing Li³ and Peigao Luo^{1,*} 

¹ Key Laboratory of Plant Genetics and Breeding at Sichuan Agricultural University of Sichuan Province, College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China

² College of Forestry, Sichuan Agricultural University, Chengdu 611130, China

³ Department of Biology and Chemistry, Chongqing Industry and Trade Polytechnic, Chongqing 408000, China

* Correspondence: lpglab@sicau.edu.cn

† These authors contributed equally to this work.

Abstract: As the largest clade of modern plants, flower plants have evolved a wide variety of flowers and fruits. MADS-box genes play key roles in regulating plant morphogenesis, while basal eudicots have an evolutionarily important position of acting as an evolutionary bridge between basal angiosperms and core eudicots. *Akebia trifoliata* is an important member of the basal eudicot group. To study the early evolution of angiosperms, we identified and characterized the MADS-Box gene family on the whole-genome level of *A. trifoliata*. There were 47 MADS-box genes (13 type I and 34 type II genes) in the *A. trifoliata* genome; type I genes had a greater gene length and coefficient of variation and a smaller exon number than type II genes. A total of 27 (57.4%) experienced whole or segmental genome duplication and purifying selection. A transcriptome analysis suggested that three and eight genes were involved in whole fruit and seed development, respectively. The diversification and phylogenetic analysis of 1479 type II MADS-box genes of 22 angiosperm species provided some clues indicating that a γ whole genome triplication event of eudicots possibility experienced a two-step process. These results are valuable for improving *A. trifoliata* fruit traits and theoretically elucidating evolutionary processes of angiosperms, especially eudicots.

Keywords: *Akebia trifoliata*; MADS-box; whole genome duplication; seed development; angiosperm



Citation: Zhong, S.; Yang, H.; Guan, J.; Shen, J.; Ren, T.; Li, Z.; Tan, F.; Li, Q.; Luo, P. Characterization of the MADS-Box Gene Family in *Akebia trifoliata* and Their Evolutionary Events in Angiosperms. *Genes* **2022**, *13*, 1777. <https://doi.org/10.3390/genes13101777>

Academic Editor: Zhiqiang Wu

Received: 23 August 2022

Accepted: 28 September 2022

Published: 1 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Throughout history, plants, especially flowering plants, have evolved ideal molecular mechanisms for morphogenesis, primarily driven by the differential growth of various tissues [1], which is regulated by intertwining network *cis*-element and *trans*-acting factors also called transcription factors [2]. Transcription factors encoded by the MADS-box gene families play fundamental roles in organogenesis control and signal transduction during the tissue growth process and have therefore been extensively studied [3]. In addition, MADS-box genes are also widely used in gene or genome duplication analyses for plant species evolution because they are highly conserved [4]. Thus, further identification of MADS-box gene families from recently published plant genomes is important work.

The MADS acronym is derived from the initial names of the first four proteins with a MADS-box to be reported: MCM1 from *Saccharomyces cerevisiae* [5], AGAMOUS from *Arabidopsis thaliana* [6], DEFICIENS from *Antirrhinum majus* [7] and SRE from *Homo sapiens* [8], which shared the highly conserved 180-bp-long motif called the MADS-box, suggesting that MADS-box genes widely exist in fungi, animals and plants [9]. The conserved MADS domain with approximately 60 amino acid residues at the N-terminus of all MADS-box proteins can bind to the CA_nG box (CC-A-rich-GC) in the promoter of their target genes [10]. Based on the sequence structure, there are two evolutionary clades of MADS-box genes: type I, including three subgroups (M α , M β and M γ), and type II, including

two subgroups (MIKC* and MIKC^C). MIKC^C can be further divided into the following groups: *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1/Tomato MADS-box gene 3* (SOC1/TM3), *Tomato MADS-box gene 8* (TM8), *APETALA3/DEFICIENS* (AP3/DEF), *PISTILLATA* (PI), *ARABIDOPSIS BSISTER* (ABS, GMM13), *AGAMOUS* (AG), *AGAMOUS-LIKE12* (AGL12), *SEPALLATA* (SEP), *SQUAMOSA* (SQUA), *ARABIDOPSIS NITRATE REGULATED 1* (ANR1), *AGAMOUS-LIKE15* (AGL15), *SHORT VEGETATIVE PHASE* (SVP), *Oryza sativa MADS-box gene 32* (OsMADS32), *AGAMOUS-LIKE6* (AGL6) and *FLOWERING LOCUS C* (FLC) [11]. Although this subfamily has been extensively researched, it is still unclear how many MIKC^C groups evolved in plants. Previous studies showed that type II genes experienced a lower rate of birth-and-death evolution than type I genes [4]; according to the ABC model, the identities of floral organs are determined by several subgroups of type II genes, and this was even extended to the ABCDE model [12,13]. Therefore, type II genes have a greater significance than type I genes in the species diversification of angiosperms [14].

Angiosperms consist of approximately 300,000 extant species, more than all other groups of land plants combined [15] and include three major evolutionary branches: basal angiosperms such as *Amborella trichopoda*, monocots such as *O. sativa* and eudicots such as *A. thaliana*. However, eudicots can be further classified into basal eudicots (also called early diverging eudicots or sister groups of core eudicots), such as *Papaver somniferum*, and core eudicots, such as *Vitis vinifera*. The basal eudicots, as the early lineage of eudicots, usually have a parallel evolutionary relative with monocots and are indeed counterparts of monocots. In addition, basal eudicots are an evolutionary bridge between basal angiosperms and core eudicots. This indicates that basal eudicots could have an important role in evolutionary studies.

A. trifoliata (Thunb.) Koidz, which is also commonly called augmelon (August melon) in various counties of China because the pericarp rapidly generates a unilateral crevice alongside the ventral suture when its fruit matures [16], belongs to the Lardizabalaceae family of basal eudicots and is a perennial climbing woody liana plant that is widely distributed in East Asia [17]. Practically, it has been widely used as a popular medicinal plant for at least 2000 years [18], and it also has great potential to be exploited as an economic crop, with uses as an edible fruit crop [19], oil crop [20] and ornamental plant [21]; thus, it has recently attracted increasing attention from commercial farmers. Theoretically, *A. trifoliata* is an important member of the basal eudicot group, so it is also of great value in studying the early evolution of angiosperms [22]. Therefore, characterizing the profile of MADS-box gene families and determining the evolutionary relationships among various species will aid in the rapid genetic improvement of economic traits and the realization of organogenesis or tissue formation, such as seed formation.

Four type II MADS-box genes (*AktAP3_1*, *AktAP3_2*, *AktAP3_3* and *AktPI*) were first isolated from *A. trifoliata*, and three *AktAP3* genes could have been putatively produced by two gene duplication events before the origin of the genus *Akebia* [23]. Then, a total of 10 type II MADS-box candidate genes were further identified from this species [22] and all of them were highly conserved and could have been saved by purifying selection during the process of early angiosperm diversification [24]. Although there are some shortages of these studies partly due to the unavailability of genomic data at that time from a systematic and integral standpoint, they afforded much useful information and a reference strategy to further comprehensively understand the characteristics and evolutionary events of MADS-box genes.

In the present study, the available high-quality assembled genome of *A. trifoliata* was employed to accomplish the following objectives: to systematically identify the candidate members of MADS-box gene families, to physically map them on chromosomal positions and to understand their profiles, such as their type, length, number of exons, selection style and putative function. In addition, we would also like to elucidate the evolutionary relatives of MADS-box gene families in whole angiosperms. Ultimately, our findings provide important information on the MADS-box families to further elucidate their functions in

A. trifoliata as well as in other angiosperms, especially basal eudicots, indirectly contributing to the genetic improvement of commercial traits such as seed yield.

2. Materials and Methods

2.1. Identification of MADS-Box Genes in *A. trifoliata*

Genome sequence and annotations of *A. trifoliata* downloaded from the Genome Warehouse of the National Genomics Data Center under the accession number GWH-BISH00000000 (<https://ngdc.cncb.ac.cn/gwh> (accessed on 8 March 2022)) were employed to identify the MADS-box genes. Despite the existence of recent publications on the *A. trifoliata* genome [25], the corresponding genomic data files are still unavailable online. Therefore, a high quality *A. trifoliata* genome assembled and submitted by our group to NCBI was employed for the identification and characterization in this study. To identify MADS-box genes, we used the MADS-box typical domain SRF-TF (PF00319) from the Pfam database to search for the predicted genes using the *hmmsearch* tool in HMMER software (v3.3.2, Sara El-Gebali, UK) with the '10e-5' e value parameter [26]. Then, the candidate genes with incomplete domains or gene structure were manually removed. To further confirm the MADS-box feature, the conserved domains and motifs were annotated using the NCBI CDD (v3.19, Aron Marchler-Bauer, USA) and MEME Suite tools (v5.3.2, Timothy L. Bailey, Australia), respectively [27,28]. The MADS-box subfamilies were classified according to sequence similarity and the phylogenetic tree branches of reported MADS-box subfamilies from the reference genome of *A. thaliana* [29]. Another phylogenetic tree was constructed based only on *A. trifoliata* MADS-box genes to further validate the classification results. These two phylogenetic trees were both constructed by using IQtree (v2.2.0.3, Lam-Tung Nguyen, Austria) software with the maximum likelihood method, automatically selecting the optimal substitution model and evaluating branch support values via UFBoot2 tests [30–32]. Multiple sequence alignment analysis of MADS-box domains was aligned by ClustalW (v2.0.10, Julie D. Thompson, UK) with default parameters [33]. Previously reported *A. trifoliata* MADS-box genes from the GenBank database and references [22,23] were anchored to our homologues using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi> (accessed on 26 March 2022)) with *blastn* mode.

2.2. Chromosomal Distribution of MADS-Box Genes

The *A. trifoliata* genome was previously assembled into 16 chromosomes based on the assembled genome data. The chromosomal positions of the MADS-box genes were assigned using the GFF3 annotation of genome data. The potential cluster of MADS-box genes was identified by sliding window analysis assuming a window size of 250 kb [34]. Then, TBtools (v1.098769, Chengjie Chen, China) was applied to visualize the positions of genes on chromosomes [35].

2.3. Synteny and Duplication Analysis of MADS-Box Genes

Both paralogous gene pairs and syntenic blocks of MADS-box genes in the *A. trifoliata* genome were identified using TBtools with default parameters of e-value 1×10^{-10} and BlastP hits number 5 for a gene based on the MCScanX tool [36]. A collinear block was defined as at least 5 paralogous gene pairs that were significantly aligned (1×10^{-5} as default). Then, the duplication types containing singleton, dispersed, proximal, tandem and WGD/segmental of the MADS-box genes were determined according to the MCScanX synteny analysis results. The nonsynonymous substitution rate (Ka) and synonymous substitution rate (Ks) of MADS-box duplicated gene pairs were calculated using TBtools with the NG method.

2.4. Construction of A Phylogenetic Tree of Type II MADS-Box Genes from 22 Species

A total of 22 species genomes (*A. trichopoda*, *N. colorata*, *O. sativa*, *Sorghum bicolor*, *Apostasia shenzhenica*, *V. vinifera*, *A. majus*, *Helianthus annuus*, *A. thaliana*, *Nelumbo nucifera*, *P. somniferum*, *Macleaya cordata*, *A. coerulea*, *Beta vulgaris*, *B. distachyon*, *Populus trichocarpa*, *Theo-*

broma cacao, *P. patens*, *Selaginella moellendorffii*, *Spirodela polyrhiza*, *Zostera marina* and *A. trifoliata*) were employed to identify MADS-box genes, in which *A. thaliana* [37], *O. sativa* [38] and *A. trichopoda* [39]) were applied as references to identify the subfamilies of the MADS-box genes. The same identified methods as for *A. trifoliata* were used to detect the MADS-box genes for 22 plant species. After classifying the MADS-box subfamilies, the phylogenetic tree of every type II subfamily was constructed and modified using IQtree and FigTree v1.4.4 (<https://github.com/4ambaut/figtree> (accessed on 21 June 2022)) [30]. The parameter settings were the same as those listed above for the phylogenetic tree of *A. trifoliata* and *A. thaliana* to detect potential duplication events. Gene trees were rooted with genes from mosses or basal angiosperms and cladogram transformation of the branches was applied.

2.5. Expression Analysis of *A. trifoliata* MADS-Box Genes

The 271.49 Gb Illumina transcriptome data of *A. trifoliata* consisted of 36 samples of three tissues (peel, flesh and seed) at four different stages (young, enlargement, colouring and maturity) with three biological replicates, respectively. Details of the data are present in NCBI BioProject PRJNA671772 (accession IDs: SAMN16551931-33, young stage of rind, flesh and seed; SAMN16551934-36, enlargement stage; SAMN16551937-39, coloring stage; SAMN16551940-42, mature stage). The sequences of these samples were aligned to the *A. trifoliata* reference genome using Hisat2 (v2.1.0, Mihaela Pertea, USA) software with default parameters [40]. Then, stringtie [41] was used to evaluate the abundance count of each gene in 36 samples. The gene expression levels in each sample were estimated using fragments per kilobase of transcript per million fragments mapped (FPKM) values according to the abundance count data by using R package DESeq2 (v1.36.0, Michael I Love, Germany) [42]. For differential gene expression, 12 groups of 4 stages and 3 tissues with 3 replicates were calculated as the count matrix. The means of each gene expression in three replicates were normalized to $\log_2(\text{FPKM} + 1)$ for visual display. After identifying the MADS-box genes from the whole genome, the expression clusters of these genes were plotted using the R package ggplot2 (v3.3) (<https://ggplot2.tidyverse.org> (accessed on 28 June 2022)) [43].

3. Results

3.1. The Number and Type of Identified MADS-Box Genes in the *A. trifoliata* Genome

A total of 47 MADS-box genes were identified from the *A. trifoliata* genome through HMM analysis (Table 1), in which there was a broad range in the gene length (from 360 bp to 80,167 bp), exon number (from one to 19), isoelectric point (from 4.67 to 11.42), amino acid number (from 81 to 545) and molecular weight (from 9.1 to 62.53) of the putative proteins encoded by the MADS-box genes. The phylogenetic tree consisting of both 108 reference MADS-box genes (consisting of both three and 14 subfamilies of type I and type II, respectively) in *A. thaliana* and 47 MADS-box genes identified in *A. trifoliata* was constructed to classify the MADS-box types in *A. trifoliata*. The phylogenetic tree (Figure 1) showed that 47 MADS-box genes of *A. trifoliata* were unevenly classified into all three subfamilies of type I and 13 subfamilies of type II reference MADS-box genes of *A. thaliana* and only FLC subfamilies of type II in *A. thaliana* were not detected in *A. trifoliata* (Figure 1).

The type I group consisted of 13 genes and the type II group consisted of the remaining 34 genes (Table 1). A comparison analysis found that type I genes have a significantly smaller length and fewer exons than type II genes (Table S1) at the $p = 0.01$ level, while the differences in isoelectric point, number of amino acids and molecular weight of the putative proteins between them were not significant at the corresponding statistical test level. For example, the largest exon number of type I was only 3 for *AktM γ _3*, while that of type II was up to 19 for *AktSOC1/TM3_2*.

In addition, two reference genes (*AT5G55690* and *AT5G58890* in *A. thaliana*) were classified into My subfamilies rather than the previously classified M β subfamilies [29], and three *A. trifoliata* MADS-box genes, *EVM0009117*, *EVM0013722* and *EVM0016918* were the paraphyletic group and classified into the same clade. A comparison analysis of the MADS-

box domain of the putative proteins encoded by 44 M β and M γ genes (6 from *A. trifoliata* and 38 from *A. thaliana*) also found that *AT5G55690*, *AT5G58890*, *EVM0009117*, *EVM0013722* and *EVM0016918* seemed closer to the M γ clade than to the M β clade (Figure S1). Conservatively, we classified the three genes into the M β /M γ clade (Table 1).

Table 1. Characteristics of the identified MADS-box gene families from the *A. trifoliata* genome.

Gene_Name	MADS-Box Type	Modified Name	Chromosome	Start Locus	Gene Length	Exon	AA	PI	MW (kDa)
<i>EVM0003126</i>	TypeI:M α	<i>AktMα_1</i>	chr2	9340686	618	2	192	8.23	21.38
<i>EVM0007393</i>	TypeI:M α	<i>AktMα_2</i>	chr4	13263148	627	2	197	8.65	22.4
<i>EVM0008161</i>	TypeI:M α	<i>AktMα_3</i>	chr3	40316320	705	1	234	10	26.3
<i>EVM0008221</i>	TypeI:M α	<i>AktMα_4</i>	chr4	13240370	1167	1	388	4.67	42.55
<i>EVM0009312</i>	TypeI:M α	<i>AktMα_5</i>	chr8	6769455	4439	2	204	9.2	23.55
<i>EVM0019163</i>	TypeI:M α	<i>AktMα_6</i>	chr4	13227911	6676	2	182	7.58	20.71
<i>EVM0019453</i>	TypeI:M α	<i>AktMα_7</i>	chr8	45540661	420	1	139	9.6	15.98
<i>EVM0009117</i>	TypeI:M β /M γ	<i>AktMβ_1</i>	chr13	27782386	3336	2	345	4.84	38.81
<i>EVM0013722</i>	TypeI:M β /M γ	<i>AktMβ_2</i>	chr16	31917540	981	1	326	5.3	37.12
<i>EVM0016918</i>	TypeI:M β /M γ	<i>AktMβ_3</i>	chr15	901697	951	1	316	8.29	36.85
<i>EVM0007317</i>	TypeI:M γ	<i>AktMγ_1</i>	chr8	15288657	782	1	251	6.8	28.49
<i>EVM0011701</i>	TypeI:M γ	<i>AktMγ_2</i>	chr4	4837008	663	1	220	5.52	25.59
<i>EVM0019125</i>	TypeI:M γ	<i>AktMγ_3</i>	chr16	32364845	3328	3	274	9.54	31.63
<i>EVM0002119</i>	TypeII:MIKC*	<i>AktMIKC*_1</i>	chr15	11731497	2402	3	87	10.41	9.64
<i>EVM0008992</i>	TypeII:MIKC*	<i>AktMIKC*_2</i>	Contig00271	36366	1726	2	82	10.81	9.1
<i>EVM0014464</i>	TypeII:MIKC*	<i>AktMIKC*_3</i>	chr2	7636528	6189	13	420	4.7	47.72
<i>EVM0015933</i>	TypeII:MIKC*	<i>AktMIKC*_4</i>	chr3	6814225	9410	12	352	7.41	40.16
<i>EVM0001820</i>	TypeII:MIKC*	<i>AktMIKC*_5</i>	chr13	2907447	7905	11	326	7.54	36.95
<i>EVM0003153</i>	TypeII:MIKC*	<i>AktMIKC*_6</i>	chr4	1021642	360	2	81	11.42	9.39
<i>EVM0007460</i>	TypeII:MIKC*	<i>AktMIKC*_7</i>	chr4	14680429	6077	10	341	6.23	38.85
<i>EVM0011606</i>	TypeII:MIKC*	<i>AktMIKC*_8</i>	chr2	12505863	4919	10	297	6.68	34.14
<i>EVM0012489</i>	TypeII:MIKC*	<i>AktMIKC*_9</i>	chr3	42029038	16357	11	287	6.72	32.76
<i>EVM0008434</i>	TypeII:GGM13	<i>AktGGM13</i>	chr1	21901520	17282	7	239	8.46	27.92
<i>EVM0009546</i>	TypeII:AG	<i>AktAG_1</i>	chr11	49371746	8551	8	256	9.59	29.72
<i>EVM0023593</i>	TypeII:AG	<i>AktAG_2</i>	chr3	33675447	37442	7	196	9.69	22.94
<i>EVM0016249</i>	TypeII:AG	<i>AktAG_3</i>	chr11	5816813	12572	9	226	9.61	26.08
<i>EVM0004910</i>	TypeII:AG/TM8	<i>AktAG/TM8</i>	chr9	26578608	5775	8	203	10.53	23.62
<i>EVM0008270</i>	TypeII:AGL12	<i>AktAGL12</i>	chr5	25729606	7325	8	238	7.84	27.26
<i>EVM0012263</i>	TypeII:AGL15	<i>AktAGL15</i>	chr14	89447	8175	8	248	8.29	28.48
<i>EVM0004219</i>	TypeII:AGL6	<i>AktAGL6_1</i>	chr12	32312628	12346	8	243	9.16	27.9
<i>EVM0022986</i>	TypeII:AGL6	<i>AktAGL6_2</i>	chr10	27960166	13564	8	243	9.21	27.78
<i>EVM0013426</i>	TypeII:ANR1	<i>AktANR1_1</i>	chr3	39522622	25664	8	235	9.76	26.98
<i>EVM0020447</i>	TypeII:ANR1	<i>AktANR1_2</i>	chr4	390010	25694	8	240	10.38	27.67
<i>EVM0022723</i>	TypeII:AP3/DEF	<i>AktAP3/DEF_1</i>	chr2	3614814	2513	7	225	9.57	26.05
<i>EVM0011008</i>	TypeII:AP3/DEF	<i>AktAP3/DEF_2</i>	chr4	42333387	6291	7	225	7.78	25.98
<i>EVM0016520</i>	TypeII:AP3/DEF	<i>AktAP3/DEF_3</i>	chr2	3598851	2268	7	227	10.42	26.52
<i>EVM0017990</i>	TypeII:PI	<i>AktPI</i>	chr3	36571931	4002	7	212	9.48	24.81
<i>EVM0006618</i>	TypeII:SEP	<i>AktSEP_1</i>	chr9	21528809	25656	8	246	8.45	28.3
<i>EVM0009403</i>	TypeII:SEP	<i>AktSEP_2</i>	chr5	8341522	80167	5	188	7.34	21.22
<i>EVM0012649</i>	TypeII:SEP	<i>AktSEP_3</i>	chr7	24138805	37115	9	254	8.93	29.22
<i>EVM0001967</i>	TypeII:SEP	<i>AktSEP_4</i>	chr7	33179065	15121	8	246	7.24	27.96
<i>EVM0001926</i>	TypeII:SOC1/TM3	<i>AktSOC1/TM3_1</i>	chr12	32400886	20703	8	231	7.94	26.59
<i>EVM0003932</i>	TypeII:SOC1/TM3	<i>AktSOC1/TM3_2</i>	chr12	32427991	30375	19	545	9.19	62.53
<i>EVM0013498</i>	TypeII:SQUA	<i>AktSQUA_1</i>	chr7	33206678	19735	8	260	9.83	30.07
<i>EVM0021295</i>	TypeII:SQUA	<i>AktSQUA_2</i>	chr3	10011542	36166	4	126	10.92	14.97
<i>EVM0004256</i>	TypeII:SVP	<i>AktSVP_1</i>	chr4	10882672	11337	8	251	6.54	28.63
<i>EVM0015068</i>	TypeII:SVP	<i>AktSVP_2</i>	chr2	6237389	9183	9	225	6.96	25.42

Abbreviations: AA, amino acid; PI, isoelectric point; MW, molecular weight.

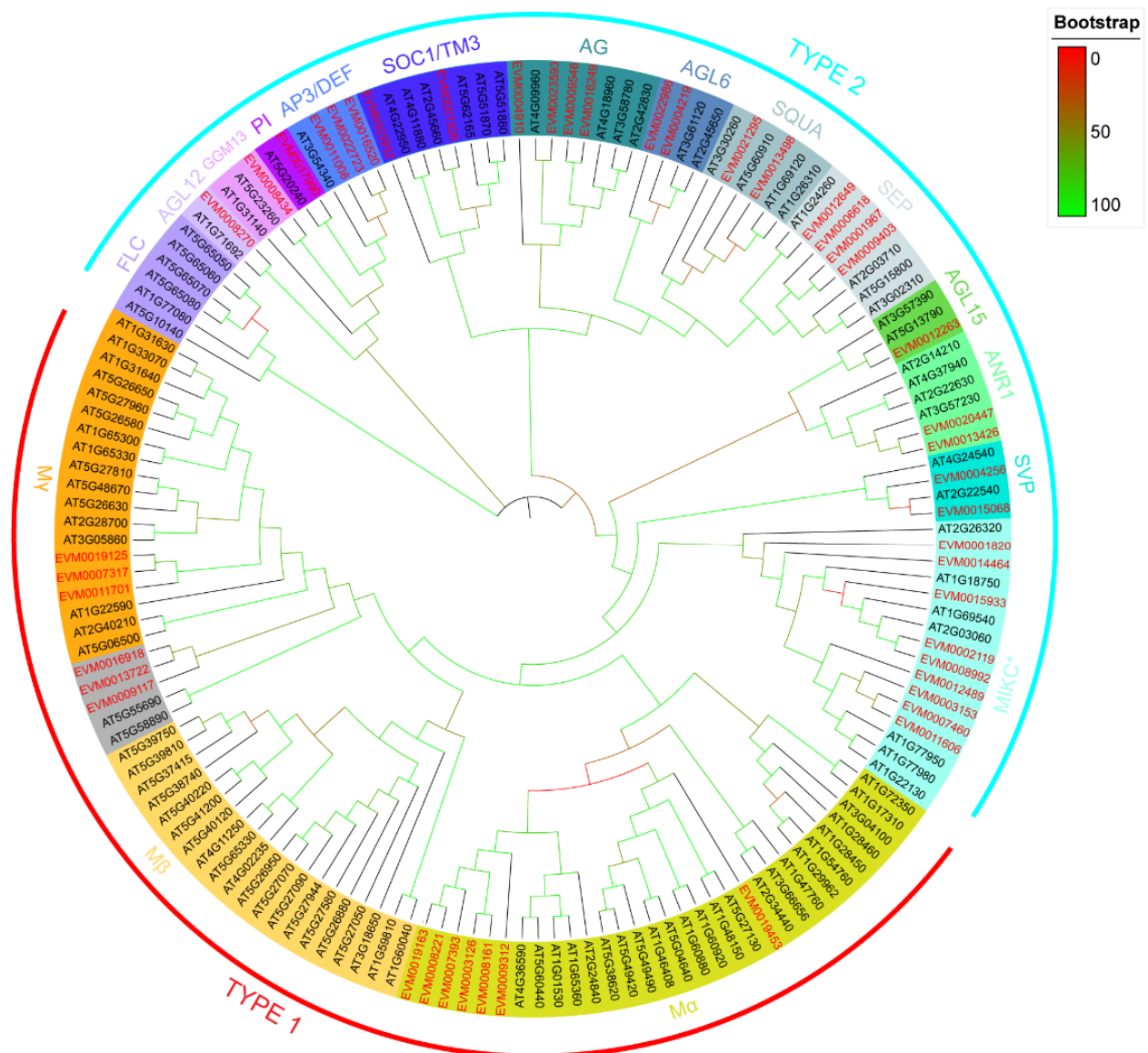


Figure 1. Classification of the *A. trifoliata* MADS-box subfamilies. Specific lineages are indicated by colours and bracketing. The type I lineages contain M α , M β , and M γ , while the type II lineages contain MIKC* and MIKC^C (all other type II subfamilies except MIKC*). The paraphyletic groups of 5 genes, including *EVM0009117*, *EVM0013722*, *EVM0016918*, *AT5G55690* and *AT5G58890*, are coloured grey because *AT5G55690* and *AT5G58890* have previously been classified as M β but are placed with M γ in our analysis.

3.2. Phylogeny and Conserved Motifs of MADS-Box Genes

Based on the *A. trifoliata* MADS-box gene phylogenetic tree (Figure 2a), 47 identified MADS-box genes in *A. trifoliata* were also classified as described above, except *EVM0004910* (*AktAG/TM8_4*). Motifs and conserved domains of typical MADS-box family members were further annotated to check the MADS domain. The results showed that 46 out of 47 identified MADS-box genes in *A. trifoliata* had the conserved motifs (motif 1) of MADS proteins, except one *Akt M β _1*. Though there was no motif annotated with the MEME Suite in *Akt M β _1*, a MADS domain could be detected with NCBI CDD (Figure 2b). In addition, all 13 type I and 9 MIKC* genes contained only the conserved MADS domain, while all 25 MIKC^C genes had both MADS and K-box domains, except for *AktSOC1/TM3_2*, which had MADS, LPLAT and EFH domains (Figure 2c). Generally, the coding DNA sequence of

type I genes with a smaller gene length had good continuity compared with that of type II genes (Figure 2d). Compared with *A. trifoliata* MADS-box genes from a previous report and the GenBank database, only a few genes (10) related to floral organ development were reported (Table S2), and most of the genes (37) identified in our study have never been deeply studied before.

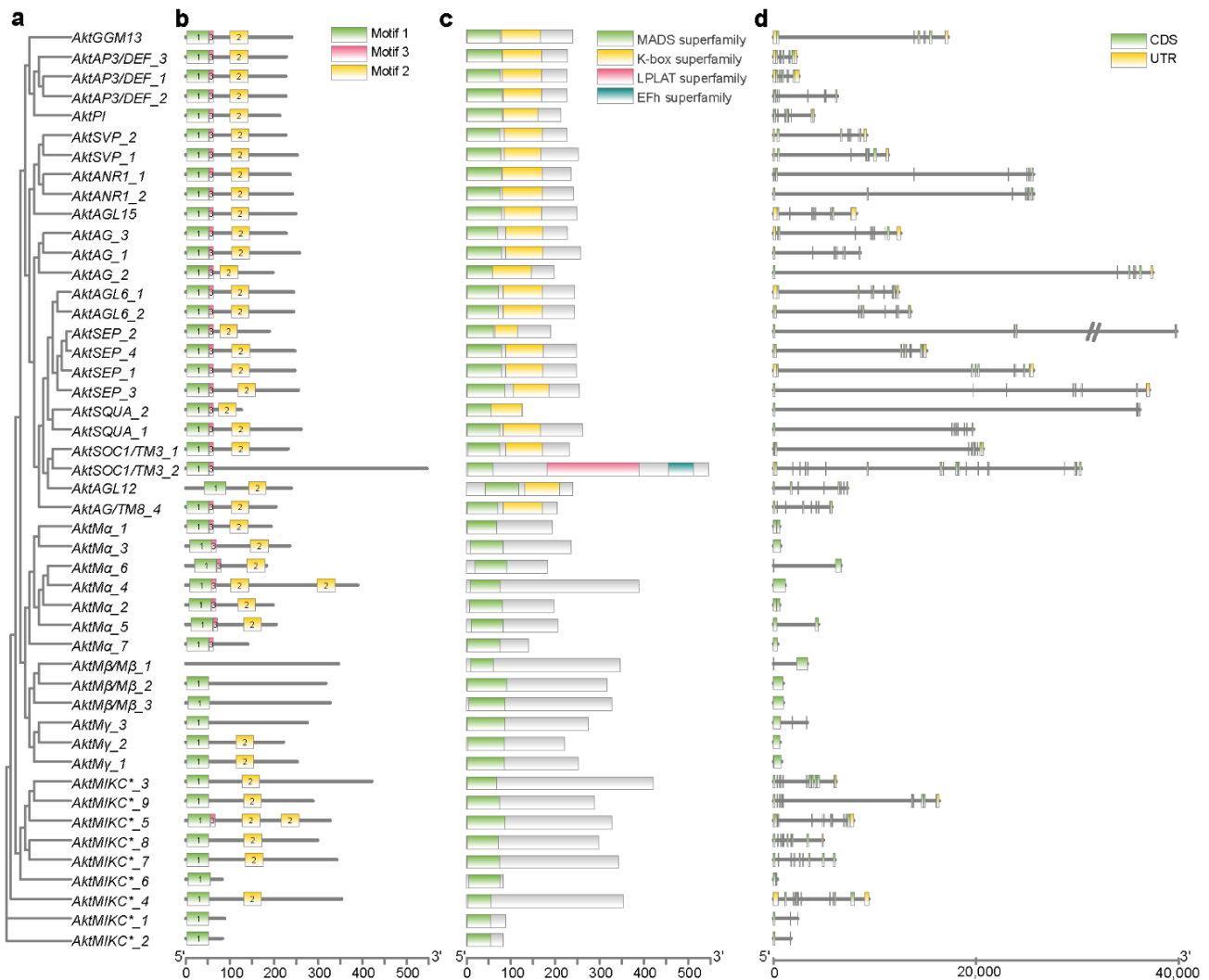


Figure 2. Gene and protein structures of the MADS-box gene families in *A. trifoliata*. (a) Phylogenetic tree of *A. trifoliata* MADS-box genes; (b) Motifs of MADS-box proteins. (c) Conserved domains of MADS-box proteins; (d) Exon-intron structures of MADS-box genes.

3.3. Chromosomal Position and Duplication Type of MADS-Box Genes

The physical position of all 47 identified MADS-box genes is ranked in Table 1, and 46 MADS-box genes were mapped to almost all 16 *A. trifoliata* chromosomes except chromosome 6 (Figure 3), while only *AktMIKC*₂* was assigned to the unassembled contig. Chromosomes 2, 3 and 4 carried 6, 7 and 9 MADS-box genes, respectively, while chromosomes 1, 10 and 14 carried only one. According to the definition of gene clusters, the genes on the chromosomes were divided into 41 loci, including 37 singletons and 4 gene clusters (Figure 3). Overall, 10 (21.3%) of the 47 MADS-box genes were distributed into 4 clusters. Most MADS-box genes were collectively distributed on chromosomes 2, 3 and 4, accounting for 22 of 47 that could represent paralogous segments resulting from ancestral polyploidization or tandem duplication events. Further synteny analysis (Figure S2) found that 27 MADS-box genes (marked in red in Figure 3) consisting of *Mα* (2), *Mβ* (1), *Mγ* (3),

*MIKC** (5), *AG* (2), *AGL6* (2), *ANR1* (2), *AP3/DEF* (2), *SEP* (2), *SOC1/TM3* (2), *SQUA* (2) and *SVP* (2) could have been putatively produced by whole genome duplication (WGD) or segmental genome duplication, while only three (*AktMα_4*, *AktMα_6* and *AktAP3/DEF_1*, marked in blue in Figure 3) could have been produced by tandem duplication.

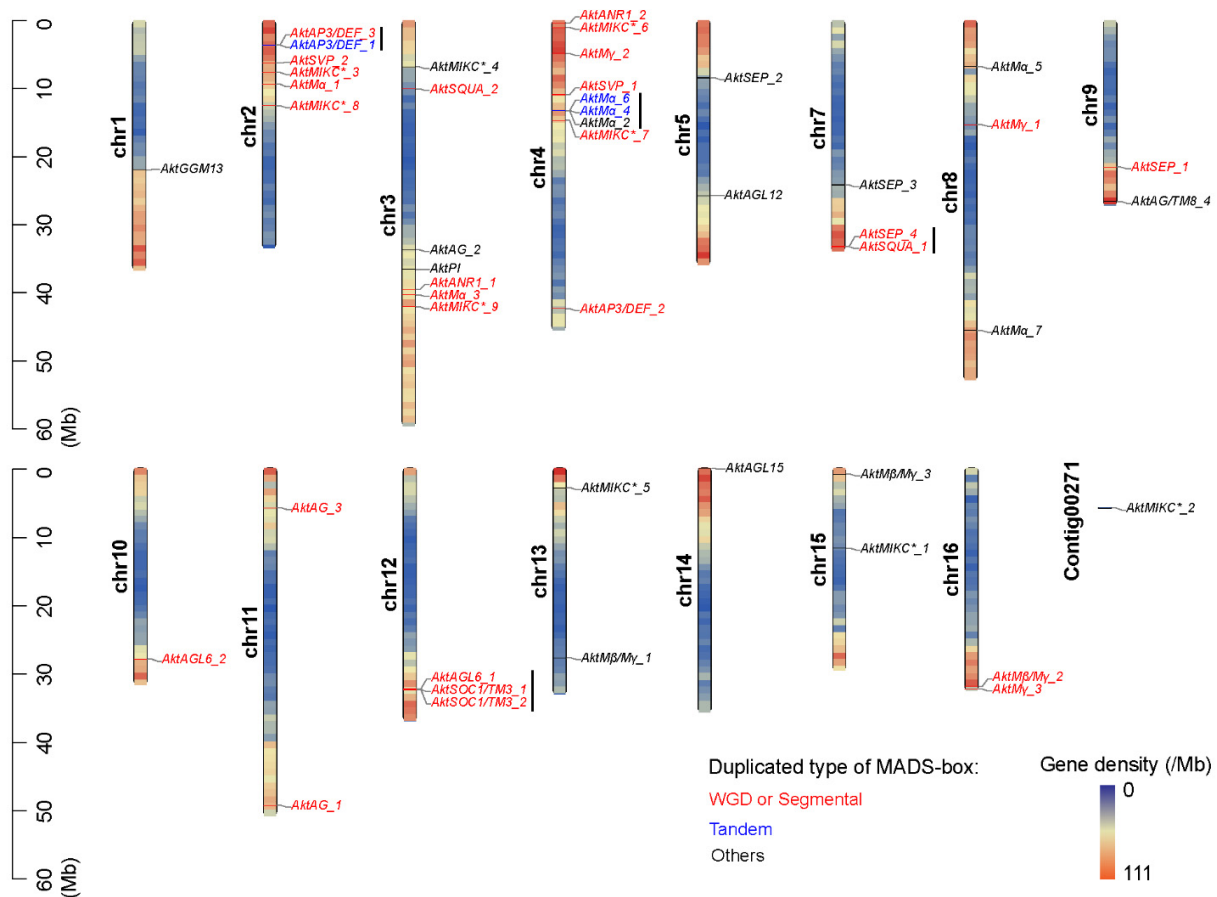


Figure 3. Physical position and duplicated type of MADS-box genes in the *A. trifoliata* genome. WGD or segmental duplicated-type genes are marked in red font, tandem duplicated-type genes are marked in blue font, and other dispersed or proximal duplicated genes are marked in black font. The black line to the right of the gene names represents cluster loci.

3.4. Estimation of K_a/K_s Value

The ratio of the nonsynonymous substitution rate (K_a) to the synonymous substitution rate (K_s) was calculated according to the paralogous pairs of 27 WGD or segmental duplication MADS-box genes, and we detected a total of 19 paralogous pairs among them. All 19 K_a/K_s values of these *A. trifoliata* MADS-box duplicated gene pairs were far less than 1 (Table S3). The largest K_a/K_s value of the paralogous MADS-box gene pair (*AktMIKC*_3* and *AktMIKC*_9*) was only 0.52, and the K_a/K_s value of *AktAG_3* and *AktAG_3* was as low as 0.11. In addition, the K_a/K_s value of the paralogous type II MADS-box gene pair was commonly lower than that of type I (Table S3).

3.5. MADS-Box Genes of Evolutionarily Important Species

The genomes of twenty two evolutionarily important species, two mosses, two basal angiosperms, six monocots, five basal eudicots and seven core eudicots (Figure 4a), were employed to identify and classify the MADS-box genes, in which three species, *A. thaliana*, *O. sativa* and *A. trichopoda*, were used as references. A total of 1469 MADS-box genes with 19 subfamilies were identified, and the average number of MADS-box genes was 67, with a range from 24 to 162 (Tables S4 and S5). There was a medium coefficient ($R = 0.52$,

$p = 0.012$) between the MADS-box gene number and genomic size among the 22 species. For type I, the number of $M\alpha$ genes was larger than that of $M\beta$ and $M\gamma$ genes in 17 out of 22 species, and for type II, there were more $MIKC^C$ genes than $MIKC^*$ genes in all 22 species except *Physcomitrella patens* (Table S5, Figure 4b). Notably, *EVM0004910* (*AktAG/TM8_4*) of *A. trifoliata* was classified into *TM8* (Table S5) rather than the *AG* clade (Figure 1). In addition, there was no $M\gamma$ in either moss, no *FLC* in all 22 species except for seven core eudicots and no *OsMADS32* clade in 12 eudicots (Table S5). Both mosses only contained the *GGM13* of $MIKC^C$ and lacked the remaining 14 subfamilies (Table S5). We also found that the coefficient of variation (CV) of type I genes was larger than that of type II genes (Figure 4b).

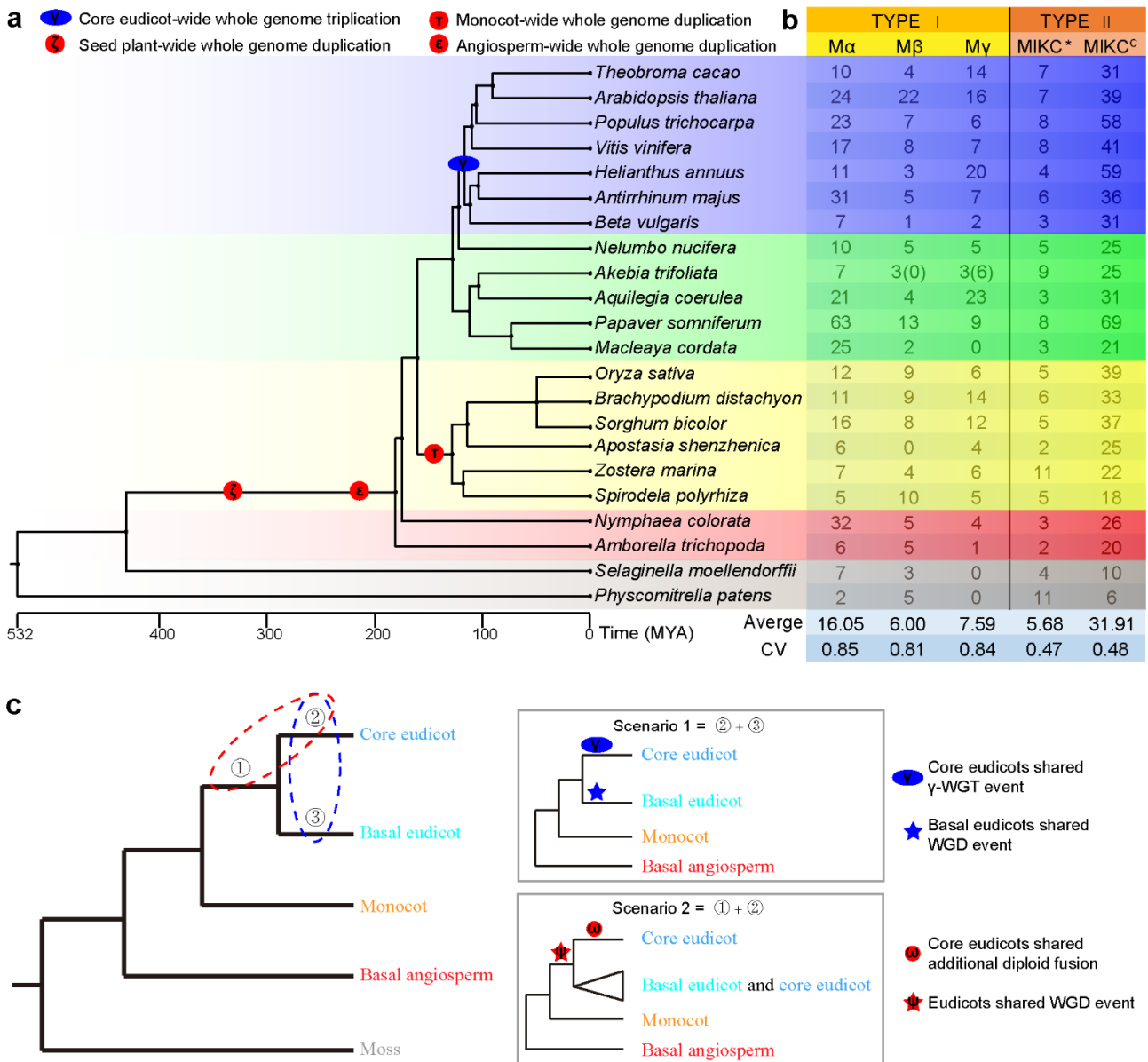


Figure 4. Diversification of MADS-box genes in land plants. (a) Phylogeny of 22 plant species. The topology of the tree was referenced from the TimeTree database, and four ancestral polyploidy events in seed plants were marked according to a previous report [44]. (b) Total number and classification of MADS-box genes in 22 plant species. CV: coefficient of variation. (c) Hypothetical duplicate genes of MADS-box origin from ancestral polyploidization events.

3.6. Possible Duplication Events on Different Divergent Angiosperm Branches

Plant MADS-box genes, especially type II MADS-box genes, are widely used in gene or genome duplication analyses; in particular, most duplicated MADS-box genes in *A. trifoliata* were derived from a WGD event (Figures 3 and S2). A total of 15 phylogenetic trees of type II MADS-box gene subfamilies in 22 angiosperm species were constructed, and only the phylogenetic tree of *OsMADS32* was not produced because of the lower number of these genes (Figure S3). Although the trace of these gene duplications could be lost in evolutionary processes, we still detected at least one potential ancestral duplication event in the phylogenetic trees of 15 subfamilies except *TM8* and *AGL6* (Figure S3 and Table S5), in which those of *MIKC**, *GGM13*, *ANR1*, *SEP* and *AG* were probably duplicated from ancestral angiosperms or seed plant-wide WGDs (ζ or ϵ) due to the duplications early in basal angiosperms species, those of *AGL12*, *PI*, *SQUA* and *SVP* supported monocot-specific WGD (τ). Interestingly, similarly to the *A. trifoliata*, other basal eudicots also had duplicated genes in most of subfamilies of type II genes including *MIKC**, *AG*, *ANR1*, *AP3/DEF*, *SEP*, *SOC1/TM3*, *SQUA* and *SVP* which highly overlap with core eudicots (Table S5 and Figure S3). These subfamilies were not only duplicated in the core eudicot lineage. Moreover, *DEF/AP3*, *SQUA*, *SOC1/TM3* and *SVP* possibility derived from a duplication event in early eudicot (ψ) and a core eudicot-wide diploid fusion event (ω) as presented in scenario 2 rather than scenario 1 in Figure 4c (Figure S3 and Table S5).

3.7. Differential Expression Analysis of MADS-Box Genes of *A. trifoliata* in Various Fruit Tissues

The expression of many MADS-box genes in the flesh, seeds and peels showed that all type I genes were expressed at very low levels in all samples (Figure 5a,b). In contrast, many type II genes have a higher expression level. Three type II genes (*AktAG_3*, *AktAG_1* and *AktSEP_3*) exhibited high expression at every stage in all tissues, especially flesh (Figure 5c). In addition, many type II genes also exhibited differential expression levels among different tissues. For example, *AktAG_2*, *AktAGL6_2*, *AktMIKC*_8*, *AktMIKC*_9*, *AktMIKC*_5*, *AktAGL15*, *AktGGM13* and *AktAGL6_1* exhibited high expression levels in seeds and low expression levels in both flesh and peel (Figure 5c). Moreover, *AktMIKC*_8* had a differential expression level among different developmental stages of seeds, with an obvious increase with growth progress (Figure 5a). Because *SEEDSTICK* (*STK*) subfamily are a sister group of *AG* genes, it is difficult to differentiate them based solely on a phylogenetic tree. We found that the *AktAG_2* gene was relatively highly expressed in all seed stages, while *AktAG_1* and *AktAG_3* genes were expressed at relatively lower levels in seed. Therefore, *AktAG_2* should be classified as part of *STK* functional genes in seed development.

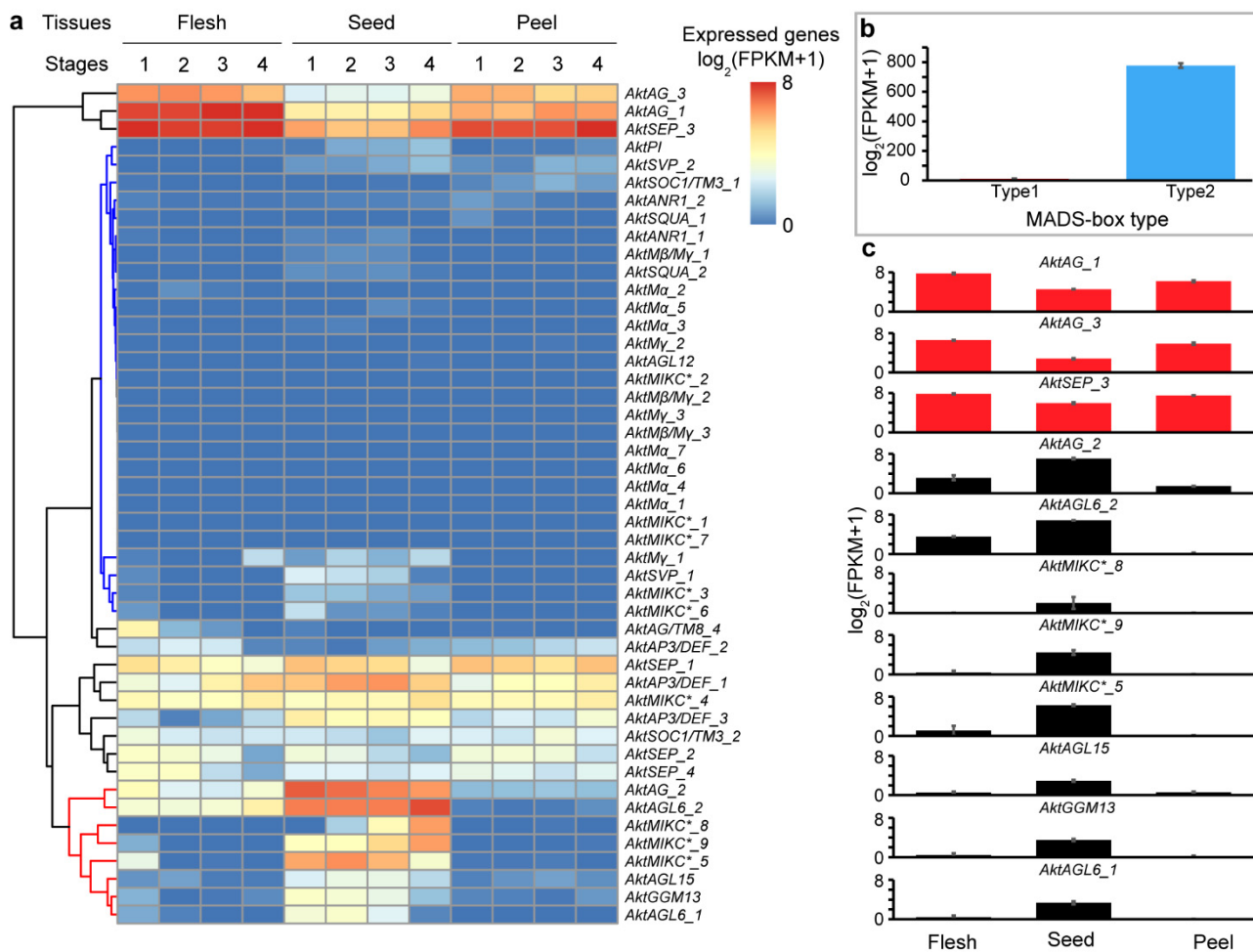


Figure 5. Expression levels of 47 MADS-box genes in different tissues and developmental stages. (a) Cluster results of gene expressions, FPKM represent fragments per kilobase of transcript per million fragments mapped. (b) Total expression levels of type I and type II genes, respectively. (c) Expressions of representative gene in three tissues.

4. Discussion

The name of MADS-box transcription factors indicates that they could exist widely and be highly conserved in plants, animals and fungi, which has been confirmed by various studies [45], and some homology has even been found in the common ancestor to prokaryotes and eukaryotes [46]. In plants, MADS-box transcription factors have evolved important regulatory functions in various biological processes, such as disease resistance [47], signal transduction [48] and morphogenesis, especially flower formation and seed development [49]. MADS-box genes were identified from all plant genomes, but whether their number is related to evolutionary divergence or genomic size was unclear.

Previous studies have shown that the number of MADS-box genes varies slightly among different species and usually ranges in the dozens, although up to 300 have been identified in the extreme case of Wheat [50]. The numbers were 70, 65, 47 and 90 in the genome basal angiosperm *Nymphaea colorata* [51], monocot *Sorghum bicolor* [52], basal eudicot *Aquilegia coerulea* [53] and core eudicot *V. vinifera* [54], respectively, and they were very close to the corresponding numbers (Table S5). This indicated that the number of MADS-box genes could be slightly related to the evolutionary clade. In the present study, we identified the numbers of these taxa in the genomes of 25 species from five different evolutionary clades: mosses, basal angiosperms, monocots, basal eudicots and core eudicots (Tables S4 and S5). Although those of moss were obviously fewer than those

of all angiosperms, their variation was slight among the four clades within angiosperms, in which their variation was smaller in monocots than in eudicots. This result further supported the view that the number of MADS-box genes was not usually related to the position on phylogenetic trees of species, and only a dozen genes indicated that they could regulate various biological processes. In addition, previous reports showed 71, 81, 162 and 300 MADS-box genes in 430 Mb *O. sativa* [55], 475 Mb *V. vinifera* [56], 2870 Mb *P. somniferum* [57] and 1700 Mb Wheat [58] genomes, respectively, which suggested that the greater the number of MADS-box genes was, the larger the genomic size was. In this study, we found that the correlation coefficient between the MADS-box gene number and genomic size was 0.52 (Table S5), indicating a moderate relationship. Therefore, we agree with the view that the number of MADS-box genes could be generally related to genomic size and scarcely related to species divergence.

Here, we identified 47 MADS-box genes from the 619 Mb *A. trifoliata* genome (Table 1), of which 10 were previously reported [22,23] and 37 were newly identified by genome-wide analysis (Table S2) which were classified into all three subfamilies of type I and 13 subfamilies of type II MADS-box genes when only the *A. thaliana* genome was used as the reference (Figure 1), while there were 14 subfamilies of type II genes when *A. thaliana*, *O. sativa* and *A. trichopoda* were used as the reference (Table S4). Our comparison analysis found that *EVM0004910* in the *AG* lineage (Figure 1) was reclassified into *TM8* (Table S4). This could be explained by the fact that there was no *TM8* clade in *A. thaliana* [29]. Therefore, there were a total of seventeen subfamilies, and only two subfamilies, *OsMADS32* and *FLC*, were not identified in the MADS-box genes of the *A. trifoliata* genome (Table 1).

Structurally, type I genes encode only one simple SRF-like MADS domain, containing one to two exons, whereas type II genes encode MEF2-like or MIKC-type MADS domains with multiple additional domains, such as the K-box (keratin-like domain), I-box (intervening domain) and variable C-terminal domain (C-terminal region) [59], and evolutionarily, type II genes are more conserved than type I genes [4]. In type I genes, *AktM γ _3* has three exons, and the other has one or two exons, while the exon number in type II was largely variable (Table 1). We found that the average amino acid number encoded by type I and type II genes was close, while the exon number of type II was significantly greater than that of type I (Table S1); additionally, the DNA sequence of type I had good continuity compared with that of type II (Figure 2d). Similar results have been reported in grape [54], rice [37] and *B. distachyon* [60].

The number of MADS-box genes was small, but the subfamilies, especially those of type II genes, were very numerous (Table S5), which indicated that they could have experienced different evolutionary events, such as different genomic duplications and natural selection styles. In addition, despite the small number of MADS-box genes, they were assigned to all 16 chromosomes except chromosome 6 (Figure 3). Both the rich subfamilies and wide chromosomal distribution suggested that whole genomic duplication (WGD) or segmental duplication could be important duplication styles for *A. trifoliata* MADS-box genes. By performing a synteny analysis, we found that 27 (57.4%) out of the 47 total identified *A. trifoliata* MADS-box genes experienced WGD or segmental duplication events. Moreover, all *Ka/Ks* values were far less than 1 (Table S3), which clearly indicates purifying selection. Thus, we propose that MADS-box genes in *A. trifoliata* mainly experienced WGD or segmental duplication and strong purifying selection in the long evolutionary period.

Previous studies have suggested that type II genes exhibit higher conservation than type I genes [4]. In the present study, we found that the coefficient of variation (CV) of all three subfamilies, *M α* , *M β* and *M γ* , of type I genes was greater than 0.8, while that of both *MIKC** and *MIKC^C* of type II genes was less than 0.5 among 22 species (Figure 4a). In addition, although *AT5G55690* and *AT5G58890* were previously classified into *M β* (Table 1), the amino acid sequence encoded by them has a high similarity with *M γ* (Figure S1), which provides a reasonable explanation for why they were classified into the *M γ* clade in Figure 1 and indicated that there was a rapid evolution from *M β* to *M γ* . Both results confirmed that type II genes were highly conserved. Type II genes also played a key role in investigating

the species diversification of flowering plants because the ABC flowering model was commonly regulated by several subfamilies of type II [12,14], even some subfamilies of type II genes involved in specification of ovule and flower development in seed plants were significant expansion in ferns [61], so we focused on the phylogenetic tree analysis for subfamilies of type II (Figure S3).

Only *GGM13* subfamilies of *MIKC^C* were identified in both moss species (Table S5), which indicated that these genes could have originated early, generally before basal angiosperm divergence. In contrast, *FLC* subfamilies were only detected in core eudicots (Table S5) and were putatively core eudicot-specific, which suggested that they could have recently originated, after core eudicots diverged from basal eudicots. Previous studies have suggested a series of WGDs in angiosperms, mainly including ancestral angiosperms or seed plant-wide WGDs (ζ or ϵ), monocot-specific WGDs (τ) and core eudicot-specific γ WGTs [44,62,63]. In our study, traces of ζ or ϵ could be detected in the phylogenetic tree of *DEF/AP3*, *SQUA*, *SOC1/TM3* and *SVP*. On the one hand, there were at least two copies of their homologues in many species among 20 angiosperms, especially in *A. trichopoda*, a model basal angiosperm (Table S5). On the other hand, their topological structures are largely consistent with the early angiosperm species phylogeny (Figure S3). Likewise, traces of monocot-specific τ could also be detected in the phylogenetic tree of the *AGL12*, *PI*, *SQUA* and *SVP* subfamilies (Figure S3). Only the traditional γ -WGT would be challenged by the phylogenetic tree of some subfamilies of type II genes.

The traditional view is that the γ event occurred approximately 117 million years ago (Mya) and at the early stages of core eudicots after the divergence of the Ranunculales and core eudicots [57,63]. In fact, both the composition and occurrence time of the γ event are still in doubt. For example, a shared WGD event on the common ancestor of basal eudicot, which is phylogenetic close to the core eudicot γ -WGT was also detected in *A. trifoliata* and many other basal eudicots based on the transcriptome data [61]. However, another genome analyses of *A. coerulea* provided some evidence suggesting that the γ -WGT event may have consisted of two steps [64]. Obviously, precisely positioning the γ -WGT within a narrow time window is difficult mainly due to the following two reasons: one is the absence of knowledge about the WGT, such as whether a single event or two independent events resulted in the triplication of the genome, and the other is the short time space from eudicot divergence to core eudicot divergence. The phylogenetic trees of *DEF/AP3*, *SQUA*, *SOC1/TM3* and *SVP* in our study supported one common eudicot-wide duplication event (ψ) and core eudicot-wide additional diploid fusion event (ω), putatively occurring before and after core eudicot divergence from basal eudicots, respectively (Figure S3), which could be explained well by scenario 2 in Figure 4c (Figures 4c and S3 and Table S5). Furthermore, the genes of the *FLC* and *AGL15* subfamilies were only duplicated in the core eudicot lineage, and two clades of homologues were also more likely to be duplicated from a potential single ω diploidization event rather than an indivisible γ triploidization event (Figures 4c and S3 and Table S5). Parts of type II MADS-box gene subfamilies, such as *AP3/DEF*, *SEP*, *SOC1/TM3* and *SQUA*, were considered to be duplicated from core eudicot-wide γ -WGT events in previous reports [14]. Unfortunately, MADS-box genes were not identified completely due to the use of transcriptomic rather than genomic data. For example, only a few MADS-box genes have been previously identified in *A. trifoliata* (Table S2) [22,23] and in *A. coerulea* (Table S5) [53]. Importantly, previous phylogenetic analyses of other basal eudicots also afforded some strong evidence of two close duplication events supporting the ψ and ω hypothesis [65]. Comprehensively, the phylogenetic trees of these subfamilies agreed well with scenario 2 rather than scenario 1 of Figure 4c, which suggested that γ could be a multiple event consisting of eudicot-wide ψ and core eudicot-specific ω rather than a single γ event. Although the origin of the γ -WGT event is currently a highly-debated field which cannot be solved simply by using the phylogenetic gene tree of the MADS-box family, our results provide an important clue to solve this problem in the future.

Differential expression analysis using comparative transcriptomics can provide important information about gene functions. We found that many MADS-box genes, especially type I genes, were expressed at extremely low levels in all three tissues (flesh, seed and peel) at all developmental stages, while *AktAG_3*, *AktAG_1* and *AktSEP_3* exhibited high expression in almost all samples (Figure 5), which indicated that only a few MADS-box genes could meet the regulatory requirement of fruit formation and that MADS-box gene expression is usually organ- or tissue-specific. Some previous reports suggested that some genes of both the *AG* and *SEP* subfamilies could be related to fruit development [66,67]; therefore, we concluded that the three genes could regulate whole fruit development. In addition, we also found that eight genes of five subfamilies (*MIKC**, *AG*, *AGL6*, *AGL15* and *GGM13*) were only expressed a high level in seeds, and in all subfamilies except *AGL15* which existed in all twenty angiosperms (Table S5). There are various reports stating that *MIKC** [68], *AG* [67], *AGL6* [69] and *GGM13* [70] participate in regulating seed development. Therefore, *AktAG_2*, *AktAGL6_2*, *AktMIKC*_8*, *AktMIKC*_9*, *AktMIKC*_5*, *AktAGL15*, *AktGGM13* and *AktAGL6_1* positively regulated seed development in *A. trifoliata*.

5. Conclusions

We identified 47 nonredundant MADS-box genes in the *A. trifoliata* genome, and 46 of them were physically assigned to 16 high-quality assembled pseudochromosomes. All 47 genes were classified into 17 subfamilies, and many of them putatively experienced WGD or segmental duplication and purification in the evolutionary process. Several candidate MADS-box genes involved in fruit or seed development were screened. Importantly, traces of major WGD events in whole angiosperms could be detected in many subfamilies of type II genes. Interestingly, the phylogenetic tree of the *DEF/AP3*, *SQUA*, *SOC1/TM3*, *SVP*, *FLC* and *AGL15* subfamilies provides some evidence that γ events are multiple events rather than a single event consisting of a eudicot-wide ψ and core eudicot-specific ω . Comprehensively, both new data resources and insight into traditional γ events would be helpful to practically improve the objective traits of *A. trifoliata* fruits and to completely elucidate the evolutionary process of early-stage eudicots.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101777/s1>, Figure S1: Comparison of MADS-box domains between $M\beta$ and $M\gamma$; Figure S2: Synteny results of the *A. trifoliata* genome and MADS-box genes; Figure S3: Phylogenetic tree of type II genes from 22 plant species; Table S1: Statistical analysis of gene structures between type I and type II MADS-box genes; Table S2: Previously reported *A. trifoliata* MADS-box genes; Table S3: Substitution rate and selection pressure of MADS-box duplicated gene pairs; Table S4: Identification of MADS-box genes in 22 plant species; Table S5: Statistics of ancestral duplication events and number of MADS-box subfamilies in 22 species.

Author Contributions: Conceptualization, S.Z., J.S. and P.L.; methodology, S.Z.; software, S.Z.; validation, J.G. and Q.L.; formal analysis, H.Y.; investigation, H.Y.; resources, P.L.; data curation, S.Z.; writing—original draft preparation, S.Z. and H.Y.; writing—review and editing, S.Z., H.Y. and P.L.; visualization, S.Z.; supervision, Z.L., T.R. and F.T.; project administration, P.L.; funding acquisition, P.L. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Department of Sichuan Province, grant number 2019YFS0020, 2020YFN0057; and the Bureau of Science and Technology of Fuling of Chongqing, grant number [FLKJ-2021ABB1016].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data analyzed during this study are included in the manuscript and supplementary information files, and transcriptomic data of *A. trifoliata* fruit tissues have been deposited in the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/sra> under the accession numbers SAMN16551931–33, SAMN16551934–36, SAMN16551937–39 and SAMN16551940–42, accessed on 28 June 2022).

Acknowledgments: We are grateful to the Science and Technology Department of Sichuan Province and Science and Technology Bureau of Yaan for supporting this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Huang, C.; Wang, Z.; Quinn, D.; Suresh, S.; Hsia, K.J. Differential growth and shape formation in plant organs. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12359–12364. [[CrossRef](#)] [[PubMed](#)]
- Sun, Y.; Dong, L.; Zhang, Y.; Lin, D.; Xu, W.; Ke, C.; Han, L.; Deng, L.; Li, G.; Jackson, D.; et al. 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol.* **2020**, *21*, 143. [[CrossRef](#)] [[PubMed](#)]
- Qu, Y.; Kong, W.; Wang, Q.; Fu, X. Genome-wide identification MIKC-type MADS-Box gene family and their roles during development of floral buds in Wheel Wingnut (*Cyclocarya paliurus*). *Int. J. Mol. Sci.* **2021**, *22*, 10128. [[CrossRef](#)]
- Nam, J.; Kim, J.; Lee, S.; An, G.; Ma, H.; Nei, M. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1910–1915. [[CrossRef](#)]
- Passmore, S.; Elble, R.; Tye, B. A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. *Gene. Dev.* **1989**, *3*, 921–935. [[CrossRef](#)] [[PubMed](#)]
- Schwarz-Sommer, Z.; Huijser, P.; Nacken, W.; Saedler, H.; Sommer, H. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **1990**, *250*, 931–936. [[CrossRef](#)]
- Sommer, H.; Beltrán, J.P.; Huijser, P.; Pape, H.; Lönig, W.E.; Saedler, H.; Schwarz-Sommer, Z. Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. *EMBO J.* **1990**, *9*, 605–613. [[CrossRef](#)]
- Norman, C.; Runswick, M.; Pollock, R.; Treisman, R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **1988**, *55*, 989–1003. [[CrossRef](#)]
- Theissen, G.; Becker, A.; Di Rosa, A.; Kanno, A.; Kim, J.T.; Münster, T.; Winter, K.; Saedler, H. A short history of MADS-box genes in plants. *Plant Mol. Biol.* **2000**, *42*, 115–149. [[CrossRef](#)]
- Shore, P.; Sharrocks, A.D. The MADS-Box family of transcription factors. *Eur. J. Biochem.* **1995**, *229*, 1–13. [[CrossRef](#)]
- Xu, Z.; Zhang, Q.; Sun, L.; Du, D.; Cheng, T.; Pan, H.; Yang, W.; Wang, J. Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in *Prunus mume*. *Mol. Genet. Genom.* **2014**, *289*, 903–920. [[CrossRef](#)] [[PubMed](#)]
- Pelaz, S.; Ditta, G.S.; Baumann, E.; Wisman, E.; Yanofsky, M.F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **2000**, *405*, 200–203. [[CrossRef](#)] [[PubMed](#)]
- Coito, J.L.; Silva, H.; Ramos, M.J.N.; Montez, M.; Cunha, J.; Amâncio, S.; Costa, M.M.R.; Rocheta, M. Vitis flower sex specification acts downstream and independently of the ABCDE model genes. *Front. Plant Sci.* **2018**, *9*, 1029. [[CrossRef](#)]
- Vekemans, D.; Proost, S.; Vanneste, K.; Coenen, H.; Viaene, T.; Ruelens, P.; Maere, S.; Van de Peer, Y.; Geuten, K. Gamma paleohexaploidy in the stem lineage of core eudicots: Significance for MADS-Box gene and species diversification. *Mol. Biol. Evol.* **2012**, *29*, 3793–3806. [[CrossRef](#)] [[PubMed](#)]
- Crane, P.R.; Friis, E.M.; Pedersen, K.R. The origin and early diversification of angiosperms. *Nature* **1995**, *374*, 27–33. [[CrossRef](#)]
- Yang, H.; Chen, W.; Fu, P.; Zhong, S.; Guan, J.; Luo, P. Developmental stages of *Akebia trifoliata* fruit based on volume. *Hortic. Sci. Technol.* **2021**, *39*, 823–831. [[CrossRef](#)]
- Li, L.; Yao, X.; Zhong, C.; Chen, X.; Huang, H. *Akebia*: A potential new fruit crop in china. *HortScience* **2010**, *45*, 4–10. [[CrossRef](#)]
- Chinese Pharmacopoeia Commission. *The Pharmacopoeia of the People's Republic of China*; China Medicine Science and Technology Press: Beijing, China, 2005.
- Guan, J.; Fu, P.; Wang, X.; Yu, X.; Zhong, S.; Chen, W.; Yang, H.; Chen, C.; Yang, H.; Luo, P. Assessment of the breeding potential of a set of genotypes selected from a natural population of *Akebia trifoliata* (Three-Leaf Akebia). *Horticultrae* **2022**, *8*, 116. [[CrossRef](#)]
- Du, Y.; Jiang, Y.; Zhu, X.; Xiong, H.; Shi, S.; Hu, J.; Peng, H.; Zhou, Q.; Sun, W. Physicochemical and functional properties of the protein isolate and major fractions prepared from *Akebia trifoliata* var. *australis* seed. *Food Chem.* **2012**, *133*, 923–929. [[CrossRef](#)]
- Zhang, G.; Chen, X. Application of vertical greening as a new model of expanding urban green spaces. *J. Landsc. Res.* **2013**, *5*, 14–16.
- Liu, C.; Zhang, J.; Zhang, N.; Shan, H.; Su, K.; Zhang, J.; Meng, Z.; Kong, H.; Chen, Z. Interactions among proteins of floral MADS-Box genes in basal eudicots: Implications for evolution of the regulatory network for flower development. *Mol. Biol. Evol.* **2010**, *27*, 1598–1611. [[CrossRef](#)] [[PubMed](#)]
- Shan, H.; Su, K.; Lu, W.; Kong, H.; Chen, Z.; Meng, Z. Conservation and divergence of candidate class B genes in *Akebia trifoliata* (Lardizabalaceae). *Dev. Genes Evol.* **2006**, *216*, 785–795. [[CrossRef](#)] [[PubMed](#)]
- Shan, H.; Zahn, L.; Guindon, S.; Wall, P.K.; Kong, H.; Ma, H.; DePamphilis, C.W.; Leebens-Mack, J. Evolution of plant MADS box transcription factors: Evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol. Biol. Evol.* **2009**, *26*, 2229–2244. [[CrossRef](#)] [[PubMed](#)]
- Huang, H.; Liang, J.; Tan, Q.; Ou, L.; Li, X.; Zhong, C.; Huang, H.; Møller, I.M.; Wu, X.; Song, S. Insights into triterpene synthesis and unsaturated fatty-acid accumulation provided by chromosomal-level genome analysis of *Akebia trifoliata* subsp. *australis*. *Hortic. Res.-Engl.* **2021**, *8*, 33. [[CrossRef](#)] [[PubMed](#)]

26. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)]
27. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)]
28. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [[CrossRef](#)]
29. Pařenicová, L.; de Folter, S.; Kieffer, M.; Horner, D.S.; Favalli, C.; Busscher, J.; Cook, H.E.; Ingram, R.M.; Kater, M.M.; Davies, B.; et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. *Plant Cell* **2003**, *15*, 1538–1551. [[CrossRef](#)]
30. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
31. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
32. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)] [[PubMed](#)]
33. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)] [[PubMed](#)]
34. Ameline-Torregrosa, C.; Wang, B.; O’Bleness, M.S.; Deshpande, S.; Zhu, H.; Roe, B.; Young, N.D.; Cannon, S.B. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* **2008**, *146*, 5–21. [[CrossRef](#)]
35. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
36. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)] [[PubMed](#)]
37. Arora, R.; Agarwal, P.; Ray, S.; Singh, A.K.; Singh, V.P.; Tyagi, A.K.; Kapoor, S. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom.* **2007**, *8*, 242. [[CrossRef](#)]
38. Kater, M.M.; Dreni, L.; Colombo, L. Functional conservation of MADS-box factors controlling floral organ identity in rice and *Arabidopsis*. *J. Exp. Bot.* **2006**, *57*, 3433–3444. [[CrossRef](#)] [[PubMed](#)]
39. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **2013**, *342*, 1241089. [[CrossRef](#)]
40. Perteau, M.; Kim, D.; Perteau, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [[CrossRef](#)]
41. Perteau, M.; Perteau, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)]
42. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
43. Wickham, H. Data Analysis. In *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Wickham, H., Ed.; Springer International Publishing: Cham, Switzerland, 2016; pp. 189–201.
44. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100. [[CrossRef](#)] [[PubMed](#)]
45. Gramzow, L.; Ritz, M.S.; Theißen, G. On the origin of MADS-domain transcription factors. *Trends Genet.* **2010**, *26*, 149–153. [[CrossRef](#)] [[PubMed](#)]
46. Mushegian, A.R.; Koonin, E.V. Sequence analysis of eukaryotic developmental proteins: Ancient and novel domains. *Genetics* **1996**, *144*, 817–828. [[CrossRef](#)]
47. Zhang, H.; Teng, W.; Liang, J.; Liu, X.; Zhang, H.; Zhang, Z.; Zheng, X. MADS1, a novel MADS-box protein, is involved in the response of *Nicotiana benthamiana* to bacterial harpin_{Xoo}. *J. Exp. Bot.* **2016**, *67*, 131–141. [[CrossRef](#)]
48. Chen, M.; Nie, G.; Yang, L.; Zhang, Y.; Cai, Y. Homeotic transformation from stamen to petal in *Lilium* is associated with MADS-box genes and hormone signal transduction. *Plant Growth Regul.* **2021**, *95*, 49–64. [[CrossRef](#)]
49. Mao, L.; Begum, D.; Chuang, H.; Budiman, M.A.; Szymkowiak, E.J.; Irish, E.E.; Wing, R.A. JOINTLESS is a MADS-box gene controlling tomato flower abscissionzone development. *Nature* **2000**, *406*, 910–913. [[CrossRef](#)]
50. Zhao, T.; Ni, Z.; Dai, Y.; Yao, Y.; Nie, X.; Sun, Q. Characterization and expression of 42 MADS-box genes in wheat (*Triticum aestivum* L.). *Mol. Genet. Genomics.* **2006**, *276*, 334–350. [[CrossRef](#)]
51. Zhang, L.; Chen, F.; Zhang, X.; Li, Z.; Zhao, Y.; Lohaus, R.; Chang, X.; Dong, W.; Ho, S.Y.W.; Liu, X.; et al. The water lily genome and the early evolution of flowering plants. *Nature* **2020**, *577*, 79–84. [[CrossRef](#)]
52. Zhao, Y.; Li, X.; Chen, W.; Peng, X.; Cheng, X.; Zhu, S.; Cheng, B. Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant Cell Tissue Organ Cult.* **2011**, *105*, 159–173. [[CrossRef](#)]
53. Sharma, B.; Kramer, E.M. The MADS-box gene family of the basal eudicot and hybrid *Aquilegia coerulea* ‘Origami’ (Ranunculaceae). *Ann. Mo. Bot. Gard.* **2014**, *99*, 313–322. [[CrossRef](#)]

54. Grimplet, J.; Martínez-Zapater, J.M.; Carmona, M.J. Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genom.* **2016**, *17*, 80. [[CrossRef](#)] [[PubMed](#)]
55. Ouyang, S.; Zhu, W.; Hamilton, J.; Lin, H.; Campbell, M.; Childs, K.; Thibaud-Nissen, F.; Malek, R.L.; Lee, Y.; Zheng, L.; et al. The TIGR rice genome annotation resource: Improvements and new features. *Nucleic Acids Res.* **2007**, *35*, D883–D887. [[CrossRef](#)] [[PubMed](#)]
56. Jaillon, O.; Aury, J.; Noel, B.; Policriti, A.; Clepet, C.; Cassagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467. [[CrossRef](#)] [[PubMed](#)]
57. Guo, L.; Winzer, T.; Yang, X.; Li, Y.; Ning, Z.; He, Z.; Teodor, R.; Lu, Y.; Bowser, T.A.; Graham, I.A.; et al. The opium poppy genome and morphinan production. *Science* **2018**, *362*, 343–347. [[CrossRef](#)] [[PubMed](#)]
58. Null, N.; Mayer, K.F.X.; Rogers, J.; Doležel, J.; Pozniak, C.; Eversole, K.; Feuillet, C.; Gill, B.; Friebe, B.; Lukaszewski, A.J.; et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **2014**, *345*, 1251788. [[CrossRef](#)]
59. Li, C.; Wang, Y.; Xu, L.; Nie, S.; Chen, Y.; Liang, D.; Sun, X.; Karanja, B.K.; Luo, X.; Liu, L. Genome-wide characterization of the MADS-box gene family in radish (*Raphanus sativus* L.) and assessment of its roles in flowering and floral organogenesis. *Front. Plant Sci.* **2016**, *7*, 1390. [[CrossRef](#)]
60. Wei, B.; Zhang, R.; Guo, J.; Liu, D.; Li, A.; Fan, R.; Mao, L.; Zhang, X. Genome-wide analysis of the MADS-box gene family in *Brachypodium distachyon*. *PLoS ONE* **2014**, *9*, e84781. [[CrossRef](#)]
61. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **2019**, *574*, 679–685. [[CrossRef](#)]
62. Jiao, Y.; Leebens-Mack, J.; Ayyampalayam, S.; Bowers, J.E.; McKain, M.R.; McNeal, J.; Rolf, M.; Ruzicka, D.R.; Wafula, E.; Wickett, N.J.; et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **2012**, *13*, R3. [[CrossRef](#)]
63. Jiao, Y.; Li, J.; Tang, H.; Paterson, A.H. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* **2014**, *26*, 2792–2802. [[CrossRef](#)] [[PubMed](#)]
64. Aköz, G.; Nordborg, M. The *Aquilegia* genome reveals a hybrid origin of core eudicots. *Genome Biol.* **2019**, *20*, 256. [[CrossRef](#)] [[PubMed](#)]
65. Shan, H.; Zhang, N.; Liu, C.; Xu, G.; Zhang, J.; Chen, Z.; Kong, H. Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Mol. Phylogenet. Evol.* **2007**, *44*, 26–41. [[CrossRef](#)] [[PubMed](#)]
66. Elitzur, T.; Yakir, E.; Quansah, L.; Zhangjun, F.; Vrebalov, J.; Khayat, E.; Giovannoni, J.J.; Friedman, H. Banana MaMADS transcription factors are necessary for fruit ripening and molecular tools to promote shelf-life and food security. *Plant Physiol.* **2016**, *171*, 380–391. [[CrossRef](#)]
67. Zhao, J.; Gong, P.; Liu, H.; Zhang, M.; He, C. Multiple and integrated functions of floral C-class MADS-box genes in flower and fruit development of *Physalis floridana*. *Plant Mol. Biol.* **2021**, *107*, 101–116. [[CrossRef](#)]
68. Kwantes, M.; Liebsch, D.; Verelst, W. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. *Mol. Biol. Evol.* **2012**, *29*, 293–302. [[CrossRef](#)]
69. Koo, S.C.; Bracko, O.; Park, M.S.; Schwab, R.; Chun, H.J.; Park, K.M.; Seo, J.S.; Grbic, V.; Balasubramanian, S.; Schmid, M.; et al. Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box Gene AGAMOUS-LIKE6. *Plant J.* **2010**, *62*, 807–816. [[CrossRef](#)]
70. Nesi, N.; Debeaujon, I.; Jond, C.; Stewart, A.J.; Jenkins, G.I.; Caboche, M.; Lepiniec, L. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **2002**, *14*, 2463–2479. [[CrossRef](#)]