

Article

MIKC-Type MADS-Box Gene Family Discovery and Evolutionary Investigation in *Rosaceae* Plants

Yue Qin ^{1,2}, Gaopu Zhu ¹ , Fangdong Li ¹, Lin Wang ¹, Chen Chen ¹ and Han Zhao ^{1,*}

¹ Research Institute of Non-Timber Forestry Chinese Academy of Forestry, Zhengzhou 450003, China; qinyue870711@163.com (Y.Q.); zhugaopu@163.com (G.Z.); lifangdong66@163.com (F.L.); 174816279@163.com (L.W.)

² College of Forestry, Nanjing Forestry University, Nanjing 210037, China

* Correspondence: zhaohan@caf.ac.cn; Tel.: +86-136-0765-0203

Abstract: MADS-box is an important transcriptional regulatory element in plant growth. The MIKC-type MADS-box genes play important roles. However, the identification and evolutionary investigation of MIKC-type MADS-box family members in *Rosaceae* have been inadequate. Therefore, based on whole-genome data from *Prunus dulcis*, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, *Prunus mira*, and *Amygdalus nana*, we depicted the evolution and divergence patterns of MIKC-type MADS-box family genes. In this study, we found 222 MIKC-type MADS-box genes from six *Rosaceae* species. These genes were classified into five clades, and only motif 1 was identified across all MIKC-type MADS-box proteins, except PdMADS42 and PmiMADS16. The structural properties of these genes significantly varied in sequence lengths between species, despite the high levels of similarity in exon lengths and numbers. MIKC-type MADS-box genes were found to have mostly been limited through purifying selection processes. Remarkably divergent regions were found inside the MIKC-type MADS-box genes' domains, where clade III displayed more conserved activities and may have retained more original functions over the evolutionary process; clade I, on the other hand, may have undergone substantial functional limitations in a specific functional role. These findings provide the groundwork for future research into the molecular evolutionary processes of the plant MIKC-type MADS-box gene family.

Keywords: selective evolutionary pressure; MIKC-type MADS-box genes; *Rosaceae* plants; gene evolution; functional differences



Citation: Qin, Y.; Zhu, G.; Li, F.; Wang, L.; Chen, C.; Zhao, H. MIKC-Type MADS-Box Gene Family Discovery and Evolutionary Investigation in *Rosaceae* Plants. *Agronomy* **2023**, *13*, 1695. <https://doi.org/10.3390/agronomy13071695>

Academic Editor: Antonio Lupini

Received: 24 April 2023

Revised: 21 June 2023

Accepted: 22 June 2023

Published: 25 June 2023



Copyright: © 2023 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

Flower organs are the basis of plant evolution and classification, and have always played important roles in plant development and biological research [1]. The MADS-box gene family is a class of transcription factors widely distributed among animals, plants, fungi, and other organisms [2,3]. MADS-box derives its name from the first initials of fore groups of genes: the mini chromosome maintenance 1 genes from yeast (MCM1), the agamous genes from *Arabidopsis thaliana* (AG), deficiens genes from snapdragon (DEF), and serum response factor genes from humans (SRF). These genes all contain a MADS-box conserved region of about 180 bp at the N-terminal, while identifying the similar DNA sequences [4,5]. The plant MADS-box genes were discovered from the mutant study of the model plants *Arabidopsis* and *Snapdragon*. MADS-box genes are widely distributed throughout the genome and expressed at different development stages. Their functions are involved in all aspects of plant growth and development, especially the development and regulation of flower organs, including initiation, differentiation, morphogenesis, and gene transcription regulation, and have become some of the most relevant transcription factor family in plants [6]. The MADS-box gene family in plants can be divided into Type I and Type II. Due to the relatively simple structure and functional redundancy of Type I genes, studies on plant MADS-box genes mostly focus on Type II genes, also known as

MIKC-type genes [1]. The encoded protein consists of four domains: MADS domain, I domain (intervening domain), K domain (keratin-like domain), and C-terminal domain. [7]. The MADS-box domain and region I are associated with the formation of protein dimers, the K-box domain mediates protein interactions, and the C-terminal is associated with transcription factor activation and the formation of higher-order protein complexes [8,9].

With the development of sequencing technology and bioinformatics, more and more plants have been thoroughly studied through genome or transcriptome sequencing. Many MADS-box gene family members and their functions within the flower organs of various species have been identified and revealed, including genes controlling flowering time and flowering inhibition (FLC, SVP, SOC1, AGL24), genes regulating floral meristem differentiation (CAL and FUL), and genes governing floral organ characteristics [10–13]. The pathway controlling flowering initiation and the genes involved, response pathway, and expression regulation mechanism were also analyzed. For example, overexpression of BpAP1 gene in European silver birch can cause early flowering [14], while overexpression of PvMADS5, a SEP homologous gene in bamboo, causes delayed flowering in *Arabidopsis* plants [15]. These studies have improved understanding of the MADS-box gene family, but many members are yet to be discovered, and the regulation mechanism of flower organ development requires further research.

Prunus dulcis, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, *Prunus mira*, and *Amygdalus nana* are important economic forestry species in *Rosaceae*. The study of genes related to flower regulation is of great value not only in phylogeny and taxonomy, but also in genetic improvement and cultivation techniques. In previous studies, researchers have conducted the identification and function analysis of the MADS-box gene family in *Prunus persica*, *Prunus armeniaca*, and *Prunus dulcis*, obtaining a number of genes, illustrating the genes' structure and sequence features, and analyzing the gene functions. However, studies investigating its evolution between different species are limited. Thus, in this study, we used genome-wide data of six *Rosaceae* plant species (*Prunus dulcis*, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, *Prunus mira*, and *Amygdalus nana*) to screen for MIKC-type MADS-box gene family members. We also looked at the evolutionary relationship, areas of functional divergence, and amino acid structures of the discovered MIKC-type MADS-box genes. These findings are not only important for elucidating the adaptive history of the *Rosaceae* species, but also for predicting the functional area of the MIKC-type MADS-box genes.

2. Materials and Methods

2.1. Gene Family Identification

Genomic sequence and annotation data for *Prunus dulcis*, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, and *Prunus mira* were obtained from the GDR database (rosaceae.org), and the *Amygdalus nana* genome assembly was obtained through our research facility (private information). We utilized the sequences of *Arabidopsis* MADS-box proteins as a search basis to retrieve MIKC-type MADS-box proteins, employing the local alignment searching modality (BLASTP) with an E value of 1×10^{-10} [16,17]. Then, HMMER v3.3.2 with the hidden Markov model (HMM) profile (PF00319 and PF01486) was employed to explore the proteomes with parameters 'hmmsearch—notextw—acc—cut_ga_' [18]. Next, we compared the MADS-box domain using SMART (smart.embl.de) with the Pfam database in the normal smart mode.

2.2. Sequence Analysis

We used MAFFT 7.310 with parameters 'mafft—maxiterate 1000—globalpair input > output' to perform multiple protein sequence alignment [19]. The conserved motifs of the MIKC-type MADS-box proteins were investigated using the MEME Suite 5.5.0 with parameters 'meme julei_6_species.fasta—protein—oc.—nostatus—time 14,400—mod anr—nmotifs 10—minw 6—maxw 100—objfun classic—markov_order 0' [20]. WebLogo 2.8.2

was used to display the sequences [21], and Tbttools software v1.106 was used to visualize the MEME and gene structure results.

2.3. Phylogenetic and Conservative Analyses

ProtTest 3.4 with model 'LG + I + G + F' was used to construct the optimum protein substitution model [22]. Using the neighbor-joining technique and MEGA Version 10.2.2 software, we built a phylogenetic tree using JTT + G with 1000 bootstrap repetitions [23]. Then, we used R-ggtree software 1.8.1 to enhance the phylogenetic tree [24–26].

2.4. Selective Evolutionary Pressure Analysis

The Codeml program EasyCodeml Version 1.4 was used to analyze the selective evolutionary pressure at the branch, site, and branch-site models (BM, SM, and BSM, respectively) [27]. Branch models are used to identify the adaptive development of several branches. Site models can discover the ω ratio between the sites and positive selected sites. Provided that the ω ratio varies between the sites and branches, BSMs may be used [28–31].

2.5. Functional Difference Analyses

The DIVERGE 3.0 tool was used to analyze the functional divergences and distance of the gene family members [32]. Type-I mutations result from functional deviations following gene duplication, subsequently causing distinct evolutionary rates to be shown by duplicate genes. The residues in duplicate genes that showed striking variations in amino acid properties across copies tested using type-II functional divergences [33,34]. Clusters *i* and *j* had a type-I functional distance equal to $dF(i,j) = -\ln(1 - \theta_{ij})$. Wang and Gu proved that the *dF* is additive under a scenario of independence, such that the functional branch dimension for a given gene cluster *x* is *bF*(*x*), and the functional branch dimension of clusters *i* and *j* is *bF*(*i*) plus *bF*(*j*). In the case of a known *ij*, the cluster branch of the functional branch length may be calculated using the normal least squares technique; if *bF* = 0, then the duplicate site of the gene has evolved at a pace very close to that of the ancestor gene [35].

3. Results

3.1. Phylogenetic and Sequence Analyses

We identified 222 MIKC-type MADS-box proteins from six *Rosaceae* species' whole genome data. Of those identified, 54, 27, 23, 69, 23, and 26 MIKC-type MADS-box genes could be detected in *Prunus dulcis* (PdMADS1-54), *Prunus salicina* (PsMADS1-27), *Prunus armeniaca* (PaMADS1-23), *Prunus persica* (PpMADS1-69), *Prunus mira* (PmiMADS1-23), and *Amygdalus nana* (AnMADS1-26), respectively. The amino acid lengths ranged from 162 aa to 670 aa.

Phylogenetic analysis revealed five distinct groups (clades I–V) within the MIKC-type MADS-box gene family. There was a significant difference in the number of genes across the five clades, with the largest and smallest clades (clade I and III) having 110 and 6, respectively (Table 1, Figure 1). The six species' genes were found in nearly every clade. The genes of *Prunus dulcis* and *Prunus persica* were the most abundant in clades I, IV and V. *Prunus persica* had most genes in clade II, but *Prunus armeniaca* was missing. The six species had the same amount of genes in clade III (Table 1, Figure 2). The results suggested that the MIKC-type MADS-box genes in *Rosaceae* are nearly identical, but serve distinct purposes.

Table 1.
 The MIKC-type MADS-box genes of six *Rosaceae* species.

Species	Common Name	MIKC-Type MADS-Box Gene Numbers	Clade				
			I	II	III	IV	V
<i>Amygdalus nana</i>	Short almond	26	14	3	1	1	7
<i>Prunus mira</i>	Light walnut	23	11	3	1	2	6
<i>Prunus persica</i>	Peach	69	26	9	1	5	28
<i>Prunus armeniaca</i>	Apricot	23	15	0	1	2	5
<i>Prunus salicina</i>	Plum	27	14	3	1	2	7
<i>Prunus dulcis</i>	Almond	54	30	3	1	7	13
Total		222	110	21	6	19	66

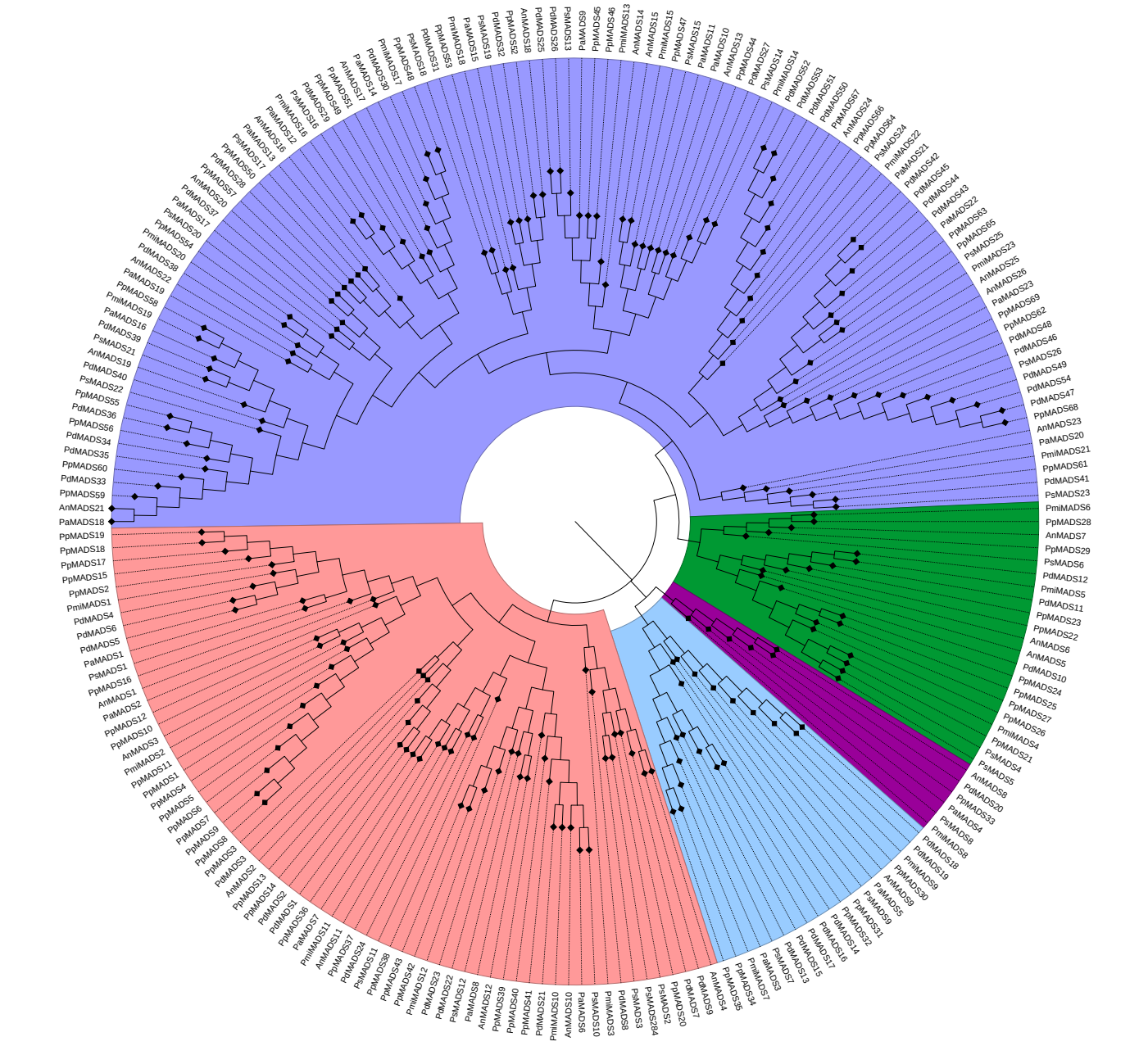


Figure 1.
 NJ phylogenetic tree of the MIKC-type MADS-box gene family from *Prunus mira* (PmiMADS), *Prunus armeniaca* (PaMADS), *Prunus persica* (PpMADS), *Prunus dulcis* (PdMADS), *Prunus salicina* (PsMADS), and *Amygdalus nana* (AnMADS) in *Rosaceae* plants. Blue is clade I, green is II, purple is III, light blue is IV, and pink is V.

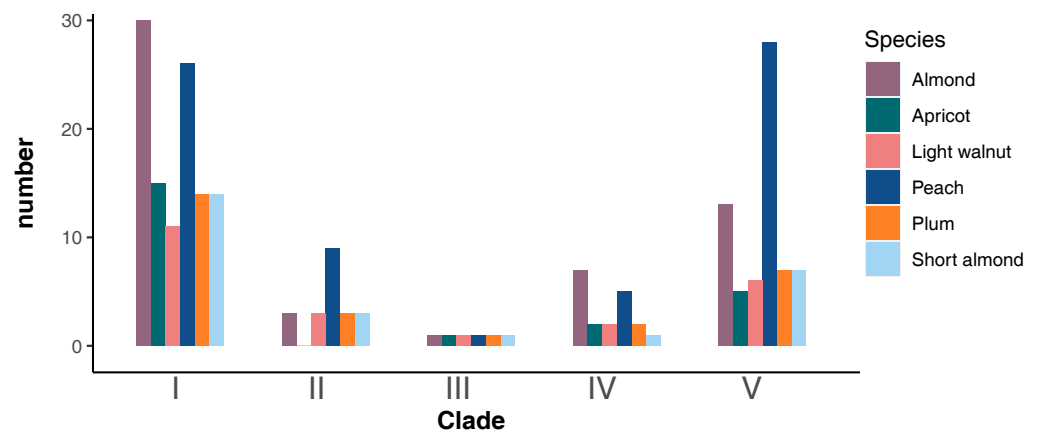


Figure 2. The distribution of MIKC-type MADS-box genes from six *Rosaceae* species in different clades. Species are represented by different colored boxes.

3.2. Motif and Structure Analyses

Ten conserved motifs, ranging from 21 to 80 amino acids, were found in this investigation. Only Motif 1 was found in all of the MIKC-type MADS-box proteins (Figure 3), except PdMADS42 and PmiMADS16. Furthermore, MIKC-type MADS-box proteins had substantially identical sequences for 10 amino acid residues (alanine, cysteine, glutamic acid, glycine, lysine, leucine, glutamine, arginine, serine, and threonine) (Figure 4). On the other hand, some MIKC-type MADS-box proteins lacked certain motifs; for example, Motifs 1 and 6 did not exist in PmiMADS16, and Motifs 4 and 6 did not exist in AnMADS23. However, there were similar or identical arrangement orders and motif types within the same clades. The results showed that the six *Rosaceae* species' MIKC-type MADS-box proteins were relatively conserved within a specific clade; the structural analysis produced similar results.

The results of the exon-intron structure analysis showed that the amino acid lengths ranged from 163 to 671. Among them, PsMADS3, PsMADS4, PsMADS7, PsMADS8, and PsMADS24 had no introns (Figure 5). The remaining MIKC-type MADS-box genes had 2–42 exon and 1–41 intron structures. Genes with 8 exons accounted for 47.75 percent of the total (106/222), most genes contained 7 or 8 exons, and only one gene contained 42 exons. This proves that the structures are specialized after gene duplication. These genes also have somewhat complicated architectures, including lengthy initial exons. In conclusion, the conserved motifs, gene architectures, and evolutionary connections of the MIKC-type MADS-box proteins were quite similar. Additionally, the diversification of gene roles was suggested due to the differences in the retained motifs and gene architectures among clades.

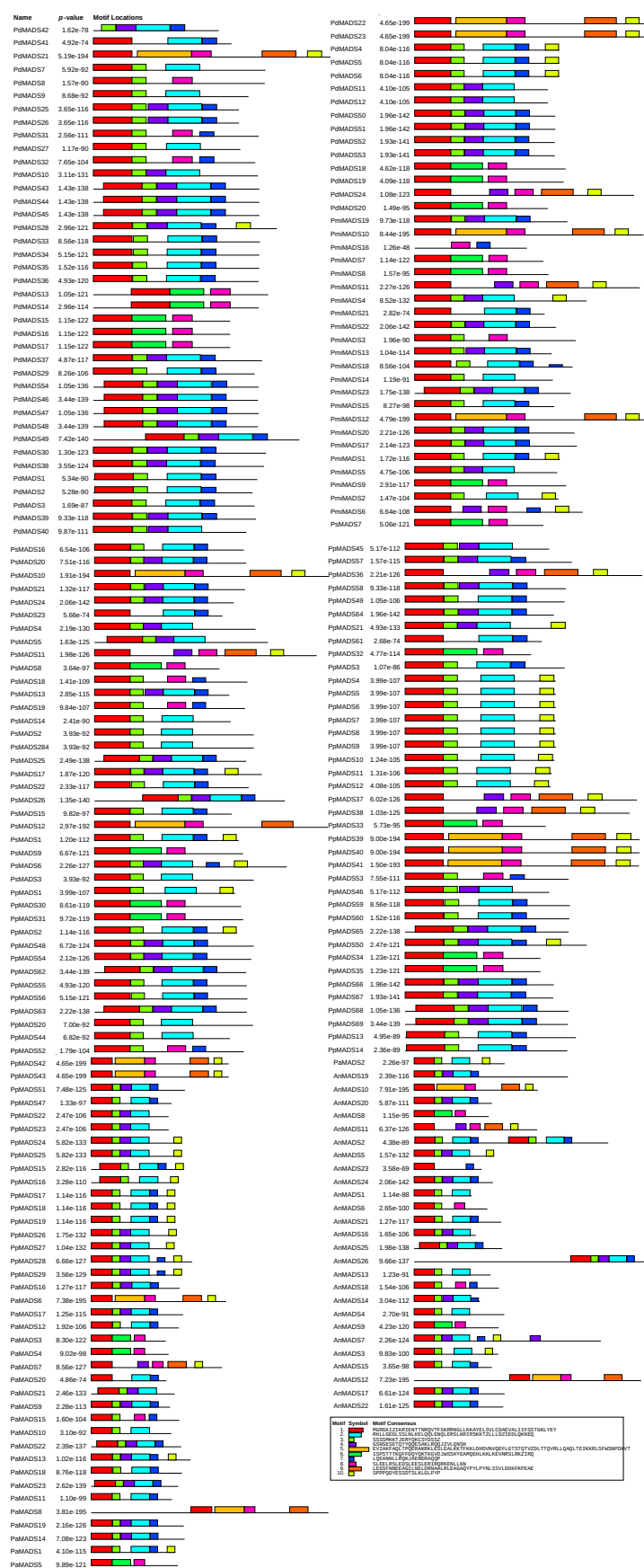


Figure 3. The conserved motifs of 222 MIKC-type MADS-box family members. Motifs 1 to 10 are represented by different colored boxes.

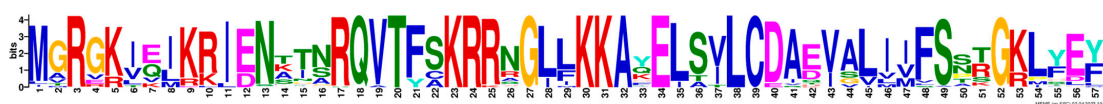


Figure 4. Sequence logo of Motif 1. Amino acid residues are represented by different colored letters.

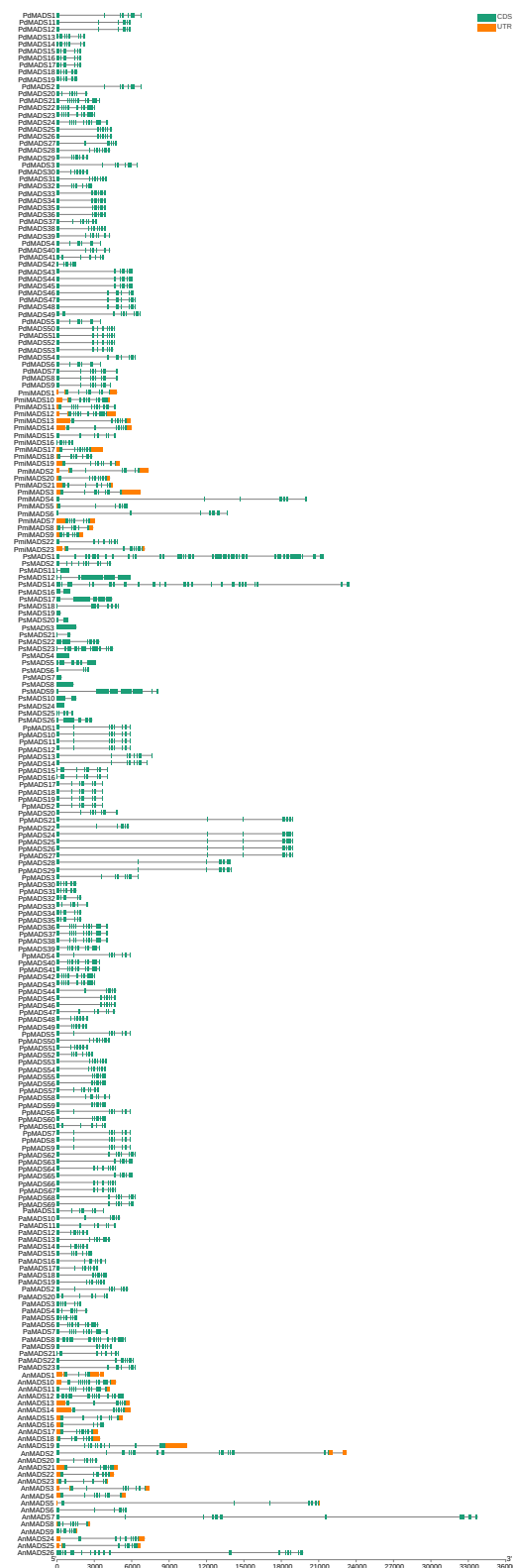


Figure 5. Gene structure of 222 MIKC-type MADS-box gene family members. The green boxes represent exons, and the gray lines represent introns.

3.3. Selective Pressure Analysis

We used EasyCodeML to analyze the selective pressure, used the branch model to test the evolutionary rate difference, and used the branch-site model to increase the detection of positive selection sites in the foreground branches. The results showed that the genetic differentiation had no significant differences ($p > 0.05$), and no positive sites were obtained according to the branch model and branch-site model results (Tables 2 and 3). Thus, we assumed the same selection pressure for each point and used the site model for testing the selection site positivity in these sequences. We found that most sites were subjected to purifying and neutral selections, and only the M8 model identified one site as having a significant positive selection site (Table 4). Although we obtained the positive selection sites, the MIKC-type MADS-box genes mainly underwent purifying selection events.

Table 2. Branch model parameters of selective pressure analysis.

Model	np/aa	ln L	Parameters Estimates			Compared Model	p-Value
Two ratio Model 2	448	−3172.993	ω :	$\omega_0 = 1.533$	$\omega_1 = 0.768$; $\omega_2 = 999.000$; $\omega_3 = 0.862$; $\omega_4 = 2.208$; $\omega_5 = 2.452$	Model 0 vs. Two ratio Model 2	>0.05
Model 0	443	−3175.884	$\omega =$	1.558			

Table 3. Branch-site model parameters of selective pressure analysis.

Model	np/aa	ln L	Parameters Estimates				Compared Model	p-Value	Positive Sites
			Site class	0	1	2a	2b		
Model A	446	−3166.858	f	0.000	0.000	0.156	0.844	Model A vs. Model A null	Not allowed
			ω_0	0.229	1.000	0.229	1.000		
			ω_1	0.229	1.000	999.000	999.000		
Model A null	445	−3165.206	1						Not allowed

Table 4. Site model parameters of selective pressure analysis.

Model	np	ln L	Parameters Estimates	Compared Model	p-Value	Positive Selected Sites
M ₀	443	−3144.848	$\omega = 0.318$	M ₀ vs. M ₃	<0.001	Not allowed
M ₃	447	−3054.877	p : 0.266, 0.519, 0.216 ω : 0.095, 0.432, 0.955			Not allowed
M _{1a}	444	−3071.511	p : 0.330, 0.670 ω : 0.193, 1.000	M _{1a} vs. M _{2a}	<0.001	Not allowed
M _{2a}	446	−3062.176	p : 0.289, 0.528, 0.183 ω : 0.204, 1.000, 2.045			Not allowed
M ₇	444	−3055.573	$p = 0.760$, $q = 1.042$	M ₇ vs. M ₈	>0.05	Not allowed
M ₈	446	−3053.688	$p_0 = 0.778$ $p = 0.852$			11 E 0.616

3.4. Functional Divergence Analysis

Gene families comprise multiple homologous genes with similar structures but different functions. In order to see whether changes in amino acid sequences within the MIKC-type MADS-box genes led to adaptive functional variation, we investigated the functional divergence of five distinct groups. The results suggest that, except for clades I/V, most type I coefficient θ values across the clades were statistically > 0 , which indicates

the presence of type I functional divergences across the clades. Six clades (I/II, I/III, II/III, II/IV, III/IV, and III/V) were remarkable ($p < 0.001$) for type I functional divergence. Clades III/V had the highest functional divergence coefficient θ ($= 1.150$), whereas clades I/V had the lowest ($= -0.002$). Other clades' type I θ values also varied, indicating that their rates of evolution were unique. In addition, type I functionally divergent sites were not detected in clades I/II, I/V, II/V, III/V, and IV/V (Table 5).

Table 5. Functional divergence analysis of MIKC-type MADS gene family in *Rosaceae*.

Clade	$\theta \pm SE$		MFE z-Scores	p -Value		Amino Acid Position with $Q(k) > 0.9$
	Type I	Type II		Type I	Type II	
I/II	0.579 ± 0.156	-0.112 ± 0.189	-4.115715	<0.001	>0.05	—
I/III	0.871 ± 0.158	0.005 ± 0.182	-6.191108	<0.001	>0.05	193,299,307,316,332,341,343, 346,359,370,371,374,386,397,400,403,407
I/IV	0.327 ± 0.225	-0.081 ± 0.207	-1.505518	>0.001	>0.05	190,301,302,310,311,316,346,349,361, 388,403
I/V	-0.002 ± 0.263	-0.660 ± 0.379	0.009046	>0.001	>0.05	—
II/III	1.058 ± 0.159	0.299 ± 0.050	-7.305373	<0.001	>0.05	190,192,193,195,272,273,290,292,295, 297,298,299,300,301,303,305,307,314, 316,318,320,321,332,342,343,345,347, 349,356,358,359,366,370,371,373,374, 385,386,387,395,396,397,403,404,405, 406,409,411
II/IV	0.874 ± 0.239	0.193 ± 0.070	-3.826513	<0.001	>0.05	193,271,272,290,296,298,299,301,302, 305,310,311,313,316,319,320,331,341, 347,350,352,361,370,371,373,382,388, 394,397,403,409
II/V	0.831 ± 0.288	-0.221 ± 0.203	-2.974702	>0.001	>0.05	—
III/IV	1.085 ± 0.232	0.232 ± 0.063	-4.882459	<0.001	>0.05	190,193,195,271,292,296,301,302,307, 310,311,313,314,319,321,331,332,341, 343,345,347,349,350,352,358,359,361, 374,382,386,387,388,394,395,404,406
III/V	1.150 ± 0.279	-0.233 ± 0.205	-4.220686	<0.001	>0.05	—
IV/V	0.205 ± 0.410	-0.425 ± 0.248	-0.50314	>0.001	>0.05	—

3.5. Functional Distance Analysis

Type I functional divergence-related cluster analyses were unable to determine which groups of duplicated genes were hampered by functional bottlenecks. Therefore, we compared the bF values across the five clades and found that they changed, indicating that unique roles evolved in the MIKC-type MADS-box gene family throughout time. Furthermore, the MIKC-type MADS-box genes tended to be conserved, as shown by bF values above 0. The lowest bF value (-2.453) was found in clade III, suggesting that the functions of the genes in this clade were more conserved and may have kept more of their original forms during evolution. With the longest functional branch length (0.405), clade I likely underwent considerable alterations to its functional restrictions in order to fulfill its specialized functional purpose (Table 6).

Table 6. Type I functional distance analysis of MIKC-type MADS gene family in *Rosaceae*.

Clade	bF
I	0.405
II	-1.270
III	-2.453
IV	-0.801
V	-0.403

4. Discussion

The MADS-box gene family plays an essential role in plant reproductive growth, and it also provides a research hotspot. However, most studies on the MADS-box gene family in *Rosaceae* mainly focus on identification and functional analysis in only one species [36–39]. In order to understand the variation pattern of MADS-box gene structure and function among different species, our study performed MIKC-type MADS-box gene identification between six key species in *Rosaceae*, and investigated their evolutionary relationship. These results are of great significance in understanding the MADS-box family function.

In the present study, we identified 222 MIKC-type MADS-box genes within six *Rosaceae* species. Among them, there were 54, 27, 23, 69, 23, and 26 genes in *Prunus dulcis*, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, *Prunus mira*, and *Amygdalus nana*, respectively, which was almost consistent with other findings in *Rosaceae* [36,37]. However, the number of MIKC-type MADS-box family members was quite different from other families. For instance, comparing with 232 members in legumes (92 genes in *Glycine max*, 45 genes in *Medicago truncatula*, 50 genes in *Vitis vinifera*, and 45 genes in *Arabidopsis thaliana*), the number was generally smaller in *Rosaceae* [40]. In the studies of conserved motifs and structure analysis, we found that no motif was shared by all MIKC-type MADS-box proteins, however Motif 1 was shared by almost all proteins, with the exception of PdMADS42 and PmiMADS16. These results suggested that Motif 1 was conserved across members of this gene family, and that PdMADS42 and PmiMADS16 had distinct purposes. Additionally, different species had varied sequence lengths, although the exon number and length were highly similar. Identified from *Prunus salicina*, the PsMADS3, PsMADS4, PsMADS7, PsMADS8, and PsMADS24 had no introns, and the gene annotation did not determine the intron structure. The result showed that intron-involving alleles or inversion might have deleted these introns due to homologous recombination, after the reverse transcription of mature mRNA [41].

Phylogenetic analysis revealed five distinct clades within the MIKC-type MADS-box gene family of the six *Rosaceae* species. The six species genes could be found in almost every clade, but *Prunus armeniaca* did not exist in clade II. The results showed that the six species could not be distinguished by them, and the genes in clade II might have some special functions which were missing in *Prunus armeniaca*. According to previous research, we found that *Prunus armeniaca* had some differences in the phenotypic characteristics of its flowers and fruit compared with the other five species. For instance, the petal was round-to-obovate, and the calyx tube was cylindrical with short fluff at base; moreover, the florescence and fruit stage were relatively early in *Prunus armeniaca* [42]. With further analysis and verification, we may be able to obtain several specific genes associated with these traits.

Using three models (branch, branch-site, and site models) to perform selective pressure analysis of the six species, we obtained only one positive site in the site model, and there were no positive sites in the branch and branch-site models. It suggested that various places undergo various selection forces, but the evolution and function of MIKC-type MADS-box genes were conserved, and purifying selection limited the MIKC-type MADS-box gene family in *Rosaceae* [43,44]. Inadequate site-related positive selection data may have been eliminated because they were reduced by the vast majority of the remainder of sites being under the evolutionary condition of neutral or purifying selection [45]. In addition, we found the results of the branch and branch-site models indicated no significant differences, which might be due to the six species being sibling species, or the conformity of the genomes and annotation level among the six species; nevertheless, further analysis is required to clearly understand the reasons.

Functional divergence analysis may be utilized to outline the putative residues of amino acids that cause the functional diversification of gene families, which occurs as a result of gene duplication. In this investigation, we found Type I functionality was shown to vary significantly among clades, and obtained locations that had functional divergence ($Q(k) > 0.90$). It suggested that the functional divergence of the current *Rosaceae* gene

family was mostly attributable to the differing proportions of evolution following genetic duplications. In addition, members of clade III kept more of their original roles throughout evolution, whereas those of clade I specialized theirs to serve a specific purpose. In future, we can research the gene functions of clade I to analyze their different functions throughout their evolution.

In order to better study the gene mutations in *Rosaceae*, our research facility has assembled the genome-wide sequence of *Amygdalus nana* for the first time. *Amygdalus nana* is a treasured wild species of the Tertiary period deciduous forests in the ancient Mediterranean [46]. At present, there are only several natural populations in China, Kazakhstan, and Russia, which are listed as endangered plants in China [47]. The kernel of *Amygdalus nana*, a vital species of non-timber forestry, has an oil content of 51.1%, similar to that of almond and higher than that of oil crops, such as rapeseed, cottonseed, and soybean [48]. In addition, having a strong ability to adapt to an extremely harsh environment, especially with cold and drought tolerance, it can grow in the hilly and mountainous areas at -35°C and annual precipitation of 50mm. Therefore, further research on *Amygdalus nana* may shed light on the evolution of and relationships between *Rosaceae* species.

5. Conclusions

In this study, the evolution of MIKC-type MADS-box gene family members in six *Rosaceae* plants was analyzed for the first time. 222 MIKC-type MADS-box family genes were identified from *Prunus dulcis*, *Prunus salicina*, *Prunus armeniaca*, *Prunus persica*, *Prunus mira*, and *Amygdalus nana*. The results of sequence and phylogenetic analyses showed that these genes were divided into five groups, the six species genes could be found in almost every clade, but *Prunus armeniaca* did not exist in clade II. Using selective evolutionary pressure and functional difference analyses, a number of positive selection sites and functional divergence locations were obtained. In addition, the conservation between different clades was evaluated, and it was concluded that the genes of clade I were specialized to serve a specific purpose. These results provide great help in the further mining of functional genes for key traits.

Author Contributions: Formal analysis, Y.Q., L.W. and C.C.; data curation, Y.Q., F.L. and H.Z.; writing—original draft preparation, Y.Q. and G.Z.; writing—review and editing, H.Z. and F.L.; visualization, L.W. and C.C.; project administration, H.Z. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Key Research and Development Program of Xinjiang Uygur Autonomous Region (2022292937) and the National Key Research and Development Program of China (2022YFD2200400).

Data Availability Statement: Required data might be provided following a proper request from the senior author.

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

References

1. Wang, Y.; Mu, Y.X.; Wang, J. Research progress of floral development regulation by MADS-box gene family. *Acta Agric. Zhejiangensis* **2021**, *33*, 1149–1158.
2. Kim, S.H.; Mizuno, K.; Fujimura, T. Isolation of MADS-box genes from sweet potato [*Ipomoea batatas* (L.) Lam.] expressed specifically in vegetative tissues. *Plant Cell Physiol.* **2002**, *43*, 314–322. [[CrossRef](#)] [[PubMed](#)]
3. Zhao, H.B.; Jia, H.M.; Wang, Y.; Wang, G.Y.; Zhou, C.C.; Jia, H.J.; Gao, Z.S. Genome-wide identification and analysis of the MADS-box gene family and its potential role in fruit development and ripening in red bayberry (*Morella rubra*). *Gene* **2019**, *717*, 144045. [[CrossRef](#)]
4. Yanofsky, M.F.; Ma, H.; Bowman, J.L.; Drews, G.N.; Feldmann, K.A.; Meyerowitz, E.M. The protein encoded by the Arabidopsis homeotic gene *agamous* resembles transcription factors. *Nature* **1990**, *346*, 35–39. [[CrossRef](#)] [[PubMed](#)]
5. Schwarzsommer, Z.; Hue, I.; Huijser, P.; Flor, P.J.; Hansen, R.; Tetens, F.; Lonig, W.E.; Saedler, H.; Sommer, H. Characterization of the *antirrhinum* floral homeotic MADS-box gene *deficiens*—evidence for DNA-binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* **1992**, *11*, 251–263. [[CrossRef](#)]

6. Dreni, L.; Zhang, D. Flower development: The evolutionary history and functions of the AGL6 subfamily MADS-box genes. *J. Exp. Bot.* **2016**, *67*, 1625–1638. [[CrossRef](#)]
7. Alvarez-Buylla, E.R.; Pelaz, S.; Liljegren, S.J.; Gold, S.E.; Burgeff, C.; Ditta, G.S.; Pouplana, L.R.D.; Martinez-Castilla, L.; Yanofsky, M.F. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5328–5333. [[CrossRef](#)]
8. Riechmann, J.L.; Meyerowitz, E.M. MADS domain proteins in plant development. *Biol. Chem.* **1997**, *378*, 1079–1101.
9. Zhao, X.Y.; Xian, D.Y.; Song, M.; Tang, Q.L. Research progress of MIKC-type MADS-box protein regulation on flowering. *Biotechnol. Bull.* **2014**, *7*, 8–15.
10. Susanne, S.; Alice, K.; Sirui, P.; Lars, S.J.; Rainer, M. Genome-wide analysis of MIKC-type MADS-box genes in wheat: Pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* **2020**, *225*, 511–529.
11. Chunmei, H.; Can, S.; Jaime, A.T.S.; Li, M.Z.; Duan, J. Genome-wide identification and classification of MIKC-type MADS-box genes in *Streptophyte* lineages and expression analyses to reveal their role in seed germination of orchid. *BMC Plant Biol.* **2019**, *19*, 223.
12. Jiang, S.C.; Pang, C.Y.; Song, M.Z.; Wei, H.L.; Fan, S.-I.; Yu, S.X. Analysis of MIKCC-type MADS-box gene family in *Gossypium hirsutum*. *J. Integr. Agric.* **2014**, *13*, 1239–1249. [[CrossRef](#)]
13. Gao, W.; Zhang, L.; Xue, C.L.; Zhang, Y.; Liu, M.J.; Zhao, J. Expression of E-type MADS-box genes in flower and fruits and protein interaction analysis in Chinese Jujube. *Acta Hort. Sin.* **2022**, *49*, 739–748.
14. Huang, H.J.; Wang, S.; Jiang, J.; Liu, G.; Li, H.; Chen, S.; Xu, H. Overexpression of BpAP1 induces early flowering and produces dwarfism in *Betula plantyphylla* × *Betula pendula*. *Physiol. Plant* **2014**, *151*, 495–506. [[CrossRef](#)] [[PubMed](#)]
15. Liu, S.; Qi, T.; Ma, J.; Ma, L.; Liu, X. Cloning and functional analysis of SEP-like gene from *Phyllostachys violascens*. *J. Nucl. Agric. Sci.* **2016**, *30*, 1453–1459.
16. Parenicová, L.; Folter, S.D.; Kieffer, M.; Horner, D.S.; Colombo, L. 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](#)]
17. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L. Sequencing of allotetraploid cotton(*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [[CrossRef](#)]
18. Prakash, A.; Jeffries, M.; Bateman, A.; Finn, R.D. The HMMER Web server for protein sequence similarity search. *Curr. Protoc. Bioinform.* **2017**, *60*, 3–15. [[CrossRef](#)]
19. Katoh, K.; Standley, D.M. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **2016**, *32*, 1933–1942. [[CrossRef](#)]
20. Timothy, L.B.; James, J.; Charles, E.G.; William, S.N. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49.
21. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)] [[PubMed](#)]
22. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. Prottest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [[CrossRef](#)] [[PubMed](#)]
23. Sudhir, K.; Glen, S.; Michael, L.; Christina, K.; Koichiro, T. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
24. Yu, G. Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinform.* **2020**, *69*, e96. [[CrossRef](#)]
25. Yu, G.; Lam, T.T.Y.; Zhu, H.; Guan, Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol. Biol. Evol.* **2018**, *35*, 3041–3043. [[CrossRef](#)]
26. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.Y. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. [[CrossRef](#)]
27. Gao, F.; Chen, C.; Arab, D.A.; Du, Z.; He, Y.; Ho, S.Y.W. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evol.* **2019**, *9*, 3891–3898. [[CrossRef](#)]
28. Yang, Z.H. PAML4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
29. Yang, Z.H.; Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **1998**, *46*, 409–418. [[CrossRef](#)]
30. Yang, Z.H.; Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [[CrossRef](#)]
31. Yang, Z.H.; Wong, W.S.; Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [[CrossRef](#)] [[PubMed](#)]
32. Gu, X.; Zou, Y.Y.; Su, Z.X.; Huang, W.; Zhou, Z.; Arendsee, Z.; Zeng, Y.W. An update of DIVERGE software for functional divergence analysis of protein family. *Mol. Biol. Evol.* **2013**, *30*, 1713–1719. [[CrossRef](#)] [[PubMed](#)]
33. Gu, X.; Velden, K.V. DIVERGE: Phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* **2002**, *18*, 500–501. [[CrossRef](#)]
34. Gu, X. A simple statistical method for estimating type-II(cluster-specific) functional divergence of protein sequences. *Mol. Biol. Evol.* **2006**, *23*, 1937–1945. [[CrossRef](#)] [[PubMed](#)]
35. Wang, Y.F.; Gu, X. Functional divergence in the caspase gene family and altered functional constraints: Statistical analysis and prediction. *Genetics* **2001**, *158*, 1311–1320. [[CrossRef](#)]

36. Christina, E.W.; Elisa, V.; Sergio, J.T.; Lgnazio, V.; Douglas, G.B. A genome-wide analysis of MADS-box genes in peach [*Prunus persica* (L.) Batsch]. *BMC Plant Biol.* **2015**, *15*, 41.
37. Hisayo, Y.; Tomomi, O.; Hiroaki, J.; Yukari, H.; Ryuta, S.; Ryutaro, T. Expressional regulation of PpDAM5 and PpDAM6, peach (*Prunus persica*) dormancy-associated MADS-box genes, by low temperature and dormancy-breaking reagent treatment. *J. Exp. Bot.* **2011**, *62*, 3481–3488.
38. Liu, X.Y.; He, Z.J.; Qiu, Y.M. Screening and Bioinformatics analysis of almond MADS-box Gene family. *Mol. Plant Breed.* **2022**, *20*, 1477–1486.
39. Chen, C.; Zhu, G.P.; Zhao, H.; Liu, H.M.; Luo, Y.; Xu, W.Y.; Huang, M.Z.; Wu, Y.T.N.; Wang, L. Genome-wide identification of MADS-box Gene family and expression analysis in *Prunus sibirica*. *Mol. Plant Breed.* **2020**, *18*, 6575–6585.
40. Zhang, Y.; Wang, J.; Yu, Z.; Xu, Q.; Zhang, L.; Pan, Y. Bioinformatics analysis of MIKC-type MADS-box gene family in legumes. *Chin. J. Oil Crop Sci.* **2022**, *44*, 798–809.
41. Wu, K.L.; Guo, Z.J.; Wang, H.H.; Li, J. The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res.* **2005**, *12*, 9–26. [[CrossRef](#)] [[PubMed](#)]
42. Yv, T.; Lu, L.; Ku, T.; Li, C.; Chen, S. *Flora Reipublicae Popularis Sinicae*; Science Press: Beijing, China, 1986; Volume 38, pp. 11–40.
43. Vallender, E.J.; Lahn, B.T. Positive selection on the human genome. *Hum. Mol. Genet.* **2004**, *13*, R245–R254. [[CrossRef](#)] [[PubMed](#)]
44. Zhang, J.Q.; Zhu, K.Y.; Shi, X.H. Adaptive evolution and identification analysis of the MADS-box gene family in *Paeonia lactiflora*. *Mol. Plant Breed.* **2019**, *17*, 6959–6966.
45. Tian, Y.X.; Wang, Q.G.; Zhang, H.; Zhou, N.N.; Yan, H.J.; Jian, H.Y.; Li, G.S.; Tang, K.X.; Qiu, X.Q. Genome-wide identification and evolutionary analysis of MLO gene family in Rosaceae plants. *Hortic. Plant J.* **2022**, *8*, 110–122. [[CrossRef](#)]
46. Ma, S.; Wang, C.; Sun, F.; Wei, B.; Nie, Y. Genetic diversity of an endangered plant *Amygdalus ledebouriana* in Xinjiang. *Sci. Silvae Sin.* **2019**, *9*, 71–80.
47. Yan, G.; Xu, Z. Study on the wild fruit tree diseases of Tianshan mountains and their distribution in Xinjiang. *Arid. Zone Res.* **2001**, *18*, 47–49.
48. Wu, Y.M.; Wu, Y.A.; Wu, Y.X.; Cao, Z. The advances of the studies on almond (*Amygdalus communis* L.) A literature review. *J. Gansu Agric. Univ.* **1996**, *1*, 86–92.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.