Genome-wide identification, evolution, and expression characterization of the pepper MADS-box gene family

zhicheng Gan[1], xingxing Wu[1], Tingting Feng[1], xiaomin Lu[1], nengbing Hu[1], ruining LI[1], and xianzhong Huang[1]

[1]Anhui Science and Technology University

November 23, 2022

Abstract

MADS domain transcription factors play roles throughout the whole lifecycle of plants from seeding to flowering and fruit-bearing. However, systematic research into MADS-box genes of the economically important vegetable crop pepper (Capsicum spp.) is still lacking. We identified 174, 207, and 72 MADS-box genes from the genomes of C. annuum, C. baccatum and C. chinense, respectively. These 453 MADS-box genes were divided into type I (Mα, Mβ, Mγ) and type II (MIKC* and MIKCC) based on their phylogenetic relationships. Collinearity analysis identified 144 paralogous genes and 195 orthologous genes in the three pepper species, and 70, 114, and 10 MADS-box genes specific to C. annuum, C. baccatum and C. chinense, respectively. Comparative genomic analysis highlighted functional differentiation among homologous MADS-box genes during pepper evolution. Tissue expression analysis revealed three main expression patterns: highly expressed in roots, stems, leaves and flowers (CaMADS93/CbMADS35/CcMADS58); only expressed in roots; and specifically expressed in flowers (CaMADS26/CbMADS31/CcMADS11). This study provides the basis for an in-depth study of the evolutionary features and biological functions of pepper MADS-box genes.

Hosted file

`essoar.10512212.1.docx` available at https://authorea.com/users/542768/articles/601054-genome-wide-identification-evolution-and-expression-characterization-of-the-pepper-mads-box-gene-family

Core Ideas:

- Genome-wide identification identified 454 *MADS-box* genes from *Capsicum* spp. Genomes.
- Genome-wide identification revealed the evolutionary relationship of MADS-box gene family in Capsicum spp.
- Most of the Capsicum *MADS-box* genes have been amplified on a large scale.
- Most of the Capsicum *MADS-box* genes have undergone purification selection during evolution.
- Tissue specific expression patterns of most *MADS-box* genes indicated their functional diversity.

Genome-wide identification, evolution, and expression characterization of the pepper MADS-box gene family

Zhicheng Gan, Xingxing Wu, Tingting Feng, Xiaoming Lu, Nengbing Hu, Ruining Li, Xianzhong Huang

College of Agriculture, Anhui Science and Technology University, Center for Crop Biotechnology, Fengyang 233100, China

Corresponding author:

College of Agriculture, Anhui Science and Technology University, Center for Crop Biotechnology, 233100 Fengyang, China

Email: huangxz@ahstu.edu.cn

College of Agriculture, Anhui Science and Technology University, Center for Crop Biotechnology, 233100 Fengyang, China

Email: lirn@ahstu.edu.cn

Abbreviations: AG, AGAMOUS; AGL, AGAMOUS-like; AP1. APETALA1; AP3, APETALA3; FLC, FLOWING LOCUS C; Ka, non-synonymous substitution rate; Ks, synonymous substitution rate; MADS, MCM1, AGAMOUS, DEFICIENS and SRF; MAF: MADS AFFECTING FLOWERING; PI, PISTILATA; qRT-PCR, Quantitative Real-Time PCR; RNA-seq, RNA sequencing; SOC1, SUPPRESSOR Of SCO COSTANS 1; SEP, SEPALLATA; SHP1, SHATTERPROOF1; STK, SEEDSTICK; SVP, SHORT VEGETATIVE PHASE.

ABSTRACT

MADS domain transcription factors play roles throughout the whole lifecycle of plants from seeding to flowering and fruit-bearing. However, systematic research into *MADS-box* genes of the economically important vegetable crop pepper (*Capsicum* spp.) is still lacking. We identified 174, 207, and 72 *MADS-box* genes from the genomes of *C. annuum*, *C. baccatum* and *C. chinense*, respectively. These 453 *MADS-box* genes were divided into type I (M , M , M) and type II (MIKC* and MIKC^{C}) based on their phylogenetic relationships.

Collinearity analysis identified 144 paralogous genes and 195 orthologous genes in the three pepper species, and 70, 114, and 10 *MADS-box* genes specific to *C. annuum*, *C. baccatum* and *C. chinense*, respectively. Comparative genomic analysis highlighted functional differentiation among homologous *MADS-box* genes during pepper evolution. Tissue expression analysis revealed three main expression patterns: highly expressed in roots, stems, leaves and flowers (*CaMADS93*/*CbMADS35*/*CcMADS58*); only expressed in roots; and specifically expressed in flowers (*CaMADS26*/*CbMADS31*/*CcMADS11*). This study provides the basis for an in-depth study of the evolutionary features and biological functions of pepper *MADS-box* genes.

Keywords: *Capsicum annuum*; *Capsicum baccatum*; *Capsicum chinense*; MADS-box family; Comparative evolution; Purifying selection

1. INTRODUCTION

MADS-box genes comprise a large family of genes distributed throughout the plant kingdom and therefore occupy an important position in plant growth and development. The MADS acronym is composed of the initials of four proteins: a yeast transcription factor (MCM1), the *Arabidopsis thaliana* floral homozygote AGAMOUS (AG), an *Antirrhinum majus* floral homozygote (DEFICIENS), and human serum response factor (SRF) (Lawton-Rauh et al., 2000; Shore et al., 1995). All four proteins have a highly conserved region consisting of 56–58 amino acids called the MADS domain.(Becker & Theissen, 2003; Schwarz-Sommer et al., 1990) Approximately one billion years ago, a duplication event occurred in the common ancestor of *MADS-box* genes, resulting in two distinct branches, type I and type II (Alvarez-Buylla et al., 2000), Type I proteins contain mainly SRF structural domains, and type I MADS-box genes can be further divided into M , M , and M ; only a few types I genes have a biological function (Smaczniak et al., 2012). Type II genes are divided into MIKC^{C} and MIKC* subtypes based on their structural features (Parenicová et al., 2003).

Replication and evolution of the type I MADS-box gene*s* appear to be faster than those of type II genes, but these observations are based on few studies, mainly on the function of type II MADS-box gene in flower development (Grimplet et al., 2016). According to the classical model of flower development "ABCDE": class A genes regulate the formation of sepals; class A and B genes together regulate petals; class B and C genes control the differentiation of stamens; class C and D genes are mainly involved in the formation of ovules; class E genes are involved in the regulation of flower organs during the formation process (Ferrario et al., 2004; Theissen et al., 1996; Weigel & Meyerowitz, 1994). In *A. thaliana*, *APETALA1* (*AP1*) represents a class A gene (Irish & Sussex, 1990); *APTALA3* (*AP3*) and *PISTILATA* (*PI*) genes belong to class B (jack et al., 1992); *AGAMOUS* (*AG*) is a representative gene with class C function (Mizukami & Ma, 1992); *SEPALLATA* (*SEP*) genes belong to class E (Pelaz et al., 2000), including *SEP1*, *SEP2*, *SEP3* and *SEP4*; and class D genes *SEEDSTICK* (*STK*) and *SHATTERPROOF1* (*SHP1*) (Favaro et al., 2003). In addition, the functions of some genes regulating flowering time, such as *FLOWING*

LOCUS C (*FLC*), *SHORT VEGETATIVE PHASE* (*SVP*), and *SUPPRESSOR Of OVEREXPRESSION Of COSTANS 1* (*SOC1*), have been confirmed in *A. thaliana*, rice (*Oryza sativa*), and wheat (*Triticum aestivum*) (Lee et al., 2008; Shimada et al., 2009; Swiezewski et al., 2009).

MADS-box genes are involved in many plant growth and development processes. *MADS-box* genes play important regulatory roles in fruit growth and development, such as the *FOREVER YOUNG FLOWER* (*FYF*) gene, the SEP-type *CMB1* gene, and auxin-related *SlIAA9* in tomato (*Solanum lycopersicon*) (Molesini et al., 2020; Xie et al., 2014; Zhang et al., 2018); *PaMADS7* of cherry (*Cerasus pseudocerasus*) (Qi et al., 2020); *MA-MADS5* of banana (*Musa nana*) (Roy et al., 2012); and *VEGETATIVE TO PRODUCTIVE TRANSITION 2* in wheat (*Triticum aestivum*) (Adamski et al., 2021). *MADS-box* genes are also involved in the plant stress response, such as *Zymoseptoria tritici ZtRlm1*; *AtAGL16*; and *SiMADS51* in *Setaria italica* (Mohammadi et al., 2020; Zhao et al., 2020). Therefore, the MADS-box family is one of the driving factors of plant diversity and plays an important role in growth and development (Lai et al., 2022; Theissen & Saedler, 2001; Yamaguchi & Hirano, 2006)

Pepper (*Capsicum* spp.) originated in Central America and the Andes mountains, growing in tropical and temperate environments (Aguilar-Meléndez et al., 2009). At present, 27 species of *Capsicum* have been identified (Albrecht et al., 2012), with five species cultivated long-term: *Capsicum annuum*, *C. baccatum*, *C. chinese*, *C. frutescens* and *C. pubescens* (Troconis-Torres et al., 2012). Evolutionary analysis shows that *C. annuum* differentiated from *C. chinense* around 1.14 million years ago, and *C. baccatum* differentiated from *C. annuum* and *C. chinense* 1.7 million years ago (Kim et al., 2017). At present, there is little research on the MADS-box gene family in pepper. A *CaMADS-box* gene is positively involved positively in low-temperature, salt, and osmotic stress s signaling pathways (Chen et al., 2019). *CanMADS1* and *CanMADS6* genes are expressed in flower buds and fruits and are highly expressed at 2 days after flowering, suggesting involvement in regulating pepper fruit development (Sung et al., 2001). However, the members of the MADS-box gene family in pepper have not been systematically identified or analyzed. In this study, we carried out genome-wide identification of the MADS-box gene family from *C. annuum*, *C. baccatum*, and *C. chinense* genome data (Liao et al., 2022; Kim et al., 2017; Qin et al., 2014). Motif distribution, gene structure, chromosomal localization, comparative evolution, and expression analysis of different tissues were performed. Colinearity analysis revealed the presence of specific *MADS-box* genes in all three *Capsicum* species, and genomic duplication events in the MADS-box transcription factor family in *C. annuum* and *C. baccatum*. The results of this study provide comprehensive insight into the characterization of *MADS-box* genes and lay the theoretical and fundamental groundwork for revealing the functions of *MADS-box* genes in peppers and for the molecular breeding of peppers.

2. MATERIAL AND METHODS

2.1 Plant materials

The seeds of *C. annuum* were planted in pots containing soil: vermiculite: perlite (2:1:1) and placed in a growth chamber under long-day conditions (16 h^{-1} light/8 h^{-1} dark, 23/20 °C day/light, 150 µmol m^{-2} s^{-1}). For tissue expression analysis, roots, stems, leaves were collected at the third true-leaf expanding stage, flower, sepal, petals, stamens and pistils were harvested at flowering stage. Pepper fruits that grew to 3 cm long were sampled. All samples were immediately snap-frozen with liquid nitrogen and then stored in a -80°C refrigerator until RNA extraction.

2.2 Identification and naming of MADS-box family genes in three peppers

MADS-box protein sequences of *A. thaliana* were downloaded from the TAIR database (http://www.arabidopsis.org), and genomic data of *C. annuum*, *C. baccatum* and *C. chinense* were downloaded from the Pepper Genome Platform (http://peppergenome.snu.ac.kr/) (Kim et al., 2017). Algorithm-based BLASTP was performed using the MADS-box protein sequence of *A. thaliana* as the query in the protein databases of *C. annuum*, *C. baccatum* and *C. chinense*, with an *E*-value < 1e-5 and other parameters as default values. The obtained candidate protein sequences were compared with Pfam (http://pfam.xfam.org/) database using HMMER (http://www.hmmer.org/). The MADS-box domain based on SRF domain (PF01486) and K domain (PF00319) was used for further comparison and screening, and the parameters were the default values. Thus, the MADS-box gene family members of three species were identified and named according to the order of their position on the chromosome. In order to view the distribution of MADS-box on the chromosomes of three pepper more directly, the online website MG2C (http://mg2c.iask.iN/mg2c_v2.0/) was used to draw the chromosome location map. The theoretical molecular weights and isoelectric points of MADS-box proteins were computed by the ExPASy (https://web.expasy.org/protparam/) proteomics server. The subcellular localization of CaMADS-box, CbMADS-box and CcMADS-box proteins were predicted by the ProtComp 9.0 (http://liNux1.softberry.com/berry.phtml) server.

2.3 Phylogenetic tree construction, gene structure and protein motif analysis

The ClustalW program was used to perform multiple sequence alignments between the MADS-box gene family protein sequences of *C. annuum*, *C. baccatum*, *C. chinense* and *A. thaliana*, with the default parameters (et al., 2002). MEGA6.0 was used to construct Neighbor-Joining phylogenetic trees and analyze the evolutionary relationship of MADS-box gene families among different species (Tamura et al., 2013). The phylogenetic trees were visualized using the EvolView server (https://evolgenius.info/evolview-v2/#login) (Zhang et al., 2012). Analysis of *MADS-box* gene exon-intron distribution based on *C. annuum*, *C. baccatum* and *C. chinense* genome gff3 files using GSDS (http://gsds.cbi.pku.edu.cn/) visualization server. The conserved motifs of the CaMADS-box, CbMADS-box and CcMADS-box family were identified using

the MEME website (http://meme-suite.org/tools/meme). The motif length range was set to 10–60 amino acid residues, the maximum number of motif discoveries was set to 10 and other parameters were set to default values.

2.4 Selective pressure analysis

Paralogous and orthologous of *MADS-box* genes in *C. annuum*, *C. baccatum*, and *C. chinense* were identified using the Ortho Venn2 online website (https://orthovenn2.bioinfotoolkits.net/home) (Xu et al., 2019). DNaSP 6.0 software was used to calculate the non-synonymous substitution rate (Ka) and the synonymous substitution rate (Ks) (Rozas et al., 2017), and the selection pressure of replicated gene pairs during evolution was evaluated by calculating the ratio *Ka/Ks*. Ka/Ks > 1, < 1 or $= 1$ represent positive, negative or neutral evolution, respectively (Yadav et al., 2015)

2.5 Pepper MADS-box gene expression analysis and qPCR validation

RNA-seq data of *C. annuum*, *C. baccatum*, and *C. chinense* transcriptomic were obtained from the BioProject database (https://www.ncbi.nlm.nih.gov/bioproject), and the transcriptome accession number of the three *Capsicum* species was PRJNA223222, PRJNA308879 and PRJNA331024, respectively (Kim et al., 2017). The fastp (Chen et al., 2018) and RSEM tools (Li & Dewey, 2011) were used to filter and compare sequencing data, and the comparison was achieved using the Bowtie2 tool (Langmead & Salzberg, 2012). Parameters are set to default values. The results were standardized using the fragments per kilobase of transcript per million mapped reads (FPKM) of a gene. After the FPKM value was converted by log2(FPKM+1), a heat map was created using the TBtools software (Chen et al., 2020), and the expression of *CaMADS-box*, *CbMADS-box* and *CcMADS-box gene*s in different tissue was analyzed.

RNA from each tissue was extracted using TRIzol kit (Life Technologies, USA); reverse transcription was performed using HiScript III RT SuperMix for qPCR (Vazyme, Nanjing) kit; qRT-PCR analysis was performed using ChamQ Universal SYBR qPCR (Vazyme, Nanjing) reagent. The PCR instrument was an ABI ViiA7 real-time fluorescence quantitative PCR machine (Life Technologies, USA). The Primer 3.0 tools (https://bioinfo.ut.ee/primer3-0.4.0/) were used to design *CaMADS-box* gene-specific amplification primers, using *CaUBI3* as the reference gene (Wan et al., 2011) (Supplemental Table S8). The qRT-PCR primers are listed in Table S8. All qRT-PCR assays were performed using three independent biological replicates, each with three technical replicates. The PCR conditions and the calculation method of relative gene expression were the same as before (Jin et al., 2019).

4.6 Protein interaction network prediction

The type II MADS-box genes between *C. annuum* and *A. thaliana*, using the AraNetV2 (http://www.inetbio.org/aranet/) and the STRING (http://string-db.org/cgi) databases and the predicted protein-protein interaction network was displayed using Cytoscape software (Su et al., 2014).

3. RESULTS

3.1 Genome-wide identification and characterization of the MADS-box gene family from *Capsicum* spp.

We identified 174, 207, and 72 MADS-box gene family members from three *Capsicum* species: *C. annuum*, *C. baccatum*, and *C. chinense*, respectively (Supplemental Tables S1), by referring to the amino acid sequences of *A. thaliana* MADS-box proteins, local BLAST comparison, and screening using the Pfam (http://pfam.xfam.org/) website. The number of genes in *C. annuum* and *C. baccatum* was more than that in *C. chinense*. All 453 MADS-box proteins possessed conserved SRF and K domains; the genes encoding these proteins were named *CaMADS1-CaMADS174*, *CbMADS1-CbMADS207*, and *CcMADS1-CcMADS72*, respectively. Analysis of physicochemical properties of MADS-box proteins in the three pepper types found that the amino acid (aa) lengths of CaMADS-box, CbMADS-box, and CcMADS-box proteins were 59-661, 49-660, and 78-1100 aa, respectively (Supplemental Tables S1). Their molecular weights were 6813-74021.51, 5691.74-74516.8 and 9052.62-125758.3 KDa, respectively. Isoelectric points ranged from 4.51-10.94, 4.89-10.23 and 4.46-10, respectively (Supplemental Tables S1). Prediction of subcellular localization showed that most MADS-box proteins were located in the nucleus, with a few located outside the cellular (Supplemental Tables S1).

3.2 Categorization, structural classification, and structure of *MADS-box* genes in pepper

A phylogenetic tree was reconstructed using 105 *AtMADS-box* and 453 pepper *MADS-box* genes to further study the phylogenetic relationships of *MADS-box* genes. Following the classification and structure of the *A. thaliana* MADS-box family, we reconstructed two separate phylogenetic trees of type I (Supplemental Figure S1) and type II (Figure 1) genes, respectively. The results showed that type I MADS-box genes were divided into three subfamilies (M , M , and M). There were 99, 6, and 16 members of M , M , and M , respectively, in *C. annuum*, 134, 8, and 7 members, respectively, in *C. baccatum*, and 34, 3, and 6 members, respectively in *C. chinense* (Supplemental Figure S1). The M subfamily in *C. baccatum* was considerably expanded, while the M subfamily in *C. chinense* was substantially contracted. Type II MADS-box genes were further divided into 10 subfamilies: SEP, AGL6, AP1, STK/AG, AGL12, SOC1, SVP, PI/AP3, FLC/MAF, and MIKC*. The "ABCDE" genes of flower development, such as *SEP*, *AP1*, *AG*, *PI*, *AP3*, and *STK*, were amplified among the type II genes of the three pepper species (Supplemental Figure. 1, Supplemental Tables S1).

Conserved motifs of MADS-box family proteins were analyzed using the online website MEME. A total of 10 motifs with a length of 15–41 amino acids were predicted (Supplemental Table S2), and their distribution trends were conserved within every subfamily (Supplemental Figures S2A, S3A, S4A). Motifs 1 and 3 were common to most MADS-box transcription factors, but motif 6 was unique to type II protein members. Motif 5 was also a unique domain to MIKC-type

protein members.

Most *MADS-box* genes belonging to the same subfamily exhibited the same pattern of gene structure, but there were great differences among different members (Supplemental Figures S2B, S3B, S4B). Most type I genes were composed of one exon, but *CbMADS28*, *CbMADS81*, *CbMADS124* and *CcMADS47* had two exons. Meanwhile, type II genes contained multiple introns and exons. Compared with type II genes, type I genes had suffered intron loss. Differences in exon and intron structure between type I and type II genes may be one of the reasons for the increase in MADS-box gene family members during evolution.

3.3 Phylogenetic relationships of *MADS-box* genes in pepper

To investigate homologous *MADS-box* genes in pepper and possible gene duplication in each pepper species, we next identified the evolutionary relationships of *MADS-box* genes in the three *capsicum* species. There were 39 groups of orthologous genes, accounting for 22.9% (40/174) of *MADS-box* genes in *C. annuum*, 19.3% (40/207) of *MADS-box* genes in *C. baccatum,* and 55.5% (40/72) of *MADS-box* genes in *C. chinense* (Figure 2). This indicated that some *MADS-box* genes were preserved during evolution of *C. annuum* and *C. baccatum*, while *MADS-box* genes were highly conserved during the evolution of *C. chinense*. In addition to 40 groups of orthologous genes shared by the three pepper species (Supplemental Table S3), 193 *MADS-box* homologs were found between any two pepper species (Supplemental Table S4). There were 85 pairs of orthologous genes between *C. baccatum* and *C. annuum*, 61 pairs of orthologous genes between *C. annuum* and *C. chinense*, and 47 pairs of orthologous genes between *C. baccatum* and *C. chinense* (Supplemental Table S4). Furthermore, *C. annuum*, *C. baccatum,* and *C. chinense* possessed 70, 117, and 10 unique *MADS-box* genes, respectively (Figure 2). There were 140 pairs of paralogous *MADS-box* genes, of which 47 pairs were tandem repeats (Supplemental table S5). Among these duplicated genes, some displayed have one-to-many relationships, such as *CaMADS40*, which was the tandem repeat gene of both *CaMADS39* and *CaMADS41*. This was reflected in *C. baccatum*, with tandem repeat genes of *CbMADS20* including *CbMADS18*, *CbMADS17*, *CbMADS9*, *CbMADS8,* and *CbMADS191*. However, there were no one-to-many situations in *C. chinense*. This indicates that the MADS-box gene family of pepper has an obvious gene replication phenomenon, which explains why the number of *MADS-box* genes in *C. annuum* and *C. baccatum* is more than that in *C. chinense*.

To further explore the evolutionary mechanisms of *MADS-box* genes in pepper, we constructed collinear circos of the three *Capsicum* species based on orthologous genes and tandem repeat genes (Figure 3). The results revealed that *MADS-box* genes are distributed on every chromosome, mainly located at the terminus of each chromosome arm. Orthologous genes in the three *Capsicum* species were very close on the chromosome, anchored in a highly conserved collinearity block. Based on the genomic information available so far, we found 39 genes not localized to chromosomes in *C. annuum*, 69 genes not localized to chromosomes in *C. baccatum*, and one gene not localized to a chromosome in

C. chinense. This phenomenon may also result from the error generated during chromosome assembly or the poor quality of assembly (Kim et al., 2017).

In conclusion, the results revealed that the MADS-box transcription factor family of pepper is somewhat conserved. The 10 pairs of genes found in *C. annuum* formed linear relationships between pairs (Figure 3), indicating that duplication occurred between *MADS-box* genes. We calculated the selection pressure of paralogous as well as orthologous genes of the MADS-box gene family in *Capsicum* spp. pepper. The results showed that among the paralogous homologs (Figure 4A), $Ka/Ks < 1$ for all paralogous genes in *C. annuum*, while in *C. baccatum* there were 32 pairs of paralogs with $Ka/Ks > 1$ (Supplemental Table S6). This indicates that *C. annuum* was subject to strong purifying selection during its evolution, whereas *C. baccatum* was susceptible to environmental changes. However, no Ka/Ks values for paralogous genes were detected in *C. chinense*. Among the orthologs in the three *Capsicum* species, *Capsicum annuum* (Ca), *C. baccatum* (Cb), and *C. chinense* (Cc) (Figure 4B), the mean *Ka/Ks* values of the orthologs of Ca-Cb, Ca-Cc, and Cb-Cc were 0.6055, 0.6003, and 0.5952, respectively, with Ca-Cb having the largest mean value, implying that the *MADS-box* homolog of Ca-Cb was subject to greater purifying selection.

3.4 Comparative evolutionary relationships of type II MADS-box genes in three *Capsicum* spp.

To study the contraction and expansion of type II MADS-box family members during evolution, the phylogenetic relationships among MIKC *MADS-box* genes in the three pepper species were explored using collinearity analysis. The results revealed 33 pairs of colinear genes between *C. annuum* and *C. baccatum*, and 18 pairs of colinear genes between *C. baccatum* and *C. chinense* (Figure 5, Supplemental Table S7). Most type II genes were located at both ends of chromosomes, such as *CaMADS8* on chromosome 1 and *CbMADS80* on chromosome 8. In addition, the homologous genes on chromosomes 1, 2, 11, and 12 of *C. annuum* were distributed on chromosomes 4 and 5 of *C. baccatum*; the type II homologous genes on chromosomes 1 and 2 of *C. annuum* were located on chromosomes 6 and 3 of *C. chinense*. In *C. baccatum*, the type II homologous genes of chromosomes 1, 5, and 8 were distributed on chromosomes 1, 2, and 6 of *C. chinense*. There was no type II MADS-box gene on chromosome 9 in the three *Capsicum* species, which may be related to interchromosome 9 translocations (Kim et al., 2017). In summary, most of the type II MADS-box genes in the three peppers showed conserved collinearity among chromosomal regions, but there was also deviation in duplicated gene pairs.

3.5 Expression characteristics of *MADS-box* genes in different pepper tissues

We next analyzed the expression profiles of the *MADS-box* genes in root, stem, leaf, and flower tissues using RNA-seq data for the three peppers. The results showed that the expression of *MADS-box* genes in 73 groups of orthologous genes was considerably different among the three

pepper species in the four tissues (Figure 6). Comprehensive analysis revealed that their expression patterns could mainly divide the genes into three categories: (1) genes expressed in all four tissues, such as *CaMADS93*/*CbMADS35*/*CcMADS58*, *CaMADS82*/*CbMADS198*/*CcMADS4* and *CaMADS116*/*CbMADS134*/*CcMADS66*, indicating that they are widely involved in the growth and development of pepper; (2) genes with high expression during flower development, such as *CaMADS26*/*CbMADS31*/*CcMADS11*, *CaMADS30*/*CbMADS33*/*CcMADS14*, and *CaMADS74*/*CbMADS83*/*CcMADS41*, belonging to type II, suggesting that they play important roles in flower development; (3) genes with high expression in roots, such as *CaMADS9*/*CbMADS81*/*CcMADS2* and *CaMADS68*/*CbMADS74*/*CcMADS36*, indicating that these genes may be involved in root development and some physiological and biochemical processes in underground plant part. In addition to the three distinct expression patterns, most of the orthologous genes showed the same expression trend in the three pepper species, but some homologous genes displayed different expressions in the same tissue. For example, *CcMADS69* was not expressed in any tissues, while its homologous gene *CaMADS127* was highly expressed not only in flowers but also in stems, indicating that orthologous genes in pepper may have gained or lost functions in the process of evolution.

To further observe the expression of the type II MADS-box genes in different tissues of pepper, an expression heat map of 39 *CaMADS-box* genes in root, stem, leaf, and flower tissues were drawn (Supplemental Figure S5). Six MADS-box genes, *CaMADS74*, *CaMADS30*, *CaMADS61*, *CaMADS26*, *CaMADS105*, and *CaMADS63*, were selected for analyzing expression levels in root, stem, leaf, flower and fruit tissues using qRT-PCR. The results revealed that the six *MADS-box* genes were differentially expressed in different tissues of pepper (Supplemental Figure S5A), but they were highly expressed in flowers, which was consistent with the results of RNA-seq data (Figure 6). Both *CaMADS26* and *CaMADS30* were highly expressed in flowers, moderately expressed in fruits, and almost not expressed in other tissues. Both *CaMADS61* and *CaMADS74* were highly expressed in flowers, with little or no expression in other tissues, but *CaMADS74* was weakly expressed in roots. Expression of *CaMADS63* was the highest in flowers, followed by fruits and leaves, and low or trace expression was found in other tissues. However, the expression of *CaMADS105* was higher in fruits than in flowers, and low or no expression was found in other tissues.

We next further analyzed the expression profiles of these six type II genes using qRT-PCR in sepal, petal, stamen, and pistil tissues (Supplemental Figure S5B). In the sepal, *CaMADS61* expression was the highest, followed by *CaMADS105* and *CaMADS74*. *CaMADS30* was moderately expressed, while *CaMADS63* and *CaMADS26* were weakly expressed. *CaMADS61* was highly expressed in petals, while the expression levels of *CaMADS105* and *CaMADS30* were relatively low. For other genes, there was little or no expression in petals. In stamen tissue, *CaMADS63*, *CaMADS26,* and *CaMADS74* showed slightly expressed; *CaMADS105* and *CaMADS30* were moderately expressed, and *CaMADS61* was

highly expressed. *CaMADS74* was the highest expression in the pistil, followed by *CaMADS105*. *CaMADS30*, *CaMADS26*, and *CaMADS63* were slightly expressed, but *CaMADS74* was not expressed in the pistil.

3.6 Interaction network of type II CaMADS proteins

To better understand the biological functions of type II MADS-box genes in pepper, we next predicted the interaction network of CaMADS proteins. Only 17 of 39 type II CaMADS-box members interacted each other (Figure 7). The interacting proteins were mainly flowering pathway proteins, among which AP1, AG, SEP3, AGL20, and AGL21 were at the core of the network. *AtAP1* regulates the transition of inflorescence meristem and the morphological development of flower organs (Mandel & Yanofsky, 1995). *AtAG* controls the stamens and carpels and inhibits the expression of *AtAP1* (Smaczniak et al., 2012). *AtSEP3* belongs to the class D gene, which involved in the process of flower development and activates the function of *AtAG* (Castillejo et al., 2005).

4. DISCUSSION

Gene duplication often accompanies plant evolution and is an important reason for the expansion of gene families (Cannon et al., 2004). The MADS-box family is one of the largest transcription factor families and plays an important role in growth and development, and signal transduction (Becker & Theissen, 2003; Kim et al., 2005). With the development of sequencing technology, MADS-box gene family members have been identified in a variety of plants in varying numbers, such as 107 *MADS-box* gene members in *A. thaliana* (Parenicová et al., 2003), 83 in *Camellia sinensis* (Zhang et al., 2021), 44 in *Nelumbo nucifera* (Lin et al., 2020), 44 in *Erigeron breviscapus* (Tang et al., 2019), 131 in *Solanum lycopersicum* (Wang et al., 2019), 54 in *Morella rubra* (Zhao et al., 2019), 42 *Phyllostachys heterocycle* (Zhang et al., 2018), 80 in *Triticum aestivum* (Ma et al., 2017), 54 in *Ziziphus jujuba* (Zhang et al., 2017), and 144 in *Raphanus sativus* (Li et al., 2016). These studies indicate that *MADS-box* genes have undergone obvious amplification and contraction, and the number and distribution in different subfamilies is also different. We study identified 174, 207, and 72 *MADS-box* genes from *C. annuum*, *C. baccatum,* and *C. chinense*, respectively (Figure 1; Supplemental Tables S1), in line with this trend. Moreover, the number of MADS-box family genes in *C. baccatum* was more than that in *C. annuum* and *C. chinense*, which may be due to the expansion of the *C. baccatum* genome caused by the amplification of retrotransposons (Kim et al., 2017). The number of *MADS-box* genes of type I and MIKC subfamilies in *C. annuum* and *C. baccatum* was more than that of the model plant *A. thaliana* and the related species tomato. These results indicate that the *MADS-box* genes in *C. annuum* and *C. baccatum* have significantly expanded, but it is strange that the number of genes identified in *C. chinense* was lower than that in *C. annuum* and *C. baccatum*, suggesting that a large number of genes have been lost during evolution. Among the MADS-box family members, those belonging to the same subfamily possessed similar motif composition and gene structure, but there was a unique motif composition and gene structure between type I and type II genes. The

MADS-box genes of type I, including M , M , and M , generally contained no or few exons (Supplemental Figures S2-S4), and may have lost multiple introns during the diversification of the MADS-box family. In addition, the distribution of introns in pepper *MADS-box* genes was also different. MIKC type genes had more introns than those of type I, which are also found in A. thaliana, tomato, and rice (Wang et al., 2017), indicating that evolution between species is conserved. However, some genes of the same subfamily showed different intron and exon arrangements, indicating the complexity of gene structure evolution, which needs further study. The same conserved motifs in the same subfamily (Supplemental Figures S2B, S3B, S4B) suggest that these motifs play an important role in gene functional specificity. Analyses of gene structure and conserved motifs provide clues for the expansion and contraction of the MADS-box gene family in pepper.

In eukaryotes, gene replication plays an important role in amplifying the number of transcription factor families and genomic complexity (Castillejo et al., 2005; Wang et al., 2019) Previous studies confirmed that *C. annuum* diverged from *C. chinense* 1.14 million years ago and *C. baccatum* diverged from *C. chinense* and *C. annuum* 1.7 million years ago (Kim et al., 2017). Our study revealed only 47 groups of orthologous genes among three pepper species, while other orthologous genes were lost to a certain extent (Figure 2). We also identified 144 groups of paralogous genes and 195 groups of orthologous genes (Figure 3, Supplemental Tables S3-S5). However, some homologous genes were lost in the three species of pepper. For example, *CcMADS43*/*CaMADS76* and *CcMADS45*/*CaMADS76* had only one homologous gene, *CbMADS101*, in *C. baccatum*; with *C. annuum* gene *CaMADS76* having no homolog in *C. baccatum*, which indicates that some homologous genes have been lost, but some homologous genes have been expanded during the evolution of pepper. These results further demonstrate both obvious contraction and expansion trends that the different subfamily members of the MADS-box gene family during the process of pepper evolution.

Genome sequencing revealed a dynamic rearrangement of chromosomes 3, 5, and 9 in *C. baccatum*, namely translocation (Kim et al., 2017). *MADS-box* genes at these loci were also changed, such as *CbMADS35* on chromosome 3 of *C. baccatum*, and its homologous gene *CcMADS58* on chromosome 9. In addition, some homologous genes were also located on different chromosomes, such as *CbMADS26*/*CcMADS30* located on chromosomes 1 and 6 respectively, *CaMADS104*/*CbMADS121* located on chromosome 11, and *CcMADS32* located on chromosome 6. Moreover, most of the orthologous genes with Ka/Ks greater than 1 displayed positive selection and may show positive changes in function under the influence of the environment (Figure 4).

Most *MADS-box* genes were differentially expressed in different tissues of pepper (Figure 6; Supplemental Figure S5), indicating their functional diversity in different tissues. Some *MADS-box* genes showed tissue-specific expression, such as *CaMADS9*/*CbMADS81*/*CcMADS2* and *CaMADS66*/*CbMADS75*/*CcMADS35*, which were mainly specifically expressed in roots, and are important candidate

genes for further functional analysis. Some *MADS-box* genes were highly expressed in fruits (Supplemental Figure S5A), such as *CaMADS30*, *CaMASDS61*, *CaMADS63*, and *CaMADS105*, suggesting important roles in controlling fruit development. Several studies have proved that *MADS-box* genes play an important role in the morphogenesis and growth of roots and fruits (Ng & Yanofsky, 2001; Riechmann et al., 1996).

The MIKC MADS-box genes play an important role in floral organ identity and flowering regulation in *A. thaliana* and *Petunia hybrida* (Heijmans et al., 2012; Ma et al., 1991). In *A. thaliana*, these genes include A genes (*AP1*, *AP2*), B genes (*AP3*, *Pl*), C genes (*AG*), D genes (*STK*, *SHP1/2*), and E genes (*SEP1/2/3/4*) (Ditta et al., 2004; Pelaz et al., 2000; Robles et al., 2005). Petunia possesses class A genes *PETUNIA FLOWERING GENE* (*PFG*), *FLORAL BINDING PROTEIN 26* (*FBP26*), and *FBP29*; class B genes *MADS-box gene 6* (*TM6*), *PMADS1/GP*, *PMADS2*, and *FBP24*; class C genes *PMADS3*, *FBP6*, and *FBP24*; class D genes *FBP1/7/2/4/5/9*; class E genes *FBP23*, *PMADS4*, and *PMADS12* (Robles et al., 2005). In this study, we predicted the possible key genes in pepper flower organ development through phylogenetic relationships, such as class A genes (*CcMADS13/CaMADS31/CbMADS34*), class B gene (*CcMADS41/CaMADS74/CbMADS83*), class C gene (*CcMADS11/CaMADS26/CbMADS31*), class D gene (*CcMADS8/CaMADS17/CbMADS27*), and class E gene (*CcMADS14/CaMADS30/CbMADS33*, *CcMADS61/CaMADS101/CbMADS114*). Based on the typical "ABCDE model", class A gene *AP1* functions in controlling sepal formation (Irish & Sussex, 1990). Phylogenetic analysis showed that *CcMADS13/CaMADS31/CbMADS34* occupies the same branch as *AP1* and *CAULIFLOWER/FRUITFUL*, suggesting that they are involved in the development of sepals. *AP3* and *PI*, class B genes, are involved in controlling petal and stamen formation (Mao et al., 2015; Sundström et al., 2006). The *MADS-box* genes of pepper in this branch include *CcMADS41/CaMADS74/CbMADS83* and *CaMADS53/CaMADS60*, suggesting functions in the development of petals and stamens. Our study showed that the pepper genes homologous to class C gene *AG* are *CcMADS11/CaMADS26/CbMADS31*, which may function in controlling stamen and ovule development (Pan et al., 2010). *CcMADS8/CaMADS17/CbMADS27* occupies the same branch as class D genes *STK* and *SHP1/2*, which is postulated to be involved in ovule development (Jack et al., 1990; Pinyopich et al., 2003). *A. thaliana SEP* genes are typical class E genes, which interact with the other four classes and are essential genes for the development of the sepals, petals, stamens, and carpels. The homologous *SEP* genes of pepper are *CcMADS14/CaMADS30/CbMADS33* and *CcMADS61/CaMADS101/CbMADS114*. Gene expression analysis using qRT-PCR showed that pepper E genes were expressed in sepal, petal, stamen, and pistil (Supplemental Figure S5B), suggesting that they plays vital roles in all stages of flower development. MIKC MADS-box genes play a central role in plant development (Schilling et al., 2018). The MIKC MADS-box genes identified in this study should be candidate genes for pepper breeding and improvement.

Acknowledgments

This work was supported by the Talent Introduction Start-up Fund Project of Anhui Science and Technology University (NXYJ202001), the University Discipline (Major) Top Talent Cultivation Funding Project (gxbjZD2021072), and the Postgraduate Science Research Project of Anhui Science and Technology University (YK202115).

SUPPLEMENTAL MATERIALS

Figure S1 Phylogenetic tree of type I MADS-box genes in *Arabidopsis thaliana*, *Capsicum annuum*, *C. baccatum*, and *C. chinense*.

Figure S2 Gene structure and motifs analysis of *MADS-box* genes in *Capsicum annuum*.

Figure S3 Gene structure and motifs analysis of *MADS-box* genes in *Capsicum baccatum*.

Figure S4 Gene structure and motifs analysis of *MADS-box* genes in *Capsicum chinense*.

Figure S5 Tissue expression profiles of six *MADS-box* genes using qRT-PCR A: Expression patterns of six *MADS-box* genes in root, stem, leaf, flower, and fruit tissues; B: Expression patterns of six *MADS-box* genes in sepal, petal, stamen, and pistil tissues.

Table S1 Information of *MADS-box* genes identified in three *Capsicum* species.

Table S2 Conserved amino acid sequences in *MADS-box* genes of three *Capsicum* species identified by MEME.

Table S3 The identified orthologous genes in three *Capsicum* species.

Table S4 The identified orthologous genes between two *Capsicum* species

Table S5 The identified paralogous genes among in *Capsicum* species.

Table S6 The Ka/Ks ratios for the paralogous and orthologous gene pairs of *MADS-box* genes in *Capsicum annuum*, *C. baccatum*, and *C. chinense*.

Table S7 Homologous II type MADS-box gene pairs in *Capsicum annuum*, *C. baccatum*, and *C. chinense*

Table S8 Primer information used for qRT-PCR.

author Contributions

Zhicheng Gan: Genome-wide identification; Bioinformatics analysis; Writing–original draft. Xingxing Wu: RNA extraction; cDNA synthesis; qRT-PCR assays. Tingting Feng: qRT-PCR validation; Review. Nengbing Hu and Xiaoming Lu: Pepper field management; Ruining Li: Project administration; Writing–review & editing; Xianzhong Huang: Conceptualization; Funding acquisition; Project administration; Supervision; Writing–review & editing.

Conflict of InterestS

The authors declare that there are no competing interests regarding the publication of this paper.

ORCID

Xianzhong Huang https://orcid.org/0000-0003-1264-8106

Ruiling Li https://orcid.org/0000-0003-1117-8800

References

Adamski, N. M., Simmonds, J., Brinton, J. F., Backhaus, A. E., Chen, Y., Smedley, M., Hayta, S., Florio, T., Crane, P., Scott, P., Pieri, A., Hall, O., Barclay, J. E., Clayton, M., Doonan, J. H., Nibau, C., & Uauy, C. (2021). Ectopic expression of *Triticum polonicum* VRT-A2 underlies elongated glumes and grains in hexaploid wheat in a dosage-dependent manner. *The Plant cell*, *33*, 2296–2319. https://doi.org/10.1093/plcell/koab119

Aguilar-Meléndez, A., Morrell, P. L., Roose, M. L., & Kim, S. C. (2009). Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; *Solanaceae*) from Mexico. A*merican journal of botany*, *96*, 1190–1202. https://doi.org/10.3732/ajb.0800155

Albrecht, E., Zhang, D., Mays, A. D., Saftner, R. A., & Stommel, J. R. (2012). Genetic diversity in *Capsicum baccatum* is significantly influenced by its eco-geographical distribution. *BMC genetics*, *13*, 68. https://doi.org/10.1186/1471-2156-13-68

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martínez-Castilla, L., & Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 5328–5333. https://doi.org/10.1073/pnas.97.10.5328

Becker, A., & Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular phylogenetics and evolution*, *29*, 464–489. https://doi.org/10.1016/s1055-7903(03)00207-0

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., & May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC plant biology*, *4*, 10. https://doi.org/10.1186/1471-2229-4-10

Castillejo, C., Romera-Branchat, M., & Pelaz, S. (2005). A new role of the Arabidopsis SEPALLATA3 gene revealed by its constitutive expression. *The Plant journal*, *43*, 586–596. https://doi.org/10.1111/j.1365-313X.2005.02476.x

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular plant*, *13*, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen, R., Ma, J., Luo, D., Hou, X., Ma, F., Zhang, Y., Meng, Y., Zhang, H., & Guo, W. (2019). CaMADS, a MADS-box transcription factor from pepper, plays an important role in the response to cold, salt, and osmotic stress. *Plant science*, *280*, 164–174. https://doi.org/10.1016/j.plantsci.2018.11.020

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., & Yanofsky, M. F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Current biology, *14*, 1935–1940. https://doi.org/10.1016/j.cub.2004.10.028

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., & Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant cell*, *15*, 2603–2611. https://doi.org/10.1105/tpc.015123

Ferrario, S., Immink, R. G., & Angenent, G. C. (2004). Conservation and diversity in flower land. *Current opinion in plant biology*, *7*, 84–91. https://doi.org/10.1016/j.pbi.2003.11.003

Grimplet, J., Martínez-Zapater, J. M., & Carmona, M. J. (2016). Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC genomics*, *17*, 80. https://doi.org/10.1186/s12864-016-2398-7

Heijmans, K., Ament, K., Rijpkema, A. S., Zethof, J., Wolters-Arts, M., Gerats, T., & Vandenbussche, M. (2012). Redefining C and D in the petunia ABC. *The Plant cell*, *24*, 2305–2317. https://doi.org/10.1105/tpc.112.097030

Irish, V. F., & Sussex, I. M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. *The Plant cell*, *2*, 741–753. https://doi.org/10.1105/tpc.2.8.741

Jack, T., Brockman, L. L., & Meyerowitz, E. M. (1992). The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*, *68*, 683–697. https://doi.org/10.1016/0092-8674(92)90144-2

Jin, Y., Liu, F., Huang, W., Sun, Q., & Huang, X. (2019). Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. *Scientific reports*, *9*, 8408. https://doi.org/10.1038/s41598-019-44849-1

Kim, S., Koh, J., Yoo, M. J., Kong, H., Hu, Y., Ma, H., Soltis, P. S., & Soltis,

D. E. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *The Plant journal* , *43*, 724–744. https://doi.org/10.1111/j.1365-313X.2005.02487.x

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., Kim, M. S., Lee, J. M., Cheong, K., Shin, H. S., Kim, S. B., Han, K., Lee, J., Park, M., Lee, H. A., Lee, H. Y., Lee, Y., Oh, S., Lee, J. H., Choi, E., Choi, E., Lee. S. E., Jeon, J., Kim, H., Choi, G., Song, H., Lee, J., Lee, S. C., Kwon, J. K., Lee, H. Y., Koo, N., Hong, Y., Kim, R. W., Kang, W. H., Huh, J. H., Kang, B. C., Yang, T. J., Lee, Y. H., Bennetzen, J. L., Choi, D. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome biology*, *18*, 210. https://doi.org/10.1186/s13059-017-1341-9

Lai, D., Yan, J., He, A., Xue, G., Yang, H., Feng, L., Wei, X., Li, L., Xiang, D., Ruan, J., Fan, Y., & Cheng, J. (2022). Genome-wide identification, phylogenetic and expression pattern analysis of MADS-box family genes in foxtail millet (*Setaria italica*). *Scientific reports*, *12*, 4979. https://doi.org/10.1038/s41598-022-07103-9

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Lawton-Rauh, A. L., Alvarez-Buylla, E. R., & Purugganan, M. D. (2000). Molecular evolution of flower development. *Trends in ecology & evolution*, *15*, 144–149. https://doi.org/10.1016/s0169-5347(99)01816-9

Lee, S., Jeong, D. H., & An, G. (2008). A possible working mechanism for rice SVP-group MADS-box proteins as negative regulators of brassinosteroid responses. *Plant signaling & behavior*, *3*, 471–474. https://doi.org/10.4161/psb.3.7.5677

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, *12*, 323. https://doi.org/10.1186/1471-2105-12-323

Li, C., Wang, Y., Xu, L., Nie, S., Chen, Y., Liang, D., Sun, X., Karanja, B. K., Luo, X., & Liu, L. (2016). Genome-Wide Characterization of the MADS-Box Gene Family in Radish (*Raphanus sativus* L.) and Assessment of Its Roles in Flowering and Floral Organogenesis. *Frontiers in plant science*, *7*, 1390. https://doi.org/10.3389/fpls.2016.01390

Liao, Y., Wang, J., Zhu, Z., Liu, Y., Chen, J., Zhou, Y., Liu, F., Lei, J., Gaut, B. S., Cao, B., Emerson, J. J., & Chen, C. (2022). The 3D architecture of the pepper genome and its relationship to function and evolution. *Nature communications*, *13*, 3479. https://doi.org/10.1038/s41467-022-31112-x

Lin, Z., Cao, D., Damaris, R. N., & Yang, P. (2020). Genome-wide identification of MADS-box gene family in sacred lotus (*Nelumbo nucifera*) identifies a

SEPALLATA homolog gene involved in floral development. *BMC plant biology*, *20*, 497. https://doi.org/10.1186/s12870-020-02712-w

Ma, H., Yanofsky, M. F., & Meyerowitz, E. M. (1991). *AGL1-AGL6*, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes & development*, *5*, 484–495. https://doi.org/10.1101/gad.5.3.484

Ma, J., Yang, Y., Luo, W., Yang, C., Ding, P., Liu, Y., Qiao, L., Chang, Z., Geng, H., Wang, P., Jiang, Q., Wang, J., Chen, G., Wei, Y., Zheng, Y., & Lan, X. (2017). Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum* L.). *PloS one*, *12*, e0181443. https://doi.org/10.1371/journal.pone.0181443

Mandel, M. A., & Yanofsky, M. F. (1995). A gene triggering flower formation in *Arabidopsis*. *Nature*, *377*, 522–524. https://doi.org/10.1038/377522a0

Mao, W. T., Hsu, H. F., Hsu, W. H., Li, J. Y., Lee, Y. I., & Yang, C. H. (2015). The C-Terminal Sequence and PI motif of the Orchid (*Oncidium Gower Ramsey*) PISTILLATA (PI) Ortholog Determine its Ability to Bind AP3 Orthologs and Enter the Nucleus to Regulate Downstream Genes Controlling Petal and Stamen Formation. *Plant & cell physiology*, *56*, 2079–2099. https://doi.org/10.1093/pcp/pcv129

Mizukami, Y., & Ma, H. (1992). Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell*, *71*, 119–131. https://doi.org/10.1016/0092-8674(92)90271-d

Mohammadi, N., Mehrabi, R., Mirzadi Gohari, A., Roostaei, M., Mohammadi Goltapeh, E., Safaie, N., & Kema, G. (2020). MADS-Box Transcription Factor *ZtRlm1* Is Responsible for Virulence and Development of the Fungal Wheat Pathogen *Zymoseptoria tritici*. *Frontiers in microbiology*, *11*, 1976. https://doi.org/10.3389/fmicb.2020.01976

Molesini, B., Dusi, V., Pennisi, F., & Pandolfini, T. (2020). How Hormones and MADS-Box Transcription Factors Are Involved in Controlling Fruit Set and Parthenocarpy in Tomato. *Genes*, *11*, 1441. https://doi.org/10.3390/genes11121441

Ng, M., & Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nature reviews. Genetics*, *2*, 186–195. https://doi.org/10.1038/35056041

Pan, I. L., McQuinn, R., Giovannoni, J. J., & Irish, V. F. (2010). Functional diversification of *AGAMOUS* lineage genes in regulating tomato flower and fruit development. *Journal of experimental botany*, *61*, 1795–1806. https://doi.org/10.1093/jxb/erq046

Parenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., Angenent, G. C., & Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the

MADS world. *The Plant cell*, *15*, 1538–1551. https://doi.org/10.1105/tpc.011544

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, *405*, 200–203. https://doi.org/10.1038/35012103

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., & Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, *424*, 85–88. https://doi.org/10.1038/nature01741

Qi, X., Liu, C., Song, L., & Li, M. (2020). PaMADS7, a MADS-box transcription factor, regulates sweet cherry fruit ripening and softening. *Plant science*, *301*, 110634. https://doi.org/10.1016/j.plantsci.2020.110634

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D. L., Tian, H., Tang, X., Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, W., Chen, Z., Feng, Y., Niu. Y., Bi, S., Yang, X., Li, W., Cai, H., Luo, X., Montes-Hernández, S., Leyva-González, M. A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, X., Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, L., Zhang, L., Zhang, Y., Liao, W., Zhang, Y., Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Y., Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P. W., Li, Y., Krogh, A., Rivera-Bustamante, RF., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K., Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 5135–5140. https://doi.org/10.1073/pnas.1400975111

Riechmann, J. L., Krizek, B. A., & Meyerowitz, E. M. (1996). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 4793–4798. https://doi.org/10.1073/pnas.93.10.4793

Robles, P., & Pelaz, S. (2005). Flower and fruit development in *Arabidopsis thaliana*. *The International journal of developmental biology*, *49*, 633–643. https://doi.org/10.1387/ijdb.052020pr

Roy Choudhury, S., Roy, S., Nag, A., Singh, S. K., & Sengupta, D. N. (2012). Characterization of an *AGAMOUS*-like MADS box protein, a probable constituent of flowering and fruit ripening regulatory system in banana. *PloS one*, *7*, e44361. https://doi.org/10.1371/journal.pone.0044361

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular biology and evolution*, *34*, 3299–3302. https://doi.org/10.1093/molbev/msx248

Schilling, S., Pan, S., Kennedy, A., & Melzer, R. (2018). MADS-box genes and crop domestication: the jack of all traits. *Journal of experimental botany*, *69*, 1447–1469. https://doi.org/10.1093/jxb/erx479

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., & Sommer, H. (1990). Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*. *Science*, *250*, 931–936. https://doi.org/10.1126/science.250.4983.931

Shimada, S., Ogawa, T., Kitagawa, S., Suzuki, T., Ikari, C., Shitsukawa, N., Abe, T., Kawahigashi, H., Kikuchi, R., Handa, H., & Murai, K. (2009). A genetic network of flowering-time genes in wheat leaves, in which an *APETALA1/FRUITFULL*-like gene, VRN1, is upstream of *FLOWERING LOCUS T*. *The Plant journal*, *58*, 668–681. https://doi.org/10.1111/j.1365-313X.2009.03806.x

Shore, P., & Sharrocks, A. D. (1995). The MADS-box family of transcription factors. *European journal of biochemistry*, 229, 1–13. https://doi.org/10.1111/j.1432-1033.1995.tb20430.x

Smaczniak, C., Immink, R. G., Angenent, G. C., & Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* , *139*, 3081–3098. https://doi.org/10.1242/dev.074674

Smaczniak, C., Immink, R. G., Muiño, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q. D., Liu, S., Westphal, A. H., Boeren, S., Parcy, F., Xu, L., Carles, C. C., Angenent, G. C., & Kaufmann, K. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 1560–1565. https://doi.org/10.1073/pnas.1112871109

Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current protocols in bioinformatics*, *47*, 8.13.1–8.13.24. https://doi.org/10.1002/0471250953.bi0813s47

Sundström, J. F., Nakayama, N., Glimelius, K., & Irish, V. F. (2006). Direct regulation of the floral homeotic *APETALA1* gene by *APETALA3* and *PISTILLATA* in *Arabidopsis*. *The Plant journal*, *46*, 593–600. https://doi.org/10.1111/j.1365-313X.2006.02720.x

Sung, S. K., Moon, Y. H., Chung, J. E., Lee, S. Y., Park, H. G., & An, G. (2001). Characterization of MADS box genes from hot pepper. *Molecules and cells*, 11, 352–359.

Swiezewski, S., Liu, F., Magusin, A., & Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, *462*, 799–802. https://doi.org/10.1038/nature08618

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Tang, W., Tu, Y., Cheng, X., Zhang, L., Meng, H., Zhao, X., Zhang, W., & He, B. (2019). Genome-wide identification and expression profile of the MADS-box gene family in *Erigeron breviscapus*. *PloS one*, *14*, e0226599. https://doi.org/10.1371/journal.pone.0226599

Theissen, G., & Saedler, H. (2001). Plant biology. Floral quartets. *Nature*, *409*, 469–471. https://doi.org/10.1038/35054172

Theissen, G., Kim, J. T., & Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of molecular evolution*, *43*, 484–516. https://doi.org/10.1007/BF02337521

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, *Chapter 2*, Unit 2.3. https://doi.org/10.1002/0471250953.bi0203s00

Troconis-Torres, I. G., Rojas-López, M., Hernández-Rodríguez, C., Villa-Tanaca, L., Maldonado-Mendoza, I. E., Dorantes-Álvarez, L., Tellez-Medina, D., & Jaramillo-Flores, M. E. (2012). Biochemical and molecular analysis of some commercial samples of chilli peppers from Mexico. *Journal of biomedicine & biotechnology*, *2012*, 873090. https://doi.org/10.1155/2012/873090

Wan, H., Yuan, W., Ruan, M., Ye, Q., Wang, R., Li, Z., Zhou, G., Yao, Z., Zhao, J., Liu, S., & Yang, Y. (2011). Identification of reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*Capsicum annuum* L.). *Biochemical and biophysical research communications*, *416*, 24–30. https://doi.org/10.1016/j.bbrc.2011.10.105

Wang, R., Ming, M., Li, J., Shi, D., Qiao, X., Li, L., Zhang, S., & Wu, J. (2017). Genome-wide identification of the MADS-box transcription factor family in pear (*Pyrus bretschneideri*) reveals evolution and functional divergence. *PeerJ*, *5*, e3776. https://doi.org/10.7717/peerj.3776

Wang, Y., Zhang, J., Hu, Z., Guo, X., Tian, S., & Chen, G. (2019). Genome-Wide Analysis of the MADS-Box Transcription Factor Family in *Solanum lycopersicum*. *International journal of molecular sciences*, *20*, 2961. https://doi.org/10.3390/ijms20122961

Weigel, D., & Meyerowitz, E. M. (1994). The ABCs of floral homeotic genes. *Cell*, *78*, 203–209. https://doi.org/10.1016/0092-8674(94)90291-7

Xie, Q., Hu, Z., Zhu, Z., Dong, T., Zhao, Z., Cui, B., & Chen, G. (2014). Overexpression of a novel MADS-box gene *SlFYFL* delays senescence, fruit ripening and abscission in tomato. *Scientific reports*, *4*, 4367. https://doi.org/10.1038/srep04367

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multi-

ple species. *Nucleic acids research*, *47*, W52–W58. https://doi.org/10.1093/nar/gkz333

Yadav, C. B., Bonthala, V. S., Muthamilarasan, M., Pandey, G., Khan, Y., & Prasad, M. (2015). Genome-wide development of transposable elements-based markers in foxtail millet and construction of an integrated database. *DNA research : an international journal for rapid publication of reports on genes and genomes*, *22*, 79–90. https://doi.org/10.1093/dnares/dsu039

Yamaguchi, T., & Hirano, H. Y. (2006). Function and diversification of MADS-box genes in rice. *TheScientificWorldJournal*, *6*, 1923–1932. https://doi.org/10.1100/tsw.2006.320

Zhang, H., Gao, S., Lercher, M. J., Hu, S., & Chen, W. H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic acids research*, *40*, W569–W572. https://doi.org/10.1093/nar/gks576

Zhang, J., Hu, Z., Yao, Q., Guo, X., Nguyen, V., Li, F., & Chen, G. (2018). A tomato MADS-box protein, SlCMB1, regulates ethylene biosynthesis and carotenoid accumulation during fruit ripening. *Scientific reports*, *8*, 3413. https://doi.org/10.1038/s41598-018-21672-8

Zhang, L., Zhao, J., Feng, C., Liu, M., Wang, J., & Hu, Y. (2017). Genome-wide identification, characterization of the MADS-box gene family in Chinese jujube and their involvement in flower development. *Scientific reports*, *7*, 1025. https://doi.org/10.1038/s41598-017-01159-8

Zhang, Y., Tang, D., Lin, X., Ding, M., & Tong, Z. (2018). Genome-wide identification of MADS-box family genes in moso bamboo (*Phyllostachys edulis*) and a functional analysis of PeMADS5 in flowering. *BMC plant biology*, *18*, 176. https://doi.org/10.1186/s12870-018-1394-2

Zhang, Z. B., Jin, Y. J., Wan, H. H., Cheng, L., & Feng, Z. G. (2021). Genome-wide identification and expression analysis of the MADS-box transcription factor family in *Camellia sinensis*. *Journal of applied genetics*, *62*, 249–264. https://doi.org/10.1007/s13353-021-00621-8

Zhao, H. B., Jia, H. M., Wang, Y., Wang, G. Y., Zhou, C. C., Jia, H. J., & Gao, Z. S., Dr (2019). 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*, *717*, 144045. https://doi.org/10.1016/j.gene.2019.144045

Zhao, P. X., Miao, Z. Q., Zhang, J., Chen, S. Y., Liu, Q. Q., & Xiang, C. B. (2020). Arabidopsis MADS-box factor AGL16 negatively regulates drought resistance via stomatal density and stomatal movement. J*ournal of experimental botany*, *71*, 6092–6106. https://doi.org/10.1093/jxb/eraa303

FIGURE LEGENDS

Figure 1 Phylogenetic tree of type II MADS-box genes in *Arabidopsis thaliana*, *Capsicum annuum*, *C. baccatum*, and *C. chinense*.

Figure 2 Number of *MADS-box* orthologs in *Capsicum annuum*, *C. baccatum*, and *C. chinense*.

Figure 3 Homologous *MADS-box* gene pairs in *Capsicum annuum* (Ca), *C. baccatum* (Cb), and *C. chinense* (Cc). Tracks from outside to inside are chromosomes numbers, gene density of the chromosome, and homologous gene pairs among the three *Capsicum* species. Blue lines connect homologous gene pairs that exist in three *Capsicum* species; red lines connect homologous gene pairs in two species; green lines connect paralogous genes.

Figure 4 Selection pressure statistics of paralogous (A) and orthologous (B) *MADS-box* genes in pepper. Ca, *Capsicum annuum*; Cb, *C. baccatum*; Cc, *C. chinense.* The 'dots' reflect the maximum and minimum Ka/Ks scores.

Figure 5 Collinearity of type II MADS-box genes in *Capsicum annuum, C. baccatum, and C. chinense.*

Figure 6 Expression profiles of *MADS-box* genes in different tissues from *Capsicum annuum* (Ca), *C. baccatum* (Cb), and *C. chinense* (Cc). Color bar indicates the variation range of $\log_{10}$(FPKM+1) values of *MADS-box* genes in different tissues. Expression of genes marked in red was verified by qRT-PCR.

Figure 7 Type II CaMADS-box protein interaction network diagram. Connection between nodes vary with the combined_score value (representing the reliability of the predicted interaction between the two proteins, the value is 0-1), thickening with an increase in score value. The size of nodes increases with the number of proteins interacting with node proteins.