

RESEARCH ARTICLE

Genome-wide analysis of MADS-box transcription factor gene family in wild emmer wheat (*Triticum turgidum* subsp. *dicoccoides*)

Ghader Mirzaghaderi *

Department of Plant Production and Genetics, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran

* gh.mirzaghaderi@uok.ac.ir

Abstract

The members of MADS-box gene family have important roles in regulating the growth and development of plants. MADS-box genes are highly regarded for their potential to enhance grain yield and quality under shifting global conditions. Wild emmer wheat (*Triticum turgidum* subsp. *dicoccoides*) is a progenitor of common wheat and harbors valuable traits for wheat improvement. Here, a total of 117 MADS-box genes were identified in the wild emmer wheat genome and classified to 90 MIKC^C, 3 MIKC*, and 24 M-type. Furthermore, a phylogenetic analysis and expression profiling of the emmer wheat MADS-box gene family was presented. Although some MADS-box genes belonging to *SOC1*, *SEP1*, *AGL17*, and *FLC* groups have been expanded in wild emmer wheat, the number of MIKC-type MADS-box genes per subgenome is similar to that of rice and Arabidopsis. On the other hand, M-type genes of wild emmer wheat is less frequent than that of Arabidopsis. Gene expression patterns over different tissues and developmental stages agreed with the subfamily classification of MADS-box genes and was similar to common wheat and rice, indicating their conserved functionality. Some TdMADS-box genes are also differentially expressed under drought stress. The promoter region of each of the TdMADS-box genes harbored 6 to 48 responsive elements, mainly related to light, however hormone, drought, and low-temperature related cis-acting elements were also present. In conclusion, the results provide detailed information about the MADS-box genes of wild emmer wheat. The present work could be useful in the functional genomics efforts toward breeding for agronomically important traits in *T. dicoccoides*.

OPEN ACCESS

Citation: Mirzaghaderi G (2024) Genome-wide analysis of MADS-box transcription factor gene family in wild emmer wheat (*Triticum turgidum* subsp. *dicoccoides*). PLoS ONE 19(3): e0300159. <https://doi.org/10.1371/journal.pone.0300159>

Editor: Muhammad Abdul Rehman Rashid, Government College University Faisalabad, PAKISTAN

Received: October 25, 2023

Accepted: February 19, 2024

Published: March 7, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0300159>

Copyright: © 2024 Ghader Mirzaghaderi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Introduction

Wheat is an important crop worldwide, occupying 17% of global cultivated lands and providing 30% of global calorie consumption [1]. However, abiotic stresses, such as drought and salinity, have a significant impact on its yield, particularly under changing climate conditions. Wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*; common name: *T. dicoccoides*), the progenitor of the A and B genome of bread wheat, has been adapted to abiotic stress during evolution and has a great potential for wheat improvement [2, 3]. Identification of genes

Funding: GM was supported by Iran National Science Foundation (INSF) grant 99014038. The funding doesn't include publication fee.

Competing interests: The author have declared that no competing interests exist.

associated with stress tolerance in wild emmer wheat, helps us to understand the mechanism underlying stress response which can be applied in wheat breeding programs.

MADS-box genes compose a regulatory family of transcription factors found in all eukaryotes and play a crucial role in controlling various aspects of plant growth and development, including flowering, fruit ripening, and seed formation. MADS-box genes have been well documented in *Arabidopsis* and rice and have been studied in common wheat [4–6] and many other plants. Genes associated with stress tolerance in wild emmer wheat have been identified [7–10]. It has been shown in model plants that some MADS-box genes modulate tolerance to drought [11–13] and cold [14]. For example, *OsMADS26*-down-regulated rice plants are more tolerant to drought without a strong impact on plant development [15]. There are evidences that the induction of *OsMADS27* mediates salt tolerance in rice [16]. In *Arabidopsis*, MADS-box genes are involved in response to water stress and drought resistance possibly by the regulation of abscisic acid (ABA) pathway [17]. Beside these evidences of the involvement of the MADS-box genes in plant growth, development and tolerance against stresses, the detailed information on MADS-box gene family is not available yet in wild emmer wheat.

It has been known for decades that the floral homeotic genes, AG (AGAMOUS) from *Arabidopsis thaliana* and DEF A (DEFICIENS A) from snap dragon (*Antirrhinum majus*), share strong sequence similarity with DNA-binding domain of SRF (SERUM RESPONSE FACTOR) transcription factor of humans and MCM1 (MINICHROMOSOME MAINTENANCE 1) of yeast. This conserved domain has since been named the MADS-box followed by the initials of MCM1, AG, DEF, and SRF. Based on the sequence of this highly conserved MADS domain which is a 58–60 amino acid DNA-binding sequence, two types of MADS-box has been distinguished [18]. The first type is known as M-type or type I MADS-box genes, which commonly contain the MADS-box domain without any other conserved domains. The second type is type II or MIKC-type MADS-box genes which harbour MADS-, I-, K-, and C-terminal domains. The additional domains downstream of the MADS-box MIKC-type proteins, especially the conserved keratin-like (K) domain play a role in protein interactions and dimerization [19, 20]. A short intervening (I) domain, separates the MADS and K domains. The I domain may also be involved in interaction with other proteins [21]. MIKC-type MADS-box proteins may also contain a variable C-terminal domain that involves in protein interaction, transcription activation or protein modification [22, 23]. Because the function of the C domain has not been clearly defined due to its variability, MIKC-type MADS-box proteins that have MADS and K domains are considered as fully functional.

The type I MADS-box proteins have been divided into $M\alpha$, $M\beta$, and $M\gamma$ clades [24]. In *A. thaliana* some members of the type I genes are important for normal development of the female gametophyte or endosperm and may be responsible for post-zygotic lethality in interspecific hybrids [25–31]. Plant Type II proteins are also divided into $MIKC^C$ and $MIKC^*$ [32]. In angiosperms and ferns, various classes of $MIKC^C$ genes have been identified while only two classes of $MIKC^*$ genes have been recognized [33, 34] based on phylogenetic relationships. Several studies imply the importance of $MIKC^*$ genes in pollen development [35–38]. On the other hand, the genes of $MIKC^C$ class play important roles in flowering time, floral organ identity, and fruit development [39–44]. The role of the MADS-box gene family is not confined to flower development. They are key components of the gene regulatory networks associated with the distinct developmental fates in the root [45] and are involved against various stress conditions [12, 46].

Here, I performed an in-silico genome-wide investigation to identify the MADS-box family members in wild emmer wheat. The phylogenetic relationship, physical localization, gene structure, conserved domain, cis-acting elements, and related micro RNAs (miRNAs) of the identified MADS-box genes were analyzed. Furthermore, the expression patterns of MADS-

box genes in different tissues and time points were investigated using publicly available RNA-seq and microarray data. This study provides information about the important candidate MADS-box genes for further wheat breeding programs.

Materials and methods

Identification of MADS-box genes in *T. dicoccoides*

Genomic DNA, protein, and transcript sequences, and the annotation file of *T. dicoccoides* were downloaded from EnsemblPlants (WEWSeq_v.1.0, <https://plants.ensembl.org/>). The Multiple Sequence Alignment for the MADS-box family was also downloaded from the plant transcription factor database [47] and used to make a Hidden Markov Model (HMM) profile by the HMMER package [48]. The HMM was used as a query to identify the MADS-box proteins of *T. dicoccoides* at the 0.001 p-value cut-off (S1 Table). To differentiate type I (M-type) and type II (MIKC-type) MADS-domain proteins, the *T. dicoccoides* MADS-box protein sequences were aligned with all MADS-box proteins of Arabidopsis [24] and rice [49] with MAFFT (L-INS-i strategy) [50] using just the MADS domain part of the sequences. A phylogenetic tree was constructed using IQTREE [51] and ModelFinder [52].

Naming of MIKC-type MADS-box genes

The identified MADS-box genes were named as follows: The name of each *T. dicoccoides* MADS-box gene is composed of the 'Td' prefix which refers to *T. dicoccoides*, plus the name of the most similar *Arabidopsis thaliana* (or *Oryza sativa* in case that the gene was not found in Arabidopsis) gene which was inferred from the phylogenetic analysis (see below), their subgenome location (A or B) and subfamily association. Identical gene names were assigned to the putative homoeologs except for the subgenome identifier (e.g. *TdAG-1A* and *TdAG-1B*). Homoeologs were identified by referring to the EnsemblPlant database. Inparalogs (i.e. duplicated copies) were indicated by consecutive numbers separated by a dash so that the name of the gene with the ID TRIDC3AG061490 is *TdFLC-3A-4* as it is the fourth *TdFLC* gene on 3A chromosome (S1 Table).

Physical characterization of MADS-box proteins

The *T. dicoccoides* annotation file was used to display the structure of the MADS-box genes using the Gene Structure Display Server (GSDS, <http://gsds.cbi.pku.edu.cn>) [53]. The conserved domains of the MADS-box proteins were identified from the Conserved Domain Database (CDD) [54] web server, and the output file was used to visualize the domain structure of the MADS-box proteins in TBtools [55]. The physical map of the MADS-box genes on *T. dicoccoides* chromosomes was generated using shinyCircos2 [56]. The intron ranges and frequencies of MADS-box genes were determined in TBtools (S2 Table).

Maximum likelihood phylogeny of MADS-box proteins

Based on the first phylogeny mentioned above, MADS-box subfamily sequences of *T. dicoccoides*, Arabidopsis [24] and rice proteins [49] were aligned using MAFFT (E-INS-i strategy). Subfamily alignments were then merged using MAFFT (E-INS-i algorithm) [50]. The resulting alignment was trimmed using the kpi-gappy strategy of the ClipKIT tool [57], and a maximum likelihood tree was inferred using the trimmed alignment with IQTREE [51]. The best amino acid substitution model was determined with the ModelFinder option based on the Bayesian information criterion (BIC) and the JTT+F+G4 was chosen [52] and 1000 ultrafast bootstraps

were applied [58]. the MIKC* subclade was set as the outgroup and the generated Newick tree file was visualized in R using the 'ggtree' package [59].

Expression of MADS-box genes

153 samples RNA-seq data generated from 20 different combinations of wild emmer wheat (genotype Zavitan) tissues and developmental stages belonging to root, leaf, flag leaf, flower (anthers and carpels), glume, lemma and palea, grain, and different stages of developing spike were downloaded from SRA database of NCBI (Accession: ERP022006) [60]. After quality control and trimming the low-quality section of reads, the read data from each sample were aligned to the *T. dicoccoides* reference genome using HISAT2, and transcripts assembling and merging were done using StringTie with default settings [61]. Normalization of abundance estimates as FPKM (fragments per kilobase of transcript per million mapped reads) values, for the MADS-box genes were extracted using the ballgown package [62]. A heatmap was produced from $\log_2(\text{FPKM}+1)$ (FPKM: fragments per kilobase of transcript per million fragments mapped) values, of MADS-box genes of *T. dicoccoides* over the developmental stages using the 'pheatmap' package. The co-expression of the MADS-box genes were analyzed by clustering using the R package WGCNA [63].

To assess the TdMADS-box gene response to drought stress, I further used microarray data (Gene Expression Omnibus (GEO) dataSets; accession: GSE31762) from a transcriptome analysis of terminal drought response applied at the inflorescence emergence stage [Zadoks 50–60, 64], after emergence of 1–2 spikes, flag leaf samples were analyzed. The microarray data belonged to two drought tolerant (Y12-3) and drought susceptible (A24-39) genotypes different in their yield and yield stability under drought stress [65]. The orthologous genes of *T. dicoccoides* were identified by Blastn of the common wheat cDNA against the *T. dicoccoides* cDNA sequences. Mean expressions were presented based on transcript per million (TPM) as $\log_2(\text{TPM} + 1)$. Mean expression of MADS-box genes between well-watered and terminal drought conditions was compared using t-test and the bar plots of the differentially expressed genes between the two conditions were produced using the 'ggplot2' package [66].

Cis-regulatory elements of MADS-box genes

The 2-Kb upstream sequences of MADS-box genes were extracted from the *T. dicoccoides* genome using TBtools [55]. The cis-acting elements of the sequences were predicted with the online PlantCARE tool [67].

MicroRNA (miRNA) target of MADS-box genes

Targeting miRNAs of MADS-box genes of *T. dicoccoides* were predicted using the Analysis page on psRNATarget website v2.0 [68]. Both the cDNA sequences of (corresponding to the longest protein variants) and intronic sequences of the TdMADS-box family were uploaded separately. The default parameters were used except that the expected value was set to 1.5. miRNA targets of cDNA and intronic sequences were separately downloaded and presented in an excel data sheet (S3 Table).

Results

Frequency and physical distribution of MADS-box genes in *T. dicoccoides*

Here, a total of 596 transcript variants belonging to 117 MADS-box genes were identified in the wild emmer using the genome assembly WEWSeq_v.1.0 [60]. Only the longest transcript variant from each gene was kept for downstream analysis. The MADS-box genes were named

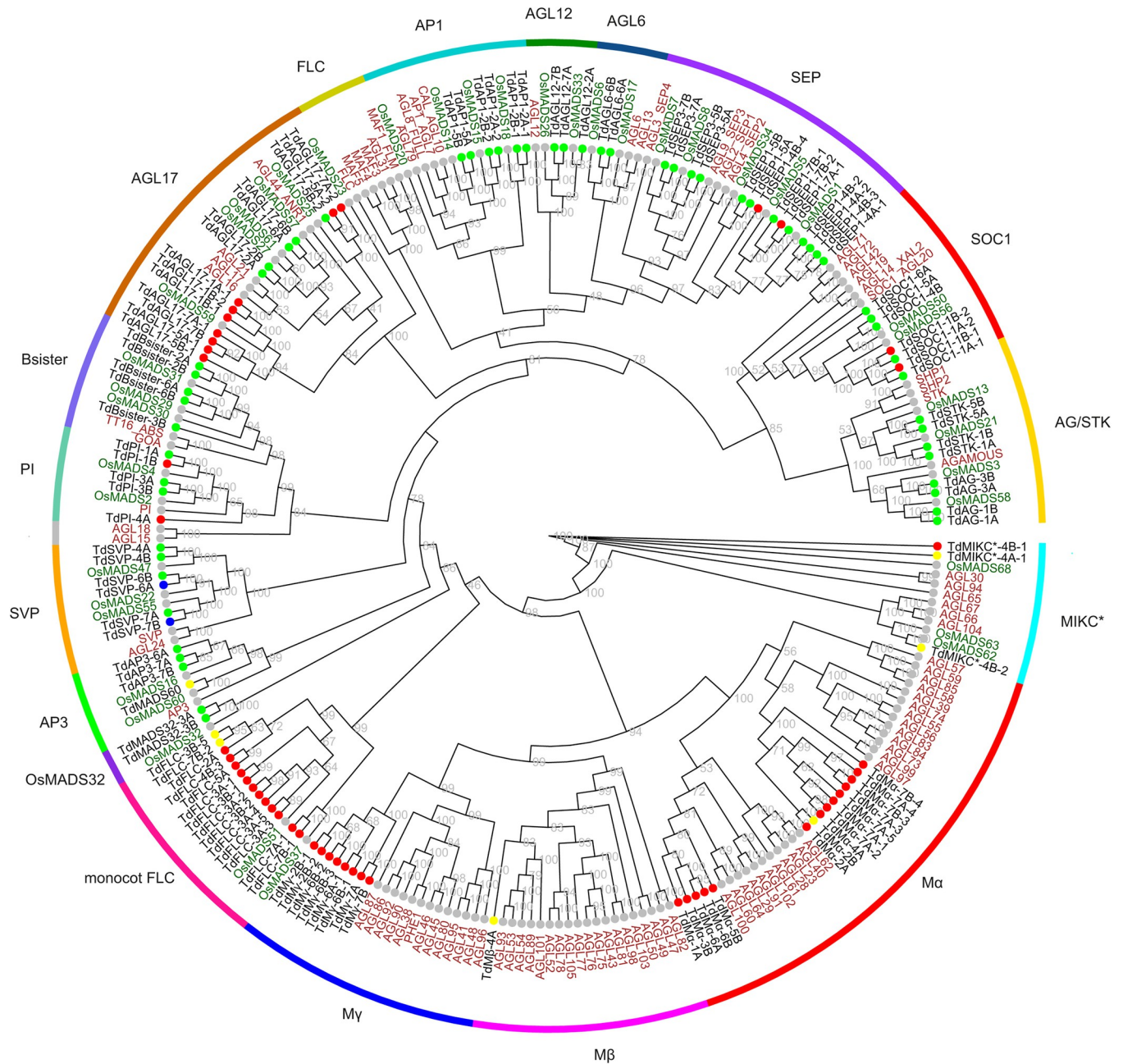


Fig 1. Maximum likelihood phylogeny of MADS-box proteins from wild emmer wheat (*Triticum dicoccoides*), rice (*Oryza sativa*), and Arabidopsis. A phylogenetic unrooted tree of MADS-box proteins from *T. dicoccoides*, rice, and Arabidopsis was inferred using MAFFT-aligned sequences and IQ-Tree [51, 52]. *T. dicoccoides* genes are colored black, whereas rice and Arabidopsis genes are in green and red, respectively. Subfamilies are indicated outside the tree. Dots next to *T. dicoccoides* gene names indicate the presence of MADS-box (red), K-box (blue), or both (green) within the coding region of the gene as detected by CDD. Yellow circles: none was detected. Accession numbers of *T. dicoccoides* genes are available in S1 Table.

<https://doi.org/10.1371/journal.pone.0300159.g001>

according to their subfamily relationship (Fig 1 and S1 Table). The corresponding 117 proteins were classified into 3 major groups i.e. 90 MIKC^C, 3 MIKC*, and 24 M-type based on phylogenetic results. The maximum number of MADS-box genes were found on chromosome 7A which harbored 14 genes followed by 7B with 12 genes, whereas, each of the other chromosomes had 7 or 8 MADS-box genes. MIKC^C-type MADS-box genes were almost randomly

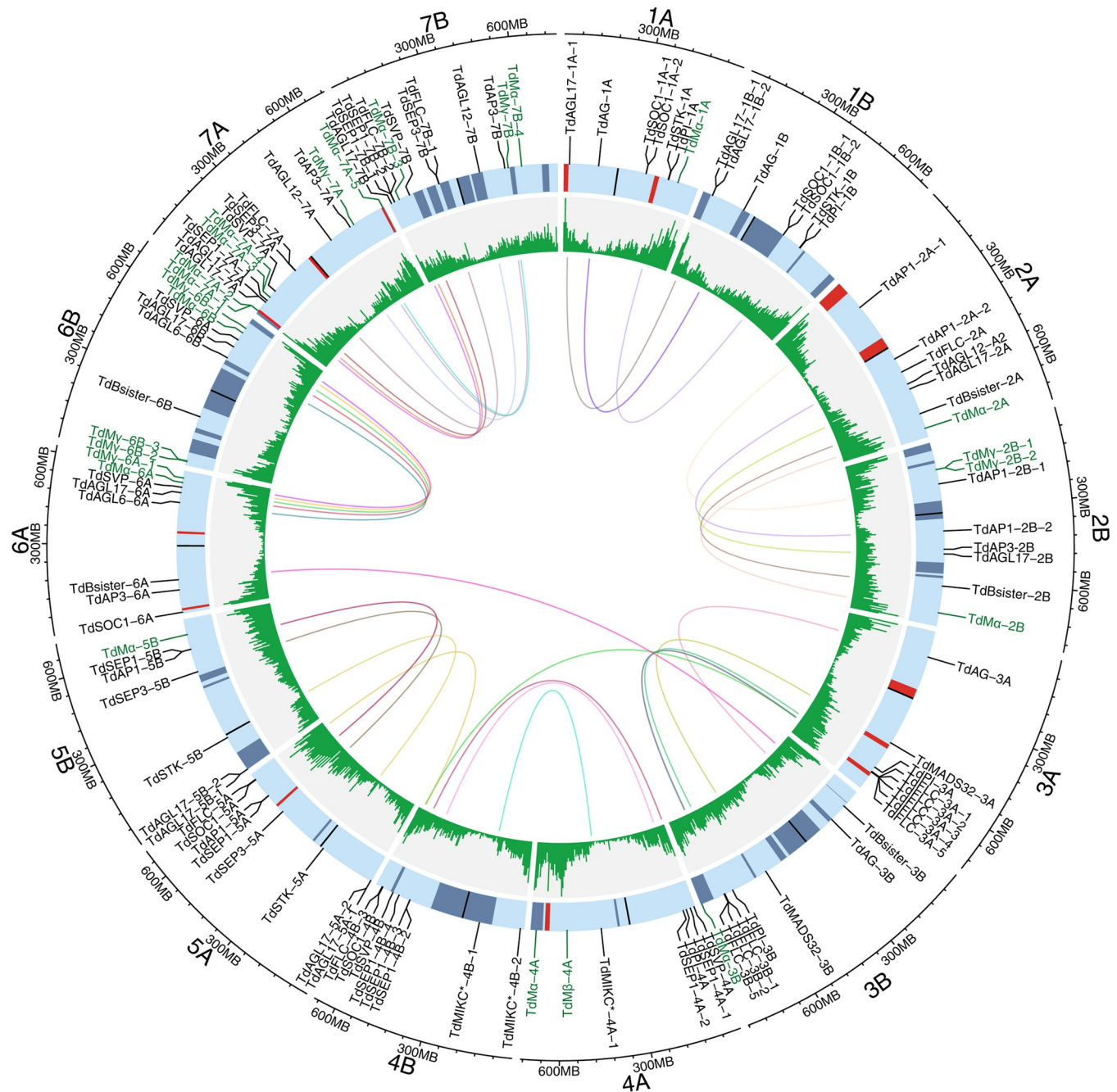


Fig 2. Chromosomal location of MADS-box genes on *T. dicoccoides* genome. The genes were mapped to 14 *T. dicoccoides* chromosomes on which the overall gene density heatmap is presented as well. Chromosome numbers are indicated outside the outer circle. Homoeologous genes are connected using central links. Chromosomes are banded according to pTa535-1 (red bands) and (GAA)₁₀ (blue bands) FISH patterns. M-type MADS-box genes are highlighted with green color.

<https://doi.org/10.1371/journal.pone.0300159.g002>

distributed on all the chromosomes. The 24 copies of M-type genes of wild emmer wheat were distributed over all the chromosomes except that 11 were predominantly located on homoeologous group 7 (Fig 2). The MIKC* genes along with the only Mβ MADS-box gene are located in homoeologous group 4 (Fig 2). None of the M-type genes contained K domain (Fig 1). As mentioned in Introduction, the functionality of the MIKC-type MADS-box genes is mostly

determined by the presence of MADS and K domains. Among the identified MADS-box genes, 58 encodes both MADS and K domains (49.57%), while 50 genes lacked K domain (42.73%), two lacked MADS box (1.71%) and 7 lacked neither MADS nor K domain based on the CDD results under the applied threshold of 0.05. None of the M-type genes contained K domain (Fig 1 and S1 Fig). Other domains also found in some MADS-box genes including DUF6119 (in *TdFLC-3A-2*), PABP (in *TdMy-2B-1*), SNAPc (*TdMy-2B-1*), ARG80 (*TdMa-2A*, *TdMa-7A-3*, *TdMa-7A-4*, *TdMy-7A* and *TdMa-7B-3*), HD-ZIP (*TdSEP1-5A* and *TdSEP1-5B*), TIM (*TdAP3-2B*), HU_IHF (*TdFLC-7B-1*), KLF8 (*TdMa-7B-4*) and SRP54 (*TdPI-4A*) were also found.

Gene structure analysis showed that first or second intron in type II MADS-box genes is considerably longer than the longest intron of MIKC* or M-type genes reaching to about 22 kb in *TdFLC-3A-1* (Fig 3A). The mean number of exons in *T. dicoccoides* MADS-box genes was 1.29 (in M-type genes), 6.47 for MIKC-type genes, and 9.67 in MIKC*-type genes (Fig 3B). However, MIKC^C-type genes were significantly longer than the M- and MIKC* type genes: the mean gene length was 1.00 kb in M-type genes, 10.53 kb in MIKC^C-type genes, and 3.07 kb in MIKC* type genes (Fig 3B). MIKC*-type genes in *T. dicoccoides* have an average number of exons (9.67) almost equal to that of Arabidopsis (10).

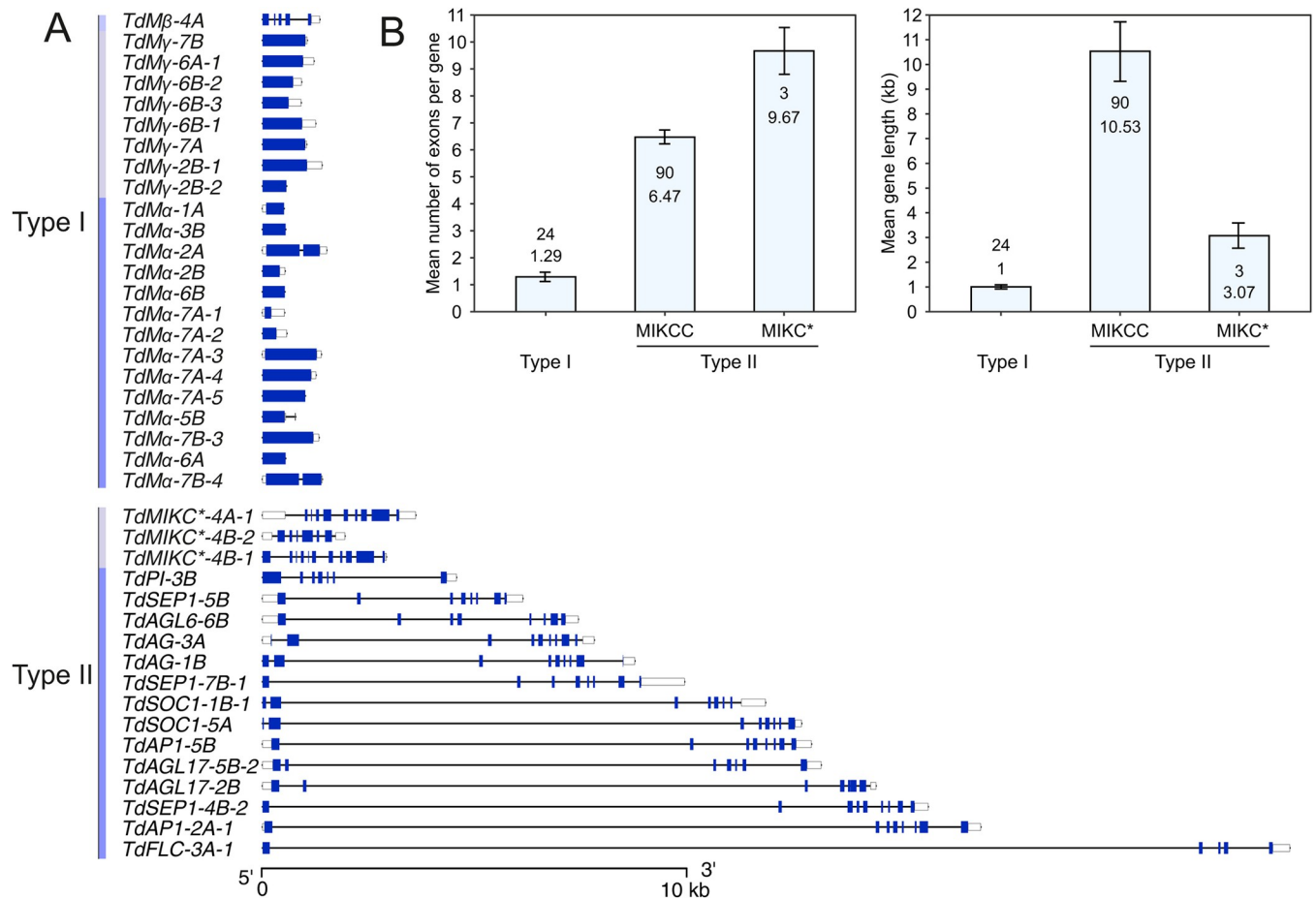


Fig 3. The structure of MADS-box genes in *T. dicoccoides*. A) Comparison of exon-intron structures between type I and a representative sample of type II genes. Exons are in blue; 3' and 5' untranslated regions (UTRs) are shown in white and introns are represented by black lines. B) The mean number of exons and the mean length of MADS-box genes (\pm standard errors). The number of genes in each group is also indicated. Almost all type I MADS-box genes were single exon genes.

<https://doi.org/10.1371/journal.pone.0300159.g003>

Phylogenetic analysis and distribution of MADS-box genes

Based on a maximum likelihood phylogenetic analysis of MADS-box genes from wild emmer wheat, rice, and Arabidopsis, 17 main grass subfamilies of MADS-box gene including *SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1*), *AG/STK* (*AGAMOUS/SEED-STICK*), *SEP1* and *SEP3* (*SEPALLATA*), *AGL6* (*AGAMOUS-LIKE6*), *AGL12*, *API*, *AGL17*, *Bsister*, *PI* (*PISTILLATA*), *SVP* (*SHORT VEGETATIVE PHASE*), *AP3*, *OsMADS32*, monocot and Arabidopsis *FLC* (*FLOWERING LOCUS C*) groups, *Mγ*, *Mβ*, and *Mα* [69, 70] were identified in wild emmer wheat. The rice and wild emmer wheat *FLC* clade composed a distinct clade different from the Arabidopsis *FLC* clade [6, 71] and hence was called monocot *FLC* (Fig 1). The phylogenetic tree shows that *AGAMOUS*, *AGL12*, *API*, *SVP*, *OsMADS32*, and *MIKC** genes of Arabidopsis have conserved sister groups in wild emmer wheat, however, some *SOC1*, *SEP1*, *AGL17*, *FLC* individuals in wild emmer wheat have gained additional copies (7:2 wild emmer wheat to rice copies for *SOC1*, 10:3 for *SEP1*, 14:5 for *AGL17* and 14:2 for monocot *FLC*) probably due to duplication events during evolution. On the other hand, the number of M-type MADS-box genes in wild emmer wheat was considerably lower than that of Arabidopsis (24:56), especially only one distantly related *Mβ* was found in wild emmer wheat compared to 19 orthologous copies of Arabidopsis (Fig 1).

T. dicoccoides contain almost two-fold *MIKC* type MADS-box genes (93) than Arabidopsis with 45 *MIKC*-type genes [24]. When considering the number of *MIKC*-type genes per subgenome, it seems that this significantly higher number is mainly the result of polyploidy because the number of *MIKC*-type MADS-box genes per subgenome in wild emmer (with the two A and B subgenomes) is $93/2 = 46.5$ which is similar to that of rice with 43 and Arabidopsis with 45 Type II MADS-box genes. On the other hand, the number of M-type genes in wild emmer (27) is lower than those of Arabidopsis (62) [24] and rice (32) [49]. None of the M-type MADS-box genes of wild emmer wheat contain K domain and most of the type-I MADS-box genes in wild emmer wheat show zero or very low expression compared to their type II homologs (S2 Fig).

Wild emmer wheat contains 14 *AGL17*-like genes, which is more than two-fold of the six *AGL17*-like genes in rice genome (Fig 1). On the other hand, this number is reasonably lower than two-thirds of the number of common wheat where 47 *AGL17* members have been identified [6]. A two-third ratio is expected in gene number of wild emmer wheat containing A and B subgenomes compared to common wheat containing A, B and D subgenomes. It seems that the higher number of *AGL17* genes in common wheat is the result of their tandem duplications mainly on chromosome 7, resulting the skewed common wheat-to-emmer wheat gene ratio of the *AGL17* genes. Five of the *AGL17* members in *T. dicoccoides* encode both MADS- and K-domain (Fig 1, green dots in *AGL17* clade), and the other nine genes only encode a MADS domain (Fig 1, red dots in *AGL17* clade). *T. dicoccoides* has 16 *FLC* members (Fig 1, monocot *FLC* clade; Fig 2), which is noticeably higher than the two *FLC* genes from rice. Most of wheat *FLC*-like genes (8 out of 14) were located on the long arm of homoeologous group 3 in close vicinity to each other, suggesting the involvement of tandem duplication.

The rice genome contains three *Bsister* paralogs including *OsMADS29*, *OsMADS30*, and *OsMADS31*. In wild emmer wheat, *OsMADS29*-like and *OsMADS31*-like genes were present in syntenic locations in both A and B subgenomes but *OsMADS30* only had one ortholog in wild emmer wheat which was located on B subgenome (Fig 1). All these five *Bsister* members in wild emmer, contained both MADS and K coding domains suggesting retention of a conserved structure and function. A wider spread for *Bsister* members has already been found in common wheat where 27 conserved or truncated *OsMADS30* homologs were dispersedly located in different chromosomes [6]. Such a dispersed distribution homologous genes might

be the result of transposon activity by capturing full or partial gene sequences and transpose them to another location.

In *SEP* clade, two *SEP3* members out of four rice orthologs were assigned to a pair of *T. dicoccoides* homoeologs, resulting in the expected 1:2 ratio. However, the *SEP1* rice genes (i.e. *OsMADS1* and *OsMADS5*) were grouped with 5 (2 + 3) and 3 (2 + 1) wild emmer wheat genes on chromosomes 4 and 7, respectively (Fig 2), suggesting occurrence of gene duplications in wild emmer wheat *SEP1* subclades.

Cis-acting elements in TdMADS-box promoters

To better understand how *T. dicoccoides* MADS-box genes regulate external stimuli, the promoter regions of the 117 *T. dicoccoides* MADS-box genes were analyzed using the PlantCARE database. The analysis detected 3135 cis-acting elements possibly responding to light, hormones, stress, endosperm meristem, etc. (Table 1 and Fig 4). Each of the TdMADS-box genes contained 6 to 48 responsive elements, mainly related to light, however hormone, drought, and low-temperature related cis-acting elements were also present. Promoter analysis further showed that TdMADS-box genes might also be involved in responses to methyl jasmonate (MeJA), ABA, auxin, gibberellin, and salicylic acid. Overall, the results suggest that the TdMADS-box family members generally respond to light and could play a role in hormone responses and abiotic stresses.

miRNAs target analysis

With stringent cut-off expectation threshold of ≤ 0.5 , psRNATarget [68] detected nine MADS-box cDNA target candidates in wild emmer wheat genome. All these cDNA sequences are predicted to be the target for miR444. Two target sites were predicted for the cDNA of each of the *TdAGL17-6B* and *TdAGL17-6A* genes while each of the remaining cDNAs contained only one target site. Furthermore, 41 different miRNA-target sequences were identified on intron sequences of 32 TdMADS-box genes at the same expectation value of 0.5 (S3 Table). At the intron level, some genes for example *TdSOC1-1A-1*, *TdSOC1-1A*, *TdAP3-7A*, *TdAP3-*

Table 1. Cis-acting elements on the promoter region of MADS-box genes in *T. dicoccoides*. Overall cis-acting elements on the 2kb upstream of MADS-box genes related to different stimuli are presented.

Stimulus	Number of elements
Light	1290
Methyl Jasmonate (MeJA)	585
Abscisic acid	419
Auxin	119
Gibberellin	84
Salicylic acid	54
Drought	88
Low temperature	92
Defense stress	33
Anaerobic induction	187
Cell cycle	11
Circadian control	25
Endosperm expression	22
Meristem expression	90
Seed	36
Total	3135

<https://doi.org/10.1371/journal.pone.0300159.t001>

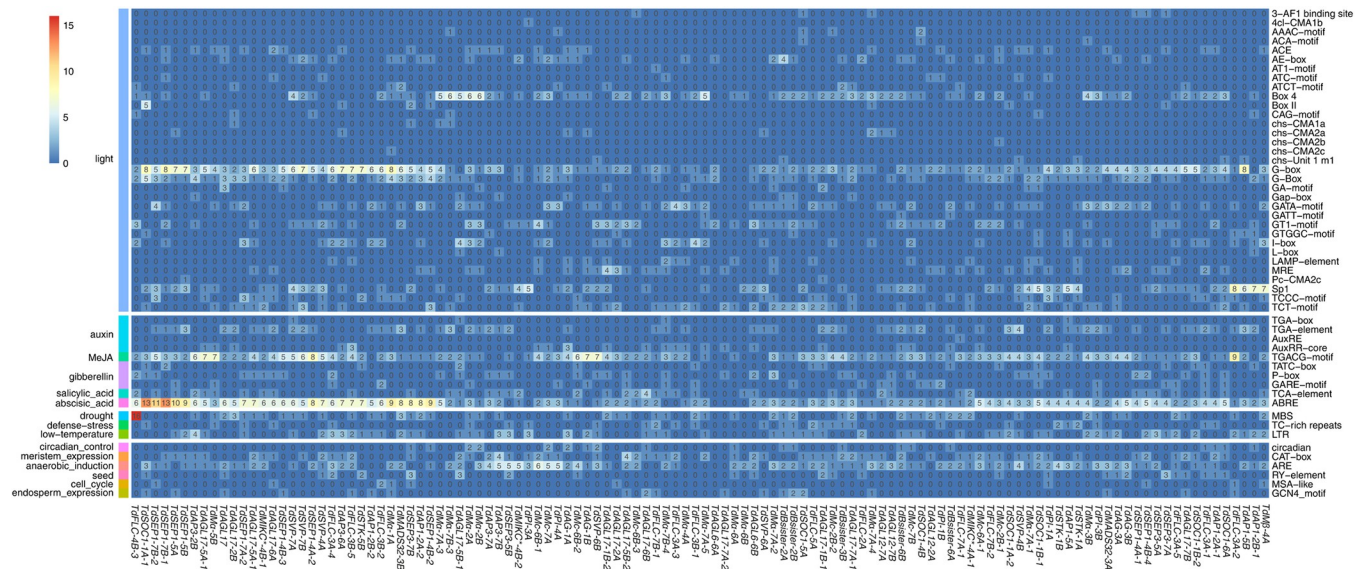


Fig 4. Potential cis-acting elements in promoter region of TdMADS-box genes of *T. dicoccoides*. The number of potential cis-acting elements in 2-kb upstream promoter region of TdMADS-box genes were predicted using the PlantCARE database [67]. The number of each cis-acting element (shown on the right side) identified for each gene is presented inside the cells.

<https://doi.org/10.1371/journal.pone.0300159.g004>

2B, *TdAGL17-2A*, *TdSOC1-1A-2*, *TdSTK-5B*, *TdAGL17-6B*, *TdSVP-4B* and *TdSEP1-4A* contained different miRNA-target sequences.

MADS-box gene expression during developmental stages

153 samples of RNA-seq data from 20 different combinations of wild emmer wheat were analyzed. The samples were from tissues and developmental stages belonging to root, leaf, flag leaf, flower (anthers and carpels), glume, lemma and palea, grain, and different stages of developing spike [60]. The resulting MADS-box gene expression values and modules are presented in S4 and S5 Tables. Out of 117 emmer wheat MADS-box genes, 85 were expressed in at least one developmental stage, with a maximum expression ranging from 1.12 to 7.97 log₂ (FPKM + 1). The maximum expression rates of the remaining 32 genes varied from 0 to 1 log₂ (FPKM + 1) (Fig 5 and S4 Table and S2 Fig). Most of the *AGL17* genes are expressed at zero to low rates except for the *TdAGL17-6A* and *TdAGL17-6B* which are expressed in root, vegetative and reproductive organs. *AG/STK* genes are mainly expressed in flower and grain, *SEP3*, *AGL6*, *PI* and *AP3* genes are expressed in flower and/or grain and to lower extents in developing spike (Fig 5A and 5B). *T. dicoccoides* contains 5 *Bsister* copies that are mainly expressed in flower and grain. It is well known that *Bsister* genes are expressed in ovule and grain with involvement in seed development [41, 72, 73]. In total, the type II MADS-box expression patterns of wild emmer wheat are similar to those of common wheat [6] and rice [49]. M-type MADS-box genes showed zero or week expression levels in wild emmer. Out of 16 M-type TdMADS-box genes, 14 expressed only in grain (123 days from sowing) at a maximum expression rate of 1.25 log₂ (FPKM + 1). *TdMβ-4A* and *TdMγ-6B-3* were also expressed in flowers (S4 Table and S2 Fig).

Nine different expression modules were detected following co-expression analysis of the MADS-box gene. The expression patterns in the resulting modules (Fig 5B) generally showed similarity to the expression of MADS-box genes subfamilies (Fig 5A). For example, *AG/STK* members were grouped into two adjacent modules. Similarly, most M-type genes were

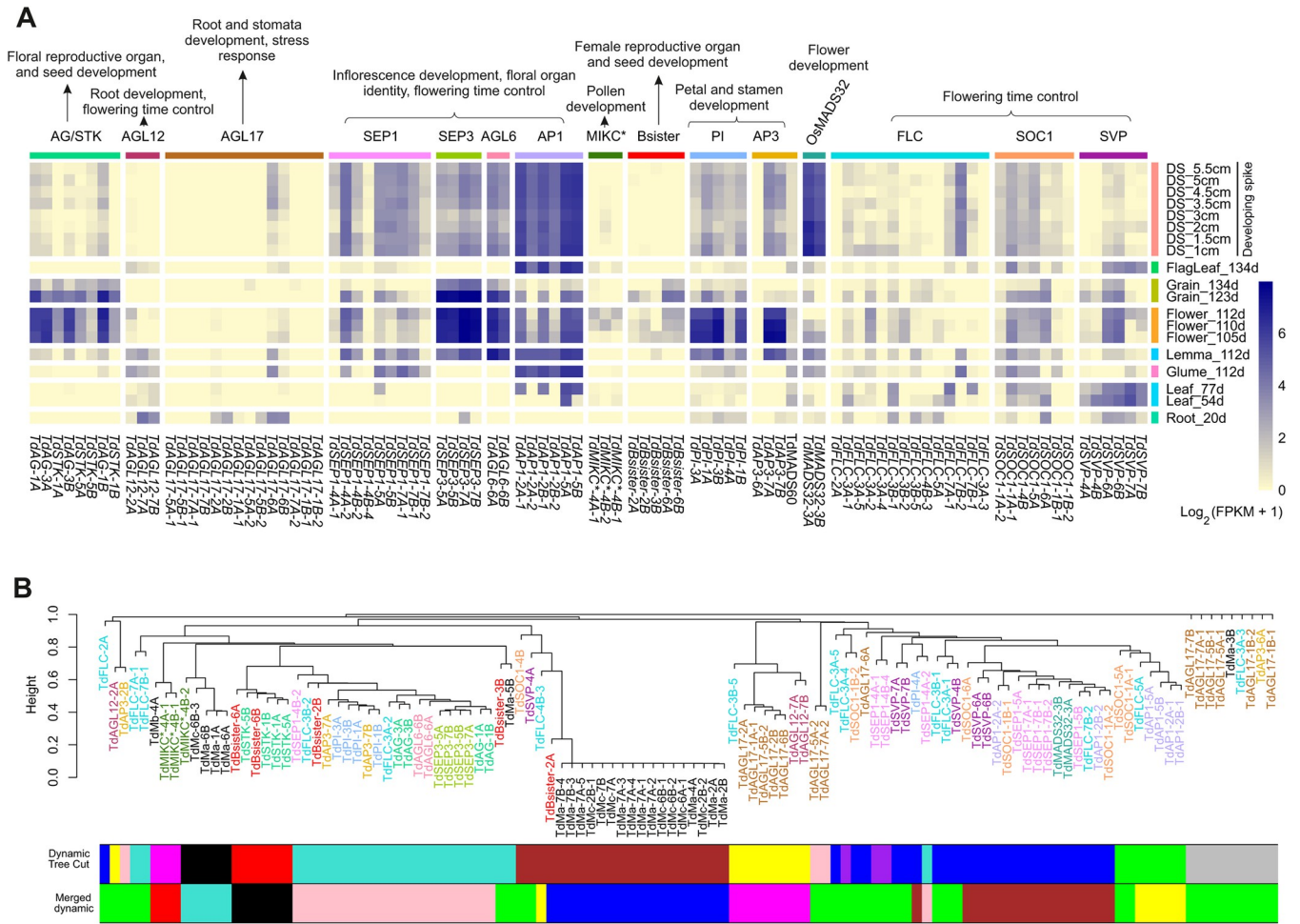


Fig 5. Expression analysis of the *T. dicoccoides* MADS-box genes as revealed by RNA-seq data. A) A heat map of mean expression of type-II MADS-box genes in different tissues and developmental stages of *T. dicoccoides*. Numbers followed by the developing stages are days from sowing (d) or spike length (cm). B) Co-expression clustering of the *T. dicoccoides* MADS-box genes based on their expression values from different tissues and developmental stages. Colors indicate the different modules. Note that M-type MADS-box genes were not presented in 'A' but they were included in co-expression pattern analysis in 'B'.

<https://doi.org/10.1371/journal.pone.0300159.g005>

grouped into a single module (Fig 5B) which indicated no or very low expression pattern (S4 Table). Genes from some subfamilies showed considerable differences in their expression patterns. For example, members of FLC and SEP subfamilies have been located in different modules.

From the microarray data, 7 differentially expressed MADS-box genes were found under drought stress in at least one of the two evaluated genotypes, among which, only *TdSOC1-6A* upregulated in both drought susceptible and drought tolerant genotypes under drought conditions while *TdPI-1A*, *TdSOC1-1A-1*, *TdSOC1-6A* and *TdAGL12-7A* differentially expressed only in the tolerant genotype (Fig 6 and S6 Table).

Discussion

The conserved function of TdMADS-box genes

MADS-box transcription factors play important roles in various processes of plant development, such as floral organ identity determination, flower development, and seed formation.

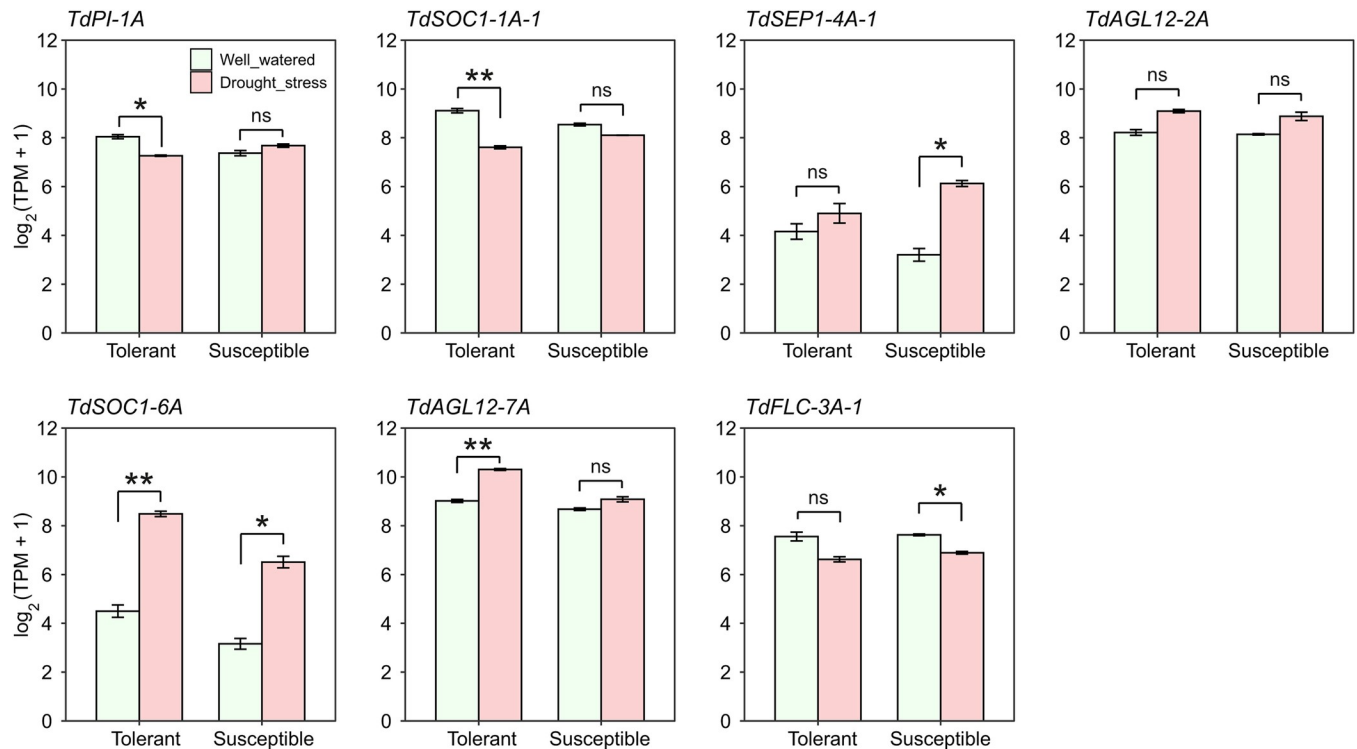


Fig 6. Expression of MADS-box genes of wild emmer wheat under drought stress. Mean expression (\pm standard error) of MADS-box genes of wild emmer wheat which differentially responded to drought stress conditions as revealed by microarray data (GEO accession: GSE31762). The microarray data belongs to the flag leaf of two wild emmer wheat genotypes contrasting in their productivity and yield stability under terminal drought stress.

<https://doi.org/10.1371/journal.pone.0300159.g006>

They are also involved in responding to environmental stresses. Here, I identified 117 MADS-box genes in the wild emmer wheat genome which is an important source for wheat improvement. Phylogenetic analysis along with the MADS-box genes of rice and Arabidopsis assigned emmer wheat MADS-box genes to 17 (14 MIKC-type and 3 M-type) subfamilies (Fig 1 and S1 Table).

In general, a high similarity in the expression pattern between wild emmer wheat MADS-box genes, common wheat [6, 74] and rice [49] orthologs was found indicating a conserved functionality of MADS-box genes between these species. In wild emmer wheat, *TRID-C5AG057030* (named *TdAPI-5A*) and *TRIDC5BG061170* (named *TdAPI-5B*) are vernalization *VRN-A1* and *VRN-B1* genes respectively. *TRIDC2AG022240* (*TdAPI-2A-1*) and *TRIDC2BG025920* (*TdAPI-2B-1*) of wild emmer wheat is co-expressed with *Vrn1-5A* (Fig 5A and S4 Table), which indicates that this gene may also be related to flowering time. Similarly, in common wheat, *TRAESCS2D02G181400* which is orthologous to *TRIDC2AG022240* (*TdAPI-2A-1*) and *TRIDC2BG025920* (*TdAPI-2B-1*) of wild emmer wheat encodes a MIKC-type MADS-box transcription factor and is co-expressed with *Vrn1-5A* [75]. Different alleles and copy number variation of *VRN* genes involve in the transition of the shoot apical meristem to the reproductive phase [76–78]. The spring forms of emmer wheat are associated with the independent emergence of a new dominant *VRN-A1* allele which resulted from changes in the promoter region and a large deletion in the first intron [77, 79]. The wild-type *VRN1* allele for winter growth habit requires long exposures to low temperatures (vernalization) to be expressed, so *VRN1* has a pivotal role in the determination of flowering time.

Interestingly, the number of M-type genes in wild emmer wheat (27) is significantly lower than that of Arabidopsis (62) [24]. None of the M-type genes contained K domain. Truncated

genes are common among M-type genes also they may be functional. In *A. thaliana* some M-type MADS-box genes are important for normal development of the female gametophyte or endosperm and may be responsible for post-zygotic lethality in interspecific hybrids [25–31]. Out of 16 M-type TdMADS-box genes of wild emmer wheat, 14 were expressed, albeit at very low rates and mostly only in grain (123 days from sowing) with a maximum expression rate of $1.25 \log_2$ (FPKM + 1). *TdMβ-4A* and *TdMγ-6B-3* were also expressed in flower (S4 Table and S2 Fig). Similarly in common wheat, almost 75% of non-expressed MADS-box genes were members of the type I clade [4]. In agreement with the results obtained here, Nam, Kim [33] found a higher proportion of nonfunctional genes among the type I MADS-box group and suggested that type I genes have undergone a higher rate of birth-and-death evolution than type II genes in angiosperms which might be the result of more frequent segmental duplications and less purifying selection of type I than in type II genes [33].

Gene duplication versus environmental stresses

Tandem duplicates may be correlated to the adaptation to different environments [80]. Duplications of large chromosomal segments i.e. segmental duplications in most cases appear to have come from one round of polyploidy [81]. *T. dicoccoides* is the oldest polyploid wheat and the ancestral species of common wheat. By comparing to diploid and hexaploid *Triticum* species, it provides an opportunity to study the MADS-box gene family members during polyploidization. Contrary to common wheat which has undergone extensive expansion of some MIKC-type subfamilies [4, 6], the number of MIKC-type MADS-box genes per subgenome in wild emmer wheat is generally comparable to that of rice and Arabidopsis. The number of MIKC-type MADS-box genes per subgenome in wild emmer is $93/2 = 46.5$ compared to that of rice (43) and Arabidopsis (45) Type II MADS-box genes. However, some subfamilies including *SOC1*, *SEPI*, *AGL17*, and *FLC* showed moderate to high rates of duplication per subgenome compared to the rice genome (7:2 copy ratio for *SOC1*, 10:3 for *SEPI*, 14:5 for *AGL17* and 14:2 for *FLC*). It has been suggested that the expansion of eudicot *FLC* genes potentially enables the ability to adapt to various environmental conditions including ambient temperatures [82]. The high level of duplication of *FLC* genes in *T. dicoccoides* may similarly contribute to its adaptation to different environments by altering its flowering time [12]. *FLC* plays a crucial role in regulating the flowering time in plants. *FLC* represses flowering transition by repressing promoters of flowering genes, such as *FT* and *SOC1*. During vernalization, *FLC* protein levels decrease and therefore flowering is induced [83–85]. The analysis showed that the Arabidopsis *FLC* clade composed a distinct clade different from the rice and emmer wheat (monocot) *FLC* clade (Fig 1) [6, 71]. The presence of *FLC-like* genes in cereals was unknown for a long time, even though there was a lot of information about Arabidopsis *FLC*. It has been suggested that mechanisms developmental and flowering time regulation in monocots compared to eudicots and thought that that *FLC* only existed in eudicot plants [Reviewed in 86]. But the synteny analysis and phylogeny has been proven that *FLC* relatives are present in cereals which are related to the *FLC* genes of Arabidopsis [71]. There are two subclades within this monocot *FLC* group, called the *OsMADS51* and *OsMADS37* subclades as these rice genes were located within each group (Fig 1).

Some MADS-box genes are expressed in response to drought stress

I observed an upregulation in response to stresses for some MIKC-type MADS-box genes. Specifically, *TdPI-1A*, *TdSOC1-1A-1*, *TdSEP1-4A-1*, *TdSOC1-6A*, *TdAGL-12-7A*, and *TdFLC-3A-1* differentially responded to drought stress condition as revealed by microarray data (Fig 6). Studies have shown that some MADS-box genes such as *AGL12* and *MBP8* have a negative

role in drought [11–13] and cold [14] tolerance by regulating the expression of genes involved in stress response pathways. In rice, overexpression of the *TdAGL12-7A* ortholog (i.e. *MADS26*) is possibly connected to response to stresses.

OsMADS26-down-regulated plants also have shown enhanced resistance against two rice pathogens. In spite of this improved resistance under biotic stresses, *OsMADS26*-down-regulated plants also showed more tolerance to drought stress in both controlled and field conditions without a strong impact on plant development [15]. Other MADS-box genes might also be involved in abiotic stress in plants. For example, it has been shown that nitrate-dependent salt tolerance is mediated by *OsMADS27* in rice (orthologous to *TdAGL17-2A* and *TdAGL17-2B* of wild emmer wheat) where the expression of *OsMADS27* was specifically induced by nitrate [16]. In Arabidopsis, *SVP* is also a major regulator of ABA catabolism and *SVP*, *CYP707A1/3*, and *AtBG1* together are involved in plant response to water stress and plant drought resistance [17].

Cis-regulatory elements and introns of MADS-box genes may contribute to environmental adaptation

The analysis of cis-acting elements in TdMADS-box promoters suggests that the TdMADS-box family generally responds to light and could play a role in hormone responses and abiotic stresses. Mutations in the *VRN1* (*API*) promoter region or deletions in its first intron results in a spring growth habit as the vernalization is not required for flowering [87]. Insertion of a GATA box like sequence at the promoter region of the *VRN-A3* locus in a cultivated emmer wheat genotype (*Triticum turgidum* L. ssp. *dicoccum*) confers early flowering trait [88].

miRNAs are 20–24 nucleotides in size and promote degradation or repression of translation of target mRNAs, hereby negatively regulate gene expression at post-transcriptional level. miR444 is a monocot-specific microRNA. It has been shown that miR444 is a key factor for virus resistance via RNA-silencing in rice. miR444 reduces the repressive roles of *OsMADS23*, *OsMADS27a*, and *OsMADS57* on *OsRDR1* transcription, thus the *OsRDR1*-dependent antiviral RNA-silencing pathway is activated [89]. Similarly, miR444 also plays a role in rice tillering [90]. miR444 and its target *OsMADS27* TF are also involved in NO₃-dependent root development [91]. NO₃⁻ depression induces miR444 expression, and the expression of a miR444 target can quench the miRNA and act as a sponge in transgenic rice lines resulting in increased total root growth [92]. At the intron level, some genes for example *TdSOC1-1A-1*, *TdSOC1-1A*, *TdAP3-7A*, *TdAP3-2B*, *TdAGL17-2A*, *TdSOC1-1A-2*, *TdSTK-5B*, *TdAGL17-6B*, *TdSVP-4B* and *TdSEPI-4A* contained different miRNA-target sequences. Intronic miRNAs are transcribed from introns of protein-coding genes. They have been shown to be involved in post-transcriptional regulation of gene expression [93]. Furthermore, intron sequences can form circular RNA. circular RNA containing miRNA sequences can regulate the expression of mRNAs by acting as miRNA sponges as well [94]. It has been shown that the miRNA444 is upregulated in *T. aestivum* under salt stress. Similarly, the miR1120 which was identified on the intron of *TdAGL17-2A* and *TdSOC1-1A-1* is upregulated under salt stress in *T. dicoccoides* [95].

Conclusion

There are evidences about the involvement of the MADS-box genes in plant growth, development and stress tolerance, however, the detailed information on MADS-box gene family was not available in wild emmer wheat. Here, a genome-wide analysis showed that the MADS-box genes in wild emmer wheat especially MIKC-type clades have retained conserved functionality. MADS-box genes in wild emmer wheat have promoters responsive to various stimuli and play important roles in growth and development and response to stresses. The specific

adjustment via gene duplication, the alterations in expression patterns under various conditions such as photoperiod, temperature, and stresses, and promoter and intronic sequence evolution have all contributed to fine-tuning the MADS-box gene functionality. The results provide comprehensive information about the MADS-box genes of wild emmer wheat that could accelerate functional genomics efforts and potentially facilitate bridging gaps toward breeding for agronomically important traits in wheat.

Supporting information

S1 Fig. Domain structure of *T. dicoccoides* MADS-box proteins. Conserved domain predictions for the 117 *T. dicoccoides* MADS-box proteins are presented.

(PDF)

S2 Fig. Expression heatmap of *T. dicoccoides* MADS-box genes. The heat map represents mean expression of 117 *T. dicoccoides* MADS-box genes in different tissues and developmental stages.

(PDF)

S1 Table. The MADS-box gene description and protein sequences. The table contains MADS-box gene description and protein sequences of *T. dicoccoides*, *Oryza sativa*, and *Arabidopsis thaliana*.

(XLSX)

S2 Table. Exon, intron, and CDS boundaries of the *T. dicoccoides* MADS-box genes.

(XLSX)

S3 Table. Predicted miRNAs targeting MADS-box gene family in *T. dicoccoides*.

(XLSX)

S4 Table. Expression data of *T. dicoccoides* MADS-box genes. *T. dicoccoides* MADS-box gene expression data based on $\log_2(\text{FPKM} + 1)$ in different tissues and developmental stages.

(XLSX)

S5 Table. Expression module data for *T. dicoccoides* MADS-box genes related to Fig 5.

(XLSX)

S6 Table. Expression data of *T. dicoccoides* MADS-box genes. The expression data of *T. dicoccoides* MADS-box gene which differentially responded to drought stress conditions as revealed by microarray data.

(XLSX)

Author Contributions

Conceptualization: Ghader Mirzaghaderi.

Formal analysis: Ghader Mirzaghaderi.

Funding acquisition: Ghader Mirzaghaderi.

Investigation: Ghader Mirzaghaderi.

Methodology: Ghader Mirzaghaderi.

Writing – original draft: Ghader Mirzaghaderi.

References

1. Shewry PR. Wheat. *J Exp Bot*. 2009; 60:1537–53. <https://doi.org/10.1093/jxb/erp058> PMID: 19386614
2. Mirzaghaderi G, Abdolmalaki Z, Ebrahimzadegan R, Bahmani F, Orooji F, Majdi M, et al. Production of synthetic wheat lines to exploit the genetic diversity of emmer wheat and D genome containing Aegilops species in wheat breeding. *Scientific Reports*. 2020; 10(1):19698. <https://doi.org/10.1038/s41598-020-76475-7> PMID: 33184344
3. Xie W, Nevo E. Wild emmer: genetic resources, gene mapping and potential for wheat improvement. *Euphytica*. 2008; 164:603–14.
4. Raza Q, Riaz A, Atif RM, Hussain B, Rana IA, Ali Z, et al. Genome-wide diversity of MADS-Box genes in bread wheat is associated with its rapid global adaptability. *Frontiers in Genetics*. 2022; 12:818880. <https://doi.org/10.3389/fgene.2021.818880> PMID: 35111207
5. Ma J, Yang Y, Luo W, Yang C, Ding P, Liu Y, et al. Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum* L.). *PLoS One*. 2017; 12(7):e0181443.
6. Schilling S, Kennedy A, Pan S, Jermiin LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytologist*. 2020; 225(1):511–29. <https://doi.org/10.1111/nph.16122> PMID: 31418861
7. Chen L, Ren J, Shi H, Zhang Y, You Y, Fan J, et al. *TdCBL6*, a calcineurin B-like gene from wild emmer wheat (*Triticum dicoccoides*), is involved in response to salt and low-K+ stresses. *Molecular breeding*. 2015; 35:1–12.
8. Kuzuoglu-Ozturk D, Cebeci Yalcinkaya O, Akpinar BA, Mitou G, Korkmaz G, Gozuacik D, et al. Autophagy-related gene, *TdAtg8*, in wild emmer wheat plays a role in drought and osmotic stress response. *Planta*. 2012; 236:1081–92.
9. Lucas S, Durmaz E, Akpinar BA, Budak H. The drought response displayed by a DRE-binding protein from *Triticum dicoccoides*. *Plant physiology and biochemistry: PPB*. 2011; 49(3):346–51. Epub 2011/02/08. <https://doi.org/10.1016/j.plaphy.2011.01.016> PMID: 21296583.
10. Lucas S, Dogan E, Budak H. TMPIT1 from wild emmer wheat: first characterisation of a stress-inducible integral membrane protein. *Gene*. 2011; 483(1–2):22–8. <https://doi.org/10.1016/j.gene.2011.05.003> PMID: 21635942
11. Zhao W, Zhang L-L, Xu Z-S, Fu L, Pang H-X, Ma Y-Z, et al. Genome-wide analysis of MADS-Box genes in foxtail millet (*Setaria italica* L.) and functional assessment of the role of SIMADS51 in the drought stress response. *Frontiers in Plant Science*. 2021; 12:659474.
12. Castelán-Muñoz N, Herrera J, Cajero-Sánchez W, Arrizubieta M, Trejo C, García-Ponce B, et al. MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. *Frontiers in Plant Science*. 2019; 10:853. <https://doi.org/10.3389/fpls.2019.00853> PMID: 31354752
13. Yin W, Hu Z, Hu J, Zhu Z, Yu X, Cui B, et al. Tomato (*Solanum lycopersicum*) MADS-box transcription factor SIMBP8 regulates drought, salt tolerance and stress-related genes. *Plant Growth Regulation*. 2017; 83(1):55–68. <https://doi.org/10.1007/s10725-017-0283-2>
14. Voogd C, Brian LA, Wu R, Wang T, Allan AC, Varkonyi-Gasic E. A MADS-box gene with similarity to FLC is induced by cold and correlated with epigenetic changes to control budbreak in kiwifruit. *New Phytologist*. 2022; 233(5):2111–26. <https://doi.org/10.1111/nph.17916> PMID: 34907541
15. Khong GN, Pati PK, Richaud F, Parizot B, Bidzinski P, Mai CD, et al. OsMADS26 negatively regulates resistance to pathogens and drought tolerance in rice. *Plant Physiology*. 2015; 169(4):2935–49. <https://doi.org/10.1104/pp.15.01192> PMID: 26424158
16. Alfatih A, Zhang J, Song Y, Jan SU, Zhang Z-S, Xia J-Q, et al. Nitrate-responsive OsMADS27 promotes salt tolerance in rice. *Plant Communications*. 2023; 4(2). <https://doi.org/10.1016/j.xplc.2022.100458> PMID: 36199247
17. Wang Z, Wang F, Hong Y, Yao J, Ren Z, Shi H, et al. The flowering repressor SVP confers drought resistance in Arabidopsis by regulating abscisic acid catabolism. *Molecular Plant*. 2018; 11(9):1184–97.
18. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences*. 2000; 97(10):5328–33. <https://doi.org/10.1073/pnas.97.10.5328> PMID: 10805792
19. Kaufmann K, Melzer R, Theißen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*. 2005; 347(2):183–98. <https://doi.org/10.1016/j.gene.2004.12.014> PMID: 15777618
20. Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes & Development*. 1991; 5(3):484–95. <https://doi.org/10.1101/gad.5.3.484> PMID: 1672119

21. Riechmann JL, Krizek BA, Meyerowitz EM. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences*. 1996; 93(10):4793–8. <https://doi.org/10.1073/pnas.93.10.4793> PMID: 8643482
22. Egea-Cortines M, Saedler H, Sommer H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *The EMBO journal*. 1999; 18(19):5370–9.
23. van Dijk ADJ, Morabito G, Fiers M, van Ham RCHJ, Angenent GC, Immink RGH. Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. *PLOS Computational Biology*. 2010; 6(11):e1001017. <https://doi.org/10.1371/journal.pcbi.1001017> PMID: 21124869
24. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *The Plant Cell*. 2003; 15(7):1538–51. <https://doi.org/10.1105/tpc.011544> PMID: 12837945
25. Köhler C, Page DR, Gagliardini V, Grossniklaus U. The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nature Genetics*. 2005; 37(1):28–30. <https://doi.org/10.1038/ng1495> PMID: 15619622
26. Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes and Development*. 2003; 17(12):1540–53. <https://doi.org/10.1101/gad.257403> PMID: 12815071
27. Portereiko MF, Lloyd A, Steffen JG, Punwani JA, Otsuga D, Drews GN. AGL80 is required for central cell and endosperm development in Arabidopsis. *Plant Cell*. 2006; 18(8):1862–72. Epub 2006/06/27. <https://doi.org/10.1105/tpc.106.040824> PMID: 16798889; PubMed Central PMCID: PMC1533969.
28. Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC. The MADS Domain protein DIANA Acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *The Plant Cell*. 2008; 20(8):2088–101. <https://doi.org/10.1105/tpc.108.058958> PMID: 18713950
29. Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L. AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *The Plant Journal*. 2008; 54(6):1037–48. Epub 2008/03/19. <https://doi.org/10.1111/j.1365-313X.2008.03485.x> PMID: 18346189.
30. Kang I-H, Steffen JG, Portereiko MF, Lloyd A, Drews GN. The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis *The Plant Cell*. 2008; 20(3):635–47. <https://doi.org/10.1105/tpc.107.055137> PMID: 18334668
31. Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L. Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Current Biology*. 2009; 19(13):1128–32. <https://doi.org/10.1016/j.cub.2009.05.068> PMID: 19559614
32. Hasebe M, Wen CK, Kato M, Banks JA. Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(11):6222–7. Epub 1998/05/30. <https://doi.org/10.1073/pnas.95.11.6222> PMID: 9600946; PubMed Central PMCID: PMC27636.
33. Nam J, Kim J, Lee S, An G, Ma H, Nei M. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(7):1910–5. Epub 2004/02/07. <https://doi.org/10.1073/pnas.0308430100> PMID: 14764899; PubMed Central PMCID: PMC357026.
34. Kwantes M, Liebsch D, Verelst W. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. *Molecular Biology and Evolution*. 2012; 29(1):293–302. Epub 2011/08/05. <https://doi.org/10.1093/molbev/msr200> PMID: 21813465.
35. Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, et al. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Molecular Biology and Evolution*. 2003; 20(12):1963–77. Epub 2003/09/02. <https://doi.org/10.1093/molbev/msg216> PMID: 12949148.
36. Verelst W, Twell D, de Folter S, Immink R, Saedler H, Münster T. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology*. 2007; 8(11):R249. Epub 2007/11/24. <https://doi.org/10.1186/gb-2007-8-11-r249> PMID: 18034896; PubMed Central PMCID: PMC2258202.
37. Adamczyk BJ, Fernandez DE. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiology*. 2009; 149(4):1713–23. Epub 2009/02/13. <https://doi.org/10.1104/pp.109.135806> PMID: 19211705; PubMed Central PMCID: PMC2663741.
38. Zobell O, Faigl W, Saedler H, Münster T. MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Molecular Biology and Evolution*. 2010; 27(5):1201–11. Epub 2010/01/19. <https://doi.org/10.1093/molbev/msq005> PMID: 20080864.
39. Li C, Lu X, Xu J, Liu Y. Regulation of fruit ripening by MADS-box transcription factors. *Scientia Horticulturae*. 2023; 314:111950.

40. Shah L, Sohail A, Ahmad R, Cheng S, Cao L, Wu W. The roles of MADS-Box genes from root growth to maturity in *Arabidopsis* and rice. *Agronomy*. 2022; 12(3):582.
41. Callens C, Tucker MR, Zhang D, Wilson ZA. Dissecting the role of MADS-box genes in monocot floral development and diversity. *Journal of Experimental Botany*. 2018; 69(10):2435–59. <https://doi.org/10.1093/jxb/ery086> PMID: 29718461
42. Goslin K, Zheng B, Serrano-Mislata A, Rae L, Ryan PT, Kwaśniewska K, et al. Transcription factor interplay between LEAFY and APETALA1/CAULIFLOWER during floral initiation. *Plant Physiology*. 2017; 174(2):1097–109. <https://doi.org/10.1104/pp.17.00098> PMID: 28385730
43. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The SEP4 gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology*. 2004; 14(21):1935–40. Epub 2004/11/09. <https://doi.org/10.1016/j.cub.2004.10.028> PMID: 15530395.
44. Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*. 2001; 409(6819):525–9. <https://doi.org/10.1038/35054083> PMID: 11206550
45. Alvarez-Buylla ER, García-Ponce B, Sánchez MdIP, Espinosa-Soto C, García-Gómez ML, Piñeyro-Nelson A, et al. MADS-box genes underground becoming mainstream: plant root developmental mechanisms. *New Phytologist*. 2019; 223(3):1143–58. <https://doi.org/10.1111/nph.15793> PMID: 30883818
46. Zhao P-X, Zhang J, Chen S-Y, Wu J, Xia J-Q, Sun L-Q, et al. *Arabidopsis* MADS-box factor AGL16 is a negative regulator of plant response to salt stress by downregulating salt-responsive genes. *New Phytologist*. 2021; 232(6):2418–39. <https://doi.org/10.1111/nph.17760> PMID: 34605021
47. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*. 2016; 45(D1):D1040–D5. <https://doi.org/10.1093/nar/gkw982> PMID: 27924042
48. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Research*. 2018; 46(W1):W200–W4. <https://doi.org/10.1093/nar/gky448> PMID: 29905871
49. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*. 2007; 8(1):1–21. <https://doi.org/10.1186/1471-2164-8-242> PMID: 17640358
50. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2017; 20(4):1160–6. <https://doi.org/10.1093/bib/bbx108> PMID: 28968734
51. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. 2014; 32(1):268–74. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
52. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. 2017; 14(6):587–9. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
53. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*. 2015; 31:1296–7. <https://doi.org/10.1093/bioinformatics/btu817> PMID: 25504850
54. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Research*. 2019; 47:D427–D32. <https://doi.org/10.1093/nar/gky995> PMID: 30357350
55. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*. 2020; 13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009> PMID: 32585190
56. Wang Y, Jia L, Tian G, Dong Y, Zhang X, Zhou Z, et al. shinyCircos-V2.0: Leveraging the creation of Circos plot with enhanced usability and advanced features. *iMeta*. 2023; 2(2):e109. <https://doi.org/10.1002/imt2.109>
57. Steenwyk JL, Buida TJ, III, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology*. 2020; 18(12):e3001007. <https://doi.org/10.1371/journal.pbio.3001007> PMID: 33264284
58. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*. 2013; 30:1188–95. <https://doi.org/10.1093/molbev/mst024> PMID: 23418397
59. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY, Evolution. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology*. 2017; 8:28–36.
60. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*. 2017; 357(6346):93–7. <https://doi.org/10.1126/science.aan0032> PMID: 28684525

61. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*. 2016; 11:1650.
62. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology*. 2015; 33:243. <https://doi.org/10.1038/nbt.3172> PMID: 25748911
63. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9(1):1–13.
64. Zadoks JC, Chang TT, Konzak CF. Decimal code for growth stages of cereals. *Weed Research*. 1974; 14:415–21.
65. Krugman T, Chagué V, Peleg Z, Balzergue S, Just J, Korol AB, et al. Multilevel regulation and signalling processes associated with adaptation to terminal drought in wild emmer wheat. *Functional and Integrative Genomics*. 2010; 10(2):167–86. <https://doi.org/10.1007/s10142-010-0166-3> PMID: 20333536
66. Wickham H. ggplot2. *Wiley interdisciplinary reviews: Computational Statistics*. 2011; 3(2):180–5.
67. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 2002; 30(1):325–7. Epub 2001/12/26. <https://doi.org/10.1093/nar/30.1.325> PMID: 11752327; PubMed Central PMCID: PMC99092.
68. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res*. 2018; 46(W1):W49–w54. Epub 2018/05/03. <https://doi.org/10.1093/nar/gky316> PMID: 29718424; PubMed Central PMCID: PMC6030838.
69. Gramzow L, Theissen G. Phylogenomics reveals surprising sets of essential and dispensable clades of MIKC-group MADS-box genes in flowering plants. *Journal of Experimental Zoology, Part B: Molecular and Developmental Evolution*. 2015; 324:353–62.
70. Ng M, Yanofsky MF. Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics*. 2001; 2(3):186–95. <https://doi.org/10.1038/35056041> PMID: 11256070
71. Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K. FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nature Communications*. 2013; 4:2280. Epub 2013/08/21. <https://doi.org/10.1038/ncomms3280> PMID: 23955420.
72. Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S, Murai K. Class D and B(sister) MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). *Plant Molecular Biology*. 2009; 71(1–2):1–14. Epub 2009/06/03. <https://doi.org/10.1007/s11103-009-9504-z> PMID: 19488678.
73. Münster T, Wingen LU, Faigl W, Werth S, Saedler H, Theißen G. Characterization of three GLOBOSA-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene*. 2001; 262(1–2):1–13. [https://doi.org/10.1016/s0378-1119\(00\)00556-4](https://doi.org/10.1016/s0378-1119(00)00556-4) PMID: 11179662
74. Ramírez-González R, Borrill P, Lang D, Harrington S, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. *Science*. 2018; 361(6403):eaar6089. <https://doi.org/10.1126/science.aar6089> PMID: 30115782
75. Yang Y, Zhang X, Wu L, Zhang L, Liu G, Xia C, et al. Transcriptome profiling of developing leaf and shoot apices to reveal the molecular mechanism and co-expression genes responsible for the wheat heading date. *BMC Genomics*. 2021; 22(1):468. <https://doi.org/10.1186/s12864-021-07797-7> PMID: 34162321
76. Muterko A. Copy number variation and expression dynamics of the dominant vernalization-A1a allele in wheat. *Plant Molecular Biology Reporter*. 2023; <https://doi.org/10.1007/s11105-023-01406-5>
77. Shcherban AB, Salina EA. Evolution of VRN-1 homoeologous loci in allopolyploids of *Triticum* and their diploid precursors. *BMC Plant Biology*. 2017; 17(1):188. <https://doi.org/10.1186/s12870-017-1129-9> PMID: 29143603
78. Würschum T, Boeven PH, Langer SM, Longin CFH, Leiser WL. Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC genetics*. 2015; 16:1–8.
79. Golovnina KA, Kondratenko EY, Blinov AG, Goncharov NP. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. *BMC Plant Biology*. 2010; 10:1–15.
80. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology*. 2008; 148(2):993–1003. Epub 2008/08/22. <https://doi.org/10.1104/pp.108.122457> PMID: 18715958; PubMed Central PMCID: PMC2556807.
81. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology*. 2004; 4(1):10. <https://doi.org/10.1186/1471-2229-4-10> PMID: 15171794

82. Theißen G, Rümpler F, Gramzow L. Array of MADS-box genes: Facilitator for rapid adaptation? Trends in Plant Science. 2018; 23(7):563–76. <https://doi.org/10.1016/j.tplants.2018.04.008> PMID: 29802068
83. Michaels SD, Amasino RM. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. The Plant Cell. 2001; 13(4):935–41.
84. Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, et al. The *FLFMADS* box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. The Plant Cell. 1999; 11(3):445–58.
85. Whittaker C, Dean C. The *FLC* locus: a platform for discoveries in epigenetics and adaptation. Annual Review of Cell and Developmental Biology. 2017; 33:555–75. <https://doi.org/10.1146/annurev-cellbio-100616-060546> PMID: 28693387
86. Kennedy A, Geuten K. The role of *FLOWERING LOCUS C* relatives in cereals. Frontiers in Plant Science. 2020; 11:617340. <https://doi.org/10.3389/fpls.2020.617340> PMID: 33414801
87. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Sciences. 2003; 100(10):6263–8. <https://doi.org/10.1073/pnas.0937399100> PMID: 12730378
88. Nishimura K, Moriyama R, Katsura K, Saito H, Takisawa R, Kitajima A, et al. The early flowering trait of an emmer wheat accession (*Triticum turgidum* L. ssp. *dicoccum*) is associated with the cis-element of the *Vrn-A3* locus. Theoretical and Applied Genetics. 2018; 131(10):2037–53. <https://doi.org/10.1007/s00122-018-3131-5> PMID: 29961103
89. Wang H, Jiao X, Kong X, Hamera S, Wu Y, Chen X, et al. A signaling cascade from miR444 to RDR1 in rice antiviral RNA silencing pathway. Plant Physiology. 2016; 170(4):2365–77. <https://doi.org/10.1104/pp.15.01283> PMID: 26858364
90. Guo S, Xu Y, Liu H, Mao Z, Zhang C, Ma Y, et al. The interaction between *OsMADS57* and *OsTB1* modulates rice tillering via *DWARF14*. Nature Communications. 2013; 4(1):1566. <https://doi.org/10.1038/ncomms2542> PMID: 23463009
91. Yan Y, Wang H, Hamera S, Chen X, Fang R. miR444a has multiple functions in the rice nitrate-signaling pathway. The Plant Journal. 2014; 78(1):44–55. <https://doi.org/10.1111/tpj.12446> PMID: 24460537
92. Pachamuthu K, Hari Sundar V, Narjala A, Singh RR, Das S, Avik Pal HC, et al. Nitrate-dependent regulation of miR444-*OsMADS27* signalling cascade controls root development in rice. Journal of Experimental Botany. 2022; 73(11):3511–30. <https://doi.org/10.1093/jxb/erac083> PMID: 35243491
93. Lin SL, Miller JD, Ying SY. Intronic microRNA (miRNA). Journal of Biomedicine and Biotechnology. 2006; 2006(4):26818. Epub 2006/10/24. <https://doi.org/10.1155/JBB/2006/26818> PMID: 17057362; PubMed Central PMCID: PMC1559912.
94. Liu R, Ma Y, Guo T, Li G. Identification, biogenesis, function, and mechanism of action of circular RNAs in plants. Plant Communications. 2022. <https://doi.org/10.1016/j.xplc.2022.100430> PMID: 36081344
95. Feng K, Nie X, Cui L, Deng P, Wang M, Song W. Genome-wide identification and characterization of salinity stress-responsive miRNAs in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*). Genes. 2017; 8(6):156. <https://doi.org/10.3390/genes8060156> PMID: 28587281