MADS-box genes galore in wheat genome: phylogenomics, evolution and stress associated functions

Qasim Raza^{a*}, Awais Riaz^a, Rana Muhammad Atif^{b,c}, Babar Hussain^{d,e}, Zulfiqar Ali^f, Hikmet Budak^g

^a Molecular Breeding Laboratory, Rice Research Institute, Kala Shah Kaku, Sheikhpura, Punjab, Pakistan.

^b Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Pakistan.

^c Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture Faisalabad, Pakistan.

^d Department of Molecular Biology & Genetics, Middle East Technical University, Ankara, Turkey.

^e Faculty of Life Sciences, University of Central Punjab, Lahore, Pakistan.

^f Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan.

^g Montana BioAgriculture, Inc., Bozeman, MT, United States.

* Correspondence: qasimnazami@gmail.com

Abstract

MADS-box gene family members play multifarious roles in regulating the growth and development of crop plants and hold enormous promise for bolstering grain yield potential under changing global environments. Bread wheat (*Triticum aestivum* L.) is a key staple food crop around the globe. Until now, the available information concerning MADS-box genes in the wheat genome has been insufficient. However, a comprehensive genome-wide analysis identified 300 high confidence MADS-box genes from the latest publicly available reference genome of wheat. Comparative phylogenetic analyses with *Arabidopsis* and rice MADS-box genes classified the wheat genes into 16 distinct subfamilies, without a single *FLOWERING LOCUS C* homolog present in the wheat genome. Gene duplications were mainly identified in subfamilies containing unbalanced homeologs, pointing towards a potential mechanism for gene family expansion. Moreover, a more recent evolutionary origin was inferred for M-type genes, as compared with MIKC-type genes, indicating their significance in understanding the evolutionary history of the wheat genome. We speculate that subfamily-specific distal telomeric duplications in unbalanced homeologs facilitate the rapid adaptation of wheat to changing environments. Furthermore, our *in-silico* expression data strongly proposed MADS-box genes as active guardians of plants against pathogen insurgency and harsh environmental conditions. In conclusion, we provide an entire complement of MADS-box genes identified in the wheat genome that will accelerate functional genomics efforts and possibly facilitate bridging gaps between genotype-to-phenotype relationships through fine-tuning of agronomically important traits.

Key words: Genome-wide analysis, *In-silico* expression, *TaMADS*'s, Transcription factors, *Triticum aestivum*, Wheat adaptability.

1. Introduction

In the fight for global food security and safety, bread wheat represents one of the largest contributing grain crops. However, following the green revolution, improvement of wheat grain production has been hindered by various bottlenecks. This includes, but is not limited to, non-availability of a reliable and fully annotated reference genome (Brenchley et al., 2012; Chapman et al., 2015; Clavijo et al., 2017; IWGSC, 2014; Zimin et al., 2017). The slow progress in harnessing a fully annotated reference genome was mainly due to the genome's allohexaploid nature, which comprises three closely related, but independently maintained, sub-genomes, known as A, B, and D. This, in combination with a high frequency of repetitive sequences, particularly hindered progress. Recently, an alliance of geneticists combined resources over a period of 13 years, to produce a fully annotated sequence of the wheat genome, which had a resultant size of ~17 Gbps. This is the largest known genome in crop plants (IWGSC, 2018). To add to this, the high-quality of the reference genome, together with large scale RNA-seq data and expression repositories (Borrill et al., 2016; Ramírez-González et al., 2018), provide a rich resource for studying evolutionary dynamics and functional characterization of important gene families, which in turn could facilitate crop improvement efforts.

MADS-box transcription factors (TFs) are a well-documented group of genes known for playing vital roles in regulating the growth and development of several important plant species (Ali et al., 2019; Smaczniak et al., 2012). These genes influence diverse biological functions, including cell development, signal transduction, biotic and abiotic stress responses, vegetative organs development, control of flowering and anthesis time, formation of meristems and flower organs, ovule development, embryo development, dehiscence zone formation, and ripening of fruits and seeds (Ali et al., 2019; Alvarez-Buylla et al., 2000; Colombo et al., 1995; Kuo et al., 1997; Messenguy and Dubois, 2003; Moore et al., 2002; Pařenicová et al., 2003; Riechmann and Meyerowitz, 1997; Rounsley et al., 1995; Saedler et al., 2001; Samach et al., 2000; Smaczniak et al., 2012). Given their importance, identification and characterization of MADS-box genes in agriculturally important species is critical for crop improvement and fine-tuning of specific-traits through genetic exploitations.

All MADS-box genes contain a highly conserved MADS (M) domain of approximately 58–60 amino acids. The M domain enables DNA binding and is located in the N-terminal region of the protein (Yanofsky et al., 1990). Members of this gene family are classified into M/type I and MIKC/type II super clades, which were generated after an ancient gene duplication event which occurred before the separation of animal and plant lineages (Becker and Theißen, 2003). M-type genes have simple intron-exon structures (zero or one intron), only contain a single M domain, and encode Serum Response Factor (SRF)-like proteins. These can be subcategorized into four clades (*Mα*, *Mβ*, *Mγ* and *Mδ*). However, the *Mδ* clade genes closely resemble those within the *MIKC*^{*} group, as previously reported (De Bodt et al., 2003).

In comparison, MIKC-type genes have four domains: (i) a highly conserved DNA binding M domain, (ii) a less conserved Intervening (I) domain of ~30 AA involved in dimer formation, (iii) a moderately conserved Keratin (K) domain of ~70 AA which regulate

heterodimerization of MADS proteins, and lastly (iv) a highly inconstant C-terminal region which contributes to transcriptional regulation and higher-order protein complex formation (Díaz-Riquelme et al., 2009; Henschel et al., 2002). MIKC-type genes encode Myocyte Enhancer Factor 2 (MEF2)-like proteins and are categorized into *MIKC^c* and *MIKC^{*}* clades (Kaufmann et al., 2005). *MIKC^{*}*-type genes have duplicated K domains and relatively longer I domains than *MIKC^c*-types (Duan et al., 2014). *MIKC^c*-type genes can be further categorized into several subclades based on their phylogenetic relationships in flowering plants (Gramzow and Theißen, 2015).

Similar to what has been observed in other model and crop plant species, in wheat MADS-box genes are known to confer drought tolerance through regulation of drought tolerance genes and micro-RNAs (Budak et al., 2015). In addition to this, *TaMADS51*, *TaMADS4*, *TaMADS5*, *TaMADS6*, and *TaMADS18* all showed up-regulation, where *TaMADAGL17*, *TaMADAGL2*, *TaMADWM31C*, and *TaMADS14* were downregulated, under phosphorous (P) starvation, and further functional analysis confirmed their role in P-deficient stress responses (SHI et al., 2016). In another study, several MADS-box genes were activated and differentially expressed after inoculation of wheat spikes with Fusarium head blight (FHB) (Kugler et al., 2013). Furthermore, a simple yet elegant ABCDE floral organ identity model explained the complex genetic interactions among MIKC-type wheat MADS-box genes for determining the fate of floral organs (Ali et al., 2019). These significant roles of MADS-box genes in biotic and abiotic stresses, fertilizer response, and flower development highlight the necessity of their comprehensive identification and characterization in the bread wheat genome.

Previously, comprehensive genome-wide analyses of MADS-box genes have been carried out in diverse plant species. However, the available information about wheat MADS-box genes is comparatively sparse. For example, previously reported genome-wide identification of MADS-box genes in wheat (Ma et al., 2017; Schilling et al., 2020) has been revealed as outdated and inadequate, as it was performed by Ma et al. (2017) using only a draft version of the wheat genome (TGACv1) which contains a lesser number of genes. To add to this, Schilling et al. (2020) studied only MIKC-type MADS-box genes. Moreover, a significant number of low confidence and/or truncated protein-coding pseudogenes were also included in the before mentioned studies.

In order to bridge the gaps in knowledge, in this study we comprehensively identify the entire complement of high confidence, full-length, protein-coding MADS-box genes present in the latest publicly available reference genome of bread wheat (IWGSC RefSeq v1.1). Through a combination of different search approaches, 300 high confidence, non-redundant, full-length MADS-box genes were identified. Comparative phylogenetic analyses with model plant MADS-box genes further classified these into at least 16 *Arabidopsis* and/or grass specific subfamilies. Moreover, homeologous and duplicated genes were identified to study probable gene family expansion and evolution mechanisms. Furthermore, expression patterns of identified MADS-box genes were studied under several biotic and abiotic stress conditions. In this way, we provide a comprehensive resource of wheat MADS-box genes which have the

1 potential to facilitate molecular breeders in fine-tuning of important traits for further
2 improvement.

3 **2. Results**

4 *2.1. MADS-box genes galore in wheat genome*

5 In this study, conserved domains and query search-based approaches identified a total of 300
6 high confidence and non-redundant MADS-box genes from the latest publicly available
7 wheat genome (**Table S2**). The NCBI CDD batch search further revealed that 167 (~55.7%)
8 encode MADS-box proteins, 125 (~41.7%) encode MADS and K-box proteins, and only 8
9 (~2.6%) encode K-box domain-containing proteins (**Table S3**). The proteins lengths,
10 molecular weight, and isoelectric points of MADS domain-containing proteins ranged from
11 58–450 amino acids, 6.483–48.304 kD, and 4.420–12.123 pI, respectively (**Table S2**). This
12 data suggests that different MADS-box genes may function within different environments.

13 Comparative phylogenetic analyses between *Arabidopsis* and wheat MADS-box genes (**Fig.**
14 **S1**), as well as between rice and wheat (**Fig. S2**), distinguished wheat genes into M- and
15 MIKC-types (**Table S2**). In wheat, 128 (~43%) MADS-box genes exhibited high sequence
16 similarity with *Arabidopsis* and rice M-type (type-I) genes, whereas 172 (~57%) revealed
17 more homology with MIKC-type (type-II) genes (**Fig. 1**). In general, MADS-box genes were
18 equally distributed among the 21 wheat chromosomes, with an exception of the three
19 homeologous chromosomes (7A, 7B and 7D), which harboured a significantly higher number
20 of genes and displayed peak gene density values (**Fig. S3**). However, M-type genes were
21 randomly distributed among chromosomes and were predominantly located on three
22 homeologous chromosomes of 3rd, 6th, and 7th linkage groups (**Fig. S4**). In comparison,
23 MIKC-type genes were equally distributed on all chromosomes. Furthermore, the percent
24 contributions of A, B, and D genomes were also comparable (**Fig. S3**). As expected, MIKC-
25 type genes were the largest group of MADS-box genes in wheat. To date, the wheat genome
26 harbours the 2nd largest number of MADS-box genes identified, following *Brassica napus*
27 (Wu et al., 2018). Possible explanations for the abundance in wheat includes the larger
28 genome size, high gene number, high rate of homeolog retention, and hexaploid nature of the
29 genome.

30 *2.2. Subfamily diversity in MADS-box genes*

31 Separate ML phylogenies among *Arabidopsis*, rice, and wheat MADS-box genes exposed 14
32 MIKC-type and 3 M-type major subfamilies. The 172 MIKC-type genes were unevenly
33 dispersed into *API* (9), *AP3* (6), *PI* (6), *AG/STK* (12), *SEP* (28), *AGL6* (3), *AGL12* (6),
34 *AGL17* (31), *Bsister* (19), *MIKC** (27), *OsMADS32* (3), *SOC1* (13), and *SVP* (9) subfamilies
35 (**Fig. 1A, Table S2**). As expected, monocot (*OsMADS32*) and eudicot (*FLC*) specific gene
36 subfamilies were also witnessed. Absence of monocot orthologs from eudicot specific
37 subclade and/or eudicot orthologs from monocot specific subclade suggest that genes of these
38 subclades might have lost after the divergence of monocot and eudicot lineages. Likewise,
39 128 M-type genes were randomly distributed into *Ma* (53), *Mβ* (28), and *Mγ* (47) subfamilies
40 (**Fig. 1B**).

41 In general, *Arabidopsis*, rice and wheat MADS-box gene subfamilies roughly followed the
42 species-specific phylogenetic clades. Triads of wheat homeologous genes exhibited close

relationships with one or more rice genes, with *Arabidopsis* genes representing a sister group association to grass genes (e.g. the *AP1*, *AP3*, *PI*, *AG/STK*, *AGL6*, *AGL12*, *OsMADS32*, *SOC1*, and *SVP* subfamilies; **Fig. 1**). Whereas, the subfamily phylogenies were more complex in case of *AGL17*, *Bsister*, *MIKC**, *SEP*, and all M-type MADS-box genes, probably due to multiple duplication events during the polyploidization of wheat genome.

2.3. Wheat genome lacks *FLC*-like genes

FLC-like genes have been confirmed to regulate flowering time and vernalization responses in plants (Andrés and Coupland, 2012; Distelfeld et al., 2009) and their wheat and rice counterparts have previously been reported (Ruelens et al., 2013; Schilling et al., 2020). However, in this study, despite employing updated resources and tools (see Materials and Method section), none of the wheat and rice MADS-box genes fell into the *Arabidopsis* specific *FLC*-clade (**Fig. 1A**). All the wheat genes which were previously reported to be *FLC*-like (**Table S4**) were grouped with *MIKC**-like MADS-box genes. Furthermore, we demonstrated that amino acid sequences of *Arabidopsis* specific *FLC* genes were significantly different from *MIKC**-like genes (**Fig. 2**). The most significant differences were detected in MADS domain region at 30th, 34th, and 50th positions where glutamic acid (E), glutamine (Q), and alanine/glycine/serine (A/G/S) residues of *FLC* genes were substituted with lysine (K), glutamic acid (E), and proline (P), respectively. Collectively, these results strongly indicate that wheat and rice genomes lack *FLC* genes, which might have been lost after the divergence of eudicot and grass lineages.

2.4. Subfamily-specific gene duplications in MADS-box genes

Overall, MADS-box genes were equally located in the interstitial and proximal regions (R2a, R2b and C) and distal telomeric (R1 and R3) ends of chromosomes (49% and 51%, respectively) (**Table S2**). However, substantial differences were observed among gene locations of M-type and *MIKC*-type genes, as well as among subfamilies. The majority of the M-type genes were located in distal telomeric segments (62%), whereas *MIKC*-type genes were more prevalent in central chromosomal segments (57%). Generally, a larger portion of genes belonging to significantly expanded subfamilies tended to be located in distal telomeric ends, whereas genes of smaller subfamilies were more clustered in central chromosomal segments (**Table S2**).

Gene duplications were identified through sequence similarities in coding sequences of all MADS-box genes. A total of 201 duplicated gene pairs with $\geq 90\%$ sequence homology were identified, which corresponded to 123 non-redundant genes (**Fig. 3; Table S6**). Two genes (*TraesCSU02G209900*, *TraesCSU02G235300*) with unknown chromosomal location information were also recognized to be duplicated. However, *TraesCSU02G235300* showed duplications with genes located on chromosome 3B only, strongly suggesting that it was also located on the 3B chromosome. The *MIKC*-type MADS-box genes were also found to be more duplicated than M-type genes (60% vs 40% of duplicated gene pairs, respectively), particularly due to expanded subfamilies (e.g. *AGL17*, *Bsister*, *MIKC**, *SEP*, and *SOC1*). Remarkably, duplicated gene pairs were subfamily-specific and particularly recognized in subfamilies containing unbalanced homeologs, except for the *AP3* subfamily (**Tables 1, S5 & S6**). Among subfamilies, *Ma*, *AGL17*, *MIKC**, and *SEP* contained 25%, 21%, 14%, and 12% of the total duplicated gene pairs, respectively, and the majority of these were located in

1 distal telomeric and sub-telomeric (one gene located on the telomeric segment and other on
2 the central segment) chromosomal regions. We also observed that the majority of the
3 duplicated gene pairs (~51%) were located in distal telomeric segments, whereas only 26%
4 and 23% of the duplicated gene pairs were located in proximal and sub-telomeric segments of
5 chromosomes. These results could be explained by higher gene density in distal vs central
6 chromosomal segments (IWGSC, 2018). Furthermore, segmental duplications were more
7 prevalent than tandem duplications (61% and 37%) in 197 of the duplicated gene pairs with
8 available chromosomal information (**Table S6**). Interestingly, >37% of all tandem
9 duplications were identified on chromosome 3B, consistent with IWGSC findings (IWGSC,
10 2018). Collectively, these results strongly suggest that unbalanced homeologs in distal
11 telomeric regions derive MADS-box subfamilies expansion through segmental duplications.
12 These results are supported by previous observations that distal chromosomal regions are
13 frequent targets of recombination events and are hotspots for faster evolution of underlying
14 genes (N. W. G. Chen et al., 2018; Glover et al., 2015; Ramírez-González et al., 2018).

15 *2.5. Faster evolution of M-type MADS-box genes*

16 To investigate evolution rates, we estimated substitutions in coding sequences and computed
17 approximate divergence time between duplicated gene pairs (**Table S6**). In M-type genes, 1st
18 and 3rd quartiles of synonymous substitutions (Ks) were considerably narrower than MIKC-
19 type genes, whereas non-synonymous to synonymous substitution ratios (Ka/Ks) of M-type
20 genes were significantly higher than 1st and 3rd quartiles of MIKC-type genes (**Fig. 4**).
21 Furthermore, the percentage of duplication events as a result of positive/Darwinian selection
22 (Ka/Ks ratio > 1) was almost doubled (23%) in M-type genes as compared with MIKC-type
23 genes (12%) (**Table S6**). Additionally, the estimated divergence time of M-type genes was
24 significantly narrower than the MIKC-type genes. The mean divergence time of M-type
25 genes was nearly half of the mean of MIKC-type genes (**Fig. 4**). Taken together, these data
26 strongly indicate a more recent, rapid evolution of M-type MADS-box genes, which could be
27 explained by their predominant localization in distal telomeric regions of chromosomes (**Fig.**
28 **S3 & S4; Table S6**).

29 *2.6. Expression patterns of MADS-box genes under abiotic and biotic stresses*

30 Extensive investigations have been carried out to study the expression patterns of MADS-box
31 genes during growth and development of crop plants. However, their transcriptional
32 regulation under stressful conditions is somewhat obscure. Therefore, we analyzed RNA-seq
33 based expression data of 300 MADS-box genes under all available biotic and abiotic stress
34 conditions in the exVIP wheat expression browser (Borrill et al., 2016; Ramírez-González et
35 al., 2018) (**Fig. 5; Table S7**). Out of the total 300 genes, nearly 57% were expressed (log₂
36 TPM 0.20–7.89) during at least one developmental stage of one or more of the stresses
37 included in this study. Whereas, the remaining 43% of genes showed no or very low
38 expression (log₂ TPM < 0.0), and were subsequently considered as not expressed.

39 Overall, MIKC-type MADS-box genes were highly expressed under all studied stresses,
40 whereas nearly 75% of the non-expressed genes belonged to the three M-type subfamilies.
41 Interestingly, the remaining 25% of the non-expressed genes belonged to MIKC-subfamilies
42 containing duplicated genes, e.g. *AGL17*, *MIKC**, and *Bsister*. As the majority of the
43 duplicated genes were located in distal telomeric and sub-telomeric segments of

chromosomes, it could be expected that promoters of these genes had undergone H3K27me3 hypermethylation which resulted in their lower or none existent expression (Ramírez-González et al., 2018).

In general, genes of all MIKC-type subfamilies (except *AGL17*) were ubiquitously expressed in fusarium head blight infected spikelets/spikes. Nevertheless, floral homeotic genes (*API*, *AP3*, *PI*, *AG/STK*, *AGL6*, and *SEP*) showed highest expression patterns. *SVP*, *MIKC*^{*}, *AGL12*, and *SOC1*-like genes were also ubiquitously expressed in all studied abiotic and biotic stresses, strongly indicating that genes belonging to these subfamilies are active guardians of the wheat genome under stressful environments. Remarkably, all *API*-like genes demonstrated high expression in *Magnaporthe oryzae* infected leaves, suggesting disease-specific expression, as well as, the involvement of these important subfamily genes in diverse aspects of wheat growth and development. Similarly, *AGL17*-like genes were expressed only in seedlings under drought, heat, and phosphate starvation conditions.

We also calculated expression based hierarchical clusters to analyze diversification in expression patterns present within subfamilies (**Fig. S5, Table S8**). The *MIKC*^{*} and *SOC1*-like genes were grouped into seven and six random clusters, respectively. Similarly, *Bsister* and *SEP*-like genes were separately grouped into five random clusters. By contrast, *API* and *Mγ*-like genes exhibited no variation in their expression patterns, as all genes were grouped into a single random cluster. Likewise, *AGL6*, *AGL17*, *AP3*, *OsMADS32*, *Mα*, and *Mβ*-like genes displayed little diversification in expression, and genes of each subfamily were grouped into two random clusters. In comparison, the majority of the genes with no or very low expression (~58%) were grouped into cluster 10 and all belonged to subfamilies with duplicated genes (**Fig. S4, Table S8**).

3. Discussions

3.1. Recent significant advancements in identification and characterization of MADS-box family members in wheat

MADS-box gene family members have been identified across all groups of eukaryotes and are known to confer diverse biological functions. In plants, these play a central role during growth and development, and are thus very important targets for crop improvement (Ali et al., 2019). Recently, significant advancements have been made in identification and characterization of MADS-box family members in wheat, which could in turn facilitate crop breeding efforts and aid in development of more resilient and higher-yielding cultivars. Ma et al. (2017) reported the first genome-wide analysis of MADS-box family members and identified a total of 180 MADS-box genes in an earlier genome version (IWGSC, 2014). Their *in-silico* expression data provided insights into the stress associated functions of MADS-box genes. To add to this, Schilling et al. (2020) conducted a comprehensive analysis of MIKC-type MADS-box genes and identified a total of 201 genes from an updated genome version (Ref. Seq 1.0) (IWGSC, 2018). They speculated that pervasive duplications, functional conservation, and putative neofunctionalization may have contributed to the adaptation of wheat to diverse environments. These advancements, along with results presented in this study, have begun to elucidate the broad landscape of wheat MADS-box genes and provide new insights into the phylogenomic, evolution, and functions of MADS-box family members.

1 3.2. *MADS-box genes are underestimated in the wheat genome*

2 Bread wheat is a hexaploid species (AABBDD, $2n=6x=42$), which originated from a series of
3 naturally occurring hybridization events among three closely related and independently
4 maintained sub-genomes. Its full genome size is ~17 Gbps, among which ~14.5 Gbps (85%)
5 is sequenced and contains ~107,891 coding genes (IWGSC, 2018). In this study, through a
6 comprehensive genome-wide analysis, we identified a total of 300 high confidence, full
7 length, non-redundant MADS-box genes from the latest publicly available reference genome
8 version (**Fig. 1, Table S2**). To date, this is the second-highest number of MADS-box genes
9 identified in a crop plant, closely following *Brassica napus* (Wu et al., 2018), which contains
10 307 full-lengths and/or incomplete (pseudo) MADS-box genes. Possible explanations for the
11 abundance found in wheat may encompass the larger genome size, higher gene number, high
12 rate of homeolog retention, and hexaploid nature of bread wheat (IWGSC, 2018). Moreover,
13 at least 35 low confidence MIKC-type MADS-box genes were recently reported in the
14 IWGSC reference genome (Schilling et al., 2020). If these low confidence genes are also
15 included with the 300 MADS-box genes identified in this study, then the final number will be
16 the highest identified in a plant genome. Furthermore, a more recent chromosome-scale
17 assembly of the bread wheat genome revealed hundreds of new genes and thousands of
18 additional gene copies (Alonge et al., 2020) which were substantially missing from the
19 IWGSC reference genome explored in the current study. Collectively, all these observations
20 strongly suggest that MADS-box genes are underestimated in wheat and their exploration
21 might help in understanding the global adaptability of wheat.

22 3.3. *Does grass genomes lack FLC-like genes?*

23 *FLOWERING LOCUS Cs* or *FLCs* are key genes conferring vernalization requirement, which
24 act as flowering repressors in *Arabidopsis* (Whittaker and Dean, 2017). High expression of
25 *FLCs* repress other flowering activator genes and causes delayed flowering. In winter
26 varieties of temperate cereals, including barley, *Brachypodium*, and wheat, vernalization is
27 regulated through *VERNALIZATION* genes (*VRN1*, *VRN2*, *VRN3*) (Greenup et al., 2009). To
28 date, strong disagreement exists in the published literature regarding the presence or absence
29 of *FLC*-like genes in cereals (Paolacci et al., 2007; Ruelens et al., 2013; Schilling et al., 2020;
30 Wei et al., 2014; Zhao et al., 2011). In this study, despite using updated resources and tools,
31 we were unable to find *FLC*-like genes within the wheat or rice genomes (**Fig. 1A**).
32 Interestingly, wheat genes which were previously reported to be *FLC*-like (Ruelens et al.,
33 2013; Schilling et al., 2020), as well as *OsMADS37* and *OsMADS51/65*, were grouped with
34 *MIKC**-like genes (**Fig. 1A, Table S4**). In comparison, all other *FLC* homologs were grouped
35 into a separate distinct *Arabidopsis*-specific *FLC* clade. Moreover, we also demonstrated that
36 multiple sequence amino acid alignments of *FLC* and *MIKC**-like genes were significantly
37 dissimilar and none of the grass genes shared sequence homology with *Arabidopsis FLC*
38 paralogs (**Fig. 2**). Altogether, these results indicate that *FLC*-like genes might have been lost
39 in cereal genomes after they diverged from eudicots. Several other studies are also in
40 agreement with current results and were unable to identify *FLC*-like genes in grass genomes
41 (Arora et al., 2007; Paolacci et al., 2007; Wei et al., 2014; Zhao et al., 2011). These
42 observations suggest that vernalization pathways in monocots and dicots evolved
43 independently and are regulated through a completely different set of genes. However, earlier
44 reports on the identification of *FLC*-like genes in grasses (Ruelens et al., 2013; Schilling et

al., 2020; Sharma et al., 2017) and functional conservation of *ODDSOC2* with *FLC* in the regulation of vernalization (Greenup et al., 2010), suggest that evolution of vernalization pathways between monocots and dicots may not be fully independent. Thus, the biological question of the presence or absence of *FLC*-like genes in grass genomes still needs to be explored. Further research efforts in this direction might address this fundamental question, and propitious outcomes could help in breeding efforts to develop winter varieties of temperate cereals which are adapted to changing environments.

3.4. Subfamily-specific distal telomeric duplications in unbalanced homeologs facilitate rapid adaptation to changing environments

Distal telomeric chromosomal segments are evolutionary hotspots for frequent recombination events and give rise to fast-evolving genes (N. W. G. Chen et al., 2018; Glover et al., 2015). Many adaptability trait related genes which are induced in response to external stimuli are found to be predominately located in distal chromosomal segments. In comparison, genes related to housekeeping and conserved developmental functions are positioned in central chromosomal segments (IWGSC, 2018; Ramírez-González et al., 2018). In this study, the majority of the identified duplicated gene pairs were located in distal telomeric (~51%) and sub-telomeric (23%) chromosomal segments (**Fig. 3, Table S6**). Remarkably, all these duplications were subfamily-specific and primarily identified in larger subfamilies containing unbalanced homeologs (*Ma*, *Mβ*, *Mγ*, *AGL17*, *Bsister*, *MIKC**, *SEP*, and *SOC1*) (**Table 1**). Interestingly, genes belonging to these subfamilies have been reported to regulate plants adaptability in changing environments. For example, *OsMADS57* (*AGL17*-like gene in rice) is induced by abscisic acid, chilling, drought, and salinity stress (Arora et al., 2007), promotes cold tolerance by interacting with a defence gene (*OsWRKY94*) (L. Chen et al., 2018), and modulates root to shoot nitrate translocation under deprived nitrate conditions (Huang et al., 2019). Similarly, downregulation of a *SEP* clade gene in pepper (*CaMADS*) caused more sensitivity to cold, salt, and osmotic stresses, whereas its overexpression in *Arabidopsis* conferred higher tolerance against these stresses (Chen et al., 2019). By contrast, almost negligible duplications were identified in a single smaller subfamily (*AP3*) with completely balanced homeologs (**Table 1**). These results strongly indicate that unbalanced homeologs of expended subfamilies have undergone subfamily-specific distal telomeric duplications, thus facilitating the rapid adaptation of bread wheat to diverse global environments. Contrarily, completely balanced homeologs of functionally conserved smaller subfamilies might have some evolutionary advantage in minimizing the developmentally detrimental gene copy number variations, due to their localization in central chromosomal segments.

3.5. A recent and faster evolutionary origin of M-type MADS-box genes

Unlike *MIKC*-type MADS-box genes, the M-type genes have not been extensively studied and little information is available about their evolutionary origin in crop plants (Masiero et al., 2011). In this study, we observed that M-type genes were predominately located in distal telomeric chromosomal segments (**Table S2**) and are evolving at a faster rate, probably due to higher frequency of tandem gene duplications and stronger purifying selections (**Fig. 3, Table S6**). Moreover, lesser synonymous substitutions and higher non-synonymous to synonymous substitution ratios of M-type duplicated genes exposed their recent divergence and evolutionary origin as compared with *MIKC*-type genes (**Fig. 4, Table S6**). These results are in agreement with a study by Nam et al. (2004), in which they also reported faster birth-

and-death evolution of M-type genes in angiosperms. This data might indicate the significance of M-type MADS-box genes in understanding the evolutionary history of the bread wheat genome.

3.6. *MADS-box genes as active guardians of plants against pathogen insurgency and harsh environments*

MADS-box genes regulate diverse developmental processes and their functions are well studied in plant morpho- and organogenesis (Ali et al., 2019; Smaczniak et al., 2012). However, several members of the MADS-box gene family are reported to be involved in the regulation of biotic and abiotic stress responses (Castelán-Muñoz et al., 2019; Q. Wang et al., 2018; Zhang et al., 2016), which point towards possible dynamic roles in stress response. In this study, we analyzed the complete atlas of their expression profiles in all publicly available wheat transcriptomic data under biotic and abiotic stresses (**Fig. 5 & S5; Tables S7 & S8**). More than 50% of the total identified MADS-box genes were expressed during at least one developmental stage of one or more stresses, and the majority of these expressed genes belonged to the MIKC-group (**Table S7**). Except for *AGL17*, genes of all other MIKC-type subfamilies were ubiquitously expressed in FHB infected spikes/spikelets (**Fig. 5 & S5**), mimicking their defensive roles against *Fusarium* infection. Previously, Yang et al. (2015) functionally characterized *FgMcm1* in the causal agent (*Fusarium graminearum*) of barley and wheat head blight disease and demonstrated that *FgMcm1* played crucial roles in cell identity and fungal development. More recently, Xu et al. (2020) also reported a close relationship between anther extrusion and field FHB resistance and pointed out that *Rht* and *Vrn* genes might have pleiotropic effects on these traits. Since FHB is a floral disease and floral organ identity is controlled by MIKC-type MADS-box genes (Ali et al., 2019), this indicates a cooperate relationship for their involvement in the pathogen response network. In future, it would be interesting to investigate how MADS-box genes are involved in fighting FHB infection.

Similarly, *SVP*, *MIKC**, *AGL12*, and *SOCI* subfamily genes were ubiquitously expressed under all studied stresses (**Fig. 5, Table S7**), pointing towards their more critical role in stress associated functions. In comparison, members of the largest MIKC group (*AGL17*) were expressed under phosphate starvation, and heat and drought stress conditions. Several functionally characterized genes belonging to all these subfamilies have also been reported to regulate different abiotic and/or biotic stresses in plants (Hwang et al., 2019; Khong et al., 2015; P. Li et al., 2020; SHI et al., 2016; Z. Wang et al., 2018). Interestingly, few M-type genes, especially *Mα*-like, were weakly expressed under different biotic stresses (**Fig. 5**) and these expression patterns are in agreement with a Guo et al. (2013) study in which they observed differential expression of an M-type gene in response to stripe rust infection in wheat. Collectively, these results strongly highlight MADS-box genes, especially MIKC-type, as key members of gene regulatory networks (Castelán-Muñoz et al., 2019) and their involvement in wheat response to possible pathogen insurgency and harsh environments.

3.7. *Towards bridging the gap between genotype and phenotype using genetic and genomic resources*

Understanding the functions of candidate genes controlling agronomically important traits is critical for the acceleration of crop improvement efforts. Following the unveiling of recent

1 wheat genetic and genomic resources (Adamski et al., 2020; Avni et al., 2017; Luo et al.,
2 2017; Maccaferri et al., 2019), rapid testing of previous model plant discoveries and their
3 cross-application in staple crops is paving the way towards bridging the gap between
4 genotype and phenotype. During recent years, several wheat genes related to flowering, yield,
5 quality, disease resistance, male sterility, and nutrient use efficiency traits have been cloned
6 and functionally characterized (briefly reviewed in Li et al. (2020)), which exposed the
7 unprecedented potential of these genetic and functional genomic resources in establishing
8 genotype-to-phenotype relationships. These resources could also facilitate in elucidating the
9 regulatory roles of neglected MADS-box subfamilies, such as *Ma*, *Mβ*, *My*, and *MIKC**,
10 during wheat growth and development, and subsequently help in deciphering whether
11 MADS-box genes confer pathogen resistance and/or abiotic stress tolerance. Furthermore,
12 many other fundamental biological questions about their phylogenomics, evolutionary origin,
13 and stress associated functions, could be addressed by exploiting these paramount resources.

14 In conclusion, MADS-box genes are critically important for wheat growth and development,
15 and hold enormous promise for bolstering yield potential under changing global
16 environments. Our results speculate that the abundance of MADS-box genes in the wheat
17 genome might be associated with its global adaptability and that subfamily-specific distal
18 telomeric duplications in unbalanced homeologs facilitate its rapid adaptation. In addition,
19 through comprehensive genome-wide and comparative analyses, we demonstrated that wheat
20 and rice genomes lack *FLC* genes, which could help molecular breeders in the identification
21 of target genes for fine-tuning of winter wheat varieties. Moreover, our *in-silico* expression
22 data strongly indicated the possibility of protective roles of MADS-box genes against
23 pathogen attacks and harsh climatic conditions. In this way, we provided an entire
24 complement of MADS-box genes identified in the wheat genome that will accelerate
25 functional genomics efforts and possibly facilitate in bridging genotype-to-phenotype
26 relationships through fine-tuning of agronomically important traits.

27 **4. Material and methods**

28 *4.1. Identification of MADS-box genes*

29 The latest publicly available genome version of wheat (IWGSC Ref. Seq v1.1) was accessed
30 through the ensemble plants database (Bolser et al., 2017) and searched using protein family
31 database (Pfam) identifiers of MADS (PF00319) and K (PF01486) domains. A total of 281
32 MADS-domain and 131 K-domain encoding genes were identified. Additionally, a query-
33 based search using “MADS” yielded 300 genes. Together, these searches identified a total of
34 712 genes (**Table S1**), among which 300 were found to be high confidence and non-
35 redundant and considered for further analyses. Detailed information of these 300 genes was
36 retrieved from the ensemble plants database and is provided in **Table S2**. Moreover, amino
37 acid sequences for the identified genes were uploaded into the NCBI conserved domain
38 database (Marchler-Bauer et al., 2015) for validation of putative protein domains (**Table S3**).
39 Gene identifiers were used as gene names for subsequent analyses; however, alternative
40 names used in previous studies (Schilling et al., 2020) were also provided in **Table S4**.

41 *4.2. Comparative phylogenetic analyses and subfamily classification*

42 Differentiation between type I (M-type) and type II (MIKC-type) MADS-box proteins were
43 achieved by separate MAFFT alignments using only the MADS domain (L-INS-i algorithm)

(Kato et al., 2018; Kato and Standley, 2013) between *Arabidopsis* (Pařenicov et al., 2003) and wheat, and rice (Arora et al., 2007) and wheat MADS-box proteins. Subsequently, maximum likelihood (ML) phylogenies were inferred using IQ-TREE (Nguyen et al., 2015) by choosing JTT+G4 best fit substitution model according to the Bayesian information criterion (BIC) (Kalyanamoorthy et al., 2017). Consistency of the ML trees was validated by setting an Ultrafast bootstrap value of 1000 (Hoang et al., 2018; Minh et al., 2013). The final phylogenetic trees were visualized with MEGA7 (Kumar et al., 2016).

Subfamily classifications were accomplished by separate MAFFT alignments within M-type (L-INS-i algorithm) and MIKC-type (E-INS-i algorithm) MADS-box protein sequences of *Arabidopsis*, rice, and wheat (Kato et al., 2018; Kato and Standley, 2013). The full-length alignments were subjected to the Gap Strip/Squeeze v2.1.0 tool (www.hiv.lanl.gov/content/sequence/GAPSTRIP/gap.html) for masking the individual residues by removing the gaps with default parameters. Then, masked alignments of M-type and MIKC-type proteins were independently subjected to the IQ-TREE software (Nguyen et al., 2015) for generating ML phylogenetic trees as described above. Subfamily names were given by following subfamily classifications in *Arabidopsis* and/or major grass species (Gramzow and Theißen, 2015) (**Table S2 & S4**).

4.3. Homeologs identification

Putative homeologs were recognized based on strong phylogenetic relationships (Ultrafast bootstrap value > 90) within different sub-families. Classifications reported in previous studies were also considered (IWGSC, 2018). The homeologs status of thirty-one genes could not be determined due to lower Ultrafast bootstrap values (**Table S5**).

4.4. Gene duplication and evolution analyses

Coding sequences (CDS) of all wheat MADS-box genes were retrieved from the ensemble plants database and blasted against each other using Sequence Demarcation Tool V1.2 (Muhire et al., 2014) for the identification of sequence identities. Gene pairs with $\geq 90\%$ identity (E value $< 1e^{-10}$) and non-homeologous status were considered as duplicated (Ning et al., 2017) (**Table S6**). If the duplicated homologous gene pair was located on the same chromosome it was defined as tandem duplication. Otherwise, when homologous gene pairs were located on different chromosomes it was defined as segmental duplication. The CDS sequences of duplicated genes were MAFFT aligned, masked, and subjected to the Synonymous Non-Synonymous Analysis Program V2.1.1 (www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) to compute the synonymous (Ks) and non-synonymous (Ka) substitution rates. To find out which type of codon selection operated during evolution, the ratio of Ka/Ks was also calculated. The approximate divergence time between duplicated gene pairs was calculated by using the formulae $T = Ks/2r \times 10^{-6}$ assuming a substitution rate (r) of 6.5×10^{-9} substitutions/synonymous site/year (El Baidouri et al., 2017) (**Table S6**).

4.5. In-silico expression analyses

Expression data under all available abiotic and biotic stress conditions were retrieved from the expVIP Wheat Expression Browser (Borrill et al., 2016; Ramrez-Gonzlez et al., 2018) as Log₂ TPM (processed expression value in transcripts per million) obtained via RNA-seq

1 analysis. Detailed information about expression levels of individual genes in tested
2 tissues/growth stages and stresses/diseases are provided in the supplementary information
3 (**Table S7**). Genesis software (Sturn et al., 2002) was used to generate heatmaps from the
4 obtained expression data. Expression based clustering of genes was achieved by following
5 the *K*-means clustering method ($K = 10$, iterations = 1000, runs = 5) (**Table S8**).

6 **Acknowledgments**

7 All authors are grateful to Dr Yuling Jiao, Institute of Genetics and Developmental Biology,
8 Chinese Academy of Sciences for his comments and suggestions on the manuscript. The
9 inspiration and support from Precision Agriculture and Analytics Lab (PAAL), affiliate of
10 National Center in Big Data and Cloud Computing (NCBC) is also gratefully acknowledged.

11 **Competing interests**

12 The authors declare that no competing interests exist.

13 **Author contributions**

14 **Qasim Raza:** Conceptualization, Data curation, Formal Analysis, Methodology, Software,
15 Validation, Visualization, Writing-original draft. **Awais Riaz:** Data curation, Formal
16 Analysis, Methodology, Software, Validation, Visualization, Writing-review & editing. **Rana**
17 **Muhammad Atif:** Data curation, Supervision, Validation, Writing-review & editing. **Babar**
18 **Hussain:** Data curation, Visualization, Writing-original draft. **Zulfiqar Ali:**
19 Conceptualization, Supervision, Writing-review & editing. **Hikmet Budak:** Supervision,
20 Writing-review & editing.

21

References:

- Adamski, N.M., Borrill, P., Brinton, J., Harrington, S.A., Marchal, C., Bentley, A.R., Bovill, W.D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., Ford, B., Ghosh, S., Harwood, W., Hassani-Pak, K., Hayta, S., Hickey, L.T., Kanyuka, K., King, J., Maccaferri, M., Naamati, G., Pozniak, C.J., Ramirez-Gonzalez, R.H., Sansaloni, C., Trevaskis, B., Wingen, L.U., Wulff, B.B.H., Uauy, C., 2020. A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *Elife*. <https://doi.org/10.7554/eLife.55646>
- Ali, Z., Raza, Q., Atif, R.M., Aslam, U., Ajmal, M., Chung, G., 2019. Genetic and molecular control of floral organ identity in cereals. *Int. J. Mol. Sci.* 20. <https://doi.org/10.3390/ijms20112743>
- Alonge, M., Shumate, A., Puiu, D., Zimin, A., Salzberg, S.L., 2020. Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. *Genetics*. <https://doi.org/10.1534/genetics.120.303501>
- Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F., Yanofsky, M.F., 2000. MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* <https://doi.org/10.1046/j.1365-313X.2000.00891.x>
- Andrés, F., Coupland, G., 2012. The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg3291>
- Arora, R., Agarwal, P., Ray, S., Singh, A.K., Singh, V.P., Tyagi, A.K., Kapoor, S., 2007. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-8-242>
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K.W., Golan, G., Deek, J., Ben-Zvi, B., Ben-Zvi, G., Himmelbach, A., Maclachlan, R.P., Sharpe, A.G., Fritz, A., Ben-David, R., Budak, H., Fahima, T., Korol, A., Faris, J.D., Hernandez, A., Mikel, M.A., Levy, A.A., Steffenson, B., Maccaferri, M., Tuberosa, R., Cattivelli, L., Faccioli, P., Ceriotti, A., Kashkush, K., Pourkheirandish, M., Komatsuda, T., Eilam, T., Sela, H., Sharon, A., Ohad, N., Chamovitz, D.A., Mayer, K.F.X., Stein, N., Ronen, G., Peleg, Z., Pozniak, C.J., Akhunov, E.D., Distelfeld, A., 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* (80-.). <https://doi.org/10.1126/science.aan0032>
- Becker, A., Theißen, G., 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* [https://doi.org/10.1016/S1055-7903\(03\)00207-0](https://doi.org/10.1016/S1055-7903(03)00207-0)
- Bolser, D.M., Staines, D.M., Perry, E., Kersey, P.J., 2017. Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomic data, in: *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-6658-5_1
- Borrill, P., Ramirez-Gonzalez, R., Uauy, C., 2016. expVIP: A customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186. <https://doi.org/10.1104/pp.15.01667>

- 1 Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L.A., D'Amore, R., Allen, A.M.,
2 McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M.,
3 Bancroft, I., Gu, Y., Huo, N., Luo, M.C., Sehgal, S., Gill, B., Kianian, S., Anderson, O.,
4 Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F.X., Edwards, K.J.,
5 Bevan, M.W., Hall, N., 2012. Analysis of the bread wheat genome using whole-genome
6 shotgun sequencing. *Nature*. <https://doi.org/10.1038/nature11650>
- 7 Budak, H., Hussain, B., Khan, Z., Ozturk, N.Z., Ullah, N., 2015. From genetics to functional
8 genomics: Improvement in drought signaling and tolerance in wheat. *Front. Plant Sci.*
9 <https://doi.org/10.3389/fpls.2015.01012>
- 10 Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-
11 Ponce, B., Sánchez, M. de la P., Álvarez-Buylla, E.R., Garay-Arroyo, A., 2019. MADS-
12 box genes are key components of genetic regulatory networks involved in abiotic stress
13 and plastic developmental responses in plants. *Front. Plant Sci.*
14 <https://doi.org/10.3389/fpls.2019.00853>
- 15 Chapman, J.A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., Strnadova, V.,
16 Jenkins, J., Sehgal, S., Olike, L., Schmutz, J., Yelick, K.A., Scholz, U., Waugh, R.,
17 Poland, J.A., Muehlbauer, G.J., Stein, N., Rokhsar, D.S., 2015. A whole-genome
18 shotgun approach for assembling and anchoring the hexaploid bread wheat genome.
19 *Genome Biol.* <https://doi.org/10.1186/s13059-015-0582-8>
- 20 Chen, L., Zhao, Y., Xu, S., Zhang, Z., Xu, Y., Zhang, J., Chong, K., 2018. OsMADS57
21 together with OsTB1 coordinates transcription of its target OsWRKY94 and D14 to
22 switch its organogenesis to defense for cold adaptation in rice. *New Phytol.*
23 <https://doi.org/10.1111/nph.14977>
- 24 Chen, N.W.G., Thareau, V., Ribeiro, T., Magdelenat, G., Ashfield, T., Innes, R.W., Pedrosa-
25 Harand, A., Geffroy, V., 2018. Common bean subtelomeres are hot spots of
26 recombination and favor resistance gene evolution. *Front. Plant Sci.*
27 <https://doi.org/10.3389/fpls.2018.01185>
- 28 Chen, R., Ma, J., Luo, D., Hou, X., Ma, F., Zhang, Y., Meng, Y., Zhang, H., Guo, W., 2019.
29 CaMADS, a MADS-box transcription factor from pepper, plays an important role in the
30 response to cold, salt, and osmotic stress. *Plant Sci.*
31 <https://doi.org/10.1016/j.plantsci.2018.11.020>
- 32 Clavijo, B.J., Venturini, L., Schudoma, C., Accinelli, G.G., Kaithakottil, G., Wright, J.,
33 Borrill, P., Kettleborough, G., Heavens, D., Chapman, H., Lipscombe, J., Barker, T., Lu,
34 F.H., McKenzie, N., Raats, D., Ramirez-Gonzalez, R.H., Counce, A., Peel, N., Percival-
35 Alwyn, L., Duncan, O., Trösch, J., Yu, G., Bolser, D.M., Namaati, G., Kerhornou, A.,
36 Spannagl, M., Gundlach, H., Haberer, G., Davey, R.P., Fosker, C., Di Palma, F.,
37 Phillips, A.L., Millar, A.H., Kersey, P.J., Uauy, C., Krasileva, K. V., Swarbreck, D.,
38 Bevan, M.W., Clark, M.D., 2017. An improved assembly and annotation of the
39 allohexaploid wheat genome identifies complete families of agronomic genes and
40 provides genomic evidence for chromosomal translocations. *Genome Res.*
41 <https://doi.org/10.1101/gr.217117.116>
- 42 Colombo, L., Franken, J., Koetje, E., Van Went, J., Dons, H.J., Angenent, G.C., Van Tunen,
43 A.J., 1995. The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell.*
44 <https://doi.org/10.1105/tpc.7.11.1859>

- 1 De Bodt, S., Raes, J., Van De Peer, Y., Theißen, G., 2003. And then there were many: MADS
2 goes genomic. *Trends Plant Sci.* <https://doi.org/10.1016/j.tplants.2003.09.006>
- 3 Díaz-Riquelme, J., Lijavetzky, D., Martínez-Zapater, J.M., Carmona, M.J., 2009. Genome-
4 wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol.*
5 <https://doi.org/10.1104/pp.108.131052>
- 6 Distelfeld, A., Li, C., Dubcovsky, J., 2009. Regulation of flowering in temperate cereals.
7 *Curr. Opin. Plant Biol.* <https://doi.org/10.1016/j.pbi.2008.12.010>
- 8 Duan, W., Song, X., Liu, T., Huang, Z., Ren, J., Hou, X., Li, Y., 2014. Genome-wide analysis
9 of the MADS-box gene family in *Brassica rapa* (Chinese cabbage). *Mol. Genet.*
10 *Genomics.* <https://doi.org/10.1007/s00438-014-0912-7>
- 11 El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., Alaux, M.,
12 Quesneville, H., Pont, C., Salse, J., 2017. Reconciling the evolutionary origin of bread
13 wheat (*Triticum aestivum*). *New Phytol.* <https://doi.org/10.1111/nph.14113>
- 14 Glover, N.M., Daron, J., Pingault, L., Vandepoele, K., Paux, E., Feuillet, C., Choulet, F.,
15 2015. Small-scale gene duplications played a major role in the recent evolution of wheat
16 chromosome 3B. *Genome Biol.* <https://doi.org/10.1186/s13059-015-0754-6>
- 17 Gramzow, L., Theißen, G., 2015. Phylogenomics reveals surprising sets of essential and
18 dispensable clades of MIKCC-group MADS-box genes in flowering plants. *J. Exp. Zool.*
19 *Part B Mol. Dev. Evol.* <https://doi.org/10.1002/jez.b.22598>
- 20 Greenup, A., Peacock, W.J., Dennis, E.S., Trevaskis, B., 2009. The molecular biology of
21 seasonal flowering-responses in *Arabidopsis* and the cereals. *Ann. Bot.*
22 <https://doi.org/10.1093/aob/mcp063>
- 23 Greenup, A.G., Sasani, S., Oliver, S.N., Talbot, M.J., Dennis, E.S., Hemming, M.N.,
24 Trevaskis, B., 2010. ODDSOC2 is a MADS box floral repressor that is down-regulated
25 by vernalization in temperate cereals. *Plant Physiol.*
26 <https://doi.org/10.1104/pp.109.152488>
- 27 Guo, J., Shi, X.X., Zhang, J.S., Duan, Y.H., Bai, P.F., Guan, X.N., Kang, Z.S., 2013. A type I
28 MADS-box gene is differentially expressed in wheat in response to infection by the
29 stripe rust fungus. *Biol. Plant.* <https://doi.org/10.1007/s10535-012-0297-6>
- 30 Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., Theißen, G., 2002. Two
31 ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella*
32 *patens*. *Mol. Biol. Evol.* <https://doi.org/10.1093/oxfordjournals.molbev.a004137>
- 33 Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2:
34 Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*
35 <https://doi.org/10.1093/molbev/msx281>
- 36 Huang, S., Liang, Z., Chen, S., Sun, H., Fan, X., Wang, C., Xu, G., Zhang, Y., 2019. A
37 transcription factor, OsMADS57, regulates long-distance nitrate transport and root
38 elongation. *Plant Physiol.* <https://doi.org/10.1104/pp.19.00142>
- 39 Hwang, K., Susila, H., Nasim, Z., Jung, J.Y., Ahn, J.H., 2019. *Arabidopsis* ABF3 and ABF4
40 Transcription Factors Act with the NF-YC Complex to Regulate SOC1 Expression and
41 Mediate Drought-Accelerated Flowering. *Mol. Plant.*
42 <https://doi.org/10.1016/j.molp.2019.01.002>

- 1 IWGSC, 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum*
2 *aestivum*) genome. *Science* (80-.). <https://doi.org/10.1126/science.1251788>
- 3 IWGSC, 2018. Shifting the limits in wheat research and breeding using a fully annotated
4 reference genome. *Science* (80-.). 361. <https://doi.org/10.1126/science.aar7191>
- 5 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermini, L.S., 2017.
6 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*.
7 <https://doi.org/10.1038/nmeth.4285>
- 8 Katoh, K., Rozewicki, J., Yamada, K.D., 2018. MAFFT online service: Multiple sequence
9 alignment, interactive sequence choice and visualization. *Brief. Bioinform.*
10 <https://doi.org/10.1093/bib/bbx108>
- 11 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
12 Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
13 <https://doi.org/10.1093/molbev/mst010>
- 14 Kaufmann, K., Melzer, R., Theissen, G., 2005. MICK-type MADS-domain proteins:
15 Structural modularity, protein interactions and network evolution in land plants. *Gene*.
16 <https://doi.org/10.1016/j.gene.2004.12.014>
- 17 Khong, G.N., Pati, P.K., Richaud, F., Parizot, B., Bidzinski, P., Mai, C.D., Bès, M., Bourrié,
18 I., Meynard, D., Beeckman, T., Selvaraj, M.G., Manabu, I., Genga, A.M., Brigidou, C.,
19 Do, V.N., Guiderdoni, E., Morel, J.B., Gantet, P., 2015. OsMADS26 negatively
20 regulates resistance to pathogens and drought tolerance in rice. *Plant Physiol.*
21 <https://doi.org/10.1104/pp.15.01192>
- 22 Kugler, K.G., Siegwart, G., Nussbaumer, T., Ametz, C., Spannagl, M., Steiner, B., Lemmens,
23 M., Mayer, K.F.X., Buerstmayr, H., Schweiger, W., 2013. Quantitative trait loci-
24 dependent analysis of a gene co-expression network associated with *Fusarium* head
25 blight resistance in bread wheat (*Triticum aestivum* L.). *BMC Genomics*.
26 <https://doi.org/10.1186/1471-2164-14-728>
- 27 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics
28 Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*
29 <https://doi.org/10.1093/molbev/msw054>
- 30 Kuo, M.H., Nadeau, E.T., Grayhack, E.J., 1997. Multiple phosphorylated forms of the
31 *Saccharomyces cerevisiae* Mcm1 protein include an isoform induced in response to high
32 salt concentrations. *Mol. Cell. Biol.* <https://doi.org/10.1128/mcb.17.2.819>
- 33 Li, J., Yang, J., Li, Y., Ma, L., 2020. Current strategies and advances in wheat biology. *Crop*
34 *J.* <https://doi.org/10.1016/j.cj.2020.03.004>
- 35 Li, P., Zhang, Q., He, D., Zhou, Y., Ni, H., Tian, D., Chang, G., Jing, Y., Lin, R., Huang, J.,
36 Hu, X., 2020. AGAMOUS-LIKE67 Cooperates with the Histone Mark Reader EBS to
37 Modulate Seed Germination under High Temperature. *Plant Physiol.*
38 <https://doi.org/10.1104/pp.20.00056>
- 39 Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T.,
40 Wang, L., Wang, Y., McGuire, P.E., Liu, S., Long, H., Ramasamy, R.K., Rodriguez,
41 J.C., Van Sonny, L., Yuan, L., Wang, Z., Xia, Z., Xiao, L., Anderson, O.D., Ouyang, S.,
42 Liang, Y., Zimin, A. V., Pertea, G., Qi, P., Bennetzen, J.L., Dai, X., Dawson, M.W.,
43 Müller, H.G., Kugler, K., Rivarola-Duarte, L., Spannagl, M., Mayer, K.F.X., Lu, F.H.,

- 1 Bevan, M.W., Leroy, P., Li, P., You, F.M., Sun, Q., Liu, Z., Lyons, E., Wicker, T.,
2 Salzberg, S.L., Devos, K.M., Dvořák, J., 2017. Genome sequence of the progenitor of the
3 wheat D genome *Aegilops tauschii*. *Nature*. <https://doi.org/10.1038/nature24486>
- 4 Ma, J., Yang, Y., Luo, W., Yang, C., Ding, P., Liu, Y., Qiao, L., Chang, Z., Geng, H., Wang,
5 P., Jiang, Q., Wang, J., Chen, G., Wei, Y., Zheng, Y., Lan, X., 2017. Genome-wide
6 identification and analysis of the MADS-box gene family in bread wheat (*Triticum*
7 *aestivum* L.). *PLoS One*. <https://doi.org/10.1371/journal.pone.0181443>
- 8 Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M.,
9 Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G., Himmelbach, A., Mascher, M.,
10 Bagnaresi, P., Faccioli, P., Cozzi, P., Lauria, M., Lazzari, B., Stella, A., Manconi, A.,
11 Gnocchi, M., Moscatelli, M., Avni, R., Deek, J., Biyiklioglu, S., Frascaroli, E., Corneti,
12 S., Salvi, S., Sonnante, G., Desiderio, F., Marè, C., Crosatti, C., Mica, E., Özkan, H.,
13 Kilian, B., De Vita, P., Marone, D., Joukhadar, R., Mazzucotelli, E., Nigro, D.,
14 Gadaleta, A., Chao, S., Faris, J.D., Melo, A.T.O., Pumphrey, M., Pecchioni, N.,
15 Milanesi, L., Wiebe, K., Ens, J., MacLachlan, R.P., Clarke, J.M., Sharpe, A.G., Koh,
16 C.S., Liang, K.Y.H., Taylor, G.J., Knox, R., Budak, H., Mastrangelo, A.M., Xu, S.S.,
17 Stein, N., Hale, I., Distelfeld, A., Hayden, M.J., Tuberosa, R., Walkowiak, S., Mayer,
18 K.F.X., Ceriotti, A., Pozniak, C.J., Cattivelli, L., 2019. Durum wheat genome highlights
19 past domestication signatures and future improvement targets. *Nat. Genet.*
20 <https://doi.org/10.1038/s41588-019-0381-3>
- 21 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer,
22 R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song,
23 J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015.
24 CDD: NCBI's conserved domain database. *Nucleic Acids Res.*
25 <https://doi.org/10.1093/nar/gku1221>
- 26 Masiero, S., Colombo, L., Grini, P.E., Schnittger, A., Kater, M.M., 2011. The emerging
27 importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*.
28 <https://doi.org/10.1105/tpc.110.081737>
- 29 Messenguy, F., Dubois, E., 2003. Role of MADS box proteins and their cofactors in
30 combinatorial control of gene expression and cell development. *Gene*.
31 [https://doi.org/10.1016/S0378-1119\(03\)00747-9](https://doi.org/10.1016/S0378-1119(03)00747-9)
- 32 Minh, B.Q., Nguyen, M.A.T., Von Haeseler, A., 2013. Ultrafast approximation for
33 phylogenetic bootstrap. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/mst024>
- 34 Moore, S., Vrebalov, J., Payton, P., Giovannoni, J., 2002. Use of genomics tools to isolate
35 key ripening genes and analyse fruit maturation in tomato. *J. Exp. Bot.*
36 <https://doi.org/10.1093/jxb/erf057>
- 37 Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: A virus classification tool based on
38 pairwise sequence alignment and identity calculation. *PLoS One*.
39 <https://doi.org/10.1371/journal.pone.0108277>
- 40 Nam, J., Kim, J., Lee, S., An, G., Ma, H., Nei, M., 2004. Type I MADS-box genes have
41 experienced faster birth-and-death evolution than type II MADS-box genes in
42 angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* <https://doi.org/10.1073/pnas.0308430100>
- 43 Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and
44 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*

- 1 Biol. Evol. <https://doi.org/10.1093/molbev/msu300>
- 2 Ning, P., Liu, C., Kang, J., Lv, J., 2017. Genome-wide analysis of WRKY transcription
- 3 factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit
- 4 condition. PeerJ. <https://doi.org/10.7717/peerj.3232>
- 5 Paolacci, A.R., Tanzarella, O.A., Porceddu, E., Varotto, S., Ciaffi, M., 2007. Molecular and
- 6 phylogenetic analysis of MADS-box genes of MIKC type and chromosome location of
- 7 SEP-like genes in wheat (*Triticum aestivum* L.). Mol. Genet. Genomics.
- 8 <https://doi.org/10.1007/s00438-007-0285-2>
- 9 Pařenicová, L., De Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E.,
- 10 Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., Colombo, L., 2003. Molecular
- 11 and phylogenetic analyses of the complete MADS-Box transcription factor family in
- 12 *Arabidopsis*: New openings to the MADS world. Plant Cell.
- 13 <https://doi.org/10.1105/tpc.011544>
- 14 Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L.,
- 15 Davey, M., Jacobs, J., Van Ex, F., Pasha, A., Khedikar, Y., Robinson, S.J., Cory, A.T.,
- 16 Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout,
- 17 C.J., Chalhoub, B., Mayer, K.F.X., Benhamed, M., Latrasse, D., Bendahmane, A.,
- 18 Wulff, B.B.H., Appels, R., Tiwari, V., Datla, R., Choulet, F., Pozniak, C.J., Provart,
- 19 N.J., Sharpe, A.G., Paux, E., Spannagl, M., Bräutigam, A., Uauy, C., 2018. The
- 20 transcriptional landscape of polyploid wheat. Science (80-.). 361.
- 21 <https://doi.org/10.1126/science.aar6089>
- 22 Riechmann, J.L., Meyerowitz, E.M., 1997. MADS domain proteins in plant development.
- 23 Biol. Chem. <https://doi.org/10.1515/bchm.1997.378.10.1079>
- 24 Rounsley, S.D., Ditta, G.S., Yanofsky, M.F., 1995. Diverse roles for MADS box genes in
- 25 *Arabidopsis* development. Plant Cell. <https://doi.org/10.1105/tpc.7.8.1259>
- 26 Ruelens, P., De Maagd, R.A., Proost, S., Theißen, G., Geuten, K., Kaufmann, K., 2013.
- 27 FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific
- 28 MADS-box genes. Nat. Commun. <https://doi.org/10.1038/ncomms3280>
- 29 Saedler, H., Becker, A., Winter, K.U., Kirchner, C., Theißen, G., 2001. MADS-box genes are
- 30 involved in floral development and evolution. Acta Biochim. Pol.
- 31 https://doi.org/10.18388/abp.2001_3920
- 32 Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F.,
- 33 Coupland, G., 2000. Distinct roles of constans target genes in reproductive development
- 34 of *Arabidopsis*. Science (80-.). <https://doi.org/10.1126/science.288.5471.1613>
- 35 Schilling, S., Kennedy, A., Pan, S., Jermini, L.S., Melzer, R., 2020. Genome-wide analysis of
- 36 MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation
- 37 and putative neofunctionalization. New Phytol. 225, 511–529.
- 38 <https://doi.org/10.1111/nph.16122>
- 39 Sharma, N., Ruelens, P., D'Hauw, M., Maggen, T., Dochy, N., Torfs, S., Kaufmann, K.,
- 40 Rohde, A., Geuten, K., 2017. A flowering locus C homolog is a vernalization-regulated
- 41 repressor in *Brachypodium* and is cold regulated in Wheat. Plant Physiol.
- 42 <https://doi.org/10.1104/pp.16.01161>
- 43 Shi, S. ya, Zhang, F. fei, Gao, S., Xiao, K., 2016. Expression pattern and function analyses of

- 1 the MADS transcription factor genes in wheat (*Triticum aestivum* L.) under
2 phosphorus-starvation condition. *J. Integr. Agric.* [https://doi.org/10.1016/S2095-3119\(15\)61167-4](https://doi.org/10.1016/S2095-3119(15)61167-4)
- 4 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam,
5 H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable
6 generation of high-quality protein multiple sequence alignments using Clustal Omega.
7 *Mol. Syst. Biol.* <https://doi.org/10.1038/msb.2011.75>
- 8 Smaczniak, C., Immink, R.G.H., Angenent, G.C., Kaufmann, K., 2012. Developmental and
9 evolutionary diversity of plant MADS-domDevelopmental and evolutionary diversity of
10 plant MADS-domain factors: Insights from recent studiesain factors: Insights from
11 recent studies. *Dev.* <https://doi.org/10.1242/dev.074674>
- 12 Stum, A., Quackenbush, J., Trajanoski, Z., 2002. Genesis: Cluster analysis of microarray
13 data. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/18.1.207>
- 14 Wang, Q., Du, M., Wang, S., Liu, L., Xiao, L., Wang, L., Li, T., Zhuang, H., Yang, E., 2018.
15 MADS-box transcription factor madsA regulates dimorphic transition, conidiation, and
16 germination of *Talaromyces marneffe*. *Front. Microbiol.*
17 <https://doi.org/10.3389/fmicb.2018.01781>
- 18 Wang, Z., Wang, F., Hong, Y., Yao, J., Ren, Z., Shi, H., Zhu, J.K., 2018. The Flowering
19 Repressor SVP Confers Drought Resistance in Arabidopsis by Regulating Absciscic Acid
20 Catabolism. *Mol. Plant.* <https://doi.org/10.1016/j.molp.2018.06.009>
- 21 Wei, B., Zhang, R.Z., Guo, J.J., Liu, D.M., Li, A.L., Fan, R.C., Mao, L., Zhang, X.Q., 2014.
22 Genome-wide analysis of the MADS-Box gene family in *Brachypodium distachyon*.
23 *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0084781>
- 24 Whittaker, C., Dean, C., 2017. The FLC locus: A platform for discoveries in epigenetics and
25 adaptation. *Annu. Rev. Cell Dev. Biol.* <https://doi.org/10.1146/annurev-cellbio-100616-060546>
- 27 Wu, Y., Ke, Y., Wen, J., Guo, P., Ran, F., Wang, M., Liu, M., Li, P., Li, J., Du, H., 2018.
28 Evolution and expression analyses of the MADS-box gene family in *Brassica napus*.
29 *PLoS One.* <https://doi.org/10.1371/journal.pone.0200762>
- 30 Xu, K., He, X., Dreisigacker, S., He, Z., Singh, P.K., 2020. Anther extrusion and its
31 association with Fusarium head blight in CIMMYT wheat germplasm. *Agronomy.*
32 <https://doi.org/10.3390/agronomy10010047>
- 33 Yang, C., Liu, H., Li, G., Liu, M., Yun, Y., Wang, C., Ma, Z., Xu, J.R., 2015. The MADS-
34 box transcription factor FgMcm1 regulates cell identity and fungal development in
35 *Fusarium graminearum*. *Environ. Microbiol.* <https://doi.org/10.1111/1462-2920.12747>
- 36 Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., Meyerowitz, E.M.,
37 1990. The protein encoded by the Arabidopsis homeotic gene *agamous* resembles
38 transcription factors. *Nature.* <https://doi.org/10.1038/346035a0>
- 39 Yu, Y., Ouyang, Y., Yao, W., 2018. ShinyCircos: An R/Shiny application for interactive
40 creation of Circos plot. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btx763>
- 41 Zhang, Z., Li, H., Qin, G., He, C., Li, B., Tian, S., 2016. The MADS-Box transcription factor
42 Bcmads1 is required for growth, sclerotia production and pathogenicity of *Botrytis*

- 1 cinerea. Sci. Rep. <https://doi.org/10.1038/srep33901>
- 2 Zhao, Y., Li, X., Chen, W., Peng, X., Cheng, X., Zhu, S., Cheng, B., 2011. Whole-genome
- 3 survey and characterization of MADS-box gene family in maize and sorghum. Plant
- 4 Cell. Tissue Organ Cult. 105, 159–173. <https://doi.org/10.1007/s11240-010-9848-8>
- 5 Zimin, A. V., Puiu, D., Hall, R., Kingan, S., Clavijo, B.J., Salzberg, S.L., 2017. The first
- 6 near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*.
- 7 Gigascience. <https://doi.org/10.1093/gigascience/gix097>
- 8

1 **Table 1**

2 Potential relationship between unbalanced homeologs and gene duplications in wheat
3 MADS-box gene family.

Subfamily	Total Homeologs ¹	Balanced Homeologs ²	Unbalanced Homeologs ³	Gene Duplications
<i>Mα</i>	53	33	20	Present
<i>Mβ</i>	28	10	18	Present
<i>Mγ</i>	47	26	21	Present
<i>AG/STK</i>	12	12	0	Absent
<i>AGL12</i>	6	6	0	Absent
<i>AGL17</i>	31	22	9	Present
<i>AGL6</i>	3	3	0	Absent
<i>AP1</i>	9	9	0	Absent
<i>AP3</i>	6	6	0	Present ⁴
<i>Bsister</i>	19	10	9	Present
<i>MIKC*</i>	27	21	6	Present
<i>OsMADS32</i>	3	3	0	Absent
<i>PI</i>	6	6	0	Absent
<i>SEP</i>	28	25	3	Present
<i>SOC1</i>	13	12	1	Present
<i>SVP</i>	9	9	0	Absent

4 ¹, Total number of gene copies from three different wheat genomes (A, B, D); ², At least one
5 gene copy is present in all diploid genomes; ³, Gene copy is missing in one or two genomes;
6 ⁴, Exception-gene duplications identified despite holding balanced homeologs.

7

Fig. 1. Comparative phylogenetic analysis based subfamily classifications of wheat MADS-box genes.

Arabidopsis, rice and wheat MADS-box proteins were MAFFT aligned (Katoh et al., 2018; Katoh and Standley, 2013), maximum likelihood phylogenies inferred using IQ-TREE software (Nguyen et al., 2015) and final trees visualized with MEGA7 (Kumar et al., 2016). Separate phylogenetic trees were inferred among *Arabidopsis*, rice and wheat MIKC-type (**A**) and M-type (**B**) proteins. Wheat MADS-box genes are highlighted with solid black circles. Only bootstrap values $\geq 50\%$, as calculated from 1000 replicates, could be displayed on the tree nodes. Subfamily-specific colouring was adopted for differentiating the different subfamilies. Subfamily names (outer bands) were given by following the subfamily classifications in *Arabidopsis* and/ or major grass species (Gramzow and Theißen, 2015).

Fig. 2. Structural differentiation between *FLC*-like and *MIKC**-like MADS-box genes.

Amino acid sequences of representative genes from both subfamilies were aligned using Clustal Omega (Sievers et al., 2011) with default parameters and multiple sequence alignment visualized with MEGA7 (Kumar et al., 2016). *FLC*-like genes (above solid black line) were separated from *MIKC**-like genes (below solid black line). The conserved and diverged residues were indicated with a star (*) and asterisk (#) symbols, respectively. Multiple sequence alignment demonstrated conservation of amino acid residues/motifs within both subfamilies, whereas diversification between subfamilies.

Fig. 3. Subfamily-specific gene duplications among wheat MADS-box genes.

Duplicated genes were plotted in a circular diagram using respective physical positions with shinyCircos (Yu et al., 2018). The outer track indicates three different subgenomes (shades of green) and inner track represents chromosomal segments (R and R3, purple; R2a and R2b, sky blue; C, light grey) (IWGSC, 2018). Duplicated genes were identified through sequence similarity (**Table S6**; see material and method section) and linked with subfamily-specific colours as in **Fig. 1A & 1B**, except for *AGL17*-like genes which were linked using chocolate colour. The linked duplicated genes positioned on different wheat chromosomes represent segmental duplications, whereas tandem duplications were indicated by incomplete links within the same chromosomes.

Fig. 4. Box-plots showing evolution patterns in M- and MIKC-type wheat MADS-box genes.

The coding sequences of the duplicated genes were MAFFT aligned (Katoh et al., 2018; Katoh and Standley, 2013), masked and subjected to SNAP V2.1.1 program for computing the synonymous (Ks) and non-synonymous (Ka) substitution rates. The divergence time between duplicated genes pairs was estimated in million years ago (MYA) by following El Baidouri et al. (2017) and box-plots were generated using Microsoft Excel 2016.

Fig. 5. Expression patterns of wheat MADS-box genes under stressful conditions.

RNA-seq based expression data of all MADS-box genes were retrieved from exVIP wheat expression browser (Borrill et al., 2016; Ramírez-González et al., 2018) and a heatmap was generated with Genesis software (Sturn et al., 2002). Processed expression levels of all genes under different abiotic (I, spikes with water stress; II, drought & heat-stressed seedlings; III, seedling treated with PEG to induce drought; IV, shoots after two weeks of cold stress; V, phosphate starvation in roots/shoots/leaves) and biotic stresses (VI, coleoptile infection with *Fusarium pseudograminearum*/crown root; VII, FHB infected spikelets (0 to 48 hours); VIII, FHB infected spikelets (30 to 50 hours); IX, spikes inoculated with FHB and ABA/GA; X, CS spikes inoculated with FHB; XI, leaves naturally infected with *Magnaporthe oryzae*; XII, PAMP inoculation of seedlings; XIII, stripe rust infected seedlings; XIV, stripe rust and powdery mildew infection in seedlings; XV, *Septoria tritici* infected seedlings; XVI, *Zymoseptoria tritici* infected seedlings) (columns) and in different subfamilies (rows) are presented as Log₂ transcripts per million (Log₂ TPM). Detailed information about expression levels of individual genes in tested tissues/growth stages and stresses/diseases are provided in **Table S7**.

Fig. S1. Comparative phylogenetic tree between *Arabidopsis* and wheat MADS-box genes.

Arabidopsis and wheat MADS-box proteins were MAFFT aligned using only MADS domain (L-INS-i algorithm), maximum likelihood phylogenies inferred using IQ-TREE software and final tree visualized with MEGA7. M- and MIKC-type genes were highlighted with green and red colour nodes, respectively.

Fig. S2. Comparative phylogenetic tree between rice and wheat MADS-box genes.

Rice and wheat MADS-box proteins were MAFFT aligned using only MADS domain (L-INS-i algorithm), maximum likelihood phylogenies inferred using IQ-TREE software and final tree visualized with MEGA7. M- and MIKC-type genes were highlighted with green and red colour nodes, respectively.

Fig. S3. Distribution pattern of MADS-box genes on wheat chromosomes and sub-genomes.

Fig. S4. Chromosomal mapping of wheat MADS-box genes.

Solid black vertical bars represent individual chromosomal lengths (Mbp). Gene names and their initial physical positions are drawn on the left and right side of the respective chromosomes, respectively. M-type genes are circled with blue boxes.