NEW RESULTS

Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications may have facilitated adaptation to different environmental conditions

Susanne Schilling[1], *Alice Kennedy*[1,2], *Sirui Pan*[1], *Lars S. Jermiin*[1,3] *and Rainer Melzer*[1*]

[1]School of Biology and Environmental Science, University College Dublin, Ireland

[2]present address: Institute for Microbiology and Plant Biology, Katholieke Universiteit Leuven, Belgium

[3]Research School of Biology, Australian National University, Australia

*corresponding author. Telephone: 00353 1 716 2290

Author emails:

susanne.schilling@ucd.ie

alice.kennedy@kuleuven.be

sirui.pan@ucdconnect.ie

lars.jermiin@anu.edu.au

rainer.melzer@ucd.ie

Date of submission: 26.03.2019

Tables and Figures: 2 Tables, 6 Figures

Word count: ca. 5408

Running title: MIKC-type MADS-box genes in wheat

Abstract

Background Wheat (*Triticum aestivum*) is one of the most important crops worldwide. Given a growing global population coupled with increasingly challenging climate and cultivation conditions, facilitating wheat breeding by fine-tuning important traits such as stress resistance, yield and plant architecture is of great importance. Since they are involved in virtually all aspects of plant development and stress responses, prime candidates for improving these traits are MIKC-type (type II) MADS-box genes.

Results We present a detailed overview of number, phylogeny, and expression of 201 wheat MIKC-type MADS-box genes, which can be assigned to 15 subfamilies. Homoeolog retention is significantly above the average genome-wide retention rate for wheat genes, indicating that many MIKC-type homoeologs are functionally important and not redundant. Gene expression is generally in agreement with the expected subfamily-specific expression pattern, indicating broad conservation of function of MIKC-type genes during wheat evolution.

We find the extensive expansion of some MIKC-type subfamilies to be correlated with their chromosomal location and propose a link between MADS-box gene duplications and the adaptability of wheat. A number of MIKC-type genes encode for truncated proteins that lack either the DNA-binding or protein-protein interaction domain and occasionally show novel expression patterns, possibly pointing towards neofunctionalization.

Conclusions Conserved and neofunctionalized MIKC-type genes may have played an important role in the adaptation of wheat to a diversity of conditions, hence contributing to its importance as a global staple food. Therefore, we propose that MIKC-type MADS-box genes are especially well suited for targeted breeding approaches and phenotypic fine tuning.

Keywords MADS-box gene, MADS-domain, wheat, *Triticum aestivum*, transcription factors, IWGSC, gene duplication, neofunctionalization, crop breeding

Background

Bread wheat (*Triticum aestivum*) is one of the most important crops worldwide, contributing a significant amount of calories and proteins to the global human diet [1, 2]. Bread wheat is hexaploid and was first domesticated some 8 - 25,000 years ago in the region that today is called the middle-east [3, 4]. Wheat originated from three diploid progenitor species: *Triticum urartu* (A-genome donor), an *Aegilops speltoides*-related grass (B-genome donor) and *Aegilops tauschii* (D-genome donor) [5]. Because of its hexaploidy and an abundance of repetitive and transposable elements, bread wheat has one of the largest crop plant genomes (approximately 16 Gbp), making it challenging to work with from a genetics, genomics and breeding perspective [6]. However, recent advances in sequencing technology have led to a high-quality genome assembly and annotation by the International Wheat Genome Sequencing Consortium (IWGSC) [7]. Further, large scale RNA-seq analyses provided insights into expression patterns of homoeologous genes in different development stages and under a variety of stress conditions; building a rich resource for more detailed analyses [8].

Transcription factors (TFs) are a major driver in evolution as well as in domestication and bear the potential for crop improvement and trait fine-tuning [9]. MADS-box genes constitute one of the largest families of plant TFs [10]. They can be divided into two phylogenetically distinct groups: type I and type II [11]. While the function of most type I MADS-box genes remains to be illuminated, several type II genes are key domestication genes in different eudicot and monocot crops (reviewed in [12]). Plant type II MADS-domain proteins possess a typical domain structure, which is composed of the MADS-, I-, K- and C-terminal domain [13]. The MADS-domain enables the DNA-binding, nuclear localization and dimerization of the TF [14, 15], while I- and K-domain facilitate dimerization and higher-order complex formation of two or more MADS-domain proteins [16, 17]. The C-terminal domain allows for transcriptional activation of some MADS-domain proteins [18]. Because of this

75 characteristic domain structure, type II genes are also referred to as MIKC-type MADS-box
76 genes [13]. MIKC-type MADS-box genes are involved in virtually all aspects of plant
77 development, including the root, flower, seed and embryo [19]. They have also been reported
78 to be involved in different stress responses [20-22]. Thus, understanding MIKC-type MADS-
79 box genes is important for understanding plant development, which is in turn crucial for plant
80 breeding and crop improvement [23].
81 MIKC-type MADS-box genes have been phylogenetically and functionally characterized in a
82 variety of model systems (Arabidopsis (*Arabidopsis thaliana*) and *Brachypodium distachyon*
83 [24, 25]) as well as important crop plants (banana, rice, brassica, cotton [21, 26-28]).
84 Individual wheat MIKC-type MADS-box genes have also been studied for almost two
85 decades. The most prominent example is probably *VERNALIZATION1* (*VRN1*), an
86 *APETALA1* (*AP1*)-like key regulator of flowering time as well as floral meristem
87 determination [29-33] and one of the most important loci that distinguishes spring from
88 winter wheat varieties [33]. Other wheat MADS-box genes have been implicated in the
89 control of flowering time, ovule development and pistilloidy [34-37]. However, a detailed
90 genome-wide phylogenetic and functional characterization of wheat MIKC-type MADS-box
91 genes is still missing.
92 To better understand the dynamics of MIKC-type gene evolution in wheat and to facilitate
93 future research on this important TF family, we provide a detailed overview of the number,
94 phylogeny and expression of MIKC-type MADS-box genes in the recently released genome
95 of *Triticum aestivum* [7]. We find a number of wheat MIKC-type subfamilies to be
96 significantly larger than expected and suggest that extensive sub- and neofunctionalization in
97 those subfamilies contributed to the global distribution of wheat.
98

Results

The wheat genome contains 201 MIKC-type MADS-box genes

A total of 439 coding sequences (including splice variants) were identified on the basis of the functional annotation (PFAM domains) in the recently released IWGSC wheat genome (Table S1) [7]. This dataset was simplified by keeping only one splice variant from each genomic locus for further analyses (Table S2). MIKC-type MADS-box genes were differentiated from type I MADS-box genes using a phylogenetic approach (Table S1 see Methods for details). An additional eight MIKC-type genes were identified using BLAST search. Altogether, we identified 201 MIKC-type (type II) MADS-box genes in wheat (Table 1, Table S2). Because MIKC-type MADS-box gene nomenclature in wheat is currently not consistent, with several genes having several synonymous names, we renamed all genes according to their subfamily association (Table 1, Table S2, see Methods for details).

The domain structure of MIKC-type MADS-domain proteins is crucial for their function, MADS- and K-domain being indispensable for DNA-binding and protein complex formation, respectively [13], although genes encoding truncated proteins may act as dominant negative versions [38, 39]. A total of 164 out of 201 MIKC-type genes encoded for both, a MADS- as well as a K-domain, while 29 genes lacked a K-box (14 %) and 7 lacked a MADS-box (3 %) (Table 1, Table S2, Figure 1). One gene encoded for a MADS- and an SRP54–domain (*TaBS.8A*, *TraesCS4A01G044400LC*) (Figure S3B, Table S2).

For a number of genes, the predicted gene structures were very long (30 kb and above, Table S2). We confirmed mRNA transcripts spanning over especially long introns (ca. 29 kbp) for two selected genes from the *FLC*-clade using RT-PCR and Sanger sequencing (Figure S1).

MIKC-type MADS-box genes belong to well-defined subfamilies

A Maximum Likelihood phylogenetic tree of all MIKC-type MADS-box genes from *Arabidopsis thaliana*, rice (*Oryza sativa*) and wheat shows that the wheat genome retained all 15 major grass MIKC-type MADS-box gene subfamilies: *AP1* (*SQUAMOSA* (*SQUA*)), *AP3 (DEF (DEFICIENS))*, *PISTILLATA* (*PI; GLOBOSA* (*GLO*)), *AGAMOUS* (*AG*) /*SEEDSTICK* (*STK*), *AG-LIKE6* (*AGL6*), *AGL12*, *AGL17*, *BSISTER* (*BS*; *GNETUM GNEMON MADS13,* (*GGM13*)), *SUPPRESSOR OF CONSTANS1* (*SOC1*), *SHORT VEGETATIVE PHASE* (*SVP*; *StMADS11*), *MIKC**, *OsMADS32*, *FLOWERING LOCUS C* (*FLC*), *SEPALLATA1*(*SEP1*) and *SEP3* (Figure 1, Figure S2) [40].

In many subclades, the gene phylogeny roughly followed species phylogeny; with the *Arabidopsis* genes displaying a sister-group relationship to the grass genes; and one or more rice MADS-box genes closely related to a triad of three wheat homoeologs (e.g. the *SVP*-, *AGL12*-, *OsMADS32*-, *MIKC**, *AP1*- and *AG*-subclades (Figure 1)). The topology in other subclades is more complex, suggesting multiple duplication events, before and/or after polyploidization of wheat (e.g. *SOC1*- and *SEP1* subclades, Figure S3A). Especially *Bsister*-, *AGL17*- and *FLC*- subclades are significantly expanded in wheat compared to Arabidopsis and rice (Figure 1, Figure 2, Figure S3B-D).

Wheat MIKC-type genes exhibit a high rate of homoeolog retention and gene duplication

In many flowering plants, the number of MIKC-type MADS-box genes is between 40 and 70 [41]. Rice and Arabidopsis for example, despite their phylogenetic distance, have a similar number of MIKC-type MADS-box genes (43 and 45, respectively) [24, 42, 43]. With 201 genes, the total number of MIKC-type MADS-box genes in wheat is among the highest of hitherto characterized flowering plant species [27, 44, 45]. This is partly due to the hexaploid

149 nature of wheat. However, even when corrected for ploidy-level, the number of MIKC-type
150 MADS-box genes in bread wheat was significantly higher than in rice (≈1.5 fold higher, χ^2
151 test $p = 0.028$, Figure 2A-D). Even when truncated genes lacking either a MADS-box or a K-
152 box were excluded, the gene number was still higher than in rice (164/3=54.7 > 42). This
153 increase in number is mainly due to the gene count in four subfamilies (*SEP1-, AGL17-*,
154 *FLC*- and *Bsister*-like genes) that are significantly larger than expected (χ^2 test $p = 0.002$,
155 $p << 0.001$, $p << 0.001$, $p << 0.001$, respectively; Figure 2A-D). The remaining subclades
156 showed the expected 3:1 ratio of wheat-to-rice genes; and some were below the expected
157 ratio (Figure 2D). In some of the latter cases, wheat orthologs of rice genes could not be
158 identified, indicating gene loss in the lineage leading to *Triticum* (e.g. *AP1*- and *AGL6*-like
159 genes, Figure 1, Figure 2A-D).
160 To better understand why MIKC-type MADS-box genes are so abundant in the wheat
161 genome, we analyzed homoeologous groups in detail (Table 2). Approximately one third of
162 all wheat genes (i.e. all genes annotated in the current version of the wheat genome) are
163 present in homoeologous groups of 3, also termed triads (1:1:1; 35.8 % of genes) [7]. In
164 contrast, almost two thirds of the 201 MIKC-type genes identified are present in triads (62.7
165 %, Table 2). If only 'full-length' MADS-box genes (defined here as containing a MADS- and
166 K-box) are considered, this ratio is even higher (72.7 %, Table 2). Also, the percentage of
167 MIKC-type genes with homoeolog-specific duplications is higher for MIKC-type genes as
168 compared to all wheat genes (8.5 % vs. 5.7 % Table 2). Loss of one homoeolog, on the other
169 hand, is less pronounced in MIKC-type genes (1 % vs. 13.2 %, Table 2). 38 genes were
170 excluded from the analysis (18.9 %; 13 % for full-length only), because the relationship of
171 the genes could not be reliably resolved. Thus, the high homoeolog retention rate can partly
172 explain the high number of wheat MIKC-type genes.
173

In addition to this, a variety of duplication patterns were observed in some subfamilies (Figure S3, Supplementary Text 1). For example, the rice *SEP1*-like genes *OsMADS1* and *OsMADS5* are sister to 10 and 7 wheat genes, respectively. In this case, the phylogenetic analysis suggests gene duplications in the lineage leading to *Triticum* but before the polyploidization of wheat (Figure S3A). The *Bsister* gene *OsMADS30* from rice is sister to 27 wheat genes, many of which lack a MADS- or K-box and are found in non-syntenic regions, indicating gene amplification occurred through transposable elements (Table 1, Figure S3B). For the *AGL17*- and *FLC*-subfamily from wheat, a number of triads can be assigned to a single rice gene (Figure S3C, D). Several genes from these two subfamilies are found in close vicinity to each other, pointing towards tandem duplications as a mechanism for subfamily expansion (Table 1, Figures 2E and S2C, S2D).

Gene duplications and truncated genes are prevalent among MIKC-type genes in subtelomeric segments

MIKC-type MADS-box genes were generally equally distributed among the chromosomes, the only exception being the three homoeologous chromosomes 7, which contained significantly more genes than expected from the chromosome lengths (χ^2 test $p << 0.001$, Figure 2E). This is mainly due to *AGL17*-like genes; with the majority of them located in tandem locations on the distal subtelomeric ends of chromosomes 7 (Figure 2E, Table S2). Overall, MIKC-type genes were located equally likely in the more central segments of the chromosomes (R2a, R2b and C) and in the subtelomeric parts of the chromosomes (R1 and R3) (48 and 50 % of genes, respectively). However, gene location varied greatly among subfamilies (Figure 3). In general, genes belonging to smaller subfamilies tended to be located in more central segments of the chromosomes; whereas a larger percentage of genes belonging to more expanded subfamilies were located in subtelomeric segments (Figure 4,

Table S2). Further, full-length MIKC-type genes were about equally likely to be located within sub-telomeric versus central segments (45 % vs. 54 %, respectively), but truncated genes (MADS- or K-box only) were two times as prevalent in subtelomeric segments (61 % vs. 31 %; 8 % of genes were not assigned to a chromosome) (Figure 3A). Genes in subtelomeric segments were often found to be in close vicinity to each other, with over half of the genes in subtelomeric segment R1 being under 1000 kbp downstream of the closest MIKC-type gene (Table S3). In contrast, only 15 % of genes in the centromere segment have a distance under 1000 kb to the next downstream MIKC-type gene (Table S3). This might be due to more frequent tandem duplication events in subtelomeric segments. These findings are in line with the observation that subtelomeric segments are targets of recombination events and that many fast evolving genes lie within these segments [46].

Conserved and divergent patterns of MIKC-type MADS-box gene expression during wheat development

To characterize the expression of wheat MIKC-type MADS-box genes, we analyzed RNA-seq data of 193 wheat MADS-box genes [8, 47]. Out of the 159 full-length genes, 83 % were expressed in at least one developmental stage, with a wide expression range with a maximum count of 1 to 424 transcripts per million (tpm_{max}) (Figure 5, Table S2). The remaining 17 % of full-length genes showed a very low expression with a tpm_{max} below 1 (Figure 5A, Table S2). Of the 33 truncated genes, encoding only for either K or MADS-domain, 30 % were expressed (4 and 6 genes, respectively, tpm_{max}1 to 51; Figure 5A, Table S2). One gene, encoding for MADS- and an SRP54-domain was expressed ubiquitously in the plant, albeit at a low level (*TaBS.8A*; tpm_{max}3.12; Table S2, Figure S3).

In general, MADS-box genes expression patterns are comparable with findings in rice [21, 48, 49] (Figure 5, Figure S4). MIKC*-type genes are expressed in anthers (Figure 5A, Table

S7). Putative floral homeotic genes, such as *PI/AP3*-like and *AG/STK*-like genes, *OsMADS29*- and *OsMADS31*-like *Bsister* genes, *AGL6*-like genes as well as *SEP1* and *SEP3*-like genes are expressed specifically during flower and seed development (Figure 5). *AGL17*-like genes are expressed in roots. *AP1*-like, *SVP*-like and some *SEP1*-like genes are expressed ubiquitously throughout the plant life cycle in different tissues.

For further analysis, genes were hierarchically clustered according to expression similarities and then grouped into different expression modules (Figure 5B, Figure S4). This analysis showed that, genes from one subfamily could differ considerably in their expression pattern. For example, members of the *SEP1*-like gene subfamily are grouped into 6 different modules and *Bsister* genes are found in 5 different modules (Figure 5B). In contrast, *SVP*- and *AP1*-like genes show relatively little variation in their expression pattern (Figure 5B). It is also noteworthy that 79 genes including representatives from almost all subclades showed no expression or only low expression under very specific conditions during the developmental time course (41 % of genes; module 9; grey, Figure 5B).

AGL17-like and *Bsister* genes are expressed in response to stress conditions

AGL17- and *OsMADS30*-like *Bsister* clades have been expanding during wheat evolution (Figure 1, Figure 2). Many of the genes from these clades are not expressed or expressed on a very low level during wheat development (Figure 5). However, some of these genes do show expression in response to distinct stress conditions (Figures 6, S4).

Bsister genes form three distinct clades of *OsMADS29*- *OsMADS30*- and *OsMADS31*-like genes in grasses [50]. Most *Bsister* genes have been described to be important for ovule and seed development with a specific expression pattern limited to female reproductive organs [51-53]. While *OsMADS29*-like and *OsMADS31*-like wheat *Bsister* genes follow this conserved expression pattern (Figure 6A, *TaBS.1* and *TaBS.2*), some *OsMADS30*-like wheat

249 genes are expressed ubiquitously during the plant life cycle (Figure 6A, *TaBS.6B1*
250 *TaBS.5A4*). In contrast, five closely related *OsMADS30*-like genes showed low or no
251 expression in any of the developmental stages. Instead, a specific upregulation in response to
252 inoculation with the pathogen *Fusarium graminearum* was detected (Figure 6A, *TaBS.4*;
253 Figure S5).
254 *AGL17*-like genes are commonly expressed in roots and leaves [54]. They have also been
255 described to be involved in osmotic and saline stress responses in rice [54]. Several wheat
256 *AGL17*-like genes are not expressed at all during developmental stages (Figure 5, grey;
257 Figure 6B). Two *AGL17*-like genes that are expressed in the root, but not in leaves under
258 control conditions, show upregulation in leaf tissue after 1 hour of heat stress (Figure 6B,
259 genes 2,3). However, gene expression is not detectable after 6 h and there is no specific
260 response to drought stress (Figure 6B, genes 2,3). Another *AGL17*-like gene is upregulated
261 after 6h of heat, but not drought stress (Figure 6B, gene 4). A fifth *AGL17*-like gene is
262 expressed in the root, and additionally upregulated in leaves in response to infection with
263 stripe rust 7 days after infection (Figure 6C).
264

Discussion

Many wheat MIKC-type MADS-box genes have evolutionarily conserved functions

MIKC-type MADS-box genes play a central role in plant development. They are therefore promising targets for crop breeding and improvement [55].

Here, we identified 201 wheat MIKC-type MADS-box genes, which we assigned to 15 conserved subfamilies (Table 1, Figure 1). At least 70 % of wheat MIKC-type MADS-box genes could be assigned to homoeologous groups with genes in every subgenome (Table 2). This is considerably above the average homoeologous retention rate in wheat (42 %, Table 2 [7]). Many MADS-domain proteins act in multi-protein complexes [56]. The composition of those protein complexes changes dynamically during development, with many proteins being part of more than one complex [57, 58]. Changes in the gene dosage may result in changes in the stoichiometry of the protein complexes which may in turn have detrimental phenotypic effects [59]. Thus, selection may act to retain homoeologs in all subgenomes.

We also found that the expression pattern of many wheat MADS-box genes is similar to that of close homologs in rice and other model plants, indicating that gene functions are broadly conserved between wheat and rice.

Together, the conservation of all major subclades, the high homoeolog retention rate and the conservation of expression patterns underline the high biological importance of the MIKC-type MADS-box gene family in general and the distinct subclades in particular. Hence, the rich knowledge about the developmental role of MIKC-type MADS-box genes from model plants, combined with a complete picture of gene number, expression data and phylogenetic analyses in wheat will aid wheat breeding and improvement.

Subfamily-specific expansion of wheat MIKC-type MADS-box genes may contribute to the high adaptability of wheat

The hexaploid nature of wheat and the large size of the MADS-box gene family provide an ideal opportunity to study the evolutionary fate of genes after gene duplication and polyploidization.

With 201 genes, wheat has one of the largest MIKC-type MADS-box gene counts among flowering plants [27, 44, 45]. In total, wheat has about 3.1 times as many transcription factors as rice, which can generally be explained by its hexaploidy [60]. However, the number of MIKC-type MADS-box genes is more than 4.5 times higher in wheat than in rice (Figure 2A-D). The strikingly high number of MIKC-type MADS-box genes observed in this study is mainly due to – leaving hexaploidy aside – the significant expansion of four subclades: *SEP1*, *Bsister*, *AGL17* and *FLC* (Figure 2).

The expansion of MIKC-type subfamilies has been reported before in different plant species, such as for *SOC1*-like genes in *Eucalyptus* as well as *SVP*-like and *SOC1*-like genes in cotton [27, 45]. Intriguingly, genes from *FLC*-, *SVP*- and *SOC1*- subfamilies are involved in the control of flowering time [12, 61, 62]. It has been hypothesized that the expansion (and contraction) of developmental control genes, more specifically eudicot *FLC*-like genes, facilitate the rapid adaptation to changes in environmental factors such as temperature [63]. In a similar manner, duplications of wheat *FLC*-like genes might have enabled the adaptation of wheat to different climatic conditions, therefore contributing to its global distribution. It will be interesting to see whether copy number variations of *FLC*-like genes can be detected in different wheat varieties.

The expansion of *Bsister* and *AGL17*-like genes may similarly be explained with an adaptive advantage. However, in those cases neofunctionalization might be involved. *Bsister* like genes form three subclades of *OsMADS29*-, *OsMADS30*- and *OsMADS31*-like genes in

grasses. Expression pattern and evolutionary analyses suggest that grass *OsMADS29*- and *OsMADS31*-like genes retain a conserved role in ovule and fruit development [48, 64], whereas *OsMADS30*-like genes may have functionally diverged [50]. For five wheat *OsMADS30*-like genes, upregulation was observed during infection with Fusarium head blight (*Fusarium graminearum*, Figure 6A). MIKC-type MADS-box genes are not typically associated with biotic stress responses and it remains speculative whether and how they might be involved in responding to a *Fusarium* infection. However, Fusarium head blight is a floral disease and *Bsister* genes are expressed in the flower. This may have facilitated a co-option of theses genes into a pathogen response network. Interestingly, the lack of synteny between *OsMADS30*-like genes might point towards transposable elements as a possible duplication mechanism, underlining the evolutionary importance of transposable elements [65].

An upregulation in response to stresses was also observed for some *AGL17*-like genes, which form the largest wheat MIKC-type subfamily. While many wheat *AGL17*-like genes are not expressed (Figure 5), a number of them are upregulated in late stages of stripe rust infection and in response to heat stress (Figure 6B, C). Another three *AGL17*-like genes were found to be upregulated in some stages of anther and grain development, a pattern unusual for *AGL17*-like genes (Figure 5A). This diversity of expression patterns and putative functions adds to the complex evolution of *AGL17*-like genes, as genes from this subfamily have been implicated in various different functions including root development, flowering time control, tillering, stomata development and stress response [54, 66-69].

OsMADS30- and *AGL17*-like genes might be involved in other stress responses as well and might be good candidates to investigate cultivar-specific resistance to biotic and abiotic stresses.

339 **Dynamic evolution of MIKC-type MADS-box genes in subtelomeric regions**

340 The cause for the expansion of the *FLC*-, *AGL17*- *SEP1*- *OsMADS30*-like subfamilies might
341 be the chromosomal position of their genes. Subtelomeric distal chromosome segments have
342 been previously described as being targets of recombination events, and many fast evolving
343 genes lie within these evolutionary hot spots [46, 70]. In wheat specifically, genes related to
344 stress response and external stimuli, notably traits with a high requirement for adaptability,
345 have been found to be located in distal chromosomal segments [7]. In contrast, genes related
346 to photosynthesis, cell cycle or translation, e.g. genes with involved in highly conserved
347 pathways are enriched in proximal chromosomal segments [7].
348 This notion is supported by our findings: genes of the larger wheat MIKC-type subclades
349 tend to be located in subtelomeric segments (Figure 4, Figure 3). Remarkably, many of these
350 expanded clades do control traits that are important for adaptation to different environments.
351 For example, *AGL17*-like genes, are involved in root development and stress response and
352 *FLC*-like genes, determine flowering time. On the other hand, smaller MIKC-type subclades
353 involved in highly conserved developmental functions, such as *AP3/PI*- or *AG/STK*-like
354 genes, which control reproductive organ identity, tend to be located more in central
355 chromosomal segments (Figure 3). The higher prevalence of duplication events in
356 subtelomeric chromosomal segments might thus have caused the expansion of certain
357 subclades, possibly facilitating rapid adaptation to different environmental conditions. On the
358 other hand, there might be an evolutionary advantage for wheat MIKC-type subclades with
359 highly conserved functions to be located in proximal chromosomal positions: this way e.g.
360 developmentally detrimental gene dosage variations are minimized.
361
362 The high prevalence of gene duplications in subtelomeric segments most likely also led to a
363 higher proportion of truncated genes lacking either the MADS- or K-box (Figure 3A). This

might render the encoded protein functionally impaired, as suggested earlier for SEP1.2A (WLHS1A), which lacks a K-domain and shows no protein-protein interactions *in vivo* [71] (Figure 1, Table 1). However, MIKC-type MADS-domain proteins without a K-domain might theoretically still be able to bind DNA [13] and compete with full-length MIKC-type proteins for target sites, thus functioning as transcriptional inhibitors. On the other hand, genes lacking the MADS-box but encoding a K-domain might act in a dominant-negative manner by binding and sequestering other MADS-domain proteins [38, 39]. Evidence that dominant-negative versions of transcriptions factors can be evolutionary and developmentally important comes from basic helix-loop-helix proteins [72, 73]. We found *OsMADS30*-like wheat genes, encoding for only MADS, only K and an unusual combination of MADS and SRP54 domain expressed ubiquitously (Figures 5A, 6A), hence deviating from canonical *Bsister* expression in the flower and again hinting towards a possible neofunctionalization of *OsMADS30*-like genes during wheat evolution.

Conclusions

MIKC-type MADS-box genes are hugely important for wheat development and hence bear immense potential for the improvement of this economically highly relevant crop. Our data indicate that MADS-box gene duplications might have been crucial for increasing the adaptability of wheat to different environmental conditions as well as for fine-tuning quantitative traits by gene duplication. By thoroughly characterizing the entire complement of wheat MIKC-type MADS box genes, we provide the basis for the development of markers for future breeding efforts as well as for the identification of gene-editing targets to improve wheat performance. Further, we frequently observe possible neofunctionalization, a requirement for understanding the emergence of new traits during evolution.

Methods

Sequence search and annotation of MIKC-type MADS-box genes

Wheat coding sequence (CDSs) predictions, refmap comparison between IWGSC and The Genome Analysis Centre (TGAC) and functional annotations of both, high and low confidence (HC and LC) wheat genes, were downloaded from the IWGSC archive v1.0 (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.0/) [7] and the CSS - TGAC comparison was downloaded from https://opendata.earlham.ac.uk/opendata/data/Triticum_aestivum/TGAC/v1/annotation/ [74]. Functional annotations were filtered for PFAM identifiers of the MADS- and K-domain (PF00319 and PF01486), respectively [75]. A total of 439 sequences were identified (see a list of all gene IDs in Table S1). Of these, 188 sequences had a MADS-box and a K-box (181 HC plus 7 LC), 240 sequences (159 HC plus 71 LC) had only a MADS-box; and 21 had only a K- -box (16 HC plus 5 LC) (Table S1). Splice variants were excluded and only the first variant was kept for further analysis, with three exceptions (see Table S2).

All CDSs were translated into amino acid sequences and aligned with all MADS-domain protein sequences of rice (*Oryza sativa*, [21]) with MAFFT (L-INS-i algorithm) [76, 77] using only the MADS-domain of each sequence. Subsequently, a phylogeny was generated using IQTREE [78, 79]. This allowed distinguishing type I and type II (MIKC-type) MADS-domain proteins (Table S1). All type I MADS-box CDSs were excluded from subsequent analyses.

We evaluated the predicted gene structure of all genes, assuming a canonical M-I-K-C domain structure. In cases where either MADS- or K-box were absent from gene predictions, we compared the sequences with closely related rice and Arabidopsis genes, TGAC gene predictions and screened the surrounding genomic regions using the NCBI conserved domain

database [80-83]. In 17 cases gene prediction was repeated with FGENESH+ or FGENESH_C [84] using rice MADS-box genes of the same clade (Table S4). In one case (*TaAGL17.2A1*, *TraesCSU01G209900*), the TGAC CDS was used instead of IWGSC prediction, because it encoded for a canonical MIKC structure as compared to the IWGSC prediction, which comprised a MADS-box. This approach yielded 193 wheat MIKC-type MADS-box sequences.

In parallel, a BLAST search was carried out by which we identified 8 additional MIKC-type genes (Table S3) (https://urgi.versailles.inra.fr/blast_iwgsc/) [7, 85].

Altogether, a total of 201 wheat MIKC-type MADS-box genes were identified (Table 1, Table S2).

Maximum likelihood phylogeny of MIKC-type MADS-box genes

Based on the first phylogeny, MIKC-type sequences were sorted into the major grass MIKC-type subfamilies (Table S2) [40]. Afterwards, subfamily alignments of MIKC-type protein sequences were created using wheat, rice and Arabidopsis protein sequences [24, 42, 43] using MAFFT (E-INS-i algorithm) [76, 77]. Subfamily alignments were then merged using MAFFT (E-INS-i algorithm;) [76, 77]. The full-length alignment of all MIKC-type MADS-domain proteins was analyzed with Homo (Version 1.3) [86] to confirm that the sequences met the phylogenetic assumption of reversible conditions. Individual residues were subsequently masked with Alistat (Version 1.3) [87], leaving only sites with a completeness score (Cc, defined as Cc = Number of unambiguous characters in the column / number of sequences) above 0.5 (a total of 207 sites) (alignment in Supplementary Text 2).

Using the masked protein alignment, a phylogenetic tree was inferred under maximum likelihood with IQ-TREE [79]. The substitution model was calculated with ModelFinder (integrated in IQ-TREE; best-fit model: JTT+R5 chosen according to BIC) [78]. Consistency

439 of the phylogenetic estimate was evaluated with Ultra Fast bootstraps as well as SH-aLRT
440 test (1000 replicates each) [88-90]. The resulting treefile was visualized with Geneious
441 version 11.1 (https://www.geneious.com) (Figure 1) and FigTree version 1.4.4
442 (http://tree.bio.ed.ac.uk/software/figtree/) (Figure S2).
443
444 **Naming of MIKC-type MADS-box genes**
445 We suggest a consistent naming pattern for all MIKC-type wheat MADS-box genes, taking
446 into account their subfamily association, phylogenetic relationships as well as their
447 subgenome location (A, B or D). Each gene name starts with an abbreviation for the species
448 name *Triticum aestivum* (*Ta*), followed by the name of the respective Arabidopsis subfamily
449 (e.g. "*SEP1*" for *SEPALLATA1*-like genes, "*AGL6*" for *AGL6*-like genes or "*BS*" for *Bsister*
450 genes). The gene names include an A, B or D, indicating the subgenome they are located in,
451 e.g. *TaAGL6B*. If more than one triad of homoeologs was found in one subfamily, these were
452 distinguished by a "." and consecutive numbers, then followed by the subgenome (e.g.
453 *TaAGL12.1A and TaAGL12.2A*). If more than one copy of a gene was present in one
454 subgenome (inparalogs, e.g. due to tandem duplications or transposition), a number was
455 added after the letter that indicates the subgenome (prevalent in *FLC*, *SEP1*, *AGL17* and
456 *Bsister* subclades). Hence, the name of the gene with the ID TraesCS7B01G020900 is
457 *TaSEP1.5B1* since it is a *SEP1*-like gene, and more precisely one of two inparalogs of the B
458 genome (Figure 1, Table 1, all gene names listed in Table S2).
459
460 **Identification of homoeologs**
461 Homoeologous genes were identified by phylogeny (Figure 1, Figure S3A-D). In some cases,
462 where Ultra Fast bootstraps and SH-aLRT were not high enough to support a clade (above 90
463 and 75, respectively), synteny and previous classifications were considered [8]. 38 genes

464 belonging to the *FLC*-, *Bsister* and *AGL17*-subclades were excluded from the analysis,
465 because their homoeolog status could not definitely be determined (Table 1, Table 2, Figure
466 S3).
467
468 **Plant cultivation, RNA isolation and RT-PCR**
469 *Triticum turgidum* cv. Kronos (tetraploid) was used for RT-PCR analysis. Plants were grown
470 in a greenhouse with no additional lighting at an ambient temperature of 20 to 24 °C until
471 heading stage. RNA was extracted from the frozen flag leaf tip using QIAGEN® RNeasy
472 Mini Kit according to manufacturer's instructions. cDNA was generated with Superscript IV®
473 Reverse Transcriptase Kit using an oligo dT primer. RT-PCR was carried out with Thermo
474 Scientific™ Phusion High-Fidelity DNA Polymerase. Primer sequences are listed in
475 Table S5.
476
477 **Expression analysis of MIKC-type MADS-box genes using RNA-seq**
478 RNA-seq data of 193 wheat MADS-box genes was analyzed using
479 www.wheatexpression.com and http://bar.utoronto.ca/efp_wheat/ [8]. For the remaining 8
480 genes, which were identified by BLAST, no expression data was available. Developmental
481 stages depicted in Figure 5 refer to 70 tissues/time points from spring wheat cv. Azhurnaya
482 [8]. Expression levels were downloaded from www.wheatexpression.com as $\log_2$tpm and a
483 heatmap was generated with heatmapper (heatmapper.ca/expression) [91]. Clustering was
484 performed with heatmapper using centroid linkage with euclidean distance measure. All
485 genes, modules and tissues are listed in Tables S6 and S7.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

We are grateful to the School of Biology and Environmental Science at the University College Dublin for general support.

Authors' contributions

RM and SS developed the analysis approach. SS and AK analyzed the data with supervision from RM and LJ. SP performed the RT-PCR experiments. SS and RM wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful for the support by the International Wheat Genome Sequencing Consortium and early access to the genome data.

Figures

Figure 1. Maximum likelihood phylogeny of MIKC-type MADS-domain proteins from bread wheat, rice and Arabidopsis. A phylogenetic tree of MADS-domain proteins from bread wheat, cultivated rice and Arabidopsis was inferred using IQ-TREE [78, 79]. Wheat genes are coloured, rice and Arabidopsis genes are in black. The Arabidopsis genes *AGL15* and *AGL18* were included in the phylogeny; however, the *AGL15*-subfamily is lost in the grass lineage [40], therefore no rice or wheat gene clustered with them. *OsMADS32*-like genes are monocot specific [40], hence in this clade there is no Arabidopsis gene. SH-aLRT and Ultrafast bootstrap values are indicated on the branches in %, values equal to 100 are designated by a -. Accession numbers of all wheat genes can be found in Table S2, a version of the tree with untransformed branches can be found in Figure S2. The tree is unrooted, the MIKC* subclade was set as the outgroup.

Figure 2. Number and location of MIKC-type genes. The number of MADS-box genes identified per MIKC-type subfamily in Arabidopsis (A), rice (B) and wheat (C) [21, 24]. The ratio of MIKC-type gene numbers total and in all subfamilies is shown for wheat to rice (black) and wheat to Arabidopsis (grey) (D). In (D) the expected ratio (3:1) is indicated by a red line, asterisks mark significant deviation from expected value (χ^2 test, $p < 0.05$).

All MIKC-type MADS-box genes were mapped to their respective locus in the wheat genome in a circular diagram using shinyCircos [92] (E). Subgenomes are indicated by different shades of blue (outer track), chromosomal segments are indicated by shades of grey (inner track) [7]. Homoeologous genes were inferred by mainly by phylogeny (details see Methods) and linked with subfamily-specific colors (inside of circle).

Figure 3. Chromosomal distribution of wheat MIKC-type MADS-box genes. The proportion of truncated (black) and full-length (white) MIKC-type genes in every chromosome segment is shown in a bar diagram (A). A schematic overview of the chromosome indicates the different segments, R1 and R3 (dark grey), R2A and R2B (light grey) and C (white) (B). The location of all genes belonging to each MIKC-type subfamily is shown as percentage of chromosome length (C). Segments have been averaged over all chromosomes, segment lengths according to [7].

Figure 4. Distribution of MIKC-type genes to subtelomeric or central chromosome segments. The number of genes per MIKC-type subclade was plotted against the fraction of genes located in central chromosome segments (R2a, R2b and C) and subtelomeric segments (R1 and R3). Subclades are indicated next to data points. Data points for some subfamilies are identical (*AGL6* and *OsMADS32*; *SVP* and *AP1; PI*, *AP3* and *AGL12; SEP3* and *MIKC**).

Figure 5. Expression of MIKC-type MADS-box genes during wheat development. Expression analysis was done for all subfamilies using RNA-seq data [8]. A heatmap shows expression level of all genes in different subfamilies (columns) and stages/tissues (rows) (A). Genes and tissues are listed in Tables S6 and S7, respectively.

Genes were clustered into different modules according to their expression and mapped to subfamilies. Colors indicate different modules: shades of brown indicate ubiquitous expression, shades of green indicate expression in the root, shades of blue indicate expression in reproductive organs and grey indicates low or no expression during development (B). The digits inside the circles indicate the number of genes in the respective module. The clustering heatmap with all modules can be found in Figure S4. A schematic representation of a wheat plant depicts colors indicating different expression modules (C).

Figure 6. Expression of *Bsister* and *AGL17*-like genes in response to stress conditions. RNA-seq data was analyzed using wheatexpression.com [47]. *Bsister* gene expression in different tissues [8] and in response to infection with Fusarium head blight [93] (A). Expression of wheat *OsMADS31*-like genes (*TaBS.2*) *OsMADS29*-like genes (*TaBS.1*) and *OsMADS30*-like genes (*TaBS.4*, *TaBS.5A4*, *TaBS.6B1*) in root, stem and grain [8] as well as spikelets during mock inoculation and infection with *Fusarium graminearum* after different time points is shown as a heatmap.

AGL17-like gene (1-4) expression in roots and seedling leaves under control conditions (c) and after 1 and 6 hours of drought (d), heat (h) and combined drought and heat stress (dh) [94] (B).

Expression of *TaAGL17.3D* within different root tissues [8] and in leaves under control conditions and after infection with stripe rust (*Puccinia striiformis*) [95] is shown as a bar diagram. SEM is indicated as error bars. Data is shown as log_2tpm (A, B) or as tpm (C).

Tables

Table 1. Subfamilies, names and numbers of wheat MIKC-type MADS-box genes.

A complete list of all wheat MIKC-type genes can be found in Table S2.

[1] asterisk indicates that the number of homeologs could not be determined due to insufficient phylogenetic resolution, genes were not included into homoeolog count (Table 2). Details see Figure S3.

[2] gene encodes for MADS but not K domain

[3] gene encodes for K but not MADS domain

[4] number in parentheses indicates a different chromosome for one of the genes; x, genes located on more than two different chromosomes

[5] parentheses indicate truncated genes, encoding for either MADS or K domain

Table 2. Groups of homoeologous MIKC-type MADS-box genes in wheat.

[1] according to [7]

[2] % calculated with 201 genes

[3] % calculated with 164 genes

[4] either n:1:n or 0:1:n

[5] see Table 1 and Figure S3

Supplementary Material

Supplementary Figures

Figure S1. RT-PCR of two FLC-like wheat genes. Schematic representation of two FLC genes and primer binding sites (red arrowheads) (A). Intron size is indicated by dotted line, boxes indicate exons, lines indicates introns. Gel images of PCR products and expected fragment sizes (B). The sequence of the PCR products was verified using Sanger Sequencing.

Figure S2. Maximum likelihood phylogeny of MIKC-type MADS-domain proteins from bread wheat, rice and Arabidopsis, branches not transformed. SH-aLRT and Ultrafast bootstrap values are indicated on the branches in %, values equal to 100 are designated by a -. The tree is unrooted, the MIKC* subclade was set as the outgroup.

Figure S3. Maximum likelihood phylogenies of four different MIKC-type subfamilies. Subphylogenies of *SEPALLATA*- (A), *Bsister*- (B), *AGL17*- (C) and *FLC*- (D) like genes from Arabidopsis, rice and wheat were generated using protein alignments and IQ-TREE [78, 79]. *AGL6*- (A), *AP3/PI*-(B) and *OsMADS32*- (C, D)-like genes were used as an outgroup. Red and blue triangles indicate truncated genes encoding only for MADS- and K-domain, respectively. Black circle indicates gene encoding for a MADS- and SRP54-domain (PF00448). Groups excluded from homoeolog analysis are indicated by asterisks.

Figure S4. Cluster expression analysis of wheat MIKC-type MADS-box genes during developmental stages. Expression analysis was done for all wheat MIKC-type genes and tissue using RNA-seq data from wheatexpression.com [8, 47]. A heatmap shows expression levels of all genes in different subfamilies (rows) and stages/tissues (columns). Genes and tissues are listed in Tables S6 and S7, respectively. Heatmap and clustering was done with heatmapper.com and a cladogram showing the result of the clustering is shown on the left. Modules 1 to 17 were assigned and borders indicated by red lines (right). Values represent $\log_2$tpm.

Figure S5. Expression of wheat *Bsister* genes during Fusarium head blight infection. RNA-seq data was analyzed using wheatexpression.com [8, 47]. *Bsister* gene expression in the donor plant and 4 near-isogenic lines (NIL) under control conditions (c) and *Fusarium graminearum* inoculation (s) (30 and 50 h) is shown as a heatmap [96]. Values represent $\log_2$tpm.

Supplementary Tables

Table S1. List of all sequences identified by PFAM domain (type I and type II).

Table S2. List of all MIKC-type MADS-box genes in wheat.

[1] (+) indicates FLC-specific K-box (not PFAM PF01486); (-) TGAC version of this gene does have PF01486, IWGSC does not have PF01486

[2] Protein length and molecular weight were determined with the protein molecular weight function on bioinformatics.org using conceptually translated CDS.

[3] According to [97]

[4] According to [8]

[5] TGAC identifiers were inferred from refmap comparison between IWGSC and TGAC [7] (details see Methods)

[6] identifiers were inferred from refmap comparison between CSS - TGAC [74] (details see Methods)

Table S3. Average distance to the nearest downstream MIKC-type gene in kbp in different chromosomal segments. Distances between start points of neighboring MIKC-type genes have been calculated and averaged over chromosomal segments.

Table S4. List of all MIKC-type MADS-box genes predicted during this study. Table includes gene name, IWGSC identifier (if available) and chromosome location, PFAM domain(s), predicted coding sequence and prediction algorithm and guiding sequence FGENESH+ and FGENESH_C predictions are guided by homolog protein or cDNA sequences, respectively.

Table S5. Primer sequences for RT-PCR.

Table S6. Expression modules and gene numbers for Figure 5A and S3.

Table S7. Tissues for expression analysis for Figure 5A and S3.

References

1. Shiferaw B, Smale M, Braun H-J, Duveiller E, Reynolds M, Muricho G: **Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security.** *Food Security* 2013, **5:**291-317.
2. Veraverbeke WS, Delcour JA: **Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality.** *Critical reviews in food science and nutrition* 2002, **42:**179-208.
3. Allaby RG, Stevens C, Lucas L, Maeda O, Fuller DQ: **Geographic mosaics and changing rates of cereal domestication.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017, **372:**20160429.
4. Smith BD, Nesbitt M: *The emergence of agriculture.* Scientific American Library New York; 1995.
5. Shewry PR: **Wheat.** *Journal of experimental botany* 2009, **60:**1537-1553.
6. Borrill P, Adamski N, Uauy C: **Genomics as the key to unlocking the polyploid potential of wheat.** *New Phytol* 2015, **208:**1008-1022.
7. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J: **Shifting the limits in wheat research and breeding using a fully annotated reference genome.** *Science* 2018, **361:**eaar7191.
8. Ramírez-González R, Borrill P, Lang D, Harrington S, Brinton J, Venturini L, Davey M, Jacobs J, Van Ex F, Pasha A: **The transcriptional landscape of polyploid wheat.** *Science* 2018, **361:**eaar6089.
9. Martínez-Ainsworth NE, Tenaillon MI: **Superheroes and masterminds of plant domestication.** *Comptes Rendus Biologies* 2016, **339:**268-273.
10. Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al: **Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes.** *Science* 2000, **290:**2105-2110.
11. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Pouplana LR, Martinez-Castilla L, Yanofsky MF: **An ancestral MADS-box gene duplication occurred before the divergence of plants and animals.** *PNAS* 2000, **97:**5328-5333.
12. Schilling S, Pan S, Kennedy A, Melzer R: **MADS-box genes and crop domestication: the jack of all traits.** *J Exp Bot* 2018, **69:**1447-1469.
13. Kaufmann K, Melzer R, Theissen G: **MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants.** *Gene* 2005, **347:**183-198.
14. Riechmann JL, Wang M, Meyerowitz EM: **DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS.** *Nucleic Acids Research* 1996, **24:**3134-3141.
15. Immink RG, Gadella TW, Ferrario S, Busscher M, Angenent GC: **Analysis of MADS box protein–protein interactions in living plant cells.** *Proceedings of the National Academy of Sciences* 2002, **99:**2416-2421.
16. Melzer R, Verelst W, Theißen G: **The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro.** *Nucleic Acids Research* 2008, **37:**144-157.
17. Yang Y, Jack T: **Defining subdomains of the K domain important for protein–protein interactions of plant MADS proteins.** *Plant molecular biology* 2004, **55:**45-59.

18. Honma T, Goto K: **Complexes of MADS-box proteins are sufficient to convert leaves into floral organs.** *Nature* 2001, **409:**525-529.
19. Gramzow L, Theissen G: **A hitchhiker's guide to the MADS world of plants.** *Genome biology* 2010, **11:**214.
20. Jia J, Zhao P, Cheng L, Yuan G, Yang W, Liu S, Chen S, Qi D, Liu G, Li X: **MADS-box family genes in sheepgrass and their involvement in abiotic stress responses.** *BMC plant biology* 2018, **18:**42.
21. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK: **MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress.** *BMC Genomics* 2007, **8**.
22. Wei M, Wang Y, Pan R, Li W: **Genome-Wide Identification and Characterization of MADS-box Family Genes Related to Floral Organ Development and Stress Resistance in Hevea brasiliensis Müll. Arg.** *Forests* 2018, **9:**304.
23. Boden SA, Østergaard L: **How can developmental biology help feed a growing population?** *Development* 2019, **146:**dev172965.
24. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, et al: **Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world.** *Plant Cell* 2003, **15:**1538-1551.
25. Wei B, Zhang R-Z, Guo J-J, Liu D-M, Li A-L, Fan R-C, Mao L, Zhang X-Q: **Genome-wide analysis of the MADS-box gene family in Brachypodium distachyon.** *PLoS One* 2014, **9:**e84781.
26. Duan W, Song X, Liu T, Huang Z, Ren J, Hou X, Li Y: **Genome-wide analysis of the MADS-box gene family in Brassica rapa (Chinese cabbage).** *MGG* 2015, **290:**239-255.
27. Nardeli SM, Artico S, Aoyagi GM, de Moura SM, da Franca Silva T, Grossi-de-Sa MF, Romanel E, Alves-Ferreira M: **Genome-wide analysis of the MADS-box gene family in polyploid cotton (Gossypium hirsutum) and in its diploid parental species (Gossypium arboreum and Gossypium raimondii).** *Plant Physiol Biochem* 2018, **127:**169-184.
28. Liu J, Zhang J, Zhang J, Miao H, Wang J, Gao P, Hu W, Jia C, Wang Z, Xu B: **Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening.** *Scientific reports* 2017, **7:**3467.
29. Li C, Lin H, Chen A, Lau M, Jernstedt J, Dubcovsky J: **Wheat VRN1 and FUL2 play critical and redundant roles in spikelet meristem identity and spike determinacy.** *bioRxiv* 2019**:**510388.
30. Murai K, Miyamae M, Kato H, Takumi S, Ogihara Y: **WAP1, a wheat APETALA1 homolog, plays a central role in the phase transition from vegetative to reproductive growth.** *Plant Cell Physiol* 2003, **44:**1255-1265.
31. Harris F, Eagles H, Virgona J, Martin P, Condon J, Angus J: **Effect of VRN1 and PPD1 genes on anthesis date and wheat growth.** *Crop and Pasture Science* 2017, **68:**195-201.
32. Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES: **MADS box genes control vernalization-induced flowering in cereals.** *PNAS* 2003, **100:**13099-13104.
33. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J: **Positional cloning of the wheat vernalization gene VRN1.** *PNAS* 2003, **100:**6263-6268.

34. 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 Mol Biol* 2009, **71:**1-14.
35. Yang W, Lou X, Li J, Pu M, Mirbahar AA, Liu D, Sun J, Zhan K, He L, Zhang A: **Cloning and functional analysis of MADS-box genes, TaAG-A and TaAG-B, from a wheat K-type cytoplasmic male sterile line.** *Frontiers in plant science* 2017, **8:**1081.
36. Meguro A, Takumi S, Ogihara Y, Murai K: **WAG, a wheat AGAMOUS homolog, is associated with development of pistil-like stamens in alloplasmic wheats.** *Sexual Plant Reproduction* 2003, **15:**221-230.
37. Zhao XY, Cheng ZJ, Zhang XS: **Overexpression of TaMADS1, a SEPALLATA-like gene in wheat, causes early flowering and the abnormal development of floral organs in Arabidopsis.** *Planta* 2006, **223**.
38. Ferrario S, Busscher J, Franken J, Gerats T, Vandenbussche M, Angenent GC, Immink RG: **Ectopic expression of the petunia MADS box gene UNSHAVEN accelerates flowering and confers leaf-like characteristics to floral organs in a dominant-negative manner.** *The Plant Cell* 2004, **16:**1490-1505.
39. Seo PJ, Hong SY, Ryu JY, Jeong EY, Kim SG, Baldwin IT, Park CM: **Targeted inactivation of transcription factors by overexpression of their truncated forms in plants.** *The Plant Journal* 2012, **72:**162-172.
40. Gramzow L, Theissen G: **Phylogenomics reveals surprising sets of essential and dispensable clades of MIKC -group MADS-box genes in flowering plants.** *J Exp Zool B* 2015, **324:**353-362.
41. Gramzow L, Theißen G: **Phylogenomics of MADS-box genes in plants - Two opposing life styles in one gene family.** *Biology (Basel)* 2013, **2**.
42. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S: **MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress.** *Bmc Genomics* 2007, **8**.
43. Verelst W, Saedler H, Münster T: **MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters.** *Plant Physiol* 2007, **143:**447-460.
44. Gramzow L, Theißen G: **Phylogenomics of MADS-box genes in plants—two opposing life styles in one gene family.** *Biology* 2013, **2:**1150-1164.
45. Vining KJ, Romanel E, Jones RC, Klocko A, Alves‐Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M: **The floral transcriptome of Eucalyptus grandis.** *New Phytologist* 2015, **206:**1406-1422.
46. Chen NW, Thareau V, Ribeiro T, Magdelenat G, Ashfield T, Innes RW, Pedrosa-Harand A, Geffroy V: **Common bean subtelomeres are hot spots of recombination and favor resistance gene evolution.** *Frontiers in plant science* 2018, **9**.
47. Borrill P, Ramirez-Gonzalez R, Uauy C: **expVIP: a customizable RNA-seq data analysis and visualization platform.** *Plant Physiol* 2016, **170:**2172-2186.
48. Yang X, Wu F, Lin X, Du X, Chong K, Gramzow L, Schilling S, Becker A, Theißen G, Meng Z: **Live and let die-the Bsister MADS-box gene OsMADS29 controls the degeneration of cells in maternal tissues during seed development of rice (Oryza sativa).** *PloS one* 2012, **7:**e51435.

49. Liu Y, Cui S, Wu F, Yan S, Lin X, Du X, Chong K, Schilling S, Theißen G, Meng Z: **Functional conservation of MIKC*-Type MADS box genes in Arabidopsis and rice pollen maturation.** *The Plant Cell* 2013, **25:**1288-1303.
50. Schilling S, Gramzow L, Lobbes D, Kirbis A, Weilandt L, Hoffmeier A, Junker A, Weigelt-Fischer K, Klukas C, Wu F, et al: **Non-canonical structure, function and phylogeny of the Bsister MADS-box gene OsMADS30 of rice (Oryza sativa).** *Plant J* 2015**:**in press.
51. Becker A, Theißen G: **The major clades of MADS-box genes and their role in the development and evolution of flowering plants.** *Molecular phylogenetics and evolution* 2003, **29:**464-489.
52. Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L: **The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat.** *The Plant Cell* 2002, **14:**2463-2479.
53. Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L: **The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development.** *The Plant Journal* 2012, **70:**409-420.
54. Puig J, Meynard D, Khong GN, Pauluzzi G, Guiderdoni E, Gantet P: **Analysis of the expression of the AGL17-like clade of MADS-box transcription factors in rice.** *Gene Expression Patterns* 2013, **13:**160-170.
55. Schilling S, Pan S, Kennedy A, Melzer R: **MADS-box genes and crop domestication: the jack of all traits.** Oxford University Press UK; 2018.
56. Theißen G, Melzer R, Rümpler F: **MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution.** *Development* 2016, **143:**3259-3271.
57. Immink RG, Kaufmann K, Angenent GC: **The 'ABC' of MADS domain protein behaviour and interactions.** *Semin Cell Dev Biol* 2010, **21:**87-93.
58. Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, Borst JW, Angenent GC: **SEPALLATA3: the 'glue' for MADS box transcription factor complex formation.** *Genome Biol* 2009, **10:**R24.
59. Birchler JA, Riddle NC, Auger DL, Veitia RA: **Dosage balance in gene regulation: biological implications.** *Trends in Genetics* 2005, **21:**219-226.
60. Borrill P, Harrington SA, Uauy C: **Genome-wide sequence and expression analysis of the NAC transcription factor family in polyploid wheat.** *G3: Genes, Genomes, Genetics* 2017, **7:**3019-3029.
61. Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH: **Role of SVP in the control of flowering time by ambient temperature in Arabidopsis.** *Genes & development* 2007, **21:**397-402.
62. Moon J, Suh SS, Lee H, Choi KR, Hong CB, Paek NC, Kim SG, Lee I: **The SOC1 MADS‐box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis.** *The Plant Journal* 2003, **35:**613-623.
63. Theißen G, Rümpler F, Gramzow L: **Array of MADS-Box Genes: Facilitator for Rapid Adaptation?** *Trends in plant science* 2018.
64. Yin L-L, Xue H-W: **The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development.** *The Plant Cell* 2012, **24:**1049-1065.

65. Dubin MJ, Scheid OM, Becker C: **Transposons: a blessing curse.** *Current opinion in plant biology* 2018, **42:**23-29.
66. Guo S, Xu Y, Liu H, Mao Z, Zhang C, Ma Y, Zhang Q, Meng Z, Chong K: **The interaction between OsMADS57 and OsTB1 modulates rice tillering via DWARF14.** *Nature communications* 2013, **4:**1566.
67. Xu N, Chu Y, Chen H, Li X, Wu Q, Jin L, Wang G, Huang J: **Rice transcription factor OsMADS25 modulates root growth and confers salinity tolerance via the ABA–mediated regulatory pathway and ROS scavenging.** *PLOS Genetics* 2018, **14:**e1007662.
68. Kutter C, Schöb H, Stadler M, Meins F, Si-Ammour A: **MicroRNA-mediated regulation of stomatal development in Arabidopsis.** *Plant Cell* 2007, **19**.
69. Hu J-Y, Zhou Y, He F, Dong X, Liu L-Y, Coupland G: **miR824-Regulated AGAMOUS-LIKE16 Contributes to Flowering Time Repression in Arabidopsis.** *Plant Cell* 2014, **26**.
70. Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, Choulet F: **Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B.** *Genome Biol* 2015, **16:**188.
71. Shitsukawa N, Tahira C, Kassai K-i, Hirabayashi C, Shimizu T, Takumi S, Mochida K, Kawaura K, Ogihara Y, Murai K: **Genetic and epigenetic alteration among three homoeologous genes of a class E MADS box gene in hexaploid wheat.** *The Plant Cell* 2007, **19:**1723-1737.
72. Jones S: **An overview of the basic helix-loop-helix proteins.** *Genome biology* 2004, **5:**226.
73. Ledent V, Vervoort M: **The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis.** *Genome research* 2001, **11:**754-770.
74. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, et al: **An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations.** *Genome Research* 2017, **27:**885-896.
75. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A: **The Pfam protein families database in 2019.** *Nucleic Acids Res* 2018, **47:**D427-D432.
76. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: Improvements in performance and usability.** *Molecular Biology and Evolution* 2013, **30:**772-780.
77. Katoh K, Rozewicki J, Yamada KD: **MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.** *Briefings in Bioinformatics* 2017.
78. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS: **ModelFinder: fast model selection for accurate phylogenetic estimates.** *Nature Methods* 2017, **14:**587.
79. 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:**268-274.

80. Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004, **32:**W327-W331.
81. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic Acids Res* 2010, **39:**D225-D229.
82. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI: **CDD: NCBI's conserved domain database.** *Nucleic Acids Res* 2014, **43:**D222-D226.
83. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR: **CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.** *Nucleic Acids Res* 2016, **45:**D200-D203.
84. Solovyev V, Kosarev P, Seledsov I, Vorobyev D: **Automatic annotation of eukaryotic genes, pseudogenes and promoters.** *Genome Biol* 2006, **7:**S10.
85. Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey C: **Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data.** *Genome Biol* 2018, **19:**111.
86. Jermiin L: **Homo version 1.3.** CSIRO: CSIRO; 2017.
87. Wong TKK, Subha; Meusemann, Karen; Yeates, David; Misof, Bernhard; Jermiin, Lars: **AliStat version 1.3. v1. CSIRO.** In *Software Collection*; 2014.
88. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS: **UFBoot2: Improving the Ultrafast Bootstrap Approximation.** *Molecular Biology and Evolution* 2018, **35:**518-522.
89. Minh BQ, Nguyen MAT, von Haeseler A: **Ultrafast approximation for phylogenetic bootstrap.** *Molecular biology and evolution* 2013, **30:**1188-1195.
90. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordjik W, Gascuel O: **New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.** *Syst Biol* 2010, **59**.
91. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS: **Heatmapper: web-enabled heat mapping for all.** *Nucleic Acids Res* 2016, **44:**W147-W153.
92. Yu Y, Ouyang Y, Yao W: **shinyCircos: an R/Shiny application for interactive creation of Circos plot.** *Bioinformatics* 2017, **34:**1229-1231.
93. Schweiger W, Steiner B, Vautrin S, Nussbaumer T, Siegwart G, Zamini M, Jungreithmeier F, Gratl V, Lemmens M, Mayer K: **Suppressed recombination and unique candidate genes in the divergent haplotype encoding Fhb1, a major Fusarium head blight resistance locus in wheat.** *Theoretical and Applied Genetics* 2016, **129:**1607-1623.
94. Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, Sun Q: **Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (Triticum aestivum L.).** *BMC plant biology* 2015, **15:**152.
95. Dobon A, Bunting DC, Cabrera-Quio LE, Uauy C, Saunders DG: **The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression.** *BMC genomics* 2016, **17:**380.

96. Kugler KG, Siegwart G, Nussbaumer T, Ametz C, Spannagl M, Steiner B, Lemmens M, Mayer KF, Buerstmayr H, Schweiger W: **Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (Triticum aestivum L.).** *BMC genomics* 2013, **14:**728.

97. Pingault L, Choulet F, Alberti A, Glover N, Wincker P, Feuillet C, Paux E: **Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome.** *Genome biology* 2015, **16:**29.

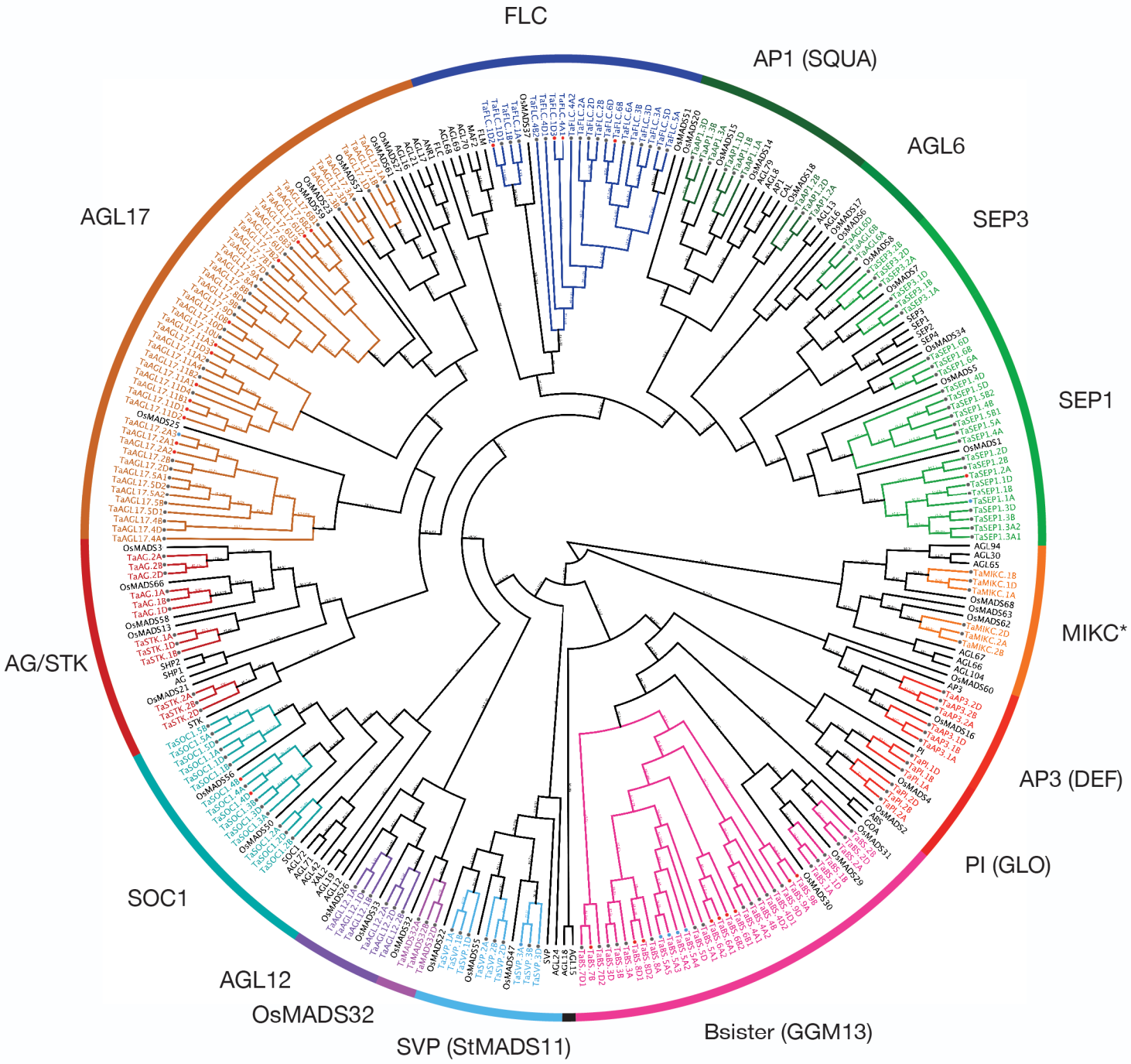

Figure 1. Maximum likelihood phylogeny of MIKC-type MADS-domain proteins from bread wheat, rice and Arabidopsis. A phylogenetic tree of MADS-domain proteins from bread wheat, cultivated rice and Arabidopsis was inferred using IQ-TREE [78, 79]. Wheat genes are coloured, rice and Arabidopsis genes are in black. The Arabidopsis genes *AGL15* and *AGL18* were included in the phylogeny; however, the *AGL15*-subfamily is lost in the grass lineage [40], therefore no rice or wheat gene clustered with them. *OsMADS32*-like genes are monocot specific [40], hence in this clade there is no Arabidopsis gene. SH-aLRT and Ultrafast bootstrap values are indicated on the branches in %, values equal to 100 are designated by a -. Accession numbers of all wheat genes can be found in Table S2, a version of the tree with untransformed branches can be found in Figure S2. The tree is unrooted, the MIKC* subclade was set as the outgroup.

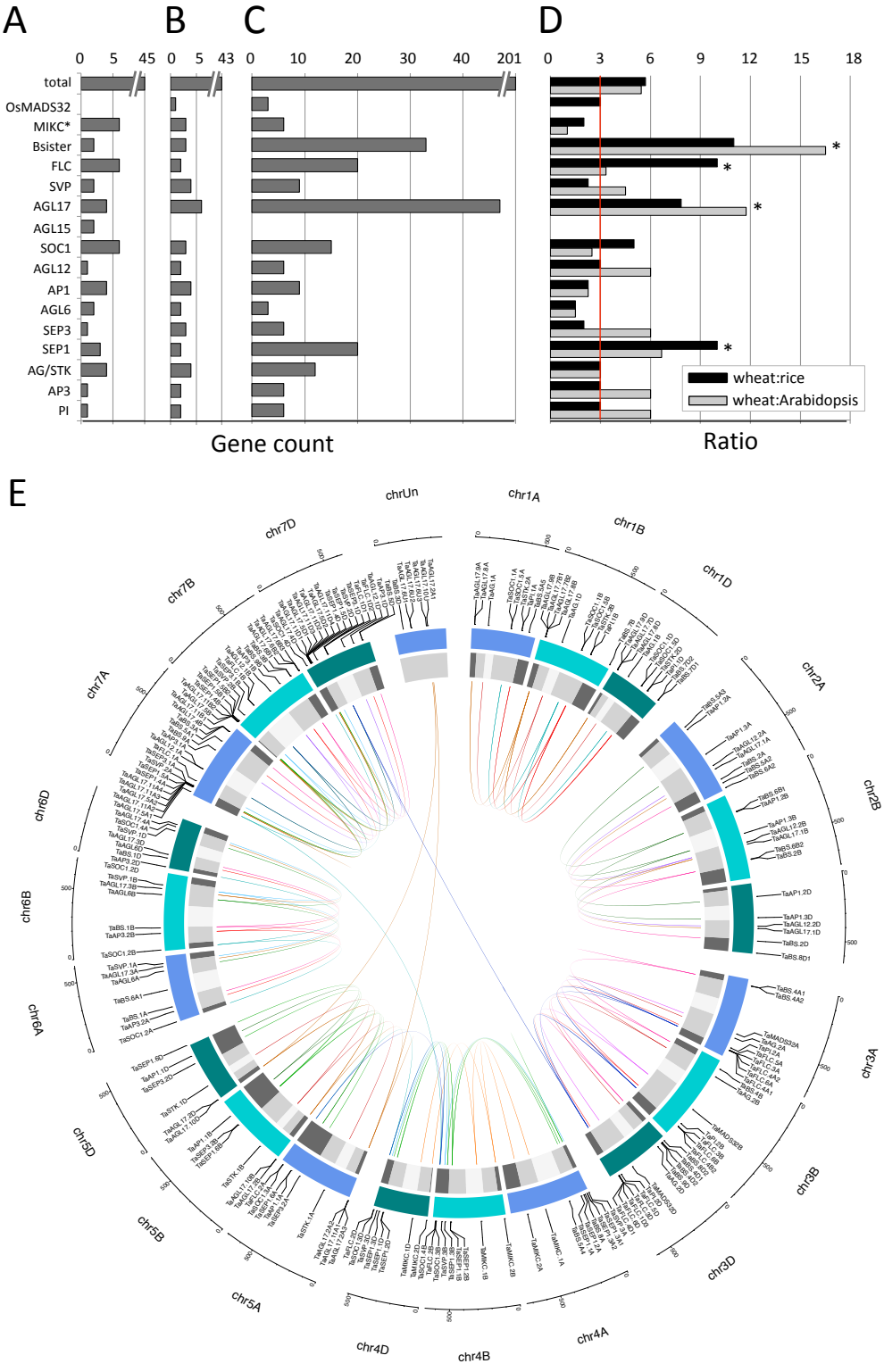

Figure 2. Number and location of MIKC-type genes. The number of MADS-box genes identified per MIKC-type subfamily in Arabidopsis (A), rice (B) and wheat (C) [21, 24]. The ratio of MIKC-type gene numbers total and in all subfamilies is shown for wheat to rice (black) and wheat to Arabidopsis (grey) (D). In (D) the expected ratio (3:1) is indicated by a red line, asterisks mark significant deviation from expected value (χ^2 test, $p < 0.05$).

All MIKC-type MADS-box genes were mapped to their respective locus in the wheat genome in a circular diagram using shinyCircos [92] (E). Subgenomes are indicated by different shades of blue (outer track), chromosomal segments are indicated by shades of grey (inner track) [7]. Homoeologous genes were inferred by mainly by phylogeny (details see Methods) and linked with subfamily-specific colors (inside of circle).

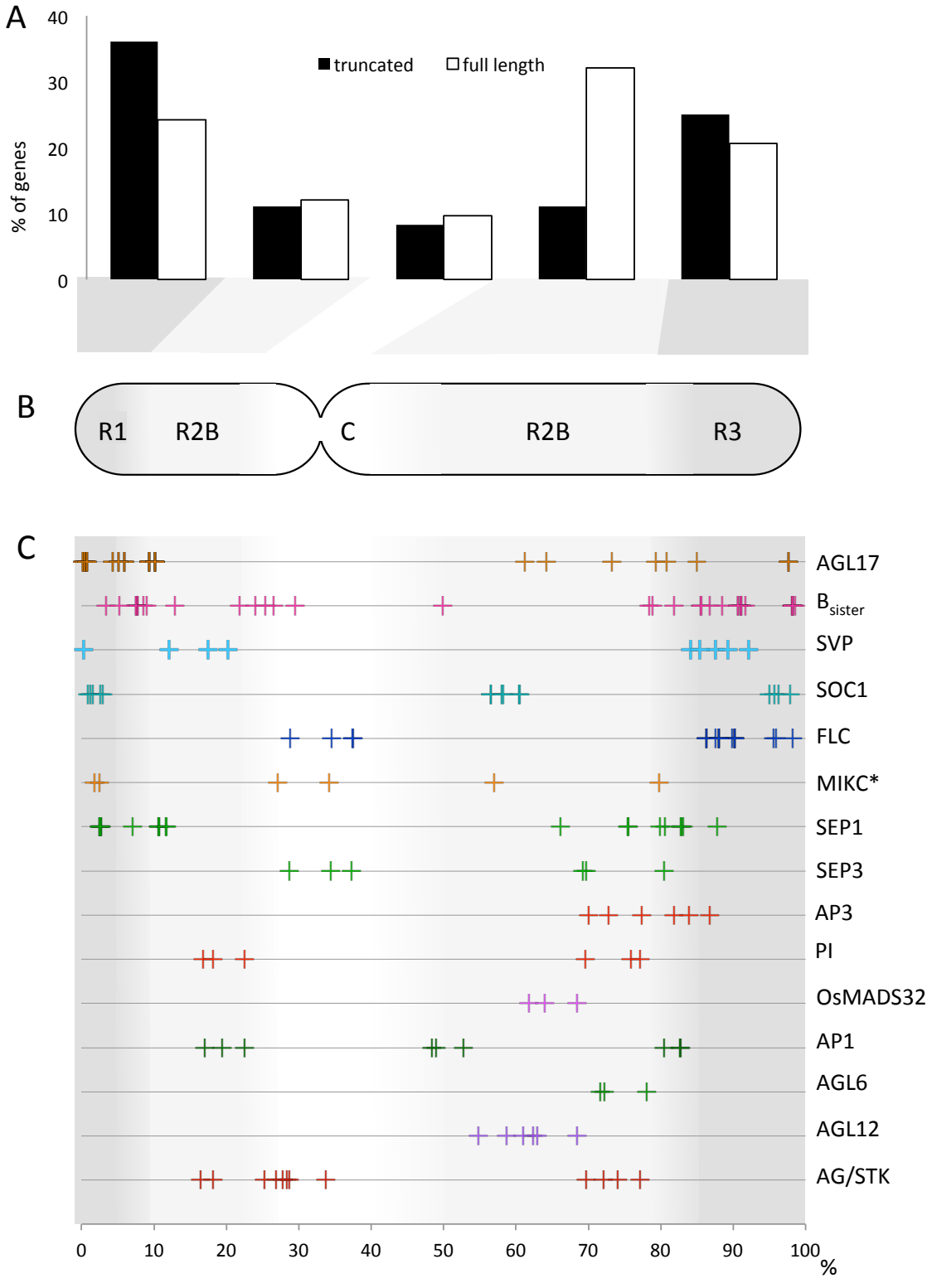

Figure 3. Chromosomal distribution of wheat MIKC-type MADS-box genes. The proportion of truncated (black) and full-length (white) MIKC-type genes in every chromosome segment is shown in a bar diagram (A). A schematic overview of the chromosome indicates the different segments, R1 and R3 (dark grey), R2A and R2B (light grey) and C (white) (B). The location of all genes belonging to each MIKC-type subfamily is shown as percentage of chromosome length (C). Segments have been averaged over all chromosomes, segment lengths according to [7].

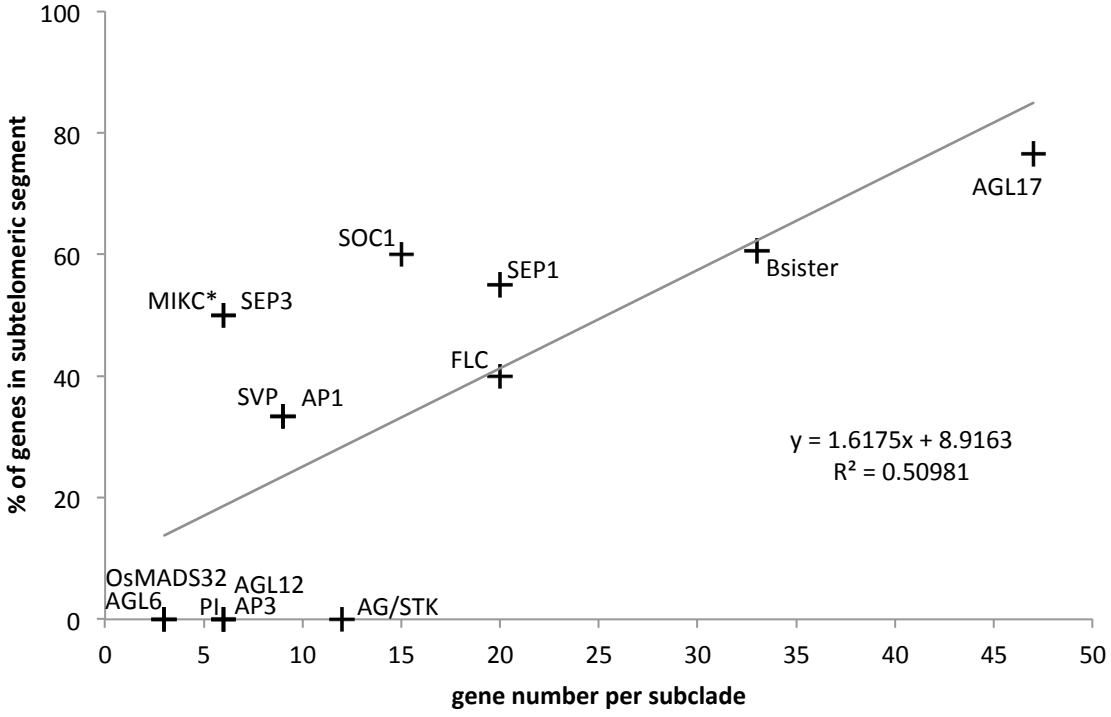

Figure 4. Distribution of MIKC-type genes to subtelomeric or central chromosome segments. The number of genes per MIKC-type subclade was plotted against the fraction of genes located in central chromosome segments (R2a, R2b and C) and subtelomeric segments (R1 and R3). Subclades are indicated next to data points. Data points for some subfamilies are identical (*AGL6* and *OsMADS32*; *SVP* and *AP1; PI*, *AP3* and *AGL12; SEP3* and *MIKC**).

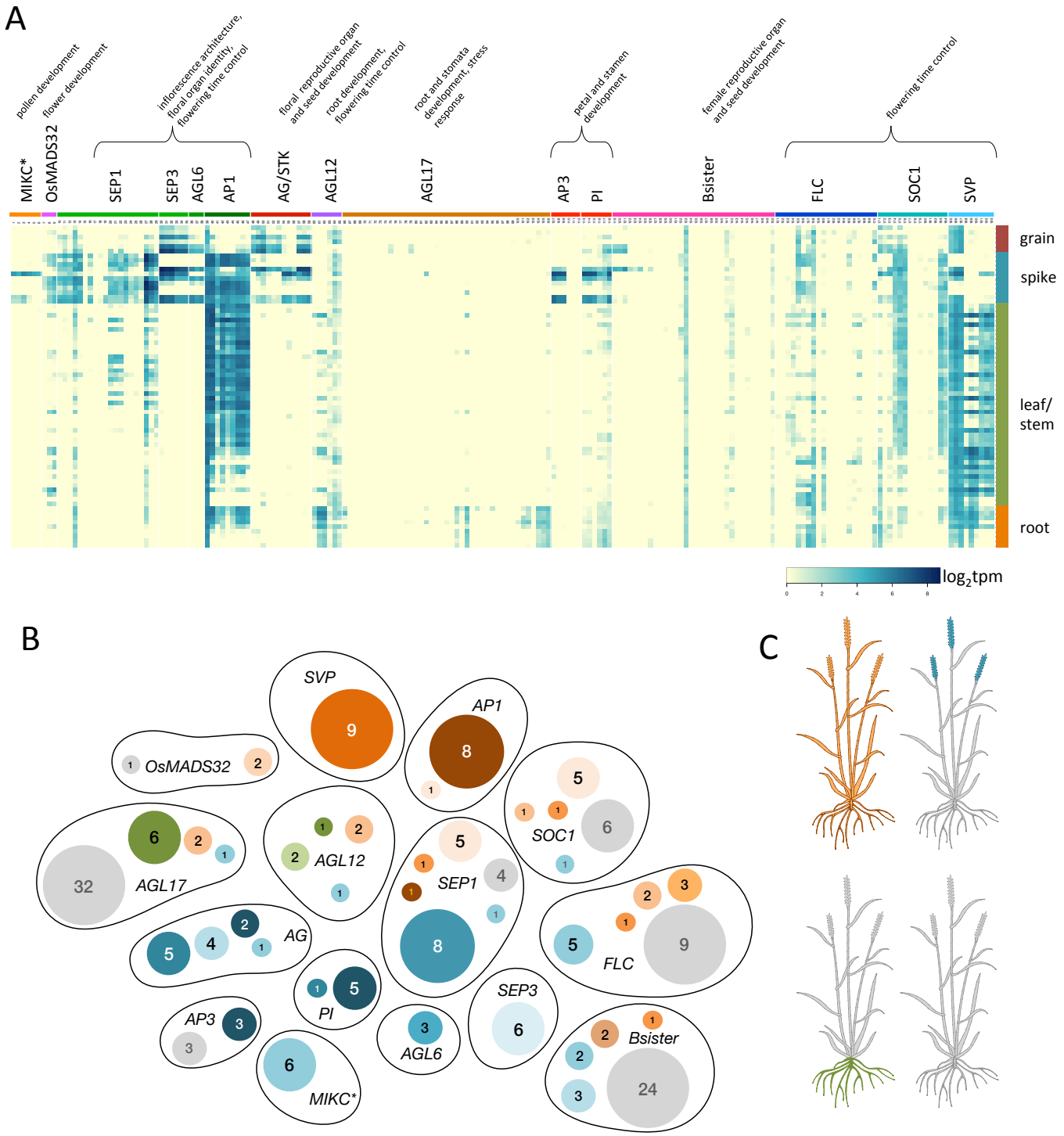

Figure 5. Expression of MIKC-type MADS-box genes during wheat development. Expression analysis was done for all subfamilies using RNA-seq data [8]. A heatmap shows expression level of all genes in different subfamilies (columns) and stages/tissues (rows) (A). Genes and tissues are listed in Tables S6 and S7, respectively.

Genes were clustered into different modules according to their expression and mapped to subfamilies. Colors indicate different modules: shades of brown indicate ubiquitous expression, shades of green indicate expression in the root, shades of blue indicate expression in reproductive organs and grey indicates low or no expression during development (B). The digits inside the circles indicate the number of genes in the respective module. The clustering heatmap with all modules can be found in Figure S4. A schematic representation of a wheat plant depicts colors indicating different expression modules (C).

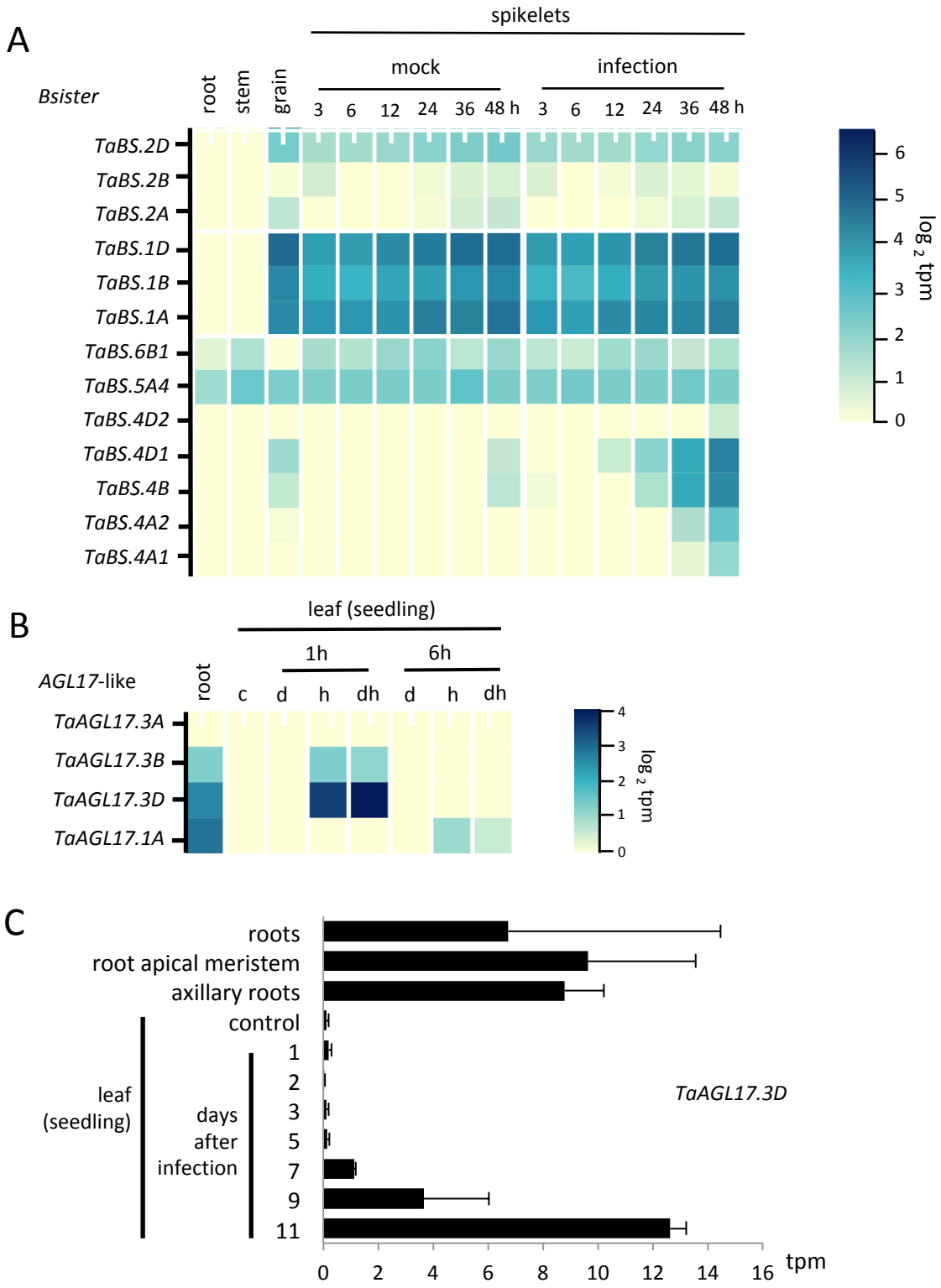

Figure 6. Expression of *Bsister* and *AGL17*-like genes in response to stress conditions. RNA-seq data was analyzed using wheatexpression.com [47]. *Bsister* gene expression in different tissues [8] and in response to infection with Fusarium head blight [93] (A). Expression of wheat *OsMADS31*-like genes (*TaBS.2*) *OsMADS29*-like genes (*TaBS.1*) and *OsMADS30*-like genes (*TaBS.4*, *TaBS.5A4*, *TaBS.6B1*) in root, stem and grain [8] as well as spikelets during mock inoculation and infection with *Fusarium graminearum* after different time points is shown as a heatmap.
AGL17-like gene (1-4) expression in roots and seedling leaves under control conditions (c) and after 1 and 6 hours of drought (d), heat (h) and combined drought and heat stress (dh) [94] (B).
Expression of *TaAGL17.3D* within different root tissues [8] and in leaves under control conditions and after infection with stripe rust (*Puccinia striiformis*) [95] is shown as a bar diagram. SEM is indicated as error bars. Data is shown as $\log_2$tpm (A, B) or as tpm (C).

Table 1. Subfamilies, names and numbers of wheat MIKC-type MADS-box genes.
A complete list of all wheat MIKC-type genes can be found in Table S2.

MIKC-type subclade	homoeolog names	x[1]	rice homolog	gene number total	full length	M[2]	K[3]	gene location Chr[4]	Genomes[5]	Named by	Alternative gene names from previous publications[6]	Reference Alt Name
MIKC*	TaMIKC*1		OsMADS68	**3**	3	0	0	**4**	**ABD**	this study	-	-
	TaMIKC*2		OsMADS63/OsMADS62	**3**	3	0	0	**4**	**ABD**	this study	-	-
OsMADS32	TaMADS32		OsMADS32	**3**	3	0	0	**3**	**ABD**	this study	TaWM16	Paolacci et al. 2007
AGL12	TaAGL12.1		OsMADS26	**3**	3	0	0	**7**	**ABD**	Paolacci et al. 2007	TaAGL32	Zhao T et al. 2006
	TaAGL12.2		OsMADS33	**3**	3	0	0	**2**	**ABD**	this study	-	-
AP1 (SQUA)	TaAP1.1		OsMADS14	**3**	3	0	0	**5**	**ABD**	Paolacci et al. 2007	VRN1; TaVRT-1	Yan et al. 2003, Fu et al. 2005; Danyluk 2003
	TaAP1.2		OsMADS18/OsMADS20	**3**	3	0	0	**2**	**ABD**	Paolacci et al. 2007	-	-
	TaAP1.3		OsMADS15	**3**	3	0	0	**2**	**ABD**	Paolacci et al. 2007	FUL2	Li et al. 2019
SVP (StMADS11)	TaSVP.1		OsMADS22	**3**	3	0	0	**6**	**ABD**	Paolacci et al. 2007	-	-
	TaSVP.2		OsMADS55	**3**	3	0	0	**7**	**ABD**	Paolacci et al. 2007	TaVRT-2	Kane et al. 2005
	TaSVP.3		OsMADS47	**3**	2	0	1	**4**	**(A)BD**	Paolacci et al. 2007	-	-
SEP1	TaSEP1.1			**3**	2	0	1	**4**	**(A)BD**	this study	TaSEP-1	Paolacci et al. 2007
	TaSEP1.2		OsMADS1	**3**	2	1	0	**4**	**(A)BD**	this study	WLHS1; TaSEP-2	Shitsukawa et al. 2007; Paolacci et al. 2007
	TaSEP1.3			**4**	4	0	0	**4**	**AABD**	this study	-	-
	TaSEP1.4		OsMADS5	**3**	3	0	0	**7**	**ABD**	this study	TaSEP-6	Paolacci et al. 2007
	TaSEP1.5			**4**	4	0	0	**7**	**ABBD**	this study	-	-
	TaSEP1.6		OsMADS34	**3**	3	0	0	**5**	**ABD**	this study	TaSEP-5	Paolacci et al. 2007
SEP3	TaSEP3.1		OsMADS7/45	**3**	3	0	0	**7**	**ABD**	this study	WSEP; TaSEP-4	Shitsukawa et al. 2007; Paolacci et al. 2007
	TaSEP3.2		OsMADS8/24	**3**	3	0	0	**5**	**ABD**	this study	TaMADS1; TaSEP-3	Zhao XY et al. 2006; Paolacci et al. 2007
AGL6	TaAGL6		OsMADS6/OsMADS17	**3**	3	0	0	**6**	**ABD**	this study	TaMADS#12	Murai et al. 1998
AG/STK	TaAG.1		OsMADS58/OsMADS66	**3**	3	0	0	**1**	**ABD**	Paolacci et al. 2007	WAG-1	Hirabayashi and Murai 2009
	TaAG.2		OsMADS3	**3**	3	0	0	**3**	**ABD**	Paolacci et al. 2007	WAG-2	Hirabayashi and Murai 2009
	TaSTK.1		OsMADS13	**3**	3	0	0	**5**	**ABD**	this study	TaAG-3; WSTK	Paolacci et al. 2007; Yamada et al. 2009
	TaSTK.2		OsMADS21	**3**	3	0	0	**1**	**ABD**	this study	TaAG-4	Paolacci et al. 2007
SOC1 (TM3)	TaSOC1.1		OsMADS56	**3**	3	0	0	**1**	**ABD**	Paolacci et al. 2007	-	-
	TaSOC1.5			**3**	3	0	0	**1**	**ABD**	this study	-	-
	TaSOC1.3		OsMADS50	**3**	3	0	0	**4(5)**	**ABD**	Paolacci et al. 2007	WSOC1	Shitsukawa et al. 2007
	TaSOC1.4			**3**	1	2	0	**7(4)**	**A(BD)**	this study	-	-
	TaSOC1.2		-	**3**	3	0	0	**6**	**ABD**	Paolacci et al. 2007	-	-
AP3 (DEF)	TaAP3.1		OsMADS16	**3**	3	0	0	**7**	**ABD**	this study	WAP3	Hama 2004
	TaAP3.2			**3**	3	0	0	**6**	**ABD**	this study	-	-
PI (GLO)	TaPI.1		OsMADS4	**3**	3	0	0	**1**	**ABD**	this study	WPI-1	Hama 2004
	TaPI.2		OsMADS2	**3**	3	0	0	**3**	**ABD**	this study	WPI-2	Hama 2004
ABS (GGM13)	TaBS.1		OsMADS29	**3**	3	0	0	**6**	**ABD**	this study	TaGGM13; WBSIS	Paolacci et al. 2007; Yamada et al. 2009
	TaBS.2		OsMADS31	**3**	3	0	0	**2**	**ABD**	this study	-	-
	TaBS.3			**3**	3	0	0	**7**	**ABD**	this study	-	-
	TaBS.4			**5**	5	0	0	**3**	**AABDD**	this study	-	-
	TaBS.5	*		**6**	2	0	4	**x**	**A(AAAA)D**	this study	-	-
	TaBS.6	*	OsMADS30	**4**	0	4	0	**2(6)**	**(AABB)**	this study	-	-
	TaBS.7	*		**3**	2	1	0	**1**	**(B)DD**	this study	-	-
	TaBS.8	*		**3**	0	3	0	**x**	**(ADD)**	this study	-	-
	TaBS.9	*		**3**	1	2	0	**7(3)**	**(AB)D**	this study	-	-
AGL17	TaAGL17.1		OsMADS27/OsMADS61	**3**	3	0	0	**2**	**ABD**	Paolacci et al. 2007	-	-
	TaAGL17.2			**5**	2	2	1	**5**	**(AA)BD(U)**	Paolacci et al. 2007	-	-
	TaAGL17.4		OsMADS25	**3**	3	0	0	**7**	**ABD**	this study	-	-
	TaAGL17.5			**5**	5	0	0	**7**	**AABDD**	this study	-	-
	TaAGL17.3		OsMADS57/OsMADS23	**3**	3	0	0	**6**	**ABD**	Paolacci et al. 2007	-	-
	TaAGL17.6	*		**6**	3	3	0	**7**	**(B)BBU(UU)**	this study	-	-
	TaAGL17.7		OsMADS59	**3**	2	1	0	**1**	**(B)BD**	this study	-	-
	TaAGL17.8			**3**	3	0	0	**1**	**ABD**	this study	-	-
	TaAGL17.9			**3**	3	0	0	**1**	**ABD**	this study	-	-
	TaAGL17.10		-	**3**	2	1	0	**5**	**(B)DU**	this study	-	-
	TaAGL17.11	*		**10**	4	6	0	**7(5)**	**(AA)AAB(B)D(DDD)**	this study	-	-
FLC	TaFLC.1		OsMADS37	**5**	3	2	0	**7(3)**	**ABD(DD)**	this study	TaAGL12	Ruelens et al. 2012
	TaFLC.2			**3**	3	0	0	**4(5)**	**ABD**	this study	-	-
	TaFLC.3			**3**	3	0	0	**3**	**ABD**	this study	-	-
	TaFLC.4		OsMADS51	**4**	3	1	0	**3**	**(A)ABD(D)**	this study	TaAGL22, TaAGL33	Appels et al., 2018; Sharma et al. 2017
	TaFLC.5			**2**	2	0	0	**3**	**AD**	this study	-	-
	TaFLC.6	*		**3**	2	1	0	**3**	**A(B)D**	this study	-	-
			total count	**201**	**164**	**30**	**7**					

[1] asterisk indicates that the number of homeologs could not be determined due to unclear phylogenetic resolution, genes were not included into homoeolog count (Table 2). Details see Figure S2.

[2] gene encodes for MADS but not K domain

[3] gene encodes for K but not MADS domain

[4] number in parentheses indicates a different chromosome for one of the genes; x, genes located on more than two different chromosomes

[5] parentheses indicate truncated genes, encoding for either MADS or K domain

Table 2. Groups of homoeologous MIKC-type MADS-box genes in wheat.

		wheat MIKC MADS (all)			wheat MIKC MADS (full-length only)		
homoeologous group (A:B:D)	all wheat genes[1]	# of groups	# of genes	% genes[2]	# of groups	# of genes	% of genes[3]
1:1:1	35.8%	42	126	62.7%	40	120	73.2%
n:1:1/1:n:1/1:1:n	5.7%	4	17	8.5%	2	8	4.9%
1:1:0/1:0:1/0:1:1	13.2%	1	2	1.0%	6	12	7.3%
other ratios[4]	8.0%	4	18	9.0%	3	10	6.1%
orphans/singletons	37.1%	0	0	0.0%	1	1	0.6%
excluded[5]	-	-	38	18.9%	-	13	7.9%
	99.8%		201	100.0%		164	100.0%

[1] according to Appels et al., 2018

[2] % calculated with 201 genes

[3] % calculated with 164 genes

[4] either n:1:n or 0:1:n

[5] see Table 1, Figure S2 and Methods