Gene organization and evolutionary history

The MADS-box genes, encoding the MADS-domain family of transcription factors, are involved in controlling all major aspects of the life of land plants. The MADS-domain family is characterized by the highly conserved DNA-binding MADS domain [1] (Figure 1a). The MADS domain is about 58 amino acids long and is encoded by a DNA sequence termed the MADS box. The first MADS-box gene to be isolated was ARG80 from Saccharomyces cerevisiae (Table 1) [2]. The term MADS-box gene, however, was coined later, after four subsequently characterized 'founding family members': MINICHROMOSOME MAINTENANCE 1 (MCM1) from S. cerevisiae, AGAMOUS (AG) from Arabidopsis thaliana, DEFICIENS (DEF) from Antirrhinum majus and SERUM RESPONSE FACTOR (SRF) from Homo sapiens (Table 1) [3]. Recent data suggest that the MADS box originated from a DNA sequence encoding a region of subunit A of the DNA-winding or -unwinding topoisomerase IIA enzymes, which are involved in, for example, DNA replication [4]. There is evidence that a gene encoding topoisomerase IIA subunit A was duplicated in a common ancestor of the extant eukaryotes. One of the duplicates accumulated sequence changes such that a domain with increased sequence specificity in DNA binding originated - the MADS domain [4]. As a result of subsequent gene duplication and divergence, the two major types of MADS-box genes recognized so far - type I or SRF-like and type II or MEF2-like (after MYOCYTE ENHANCER FACTOR 2) - were established in the most recent common ancestor (MRCA) of extant eukaryotes [4]. These two types are distinguished on the basis of sequence dissimilarity and also differ in sequence specificity in DNA binding and the amount of DNA bending that they induce [1].

The MADS-box family subsequently diversified in remarkably different ways in the various eukaryotic lineages during evolution. The number of MADS-box genes has remained quite small in protists, animals and fungi. In contrast, their numbers greatly increased in some plant lineages, so that there is now just one MADS-box gene in extant green algae of the chlorophyte lineage, but more than 20 in mosses and around 100 in angiosperm plants - flowering plants such as A. thaliana, Populus trichocarpa (poplar) and Oryza sativa (rice) [5-7]. MADS-box transcription factors contribute to a wide range of biological processes, ranging from muscle development, cell proliferation and differentiation in...
**Table 1. Summary of MADS-box genes and their functions**

| Name and synonyms | Type | Group | Organism | Function |
|-------------------|------|-------|----------|----------|
| AGAMOUS (AG)      | Type II | MIKC<sup>C</sup> | A. thaliana | Stamen and carpel development |
| AGAMOUS-LIKE 15 (AGL15) | Type II | MIKC<sup>C</sup> | A. thaliana | Flowering time determination |
| AGAMOUS-LIKE 23 (AGL23) | Type I | Ma | A. thaliana | Female gametophyte development |
| AGAMOUS-LIKE 24 (AGL24) | Type II | MIKC<sup>C</sup> | A. thaliana | Flowering time determination |
| AGAMOUS-LIKE 28 (AGL28) | Type I | Ma | A. thaliana | Potentially flowering time determination |
| AGAMOUS-LIKE 37 (AGL37), PHERES1 | Type I | Ma | A. thaliana | Seed development |
| AGAMOUS-LIKE 61 (AGL61), DIANA | Type I | Ma | A. thaliana | Central cell development |
| AGAMOUS-LIKE 62 (AGL62) | Type I | Ma | A. thaliana | Central cell development |
| AGAMOUS-LIKE 66 (AGL66) | Type II | MIKC<sup>*</sup> | A. thaliana | Pollen development |
| AGAMOUS-LIKE 80 (AGL80) | Type I | Ma | A. thaliana | Central cell development |
| AGAMOUS-LIKE 104 (AGL104) | Type II | MIKC<sup>*</sup> | A. thaliana | Pollen development |
| APETALA1 (AP1) | Type II | MIKC<sup>C</sup> | A. thaliana | Sepal and petal development |
| APETALA3 (AP3) | Type II | MIKC<sup>C</sup> | A. thaliana | Petal and stamen development |
| ARABIDOPSIS BISISTER (ABS) | Type II | MIKC<sup>C</sup> | A. thaliana | Endothelium development |
| ARG80 | Type I | NA | Saccharomyces cerevisiae | Arginine metabolism |
| CAULIFLOWER (CAL) | Type II | MIKC<sup>C</sup> | A. thaliana | Floral meristem development |
| CerMADS2 | Type II | MIKC<sup>C</sup> | Ceratopteris reinhardtii | NA |
| CerMADS3 | Type II | MIKC<sup>C</sup> | C. reinhardtii | NA |
| CMADS1 | Type II | MIKC<sup>C</sup> | C. reinhardtii | NA |
| DEFICIENS (DEF) | Type II | MIKC<sup>C</sup> | Antirrhinum majus | Petal and stamen development |
| FLOWERING LOCUS C (FLC) | Type II | MIKC<sup>C</sup> | A. thaliana | Vernalization response |
| FRUITFULL (FUL) | Type II | MIKC<sup>C</sup> | A. thaliana | Floral meristem development, fruit formation |
| GmSEP1 | Type II | MIKC<sup>C</sup> | Glycine max | Reproductive organ development |
| MADS AFFECTING FLOWERING 1 (MAF1) | Type II | MIKC<sup>C</sup> | A. thaliana | Flowering time determination |
| MINICHROMOSOME MAINTENANCE 1 (MCM1) | Type I | NA | S. cerevisiae | Arginine metabolism, cell viability, mating, minichromosome maintenance, recombination, osmotolerance |
| MASAKO C1-C6 | Type II | MIKC<sup>C</sup> | Rosa rugosa | Stamen and carpel development |
| OsMADS22 | Type II | MIKC<sup>C</sup> | Oryza sativa | Plant hormone response |
| OsMADS47 | Type II | MIKC<sup>C</sup> | O. sativa | Plant hormone response |
| OsMADS50 | Type II | MIKC<sup>C</sup> | O. sativa | Flowering time determination |
| OsMADS58 | Type II | MIKC<sup>C</sup> | O. sativa | Floral meristem determination |
| OsMADS60 | Type II | MIKC<sup>C</sup> | O. sativa | NA |
| OsMADS88 | Type I | Ma | O. sativa | NA |
| OsMADS99 | Type I | Ma | O. sativa | NA |
| PISTILLATA (PI) | Type II | MIKC<sup>C</sup> | A. thaliana | Petal and stamen development |
| PPM1 | Type II | MIKC<sup>C</sup> | Physcomitrella patens | NA |
| PpMADS1 | Type II | MIKC<sup>C</sup> | P. patens | NA |
| PPMCS | Type II | MIKC<sup>C</sup> | P. patens | NA |
| SEPALLATA1-4 (SEP 1-4) | Type II | MIKC<sup>C</sup> | A. thaliana | Floral organ development |
| SERUM RESPONSE FACTOR (SRF) | Type I | NA | Homo sapiens | Immediate–early response |
| SHATTERPROOF 1 (SHP1) and SHATTERPROOF 2 (SHP2) | Type II | MIKC<sup>C</sup> | A. thaliana | Fruit dehiscence |
| SHORT VEGETATIVE PHASE (SVP) | Type II | MIKC<sup>C</sup> | A. thaliana | Flowering time determination |
| SUPPRESSOR OF CONSTANS 1 (SOC1) | Type II | MIKC<sup>C</sup> | A. thaliana | Flowering time determination |
| VERNALISATION 1 (VRN1) | Type II | MIKC<sup>C</sup> | Triticum aestivum | Vernalization response |

For brevity, a processes or developmental context (rather than a specific function) is given that is controlled or specified by the corresponding MADS-domain transcription factor. NA, functional information is not available.
animals to pheromone responses in fungi [1]. This review will focus on the family of plant MADS-box genes, which is involved in controlling all major aspects of the life of land plants [8] but which still has a large number of uncharacterized members (see, for example, [5-7]).

In line with recent phylogenetic analyses [4], the type I and type II MADS-box genes of plants have been hypothesized to be orthologous to the SRF-like and MEF2-like MADS-box genes, respectively, in animals and fungi but whether all plant genes annotated as type I or II really constitute two distinct clades is not completely clear [4,9,10]. Type I and type II genes in plants differ in a number of features. Whereas type I MADS-box genes have usually one or two exons, type II genes have an average of seven [11]. Type I and type II genes also differ in their evolutionary rates, with evidence that type I MADS-box genes experience faster rates of gene birth and death [12]. Type I genes seem to have experienced more small-scale duplications, probably as a consequence of their shorter length [13], whereas type II MADS-box genes were preferentially retained after whole-genome and large-scale duplications, probably owing to their potential for evolving new functions or subdividing functions of the ancestral gene after duplication, or because of a requirement for a balanced number of proteins in multimeric complexes [12]. Consequently, the numbers of type I MADS-box genes is more variable in different angiosperms than those of type II genes.

At the protein level, the two types of MADS-domain proteins differ in domain structure (Figure 1b). Type II MADS-domain transcription factors of plants are characterized by the presence of a keratin-like (K) domain and are commonly referred to as MIKC-type proteins after their domain structure: MADS, intervening (I), keratin-like (K) and carboxy-terminal (C) domains [5,8]. The type I proteins have no K domain (Figure 1b).

The type I MADS-domain proteins of plants can be further subdivided into the three groups Mα, Mβ and Mγ on the basis of their phylogeny and the presence or absence of conserved motifs in the carboxy-terminal region [5,11,14]. Type II MADS-domain proteins are subdivided into the groups MIKC+ and MIKC* as judged from the number of exons that encode the I domain and structural differences in the K domain [15]. MIKC+ group proteins can then be further subdivided into quite a number of ancient clades as revealed by phylogenetic reconstruction [8]. Likewise, two classes of MIKC* proteins are distinguished - S and P [16].

The increasing amount of whole-genome information for different plant species allows insights into the evolution of these different types and groups of MADS-box genes. Ostreococcus tauri [17] and Ostreococcus lucimarinus [18] - green algae belonging to the chlorophyte lineage - each contain only one MADS-box gene (see O. tauri in Figure 2). The predicted proteins lack a K domain and could represent type I proteins. Tanabe et al.
characterized type II MADS-box genes in three charophyte green algae, the most basal of the streptophytes - a group that includes the charophyte and land plant (embryophyte) lineages (Figure 2). Thus, the K domain joined an ancestral type II MADS-domain protein to form the MIKC-group proteins near the base of the streptophyte lineage more than 700 million years ago (Figure 2) [20]. The MIKC-type MADS-box genes from charophyte algae do not form a clade with the MIKC* genes with high support, and thus might represent ancestral homologs of MIKC-type genes [21]. This ancestral MIKC-type gene was probably duplicated in the lineage that led to extant land plants, which gave rise to the MIKC* and MIKC* groups (Figure 2). The earliest-branching species of land plants for which whole-genome information is currently available is the moss Physcomitrella patens [22]. The family of MADS-box genes has expanded to about 23 members in the lineage that led to this moss (Figure 2). MIKC* and MIKC* genes, as well as Mα and Mβ genes, have been annotated for P. patens, suggesting that the MRCA of extant mosses and vascular plants about 450 million years ago already had at least one representative of all of these clades: that is, four different MADS-box genes (Figure 2). The major clades of MIKC* genes were established more than 300 million years ago, so that the MRCA of extant seed plants had at least 10 MIKC* group genes [23].

Extensive genome-wide annotation and phylogenetic studies of MADS-box genes have been conducted for A. thaliana [5], poplar [7] and rice [6] and allow predictions of the MADS-box gene constitution of the MRCA of extant monocotyledons (monocots) and eudicotyledons (eudicots) - the two major groups of flowering plants. At least three type I (one of each group Mα, Mβ and Mγ), eleven MIKC* group and two MIKC*-group MADS-box genes (one of each of the classes S and P) were present in the MRCA of monocots and eudicots (Figure 2) [5-7]. Since then, the total number of MADS-box genes has at least doubled in all angiosperm species analyzed so far. Expansions in the Mα, Mβ and My
groups seem to be lineage-specific, whereas the major clades of the MIKCc group and the two classes of the MIKC* group show only occasional lineage-specific expansions (Figure 3) [5-7]. In the case of monocot genomes, a total of 61 MADS-box genes has been reported for Zea mays (maize) [24] and 75 in O. sativa [6], whereas the total number of MADS-box genes in eudicots ranges from 105 in P. trichocarpa [7], through 107 in A. thaliana [5] to 212 in Glycine max (soybean) [25]. The number of MADS-domain transcription factors could be significantly higher than the gene numbers in all species because of alternative splicing [26], for example, but there is only limited evidence that alternative splicing is of functional importance for MADS-box genes in plants.

Characteristic structural features

The DNA-binding MADS domain of the protein is encoded by the MADS box, which is usually located on one exon [15]. Crystal structures of plant MADS domains are not known. However, structures of SRF and MEF2 from humans and MCM1 from S. cerevisiae [27-29] reveal that the MADS domain folds into an amino-terminal extension of 14 amino acids, followed by a long amphipathic α-helix and two β-strands (the structure of SRF is shown in Figure 4a) [27]. Contact with the minor groove of DNA is mediated by the amino-terminal extension, while one face of the α-helix contacts the major groove. MADS-domain proteins, including the plant proteins, bind to DNA as homo- or heterodimers [20], where the heterodimeric interaction partner is usually another MADS-domain protein. The two β-strands, which form a β-sheet with the two β-strands of the interaction partner, are required for dimerization [27].

Dimers of MADS-domain transcription factors bind to CArG-boxes, stretches of DNA with the consensus sequence 5’-CC[A/T]G-G-3’ [20], or very similar sequences [21]. CArG-box motifs are very common in the genome, as they are short and variable, and target gene prediction based solely on these motifs is difficult [30]. How MADS-domain proteins achieve target gene specificity is thus still unclear [31].

For most plant MADS-domain proteins, the MADS domain represents the amino-terminal domain of the protein [5,6]. However, some plant MADS-domain transcription factors have distinct amino-terminal regions upstream of the MADS domain. Their sequences are very...
diverse and no function has been assigned to these amino-terminal regions so far. Examples of MADS-domain proteins with an amino-terminal extension include the type I protein AGAMOUS-LIKE61 (AGL61, DIANA) [32], the MIKC C-group (type II) protein AG of *A. thaliana* (Table 1) and many of its close relatives [33], some MADS-domain proteins of both types from *O. sativa* (for example, OsMADS58, OsMADS60, OsMADS88 and OsMADS99) [6], as well as the MIKC C-group proteins CMADS1, CerMADS2 and CerMADS3 of the fern *Ceratopteris richardii* [34] (Table 1). Apart from the MADS domain, no conserved domains are present in type I MADS-domain proteins (Figure 1b). Several conserved sequence motifs specific for the Mα, Mβ and Mγ groups have been identified [5,6,11]. However, they are not related to any other known motifs and no structure and functions can be assigned to them.

As mentioned earlier, the type II MIKC-type proteins of plants have a much more defined and conserved domain structure (Figure 1b) than the type I proteins. The K domain is, after the MADS domain, the best-conserved domain of type II proteins. It is usually encoded in three exons, is approximately 70 amino acids long [15] and is subdivided into three subdomains, K1, K2 and K3 (Figure 1b) [20]. The subdomains largely coincide with the exons. Each subdomain is characterized by a heptad repeat [abcdefg], where positions a and d usually contain hydrophobic amino acids [35]. The subdomains form amphipathic α-helices that are predicted to form coiled coils and thereby mediate protein-protein interactions of MADS-domain proteins [35]. More specifically, in some cases, K1 is required for DNA-binding dimer formation. K1 and K2 generally support the formation of DNA-binding dimers, while K3 may contribute to multimerization [20].

The I domain and the C domain are the least conserved domains of MIKC-type proteins [5]. Besides structural differences in the K domain, the length of the sequence and the number of exons encoding the I domain distinguish MIKC type proteins [15]. Whereas MIKC C-type proteins have a short I domain encoded by only one or two exons, MIKC type proteins have a longer I domain encoded by four or five exons [15,21]. The I domain influences the specificity of DNA-binding dimer formation. Together with the MADS domain, it is often sufficient for the formation of DNA-binding dimers. The C domain is encoded by a variable number of exons and in some proteins, it is important for the activation of transcription of target genes [36] and may also be important for the formation of multimeric complexes. However, a recent study of SEP3 of *A. thaliana* (Table 1) revealed that the C domain is not required for multimerization in...
this case, whereas subdomain K3 is essential, at least in the absence of the C domain [37].

Multimeric complexes of MIKC-type proteins have been suggested to be required for the specification of the floral organs - sepals, petals, stamens and carpels [38]. The 'floral quartet model' (Figure 4b) hypothesizes that two dimers of MADS-domain proteins bind to neighboring CArG-boxes and interact with each other. This interaction leads to loop formation of the intervening DNA and finally to differential regulation of target genes by different complexes (Figure 4b).

A number of MADS-box gene primary transcripts are known to be alternatively spliced. Alternative splicing has been demonstrated in two AG homologs in cotton [26], MASAKO C1-C6 from rose (Table 1), and around 20 MADS-box genes in A. thaliana [5]. The relevance of these observations has remained elusive, however, as differential functions for the alternative splice products have yet to be revealed.

**Localization and function**

**Cellular localization**

As transcription factors, MADS-domain proteins are assumed to be localized in the nucleus. Several stretches rich in basic residues in the MADS domain have been identified as nuclear localization signals (NLS) [39-41]. The most prominent signal for translocation into the nucleus is the motif KR[K/R]XKK at positions 22 to 30 of the MADS domain (Figure 1a). Subcellular localization was analyzed for two type I proteins from plants, AGL61 [32] and AGL80 [42] from A. thaliana, and for several type II MADS-domain transcription factors, such as GmSEP1 from soybean [43], OsMADS22, OsMADS47 and OsMADS50 [44] from rice and PISTILLATA (PI), APETALA3 (AP3) [40], ARABIDOPSIS BSISTER (ABS) [45], AGL24 [46] and AGL15 [47] from A. thaliana (Table 1), all of which were shown to be indeed localized in the nucleus.

**Function**

For a better understanding of MADS-domain protein function, one should remember that all land plants have a complex life cycle with the alternation of multicellular sexual and asexual phases (haploid gametophyte and diploid sporophyte, respectively; Figure 5). Land plants are very probably monophyletic and originated from haploid streptophyte algae [21].

**Type I genes**

For quite a long time little was known about type I MADS-box genes in plants, so that they were somewhat the ‘dark matter of the MADS universe.’ This is in part due to the fact that the type I genes were first identified by genomic studies rather than by forward genetics, and no mutant phenotypes were known [32]. Moreover, plant type I genes are only weakly expressed in all species analyzed so far [5-7,10,14,32,42,48-50]. One could assume, therefore, that limited functional importance or functional redundancy contributed to the fact that their roles remained elusive [5-7,11].

In recent years, however, the situation has changed dramatically. Several pioneering studies on type I MADS-box genes from A. thaliana revealed that they are important for female gametophyte, embryo sac and seed development [13,14,32,48,49]. Four A. thaliana genes of the Ma group have been functionally characterized, namely AGL23 [48], AGL28 [14], AGL61 [32] and AGL62 (Table 1) [49]. Plants with an insertion of Agrobacterium T-DNA in the MADS-box of AGL23 (probably resulting in a loss of AGL23 function) showed arrest of female gametophyte development and persistence of the megaspore during subsequent phases of ovule development. In addition, agl23 mutants develop albino seeds that have no chloroplasts and so do not develop into viable plants [48]. Thus, AGL23 plays an important role in the development of the female gametophyte and, in addition, is involved in controlling the biogenesis of organelles during embryo development. AGL61 functions together with AGL80 to differentiate the central cell in the gametophyte. In agl61 mutant ovules, the polar nuclei do not fuse and central cell morphology is aberrant. In addition, the central cell begins to degenerate before fertilization, and so no zygote or endosperm is formed [32]. Similarly, the seeds of agl62 mutants suffer from premature formation of cell walls in the endosperm [49]. Loss of function of agl28, the closest homolog of AGL23, has no obvious mutant phenotype [14]. Precocious over-expression of AGL28 leads to early flowering, suggesting (but not conclusively demonstrating) that this gene functions in the promotion of flowering [14].

No MADS-box gene of the Mb group has been functionally characterized so far, and AGL37 (PHERES1) and AGL80 of A. thaliana are the only two functionally analyzed genes of the Mg group (Table 1). AGL37 is regulated epigenetically and has a key role in seed development [50]. The AGL80 protein can form a heterodimer with AGL61, and an insertion of T-DNA into the MADS-box of AGL80 (probably resulting in loss of function) leads to altered development of the central cell and endosperm development is not initiated, similar to the AGL61 phenotype [42].

Even though only a small fraction of all type I genes of A. thaliana and no genes of other species have been functionally characterized yet, it is tempting to speculate that the plant type I genes in general have a functional focus on female gametophyte, embryo and seed development, that is, in controlling the ‘female side’ of plant life (Figure 5).
Type II genes

Much more is known about the type II MIKC-type MADS-box genes. In fact, they are one of the most intensively and comprehensively studied families of plant genes in terms of both their developmental genetics and their phylogenomics [5-10,20,21,23,51,52]. MIKC-type proteins of charophyte green algae are only expressed in gametangial cells (cells of the structures that produce the gametes) and thus might have a function in the differentiation of these cells [19]. Some charophyte algae are the sister group of land plants, but are purely haploid organisms (with only the zygote being diploid). The importance of MIKC-type genes for the development of diploid sporophytes in land-plant life cycles thus probably originated in the lineage that led to land plants.

MIKC\textsuperscript{C}-group genes represent the best-studied group of MADS-box genes and their evolution and functions have been extensively reviewed (see, for example, [8,10,20,23,51-53]). MIKC\textsuperscript{C}-group genes are expressed in both gametophytes and sporophytes in mosses and ferns [21], but almost exclusively in the sporophyte in \textit{A. thaliana} (Figure 5). Expression patterns of the genes of different clades of MIKC\textsuperscript{C}-group genes are often similar for rice and \textit{A. thaliana}, especially in reproductive tissues, indicating a conservation of function [6]. In contrast to MIKC\textsuperscript{*}-group genes, \textit{AGL18} is the only MIKC\textsuperscript{C}-group gene of \textit{A. thaliana} to be expressed in the gametophyte [21] (Table 1). Some MIKC\textsuperscript{C}-group genes of rice, tomato and \textit{A. thaliana}, among others, are affected by stress treatment, indicating that they are involved in regulating flowering time in response to stress [6].

MIKC\textsuperscript{C}-group genes control various aspects of sporophyte development (Figure 5) [5-8,20,23,52]. These genes are especially prominent on almost all levels of the gene regulatory network that controls reproductive development in flowering plants such as \textit{A. thaliana}. Accordingly, MIKC\textsuperscript{C}-group genes determine flowering time (for example, \textit{SUPPRESSOR OF CONSTANS 1 (SOC1), FLOWERING LOCUS C (FLC), AGL24, MADS AFFECTING FLOWERING 1 (MAF1), and SHORT VEGETATIVE PHASE (SVP)}) and specify floral meristem identity (for example, \textit{AP1, FRUITFULL (FUL), CAULIFLOWER (CAL)}) , floral organ identity (for example, \textit{AP1, SEP1 to 4, AP3, PI, and AG; Figure 4b}), fruit formation (for example, \textit{SHATTERPROOF 1 (SHP1) and SHP2, and FUL}) and seed pigmentation (for example, \textit{ABS}) [5,20,45] (Table 1).
MIKC*-group genes were detected much later than MIKCc-group genes because of their functional redundancy and their restricted functional importance in the plant life cycle [15]. From mosses to A. thaliana, the expression of MIKCc-group genes is largely restricted to the gametophytic generation; in A. thaliana, expression is confined to male gametophytes [21]. Consequently, MIKC*-group genes have been suggested to have in general critical roles in gametophyte development in all land plants [21]. More specifically, in the case of AGL66 and AGL104 of A. thaliana, these genes regulate pollen maturation; that is, male gametophyte development (Figure 5) [16,54].

Control of MADS-box gene expression
The MADS-box genes are regulated in various ways: transcriptional regulation by transcription factors, often constituting feedback and feed-forward loops, epigenetic control, and regulation by microRNAs (miRNAs) have all been identified. The best-studied examples of epigenetic control are in the two type II MIKCc-group genes FLC of A. thaliana and VERNALISATION 1 (VRN1) of Triticum aestivum (common wheat) (Table 1), both of which channel the response to vernalization (exposure to a prolonged period of cold) and thereby regulate flowering time [55]. Whereas FLC is repressed after vernalization by the deacetylation of histones at its locus, the reduction of repressive histone methylations and increase of activating histone methylations at VRNJ lead to activation of this gene by vernalization [55]. The Polycomb-group protein CURLY LEAF represses a MIKCc-group gene involved in flower development, AG of A. thaliana, by increasing repressive histone methylations [56,57], and the type I gene AGL37 [50] is repressed by MEDEA, which confers repressive histone methylations. The trithorax-group protein ATX1 of A. thaliana maintains the active state of the floral homeotic genes AP1 and AG by inducing activating histone methylations and thereby ensures the normal development of floral organs [53].

Several MIKCc-group (type II) MADS-box genes are targeted by miRNAs. Specific miRNAs bind to the mRNAs of these genes, leading to the cleavage and subsequent degradation of the mRNAs. The miR-444 family of miRNAs is specific to the Poaceae (the grasses) and targets mRNAs of AGL17-like MADS-box genes [58]. Overexpression or loss-of-function phenotypes for the miR-444 family are not known. AGL16 determines the density and distribution of stomata on leaves of A. thaliana and is regulated by miR-824, which is specific to the Brassicaceae (the mustard family) [59]. There are two different alleles of the precursor of miR-824 in A. thaliana that differ in their thermostability and are maintained by balancing selection [59]. Finally, the miR-538 family from the moss P. patens is predicted to target three MIKCc-group MADS-box genes, namely PPM1, PpMADS1 and PPMC5 [60]. Again, overexpression or loss-of-function phenotypes for the miR-538 family are not known. The target genes of the miR-444 family (the AGL17-like genes) and of miR-824 (AGL16), lie within one clade of MIKCc-group genes, whereas the target genes of miR-538 (PPM1, PpMADS1 and PPMC5) do not seem to have orthologs in angiosperms [8,15] (Table 1). In the A. thaliana Landsberg erecta ecotype, short interfering RNAs (siRNAs) recruit a methylase to the promoter of FLC, which initiates heterochromatization and thus inhibition of the FLC promoter and thereby promotes flowering, as FLC is a vernalization-dependent repressor of flowering (see above) [61].

MADS-domain transcription factors themselves form dimers and multimeric complexes that bind to DNA and thereby regulate their target genes by direct transcriptional activation or repression [31,38,62]. Complex formation provides a basis for the formation of feedback and feed-forward loops. These loops constitute regulatory mechanisms that have in plants, so far, only been shown for MIKCc-group genes, but probably also have a role in controlling the expression of other plant MADS-box genes. A positive autoregulatory feedback loop was identified, for example, for the floral organ identity genes AP3 and PI of A. thaliana [53]. In this case, the AP3 and PI proteins form obligate heterodimers that upregulate the expression of their own genes. This intriguing kind of gene interdependence may have helped to canalize the structure of the flower during evolution and may confer robustness during development [63]. In A. thaliana, a feed-forward loop regulates flowering time and includes LEAFY (LFY), which is not a MADS-box gene but encodes another kind of transcription factor, and the two MIKCc-group MADS-box genes AP1 and SEP3. LFY activates AP1, which in turn activates SEP3. SEP3 then, together with LFY, activates AG, AP3 and PI [53] (Table 1). This feed-forward loop prevents precocious differentiation of the floral organs.

Plant MADS-box genes provide excellent and widely recognized examples of functional redundancy, often involving several genes. Prominent examples are the MIKCc-group class E floral organ identity genes SEP1, SEP2, SEP3 and SEP4, which are largely redundant, as are SHP1 and SHP2 (redundantly involved in seed dehiscence) and API, CAL and FUL (flowering time) (Table 1) [5,64]. Redundancy of MADS-box genes is thought to confer developmental robustness [64]. Functional redundancy has also been revealed for A. thaliana MIKC*-group genes involved in male gametophyte development [16,54] and has also been suggested for type I MADS-box genes because of the low number of phenotypic mutants [5,13].
Frontiers

The critical role of MADS-domain proteins in plant development was shown by comprehensive functional studies. However, transcription factors mainly belonging to the MIKC\(^*\)-group of some model species have been characterized in detail so far. It will be revealing to elucidate the functions of MADS-box genes from a greater variety of plants and from more type I genes to obtain a more representative picture of MADS-domain protein functions.

More and more MADS-box genes are being identified by sequencing of whole genomes and transcriptomes from a wide range of plants. It will be interesting to follow their evolution to infer the ancestral functions of the MADS-box gene(s) in the MRCA of, for example, streptophytes and land plants. Furthermore, increasing knowledge of MADS-box genes in green plants will help to better understand the role of their expansion in plant evolution.

MADS-domain transcription factors bind to CaR\(G\)-boxes of their target genes. However, the number of CaR\(G\)-boxes in the genome is enormous, and different MADS-domain proteins recognize different sets of target genes. Further studies on direct target genes of these transcription factors will help to elucidate how MADS-domain proteins recognize the promoters of their target genes and might even enable the development of algorithms to predict target genes.

Given the impression that MIKC-type genes have more prominent functions in land plants than type I genes, the acquisition of the K domain of MADS-domain transcription factors and the subsequent diversification of the emerging MIKC-type MADS-box genes seems to have played a key role in the evolution of land plants. However, no clear homologous domain has been identified in eukaryotes other than plants or in bacteria. Thus, it will be exciting to elucidate the origin of the K domain.

Acknowledgements

We are grateful to the Friedrich Schiller University in Jena for general support and to all members of the Theissen laboratory for valuable discussions.

References

1. Messenguy F, Dubois E. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene 2003, 316:1-21.
2. Dubois E, Bercy J, Descamps F, Messenguy F. Characterization of two new genes essential for vegetative growth in Saccharomyces cerevisiae: nucleotide sequence determination and chromosome mapping. Gene 1987, 55:265-275.
3. Schwarz-Sommer Z, Huisjes P, Nacken W, Saedler H, Sommer H. Genetic control of flower development by homeotic genes in Antirrhinum majus. Science 1990, 250:931-936.
4. Gramzow L, Ritz MS, Theissen G. On the origin of MADS-domain transcription factors. Trends Genet 2010, 26:149-153.
5. Pirescova I, de Folter S, Kieffer M, Horner DS, Fasoli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies S, Angenent GC, Colombo L. Molecular and phylogenetic analysis of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 2003, 15:1538-1551.
6. 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, 8:242.
7. Leseburg CH, Li A, Kang H, Duval M, Mao L. Genome-wide analysis of the MADS-box gene family in Populus trichocarpa. Gene 2006, 378:84-94.
8. Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 2002, 29:464-489.
9. Alvarez-Buylla ER, Pelaz S, Liljenberg SJ, Gold SE, Burgeff C, Ditta GS, de Pousplana LR, Martinez-Castilla L, Yanofsky MF. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci USA 2000, 97:5328-5333.
10. De Bodt S, Raes J, De Peere Y, Theissen G. And then there were many: MADS goes genomic. Trends Plant Sci 2003, 8:475-483.
11. De Bodt S, Raes J, Floquin K, Rombouts S, Rouze P, Theissen G, Van de Peer Y. Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol 2003, 56:573-586.
12. Nam J, Kim J, Lee S, An G, Ma H, Noh M. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc Natl Acad Sci USA 2004, 101:1910-1915.
13. Bemer M, Gordon J, Waterings K, Angenent GC. Divergence of recently duplicated My-type MADS-box genes in Petunia. Mol Biol Evol 2010, 27:481-495.
14. Yoo SK, Lee JS, Ahn JH. Overexpression of AGAMOUS-LIKE 28 (AGL28) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. Biochem Biophys Res Commun 2005, 348:929-936.
15. Henschel K, Kofugi R, Hasebe M, Saedler H, Munster T, Theissen G. Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Mol Biol Evol 2002, 19:801-814.
16. Adamczyk BJ, Fernandez DE. MIKC\(^*\) MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. Plant Physiol 2009, 149:1713-1723.
17. Derelle E, Ferrer C, Rombouts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroove S, Echeyne S, Cooke R, Saiyes Y, Wayts J, Jabbari K, Bowler C, Panaud O, Pürg U, Ball SG, Bali J, Rouzet P, Picard A, Lelsius M, Demaille J, Van de Peer Y, Moreau H. Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc Natl Acad Sci USA 2006, 103:11647-11652.
18. Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rouze P, Zhou K, Chilera R, Merchant SS, Podell S, Gaasterland T, Nappi C, Gendler K, Manuel A, Tai V, Vallon O, Piganee G, Janecz S, Heidej M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchek I, Pazour GJ, et al. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA 2007, 104:7705-7710.
19. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Munster T, Theissen G, Nozaki H, Ito M. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc Natl Acad Sci USA 2005, 102:4246-4241.
20. Kaufmann K, Melzer R, Theissen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and networks evolution in land plants. Gene 2005, 347:183-198.
21. Zobel O, Faigl W, Saedler H, Munster T, Theissen G. MIKC\(^*\) MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol Biol Evol 2010, 27:1201-1211.
22. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Pouplana LR, Martinez-Castilla L, Yanofsky MF. Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol 2003, 56:573-586.
23. Melzer R, Wang YQ, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 2002, 29:464-489.
24. Soderlund C, Descour A, Kudrna D, Bomhoff M, Boyd L, Currie J, Angelova A, Collura K, Wissotski M, Ashley E, Morrow D, Fernandes J, Walbot V, Yu Y.
Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. PLoS Genet 2009, 5:e1000740.

25. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Telenius H, Cheng J, Xu D, Helentjaris T, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharjya MK, Sandhu D, Vallyathan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Briscoe M, Abernathy B, Dui J, Tian Z, Zhu L, et al.: Genome sequence of the palaeopolyploid soybean. Nature 2010, 461:78-183.

26. LIGHTFOOT DJ, MALONE KM, TIMMIS JN, ORFORD SJ: Evidence for alternative splicing of MADS-box transcripts in developing cotton fibre cells. Mol Genet Genomics 2008, 279:75-85.

27. Pellegrini L, Song T, Richmond TJ: Structure of serum response factor core domain bound to DNA. Nature 1995, 376:490-498.

28. Santelli E, Richmond TJ: Crystal structure of MEF2A core bound to DNA at 1.5 angstrom resolution. J Mol Biol 2000, 297:437-449.

29. Tan S, Richmond TJ: Crystal structure of the yeast MAT alpha 2/CM1/DNA ternary complex. Nature 1998, 391:660-666.

30. Verelst W, Twel D, de Folter S, Immink R, Saedler H, Munster T: MADS-box complexes regulate transcription dynamics during pollen maturation. Genome Biol 2007, 8:R249.

31. Melzer R, Theissen G: Reconstitution of ‘floral quartets’ in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Res 2009, 37:2723-2736.

32. Berner M, Wolters-Arts M, Grossniklaus U, Angenent GC: The MADS domain protein DIANA acts together with AGAMOUS-LIKE8 to specify the central cell in Arabidopsis ovules. Plant Cell 2008, 20:2088-2101.

33. Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF: Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 1995, 140:345-356.

34. Hasebe M, Wem CK, Kato M, Banks JA: Characterization of MADS homeotic genes in the fern Ceratopteris richardii. Proc Natl Acad Sci USA 1998, 95:6222-6227.

35. Yang Y, Fanning L, Jack T: A possible working mechanism for rice SVP-group floral organ identity proteins, APETALA3 and PISTILLATA. Plant Cell Physiol 2004, 45:571-581.

36. Homma T, Goto K: Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature 2001, 409:525-529.

37. Melzer R, Verelst W, Theissen G: The class E floral homeotic protein AGL80 is specifically bound to and phosphorylated by serine/threonine-dependent protein kinase. Plant Cell Physiol 2003, 44:735-742.

38. Perry SE, Lehti MD, Fernandez DE: The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. Plant Physiol 1999, 120:121-130.

39. Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L: AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. Plant J 2008, 54:1037-1048.

40. KANG HS, STEFFEN JG, PORTERIEKO MF, LLOYD A, DREWS GN: The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. Plant Cell 2008, 20:635-647.

41. Kóhler C, Hennig L, Spillane C, Pien S, Grassme W, Grossniklaus U: The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHENEST. Gene Dev 2003, 17:1540-1553.

42. Irish VF, Litt A: Flower development and evolution: gene duplication, diversification and recombination. Curr Opin Genet Dev 2005, 15:454-460.

43. Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KL, Saedler H: A short history of MADS-box genes in plants. Plant Mol Biol 2000, 42:115-149.

44. Liu Z, Mara C: Regulatory mechanisms for floral homeotic gene expression. Semin Cell Dev Biol 2010, 21:80-86.

45. Verelst W, Twel D, de Folter S, Immink R, Saedler H, Munster T: MADS-box complexes regulate transcriptome dynamics during pollen maturation. Genome Biol 2007, 8:R249.

46. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, Goodrich J: Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J 2006, 25:4638-4649.

47. Perry SE, Lehti MD, Fernandez DE: The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. Plant Physiol 1999, 120:121-130.