Title page

Classification:
Biological Sciences – Evolution

Title:
MIKC* MADS-box genes: Conserved regulators of the gametophytic generation of land plants

Authors:
Oliver Zobell,[a,b] Wolfram Faigl,[a] Heinz Saedler,[a] and Thomas Münster [a]

Author affiliations:
[a] Department of Molecular Plant Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, D-50829 Cologne, Germany
[b] Department of Botany, University of Osnabrück, Barbarastrasse 11, D-49076 Osnabrück, Germany

Corresponding authors:
Oliver Zobell
Department of Botany
University of Osnabrück
Barbarastrasse 11
D-49076 Osnabrück
Germany
Tel. +49-541-9692272
Fax +49-541-9692845
oliver.zobell@biologie.uni-osnabrueck.de

Thomas Münster
Department of Molecular Plant Genetics
Max Planck Institute for Plant Breeding Research
Carl-von-Linne-Weg 10
D-50829 Cologne
Germany
Tel. +49-221-5062122
Fax +49-221-5062113
muenster@mpiz-koeln.mpg.de

Manuscript information:
Manuscript: 18 text pages
Supplementary information: 6 text pages
Figures: 6
Supplementary figures: 4
Supplementary tables: 2

Abbreviations:
MADS-box, *MCM1-AGAMOUS-DEFICIENS-SRF*-box; *MpMADS1*, *Marchantia polymorpha* MADS-box gene 1; *SsMADS1* through *SsMADS4*, *Sphagnum*

subsecundum MADS-box genes 1 through 4; *AuMADS1*, *Atrichum undulatum* MADS-box gene 1; *FhMADS1* through *FhMADS11*, *Funaria hygrometrica* MADS-box genes 1 through 11.

Data deposition footnote:
The sequences reported in this paper have been deposited in the GenBank database (accession nos. GQ334454-GQ334470).

Author contributions footnote:
O.Z. and T.M. designed research; O.Z. and W.F. performed experiments; O.Z. analyzed data and wrote the manuscript; T.M. and H.S. edited the manuscript.

Abstract

Land plants (embryophytes) are characterized by an alteration of two generations, the haploid gametophyte and the diploid sporophyte. The development of the small and simple male gametophyte of the flowering plant Arabidopsis (*Arabidopsis thaliana*) critically depends on the action of five MIKC* group MADS-box proteins. In this study, these MIKC* MADS-box genes were isolated from land plants with relatively large and complex gametophyte bodies, namely the bryophytes. We found that while the gene family expanded in the mosses *Sphagnum subsecundum*, *Physcomitrella patens* and *Funaria hygrometrica* only a single homologue, *MpMADS1*, has been retained in the liverwort *Marchantia polymorpha*. Liverworts are the earliest-divergent land plants, and so *MpMADS1* may be regarded a relict of the ancestral land plant homologue. MpMADS1 forms a homodimeric DNA-binding complex, which is in contrast to the Arabidopsis proteins that are functional only as heterodimeric complexes. The *M. polymorpha* homodimer, nevertheless, recognizes the same DNA sequences as its angiosperm counterparts, and can functionally replace endogenous MIKC* complexes to a significant extent when heterologously expressed in Arabidopsis pollen. The 11 MIKC* homologues from the moss *F. hygrometrica* are highly and almost exclusively expressed in the gametophytic generation, suggesting that the gametophytic role of MIKC* MADS-box genes has been preserved throughout the land plant lineage. Taken together, these findings imply that MIKC* MADS-box proteins act as conserved components of land plant gametophyte developmental networks.

Introduction

Land plants (embryophytes) possess a life cycle that involves the alternation of multicellular sexual (haploid gametophyte) and asexual (diploid sporophyte) phases. All land plants are descended from a single aquatic ancestor that is most closely related to present-day streptophyte algae (1). They have inherited the multicellular gametophyte from this ancestor. The multicellular sporophyte, on the other hand, is a novelty that evolved during the transition to land (2). Among present-day land plants, bryophytes are thought to bear the strongest morphological resemblance to the earliest land plants (3). Bryophytes form a paraphyletic group that is composed of three monophyletic clades: liverworts, mosses and hornworts. Phylogenetic evidence indicates that the four major lineages of land plants have successively diverged in the following order: liverworts, mosses, hornworts, and vascular plants (4). All bryophytes develop an elaborate, dominating gametophyte and a simpler dependent sporophyte. This is in sharp contrast to vascular plants, which evolved increasingly complex sporophytes at the expense of gametophyte complexity, size and, ultimately, independence (5). In flowering plants, for instance, the gametophyte has been reduced to just a few cells in pollen grains and ovules, encapsulated in sporophytic tissue.

The development of the flowering plant Arabidopsis (*Arabidopsis thaliana*) critically depends on the action of different MADS-box transcription factors. These proteins form dimeric complexes that bind specific DNA sequences known as CArG motives (6, 7). The MADS-box gene family has diverged into two broad clades, termed type I and type II, which already existed before the divergence of plants, animals, and fungi (8). In land plants, type II MADS-box genes have further diverged into two groups, MIKC^c and MIKC* (9). The MIKC^c MADS-box genes of Arabidopsis regulate floral organ identity determination as well as several other aspects of sporophyte development (6). Only one MIKC^c MADS-box gene, *AGL18*, is expressed in the gametophyte (10-12), although the loss of *AGL18* function is apparently without effect (12, 13). The MIKC* MADS-box genes, on the other hand, are mainly expressed in the Arabidopsis male gametophyte (11, 14), with only one being expressed in embryonic tissue (15). The five gametophytic MIKC* MADS-box proteins function as heterodimeric complexes, with one binding partner being AGL66 or AGL104, and the other being AGL30, AGL65, or AGL94 (14). The incremental depletion of different complexes from pollen increasingly impairs *in vitro* pollen germination and *in vivo* pollen performance (12, 14, 16). Accordingly, MIKC* MADS-box transcription factors are major, partially redundant regulators of the male gametophytic developmental program.

The type II MADS-box genes of streptophytes (land plants + streptophyte algae) typically have a modular structure, consisting of a highly conserved MADS-box (M), a moderately conserved intervening domain (I), a well-conserved keratin-like domain (K), and a highly divergent C-terminal domain (C), hence the name MIKC. A characteristic difference between the genes of the MIKC^c and the MIKC* clade is the length of their I domain. While this it is rather short and encoded by a single exon in MIKC^c MADS-box genes, it is long and encoded by four to five exons in MIKC* MADS-box genes (9). The I, K, and C domains are not recognized in the type II MADS-box genes of either animals, fungi or chlorophyte algae (8). Accordingly, type II MADS-box proteins probably acquired a novel mode of functioning in the streptophyte lineage. The study of three different species of streptophyte algae revealed that each of them has only one type II MADS-box gene

(17). Because a phylogenetic relationship with neither the MIKC* clade nor the MIKC^c clade was well supported, the algal genes could be viewed as the ancestral MIKC homologues before their divergence into clades MIKC* and MIKC^c. Their expression in gametangial cells in both *Chara globularis* and *Closterium peracerosum-strigosum-littorale* further suggested that they play a role in gamete differentiation (17). Since the genes of the MIKC* clade are the only type II MADS-box genes that are predominantly expressed in the Arabidopsis gametophyte, it has been suggested that MIKC* MADS-box proteins retain the original type II MADS-box gene function from algae (11). However, little is known about MIKC* MADS-box genes in land plants other than Arabidopsis. As many as 11 genes are present in the moss *P. patens* (18), though only two of them have been analyzed up to now (19). Herein we report on the cloning and characterization of the MIKC* MADS-box genes from four bryophyte species: the mosses *Funaria hygrometrica*, *Atrichum undulatum*, and *Sphagnum subsecundum*; and the liverwort *Marchantia polymorpha*. Evidence is provided that the members of this clade of MADS-box genes have conserved functions in land plant gametophytes.

Results

Phylogeny and expression of MIKC* MADS-box genes in bryophytes

MIKC* MADS-box genes were isolated using a PCR based approach which exploits the high degree of sequence conservation within the MADS-box. The completed genome sequences of the flowering plants Arabidopsis and *Oryza sativa*, the lycophyte *Selaginella moellendorfii*, and the moss *P. patens* provided us with full gene complements from three disparate land plant lineages. Deduced amino acid sequences from all of the respective homologues were used to generate an alignment. The most highly conserved positions of the MADS-box were then identified, and the corresponding nucleic acid positions taken as a basis for primer design. Fragments of MADS-box sequences were amplified with a set of non-degenerate primers. Applying these at a low annealing temperature effectively permitted a large degree of sequence degeneracy, as indicated by *a posteriori* analysis of isolated sequences. Specificity was nevertheless preserved, since only MADS-box genes that belong to the MIKC* clade were isolated. Exhaustive PCR reactions were performed using both cDNA and genomic DNA as templates. In earlier research, 11 MIKC* MADS-box genes had been identified in the moss *Physcomitrella patens* (9, 18, 19). In this study, 17 new bryophyte homologues were cloned: one from the liverwort *Marchantia polymorpha* (*MpMADS1*), four from the basal moss *Sphagnum subsecundum* (*SsMADS1* through *SsMADS4*), one from the later-divergent moss *Atrichum undulatum* (*AuMADS1*), and 11 from a close relative of *P. patens*, the moss *Funaria hygrometrica* (*FhMADS1* through *FhMADS11*). Attempts to isolate homologues from the hornworts *Anthoceros agrestis* and *Anthoceros formosae* were unsuccessful, indicating that MIKC* MADS-box genes have either been lost in this lineage, or failed to be amplified with our primer set. Efforts to amplify sequences from the streptophyte algae *Chara spec.*, *Coleochaeta scutata*, *Klebsormidium nitens*, and *Mesostigma viride* were also unsuccessful. This is in line with the previous finding that streptophyte algae lack type II MADS-box genes of the MIKC* clade (17). Finally, complete cDNA sequences of each novel gene were obtained by 3'-RACE (rapid amplification of cDNA ends) and 5'-RACE.

In a phylogenetic reconstruction of aligned amino acid sequences, the bryophyte genes constitute a monophyletic clade that is sister to the Arabidopsis homologues (Figure 1). The liverwort gene *MpMADS1* is sister to all of the moss homologues. The moss clade further diverges into two monophyletic clades. The first of these contains all of the homologues from the basal moss *S. subsecundum* (order Sphagnales). The second clade harbors homologues from the later-divergent mosses *A. undulatum*, *F. hygrometrica*, and *P. patens*. Next, the Arabidopsis sequences were excluded from the alignment to allow more informative amino acid positions to be included in the phylogeny reconstruction (Figure S1). This improved the bootstrap support for the aforementioned relationships within the bryophyte lineage (Figure S3). For instance, there was significant support for an orthological relationship between the 11 homologues of *F. hygrometrica* and *P. patens*, which is in accordance with the fact that both species are members of the same family, namely the Funariaceae (order Funariales). The precise phylogenetic position of *AuMADS1* from *A. undulatum* (order Polytrichales) could not be resolved. A Southern blot analysis was then performed to probe the genomes of *F. hygrometrica*, *P. patens*, and *M. polymorpha*. In accordance with the high sequence identity levels between the *F. hygrometrica* and *P. patens* orthologues, an *FhMADS1* probe hybridized strongly with *P. patens'*

genomic DNA (Figure 2A). The same probe, albeit containing the highly conserved MADS-box, did not hybridize with *M. polymorpha's* genomic DNA. In the reverse experiment, *F. hygrometrica* and *P. patens'* genomic DNA also failed to bind an *MpMADS1* probe. Probing genomic DNA of male and female *M. polymorpha* with the *MpMADS1* probe produced a single band, which corresponds to the *MpMADS1* gene, even at low stringency (Figure 2B). The gene is, therefore, likely to be the only homologue. Finally, the genomic loci of all newly isolated genes were cloned and sequenced. Each homologue has the characteristic exon/intron structure of the MIKC* MADS-box gene clade, with a long I domain encoded by four to five exons (Figure S2) (9).

Real-time PCR expression analysis was performed to measure the transcript levels of the MIKC* MADS-box genes in the gametophyte and the sporophyte. The gametophytes of *P. patens* and *F. hygrometrica* are indistinguishable. However, unlike *P. patens*, the *F. hygrometrica* sporophyte grows a long seta, which facilitates the harvesting of sporophytic tissue that is free of any contaminating gametophytic material. For other species, the attempts to either harvest pure sporophytic tissue (*M. polymorpha*) or stimulate the induction of sexual organs (*S. subsecundum*) failed. Accordingly, *F. hygrometrica* was selected for a detailed expression analysis. The fact that complete transcripts could be amplified from the gametophytic cDNA of each species indicated that all homologues are, in any event, expressed in the gametophyte. Transcript levels were then determined for all *F. hygrometrica* MIKC* MADS-box genes in the following tissues: protonemata consisting of chloronema cells, caulonema cells and occasional gametophore buds (Figure 3A); gametophores (Figure 3B); young sporophytes (Figure 3C); and mature sporophytes (Figure 3D). Firstly, each gene was expressed, which suggests that each gene is functional. This is also supported by the fact that the ratio of nonsynonymous to synonymous nucleotide substitutions between individual *F. hygrometrica* and *P. patens* orthologues is always less than one. This proves that they are under the effect of purifying selection and, therefore, functional. Although the expression levels vary considerably between different genes, all of the homologues are strongly expressed in the gametophyte and only residually in the sporophyte (Figure 3E). The biggest difference in transcript abundance between both generations was observed for *FhMADS9* (3.500x higher in protonemata than in young sporophytes), and the smallest for *FhMADS11* (8x higher in gametophores than in mature sporophytes). Most genes were expressed at higher levels in protonemata than in gametophores, apart from *FhMADS1* and *FhMADS4*, which were expressed at similar levels in both stages, and *FhMADS10*, which was expressed more strongly in gametophores.

Functional analysis of *MpMADS1* from the liverwort *Marchantia polymorpha*

The Southern blot analysis indicated the existence of only a single *M. polymorpha* homologue. Since MADS-box proteins are known to bind DNA as dimers, whether the MpMADS1 protein is able to form homodimers was tested. This was firstly carried out in the yeast two-hybrid system. The growth of yeast cells which express a fusion between the GAL4 DNA-binding domain (BD) and the MpMADS1 protein in the absence of a GAL4 activation domain (AD)-containing binding partner implied that MpMADS1 has transcriptional activation potential. This property resides in the C-terminal half of the C domain, since it is lost when the C domain is truncated at position 303 (Figure 4A). Yeast cells which co-express the C-terminally truncated MpMADS1 protein that is fused to the GAL4 DNA-BD, and the GAL4 AD that is

fused to the complete MpMADS1 protein, are able to grow on selective dropout medium (Figure 4A). Thus, the homodimerization of the MpMADS1 protein occurs in yeast cells. Whether it also occurs *in planta* was tested in tobacco (*Nicotiana tabacum*) with the bimolecular fluorescence complementation (BiFC) system (20). The transient expression of the MpMADS1 protein, which was fused with either an N-terminal or a C-terminal YFP fragment, did not result in YFP fluorescence in tobacco leaf cells (Figure 4B-C). YFP fluorescence was, however, reconstituted when both fusion proteins were co-expressed, implying that a physical interaction occurs between the MpMADS1 proteins (Figure 4D). The subcellular localization of the fluorescent signal further testified that this interaction takes place in the nucleus, which confirms the protein's function as a transcription factor.

The MpMADS1 homodimer's DNA-binding ability and preference were investigated *in vitro* by random binding site selection. The protein was synthesized *in vitro*, and then used in an electrophoretic mobility shift assay (EMSA) to bind oligonucleotides from a randomized pool. After five iterative rounds of binding, a strong preference was apparent for the CArG motives of the so-called myocyte enhancer factor-2 (MEF2)-type, which have a strict CTA(A/T)$_4$TAG consensus sequence (7). The sequences CTA(TTTT)TAG and CTA(TATA)TAG together represented around 70% of all bound sequences (Figure 5A). Little more than 13% of bound sequences contain the related, but one nucleotide shorter, N9-type motif (14). The serum response element (SRE)-type CArG motif, CC(A/T)$_6$GG, which is the motif preferred by most of the MADS-box proteins tested to date (21), was not present in any of the bound oligonucleotides (Figure S4B). Competitive EMSA experiments were then performed to corroborate these findings. The binding of MpMADS1 to a labeled probe containing the MEF2-type motif CTA(TATA)TAG was strongly contested by an excess of a cold competitor probe containing the same motif or the MEF2-type motif CTA(TTTT)TAG (Figure 5B). On the other hand, an excess of a cold competitor probe containing the N9-type motif CTA(TTT)TAG could not compete as well, and the SRE-type motif CCTATTTAGG even less so. *In vitro*, the MpMADS1 homodimer thus exhibits DNA-binding properties that are very similar to those of the heterodimeric MIKC* MADS-box protein complexes of Arabidopsis, which have a preference for the same MEF2-type CArG motives (14).

Given that homologues from both plants have maintained the ability to recognize the same DNA sequences, whether the MpMADS1 protein can functionally replace its Arabidopsis counterparts was tested. To do so, *MpMADS1* was heterologously expressed in *agl66/agl104-1* double mutant pollen. This mutant's pollen has strongly impaired *in vitro* germination competence (Figure 6) (14). The *AGL65* promoter was used to drive the pollen-specific expression of either GFP, an AGL66:GFP fusion protein, or an MpMADS1:GFP fusion protein (Figure 6C-E). The localization of the MpMADS1:GFP fusion protein to the nucleus corroborated previous observations made with the BiFC system in tobacco leaf cells (Figure 6E). The expression of the AGL66:GFP fusion protein resulted in the full rescue of the pollen germination phenotype (Figure 6D, F). This indicated that the AGL66 protein, homologously expressed under the control of the *AGL65* promoter and C-terminally tagged with a GFP moiety, is fully functional. Pollen-expressed MpMADS1:GFP partially restored pollen tube formation, increasing germination efficiency from 4% to around 22%, compared to 40% in the wild type (Figure 6E, F). Therefore, the MpMADS1:GFP fusion protein must be capable of addressing at least a subset of target genes in pollen.

Discussion

The phylogenetic reconstructions presented here reveal that only a single MIKC* MADS-box gene was present before the divergence of the major bryophyte lineages. They also demonstrated that a single homologue has been retained in a member of the most ancient land plant lineage, the liverwort *M. polymorpha*. The gene clade then expanded independently in at least two moss lineages, namely the basal lineage leading to the peat moss *S. subsecundum* and the later divergent lineage leading to *F. hygrometrica* and *P. patens*. A total of 11 homologues are encoded in the genomes of *P. patens* and *F. hygrometrica*. Only six homologues are present in Arabidopsis (22). These numbers are in contrast to the general notion that transcription factor families expanded more in Arabidopsis than in *P. patens* (18). The MIKC^c MADS-box gene clade, for instance, comprises 39 homologues in Arabidopsis compared to just six in *P. patens* (18, 22). In fact, it is thought that the dramatic explosion of sporophytically expressed MIKC^c MADS-box genes in the angiosperm lineage played a crucial role in the origin of complex sporophytic morphologies, such as the flower (23). By analogy, the specific expansion of gametophytically expressed MIKC* MADS-box genes in mosses could have been interpreted as forming the genetic basis of a highly complex gametophytic body. However, our finding of only a single gene in the liverwort *M. polymorpha*, whose gametophyte body is arguably just as complex as that of mosses, refutes such a hypothesis.

The five pollen expressed MIKC* MADS-box proteins of Arabidopsis form five distinct heterodimeric complexes (14). This mode of obligatory heterodimerization probably evolved from homodimerization through gene duplication, subsequent sequence divergence, and the eventual loss of the ability to homodimerize. In the seed plant lineage, precisely such an evolutionary course has previously been traced for class B MIKC^c MADS-box proteins, which specify petal and stamen identity during flower development (24). In this case, the gaining of obligatory heterodimerization in angiosperms is thought to have provided a robustness against mutational and stochastic perturbations, and to have thereby played a role in the evolution of the flower (24, 25). Analysis of single, double and triple Arabidopsis MIKC* MADS-box gene mutants revealed that individual complexes redundantly regulate partially overlapping sets of genes, and that a transcriptional feedback mechanism compensates for the loss of one complex through the augmentation of the transcript levels of another (12). The MIKC* heterodimers of Arabidopsis, thus, also constitute a highly robust system. Conversely, gene duplication has not been exploited as a source of mutational robustness in the liverwort *M. polymorpha*. It is intriguing that as many as 11 MIKC* MADS-box genes have been retained in the mosses *P. patens* and *F. hygrometrica*. Unraveling their function is likely to reveal complex patterns of both conservation and divergence of gene expression, protein interaction, and DNA-binding specificity.

MIKC* MADS-box genes play a key role during the development of the Arabidopsis male gametophyte (12, 14, 16). In this study, the transcript levels of MIKC* MADS-box genes were determined in the moss *F. hygrometrica*. Each of 11 homologues turned out to be highly expressed in the gametophyte and only residually so in the sporophyte. Therefore, genes of this clade are likely to function specifically in the gametophytic generation of bryophytes, just as their counterparts do in Arabidopsis. It was further shown that the liverwort protein binds the same DNA sequences as its flowering plant homologues, and that MpMADS1 can functionally

replace – at least to a significant extent – the endogenous MIKC* complexes in the male gametophyte of Arabidopsis. Taken together, these findings suggest that MIKC* MADS-box proteins were important components of gametophyte regulatory networks during the early stages of land plant evolution, and that this function has been retained ever since. Interestingly, similar observations have been made for type II MADS-box genes in metazoans. The *myocyte enhancer factor-2* (*MEF2*) counterparts of flies and humans are functionally conserved regulators of muscle development, which are similarly able to bind the same DNA sequences and activate the same target genes (26). Type II MADS-box transcription factors have thus experienced conservative molecular evolution in both plants and metazoans.

A type II MADS-box gene probably plays a role in gamete formation in the closest living relatives of the aquatic ancestors of land plants, the streptophyte algae (17). This ancestral homologue was duplicated after the transition to land and diverged into the MIKC^c and MIKC* clades (9). All land plants inherited a multicellular, gamete-producing gametophytic generation from their aquatic ancestors, but a novelty of these plants is the development of a multicellular sporophyte body (2). Comparative studies between *P. patens* and Arabidopsis have indicated that the consecutive evolution of this new structure involved the modification of sporophyte-specific developmental programs (27, 28), as well as a recruitment of gametophytic developmental programs into the sporophytic generation (29, 30). MIKC^c MADS-box genes seem to have undergone the latter fate because they are expressed in both the gametophyte and the sporophyte in mosses and ferns (31-33), and almost exclusively in the sporophyte in Arabidopsis (11). On the other hand, our data suggest that MIKC* MADS-box genes have retained the ancestral gametophytic function.

Materials and methods

Plant material and culture conditions. *F. hygrometrica* spores were collected from a local garden center in Cologne, Germany. The standard *M. polymorpha* strains Takaragaike-1 (♂) and Takaragaike-2 (♀) have been used in this study (34). *S. subsecundum* and *A. undulatum* were collected at the Botanical Garden in Dresden, Germany. Plant culture conditions are described in SI Materials and methods.

Gene isolation, sequence analysis and blotting techniques. The isolation of genomic DNA and the PCR based isolation of MIKC* MADS-box genes are described in SI Materials and methods. *P. patens* and Arabidopsis MIKC* MADS-box gene sequences were collected from DNA databases (see Table S2 for accession numbers), and aligned with newly isolated gene sequences using ClustalW in the MacVector 9.0 software (Accelrys). Alignments were manually adjusted. Calculations and bootstrapping (1000 replicates) of phylogenetic trees were performed with the neighbor-joining method using the programs PROTDIST and NEIGHBOR, the parsimony method using the program PROTPARS, and the maximum likelihood method using the program PROML (JTT probability model; jumble 10x), all of which are included in the Phylip 3.68 software package (35). Blotting techniques are described in SI Materials and methods.

Real-time PCR. Total RNA was isolated from *F. hygrometrica* protonemata, gametophores, young sporophytes, and mature sporophytes with the RNeasy Plant Mini Kit (Qiagen), followed by a cDNA synthesis with random hexamers. Real-time PCR experiments were performed on a Bio-Rad iCycler using the manufacturer's SYBR-Green I chemistry. The data were subsequently analyzed with Bio-Rad's iQ5 Optical System Software. The primers used for the analysis are described in SI Materials and methods. The quality of sporophytic cDNA preparations was confirmed by amplification of 18S rDNA and *RAN* (*Ras-related nuclear protein*) transcripts. Each real-time PCR reaction was run in triplicate, and the resulting mean Ct values were used to calculate the mean normalized expression (MNE) for the 18S reference gene (Ref) and the gene of interest (GOI) using the formula $E(Ref)^{Ct(Ref)}/E(GOI)^{Ct(GOI)}$, where E is the primer efficiency and Ct the threshold cycle. Two biological replicates were conducted, and mean MNE values and standard errors were calculated. The highest mean MNE value was set to 100 and all other values were scaled relative to it.

Yeast two-hybrid analysis. Details of the construction of vectors are given in SI Materials and methods. All combinations of pGADT7 and pGBKT7 vectors, i.e. empty vectors and vectors encoding full-length or truncated MpMADS1 proteins, were co-transformed in the yeast (*Saccharomyces cerevisiae*) strain AH109 (Clontech Laboratories) according to the manufacturer's lithium acetate-mediated method, and plated on a synthetic dropout (SD) medium lacking Leu and Trp to select for co-transformants. Three independent co-transformants were grown to similar densities in a liquid Yeast Peptone Dextrose (YPD) medium and plated onto SD medium lacking Leu, Trp, Ade, and His to select for protein-protein interactions.

Bimolecular fluorescence complementation (BiFC). The primers z330/z335 (Table S1) were used to amplify the complete *MpMADS1* coding sequence from *M. polymoprha* cDNA, followed by Gateway-based cloning (Invitrogen), first into the

entry vector pENTR/D-TOPO (Invitrogen), and then into the destination vectors pBaTL-YFPc and pBaTL-YFPn (36). Transformation into *Agrobacterium tumefaciens* strain ABI (37) and consecutive infiltration into *Nicotiana benthamiana* leaves were as described (14). After three days, infiltrated leaves were examined for fluorescence using a Zeiss Axiophot microscope equipped with a YFP filter [500/20 (exciter) / 515LP (beamsplitter) / 535/30 (emitter)].

Random binding site selection (RBSS) and competitive electrophoretic mobility shift assay (EMSA). The RBSS procedure is described in detail in SI Materials and methods. After binding site enrichment through five consecutive rounds of RBSS, the recovered fragments were subjected to eight PCR cycles, and then cloned into the pGEM-T easy vector (Promega). Eighty-nine clones were sequenced in total. To confirm the RBSS results, a competitive EMSA experiment was designed as follows. Oligo z389, containing one of the two preferentially bound MEF2-type motives, was radioactively labeled by PCR with primers z289/z290 (Table S1) as described (14). Unlabeled, double-stranded competitor probes were obtained through boiling and gradual cooling of the complementary oligo pairs z389/z390, z391/z392, z393/z394, and z395/z396 (Table S1). The cold competitor probes were used at 5- and 50-fold molar excesses relative to the labeled probe. Competitive protein-DNA binding reactions were commenced by incubating the protein on ice with an unlabeled competitor probe for 15 minutes, prior to the addition of the labeled probe, followed by another 15 minutes of incubation on ice. Subsequent electrophoretic separation as well as visualization of bands was conducted as previously.

Arabidopsis complementation and *in vitro* pollen germination assays. The *agl66/agl104* double mutant (14) that was used in these experiments is referred to as *agl66/agl104-1*, after the recent report of a second *AGL104* mutant allele termed *agl104-2* (16). Details about the construction and transformation of vectors are given in SI Materials and methods. For each construct, two independent homozygous T3 transformants were selected which carry T-DNA insertions that segregate as a single locus, as indicated by the segregation of barnase resistance and the pollen GFP fluorescence in the T2 generation. *In vitro* germination of pollen harvested from 8-10 flowers was performed according to an optimized protocol (38). Pollen germination was documented with a Leica MZFLIII stereomicroscope. For quantification, at least 400 pollen grains on two different microscopy slides were scored for each line. The experiment was conducted in duplicate and mean values and standard errors were calculated. GFP fluorescence was observed using a Zeiss Axiophot microscope equipped with a GFP filter [480/20 (exciter) / 495LP (beamsplitter) / 510/20 (emitter)].

Acknowledgments

We thank Wim Verelst and Michiel Kwantes for generously providing us with the pGJ2364-AGL65::GFP and pGJ2364-AGL65::AGL66:GFP constructs, and Daniela Liebsch for providing *P. patens* gene sequences and 18S primers. Sabina Wodniok and Burkhard Becker are gratefully acknowledged for donating nucleic acids of streptophyte algae, and Ute Wahrmund, Maike Petersen, and Katsuyuki Yamato for hornwort and liverwort material. We also want to thank Dietmar Quandt and Jan-Peter Frahm for *Sphagnum subsecundum* species determination. O.Z. expresses his gratitude to Sabine Zachgo and Wim Verelst for their valuable comments on the manuscript. This work was supported by the Max Planck Society and by a grant from the Collaborative Research Centre SFB680 of the Deutsche Forschungsgemeinschaft (DFG) (to T.M.).

References

1. Graham LE (1993) *Origin of land plants* (John Wiley & Sons, Inc.).
2. Kenrick P & Crane PR (1997) The origin and early evolution of plants on land. *Nature* 389:33-39.
3. Gray J (1985) The microfossil record of early land plants - Advances in understanding of early terrestrialization. *Philos Trans R Soc Lond B Biol Sci* 309:167-195.
4. Qiu YL*, et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103:15511-15516.
5. Graham LE, Cook ME, & Busse JS (2000) The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci USA* 97:4535-4540.
6. Becker A & Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29:464-489.
7. Shore P & Sharrocks AD (1995) The MADS-xox family of transcription factors. *Eur J Biochem* 229:1-13.
8. Alvarez-Buylla ER*, et al.* (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* 97:5328-5333.
9. Henschel K*, et al.* (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. *Mol Biol Evol* 19:801-814.
10. Alvarez-Buylla ER*, et al.* (2000) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* 24:457-466.
11. Kofuji R*, et al.* (2003) Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* 20:1963-1977.
12. Verelst W*, et al.* (2007) MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology* 8:R249.
13. Adamczyk BJ, Lehti-Shiu MD, & Fernandez DE (2007) The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis. *Plant J* 50:1007-1019.
14. Verelst W, Saedler H, & Munster T (2007) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiol* 143:447-460.
15. Lehti-Shiu MD, Adamczyk BJ, & Fernandez DE (2005) Expression of MADS-box genes during the embryonic phase in Arabidopsis. *Plant Mol Biol* 58:89-107.
16. Adamczyk BJ & Fernandez DE (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol* 149:1713-1723.
17. Tanabe Y*, et al.* (2005) Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proc Natl Acad Sci USA* 102:2436-2441.
18. Rensing SA*, et al.* (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319:64-69.

19. Riese M, *et al.* (2005) Isolation and characterization of new MIKC*-type MADS-box genes from the moss Physcomitrella patens. *Plant Biol (Stuttg)* 7:307-314.
20. Hu CD, Chinenov Y, & Kerppola TK (2002) Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* 9:789-798.
21. Ito Y, *et al.* (2008) DNA-binding specificity, transcriptional activation potential, and the rin mutation effect for the tomato fruit-ripening regulator RIN. *Plant J* 55:212-223.
22. Parenicova L, *et al.* (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. *Plant Cell* 15:1538-1551.
23. Irish VF (2006) Duplication, diversification, and comparative genetics of angiosperm MADS-box genes. *Adv Bot Res* 44:129-161.
24. Winter KU, *et al.* (2002) Evolution of class B floral homeotic proteins: Obligate heterodimerization originated from homodimerization. *Mol Biol Evol* 19:587-596.
25. Lenser T, Theissen G, & Dittrich P (2009) Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Comput Biol* 5:7.
26. Black BL & Olson EN (1998) Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol* 14:167-196.
27. Maizel A, *et al.* (2005) The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308:260-263.
28. Tanahashi T, Sumikawa N, Kato M, & Hasebe M (2005) Diversification of genie function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens. *Development* 132:1727-1736.
29. Menand B, *et al.* (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316:1477-1480.
30. Nishiyama T, *et al.* (2003) Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: Implication for land plant evolution. *Proc Natl Acad Sci USA* 100:8007-8012.
31. Hasebe M, Wen CK, Kato M, & Banks JA (1998) Characterization of MADS homeotic genes in the fern Ceratopteris richardii. *Proc Natl Acad Sci USA* 95:6222-6227.
32. Munster T, *et al.* (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc Natl Acad Sci USA* 94:2415-2420.
33. Quodt V, Faigl W, Saedler H, & Munster T (2007) The MADS-domain protein PPM2 preferentially occurs in gametangia and sporophytes of the moss Physcomitrella patens. *Gene* 400:25-34.
34. Ishizaki K, Chiyoda S, Yamato KT, & Kohchi T (2008) Agrobacterium-mediated transformation of the haploid liverwort Marchantia polymorpha L., an emerging model for plant biology. *Plant Cell Physiol* 49:1084-1091.
35. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164-166.
36. Hackbusch J, Richter K, Muller J, Salamini F, & Uhrig JF (2005) A central role of Arabidopsis thaliana ovate family proteins in networking and

subcellular localization of 3-aa loop extension homeodomain proteins. *Proc Natl Acad Sci USA* 102:4908-4912.

37. Koncz C & Schell J (1986) The promoter of Tl-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of Agrobacterium binary vector. *Mol Gen Genet* 204:383-396.
38. Boavida LC & McCormick S (2007) Temperature as a determinant factor for increased and reproducible in vitro pollen germination in Arabidopsis thaliana. *Plant J* 52:570-582.

Figure legends

Figure 1: Neighbor-joining phylogenetic tree of MIKC* MADS-box genes from bryophytes and Arabidopsis, using 147 amino acids from the MADS, I, and K domains as sequence data. For the major nodes, the bootstrap values obtained from 1000 replicates with the neighbor-joining and the parsimony methods, respectively, are indicated. Scale bar = 0.1 substitutions per site.

Figure 2: Southern blot analysis. Genomic DNA of *Marchantia polymorpha* (♂ and ♀), *Funaria hygrometrica*, and *Physcomitrella patens* was digested with the enzymes *ClaI* (C), *NcoI* (N), *BglII* (B), and *EcoRI* (E), and the blot was subsequently hybridized at a reduced stringency with an *FhMADS1* probe (*A*) and an *MpMADS1* probe (*B*).

Figure 3: Expression of *Funaria hygrometrica* MIKC* MADS-box genes. (*A-D*) Successive stages of the life cycle of *F. hygrometrica* that were used for RNA sampling: protonema (*A*), gametophore (*B*), young sporophyte (*C*), and mature sporophyte (*D*). The white arrowhead, black arrowhead, and asterisk in A indicate a chloronema cell, a caulonema cell, and a gametophore bud, respectively. (*E*) Transcript levels of *FhMADS1* through *FhMADS11* in the tissues illustrated in A-D, as determined by real-time PCR quantification. The highest signal value is set to 100, and all other values are calculated relative to that. The values represented in the graph are listed in the table below. Scale bars = 0.2 mm (A) and 2 mm (B-D).

Figure 4: MpMADS1 homodimerization in yeast and *in planta*. (*A*) MpMADS1 homodimer formation tested by yeast-two hybrid analysis. Yeast cells co-expressing MpMADS1 that was fused to the GAL4 DNA-binding domain (BD) and MpMADS1 that was fused to the GAL4 activation domain (AD) were analyzed for growth on SD/-Leu/-Trp/-His/-Ade medium. To abolish the inherent transcriptional activation potential of MpMADS1, a C-terminally truncated version of MpMADS1 (MpMADS1ΔC) that was fused to the GAL4 DNA-BD was utilized. (*B-D*) MpMADS1 homodimer formation tested by bimolecular fluorescence complementation (BiFC). YFP fluorescence was analyzed in tobacco leaf cells transiently expressing MpMADS1:YFPn (*B*), MpMADS1:YFPc (*C*), or MpMADS1:YFPn and MpMADS1:YFPc together (*D*). Scale bars = 50 μm.

Figure 5: MpMADS1 DNA-binding specificity. (*A*) Summary of CArG motives found in 89 oligonucleotides isolated after five rounds of random binding site selection with the MpMADS1 protein. (*B*) DNA-binding specificity of MpMADS1 tested by competitive EMSA. Lane 1, no protein; lanes 2 to 10, MpMADS1. The binding of MpMADS1 to a labeled probe containing the CTA(TATA)TAG motif was contested by the addition of none (lane 2), 5x (lanes 3, 5, 7 and 9), and 50x (lanes 4, 6, 8 and 10) molar excesses of a cold competitor probe containing the CTA(TATA)TAG motif (lanes 3 and 4), the CTA(TTTT)TAG motif (lanes 5 and 6), the CTA(TTT)TAG motif (lanes 7 and 8), or the CCTATTTAGG motif (lanes 9 and 10).

Figure 6: Complementation of the Arabidopsis *agl66/agl104-1* mutant with MpMADS1. (*A-E*) *In vitro* germination assay of pollen grains from wild type (*A*), *agl66/104-1* (*B*), *agl66/104-1 AGL65::GFP* (*C*), *agl66/104-1 AGL65::AGL66:GFP* (*D*), and *agl66/104-1 AGL65::MpMADS1:GFP* (*E*) anthers of Arabidopsis. White arrowheads indicate pollen tubes. Insets in C-E illustrate pollen-specific fluorescence of GFP and GFP fusion proteins. (*F*) Quantification of *in vitro* pollen germination efficiency in each of the genetic backgrounds. Two independent transformants (#1 and #2) are shown for each complementation construct. Scale bars = 200 μm (A-E) and 2 μm (insets in C-E).

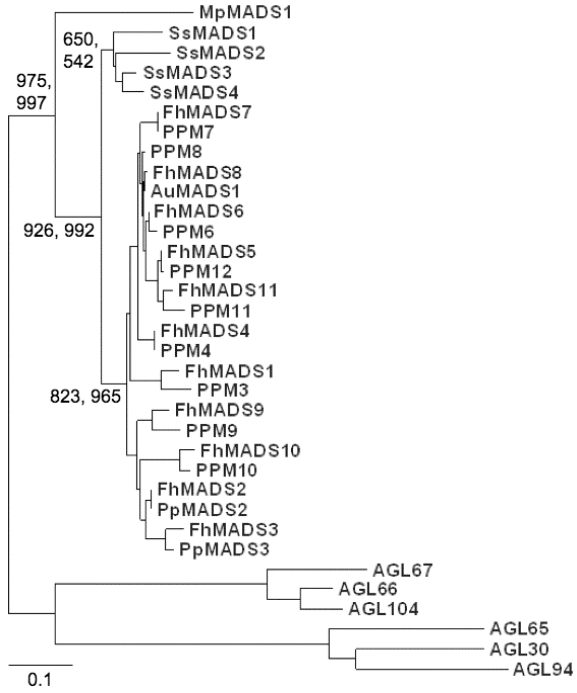


Figure 1

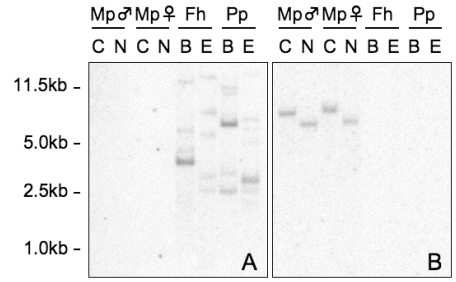
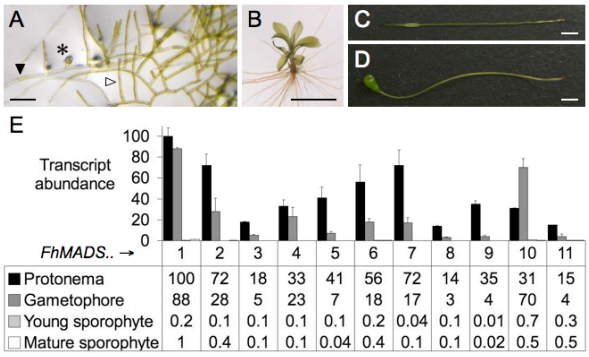


Figure 2



FhMADS.. →	1	2	3	4	5	6	7	8	9	10	11
Protonema	100	72	18	33	41	56	72	14	35	31	15
Gametophore	88	28	5	23	7	18	17	3	4	70	4
Young sporophyte	0.2	0.1	0.1	0.1	0.1	0.2	0.04	0.1	0.01	0.7	0.3
Mature sporophyte	1	0.4	0.1	0.1	0.04	0.4	0.1	0.1	0.02	0.5	0.5

Figure 3

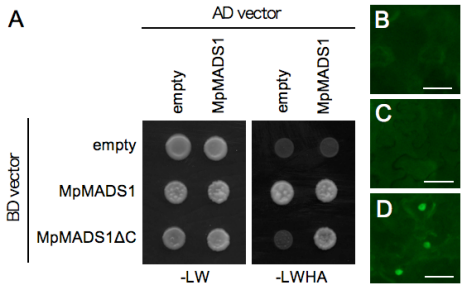


Figure 4

A	motif type	sequence (5'-3')	frequency (n=89)
	MEF2	CTA(TTTT)TAG	38
		CTA(TATA)TAG	25
		other CTA(WWWW)TAG	11
	MEF2 w/ one G:C bp	CTN(NNNN)TAG	3
	N9	CTA(TTT)TAG	4
		other CWW(TTT)WWG	4
	N9 w/ one G:C bp	CTA(TTN)NNG	4

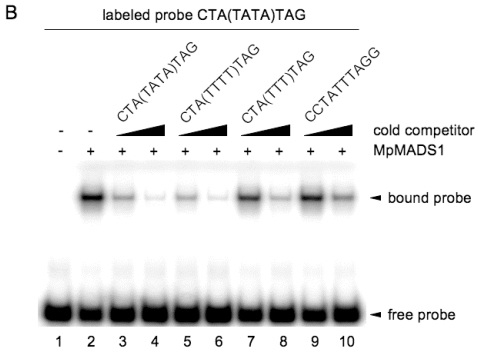


Figure 5

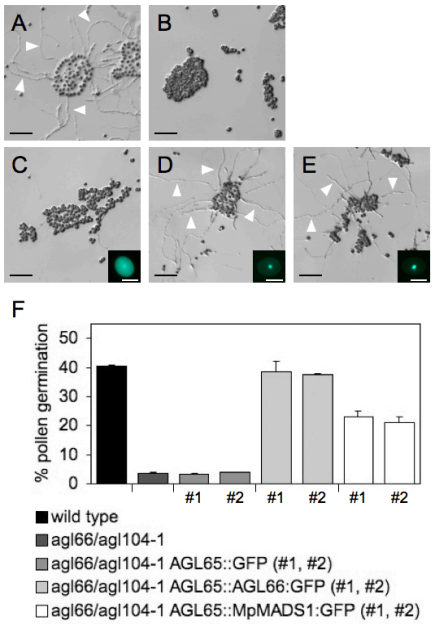


Figure 6