Molecular analyses of *Physcomitrella patens* MIKC^{c}-type MADS-box genes establish an evolutionary context in plants

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Vanessa Quodt

aus Köln

2006

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln durchgeführt.

Berichterstatter: Prof. Dr. Heinz Saedler
Prof. Dr. Diethard Tautz

Tag der mündlichen Prüfung: 14.02.2007

Table of contents

1 Introduction

1.1 *MADS-box genes and the evo-devo concept*

The major patterns of the evolution of land plants have been revealed through comparative studies on morphology and phylogenetic relationships, however, the molecular mechanisms responsible for the enormous diversity of land plant body plans are still a promising focus of evolutionary biology. The basic principle of evolutionary developmental biology ("evo-devo") is that genes controlling a developmental process also play a crucial role in its evolution (Gilbert et al., 1996; Vergara-Silva et al., 2000). Every living multicellular organism is generated anew from a single cell in each generation, driven by developmental processes. Thus, evolutionary changes in these processes are closely related to the morphological evolution of multicellular organisms. Based on the evo-devo concept, the genetic control of developmental mechanisms has been a major research focus in many different model organisms (Theißen et al., 2000; Munoz-Chapuli et al., 2005; Joron et al., 2006), because analyzing phylogeny and function of developmental control genes may help to understand the evolution of morphology.

Many developmental control genes are members of multigene families and encode transcription factors that play critical roles at key nodes of gene networks, regulating cellular processes. A very well-known family of transcription factors is the family of MADS-box genes that has been identified in a diverse range of eukaryotic organisms including yeasts, plants, and mammals (reviewed in: Messenguy and Dubois, 2003). Their name represents an acronym of the four founding members *MINI-CHROMOSOME MAINTENANCE 1* (*MCM1*) in yeast, *AGAMOUS* (*AG)* in Arabidopsis, *DEFICIENS* (*DEF)* in Antirrhinum and Serum Response Factor (*SRF)* in humans (Schwarz-Sommer et al., 1990). Thus, at least one MADS-box gene existed before the divergence of plants and animals (Alvarez-Buylla et al., 2000b).

All MADS domain proteins share a highly conserved 56 amino acid motif, the MADS domain, which is responsible for DNA binding and dimerization (Shore and Sharrocks,

1995). Outside the DNA-binding domain, the proteins are rather diverse (Alvarez-Buylla et al., 2000b).

MADS domain proteins regulate expression of target genes by binding to CArG-box motifs in their promoter regions (Egea-Cortines et al., 1999; Honma and Goto, 2001; Theißen and Saedler, 2001). Most MADS domain proteins prefer the so-called Serum Response Element (SRE-) type CArG-box with a consensus of $CC(A/T)_6GG$ (Hayes et al., 1988; Riechmann et al., 1996).

Based on their sequence similarity within the MADS domain, eukaryotic non-plant MADS-box genes have been subdivided into two groups, Serum Response Factor (SRF-) type and Myocyte Enhancer Factor-2 (MEF2-) type with various regulatory functions (Theißen et al., 1996; reviewed in: Theißen et al., 2000). By contrast, the majority of the described plant MADS domain proteins belong to the MIKC-type that is characterized by a conserved structural organization. Downstream of the highly conserved MADS domain they feature an intervening (I-) domain, responsible for specification of dimerization (Fan et al., 1997), followed by a conserved keratin-like (K-) domain that promotes protein dimerization (Riechmann and Meyerowitz, 1997; Kaufmann et al., 2005), and a C-terminal domain that contributes to the formation of multimeric complexes (Egea-Cortines et al., 1999; Honma and Goto, 2001). Some MIKC proteins possess an additional N-terminal extension (Theißen et al., 1996).

The remaining plant MADS-box genes are of the M-type and lack the plant-specific K-domain. Their functions are unknown. Alvarez-Buylla et al. (2000b) divided SRF-, MEF2-, M- and MIKC-type MADS-box genes into two groups, type I and II, according to a phylogenetic analysis of the MADS domain. According to this, type I includes animal and fungal SRF-type genes as well as a group of largely uncharacterized genes, among them all plant M-genes. Type II consists of MEF2-type genes and the MIKC-type genes that are unique to plants. Parenicova et al. (2003) suggested a further subdivision of all plant M-type MADS-box genes into four groups, termed Mα-Mδ, while Mδ-type genes consist exclusively of MIKC*-type genes (Kofuji et al., 2003). Phylogenetic analyses revealed that, due to a remarkable similarity between the $MIKC^c$-

("classic") and the MIKC*-type genes, the latter belong within the type II group instead of type I.

MIKC*-type genes differ from the MIKC^c-type in the length of the I-domain and composition of the K-domain (Henschel et al., 2002). So far, MIKC*-type genes have been identified in many different species, including Physcomitrella (Henschel et al., 2002; Riese et al., 2004), lycopods, ferns (Kwantes, unpublished results), monocotyledons (Nam et al., 2004) and eudicotyledons such as Arabidopsis (Parenicova et al., 2003; Verelst et al., in press). Their functions are largely unknown. Contrary to MIKC^c-type genes, their expression is primarily restricted to the gametophyte among all identified species. Interestingly, the only MIKC MADS-box gene that was described for each of three Charophycean green algae, representatives of the closest living relatives of land plants (Karol et al., 2001), was of the classic type (Tanabe et al., 2005). The MIKC^c-type genes are therefore most likely ancestral to the MIKC*-type genes. Alternatively, the MIKC*-type may have been lost in the lineage leading to extant Charophyceae.

Among MIKC^c-type genes, a further phylogenetic subdivision into defined clades is possible. Most members of one clade share similar expression patterns and highly related functions (Theißen et al., 1996; Theißen et al., 2000). The best-characterized MIKC^c-type MADS-box genes are those that function as floral organ identity genes in angiosperms, known as the "ABC model" (Schwarz-Sommer et al., 1990; Weigel and Meyerowitz, 1994). These genes encode transcription factors that activate or repress target genes in their distinct expression domains, thereby defining the whorls of the angiosperm flower. Orthologs of some of the floral organ identity genes have also been identified in monocotyledons (Kang et al., 1995; Benedito et al., 2004) and gymnosperms (Shindo et al., 1999; Winter et al., 2002; Becker et al., 2003), but not in ferns (Münster et al., 1997; Theißen et al., 2000). However, MADS-box gene functions are not restricted to reproductive structures. In Arabidopsis, for example, they play many different roles, including regulatory functions in vegetative tissues (Rounsley et al., 1995; Alvarez-Buylla et al., 2000a; Aswath and Kim, 2005).

Knowledge about MIKC^c-type genes in non-seed plants is limited. Some genes have been isolated from green algae, the moss Physcomitrella, lycopods and ferns (Münster et al., 1997; Hasebe et al., 1998; Krogan and Ashton, 2000; Henschel et al., 2002; Hohe et al., 2002; Svensson and Engström, 2002; Münster et al., 2002; Tanabe et al., 2003; Tanabe et al., 2005). So far, gene functions are unknown, and the available data do not produce a well-resolved phylogenetic topology concerning the deep branching of the subfamilies. However, gaining knowledge about expression and functionality of MIKC^c-type genes in basal land plants will contribute largely to understanding the evolution of land plants. So far, the origin of the plant-specific MIKC domain structure has been revealed based on the analyses of freshwater green algae. One MIKC^c-type MADS-box gene has been identified in each of three species of Charophycean green algae (Tanabe et al., 2005), but not in the analyzed red alga *Cyanidioschyzon merolae* (Matsuzaki et al., 2004) or the recently sequenced genome of *Chlamydomonas reinhardtii*. Thus, MIKC^c-type MADS-box genes originated in the last common ancestor of Charophyceae and all land plants (Graham et al., 2000; Tanabe et al., 2005), more than 470 million years ago (Kenrick and Crane, 1997).

1.2 *The bryophyte* **Phycomitrella patens** *as a model plant for evolutionary questions*

It is widely accepted that bryophyte-like organisms were the first plants that colonized land more than 450 million years ago (Kenrick and Crane, 1997; Purugganan, 1997). Therefore, recent bryophytes represent ancestral systems and are thus key to understanding the evolution of land plants. However, phylogenetic relationships among extant bryophytes, i.e. liverworts, hornworts and mosses, are unclear (Nickrent et al., 2000) - while the most basal position within land plant phylogeny is often attributed to liverworts (Qiu et al., 1998; Qiu et al., 2006), some studies present evidence to place hornworts as the first branch of the land plant tree (Kenrick and Crane, 1997; Nickrent et al., 2000). Furthermore, cladistic analyses of bryophyte chloroplast genes strongly

suggest a monophyletic origin of all bryophytes (Nishiyama et al., 2004). To resolve relationships among non-vascular plants thoroughly, further extensive phylogenetic analyses are needed.

While hornworts and liverworts exhibit a simple thalloid morphology, leafy moss plants resemble the structure of kormophytes. They possess root-like structures, called rhizoids, simple stems that lack highly developed conducting systems, and leaf-like structures that consist of only one cell layer and do not possess the sophisticated morphology of a higher plant leaf.

Physcomitrella patens (Hedw.) B. S. G., a prominent representative of mosses (musci), has been utilized exceedingly as a model system for basal land plants, contributing considerably to understanding the evolution of plant development processes (Cove and Knight, 1993; Reski, 1998; Cove et al., 2006). It constitutes a relatively small genome of approximately 511 Mbp of DNA (Schween et al., 2003), distributed among 27 chromosomes, that has recently been sequenced by an international collaboration involving several laboratories worldwide and is being annotated at present. So far, Physcomitrella is the only plant that features highly efficient allelic replacement by homologous recombination (Schaefer and Zryd, 1997; Schaefer, 2001; Schaefer, 2002; Kamisugi et al., 2005; Kamisugi et al., 2006). This mechanism provides an extremely useful tool for molecular analyses, especially since the dominant generation of the moss life cycle is the haploid gametophyte (Schaefer and Zryd, 1997). Thus, a potential phenotype caused by the replacement of a wild type gene locus will manifest immediately.

A characteristic unique to plants is the alternation between a haploid and a diploid generation. In mosses, the alternation of generations is heteromorphic, constituting a dominant haploid phase, the gametophyte, and a short-lived sporophyte that depends on nutrient supply by the gametophyte (Bold, 1940). The haploid phase is initiated with a meiospore that germinates and produces a filamentous structure called protonema. The leafy moss plant, termed gametophore, develops from a bud. On each gametophore apex, sexual organs develop from apical cell divisions (Lal and Bhandari, 1968). Physcomitrella is a monoecious moss and thus develops both male and female

gametangia on each gametophore, induced by autumnal conditions. Each apex carries several organs of each type, arranged in bundles. The male organs, called antheridia, are club-shaped and contain the male gametes (spermatozoids), surrounded by a single cell layer. In the female reproductive organs, the flask-shaped archegonia, an egg cell awaits fertilization. When the egg cell is mature, the cells filling the archegonium neck tube disintegrate to form a mucous material that will facilitate the entry of spermatozoids. Fertilization in mosses is favored in a moist environment, when mature spermatozoids swim to an archegonium in a continuous film of water and fertilize the egg cell. The zygote constitutes the first diploid cell within the life cycle of a moss and gives rise to the sporophyte that consists of a foot (haustorium), a seta, and a sporangium. The Physcomitrella sporophyte possesses a relatively short seta and lacks specialized structures for dehiscense, releasing spores by disintegration of the sporangium wall. At the interface between the haploid and the diploid generation of liverworts and mosses, the foot of the sporophyte is embedded within the vaginula of the gametophyte, but no direct connection between the cells exists (Frahm, 2001). Nutrients are transported through the apoplastic placental gap from the haploid gametophore to the diploid generation to sustain its proper development. In *Funaria hygrometrica*, for instance, transfer cells of the haustorium play a key role in sugar absorption by the sporophyte (Browning and Gunning, 1979a; 1979b). In the mature sporangium, finally, meiosis takes place and produces haploid spores.

Taking all described features into consideration, the moss *Physcomitrella patens* constitutes an ideal model organism for the investigation of basal land plant evolution.

1.3 *MIKC^c-type MADS-box genes in* Physcomitrella patens

In *Physcomitrella patens*, four MIKC^c-type MADS-box genes have been characterized (Krogan and Ashton, 2000; Henschel et al., 2002; Hohe et al., 2002). Furthermore, the recently completed genome sequencing revealed the existence of two additional genes of the c-type (In this work, a new nomenclature will be applied; see results).

Phylogenetic analysis by the neighbor-joining method suggests that *PPMC1*, *PPMC2* and *PPMC3* are more closely related to each other than to *PPMC4* (Hohe et al., 2002). However, the moss genes could not be assigned non-ambiguously to any of the well-defined MIKC-type subfamilies known from seed plants (Henschel et al., 2002). Orthology relationships between the Physcomitrella MADS-box genes and those from seed plants are thus still unclear, whereas orthologous genes have been reported between gymnosperms and angiosperms (Winter et al., 1999; Winter et al., 2002).

Structural analysis of *PPMC2* has revealed three alternative splice forms of the corresponding mRNA (Krogan and Ashton, 2000; Henschel et al., 2002). Splice variant "a" results from 11 exons, however, variant "b" retains intron 2 (Krogan and Ashton, 2000). As a result, the mRNA exhibits a stop codon immediately downstream of the MADS-domain and thus most likely leads to a non-functional peptide. Splice form "c" of *PPMC2* differs from variant "a" only in the 3' UTR, where a portion of intron 10 is retained, however, exon 11 is missing (Henschel et al., 2002).

Furthermore, a 64 amino acid putative N-terminal domain has been reported for *PPMC2* (Krogan and Ashton, 2000). A comparable N-terminal extension of the MADS-box has so far only been reported for some *AG*-like and *CRM6*-like genes (Theißen et al., 1996; Hasebe et al., 1998). An additional upstream AUG in the putative 5' UTR of *PPMC2*, leading to a small open reading frame that overlaps the possible N-terminus, has also been described by Krogan and Ashton (2000).

RT-PCR of the three genes *PPMC1*, *2* and *3* reveals that expression is found during all developmental stages of the moss life cycle (Faigl and Münster, unpublished results). All three genes possess a broad expression pattern including protonema, gametophore and sporophyte tissues. Comparably, *PPMC4* has been shown to be expressed in protonema and gametophores (Hohe et al., 2002).

1.4 *Goal of this thesis*

The goal of this thesis was to contribute to understanding the evolution of MADS domain transcription factors in non-seed plants, and, thereby, provide insight concerning the evolution of land plant morphology. For this purpose, function and expression domains of *Physcomitrella patens* MIKC classic type MADS-box genes *PPMC1* and *PPMC2* were analyzed. Protein localization was revealed via translational reporter gene fusions, and gene disruption lines were produced to allow conclusions regarding functionality. An additional investigation of a putative regulatory function of the highly complex and unusually long 5' UTR of the gene *PPMC2* was performed to reveal the level of gene regulation. Based on reporter gene fusion data and gene knock-out lines, putative functions for *PPMC1* and *PPMC2* will be discussed. Finally, the results will be embedded into an evolutionary context.

2 Material and Methods

2.1 *Sources of supply for commercially available materials*

The chemical products used for the preparation of buffers, media and solutions were drawn from Biozym (Rockland, USA), Invitrogen (Karlsruhe), Merck (Darmstadt), Promega (Heidelberg), Qiagen (Hilden), Roche (Mannheim), Roth (Karlsruhe), Serva (Heidelberg) and Sigma (München).

The enzymes were purchased from KMF Laborchemie Handels GmbH (Lohmar), New England Biolabs (Schwalmbach), Takara Bio Inc. (Otsu, Japan), Roche (Mannheim) and Invitrogen (Karlsruhe). All enzymatic reactions were performed according to the manufacturer's instructions.

The radioisotopes [^{35}S]-Methionine (>37 TBq/mmol) and [γ^{32}P]-dATP (30 Tbq/mmol) were obtained from Hartmann-Analytik (Braunschweig).

The commercial kits used for clean-up procedures of nucleic acids, isolation of plasmid DNA and 5'RACE were obtained from Qiagen (Hilden) and Roche (Karlsruhe).

Molecular weight markers for nucleic acid and protein gel electrophoresis were obtained from Eurogentec Deutschland GmbH (Köln) and New England Biolabs (Frankfurt a. M.).

Sigma-Genosys (Steinheim), Invitrogen (Karlsruhe), Operon (Köln), MWG-Biotech AG (Ebersberg) and Metabion International AG (Martinsried) synthesized the oligonucleotide primers used in this work. They are listed in the supplement section.

2.2 *Media, buffers, solutions*

All media, buffers and solutions were prepared according to Sambrook and Russell (2001) unless otherwise noted.

2.3 *Scientific software and online tools*

Software:

MacVector 9.0 and AssemblyLIGN	Accelrys Inc.
gcg/Wisconsin Package	University of Wisconsin
Image Quant	Molecular Dynamics, Krefeld

Tool	**application**	**URL**
Rebase	restriction enzymes	http://rebase.neb.com/rebase/rebase.html
BLAST Assemble Data Submission	assembly of sequences	http://moss.nibb.ac.jp/cgi-bin/ blast-assemble
NCBI (Blast and Pubmed)	database for sequence information and literature	http://www.ncbi.nlm.nih.gov/
Physcobase	Physcomitrella sequence database	http://moss.nibb.ac.jp
NetPlantGene Server	splice site prediction	http://www.cbs.dtu.dk/services/NetPGene/ index.php

Table 1:

Online tools used for *in silico* analyses of nucleic acid and protein sequences

2.4 *Bacteria and plasmids*

All cloning vectors were amplified using the *E. coli* strain TOP 10 (Invitrogen).

F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG

Plasmid	Application	Source
pCR^{R}2.1 TOPO^R	cloning and sequencing	Invitrogen
pCR^{R}2.1	cloning and sequencing	Invitrogen
pSPUTK	in vitro transcription and translation	Promega
pPpMADS1_GUS_NptII	source for *GUS* and *nptII* sequences	W. Faigl, MPIZ
pARLAK	cloning of moss transformation constructs	W. Faigl, MPIZ
pPPMC2_dsRED_aph4	transformation of moss	W. Faigl, MPIZ
pPPMC1_Zeo^R	transformation of moss	W. Faigl, MPIZ

Table 2: DNA vectors

2.5 *Isolation of nucleic acids*

The isolation of DNA was performed according to "Moss gene technology" (Knight et al., 2002); for RNA isolation from moss tissues Biomol reagent (Biomol GmbH, Hamburg) was applied according to the manufacturer's instructions.

2.6 *Sequencing*

DNA sequencing reactions were performed by the ADIS Sequencing Service Unit of the Max-Planck-Institute for Plant Breeding Research, Cologne.

2.7 *Plant material*

The transgenic *Physcomitrella patens* lines M49 and M50, resulting from the transformation of a translational fusion of *PPMC2* and *GUS* (see Table 4 and Fig. S2), as well as the lines M143 and M144, expressing a fusion of *PPMC1* and *GUS* (see Table 4 and Fig. S2), were kindly provided by Wolfram Faigl, MPIZ.

2.8 *Maintenance of moss cultures*

An established wildtype *Physcomitrella patens* (Hedw.) B.S.G. lab strain was cultivated as described by Schaefer (www2.unil.ch/lpc/docs/pdf/PPprotocols2001.pdf).

2.9 *Selective cultivation*

Transgenic moss lines carrying selection cassettes for Paromomycin, Hygromycin or Zeocin were cultivated on medium containing established concentrations of the respective antibiotic. The antibiotics were obtained from Duchefa Biochemie B.V. (Haarlem, The Netherlands).

Antibiotic	**Concentration**	**Selection Cassette**	**Source**
Paromomycin	40mg/l	35S::*nptII*::polyA	pPpMADS1_GUS_NptII W. Faigl, MPIZ
Hygromycin	15mg/l	nos::*aph4*::polyA	pUC18_Hygro^{R}, B. Reiss
Zeocin	50mg/l	35S::*ble*::polyA	P35S-Zeo, M. Hasebe

Table 3:

Antibiotics for the selection procedure of transformed moss protoplasts

2.10 *Moss transformation for gene knock-out and gene-reporter fusion-loci*

In two translational fusions of *PPMC2* a reporter gene was fused in frame to exon 7, abolishing alternative splicing events that have been shown to occur further downstream in the *PPMC2* transcript (Krogan and Ashton, 2000; Henschel et al., 2002). In the DNA construct *PPMC2* $_{exon7}$*:GUS nptII* (Henschel, unpublished data), the *ß-Glucuronidase* (*GUS*) open reading frame was fused to *PPMC2* via a naturally occurring *Aat*II restriction site, followed by a 35S::*nptII*::PolyA selection cassette. The core construct is flanked by stretches of DNA homologous to the wild type *PPMC2* locus in order to ensure homologous recombination upon transformation (see Table 4 and Fig. S2). Fusion of *PPMC2* and the *dsRED* ORF was achieved by ligating the *dsRED* ORF via *Not*I to exon 7 of *PPMC2* (see Table 4 and Fig. S2). The *aph4* gene conveying resistance to Hygromycin under control of the *nos* promoter and terminator was connected downstream in reverse orientation via *Sal*I. The construct is flanked by stretches of homologous *PPMC2* genomic sequence.

For the replacement of the *PPMC2* coding region by the *GUS* ORF, a 1674 bp fragment of the *PPMC2* promoter was cloned upstream of the *GUS* ORF, followed by a 35S::*nptII*::PolyA selection cassette and a 1330 bp 3' flanking region, consisting of the *PPMC2* coding region up to intron 4 (see Table 4 and Fig. S2). *PPMC2::DEF*$_{5'UTR}$*:GUS* was synthesized from the resulting construct by replacing the putative 5' UTR region of *PPMC2* (-730 to -1) by a PCR product representing the 5' UTR of the *DEFICIENS* gene from *Antirrhinum majus* (Schwarz-Sommer et al., 1992; see Table 4 and Fig. S2). This led to a shorter 5' flanking homologous region of 1049 bp.

In order to reveal expression pattern and function of *PPMC1*, a translational fusion with *GUS* and a gene disruption construct were synthesized. A *Not*I-flanked fragment of both the *GUS* ORF and the *nptII* selection cassette was ligated to exon 7 of *PPMC1* and completed by a 3' homologous DNA stretch (see Table 4 and Fig. S2). For the knock-out of *PPMC1*, the 35S::*ble*::polyA selection cassette was inserted between flanking

homologous regions of the genomic sequence, leading to disruption of the wild type locus in intron 1 upon homologous recombination (see Table 4 and Fig. S2).

target locus	**construct**	**5' flanking homologous sequence**	**reporter**	**resistance cassette**	**3' flanking homologous sequence**	**plasmid backbone**	**resulting transgenic moss lines**
PPMC2	*PPMC2*$_{exon7}$*:GUS_nptII*; translational fusion	+372 to +2065 (1694 bp)	*GUS*	35S::*nptII*::polyA	+2066 to +3004 (939 bp)	pBSCII SK (+)	M49, M50
PPMC2	*PPMC2*$_{exon7}$*:dsRED_aph4*; translational fusion	+478 to +2023 (1546 bp)	*dsRED*	nos::*aph4*::polyA	+2029 to +2746 (718 bp)	pCR 2.1 TOPO	MVQ3, MVQ4
PPMC2	*PPMC2::GUS_nptII*; promoter fusion; gene disruption	-1674 to -1 (1674 bp)	*GUS*	35S::*nptII*::polyA	+1 to +1330 (1330 bp)	pARLAK	MVQ1, MVQ2
PPMC2	*PPMC2::DEF*$_{5'UTR}$*:GUS_nptII*; promoter fusion with *DEFICIENS* 5' UTR	-1674 to -731 of *PPMC2* (944 bp) and -105 to -1 of *DEFICIENS* (105 bp) = 1049 bp	*GUS*	35S::*nptII*::polyA	+1 to +1330 (1330 bp)	pARLAK	MVQ5, (MVQ6, MVQ7), MVQ8-12
PPMC1	*PPMC1* $_{exon7}$*:GUS_nptII*; translational fusion	+376 to +2072 (1697 bp)	*GUS*	35S::*nptII*::polyA	+2081 to +2823 (743 bp)	pBSCII SK (+)	M143, M144, MVQ17
PPMC1	*PPMC1_ble*; gene disruption	-3 to +655 (660 bp)	-	35S::*ble*::polyA	+664 to +1678 (1021 bp)	pCR 2.1 TOPO	MVQ13, MVQ14

Table 4:

DNA constructs for allelic replacement in *Physcomitrella patens.* Schematic drawings of the constructs including restriction sites are shown in the supplement Fig. S2.

2.11 *Transformation procedure and selection*

The transformation procedure was performed according to Schaefer (2001). Each construct was excised from the plasmid via the respective restriction enzymes (see Fig. S2) and separated from the vector backbone by gel electrophoresis. 12µg of purified construct DNA were used for the transformation of 400.000 protoplasts via polyethylene-glycol-mediated transfer. After recovery, the protoplasts were cultivated on antibiotic-containing medium (see Tables 3 and 4) for two rounds of selection.

2.12 *Molecular characterization of transgenic moss lines*

Putative positive transformants were identified by DNA isolation and subsequent diagnostic PCR as described in the online "PHYSCOmanual 1.2" by Hiwatashi and colleagues (http://www.nibb.ac.jp/~evodevo/PHYSCOmanual/00Eindex.htm). For this purpose, young tissue from the edge of each clone was transferred to 30µl of PCR buffer, frozen twice in liquid nitrogen, incubated at 68°C for 10min and centrifuged 5min at 5000rpm. 2µl of this DNA preparation were immediately used for PCR analysis.

The fusion cDNAs of *PPMC2*$_{exon7}$*:GUS*, *PPMC2*$_{exon7}$*:DsRED* and *PPMC1*$_{exon7}$*:GUS* were partially amplified by RT-PCR and verified by sequencing. The presence of a *GUS* transcript as well as the absence of a *PPMC2* transcript in *PPMC2::GUS* (see Fig. 10a) and *PPMC2::DEF*$_{5'UTR}$*:GUS* lines was also determined by RT-PCR. Accordingly, a successful gene disruption for *PPMC1* knock-out lines was demonstrated (see Fig. 12).

2.13 *Histochemical detection of GUS activity*

The histochemical GUS activity was assayed as described by Nishiyama et al. (2000).

2.14 *Semi-thin sections of moss tissue and microscopic documentation*

GUS-stained gametophores of the *PPMC2$_{exon7}$:GUS* line were embedded in Araldite according to Sorensen et al. (2002). After vacuum infiltration and overnight incubation in a 5% glutaraldehyde fixative, the samples were carefully dehydrated through an ethanol series (10%, 30%, 50%, 70%, 90%, 2 x 100%) on ice. Eosin yellow was added to the last steps (90% and 100%) of the ethanol series at a final concentration of 0.1% to stain all moss tissues lacking a GUS signal. This step was crucial to retain visibility of the minuscule gametophores in the synthetic resin. After a 12h incubation at 4°C, the samples were transferred to a 1:1 ethanol:propylene oxide mix and infiltrated at 4°C for 5h. Subsequently, the mixture was replaced by absolute propylene oxide and kept at 4°C over night. The gametophores were then exposed to a 1:1 solution of propylene oxide and Araldite mix (prepared from the Epoxy Resin Kit from Agar Scientific) for 2h, followed by vacuum infiltration and overnight incubation in pure Araldite mix. The next day, the infiltrated gametophores were positioned upside down in conical polyethylene capsules, submerged in freshly prepared Araldite mix and vacuum infiltrated. After polymerization at 60°C for 48h, semi-thin sections of typically 800nm were cut using glass knives. The sections were embedded in fresh Araldite on microscopic slides and examined using dark-field microscopy.

2.15 *Protein synthesis and promoter binding assay*

The full-size proteins PPMC1 and PPMC2 (Fig. 3a) were synthesized in vitro from 500ng of template plasmid (kindly provided by R. Hallinger, MPIZ) using the TNT SP6 Coupled Reticulocyte Lysate System (Promega) as described previously (Egea-Cortines

et al., 1999; Winter et al., 2002). The [^{35}S] methionine labelled proteins were verified using SDS-PAGE by exposing a storage phosphor screen to a 12% SDS-PAGE gel after gel run. Subsequently, the proteins were used for a DNA binding assay with two fragments (-552 to -1) of the *PPMC2* promoter labelled by polynucleotide kinase with 5' phosphate derived from [γ^{32}P] ATP. One of the DNA fragments contained the SRE-type CArG-box, in the other DNA probe this motif was removed by PCR using composite primers (see Supplement). Binding conditions and gel retardation assays were as described by Egea-Cortines et al. (1999). Detection of protein-DNA complexes was achieved by exposition of a native polyacrylamide gel to a phosphor storage screen and subsequent scanning of the signals by a PhosphoImager.

3 Results

3.1 *The* Physcomitrella patens *genome contains six MIKC^c-type MADS-box genes*

In addition to the four published *Physcomitrella patens* MIKC^c-type MADS-box genes *PPM1*, *PPM2*, *PpMADS-1* and *PpMADS-S*, now designated *PPMC1*, *PPMC2*, *PPMC3* and *PPMC4*, two more genes could be found in the database of the Physcomitrella genome of the Gransden lab strain using BLAST and a contig assembly software. After personal communication with other scientists working on MADS-box genes in Physcomitrella, the two new gene loci were termed *PPM5* (*PPMC5*) and *PPM14* (*PPMC6*). All six genes encode a highly conserved MADS-domain, a short I-domain, a K-box and a C-terminal domain (Fig. 1). An additional upstream AUG in the mRNA sequences of all genes indicates a putative N-terminal domain preceding the MADS domain. This is also supported by the high degree of sequence conservation compared to the remaining putative 5' UTR (data not shown). Only *PPMC2* and *PPMC3* possess an SRE-type CArG-box in their putative promoters at -394/-403 and -738/-729, respectively. Based on their structure, the six genes can be subdivided into two groups (Fig.1). The first group comprises the genes *PPMC1*, *PPMC2* and *PPMC3* which share a longer C-terminal domain and a shorter first intron than the genes *PPMC4*, *PPMC5* and *PPMC6* of the second group, where the C-terminus is encoded by no more than two exons (Fig. 1). Within each group, sequence and structural similarity decrease versus the 3' end. In *PPMC1*, *2* and *3* exons 8 and 9 vary in composition and length (Fig. 1). However, major differences are restricted to the 3' UTR. Exon 5 of *PPMC6* includes the coding region that is subdivided into exons 5 and 6 in all other genes and therefore encodes part of the K- and also of the C-domain. The C-terminus of *PPMC4* is not known beyond position 501 (exon 6) because a stop-codon could not be identified. Furthermore, the available 3' sequence shows no homology to any of the other five c-type genes.

The neighbor-joining tree (Fig. 2) is based on a cDNA alignment of bases 1-501 of the open reading frames of all six *Physcomitrella patens* MIKC^c MADS-box genes and the two *Arabidopsis thaliana* c-type MADS-box genes *APETALA1* (*AP1*) and

SEPALLATA1 (*SEP1*) as outgroup (see Fig. S1). As suggested by the structural differences among Physcomitrella c-type MADS-box genes (Fig. 1), the phylogenetic tree supports a subdivision into two subclades (Fig. 2). High bootstrap values undermine the branching that places *PPMC2* in the basal position within the *PPMC2*-like subclade (Henschel et al., 2002) and, accordingly, *PPMC4* in the *PPMC4*-like subclade (Fig. 2).

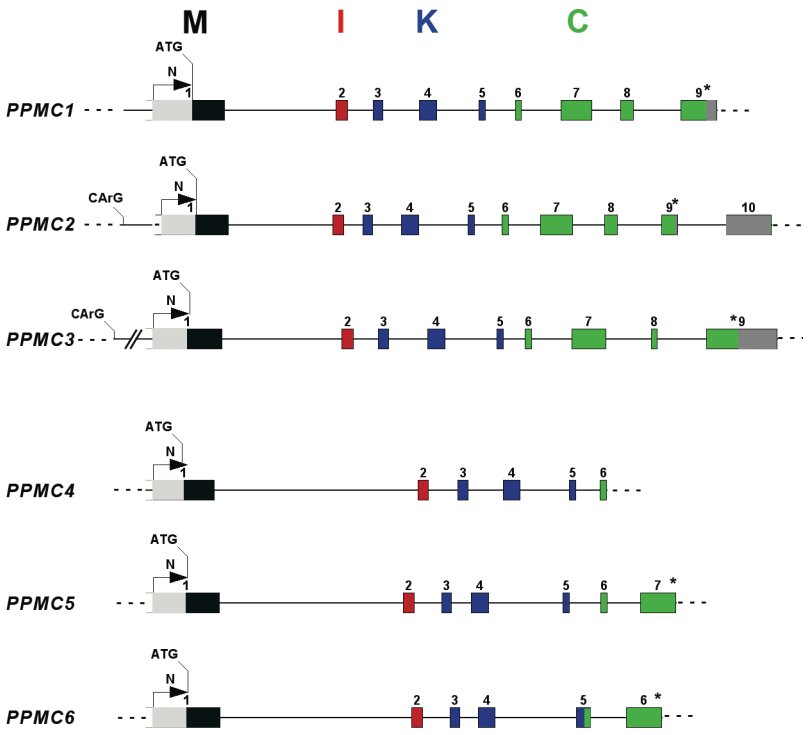


Fig. 1

Exon-intron structure of the six MIKC^c-type MADS-box genes of *Physcomitrella patens*. Exons are shown in boxes, introns are represented by the lines between the boxes. The colors indicate the affiliation of each exon to the above indicated protein domain M(ADS), I(ntervening), K(eratin-like) or C(-terminal); exonic sequence of the 3' UTR is gray. The coding sequences are highlighted by both start codon (ATG) and stop codon (*). Each sequence contains a putative N-terminal domain, indicated by a light gray box. SRE-type CArG-boxes in the putative promoters of *PPMC2* and *PPMC3* are shown.

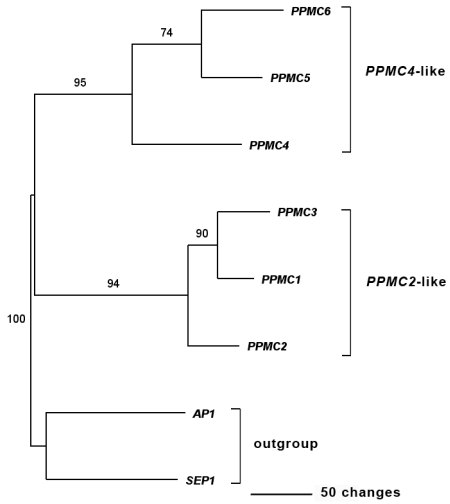


Fig. 2

Unrooted phylogenetic neighbor-joining tree based on a cDNA alignment (see supplement) of all six *Physcomitrella patens* MIKC^c-type MADS-box genes and the *Arabidopsis thaliana* MIKC^c-type MADS-box genes *APETALA1* and *SEPALLATA1* as outgroup. (accession numbers AF150931 (*PPMC1*), AJ419328 (*PPMC2*), AB067688 (*PPMC3*), DQ191323 (*PPMC4*), NM_105581 (*AP1*), NM_121585 (*SEP1*)). The subclade with *PPMC1*, *PPMC2* and *PPMC3* is called *PPMC2*-like, the other subclade, consisting of *PPMC4*, *PPMC5* and *PPMC6*, is designated *PPMC4*-like. Bootstrap values are shown.

3.2 *PPMC2 protein binds to its own putative promoter region*

An SRE-like CArG-box, the established common DNA binding motif of most MIKC^c-type proteins (Hayes et al., 1988; Riechmann et al., 1996), has been identified upstream of the *PPMC2* open reading frame at position -403/-394 (Henschel et al., 2002). A CArG-box containing DNA fragment -552 to -1 upstream of the ATG of the coding sequence of *PPMC2* ("prom*PPMC2* $_{(-552/-1)}$") was used to test for *in vitro* binding of different *Physcomitrella* MIKC^c- and MIKC*-type MADS-domain proteins. With the

exception of PPMC2 protein itself, binding could not be detected for other proteins in electrophoretic mobility shift assays (Fig. 3b and data not shown).

Binding of PPMC2 protein was further analyzed using a modified DNA binding probe where the CArG-box motif was removed using composite primers (see Supplement/Oligonucleotide primers). As a result of this modification, binding of PPMC2 protein was no longer verifiable in the *in vitro* assay, identifying the CArG-box as an essential recognition motif for DNA binding of PPMC2 (Fig. 3b).

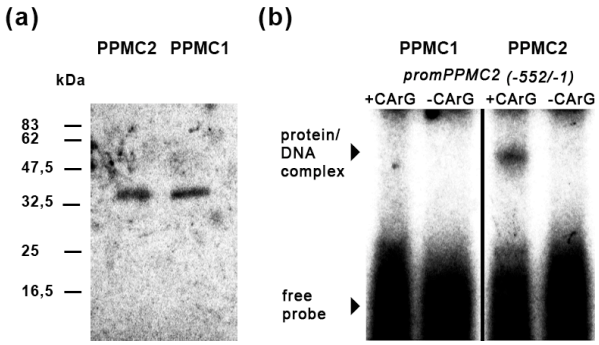


Fig. 3:

***In vitro* binding study of PPMC2 protein to the putative promoter region of the *PPMC2* gene.**

a. Autoradiography of recombinant PPMC1 and PPMC2 protein radioactively labelled with [^{35}S] and loaded onto an SDS-PAGE gel.

b. Native PAGE analysis of electrophoretic mobility shift assays to test for binding of PPMC1 and PPMC2 protein to a DNA probe representing part of the putative *PPMC2* promoter (-552 to -1) with or without the SRE-type CArG-box.

3.3 *PPMC2 protein is strongly expressed in archegonia, spermatozoids and sporophyte feet*

While the temporal expression pattern of *PPMC2* transcript was demonstrated by RT-PCR (Faigl and Münster, unpublished results), tissue or organ specific localization of PPMC2 protein was analyzed by *in planta* expression of reporter gene fusions. Connection of the *ß-Glucuronidase* open reading frame to exon 7 of *PPMC2* (see Table 4 and Fig. S2) leads to intense GUS activity in female reproductive organs, spermatozoids and sporophyte feet (Fig. 4a-d) in the two independent knock-in lines M49 and M50. Expression in archegonia is mostly restricted to the ventral area, but cell-specific distribution, especially in the egg cell or zygote, cannot be resolved studying entire organs. Spermatozoids show increasing intensity of GUS activity that can be correlated with maturation (Fig. 4c). However, expression of $PPMC2_{exon7}$*:GUS* exhibits a high degree of variability according to the level of intensity and distribution among gametophore apices. While in some bundles of gametangia every single organ produces a blue signal upon staining, in other cases only few or none of the reproductive organs do so (data not shown). This is also true for sporophyte feet which vary strikingly in appearance and strength of expression.

Semi-thin sections of stained gametophores carrying gametangia reveal that the fusion protein $PPMC2_{exon7}$:GUS is found in all of the cell layers composing the ventral part of an archegonium (Fig. 5b, c, e, f). The unequal distribution of ß-Glucuronidase product in the 800nm sections is caused by the large vacuoles of the respective cells. Depending on the section plane, GUS product may not be visible because the vacuole eventually takes up all the space within a cell. However, the sections do not contribute to elucidating protein localization in egg cells or zygotes. The delicate protoplasts were likely disrupted during the experimental procedure. GUS activity can also be associated with antheridia carrying mature spermatozoids (Fig. 5d). In contrast, male sexual organs that have already released their reproductive sperm cells no longer show any GUS activity (Fig. 5c).

To observe PPMC2 localization *in vivo* with emphasis on the temporal aspects of development on a single gametophore, knock-in moss lines expressing a fusion of

PPMC2 with the fluorescent protein dsRED were produced (Table 4 and Fig. S2). DsRED fluorescence was verified by spectral analysis (data not shown). Analysis of the two independent lines MVQ3 and MVQ4 did not only confirm the results of the *PPMC2*$_{exon7}$*:GUS* fusion, but revealed additional aspects of *PPMC2* expression (Fig. 6). While strong fluorescent signals in gametangia and sporophyte feet (Fig. 6i, k, m, o, q, s, cc) support that PPMC2 is localized in these organs, a weak ubiquitous signal in all the remaining tissues during the complete life cycle was also observed (Fig. 6a, c, g, e). Apparently, *PPMC2* is not only expressed during all stages of the moss life cycle as shown by RT-PCR (Faigl and Münster, unpublished results), but also the protein is found ubiquitously. However, the variability of *PPMC2*$_{exon7}$*:dsRED* expression in archegonia, spermatozoids and sporophyte feet matches that of *PPMC2*$_{exon7}$*:GUS* in terms of intensity and distribution. Observation of gametophore apices *in vivo* reveals that single reproductive organs or entire bundles regularly lack a fluorescent signal (data not shown). In spite of this fact, the respective apices develop normal sporophytes (without the intense dsRED signal in the foot) with viable spores.

While the analysis of lines expressing *PPMC2*$_{exon7}$*:GUS* does not clarify the localization of PPMC2 protein in egg cells and zygotes, the possibility to observe the *dsRED* reporter gene product *in vivo* allows a thorough examination without the risk of destroying the easily damageable cells. Both the haploid egg cell (not shown) and the diploid zygote (Fig. 6m), discernable by the open archegonium neck, show a strong fluorescent signal.

Furthermore, the application of protein fusion with dsRED allows detection of the subcellular distribution of PPMC2. The transcription factor obviously allocates within the nucleus (Fig. 6a, c, g), which was verified by DAPI staining (data not shown), but is also present in the cytosol.

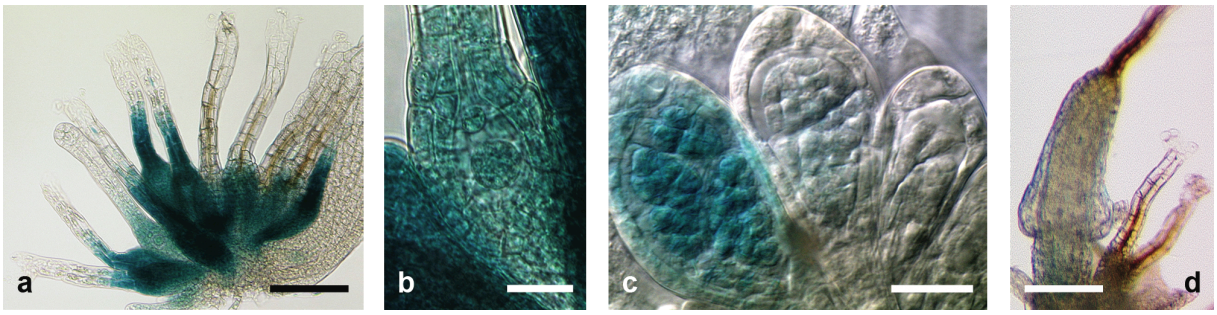


Fig. 4:

a-d. Expression of a PPMC2$_{exon7}$:GUS fusion protein in a bundle of archegonia (a), the ventral part of a single archegonium with egg cell (b), a group of antheridia of different developmental stages (c) and in a young sporophyte (d). Scale bars correspond to 200µm (a, d) and 20µm (b, c).

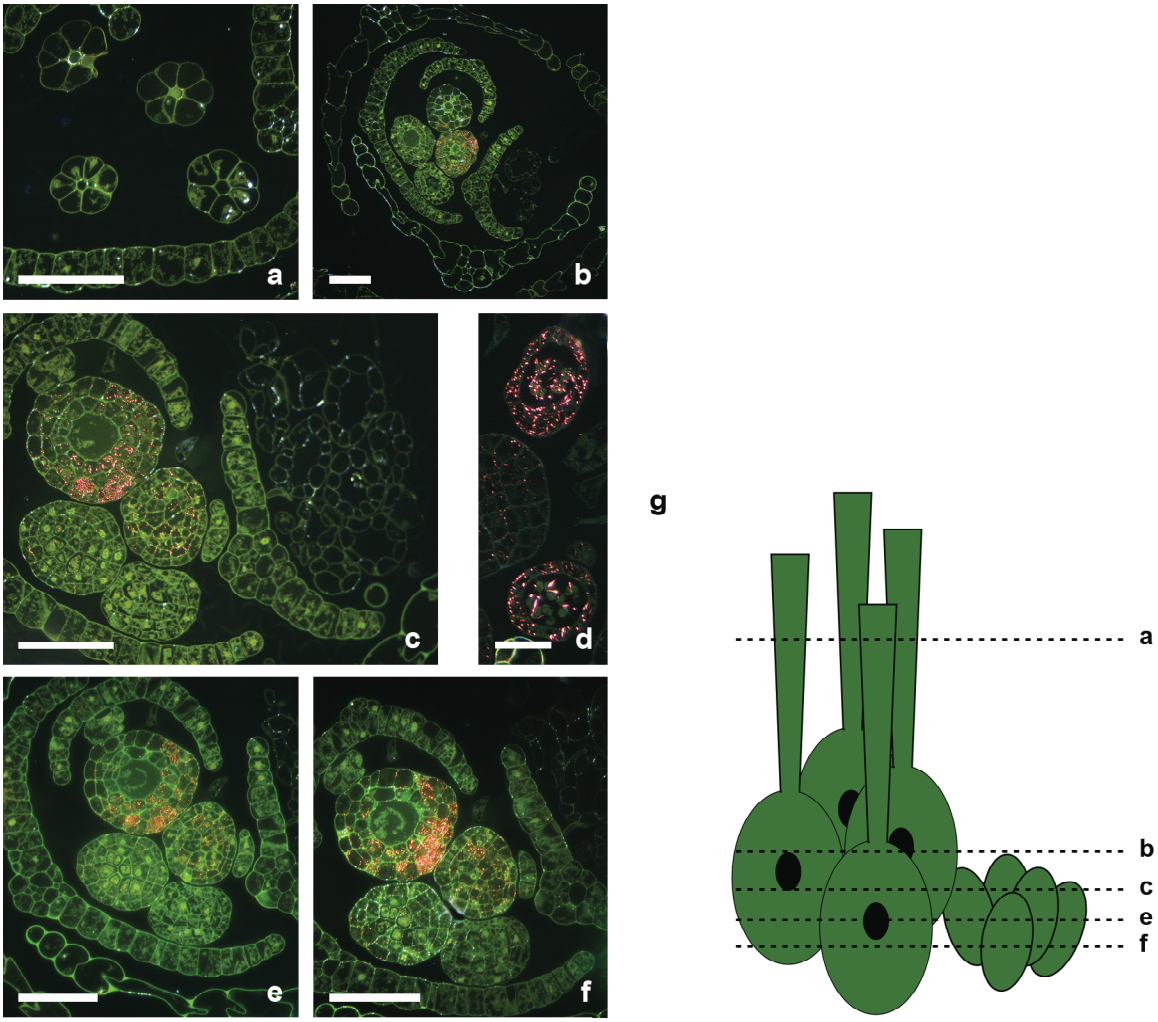
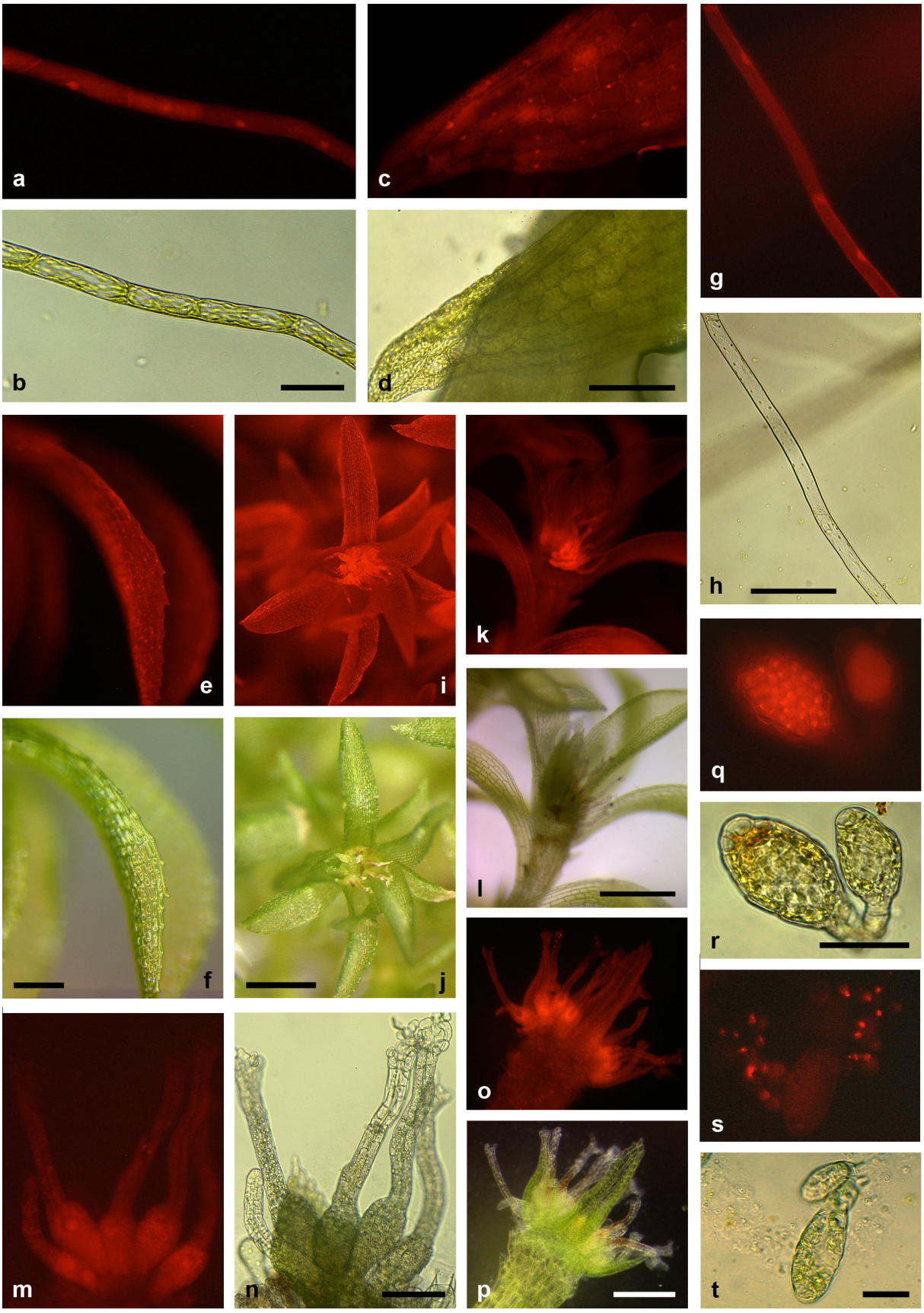


Fig. 5:

Semi-thin sections showing the localization of PPMC2 $_{exon7}$:GUS fusion protein in the apical region of a Physcomitrella gametophore bearing both types of gametangia. The images were taken during dark-field microscopy and display GUS activity product in pink.

a-g. Section through archegonia necks (a), ventral parts of archegonia (b, c, e, f), and empty antheridia (b, c, f) as depicted in the schematic drawing of the organs (g). A section of antheridia containing spermatozoids is shown in (d). The scale bars indicate 50µm (a-c, e, f) and 20µm (d), respectively.


a
c
g
b
d
h
e
i
k
q
l
f
j
r
o
s
m
n
p
t

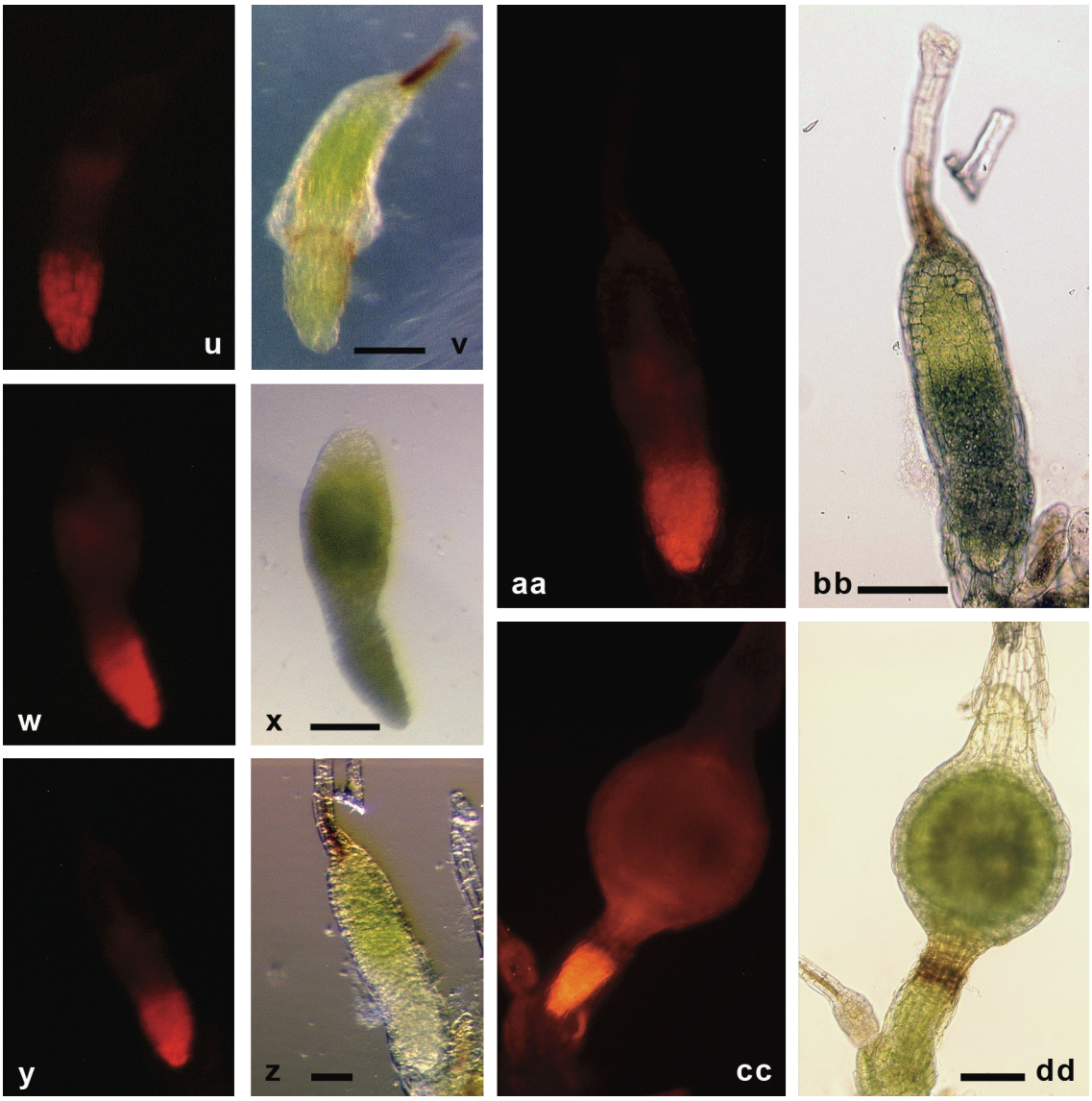


Fig. 6:

a-dd. Fluorescent signal of a PPMC2$_{exon7}$:DsRED fusion protein and the corresponding light microscopic image in protonema (a, b), leaflets (c-f), rhizoids (g, h), gametangia on a gametophore apex (i-l, o, p), a bundle of archegonia of different developmental stages (with zygote) (m, n), a young and a fully developed antheridium with spermatozoids (q, r), spermatozoids released from a mature antheridium (s, t) and five developmental stages of a sporophyte (u-dd). Scale bars correspond to 500µm (e, f, i-l), 250µm (o, p), 100µm (c, d, u-z, aa-dd) and 50 µm (a, b, g, h, m, n, q-t), respectively.

3.4 *Variable splicing of the* PPMC2 *5' UTR*

While the available RT-PCR data of *PPMC2* do not specify tissue- or organ-specific expression, the protein pattern clearly gives evidence of differential regulation. Since the analyses via *in situ* hybridization and single-cell or tissue-specific RT-PCR to reveal distinct transcript patterns are not established in Physcomitrella, it is not possible to clarify directly if *PPMC2* regulation is executed on the transcript level. However, it has been reported that many transcription factors are subject to translational control rather than transcriptional control via their 5' UTR where different mechanisms have been described that lead to stalling or dissociation of scanning ribosomes (Gallie, 1993). Therefore, a structural analysis of the *PPMC2* 5' UTR was performed to investigate its potential role in translational regulation.

5' RACE experiments (see Supplement/Oligonucleotide primers) revealed a putative transcription initiation site at -730 for *PPMC2* mRNA (data not shown). The unusually long 5' UTR exhibits an additional upstream intron, a feature that has been associated with translational control of gene expression (Weise et al., 2005). The intron 0 comprises three splice variants a, b and c (Fig. 7a and Fig. S3) with a common 3' splice site. Their 5' splice sites differ and thus generate 5' UTR versions of unequal lengths. Additionally, four short upstream open reading frames (uORFs) have been identified within the putative 5' UTR of *PPMC2* (Fig. 7a and Fig. S3). Splicing of intron 0b leads to removal of uORF3 and disruption of uORF2. However, when intron 0a or 0c are spliced, uORF4 and 2 or uORF3 and 2 are combined to a new open reading frame, respectively. All these features are highly plausible candidates for a regulatory mechanism.

RT-PCR with cDNAs representing different stages of the moss life cycle demonstrates that all splice variants, including the non-spliced version, exist during all developmental stages in a comparable ratio (Fig. 7b). In order to investigate the composition of 5' UTR splice variants in those tissues with strong *PPMC2* expression, the apical parts of gametophores with reproductive organs were harvested to produce an apex-enriched fraction since collecting gametangia only was technically impractical. Fig. 7b shows that the distribution of 5' UTR versions in the apex-enriched fraction is comparable to all

other samples, including gametophores of the same developmental stage that were not enriched. Apparently, there is no major difference between the composition and weighting of splice variants throughout the moss life cycle. Translational control by splicing of the 5' UTR is thus unlikely to be directly responsible for the differential expression pattern of *PPMC2*. However, involvement of the *PPMC2* 5' UTR in translational regulation by a different mechanism cannot be excluded and was subject to further investigation.

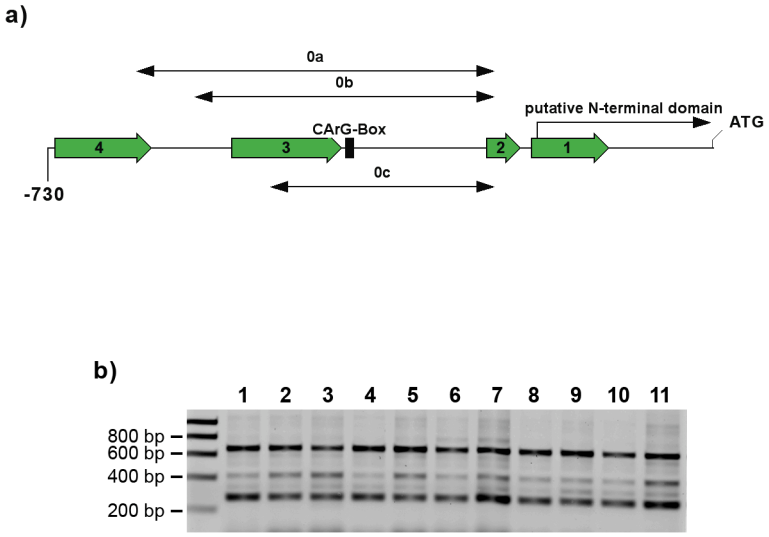


Fig. 7:

Structural and functional analysis of the *PPMC2* 5' UTR.

a. Schematic representation of the 5' UTR of *PPMC2* with the putative transcription start site at -730. Introns 0a, 0b and 0c, which share a 3' splice site at -239 but have different 5' splice sites at -633, -569 and -486, respectively, are shown. Upstream open reading frames 1-4 are represented by green arrows. Moreover, the position of the SRE-type CArG-box (-403 to -394) is indicated.

b. RT-PCR analysis of the *PPMC2* 5' UTR from cDNA pools representing different stages of the moss life cycle (1-11). (1) 1 week old protonema, (2) 6 week old protonema, (3) protonema with gametophores, (4) young

gametophores, (5) gametophores, (6) gametophores 24h after induction at 17°C, (7) gametophores 3 weeks after induction at 17°C, (8) gametophores with gametangia, (9) apex enriched fraction of gametophores carrying mature gametangia, (10) young sporophytes and (11) sporophytes, respectively. Apparent sizes of the bands are indicated on the left.

3.5 *Replacement of the* PPMC2 *5' UTR region has no effect on the PPMC2 protein expression*

To elucidate putative regulatory functions in protein synthesis, the 5' UTR region of the *PPMC2* genomic locus was replaced by the respective 5' UTR of the MADS-box gene *DEFICIENS* of *Antirrhinum majus*. The *DEFICIENS* mRNA is known to have no translational control activity (Schwarz-Sommer et al., 1992; Zachgo et al., 1995). Therefore, its 5' UTR was selected for a fusion with the coding region of the *GUS* reporter gene. The *DEFICIENS*$_{5'UTR}$*:GUS* fusion was placed via homologous recombination at the *PPMC2* locus in the Physcomitrella genome directly downstream of the native *PPMC2* promoter region (for details see Table 4 and Fig. S2). Molecular analysis of moss lines with a correct integration of the *DEFICIENS*$_{5'UTR}$*:GUS* fusion lead to the identification of eight independent lines called MVQ5-12. A GUS pattern identical to that of the translational fusions in M49 and M50 could be observed among the six lines MVQ5 and MVQ8-12 (Fig. 8), however, the lines MVQ6 and MVQ7 differed slightly in their expression pattern in gametangia. Unlike the other lines, here GUS product showed irregular distribution within archegonia as well as a signal in the apical tissues underneath the reproductive organs. These lines were omitted.

As a control for the GUS expression pattern obtained with the *DEFICIENS* 5' UTR, transgenic moss lines carrying a *PPMC2*::*GUS* reporter fusion at the native gene locus were used. Two independent lines called MVQ1 and MVQ2 were identified and analyzed (Fig. 9).

GUS expression in these lines matches that of MVQ5 and MVQ8-12; the protein is localized in archegonia (Fig. 9a, b), spermatozoids (Fig. 9c) and sporophyte feet (Fig. 9d). The results clearly demonstrate that *PPMC2* is not translationally regulated by its 5' UTR.

Fig. 8:

a-d. GUS expression pattern resulting from a *DEF*$_{5'UTR}$*GUS* transcript under control of the *PPMC2* promoter. The transgenic moss line exhibits GUS signals in antheridia (a, c), archegonia (a, b) and sporophytes (d). Scale bars represent 500µm (a, d) and 100µm (b, c), respectively.

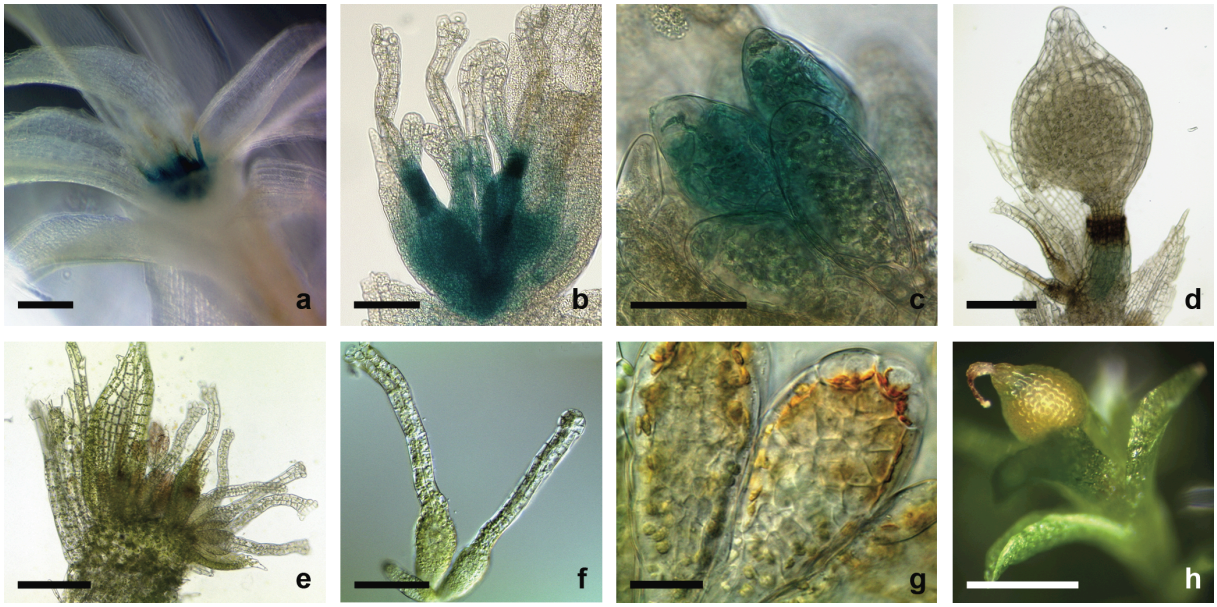


Fig. 9:

a-h. Microscopic and binocular images depicting the expression of *GUS* under control of the *PPMC2* promoter. Staining occurs in archegonia (a, b), antheridia (a, c) and the sporophyte foot (d). The corresponding wild type organs are shown (e-h). Scale bars correspond to 500µm (h), 200µm (a, d, e), 100µm (b, f), 50µm (c) and 20µm (g).

3.6 *Disruption of the* PPMC2 *genomic locus caused no obvious phenotypical changes*

Targeted integration of the *PPMC2::GUS_nptII* construct (see Table 4 and Fig. S2) at the genomic locus of *PPMC2* did also result in disruption of the gene. RT-PCR verified the loss of *PPMC2* transcripts in the transgenic lines MVQ1 and MVQ2 (Fig. 10a). Thus, growth and development of these lines were closely observed to elucidate the function of *PPMC2* in the moss. Based on the *PPMC2* expression pattern, putative developmental or morphological changes were particularly expected to affect the

development of gametangia and gametes, the efficiency of fertilization and the early phases of embryo and sporophyte formation.

Both moss lines did not show any obvious deviations from the wild type morphology (Fig. 10b-g); timing and course of the life cycle did not change. All egg cells and spermatozoids observed looked normal; the mobility of the spermatozoids was not reduced. Also, the number and morphology of sporophytes and developing spores in the sporangia was in the normal range of Physcomitrella wild type moss.

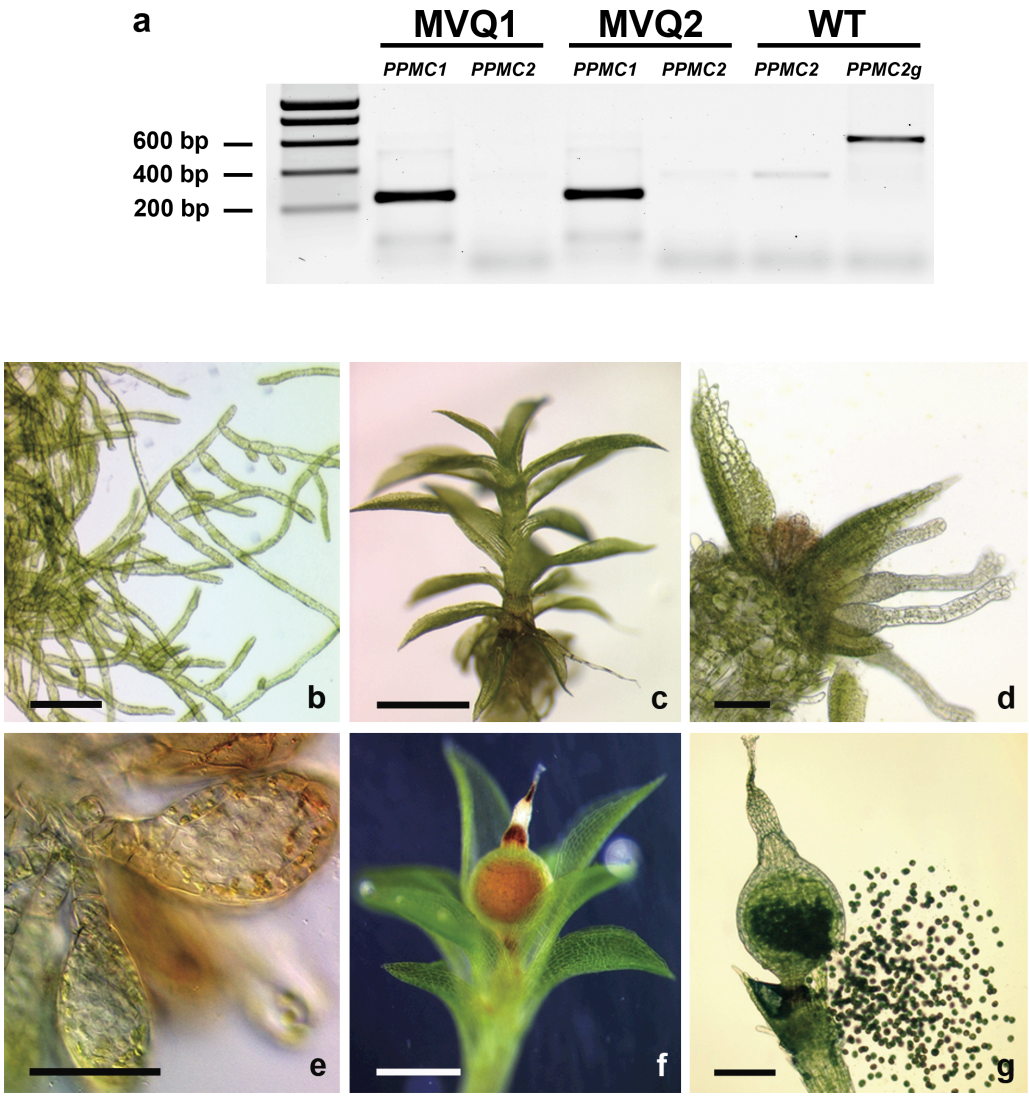


Fig. 10:

Analysis of transgenic knock-out moss lines with a disrupted *PPMC2* locus.

a. **Molecular verification of the targeted *PPMC2* gene disruption in the lines MVQ1 and MVQ2. A 33 cycle RT-PCR with cDNA pools representing young protonema and gene-specific primers for *PPMC1* and *PPMC2*. *PPMC2* primers with wild type (WT) protonema cDNA and genomic DNA (*PPMC2g*) as positive control are shown on the right. Apparent sizes of the product bands are indicated.**

b-g. **Phenotypical characterization of the *PPMC2* knock-out lines. The images depict protonema (b), a young gametophore (c), a gametophore apex with**

male and female gametangia (d), antheridia (e), a sporophyte (f) and a mature sporophyte releasing spores (g). Scale bars correspond to 1000µm (c), 500µm (f), 200µm (b, d, g) and 50µm (e), respectively.

3.7 *PPMC1 is expressed in the gametophore apex and in young sporophytes*

In order to investigate redundancy between *PPMC2* and the other highly similar *Physcomitrella patens* MIKC^c MADS-box genes, the protein expression pattern of a closely related family member (see Fig. 2), *PPMC1*, was analyzed. The molecular analysis of putative transformants of a $PPMC1_{exon7}$*:GUS* translational fusion (see Table 4 and Fig. S2) lead to the identification of three independent moss lines called MVQ17, M143 and M144. While RT-PCR with wild type cDNA and gene specific primers demonstrates that *PPMC1* transcript is present during all developmental stages of the moss life cycle (W. Faigl and T. Münster, unpublished results), GUS staining revealed a temporally and spatially specific protein localization in both the gametophyte and the sporophyte generation of Physcomitrella. It was detected weakly but regularly in the apices of gametophores carrying mature gametangia (Fig. 11a, b), but also in the spore capsules of young sporophytes (Fig. 11c, d). In mature sporophytes GUS activity was mostly no longer traceable, but in rare cases weak expression was found in the short seta above the brown ring and in the very basal part of the spore capsule (data not shown). Comparable to *PPMC2*, a certain variability of the localization signal was observed. Not all apices with mature reproductive organs showed GUS activity. Moreover, staining of sporophytes was even more variable. The intensity and distribution of the signal in young sporophytes was mostly uneven, however, few mature sporophytes with weak expression in the seta suggest that *PPMC1* is first translated throughout the complete upper embryo (with the exception of the foot) and later restricted to the seta. Altogether, PPMC1 expression is always adjacent to PPMC2 expression (Fig. 4 and 11), but does not seem to overlap.

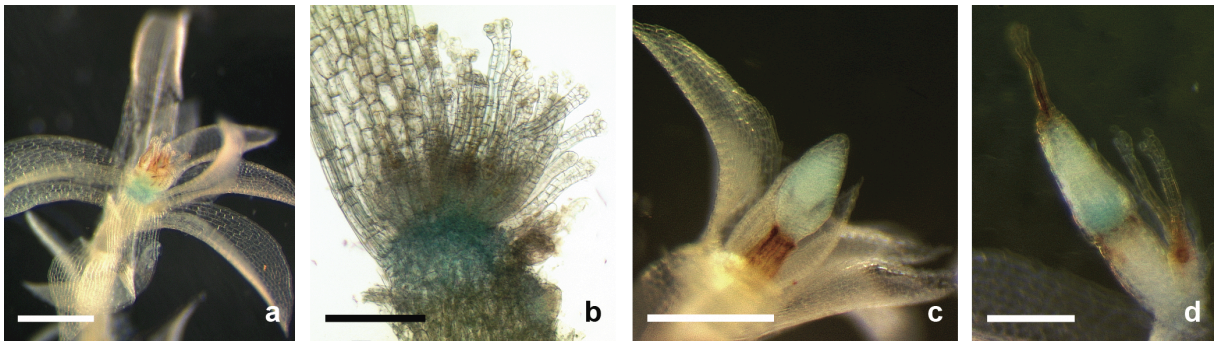


Fig. 11:

a-d. GUS expression pattern in a transgenic line expressing a PPMC1$_{exon7}$:GUS fusion protein. Staining occurs in the apical gametophore region underneath the gametangia (a, b) and in young sporophytes (c, d). Scale bars represent 500µm (a, c) and 200µm (b, d).

3.8 *A* PPMC1 *gene knock-out is without obvious phenotype*

Two independent gene disruption lines for *PPMC1* (see Table 4 and Fig. S2), called MVQ13 and MVQ14, were molecularly identified by RT-PCR (Fig. 12). Close observation of the complete life cycle under regular culture conditions did not reveal any phenotypical changes (data not shown).

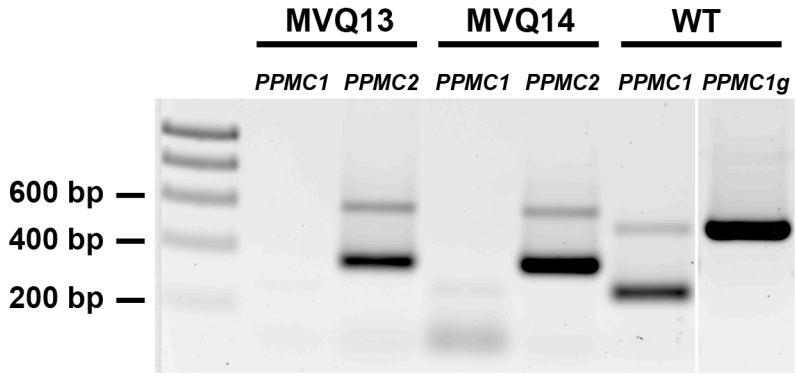


Fig. 12:

RT-PCR analysis of the *PPMC1* gene disruption lines MVQ13 and MVQ14 with cDNA derived from young protonema. Results after 33 PCR cycles using gene specific primers for *PPMC1* and *PPMC2* are shown. Wild type protonema cDNA and genomic DNA (*PPMC1g*) served as template for *PPMC1* primers as positive control. Apparent band sizes are indicated on the left.

Additional bands representing the respective non-spliced product are caused by alternative splicing events, since contamination with genomic DNA was excluded experimentally (data not shown).

3.9 *Double transformants expressing a* PPMC2$_{exon7}$:dsRED *translational fusion in a* PPMC1 *gene disruption background show no phenotype*

The transformation of protoplasts from MVQ3 with the gene knock-out construct *PPMC1_ble* resulted in two independent lines MVQ15 and MVQ16 which express a translational fusion of *PPMC2* and *dsRED* in a *PPMC1* gene disruption background (data not shown). The lines were studied under regular growth conditions and tested for changes in the fluorescent signal representing PPMC2 expression. Neither changes in morphology or development nor in the fluorescence of the fusion protein PPMC2$_{exon7}$:dsRED were detected (data not shown).

4 Discussion

The results demonstrate the protein patterns of the two *Physcomitrella patens* MIKC^c-type MADS-box genes *PPMC1* and *PPMC2*, allowing conclusions regarding their functions. Gene disruption mutants for both genes, however, lack an apparent phenotype. Furthermore, experiments clearly exclude a regulatory function of both the *PPMC2* 5' UTR and the SRE-type CArG-box in the putative *PPMC2* promoter under standard culture conditions. Based on the experimental data, putative functions of the encoded proteins and regulatory mechanisms controlling gene expression are discussed. Finally, the results are embedded in an evolutionary context to explain the emergence of MIKC^c-type MADS-box genes in basal land plants.

4.1 PPMC1, PPMC2 *and* PPMC3 *constitute the* PPMC2*-like subclade of* Physcomitrella patens *MIKC^c-type MADS-box genes*

The phylogenetic tree (Fig. 2) suggests two subclades for the six MIKC^c-type MADS-box genes in the Physcomitrella genome, strongly supported by high bootstrap values. One of the most prominent structural features, the different length of the C-terminus in the *PPMC2*-like and the *PPMC4*-like subclade, additionally justifies the topology of the suggested phylogeny.

The basal position of *PPMC4* is supported by a bootstrap value of 74, but due to the lack of further sequence information for *PPMC4* beyond exon 6, the phylogeny cannot be resolved completely. However, since the calculations based on 501 bp of cDNA sequence convincingly group *PPMC4* with *PPMC5* and *6*, it is likely to expect that *PPMC4* also possesses a short C-terminal domain and is therefore represented by a major portion of its coding region. It is apparent that the neighbor-joining tree reflects a highly plausible phylogeny regarding the *PPMC4*-like subclade, but even more so within the *PPMC2*-like subclade. The very high bootstrap value of 90 strongly supports the assumption that *PPMC1* and *3* are young paralogs with *PPMC2* in a basal position.

The tree topology suggests that the two subclades evolved from gene duplication from a common ancestor that possessed only one MIKC^c-type MADS-box gene in its genome. Only one MADS-box gene of this type has indeed been described in each of three species of charophycean green algae, the closest relatives of land plants and, as such, more basal plants than Physcomitrella (Karol et al., 2001; Tanabe et al., 2005). Therefore, the ancestor of Physcomitrella and these algae may have had only one gene as well. A focus on the C-terminal domain reveals that the described charophycean MADS-box genes feature a short C-terminus (Tanabe et al., 2005) like the moss genes from the *PPMC4*-like subclade. Therefore, the *PPMC4*-like subclade may represent a more ancestral state than the *PPMC2*-like subclade. Possibly, after gene duplication and diversification, the two clades evolved from one gene with a short and one with a long C-terminus. Alternatively, the shorter C-terminal domain might be the result of a secondary loss of coding region. A thorough phylogenetic analysis including all basal plant MIKC^c-type MADS-box genes may help to clarify the matter. Irrespectively, mutations in the C-terminal domain have been discussed to be correlated with diversification and neo-subfunctionalization (Vandenbussche et al., 2003).

A common observation regarding MIKC^c-type genes in basal representatives of plants is the increasing number of gene family members (Kaufmann et al., 2005). While the previously mentioned charophycean green algae posses only one gene of the c-type each (Tanabe et al., 2005), Physcomitrella features six genes and estimates for some analyzed fern representatives suggest up to 30 family members (Münster, personal communication). More derived land plants possess even higher c-type MADS-box gene numbers. In gymnosperms more than 30 genes have been predicted (Theißen, personal communication), and in the angiosperm *Arabidopsis thaliana* 42 of all identified MADS-box genes are of the MIKC^c-type (Parenicova et al., 2003). Apparently, MIKC^c-type gene evolution is characterized by numerous duplication and diversification events, leading to large groups of genes that provide the basis for specialization.

4.2 PPMC2 *is ubiquitously expressed and transcriptionally upregulated in gametangia and sporophyte feet*

PPMC2 protein is found in in archegonia, spermatozoids and sporophyte feet

Based on localization of a PPMC2$_{exon7}$:dsRED fusion protein, the MIKC^c MADS-box gene *PPMC2* is most likely expressed ubiquitously in the haploid and the diploid generation of *Physcomitrella patens*. It cannot be excluded, however, that translation only occurs in distinct tissues of the moss and the protein is subsequently dispersed by protein transport. The analysis of knock-in lines expressing a fusion of *PPMC2* with either *GUS* or *dsRED* reveals that the protein is found in all cells, but is strongly upregulated in the ventral area of archegonia including egg cell and zygote, in developing and mature spermatozoids and in sporophyte feet. Moreover, observation of the fluorescent dsRED signal leads to the insight that the transcription factor accumulates within the nucleus. It is also found in the cytoplasm where it is translated before it is transported into the nuclear compartment.

Since alternative splicing events in those exons encoding the C-terminal domain of *PPMC2* have been reported (Krogan and Ashton, 2000; Henschel et al., 2002), both reporters were fused to exon 7 of *PPMC2* in order to abolish these events. As a consequence, both fusion proteins represent the complete *PPMC2* expression domains. Still, there are differences. While GUS activity can only be detected in those tissues where expression is very strong, the fusion of *PPMC2* and *dsRED* produces a fluorescent signal also in cells of weak expression. Apart from this potential difference in sensitivity, both protein fusion lines show the same pattern of PPMC2 localization throughout the life cycle. This includes an apparent variability of expression in single gametophores. Regularly, high expression of the reporter is restricted to only some of the archegonia and antheridia on one apex; in other cases the apical part of a gametophore completely lacks a signal. Observing such gametophores of the *PPMC2$_{exon7}$:dsRED* lines, cultivated under the same conditions, revealed that their gametangia lead to normal development of a sporophyte producing vial spores. However, these sporophytes do not express the fusion protein in the foot. In many cases

the fluorescent signal was completely absent, including the whole gametophore. These findings emphasize that *PPMC2* is not a classical developmental control gene. In flowering plants, for example, MADS-box genes of the ABC-model function as homeotic selector genes (Theißen et al., 2000), regulating target genes in those tissues where they are expressed, and thus defining the identity of the floral whorls. Loss-of-function mutants exhibit obvious phenotypes (Yanofsky et al., 1990; Jack et al., 1992; Goto and Meyerowitz, 1994; Pelaz et al., 2000). The *PPMC2* expression pattern also suggests a high degree of specificity, however, the variability of GUS activity and dsRED fluorescent signal, as well as the lack of phenotype in the gene disruption mutants, support the assumption that the gene function may not be essential. Instead, the influence of endogenous or environmental factors on *PPMC2* expression seems likely. Initial experiments with a few selected environmental parameters did not provide evidence to explain the variation in protein localization.

PPMC2 is *regulated on the transcriptional level*

Structural features of *PPMC2* may help elucidate the differential regulation of expression. It is now widely acknowledged that the 5' untranslated region of mRNA is involved in many post-transcriptional regulatory pathways that control the level of gene expression (Gallie, 1993) A common regulatory mechanism is folding of the mRNA 5' UTR into a stable secondary structure, impeding the association of scanning ribosomes fo translation (Kozak, 1986). However, *in silico* predictions for the secondary structure of the *PPMC2* 5' UTR did not provide data supporting this hypothesis (data not shown). Among the many mechanisms described is also alternative splicing of the 5' UTR that leads to differentially regulated transcript versions (Procissi et al., 2002). RT-PCR and 5' RACE revealed the existence of an additional intron with three different 5' splice sites and a common 3' splice site in the 5' UTR of *PPMC2*. Additionally, four short upstream open reading frames (uORFs) have been detected within the putative 5' UTR. Many cases have been reported where one or more uORFs cause dissociation or stalling of ribosomes before reaching the main

ORF, leading to inhibition of translation (Morris and Geballe, 2000; Meijer et al., 2002). The combination of splice variants and uORFs that are destroyed or recombined through splicing events, as was shown for *PPMC2* (see results), offers an immense regulatory potential via the 5' UTR. Furthermore, the other five MIKC^c-type MADS-box genes exhibit several putative uORFs in their 5' leader regions as well, suggesting a conserved mechanism. The total number of uORFs cannot be predicted, because the sites of transcriptional initiation for each gene are unknown. Conspicuously, uORFs1, those closest to the main ATG, are highly conserved in sequence and position among all six genes, partly because they overlap to a great extent with the putative N-terminal domains (data not shown). Therefore, a conserved functional relevance of uORF1 seems likely. The other uORFs further upstream are very different in sequence and position, suggesting different functionality among the genes, or no function at all.

Splicing of 5' UTRs, besides in *PPMC2*, has not been tested experimentally. However, the online tool NetPlantGene Server (see Material and Methods) offers a splice site prediction software that was able to identify intron 0 in *PPMC2* and a a similar scenario for *PPMC3*. Here, one 3' splice site with a plausible branch point consensus and three 5' splice sites that each may cooperate with the one 3' splice site are predicted (data not shown). Like *PPMC2*, intron splicing would lead to new uORF combinations. Still, the scenario seems unlikely since already the 5' splice site that is closest to the main ORF is positioned at -807 and thus would require a 5' UTR that is even longer than the predicted 5' UTR of *PPMC2*. Since transcription initiation sites are not available, this matter needs further analysis.

However, RT-PCR has proven the existence of three splice variants and one non-spliced version of the *PPMC2* 5' UTR. The ratio of the four variants appears to be comparable in all analyzed stages of the moss life cycle. Still, a final evaluation is not possible. Due to the lack of suitable methodology, it has not been feasible to clarify whether differences in splice variant distribution, potentially caused by cell- or tissue-specific factors (Procissi et al., 2002), are merely too subtle to be identified. Alternatively, splicing of the 5' UTR may not be the mechanism responsible for the *PPMC2* expression pattern. However, 5' UTRs possess numerous other structural and

compositional features that have been shown to play a role in post-transcriptional control of gene expression (Gallie, 1993). Therefore, a different approach was pursued to identify a general role of the 5' UTR in the regulation of *PPMC2*.

In planta studies with moss lines expressing *GUS* fused to the 5' UTR of the *Antirrhinum majus* gene *DEFICIENS* under the control of the *PPMC2* promoter lead to a GUS pattern identical to that of knock-in lines with the original *PPMC2* 5' UTR. Thus, the *PPMC2* 5' UTR is obviously not responsible for the regulation of *PPMC2* translation. A potential role of its 3' UTR in translational regulation can also be excluded: in the protein fusion lines $PPMC2_{exon7}$:GUS and $PPMC2_{exon7}$:DsRED the 3' UTR was derived from the respective fused reporter gene. The defined protein pattern still remains consistent in all lines expressing a reporter gene under control of the *PPMC2* promoter. Taking all experimental results into account, control of *PPMC2* obviously functions on the transcriptional, not on the translational level. However, a potential relevance of the 5' UTR splice variants, for instance under stress conditions, might possibly exist (Meijer and Thomas, 2002).

4.3 *Binding of PPMC2 to its own promoter does not influence the expression pattern*

Binding of PPMC2 to the CArG-box in its own promoter was shown using EMSA with *in vitro* synthesized proteins. If binding occurs *in vivo* is not known. However, replacing the *PPMC2* coding region by *GUS* and thus producing a knock-out does not change the GUS pattern compared to that of the moss line expressing a fusion protein. Obviously, PPMC2 protein is not needed for regulation of the corresponding gene locus under standard culture conditions. The additional removal of 730 bp of the putative *PPMC2* 5' UTR, containing the CArG-box, and subsequent replacement by the *DEFICIENS* 5' UTR did not change the pattern of GUS activity either. This leads to the conclusion that, under standard conditions, the CArG-box is not involved in the control of *PPMC2* expression at all. The results suggest that binding either does not occur *in vivo* or may

not play a role as long as Physcomitrella is not exposed to biotic or abiotic stress. Therefore, one plausible scenario is binding of PPMC2 to its own CArG-box for auto-regulation during stress conditions. Since MADS domain proteins are known to form dimers or even higher-order complexes that bind to CArG-boxes (Kaufmann et al., 2005), PPMC2 may bind as a homodimer, a heterodimer or even a heteromultimer (Hallinger, 2004). Due to the ATG-proximal position of the CArG-box, protein binding may induce the transcription of mRNA with a short 5' UTR that might lead to an increase of translation. Additional experiments are needed to answer this question.

4.4 PPMC2 *may function in the definition of sink tissues*

While in flowering plants MADS-box genes generally fulfil functions in those tissues or organs where they are expressed (Theißen et al., 2000), this correlation is difficult to establish in non-flowering plants: here MADS-box gene expression is mostly ubiquitous (Münster et al., 1997; Svensson and Engström, 2002; Münster et al., 2002; Tanabe et al., 2003; Tanabe et al., 2005). It has been postulated that a broader expression pattern of a transcription factor may be correlated with a more basal function in plant development (Hasebe et al., 1998; Theißen et al., 2000), i. e. a function that is generally required in all tissues of a plant. By contrast, MADS-box genes in derived seed plants fulfil highly specific tasks in a restricted area of the plant body (Theißen et al., 2000).

PPMC2 exhibits a weak ubiquitous expression in both generations of the Physcomitrella life cycle, but is conspicuously upregulated in the reproductive organs of the gametophyte and the foot of the sporophyte. Both types of expression domain feature highly important developmental and metabolic processes, suggesting that *PPMC2*, even if its function is basal, is especially required during growth and fertilization of gametangia as well as for embryo development and sporophyte maintenance. A certain structural and sequence similarity between *PPMC2* and the B-sister gene clade leads to the assumption that the encoded proteins may fulfil similar functions (Henschel et al., 2002). B-sister genes have been identified in several angiosperm and gymnosperm

representatives (Becker et al., 2002), but not in ferns (Münster et al., 1997; Theißen et al., 2000). *TRANSPARENT TESTA16*, for instance, is expressed mainly in female reproductive organs of *A. thaliana* and is required for the accumulation of proanthocyanidins in the endothelium of the seed coat (Nesi et al., 2002), possibly protecting the seed from harmful UV irradiation. Accordingly, PPMC2 may be responsible for the protection of female reproductive organs (and spermatozoids) from UV light, potentially by inducing flavonoid synthesis or accumulation in the vacuoles of the respective cells (Harborne and Williams, 2000; Winkel-Shirley, 2001). Bryophyte UV protective flavonoids have primarily been analyzed in Antarctic mosses that are exposed to extreme levels of UV irradiation due to the ozone hole (Newsham, 2003; Green et al., 2005), but they are also found in many other moss species (Bates, 2000). While anthocyanins have not been reported in Physcomitrella, a core enzyme of the flavonoid biosynthesis pathway, the chalcone synthase, has been characterized (Jiang et al., 2006). In this context, the weak ubiquitous signal obtained from a $PPMC2_{exon7}$:dsRED fusion can be explained by assuming that UV light protection is crucial in all tissues, but to a different degree. Interestingly, it has been argued that the ancestral function of flavonoids in the first land plants may have been a regulatory one instead of UV light protection (Stafford, 1991), suggesting that they were not as plentiful or effective as present day forms. Stafford (1991) speculates that flavonoids modulated auxin concentrations via an IAA oxidase. A comparable mechanism could still be active in Physcomitrella, however, the expression domains of auxin-responsive genes and *PPMC2* do not overlap (Bierfreund et al. 2003).

The foot of the Physcomitrella sporophyte may also be in need for protective pigments, but it is more likely to assume a different function of the transcription factor or the potentially accumulated flavonoids in the diploid generation. In terms of auxin regulation, a possible indirect regulatory function of *PPMC2* in the foot remains speculative since auxin response has not been described in the moss sporophyte.

The *PPMC2* loss-of-function mutants show no phenotype, but the lack of a UV protective mechanism will most certainly lead to the accumulation of gene mutations, especially in reproductive cells, and thus cause a potentially lethal phenotype after

several generations only. However, experiments conducted with the knock-out lines in order to detect anthocyanins and derivatives in different tissues were not successful. Furthermore, exposition of the transgenic lines MVQ3 and 4 (expressing a translational fusion of *PPMC2* and *dsRED*) to UV-light did not lead to any visible changes in the distribution or intensity of fluorescence (data not shown). Thus, the general variability of the observed fluorescent signal is also not caused by light stimuli. However, the function of *PPMC2* may be a different one altogether.

Despite the fact that gametangia and the sporophyte foot differ immensely in functionality and structure, a common feature is their evident requirement for metabolic energy. Sufficient nutrient supply of cells that develop into reproductive cells such as spermatozoids and egg cells, as well as tissues surrounding an egg cell and providing optimal conditions for fertilization, is indispensable. It has been demonstrated that optimal sucrose and mineral salt concentrations are essential for gametangial induction and fertility in the moss Bartramidula (Chopra and Rahbar, 1982) and the liverwort Riccia (Chopra and Sood, 1973; Sood, 1974). The function of the sporophyte foot in energy supply is even more obvious. In mosses, the diploid generation completely depends on nutrition by the dominant gametophyte (Renault et al., 1992; Schofield, 2001). Nutrients such as carbon, nitrogen or phosphorus compounds (Renault et al., 1992; Mues, 2000) are transported across the placental gap between the gametophore apex and the sporophyte foot. In the moss *Polytrichum formosum*, the highest concentration of the main soluble sugar, sucrose, was found in the sporophyte foot (Renault et al., 1992). Furthermore, the epidermal transfer cells of the *P. formosum* haustorium have shown to create a large proton motive force to energize the uptake of amino acids released into the placental gap by the gametophyte (Renault et al., 1989). A high expression of *PPMC2* in the aforementioned tissues of *Physcomitrella patens* may therefore function in the definition of sink tissues to ensure proper development of those organs that are required for the transition into the next generation of the life cycle. However, disruption of the gene is without obvious effect. Presuming that *PPMC2* fulfils the suggested function also in all remaining tissues where it is only weakly expressed, it seems obvious that it must be much less significant there. Alternatively, the

transcription factor may be involved in different tissue-, organ- or generation-specific gene networks and consequently play diverse roles. However, if *PPMC2* fulfilled various cell-specific functions, a knock-out would clearly not be able to cope with the loss in such an unaffected way, even under standard conditions. Instead, a supportive but not crucial general function, for instance the designation of sink tissues, seems likely. The morphological analysis of the *PPMC2* knock-out lines suggests that a potential phenotype may be very subtle and show under conditions deviant from standard culture conditions. Various experiments with the disruption lines to elucidate the function of *PPMC2* still need to be conducted. Presuming a function in sink tissue definition, analyses involving nutrient starvation and transport inhibitors are likely to provide the necessary evidence.

4.5 *Gene functions among the* PPMC2*-like subclade are non-redundant*

The high amino acid sequence similarity between PPMC1, 2 and 3 suggests redundant functionality among these proteins. PPMC1 as one of the possible candidates was selected to reveal a possible redundancy with PPMC2. The two corresponding genes are apparently highly comparable in several aspects. *PPMC1* transcript was also found during all developmental stages of the moss life cycle (Faigl and Münster, unpublished results), however, the translational fusion revealed a distinct protein pattern in both generations. In the gametophyte, GUS product was detected weakly in the apices of gametophores carrying mature gametangia, in the sporophyte the signal was located in young sporangia. It even seems plausible to assume that a fusion with dsRED might reveal an additional weak ubiquitous signal for *PPMC1*. It would not be surprising because it is likely to assume a similar regulatory mechanism for two highly similar genes. Like *PPMC2*, the expression of *PPMC1* shows a high degree of variability. Most, but not all gametophore apices with gametangia carry a blue signal, and the distribution and intensity in sporophytes is comparably uneven.

In spite of the striking compliances, the results strongly suggest that the two proteins do not fulfil redundant functions. The reporter fusions clearly show that the proteins are expressed in neighboring tissues, but never co-localize. While PPMC2 is found in archegonia and spermatozoids in the gametophyte, PPMC1 is expressed in the apical tissue underneath the reproductive organs. In the sporophyte, PPMC2 is found in the foot, and PPMC1 in the remaining diploid tissues, separated from PPMC2 by the brown ring at the upper part of the seta (Fig. 11c, d and data not shown). The lack of phenotype in the knock-out lines for both genes may imply that either gene can take over the function of the other gene if needed, however, two independent lines (MVQ15 and 16) expressing the translational fusion *PPMC2*$_{exon7}$*:dsRED* in a *PPMC1* knock-out background prove that this is not the case (data not shown). The fusion protein exhibits the same expression pattern as in the wild type background and is therefore not upregulated in those tissues that have shown to express *PPMC1*.

Moreover, reporter gene expression under control of the *PPMC3* promoter shows that the transcription factor is neither found in archegonia, spermatozoids or the sporophyte foot nor in *PPMC1* expression domains (Faigl and Münster, unpublished results). Obviously, even the young paralogs *PPMC1* and *PPMC3* do not share expression domains. Therefore, redundancy among the genes of the *PPMC2*-like subclade is very unlikely, even though the ubiquitous weak expression of *PPMC2* necessarily produces some overlap.

4.6 PPMC1 *functions in the apices of fertile gametophores and in young sporophytes*

Even though alternative splicing of the C-terminal region of *PPMC1* has not been reported, the high similarity between *PPMC1* and *PPMC2* lead to the decision to make the translational fusion to *GUS* also in exon 7 to ensure a complete picture of its expression pattern. Unlike *PPMC2*, there are no further transgenic lines available that express *GUS* under control of the *PPMC1* promoter and thus validate the pattern,

however, the construct was made accordingly and the three independent lines M143, M144 and MVQ17 have identical GUS signals (Fig. 11 and data not shown). The gene disruption mutants do not show a phenotype under standard culture conditions, therefore, a putative function of *PPMC1* must be discussed mainly based on the obtained GUS pattern.

It is a striking fact that *PPMC1* is expressed adjacent to *PPMC2* in both the haploid and the diploid moss generation. It has been shown that the two encoded proteins preferentially heterodimerize *in vitro* (Hallinger, 2004), however, based on the expression patterns, this is unlikely to happen *in vivo* or must be restricted to few cells. The neighboring expression patterns also suggest that *PPMC1* and *2* act as cadastral genes, defining certain areas of the plant body in a developmental context (Theißen et al., 2000). The fact that gene disruption is without effect clearly shows that the two moss transcription factors are no typical organ identity genes (Theißen et al., 2000). However, they might still function in the definition of tissues on a more basal level.
PPMC1, even though expressed in other tissues than *PPMC2*, might still fulfil the same function. The two proteins are highly similar and may have diverged in terms of expression domains, but possibly not regarding their functionality. Considering the suggested involvement of PPMC2 protein in UV-light protection, PPMC1 may play the same role in its own distinct expression domains. This seems plausible regarding the sporophytic generation, where *PPMC1* expression might represent protection of the sporogenous tissues from UV irradiation. However, in the apical cells of the gametophores this function seems less likely because expression only occurs weakly during late stages of fertilization, not earlier when apical cells start to develop into gametangia. Therefore, *PPMC1* might either provide general UV-light protection on a very basic level in the gametophyte, or fulfil a different function there.

Alternatively, *PPMC1* may be involved in sink tissue definition as suggested for *PPMC2*. In this context, however, it is more likely to generally assume different functionality of the two proteins due to the nature of the expression domains. Thus,

PPMC1 might establish a metabolite gradient to ensure proper provisioning of sink tissues defined by PPMC2. If so, the apical area of a gametophore would apply this gradient to ensure fertilization, and in the sporophyte PPMC1 would aid in directing nutrients delivered from the gametophyte via the placental gap and through the cells of the foot. Alternatively, it may lead to the formation of a sink for other nutrients than those attracted by the expression of *PPMC2*. After all, the expression domains possess distinct features that require individual metabolite supply (Mues, 2000). In the developing sporophyte, PPMC1 protein is first found ubiquitously with the exception of the foot, and later only in a restricted area above the brown ring of the seta, suggesting changing metabolite requirements throughout maturation. The weak expression of *PPMC1* in apices with gametangia may represent a requirement for nutrients to support fertilization in the neighboring reproductive organs. Concentrations of nitrogen, phosphorus and potassium were found to be highest in the young shoot apices of the moss Hylocomium (Tamm, 1953), and movement of photosynthates from old to young tissues of Sphagnum have been demonstrated (Rydin and Clymo, 1989), supporting this hypothesis. In any case, *PPMC1* may fulfil different functions in the gametophyte and the sporophyte, potentially caused by different available binding partners and differential functional contexts (Kaufmann et al., 2005). On all accounts, it seems likely that PPMC1 and PPMC2 are never found in the same cells, leading to a clear separation of expression domains. This scenario would even include the heterodimerization of the two proteins, as shown to take place *in vitro* (Hallinger, 2004), in those cells where expression of both accidentally occurs, eventually leading to their mutual inactivation.

The lack of a gene disruption phenotype supports the assumed similarity in function or functional relevance of *PPMC1* and *2*, while redundancy among the three genes *PPMC1*, *PPMC2* and *PPMC3* could convincingly be excluded. The loss-of-function mutants might reveal a putative function of *PPMC1* under culture conditions that cause stress and thus induce differential gene expression. Experiments with a number of classical stress candidates such as UV-light or nutrient deprivation (data not shown), however, did not reveal a stress-related function of either *PPMC1* or *PPMC2* so far.

Alternatively, the phenotype of the *PPMC1* knock-out lines might show under standard conditions, but is too subtle to be detected by morphological analyses only. Moreover, the suggested subtlety may not allow identification of a detectable difference between wild type and gene disruption mutant until after several generations. Based on the expression patterns of *PPMC1* and *PPMC2*, this difference may be associated with either the haploid or the diploid generation of the life cycle, or even both. Considering the vital importance of these transcription factors in seed plants (Theißen et al., 2000), it is likely to assume a disadvantage for the loss-of-function mutants. Close observation of the existing gene disruption lines over several generations is necessary to reveal a putative temporally accumulative effect.

Based on the data for *PPMC1* and *2*, it is reasonable to assume a rather basal function in distinct expression domains also for the other four MIKC^c-type MADS-box genes of *Physcomitrella patens*. However, this is in clear contrast to the highly specialized MADS-box genes in seed plants that fulfil specific functions in restricted tissues or organs (Theißen et al., 2000). Unfortunately, knowledge about MADS-box gene functions in non-seed plants is merely based on conclusions drawn from expression patterns in few representatives (Münster et al., 1997; Hasebe et al., 1998; Münster et al., 2002; Tanabe et al., 2003; Tanabe et al., 2005) and thus remains speculative. Microarray analyses will be able to clearly associate the genes with developmental or metabolic processes, aiding in elucidating their roles.

4.7 PPMC1 *and* PPMC2 *represent the evolutionary transition state of gene recruitment from gametophyte to sporophyte*

The life cycle of plants features an alternation of a haploid generation, called gametophyte, and a diploid generation, called sporophyte. While in charophycean green algae, the closest living relatives of land plants (Karol et al., 2001; Tanabe et al., 2005), the diploid generation of the life cycle is restricted to the zygote, land plants develop multicellular sporophytes. In representatives of basal plants like *Physcomitrella patens*

the sporophyte is dominated by the haploid generation that makes up the greater part of the plant body. However, in more derived plants the sporophyte is usually the dominant phase of the life cycle. In flowering plants, the gametophyte is merely represented by a few cells: the male gametophyte is the three-cellular pollen, the female gametophyte consists of seven cells (eight nuclei) that constitute the embryo sac (reviewed in: Boavida et al., 2005).

The fact that sporophyte-dominant plants evolved from gametophyte-dominant ancestors suggests that the precursors of land plant MADS-box genes originated from the haploid generation and were recruited into the diploid phase during the evolution of land plants (Kofuji et al., 2003; Nishiyama et al., 2003; Tanabe et al., 2005). The increasing relevance of the sporophytic generation throughout the evolution of land plants was accompanied by a need for the required "genetic equipment" for its sustenance. This is supported by current data that depict an evolutionary trend of MADS-box genes to adopt new functions in the diploid body of more derived species.

Three species of the closest living relatives of land plants, the freshwater charophycean green algae, display what may be considered as the ancestral state of MADS-box gene expression (Tanabe et al., 2005). The only MADS-box gene *CpMADS1* found in the unicellular charophycean green alga *Closterium peracerosum-strigosum-littorale* complex is restricted to haploid cells. However, it is upregulated in gametangial cells which are homologous to the moss gametangia (Tanabe et al., 2005). By contrast, the transcript is hardly traceable in the only diploid cell type, the zygote. A very similar distribution was shown for the MADS-box genes of the multicellular charophycean green alga *Chara globularis*: *CgMADS1* expression could not be detected in the zygote, but was found in the egg cell and more weakly in the surrounding tube cells of the oogonium. Additionally, the transcript was verified in the filaments that give rise to spermatozoids as well as in the outermost layer of the antheridium (Tanabe et al., 2005). Unfortunately, no expression data are available for the only c-type MADS-box gene *CsMADS1* identified in *Coleochaete scutata* (Tanabe et al., 2005), however, evidence of transfer cells surrounding the zygote and providing sugars for its development suggest a correlation.

Like *CpMADS1*, *CgMADS1* expression obviously shows remarkable homology to that of *PPMC2* in Physcomitrella. While *PPMC1* expression was detected in the gametophyte generation also, there is no obvious homology to the algal genes in terms of organ- or tissue-specificity.

Among more derived lineages of non-seed plants, five MIKC^c MADS-box genes have been isolated from the club moss *Lycopodium annotinum* (Svensson and Engström, 2002). The broad expression patterns of four of these genes, *LAMB2*, *4*, *5* and *6*, in both vegetative and reproductive tissues of the sporophytic generation support the assumption by Theißen et al. (2000) that, during the evolution of basal land plants, an unspecific localization of MADS-box genes represents the ancestral state. While data on mRNA expression in the gametophyte still lack, *LAMB2*, *4* and *6* transcript distribution demonstrates a certain level of homology with both *PPMC1* and *PPMC2* expression in the moss sporophyte. The transcripts are differentially expressed in diploid vegetative and reproductive tissues (Svensson & Engström, 2002).

Transcript patterns in the gametophytes of several lycopods are still missing and might complete the picture. However, data allocating the MADS-box gene transcripts in the dominant sporophyte of the spike moss *Selaginella remotifolia* are available. One MADS-box gene called *SrMADS1* has been identified so far (Tanabe et al., 2003). While its occurrence in the haploid micro- or megagametophyte is unknown, it is expressed almost ubiquitously in the diploid generation. This rather broad expression pattern resembles the weak ubiquitous signal of PPMC2$_{exon7}$:dsRED (in the sporophyte), but also the GUS signal of PPMC1$_{exon7}$:GUS in the young diploid plant body.

The pteridophyte *Ophioglossum pedunculosum* has also been subject to investigation in terms of MADS-box gene expression. The transcript patterns of four genes isolated from the eusporangiate fern are generally broad in the sporophyte (Münster et al., 2002). While *OPM1*, *3* and *5* are found in both vegetative and reproductive tissues, *OPM4* is restricted to reproductive tissues. As in Selaginella, gametophytes were not available for mRNA analyses.

In a representative of leptosporangiate ferns, *Ceratopteris richardii*, data on five MADS-box genes demonstrate a possibly more derived state of MADS-box gene evolution (Hasebe et al., 1998). The differential expression patterns suggest a certain degree of specialization, accompanied by a trend in favor of the sporophyte. While *CMADS2* and *CMADS3* are predominantly found in the gametophyte and only weakly in the sporophyte, *CMADS1* exhibits a major distribution in meristematic sporophyte tissues in contrast to a weak expression in gametophytes (Hasebe et al., 1998). Moreover, *CMADS6* (also called *CRM3*) has been reported to execute a potentially specific function in hermaphroditic gametophytes (Hasebe et al., 1998) and also in spermatides (Di Rosa, 1998), while it has additionally been identified in the sporophyte (Münster et al., 1997). By contrast, *CMADS4* is restricted to the sporophyte and particularly upregulated in the root.

Among the described non-seed plants, a striking level of homology of expression domains between algal genes and *PPMC2* can be observed. The genes are predominantly expressed in haploid reproductive structures. Localization homologies between *PPMC1* and other non-seed plant MADS-box genes are less obvious. In the sporophytes of the non-seed plants, homologies are generally difficult to assign due to the variation in morphological complexity. In any case, the given data collectively support the general notion that in plant groups branching off from the land plant lineage after mosses MADS-box genes tend to be primarily expressed in the dominant sporophytes and were lost in the gametophyte. The only MIKC^c-type MADS-box gene specifically expressed in the haploid tissues of *A. thaliana* is *AGL18*, however, it has been speculated that it was re-recruited from the sporophyte (Kofuji et al., 2003). While the genes obviously diverged gradually in terms of expression pattern, the data on *PPMC1* and *PPMC2* contradict a general correlation between expression domain and functional relevance. Not only seem their protein patterns limited to certain organs or tissues, already the only MADS-box gene of each of the three analyzed Charophyceae shows temporally and spatially defined expression to a certain degree (Tanabe et al., 2005). Still, their suggested functions are basal. Moss MIKC genes obviously emphasize the complexity of functional evolution.

Taken together, the MIKC^c-type MADS-box genes *PPMC1* and *PPMC2* represent an evolutionary transition state of gene recruitment from gametophyte to sporophyte. They are found in both generations of the life cycle, and the expression patterns reveal that they may not function as typical developmental control genes in Physcomitrella, but fulfil more basal functions (Nishiyama et al., 2003) that are not crucial for normal plant development under standard conditions.

In the gametophyte-dominant moss *Physcomitrella patens*, a basal representative of land plants, the expression patterns of two out of six MIKC^c MADS-box genes thus convincingly support an observed evolutionary trend of gene recruitment from the gametophyte into the sporophyte.

4.8 *MIKC^c-type MADS-box genes in non-seed plants*

In the context of the discussed data on non-seed plants, *PPMC1* and *PPMC2* emphasize one of the key events in the evolution of plants. The development of an extended diploid generation in the plant life cycle conferred on plants a major means of protection against deleterious mutations. With the emergence of the multicellular sporophyte, the need for genetic control probably lead to the recruitment of gametophytic MADS-box genes into the diploid plant body during the course of land plant evolution (Kofuji et al., 2003; Nishiyama et al., 2003). This gene recruitment is represented by the expression patterns and the proposed functionality of the two moss MIKC^c-type genes. They are present in both generations of the moss life cycle and most likely fulfil basal functions in dinstinct domains. Furthermore, their expression patterns suggest that different functional contexts in the gametophyte compared to the sporophyte are possible. As transcription factors, PPMC1 and PPMC2 may be involved in different networks as potential components of di- or tetrameric protein complexes (Theißen et al., 2000; Kaufmann et al., 2005). The recruitment into a novel environment, the diploid generation of the life cycle, may thus lead to a change in gene function itself or in the significance of the retained function in a new context.

MIKC^c-type MADS-box genes obviously constitute a family of highly important genes that played crucial roles during the evolution of plants (Theißen et al., 2000). A cDNA with a MADS-box, the highly conserved core motif of these genes, was isolated from a representative of the putative sister group of the green plants, the red alga *Cyanidioschyzon merolae* (Matsuzaki et al., 2004), however, the I-, K- and C-domains were missing. The MIKC-type MADS-box genes obviously evolved in the lineage consisting of charophycean green algae and land plants (Kaufmann et al., 2005).

While only one gene of the MIKC structure has been identified in each of three Charophyceae (Tanabe et al., 2005), the number of genes in more derived land plants is much higher and originated from duplication and diversification of fewer ancestral genes (Theißen et al., 1996; Theißen et al., 2000). One MADS-box gene is obviously still sufficient in freshwater green algae, but the colonization of the land must have required more than one gene to cope with the terrestrial environment. With the exception of *Selaginella remotifolia* (Tanabe et al., 2003), in all analyzed non-seed plants so far more than one MADS-box gene has been identified (Münster et al., 1997; Hasebe et al., 1998; Münster et al, 2002; Svensson and Engström, 2002).

Even though their functions in non-seed plants are still unknown, the temporal and spatial differentiation of MIKC genes suggest a more specialized function in more derived plants compared to Physcomitrella. Obviously, a higher number of MADS transcription factors leads to task sharing among the available family members. In flowering plants, MADS-box genes are highly specific and are, among many other functions, involved in all stages of the morphogenetic process of flower development (Rounsley et al., 1995; Alvarez-Buylla et al., 2000a; Aswath and Kim, 2005; Kaufmann et al., 2005).

The outstanding evolutionary success of MADS-box genes certainly calls for an explanation. Kaufmann et al. (2005) suggest a crucial role of the typical MIKC domain structure. The ability to bind to different partners and, especially, form higher-order complexes, is based on the structural properties of the transcription factors and probably facilitated a rapid functional diversification (Kaufmann et al. 2005). This is supported by diversification of the K-domain between subfamilies, representing

family-specific functional constraints on possible multimers (Kaufmann et al., 2005). These features are unique to MIKC genes and may have provided a basis for the recruitment into novel environments to develop key innovations such as reproductive organs.

Collectively, the presented data support the general notion of a progressive diversification of MIKC^c-type MADS-box genes during the evolution of plants, represented by genes of recent algae, a moss, lycopods and ferns. However, it must be kept in mind that the plant representatives are indeed recent and have evolved since their first appearance. The MIKC genes of those plants thus do not truly represent an ancestral state, instead, the conspicuously specific expression patterns strongly suggest a derived state. Contrary to derived seed plants, evolutionary changes in non-seed plants obviously did not lead to higher morphological complexity. Instead, advantageous traits such as flexibility and durability may have been the focus to ensure survival of the plants. In fact, a simple organization may be the result of structural specialization and thus be considered an advantage (Frey, 1981). In this context, the simple morphology of Physcomitrella justifies the suggested basal functions for *PPMC1* and *2*. After all, the evo-devo concept is based on a correlation between the development and the evolution of morphology (Gilbert et al., 1996; Vergara-Silva et al., 2000). Thus, the need for a higher number of MIKC^c-type MADS-box genes with more specialized functions arises in more complex plants. Still, the genes in recent representatives of non-seed plants seem no less significant. Since basal plants have had the same time span to evolve protective mechanisms against harmful mutations as seed plants, they probably react to gene loss via alternative strategies, explaining the lack of obvious loss-of-function phenotypes in Physcomitrella. Redundancy can most likely be excluded among the MIKC^c-type MADS-box genes of the moss, but redundancy beyond a gene family, to mention one possible alternative mechanism, could generally exist.

Obviously, the number of gene family members and their speciation in recent non-seed plant representatives still reflect the evolution of plants, however, their significance might be comparable in the respective systems. The identification of MIKC^c-

type MADS-box gene functions in the described organisms as well as in additional representatives of non-seed plants will further elucidate their evolution in land plants.

5 Abstract

MADS-box genes encode a family of highly conserved transcription factors involved in numerous developmental processes in higher eukaryotes. While the members of this transcription factor family have a mostly tissue-specific expression pattern in seed plants, MADS-box genes in non-seed plants show broad expression domains often comprising the gametophytic as well as the sporophytic phase of the plant life cycle. Their functions in non-seed plants remain elusive, although a number of genes could be isolated and characterized from lycophytes, pteridophytes, bryophytes and green algae. To investigate MADS-box genes in a representative of early land plants, the moss *Physcomitrella patens* was selected.

PPMC1 and *PPMC2* are two out of six members of the small *Physcomitrella patens* gene family of MIKC^c-type MADS-box genes. While their transcripts are detected equally during all stages of the *P. patens* life cycle, translational reporter gene fusions reveal distinct expression patterns in both generations of the moss life cycle.

In addition to a weak ubiquitous signal, *PPMC2* shows strong expression in both male and female gametangia of the gametophyte as well as in the haustorium of the diploid sporophyte. As a major suspect of regulation, the unusually long 5' UTR with features known to be involved in translational activation or repression was investigated. However, replacement of the *PPMC2* 5' UTR by the 5' UTR of the Antirrhinum gene *DEFICIENS* in a promoter::GUS fusion disproved regulation on the translational level and, instead, confirmed transcriptional regulation.

The translational reporter gene fusion of *PPMC1* provides evidence that the protein is first expressed in the apices of gametophores carrying mature gametangia. Later, developing sporophytes show broad protein distribution that is gradually restricted to the upper half of the short seta in older stages and finally disappears. Several gene disruption lines have been generated to help reveal the functions of *PPMC1* and *PPMC2*, but so far no obvious phenotype has been detected. Furthermore, a translational

reporter gene fusion of *PPMC2* in a *PPMC1* gene disruption background convincingly supports that the genes are non-redundant.

Based on these findings, putative functions for *PPMC1* and *PPMC2* in the moss are discussed. The genes likely participate in UV-light protection or sink tissue definition in both generations of the moss life cycle, however, gene functions may have diverged between the haploid and the diploid generation. These findings strongly support the hypothesis of MADS-box gene recruitment from the gametophyte into the sporophyte during the evolution of land plants.

6 Zusammenfassung

MADS-Box-Gene kodieren für eine hochkonservierte Familie von Transkriptionsfaktoren, die zahlreiche Entwicklungsprozesse in höheren Eukaryoten steuern. Während die Mitglieder dieser Genfamilie in Samenpflanzen meist eine gewebespezifische Expression aufweisen, sind die Expressionsdomänen in samenlosen Pflanzen eher breit und umfassen sowohl die gametophytische als auch die sporophytische Phase des Lebenszyklusses. Die Funktionen der MADS-Box-Gene in samenlosen Pflanzen sind noch nicht aufgeklärt, obwohl einige Gene aus Lycophyten, Pteridophyten, Bryophyten und Grünalgen isoliert und charakterisiert wurden. Um MADS-Box-Gene in einem Repräsentanten früher Landpflanzen zu untersuchen, wurde das Laubmoos *Physcomitrella patens* ausgewählt.

PPMC1 und *PPMC2* sind zwei von sechs Genen aus der Familie der MIKC^c-Typ MADS-Box-Gene aus *Physcomitrella patens*. Während die Transkripte dieser Gene in jedem Stadium des Physcomitrella-Lebenszyklusses nachgewiesen wurden, ist es gelungen, anhand translationaler Reportergenfusionen zu zeigen, dass sie sowohl in der haploiden als auch in der diploiden Generation definierte Expressionsmuster besitzen.

Neben einem schwachen ubiquitären Signal weist *PPMC2* starke Expression in den männlichen und weiblichen Gametangien des Gametophyten sowie im Fuß des diploiden Sporophyten auf. Da der 5' UTR ungewöhnlich lang ist und außerdem zahlreiche, für translationale Kontrolle typische Eigenschaften besitzt, wurde seine potenzielle Beteiligung an einem Regulationsmechanismus untersucht. Jedoch widerlegte der Austausch des *PPMC2* 5' UTR durch den 5' UTR des Gens *DEFICIENS* aus Antirrhinum in einer Promotor-GUS-Fusion eine Regulation auf Translationsebene und bestätigte stattdessen Transkriptionsregulation.

Mittels einer translationalen Reportergenfusion konnte PPMC1 in den Apices gametangientragender Gametophoren und in jungen Sporophyten nachgewiesen werden. Während das Protein in jungen Sporophyten mit Ausnahme des Fußes nahezu ubiquitär

exprimiert wird, konzentriert es sich im reifenden Sporophyten in der oberen Hälfte der Seta und ist schließlich nicht mehr nachzuweisen.

Um die Funktionen der Gene *PPMC1* und *PPMC2* zu klären, wurden mehrere Verlustmutantenlinien hergestellt. Bisher wurde jedoch kein deutlicher Phänotyp identifiziert. Durch eine translationale Reportergenfusion von *PPMC2* in einem *PPMC1*-mutanten Hintergrund konnte, zusätzlich zu den voneinander abweichenden Expressionsmustern, Redundanz der beiden Gene ausgeschlossen werden.

Basierend auf diesen Ergebnissen werden potenzielle Funktionen von *PPMC1* und *PPMC2* im Moos diskutiert. Möglicherweise sind sie am Schutz vor schädlicher UV-Strahlung oder an der Definition von "Sink"-Gewebe in beiden Generationen des Lebenszyklusses von Physcomitrella beteiligt; zudem könnten sich in den beiden Generationen voneinander abweichende Funktionen entwickelt haben. Die vorliegenden Ergebnisse bestätigen die Hypothese, nach welcher MADS-Box-Gene während der Evolution aus dem Gametophyten in den Sporophyten rekrutiert wurden.

7 References

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF. 2000a. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* **24**: 457-466.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, De Pouplana LR, Martinez-Castilla L, Yanofsky MF. 2000b. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* **97**: 5328-5333.

Aswath, CR, Kim, SH. 2005. Another story of MADS-box genes – their potential in plant biotechnology. *Plant Growth Regul* **46**: 177-188.

Bates JW. 2000. Chemical constituents and biochemistry. In: Shaw AJ, Goffinet B, Editors. Bryophyte Biology, Cambridge University Press.

Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li M-A, Saedler H, Theißen G. 2002. A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol Genet Genomics* **266**: 942-950.

Becker A, Saedler H, Theißen G. 2003. Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev Genes Evol* **213**: 567-572.

Benedito VA, Visser PB, van Tuyl JM, Angenent GC, de Vries SC, Krens FA. 2004. Ectopic expression of *LLAG1*, an *AGAMOUS* homologue from lily (*Lilium longiflorum* Thunb.) causes floral homeotic modifications in Arabidopsis. *J Exp Bot* **55**: 1391-1399.

Bierfreund NM, Reski R, Decker E. 2003. Use of an inducible reporter gene system for the analysis of auxin distribution in the moss *Physcomitrella patens*. *Plant Cell Rep* **21**: 1143-1152.

Boavida LC, Vieira AM, Becker JD, Feijo JA. 2005. Gametophyte interaction and sexual reproduction: how plants make a zygote. *Int J Dev Biol* **49**: 615-632.

Bold HC. 1940. The nutrition of the sporophyte in the Musci. *Am J Bot* **27**: 318-322.

Browning AJ, Gunning BES. 1979a. Structure and function of transfer cells in the sporophyte haustorium of *Funaria hygrometrica* Hedw. I. The development and ultrastructure of the haustorium. *J Exp Bot* **30**: 1233-1246.

Browning AJ, Gunning BES. 1979b. Structure and function of transfer cells in the sporophyte haustorium of *Funaria hygrometrica* Hedw. II. Kinetics of uptake of labelled sugars and localization of absorbed products by freeze-substitution and autoradiography. *J Exp Bot* **30**: 1247-1264.

Chopra RN, Rahbar K. 1982. Temperature, light and nutritional requirements for gametangial induction in the moss *Bartramidula bartramioides*. *New Phytol* **2**: 251-257.

Chopra RN, Sood S. 1973. *In vitro* studies in Marchantiales. I. Effects of some carbohydrates, agar, pH, light and growth regulators on the growth and sexuality in *Riccia crystalline*. *Phytomorphology*. **23**: 230-244.

Cove DJ, Knight CD. 1993. The moss *Physcomitrella patens*, a model system with potential for the study of plant reproduction. *Plant Cell* **5**: 1483-1488.

Cove D, Bezanilla M, Harries P, Quatrano R. 2006. Mosses as model systems for the study of metabolism and development. *Annu Rev Plant Biol* **57**: 497-520.

Di Rosa A. 1998. Molekularbiologische Untersuchungen zum Ursprung homöotischer Gene in Pflanzen am Beispiel der MADS-Box-Genfamilie aus dem Farn *Ceratopteris richardii*. Promotionsschrift, Universität zu Köln.

Egea-Cortines M, Saedler H, Sommer H. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J* **18**: 5370-5379.

Fan HY, Hu Y, Tudor M, Ma H. 1997. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J* **12**: 999-1010.

Frahm J-P. 2001. Biologie der Moose, Spektrum akademischer Verlag GmbH.

Frey W. 1981. Morphologie und Anatomie der Laubmoose. *Adv Bryol* **1**: 399-477.

Gallie DR. 1993. Post-transcriptional regulation of gene expression in plants. *Annu Rev Plant Physiol Plant Mol Biol* **44**: 77-105.

Gilbert SF, Opitz JM, Raff RA. 1996. Resynthesizing evolutionary and developmental biology. *Dev Biol* **173**: 357-372.

Goto K, Meyerowitz M. 1994. Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev* **8**: 1548-1560.

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.

Green TGA, Kulle D, Pannewitz S, Sancho LG, Schroeter B. 2005. UV-A protection in mosses growing in continental Antarctica. *Polar Biol* **28**: 822-827.

Hallinger R. 2004. Interaktionen von MADS-Domänen-Proteinen aus dem Laubmoos *Physcomitrella patens*. Diplomarbeit, Fachhochschule Darmstadt.

Harborne JB, Williams CA. 2000. Advances in flavonoid research since 1992. *Phytochemistry* **55**: 481-504.

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

Hayes TE, Sengupta P, Cochran BH. 1988. The human c-FOS serum response factor and the yeast factors GRM/pRTF have related DNA-binding specificities. *Genes Dev* **2**: 1713-1722.

Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theißen G. 2002. Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol Biol Evol* **19**: 801-814.

Henschel K. 2002. Strukturelle und funktionelle Charakterisierung von MADS-Box-Genen aus dem Laubmoos *Physcomitrella patens* (Hedw.) B. S. G. Promotionsschrift, Universität zu Köln.

Hiwatashi Y, Fujita, T, Sato Y, Tanahashi T, Nishiyama T, Sakakibara K, Kofuji R, Aono N, Sumikawa N, Murata T, Hasebe M. PHYSCOmanual ver. 1.2. URL: http://www.nibb.ac.jp/evodevo/titleE.html.

Hohe A, Rensing SA, Mildner M, Lang D, Reski R. 2002. Day length and temperature strongly influence sexual reproduction and expression of a novel MADS-box gene in the moss *Physcomitrella patens*. *Plant Biol* **4**: 595-602.

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

Jack T, Brockman LL, Meyerowitz EM. 1992. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* **68**: 683-697.

Jiang C, Schommer CK, Kim SY, Suh D-Y. 2006. Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. *Phytochemistry* **67**: 2531-2540.

Joron M, Jiggins CD, Papanicolaou A, McMillan WO. 2006. Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity* **97**: 157-167.

Kamisugi, Y, Cuming AC, Cove D. 2005. Parameters determining the efficiency of gene targeting in the moss *Physcomitrella patens*. *Nucl Acids Res* **33**: e173.

Kamisugi Y, Schlink K, Rensing SA, Schween G, von Stackelberg M, Cuming AC, Reski R, Cove D. 2006. The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucl. Acids Res* **34**: 6205-6214.

Kang HG, Noh YS, Chung YY, Costa MA, An K, An G. 1995. Phenotypic alterations of petal and sepal by ectopic expression of a rice MADS-box gene in tobacco. *Plant Mol Biol* **29**: 1-10.

Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* **294**: 2351-2353.

Kaufmann K, Melzer R, Theißen G. 2005. MIKC-type MADS domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**: 183-198.

Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* **389**: 33-39.

Knight CD, Cove DJ, Cuming AC, Quatrano RS. 2002. Moss gene technology. In: Gilmartin P, Bowler C, Editors. Molecular Plant Biology, Vol. 2, Oxford University Press.

Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M. 2003. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* **20**: 1963-1977.

Kozak M. 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci USA* **83**: 2850-2854.

Krogan NT, Ashton NW. 2000. Ancestry of plant MADS-box genes revealed by bryophyte (*Physcomitrella patens*) homologues. *New Phytol* **147**: 505-517.

Lal M, Bhandari NN. 1968. The development of sex organs and sporophyte in *Physcomitrium cyathicarpum* Mitt. *Bryologist* **71**: 11-20.

Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*. **428**: 653-657.

Meijer HA, Thomas AAM. 2002. Control of eukaryotic protein synthesis by upstream open reading frames in the 5' untranslated region of an mRNA. *Biochem J* **367**: 1-11.

Messenguy F, Dubois E. 2003. Role of MADS-box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* **316**: 1-21.

Morris DR, Geballe AP. 2000. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**: 8635-8642.

Mues R. 2000. Mineral nutrition, substratum ecology, and pollution. In: Shaw AJ, Goffinet B, Editors. Bryophyte Biology, Cambridge University Press.

Munoz-Chapuli R, Carmona R, Guadix JA, Macias D, Perez-Pomares JM. 2005. The origin of the endothelial cells: an evo-devo approach for the invertebrate/vertebrate transition of the circulatory system. *Evol Dev* **7**: 351-358.

Münster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theißen G. 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.

Münster T, Faigl W, Saedler H, Theißen G. 2002. Evolutionary aspects of MADS-box genes in the eusporangiate fern Ophioglossum. *Plant Biol* **4**: 474-483.

Nam J, Kim J, Lee S, An G, Ma H, Nei M. 2004. 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* **101**: 1910-1915.

Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L. 2002. The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **14**: 2463-2479.

Newsham KK. 2003. UV-B radiation arising from stratospheric ozone depletion influences the pigmentation of the Antarctic moss *Andreaea regularis*. *Oecologia* **135**: 327-331.

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol Biol Evol* **17**: 1885-1895.

Nishiyama T, Hiwatashi Y, Sakakibara K, Kato M, Hasebe M. 2000. Tagged mutagenesis and gene-trap in the moss *Physcomitrella patens* by shuttle mutagenesis. *DNA Res* **7**: 9-17.

Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M. 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proc Natl Acad Sci USA* **100**: 8007-8012.

Nishiyama T, Wolf PG, Kugita M, Sinclair RB, Sugita M, Sugiura C, Wakasugi T, Yamada K, Yoshinaga K, Yamaguchi K, Ueda K, Hasebe M. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol Biol Evol* **21**: 1813-1819.

Parenicova L, De Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L. 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.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**: 200-203.

Purugganan, MD. 1997. The MADS-Box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J Mol Evol* **45**: 392-396.

Procissi A, Piazza P, Tonelli C. 2002. A maize r1 gene is regulated posttranscriptionally by differential splicing of its leader. *Plant Mol Biol* **49**: 239-248.

Qiu Y-L, Cho YR, Cox JC, Palmer JD. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**: 671-674.

Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* **103**: 15511-15516.

Renault S, Despeghel-Caussin C, Bonnemain JL, Delrot S. 1989. The proton electrochemical transmembrane gradients generated by the transfer cells of the haustorium of *Polytrichum formosum* and their use in the uptake of amino acids. *Plant Physiol* **90**: 913-920.

Renault S, Bonnemain JL, Faye L, Gaudillere JP. 1992. Physiological aspects of sugar exchange between the gametophyte and the sporophyte of *Polytrichum formosum*. *Plant Physiol* **100**: 1815-1822.

Reski R. 1998: Molecular genetics of *Physcomitrella*. *Planta* **208**: 301-309.

Riechmann JL, Wang MQ, Meyerowitz EM. 1996. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucl Acids Res* **24**: 3134-3141.

Riechmann JL, Meyerowitz EM. 1997. MADS domain proteins in plant development. *Biol Chem* **378**: 1079-1101.

Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Münster T. 2004. Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biol* **7**: 307-314.

Rounsley SD, Ditta GS, Yanofsky MF. 1995. Diverse roles for MADS-box genes in Arabidopsis development. *Plant Cell* **7**: 1259-1269.

Rydin H, Clymo RS. 1989. Transport of carbon and phosphorus compounds about Sphagnum. *Proc Royal Soc London, Series B, Biological Sciences* **237**: 63-84.

Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual. Third edition. Cold Spring Harbor Laboratory Press**.**

Schaefer DG, Zryd J-P. 1997. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J* **11**: 1195-1206.

Schaefer DG. 2001. Gene targeting in *Physcomitrella patens*. *Curr Opin Plant Biol* **4**: 143-150.

Schaefer DG. 2002. A new moss genetics: targeted mutagenesis in *Physcomitrella patens*. *Annu Rev Plant Biol* **53**: 477-501.

Schofield WB. 2001. Introduction to Bryology, The Blackburn Press.

Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**: 931-936.

Schwarz-Sommer Z, Hue I, Huijser P, Flor P, Hansen R, Tetens F, Lönnig WE, Saedler H, Sommer H. 1992. Characterization of the *Antirrhinum* floral homeotic MADS-box gene *DEFICIENS*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* **11**: 251-263.

Schween G, Gorr G, Hohe A, Reski R. 2003. Unique tissue-specific cell cycle in *Physcomitrella*. *Plant Biol* **5**: 1-9.

Shindo S, Ito M, Ueda K, Kato M, Hasebe M. 1999. Characterization of MADS genes in the gymnosperm *Gnetum parvifolium* and its implication on the evolution of reproductive organs in seed plants. *Evol Dev* **1**: 180-190.

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *Eur J Biochem* **229**: 1-13.

Sood S. 1974. *In vitro* studies in Marchantiales-II. Effect of mineral nutrients, chelates and organic nitrogenous sources on the growth and sexuality in *Riccia crystalline*. *Phytomorphology* **24**: 186-197.

Sorensen A, Guerineau F, Canales-Holzeis C, Dickinson HG, Scott RJ. 2002. A novel extinction screen in *Arabidopsis thaliana* identifies mutant plants defective in early microsporangial development. *Plant J* **29**: 581-594.

Stafford HA. 1991. Flavonoid evolution: an enzymic approach. *Plant Physiol* **96**: 680-685.

Svensson ME, Engström P. 2002. Closely related MADS-box genes in club moss (*Lycopodium*) show broad expression patterns and are structurally similar to, but phylogenetically distinct from, typical seed plant MADS-box genes. *New Phytol* **154**: 439-450.

Tamm CO. 1953. In: Sjors H. Growth, yield and nutrition in carpets of a forest moss (*Hylocomium splendens*). *Medd. Fr. Statens Skogsforskningsinst. Stockholm* **43**: 1.

Tanabe Y, Uchida M, Hasebe M, Ito M. 2003. Characterization of the *Selaginella remotifolia* MADS-box gene. *J Plant Res* **116**: 71-75.

Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Münster T, Theißen G, Nozaki H, Ito M. 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.

Theißen G, Kim JT, Saedler H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol* **43**: 484-516.

Theißen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H. 2000. A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149.

Theißen G, Saedler H. 2001. Floral quartets. *Nature* **409**: 469-471.

Vandenbussche M, Theißen G, Van de Peer Y, Gerats T. 2003. Structural diversification and neo-subfunctonalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucl Acids Res* **15**: 4401-4409.

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* (in press).

Vergara-Silva F, Martinez-Castilla L, Alvarez-Buylla ER. 2000. MADS-box genes: development and evolution of plant body plans. *J Phycol* **36**: 803-812.

Weigel D, Meyerowitz EM. 1994. The ABCs of floral homeotic genes. *Cell* **78**: 203-209.

Weise A, Rodriguez-Franco M, Timm B, Hermann M, Link S, Jost W, Gorr G. 2005. Use *of Physcomitrella patens* actin 5' regions for high transgene expression: importance of 5' introns. *Appl Microbiol Biotechnol* **70**: 337-345.

Winkel-Shirley B. 2001. Flavonoid Biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* **126**: 485-493.

Winter KU, Becker A, Münster T, Kim JT, Saedler H, Theißen G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc Natl Acad Sci* **96**: 7342-7347.

Winter KU, Saedler H, Theißen G. 2002. On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in Arabidopsis by expression of an orthologue from the gymnosperm Gnetum. *Plant J* **31**: 457-475.

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. 1990. The protein encoded by the Arabidopsis homeotic gene *AGAMOUS* resembles transcription factors. *Nature* **346**: 35-39.

Zachgo S, de Anrade Silva E, Motte P, Tröbner W, Saedler H, Schwarz-Sommer S. 1995. Functional analysis of the *Antirrhinum* floral homeotic *DEFICIENS* gene in vivo and in vitro using a temperature-sensitive mutant. *Development* **121**: 2861-2875.

8 Supplement

8.1 Supplementary figures

Fig. S1:

Alignment of bases 1-501 of MIKC^c-type MADS-box gene cDNA sequences *PPMC1, PPMC2, PPMC6, PPMC4, PPMC5* and *PPMC3* of *Physcomitrella patens* and *APETALA1* and *SEPALLATA1* of *Arabidopsis thaliana* serving as outgroup. The alignment was used for the calculation of a phylogenetic neighbor-joining tree as shown in Fig. 2.

```
PPMC1     ATGGGTCGCG GTAAAATTGA GATTAAGAAG ATTGAGAATA CCACAAGCAG
PPMC2     ATGGGTCGGG GGAAAATTGA GATTAAGAAG ATTGAGAATA CTACGAGCAG
PPMC6     ATGGGTCGCG GAAAAATAGA AATTAAGAAG ATTGAGAATC CCACCAGCAG
PPMC4     ATGGGTCGCG GTAAAATTGA AATCAAGAAG ATAGAAAATC CTACTAGTAG
PPMC5     ATGGGTCGCG GGAAAATAGA AATTAAGAAG ATTGAGAATC CCACCAGCAG
PPMC3     ATGGGTCGCG GCAAAATTGA GATCAAGAAG ATTGAGAATA CAACCAGCAG
SEP1      ATGGGAAGAG GAAGAGTAGA GCTGAAGAGG ATAGAGAACA AAATCAACAG
AP1       ATGGGAAGGG GTAGGGTTCA ATTGAAGAGG ATAGAGAACA AGATCAATAG

PPMC1     GCAGGTGACA TTCTCCAAGA GGCGTGGAGG TCTATTGAAG AAGGCACACG
PPMC2     GCAGGTGACG TTCTCCAAGA GGCGAGGAGG GCTTCTGAAG AAAGCGCACG
PPMC6     GCAGGTCACA TTTTCCAAGA GGCGTGGGGG CCTGCTCAAG AAGGCCCATG
PPMC4     GCAGGTAACC TTTTCTAAGA GGCGAGGAGG CCTTTTGAAG AAAGCACACG
PPMC5     GCAGGTCACA TTTTCTAAGA GACGCGGAGG GCTGCTCAAG AAGGCTCATG
PPMC3     GCAGGTGACA TTCTCCAAGA GGCGCGGTGG TCTTTTGAAG AAGGCGCACG
SEP1      ACAAGTAACG TTTGCAAAGC GTAGGAACGG TTTGTTGAAG AAAGCTTATG
AP1       ACAAGTGACA TTCTCGAAAA GAAGAGCTGG TCTTTTGAAG AAAGCTCATG

PPMC1     AACTTGCGGT TCTGTGCGAT GCCGAGGTGG CGCTGGTTAT TTTCTCCAGC
PPMC2     AGCTTGCGGT TCTGTGCGAT GCGGAAGTGG CACTTGTTAT TTTCTCCAGC
PPMC6     AGCTCGCCGT ACTCTGCGAT GCCGAGGTGG CCCTCATCAT TTTCTCCAGC
PPMC4     AGCTTGCAGT GCTATGCGAT GCAGAGGTGG CACTCATCAT CTTTTCCAGC
PPMC5     AGCTGGCTGT GCTCTGCGAT GCCGATGTGG CCCTCATCAT TTTCTCCAGC
PPMC3     AGCTTGCGGT TCTGTGTGAT GCCGAGGTGG CGCTGGTTAT TTTCTCCAGC
SEP1      AATTGTCTGT TCTCTGTGAT GCTGAAGTTG CTCTCATCAT CTTCTCCAAC
AP1       AGATCTCTGT TCTCTGTGAT GCTGAAGTTG CTCTTGTTGT CTTCTCCCAT

PPMC1     ACTGGAAAGC TCTTCGAGTA TGCCAGCTCG GGCAGCATGC GAGACATCAT
PPMC2     ACCGGGAAGC TCTTTGAGTA TGCCAGCTCA GGCAGCATTC GAGACATCAT
PPMC6     ACAGGGAAGC TGTTCGAATT CGCCAGCTCA GGCAGTATGC GCGACATTTT
PPMC4     ACAGGAAAGC TATTTGAATT CGCCAGCTCA GGCAGCATGC GCGATATTCT
PPMC5     ACAGGAAAGC TGTTCGAATT CGCCAGCTCA GGCAGCATGC GCGACATTCT
PPMC3     ACTGGAAAGC ACTTTGAGTT TGCCAGTTCA GGCAGCATGC GGGACATCAT
SEP1      CGTGGAAAGC TCTATGAGTT TTGCAGCTCC TCAAACATGC TCAAGACACT
AP1       AAGGGAAAAC TCTTCGAATA CTCCACTGAT TCTTGTATGG AGAAGATACT

PPMC1     CGAGCGCTAT AAGAAGAGCC CGAATG---- --------GC GCAATGAAGT
PPMC2     CGACCGGTAC AAGAAGGGCT CGGATG---- --------G- --AATGCAAA
PPMC6     GGAGCGATAC AGCAAGTGTC CGGATG---- --------G- --AGTGCAAA
PPMC4     GGAGCGATAC AGCAAGTGTC CAGACG---- --------G- --AGTGCAGA
PPMC5     GGAGCGATAT AGTAAGTGCC CTGATG---- --------G- --ATCCCAGA
PPMC3     TGAGCGGTAC AGGAAGAGCT CGGATG---- --------GT GCAGTGAAGC
SEP1      TGATCGGTAC CAGAAATGCA GCTATGGATC CATTGA-AGT CAACAACAAA
AP1       TGAACGCTAT GAGAGGTACT CTTACG--CC GAAAGACAGC TTATTGCACC
```

```
PPMC1     CTGGCGCCAG CACTGATTTC CTGGGTCGCG AGGTCGT--G AAGTTACAGG
PPMC2     ATGGCGCCAG AAATGATTTC ATGGGTTGTG AAGTAGT--A AAGTTACGCG
PPMC6     CTGATGGCAA TAGCGACTTC ATGGGTCGAG AAGTGGT--G AAGCTACGGC
PPMC4     CTACCGGGA- -----ACTTC ATGGGTCGTG AAGTCGT--G AAGCTACGAC
PPMC5     CCGGCGTTAA CAGTGACTTC CTGGGTCGGG AAGTGGT--G AAGCTGCGAC
PPMC3     GTGGCACCAA TACTGATTTA CTTGGTCGGG AGGTGAT--T AAGTTAAAAC
SEP1      CCTGCCAAAG AACT-TGAGA ACAGCTACAG AGAATATCTG AAGCTTAAGG
AP1       TGAGTCCGAC GTCAATACAA ACTGGTCGAT GGAGTATAAC AGGCTTAAGG

PPMC1     AGCAAGTGGA GCGGTTGAAA AGCTCTCAAA GGCGCATGCT TGGCGAGGAT
PPMC2     AGCAATTGGA GCAACTAAAA GCCTCTCACA GGCACATGCT CGGTGAAGAT
PPMC6     AACAGTTAGA GCGATTGCAG CATTCTCAAA GGCACATGCT TGGTGAGGAT
PPMC4     AGCAGTTGGA GCGGATGCAG CATTCGCAAA GGCAAATGCT TGGTGAGGAT
PPMC5     AGGAGTTGGA GCGGTTGCAG CATTCTCAAA GACACATGCT TGGTGAGGAT
PPMC3     AGCAAGTAGA ACGATTGGAA AGCTCTCAAA GGCATATGCT TGGTGAGGAT
SEP1      GTAGATATGA GAACCTTCAA CGTCAACAGA GAAATCTTCT TGGGGAGGAT
AP1       CTAAGATTGA GCTTTTGGAG AGAAACCAGA GGCATTATCT TGGGGAAGAC

PPMC1     CTTTCTGCGC TTAAGGTGCC TGACCTGTTG CAGTTGGAAC AGCAACTCGA
PPMC2     CTGTCGCTGC TTAAGGTGCC TGATCTATTG CAACTGGAGC AACAACTAGA
PPMC6     CTTCAGGTTC TGACGGTGCC TGACCTTCTG CAATTGGAGC AGCAACTGGA
PPMC4     CTCCAGGTTT TAACAGTATC CGACTTGCTG CAATTGGAGC AGCAACTGGA
PPMC5     CTCCAAGTGT TAACTGTGCC TGACCTTCTG CAATTGGAGC AGCAACTGGA
PPMC3     CTTTCAGCTT TGAAGGTATC TGACCTTTTG GAGCTGGAGC AGCAGCTTGA
SEP1      TTAGGACCTT TGAATTCAAA GGAGTTAGAG CAGCTTGAGC GTCAACTGGA
AP1       TTGCAAGCAA TGAGCCCTAA AGAGCTTCAG AATCTGGAGC AGCAGCTTGA

PPMC1     TCTGGGTGCA TCAAGAGTGA GAGCAAGAAA GAATCAACTC ATTCTGGAAG
PPMC2     TTTGGGTGCC TCTCGAGTTC GAGCAAGGAA AAATCAACTT ATATTGGAAG
PPMC6     CATGGGTGTG TCTCGAGTTC GAGCAAGGAA GAATCAACTT TTACTTGAAG
PPMC4     TGTCGGTGCC TCTAGAGTAC GAGCAAGGAA GAACCAGCTT TTATTGGAAG
PPMC5     CATGGGTGCT TCTCGAGTTC GAGCGAGGAA GAACCAACTT TTACTGGAAG
PPMC3     TCAGGGTGCT TCACGAGTGA GAGCAAGGAA GAATCAACTC ATTTTAGAAG
SEP1      CGGCTCTCTC AAGCAAGTTC GGTCCATCAA GACACAGTAC ATGCTTGACC
AP1       CACTGCTCTT AAGCACATCC GCACTAGAAA AAACCAACTT ATGTACGAGT

PPMC1     AGATCGAGGG ATTGCAGAAA AAGGAACAGG AACTGATGGT TGCAAATGAG
PPMC2     AGGTCGAGTC ATTGCGGAGA AAGGAGCACG AGCTGCTAAT TGCAAATGAG
PPMC6     AGGTTGAGGA ATTGCGGCGA AAGGAGCACG ACTTACAGGC CGCAAATGAA
PPMC4     AAATTGAACA ATTAAGACAA AAGGAGCTTG ATTTACAGGC CGAAAATGAA
PPMC5     AGATTGAAGA GTTGCGTAGA AAGGAGCATG ACCTGCATGC CGTAAACGAG
PPMC3     AGATCGAAGA CTTGCGGAGA AAGGAGCATG AACTGATGAT TGCAAACGAG
SEP1      AGCTCTCGGA TCTTCAAAAT AAAGAGCAAA TGTTGCTTGA AACCAATAGA
AP1       CCATCAATGA GCTCCAAAAA AAGGAGAAGG CCATACAGGA GCAAAACAGC

PPMC1     GATCTTCGCA AGAAG---
PPMC2     GACCTCCGCC AGAAGCTT
PPMC6     GAATTGCGTC AGAAGCTT
PPMC4     GATTTGCGTA AAAAG---
PPMC5     GAGTTGCGTC AGAGGCTT
PPMC3     GCTCTTCGCA AGAAG---
SEP1      GCT------- --------
AP1       ATG------- --------
```

Fig. S2:

DNA constructs used for transformation of Physcomitrella protoplasts. Homologous flanking regions are indicated by thin black arrows and labeled according to their origin. *GUS* reporter gene is represented by a blue arrow and *dsRED* by a red arrow. Selection cassettes are shown as boxes in different colors classifying the respective gene: *nptII* (dark green), *aph4* (light green) and *ble* (yellow). In c and d, 5' UTR sequences are highlighted by a red box. Restriction sites used for cloning are indicated.

a. *PPMC2$_{exon7}$:GUS_nptII*

b. *PPMC2$_{exon7}$:dsRED_aph4*

c. *PPMC2::GUS_nptII*

d. *PPMC2::DEF$_{5'UTR}$:GUS_nptII*

e. *PPMC1$_{exon7}$:GUS_nptII*

f. *PPMC1_ble*

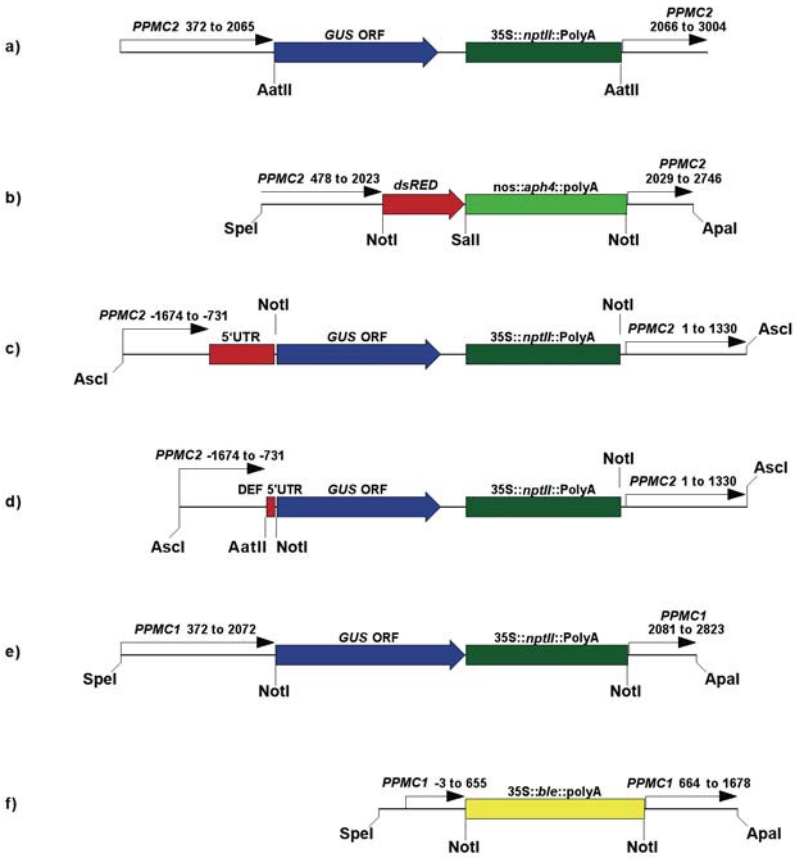


Fig. S3:

Alignment of the putative *PPMC2* 5' UTR (-730 to -1 of *PPMC2* genomic sequence) with sequenced RT-PCR products representing the non-spliced UTR and the three splice variants 0a, 0b and 0c. RT-PCR primers were Q127 (-659) and Q021 (-1). (For the purpose of comprehensibility the start codon ATG was added 3' to the genomic sequence.) The upstream open reading frames are labeled and shown in bold. ClustalW alignment parameters are listed.

```
ClustalW (v1.4) multiple sequence alignment

5 Sequences Aligned           Alignment Score = 33503
Gaps Inserted = 3             Conserved Identities = 264

Pairwise Alignment Mode: Slow
Pairwise Alignment Parameters:
    Open Gap Penalty = 7.0    Extend Gap Penalty = 4.0

Multiple Alignment Parameters:
    Open Gap Penalty = 6.0    Extend Gap Penalty = 3.0
    Delay Divergent = 40%     Transitions: Weighted

Processing time: 1.3 seconds

                        → uORF4
PPMC2 5'UTR       1 ACAAGTATGTGGTTCGCGTCGTACAGAAGCTCAATTGGGAAGCGGCAGGG  50
non-spliced       1                                                    0
splice var. 0a    1                                                    0
splice var. 0b    1                                                    0
splice var. 0c    1                                                    0


PPMC2 5'UTR      51 TGGGCATTTACCCCATTCGGAGGCCCTCCCCCAGGGTCCTGATCTTGGTG 100
non-spliced       1                      GGCCCTCCCCCAGGGTCCTGATCTTGGTG  29
splice var. 0a    1                      GGCCCTCCCCCAGGGTCCTGATCTTG---  26
splice var. 0b    1                      GGCCCTCCCCCAGGGTCCTGATCTTGGTG  29
splice var. 0c    1                      GGCCCTCCCCCAGGGTCCTGATCTTGGTG  29
                                         *************************


PPMC2 5'UTR     101 AGCATATTTGGTAGTTCATCGTTGTCAGGCATTAGGTGTCGTAACGCGCC 150
non-spliced      30 AGCATATTTGGTAGTTCATCGTTGTCAGGCATTAGGTGTTGTAACGCGCC  79
splice var. 0a   27 --------------------------------------------------  26
splice var. 0b   30 AGCATATTTGGTAGTTCATCGTTGTCAGGCATTAGGTGTCGTAACGCGCC  79
splice var. 0c   30 AGCATATTTGGTAGTTCATCGTTGTCAGGCATTAGGTGTCGTAACGCGCC  79


PPMC2 5'UTR     151 GGTGTCGGCAGGTGGGTGGCTGGGGGCCTGCGGTACTGGTGGAGTCTCTT 200
non-spliced      80 GGTGTCGGCAGGTGGGTGGCTGGGGGCCTGCGGTACTGGTGGAGTCTCTT 129
splice var. 0a   27 --------------------------------------------------  26
splice var. 0b   80 GGTGTCGGCAG------------------------------------------  90
splice var. 0c   80 GGTGTCGGCAGGTGGGTGGCTGGGGGCCTGCGGTACTGGTGGAGTCTCTT 129

                    → uORF3
PPMC2 5'UTR     201 AATGACGACGACTTGCCTTGTATTTGCTAATCAGCTCTCAGCAGGTCTCA 250
non-spliced     130 AATGACGACGACTTGCCTTGTATTTGCTAATCAGCTCTCAGCAGGTCTCA 179
splice var. 0a   27 --------------------------------------------------  26
splice var. 0b   91 --------------------------------------------------  90
splice var. 0c  130 AATGACGACGACTTGCCTTGTATTTGCTAATCAGCTCTCAGCAG------ 173
```

```
PPMC2 5'UTR     251 TCAGCTCCAAGCTCCCCACCCGGCACTCGCTGCCAGACCGTGAGTCAGGA 300
non-spliced     180 TCAGCTCCAAGCTCCCCACCCGGCACTCGCTGCCAGACCGTGAGTCAGGA 229
splice var. 0a   27 -------------------------------------------------  26
splice var. 0b   91 -------------------------------------------------  90
splice var. 0c  174 ------------------------------------------------- 173

PPMC2 5'UTR     301 GTAGGGTCAGGGGAGCAGTTGTAGCATCCAAAAAAGGACTGAAGTGGAAG 350
non-spliced     230 GTAGGGTCAGGGGAGCAGTTGTAGCATCCAAAAAAGGACTGAAGTGGAAG 279
splice var. 0a   27 -------------------------------------------------  26
splice var. 0b   91 -------------------------------------------------  90
splice var. 0c  174 ------------------------------------------------- 173

PPMC2 5'UTR     351 GGATTCGTGTCTGGCAGCGTATCTCTACACGTTCTGGGTTTCTGTTGCTA 400
non-spliced     280 GGATTCGTGTCTGGCAGCGTATCTCTACACGTTCTGGGTTTCTGTTGCTA 329
splice var. 0a   27 -------------------------------------------------  26
splice var. 0b   91 -------------------------------------------------  90
splice var. 0c  174 ------------------------------------------------- 173

PPMC2 5'UTR     401 GATTGAGCTAGCGGCCCACTTTCTTAACCGCTTGTTGCGAGGCTGCTCTT 450
non-spliced     330 GATTGAGCTAGCGGCCCACTTTCTTAACCGCTTGTTGCGAGGCTGCTCTT 379
splice var. 0a   27 -------------------------------------------------  26
splice var. 0b   91 -------------------------------------------------  90
splice var. 0c  174 ------------------------------------------------- 173

                                                 → uORF2
PPMC2 5'UTR     451 GTAGATCATCCTAGCTTACTATTTGATTTTGAATGGTGCTAGGTTGTGAG 500
non-spliced     380 GTAGATCATCCTAGCTTACTATTTGATTTTGAATGGTGCTAGGTTGTGAG 429
splice var. 0a   27 -------------------------------------------GTTGTGAG  34
splice var. 0b   91 -------------------------------------------GTTGTGAG  98
splice var. 0c  174 -------------------------------------------GTTGTGAG 181
                                                        ********
                                                 → uORF1
PPMC2 5'UTR     501 AAGCAGACTCGTAAGGTGTAGAGCGGGCTATGGACGATGACTGCGTCT 550
non-spliced     430 AAGCAGACTCGTAAGGTGTAGAGAGCGGGCTATGGACGATGACTGCGTCT 479
splice var. 0a   35 AAGCAGACTCGTAAGGTGTAGAGAGCGGGCTATGGACGATGACTGCGTCT  84
splice var. 0b   99 AAGCAGACTCGTAAGGTGTAGAGAGCGGGCTATGGACGATGACTGCGTCT 148
splice var. 0c  182 AAGCAGACTCGTAAGGTGTAGAGAGCGGGCTATGGACGATGACTGCGTCT 231
                    **************************************************

PPMC2 5'UTR     551 GGACAATCCTGCAAGCAGTGGCAGTTCGGGGCTGTACCCAGTACTGCCAC 600
non-spliced     480 GGACAATCCTGCAAGCAGTGGCAGTTCGGGGCTGTACCCAGTACTGCCAC 529
splice var. 0a   85 GGACAATCCTGCAAGCAGTGGCAGTTCGGGGCTGTACCCAGTACTGCCAC 134
splice var. 0b  149 GGACAATCCTGCAAGCAGTGGCAGTTCGGGGCTGTACCCAGTACTGCCAC 198
splice var. 0c  232 GGACAATCCTGCAAGCAGTGGCAGTTCGGGGCTGTACCCAGTACTGCCAC 281
                    **************************************************

PPMC2 5'UTR     601 GCAAGAGTCATCGAGTGAGATAATTCTGCACGGCGATTGTGAAGTGCCGA 650
non-spliced     530 GCAAGAGTCATCGAGTGAGATAATTCTGCACGGCGATTGTGAAGTGCCGA 579
splice var. 0a  135 GCAAGAGTCATCGAGTGAGATAATTCTGCACGGCGATTGTGAAGTGCCGA 184
splice var. 0b  199 GCAAGAGTCATCGAGTGAGATAATTCTGCACGGCGATTGTGAAGTGCCGA 248
splice var. 0c  282 GCAAGAGTCATCGAGTGAGATAATTCTGCACGGCGATTGTGAAGTGCCGA 331
                    **************************************************

PPMC2 5'UTR     651 CCTACTTGTCACTAGGCGTGGCGTGCGGCACTGGAGGGGATTCGACCTAT 700
non-spliced     580 CCTACTTGTCACTAGGCGTGGCGTGCGGCACTGGAGGGGATTCGACCTAT 629
splice var. 0a  185 CCTACTTGTCACTAGGCGTGGCGTGCGGCACTGGAGGGGATTCGACCTAT 234
splice var. 0b  249 CCTACTTGTCACTAGGCGTGGCGTGCGGCACTGGAGGGGATTCGACCTAT 298
splice var. 0c  332 CCTACTTGTCACTAGGCGTGGCGTGCGGCACTGGAGGGGATTCGACCTAT 381
                    **************************************************

PPMC2 5'UTR     701 TCGACAGAGTCGCGGTCCAAGGGTCGGGTTATG 733
non-spliced     630 TCGACAGAGTCGCGGTCCAAGGGTCGGGTT    659
splice var. 0a  235 TCGACAGAGTCGCGGTCCAAGGGTCGGGTT    264
splice var. 0b  299 TCGACAGAGTCGCGGTCCAAGGGTCGGGTT    328
splice var. 0c  382 TCGACAGAGTCGCGGTCCAAGGGTCGGGTT    411
                    ******************************
```

8.2 Abbreviations

al.	alii
BLAST	Basic Local Alignment Search Tool
CArG	C A/T-rich G motif
cDNA	complementary DNA
c-type	classic-type
*-type	star-type
DAPI	4',6-diamidino-2-phenylindole
dsRED	red fluorescent protein of *Discosoma spec*.
Fig.	Figure
GUS	ß-Glucuronidase
kb	kilobase
MADS	acronym for the four founder proteins MCM1 (*S. cerevisiae*), AGAMOUS (*A. thaliana*), DEFICIENS (*A. majus*) and SRF (*H. sapiens*)
MIKC	domain structure consisting of MADS, intervening, keratin-like and C-terminal domain
ORF	open reading frame
PCR	polymerase chain reaction
5'RACE	5' rapid amplification of cDNA ends
rpm	rounds per minute
RT-PCR	reverse transcriptase PCR
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SRE	serum response element
uORF	upstream open reading frame
UTR	untranslated region
WT	wild type

8.3 Oligonucleotide primers

5’ RACE nested primers:

Q102 5’-TGGAGAACGTCACCTGCCTGCTCGTA-3’

Q103 5’-TCTCAATCTTCTTAATCTCAATTTTC-3’

Q104 5’-CCCCGACCCATAACCCGACCCTTGGA-3’

RT-PCR 5’ UTR *PPMC2*:

Q127 5’-GGCCCTCCCCCAGGGTCCTGATC-3’

Q021 5’-AACCCGACCCTTGGACCGCGACTC-3’

Synthesis of *PPMC2* promoter binding probes:

Flanking primers

Q001 5’-TGCGGTACTGGTGGAGTCTCTTAA-3’

Q021 5’- AACCCGACCCTTGGACCGCGACTC-3’

Composite primers for removal of the CArG-box

Q022 5’-CAGGGGAGCAGTTGTAGCATACTGAAGTGGAAGGGATTCG-3’

Q023 5’-CGAATCCCTTCCACTTCAGTATGCTACAACTGCTCCCCTG-3’

Identification of positive moss transformants:

PCR

PPMC1_ble

W359 5’-GGAGTGGGTATGGATGATGACTGCG

Q107 5’-GTCCCGGAAGTTCGTGGACACGA-3’

PPMC2::GUS* and *PPMC2::DEF$_{5'UTR}$:GUS

Q117 5’-TGTGCGAGGGCAGGATTCTGGCTA-3’

Q017 5’-CACTGCGTCTGGACAATCCTGCA-3’

W327 5’-GAAACGCAGCACGATACGCTGGCC-3’

PPMC2$_{exon7}$:dsRED

Q143 5’-ATATGCGGCCGCATGGGTCGGGGGAAAATTGAGATTA-3’

W338 5’-GACATATACCTTGCTTCCATACTG-3’

RT-PCR

***PPMC2* gene disruption**

W362 5’-GCGTCTTCATATCAGGATTCAAAACTCAAG-3’

W363 5’-GGTTGTGCAGGTGAATCTGCCATG-3’

PPMC1_ble

W364 5’-CGAGAAACAACACCGGAGACGATAGGCTAG-3’

W365 5’-CATATCCCCCAAGCTCTAGCAGATG-3’

PPMC2 $_{exon7}$:dsRED

Q164 5’-CGAGCTGCTAATTGCAAATGAGGAC-3’

W338 5’-GACATATACCTTGCTTCCATACTG-3’

PPMC2::GUS* and *PPMC2::DEF$_{5'UTR}$:GUS

Q017 5’-CACTGCGTCTGGACAATCCTGCA-3’

W327 5’-GAAACGCAGCACGATACGCTGGCC-3’

PPMC1 $_{exon7}$:GUS

W297 5’-GAACAGGAACTGATGGTTGCAAATG-3’

W327 5’-GAAACGCAGCACGATACGCTGGCC-3’

Detailed information on sequences and PCR conditions are deposited in the Münster group, MPIZ.

8.4 Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Heinz Saedler betreut worden.

8.5 Lebenslauf

Name	Vanessa Quodt
Adresse	Emser Straße 2, 51105 Köln
Geburtsdatum	1. November 1976
Geburtsort	Köln
Nationalität	deutsch

Schulausbildung

1983-1986	Grundschule in Frechen-Königsdorf
1986-1987	Grundschule in Waldkraiburg (Bayern)
1987-1994	Ruperti-Gymnasium in Waldkraiburg bzw. Mühldorf
1994-1995	Stipendium für ein Jahr Auslandsaufenthalt mit Besuch der High School in Niagara Falls/ New York State, USA
1995	High School Graduation
1995-1997	Albertus-Magnus-Gymnasium Bensberg
Mai 1997	Allgemeine Hochschulreife

Studium

1997-2002	Studium der Biologie an der Universität zu Köln
2002-2003	Diplomarbeit am Botanischen Institut der Universität zu Köln, Betreuung durch Prof. Dr. U.-I. Flügge. Thema: Analyse des Phospho*enol*pyruvat-Stoffwechsels an *Arabidopsis thaliana PPT*-Mutanten
seit November 2003	Doktorarbeit am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung Molekulare Pflanzengenetik (Prof. Dr. H. Saedler). Thema: Molecular analyses of *Physcomitrella patens* MIKC^c-type MADS-box genes establish an evolutionary context in plants

Köln, den 13. Dezember 2006

Vanessa Quodt