Plant Cell Rep (2013) 32:1161–1177
DOI 10.1007/s00299-013-1411-8

ORIGINAL PAPER

A parsimonious model of lineage-specific expansion of MADS-box genes in *Physcomitrella patens*

E. I. Barker · N. W. Ashton

Received: 18 February 2013 / Accepted: 1 March 2013 / Published online: 23 March 2013

Abstract

Key message **The MADS-box gene family expanded in the lineage leading to the moss, *Physcomitrella patens*, mainly as a result of polyploidisations and/or large-scale segmental duplication events and to a lesser extent by tandem duplications.**

Abstract Plant MADS-box genes comprise a large family best known for the roles of type II *MIKC*^{C} genes in floral organogenesis, but also including type II *MIKC** genes, some of which have been implicated in male gametophytic development, and type I genes, a few of which are involved in ontogeny of female gametophytes, seeds and embryos. Genome-wide analyses of the MADS-box family in angiosperms have revealed numeric predominance of type I and *MIKC*^{C} genes and cross-species phylogenetic clustering of the Mα, Mβ and Mγ subtypes of type I genes and of 12 major subgroups of *MIKC*^{C} genes. The genome sequence of *Physcomitrella patens* has facilitated investigation of its full complement of 26 MADS-box genes, including 6 *MIKC*^{C} genes, 11 *MIKC** genes, seven type I genes and two pseudogenes. A much higher degree of similarity in sequence and architecture within the *MIKC*^{C} and *MIKC** gene subtypes exists in *Physcomitrella* than in *Arabidopsis*. Furthermore, MADS-box and K-box sequence is highly conserved between the *MIKC*^{C} and *MIKC** subgroups in *Physcomitrella*. Nine *MIKC** genes and two *MIKC*^{C} genes are located in pairs or triplets on individual DNA scaffolds. Phylogenetic gene clustering, gene architectures and gene linkages (directly determined from examination of the genome sequence) underpin a parsimonious model of two tandem duplications and three segmental duplication events, which can account for lineage-specific expansion of the MADS-box gene family in *Physcomitrella* from 4 members to 26. Two of these segmental duplication events may be indicative of polyploidisations, one of which has been postulated previously.

Keywords Physcomitrella · Moss · MADS-box gene complement · Comparative genomics · Gene duplication

Abbreviations

CLASP	Cytoplasmic linker associating protein
EST	Expressed sequence tag
ExPASy	Expert protein analysis system
JGI	Joint genome institute
MEF2	Myocyte enhancer factor 2
MEME	Multiple expectation maximisation for motif elicitation
ML	Maximum likelihood
MYA	Million years ago
NCBI	National center for biotechnology information
PIP	Plasma membrane intrinsic protein
RDP3	Recombination detection program version 3
SRF	Serum response factor
TAIR	The *Arabidopsis* information resource
WMP	Weighted maximum parsimony

Communicated by W. Harwood.

Electronic supplementary material The online version of this article (doi:10.1007/s00299-013-1411-8) contains supplementary material, which is available to authorized users.

E. I. Barker · N. W. Ashton (✉)
Department of Biology, University of Regina, Regina, SK S4S 0A2, Canada
e-mail: physcomitrella@gmail.com

Introduction

Availability of the *Physcomitrella patens* genome sequence (Rensing et al. 2008) has afforded an opportunity to examine the full complement of MADS-box genes in a member of the bryophytes, which diverged from the rest of the green plant lineage (after separation of the charophytes) at least 420 MYA (reviewed in Sanderson et al. 2004; Zimmer et al. 2007) and thus hold a phylogenetically informative position.

MADS-box genes comprise a large family of genes, characterised by a well-conserved MADS-box of approximately 180 bp, are found in plants, animals and fungi and encode transcription factors (for a recent review of plant MADS-box genes, see Gramzow and Theiβen 2010). In plants, MADS-box genes are best known for specifying floral organ identities and it has been argued that rapid expansion and diversification of this gene family were critical factors for evolution of angiosperms and the organs that define them (Theiβen et al. 2000).

Expansion of the seed plant MADS-box family has involved tandem duplication (copying of a gene in proximity to the original gene through unequal crossing over) and segmental duplication (copying and translocation of a lengthy DNA section or duplication of an entire genome [polyploidization]) (Pařenicová et al. 2003; Veron et al. 2007; Lee and Irish 2011). In addition, transposable elements are thought to have distributed copies of an *AG*-like MADS-box throughout the maize genome (Fischer et al. 1995; Montag et al. 1996).

Type I and II genes were originally classified on the basis of their proposed monophyletic relationships to animal *SRF*- (*SERUM RESPONSE FACTOR*-) like and *MEF2*- (*MYOCYTE ENHANCER FACTOR 2*-) like MADS-box genes, respectively (Alvarez-Buylla et al. 2000). However, more recent genome-wide analyses of MADS-box genes in *Arabidopsis* have not supported these relationships (De Bodt et al. 2003b; de Folter et al. 2004; Kofuji et al. 2003; Pařenicová et al. 2003). Nevertheless, an artificial, polyphyletic type I grouping (De Bodt et al. 2003b; Kofuji et al. 2003) and a monophyletic type II grouping (Kofuji et al. 2003) were distinguished by the absence or presence, respectively, of a conserved keratin-like (K-) box and by differences in exon–intron architecture. Type I genes in angiosperms usually contain one or two exons (De Bodt et al. 2003a; Pařenicová et al. 2003) and have been classified as Mα, Mβ or Mγ based on the phylogenetic analysis (Pařenicová et al. 2003).

A few type I MADS-box genes have been characterised functionally [Mα genes: *AGL62* (Kang et al. 2008), *DIANA* (*AGL61*) (Bemer et al. 2008; Steffen et al. 2008), *AGL23* (Colombo et al. 2008); Mγ genes: *PHERES1* (*PHE1*) (Köhler et al. 2003), *AGL80* (*FEM111*) (Portereiko et al. 2006)]. These genes play roles in the ontogeny of female gametophytes, embryos and seeds. Furthermore, all of the 38 *Arabidopsis* type I genes for which expression has been detected, using transgenic plants harbouring GUS-GFP reporter constructs, are active in female gametophytes and seeds (Bemer et al. 2010).

Type II genes encode proteins with the canonical MIKC structure consisting of the DNA-binding MADS domain, a weakly conserved intervening (I-) domain, the K-domain, which is predicted to form a coiled-coil structure, and a variable C-terminal domain (Ma et al. 1991; Theißen et al. 2000). An N-terminal domain may precede the MADS domain.

Type II genes are subdivided into $MIKC^C$ genes and *MIKC** genes on the basis of the expanded I region and less well-conserved K-box in *MIKC** genes (Svensson et al. 2000; Henschel et al. 2002). The $MIKC^C$ subtype includes many of the genes that control floral organogenesis. Forward genetics studies of angiosperms displaying homeotic floral phenotypes led to the well-known ABC model of floral morphogenesis (Coen and Meyerowitz 1991). Further investigation resulted in extension of this model to the ABCD model (Colombo et al. 1995), followed by the ABCDE and related protein-based, floral quartet models (Theiβen 2001; Theißen and Saedler 2001). $MIKC^C$ genes are also involved in floral meristem development, floral transition, senescence and abscission of flowers, embryonic development (Fernandez et al. 2000), leaf and root morphogenesis (Tapia-López et al. 2008), nodulation (Heard and Dunn 1995; Heard et al. 1997; Zucchero et al. 2001), and fruit development and dehiscence in flowering plants (for a summary of $MIKC^C$ functions, see Rijpkema et al. 2007) and development of reproductive structures in non-flowering spermatophytes (reviewed in Theiβen et al. 2000). The functions of $MIKC^C$ genes in cryptogams remain elusive. Relatively ubiquitous expression patterns have been observed in ferns (Hasebe et al. 1998; Münster et al. 1997, 2002), clubmoss (Svensson and Engström 2002) and moss (Singer et al. 2007; Quodt et al. 2007), suggesting that $MIKC^C$ gene functions are less organ-specific in non-seed plants than in seed plants. $MIKC^C$ gene knockouts show that functional redundancy characterises some members of this gene group in *Physcomitrella*, while gene knock-downs display a multifaceted mutant phenotype affecting both the gametophyte and sporophyte and implicating at least some $MIKC^C$ genes in reproductive functions (Singer et al. 2007).

Five of the *MIKC** genes in *A. thaliana* are expressed in pollen (Kofuji et al. 2003; Pina et al. 2005; Verelst et al. 2007a, b; Adamczyk and Fernandez 2009). *AGL66* and *AGL104* function redundantly in pollen germination and

their protein products form heterodimers with those of the remaining three genes. The sixth *MIKC** gene is expressed in siliques (de Folter et al. 2004). Little is known about functions of *MIKC** genes in other tissues. The expression of all six of the *Arabidopsis MIKC** genes in embryos, of five *MIKC** genes (all but *AGL67*) in inflorescences, four *MIKC** genes (excepting *AGL66* and *AGL104*) in seedlings (Lehti-Shiu et al. 2005) and *AGL67* in siliques (de Folter et al. 2004) suggests that *MIKC** functions are diverse in *Arabidopsis* and not restricted to male gametophytic tissue. In ferns, *MIKC** genes are expressed in both the gametophyte and the sporophyte generations (Kwantes et al. 2011). In the lycophyte, *Selaginella moellendorffii*, two of the *MIKC** genes are expressed exclusively in microsporangia while the third is also expressed in vegetative tissues. However, in another lycophyte, *Lycopodium annotinum*, the *MIKC** gene, *LAMB1*, is expressed exclusively in strobili, the reproductive structures of the sporophyte generation (Svensson et al. 2000). In the moss, *Funaria hygrometrica*, *MIKC** genes are expressed primarily in gametophytes, particularly protonemata (Zobell et al. 2010).

Here, we describe the sequences and architectures of the 26 genes that comprise the complete complement of MADS-box genes in *Physcomitrella*. We draw attention to the unusually high degree of conservation within and between the *MIKC*^{C} and *MIKC** subtypes and provide evidence that gene conversion has not played a significant role in maintaining sequence similarity. Using the tools of phylogenetic analysis, we have attempted to discern the evolutionary relationships among these genes as well as their relationships to MADS-box genes in other plant taxa. From our investigation of the scaffold locations of closely related MADS-box genes and neighbouring genes, we provide evidence of the gene duplications responsible for expansion of the MADS-box family in the bryophyte lineage leading to *Physcomitrella patens*.

Materials and methods

Identification and annotation of genes

MADS-box genes in the *Physcomitrella patens* genome were identified using the keyword "MADS" and the Advanced Search tool in the JGI (US Department of Energy's Joint Genome Institute) database. In addition, tblastn (Altschul et al. 1990) searches of JGI's database were performed using the default settings and, as queries, the amino acid sequences of each known MADS-box gene in *Physcomitrella* and of each novel gene as its sequence was discovered. Similar searches were performed to identify MADS-box genes in JGI's genome databases for *Ostreococcus lucimarinus* and *O. tauri* and the existence of one MADS-box gene in each species has been verified by Palenik et al. (2007). EST evidence for each gene was sought in Unigene and the Cosmoss *Physcomitrella* genome database.

Coding sequences of MADS-box genes were derived by virtual translation of the 4-kb nucleotide sequence downstream from the 5′ end of the MADS-box using the ExPASy translation tool (Gasteiger et al. 2003) and meticulous comparison of these DNA and amino acid sequences with JGI's predicted gene models, ESTs (expressed sequence tags) representing *Physcomitrella* MADS-box genes and also genomic sequences of already identified MADS-box genes of *Physcomitrella* and other plants. Following release of the Cosmoss v1.6 gene models, conserved N-terminal sequences were added to our MIKC^{C} sequences. In addition, motif searches were performed using the *Physcomitrella* sequences and various sets of representative MADS domain protein sequences from green algae and vascular plants as input for MEME, version 3.5.4 (Bailey and Elkan 1994). Exon–intron boundaries were determined by identifying splice sites in the genomic DNA sequences that conformed to the *Physcomitrella* consensus splice sites (Rensing et al. 2005) and that resulted in coding sequences that matched EST evidence and conserved motifs.

All 26 genes were annotated manually in the JGI database.

Sequence alignment and phylogenetic analysis

Sequences were aligned in Clustal W (Thompson et al. 1994) and adjusted manually by eye in MacClade (Maddison and Maddison 2001) where necessary. For phylogenetic tree construction, WMP and ML trees were constructed using PAUP* (Swofford 1998) and Bayesian analyses were performed using MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Model testing for Bayesian and ML analyses of DNA sequences was performed using Modeltest3.7 (Posada and Crandall 1998) or jModelTest (Guindon and Gascuel 2003; Posada 2008) and the best models were selected according to the Akaike Information Criterion. Burn-in for all Bayesian trees was 25 % of the samples. For all phylogenetic trees, gaps were treated as missing data. Bootstrap support and posterior probabilities are reported as follows: high ≥85 %, moderate 70–84 % or low 50–69 %. Branches with <50 % support were collapsed into polytomies.

Unrooted Bayesian and WMP trees were constructed using the 60 amino acid MADS domain sequences of 143 genes from representative, phylogenetically informative plant taxa including the full complement of MADS-box genes from *Physcomitrella* (except for the pseudogenes,

PPMA5 and *PPTIM6*), 99 genes from *Arabidopsis*, eight genes from the ferns, *Ceratopteris richardii* and *Ceratopteris pteroides*, five genes from the clubmoss, *Lycopodium annotinum*, one gene each from the spikemosses, *Selaginella remotifolia* and *S. moellendorffii*, the charophycean green algae, *Chara globularis, Coleochaete scutata,* and *Closterium peracerosum-strigosum-littorale*, and the chlorophyte green algae, *Ostreococcus lucimarinus* and *O. tauri*. For the Bayesian tree, the mixed model and the default settings except for nchains = 8 were used and three million generations were performed. For the WMP analysis, MaxTrees was set at 600 and support for the inferred tree was measured using 500 bootstrap replicates.

Sequences used for the *Physcomitrella* and *Arabidopsis* genes are available from JGI's database (See Electronic Supplementary Material S1 for Protein ID numbers) and The *Arabidopsis* Information Resource (TAIR) (A hyperlinked list of *Arabidopsis* MADS-box genes is available at http://www.arabidopsis.org/browse/genefamily/MADSlike.jsp), respectively. GenBank accession numbers for ferns, lycophytes and charophytes are as follows: *CRM1* (Y08014), *CRM3* (Y08239), *CMADS1* (U91415), *CMADS2* (U91416), *CMADS4* (U95609), *CerMADS1* (D89670), *CerMADS2* (D89671), *CerMADS3* (D89672), *LAMB1* (AF232927), *LAMB2* (AF425598), *LAMB3* (AF425599), *LAMB4* (AF425600), *LAMB6* (AF425602), *SrMADS1* (AB086021), *CgMADS1* (AB035567), *CpMADS1* (AB091476), *CsMADS1* (AB035568). A Mα gene (BXZC46793; ti/1415749262), from *S. moellendorffii*, was found by a blastn search of the whole genome sequence collection in the GenBank trace archive, using the MADS-box sequence of *PPTIM2* as query, and has since been named *MADS15* (Banks et al. 2011; Gramzow et al. 2012). Sequences for *OlTIM1* from *O. lucimarinus* (Protein ID 120540) and *OtTIM1* from *O. tauri* (Protein ID 38053) are from the respective JGI databases (http://genome.jgi.doe.gov/Ost9901_3/Ost9901_3.home.html and http://genome.jgi.doe.gov/Ostta4/Ostta4.home.html).

Because the relationships among the type II genes of *Physcomitrella* were not fully resolved in this comprehensive tree, separate rooted trees of *MIKC*^{C} and *MIKC** genes of *Physcomitrella* were constructed by WMP, Bayesian and ML methods. DNA sequences of *MIKC*^{C} and *MIKC** genes used for the respective trees included the complete coding sequences except for small portions of sequence that could not be unambiguously aligned. We used *Physcomitrella* type II genes in preference to genes from other taxa to root the trees since, in addition to the MADS box, a large portion of the I region and the extended K-box (Krogan and Ashton 2000) could be aligned unambiguously. This maximised the resolution and robustness of the trees. We chose two genes, one from each major clade of *MIKC** genes (*PpMADS2* and *PPM6*), and two genes, one from each major clade of *MIKC*^{C} genes (*PPM1* and *PPMC6*), to root the *MIKC*^{C} and *MIKC** trees, respectively.

For completeness and clarity of presentation, trees of *Physcomitrella* type I genes were similarly constructed. Type I trees were rooted with the sole MADS-box gene present in each of *O. lucimarinus* and *O. tauri*. All of the sequences used for the type I trees consisted of the most conserved middle portion of the MADS box, comprising 150 nucleotides.

Alignments are available upon request. In the Bayesian analyses 500,000 generations, 800,000 generations and 100,000 generations were performed for the *MIKC*^{C}, *MIKC** and type I trees, respectively. The robustness of each *Physcomitrella* gene tree was measured using 1,000 replicates for both WMP and ML. In the Discussion, we have substituted "closely related" for "close phylogenetic relatedness is inferred" to avoid repetitive unwieldy phrasing.

Detection of putative gene conversion events

Putative gene conversion tracts were sought with RDP3 software (Heath et al. 2006), which uses several methods to detect recombination: RDP (Martin and Rybicki 2000), BOOTSCAN (Martin et al. 2005), GENECONV (Sawyer 1989; Padidam et al. 1999), Maximum Chi-square (Smith 1992; Posada and Crandall 2001), CHIMAERA (Posada and Crandall 2001), sister scanning (Gibbs et al. 2000), and 3SEQ (Boni et al. 2007). Default settings were used.

Analysis of presumptive duplications

Physical separations and relative orientations of pairs of MADS-box genes were investigated to evaluate the significance of tandem and segmental duplications in *P. patens*. All genes located between linked MADS-box genes or within 50 kb segments flanking each MADS-box gene were identified, using JGI's Genome Browser page and linked Protein pages.

To investigate whether MADS-box genes in *Physcomitrella* may have been duplicated by transposition, the JGI Genome Browser was used to search for transposons near MADS-box genes. In addition, 8 kb of DNA flanking each MADS-box gene was searched manually for evidence of polyadenylate sequences that might indicate the involvement of non-viral retrotransposons or some form of reverse transcription.

The occurrence of a paleopolyploidisation event in *Physcomitrella* between 30 and 60 MYA was postulated by Rensing et al. (2007) on the basis of a clear peak in rates of synonymous substitution (K_s) in ESTs representing gene paralogues. The peak in K_s values was confirmed in a

similar study of genomic sequences of gene paralogues (Rensing et al. 2008). To identify MADS-box gene paralogues that may have been generated during the proposed polyploidisation period, K_s values were calculated for all pair-wise alignments of *MIKC*^{C} genes and *MIKC** genes using the method described by Rensing et al. (2007). Cutoff values of $0.5 < K_s < 1.1$ were chosen to encompass the ranges of K_s values that defined the peaks for the ESTs ($0.6 < K_s < 1.1$) and the genomic sequences ($0.5 < K_s < 0.9$). Because some pairs of type I genes could not be aligned unambiguously, K_s values were calculated only for pairwise alignments of genes that clustered closely together at the extremities of the phylogenetic tree.

Results

A search of JGI's database for *Physcomitrella* revealed twenty-six MADS-box genes of which one type I gene, five *MIKC*^{C} genes and six *MIKC** genes were known before sequencing of the genome (Electronic Supplementary Material S1).

MADS-box hits not accompanied by K-box hits were classified as type I genes. The type I genes were named *P. patens* Type I *MADS1-8* (*PPTIM1-8*). *PPTIM2* and *PPTIM3* were classified as Mα genes because they encode the motif **FSFGHPSIDYV**, which closely resembles the consensus sequence **YSFGHP(F)DAV** characteristic of Mα proteins in *Arabidopsis* and rice (De Bodt et al. 2003a; Pařenicová et al. 2003).

Novel *MIKC* MADS-box genes revealed by our searches were categorised as *MIKC*^{C} or *MIKC** by comparing their sequences and architectures with those of previously evaluated and classified *Physcomitrella* type II genes (Krogan and Ashton 2000; Henschel et al. 2002; Hohe et al. 2002; Riese et al. 2005; Singer et al. 2007). A previously unnamed *MIKC*^{C} gene and a novel gene were named *Physcomitrella patens MIKC*^{C}*5* (*PPMC5*) and *PPMC6*, respectively. The novel *MIKC** genes were termed *P. patens MIKC Asterisk5* (*PPMA5*), *PPMA8*, *PPMA9*, *PPMA10*, *PPMA11* and *PPMA12*. EST data are available for all the novel type II genes and three of the novel type I genes.

Virtual sequences comprising 138 amino acid residues beginning with the MADS domains of PPTIM6, PPTIM7 and PPTIM8 are 50 % identical although the corresponding DNA coding sequence of *PPTIM6* is interrupted by five in-frame nonsense codons. In addition, the MADS-box of *PPMA5* contains two putative insertions that perturb the translational reading frame. Scrutiny of the genomic sequences of *PPTIM6* and *PPMA5* failed to reveal potential splice sites that would allow the joining of conserved sense sequences. Thus, both genes were classified as pseudogenes. *PPTIM6* was included only in the duplication analysis and *PPMA5* was excluded from further analyses except where noted.

The Advanced Search tool in the JGI database yielded a 27th putative MADS-box gene (Protein ID 121924). Its amino acid sequence was only 17 % identical to the sequence of PPM1 when aligned in Clustal W (Thompson et al. 1994) and, when used as a query sequence in a tblastn search of the NCBI database, yielded no MADS-box gene hits. In addition, MEME did not detect a MADS domain in the sequence. Therefore, in our opinion, this gene is not a MADS-box gene and we did not investigate it further.

Conservation of type II MADS-box gene sequences and architectures in *Physcomitrella*

Amino acid residues in the MADS domains of type II proteins are identical at 35 of 60 positions (Fig. 1a). Conserved and semi-conserved substitutions (conservation of amino acid groups with strongly or weakly similar properties, respectively, as defined by Clustal W) exist at another 20 positions. The amino acid residues at 11 positions are perfectly conserved within the *MIKC*^{C} and *MIKC** subtypes but differ between the two. In addition, the two subtypes may be distinguished by their different motives at the C-terminal end of the MADS domain.

We have used a traditional definition of the extended K domain, although Kwantes et al. (2011) have provided some evidence that a large section of the I domain may have resulted from duplication of a portion of the K domain. The K domain sequences of MIKC^C and MIKC* proteins are identical at 14 of 89 positions and display conserved or semi-conserved substitutions at an additional 23 positions (Fig. 1b). The motif **RVRARK** in the K domain is identical in 15 of the 17 type II proteins and differs at only one position in the other two. The amino acid residues at 10 positions are identical within the *MIKC*^{C} and *MIKC** subtypes but differ between the two groups.

The positions of hydrophobic amino acid residues in the heptad repeats of K1, K2 and K3 and of two hydrophobic amino acid residues, which lie outside the heptad repeats but are important for protein interactions in SEP3 (Kaufmann et al. 2005), are identical in the six *Physcomitrella* MIKC^C sequences (Fig. 1b). The pattern of hydrophobic amino acid residues in K1 and K2 of PI (Kaufmann et al. 2005) is identical to that in the *Physcomitrella* MIKC* sequences. However, the positions of hydrophobic amino acid residues in K3 and of the other two hydrophobic amino acid residues mentioned above are not conserved in the *Physcomitrella* MIKC* proteins.

Proteins encoded by *PpMADS-S*, *PPMC5* and *PPMC6* lack the motif, **NRLHANIS/LPSVRI**, corresponding to DNA at or very near the 3′ end of the coding sequence of the

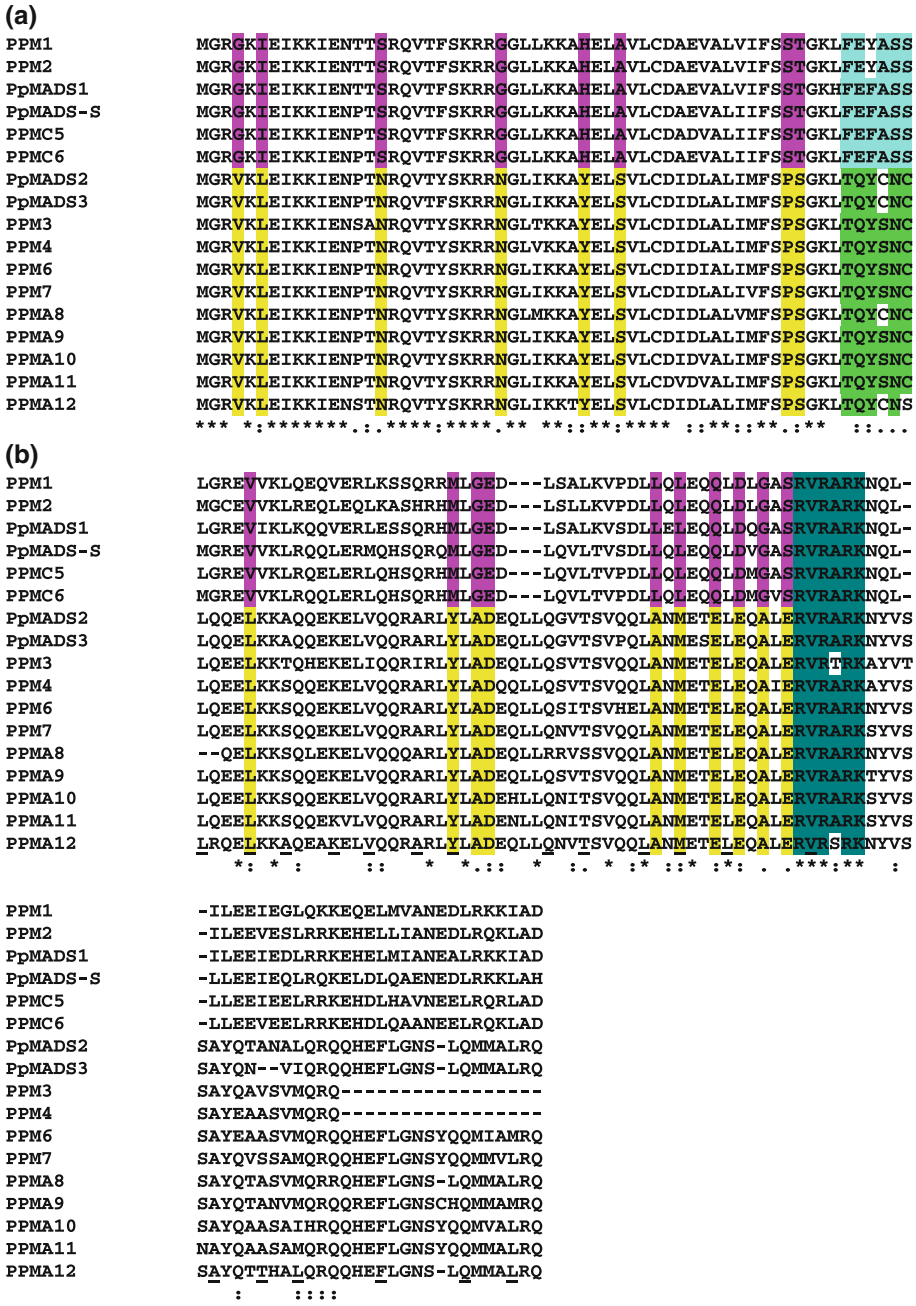

Fig. 1 Multiple sequence alignment of type II MADS domains (**a**) and extended K domains (**b**) in *P. patens*. The pseudogene *PPMA5* has been omitted. Highlighting indicates the first eight conserved amino acid residues that differ between MIKC^C proteins (*pink*) and MIKC* proteins (*yellow*). A six-amino acid motif at the C-terminal end of the MADS domain makes distinguishing the sequences of MIKC^C (*blue*) and MIKC* (*green*) proteins facile. Another six amino acid motif (*teal*) at the C-terminal end of the K domain is almost perfectly conserved in all type II proteins. Amino acid residues occupying positions corresponding to the a and d positions of heptad repeats in *Arabidopsis* MIKC proteins or identified as important for protein interactions (Kaufmann et al. 2005) are underlined in the sequence of PPMA12

other three *MIKC^C* genes. Interestingly, however, vestiges (underlined in the sequences below) of this motif may be discernible by virtually translating portions of sequence sequestered in the flanking 3′ untranslated regions (UTRs) of these genes. Thus, ignoring the last nucleotide before the stop codons in *PpMADS-S*, *PPMC5* and *PPMC6* and continuing translation to the next available stop codon in each 3′ UTR generates the motifs, **NRVHANFP**, **NRLHAIFPPQGK-QYKLHCSFGE** and **NRLHATFQPRGK,** respectively.

Exon–intron architecture is highly conserved in the *Physcomitrella* MADS-box genes. The *MIKC^C* genes contain 9 exons except for *PPMC6*, which lacks Intron 5 (Fig. 2). Furthermore, intron phases are identical in all six *MIKC^C* genes with the exception of Intron 8 and the missing intron in *PPMC6* (Electronic Supplementary Material S2).

In the *MIKC** genes, one I-region exon is absent in *PPM3* and *PPM4* (Fig. 2) as noted by Henschel et al. (2002). Intron 7 is missing in *PPM6*, *PPM7* (Riese et al. 2005), *PPMA9*, *PPMA10* and *PPMA11*. The first exonic sequence in the C-terminal region is absent in *PPM3* and *PPM4*, fused to the K-box in *PpMADS2*, *PpMADS3*, *PPM6*, *PPMA8*, *PPMA9* and *PPMA12*, and continuous with the next downstream exonic sequence in *PPM7*. In *PpMADS2*

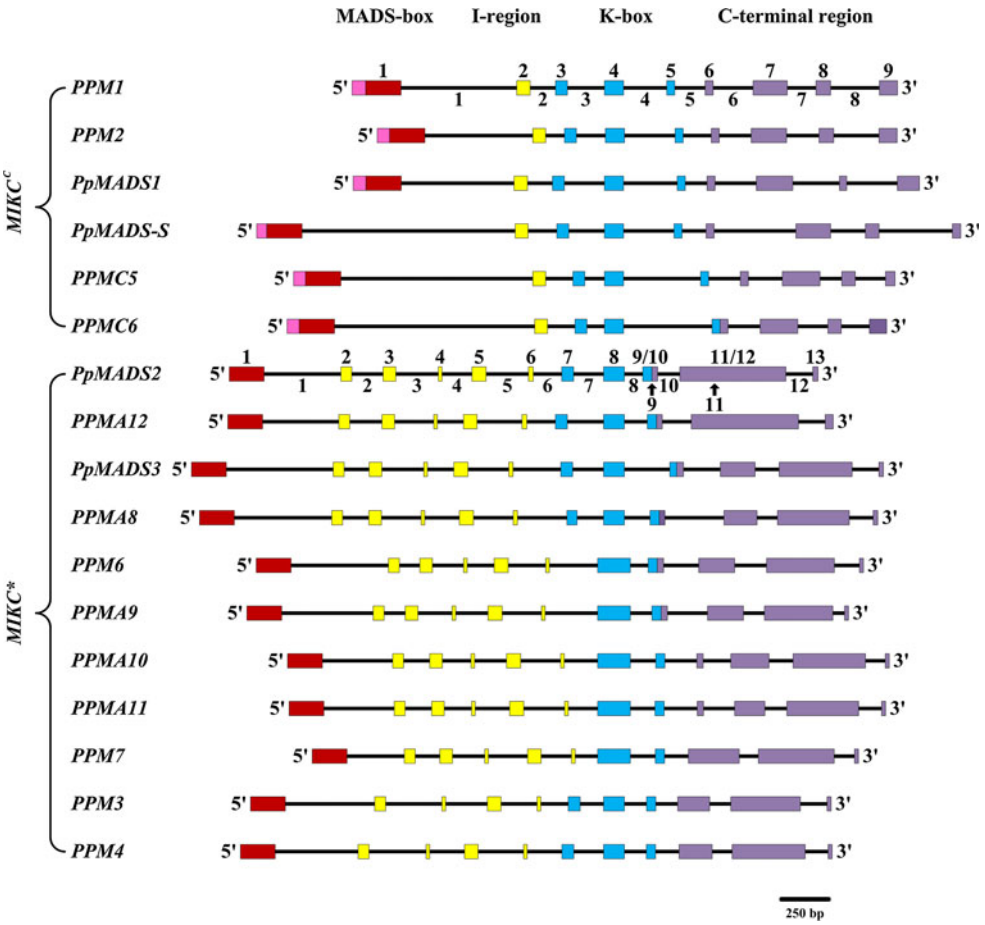

Fig. 2 Architecture of the type II MADS-box genes in *P. patens*. Coloured rectangles (N-terminal region, *pink* MADS-box, *red* I-region, *yellow* K-box, *blue* C-terminal region, *purple*) and *black lines* represent exons and introns respectively. Exons and introns are numbered above and below, respectively, *PPM1* and *PpMADS2*. The position of Intron 9 and the approximate position of Intron 11, both absent in *PpMADS2*, are indicated by *arrows*

and *PPMA12*, one continuous exonic sequence corresponds to the two long C-terminal exons in the remaining *MIKC** genes. Excluding the fact that certain introns are absent in some of the *MIKC** genes, all intron phases are conserved in *MIKC** genes and positions of introns are conserved except that, when compared with the other *MIKC** genes, the position of Intron 11 differs by a single codon in *PPM6* and *PPMA9*. Similarly, the position of Intron 12 differs by one codon in *PpMADS2*, *PpMADS3*, *PPMA8* and *PPMA12* (Electronic Supplementary Material S3).

Phylogenetic analyses

Our comprehensive (multi-taxon) Bayesian tree (Electronic Supplementary Material S4) was consistent overall with trees published by others (see, for example, Becker and Theiβen 2003; Kofuji et al. 2003; Pařenicová et al. 2003, Gramzow et al. 2012). The *MIKC*^{C} genes from *Physcomitrella* formed a single cluster supported by a high posterior probability. The *Physcomitrella MIKC** genes clustered together in a highly supported group within a moderately supported larger cluster that included the *MIKC** genes from *Arabidopsis*. *LAMB1* appeared separately from the other genes.

The *Physcomitrella* type I genes were grouped into three separate clusters. *PPTIM1* with *PPTIM4* and *PPTIM5* formed a cluster supported by a high posterior probability. *PPTIM2* and *PPTIM3* formed a group that clustered, with a high posterior probability, with the *S. moellendorffii* gene *MADS15* as a sister, within a cluster that included the majority of the Mα genes from *Arabidopsis*. The posterior probability for the larger cluster was low, however, and the remaining Mα genes from *Arabidopsis* formed a separate group. *PPTIM7* and *PPTIM8* clustered with all but two of the Mβ genes from *Arabidopsis* in a cluster within a larger cluster that included the remaining Mβ genes and all of the Mγ genes from *Arabidopsis* as well as the unclassified type I gene, *AGL33*. The posterior probability for this cluster was low.

The comprehensive WMP tree was generally consistent with the Bayesian tree although the WMP tree provided less resolution and the majority of the bootstrap values were numerically lower than the Bayesian posterior probabilities. The *Physcomitrella MIKC** genes formed a cluster with moderate bootstrap support, separate from the *Arabidopsis MIKC** genes. *PPTIM2* and *PPTIM3* formed a cluster with *SmMADS15*, with high bootstrap support, but the Mα genes from *Arabidopsis* formed several separate

clusters. *PPTIM7* and *PPTIM8* formed a cluster within a larger cluster that included, with low bootstrap support, all of the *Arabidopsis* Mβ and Mγ genes.

The topologies of the WMP, Bayesian and ML trees of *Physcomitrella* genes were identical for the *MIKC** genes but somewhat different for the *MIKC*^{C} genes and the type I genes. The *MIKC** genes (Fig. 3a) were resolved, with high support, into two main clades each of which contained smaller, highly supported clades. Thus, one of the main clades comprised two smaller clusters, (*PpMADS2*, *PPMA12*) and (*PpMADS3*, *PPMA8*), and the other incorporated two subclades, the first one containing *PPM6* and *PPMA9* and the second containing two smaller clusters. One of these comprised *PPM3* and *PPM4* and the other included *PPM7* as sister to a clade containing *PPMA10* and *PPMA11*. A second WMP analysis (not shown), including the pseudogene, *PPMA5*, its sequence having been aligned with the other *MIKC** sequences by manually correcting for two presumed indels in the MADS domain, produced a tree with a subclade consisting of *PPMA5* and *PPM7* and otherwise identical topology.

In the WMP tree, the *MIKC*^{C} genes were resolved into two highly supported clusters, (*PPM1*, *PpMADS1*, *PPM2*) and (*PpMADS-S*, *PPMC5*, *PPMC6*) (Fig. 3b). Furthermore, the former clade contained a highly supported subclade consisting of *PPM1* and *PpMADS1* and the latter clade also included a highly supported subclade comprising *PPMC5* and *PPMC6*. The Bayesian and ML trees were consistent with the WMP tree with respect to the first clade, although support was moderate or low for some branches. However, in the Bayesian tree, the second clade and the *PPMC5-PPMC6* subclade within it were moderately supported. In the ML tree, *PpMADS-S*, *PPMC5* and *PPMC6* were unresolved.

The *Physcomitrella* type I gene trees (Fig. 3c) revealed identical relationships to those seen in our comprehensive tree (Electronic Supplementary Material S4) except that, in the WMP tree, the clade containing the Mα genes, *PPTIM2*

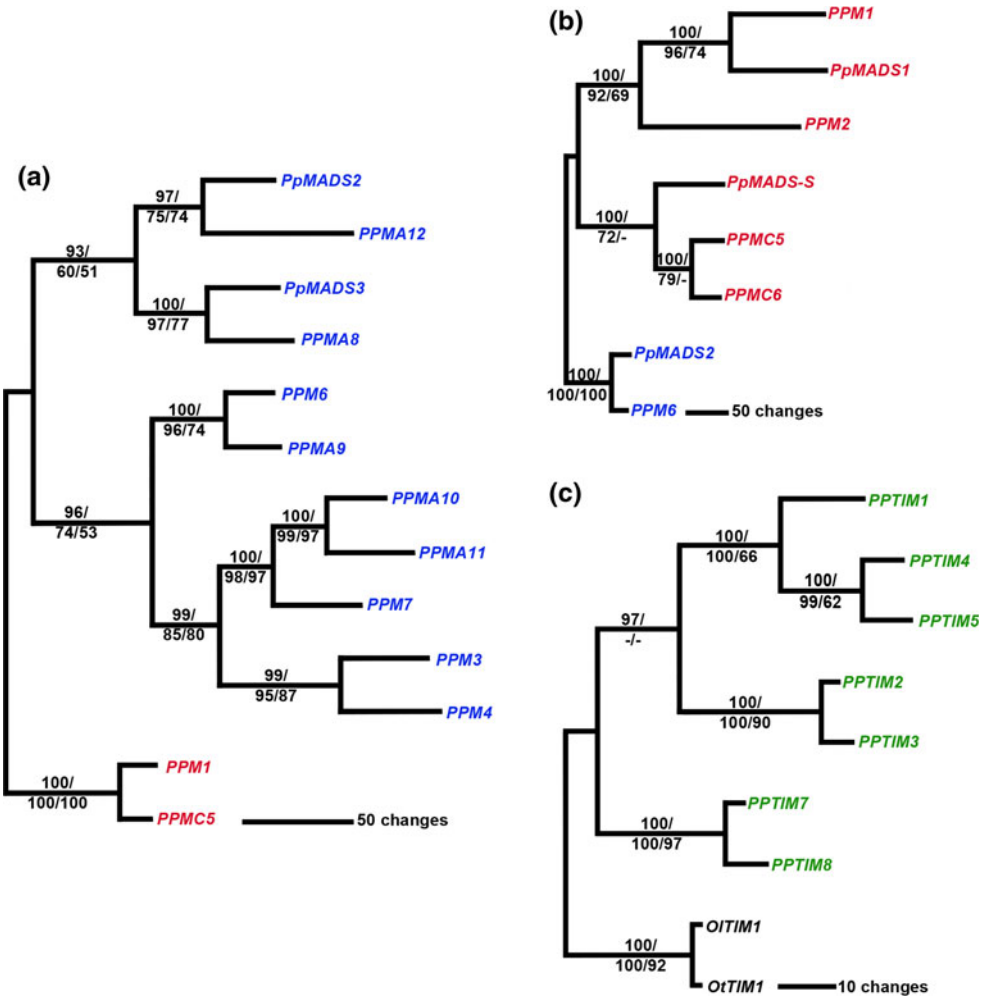

Fig. 3 Rooted weighted maximum parsimony (WMP) trees of the major groups of *P. patens* MADS-box genes: **a** *MIKC**, **b** *MIKC*^{C} and **c** type I. *MIKC**, *MIKC*^{C} and type I genes are shown in *blue*, *red* and *green* respectively. Bootstrap values and posterior probabilities from the WMP, Bayesian and ML trees are shown *top to bottom*, *left to right*

and *PPTIM3* was sister to the clade containing *PPTIM1, PPTIM4* and *PPTIM5*.

Duplications

One triplet and four pairs of type II MADS-box genes are located on five DNA scaffolds (Fig. 4a), with a combined length of approximately 10.9 Mb, from a total of 2,106 scaffolds, corresponding to approximately 480 Mb. Pairs comprising linked *MIKC** genes are separated by a minimum of 6 kb and a maximum of 24 kb. Two pairs each consist of a $MIKC^C$ gene and a *MIKC** gene, separated by 83 and 224 kb.

The type I genes, *PPTIM4* and *PPTIM5*, are also physically linked in a tail-to-tail arrangement with approximately 3 kb separating their respective MADS-boxes (Fig. 4b).

PPTIM2 and *PPTIM3* are located in syntenic arrangements with four other genes encoding a mitochondrial

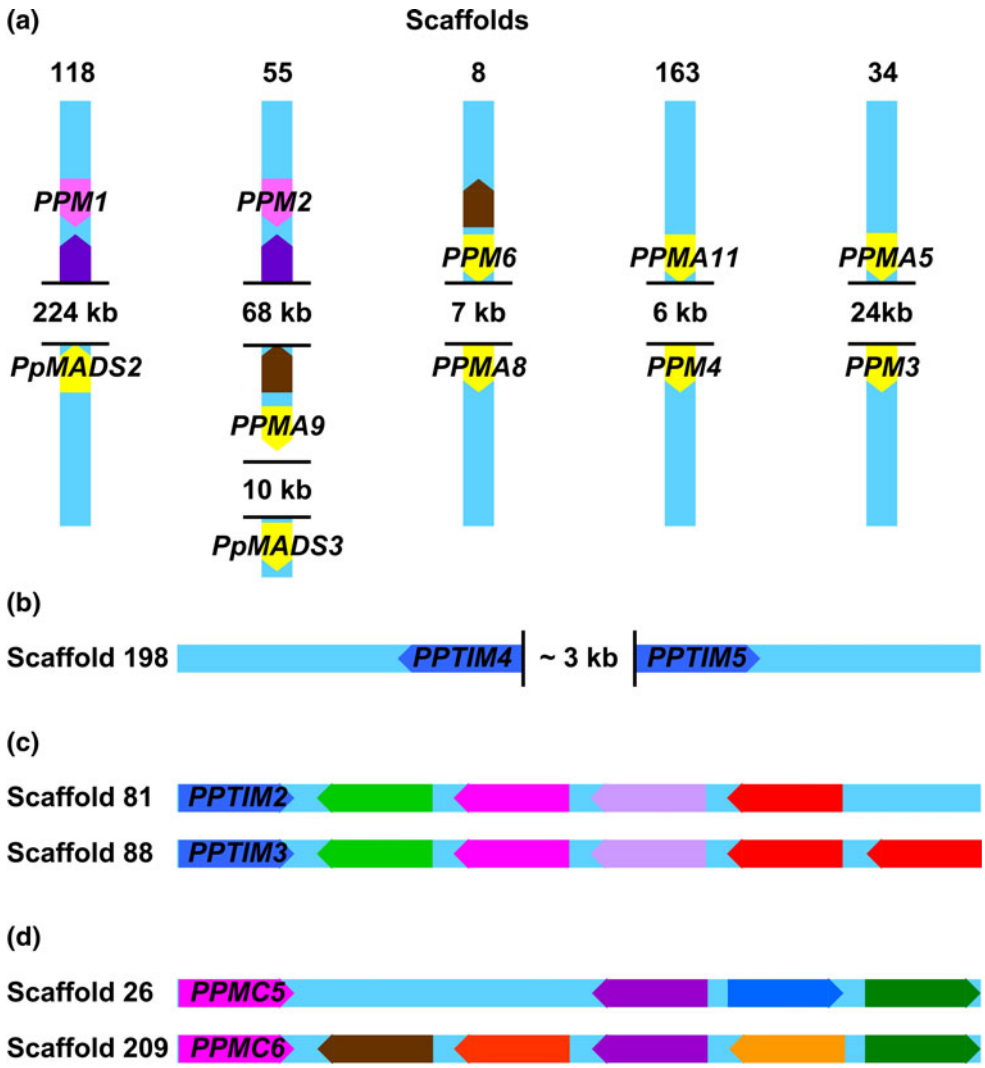

Fig. 4 Evidence for tandem and segmental duplications in *P. patens*. *Light blue* rectangles represent segments of scaffolds containing duplicated MADS-box genes and linked genes (pentagons pointing in the 5′ → 3′ direction). Spaces between the *black lines* represent DNA of the lengths indicated. (Diagrams are not to scale.) **a** Locations of one triplet and four pairs of type II MADS-box genes on five scaffolds. $MIKC^C$ genes are shown in *pink* and *MIKC** genes in *yellow*. *PIP* genes and *CLASP* genes are shown in *purple* and *brown* respectively. **b** Segment of DNA containing *PPTIM4* and *PPTIM5*, which are oriented in opposite directions with approximately 3 kb separating their MADS-boxes. **c** *PPTIM2* and *PPTIM3* (*dark blue*) and regions (approximately 27 and 46 kb downstream from *PPTIM2* and *PPTIM3* respectively) containing genes encoding mitochondrial transcription termination factors (*green*), two similar predicted proteins of unknown family (*pink*), Trehalose-6-phosphate synthase components TPS1 and related subunits (*purple*), and the catalytic subunits of a Serine/Threonine protein phosphatase 2A (*red*). (d) *PPMC5* and *PPMC6* (*pink*) and segments (approximately 37 and 19 kb downstream from *PPMC5* and *PPMC6* respectively) containing genes encoding DDHC-type zinc-finger proteins (*purple*), RNA binding proteins (*green*), a GTP-binding signal recognition particle SRP54 (*blue*), an HNH endonuclease (*brown*), an asparaginase (*red*) and an unknown predicted gene (*orange*)

transcription termination factor, an unclassified predicted protein, the trehalose-6-phosphate synthase component TPS1 and related subunits, and the catalytic subunit of serine/threonine protein phosphatase 2A within approximately 27 and 40 kb, respectively (Fig. 4c). A duplicate of the fourth gene is located immediately downstream from the first copy on the scaffold containing *PPTIM3*. Similarly, duplicate sets of DDHC-type zinc-finger genes and RNA binding protein encoding genes are linked, within 35 kb, to *PPMC5* and *PPMC6* in the same order and relative orientations (Fig. 4d).

Search for transposable elements located within or near MADS-box genes

No transposable element was found overlapping a MADS-box gene and no putative DNA transposase or helitron was detected on any scaffold containing a MADS-box gene. No polyadenylate sequence was found within 500 nucleotides (the approximate maximum length of a SINE) on either side of a MADS-box gene or within 8 kb (the approximate maximum length of a LINE) and accompanied by a putative reverse transcriptase gene.

Discussion

Gene duplication

Gene families expand in number and roles by tandem and segmental duplications with subsequent nonfunctionalisation (pseudogenisation) of one copy of each duplicate pair or retention of both genes (Ohno 1970; reviewed in Zhang 2003). Duplicate copies may be preserved as functionally redundant genes resulting in increased amounts of gene product (Kondrashov and Koonin 2004) or, in the case of large-scale duplications, such as diploidisation, preservation of the stoichiometry of dimerisation or complexing of the gene products (Lynch and Conery 2000). Alternatively, duplicate genes may diversify in sequence and expression with concomitant neofunctionalisation (Ohno 1970) or they may partition the functions of the ancestral gene (subfunctionalisation) (Force et al. 1999).

PPTIM4 and *PPTIM5* are tandemly arrayed genes (Fig. 4b) that are closely related (Fig. 3c), suggesting that they are the result of tandem duplication. Conversely, pairs of MADS-box genes, in some cases linked to homologues of other genes, appear to have been copied during whole genome duplication or other large-scale segmental duplication events. The synteny involving *PPTIM2* and *PPTIM3* and linked homologues of four other genes on scaffolds 81 and 88, respectively, implies that these linkage groups arose by segmental duplication.

Although the subclade comprising *PPMC5* and *PPMC6* with *PpMADS-S* as sister is strongly supported only in the WMP tree (Fig. 3b), other evidence corroborates these relationships. The synteny surrounding *PPMC5* and *PPMC6* indicates that they are a duplicate gene pair (Fig. 4d). In addition, in *PpMADS-S*, *PPMC5* and *PPMC6* the first intron is significantly longer (1,005–1,116 bp) than the corresponding intron in *PPM1*, *PPM2* and *PpMADS1* (565–652 bp). Finally, the three genes of the former clade all possess nonsense (translation stop) codons at the same upstream position relative to genes of the latter clade (Electronic Supplementary Material S2). Because the only putative gene conversion tract detected by RDP3 consisted of the first 80 amino acids in PpMADS-S and PPMC6, tree construction may have been confounded by this tract.

In most instances, the component genes of physically linked *MIKC* gene pairs (Fig. 4a) are more closely related to genes within one or several other linked pairs (on different scaffolds) than they are to each other (Fig. 3a, b), suggesting that the genes have been duplicated together during segmental duplication events. For example, the linked genes, *PPMA9* and *PpMADS3*, are closely related phylogenetically to the linked genes, *PPM6* and *PPMA8*, respectively. *PPM3*, which is linked to the pseudogene *PPMA5*, is closely related to *PPM4*, itself linked to *PPMA11*.

Similarly, *PPM2* and *PpMADS3* are linked genes which are closely related to the linked genes *PPM1* and *PpMADS2*, respectively. Genes encoding plasma membrane intrinsic protein (PIP) subfamily aquaporins, *PpPIP2;4* and *PpPIP2;2*, are situated within 8 and 22 kb, respectively of the $MIKC^C$ genes, *PPM1* and *PPM2* (Fig. 4a). Interestingly *PpMADS1*, the gene most closely related to *PPM1* and *PPM2*, is also linked to a nearby PIP gene, *PpPIP2;3*. A second copy of *PpPIP2;4* is located approximately 27 kb upstream from the first copy. These are four of the five *PIP* genes that comprise one of three clades of *PIP* genes in *Physcomitrella* (Danielson and Johanson 2008).

PpMADS2 is oriented in the opposite direction from *PPM1*, *PPM2*, *PPMA9* and *PpMADS3* and is separated from *PPM1* by approximately 224 kb, whereas the distance between *PpMADS3* and *PPM2* is only approximately 92 kb. Therefore, *PpMADS2* plus a flanking DNA segment were probably inverted during or subsequent to the duplication. To investigate this possibility, genes located in the 100 kb segment immediately 5′ to *PpMADS2* were identified and compared with the genes situated between *PPM2* and *PpMADS3* (data not shown). No similarity was found indicating either the initial chromosomal rearrangement was not a simple inversion or subsequent structural reorganization destroyed or relocated the expected synteny.

Recent peaks of retrotransposon activity have occurred in *Physcomitrella* (Rensing et al. 2008). However, we found no evidence that duplication of MADS-box genes had occurred by transposition of any kind. Although it remains possible that (retro) transposon-mediated duplication of MADS-box genes in *Physcomitrella* occurred and can no longer be detected, we suggest that duplication of MADS-box genes by unequal crossing over between repetitive DNA elements, possibly including transposons, within the genome and polyploidisation are more likely explanations.

The K_s values of most pairs of closely related MADS-box genes in *Physcomitrella* (Electronic Supplementary Material S5) fall within the range of K_s values ($0.5 < K_s < 1.1$) representing the polyploidisation period proposed by Rensing et al. (2007), suggesting that these sets of gene duplicates may have been generated during this event. Since the K_s values for two of three pairwise comparisons of genes in each of the two major clades of $MIKC^C$ genes and in all three pairwise comparisons of genes in the clade comprising *PPM7*, *PPMA10* and *PPMA11* are within this range, it is possible that a large scale segmental duplication occurred just before or very soon after the proposed diploidisation.

Model of MADS-box gene duplication in *Physcomitrella*

We propose a parsimonious model of two tandem and three segmental duplications, which can account for the expansion of the MADS-box gene family from 4 members to 26 in the *P. patens* lineage. In our model (Fig. 5), the names of extant genes have been used for the sake of simplicity, but it should be noted that the genes, which were actually duplicated, are the closest ancestors of those named.

A plausible sequence of events is that a $MIKC^C$ gene, a *MIKC** gene, a Mα and a Mβ gene, existing prior to the divergence of the bryophytes and the tracheophytes, passed into the *P. patens* lineage as *PPM2*, *PpMADS3*, *PPTIM2* and *PPTIM7*, respectively. The following steps then occurred in sequence.

Step 1

Tandem duplication of *PpMADS3* gave rise to *PPMA9* and, sometime later, Intron 7 of *PPMA9* was lost.

Step 2

Segmental duplication of *PPM2*, *PPMA9* and *PpMADS3* gave rise to *PpMADS-S*, *PPMA11* and *PPM4*. During or following the duplication, chromosomal rearrangement separated *PpMADS-S* from *PPMA11* and *PPM4*. *PPMA11* gained Intron 9 and *PPM4* lost Exon 3 and Exon 10. An indel resulted in introduction of a stop codon and truncation of the coding sequence of *PpMADS-S*. *PPTIM2* and *PPTIM7* were not copied during this step or their duplicates were subsequently lost.

Step 3

A second, large-scale segmental duplication resulted in copying of *PPM2* and *PpMADS3* to give rise to *PPM1* and *PpMADS2*. During this step or subsequently, a lengthy DNA segment containing *PpMADS2* was inverted. *PpMADS2* lost Intron 11 at some point. *PPMA9* either was not duplicated or its duplicate was lost, possibly during this inversion. *PpMADS-S* was duplicated giving rise to *PPMC5*. *PPMA11* and *PPM4* were also copied to produce *PPMA5* (which subsequently lost Intron 10) and *PPM3*. *PPTIM2* was duplicated resulting in *PPTIM1*, which subsequently diverged in sequence such that it is no longer recognisable as an Mα gene. *PPTIM7* was duplicated giving rise to *PPTIM6*, which degenerated, becoming a pseudogene.

Step 4

In a third segmental duplication, possibly the polyploidisation proposed by Rensing et al. (2007), *PPMA9* and *PpMADS3* were duplicated, giving rise to *PPM6* and *PPMA8*. *PPM1* was copied to produce *PpMADS1*. *PpMADS2* was duplicated giving rise to *PPMA12*. In addition, *PPMA11* was duplicated to give rise to *PPMA10*. *PPMC5* was copied, producing *PPMC6*, which subsequently lost Intron 5. *PPMA5* was duplicated to produce *PPM7*. Later, *PPMA5* lost Exon 2 and deteriorated further, becoming a pseudogene through the introduction of frameshifts caused by indels. *PPTIM2* was copied to give rise to *PPTIM3* and duplication of *PPTIM1* produced *PPTIM5*. *PPTIM7* was duplicated giving rise to *PPTIM8*.

Step 5

PPTIM5 gave rise to *PPTIM4* by means of a recent tandem duplication (not shown in Fig. 5).

Our phylogenetic tree of *MIKC** genes does not display a node representing a hypothetical gene that is ancestral to *PpMADS3* and *PPM4* but not to *PPMA9* (Fig. 3a) and, thus, does not support our model with respect to Step 2. In sharp contrast, the close linkages of *PPMA9* to *PpMADS3* and *PPMA11* to *PPM4* as well as the presence of an intron in *PPM4* and *PPM3*, which is also present in *PpMADS3* and its proposed descendants but is lacking in *PPMA9* and its putative descendants do support Step 2.

This model of MADS-box gene expansion in the *Physcomitrella* lineage from 4 to 26 genes is parsimonious,

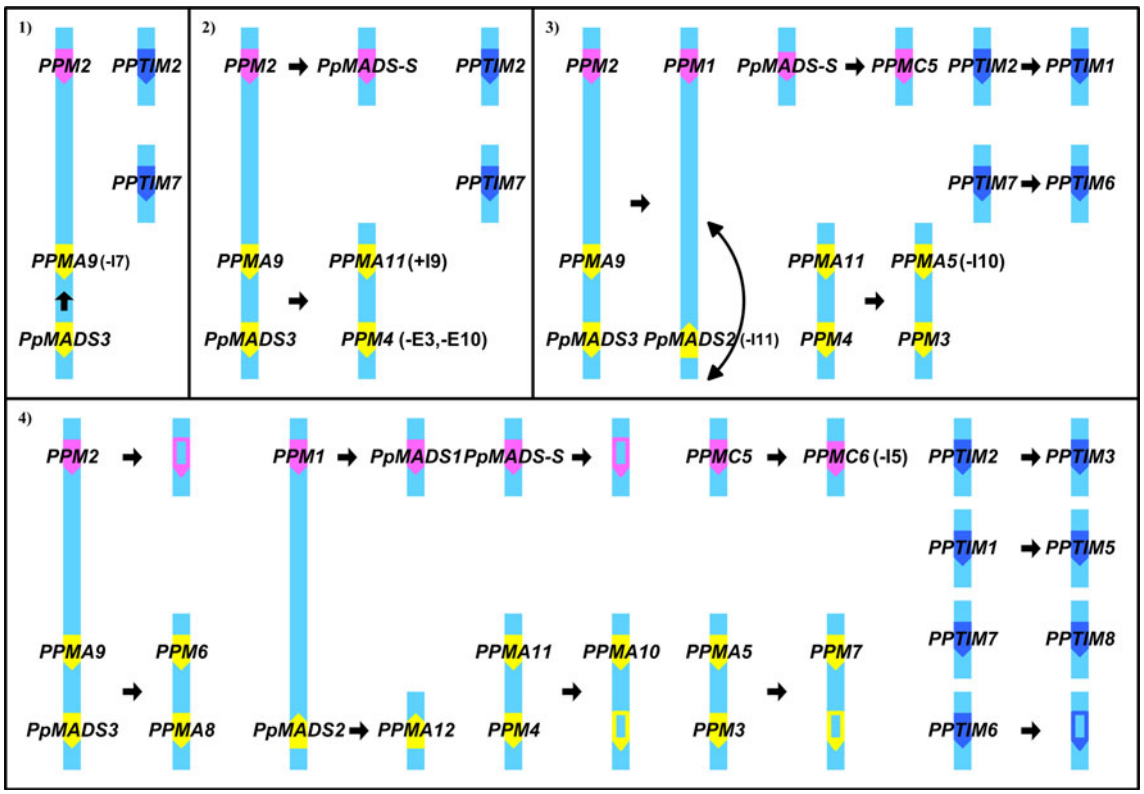

Fig. 5 A parsimonious duplication model rationalising expansion of the MADS-box gene family. All *MIKC*^{C} genes (*pink*), *MIKC** genes (*yellow*) and type I genes (*blue*) existing at the end of each of steps 1 through 4 (proposed diploidisation) are shown in the corresponding numbered panels. Proposed gene losses are shown as pentagonal outlines. *Horizontal* and *vertical arrows* represent segmental and tandem duplications respectively; the *curved arrow* represents the proposed inversion of a DNA segment. Losses (-) and gains (+) of individual exons (E) and introns (I), denoted by numbers, are indicated in parentheses after gene names

requiring only five steps: a tandem duplication followed by three multigene segmental duplications (the last of which is consistent with polyploidisation) and a recent event, in which a single gene was copied. Overall agreement of evidences from phylogenetic trees, chromosomal linkages and gene architectures indicates that our model is robust. A small discrepancy is that the K_s values for one *MIKC*^{C} and two paralogous *PPTIM* gene pairs generated in Step 4, namely *PPM1-PpMADS1* ($K_s = 0.48$), *PPTIM7-PPTIM8* ($K_s = 0.40$) and *PPTIM1-PPTIM5* ($K_s = 3.0$) fall, respectively, very slightly, slightly and significantly outside the range of values corresponding to the polyploidisation period proposed by Rensing et al. (2007). However, all other *MIKC*^{C} and *MIKC** paralogues and the *PPTIM2-PPTIM3* gene pair produced in Step 4 have K_s values within the expected range. Moreover, the K_s value for *PPTIM4-PPTIM5* is low (0.49), consistent with this paralogous gene pair being produced by a recent duplication as proposed in our model. It should be noted that K_s values >1.0 are generally interpreted cautiously because they are error-prone due to the occurrence of multiple synonymous substitutions at each synonymous site (Blanc and Wolfe 2004).

This model is seductive because of its simplicity and since it implies that Step 3 may represent a second polyploidisation. If it was not a polyploidisation, the ancestral genes duplicated within it must have been linked, or duplicated more or less simultaneously during a burst of transposon activity (which we think is unlikely), in order for Step 3 to be considered a single event. A less attractive option is that Step 3 is a collection of non-simultaneous but sequentially equivalent duplications, which have in common only that they preceded the polyploidisation hypothesized by Rensing et al. (2007).

Based on an estimate of 172 million years for the age of the Funariidae (Newton et al. 2007) and chromosome numbers reported for *Funaria* (4, 14, 21, 28, 42, 56) and *Physcomitrella* (14, 27, 28), Rensing et al. (2007) proposed that independent polyploidisation events have occurred in the Funariaceae and that the whole genome duplication in the *Physcomitrella* lineage probably occurred after speciation among the Funariaceae. However, pairwise orthology between eleven *MIKC** genes in *Funaria hygrometrica* and eleven *MIKC** genes in *Physcomitrella* (Zobell et al. 2010) provides compelling evidence that expansion of the *MIKC** gene complement to 11 genes occurred before divergence

of *Funaria* and *Physcomitrella*. If whole genome duplication occurred in the *Physcomitrella* lineage after speciation, 22 *MIKC** genes should have resulted. The pseudogene *PPMA5* might be the product of subsequent deterioration of one of these genes, still leaving 10 genes unaccounted for. Therefore, we presume that the polyploidisation proposed by Rensing occurred in a common ancestor of the two moss genera.

Functional significance of gene duplication in *Physcomitrella*

Sequence conservation

Type II MADS-box genes in *Physcomitrella* are highly conserved in both sequence and architecture. Clustal W alignments reveal identity of 35 amino acid residues in the MADS domain and 14 in the K domain of *Physcomitrella* type II proteins (Fig. 1). In sharp contrast, amino acid residues are identical at seven positions in the MADS domain of type II proteins in *Arabidopsis* and sequence identity is not found at any position outside the MADS domain (not shown). A possible explanation of these observations is that gene conversion has occurred frequently within the MADS-box gene family in *Physcomitrella*. However, recombination detection software failed to provide evidence that this is the case.

Conservation of gene sequences and EST evidence suggest that the majority of MADS-box genes in *Physcomitrella* are functional. However, the significance of retention of highly similar MADS-box gene homologues is unclear since, in general, duplicate transcription factor genes appear to have been preferentially retained following whole genome duplications in *Arabidopsis*, but not in *Physcomitrella* (Rensing et al. 2007 and references within).

According to the gene dosage hypothesis, duplicate genes that are retained in a genome provide an enhanced gene dosage effect that is beneficial to the organism (Kondrashov and Koonin 2004). Alternatively, the gene balance hypothesis predicts that duplicates of genes that encode interacting proteins are preferentially retained after a large-scale duplication event such as polyploidisation to preserve the stoichiometry of interaction (reviewed in Birchler and Veitia 2007). However, the results from gene knockouts, albeit involving a limited number of *MIKC* genes, suggest that gene dosage and/or gene balance cannot be the only explanations for retention of duplicated *MIKC* genes in the *Physcomitrella* genome (Singer et al. 2007; Singer and Ashton 2009). A third possibility is that the retention of duplicated gene copies may contribute to robustness (Gu et al. 2003; Félix and Wagner 2008). Evidence exists for selective retention of the *SEPALLATA 1* (*SEP1*)-*SEP2* and *SHATTERPROOF 1* (*SHP1*)-*SHP2* duplicate pairs of MADS-box genes in *Arabidopsis* (Moore et al. 2005). In addition, retention of duplicate genes may allow for the evolution of differential expression and/or an expanded repertoire of protein complexes, thereby contributing to morphological elaboration (Kaufmann et al. 2005; Veron et al. 2007). Future investigation of MADS domain protein interactions in *Physcomitrella* will be particularly interesting since it holds the prospect of revealing parallels between retention of pairs of duplicated genes and patterns of co-expression and protein dimerisation.

Evolution of the MADS-box gene family and the land plant body plan

The MADS-box complement of 26 genes in *P. patens* is intermediate (Rensing et al. 2008) between that found in green algae and angiosperms and similar to that found in the relatively simple vascular plant *Selaginella* (Gramzow et al. 2012) (Fig. 6). This suggests a possible relationship between expansion of the MADS-box gene family and elaboration of both the gametophytic and sporophytic plant body plans. Our analysis provides strong evidence that much of the expansion of the MADS-box gene family in *Physcomitrella* to its current size occurred within the lineage leading to *Physcomitrella* after its divergence from the tracheophyte lineage. A striking difference between the MADS-box gene family in *Physcomitrella* and that in vascular plants is the preponderance of *MIKC** genes in *Physcomitrella*. Therefore, expansion of *MIKC** genes in the moss lineage may have been related to elaboration of the gametophytic plant body plan, as has been suggested by Gramzow et al. (2012).

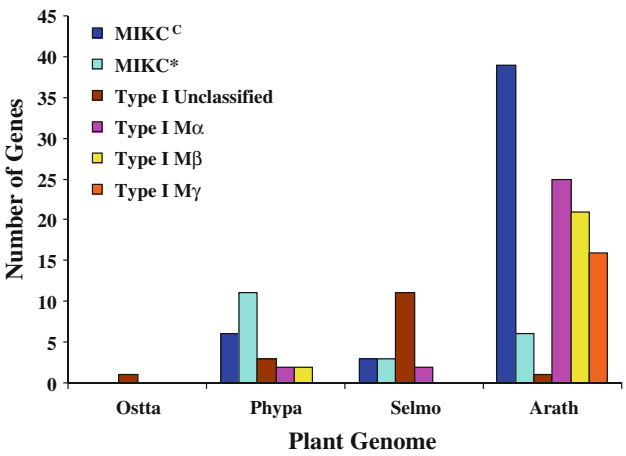

Fig. 6 Histogram of the numbers of type I and type II MADS-box genes in the genomes of *Ostreococcus tauri* (Ostta), *Physcomitrella patens* (Phypa), *Selaginella moellendorffii* (Selmo) and *Arabidopsis thaliana* (Arath)

Type I MADS-box genes, *MIKC*^{C} genes and *MIKC** genes have all been implicated in seed plant reproductive development, and some *MIKC*^{C} genes have a reproductive function in *Physcomitrella* (gametangia formation) (Quodt et al. 2007; Singer et al. 2007) and in charophycean algae (haploid reproductive cell differentiation) (Tanabe et al. 2005). It is plausible, therefore, that an ancestral regulator–target relationship between MADS-domain transcription factors and effector gene regulatory elements has been conserved during land plant evolution while expansion and divergence of the MADS-box family has paralleled elaboration of both gametophytic and sporophytic body plans.

Further progress in understanding the evolution of MADS-box genes in *Physcomitrella* will require continuing the functional characterisation of *MIKC* genes, and extending it to include type I genes. While this is necessary, it is also daunting since the high level of sequence conservation within each of the three groups of *Physcomitrella* MADS-box genes raises the prospect that single gene knockouts will be rendered useless for determining gene functions because of functional redundancy as has been shown already for the three *MIKC*^{C} genes in the *PPM2*-like clade (Singer et al. 2007; Singer and Ashton 2009). This study provides information about gene sequences, phylogenetic relationships and chromosomal linkages that can guide the choice of optimal subsets of genes for multiple gene targeting experiments and thereby maximise the likelihood of successfully determining MADS-box gene functions in this bryophyte.

Acknowledgments This study was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant awarded to N.W. Ashton and NSERC postgraduate scholarships (CGS-M and PGS-D) awarded to E.I. Barker. We wish to thank W. Chapco and D. Contreras for their help with phylogenetic analysis.

References

Adamczyk BJ, Fernandez DE (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. Plant Physiol 149:1713–1723. doi:10.1104/pp.109.135806

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martinez-Castilla L, Yanofsky MF (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. PNAS 97:5328–5333. doi:10.1073/pnas.97.10.5328

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximisation to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. AAAI Press. Menlo Park, pp 28–36

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, de Pamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, Ashton NW, Axtell MJ, Barker E, Barker MS, Bennetzen JL, Bonawitz ND, Chapple C, Cheng C, Correa LGG, Dacre M, DeBarry J, Dreyer I, Elias M, Engstrom EM, Estelle M, Feng L, Finet C, Floyd SK, Frommer WB, Fujita T, Gramzow L, Gutensohn M, Harholt J, Hattori M, Heyl A, Hirai T, Hiwatashi Y, Ishikawa M, Iwata M, Karol KG, Koehler B, Kolukisaoglu U, Kubo M, Kurata T, Lalonde S, Li K, Li Y, Litt A, Lyons E, Manning G, Maruyama T, Michael TP, Mikami K, Miyazaki S, Morinaga S-i, Murata T, Mueller-Roeber B, Nelson DR, Obara M, Oguri Y, Olmstead RG, Onodera N, Petersen BL, Pils B, Prigge M, Rensing SA, Riaño-Pachón DM, Roberts AW, Sato Y, Scheller HV, Schulz B, Schulz C, Shakirov EV, Shibagaki N, Shinohara N, Shippen DE, Sørensen I, Sotooka R, Sugimoto N, Sugita M, Sumikawa N, Tanurdzic M, Theißen G, Ulvskov P, Wakazuki S, Weng J-K, Willats WWGT, Wipf D, Wolf PG, Yang L, Zimmer AD, Zhu Q, Mitros T, Hellsten U, Loqué D, Otillar R, Salamov A, Schmutz J, Shapiro H, Lindquist E, Lucas S, Rokhsar D, Grigoriev IV (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 20:960–963. doi:10.1126/science.1203810

Becker A, Theißen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 29:464–489. doi:10.1016/S1055-7903(03)00207-0

Bemer M, Wolter-Arts M, Grossniklaus U, Angenent GC (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in *Arabidopsis* ovules. Plant Cell 20(8):2088–2101. doi:10.1105/tpc.108.058958

Bemer M, Heijmans K, Airoldi C, Davies B, Angenent GC (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. Plant Physiol 154(1):287–300. doi:10.1104/pp.110.160770

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402. doi:10.1105/tpc.106.049338

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16(7):1667–1678. doi:10.1105/tpc.021345

Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176:1035–1047. doi:10.1534/genetics.106.068874

Coen ES, Meyerowitz EM (1991) War of the whorls: genetic interactions controlling flower development. Nature 353:31–37. doi:10.1038/353031a0

Colombo L, Franken J, Koetje E, van Went J, Dons HJM, Angenent GC, van Tunen AJ (1995) The petunia MADS-box gene *FBP11* determines ovule identity. Plant Cell 7:1859–1868

Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L (2008) *AGL23*, a type I MADS-box gene that controls female gametophyte and embryo development in *Arabidopsis*. Plant J 54:1037–1048. doi:10.1111/j.1365-313X.2008.03485.x

Danielson JÅH, Johanson U (2008) Unexpected complexity of the aquaporin gene family in the moss *Physcomitrella patens*. BMC Plant Biol 8:45–59. doi:10.1186/1471-2229-8-45

De Bodt S, Raes J, Florquin K, Rombauts S, Rouzé P, Theißen G, Van de Peer Y (2003a) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol 56:573–586. doi:10.1007/s00239-002-2426-x

De Bodt S, Raes J, Van de Peer Y, Theißen G (2003b) And then there were many: MADS goes genomic. Trends Plant Sci 8(10):475–483. doi:10.1016/j.tplants.2003.09.006

de Folter S, Busscher J, Colombo L, Losa A, Angenent GC (2004) Transcript profiling of transcription factor genes during development in *Arabidopsis*. Plant Mol Biol 56:351–366. doi:10.1007/s11103-004-3473-z

Félix M-A, Wagner A (2008) Robustness and evolution: concepts, insights and challenges from a developmental model system. Heredity 100:132–140. doi:10.1038/sj.hdy.6800915

Fernandez DE, Heck GR, Perry SE, Patterson SE, Bleecker AB, Fang S-C (2000) The embryo MADS domain factor AGL15 acts postembryonically: inhibition of perianth senescence and abscission via constitutive expression. Plant Cell 12:183–197. doi:10.1105/tpc.12.2.183

Fischer A, Baum N, Saedler H, Theiβen G (1995) Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. Nucl Acids Res 23(11):1901–1911. doi:10.1093/nar/23.11/1901

Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucl Acids Res 31(13):3784–3788. doi:10.1093/nar/gkg563

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a monte carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582. doi:10.1093/bioinformatics/16.7.573

Gramzow L, Theiβen G (2010) A hitchhiker's guide to the MADS world of plants. Genome Biol 11:214. doi:10.1186/gb-2010-11-6-214

Gramzow L, Barker E, Schulz C, Ambrose B, Ashton N, Theiβen G, Litt A (2012) *Selaginella* genome analysis—entering the "homoplasy heaven" of the MADS world. Front Plant Sci 3:214. doi:10.3389/fpls.2012.00214

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421:63–66. doi:10.1038/nature01198

Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704. doi:10.1080/10635150390023552O

Hasebe M, Wen C-K, Kato M, Banks JA (1998) Characterisation of MADS homeotic genes in the fern *Ceratopteris richardii*. PNAS 95:6222–6227

Heard J, Dunn K (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. PNAS 92: 5273–5277

Heard J, Caspi M, Dunn K (1997) Evolutionary diversity of symbiotically induced nodule MADS box genes: characterisation of *nmhC5*, a member of a novel subfamily. MPMI 10:665–676. doi:10.1094/MPMI.1997.10.5.665

Heath L, van der Walt E, Varsani A, Martin DP (2006) Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J Virol 80:11827–11832. doi:10.1128/JV.01100-06

Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theiβen G (2002) Two ancient classes of *MIKC*-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol Biol Evol 19(6):801–814

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. doi:10.1055/s-2002-35440

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Kang I-H, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The AGL62 MADS domain protein regulates cellularisation during endosperm development in *Arabidopsis*. Plant Cell 20:635–647. doi:10.1105/tpc.107.055137

Kaufmann K, Melzer R, Theiβen G (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347:183–198. doi:10.1016/j.gene.2004.12.014

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(12):1963–1977. doi:10.1093/molbev/msg216

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes Dev 17(12):1540–1553. doi:10.1101/gad.257403

Kondrashov FA, Koonin EV (2004) A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. Trends Genet 20:287–291. doi:10.1016/j.tig.2004.05.001

Krogan NT, Ashton NW (2000) Ancestry of plant MADS-box genes revealed by bryophyte (*Physcomitrella patens*) homologues. New Phytol 147(3):505–517. doi:10.1046/j.1469-8137.2000.00728.x

Kwantes M, Liebsch D, Verelst W (2011) How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol Biol Evol. doi: 10.1093/molbev/msr200

Lee H-L, Irish VF (2011) Gene duplication and loss in a MADS box gene transcription factor circuit. Mol Biol Evol 28(12):3367–3380. doi:10.1093/molbev/msr169

Lehti-Shiu MD, Adamczyk BJ, Fernandez DE (2005) Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. Plant Mol Biol 58:89–107. doi:10.1007/s11103-005-4546-3

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155. doi:10.1126/science.290.5494.1151

Ma H, Yanofsky MF, Meyerowitz EM (1991) *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev 5:484–495. doi:10.1101/gad.5.3.484

Maddison JR, Maddison WP (2001) MacClade 4: analysis of phylogeny and character evolution. Version 4.05. Sinauer Associates, Sunderland, Massachusetts

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16(6):562–563. doi:10.1093/bioinformatics/16.6.562

Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified BOOTSCAN algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21(1):98–102. doi:10.1089/aid.2005.21.98

Montag K, Salamini F, Thompson RD (1996) The ZEM2 family of maize MADS-box genes possess features of transposable elements. Maydica 41:241–254

Moore RC, Grant SR, Purugganan MD (2005) Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Mol Biol Evol 22(1):91–103. doi:10.1093/molbev/msh261

Münster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theiβen G (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. PNAS 94:2415–2420. doi:10.1073/pnas.94.6.2415

Münster T, Faigl W, Saedler H, Theiβen G (2002) Evolutionary aspects of MADS-box genes in the eusporangiate fern *Ophioglossum*. Plant Biol 4(4):474–483. doi:10.1055/s-2002-34130

Newton AE, Wikström N, Bell N, Forrest LL, Ignatov MS (2007) Dating the diversification of the pleurocarpous mosses. In: Newton AE, Tangney RS (eds) Pleurocarpous mosses: systematics and evolution. CRC Press, Boca Raton, pp 337–366

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Heidelberg

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225. doi:10.1006/viro.1999.0056

Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otillar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H, Grigoriev IV (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. PNAS 104(18):7705–7710. doi:10.1073/pnas.0611046104

Pařenicová 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(7):1538–1551. doi:10.1105/tpc.011544

Pina C, Pinto F, Feijó JA, Becker JD (2005) Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744–756. doi:10.1104/pp.104.057935

Portereiko MF, Lloyd A, Steffen JG, Punwani JA, Otsuga D, Drews GN (2006) *AGL80* is required for central cell and endosperm development in *Arabidopsis*. Plant Cell 18:1862–1872. doi:10.1105/tpc.106.040824

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256. doi:10.1093/molbev/msn083

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14(9):817–818. doi:10.1093/bioinformatics/14.9.817

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. PNAS 98(24):13757–13762. doi:10.1073/pnas.241370698

Quodt V, Faigl W, Saedler H, Münster T (2007) The MADS-domain protein PPM2 preferentially occurs in gametangia and sporophytes of the moss *Physcomitrella patens*. Gene 400:25–34. doi:10.1016/j.gene.2007.05.016

Rensing SA, Fritzowsky D, Lang D, Reski R (2005) Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, Physcomitrella patens. BMC Genomics 6:43. doi:10.1186/1471-2164-6-43

Rensing SA, Ick J, Vawcett JA, Lang D, Zimmer A, van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol Biol 7:130–139. doi:10.1186/1471-2148-7-130

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujits T, Oishi K, Shin-I T, Kruoki Y, Toyoda A, Suzuki Y, Hashimoto S-I, Yamaguchi K, Sugano S, Kohara A, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S-H, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319(5859):64–69. doi:10.1126/science.1150646

Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Münster T (2005) Isolation and characterisation of new MIKC*-type MADS-box genes from the moss Physcomitrella patens. Plant Biol 7:307–314. doi:10.1055/s-2005-865640

Rijpkema AS, Gerats T, Vandenbussche M (2007) Evolutionary complexity of MADS complexes. Curr Opin Plant Biol 10:32–38. doi:10.1016/j.pbi.2006.11.010

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. doi:10/1093/bioinformatics/btg180

Sanderson MJ, Thorne JL, Wikström N, Bremer K (2004) Molecular evidence on plant divergence times. Am J Bot 91(10):1656–1665. doi:10.3732/ajb.91.10.1656

Sawyer SA (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Singer SD, Ashton NW (2009) MADS about MOSS. Plant Signal Behav 4(2):111–112. doi:10.4161/psb.4.2.7479

Singer SD, Krogan NT, Ashton NW (2007) Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. Plant Cell Rep 26(8):1155–1169. doi:10.1007/s00299-007-0312-0

Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129. doi:10.1007/BF00182389

Steffen JG, Kang I-H, Portereiko MF, Lloyd A, Drews GN (2008) AGL61 interacts with AGL80 and is required for central cell development in *Arabidopsis*. Plant Physiol 148(1):259–268. doi:10.1104/pp.108.119404

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. doi:10.1046/j.1469-8137.2002.00392.x

Svensson ME, Johannesson H, Engström P (2000) The *LAMB1* gene from the clubmoss, *Lycopodium annotinum*, is a divergent MADS-box gene, expressed specifically in sporogenic structures. Gene 253:31–43. doi:10.1016/S0378-1119(00)00243-2

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland Massachusetts

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. doi:10.1073/pnas.0409860102

Tapia-López R, García-Ponce B, Dubrovsky JG, Garay-Arroyo A, Pérez-Ruíz RV, Kim S-H, Acevedo F, Pelaz S, Alvarez-Buylla ER (2008) An *AGAMOUS*-related MADS-box gene, *XAL1* (*AGL12*), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. Plant Physiol 146:1182–1192. doi:10.1104/pp.107.108647

Theißen G, Saedler H (2001) Plant biology: floral quartets. Nature 409:469–471. doi:10.1038/35054172

Theiβen G (2001) Development of floral organ identity; stories from the MADS house. Curr Opin Plant Biol 4(1):75–85. doi:10.1016/S1369-5266(00)00139-4

Theiβen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter K-U, Saedler H (2000) A short history of MADS-box genes in plants. Plant Mol Biol 42(1):115–149. doi:10.1023/A:1006332105728

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22(22):4673–4680. doi:10.1093/nar/22.22.4673

Verelst W, Saedler H, Münster T (2007a) *MIKC** MADS-protein complexes bind motifs enriched in the proximal region of

late-pollen-specific *Arabidopsis* promoters. Plant Physiol 143: 447–460. doi:10.1104/pp.106.089805

Verelst W, Twell D, de Folter S, Immink R, Saedler H, Munster T (2007b) MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol 8(11):R249. doi: 10.1186/qb-2007-8-11-r249

Veron AS, Kaufmann K, Bornberg-Bauer E (2007) Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. Mol Biol Evol 24(3): 670–678. doi:10.1093/molbev/msl197

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18(6):292–298. doi:10.1016/S0169-5437(03)00033-8

Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol Genet Genomics 278(4):393–402. doi:10.1007/s00438-007-0257-6

Zobell O, Faigl W, Saedler H, Münster T (2010) MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol Biol Evol 27(5):1201–1211. doi:10.1093/molbev/msq005

Zucchero JC, Caspi M, Dunn K (2001) ngl9: a third MADS box gene expressed in alfalfa root nodules. Mol Plant-Microbe Interact 14(12):1463–1467. doi:10.1094/MPMI.2001.14.12.1463