Conservation and diversification of MIKC* MADS-domain transcription factors during the evolution of vascular land plants

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Michiel Kwantes

aus Tilburg, Niederlande

Köln, 2010

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

Berichterstatter: Prof. Dr. Heinz Saedler

                             Prof. Dr. Jonathan C. Howard

Tag der mündlichen Prüfung: 25.06.2009

"So I have just one wish for you - the good luck to be somewhere where you are free to maintain the kind of integrity I have described […that is not lying, but bending over backwards to show how you're maybe wrong…] and where you do not feel forced by a need to maintain your position in the organization, or financial support, or so on, to lose your integrity. May you have that freedom."

Richard P. Feynman

Adapted from the Caltech commencement address given in 1974.

Abstract

The morphological diversity of land plants is astounding. However, what we see mostly is the sporophytic phase that is dominant in the majority of land plants. In contrast, the diversity of the secret gametophytic phase is unseen and rather uninvestigated. Recently, evidence has accumulated that the so-called MIKC* group of MADS-domain transcription factors is important for the proper functioning of the Arabidopsis male gametophyte (pollen). Already earlier, MIKC* genes were identified in the moss *Physcomitrella patens*, which has a dominant gametophytic phase. MADS-domain proteins are well known for the roles they have in flower development and thus for the establishment of the sporophytic body plan. That MIKC* genes have a similar role in the gametophytic phase is not granted, but a tempting hypothesis. To study the function of MIKC* genes and their possible role in land plant gametophyte development and its evolution, they were isolated from a broad variety of vascular land plants, namely, the lycophyte *Selaginella moellendorffii*, the fern *Ceratopteris richardii*, the basal eudicot *Eschscholzia californica,* the monocot *Oryza sativa* and the basal angiosperm *Aristolochia fimbriata*. Sequence comparison showed that MIKC* MADS-box genes probably evolved from classical MIKC^c genes by a duplication event in the Keratin-like domain. Further phylogenetic analysis revealed that 2 phylogenetic subclades emerged early in the evolution of vascular plants and indications were found for a recent subfunctionalization of one of the subclades in angiosperms. MIKC* genes from different, remote, plant lineages were heterologously expressed in an Arabidopsis MIKC* mutant and it could be shown that they were able to perform the same function as Arabidopsis MIKC* genes. This information plus the results that were gathered by performing expression and yeast-2-hybrid interaction studies, were unified in a hypothesis concerning the function of MIKC* genes during land plant evolution.

Zusammenfassung

Die morphologische Vielfalt der Landpflanzen ist staunenerregend. Dabei ist was wir sehen meist die sporophytische Phase, die bei der Mehrheit der Landpflanzen dominant ist. Die Vielfalt der geheimnisvollen gametophytischen Phase dagegen, ist verborgen und eher unerforscht. In jüngster Zeit haben sich die Hinweise verdichtet, dass die sogenannte MIKC*-Gruppe von MADS-Domänen-Transkriptionsfaktoren in Arabidopsis wichtig für das normale Funktionieren des männlichen Gametophyten (des Pollens) ist. Bereits zuvor wurden MIKC*-Gene in dem Moos *Physcomitrella patens*, das eine dominante haploide Phase besitzt, gefunden. MADS-Domänen-Proteine sind sehr bekannt für ihre Rolle in der Blütenentwicklung und damit in der Realisierung des sporophytischen Bauplans. Dass MIKC*-Gene eine ähnliche Rolle in der gametophytischen Generation spielen, ist nicht gewiss, jedoch eine verlockende Hypothese. Um die Funktion von MIKC*-Genen und ihre mögliche Rolle in der Evolution von Landpflanzen-Gametophyten zu untersuchen, wurden sie aus den verschiedensten Gruppen der Gefäßpflanzen isoliert, nämlich aus dem Bärlappgewächs *Selaginella moellendorffii*, dem Farn *Ceratopteris richardii*, der basalen Eudicotylen *Eschscholzia californica*, der einkeimblättrigen Art *Oryza sativa* und aus der basalen Angiosperme *Aristolochia fimbriata*. Sequenzvergleiche zeigten, dass MIKC*-MADS-Box-Gene sich vermutlich aus klassischen MIKC^c-Genen durch ein Duplikationsereignis in der Keratin-ähnlichen Domäne entwickelt haben. Weitere phylogenetische Untersuchungen zeigten, dass früh in der Evolution der Gefäßpflanzen zwei phylogenetische Untergruppen entstanden und es wurden Hinweise auf eine rezente Subfunktionalisierung einer der beiden Sub-Kladen in Angiospermen gefunden. MIKC*-Gene aus verschiedenen, entfernt verwandten Pflanzen-Linien wurden heterolog in einer Arabidopsis MIKC*-Mutante exprimiert, und es konnte gezeigt werden, dass sie in der Lage waren, dieselbe Funktion wie Arabidopsis MIKC*-Gene auszuführen. Diese Erkenntnisse, sowie die Ergebnisse von Expressions- und Yeast-2-Hybrid-Interaktions-Studien wurden in einer Hypothese bezüglich der Funktion von MIKC*-Genen in der Evolution der Landpflanzen vereint.

Table of Contents

1   Introduction

1.1   On the alternation of generations during land plant evolution

All land plants have in common that their life cycle consists of two alternating generations: the diploid sporophyte that produces the spores and the haploid gametophyte that produces the sperm and egg cells. In charophycean algae, which share the most recent common ancestor with land plants, the sporophytic phase consists of a single cell, the zygote, which is the direct product of fertilization. Also in land plants a zygote is formed but in contrast to the charophytes, these zygotes divide mitotically and form an embryo (reviewed in Graham 1996). This trait has provided the land plants with the name embryophytes.

All land plants have both multicellular gametophytes and sporophytes but in different plant lineages they differ in size and complexity (figure 1). The most basal land plants, the bryophytes, have a dominant gametophyte. The gametes of bryophytes are
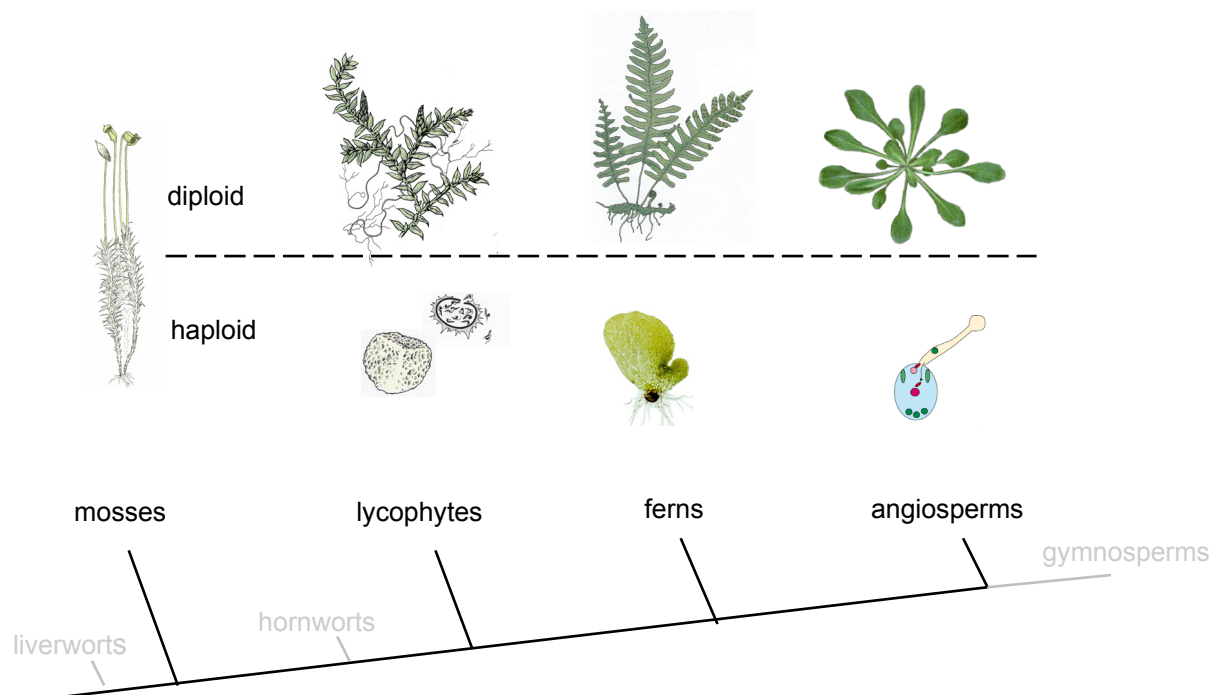
Figure 1. Different sporophytes and gametophytes from a selection of land plants. Bryophytes (liverworts, mosses and hornworts) have a dominant haploid phase. In vascular land plants the gametopyhte is not dominant and the diploid sporophyte is more complex than the haploid generation. Sizes are not proportional (Adapted from Raven (1992) and www.C-fern.org).

produced in specialized organs of the haploid phase. Archegonia are the sites where egg cells are produced and fertilization occurs. Motile sperm cells are produced by the antheridia and need water as a medium to reach the eggs. Apart from this task, the haploid phase also provides nutrition for the embryo and the mature sporophyte into which it develops.

Approximately 350-400 million years ago plants evolved that did not have a dominant haploid but a dominant diploid phase. The new body plan of the sporophyte was characterized, among many other innovations, by the presence of vascular tissue and showed a high complexity (Kenrick and Crane, 1997). Although the sporophyte of vascular plants is referred to as dominant, it must be remarked that it is still dependent upon the gametophyte in the first stages of embryo development.

The first vascular plant group to appear of which the descendants still exist today, are the lycophytes. They use spores as a means of dispersal, a trait that is also found in ferns and their allies. The production of drought tolerant spores is a basal feature that was inherited from their bryophyte-like ancestor. Despite the common ancestry, the gametophytes that develop out of the spores of these early vascular land plants are, compared to those of bryophytes, very reduced and have as primary task "only" the production of gametes.

The gametophytes of angiosperms, flowering plants, can be seen as even more reduced. However, in the lineage leading to the seed plants (angiosperms and gymnosperms) also some novelties evolved. Perhaps the most important is that the male gametophyte (pollen) produces sperm cells that have no flagella, in contrast to the sperm cells of all non-seed plants. Sperm cells are delivered to the female gametophyte, in angiosperms called the embryo sac, by the pollen tube. As a consequence, seed plants are not dependent upon water for the fertilization process, which probably added to their success during evolution.

1.2   MADS-box genes are important for the angiosperm body plan

Evolution of the body plan is tightly linked to the evolution of transcription factors controlling the developmental programs that guide the proper ontogeny (reviewed in Theissen et al. 2000). For the evolution of land plants, it has been noted that a positive correlation can be seen between the complexity of the sporophytic generation and the number of MADS-box transcription factors in the genome (Nam

et al. 2004). For example, Arabidopsis has 107 MADS-box transcription factors (Parenicova et al. 2003) while there are only 20 known from the moss *Physcomitrella patens* (Rensing et al. 2008). However, the relation between the increased number of MADS-domain transcription factors and the complexity of angiosperms is not clear. Did the rise in the number of MADS-domain transcription factors precede the increase in complexity or was it an effect? As in many organisms, the architecture of flowering plants is characterized by its modularity. For example, and most strikingly, the organs of the flower, from sepal to carpel are all thought to be modified leaves (reviewed in Theissen et al. 2000). The study of MADS-box genes that control organ identity has provided a lot of insight into how the flower body evolved and is being built (Sommer et al. 1990, Coen and Meyerowitz 1991, Theissen and Saedler 2001). However, the flower is a relatively recent innovation, and studying MADS-domain transcription factors in this derived organ might not be sufficient to gain understanding into how MADS-box genes and plant developmental programs co-evolved. Furthermore, knowledge about ancestral functions and molecular features that made MADS-box genes important determinants of development is possibly easier to obtain within another context and might prove to be fundamentally different from what is currently known.

1.3   MIKC* transcription factors are important for the development of the Arabidopsis male gametophyte

MIKC* proteins are very similar to the so-called classical MIKC (MIKC^c) MADS-domain transcription factors that are well known for their roles in floral development. "MIKC" indicates the modular structure of both groups of proteins, which consists of 4 domains, namely, the MADS-domain (M) that functions in DNA binding and dimerization; the intervening domain (I) separates the MADS-domain from the K-domain and specifies dimerization (Riechmann and Meyerowitz 1997); the Keratin-like domain (K) has homology to the coiled-coil domain of Keratin (Ma et al. 1991) and functions in protein-protein interaction (Davies et al. 1996); the C-terminal (C) domain is involved in transcription activation and in higher order complex formation (Honma and Goto 2001, Theissen and Saedler 2001, Tonaco et al. 2006, Melzer and Theissen 2009). It has been reported that the most prominent differences between MIKC^c and MIKC* proteins are that the latter have an elongated I-domain, less

regularly spaced hydrophobic residues in the K-domain and a more sophisticated exon-intron structure than the (classical) MIKC^c MADS domain proteins. (Henschel et al. 2002; Parenicova et al. 2003; Riese et al. 2005 and Tanabe et al. 2005).

MIKC* transcription factors were first discovered in the moss *Physcomitrella patens* (Physcomitrella), which has a dominant gametophytic phase (Henschel et al. 2002). Not long afterwards, they were also recognized in *Arabidopsis thaliana* (Arabidopsis) (Kofuji et al. 2003). In Arabidopsis the MIKC* MADS-domain transcription factor family consists of 6 members, which can be subdivided in two monophyletic subgroups designated as "S" and "P" (Nam et al. 2004). MADS-domain transcription factors bind to DNA as dimers and for Arabidopsis MIKC* proteins it was demonstrated that only heterodimers consisting of a member of each phylogenetic subclade exist (Verelst et al. 2007a). Five of the 6 Arabidopsis MIKC* genes are expressed in pollen. Mutant analysis has shown that the members of each clade are highly redundant and a strong phenotypic aberration only becomes apparent when loss-of function alleles are combined. Double and triple mutants produced by W. Verelst (Verelst et al. 2007b) showed an *in vitro* pollen germination defect (exemplified in figure 2). Subsequent microarray analysis revealed that the
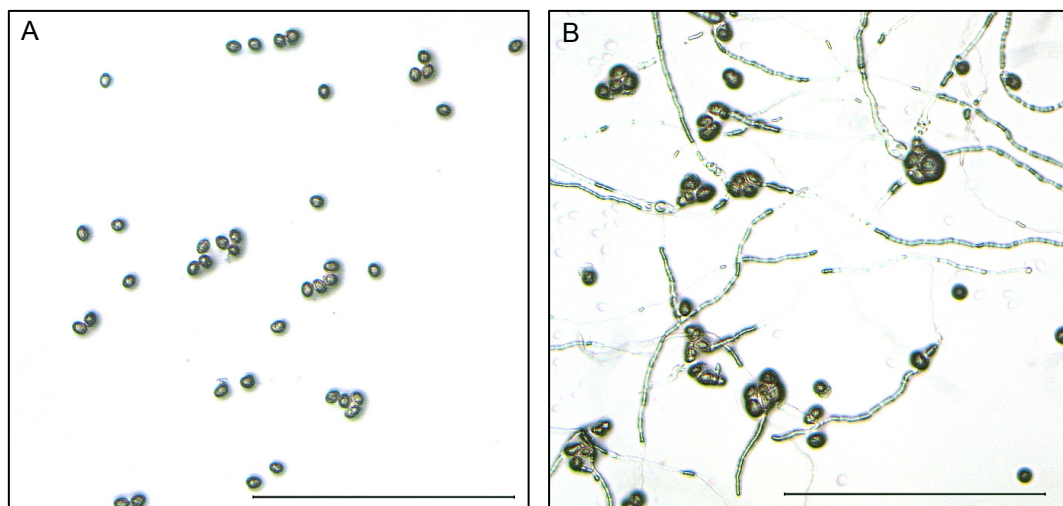
Figure 2. Comparison between the *in vitro* pollen germination of wild type pollen and MIKC* double mutant pollen. **(A)** Pollen from the Arabidopsis double mutant *agl66/104-1* shows no germination after incubation on pollen germination medium. **(B)** Wild type pollen is able to germinate and pollen tubes can be seen that extend from the pollen grains. Scale bar indicates 500µm.

transcription of more than 1300 genes, of roughly 7000 expressed genes, is affected in triple mutant pollen. Furthermore, the analysis showed that MIKC* complexes repress immature pollen genes and activate mature pollen genes during development. Adamczyk and Fernandez (2009) confirmed these results by using mutants with stronger loss-of function alleles. Additionally, they convincingly showed that the stronger mutant has defects in pollen tube formation *in vivo*, which ultimately led to a decrease in seed set. Pollen grains consist of 3 cells, two sperm cells and a larger vegetative cell that contains the sperm cells and will form the pollen tube. No evidence has been found that sperm cells are affected in MIKC* mutants nor that MIKC* proteins are expressed in sperm cells (Borges et al. 2009). Taken together, it can be concluded that recently a lot of evidence has accumulated that underlines the importance of MIKC* transcription factors for the development of the male gametophyte (pollen) of Arabidopsis. Therefore, it was decided to investigate the putatively gametophyte specific MIKC* MADS-domain transcription factor family in different vascular land plant lineages. Moreover, clues about the evolution of the gametophyte body plan and the function(s) that MIKC* proteins fulfilled during evolution, which are currently hardly understood, may become apparent.

Here will be described how MIKC* genes were isolated out of a broad selection of vascular land plants (see figure 3). *Selaginella moellendorffii* (Selaginella) was chosen as a representative of the lycophytes, a plant clade that diverged from other vascular plants around 400 million years ago and is the most basal vascular plant lineage (Weng et al. 2005). The fern *Ceratopteris richardii* (Ceratopteris) was chosen in the monilophyte clade, which is the group of plants considered to be most closely related to seed plants (Pryer et al. 2001). For the seed plant clade only representatives of angiosperms were selected, namely the monocot *Oryza sativa* (Oryza), the basal eudicot *Eschscholzia californica* (Eschscholzia) and the basal angiosperm *Aristolochia fimbriata* (Aristolochia). Sequence comparison combined with a phylogenetic approach and protein-protein interaction analysis is used to describe how the proteins have diversified during evolution. Furthermore, a subset of MIKC* proteins from the different plant lineages mentioned above were tested for their ability to complement an Arabidopsis MIKC* mutant. Together with expression data the point will be made that all MIKC* proteins share the same, conserved, function.

2   Material and Methods

All procedures described below were performed by the author. Transgenic Arabidopsis plants expressing PPM3, PPM4, AGL15, AGAMOUS and MEF2A were produced in cooperation with W. Faigl (MPIZ Cologne).

2.1   Plant materials and cultivation

Arabidopsis thaliana agl66/104-1 double mutant plants were kindly provided by W. Verelst, who described them in Verelst et al. 2007b. The *agl66/104-1* double mutant plants are homozygous for the T-DNA insertions from SALK line 098698 (*agl104-1*) and SALK line 072108 (*agl66*). Arabidopsis plants were grown in a greenhouse at 22ºC in long day conditions (16 hours light, 8 hours dark).

For the study of MIKC* transcription factors, model species were chosen that represent different vascular plant lineages and that are also easy to culture. Furthermore, for Oryza and Selaginella the genome sequence was available.

Eschscholzia californica seeds were donated by Andrea Scholz and Stefan Gleissberg (University Mainz) and grown in similar conditions as Arabidopsis.

Aristolochia fimbriata seeds were obtained from Christoph Neinhuis (TU Dresden) and grown in long day conditions in a standard greenhouse at 22-26ºC.

Selaginella moellendorffii plants were acquired from Plant Delights Nursery, Inc., Raleigh, NC and grown in long day conditions under a cover to create a moist environment and under the canopy of larger plants to reduce light intensity.

Ceratopteris richardii strain Hn spores were obtained from William Martin and are described in Scott and Hickock (1987). Spores were sterilized for 20' in a 0.875 % (w/v) NaOCl solution, washed 3 times with sterile water and imbibed for one day. Then the sterilization procedure was repeated and spores were dispersed on solid

medium consisting of 2.2 g l^{-1} MS salt with vitamins (Duchefa, Haarlem, The Netherlands) supplemented with 30 g l^{-1} sucrose, 2.7 ml of a chelated iron solution (Hickock et al. 1997) and 9 g l^{-1} agar. The pH was adjusted to 5.9 using KOH. Gametophytes were cultured at 29ºC under a light regime of 18 hours of light and 6 hours darkness and a humidity of 95%. After 2 weeks gametophytes were collected for RNA preparation.
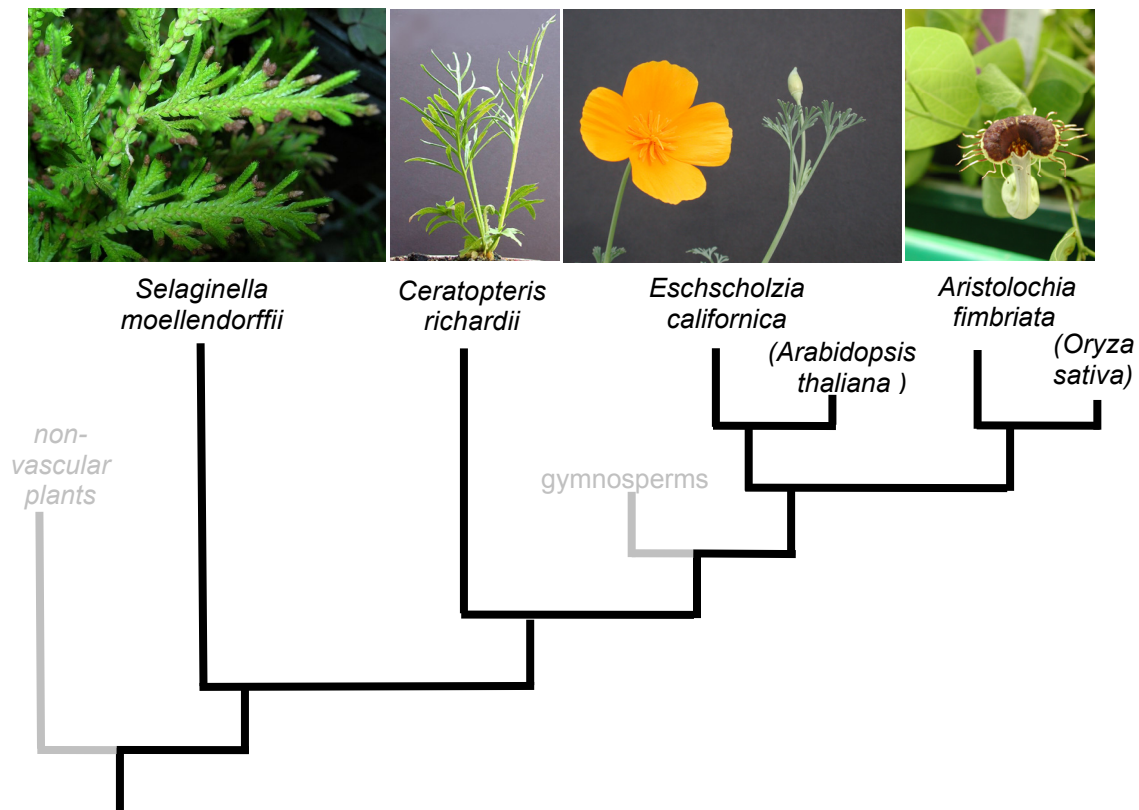
Figure 3. A simplified phylogenetic tree showing the evolutionary relationship between the model species used in this study.

2.2   Identification of MIKC* sequences in model species

The TIGR rice genome annotation sequence database release 5 (available at www.gramene.org) and the assembled trace files of *Selaginella moellendorffii* (http://moss.nibb.ac.jp, sequencing performed by the DOE Joint Genomic Institute) were screened for putative MIKC* genes using a tblastn search (Altschul et al. 1997)

with AGL30 and AGL66 as a query. To look for already available MIKC* genes in the other model species used here, a similar, species-specific search was performed against the plantGDB DNA and RNA collection (www.plantGDB.org).

2.3   Cloning of MIKC* sequences

MIKC* genes were expected to be expressed in gametophytic tissues. Therefore, RNA was extracted from angiosperm male gametophytes (pollen), a mixture of male and hermaphroditic *Ceratopteris richardii* gametophytes and the strobili of *Selaginella moellendorffii*, which are cone-shaped organs that carry the sporangia and harbour the developing gametophytes. Plant material for RNA extraction was harvested at identical time points and grinded using liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (www.qiagen.com). Subsequently, cDNA pools were created using the Superscript II Reverse Transcriptase kit (www.invitrogen.com) with an oligo d(T)-anchor primer from the 2nd generation 5'/3' RACE kit (www.roche.com). DNA was extracted using the DNeasy Plant Mini kit (www.qiagen.com).

MIKC* cDNAs that were already identified in the database were amplified using (nested) PCR and primers based on the putative 5' and 3' UTRs. A list of primers can be found in supplemental table S1.

Nested PCR with degenerate primers and Apliqon taq polymerase (www.ampliqon.com) was used to search for new MIKC* cDNAs from Ceratopteris, Eschscholzia and Aristolochia. Primers were designed in conserved parts of MIKC* proteins at positions that differed from classical MIKC^c MADS domain proteins. The partial cDNAs obtained from degenerate PCRs were completed using the 5'/3' RACE kit, 2nd generation (www.roche.com).

All full-length cDNAs and genomic loci were amplified with the proofreading enzyme Primestar HS DNA Polymerase (www.takara.com) and cloned in the pGEM-

T-Easy Vector System (www.promega.com). All MIKC* cDNA sequences were verified against an independent second clone or the genomic locus.

2.4   Phylogenetic analysis of MIKC* sequences

A tblastn search (Altschul et al. 1997) was performed against the NCBI nr/nt database and the expressed sequence tags database (http://www.ncbi.nlm.nih.gov) using the MIKC* proteins AGL30 and AGL66 of Arabidopsis as a query. To obtain a dense taxon sampling searches were performed separately against all the orders of angiosperms plus all gymnosperm, fern and allies, hornwort, liverwort and streptophyte algae sequences. For the orders Brassicales and Poales, only the sequences of *Arabidopsis thaliana* and *Oryza sativa* were used, respectively. Putative positive hits were identified based on the presence of conserved peptide motifs in the MADS-domain. The total collection of sequences that were used for making an alignment consisted of the derived peptide sequences of all plantGDB-assembled unique transcripts (www.plantGDB.org) that corresponded to the cDNAs identified by the BLAST search and that coded for at least the M-, I- and K-domain. That dataset was supplemented with the proteins sequences derived from all the MIKC* genes that were isolated in this study, 5 of the 11 MIKC* genes from *Physcomitrella patens*, 11 classical MIKC^c genes from a broad range of land plants and 2 MIKC^c genes from streptophyte algae. Preliminary alignments were made by Clustal W using the BLOSUM30, PAM350 and MD350 matrices with an open gap penalty of 10 and an extend gap penalty of 0.01 for the pairwise alignment and an 0.05 extend gap penalty and 40% divergence delay for the multiple alignment. Alignments using the BLOSUM30 matrix were further manually optimized, gaps were removed and the MADS-domain and parts of the I- and K-domain were used for phylogenetic analysis using the neighbour joining method. Bootstrap values were calculated from 2000 repeats. All phylogenetic analyses were performed using the MacVector 7.2.3 software package (http://accelrys.com).

2.5 Expression study of *Selaginella moellendorffii* MIKC* genes using quantitative real-time PCR

Selaginella moellendorffii total RNA was extracted as described for cloning of MIKC* sequences. Real-time quantitative PCR (qPCR) primer pairs were designed such that at least one of each pair crossed an exon junction, except for *6-phosphogluconate dehydrogenase* (*6PGD)*, which is encoded by a single exon. Expression data was normalized using *6PGD*, which has been used previously for semiquantative analysis in *Selaginella remotifolia* (Tanabe et al. 2003), and the homologs of the Arabidopsis genes *PDF1* and *At5g46630*, which code for the 65 kDa regulatory subunit of protein phosphatase 2A and a subunit of the clathrin adapter complex, respectively. The latter two genes have been reported to be superior reference genes in Arabidopsis (Czechowski et al. 2002). Quantitative real time PCR was performed on an iQ5 Real Time PCR Detection System (www.bio-rad.com) with a cycling protocol of: 2' 50°C, 10' 95°C and 40-50 cylces of 15'' 95 and 1' 60°C. Each reaction contained 12.5 µl 2x IQ SYBR premix (www.bio-rad.com), 10 µl template and 200nM of primers in a final volume of 25µl. Melting curve and data analyses were performed using the Bio-Rad iQ5 Optical System Software. Primer efficiencies were determined from dilution curves on cDNA mixtures from all tissues using the Pfaffl method (Tichopad et al. 2003). Following analyses, PCR products were checked on 2.5% agarose gels.

2.6 Yeast-2-hybrid MIKC* interaction study

The full length coding sequences of all isolated MIKC* genes were cloned in the pGADT7 prey and pGBKT7 bait vectors (www.clontech.com), which carry the *LEU2* and *TRP1* nutritional markers, respectively, using *Nco*I and *BamH*I restriction sites. For all Oryza and Eschscholzia MIKC* bait constructs and the *SmMADS3* bait construct strong autoactivation was observed. For each bait protein fusion, one or 2 different C-terminal deletion constructs were produced. Primers used for cloning can be found in supplemental table S1.

Per species all combinations of bait and prey vector were cotransformed in yeast strain *Saccharomyces cerevisiae* AH109 according to the Matchmaker *GAL4* Two-Hybrid System 3 protocol (www.clontech.com). Cotransformants were selected on synthetic dropout medium lacking Leu and Trp (SD/-L/-W). To test for interactions, single yeast colonies from the SD/-L/-W selection were grown in liquid SD/-L/-W medium and a spot dilution assay was performed on solid synthetic dropout medium lacking the amino acids Leu, Trp and Ade (SD/-L/-W/-A). To validate interactions and test relative interaction strength, a non-lethal ß-galactosidase assay was performed as described by Duttweiler (1996).

2.7   Transformation of Arabidopsis

All MIKC* protein coding sequences isolated here were cloned in binary vector pGJ2364 that contains the first upstream 1555 bp of the *AGL65* promoter fused to GFP (kindly provided by W. Verelst, MPIZ, Cologne). For cloning purposes, position −1 and −2 of the *AGL65* promoter had been substituted in cytosine residues. In addition, also constructs were prepared that contained the coding sequences of the MIKC* genes *PPM3* and *PPM4* of *Physcomitrella patens* (kindly provided by Daniela Liebsch, MPIZ, Cologne), the classical MIKC^c genes *AGAMOUS* and *AGL15* from *Arabidopsis thaliana* (kindly provided by Richard Immink, Plant Research International, Wageningen) and the human MADS-domain protein MEF2A (kindly provided Andrew D. Sharrocks, University of Manchester, Manchester). Coding sequences without the stop codon were cloned into pGJ2364 digested with *Nco*I to obtain a C-terminal GFP fusion. Inserting the complete CDS in pGJ2364 digested with *Nco*I and *BamH*I, which were used to excise the GFP tag, produced non-GFP constructs. Constructs were introduced in *Agrobacterium* strain GV3101 or ABI using electroporation (Koncz and Schell, 1986, Koncz et al. 1984) and then in Arabidopsis *agl66/104-1* double mutants using floral dipping (Clough and Bent, 1998). A nos-bar cassette in pGJ2364 allowed for selection by spraying with BASTA.

2.8 Selection of transgenic plants

After screening with BASTA, putative transgenic plants (generation T1) were genotyped. DNA was extracted using the Plant DNeasy Mini kit (www.qiagen.com) and PCRs were performed to confirm the presence of the transgene. Each line was also tested for the presence of the SALK T-DNA integrations in the *AGL66* and *AGL104* loci using a primer in the left border of the T-DNA and a gene specific primer. Additionally, the absence of the wild type loci was confirmed by PCR with primers flanking the SALK T-DNA insertions and an elongation time such that wild type loci could be amplified but not the longer T-DNA inserted loci.

To obtain putative single copy integration lines for the heterologous MIKC* genes, a qPCR approach was designed. One primer pair was targeted against a part of the promoter of *AGL65* that is also used in the binary constructs and another primer pair against a single copy gene to normalize for the DNA input in the qPCR reaction. In this way the number of inserted transgenes could be derived from subtracting the number of native *AGL65* promoters in the Arabidopsis genome, which is 2, from the number derived from the qPCR data. Further, the same procedures as for expression analysis were used with the exception that primer efficiencies were based on qPCR from dilution series of genomic DNA. Unfortunately, the qPCR method proved to be practically unsuited for isolating *single copy* lines, as described in the results, hence, *single integration* lines were chosen based upon the pollen fluorescence patterns of segregating T2 lines.

2.9 *In vitro* pollen germination assays

At first, transgenic lines were screened qualitatively in the hemizyous T1 generation for enhanced pollen germination. Lines that showed complementation were selected for a quantitative *in vitro* pollen germination assay in homozygous T3 plants. For the germination test, pollen was incubated on germination medium for one day in a moist germination chamber at 22°C as described by Boavida and McCormick (2007). Pollen from a single flower was transferred to an agarose pad on a microscope slide.

Pollen was spread evenly across the surface not using the flower itself but by using a human hair bend in a hairpin shape attached to a toothpick, preventing sporophytic material, which has a positive effect on pollen germination, to hamper the experiment. For each agarose pad at least 2 different areas were examined. All experiments were performed a minimum of 2 times and a total of at least 300 pollen grains were scored for germination using differential interference contrast microscopy. A positive germination call was awarded if a pollen tube was at least as long as the diameter of the pollen grain (Boavida and McCormick 2007). Germination rates of wild type and *agl66/104-1* mutant pollen were compared to the pollen from the transgenic lines.

3 Results

3.1 MIKC* genes were identified in all model species

Searching the Oryza database delivered 3 candidate MIKC* loci: *Os11g43740*, *Os08g38590* and *Os06g11970*, also known as *OsMADS68*, *OsMADS62* and *OsMADS63*, respectively. In addition, a cDNA corresponding to *OsMADS63* was identified (accession number AK111776). The 3 *Oryza sativa* MIKC* loci identified here are the same as found in a phylogenetic analysis of Oryza and Arabidopsis MADS-box genes performed by Nam et al. (2004). In contrast to Nam et al. (2004) and the findings reported here, Arora and coworkers (2007) communicated that the Oryza genome holds 6 MIKC* MADS-box genes. That conclusion was based upon a study of 3 phylogenetic trees created from alignments of the MADS-domain, the complete protein or the full-length coding sequence of Arabidopsis and Oryza MADS-box genes. A re-examination of their phylogenetic tree derived from an alignment of exclusively MADS-domains, however, clearly showed that the MIKC* clade included only 3 Oryza MIKC* MADS proteins, the same ones as recognized here. In the other 2 phylogenetic trees by Arora, surprisingly, no other Oryza MADS gene clusters with Arabidopsis MIKC* genes in a monophyletic clade. Therefore, the number of MIKC* genes in *Oryza sativa* is limited to 3.

In the database of *Selaginella moellendorffii* another 6 putative MIKC* sequences were found. These most likely represent 3 pairs of alleles, since *Selaginella moellendorffii* is diploid (Wang et al. 2005) and the cDNA sequences were extremely similar. The genes were designated as *SmMADS1*, *SmMADS2* and *SmMADS3* and the corresponding full-length cDNAs were cloned.

For *Ceratopteris richardii* the search at PlantGDB delivered one expressed sequence tag (accession number BE643398) coding for a partial putative MIKC* gene designated as *CRM13*, continuing already existing nomenclature. Another 3 new MIKC* genes (*CRM14*, *CRM15* and *CRM16*) were found by PCR using degenerate primers. For all sequences, full-length cDNAs as well as the genomic loci were amplified. Efforts to amplify the genomic locus of *CRM16* were not successful.

For *Aristolochia fimbriata* and *Eschscholzia californica* no MIKC* sequences were found in the databases. PCRs with degenerate primers yielded 2 MIKC* genes for each species. The *Eschscholzia californica* MIKC* genes were designated as *EcMADS1* and *EcMADS2* and the corresponding full-length cDNAs were amplified. Only the locus of *EcMADS1* could be isolated as a continuous sequence. For *EcMADS2*, 3 overlapping fragments were amplified that cover the whole locus. Since the DNA template was from a pool of individuals it is unsure whether the assembly of the 3 fragments represents a naturally occurring locus. For *Aristolochia fimbriata*, one complete cDNA could be isolated (*AfMADS1*). Another cDNA (*AfMADS2*) missed 6 triplets encoding the N-terminal amino acids of the MADS-domain. Sequencing 2 independent 5' RACE products retrieved this sequence. Only the locus of *AfMADS1* was isolated.

Table 1. Overview of the isolated sequences reported in this manuscript. Selaginella and Oryza loci are available from their respective databases.

species	Protein name	Length (aa)	Accession cDNA	Accession protein	Accession locus
Oryza sativa	OsMADS62	339	FM956504	CAX11684	-
	OsMADS68	383	FM956505	CAX11685	-
Eschscholzia californica	EcMADS1	361	FM958508	CAX16991	FM958507
	EcMADS2	339	FM958509	CAX16992	FM992874
					FM992875
					FM992876
Aristolochia fimbriata	AfMADS1	349	FN386479	CAY39417	FN386478
	AfMADS2	366	N/A	N/A	N/A
Ceratopteris richardii	CRM13	254	FM995267	CAX33868	FM995268
	CRM14	448	FM995269	CAX33870	FM995270
	CRM15	423	FM995271	CAX33872	FM995272
	CRM16	238	FM995273	CAX33874	N/A
Selaginella moellendorffii	SmMADS1 allele 1	371	FM999804	CAX46406	-
	SmMADS1 allele 2	371	FM999805	CAX46407	-
	SmMADS2 allele 1	230	FM999806	CAX46408	-
	SmMADS3 allele 1	447	FM999807	CAX46409	-
	SmMADS3 allele 2	447	FM999808	CAX46410	-

All sequence data from this manuscript has been submitted to the EMBL nucleotide sequence database. An overview of all isolated sequences, nomenclature and protein length can be found in table 1.

3.2 Sequence comparison of the K-domain of MIKC* and MIKC^c genes reveals an ancient duplication

Clustal W analyses with different protein weight matrices all aligned 3 regions C-terminal of the MADS-domain of MIKC* proteins with the K1, K2 and K3 subdomains of the Keratin-like domain of classical MIKC proteins that contain heptad repeats of the hydrophobic residues V, I, L and M (data not shown). A certain similarity of these 3 regions with the MIKC^c K-domain has been noticed before but
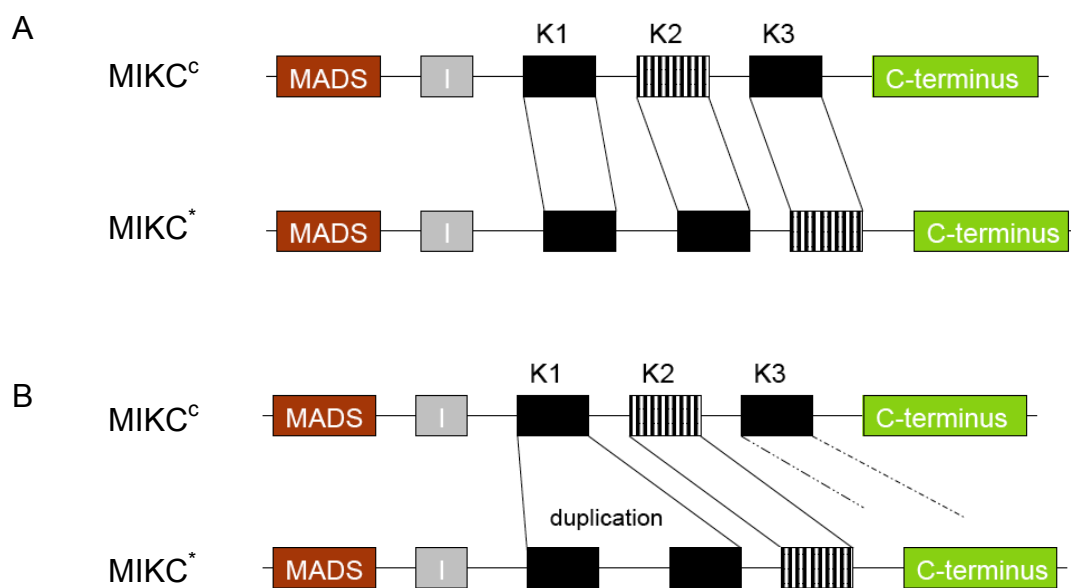
Figure 4. A comparison of 2 alternative alignments between MIKC^c and MIKC* genes with a simplified gene exon-intron structure. Striped boxes have a high homology to each other. **(A)** depicts the alignment preferred by Clustal W analyses. **(B)** shows that it is unclear to which MIKC* K-domain the MIKC^c K1 region should be aligned.

the K1-subdomain of MIKC* genes was nonetheless interpreted as elongated Intervening region (Henschel et al., 2002). Although the recognition of the "new" K1 in MIKC* genes has made the number of heptad repeat regions between MIKC* and MIKC^c become equal, the order in which they should be aligned is *not* straightforward. Two types of alignments were produced, which are schematically depicted in figure 4. A benefit of the first alignment (figure 4A is that it has the least gaps, but, there the fact is dismissed that the K2 subdomain of MIKC^c proteins has a much higher homology to K3 of MIKC* genes than to the K2 of MIKC* proteins. In
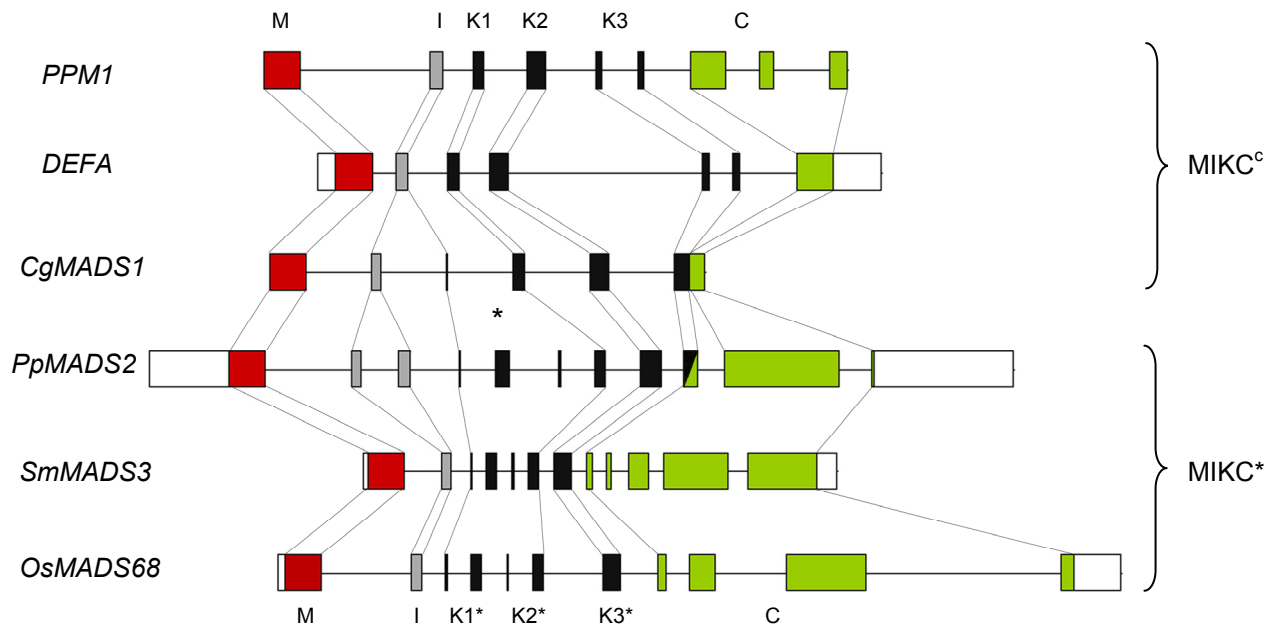
Figure 5. Comparison of the exon-intron structures of MIKC^c and MIKC* genes. UTRs are in white boxes and protein domains in coloured boxes. The asterisk indicates a duplication event. It must be noticed that the third exon of *CgMADS1* has been postulated to belong to its I-domain but that the homologous exon in MIKC* genes has been recruited to the K-domain. The half-shaded box in *PpMADS2* indicates that homology between MIKC^c and MIKC* genes is not recognizable in that exon. Since the lengths of the introns of *CgMADS1* have not been published, an arbitrary length of 300 bp has been assigned here. An alignment showing the positioning of the heptad repeat domains is given in supplemental figure S1.)

the second alignment (figure 4B the more homologous MIKC^c K2 and MIKC* K3 subdomains are aligned. Naturally, only one of these alignments reflects evolutionary history correctly. When also the exon-intron structure in this region is taken into account, shown in figure 5, it can be seen that the K1 and K2 region of MIKC* proteins are actually encoded by 4 exons: 2 modules of a smaller and a relatively longer exon. This suggests that a duplication event occurred early in the evolution of MIKC* genes that gave rise to K1 and K2 regions of MIKC* genes.

3.3   Land plants have multiple classes of MIKC* proteins

Unfortunately, the duplication scenario that was launched in the previous section does not offer an answer to the question whether the K1 of MIKC^c proteins should be aligned to the K1 or the K2 region of MIKC* proteins. A phylogenetic tree based on an alignment where these paralogous regions were completely left out, did not result in a tree with high bootstrap support (data not shown). It was therefore opted to make

2 different alignments. In the first alignment the MIKC* K1 subdomain was aligned with the MIKC^c K1 and in the second alignment with MIKC^c K2. The latter possibility gave the highest degree of homology (data not shown). Previous authors also preferred this type of alignment (Henschel et al. 2002, Tanabe et al. 2005). Additionally, the structure within the MIKC*-clade of the tree derived from this alignment did not deviate from the topology of a phylogenetic tree derived from an alignment of MIKC* sequences alone (data not shown).

Figure 6 shows the unrooted neigbour joining tree from the alignment of MIKC^c and MIKC* proteins. MIKC* and MIKC^c proteins clearly form 2 separate clades. Within the MIKC* clade, 3 highly supported groups can be observed. One contains the MIKC* proteins of Physcomitrella and the other 2 contain all the MIKC* sequences from Ceratopteris and the angiosperms. Noticeably, the relationship of the MIKC* proteins of the lycophyte *Selaginella moellendorffii*, to the other MIKC* sequences or among themselves stayed unresolved. Perhaps a more thorough sampling in the basal land plant lineages, which is at the moment hampered by the lack of available sequences, will clear up that relationship.

The MIKC* genes from ferns and seed plants are found in 2 subclades, which correspond to the S- and P-clade of MIKC* proteins as determined by an analysis of Nam and coworkers (2004) based on Oryza, Arabidopsis and Tobacco MADS-box genes. It can be seen that not all species have representatives in both clades though. Probably, this is due to the incompleteness of the expressed sequence tag database since representatives from both clades were found in all the model species investigated in this manuscript. Furthermore, if the genomes of all the currently available angiosperms are searched, at least 2, divergent, MIKC* genes can be recognized (data not shown).

Within the angiosperm S-clade of MIKC* proteins another clear bifurcation is present. Figure 7 shows how MIKC* sequences from these 2 S-subclades are distributed across the orders of extant angiosperms. Obviously, no plants were found that carry both kinds of S-subclade genes. However, plants in similar groups of angiosperms can possess different S-subclade genes.
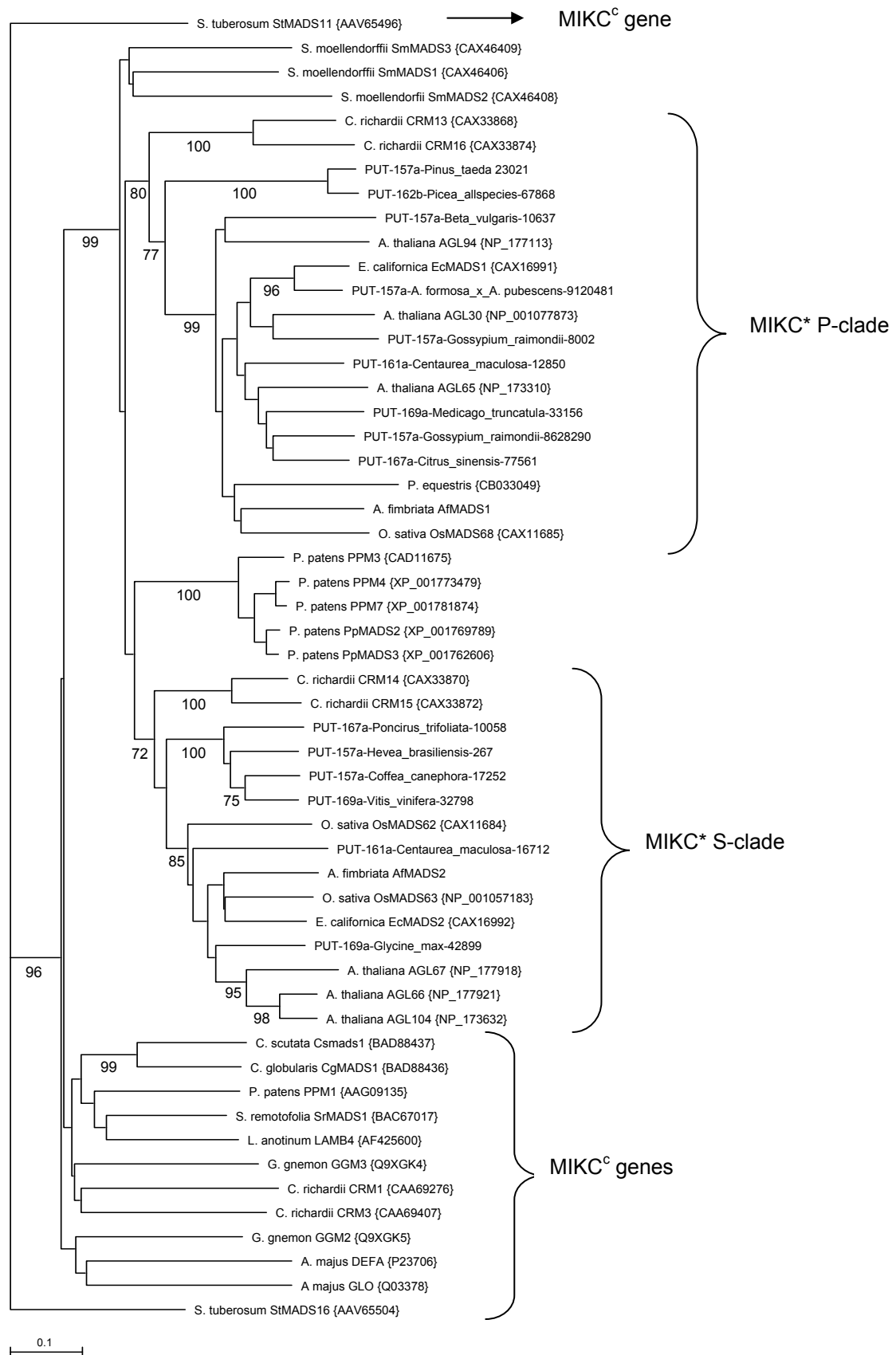
Figure 6. Unrooted neighbour joining tree of MIKC* and MIKC^c genes based on the alignment given in supplemental figure S1. Bootstrap values are shown below branches only when support was higher than 70%. Bootstrap support within the *Physcomitrella patens* subclade is not shown. Accession numbers of the used sequences are shown in brackets or their plantGDB putative transcript is given.
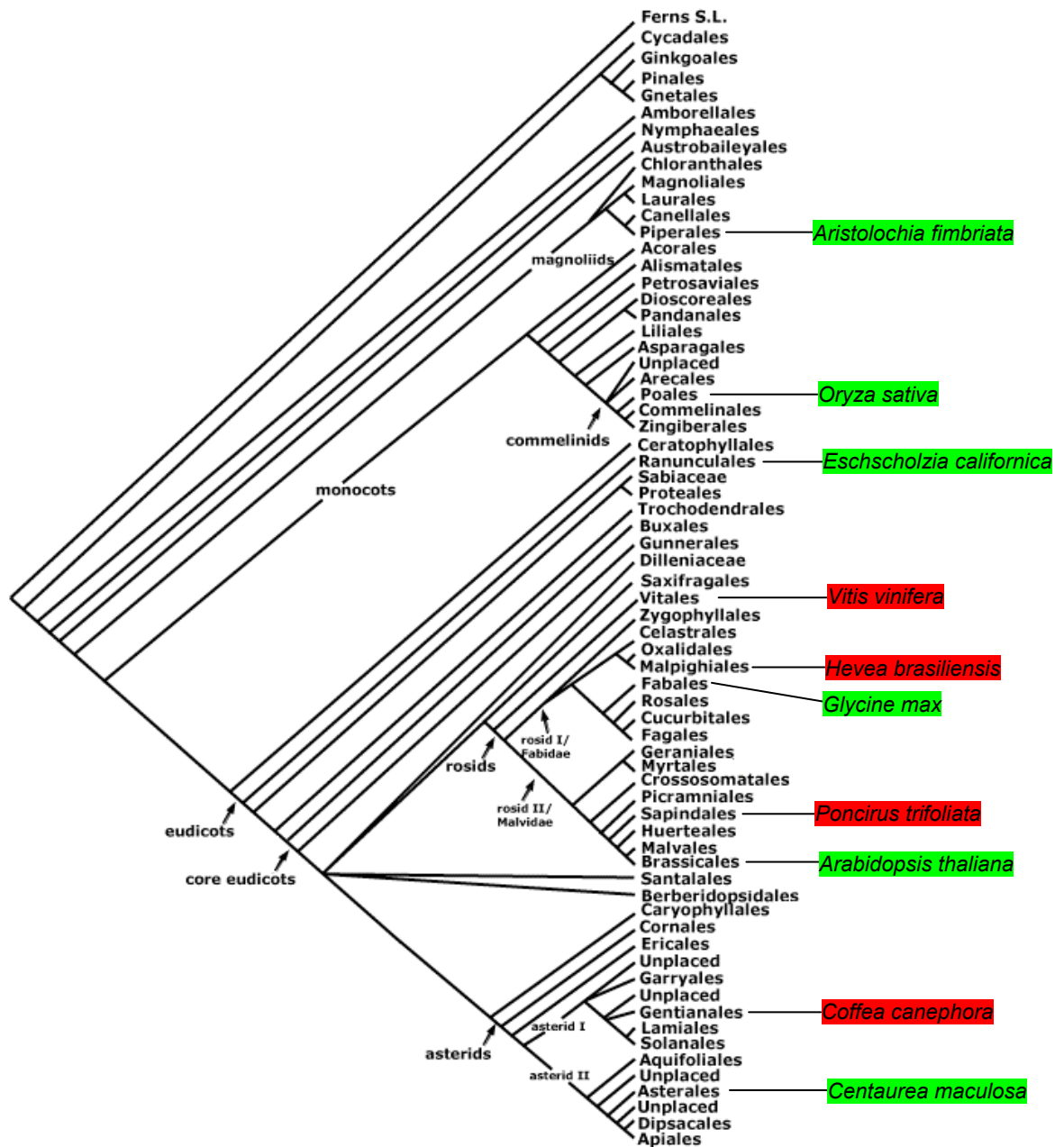
Figure 7. Distribution of the 2 MIKC* S-sublades over extant groups of angiosperms. Angiosperms species from which the different S-clade genes were isolated are in red and green. The ancestors of the 2 rosid groups Fabidae and Malvidae as well as the ancestor of the asterids must have had 2 types of S-subclade MIKC* genes. Branches of the phylogenetic tree terminate in taxonomic orders. Adapted from the Angiosperm Phylogeny Group website (http://www.mobot.org/mobot/research/apweb/welcome.html).

3.4 MIKC* genes are expressed specifically in gametophytes (and roots?)

Most of the identified MIKC* expressed sequence tags from the database originated from cDNA collections obtained from tissues that contained gametophytes. Remarkably, the EST from *Beta vulgaris* (sugar beet) came from a cDNA library of roots and the one of *Pinus taeda* (Loblolly pine) from flooded roots. Interestingly, according to the TAIR Arabidopsis eFP Browser (http://www.arabidopsis.org), which is based on microarray expression data, the Arabidopsis MIKC* gene *AGL67* is expressed in the root of the embryo.
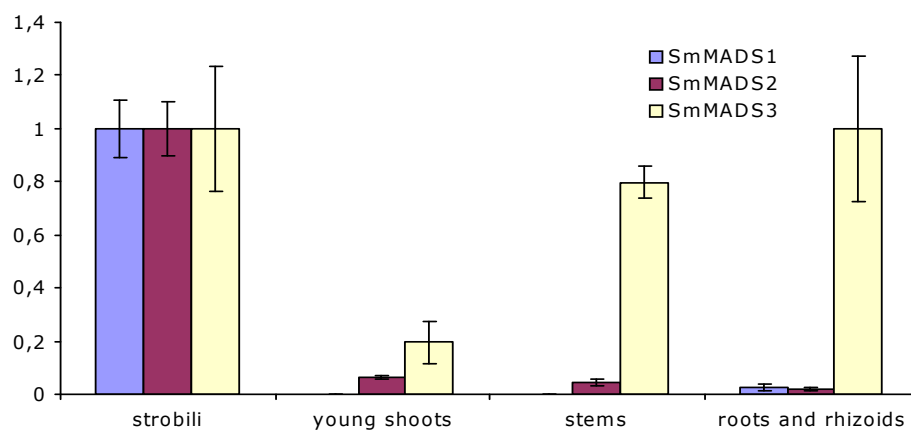
Figure 8. Relative expression levels of the Selaginella MIKC* genes obtained by qPCR. *SmMADS1* and *SmMADS2* are almost exclusively expressed in strobili. Expression of *SmMADS3* is more ubiquitous, and highest in stobili and roots and rhizoids. Bars represent standard deviation.

All the MIKC* cDNAs that were isolated and reported on in this manuscript originated from gametophytic tissues. Only for the model species *Selaginella moellendorffii* a detailed expression analyses was performed.

Quantitative real-time PCR was used to determine the relative expression levels of MIKC* genes in different plant tissues of *Selaginella moellendorffii*. Figure 8 shows that *SmMADS1* and *SmMADS2* are highest expressed in the gametophyte-containing strobili. *SmMADS3* expression was equally high in strobili and roots and rhizoids. Also, substantial expression of *SmMADS3* was detected in young shoots and stems. Similar expression patterns were found in 3 individual plants for *SmMADS1* and *SmMADS2*. In one out of 3 plants, however, expression of *SmMADS3* could not be

detected in stem tissue, although RNA input and the expression level of reference genes was similar (data not shown).

3.5   MIKC* proteins from different species show variable interactions

MADS-domain proteins need to form dimers to be able to bind DNA and work as a transcription factor. Here, the Y2H system was used to asses which dimeric transcription factor complexes could be functioning in the species from which MIKC* transcripts were isolated. W. Verelst (Verelst et al. 2007a) has shown that Arabdopsis MIKC* dimers always contain one protein of the S- and one protein of
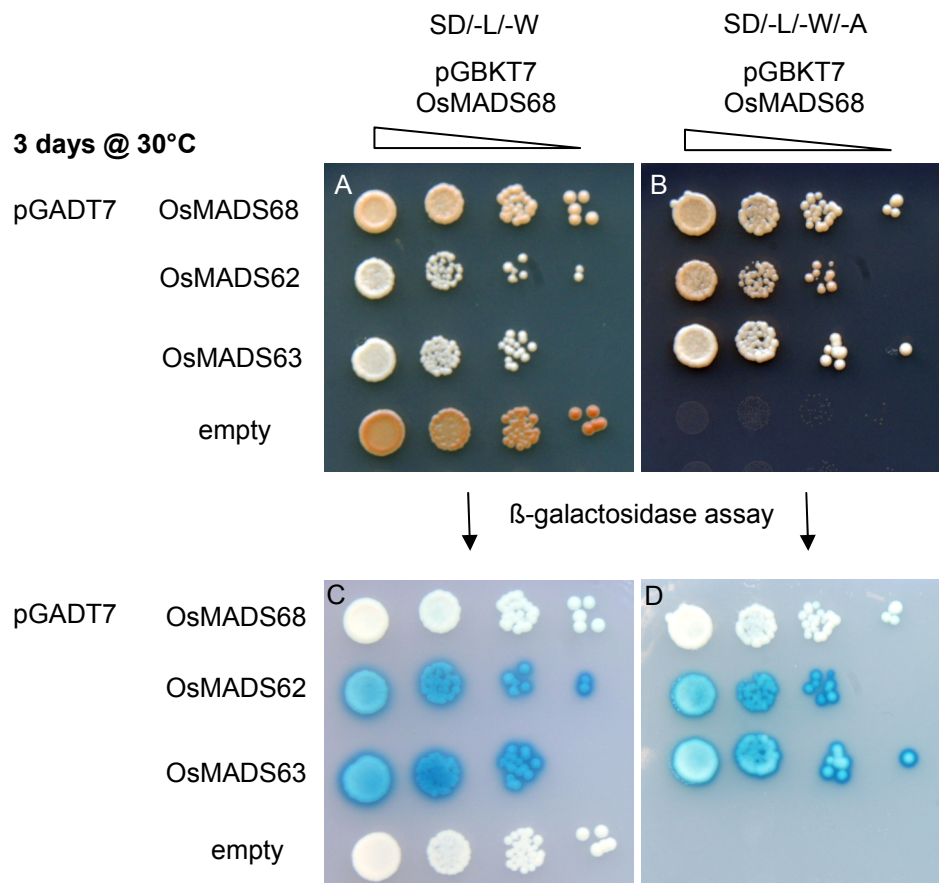
Figure 9. A part of the spot dilution assay performed for *Oryza sativa* MIKC* proteins. **(A)** All cotransformants were able to grow on SD/-L/-W medium. **(B)**. Cotransformants with an empty pGADT7 vector showed no growth on SD/-L/-W/-A medium. **(C, D)** The same yeast colonies as **(A)** and **(B)** after a ß-galactosidase assay. A stonger blue colouring indicates a stronger ß-galactosidase activity caused by a strong interaction of the Oryza MIKC*-GAL4 DNA binding and activation domain protein fusions. The complete matrix can be found in supplemental figures S2-S7.

the P-clade. The MIKC* protein interactions of the model species tested below, showed that variations from this strict heterodimerization are common.

3.5.1 *Oryza sativa* MIKC* protein interactions

Since all the yeast cotransformants were able to grow on selective SD/–L/-W/-A medium it must be concluded that all interactions between the fusions of the Oryza MIKC* proteins with the GAL4 DNA binding and activation domains were possible. The interactions were not equally strong for all combinations of protein fusions though. This can already be observed in figure 9A and 9B that show growth on SD/-L/-W medium, where there is selected for cotransformants and not for protein

prey \ bait	P-clade OsMADS68		S-clade OsMADS62		OsMADS63		pGBKT7
	Δ262-383	Δ299-383	Δ229-339	Δ266-339	Δ214-360	Δ294-360	
OsMADS68							
OsMADS62							
OsMADS63							
pGADT7							

Figure 10. Schematic depiction of all tested *Oryza sativa* fusion protein interactions. A darker colouring of the boxes represents stronger interactions. Homodimers are formed but the heterodimers consisting of S- and P-clade proteins showed the strongest interaction in a ß-galactosidase assay. The complete matrix can be found in supplemental figure S2-S7.

interaction. Some cotransformants show a reddish coloring caused by a low adenine production by the ADE2 selection marker, although there is adenine in the medium. As expected, the reddest colonies are from a cotransformation with an empty pGADT7 vector, where the activation domain is not fused to any Oryza protein. From this can be concluded that the activation domain alone probably cannot interact with the DNA binding fusion to drive ADE2 expression and it excludes the possibility for false positives by so-called auto-activation to occur. Indeed, it can be seen in figure 9B that the yeast transformed with the bait fusion construct in

combination with the empty prey vector is not able to grow on selective SD/-L/-W/-A medium in contrast to all the other plated cotransformants.

Differences in interaction strength were confirmed by the ß-galactosidase assay as exemplified in figure 9C and 9D. A much stronger ß-galactosidase activity was observed from yeast colonies that harbored heterodimeric MIKC* complexes. Moreover, when all interactions are taken into account, shown in figure 10, it can be noticed that the strongest interactions occurred when heterodimers were formed between the members of the phylogenetically distinct S- and P-clades. Additionally, homodimers of S- and P-clade protein fusions were observed to show a different interaction strength depending on the length of the C-terminal deletions of the bait. Interactions tended to be less strong when the C-terminal domain was shorter. This decrease is most notable for the OsMADS63 Δ214-360 fusion, which is not able to homodimerize anymore on SD/-L/-W/-A selective medium (supplemental figure S7). These results point to a function of the C-terminus in protein-protein interaction for Oryza MIKC* proteins.

3.5.2 *Eschscholzia californica* MIKC* protein interactions

Figure 11 summarizes the interaction data for Eschscholzia. It can be seen that EcMADS1 can form obligatory heterodimers with EcMADS2. EcMADS2 could also form homodimers. In contrast to the the Oryza interactions, no large differences

	P-clade	S-clade	
bait / prey	EcMADS1	EcMADS2	pGADT7
EcMADS1 Δ 243-361		▓▓▓	
EcMADS2 Δ203-339		░░░	
EcMADS2 Δ278-339		▒▒▒	
pGBKT7			

Figure 11. Schematic depiction of Eschscholzia fusion protein interactions. EcMADS2 can form homodimers and EcMADS1 only heterodimers. Darker colouring indicates a stronger interaction. Original data can be found in supplemental figure S8-9.

between the interaction strength of homo- and heterodimers were observed in the ß-galactosidase assay on SD/-L/-W/-A medium (supplemental figures S8-S9). Also no difference between interactions with the 2 different C-terminal deletion proteins of EcMADS2 could be observed.

3.5.3   *Ceratopteris richardii* MIKC* protein interactions

In contrast to all other MIKC* proteins tested, fusions of Ceratopteris MIKC* genes with the GAL4 DNA binding domain were not able to (auto) activate ADE2 expression. This made the production of C-terminal deletion constructs unnecessary. In figure 12, it can be seen that Ceratopteris S-clade proteins cannot form homodimers, unlike the Eschscholzia S-clade protein. The Ceratopteris P-clade proteins do form homodimers and, additionally, are also able to form heterodimers within their own clade, again contrasting the interactions observed for Eschscholzia MIKC* proteins. Heterodimers between S- and P-clade members could also be detected.

bait / prey	P-clade		S-clade		
	CRM13	CRM16	CRM14	CRM15	pGBK
CRM13	▨ (light)	▨ (light)			
CRM16	■ (dark)	■ (dark)	■ (dark)	▢ (very light)	
CRM14					
CRM15					
pGBK					

Figure 12. Schematic summary of *Ceratopteris richardii* MIKC* proteins interactions. Darker colouring indicates a stronger interaction.

3.5.4 *Selaginella moellendorffii* MIKC* protein interactions

For the Selaginella MIKC* genes phylogenetic analysis did not cluster them in the S- or P-clades, which makes a comparison to the MIKC* genes of the other organisms difficult. All 3 proteins showed a different dimerization behaviour. SmMADS1 could only form heterodimers while SmMADS2 formed homo- and heterodimers, albeit not all interactions could be reciprocally verified and the interaction of SmMADS1 with SmMADS3 was very weak. SmMADS3 further was capable of forming homodimers.

prey bait	SmMADS1	SmMADS2	SmMADS3	pGADT7
SmMADS1		■		
SmMADS2	■	■		
SmMADS3 Δ 240-447				
SmMADS3 Δ 294-447	░	■	■	
pGBKT7				

Figure 13. Schematic summary of *Selaginella moellendorffii* MIKC* proteins interactions. Darker colouring indicates a stronger interaction.

3.6   Arabidopsis T1 transgenic lines show high copy numbers

Arabidopsis *agl66/104* double mutant plants were transformed with MIKC* genes from a broad variety of land plants to assess whether the mutant phenotype could be rescued. Initially, a qPCR strategy was designed to select single copy T-DNA integration lines in the T1 generation. Single copy lines have the benefit that a minimum of disturbance due to the integration of the T-DNA in the host genome occurs and that the expression levels of the transgene is theoretically similar to expression of the gene of which the promoter was used. In figure 14, it can be seen that the overall copy number of *AGL65* promoters in 6 tested T1 lines was very high compared to the situation in wild type Arabidopsis. Still, in many of these lines only half of the pollen was fluorescent, which is characteristic for a single integration site (data not shown). This indicated that many copies were integrated at the same position. Because *single copy* lines were expected to be very rare and selecting them thus very laborious, *single integration* lines were chosen for further analysis.
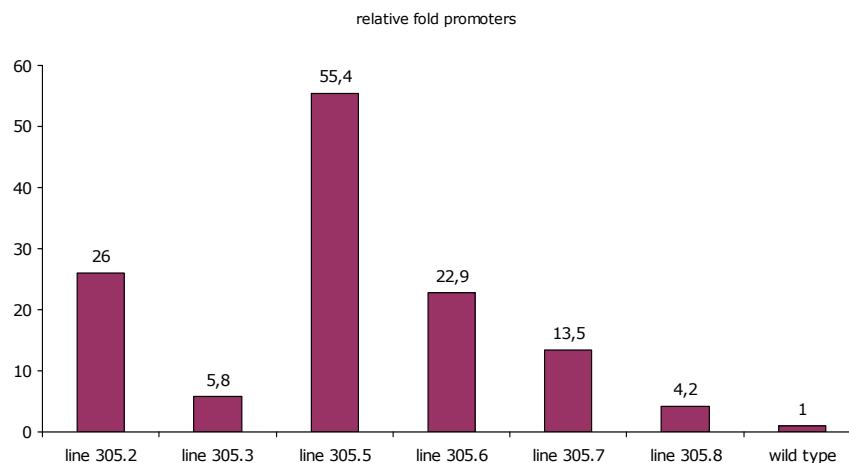
Figure 14. Relative fold of *AGL65* promoters in 6 independent T1 lines transformed with pGJ2364 containing the OsMADS62-GFP fusion under control of the *AGL65* promoter. In a diploid wild type genome 2 *AGL65* promoters are present, corresponding to a value of 1 on the relative fold promoters axis.

3.7   The pollen germination phenotype of the Arabidopsis *agl66/104* double mutant can be rescued by heterologous expression of MIKC* genes from distantly related species

To test the functionality of the *AGL65* promoter that was used in the binary vector, the promoter-*AGL65-GFP* fusion was transformed in Arabidopsis *agl66/104* double mutants, which have T-DNA insertions in the 2 of the 3 S-clade genes, namely *AGL66* and *AGL104*. Figure 15 shows that these lines exhibit a strong GFP signal in the cytoplasm whereas, for example, the pollen of plants that express a protein fusion
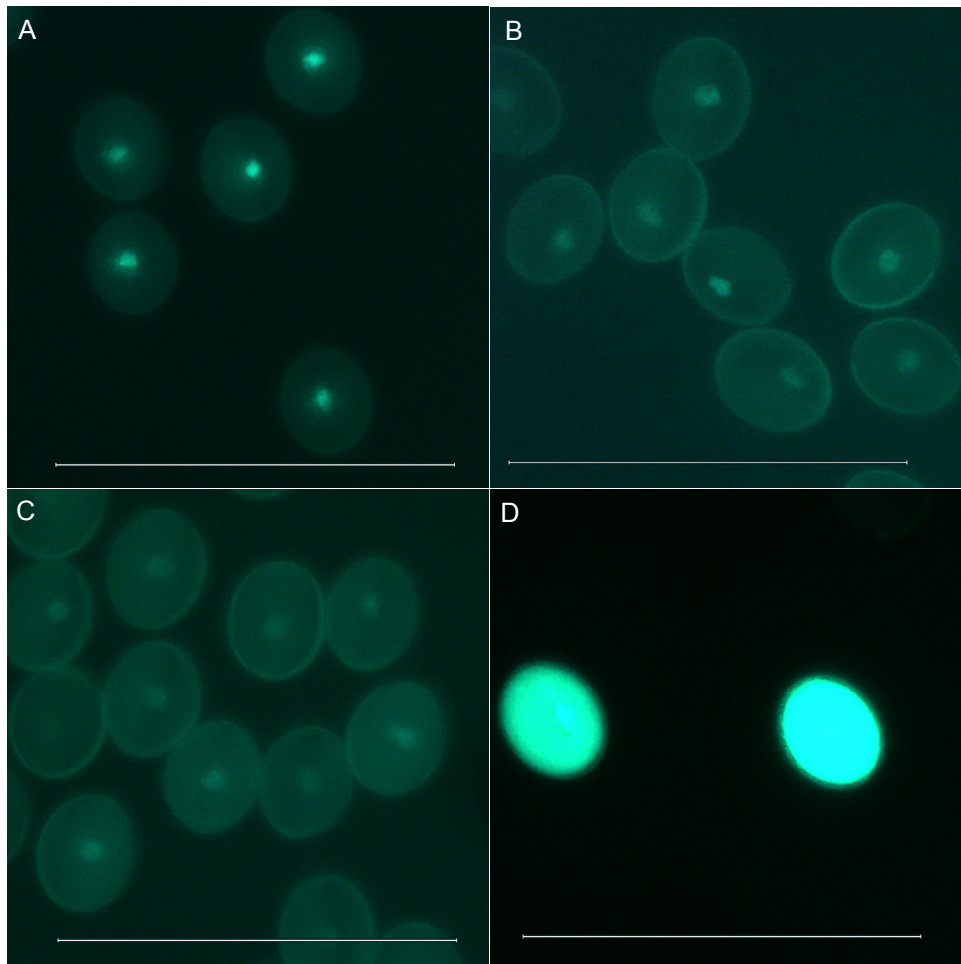
Figure 15. GFP fluorescence of pollen from the T3 generation of Arabidopsis transformants. **(A)** Pollen from plants transformed with the *AGL66-GFP* construct. **(B, C)** Pollen from the *SmMADS3-GFP* line 1 and 2 both shows a GFP signal in the vegetative nucleus. The picture in **(C)** has been taken using a 4x longer exposure time than **(B)**. **(D)** Arabidopsis pollen that only possessed a promoter-*AGL65-GFP* fusion showed a signal in the complete cytoplasm of the vegetative cell. Scale bars represent 100 μm.

of SmMADS3 and GFP, the GFP signal is restricted to the nucleus of the vegetative cell. Also the ovules were checked for a GFP signal but this could not be detected (data not shown), thus indicating that the *AGL65* promoter is specific for the male gametophyte.
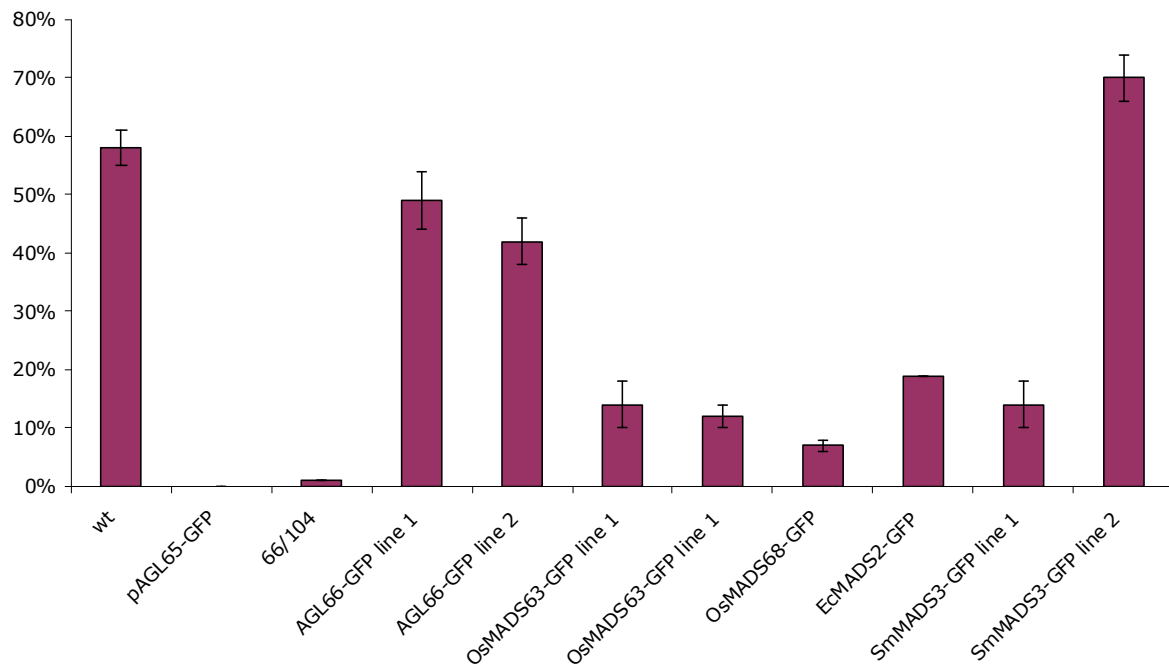
Figure 16. *In vitro* pollen germination rates of wild type, mutant and complemented Arabidopsis plants. Bars indicate standard error.

To test whether it is possible to complement the mutant phenotype using MIKC* sequences driven by the *AGL65* promoter, Arabidopsis double mutants were transformed with a construct containing the Arabidopsis S-clade MIKC* gene *AGL66* fused to GFP. Indeed it was possible to rescue the *in vitro* pollen germination defect of the mutant, in contrast to double mutants that carry only a promoter-*AGL65-GFP* fusion. Furthermore, the GFP signal from the AGL66-GFP fusions was observed in the vegetative nucleus indicating a proper targeting of the transcription factor (figure 15A). The ability to complement the pollen germination defect was investigated more precisely by a quantitative pollen germination assay. Pollen from 2 independent single integration *AGL66-GFP* homozygous T3 lines was applied to agarose pads and the number of pollen tubes that developed was counted. In figure

16 it can be seen that both lines carrying the *AGL66-GFP* gene restored the pollen germination close to wild type levels. Pollen germination rates of the double mutant and plants transformed with only the promoter-GFP fusion showed minimal germination (figure 16).

To investigate if MIKC* transcription factors of the other (vascular) plants can perform the same function as the MIKC* genes of Arabidopsis their ability to complement the Arabidopsis S-clade mutant was tested. Double mutant Arabidopsis plants were transformed independently with all the MIKC* homologs that were isolated except for the MIKC* genes of Aristolochia. Quantitative pollen germination assays were performed with putative homozygous single integration T3 lines.

In Oryza, 3 MIKC* proteins are present of which 2 belong to the S-clade and one to the P-clade (Nam et al. 2004, this study). Pollen germination assays indicated that all 3 members were found to complement the *in vitro* germination defect. Subsequently, the germination rate of 2 independent *OsMADS63-GFP* Arabidopsis lines was quantified, shown in figure 16. Clearly the germination defect is complemented but to a less strong degree in comparison to the AGL66-GFP fusion line. Also the P-clade member *OsMADS68* could rescue the mutant phenotype but not as well as the OsMADS63-GFP fusion (figure 16). Fluorescence of the OsMADS68-GFP fusion was localized to the nucleus, similar as for the S-clade fusions (data not shown).
For Eschscholzia only transformants that carried the S-clade gene *EcMADS2* showed an enhanced pollen germination (figure 16) although both S- and P-clade MIKC* genes were targeted to the nucleus (data not shown).

Phylogenetic analysis has shown that the MIKC* genes of Ceratopteris are also divided over the S- and P-clades. However, neither transformants containing S- or P-clade members showed an increase in pollen germination (data not shown). Furthermore, pollen fluorescence was very low and cytoplasmically localized or not observable. Possibly the GFP moiety rendered the Ceratopteris MIKC* proteins dysfunctional thereby preventing complementation of the pollen germination phenotype. To verify this, also non-GFP lines were tested but also showed no rescue

(data not shown). Sterical hindrance by the GFP peptide as a cause for non-complementation could therefore be excluded.

The Selaginella SmMADS3-GFP protein fusion showed proper targeting to the nucleus (figure 15) and rescue of the pollen germination phenotype (figure 16). Interestingly, the germination efficiency of pollen from *SmMADS3-GFP* line 2 showed a higher germination rate than wild type pollen as opposed to the performance of *SmMADS3-GFP* line 1. Furthermore the GFP signal in pollen from *SmMADS3-GFP* line 2 was higher than that of line 1 (Figure 15). *SmMADS1*- and *SmMADS2-GFP* lines behaved similar as the Ceratopteris MIKC*-GFP lines with low to non-existent GFP signals (data not shown). Influence of the GFP fusion was also tested for SmMADS1 and it could be shown that when not fused to GFP, it can complement the mutant phenotype (data not shown). For SmMADS2 without GFP no data is available.

Because MIKC* genes as distantly related as those of Selaginella still complemented the Arabidopsis mutant, it was decided to investigate if even more distantly related proteins were similarly successful. Therefore the MIKC* genes *PPM3* and *PPM4* from *Physcomitella patens*, the classical MIKC^c genes *AGL15* and *AGAMOUS* from Arabidopsis, as well as the human Myocyte Enhancer Factor *MEF2A* were transformed in the *agl66/104–1* double mutant as a fusion with GFP or unfused. Although fluorescence of the GFP signal was localized in the vegetative nucleus, none of plants tested showed an increase in pollen germination efficiency (data not shown).

4   Discussion

4.1   A new origin for MIKC* genes

Traditionally, MIKC* transcription factors have been defined as MADS-domain proteins with a "classical" MIKC domain structure but with a longer Intervening region and a K-domain that has less regularly spaced hydrophobic residues than classical MIKC^c genes (Henschel et al. 2002, Riese et al. 2005). Since the first discovery of MIKC* genes in the moss *Physcomitrella patens*, the origin of this group has remained unclear (Henschel et al. 2002, Parenicova et al. 2003, Tanabe et al. 2005, Riese et al. 2005) and hypotheses ranging from duplication of the Intervening region of the MIKC^c genes to a *de novo* insertion in the K–domain have been drawn up as an explanation (Henschel et al. 2002). This controversy is now resolved. By aligning members of the MIKC^c and MIKC* gene family from a broad variety of land plants, a part of the elongated I-domain could be shown, in fact, to contain homology to a heptad repeat motif, a hallmark of the K-domain (Ma et al. 1991, Jack 2004, reviewed in Kaufman et al. 2005). Furthermore, the exon-intron structure in this part of the gene supports that a duplication event gave rise to this "new" K-domain. In the MIKC^c gene *CgMADS1* from *Chara globularis* the ancestral unduplicated status is still apparent.

Possibly, the genesis of MIKC* genes started with a duplication event of a classical-like MIKC gene. A subsequent intragenic duplication in one of the copies in the region that encodes the K-domain may potentially have prevented that its translated protein could dimerize with the ancestral MIKC protein from the structurally unchanged gene. In support of this, a study of the dimerization behaviour of pollen-specific Arabidopsis MIKC* proteins pointed out that they cannot interact with pollen expressed MIKC^c genes (Verelst et al. 2005a). The putative newly attained "isolation" could have allowed for mutations to accumulate without the danger that they result in interference with the proper functioning of the paralog, achieving more freedom for sub- or neofunctionalization.

Interestingly, MIKC* genes have only been identified in land plants so far, suggesting that they evolved from MIKC^c genes early in land plant evolution and opening the

possibility that they might have played a considerable role in establishing the land plant life style.

4.2   The ancestor of ferns and seed plants had 2 divergent MIKC* genes that can form heterodimers

Due to the described duplication in the K-domain of MIKC* genes caution was needed when making an alignment because only a small part of the classical K-domain had a true orthologous relationship with MIKC* genes. Unfortunately, a phylogenetic tree calculated using the neighbour joining method and based on a multiple alignment of only unambiguously alignable and orthologous stretches of MIKC^c and MIKC* genes did not lead to a tree supported by high bootstrap values. Therefore, a phylogenetic tree was constructed using an alignment that also contained a part of the duplicated region. This tree showed that the sequences from the fern *Ceratopteris richardii* and those of angiosperms can be divided in 2 monophyletic subgroups indicating that the ancestor of ferns and seed plants had 2 divergent MIKC* genes. Previously, this subdivision had already been recognized in monocots and eudicots (Nam et al. 2004). The phylogenetic evidence presented here argues that the duplication event leading to the S- and P-clade already took place in the ancestor of ferns and seed plants. Furthermore, in all the model species investigated here, except Selaginella, and in all fully sequenced angiosperms, members of each group could be identified. Probably, the conservation of both members in the genomes has been beneficial for the survival of ferns and angiosperms. Evidence for the molecular functioning of these 2 subgroups comes from yeast-2-hybrid experiments, although it must be mentioned that it cannot be excluded that some interactions for Eschscholzia and Ceratopteris were missed, since possibly not all their MIKC* genes were retrieved. However, the biological relevance of the interactions can be warranted because all cDNAs were extracted from gametophytic tissues. With this prerequisite settled, a quick glance on the yeast-2-hybrid data already points out that different networks of MIKC* interactions have evolved in all species. However, it is striking that all MIKC* networks are marked by the possibility of heterodimerization between S- and P-clade members. Homodimerizing capacities in the different

organisms might reflect a derived adaptation of the MIKC* network to the demands of that specific species. On the other hand, the ancestral, single, MIKC* gene must have been able to form homodimers. Perhaps that from this starting point a more sophisticated network evolved where many more combinations of dimerization were possible. Such a network could also exist, for example, in the moss *Physcomitrella patens*, which has 11 MIKC* genes (Rensing et al. 2008). The emergence of the S- and P-clades during the early evolution of vascular plants might reflect a reduction of that network to its core function.

4.3 Angiosperms evolved 2 kinds of S-clade genes

Superimposing the phylogenetic tree on the species tree further revealed that the S-clade of angiosperms is split in yet another 2 subclades. Members of one clade are only found in core eudicots but again, it must be kept in mind that similar sequences might also be present in more basal plants but have escaped identification so far. Also, it must still be noted that no species has been identified that carries both S-subclade members. It is unlikely though that the presence of a member of one clade excludes the presence of the member of the other clade since multiple ancestors must have had members of both groups to explain the observed distribution of the S-subclades genes in extant eudicots. Still the question remains whether these subdivisions reflect a functional divergence. Unfortunately, the S-clade genes from the species tested in the yeast-2-hybrid study all belonged to the same subclade, so insight in the dimerization properties between the distinctive clades is not available. Also, no specific traits could be discovered that are associated with the differences of S-subclade members among plants. Since the morphology of MIKC* mutant Arabidopsis pollen did not change compared to wild type (Verelst et al. 2007a,b, Adamczyk and Fernandez 2009) it is doubtful whether the proposed functional divergence between the pollen of plants that carry different S-subclade MIKC* genes would result in phenotypic differences.

4.4 Bryophyte and lycophyte MIKC* genes share characteristics with those of fern and angiosperms

So far, the interpretation of phylogenetic and yeast-2-hybrid data of the more basal MIKC* genes presented in the results has been omitted. Unfortunately, the phylogenetic tree shown in figure 6 is not informative concerning which MIKC* group is basal. Dimerization properties of the Selaginella proteins could support that the *SmMADS1* and *SmMADS2* genes actually belong to the S- and P-clades since the yeast-2-hybrid experiments showed that they display a strong heterodimeric interaction. The distant relationship to the other MIKC* sequences might have prevented them to be recognized as such by phylogenetic approaches. The homodimerization of SmMADS3 might be linked to its expression that is also apparent outside of the gametophytic phase, where it possibly has no other interaction partners since *SmMADS1* and *SmMADS2* were predominantly expressed in the gametophyte-bearing strobili. From Physcomitrella MIKC* proteins no interaction data is available in the literature. Interestingly, the exon-intron structure of *PpMADS2* and others show an extra exon in the I-domain, which is probably derived from a duplication event (described in Henschel et al. 2002 and not to be confused with the previously discussed duplication inside of the K-domain). Such an extra exon is also present in the S-clade genes of Arabidopis as well as in the S-clade gene of *Glycine max* identified here. Furthermore, *CRM13* and *CRM14* of Ceratopteris have an N-terminal extension of the third exon. Whether these are traces of an orthology relationship is unsure but it certainly suggests that bryophyte MIKC* genes are more related to MIKC* genes of vascular plants than the phylogenetic analyses revealed.

4.5 How is complementation of the *in vitro* pollen germination defect accomplished?

In Arabidopsis, MIKC* dimers are exclusively formed between members of the S- and P-clade. Consequently, when the expression of MIKC* genes from one of the clades is severely reduced, like in the *agl66/104-1* double mutant used here, only a

small amount of dimers can be formed and imported into the nucleus. This reduction of transcription factors is the likely cause of the *in vitro* pollen germination defect (Verelst et al. 2007a, b; Adamczyk and Fernandez 2009).

As a positive control in the experiments described in this manuscript the *agl66/104-1* double mutant was transformed with the coding sequence of the S-clade gene *AGL66* fused to GFP and driven by the *AGL65* promoter. It was expected that the introduction of this transgene would lead again to elevated levels of one of the S-clade proteins allowing for MIKC* dimers to be formed again and hence rescuing the phenotype. Indeed, the *AGL66-GFP* lines showed a strong nuclear localized signal and pollen germination assays revealed that the germination rate of the *AGL66-GFP* complementation lines was restored, however, not entirely to wild type levels. Apparently, the construct sequence that was transformed in the mutant was not sufficient for a full complementation. At the protein level, the only difference between the wild type and the lines complemented with the *AGL66-GFP* fusion is that the latter has a reduced expression level of AGL104 from the *agl104-1* locus. Can this reduced expression level cause the small discrepancy? In Verelst et al. (2007a) the germination rate of the *agl104-1* mutant is reported not to be different from wild type levels. Also the transcriptome of *agl104-1* pollen only has a different expression of 4 out 1373 genes that are putatively regulated by MIKC* proteins (Verelst et al. 2007b). Although these genes could be of significance, the cause for the discrepancy might rather lie in the use of the *AGL65* promoter to drive expression of the *AGL66-GFP* fusion. Although both *AGL65* and *AGL66* are specifically expressed in pollen (Verelst et al 2007a) the *AGL65* promoter is active already much earlier during pollen development. Furthermore, *AGL65* transcript levels are approximately 14 times higher in mature pollen than *AGL66* levels (Adamczyk and Fernandez 2009). This putative difference in promoter activity might potentially disturb the equilibrium between different MIKC* complexes, which regulate partly overlapping target genes (Verelst et al. 2007b). Moreover, using qPCR it was also observed that the transformed Arabidopsis lines can have a large number of constructs in their genome, thereby boosting the expression level of the transgene. Additionally, because the MIKC* network has strong features of autoregulation (Verelst et al. 2007b, Adamczyck and Fernandez 2009) the overexpression of one

member by the *AGL66* rescue construct is likely to feedback on the whole network. In conclusion, the mutant pollen defect can almost fully be complemented, but, there are multiple indications that the expression pattern of *AGL66-GFP* in mutant pollen might differ from that of *AGL66* in wild type, possibly having a small detrimental effect on pollen performance.

4.6 MIKC* genes from different vascular lineages are able to perform the same function in Arabidopsis

After it was established that it is possible to rescue the germination phenotype of the Arabidopsis double mutant, the ability of homologous MIKC* genes from different plants to perform the same function was tested. In the discussion below it will be shown that often multiple mechanisms are possible through which heterologous expression of MIKC* genes can complement the mutant phenotype. For the sake of simplicity it will be assumed that the interactions observed in yeast are valid in Arabidopsis and that the Arabidopsis proteins follow the rules of dimerization that have been described for them by Verelst (2007a): Arabidopsis proteins can only form heterodimers if the monomers of each dimer are from the 2 phylogenetic different subclades S and P. Moreover, the postulation of interactions of Arabidopsis proteins that do not obey these rules are not needed to explain the complementations that were observed.

4.6.1 Can Oryza MIKC* proteins rescue the Arabidopsis double mutant phenotype through 2 distinct types of interactions?

MIKC* genes from the monocot *Oryza sativa* were able to partly rescue the pollen germination defect. Since all Oryza proteins displayed the capacities to form both homo- and heterodimers (in yeast) it is uncertain through what kind of MIKC* dimer complementation occurs. The Arabidopsis P-clade proteins that are still available in the Arabidopsis S-clade mutant might be able to interact with an S-clade protein of Oryza leading to dimers consisting of MIKC* proteins of both species that could regulate the proper target genes for a partial complementation. On the other hand, it

is very likely that the heterologous Oryza MIKC* proteins form homodimers in Arabidopsis when expressed there. If so, they might be able to control partly the same genes as Arabidopsis MIKC* heterodimeric complexes do in wild type pollen. That the latter scenario occurs is supported by the observation that also the Oryza P-clade protein could rescue the mutant phenotype. In the end, the truth might be that complementation is accomplished through both kind of mechanisms described above. That also would clarify why the Oryza P-clade member is less successful in complementation than the Oryza S-clade member: complementation by the S- clade member shows the additive effect of its capacities to form heterodimers with Arabidopsis proteins and to form homodimers that autonomously can regulate some of the target genes that rescue the mutant phenotype.

4.6.2 Lessons from a non-complementing Eschscholzia MIKC* gene

A partially different complementation behaviour is shown by the MIKC* genes of Eschscholzia. EcMADS2 acted similar to Oryza S-clade proteins, both in the yeast-2-hybrid study and in the Arabidopsis complementation; therefore a detailed description will be omitted here. The more interesting was the observation that EcMADS1, the P-clade protein of Eschscholzia was not able to complement the Arabidopsis phenotype when expressed in pollen. Fluorescence imaging did show that the transcription factor was transported to the nucleus. For petunia MADS-domain proteins it has been demonstrated that they have to be imported into the nucleus as dimers (Immink et al. 2002). Yeast-2-hybrid studies showed that EcMADS1 could not form homodimers. What can then clarify the apparent nuclear targeting of EcMADS1 protein in the transformants? Possibly, EcMADS1 forms dimers with the residual AGL104 S-clade protein still produced by the knockdown *agl104-1* allele. However, to accomplish this they have to outcompete the native Arabidopsis MIKC* P-clade proteins which also have no interaction partner and that can be assumed to have a higher affinity for AGL104. Also, it is not expected that P-clade EcMADS1 is imported as a dimer with an Arabidopsis P-clade protein for reasons described at the start of this section. An alternative explanation might lie in the nature of the *agl66/104-1* T-DNA insertions in the double mutant. Although the

presence of AGL66 full-length transcript could not be detected (Verelst et al. 2007a) Adamczyk and Fernandez showed that the *agl66* T-DNA inserted locus is transcribed at least until the site where the T-DNA is inserted. Since it is inserted in the ninth exon, a truncated AGL66 protein with a complete MADS-, I- and K-domain might still be expressed. From deletion experiments with Arabidopsis MIKC^c genes (Huang et al. 1996) it was shown that the MADS- and I-domain can be sufficient for dimerization. Moreover, in the yeast-2-hybrid experiments performed for the Arabidopsis MIKC*genes (Verelst et al. 2007a) a truncated bait protein was used which has a larger C-terminal deletion than the one putatively caused by the T-DNA insertion in the *agl66* mutant locus. It is thus very plausible that the EcMADS1 protein is imported in the nucleus as a dimer with a truncated AGL66 partner. The existence of cross-species hybrid MIKC* complexes proposed to explain the complementation in Arabidopsis by Oryza proteins and EcMADS2, is thus supported by the nuclear targeting of EcMADS1.

4.6.3 Ceratopteris MIKC* proteins are too divergent to substitute for Arabidopsis MIKC* proteins

Two pairs of orthologous S- and P-clade MIKC* proteins were identified in the fern *Ceratopteris richardii*. It was expected that they could enhance the pollen germination rate of the mutant similar to the MIKC* genes from the other species discussed until now. However, the GFP-fusion lines often showed only weak fluorescent signals. Furthermore, the little amount of GFP fusion protein that was present was not transported to the vegetative nucleus of pollen. Perhaps the fusion with the GFP caused a hindrance, as observed for SmMADS1 of Selaginella, which will be discussed later. However, transformation of non-GFP constructs also showed no complementation. Whether the non-GFP-fusions are transported to the nucleus, remains unknown. If so, they obviously are incapable of regulating the appropriate Arabidopsis MIKC* target genes needed for complementation. Already in the yeast-2-hybrid experiments it was shown that Ceratopteris MIKC* proteins must have some different biochemical properties than the tested S- and P-clade MIKC* genes from other plants. Namely, all the Ceratopteris MIKC* genes did not show

autoactivation of the yeast-2-hybrid system used here while the angiosperm S- and P-clade proteins did. Autoactivation is accomplished by a direct interaction of the protein that is fused to the GAL4 DNA binding domain with the transcription machinery of the yeast host without the need for dimerization with the GAL4 activation domain fusion. Maybe Ceratopteris MIKC* proteins evolved different biochemical properties as an adaptation to a different interactome. If so, it still is possible that the MIKC* genes of Ceratopteris and angiosperms do fulfil the same function in their respective biological context. On the other hand, the gametophyte of Ceratopteris is definitely much more complex than that of angiosperms. Perhaps that this also is reflected in the functional divergence between Ceratopteris and angiosperm MIKC* genes.

4.6.4 Selaginella MIKC* genes have conserved properties that allow them to function in Arabidopsis

The ancestor of *Selaginella moellendorffii* diverged from all other land plants more than 400 million years ago. While the ancestor of all other vascular land plants evolved S- and P-clade MIKC* genes, the MIKC* genes in the lycophyte lineage diverged such that phylogenetic analysis using the genes that were identified here, could not resolve the relationship to the other MIKC* genes. Surprisingly, the large separation in time did not have its effect on the potential to complement the Arabidopsis mutant. In 2 independent lines expressing SmMADS3-GFP protein fusions the pollen germination defect was rescued, albeit to very different extent *SmMADS3-GFP* line 1 showed a pollen germination efficiency of a mere 10% while pollen from *SmMADS3-GFP* line 2 germinated even more efficiently than the wild type. Although this difference seems enigmatic it is more understandable when also the expression levels of the proteins are taken into account. Fluorescence signals indicated that there is more SmMADS3-GFP fusion protein in the pollen of the line that showed the biggest degree of complementation. It must be concluded then that the protein concentration can also have an influence on complementation. The possible importance of dosage effect in the functioning of MADS-domain proteins has been well described in the literature. For example, Arabidopsis plants

homozygous for the *sep2* and *sep3* mutations and heterozygous for the *sep1* mutation do not develop ovules. The phenotype is not present when 2 functional alleles of SEP1 are present in the same background (Favaro et al. 2003). It cannot be neglected that the germination efficiency of the pollen of one of the Arabidopsis plants transformed with the SmMADS3-GFP fusion was higher than that of wild type pollen. Perhaps the rescue of the pollen germination could be the effect of the expression of a more or less random set of genes or be the result of a stress reaction caused by the high expression of the transgene, which maybe influences the ability to form a pollen tube positively. In the next section evidence will be provided that this is not the case and that the complementation by MIKC* genes is indeed specific. Another explanation must then be found for the "over-complementation" that was observed. A closer look to the mutant phenotype is therefore helpful. Importantly, the kind of rescuing described here concerns a discrete value: germination or no germination. The quality of the pollen tube however, could not be distinguished from wild type. Although the expression of over a thousand genes is disturbed in the double mutant used here (Verelst et al. 2007b) it seems to affect the ability to initiate a pollen tube more than the ability to actually produce it. This becomes apparent because the *in vitro* pollen germination defect cannot only be rescued through complementation by MIKC* genes, complementation also occurs if the culture medium is contaminated with sporophytic material. For example, if an anther or a style broke of during the application of the pollen, the mutant pollen surrounding that tissue will show an increased germination rate. Somehow the mutations in the *agl66/104-1* mutant seem to cause an altered sensitivity to abiotic signals. Perhaps an "over rescue" of the Arabidopsis target gene expression levels by the phylogenetically remote SmMADS3-GFP protein can cause an over-sensitivity for the reception of positive cues from the environment.

In addition to the SmMADS3 protein also SmMADS1 showed complementation, in contrast to the SmMADS1 fused to GFP. Evidently, the GFP fusion can in some cases perturb the functioning of the complete fusion protein. The yeast-2-hybrid data indicated that SmMADS1 can only form heterodimers. If this also is the case in

Arabidopsis it must interact with one of the Arabidopsis P-clade proteins. This suggests that it may have evolved from an ancient S-clade like protein.

Since the distantly related MIKC* genes of Selaginella were able to complement the Arabidopsis double mutant, it was also tested if the even more remote MIKC* genes *PPM3* and *PPM4* of the moss Physcomitrella could also perform the same function as Arabidopsis MIKC* proteins. However, no complementation was found in these cases. Off course, it cannot be excluded that perhaps some of the other Physcomitrella MIKC* genes would be more successful.

As mentioned in the introduction, the understanding of classical MIKC^c genes is much more advanced than those of MIKC* genes and their functional evolution has been discussed in detail. Potential orthologs of classical MIKC^c genes were identified in gymnosperms but not in more ancient land plant lineages (Theissen, 2000). It has also been shown that an Arabidopsis floral homeotic MADS mutant could be partly complemented by an orthologous gene from a gymnosperm (Winter et al 2002). In this manuscript an orthology relationship between the MIKC* genes of the fern *Ceratopteris richardii* and angiosperm MIKC* genes has been described, which is unprecedented for the MADS superfamily of transcription factors in land plants. A functional equivalence between these orthologs could not be shown, however. Surprisingly, MIKC* proteins from the lycophyte *Selaginella moellendorffii* could complement the MIKC* Arabidopsis mutant. This is the first time that an Arabidopsis MADS-domain transcription factor mutant has been complemented by a homologous gene from a species so distantly related.

4.7 What makes an MIKC* gene?

MADS-domain proteins have been described to bind to DNA sequences with the consensus $CC(A/T)_6GG$ (Hayes et al. 1988) or $C(A/T)_8G$ (Pollock and Treisman 1991); the so-called C A/T-rich G (CArG) boxes. Through random binding site selection and competitive EMSA experiments it has been shown that the Arabidopsis MIKC* dimers prefer the more restricted $CTA(A/T)_4TAG$ motifs, which are also enriched in the proximal regions of late pollen specific promoters (Verelst et al. 2007a). In the previous sections a lot of attention has been given to the requirement of clade-specific dimerization of MIKC* proteins for the complementation of the

Arabidopsis double mutant. Unexpectedly, atypical MIKC* protein dimers from the remotely related lycophyte *Selaginella moellendorffii* were also able to rescue the mutant. Therefore it remains the question whether rescue is simply accomplished only because the complementing MIKC* genes bind to the same *cis* regulatory elements as the native Arabidopsis proteins do, or to put it in a different way: are all transcription factors that bind to these motifs able to rescue the double mutant phenotype?

To answer that question Arabidopsis plants were transformed with the human MEF2A transcription factor which has a general preference for the $CTA(A/T)_4TAG$ motif (Pollock and Treisman 1991, Shore and Sharrocks, 1995). Also, *AGL15* and *AGAMOUS*, which are classical MIKC^c genes that are normally not expressed in Arabidopsis pollen, were expressed under control of the *AGL65* promoter in the mutant background. Both AGL15 and AGAMOUS have an affinity for the MEF2A motif but AGAMOUS also can bind to the less specific $C(A/T)_8G$ consensus sequence (Tang and Perry, 2003). If it was observed that one of these non-MIKC* transcription factors could rescue the mutant, the conclusion would be that the functioning of MIKC* MADS-domain proteins is simply a case of DNA binding specificity. Although GFP signals could be detected in the vegetative nuclei of pollen from plants transformed with the *AGL15*, *AGAMOUS* and *MEF2A-GFP* fusion constructs; pollen germination assays revealed that the mutant phenotype was not rescued. The hypothesis that only the preference for the right DNA motif is sufficient for complementation of the *agl66/104-1* double mutant can therefore be rejected. Furthermore, this observation made it also unlikely that rescue takes place through a misregulation of random genes, as hypothesized before, to explain the increased pollen germination of Arabidopsis plants transformed with the *SmMADS3-GFP* construct. In conclusion, not only the capacity to bind a specific CArG motif is the determinant that makes an MIKC* gene. Perhaps the true nature of MIKC* proteins lies in their *trans* acting elements. The most uncharacterized part of MIKC-like MADS-domain proteins is the C-terminus (reviewed in Kaufmann 2005). Using the yeast-2-hybrid system it was shown that the C-terminal domain of the Oryza OsMADS62 MIKC* protein is needed for the formation of homodimers. A function that has previously also been described for a MADS-domain protein from *Lilium*

longiflorum (Tzeng et al. 2004). However, the C-terminal domain is mostly known for its role in transcriptional activation (Honma and Goto 2001). The C-terminal domain of MIKC* proteins is generally much longer than that of classical MIKC^c proteins, clearly visible in the alignment in Supplemental figure S1. Perhaps the function that distinguishes MIKC* from MIKC^c proteins is hidden there.

4.8   On the evolution of the function of MIKC* genes

Because the body plans of gametophytes have changed a lot during evolution it can be expected that this is reflected in the functional diversification of the transcription factors that control gametophytic development. This implies that if MIKC* proteins are such developmental regulators, it is required that their function has changed during evolution. The experiments described in this manuscript have shown that MIKC* genes from distantly related species can perform the same function as Arabidopsis MIKC* genes. However, this does not necessarily mean that the function of MIKC* genes has been conserved during evolution. The target genes that are regulated by Selaginella MIKC* genes in Selaginella do not have to be the same as the target genes that Selaginella MIKC* genes can regulate in Arabidopsis. However, there is circumstantial evidence that these target genes are the same and that as a consequence, MIKC* genes have a conserved function, leading to the conclusion that MIKC* genes have not been important regulators of developmental programs during evolution.

Firstly, all the MIKC* genes that could complement the pollen germination defect of Arabidopsis *agl66/104-1* double mutants were isolated from cDNA collections from gametophytic tissues, which supports that a conserved *cis*- and *trans*-regulatory machinery is the cause of their expression there. Interestingly, expression of the Selaginella MIKC* gene *SmMADS3* was reported also in roots and apical sporophytic tissues. Possibly the gene was adopted from the gametophytic generation to function there. Indications for the expression of MIKC* genes in roots have also been found in *Pinus taeda*, *Beta vulgaris* and the Arabidopsis embryo. It is generally accepted that the roots of lycophytes evolved independently from those of ferns and

angiosperms (reviewed in Dolan, 2009). To explain the presence of ectopic expression of MIKC* genes in roots of the lycophyte Selaginella and in seed plants, the observed expression in roots has to have evolved 2 times independently. There is only a small chance that a similarly functioning organ (the root) recruits a homologous transcription factor from the gametophytic phase if it has not got the same function. The chance *is* small because if the function in the gametophyte would be different, it means that both of the presumed different processes regulated by MIKC* genes in the gametophyte of Selaginella and for example, *Pinus taeda*, should have evolved in such a way that they would be beneficial for the functioning of the root.

As a second point, it cannot be overlooked that the presence of dimers between MIKC* proteins of different species, for example, Eschscholzia and Arabidopsis, must have occurred during the complementation experiments, else, it cannot so easily be understood how complementation was accomplished. Somehow, the diversification of dimerization properties must have been constrained. The most plausible explanation for such a constraint is that the function of MIKC* proteins does not allow them to diversify. Their function is thus most likely conserved.

Through the probabilistic argumentation given above, the conclusion must be drawn that MIKC* genes have a conserved function in gametophytes. The complementation of the Arabidopsis MIKC* mutant by distantly related MIKC* genes can be seen as a necessary consequence of this conservation.

For which function have MIKC* genes then been conserved? One clue about the function of MIKC* genes has been provided in the first section of the discussion, where it was mentioned that MIKC* genes seem to have come into existence together with land plants. Another clue came to light by their conserved expression in gametophytes. However, expression in roots was also observed. Can all these observations be united in a hypothesis for the function of MIKC* genes?

When plants invaded land they were probably subject to a stronger fluctuation of nutrients (and water) than their ancestor that lived in a freshwater environment. Perhaps that early land plants, which had a dominant gametophytic phase, would

have benefited from a quick response to take up necessary compounds if they are only temporarily available. The evolution of a transcriptional network to mediate this response could be beneficial. Such a function also fits to the life style of the very derived male gametophytes of angiosperms (pollen). Since they have no chloroplasts, they are completely dependent upon the sporophytic generation for their nutrition, both during development and after they land on the stigma and form a pollen tube. A rapid response when getting in touch with the stigma might aid in a quicker formation of the pollen tube and a higher chance of reproductive success.

The roots of vascular plants evolved much later than gametophytes. The evolution of roots may have been facilitated by the presence of pre-existing functional modules from other tissues that could be recruited to function in the root. MIKC* genes might control such a functional module because it is their basal function. In the end, the expression of MIKC* genes in roots, next to gametophytes, might have uncovered their true nature.

5   Conclusion

The results presented here have given a rewarding insight in the evolution of MIKC*
transcription factors. It has been shown that MIKC* genes probably evolved from
MIKC^c genes through a duplication event in the K-domain. Furthermore, through the
reconstruction of the evolutionary history using a phylogenetic approach, events
could be pinpointed that might reflect important functional changes. An example of
such an event is the coming into existence of 2 divergent MIKC* subclades in the
early evolution of vascular land plants. Additional experiments were performed in
which it was determined that MIKC* genes from distantly related species, for
example, *Selaginella moellendorffii,* can perform the same function as Arabidopsis
MIKC* genes. Based on the conserved expression in gametophytic tissue and
expression in roots the hypothesis was drawn up that MIKC* genes have a basal role
which is associated with a root-related function. Unfortunately, no more information
about the specific function can be expected to come from a study using a
heterologous system. The time has come to test the proposed hypothesis in a
homologous system.

6 Literature cited

Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402 (1997).

Arias, T. & Williams, J. H. Embryology of Manekia naranjoana (Piperaceae) and the origin of tetrasporic, 16-nucleate female gametophytes in Piperales. *American Journal of Botany* **95**, 272-285 (2008).

Arora, R. et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* **8**, - (2007).

Boavida, L. C. & McCormick, S. Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant Journal* **52**, 570-582 (2007).

Borges, F. et al. Comparative transcriptomics of *Arabidopsis thaliana* sperm cells. *Plant Physiology* **148**, 1168-1181 (2008).

Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* **16**, 735-743 (1998).

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology* **139**, 5-17 (2005).

Davies, B., EgeaCortines, M., Silva, E. D., Saedler, H. & Sommer, H. Multiple interactions amongst floral homeotic MADS box proteins. *Embo Journal* **15**, 4330-4343 (1996).

Dolan, L. Body building on land - morphological evolution of land plants. *Current Opinion in Plant Biology* **12**, 4-8 (2009).

Duttweiler, H. M. A highly sensitive and non-lethal beta-galactosidase plate assay for yeast. *Trends in Genetics* **12**, 340-341 (1996).

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

Favaro, R. et al. MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **15**, 2603-2611 (2003).

Felsenstein, J. Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166 (1989).

Graham, L. E. Green algae to land plants: An evolutionary transition. *Journal of Plant Research* **109**, 241-251 (1996).

Hayes, T. E., Sengupta, P. & Cochran, B. H. The Human C-Fos Serum Response Factor and the Yeast Factors Grm/Prtf Have Related DNA-Binding Specificities. *Genes & Development* **2**, 1713-1722 (1988).

Henschel, K. et al. Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Molecular Biology and Evolution* **19**, 801-814 (2002).

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

Honys, D. & Twell, D. Comparative analysis of the Arabidopsis pollen transcriptome. *Plant Physiology* **132**, 640-652 (2003).

Huang, H. et al. DNA binding properties of two *Arabidopsis* MADS domain proteins: Binding consensus and dimer formation. *Plant Cell* **8**, 81-94 (1996).

Immink, R. G. H., Gadella, T. W. J., Ferrario, S., Busscher, M. & Angenent, G. C. Analysis of MADS box protein-protein interactions in living plant cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2416-2421 (2002).

Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389**, 33-39 (1997).

Kofuji, R. et al. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Molecular Biology and Evolution* **20**, 1963-1977 (2003).

Koncz, C., Kreuzaler, F., Kalman, Z. & Schell, J. A Simple Method to Transfer, Integrate and Study Expression of Foreign Genes, Such as Chicken Ovalbumin and Alpha-Actin in Plant Tumors. *Embo Journal* **3**, 1029-1037 (1984).

Koncz, C. & Schell, J. The Promoter of Tl-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of Agrobacterium Binary Vector. *Molecular & General Genetics* **204**, 383-396 (1986).

Little, D. P., Moran, R. C., Brenner, E. D. & Stevenson, D. W. Nuclear genome size in Selaginella. *Genome* **50**, 351-356 (2007).

Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. AGL1-AGL6, an Arabidopsis Gene Family with Similarity to Floral Homeotic and Transcription Factor Genes. *Genes & Development* **5**, 484-495 (1991).

Melzer, R., Kaufmann, K. & Theissen, G. Missing links: DNA-binding and target gene specificity of floral homeotic proteins. *Advances in Botanical Research: Incorporating Advances in Plant Pathology, Vol 44* **44**, 209-236 (2006).

Melzer, R. and Theissen, G. Reconstitution of "floral quartets" *in vitro* involving class B and class E floral homeotic proteins. *Nucleic Acid Research*, Vol 37 **8**, 2723-2736.

Munster, T. et al. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 2415-2420 (1997).

Nam, J. et al. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1910-1915 (2004).

Ng, M. & Yanofsky, M. F. Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics* **2**, 186-195 (2001).

Parenicova, L. et al. 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 (2003).

Pina, C., Pinto, F., Feijo, J. A. & Becker, J. D. Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology* **138**, 744-756 (2005).

Pollock, R. & Treisman, R. Human Srf-Related Proteins - DNA-Binding Properties and Potential Regulatory Targets. *Genes & Development* **5**, 2327-2341 (1991).

Pryer, K. M. et al. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**, 618-622 (2001).

Qiu, Y. L. et al. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15511-15516 (2006).

Qiu, Y. L. & Palmer, J. D. Phylogeny of early land plants: insights from genes and genomes. *Trends in Plant Science* **4**, 26-30 (1999).

Raven, P.H., Evert, R.F., Eichhorn, S.E. Biology of plants. 5th edition. Worth Publishers 1992.

Rensing, S. A. et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64-69 (2008).

Riechmann, J. L. & Meyerowitz, E. M. MADS domain proteins in plant development. *Biological Chemistry* **378**, 1079-1101 (1997).

Riechmann, J. L., Wang, M. Q. & Meyerowitz, E. M. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Research* **24**, 3134-3141 (1996).

Riese, M. et al. Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biology* **7**, 307-314 (2005).

Shiu, S. H., Shih, M. C. & Li, W. H. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiology* **139**, 18-26 (2005).

Shore, P. & Sharrocks, A. D. The Mads-Box Family of Transcription Factors. *European Journal of Biochemistry* **229**, 1-13 (1995).

Schwarz-Sommer, Z. et al. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**, 931-936 (1990).

Sommer, H. et al. Deficiens, a homeotic gene involved in the control of flower development in *Anthirrhinum majus*: the protein shows homolgy to transcription factors. *EMBO journal* **9**, 605-613 (1990)

Tanabe, Y. et al. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2436-2441 (2005).

Tang, W. N. & Perry, S. E. Binding site selection for the MADS-domain protein AGL15: an in vitro and in vivo study. *Faseb Journal* **17**, A185-A185 (2003).

Theissen, G. et al. A short history of MADS-box genes in plants. Plant Molecular Biology **42**, 115-149 (2000)

Theissen, G. & Saedler, H. Plant biology - Floral quartets. *Nature* **409**, 469-471 (2001).

Tichopad, A., Dilger, M., Schwarz, G. & Pfaffl, M. W. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Research* **31**, - (2003).

Tonaco, I. A. N., Borst, J. W., de Vries, S. C., Angenent, G. C. & Immink, R. G. H. In vivo imaging of MADS-box transcription factor interactions. *Journal of Experimental Botany* **57**, 33-42 (2006).

Tzeng, T. Y., Liu, H. C. & Yang, C. H. The C-terminal sequence of LMADS1 is essential for the formation of homodimers for B function proteins. *Journal of Biological Chemistry* **279**, 10747-10755 (2004).

Verelst, W., Saedler, H. & Munster, T. MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiology* **143**, 447-460 (2007a).

Verelst, W. et al. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology* **8**, - (2007b).

Wang, W. M. et al. Construction of a bacterial artificial chromosome library from the spikemoss *Selaginella moellendorffii*: a new resource for plant comparative genomics. *BMC Plant Biology* **5**, - (2005).

Weng, J. K., Tanurdzic, M. & Chapple, C. Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. BMC *Genomics* **6**, - (2005).

Winter, K. U., Saedler, H. & Theissen, G. 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 Journal* **31**, 457-475 (2002).

Supplement

Table S1. List of primers

	location	Sequence 5' ->3'	name
Oryza sativa	Oryza cDNA synthesis	CCG GAT CCT CTA GAG CGG CC GC (T)$_{17}$	
	Oryza anchor primer 1	CGG ATC CTC TAG AGC GGC	pr44
	Oryza anchor primer 2	CTA GAG CGG CCG CTT TTT TTT TTT TTT	pr46
	OsMADS68 5'UTR for cDNA	ATT GGA CTG CAG ATT CGT CGC C	pr17
	OsMADS68 3'UTR for cDNA	TCA TGG CAA ATT GTA ACA TTC TTT ACA GGA GAG	pr55
	OsMADS68 5'NcoI, restriction with BsaI yields an NcoI site	TAT AGG TCT CCC ATG GGG AGG GTC AAG CTC AAG	pr94
	OsMADS68 3'NcoI, restriction with BsaI yields an NcoI site	AAA GGT CTC CCA TGA TCA TGA GCT GCC GGT GTG G	pr220
	OsMADS68 3'BamHI, restriction with BsaI yields a BamHI site	TAT AGG TCT CCG ATC CTC AAA TCA TGA GCT GCC GGT GTG	pr169
	OsMADS68 Δ 262-383 3'BamHI, restriction with BsmBI yields a 3'BamHI site	AAA CGT CTC GGA TCC TCA CTC GCC GCC GCC GGC GTT G	pr313
	OsMADS68 Δ 299-383 3'BamHI, restriction with BsmBI yields a BamHI site	AAA CGT CTC GGA TCC TCA GGC GTT GTC GAA CGC CGA C	pr314
	OsMADS62 5'UTR for cDNA, used with anchor primer 2	CGA TCG AG C TAT CTC GCT CTG	pr12
	OsMADS62 5'UTR nested primer for cDNA	GAG ATT TTT GAC GTC GTG AAT ATA TCG G	pr22
	OsMADS62 3'UTR nested primer for cDNA	GTT TGT AGG TTA GTT AGG TGA GGT CG	pr24
	OsMADS62 5' NcoI	TAT ACC ATG GGG AGG GTG AAG CTG C	pr81
	OsMADS62 3' NcoI	AAA CCA TGG CGA TGT TCG CCG GCG	pr166
	OsMADS62 3' BamHI, restriction with BsmBI yields a BamHI site	AAA CGT CTC GGA TCC TCA GGC GAT GTT CGC CG	pr176
	OsMADS68 Δ 229-339 3'BamHI, restriction with BsmBI yields a BamHI site	AAA CGT CTC GGA TCC TCA ACT GAA CAT CGG ATG GCC AG	pr315
	OsMADS68 Δ 266-339 3'BamHI, restriction with BsmBI yields a BamHI site	AAA CGT CTC GGA TCC TCA GCA GGT GTA CGC CTG CCT C	pr316
	OsMADS63 5'UTR for cDNA, used with anchor primer 1	CGG ACG ACG GCC GGA ACA TAG	pr13
	OsMADS63 5' NcoI, restriction with BsmBI yields an NcoI site	TTT CGT CTC CCA TGG GAC GGG TGA AG	pr168
	OsMADS63 3' NcoI, restriction with BsmBI yields an NcoI site	CGT CTC ACA TGC CAA CGT TAA CCG GAG CAA TG	pr167
	OsMADS63 3'BamHI	TT GGA TCC TTA ACC AAC GTT AAC CGG AGC AAT G	pr175
	OsMADS63 Δ 214-360 3' BamHI	TTT GGA TCC TTA GCT TCC AAA CCC ATT ATC AGT GAT GTC	pr 317
	OsMADS63 Δ 294-360 3' BamHI	TTT GGA TCC TTA CGC GTT GAG GAA CTC CGC	pr318
Selaginella moellendorffii	Selaginella cDNA synthesis primer from Invitrogen	GAC GAC GCG TAT CGA TGT CGA C (T)$_{16}$V	pr08

Selginella anchor primer	ACC ACG CGT ATC GAT GTC	pr09
SmMADS1 allele 1 and allele 2 5'UTR for cDNA, allele 2 was amplified with the anchor primer	GGT GAT AAG GAA TGA AGA ATA GGA GTG TC	pr37
SmMADS1 allele 1 3'UTR for cDNA	AAC AAG AAA TAA GAA CTT CAA TTT ATC ACA AAC TTG C	pr135b
SmMADS1 5' NcoI, restricton with BsmBI yields an NcoI site	TTT CGT CTC CCA TGG GGA GGG TTA AGC TTG AGA TT	pr153
SmMADS1 3' NcoI,	AAA CGT CTC CCA TGG ATT CAT CGC TAT GAA ACA CAT TGT CCA AAT C	pr186
SmMADS1 3' BamHI	AAA GGA TCC GTC AAG ATT CAT CGC TAT GAA ACA CAT TGT C	pr154
SmMADS1 qPCR FW	CAA GTT TTA TCA GGG TGA TAC TCA AAT TCT AG	pr413
SmMADS1 qPCR RV	GGG AGG AAA TGT TCA GTC TGT TTT TG	pr414
SmMADS2 5' UTR for cDNA was used in combination with selaginella anchor primer	GCC TTT TAA TGC TCT TTC GAA AGC TTT C	pr38
SmMADS2 5' NcoI	TTT CCA TGG GGA GAG TGA AAT TAG AGA TAA AGA GAA TAG AG	pr155
SmMADS2 3' NcoI	TTT CCA TGG TGG GAG TAT AAT CGC TAG AAT CAA AGA GAT G	pr200
SmMADS2 3' BamHI	AAA GGA TCC ATC ACG TGG GAG TAT AAT CGC TAG AAT CAA AG	pr156
SmMADS2 qPCR FW	GGT TTT ACA GAA GGA GCT AGC AAG	pr415
SmMADS2 qPCR RV	CAA GGA TCT TGA GCA AGC TAG C	pr372
SmMADS3 allele 1 5'UTR for cDNA	GGC ACG ACA TTA ATG TCA AAT GGT TG	pr39
SmMADS3 allele 1 3'UTR for cDNA	CTT TTC TTT TGG CGT CTT TTT CTT CTT C	pr136
SmMADS3 allele 2 5' UTR used with the anchor primer	CGG TTG CCA GAG AAC AGA AAC	pr 90
SmMADS3 5' NcoI, restriction with BsmBI yields an NcoI site	TTT CGT CTC CCA TGG GTC GGG TGA AGC TGG	pr157
SmMADS3 3' NcoI restriction with BsmBI yields an NcoI site	AAA CGT CTC CCA TGT TTT CAT AGC TTG TTG ACA GTG TGA GC	pr188
SmMADS3 3' BamHI	AAA GGA TCC GTC AGT TTT CAT AGC TTG TTG ACA GTG TGA G	pr158
SmMADS3 qPCR FW	GGT CAA CTG CAA CAG ATG TAT CC	pr416
SmMADS3 qPCR RV	GGC ATC TCC ATT GTC CCT TC	pr374
Sm6PGD qPCR FW	GCT CAT GGA TCC CGA GTT TG	pr408
Sm6PGD qPCR RV Amplicon 210 bp	TCC ACT CGC TCG TAG GTG	pr364
SmCLATH qPCR FW	GTC CCA ATG TTC ACT GCC TC	pr367
SmCLATH qPCR RV Amplicon 58 bp	TCT TTT CCC ACA CCT TGA GAA AC	pr428
Sm1 qPCR FW	CAA GTT TTA TCA GGG TGA TAC TCA AAT TCT AG	pr413
Sm1 qPCR RV Amplicon 162 bp	GGG AGG AAA TGT TCA GTC TGT TTT TG	pr414
Sm2 qPCR FW	GGT TTT ACA GAA GGA GCT AGC AAG	pr415
Sm2 qPCR RV Amplicon 176 bp	CAA GGA TCT TGA GCA AGC TAG C	pr372
Sm3 qPCR FW	GGT CAA CTG CAA CAG ATG TAT CC	pr416
Sm3 qPCR RV Amplicon 141	GGC ATC TCC ATT GTC CCT TC	pr374
SmPDF1 FW	ACT GAG ACT TTC GGC GGT AG	pr481

54

SmPDF1 RV Amplicon 119 bp	CAC GCC AGG ACT TAT CCT GAG	pr483
All other primers are available upon request		

Figure S1. Alignment of MIKC^c and MIKC* genes. Exon information, when available, is indicated by alternating bold/non-bold font. Parts of the alignments that were used for the phylogenetic tree in figure 7 are indicate by (*) signs above the alignment. Heptad repeat regions in MIKC^c proteins are according to Kaufmann et al. 2005 and shown above the alignment and for MIKC* proteins below the alignment. Heptad repeat hydrophobic amino acids (V, I, L, M) in "A" and "D" positions are in blue. The 42 aminoacid N-terminal extension of the MADS-domain of SrMADS1 was removed.

```
                        |----------------------MADS-domain-------------------------||------
                        ************************************************************    *****
DEFA             1 MARGKIQIKRIENQTNRQVTYSKRRNGLFKKAHELSVLCDAKVSIIMISSTQKLHEYISPTTA--TKQLF  68
GLO              1 MGRGKIEIKRIENSSNRQVTYSKRRNGIMKKAKEISVLCDAHVSVIIFASSGKMHEFCSPSTT--LVDML  68
StMADS11         1 MVRQKIQIKKIDNLTARQVTFSKRRRGLFKKAQELSTLCDADIGLIVFSATGKLFEYSSSS----MMQLI  66
StMADS16         1 MAREKIKIKKIDNITARQVTFSKRRRGLFKKAEELSVLCDADVALIIFSATGKLFDFASTS----MKDIL  66
GGM2             1 MGRGKIEMKKIENTNNRQVTFSKRRNGLMKKAQELAVLCDAEVGLIIFSSTGKLFQYCNTS----MSQVL  66
GGM3             1 MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSSRGRLYEFANNS----VKRTI  66
CRM1             1 MVRRKIKIKRIENATTRQVTFSKRRGGLLKKAHDLSVLCDAEVAVIIFSSKGKLFQFANPS----METVL  66
CRM3             1 MGRGKIEMKRIENRTTRQVTFCKRRAGLVKKAHELSVLCDADVALIVFSSSGRLFEYAGSRS---MREII  67
SrMADS1          1 VGRGKIEIKRIENATSRQVTFSKRRGGLLKKAHELSVLCDAQVALIIFSSTGKLFEYASTS----MKEIL  66
LAMB4            1 MGRGKIEIKRIENATCRQVTFSKRRSGLLKKAHELSVLCDAQVAVIIFSSTGKLFQFASTR----MKEIL  66
PPM1             1 MGRGKIEIKKIENTTSRQVTFSKRRGGLLKKAHELAVLCDAEVALVIFSSTGKLFEYASSGS---MRDII  67
Csmads1          1 MGRGKIEINRIENATSRQVTFSKRRNGLLKKAYELSVLCDVDIAVIVFSPTGKLFQYASSS----MKEIL  66
CgMADS1          1 MGRAKIEIKRIDNATSRQVTFSKRRNGLLKKAYELSVLCDADIAVIMFSPTGKLFEYANSS----MKEIL  66
PPM3             1 MGRVKLEIKKIENSANRQVTYSKRRNGLTKKAYELSVLCDIDLALIMFSPSGKLTQYSNCS----IEDII  66
PPM4             1 MGRVKLEIKKIENPTNRQVTYSKRRNGLVKKAYELSVLCDIDLALIMFSPSGKLTQYSNCS----IEDII  66
PPM7             1 MGRVKLEIKKIENPTNRQVTYSKRRNGLIKKAYELSVLCDIDLALIMFSPSGKLTQYSNCS----IEDVI  66
PpMADS2          1 MGRVKLEIKKIENPTNRQVTYSKRRNGLIKKAYELSVLCDIDLALIMFSPSGKLTQYCNCS----IEEVI  66
PpMADS3          1 MGRVKLEIKKIENPTNRQVTFSKRRNGLIKKAYELSVLCDIDLALIMFSPSGKLTQYCNCS----IEEVI  66
CRM13            1 MGRVKLAIKTIETAVGRQVTYSKRRAGLEKKAQELATLCDIDLLLILFSPAGKPSWINGSKS--PPEEVL  68
CRM16            1 MGRVKLAIKKIETPVGKQVTYSKRKAGLEKKAKELATLCDIDLLLILFSPAGKLSWINGSKS--PPEEVL  68
P. taeda 23021   1 MGRMKLAIRRLEHNNNRQITFSKRRNGIIKKAKELAVLCDIEILLIMFSPSGKPTVCVGEKS--VMEDVI  68
Picea 67868      1 MGRMKLAIRRLEHNNNRQITFSKRRNGIIKKAKEIAILCDIDILLMMFSPSGKPTVCVGEKS--VMEDVI  68
P. equestris     1 MGRAKLQIKKLESNSARNATYTKRRAGILKKAKELSILCDIEVLLLMFSPNGNPTLCLGEQS--NLVQMI  68
Med. tru. 331    1 MGRVKLKIKKLESTSNRHVTYSKRRSGILKKAKELSILCDIDILLLMFSPTGKPTLLQGERS--NMEEII  68
Gos. rai. 862    1 MGRVKLKIKRLESYSNRQVTYSKRRTGILKKAKELSILCDIHIILLMFSPTGKPTLFHGERS--NIEEVI  68
Cen. mac. 1285   1 MGRVKLKIKRLENISNRQVTFSKRRNGILKKAKELSVLCDIDIILLMFSPTGKPTLFTGQRS--NIDEVI  68
AGL65            1 MGRVKLKIKRLESTSNRQVTYTKRKNGILKKAKELSILCDIDIVLLMFSPTGRATAFHGEHS--CIEEVI  68
Cit. sin. 77561  1 MGRVKLKIKRLESTSNRQVTYSKRRNGILKKARELSILCDIDIVLLMFSPTGRSTLFHGQRS--NIEEVI  68
AfMADS1          1 MGRVKLKIKKLESSNNRCVTYSKRKTGIMKKAKELSILCDIDIALLMFSPNGKPSLCLGERSSSNIEDVI  70
OsMADS68         1 MGRVKLKIKKLENSSGRHVTYSKRRSGILKKAKELSILCDIPLILLMFSPNDKPTICVGDHS--SIEDVI  68
EcMADS1          1 MGRVKLKIKRLENTSGRQVTYSKRRAGILKKARELSILCDIDIALLMFSPTGKPTLCLGDRS--NIEDVI  68
A. for. x pub.   1 MGRVKLKIKRLENTSGRQVTYSKRRAGILKKARELSILCDIDIALLMFSPTGKPTLCLGERS--TIEEVI  68
Bet. vul. 10637  1 MGRVKLDIKKLENSSSRQTTYSKRKHGILKKAQELSILCDIDLALVMFSPSGKPSLCCGKHS---IEEVI  67
AGL94            1 MGRVKLKIKKLQNMNGRQCTYTKRRHGIMKKAKELSILCDIDVVLLMFSPMGKASICIGKHS---IGEVI  67
AGL30            1 MGRVKLKIKKLENTNGRQSTFAKRKNGILKKANELSILCDIDIVLLMFSPTGKAAICCGTRS--SMEEVI  68
Gos. rai. 800    1 MGRGKLKIKKLENTNGRQATYAKRKHGIMKKANELSILCDVEIILLMFSPTNKPSVCIGKRS--SIEEII  68
SmMADS1          1 MGRVKLKIKIENATNRQVTYSKRRNGLVKKAYELSILCDIDIALIMFSPTGKLTQYATDMR---VEDVI  67
SmMADS2          1 MGRVKLEIKRIENSVSRHATFAKRKIGLVKKAQELATLCDIDIALIMFSPVDHLIHYPSDLK---IQEII  67
SmMADS3          1 MGRVKLEIKKIENHQARQVTYSKRRNGLMKKAFELSTLCDTDVALIMFSPAGKLSIHPNDGR---IEEII  67
CRM14            1 MGRVKLEIRRIENPTNRQVTYSKRRNGLIKKAYELSVLCDIDIALIMFSPSGRLNHFSGKKR---IEDVI  67
CRM15            1 MGRVKLEIRRIENPTNRQVTYSKRRNGLIKKAYELSVLCDVDIIIFSPSGRLDHFSGRKR---IEDVI  67
Hev. bra. 267    1 MGRVKLQIKRIENTTNRQVTYSKRRNGLIKKAYELSVLCDVDVALIMFSPSGRLSLFSGNKS---IEEIL  67
Cof. can. 17252  1 MGRVKLQIKKIESTTNRQVTFSKRRNGLIKKAYELSVLCDVDVALIMFSPSGRLSVFSGNKS---LEEIM  67
Vit. vin. 32798  1 MGRVKLQIKKIENTTNRQVTFSKRRNGLIKKAYELSVLCDVDVALIMFSPSGRISLFSGNKS---IEEIM  67
Cen. mac. 1671   1  GRNKLPMKRIENNTSRLVTFCKRRNGLIKKAYELSVLCDIDIALIMFSPSGRLNHFSCKRR---TEDVL  66
Gly. max 42899   1 MGRVKLEIRRIENPTNRQVTFSKRRNGLIKKAYELSILCDIDIAVIMFSPSGRLNHFSGRRR---IEDVF  67
Pon. tri. 100    1 MGRVKVQIKRIENATNRQVTFSKRRNGLVKKAYELSVLCDVDVALIMFSPSGRVSLFSGNKS---MEEIL  67
AGL67            1 MGRVKLELKRIEKSTNRQITFSKRKKGLIKKAYELSTLCDIDLALLMFSPSDRLCLFSGQTR---IEDVL  67
AGL66            1 MGRVKLEIKRIENTTNRQVTFSKRRNGLIKKAYELSILCDIDIALLMFSPSDRLSLFSGKTR---IEDVF  67
AGL104           1 MGRVKLEIKRIENTTNRQVTFSKRRNGLIKKAYELSILCDIDIALLMFSPSDRLSLFSGKTR---IEDVF  67
OsMADS62         1 MGRVKLPIKRIENTTNRQVTFSKRRNGLIKKAYELSVLCDIDVALLMFSPSGRLSHFSGRRG---VEDVI  67
OsMADS63         1 MGRVKLQIKRIENIPNRQVTFSKRRNGLIKKAYELSVLCDIDIALLMFSPSGRLSHFSGRRR---IEDVL  67
EcMADS2          1 MGRVKLQIKRIENNTNRQVTFSKRRNGLIKKAYELSILCDIDIALIMFSPSGRLSHFSGKRR---IEDVL  67
AfMADS2          1 MGRVKLQIKRIENTTNRQVTFSKRRNGLIKKAYELCILCDIDIALIMFSPSGRLSHFSGKKR---IEDVL  67
```

```
                     ----I-domain----------------------||-----------------K-domain-------
                     **************
DEFA            69 DQYQKAVGVDLWS---------------------------------------------------- 81
GLO             69 DHYHKLSGKRLWD---------------------------------------------------- 81
StMADS11        67 EKHKMQSERDSMDNPEQ------------------------------------------------ 83
StMADS16        67 GKYKLQSASLEKV--------------------------------------------------- 79
GGM2            67 EKYHKSPGVDHWD--------------------------------------------------- 79
GGM3            67 ERYRKTCADNNQGGA------------------------------------------------- 81
CRM1            67 GRYVKASRDPEAG--------------------------------------------------- 79
CRM3            68 QAYVDAHEDSSSLLQLRS----------------------------------------------- 85
SrMADS1         67 DRYGKYPESVQGGNMA------------------------------------------------ 82
LAMB4           67 ERYERNSDPLQIDA-------------------------------------------------- 80
PPM1            68 ERYKKSPNGAMKS--------------------------------------------------- 80
Csmads1         67 ERYEQVPPEQKEKGS------------------------------------------------- 81
CgMADS1         67 DRYHSCPPEQREK--------------------------------------------------- 79
PPM3            67 DRFANLPTQERNKR--------------------KIENLEYLQKAL-RKLTGEKEWVPNQIISGSKS- 112
PPM4            67 GRFANLPMHERNKR--------------------KLENLEYLHKAL-KKLAGDKELVSNQLISGSKS- 112
PPM7            67 SRFANLPMHERNKSFEDMLTRFANFHMHDRTKYNRKIENLEYLHKSL-KKLNGEKDSASNQLLLGNKG- 134
PpMADS2         67 ARFANLPMHERNKSFEDMLTRFANNQMHHDRSKYTRKIENLEYLHKAL-KKLAGEKDLVSNQQLLSGSKS 135
PpMADS3         67 ARFANLPMHERNKSFEDMLARFSNNHMHHDRSKYNRKVENLEQLQRAL-KKLAEEHGLAASQDHLSGSKS 135
CRM13           69 MRFANLSPQDRAKR--------------------KVEALEQLKRVF-KKLDHNVNVEDYTRVPT---- 111
CRM16           69 MRFASLSPQERAKR--------------------KIEALEQLKRTF-KKLDHDVNVEEFTRQNVTA-- 113
P. taeda 23021  69 AKFCELTPQERCKR--------------------KLESLEVLKKTF-KRLDHDVNIDDFLEARCSSEK 115
Picea 67868     69 AKFCELTPQERCKR--------------------KLESLEVLKKTF-KRLDHDVNIDDFLGERCSSEK 115
P. equestris    69 SKFYQLSPQGRENS--------------------KLEILETLRKTF-KNLDHDVKIEDFIGSSA---- 111
Med. tru. 331   69 AKFAQLSPQERAKR--------------------KMESLEALKKTF-KKLDHDVKIQDFLGSNS---- 111
Gos. rai. 862   69 VKFAQLTPQERAKR--------------------KLESLEALKKTF-MKLDHDLNIHDFLGANS---- 111
Cen. mac. 1285  69 AKFARLTPQERAKR--------------------KLESLEALKKTF-KKLDHDVNIQDFAGASS---- 111
AGL65           69 SKFAQLTPQERTKR--------------------KLESLEALKKTF-KKLDHDVNIHDFLGARN---- 111
Cit. sin. 77561 69 ARFAQLTPQERAKR--------------------KLESLEALKKTF-KKLDHDVNIQDFLGAST---- 111
AfMADS1         71 SRFAQLTPQERAKR--------------------KLESLETLKKTF-KKLDHDVNVQDFLGTGT---- 113
OsMADS68        69 TKYAQQTPQERAKR--------------------KLESLEALKKTF-KKLDHDVNIQDFLGSGG---- 111
EcMADS1         69 AKFAQLTPQERAKR--------------------KLESLEALKKTF-KKLEHDVNIQDFLGTST---- 111
A. for. x pub.  69 AKFAQLTPQERAKR--------------------KLESLEALKKTF-KKLDHDVNIQEFLGNST---- 111
Bet. vul. 10637 68 ARLNQLPPQERTKR--------------------KLEGLEALKKTYRKKSDHAVNIEDFLDTSA---- 111
AGL94           68 AKFAQLSPQERAKR--------------------KLENLEALRKTF-MKANHDIDISKFLDRIST--- 111
AGL30           69 AKFSQVTPQERTKR--------------------KFESLENLKKTF-QKLDHDVNIREFIASSN---- 111
Gos. rai. 800   69 EKFAQLTPQERAKR--------------------KLESLEALKKTF-KKLDHDVNIHEFLGSST---- 111
SmMADS1         68 LRYANVSEAERSKR--------------------KMENWEQLSRAI-KKLKSETDGSQISRNTIDF-- 112
SmMADS2         68 MRYANVPLAERIKR--------------------KKENLEQLNRGI-RKLKGDNEIARGMGF------ 108
SmMADS3         68 LRFISLPENERTKR--------------------RLDSEEIINKIR-RSYPSGSEITNGETSNSNY-- 112
CRM14           68 ARFANLSDHERAKS--------QCLKWCNSHNLPPKIANSEMLMRAA-KRLKQETYMAIQLSSSCSRV- 126
CRM15           68 ARFVNLSDHERTKS--------QCLKWCDKRSLLPMIANTEMLMRAA-KRLKQESYMAVQLSTNSPRV- 126
Hev. bra. 267   68 TRYVNLPEHERGR-----------------------LHKQEFLEKAL-GKLKAEGDRNHQAASNPAITD 112
Cof. can. 17252 68 ARYLNLPEHERGR-----------------------LHNQEYLEKAL-GKLKSEADRTNQDVSPVSVD- 111
Vit. vin. 32798 68 TRYVNLPEHERGR-----------------------LHNQEFLQRAL-GKLKSEANRTYQAAASPMSTD 112
Cen. mac. 1671  68 NRYVNLSDSERGS-----------------------VIQHREYLNSLL-SKIRSENDLAIHIPCPGAID- 111
Gly. max 42899  68 TRYINLPDQERDN-----------AVGFPELPYRRSIQNKEYLLRTL-QQLRSENDIALQLANPGDIN- 123
Pon. tri. 100   68 ARYVNLPEHDRGR-----------------------LRNQEFLQRAL-GKLRSEANQTYEATSNPTTNT 112
AGL67           68 ARYINLPDQERENAI-----------VFPDQSKRQGIQNKEYLLRTL-EKLKIEDDMALQINEPRPEAT 124
AGL66           68 SRYINLSDQERENAL-----------VFPDQSRRPDFQSKEYLLRTL-QQLKAENDIALQLTNPTAIN- 123
AGL104          68 SRFINLPKQERESAL-----------YFPDQNRRPDIQNKECLLRIL-QQLKTENDIALQVTNPAAIN- 123
OsMADS62        68 LRYMNLSEHDRGE--------------------AIQNREYLISML-QRLKRESDMATQLANPGALN- 112
OsMADS63        68 TRYINLPESDRGG--------------------TIQNREYLINML-TQLKCESDVTEDLTNTSSKAP 113
EcMADS2         68 TRYINLPEDHRGG--------------------IIQNREYLIRTL-KKLKCESDIALQLANPAAIN- 112
AfMADS2         68 TRFISLPDHDREG--------------------AIHNREYLIRTL-KKLKCESDVAAQLANPGAVN- 112
                                                    a   a   d    a    d   a   d   a
```

MIKC* K1

```
                   |-------------------------||----
```

duplicated segment I

```
                    ----------------------K-domain-continued--------------------------
                    ****************************         **********************
                         MIKCᶜ K1                           MIKCᶜ K2
                    d    ad  ad    d            a    ad     ad   ad   ad
DEFA              82 ---SHYEKMQEHLKKLNEVNRNLRREIRQRMGES---LNDLGYEQIVNLIEDMDNSLKLITRERKYKVISN 145
GLO               82 ---PKHEHLDNEINRVKKENDSMQIELRHLKGED---ITTLNYKELMVLEDALENGTSALKNKQMEFVRM 145
StMADS11          84 LHSSNLLSEKKTHAMLSRDFVEKNRELRQLHGEE---LQGLGLDDLMKLEKLVEGGISRVLRIKGDKFMK 150
StMADS16          80 DEPSLDLQLENSLNMRLSKQVADKTRELRQMRGEE--LEGLSLEELQQIEKRLEAGFNRVLEIKGTRIMD 147
GGM2              80 ---IELQIMGQELIKERRENEKLRSKLRYMMGED---IGELKIAQLEKLEHDLESALRLVRRKKDHAWDY 143
GGM3              82 IAESNAQYWQQEAVKLKQQIDVLNNQIRHYMGECL---QSMTIKELKQLEGKLEKGLGRVRSKRNEKLLE 148
CRM1              80 --DNGSSTDNVEADRLTVFTEKLKMLQSNVIGDD---LERLSLRDLIHLEQQVHESLGHIRAKKDELILE 144
CRM3              86 EEACVSQDLRAELTELRKEVESLRQEKRRKDGDIH-DLKLLSADELDSLEGEVETSLCSIRKRQKQLYRE 154
SrMADS1           83 SHHEASDFISHEIRRLKQQLQRSQQSRRHLLGDD---LSHLPIKDLQNLEQQLEVGLSRVRSRKDQVLMD 149
LAMB4             81 VIPRDLEYWSREAAKAKDELDRCHQKNRHMLGED---LSALNLNELEELEQQLDSGLRRVCRKDQVLRD 147
PPM1              81 --GASTDFLGREVVKLQEQVERLKSSQRRMLGED---LSALKVPDLLQLEQQLDLGASRVRARKNQLILE 145
Csmads1           82 QRLDNMDYLNREVAKLRNEVEHKYHEARQLEGED---LDRLGVYELEQLEQKLSNSMRRIRGRKDELMKA 148
CgMADS1           80 RKFDNTDYLSKEAKRLRHEVELAKQENRHLSGED---LNAVQMPELDGLELKLEDALRKIRFRKREVMQM 146
PPM3             113 ---EEVELLQEELKKTQHEKELIQQRIRLYLADEQLLQSVTSVQQLANMETELEQALERVRTRKAYVTSA 179
PPM4             113 ---YEVGLLQEELKKSQQEKELVQQRLYLADQQLLQSVTSVQQLANMETELEQAIERVRARKAYVSSA 179
PPM7             135 ---YEVGLLQEELKKSQQEKELVQQRARLYLADEQLLQNVTSVQQLANMETELEQAIERVRARKSYVSSA 201
PpMADS2          136 S-SYEAEVLQQELKKAQQEKELVQQRARLYLADEQLLQGVTSVQQLANMETELEQALERVRARKNYVSSA 204
PpMADS3          136 --TYETELLQQELKKAQQEKELVQQRARLYLADEQLLQGVTSVPQLANMESELEQALERVRARKNYVSSA 203
CRM13            112 ---NSIEELHQHLQNLQLQFLHLQQRLRLYETY-------QTLEEANSVERVLQKRLEEVRSRKQQLSNL 171
CRM16            114 ---QNIEESHLYLQNIQIECLQLQQKLRAYESF-------QNIEEADKLEEFLQRRIQEVRVHKELLGSR 173
P. taeda 23021   116 ---LEVQELHNELRRIQTKVLSLQEKTWYFSDE--NQTNILNLDQTRDVEKGLMEALARVRMRKEALKND 180
Picea 67868      116 LEKLEVQELHNELTRIQTKVLNLREKTWYFRDEI-NETNILNLDQQARDVEERLMEALGRVRMRKEALKND 184
P. equestris     112 ---RTEEELTNHLQTLKAQLAELELRLSYWTDP----GKVQSIDQTQRMEHTLRASLNQIQIQKASFGNQ 174
Med. tru. 331    112 ---QTVEDLSHQVRVLQARLAEVQQRLSYWSNL----EKINNLEHLRQMEDSLRESINRLSIQKENLEKH 174
Gos. rai. 862    112 ---QSIEEMTKEVSRFRAQLAEVHKRLSYWSNP----DKIDNIEHLRQMEDSLRESIERVRIHKENFGKH 174
Cen. mac. 1285   112 ---QSAEDLSNHAMMLRSQLADIHKRLSYWSNP----DKIDNIEHLKQMEDSLRKSLDRIRIHKENFGQQ 174
AGL65            112 ---QTIEGLSNQVAIYQAQLMECHRRLSCWTNI----DRIENTEHLDLLEESLRKSIERIQIHKEHYRKN 174
Cit. sin. 77561  112 ---QTVEELTHQVRILQAQLTEVQQRLSYWSNP----GNIESTEHLRQMENSLRESINQICLHKENFGKQ 174
AfMADS1          114 ---QTVEELTNQSRLLQAQVAEMHKKLSYYSNP----EKINNLDQIRLMEETLRDSINRVQTHKVNFEKQ 176
OsMADS68         112 ---QTVEELSSHLGALQCQMADVEKRLSYWSDP----EKVENIDHIRAMEQSLKESLNRIRIHKENFAKQ 174
EcMADS1          112 ---QTVEDLTNQRNLLQRQLSEMVERLRCWTDP----EKINNVDHLNAMEQSIKESLNRIHTHKENLGKQ 174
A. for. x pub.   112 ---QTVEDLTNQSRLLEAQLSEMHKRLSCWTEP----EKINNLRAMEEDNLKDSLNRIRIHKDTFGKP 174
Bet. vul. 10637  112 ---PTLKDVTEQASKLQTRLAEIHERLWYWTNP----DNVNDFGLLMQMEDSLEKSLDQIRQRQKENFQKQ 174
AGL94            112 ---PTVEVLSEKIRFLQTQLSDIHTRLSYWTDV----DNIDSVDVLQQLEHSLRQSLAQIYGRKASMPQR 174
AGL30            112 ---STVEDLSTQARILQARISEIHGRLSYWTEP----DKINNVEHLGQLEISIRQSLDQLRAHKEHFGQQ 174
Gos. rai. 800    112 ---QTIEDLTNQARLLQARLSEIHRRLSCWTDV----DKINNVEHLGQMEDSLKDYLNQIRAHKENLGKQ 174
SmMADS1          113 ---NNLGMIQEEIARLQLENEQLIERFKFYQGDT-QILEHLSYEDLAKLEEELATTLHHIRIRKHNMDVE 178
SmMADS2          109 ---EDIEVLQKEIARLQQENQILQNHLRLFQGDCAS-LDSLNYEELHQMELHIQTVMEKASLLKILDYNK 174
SmMADS3          113 ---ESLVAHYRDVSKIQFENEALRTRLRFFEGD----IGGLGIEELLQLEECLKSTLQKIRTRRKQELQNH 175
CRM14            127 ---TNLETLQYEIQRTQVQLQDIETRLRIYEGDL---KCMHSMEQLDSLEQHLQTALESLRIRKQALDQN 190
CRM15            127 ---TSFETLQFEVQRSQVQVQDMENRLRMYEGDL---KYVNSMQHLEIIEQQLQTALETLRIRKQATEGC 190
Hev. bra. 267    113 ---SQLEEFQQEIVRFKSQVEDMEKQIRILEGNL---SHTTTLPEAEYQEQILEEALKRVQMRKQVLEEK 176
Cof. can. 17252  112 ---SQIEEIQQEILRYKSQMEDMEKKLRIYEGDP---WEINTICEAEYREQILEETLNQVRARKVNLAQQ 175
Vit. vin. 32798  113 ---SQLEEIQQEILKYKSQLDDMQKRLRIFEGDP---SEITTVREVEYREQILEETLKQVRFQKQALEEK 176
Cen. mac. 1671   113 ---MQSEEIHQEISTLQHQLHMAKEQLRIFEDDP---TTFTSLYDYESYEKLLYKTLDRVTQRKEYLLSN 175
Gly. max 42899   124 ---SEIEELQQQEVNRLQQQLQMAEEQIRLYEPDD---LKMSSMADLENSEKHLVDVLTRVIQRKEYLLSN 187
Pon. tri. 100    113 D--SHLEEVQQEIVKCKSQIQNMEKRLRIFEEDP---SEMTTLYEIKHREQILEETLKQVRIRKHLLEEN 177
AGL67            125 N--SNVEELEQEVCRLQQQLQISEEELRKFEPDP---MRLTSMEEIEACEANLINTLTRVVQRREHLLRK 189
AGL66            124 ---SDVEELEHEVYKLQQQLLMAEEELRKYEPDP---IRFTTMEEYETCEKQLMDTLTRVNQRREHILSQ 187
AGL104           124 ---SDVEELEHEVCRLQQQLQMAEEELRRYEPDP---IRFTTMEEYEVSEKQLLDTLTHVVQRRDHLMSN 187
OsMADS62         113 ---EKIEIQQEIYSSQQQLQITEDRLRMFEPDP---AAFGTSSEVDGCEKYLMELLTRVVERKNNLLSS 176
OsMADS63         114 V-NSNIEELQQEIRRCQHQMQLTEEQLRMFEPDP---ARSASMEDVEASEKFIAGILSRVEERKRYLLCS 179
EcMADS2          113 ---SNIEDLHQEISRHQHEIQMIEERLRIYEPDP---LKISSMADIESCEKNLFDTLTRITERKKYLLSN 176
AfMADS2          113 ---AQIEELQQQEINGHQHQIQMAEEEQLTVFEPDP---LRFTSMEEFDDSEKFLVDALTRVTQRKKFLLSN 176
                    a    ad   ad  ad     ad    ad
                         MIKC* K2                           MIKC* K3

                    -------------------------|
                    duplicated segment II
```

```
                     ----------------------||----------MIKCᶜ C-terminal domain --->>
                          MIKCᶜ K3
                        a  a   d  a  d   a
DEFA            146 QIDTSKKKVRNVEEIHRNLVLEFDARREDPHFGLVDNEGDYNSVLGFPNGGPRIIALRLPTNHHPTLHSG 215
GLO             146 MRKHNEMVEEENQSLQFKLRQMHLDPMNDNVMESQAVYDHHHHQNIADYEAQMPFAFRVQPMQPNLQERF 215
StMADS11        151 EISSLKKKEAQLQEENSQLKQQSQARLNEEGQNVIEQGHSADSITNNRSLVNSHQDYNDSDTSLKLCLAF 220
StMADS16        148 EITNLQRKGAELMEENKQLKHKMEIMKKGKFPLLTDMVMEEGQSSESIITTNNPDQDDSSNASLKLGGTT 217
GGM2            144 QRTILLKKVKLQYALHEQMSRQLPMEVLAKAEEEARQILASTAEEEARQNMTFSFLPNASTQYARIA    210
GGM3            149 DIDTLQRREDNLIRENEYIRNKIAECQSHQHANMLTAAAVEYDAIPAAYDSRNFMHANLIEAAAAHHHYA 218
CRM1            145 QIDEFKQKMADTRRTTNANTSMLDRLVDFCSSGITGSQNIVESEPGYVVGVPRTDTLRAGWIATSDQSND 214
CRM3            155 KMNETFRKEEELRKENEKLRKQVEELQCARWEICCAQGFEASATSMCTSHGMTSERSDSFLDLRLN     220
SrMADS1         150 QVDELRRRELTLHKDNEMLRRRLSDVQGMAESGRSTLIVNPWRQGNLLQTWQMYPCHFLDTKAAA     214
LAMB4           148 RIEELTKQESFLREENRMLRSKIAVPKESTEPILDEMNMETREPPSMAIVEEELSLKLNSKTWPDIHFSL 217
PPM1            146 EIEGLQKKEQELMVANEDLRKKIADAEAVARANLSEARPESPRHLARTLSRDVSASSHPAATVYPHPNLR 215
Csmads1         149 ELEGLRKQVADMETALVGAASTDGRPLSGSSNYLLQSIPGIRTMPPSSLGGMNPASTSLQLGSDRLFGNR 218
CgMADS1         147 EIDRLQQKVWTSSPQFLLAWRSCHPAVVQIIIFCSHPAFVPKFILQ                        192
PPM3            180 YQAVSVMQRQMLALQAQQQKQQRDTLGQQPLPLQWNNPECGHSMLQEFMDQHANPQAIVPVQHMNREMGS 249
PPM4            180 YEAASVMQRQVMALQVRAQQQQQQQGLSGSTHPPFLQWNNPERGEAILQDFMEQQTNSQSIVPVQMNREI 249
PPM7            202 YQVSSAMQRQQHEFLGNSYQQMMVLRQQQAHQHHQQQQQQQKAQHQHQQQGGIPAAAQQLPFLQWNGA    271
PpMADS2         205 YQTANALQRQQHEFLGNSLQMMALRQQQQQQQQQAGMGGTPSFLHWTMQSERPEPTQDFMEQQTNSSASL 274
PpMADS3         204 YQNVIQRQQHEFLGNSLQMMALRQQQSQQAQQQQAGMTPSQVPYMHWTMQQERPEPTPDFMEQQTNPAA  273
CRM13           172 CVAPYNESMEQQIYADVVMPSSVQQHQGANWDANGSRQQTTTWDFEGNRQHSSSNWDTDGTRQLNNIALS 241
CRM16           174 QCIVPCGPSASQQVYSDGVTNSNFPAQGAWESHHDRLQDISELQAKYYCQRYGEELPEAWFNDNT      238
P. taeda 23021  181 QILTSAERSEKLYHNNICFPSAPEGTEQSQLMSWMPTDGQPFMVHEDPTLGLCQRDVKCAAEASLPVPSG 250
Picea 67868     185 QILTPAERSEKLYHNNICFPSAPEGTEQSQPMSWMPTDGQLFMVHEDPTLGLSQRDVNCAADASAPVPSG 254
P. equestris    175 LMPLCNSGQNSIPLPMAMSNDQQSPHMPWIHGCDGHHLMPFRKMLNLMQKEIRQLPRPYTFKGFLTLKDP 244
Med. tru. 331   175 QLMSLECVNQLPEGMSLPLMMSGLQESQPLSWLLSGDNHQLMLPSEPKFMPFSDNGNRDVECSTDISLPS 244
Gos. rai. 862   175 HLMSLECCNQFQNRIPLSVMIGAVQEAQPVMWLPNNENHHTLLHNELNFLPHRDAECSTDCSLAGYSGFF 244
Cen. mac. 1285  175 KLIPLDCMSQFQNGLHLPLMMANTQEDQALPWHPNNENQNLILPEKQNYMPHRDGECSGMSITNFSGLFG 244
AGL65           175 QLLPIECATTQFHSGIQLPMAMGGNSSMQEAHSMSWLPDNDHQQTILPGDSSFLPHREMDGSIPVYSSCF 244
Cit. sin. 77561 175 QLMSLEFAGQSGMHLPWMMNVMQEKQSLSWLPNNDNQHMLVPNDPASCPQEIWDVQEDATFQAILGSWAP 244
AfMADS1         177 QLMSLECGQFQNGMHLPLAMGGEQQAQTLSWLANNDGRAILMTENPNLVSHRDIECSSSAPVPSFSSYFN 246
OsMADS68        175 HLMSLQCAAAQFQNDMKLPLGLTGDPNTSSWFHGGGGAEAQQPMMLPEDPSLLHQRDIGCSASTSLQSYP 244
EcMADS1         175 QLMSLECAQFQNGMQIPLGMGGEQQPQSLSWIPNHDSQHLMLSEDTSLLPQRDIECSTDPSLQNYSSYLG 244
A. for. x pub.  175 QLIPLECTSSQFQNSIHLPLGMGGEQQAQNMTWIPHNGGQHLMFSGDPNLLHQRDMECSTDASVPSYSGY 244
Bet. vul. 10637 175 HCLALVNEENFQKEHRLA                                                     192
AGL94           175 QQQQLMSSQCKNQLQTEIDIDFGMEMEQQLENFSWVRTDEN-MNVPIEEEDPNLQLHHMYKDITCSASSA 243
AGL30           175 QQAMQIENANFVKDWSTCSMQDGIQIPLEQQLQSMSWILNSNTTNIVTEEHNSIPQREVECSASSSFGSY 244
Gos. rai. 800   175 QLLPIECTSQFQNEMHVPFRMGIEQQLQSLAWMPNNDSRHMALPEDPNL                      223
SmMADS1         179 FCMRQKQTEHFLPQQVYHLPLHASGINSFSSHWSDRDPQIIKFLTEETSVGSSTDTLYFEQQNNIPYQTT 248
SmMADS2         175 LHPDMNLSLFRDEQHQALFMGDHLNPMSHEPQVGFPEVQNLQQEPHLFDSSDYTPT               230
SmMADS3         176 GMGGMMQNQVVPYGQLQQMYPMMAPNGSMGDVQIPSSWSDSASQYNELSFFPQRRDNGDASSPSSSMSQA 245
CRM14           191 THLSMPLDSARAQIYGIMQSQPGGMLQHATPAHIWLSVLRDYQIPSTHNMAEHPNSILSLRDCQDGDCPS 260
CRM15           191 NFLAHGAAREQMYGMMNIHAQTGKMMEQTISPHWMQWLAVMQDSQMASSQSIVENPGPTPILSLRESQTG 260
Hev. bra. 267   177 YNSSGAPTVSQCNQAHLPPKNAYVNDLVTEESPKQCLGLASSERATSSDLEFFGLQWPSSLKKPGAARS  245
Cof. can. 17252 176 AAQVNGFTTTRSASSILDWFPHQRDQDQIPMLNFLDPTGLIPLRAGQADQRIENMVPASLTLPPQSHMDTA 245
Vit. vin. 32798 177 YNTTEPQPTAQVHLPPETANVNGFISGNPNNMLDWLPPRDPQVQILNFLDSNGLLPMREQPQRMVEILPP 246
Cen. mac. 1671  176 QPPAYHLRVLQGESSSKNEDINCWPENVCDGANQMNGQSHQNMFTAASPENTFEATRNQLSNALFDPLPK 245
Gly. max 42899  188 HLSSYDPSGIQGIPTSFENVGWLQDGS                                           214
Pon. tri. 100   178 CNSSSTPATSQARLPPAETSDHMNGFVTESPNNILEWLPHRNSQVHNLNILEFKWPPFPLKKEPQEANKN 247
AGL67           190 SCEAQSNQQSMDGILLNDIVEDWGPEPEPKQAHMIANSAHHSNQPSYDLLLRRSNSSSNQNPK        252
AGL66           188 DQLSSYEASALQQQQSMGGPFGNDVVGGWLTENGPNEAHLFDASAHSAMYETLLQGSSSSSNQNNIMGES 257
AGL104          188 HLSSYEASTMQPNIGGPFVNDVVEGWLPENGTNQTHLFDASAHSNQLRELSSAMYEPLLQGSSSSSNQNN 257
OsMADS62        177 HMAPFDATTAAMQGADGTQMYVSQADGLATFGGDAAMWGPDGGADPGHPMFSASDPLIYLRDHDVYDANS 246
OsMADS63        180 MGSFDVTASTSAMQHLYLPQQHQHGDITDNGFGSDEVASWVSEGMPPTTSSVASIFAGTSDSMMSFRDQA 249
EcMADS2         177 QQIASYDPSDIQMYMDPQEGMGSTSFRNEVVNWMPDHQHNPNQIYVESHSLSPLRDHQSNMFHEQLQHGS 246
AfMADS2         177 -HLPSYDPSSIPVYLDSQEGLAPSFESEVANWLPDQNAPNPGQIYVGSDPLIHLRDQQSSVYDPLSHPEG 245
```

```
DEFA               216 GGSDLTTFALLE                                                    227
GLO                215                                                                 215
StMADS11           220 P                                                               221
StMADS16           218 AVEDECSITSLKLGLPFS                                              235
GGM2               210                                                                 210
GGM3               219 QQEQTALHLGSEHKYSVHYPDPQAHKFQT                                    247
CRM1               215 GPEYVPCAKRFRITEDLNESPASE                                        238
CRM3               220                                                                 220
SrMADS1            214                                                                 214
LAMB4              218 QLGQYQPPRTETPTANHRPN                                            237
PPM1               216 DVQRSQTSLQLGMFSSESYPPSSSRWPSEQQFPSASEGCAGESSMKWDHPHYHIQNRLHANILPSVRI 283
Csmads1            219 GVELHDRSASDESPVMTNRMSVDFAQAPREMSGVDLSGSPVPPWKSQAAAAAQQEWKNQASSPTDWKVTN 288
CgMADS1            192                                                                 192
PPM3               250 NSEASPSSFFSQSGSQSNLHNLAGQLNLVGANRGLNIHDMALERSGSGQYDEQKKSKGGANAGFASSSSS 319
PPM4               250 GSSNEAGPSTFFPQTGSQSNLHSLAGQMSLVGVNRGLSIHDMALERSRGQYEDRKKSKGESNTGFGASSS 319
PPM7               272 QERPEATLQDFMEQQTNTSQAIVPSQMSREVVSQREVSPSNNGFFPGTASQSTLHNLGQIGQLVGANRGI 341
PpMADS2            275 MPAPLSIREPGSNGEANPAAAAYFPAGPQPGALESLRQMGMAAASRGLTMHEMSLEQPNGQYIEDQQKK 343
PpMADS3            274 PLMHASLNREPGSNGESNPGVGYFPSGPQPSNMETLRQIGMGGVSRGLTMHEMSLEQPNVAQYSEEQKK 342
CRM13              242 QLSSPMFMQSNLL                                                   254
CRM16              239                                                                 238
P. taeda 23021     251 FMFNCKQGNMEDISSQLGTFSNPLNRRGLSLHLEYQDSSYHTDCHYTCSDVTEEKESKPELNLQGTTMDY 320
Picea 67868        255 FMYNCKHANMEDTSSQLATFSNPLNRCGLSLHLEYQGSSYHPDCQYTCSDVTEEKESKPELSLQGTTAVY 324
P. equestris       245 XCYQTRTWECLNXFEPKLLXASATWGSIIPFQAL                              278
Med. tru. 331      245 YSGYTGNCKLGGGKLSTCHYFGSRRWYLE                                   273
Gos. rai. 862      245 GSGKQTEISSSGQVDNVVQECNALNELGSNACLNLEPGEQYFYQPYSASNYQDDEKLKTEMEVNLQGNPV 314
Cen. mac. 1285     245 GGKQLEMDGAGKVDTTRQDGGLAELCSTSNLRPQFSEQFPFHPYGNLNFQQSKEAKPETATSLQGFLDFS 314
AGL65              245 FESTKPEDQICSNPGQQFEQLEQQGNGCLGLQQLGEEYSYPTPFGTTLGMEEDQEKKIKSEMELNNLQQQ 314
Cit. sin. 77561    245 GNR                                                             247
AfMADS1            247 NGKTEMDTSGQEGGTLNELNPNECLRLQLGGQFPYEPYNLNLIDEKKFKPEGDMLHEQSVDYHVSGFEQS 316
OsMADS68           245 GYFSMGKQSTDNAGGGEQHHHAAVQQQPEFSQADCLTSLQLGAQFPYPSAFDNAGLLSDRLFDNAAAAAA 314
EcMADS1            245 MGKQVEIDNSGQDGIGSSLPDFSANACLSLQIGGQFPYQPYDFDLVNNNKYKNDSEMSLQEPRLDYQVGN 314
A. for. x pub.     245 YGEGKHAEIDCSGQNGGAFHELNGNACLRLQLNGQLPLQQYGLNFLFDKNKSELKMNVQEPPLDYQGGNS 314
Bet. vul. 10637    193                                                                 192
AGL94              244 LGNYSGLFSKSSDILQKLETGSIPGTSADPNQQFSNLSFLNDQKLKQLAEWNLLGSPADYYVSQILEASY 313
AGL30              245 PGYFGTGKSPEMTIPGQETSFLDELNTGQLKQDTSSQQQFTNNNNITAYNPNLHNDMNHHQTLPPPPLP 313
Gos. rai. 800      224                                                                 223
SmMADS1            249 QLNLFGTHGLSFDNANGRGHCYIGDNKNPIIYISPNPNTNLEHSMEQPQNEQHQREISLNWNLASHPNNQ 318
SmMADS2            231                                                                 230
SmMADS3            246 YTNQMRMNQVQQLKSPFGMQQAMMSEAKMFPLYYEEMETQPTIYKEEKADLIPQQQEEEHKESLVFDSTE 315
CRM14              261 ASSSLYSRTLCHVSDMQGTAPEADHSGEQQAMPQFRAGQIYQVDSQSIQEVEGKVEGVDCREHKYSTTKV 330
CRM15              261 GECSSASSSQYPKALCHLNEMQARPPQSDQSSAQHAFSHFRLEQMYRLASQDMQSVKNDKEEMDSREHNH 330
Hev. bra. 267      246                                                                 245
Cof. can. 17252    246 ADMQQQLNDHNSPISKIENDPQVLQRPHDDHFGQIIDVNLSPWTHLYPTDQLQMIPQRSSFMALFN 311
Vit. vin. 32798    247 SSTLLPGQNIHLEDHMSPSSGMEEDNNVQRPQFGPAIDVNLSPWAEFYPTAGNGSYPAAQPRGPALLE-- 314
Cen. mac. 1671     246 STMNEETQSIEGFQGCQQTEESLELWHQSTDDFLSSLITHD                        286
Gly. max 42899     215                                                                 214
Pon. tri. 100      248 NWLPPLKYFPGGTQY                                                 262
AGL67              253                                                                 252
AGL66              258 NVSNHNGDMFQEWAQAYNSTTAHNPSTLFPPMQHQHGLVVVDPNIEEIEIPVMKKDAQADHEVSDYDIRMP 327
AGL104             258 MSECHVTNHNGEMFPEWAQAYSSSALFASMQQQHEGVGPSIEEMMPAQQSDIPGVTAETQVDHEVSDYET 327
OsMADS62           247 QVAGGHAAAADAWRQAYTCTELLSTLIPTTPFPLMPHCLGPEDQYLSMEHGMVAAAQEPVEASTASCSYV 316
OsMADS63           250 VYDTMRQDACVDQTVVPEMGMCHVDQQNQSDDWQAYTSAEFLNALIPPTPFPLDDEDTMGPMLASSPLLM 319
EcMADS2            247 SLNVDPRSMGECHVNNQTEGNLPPWNQTYTSAELFSTLLPSSFSLIQHDGLVGSNMQPMMQQHEQVETPI 316
AfMADS2            246 LQVDPTMGECHISNSSDRNGNLPHWQPFLHSPDLLSNLNLISPGSFPFQQGLTGHEIQSIMDPGDRMSAA 315
```

60

```
DEFA             228                                                                                     227
GLO              216                                                                                     215
StMADS11         222                                                                                     221
StMADS16         236                                                                                     235
GGM2             211                                                                                     210
GGM3             248                                                                                     247
CRM1             239                                                                                     238
CRM3             221                                                                                     220
SrMADS1          215                                                                                     214
LAMB4            238                                                                                     237
PPM1             284                                                                                     283
Csmads1          289 TEHLDSWPKAPAPTPEWKSTSVQPEWKNQSSPSSEWKPLDWMYHGPQD                                     336
CgMADS1          193                                                                                     192
PPM3             320 GAAAEGSAETFTGQSEPNASNWQLSHQAHHHMHAYSSAQYPTGFSNQNEEAWK                                372
PPM4             320 SSGAAADQGTGETFTGQSEPAHGSAPSASNWQLSHQAHHAQAYSAAHYPTGYFNQNTDTWK                        380
PPM7             342 SMHDMALEQRGGQHEEQKKAKLEITKNMGFASSSSSGGAAAEGSAETNLNQTFNNQTEQVHGNANASNWH               411
PpMADS2          344 AKVEGSMGFATSSSAAEAVSENTVTRDHQQSEQTDVNGGAADWPPHPANHHPNAFTNGGLFSSFSQNGHH               413
PpMADS3          343 AKVEASMGFVTSTSAGPVGAEAVSKTNFSREQQPEQAHGTVSASEWPPHPANHHPHAYGNGSLFFSQNEH                412
CRM13            255                                                                                     254
CRM16            239                                                                                     238
P. taeda 23021   321 DTHLIGTGFDGLPQQWPSSYNQTAGSLFDTYVCSWQQSMSMCGSASASTSHQISTCEQMEQPLGTQMHAS               390
Picea 67868      325 DTQSIGMGFDGLHQQWSSSCNHVVGSLFDTCLDSWSMPMFGSASASTSHQMNTCEQVEQPLGTQMQAAYE               394
P. equestris     279                                                                                     278
Med. tru. 331    274                                                                                     273
Gos. rai. 862    315 GNQVISNFENPRPMYNNGHQAWVLSSGPCGIAMFDGNSYHQTPSSGQCHDLLKDQTCAEILFHGSWSFLV               384
Cen. mac. 1285   315 MNCNFEMPRPVYNDNVCHAWNPEPGPCPLPMIDGNSYSQPQQPEHHLMTNNQN                                367
AGL65            315 QQQQQQQQQQDPSMYDPMANNNGGCFQIPHDQSMFVNDHHHHHHHHHQNWVPDSMFGQTSYNQVCVFTPP               384
Cit. sin. 77561  248                                                                                     247
AfMADS1          317 RPEYDANHHSWASTSGHCGVPMFEEHTFQQQPN                                                    349
OsMADS68         315 AMDFGGHYDLPRPGDEASFQNWASAACGATMYDHQQQQQQQQQPAQLPAAATVEAPSFNHPSPHRQLMI                383
EcMADS1          315 INSFGPPRHRYDASDQAWASTSGQCSVPMFDGQPFPTHSNQIEYHHE                                      361
A. for. x pub.   315 FEPPRHGYDINHPSWESNSVPCPNDIFDEHSFPQQNS                                                351
Bet. vul. 10637  193                                                                                     192
AGL94            314 KPQIGGKNNGASSETLPYVAVFDDPLYFWPN                                                      344
AGL30            314 LTLPHAQVYIPMNQREYHMNGFFEAPPPDSSAYNDNTNQTRFGSSSSSLPCSISMFDEYLFSQVTKTKLS               383
Gos. rai. 800    224                                                                                     223
SmMADS1          319 MPDLYSSSQGMNEAEDESTESRLGEDDEPEDKDDLDNVFHSDES                                         362
SmMADS2          231                                                                                     230
SmMADS3          316 MQPVWNAYRNSEQNPSFYPGQQEAISDSNFSRGGAGNDSTQNNMMVSFDCSHHGIGETGLDSSNMLCGQQ               385
CRM14            331 DSQTHPQREMDVNAGTCVSSSSDWQSAFHSVHPSSIAPATSPLLNPISLRQHAASMGMMLAHQRHDQQSI               400
CRM15            331 TTENNYAHTANPGVDVNVGSSVISAPTEWQSIFNSVHPSMSNRASSGLLYPAALRQNALLGIMLGCEKRD               400
Hev. bra. 267    246                                                                                     245
Cof. can. 17252  312                                                                                     311
Vit. vin. 32798  315 LFLSQFTP                                                                             322
Cen. mac. 1671   287                                                                                     286
Gly. max 42899   215                                                                                     214
Pon. tri. 100    263                                                                                     262
AGL67            253                                                                                     252
AGL66            328 QLSSQ                                                                               332
AGL104           328 KVPQLSSQ                                                                            335
OsMADS62         317 PSDENSGTPVMAYDSNPPPANIA                                                             339
OsMADS63         320 PGIHDQQPPVEDMATAGCSQAPANDGNGLYAAEDIAPVNVG                                            360
EcMADS2          317 NCSQGQNINSVDVADYNTNVHQQIHNNNINFE                                                     348
AfMADS2          316 TTSNCTQIQTKIEENSTWQS                                                                335
```

```
DEFA            228                                                                                  227
GLO             216                                                                                  215
StMADS11        222                                                                                  221
StMADS16        236                                                                                  235
GGM2            211                                                                                  210
GGM3            248                                                                                  247
CRM1            239                                                                                  238
CRM3            221                                                                                  220
SrMADS1         215                                                                                  214
LAMB4           238                                                                                  237
PPM1            284                                                                                  283
Csmads1         337                                                                                  336
CgMADS1         193                                                                                  192
PPM3            373                                                                                  372
PPM4            381                                                                                  380
PPM7            412 QANHQAQAYINSQYPNGFFNQNADAWK                                                      438
PpMADS2         414 PAWK                                                                             417
PpMADS3         413 HAWK                                                                             416
CRM13           255                                                                                  254
CRM16           239                                                                                  238
P. taeda 23021  391 YECRAGGSCAASLM                                                                   404
Picea 67868     395 CHAGASCISSLMQEPWKNMQVQEYPVL                                                      421
P. equestris    279                                                                                  278
Med. tru. 331   274                                                                                  273
Gos. rai. 862   385 IEVKHTAHEALTQPVFAAGHVSCIFSVTL                                                    413
Cen. mac. 1285  368                                                                                  367
AGL65           385 LELSR                                                                            389
Cit. sin. 77561 248                                                                                  247
AfMADS1         350                                                                                  349
OsMADS68        384                                                                                  383
EcMADS1         362                                                                                  361
A. for. x pub.  352                                                                                  351
Bet. vul. 10637 193                                                                                  192
AGL94           345                                                                                  344
AGL30           384 QRF                                                                              386
Gos. rai. 800   224                                                                                  223
SmMADS1         363                                                                                  362
SmMADS2         231                                                                                  230
SmMADS3         386 VMCGSGAGGAGGAGAGGGEEDGAANASQQQPQQQQQQPHAQVSAPPISTSQSSLTLSTSYEN 447
CRM14           401 MPFASQQTQVEPAASRSMLNDVDSGAIQMADGQNVLSLATSFGGSSYN                                  448
CRM15           401 QQELMPFEVTNVHRGEAGAEAPD                                                          423
Hev. bra. 267   246                                                                                  245
Cof. can. 17252 312                                                                                  311
Vit. vin. 32798 323                                                                                  322
Cen. mac. 1671  287                                                                                  286
Gly. max 42899  215                                                                                  214
Pon. tri. 100   263                                                                                  262
AGL67           253                                                                                  252
AGL66           333                                                                                  332
AGL104          336                                                                                  335
OsMADS62        340                                                                                  339
OsMADS63        361                                                                                  360
EcMADS2         349                                                                                  348
AfMADS2         336                                                                                  335
```
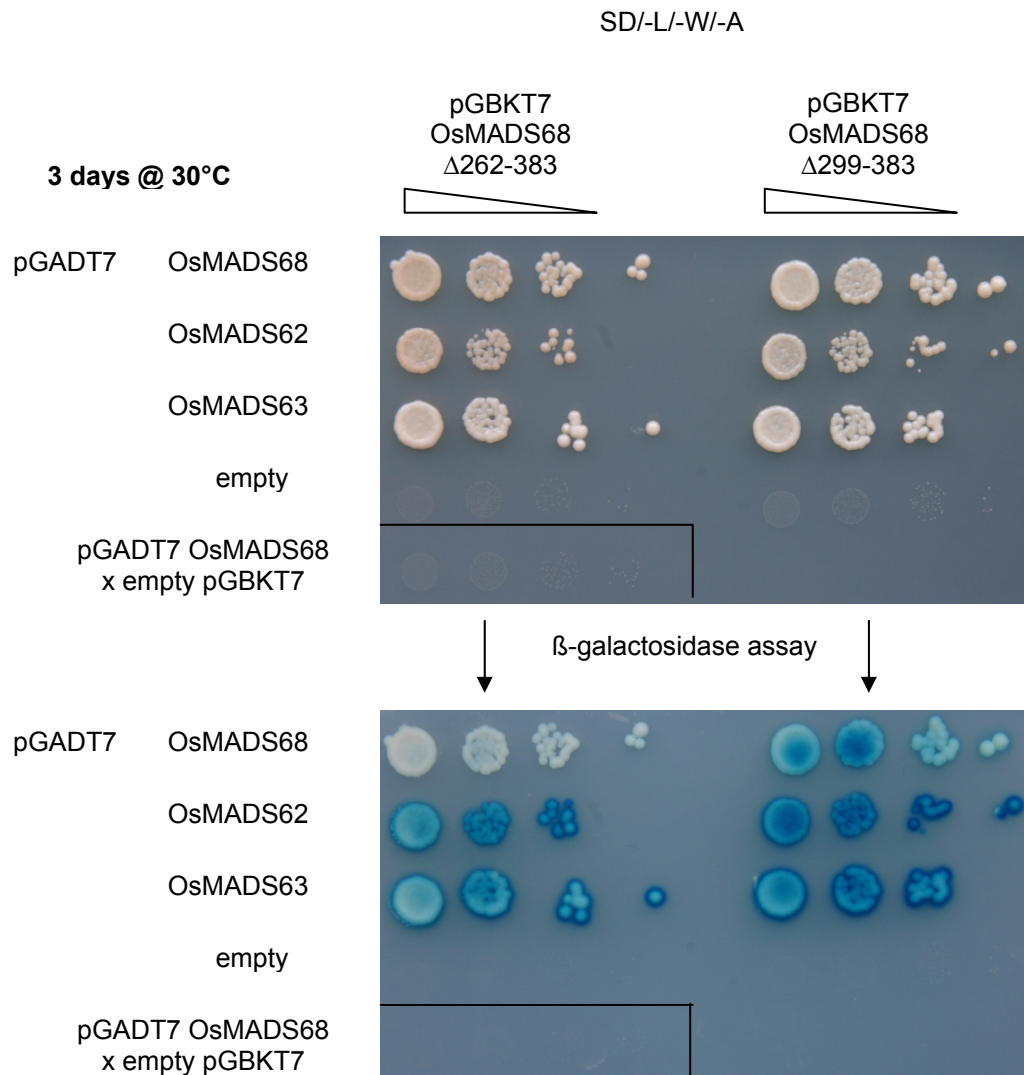
Figure S2. Data on the growth of cotransformants containing pGBKT7 OsMADS68 constructs combined with all Oryza prey constructs plated on non-selective SD/-L/-W medium. Cotransformants with the full-length OsMADS68 pGBKT7 vector that showed autoactivation are not shown.
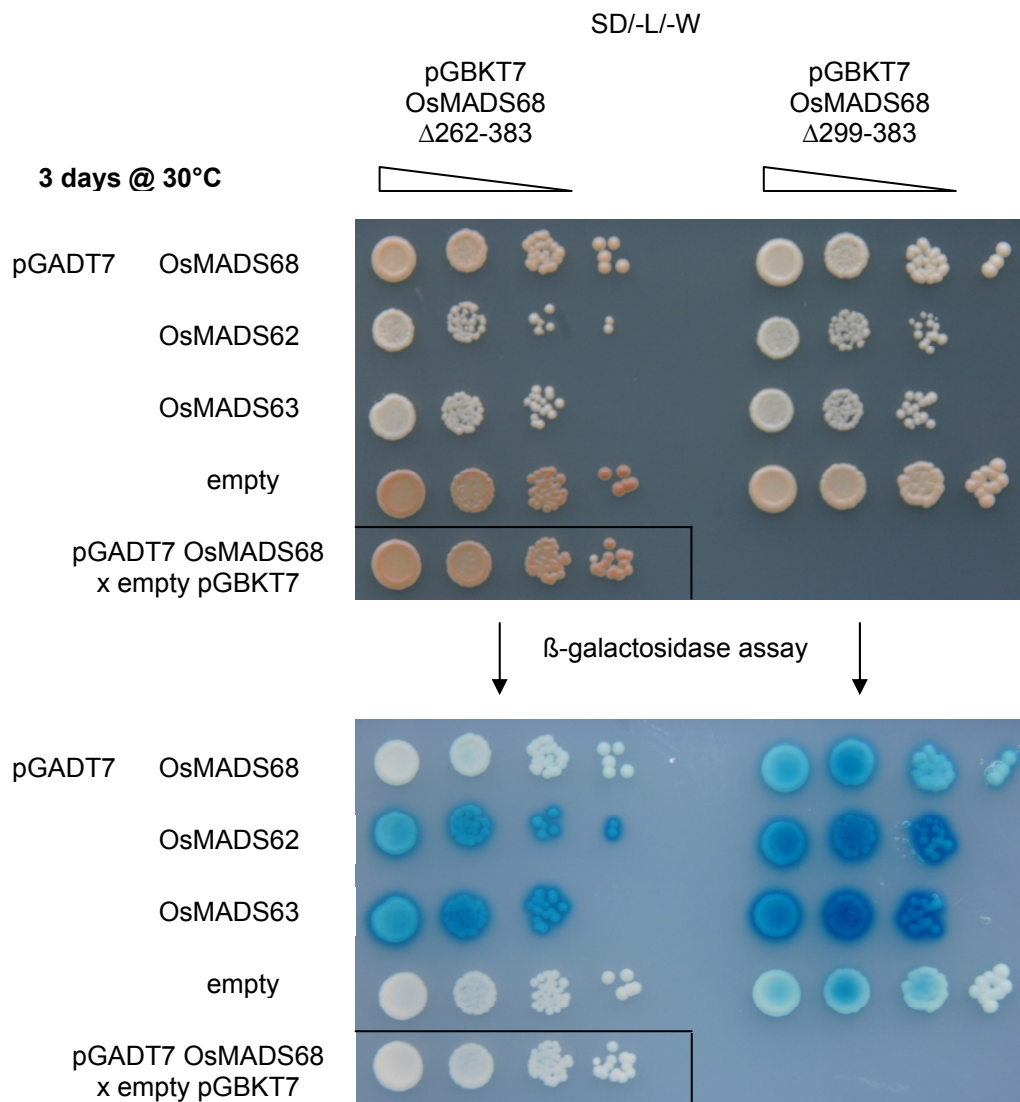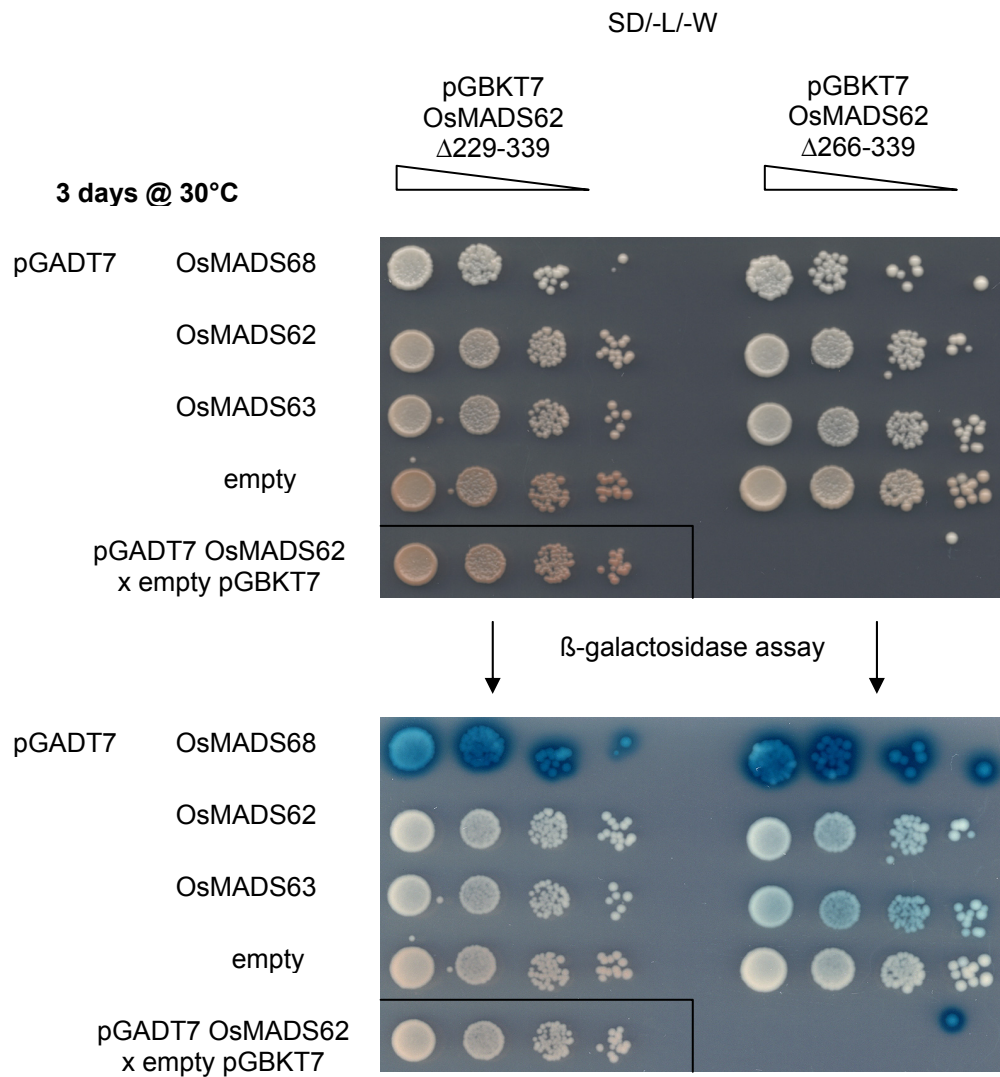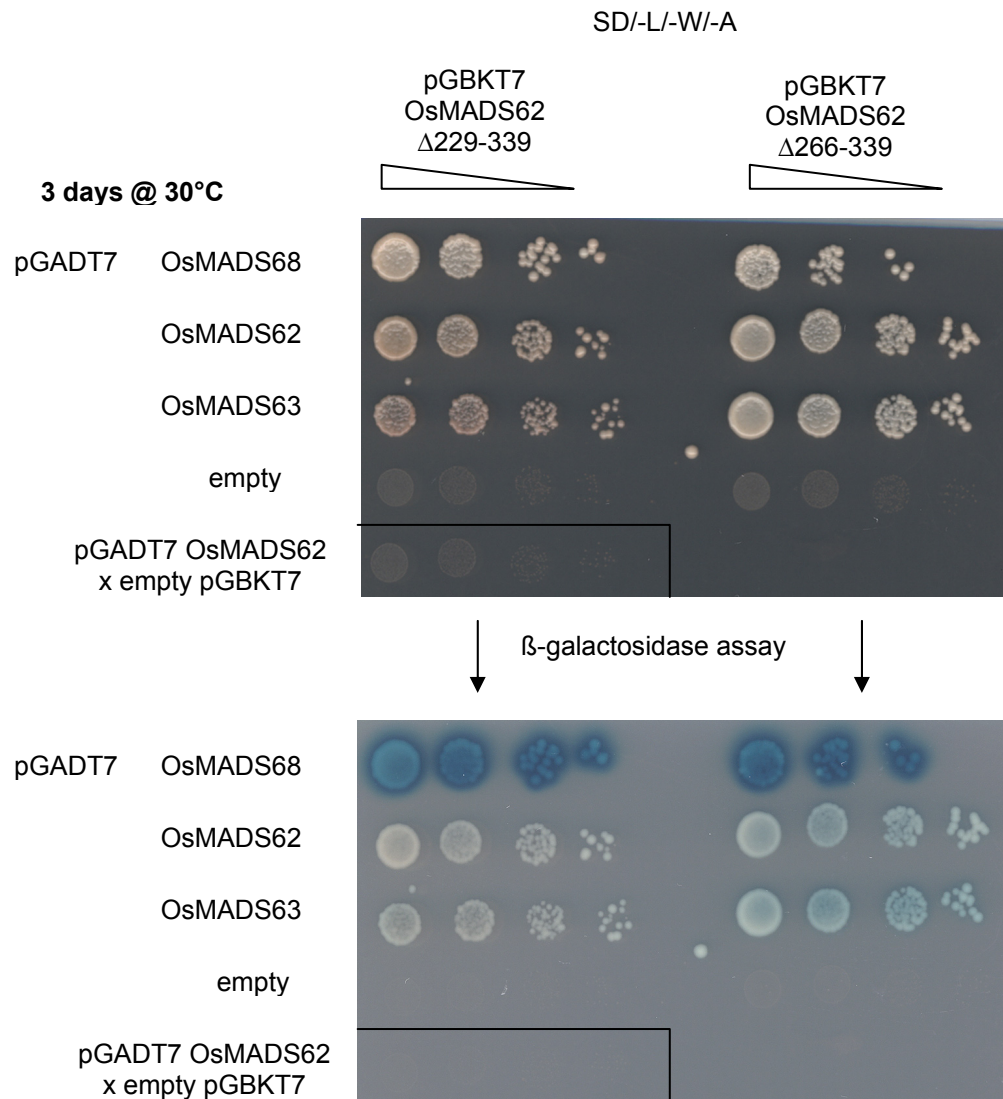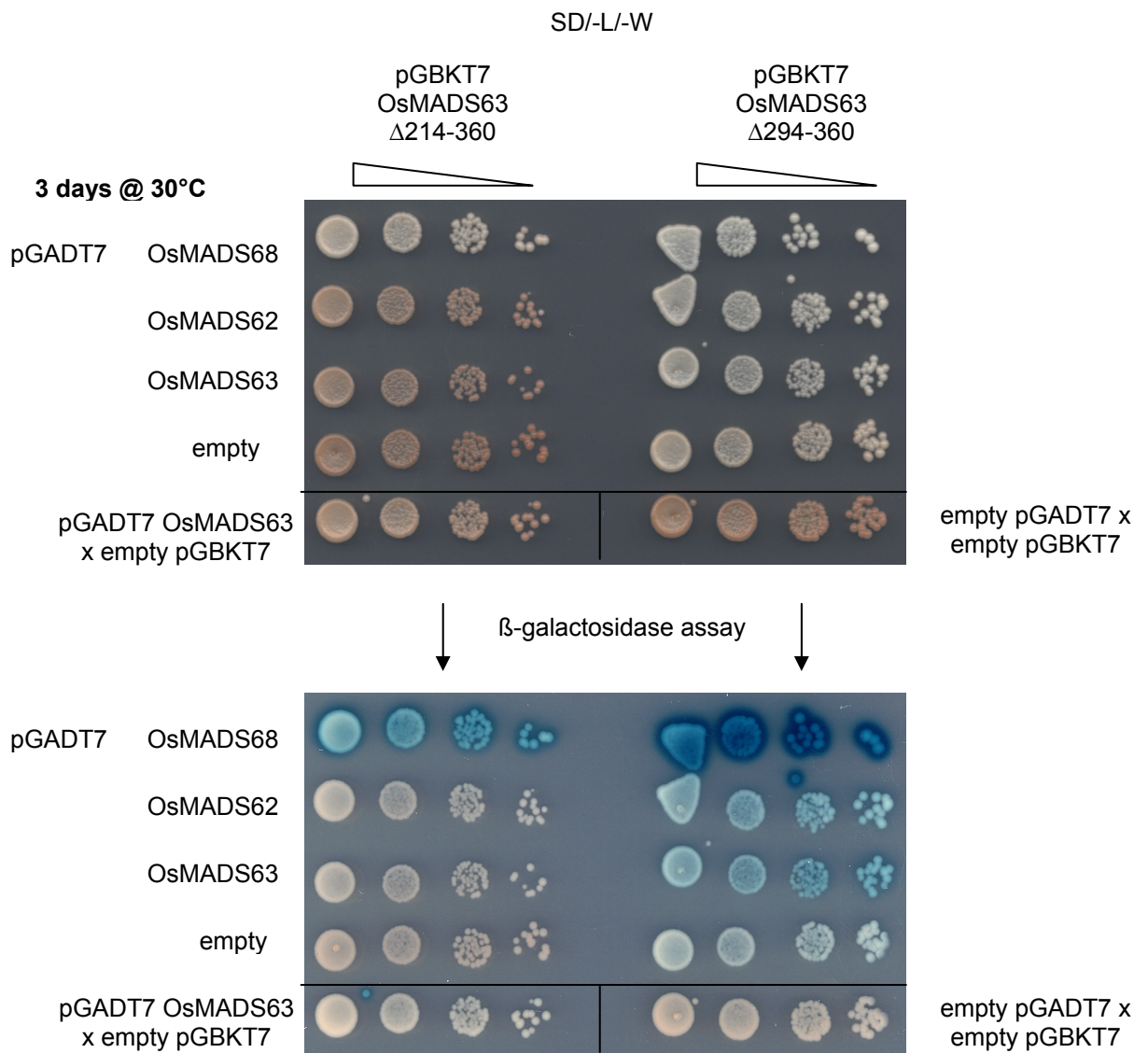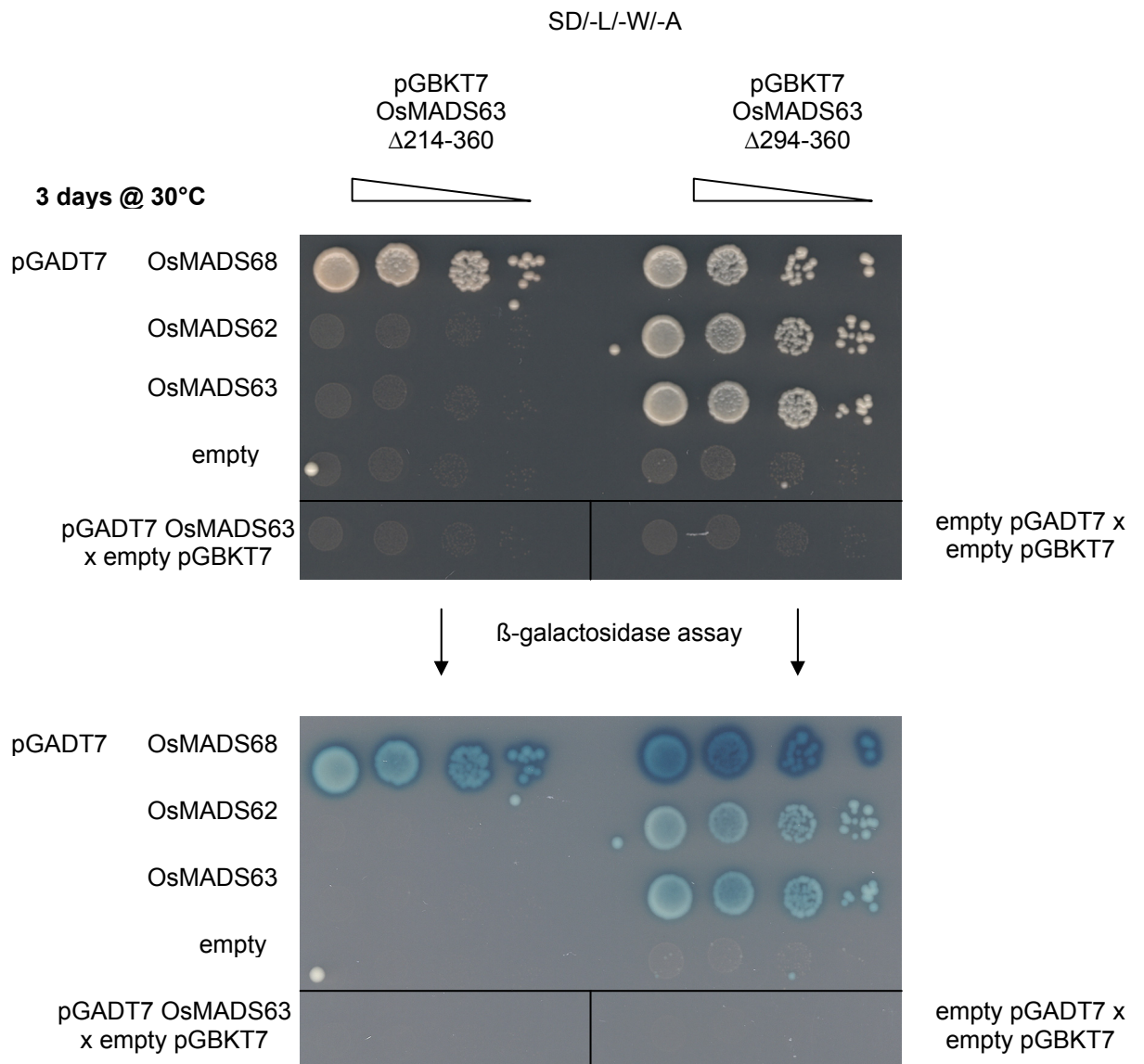
Figure S3. Yeast-2-hybrid interaction data of cotransformants containing pGBKT7 OsMADS68 constructs combined with all Oryza prey constructs plated on selective SD/-L/-W/-A medium. Cotransformants with the full-length OsMADS68 pGBKT7 vector that showed autoactivation are not shown.

Figure S4. Data on the growth of cotransformants containing pGBKT7 OsMADS62 constructs combined with all Oryza prey constructs plated on non-selective SD/-L/-W medium. Cotransformants with the full-length OsMADS62 pGBKT7 vector that showed autoactivation are not shown.

Figure S5. Yeast-2-hybrid interaction data of cotransformants containing pGBKT7 OsMADS62 constructs combined with all Oryza prey constructs plated on selective SD/-L/-W/-A medium. Cotransformants with the full-length OsMADS62 pGBKT7 vector that showed autoactivation are not shown.

Figure S6. Data on the growth of cotransformants containing pGBKT7 OsMADS63 constructs combined with all Oryza prey constructs plated on non-selective SD/-L/-W medium. Cotransformants with the full-length OsMADS63 pGBKT7 vector that showed autoactivation are not shown.

Figure S7. Yeast-2-hybrid interaction data of cotransformants containing pGBKT7 OsMADS63 constructs combined with all Oryza prey constructs plated on selective SD/-L/-W/-A medium. Cotransformants with the full-length OsMADS63 pGBKT7 vector that showed autoactivation are not shown.

Figure S8. Data on the growth of all Eschscholzia MIKC* protein cotransformants plated on non-selective SD/-L/-W medium.
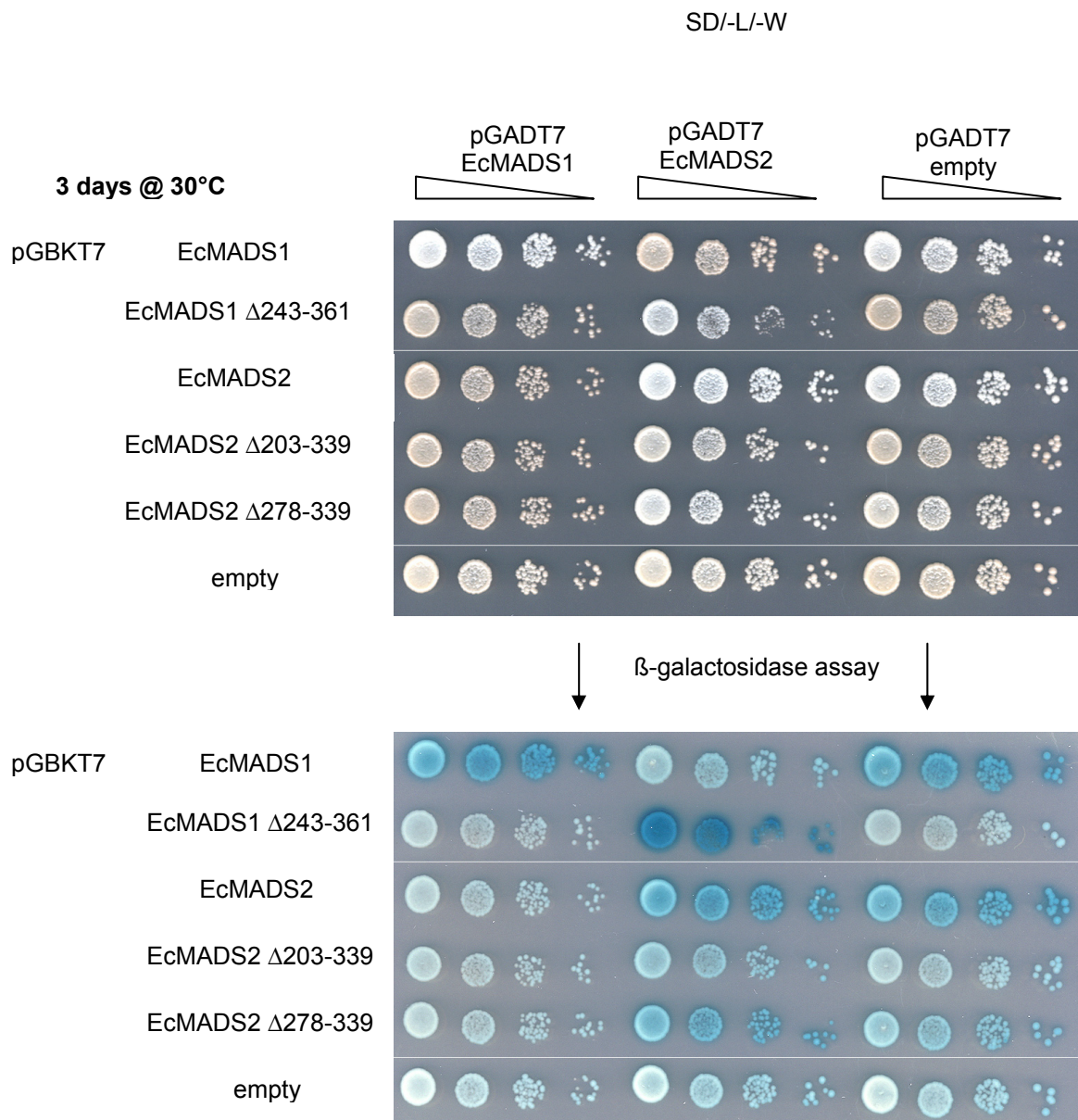
Figure S9. Yeast-2-hybrid interaction data of all Eschscholzia MIKC* protein cotransformants plated on selective SD/-L/-W/-A medium
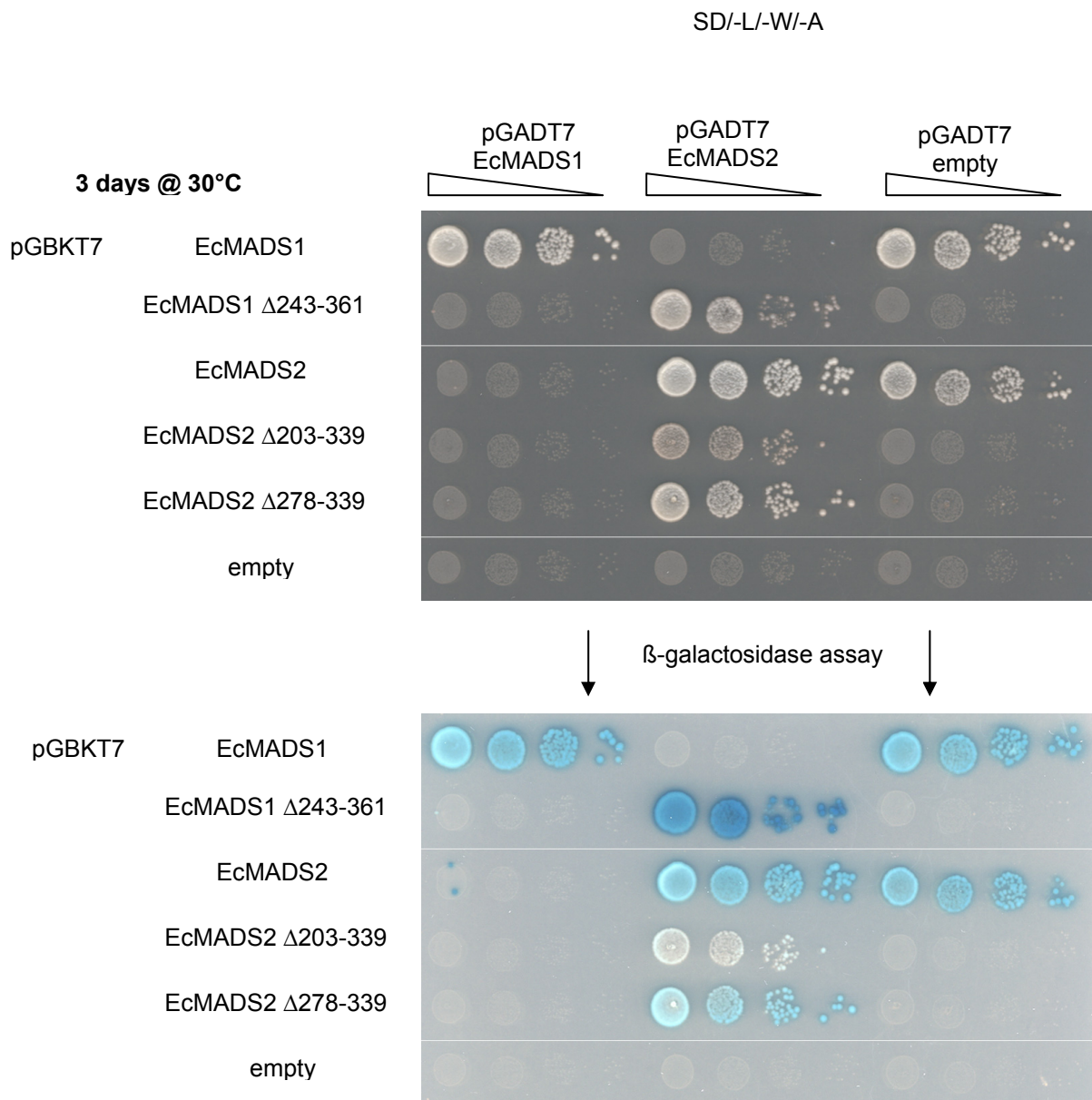
Eidesstattliche Erklärung

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.

Acknowledgement

I would like to thank Heinz Saedler and Thomas Muenster for their confidence, for offering me a position to work on my PhD project in their laboratory and greenhouses and the accompanying supervision. Furthermore, I thank Maarten Koornneef for willing to be my second supervisor and Jonathan C. Howard, Michael Melkonian and Wim Soppe for their kindness to join the defence committee.

Although many people were present during my job interview about three and a half years ago, I have the suspicion that the voices of Wim Verelst and Ralf Petri have spoken strongly in my benefit for which I am grateful. My appreciation goes also to you, Wim, for taking the effort of spreading the job advertisement and I also realize that my project could not have been prosperous without your work.

This PhD project would have been a lonely endeavour without the people in my lab, especially Simone and Ramzan who sat next to me, sharing the good times and the bad. Perhaps bad times would have even gained the upper hand if it had not been for you, Daniela, always there with a true heart and mind, to rescue the scientific spirit and, let's not forget, preventing -80°C freezers from unfreezing.

Only when experiments allowed, of course, I occasionally took a trip back home to the Netherlands. I know I am very fortunate with the safe haven that awaits me there.

While I am writing this I do not know yet where my future will unfold. I can say though that I only hope to meet again some of the people as I met here.

Curriculum vitae

Michiel Kwantes was born in Tilburg in the Netherlands on the 7th of September 1979. After growing up in Tilburg, he went to the University of Nijmegen to study biology. He received his first scientific training in the lab of evolutionary microbiology, led by Johannes Hackstein. There he studied the evolution of enzymes that are used by the ciliate *Nyctotherus ovalis* to produce energy (and hydrogen). The degree Master of Science was obtained in June 2003 after completing a second internship in Titi Mariani's lab of Plant Cell Biology, where he investigated the signalling pathway of the plant hormone ethylene in *Arabidopsis thaliana*. Positions for graduate students with emphasis on evolution are not plentiful. Fortunately, in 2006, he was offered a position at the Max Planck Institute for Plant Breeding Research in Cologne, in the group of Thomas Muenster, which belongs to the department of Molecular Plant Genetics, led by Heinz Saedler. On the 25th of June 2009 he successfully finished his task as PhD student by defending this manuscript.