About sisters and starlets:

B_{sister} and MIKC*-type MADS-box genes in rice development

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät

der Friedrich-Schiller-Universität Jena

von Diplom-Biologin Susanne Schilling

geboren am 5. März 1985 in Erfurt

Gutachter

1. Prof. Dr. Günter Theißen (Jena)
2. Prof. Dr. Ralf Oelmüller (Jena)
3. Prof. Dr. Marcel Quint (Halle)

Tag der öffentlichen Verteidigung

Dienstag, den 13. Dezember 2016

Table of contents

4

1. Introduction

Gene duplications provide the 'raw materials' that are utilized by natural selection in order to generate new genetic and eventually phenotypic traits (Conant and Wolfe 2008). Single genes are more often than not under a high negative selection pressure, thus unable to acquire new functions. The genetic 'backup' that is formed by gene duplications and usually not under the same selective pressure is therefore regarded crucial for the emergence of new genetic functions (Taylor and Raes 2004; Ohno 1970; He and Zhang 2005; Flagel and Wendel 2009). Three possible outcomes have been suggested after a gene duplication event: sub-, neo- and non-functionalization (Figure 1). In case of a sub-functionalization, both of the genes resulting from a duplication event remain in the genome while adopting different parts of the ancestral function (Force et al. 1999). Neo-functionalization occurs if one of the two copies acquires beneficial changes that result in a new function whilst the other gene copy maintains the ancestral function (Ohno 1970). In case of non-functionalization, one of the two genes retains its function whilst the other becomes pseudogenized and/or eventually gets completely lost (Figure 1) (Jacq et al. 1977).
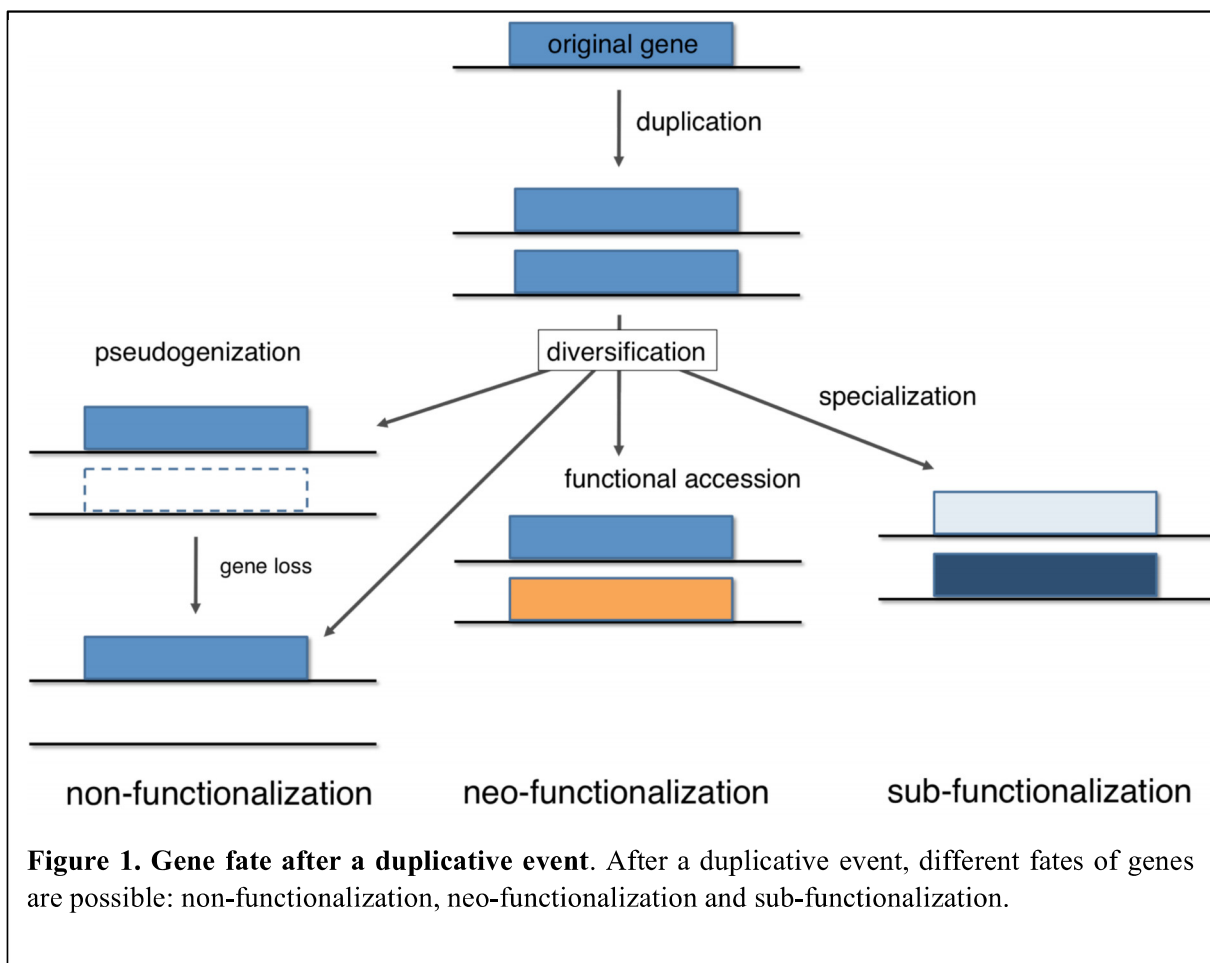

Figure 1. Gene fate after a duplicative event. After a duplicative event, different fates of genes are possible: non-functionalization, neo-functionalization and sub-functionalization.

1.1. The Innan-Kondrashov model for gene fate after duplication

Innan and Kondrashov (2010) classified the fates of duplicated genes in an even more detailed way. They categorize possible outcomes of gene duplications by the selection pressure that acts upon the newly arisen copy during the phase of fixation, hence in the time frame after the duplicative event. In category I they group the models of neo-functionalization and sub-functionalization, arguing that in all these models the duplicated genes get fixed within the population under neutral evolution (Innan and Kondrashov 2010). Duplications in this respective group result in copies that are identical and evolutionary neutral. Hence, in this case there is no selection pressure for keeping or losing the duplicated gene and often gene duplicates will be lost due to drift (non-functionalization). Fixation of the gene pair within the population can be an evolutionary neutral process, so that after the fixation within the population, the two genes acquire different mutations and subsequently become neo- or sub-functionalized (fate-determination). However, a mutation determining the fate of the gene (sub-, neo- or non-functionalization) can also arise before the fixation in the population. Hence fixation and fate determination phase can also be overlapping.

If the duplicative event itself does have an advantageous effect directly after the duplicative event, e.g. due to beneficial gene dosage increase or shielding of deleterious mutations by gene redundancy, positive selection applies during the fixation period. Innan and Kondrashov group these events into category II (Innan and Kondrashov 2010).

Models which propose that positive selection acts on the gene even before the duplicative event fall into category III. In this case many different alleles are maintained in a population as genetic variety, e.g. through balancing selection. In these cases the duplicated gene is often fixed immediately after the duplication, because it is beneficial to have diverse copies of the gene (Innan and Kondrashov 2010).

1.2. MADS-box genes are key regulators of plant development

MADS-box genes can be found in almost all eukaryotes (Gramzow et al. 2010). They encode MADS-domain transcription factors that control developmental processes through the regulation of gene expression (Krizek et al. 2005; Ng et al. 2001; Schwarz-Sommer et al. 1990). Likely, MADS-box genes evolved from a type II topoisomerase in a common ancestor of all eukaryotes (Gramzow et al. 2010). While animals and fungi usually possess only a few different MADS-box genes (Riechmann et al. 2000), plant MADS-box genes have undergone

several gene duplication events during evolution leading to 107 different MADS-box genes in *Arabidopsis thaliana* and 75 in *Oryza sativa* (Arora et al. 2007; Gramzow et al. 2010; Parenicova et al. 2003). Thus, plant MADS-box genes represent an excellent case to study the fate of duplicated genes.

During plant development, MADS-domain proteins confer a whole variety of different functions. They act during male and female gametophyte development as well as embryo and seed development (Bemer et al. 2008; Gramzow and Theissen 2010; Mizzotti et al. 2012; Verelst et al. 2007b). Further, MADS-box genes control flowering time, flower and fruit development and regulate the development of the root (Gramzow and Theissen 2010; Michaels and Amasino 1999; Scortecci et al. 2001). MADS-box genes can be categorized in type I or *SERUM-RESPONSE-FACTOR*-like (*SRF*-like) and type II or *MYOCYTE-ENHANCER-FACTOR2*-like (*MEF2*-like) MADS-box genes (Nam et al. 2004). Plant *MEF2*-like MADS-box genes are also known as MIKC-type MADS-box genes. This relates to their typical domain structure that comprises the highly conserved MADS-domain (M) which confers DNA-binding, the short and relatively diverse intervening (I) domain, the medium conserved keratin-like (K) domain responsible for forming dimers and higher order complexes and the highly variable C-terminal (C) domain (Cho et al. 1999; Fan et al. 1997; Riechmann and Meyerowitz 1997).

The most prominent and well-studied representatives of MIKC-type MADS-box genes are probably the floral homeotic genes that are determining floral organ identity according to the ABC model of flower development (Coen and Meyerowitz 1991; Weigel and Meyerowitz 1994). In this model (Figure 2), class A genes determine sepal development in the outermost whorl of the flower. A combinatory expression of class A and B genes leads to the development of petals in the second whorl, whilst class B and C expressed together will determine stamens in the third whorl. Class C genes alone specify the development of carpels in the fourth and innermost floral whorl (Weigel and Meyerowitz 1994). In an extension of the ABC model, class D genes determine the development of ovules (Angenent et al. 1995; Favaro et al. 2003; Pinyopich et al. 2003) and class E genes are obligatory expressed in all whorls as a kind of switch distinguishing between floral and non-floral tissue (Ditta et al. 2004; Pelaz et al. 2000; Theissen 2001).

Figure 2. The ABCDE-model. The expression of class A, B, C, D, E genes determines the identity of the organs in the respective whorl. Closely related B_{sister} and MIKC*-type genes are expressed in the fourth and third whorl, respectively, and are obligatory for proper carpel/seed and pollen development and maturation.
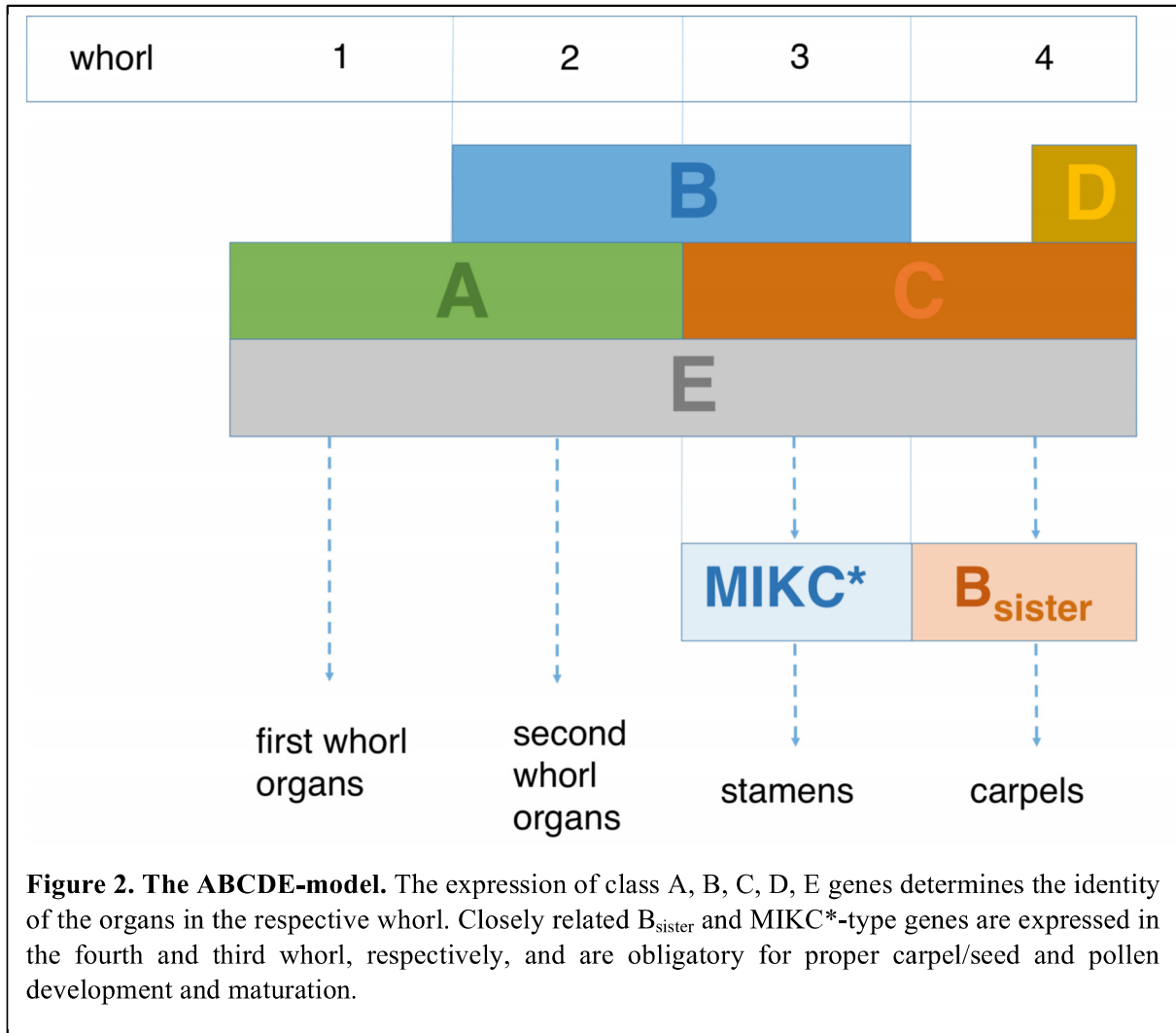
1.3. B_{sister} genes MADS-box play a key role in ovule and seed development

As mentioned above, class B genes, e.g. *APETALA3* and *PISTILLATA* from *A. thaliana,* determine the fate of second and third whirl organs (Goto and Meyerowitz 1994; Jack et al. 1992; Sommer et al. 1990; Tröbner et al. 1992). Their phylogenetically closely related sister group are the so-called B_{sister} genes (Becker et al. 2002). The founding member of this clade is *GNETUM GNEMON MADS13* (*GGM13*) from the gymnosperm *Gnetum gnemon. GGM13* is expressed predominantly in the developing female reproductive organs (Becker et al. 2002). This expression pattern, restricted to female reproductive organs as well as the developing seed, turned out to be a key pattern of B_{sister} gene expression in all major groups of seed plants (Figure 2) (Becker et al. 2002; de Folter et al. 2006; Nesi et al. 2002; Yamada et al. 2009).

Figure 3. Schematic representation of a mature ovule (left) and a developing seed (right). Maternal tissue is depicted in blue, filial tissue is depicted in green. The double fertilization of the haploid egg cell (1n) and the diploid central cell (2n) yields a diploid embryo and a triploid endosperm (3n). Antipods, integuments, synergids and nucellus degenerate during the process of seed development.
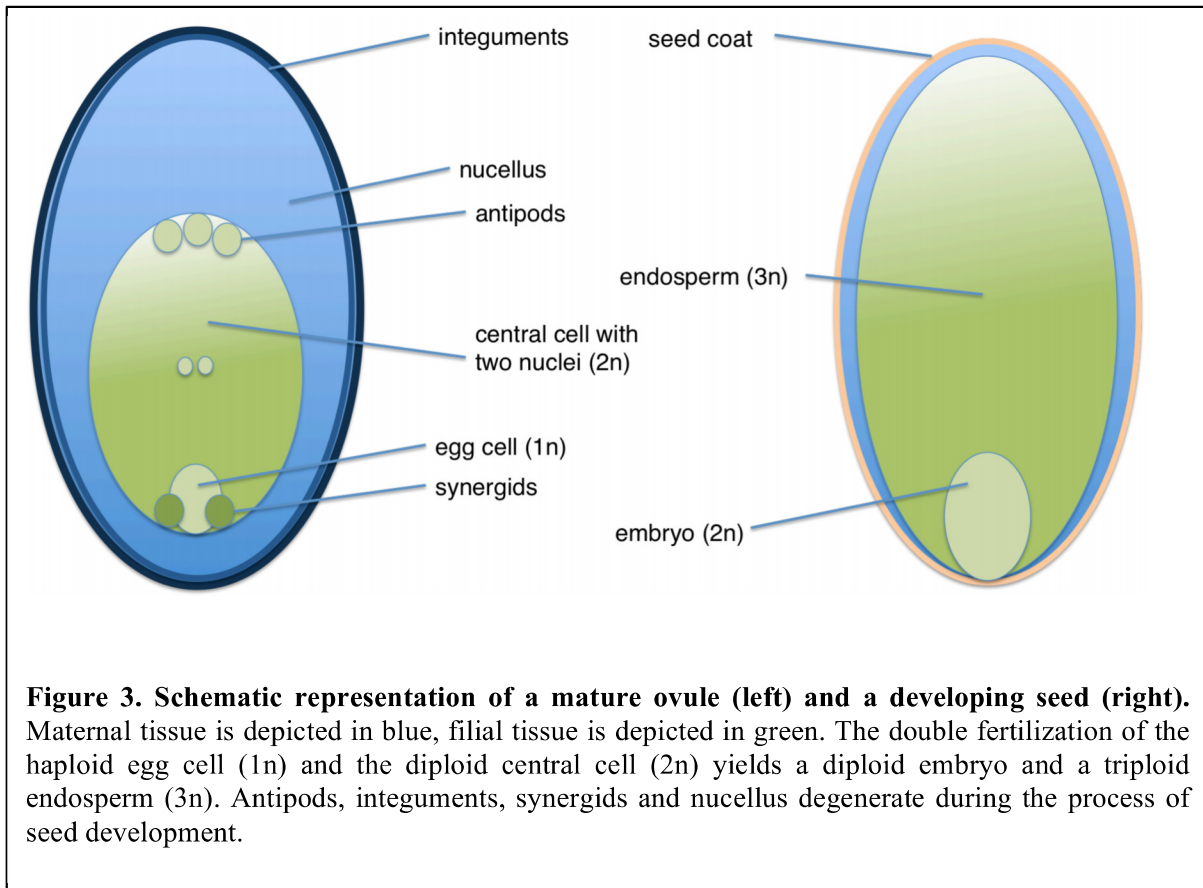
In *A. thaliana*, two B_{sister} genes have been isolated, *ARABIDOPSIS BSISTER* (*ABS*; also known as *TRANSPARENT TESTA* (*TT16*) or *AGAMOUS-LIKE32* (*AGL32*)) and *GORDITA* (*GOA*; also known as *AGL63*) (Erdmann et al. 2010; Kaufmann et al. 2005; Nesi et al. 2002; Prasad et al. 2011). *ABS* is expressed in the ovule, more precisely in the endothelial layer of the inner integuments (Figure 3) (Mizzotti et al. 2012). In line with this finding, *abs* mutant seeds show a lack in anthocyanidin synthesis, resulting in a yellowish colour in contrast to brown wild type seeds (Kaufmann et al. 2005; Nesi et al. 2002). Further, it was shown that *ABS* acts redundantly with the class D gene *SEEDSTICK* (*STK*) during ovule development; hence *abs/stk* double mutants produce a significantly lower number of seeds (Mizzotti et al. 2012).

GOA, the *ABS* paralog in *A. thaliana*, has a much broader expression pattern than *ABS* and most other B_{sister} genes. *GOA* expression is not restricted to the ovule and seed, but is also detected in stamens and petals (Erdmann et al. 2010; Prasad et al. 2010). The *goa* mutant phenotype is described to have shorter fruits than the wild type, hence *GOA* might regulate longitudinal growth of the fruits (Prasad et al. 2010). It is currently discussed whether GOA

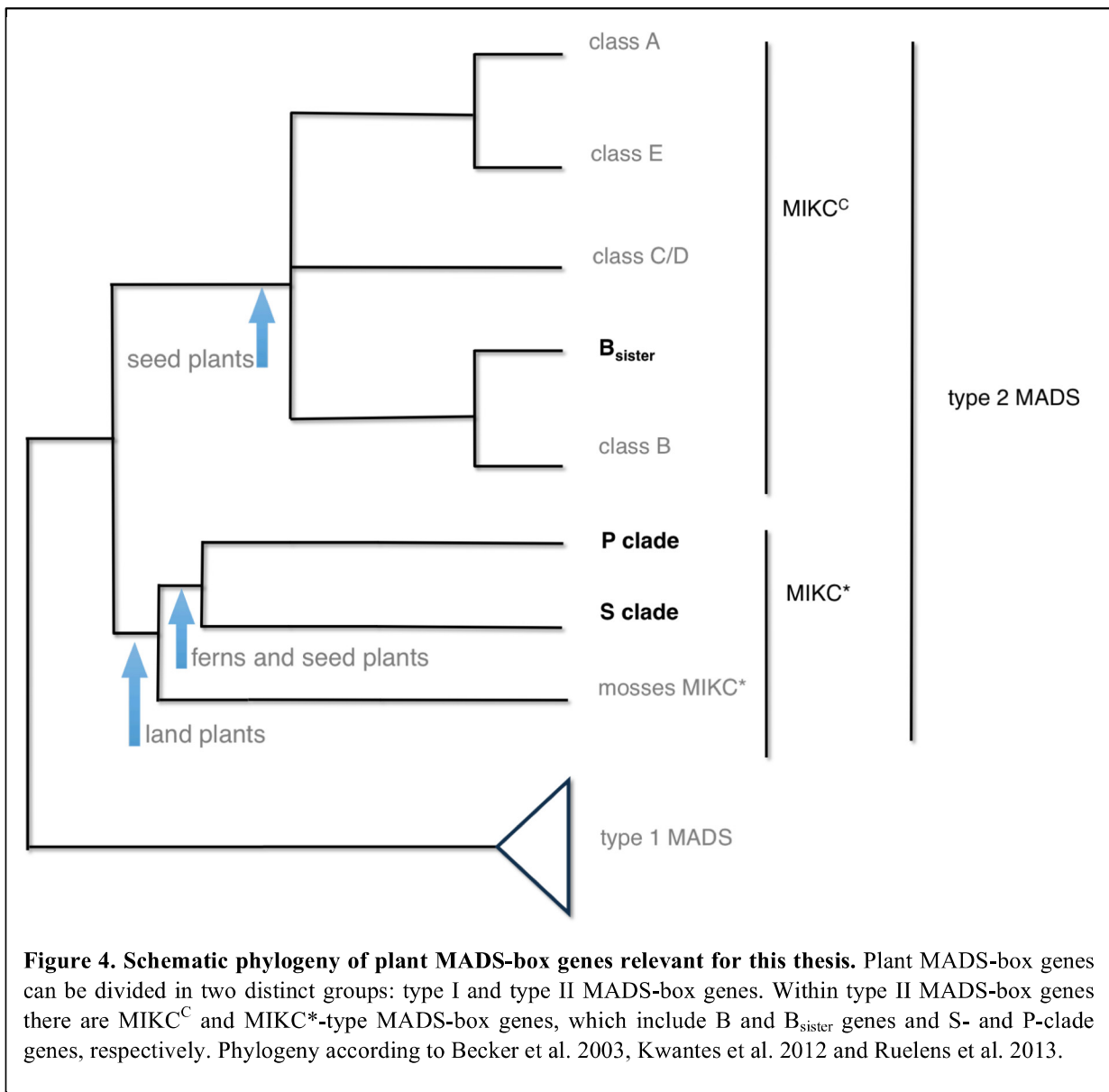underwent neo-functionalization during its evolution (Erdmann et al. 2010; Andrea Hoffmeier, personal communication).

The *Petunia* B_{sister} gene *FLORAL BINDING PROTEIN 24* (*FBP24*) is expressed like most of the B_{sister} genes in the developing and the mature ovule (de Folter et al. 2006). *fbp* knock-down co-suppression lines are affected in the development of the endothelial layer, similar to *abs* mutant plants, suggesting that *FBP24* might play a similar role as does *ABS* (de Folter et al. 2006). However, these findings are not uncontroversial since an *fbp* knock-out mutant did not show any phenotypic alterations (de Folter et al. 2006).

The expression patterns of monocot B_{sister} genes have been studied as well. The wheat (*Triticum aestivum*) B_{sister} gene *WHEAT BSISTER* (*WBsis*) has been shown to be expressed in the inner integument of the ovule, similar to *ABS*, but also in the outer integument and the nucellus (Figure 3) (Yamada et al. 2009). The maize B_{sister} gene *ZEA MAYS MADS17* (*ZMM17*) is expressed in the primordia of female reproductive organs, and in later stages of development its expression is restricted to the ovule and the silk (Becker et al. 2002). The three B_{sister} genes from rice (*Oryza sativa*) have not been studied in much detail, however large scale analyses showed that while *OsMADS29* was expressed similar to other B_{sister} genes, *OsMADS30* and *OsMADS31* have a much broader expression pattern (Arora et al. 2007). However, beyond the expression data, the functional significance of B_{sister} genes in monocots remained to be illuminated.

1.4. MIKC* -type MADS-box genes function in male gametophyte development

There are two types of MIKC-type MADS-box genes that are phylogenetically closely related: the classic MIKC-type (MIKC^C) genes (to which also class B and B_{sister} genes belong) and the MIKC*-type MADS-box genes (Figure 4) (Henschel et al. 2002). MIKC*-type MADS-box genes differ from the MIKC^C-type in that they encode for a longer I domain (Henschel et al. 2002). Both, MIKC^C- and MIKC*-type genes are found in representatives of all groups of extant land plants: bryophytes, lycophytes, ferns and seed plants. Therefore, the two types likely originated by a duplication of an ancient MIKC-type gene (Henschel et al. 2002; Kwantes et al. 2012). In all groups investigated so far MIKC*-type genes are predominantly expressed during the gametophytic stage, suggesting an ancestral role during the haploid phase of the life cycle (Gramzow et al. 2012; Henschel et al. 2002; Kofuji et al. 2003; Kwantes et al. 2012; Rensing et al. 2008; Riese et al. 2005; Zobell et al. 2010).

Accordingly, MIKC*-type MADS-box genes play a key role also in the development of the male gametophyte of *A. thaliana* (Figure 2) (Adamczyk and Fernandez 2009; Verelst et al. 2007a; Verelst et al. 2007b). *A. thaliana* possesses six different MIKC*-type genes, five of which are expressed during pollen development: *AGL30*, *AGL65*, *AGL66*, *AGL94* and *AGL104* (Honys and Twell 2004; Kofuji et al. 2003; Verelst et al. 2007a). They have shown to be regulating the process of pollen maturation (Verelst et al. 2007a, 2007b) probably by increasing the expression of genes acting in late pollen development and repressing early-stage genes at later stages of pollen development (Verelst et al. 2007b). No aberrant phenotype has been reported for an *A. thaliana* MIKC*-type gene single mutant, thus the genes act highly redundant (Verelst et al. 2007a; Adamczyk and Fernandez 2009).

Figure 4. Schematic phylogeny of plant MADS-box genes relevant for this thesis. Plant MADS-box genes can be divided in two distinct groups: type I and type II MADS-box genes. Within type II MADS-box genes there are MIKC^C and MIKC*-type MADS-box genes, which include B and B_sister genes and S- and P-clade genes, respectively. Phylogeny according to Becker et al. 2003, Kwantes et al. 2012 and Ruelens et al. 2013.

In extant seed plants and ferns two different lineages of MIKC*-type genes have been found, termed P-clade and S-clade (Gramzow et al. 2012; Kwantes et al. 2012; Nam et al. 2004). Those could not be identified in bryophytes and lycophytes and hence probably descended from a duplication of an ancestral MIKC*-type gene in the lineage leading to extant seed plants and ferns (Kwantes et al. 2012).

MIKC*-type protein interaction has been studied for representatives of all major groups of land plants (Kwantes et al. 2012; Verelst et al. 2007a; Zobell et al. 2010). *A. thaliana* S-clade proteins form obligate heterodimers with P-clade proteins (Verelst et al. 2007a). Similar findings were made for the early diverging eudicot *Eschscholzia californica* and the fern *Ceratopteris richardii*, with the addition that here also homodimerization and intraclade heterodimerization was observed (Kwantes et al. 2012). However, it was proposed that the major functional significance of MIKC*-type proteins in male gametophytes of euphyllophytes (ferns and seed plants) is mediated through interclade heterodimers and that this is the main cause o the high conservation of the P- and the S-clade over the course of about 400 million years of evolution (Kwantes et al. 2012). Homodimerization has been observed for the MIKC*-type protein MpMADS1 from the moss *Marchantia polymorpha*, suggesting that a common ancestor of P- and S-clade proteins has had the ability to form homodimers (Zobell et al. 2010). After the duplication event, interaction between P- and S-clade proteins evolved into obligate heterodimerization in the lineage leading to seed plants and eventually higher eudicots (Zobell et al. 2010).

In mosses, the number of MIKC*-type genes varies from only one gene in the liverwort *Marchantia polymorpha* to eleven in the moss *Funaria hygrometrica* (Zobell et al. 2010). Here, a clear phylogenetic pattern with conserved clades is not present, suggesting independent duplication events. Homo- as well as heterodimerization has been reported for MIKC*-type proteins of bryophytes and lycophytes (Kwantes et al. 2012; Zobell et al. 2010). However, already in these early diverging land plants, a tendency towards gametophytic expression is quite clear, suggesting an ancestral significance during this phase of land plant life cycle for MIKC*-type genes (Zobell et al. 2010).

MIKC*-type proteins do not only differ from MIKC^C-type proteins with respect to the length of their I domain, but also have distinct DNA-binding properties. MADS-domain proteins in general bind as dimers to specific DNA sequences, termed CArG-boxes (C-A-rich-G-boxes). Most MIKC^C-type MADS-domain proteins prefer CArG-boxes with the consensus 5'-CC(A/T)$_6$GG-3', so called SRE-type CArG-boxes (de Folter and Angenent 2006; Hayes et al.

1988; Riechmann et al. 1996). However, MIKC*-type MADS-domain proteins preferentially bind to a second type of CArG-box, termed N10-type, with the consensus 5'-CTA(A/T)$_4$TAG-3', or more relaxed 5'-C(A/T)$_8$G-3' (Pollock and Treisman 1991; Shore and Sharrocks 1995; Verelst et al. 2007a; Wu et al. 2011; Zobell et al. 2010). This binding behaviour is highly conserved amongst MIKC*-type proteins from extant land plants (Kwantes et al. 2012; Verelst et al. 2007a; Zobell et al. 2010).

1.5. Aims of this thesis

Gene duplications can be seen as a way to generate a genetic 'backup' - the copied gene - that afterwards has the potential to adapt new biological functions (Taylor and Raes 2004). The aim of this thesis is to provide a greater insight of the processes that occur during the evolution of pairs of genes that were generated by gene duplications and depicting the potentially very different fates of the duplicated genes. Since MADS-box genes are so highly diversified in the plant kingdom they are an excellent gene family to study for this purpose. *O. sativa* is a good model for this investigations, as it is a well studied monocot crop plant which allows for comparison with eukaryotic model plants such as *A. thaliana* over a great evolutionary distance. By focusing on rice B$_{sister}$ genes we were able to study the function two paralogous genes with evolutionary quite distinct fates. In addition, I studied the rice MIKC*-type MADS-box genes as an remarkable example for a highly conserved and functionally constrained subclade of MADS-box genes.

The thesis follows three distinct aims:

I: Illuminating the evolution and function of the B$_{sister}$ MADS-box gene *OsMADS30* from *O. sativa*. Phylogenetic relationships within the monocot B$_{sister}$ clade were analysed in detail and selection analyses were carried out to compare evolutionary pressures acting on the different clades of B$_{sister}$ genes in grasses. Furthermore, *OsMADS30* expression and gene structure as well as the *osmads30* mutant phenotype was investigated in order to compare the functional significance of *OsMADS30* to other members of the B$_{sister}$ clade. Also, the gene structure and expression of close relatives were investigated to gain insight into the evolutionary history of *OsMADS30* (manuscript I).

II: Clarifying the functional importance of the rice B$_{sister}$ gene *OsMADS29,* the paralog of *OsMADS30*. For this purpose, a meta-analysis of B$_{sister}$ gene expression patterns and the *OsMADS29* expression pattern was investigated in detail. My co-workers also investigated the mutant phenotype of *OsMADS29* knock-down plants and the phylogeny of B$_{sister}$ genes of the grass family (manuscript II).

III: Analysing how the MIKC*-type MADS-box genes *OsMADS62* and *OsMADS63* from the S-clade and *OsMADS68* from the P-clade evolved and determine their role in male reproductive organ development in rice. Therefore I analysed the DNA binding of OsMADS62, OsMADS63 and OsMADS68. My co-workers analysed the expression of all genes and studied single and double mutants and protein-protein interaction. Altogether we aimed to reveal the level of functional conservation between MIKC*-type MADS-box genes in flowering plants (manuscript III).

2. Manuscripts

2.1. Overview of the manuscripts

I. S. Schilling, L. Gramzow, D. Lobbes, A. Kirbis, L. Weilandt, A. Hoffmeier, A. Junker, K. Weigelt-Fischer, C. Kluklas, F. Wu, Z. Meng, T. Altmann and G. Theißen. (2015): Non-canonical structure, function and phylogeny of the B_{sister} MADS-box gene *OsMADS30* of rice (*Oryza sativa*). *Plant Journal*, Volume 84, Issue 6, 1059–1072.

In this manuscript the authors report on the evolution of the B_{sister} gene *OsMADS30* from rice and point out differences to previously described B_{sister} MADS-box genes. They present evidences that suggest that *OsMADS30* might have undergone neo-functionalization by the gain of a heterologous C-terminal domain.

Author contributions:

Susanne Schilling performed cloning, analysis of the gene structures as well as selection analysis, quantitative RT-PCR experiments and *in situ* hybridization, wrote the draft and prepared the figures of the manuscript. Lydia Gramzow prepared the phylogenetic tree and introduced Susanne Schilling to selection analysis. Andrea Hoffmeier and Lisa Weilandt performed preparative experiments. Alexander Kirbis performed quantitative RT-PCR and germination experiments. Astrid Junker, Kathleen Weigelt-Fischer, Christian Kluklas and Thomas Altmann perfomed and analyzed the imaging-based phenotyping. Zheng Meng and Feng Wu contributed semi-quantitative RT-PCR and GUS assay. Günter Theißen and Dajana Lobbes designed and supervised the project. All authors contributed in improving and revising manuscript.

Overall contribution of Susanne Schilling: 65 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Günter Theißen

II. X. Yang, F. Wu, X. Lin, X. Du, K. Chong, L. Gramzow, S. Schilling, A. Becker, G. Theißen and Z. Meng (2012): Live and Let Die - The B_{sister} MADS-Box Gene *OsMADS29* Controls the Degeneration of Cells in Maternal Tissues during Seed Development of Rice (*Oryza sativa*). *PLoS One*, **7/12**, e51435-e51435.

In this manuscript the authors report on the phylogeny and function of the rice B_{sister} gene *OsMADS29*. *OsMADS29* plays a key role for the proper development of the rice seed by regulating cell degeneration of maternal tissues. Further, the ancestral function of B_{sister} genes is investigated.

Author contributions:

Xuelian Yang and Feng Wu performed the expression analyses and RNAi knock-down of *OsMADS29*, analyzed the data and wrote the manuscript. Lydia Gramzow contributed the phylogenetic analysis. Susanne Schilling performed the comparative meta-analysis of B_{sister} gene expression and contributed in writing and improving the manuscript. Xuelei Lin and Xiaoqiu Du contributed reagents, materials and analysis tools. Kang Chong gave suggestions to the manuscript. Zheng Meng designed and supervised the project and wrote the manuscript. Annette Becker and Günter Theißen contributed by improving the manuscript.

Overall contribution of Susanne Schilling: 5 %

I hereby certify the accuracy of the statements on the contributions of the authors.

Günter Theißen

III. Y. Liu, S. Cui, F. Wu, S. Yan, X. Lin, X. Du, K. Chong, <u>S. Schilling</u>, G. Theißen and Z. Meng (2013): Functional Conservation of MIKC*-Type MADS Box Genes in *Arabidopsis* and Rice Pollen Maturation. *The Plant Cell*, **25**, 1288–1303.

In this manuscript the authors report on the three MIKC*-type MADS-box genes from rice. It is found that they are highly conserved in both, function as well as molecular interaction, compared to *A. thaliana* MIKC*-type MADS-box genes.

<u>Author contributions:</u>

Yuan Liu designed and performed the expression analyses and RNAi knock-down experiments and wrote the manuscript. Shaojie Cui and Shuo Yan assisted with the experiments. Xuelei Lin, Feng Wu, Xiaoqiu Du and Kang Chong contributed reagents, materials and analysis tools and suggestions to the manuscript. <u>Susanne Schilling</u> performed part of the EMSA experiments, wrote the introduction and contributed in improving the manuscript. Zheng Meng designed and supervised the project and wrote the manuscript. Günter Theißen contributed in writing the manuscript.

<u>Overall contribution of Susanne Schilling: 15 %</u>

I hereby certify the accuracy of the statements on the contributions of the authors.

Günter Theißen

2.2. Manuscript I

S. Schilling, L. Gramzow, D. Lobbes, A. Kirbis, L. Weilandt, A. Hoffmeier, A. Junker, K. Weigelt-Fischer, C. Kluklas, F. Wu, Z. Meng, T. Altmann and G. Theißen. (2015): **Non-canonical structure, function and phylogeny of the B$_{sister}$ MADS-box gene *OsMADS30* of rice (*Oryza sativa*).** *Plant Journal*, *84*(6), 1059-1072.

Non-canonical structure, function and phylogeny of the B$_{sister}$ MADS-box gene *OsMADS30* of rice (*Oryza sativa*)

Susanne Schilling[1], Lydia Gramzow[1], Dajana Lobbes[1], Alexander Kirbis[1], Lisa Weilandt[1], Andrea Hoffmeier[1], Astrid Junker[2], Kathleen Weigelt-Fischer[2], Christian Klukas[2], Feng Wu[3], Zheng Meng[3], Thomas Altmann[2] and Günter Theißen[1,*]

[1]*Department of Genetics, Friedrich Schiller University Jena, Jena D-07743, Germany,*
[2]*Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben D-06466, Germany, and*
[3]*Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*

SUMMARY

B$_{sister}$ MADS-box genes play key roles in female reproductive organ and seed development throughout seed plants. This view is supported by their high conservation in terms of sequence, expression and function. In grasses, there are three subclades of B$_{sister}$ genes: the *OsMADS29-*, the *OsMADS30-* and the *OsMADS31*-like genes. Here, we report on the evolution of the *OsMADS30*-like genes. Our analyses indicate that these genes evolved under relaxed purifying selection and are rather weakly expressed. *OsMADS30*, the representative of the *OsMADS30*-like genes from rice (*Oryza sativa*), shows strong sequence deviations in its 3′ region when compared to orthologues from other grass species. We show that this is due to a 2.4-kbp insertion, possibly of a hitherto unknown helitron, which confers a heterologous C-terminal domain to *OsMADS30*. This putative helitron is not present in the *OsMADS30* orthologues from closely related wild rice species, pointing to a relatively recent insertion event. Unlike other B$_{sister}$ mutants *O. sativa* plants carrying a T-DNA insertion in the *OsMADS30* gene do not show aberrant seed phenotypes, indicating that *OsMADS30* likely does not have a canonical 'B$_{sister}$ function'. However, imaging-based phenotyping of the T-DNA carrying plants revealed alterations in shoot size and architecture. We hypothesize that sequence deviations that accumulated during a period of relaxed selection in the gene lineage that led to *OsMADS30* and the alteration of the C-terminal domain might have been a precondition for a potential neo-functionalization of *OsMADS30* in *O. sativa*.

Keywords: *Oryza*, neo-functionalization, neutral evolution, helitron, domestication, imaging-based plant phenotyping.

INTRODUCTION

Gene duplication events are common in evolution and three major fates are distinguished for the resulting duplicated genes. In the first scenario, one of the two genes retains the ancestral function while the second one acquires deleterious mutations and therefore becomes a pseudogene (non-functionalization). Alternatively, one gene can retain the ancestral function while the other one acquires beneficial mutations which are positively selected for and result in a new function (neo-functionalization) (Ohno, 1970; Lynch and Conery, 2000). Finally, both genes can acquire partially destructive but complementary mutations so that both now 'share' the function that was originally conferred by the precursor gene (sub-functionalization) (Lynch and Conery, 2000).

MADS-box genes are suited to study sub-, neo- and non-functionalization because they underwent numerous duplications during seed plant evolution (Theissen *et al.*, 2000; Arora *et al.*, 2007; Gramzow and Theissen, 2010; Airoldi and Davies, 2012). One particularly interesting subfamily of MADS-box genes are B$_{sister}$ genes, which constitute the phylogenetic sister group of the class B (*DEFICIENS-* and *GLOBOSA*-like) floral homeotic genes (Becker *et al.*, 2002). Like other groups of MADS-box genes, B$_{sister}$ genes encode for proteins that comprise a MIKC structure, in which the MADS domain (M) precedes a shorter Intervening-domain (I), the Keratin-like domain (K) and the C-terminal domain (C). The C-terminal domain of B and B$_{sister}$ genes contains two characteristic motifs, a PI-derived motif

and a Paleo-AP3 motif (Becker *et al.*, 2002). B_{sister} genes have been proposed to be essential for ovule and seed development throughout all seed plants (Becker *et al.*, 2002; Nesi *et al.*, 2002; de Folter *et al.*, 2006; Yamada *et al.*, 2009; Yang *et al.*, 2012; Yin and Xue, 2012; Chen *et al.*, 2013a). Their sequences are highly conserved and the majority of the B_{sister} genes is expressed very specifically within female reproductive organs and developing seeds in gymnosperms, monocots and eudicots (Becker *et al.*, 2002; Nesi *et al.*, 2002; de Folter *et al.*, 2006; Yamada *et al.*, 2009; Yang *et al.*, 2012).

In line with their expression pattern, B_{sister} MADS-box genes play a central role in the development of female reproductive organs (Mizzotti *et al.*, 2012; Yin and Xue, 2012). The B_{sister} gene *ARABIDOPSIS BSISTER* (*ABS*) of *Arabidopsis thaliana* has been shown to regulate endothelium and endosperm formation by acting redundantly with the D class MADS-box gene *SEEDSTICK* (Mizzotti *et al.*, 2012).

The B_{sister} gene *OsMADS29* from cultivated rice, *Oryza sativa*, plays a key role in seed development: knock out and knock down of *OsMADS29* results in shriveled or even aborted seeds that lack starch granules almost completely (Yang *et al.*, 2012; Yin and Xue, 2012). Most probably *OsMADS29* regulates degeneration of maternal tissue like the ovular vascular trace, the integuments and nucellar projection and epidermis. This is conferred by programmed cell death, which is obligate for endosperm development and grain filling (Yang *et al.*, 2012; Yin and Xue, 2012). The OsMADS29 protein interacts with a variety of other seed-expressed MADS-domain proteins. Interactions with SEPALLATA-like, AG-like, AGL6-like, StMADS11-like and AP1-like proteins were detected amongst others (Nayar *et al.*, 2014). OsMADS29 has also been shown to form homodimers, and heterodimers with its closely related paralog OsMADS31 (Nayar *et al.*, 2014).

Beyond *OsMADS29*, the *O. sativa* genome encodes two other B_{sister} genes: *OsMADS30* and *OsMADS31*. In contrast to *OsMADS29*, information about the function and evolution of *OsMADS30* and *OsMADS31* remains scarce. Here, we investigate the B_{sister} gene *OsMADS30* and show that it encodes canonical MADS-, I- and K-domains, but a quite unusual C-terminal domain and that it is expressed not only in female reproductive organs but also in vegetative tissues. In *OsMADS30* T-DNA insertion lines no deviations in seed set and seed phenotype were detected. We present evidence indicating that the subclade of *OsMADS30* orthologues, including genes from maize (*Zea mays* ssp. *mays*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*) and other grass species, evolved under relaxed purifying selection. These results indicate that *OsMADS30* does not confer a function similar to previously described B_{sister} genes and hence is not a canonical B_{sister} gene. However, *OsMADS30* T-DNA insertion lines display a reduced shoot size (projected area) and altered architecture (fewer and

shorter tillers and leaves). We hypothesize that *OsMADS30* may have evolved a new function as compared to 'traditional' B_{sister} genes. A reduced selection pressure during the evolution of *OsMADS30* orthologues might have paved the way for a potential neo-functionalization of *OsMADS30*.

RESULTS

The subclades of *OsMADS29-*, *OsMADS30-* and *OsMADS31*-like genes evolved differentially in Poales

B_{sister} genes have been reported to form three subclades in monocots: the *OsMADS29-*, *OsMADS30-* and *OsMADS31*-like genes (Yang *et al.*, 2012). To gain a better understanding of their evolution, we used publicly available genome sequence data (Al-Dous *et al.*, 2011; D'Hont *et al.*, 2012; Al-Mssallem *et al.*, 2013; Chen *et al.*, 2013b; Jacquemin *et al.*, 2013; Singh *et al.*, 2013; Wang *et al.*, 2014) and surveyed RNA sequencing data from the 1kp project (http://onekp.com/) to identify monocot B_{sister} genes. Using these sequences we reconstructed a phylogenetic tree (Figures 1 and S1). In our phylogeny, the B_{sister} genes from grasses are monophyletic and form three large subclades, which correspond to the previously described *OsMADS29-*, *OsMADS30-* and *OsMADS31*-like genes. The B_{sister} genes from *Musa accuminata* are sister to all Poaceae B_{sister} genes in our phylogeny (Figure 1). Thus, the three subclades of *OsMADS29-*, *OsMADS30-* and *OsMADS31*-like genes originated after the lineage that led to extant Zingiberales had split-off from the lineage that led to Poales. Possibly, the three subclades originated by the two consecutive genome duplications that were predicted for the ancestor of grasses (Poaceae) followed by a loss of one of the four resulting genes (D'Hont *et al.*, 2012). The split of Panicoideae and Pooideae dates back approximately 45–60 million years ago (MYA) (International Brachypodium Initiative I, 2010; Zhang *et al.*, 2012), thus the three subclades are at least 45 million years old. Within the B_{sister} genes of grasses the *OsMADS29* subclade is sister to the *OsMADS30* subclade (Figure 1). Within the *OsMADS29* subclade, the gene phylogeny resembles the species phylogeny with all genes from the Pooideae being the sister clade to genes from the *Oryza* genus and the clade of these two being the sister clade to all genes from Panicoideae. Topology within both, the *OsMADS30* and the *OsMADS31* subclade also resembles the species phylogeny rather well (Kellogg, 1998, 2001).

Whilst the branch leading to the *OsMADS29* subclade and the branches within this subclade are short, indicating only few substitutions in the *OsMADS29*-like genes, the branch leading to the *OsMADS30* subclade as well as the branches within the subclade are rather long, pointing to a relatively high nucleotide substitution rate affecting the *OsMADS30*-like genes of the corresponding species (Fig-
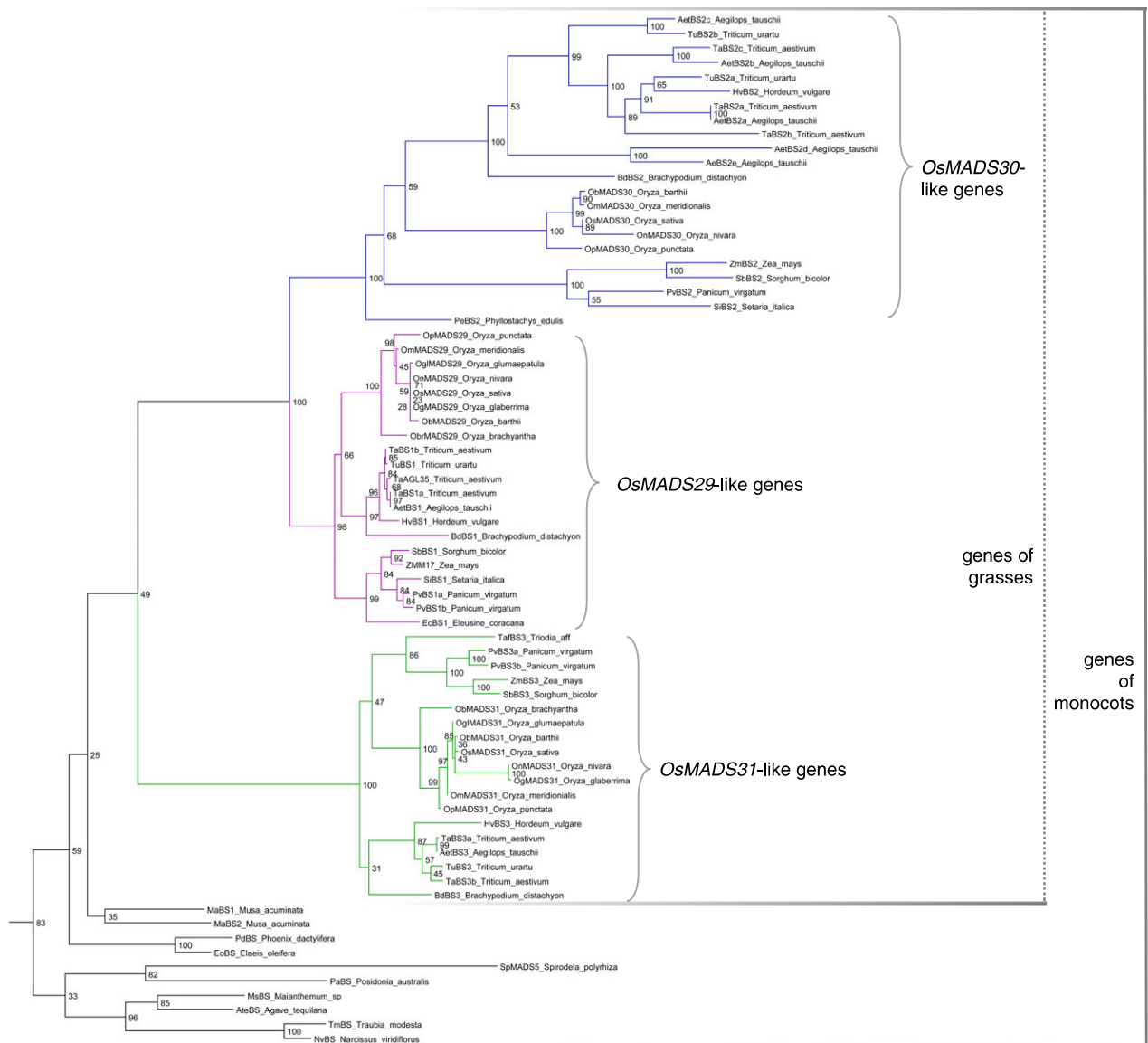
Figure 1. Phylogeny of monocot B$_{sister}$ genes. Maximum likelihood tree including all known monocot B$_{sister}$ genes with bootstrap values given on the nodes. *OsMADS29-*, *OsMADS30-* and *OsMADS31-*subclades are indicated by purple, blue and green lines, respectively. The tree is part of a phylogeny comprising angiosperm B$_{sister}$ genes where *GGM13* from the gymnosperm *Gnetum gnemon* served as the representative of the outgroup (Figure S1).

ure 1). This observation is substantiated in an alignment of amino acid sequences of grass B$_{sister}$ proteins, where OsMADS30 orthologues are much more heterogeneous than OsMADS29 and OsMADS31 orthologues even within the usually highly conserved MADS domain (Figure S2).

The *OsMADS30*-like genes evolved under relaxed purifying selection

Our phylogeny (Figure 1) and alignment (Figure S2) suggest variation in selection pressure acting on the three subclades of B$_{sister}$ genes in grasses. To further test this hypothesis, we analyzed the non-synonymous to synonymous substitution ratios ($\omega = dN/dS$) for the different sub-

clades. Different models were applied that allowed either one ω-value or different ω-values for two and four subclades, respectively (see Methods S1 for details). In all cases, models allowing different ω-values for different subclades of B$_{sister}$ genes fitted the data significantly better than the model assuming the same ω-value for all subclades ($P < 0.001$).

For models allowing different ω-values, the *OsMADS30* subclade has a higher ω-value than the *OsMADS29* and the *OsMADS31* subclade, indicating that less selection pressure acted on the *OsMADS30*-like genes in comparison to the genes of the *OsMADS29* and the *OsMADS31* subclades ($\omega_{30,MIKC} = 0.75$ versus $\omega_{other,MIKC} = 0.35$, two-ratio model;

Table 1). For the *OsMADS29* and *OsMADS31* subclade as well as the branch leading to the *OsMADS31* subclade ω-values decreased when the sequences encoding for the C-terminal domain were eliminated ($\omega_{other,MIK}$ = 0.20, two-ratio model; Table 1), which was expected since the C-domain of MIKC type MADS-domain proteins is known to evolve under more relaxed selection pressure than the MADS- and K-domains (Mondragon-Palomino *et al.*, 2009). When only the MADS box was used for the four ratio test, the ω-values of the *OsMADS29* and *OsMADS31* subclades decreased further ($\omega_{other,M}$ = 0.05, two-ratio model; Table 1). Intriguingly, ω-values calculated for the *OsMADS30* subclade remained similar no matter which parts of the sequences were used for calculations ($\omega_{30,MIKC}$ = 0.75, $\omega_{30,MIK}$ = 0.69, $\omega_{30,MADS}$ = 0.75, $\omega_{30,K}$ = 0.58, two-ratio model; Table 1). Similar values have been found for some pseudogenes (Wang *et al.*, 2012), thus they may indicate that the entire sequence of the *OsMADS30*-like genes including the MADS box has evolved under relaxed, almost neutral, selection pressure. However, at least outside of the *Oryza* lineage we did not find common features of pseudogenization, like translational frame shift mutations or premature stop codons amongst *OsMADS30*-like genes (Figures S2 and 2).

The OsMADS30 protein has an aberrant C-terminal domain caused by the insertion of a mobile element

The C-terminal sequence of OsMADS30 is different from that of its orthologues and paralogues from other grass species (starting at alignment position 188 in Figure S2). A database search using the BLAST algorithm (Altschul *et al.*, 1990) revealed that the entire part of the gene which is encoding for this deviant C-terminal domain as well as the *OsMADS30* 3'UTR and the downstream DNA sequence can be found three times in the genome of *O. sativa* ssp. *japonica*: within *OsMADS30* on chromosome 6, at another position on chromosome 6 and on chromosome 3 (Figure 2a). Due to its putatively 'mobile' character we termed the region *Oryza sativa* Mobile Element (*OsME*). *OsME* has a total size of 2364 bp with a similarity of 97–99% when the

different loci are compared to each other. We termed the *OsME* copies *OsME1* (at the *OsMADS30* locus), *OsME2* (the second locus on chromosome 6) and *OsME3* (on chromosome 3).

By BLAST searches we found, that the protein sequence encoded by the *OsME* region of the *OsMADS30* mRNA is not homologous to any known protein family. However, OsMADS30 had previously been predicted to possess a transmembrane (TM) motif, which is very unusual for MADS-domain transcription factors (Kim *et al.*, 2010). Remarkably, this putative TM motif is encoded by the *OsME* part of the *OsMADS30* mRNA (Figure S3) as predicted by the 'DAS' transmembrane prediction server (Cserzo *et al.*, 1997).

To estimate the time point of the *OsME* insertion at the *OsMADS30* locus we compared *OsMADS30* to its closely related orthologues from different wild rice species (Figure 2b). We found that in every species an *OsMADS30* orthologue was present; however none of them possessed the *OsME* insertion (Figure 2c). Instead, they all encode for a conventional C-terminal domain which is similar to those from OsMADS30 orthologues in non-rice species (Figure S2). We confirmed this by sequencing the *OsMADS30* orthologue of *O. rufipogon* W360 (henceforth called *OrMADS30*). However, we found that the integrity of *OsMADS30* orthologues is also disturbed in other rice species: in *O. meridionalis* and *O. glaberrima*, the *OsMADS30* orthologue is potentially impaired by a transposon insertion into the second intron and in *O. brachyantha* we found the entire I- and K-box deleted (Figure 2b). Similar rearrangements have neither been found in *OsMADS29*- nor in *OsMADS31*-like genes of the *Oryza* species.

Even though an *OsME* was not inserted in the *OsMADS30* orthologues in wild rice species, *OsME* was detected at other loci in the genomes of these species. Using a BLAST search we detected a copy of *OsME* homologues in all of the investigated wild rice species, except for *O. brachyantha* and *O. punctata*, which are more distantly related to *O. sativa* than the remaining species (Ge *et al.*, 1999). *OsMEs* in other *Oryza* species were localized at positions homologous to the positions of both, *OsME2* and *OsME3* in *O. sativa* (Figure 2c).

The OsME resembles both helitrons and segmental duplications

Remarkably, cDNAs mapping to *OsME2* and *OsME3* have been detected, indicating that the two other *OsME* copies remote from *OsMADS30* are being expressed as well (Rice Full-Length cDNA Consortium 2003). Different splice sites are recognized at these loci despite of their high sequence similarity (Figure 2a). Using 5' and 3' RACE we confirmed the published cDNAs from *OsME2* and *OsME3* and also found a transcript from *OsME1* which shares its 3' sequence with the *OsMADS30* mRNA and has a sequence

Table 1 Results of selection analysis. ω-values are given for one, two or four ratio tests and different parts of the sequences used

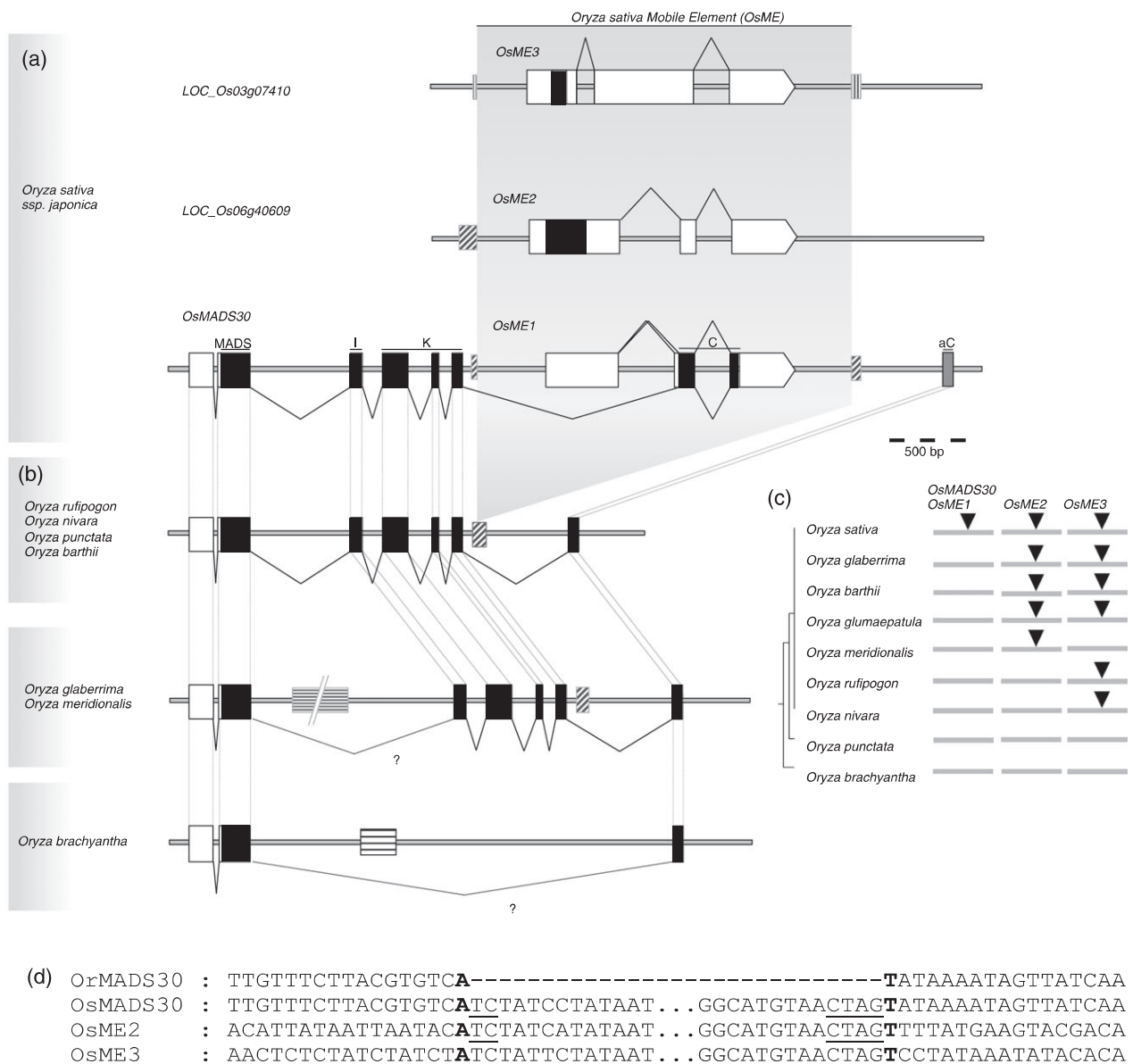| Test | Clade | Sequence | | | |
		MIKC	MIK	MADS	K
One ratio		0.5	0.38	0.25	0.40
Two ratio	*OsMADS30*	0.75	0.69	0.75	0.58
	Other	0.35	0.20	0.06	0.26
Four ratio	*OsMADS30*	0.75	0.69	0.75	0.58
	Branch to *OsMADS31*	0.57	0.45	0.15	0.59
	OsMADS31	0.45	0.16	0.05	0.19
	OsMADS29	0.25	0.22	0.05	0.28

Figure 2. Genomic structure of the region of *OsMADS30*, *OsME2* and *OsME3* and homologous sequences in *Oryza* spp.

(a) Gene structure of *OsME* and *OsMADS30*. *OsMADS30* has a typical MIKC structure with the *OsME1* inserted downstream of the K-box. Two more *OsME* loci, *OsME2* and *OsME3*, are contained in the *Oryza sativa* genome. The *OsME* is highlighted by a grey background. The ancient C-terminal domain coding region (aC) of *OsMADS30* is downstream of *OsME1*.

(b) Gene structures of the *OsMADS30* orthologues from wild rice species. In (a) and (b) the following sequence features are drawn as boxes: coding sequences are filled black, UTRs filled white, ancestral C (aC) filled dark grey, transposable elements are in grey boxes and filled with vertical stripes (SINE), diagonal stripes (MITE) and horizontal stripes (other transposable elements). Introns and intergenic regions are depicted by grey lines.

(c) *OsME* loci in different rice species. Phylogenetic relationship between rice species is given on the left, different loci and their orthologues are shown on the right. Insertion of *OsME* is indicated by a black triangle.

(d) Alignment of the insertion sites of all *OsME* and the homologue region in *OrMADS30*. ApT insertion site is indicated in bold. Helitron consensus sequences are underlined.

partially overlapping the *OsME2/OsME3* transcripts (Figures 2a and S4). However, predicted protein sequences of all of the *OsME* transcripts do not have similarity to known proteins, including transposases or any other proteins known to be encoded by transposons. Further, the predicted open reading frames are rather short (88 and 23 amino acids for *OsME2* and *OsME3*, respectively). However, short similarities with DNA sequences of transposons could be identified within *OsME* (Figure S5a).

Dot plot analysis showed that *OsME1* does not have a sequence resembling any type of a repeat characteristic for transposons, such as long terminal repeats (LTRs), charac-

teristic for LTR retrotransposons, or terminal inverted repeats (TIRs), typical for DNA transposable elements (Figure S5b,c). Furthermore, the *OsME* copies do not contain a PolyA stretch, which is a typical feature of non-LTR retrotransposons.

However, the *OsME* has certain characteristics of a helitron, DNA transposons that replicate by a rolling-circle mechanism (Kapitonov and Jurka, 2001). Comparing the *OsMADS30*, *OsME2*, *OsME3* and *OrMADS30* genomic DNA sequences, we found that *OsME* very likely was inserted without target site duplication within an ApT-site (Figure 2d), which is one common feature of helitrons (for review see (Kapitonov and Jurka, 2007)). *OsME* also has the common 5′-TC and 3′-CTRR termini (where R stands for A or G). Furthermore, *OsME* shares high sequence similarities with a number of different loci in the *Oryza* genome over several parts of the sequence, suggesting that it 'captured' foreign DNA, which is common for helitrons (Kapitonov and Jurka, 2007). However, unlike *OsME* most helitrons encode for a palindromic sequence close to the 3′ end, which forms a stem–loop structure during transposition and acts sterically as a termination signal for the transposition (Kapitonov and Jurka, 2007).

The composite structure of *OsME* might also originate from segmental duplications (for review see (Bailey and Eichler, 2006)). For humans, it has been reported that segmental duplications preferably insert into *Alu* elements, a subgroup of short interspersed nuclear elements (SINEs) (Zhou and Mishra, 2005). *OsME1* at the *OsMADS30* site is inserted into a miniature inverted-repeat transposable element (MITE) (Figure 2a). Like SINEs, MITEs are small non-autonomous transposable elements (Jiang *et al.*, 2004). This MITE is also present in *OrMADS30*, suggesting its insertion prior to the *OsME1* (Figure 2b). *OsME2* is inserted right next to a MITE while *OsME3* is inserted within a SINE element (Figure 2a).

OsMADS30 does not show the typical expression pattern of B$_{sister}$ genes

B$_{sister}$ genes generally have a very narrow expression domain restricted to developing flowers and seeds (Becker *et al.*, 2002; de Folter *et al.*, 2006; Yamada *et al.*, 2009; Yang *et al.*, 2012). Using RT-PCR we found comparatively weak signals of *OsMADS30* in seedlings, roots, developing panicles and developing seeds (Figure 3a). RT-PCR was also carried out for different tissues within the mature flower, showing rather weak signals in palea, lemma and the lodicules, but stronger signals in the pistil (Figure 3a). Promoter–GUS fusion lines and *in situ* hybridization experiments confirmed this broad expression pattern, showing signals in roots and leaves (tested for GUS only), in palea, lemma and stamina of developing flowers, in the ovary wall, the nucellus and the embryo sac of mature flowers and the pericarp of the developing caryopses (Figures 3b–

h and 4a–c,k,l). Remarkably, *in situ* hybridization also showed signals when a 'sense probe' with the sequence of the *OsMADS30* mRNA was used (Figure 4f–h,n,o). Signals obtained by utilizing the 'sense probe' were matching the signals generated with the 'antisense probe'. We hypothesized that a piece of RNA antisense to the *OsMADS30* mRNA caused these signals. To clarify the nature of the antisense RNA (asRNA), a primer was used to specifically reverse transcribe the asRNA. We amplified and cloned small fragments complementary to the *OsMADS30* mRNA (Figure S6). However, we did not get an asRNA as long as the *OsMADS30* mRNA. Intriguingly, we found the exon-exon borders of the cloned asRNA fragments to be exactly the same as for the *OsMADS30* sense mRNA, suggesting that the *OsMADS30* asRNA fragments are the product of an RNA-dependent RNA polymerase (RDR) rather than generated by antisense transcription from the *OsMADS30* locus.

Using quantitative RT-PCR (qRT-PCR) we found that in tissues that exhibited expression of the *OsMADS30* sense mRNA, an amount of asRNA in the same order of magnitude was present (Figure 3i).

The expression of *OrMADS30* from *O. rufipogon* W360 was also investigated with qRT-PCR (Figure 3j). We found that *OrMADS30* was much lower expressed than *OsMADS30*. In contrast to *OsMADS30*, expression was predominantly found in developing flowers and caryopses (Figure 3j). An *OrMADS30* asRNA could also be detected (Figure 3j). *In situ* hybridization showed that during reproductive development *OrMADS30* was expressed throughout flower and caryopsis development (Figure 4d,e,m) in roughly the same temporal and spatial pattern as *OsMADS30* in *O. sativa*. A signal from a putative asRNA was detected with this method as well (Figure 4i,j,p).

Surveying published expression patterns of grass B$_{sister}$ genes, we found that *OsMADS30* orthologues in general are expressed to a lower extent than *OsMADS29* orthologues (Figure S7 and Table S1). However, expression of *OsMADS30* in vegetative tissue seems to be unique amongst its orthologues.

OsMADS30 T-DNA insertion lines show alterations in plant size and architecture

To assess the function of *OsMADS30* based on a mutant phenotype, we investigated T-DNA insertion lines, which we named *osmads30-1* and *osmads30-2*. The lines possess T-DNA insertions in the 6th intron (*osmads30-1*) and the 4th exon of *OsMADS30* (*osmads30-2*), respectively (Figure 5a). For *osmads30-1* we could detect a chimeric mRNA of *OsMADS30-MIK* fused to green fluorescent protein (GFP)) mRNA, as GFP is encoded by the T-DNA (Figure S8a). For *osmads30-2* no cDNA could be amplified using a primer pair surrounding the T-DNA insertion site;

Figure 3. Expression of *OsMADS30* and *OrMADS30*.

(a) RT-PCR of *OsMADS30* for different tissues, developmental stages and floral organs. *OsMADS30* cDNA was amplified for 35 (35c) and 40 cycles (40c), respectively. Amplifications of *APT1* and *ACTIN1* cDNA for 25 cycles served as loading controls.

(b–h) GUS assay for *OsMADS30* promoter, scale bars 1 mm. (b) Seedling roots 7 days after germination. (c) Mature leaf. (d) Panicles at about 5 cm in size. (e) Developing floret. (f) Third and forth whorls of the florets; c, carpel; s, stamen. (g) Developing caryopses at 5 days after pollination. (h) Developing caryopses at 8 days after pollination.

(i–j) qRT-PCR of *OsMADS30* (i) and *OrMADS30* (j) in different tissues and developmental stages. Relative expressions are given for the *OsMADS30* mRNA and the asRNA in black and white bars, respectively, standard deviation is indicated by lines. Statistical differences between leaf tissue and all other tissues (i) or the respective tissue from *O. rufipogon* (j) are marked with asterisks [one-factorial analysis of variance (ANOVA)] and plus signs (Mann–Whitney *U*-test). L, leaf; SD, 7-day-old seedling; R, root; DF, developing flower; MF, mature flower; DC1, developing caryopses 2–5 days after pollination; DC2, developing caryopses 7–10 days after pollination.
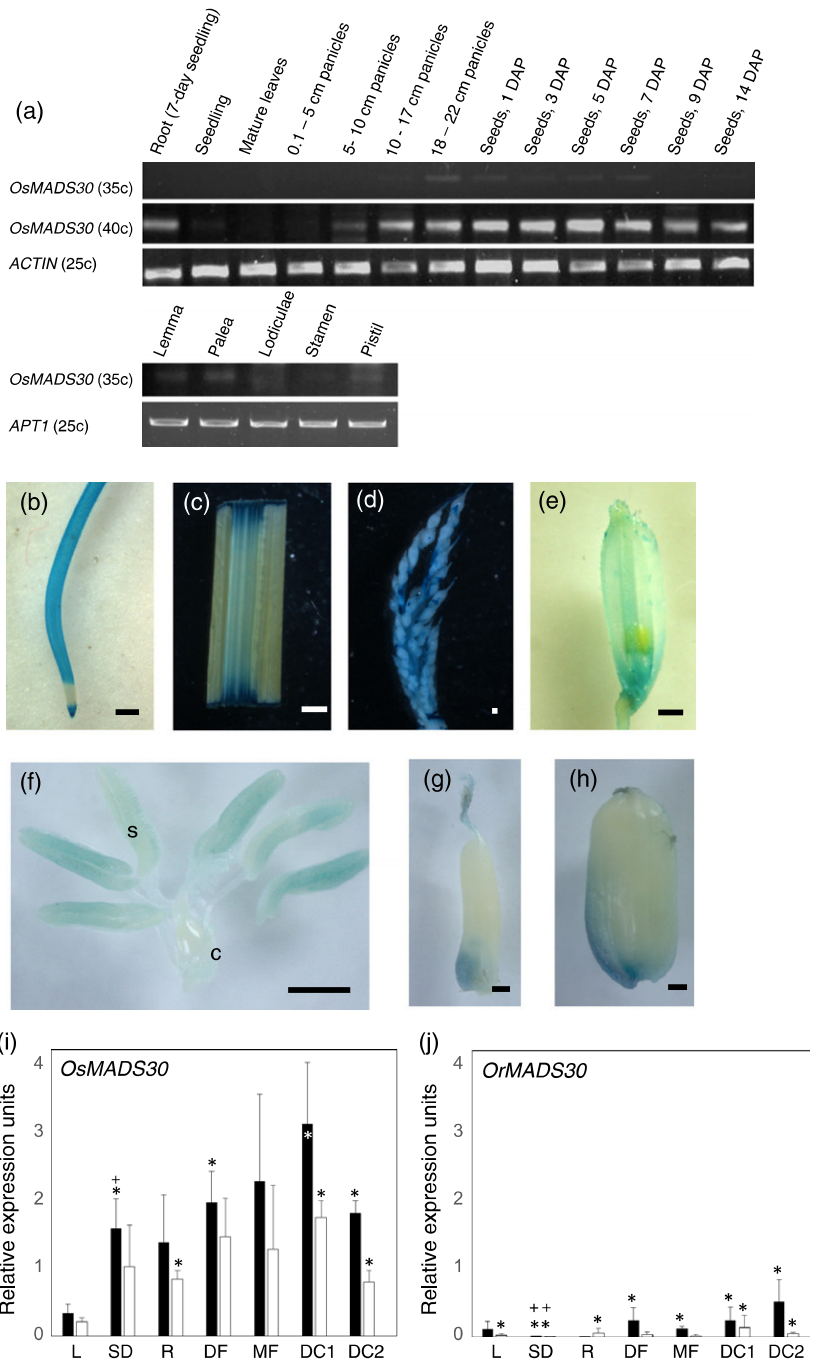
however an amplicon with primers upstream of the insertion site was detectable, pointing to a truncated transcript (Figure S8b). T_3 generation plants from both T-DNA insertion lines and wild type *O. sativa* ssp. *japonica* control plants were grown in a greenhouse to monitor their phenotypes. Mendelian T-DNA segregation observed in progeny of heterozygous plants of both lines suggests that gametophytic and sporophytic heredity transmissions were not impaired (Table 2). When compared to the wild type, no difference in the appearance of flowers and the mature ovules (Figure S9) as well as in the seed set could be observed (with 108 and 101% of the seed set in the mutant line compared with respective wild type sibling for *osmads30-1* and *osmads30-2*; Figure S10a). Furthermore, grains of both mutant lines did not show any differences regarding grain length, width and weight as compared with wild type *O. sativa* (Figure S10b–d). Also, no difference between wild type and both of the T-DNA insertion lines could be observed regarding the germination rate of the grains (Figure S11).
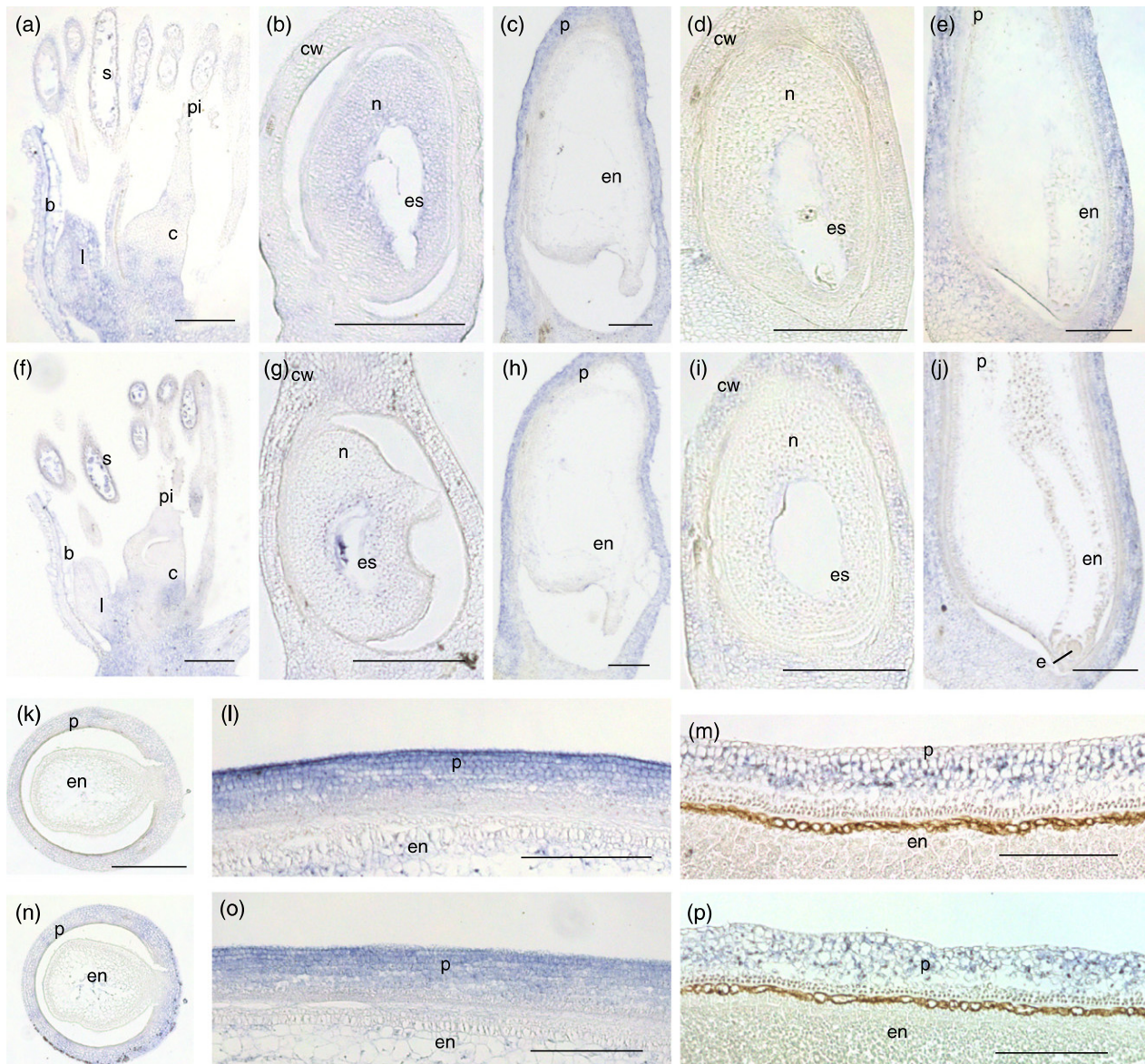
Figure 4. *OsMADS30* and *OrMADS30 in situ* hybridization.
Expression pattern of the sense mRNA of *OsMADS30* (a–c, k, l) and of the *O. rufipogon* orthologue *OrMADS30* sense mRNA (d, e, m) is shown in comparison with the expression pattern of the antisense RNAs complementary to *OsMADS30* (f–h, n, o) and *OrMADS30* (i, j, p). Sections from the following developmental stages were used for hybridization: developing flower (a, f), mature ovule (b, d, g, i), developing caryopsis at 2 DAP (c, h), a transverse section of a developing caryopsis at 3 DAP (k, n), a developing caryopsis at 4 DAP (e, j) and the pericarp of a developing caryopsis at 7 DAP (l, o) and at 10 DAP (m, p). Scale bars = 200 μm. b, bract; c, carpel; cw, carpel wall; e, embryo; en, endosperm es, embryo sac; l, lodicule; n, nucellus; p, pericarp; pi, pistil; s, stamen.

As we could not detect any aberrant flower or grain phenotype in either of the T-DNA lines, we extended our analyses by applying an imaging-based plant phenotyping approach. Figure 5(b–d) shows a panel of visible light side view images of representative wild type and mutant plants at day 131.

Intriguingly, *osmads30-1* and *osmads30*-2 plants showed significantly smaller projected shoot area (mainly representing leaf area) when compared with the wild type of *O. sativa* ssp. *japonica* (Figure 5e). This was observed throughout the whole time of growth monitoring (day 117-147), although *osmads30-1* displayed a more severe reduction in projected shoot area compared with *osmads30-2*, which was statistically significant during late vegetative development (Tables S3 and S4 after day 140). Relative growth rates of both mutants were similar to that of wild type plants except for the last days of phenotypic analysis (Figure S12a). After day 140, plants of both

mutant lines grew significantly faster with *osmads30-2* plants showing a further increase in growth rate after day 143. Especially *osmads30-1* mutants also displayed alterations in plant architecture. The side view convex hull area of both mutants is substantially and significantly lower than that of the wild type, probably due to a more erect growth habit (Figure 5f). However, the two mutant lines also differed to some extant from each other, e.g. plant height (Figure S12b,c and Tables S3 and S4).

DISCUSSION

OsMADS30 does not confer a canonical B$_{sister}$ function

In this work we analyzed the phylogeny and signatures of selection of B$_{sister}$ genes within the grass lineage (Poaceae) in order to illuminate their evolutionary history.

The longer branch lengths (Figure 1) as well as the higher ω-values that resemble those of pseudogenes (Wang *et al.*, 2012) within the *OsMADS30* subclade (Table 1) compared with the two other B$_{sister}$ subclades indicate that purifying selection has been considerably relaxed for *OsMADS30*-like genes. This is in line with the lower pairwise identity within the usually highly conserved MADS domain and the degeneration of the C-terminal Paleo-AP3 and PI-derived motifs (Figure S2). An explanation for this finding might be that early after the duplication that gave rise to the *OsMADS29* and the *OsMADS30* subclade, the canonical B$_{sister}$ function was conferred by the *OsMADS29* ancestor to an almost exclusive extent, while the ancestral *OsMADS30* gene became a target for relaxed selection. This novel evolutionary trajectory might have been triggered by a mutation changing a pivotal amino acid within the MADS- or K-domain that resulted in a severe impairment of the OsMADS30 ancestor, for example by a change in the capability of dimerization of the protein. Also, a mutation within the promoter of the *OsMADS30* ancestor, which could have diminished expression, is conceivable. This would be in line with the diminished expression level of the *OsMADS30*-like genes (Figure S7).

However, even though the *OsMADS30*-like genes exist for at least 45 million years, typical traits of pseudogenes such as premature stop codons, translational frameshift mutations or mutated splice sites are missing outside of the *Oryza* lineage. Thus, a completely neutral evolution of all members of the subclade is rather unlikely. Hence, *OsMADS30*-like genes might have evolved within a 'twilight zone' between purifying selection and neutral evolution.

We further showed that *OsMADS30* expression domains are more unspecific than those of other B$_{sister}$ genes that have been studied so far (Figures 3 and 4) (Becker *et al.*, 2002; Nesi *et al.*, 2002; Yamada *et al.*, 2009). *OrMADS30*, the *OsMADS30* orthologue from the closely related wild rice *O. rufipogon* and other orthologues from grasses are expressed rather weakly in comparison to the *OsMADS29* orthologues (Figure S7). Furthermore, *OsMADS30* T-DNA insertion lines show no aberrant flower or seed phenotype. Taken together, these data argue against a function of *OsMADS30* in flower and caryopsis development, in contrast with what would be expected for a typical B$_{sister}$ gene. However, we cannot exclude that a function of *OsMADS30* in caryopsis development was obscured due to redundancy with another B$_{sister}$ gene or any other gene expressed in flowers or caryopses. Nevertheless, our findings support the view that *OsMADS30* and possibly also other *OsMADS30*-like genes evolved in an unusual way and very likely lost their ancestral function. Thus, we consider them being non-canonical B$_{sister}$ genes.

The C-terminal domain of MADS-domain proteins confers important functions, e.g. in some cases transcriptional activation (Honma and Goto, 2001). A possible function of the ancestral OsMADS30 protein might have been impaired by a complete exchange of the C-terminal domain due to the *OsME* insertion. Furthermore, *OsMADS30*-like genes in the rice lineage may have a minor functional significance as indicated by the relaxed selection pressure, the transposon insertion and deletion of exons within three of the seven investigated wild rice *OsMADS30* orthologues (Figure 2), and the very low expression of the *O. rufipogon* orthologue *OrMADS30*. In extreme, the ancestral *OsMADS30* before the *OsME* insertion might have been a non-functional gene (pseudogene).

The *OsME* might be a so far unrecognized transposable element

We found some evidence that *OsME* represents a special type of transposon, a helitron (Figure 2). Helitrons are most likely replicating by a rolling-circle mechanism and therefore do not induce target site duplications while inserting into new host sites (Kapitonov and Jurka, 2001), which is also true for the *OsME*. While *OsME* resembles also other features of helitrons, it does not show the typical stem–loop structure on its 3′ terminus. However, helitrons in *Aspergillus* also have been found to lack this characteristic trait (Galagan *et al.*, 2005), indicating that such a stem–loop structure is not an absolute requirement for a helitron.

However, it cannot be excluded that the three *OsME* loci are the outcome of segmental duplications rather than transposition. Segmental duplications are formed during a two-step process: first ancestral duplication units aggregate in seeding blocks which later on spread through the genome as duplication blocks (Bailey and Eichler, 2006; Kahn and Raphael, 2008). This would explain the high sequence similarity of partial sequences of *OsME* with a multitude of other loci in the rice genome.
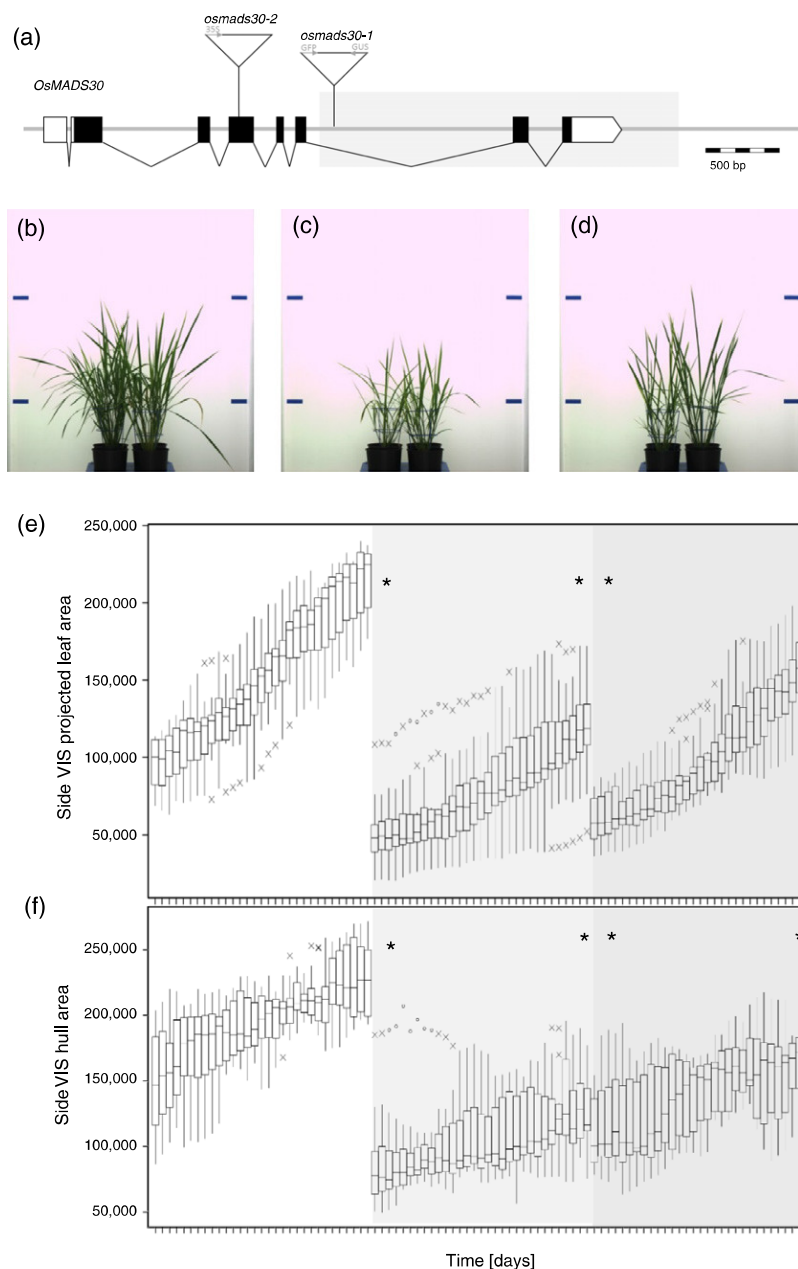
Figure 5. Genetic and phenotypic parameters of *Oryza sativa OsMADS30* T-DNA insertion lines and control plants.

(a) Location of T-DNA insertions. The locations of T-DNA insertions are depicted by triangles.

(b–d) Side view of wild type plants (b), plants of *osmads30-1* (c), and *osmads30-2* (d) T-DNA lines on day 131.

(e, f) Boxplots for each day of imaging (117-147 days) of the side visible area (e) and the side visible hull area (f) of wild type control *Oryza sativa* plants of the cultivar Dongjin (white background) and *osmads30-1* (light grey background) and *osmads30-2* (dark grey background). Significant differences of *osmads30-1* and *osmads30-2* to *O. sativa* cv. *Dongjin* are depicted by asterisks on day 120 and 147. Reason(s) for the apparent higher severity of the *osmads30-1* allele as compared with *osmads30-2* cannot be directly deduced from the presently available data. This deduction would require a detailed quantitative analysis of potentially formed partial or aberrant gene products to elucidate which allele represents a full or partial loss or possible gain of function.

OsMADS30 evolved in a twilight zone between non- and neo-functionalization

Our analyses suggested that the *OsME* is not present in the wild rice orthologues of *OsMADS30*, including *OrMADS30* from *O. rufipogon*. However, according to the recently published *O. rufipogon* genome (Huang *et al.*, 2012), the *OsMADS30* orthologue contains the *OsME* element, thus the parallel existence of both alleles (with and without the *OsME*) in the extant *O. rufipogon* population seems conceivable. The same could be true for the ancestral *O. rufipogon* population, from which *O. sativa* most likely was domesticated (Huang *et al.*, 2012). Hence, one possible scenario is that *O. sativa* descended from a subset of the population carrying the *OsMADS30* allele with the *OsME*. The insertion of *OsME* into an ancestral *OsMADS30*-like gene in the *Oryza* lineage may have been responsible for the widened expression pattern and the increased expression level of *OsMADS30* compared to *OrMADS30* (Figures 3 and 4), since regulatory sequences, including enhancers, are not necessarily located upstream but can also be located within and downstream of the transcribed region (Jeong *et al.*, 2002; Kaufmann *et al.*, 2009). The heterologous 3′ part of the mRNA might also have been responsible for altered RNA stability (LaGrandeur and

Table 2 Genotypes of F1 generations of OsMADS30 T-DNA insertion lines

	Wildtype (*OsMADS30/ OsMADS30*)	Heterozygous (*OsMADS30/ osmads30*)	Homozygous (*osmads30/ osmads30*)	Total
osmads30-1	109 (24.9)	196 (44.7)	133 (30.4)	438 (100)
osmads30-2	60 (26.8)	110 (49.1)	54 (24.1)	224 (100)

Total numbers of wild type, heterozygous and homozygous plants in the F1 generation obtained from selfing heterozygous T-DNA insertion plants. Relative numbers (%) are given in parentheses.

Parker, 1999; Leipuviene and Theil, 2007; Clark *et al.*, 2009). The resulting chimeric gene encodes a putative transmembrane domain, which is very rare for MADS-box genes (Kim *et al.*, 2010). A variety of the reported plant membrane-bound transcription factors is related to abiotic stress responses (for a review see (Seo *et al.*, 2008)). An improved stress response, for example within roots, could therefore have been an advantage conferred by the newly acquired *OsMADS30* allele.

We identified an asRNA, which is complementary to and shares exon-exon junctions with the *OsMADS30* mRNA (Figure S6). It is also expressed within the same order of magnitude as the *OsMADS30* mRNA (Figures 3 and 4). It has been shown for Arabidopsis that the presence of an asRNA and a similar expression level of sense and anti-sense RNA is common for a substantial number of genes and that this may be related to stress responses within the plant (Matsui *et al.*, 2008). AsRNAs contribute to a huge variety of regulatory mechanisms such as RNA editing, alternative splicing, methylation and RNA stability (Tufarelli *et al.*, 2003; Nishikura, 2006; Allo *et al.*, 2009; Fabian *et al.*, 2010). Gene silencing as well as activation are also associated with asRNA (Lee *et al.*, 1993; Palatnik *et al.*, 2003; Vasudevan *et al.*, 2007). However, the mechanism of the interaction of *OsMADS30* asRNA with the mRNA remains to be investigated. Most likely the presence of an asRNA has not been induced by the insertion of the *OsME*, since also in the *O. rufipogon*, which has no *OsME* in its *OrMADS30*, an asRNA was found.

Through imaging-based phenotypic analyses we found that homozygous individuals of two *OsMADS30* T-DNA insertion lines were significantly smaller than the respective wild types. However, growth rates did not differ significantly. This suggests that effects responsible for the observed lower shoot sizes of the mutants may have occurred before the detailed monitoring of the plants.

Plant height and compactness at the late booting stage in rice have been shown to have predictive power for plant biomass (Yang *et al.*, 2014). Alterations in growth dynamics and plant morphology in the mutant(s) indicate a function of *OsMADS30* during vegetative development, most likely in root development, since expression in root is most

prevalent (Figure 3). Potential functional implications for plant architecture remain to be elucidated.

In principle it could be that the diminished shoot sizes are the result of mutations or modifications other than the T-DNA insertion. We consider this unlikely, however, since both *osmads30-1* and *osmads30-2*, two independently generated lines with T-DNA insertions show similar mutant phenotypes. However, we cannot completely rule out yet that the mutant phenotypes are caused by the presence of the T-DNA itself, irrespective of its genomic position.

Our data provide an initial clue that *OsMADS30* acts in vegetative tissues and thus acquired a new function in cultivated rice. It is possible that this process of neo-functionalization has been primed by a decreased requirement of functional *OsMADS30*-like genes within the rice lineage. However, the insertion of *OsME* may have led to functional 'resurrection' of *OsMADS30* in rice. It has been reported, that pseudogenes can readopt a function, but convincing examples remain scarce (Trabesinger-Ruef *et al.*, 1996; Sharon *et al.*, 1999). Further studies, e.g. employing recombinant inbred lines arising from crosses of *O. sativa* and *O. rufipogon* or gene targeting approaches (Shimatani *et al.*, 2015), remain to show to which extant the *OsMADS30* locus is capable of modulating vegetative parameters of *O. sativa.* This will be needed to substantiate the hypothesis that *OsMADS30* is a precedence case of a resurrected neo-functionalized plant MADS-box gene.

EXPERIMENTAL PROCEDURES

Cultivation of rice plants and validation of T-DNA lines

Oryza sativa ssp. *japonica* cv. *Dongjin* and cv. *Tainung* were grown under short day conditions in a greenhouse at an average day temperature of 28°C. T-DNA lines were obtained from the Rice T-DNA Insertion Sequence Database (RISD (Ryu *et al.*, 2004)) and the Taiwan Rice Insertional Mutants (TRIM (Hsing *et al.*, 2007)) Database and verified by PCR with gene and T-DNA-specific primers. The number of T-DNAs in a plant genome was estimated by Southern blot and only plants with one T-DNA were used for further investigations. The T-DNA line PFG_1B-15231 contains the T-DNA of pGA2717 within the sixth intron of *OsMADS30* (*Os06g45650*); the mutant allele was named *osmads30-1*. The T-DNA line TRIM_M0035658 contains the T-DNA of pTag8 within the fourth exon of *OsMADS30;* the mutant allele was named *osmads30-2.*

In situ hybridization

Fresh wild-type flowers and developing caryopes were fixed in 4% paraformaldehyde or FAA (50% ethanol, 5% acetic acid, 3.7% formaldehyde), dehydrated and embedded in paraffin (Roth, https://www.carlroth.com/de/de); 8 μm-thick sections were hybridized on object slides with two different probes specific for *OsMADS30* (Figure S6) as described elsewhere (Kouchi and Hata, 1993). The same probe was used for sections of *O. rufipogon.*

RNA isolation and RT-PCR

Total RNA was isolated using the RNeasy Mini Plant Kit (Qiagen, https://www.qiagen.com/) according to manufacturer's instructions for most of the tissues. Total RNA from caryopses 10 days after pollination was isolated as described elsewhere (Li and Trick 2005). Reverse transcription reactions were carried out with Superscript-III First Strand Synthesis System for RT-PCR (Invitrogen, https://www.thermofisher.com/ie/en/home/brands/invitrogen.html). Gene-specific primers were used for direction specific reverse transcription of antisense RNA. The cDNA samples were used as templates for cloning, RT-PCR and quantitative RT-PCR. Real-time PCRs were performed using Maxima Probe qPCR Master Mix and Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific, http://www.thermoscientific.com/en/home.html) on an MXPRO 3000x (Stratagene, http://www.genomics.agilent.com/en/product.jsp?cid=AG-PT-169&tabId=AG-PR-1126&_requestid=85924) system and software according to the manufacturer's instructions. The housekeeping genes *ACTIN1* and *UBC* were used for normalization. Gene-specific probes and primers are shown in Table S5. Two to three biological replicates and three technical replicates were analyzed for each sample of qRT-PCR analysis. Statistical analysis was carried out with IBM SPSS statistics 23.

Phylogenetic analyses

B_{sister} genes from monocots were identified (for details see Methods S1) and combined with B_{sister} genes from Yang *et al.*, 2012 (Table S2) to construct a phylogenetic tree using the gymnosperm B_{sister} gene *GGM13* from *Gnetum gnemon* as a representative of the outgroup. The protein sequences were aligned using Probalign (Roshan and Livesay, 2006) and the resulting alignment was reverse translated into a nucleotide alignment using RevTrans (Wernersson and Pedersen, 2003). A Maximum Likelihood phylogeny was constructed using RAxML (Stamatakis, 2006) with the substitution model GTRGAMMA and conducting a rapid bootstrap analysis with 1000 replicates and the search for the best-scoring phylogeny in one program run.

Selection analyses

Selection analysis was carried out using some of the grass B_{sister} genes that had been used for the phylogeny (for details see Methods S1). Protein sequences were aligned using ClustalW (Thompson *et al.*, 1994). The resulting alignment was reverse translated using RevTrans (Wernersson and Pedersen, 2003). This nucleotide alignment and the previously obtained phylogeny were used as input for PAML (Yang, 2007). Different subsets of the alignment and different models were used for selection analyses (see Methods S1).

Promoter–GUS fusion and histochemical analysis of GUS activity

A sequence of 3178 bp containing 2295 bp upstream of the translation start site of *OsMADS30*, and the MADS-box (188 bp + 38 bp) and the second intron (658 bp) of that gene was amplified with specific pairs of primers. This fragment was then recombined into the vector pCAMBIA1301-Ubi (Cui *et al.*, 2010) upstream of the GUS (*uidA*) gene using the restriction enzymes *Kpn*I and *Nco*I. The resulting construct was introduced into the cultivar Zhonghua 11 by *Agrobacterium tumefaciens*-mediated transformation, as described previously (Hiei *et al.*, 1994). GUS staining was performed as previously described (Jefferson *et al.*, 1987).

Imaging-based plant phenotyping

After a growth period of in total 4 months, plants were transferred to the climate-controlled greenhouse for high-throughput plant phenotyping (for details see Methods S1). During all steps of plant growth genotypes and replicates were randomized within the plant growth area. During cultivation in the high-throughput plant phenotyping system, plant morphological parameters (extracted from RGB imaging, top and side view) and bulk fluorescence signals (FLUOR) were recorded (Junker *et al.*, 2014). The IAP software tool was used for image analysis and feature extraction (Klukas *et al.*, 2014). For details see Methods S1.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phylogeny of B_{sister} genes.

Figure S2. Alignment of B_{sister} proteins from grasses (Poaceae), *Amborella* and *Gnetum*.

Figure S3. Transmembrane prediction for OsMADS30.

Figure S4. Alignment of the *OsMADS30* genomic locus.

Figure S5. Sequence features of *OsME*.

Figure S6. *OsMADS30* asRNA and *in situ* hybridization probes.

Figure S7. Expression of B_{sister} genes in grasses.

Figure S8. *OsMADS30* cDNA of T-DNA lines.

Figure S9. Mature carpels of *Oryza sativa.*

Figure S10. Seeds of *OsMADS30* T-DNA insertion lines.

Figure S11. Grain germination of *Oryza sativa OsMADS30* T-DNA insertion lines and control plants.

Figure S12. Growth rate (a), compactness (b) and plant height (c) of *OsMADS30* T-DNA insertion lines and wild type *Oryza sativa.*

Table S1. Expression data of B_{sister} genes in grasses.

Table S2. Genes used for phylogeny.

Table S3. Extracted values of selected phenotypic traits.

Table S4. Imaging derived phenotypic traits of 3 rice genotypes for day 117 and day 147 after sowing.

Table S5. Primer and probes for quantitative RT-PCR.

Methods S1. Supplementary Experimental Procedures.

REFERENCES

Airoldi, C.A. and Davies, B. (2012) Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics*, **39**, 157–165.

Al-Dous, E.K., George, B., Al-Mahmoud, M.E. *et al.* (2011) *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* **29**, 521–527.

Allo, M., Buggiano, V., Fededa, J.P. *et al.* (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* **16**, 717–724.

Al-Mssallem, I.S., Hu, S., Zhang, X. *et al.* (2013) Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* **4**, 2274.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Arora, R., Agarwal, P., Ray, S., Singh, A.K., Singh, V.P., Tyagi, A.K. and Kapoor, S. (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom.* **8**, 242.

Bailey, J.A. and Eichler, E.E. (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**, 552–564.

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

Chen, G.Q., Deng, W., Peng, F., Truksa, M., Singer, S., Snyder, C.L., Mietkiewska, E. and Weselake, R.J. (2013a) Brassica napus TT16 homologs with different genomic origins and expression levels encode proteins that regulate a broad range of endothelium-associated genes at the transcriptional level. *Plant J.* **74**, 663–677.

Chen, J., Huang, Q., Gao, D. *et al.* (2013b) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying Oryza genome evolution. *Nat. Commun.* **4**, 1595.

Clark, A., Dean, J., Tudor, C. and Saklatvala, J. (2009) Post-transcriptional gene regulation by MAP kinases via AU-rich elements. *Front. Biosci. (Landmark Ed)*, **14**, 847–871.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G. and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673–676.

Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., Xu, Q., Chong, K., Theissen, G. and Meng, Z. (2010) Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant J.* **61**, 767–781.

D'Hont, A., Denoeud, F., Aury, J.M. *et al.* (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature,* **488**, 213–217.

Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379.

de Folter, S., Shchennikova, A.V., Franken, J., Busscher, M., Baskar, R., Grossniklaus, U., Angenent, G.C. and Immink, R.G. (2006) A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. *Plant J.* **47**, 934–946.

Galagan, J.E., Calvo, S.E., Cuomo, C. *et al.* (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature,* **438**, 1105–1115.

Ge, S., Sang, T., Lu, B.R. and Hong, D.Y. (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl Acad. Sci. USA,* **96**, 14400–14405.

Gramzow, L. and Theissen, G. (2010) A hitchhiker's guide to the MADS world of plants. *Genome Biol.* **11**, 214.

Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282.

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

Hsing, Y.I., Chern, C.G., Fan, M.J. *et al.* (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.* **63**, 351–364.

Huang, X.H., Kurata, N., Wei, X.H. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501.

International Brachypodium Initiative I (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature,* **463**, 763–768.

Jacquemin, J., Bhatia, D., Singh, K. and Wing, R.A. (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr. Opin. Plant Biol.* **16**, 147–156.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K. and An, G. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* **130**, 1636–1644.

Jiang, N., Feschotte, C., Zhang, X. and Wessler, S.R. (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.* **7**, 115–119.

Junker, A., Muraya, M.M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A.E., Meyer, R.C., Riewe, D. and Altmann, T. (2014) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant. Sci.* **5**, 770.

Kahn, C.L. and Raphael, B.J. (2008) Analysis of segmental duplications via duplication distance. *Bioinformatics,* **24**, i133–i138.

Kapitonov, V.V. and Jurka, J. (2001) Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA,* **98**, 8714–8719.

Kapitonov, V.V. and Jurka, J. (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* **23**, 521–529.

Kaufmann, K., Muino, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P. and Angenent, G.C. (2009) Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol.* **7**, e1000090.

Kellogg, E.A. (1998) Relationships of cereal crops and other grasses. *Proc. Natl Acad. Sci. USA,* **95**, 2005–2010.

Kellogg, E.A. (2001) Evolutionary history of the grasses. *Plant Physiol.* **125**, 1198–1205.

Kim, S.G., Lee, S., Seo, P.J., Kim, S.K., Kim, J.K. and Park, C.M. (2010) Genome-scale screening and molecular characterization of membrane-bound transcription factors in Arabidopsis and rice. *Genomics,* **95**, 56–65.

Klukas, C., Chen, D. and Pape, J.M. (2014) Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant Physiol.* **165**, 506–518.

Kouchi, H. and Hata, S. (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* **238**, 106–119.

LaGrandeur, T. and Parker, R. (1999) The cis acting sequences responsible for the differential decay of the unstable MFA2 and stable PGK1 transcripts in yeast include the context of the translational start codon. *RNA,* **5**, 420–433.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* **75**, 843–854.

Leipuviene, R. and Theil, E.C. (2007) The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. *Cell. Mol. Life Sci.* **64**, 2945–2955.

Li, Z. and Trick, H.N. (2005) Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *BioTechniques,* **38**, 872–876.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science,* **290**, 1151–1155.

Matsui, A., Ishida, J., Morosawa, T. *et al.* (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol.* **49**, 1135–1149.

Mizzotti, C., Mendes, M.A., Caporali, E., Schnittger, A., Kater, M.M., Battaglia, R. and Colombo, L. (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J.* **70**, 409–420.

Mondragon-Palomino, M., Hiese, L., Harter, A., Koch, M.A. and Theissen, G. (2009) Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Evol. Biol.* **9**, 81.

Nayar, S., Kapoor, M. and Kapoor, S. (2014) Post-translational regulation of rice MADS29 function: homodimerization or binary interactions with other seed-expressed MADS proteins modulate its translocation into the nucleus. *J. Exp. Bot.* **65**, 5339–5350.

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

Nishikura, K. (2006) Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat. Rev. Mol. Cell Biol.* **7**, 919–931.

Ohno, S. (1970) *Evolution by Gene Duplication Heidelberg.* Heidelberg, Germany: Springer Verlag.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature,* **425**, 257–263.

Rice Full-Length cDNA Consortium (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science,* **301**, 376–379.

Roshan, U. and Livesay, D.R. (2006) Probalign: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics*, **22**, 2715–2721.

Ryu, C.H., You, J.H., Kang, H.G., Hur, J., Kim, Y.H., Han, M.J., An, K., Chung, B.C., Lee, C.H. and An, G. (2004) Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. *Plant Mol. Biol.* **54**, 489–502.

Seo, P.J., Kim, S.G. and Park, C.M. (2008) Membrane-bound transcription factors in plants. *Trends Plant Sci.* **13**, 550–556.

Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T. and Lancet, D. (1999) Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics*, **61**, 24–36.

Shimatani, Z., Nishizawa-Yokoi, A., Endo, M., Toki, S. and Terada, R. (2015) Positive-negative-selection-mediated gene targeting in rice. *Front. Plant. Sci.* **5**, 748.

Singh, R., Ong-Abdullah, M., Low, E.T. *et al.* (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, **500**, 335–339.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Munster, T., Winter, K.U. and Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Trabesinger-Ruef, N., Jermann, T., Zankel, T., Durrant, B., Frank, G. and Benner, S.A. (1996) Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *FEBS Lett.* **382**, 319–322.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G. and Higgs, D.R. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* **34**, 157–165.

Vasudevan, S., Tong, Y. and Steitz, J.A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science*, **318**, 1931–1934.

Wang, L., Si, W., Yao, Y., Tian, D., Araki, H. and Yang, S. (2012) Genome-wide survey of pseudogenes in 80 fully re-sequenced *Arabidopsis thaliana* accessions. *PLoS ONE*, **7**, e51769.

Wang, W., Haberer, G., Gundlach, H. *et al.* (2014) The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat. Commun.* **5**, 3311.

Wernersson, R. and Pedersen, A.G. (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31**, 3537–3539.

Yamada, K., Saraike, T., Shitsukawa, N., Hirabayashi, C., Takumi, S. and Murai, K. (2009) Class D and B(sister) MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (Triticum aestivum L.). *Plant Mol. Biol.* **71**, 1–14.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Yang, X., Wu, F., Lin, X., Du, X., Chong, K., Gramzow, L., Schilling, S., Becker, A., Theissen, G. and Meng, Z. (2012) Live and let die - the B(sister) MADS-box gene OsMADS29 controls the degeneration of cells in maternal tissues during seed development of rice (Oryza sativa). *PLoS ONE*, **7**, e51435.

Yang, W., Guo, Z., Huang, C. *et al.* (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **5**, 5087.

Yin, L.L. and Xue, H.W. (2012) The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. *Plant Cell*, **24**, 1049–1065.

Zhang, G., Liu, X., Quan, Z. *et al.* (2012) Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* **30**, 549–554.

Zhou, Y. and Mishra, B. (2005) Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. *Proc. Natl Acad. Sci. USA*, **102**, 4051–4056.
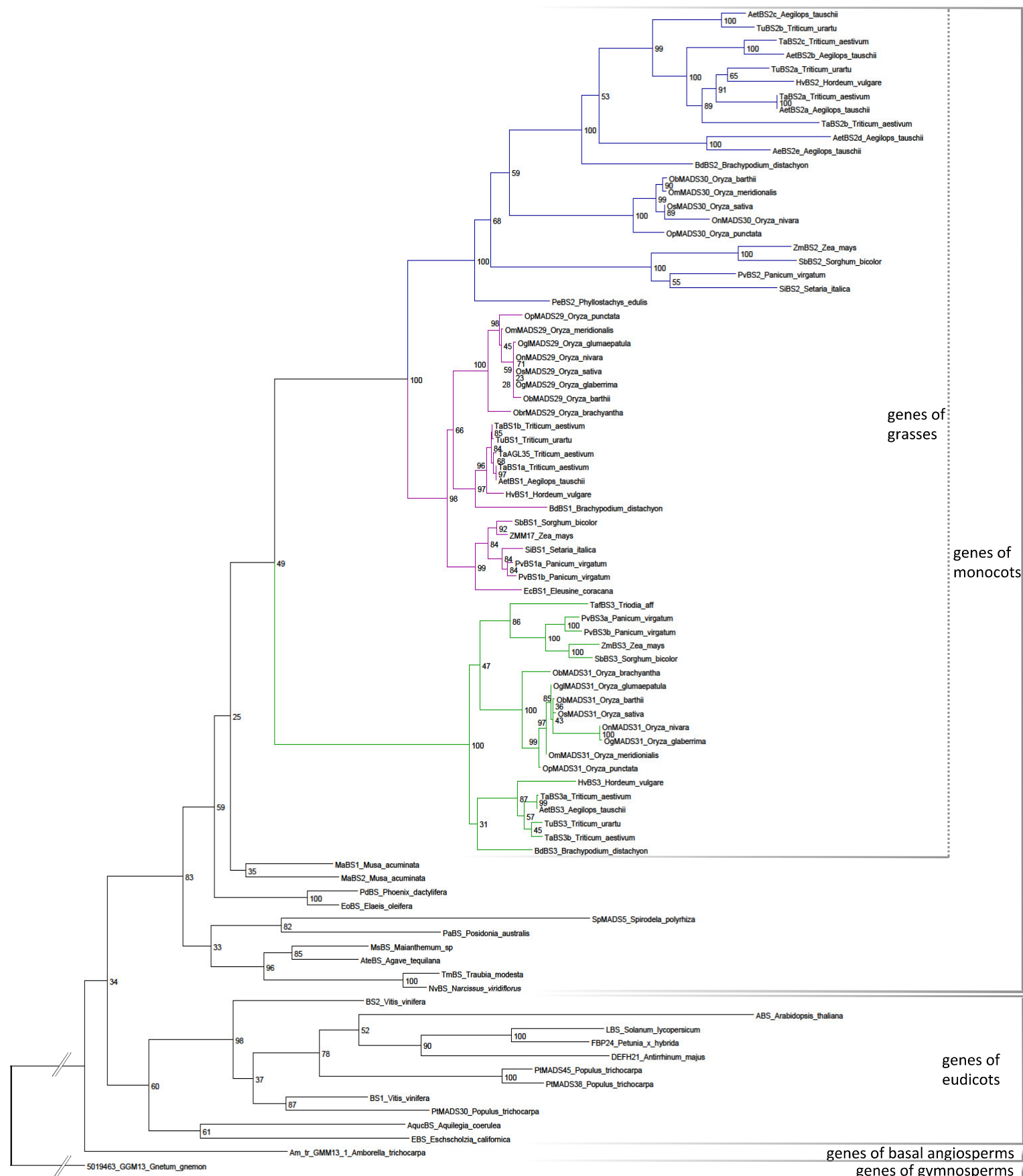
Figure S1. Phylogeny of B_sister genes. A RAxML tree was calculated for all known monocot and representative eudicot B_sister genes as well as the *Amborella trichocarpa* homologs with bootstrap values given on the nodes. *GGM13* from *Gnetum gnemon* served as representative of the outgroup. *OsMADS29*-, *OsMADS30*- and *OsMADS31*-subclades are indicated by purple, blue and green branches, respectively. Gene phylogeny within subclades follows species phylogeny with two exceptions: The ortholog of *OsMADS30* from *Phyllostachys* is basal within the *OsMADS30* subclade in our phylogeny but would be expected to be closely related to the *OsMADS30*-like genes from *Oryza* species (Kellogg 2001, 1998). Also, the *OsMADS30* ortholog of *Hordeum vulgare* would be expected to be the sister gene to the subclade of *OsMADS30*-like genes of *Triticum* and *Aegilops* species, but instead it appears to be closely related to one *OsMADS30*-like gene from *Triticum urartu*. The branch leading to the *OsMADS29* subclade and the branches within this subclade are short, indicating only few substitutions in the *OsMADS29*-like genes, whereas the branch leading to the *OsMADS31* subclade is much longer, which implies considerable sequence changes before species divergence.
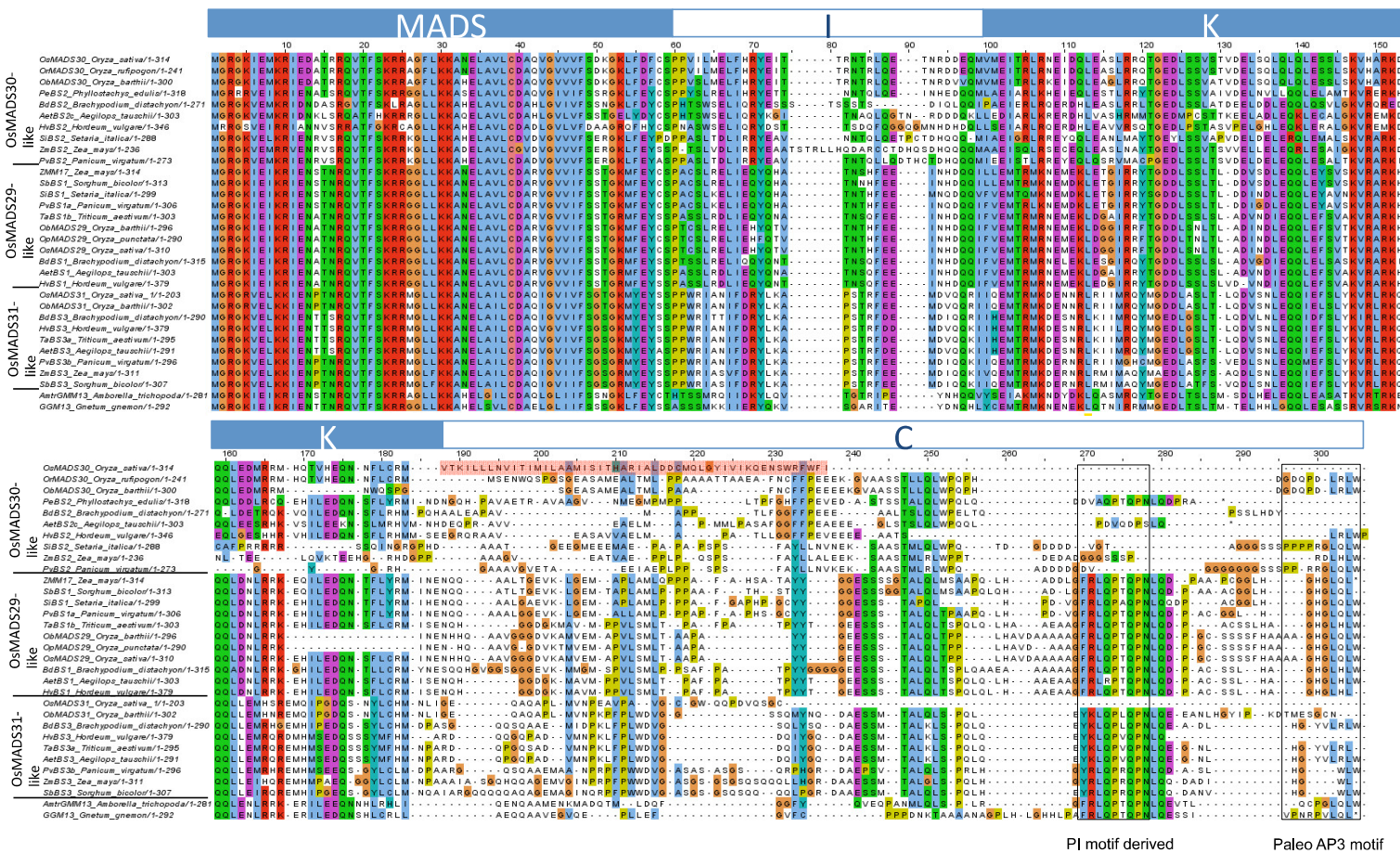
Figure S2. Alignment of B_sister proteins from grasses (Poaceae), _Amborella_ and _Gnetum_. Protein sequences of OsMADS29-, OsMADS30- and OsMADS31-like proteins as well as B_sister protein sequences from _Gnetum genom_ and _Amborella trichopoda_ were aligned. MADS, I, K and C-terminal domain are designated by bars above the alignment. The heterologous C-terminal domain of OsMADS30 encoded by the _OsME_ is highlighted in red. PI-derived and Paleo-AP3 motif are indicated by boxes. The alignment is colored according to the ClustalX coloring scheme. The mean pairwise identity within the MADS domain of OsMADS29-like proteins and OsMADS31-like proteins is 97 % and 93 %, respectively, while for OsMADS30-like proteins it is only 73 %. Furthermore, almost all OsMADS30-like proteins lack the PI derived motif and have a distorted Paleo AP3 motif.
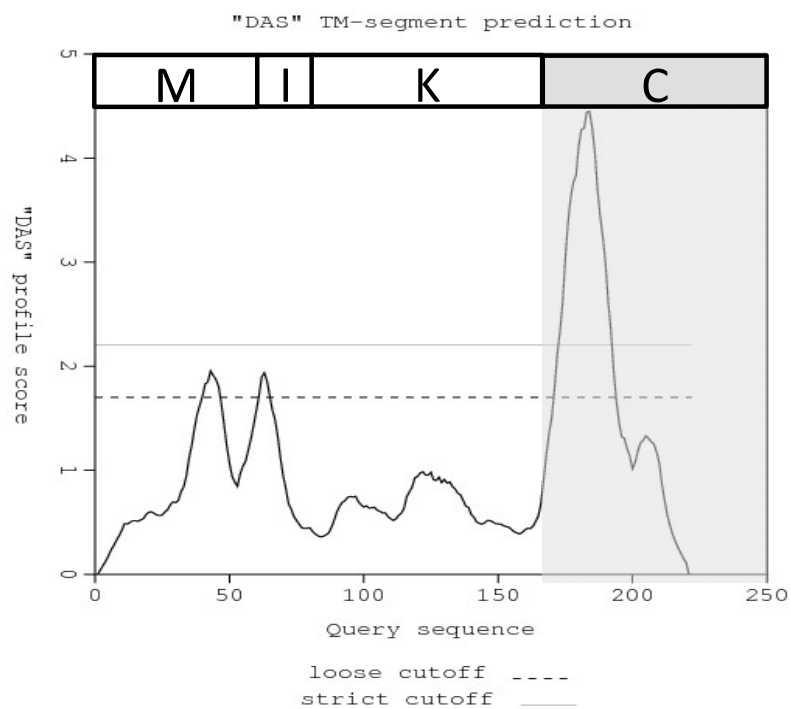
Figure S3. Transmembrane prediction for OsMADS30. "DAS" (Dense Alignment Surface) profile score of OsMADS30 predicted on the basis of its amino acid sequence. Protein domain structure is depicted on top. Calculated at http://www.sbc.su.se/~miklos/DAS/

```
                    *        20         *        40         *        60         *        80         *       100         *       120         *       140
mRNA_OsM30 : GGATTGGAGCTAGCACTTCTTTGACTGCCCTTGTTCATCGAGGC------------------------------------------------------------------------------------------------GGCGAGGCGAGGTC :   58
chr_OsM30  : GGATTGGAGCTAGCACTTCTTTGACTGCCCTTGTTCATCGAGGCAAGTGCAGAAATCTTTTCAGTGTGGCTGGATCCCGTTTAATTCAAGCGATTTGATTGGGGTTTTTTGGTTTTTTCGATGGACAGGCGGCGAGGCGAGGTC :  144
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                    *       160         *       180         *       200         *       220         *       240         *       260         *       280
mRNA_OsM30 : GCGATGGGGCAAGGGAAGATCGAGATGAAGAGGATCGAGGACGCGACGAGGCGGCAGGTGACGTTCAGCAAGCGCAGGGCTGGGTTTCTCAAGAAGGCGAACGAGCTCGCCGTGCTGTGCGATGCGCAGGTCGGCGTCGTCGTC :  202
chr_OsM30  : GCGATGGGGCAAGGGAAGATCGAGATGAAGAGGATCGAGGACGCGACGAGGCCGCAGGTGACGTTCAGCAAGCGCAGGGCTGGGTTTCTCAAGAAGGCGAACGAGCTCGCCGTGCTGTGCGATGCGCAGGTCGGCGTCGTCGTC :  288
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

               *       300         *       320         *       340         *       360         *       380         *       400         *       420         *
mRNA_OsM30 : TTCTCCGACAAGGGCAAGCTCTTCGACTTCTGCAGCCCGCCGGTCAT-------------------------------------------------------------------------------------------------------- :  249
chr_OsM30  : TTCTCCGACAAGGGCAAGCTCTTCGACTTCTGCAGCCCGCCGGTCATGTGCGCCCGCCGCCGCCGGCCCTAGCTTCTATCTCTGTTTCTTGTTAATTTGTTTATGACGAAATGCCAATGAAAAATCTGCTTTAAAAGAAACAAG :  432
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                  440         *       460         *       480         *       500         *       520         *       540         *       560         *
mRNA_OsM30 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
chr_OsM30  : AAAGAGAGAGAGAGTGTGTGTGTGTGCTGATCTCTAATCTTGATTTCTATATATCTGCATGCCTTAGCTGAGTTTTCAGTTCTTTGTGGTGATTGGTGGTGTGTGCGATGATTCAGAAATTTAGCATTGCATTTATTTGTTG :  576
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                  580         *       600         *       620         *       640         *       660         *       680         *       700         *       720
mRNA_OsM30 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
chr_OsM30  : CAGTTGATATTTCACTCAGCATTCCAGATAAATGTTTGGGGGTACTGTTCGATTAACATCCCGATCAAATGGGTAGTGTTCTTCATGATTGGTAGTGGATTGGGAAATTGGGATCGCTTTCCTTGCATTCAACATTATCACTA :  720
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                       *       740         *       760         *       780         *       800         *       820         *       840         *       860
mRNA_OsM30 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
chr_OsM30  : GAAATCTTTCGATGGTTCATGTTCTAGAGCTTGCTTTTCTTTCTGATTTAATCTTTTCTGGCATCCTGAAATTTTGATTTCAGGATCTATTCCTATAGGACTTGTGCTTGCTGCATAGTGCATTCTTCAATTAAGCAGTGCCCA :  864
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                       *       880         *       900         *       920         *       940         *       960         *       980         *      1000
mRNA_OsM30 : --------------------------------------------------------------------------------------------------------------------------------------CTTGATGGAGCTGTT :  264
chr_OsM30  : TGACCACATCGATTAGTACTACTAGAAGCTAATACTCCTTTTACATGTATAAGTTCACTTCATTGGTCCAATGTACTAGTAATACCATATTGTTTCTGGATGTAACATATTTGATTATGCTTTGTTAAGCTTGATGGAGCTGTT : 1008
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

               *      1020         *      1040         *      1060         *      1080         *      1100         *      1120         *      1140         *
mRNA_OsM30 : TCACCGTTATGAGATCACCACCAGAAACACTCGGCTTCAGGAGACAAACCGTGATGATGAG------------------------------------------------------------------------------------ :  325
chr_OsM30  : TCACCGTTATGAGATCACCACCAGAAACACTCGGCTTCAGGAGACAAACCGTGATGATGAGGTAAGCGAACTGAATCTAACACATGTTATTTTTTTCTTTTTTACATTTGAAAGGCAATTGCATGTCCCACACCAAATTCCTAA : 1152
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                  1160         *      1180         *      1200         *      1220         *      1240         *      1260         *      1280         *
mRNA_OsM30 : --------------------------------------------------------------CAAATGGTCATGGAGATCACAAGGCTAAGGAATGAGATCGACCAGCTCGAAGCCAGTTTAAGGAGGCAAACTGGAGAAGACCTGTCATCT :  415
chr_OsM30  : TCACTCAAAGCTTCATTTCTACAACTAATCTACATGTGTGATGCTGCCCTTGCAGCAAATGGTCATGGAGATCACAAGGCTAAGGAATGAGATCGACCAGCTCGAAGCCAGTTTAAGGAGGCAAACTGGAGAAGACCTGTCATCT : 1296
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -

                  1300         *      1320         *      1340         *      1360         *      1380         *      1400         *      1420         *      1440
mRNA_OsM30 : GTGTCCACAGTGGATGAGCTCAGCCAGCTGCAGCTGCAGCTTGAATCATCTCTCAGCAAAGTTCATGCGAGGAGG------------------------------------------------------------------------ :  491
chr_OsM30  : GTGTCCACAGTGGATGAGCTCAGCCAGCTGCAGCTGCAGCTTGAATCATCTCTCAGCAAAGTTCATGCGAGGAAGGTACTGACACTTGCTTAAATATGGCAGAAGATTAGTGCAGATCAGCATCAAATCCTTATGGAATATACA : 1440
ancstrC    : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME1      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME2      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
OsME3      : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J013110N09 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
J090038P19 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
001-033-F0 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
```

Figure S4. Alignment of the *OsMADS30* genomic locus. The nucleotide sequence of the genomic locus of *OsMADS30* (chr_OsM30) was aligned to the *OsMADS30* mRNA (mRNA_OsM30), *OsME1*, *OsME2*, *OsME3*, the putative ancestral C-terminal region (ancstrC) and published mRNA transcribed from *OsME2* (J090038P19) and *OsME3* (001-033-F05; J013110N09).
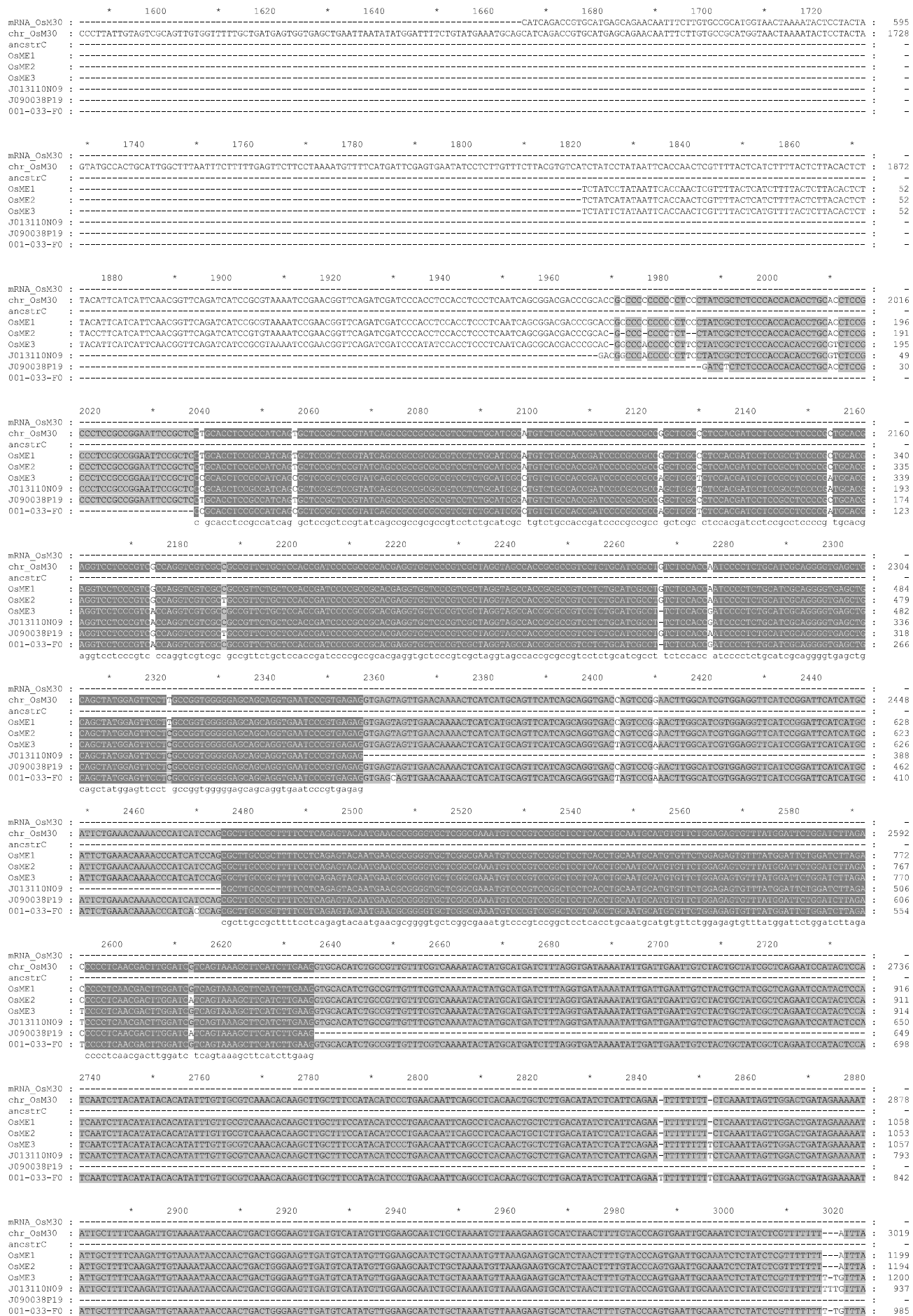
Figure S4. Alignment of the *OsMADS30* genomic locus. continued

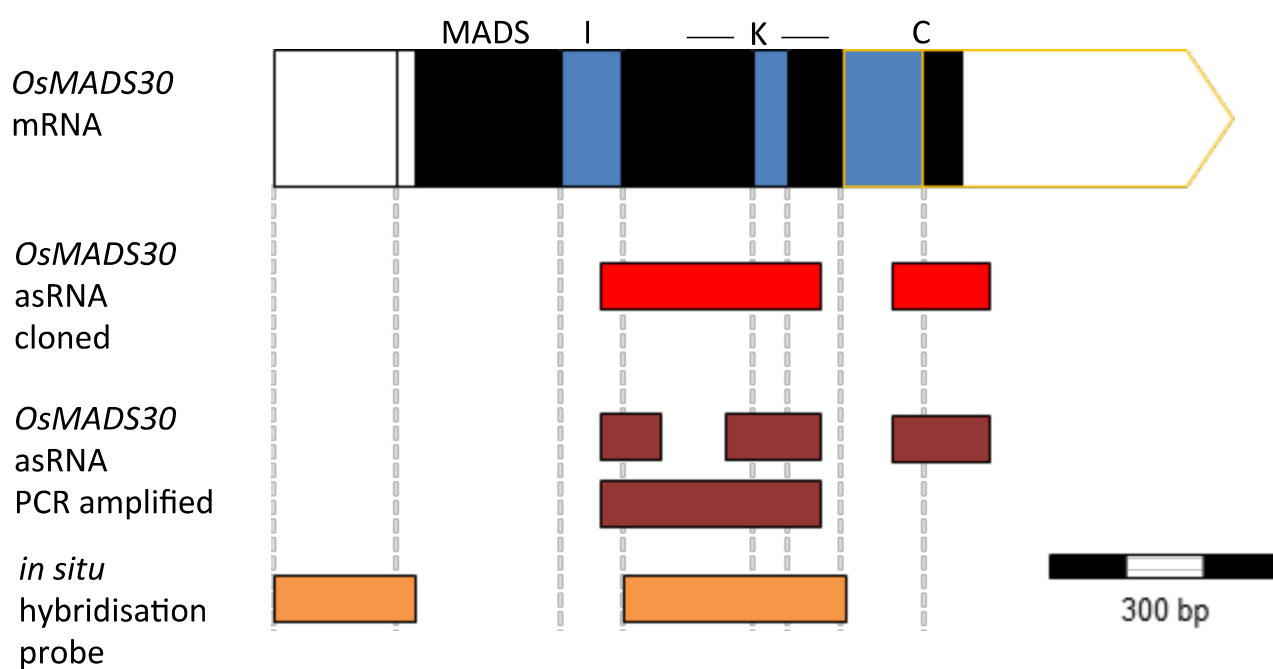Figure S4. Alignment of the *OsMADS30* genomic locus. continued

a

Name	From	To	Name	From	To	Dir	Sim	Pos	Score
31150.tmp/data.ori	158	456	ENSPM4_OS	4486	4779	d	0.7083	0.7083	404

```
CCCCCGTTCCTATCGCTCTCCCACGACACCTGCGTCTCCGCC-CTCCGCCGGAATTCCGCTCCCGCACCT
||  ||---|||-||:||||: :|-||||  |:|||| :||-|||||:|||--|: |||-|||:||||||
CCGCCCT--CTA-CGCCTCCTCTC-CACCGCCGCCTCGACCTCTCCACCG--ACGCCG-TCCTGCACCA

CCGCCATCAGCGCGTC-CGGTCCGTATCAGCCGCCGC-----------GCCGTCCTGTGCATCGGCC-----
||||||| :-|:||||| ||:||||||||||-|||||-----------||||:||| ||:|||||-----
CCGCCGCC-GCACTCACTCTC----TCGACCGCCGCCATCCATCTCCGCCGCCATCTCCACCGCCGCCGC

------TGTCTGCCACCGATCCCCGCCGCCAGCTCG-CTCTCCACGATCCTCCGCCTCCCCGATGC-AC
------| ||:|||-||:|||||||||||||| :|-||||| ||--|| ||||||:|||--|||-::|
CTCGACCTCTCCGCCGCC-ATCTCGCCGCCATCCAGACCCTCCTCG--CCGCCGCCGCTCTC--TGCCGC

GAGGTCCTCCCGTCACCAGGTCG---TCGCCGCCGGTTCTGCTCCACCGATCCCCGCCGCACGAGGTGCTC
|  ||| |||-||| ||--|:||---|||||||||| ::|:-||||||:|||  || ||| ---:|||
CATCTCCGCCC-TCTCC--GCCGCCGTCGACCTCCCT-CTCGCACCG-CCGCCTCCTCACT---CACTC

CCGTCGGTAGGTAGCCACCGC-GCCGTCCTCTGCATCGGCCTTCTCCACCG
:|-|||----------|||||-||||| |:|||| ||||||| |||||||
TC-TCG---------ACCGCCGCCGTCCTCTCCACCGCCATCTCCACCG
```

Name	From	To	Name	From	To	Dir	Sim	Pos	Score
31150.tmp/data.ori	619	658	MuDR-5_ZM	1476	1516	c	0.8049	0.8049	204

```
GATTCATCATGCATT-CTGAAACAAAACCCATCATCCAGCG
|:|:|||||||||||-|||||||||||| ||||||| ||:|
GGTCCACCATGCATTTCTGAAACAAAATCCATAATACAGTG
```

b

c

Figure S5. Sequence features of *OsME*. (a) Results from a RepBase search using *OsME1* as a query. Searching RepBase (Jurka, et al. 2005), we found that a region of 298 bp of the *OsME* shows a similarity of about 70 % to a non-coding region of an En/Spm transposable element and a 40 bp region with similarity to a MuDR transposable element. (b, c) A dot plot comparison of *OsME1* to itself was conducted with seaview (Gouy et al. 2010). For LTRs and TIRs diagonal lines would be expected in the corners of the diagram. Window size and identical sites per window were 4/4 (b) and 11/20 (c).

Figure S6. *OsMADS30* **asRNA and** *in situ* **hybridization probes**. *OsMADS30* mRNA is schematically drawn on top. The encoded domains are given above. Coding regions are filled black and blue, UTRs filled white. Regions encoded by *OsME* are bordered orange. The exons are indicated by broken grey lines. Red, brown and orange boxes below the *OsMADS30* mRNA indicate cloned or PCR amplified fragments of complementary asRNA and probes used for *in situ* hybridization, respectively.

		seedling	leaf	root	flower	seed
OsMADS29-subclade	Sb04g004736 (*Sorghum bicolor*)				■ (black)	■ (black)
	TaAGL35 (*Triticum aestivum*)				■ (black)	■ (black)
	MLOC_65966 (*Hordeum vulgare*)				■ (black)	■ (black)
	ZMM17 (*Zea mays*)				■ (black)	■ (black)
	OsMADS29 (*Oryza sativa*)				■ (black)	■ (black)
OsMADS30-subclade	Sb10g026690 (*Sorghum bicolor*)				▩ (grey)	▩ (grey)
	MLOC_64819 (*Hordeum vulgare*)				▩ (grey)	
	ZMM51 (*Zea mays*)					
	OsMADS30 (*Oryza sativa*)	▩ (grey)			▩ (grey)	▩ (grey)
	OrMADS30 (*Oryza rufipogon*)				▩ (grey)	▩ (grey)
OsMADS31- subclade	Sb06g028420 (*Sorghum bicolor*)				■ (black)	■ (black)
	ZMM20 (*Zea mays*)				▩ (grey)	▩ (grey)
	ZMM9 (*Zea mays*)	▩ (grey)	▩ (grey)		■ (black)	■ (black)
	ZMM27 (*Zea mays*)		▩ (grey)		▩ (grey)	▩ (grey)
	ZmMADS21 (*Zea mays*)		▩ (grey)		▩ (grey)	▩ (grey)
	ZMM28 (*Zea mays*)		▩ (grey)		▩ (grey)	▩ (grey)
	ZMM48 (*Zea mays*)		▩ (grey)		▩ (grey)	▩ (grey)
	OsMADS31 (*Oryza sativa*)	▩ (grey)	▩ (grey)	▩ (grey)	▩ (grey)	▩ (grey)
	HvBS3 (*Hordeum vulgare*)	nd	nd	nd	▩ (grey)	nd

Figure S7. Expression of B$_{sister}$ genes in grasses. Data obtained from genevestigator, qTeller and ESTs. White, grey and black boxes indicate no, weak and medium/strong expression, respectively. See Supplementary Table 1 for details and references. nd, no data.
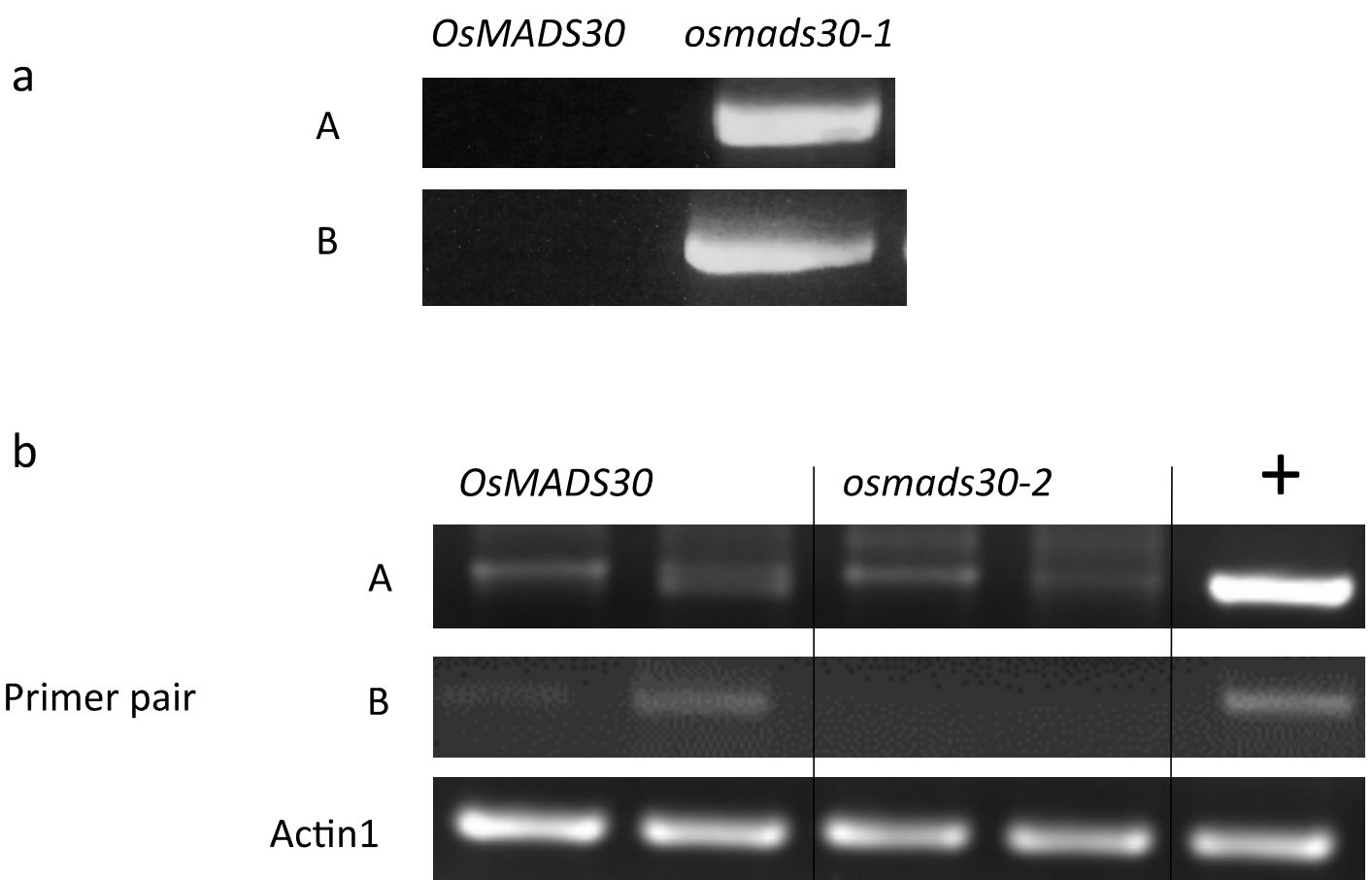
Figure S8. *OsMADS30* **cDNA of T-DNA lines.** (a) cDNA from rice plants with and without T-DNA insertion was used to amplify PCR products. These were used as a template for nested PCRs with two different primer combinations (A, B) located within *OsMADS30* (forward primer, upstream T-DNA insertion) and within the *GFP* gene (reverse primer, part of T-DNA). The PCR product was verified through sequencing. (b) RNA extracted from 7 day old seedlings with and without T-DNA insertion was used for cDNA synthesis. Primer pairs were located both upstream (A), and up and downstream (B) of the T-DNA insertion of *osmads30-2*. Actin1 was used as control. + , plasmid DNA with the respective cloned cDNA.
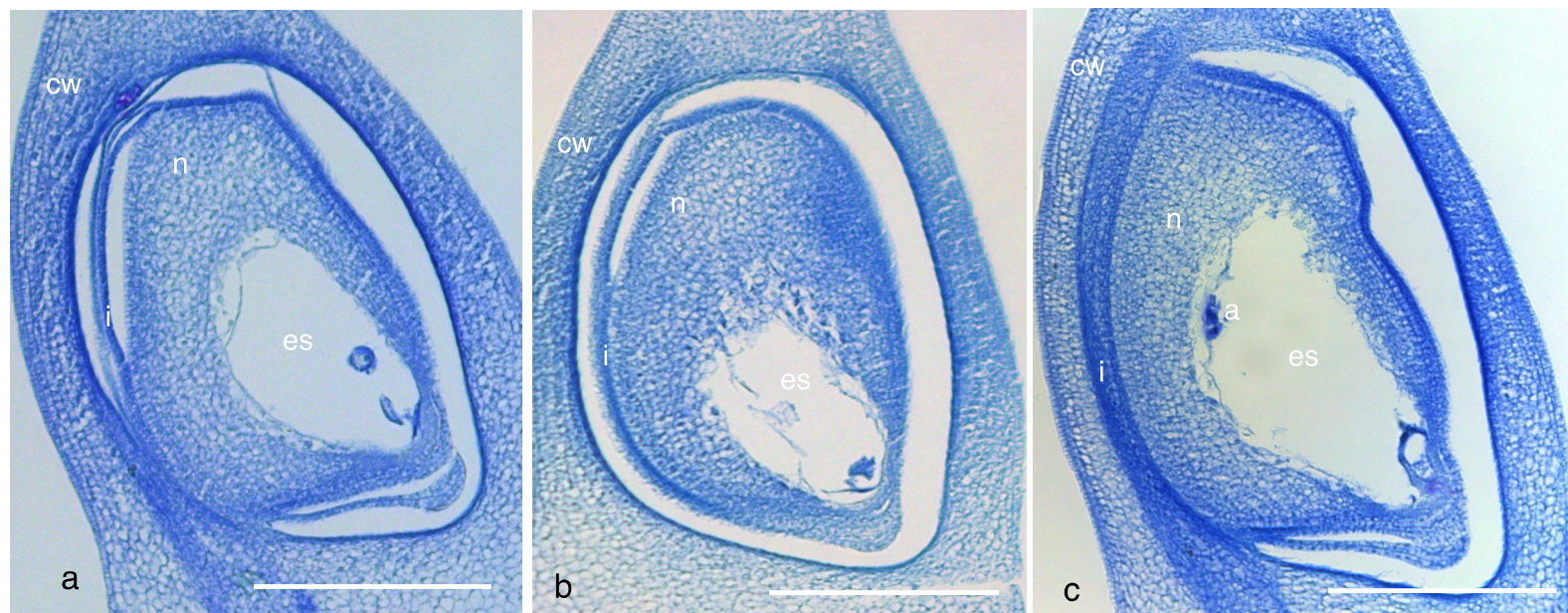
Figure S9. Mature carpels of *Oryza sativa*. Toluidine stained mature carpels of the *Oryza sativa japonica* wild type (a) and the T-DNA insertion lines *osmads30-1* (b) and *osmads30-2* (c). Scale bars = 200 µm. a, antipods; cw, carpel wall; es, embryo sac; n, nucellus.
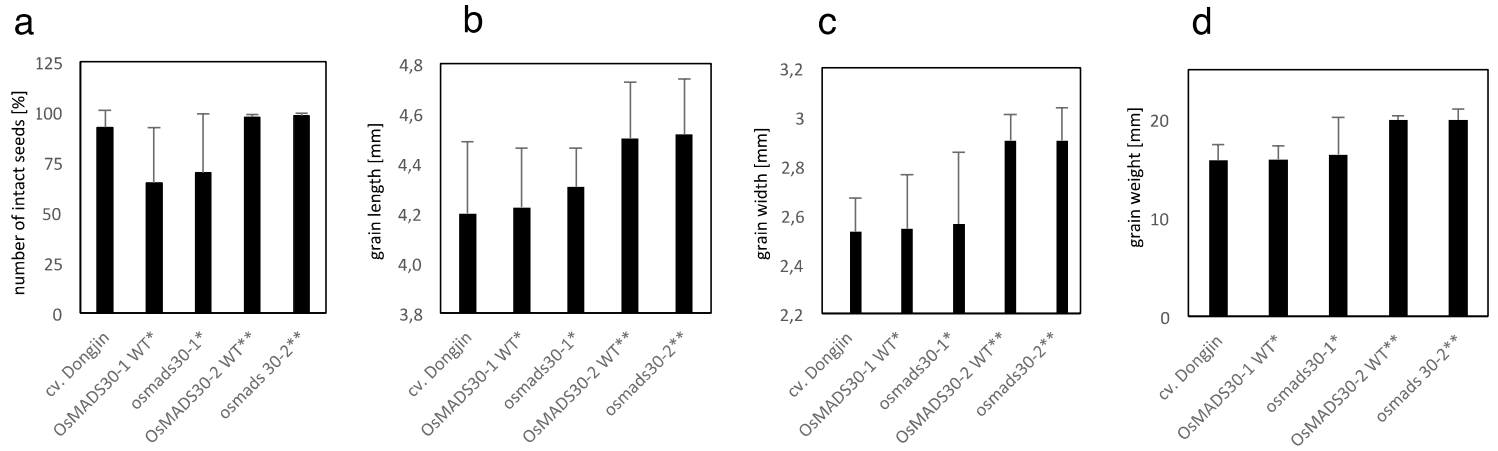
Figure S10. Seeds of *OsMADS30* T-DNA insertion lines. Seed set (a), length (b), width (c) and weight (d) of T-DNA insertion lines and wild type control plants. * and ** indicate sibling plants obtained from a selfed heterozygous plant. Differences between non-transformed Donjing cv. plants and mutant lines did occur. However, they were not correlated with the mutant allele (*osmads30-1* and *osmads30-2*) but rather with the T-DNA insertion line itself, as they were found in wild type siblings of the respective mutant line as well. Hence we do not consider the mutation of *OsMADS30* to be the cause of the variation of respective traits.
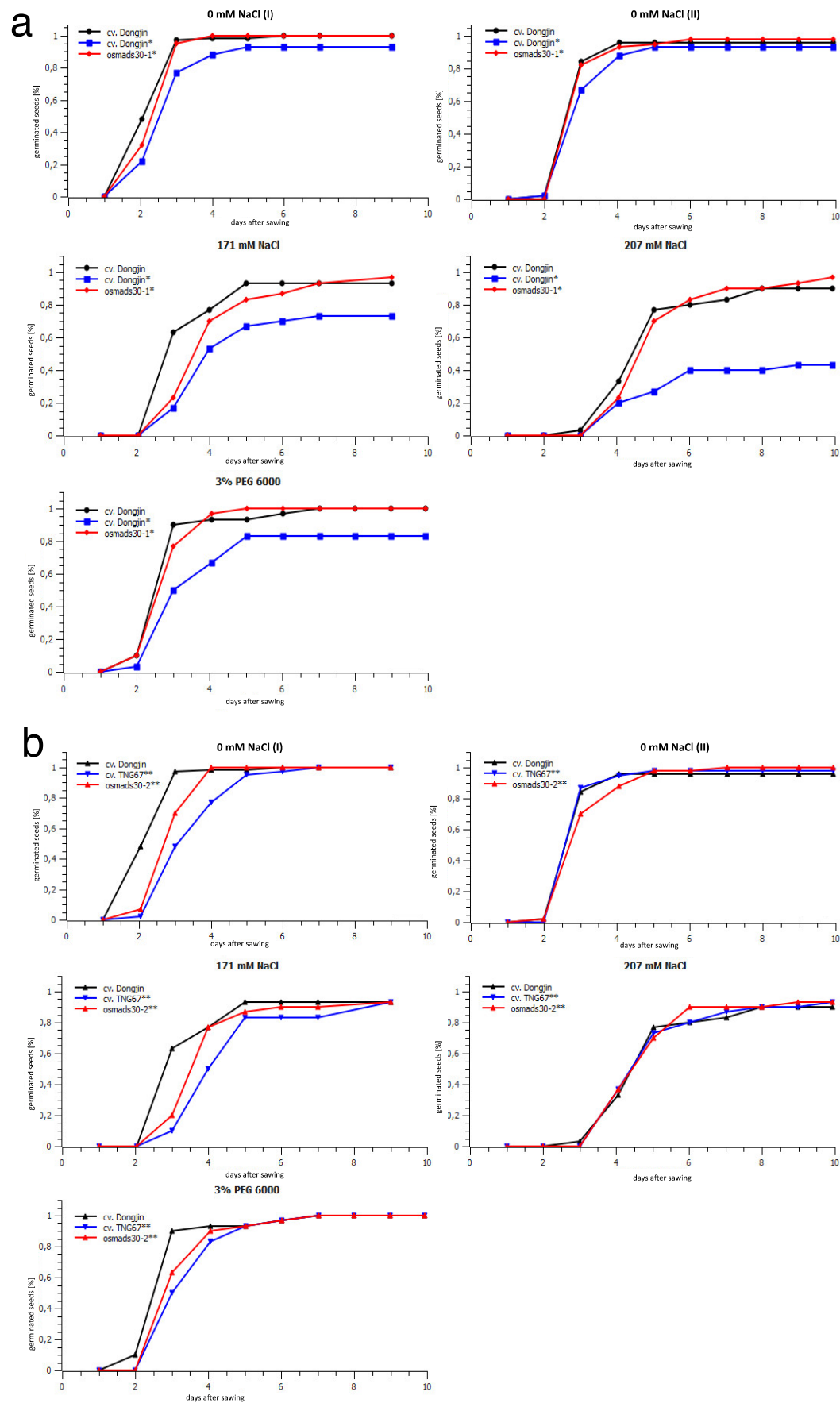
Figure S11. Grain germination of *Oryza sativa OsMADS30* T-DNA insertion lines and control plants. The fraction of germinated grains was determined for *osmads30-1* (a) and *osmads30-2* (b) under different concentrations of NaCl (0 mM (duplicated), 171 mM, 207 mM) and PEG 6000 (3%) daily over the course of ten days. Significant differences between the germination rate of the *Donjing* cv. and both *osmads30-1* and *osmads30-2* grains could not be detected. * and ** indicate that seeds were taken from sibling plants.

The differences in germination rate of *osmads30-1** and *Dongjin** under high salt conditions (a) is likely not correlated with *OsMADS30* since the effect cannot be seen in the second T-DNA line.
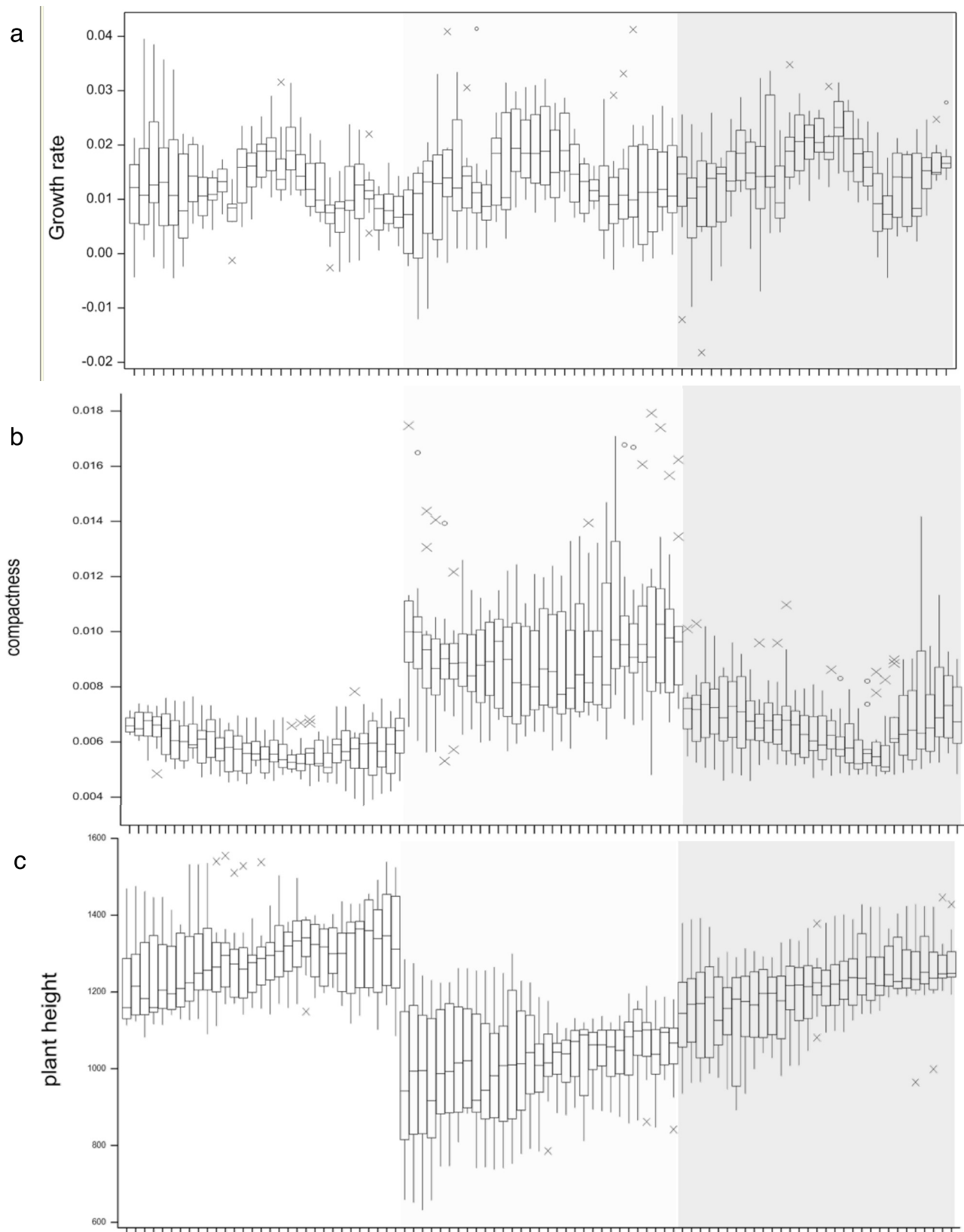
Figure S12. Growth rate (a), compactness (b) and plant height (c) of *OsMADS30* T-DNA insertion lines and wild type *Oryza sativa*. Boxplots for each day of imaging (117-147 days) of wild type control *Oryza sativa* plants of the cultivar Dongjin (white background) and *osmads30-1* (light grey background) and *osmads30-2* (dark grey background). (a) Growth rates did not significantly differ between wild type and mutant plants. (b, c) *osmads30-1* plants were found to have a significantly lower plant height (although being comparable in terms of projected leaf area) while being significantly more compact than *osmads30-2* and wild type plants which could be due to different plant morphological features such as differences in tiller or leaf number (see also Table S3, S4). Probably due to the lower plant height (with shorter leaves/tillers) the hull area of *osmads30-1* is furthermore significantly smaller than that of *osmads30-2* at almost all observed days (Supplementary Table 3). Since the projected shoot area (side) is only weakly different between the two mutants (significant only at days 145 and 147), the calculated compactness value (the fraction of the convex hull area filled with plant pixels) of *osmads30-1* is significantly higher than that of *osmads30-2*. This feature distinguishes *osmads30-1* also clearly (and significantly) from the wild type.

Supplementary Experimental Procedures

Phylogenetic Analyses – Identification of B$_{sister}$ genes

The B$_{sister}$ genes from *Musa accuminata* were taken from Gramzow and Theißen, 2013 (Gramzow and Theissen 2013) and named *MUSA ACUMINATA BSISTER1* and *2*, (*MaBS1*, and *MaBS2*). The B$_{sister}$ genes from *Elais oleifera* and *Phoenix dactylifera* were identified via BLAST searches (Altschul et al. 1990) of the genomes using *OsMADS29* as query (Al-Mssallem et al. 2013; Singh et al. 2013). They were termed *ELAIS OLEIFERA BSISTER* (*EoBS*) and *PHOENIX DACTYLIFERA BSISTER* (*PdBS*), respectively. The B$_{sister}$ gene of *Spirodela polyrhiza* was taken from Wang *et al.*, 2014 (Wang et al. 2014). Furthermore, we searched the transcriptome data of the 1kp project (http://onekp.com/) to identify monocot B$_{sister}$ genes. In frame of another project we identified all MADS-box genes in the transcriptome data using Hidden Markov Model searches (Gramzow and Theissen 2013) and identified B$_{sister}$ genes using phylogenies. Complete or almost complete B$_{sister}$ gene sequences were identified from *Posidonia australis*, *Maianthemum* spec., *Agave tequilana*, *Traubia modesta*, *Narcissus viridiflorus* and *Triodia aff* and termed *PaBS*, *MsBS*, *AteBS*, *TmBS*, *NvBS* and *TafBS*, respectively. The B$_{sister}$ genes from *Oryza* spec. were identified by BLAST (Altschul et al. 1990) searching the genomes of *O. glabberima*, *O. punctata*, *O. brachyantha*, *O. barthii*, *O. rufipogon*, *O. meridionalis* and *O. nivara* using *OsMADS29*, *OsMADS30* and *OsMADS31* as queries. Genes were named according to the *O. sativa* orthologs. The *OsMADS30* orthologs of *O. glaberrima* and *O. glumaepatula* did not contain a complete MADS box due to incomplete sequencing and possibly incorrect gene prediction, respectively and were hence not included in the phylogenetic analyses. The B$_{sister}$ genes from *Aegilops tauschii*, *Triticum urartu* and *Triticum aestivum* were identified by BLAST search (Altschul et al. 1990) of the respective genome and named according to species name and their position in

the phylogeny, BS1, BS2 or BS3 if they were orthologous to *OsMADS29, OsMADS30* or *OsMADS31*, respectively.

Selection analysis

Selection analysis was carried out using all grass B_{sister} genes that were used for the phylogeny. Due to high sequence similarity to other genes in the dataset all *Oryza* sequences except the ones from *O. sativa*, and all sequences from *Aegilops tauschii* and *Triticum urartu* were excluded from the analysis. Remaining protein sequences were aligned using ClustalW (Thompson et al. 1994). The resulting alignment was reverse translated using RevTrans (Wernersson and Pedersen 2003). This nucleotide alignment and the previously obtained phylogeny were used as input for PAML (Yang 2007). We used different subsets of the alignment, corresponding to the full alignment except the part encoding for the OsMADS30 C-terminal domain (MIKC), the full alignment without the part encoding for the C-terminal domain of all genes in the analysis (MIK), only the MADS-box and only the K-box. We also used different branch models, the one-ratio model allowing only one ratio for all branches in the phylogeny, a two-ratio model allowing the *OsMADS30* subclade to have a different ω-value to all the other branches and a four ratio model which has different ω-values for each of the three subclades, *OsMADS29, OsMADS30* and *OsMADS31* and the branch leading to the *OsMADS31* subclade. To find out which model fits the data best, Likelihood ratio tests were carried out, comparing the two ratio model with the one ratio model and the four ratio model with the one ratio model.

Imaging-based plant phenotyping

Seeds of the *osmads30-1* and *osmads30-2* mutant lines as well as *Oryza sativa ssp. japonica* cv. Dongjin wildtype were sown in trays filled with a 2:1 mixture of clay substrate (Klasmann-Deilmann GmbH, Germany) and vermiculite (1/2mm, Gärtnereibedarf Kammlott,

Germany) . Seedling were cultured in a climate controlled glass house at 24/17°C day/night, 60% relative air humidity, and 205–245 μmol m−2s−1 PAR with the light period set to 16 h (06:00–22:00 h). Due to the use of shading (when outside sun light exceeded 65 klux), total light intensity (natural sunlight + supplemental illumination) only rarely exceeded 380 μmol m−2s−1 PAR. For the analysis of growth dynamics and plant architectural as well as physiological features expressed upon cultivation in an acclimatized greenhouse. After an initial cultivation period of four months and before reaching the booting stage, ten plants per mutant and the *Oryza sativa* wild type were transferred to an automated phenotyping platform. After eleven weeks of pre-culture seedlings were transferred to 5.5l pots with clay/vermiculite mixture, stabilized with blue cages and the substrate mixture was covered with a blue rubber mat. Watering from the bottom was performed every second day and fertilizer was applied once per week (40ml/10l Wuxal (Manna, Germany)).

The plants were grown for further six weeks with daily imaging, weighing and watering. The relative growth rate was calculated as RGR = (log(LA[tn+4]) - log(LA[tn])) / ((tn+4) - tn) in a sliding window manner to represent the increase in projected leaf area (side view) every 4 days as described elsewhere (Poorter and Lewis 1986). Plant compactness was calculated as Compactness = 4∏ (projected shoot area/border pixels²) and the hull area represents the area (in pixels) of the convex hull, which is the shortest convex line drawing around the plant. Extracted values of selected traits were subjected to statistical analysis using GenStat (ANOVA and subsequent post-hoc analysis by Tukey's range test).

Additional References

Al-Mssallem, I.S., Hu, S., Zhang, X. et al. (2013) Genome sequence of the 11 date palm Phoenix dactylifera L. *Nat. Commun.* **4**, 2274.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–12 410.

Gramzow, L. and Theissen, G. (2013) Phylogenomics of MADS-box genes in plants – two opposing life styles in one gene family. *Biology* (Basel), **2**, 1150–14 1164.

Poorter, H. and Lewis, C. (1986) Testing differences in relative growth-rate – a method avoiding curve fitting and pairing. *Physiol. Plant.* **67**, 223–16 226.

Overview supplementary tables manuscript I

Table S1. Expression data of Bsister genes in grasses.

Table S2. Genes used for phylogeny.

Table S3. Extracted values of selected phenotypic traits.

Table S4. Imaging derived phenotypic traits of 3 rice genotypes for day 117 and day 147 after sowing.

Table S5. Primer and probes for quantitative RT-PCR.

(please find the tables on the CD attached to the thesis or online: http://onlinelibrary.wiley.com/doi/10.1111/tpj.13055/abstract)

2.3. Manuscript II

X. Yang, F. Wu, X. Lin, X. Du, K. Chong, L. Gramzow, <u>S. Schilling</u>, A. Becker, G. Theißen and Z. Meng (2012): **Live and Let Die – The B_{sister} MADS-Box Gene *OsMADS29* Controls the Degeneration of Cells in Maternal Tissues during Seed Development of Rice (*Oryza sativa*).** *PLoS One*, **7/12**, e51435-e51435.

PLOS | ONE

Live and Let Die - The B$_{sister}$ MADS-Box Gene *OsMADS29* Controls the Degeneration of Cells in Maternal Tissues during Seed Development of Rice (*Oryza sativa*)

Xuelian Yang[1,2,9], Feng Wu[1,9], Xuelei Lin[1,2], Xiaoqiu Du[1], Kang Chong[1], Lydia Gramzow[3], Susanne Schilling[3], Annette Becker[4], Günter Theißen[3]*, Zheng Meng[1]*

1 Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing, People's Republic of China, 2 Graduate School, Chinese Academy of Sciences, Beijing, People's Republic of China, 3 Department of Genetics, Friedrich Schiller University Jena, Jena, Germany, 4 Plant Evo-Devo Group, The Institute of Botany, Justus-Liebig-University Gießen, Gießen, Germany

Abstract

B$_{sister}$ genes have been identified as the closest relatives of class B floral homeotic genes. Previous studies have shown that B$_{sister}$ genes from eudicots are involved in cell differentiation during ovule and seed development. However, the complete function of B$_{sister}$ genes in eudicots is masked by redundancy with other genes and little is known about the function of B$_{sister}$ genes in monocots, and about the evolution of B$_{sister}$ gene functions. Here we characterize *OsMADS29*, one of three MADS-box B$_{sister}$ genes in rice. Our analyses show that *OsMADS29* is expressed in female reproductive organs including the ovule, ovule vasculature, and the whole seed except for the outer layer cells of the pericarp. Knock-down of *OsMADS29* by double-stranded RNA-mediated interference (RNAi) results in shriveled and/or aborted seeds. Histological analyses of the abnormal seeds at 7 days after pollination (DAP) indicate that the symplastic continuity, including the ovular vascular trace and the nucellar projection, which is the nutrient source for the filial tissue at early development stages, is affected. Moreover, degeneration of all the maternal tissues in the transgenic seeds, including the pericarp, ovular vascular trace, integuments, nucellar epidermis and nucellar projection, is blocked as compared to control plants. Our results suggest that *OsMADS29* has important functions in seed development of rice by regulating cell degeneration of maternal tissues. Our findings provide important insights into the ancestral function of B$_{sister}$ genes.

Background

Genetic and functional analyses of floral homeotic mutants in the model eudicot plants *Arabidopsis thaliana* and *Antirrhinum majus* led to the formulation of the ABC model, which was proposed to explain the determination of floral organ identities [1–2]. According to this model, class A genes specify the identity of sepals, class A and B genes specify petal identity, class B and C genes determine stamen identity, and class C genes determine carpel identity. Most of the floral homeotic genes that play a role in the ABC model are MIKC-type MADS-box genes. These genes were named after the domain structure of the encoded transcription factors consisting of a conserved DNA-binding MADS (M) domain, a less conserved intervening (I) region, a moderately conserved keratin-like (K) domain and a highly variable C-terminal (C) region [3–5].

In the ABC model, class B genes specify petal and stamen identity in combination with class A and C genes, respectively [2]. Several years ago, the sister clade of class B genes (*DEF/GLO*-like genes, also known as *AP3/PI*-like genes), termed B$_{sister}$ genes, was

identified in both angiosperms and gymnosperms [6]. Protein sequence alignments with other MADS-domain proteins indicated that compared to other MIKC-type proteins, the proteins encoded by B$_{sister}$ genes share a shorter I domain, a sub-terminal PI Motif-derived sequence and, in some cases, also a PaleoAP3 Motif in the C-terminal region with the AP3/PI-like proteins of gymnosperms and angiosperms [6].

In contrast to class B genes, which are predominantly expressed in male reproductive organs (and in angiosperm petals), B$_{sister}$ genes were found to be mainly transcribed in female reproductive organs (ovules, carpel walls) and in developing seeds [7–9]. Furthermore, the sequences of B$_{sister}$ genes are highly conserved. Together these findings suggest that B$_{sister}$ genes play an important role in ovule and/or seed development, which has been conserved for about 300 million years [6].

In *Arabidopsis*, two B$_{sister}$ genes, *ARABIDOPSIS BSISTER* (*ABS*; also known as *TRANSPARENT TESTA 16* (*TT16*) and *AGL32*) and *GORDITA* (*GOA*, formerly known as *AGL63*), were identified [9–12]. Expression analyses showed that *ABS* is expressed in the endothelial layer of the inner integuments of mature ovules [13].

In line with this, the *abs* (*tt16*) mutant seeds showed a loss of pigmentation and defect inner integument, indicating that *ABS* (*TT16*) regulates anthocyanidin accumulation and inner endothelial cell differentiation in *Arabidopsis* [9,12]. Recently it was shown that *ABS* acts redundantly with the class D floral homeotic gene *SEEDSTICK* (*STK*); *abs/stk* double mutants are characterized by a total absence of the endothelium and by massive starch accumulation in the embryo sac resulting in a low number of viable seeds [13].

The closest relative of *ABS* in *Arabidopsis* is *GOA*. *GOA* has a broader expression pattern compared to *ABS*. It has undergone neofunctionalization and has non-redundant functions to *ABS* in ovule outer integument development and the regulation of fruit longitudinal growth [10,11]. In *Petunia*, the *FLORAL BINDING PROTEIN 24* gene (*FBP24*) is expressed in ovule primodia, nucelli and integuments. Later, its expression is restricted to the endothelium of mature ovules and seeds. An *fbp24* knock-out mutant did not show alterations in development, indicating that *FBP24* acts redundantly with other genes. *fbp24* knock-down lines in which so far unknown other genes are likely co-suppressed, are affected in endothelium layer development similar to *abs* mutants, which suggests that the *FBP24* function is similar to that of *ABS* in *Arabidopsis* [8].

So far, functional studies have been conducted for B_{sister} genes from core eudicots only. However, expression patterns have been determined also for B_{sister} genes from a number of other species, as summarized in Figure 1. The B_{sister} gene *GGM13* of the gymnosperm *Gnetum gnemon* is expressed in the developing nucellus and inner envelope of female reproductive units [6]. Expression analyses of B_{sister} genes in monocots were carried out in several species. The wheat B_{sister} gene, *WBsis*, is expressed in the endothelial layer of the inner integument of the ovule [7], resembling the expression pattern of *ABS* in *Arabidopsis* [13]. However, weak expression of *WBsis* was detected also in the nucellus and in the outer integument. The maize B_{sister} gene *ZMM17* is initially expressed broadly in all organ primordia of the female spikelet, but at later developmental stages expression is restricted to the ovule and the developing silk [6]. Three B_{sister} genes, *OsMADS29*, *OsMADS30*, and *OsMADS31*, were identified in the rice genome [14]. Both RT-PCR and microarray analyses showed that transcription of *OsMADS29* is restricted to developing seeds, *OsMADS30* is expressed throughout all organs of the rice plant, and that expression of *OsMADS31* is below detection limit of the methods being used [14,15].

Since within angiosperms the monocot rice is only distantly related to the core eudicots *Arabidopsis* and *Petunia*, functional conservation and diversification of B_{sister} genes between monocots and eudicots can be investigated employing these species and along with this provide clues about the ancestral B_{sister} function. Here we investigate *OsMADS29*, because among the three B_{sister} genes of rice it is the only one exclusively and highly expressed during ovule and seed development resembling the expression pattern of B_{sister} genes of other species. We performed detailed expression analyses of *OsMADS29* and explored its functions in rice by a reverse genetics approach employing RNAi. Our results show that *OsMADS29* is mainly expressed in the developing ovule and seed. Knock-down of *OsMADS29* leads to aborted and/or shriveled seeds with deficient accumulation of starch in the endosperm. The severe phenotype of the *OsMADS29* knock-down lines suggests that, it plays an important role during rice seed development and controls at least some aspects non-redundantly with other genes, which is in contrast to typical findings about B_{sister} genes in eudicots. Our results thus corroborate many of the findings recently reported by Yin and Xue (2012) [16]. While these authors, however, focused on transgenic lines with a quite moderate reduction in *OsMADS29* mRNA levels, we report here in more detail the phenotypes of lines with stronger knock-down effects. Moreover, we provide novel insights into the evolution of the B_{sister} gene function and expression in flowering plants and reveal the crucial role of B_{sister} genes in the seed development.

Results

Phylogeny of B_{sister} Genes from Monocots

In order to gain insight into the evolutionary relationships between *OsMADS29* and other B_{sister} genes, a comprehensive search for B_{sister} genes of monocots was carried out in GenBank and Phytozome (www.phytozome.net) (Figure S1). A phylogenetic tree was then constructed with the identified genes and representative B_{sister} genes from core eudicots (Figure 1). The topology of the phylogenetic tree shows that B_{sister} genes of core eudicots and monocots constitute two different clades, suggesting that the most recent common ancestor of both taxa contained one B_{sister} gene only. The clade of grass B_{sister} genes is divided into three subclades, termed the *OsMADS29*, *OsMADS30*, and *OsMADS31* subclade. These clades comprise rice B_{sister} genes together with putative orthologs from other grass species such as barely (*Hordeum vulgare*), wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), maize (*Zea mays*), *Brachypodium distachyon*, and *Setaria italica* (Figure 1), indicating that the gene duplication events that generated the three clades occurred before grass diversification.

OsMADS29 is Specifically Expressed in Ovules and Developing Seeds

To analyze the expression patterns of *OsMADS29* in detail, RT-PCR, quantitative reverse transcription PCR (qRT-PCR), and RNA *in situ* hybridization analyses were performed using diverse tissues at different developmental stages.

RT-PCR analyses show that the expression of *OsMADS29* initiates at panicles 0.1~5 cm, then increases gradually with the development of panicles at 6–22 cm. After pollination, the expression level reaches the maximum value at 5 DAP-7 DAP, then decreases from 9 DAP. No expression of *OsMADS29* is detected in the vegetative organs (Figure 2A). Results of qRT-PCR show similar expression profiles of *OsMADS29* to those of the RT-PCR (Figure 2B).

RT-PCR was also performed to examine the expression of *OsMADS29* in different floral organs. Like other known B_{sister} genes, *OsMADS29* is specifically expressed in the pistils (Figure 2C).

In situ hybridization was performed to locate the *OsMADS29* expression at the cellular level (Figure 2D–L). Transcripts of *OsMADS29* are localized in the ovule, including integuments (IN) and nucellus (NU) throughout ovule development, from the primordial stage to maturity (Figure 2D and H), as well as in the ovule vasculature. After pollination, *OsMADS29* is expressed evenly throughout the whole seed (Figure 2I and K) and the embryo (EM) (Figure 2E, F and G) except for some cell layers in the outer epidermis of the pericarp. Notably, the gene is expressed in some tissue types which degenerate at later stages of seed development. These tissues include the inner epidermis of the pericarp (Figure 2H and I, indicated by an arrow), the tube-cells (TC) and cross-cells (CC) which differentiate from the inner epidermis at 7 DAP (Figure 2L), the inner and outer integument II (Figure 2D and H), the ovular vascular trace (OV) (Figure 2D, E, H, I and K), the nucellar projection (NP) (Figure 2I, K), and the nucellar epidermis (NE) (Figure 2I, K and L).
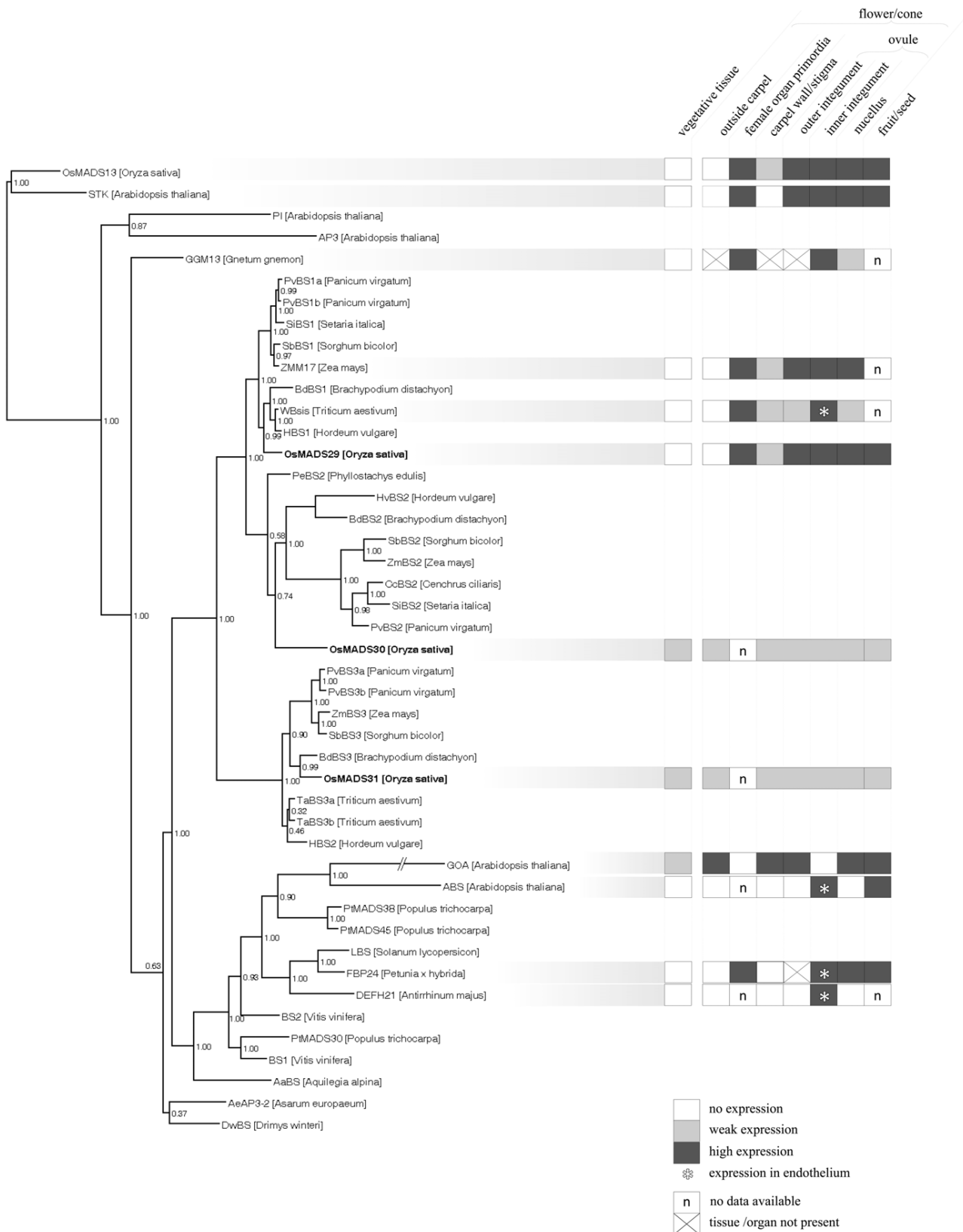
Figure 1. Phylogeny and expression patterns of B_sister genes. Bayesian phylogeny of class D, class B and B_sister genes in seed plants. Posterior probabilities are indicated on the nodes. The expression patterns of B_sister and class D genes are depicted as far as they are known. Colors indicate expression intensity: white, no expression; light grey, weak expression; dark grey, strong expression; asterisk, expression was detected particularly in endothelial cells of inner integument; n, no expression data available; cross, tissue or organ is not present in respective species.
doi:10.1371/journal.pone.0051435.g001
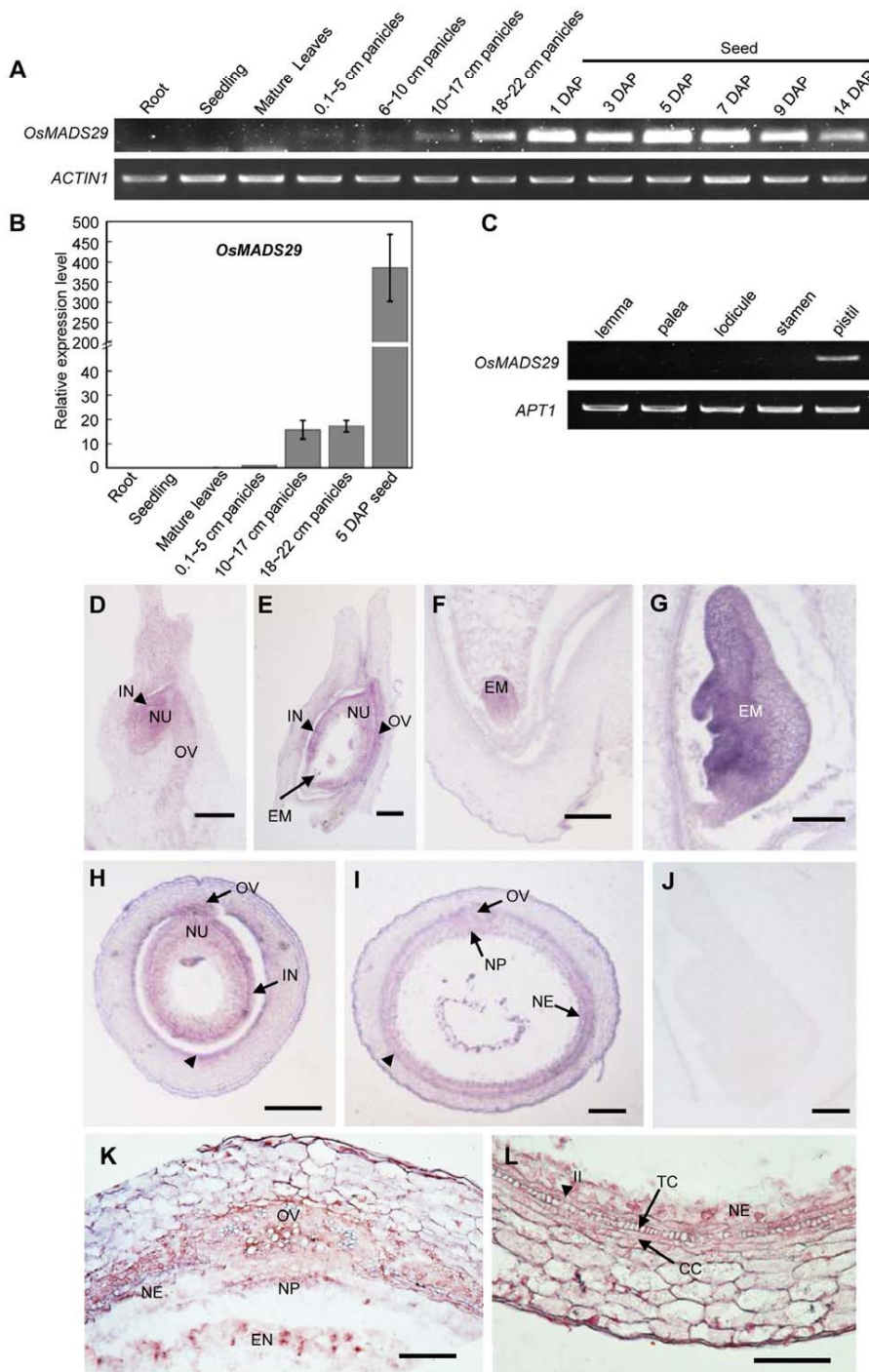
Figure 2. Expression analyses of *OsMADS29*. (A). RT-PCR analyses of *OsMADS29* expression at different developmental stages of wild type plants. DAP, days after pollination. *ACTIN1* was used as control. **(B).** Relative expression levels of *OsMADS29* at different developmental stages. Amounts of transcripts are shown as relative values to those of *UBQ*. Error bars show SD (n = 3). **(C).** RT-PCR analyses of *OsMADS29* expression in various floral organs of wild type plants at heading date stage. *APT1* was used as a control [47]. **(D–L).** *In situ* hybridization analyses of *OsMADS29* in ovules and developing seeds. **(D).** Longitudinal section of the ovule. **(E–G).** Longitudinal sections of seed at 1 DAP (E), and embryo of seed at 3 DAP (F) and 5 DAP (G). **(H–I).** Transverse sections of the mid-region of the ovary at a few hours before anthesis (H) and seed 3 DAP (I). The arrows head in (H) and (I) indicate the inner layer cells of the pericarp. **(J).** Negative control with sense probes of *OsMADS29*. **(K–L).** A magnification of the transverse section of a seed at 7 DAP. (K). The ovular vascular trace in the seed at 7 DAP. (L). Part of pericarp of seed at 7 DAP. CC, cross-cells; EM, embryo; EN, endosperm; IN, integuments; II, inner layer of the inner integument; NE, nucellar epidermis; NP, nucellar projection; NU, nucellus; OV, ovular vascular trace; TC, tube-cells. Bars = 100 μm in (D) to (J), and 50 μm in (K) and (L).
doi:10.1371/journal.pone.0051435.g002

Silencing of *OsMADS29* Results in Seed Abortion and Deficiency in Starch Accumulation in the Endosperm

To explore the functions of *OsMADS29*, we generated transgenic plants in which *OsMADS29* expression is specifically silenced at different levels by RNAi (Figure 3J and K). Statistical evaluations of data from mature grains shows that there is a usual abortion rate of about 9% in the control plants (Figure 3L). In contrast, of fifteen independent transgenic lines, five lines with strong *OsMADS29* down regulation produced strong phenotypes with 100% abnormal seeds, including, on average, 66% shriveled seeds (mildly affected seeds with deficient starch accumulation in the endosperm, defined as m-seeds) (Figure 3D, E, F, L and Figure 4B) and 34% aborted seeds (severely affected seeds which are fully aborted before 7 DAP, s-seeds) (Figure 3G, H, I, L and Figure 4C). Seven lines with moderate *OsMADS29* down regulation exhibited phenotypes with 63% abnormal seeds, including, on average, 25% m-seeds and 38% s-seeds, and 37% normal grains (n-seeds) (Figure 3L). In the three lines with weak down regulation, most of the grains are n-seeds except for, on average, 17% s-seeds (Figure 3L).

Considering that rice seeds reach their maximum length at about 5–6 DAP [17], 7 DAP were chosen as first time point for investigation of seeds. At 7 DAP, the s-seeds are aborted already (Figure 3G), while the m-seeds are still similar to the control seeds (Figure 3A) except for being slightly smaller in size (Figure 3D). At 10 DAP, unlike n-seeds that show an increasing width, s-seeds did

not show any further development (Figure 3H), while the apical part of the m-seeds reveals initial signs of shrinking possibly caused by a failure to accumulate sufficient levels of starch (Figure 3E). One month later, compared with the fully developed n-seeds, the s-seeds did not show much further development compared to 7 DAP (Figure 3I), and the mature m-seeds appear shriveled with wrinkles on their pericarps (Figure 3F).

The agronomic traits of seeds of control and *OsMADS29* RNAi transgenic lines were analyzed except for the s-seeds that had been aborted at 7 DAP (Table 1). In line 5 (L5), in which the *OsMADS29* expression was severely down-regulated, the length of m-seeds is 4.84 mm on average, while control seeds are about 16.7% longer (5.65 mm). The average width of control seeds was 3.15 mm, about 18.9% wider than that of the m-seeds (2.65 mm). The average thickness of m-seeds (1.00 mm) was reduced to about 50% of that of the control (1.98 mm) (P-value <0.01). These differences in length, width, and thickness are probably caused by a deficiency in starch filling. As expected, the 1000-grain-weight is also extremely reduced in m-seeds (2.15 g), with the weight being only about 10% of that of the control seeds (23.48 g) (P-value <0.01).

To correlate the phenotypes with *OsMADS29* transcript abundance, qRT-PCR experiments were performed. As shown in Figure 3J, the transcript level of *OsMADS29* was highly reduced to 26.4% of that of the wild-type in L5 exhibiting a strong phenotype, significantly reduced to 35.3% in L8 with a moderate
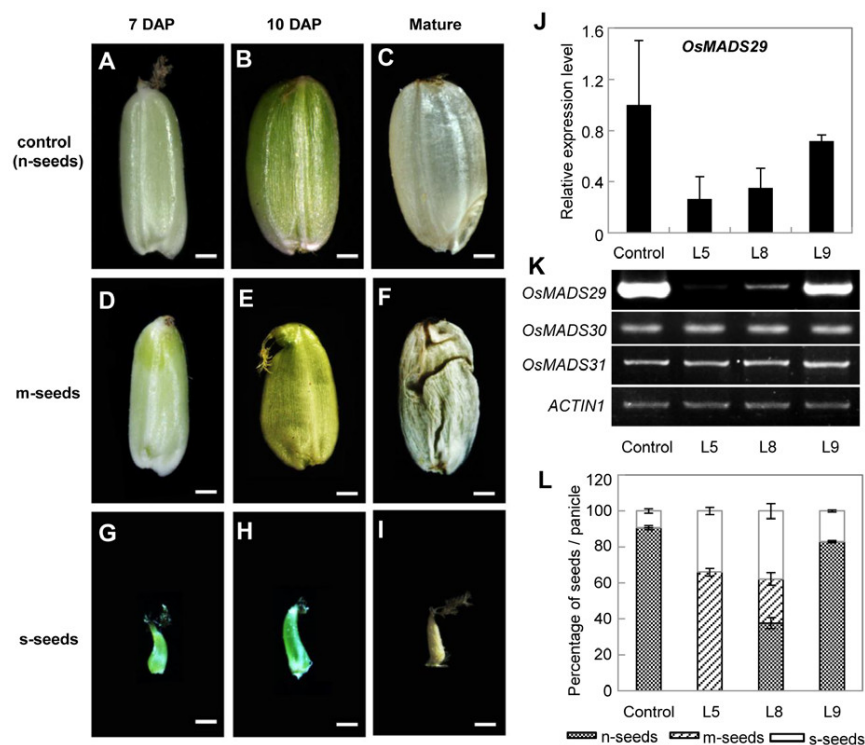
Figure 3. Phenotypes of *OsMADS29* RNAi lines and molecular detections. (A–I). Profiles of seeds at 7 DAP, 10 DAP and mature stages, respectively. (A–C), n-seeds of control plants. (D–F), m-seeds with deficient starch accumulation in the endosperm. (G–I), s-seeds that are aborted before 7 DAP. Bars = 1 mm. **(J).** Relative expression levels of *OsMADS29* in control plants and three typical *OsMADS29* RNAi lines (L5, L8 and L9). Total RNA was pooled from stage Ov10 [48]. The *ACTIN1* was used as the internal control. Error bars show SD (n = 3). **(K).** RT-PCR analyses of *OsMADS29* and the other two related B$_{sister}$ genes of rice, *OsMADS30* and *OsMADS31*, in control plants and RNAi plants. Total RNA was pooled from stage Ov10. The *ACTIN1* was used as the internal control. **(L).** Percentage of different kinds of seeds in one panicle in control plants and different *OsMADS29* RNAi lines. The error bar indicates the SE (n = 3).
doi:10.1371/journal.pone.0051435.g003

Table 1. Agronomic traits of seeds in control plants and *OsMADS29* RNAi transgenic lines.

	Length (mm)	Width (mm)	Thinkness (mm)	1000-grain-weight (g)
Control (n-seeds)	5.65±0.17	3.15±0.13	1.98±0.06	23.48±0.14
L5 (m-seeds)	4.84±0.23**	2.65±0.17	1.00±0.04**	2.15±0.10**
L8 (m-seeds)	5.61±0.26	2.86±0.18	1.31±0.05**	5.45±0.06**

The 1000-grain-weight of L5 and L8 was measured by using the deficient filled seeds. **$P < 0.01$ when compared with the control seeds.
doi:10.1371/journal.pone.0051435.t001

phenotype, and slightly reduced to 71.7% in L9 showing only a weak phenotype. In contrast, the expression levels of the other two B_{sister} genes in rice, *OsMADS30* and *OsMADS31* remained unchanged in the *OsMADS29* knock-down lines (Figure 3K). These results indicate that the observed phenotypes of the transgenic plants are caused by specific silencing of *OsMADS29*.

Given that the starch content is substantially decreased in mature seeds of transgenic plants (Figure 4B and C), the expression levels of four genes, including *OsAGPS1*, *OsAGPL2*, *OsAGPL3*, and *OsAGPS2a* which encode the rate-limiting-step enzymes in the starch synthesis pathway [18], were examined in developing seeds of control plants and *OsMADS29* RNAi lines at 10 DAP. qRT-PCR results showed that the expression levels of those four genes in transgenic seeds are similar to the ones of the control seeds (Figure S2), indicating that starch synthesis itself is not affected.

Silencing of *OsMADS29* Affects Vascular Trace Development and the Degeneration of Tissues in Rice Grains

Down regulation of *OsMADS29* does not affect ovule development but causes seeds to abort or shrivel as they are most likely deficient in endosperm development and starch accumulation (Figure S3F, Figure 3A–I, Figure 4B and C). To observe the internal organization of developing seeds, mid-region transverse sections of seeds at 7 DAP were made. While the seeds of control plants are entirely filled with endosperm cells that produce starch granules (Figure 4A), no endosperm cells can be observed in the m- and s-seeds of the transgenic lines at 7 DAP (Figure 4B and C). In control plants, the xylem cells are large, have a circular shape and are centered in the middle of the ovular vascular trace (OV), and the phloem cells are small and arranged in a cluster at the periphery of the OV (Figure 4D). In the s-seeds, the xylem cells are smaller in size and fewer in number than that of the control seeds. Also a lower number of phloem cells can be seen at the margin of the OV (Figure 4F). In m-seeds, the size and number of xylem cells are similar to those of the s-seeds, whereas arrangement style of phloem cells is scattered in contrast to the phloem cell clusters of the control plants (Figure 4E). Moreover, the number of cell layers of the OV in the transverse orientation is higher in the *OsMADS29* RNAi transgenic lines than in the control lines (marked by double arrows brackets in Figure 4D–F), suggesting that adjacent cells in the OV probably fail to integrate into the xylem and phloem. Another tissue type recognized as abnormal in the transgenic lines is the nucellar projection (NP), which is a small zone of persisting nucellar cells attached to the chalaza. The NP cells degenerate in the control seeds at 7 DAP such that of originally 4–5 cell layers only 2–3 layers remain. The respective cells are small and flat, and constitute a tissue required for nutrient transfer to the endosperm and embryo (Figure 4D; Figure S4). In contrast, the NP cells of m-

and s-seeds at 7 DAP remain large and circular, without any obvious indication of cell degeneration (Figure 4E and F).

In mature grains, the structure of the embryo was not affected in the m-seeds, except for the fact that the embryos were considerably smaller than in control seeds (Figure 5E and H). Notably in the n-seeds at 30 DAP, the maternal tissues (pericarp, ovular vascular trace, integuments, and nucellar epidermis) were almost degenerated to form the cuticula (Figure 4G, I and K) along with the maturity of filial tissue (endosperm and embryo). In contrast, all the maternal tissues remained in the mature seeds in the *OsMADS29* RNAi transgenic lines at this stage (Figure 4H, J and L).

Longitudinal histological sections of seeds at 7 DAP showed that development of the embryo in transgenic seeds severely lagged behind that of the wild-type. Compared to the completed embryo morphology in control seeds at 7 DAP (Figure 5A), embryos of s-seeds at 7 DAP appear to cease development at a stage corresponding to 3 DAP of wild-type seeds, and remain in this stage (Figure 5B and Figure S3C); embryos of m-seeds at 7 DAP grew retarded to a stage corresponding to almost 5 DAP of control seeds (Figure 5C and Figure S3D). In mature grains, the embryo of m-seeds is considerably smaller than the one in control seeds (Figure 5E and H). Considering that the mature embryo of m-seeds failed to germinate even on the 1/2 MS culture medium (Figure 5G), mid-region transverse sections of embryos were carried out to thoroughly observe the structure of the vascular system, which also developed abnormally in m- and s-seeds at 7 DAP. While vascular bundles are obvious between the shoot apical meristem and radicle of control embryos, no vascular bundles were observed in the embryos of m-seeds (Figure 5F and I).

Previous studies have shown that the differentiation of the tracheary elements as well as the degeneration of nucellus and pericarp cells are also brought about by PCD [19–21]. To analyze the relationship between *OsMADS29* and cell degeneration caused by PCD during seed development, an Evan's blue staining experiment, which can dye the dead cells to be blue, was performed. Results showed that cell death of the NP and endosperm tissues in the m-seeds obviously occurs later than that of control seeds during the stage from 15 DAP to 26 DAP (Figure 6A–H). In addition, qRT-PCR expression analyses were carried out with three classes of genes known to positively regulate PCD through three different pathways, including *VPE1-4* [22–24], *VDAC1-3* [25,26], and *PBZ1* [27,28]. Our results show that *OsVPE1* (Os04g45470) is the only gene which is significantly down-regulated in *OsMADS29* knock-down transgenic seeds (Figure 6I). A putative CArG-box with the consensus sequence $C(A/T)_8G$ was found at position −1827 in the upstream region of *OsVPE1* (Figure S5). As MADS-domain proteins have been shown to bind to CArG-boxes of their target genes to regulate expression [29], this may indicate direct regulation of *OsVPE1* by *OsMADS29*.
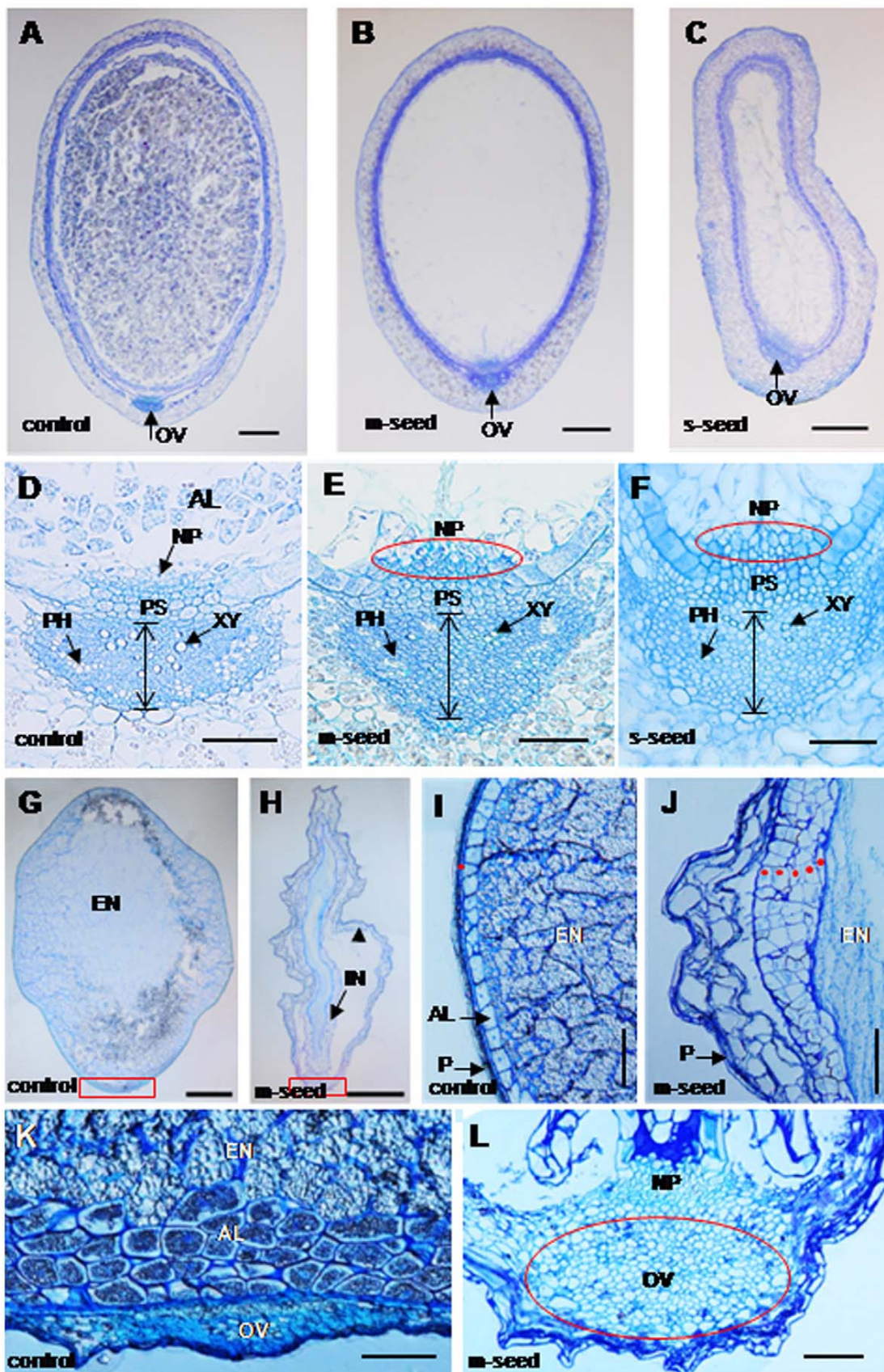
Figure 4. Histological analyses of control and transgenic seeds. (A). Starch granules were seen in control seeds at 7 DAP. **(B, C).** m- and s-seeds at 7 DAP without starch granules. **(D–F).** A higher magnification of the OV at 7 DAP in the control seeds (D), m-seeds (E) and s-seeds (F), respectively. The double arrow between the two transverse lines indicates the cell layers of the OV. **(G, H).** Semi-thin transverse sections from the mid-region of mature control seeds (G) and m-seeds (H) at 30 DAP. The red rectangle in (G) and (H) indicates the position of ovular vascular trace. The arrowhead in (H) indicates the remained pericarp. **(I, J).** Higher magnification for part of the pericarp in mature control seeds (I) and m-seeds (J) at 30 DAP. It shows that the pericarp cuticular layer in (I) and persisted pericarp cells in (J). The red dots indicate the cuticular seed coat in control seeds (I) and the cells of integuments and nucellar epidermis in m-seeds are not degenerated (J). **(K, L).** Higher magnification of the OV in control seeds (K) and m-seeds (L), which showed in the red rectangle in (G) and (H). The red circle in (L) indicates the large OV in m-seeds. AL, aleurone; NP, nucellar projecton; OV, ovular vascular trace; P, pericarp; PH, phloem; PS, the pigment strand; XY, xylem; EN, endosperm. Bars = 200 µm in (A–C), 50 µm in (D–F), 500 µm in (G) and (H), 100µm in (I) and (J), and 50 µm in (K) and (L).
doi:10.1371/journal.pone.0051435.g004

Discussion

OsMADS29 is a Canonical B$_{sister}$ Gene in Rice

As revealed by phylogenetic analyses previously, B$_{sister}$ genes are the sister clade of the *DEF/GLO*-like (or *AP3/PI*-like) genes comprising class B floral organ identity genes [6]. Our phylogeny suggest that the two gene duplications that gave rise to the three B$_{sister}$ genes *OsMADS29*, *OsMADS30* and *OsMADS31* in rice occurred after the divergence of the lineages that led to extant eudicots and monocots about 150 MYA [30], but predated the diversification of grasses from a common ancestor about 55–70 MYA [31–33]. In our phylogeny, the branch lengths are relatively short for all genes belonging to the *OsMADS29* clade as compared to the branch lengths for the genes belonging to the other two monocot B$_{sister}$ clades (Figure 1). This suggests that purifying selection is acting on the genes in the *OsMADS29* clade. Furthermore, the expression pattern of the genes in the *OsMADS29* subclade seems to be conserved, as the expression of *OsMADS29* in the developing ovule including integuments and nucellus is similar to that of the previously described expression patterns of *WBsis* and *ZMM17* (Figure1) [6,7]. Taken together, this provides evidence for the functional importance of the genes in the *OsMADS29* clade. Protein sequence alignments of the C-terminal domains show that OsMADS29 and OsMADS31 are typical B$_{sister}$ proteins with respect to their protein sequence as they both share a conserved sub-terminal 'PI Motif-Derived sequence' and a terminal 'Paleo AP3 Motif' with other B$_{sister}$ proteins. In contrast, OsMADS30 does not include these motifs characteristic for B$_{sister}$ proteins (Figure S6). Expression analyses reveal that *OsMADS31b* (the alternative splice isoform of *OsMADS31* reported by Lee *et al.* (2003) [15]) is broadly expressed during all developmental stages of rice, whereas *OsMADS30* expression is hardly detectable at all (Figure S7), which are unusual expression patterns for B$_{sister}$ genes. Only *OsMADS29*, which is specifically expressed in mature florets, especially in pistils and in developing seeds, exhibits an expression pattern typical for B$_{sister}$ genes (Figure 2A–C). After pollination, *OsMADS29* is highly expressed in developing seeds (Figure 2D–L). Expression during seed development has been reported for *ABS*, *GOA* and *FBP24* as well [8–10]. These results suggest functional conservation of B$_{sister}$ genes between eudicots and monocots in ovule and seed development. Furthermore, our study shows that *OsMADS29* is probably the canonical representative of these functions in rice.

Conserved Functions of B$_{sister}$ Genes in Ovule and Seed Development

Given that all B$_{sister}$ genes investigated so far, including *OsMADS29*, are expressed during early ovule development, it was hypothesized that members of the B$_{sister}$ subfamily play important roles in female reproductive organ (ovule) development [34]. In our study, down-regulation of *OsMADS29* does not affect the structure of the ovule (Figure S3A and F). Accordingly, either *OsMADS29* is not required for ovule specification in rice or its

function in ovule development is obscured by redundancy with another gene, e.g. with *OsMADS21*, which is a class D gene. The expression of *OsMADS21* overlaps with *OsMADS29* and single mutants of *osmads21* do not have a mutant phenotype [35]. Redundancy between B$_{sister}$ and D class genes would resemble recent findings for *ABS/TT16* and *STK* in *Arabidopsis* [13]. Redundancy with the other two B$_{sister}$ genes *OsMADS30* and *OsMADS31* may also obscure the spectrum of functional significance of *OsMADS29* during ovule development. Recently, some microarray data about ovule development can be used from the public database, which can give us some new clues for the ovule development of *OsMADS29* in our future work [36].

In contrast to ovule development, the effect on seed development is stronger for *OsMADS29* than for the *Arabidopsis* B$_{sister}$ genes. While the phenotypes of the single mutants *abs* (*tt16*) and *goa* of *Arabidopsis* are mild, given that they still produce seeds which germinate properly [9,11,12], silencing of *OsMADS29* in rice leads to a severe phenotype with sterile seeds which are aborted or shriveled. The mild phenotype of the *abs* single mutant may again be explained by redundancy of *ABS* (*TT16*) and *STK* in *Arabidopsis* [13]. In rice, this redundancy between B$_{sister}$ and class D genes in seed development may have been lost or not gained in evolution, leading to a severe phenotype in *OsMADS29* knock down lines. Compared to control seeds which are full of starch granules, cross-sections of abnormal seeds at 7 DAP show no visible starch granules. Interestingly, starch accumulation is also deregulated in the *abs/stk* double mutant. However, in this case, an excess of starch was observed in the embryo sac and the developing seed [13]. Although some shriveled seeds of the *OsMADS29* knock down lines have embryos with normal shape, they have no vascular bundles and fail to germinate, possibly because of the lack of nutrition transport (Figure 5D–I).

The mutant phenotypes of B$_{sister}$ genes from eudicots and rice differ during ovule and seed development. This may indicate that the function of *OsMADS29* has diverged from that of B$_{sister}$ genes in eudicots during the evolution of monocots. However, the differences in the mutant phenotypes can also, at least partially, be explained by redundancy to other genes. Hence, it is also possible that the function of *OsMADS29* resembles the ancestral function of B$_{sister}$ genes while this ancestral function is masked in eudicots due to other genes with redundant functions.

OsMADS29 is Required for Vascular Trace Development and Cell Degeneration in Rice Seeds

Previous studies have shown that there is a symplastic continuity, including the ovular vascular trace (OV), pigment strand (PS) and nucellar projection (NP), which is the nutrient source for the filial tissues (endosperm and embryo), between the rice maternal tissues and the filial tissues [37–39]. Therefore the path among the OV, PS and NP plays an important role in grain filling and seed development. In control plants, the nucellus was absorbed by the filial tissues during early seed development (Figure 4D). In contrast, the structures of the nucellus, including
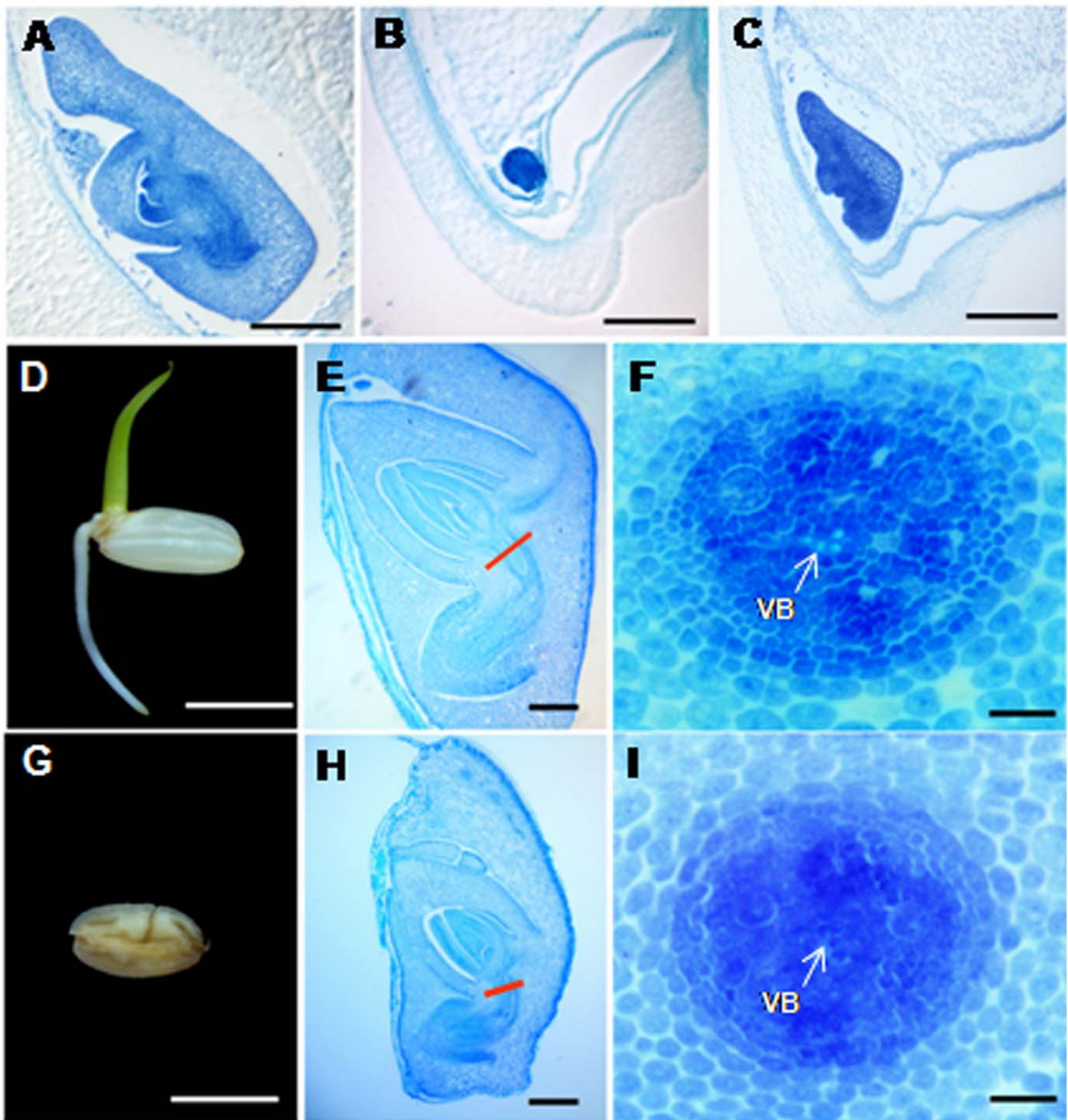
Figure 5. Histological structure of embryo and germination of mature seeds. (A). Longitudinal sections of control embryo at 7DAP. **(B, C)** Longitudinal sections of embryo of s-seeds (B) and m-seeds (C) at 7 DAP, respectively. **(D).** The normally germinated seed of control plant. **(E).** Longitudinal sections of the mature embryo in control seed. The red line referred to the position of vascular bundle. **(F).** Magnifying observation of cross sections from the red line marked parts in (E). The arrow showed the clear vascular bundle. **(G).** The m-seeds failed to germinate. **(H).** Longitudinal sections of the mature embryo in m-seeds. **(I).** Magnifying observation of cross sections from the red line marked parts in (H). No obvious vascular bundles were observed at the arrow referred place. VB, vascular bundles. Bars = 20 μm in (A–C), 5 mm in (D) and (G), 0.2 mm in (E) and (H), 20 μm in (F) and (I).
doi:10.1371/journal.pone.0051435.g005

NP and NE are still visible in the transgenic grains at 7 DAP and in mature seeds (Figure 4E, F, H, J and L). This indicates that the process of degeneration of their cells is partially suppressed. Similar results concerning the defect structure of NE and NP have recently been reported by Yin and Xue (2012) when this

manuscript was in preparation [16]. However, we also observed defects in other tissues, such as the OV, which may lead towards a more comprehensive understanding of the function of *OsMADS29*. In the transgenic abnormal seeds at 7 DAP, the shapes of the OV are irregular and the cell numbers of xylem and phloem are less
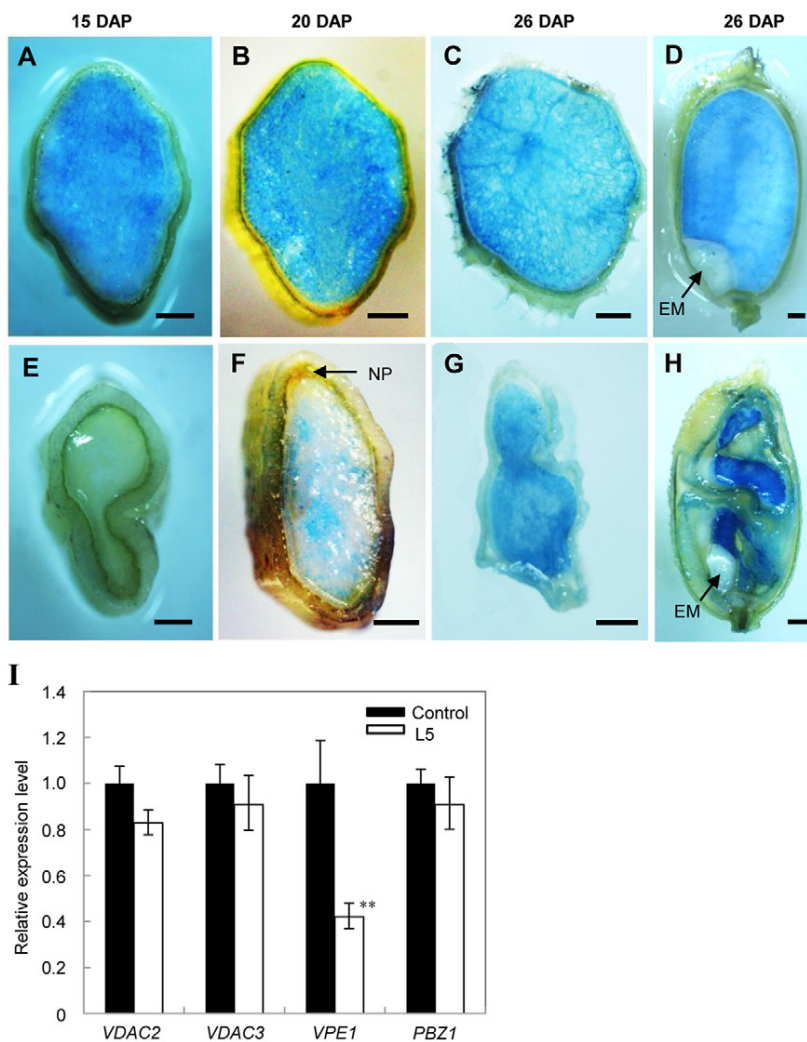
Figure 6. Detections of tissue degeneration. (A–D). Evan's Blue staining of control seeds. (A-C) Transverse sections at 15 DAP, 20 DAP, and 26 DAP, respectively. (D) The longitudinal section at 26 DAP. The blue tissues indicate dead endosperm cells. **(E–H).** Evan's Blue staining of m-seeds. (E–G) Transverse sections at 15 DAP, 20 DAP, and 26 DAP respectively. H Longitudinal sections at 26 DAP. Bar = 500 μm in (A) to (H). EM, embryo. **(I).** Quantitative real-time PCR analyses of PCD-related genes in 7 DAP seeds of control and transgenic plants. *ACTIN1* was used as an internal control. Error bar indicated the SD (n = 3). ** indicates an extremely significant difference (P<0.01) between control and transgenic lines. doi:10.1371/journal.pone.0051435.g006

than those in the control plants (Figure 4E, F), indicating that transport paths for nutrients are strongly affected. Xylem and phloem are differentiated from the cambium cells, revealing that their formation is a process involving PCD [21]. Hence, fewer cells in the xylem and phloem of transgenic abnormal seeds at 7 DAP imply that cell degeneration of the tracheary tissue (including xylem and phloem) is suppressed to a great extent. This further indicates that the defect tracheary system may be another important cause, besides the failed degradation of NE and NP, for the deficient starch accumulation, as no tracheary system is observed in the m-seeds at mature stage. Therefore, the loss of nutrient transport may be the possible direct reason for the small embryo and the germination failure. This result also shows that *OsMADS29* affects not only the cell degeneration of maternal tissues but also that of filial tissues. Furthermore, the pericarp cells in the *OsMADS29* transgenic m-seeds remain in their original state without any sign of degeneration (Figure 3B and E). However, the pericarp of the wild type mature seed is degenerated in a PCD

dependent process to become the cuticula which protects the filial tissue (Figure 4I) [19]. Collectively, these results show that the degeneration of cells is localized in those tissues of seeds where gene expression of *OsMADS29* is found.

As expected, cell degeneration was indeed shown to be suppressed during seed development of transgenic plants. Evan's blue staining indicates that cell death of NP and endosperm tissue in transgenic seeds is obviously delayed as compared to that of control seeds (Figure 6A–H), suggesting that the process of PCD is delayed and suppressed to some extent. Since molecular markers for cell death in seeds are not available at present, the expression of three kinds of genes, *VPEs*, *VADC*, and *PBZ1*, known to be involved in PCD through different pathways, was investigated. Previous studies showed that the *VPEs* are involved in vacuole mediated cell death in both defense and development, regulating PCD in *Arabidopsis*, rice, and barley [22–24]. Only four *VPE* genes were found in rice, among which *OsVPE1* and *OsVPE4* are seed-type genes,

whereas *OsVPE2* and *OsVPE3* are vegetative-type genes [22]. In this study, only the transcript amount of *OsVPE1* was found to be substantially reduced in abnormal transgenic seeds (Figure 6I). In contrast, the expression levels of the *VDAC* genes, which are elements of the mitochondrial death machinery [25,26], and *OsPBZ1*, which is involved in PCD induced by abiotic stresses [27,28], were the same in abnormal transgenic as in control seeds (Figure 6I). A Cys protease involved in regulating PCD of maternal tissues has been identified as a direct target gene of *OsMADS29* by Yin and Xue (2012), but the pathway of PCD regulated by *OsMADS29* has remained unknown [16]. Our molecular analyses together with the histological structures of the abnormal seeds in our transgenic plants suggest that PCD of the tracheary system and pericarp in rice caryopses regulated by *OsMADS29* occurs through a pathway mediated by vacuolar processing. The expression levels of genes encoding the rate-limiting-step enzymes of the starch synthesis pathway, *OsAGPS1*, *OsAGPS2a*, *OsAGPL2*, and *OsAGPL3*, are not affected (Figure S2). Taken together, our results reveal that the failure of the degeneration process in maternal tissues and thus insufficient development of xylem and phloem leading to an undersupply of filial tissue, rather than the starch synthesis process itself, is the original cause for the abnormal development of transgenic seeds. Interestingly, as mentioned above, also the eudicot B$_{sister}$ genes *ABS* and *FBP24* are functioning in the maternal tissue that supplies the filial tissue with nutrients [8,9,12,13]. Hence, developmental control of the tissue responsible for nutrient supply to the embryo may be one of the ancestral functions of B$_{sister}$ genes.

Our molecular analyses and structural observations provide evidence for the function of *OsMADS29* in cell degeneration during seed development and may enable a better understanding of the process of PCD during seed development. The severe phenotype of *OsMADS29* knockdown lines indicates that this B$_{sister}$ gene at least in part functions non-redundantly to other genes during seed development which is in contrast to the B$_{sister}$ genes of eudicots which have been studied so far. Furthermore, the expression pattern of *OsMADS29* is similar to that of the gymnosperm B$_{sister}$ gene *GGM13* and resembles, more or less, a combination of the expression patterns known from eudicot B$_{sister}$ genes. Hence, *OsMADS29* may have largely retained the ancestral expression pattern and its functions as revealed here may represent the ancestral function of B$_{sister}$ genes. On the other hand, different functions and regulatory mechanisms of eudicot B$_{sister}$ genes in cell expansion, pigment accumulation or integument specification suggest that B$_{sister}$ genes may have undergone sub- and neo-functionalization in various species during the evolution of flowering plants to control different aspects of seed development.

Methods

Plant Materials

The rice cultivar 'Zhonghua 11' (*Oryza sativa* L. ssp. *japonica*) was planted in local paddy-fields of the Institute of Botany, the Chinese Academy of Sciences.

Knockdown Vector Construction of *OsMADS29* and Rice Transformation

The *OsMADS29* full length cDNA clone (Accession NO. AK109522) was obtained from the Japan NIAS KOME stock center (http://cdna01.dna.affrc.go.jp/cDNA/). To generate the *OsMADS29* knock-down vector, a 488 bp long specific region (nucleotide positions 474 - 962 counted from ATG) was amplified, using the PCR primer pairs 29RNAiF (*Bam* HI and *Spe* I) and

29RNAiR (*Nco* I and *Hpa* I). The intron and nos terminator cassette of pJawohl3-RNAi (GenBank Accession No. AF404854) was transferred with *Bam* HI/*Not* I site to the pBluescript SK (Stratagene, La Jolla, CA, USA), termed pBJWl3 as an intermediate vector. The RNAi-fragment of *OsMADS29* was cloned into two sides of the intron region of the pBJWl3 vector in sense (*Bam* HI/*Nco* I) and antisense (*Spe* I/*Hpa* I) orientation, respectively. Finally, the dsRNAi cassette containing two oppositely orientated coding sequences and the nos terminator was mobilized with restriction enzymes *Bam* HI and *Sac* I, and was introduced into the vector pCAMBIA1301-Ubi, in which the maize (*Zea mays*) *Ubiquitin* promoter was inserted using the *Hind* III and *Bam* HI sites, resulting in the final plasmid pUOsM29I. The construct was verified by sequencing and restriction mapping. Rice transformation was performed according to the previous methods [40].

In situ Hybridization

Fresh wild-type flowers and seeds (1–7 DAP) were fixed immediately and embedded in paraffin (Sigma). 8 µm-thick sections were hybridized with the *OsMADS29* specific sequences (nucleotides 695–962 counted from the start codon ATG) according to the method described by Cui *et al.*, (2010) [40].

RT-PCR and Quantitative Real-time RT-PCR Analyses

Total RNA was isolated from different rice tissues using TRIzol reagent (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. Reverse transcription was performed using Superscript-III Reverse Transcriptase (Invitrogen, Carlsbad, USA). The diluted cDNA samples were used as templates for RT-PCR and real-time PCR. The internal control genes for RT-PCR were *ACTIN1* and *APT1* (*Adenine Phosphoribosyltransferase1*). Real-time PCR were performed using SYBR Premix Ex Taq (Takara, Dalian, China) on a Rotor-Gene 3000 (Corbett Research, QIAGEN, Hilden, Germany) detection system and software according to the manufacturer's instructions. The *ACTIN1* and *UBQ* were used as an internal control. Gene-specific primers are shown in Table S1.

Histochemical Analyses

Samples were embedded in paraffin (Sigma-Aldrich, St. Louis, USA) for histological sections (8-µm thickness) and in LR white resin (Sigma-Aldrich, St. Louis, USA) for semi-thin slices (1-µm thickness), respectively. Slices were stained with 0.1% toluidine blue and observed with a light microscope.

Evans Blue Staining

Fresh caryopses at different days after pollination were cut into longitudinal and cross-sections with a sharp double-edged blade. The sections were stained with Evans Blue solution (0.1% in H$_2$O) for 2 min and washed with water for 60 min, then photographed using a Leica microscope (Leica DM4500B).

Phylogenetic Analyses

B$_{sister}$ genes from monocotyledonous plants were searched using the blastn algorithm (Figure S1). The amino acid sequences of the three B$_{sister}$ proteins from *Oryza sativa*, OsMADS29, OsMADS30 and OsMADS31, were used as query sequences. The databases "Nucleotide collection (nr/nt)" and "non-human, non-mouse ESTs (est_others)" were searched. Furthermore, the whole genome sequences of *Setaria italica*, *Brachypodium distachyon*, *Sorghum bicolor* and *Panicum virgatum* were searched at phytozome (www. phytozome.net). The BLAST results were combined with known

B_{sister} genes and representative MIKC-type MADS-box genes of other clades in a neighbor joining tree constructed using PAUP* version 4.0b10 [41] to identify B$_{sister}$ genes and to remove redundancies. Amino acid sequences of known B$_{sister}$ proteins and the newly identified B$_{sister}$ proteins were aligned with ProbCons [42]. The ProbCons alignment was reverse translated into a nucleotide alignment using RevTrans1.4 [43]. The best fitting nucleotide substitution model to this alignment was determined using PAUP* version 4.0b10 [41] and Modeltest3.7 [44] where positions 1 to 24 and 508 to the end of the alignment were excluded. With the best-fitting model of nucleotide substitution (GTR+I+G) [45], a MrBayes v3.1.2 [46] phylogeny was determined, excluding the same positions as for the Modeltest, using AP3 of *Arabidopsis thaliana* as outgroup, generating 3,000,000 trees, sampling every 100th generation and discarding 7,500 of the sampled trees. The final phylogeny is a completely resolved consensus tree with Bayesian posterior probabilities.

Supporting Information

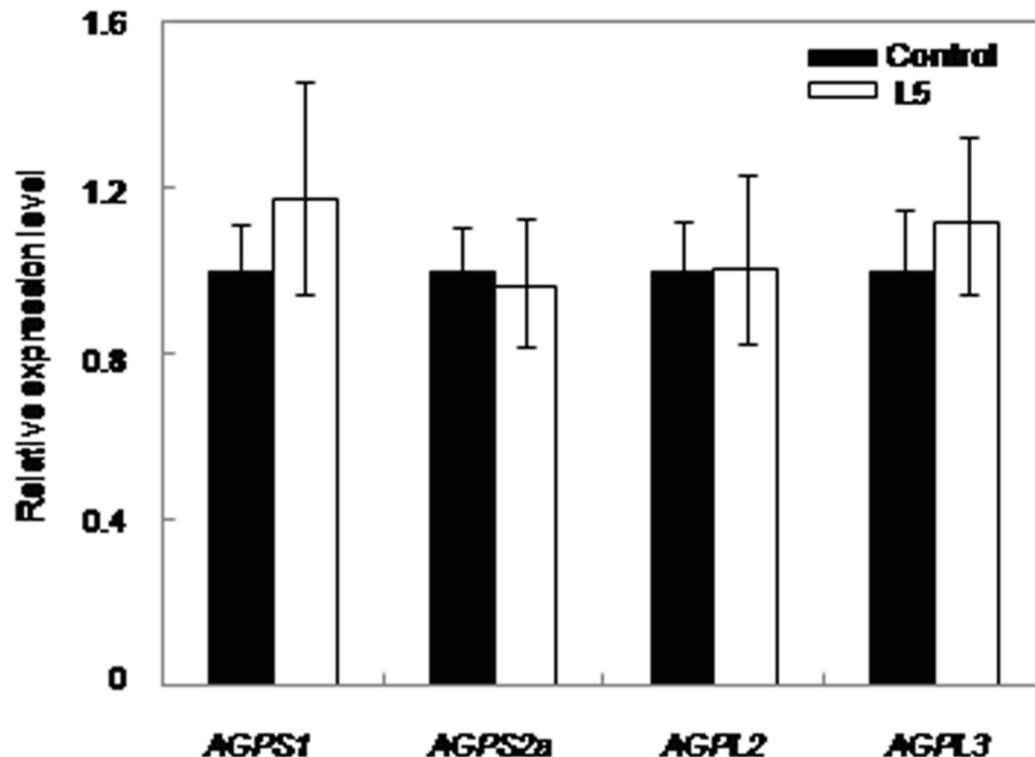**Figure S1**   Accession numbers of genes used in the phylogeny trees.
(TIF)

Figure S2   Expression detections for the selected genes related to the starch synthesis. Quantitative real-time PCR analyses of ADP-glucose pyrophosphorylase genes in young seeds (10 DAP) of control and RNAi transgenic plants. *ACTIN1* was used as an internal control. Error bars indicate the SD (n = 3).
(TIF)

Figure S3   Longitudinal sections of ovule and embryo of control and RNAi transgenic plants. (**A**). Longitudinal sections of mature ovule (OV10) in control plant. (**B–E**). Longitudinal sections of embryo of developing seeds in control plants at 1, 3, 5, and 7 DAP, respectively. (**F**). Longitudinal sections of mature ovule (OV10) in RNAi transgenic plant. Bar = 20 μm.
(TIF)

Figure S4   Transverse sections in the mid-region of control seeds at 0–7 DAP. (**A–C**). Transverse sections of control seeds at stages 0, 1, 2 DAP. (**D–F**). Magnification of OV and NU at stages 0, 1, 2 DAP, respectively. (**G–H**). Transverse sections of control seeds at stages 3, 5, 7 DAP, respectively. (**J–I**). Magnification of OV and NP at stages 3, 5, 7 DAP, respectively. NP, nucellar projection; NU, nucellus; OV, ovular vascular trace. Bar = 50 μm in (A–C), 25 μm in (D–F), 100 μm in (G–H), 200 μm in I, and 50 μm in (J–L), respectively.
(TIF)

Figure S5   The promoter of *OsVPE1* upstream about 3000 bp. A putative CArG-box at position -1827 in the upstream region of *OsVPE1*.
(TIF)

Figure S6   C-terminal sequence alignment of B$_{sister}$ Proteins.
(TIF)

Figure S7   RT-PCR analyses of *OsMADS30* and *OsMADS31*. (**A**). RT-PCR analyses of *OsMADS30* and *OsMADS31* at different development stages. DAP, days after pollination. *ACTIN1* was used as control. (**B**). RT-PCR analyses of *OsMADS30* and *OsMADS31* expression in various floral organs of wild type plants at heading date stage. *APT1* was used as a control.
(TIF)

Table S1   Gene specific primers used in this study.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: XY ZM. Performed the experiments: XY FW LG. Analyzed the data: XY FW. Contributed reagents/materials/analysis tools: XL XD. Wrote the paper: XY FW ZM GT. Gave suggestions for the draft manuscript: KC SS AB.

References

1. Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. Cell 78: 203–209.
2. Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353: 31–37.
3. Cho S, Jang S, Chae S, Chung K, Moon YH, et al. (1999) Analysis of the C-terminal region of *Arabidopsis thaliana APETALA1* as a transcription activation domain. Plant Mol Biol 40: 419–429.
4. Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378: 1079–1118.
5. Fan HY, Hu Y, Tudor M, Ma H (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J 12: 999–1010.
6. Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li MA, et al. (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol Genet Genomics 266: 942–950.
7. Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S, et al. (2009) Class D and B$_{sister}$ MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). Plant Mol Biol 71: 1–14.
8. de Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, et al. (2006) A B$_{sister}$ MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. Plant J 47: 934–946.
9. Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, et al. (2002) The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14: 2463–2479.
10. Prasad K, Zhang X, Tobón E, Ambrose BA (2010) The Arabidopsis B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. Plant J 62: 203–214.
11. Erdmann R, Gramzow L, Melzer R, Theißen G, Becker A (2010) *GORDITA* (*AGL63*) is a young paralog of the *Arabidopsis thaliana* B$_{sister}$ MADS box gene *ABS* (*TT16*) that has undergone neofunctionalization. Plant J 63: 914–924.
12. Kaufmann K, Anfang N, Saedler H, Theissen G (2005) Mutant analysis, protein–protein interactions and subcellular localization of the Arabidopsis B$_{sister}$ (ABS) protein. Mol Genet Genomics 274: 103–118.
13. Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, et al. (2012) The MADS box genes *SEEDSTICK* and *ARABIDOPSIS* B$_{sister}$ play a maternal role in fertilization and seed development. Plant J 70: 409–420.
14. Arora R, Agarwal P, Ray S, Singh A, Singh V, et al. (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8: 242.
15. Lee S, Kim J, Son JS, Nam J, Jeong DH, et al. (2003) Systematic reverse genetic screening of T-DNA tagged genes in rice for functional genomic analyses: MADS-box genes as a test case. Plant Cell Physiol 44: 1403–1411.
16. Yin L, Xue H (2012) The *MADS29* transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. Plant Cell 24: 1049–1065.
17. Hoshikawa K (1993) Anthesis, fertilization and development of caryopsis. Science of the rice plant 1: 339–376.
18. Ohdan T, Francisco PB, Sawada T, Hirose T, Terao T, et al. (2005) Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J Exp Bot 56: 3229–3244.
19. Zhou Z, Wang L, Li J, Song X, Yang C (2009) Study on programmed cell death and dynamic changes of starch accumulation in pericarp cells of *Triticum aestivum* L. Protoplasma 236: 49–58.
20. Dominguez F, Moreno J, Cejudo FJ (2001) The nucellus degenerates by a process of programmed cell death during the early stages of wheat grain development. Planta 213: 352–360.
21. Fukuda H (1997) Programmed cell death during vascular system formation. Cell Death Differ 4: 684–688.
22. Deng M, Bian H, Xie Y, Kim Y, Wang W, et al. (2011) Bcl-2 suppresses hydrogen peroxide-induced programmed cell death via *OsVPE2* and *OsVPE3*, but not via *OsVPE1* and *OsVPE4*, in rice. FEBS J 278: 4797–4810.

23. Radchuk V, Weier D, Radchuk R, Weschke W, Weber H (2010) Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. J Exp Bot 62: 1217–1227.

24. Ikuko HN, Noriyuki H, Satoru N, Miwa K, Mikio N (2005) Vacuolar processing enzyme: an executor of plant cell death. Curr Opin Plant Biol 8: 404–408.

25. Varda SB, Nurit K, Hilal Z (2008) Uncovering the role of VDAC in the regulation of cell life and death. J Bioenerg Biomembr 40: 183–191.

26. Godbole A, Varghese J, Sarin A, Mathew MK (2003) VDAC is a conserved element of death pathways in plant and animal systems. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1642: 87–96.

27. Kim S, Kim S, Wang Y, Yu S, Choi I, et al. (2011) The RNase activity of rice probenazole-induced protein1 (PBZ1) plays a key role in cell death in plants. Mol Cells 31: 25–31.

28. Jwa NS, Kumar Agrawal G, Rakwal R, Park CH, Prasad Agrawal V (2001) Molecular cloning and characterization of a novel jasmonate inducible pathogenesis-related class 10 protein gene, *JIOsPR10*, from rice (*Oryza sativa* L.) seedling leaves. Biochem Bioph Res Co 286: 973–983.

29. Riechmann JL, Wang M, Meyerowitz EM (1996) DNA-binding properties of arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res 24: 3134–3141.

30. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proc R Soc Lond B Biol Sci 268: 2211–2220.

31. Prasad V, Strömberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. Science 310: 1177–1180.

32. Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol 154: 15–28.

33. Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125: 1198–1205.

34. Becker A, Winter KU, Meyer B, Saedler H, Theißen G (2000) MADS-Box gene diversity in seed plants 300 Million Years Ago. Mol Biol Evol 17: 1425–1434.

35. Dreni L, Jacchia S, Fornara F, Fornari M, Ouwerkerk PBF, et al. (2007) The D-lineage MADS-box gene *OsMADS13* controls ovule identity in rice. Plant J 52: 690–699.

36. Fujita M, Horiuchi Y, Ueda Y, Mizuta Y, Kubo T, et al. (2010) Rice expression atlas in reproductive development. Plant Cell Physiol 51: 2060–2081.

37. Krishnan S, Dayanandan P (2003) Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). J Bioscience 28: 455–469.

38. Oparka KJ, Gates P (1984) Sink anatomy in relation to solute movement in rice (*Oryza sativa* L.): A summary of findings. Plant Growth Regul 2: 297–307.

39. Oparka KJ, Gates P (1981) Transport of assimilates in the developing caryopsis of rice (*Oryza sativa* L.). Planta 151: 561–573.

40. Cui R, Han J, Zhao S, Su K, Wu F, et al. (2010) Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). Plant J 61: 767–781.

41. Swofford D (1993) PAUP: A computer program for phylogenetic inference using maximum parsimony. J Gen Physiol 102: 9A.

42. Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res 15: 330–340.

43. Wernersson R, Pedersen AG (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res 31: 3537–3539.

44. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818.

45. Rodríguez F, Oliver JL, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142: 485–501.

46. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

47. Ohmori S, Kimizu M, Sugita M, Miyao A, Hirochika H, et al. (2009) *MOSAIC FLORAL ORGANS1*, an *AGL6*-Like MADS box gene, regulates floral organ identity and meristem fate in rice. Plant Cell 21: 3008–3025.

48. Itoh JI, Nonomura KI, Ikeda K, Yamaki S, Inukai Y, et al. (2005) Rice plant development: from zygote to spikelet. Plant Cell Physiol 46: 23–47.

Sequence data for this article can be found in the GenBank/EMBL/DDBJ/Phytozome databases under the following accession numbers: *OsMADS2* (L37526), *OsMADS4* (L37527), *OsMADS13* (AF151693), *OsMADS16* (AF077760), *OsMADS29* (AK109522), *OsMADS30* (AY174093), *OsMADS31* (AY177698), *APT1* (AK073627), and *ACTIN1* (AK100267) from rice; *AP3* (M86357), *STK* (NM_001203767), *PI* (D30807), *GOA* (AY141243) and *ABS* (AJ318098) from *Arabidopsis thaliana*; *PtMADS38* (POPTR_0012s14770), *PtMADS45* (POPTR_0015s14950) and *PtMADS30* (POPTR_0007s07620) from *Populus trichocarpa*; *DEFH21* (AJ307056) from *Antirrhinum majus*; *FBP24* (AF335242) from *Petunia hybrida*; *LBS* (AI486443, AI899235) from tomato; *ZMM17* (AJ271208), *ZmBS2* (EB160486) and *ZmBS3* (DR811323) from maize; *WBSis* (AM502893, DQ512369), *TaBS3a* (GH730902), and *TaBs3b* (GH731782) from *Triticum aestivum*; *HBS1* (BQ764751), *HvBS2* (AK373226) and *HBS2* (CK122890) from *Hordeum vulgare*; *BdBS1* (BRADI3G05260), *BdBS2* (BRADI1G32210), and *BdBS3* (BRADI5G21700) from *Brachypodium distachyon*; *PvBS1a* (Pavirv00018401m), *PvBS1b* (Pavirv00035216m), *PvBS2* (Pavirv00046240m), *PvBS3a* (Pavirv00008728m) and PvBS3b (Pavirv00020904m) from *Panicum virgatum*; *SiBS1* (Si018183m) and *SiBS2* (Si007187m) from *Setaria italica*; *SbBS1* (XM_002453325), *SbBS2* (XM_002437376) and *SbBS3* (XM_002447045) from *Sorghum bicolor*; *PeBS2* (FP092548) from *Phyllostachys edulis*; *CcBS2* (EB667398) from *Cenchrus ciliaris*; *AaBS* (AY436713) from *Aquilegia alpine*; *DwBS* (AY436726) from *Drimys winteri*; *AeAP3-2* (AF230698) from *Asarum europaeum*; *GGM13* (AJ132219) from *Gnetum gnemon*.

Figure S1. Accession numbers of genes used in the phylogeny trees.

Figure S2. Expression detections for the selected genes related to the starch synthesis. Quantitative real-time PCR analyses of ADP-glucose pyrophosphorylase genes in young seeds (10 DAP) of control and RNAi transgenic plants. ACTIN1 was used as an internal control. Error bars indicate the SD (n=3).

Figure S3. Longitudinal sections of ovule and embryo of control and RNAi transgenic plants. (A). Longitudinal sections of mature ovule (OV10) in control plant. (B–E). Longitudinal sections of embryo of developing seeds in control plants at 1, 3, 5, and 7 DAP, respectively. (F). Longitudinal sections of mature ovule (OV10) in RNAi transgenic plant. Bar=20 µm.

Figure S4. Transverse sections in the mid-region of control seeds at 0–7 DAP. (A–C). Transverse sections of control seeds at stages 0, 1, 2 DAP. (D–F). Magnification of OV and NU at stages 0, 1, 2 DAP, respectively. (G–H). Transverse sections of control seeds at stages 3, 5, 7 DAP, respectively. (J–I). Magnification of OV and NP at stages 3, 5, 7 DAP, respectively. NP, nucellar projection; NU, nucellus; OV, ovular vascular

trace. Bar=50 μm in (A–C), 25 μm in (D–F), 100 μm in (G–H), 200 μm in I, and 50 μm in (J–L), respectively.


```
>OsVPE1_upstream_3000bp
GATTTTCCTCATCAATTTGACTATATCACAAATATACTCCATACATTAATTTTCAAAGCATAGGCATAAAGCTTCAAAAA
GCAAGCCATAGATAAACACTTATAGTGCCCAAAGCACAAATTTATTCCATCACATGTCAAAATAGAGATCATTATATTTT
CTCTAAATAGCAAATATGGGGTAAATAAATGAATAAAGTTCAGAAGTGAAACAACCCTTGCCAAAGCCCGCCAAACAACT
ATGCATTCCAACATTGTGTACATTATTGAGCTCAGAACTAACATATGAATTTCAGTCGATATAACACATGAATCTGTCAT
GTATTGCAAACTGAAATGCTAATTTTGTAACTCCAATCCAATATGAAGAATCGATGTTTGTTGGTGACAAGAGCGAAGTG
GTATTTAGGAGCAAATATAAACAACTTAACCGGGAGATAAGCAACAACCCAAGCGTAGAACCTATCAGGAATATCCTTAA
CCTTTAGCTACATCTACCTTCCAGCACATCTGAAAAAGTTGCTCCCAGGTTGTACACTGGTGGATAGCAGAGGGCTTCGT
GCAAGGGAGAGGTGGAAACAAAATGCAGGAATTGAGAAACACTGTATGAAGGCAGTGACCTGATGTTGGCAAGCCAGCGC
ACTTGATGACTGATAGCATACTTTATGGAGTAAAATGGACAATTTCCTTAAAAAGAAGAACTGATGAAGAACCACTACTT
TAAGTGCATCAGTTGTTAACCAATAGTATTAAATTTAAAAGTAGCAAATTATCATGAGTTGCTCATATAATGCTGAATGT
TACCTGACCAACATCGGCACGGTAGAAGTAGGAAGGGCCTCGCGGAATCAAAGCTGCATCTCCACGGGTTTCTAAGACCCA
ATGCCCAAACTTCTAGTTCTAAGTCACTGTCATCAGTGTATGGGTTGTCTTTTGGAATGGAGTAGTTACCCCACAAGCTG
GTTAGCAATTTCACTCTGTCCTGTAAAGGGAAAAAAGGGAACAAATTATCAAATACAGTACATGCTGATGCTCCATATAT
TGCTAATTGTTTGTCTTATGTGAGAGGAGATTTATCACAAATAAATATGTGAAGTCGTTTGCAAAAAATTCAGTTCGTAT
GTTTGAAGTTCATCATAAACATATTTATCTACCCACCCCATTACCTACTTACACAAATATAAGTAGTGAATCCAGATAAC
TGGGTTGTTGCTATATAGGAAAAAATTGCAGAAGTAGCAGTACCTATCTCATGACGGTGGAAAAAAGACACTCAAATAAA
TTTGATTACCTGTATCACCATCCTGTTTTGCAACTCGTAGCAAAATGGAGTACAGTAATAAAATCCTATTACCAGTTGCA
ATAGTACAATGGAACTTAGCTGATTCCCATAATTTGTAGTATACACATGATGACACAAAGGGATGGATGGGGTCTAGTTG
TACAAGAAGGGCTCTGCAACACTATTTTATCTCCATTACCAAATTCTCTGTAACACTATTCTATGTGCATTACCAAAATC
AGAACTCAAAAAAAGCAGGAATTCATTCAGATCGTAACCTGGTCAAATTCTTGTGAGCAAAAATTTACCTTGGGCACAGA
GGAGATCGGCAAAGGTGGCCACCGGTGACGATGCGGACGGGAGAGATCAACGCTAGCGGAGCAGATCGGTGAGCGACCGT
CGGTGGCGATGCGGACGGGAGAGATCGGCTCGGGCCGGAGGCGGAGATCGACGACGTTGGTGACGCTGGCCCGACGG
GAGAGGGATGGGCGTTGTTGGCGTGGCGGCGTCTCGATGGCGGATGGCGCTGGATAGGCGATGGAGCGATGCGAGAGAG
AGGAGGCGACGTCGCGATGGCTGATGGCGCTGGATAGAGCGATGGCGGCGGCGTGAGGGTGCAGCGTCGGTGGCGTGCGA
GAGAGAGGAGGATGCGTGTGAGGAGCGTGATTGGGGGGATGCGCGTGAGGAAAGGGGAGAGGAGGACGCGCGTGAGGAGC
GCGTGCCGGAGGGGAGGGTCCGCAACGATTTTGAGGCGTTGGATCAGATGATTCTGGACCGTCGGATTTGATGAATGGGGT
TTGTAGGAGTGTAGTTGGTGAGTTATAGGGTAGATGTGTGTCGTGCATACTTCTCTGTGCTCACTACAGTACATATA
GAGAGCCCTTTTACAACACACAAGTTTGATTGAGTCAATGGTTTTATGGGCTCGTGGCCTGTGTGTGGGCCCAATTTCAT
TTTTGGTCCCCTTATCATCCACATCCAGTTGACCAGTCCCTCCTAGCCCATACGGCCTATATTCTGAAGGCCCCAGACCA
AAGTACTGATGGGCCATGTTCCCAACTCTGTAGACTGTAGTCTGGACTTCCACGTCACGTCCACAATTCCCATTTCACAC
ATTTCATTTGGGCCTCTAAAGTAGAAAGGTCTCAACTTTTGACGTTCCACCGGTAAACAGCAGCAGCAGCAGCAAGGCAC
CTACTCCTAATCGTCTCATGTAACACGCGAGTTGATCGTGATCATTCCTCCATCACGAAAATATCATCTCACGCAAAAGCA
CGAAGTCCACCTGCCACGGTTCTGGAGCTCCGCTCCTTTTTCCCAAACACAGCAGAGTTTCAGACTTTTCAGCTTCGGTT
TCCCTTGTTTCTGATCTTTCTTTCTCCTTCATCAAGGCACGGAACAATGCGATCTCCGCGCGGTGTTTGCATTTTGCAA
CCGAAATCTCGCGAAAACCCAAACCCCTCCCGTTCACCGTGCCGGGACTGTACACGTTCTGCGGCCACAAGCCATGGGAT
TTCCAAATCACGAAGTACAATAAATGGAAGAAACCACAAGTTTGTTCCACGCAATGCCCCGCGTGATCCACGCGACCAC
ACTTGCGAGGAGCGCGCGAGCGAAAGCTGCCGCGATGGCGAGACGACGTGCCGCGGGGTCACGTGTCCTCGTGGCACGCTA
ACTCGGGAGTGTACTTAAGGCCGCGTCTCACCAACTACTC
```

Figure S5. The promoter of OsVPE1 upstream about 3000 bp. A putative CArG-box at position -1827 in the upstream region of OsVPE1.

Figure S6. C-terminal sequence alignment of B$_{sister}$ Proteins.

Figure S7. RT-PCR analyses of *OsMADS30* and *OsMADS31*. (A). RT-PCR analyses of *OsMADS30* and *OsMADS31* at different development stages. DAP, days after pollination. *ACTIN1* was used as control. (B). RT-PCR analyses of *OsMADS30* and *OsMADS31* expression in various floral organs of wild type plants at heading date stage. *APT1* was used as a control.

Text S1. Accession numbers

Sequence data for this article can be found in the GenBank/EMBL/DDBJ/Phytozome databases under the following accession numbers: *OsMADS2* (L37526), *OsMADS4* (L37527), *OsMADS13* (AF151693), *OsMADS16* (AF077760), *OsMADS29* (AK109522), *OsMADS30* (AY174093), *OsMADS31* (AY177698), *APT1* (AK073627), and *ACTIN1* (AK100267) from rice; *AP3* (M86357), *STK* (NM_001203767), *PI* (D30807), *GOA* (AY141243) and *ABS* (AJ318098) from *Arabidopsis thaliana*; *PtMADS38* (POPTR_0012s14770), *PtMADS45* (POPTR_0015s14950) and *PtMADS30* (POPTR_0007s07620) from *Populus trichocarpa*; *DEFH21* (AJ307056) from *Antirrhinum majus*; *FBP24* (AF335242) from *Petunia hybrida*; *LBS* (AI486443, AI899235) from tomato; *ZMM17* (AJ271208), *ZmBS2* (EB160486) and *ZmBS3* (DR811323) from maize; *WBSis* (AM502893, DQ512369), *TaBS3a* (GH730902), and *TaBs3b* (GH731782) from *Triticum aestivum*; *HBS1* (BQ764751), *HvBS2* (AK373226) and *HBS2* (CK122890) from *Hordeum vulgare*; *BdBS1* (BRADI3G05260), *BdBS2* (BRADI1G32210), and *BdBS3* (BRADI5G21700) from *Brachypodium distachyon*; *PvBS1a* (Pavirv00018401m), *PvBS1b* (Pavirv00035216m), *PvBS2* (Pavirv00046240m), *PvBS3a* (Pavirv00008728m) and PvBS3b (Pavirv00020904m) from *Panicum virgatum*; *SiBS1* (Si018183m) and *SiBS2* (Si007187m) from *Setaria italica*; *SbBS1* (XM_002453325), *SbBS2* (XM_002437376) and *SbBS3* (XM_002447045) from *Sorghum bicolor*; *PeBS2* (FP092548) from *Phyllostachys edulis*; *CcBS2* (EB667398) from *Cenchrus ciliaris*; *AaBS* (AY436713) from Aquilegia alpine; *DwBS* (AY436726) from Drimys winteri; *AeAP3-2* (AF230698) from *Asarum europaeum*; *GGM13* (AJ132219) from *Gnetum gnemon*.

Table S1. Gene specific primers used in this study.

Gene name	Primer name	Sequences (5' to 3')
RT-PCR analysis		
OsMADS29	OsMADS29-F	CACGA TCAGC AAATA TTTGT GG
	OsMADS29-R	ACGAA GGTTG TCCAG CTGCT
OsMADS30	OsMADS30-F	CAGTG GATGA GCTCA GCCAG
	OsMADS30-R	TCCTA CTGCT TCCAG GAAGT
OsMADS31	OsMADS31-F	GGTGA TGACT GGCT TCACT GAC
	OsMADS31-R	TGGTT GCTCA GTTGC ATCCA GAC
APT1 (Control)	APT1-F	ATTCA TTTTT GGTCC GCCC
	APT1-R	CCCAA ATAAC TCATG TGCCT AC
ACTIN1 (Control)	ACTIN1-F	CCAAT CGTGA GAAGA TGACC CA
	ACTIN1-R	CCATC AGGAA GCTCG TAGCT CT
RNAi vecotor construct		
OsMADS29	OsMADS29RNAi-R	GCA <u>GGATCC</u> <u>ACTAGT</u> GGAAG ACCAG AACAG C
	OsMADS29RNAi-F	GGA <u>CCATGG</u> <u>GTTAAC</u> AACAC AGCAA CCC
Quantitative real time RT-PCR		
OsMADS29	qOsMADS29-F	GATGA CTCGG ATGAG GAACG
	qOsMADS29-R	ACGAA GGTTG TCCAG CTGCT
OsAGPS1	qOsAGPS1-F	GTGCCACTTAAAGGCACCATT
	qOsAGPS1-R	CCCACATTTCAGACACGGTTT
OsAGPS2a	qOsAGPS2a-F	ACTCCAAGAGCTCGCAGACC
	qOsAGPS2a-R	GCCTGTAGTTGGCACCCAGA
OsAGPL2	qOsAGPL2-F	AGTTCGATTCAAGACGGATAGC
	qOsAGPL2-R	CGACTTCCACAGGCAGCTTATT
OsAGPL3	qOsAGPL3-F	AAGCC AGCCA TGACC ATTTG
	qOsAGPL3-R	CACAC GGTAG ATTCA CGAGA CAA
OsVDAC2	qVDAC2-F	TCACT GTTGC TGGCA CGAAG A
	qVDAC2-R	CTGGT AGGGA AAAGG AATGG ATAG
OsVDAC3	qVDAC3-F	GGCTC TACAC CGACA TCGGC AAGA
	qVDAC3-R	GCAGC TGTGA TAGTG ACGCC CTCG
OsVPE1	qVPE1-F	AAGTG GGAGC CGCTG ATTCG
	qVPE1-R	AGAAT CTGGT ACGCA TGGCA C
OsPBZ1	qPBZ1-F	TGTCC TAAAG TCGGA TGTGC T
	qPBZ1-R	TGCCA TAGTA GCCAT CCACG A
ACTIN1 (Control)	qACTIN1-F	TGCTA TGTAC GTCGC CATCC AG
	qACTIN1-R	AATGA GTAAC CACGC TCCGT CA

2.4. Manuscript III

Y. Liu, S. Cui, F. Wu, S. Yan, X. Lin, X. Du, K. Chong, <u>S. Schilling</u>, G. Theißen and Z. Meng (2013): **Functional Conservation of MIKC*-Type MADS Box Genes in *Arabidopsis* and Rice Pollen Maturation.** *The Plant Cell*, **25**, 1288–1303.

Functional Conservation of MIKC*-Type MADS Box Genes in *Arabidopsis* and Rice Pollen Maturation[C][W]

Yuan Liu,[a,b] Shaojie Cui,[a,b] Feng Wu,[a] Shuo Yan,[a,b] Xuelei Lin,[a,b] Xiaoqiu Du,[a] Kang Chong,[a] Susanne Schilling,[c] Günter Theißen,[c] and Zheng Meng[a,1]

[a] Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Department of Genetics, Friedrich Schiller University Jena, D-07743 Jena, Germany

There are two groups of MADS intervening keratin-like and C-terminal (MIKC)-type MADS box genes, MIKC[C] type and MIKC* type. In seed plants, the MIKC[C] type shows considerable diversity, but the MIKC* type has only two subgroups, P- and S-clade, which show conserved expression in the gametophyte. To examine the functional conservation of MIKC*-type genes, we characterized all three rice (*Oryza sativa*) MIKC*-type genes. All three genes are specifically expressed late in pollen development. The single knockdown or knockout lines, respectively, of the S-clade *MADS62* and *MADS63* did not show a mutant phenotype, but lines in which both S-clade genes were affected showed severe defects in pollen maturation and germination, as did knockdown lines of *MADS68*, the only P-clade gene in rice. The rice MIKC*-type proteins form strong heterodimeric complexes solely with partners from the other subclade; these complexes specifically bind to N10-type C-A-rich-G-boxes in vitro and regulate downstream gene expression by binding to N10-type promoter motifs. The rice MIKC* genes have a much lower degree of functional redundancy than the *Arabidopsis thaliana* MIKC* genes. Nevertheless, our data indicate that the function of heterodimeric MIKC*-type protein complexes in pollen development has been conserved since the divergence of monocots and eudicots, roughly 150 million years ago.

INTRODUCTION

Development of the male gametophyte (pollen) is highly conserved among angiosperms and results in one vegetative and two sperm cells that act together in double fertilization of the female gametophyte. Monocots and eudicots, despite their evolutionary differences, express a huge number of homologous transcripts during pollen development (Wei et al., 2010). However, hard evidence that this is due to conserved developmental processes remains scarce. Since MADS box genes play crucial roles in plant development, they are promising candidates for comparative analyses of key regulatory networks.

In plants, MADS box genes, encoding MADS domain transcription factors, have acquired considerable phylogenetic diversity (Becker and Theissen, 2003; Nam et al., 2004). Type II MADS domain proteins in plants all share the same domain structure, comprising a MADS (M), intervening (I), keratin-like (K), and C-terminal (C) domain; therefore, they are also known as MIKC-type MADS domain proteins (Ma et al., 1991; Fan et al., 1997; Riechmann and Meyerowitz, 1997; Alvarez-Buylla et al., 2000). Two different types of MIKC-type genes have been defined: the classic MIKC-type (MIKC[C]-type) and the MIKC*-type genes (Henschel et al., 2002). MIKC*-type genes probably originated more than 450

million years ago in the lineage that led to extant land plants, by duplication of an ancestral MIKC[C]-type gene, followed by an elongation of the I-region or a duplication of parts of the K-box (Henschel et al., 2002; Kwantes et al., 2012). MIKC[C]-type MADS box genes have been shown to regulate diverse aspects in development of the sporophyte of higher plants (Gramzow and Theissen, 2010; Smaczniak et al., 2012), and MIKC*-type genes are crucial for development of the male gametophyte in the model plant *Arabidopsis thaliana* (Verelst et al., 2007a, 2007b; Adamczyk and Fernandez, 2009). The *Arabidopsis* genome encodes six MIKC*-type genes (*AGAMOUS-LIKE30* [*AGL30*], *AGL65*, *AGL66*, *AGL67*, *AGL94*, and *AGL104*), all of which, except *AGL67*, are almost exclusively expressed during pollen development (Kofuji et al., 2003; Honys and Twell, 2004; Verelst et al., 2007a). They activate late pollen developmental genes and repress early genes at late stages of pollen development and thus are crucial for pollen maturation (Verelst et al., 2007b). In *Arabidopsis*, MIKC*-type genes act in a highly redundant manner, with no aberrant phenotype being described for any single loss-of-function mutant (Verelst et al., 2007a; Adamczyk and Fernandez, 2009).

MIKC*-type genes are not only present in higher eudicots but have also been identified in representatives of all major groups of land plants, including bryophytes, lycophytes, ferns, gymnosperms, basal angiosperms, eudicots, and monocots (Henschel et al., 2002; Kofuji et al., 2003; Riese et al., 2005; Rensing et al., 2008; Zobell et al., 2010; Gramzow et al., 2012; Kwantes et al., 2012). In ferns and seed plants, two different monophyletic lineages of MIKC*-type genes could be identified, the P- and S-clades, which are not present in bryophytes or lycophytes (Nam et al., 2004; Gramzow et al., 2012; Kwantes et al., 2012). This high degree of conservation of different subgroups is not observed for MIKC[C]-type genes (Kwantes et al., 2012). Also,

MIKC*-type genes were shown to be expressed predominantly during the gametophytic stage in all land plants that have been studied so far, suggesting an ancient function during the haploid phase of the life cycle which was progressively restricted to the male gametophyte in the lineage leading to higher eudicots (Kwantes et al., 2012). However, when and in which common ancestors that happened is unknown.

Protein–protein and protein–DNA interactions of MIKC*-type proteins have been studied in *Arabidopsis* and in the liverwort (bryophyte) *Marchantia polymorpha*, the lycophytes *Selaginella moellendorffii* and *Selaginella pallescens*, the fern *Ceratopteris richardii*, and the basal eudicot *Eschscholzia californica* (Verelst et al., 2007a; Zobell et al., 2010; Kwantes et al., 2012). Generally, MADS domain proteins bind to DNA sequences called CArG-boxes (for C-A-rich-G). While most MADS domain proteins prefer the so-called serum response element (SRE)-type CArG-box, $5'$-CC(A/T)$_6$GG-$3'$, for binding (Hayes et al., 1988; Riechmann et al., 1996; de Folter and Angenent, 2006), some others, such as the mammalian MYOCYTE ENHANCER FACTOR2A (MEF2A) and the MIKC*-type proteins of all species investigated so far, preferentially bind to the N10-type CArG-box or MEF2 consensus binding site. This motif has the consensus sequence $5'$-CTA(A/T)$_4$TAG-$3'$ but might be more loosely defined as $5'$-C(A/T)$_8$G-$3'$ (Pollock and Treisman, 1991; Shore and Sharrocks, 1995; Verelst et al., 2007a; Zobell et al., 2010; Wu et al., 2011).

Arabidopsis MIKC*-type proteins form obligate heterodimers exclusively with one interaction partner from the P-clade and one from the S-clade (Verelst et al., 2007a). Interclade heterodimerization was also observed for MIKC*-type proteins from more basal species, such as the fern *C. richardii* and the basal eudicot *E. californica*. However, MIKC*-type proteins isolated from these species also showed homodimerization and intraclade heterodimerization (Kwantes et al., 2012).

In contrast with their orthologs from *Arabidopsis*, MIKC*-type genes from monocotyledonous plants have been little investigated so far. Hence, it remains unclear whether the high sequence conservation between monocot and eudicot MIKC*-type genes is also reflected in conserved gene expression and protein interaction patterns and function in pollen maturation.

Here, we provide a comprehensive investigation of MIKC*-type MADS box genes from a monocot species. We characterize all three MIKC*-type genes from rice (*Oryza sativa*): the S-clade genes *MADS62* and *MADS63* and the P-clade gene *MADS68*. Expression of all three MIKC*-type genes was detected in late developmental stages of pollen and proteins of different subclades form heterodimers, features highly similar to the situation in *Arabidopsis*. Furthermore, we report loss-of-function phenotypes of the rice MIKC*-type genes, including the single gene knockdown of a MIKC*-type gene that resulted in a mutant phenotype (i.e., severe defects in pollen maturation and germination). Our results suggest a highly conserved role of MIKC*-type genes in monocots and eudicots.

RESULTS

Phylogenetic Analysis of MIKC*-Type Genes

To gain insight into the evolutionary relationships of MIKC*-type genes within land plants and especially monocots, a comprehensive

search was performed in the National Center for Biotechnology Information, Phytozome, and PlantGDB databases. Homologous sequences were retrieved from a broad variety of land plants, including bryophytes, lycophytes, ferns, gymnosperms, monocots, basal eudicots, and core eudicots (see Supplemental Table 1 online). In particular, the MIKC*-type genes of different Poaceae (grass) species were obtained through BLAST for the available genome sequences, including those of barley (*Hordeum vulgare*), sorghum (*Sorghum bicolor*), maize (*Zea mays*), purple false brome (*Brachypodium distachyon*), switchgrass (*Panicum virgatum*), and foxtail millet (*Setaria italica*). In rice, even though six genes grouped with MIKC*-type genes of *Arabidopsis* in some previous analyses (Arora et al., 2007), only three of them, *MADS62*, *MADS63*, and *MADS68*, were identified as genuine MIKC*-type genes (Gramzow et al., 2012; Kwantes et al., 2012). All the obtained sequences were aligned (see Supplemental Data Set 1 online) and then phylogenetic trees were constructed using maximum likelihood and Bayesian methods (Figure 1). Four highly supported clades can be observed within the superclade of MIKC*-type genes. One contains the MIKC*-type genes from mosses and one consists of sequences from lycophytes, while the other two monophyletic subgroups, the S- and P-clades, include sequences from both ferns and seed plants. In grasses, MIKC*-type genes of the S-clade constitute two subclades, *MADS62*-like and *MADS63*-like genes. These subclades comprise MIKC*-type genes from rice together with putative orthologs from other grass species. Only a single P-clade MIKC*-type gene, *MADS68* and its orthologs, could be found in each of the grass species (Figure 1).

Rice MIKC*-Type Genes Are Specifically Expressed in Pollen

To determine the temporal and spatial expression patterns of rice MIKC*-type genes, diverse tissues at different developmental stages were analyzed by RT-PCR, quantitative RT-PCR (qRT-PCR), promoter-β-glucuronidase (GUS) fusion, and in situ hybridization.

RT-PCR analysis revealed that transcripts of *MADS62*, *MADS63*, and *MADS68* are specifically present in large quantities in the anthers and pollen at stage 13, when pollen achieves maturation, as classified previously (Zhang and Wilson, 2009). No expression was detectable in other tissues, including roots, stems, leaves, a mixture of rice panicles at stages P1 (1 to ~3 cm in length) and P2 (3 to ~6 cm in length), developing seeds, and the other three whorls of mature rice floral organs (lemma/palea, lodicule, and pistil) (Figure 2A).

To further analyze the expression profiles of rice MIKC*-type genes in anthers, qRT-PCR analyses were performed. Results showed that the three genes are only expressed at very low levels or not expressed at all in a mixture of rice anthers of stages 1 to 7 and stages 8 and 9. Expression of the three MIKC*-type genes was first detected in anthers at stage 10, when vacuolated microspores form. Later during pollen development, the expression increased continuously through stage 11, when bicellular pollen forms, and reached the highest levels in anthers at stage 12 (Figure 2B). However, the expression levels of the three genes differ, with *MADS68* showing the highest and *MADS63* the lowest level of expression (Figure 2B).

Additionally, we constructed reporter constructs by fusing the *gusA* (also known as *uidA*) gene encoding GUS to 3-kb
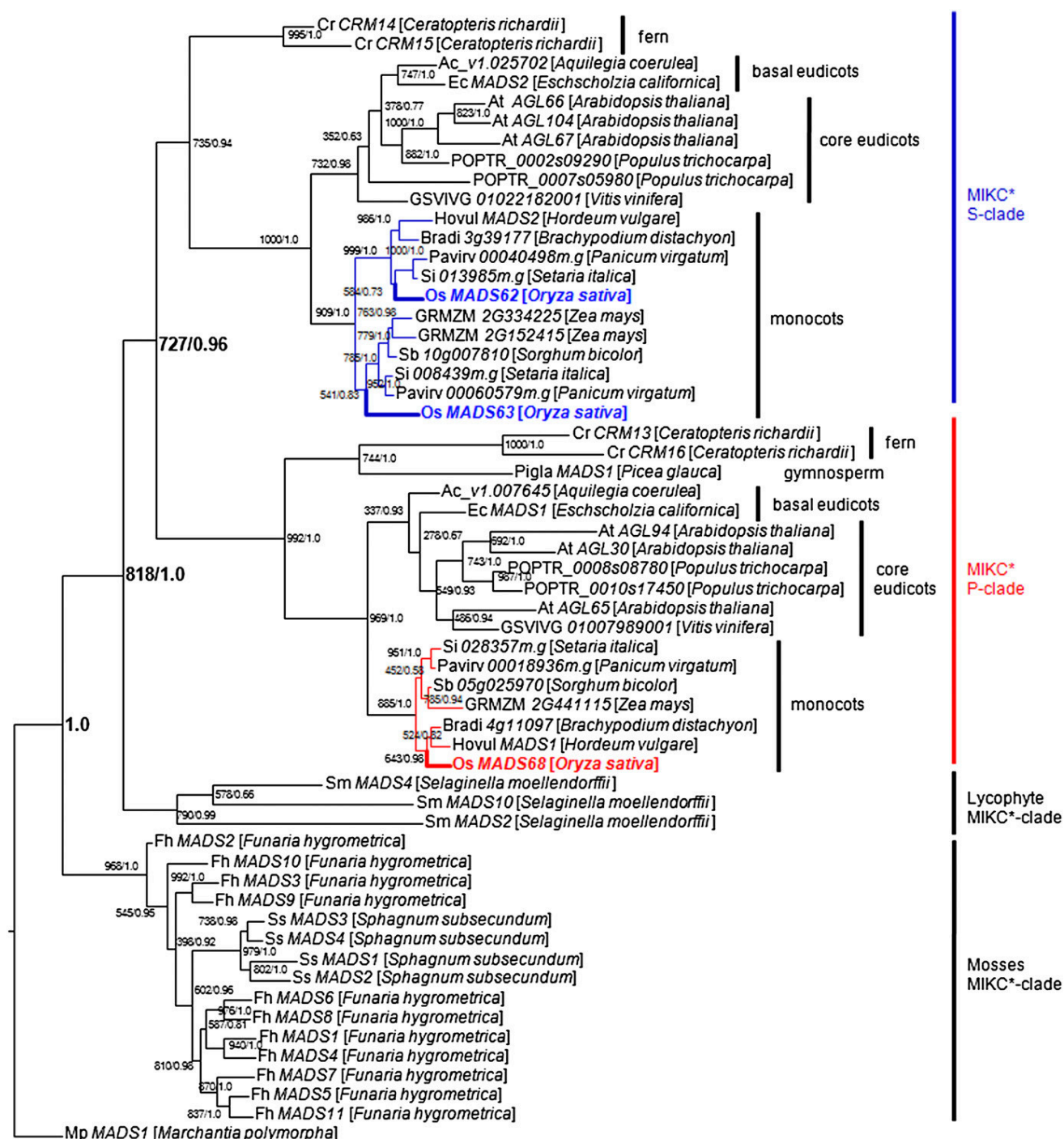
Figure 1. Phylogenetic Tree of MIKC*-Type Genes.

Maximum likelihood and Bayesian phylogenies were conducted from an alignment of 58 cDNAs of MIKC*-type genes from a broad variety of land plants (an overview of all sequences can be found in Supplemental Table 1 online). Mp *MADS1*, the MIKC*-type gene from the liverwort *M. polymorpha*, representing the most early branching bryophyte lineage, was used as the outgroup. Maximum likelihood bootstrap/posterior probability values are shown at the branching points. Terminal nodes are labeled with a gene name followed by the species name. The S-clade genes are highlighted in blue and include the two rice genes *MADS62* and *MADS63*. The P-clade genes are highlighted in red and include the only rice P-clade gene *MADS68*.
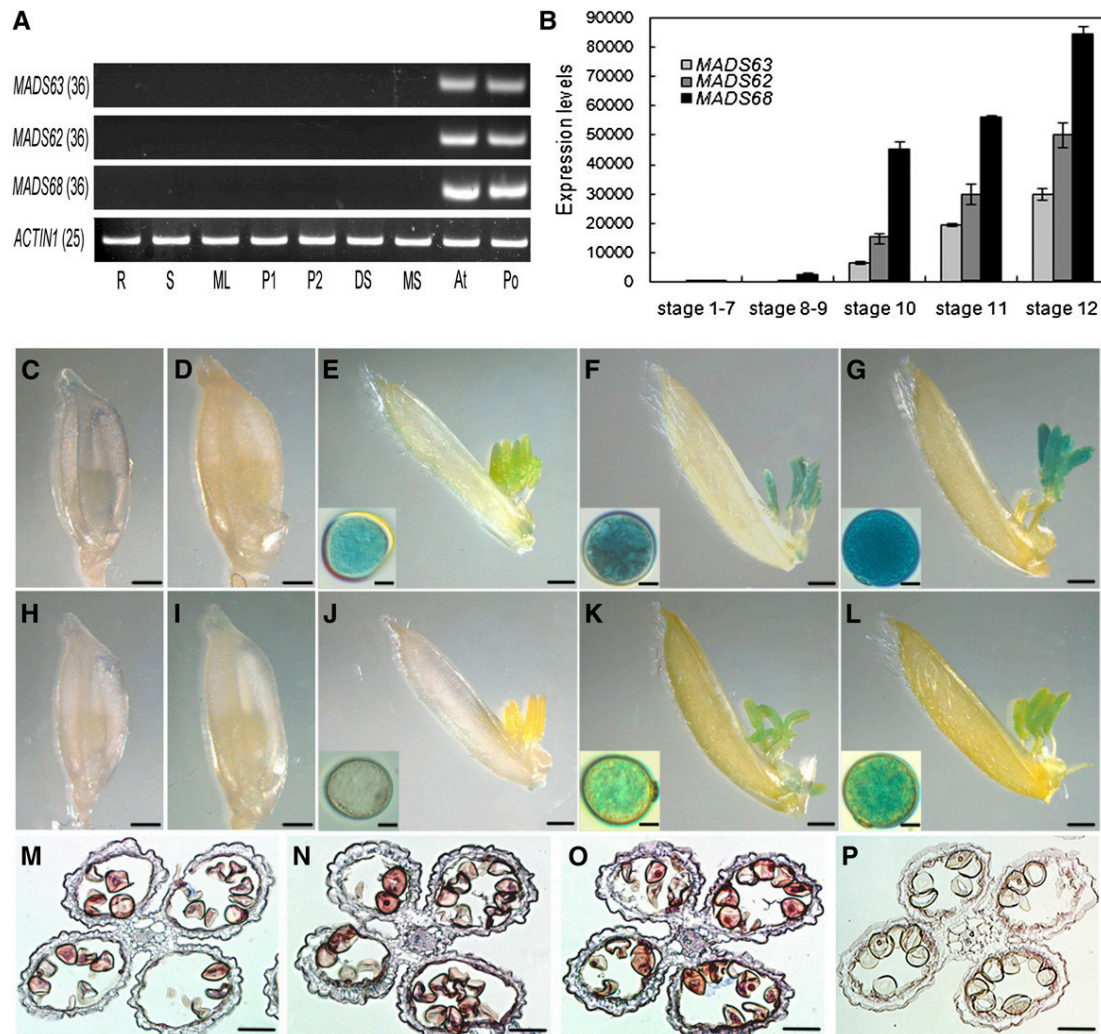
Figure 2. Expression Pattern of Rice MIKC*-Type Genes.

(A) Expression analyses of three rice MIKC*-type genes in various tissues of rice by RT-PCR. Total RNA was isolated from of 7-d-old roots (R), 7-d-old stems (S), mature leaves (ML), 1- to 3-cm panicles (P1), 3- to 6-cm panicles (P2), a mixture of developing seeds (DS), a mixture of lemmas/paleas, lodicules, and pistils of mature spikelets (MS), anthers at stage 13 (An), and mature pollen (Po). *ACTIN1* was used as control. The cycle number of PCR is shown in parentheses for each gene.
(B) Expression levels of three rice MIKC*-type genes in the anther at different developmental stages as revealed by qRT-PCR analysis. Total RNA was pooled from anthers of stage 1 to 7, stage 8 and 9, stage 10, stage 11, and stage 12. Data show the mean $\pm$ sd (n = 3).
(C) to **(G)** GUS staining analyses of florets and pollen grains at different developmental stages in *MADS68$_{pro}$:GUS* transgenic plants. A stained floret with anther before stage 8 **(C)**, at stage 9 **(D)**, at stage 10 after the removal of lemma **(E)**, at stage 11 after the removal of lemma **(F)**, and at stage 12 after the removal of lemma **(G)**. Insets (**[E]** to **[G]**) show a magnified pollen grain at the corresponding stage.
(H) to **(L)** GUS staining analyses of florets and pollen grains at different developmental stages in *MADS63$_{pro}$:GUS* transgenic plants. A stained floret with anther before stage 8 **(H)**, at stage 9 **(I)**, at stage 10 after the removal of lemma **(J)**, at stage 11 after the removal of lemma and insets **(K)**, and at stage 12 after the removal of lemma **(L)**. Insets (**[J]** to **[L]**) show a magnified pollen grain at the corresponding stage.
(M) to **(P)** In situ hybridization of *MADS63* transcripts in anther at stage 11 **(M)**, *MADS62* transcripts in anther at stage 11 **(N)**, *MADS68* transcripts in anther at stage 11 **(O)**, and negative control with sense probes of *MADS68* **(P)**.
Bars = 1 mm in **(C)** to **(L)**, 10 µm in insets, and 50 µm in **(M)** to **(P)**.

sequences upstream of the translation start sites, which we assumed to contain the promoters of the different MIKC*-type genes, and transformed these constructs into rice plants. GUS activity from the *MADS68$_{pro}$:GUS* fusion was first detectable in anthers at stage 10, increasing gradually from stage 11 to 12

(Figures 2C to 2G), consistent with the RT-PCR and the qRT-PCR results. For the *MADS63$_{pro}$:GUS* transgenic lines, GUS expression was hardly observed in anthers before stage 10 (Figures 2H to 2J) but became obvious in anthers at stages 11 and 12 (Figures 2K and 2L). Similar results were observed for

the MADS62~pro~:GUS constructs (see Supplemental Figure 1 online).

To determine whether rice MIKC*-type genes are also expressed in the anther walls, RNA in situ hybridization analyses were performed. The results showed that transcripts of *MADS62, MADS63*, and *MADS68* are specifically localized in pollen but not in anther walls (Figures 2M to 2P). Taken together, these results reveal that rice MIKC*-type genes are exclusively expressed in pollen at late developmental stages, suggesting that they may be involved in pollen maturation.

Dimerization and DNA Binding of Rice MIKC*-Type Proteins

Given that MADS domain transcription factors need to form dimeric complexes to bind DNA (Riechmann et al., 1996), the interaction patterns between the three MIKC*-type proteins in rice were examined by yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) assays. In the Y2H system, strong interactions between MADS62 and MADS68 as well as between MADS63 and MADS68 were observed. Interactions between MADS62 and MADS63 as well as MADS68 homodimerization were of medium strength, while MADS62 and MADS63 each were observed to homodimerize only weakly (Figure 3A; see Supplemental Figure 2 online).

To determine rice MIKC*-type protein interactions in planta, BiFC was performed using the full-length coding sequences. The fluorescence intensity and percentage of BiFC signal-emitting protoplasts, which reflect the strength of interactions, were measured and determined. The strongest interaction could be detected in rice protoplasts transiently transformed with *MADS62/MADS68* and *MADS63/MADS68* combinations (Figures 3B and 3C; see Supplemental Figure 3 online), which is in agreement with the results of the Y2H assays. However, considerably weaker interactions were observed when *MADS62* together with *MADS63* was transformed as well as *MADS62, MADS63*, and *MADS68* alone, thereby reflecting the ability of weak heterodimerization of MADS62/MADS63 and homodimerization of all the rice MIKC*-type proteins in vivo.

To determine their ability to bind DNA, the rice MIKC*-type proteins were synthesized in vitro in a cell-free system and tested in electrophoretic mobility shift assay (EMSA) experiments. In vitro, only MADS63/MADS68 and MADS62/MADS68 heterodimeric complexes were found to bind to DNA probes containing CArG-boxes of the N10 or SRE type (Figure 3D; see Supplemental Figure 4 online). However, MADS62/MADS63, MADS62/MADS62, MADS63/MADS63, and MADS68/MADS68 complexes neither bound to N10-type nor to SRE-type CArG-boxes (Figure 3D; see Supplemental Figure 4 online). Competitive EMSA experiments were further performed to test the binding specificity of these complexes. The binding of MADS62/MADS68 heterodimer to a labeled probe containing the N10-type motif 5'-CTATATATAG-3' was strongly competed by an excess of a probe containing the same motif or the N10-type motif CTATTTTTAG. However, an excess of a competitor probe containing the SRE-type motif CCATATATGG showed very little competition for binding (see Supplemental Figure 5 online). Thus, rice MIKC*-type heterodimers have a strong preference for the N10-type compared with the SRE-type motifs. Our findings suggest that even though rice MIKC*-type proteins are able to form homodimers and

heterodimers among S-clade proteins, these might not function as transcription factors, since these dimers are not able to bind to DNA (Figures 3A to 3D; see Supplemental Figures 2 to 4 online). Thus, the heterodimers formed between members of the P- and S-clades may be the only dimers of rice MIKC*-type proteins that participate in directly controlling target genes and pollen development.

Loss of Function of Single Rice S-clade Genes Does Not Cause a Visible Mutant Phenotype

To explore the functions of the rice S-clade MIKC*-type genes, for *MADS63*, a mutant (PFG_ 2D-10691) that harbors a T-DNA insertion in the seventh intron (see Supplemental Figure 6A online) was obtained from the Postech collection. For *MADS62*, RNA interference (RNAi) transgenic lines were generated. Homozygous *mads63* mutant plants were confirmed by PCR of genomic DNA (see Supplemental Figure 6B online) and used for further analysis. Transcripts of *MADS63* could not be detected by RT-PCR analysis in mutant plants, suggesting that *mads63* is a null mutant (see Supplemental Figure 6C online).

Genetic analyses were performed to determine whether the *mads63* mutant showed gametophytic defects. When heterozygous lines were selfed, plants with wild-type, heterozygous, and homozygous genotypes were produced with a ratio of nearly 1:2:1, thus indicating Mendelian segregation. When heterozygous lines were test-crossed with wild-type pollen, two types of offspring resulted: plants with homozygous wild-type and heterozygous genotypes with a ratio of ~1:1. Also, homozygous mutants pollinated with wild-type pollen all generated heterozygous progeny (see Supplemental Table 2 online). These results demonstrate that the male and female gametophytes, as well as the surrounding sporophytic tissues, are properly functioning. Moreover, I_2-KI staining, Alexander staining, transmission electron microscopy (TEM), and pollen germination assays in vitro all showed that *mads63* pollen is well developed and viable (see Supplemental Figures 6D to 6G online).

Likewise, no aberrant morphologies were found in the *MADS62*_RNAi transgenic lines.

Downregulation of the P-Clade Gene *MADS68* Affects Pollen Maturation and Germination

To elucidate the role of the P-clade gene *MADS68* during rice pollen development, transgenic RNAi knockdown lines of *MADS68* were generated. Forty-nine independent transgenic lines were obtained. No differences in vegetative growth and floral organ development were observed between knockdown plants and control plants transformed with empty vector. However, 26 RNAi lines had obviously reduced pollen fertility compared with control plants. Out of these, we selected five RNAi lines (lines 7, 13, 26, 31, and 41) for further phenotype analysis and measurement of *MADS68* expression levels by qRT-PCR. Viability of pollen, tested with I_2-KI staining and Alexander staining, was lower in *MADS68* RNAi plants (31 to 64%) compared with control plants (97%; Figure 4A), and the frequency of in vitro pollen germination (12 to 30%) was much lower than that of control plants (71%; Figure 4A). Moreover, the reduced expression level of *MADS68* (Figure 4B) was overall consistent with
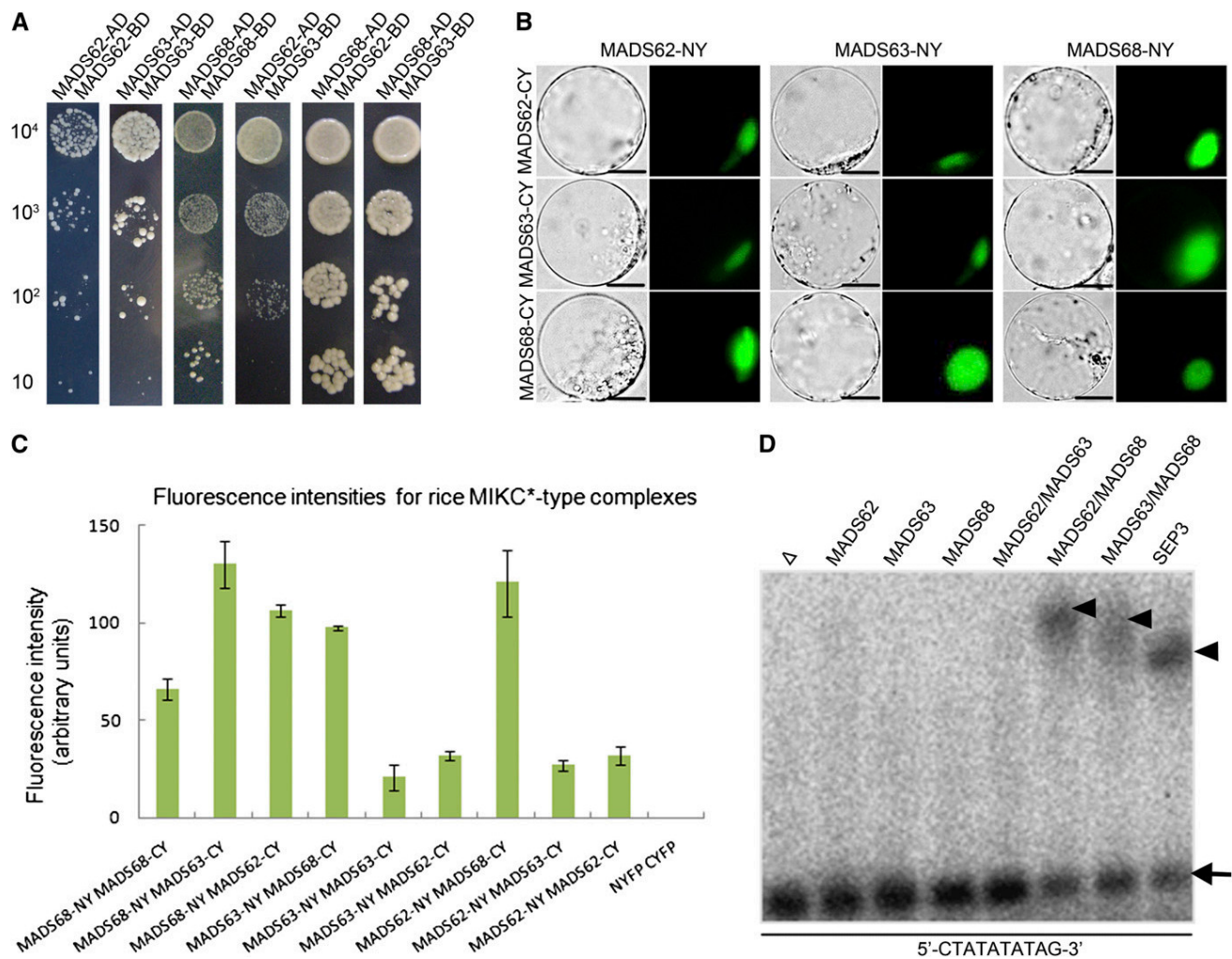
Figure 3. Interaction Patterns and DNA Binding Ability of the Rice MIKC*-Type Proteins.

(A) Rice MIKC*-type protein interactions as revealed by Y2H. Each MIKC*-type protein was fused to the activation domain (AD) as a prey and the DNA binding domain (BD) as a bait. Serial dilutions of 10^4 to 10^1 AH109 cells containing different constructs combinations indicated were grown on the selective medium.

(B) BiFC yellow fluorescent protein fluorescence and bright-field images of rice protoplasts cotransfected with constructs encoding the indicated fusion proteins. Each MIKC*-type protein fused with the N-fragment or C-fragment of yellow fluorescent protein is labeled with its gene name followed by -CY and -NY, respectively. Bars = 10 μm

(C) Quantification of BiFC fluorescence intensities in transiently transfected rice protoplasts. Fluorescence intensity (arbitrary units) for each combination was determined to assess the strength of protein interactions. The mean and SD of six independent measurements are shown. Empty NYFP (N-terminal fragment of yellow fluorescent protein) and CYFP (C-terminal fragment of yellow fluorescent protein) were used as negative control.

(D) EMSA assay for rice MIKC*-type protein complexes binding N10-type CArG-box DNA. A probe containing an N10-type CArG-box (5′-CTATA-TATTAG-3′) was incubated with in vitro–translated MADS62, MADS63, MADS68, and combinations of these proteins. Free DNA is indicated by an arrow and shifted complexes by arrowheads. In vitro translation with SEPALLATA3 (SEP3) and an empty vector (Δ) served as positive and negative control, respectively.

[See online article for color version of this figure.]

the degree of pollen viability and germination rate in the *MADS68* RNAi transgenic plants. No significant decrease in *MADS62* or *MADS63* expression was observed in the *MADS68* RNAi plants (Figure 4B). These results indicate that the observed phenotype of abnormal pollen in the transgenic plants was caused by specific silencing of *MADS68*.

Compared with the pollen grains of control plants, which were deeply stained with I$_2$-KI and enriched with starch (Figures 4C and 4E), we observed only weak staining and defects of starch accumulation in the abnormal pollen grains of RNAi plants (Figures 4D and 4I). To further determine whether mitotic cell division was affected, 4′,6-diamidino-2-phenylindole (DAPI)
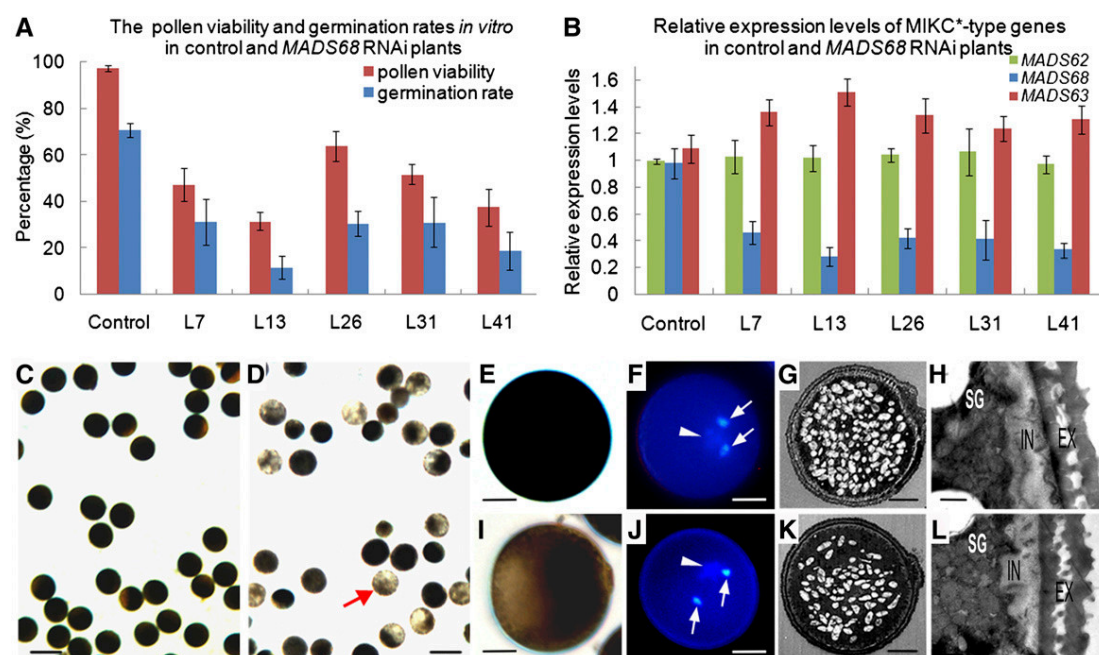
Figure 4. Molecular Detection and Pollen Phenotypes of *MADS68* RNAi Lines.

(A) The percentage of viable pollen grains and the germination rates in vitro of control plants and different *MADS68* RNAi lines. The data show mean ± SD of three independent experiments.

(B) Relative expression levels of *MADS68*, *MADS62*, and *MADS63* in control plants and five selected *MADS68* RNAi lines (L7, L13, L26, L31, and L41). Total RNAs were pooled from anthers at stage 13. The samples were quantified by qRT-PCR using *UBIQUITIN* as a reference gene. The data are presented as mean ± SD ($n = 3$).

(C) to **(L)** Pollen phenotypes at the maturation stage. I_2-KI staining showing pollen viability in control plant **(C)** and *MADS68* RNAi L7 **(D)**. The higher magnification of I_2-KI staining pollen of control plant **(E)** and abnormal pollen of line *MADS68* RNAi L7 **(I)**. DAPI staining showing three nuclei of pollen grain in the control plant **(F)** and line *MADS68* RNAi L7 **(J)**. TEM showing starch accumulation in pollen of control plant **(G)** and abnormal pollen of line *MADS68* RNAi L7 **(K)**; wall structure in pollen of control plant **(H)** and abnormal pollen of line *MADS68* RNAi L7 **(L)**. Arrowheads indicate the vegetative nucleus, and arrows indicate sperm-cell nuclei in pollen of control plant **(F)** and abnormal pollen of line *MADS68* RNAi L7 **(J)**.

EX, extine; IN, intine; SG, starch granules. Bars = 100 µm in **(C)** and **(D)**, 12.5 µm in **(E)**, **(F)**, **(I)**, and **(J)**, 5 µm in **(G)** and **(K)**, and 0.5 µm in **(H)** and **(L)**.

staining and fluorescence microscopy of RNAi pollen grains were performed. Nuclear staining showed two brightly stained small organelles, the sperm nuclei (Figure 4J, arrow), and a larger and diffusely stained structure, the vegetative cell nucleus (Figure 4J, arrowhead) in the abnormal pollen grains of the RNAi plants, similar to those of pollen grains of control plants (Figure 4F). Pollen morphology was investigated in detail by TEM to check for structural mutant phenotypes. A decrease of starch grains in the mutant pollen compared with pollen of the control plants was observed (Figures 4G and 4K). Beyond that, no obvious morphological differences in exine and intine structures of the abnormal pollen were observed in comparison to those of control plants (Figures 4H and 4L). These analyses demonstrate that knockdown of *MADS68* causes defects in pollen maturation and germination but did not affect pollen mitosis or pollen wall development.

Knockdown of *MADS68* and *MADS62* in the *mads63* Background Generates More Severely Affected Pollen

Because the *mads63* mutant and *MADS62* single RNAi plants displayed no obvious morphological alteration, and the three rice MIKC*-type genes exhibited similar expression patterns, the possible functional redundancy among the three genes was investigated.

Given that the *mads63* mutant investigated here was hygromycin resistant, pCam23A-Ubi, a vector conferring G418 resistance, was used for silencing *MADS62* and *MADS68* in the *mads63* mutant background (see Supplemental Figure 7A online). The double-transgenic lines were confirmed by PCR of genomic DNA (see Supplemental Figures 7B and 7C online). RNAi knockdown of either *MADS68* or *MADS62* in the *mads63* background exhibited aberrant pollen phenotypes in 22 of 28 transgenic lines and in nine of 16 transgenic lines, respectively. The mutant phenotypes were similar to those observed in the *MADS68* RNAi lines, but to a more severe extent. Six RNAi lines (lines 7, 23, and 28 for *MADS68i/mads63*; lines 2, 6, and 13 for *MADS62i/mads63*) were selected for more detailed phenotypic analyses and quantitative analysis of gene expression levels (Figure 5). Expression of the genes was diminished as expected from single knockdown and knockout lines (Figure 5A). Using I_2-KI staining and Alexander staining, we observed that a certain proportion of pollen grains (56 to 66% for *MADS68i/mads63*; 35 to 53% for *MADS62i/mads63*) were small

and very weakly stained, indicating little starch accumulation at the mature stage compared with control plants (Figures 5B to 5K). Furthermore, the mutant pollen displayed an in vitro germination rate (14 to 25% for *MADS68i/mads63*; 26 to 39% for *MADS62i/mads63*; Figure 5B) that was much lower than in the case of control plants (71%). In contrast with control plants, only two nuclei, a vegetative-like nucleus and a generative-like nucleus (Figures 5J to 5N), could be identified in the abnormal pollen of *MADS68i/mads63* and *MADS62i/mads63* lines by Alexander and DAPI staining. These closely resembled the vegetative and generative nucleus, respectively, of wild-type binucleate pollen, thus suggesting the abnormal pollen of both *MADS68i/mads63* and *MADS62i/mads63* plants is unable to complete the second mitosis to generate three nuclei.

Arrest in Pollen Development Occurs during the Bicellular Microspore Stage When MIKC*-Type Genes Are Affected

To determine exactly when pollen development is aborted in RNAi plants, pollen collected at different developmental stages was examined by light microscopy.

For single knockdown *MADS68* RNAi lines, no obvious phenotypic alterations were observed in the pollen of RNAi plants until stage 11 (see Supplemental Figure 8D online), compared with control plants (see Supplemental Figure 8A online). At stage 12, bicellular pollen of control plants was enriched with starch granules (see Supplemental Figure 8B online), whereas the development of nearly half of the pollen grains of *MADS68* RNAi plants did not have as many inclusions as pollen of control plants (see Supplemental Figure 8E online). Eventually, a significant number of the abnormal pollen grains failed to reach the mature stage (see Supplemental Figure 8F online).

For *MADS68i/mads63* and *MADS62i/mads63* plants, microscopy showed that microspores of RNAi and control plants before stage 10 were similar in appearance. However, alterations became apparent at late stage 11, when starch accumulation started in bicellular pollen of control plants. Two types of pollen grains could be found in the anthers of *MADS68i/mads63* and *MADS62i/mads63* plants. Whereas some of the grains (34.8% for *MADS68i/mads63*; 52.7% for *MADS62i/mads63*) developed as normally as those in the wild type, the rest displayed apparent abnormalities: smaller pollen and delayed formation of starch granules (see Supplemental Figures 8G and 8J online), compared with control plants (see Supplemental Figure 8A online). Also, the delayed and aberrant pollen was evident at stages 12 and 13, and pollen development eventually arrested or aborted at the bicellular stage (Figure 5; see Supplemental Figures 8H, 8I, 8K, and 8L online).

Late-Stage Pollen Genes with N10-Type CArG-Boxes in Their Putative Regulatory Regions Are Potential Targets of Rice MIKC*-Type Genes

We reanalyzed the rice pollen transcriptome data set from Wei et al. (2010), containing the expression levels of the rice genes in the five stages of pollen development represented on the Affymetrix microarray chip. Excluding transposon and retrotransposon elements, 437 transcripts were identified that were enriched in late stage pollen, including 113 in bicellular pollen, 140 in immature tricellular pollen, 68 in mature pollen grains, and 116 in germinated pollen grains. Subsequently, we screened the 3-kb upstream regions of these genes for the presence of N10-type CArG-boxes using the PLACE tool (Higo et al., 1999). In total, 62 genes possessed N10-type motifs in their putative regulatory regions: 18 bicellular pollen–enriched, 21 tricellular pollen–enriched, four mature pollen grain–enriched, and 19 germinated pollen grain–enriched transcripts (listed in Supplemental Data Set 2 online). These genes are hence potential targets of rice MIKC*-type transcription factor complexes.

RT-PCR and qRT-PCR were further performed in the wild-type plants, *mads63*, *MADS68* single RNAi, *MADS62i/mads63*, and *MADS68i/mads63* plants to test the alteration in expression of a selected nine of these in silico predicted target genes. Results indicate that four genes (Os05g45370, Os02g02460, Os05g11790, and Os06g49860) are substantially downregulated, and two (Os01g47050 and Os10g26470) are upregulated in *MADS68i/mads63* and/or *MADS62i/mads63* plants (see Supplemental Figures 9 and 10 online).

To further test whether these genes are targets of rice MIKC*-type genes, EMSA experiments and protoplast transient transfection assays were performed. The EMSA experiments showed that MADS62/MADS68 and MADS63/MADS68 heterodimeric complexes directly bind to the putative promoter regions of Os10g26470, Os05g11790, Os01g47050, and Os06g49860 and that these interactions are outcompeted by the addition of excess unlabeled fragments (Figures 6A to 6D). Transient transfection assays confirmed that MADS62/MADS68 and MADS63/MADS68 combinations can activate Os05g11790$_{pro}$:LUC and Os06g49860$_{pro}$:LUC expression in vivo. By contrast, MADS62/MADS63 combinations and any of the three proteins alone could not activate Os05g11790$_{pro}$:LUC nor Os06g49860$_{pro}$:LUC expression (Figures 6E to 6G). Therefore, these genes may be primary targets of MADS62/MADS68 and MADS63/MADS68 complexes during pollen development.

DISCUSSION

The Function of MIKC*-Type Genes Is Conserved between *Arabidopsis* and Rice

Phylogenetic studies revealed that there are four well-supported clades within the superclade of MIKC*-type genes, comprising genes of mosses, lycophytes, and the P-clade and S-clade genes of ferns and seed plants, respectively. The latter ones probably originated by a duplication event in the lineage that led to the ferns and seed plants after the lineage that led to mosses and lycophytes had branched off (Figure 1; Gramzow et al., 2012; Kwantes et al., 2012). Within the S- and the P-clades, genes from ferns are basal to angiosperm MIKC*-type genes, which further show a split into a monocot and a eudicot subgroup, thus resembling species phylogeny. In rice and other grass species, two different subclades, *MADS62*-like and *MADS63*-like genes can be found within the S-clade. In combination with the colinearity found between the two rice genes and their corresponding orthologs (Tang et al., 2008a, 2008b), this finding suggests that the duplication event that led to the two subclades had already occurred
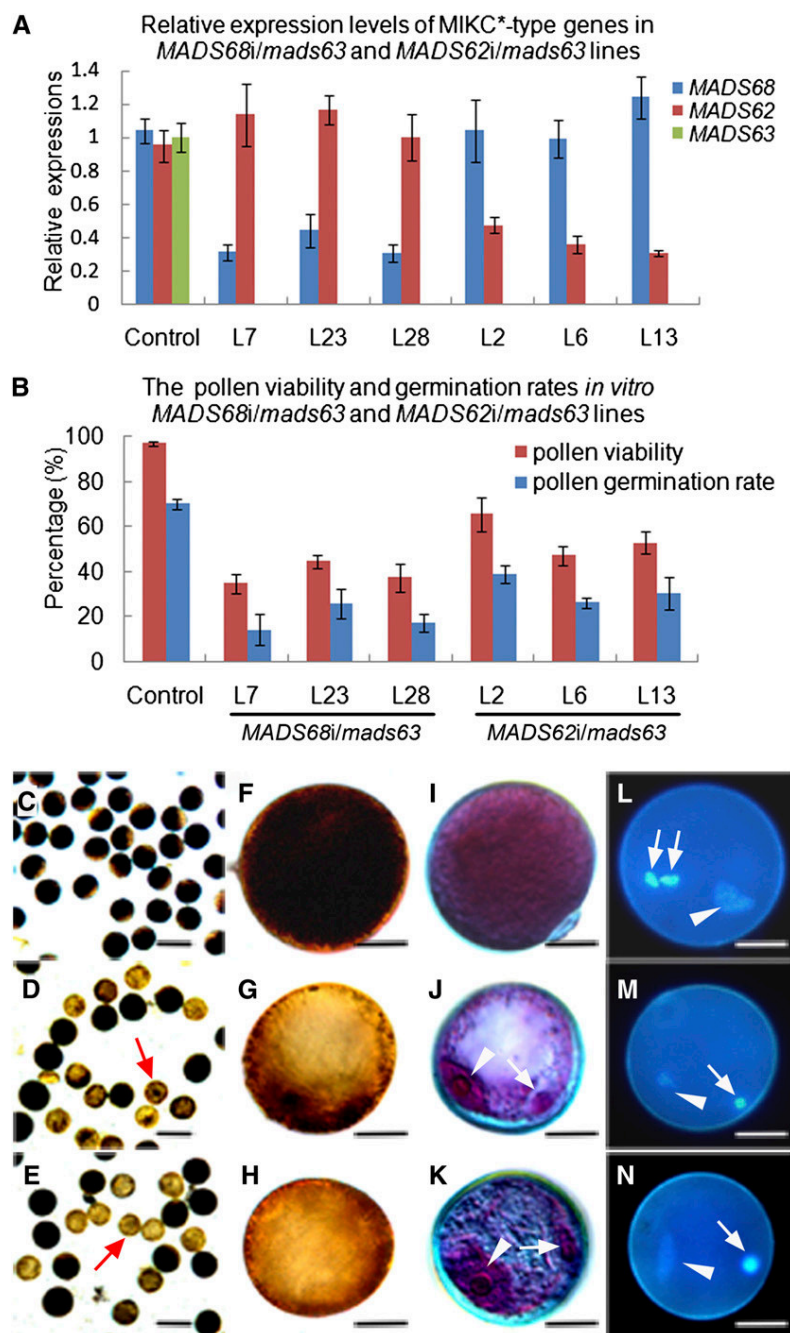
Figure 5. Severe Defects in the Pollen Maturation and Germination for *MADS68i/mads63* Plants and *MADS62i/mads63* Plants.

(A) Relative expression levels of *MADS68*, *MADS62*, and *MADS63* in control plants, three selected *MADS68i/mads63* lines (L7, L23, and L28), and three selected *MADS62i/mads63* lines (L2, L6, and L13). Total RNAs were pooled from anthers at stage 13. The samples were quantified by qRT-PCR using *UBIQUITIN* as a reference gene. Data show mean ± SD ($n = 3$).

(B) The percentage of viable pollen grains and the germination rates in vitro of control plants and different *MADS68i/mads63* lines (L7, L23, and L28) and *MADS62i/mads63* lines (L2, L6, and L13). The data show mean ± SD of three independent experiments.

(C) to **(E)** I_2-KI staining showing starch accumulation of pollen from a control plant **(C)**, line *MADS68i/mads63* L23 **(D)**, and line *MADS62i/mads63* L6 **(E)**.

(F) to **(H)** A higher magnification of I_2-KI stained pollen from a control pollen **(F)** and pollen of line *MADS68i/mads63* L23 **(G)** and line *MADS62i/mads63* L6 **(H)**.

(I) to **(K)** Alexander staining of pollen from the control plant **(I)** and lines *MADS68i/mads63* L23 **(J)** and *MADS62i/mads63* L6 **(K)**.

before the diversification of grass species. The existence of S-clade genes only from either the *MADS62* or the *MADS63* subclade in barley, sorghum, maize, and *B. distachyon* is probably due to gene loss after duplication or lack of comprehensive genome data. Only a single P-clade MIKC*-type gene could be found in each of the grass species, which is in contrast with eudicots, where several recent duplications could be observed within the P- and S-clades (Figure 1).

Expression studies in several representatives of the major land plant lineages revealed that most MIKC*-type genes are highly expressed in the male gametophyte (Verelst et al., 2007a; Zobell et al., 2010; Kwantes et al., 2012). Five of the six *Arabidopsis* MIKC*-type genes (*AGL30*, *AGL65*, *AGL94*, *AGL66*, and *AGL104*) are predominantly expressed in pollen, while the other one (*AGL67*) is expressed in embryonic tissue (Kofuji et al., 2003; Honys and Twell, 2004; Lehti-Shiu et al., 2005; Verelst et al., 2007a). Within the P-clade genes, different temporal expression profiles were observed between *AGL30* and *AGL65*, which represent a pair of paralogs that originated from a recent gene duplication event (Veron et al., 2007). Like the two S-clade genes *AGL66* and *AGL104*, *AGL30* is expressed already at the unicellular microspore stage, whereas *AGL65* is only activated during the tricellular stage (Honys and Twell, 2004; Verelst et al., 2007a, 2007b). In this study, the three rice MIKC*-type genes *MADS62*, *MADS63*, and *MADS68* were observed to be expressed exclusively during pollen development, from the unicellular microspore stage to pollen maturation (Figure 2), which resembles the expression of *AGL66*, *AGL104*, and *AGL30* in *Arabidopsis*. This suggests that a pollen-specific expression pattern was already established in the most recent common ancestor of monocots and eudicots.

It has been proposed that MIKC*-type genes may provide conserved and essential functions to the male gametophytes of all vascular land plants, including lycophytes, ferns, and seed plants (Kwantes et al., 2012). Mutant analyses also confirmed that in the eudicot *Arabidopsis*, the MIKC*-type genes are required for pollen maturity and tube growth (Verelst et al., 2007a, 2007b; Adamczyk and Fernandez, 2009). The *mads63* mutant and the *MADS62* RNAi transgenic plants display no aberrant pollen morphologies, but in the *MADS62i/mads63* double mutants, pollen maturation and germination is impaired (Figure 5; see Supplemental Figures 6 and 8 online). This indicates that the two S-clade genes of *MADS62* and *MADS63* act largely redundantly in pollen development, as their homologs *AGL66* and *AGL104* in *Arabidopsis* do (Verelst et al., 2007a; Adamczyk and Fernandez, 2009). Previously, neither in *Arabidopsis* nor in any other species was a complete knockout or a knockdown of P-clade genes achieved. However, the downregulation of the sole rice P-clade gene *MADS68* in RNAi transgenic lines resulted in defects in pollen maturation and germination

(Figure 4; see Supplemental Figure 8 online), similar to that of *MADS62i/mads63* double mutants. These data illustrate that P- and S-clade genes are not functionally redundant and that both clades are indispensable for pollen development.

However, in the double loss-of-function line *MADS62i/mads63*, the extent of pollen abnormality is more severe than in the *MADS68* single RNAi lines. This could possibly be due to incomplete downregulation of *MADS68* via RNAi (Figure 4), meaning that a certain degree of *MADS68* function may have remained intact even in the knockdown lines. Furthermore, the *MADS68i/mads63* mutant lines show a more severe phenotype than the single knockdown *MADS68* transgenic lines do. This could hint at an incomplete redundancy of *MADS62* and *MADS63*, since a complete redundancy of both genes should enable the intact *MADS62* to replace *MADS63* function completely in this double mutant, even though neither the *mads63* mutant nor the *MADS62* RNAi transgenic plants showed a mutant phenotype. However, the more severe mutant phenotype of *MADS68i/mads63* lines compared with *MADS68* RNAi lines may also suggest that there is some redundancy between the S- and P-clade genes that escaped previous analyses.

In *MADS68* RNAi, *MADS68i/mads63*, and *MADS62i/mads63* plants, starch accumulation during pollen development is affected (Figures 4 and 5; see Supplemental Figure 8 online). Starch filling is a prominent metabolic process during pollen maturation. Starch can serve as an energy resource for subsequent germination, and starch levels hence can be considered as a checkpoint for pollen maturity (Datta et al., 2002). The degree of starch accumulation below a certain threshold can result in premature termination of pollen development and failure of pollen to complete second mitosis (Wen and Chase, 1999; Zhang et al., 2001; Datta et al., 2002; Han et al., 2006). Accordingly, the fact that pollen development of *MADS68i/mads63* and *MADS62i/mads63* plants is delayed and eventually arrested at the bicellular pollen stage while *MADS68* single RNAi can finish the second mitosis to produce three nuclei is assumed to be related to the apparently different degrees of starch deficiency, which are much more severe in the double loss-of-function plants than in *MADS68* single RNAi lines (Figures 4 and 5; see Supplemental Figure 8 online). By contrast, many of *agl66 agl104-2* double mutant pollen grains showed signs of structural damage and ~40% of the double mutant pollen grains had at least one visible nucleus compared with 70% of wild-type pollen in *Arabidopsis* (Adamczyk and Fernandez, 2009). However, it is unknown as to precisely when the structural abnormalities begin to appear and whether pollen second mitosis occurs in *agl66 agl104-2* double mutant pollen. Therefore, our findings provide clues at the cellular level for causes of pollen defects in loss-of-function mutants of MIKC*-type genes, at least in rice.

Figure 5. (continued).

(L) to **(N)** DAPI staining showing three nuclei in pollen of control plants **(L)**, but only two nuclei in abnormal pollen of lines *MADS68i/mads63* L23 **(M)** and *MADS62i/mads63* L6 **(N)**.

Arrowheads indicate the vegetative nucleus in pollen of control plants **(L)** and in abnormal pollen of *MADS68i/mads63* L23 (**[J]** and **[M]**) and *MADS62i/ mads63* L6 (**[K]** and **[N]**); arrows indicate sperm-cell nuclei in pollen of a control plant **(L)** and the single generative-like cell nucleus in an abnormal pollen of *MADS68i/mads63* L23 (**[J]** and **[M]**) and *MADS62i/mads63* L6 (**[K]** and **[N]**). Bars = 100 μm in **(C)** to **(E)** and 12.5 μm in **(F)** to **(N)**.
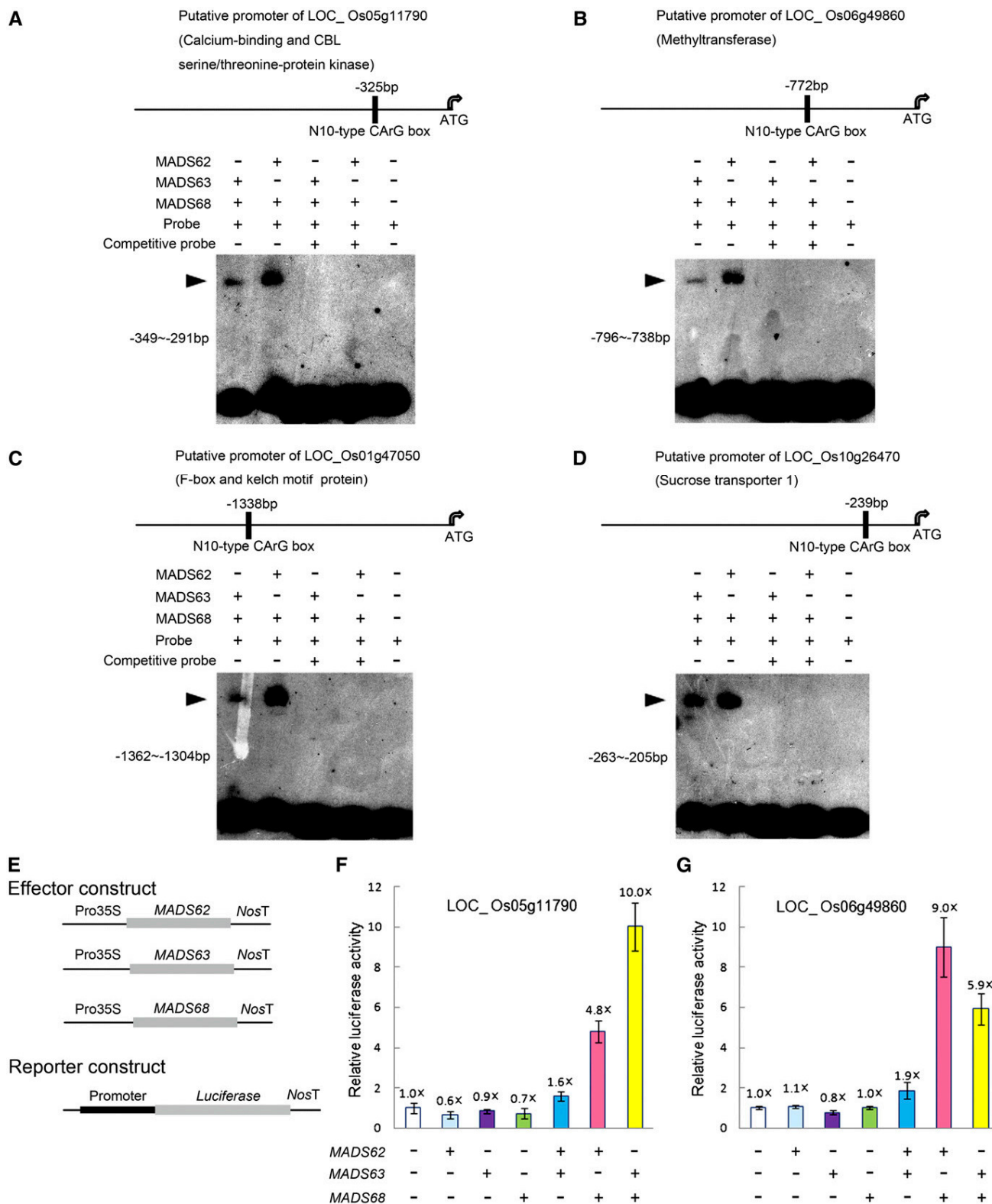
Figure 6. EMSA and Transient Transfection Analyses for Heterodimeric Complexes of Rice MIKC*-Type Proteins.

Taken together, our data underline the fact that despite independent duplications in both P- and S-clade genes, and sequence divergence after the separation of monocots and dicots more than 150 million years ago (Wikström et al., 2001), MIKC*-type genes confer an indispensable and highly conserved function in pollen maturation and germination in the monocot rice as well as in the higher eudicot *Arabidopsis*. This strongly suggests that P- and S-clade genes already had such a role in pollen development in the most recent common ancestor of monocots and eudicots. Further investigation of MIKC*-type genes in gymnosperms and basal angiosperms may reveal whether functional conservation of MIKC*-type genes for male gametophytes could be extended to all angiosperms or seed plants.

The Formation of Heterodimeric Complexes between S- and P-Clade Proteins Is Crucial for the Function of MIKC*-Type Proteins

It was hypothesized that homodimerization might be an ancestral state of MIKC*-type complexes, and obligatory heterodimerization probably evolved from homodimerization after gene duplication and subsequent sequence divergence (Zobell et al., 2010). Shortly after the divergence of the S- and P-clade genes, MIKC*-type homodimers and heterodimers between members of the same clade might still have been functional, as one still finds in ferns. Based on expression and interaction data, it is suggested that the MIKC*-type protein complexes might function not only in male and female gametophytes but also in sporophytic tissues (Kwantes et al., 2012). However, in basal eudicots, the functional MIKC*-type protein complexes in the form of S- and P-clade heterodimers might only exist in male gametophytes (Kwantes et al., 2012). In the higher eudicot *Arabidopsis*, the MIKC*-type proteins form obligate heterodimeric complexes between S- and P-clade members and function exclusively in pollen (Verelst et al., 2007a, 2007b; Adamczyk and Fernandez, 2009). Our protein–protein interaction analyses revealed that the MIKC*-type proteins in rice are able to form both homodimeric and heterodimeric complexes, but our EMSA analyses demonstrated that only heterodimeric complexes between S- and P-clade proteins are able to bind to DNA (Figure 3). Therefore, our study suggests that only heterodimeric complexes of S- and P-clade members are of functional relevance, even though a regulatory role of protein dimerization that does not lead to DNA binding complexes cannot be completely ruled out at the moment.

Obligate heterodimerization of proteins from the S- and P-clades is in line with the severe phenotypes that were observed when all genes from one phylogenetic clade were knocked down, as in *MADS62i/mads63* and *MADS68i* lines. In these cases, putatively heterodimeric complexes cannot form, thus leading to aberrant pollen development. By contrast, the normal phenotype of the single mutants of S-clade genes *MADS62* and *MADS63* might be due to redundancy of the MADS62/MADS68 and the MADS63/MADS68 complexes.

Furthermore, we showed that rice MIKC*-type protein heterodimers are very strict with respect to their binding motif, N10-type CArG-boxes, which is very similar to the behavior of MIKC*-type proteins heterodimers of *Arabidopsis* (Verelst et al., 2007a). In contrast with SEP3, which is a MIKC^C-type protein, MIKC*-type heterodimers did not bind to SRE-type CArG-boxes in competition assays. This suggests a very high target specificity for MIKC*-type genes, not only in rice but very likely also in *Arabidopsis*, which might be essential for their exclusive and crucial role in pollen development.

Taken together, our data suggest that MIKC*-type protein complexes have confined their function to pollen development in the most recent common ancestor of monocots and eudicots.

The Heterodimeric MIKC*-Type Protein Complexes Are Involved in Pollen Maturation and Germination via Binding N10-Type CArG-Box Motifs of Target Genes

A considerable number of genes have been demonstrated to be associated with the process of pollen maturation and subsequent pollen germination (Honys and Twell, 2003, 2004). Gene expression analyses employing microarrays revealed that in *Arabidopsis* heterodimeric complexes of MIKC*-type proteins regulate pollen maturation and germination by activating or repressing many pollen-specific genes (Verelst et al., 2007b; Adamczyk and Fernandez, 2009). However, no direct biochemical or molecular genetic experiments were performed to verify that these genes are direct targets of heterodimeric complexes of MIKC*-type proteins. In rice, binding sites typical for MIKC*-type protein complexes are enriched in putative promoters of genes expressed during late pollen development (see Supplemental Data Set 2 online). These genes may well be involved in pollen maturation and germination and are potential targets involved in the pollen regulatory network of MIKC*-type protein complexes. Several of these genes have been demonstrated to be substantially up- or downregulated in the double loss-of-function lines (see Supplemental Figures 9 and 10 online), and EMSA assays confirmed that MADS62/MADS68 and

---

Figure 6. (continued).

(A) to **(D)** EMSA assays show that MADS62/MADS68 and MADS63/MADS68 dimers directly bind to fragments of putative promoter regions of Os05g11790 **(A)**, Os06g49860 **(B)**, Os01g47050 **(C)**, and Os10g26470 **(D)**. The positions of the N10-type CArG-boxes site in the promoter regions are shown (top panels). DNA fragments (each shown in left of bottom panel) were labeled as probes and unlabeled DNA fragments in 20-fold excess were added as competition probes. The shifted bands are indicated with arrowheads.

(E) to **(G)** Transient transfection assays indicate that heterodimers of MIKC*-type proteins activate LUC reporter gene expression under the control of putative promoters of Os05g11790 **(F)** and Os06g49860 **(G)** in *Arabidopsis* protoplasts. The luciferase activity of protoplasts transfected with empty effector and reporter vectors **(E)** was set to 1×, and other values for the different constructs combinations were normalized relative to this value. Data show mean ± SD (n = 3).

[See online article for color version of this figure.]

MADS63/MADS68 complexes can directly bind to N10-type CArG-boxes that are present in the putative promoter regions of these genes (Figures 6A to 6D). Furthermore, two of these genes, Os05g11790 and Os06g49860, can be activated by MADS62/MADS68 and MADS63/MADS68 combinations in vivo in transient transfection assays (Figures 6E to 6G). Os05g11790 (CIPK20) is annotated as encoding a calcium binding and CBL-interacting protein kinase (CIPK), which has not yet been functionally characterized. However, RNAi suppression of CIPK23, another member of the rice CIPK family, results in partial male sterility due to a proportion of pollen without starch accumulation (Yang et al., 2008), which is similar to the pollen phenotype of rice plants with MIKC*-type gene loss of function. Intriguingly, seven members of the Arabidopsis CIPK family are significantly downregulated in agl66 agl104-2 double, agl65 agl66 agl104-1 triple, and agl65 agl66 agl94 agl104-1 quadruple mutants (Adamczyk and Fernandez, 2009). For AT45820 (CIPK20), which is the putative ortholog of rice CIPK20, a N10-type CArG-box has been found within its 1-kb putative promoter region (see Supplemental Data File S1 in Adamczyk and Fernandez, 2009). Given that MIKC*-type heterodimeric complexes preferentially bind N10-type motifs both in rice and Arabidopsis (Verelst et al., 2007a; this study), it is conceivable that they regulate orthologous targets (for example, rice and Arabidopsis CIPK20). Hence, the gene regulatory network involving MIKC*-type protein complexes is likely conserved between rice and Arabidopsis.

Additionally, Os06g49860 is a member of a methyltransferase family. These enzymes play crucial roles in many development processes, and some of them have been reported to be associated with pollen development (Mou et al., 2002; Grini et al., 2009). These data thus provide some clues concerning the molecular mechanisms of pollen development, in which a large number of genes are regulated by complexes of MIKC*-type proteins.

Taken together, complexes of MIKC*-type proteins are essential components of the gene regulatory network for pollen maturation and germination in rice, in which they activate or repress the expression of many downstream target genes via binding to N10-type CArG-boxes in regulatory regions of their DNA. In combination with previous studies in Arabidopsis, we conclude that the function and underlying regulatory network of MIKC*-type protein complexes in pollen developmental processes was already established in the most recent common ancestor of monocots and eudicots and hence has been strongly conserved for at least ~150 million years.

METHODS

Plant Materials and Growth Conditions

The rice cultivars 'Zhonghua 11' and 'Hwayoung' (Oryza sativa ssp japonica) were grown under natural conditions in an experimental field of the Institute of Botany, Chinese Academy of Sciences, Beijing, China.

Mutant Identification and Cultivation

The T-DNA mutant line PFG_2D-10691 for MADS63, derived from cultivar Hwayoung, was obtained from the Postech collection (Jeong et al., 2002; An et al., 2003). Homozygous lines were identified by PCR screening with a pair of MADS63 specific primers and a T-DNA border primer (see Supplemental Table 3 online). Then, the homozygous lines were backcrossed to wild-type plants to generate F1 seeds to make the genetic background uniform. F1 plants were cultivated for two generations, and homozygous descendants (F3 generation) were isolated as described above for further analysis. Total RNAs from the anthers of F3 homozygous mutants were extracted and RT-PCR was performed to analyze the expression of MADS63.

Generation of Knockdown Plants in Wild-Type and mads63 Backgrounds

To construct the MADS62 and MADS68 knockdown vectors, gene-specific DNA fragments of 307 bp (MADS62; nucleotides 745 to 1051 counted from ATG) or 402 bp (MADS68; nucleotides 179 to 581 from ATG) were amplified as RNAi fragments. The intron and cauliflower mosaic virus 35S terminator cassette of pJawohl3-RNAi (GenBank accession number AF404854) was cloned into the HindIII-NotI site of the pBluescript SK vector (Stratagene), thus yielding pBIntron as an intermediate vector. The RNAi fragments of MADS62 and MADS68 were subcloned into the pBIntron vector in both sense (EcoRI-SmaI) and antisense (HindIII-XhoI) orientations. Finally, the RNAi cassettes were mobilized with restriction enzymes KpnI and SacI and introduced into the vector pCAMBIA1301-Ubi (Cui et al., 2010), thus yielding the final binary vectors. These constructs or an empty vector were introduced into cultivar Zhonghua 11 by Agrobacterium tumefaciens–mediated transformation, as described previously (Hiei et al., 1994).

To generate the MADS68 and MADS62 RNAi binary vectors for transformation of the mads63 mutant, the RNAi cassettes described above were digested with restriction enzymes KpnI and SacI and finally ligated into the vector pCam23A-Ubi (a derivative of pCAMBIA 2300 carrying the maize [Zea mays] UBIQUITIN promoter). A schematic diagram of the resulting RNAi constructs can be found in Supplemental Figure 7 online. Then, F3 homozygous mads63 mutants were used for subsequent transformation with MADS68 and MADS62 RNAi constructs under selection with 120 mg/mL G418, as described previously (Huang et al., 2001). The empty vector pCam23A-Ubi was used in parallel transformations as control for vector-dependent effects.

Promoter-GUS Fusion Studies and Histochemical Analysis of GUS Activity

The sequences ~3.0 kb upstream of the translation start sites of MADS62, MADS63, and MADS68 were amplified with specific pairs of primers (see Supplemental Table 3 online) and recombined into the vector pCAMBIA1301 upstream of the GUS (uidA) gene. The resultant constructs were transformed into rice by the method described above. GUS staining was performed as previously described (Jefferson et al., 1987).

Phylogenetic Analyses

All sequences for phylogenetic analyses were retrieved from the databases of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), Phytozome v7.0 (http://www.phytozome.net/), and PlantGDB (http://www.plantgdb.org/) using the BLAST program (Altschul et al., 1990). Fifty-eight full-length amino acid sequences of MIKC*-type proteins were aligned with the MUSCLE program (http://www.ebi.ac.uk/Tools/msa/muscle/) and manually optimized using Genedoc software (Nicholas and Nicholas, 1997). Then, the alignment was reverse translated using aa2dna (http://www.personal.psu.edu/nxm2/software.htm) into the corresponding nucleotide alignment, which is provided in Supplemental Data Set 1 online. Phylogenetic analyses were performed based on the alignment of DNA sequences that encode MIK regions using maximum likelihood and Bayesian methods. Maximum likelihood analysis was performed using PHYML version 3.0 with GTR + I + Γ model and 1000

bootstrap replicates (Guindon and Gascuel, 2003). Bayesian phylogeny was determined using MrBayes version 3.2.1 (Ronquist et al., 2012), sampling every 500 generations for 1,000,000 generations and discarding 500 of the sampled trees; the posterior probability was used to estimate nodal robustness.

In Situ Hybridization Analyses

MADS62-, *MADS63-*, and *MADS68-*specific regions were amplified using corresponding primer pairs (see Supplemental Table 3 online) and then transcribed in vitro as probes using the Digoxigenin RNA labeling kit (Roche). Fresh wild-type flowers from different developmental stages were fixed immediately, embedded in paraffin (Sigma-Aldrich), and sectioned at 8-μm thickness. Hybridization and immunological detection were performed according to the previous description (Cui et al., 2010).

RT-PCR and qRT-PCR Analyses

Total RNAs were isolated from various wild-type rice tissues and anthers at different developmental stages of wild-type, mutant, and transgenic lines using TRIzol reagent (Invitrogen) digested with DNase I and reverse transcribed using Superscript-III reverse transcriptase (Invitrogen) into cDNA samples as templates for subsequent RT-PCR and qRT-PCR. For RT-PCR, the internal standard *ACTIN1* was used. qRT-PCR analyses were performed using SYBR Premix Ex Taq Mix (Takara) on a Rotor-Gene 3000 (Corbett Research) according to the manufacturer's instructions. For analysis of the expression of *MADS62*, *MADS63*, and *MADS68* in wild-type anthers at different stages, a dilution series and standard curve were created to determine the initial quantity of each gene and then their expressions levels were normalized to the geometric average the expression of internal genes *UBIQUITIN* and *ACTIN1* (Vandesompele et al., 2002). For other qRT-PCR analyses, relative expression levels were calculated using *UBIQUITIN* as reference gene. The primer sequences are listed in Supplemental Table 3 online.

Y2H and BiFC assays

For Y2H assays, the full-length coding sequence of each rice MIKC*-type gene was amplified and fused into the GAL4 activation domain vector pGADT7 or the GAL4 binding domain vector pGBKT7. Yeast cells containing different construct combinations were grown on the selective medium SD-LTHA + 12.5 mM 3-amino-1,2,4-triazole (L, Leu; T, Trp; H, His; A, adenine). Yeast transformations and interaction analyses were conducted as previously described (Cui et al., 2010). For BiFC assays, full-length coding sequences of three rice MIKC*-type genes were amplified and fused into the pUC-SPYNE and the pUC-SPYCE vectors (Walter et al., 2004). Rice protoplast isolation and transfection was performed essentially as described (Chen et al., 2006). The experiments were repeated six times. Fluorescence intensity (arbitrary units) was determined using Image J software. More than 450 rice protoplasts for each complex were counted to determine the percentage of BiFC fluorescence signal-emitting protoplasts after transient transfection.

EMSA Analyses

For cell-free expression, the coding sequences of *MADS62*, *MADS63*, and *MADS68* were subcloned into the pTNT vector via digestion and ligation. In vitro translation was done with a SP6 TNT QuickCoupled Transcription/Translation mix (Promega) according to the manufacturer's instructions. DNA fragments were labeled using polynucleotide kinase (Fermentas) and then used as probes (see Supplemental Table 4 online); unlabeled DNA fragments were used as competition probes. The protein-DNA binding and competition reactions, electrophoretic separation, and

visualization of bands were conducted as described (Verelst et al., 2007a). The experiments were performed three times independently.

I$_2$-KI, Alexander, and DAPI Staining and Pollen Germination in Vitro

To assess pollen viability, mature anthers were incubated with 1% I$_2$-KI staining and Alexander staining solution (Alexander, 1969). For pollen germination assays, mature pollen grains were gently applied to plates in a medium consisting of 20% Suc, 40 mg/L boric acid, 3 mM/L Ca(NO$_3$)$_2$, 5 mg vitamin B$_1$, and 10% polyethylene glycol 4000 and cultured at room temperature for 10 min before observation. The percentage of pollen viability and pollen germination rate for each line was calculated using three independent biological samples, and at least 500 pollen grains for each sample were scored. For DAPI staining, pollen grains were fixed in an ethanol:acetic acid (3:1) solution, rehydrated, and then incubated in 1 μg/mL DAPI solution for 1 h at room temperature. Photography was performed using a Leica microscope (DM4500B) under UV light.

TEM for Pollen Ultrastructural Analyses

Anthers, sampled from wild-type, *mads63* mutant, and *MADS68* RNAi transgenic plants at developmental stage 13, were fixed with 3% glutaraldehyde and 2% osmic acid, dehydrated, and embedded in Spurr's resin (Sigma-Aldrich). Then, ultrathin sections (50-nm thick) were mounted on a grid, stained with 4% uranylacetate, and observed under a transmission electron microscope (JEM-1230; JEOL).

Protoplast Transient Transfection Assays

To generate effector constructs, full-length sequences of *MADS62*, *MADS63*, and *MADS68* were amplified and inserted into the p2GW7,0 vector (Karimi et al., 2002). About 2-kb (upstream of ATG) putative promoter regions of Os05g11790 and Os06g49860 were amplified and inserted into pUC-35sLUC (Tang et al., 2012) to produce the corresponding promoter-LUC reporter vectors. The primers used are listed in Supplemental Table 3 online.

Arabidopsis thaliana protoplasts were isolated and transfected as described (Yoo et al., 2007). The reporter and effector constructs as well as 35S:GUS vector (internal control) were cotransformed into protoplasts. After cotransformation, samples were incubated for 12 h in darkness and then the luciferase and GUS enzymatic assays were performed as previously described (Tang et al., 2012). The reporter gene expression levels were determined as the relative ratio of luciferase to GUS activity.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL databases under the following accession numbers: *MADS62* (Os08g38590), *MADS63* (Os06g11970), *MADS68* (Os11g43740), *UBIQUITIN* (NM_001056014, Os03g13170), and *ACTIN1* (AK100267, Os03g50890); tested predicted target genes (Os05g45370, Os02g02460, Os05g11790, Os06g49860, Os01g47050, Os10g26470, Os03g25460, Os03g27900, and Os12g37690) from rice; and *SEP3* (NM_102272, AT1G24260) from *Arabidopsis*.

Supplemental Data

The following materials are available in the online version of this article.

   Supplemental Figure 1. GUS Staining Analyses of Florets and Pollen Grains in *MADS62$_{pro}$:GUS* Transgenic Plants.

   Supplemental Figure 2. Schematic Depiction of All Rice MIKC*-Type Protein Interactions as Revealed by Y2H Analyses.

   Supplemental Figure 3. The Percentage of BiFC Fluorescence Signal-Emitting Protoplasts in Rice Protoplasts after Transient Transfection.

Supplemental Figure 4. EMSA Assay for Rice MIKC*-Type Heterodimeric Complexes Binding SRE-Type CArG-Box DNA.

Supplemental Figure 5. Competitive EMSA for MIKC*-Type Heterodimers and SEP3.

Supplemental Figure 6. The Identification and Microscopy Analyses of Pollen of *MADS63* T-DNA Insertion Mutants.

Supplemental Figure 7. Schematic Diagram of RNAi Constructs and Identification of *MADS68i/mads63* and *MADS68i/mads63* Transgenic Plants.

Supplemental Figure 8. Delay and Arrest of Pollen Development Occurs from the Bicellular Stage.

Supplemental Figure 9. RT-PCR Analyses for a Selection of Nine in Silico Predicted Target Genes of the Rice MIKC*-Type Protein Complexes.

Supplemental Figure 10. qRT-PCR Analyses of the Expression of Putative Target Genes of the Rice MIKC*-Type Protein Complexes.

Supplemental Table 1. Detailed Description of MIKC*-Type Genes Used for Alignment and Phylogenetic Analyses.

Supplemental Table 2. Genetic Analyses of the *mads63* Mutant.

Supplemental Table 3. Gene-Specific Primers Used in This Study.

Supplemental Table 4. EMSA Probes Used in This Study.

Supplemental Data Set 1. Text File of Alignment of Nucleotide Sequences of cDNAs of 58 MIKC*-Type Genes Used for the Phylogenetic Analyses Shown in Figure 1.

Supplemental Data Set 2. List of Rice Late Pollen-Enriched Genes with Their Expression Levels and Spatial Distribution of N10-Type CArG-Box Motifs in Their Putative Promoters.

AUTHOR CONTRIBUTIONS

Y.L. and Z.M. designed the research. Y.L., S.C., S.S., and S.Y. performed the research. Y.L. and S.S. analyzed the data. X.L., F.W., and X.D. contributed reagents/materials/analysis tools. Y.L., Z.M., S.S., and G.T. wrote the article. F.W. and K.C. gave suggestions for the draft article.

REFERENCES

Adamczyk, B.J., and Fernandez, D.E. (2009). MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. Plant Physiol. **149:** 1713–1723.

Alexander, M.P. (1969). Differential staining of aborted and nonaborted pollen. Stain Technol. **44:** 117–122.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martínez-Castilla, L., and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97:** 5328–5333.

An, S., et al. (2003). Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol. **133:** 2040–2047.

Arora, R., Agarwal, P., Ray, S., Singh, A.K., Singh, V.P., Tyagi, A.K., and Kapoor, S. (2007). MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics **8:** 242.

Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. **29:** 464–489.

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M.E., Umemura, K., and Wang, G.L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. Mol. Plant Pathol. **7:** 417–427.

Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., Xu, Q., Chong, K., Theissen, G., and Meng, Z. (2010). Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). Plant J. **61:** 767–781.

Datta, R., Chamusco, K.C., and Chourey, P.S. (2002). Starch biosynthesis during pollen maturation is associated with altered patterns of gene expression in maize. Plant Physiol. **130:** 1645–1656.

de Folter, S., and Angenent, G.C. (2006). *trans* meets *cis* in MADS science. Trends Plant Sci. **11:** 224–231.

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

Gramzow, L., Barker, E., Schulz, C., Ambrose, B., Ashton, N., Theissen, G., and Litt, A. (2012). *Selaginella* genome analysis - Entering the "homoplasy heaven" of the MADS world. Front. Plant Sci. **3:** 214.

Gramzow, L., and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. Genome Biol. **11:** 214.

Grini, P.E., Thorstensen, T., Alm, V., Vizcay-Barrena, G., Windju, S.S., Jørstad, T.S., Wilson, Z.A., and Aalen, R.B. (2009). The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis*. PLoS ONE **4:** e7817.

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

Han, M.J., Jung, K.H., Yi, G., Lee, D.Y., and An, G. (2006). *Rice Immature Pollen 1* (*RIP1*) is a regulator of late pollen development. Plant Cell Physiol. **47:** 1457–1472.

Hayes, T.E., Sengupta, P., and Cochran, B.H. (1988). The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. Genes Dev. **2** (12B)**:** 1713–1722.

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

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. **6:** 271–282.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. **27:** 297–300.

Honys, D., and Twell, D. (2003). Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol. **132:** 640–652.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol. **5:** R85.

Huang, J.Q., Wei, Z.M., An, H.L., and Zhu, Y.X. (2001). *Agrobacterium tumefaciens*-mediated transformation of rice with the spider insecticidal gene conferring resistance to leaffolder and striped stem borer. Cell Res. **11:** 149–155.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K., and An, G. (2002). T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. **130:** 1636–1644.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci. **7:** 193–195.

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

Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol. Biol. Evol. **29:** 293–302.

Lehti-Shiu, M.D., Adamczyk, B.J., and Fernandez, D.E. (2005). Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. Plant Mol. Biol. **58:** 89–107.

Ma, H., Yanofsky, M.F., and Meyerowitz, E.M. (1991). *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. **5:** 484–495.

Mou, Z., Wang, X., Fu, Z., Dai, Y., Han, C., Ouyang, J., Bao, F., Hu, Y., and Li, J. (2002). Silencing of phosphoethanolamine N-methyltransferase results in temperature-sensitive male sterility and salt hypersensitivity in *Arabidopsis*. Plant Cell **14:** 2031–2043.

Nam, J., Kim, J., Lee, S., An, G., Ma, H., and Nei, M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc. Natl. Acad. Sci. USA **101:** 1910–1915.

Nicholas, K.B., and Nicholas, H.B.J. (1997). GeneDoc: A tool for editing and annotating multiple sequence alignments. EMBNEW. NEWS **4:** 14.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5** (12A): 2327–2341.

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

Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. Biol. Chem. **378:** 1079–1101.

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

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

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61:** 539–542.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur. J. Biochem. **229:** 1–13.

Smaczniak, C., Immink, R.G., Angenent, G.C., and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. Development **139:** 3081–3098.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008a). Synteny and collinearity in plant genomes. Science **320:** 486–488.

Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M., and Paterson, A.H. (2008b). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res. **18:** 1944–1954.

Tang, W., Wang, W., Chen, D., Ji, Q., Jing, Y., Wang, H., and Lin, R. (2012). Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in *Arabidopsis*. Plant Cell **24:** 1984–2000.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. **3:** RESEARCH0034.

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

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H., and Münster, T. (2007b). MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol. **8:** R249.

Veron, A.S., Kaufmann, K., and Bornberg-Bauer, E. (2007). Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. Mol. Biol. Evol. **24:** 670–678.

Walter, M., Chaban, C., Schütze, K., Batistic, O., Weckermann, K., Näke, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K., and Kudla, J. (2004). Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. **40:** 428–438.

Wei, L.Q., Xu, W.Y., Deng, Z.Y., Su, Z., Xue, Y., and Wang, T. (2010). Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. BMC Genomics **11:** 338.

Wen, L.Y., and Chase, C.D. (1999). Mitochondrial gene expression in developing male gametophytes of male-fertile and S male-sterile maize. Sex. Plant Reprod. **11:** 323–330.

Wikström, N., Savolainen, V., and Chase, M.W. (2001). Evolution of the angiosperms: Calibrating the family tree. Proc. Biol. Sci. **268:** 2211–2220.

Wu, W., Huang, X., Cheng, J., Li, Z., de Folter, S., Huang, Z., Jiang, X., Pang, H., and Tao, S. (2011). Conservation and evolution in and among SRF- and MEF2-type MADS domains and their binding sites. Mol. Biol. Evol. **28:** 501–511.

Yang, W., Kong, Z., Omo-Ikerodah, E., Xu, W., Li, Q., and Xue, Y. (2008). Calcineurin B-like interacting protein kinase OsCIPK23 functions in pollination and drought stress responses in rice (*Oryza sativa* L.). J. Genet. Genomics **35:** 531–543, S1–S2.

Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: A versatile cell system for transient gene expression analysis. Nat. Protoc. **2:** 1565–1572.

Zhang, D., and Wilson, Z. (2009). Stamen specification and anther development in rice. Chin. Sci. Bull. **54:** 2342–2353.

Zhang, Y., Shewry, P.R., Jones, H., Barcelo, P., Lazzeri, P.A., and Halford, N.G. (2001). Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. Plant J. **28:** 431–441.

Zobell, O., Faigl, W., Saedler, H., and Münster, T. (2010). MIKC* MADS-box proteins: Conserved regulators of the gametophytic generation of land plants. Mol. Biol. Evol. **27:** 1201–1211.

**Supplemental Figure 1. GUS staining analyses of florets and pollen grains in *MADS62$_{pro}$*:*GUS* transgenic plants.**
A stained floret with anther before stage 8 **(A)**, at stage 8 **(B)**, at stage 9 **(C)**, at stage 10 **(D)**, at stage 11 **(E)** and a magnified pollen grain **(G)**, at stage 12 **(F)** and a magnified pollen grain **(H)**.
Bars = 1 mm in **(A)** to **(F)**, and 10 μm in **(G)** and **(H)**.

prey \ bait	S-clade		P-clade	
	MADS62	MADS63	MADS68	pGBKT7
MADS62	░░ (light)	▒▒ (medium)	██ (black)	
MADS63	▒▒ (medium)	░░ (light)	██ (black)	
MADS68	██ (black)	██ (black)	▒▒ (gray)	
pGADT7				

Supplemental Figure 2. Schematic depiction of rice MIKC*-type fusion protein interactions as revealed by Y2H analyses.

Each MIKC*-type protein was fused to the activation domain (AD) as a prey and the DNA-binding domain (BD) as a bait. Color intensity corresponds to interaction strength and a darker color of the boxes represents stronger interactions. Empty vectors pGADT7 and pGBKT7 were used as a negative control.
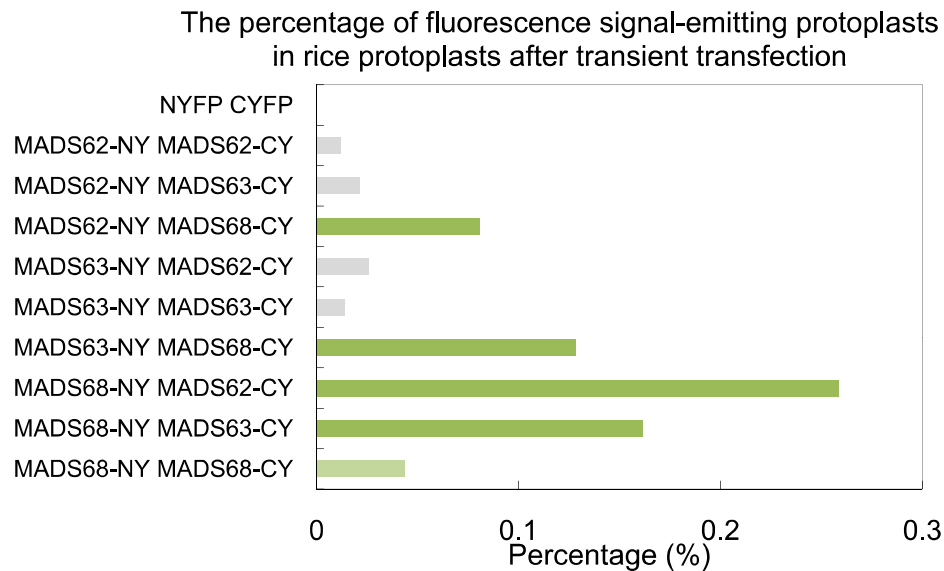
The percentage of fluorescence signal-emitting protoplasts in rice protoplasts after transient transfection

Supplemental Figure 3. The percentage of BiFC fluorescence signal-emitting protoplasts in rice protoplasts after transient transfection.
The mean of six independent experiments for each combination are shown. Empty NYFP and CYFP were used as negative control.
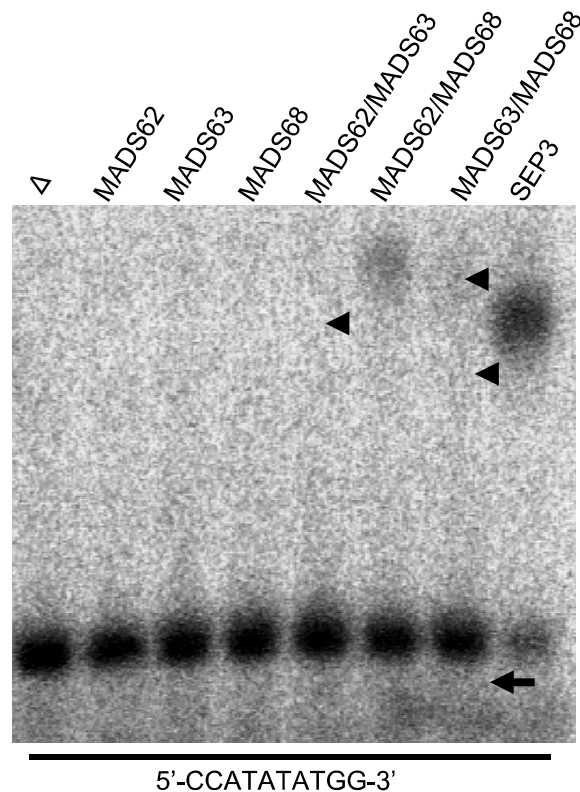
5'-CCATATATGG-3'

Supplemental Figure 4. EMSA assay for rice MIKC*-type heterodimeric complexes binding SRE-type CArG-box DNA.
A probe containing a SRE-type CArG-box (5'-CCATATATTGG-3') was incubated with *in vitro* translated MADS62, MADS63, MADS68 and combinations of these proteins. Free DNA is indicated by an arrow, shifted complexes by arrowheads. *In vitro* translation with SEP3 and an empty vector (Δ) served as positive and negative control, respectively.
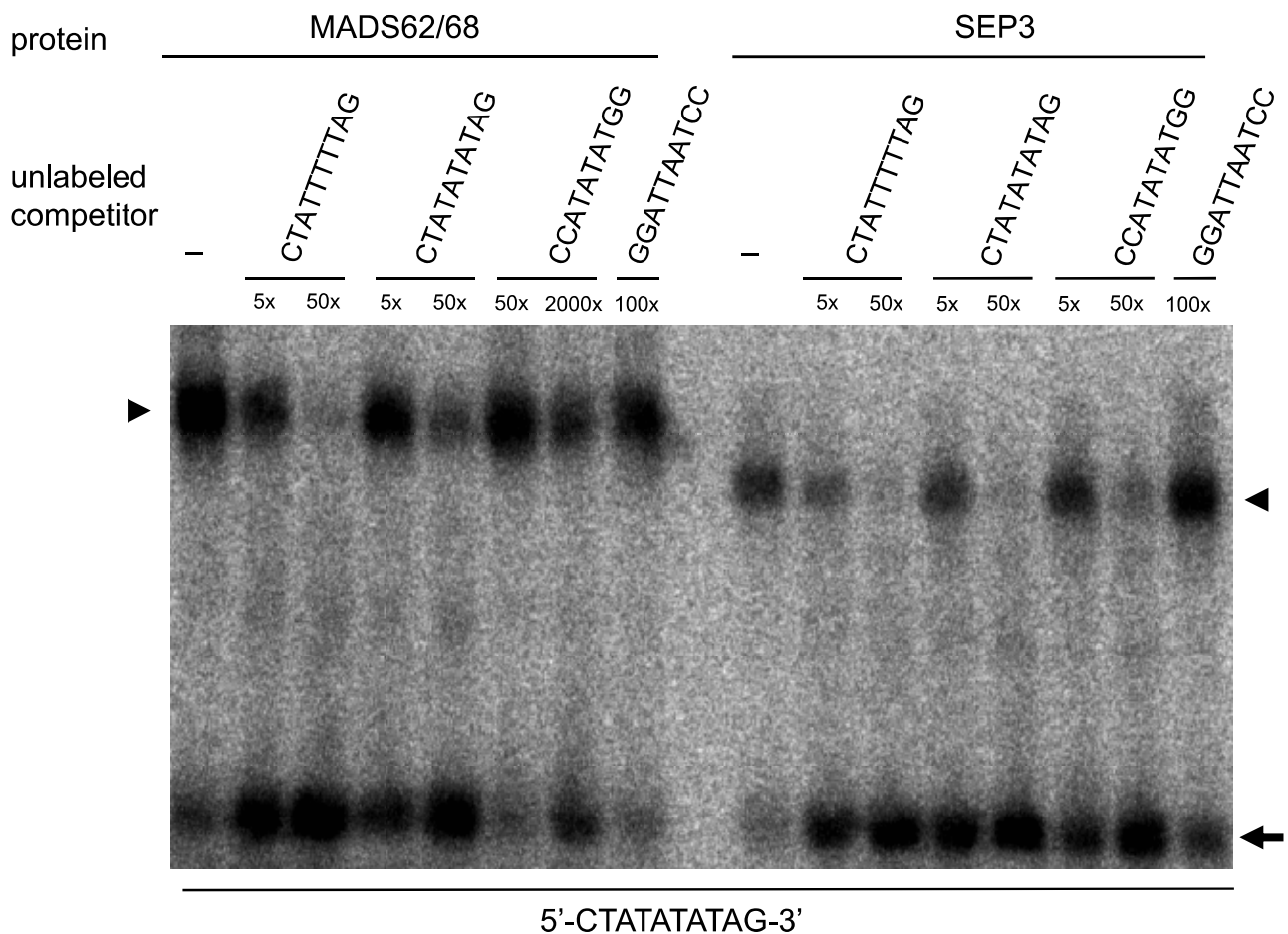
Supplemental Figure 5. Competetive EMSA for MIKC*-type heterodimers and SEP3.

In vitro co-translated MADS62/MADS68 or SEP3 were incubated with a labeled containing a N10-type CTATATATAG probe (-), and the different competitor probes with the excess at which they were used in each lane. Free DNA is indicated by an arrow, shifted complexes by arrowheads.

Left panel: An unlabeled N10-type CTATTTTTAG probe successfully competes with a labeled N10-type CTATATATAG probe for binding to the MADS62/MADS68 protein complex (complete competition at 50-fold excess) and another unlabeled N10-type CTATATATAG probe compete to a lesser extent. In contrast, almost no competition was observed with an unlabeled probe containing the SRE-type motif CCATATATGG, even when a 2000-fold excess was supplied. Similarly, a 100-fold excess of an unlabeled probe containing the derived motif GGATTAATCC was not able to compete for binding.

Right panel: For N10-type competitor probes, SEP3 showed competition behavior similar to the MADS62/MADS68 heterodimer. But the binding of SEP3 to N10-type probe could be almost completely outcompeted with only 50-fold excess of the SRE-type competitor. An unlabeled probe containing the derived motif GGATTAATCC was used for a negative control.
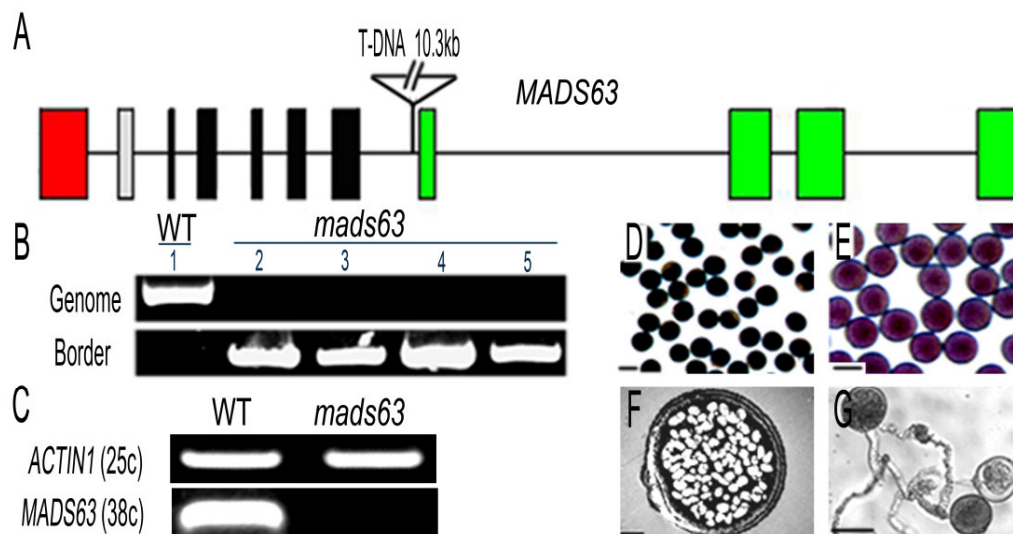
Supplemental Figure 6. The identification and microscopic analyses of pollen of *MADS63* T-DNA insertion mutants.
(A) Schematic diagram of the genomic structure of *MADS63* and T-DNA insertion position. Red box corresponds to exon encoding MADS-box domain, whereas grey box represents exon encoding I region; black boxes, K domain and light green boxes, C terminal, with lines indicating introns. The 10.3 kb T-DNA insertion disrupts the seventh intron of *MADS63* in the mutant line 2D-10691.
(B) The molecular identification of *mads63* T-DNA mutant. Four homozygous plants from F3 generation (lane 2-5) were determined by genomic PCR using T-DNA border primer combined with *MADS63* specific primer pairs (listed in Supplemental Table 3 online).
(C) RT-PCR analyses of *MADS63* transcript in homozygous mutant. Total RNA samples were prepared from anthers at stage 13 of wild-type and *mads63* homozygous mutant of F3 generation. The *ACTIN1* serves as standard control. The cycle number of PCR is shown in parentheses.
(D) to **(G)** The *mads63* pollen phenotypes investigated by I$_2$-KI staining **(D)**, Alexander staining **(E)**, transmission electron microscopy **(F)** and the germination assays *in vitro* **(G)**. Scale bars = 50 μm in **(D)**, **(E)** and **(G)**, and 3 μm in **(F)**.

Supplemental Figure 7. Schematic diagram of RNAi constructs and identification of *MADS68*i/*mads63* and *MADS68*i/*mads63* transgenic plants.

(A) Schematic representation of the *MADS62* and *MADS68* RNAi constructs for transformation into *mads63* background rice. The dsRNAi cassette containing two oppositely orientated fragments and an intron was introduced into pCam23A-Ubi bone vector. *Pubi*, Maize *UBIQUITIN* promoter. *Npt II*, G418-resistant selectable marker gene. *35S ter*, CaMV35S terminator. *P35S*, CaMV35S promoter. LB, left border. RB, right border.

(B) The molecular identification of *MADS68*i/*mads63* transgenic lines. The homozygous status for *MADS63* allele in *MADS68*i/*mads63* plants were determined by genomic PCR using T-DNA border primer combined with *MADS63* specific primer pairs (see Methods). The *MADS68* RNAi positive lines were identified by genomic PCR using *MADS68* specific primer pairs (*MADS68*-F and *MADS68*-R, see Supplemental Table 3 online; the length of the foreign RNAi fragment and the corresponding endogenous genomic sequence is 402 and 1345 bp, respectively). Six selected lines L3, L7, L14, L17, L23 and L28 were showed.

(C) The molecular identification of *MADS62*i/*mads63* transgenic lines. The homozygous status for *MADS63* allele in *MADS62*i/*mads63* plants were determined by genomic PCR using T-DNA border primer combined with *MADS63* specific primer pairs. The *MADS62* RNAi positive lines were identified by genomic PCR using *MADS62* specific primer pairs (*MADS62*-F and *MADS62*-R, see Supplemental Table 3 online; the length of the foreign RNAi fragment and the corresponding endogenous genomic sequence is 307 and 401 bp, respectively). Four selected lines L2, L6, L8, and L13 were showed.
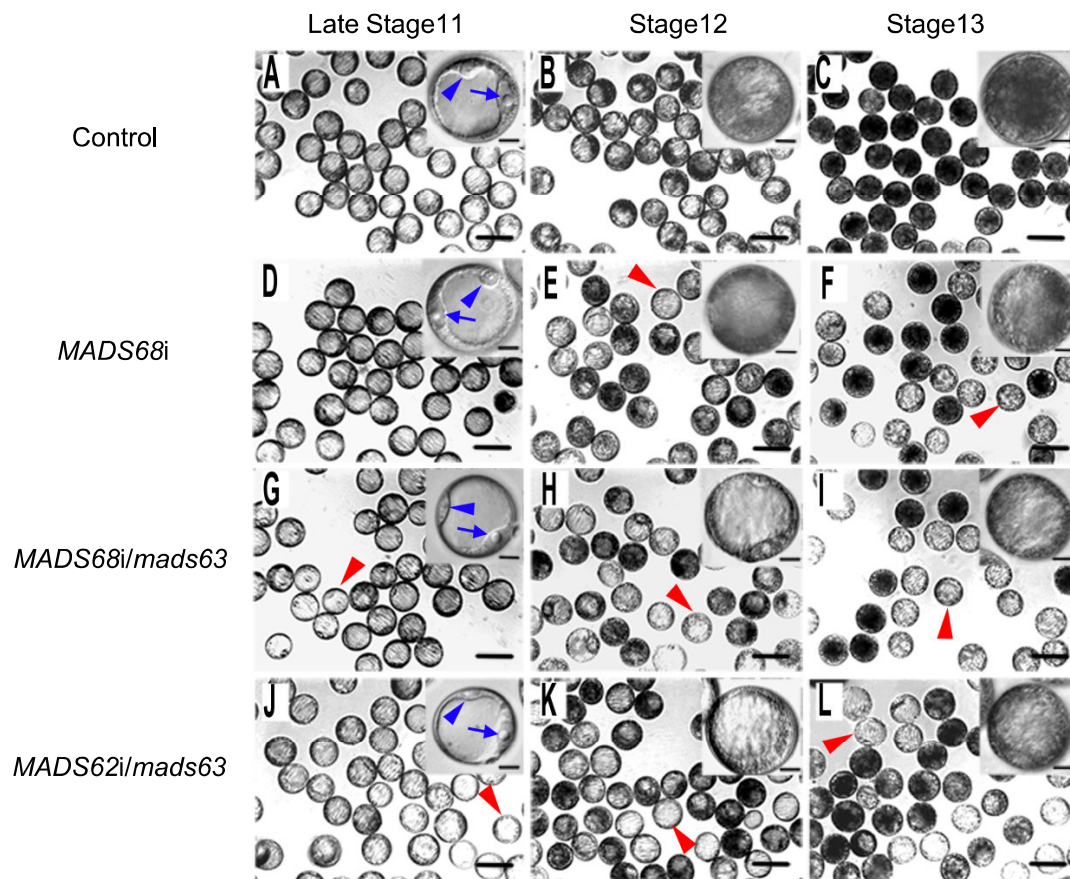
Supplemental Figure 8. Delay and arrest of pollen development occurs from bicellular stage.
(A) to **(C)** The development process of pollen from anther at late stage 11 to 13 of a control plant.
**(D)** to **(F)** The development process of pollen from anther at late stage 11 to 13 of line *MADS68* RNAi L7.
(G) to **(I)** The development process of pollen from anther at late stage 11 to 13 of line *MADS68i/mads63* L7.
**(J)** to **(L)** The development process of pollen from anther at late stage 11 to 13 of line *MADS62i/mads63* L13.
The pollen were collected and prepared from anthers at the late stage 11 (**[A]**, **[D]**, **[G]**, and **[J]**), when starch granules started to accumulate in bicellular pollen of control plants; anthers at the stage 12 (**[B]**, **[E]**, **[H]**, and **[K]**), when starch were filled in the most space of pollen in control plants; and anthers at the stage 13 (**[C]**, **[F]**, **[I]**, and **[L]**), when pollen achieved maturation with three nuclei in control plants. Insets show magnified pollen with apparent defects at the corresponding panel. Red arrowheads indicate abnormal pollen from *MADS68* RNAi, *MADS68i/mads63*, and *MADS62i/mads63* lines. Blue arrows indicate the vegetative nucleus and arrowheads indicate generative cell in insets (**[A]**, **[D]**, **[G]**, and **[J]**).
Scale bars = 100 μm in **(A)** to **(L)**, and 12.5 μm in the corresponding insets.
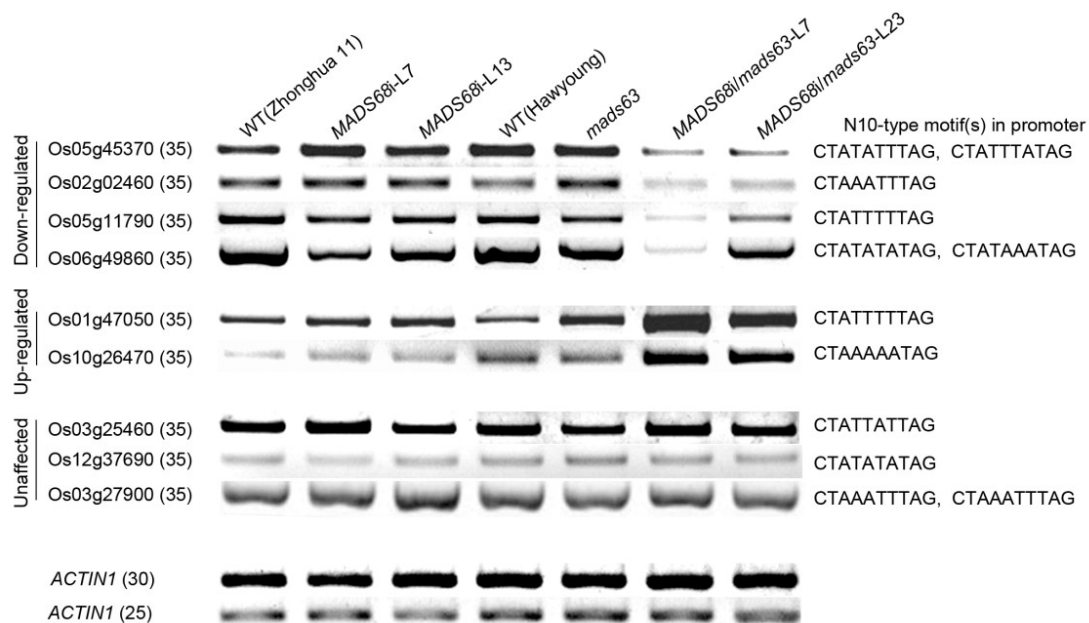
Supplemental Figure 9. RT-PCR analyses for a selection of nine *in silico* predicted target genes of the rice MIKC*-type protein complexes.

Total RNAs were pooled from anthers at stage 13 of wild-type (Zhonghua 11 and Hwayoung), *mads63* mutant, *MADS68* RNAi, and *MADS68i*/*mads63* lines. *ACTIN1* is amplified as the standard control. For each of the putative target genes, N10-type motif(s) present in the promoter region are listed in right. The cycle number of PCR is shown in parentheses for each gene.
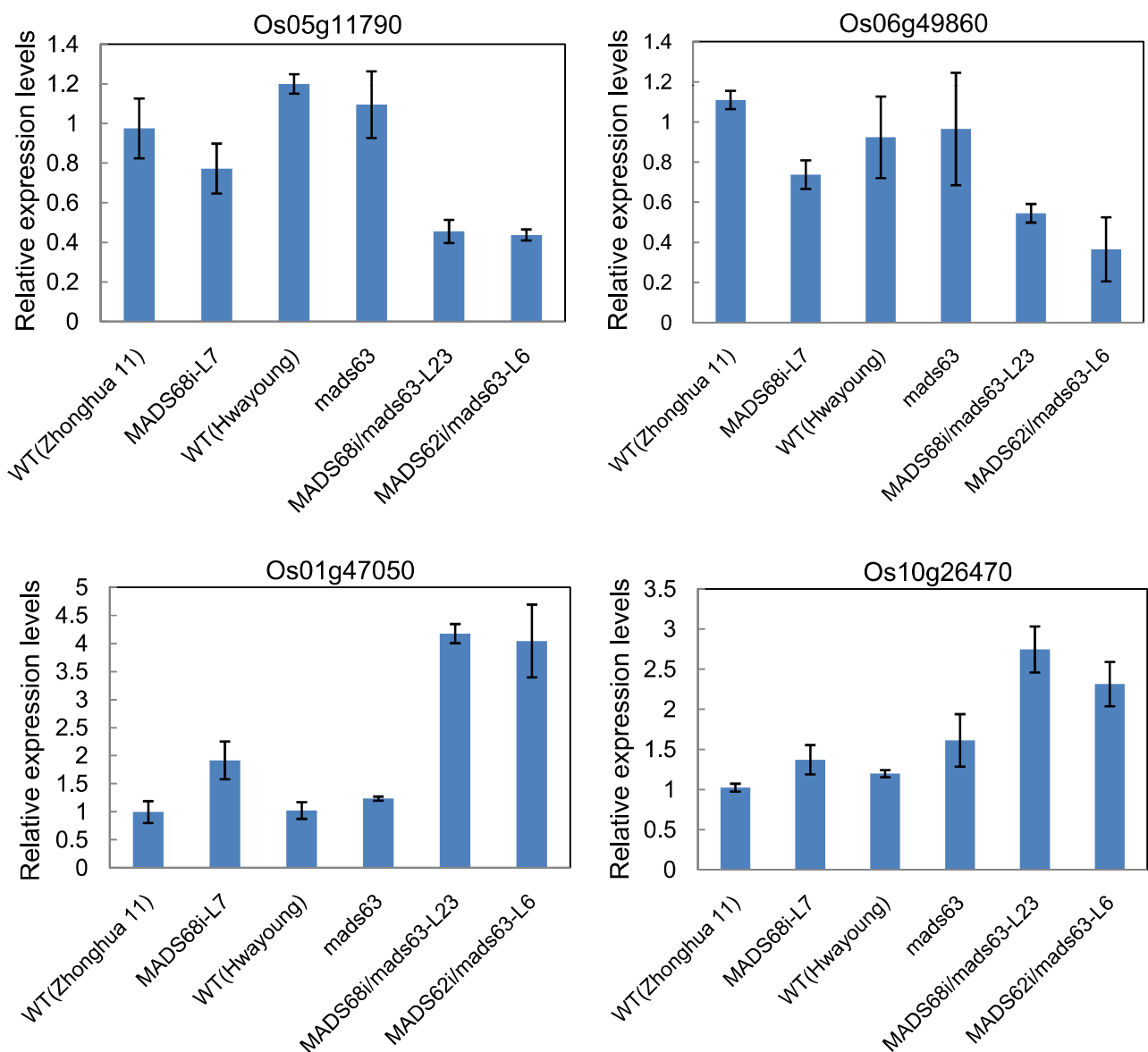
Supplemental Figure 10. qRT-PCR analyses of the expressions of putative target genes of the rice MIKC*-type protein complexes.
Total RNAs were pooled from anthers at stage 13 of wild-type (Zhonghua 11 and Hwayoung), *mads63* mutant, *MADS68* RNAi, *MADS68i/mads63*, and *MADS62i/mads63* lines. The samples were quantified using *UBIQUITIN* as a reference gene and the data are presented as mean $\pm$ SD (standard deviation, n = 3).

Supplemental Table 1. Detailed Description of MIKC*-Type Genes

Used for Alignment and Phylogenetic Analyses.

Supplemental Table 2. Genetic Analyses of the mads63 Mutant.

Supplemental Table 3. Gene-Specific Primers Used in This Study.

Supplemental Table 4. EMSA Probes Used in This Study.

Supplemental Data Set 1. Text File of Alignment of Nucleotide

Sequences of cDNAs of 58 MIKC*-Type Genes Used for the

Phylogenetic Analyses Shown in Figure 1.

Supplemental Data Set 2. List of Rice Late Pollen-Enriched Genes

with Their Expression Levels and Spatial Distribution of N10-Type

CArG-Box Motifs in Their Putative Promoters.

(please find the tables on the CD attached to the thesis or online: http://www.plantcell.org/content/early/2013/04/22/tpc.113.110049/suppl/DC1)

3. Discussion

Gene duplications generate the 'raw material' – literally new genetic material – for new genes. In many cases these new genes are not as selectively constrained, as the original function is continuously executed through the sister gene. Furthermore the copied gene already is a functional gene with e.g. a DNA-binding domain, interaction domains etc. So instead of generating a complex genetic function 'from scratch', a duplicated gene already has features that it can build up on.

The aim of this thesis was to follow up different outcomes of gene duplications and therefore obtain insights into the emergence of new genetic functions. I participated in three different publications in order to characterize MIKC-type MADS-box genes of two different subclades, which originated by gene duplication. The investigations were predominantly undertaken in the model crop plant rice (*Oryza sativa*) and close relatives (*O.* spec.). Two manuscripts discuss the significance of the *O. sativa* B_{sister} genes *OsMADS30* (manuscript I) and *OsMADS29* (manuscript II). The last one illuminates the role of the three MIKC*-type MADS-box genes *OsMADS62*, *OsMADS63* and *OsMADS68* (manuscript III).

3.1. Two unequal sisters: the function and evolution of the non-canonical B_{sister} gene *OsMADS30* and its canonical paralog *OsMADS29*

Prior to this thesis, only little was known about B_{sister} genes of monocots. The maize B_{sister} gene *ZMM17* is expressed in the ovule and its primordial tissues (Becker et al. 2002), the wheat B_{sister} gene *WBsis* is expressed in the ovule (Yamada et al. 2009). The three paralogous rice B_{sister} genes *OsMADS29*, *OsMADS30* and *OsMADS31* had been roughly characterized regarding their expression by microarray analyses and quantitative RT-PCR (*OsMADS30* only) (Arora et al. 2007). From this analysis it became clear already that *OsMADS29* has an expression pattern similar to other B_{sister} genes while *OsMADS30* is expressed throughout the plant life cycle (Arora et al. 2007). My co-workers could precisely determine the expression of *OsMADS29*, which is expressed exclusively during flower and seed development within the ovule and the developing seed (manuscript II). Expression was undetectable during the vegetative phase of the rice life cycle. This resembles the expression patterns of other B_{sister} genes like *TT16/ABS* from *A. thaliana* and *GGM13* from the gymnosperm *Gnetum gnemon*, which was further underlined through my analysis of all available expression data of B_{sister} genes at this time (manuscript II; Becker et al. 2002; Nesi et al. 2002). Further, my co-

workers could also predict that one function of *OsMADS29* is the mediation of cell death during seed development (manuscript II). In developing seeds of *OsMADS29* knock-down mutants the symplastic continuity was disturbed, which is supposed to be the nutrient source of the filial tissue (Krishnan and Dayanandan 2003; Oparka and Gates 1981). Furthermore, the degeneration of maternal tissue e.g. the nucellar projection and the pericarp, which is pivotal for grain filling and seed growth, was disturbed in transgenic plants (manuscript II). Hence, *OsMADS29* plays a key role during seed development, as do its homologs in higher eudicots (Mizzotti et al. 2012; de Folter et al. 2006).

However, whether *OsMADS29* has a function also during ovule development, which is implied by its expression during this process and suggested due to the function that was defined for *ABS/TT16* (Mizzotti et al. 2012), remains to be illuminated. Other research groups have studied *OsMADS29* in detail as well. The gene is most likely activated by the auxin response pathway (Yin and Xue 2012) and may also be involved in the cytokinin biosynthesis pathway (Nayar et al. 2013). OsMADS29 is forming homodimers as well as heterodimers with the B_{sister} protein OsMADS31 and is interacting with MADS-box proteins from many different subclades such as SEPALLATA-like, AGL6-like, AGAMOUS-like, SVP-like, AGL17-like, TM3-like and SQUAMOSA-like proteins (Nayar et al. 2014). Conclusively, *OsMADS29* very likely plays a key role during seed development in *O. sativa* within a huge regulatory network. The general significance of B_{sister} genes even beyond higher eudicots and their high functional conservation becomes evident.

Despite the overall similarities, *ABS* and *OsMADS29* differ in their functional details, but this can be expected considering their evolutionary distance. One could speculate that this might be due to the differences in seed morphology between grasses and higher eudicots. Whilst in grasses, the main storage of energy for the seedling is in the endosperm, seeds of eudicots have a tendency to store a higher percentage of their energy in the two cotyledons (Bewley et al. 2012). Ripe *A. thaliana* seeds have a highly reduced endosperm; as the endosperm is almost entirely consumed in the process of development. In contrast, seeds of *O. sativa* posses a large endosperm for storing nutrients for the developing embryo. The B_{sister} genes might well have been adapted for different purposes during seed development since the strategy of e.g. storing energy is so diverse. While *OsMADS29* contributes to the decline of maternal tissue by activating programmed cell death enabling grain filling, *ABS* contributes to seed protection (anthocyanidin synthesis) and ovule development (manuscript II; Nesi et al. 2002; Mizzotti et al. 2012). Considering the wide target gene variety amongst MADS-domain

proteins in general (Kaufmann et al. 2009) this range in target gene specificity is quite reasonable. Both, the recent role in monocots and eudicots, could be derived from a broader role in the most recent common ancestor.

Phylogenetically closest to *OsMADS29*-like genes are *OsMADS30*-like genes. In a meta-analysis of expression pattern of *Poaceae* B_{sister} genes I could show that while *OsMADS29*-like genes are exclusively and strongly expressed throughout the reproductive phase, the *OsMADS30*-like genes have a tendency towards a wider expression pattern. Further, their expression level on average is much lower than those of *OsMADS29*-like genes (manuscript I, II). The comparison of the expression of *OsMADS30* from *O. sativa* with its wild rice ortholog from *O. rufipogon OrMADS30*, revealed that the wild rice ortholog is around ten times weaker expressed than the ortholog of the cultivated rice, but the expression is restricted to the reproductive organs, more precisely to the developing floret and seed (manuscript I).

Despite considerable efforts, we could not find an aberrant flower or seed phenotype for *OsMADS30* T-DNA insertion lines (manuscript I). This might be due to a non-functionalization of *OsMADS30*, at least during seed development. However, it also is possible that the significance of *OsMADS30* is obscured by a genetic redundancy with (an)other gene(s).

However, I could find a number of evidences that argue in favour of a significance of *OsMADS30* during the vegetative growth phase of *O. sativa* (manuscript I). *OsMADS30* is expressed during the vegetative growth phase (especially the seedling and the root) of *O. sativa*, in contrast to almost all other B_{sister} genes that have been studied (Becker et al. 2002; de Folter et al. 2006; Nesi et al. 2002; Yamada et al. 2009). Further, I found that the sequence of the *OsMADS30* wild rice orthologs differed fundamentally from *OsMADS30* itself (manuscript I). Very likely, a region of 2.4 kb inserted downstream of the K-box of *OsMADS30*. This insertion encodes for the C-terminal domain of the extant OsMADS30 as well as intronic parts of the gene and the 3' UTR. Hence, there is no sequence homology between the C-terminal domain of OsMADS30 and all other OsMADS30-like proteins. This insertion is most likely caused by a helitron, a special kind of DNA transposable element, which spreads through a rolling circle mechanism (manuscript I; Kapitonov and Jurka 2007).

The C-terminal domain of MADS-domain proteins has been reported to confer transcriptional activation (Honma and Goto 2001). The modification of the C-terminal domain might have altered the function of the protein significantly. Hence, the heterologous C-terminal domain of

OsMADS30 might be capable of entirely new functions compared to OsMADS30-like proteins from wild rice as well as other grass species (manuscript I). For example, I could predict that the heterologous C-terminus has the propensity to form a transmembrane structure, which is very rare for MADS-box genes in general (manuscript I). These transmembrane properties might potentially enable a new function for OsMADS30.

We demonstrate that *OsMADS30* T-DNA insertion lines show a significantly reduced growth size in comparison to wild type plants (manuscript I) that could hint to a function in vegetative growth. However, further investigations are necessary to determine the true significance of *OsMADS30* in this respect, since it is not entirely clear whether the effect seen in the imaging-based phenotyping is due to the mutation in *OsMADS30* or an unspecific by-product of the T-DNA transformation. If *OsMADS30* was indeed revealed to definitely have an impact on vegetative growth of rice plants, it would be considerable that *OsMADS30* might have played a role during rice domestication by e.g. generating an advantage during vegetative growth that could have been selected for during the artificial selection of the domestication process.

Employing phylogenetic analysis, we could show that *OsMADS29*-like and *OsMADS30*-like genes constitute phylogenetic sister groups which most likely originated in the lineage leading to the *Poaceae* after Zingiberales had branched off (manuscript I; manuscript II). Using selection analyses, we could show that *OsMADS29*-like genes and *OsMADS30*-like genes evolved quite differently from each other. These results suggest that while *OsMADS29*-like genes evolved under negative selection, *OsMADS30*-like genes evolved under a more relaxed selection, close to neutral evolution (manuscript I). In line with this, OsMADS30-like proteins from different *Poaceae* species have more heterogeneous MADS-domains and less conserved paloea-AP3 and PI motifs than OsMADS29-like proteins from the respective species (manuscript I). This evolutionary pattern is in line with two inherently different functions of the two sister groups in grasses. *OsMADS29* fulfils a highly conserved role during seed development, also other *OsMADS29*-like genes are supposed to mediate crucial functions during ovule and seed development (Yamada et al. 2009; Becker et al. 2002). The evolutionary less conserved *OsMADS30* could be non-functionalized or acquired a new function well beyond its B_{sister} origin. Follow up experiments have to illuminate the actual function in vegetative development before further conclusions can be drawn.

Either way, the emergence of the two subclades through gene duplication yielded two entirely different outcomes.

3.2. MIKC*-type MADS-box genes in rice: when phylogeny mirrors biochemistry

The great conservation of pollen developmental processes within angiosperms is remarkable (Wei et al. 2010). MIKC*-type MADS-box genes have been found to be key regulators of late pollen development in *A. thaliana* (Verelst et al. 2007a, b). However, formerly it was not known whether this regulatory function is also conserved in monocots. Hence, we employed detailed expression, mutant and biophysical interaction analysis to illuminate the significance of MIKC*-type genes in *O. sativa*, as a representative of monocots (manuscript III).

O. sativa has three MIKC*-type MADS-box genes: *OsMADS62* and *OsMADS63* from the S-clade and *OsMADS68* from the P-clade. In higher eudicots MIKC*-type MADS-box genes are expressed predominantly during late stages of pollen development and P- and S-clade proteins form obligatory heterodimers (Adamczyk and Fernandez 2009; Verelst et al. 2007a). We could show that not only are the three rice paralogs expressed alike their eudicot orthologs from *A. thaliana* (Verelst et al. 2007a), also their mutant phenotype mimics those of certain *A. thaliana* MIKC*-type mutants (Adamczyk and Fernandez 2009; Verelst et al. 2007a). Whilst a knock-out/knock-down of one of the two S-clade genes does not result in an altered phenotype, the knock-out/ knock-down of either both S-clade genes or the one P-clade gene was followed by a massive impairment of male fertility due to impaired pollen ripening (manuscript III). This is very well explained by the interaction patterns of the MIKC*-type proteins. Just as in eudicots, we found S-clade proteins binding to DNA only in the presence of P-clade proteins and *vice versa* (manuscript III). Both S-clade proteins are able to bind DNA as a heterodimer with the P-clade protein OsMADS68. The presence of OsMADS68 as the only P-clade member in *O. sativa* is obligatory as suggested by the strong mutant phenotype.

The obligatory heterodimerization of P- and S-clade proteins is also mirrored within the phylogeny. In almost all species of ferns and seed plants that have been investigated, at least one P-clade as well as at least one S-clade protein has been observed (Kwantes et al. 2012; manuscript III).

As expected, the MIKC*-type heterodimers that we found via EMSA experiments have had a higher affinity for a N10-type CArG-box (consensus 5'-CAT(A/T)$_4$ATG-3'), as opposed to the "standard" SRE-type CArG-box bound by floral homeotic proteins (consensus 5'-CC(A/T)$_6$GG-3') (manuscript III). Similar observations have been made for *A. thaliana* MIKC*-type proteins (Verelst et al. 2007a). Thus, the binding specifics of MIKC*-type

proteins probably has been established in an ancestor of monocots and eudicots and remained highly conserved since then.
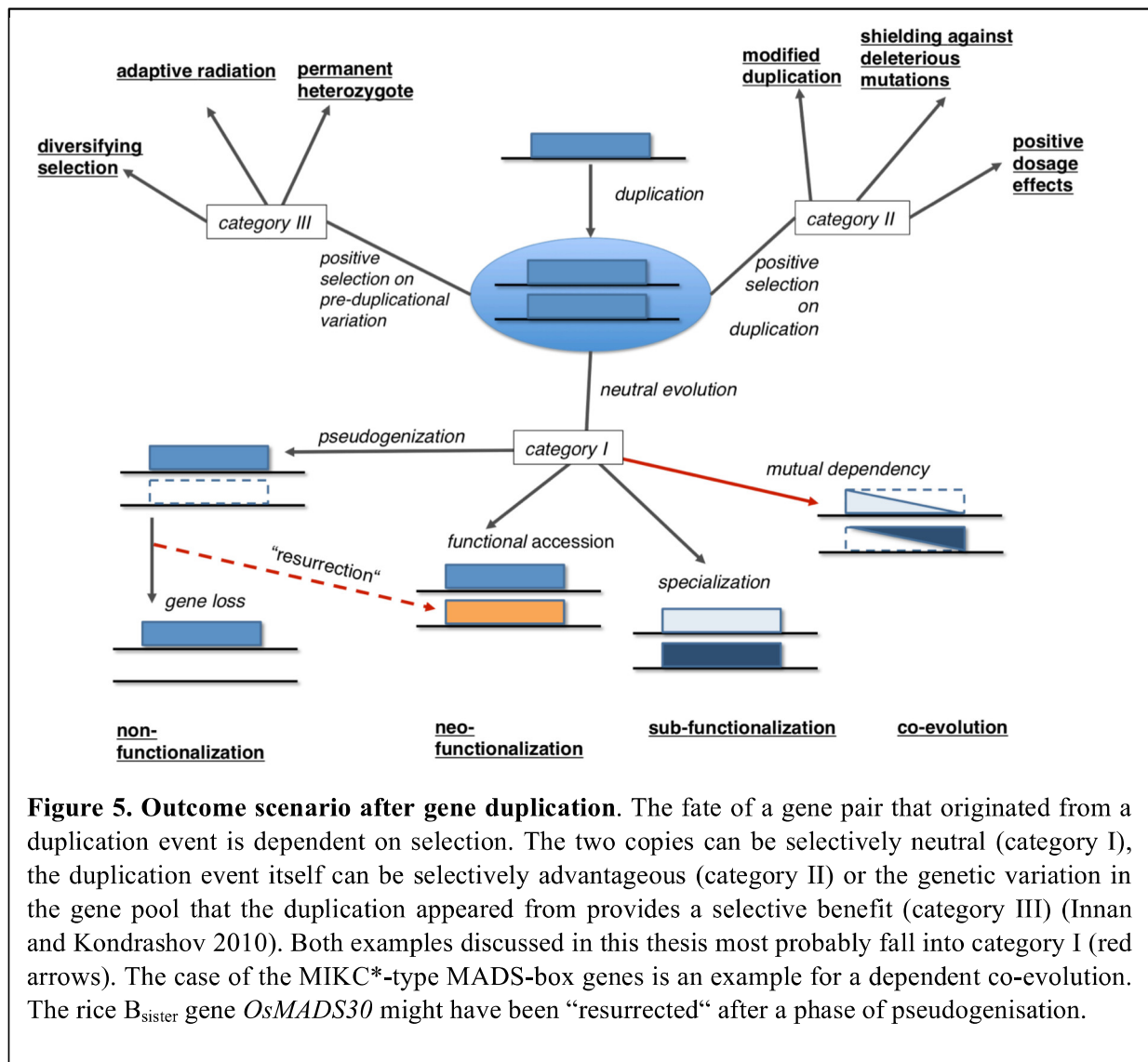
3.3. Two different outcomes of gene duplication: evolution of MIKC*-type and B$_{sister}$-genes

Both clades, B$_{sister}$ genes and MIKC*-type genes themselves were derived by gene duplication of an ancient gene: a B-class/B$_{sister}$ precursor gene and an ancestral MIKC–type gene, respectively (Becker et al. 2002; Kwantes et al. 2012). Both gene duplications appear to be quite ancient, however, the origin of MIKC*-type genes most probably dates back even farther than the emergence of B$_{sister}$ genes. Whilst the former are found in all land plants, the latter ones have been observed in seed plants exclusively (Figure 4) (manuscript I; manuscript II; Kwantes et al. 2012). The split into three different paralogous subclades of B$_{sister}$ genes that can be observed in grasses only is even more recent (manuscript I; manuscript III) whilst the diversification into P- and S-clade MIKC*-type MADS-box genes is at least as old as the lineage leading to ferns and seed plants (Kwantes et al. 2012). The outcomes of these gene duplications that are resulting in whole clades of genes today are remarkably different. On the one hand the split in P- and S-clade resulted in an intriguing interdependence of the two encoded proteins. This process was not "traditional" sub-functionalization, where parts of the biological function of the original gene are passed on to the two genes derived by the gene duplication. More so the two resulting proteins evolved "hand in hand", as a form of co-evolution (Figure 5). Most likely the protein encoded by the ancestral MIKC*-type gene was able to form a homodimer just as the single MIKC*-type protein MARCHANTIA POLYMORPHA MADS-BOX GENE 1 (MpMADS1) from the extant moss *M. polymorpha* does today (Zobell et al. 2010). The two copies generated by gene duplication were able to interact and most likely evolved afterwards into a state of obligate heterodimerisation. For this process to take place the biochemical properties of both proteins changed in a way that both proteins became mutually dependent on each other for conferring the biological function (Figure 5). By this mechanism it is ensured that a false expression of one of the genes does not lead to unwanted target gene regulation.

This scenario can be grouped into category I of the Innan and Kondrashov model (Figure 5) (Innan and Kondrashov 2009). B class genes in evolved in a similar manner. In this case, a duplication resulted *DEFICIENS*- and *GLOBOSA*-like genes and the corresponding proteins

form obligate heterodimers with each other in core eudicots, ensuring a confined determination of petals and stamens in the second and third floral whorl (Melzer et al. 2014).

In contrast, the consecutive duplication that led to three paralogous subclades of B_{sister} genes in the family of grasses yielded a completely different outcome (manuscript I; manuscript II). The expression patterns and functional similarities between different species of angiosperms make it likely that the ancestral grass B_{sister} gene had a function within ovule and/or seed development (Becker et al. 2002; Yamada et al. 2009; Nesi et al. 2002). As far as we can judge now, this function was later (at least in grasses) taken over by only one of the three paralogs, the *OsMADS29*-like genes. The expression of several *OsMADS29*-like genes has been found to resemble the expression pattern of *OsMADS29* and *ABS* what makes it likely that they also play similar roles during female organ and seed development (Becker et al. 2002; Nesi et al. 2002; Mizzotti et al. 2012; Yamada et al. 2009; manuscript I; manuscript II).

Figure 5. Outcome scenario after gene duplication. The fate of a gene pair that originated from a duplication event is dependent on selection. The two copies can be selectively neutral (category I), the duplication event itself can be selectively advantageous (category II) or the genetic variation in the gene pool that the duplication appeared from provides a selective benefit (category III) (Innan and Kondrashov 2010). Both examples discussed in this thesis most probably fall into category I (red arrows). The case of the MIKC*-type MADS-box genes is an example for a dependent co-evolution. The rice B_{sister} gene *OsMADS30* might have been "resurrected" after a phase of pseudogenisation.

Hence, the canonical B_{sister} function in grasses (i.e. seed development) is most likely conferred by the *OsMADS29*-like genes.

Although we cannot exclude a function of this gene during seed development, the low sequence conservation and the broad expression pattern may be indicators that the entire *OsMADS30* clade does not confer a canonical B_{sister} function and might even be on the way to non-functionalization (manuscript I). In this respect, the duplicative event generating the *OsMADS29/OsMADS30*-clade can also be grouped into category I according to Innan and Kondrashov (Figure 5) (Innan and Kondrashov 2009).

Given the possible minor role during seed development, the ancestral *OsMADS30* may have been a good candidate for neo-functionalization: the MADS domain provided a DNA-binding property and by the helitron insertion it received a quite unique property within its C-terminal domain among MADS-domain proteins: a transmembrane subdomain. There is evidence that *OsMADS30* confers a function within the vegetative phase of the plant growth; the newly acquired C-terminal domain might have enabled OsMADS30 to confer this new function, possibly due to the transmembrane features (manuscript I). Neo-functionalization and functional variation of genes via transposable elements might be a common evolutionary mechanism. Neo-functionalization due to frameshifts within the C-terminal domain has been proposed before for floral homeotic MADS-domain proteins (Vandenbussche et al. 2003), but a neo-functionalization of a MADS-box gene via a transposable element was to the best of my knowledge not described until now. I speculate that due to its new role in vegetative growth *OsMADS30* was possibly involved in controlling a trait that was selected for during rice domestication.

Hence, *OsMADS30* might be one example for neo-functionalization and even a "resurrective" process where the neo-functionalization is achieved by a "detour" via pseudogenization or a period of diminished significance of the respective gene (Figure 5). This might well have been facilitated via the insertion of a transposable element. The potential of transposable elements of influencing evolution is well recognized today (Feschotte and Pritham 2007) and transposable elements may in general be one important facilitator for neo-functionalization of pseudogenes. There are a few examples, which actually show that pseudogenes can readopt a function, e.g. by gene conversion (Trabesinger-Ruef et al. 1996). Neo-functionalization via sub-functionalization has been shown before to be a promising concept as well (Rastogi and Liberles 2005). The neo-functionalization of pseudogenes could be a common and hitherto underestimated evolutionary mechanism. Pseudogenes have even been proposed to be

renamed into "*pot*ogenes" due to their *pot*ential to regain functional significance and *OsMADS30* might be one example for this (Balakirev and Ayala 2003).

3.4. Conclusions and scientific significance

This thesis provided evidences for the non-canonical properties of the B$_{sister}$ gene *OsMADS30* as well as an understanding of the canonical role of *OsMADS29*. Further it underlines the high evolutionary conservation in terms of biophysical interaction between MIKC*-type proteins. I present a hypothesis according to which a putatively non-functionalized MADS-box gene underwent subsequently neo-functionalization. This work highlights the diverse fates genes can experience after duplication. It is becoming evident that constraints in gene regulatory circuits (like the heterodimerization of MIKC*-type MADS-box genes) and patterns of non- and/or neo-functionalization (as in the B$_{sister}$ lineage) cannot be fully understood without a thorough comprehension of the underlying gene duplication patterns.

3.5. Outlook

OsMADS30 might have been subject to neo-functionalization. However, despite the evidence presented in manuscript I, a number of questions remain to be answered. To further substantiate a role for *OsMADS30* in vegetative development, the analysis of additional mutant alleles, ideally loss of function mutants e.g. created by CRISPR/Cas9 could prove fruitful (Miao et al. 2013). Further, it is worth investigating whether the role of *OsMADS30* during flower and/or seed development is obscured by redundancy with another gene, e.g. another B$_{sister}$ gene or a class D gene. Thus, double mutants with *osmads29* or the D-class genes *osmads13* or *osmads21*, all of which have functions in ovule and seed development (manuscript II) (Dreni et al. 2007; Yin and Xue 2012), would be good candidates for further investigations. Since the C-terminal domain of class B MADS-domain proteins is known to influence higher order complex formation (Egea-Cortines et al. 1999; Fan et al. 1997; Pelaz et al. 2001), it is possible that OsMADS30 has other interaction properties than its ancestor without the heterologous C-terminal domain. Studying protein interactions between OsMADS30 and other MADS-domain proteins or even other transcription factors that are known to overlap in expression with *OsMADS30* could hence be promising. The comparison regarding protein interaction and transmembrane properties of the ancient and the extant

OsMADS30 protein would certainly be interesting as well and could enable further insights into the functional significance of *OsMADS30* and its exciting evolution.

The function of *OsMADS29* during seed development has been investigated in detail (Yin and Xue 2012; manuscript II), however it is also expressed quite early during ovule and carpel development. Given the developmental role of B_{sister} genes in eudicots (Becker et al. 2002, Nesi et. al 2002; de Folter et al. 2006) and its expression pattern, it is conceivable that *OsMADS29* also plays a role during ovule development. Again, this function could be obscured by redundancy with other MADS-box genes, e.g. the B_{sister} gene *OsMADS31* or a class D gene or even other transcription factors. The *A. thaliana* B_{sister} gene *ABS/TT16* was shown to act redundantly with the class D gene *SEEDSTICK* in ovule and seed development (Mizzotti et al. 2012). Before this redundancy was revealed it was hypothesized to play only a minor role in female organ development and seed pigmentation (Nesi et al. 2002). The same could be true for the role of *OsMADS29* during ovule development and further research could have the potential to unravel this.

MIKC*-type MADS-box genes have been well studied in *A. thaliana* and also in *O. sativa*. Since *O. sativa* has only three paralogs of MIKC*-genes, much less than *A. thaliana*, it would be a better candidate for further basic research, e.g. addressing detailed mutant phenotypes, double mutants, biochemical properties such as protein-DNA and protein-protein interaction and pollen ripening pathways. Since basic biochemical properties and expression are virtually the same in rice and *A. thaliana*, effects of genetic redundancy would be much easier to avoid in rice. MIKC*-type MADS-box genes are essential for pollen maturation and hence knock-out plants are sterile. This property could theoretically be utilised to generate male sterile plants, a trait often desired in GMO plant generation to avoid cross contamination of non-GMO crops and hybrids with wild species in close vicinity.

4. Summary

4.1. Summary

Gene duplications are a source for new genes (paralogs) with derived or entirely new functions. Hence gene duplications are one important requirement for organisms to evolve and to adapt to new environments. Here, I investigate the function and evolution of different rice MIKC-type MADS-box genes, which due to their high variety and significance in plant development are excellent objects to study the outcome of gene duplications.

B_{sister} MADS-box genes are known for their predominant role in plant female organ development. The rice B_{sister} gene *OsMADS29* confers the 'canonical' B_{sister} function in rice. With an expression exclusively during ovule and seed development and a seed impaired knock-down phenotype is similar to its co-orthologs in eudicots. With a detailed comparative meta-analysis of all available B_{sister} gene expression patterns I underline the significance of *OsMADS29*-like genes within grass ovule and seed development. These genes are all very strongly and more or less exclusively expressed during ovule and/or seed development thus resembling eudicot and gymnosperm B_{sister} gene expression patterns.

I further performed a detailed analysis of the rice gene *OsMADS30* showing that it is a non-canonical B_{sister} gene: it has a comparatively weak and broad expression and T-DNA-insertion lines show no aberrant seed phenotype. I further demonstrated that *OsMADS30* evolved under relaxed purifying selection and has a 3' region that strongly deviates from other grass B_{sister} genes, including closely related *OsMADS30* orthologs from wild rice species. This is due to the recent insertion of a hitherto unknown transposable element. T-DNA insertion lines of *OsMADS30* show diminished vegetative growth, which might be first evidence for a neo-functionalization of *OsMADS30*.

Beyond B_{sister} genes that function predominantly in female organ development, I also studied MIKC*-type MADS-domain proteins that function in pollen development: OsMADS62, OsMADS63 and OsMADS68. I demonstrate that they bind not as homodimers but only as heterodimers to DNA, an interaction behaviour that strongly resembles that of eudicot MIKC*-type MADS-domain proteins. This is congruent with the findings of my co-workers concerning phylogeny, protein-protein interaction, expression pattern and mutant analysis.

This thesis contributes to a deeper understanding of the developmental and evolutionary significance of various rice MADS-box genes. Eventually it shows how fundamentally different the results of gene duplications even within the same gene family can turn out.

4.2. Zusammenfassung

Genduplikationen sind eine Quelle für neue Gene (Paraloge) mit abgeleiteter oder neuer Funktion. Daher sind sie eine wichtige Voraussetzung für die Evolution von Organismen und deren Adaption an neue Umwelteinflüsse. In dieser Arbeit untersuche ich die Funktion und Evolution von verschiedenen MIKC-Typ-MADS-Box-Genen aus Reis, die sich dank ihrer hohen Diversität, Zahl und ihrer großen Bedeutung während pflanzlicher Entwicklungsprozesse gut eignen, um Genduplikation zu studieren. B_{sister}-MADS-Box-Gene sind für ihre bedeutende Rolle in der Entwicklung von weiblichen Reproduktionsorganen in Pflanzen bekannt. Das B_{sister}-Gen *OsMADS29* vermittelt die ‚kanonische' B_{sister}-Rolle in Reis. Mit einer Expression, die ausschließlich auf Ovulen- und Samenentwicklung beschränkt ist und einem mutanten Phänotyp bei dem die Samenentwicklung gestört ist, ähnelt es sehr seinen eudikotylen Co-Orthologen. Mittels einer komparativen Meta-Analyse aller verfügbaren B_{sister}-Gene unterstreiche ich die Signifikanz von *OsMADS29*-ähnlichen Genen während Ovulen- und Samenentwicklung in Gräsern. Diese Gene sind alle mehr oder weniger exklusiv während Ovulen- und/oder Samenentwicklung exprimiert und gleichen damit den eudikotylen und Gymnospermen-B_{sister}-Genen. Weiterhin habe ich eine detaillierte Analyse des Reisgens *OsMADS30* durchgeführt und gezeigt, dass es sich um ein nicht-kanonisches B_{sister}-Gen handelt: es hat eine vergleichsweise schwache und breitere Expression und T-DNA-Insertionslinien zeigen keinen abnormen Samenphänotyp. Außerdem konnte ich zeigen, dass *OsMADS30* unter abgeschwächtem Selektionsdruck evolviert ist und seine 3'-Region stark von der anderer Gras-B_{sister}-Gene abweicht, auch derer von sehr nahe verwandten Wildreisarten, was durch eine evolutionär junge Insertion eines bislang unbekannten Transposons verursacht wird. *OsMADS30*-T-DNA-Insertionslinien zeigen ein reduziertes Wachstum, was als erste Evidenz für eine Neofunktionalisierung von *OsMADS30* gedeutet werden kann. Über die B_{sister}-Gene hinaus, habe ich auch MIKC*-Typ-MADS-Domänen-Proteine in Reis untersucht, die eine Funktion während der Pollenentwicklung besitzen: OsMADS62, OsMADS63 und OsMADS68. Ich zeige, dass sie als obligate Heterodimere an DNA binden, ein Interaktionsmuster, wie von eudikotylen MIKC*-Typ-MADS-Domänen-Proteinen. Dies deckt sich mit den Ergebnissen meiner Co-Autoren bezüglich Phylogenie, Protein-Protein-Interaktion, Expressionsmuster und Mutantenanalyse. Diese Arbeit trägt zu einen tieferen Verstehen der Bedeutung von verschiedenen MADS-Box-Genen während der Entwicklung und ihrer Evolution an sich bei. Sie zeigt, wie fundamental verschieden die Resultate nach einer Genduplikation auch innerhalb derselben Genfamilie sein können.

5. Bibliography

Adamczyk BJ, Fernandez DE. 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiology,* **149**: 1713-1723.

Airoldi CA, Davies B. 2012. Gene duplication and the evolution of plant MADS-box transcription factors. *Journal of Genetics and Genomics,* **39**: 157-165.

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, *et al.* **2011**. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology,* **29**: 521-U84.

Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, *et al.* **2013**. Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications,* **4**.

Alexander MP. 1969. Differential staining of aborted and nonaborted pollen. *Stain Technology,* **44**: 117-122.

Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, *et al.* **2009**. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nature Structural & Molecular Biology,* **16**: 717-724.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology,* **215**: 403-410.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Pouplana LR, *et al.* **2000**. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America,* **97**: 5328-5333.

An SY, Park S, Jeong DH, Lee DY, Kang HG, Yu JH, Hur J, *et al.* **2003**. Generation and analysis of end sequence database for T-DNA tagging lines in rice. *Plant Physiology,* **133**: 2040-2047.

Angenent GC, Franken J, Busscher M, van Dijken A, van Went JL, Dons HJ, van Tunen AJ. 1995. A novel class of MADS box genes is involved in ovule development in petunia. *The Plant Cell,* **7**: 1569-82.

Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S. 2007. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics,* **8**.

Bailey JA, Eichler EE. 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Reviews Genetics,* **7**: 552-564.

Balakirev ES, Ayala FJ. 2003. Pseudogenes: Are they "junk" or functional DNA? *Annual Review of Genetics,* **37**: 123-151.

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

Becker A, Theißen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution,* **29**: 464-489.

Becker A, Winter KU, Meyer B, Saedler H, Theißen G. 2000. MADS-box gene diversity in seed plants 300 million years ago. *Molecular Biology and Evolution,* **17**: 1425-1434.

Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC. 2008. The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in *Arabidopsis* ovules. *The Plant Cell,* **20**: 2088-2101.

Bewley JD, Bradford K, H. H. 2012. *Seeds: physiology of development, germination and dormancy*: Springer Science & Business Media.

Chen G, Deng W, Peng F, Truksa M, Singer S, Snyder CL, Mietkiewska E, *et al.* **2013a**. *Brassica napus TT16* homologs with different genomic origins and expression levels encode proteins that regulate a broad range of endothelium-associated genes at the transcriptional level. *Plant Journal,* **74**: 663-677.

Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, *et al.* **2013b**. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nature Communications,* **4**: 1595.

Chen S, Tao L, Zeng L, Vega-Sanchez ME, Umemura K, Wang G-L. 2006. A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. *Molecular Plant Pathology,* **7**: 417-427.

Cho SC, Jang SH, Chae SJ, Chung KM, Moon YH, An GH, Jang SK. 1999. Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. *Plant Molecular Biology,* **40**: 419-429.

Clark A, Dean J, Tudor C, Saklatvala J. 2009. Post-transcriptional gene regulation by MAP kinases via AU-rich elements. *Frontiers in Bioscience,* **14**: 847-871.

Coen ES, Meyerowitz EM. 1991. The war of the whorls - genetic interactions controlling flower development. *Nature,* **353**: 31-37.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, **9**: 938-950.

Cserzo M, Wallin E, Simon I, vonHeijne G, Elofsson A. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering*, **10**: 673-676.

Cui R, Han J, Zhao S, Su K, Wu F, Du X, Xu Q, et al. 2010. Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant Journal*, **61**: 767-781.

D'Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, Noel B, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, **488**: 213-217.

Datta R, Chamusco KC, Chourey PS. 2002. Starch biosynthesis during pollen maturation is associated with altered patterns of gene expression in maize. *Plant Physiology*, **130**: 1645-1656.

de Folter S, Angenent GC. 2006. trans meets cis in MADS science. *Trends in Plant Science*, **11**: 224-231.

de Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, Grossniklaus U, Angenent GC, et al. 2006. A B-sister MADS-box gene involved in ovule and seed development in petunia and *Arabidopsis*. *Plant Journal*, **47**: 934-946.

Deng M, Bian H, Xie Y, Kim Y, Wang W, Lin E, Zeng Z, et al. 2011. *Bcl-2* suppresses hydrogen peroxide-induced programmed cell death via *OsVPE2* and *OsVPE3*, but not via *OsVPE1* and *OsVPE4*, in rice. *FEBS Journal*, **278**: 4797-4810.

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology*, **14**: 1935-1940.

Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Research*, **15**: 330-340.

Dominguez F, Moreno J, Cejudo FJ. 2001. The nucellus degenerates by a process of programmed cell death during the early stages of wheat grain development. *Planta*, **213**: 352-360.

Dreni L, Jacchia S, Fornara F, Fornari M, Ouwerkerk PBF, An G, Colombo L, et al. 2007. The D-lineage MADS-box gene *OsMADS13* controls ovule identity in rice. *Plant Journal*, **52**: 690-699.

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

Erdmann R, Gramzow L, Melzer R, Theissen G, Becker A. 2010. *GORDITA* (*AGL63*) is a young paralog of the *Arabidopsis thaliana* B-sister MADS box gene *ABS* (*TT16*) that has undergone neofunctionalization. *Plant Journal*, **63**: 914-924.

Fabian MR, Sonenberg N, Filipowicz W. 2010. Regulation of mRNA translation and stability by microRNAs. In: Kornberg RD, Raetz CRH, Rothman JE, Thorner JW, eds. *Annual Review of Biochemistry, Vol 79*.

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

Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, et al. 2003. MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell*, **15**: 2603-2611.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, **41**: 331-368.

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist*, **183**: 557-564.

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

Fujita M, Horiuchi Y, Ueda Y, Mizuta Y, Kubo T, Yano K, Yamaki S, et al. 2010. Rice expression atlas in reproductive development. *Plant and Cell Physiology*, **51**: 2060-2081.

Fukuda H. 1997. Programmed cell death during vascular system formation. *Cell Death and Differentiation*, **4**: 684-688.

Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, et al. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, **438**: 1105-1115.

Gaut BS. 2002. Evolutionary dynamics of grass genomes. *New Phytologist*, **154**: 15-28.

Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences of the United States of America*, **96**: 14400-14405.

Godbole A, Varghese J, Sarin A, Mathew MK. 2003. VDAC is a conserved element of death pathways in plant and animal systems. *Biochimica et Biophysica Acta-Molecular Cell Research*, **1642**: 87-96.

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

Gramzow L, Barker E, Schulz C, Ambrose B, Ashton N, Theißen G, Litt A. 2012. *Selaginella* genome analysis - entering the "homoplasy heaven" of the MADS world. *Frontiers in Plant Science,* **3**: 214.

Gramzow L, Ritz MS, Theißen G. 2010. On the origin of MADS-domain transcription factors. *Trends in Genetics,* **26**: 149-153.

Gramzow L, Theissen G. 2010. A hitchhiker's guide to the MADS world of plants. *Genome Biology,* **11**: 214.

Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jorstad TS, Wilson ZA, *et al.* 2009. The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis. PloS One,* **4**: e7817.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology,* **52**: 696-704.

Han M-J, Jung K-H, Yi G, Lee D-Y, An G. 2006. Rice immature pollen 1 (RIP1) is a regulator of late pollen development. *Plant and Cell Physiology,* **47**: 1457-1472.

Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M, Nishimura M. 2005. Vacuolar processing enzyme: an executor of plant cell death. *Current Opinion in Plant Biology,* **8**: 404-408.

Hayes TE, Sengupta P, Cochran BH. 1988. The human c-fos SERUM RESPONSE FACTOR and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes & Development,* **2**: 1713-1722.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics,* **169**: 1157-1164.

Henschel K, Kofuji R, Hasebe M, Saedler H, Munster T, Theissen G. 2002. Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens. Molecular Biology and Evolution,* **19**: 801-814.

Hiei Y, Ohta S, Komari T, Kumashiro T. 1994. Efficient transformation of rice (*Oryza sativa* L) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant Journal,* **6**: 271-282.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research,* **27**: 297-300.

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

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

Honys D, Twell D. 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis. Genome Biology,* **5**.

Hsing Y-I, Chern C-G, Fan M-J, Lu P-C, Chen K-T, Lo S-F, Sun P-K, *et al.* 2007. A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Molecular Biology,* **63**: 351-364.

Huang JQ, Wei ZM, An HL, Zhu YX. 2001. *Agrobacterium tumefaciens*-mediated transformation of rice with the spider insecticidal gene conferring resistance to leaffolder and striped stem borer. *Cell Research,* **11**: 149-155.

Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature,* **490**: 497-501.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics,* **11**: 97-108.

International Brachypodium Initiative I. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature,* **463**: 763-768.

Itoh J, Nonomura K, Ikeda K, Yamaki S, Inukai Y, Yamagishi H, Kitano H, *et al.* 2005. Rice plant development: from zygote to spikelet. *Plant and Cell Physiology,* **46**: 23-47.

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

Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of *Xenopus laevis. Cell,* **12**: 109-120.

Jacquemin J, Bhatia D, Singh K, Wing RA. 2013. The international *Oryza* map alignment project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Current Opinion in Plant Biology,* **16**: 147-156.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO Journal,* **6**: 3901-3907.

Jeong DH, An SY, Kang HG, Moon S, Han JJ, Park S, Lee HS, *et al.* 2002. T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiology,* **130**: 1636-1644.

Jiang N, Feschotte C, Zhang XY, Wessler SR. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology,* **7**: 115-119.

Junker A, Muraya MM, Weigelt-Fischer K, Arana-Ceballos F, Klukas C, Melchinger AE, Meyer RC, *et al.* 2014. Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Frontiers in Plant Science,* **5**: 770.

Jwa NS, Agrawal GK, Rakwal R, Park CH, Agrawal VP. 2001. Molecular cloning and characterization of a novel jasmonate inducible pathogenesis-related class 10 protein gene, JIOsPR10, from rice (*Oryza sativa* L.) seedling leaves. *Biochemical and Biophysical Research Communications*, **286**: 973-983.

K H. 1993. Anthesis, fertilization and development of caryopsis. *Science of the rice plant. Volume 1: Morphology*. Tokyo: Food and Agriculture Policy Research Center.

Kahn CL, Raphael BJ. 2008. Analysis of segmental duplications via duplication distance. *Bioinformatics*, **24**: 1133-1138.

Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, **98**: 8714-8719.

Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics*, **23**: 521-529.

Karimi M, Inze D, Depicker A. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science*, **7**: 193-195.

Kaufmann K, Anfang N, Saedler H, Theissen G. 2005. Mutant analysis, protein-protein interactions and subcellular localization of the Arabidopsis B-sister (ABS) protein. *Molecular Genetics and Genomics*, **274**: 103-118.

Kaufmann K, Muino JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P, Angenent GC. 2009. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PloS Biology*, **7**: 854-875.

Kellogg EA. 1998. Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences of the United States of America*, **95**: 2005-2010.

Kellogg EA. 2001. Evolutionary history of the grasses. *Plant Physiology*, **125**: 1198-1205.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*, **301**: 376-379.

Kim S-G, Lee S, Seo PJ, Kim S-K, Kim J-K, Park C-M. 2010. Genome-scale screening and molecular characterization of membrane-bound transcription factors in *Arabidopsis* and rice. *Genomics*, **95**: 56-65.

Kim SG, Kim ST, Wang Y, Yu S, Choi IS, Kim YC, Kim WT, et al. 2011. The RNase activity of rice probenazole-induced protein1 (PBZ1) plays a key role in cell death in plants. *Molecules and Cells*, **31**: 25-31.

Klukas C, Chen D, Pape J-M. 2014. Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant Physiology*, **165**: 506-518.

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

Kouchi H, Hata S. 1993. Isolation and characterization of novel nodulin cDNAs pepresenting genes expressed at early stages of soybean nodule development. *Molecular & General Genetics*, **238**: 106-119.

Krishnan S, Dayanandan P. 2003. Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). *Journal of Biosciences*, **28**: 455-469.

Krizek BA, Fletcher JC. 2005. Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics*, **6**: 688-698.

Kwantes M, Liebsch D, Verelst W. 2012. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. *Molecular Biology and Evolution*, **29**: 293-302.

LaGrandeur T, Parker R. 1999. The cis acting sequences responsible for the differential decay of the unstable MFA2 and stable PGK1 transcripts in yeast include the context of the translational start codon. *RNA*, **5**: 420-433.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *LIN-4* encodes small RNAs with antisense complementarity to *LIN-14*. *Cell*, **75**: 843-854.

Lee S, Kim J, Son JS, Nam J, Jeong DH, Lee K, Jang S, et al. 2003. Systematic reverse genetic screening of T-DNA tagged genes in rice for functional genomic analyses: MADS-box genes as a test case. *Plant and Cell Physiology*, **44**: 1403-1411.

Lehti-Shiu MD, Adamczyk BJ, Fernandez DE. 2005. Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. *Plant Molecular Biology*, **58**: 89-107.

Leipuviene R, Theil EC. 2007. The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. *Cellular and Molecular Life Sciences*, **64**: 2945-2955.

Li ZW, Trick HN. 2005. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *Biotechniques*, **38**: 872-876.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*, **290**: 1151-1155.

Ma H, Yanofsky MF, Meyerowitz EM. 1991. *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes & Development*, **5**: 484-495.

Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M*, et al.* **2008**. *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant and Cell Physiology,* **49**: 1135-1149.

Melzer R, Härter A, Rümpler F, Kim S, Soltis PS, Soltis DE, Theißen G. 2014. DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Annals of Botany,* **114**: 1431-1443.

Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, Wan J*, et al.* **2013**. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res,* **23**: 1233-1236.

Michaels SD, Amasino RM. 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell,* **11**: 949-956.

Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L. 2012. The MADS box genes *SEEDSTICK* and *ARABIDOPSIS BSISTER* play a maternal role in fertilization and seed development. *Plant Journal,* **70**: 409-420.

Mondragon-Palomino M, Hiese L, Haerter A, Koch MA, Theissen G. 2009. Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Evolutionary Biology,* **9**: 81.

Mou ZL, Wang XQ, Fu ZM, Dai Y, Han C, Ouyang J, Bao F*, et al.* **2002**. Silencing of phosphoethanolamine N-methyltransferase results in temperature-sensitive male sterility and salt hypersensitivity in *Arabidopsis*. *Plant Cell,* **14**: 2031-2043.

Nam J, Kim J, Lee S, An GH, Ma H, Nei MS. 2004. 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.

Nayar S, Kapoor M, Kapoor S. 2014. Post-translational regulation of rice MADS29 function: homodimerization or binary interactions with other seed-expressed MADS proteins modulate its translocation into the nucleus. *Journal of Experimental Botany,* **65**: 5339-5350.

Nayar S, Sharma R, Tyagi AK, Kapoor S. 2013. Functional delineation of rice MADS29 reveals its role in embryo and endosperm development by affecting hormone homeostasis. *Journal of Experimental Botany,* **64**: 4239-4253.

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

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

Nicholas KB, Nicholas HB, Deerfield DW. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW. NEWS,* **4**: 14.

Nishikura K. 2006. Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nature Reviews Molecular Cell Biology,* **7**: 919-931.

Ohdan T, Francisco PB, Sawada T, Hirose T, Terao T, Satoh H, Nakamura Y. 2005. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *Journal of Experimental Botany,* **56**: 3229-3244.

Ohmori S, Kimizu M, Sugita M, Miyao A, Hirochika H, Uchida E, Nagato Y*, et al.* **2009**. *MOSAIC FLORAL ORGANS1*, an *AGL6*-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell,* **21**: 3008-3025.

Ohno S. 1970. *Evolution by gene duplication.* Heidelberg, Germany: Springer.

Oparka KJ, Gates P. 1981. Transport of assimilates in the developing caryopsis of rice (Oryza sativa L) - ultrastructure of the pericarp vascular bundle and its connections with the aleurone layer. *Planta,* **151**: 561-573.

Oparka KJ, Gates P. 1984. Sink anatomy in relation to solute movement in rice (*Oryza sativa* L) - a summary of findings. *Plant Growth Regulation,* **2**: 297-307.

Palatnik JF, Allen E, Wu XL, Schommer C, Schwab R, Carrington JC, Weigel D. 2003. Control of leaf morphogenesis by microRNAs. *Nature,* **425**: 257-263.

Pařenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE*, et al.* **2003**. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *The Plant Cell,* **15**: 1538-1551.

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

Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF. 2001. APETALA1 and SEPALLATA3 interact to promote flower development. *The Plant Journal,* **26**: 385-394.

Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature,* **424**: 85-88.

Pollock R, Treisman R. 1991. Human SRF-related proteins - DNA-binding properties and potential regulatory targets. *Genes & Development,* **5**: 2327-2341.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics,* **14**: 817-818.

Prasad K, Zhang X, Tobon E, Ambrose BA. 2010. The *Arabidopsis* B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant Journal,* **62**: 203-214.

Prasad V, Stromberg CAE, Alimohammadian H, Sahni A. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. *Science,* **310**: 1177-1180.

Prasad V, Stromberg CAE, Leache AD, Samant B, Patnaik R, Tang L, Mohabey DM, *et al.* 2011. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications,* **2**: 480.

Radchuk V, Weier D, Radchuk R, Weschke W, Weber H. 2011. Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. *Journal of Experimental Botany,* **62**: 1217-1227.

Rastogi S, Liberles DA. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology,* **5**: 28-28.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, *et al.* 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science,* **319**: 64-69.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, *et al.* 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science,* **290**: 2105-2110.

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

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

Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Munster T. 2005. Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biology,* **7**: 307-314.

Rodriguez F, Oliver JL, Marin A, Medina JR. 1990. The general stochastic-model of nucelotide substitution. *Journal of Theoretical Biology,* **142**: 485-501.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* **19**: 1572-1574.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, *et al.* 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology,* **61**: 539-542.

Roshan U, Livesay DR. 2006. Probalign: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics,* **22**: 2715-2721.

Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K. 2013. *FLOWERING LOCUS C* in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat Commun,* **4**.

Ryu CH, You JH, Kang HG, Hur JH, Kim YH, Han MJ, An KS, *et al.* 2004. Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. *Plant Molecular Biology,* **54**: 489-502.

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

Scortecci KC, Michaels SD, Amasino RM. 2001. Identification of a MADS-box gene, *FLOWERING LOCUS M*, that represses flowering. *The Plant Journal,* **26**: 229-236.

Seo PJ, Kim S-G, Park C-M. 2008. Membrane-bound transcription factors in plants. *Trends in Plant Science,* **13**: 550-556.

Sharon D, Glusman G, Pilpel Y, Khen M, Gruetzner F, Haaf T, Lancet D. 1999. Primate evolution of an olfactory receptor cluster: Diversification by gene conversion and recent emergence of pseudogenes. *Genomics,* **61**: 24-36.

Shimatani Z, Nishizawa-Yokoi A, Endo M, Toki S, Terada R. 2015. Positive-negative-selection-mediated gene targeting in rice. *Frontiers in Plant Science,* **5**: 748.

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *European Journal of Biochemistry,* **229**: 1-13.

Shoshan-Barmatz V, Keinan N, Zaid H. 2008. Uncovering the role of VDAC in the regulation of cell life and death. *Journal of Bioenergetics and Biomembranes,* **40**: 183-191.

Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, Ooi LC-L, *et al.* 2013. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature,* **500**: 335-339.

Smaczniak C, Immink RGH, Angenent GC, Kaufmann K. 2012. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development,* **139**: 3081-3098.

Sommer H, Beltrán JP, Huijser P, Pape H, Lönnig WE, Saedler H, Schwarz-Sommer Z. 1990. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *The EMBO Journal,* **9**: 605-613.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* **22**: 2688-2690.

Swofford DL. 1993. PAUP - a computer-program for phylogenetic inference using maximum parsimony. *Journal of General Physiology,* **102**: A9-A9.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008a. Perspective - synteny and collinearity in plant genomes. *Science,* **320**: 486-488.

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008b. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research,* **18**: 1944-1954.

Tang W, Wang W, Chen D, Ji Q, Jing Y, Wang H, Lin R. 2012. Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in *Arabidopsis. Plant Cell,* **24**: 1984-2000.

Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics,* **38**: 615-643.

Theißen G. 2001. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology,* **4**: 75-85.

Theißen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, et al. 2000. A short history of MADS-box genes in plants. *Plant Molecular Biology,* **42**: 115-149.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* **22**: 4673-4680.

Trabesinger-Ruef N, Jermann T, Zankel T, Durrant B, Frank G, Benner SA. 1996. Pseudogenes in ribonuclease evolution: A source of new biomacromolecular function? *FEBS Letters,* **382**: 319-322.

Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönnig WE, Saedler H, et al. 1992. *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *The EMBO Journal,* **11**: 4693-4704.

Tufarelli C, Stanley JAS, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR. 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature Genetics,* **34**: 157-165.

Vandenbussche M, Theissen G, Van de Peer Y, Gerats T. 2003. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Research,* **31**: 4401-4409.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology,* **3**: research0034.

Vasudevan S, Tong Y, Steitz JA. 2007. Switching from repression to activation: MicroRNAs can up-regulate translation. *Science,* **318**: 1931-1934.

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

Verelst W, Twell D, de Folter S, Immink R, Saedler H, Muenster T. 2007b. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology,* **8**: R249.

Veron AS, Kaufmann K, Bornberg-Bauer E. 2007. Evidence of interaction network evolution by whole-genome duplications: A case study in MADS-box proteins. *Molecular Biology and Evolution,* **24**: 670-678.

Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, et al. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant Journal,* **40**: 428-438.

Wang L, Ge H, Hao C, Dong Y, Zhang X. 2012. Identifying loci Influencing 1,000-kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PloS One,* **7**: e29432.

Wang W, Haberer G, Gundlach H, Glaesser C, Nussbaumer T, Luo MC, Lomsadze A, et al. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nature Communications,* **5**: 3311.

Wei LQ, Xu WY, Deng ZY, Su Z, Xue Y, Wang T. 2010. Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa. BMC Genomics,* **11**: 338.

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

Wen LY, Chase CD. 1999. Mitochondrial gene expression in developing male gametophytes of male-fertile and S male-sterile maize. *Sexual Plant Reproduction,* **11**: 323-330.

Wernersson R, Pedersen AG. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Research,* **31**: 3537-3539.

Wikstrom N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society B-Biological Sciences,* **268**: 2211-2220.

Wu W, Huang X, Cheng J, Li Z, de Folter S, Huang Z, Jiang X*, et al.* **2011**. Conservation and evolution in and among SRF- and MEF2-type MADS domains and their binding sites. *Molecular Biology and Evolution,* **28**: 501-511.

Yamada K, Saraike T, Shitsukawa N, Hirabayashi C, Takumi S, Murai K. 2009. Class D and B-sister MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). *Plant Molecular Biology,* **71**: 1-14.

Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W*, et al.* **2014**. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications,* **5**: 5087.

Yang W, Kong Z, Omo-Ikerodah E, Xu W, Li Q, Xue Y. 2008. Calcineurin B-like interacting protein kinase OsCIPK23 functions in pollination and drought stress responses in rice (*Oryza sativa* L.). *Journal of Genetics and Genomics,* **35**: 531-543.

Yang X, Wu F, Lin X, Du X, Chong K, Gramzow L, Schilling S*, et al.* **2012**. Live and let die - The B-sister MADS-box gene *OsMADS29* controls the degeneration of cells in maternal tissues during seed development of rice (*Oryza sativa*). *Plos One,* **7**: e51435.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* **24**: 1586-1591.

Yin L-L, Xue H-W. 2012. The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. *Plant Cell,* **24**: 1049-1065.

Yoo S-D, Cho Y-H, Sheen J. 2007. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols,* **2**: 1565-1572.

Zhang D, Wilson ZA. 2009. Stamen specification and anther development in rice. *Chinese Science Bulletin,* **54**: 2342-2353.

Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M*, et al.* **2012**. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. *Nature Biotechnology,* **30**: 549-+.

Zhang YE, Long M. 2014. New genes contribute to genetic and phenotypic novelties in human evolution. *Current Opinion in Genetics & Development,* **29**: 90-96.

Zhang YH, Shewry PR, Jones H, Barcelo P, Lazzeri PA, Halford NG. 2001. Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. *Plant Journal,* **28**: 431-441.

Zhou Y, Mishra B. 2005. Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. *Proceedings of the National Academy of Sciences of the United States of America,* **102**: 4051-4056.

Zhou Z, Wang L, Li J, Song X, Yang C. 2009. Study on programmed cell death and dynamic changes of starch accumulation in pericarp cells of Triticum aestivum L. *Protoplasma,* **236**: 49-58.

Zobell O, Faigl W, Saedler H, Muenster T. 2010. MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Molecular Biology and Evolution,* **27**: 1201-1211.

6. Acknowledgements

Zu guter Letzt und ganz besonders danke ich meinen Mann, Kollegen und nicht zuletzt bestem Freund Rainer Melzer, ohne dessen bedingungslosen Rückhalt und unendliche Geduld, ohne dessen unbezahlbare fachliche und unbegrenzte emotionale Unterstützung, diese Arbeit so nicht zustande gekommen wäre. Danke für einfach alles.

7. Declaration of authorship (Ehrenwörtliche Erklärung)

Hiermit erkläre ich ehrenwörtlich, dass mir die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist und ich die vorliegende Arbeit selbst angefertigt habe. Alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen habe ich in meiner Arbeit angegeben. Bei der Auswertung des Materials sowie beim Verfassen des Manuskriptes haben mich die in der Danksagung dieser Arbeit genannten Personen unterstützt.

Ferner erkläre ich ehrenwörtlich, für die Anfertigung der Arbeit keinen Promotionsberater in Anspruch genommen zu haben und dass Dritte weder mittelbar noch unmittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Ich habe die vorliegende Dissertation bisher nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung vorgelegt. Auch habe ich weder diese Dissertation noch eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Susanne Schilling

Jena, 30. Juni 2016

8. Curriculum vitae

Particulars

Name: Susanne Schilling

Birth: 05.03.1985, Erfurt

Family status: married, two kids

Education

1991 – 1995   Hans-Seiler-Schule Erfurt

1995 – 2001   Martin-Luther-Gymnasium Erfurt

2001 – 2003   Albert-Schweizer-Gymnasium Erfurt

                         school leaving certificate: Abitur (grade 1.0)

Higher education

2003 – 2009   Study of Biology at the Friedrich Schiller University Jena

                         leaving certificate: Diplom (grade 1.2)

2006 – 2007   Student research assistent at the Genetic Chairs, University Jena

                         Project: Glutamine-repeats in AGAMOUS-Orthologs

2008 – 2009   Diplomarbeit at the Genetics Chair, University Jena

                         Topic: Funktionelle und evolutionäre Charakterisierung einer polymorphen glutaminreichen Region im C-terminalen Bereich von AGAMOUS-Orthologen

2009 – to date   PhD student at the Genetics Chair, University Jena

                         *Topic: **About sisters and starlets:***

                         *B_{sister} **and MIKC*-type MADS-box genes in rice development***

9. Conferences and Publications

Publications

X. Yang, F. Wu, X. Lin, X. Du, K. Chong, L. Gramzow, <u>S. Schilling</u>, A. Becker, G. Theißen and Z. Meng. (2012). Live and Let Die - The B_{sister} MADS-Box Gene *OsMADS29* Controls the Degeneration of Cells in Maternal Tissues during Seed Development of Rice (Oryza sativa). *PLoS One* 2012, **7/12**:e51435-e51435.

Y. Liu, S. Cui, F. Wu, S. Yan, X. Lin, X. Du, K. Chong, <u>S. Schilling</u>, G. Theißen and Z. Meng. (2013). Functional Conservation of MIKC*-Type MADS Box Genes in Arabidopsis and Rice Pollen Maturation. *The Plant Cell* 2013, **25**:1288–1303.

<u>S. Schilling</u>, L. Gramzow, D. Lobbes, A. Kirbis, L. Weilandt, A. Hoffmeier, A. Junker, K. Weigelt-Fischer, C. Kluklas, F. Wu, Z. Meng, T. Altmann and G. Theißen. (2015). Non-canonical structure, function and phylogeny of the B_{sister} MADS-box gene *OsMADS30* of rice (*OsMADS30*). *Plant Journal*, *84*(6), 1059-1072.

Poster

<u>S. Schilling</u>, D. Lobbes, L. Weilandt, A. Hoffmeier, L. Gramzow and G. Theißen: The black sheep amongst the sisters: On the Evolution and Function of the Rice B_{sister} gene *OsMADS30*. Botanikertagung (2011).