UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Molecular and Cellular Biology

XXIX Ciclo

New insights on the *AGAMOUS* subfamily genes in *Oryza Sativa*

Andrea Ravasio

PhD Thesis

Scientific tutor: Prof. Martin F. Kater

Academic year: 2015-2016

SSD: BIO/18; BIO/01

Thesis performed at:

Università degli Studi di Milano, Dipartimento di Bioscienze

Table of contents

PART 1

ABSTRACT

More than 30 years of plant biology have assessed the tight relationship existing between MADS-box genes and plant development and reproduction. Transcription factors of the MADS-box family have been associated with crucial processes throughout the whole life plant: germination, hormonal signalling, vegetative growth, transition from vegetative to reproductive phase, flower, fruit and seed development, embryogenesis. MADS-box are key regulators of the gene networks behind all of these processes; therefore, they have been (and continue to be) the subject of a number of molecular and functional studies which aim to add more and more details on the overall picture of how plants are made.

Among MADS-box, the *AGAMOUS* subfamily plays a fundamental role in plant reproduction; genes belonging to this subfamily have been associated with reproductive organ development and seed and fruit development in many different species. During evolution, the *AGAMOUS* subfamily underwent various events of duplication which led to the presence, in the extant plant species, of multiple *AGAMOUS*-like genes, often acquiring different functions thanks to subfunctionalization and neofunctionalization.

The aim of this thesis is to shed some more light upon the *AGAMOUS*-like genes present in rice (*Oryza sativa*).

In the first part, we demonstrate how the alternative splicing of a single aminoacid in rice *AGAMOUS* ortholog *OsMADS3* can alter its functionality, as proven by the different ability of the two isoforms to interact with other MADS-box proteins and by the different phenotype generated by their ectopic expression in *A. thaliana*.

In the second part, we present preliminary data about the role of rice *AGAMOUS* subfamily genes in controlling rice seed development. *AGAMOUS*-like genes have been associated with fruit and seed development in various species, but functional characterization experiments are difficult to perform on fruit plants because of their long life cycle; thus, rice represent the perfect model species to start with. Our experiments show that in multiple mutants in *AGAMOUS*-like genes fertility and seed development are severely impaired.

CHAPTER 1: INTRODUCTION

The Model Species: Arabidopsis Thaliana And Oryza Sativa

In the field of plant biology, two important model species that are used in laboratories all around the world are *Arabidopsis thaliana* (thale cress) and *Oryza sativa* (the common rice).

A. thaliana is a small annual (or rarely biennial) plant belonging to the Brassicaceae family and originating from Europe, Asia and northern Africa.

After full maturation, the stem is 20-25 cm tall. Leaves are mainly found at the base of the plant; they are 1.5-5 cm long and they are organized in a rosette with a spiral phyllotaxis and very short internodes. Few leaves (called cauline leaves) are also present on the inflorescence stem.

From the inflorescence meristem, following once again a spiral phyllotaxis, the floral meristems arise, each of whom develops into a single flower. The fruit is a 20 mm long silique containing about 50 seeds. Roots have a very simple structure, with a single main root growing vertically downwards and small secondary roots originating from the main one.

Despite not having any economical or agricultural relevance, *A. thaliana* has become since many years the main model species for plant biology because of its numerous advantages. First of all, its relatively short life cycle: 6 weeks from germination to mature seeds. The growth can be performed both in soil in the greenhouse or in Petri dishes. Small dimension, high number of seeds produced by each individual and self-pollination make Arabidopsis a perfect model for large-scale mutagenesis and genetic studies. Furthermore, Arabidopsis transgenic lines are routinely generated thanks to the use of *Agrobacterium tumefaciens* and to the "floral dip" protocol[1]. Moreover, the *A. thaliana* genome is one of the smallest plant genomes known (~157 Mb divided on 5 chromosomes) and also the first to be completely sequenced (in 2000[2]).

Oryza sativa is an annual plant belonging to the Poaceae family and it is the most widespread of the two species from which rice is commonly produced (the other is the African variety *O. glaberrima*), covering around 95% of the world rice cultivations. *O. sativa* species are further divided into two main subspecies: *O. sativa indica*, which is typical of tropical climates, and *O. sativa japonica*, which is typical of temperate climates (this is indeed the most cultivated variety in Italy).

Figure 1. *Arabidopsis thaliana* (left) and *Oryza sativa* (right).

After full maturation, rice stem, called culm, can reach between 120 and 195 cm, with long empty internodes and full nodes. Leaves are dark green and are covered by a rough layer of short hairs. The stem terminates with the inflorescence, formed by a main rachis from which multiple ramifications originate, each of which terminates with a spike bearing an average of 80 spikelets. The fruit consist of a flat elliptic caryopsis, enclosed by two glumes.

Rice has an agricultural and commercial, as well as scientific, value. It is largely cultivated and consumed all around the world, especially in Asian countries (China and India cover by themselves more than 50% of world production), but also in Europe, where Italy is the leading country in rice production. Similarly to *Arabidopsis*, it has a short life cycle (about 4 months) and a relatively small genome: 460 Mb for the *indica* variety (sequenced in 2001[3]) and 390 Mb for the *japonica* (sequenced in 2002[4]). It can be cultivated in high density in greenhouses without risking cross-contamination because it is almost entirely self-pollinating. Moreover, there's an efficient protocol for Agrobacterium-mediated rice transformation through in-vitro tissue culture[5]. Another very interesting aspect is that rice belongs to the same family as wheat, barley, maize and sorghum, meaning the most cultivated cereals worldwide. Thanks to all of these elements, rice has become in the past years a widely used model species for graminaceae and monocotyledons research in general.

Flower Structures

Flowers of *A.thaliana* and *O.sativa* appear to be very different, as one may expect from two species whose evolutionary divergence dates back to more than 150 mya. However, their structure is actually very similar (Figure 2); this is not only true for these two species, but for angiosperms in general.

In the vast majority of angiosperms, the floral organs are organized into four concentric regions called whorls. The first two outer whorls contain the non-reproductive (or perianth) organs. In

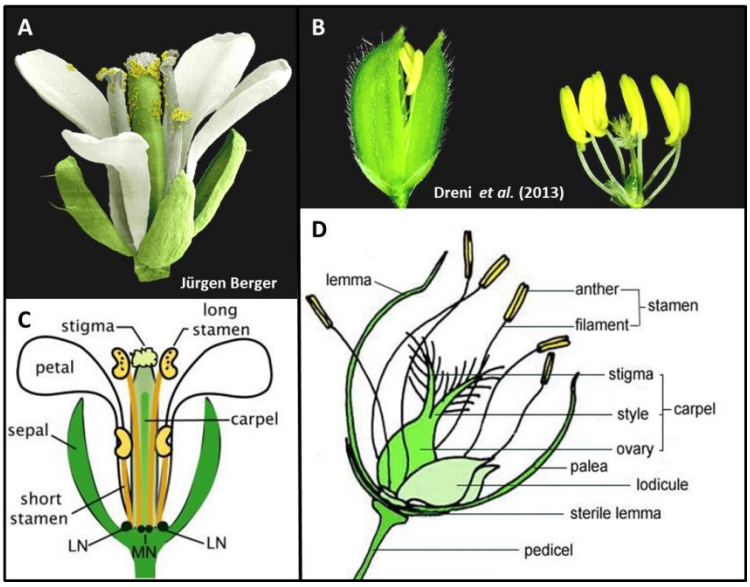


Figure 2. Flowers of *A. thaliana* (A, C) and *O. sativa* (B, D).

Arabidopsis these organs are leaf-like organs called sepals (in the first whorl) and the petals (in the second whorl), arranged in a cross-like shape and alternated to the sepals. The two inner whorls contain the reproductive organs: the stamens, the male reproductive organs, in the third whorl and the pistil, the female reproductive organ, in the fourth whorl. In Arabidopsis we find six stamens, of which four longer called medial and two shorter called lateral, and one pistil, resulting from the fusion of two carpels[6] (Figure 2C).

The rice flower is part of the spikelet, a typical structure of graminaceae, composed of the flower itself and by four bract-like organs called glumes: two small rudimentary glumes at the very base of the spikelet and two other glumes called empty or sterile glumes (Figure 2D).

The flower looks morphologically quite different but is structurally very similar to the Arabidopsis flower. In the first whorl we find two distinct glumes called palea and lemma. The second whorl contains the lodicules, two small organs located at the base of the flower, on the lemma side; they can be considered as modified petals[7] and, indeed, they have in rice the function of petals, which is to cause anthesis by separating the organs enclosing the flower, in this case palea and lemma. In the inner whorls we find, as in Arabidopsis, six stamens and one pistil (a monocarpellary one, in this case) (Figure 2D).

The MADS-Box Family

In 1990 the *A. thaliana* gene ***A****GAMOUS*[8] and the *Anthirrinum majus* gene ***D****EFICIENS*[9] have been identified as regulators of floral organ identity. The DNA-binding sequence of about 60 amino acids of these transcriptional factors showed a high similarity to that of two other genes previously identified, the human gene ***S****ERUM RESPONSE FACTOR (****S****RF)*[10] and the *Saccharomyces cerevisiae* gene ***M****INICHROMOSOME MAINTENANCE 1 (****M****CM1)*[11]. Therefore, transcription factor genes that

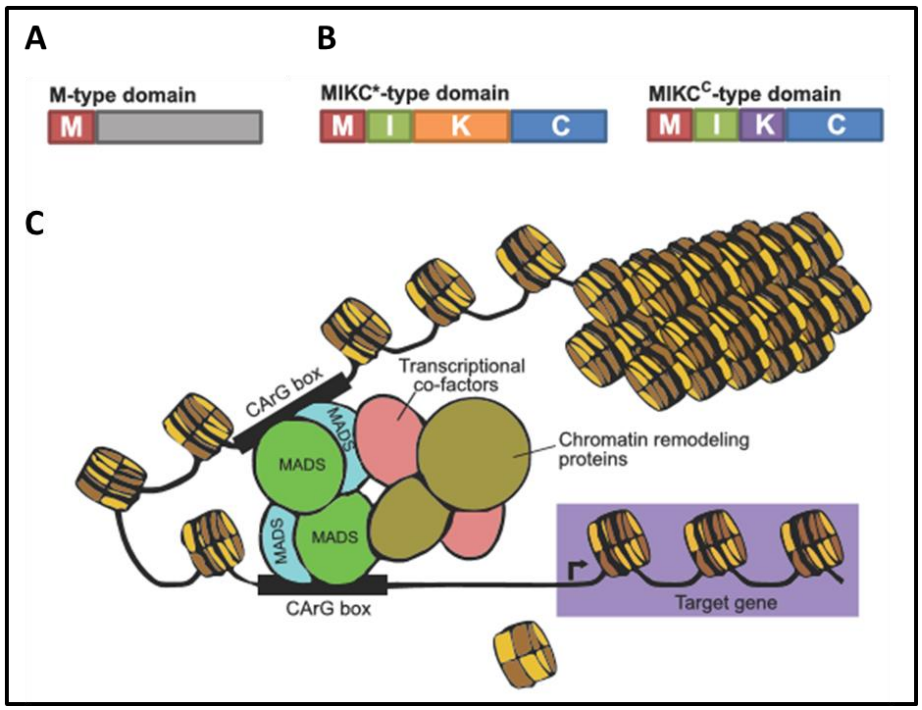


Figure 3. A) Schematic structure of type I MADS-box factors; B) Schematic structures of type II MADS-box factors; C) Model for MADS-box complexes formation and transcriptional activity. (Images modified from Smaczniak *et al.* (2012)[14])

encode such a DNA binding domain are called MADS-box genes (from the first letters of the four founding members)[9].

Based on other protein domains, the MADS gene family can be divided in two big subfamilies, type 1 MADS-box (or *SRF*-like) and type 2 MADS-box genes (or *MEF2*-like)[12] (Figure 3A,B). Type 1 MADS-box genes are more abundant in plants, but no function was assigned to them until 2008, when they were found to be expressed in gametophytes and in the first stages of embryo and endosperm development[13,14]. Type 2 MADS-box genes have been instead deeply studied since early '90s in plant biology because many of them are related with embryogenesis, flowering time and plant, flower and fruit development[14,15]. They are also known as MIKC MADS because of their conserved molecular structure, formed by four domains: the N-terminal MADS DNA-binding domain (M), the intervening (I) and keratin-like (K) domains, involved in dimerization and formation of higher order complexes, and the C-terminal domain (C), involved in transcriptional regulation and probably also in complex formation. MIKC MADS factors can be further divided in MIKC^C and MIKC* based on K-domain length, which is longer in the second group[15] (Figure 3B).

Despite the great amount of studies on MADS proteins, their molecular mechanisms of action have not being fully uncovered yet. The MADS domain recognizes a DNA consensus sequence like CC(A/T)$_6$GG, called CArG-box. In 1992 it was shown that MADS proteins can bind the DNA as dimers[16]; ten years later, they have also been proven capable of forming higher order complexes, thus leading to the "floral quartet" model. According to this model, a tetramer made of two MADS dimers can bind a DNA region with two CArG-boxes and create a loop between the two binding

sites[17]; more recently it has been shown that the complex can also bind regions with one proper CArG-box and other sites with less affinity[14] (Figure 3C).

Recent studies suggest that MADS genes had a fundamental role in the evolution of plants in general and for angiosperms in particular. As a matter of fact, many subfamilies of MIKC^C genes seem to have evolved starting from a common ancestor of angiosperms and gymnosperms and expression analysis indicate that many basic functions of these genes could have been conserved between the two classes. Moreover, MADS-box genes are involved in almost all relevant aspects of plant development and reproduction; some of them, like the *SQUAMOSA* and *SEPALLATA* subfamilies, can only be found in flowering plants, suggesting a key role in the origin and evolution of the flower itself[14,18].

The ABC Model

In 1991 Coen and Meyerowitz proposed a genetic model for the control of the floral organ identity, called the ABC model (Figure 4), after the observation of similar homeotic mutants in *A. thaliana* and *Anthirrinum majus*[19]. Being these two species quite distant in terms of evolution, they hypothesized the existence of a conserved structural logic.

According to the ABC model, the development of the different floral organs is regulated by three classes of genes (mainly MADS-box genes): class A genes are expressed in the outer two whorls; class B genes are expressed in the second and third whorl; class C genes are expressed in the inner two whorls. In the first whorl only A function genes are expressed which leads to sepal identity. In the most inner whorl only the C function is expressed which leads to carpel identity. The interaction between A and B genes in the second whorl and between B and C genes in the third, instead, respectively lead to petal and stamen identity. Another fundamental rule in the model is that A and C class genes are mutually exclusive expressed[19].

In *A.thaliana* there are two A class genes, *APETALA1 (AP1)* and *AP2*, two B class genes, *AP3* and *PISTILLATA (PI)* and one C class gene, *AGAMOUS (AG)*. They are all type 2 MADS-box factors, with the exception of *AP2*, which is an ERF-like gene[20].

AP1 belongs to the *SQUAMOSA* subfamily, together with *CAULIFLOWER (CAL)* and *FRUITFULL (FUL)*; these genes have a role in floral meristem determinacy, as shown by the mutants *ap1 cal* and *ap1 cal ful*, where the inflorescence meristem proliferates abnormally, forming a cauliflower-like curd instead of terminating in a single flower[21,22]. According to phylogeny, it looks like this was

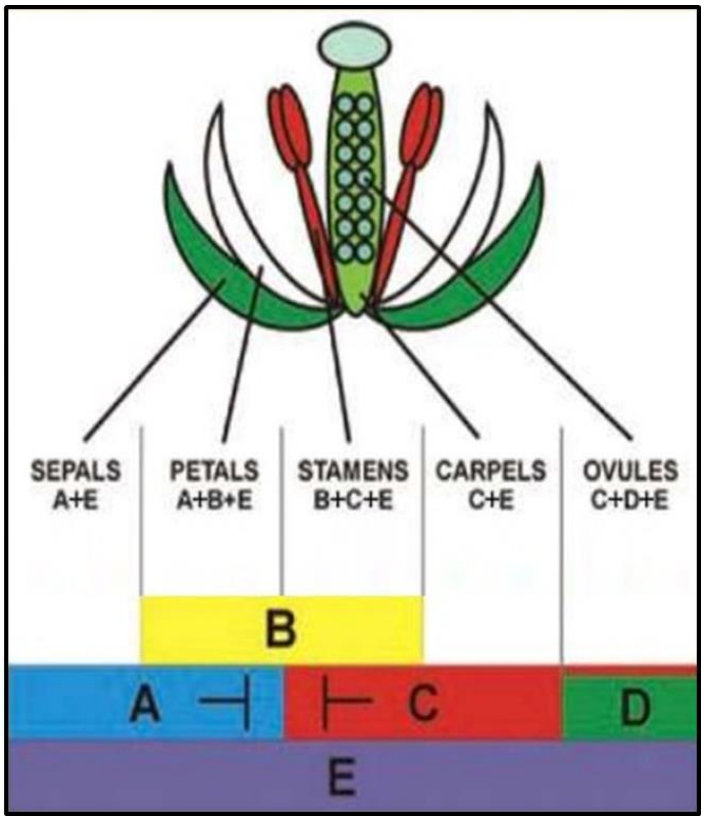


Figure 4. The ABCDE model.

the ancestral function of these genes and that only later *AP1* has become responsible for sepals and petals identity. In the *ap1* single mutant sepals are converted into bract-like organs and chimeric organs or no organs at all can be found in the second whorl; it is not rare to observe new flowers arising from the first whorl, thus confirming the role of *AP1* in meristem determinacy[23]. The phenotype of the *ap2* single mutant is instead more clearly that of an *a* function mutant: sepals are converted into carpels or carpelloid leaves and petals into stamens[20,21,24].

AP3 and *PI* belong to the *DEFICIENS* and *GLOBOSA* subfamilies, respectively. Phylogenetic analysis on extant gymnosperms suggests an ancestral role of B class genes in male reproductive organs identity and development and, only later, their recruitment for petal identity. Single mutants *ap3* and *pi* show clear petal-to-sepal conversion and stamen-to-carpel conversion, thus confirming their B function and, indirectly, also the other two functions[20,25].

In the *ag* single mutant, reproductive organs are replaced by perianth organs and also floral meristem determinacy is lost, so the flower continues to produce new whorls of perianth organs in its centre. This indicates a role of *AG* in floral meristem determinacy, in addition to its reproductive organ identity function[8,20].

In 1997 Colombo and colleagues discovered two genes in petunia (*Petunia hybrida*), *FLORAL BINDING PROTEIN 7 (FBP7)* and *FBP11,* whose expression was localized in the gynoecium just before ovule primordia formation and, later on, only in developing ovules; moreover, their co-suppression led to the conversion of ovules into carpelloid organs. Ectopic expression of *FBP11* could produce ectopic ovules on sepals and petals without altering the expression of other MADS-box genes, in particular C-class genes, thus confirming itself as the only responsible for the development of those ovules. All of this, together with the observation of mutants in which ovules and carpels arise independently and with the fact that there are plants (like all gymnosperms) in

which ovules naturally form without a carpel, brought to the introduction in the ABC model of a fourth class, named D-class, whose members are directly responsible for ovule identity and development[26–28] (Figure 4). In 2003 the *A. thaliana* gene *SEEDSTICK* was identified as homologous of *FBP7*; its role mainly concerns the correct seed dispersion, but it has been shown that it is also responsible for ovule identity determination. Despite sharing this function with two other genes, *SHATTERPROOF1/2 (SHP1/2)*, *STK* is the only official D-class gene in Arabidopsis[29,30].

In 2001 a group of *A. thaliana* genes, named *SEPALLATA1/2/3/4 (SEP1/2/3/4)*, was discovered. These genes are involved in the development of all floral organs, since in the triple mutant *sep1 sep2 sep3* all floral organs are converted to sepals and in the quadruple mutant *sep1 sep2 sep3 sep4* they are all converted to leaf-like organs. Successive studies demonstrated the ability of the SEP proteins to interact and create tetramers with members of the A, B, C and D classes and that these interactions are fundamental for their functionality; thus, a new class, the E class, was added to the model[31,32](Figure 4).

In the past years, the development of NGS technologies, the increased availability of completely sequenced genomes and the development of protocols to perform functional studies in many other species allowed to demonstrate the widespread validity of the ABCDE model in both dicotyledons and monocotyledons.

Recently, Causier and colleagues proposed to modify the ABCDE model into an (A)BCD model, where the (A) function consists in the establishment of the correct floral context in which the other classes will develop their function (thus including also the E function). This proposal came from the observation that the A function as originally described only exists in *Arabidopsis*, but it is difficult to be applied in other species[33].

The AGAMOUS Subfamily

The studies presented in this thesis mainly revolve around members of C and D classes.

Phylogenetic studies have clearly shown how all known members of C and D classes belong to a large subfamily of MADS-box genes called the *AGAMOUS* subfamily, from the name of the Arabidopsis gene which was founding member of the family[34].

As previously said, C-class genes have three main functions in flower development: the establishment of stamen and carpel identity in the inner whorls, the confinement of A-class genes

expression in the outer whorls, and the floral meristem determinacy. D-class genes, instead, control ovule identity, often redundantly with C-class genes.

Within the *AGAMOUS* subfamily, many distinct phylogenetic lines can be found, originated by various duplication events. The first division is between *AG* and *AGL11* lineages, also simply known as C and D lineages. The *AG* lineage can be further divided into the *euAG* and *PLE* lineages[35].

In *A. thaliana,* four genes belong to the *AGAMOUS* subfamily: *AGAMOUS* itself, *STK, SHP1* and *SHP2. AGAMOUS* is the only *euAG* gene and it is also the only official C-class gene of Arabidopsis. *SHP1* and *SHP2* belong to the *PLE* lineage. Despite they retain an ancestral ability to develop a C function, their different temporal and spatial expression profile prevents them from cooperating with *AGAMOUS* in the C function. Their specific function is instead mainly linked to seed dispersion: in the double mutant *shp1 shp2* the siliques can't open because the valve margins are not lignified and, thus, seed cannot be dispersed[36].

STK is the only *AGL11* gene in Arabidopsis and founding member of this sub-lineage (*AGL11* is another name for *STK*). Despite being the only D-class gene of Arabidopsis, the proper *d* mutant can only be seen when *stk* mutant is combined with *shp1/2* mutants: in the *stk shp1 shp2* triple mutant ovule identity is lost and ovule integuments are converted into carpelloid organs. The phenotype is enhanced when a mutation in *BEL1* is combined to the *stk shp1 shp2* triple mutant: ovule integuments are converted into leaf-like tissues[37]. Also *AG* has a role in determining ovule identity. *STK*, differing from *SHPs*, does not retain any C function[29,30].

Observations on mutants in other dicotyledonous show that the association between the *AG* subfamily and the C and D functions is highly conserved, with various degrees of redundancy, sub-functionalization and neo-functionalization. In *Anthirrinum majus* we find two *AG*-like genes, the *PLE* gene *PLENA* and the *euAG* gene *FARINELLI*[38]; opposite to Arabidopsis, the C-function is associated to the *PLE* gene and not to the *euAG* one. Also in *Petunia hybrida* there are two members of the *AG* lineage, the *euAG* gene *FBP6* and the *PLE* gene *pMADS3*[39]; in this case, the analysis on the *fbp6* and *pMADS3-RNAi* mutant lines indicates that *FBP6* has a more pronounced C-function compared to *pMADS3*, but the complete loss of stamen and carpel identity is obtained only when the two mutations are combined[39]. Recently, mutations in the *AGAMOUS*-like genes *NbAG (euAG)* and *NbSHP (PLE)* of *Nicotiana benthamiana* have been generated by virus-induced gene silencing (VIGS)[40]. Again, the single mutants display various defects in reproductive organ identity and in meristem determinacy, but the true *ag*-like phenotype is best observed when the two mutations are induced simultaneously[40].

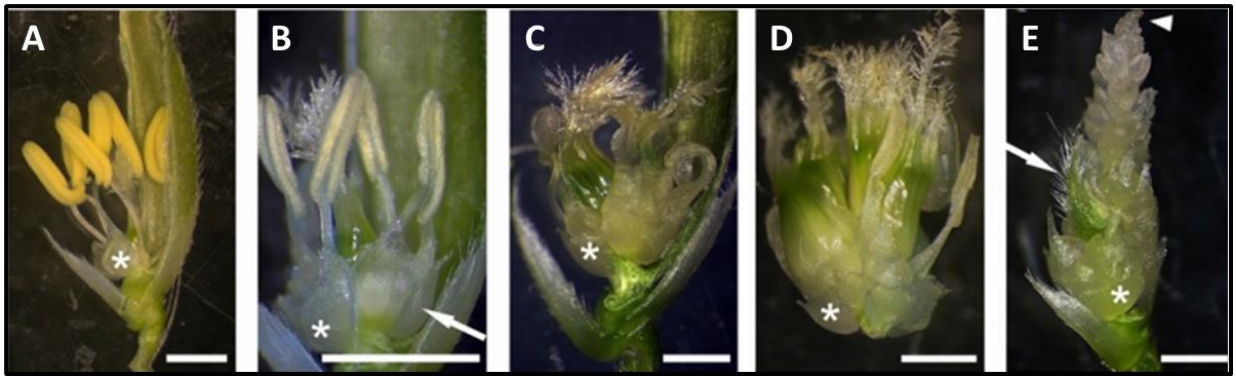


Figure 5. A) WT rice flower; B, C) *osmads3-3* mutants; D) *osmads3-3 osmads58/+*; E) *osmads3-3 osmads58*. Second whorl lodicules are indicated by the stars. The arrows indicate ectopic lodicules in B and the ectopic palea emerging from the fourth whorl in E. Bars: 100 µm (images modified from Dreni *et al.* (2011)[43]).

Moving to monocotyledonous, the main data about *AG* subfamily come from rice and maize *(Zea maize)*. In rice there are four members of the *AGAMOUS* subfamily: *OsMADS3, OsMADS58, OsMADS13* and *OsMADS21*. It is not uncommon in monocotyledonous to find two homologous genes for each *AG*-like gene from the dicotyledonous; this is due to a whole genome duplication event which took place in the monocots lineage after the divergence between monocots and dicots[35].

OsMADS3 and *OsMADS58* are direct homologous of *AGAMOUS*. This is not only true in terms of protein sequence, but also in terms of function: the double mutant *osmads3-3 osmads58* completely lacks reproductive organs, which are replaced by an indeterminate proliferation perianth organs, especially lodicules[20,41](Figure 5E). This phenotype can only be observed in the double mutant, thus proving the two genes redundantly share C function. However, observation of the single mutants also reveals a partial sub-functionalization. In the *osmads3-3* single mutant, stamens are partially or totally converted to lodicules and they are generally sterile; the carpel is also affected, showing an altered morphology and multiple stigmas on top; floral meristem determinacy is only slightly impaired (Figure 5B,C). On the contrary, *osmads58* single mutant does not show any phenotype[20]. It is interesting to notice that when *osmads58* heterozygous mutant is combined with *osmads3-3* knock-out, the phenotype of this one already gets more severe, thus furtherly proving the redundancy between the two genes[20] (Figure 5D). Also for *STK* there are in rice two homologous genes: *OsMADS13* and *OsMADS21*. In this case, the role of D-class gene is assigned to *OsMADS13* alone. In *osmads13* single mutant, floral organs are correctly formed, but inside the carpel the ovule is converted into another carpel with multiple styles and stigmas or in other less specific tissues; in both cases, *in situ* analysis detect the presence of *DROOPING LEAF* (*DL*), which is a marker of carpel identity[42,43]. Moreover, when *osmads13* is combined with *osmads3-3* or *osmads58* single mutants, the resulting flowers display a stronger indeterminacy of

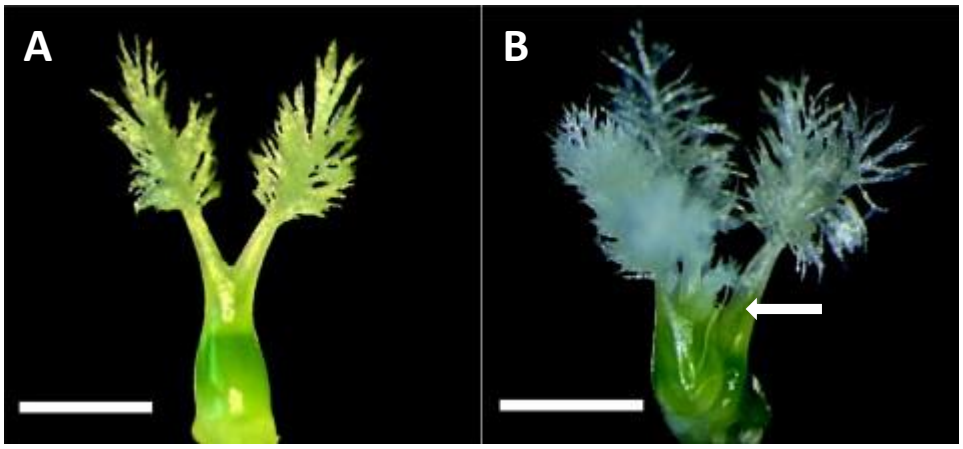


Figure 6. A) WT pistil; B) *osmads13* pistil. The arrow indicates the new carpel emerging from inside the primary carpel. Bars: 500 µm (images modified from Dreni *et al.* (2013)[88]).

the fourth whorl which is not observed in the single mutants, thus indicating a role of *OsMADS13* in floral meristem determinacy[43]. Such a redundancy between AG and AGL11 lineages has also been proven to exist in petunia; on the contrary, in Arabidopsis, *STK* has no role in meristem determinacy[29]. This difference is probably due to the different origin of the ovules in these species: Arabidopsis has a parietal placentation, while rice and petunia have a central placentation. This means that, while in Arabidopsis the floral meristem terminates in carpel formation and ovules arise from a new meristematic tissue (the placenta), in rice and petunia ovules arise directly from the floral meristem.

OsMADS21 is expressed at very low levels compared to *OsMADS13*; it is not involved in the C function and the *osmads13 osmads21* double mutant has the same phenotype as the *osmads13* single mutant, so it looks like *OsMADS21* is not even involved in D function[42]. However, when *OsMADS21* is expressed under the promoter of *OsMADS13* it is able to rescue the *osmads13* mutant, demonstrating that it retains a limited D function[43].

In *Zea mays* the same subfunctionalization among C-class genes has been observed. Maize has three genes which have been classified as *AGAMOUS* homologous: *ZMM2, ZMM23* and *ZAG1*.

ZAG1 was originally identified as the direct ortholog of *AGAMOUS*[44]. Its expression is detectable in stamen and carpel primordia, but subsequently it strongly diminishes in stamens (thus differing from *AGAMOUS*). In 1996, a null allele of *ZAG1* was generated[45]. The main phenotype was an alteration in carpel morphology and a slight loss of meristem determinacy; however, differing from the *ag* mutant in *Arabidopsis*, since stamens were almost normal and also carpelloid features were still observable in the fourth whorl[45].

An analysis of the expression profile of another *AGAMOUS*-like gene, *ZMM2*, revealed a potential redundancy with *ZAG1,* in that *ZMM2* initially shares the same expression profile of *ZAG1*, but then its signal gets stronger in the stamens. The behaviour of *ZMM2* and *ZAG1*, based on

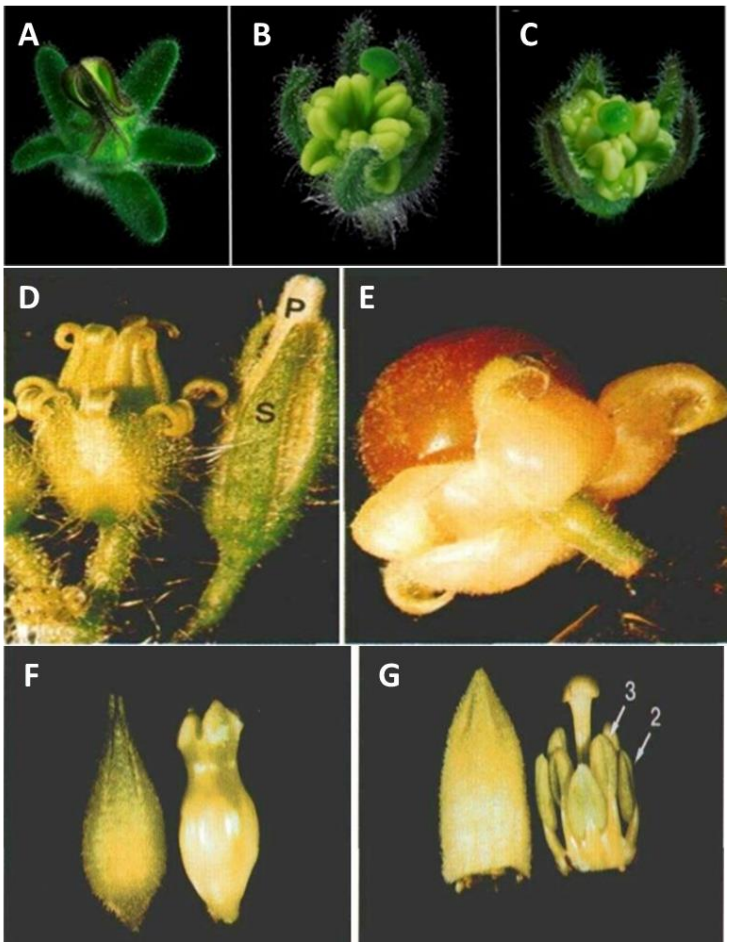


Figure 7. A-C) Petunia flower buds of WT (A), *35S::pMADS3* (B) and *35S::FBP6* (C) plants; D) WT (right) and *35S::TAG1* flowers of tomato; petals (P) and sepals (S) are indicated on WT flower; E) Transgenic tomato fruit of a *35S::TAG1* plant; F-G) Flowers of WT (left) and *35S::NAG1* (right) plants of tobacco, with sepals (F) and without sepals (G). Second and third whorl stamens are indicated in G. Images modified from Hejimans *et al.* (2012)[39], Pnueli *et al.* (1994)[62], Kempin *et al.* (1993)[63].

expression profiles and on the *zag1* mutant phenotype, resembles closely that of *OsMADS3* and *OsMADS58*, but the separation of roles looks stronger in maize than in rice[45]. Recently, the wheat gene *WAG-1* has been characterized: its expression is associated with pistil development and with pistillody (the partial conversion of stamens into pistils)[46,47].

AGAMOUS Activity

As previously said, *AGAMOUS* has three major roles in Arabidopsis flower development: establishment of stamen identity, establishment of carpel identity and floral meristem termination[23,48].

During the first stages of flower development, floral meristem has to maintain an indeterminate state; therefore, *AGAMOUS* is repressed by the AP1-SVP-AGL24 complex, which interacts with the co-repressor complex SEUSS-LEUNIG[49,50].

Around stage 3 of flower development, floral organs specification begins: B and C class genes (thus including *AGAMOUS*) are activated by AP1, SEP3, LEAFY (LFY) and WUSCHEL (WUS)[23]. Recently, trithorax factors like *ULTRAPETALA (ULT)* and *ARABIDOPSIS HORTOLOG OF TRITHORAX 1 (AHT1)* have also been found to activate *AGAMOUS* expression[23].

AGAMOUS form a complex with SEP3, AP3 and PI to promote stamen identity. Initial targets of the complex are genes like *SUPERMAN (SUP)*[51,52] and *CUP-SHAPED COTYLEDON 2 (CUC2)*[53], which are involved in setting the third whorl position and boundaries, and *UNUSUAL FLORAL ORGANS (UFO)*, which controls meristem differentiation and organ primordia development. Later on, genes that

control sporogenic tissues development, like *NOZZLE/SPOROCYTELESS (NZZ/SPL)*[54] are also activated.

Carpel development is regulated by a complex only formed by AGAMOUS and SEP3. Expression of the genes involved in the development of the different carpel tissues, like the *HECATE* genes[55], *CRABS CLAW (CRC), SPATULA (SPT)*[56] and the *SHPs*, is regulated by this complex.

Once floral organs have been specified, the floral meristem has to be terminated. Thus, AGAMOUS represses the activity of WUSCHEL, a transcription factor responsible of maintaining the stem cell niche in the flower meristem. The repression occurs both directly, by inducing histone modifications on his promoter, and indirectly, by activating the C2H2-zinc finger gene *KNUCKLES (KNU)*[57,58].

Recently, it has been reported that AGAMOUS is also able to repress the leaf developmental program, for example inhibiting trichomes formation[48], thus supporting Goethe's idea that flowers are "modified leaves".

Ectopic Expression Of AGAMOUS Genes

Apart from the generation of mutant lines, another common experiment to test the functionality of a gene is to check the effects of its ectopic expression, typically under the control of a constitutive promoter, like the CaMV 35S promoter[59]. This is particularly efficacious to the homeotic ABC floral identity genes. Thanks to the combinatorial nature of the ABC model and to the almost perfect spatial separation of the different roles, allows predicting the outcome of the ectopic expression. There are several examples of these kinds of experiments using *AGAMOUS* subfamily genes, since it is an efficient way to test whether a newly discovered, putative *AG* gene, has C-function activity or not. The phenotype one may expect from the ectopic expression of a C-class gene is the exact opposite of a *c* mutant, meaning conversion of perianth organs into reproductive organs; in particular, based on the ABC model, one may expect a conversion of petals (or second whorl organs) into stamens or staminoid organs in the second whorl and of sepals into carpels or carpelloid organs in the first whorl. This is exactly what happens when a *35S::AG* is introduced in a WT plant of Arabidopsis, as shown by Mizukami *et al.* (1992)[60]. The same kind of experiment was also used to demonstrate that the *SHP* genes, despite not being involved in the C function because of their spatial-temporal expression, retain the ability of performing this function when ectopically expressed in flowers[29].

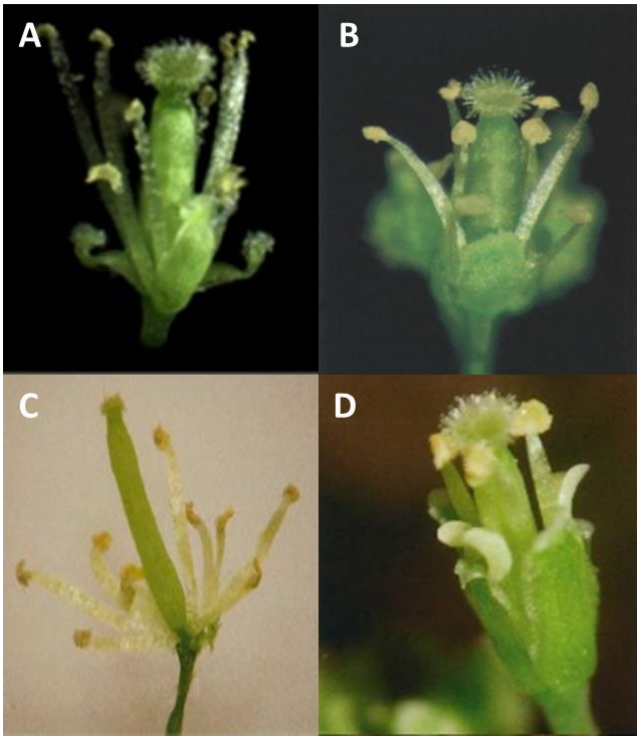


Figure 8. Arabidopsis flowers of plants expressing *35S::FBP6* (A), *35S::CaMADS1* (B), *35S::MASAKOC1* (C) and *35S::MASAKOD1* (D), all displaying homeotic conversion of petals into stamens. Images modified from Heijmans *et al.* (2012)[39], Rigola *et al.* (2000)[64], Kitahara *et al.* (2004)[65].

Ectopic expression of *AGAMOUS*-like genes has also been performed in other species apart from Arabidopsis. In transgenic plants of *Petunia hybrida* transformed with *35S::pMADS3* (Figure 7B) or *35S::FBP6* (Figure 7C) the corolla limbs were converted into anteroid tissues and the sepals morphology was altered[39,61]. In tomato, *TAG1* was identified as a putative orthologous of *AGAMOUS*; the introduction of *35S::TAG1* in transgenic tomato plants caused sterile stamens to appear in place of petals in the second whorl (Figure 7D) and sepals to convert into fleshy pericarp leaves (no other carpelloid features were observed)[62] (Figure 7E). Something similar happens when the tobacco (*Nicotiana tabacum*) gene *NAG1* is ectopically expressed in transgenic tobacco plants: petals and sepals acquire staminoid and carpelloid features, respectively (the number of organs is generally maintained)[63] (Figure 7F,G).

Thanks to the simple "floral dip" transformation protocol[1], *A. thaliana* is also the perfect model organism to perform functional studies on heterologous genes. Since the development of this method, many experiments have been performed where an *AGAMOUS*-like gene from another plant species has been introduced in Arabidopsis under the control of the 35S promoter, to test if it was able to phenocopy the *35S::AGAMOUS* overexpression as described by Mizukami *et al.* (1992). In most of the cases, the experiments produced very similar results: the observed phenotypes ranged from shorter petals and sepals in plants with a mild expression to petal-to-stamen conversion and presence of carpelloid features (cell shape, stigmatic tissue, sometimes ovules) on sepals in the first whorl. This happened, for example, with the ectopic expression of *FBP6* from petunia (*Petunia hybrida)*[39] (Figure 8A), or *CaMADS1* from hazelnut (*Corylus avellana*)[64] (Figure 8B)and with *MASAKO C1* and *MASAKO D1* from wild rose (*Rosa rugosa*)[65] (Figure 8C,D). The expression of *PpAG1* from peach (*Prunus persica*) generated mainly carpelloid sepals, while the petals were only mildly converted or just disappeared from the flower[66]; the same phenotype

can be observed with the ectopic expression of the the Japanese gentian *(Gentiana scabra) PLE* gene *GsAG1*, whose mutant has been associated with the so called "double flower" phenotype (which is actually the common *agamous* phenotype).

It is interesting to discover that similar effects can be obtained when the heterologous gene comes from a gymnosperm (a non-flowering plant), considering that the last common ancestor of angiosperms and gymnosperms dates back to 285-350 mya[67,68]. When the *SAG1* gene from the conifer black spruce (*Pecia mariana*) is expressed in *Arabidopsis*, the result is, once again, a range of phenotypes going from reduction of the outer organ length to their conversion into reproductive-organ-like structures[69]. Even more surprisingly, the ectopic expression of the *Cycas edentata* gene *CyAG* is able to completely rescue the *ag-2* mutant phenotype in Arabidopsis[70].

Role In Fruit Development

MADS-box genes belonging to the *AGAMOUS* subfamily have been extensively studied in the past 30 years mainly for their roles in flower development. However, many recent publications show that these genes also have an active role after fertilization, during the development of seeds and fruits.

In *A. thaliana*, *STK, SHP1* and *SHP2* redundantly control ovule identity and development[29,30]. However, these three genes were initially characterized for their roles in fruit development, especially in the correct formation of the dehiscence zones. The *stk* mutant has smaller seeds but longer funiculi and the seeds do not detach from the silique due to defects in dehiscence zone formation[29]. *SHP1* and *SHP2* are involved in the lignification of the valve margins of the silique, a process that allows the opening of the silique and the seed dispersal; in the *shp1 shp2* double mutant the lignification does not occur and the fruit is not able to break[36]. *SHP* genes mainly define valve margin identity by regulating the expression of three bHLH genes: *INDEHISCENT (IND), ALCATRAZ (ALC)* and *SPATULA (SPT)*[71–73]. Recently, it has been demonstrated that *STK* perform a similar action in seed abscission zone by regulating *ALC*, *SPT* and the *IND*-like gene *HECATE3 (HEC3)*; these are however repressed by *STK* and not activated (as it happens with the *SHPs*) due to *STK* interaction with *SEUSS*[74].

Recently, the interaction between *STK* and *ABS* during seed development has been proven: the *stk abs* double mutant displays a variety of defects in seeds, including high rate of abortion and developmental arrest, lack of the endothelium and alterations in the morphology of the other

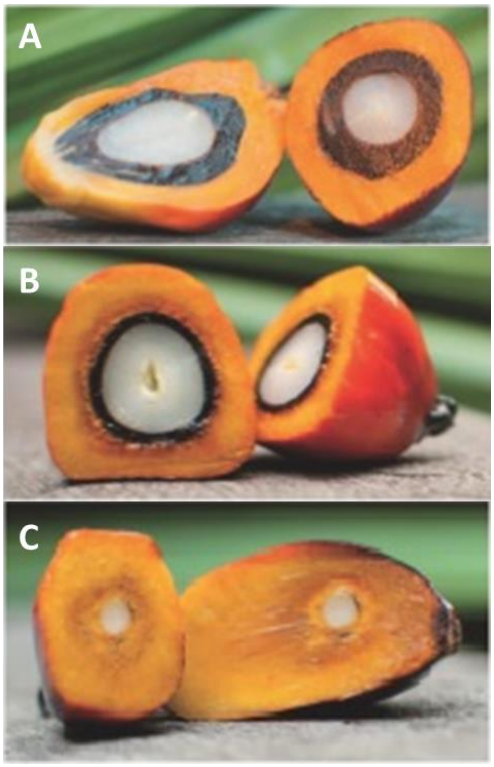


Figure 9. Oil palm coconuts of *dura* (A), *tenera* (B) and *pisifera* (C) varieties. Images modified from Singh *et al*. (2013)[77].

integument layers, massive accumulation of starch in the embryo sac which is not consumed during embryo development and defects in sperm release[75]. None of these defects has been observed in the *stk* or *abs* single mutants.

Apart from Arabidopsis, many *AG*-related genes have been assigned a role in seed development. In petunia, the D-class genes *FBP7* and *FBP11* expression in still high in the ovary after pollination, then it decreases when the fruit is mature; their expression is mainly localized in the seed coat[27]. In the double mutant *fbp7 fbp11* the seeds appear shrunken compared to WT seeds; a more detailed analysis revealed a partial or total degeneration of the endosperm and, quite often, the arrest of embryo development at various stages[27].

In *Nicotiana benthamiana*, the VIGS silencing of *NbSHP*, apart from flower defects, also causes fruit indehiscence due to a lack of lignification on top of the fruit, similarly to what happens in Arabidopsis[40]. Ectopic expression of the soybean (*Glycine maxima*) *PLE*-like gene *GmAGL1* in Arabidopsis causes petal disappearance and early seed release due to alterations in the valve dehiscence zone[76].

In oil palm (*Elaeis guineensis*), three fruit forms exist which differ in the thickness of the ligneous, coconut-like shell that surrounds the kernel: the *dura* form has a very thick shell (Figure 9A), which is absent in the *pisifera* form (Figure 9C); the *tenera* form, a hybrid between the other two forms, has a thinner shell (Figure 9B) and produces much more oil than the *dura* form. Singh *et al.* (2013) demonstrated a link between these features and the segregation of *SHELL*, an ortholog of *STK*: the mutant *shell* allele causes a reduction in shell lignification, leading to the *tenera* (*SH/sh*) and *pisifera* (*sh/sh*) phenotypes[77].

In peach, the *SHP*-like gene *PPERSHP* is involved in the pit hardening process and is hold responsible for the so called split-pit (the disruption of the pit along the side suture line), which is a common problem in stone fruits[78]. Moreover, the *PLE*-like gene *PpPLE*, when ectopically expressed in tomato, is able to induce the conversion of sepals into fleshy, reddish fruit-like

organs[79]. A similar phenotype has been reported for the *Gingko biloba AG*-like gene *GBM5*, which is expressed in the fleshy fruit-like structure of this plant[80,81].

In tomato, mutations in the *TAGL1* gene cause defects in fruit ripening. In *TAGL1-RNAi* lines, fruits have a darker shade of green compared to WT at the beginning of maturation, then at a stage that they are fully ripened they appear orange or, in general, less red than the WT fruit; this has been associated with significant variations in carotenoid and β-chlorophyll synthesis[82].

Ectopic expression in tomato of the *Musa acuminata* gene *MaMADS7*, an *AG*-like gene belonging to the *AGL11* subclade, induces precocious fruit ripening; the gene has been associated with ethylene biosynthesis and promotion of ripening[83].

Conservation Of Protein Domains

As previously mentioned, type II MADS-box proteins, which represents the vast majority of MADS-box TFs involved in plant development and reproduction, are also called MIKC from the four conserved protein domains they have: the MADS (M), the intervening (I), the keratin-like (K) and the C-terminal (C) domains. The structure of these four domains has been highly conserved during evolution because they are fundamental for MADS-box TFs functionality, in particular for DNA binding specificity and for protein-protein interactions.

Airoldi and colleagues (2010) nicely demonstrated how small changes inside one of these domains can strongly alter protein function. In snapdragon, the C-function has been assigned to *PLENA*, while the *euAG* gene *FARINELLI* is able to specify stamen, but not carpel, identity, as demonstrated by their ectopic expression in Arabidopsis[84]. Alignment of the two protein sequences with that of AGAMOUS reveal that a glutamine residue (Q173) is present in the 3rd α-helix of the K domain (K3) but is absent in AGAMOUS and PLENA. Ectopic expression of a mutated variant of FAR lacking Q173 in Arabidopsis produced a phenotype similar to PLE expression; conversely, the *35S::FAR* phenotype could be reproduced with a mutated AG where a glutamine was artificially inserted in the same position as that of FAR[84]. Interestingly, the position of the insertion seems to be more important than the insertion itself: as a matter of fact, the same phenotype could be obtained with the insertion of an arginine, instead of a glutamine, at the same position in the AG protein sequence[84]. This happens because the insertion falls inside a region which is known to be involved in protein complexes formation: yeast 2-hybrid assays demonstrated that while AGAMOUS can

interact with SEP1/2/3, FAR can interact only with SEP3, thus limiting its spatial activity to the three inner domains and explaining the lack of sepal conversion[84].

Another example of the importance of MADS domain conservation comes from the oil palm gene *SHELL*. An extensive analysis on many varieties carrying the *dura* and *pisifera* phenotypes (described above) revealed the existence of two versions of the mutant allele *shell*, each carrying an aminoacidic substitution in the MADS domain: a lysine-to-arginine in one case, a leucine-to-proline in the other, occurring at a distance of 3 aminoacids one from each other[77]. The first substitution removes a lysine which is directly involved in DNA binding[85]; the second, introducing a proline, causes a disruption in an α-helix involved in DNA binding and MADS dimerization[86]. As a matter of fact, a yeast hybrid assay demonstrated that this second mutant variety, opposite to the WT form of SHELL, is not able to interact *in vitro* with OsMADS24 (a rice ortholog of SEP3)[77].

Recently, the dimeric and tetrameric crystal structure of SEPALLATA3 K-domain have been reported[87]. In the paper it is also reported how single amino acid changes in the α-helices of SEP3 K-domain severely affect tetramer formation in favour of dimer formation *in vitro* and how the removal of part of the second helix almost eliminates tetramer formation[87].

AIM OF THE PROJECT

Alternative Splicing In Rice OsMADS3

Investigation in online databases and available RNA-Seq data revealed the existence of an alternative splicing in the rice *AGAMOUS* ortholog *OsMADS3* which gives rise to two protein isoforms only differing one from each other for the presence or absence of a serine residue (S109). The isoform bearing S109 (*OsMADS3*^{S109}) is the one that has been previously characterized for its functions in flower development; on the contrary, nothing is known about the isoform lacking S109 (*OsMADS3*$^{\Delta S109}$). Given the conservation of this particular serine in the *OsMADS3* clade in graminaceae and its position inside the highly conserved K-domain, we hypothesized its presence or absence could be relevant for the protein functionality. Therefore, we expressed both isoforms in *A. thaliana* under the control of the CaMV 35S promoter to evaluate the ability of each form to induce ectopic changes in floral organs and to rescue the *ag-3* mutant phenotype.

Role Of Rice AGAMOUS Genes In Seed Development

As described by Arora *et al.* (2007), all four members of the rice *AGAMOUS* subfamily continue to be expressed after fertilization, during all stages of seed development. In other species *AGAMOUS* family genes have been reported to be involved in a variety of post-fertilization processes, including seed integuments and vasculature development, fruit ripening, nutrient accumulation and seed abscission. The available data are, however, mainly restricted to the traditional model species, like Arabidopsis, petunia, tomato and tobacco; little is known about what happens in more economically relevant species, like cereals.

We hypothesized rice *AGAMOUS* genes could have a similar role during seed development. In order to prove our hypothesis, multiple mutants in *OsMADS3, OsMADS58, OsMADS13* and *OsMADS21* have been generated and the effects of the multiple mutations on fertility and seed development were evaluated.

MAIN RESULTS

Different Functionality Of OsMADS3 Isoforms

Ectopic expression of *OsMADS3*^{S109} and *OsMADS3*$^{\Delta S109}$ in Arabidopsis reveals that the two isoforms actually behave differently. *OsMADS3*$^{\Delta S109}$ produces on average stronger homeotic changes on perianth organs than *OsMADS3*^{S109}, closely resembling those produced by *AGAMOUS* ectopic expression; moreover, *OsMADS3*$^{\Delta S109}$ is also able to rescue floral meristem indeterminacy in the *ag-3* mutant background (even if plants remain sterile), while *OsMADS3*^{S109} does not show this ability.

Yeast 2-hybrid analysis also revealed different interaction ability of the two isoforms. Surprisingly, OsMADS3^{S109} appears to be able to interact with SEP1 and SEP3, while OsMADS3$^{\Delta S109}$ isn't; as a confirmation, OsMADS3$^{\Delta S109}$ also was not able to interact with OsMADS8, a rice ortholog of SEP3 which is the usual partner of OsMADS3.

See Chapter 2 for a more detailed report of the results mentioned above.

Mutations In Rice AGAMOUS Genes Strongly Affect Seed Development And Fertility.

In our growth conditions, WT rice plants have a fertility of almost 100% and the average seed weight is around 190 mg.

Given the fact that *OsMADS3* and *OsMADS13* have to be kept in a heterozygous state in order for the plant to be able to produce seeds, mutations in one or two of the four *AGAMOUS* family members in rice do not apparently affect seed development or fertility. When mutations are combined in triple or even quadruple mutant lines, effects on seed yield get more drastic: in the most severe genotype, the quadruple mutant *osmads3-3/+ osmads13/+ osmads21 osmads58*, both seed weight and fertility are reduced by almost 50%.

An analysis on aborted seeds in triple and quadruple mutants reveals that in almost 50% of the cases the fertilization had occurred and seed development had initiated. However, the process was arrested at intermediate stages, thus suggesting that *AG* genes have an active role in guaranteeing correct seed development.

Further analyses are being currently carried out which are revealing defects in seed coat and vasculature of the mutant seeds, thus indicating an incorrect intake of nutrients and defective communication between the inner and outer environment as main reasons for the seed impaired development.

See Chapter 3 for a more detailed report of the results mentioned above.

BIBLIOGRAPHY

1. Clough, S. J. & Bent, A. F. Floral dip: A simplified method for Agrobacterium-mediated transformation of *A. thaliana*. *Plant J.* **16,** 735–743 (1998).
2. Arabidopsis, T. & Initiative, G. Analysis of the genome sequence of the flowering plant *A. thaliana*. **408,** (2000).
3. Yu, J. *et al.* A draft sequence of the rice (Oryza sativa ssp indica) genome. *Chin. Sci. Bull.* **46,** 1937–1942 (2001).
4. Goff, S. A. *et al.* A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science (80-.).* **296,** 92–100 (2002).
5. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (Oryza sativa L.)

mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6,** 271–82 (1994).

6. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early Flower Development in Arabídopsis. *Plant Cell* **2,** 755–767 (1990).
7. Whipple, C. J., Zanis, M. J., Kellogg, E. a & Schmidt, R. J. Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1081–6 (2007).
8. Yanofsky, M. F. *et al.* The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **346,** 35–39 (1990).
9. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H. Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. *Science (80-.).* **250,** 931–936 (1990).
10. Norman, C., Runswick, M., Pollock, R. & Treisman, R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* **58,** 1173–1181 (1989).
11. Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B. K. Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *J. Mol. Biol.* **204,** 593–606 (1988).
12. Alvarez-Buylla, E. R. *et al.* An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5328–33 (2000).
13. Masiero, S., Colombo, L., Grini, P. E., Schnittger, A. & Kater, M. M. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **23,** 865–72 (2011).
14. Smaczniak, C., Immink, R. G. H., Angenent, G. C. & Kaufmann, K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139,** 3081–98 (2012).
15. Arora, R. *et al.* MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* **8,** 242 (2007).
16. Schwarz-Sommer, Z. *et al.* Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* **11,** 251–63 (1992).

17. Theissen, G. & Saedler, H. Floral quartets. *Nature* **409,** 469–471 (2001).
18. Becker, A. & Theissen, G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **29,** 464–489 (2003).
19. Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353,** 31–37 (1991).
20. Kater, M. M., Dreni, L. & Colombo, L. Functional conservation of MADS-box factors controlling floral organ identity in rice and Arabidopsis. *J. Exp. Bot.* **57,** 3433–44 (2006).
21. Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. & Smyth, D. R. Control of flower development in *A. thaliana* by APETALA1 and interacting genes. *Development* **119,** 721–743 (1993).
22. Ferrándiz, C., Gu, Q., Martienssen, R. & Yanofsky, M. F. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* **127,** 725–734 (2000).
23. Wils, C. R. & Kaufmann, K. Gene-regulatory networks controlling inflorescence and flower development in *A. thaliana*. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860,** 95–105 (2017).
24. Bowman, J. L., Smyth, D. R. & Meyerowitzt, E. M. Genetic interactions among floral homeotic genes of Arabidopsis. *Development* **20,** 1–20 (1991).
25. Winter, K. *et al.* Evolution of Class B Floral Homeotic Proteins : Obligate Heterodimerization Originated from Homodimerization. 587–596 (2000).
26. Colombo, L. *et al.* The Petunia MADS Box Gene FBP11 Determines Ovule Identity. *Plant Cell* **7,** 1859–1868 (1995).
27. Colombo, L. *et al.* Downregulation of Ovule-Specific MADS Box Genes from Petunia Results in Maternally Controlled Defects in Seed Development. *Plant Cell* **9,** 703–715 (1997).
28. Angenent, G. C. *et al.* A Novel Class of MADS Box Genes is lnvolved in Ovule Development in Petunia. *Plant Cell* **7,** 1569–1582 (1995).
29. Pinyopich, A. *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424,** 85–88 (2003).
30. Favaro, R. *et al.* MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* **15,** 2603–2611 (2003).
31. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405,** 200–203 (2000).

32. Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. & Yanofsky, M. F. The SEP4 Gene of *A. thaliana* Functions in Floral Organ and Meristem Identity. *Curr. Biol.* **14,** 1935–1940 (2004).
33. Causier, B., Schwarz-Sommer, Z. & Davies, B. Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* **21,** 73–9 (2010).
34. Dreni, L. & Kater, M. M. MADS reloaded: Evolution of the AGAMOUS subfamily genes. *New Phytol.* **201,** 717–732 (2014).
35. Kramer, E. M., Jaramillo, M. A. & Di Stilio, V. S. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166,** 1011–23 (2004).
36. Liljegren, S. J. *et al.* SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* **404,** 766–770 (2000).
37. Brambilla, V. *et al.* Genetic and Molecular Interactions between BELL1 and MADS Box Factors Support Ovule Development in Arabidopsis. *Plant Cell Online* **19,** 2544–2556 (2007).
38. Davies, B. *et al.* PLENA and FARINELLI: redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. *EMBO J.* **18,** 4023–34 (1999).
39. Heijmans, K. *et al.* Redefining C and D in the petunia ABC. *Plant Cell* **24,** 2305–17 (2012).
40. Fourquin, C. & Ferrándiz, C. Functional analyses of AGAMOUS family members in Nicotiana benthamiana clarify the evolution of early and late roles of C-function genes in eudicots. *Plant J.* **71,** 990–1001 (2012).
41. Yamaguchi, T. *et al.* Functional diversification of the two C-class MADS box genes OSMADS3 and OSMADS58 in Oryza sativa. *Plant Cell* **18,** 15–28 (2006).
42. Dreni, L. *et al.* The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. *Plant J* **52,** 690–699 (2007).
43. Dreni, L. *et al.* Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. *Plant Cell* **23,** 2850–63 (2011).
44. Schmidt, R. J. *et al.* Identification and molecular characterization of ZAG1, the maize homolog of the Arabidopsis floral homeotic gene AGAMOUS. *Plant Cell* **5,** 729–737 (1993).
45. Mena, M. *et al.* Diversification of C-function activity in maize flower development. *Science (80-.).* **274,** 1537–1540 (1996).
46. Meguro, A., Takumi, S., Ogihara, Y. & Murai, K. WAG, a wheat AGAMOUS homolog, is

associated wiht development of pistillike stamens in alloplasmic wheats. *Sex. Plant Reprod.* **15,** 221–230 (2003).

47. Zhang, L., Yang, Z., Peng, Z., Yu, Y. & Sun, Q. Molecular Cloning, Characterization, and Expression Analysis of WAG-1 gene in the Pistillody Line of Common Wheat. *Mod. Appl. Sci.* **9,** 44 (2015).
48. O'Maoileidigh, D. S. *et al.* Control of Reproductive Floral Organ Identity Specification in Arabidopsis by the C Function Regulator AGAMOUS. *Plant Cell* **25,** 2482–2503 (2013).
49. Gregis, V., Sessa, A., Colombo, L. & Kater, M. M. AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. *Plant J.* **56,** 891–902 (2008).
50. Sridhar, V. V., Surendrarao, A., Gonzalez, D., Conlan, R. S. & Liu, Z. Transcriptional repression of target genes by LEUNIG and SEUSS, two interacting regulatory proteins for Arabidopsis flower development. *Proc. Natl. Acad. Sci.* **101,** 11494–11499 (2004).
51. Bowman, J. L. *et al.* SUPERMAN, a regulator of floral homeotic genes in Arabidopsis. *Development* **114,** 599–615 (1992).
52. Sakai, H., Medrano, L. J. & Meyerowitz, E. M. Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries. *Nature* **378,** 199–203 (1995).
53. Kamiuchi, Y., Yamamoto, K., Furutani, M., Tasaka, M. & Aida, M. The CUC1 and CUC2 genes promote carpel margin meristem formation during Arabidopsis gynoecium development. *Front. Plant Sci.* **5,** (2014).
54. Liu, X. *et al.* The SPOROCYTELESS/NOZZLE Gene Is Involved in Controlling Stamen Identity in Arabidopsis. *PLANT Physiol.* **151,** 1401–1411 (2009).
55. Schuster, C., Gaillochet, C. & Lohmann, J. U. Arabidopsis HECATE genes function in phytohormone control during gynoecium development. *Development* **142,** 3343–3350 (2015).
56. Alvarez, J. & Smyth, D. R. CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126,** 2377–2386 (1999).
57. Liu, X. *et al.* AGAMOUS Terminates Floral Stem Cell Maintenance in Arabidopsis by Directly Repressing WUSCHEL through Recruitment of Polycomb Group Proteins. *Plant Cell* **23,** 3654–3670 (2011).
58. Sun, B. *et al.* Timing Mechanism Dependent on Cell Division Is Invoked by Polycomb Eviction in Plant Stem Cells. *Science (80-.).* **343,** 1248559–1248559 (2014).
59. Odell, J. T., Nagy, F. & Chua, N.-H. Identification of DNA sequences required for activity of

the cauliflower mosaic virus 35S promoter. *Nature* **313,** 810–812 (1985).

60. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell* **71,** 119–131 (1992).
61. Tsuchimoto, S., Krol, A. R. Van Der & Chua, N. Ectopic Expression of pMADS3 in Transgenic Petunia Phenocopies the Petunia blind Mutant. **5,** 843–853 (1993).
62. Pnueli, L., Hareven, D., Rounsley, S. D., Yanofsky, M. F. & Lifschitz, E. Isolation of the Tomato AGAMOUS Gene TAG1 and Analysis of Its Homeotic Role in Transgenic Plants. *Plant Cell Online* **6,** 163–173 (1994).
63. Kempin, S. A., Mandel, M. A. & Yanofsky, M. F. Conversion of Perianth into Reproductive Organs by Ectopic Expression of the Tobacco Floral Homeotic Gene NAG1. *Plant Physiol* **103,** 1041–1046 (1993).
64. Rigola, D.; Pè, M. E.; Mizzi, L; Ciampolini, F.; Sari-Gorla, M. CaMADS1 , an AGAMOUS homologue from hazelnut , produces floral homeotic conversion when expressed in Arabidopsis. **2,** 185–191 (2001).
65. Kitahara, K., Hibino, Y., Aida, R. & Matsumoto, S. Ectopic expression of the rose AGAMOUS-like MADS-box genes 'MASAKO C1 and D1' causes similar homeotic transformation of sepal and petal in Arabidopsis and sepal in Torenia. *Plant Sci.* **166,** 1245–1252 (2004).
66. Martin, T. *et al.* PpAG1, a homolog of AGAMOUS, Expressed in Developing Peach Flowers and Fruit. *Can. J. Bot.* **84,** 767–776 (2006).
67. Savard, L. *et al.* Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5163–5167 (1994).
68. Munster, T. *et al.* Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Evolution (N. Y).* **94,** 2415–2420 (1997).
69. Rutledge, R. *et al.* Characterization of an AGAMOUS homologue from the conifer black spruce (Picea mariana) that produces floral homeotic conversions when expressed in Arabidopsis. *Plant J.* **15,** 625–634 (1998).
70. Zhang, P., Tan, H. T. W., Pwee, K. H. & Kumar, P. P. Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* **37,** 566–577 (2004).
71. Liljegren, S. J. *et al.* Control of fruit patterning in Arabidopsis by INDEHISCENT. *Cell* **116,**

843–853 (2004).

72. Girin, T. *et al.* INDEHISCENT and SPATULA Interact to Specify Carpel and Valve Margin Tissue and Thus Promote Seed Dispersal in Arabidopsis. *Plant Cell* **23,** 3641–3653 (2011).
73. Groszmann, M., Paicu, T., Alvarez, J. P., Swain, S. M. & Smyth, D. R. SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. *Plant J.* **68,** 816–829 (2011).
74. Balanzà, V., Roig-Villanova, I., Di Marzo, M., Masiero, S. & Colombo, L. Seed abscission and fruit dehiscence required for seed dispersal rely on similar genetic networks. *Development* **143,** 3372–3381 (2016).
75. Mizzotti, C. *et al.* The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J.* **70,** 409–20 (2012).
76. Chi, Y. *et al.* GmAGL1, a MADS-Box Gene from Soybean, Is Involved in Floral Organ Identity and Fruit Dehiscence. *Front. Plant Sci.* **8,** 1–11 (2017).
77. Singh, R. *et al.* The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500,** 340–4 (2013).
78. Tani, E., Polidoros, A. N. & Tsaftaris, A. S. Characterization and expression analysis of FRUITFULL- and SHATTERPROOF-like genes from peach (Prunus persica) and their role in split-pit formation. *Tree Physiol.* **27,** 649–659 (2007).
79. Tadiello, A. *et al.* A PLENA-like gene of peach is involved in carpel formation and subsequent transformation into a fleshy fruit. *J. Exp. Bot.* **60,** 651–661 (2009).
80. Jager, M., Hassanin, A., Manuel, M., Le Guyader, H. & Deutsch, J. MADS-box genes in Ginkgo biloba and the evolution of the AGAMOUS family. *Mol. Biol. Evol.* **20,** 842–854 (2003).
81. Lovisetto, A., Baldan, B., Pavanello, A. & Casadoro, G. Characterization of an AGAMOUS gene expressed throughout development of the fleshy fruit-like structure produced by Ginkgo biloba around its seeds. *BMC Evol. Biol.* **15,** 139 (2015).
82. Pan, I. L., McQuinn, R., Giovannoni, J. J. & Irish, V. F. Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development. *J. Exp. Bot.* **61,** 1795–1806 (2010).
83. Liu, J. *et al.* Role for the banana AGAMOUS-like gene MaMADS7 in regulation of fruit ripening and quality. *Physiol. Plant.* **155,** 217–231 (2015).
84. Airoldi, C. a, Bergonzi, S. & Davies, B. Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18898–902 (2010).

85. Huang, H. DNA Binding Properties of Two Arabidopsis MADS Domain Proteins: Binding Consensus and Dimer Formation. *Plant Cell Online* **8,** 81–94 (1996).

86. Immink, R. G. H., Kaufmann, K. & Angenent, G. C. The 'ABC' of MADS domain protein behaviour and interactions. *Semin. Cell Dev. Biol.* **21,** 87–93 (2010).

87. Puranik, S. *et al.* Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26,** 3603–3615 (2014).

88. Dreni, L., Osnato, M. & Kater, M. M. The Ins and Outs of the Rice AGAMOUS Subfamily. *Mol. Plant* **6,** 650–664 (2013).

PART 2

CHAPTER 2: ALTERNATIVE SPLICING OF A SINGLE AMINOACID IN RICE *OSMADS3* ALTERS PROTEIN FUNCTIONALITY

Andrea Ravasio, Ludovico Dreni, Nahuel Gonzalez-Schain, Veronica Gregis, Martin M. Kater
(Manuscript in preparation)

ABSTRACT

Within the MADS-box gene family the *AGAMOUS*-subfamily genes are particularly important for plant reproduction because they are directly involved in flower development, specifically in establishing the identity of reproductive organs and in controlling their overall development.
A number of studies in the last 30 years have demonstrated how *AGAMOUS* genes have been conserved during land plants evolution, since their function has been retained across very different species. However, duplication events and alternative splicing mechanisms have led to subfunctionalization and neofunctionalization among multiple *AGAMOUS*-like genes in many species.
Here we show how alternative splicing produces in rice two types of transcripts of the *AGAMOUS* ortholog *OsMADS3* which differ one from the other in only one serine residue. This alternative splicing appears to be retained not only in rice, but in the whole Poales group.
Expression in *A. thaliana* of the two isoforms revealed that the serine-lacking isoform OsMADS3$^{\Delta S109}$ has a stronger ability to induce homeotic changes in floral organs compared to the serine-carrying isoform OsMADS3^{S109}. Moreover, OsMADS3$^{\Delta S109}$ is partially able to complement the *ag-3* mutant phenotype of Arabidopsis. Yeast-2-hybrid assays also demonstrate that the two isoforms interact differently with SEPs proteins from both Arabidopsis and rice.

INTRODUCTION

The evolution of the flower as new reproductive structure is the most remarkable specific feature of angiosperms (flowering plants), which currently are by far the most dominating group of plants on Earth, counting over 350,000 species (http://www.theplantlist.org/). In its basic model, the flower is a bisexual structure composed of perianth organs surrounding the reproductive organs, the male stamens and the female gynoecium, which are located in the center of the flower. All

organs are arranged either in a spiral phyllotaxy or in adjacent whorls, depending on different angiosperm lineages. The gynoecium is composed of a variable number of pistils, and each pistil can be formed by one carpel or more carpels fused together. It differentiates into a basal swollen part, which contains the ovule(s) and is called ovary; on top of this structure the style(s) and stigma(s) develop. The pistil upon pollination eventually becomes a fruit, and the ovules within develop into seeds[1].

The two most evolved and successful angiosperm *taxa* are referred to as eudicots and monocots, comprising *ca*. 75% and 20% of the living species, respectively. It is estimated that divergence between eudicot and monocot species occurred about 150 million years ago (MYA)[2]. Plants belonging to the two lineages evolved inflorescences and flowers having very different complexity and shapes, probably driven by different selective pressure. This can be clearly seen when we compare the morphologies of two model species representing each lineage, the core eudicotyledon *A. thaliana* (L.) Heynh. and the monocotyledon *Oryza sativa* L. However, if we compare the basic floral structure of these two species, a conserved organization emerges, in which four different types of floral organs are arranged into concentric whorls. The outermost whorls bear so-called perianth organs: (i) leaf-like organs represented by four sepals in Arabidopsis and lemma and palea in rice (even if the majority of literature does not consider lemma as a first whorl organ[3]), and (ii) in a more inner whorl the petals or lodicules in Arabidopsis and rice, respectively. The male and female reproductive organs are located in the most inner two whorls. Both Arabidopsis and rice flowers develop six stamens; in Arabidopsis a bicarpellate pistil develops in the most inner whorl, whereas in rice the pistil is tricarpellate[4,5].

AGAMOUS (AG), a member of the MADS-box TFs family[6], is essential for stamen and carpel development in Arabidopsis. The *ag* loss-of-function mutant completely lacks reproductive organs; petals replace stamens in the third whorl and a new *ag* flower replaces the carpel in the fourth whorl, resulting in a complete loss of floral meristem determinacy (FMD)[7]. AGAMOUS acquires its reproductive organ identity function through complex formation with SEPALLATA (SEP) MADS-domain proteins, in particular SEP3[8]. Complexes formed by only AG and SEP proteins direct carpel development, whereas in the third whorl SEPs mediate the interaction between AG, APETALA3 (AP3) and PISTILLATA (PI) to form the MADS stamen identity complex[9,10].

The *AGAMOUS* subfamily is typically represented by two or more genes in each angiosperm species having various degrees of functional redundancy and sub-functionalization, and is further divided into many paraphyletic lineages. The main division which is important to mention here is

the one between the *AG* and *AGL11* lineages[11] (also known as C and D lineages[12]), which diverged before the most recent common ancestor of extant angiosperms. All members of the *AGAMOUS* subfamily cluster into one of these two groups. Generally speaking, the genes providing the carpel and stamen identity function belong to the *AG* lineage, whereas the *AGL11*-like genes provide ovule identity[6,11,12]. However, despite their ancestral origin, functional redundancy between the two clades is common. For example, all the *AG*- and *AGL11*-like genes of Arabidopsis and petunia (*Petunia x hybrida* hort. ex E.Vilm.) participate in the ovule identity function[13–15]; in rice, the *AGL11* lineage gene *OsMADS13* supports the *AG* lineage genes in FMD, besides regulating ovule identity[16,17].

Besides *AGAMOUS*, two other members of the *AG* lineage are present in Arabidopsis, *SHATTERPROOF1/2 (SHP1/2)*. Despite they have been proven to retain a potential reproductive organ identity function, they do not cooperate in the C-function with *AGAMOUS* due to their different spatial-temporal expression and regulation[13,18].

Interestingly, in grasses, the *AGAMOUS* and *AGL11* lineages are further duplicated to form four conserved clades which are named after their representative genes of rice. Among them, the *OsMADS3* and *OsMADS58* clades belong to the *AGAMOUS* lineage whereas the *OsMADS13* and *OsMADS21* clades belong to the *AGL11* lineage. Therefore, the rice genes *OsMADS3* and *OsMADS58* are both direct orthologous of *AG* with a similar expression domain, and they redundantly control the reproductive organ identity function in rice. Like the Arabidopsis *ag* mutant, the *osmads3 osmads58* double mutant shows a complete loss of reproductive organ identity; stamens are replaced by lodicules and an ectopic palea emerges in place of the carpel. Moreover, FMD is also lost and the flower continues to produce ectopic lodicules from its center[17].

The phenotypes of the single mutants suggest that *OsMADS3* has an important role in stamen development, whereas the two genes seem to be equally important for pistil development. In the *osmads3-3* knock-out mutant, the most evident phenotype is a variable stamen-to-lodicule conversion, while carpel identity and FMD are only mildly affected; this is even more evident in the *osmads3-4* mild allele, where FMD is retained and most of the pistils develop normally, while there's a slight stamen-to-lodicule conversion[17,19]. In the *osmads58* mutant, instead, the flowers develop normally, with only 5% of the flowers showing supernumerary stigmas[17].

Like the eudicot AG protein, rice OsMADS3 and OsMADS58 form complexes with SEP-like factors. In rice there are five members of the *SEP* subfamily: *OsMADS1, OsMADS5, OsMADS24* (allelic to *OsMADS8*), *OsMADS34* and *OsMADS45* (allelic to *OsMADS7*)[20].

In the ancient tetraploid maize *(Zea mays* L.*)*, *Z. mays MADS2 (ZMM2)* and *ZMM23* are orthologous to *OsMADS3* whereas *ZAG1* is orthologous to *OsMADS58.* The *zag1* mutant shows normal male organs and defective pistils; no mutants for the *OsMADS3* homologs have been isolated so far, but *ZMM2* is highly expressed in the stamens, suggesting a role in stamen development[21,22]. A similar expression pattern can be observed in *Brachypodium distachyon* (L.) P.Beauv. and barley (*Hordeum vulgare* L.), where the orthologs of *OsMADS3* are preferentially expressed in stamens and young seeds, while the orthologs of *OsMADS58* are expressed more broadly throughout the flower[23,24]. In alloplasmic wheat (*Triticum aestivum* L.), the *OsMADS58* ortholog *WAG* has been associated with pistillody (formation of pistil-like stamens)[25].

Ectopic *AGAMOUS* expression in Arabidopsis results in a homeotic conversion of sepals into carpel-like organs with ectopic ovules, and of petals into stamens[26]. Several functional studies, using *AGAMOUS* orthologous genes from other eudicot, monocot or even gymnosperm species, showed that when ectopically expressed in Arabidopsis, they mimic the ectopic expression of *AGAMOUS,* or in some cases they even rescue the *ag* loss of function mutants[14,27–31]. Stunningly, the *AGAMOUS* gene from the gymnosperm *Cycas edentata* de Laub. was able to rescue the Arabidopsis *ag* mutant, despite the two species likely diverged about 300 MYA[28]. Another study shows very clearly the importance of the conservation of the domains required for protein complex formation. The ectopic expression of the snapdragon C-class genes *PLENA (PLE)* and *FARINELLI (FAR)* either in Arabidopsis or snapdragon revealed that the two genes have different abilities in organ specification: *PLE* can convert petals into stamens and induces carpelloid features on the sepals, similar to *AGAMOUS* ectopic expression, while *FAR* only affects petal identity[27,32]. Airoldi and colleagues (2010) showed that this different ability of PLE and FAR is due to different abilities to interact with SEP proteins. Indeed FAR is only able to interact with Arabidopsis SEP3, which is not expressed in the sepals, but not with the other Arabidopsis SEPs[27]. Hence, when FAR is expressed in Arabidopsis sepals it can't find any SEP interaction partner and remains therefore inactive. The restricted interaction ability of FAR is the result of a single glutamine insertion (Q173) in the predicted α-helical region spanning the junction of the K and C domains of FAR[27]. This additional glutamine in the FAR protein of snapdragon is the result of a three-base duplication of the splice acceptor site (CAG) at the 5' end of exon seven[27].

Extensive duplications of homeotic genes have likely been a driving force for the evolution of floral structures and their diversification in angiosperms. Besides gene duplication, variants of homologous proteins can also arise via alternative splicing of transcripts from the same gene. For

MADS-box genes of plants, a few cases of alternative splicing have been reported. Considering only the *AGAMOUS* subfamily, exon skipping gives rise to the double-flowered mutant of *Prunus lannesiana* (Carrière) E.H. Wilson (Rosaceae)[33]. Exon skipping also causes alternative splicing isoforms from *AG* genes of *Magnolia stellata* (Siebold & Zucc.) Maxim. and in the developing fibre cells of cotton (*Gossypium hirsutum* L.)[34,35]. Alternative splicing of *AGAMOUS* genes have also been reported in wild rose (*Rosa rugosa* Thunb.) and *Crocus sativus* L.[31,36], but the functional significance of these events is unknown.

To date, a complete functional characterization of splicing variants has only been carried out for the *AG* lineage gene *PapsAG* from opium poppy (*Papaver somniferum* L.)[37]. *PapsAG* undergoes alternative splicing at the 3' end generating two different transcripts, *PapsAG-1* and *PapsAG-2,* which are co-expressed in stamens and carpels. These two variants only differ in the last part of the C-terminal region: the predicted PapsAG-1 protein terminates, like most AG lineages proteins, with the conserved AG motif II[12] whereas PapsAG-2 has 24 extra amino acidic residues beyond this point. A Virus Induced Gene Silencing (VIGS) approach specifically targeting each one or both these variants, indicated that they share redundantly the C function, however *PapsAG-2* has distinctive roles in gynoecium development, particularly in the septum, ovule and stigma[37].

Here, we describe the identification of two alternative splicing isoforms of the rice *OsMADS3* gene, which differ for only three nucleotides in length, resulting from a three-base pair duplication of a splice acceptor site. The longer transcript encodes for an additional serine residue (S109) within the first predicted α-helix of the K-box, a domain whose length and amino acidic composition are highly conserved among angiosperms[38]. We provide evidence that the two splicing variants are broadly co-expressed in reproductive tissues, and probably ubiquitous within the whole *OsMADS3* clade of grass family. Functional analysis reveals differences in their ability to interact with SEP proteins and their ability to rescue the *ag* mutant phenotype in Arabidopsis.

MATERIALS AND METHODS

Plant materials

A. thaliana L*er*. plants and *Oryza sativa ssp. japonica* cv. Nipponbare were used for our experiments. Arabidopsis plants carrying the *AGAMOUS* mutant allele *ag-3* (Bowman *et al.*, 1989) were already available in our lab.

Mature WT inflorescences of both species were used for total RNA isolation and gene cloning; mature inflorescences from *A. thaliana* transgenic lines were used for total RNA isolation and Real Time PCR analysis.

Phylogenetic analyses

To find suitable RNA-Seq samples from grasses and other Poales, we first interrogated the NCBI SRA database with the name of species of interest, and we eventually selected the most suitable samples (i.e. reproductive stages, when available). Then, for each gene / species, we used the nucleotide BLAST tool (settings: optimize for somewhat similar sequences), to screen those data with a query sequence of 37-47 nucleotides, comprising the eson4-exon5 splicing region, and downloaded all the homologous reads. Control query and read homologous sequences where then aligned with MAFFT, and edited manually with Genedoc. After removing the dubious or low quality reads (i.e. those with more than one or two different bases) the alternative splicing events were counted manually.

Open reading frames were controlled by Gene Runner. Images were processed with Gimp and Paint.NET

RNA isolation

Total RNA was extracted using the LiCl extraction method. Total RNAs were converted into first-strand cDNA by using the ImProm-II Reverse Transcription System (Promega, Madison, USA).

Gene cloning

The CDS of *OsMADS3*^{S109}, *OsMADS3*$^{\Delta S109}$, *OsMADS58* and *OsMADS8* were amplified in a thermocycler (MasterCycler EPGradient S, Eppendorf, Hamburg, Germany) using Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, Waltham, USA) using primers OsP337/OsP118 (for both *OsMADS3* isoforms), OsP119/OsP120 and OsP781/OsP782, respectively. The PCR products were isolated by gel electrophoresis, purified using NucleoSpin Gel Purification Kit (Macherey-Nagel, Duren, Germany); the CDS of *OsMADS3*^{S109}, *OsMADS3*$^{\Delta S109}$ and *OsMADS58* were inserted into a pTOPO vector following the related protocol; the CDS of *OsMADS8* was inserted into a pDONR207 entry vector via Gateway BP Clonase II Enzyme Mix in accordance with the related protocol (ThermoFisher Scientific, Waltham, USA). The recombination products were used to transform electrocompetent *E. coli* cells (Dh10b strain) and colonies positive for the desired

plasmids were isolated on LB-agar plates containing gentamycin (30 mg/L). Colony PCR was performed to confirm the presence of the constructs; in particular, colony PCR was performed to discriminate between colonies carrying *OsMADS3*^{S109} and colonies carrying *OsMADS3*$^{\Delta S109}$, using the reverse primers OsP428 and OsP429 with the forward primer OsP17. The different constructs have been sent for sequencing to the StarSEQ facility (Mainz, Germany).

AGAMOUS CDS had previously been cloned in our laboratory and was already available in a pDONR207 entry vector.

Transgenic plants generation

The CDS of *AGAMOUS*, *OsMADS3*^{S109}, *OsMADS3*$^{\Delta S109}$and *OsMADS58* were inserted, using the Gateway LR Clonase II Enzyme (ThermoFisher Scientific, Waltham, USA), in the pB2GW7 destination vector for expression under the control of the CaMV 35S promoter. The recombination product was used to transform electrocompetent *E. coli* cells (DH10b strain) and positive colonies were isolated on LB agar plates containing spectinomycin (100 mg/L). Colony PCR was performed to confirm the presence of the constructs. Plasmids were re-extracted using NucleoSpin Plasmid Isolation kit (Macherey-Nagel, Duren, Germany) and used to transform electrocompetent *A. tumefaciens* cells (EHA105 strain). Positive colonies were selected on LB-agar plates containing rifampicin and spectinomycin (100 mg/L). Colony PCR was performed to confirm the presence of the constructs.

Agrobacterium strains carrying *pB2GW7-AGAMOUS, pB2GW7-OsMADS3*^{S109}*, pB2GW7-OsMADS3*$^{\Delta S109}$ and *pB2GW7-OsMADS58* were used to transform *ag-3/+* heterozygous plants of *A. thaliana* L*er*. (T_0) with the floral dip method (Clough & Bent, 1998). The T_1 of these plants was harvested and transformed plants were isolated through multiple rounds of BASTA selection. DNA extraction and PCR were performed on selected plants to confirm the presence of the transgene.

Real-Time PCR

Inflorescences of *A. thaliana* were pooled by construct and background (WT/heterozygous or *ag-3*) and total RNA was extracted using the LiCL method and converted into first-strand cDNA by using the ImProm-II Reverse Transcription System (Promega, Madison, USA).

The expression levels of *35S::AGAMOUS, 35S::OsMADS3*^{S109}*, 35S::OsMADS3*$^{\Delta S109}$ and *35S::OsMADS58* were evaluated by Real Time PCR using StoS Quantitative Master Mix (GeneSpin,

Milano, Italy) and primers RT1898/RT1899, RT973/RT974 (for both *OsMADS3* isoforms) and RT975/RT976, respectively.

The expression levels of *SPOROCYTELESS (At4g27330), REM22 (At3g17010), CRABS CLAW (At1g69180)* and *SHATTERPROOF1 (At3g58780)* were also evaluated using primers RT1674/RT1675, RT1672/RT1673, RT1676/RT1677 and AtP650/AtP651, respectively.

Arabidopsis housekeeping gene ubiquitin (*At4g36800*) was used as an internal reference during the experiments.

Yeast-2-hybrid assay

The CDS of *AGAMOUS*, *OsMADS8, OsMADS3*^{S109} and *OsMADS3*$^{\Delta S109}$ were inserted, using the Gateway LR Clonase II Enzyme (ThermoFisher Scientific, Waltham, USA), in the *pGADT7* and *pGBKT7* plasmids for C-terminal fusion to the GAL4 activation domain (AD) and binding domain (BD), respectively. The recombination product was used to transform electrocompetent *E. coli* cells (DH10b strain) and positive colonies were isolated on LB agar plates containing ampicillin (100 mg/L) for *pGADT7* or kanamycin (50 mg/L) for *pGBKT7*. Colony PCR was performed to confirm the presence of the constructs.

The *pGADT7-SEP1, pGADT7-SEP3*$^{\Delta 192}$*, pGBKT7-SEP1* and *pGBKT7- SEP3*$^{\Delta 192}$ were already available in our laboratory.

Combinations of *pGADT7* and *pGBKT7* constructs (see table 2) have been simultaneously used to transform chemo-competent *S. cerevisiae* cells (AH109 strain). Positive colonies for both plasmids were isolated on YSD-agar plates lacking leucine and tryptophan (YSD-L-W). Interaction assays were performed on YSD-agar plates lacking leucine, tryptophan and histidine and containing increasing concentrations of 3-amino-1,2,4-triazole (YSD-L-W-H + 3AT 1mM/2,5mM/5mM) or on YSD-agar plates lacking leucine, tryptophan and adenine (YSD-L-W-A). Growth for interaction assay was performed at 28°C for at least two weeks. As a positive control for interaction, *pGADT7-OsMADS6* and *pGBKT7-OsMADS13* constructs were used (already available in the lab).

PRIMER NAME	SEQUENCE
OsP 17	5’ – TAATACGACTCACTATAGGGTGTGTGTACG-TACGGTGTCTACAC – 3’
OsP 118	5’ – CTGGTTATGATGATCTGACG – 3’
OsP 119	5’ – CACCGAACACACCCCTCTCTTCC – 3’
OsP 120	5’ – CTGATCAGTCTCCTTGAAGAAGG – 3’
OsP 337	5’ – CACCAGTTAGCATACCCATCCATG – 3’
OsP 781	5’ – GGGGACAAGTTTGTACAAAAAAGCAG-GCTTGATGGGGAGAGGGAGGGTGGAGCTG – 3’
OsP 782	5’ – GGGGACCACTTTGTACAAGAAAGCTG-GGTTCAGGAGTTCAGGGTAGCCATGTC – 3’
OsP 428	5’ – AATCCCCCACTATGGTCCTGT – 3’
OsP 429	5’ – AATCCCCCACTATGGTCCTAC – 3’
AtP 650	5’ – GCTGTTAGTGGCAGAGATAGAG – 3’
AtP 651	5’ – CCGGATTCGTAAACTGTCGTC – 3’
RT 147	5’ – CTGTTCACGGAACCCAATTC – 3’
RT 148	5’ – GGAAAAAGGTCTGACCGACA – 3’
RT 973	5’ – GGGATTCTATCAACACCATGAG – 3’
RT 974	5’ – CTCAACTTCAGCATATAACAGC – 3’
RT 975	5’ – CTGCTAAGCTGAAGCAACAG – 3’
RT 976	5’ – CTTCCAGCTGCTTAAGTTCTC – 3’
RT 1672	5’ – AAGGTGCTTTGCTTGGTGAG – 3’
RT 1673	5’ – CACCAACCATCAAAGGGAAC – 3’
RT 1674	5’ – GACACTTGCTTCAAGAAGAAACGTTTGG – 3’
RT 1675	5’ – GAAGAAGATACTGATCGTAGCCGTTCAT – 3’
RT 1676	5’ – CCACACCGTGAAGCTTTCAGTGCT – 3’
RT 1677	5’ – TCACCGAATCCCAAGCCATGGATC – 3’
RT 1898	5’ – AACTTGGAAGGCAGATTAGAGAG – 3’
RT 1899	5’ – GTGGCATAAGCTGCTCGTAG – 3’

Table 1. Primers used in the experiments.

RESULTS

OsMADS3 encodes two protein isoforms differing only for one serine residue in the K-box

OsMADS3 (*LOC_Os01g10504*) locates on the short arm of rice chromosome 1, and in the current MSU annotation is represented by three gene models (Figure A), whose predicted protein products differ only for the few C terminal amino acidic residues after the conserved AG motif II. The variants no. 2 and no. 3 have the same predicted protein product (Figure C), however no. 2 has at least one additional intron in the 3'UTR sequence, in addition of the 8 conserved introns which are typical of *AG* lineage genes[39] (Figure A). The analysis of peaks of mRNA-seq reads and cDNA clones from public databases suggest that *LOC_Os01g10504.2* is the only correct and transcribed gene model (Figure A). However, when we cloned and sequenced the *OsMADS3* coding sequence, we noticed that some clones just lacked 3 nucleotides (TAG) compared to the annotated transcript. These variants differ at the splicing junction between exon four and exon five and caused a loss of a Serine (S) residue in the predicted OsMADS3 peptide (Fig. 1B). Because the length of the N terminal region of AG lineage proteins has been experimentally determined only for the Arabidopsis AGAMOUS[40], here we refer to this serine amino acidic position as S109, counting from the more conserved Methionine residue just preceding the MADS domain (M1, Figure C). Therefore, we named the two protein variants as OsMADS3^{S109} (having S109) and OsMADS3$^{\Delta S109}$ (lacking S109).

S109 potentially alters the organization of the K-box first α-helix, which normally is highly conserved in structure and length, due to its importance in mediating the interaction between MADS-domain proteins[38]. Interesting, in this regard, is that the isoform lacking S109 has the K-box just similar to the other AGAMOUS subfamily members. A deeper analysis revealed that OsMADS3 S109 is encoded by an AGT codon which, in the pre-mRNA, is the 3'-splice acceptor site of the fourth intron (5'-AG-intron-T-3') (Figure B). This 3' acceptor site is duplicated into a TAGTAG repetition (Figure B). The splicing machinery often recognizes the first AG as acceptor site (*OsMADS3^{S109}*), as confirmed by online annotations, but in a significant fraction of the transcripts the second AG site is taken, thus removing a TAG from the mature transcript (*OsMADS3$^{\Delta S109}$*). The frame is thus maintained in *OsMADS3$^{\Delta S109}$*, but the resulting protein lacks S109 (Fig. 1B). The existence of such two transcript variants was also confirmed by analyzing several mRNA-seq data

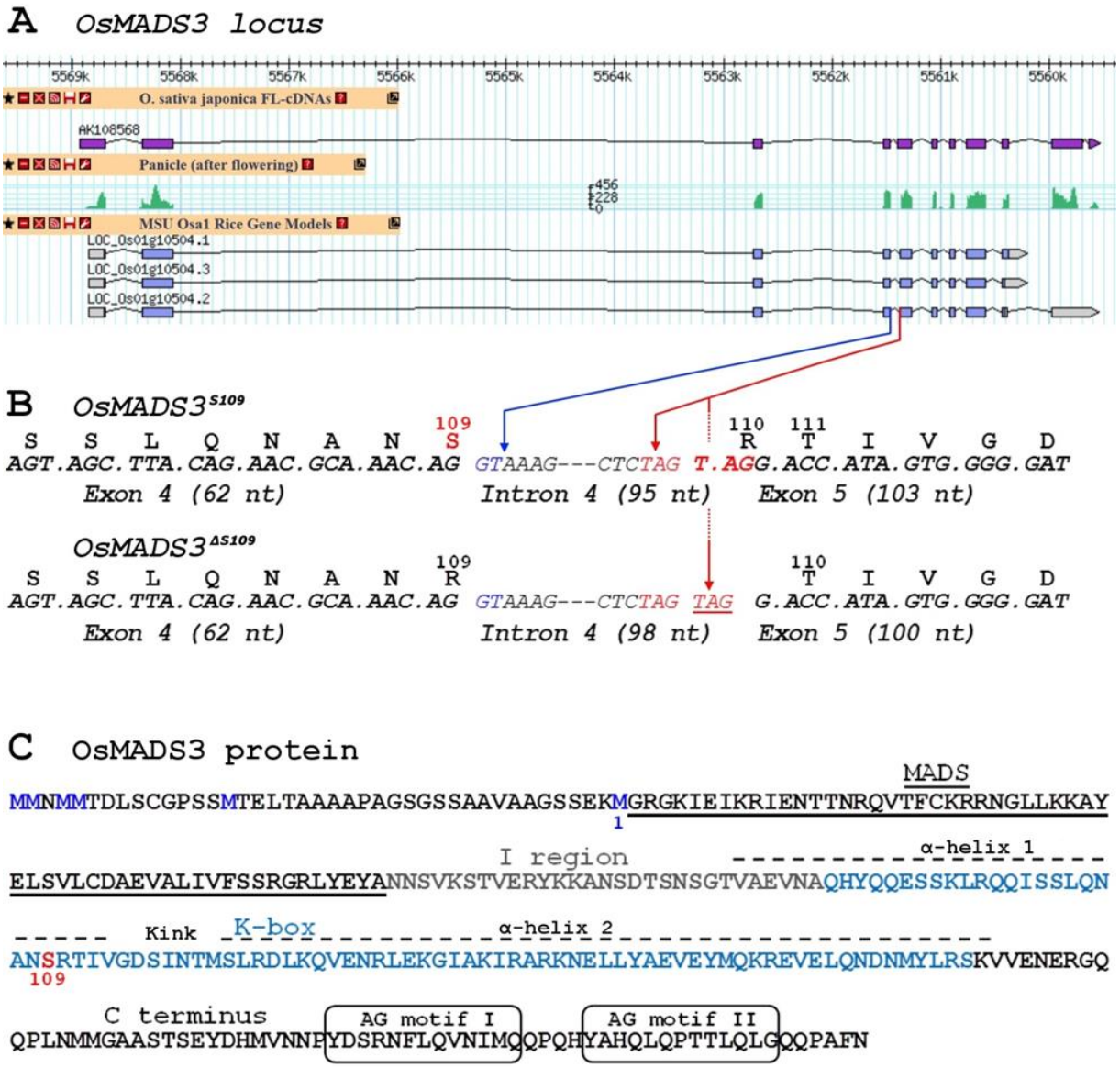


Figure 1. A) Overview of available annotations for OsMADS3 locus (LOC_Os01g10504). B) Detail of the OsMADS3 locus around the 4th intron. The blue arrow indicates the splicing donor site at the 5' of 4th intron; the red arrows indicate the two alternative acceptor sites at the 3' of 4th intron. C) OsMADS3 protein sequence. The putative first methionine (blue), the serine at position 109 (red) and the main conserved motifs are highlighted.

from rice reproductive tissues which are available on the NCBI Sequence Read Archive (SRA). These data clearly show that wherever *OsMADS3* is expressed, roughly 39% of the transcripts carry this 3-nt deletion (Table 1). This amount ranged significantly between different samples (approx. 29-47%), but we could not find any obvious link with tissue type nor stage, as the two transcripts coexisted in early and late panicles, stamens, ovaries and stigmas.

The splicing isoforms involving S109 are ubiquitous in the grass family OsMADS3-like proteins

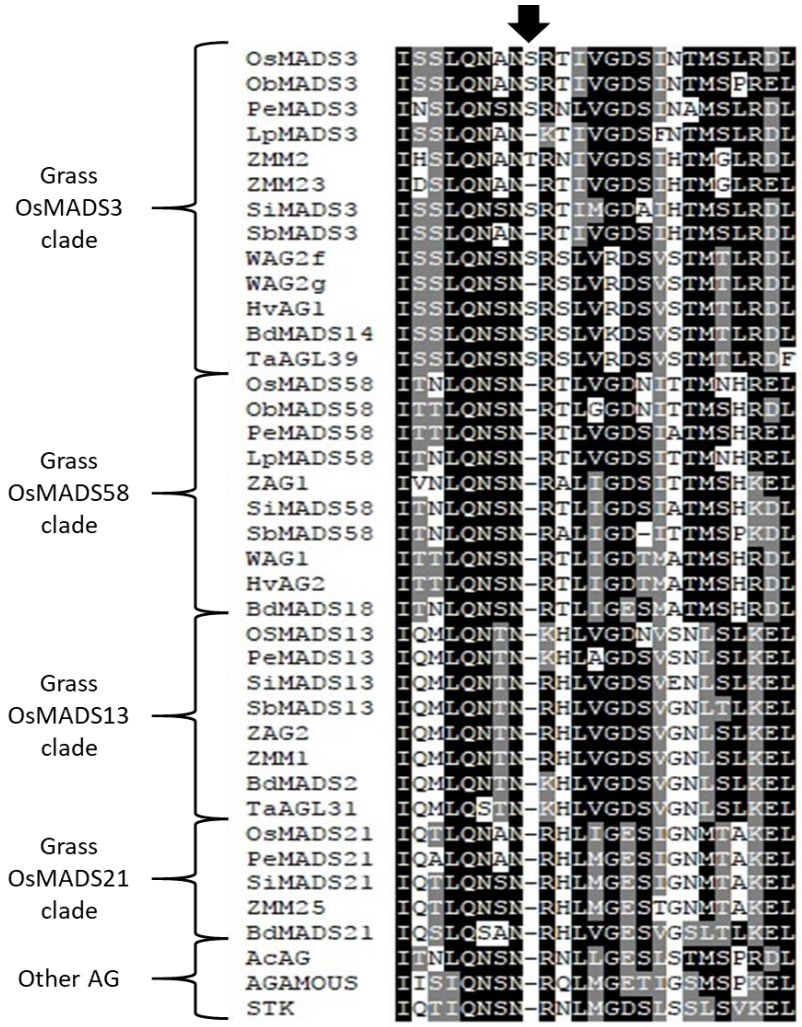


Figure 2. Protein alignment of C- and D-class AGAMOUS-like proteins from different grasses: Os, Oryza sativa; Ob, Oryza brachyantha; Pe, Phyllostachys edulis; Lp, Leersia perrieri; ZM, Zea Mays; Si, Setaria Italica; Sb, Sorghum bicolor; Ta or W, wheat (Triticum aestivum); Hv, Hordeum vulgare; Bd, Brachypodium distachyon; Ac, Ananas comosus. The alignment includes the first part of the K-domain of each protein. The arrow highlights the aminoacidic residue 109 (referred to OsMADS3 sequence). The protein sequences of STK and AGAMOUS from A. thaliana are included in the alignment as representatives of the dicotyledons.

Alignments of class C and D *AGAMOUS* subfamily protein sequences, from several different angiosperms and gymnosperms, suggested that S109 is conserved in many OsMADS3 homologs from other grass species, but it is instead absent in the other members of the AGAMOUS subfamily within the same grass species, including the OsMADS58 subgroup, which is sister to the OsMADS3 clade[41]. It appears to be absent as well outside the Poaceae (grass) family from all the other flowering plants and gymnosperms (Figure 2).

Within the grass *OsMADS3* clade, a search in the Phytozome and Gramene databases revealed that both splicing variants are currently annotated only for rice *OsMADS3*, the *OsMADS3*-like gene of *Brachypodium* and in maize *ZMM2*. The isoform with S109 is the only annotation for *Oryza brachyantha* A.Chev. & Roehr., *O. glaberrima* Steud., *O. longistaminata* A.Chev. & Roehr, *Setaria italica* (L.) P.Beauv., *Sorghum bicolor* (L.) Moench and maize *ZMM23*. The isoform lacking S109 is the only one annotated for *Leersia perrieri* (A.Camus) Launert, *O. barthii* A.Chev., *O. glumaepatula* Steud., *O. meridionalis* N.Q.Ng, *O. nivara* S.D.Sharma & Shastry, *O. rufipogon* Griff., and *O. sativa ssp. indica*. Both the variants of rice and *Brachypodium* are also reported on NCBI (XM_015777004, XM_015777012, XM_010232289 and XM_010232291). Furthermore, we noticed that this alternative splicing corresponds to the one recently described in the *OsMADS3* homolog of *Aegilops tauschii*[42], which is the D-genome donor

of bread wheat, also if the author interpreted the alternative splicing as involving an AGT instead of a TAG sequence, probably for the ambiguity in aligning the last two nucleotides 'AG' of the fourth exon in the variant lacking TAG. However, when we analyzed the genomic *loci* of various *OsMADS3* orthologs, we noticed that the TAGTAG duplication in the 3' splice acceptor site of intron four, is conserved throughout the whole Poaceae family, and just changes into CAGTAG in *Brachypodium* and barley (a selection of species is shown in Figure 3). The duplication is not observed in any of the three other clades of the *AGAMOUS* subfamily in grasses, i.e. *OsMADS13, OsMADS21* and *OsMADS58*. Therefore, the alternative splicing might occur ubiquitously in *OsMADS3* genes of grasses, without being fully represented in current databases. To verify this hypothesis, we extended the analysis of published mRNA-Seq reads to other species representative of the main grass evolutionary clades: *Streptochaeta angustifolia* Soderstr. (from the basalmost subfamily Anomochlooideae), *Phyllostachys edulis* (Carrière) J. Houz. (Bambusoideae), *Leersia perrieri* (another species from Oryzoideae), *Brachypodium* and barley (Pooideae), *Sorghum* and maize (Panicoideae). Stunningly, the analysis revealed that the S109 alternative splicing is indeed conserved across grasses. Male, female and seed specific samples were available for *Brachypodium*, *Sorghum* and maize which indicated that, like in rice, the two transcript variants do not have an obvious tissue or stage specificity. The isoform lacking S109 was the predominant one only in *Streptochaeta*, *Leersia*, *Sorghum* and in maize *ZMM23*, on the contrary to *ZMM2* (Table 2).

Outside the Poaceae group, the TAGTAG duplication allowing the alternative splicing seems not to exist, leaving only a typical *AGAMOUS*-like transcript to form, which always lacks the additional aminoacid from this region. This is the case of other commelinids like *Musa acuminata* Colla*, Elaeis guineensis* Jacq. and *Ananas comosus* (L.) Merr., as well as any of the other angiosperms and gymnosperms analyzed (Figure 3).

In the order Poales, *Ananas comosus* (pineapple) is an excellent sister group of Poaceae to study, because of its high quality whole genome sequence[43], and because it diverged before the whole genome duplication (WGD) event that left obvious traces in any grass genome studied so far[44]. This grass-specific WGD caused the split of the *AG* lineage into the grass *OsMADS3* and *OsMADS58* clades, which are therefore sister paralogous clades equally related to the pineapple *AGAMOUS* gene, *AcAG*. In *AcAG*, there is no TAGTAG duplication in the fourth intron 3' splicing acceptor site (Figure 3) and, accordingly, all the mRNA-Seq reads miss S109 (**Table 2**). Despite no genomic data

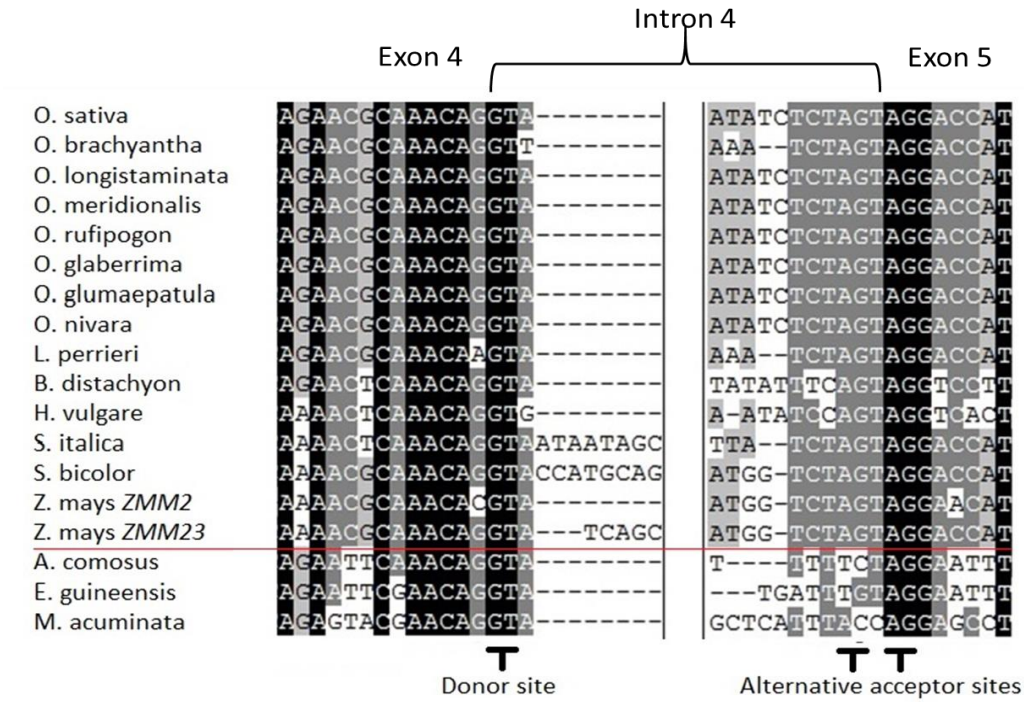


Figure 3. Alignment of corresponding splicing sites from the pre-mRNA of *OsMADS3* and its direct orthologs from different monocotyledon species. The red line divides grasses (above) and non-grasses (below).

are yet available from Poales other than grasses and pineapple, a limited amount of mRNA-Seq reads allowed us to test the presence of the alternative splicing in the lineages closest to grasses.

Nearly 2000 reads from *Luzula elegans* (Juncaceae) support only the conserved splicing tipical of *AGAMOUS* subfamily genes (Table 2). The so called graminid clade comprises grasses plus very few species included in the families Flagellariaceae, Joinvilleaceae and Ecdeiocoloeaceae (Supplemental Figure 1), where we have found clear evidence of S109 alternative splicing at least in the latter (Table 2). Sister to the graminid clade is the restiid clade (Supplemental Figure 1), of which all previous families have recently been located into the Restionaceae according to the current APG IV phylogeny[45]. Surprisingly, a consistent amount of reads suggested that in this family an alternative splicing may remove, rather than adding, an amino acidic residue from *AGAMOUS* transcripts (data not shown).

In conclusion, the alternative splicing of *OsMADS3* exons four and five is found ubiquitously in grasses, and probably arose before their origin. This would imply that it originated in the pre-grass *AGAMOUS* lineage and that, after the split of *OsMADS3* and *OsMADS58* clades, it remained conserved only in the former. However, an *OsMADS3* clade-specific origin cannot be excluded, until the grass WGD event can be exactly placed in relation to grasses and its poorly studied sister families.

NCBI SRA Experiment	**Species**	Query sequence	Total good reads	**Long splicing S109**	**TAG deletion ΔS109**
SRX1332256 SRX507920 DRX000335 DRX000334 DRX000333 SRX100746	***Oryza sativa ssp. japonica***	TAGCTTACAGAACGCA AACAGTAGGACCATAG TGGGGGATTC	2021	**1238 61.3%**	**783 38.7%**
SRX472914	***Leersia perrieri***	CAGTAGCTTACAGAAC GCAAACAAGACCATAG TGGGGGATTC	1205	**562 46.6%**	**643 53.4%**
SRX2342718 SRX2342716 SRX1583837 SRX1583838 SRX1583839 SRX100693 SRX100694 SRX100691	***Brachypodium distachyon***	AGCCTGCAGAACTCAA ACAGTAGGTCCTTAGT GAAGG	391	**324 82.9%**	**67 17.1%**
SRX375649 SRX378862	***Hordeum vulgare***	CTTGCAAAACTCAAAC AGTAGGTCACTGGTGA GAGATT	39	**23 59%**	**16 41%**
SRX099185 SRX099141 SRX099021	***Sorghum bicolor***	GCTTGCAAAACGCAAA CAGTAGGACCATAGTG GGAGAT	223	**40 17.9%**	**183 83.1%**
SRX058598 SRX058600 SRX058607 SRX058605 SRX058604 SRX058601 SRX058599 SRX058597	***Zea mays***	ZMM2: CAGCTTGCAAAACGCA AACACTAGGAACATAG TGGGAGATTC	496	**372 75%**	**124 25%**
		ZMM23: TAGCTTGCAAAACGCA AACAGTAGGACCATAG TGGGAGATTC	118	**23 19.5%**	**95 80.5%**
SRX2375352 SRX2375353 SRX2375354 SRX567871	***Phyllostachys edulis***	TAGCTTGCAGAACTCA AACAGTAGGAACTTAG TGGGGGATTC	83	**22 26.5%**	**61 73.5%**
SRX1639030	***Streptochaeta angustifolia***	TTACCACCTTGCAGAA CAACAACAGGACCATA ATGGGGGATTCTGTA	11	**1 9.1%**	**10 90.9%**
SRX1639019	***Ecdeiocolea monostachya (Ecdeiocoleaceae)***	TCACCAACTTGCAGAA CTCCAATAGCAGGACT ATACAGGCAGGGGA	100	**24 24%**	**76 76%**
SRX1639028	***Joinvillea ascendens (Joinvilleaceae)***	NO READS AVAILABLE FOR THE AGAMOUS EXON4-EXON5 REGION			
SRX1639020	***Flagellaria indica (Flagellariaceae)***	TACCAACTTGCAGAAC TCAAATAGGACTCTAC TGGGGGATTCA	23	**0 0%**	**23 100%**
ERX1349704	***Luzula elegans (Juncaceae)***	AGCACATTAAATAACA GTAATAGGAATTTGTT GGGCGAG	1921	**0 0%**	**1921 100%**
SRX1465570 SRX1465595	***Ananas comosus (Bromeliaceae)***	CAACCTCCAGAATTCA AACAGGAATTTACTGG GTGAGTCTCT	232	**0 0%**	**232 100%**

Table 2. Summary of the number of reads mapping on the exon 4-exon 5 junction of the *MADS3*-like gene from various monocots species. The percentage of reads indicating the presence or absence of the second TAG at the splicing acceptor site of intron 4 (or correspondent intron) is reported.

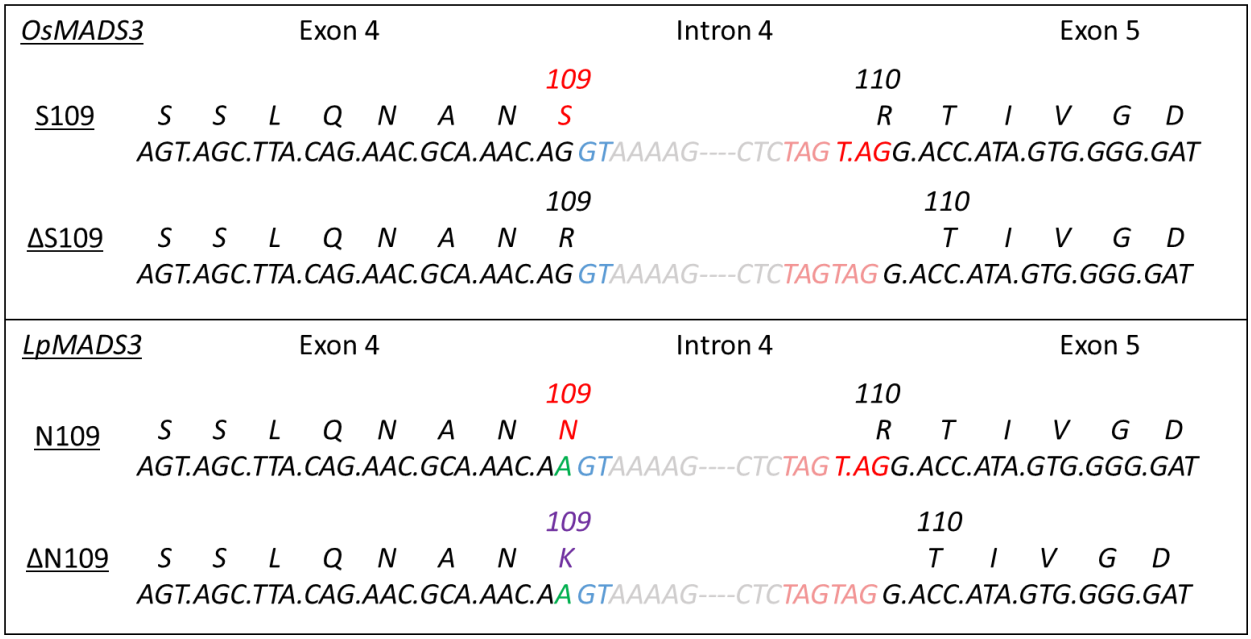


Figure 4. Schematic representation of alternative splicing possible outcomes in *OsMADS3* from *Oryza sativa* and its homolog from *Leersia perrieri.* The TAG repetition where the alternative splicing occurs is highlighted in red; the G-A substitution in *L. perrieri* genomic sequence is highlighted in green.

In most cases the additional amino acidic residue 109 is a Serine, some species show other amino acids in this specific location which further suggest that the conservation of this position is crucial. In maize ZMM2 and in *Leersia* LpMADS3, the Serine is replaced by Threonine and Asparagine respectively (T109 and N109), other amino acids with polar uncharged side chains. This suggests that this residue is important not only for its position, but also for it physical/chemical properties, on the contrary to Q173 in snapdragon[27].

In almost all the species that we considered, the presence or absence of the codon for S109 does not alter the translation of the surrounding codons, which usually encode for an asparagine (upstream) and an arginine (downstream). Due to a G to A substitution in the last nucleotide of exon four, in *Leersia perrieri*, a species very close to the *Oryza* subgroup, not only S109 is replaced by N109 in the long splicing variant, but in the short variant missing N109, the downstream arginine (R) is replaced by a lysine (K) (Figure 4). Yet, this substitution doesn't change the electrical and biochemical features of the side chain, which keeps being positively charged (they both have a basic NH_2^+ group). A different case is maize ZMM2 where, in the short variant, splicing out the codon for T109 generates another codon for T, rather than for the highly conserved downstream R. However, the presence of two OsMADS3 homologs in maize probably allows a much more relaxed selective pressure and more possibilities of subfunctionalizaition.

OsMADS3 isoforms have different ability to induce homeotic changes in Arabidopsis flowers

In order to compare the functionality of the two OsMADS3 isoforms, which we named as OsMADS3^{S109} (having S109) and OsMADS3$^{\Delta S109}$ (lacking S109), we first expressed them in Arabidopsis as heterologous model system, along with OsMADS58 and Arabidopsis AGAMOUS as control. To do so, we cloned the coding sequences of *AGAMOUS*, *OsMADS3^{S109}*, *OsMADS3$^{\Delta S109}$* and *OsMADS58* downstream the CaMV 35S promoter and we introduced these constructs into Arabidopsis plants which were heterozygotes for the *ag-3* mutation[46]. In many previous publications, the effects of ectopic heterologous C-class gene expression have only been evaluated in a WT background[27,29,47]. In this study we used the progeny of a T_0 segregating for the *ag-3* allele, because we were interested to observe the ability of rice proteins to compensate for the lack of AG activity in Arabidopsis.

In different rounds of transformation and BASTA selection performed, we could select about 200 resistant plants for each construct, where the presence of the transgene was confirmed by PCR. All plants showed dwarfism, curled rosette leaves and early flowering, even in short day conditions; these are common phenotypes due to the ectopic expression of AG-like proteins. In addition to this, about 50% of the selected plants for each construct (with the exception of plants expressing OsMADS58, as explained in a following paragraph) displayed a variety of phenotypes related to flower morphology; the other half, despite bearing the construct, did not show any phenotype and appeared identical to the WT (or to the *ag-3* mutant, depending on the background).

We selected and followed 20 T_1 plants for each construct, 10 with evident morphological phenotypes and 10 with no evident phenotype. All these plants had an *ag-3/+* heterozygous background, so that in the T_2 generation we could once again observe the effect of the constructs in all different backgrounds.

Our analysis on the transgenes segregation and expression demonstrates the link between phenotypes observed and transgene expression: whenever the transgene is there, but it is not expressed, plants are again identical to the untransformed controls (Figure 5).

Morphological analysis of flowers of the different transgenic lines

Plants transformed with the *35S::AGAMOUS* construct represented our positive controls. In the WT background we could observe a mild to strong conversion of second whorl petals towards

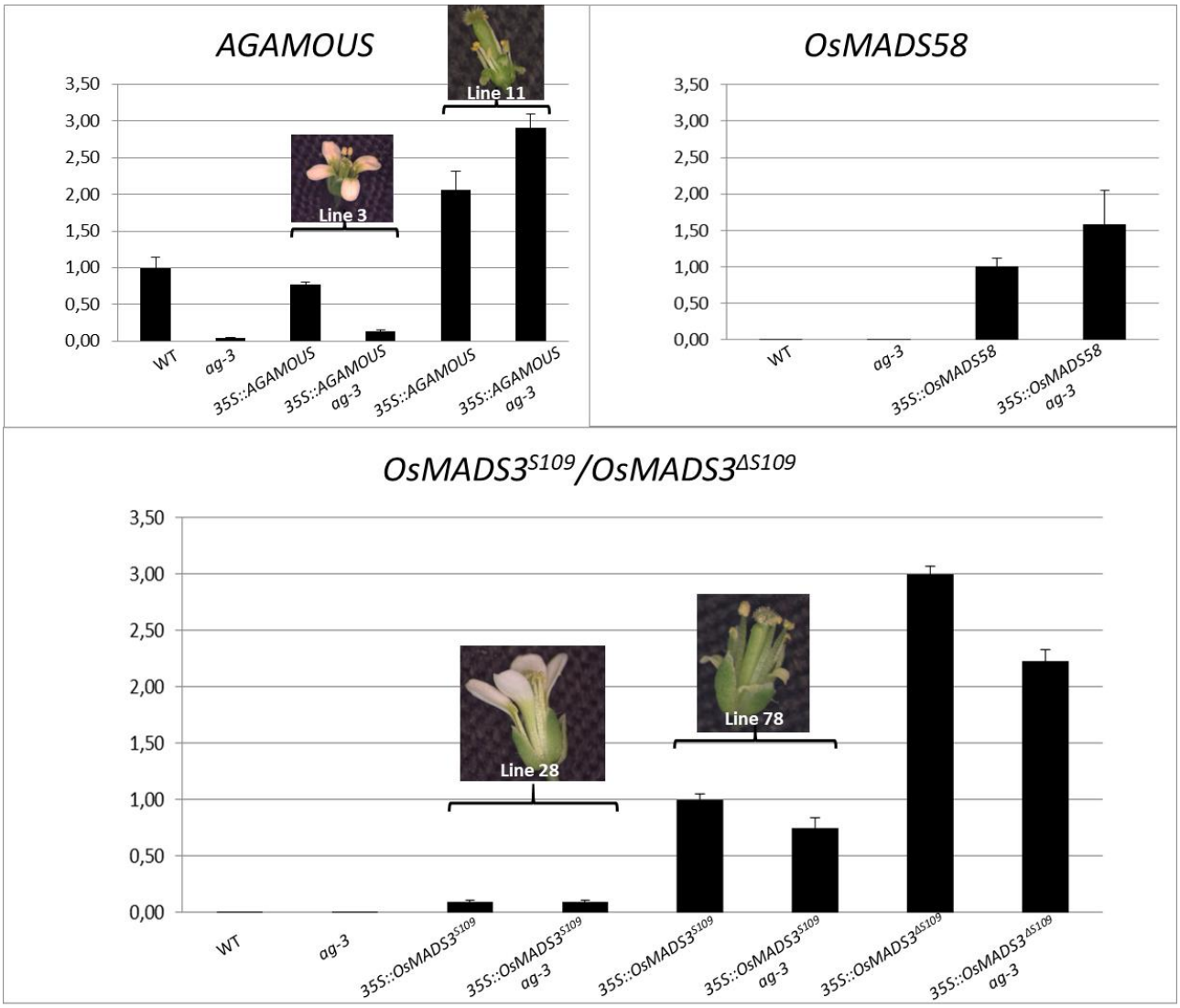


Figure 5. Expression levels of *35S::AGAMOUS, 35S::OsMADS58, 35S::OsMADS3^{S109}* and *35S::OsMADS3$^{\Delta S109}$* in WT and *ag-3* background, compared to non-transgenic lines. *AGAMOUS* levels have been normalized against the WT level, where the endogenous *AGAMOUS* is also detected; *OsMADS58* and *OsMADS3* levels have been normalized against the expression level of the transgene in the WT background. An example of a line where *35S::AGAMOUS* (line 3) and *35S::OsMADS3* (line 28) are expressed at low levels is also reported. Standard deviation is indicated.

stamens or staminoid-like structures (Figure 6B,C and E). These phenotypes were also observed in plants with an *ag-3/+* heterozygous background. SEM images of anther-like petals in *35S::AGAMOUS* plants showed that the base looks very similar to the actual anther filament (Fig. 5C), while the upper part of the organ resembles the shape of an anther, but the surface cells look spherical and resemble those of the petal surface (Figure 7E).

In the *ag-3* mutant background, all petals acquired a yellowish appearance (Fig. 5I) and stigmatic tissues appeared on the sepals (Figure 6G). In addition, 30% of the plants recovered FMD almost

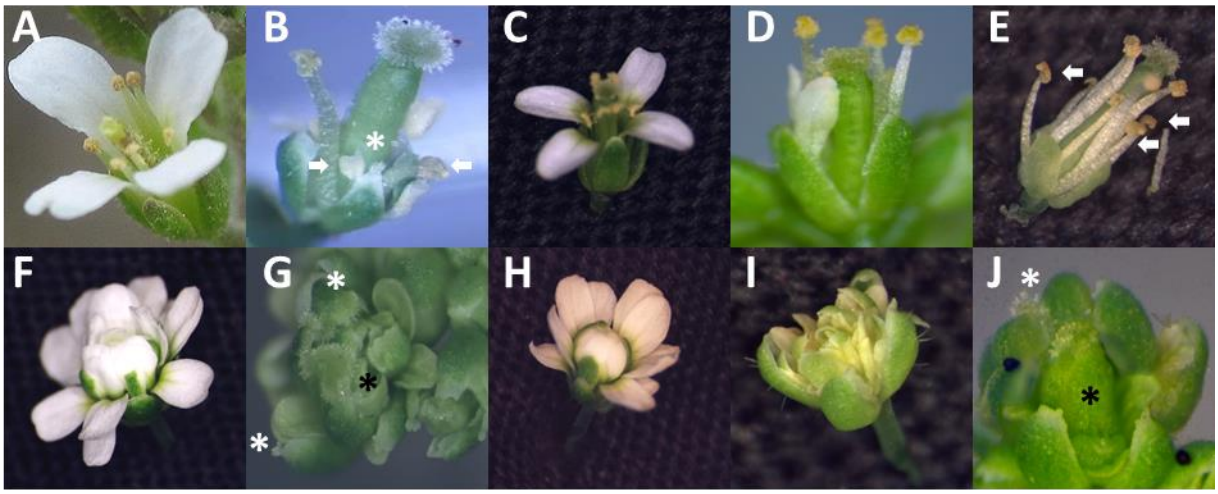


Figure 6. Flowers of plants expressing *35S::AGAMOUS* (B,G), *35S::OsMADS58* (C, H), *35S::OsMADS3*^{S109} (D, I) and *35S::OsMADS3*$^{\Delta S109}$ (E, J) in WT (B-E) and *ag-3* (G-J) background, compared to WT (A) and *ag-3* mutant (F). Ectopic stigmas on sepals (white stars), petal-to-stamen conversion (white arrows) and meristem determinacy recovery with formation of carpel-like organs (black stars) are highlighted.

completely. In these plants, the meristem terminated with a carpelloid organ arising from the fourth whorl, closed by stigmatic tissue at the tip (Figure 6G; Figure 7F). None of the *35S::AG* lines in the *ag-3* background were able to produce seeds.

Petals of *35S::AG ag-3* plants showed a mild conversion to anthers when observed by SEM (Figure 7H). The cell shape mostly recalled that of a normal petal except some group of cells which acquired a lobate shape tipical of anther cells. Moreover the overall shape of the terminal part of these organs look like stamens. Another detail to notice is that, in the first whorl, the nature of the cells surrounding the ectopic stigmas in *35S::AG* lines with *ag-3* background varied from irregular sepal cells to smoother and longer cells which mostly recalled those of the style (Figure 7D and G). In some plants, ectopic ovules were visible near the stigmatic tissue on sepals (Figure 7G).

Plants transformed with the *35S::OsMADS58* construct never showed any change in phenotype, neither in WT nor in the *ag-3* background, regardless of the transgene expression level (Figure 6C and H). It looks like *OsMADS58* is not able to produce any effect when expressed in Arabidopsis. It is worth to mention that the OsMADS58 protein lacks a conserved C-terminal AG II motif, which might affect its functionality[12,48].

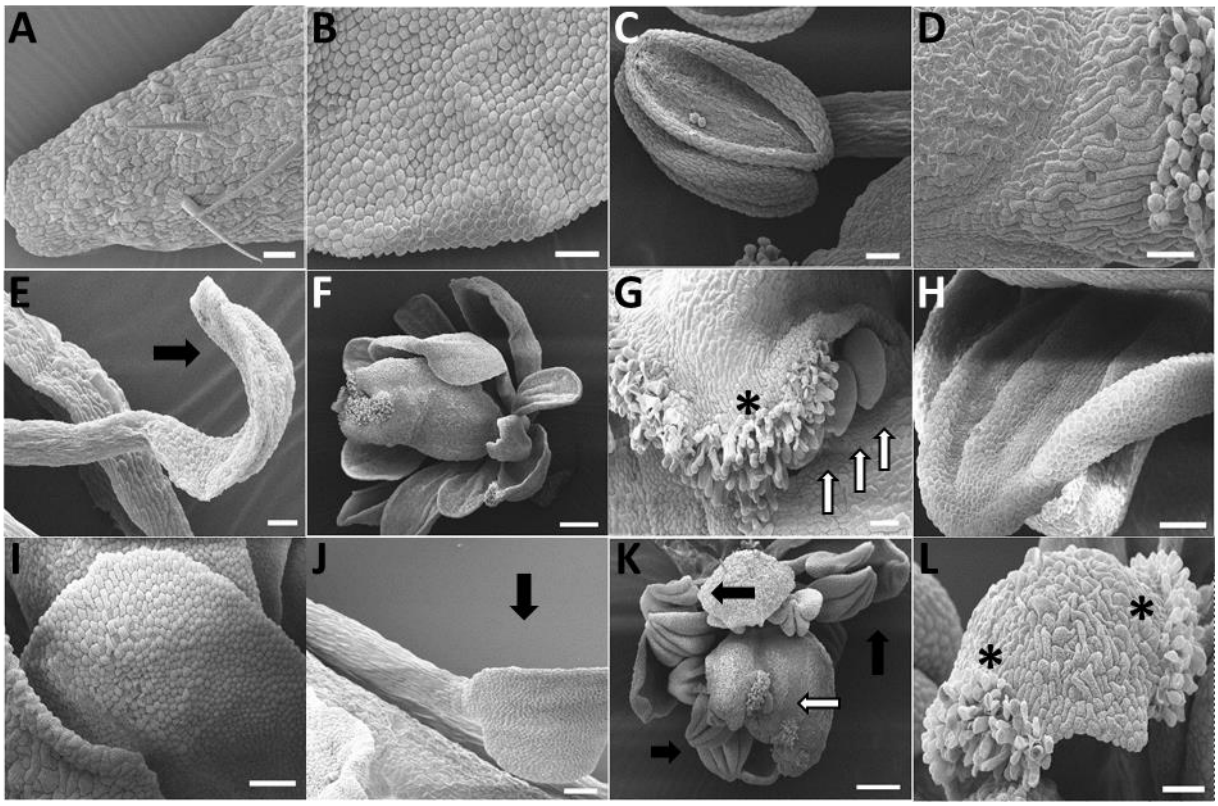


Figure 7. SEM analyses of ectopic organs.

- **A) WT sepal**
- **B) WT petal**
- **C) WT stamen**
- **D) WT carpel**
- **E) Petal-to-stamen conversion in *35S::AGAMOUS* (WT background)**
- **F) Meristem determinacy recovery in *ag-3 35S::AGAMOUS***
- **G) Ectopic stigmas and ovules on sepals in *ag-3 35S::AGAMOUS***
- **H) Mild petal-to-stamen conversion in *ag-3 35S::AGAMOUS***
- **I) Mild petal-to-stamen conversion in *ag-3 35S::OsMADS3^{S109}***
- **J) Petal-to-stamen conversion in *35S::OsMADS3$^{\Delta S109}$***
- **K) Staminoid petals and meristem determinacy recovery in *ag-3 35S::OsMADS3$^{\Delta S109}$*; an ectopic ovule is also visible**
- **L) Ectopic stigmas on sepals in *ag-3 35S::OsMADS3$^{\Delta S109}$***

Ectopic stigmas on sepals (black stars), petal-to-stamen conversion (black arrows) and ectopic ovules white arrows) are highlighted. Reference bar for F, K: 200 µm; reference bar for the other pictures: 50 µm.

In Arabidopsis plants expressing *35S::OsMADS3^{S109}* in a WT background, the main phenotype was a strong reduction in sepals and petals length; as a result, reproductive organs get uncovered much earlier than usual, around stage 6-7 of flower development, while in WT flowers the bud opens around stage 13 (Figure 6D; stages indicated according to Meyerowitz *et al.*[4]). Occasionally we observed a mild petal-to-stamen conversion (not shown).

35S::OsMADS3^{S109} expression in the *ag-3* mutant background caused that all petals acquired a yellowish appearance (Figure 6I). SEM images show that the overall identity of these organs is that

of a petal, even if some cells acquired a lobate shape tipical of an anther-like cell (Figure 7C and I). Notably none of the plants were able to recover the FMD.

At the morphological level, WT plants expressing *35S::OsMADS3$^{\Delta S109}$* behave very similarly to those expressing *35S::AGAMOUS*. In the WT background they show a strong petal-to-stamen conversion with cells having a lobate shape tipical of anther-like cells (Figure 6E; Figure 7B, C and J).

In the *ag-3* background, stigmatic tissue can be observed on outer sepals and about 60% of the plants also show FMD recovery and formation of carpel-like organs in the fourth whorl (Figure 6J; Figure 7K and L). SEM images reveals that also in the *ag-3* background all petals are converted to stamens in an even more pronounced way than in plants expressing *35S::AGAMOUS* (Figure 7K).

Homeotic conversions at the molecular level

In order to demonstrate at a molecular level the identity of ectopic organs that we found in the transgenic inflorescences we performed qRT-PCR experiments to check the expression of four genes: *SPOROCYTELESS (SPL)*[49] and *REPRODUCTIVE MERISTEM 22 (REM22)*[50], as markers for stamen identity, and *CRABS CLAW (CRC)*[51] and *SHATTERPROOF1 (SHP1)*[18], as markers for carpel identity. All of these genes have been described as direct targets of *AGAMOUS* and they all appear strongly downregulated in the inflorescence of *ag-3*. The analysis has been performed on plants expressing the transgenic constructs both in WT and the *ag-3* mutant background (Figure 8; Figure 9). *SHP1* and *REM22* expression levels remain mostly unaltered in transgenic lines with a WT background, except for plants expressing *35S::OsMADS3$^{\Delta S109}$*, where *REM22* appears upregulated (Figure 8); *SPL* and *CRC* levels are comparable to WT levels in plants expressing *35::AGAMOUS* and *35S::OsMADS58*, but they are surprisingly downregulated in plants expressing *35S::OsMADS3^{S109}* and *35S::OsMADS3$^{\Delta S109}$* (Figure 8).

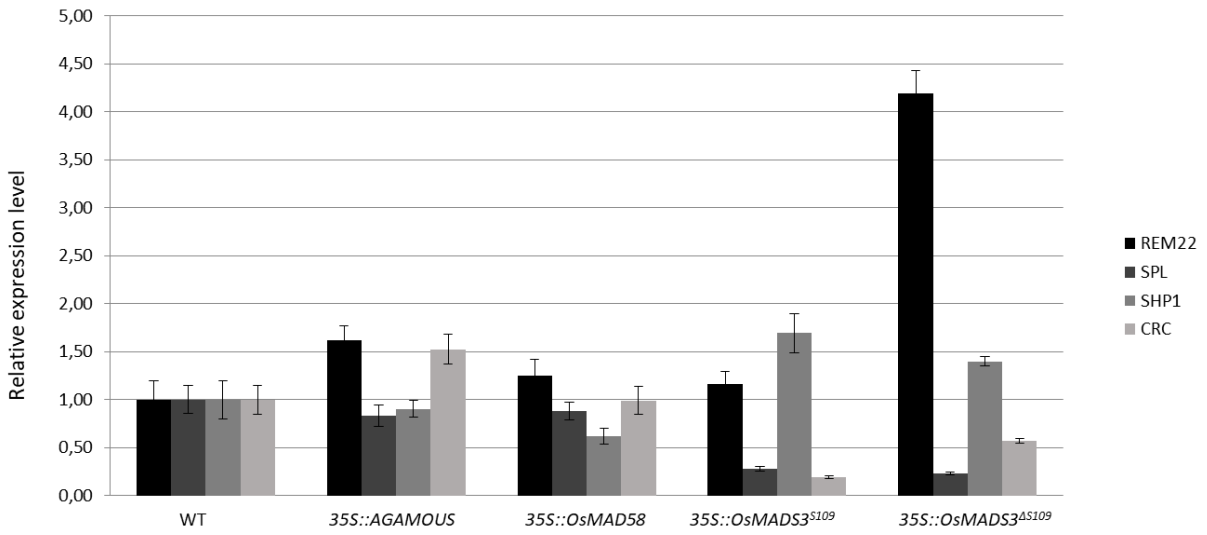


Figure 8. Expression of *REM22, SPL, SHP1* and *CRC* in transgenic lines with WT background. The expression levels have been normalized against the WT. Standard deviation is indicated.

In plants expressing the *35S::AGAMOUS* in the *ag-3* background, all marker genes appeared to be upregulated compared to non-transformed plants (Figure 9). This proves that, even if the construct was not able to perfectly recover the mutant phenotype, it was able to promote part of the floral molecular program; moreover, it confirms our hypotheses about the mild petal-to-stamen conversion observed in the *ag-3* background (Figure 7H).

As expected, expression level of all marker genes in plants expressing *35S::OsMADS58* remain unaltered compared to the untransformed mutant (Figure 9); this perfectly correlates with the fact that no alteration in the flowers of these plants has been observed.

In plants expressing *35S::OsMADS3^{S109}*, qRT-PCR analysis showed that *REM22* and *SPL,* markers of stamen identity, were upregulated in plants with *ag-3* mutant background, despite the fact no full stamens are formed in those plants (Figure 9). Intriguingly, this seems to correlate with the partial subfunctionalization of *OsMADS3* for stamen specification in rice and, again, it confirms our observations of the SEM analysis (Figure 7I). The two markers for carpel identity, on the contrary, remain unaltered and this correlates with the lack of ectopic stigmas.

Plants expressing *35S::OsMADS3$^{\Delta S109}$*, as explained above, have a phenotype very similar to those expressing *35S::AGAMOUS.* This similarity is confirmed by the qRT-PCR analyses: markers for stamen identity appear upregulated, same as in the other transgenic lines, but also *SHP1* is upregulated in the *ag-3* background (Figure 9) and this could account for the formation of stigmatic tissues which is not observed in plants expressing *35S::OsMADS3^{S109}*. The expression level of *CRC*, instead, remains unaltered, while in *35S::AGAMOUS* plants this gene is also upregulated (Figure 9).

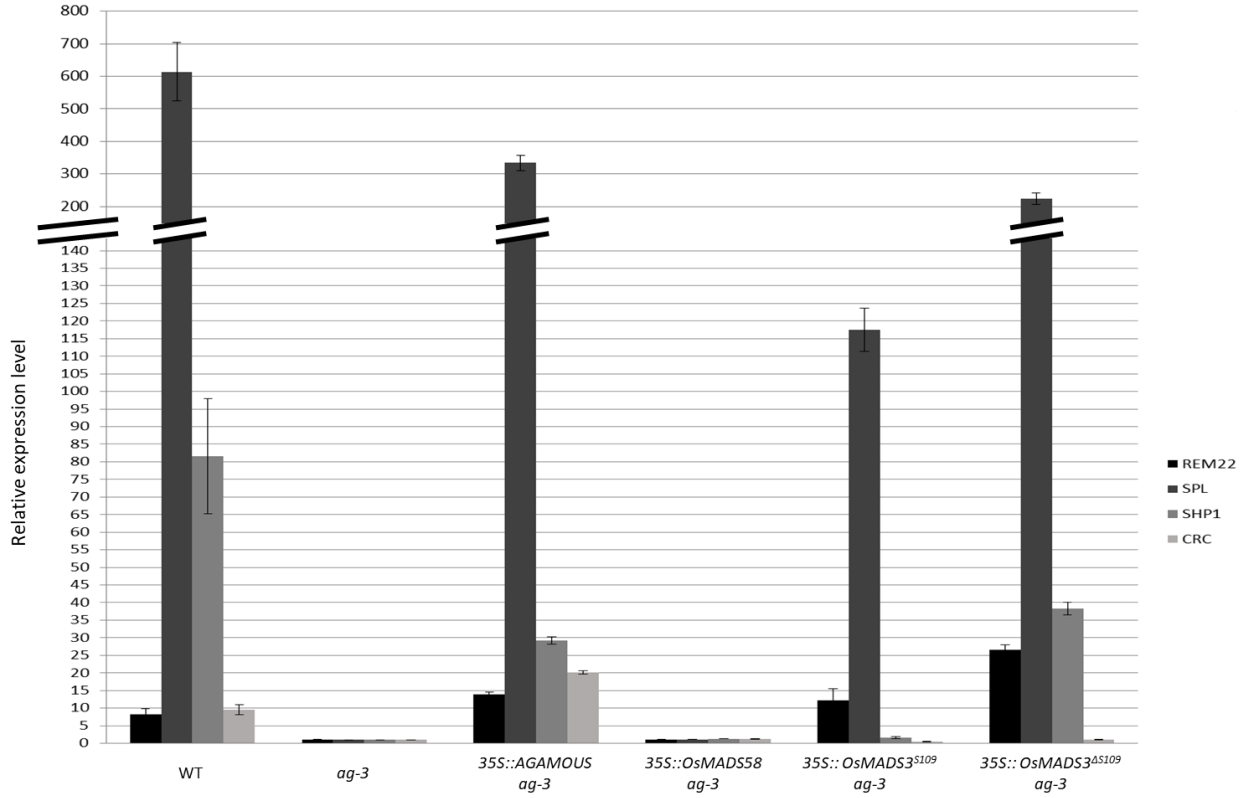


Figure 9. Expression of *REM22, SPL, SHP1* and *CRC* in transgenic lines with *ag-3* background. The expression levels have been normalized against *ag-3* mutant. Standard deviation is indicated.

Lack of S109 weakens the interaction between OsMADS3 and SEPs in yeast-2-hybrid assay

The activity of C-class proteins takes place through protein-protein interactions with members of the E-class. Since the difference between the two OsMADS3 isoforms is one aminoacid in the K region, which is known to be involved in protein-protein interaction in MADS-box genes, we hypothesized that different interaction ability could account for the different ability of the two isoforms in recovering the *ag-3* mutant.

We therefore tested the interaction of AGAMOUS, OsMADS3^{S109} and OsMADS3$^{\Delta S109}$ with Arabidopsis SEP1/2/3 proteins using the yeast-2-hybrid technique (Table 3). The result was that both AGAMOUS (as expected) and OsMADS3^{S109} could interact well with SEP1 and SEP3, while OsMADS3$^{\Delta S109}$ interactions were surprisingly weaker or completely absent (Figure 10A, B). This is in contrast with the fact that OsMADS3$^{\Delta S109}$ and AGAMOUS ectopic expression produce very similar morphological effects and that the loss of S109 should make the OsMADS3 structure more similar to AGAMOUS.

Interestingly, replacing the Arabidopsis SEPs with OsMADS8, a rice ortholog of SEP3 and a known interactor of OsMADS3[52], we observe once again that OsMADS3$^{\Delta S109}$ interactions are weaker or absent, compared to OsMADS3. (Figure 10C). This last experiment in particular suggests that S109 is important to guarantee the correct protein-protein interactions and that OsMADS3 protein has evolved to be functional with the addition of this aminoacid.

DISCUSSION

In 2014, the crystal structure of a portion of the SEPALLATA3 protein from *A. thaliana* was resolved by Puranik and colleagues[53]. This portion includes part of the intervening domain and the whole K-domain. The structure clearly shows how the K-domain folds into two α-helices rather than three as previously suggested[53], and they are separated by a rigid kink.

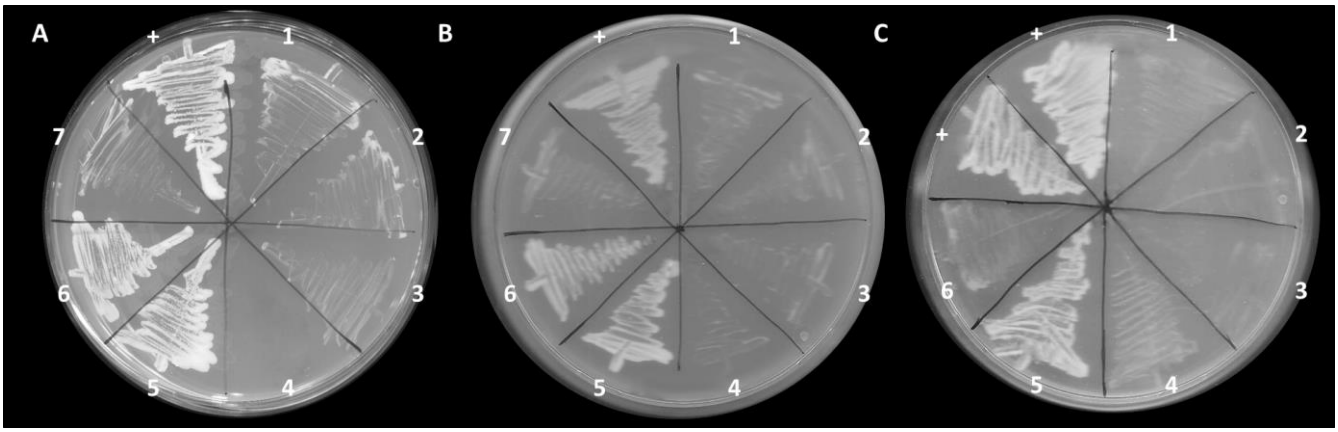
The OsMADS3 protein shares 43% similarity in the K-domain with SEP3, thus allowing us to hypothesize a similar 3D structure. Considering this hypothesis, serine residue S109 falls in the first α-helix (K1) of the K-domain. In order to fold correctly, α-helices need to have a quite constrained protein sequence; the absence or presence of a single amino acid (regardless of its nature) could interfere or even disrupt the correct folding, thus altering protein functionality. A good example is shown in literature by Airoldi and colleagues, where the mutation was related to the K-domain[27].

All the proteins in the AGAMOUS subfamily from both ancestral and evolved families lack this serine residue (or any other residue) in the considered position, indicating the insertion has

Figure 10. Yeast-2-hybrid assays.

- **A) Interaction test between AD-SEP1 and BD-AG (5), BD-OsMADS3^{S109} (6) and BD-OsMADS3$^{\Delta S109}$ (7).**
- **B) Interaction test between AD-SEP3Δ192 and BD-AG (5), BD-OsMADS3^{S109} (6) and BD-OsMADS3$^{\Delta S109}$ (7).**
- **C) Interaction test between AD-OsMADS8 and BD-OsMADS3^{S109} (5) and BD-OsMADS3$^{\Delta S109}$ (6).**

In each plate, negative controls (1-4) and positive control (+) are shown.



	YSD -L-W-H	YSD -L-W-H + 3AT 1mM	YSD -L-W-H + 3AT 2.5mM	YSD -L-W-H + 3AT 5mM	YSD -L-W-A
AGAMOUS – SEP1	++	++	++	++	nt
AGAMOUS – SEP2	++	-	-	-	nt
AGAMOUS – SEP3$^{\Delta 192}$	++	++	++	++	++
SEP1 – AGAMOUS	++	++	++	++	nt
SEP2 – AGAMOUS	++	-	-	-	nt
SEP3$^{\Delta 192}$ – AGAMOUS	++	++	++	++	nt
OsMADS3^{S109} – SEP1	++	-	-	-	nt
OsMADS3^{S109} – SEP2	++	-	-	-	nt
OsMADS3^{S109} – SEP3$^{\Delta 192}$	++	-	-	-	-
SEP1 – OsMADS3^{S109}	++	++	++	++	nt
SEP2 – OsMADS3^{S109}	++	-	-	-	nt
SEP3$^{\Delta 192}$ – OsMADS3^{S109}	++	++	++	+	nt
OsMADS3$^{\Delta S109}$ – SEP1	+	-	-	-	nt
OsMADS3$^{\Delta S109}$ – SEP2	++	-	-	-	nt
OsMADS3$^{\Delta S109}$ – SEP3$^{\Delta 192}$	++	-	-	-	-
SEP1 – OsMADS3$^{\Delta S109}$	++	-	-	-	nt
SEP2 – OsMADS3$^{\Delta S109}$	++	-	-	-	nt
SEP3$^{\Delta 192}$ – OsMADS3$^{\Delta S109}$	++	-	-	-	nt
OsMADS3^{S109} – OsMADS8	++	-	-	-	-
OsMADS8 – OsMADS3^{S109}	++	++	++	++	++
OsMADS3$^{\Delta S109}$ – OsMADS8	++	-	-	-	-
OsMADS8 – OsMADS3$^{\Delta S109}$	+	-	-	-	-

Table 3. Summary of yeast-2-hybrid tests. In each couple, the first protein is fused to GAL4 activation domain (AD) and the second protein is fused to GAL4 binding domain (BD). The symbols indicate strong growth (++), mild growth (+) or no growth (-) of yeast colonies on the different media. Some interactions have not been tested (nt) on YSD -L-W-A.

occurred later during evolution. In particular, we provide strong evidence that it occurred after the differentiation between monocots and dicots and shortly before or after the WGD event which occurred in grasses. It looks like serine S109 has acquired an active and fundamental role in the functionality of OsMADS3-like proteins during evolution, thus selective pressure works to keep it in place.

Despite the majority of OsMADS3-like genes are annotated only for one of the two isoforms, we show that the alternative splicing occurs widely in grasses. In most cases, the additional amino acidic residue is Serine, or anyway an amino acid with a polar uncharged side chain. Then, the physical/chemical properties of this residue are likely to be important, on the contrary to Q173 in snapdragon[27].

The TAG duplication that we report in grasses, is a similar event of the CAG duplication of a splice acceptor in the snapdragon gene *FAR*, which results in an additional glutamine (Q173) in the last α-helix of the K-box[27]. However, there is an important difference between the two cases: S109 is not present in all the OsMADS3 peptides because both the (T)AG sequences can work as splice acceptor sites, whereas in snapdragon *FAR* only the 'new' (C)AG repetition seems to be recognized as splice acceptor site, whereas the conserved one seems to have lost its function[27]. To our knowledge, the alternative splicing that we reveal here is the first case, within MADS-box genes, of an alternative splicing which is highly conserved in a whole plant family and which causes a difference of just one amino acidic residue in the derived peptide.

The surprising level of conservation of the alternative splicing in grasses, suggests that *OsMADS3*^{S109} and *OsMADS3*$^{\Delta S109}$ are both translated into proteins having important specific functions. However, the results of our experiments are partly contradictory. First of all, mRNA-seq data show that the two alternative transcripts share the same expression profile, thus apparently excluding tissue- or stage-specific functions for them. Moreover, we were able to complement almost completely the *osmads3-3 osmads58* double mutant in *O. sativa* by constitutive expression of *OsMADS3*^{S109}, though the plants remained almost completely sterile (unpublished data), so *OsMADS3*$^{\Delta S109}$ might be not strictly necessary for flower development. However, we haven't got plants for the reciprocal constitutive expression of *OsMADS3*$^{\Delta S109}$.

The first α-helix is important for protein dimerization[53], and indeed our yeast-2-hybrid assay indicates that the lack of S109 greatly weakens the ability of OsMADS3$^{\Delta S109}$ to interact with the known SEP3-like partner OsMADS8, and also with SEP1 and SEP3 in *A. thaliana*.

The results we got by expressing *OsMADS3*^{S109} and *OsMADS3*$^{\Delta S109}$ in *A. thaliana*, however, seem to indicate the opposite: *OsMADS3*$^{\Delta S109}$ is clearly more functional in mimicking *AGAMOUS* function in Arabidopsis than *OsMADS3*^{S109}. Here, it is important to mention that heterologous expression cannot predict the real function of a gene in its original genetic and molecular context[54]. However, transcription factors can show exceptional affinity of some known target binding sites in heterologous systems, while losing partially or completely the affinity for most of the others. This is the case of LFY homologs versus the LFY target *AP1* in Arabidopsis[55].

It has already been demonstrated that the ectopic expression of an *AGAMOUS*-like gene in Arabidopsis can produce the typical C-overexpression phenotype (petal-to-stamen conversion, carpelloid sepals) also by activating the expression of the endogenous *AGAMOUS* and of the *SHATTERPROOF* genes. This is the case of the D-class gene *SEEDSTICK (STK)*: its overexpression produces phenotypes in Arabidopsis flower which are actually due to the ectopic activation of the *SHPs*[13,18].

Our hypothesis is that the inability of OsMADS3$^{\Delta S109}$ to interact with AGAMOUS partners such as the SEPs proteins that we tested, in Arabidopsis could actually be the key to its functionality. While both OsMADS3^{S109}and OsMADS3$^{\Delta S109}$ may retain the DNA-binding function (the MADS DNA-binding and C-terminal domains are unaltered in OsMADS3$^{\Delta S109}$), OsMADS3^{S109} is also able to interact with AGAMOUS partners like SEPs, while OsMADS3$^{\Delta S109}$ isn't. Furthermore, these complexes could be not as functional as they should, due to the fact that OsMADS3 is not their natural partner. The final effect would be that OsMADS3^{S109} subtracts SEP proteins and prevents the possibility of forming other functional complexes, for example with the SHPs, which are still present in our mutant lines. This would not happen in lines expressing *OsMADS3*$^{\Delta S109}$, which would only act as an activator but not as an interactor with SEP, thus leaving the SEP-SHP complex free to replace AGAMOUS in its C-function. Therefore, OsMADS3$^{\Delta S109}$ (possibly cooperating with other proteins) could at least be able to activate *SHPs*, which would be enough to explain its ability to partially rescue *ag-3* mutant. However, the dramatically higher expression of *REM22* in *OsMADS3*$^{\Delta S109}$ *ag-3* mutant background suggests that *OsMADS3*$^{\Delta S109}$ might be able to recognize more targets of AGAMOUS besides *SHPs*.

The difference in ability to interact with SEP1, SEP3 and OsMADS8 in yeast-2-hybrid assays may indicate that, in their native background, OsMADS3^{S109} and OsMADS3$^{\Delta S109}$ have different affinity to form different complexes. The five SEPs and two AGL6-like proteins which are present in rice are good candidates for this. The functional characterization of these two protein isoforms will need

further studies in rice to understand their roles in flower development and to explain their evolutionary conservation in grasses.

ACKNOWLEDGMENTS

We thank Anne Dievart for valuable suggestions about some of the experiments.
We also thank Stefano Ricagno for valuable discussion about protein structure alterations and Elisabetta Caporali for SEM images.

BIBLIOGRAPHY

1. Bowman, J. Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. *J Biosci.* **22,** 515–527 (1997).
2. Magallon, S., Gomez-Acevedo, S., Sanchez-Reyes, L. L. & Hernandez-Hernandez, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207,** 437–453 (2015).
3. Lombardo, F. & Yoshida, H. Interpreting lemma and palea homologies: a point of view from rice floral mutants. *Front. Plant Sci.* **6,** 1–6 (2015).
4. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early Flower Development in Arabídopsis. *Plant Cell* **2,** 755–767 (1990).
5. Yoshida, H. & Nagato, Y. Flower development in rice. *J. Exp. Bot.* **62,** 4719–30 (2011).
6. Dreni, L. & Kater, M. M. MADS reloaded: Evolution of the AGAMOUS subfamily genes. *New Phytol.* **201,** 717–732 (2014).
7. Yanofsky, M. F. *et al.* The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **346,** 35–39 (1990).
8. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405,** 200–203 (2000).
9. Theissen, G. Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **4,** 75–85 (2001).
10. Theissen, G. & Saedler, H. Floral quartets. *Nature* **409,** 469–471 (2001).
11. Zahn, L. M. *et al.* Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: Evidence of independent sub- and neofunctionalization events. *Evol. Dev.* **8,** 30–45 (2006).

12. Kramer, E. M., Jaramillo, M. A. & Di Stilio, V. S. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166,** 1011–23 (2004).
13. Pinyopich, A. *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424,** 85–88 (2003).
14. Heijmans, K. *et al.* Redefining C and D in the petunia ABC. *Plant Cell* **24,** 2305–17 (2012).
15. Brambilla, V. *et al.* Genetic and Molecular Interactions between BELL1 and MADS Box Factors Support Ovule Development in Arabidopsis. *Plant Cell Online* **19,** 2544–2556 (2007).
16. Dreni, L. *et al.* The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. *Plant J* **52,** 690–699 (2007).
17. Dreni, L. *et al.* Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. *Plant Cell* **23,** 2850–63 (2011).
18. Favaro, R. *et al.* MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* **15,** 2603–2611 (2003).
19. Hu, L. *et al.* Rice MADS3 Regulates ROS Homeostasis during Late Anther Development. *Plant Cell* **23,** 515–533 (2011).
20. Cui, R. *et al.* Functional conservation and diversification of class E floral homeotic genes in rice (Oryza sativa). *Plant J* **61,** 767–781 (2010).
21. Mena, M. *et al.* Diversification of C-function activity in maize flower development. *Science (80-.).* **274,** 1537–1540 (1996).
22. Schmidt, R. J. *et al.* Identification and molecular characterization of ZAG1, the maize homolog of the Arabidopsis floral homeotic gene AGAMOUS. *Plant Cell* **5,** 729–737 (1993).
23. Wei, B. *et al.* Genome-wide analysis of the MADS-Box gene family in Brachypodium distachyon. *PLoS One* **9,** (2014).
24. Druka, A. *et al.* An atlas of gene expression from seed to seed through barley development. *Funct. Integr. Genomics* **6,** 202–211 (2006).
25. Meguro, A., Takumi, S., Ogihara, Y. & Murai, K. WAG, a wheat AGAMOUS homolog, is associated wiht development of pistillike stamens in alloplasmic wheats. *Sex. Plant Reprod.* **15,** 221–230 (2003).
26. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell* **71,** 119–131 (1992).

27. Airoldi, C. a, Bergonzi, S. & Davies, B. Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18898–902 (2010).
28. Zhang, P., Tan, H. T. W., Pwee, K. H. & Kumar, P. P. Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* **37,** 566–577 (2004).
29. Martin, T. *et al.* PpAG1, a homolog of AGAMOUS, Expressed in Developing Peach Flowers and Fruit. *Can. J. Bot.* **84,** 767–776 (2006).
30. Rutledge, R. *et al.* Characterization of an AGAMOUS homologue from the conifer black spruce (Picea mariana) that produces floral homeotic conversions when expressed in Arabidopsis. *Plant J.* **15,** 625–634 (1998).
31. Kitahara, K. & Matsumoto, S. Rose MADS-box genes 'MASAKO C1 and D1' homologous to class C floral identity genes. *Plant Sci.* **151,** 121–134 (2000).
32. Davies, B. *et al.* PLENA and FARINELLI: redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. *EMBO J.* **18,** 4023–34 (1999).
33. Liu, Z., Zhang, D., Liu, D., Li, F. & Lu, H. Exon skipping of AGAMOUS homolog PrseAG in developing double flowers of Prunus lannesiana (Rosaceae). *Plant Cell Rep.* **32,** 227–237 (2013).
34. Zhang, B., Liu, Z. X., Ma, J., Song, Y. & Chen, F. J. Alternative splicing of the AGAMOUS orthologous gene in double flower of Magnolia stellata (Magnoliaceae). *Plant Sci.* **241,** 277–285 (2015).
35. Lightfoot, D. J., Malone, K. M., Timmis, J. N. & Orford, S. J. Evidence for alternative splicing of MADS-box transcripts in developing cotton fibre cells. *Mol. Genet. Genomics* **279,** 75–85 (2008).
36. Tsaftaris, A. S., Pasentsis, K. & Polidoros, A. N. Isolation of a differentially spliced C-type flower specific AG-like MADS-box gene from Crocus sativus and characterization of its expression. *Biol. Plant.* **49,** 499–504 (2005).
37. Hands, P., Vosnakis, N., Betts, D., Irish, V. F. & Drea, S. Alternate transcripts of a floral developmental regulator have both distinct and redundant functions in opium poppy. *Ann. Bot.* **107,** 1557–1566 (2011).
38. Kaufmann, K., Melzer, R. & Theißen, G. MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene* **347,** 183–198

(2005).

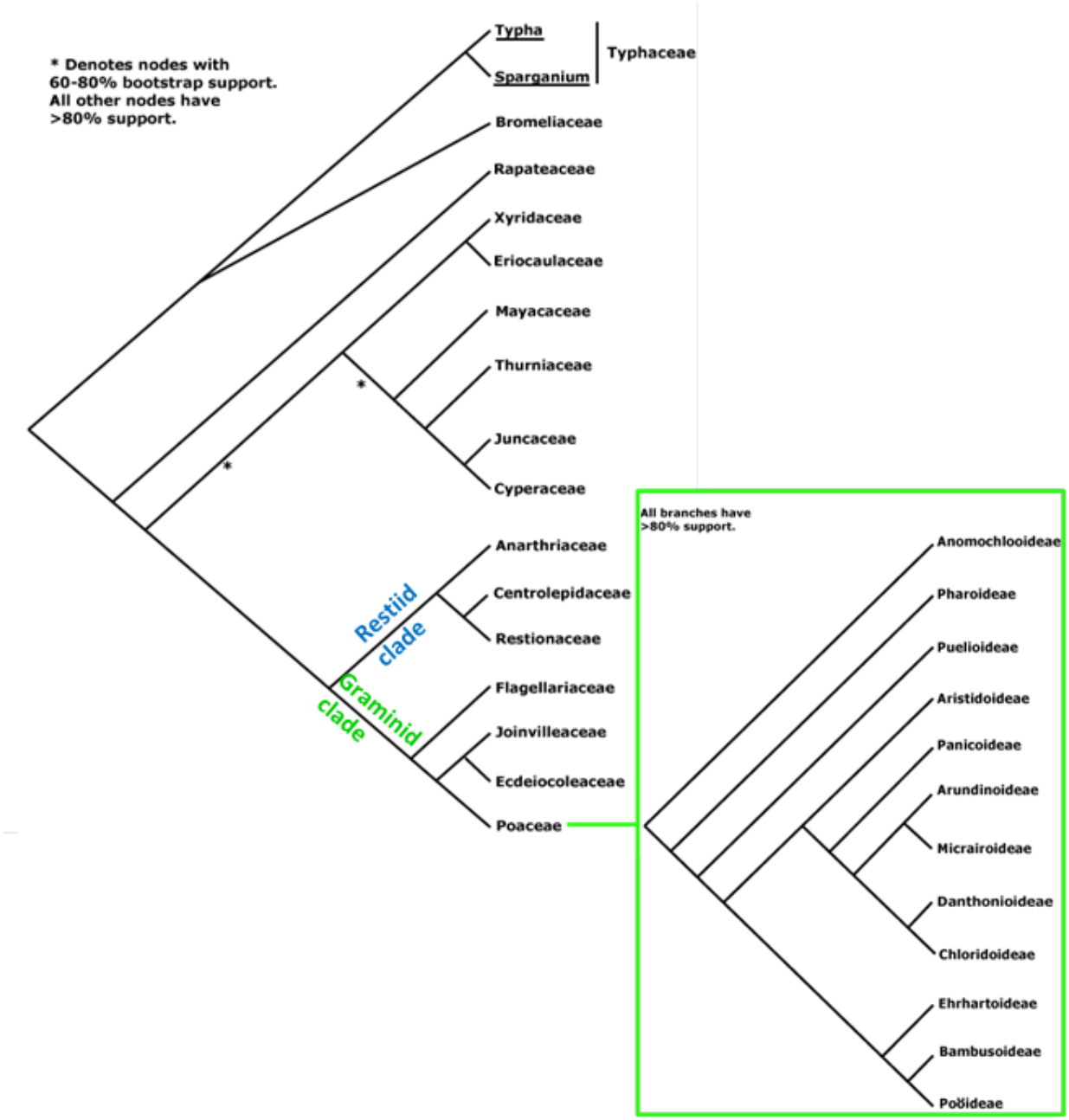
39. Johansen, B., Pedersen, L. B., Skipper, M. & Frederiksen, S. MADS-box gene evolution-structure and transcription patterns. *Mol. Phylogenet. Evol.* **23,** 458–480 (2002).
40. Riechmann, J. L., Ito, T. & Meyerowitz, E. M. Non-AUG initiation of AGAMOUS mRNA translation in *A. thaliana*. *Mol. Cell. Biol.* **19,** 8505–8512 (1999).
41. Yamaguchi, T. *et al.* Functional diversification of the two C-class MADS box genes OSMADS3 and OSMADS58 in Oryza sativa. *Plant Cell* **18,** 15–28 (2006).
42. Wei, S. Molecular cloning, characterization and expression of WAG-2 alternative splicing transcripts in developing spikes of Aegilops tauschii. *J. Genet.* **95,** 581–585 (2016).
43. Ming, R. *et al.* The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47,** 1435–1442 (2015).
44. Paterson, A. H., Bowers, J. E. & Chapman, B. A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci.* **101,** 9903–9908 (2004).
45. Chase, M. W. *et al.* An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181,** 1–20 (2016).
46. Bowman, J. L., Smyth, D. R. & Meyerowitzt, E. M. Genetic interactions among floral homeotic genes of Arabidopsis. *Development* **20,** 1–20 (1991).
47. Zhang, J., Nallamilli, B. R., Mujahid, H. & Peng, Z. OsMADS6 plays an essential role in endosperm nutrient accumulation and is subject to epigenetic regulation in rice (Oryza sativa). *Plant J.* **64,** 604–17 (2010).
48. Dreni, L., Osnato, M. & Kater, M. M. The Ins and Outs of the Rice AGAMOUS Subfamily. *Mol. Plant* **6,** 650–664 (2013).
49. Liu, X. *et al.* The SPOROCYTELESS/NOZZLE Gene Is Involved in Controlling Stamen Identity in Arabidopsis. *PLANT Physiol.* **151,** 1401–1411 (2009).
50. Romanel, E. *et al.* Reproductive meristem22 is a unique marker for the early stages of stamen development. *Int. J. Dev. Biol.* **55,** 657–664 (2011).
51. Alvarez, J. & Smyth, D. R. CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126,** 2377–2386 (1999).
52. Malcomber, S. T. & Kellogg, E. a. SEPALLATA gene diversification: brave new whorls. *Trends Plant Sci.* **10,** 427–435 (2005).
53. Puranik, S. *et al.* Structural Basis for the Oligomerization of the MADS Domain Transcription

Factor SEPALLATA3 in Arabidopsis. *Plant Cell* **26,** 3603–3615 (2014).

54. Kramer, E. M. A stranger in a strange land: the utility and interpretation of heterologous expression. *Front. Plant Sci.* **6,** (2015).

55. Maizel, A. *et al.* The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science (80-.).* **308,** 260–263 (2005).

SUPPLEMENTARY MATERIALS



Supplemental Figure 1. Phylogenetic tree of Poales.

CHAPTER 3: ROLE OF AGAMOUS-SUBFAMILY TRANSCRIPTION FACTORS DURING SEED DEVELOPMENT IN RICE

Ludovico Dreni, Andrea Ravasio, Martin M. Kater

ABSTRACT

AGAMOUS-subfamily genes have been extensively studied in the past 30 years for their central role in flower development. Their activity has mainly been associated with the C and D functions of the so-called ABC model for floral organs specification. However, a number of studies demonstrated that *AGAMOUS* genes have also a post-fertilization role and that they are directly involved, at various stages, in the correct development of seeds and fruits.

In rice (*Oryza sativa*), four members of the *AGAMOUS* subfamily are present, namely *OsMADS3, OsMADS13, OsMADS21* and *OsMADS58*. Their function in flower development has been deeply studied, however, little is known about their role in seed development. The expression profile of these four *AGAMOUS* genes suggests an active role in seed development since all of them are expressed in the seed coat and in the vasculature starting from the first developmental stages. Multiple mutants in these genes display severe defects in seed development. Fertility is also impaired, with a high number of seeds aborting in the initial developmental phases.

INTRODUCTION

Rice seed development

Flowers of the Graminaceae have a typical structure called spikelet, in which the flower is surrounded by bract-like organs called glumes. In rice, the flower is subtended by four glumes, two rudimentary glumes and two sterile glumes. Two more glumes, the lemma and the palea, surround the other organs: two lodicules (homologues of the petals), six stamens and a monocarpellary pistil.

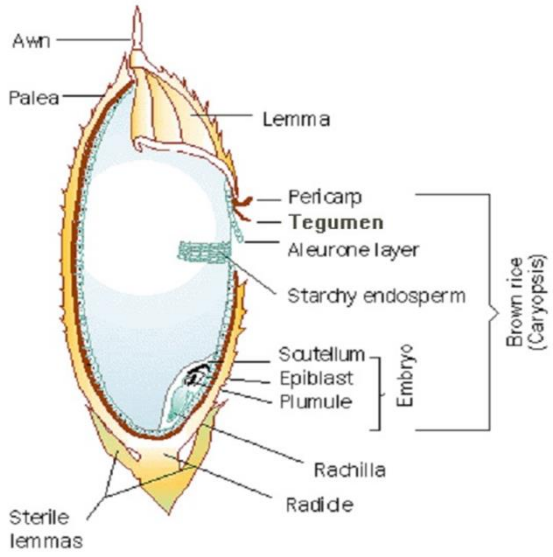


Figure 1. Schematic representation of a rice seed.

Rice spikelets, like in other Graminaceae, produce a dry indehiscent fruit called caryopsis, which develops one single seed. Most of the volume in rice seeds is occupied by the starchy endosperm, enclosed in a seed coat composed of different layers: from outside: the pericarp, the tegmentum and the aleurone layer. At the basis of the seed lays the embryo.

After spikelet anthesis and pollination, a double fertilization event takes place inside the ovary: one of the two sperm cells released by the pollen tube fertilizes the egg cell, generating the zygote; the other sperm cell fuses with the polar nuclei, generating the triploid nucleus which will give rise to the endosperm. The first day after pollination (DAP), ovary, embryo and endosperm simultaneously start their development[1].

The ovary elongates and then expands inside the glumes, completely filling the inner space, in about 30 days. Maturation is completed around 45 DAP[1].

The ovary wall, right after the anthesis, is composed of different layers: outer epidermis, various central layers, inner epidermis and outer and inner integument of the ovule. During seed maturation, all of these layers undergo changes. Cells of the outer epidermis multiply and elongates vertically; cells of the inner central layers elongates laterally, at a right angle respect to the seed vertical axis (for this reason they are called cross cells); cells of the inner epidermis also elongate vertically, then they lignify (they are called tube cells because they have a tube shape if observed longitudinally). The ovule outer integument degenerates, while the inner integument cells divide and elongate, developing into a transparent layer called testa, which later on fuses with the esosperm and becomes part of the seed coat.

When the zygote begins its mitotic divisions, the bud primordium appears on the ventral side and the root primordium on the other side. Between 5 and 10 DAP, three leaf primordia differentiate and the vasculature connecting the embryo and the scutellum is completed. After 12 DAP, cell division stops and the third leaf primordium arrests its development. The embryo reaches complete formation around 25 DAP[1].

After fertilization the primary triploid nucleus of the endosperm starts dividing and the first cells cover the inner wall of the embryo sac in a few hours. After the formation of cell walls around these first cells (approximately 3 DAP), new cells continue to form, filling the inner space of the embryo sac. Cell division continues until about 30 DAP. After that, the outer endosperm layer differentiates into the aleurone layer, while the inner cells accumulate starch. Cells of the aleurone layer do not accumulate starch, but protein granules and lipids. The aleurone layer main function is to synthesize enzymes that can hydrolyse the starch content of the inner cells when seed germination takes place[1].

Also lemma and palea are modified during seed development, becoming thicker and lignified.

Role of AGAMOUS subfamily genes

AGAMOUS-like genes are involved in floral organ formation and in various processes concerning flower development before fertilization. This is true also in the case of rice *AGAMOUS*-like genes *OsMADS3, OsMADS13, OsMADS21* and *OsMADS58*. However, as shown by the analyses of Arora et al. (2007)[2], the rice *AGAMOUS* genes continues to be expressed even when flower development has terminated and fertilization has occurred. In particular, *OsMADS3* and *OsMADS13*, whose expression increases throughout the whole panicle development, are expressed at even higher levels during the first stages of seed development, then their expression start to decrease in later stages[2]. *OsMADS58* expression profile follows that of other MADS-box genes (like the E-class genes *OsMADS1, OsMADS5, OsMADS7, OsMADS8* and the *AGL6*-like gene *OsMADS6*) which show an almost steady level of expression from panicle development to the late stages of seed development, where their expression level decreases[2]. Even *OsMADS21* expression, which is almost absent during panicle development, increases right after fertilization, reaches a peak during mid stages of seed development and then decreases again[2]. This is surprising, since *OsMADS21*, despite being a paralogue of *OsMADS13*, seems to have no function in ovule development and was considered due to its very low expression in the ovule more like a pseudogene[3,4].

These expression profiles have been further confirmed by RNA-Seq and *in situ* hybridization experiments performed in our laboratory. From RNA-Seq data emerges that all four genes have their peak of expression immediately after fertilization has occurred (0 DAF), then their expression levels decrease.

From *in situ* experiments previously performed in our laboratory it emerges that the expression of all four genes goes in parallel throughout the whole seed development (Figure 2). During the first stages, the signal is restricted to seed coat and dorsal vasculature: *OsMADS3* and *OsMADS58* also show a faint expression signal in the pericarp, which soon disappears (Figure 2A,C,E,G); at 8 days after fertilization (DAF), the signal is much stronger in the dorsal vasculature than in the seed coat (Figure 2B,D,F,H); by 15 DAF, the signal disappears completely from the seed coat and it is still faintly visible in the dorsal vasculature (data not shown).

These observations, taken together, strongly suggest the possibility that the four genes have an active role in rice seed development.

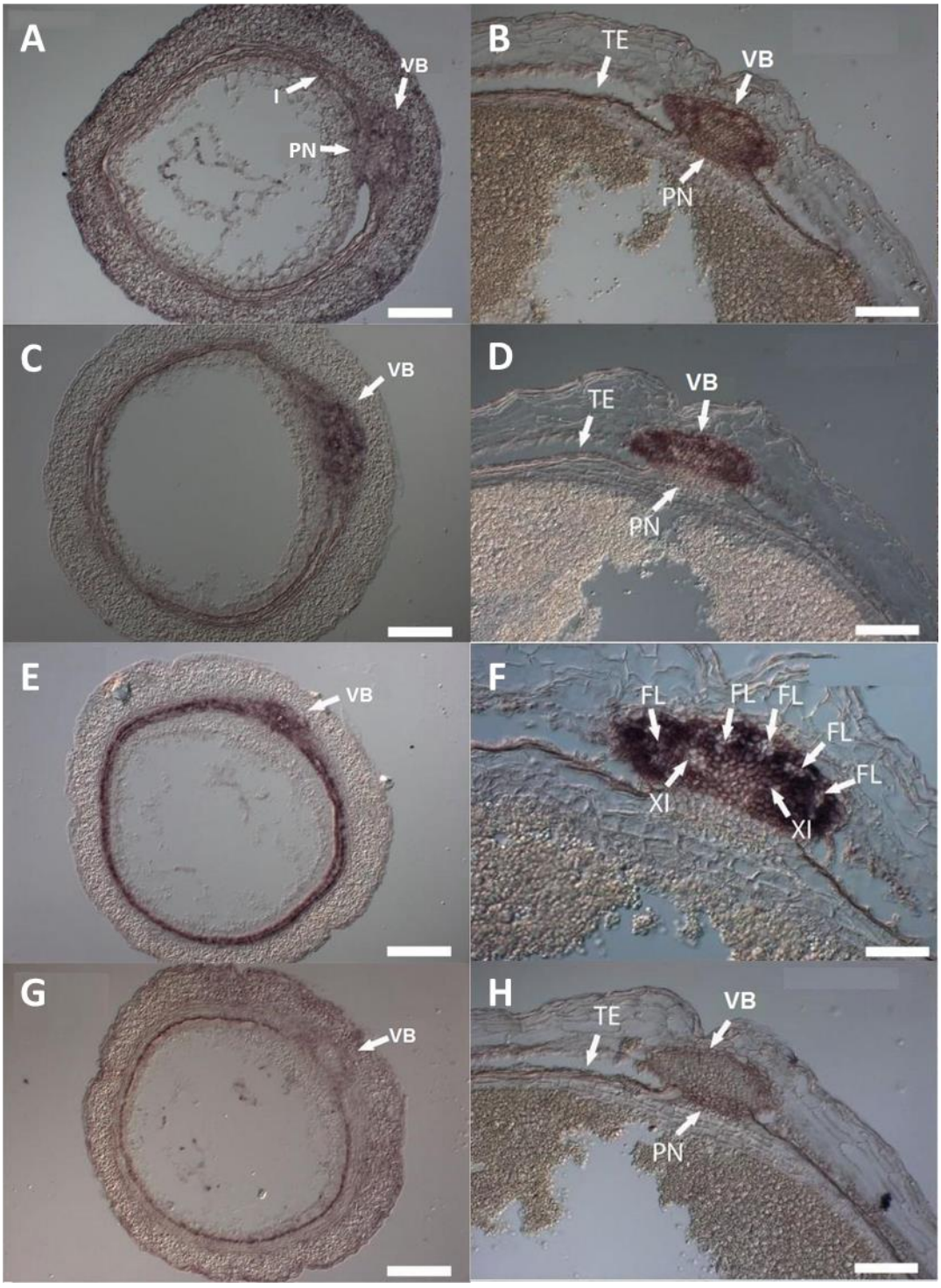


Figure 2. In situ hybridization of OsMADS3 (A-B), OsMADS13 (C-D), OsMADS21 (E-F) and OsMADS58 (G-H) in rice ovules at 2 DAP (left) and 8 DAP (right). Signal of the four transcripts is visible in the integument and in the vascular bundle.

Legenda: FL: phloem; I: integument; PN: Nucellar projection; TE: testa; VB: vascular bundle; XI: xylem. Reference bar: 100 µm (50 µm in F)

RESULTS

Generation of multiple mutants

In order to better understand the role of rice *AGAMOUS-like* genes during seed development, various combinations of double, triple and quadruple mutants were generated and analysed.

Single mutants for *OsMADS21* and *OsMADS58* do not show any defect in flower or seed development[4,5]. Seeds of single mutants for *OsMADS3* and *OsMADS13* cannot be analysed because these mutants are completely sterile due to defects in reproductive organs; however, the heterozygous lines *osmads3-3/+* and *osmads13/+* have a normal fertility and produce normal seeds[3–5]. Thus, we hypothesized that the four genes could act redundantly during seed development; this is often the case in Graminaceae, where a duplication event led to the presence of multiple paralogs for many MADS-box genes[6], allowing for large redundancy, subfunctionalization and neofunctionalization.

Based on this hypothesis, we decided to generate and analyse multiple mutant combinations of the four genes. Single mutants for *OsMADS3, OsMADS13*, *OsMADS21* and *OsMADS58* were already available in our laboratory[3–5]; the multiple mutants were generated by crossing and subsequent analysis of the segregating progeny (see **Table 4**).

It is important to point out that, in all combinations analysed, *OsMADS3* and *OsMADS13* were always kept in a WT or heterozygous state. This is because while *osmads21* and *osmads58* single mutants show no defect in flower or seed development, *osmads3-3* and *osmads13* single mutants

NAME	GENOTYPE
Double mutant 1	*OsMADS3 osmads13/+ OsMADS21 osmads58*
Double mutant 2	*osmads3-3/+ OsMADS13 OsMADS21 osmads58*
Triple mutant 1	*OsMADS3 osmads13/+ osmads21 osmads58*
Triple mutant 2	*osmads3-3/+ OsMADS13 osmads21 osmads58*
Quadruple mutant	*osmads3-3/+ osmads13/+ osmads21 osmads58*

Table 4. Genotypes of mutant lines.

are completely male and female sterile[4]; the heterozygous lines *osmads3-3/+* and *osmads13/+*, on the contrary, are fully capable of producing fertile seeds, so at least one functional copy of each of the two genes has to be maintained to be able to produce seeds.

Analysis of seeds from mutant lines

Triple and quadruple mutants show a decrease in seed weight

The first analysis we performed was a comparison of the average seed weight among the different mutant lines. A minimum of 100 seeds from at least two independent individuals for each mutant line were collected, cleaned from the glumes and individually weighed.

The results of the analysis are shown in Figure 3. The average weight of a WT seed is around 190 mg. Seed from double mutant lines do not differ much from WT seed, in terms of weight. In the triple mutant lines, seed weight is reduced by 25 – 30% (average of about 130 – 150 mg), while in the quadruple mutant the reduction is more severe, around 50% (average of 80 – 90 mg). The seeds also appear visibly smaller and defective (Figure 4).

Seed abortion in mutant lines

Another aspect we decided to take into consideration was the fertility of the different lines.

The abortion rate of each line was measured as a ratio between the number of aborted seeds and the total number of seeds produced by the plant (Figure 5).

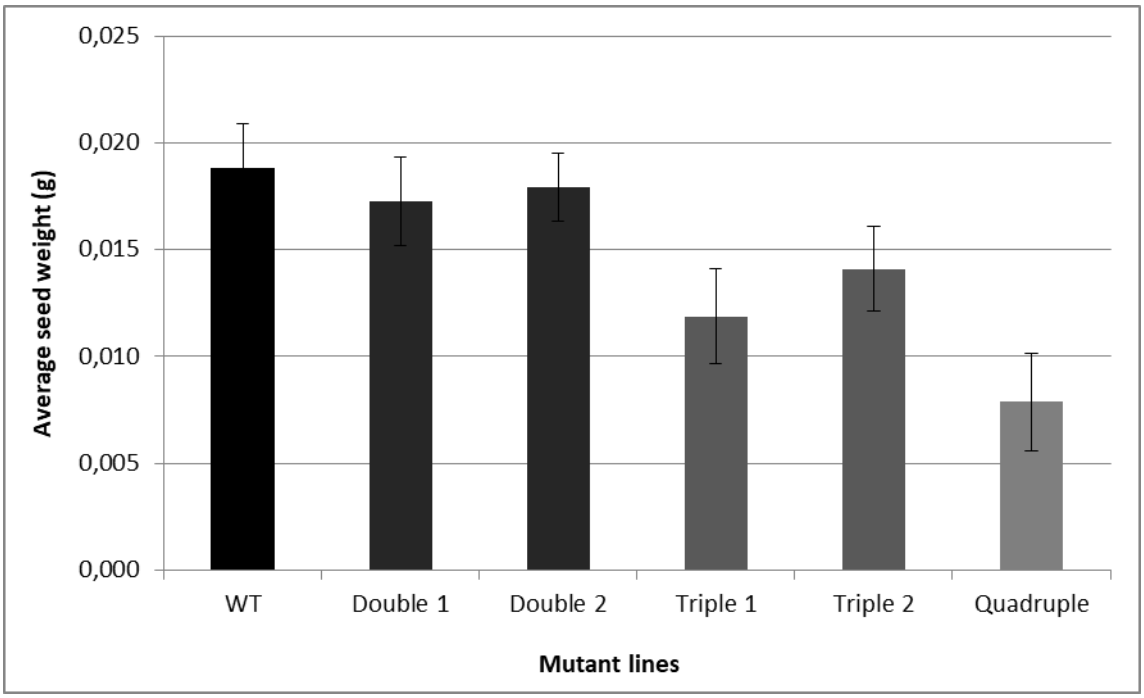


Figure 3. Average seed weight of the «double», «triple» and «quadruple» mutant lines, compared to WT (see Table 4 **for genotype details). Standard deviation is indicated.**

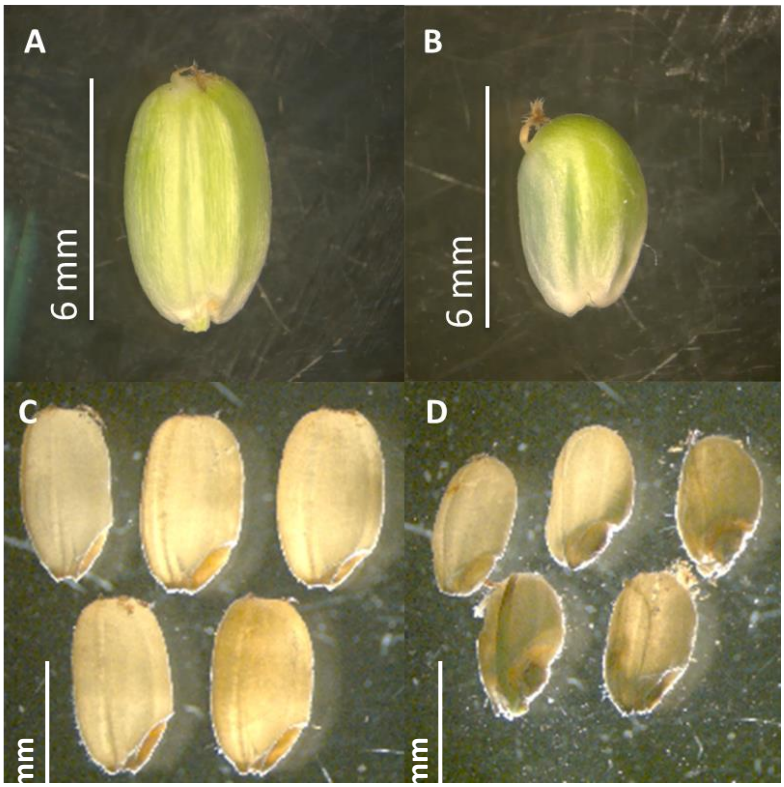


Figure 4. Comparison of WT (left) and "quadruple" mutant (right) seeds at 10 DAP (A-B) and at complete maturation (C-D).

The abortion rate of the triple mutants is around 15-20%, which is not particularly severe, but higher than that of the WT (and double mutant) lines, which in our growth conditions is very low (around 5%). The quadruple mutant lines show once again the most severe phenotype, with an abortion rate of about 50% (Figure 5).

We decided to investigate also if the abortion was occurring randomly or in any particular stage during seed development. We divided the aborted seeds into three categories,

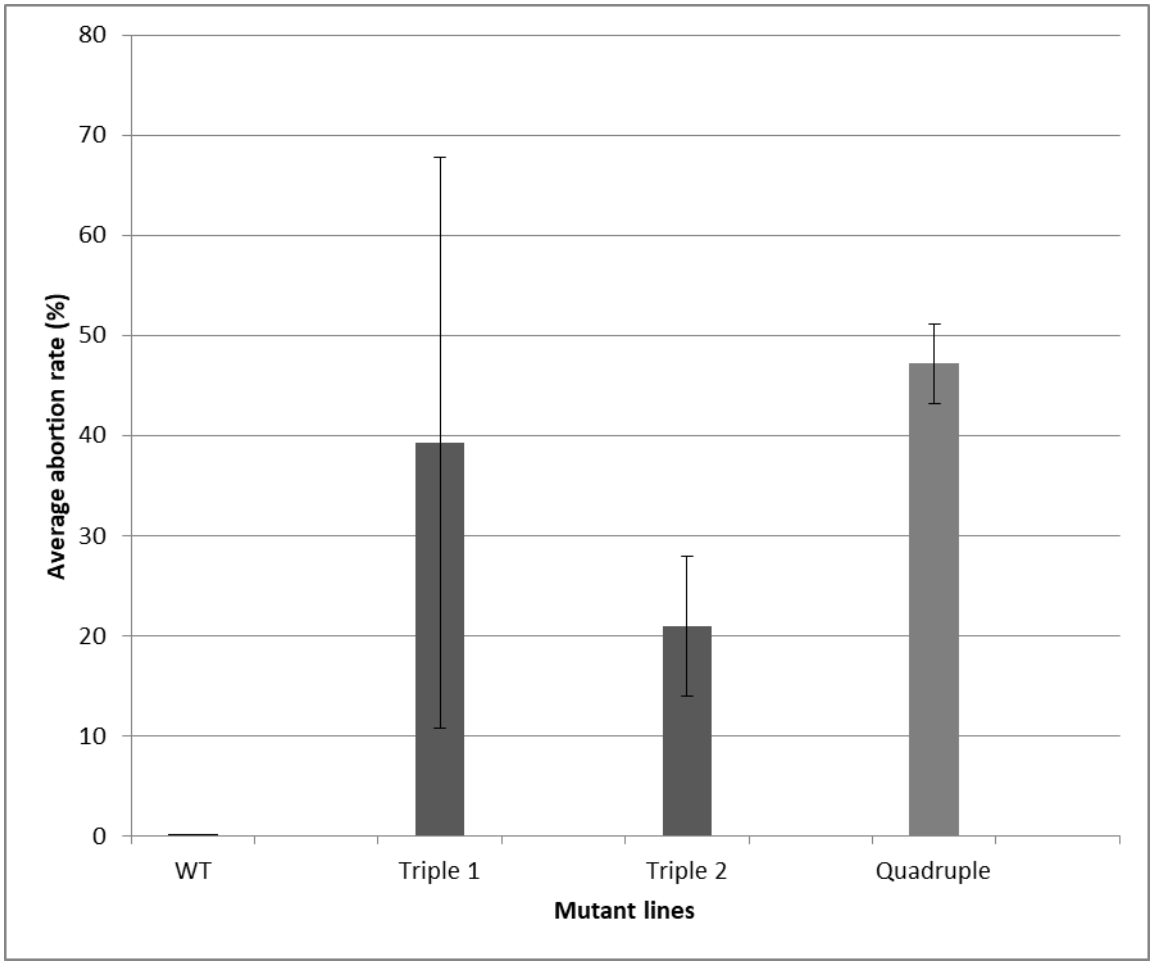


Figure 5. Average abortion rate of «triple» and «quadruple» mutants, compared to WT (see Table 4 **for genotype details). Standard deviation is indicated.**

Figure 6. A) Ovaries at 0 DAF (left) and 1 DAF (right); B) ovaries at 2 DAF (left) and 3 DAF (right); C) ovaries at 4 DAF (left) and 5 DAF (right). These stages have been used as reference to classify "non fertilized" (A), "early aborted" (B) and "late aborted" (C) seeds, respectively.

based on the appearance of the ovary compared to the seed developmental stages[1]. The "non-fertilized" (NF) seeds are those where the ovary does not even start to develop and remains small, at the base of the glumes; there's a high chance that these seeds have not been pollinated (Figure 6A). The "early aborted" (EA) seeds are those which reached at least half of the glume length before their development was arrested (Figure 6B). The "late aborted" (LA) seeds are those where the development stopped after the ovary has reached the full length of the glumes and has already started the filling phase (Figure 6C). The analysis revealed that, while the few aborted seeds from WT lines fall in the NF category, in the mutant lines there is a clear increase of seeds aborted in later stages, especially those we call early aborted (EA) (Figure 7). This can be considered an indirect proof of the active role of rice *AGAMOUS-like* genes during the first stages of seed development.

Yeast-2-hybrid assays

Hypothesizing the formation of higher order complexes

MADS-domain factors, particularly those involved in flower development, are known to function in higher order MADS-complexes[7]. We hypothesized protein-protein interaction could also take place among rice *AGAMOUS*-like genes during their active role in seed development. Since OsMADS3 and OsMADS13 are not able to physically interact *in vivo*[8], we hypothesized that other factors could be involved in complex formation.

OsMADS7 and *OsMADS8*, also known as *OsMADS45* and *OsMADS24*[9], are the rice orthologs of *SEP3*, an E-class gene from *A. thaliana*. Proteins encoded by these two genes are known to be able to interact with both OsMADS3 and OsMADS13, similar as has been described for their orthologs SEP3, AGAMOUS and SEEDSTICK in Arabidopsis[10–13].

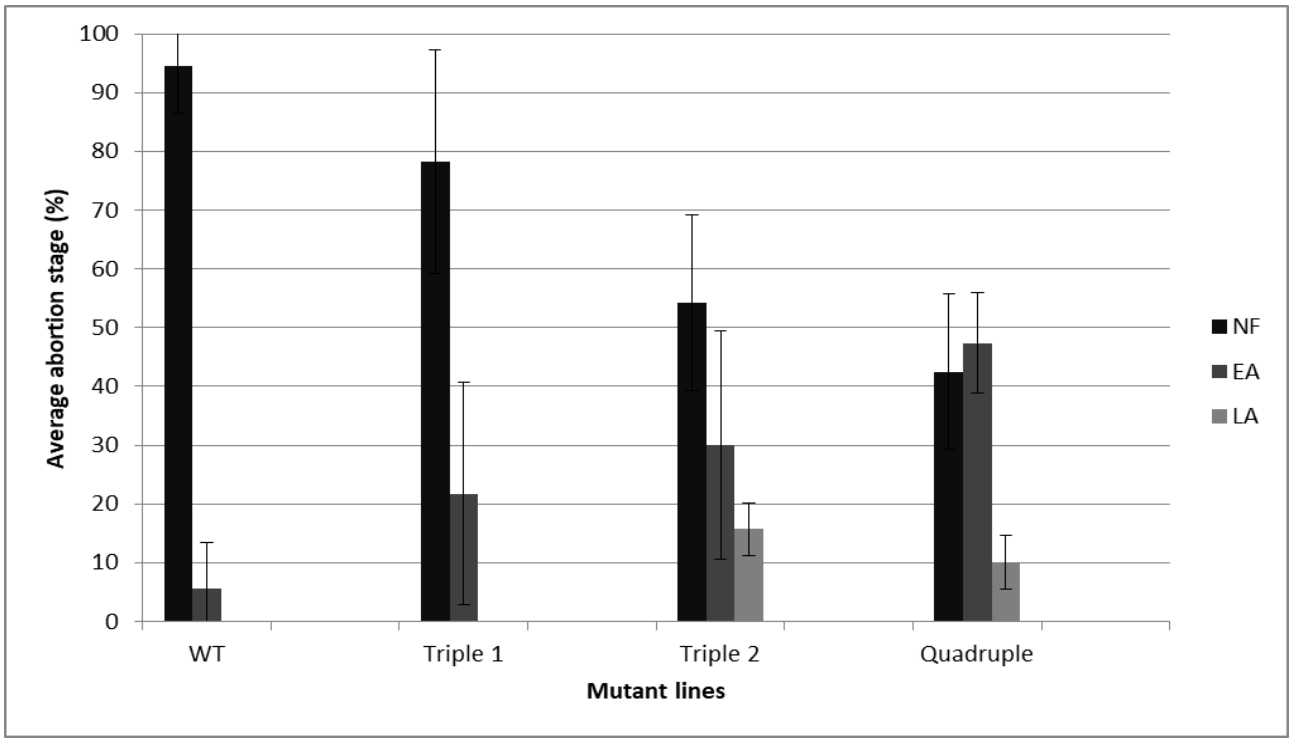


Figure 7. Average number of aborted seeds per line, divided by the categories of «non fertilized» (NF), «early aborted» (EA) and «late aborted» (LA), expressed as percentages of the total number of aborted seeds, compared to WT (see Table 4 **for genotype details). Standard deviation among replicates is indicated.**

OsMADS6 belongs to the *AGL6* gene family, which is a sister clade of the *SEP* family[14–16]. As reported by Zhang *et al.* (2010), *OsMADS6* expression is detectable in flower, embryo and endosperm tissues[17]. Mutations in *OsMADS6* produce defects in all four floral whorls and alteration in organ number and morphology, supporting the hypothesis that *OsMADS6*, besides being phylogenetically related to the *SEP* family, also retains some *SEPALLATA*-like functions[17,18]. Another interesting aspect to consider is that *osmads6* mutant alleles display various defects in seed development, particularly a high rate of abortion and the production of smaller seeds due to a reduction of about 25% in starch filling[17].

OsMADS29 belongs to the B_{sister} family, which is a sister clade of the B class[19]. Expression of *OsMADS29* is already detectable during panicle development, but is particularly strong in developing ovules and developing seeds[20]. Again, *osmads29* mutants show severe defects in seed development and starch filling; in this case, the reason is a delayed development of mutant seeds compared to WT seeds and multiple defects in the programmed maternal tissues degeneration and in the vasculature formation, both fundamental for the correct intake of nutrients by the developing seed[20,21]. Moreover, it has already been proven that in Arabidopsis the B_{sister} gene *ABS* can interact with *SEEDSTICK* during fertilization and seed development[22].

All of this information taken together and considering the similarities between the seed-related phenotypes of *OsMADS6* and *OsMADS29* and those of our multiple mutants, we hypothesized that OsMADS6*,* OsMADS7*,* OsMADS8 and OsMADS29 could be good candidates to cooperate with the *AGAMOUS*-like factors in protein complexes formation during seed development.

Yeast screening confirmed binary interactions, but could not proof higher order complex formation

To test our hypothesis, we cloned the CDS of *OsMADS3, OsMADS13, OsMADS6, OsMADS8* and *OsMADS29* into the plasmids for yeast-2-hybrid, 3-hybrid and 4-hybrid assays.

Our experiments initially confirmed the following interactions: OsMADS13-OsMADS6, OsMADS13-OsMADS8, and OsMADS3-OsMADS8, which were already published in literature[23,24]. In addition, we showed the interactions, OsMADS29-OsMADS3 and OsMADS6-OsMADS8; no interaction between OsMADS13 and OsMADS29 could be detected. Our results were confirmed by a paper published by Nayar *et al.* (2014)[25] (see Table 5).

In our experiments, we also tested the possibility of a ternary or quaternary complex formation by means of yeast-3-hybrid and yeast-4-hybrid assays, but none of our tests could actually prove the formation of a higher order complex because of an auto-activation problem: it seems that

	YSD -L-W-H	YSD -L-W-H + 3AT 1mM	YSD -L-W-H + 3AT 2.5mM	YSD -L-W-H + 3AT 5mM
OsMADS6 – OsMADS3	++	-	-	-
OsMADS6 – OsMADS13	++	++	++	++
OsMADS6 – OsMADS8	++	++	++	++
OsMADS8 – OsMADS3	++	++	++	++
OsMADS8 – OsMADS13	++	++	++	++
OsMADS29 – OsMADS13	++	-	-	-
OsMADS29 – OsMADS3	++	++	++	++
OsMADS6 – OsMADS29	++	+	+	+
OsMADS8 – OsMADS29	++	++	++	++

Table 5 (previous page). Summary of yeast-2-hybrid tests. In each couple, the first protein is fused to GAL4 activation domain (AD) and the second protein is fused to GAL4 binding domain (BD). The symbols indicate strong growth (++), mild growth (+) or no growth (-) of yeast colonies on the different media.

OsMADS6, when used as a "bridge" (meaning that it is not fused to GAL4 AD or BD domain, but only to a nuclear localization signal), is able to activate transcription even in the absence of a protein fused to the GAL4 AD domain (see Table 6).

pGADT7	pTFT	pRED	pGBKT7	YSD - LWAH	YSD – LWAH + 3AT 1mM	YSD – LWAH + 3AT 2,5mM	YSD – LWAH + 3AT 5mM
OsMADS13	**OsMADS6**		**OsMADS29**	**++**	**-**	**-**	**-**
(empty)	OsMADS6		OsMADS29	++	+	+	-
OsMADS13	(empty)		OsMADS29	++	-	-	-
OsMADS13	OsMADS6		(empty)	++	-	-	-
OsMADS29	**OsMADS6**		**OsMADS13**	**++**	**-**	**-**	**-**
(empty)	OsMADS6		OsMADS13	++	+	+	+
OsMADS29	(empty)		OsMADS13	++	-	-	-
OsMADS29	OsMADS6		(empty)	++	-	-	-
OsMADS13	**OsMADS3**	**OsMADS6**	**OsMADS29**	**++**	**+**	**-**	**-**
OsMADS13	OsMADS3	OsMADS6	(empty)	++	-	-	-
OsMADS13	OsMADS3	(empty)	OsMADS29	++	-	-	-
OsMADS13	(empty)	OsMADS6	OsMADS29	++	+	+	+
(empty)	OsMADS3	OsMADS6	OsMADS29	++	+	-	-
OsMADS29	**OsMADS3**	**OsMADS6**	**OsMADS13**	**++**	**+**	**+**	**+**
OsMADS29	OsMADS3	OsMADS6	(empty)	++	-	-	-
OsMADS29	OsMADS3	(empty)	OsMADS13	++	-	-	-
OsMADS29	(empty)	OsMADS6	OsMADS13	++	+	+	+
(empty)	OsMADS3	OsMADS6	OsMADS13	++	+	+	+

Table 6. Summary of yeast-3-hybrid (blue) and yeast-4-hybrid (green) tests. The symbols indicate strong growth (++), mild growth (+) or no growth (-) of yeast colonies on the different media.

DISCUSSION

The spatial and temporal expression pattern of rice *AGAMOUS*-like genes in seed already hinted towards a possible active role of these genes in supporting the seed development. The analysis on fertility and seed weight of the "triple" and "quadruple" mutants gave further proof of the actual existence of this role. The experiments demonstrated that, whatever function the four genes have in seed development, they act in large redundancy with each other, since the most severe phenotype can be observed only when three or more genes are in homozygous or heterozygous state. Moreover, the multiple mutant lines show a higher percentage of seeds whose development was arrested at intermediate stages, compared to the WT, where the few aborted seeds mainly result from a failed fertilization, thus suggesting that the *AGAMOUS*-like genes are actively needed in the initial phases of seed development.

Yeast-2-hybrid experiments performed in our laboratory and also published in literature[25] confirm the interaction between the AGAMOUS proteins and other MADS-domain factors potentially involved in seed development, like OsMADS6 and OsMADS29. The various combinations of dimers also suggest the possibility that higher order complexes could be formed, despite we weren't able to obtain clear evidences in this sense. Experiments in *A. thaliana* strongly support this hypothesis. STK and ABS (ortholog of OsMADS29) were reported to act cooperatively (also by physical interaction) during seed integuments development[22,26]. Recently, also SHP1 and SHP2 have been reported to have a role in post-fertilization seed development, by acting antagonistically to ABS in the control of cell division in the integuments and by ensuring the synchronization between ovule and embryo development[27].

Preliminary results obtained by dr. Ludovico Dreni, not reported in this chapter, demonstrated that the cutin layer which covers the rice seed is deposited earlier in the multiple mutants and this could affect the molecular cross talk and the exchange of nutrients, which are fundamental in the early developmental stages. Moreover, the dorsal vasculature structure is also affected, thus creating an even stronger interference in the initial nutrient uptake (data not shown).

RNA-Seq analyses performed by dr. Dreni on 2 DAF quadruple mutants reveal deregulation in genes related to cell wall formation and lipid metabolism. Surprisingly, the B class MADS-box genes *OsMADS2* and *OsMADS16*/*SUPERWOMAN (SPW)* and the E-class gene *OsMADS1* appear upregulated in the mutant; a more accurate analysis displayed that *spw1-1* mutant, besides being male sterile, is also female sterile and shows embryo abortion and that *osmads1* mutant seeds

display a reduction in seed thickness, whereas our mutants mainly display a reduction in length. These observations point towards an involvement of these genes in seed development.

Having a precise idea of the processes leading to seed formation has a great economical value: on one side, it could be a starting point for yield improvement, especially in crop plants, where the "commercial" part is the seed itself; but on the other side, it is also a starting point to understand how to interfere with seed development, especially in fleshy fruit cultivars, where it is generally preferable to favour fruit dimension and fleshy pulp content respect to seeds. In oil palm *(Elaeis guineensis*) a mutation in the *STK* ortholog *SHELL* has been associated with a reduction in the coconut kernel, resulting in a fruit with a higher oil content[28]. Many MADS-box genes have been found to be differentially expressed in seeded and seedless varieties of grapevine *(Vitis vinifera)*[29]; in particular, structural differences in the regulatory regions of the *STK* ortholog *VviAGL11* have been found comparing a seedless grapevine (Thompson seedless) with a seeded variety (Sultanine Monococco)[30,31]. The same effect has been found suppressing the *SlyAGL11* gene in tomato[31]. Moreover, the tomato AGL6-like gene *SlAGL6* has been associated with parthenocarpy (fertilization-independent fruit production, giving rise to seedless fruits)[32].

These are just some examples of how important it is to unravel the pathways leading to seed and fruit formation, in scientific but also agro-economical terms. Functional analyses on a broadly diffused model species like *Oryza sativa* could pave the way to better understand how homologous genes could act in fruit and seed development in other, even more interesting and agronomically important species.

BIBLIOGRAPHY

1. Hoshikawa, K. *The growing rice plant. An anatomical monograph*. (1989).
2. Arora, R. *et al.* MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* **8,** 242 (2007).
3. Dreni, L. *et al.* The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. *Plant J* **52,** 690–699 (2007).
4. Dreni, L. *et al.* Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. *Plant Cell* **23,** 2850–63 (2011).

5. Kater, M. M., Dreni, L. & Colombo, L. Functional conservation of MADS-box factors controlling floral organ identity in rice and Arabidopsis. *J. Exp. Bot.* **57,** 3433–44 (2006).
6. Kramer, E. M., Jaramillo, M. A. & Di Stilio, V. S. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166,** 1011–23 (2004).
7. Theissen, G. & Saedler, H. Floral quartets. *Nature* **409,** 469–471 (2001).
8. Li, H., Liang, W., Yin, C., Zhu, L. & Zhang, D. Genetic Interaction of OsMADS3, DROOPING LEAF, and OsMADS13 in Specifying Rice Floral Organ Identities and Meristem Determinacy. *Plant Physiol.* **156,** 263–274 (2011).
9. Greco, R. *et al.* MADS box genes expressed in developing inflorescences of rice and sorghum. *Mol. Gen. Genet.* **253,** 615–623 (1997).
10. Zahn, L. M. *et al.* The evolution of the SEPALLATA subfamily of MADS-box genes: A preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169,** 2209–2223 (2005).
11. Malcomber, S. T. & Kellogg, E. a. SEPALLATA gene diversification: brave new whorls. *Trends Plant Sci.* **10,** 427–435 (2005).
12. Honma, T. & Goto, K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409,** 525–529 (2001).
13. Pelaz, S., Tapia-López, R., Alvarez-Buylla, E. R. & Yanofsky, M. F. Conversion of leaves into petals in Arabidopsis. *Curr. Biol.* **11,** 182–184 (2001).
14. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5,** 484–495 (1991).
15. Reinheimer, R. & Kellogg, E. A. Evolution of AGL6-like MADS Box Genes in Grasses (Poaceae): Ovule Expression Is Ancient and Palea Expression Is New. *Plant Cell Online* **21,** 2591–2605 (2009).
16. Dreni, L. & Zhang, D. Flower development: The evolutionary history and functions of the AGL6 subfamily MADS-box genes. *J. Exp. Bot.* **67,** 1625–1638 (2016).
17. Zhang, J., Nallamilli, B. R., Mujahid, H. & Peng, Z. OsMADS6 plays an essential role in endosperm nutrient accumulation and is subject to epigenetic regulation in rice (Oryza sativa). *Plant J.* **64,** 604–17 (2010).
18. Moon, Y. H. *et al.* Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. *Plant Physiol.*

120, 1193–1204 (1999).

19. Becker, A. *et al.* A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics* **266,** 942–950 (2001).
20. Yang, X. *et al.* 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 (2012).
21. Yin, L.-L. & Xue, H.-W. The MADS29 Transcription Factor Regulates the Degradation of the Nucellus and the Nucellar Projection during Rice Seed Development. *Plant Cell* **24,** 1049–1065 (2012).
22. Mizzotti, C. *et al.* The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J.* **70,** 409–20 (2012).
23. Favaro, R. *et al.* Ovule-specific MADS-box proteins have conserved protein-protein interactions in monocot and dicot plants. *Mol. Genet. Genomics* **268,** 152–159 (2002).
24. Cooper, B. *et al.* A network of rice genes associated with stress response and seed development. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 4945–50 (2003).
25. Nayar, S., Kapoor, M. & Kapoor, S. 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 (2014).
26. Coen, O. *et al.* Developmental patterning of the sub-epidermal integument cell layer in *Arabidopsis* seeds. *Development* dev.146274 (2017). doi:10.1242/dev.146274
27. Ehlers, K. *et al.* The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and Seed coat development and for endosperm formation in *A. thaliana*. *PLoS One* **11,** 1–29 (2016).
28. Singh, R. *et al.* The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500,** 340–4 (2013).
29. Wang, L. *et al.* Evolutionary and expression analysis of a MADS-box gene superfamily involved in ovule development of seeded and seedless grapevines. *Mol. Genet. Genomics* **290,** 825–846 (2014).
30. Malabarba, J. *et al.* The MADS-box gene Agamous-like 11 is essential for seed morphogenesis in grapevine. *J. Exp. Bot.* **68,** 1493–1506 (2017).
31. Ocarez, N. & Mejia, N. Suppression of the D-class MADS-box AGL11 gene triggers seedlessness in fleshy fruits. *Plant Cell Rep.* **35,** 239–254 (2016).

32. Klap, C. *et al.* Tomato facultative parthenocarpy results from SlAGAMOUS-LIKE 6 loss of function. *Plant Biotechnol. J.* **15,** 634–647 (2017).

CHAPTER 4: CONCLUSIONS

Rice and other cereals are of enormous agricultural and economical importance because they represent a primary food source all around the globe and especially in developing regions like Eastern Asia. Thus, one of the main challenges, in a world where the population has reached 7,5 billion people and land and water available for agriculture decreases day by day, is to increase yield/ha in a sustainable way. Therefore, crop improvement is drastically needed in a short time period. Crop improvement by classical breeding approaches is rather slow and can't keep pace with the world food demand. Therefore, our crops will have to be improved through accelerated breeding methods by taking advantage of the latest technologies that are available .

Advancements in sequencing technologies and development of new techniques and protocols for site-directed mutagenesis are greatly expanding our knowledge about the evolution of plants and the molecular mechanisms behind their growth, reproduction, diffusion and ability to cope with stresses. This is fundamental for us, because knowing how, at a molecular level, a plant is able to ensure the correct seed development or to respond to drought, temperature shocks or other environmental stresses allow us to isolate the genes responsible or involved in these processes and to exploit them to our advantage, for example following the character during breeding processes to ensure the desired outcome. This is something that has been done for centuries by farmers; the only difference is that we are now able to link the phenotypic character to specific alleles that can be followed molecularly during the breeding process.

The number of species that can be used in a laboratory as models has increased during the last decades thanks to the development of new mutagenesis techniques like the Virus Induced Gene Silencing (VIGS)[1] or the most recent CRISPR-CAS9 technology[2]. Rice, however, still remains one of the most exploited model species for functional gene analysis in a monocotyledonous plant species, thanks to its agricultural relevance and its ease for using it in molecular studies (it was one of the first plant genomes to be completely sequenced[3] and one of the first monocot species for which an efficient transformation protocol was set[4]).

Almost 30 years of functional studies have assessed the fundamental role of *AGAMOUS* subfamily genes in all phases of plant reproduction. Besides their role in the ABCDE model for flower development, which is continuously reconfirmed by more and more findings in new model species, it is by now clear that their activity is not restricted to flower development only, but continues after fertilization to ensure the correct development of fruits and seeds. Since

AGAMOUS subfamily members are important transcription factors regulating processes related to reproduction, a precise understanding of their mechanisms of action and of the processes they regulate will make an important contribution to improve plant productivity in the future. Doing this research with the model crop rice is important since it opens the way to understanding these processes also in other important cereals.

How S109 influences the activity of rice OsMADS3?

MADS-box proteins have maintained a highly conserved structure during evolution and there are evidences demonstrating that small variation in their conserved domain can greatly affect their functionality. It was interesting to find that in rice the *AGAMOUS* ortholog *OsMADS3* has two isoforms, both normally expressed, which differ only in the presence or absence of a single serine residue in one of the most conserved domains, the K-domain; even more interesting is that this situation was not restricted to rice only but also in *MADS3* orthologs from many other graminaceae species.

Our experiments revealed that the two isoforms actually behave differently, when ectopically expressed in Arabidopsis; the serine-lacking isoform *OsMADS3*$^{\Delta S109}$, in particular, appears to work better than the other isoforms in generating homeotic changes in Arabidopsis flowers. This result is curious, since the C-function of the ABC model in rice was originally associated to the other isoform (together with the paralogous gene *OsMADS58*); moreover, our yeast assays demonstrated that OsMADS3$^{\Delta S109}$ cannot interact with OsMADS3^{S109} partners, as one may expect from an alteration in the conserved K-domain, which is fundamental for protein-protein interaction.

Expression of *OsMADS3*$^{\Delta S109}$ transcript and conservation of this isoform throughout *Poaceae*, however, suggest it could have an active role. With this regard, it would be interesting to exploit CRISPR-CAS9 technology to generate a rice variety that can only express *OsMADS3*$^{\Delta S109}$, in order to determine if its functionality can compensate for the lack of the *OsMADS3*^{S109} isoform.

A discovery in this sense could open the way to experiments and analyses in other *Poaceae* where the two isoforms are also present.

Which processes are regulated by AGAMOUS-like genes during rice seed development?

AGAMOUS-like genes are actively involved in the correct development of seeds and fruits, as demonstrated by functional analyses of mutants generated for example in *A.thaliana*[5–11], *Petunia hybrida*[12], *Nicotiana benthamiana*[13] and tomato[14]. However, we have still little information about their functions in fruit plants or crop plants, due to the difficulty on generating mutants; some data have been obtained by analysing natural varieties of fruit plants like peach[15,16] and oil palm[17].
The availability of rice mutants in *AGAMOUS* genes and the previous knowledge about their role in flower development makes once again rice the perfect model organism to use as a starting point to better understand the role of *AGAMOUS* genes during seed development.
Our preliminary analyses clearly demonstrate that the four rice *AGAMOUS* genes *OsMADS3, OsMADS13, OsMADS21* and *OsMADS58* are redundantly involved in controlling grain development: severe defects in seeds can indeed be observed only when multiple mutations are combined in a single rice plant.
Defects appear to be mainly related to seed coat formation and nutrient intake; this correlates with the expression profile of the four genes in seed, which is restricted to integuments and vasculature. These data also correlate with what has been reported in other species, where defects are mainly seen in seed/fruit filling, integument formation and control of lignification processes.
The results obtained are preliminary and require further experimentation, but they are of valuable importance since they represent a starting point to uncover new factors involved in fruit and seed development, in rice as in other crops and fruit plants; this, in turn, could open new possibilities in terms of yield and quality improvement.

BIBLIOGRAPHY

1. Baulcombe, D. C. Fast forward genetics based on virus-induced gene silencing. *Curr. Opin. Plant Biol.* **2,** 109–113 (1999).
2. Miao, J. *et al.* Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* **23,** 1233–1236 (2013).
3. Goff, S. A. *et al.* A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science (80-.).* **296,** 92–100 (2002).

4. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6,** 271–82 (1994).
5. Pinyopich, A. *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424,** 85–88 (2003).
6. Favaro, R. *et al.* MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* **15,** 2603–2611 (2003).
7. Liljegren, S. J. *et al.* SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* **404,** 766–770 (2000).
8. Mizzotti, C. *et al.* The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J.* **70,** 409–20 (2012).
9. Mizzotti, C. *et al.* SEEDSTICK is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat. *PLoS Genet.* **10,** (2014).
10. Orozco-Arroyo, G., Paolo, D., Ezquer, I. & Colombo, L. Networks controlling seed size in Arabidopsis. *Plant Reproduction* **28,** 17–32 (2015).
11. Ezquer, I. *et al.* Restricted Access The developmental regulator STK controls the structure and mechanical properties of the Arabidopsis seed coat. *Plant Cell* **29,** (2016).
12. Colombo, L. *et al.* Downregulation of Ovule-Specific MADS Box Genes from Petunia Results in Maternally Controlled Defects in Seed Development. *Plant Cell* **9,** 703–715 (1997).
13. Fourquin, C. & Ferrándiz, C. Functional analyses of AGAMOUS family members in Nicotiana benthamiana clarify the evolution of early and late roles of C-function genes in eudicots. *Plant J.* **71,** 990–1001 (2012).
14. Pan, I. L., McQuinn, R., Giovannoni, J. J. & Irish, V. F. Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development. *J. Exp. Bot.* **61,** 1795–1806 (2010).
15. Tadiello, A. *et al.* A PLENA-like gene of peach is involved in carpel formation and subsequent transformation into a fleshy fruit. *J. Exp. Bot.* **60,** 651–661 (2009).
16. Tani, E., Polidoros, A. N. & Tsaftaris, A. S. Characterization and expression analysis of FRUITFULL- and SHATTERPROOF-like genes from peach (Prunus persica) and their role in split-pit formation. *Tree Physiol.* **27,** 649–659 (2007).
17. Singh, R. *et al.* The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500,** 340–4 (2013).