Endosperm evolution by duplicated and neofunctionalized Type I MADS-box transcription factors

Yichun Qiu 1,2, Claudia Köhler 1,2, *

1 Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Centre for Plant Biology, Uppsala, 75007, Sweden

2 Max Planck Institute of Molecular Plant Physiology, Potsdam Science Park, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

*Correspondence: Claudia Köhler koehler@mpimp-golm.mpg.de
Abstract

MADS-box transcription factors (TFs) are present in nearly all major eukaryotic groups. They are divided into Type I and Type II that differ in domain structure, functional roles, and rates of evolution. In flowering plants, major evolutionary innovations like flowers, ovules and fruits have been closely connected to Type II MADS-box TFs. The role of Type I MADS-box TFs in angiosperm evolution remains to be identified. Here, we show that the formation of angiosperm-specific Type I MADS-box clades of $\text{M}_\gamma$ and $\text{M}_\gamma$-interacting $\text{M}_\alpha$ genes ($\text{M}_\alpha^*$) can be tracked back to the ancestor of all angiosperms. Angiosperm-specific $\text{M}_\gamma$ and $\text{M}_\alpha^*$ genes were preferentially expressed in the endosperm, consistent with their proposed function as heterodimers in the angiosperm-specific embryo-nourishing endosperm tissue. We propose that duplication and diversification of Type I MADS-genes underpins the evolution of the endosperm, a developmental innovation closely connected to the origin and success of angiosperms.

Introduction

MADS-box transcription factors (TFs) are an evolutionary ancient class of TFs and major developmental regulators present in nearly all major eukaryotic groups (Alvarez-Buylla et al. 2000). They have largely amplified during land plant evolution and play important roles in regulating organ patterning and timing of reproductive developmental programs (Nam et al. 2003; Gramzow and Theißen 2013). The loosely conserved DNA binding MADS-domain is located at the N-terminus of MADS-box proteins, while based on the C-terminal sequences two types of MADS-box TFs are distinguished, Type I and Type II (Alvarez-Buylla et al. 2000; Schwarz-Sommer et al. 1990). The duplication and divergence of Type II MADS-box genes, or MIKC-type, has been linked to the evolution of floral organs in angiosperms, including flowers, ovules and fruits (Nam et al. 2003; Becker and Theissen 2003; Ruelens et al. 2013; Ruelens et al. 2017). Compared to Type II, Type I MADS-box genes are underrepresented in gymnosperms and have experienced more frequent lineage-specific duplications in angiosperms, followed by fast pseudogenization and gene loss (Nam et al. 2004; Gramzow and Theißen 2013). Nevertheless, the role of Type I MADS-box TFs in angiosperm evolution remains to be identified. Emerging studies suggest a role for Type I MADS-box genes in the regulation of female gametophyte and endosperm development in Arabidopsis and grasses (Bemer et al. 2008; Colombo et al. 2008; Steffen et al. 2008; Shirzadi et al. 2011; Roszak and Köhler 2011; Hehenberger et al. 2012; Chen et al. 2016; Batista et al. 2019; Zhang et al. 2020; Paul et al. 2020).

The endosperm is a reproductive novelty of angiosperms that develops as the second product of double fertilization alongside the embryo to support its growth. This nourishing behaviour of endosperm starts only after fertilization; in contrast to gymnosperms, where the large female gametophyte stores nutrients independently of
the fertilization status of the gametophyte (Baroux et al., 2002). The endosperm is furthermore establishing reproductive barriers between closely related species, fueling plant speciation (Köhler et al. 2021). Considering the contribution of the endosperm to the evolutionary success of angiosperms, understanding the genetic basis of endosperm evolution is of key importance. In this study, we establish a link between the evolution of Type I MADS-box genes and the origin of the endosperm in flowering plants. We hypothesize that through gene duplication and neofunctionalization, novel subfamilies of Type I MADS-box TFs acquired endosperm-specific function in the shared common ancestor of all extant angiosperms after its divergence from gymnosperms. This process likely underpinned the evolution of the endosperm in angiosperms.

Results and Discussion

Duplication of Mβ and My MADS-box TF genes is concerted with the evolution of angiosperms

We identified Type I MADS-box genes in 40 species, representing all major lineages of angiosperms and other land plants as outgroups (Supplementary Table S1). The phylogeny of Type I MADS-box genes in all angiosperms revealed three major clades (Fig. 1; Supplementary Figure S1), corresponding to the previously defined groups Mα, Mβ and My (Parenicová et al. 2003; Arora et al. 2007; Gramzow and Theißen 2013). Specifically, we found My type genes in all angiosperms we assayed (Supplementary Table S2), including Amborella trichopoda, the species sister to all other angiosperms, suggesting the presence of an ancestral My MADS-box gene in the most recent common ancestor of all angiosperms. Mβ genes in angiosperms are sister to the angiosperm My clade, while the most closely related homologs in three major lineages of gymnosperms, Picea abies, Ginkgo biloba, and Gnetum luofuense (previously identified as Gnetum montanum in the genome project) (Wan et al. 2018; Hou et al. 2020), form a clade that is the outgroup of the angiosperm My / Mβ clade, followed successively by Mβ-like genes in the fern Salvinia cucullata, the clubmoss Selaginella moellendorffii and the mosses Physomitrella patens and Sphagnum fallax. Supporting previous findings (Gramzow et al. 2014), ancestral seed plants probably possessed only Mβ-like genes, in form of pre-duplicated Mβ/γ genes (Fig. 1). After the divergence from the ancestral gymnosperms, a gene duplication event in the common ancestor of all angiosperms gave rise to the My clade, thus most likely there was at least one ancestral angiosperm Mβ gene and one ancestral angiosperm My gene inherited in all the descendant lineages of angiosperms (Fig.1).

Expression of My MADS-box TF genes in the endosperm is ubiquitous across the phylogeny of angiosperms

We investigated the expression patterns of the duplicated Type I MADS-box genes to pinpoint their regulatory roles in certain tissue types. Based on transcriptome data across different organs and developmental stages in Arabidopsis thaliana (Klepikova et al. 2016), My genes were preferentially expressed in seeds and siliques, but rarely in vegetative tissues (Fig. 2a). Using available microarray data from dissected seed tissues (Belmonte et al.
Fig. 1. Phylogeny of Type I MADS-box TFs in selected land plants. Gene identifiers as in Supplementary Table S1-2.

2013), we inferred that several Mγ genes were mainly expressed in the early developing endosperm, but less or absent in the other compartments of seeds, such as seed coat or embryo (Fig. 2b). These data suggest that Mγ MADS-box TFs have endosperm-specific functions in A. thaliana. Consistent with this notion, the Mγ MADS-box TF PHERES1 is a master regulator of a gene regulatory network controlling endosperm development (Batista et al. 2019).
Fig. 2. Expression of Type I MADS-box genes and interaction of Type I MADS-box proteins in *Arabidopsis thaliana*. A. Expression of *Arabidopsis* Type I MADS-box genes across different organ types and developmental stages (data from Klepikova et al. 2016). B. Expression of *Arabidopsis* Type I MADS-box genes across seed tissue types and developmental stages (data from Belmonte et al. 2013). end: endosperm; em: embryo; sc: seed coat; micro: micropylar; peri: peripheral; chal: chalazal; preglob: preglobular; glob: globular. C. Interaction between *Arabidopsis* Mα TFs and Mβ or Mγ TFs (based on yeast-two-hybrid data from de Folter et al. 2005 and Bemer et al. 2008). Phylogeny of Type I MADS-box genes is shown as ML trees with bootstrap values supporting the branches. AGL37 is also known as PHE1, and AGL38 as PHE2.
We also investigated the endosperm transcriptomes at early developing stages of maize, coconut, castor bean, soybean, and tomato and found at least one of the Mγ genes to be expressed in the endosperm of each species, consistent with their proposed roles in endosperm development (Fig. 3). The Mγ genes of maize and soybean had either none or minimal expression in the embryo, supporting an endosperm-specific function (Supplementary Figure S2). Mγ gene expression was also detected in whole-seed transcriptomes of rice, avocado and monkeyflower (Fig. 3). Since the orthologous Mγ genes were primarily expressed in the endosperm in other species, we infer that the observed Mγ expression in whole-seed transcriptomes likely reflects transcription predominantly in the endosperm. Thus, Mγ genes are ubiquitously expressed in the early endosperm of various species representing major lineages of angiosperms, including eudicots, monocots and magnoliids, indicating that endosperm expression of Mγ genes is a conserved feature of angiosperms. Among those expressed Mγ genes, OsMADS87/89 in rice have been characterized as transcription factors regulating endosperm development similar to PHERES1 in A. thaliana, suggesting that the expressed Mγ genes in diverse angiosperm lineages may function similarly (Chen et al. 2016; Paul et al. 2020).

In contrast, Mβ genes in A. thaliana were barely expressed in the endosperm or other seed tissues, only one of them had low expression in the seed coat (Fig. 2b). Similarly, in maize transcriptomes, Mβ expression was not detected in the endosperm (Fig. 3). While Mβ expression was detectable at variable levels in the endosperm transcriptomes of coconut, soybean, castor bean and tomato (Fig. 3), the expression level of Mβ genes was lower compared to the corresponding Mγ expression. Based on whole-seed transcriptomes, Mβ genes in avocado were nearly not expressed, Mβ genes in rice were expressed at low level, while some Mβ genes in monkeyflower were active at later stages of seed development compared to Mγ genes. The sporadic occurrence of Mβ gene expression in the endosperm or other seed tissues across the phylogeny of angiosperms suggests that the function of Mβ is dispensable in the context of endosperm regulation. In support of this notion, Type I MADS-box genes with known functional roles in the endosperm are either Mγ or Mα type genes (Bemer et al. 2008; Colombo et al. 2008; Steffen et al. 2008; Shirzadi et al. 2011; Roszak and Köhler 2011; Hehenberger et al. 2012; Chen et al. 2016; Batista et al. 2019; Zhang et al. 2020; Paul et al. 2020). Absence of Mβ genes was previously reported for the orchids Apostasia shenzhenica, Phalaenopsis equestris and Dendrobium catenatum, and the loss of Mβ genes was proposed connected to the deficiency of endosperm in orchids (Zhang et al. 2017). Nevertheless, some orchid species undergo double fertilization and form a rudimentary endosperm (Pace 1907; Sood and Mohana Rao 1988), suggesting that loss of Mβ is not directly related to the loss of endosperm formation in orchids. In agreement with this view, transcripts of Mα and Mγ are present in developing seeds of A. shenzhenica and P. equestris (Zhang et al. 2017), likely derived from the arrested endosperm. In A. thaliana, expression of some Mβ genes could be detected in the female gametophyte (Bemer et al. 2010), raising the hypothesis that their functional role is restricted to maternal tissues, rather than the endosperm. We tested this hypothesis by investigating the
Expression of Type I MADS-box genes in flowering plants. Upper panels: endosperm transcriptomes of tomato, soybean, coconut, castor bean and maize. Lower panels: whole-seed transcriptomes of avocado, rice and monkeyflower. In each panel, the expression values were normalized into a 0-1 spectrum, with the max value set as 1. For soybean, maize, avocado and rice, gene expression levels in two genotypes (LA1589 / SU1 genotypes) are shown. In rice, gene expression levels in two reciprocal crosses between lines ZB107 and ZB306 are shown. In castor bean, gene expression levels in two genotypes (Aromatic / non-aromatic) are shown. For tomato and coconut, three stages are shown: dap: days after pollination; daa: days after anthesis. For tomato and coconut, three stages are shown; 26 dap, 40 dap and 60 dap. In each panel, the expression values were normalized into a 0-1 spectrum, with the max value set as 1.
transcriptomes of species with perispermic seeds, in which the maternally-derived perisperm rather than the endosperm provides nutrients to the embryo. Consistent with the proposed functional role of Mβ genes in maternal tissues, we detected Mβ transcripts in the transcriptome assembly from perisperm of Coffea arabica. Likewise, in Nymphaea thermarum perispermic seeds, transcript levels of Mβ genes were much higher compared to the barely detectable My gene transcripts (Supplementary Fig. S3), consistent with the perisperm accounting for the majority of the seed volume in Nymphaea (Povilus et al. 2015). We also investigated transcriptomes of gymnosperm reproductive tissues to infer the functional role of pre-duplicate Mβ/γ orthologs (Supplementary Fig. S3). Mβ/γ orthologous genes were expressed in female cones of Picea abies and ovules of Gnetum luofuense and expression of some Mβ/γ orthologous genes could also be detected in developing seeds of Gnetum luofuense, suggesting these genes perform important roles in the maternal reproductive tissue and possibly regulate the maternal nourishing behaviour supporting the development of seeds. In gymnosperms, the large female gametophyte nourishes the embryo after fertilization; while in angiosperms this role has been adopted by the endosperm which develops alongside the embryo after fertilization (Baroux et al. 2002). Based on our data we propose that the function of pre-duplicated Mβ/γ genes was to control nutrient provisioning in the female gametophyte, a function that is maintained by angiosperm Mβ genes acting in the female gametophyte and perisperm, while My genes neofunctionalized and adopted an endosperm-specific function, likely enabling endosperm development.

**Duplication of Mα genes and specialization of interaction with My and Mβ**

MADS-box TFs usually form homo- or heterodimers (Kaufmann et al. 2005). In A. thaliana, an atlas of MADS-box interactions based on yeast two-hybrid data revealed distinct interaction patterns between Type II and Type I TFs (de Folter et al. 2005). Some Type II MADS-box TFs can homodimerize, but many typically heterodimerize only with other Type II TFs. In contrast, Type I TFs unlikely form homodimers, nor do they heterodimerize within the Mα, Mβ and My subgroups. Instead, Mα TFs interact with members of the Mβ and My subgroups, while Mβ TFs and My TFs barely interact, consistent with their intrinsic relatedness inherited from pre-duplicate Mβ/γ ancestors, which only dimerize with ancestral Mα TFs. Notably, we found that the Mα TFs (AGL62, 40, 28, 23, 61) that mainly interact with My TFs clustered in a single clade (Fig. 2c). Another cluster contained Mα TFs that interact specifically with Mβ TFs and Mα TFs that have the potential to interact with both, Mβ and My. Genes encoding for the obligate My-interacting Mα TFs (Mα* hereafter) were preferentially expressed in reproductive tissues and co-expressed with PHE1/2 and other genes encoding for My TFs in the endosperm (Fig. 2). In contrast, genes encoding for Mβ-interacting Mα TFs, as well as Mβ genes were not expressed in the endosperm. Those Mα TFs that were able to
interact with both, Mβ and Mγ TFs, did not co-express with My TFs in the endosperm, making it unlikely that they are able to form functional heterodimers with My TFs.

We next investigated if there are Mα TFs specialized to be Mα* in other angiosperms. A bona fide Mα* TF is expected to have central cell / endosperm-enriched or endosperm-specific expression and interacts with My TFs. Based on these predictions, Arabidopsis AGL62, 40, 28, 23, 61 classify as Mα* TFs. In rice, the Mα type TFs MADS78 and 79 interact with the My type TFs MADS87 and MADS89 and the interaction between the two Mα TFs and My TFs is required for endosperm development (Paul et al. 2020). The two rice Mα genes as well as the two My genes are barely expressed in non-endosperm tissues (Sakai et al. 2011; Davidson et al. 2012). We found that the two rice Mα genes are closely related with each other in the same subclade (Fig. 1; Fig. S1). Knockout of both, MADS78 and 79 genes, results in endosperm failure and seed lethality (Paul et al. 2020), revealing that other Mα TFs that putatively interact with Mβ TFs cannot complement the My-interacting function in the endosperm.

To test whether the functional divergence of Mα genes can be detected in other angiosperm species, we analyzed the expression of Mα genes in the transcriptomes of endosperm or seeds where My expression could be detected. We also found Mα genes to be highly expressed specifically in the endosperm or seeds in those species, suggesting that the regulatory divergence between the Mα* genes and other Mα genes took place across the angiosperm phylogeny (Fig. 3). We hypothesize that in response to the duplication of Mβ and My genes, the duplicated Mα genes specialized in protein-protein interactions and subsequently the novel interacting pairs, Mα* and My, together occupied the endosperm regulatory niche.

While the phylogeny of Mα group Type I MADS-box TFs in land plants was difficult to resolve, there is only a single cluster of Mα genes in non-flowering plants (Fig. 1). Thus, the Mα-like genes in non-flowering plants have not undergone the diversification observed in angiosperms, so they likely represent the ancestral interacting partners of the pre-duplicated Mβ-like genes (Fig. 1). In contrast, several rounds of duplications gave rise to angiosperm-specific Mα TF clades that could diverge to Mα* genes (Fig. 1), in concert with the duplication of Mβ and My clades.

We observed that many angiosperm species have at least two clusters of divergent Mα genes, including the groups representing the successive sister lineages to all other angiosperms, Amborella and Nymphaeales. Furthermore, the Mα gene phylogeny of all major angiosperm groups is largely, although imperfectly, reflected by a two-clade pattern, despite the uncertainty at the basal nodes with quite short branches (Fig.1; Fig. S1). A parsimonious model to describe the evolution of Mα type genes in angiosperms is that ancestral angiosperms most likely already possessed two, if not multiple types of Mα genes that arose from angiosperm-specific duplication. These could then have subfunctionalized by forming heterodimeric complexes with either Mβ or My interacting partners. Another requirement for the specialization of bona fide Mα* TFs was the acquisition of novel expression in the endosperm. We hypothesize this two-step specialization restrained the occurrence of Mα* precursors and
propose that one group of ancestral Mα TFs initiated the subfunctionalization and gave rise to a single cluster of potential Mα* TFs, which were capable to specialize into Mα*, while the other Mα TFs did not gain this competence. We observed that in all the eudicot species we surveyed, there are Mα genes closely related to the AGL62 clade of Arabidopsis and expressed in the endosperm or seed transcriptomes; likewise, the expressed Mα genes in maize and coconut are in the same clade as MADS78/79 of rice (Supplementary Figure S4). These putative Mα* genes may have the same Mα* origin. Alternatively, it is also possible that several events of Mα* specialization took place in different Mα subclades convergently in angiosperms. Based on approximately unbiased (AU) tests (Shimodaira 2002) it is not possible to differentiate between the two hypotheses (Supplementary Figure S5). Nevertheless, following the specialization of an ancestral Mα*, some descendant genes that duplicated subsequently in the clade may have lost the function and pseudogenized, consistent with previous predictions (Nam et al. 2004). In consequence, the retained functional Mα* genes appear scattered in the phylogeny, obscuring a possible shared origin. In summary, we conclude that duplication of Mα genes and subsequent specialization of Mα* in angiosperms enabled the formation of heteromeric Type I MADS TF complexes required for the regulation of endosperm development.

Conclusion

Angiosperms are the most abundant and diverse group among land plants. The success of angiosperms is closely connected to the developmental innovations of flowers and fruits, as well as the process of double fertilization, coupling fertilization to the formation of the embryo nourishing endosperm tissue (Baroux and Grossniklaus 2019). Duplication and diversification of type II MADS-box genes underpin the evolution of flowers and fruits in angiosperms (Irish and Litt 2005; Ruelens et al. 2017), while the role of type I MADS-box genes for angiosperm evolution remained obscure. Based on our data, we propose that the origin of the embryo nourishing endosperm tissue is linked to the angiosperm-specific duplication of Type I MADS-box genes (Fig. 4). In the earliest land plants, ancestral Mα and Mβ/γ-like TFs likely formed heterodimers that had reproductive function based on the expression of gymnosperm Mα and Mβ/γ TFs in female cones and seeds. After the angiosperm lineage diverged from the gymnosperms, true Mγ TFs arose by gene duplication, experienced neofunctionalization, and drove the concerted divergence of some Mα TFs formed by angiosperm-specific gene duplication events. These novel Mγ-Mα heterodimers adopted a function as master regulators of the endosperm developmental network in flowering plants. This proposed scenario is strongly supported by the specific or preferential expression of Mγ and Mα* genes in the endosperm of all sampled angiosperm species as well as functional data in A. thaliana and rice, revealing that Mγ and Mα* TFs are required for endosperm development (Chen et al. 2016; Batista et al. 2019; Paul et al. 2020). In contrast to gymnosperms that only have few Type I MADS-box genes (Gramzow et al. 2014); in angiosperms, their number strongly amplified, correlating with the evolution of the embryo-nourishing endosperm. The link between Mγ TFs and endosperm evolution was furthermore supported by the negligible expression of Mγ
genes in perispermic seeds, in which the maternal perisperm instead of the endosperm supports embryo growth (Lu and Magnani 2018). The maternal nourishing function in perispermic seeds correlates with the expression of Mβ genes, consistent with the proposed ancestral role of pre-duplicated Mβ/γ genes in regulating nutrient transfer from the maternal tissues to the embryo.

Together, our work provides new insights into the role of Type I MADS-box proteins in the origin and evolution of the endosperm, a developmental novelty associated with the rise and diversification of angiosperms.

![Diagram showing the evolution of MADS-box proteins in seed plants](https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab355/6460346)
Methods

Phylogenetic analyses

Amino acid sequences of Type I and Type II MADS-box proteins of A. thaliana obtained from TAIR10 were used as queries to search for MADS-box proteins in other plant species. The sequences of coding genes in land plant lineages were obtained from PLAZA 4.0 (https://bioinformatics.psb.ugent.be/plaza/; Van Bel et al. 2018), Phytozome v.12 (https://phytozome.jgi.doe.gov/; Goodstein et al., 2012), CoGe (https://genomevolution.org/coge/; Grover et al. 2017) or other taxon-themed databases (Supplementary Table S1). MADS-box genes were obtained through reciprocal best BLASTP hits with A. thaliana MADS-box genes. The presence of MADS domain in the BLASTP output sequences was further confirmed by the conserved domain search tool, CD-Search (Marchler-Bauer and Bryant 2004) by aligning to the MADS domain entries in the Conserved Domain Database (Lu et al. 2020).

MUSCLE was used to generate the amino acid alignments of MADS-box domains extracted from the identified genes with default settings (Edgar 2004). IQ-TREE 1.6.7 was applied to perform phylogenetic analyses for maximum likelihood trees (Nguyen et al. 2015). The implemented ModelFinder determined LG amino acid replacement matrix (Le and Gascuel 2008) to be the best substitution model in the tree inference (Kalyaanamoorthy et al. 2017). 1000 replicates of ultrafast bootstraps were applied to estimate the support for reconstructed branches (Hoang et al. 2018). The Mα, Mβ and Mγ type genes were curated from the phylogenetic position with the defined Arabidopsis MADS-box genes. Specifically, for elucidating the evolutionary trajectory of putative Mα* TFs, we compared the topology of constrained phylogenetic trees based on different hypotheses by approximately unbiased (AU) tests (Shimodaira 2002).

Expression analyses

The expression data of Type I MADS-box genes in A. thaliana were extracted from Klepicova et al. (2016) for a spectrum of different organ types and developmental stages and Belmonte et al. (2013) for specific compartments in developing seeds. The other transcriptomes used in this study were retrieved from maize (Chen et al. 2014; Walley et al. 2016), rice (Sakai et al. 2011; Davidson et al. 2012; Paul et al. 2020), soybean (Chen et al. 2021), castor bean (Xu et al. 2014), tomato (Pattison et al. 2015), coconut (Saensuk et al. 2016), avocado (Ge et al. 2019), monkeyflower (Flores-Vergara et al. 2019), coffee (Ivamoto et al. 2017), Nymphaeathermarum (Povilus and Friedman 2021), Picea (Nystedt et al. 2013), and Gnetum (Hou et al. 2019; Deng et al. 2020).

Acknowledgements
We thank Dr. Rebecca Povilus and Dr. William Friedman for sharing the seed transcriptome data of *Nymphaea thermarum*. We thank Dr. Qin Li for the comments on the data analyses and visualization. This work was supported by a grant from the Swedish Research Council to CK, a grant from the Knut and Alice Wallenberg Foundation to CK, and support from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine to CK.

**Author Contributions**

Conceptualization, Y.Q. and C.K.; Methodology, Y.Q.; Investigation, Y.Q.; Validation, Y.Q. and C.K.; Formal Analysis, Y.Q.; Data Curation, Y.Q. and C.K.; Writing – Original Draft, Y.Q. and C.K.; Writing – Review & Editing, Y.Q. and C.K.; Visualization, Y.Q.; Funding Acquisition, C.K.; Supervision, C.K.

**Declaration of Interests**

The authors declare no competing interests.

**Figure legends**

Fig. 1. Phylogeny of Type I MADS-box TFs in selected land plants. Gene identifiers as in Supplementary Table S1-2.

Fig. 2. Expression of Type I MADS-box genes and interaction of Type I MADS-box proteins in *Arabidopsis thaliana*. A. Expression of *Arabidopsis* Type I MADS-box genes across different organ types and developmental stages (data from Klepikova et al. 2016). B. Expression of *Arabidopsis* Type I MADS-box genes across seed tissue types and developmental stages (data from Belmonte et al. 2013). end: endosperm; em: embryo; sc: seed coat; micro: micropylar; peri: peripheral; chal: chalazal; preglob: preglobular; glob: globular. C. Interaction between *Arabidopsis* Ma TFs and My or Mb TFs (based on yeast-two-hybrid data from de Folter et al. 2005 and Bemer et al. 2008). Phylogeny of Type I MADS-box genes is shown as ML trees with bootstrap values supporting the branches. AGL37 is also known as PHE1, and AGL38 as PHE2.

Fig. 3. Expression of Type I MADS-box genes in flowering plants. Upper panels: endosperm transcriptomes of tomato, soybean, coconut, castor bean and maize. Lower panels: whole-seed transcriptomes of avocado, rice and monkeyflower. In each panel the expression values were normalized into a 0-1 spectrum, with the max value set as 1. For soybean, maize, avocado and rice, gene expression levels at two developmental stages; for monkeyflower, three stages are shown; dap: days after pollination; daa: days after anthesis. For tomato and coconut, gene expression levels in two genotypes (LA1589 / SUN) or varieties (aromatic / non-aromatic) are shown. For castor bean, gene expression levels in two reciprocal crosses between lines ZB306 and ZB107 are shown. In rice, LOC_Os09g02830 is OsMADS78; LOC_Os01g74440 is OsMADS79; LOC_Os03g38610 is OsMADS87; LOC_Os01g18440 is OsMADS89.
Fig. 4. Model depicting the role of Type I MADS-box TFs in the evolution of embryo-nourishing tissues in seed plants. In early seed plants, ancestral Mα and Mβ/γ TFs likely formed heterodimers that regulated the embryo-nourishing behaviour. The gymnosperm Mα and Mβ-like TFs dimerize and function in maternal tissues. Angiosperms express Mα and Mβ heterodimers in maternal tissues, while Mα* and My TFs that arose from angiosperm-specific duplications function in the endosperm and likely enabled endosperm evolution.

Supplementary Figures and Tables

Fig. S1. Phylogeny of Type I MADS-box TFs in several focused groups of land plants with Arabidopsis thaliana MADS-box genes (highlighted in red) as references, shown by ML trees with bootstrap values supporting the branches. Gene identifiers in Supplementary Table S1-2. Suffixes of “a”, “b”, “g” denote Mα, Mβ or My TFs, respectively; ‘x’ after gene identifier denotes the second MADS domain in the gene. Genes from selected species were colored to demonstrate the presence of My genes and the two or more clusters of Mα genes. A. ANA, Magnoliids, Ceratophyllales and Proteales; Amborella trichopoda (yellow), Nymphaea thermarum (purple), Cinnamomum camphora (blue). B. Monocots; Oryza sativa (yellow), Zoster marina (purple), Asparagus officinalis (blue). C. Asterids and Caryophyllales; Chrysanthemum nankingense (yellow), Erythranthe guttata (purple), Solanum lycopersicum (blue). D. Rosids; Eucalyptus grandis (yellow), Theobroma cacao (purple), Glycine max (blue).

Fig. S2. Expression of Type I MADS-box genes in reproductive tissues of maize (A), soybean (B) and tomato (C), showing endosperm-expression of My genes. The expression values were normalized into a 0-1 spectrum, with the max value set as 1. A. Gene expression levels at two developmental stages of seed tissues in maize and average expression level in non-endosperm tissues across the whole plant. B. Gene expression levels at two developmental stages of seed tissues in soybean. C. Gene expression levels in two genotypes (LA1589 / SUN) of seed and fruit tissues in tomato, before and after fertilization. en/end: endosperm; em/emb: embryo; dap: days after pollination.

Fig. S3. Expression of Type I MADS-box genes in reproductive tissues of Nymphaea thermarum (A), Picea abies (B) and Gnetum luofuense (C), showing Mβ expression. The expression values were normalized into a 0-1 spectrum, with the max value set as 1. A. Gene expression levels at unfertilized ovule and two developmental stages of seeds in Nymphaea thermarum; daa: days after anthesis. B. Gene expression levels of female and male cones in Picea abies and average expression level in vegetative tissues; veg: vegetative. C. Gene expression levels at mature vs immature seeds in Gnetum luofuense and fertile vs sterile ovules.

Fig. S4. The evolution of Mα* genes. A. Phylogeny of Mα genes from species with expression data from endosperm or seeds. The red clade has the confirmed Mα* genes in Arabidopsis thaliana, and the yellow clade has the confirmed Mα* genes in rice. Genes in the red and yellow clades are considered putative Mα* if they are expressed
preferentially in the endosperm. The avocado genes in the red and yellow clades are weakly expressed (Fig.3), but
the gene PerAmScf00637_936915306 (with two MADS domains annotated and shown in the phylogeny) in the
sister clade was predicted as functional Mα*. B. A replicated tree of cladogram form labelled with the bootstrap
support values for the branches. The internal clustering pattern within clades are better supported with values
above 90. The basal branches connecting the major clades generally have supportive values below 80, suggesting
that the backbone topology might not be fully resolved, and alternative topologies are equally possible, see Fig.
S5.

Fig. S5. Topology of different phylogenetic trees tested by IQ-tree representing two hypotheses of Mα* evolution.
Amino acid sequences of MADS domains were aligned as input. In the constrained trees, Arabidopsis Mβ and Mγ
were always used as outgroups. Three subclades of Arabidopsis Mα were shuffled due to uncertain relatedness.
Rice Mα* were swamped to cluster with different Arabidopsis clades, illustrating two hypotheses: a shared origin
of Mα* (blue clade) as in Trees 1-3, 6-8 and 11; independent rise of Mα* in Arabidopsis (pink clade) and in rice
(green clade) as in Trees 4-5, 9-10 and 12-15. Red arrows point to the putative branches giving rise to Mα* clade.
Neither hypothesis is rejected, nor is significantly better than the other. As negative controls, Trees 16 and 17,
with rice Mα* forced to cluster with Arabidopsis Mβ or Mγ, are rejected. Stats of tree topology tests from IQ-tree
are listed.

Table. S1. List of surveyed species.

Table. S2. List of identified Type I MADS-box genes in the surveyed species.

Reference

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martínez-
Castilla, L., and Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the
divergence of plants and animals. Proceedings of the National Academy of Sciences of the United States of
America, 97(10), 5328–5333. https://doi.org/10.1073/pnas.97.10.5328

Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., and Kapoor, S. (2007). MADS-box gene
family in rice: genome-wide identification, organization and expression profiling during reproductive
development and stress. BMC Genomics, 8, 242. https://doi.org/10.1186/1471-2164-8-242

Baroux, C., and Grossniklaus, U. (2019). Seeds - An evolutionary innovation underlying reproductive success in
flowering plants. Current Topics in Developmental Biology, 131, 605–642. https://doi.org/10.1016/bs.ctdb.2018.11.017
Baroux, C., Spillane, C., and Grossniklaus, U. (2002). Evolutionary origins of the endosperm in flowering plants. Genome Biology, 3(9), reviews1026. https://doi.org/10.1186/gb-2002-3-9-reviews1026

Batista, R. A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D. D., and Köhler, C. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. eLife, 8, e50541. https://doi.org/10.7554/eLife.50541

Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molecular Phylogenetics and Evolution, 29(3), 464–489. https://doi.org/10.1016/s1055-7903(03)00207-0

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proceedings of the National Academy of Sciences of the United States of America, 110(5), E435–E444. https://doi.org/10.1073/pnas.1222061110

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

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

Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., and Walia, H. (2016). Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. Plant Physiology, 171(1), 606–622. https://doi.org/10.1104/pp.15.01992

Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A., and Lai, J. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiology, 166(1), 252–264. https://doi.org/10.1104/pp.114.240689

Chen, M., Lin, J. Y., Wu, X., Apuya, N. R., Henry, K. F., Le, B. H., Bui, A. Q., Pelletier, J. M., Cokus, S., Pellegrini, M., et al. (2021). Comparative analysis of embryo proper and suspensor transcriptomes in plant embryos with different morphologies. Proceedings of the National Academy of Sciences of the United States of America, 118(6), e2024704118. https://doi.org/10.1073/pnas.2024704118
Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. M., and Colombo, L. (2008). AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. The Plant Journal, 54(6), 1037–1048. https://doi.org/10.1111/j.1365-313X.2008.03485.x

Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., Jiang, N., and Robin Buell, C. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. The Plant Journal, 71(3), 492–502. https://doi.org/10.1111/j.1365-313X.2012.05005.x

de Folter, S., Immink, R. G., Kieffer, M., Parenicová, L., Henz, S. R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M. M., Davies, B., and Angenent, G. C. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. The Plant Cell, 17(5), 1424–1433. https://doi.org/10.1105/tpc.105.031831

Deng, N., Hou, C., He, B., Ma, F., Song, Q., Shi, S., Liu, C., and Tian, Y. (2020). A full-length transcriptome and gene expression analysis reveal genes and molecular elements expressed during seed development in Gnetum luofuense. BMC Plant Biology, 20(1), 531. https://doi.org/10.1186/s12870-020-02729-1

Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Flores-Vergara, M. A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis Balaguer, M. A., Coimbra, S., Willis, J., and Franks, R. G. (2020). Developmental analysis of Mimulus seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. Frontiers in Plant Science, 11, 132. https://doi.org/10.3389/fpls.2020.00132

Ge, Y., Cheng, Z., Si, X., Ma, W., Tan, L., Zang, X., Wu, B., Xu, Z., Wang, N., Zhou, Z., et al. (2019). Transcriptome Profiling Provides Insight into the Genes in Carotenoid Biosynthesis during the Mesocarp and Seed Developmental Stages of Avocado (Persea americana). International Journal of Molecular Sciences, 20(17), 4117. https://doi.org/10.3390/ijms20174117

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research, 40(Database issue), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Gramzow, L., Weilandt, L., and Theißen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. Annals of Botany, 114(7), 1407–1429. https://doi.org/10.1093/aob/mcu066
Gramzow, L., and Theißen, G. (2013). Phylogenomics of MADS-Box genes in plants - two opposing life styles in one gene family. Biology, 2(3), 1150–1164. https://doi.org/10.3390/biology2031150

Grover, J.W., Bomhoff, M., Davey, S., Gregory, B.D., Mosher, R.A., Lyons, E. (2017). CoGe LoadExp+: A web-based suite that integrates next-generation sequencing data analysis workflows and visualization. Plant Direct, 00: 1–7. https://doi.org/10.1002/pld3.8

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. Development, 139(11), 2031–2039. https://doi.org/10.1242/dev.077057

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hou, C., Li, L., Liu, Z., Su, Y. and Wan, T. (2020) Diversity and expression patterns of MADS-Box genes in Gnetum luofuense — Implications for functional diversity and evolution. Tropical Plant Biology, 13, 36–49. https://doi.org/10.1007/s12042-019-09247-x

Hou, C., Saunders, R., Deng, N., Wan, T., and Su, Y. (2019). Pollination drop proteome and reproductive organ transcriptome comparison in Gnetum reveals entomophilous adaptation. Genes, 10(10), 800. https://doi.org/10.3390/genes10100800

Irish, V. F., and Litt, A. (2005). Flower development and evolution: gene duplication, diversification and redeployment. Current Opinion in Genetics and Development, 15(4), 454–460. https://doi.org/10.1016/j.gde.2005.06.001

Ivamoto, S. T., Reis, O. Júnior, Domingues, D. S., Dos Santos, T. B., de Oliveira, F. F., Pot, D., Leroy, T., Vieira, L. G., Carazzolle, M. F., Pereira, G. A., et al. (2017). Transcriptome analysis of leaves, flowers and fruits perisperm of Coffea arabica L. reveals the differential expression of genes involved in raffinose biosynthesis. PloS One, 12(1), e0169595. https://doi.org/10.1371/journal.pone.0169595

Kalyaanamoorthy, S., Minh, B. Q., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods, 14(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene, 347(2), 183–198. https://doi.org/10.1016/j.gene.2004.12.014
Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. The Plant Journal, 88(6), 1058–1070. https://doi.org/10.1111/tpj.13312

Köhler, C., Dziasek, K., and Del Toro-De León, G. (2021). Postzygotic reproductive isolation established in the endosperm: mechanisms, drivers and relevance. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 376(1826), 20200118. https://doi.org/10.1098/rstb.2020.0118

Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. Molecular Biology and Evolution, 25(7), 1307–1320. https://doi.org/10.1093/molbev/msn067

Lu, J., and Magnani, E. (2018). Seed tissue and nutrient partitioning, a case for the nucellus. Plant Reproduction, 31(3), 309–317. https://doi.org/10.1007/s00497-018-0338-1

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research, 48(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Research, 32(Web Server issue), W327–W331. https://doi.org/10.1093/nar/gkh454

Nam, J., dePamphilis, C. W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. Molecular Biology and Evolution, 20(9), 1435–1447. https://doi.org/10.1093/molbev/msg152

Nam, J., Kim, J., Lee, S., An, G., Ma, H., and Nei, M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proceedings of the National Academy of Sciences of the United States of America, 101(7), 1910–1915. https://doi.org/10.1073/pnas.0308430100

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature, 497(7451), 579–584. https://doi.org/10.1038/nature12211

Pace, L. (1907) Fertilization in Cypripedium. Botanical Gazette, 44(57), 353–374. https://doi:10.1086/329378
Parenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. The Plant Cell, 15(7), 1538–1551. https://doi.org/10.1105/tpc.011544

Pattison, R. J., Csukasi, F., Zheng, Y., Fei, Z., van der Knaap, E., and Catalá, C. (2015). Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. Plant Physiology, 168(4), 1684–1701. https://doi.org/10.1104/pp.15.00287

Paul, P., Dhatt, B. K., Miller, M., Folsom, J. J., Wang, Z., Krassovskaya, I., Liu, K., Sandhu, J., Yu, H., Zhang, C., et al. (2020). MADS78 and MADS79 are essential regulators of early seed development in rice. Plant Physiology, 182(2), 933–948. https://doi.org/10.1104/pp.19.00917

Povilus, R. A., and Friedman, W. E. (2021) Transcriptomes across fertilization and seed development in the water lily Nymphaea thermarum (Nymphaeales) reveal dynamic expression of DNA and histone methylation modifiers. bioRxiv, https://doi.org/10.1101/2021.04.04.438399

Povilus, R. A., Losada, J. M., and Friedman, W. E. (2015). Floral biology and ovule and seed ontogeny of Nymphaea thermarum, a water lily at the brink of extinction with potential as a model system for basal angiosperms. Annals of Botany, 115(2), 211–226. https://doi.org/10.1093/aob/mcu235

Roszak, P., and Köhler, C. (2011). Polycomb group proteins are required to couple seed coat initiation to fertilization. Proceedings of the National Academy of Sciences of the United States of America, 108(51), 20826–20831. https://doi.org/10.1073/pnas.111711108

Ruelens, P., de Maagd, R. A., Proost, S., Theißen, G., Geuten, K., and Kaufmann, K. (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nature Communications, 4, 2280. https://doi.org/10.1038/ncomms3280

Ruelens, P., Zhang, Z., van Mourik, H., Maere, S., Kaufmann, K., and Geuten, K. (2017). The origin of floral organ identity quartets. The Plant Cell, 29(2), 229–242. https://doi.org/10.1105/tpc.16.00366

Saensuk, C., Wanchana, S., Choowongkomon, K., Wongpornchai, S., Kraithong, T., Imsabai, W., Chaichoompu, E., Ruanjaichon, V., Toojinda, T., Vanavichit, A., et al. (2016). De novo transcriptome assembly and identification of the gene conferring a "pandan-like" aroma in coconut (Cocos nucifera L.). Plant Science, 252, 324–334. https://doi.org/10.1016/j.plantsci.2016.08.014
Sakai, H., Mizuno, H., Kawahara, Y., Wakimoto, H., Ikawa, H., Kawahigashi, H., Kanamori, H., Matsumoto, T., Itoh, T., and Gaut, B. S. (2011). Retrogenes in rice (Oryza sativa L. ssp. japonica) exhibit correlated expression with their source genes. Genome Biology and Evolution, 3, 1357–1368. https://doi.org/10.1093/gbe/evr111

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in Antirrhinum majus. Science, 250(4983), 931–936. https://doi.org/10.1126/science.250.4983.931

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Systematic Biology, 51(3), 492–508. https://doi.org/10.1080/10635150290069913

Shirzadi, R., Andersen, E. D., Bjerkman, K. N., Gloeckle, B. M., Heese, M., Ungru, A., Winge, P., Koncz, C., Aalen, R. B., Schnittger, A., et al. (2011). Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. PLoS Genetics, 7(2), e1001303. https://doi.org/10.1371/journal.pgen.1001303

Sood, S.K., and Mohana Rao, P.R. (1988) Studies in the embryology of the diandrous orchid Cypripedium cordigerum (Cypripedieae, Orchidaceae). Plant Systematics and Evolution, 160, 159–168. https://doi.org/10.1007/BF00936043

Steffen, J. G., Kang, I. H., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. Plant Physiology, 148(1), 259–268. https://doi.org/10.1104/pp.108.119404

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Research, 46(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002

Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urich, M. A., Nery, J. R., Smith, L. G., Schnable, J. C., Ecker, J. R., et al. (2016). Integration of omic networks in a developmental atlas of maize. Science, 353(6301), 814–818. https://doi.org/10.1126/science.aag1125

Wan, T., Liu, Z. M., Li, L. F., Leitch, A. R., Leitch, I. J., Lohaus, R., Liu, Z. J., Xin, H. P., Gong, Y. B., Liu, Y., et al. (2018). A genome for gnetophytes and early evolution of seed plants. Nature Plants, 4(2), 82–89. https://doi.org/10.1038/s41477-017-0097-2

Xu, W., Dai, M., Li, F., and Liu, A. (2014). Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. Nucleic Acids Research, 42(11), 6987–6998. https://doi.org/10.1093/nar/gku375
Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R., Hsiao, Y. Y., Niu, S. C., Wang, J. Y., Lin, Y. C., Xu, Q., Chen, L. J., et al. (2017). The Apostasia genome and the evolution of orchids. Nature, 549(7672), 379–383. https://doi.org/10.1038/nature23897

Zhang, M. X., Zhu, S. S., Xu, Y. C., Guo, Y. L., Yang, W. C., and Li, H. J. (2020). Transcriptional repression specifies the central cell for double fertilization. Proceedings of the National Academy of Sciences of the United States of America, 117(11), 6231–6236. https://doi.org/10.1073/pnas.1909465117