Endosperm Evolution by Duplicated and Neofunctionalized Type I MADS-Box Transcription Factors

Yichun Qiu[1,2] and Claudia Köhler*[,1,2]

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

[2]Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

***Corresponding author:** E-mail: koehler@mpimp-golm.mpg.de.

Associate editor: Juliette de Meaux

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 Mγ and Mγ-interacting Mα genes (Mα*) can be tracked back to the ancestor of all angiosperms. Angiosperm-specific Mγ and Mα* 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 underpin the evolution of the endosperm, a developmental innovation closely connected to the origin and success of angiosperms.

Key words: **plant reproduction, endosperm evolution, MADS-box transcription factors, gene duplication.**

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 (Schwarz-Sommer et al. 1990; Alvarez-Buylla et al. 2000). The duplication and divergence of Type II MADS-box genes, or MIKC-type, have been linked to the evolution of floral organs in angiosperms, including flowers, ovules, and fruits (Becker and Theissen 2003; Nam et al. 2003; Ruelens et al. 2013, 2017). Compared with 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; Roszak and Köhler 2011; Shirzadi et al. 2011; Hehenberger et al. 2012; Chen et al. 2016; Batista et al. 2019; Paul et al. 2020; Zhang 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 behavior 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 Mγ 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, Supplementary Material online). The phylogeny of Type I MADS-box genes in all

Article

angiosperms revealed three major clades (fig. 1; supplementary fig. S1, Supplementary Material online), corresponding to the previously defined groups Mα, Mβ, and Mγ (Parenicová et al. 2003; Arora et al. 2007; Gramzow and Theißen 2013). Specifically, we found Mγ type genes in all angiosperms we assayed (supplementary table S2, Supplementary Material online), including *Amborella trichopoda*, the species sister to all other angiosperms, suggesting the presence of an ancestral Mγ MADS-box gene in the most recent common ancestor of all angiosperms. Mβ genes in angiosperms are sister to the angiosperm Mγ 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 Mγ/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 preduplicated 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 Mγ clade, thus most likely there was at least one ancestral angiosperm Mβ gene and one ancestral angiosperm Mγ gene inherited in all the descendant lineages of angiosperms (fig.1).

Expression of Mγ 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), Mγ 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. 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).

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 fig. S2, Supplementary Material online). Mγ gene expression was

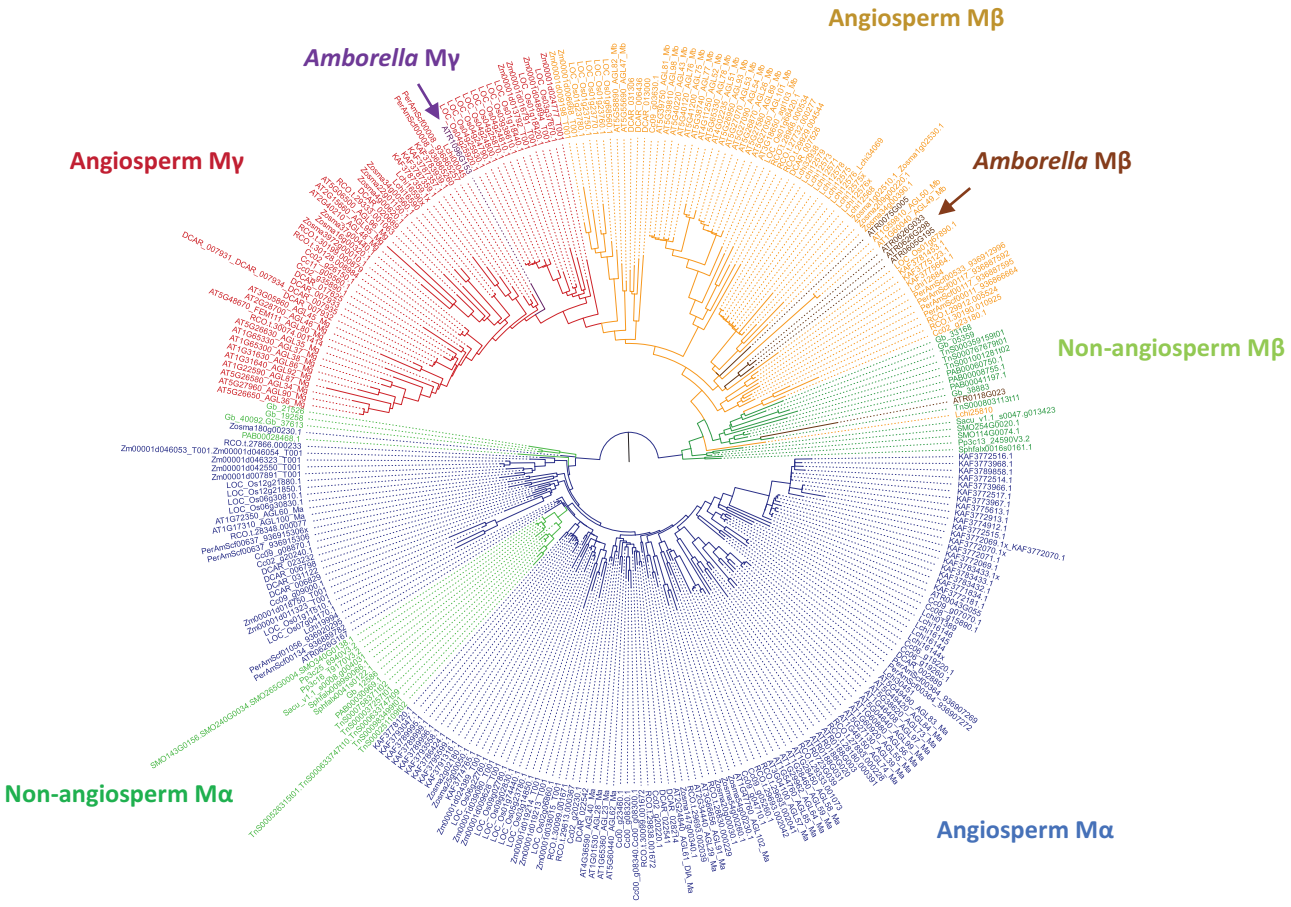

FIG. 1. Phylogeny of Type I MADS-box TFs in selected land plants. Gene identifiers as in supplementary tables S1 and S2, Supplementary Material online.

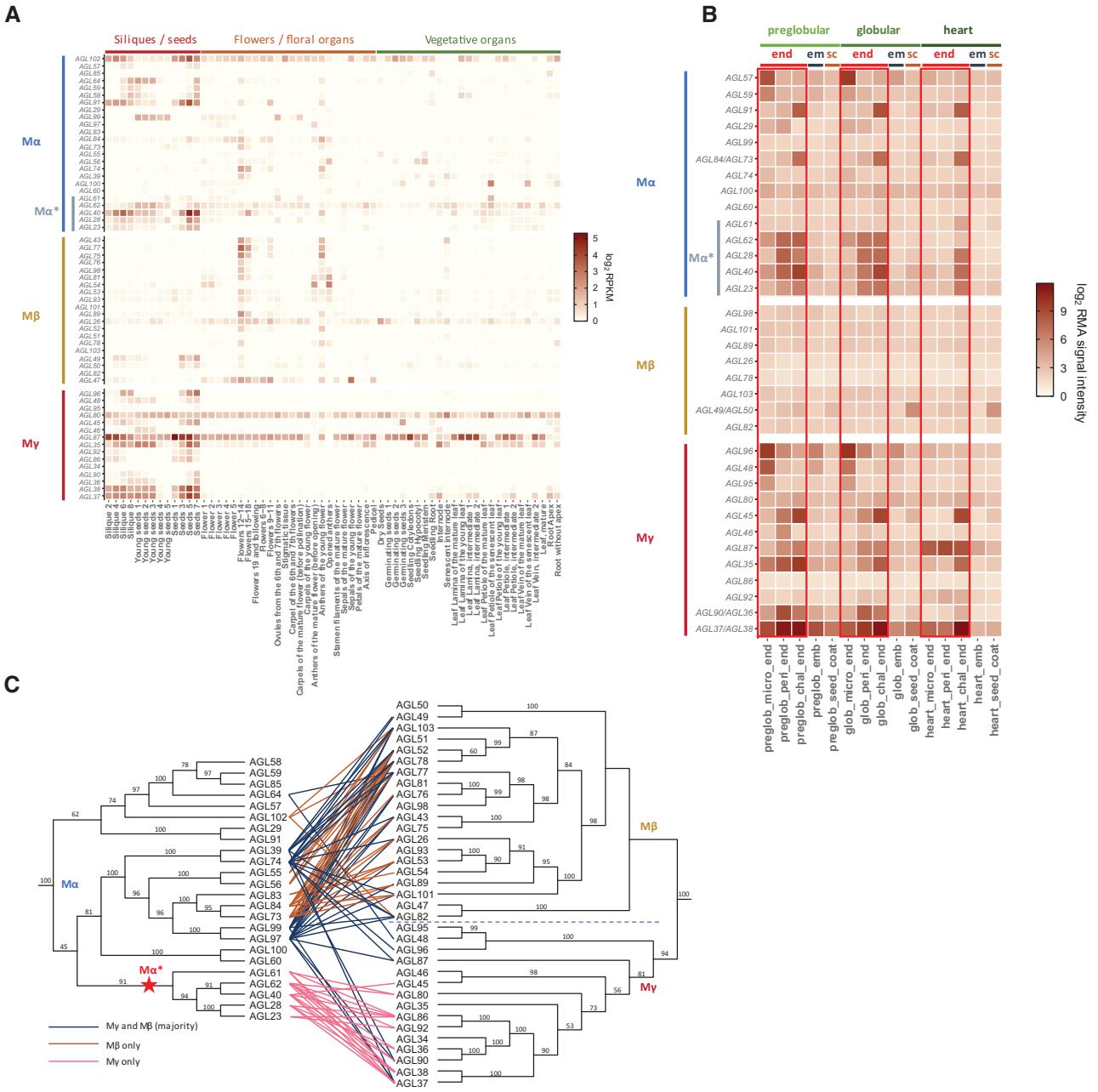

Fig. 2. Expression of Type I MADS-box genes and interaction of Type I MADS-box proteins in *Arabidopsis thaliana*. (*A*) Expression of *Arabidopsis* Type I MADS-box genes across different organ types and developmental stages (data from Klepikova et al. [2016]). (*B*) Expression of *Arabidopsis* Type 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 Mγ or Mβ 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*.

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 TFs 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. 2*B*). Similarly, in maize transcriptomes, Mβ expression was not detected in the endosperm (fig. 3). Although Mβ expression was detectable at

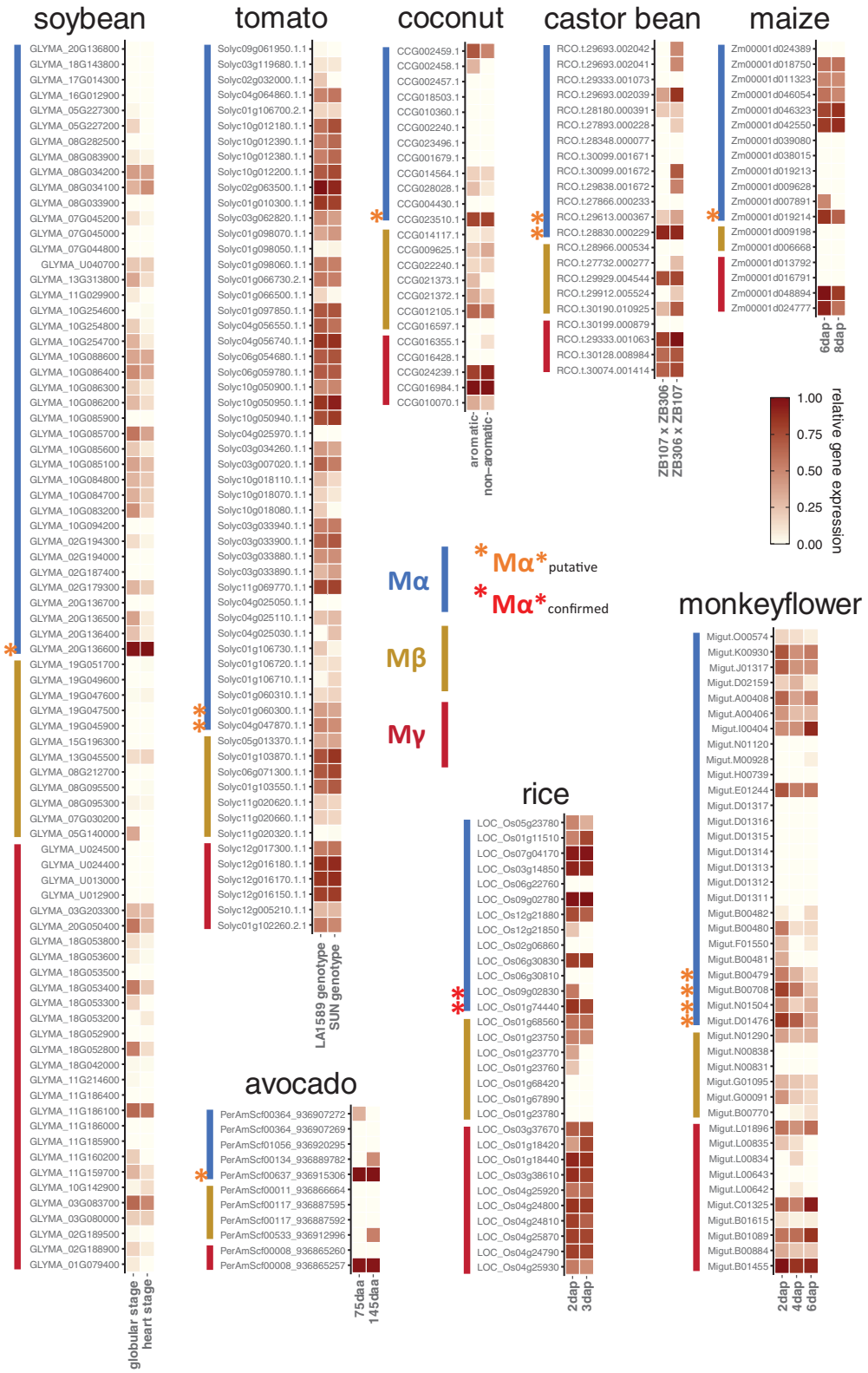

Fig. 3. Expression of Type I MADS-box genes in flowering plants. Upper panels: endosperm transcriptomes of tomato, soybean, coconut, castor bean, and maize. Lower panels: whole-seed transcriptomes of avocado, rice, and monkeyflower. In each panel, the expression values were normalized into a 0–1 spectrum, 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/nonaromatic) 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*; and LOC_Os01g18440 is *OsMADS89*.

variable levels in the endosperm transcriptomes of coconut, soybean, castor bean, and tomato (fig. 3), the expression level of Mβ genes was lower compared with 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, whereas some Mβ genes in monkeyflower were active at later stages of seed development compared with 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; Roszak and Köhler 2011; Shirzadi et al. 2011; Hehenberger et al. 2012; Chen et al. 2016; Batista et al. 2019; Paul et al. 2020; Zhang et al. 2020). The 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 to be 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 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 with the barely detectable Mγ gene transcripts (supplementary fig. S3, Supplementary Material online), 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 preduplicated Mβ/γ orthologs (supplementary fig. S3, Supplementary Material online). Mβ/γ orthologous genes were expressed in female cones of *P. abies* and ovules of *G. luofuense* and expression of some Mβ/γ orthologous genes could also be detected in developing seeds of *G. luofuense*, suggesting these genes perform important roles in the maternal reproductive tissue and possibly regulate the maternal nourishing behavior supporting the development of seeds. In gymnosperms, the large female gametophyte nourishes the embryo after fertilization; whereas 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 preduplicated 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, whereas Mγ genes neofunctionalized and adopted an endosperm-specific function, likely enabling endosperm development.

Duplication of Mα Genes and Specialization of Interaction with Mγ 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 Mγ subgroups. Instead, Mα TFs interact with members of the Mβ and Mγ subgroups, whereas Mβ TFs and Mγ TFs barely interact, consistent with their intrinsic relatedness inherited from preduplicate 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 Mγ 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 Mγ. Genes encoding for the obligate Mγ-interacting Mα TFs (Mα* hereafter) were preferentially expressed in reproductive tissues and coexpressed with *PHE1/2* and other genes encoding for Mγ 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 coexpress with Mγ TFs in the endosperm, making it unlikely that they are able to form functional heterodimers with Mγ 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 Mγ 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 Mγ type TFs MADS87 and MADS89 and the interaction between the two Mα TFs and Mγ TFs is required for endosperm development (Paul et al. 2020). The two rice Mα genes as well as the two Mγ 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; supplementary fig. S1, Supplementary Material online). 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 Mγ-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 Mγ 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 Mγ genes, the duplicated Mα genes specialized in protein–protein interactions and subsequently the novel interacting pairs, Mα* and Mγ, together occupied the endosperm regulatory niche.

Although 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 nonflowering plants (fig. 1). Thus, the Mα-like genes in nonflowering plants have not undergone the diversification observed in angiosperms, so they likely represent the ancestral interacting partners of the preduplicated 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 Mγ 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; supplementary fig. S1, Supplementary Material online). 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 Mγ 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*, whereas 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 fig. S4, Supplementary Material online). 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 fig. S5, Supplementary Material online). 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), whereas 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 preduplicated 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.

Materials and Methods

Phylogenetic Analyses

Amino acid sequences of Type I and Type II MADS-box proteins of *A. thaliana* obtained from TAIR10 were used as

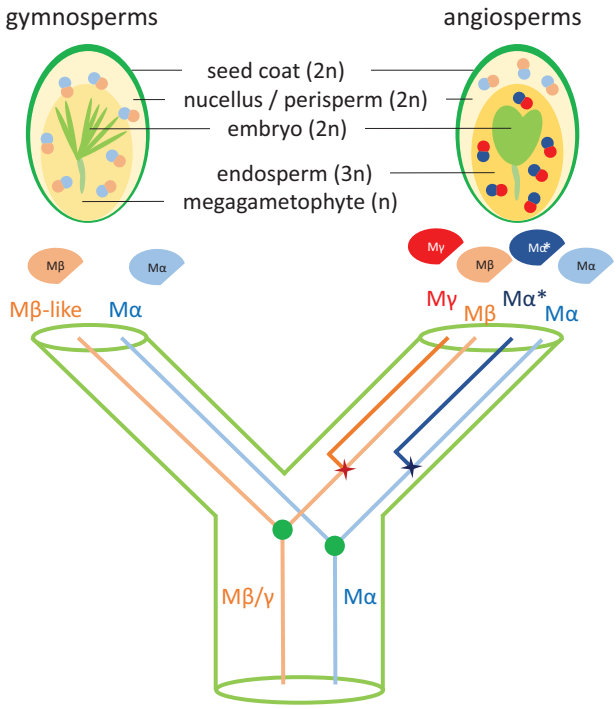

Fig. 4. Model depicting the role of Type I MADS-box TFs in the evolution of embryo nourishing tissues in seed plants. In early seed plants, ancestral Mα and Mβ/γ TFs likely formed heterodimers that regulated the embryo nourishing behavior. The gymnosperm Mα and Mβ-like TFs dimerize and function in maternal tissues. Angiosperms express Mα and Mβ heterodimers in maternal tissues, whereas Mα* and Mγ TFs that arose from angiosperm-specific duplications function in the endosperm and likely enabled endosperm evolution.

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/, last accessed December 18, 2021; Van Bel et al. 2018), Phytozome v.12 (https://phytozome.jgi.doe.gov/, last accessed December 18, 2021; Goodstein et al. 2012), CoGe (https://genomevolution.org/coge/, last accessed December 18, 2021; Grover et al. 2017), or other taxon-themed databases (supplementary table S1, Supplementary Material online). 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 (ML) 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). One thousand replicates of ultrafast bootstraps were applied to estimate the support for reconstructed branches (Hoang et al. 2018). The Mα, Mβ, and Mγ Type I 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 AU tests (Shimodaira 2002).

Expression Analyses

The expression data of Type I MADS-box genes in *A. thaliana* were extracted from Klepikova 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. 2020), coffee (Ivamoto et al. 2017), *N. thermarum* (Povilus and Friedman 2021), *Picea* (Nystedt et al. 2013), and *Gnetum* (Hou et al. 2019; Deng et al. 2020).

Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

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 (2017-04119) to C.K., a grant from the Knut and Alice Wallenberg Foundation

(2018-0206) to C.K., and support from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine to C.K.

Author Contributions

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

Data Availability

All data are incorporated into the article and its online supplementary material.

References

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA*. 97(10):5328–5333.

Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S. 2007. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8:242.

Baroux C, Grossniklaus U. 2019. Seeds—an evolutionary innovation underlying reproductive success in flowering plants. *Curr Top Dev Biol*. 131:605–642.

Baroux C, Spillane C, Grossniklaus U. 2002. Evolutionary origins of the endosperm in flowering plants. *Genome Biol*. 3(9):reviews1026.

Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, Köhler C. 2019. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* 8:e50541.

Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*. 29(3):464–489.

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci USA*. 110(5):E435–E444.

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

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

Chen C, Begcy K, Liu K, Folsom JJ, Wang Z, Zhang C, Walia H. 2016. Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol*. 171(1):606–622.

Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J. 2014. Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol*. 166(1):252–264.

Chen M, Lin JY, Wu X, Apuya NR, Henry KF, Le BH, Bui AQ, Pelletier JM, Cokus S, Pellegrini M, et al. 2021. Comparative analysis of embryo proper and suspensor transcriptomes in plant embryos with different morphologies. *Proc Natl Acad Sci USA*. 118(6):e2024704118.

Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L. 2008. AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J*. 54(6):1037–1048.

Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, Jiang N, Robin Buell C. 2012. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J*. 71(3):492–502.

de Folter S, Immink RG, Kieffer M, Parenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, et al. 2005. Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *Plant Cell*. 17(5):1424–1433.

Deng N, Hou C, He B, Ma F, Song Q, Shi S, Liu C, 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 Biol*. 20(1):531.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Flores-Vergara MA, Oneal E, Costa M, Villarino G, Roberts C, De Luis Balaguer MA, Coimbra S, Willis J, Franks RG. 2020. Developmental analysis of Mimulus seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. *Front Plant Sci*. 11:132.

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*). *Int J Mol Sci*. 20(17):4117.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 40(Database issue):D1178–D1186.

Gramzow L, Theißen G. 2013. Phylogenomics of MADS-Box genes in plants—two opposing life styles in one gene family. *Biology (Basel)* 2(3):1150–1164.

Gramzow L, Weilandt L, Theißen G. 2014. MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann Bot*. 114(7):1407–1429.

Grover JW, Bomhoff M, Davey S, Gregory BD, Mosher RA, Lyons E. 2017. CoGe LoadExp+: a web-based suite that integrates next-generation sequencing data analysis workflows and visualization. *Plant Direct* 1(2):7.

Hehenberger E, Kradolfer D, Köhler C. 2012. Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139(11):2031–2039.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35(2):518–522.

Hou C, Li L, Liu Z, Su Y, Wan T. 2020. Diversity and expression patterns of MADS-Box genes in *Gnetum luofuense*—implications for functional diversity and evolution. *Trop Plant Biol*. 13(1):36–49.

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

Irish VF, Litt A. 2005. Flower development and evolution: gene duplication, diversification and redeployment. *Curr Opin Genet Dev*. 15(4):454–460.

Ivamoto ST, Reis OJ, Domingues DS, Dos Santos TB, de Oliveira FF, Pot D, Leroy T, Vieira LG, Carazzolle MF, Pereira GA, 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.

Kalyaanamoorthy S, Minh BQ, Wong T, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.

Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347(2):183–198.

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J*. 88(6):1058–1070.

Köhler C, Dziasek K, Del Toro-De León G. 2021. Postzygotic reproductive isolation established in the endosperm: mechanisms, drivers and relevance. *Philos Trans R Soc Lond B Biol Sci*. 376(1826):20200118.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25(7):1307–1320.

Lu J, Magnani E. 2018. Seed tissue and nutrient partitioning, a case for the nucellus. *Plant Reprod*. 31(3):309–317.
Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 48(D1):D265–D268.
Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res*. 32(Web Server issue):W327–W331.
Nam J, dePamphilis CW, Ma H, Nei M. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol*. 20(9):1435–1447.
Nam J, Kim J, Lee S, An G, Ma H, Nei M. 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci USA*. 101(7):1910–1915.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, 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.
Pace L. 1907. Fertilization in Cypripedium. *Bot Gaz*. 44(5):353–374.
Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, 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. *Plant Cell*. 15(7):1538–1551.
Pattison RJ, Csukasi F, Zheng Y, Fei Z, van der Knaap E, Catalá C. 2015. Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. *Plant Physiol*. 168(4):1684–1701.
Paul P, Dhatt BK, Miller M, Folsom JJ, 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 Physiol*. 182(2):933–948.
Povilus RA, Friedman WE. 2021. Transcriptomes across fertilization and seed development in the water lily *Nymphaea thermarum* (Nymphaeales) reveal dynamic expression of DNA and histone methylation modifiers. bioRxiv. doi:10.1101/2021.04.04.438399.
Povilus RA, Losada JM, Friedman WE. 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. *Ann Bot*. 115(2):211–226.
Roszak P, Köhler C. 2011. Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proc Natl Acad Sci USA*. 108(51):20826–20831.
Ruelens P, de Maagd RA, Proost S, Theißen G, Geuten K, Kaufmann K. 2013. FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat Commun*. 4:2280.
Ruelens P, Zhang Z, van Mourik H, Maere S, Kaufmann K, Geuten K. 2017. The origin of floral organ identity quartets. *Plant Cell*. 29(2):229–242.
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 Sci*. 252:324–334.
Sakai H, Mizuno H, Kawahara Y, Wakimoto H, Ikawa H, Kawahigashi H, Kanamori H, Matsumoto T, Itoh T, Gaut BS. 2011. Retrogenes in rice (*Oryza sativa* L. ssp. japonica) exhibit correlated expression with their source genes. *Genome Biol Evol*. 3:1357–1368.
Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250(4983):931–936.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. 51(3):492–508.
Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, Winge P, Koncz C, Aalen RB, Schnittger A, et al. 2011. Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. *PLoS Genet*. 7(2):e1001303.
Sood SK, Mohana Rao PR. 1988. Studies in the embryology of the diandrous orchid *Cypripedium cordigerum* (Cypripedieae, Orchidaceae). *Plant Syst Evol*. 160(3–4):159–168.
Steffen JG, Kang IH, Portereiko MF, Lloyd A, Drews GN. 2008. AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol*. 148(1):259–268.
Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res*. 46(D1):D1190–D1196.
Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, Urich MA, Nery JR, Smith LG, Schnable JC, Ecker JR, et al. 2016. Integration of omic networks in a developmental atlas of maize. *Science* 353(6301):814–818.
Wan T, Liu ZM, Li LF, Leitch AR, Leitch IJ, Lohaus R, Liu ZJ, Xin HP, Gong YB, Liu Y, et al. 2018. A genome for gnetophytes and early evolution of seed plants. *Nat Plants*. 4(2):82–89.
Xu W, Dai M, Li F, Liu A. 2014. Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. *Nucleic Acids Res*. 42(11):6987–6998.
Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ, et al. 2017. The Apostasia genome and the evolution of orchids. *Nature* 549(7672):379–383.
Zhang MX, Zhu SS, Xu YC, Guo YL, Yang WC, Li HJ. 2020. Transcriptional repression specifies the central cell for double fertilization. *Proc Natl Acad Sci USA*. 117(11):6231–6236.