Comparative transcriptomics of seed nourishing tissues: uncovering conserved and divergent pathways in seed plants

Ana Marcela Florez-Rueda[1,2], Célia M. Miguel[3] and Duarte D. Figueiredo[1*]

[1] Max Planck Institute of Molecular Plant Physiology, Potsdam Science Park, Am Mühlenberg 1, 14476 Potsdam, Germany

[2] University of Potsdam, Karl-Liebknechts-Str. 24-25, Haus 26, 14476 Potsdam, Germany

[3] Biosystems & Integrative Sciences Institute (BioISI), Faculty of Sciences, University of Lisbon, Portugal

* figueiredo@mpimp-golm.mpg.de, +49-331-5678166

Summary

The evolutionary and ecological success of Spermatophytes is intrinsically linked to the seed habit, which provides a protective environment for the pre-development of the new generation. This environment includes an ephemeral nourishing tissue that supports embryo growth. In gymnosperms this tissue originates from the asexual proliferation of the maternal megagametophyte, while in angiosperms it is a product of fertilization, and is called the endosperm. The emergence of these nourishing tissues is of profound evolutionary value, and they are also food staples for most of the world's population. Here, using orthogroup inference, we provide a comparative transcriptomic analysis of seed nourishing tissues from representative species of all main angiosperm clades, including those of early diverging basal angiosperms, and of two gymnosperm representatives. Our results show that, although the structure and composition of seed nourishing tissues has seen significant divergence along evolution, there are signatures that are conserved throughout the phylogeny. Conversely, we identified processes that are exclusive to each clade, and thus illustrate their functional divergence. With this, we aimed to provide a foundation for future studies on the evolutionary history of seed nourishing structures, as well as a resource for gene discovery in new functional studies.

Significance Statement

Within seeds a specialized structure is responsible for nourishing the embryo, and is thus analogous to the mammalian placenta. Moreover, these nourishing tissues are also important sources of staple foods and feed. Here, we provide novel gene expression datasets of nourishing tissues of early diverging angiosperms, and use this information for a meta-analysis to identify pathways conserved, or divergent, throughout evolution. Thus, we aim to provide a resource for gene discovery for seed biology studies.

Keywords: seed, endosperm, megagametophyte, basal angiosperm, gymnosperm, orthogroup inference, transcriptomics.

Introduction

The seed nourishing tissues are specialized structures that provide nutrients and support for the developing embryo or for the germinating seedling. There is an outstanding diversity of plant seeds and strategies for nutrient transfer and storage (Linkies *et al.*, 2010; Chen *et al.*, 2017). In most angiosperms, these functions are carried out by the endosperm, the development of which, like that of the embryo, is derived from a fertilization event. However, this coupling to fertilization is an innovation of the angiosperms: gymnosperm seeds do not produce an endosperm, but their embryos are surrounded by a nutritive tissue that results from the proliferation of the megagametophyte (Linkies *et al.*, 2010; Rodrigues *et al.*, 2018).

These seed nourishing structures can be persistent or be consumed by the developing embryo. Seedlings of species with persistent endosperms, like cereals, rely on it for nourishment. Morphologically, those cereal endosperms are often larger and more prominent than those of other angiosperms, occupying a larger proportion of the mature seed. Contrastingly, in most eudicots, although serving a critical role during early seed development, the endosperm is much smaller, and is consumed as seeds mature. This leaves the cotyledons as the main storage tissue for the germinating seedling. In monocots, because the endosperm is typically the primary source of nutrients for the developing embryo, the cotyledon is often very small or absent (Sabelli P A *et al.*, 2005; Sabelli & Larkins, 2009). Like in most eudicots, in early divergent angiosperms such as *Amborella* and *Nymphaea*, the endosperm is not involved in nutrient storage but rather its main role is transfer (Floyd & Friedman, 2000; Povilus & Friedman, 2022). In fact, in *Nymphaea*, the storage function is handled by the perisperm, which derives from the ovule nucellus (Povilus *et al.*, 2015). This is also the case in other species with perispermic seed habits, like those of the Amaranthaceae family (Vandelook *et al.*, 2021).

The transcriptional landscape of seed endosperm is expected to vary between various plant clades and to reflect their diverse evolutionary histories and functional adaptations. Evolutionary convergence and divergence at the transcriptional level refer to the similarities and differences in gene expression patterns between different species (Ran *et al.*, 2018; García de la Torre *et al.*, 2021). Recently, evolutionary transcriptomics has greatly benefited from advances in orthologous inference pipelines such as OrthoFinder (Emms & Kelly, 2019). One major advantage is to accurately identify orthologous genes across large numbers of species, allowing for a more comprehensive comparative gene

expression analysis, and thus the potential to uncover kingdom-wide mechanisms (Julca *et al.*, 2021, 2023). Such studies have led to substantial insights into the evolution of gene regulation and the role of gene expression in phenotypic divergence between species (Ferrari *et al.*, 2019; Yu *et al.*, 2020; Gao *et al.*, 2021; García de la Torre *et al.*, 2021). The availability of large-scale transcriptomic datasets, combined with advances in computational and analytical tools, has enabled researchers to identify co-expression networks and functional modules across different species (Hansen *et al.*, 2014; Qiao *et al.*, 2016; Vercruysse *et al.*, 2020). The integration of the large amount of publicly available datasets with novel analytical methods is opening the way for the community to generate and test powerful hypotheses about gene function and evolution (Leebens-Mack *et al.*, 2019; Julca *et al.*, 2023).

In this study we leverage the diversity of seed-nourishing tissues and perform a multispecies comparative transcriptomic study. Our main question is what the transcriptional signatures are, that are conserved in different nourishing tissues across the plant phylogeny. Additionally, we aimed to identify particular transcriptional features of the nourishing tissues of divergent plant clades. We chose to examine distinct clades with different endosperm conformations to maximize the power of our inferences. Specifically, we used early divergent angiosperms represented by *Amborella trichopoda* (Amborella) and *Nymphaea caerulea* (water lily). We included maize (*Zea mays*) and rice (*Oryza sativa*) which possess a large and persistent starchy endosperm. As core eudicots we included the well-studied endosperm of *Arabidopsis thaliana* (Arabidopsis), and those of *Solanum peruvianum* (wild tomato) and *Mimulus guttatus* (monkeyflower). This selection also covers species with distinct patterns of endosperm development: nuclear, for Arabidopsis, rice and maize; and *ab initio* cellular, for the remaining species. As representatives of the gymnosperms, we studied the megagametophyte transcriptome of *Pinus pinaster* (maritime pine) and *Ginkgo biloba* (ginkgo). Our analyses provide a comparative framework to identify potential orthologs of genes of interest that are expressed in the seed nourishing structures of land plants. With this, we aim to provide a resource for future functional studies in seed biology.

Experimental Procedures

Extended materials and methods are provided as Supporting Information (Methods S1).

We sampled transcriptomes of the earliest available stages of nourishing tissues in species of all main angiosperm clades. We generated transcriptomes of laser microdissected endosperms and of leaves of *Nymphaea caerulea* and *Amborella trichopoda* (Figure S1), and of manually dissected megagametophytes of *Pinus pinaster*. We obtained transcriptomes of nourishing tissues and leaves from public repositories for *Gingko biloba*, *Arabidopsis thaliana*, *Solanum peruvianum*, *Mimulus guttatus, Oryza sativa* and *Zea mays*. Details of data provenance and availability are in Table S1. We inferred Orthogroups for all species using Orthofinder (Emms & Kelly, 2019). Mapping to reference genomes was performed with Hisat2 (Kim *et al.*, 2015), and the genome versions are described in Table S1. We performed differential gene expression analyses on each species comparing nourishing tissues vs. leaves using DESEQ (Love *et al.*, 2014). DEGs per species were translated to OG assignments, resulting in differentially expressed orthogroups (DEOGs). Using DEOGs we performed set operations to identify conserved and divergent OG sets among the species studied. Enrichment analyses and protein network inferences were performed in STRING (Szklarczyk *et al.*, 2017) using the best Arabidopsis ortholog Blast hit for the OGs. All additional data plotting and statistical tests were performed in R (version 4.2.0).

Results

Broad gene expression patterns discriminate the larger angiosperm clades: Monocots set the difference

Seed nourishing tissues have substantially diverged in morphology and developmental patterns throughout the more than 300 million years since angiosperms diverged from their common gymnosperm ancestor (Zimmer et al., 2007; Ran et al., 2018; Lubna et al., 2021). Nevertheless, we hypothesized that there should be common transcriptomic signatures that remained unchanged throughout evolution. To test this hypothesis, we compared the gene expression profiles of seed nourishing tissues to those of somatic ones (in this case leaves). We did this for species of all main angiosperm clades: monocots, eudicots and basal angiosperms; and included two diverging gymnosperm representatives, as we hypothesized that some of those transcriptomic signatures could already be present in the gymnosperm megagametophyte, which is the evolutionary precursor of the endosperm. We also included in the analysis the perisperm of *Nymphaea*,

because in this species it is the perisperm, and not the endosperm, that functions as a storage tissue. We used either publicly available datasets for the clades chosen, or generated our own, as described in the Experimental Procedures and Methods S1. For all species we used, or produced, datasets from nourishing tissues at the earliest available stages of seed development (Table S1, Figure S1).

To compare the transcriptional landscapes among tissues of divergent clades, we used Orthofinder (Emms & Kelly, 2019), and assigned orthologous genes to orthogroups (OGs). These OGs represent the set of genes from the species used in this study, which descended from a single gene in the last common ancestor of this set of species (Emms & Kelly, 2019). Eighty nine percent of the input genes were assigned to 34,041 OGs. Rice had the largest percentage of unassigned genes (16.7%) and the largest number of duplications were detected on the monocot node clade in the phylogenetic tree produced (3,405, Figure 1A-B). The species tree of Figure 1B reflects the individual OGs gene trees and the gene duplication events that were identified using our datasets. On the other hand, *Ginkgo* and *Pinus* revealed the smallest OG overlap with the other species included in the study while also displaying numerous gene duplication events (Figure 1 B-C). The largest OG overlap is between the two monocot species analyzed, sharing 13,291 OGs, followed by overlaps within the eudicot species (Figure 1C). Rice and maize displayed striking similarities and overlapping OG sets (total and differentially expressed), despite maize having almost twice the number of genes annotated and assigned to OGs, compared to rice (Figure 1A). Details of the OGs identified and the species gene correspondence can be found in Table S2.

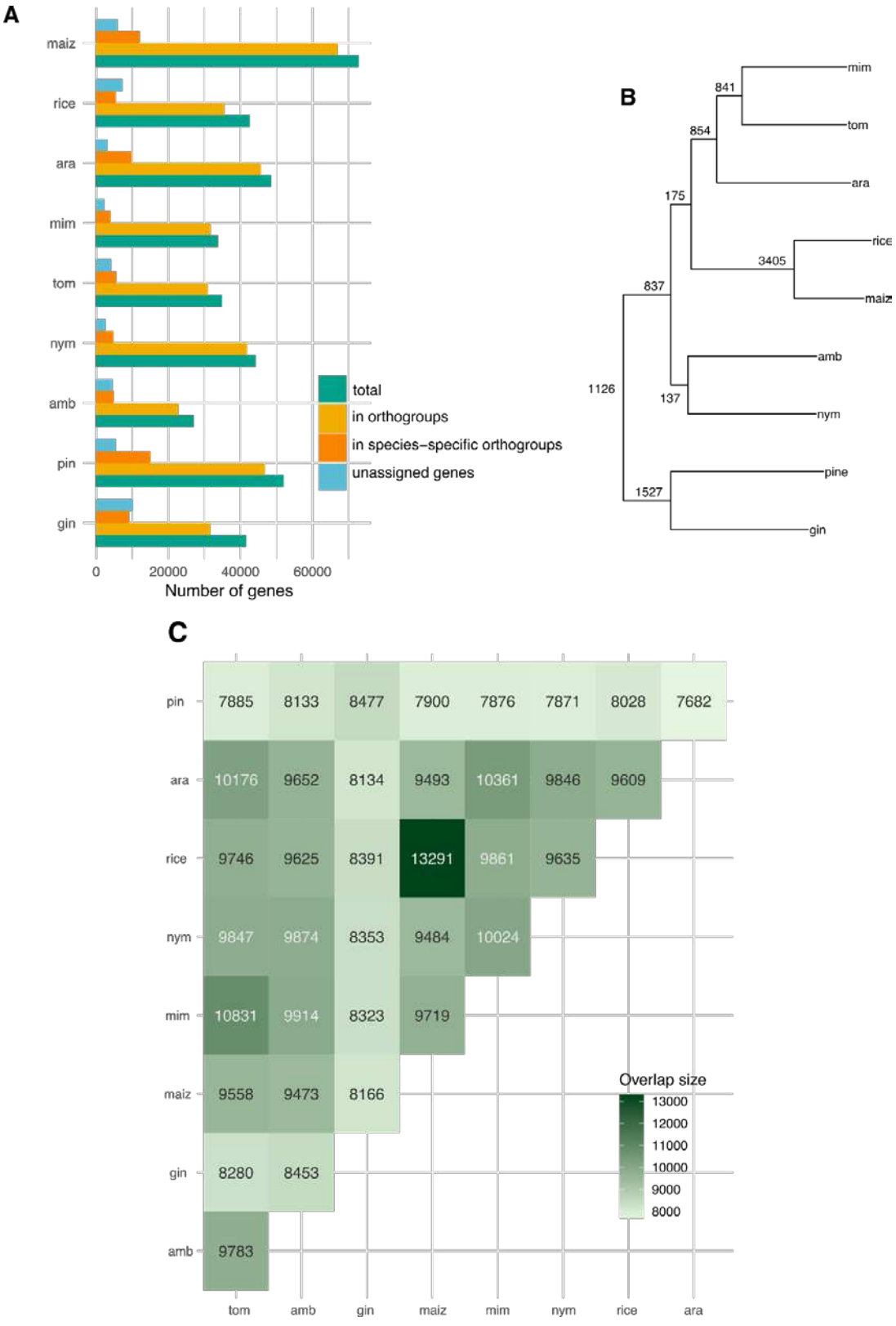


Figure 1. Orthology inference pipeline descriptive results. A. Overall statistics per species. B. Species trees with nodes displaying gene duplication events C. Overlap in the number of OGs in each pair of species. mim=*Mimulus*, nym=*Nymphaea*, amb=*Amborella*, gin=*Ginkgo*, pin=*Pine*, maiz=*Zea*, tom=*Solanum*, rice=*Oryza*, ara=*Arabidopsis.*

Having assigned genes into OGs, we used expression estimates for leaves and nourishing tissues and ran a PCA analysis to assess the variation among our datasets. Most of the variation in overall gene expression is captured by PC1, accounting for 34.83% of the total variance and exhibiting a significant correlation (0.75) with the corresponding clade (Figure 2A and Figure S2): the transcriptomic diversity of a tissue is determined by whether it belongs to a gymnosperm, early divergent angiosperm, eudicot, or monocot. Monocots are prominently located on the left side of the PCA (Figure 2A) and are distinct from other angiosperms, including from the more divergent plant clades. This pattern likely indicates the specialization of the large, starchy endosperm of cereals (Poaceae), which are the only representative of the clade in this study, although the divergence also extends to the somatic tissues. The PC6, which accounts for 3.49% of the variance, exhibits the second strongest and significant correlation (0.71) with the tissue variable (Figure S2). The top loadings of PC6 therefore define the transcriptional network of angiosperm nourishing tissues, and correspond to OGs annotated with functions related to nutrient storage: CRUCIFERIN 2 (CRU2), RMLC-LIKE CUPIN, CYSTEINE PEROXIDOREDOXIN 1 (PER1) and VICILIN-LIKE SEED STORAGE PROTEIN (PAP85; Figure S2 and Table S3).

Interestingly, the gymnosperm nourishing tissues cluster together with all angiosperm endosperms, regardless of their intrinsic biological differences, such as ploidy and parental origin. Although all eudicot nourishing tissues cluster together in the upper right quadrant of the PCA, there appears to be a close relationship in the overall patterns of gene expression of the basal endosperms *Amborella* and *Nymphaea*. This is in contrast to the more divergent samples of *Solanum*, *Arabidopsis*, and *Mimulus*, which probably reflects specialization of the nourishing tissues in the later diversified clades of Angiosperms.

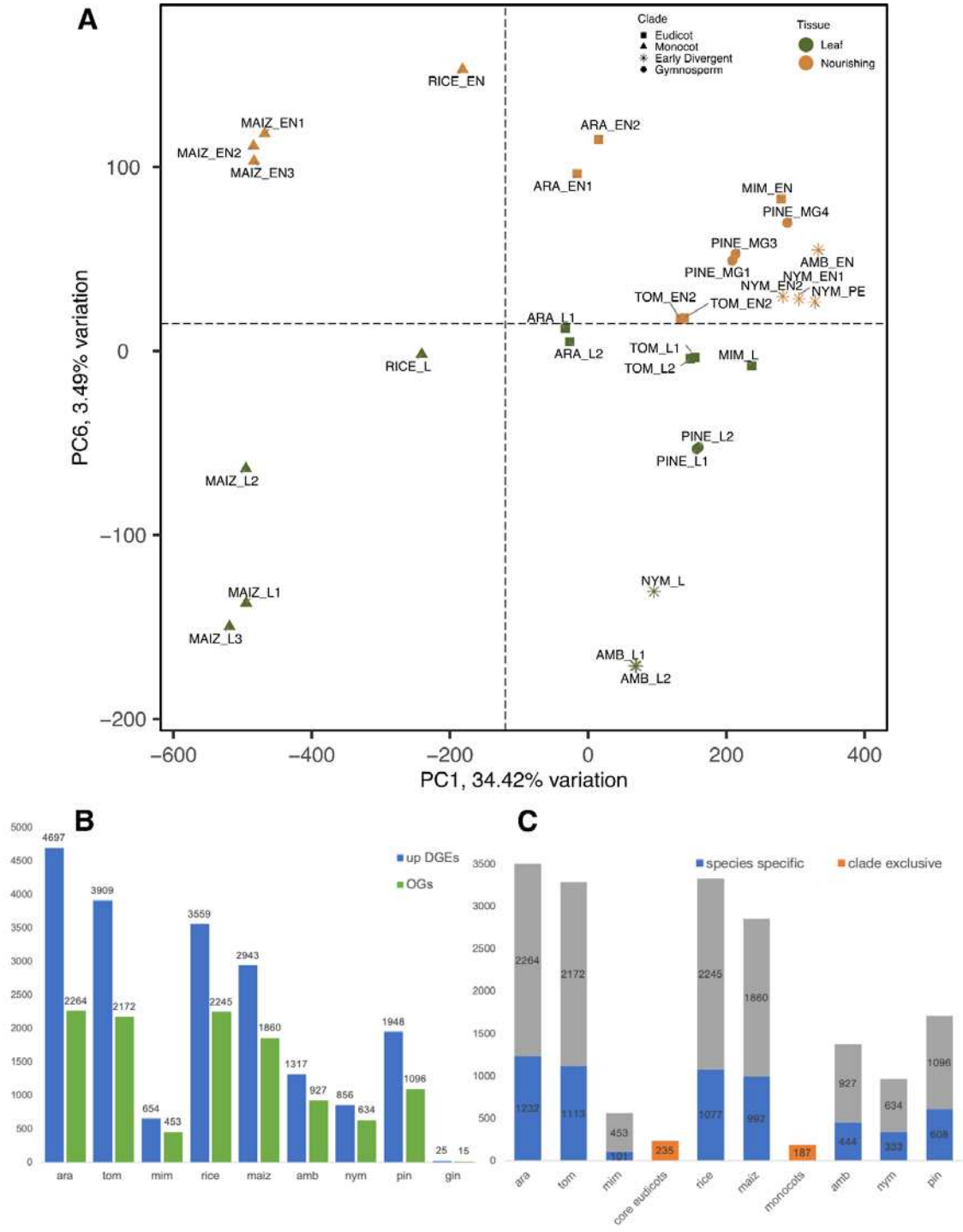


Figure 2. A. PCA describes broad gene expression patterns of photosynthetic vs. nourishing tissues in a broad range of land plants. B. Differentially expressed genes (DEGs, in blue) and corresponding differentially expressed orthogroups (DEOG in green) per species. C. DEOGs set description. Fractions of species specific OGs are depicted in blue, shared OGs in grey. mim=*Mimulus*, nym=*Nymphaea*, amb =*Amborella*, gin=*Ginkgo*, pin=*Pine*, maiz=*Zea*, tom=*Solanum*, rice=*Oryza*, ara=*Arabidopsis*, L= leaf, MG=megagametophyte, EN=endosperm, PE=perisperm.

Cross-species DGE analyses using OGs

To find transcriptional signatures that are specific for the seed nourishing tissues, for each species we performed an analysis of differential gene expression (DGE) between the transcriptomes of the leaves and of the nourishing tissues (Table S3). To enable cross-species comparisons, an OG identity was assigned to each DEG, delimiting Differentially Expressed OrthoGroups (DEOGs; Table S3). Some OGs consisted of multiple genes and some genes did not have OG assignment (Figure 1), resulting in less DEOGs than actual DEGs (Figure 2B). It is important to note that the number of identified DEOGs can be determined not exclusively by the biological properties of the taxa and tissue but also by technical limitations, such as genome annotation or library complexity, resulting in low numbers of identifiable DEOGs in the tissue. A significant proportion of the DEOGs were private, meaning they did not overlap between species (Figure 2C, Figure S2). DEOGs in the endosperms of rice and maize show the highest degree of overlap. This can be attributed to their similar transcriptional strategies, which also explains why they are grouped together on PC1 (Figure 2A). The second largest set of overlapping DEOGs is found in *Solanum* and *Arabidopsis* (Figure 2C, Figure S2), which also entail the best annotated eudicot genomes. In contrast, overlaps of *Mimulus* DEOGs with the other eudicot taxa were circa one order of magnitude lower (Figure S2). Due to this low number of shared features, the DEOGs of *Mimulus* were not included in the overlapping eudicot set (described below). Among all species tested, *Arabidopsis*, *Solanum*, *Oryza*, and *Zea* had the highest number of DEOGs in their endosperms (Figure 2B-C, Figure S2, Table S1, Table S3). We therefore focused our eudicot and monocot clade analyses on the shared sets of these taxa. Next, we investigate the identity of shared, non-shared, and private DEOGs between and within the main plant clades to describe the transcriptional networks of their nourishing tissues. For the sake of simplicity, the full transcriptional networks are provided as Supplementary Figures. Below we focus on a pair of relevant DEOG networks per comparative analysis.

The conserved transcriptional network of the angiosperm seed nourishing tissue

First, we identified 175 OGs that were enriched in angiosperm endosperms in at least one monocot, one dicot, and one early diverging taxon (Hypergeometric test p-value: 2,70E-02). We constructed functional clusters out of these OGs, to assess what biological functions are specifically enriched in seed nourishing tissues. Sixteen functional clusters

were identified, two of which were prominent and we describe here in more detail (Figure S3, Table S4). First, one cluster involved in cell cycle and nucleosome architecture drives the enrichment of the KEGG term “Nucleosome Core" (KW-0544; Figure S3, Table S4). At the core of this cluster are five histone-annotated OGs (Figure 3, Figure S3 and Table S3), which are functionally linked to two OGs representing subunits of the RNA Polymerase IV (RNAP IV). Also interacting with the main histone cluster are three OGs that represent A2 and type B cyclins, as well as UV-B-INSENSITIVE 4 (UVI4), a regulator of cyclin turnover. Additionally, there are four OGs representing components of the microtubule machinery. These OGs and the species-specific genes they encompass represent a gene network that is involved in cell cycle transitions and cell division in the angiosperm nourishing tissues. Interestingly, an OG annotated as SYP111/KNOLLE is also enriched. SYP111/KNOLLE is necessary for endosperm cellularization in Arabidopsis (Tiwari *et al.*, 2010; Park *et al.*, 2018), and our results thus indicate that its function is likely evolutionarily conserved.

To illustrate the structure of OGs in each of our target species, we created heatmaps displaying the expression values for all DEGs belonging to the OGs in the cluster (Figure 3 and Figure S4). OGs annotated as histones and cyclins, in particular, are composed of multiple genes per species. This pattern indicates their diversification in the endosperm. Other OGs in this cluster do not include a significant number of DEGs. For instance, those representing RNAP IV subunits or components of the microtubule assembly machinery typically have only one DEG in most species examined (Figure 3 and Figure S4). Importantly, histone variants are emerging as modulators of chromatin architecture (Borg *et al.*, 2021; Bhagyshree *et al.*, 2023). The histone proteins that we identified in this conserved cluster of enriched OGs are therefore strong candidates for determining the specific chromatin state of the endosperm.

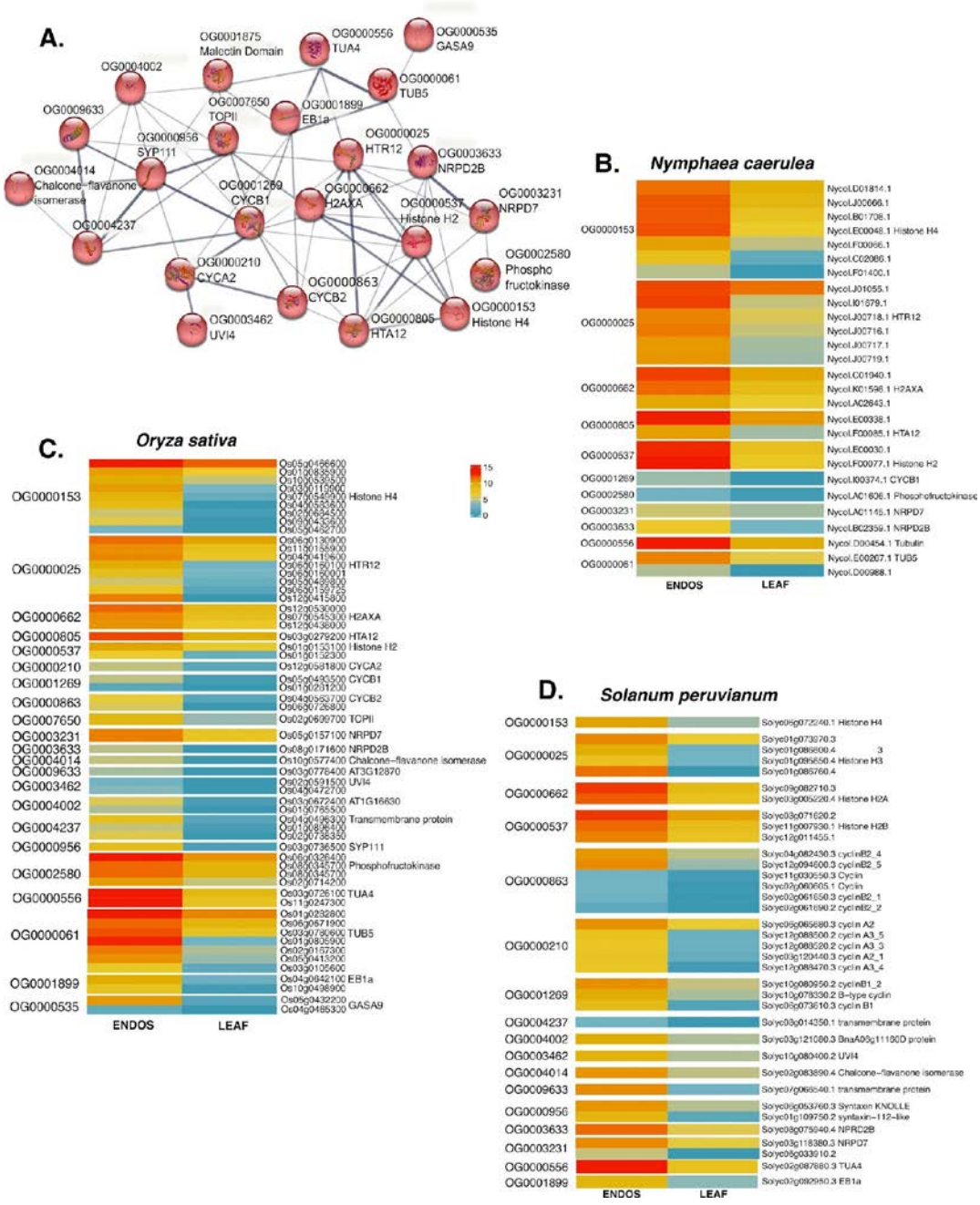


Figure 3. Conserved transcriptional network of the angiosperm seed nourishing tissue. A. Cell cycle and nucleosome architecture interacting network. B-D Orthogroup to DEG correspondence and expression values in CPM for B. *Nymphaea caerulea.* C. *Oryza sativa* and D. *Solanum peruvianum*.

A second important cluster conserved in all clades relates to nutrient storage (Figure 4A). The term "nutrient reservoir activity, and lipid droplet" was found enriched in the conserved sets of 175 OGs (STRING cluster CL:39149). This cluster comprises the OGs that drive most of the variation of the PC6 (Figure S2). These OGs are mostly annotated

as CUPINS, which are non-enzymatic proteins involved in seed storage (Dunwell, 1998). One example being CRU2 (Lin *et al.*, 2013). To display the relative expression of these genes in the endosperm we plotted heatmaps of expression of DGEs belonging to these OGs in all species studied (Figure 4B-D, Figure S5). Although most OGs in this cluster are single gene OGs, OG0000230, annotated as CRU2, shows multiple members in all species, evidencing diversification of cruciferin storage proteins in angiosperms. Interestingly, a well-defined set of rice storage glutelins also belong to this OG. The OG0007276 is best annotated as a Late Embryogenesis Abundant (LEA), LEA4-5, and is similarly present in the nutrient transfer cluster (Figure 4). Furthermore, three more OGs annotated as LEA proteins are present in the larger set of 175 OGs conserved expressed in endosperms (Table S4).

Interestingly, OGs related to Abscisic acid (ABA) signaling were also found on this interacting cluster. Specifically, ABSCISIC ACID INSENSITIVE 3 (ABI3), a transcription factor that participates in ABA-regulated gene expression during seed development (Mönke et al., 2012), and regulates seed storage genes (Lara *et al.*, 2003). The same was true for GLUTATHIONE S-TRANSFERASE 18 (GSTU18), as glutathione metabolism plays roles in seed dormancy and germination (Koramutla *et al.*, 2021). Other OGs of relevance in this cluster encode seed transporter proteins, such as AMINO ACID PERMEASE 8 (AAP8) and the sugar transporter SWEET9 (Table S4).

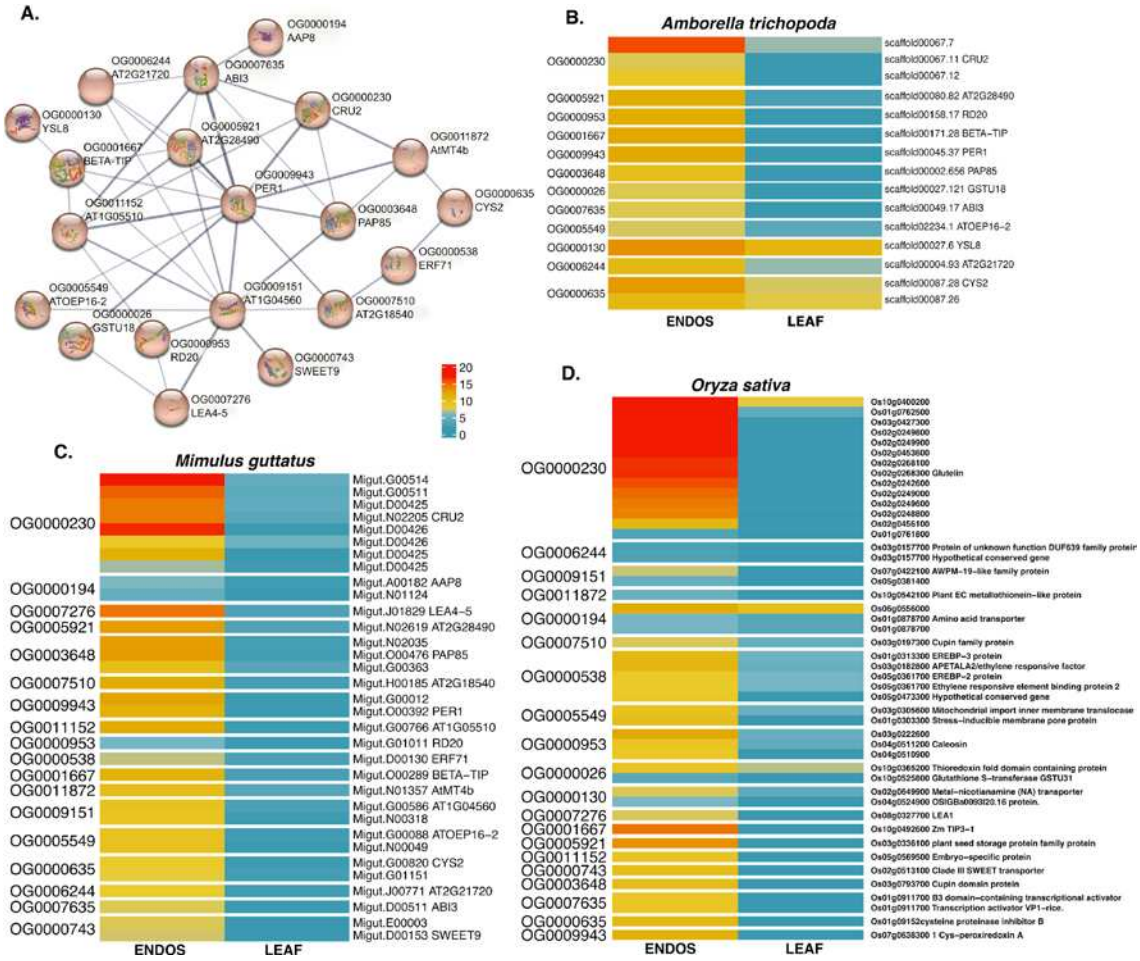


Figure 4. Conserved transcriptional network of the angiosperm seed nourishing tissue. A. Nourishing cluster protein interacting network. Orthogroup to DEG correspondence and expression values in CPM for B. *Amborella trichopoda*. C. *Mimulus guttatus* and D. *Oryza sativa*.

The shared transcriptional network of angiosperm nourishing tissues also includes four OGs that form a chaperone/heat shock cluster (Table S4, Figure S3). Moreover, there are a number of OGs encoding transcription factors (TFs) that are shared among angiosperms. These include the MYB-type TFs MYB65 and MYB100. Interestingly, only one OG annotated as a MADS-box transcription factor was present in the angiosperm common set, which was composed of a large set of SEPALLATA-like proteins across the angiosperms.

Likewise present was RGE1, also known as ZHOUPI. Moreover, two OGs annotated as subtilisin proteases (SBT1.1, SBT1.7) are consistently present in angiosperm endosperms and likely fulfill functions similar to that of another subtilase protein, ALE1 which together with ZHOUPI acts in an intercompartment signaling pathway between endosperms and embryos (Xing *et al.*, 2013; Moussu *et al.*, 2017). Other conserved OGs

in the angiosperm nourishing tissues but not forming clusters in the STRING protein network are proteins involved in hormone biosynthesis and signaling. These proteins include YUCCA11, involved in auxin biosynthesis, and the brassinosteroid response regulator BRASSINOSTEROID-RELATED HOMEOBOX 2 (Table S4).

The transcriptional network of the gymnosperm megagametophyte

As non-angiosperm seed plants, the gymnosperms are expected to exhibit some degree of evolutionary divergence. Their seeds contain a nourishing tissue resulting from the proliferation of the haploid female megagametophyte, which is not the product of fertilization. We aimed to uncover similarities and differences of the nourishing tissues of angiosperms and gymnosperms, using *Pinus pinaster* as the model gymnosperm species. For this, we intersected the set of exclusive DEOGs of the conserved angiosperm nourishing tissue (see above and in Table S4) and compared it to the full set of DEOGs of the pine megagametophyte. Sixty-four OGs that were conserved in angiosperms were also found as DEOGs in pine (Table S5). We found that members of the two main groups of conserved OGs in the angiosperm private network were partly present in the gymnosperm DEOG set. This included OGs annotated as histones, RNAP IV subunits and microtubule machinery (Table S5). The only OG corresponding to a MADS box TF that was found conserved in all angiosperms, annotated as SEP3, was not present in the DEOGs uncovered in pine. This was expected, given that *SEP* genes are specific to the angiosperms (Zahn *et al.*, 2005). Other sets of DEOGs conserved in angiosperms and absent in gymnosperm include those annotated as gibberellin regulated proteins, GASAs, and the subtilisin-like proteases SBT1.1 and SBT1.7.

Next, we identified 608 OGs that are unique to the nourishing tissue of gymnosperms (Hypergeometric test p-value: 1,49E-21, Figure 2C, Figure S2; Figure S6). Enrichment of several terms shows the specialization of pine megagametophyte in the biosynthesis of specific secondary metabolites (ATH01110 KEGG term “Biosynthesis of secondary metabolites"). Moreover, we found pine private OGs involved in pathways related to jasmonic acid (JA), alpha linoleic acid, ascorbic acid, and phenylpropanoid and flavonoid biosynthesis. Regarding JA, OGs related to both its biosynthesis and signaling were enriched (Fig. 5A). For example, this includes OGs annotated as LIPOXYGENASE 3 (LOX3), which is involved in JA biosynthesis, as well as for the JA effector JASMONATE-ASSOCIATED 10 (JAZ10). Importantly, JA has been shown to be

involved in the regulation of seed size and germination (Linkies & Leubner-Metzger, 2012; Pan *et al.*, 2020a; Hu *et al.*, 2021). Moreover, we identified a cluster involved in phenylpropanoid biosynthesis (Fig. 5B), a process tightly linked to lignin biosynthesis (Douglas, 1996). Although lignin has been found to play a role in angiosperm seed coats, the private occurrence of this cluster in the pine megagametophyte may be related to the exposed nature of the gymnosperm seed, and this particular cluster of genes may be involved in mediating its protection (Emonet & Hay, 2022). Flavonoids are also a product of the phenylpropanoid pathway, and we identified a cluster involved in their biosynthesis (Figure 5B). Finally, still regarding metabolite biosynthesis, the term "L-ascorbic acid biosynthesis process" was enriched in our analyses. Key OGs in this pathway include those annotated as GLUTATHIONE S-TRANSFERASE 20 (GSTU20) and DEHYDROASCORBATE REDUCTASE 2 (DHAR2; Figure 5B). In addition to their roles in ascorbic acid biosynthesis, these OGs may be involved in gymnosperm seed dormancy and germination, as part of the glutathione metabolism machinery (Koramutla *et al.*, 2021).

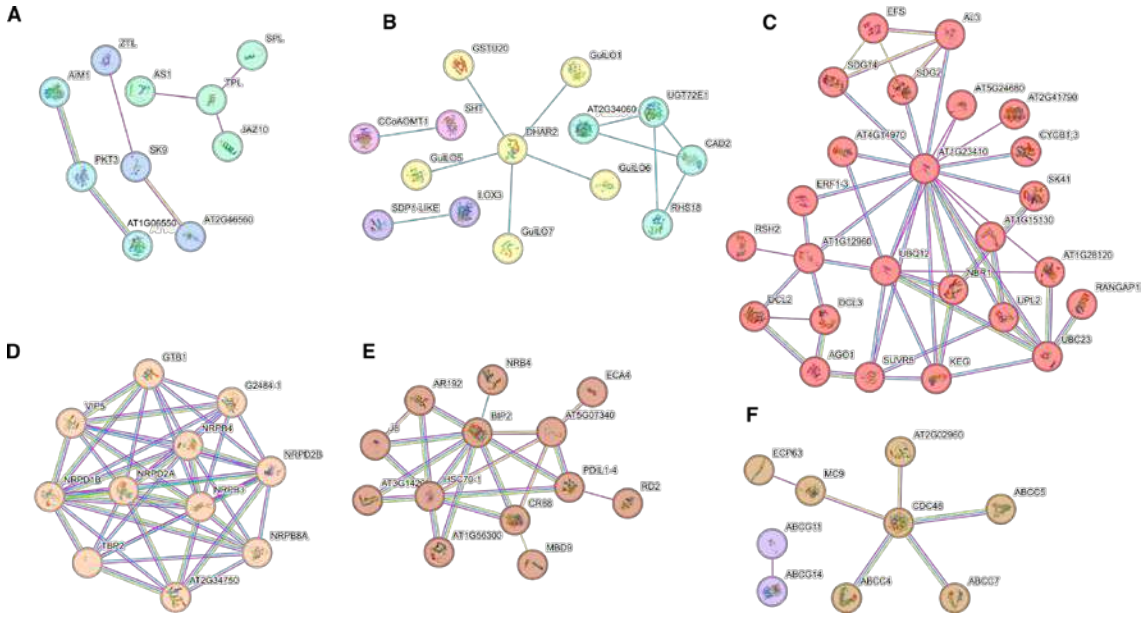


Figure 5. Orthogroup-based putative protein clusters private to the pine megagametophyte transcriptional network. A. Jasmonate related. B. Metabolite biosynthesis C. Epigenetic machinery D. RNA polymerases. E. Chaperones F. ABC-type transporters.

The gymnosperm nourishing tissues also entails a specific set of epigenetic machinery, as illustrated by the largest cluster identified in our analyses (Figure 5C). This

includes, for example, two OGs annotated as DICER-LIKE (DCL) proteins, and as ARGONAUTE 1 (AGO1). Moreover, this large interacting cluster of proteins includes four SET domain-containing OGs annotated as histone methyltransferases, namely, SET DOMAIN GROUP 2 (SDG2), EARLY FLOWERING IN SHORT DAYS (EFS), TRITHORAX 3 (SDG14) and SU(VAR)3-9-RELATED PROTEIN 5 (SUVR5; Figure 5C). This cluster also contains several OGs annotated as proteins belonging to the ubiquitin conjugation pathway, such as UBIQUITIN 12 (UBQ12) and UBIQUITIN-CONJUGATING ENZYME 23 (UBC23), among others (Figure 5C). These, together with a smaller cluster containing OGs annotated as SKP1-LIKE 9 (SK9) and ZEITLUPE (ZTL), represent the ubiquitination pathway in the pine megagametophyte. This pathway may be particularly important for the programmed cell death events that occur in *Pinus*, as in other conifers, whereby all the embryos resulting from cleavage polyembryony except the dominant one are eliminated (Filonova *et al.*, 2002), in a process in which the megagametophyte may play a role (Williams, 2009).

Further related to epigenetic processes, we detected a separate large cluster in the protein interaction network which is composed of eight OGs encoding subunits of the RNAPs II, IV, and V (Figure 5D). The identification of OGs highlighting crucial components associated with 24-nt siRNA production and the silencing machinery, such as DCL3, AGO1 and RNAPs IV and V, along with previous findings showing the presence of these small RNAs in these tissues (Rodrigues *et al.*, 2019), supports a role of the gymnosperm megagametophyte in RdDM. In angiosperms, 24-nt siRNAs, likely originating maternally, accumulate in endosperm tissues initially (Mosher *et al.*, 2009), while at later stages of development these are produced from both parental genomes (Rodrigues *et al.*, 2013; Xin *et al.*, 2014). Despite the distinctions between both nourishing tissue types, it appears that active silencing pathways associated with siRNAs are key, possibly to ensure genome stability and reproductive success.

One of the largest sets of interacting proteins uncovered in our analyses comprises eight OGs annotated as chaperones, as well as five OGs annotated as chaperones with DnaJ domains (Figure 5E). Interestingly, eight ATPase coupled ABC transporters are present in the OGs that are enriched in the pine megagametophyte (Figure 5F). They drive the enrichment of the GO term “ATPase-coupled transmembrane transporter activity”, and include OGs annotated as ABC transporters of Type 1 and Type 2, such as ABCG11 and ABCG14 (Figure 5F). Among other functions, ABC transporters have been

implicated in the mobilization of surface lipids, accumulation of phytate in seeds, and in transporting the phytohormones auxin and abscisic acid (Kang *et al.*, 2011; Do *et al.*, 2018).

The private transcriptional networks of the nourishing tissues of the early divergent angiosperm lineages, *Amborella and Nymphaea.*

The basal angiosperms investigated in this study are *Amborella* and *Nymphaea*. Although both are considered early divergent angiosperms, *Amborella* is regarded as the most basal of all angiosperms and forms a sister clade to all remaining extant flowering plants (Amborella Genome Project, 2013; The Angiosperm Phylogeny Group, 2016). The endosperms of both species, however, exhibit striking differences: water lilies exhibit a diploid endosperm with a reduced size, suggesting a more derived state compared to the larger triploid endosperm of *Amborella*. We examined the private transcriptional networks of their nourishing tissues separately due to these distinctive characteristics.

For *Amborella*, we detected 437 OGs that were expressed in its nourishing tissue but not in any of the other angiosperms studied here (Hypergeometric test p-value: 3,62E-14, Figure 2C, Figure S2). The most prominent cluster in the private transcriptional network of the *Amborella* endosperm involves proteins related to cell cycle processes and chromosome organization (Figure 6A, Figure S7, Table S6). Particularly relevant OGs are three annotated as cyclins, and two annotated as components of RNAPs, namely, INCURVATA2 (ICU2) and REVERSIONLESS 1 (REV1). Additionally, there are OGs annotated as proteins involved in chromosome organization, such as MINICHROMOSOME MAINTENANCE 5 (MCM5), MUTS HOMOLOG 7 (MSH7) and CURLY LEAF (CLF), a Polycomb Group protein (PcG). Another set of interacting OGs present in the protein interaction network are involved in transcriptional regulation and include DNA-DIRECTED RNA POLYMERASES SUBUNITS (Figure 6B). Interacting with those, there are two OGs annotated as TRANSCRIPTION INITIATION FACTORS, TFIIB2 and TFIIE (Table S6).

Metabolism-related OGs were also found enriched in the *Amborella* endosperm. For instance, there is a large cluster of proteins related to lipid biosynthesis (Figure 6C). Moreover, three OGs were annotated as lipoxygenases (LOX; Table S6, Figure 6G), which are typical of seeds and catalyze the oxidation of polyunsaturated fatty acids into functionally diverse oxylipins.

Importantly, a set of four OGs annotated as receptors harboring protein kinase domains were found exclusive in *Amborella*. These are STRUBBELIG-RECEPTOR FAMILY 3 (SRF3), RECEPTOR-LIKE PROTEIN KINASE 1 (RPK1), GASSHO1 (GSO1) and the putative brassinosteroid receptor BRI1-LIKE 1 (BRL1; Figure 6D). Other hormone-related OGs were also found enriched in the *Amborella* endosperm, namely those related to auxin and ethylene signaling. Some examples are AUXIN RESPONSE FACTOR 2 (ARF2), ETHYLENE-INSENSITIVE3-LIKE 1 (EIL1) and ETHYLENE RESPONSE FACTOR 71 (ERF71; Table S6).

A specialization of a subset of LEA proteins seems to be shared among all the clades studied. In the case of *Amborella,* OGs representing LEA proteins were found in the exclusive set of OGs in the endosperm, such as LEA3 (Figure 6F).

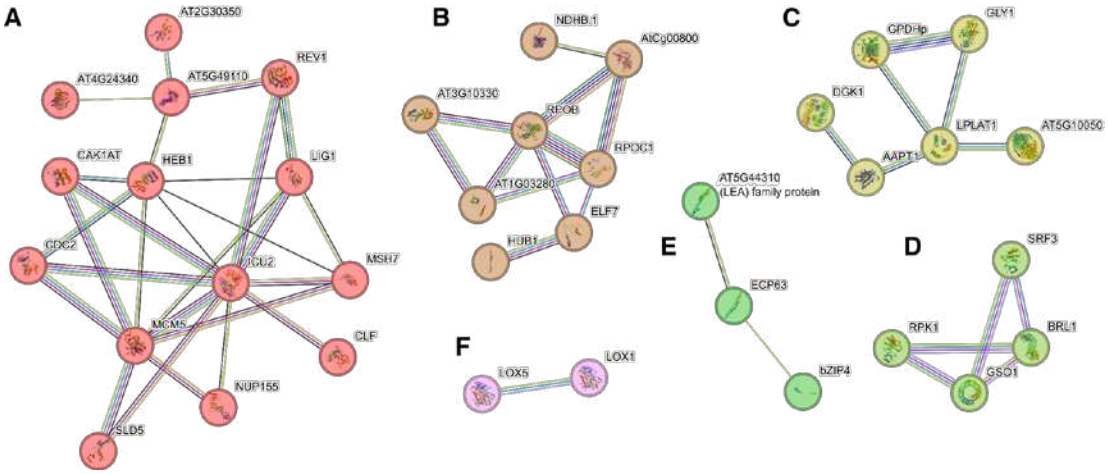


Figure 6. Orthogroup-based putative protein clusters private to the *Amborella* endosperm transcriptional network. A. Cell cycle and chromosome organization. B. DNA-directed RNA polymerases and associated proteins. C. Glycerolipid biosynthesis. D. Receptor Kinases. E. LEA proteins. F. Lipooxigenases.

Next, we identified the DEGs that are specific to the nourishing tissues of the *Nymphaea* seeds, the endosperm and the perisperm. At the time of collection, around a week after pollination, the endosperm was significantly smaller than the perisperm, which was filled with starch and oil deposits (Fig. S1). Our analyses of differential gene expression between these two tissues showed that the endosperm had more DEGs (2214) than the perisperm (826), likely reflecting a more active transcriptional state. There were commonalities in the sets of enriched terms in both the perisperm and endosperm, when compared to somatic tissues (Table S7). Shared functions include hormone metabolism and signaling, as illustrated by the enriched terms “ABA-activated signaling pathway”,

"Response to auxin" and "Hormone biosynthetic process" (Table S7). Also development-related terms were found enriched in both DEG sets, like "Meristem development" and "Plant organ development" (Table S7). The genes that are privately enriched in each tissue separately give us an overview of their specific functions in the *Nymphaea* seed. The perisperm was enriched in several secondary metabolism related processes, including "Flavonoid biosynthetic process", "Phenylpropanoid metabolic process" and "Ethylene biosynthetic process", pointing to a pivotal role for the perisperm in the biosynthesis of hormones and secondary metabolites. Likewise, peroxisome related terms were specific to the perisperm. Roles for the peroxisome in the perisperm may be related to hormone biosynthesis and fatty acid-oil body interactions (Pan et al., 2020). The endosperm had however a larger set of enriched terms correlating with the larger number of DGEs (Table S7). Among them we can highlight those with transcriptional regulation and epigenetic functions, like "Chromatin assembly", "RNA methylation", as well as those highlighting specific cell components, such as "Plasmodesmata", "Plant-type cell wall", "Plant-type vacuole" and "Microtubule". These DEGs between endosperm and perisperm clearly indicate a sub-functionalization of these two structures, with the endosperm being enriched in processes related to gene regulation and cellular dynamics, while the perisperm is mostly enriched in metabolism-related processes.

In addition to the DEOGs shared with other Angiosperms (Table S4), we identified transcriptional networks that were specific to the water lily nourishing tissues. We detected 334 OGs that were exclusively DE in the endosperm and perisperm (Hypergeometric Test p-value: 3,12E-09, Figure 2C, Figure S2). The protein interaction network of the *Nymphaea* endosperm displayed proteins involved in chromatin organization (Figure S8, Table S7). This was evidenced by a large cluster that includes OGs encompassing proteins with epigenetic roles, like HISTONE DEACETYLASE 9 (HDA9), PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1 (PIE1), a chromatin remodeler of the SWI2/SNF2 family, and the Flowering Locus VE (FVE), an MSI1 family protein and putative subunit of Polycomb Repressive Complexes (PRC) (Pazhouhandeh *et al.*, 2011). Members of the Cul4-RING E3 ubiquitin ligase complex are also present in this interaction cluster (Figure 7A), which may reflect interactions of endosperm-specific epigenetic machinery that involves PRC2 and Cul4-RING E3 interactions, as reported in Arabidopsis (Pazhouhandeh *et al.*, 2011). Still related to the

ubiquitination machinery, we identified a protein cluster with several components of the proteasome (Figure 7D).

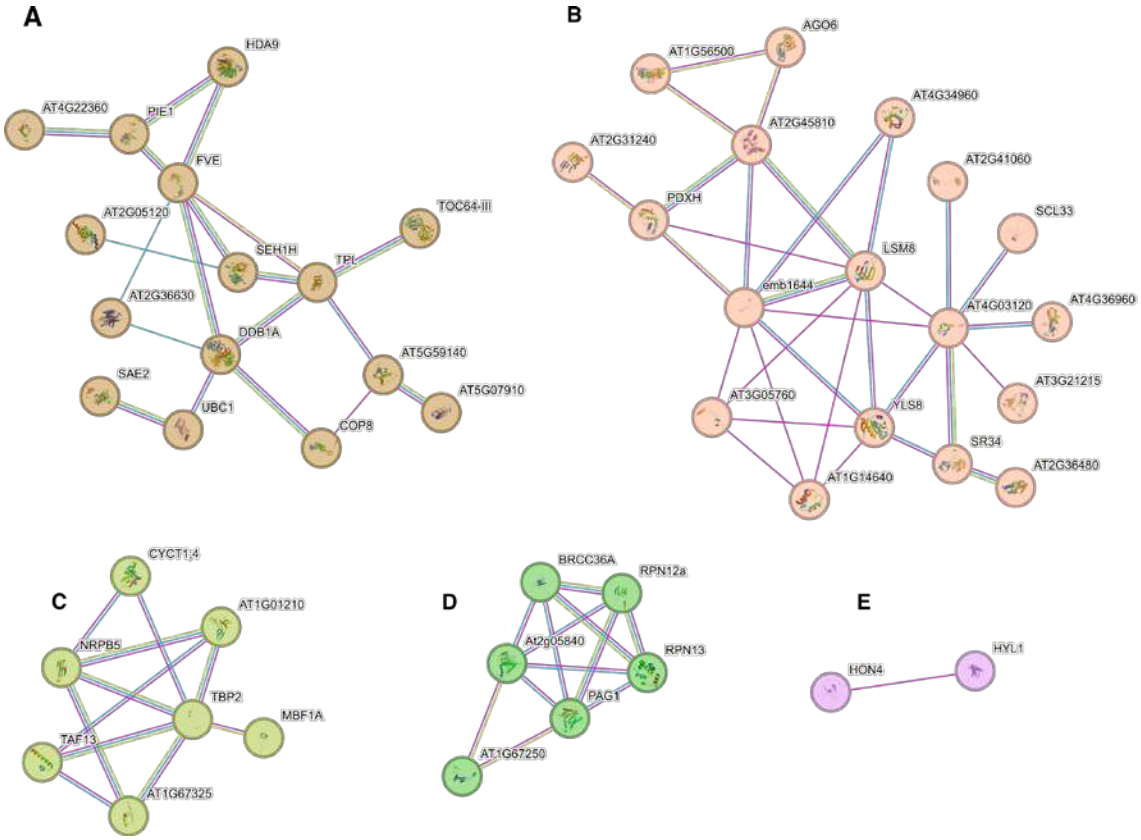


Figure 7. Orthogroup-based putative protein clusters private to the *Nymphaea* endosperm transcriptional network. A. Chromatin organization. B. Spliceosome. C. RNA Pol II. D. Proteasome. E. RdDM.

Another prominent functional cluster includes members of the spliceosome, driving the enrichment of the term “Interchromatin granule”, which in animal cells is linked to spliceosome activity (Mintz *et al.*, 1999). Another interesting cluster is composed of two OGs representing RNAPs, namely, DNA-DIRECTED RNA POLYMERASE, SUBUNIT M (ARCHAEAL) and RNA POLYMERASE II SUBUNIT 5 (NRPB5) (Figure 7C). Interacting with them are three OGs that encode accompanying proteins of the Polymerase II holoenzyme (Figure 7C). Additionally, two OGs putatively involved in RdDM were present in the private *Nymphaea* set, including the linker histone HON4/HYPONASTIC LEAVES 1 (HYL1) and ARGONAUTE 6 (AGO6; Figure 7E).

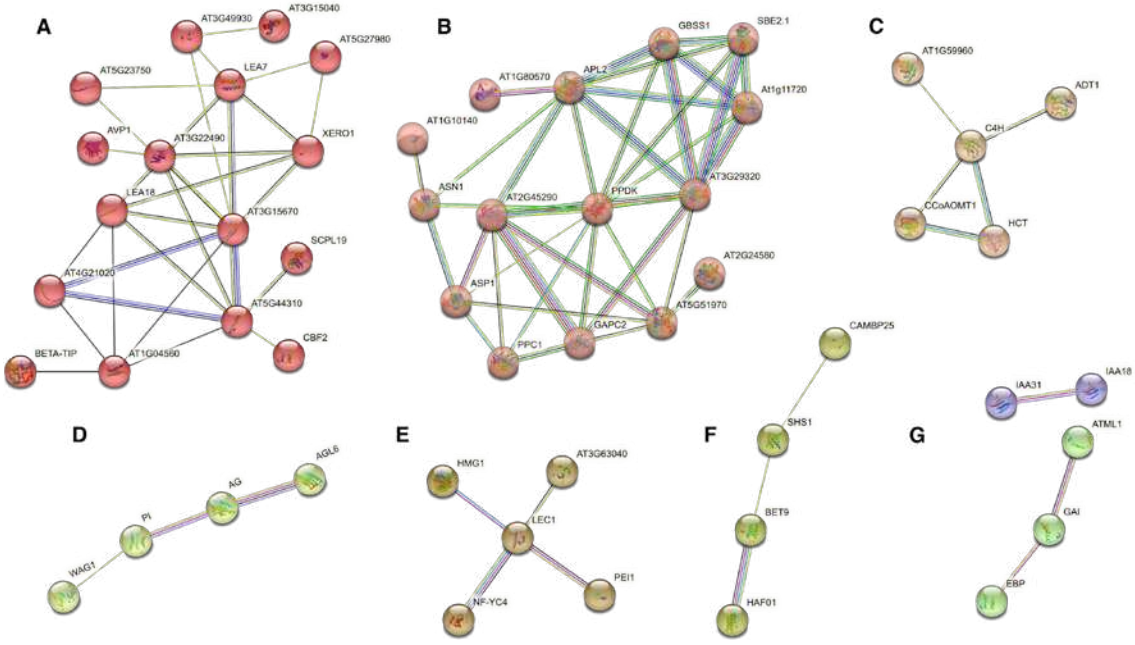
Transcriptional Landscapes in Monocots and Eudicots: Shared Features and Unique Patterns

We then proceeded to the realm of core angiosperms, monocots and eudicots. Here, we focus on describing the transcriptional signatures inherent to the endosperms of both clades, elucidating distinctive and common gene expression patterns.

The transcriptional signatures of the monocot endosperm showed strong differentiation to other seed plants, as illustrated by the PCA analysis (Figure 2A). It is worth highlighting that the monocot dataset is restricted to the Poaceae, which displays the large persistent endosperm of cereals. With the aim to describe its transcriptional network we identified DEOGs exclusively expressed in the monocot endosperms and, secondly, looked for OGs that were absent in monocots but present in eudicots. We thus identified 235 DEOGs private to the rice and maize endosperms (Hypergeometric test p-value: 6,45E-06, Figure 2C, Figure S2, Table S8). Interestingly, we identified a monocot-exclusive nutrient storage cluster (Figure 8A). At least seven of the OGs encode LEA proteins and dehydrins, which are LEA2-type proteins, evidencing a definitive role for these in monocot endosperm nutrient storage. Another important cluster includes many genes involved in processes related to the Krebs cycle and starch biosynthesis, processes that take place in the amyloplast of cereal endosperms (Figure 8B, Figure S9 and Table S8). These OGs drive the enrichment of the GO process terms “Starch biosynthetic process”, “Glycogen metabolic process”, and the GO Component term “Amyloplast” (Table S8). Another relevant cluster is related to flavonoid biosynthesis and other secondary metabolites (Figure 8C, Figure S9). The accumulation of flavonoids has been proposed to inhibit auxin transport and modulate seed development (Peer & Murphy, 2007). Most of these OGs had their best Orthofinder assigned target in the *Oryza* genome (Table S8).

Monocots appear to have a transcriptional regulatory machinery specialized in the endosperm. We found enriched the terms: “Regulation of transcription, DNA-templated” and “DNA-binding transcription factor activity” (Table S8). Moreover, we identified clusters of OGs that include putative transcription factors and other transcriptional regulators (Figure 8D-F, Figure S9, Table S8). Among those are three MADS domain containing OGs and two putative subunits of Nuclear Factor Y, the latter being a best hit for LEAFY COTYLEDON 1 (LEC1). Among other sets of transcriptional regulators found in this OG set, are bromodomain containing proteins which likely function as epigenetic regulators (Jarończyk *et al.*, 2021): two OGs containing bromodomains, the histone acetyltransferase TBP-ASSOCIATED FACTOR 1 (HAF01) and

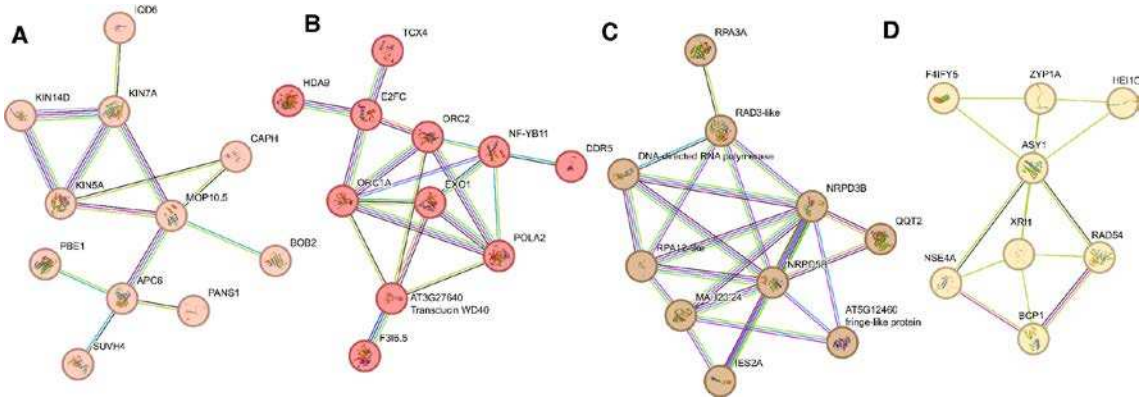
539 BROMODOMAIN AND EXTRATERMINAL DOMAIN PROTEIN 9 (BET9), were
540 found exclusively in the monocot endosperm (Figure 8F).
541 Additionally, we found OGs related to hormone signaling in the monocot nourishing
542 tissue (Figure 8G): two OGs corresponding to the INDOLE-3-ACETIC ACID
543 INDUCIBLE (IAA) family of proteins, implicated in auxin signaling, and OGs related to
544 gibberellin and ethylene signaling, like GIBBERELLIC ACID INSENSITIVE (GAI) and
545 ETHYLENE-RESPONSIVE ELEMENT BINDING PROTEIN (EBP).
546
547



548

549 Figure 8. Orthogroup-based putative protein clusters private to the monocot endosperm transcriptional
550 network. A.Storage. B. Starch biosynthesis / Amyloplast. C. Secondary metabolite biosynthesis. D. MADS-
551 box proteins. E. Other TFs. F. Bromodomain-containing proteins. G. Hormone signaling.
552
553 Likewise, we detected 187 DEOGs that were exclusive to the eudicot endosperms
554 (Hypergeometric test p-value: 8,10E-03, Figure 2C, Figure S2; Table S9). The term
555 "Microtubule cytoskeleton" was enriched in this DEOG set. Genes that drive its
556 enrichment and that appear to form a cluster of interacting proteins include three kinases,
557 KIN5A, KIN7A, KIN14D, and other important components of the cytoskeleton (Figure
558 9A, Figure S20). Specialization of the replication and transcription machinery in the
559 eudicot endosperm is also evidenced by the presence of two related protein clusters
560 (Figure 9B-C). One of them encompasses DEOGs involved in DNA replication (Figure
561 9B), which includes two origin-of-replication complexes, ORC1 and ORC2. Likewise,

562 POLA2, an alpha subunit of the DNA polymerase, and EXO1, an exonuclease, were
563 present in this replication cluster (Figure 9B). The other cluster is composed of several
564 DEOGs that are annotated as subunits of DNA-directed RNAPs and their accompanying
565 proteins (Figure 9C). This includes for example: NRPD3B, NRPD5B, RPA12-like and
566 RPA3A. Additionally, the term "Condensed chromosome" was found enriched. Its
567 corresponding DEOGs form a cluster of interacting proteins, which include
568 ASYNAPTIC1 (ASY1) and the SWI/SNF member RAD54 (Figure 9D). Other DEOGs
569 of importance are best annotated as transcription factors, including AGAMOUS-like 18
570 (AGL18), Homeobox-leucine zipper protein (ATHB-9), GLABRA 2 (GL2) and
571 REGULATOR OF AXILLARY MERISTEMS 2 (RAX2; Table S9).
572



573
574 Figure 9. Orthogroup-based putative protein clusters private to the eudicot endosperm transcriptional
575 network. A.Cytoskeleton. B. Replication. C. Transcription. D. Chromosome condensation.
576

577 With the aim to detect differences between monocots and eudicots we identified non-
578 exclusive DEOGs sets of the monocot and eudicot taxa. We found 568 DEOGs shared
579 between maize and rice and 574 DEOGs shared between the eudicots. We performed set
580 operations to disentangle the differences in the transcriptomic landscape of those taxa.
581 This analysis clarifies angiosperm differences, disregarding patterns predating the core
582 angiosperms that might obscure diversification within clades, if gymnosperms and early
583 divergent angiosperm DEOG sets were included. With the aim to determine which
584 DEOGs were specific to each of the two main angiosperm groups, monocots and eudicots,
585 we determined the complement DEOG sets of both clades. A total of 464 DEOGs were
586 found in eudicots and not in monocots (Table S9). Replicating the trend of the exclusive
587 set of DEOGs of core eudicots, we found five DEOGs annotated as subunits of RNAPs,
588 as well as nine members of the kinesin complex. Although microtubule cytoskeleton

proteins were part of the conserved transcriptional network of the angiosperm seed nourishing tissue (Table S4), the large diversification of kinesins was observed only in the core eudicots and not in monocots. The larger number of DEOGs belonging to these families, in this broader comparison (non-exclusive to other taxa), indicates that these OGs likely share ancestry with non-core angiosperm taxa such as gymnosperms and early divergent groups, rather than the more closely related monocots. This suggests an ancestral evolutionary role in the nourishing tissues for these OGs and their diversification within the eudicot lineage. Other DEOGs show the same trend of increased diversification in the eudicots in comparison to monocots. For example, five ABC-type transporters, two WUSCHEL RELATED HOMEOBOX (WOX) TFs and three DNA repair RAD-annotated DEOGs were found only in eudicots and not in monocots (Table S9).

Comparing only the DEOGs of core angiosperms, we then identified 459 monocot-specific DEOGs, including families such as ATL RING-H2 finger proteins, ethylene-responsive transcription factors (ERFs), Subtilisin-like proteases (SBTs), UDP-glycosyltransferases (UGTs), and ZAT family TFs (ZATs), showcasing examples of diversifying evolution within monocots. These families were notably absent in the core eudicots complement set of DEOGs (Table S8).

Discussion

The evolutionary advent of seeds is one of the factors that allowed land plants to reproduce in dry environments and, therefore, to occupy new ecological niches. In part, this is due to the protection of the embryo against environmental conditions by the sporophytic tissues of the maternal plant. Moreover, seeds allow for a state of dormancy, and only germinating when the conditions are favorable. But a major advance of the seed habit is the pre-development of the new sporophytic generation within the maternal tissues, where it is nourished by a specialized structure. Interestingly, already in seed ferns there were structures with a nourishing function, probably derived from the megagametophyte (Spencer *et al.*, 2013). In the gymnosperms the nourishing function is carried out by the proliferating megagametophyte, while in the angiosperms, this is done by the endosperm. The main difference being that the nourishing structure of the gymnosperms develops autonomously, i.e., its proliferation is not coupled to a

fertilization event, while the development of the angiosperm endosperm is coupled to fertilization. This coupling has obvious advantages, such as ensuring that nutrients are not allocated to "empty" seeds (without embryos), but it also allows for a coordinated development between the embryo and the endosperm (Ingram, 2020). It is also worth noting that the nourishing function in some angiosperms can also be provided by the perisperm, which originates from the sporophytic nucellus of the ovule. This is the case in *Nymphaea*, which we studied here, and in species of Amaranthaceae, like amaranth and quinoa (Burrieza *et al.*, 2014; Povilus *et al.*, 2015). Interestingly, it seems that in endospermic seeds, like those of Arabidopsis, the endosperm develops antagonistically with the nucellus, instructing its degeneration (Xu *et al.*, 2016). In seeds where the two structures co-exist, it is reasonable to hypothesize that they carry out complementary functions. This is supported by our comparative analysis of DEOGs between the endosperm and the perisperm of *Nymphaea*. Indeed, while the perisperm is mostly enriched in metabolism-related processes, the endosperm is enriched in gene regulation and cell-related processes, likely linked to its proliferative nature and it being a site for parental conflict (Povilus *et al.*, 2018; Köhler *et al.*, 2021). This suggests that in perispermic seeds, there is a subfunctionalization of these two structures, whereas in endospermic species these functions are fully fulfilled by the endosperm.

The fact that the seed nourishing structures of different clades share common functionalities, suggests that they probably also share some transcriptomic signatures. However, it is also expected that there are signatures that are specific to certain clades, since there is significant divergence in the form and function of some of those structures. This hypothesis is supported by our data, which shows that certain gene expression signatures are conserved throughout evolution, while others are private to the nourishing structures of each clade.

Conserved transcriptional features of the angiosperm endosperm

Seed storage proteins are a vital component of plant seeds and play a crucial role in ensuring the successful germination and growth of new plants. OGs encompassing genes that encode storage proteins were some of the main drivers of the differentiation of the nourishing tissue transcriptomes. Proteins assembled in these OGs and consistently present in all angiosperms studied are annotated as cruciferins, cupins and gluteins, but also as caleosins and peroxygenases, which are involved in the biogenesis and

maintenance of lipid bodies (Rahman et al. 2018). LEA proteins, although not fulfilling a specific storage function, were also part of the conserved angiosperm nutrient storage cluster. These proteins primarily function to safeguard cellular components and maintain seed viability under desiccation and dehydration stress (Cuming, 1999). Interestingly, diversification of LEA proteins in the monocot endosperm was evidenced by our analyses. It is important to note that some overlap or interactions between LEA proteins and nutrient storage components may exist (Dirk *et al.*, 2020). For instance, during seed desiccation and maturation, LEA proteins may interact with and stabilize storage proteins and lipids, helping to maintain their integrity and functionality (Cuming, 1999; Dirk *et al.*, 2020). Another group of seed proteins enriched in our analyses were lipoxygenases (LOX), which were first discovered in legumes and are prominent seed proteins (Casey, 1999). LOX-generated oxylipins, particularly JA and its derivatives, can modulate seed dormancy by influencing hormone signaling pathways and interacting with other dormancy-related regulators (Casey, 1999; Chauvin *et al.*, 2013). LOX specialization in the nourishing tissue as revealed by our analyses was not a conserved feature of angiosperms but rather appeared to be an ancestral trait present only in gymnosperms and *Amborella*. Instead of playing a critical role in core angiosperm endosperms, LOX proteins seem to play roles in plant defense, development and stress responses in vegetative tissues (Chauvin *et al.*, 2013).

A protein cluster involved in cell cycle and nucleosome assembly was also a highlight of the clusters shared by angiosperms. Strikingly, OGs annotated as histone variants were enriched in all angiosperm clades. Histone variants contribute to the definition of chromatin states and gene expression patterns. By modifying chromatin structure, they regulate the accessibility of DNA, influence gene expression, contribute to epigenetic inheritance, and participate in plant responses to environmental cues and developmental processes (Borg *et al.*, 2021; Bhagyshree *et al.*, 2023). Interestingly, histone variants were exclusively present in the shared protein network and not in any of the private sets of proteins specific to particular plant clades. This result may signify that the identified histone variant OGs represent the entire range of variants that play crucial, and therefore evolutionarily conserved, roles in angiosperm endosperms. Interestingly, genes encoding histone variants have been shown to be determinant for parental effects during embryogenesis in mammals (Molaro *et al.*, 2020). Moreover, histone variants have been shown to correlate with different chromatin states (Borg *et al.*, 2021), and, for instance,

687 the paternal expression of an H3 variant which is insensitive to PcG function is required
688 for reprogramming of the paternal germline (Borg *et al.*, 2020). This is particularly
689 relevant because the angiosperm endosperm is a site for genomic conflict between the
690 mother and the father, and the chromatin landscape of the endosperm underlies
691 reproductive barriers in angiosperms (Jiang *et al.*, 2017). Thus, it is tempting to
692 hypothesize that the differential expression of histone variants in the endosperm
693 correlates with the arisal of a parental conflict for nutrient allocation.
694
695 *Private transcriptional networks shed light into clade specializations*
696 The sets of private transcriptional networks in each major clade of seed plants potentially
697 represent genes that contributed to specialization within those clades. Genes that are
698 expressed in only a subset of plants within a restricted phylogenetic group should be
699 considered as potential candidates for functional studies. The private network of the
700 *Amborella* endosperm revealed OGs involved in epigenetic processes. Similarly, the
701 private network of *Nymphaea* consists of OGs that are annotated as RNA polymerases
702 and putative PcG components. These results highlight the evolutionary divergence of
703 these taxa from core angiosperms and precisely pinpoint the differences in the epigenetic
704 machinery that is active in their endosperms. In Arabidopsis, expression of components
705 of the FERTILIZATION-INDEPENDENT SEED PRC2 (FIS-PRC2), which are specific
706 to the gametophyte and the endosperm, progressively declines during seed development.
707 After a few mitotic rounds, FIS-PRC2 components like FIS2, MEA and FIE become
708 undetectable or limited to the chalazal cyst (Luo *et al.*, 2000). Given that we do not detect
709 differential expression of PRC2-encoding genes in the endosperms of other eudicots and
710 monocots, it is likely that those genes are also downregulated after fertilization, like in
711 Arabidopsis. However, the same may not be true for early diverging lineages, where we
712 do detect those genes expressed in the endosperm. This suggests that constitutive PRC2
713 activity in the endosperm may be the ancestral state, and its downregulation during
714 endosperm formation arose later during angiosperm speciation.
715 An evident hallmark of the nourishing tissue specialization was the biosynthesis of
716 secondary metabolites. For example, the term "Biosynthetic process" was enriched in the
717 exclusive sets of pine and in monocots. Regarding the transcriptional network of pine, it
718 demonstrated a wide range of biosynthetic pathways related to hormone signaling, such
719 as JA, and to structural characteristics of the seed. JAs are derived from linolenic acid

(Wasternack & Strnad, 2018) and, together with other phytohormones, like auxin and ABA, it has been shown to affect seed germination in Arabidopsis (Pan *et al.*, 2020a; Mei *et al.*, 2023). Furthermore, flavonoids have been suggested to play a regulatory role in phytohormone signaling (Appelhagen *et al.*, 2011; Li & Zachgo, 2013; Brunetti *et al.*, 2018). It is tempting to hypothesize that the specific proteins assembled in the OGs identified with secondary metabolite biosynthesis in monocots and pine may act as regulators of phytohormone signaling in their nourishing tissues.

The persistent nature of the monocot cereal endosperm leaves a distinct mark on their exclusive transcriptional network. A large protein cluster involved in processes that take place in the amyloplast of cereal endosperms, such as starch biosynthesis, was conspicuous in our analysis. We also identified two OG clusters annotated as TFs in the private transcriptional network of monocots. This highlights their functional role and possibly signals the diversification of these TFs in the monocot endosperm. The MADS-box and AP2-B3 protein families are well-studied transcription factor families that play important roles in regulating endosperm development (Lu *et al.*, 2012; Batista *et al.*, 2019; Yang *et al.*, 2020; Song *et al.*, 2021). Although we identified a single DEOG grouping a large set of AGAMOUS-LIKE proteins, AGL9, in the shared set of expressed OGs among all angiosperms, and AGL18 in the private transcriptional network of the eudicot endosperm, we did not observe any strong trend of diversification of OGs containing these TFs in the various endosperm of angiosperms. The pattern may be restricted to monocots, in which we identified three MADS domain containing DEOGs. This could suggest that remaining OGs encompassing additional MADS-box TFs have functional roles throughout the entire plant body and indeed play roles in the endosperms, but not specifically in it. Alternatively, it is very likely that genes encoding endosperm-specific MADS-box TFs were no longer expressed in the samples that we analyzed. Type I MADS-box TFs have been shown to play prominent roles in endosperm development, and their downregulation at the time of endosperm cellularization is crucial for this developmental transition (Erilova *et al.*, 2009; Hehenberger *et al.*, 2012; Batista *et al.*, 2019). Because the datasets that we analyzed originated in endosperms that were in early developmental stages, either cellular or close to undergoing cellularization, this likely explains why this family of genes is not prominent in our analysis.

Our work provides the first comparative transcriptomic analysis of early seed nourishing tissues, which includes representatives of all main angiosperm clades. We also

753 provide the first transcriptomic datasets of isolated endosperms of basal angiosperms. In
754 addition to the newly generated datasets, we present an orthogroup database
755 encompassing the main phylogenetic plant clades. These catalogs (private and shared
756 sets) will streamline the selection of candidate genes for functional genetic studies in
757 nourishing tissues across the plant kingdom, enabling targeted investigations and
758 enhancing our understanding of gene function and evolution.
759

760 ## Acknowledgements

761 We thank Bernhard Reiken at the Bonn botanical Garden for providing the *Amborella*
762 materials, and Ingo Kallemeier and Michael Burkart at the Potsdam Botanical Garden for
763 the *Nymphaea* materials. We are grateful to Barbara Schmidt and Prof. Julien Bachelier
764 at the Berlin Botanical Garden for guidance with *Nymphaea* material collection. We also
765 thank Arun Sampathkumar and Anja Froehlich for help with the LCM, and Andreia
766 Rodrigues for help with *Pinus* megagametophyte dissection.
767 This work was funded by a Humboldt Postdoctoral Fellowship to AMFR, by the Max
768 Planck Society to DDF, and by Fundação para a Ciência e a Tecnologia to CMM
769 (IF/01168/2013).
770

771 ## Competing interests

772 The authors have no competing interests.
773

774 ## Author contributions

775 AMFR and DDF designed the research. AMFR, CMM and DDF collected the plant
776 materials. AMFR performed the experiments and analyzed the data. AMFR wrote the
777 first draft of the manuscript, and all authors contributed to and approved the final version.
778

779 ## Data statement

780 Newly generated data for the endosperms and leaves of *Nymphaea caerulea* and
781 *Amborella trichopoda* as well as the megagametophyte transcriptomes of *Pinus pinaster*
782 have been posted in the NCBI SRA database under submission SUB13828753. All other
783 data used in this manuscript is publicly available and their sources described in Table S1.

784

785 # References

786 **Amborella Genome Project**. **2013**. The Amborella Genome and the Evolution of
787 Flowering Plants. *Science* **342**: 1241089.
788 **Appelhagen I, Lu G-H, Huep G, Schmelzer E, Weisshaar B, Sagasser M**. **2011**.
789 TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and
790 late steps of flavonoid biosynthesis in the endothelium of Arabidopsis thaliana
791 seeds. *The Plant Journal* **67**: 406–419.
792 **Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo
793 DD, Köhler C**. **2019**. The mads-box transcription factor pheres1 controls imprinting
794 in the endosperm by binding to domesticated transposons. *eLife* **8**.
795 **Bhagyshree J, J LZ, Elin A, Akihisa O, Vikas S, Ramesh Y, Svetlana A, Luisa KA,
796 Frédéric B**. **2023**. Histone variants shape the chromatin states in Arabidopsis. *eLife*
797 **12**.
798 **Borg M, Jacob Y, Susaki D, LeBlanc C, Buendía D, Axelsson E, Kawashima T,
799 Voigt P, Boavida L, Becker J, *et al.*2020**. Targeted reprogramming of H3K27me3
800 resets epigenetic memory in plant paternal chromatin. *Nature Cell Biology* **22**: 621–
801 629.
802 **Borg M, Jiang D, Berger F**. **2021**. Histone variants take center stage in shaping the
803 epigenome. *Current Opinion in Plant Biology* **61**: 101991.
804 **Brunetti C, Fini A, Sebastiani F, Gori A, Tattini M**. **2018**. Modulation of
805 Phytohormone Signaling: A Primary Function of Flavonoids in Plant–Environment
806 Interactions. *Frontiers in Plant Science* **9**.
807 **Burrieza HP, López-Fernández MP, Maldonado S**. **2014**. Analogous reserve
808 distribution and tissue characteristics in quinoa and grass seeds suggest convergent
809 evolution. *Frontiers in Plant Science* **5**.
810 **Casey R**. **1999**. Lipoxygenases. In: Shewry PR, Casey R, eds. Seed Proteins.
811 Dordrecht: Springer Netherlands, 685–708.
812 **Chauvin A, Caldelari D, Wolfender J-L, Farmer EE**. **2013**. Four 13-lipoxygenases
813 contribute to rapid jasmonate synthesis in wounded Arabidopsis thaliana leaves: a
814 role for lipoxygenase 6 in responses to long-distance wound signals. *New
815 Phytologist* **197**: 566–575.
816 **Chen F, Liu X, Yu C, Chen Y, Tang H, Zhang L**. **2017**. Water lilies as emerging
817 models for Darwin's abominable mystery. *Horticulture Research* **4**.
818 **Cuming AC**. **1999**. LEA Proteins. In: Shewry PR, Casey R, eds. Seed Proteins.
819 Dordrecht: Springer Netherlands, 753–780.
820 **Dirk LMA, Abdel CG, Ahmad I, Neta ICS, Pereira CC, Pereira FECB, Unêda-
821 Trevisoli SH, Pinheiro DG, Downie AB**. **2020**. Late Embryogenesis Abundant
822 Protein–Client Protein Interactions. *Plants* **9**: 814.
823 **Do THT, Martinoia E, Lee Y**. **2018**. Functions of ABC transporters in plant growth
824 and development. *Current Opinion in Plant Biology* **41**: 32–38.
825 **Douglas CJ**. **1996**. Phenylpropanoid metabolism and lignin biosynthesis: from weeds to
826 trees. *Trends in Plant Science* **1**: 171–178.
827 **Dunwell JM**. **1998**. Cupins: A New Superfamily of Functionally Diverse Proteins that
828 Include Germins and Plant Storage Proteins. *Biotechnology and Genetic
829 Engineering Reviews* **15**: 1–32.
830 **Emms DM, Kelly S**. **2019**. OrthoFinder: Phylogenetic orthology inference for

comparative genomics. *Genome Biology* **20**.

Emonet A, Hay A. **2022**. Development and diversity of lignin patterns. *Plant Physiology* **190**: 31–43.

Erilova A, Brownfield L, Exner V, Rosa M, Twell D, Scheid OM, Hennig L, Kohler C. **2009**. Imprinting of the polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. *PLoS Genetics* **5**: e1000663.

Ferrari C, Proost S, Janowski M, Becker J, Nikoloski Z, Bhattacharya D, Price D, Tohge T, Bar-Even A, Fernie A, *et al.*2019. Kingdom-wide comparison reveals the evolution of diurnal gene expression in Archaeplastida. *Nature Communications* **10**: 737.

Filonova LH, Von Arnold S, Daniel G, Bozhkov PV. **2002**. Programmed cell death eliminates all but one embryo in a polyembryonic plant seed. *Cell Death & Differentiation* **9**: 1057–1062.

Floyd SK, Friedman WE. **2000**. Evolution of Endosperm Developmental Patterns among Basal Flowering Plants. *International Journal of Plant Sciences* **161**: S57–S81.

Gao P, Quilichini TD, Yang H, Li Q, Nilsen KT, Qin L, Babic V, Liu L, Cram D, Pasha A, *et al.*2021. Evolutionary divergence in embryo and seed coat development of U's Triangle Brassica species illustrated by a spatiotemporal transcriptome atlas. *New Phytologist*.

García de la Torre VS, Majorel-Loulergue C, Rigaill GJ, Alfonso-González D, Soubigou-Taconnat L, Pillon Y, Barreau L, Thomine S, Fogliani B, Burtet-Sarramegna V, *et al.*2021. Wide cross-species RNA-Seq comparison reveals convergent molecular mechanisms involved in nickel hyperaccumulation across dicotyledons. *New Phytologist* **229**: 994–1006.

Hansen BO, Vaid N, Musialak-Lange M, Janowski M, Mutwil M. **2014**. Elucidating gene function and function evolution through comparison of co-expression networks of plants. *Frontiers in Plant Science* **5**.

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

Hu S, Yang H, Gao H, Yan J, Xie D. **2021**. Control of seed size by jasmonate. *Science China Life Sciences* **64**: 1215–1226.

Ingram GC. **2020**. Family plot: the impact of the endosperm and other extra-embryonic seed tissues on angiosperm zygotic embryogenesis. *F1000Research* **9**: F1000 Faculty Rev-18.

Jarończyk K, Sosnowska K, Zaborowski A, Pupel P, Bucholc M, Małecka E, Siwirykow N, Stachula P, Iwanicka-Nowicka R, Koblowska M, *et al.*2021. Bromodomain-containing subunits BRD1, BRD2, and BRD13 are required for proper functioning of SWI/SNF complexes in Arabidopsis. *Plant Communications* **2**: 100174.

Jiang H, Moreno-Romero J, Santos-González J, De Jaeger G, Gevaert K, Van De Slijke E, Köhler C. **2017**. Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in Arabidopsis thaliana. *Genes & Development* **31**: 1272–1287.

Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner A-C, Hackenberg D, Steinbachová L, Michaelidis C, Gomes Pereira S, Misra CS, *et al.*2021. Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nature Plants* **7**: 1143–1159.

Julca I, Tan QW, Mutwil M. 2023. Toward kingdom-wide analyses of gene expression. *Trends in Plant Science* **28**: 235–249.

Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y, Martinoia E. 2011. Plant ABC Transporters. *The Arabidopsis Book / American Society of Plant Biologists* **9**: e0153.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods* **12**: 357–360.

Köhler C, Dziasek K, Del Toro-De León G. 2021. Postzygotic reproductive isolation established in the endosperm: mechanisms, drivers and relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**: 20200118.

Koramutla MK, Negi M, Ayele BT. 2021. Roles of Glutathione in Mediating Abscisic Acid Signaling and Its Regulation of Seed Dormancy and Drought Tolerance. *Genes* **12**: 1620.

Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J. 2003. Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *The Journal of Biological Chemistry* **278**: 21003–21011.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, *et al.*2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Li S, Zachgo S. 2013. TCP3 interacts with R2R3-MYB proteins, promotes flavonoid biosynthesis and negatively regulates the auxin response in Arabidopsis thaliana. *The Plant Journal* **76**: 901–913.

Lin Y, Pajak A, Marsolais F, McCourt P, Riggs CD. 2013. Characterization of a Cruciferin Deficient Mutant of Arabidopsis and Its Utility for Overexpression of Foreign Proteins in Plants. *PLoS ONE* **8**: e64980.

Linkies A, Graeber K, Knight C, Leubner-Metzger G. 2010. The evolution of seeds. *New Phytologist* **186**: 817–831.

Linkies A, Leubner-Metzger G. 2012. Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. *Plant Cell Reports* **31**: 253–270.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lu J, Zhang C, Baulcombe D, Chen Z. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 5529–5534.

Lubna, Asaf S, Khan AL, Jan R, Khan A, Khan A, Kim K, Lee I. 2021. The dynamic history of gymnosperm plastomes: Insights from structural characterization, comparative analysis, phylogenomics, and time divergence. *The Plant Genome* **14**: e20130.

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A. 2000. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proceedings of the National Academy of Sciences* **97**: 10637–10642.

Mei S, Zhang M, Ye J, Du J, Jiang Y, Hu Y. 2023. Auxin contributes to jasmonate-mediated regulation of abscisic acid signaling during seed germination in Arabidopsis. *The Plant Cell* **35**: 1110–1133.

Mintz PJ, Patterson SD, Neuwald AF, Spahr CS, Spector DL. 1999. Purification and biochemical characterization of interchromatin granule clusters. *The EMBO Journal*

18: 4308–4320.

Molaro A, Wood AJ, Janssens D, Kindelay SM, Eickbush MT, Wu S, Singh P, Muller CH, Henikoff S, Malik HS. **2020**. Biparental contributions of the H2A.B histone variant control embryonic development in mice. *PLoS Biology* **18**: e3001001.

Mönke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hähnel U, Junker A, Weisshaar B, Conrad U, Bäumlein H, *et al.*2012. Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Research* **40**: 8240–8254.

Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC. **2009**. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* **460**: 283–286.

Moussu S, Doll NM, Chamot S, Brocard L, Creff A, Fourquin C, Widiez T, Nimchuk ZL, Ingram G. **2017**. ZHOUPI and KERBEROS Mediate Embryo/Endosperm Separation by Promoting the Formation of an Extracuticular Sheath at the Embryo Surface. *The Plant Cell* **29**: 1642–1656.

Pan J, Hu Y, Wang H, Guo Q, Chen Y, Howe GA, Yu D. **2020a**. Molecular Mechanism Underlying the Synergetic Effect of Jasmonate on Abscisic Acid Signaling during Seed Germination in Arabidopsis. *The Plant Cell* **32**: 3846–3865.

Pan R, Liu J, Wang S, Hu J. **2020b**. Peroxisomes: versatile organelles with diverse roles in plants. *New Phytologist* **225**: 1410–1427.

Park M, Krause C, Karnahl M, Reichardt I, El Kasmi F, Mayer U, Stierhof Y-D, Hiller U, Strompen G, Bayer M, *et al.*2018. Concerted Action of Evolutionarily Ancient and Novel SNARE Complexes in Flowering-Plant Cytokinesis. *Developmental Cell* **44**: 500-511.e4.

Pazhouhandeh M, Molinier J, Berr A, Genschik P. **2011**. MSI4/FVE interacts with CUL4–DDB1 and a PRC2-like complex to control epigenetic regulation of flowering time in Arabidopsis. *Proceedings of the National Academy of Sciences* **108**: 3430–3435.

Peer WA, Murphy AS. **2007**. Flavonoids and auxin transport: modulators or regulators? *Trends in Plant Science* **12**: 556–563.

Povilus RA, Diggle PK, Friedman WE. **2018**. Evidence for parent-of-origin effects and interparental conflict in seeds of an ancient flowering plant lineage. *Proceedings of the Royal Society B: Biological Sciences* **285**: 20172491.

Povilus RA, Friedman WE. **2022**. Transcriptomes across fertilization and seed development in the water lily Nymphaea thermarum (Nymphaeales): evidence for epigenetic patterning during reproduction. *Plant Reproduction*.

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. *Annals of Botany* **115**: 211–226.

Qiao Z, Pingault L, Nourbakhsh-Rey M, Libault M. **2016**. Comprehensive Comparative Genomic and Transcriptomic Analyses of the Legume Genes Controlling the Nodulation Process. *Frontiers in Plant Science* **7**: 34.

Ran J-H, Shen T-T, Wang M-M, Wang X-Q. **2018**. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proceedings of the Royal Society B: Biological Sciences* **285**: 20181012.

Rodrigues AS, Chaves I, Costa BV, Lin Y-C, Lopes S, Milhinhos A, Van De Peer Y, Miguel CM. **2019**. Small RNA profiling in Pinus pinaster reveals the

transcriptome of developing seeds and highlights differences between zygotic and somatic embryos. *Scientific Reports* **9**: 11327.

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D. **2013**. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proceedings of the National Academy of Sciences* **110**: 7934–7939.

Rodrigues AS, Vega JJ, Miguel CM. **2018**. Comprehensive assembly and analysis of the transcriptome of maritime pine developing embryos. *BMC Plant Biology* **18**: 1–20.

Sabelli PA, Larkins BA. **2009**. The Development of Endosperm in Grasses. *Plant Physiology* **149**: 14–26.

Sabelli P A, Leiva-Neto J T, Dante R A, Nguyen H, Larkins B A. **2005**. Cell cycle regulation during maize endosperm development. *Maydica* **50**: 485–496.

Song J, Xie X, Chen C, Shu J, Thapa RK, Nguyen V, Bian S, Kohalmi SE, Marsolais F, Zou J, *et al.*2021. LEAFY COTYLEDON1 expression in the endosperm enables embryo maturation in Arabidopsis. *Nature Communications* **12**.

Spencer ART, Wang S-J, Dunn MT, Hilton J. **2013**. Species of the medullosan ovule Stephanospermum from the Lopingian (late Permian) floras of China. *Journal of Asian Earth Sciences* **76**: 59–69.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, *et al.*2017. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45**: D362–D368.

The Angiosperm Phylogeny Group. **2016**. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.

Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, Salazar A, Zhang K, Pennell R, Scott RJ. **2010**. Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. *BMC Plant Biology* **10**: 72.

Vandelook F, Newton RJ, Bobon N, Bohley K, Kadereit G. **2021**. Evolution and ecology of seed internal morphology in relation to germination characteristics in Amaranthaceae. *Annals of Botany* **127**: 799–811.

Vercruysse J, Van Bel M, Osuna-Cruz CM, Kulkarni SR, Storme V, Nelissen H, Gonzalez N, Inzé D, Vandepoele K. **2020**. Comparative transcriptomics enables the identification of functional orthologous genes involved in early leaf growth. *Plant Biotechnology Journal* **18**: 553–567.

Wasternack C, Strnad M. **2018**. Jasmonates: News on Occurrence, Biosynthesis, Metabolism and Action of an Ancient Group of Signaling Compounds. *International Journal of Molecular Sciences* **19**: 2539.

Williams CG (Ed.). **2009**. Syngamy, Embryo Development and Seed Dispersal. In: Conifer Reproductive Biology. Dordrecht: Springer Netherlands, 107–121.

Xin M, Yang R, Yao Y, Ma C, Peng H, Sun Q, Wang X, Ni Z. **2014**. Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. *BMC Plant Biology* **14**: 192.

Xing Q, Creff A, Waters A, Tanaka H, Goodrich J, Ingram GC. **2013**. ZHOUPI controls embryonic cuticle formation via a signalling pathway involving the subtilisin protease ABNORMAL LEAF-SHAPE1 and the receptor kinases GASSHO1 and GASSHO2. *Development (Cambridge, England)* **140**: 770–779.

Xu W, Fiume E, Coen O, Pechoux C, Lepiniec L, Magnani E. **2016**. Endosperm and

1027 Nucellus Develop Antagonistically in Arabidopsis Seeds. *The Plant Cell*:
1028 tpc.00041.2016.
1029 **Yang T, Guo L, Ji C, Wang H, Wang J, Zheng X, Xiao Q, Wu Y**. **2020**. The B3
1030 domain-containing transcription factor ZmABI19 coordinates expression of key
1031 factors required for maize seed development and grain filling. *The Plant Cell* **33**:
1032 104–128.
1033 **Yu Y, Hu H, Doust AN, Kellogg EA**. **2020**. Divergent gene expression networks
1034 underlie morphological diversity of abscission zones in grasses. *The New*
1035 *Phytologist* **225**: 1799–1815.
1036 **Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE,**
1037 **dePamphilis CW, Ma H**. **2005**. The Evolution of the SEPALLATA Subfamily of
1038 MADS-Box GenesSequence data from this article have been deposited with the
1039 EMBL/GenBank Data Libraries under accession nos. AY850178, AY850179,
1040 AY850180, AY850181, AY850182, AY850183, AY850184, AY850185,
1041 AY850186. *Genetics* **169**: 2209–2223.
1042 **Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA**. **2007**. Dating the
1043 early evolution of plants: detection and molecular clock analyses of orthologs.
1044 *Molecular Genetics and Genomics* **278**: 393–402.
1045

Supplementary Information

Methods S1

Table S1. Data provenance, tissue details, number of total mapped reads, total number of genes detected, DEGs and DEOGs.

Table S2. Orthogroup description.

Table S3. Results of differential gene expression between nourishing tissues and leaves of all species studied.

Table S4. Conserved transcriptional network of the angiosperm seed nourishing tissue.

Table S5. The transcriptional network of the *Pinus pinaster* megagametophyte.

Table S6. The private transcriptional network of the *Amborella trichopoda* endosperm. Table S7 – The transcriptional network of the *Nymphaea caerulea* seed.

Table S8. The monocot endosperm transcriptional network.

Table S9. The transcriptional network of the eudicot endosperm.

Figure S1. Examples of seeds of *Amborella trichopoda* and *Nymphaea caerulea* used for microdissection and transcriptome generation.

Figure S2. Details of PCA analyses and Upset plot of DEOGs showing intersections

Figure S3. The conserved transcriptional network of the angiosperm endosperm.

Figure S4. The conserved transcriptional network of the angiosperm endosperm, the cell cycle and nucleosome architecture interacting network

Figure S5. Conserved transcriptional network of the angiosperm seed nourishing tissue, the Nourishing cluster protein interacting network

Figure S6. The *Pinus pinaster* megagametophyte private protein network.

Figure S7. The *Amborella* endosperm private protein network.

Figure S8. The *Nymphaea caerulea* endosperm private protein network.

Figure S9. The monocot endosperm private protein network.

Figure S10. The eudicot endosperm private protein network.