Imprinting and DNA methylation in water lily endosperm: implications for seed evolution

Rebecca A. Povilus[1], Caroline A. Martin[1,2], Lindsey L. Bechen[1], and Mary Gehring[1,2,3,4*]

[1] Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge MA 02142

[2] Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139

[3] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge MA 02139

[4] Howard Hughes Medical Institute

*Corresponding author: Mary Gehring, mgehring@wi.mit.edu

Author Contributions: R.A.P. and M.G. conceived of the original premise of the study. R.A.P. conducted experiments and wrote the paper with input from M.G. and C.A.M. C.A.M. assisted in processing genomic data. L.L.B performed methylation profiling on leaf tissue.

Competing Interest Statement: The authors have no competing interests to declare.

Keywords: Endosperm, Imprinting, Early Angiosperms, DNA methylation

This PDF file includes:

Main Text

Figures 1-4

Legends/Titles for Supplementary Figures 1-5 and Supplementary Tables 1-5

Abstract

Endosperm is a key evolutionary innovation associated with the origin of angiosperms (flowering plants). This altruistic seed tissue supports the growth and development of the embryo by mediating the relationship of the mother plant as a nutrient source to the compatriot embryo as a nutrient sink. The endosperm is the primary site of gene imprinting in plants (where expression of an allele in offspring depends on which parent it was inherited from) and of parent-specific epigenetic modifications like DNA methylation, which are differentially patterned during male and female gamete development. Experimental results from a phylogenetically-wide array of monocot and eudicot plants suggest these parent-of-origin effects are a common feature across angiosperms. However, information about genetic imprinting and epigenetic modifications in seeds of angiosperm lineages whose origins predate the monocot-eudicot divergence (such as Nymphaeales, water lilies) is extremely limited. Additionally, Nymphaeales are an intriguing lineage in which to investigate seed genetic and epigenetic phenomena, as they are characterized by diploid endosperm and a maternal storage tissue (perisperm), both of which are unusual across angiosperm diversity. Here, we examined DNA methylation and genetic imprinting using two reproductively compatible water lily sister-species, *Nymphaea thermarum* and *N. dimorpha*. Our results suggest that maternally-expressed imprinted genes and differential DNA methylation of maternally and paternally inherited endosperm genomes are an ancestral condition for endosperm, while other seed characters like seed provisioning strategies, endosperm ploidy, and paternally-expressed imprinted genes might have evolved as coinciding, opposing strategies in the evolutionary dialogue over parental control of offspring development.

Main Text

Introduction

The evolutionary origin of endosperm, a second fertilization product in the seeds of flowering plants, fundamentally altered the relationship between an embryo and its mother during seed development. In non-flowering seed plants, the embryo is directly connected to tissue that only contains genome complements from its mother. However, in angiosperm seeds, endosperm largely separates the embryo and its mother and is the product of a fertilization event and thus biparental, with both maternal and paternal genome contributions. Endosperm is widely recognized as a key mediator of developmental and nutritional relationships between an embryo and its mother (Povilus and Gehring 2022). The balance of maternal and paternal genomes is important for endosperm (and seed) viability, as evidenced by the phenomena of parental genome dosage sensitivity ((Haig and Westoby 1991) and references therein). When extra paternal genome complements are added to the endosperm, it over-proliferates – often leading to initially larger but ultimately collapsed, inviable seeds. Conversely, when extra maternal genome complements are added to the endosperm, reduced endosperm proliferation is observed, resulting in smaller seeds with fewer invested maternal resources (Haig and Westoby 1991; Birchler 1993; Scott et al. 1998). The endosperm is subject to other parent-of-origin effects such as imprinted gene expression (where expression of an allele depends on which parent it was inherited from) and parent-of-origin specific epigenetic modifications like DNA and histone methylation, which are differentially patterned during male and female gamete development (Gehring et al. 2006; Park et al. 106; Moreno-Romero et al. 2019; Borg et al. 2020). Our knowledge of endosperm gene imprinting and its underlying mechanisms is built from experiments performed in a phylogenetically wide array of monocot and eudicot plants (Povilus and Gehring 2022; Picard and Gehring 2020). However, information from lineages whose origins predate the monocot-eudicot divergence is extremely limited.

71 The order Nymphaeales (water lilies) is sister to all other angiosperms except for
72 *Amborella tricopoda*. Endosperm parental genome dosage sensitivity has been documented in
73 a species of water lily (Povilus et al. 2018), but nothing is known about patterns of genetic
74 imprinting or endosperm epigenetic patterning in this or any other ANA-grade lineage
75 (Amborella, Nymphaeales, Austrobaileyales), magnollids, or Chloranthales. In addition to their
76 relationship to other angiosperms, Nymphaeales are a particularly intriguing system in which to
77 investigate genetic imprinting and associated epigenetic patterning given the unique
78 combination of seed characters found in this lineage. First, endosperm of the Nymphaeales is
79 *ab initio*-cellular (the first nuclear division of the endosperm is accompanied by cellular division)
80 and diploid with a 1:1 maternal:paternal genome ratio (Orban and Bouharmont 1998; Williams
81 and Friedman 2002; Friedman 2006; Friedman 2008; Rudall et al. 2008; Povilus et al. 2015),
82 whereas triploid endosperm (2:1 ratio) characterizes the majority of angiosperms and all taxa in
83 which endosperm epigenetic patterning and genetic imprinting have been studied (Haig and
84 Westoby 1991). Diploidy has been suggested to represent the ancestral ploidy of endosperm
85 (Williams and Friedman 2002) and thus Nymphaeales are an opportunity to test how these
86 processes operate in the context of different base maternal-paternal genome/gene dosage
87 ratios. Second, in seeds of Nymphaeales nutrients are primarily stored in a perisperm (which is
88 derived from maternal tissue and contains no paternal genome contribution) instead of in
89 offspring tissues, in contrast to the vast majority of flowering plants (Lersten 2004; Patten et al.
90 2014). Therefore, Nymphaeales is an excellent clade in which to investigate the suggested
91 connection between genetic imprinting in endosperm and control of nutrient storage (Haig and
92 Westoby 1991; Patten 2014). Nutrient storage in perisperm is only initiated after fertilization
93 (Povilus et al. 2015), suggesting influence of offspring tissues on this process.

94 Characterizing genetic imprinting and epigenetic modifications in water lilies therefore
95 offers a unique perspective on the evolution of key endosperm traits and processes that are

96 associated with the origin of angiosperms. Here, we sought to determine whether gene
97 imprinting and parent-of-origin effects on DNA methylation, which has been mechanistically
98 linked to gene imprinting, preceded the origin of parental dosage-imbalanced (triploid)
99 endosperm.

100

101 **Results**

102 *Nymphaea thermarum* has been developed as an experimental system for the
103 Nymphaeales (Povilus et al. 2015; Povilus et al. 2018; Povilus et al. 2020). Assessing parent-
104 of-origin effects at the molecular level requires sequence polymorphisms, of which there are few
105 within the highly-inbred extant populations of *N. thermarum* in cultivation (Povilus et al. 2015).
106 We therefore assessed parent-of-origin effects in Nymphaeales by examining F1 tissue from
107 crosses between *N. thermarum* and *Nymphaea dimorpha* (which was formerly known as *N.*
108 *minuta*). These two species are estimated to have diverged roughly 20 million years ago
109 (Borsch et al. 2011). We confirmed the internal structure of young seeds of *N. thermarum*, *N.*
110 *dimorpha*, and of F1 reciprocal crosses and determined that we could ensure consistency in
111 developmental stage among crosses (Figure 1A). The hybrid F1 seeds are fully viable and
112 germinate to give rise to viable F1 plants (Supplementary Figure 1), suggesting no large-scale
113 divergence in endosperm developmental programs that would lead to failure in seed
114 development. We performed long-read based, *de novo* genome assembly and annotation for *N.*
115 *dimorpha* (248 contigs with an NG50 of 13,941,033 bp, representing 83% of the estimated
116 genome size, with the set of 40,850 annotated genes having a BUSCO score of 85% for the
117 Embryophyta gene set) and an improved genome assembly and annotation for *N. thermarum*
118 (1,553 contigs with and NG50 of 4,352,861, representing 86% of the estimated genome size,
119 with the set of 42,431 annotated genes having a BUSCO score of 83% for the Embryophyta
120 gene set) (Supplementary Table 1, Supplementary Figure 2). To allow direct comparisons of

121 genomic regions, the *N. thermarum* and *N. dimorpha* genomes were aligned and re-annotated
122 to create “reorganized” genomes for each species (each 358,929,111 bp in length and with a
123 resolved annotation having 39,608 genes and a BUSCO score of 72% for the Embryophyta
124 gene set).

125 To examine parent-of-origin specific gene expression, we made use of reciprocal
126 crosses between *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*) (2 samples each from *N. thermarum*
127 *x N. dimorpha* and *N. dimorpha x N. thermarum* crosses) as well as self-fertilized seeds (3
128 samples each of *N. thermarum* and *N. dimorpha*), and isolated RNA from young endosperm at
129 9-10 days after pollination/anthesis (Supplementary Table 1). By performing mRNA-seq, we
130 detected expression of a total of 22,984 genes with a TPM >= 1 (expression averaged across all
131 samples). A principal component analysis (PCA) of total gene expression revealed that
132 biological replicates clustered together according to cross type, with hybrid endosperm samples
133 midway along PC1 (54.85%) between endosperm from *N. thermarum* self-fertilized seeds and
134 *N. dimorpha* self-fertilized seeds (Figure 2A). Differential gene expression analysis between
135 sets of hybrid and non-hybrid seeds revealed that 879 genes were consistently significantly
136 differentially expressed in all comparisons; this set of genes was not significantly enriched for
137 any KEGG pathways, but was enriched for the GO biological process term “RNA-dependent
138 DNA biosynthetic processes” (FDR = 1.9e-2, n=8, fold enrichment = 7.1). Importantly, while
139 expression differences between hybrid and non-hybrid endosperm existed, endosperm of the
140 two hybrid cross directions were more similar to each other than they were to endosperm of the
141 parental lines (Figure 2A). We implemented a previously developed pipeline to evaluate
142 imprinted gene expression (Gehring et al, 2011) (see Methods). In each possible comparison of
143 a *Nt* x *Nd* and *Nd* x *Nt* cross, we identified transcripts that showed a significant bias in the
144 number of reads mapping uniquely to either the maternally- or paternally-inherited alleles, in
145 both cross directions (Figure 2B, Supplementary Table 2). For these imprinting tests, 26,465

146 genes had at least one read that could be assigned to a parent-of-origin and 16,647 genes
147 passed our minimum allele-specific read count cut-off of 50 reads and were assessed for
148 imprinting. Our analysis revealed the presence of imprinted genes in *Nymphaea*. We identified
149 small numbers of paternally-biased genes in individual cross comparisons, but only 1 paternally
150 expressed imprinted gene (PEG) was consistent in at least 75% of comparisons (3 of 4 total
151 possible cross comparisons) (Figure 2C); this PEG is a homolog of *CELLULOSE SYNTHASE*
152 *LIKE G2* (*ATCLSG2*) in *Arabidopsis thaliana*. A handful of PEGs have previously been identified
153 as conserved between monocots and dicots (Pignatta and Gehring 2012). We examined the
154 expression of homologs of these specific genes in *Nymphaea* endosperm. Although there was
155 some evidence for paternally-biased expression, there were also large *cis* or species effects on
156 transcription, and these genes did not meet all of our criteria for imprinting (Supplementary
157 Table 2). In contrast to PEGs, 157 MEGs were consistently identified in at least 75% of
158 comparisons, with 147 being identified as MEGs in all comparisons (Figure 2C) (Supplementary
159 Table 2). Previous studies have shown that imprinting can be altered or obscured in
160 interspecies hybrids (Josefsson et al. 2006; Burkart-Waco et al. 2015; Florez-Rueda et al.
161 2016), and we evaluated whether the identified MEGs were differentially expressed in the F1
162 hybrid endosperm. When endosperm gene expression profiles were compared between each
163 hybrid type and each parental species type (Figure 2D, Supplementary Figure 3, Supplementary
164 Table 3), only 10 MEGs were consistently (in 75% or more of comparisons) significantly
165 differentially expressed, including only 1 MEG that was significantly differentially expressed in all
166 comparisons. This indicates low overlap between identified MEGs and genes that are mis-
167 regulated in hybrid crosses. We conclude that imprinted expression exists in *Nymphaea*
168 endosperm, but is largely restricted to MEGs and mostly does not include genes whose
169 expression is altered in hybrid endosperm.

170 The identification of MEGs can be influenced by contamination with maternal tissue or
171 transcripts in endosperm samples and thus we evaluated the extent to which this might be
172 affecting our results. Notably, PCA results indicate that hybrid endosperm from both cross
173 directions was more like each other transcriptionally than endosperm of their mother species,
174 suggesting no significant, wide-spread maternal contamination (Figure 2A). We additionally
175 performed RNA *in situ* hybridizations on seeds of self-fertilized *N. thermarum* for a set of
176 identified MEGs to test whether they showed substantial expression in maternal tissue or lacked
177 expression in the endosperm (both of which would indicate potential for maternal tissue
178 contamination). Target genes were selected for high maternal expression bias and high
179 expression (Figure 2E), including a gibberellic acid oxidase homolog (*GA2OX2*) (Figure 2F) and
180 a sucrose synthase 3 homolog (*SUS3*) (Figure 2G). We furthermore confirmed that the high
181 starch and carbohydrate content of the perisperm was not interfering with the *in situ*
182 hybridization experimental protocol, as we were able to detect expression of a subfamily of
183 terpene synthases in the perisperm, as well as in the endosperm (Supplementary Figure 4). For
184 both *GA2OX2* and *SUS3*, we detected expression in the endosperm and not in the perisperm,
185 indicating that for these and likely other MEGs, the identified maternal expression bias is
186 unlikely to be caused by contamination with maternal tissue. For *SUS3*, we also performed *in*
187 *situ* hybridizations on seeds from reciprocal crosses and from *N. dimorpha* self-fertilizations, and
188 found similar endosperm expression patterns (Supplementary Figure 4).

189 Additionally, we took advantage of whole-seed expression datasets from *N. thermarum*
190 (Povilus and Friedman 2022) to explore the potential impact of maternal tissue transcript
191 contamination on our imprinting results. For each transcript expressed in the endosperm, we
192 calculated corrected endosperm maternal read counts based on assuming that 50% or 25% of
193 the isolated endosperm transcript pool was comprised of transcripts from a whole-seed pool
194 (Tonosaki et al. 2024). We found that correcting for an assumed 50% sample contamination

195 resulted in identification of 112 MEGs and correcting for an assumed 25% contamination
196 resulted in identification of 134 MEGs, compared to the originally identified 157 MEGs
197 (Supplementary Table 4). Both sets of 112 MEGs and 134 MEGs were subsets of the originally
198 identified 157 MEGs. The “corrected” MEGs were enriched for KEGG or GO biological process
199 terms related to response and development, carbohydrate processing, and secondary
200 metabolism (Supplementary Table 4), similar to the enrichments for the original, uncorrected set
201 of MEGs (Figure 2G). Furthermore, in the corrected datasets *SUS3* was not identified as a
202 MEG, although our *in situ* hybridizations demonstrated that *SUS3* was not highly expressed in
203 maternal seed tissues but was expressed in endosperm (Figure 2F, Supplementary Figure 4).
204 We concluded that the correction for extensive, assumed whole-seed transcript contamination
205 was likely inappropriately removing true MEGs, and therefore proceeded to use the uncorrected
206 set of MEGs for further analysis.

207 Overall, *Nymphaea* MEGs were significantly enriched for GO annotations associated
208 with response and development, carbohydrate processing, and secondary metabolism (Figure
209 2H, Supplementary Table 4). The enrichment for processes integral to development and nutrient
210 dynamics in seeds is similar to what has been found in other species (Xin et a. 2013; Picard et
211 al. 2021). Notably, *SUS3* is a maternally expressed imprinted gene in *Nymphaea* and is a key
212 part of nutrient processing in the endosperm in other species (Angeles-Núñez and Tiessen
213 2010). These results are congruent with genetic imprinting being associated with nutrient
214 dynamics during seed development.

215 Having found evidence for parent-of-origin effects on gene expression, we next
216 investigated parent-of-origin effects on DNA methylation by performing endosperm whole-
217 genome enzymatic methyl-sequencing (EM-seq). We again made use of reciprocal crosses
218 between *N. thermarum* and *N. dimorpha* to permit allele-specific characterization of DNA
219 methylation patterning in young F1 hybrid endosperm; biological replicates from two *N.*

220 *thermarum x N. dimorpha* crosses and from three *N. dimorpha x N. thermarum* crosses were
221 analyzed (Supplementary Table 1). We also obtained methylation profiles from single samples
222 of *N. thermarum* and *N. dimorpha* leaves and the leaves of an F1 *N. dimorpha x N. thermarum*
223 hybrid. In both endosperm and leaves, the average methylation profiles of *N. thermarum* and *N.*
224 *dimorpha* alleles of genes and repeats (transposable elements) (Figure 3A) was similar to other
225 angiosperms (Niederhuth et al. 2016), with CG methylation occurring in gene bodies and CG
226 and non-CG methylation in repeats. We identified differentially methylated regions (DMRs)
227 between *N. thermarum* and *N. dimporpha* genomes in F1 hybrid leaves and between *N.*
228 *thermarum* and *N. dimorpha* leaves. In hybrid leaves, the majority of DMRs occurred in the CG
229 context, and similar numbers of regions were more methylated in one species or one genome
230 versus the other (Figure 3B, Supplementary Figure 5). We then identified DMRs between
231 maternal and paternal alleles in F1 endosperm. The majority of CG and CHH DMRs were less
232 methylated on maternal alleles than on paternal alleles, regardless of which species was the
233 maternal parent in the F1 hybrid endosperm (Figure 3B). Additionally, maternal alleles of both
234 species were consistently hypermethylated in the CHG context and were generally more
235 methylated in endosperm than in hybrid or parental species' leaf tissue, for both genic flanking
236 regions and within repeats (Figure 3A). The higher methylation of maternal alleles in the CHG
237 context was also borne out by analyzing the number of DMRs between maternal and paternal
238 alleles (Figure 3B, Supplementary Table 5): the CHG context had the highest difference
239 between the portion of maternally- or paternally-hypermethylated DMR windows, with both
240 species having more maternally hypermethylated windows. These methylation patterns were
241 not found in leaf tissue (Figure 3B; Supplementary Figure 5) and may therefore be unique to
242 endosperm. Together, these findings parallel observations of methylation patterning in
243 endosperm of monocots and dicots, in which endosperm is CG hypomethylated on maternal
244 alleles due to active DNA demethylation that occurs in the central cell (the female gamete that is
245 the progenitor of the endosperm) before fertilization. Maternal allele CG hypomethylation has

246 been noted in rice, Arabidopsis, and other species (Gehring et al. 2009; Hsieh et al. 2009;
247 Waters et al. 2011; Rodrigues et al. 2013; Park et al. 2016; Zhang et al. 2016; Zhang et al.
248 2021). Endosperm maternal allele CHG hypermethylation has been observed in *Arabidopsis*
249 *lyrata* (Klosinska et al. 2016) and to a lesser extent in *A. thaliana* (Moreno-Romero et al. 2019),
250 although this occurs primarily in gene bodies, unlike the observations here where it occurs in
251 gene flanking regions and in repeats.

252 We further examined the set of genes that overlapped a maternal- or- paternal
253 hypermethylated DMR consistently across both species (had a DMR that overlapped the gene
254 region +/- 2kb in at least 75% of pairwise comparisons, while not overlapping a DMR of the
255 opposite type in more than 25% of comparisons) (Figure 3C, Supplementary Table 5). Similar to
256 looking at numbers of DMRs associated with genes separately in each species, more genes
257 were associated with paternally hypermethylated DMRs in the CG and CHH contexts, while in
258 the CHG context more genes overlapped with maternally hypermethylated DMRs. Few
259 imprinted loci were consistently associated with allele-specific DMRs (Supplementary Table 5).
260 Four MEGs were consistently associated with maternal allele hypomethylated DMRs in the CG
261 context, including homologs of *TARGET OF MONOPTEROS 6* (*TMO6*) and *GA2OX2* (Figure
262 3D), and 2 MEGs were associated with maternally hypomethylated DMRs in the CHH context.
263 These regions were not differentially methylated in F1 hybrid leaves, indicating a parent-of-
264 origin effect on methylation that is specific to the endosperm (Figure 3D). The one identified
265 PEG was not associated with any significant methylation differences between parental alleles.
266 Thus, we conclude that there are parent-of-origin specific differences in DNA methylation in
267 *Nymphaea* endosperm, of a similar nature (maternal CG hypomethylation) to those observed in
268 monocots and eudicots. The majority of *Nymphaea* MEGs are not associated with differential
269 DNA methylation. For comparison, in maize and Arabidopsis approximately 50% and 40% of

270 MEGs, respectively, are associated with differential DNA methylation (Pignatta et al. 2014; Gent
271 et al. 2022).

272

273 **Discussion**

274 Our results illuminate the evolution of imprinting and potential mechanisms facilitating the
275 emergence of gene imprinting. In summary, we found that genetic imprinting and parent-of-
276 origin effects on DNA methylation patterning occur in endosperm of Nymphaea seeds. Both
277 DNA methylation and genetic imprinting have been suggested to be strategies that can alter the
278 effective maternal or paternal genome/gene dosage in endosperm. Changes to absolute
279 parental genome dosage (and ploidy) of endosperm have also occurred repeatedly during
280 angiosperm evolution. Our discovery of endosperm genetic imprinting and parent-of-origin
281 effects on DNA methylation in *Nymphaea* suggests that these characters/processes predate the
282 evolution of triploid endosperm and are likely to be have been either co-opted from preexisting
283 ancestral molecular programs or are novelties associated with the origin of endosperm itself
284 (Figure 4). In either case, these findings demonstrate that a 2:1 maternal:paternal genome
285 dosage ratio is not a requirement for either endosperm maternal allele CG hypomethylation and
286 CHG hypermethylation, or for genetic imprinting. Furthermore, our results suggest that parent-
287 of-origin effects on endosperm development in *Nymphaea* (Povilus et al. 2018) could be linked
288 to parent-of-origin-specific DNA methylation patterning or maternally-expressed imprinted
289 genes, but not extensively to paternally-expressed imprinted genes. This is perhaps surprising
290 given that Povilus et al. (2018) observed paternal effects when diploids were pollinated by
291 tetraploids: in mature seeds, endosperm of both 4n x 2n (maternal excess) and 2n x 4n
292 (paternal excess) is larger than endosperm of 2n x 2n crosses. However, the developmental
293 timing by which larger endosperm is achieved differs between maternal and paternal excess
294 crosses. During later development (7-32 DAA), the endosperm of 2n x 4n crosses grows

295 significantly faster than that of 2n x 2n crosses. By contrast, in 4n x 2n crosses, the endosperm
296 grows faster earlier (1-7 DAA), and then decelerates at later stages. The observed maternal and
297 paternal effects in *N. thermarum* endosperm are therefore distinct from those typically observed
298 in maize or Arabidopsis, where maternal excess seeds undergo early endosperm cellularization
299 and are typically smaller than 2n x 2n seed at maturity and paternal excess seeds undergo
300 extended endosperm proliferation and are larger and dead (Arabidopsis) or smaller and dead
301 (maize) at maturity (Scott et al. 1998; Pennington et al. 2008). Although it has been proposed
302 that an increased dosage of PEGs is the cause of interploidy paternal excess phenotypes, direct
303 evidence is limited. Indeed, in Arabidopsis it has been shown that PEG expression is increased
304 in both viable and non-viable seeds from Arabidopsis paternal excess crosses (Satyaki and
305 Gehring 2019), suggesting that PEG expression is not the determining factor, or sole
306 determining factor, of paternal excess interploidy phenotypes. Finally, a single asymmetry
307 between parental genomes, such as a MEG, has the potential to cause both maternal and
308 paternal parental effects as the endosperm dosage of the gene would differ between interploidy
309 crosses.

310 It remains possible that we failed to detect PEGs because of our use of inter-species
311 comparisons or because of the developmental stage at which we analyzed imprinting. Inter-
312 specific and inter-ploidy crosses can result in misregulation of imprinted gene expression. In
313 hybrid crosses between *A. thaliana* and *A. arenosa*, several PEGs gain expression from the
314 maternally-inherited allele in whole seed data, rendering them biallelic or even maternally
315 biased (Josefsson et al. 2006; Burkart-Waco et al. 2015). Other genes become newly paternally
316 biased. Similarly, endosperm of wild tomato hybrids exhibit a genome-wide increase in maternal
317 transcriptome contribution and reduced paternal bias of PEGs (Florez-Rueda et al. 2016).
318 However, in both of these examples, endosperm development is defective and the hybrid seeds
319 are inviable. This is not the case in crosses between *N. thermarum* and *N. dimorpha* – the

320 seeds and F1 plants are fully viable (Figure 1, Supplementary Figure 1), suggesting no defects
321 in endosperm development of the kind that is often correlated with misregulated genomic
322 dosage or misregulated imprinting. Thus, although we cannot exclude it, we think it unlikely that
323 PEGs were not detected in *Nymphaea* endosperm because of the genetic material used in our
324 crosses. We also cannot exclude the possibility that PEGs might be active at earlier stages of
325 Nymphaea seed development than we sampled, before significant development/functioning of
326 the perisperm, as has been suggested by a study on endosperm of a different Nymphaea
327 species (Florez-Rueda et al, 2024). Yet, in other species PEGs have been detected across all
328 assayed stages of endosperm development, including in seeds with similar stages of embryo
329 development as sampled in our study.

330 The observed differences in CG DNA methylation between endosperm maternal and
331 paternal alleles (Figure 3B,D) are consistent with the activity of a 5-methylcytosine DNA
332 glycosylase in the central cell before fertilization, which would be predicted to cause maternal
333 allele CG hypomethylation in endosperm after fertilization. Homologues of DME are present in
334 all angiosperms (Pei et al. 2019), including *Nymphaea* (Povilus and Friedman 2022). The lack of
335 high congruence between CG maternal allele hypomethylated DMRs and *Nymphaea* MEGs is
336 not inconsistent with data from other species. In Arabidopsis and maize, there are many more
337 endosperm CG hypomethylated regions than there are imprinted genes (Gehring et al. 2009;
338 Pignatta et al. 2014; Gent et al. 2022). Although the imprinting of individual Arabidopsis MEGs
339 like *FWA* and *SDC* is correlated with differential methylation between maternal and paternal
340 alleles (Kinoshita et al. 2004), as a group MEGs are not enriched for differential methylation
341 compared to non-imprinted genes (Pignatta et al. 2014). This is in contrast to PEGs, where
342 differential CG methylation of upstream or downstream regions is enriched compared to non-
343 imprinted genes (Pignatta et al. 2014; Moreno-Romero et al. 2019). We also observed
344 endosperm maternal allele CHG hypermethylation in *Nymphaea* endosperm in repeats and

345 gene-flanking regions (Figure 3A-C). In other angiosperm species, CHG methylation is
346 frequently associated with H3K9me2 and transcriptional silencing. The maternal alleles of PEGs
347 in Arabidopsis species are significantly associated with CG hypomethylation and CHG
348 hypermethylation in endosperm (Klosinska et al. 2016; Moreno-Romero et al. 2019). Although
349 we observed maternal allele CG hypomethylation and CHG hypermethylation in *Nymphaea*
350 endosperm, they were not associated with PEGs, suggesting that paternal expression bias is
351 not an inevitable outcome of these types of epigenetic patterns. Other epigenetic mechanisms
352 could regulate gene imprinting in *Nymphaea*. In mammals, imprinting of a small number of
353 genes is regulated by H3K27me3, without an apparent role for DNA methylation (Lewis et al.
354 2004; Inoue et al. 2017; Santini et al. 2021; Raas et al. 2022), and has been proposed as an
355 ancestral mechanism of imprinting in the placenta. H3K27me3 also plays an important role in
356 plant gene imprinting and is often coincident with differential DNA methylation (Gehring et al.
357 2006; Zhang et al. 2014; Moreno-Romero et al. 2019). Histone modification patterns in
358 *Nymphaea* endosperm is a potential area of future investigation.

359 In the context of parental conflict over the investment of maternal resources in offspring
360 and maternal-offspring coadaptation, the evolution of maternal storage tissues and the notable
361 paucity of PEGs in Nymphaea endosperm give rise to two distinct hypotheses about the early
362 evolution of angiosperm seed development and endosperm molecular/genetic processes
363 (Figure 4). For the first hypothesis, if MEGs and PEGs were both present in the ancestrally
364 diploid endosperm of the earliest flowering plants, then PEGs were largely lost in association
365 with the evolution of perisperm in the Nymphaeales (Figure 4, left). If genetic imprinting is a tool
366 to establish maternal or paternal control over resource investment in offspring (Haig and
367 Westoby 1991), perhaps PEGs are no longer an effective paternal strategy when the mother
368 establishes primary control by storing invested resources in a maternally-derived tissue that
369 does not have a paternal genome contribution. A second hypothesis is that genetic imprinting

370 has evolved in stepwise fashion along with endosperm ploidy changes (Figure 4, right). In this
371 case, MEGs may have evolved as a maternal strategy to balance the addition of a paternal
372 genome – and potential for paternal influence on seed development – that resulted in the origin
373 of endosperm. Subsequent addition of a second maternal genome complement with the
374 evolution of triploid endosperm may then be similarly associated with the evolution of PEGs.
375 Studying other members of ANA-grade lineages with diploid endosperm that lack perisperm
376 (such as in Austrobaileyales (Losada et al. 2017)) or taxa with triploid endosperm and perisperm
377 (as can be found in Amaranthaceae (Coimbra and Salema 1999; López-Fernández and
378 Maldonado 2013), within the eudicots) would help distinguish between these hypotheses by
379 specifically testing the relationship between maternally-derived storage tissues and the absence
380 of PEGs. Thus, while characterizing parent-of-origin effects on gene expression and epigenetic
381 modifications in Nymphaea endosperm is an important step in understanding the evolution of
382 molecular processes in endosperm, the genetic and developmental diversity across angiosperm
383 seeds deserves further attention.

384

385

386 **Materials and Methods**

387 **Data availability**

388 Data generated as part of this study are available as part of NCBI BioProjects PRJNA1085993,
389 PRJNA1086866, PRJNA1086863, PRJNA1085992, and PRJNA1087317, including raw
390 sequence data deposited in SRA, and genome assemblies deposited in NCBI WGS.

391

392 **Plant growth and sample collection**

393 Seeds of *N. thermarum* and *N. dimorpha* were sourced from the Arnold Arboretum of Harvard
394 University (Boston, MA, USA) and grown at the Whitehead Institute for Biomedical Research
395 (Cambridge, MA, USA) (Supplementary Materials and Methods).

396 Controlled pollinations and self-fertilizations were performed as previously described
397 (Povilus et al. 2015; Povilus et al. 2018). For collection of seeds from crosses and self-
398 fertilizations, first day of anthesis (time of female receptivity and fertilization (Povilus et al.
399 2015)) was defined by presence of stigmatic fluid. Fruits were collected at 10 days after
400 anthesis (DAA) and seeds were immediately removed and dissected with fine forceps in
401 dissection buffer. Endosperm tissue was washed with dissection buffer multiple times and
402 frozen in liquid nitrogen (Supplementary Materials and Methods).

403

404 **Whole genome sequencing, assembly, and annotation**

405 For long-read and short-read DNA sequencing, high molecular weight DNA was
406 extracted from > 1 g young leaf samples from a single individual plant using a modified CTAB-
407 based protocol (Supplementary Materials and Methods). Samples were prepared for PacBio
408 sequencing (PacBio Sequel SMRTcell, 20h, v3 chemistry) and were sequenced at the MIT
409 BioMicroCenter. *Nymphaea thermarum* was sequenced in both LR (long-read) and HiFi (high-
410 fidelity) modes; *N. dimorpha* was sequenced in LR mode. *N. dimorpha* short-read data was
411 obtained from the same sample, using 1 lane of an Illumina HiSeq2000 flow-cell (40 bp, paired-
412 end reads) at the MIT BioMicroCenter. Short-read genomic data for *N. thermarum* was
413 downloaded from BioProject PRJNA508901.

414 Genome assembly for *N. thermarum* and *N. dimorpha* was performed separately using
415 long-reads as input for Canu (version 2.1.1) (Koren et al. 2017); short-reads were used to polish
416 the resulting assemblies using POLCA (from MaSuRCA version 3.4.2) (Zimin et al. 2013).

417 Genome assemblies were visualized with Bandage (Wick et al. 2015). The resulting original
418 genome assemblies were separately annotated with MAKER (version 2.31.10)(Campbell 2014)
419 for both species, using an iterative approach to train AUGUSTUS (version 3.3.3)(Stanke 2006)
420 and SNAP (version 2006.07.28-1)(Korf 2004) gene-model predictors; initial input for all
421 annotation pipelines included the set of transcript and protein sequences from the published *N.*
422 *thermarum* genome assembly/annotation(Povilus 2020), the set of all protein sequences from
423 Nymphaeaceae available on NCBI, protein sequences from the *N. colorata* genome assembly
424 and annotation (Zhang 2020), all basal Magnoliophyta protein sequences on Uniprot, Amborella
425 protein sequences, and TAIR10 protein sequences. Three rounds of annotation and gene model
426 predictor training were performed for annotation of both species. Repeat identification and
427 masking was performed with RepeatMasker (version 4.0.5)(Chen 2004) using Spermatophyta
428 as the specified query clade and the Embryophyta repeat database.

429 To create genome assemblies of *N. thermarum* and *N. dimorpha* that shared positional
430 homology, the *N. thermarum* contigs were mapped to *N. dimorpha* contigs using minimap2 (Li
431 2018) (the *N. dimorpha* assembly was used as the reference as it had the fewest contigs;
432 genome alignment visualized using D-Genies (Cabanettes and Klopp 2018). The resulting
433 reorganized genomes for both species were separately re-annotated as described above, and
434 the final annotations were resolved using MAKER to the reorganized *N. thermarum* genome,
435 with positional homology used to apply the annotation to the *N. dimorpha* genome assembly.
436 For each species, assembled transcripts were then generated using the resolved annotation
437 and genome assembly of each species, resulting in a set of homologous *N. thermarum*
438 transcripts and a set of *N. dimorpha* transcripts.

439 For *N. thermarum* and *N. dimorpha* transcripts, homology to *Arabidopsis thaliana* was
440 determined by blastx searching the *N. thermarum* transcripts against the TAIR11 protein set (e-
441 value cut-off set at 1e-4). The top *A. thaliana* blastx hit for each *N. thermarum* transcript was

442 selected as the putative homolog. For each *N. dimorpha* transcript, the putative *N. thermarum*
443 homolog was similarly identified with a blastx search (e-value cut-off set at 1e-4), and the
444 corresponding *A. thaliana* homologs was assigned to the *N. dimorpha* transcript.

445

446 **DNA methylation-sensitive sequencing and analysis**

447 DNA was extracted from endosperm using the QIAamp micro kit (Qiagen, cat# 56304).
448 For enzymatic methyl conversion sequencing and library preparation, an NEBNext Enzymatic
449 Methyl-seq kit was used; one additional AMPure bead clean-up was performed on libraries to
450 remove primer dimer. Sequencing was performed at the Whitehead Institute Genome
451 Technology Core. Libraries were pooled and sequenced across 2 lanes of a NovaSeq SP
452 flowcell (50 bp, paired-end reads) (endosperm samples) or 2 lanes of a NovaSeq S4 flowcell
453 (150 bp, paired-end) (leaf samples) to give ~14x genome coverage. Enzymatic-methyl
454 sequencing conversion rate was assessed prior to sequencing (Supplementary Materials and
455 Methods). Conversion rates were calculated using CyMATE (Hetzl et al. 2007). Sample
456 conversion rate averaged 99.85%.

457 Reads from enzymatic-converted samples were first mapped to a concatenation of the
458 originally produced *N. thermarum* and *N. dimorpha* assemblies and annotations, using Bismark
459 (version 0.22.3) (Krueger and Andrews 2011). 150 bp reads of leaf samples were broken into 40
460 bp segments and all reads were treated as single-end during mapping to ensure consistency in
461 data processing. The reads that uniquely mapped to either species' genome were sorted into
462 separate sets of *N. thermarum* or *N. dimorpha* reads and mapped to their respective species'
463 reorganized genome annotation with Bismark and methylation data was extracted. Analysis of
464 average DNA methylation 5', 3' and interior of features was performed using previously
465 developed custom pipelines (Pignatta et al. 2014). Differentially methylated regions (DMRs)

466 between samples were identified in the CG, CHG, and CHH contexts using a previously
467 developed pipeline (Pignatta et al. 2014). DMRs were defined as 300-bp windows for which 3 or
468 more cytosines with a coverage of 5 or more reads had a methylation difference of 35% or
469 greater between samples for CG and CHG contexts and 10% or greater for the CHH context,
470 with a Fisher's exact test with Benjamini-Hochberg correction p-value cutoff of 0.01 to determine
471 significance. DMRs were called between all combinations of biological replicates. For total
472 number of DMRs between endosperm maternal and paternal alleles, the number of DMRs was
473 averaged across all replicate comparisons. Genes and repeat regions were identified as
474 associated with a DMR if the gene or repeat region had a DMR within the annotated region or
475 +/-2 kb.

476 GO enrichment analysis was performed using ShinyGO 0.77 (Ge et al. 202). Putative
477 Arabidopsis homologs of all transcripts were used, and the set of putative Arabidopsis homologs
478 of all transcripts expressed during seed development (TPM > 1) (Povilus and Friedman 2022)
479 was used as the background set.

480

481 **RNA sequencing and data analysis**

482 For mRNA sequencing, RNA was extracted from frozen endosperm samples via
483 RNAqueous Total RNA Isolation Kit (Invitrogen) according to the kit protocol. Libraries were
484 prepared and sequenced at the MIT BioMicroCenter via NEBNext Ultra II Directional RNA
485 Library Prep Kit for Illumina (polyA-based isolation). Samples were pooled and sequenced on
486 one NovaSeq S4 flowcell (50bp, single end reads).

487 For full analysis methods, see Supplementary Materials and Methods. Briefly, for initial
488 analysis of gene expression, reads from all hybrid samples were mapped to the concatenated
489 genomes of the originally produced *N. thermarum* and *N. dimorpha* assemblies and

490 annotations; reads from non-hybrid samples were mapped to the reorganized genome of their
491 respective species. For identification of imprinted genes in hybrid samples, the reads that
492 uniquely mapped to either species' genome were sorted into separate sets of *N. thermarum* or
493 *N. dimorpha* reads and were used for subsequent analysis. *N. thermarum* reads were mapped
494 to the reorganized *N. thermarum* genome annotation, *N. dimorpha* reads were mapped to the
495 reorganized *N. dimorpha* genome annotation. Resulting allele-specific count tables for each
496 transcript were used for calling genetic imprinting. Genetically imprinted genes were called as
497 previously described (Gehring et al. 2011; Pignatta et al. 2014), using a pairwise comparison of
498 all possible combinations of each hybrid cross sample. For each gene we tested whether there
499 was a significant difference (Benjamini-Hochberg adjusted p-value < 0.01) between p_1 and p_2,
500 where p_1 is the portion of Nt reads for a gene in a Nt x Nd cross and p_2 is the portion of Nt reads
501 for the same gene in a Nd x Nt cross. While mapping reads, a slight maternal expression bias
502 was noted for both cross directions (Supplemental Table 1). Therefore, when calling imprinted
503 genes, the expected maternal : paternal expression ratio was adjusted from 1 (the anticipated
504 null ratio for diploid endosperm) to the average maximum observed maternal expression bias of
505 1.32 (null hypothesis: p_1=1.32, p_2=0.57). To increase stringency, minimum allelic-specific read
506 count was set to 50, a minimum imprinting factor was set to 2, and a maximum cis-effect factor
507 was set to 15. The imprinting factor is a measure of the magnitude of imprinting. For each gene
508 in a sample, a 95% confidence interval was determined around the Nt/Nd read ratio; the
509 imprinting factor is the low value of the high confidence interval divided by the high value of the
510 low confidence interval for the reciprocal cross (Gehring et al. 2011). The *cis*-effect factor is
511 calculated in a similar manner. In addition to these specifications, MEGs were required to have
512 a minimum of 70% maternal allele reads and PEGs were required to a maximum of 30%
513 maternal allele reads in both cross directions. In order for a gene to be considered as
514 consistently imprinted, it had to be called as imprinted in at least 75% (3 of 4) of pairwise
515 comparisons.

516 For correction of endosperm reads to account for potential maternal tissue
517 contamination, we mapped reads from endosperm samples (this study) and whole-seed
518 samples (Povilus and Friedman 2022) to the reorganized *N. thermarum* genome and proceeded
519 as described in (Tonosaki et al. 2024) (Supplementary Materials and Methods).

520 Differential gene expression analysis between endosperm of hybrid crosses and non-
521 hybrid endosperm was performed by mapping reads to the concatenated genomes as described
522 above, and then mapping uniquely mapping reads to their respective reorganized genome using
523 kallisto (v 0.46.1) (Bray et al. 2016). Differential gene expression analysis was performed using
524 DEseq2 (Love et al. 2014) using mostly default settings and filtering for loci with adjusted p
525 value less than or equal to 0.01 and mean TPM (of all samples) greater than or equal to 10.

526 GO enrichment analysis was performed using ShinyGO 0.77 (Ge et al. 2020), using
527 default settings. Putative Arabidopsis homologs of all transcripts in the test set were used, and
528 the set of putative Arabidopsis homologs of all transcripts expressed during seed development
529 (TPM > 1) (Povilus and Friedman 2022) was used as the background set.

530

531 **In situ hybridizations and histology**

532 *In situ* hybridizations were performed as previously described (Pignatta et al. 2018) (see
533 Supplementary Materials and Methods for probe information). Preparation of seed samples for
534 histological analysis was performed as previously described for seeds of *A. thaliana* (Pignatta et
535 al. 2018) and stained with toluidine blue, with adaptations of incubation times as necessary. All
536 samples were sectioned on a Leica RM 2065 rotary microtome at a thickness of 8 µm and
537 imaged using a Zeiss Axio Imager M2. Image tiling, color and brightness/contrast adjustments
538 and Smart Sharpen were applied to whole images, with particular attention to having even
539 contrast and white-balance across different images (Adobe Photoshop).

540

541 **Acknowledgments**

542 Funding was provided by the National Science Foundation grants MCB 2101337 to M. G., MCB
543 1453459 to M. G., IOS 1812116 to R.A.P, and a NSF Graduate Research Fellowship to C.A.M.
544 Sequencing was performed at the MIT BioMicro Center and the Whitehead Institute Genome
545 Technology Core.

546

547 **References**

548 Angeles-Núñez JG, Tiessen A. 2010. Arabidopsis sucrose synthase 2 and 3 modulate
549 metabolic homeostasis and direct carbon towards starch synthesis in developing seeds. *Planta*
550 232(3): 701-718. 10.1007/s00425-010-1207-9.

551 Birchler JA. 1993. Dosage analysis of maize endosperm development. *Annu. Rev. Genet.* 27:
552 181-204. 10.1146/annurev.ge.27.120193.001145.

553 Borg M, Jacob Y, Susaki D, LeBlanc C, Buendía D, Axelsson E, Kawashima T, Voigt P, Boavida
554 L, Becker J, et al. 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in
555 plant paternal chromatin. *Nat. Cell Biol.* 22(6): 621-629. 10.1038/s41556-020-0515-y.

556 Borsch T, Löhne C, Mbaye M, Wiersema L. 2011. Towards a complete species tree of
557 *Nymphaea*: shedding further light on subg. Brachyceras and its relationship to the Australian
558 water-lilies. *Telopea* 13: 193-217. ISSN (Electronic): 2200-4025.

559 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-sq
560 quantification. *Nature Biotech.* 34: 525-547. 10.1038/nbt.3519.

561 Burkart-Waco D, Ngo K, Liberman M, Comai L. 2015. Perturbation of parentally biased gene
562 expression during interspecific hybridization. *PLoS One* 10(2): e0117293.
563 10.1371/journal.pone.0117293.

564 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and
565 simple way. *PeerJ* 6 :e4958. 10.7717/peerj.4958.

566 Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome Annotation and Curation Using
567 MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* 48: 4.11.1–4.11.39.
568 10.1002/0471250953.bi0411s48.

569 Coimbra S, Salema R. 1999. Ultrastructure of the developing and fertilized embryo sac of
570 *Amaranthus hypochandriacus L*. *Ann. Bot.* 84: 781-789. 10.1006/anbo.1999.0978.

571 Florez-Rueda AM, Paris M, Schmidt A, Widmer A, Grossniklaus U, Städler T. 2016. Genomic
572 imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol.*
573 *Biol. Evo.* 33(11): 2935-2946. 10.1093/molbev/msw175.

574 Florez-Rueda AM, Sharmann M, de Souza LP, Fernie AR, Bachelier JB, Figueiredo DD. 2024.
575 Genomic imprinting in an early-diverging angiosperm reveals ancient mechanisms for seed
576 initiation in flowering plants. *bioRxiv*. 10.1101/2024.09.18.613646.

577 Friedman WE. 2006. Embryological evidence for developmental lability during early angiosperm
578 evolution. *Nature* 441: 337-340. 10.1038/nature04690.

579 Friedman WE. 2008. Hydatellaceae are water lilies with gymnospermous tendencies. *Nature*
580 453: 94-97. 10.1038/nature06733.

581 Ge SX, Jung D, Yao R. 2020. ShinyGo: a graphical gene-set enrichment tool for animals and
582 plants. *Bioinformatics* 36(8): 2628-2629. 10.1093/bioinformatics/btz931.

583 Gehring M, Huh JH, Hsieh T, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. 2006.
584 DEMETER RNA glycosylase established MEDEA polycomb gene self-imprinting by allele-
585 specific demethylation. *Cell* 124(3):495-506. 10.1016/j.cell.2005.12.034.

586 Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during
587 seed development underlies gene imprinting. *Science* 324(5933): 1447-1451.
588 10.1126/science.1171609.

589 Gehring M, Missirian V, Henikoff S. 2011. Genomic analysis of parent-of-origin allelic
590 expression in Arabidopsis thaliana seeds. *PLoS ONE* 6(8): e23687l.
591 10.1371/journal.pone.0023687.

592 Gent JI, HIggins KM, Swentowsky KW, Fu F, Zeng Y, Kim DW, Dawe RK, Springer NM,
593 Anderson SN. 2022. The maize gene maternal derepression of r1 encodes a DNA glycosylase
594 that demethylates DNA and reduces siRNA expression in the endosperm. Plant Cell 34(10):
595 3685-3701. 10.1093/plcell/koac199.

596 Haig D, Westoby M. 1991. Genomic imprinting in endosperm: its effect on seed development in
597 crosses between species, and its implications for the evolution of apomixis. *Philos. Trans. R.*
598 *Soc. B* 333(1266): 1-13. 10.1098/rstb.1991.0057.

599 Hetzl J, Foerster AM, Raidl G, Scheid OM. 2007. CyMATE: a new tool for methylation analysis
600 of plant genomic DNA after bisulphite sequencing. *Plant J.* 51(3): 526-536. 10.1111/j.1365-
601 313X.2007.03152.x.

602 Hsieh T, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D. 2009.
603 Genome-wide demethylation of Arabidopsis endosperm. *Science* 324(5933): 1451-1454.
604 10.1126/science.1172417.

605 Inoue A, Jiang L, Falong L, Suzuki T, Zhang Y. 2017. Maternal H3K27me3 controls DNA
606 methylation-independent imprinting. *Nature* 547:419-424. 10.1038/nature23262.

607 Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during
608 interpsecies hybridization. *Curr. Biol.* 16(13): 1322-1328. 10.1016/j.cub.2006.05.045.

609 Kinoshita T., Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T. 2004.
610 One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science*
611 303(5657): 521-523. 10.1126/science.1089835.

612 Klosinska M, Picard CL, Gehring M. 2016. Conserved imprinting associated with unique
613 epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants* 2: 16145.
614 10.1038/nplants.2016.145.

615 Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. 2017. Canu: scalable and accurate long-
616 read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27: 722-736.
617 10.1101/gr.215087.116.

618 Korf I. 2004. Gene finding in novel genomes. *BMC Bioinform.* 5: 59. 10.1186/1471-2105-5-59.

619 Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-
620 Seq applications. *Bioinformatics* 27(11): 1571-1572. 10.1093/bioinformatics/btr167.

621 Lersten NR. 2004. Endosperm. In: Lersten NR, editor. In Flowering Plant Embryology. Blackwell
622 Publishing. p. 161–163.

623 Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, Walter J, Higgins M, Feil R, Reik W. 2004.
624 Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation
625 independent of DNA methylation. *Nature Genet.* 36:1291-1295. 10.1038/ng1468.

626 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-
627 3100. doi:10.1093/bioinformatics/bty191.

628 López-Fernández MP, Maldonado S. 2013. Programmed cell death during quinoa perisperm
629 development. *J. Exp. Bot.* 64(11): 3313-3325. 10.1093/jxb/ert170.

630 Losada JM, Bachelier JB, Friedman WE. 2017. Prolonged embryogenesis in *Austrobaileya*
631 *scandens* (Austrobaileyaceae): its ecological and evolutionary significance. *New Phyol.* 215(2):
632 851-864. 10.1111/nph.14621.

633 Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for
634 RNA-seq data with DESeq2. *Genome Biology*, **15**, 550. 10.1186/s13059-014-0550-8.

635 Moreno-Romero J, Del Toro-De León G, Yadav VK, Santos-González J, Köhler C. 2019.
636 Epigenetic signatures associated with imprinted paternally expressed genes in the Arabidopsis
637 endosperm. *Genom. Biol.* 20(1): 41.10.1186/s13059-019-1652-0.

638 Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM,
639 Udall JA, et al. 2016. Widespread natural variation of DNA methylation within angiosperms.
640 *Genom. Biol.* 17: 194. 10.1186/s13059-016-1059-0.

641 Orban I, Bouharmont J. 1998. Megagametophyte development of *Nymphaea nouchali Burm. F.*
642 (Nymphaeaceae). *Bot. J. Linn. Soc.* 126: 339-348. 10.1111/j.1095-8339.1998.tb01386.x.

643 Park K, Kim MY, Vickers M, Park J, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi
644 Y, Scholten S. 2016. DNA demethylation is initiated in the central cells of Arabidopsis and rice.
645 *Proc. Natl. Acad. Sci. U S A* 113(52): 15138-15143. 10.1073/pnas.1619047114.

646 Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R, Wolf JB. 2014. The evolution of
647 genomic imprinting: theories, predictions and empirical tests. *Heredity* 113: 119-128.
648 10.1038/hdy.2014.29.

649 Pei L, Zhang, L Li J, Shen C, Qiu P, Tu L, Zhang X, Wang M. 2019. Tracing the origin and
650 evolution history of methylation-related genes in plants. *BMC Plant Biol.* 19(1): 307.
651 10.1186/s12870-019-1923-7.

652 Pennington PD, Costa LM, Gutierrez-Marcos JF, Greenland AJ, Dickinson HG. 2008. When
653 Genomes collide; aberrant seed development following maize interploidy crosss. *Ann. Bot.* 101:
654 833-843. 10.1093/aob/mcn017.

655 Picard CL, Gehring M. 2020. Identification and comparison of imprinted genes across plant
656 species. In: Spillane C, McKeown P, editors. Plant Epigenetics and Epigenomics: Methods and
657 Protocols, Methods in Molecular Biology volume 2093. Springer Science+Business Media, LLC,
658 part of Springer Nature, p. 173-201. 10.1007/978-1-0716-0179-2_13.

659 Picard CL, Povilus RA, Williams BP, Gehring M. 2021. Transcriptional and imprinting complexity
660 in Arabidopsis seeds at single-nucleus resolution. *Nat. Plants* 7: 730-738. 10.1038/s41477-021-
661 00922-0.

662 Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. 2014. Natural epigenetic
663 polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife*, 3: e03198.
664 10.7554/eLife.03198.

665 Pignatta D, Gehring M. 2012. Imprinting meets genomics new insights and new challenges.
666 *Curr. Opin. Plant Biol.* 15(5): 530-535. 10.1016/j.pbi.2012.09.004.

667 Pignatta D, Novitzky K, Satyaki PRV, Gehring M. 2018. A variably imprinted epiallele impacts
668 seed development. *PLoS Genet.* 14(11): e1007469. 10.1371/journal.pgen.1007469.

669 Povilus RA, DaCosta JM, Grassa C, Satyaki PRV, Moeglein M, Jaenisch J, Xi Z, Mathews S,
670 Gehring M, Davis CC, Friedman WE. 2020. Water lily (*Nymphaea thermarum*) genome reveals
671 variable genomic signatures of ancient vascular cambium losses. *Proc. Natl. Acad. Sci. U S A*
672 117(15): 8649-8656. Doi.org/10.1073/pnas.1922873117.

673 Povilus RA, Diggle PK, Friedman WE. 2018. Evidence for parent-of-origin effects and
674 interparental conflict in seeds of an ancient flowering plant lineage. *Proc. Royal. Soc. B*
675 285(1872): 20172491. 10.1098/rspb.2017.2491.

676 Povilus RA, Friedman WE. 2022. Transcriptomes across fertilization and seed development in
677 the water lily *Nymphaea thermarum*: evidence for epigenetic patterning during reproduction.
678 *Plant Reprod.* 35(5): 161-178. 10.1007/s00497-022-00438-3.

679 Povilus RA, Gehring M. 2022. Maternal-filial transfer structures in endosperm: a nexus of
680 nutritional dynamics and seed development. *Curr. Opin. Plant Biol.* 65: 102121.
681 10.1016/j.pbi.2021.102121.

682 Povilus RA, Losada JM, Friedman WE. 2015. Floral biology and ovule and seed ontogeny of
683 Nymphaea thermarum, a water lily at the brink of extinction with potential as a model system for
684 basal angiosperms. *Ann. Bot.* 115: 211–226. 10.1093/aob/mcu235.

685 Raas MWD, Zijlmans DW, Vermeulen M, Marks H. 2022. There is another: H3K27-mediated
686 genomic imprnting. *Trends Genet.* 38(1):82-96. 10.1016/j.tig.2021.06.017.

687 Rodrigues JA, Ruan R, Nishimura T, Sharma MJ, Sharma R, Ronald PC, Fischer RL, Zilberman
688 D. 2013. Imprinted expression of genes and small RNA is associated with localized
689 hypomethylation of the maternal genome in rice endosperm. Proc. *Natl. Acad. Sci. U S A*
690 110(19): 7934-7939. 10.1073/pnas.1306164110.

691 Rudall PJ, Remizowa MV, Beer AS, Bradshaw E, Stevenson DW, Macfarlane TD, Tuckett RE,
692 Yadav SR, Sokoloff DD. 2008. Comparative ovule and megagametophyte development in
693 Hydatellaceae and water lilies reveal a mosaic of features among the earliest angiosperms.
694 *Ann. Bot.* 101(7): 941-956. 10.1093/aob/mcn032.

695 Santini L, Halbritter F, Titz-Teixeira F, Suzuki T, Asami M, Ma X, Ramesmayer J, Lackner A,
696 Warr N, Pauler F, Hippenmeyer S, Laue E, Karlik M, Bocj C, Beyer A, Perry ACF, Leeb M.
697 2021. Genomic imprinting in mouse blastocysts is predominantly associated with H3K27me3.
698 *Nature Commun.* 12:3804. 10.1038/s41467-021-23510-4.

699 Satyaki PRV, Gehring M. 2019. Paternally acting canonical RNA-directed DNA methylation
700 pathwa genes sensitize Arabidopsis endosperm to paternal genome dosage. *The Plant Cell*
701 31(7):1563-1578. 10.1105/tpc.19.00047.

702 Scott RJ, Spielman M, Bailey J, Dickinson HG, Parent-of-origin effects on seed development in
703 Arabidopsis thaliana. *Development* 125(17): 3329-3341. 10.1242/dev.125.17.3329.

704 Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio
705 prediction of alternative transcripts. *Nucleic Acids Res.* 34: W435–W439. 10.1093/nar/gkl200.

706 Tonosaki K, Ono A, Nagata H, Furuumi H, K Nonomura, Sato Y, Comai L, Hatakeyama K,
707 Kamakatsu T, Kinoshita T. 2024. Multi-layered epigenetic control of persistent and stage-
708 specific imprinted genes in rice endosperm. *Nat. Plants* 10: 1231–1245. 10.1038/s41477-024-
709 01754-4.

710 Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C, Xu W, Schnable PS,
711 Vaughn MW, Gehring M, Springer NM. 2011. Parent-of-origin effects on gene expression and
712 DNA methylation in the maize endosperm. *Plant Cell* 23(12): 4221-4233.
713 10.1105/tpc.111.092668.

714 Wick RR, Schultz MB, Zobel J, KE Holt. 2015. Bandage: interactive visualization of de novo
715 genome assemblies. *Bioinformatics* 31(20): 3350-3352. 10.1093/bioinformatics/btv383.

716 Williams JH, Friedman WE. 2002. Identification of diploid endosperm in an early angiosperm
717 lineage. *Nature* 415: 522-526. 10.1038/415522a.

718 Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, et al. 2013.
719 Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation
720 during maize endosperm development. *Plant Cell* 25(9): 3212-3227. 10.1105/tpc.113.115592.

721 Zhang M, Li N, He W, Zhang H, Yang W, Liu B. 2016. Genome-wide screen of genes imprinted
722 in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J.* 85(3):
723 424-436. doi: 10.1111/tpj.13116.

724 Zhang M, Shaojun X, Dong X, Zhao X, Zeng B, hen J, Li H, Yang W, Zhao H, Wang G, Chen Z,
725 SUn S. 2014. Genome-size high resolution parental-specific DNA and histone methylation maps
726 uncover patterns of imprinting regulation in maize. *Genome Res.* 24:167-176.
727 10.1101/gr.155879.113

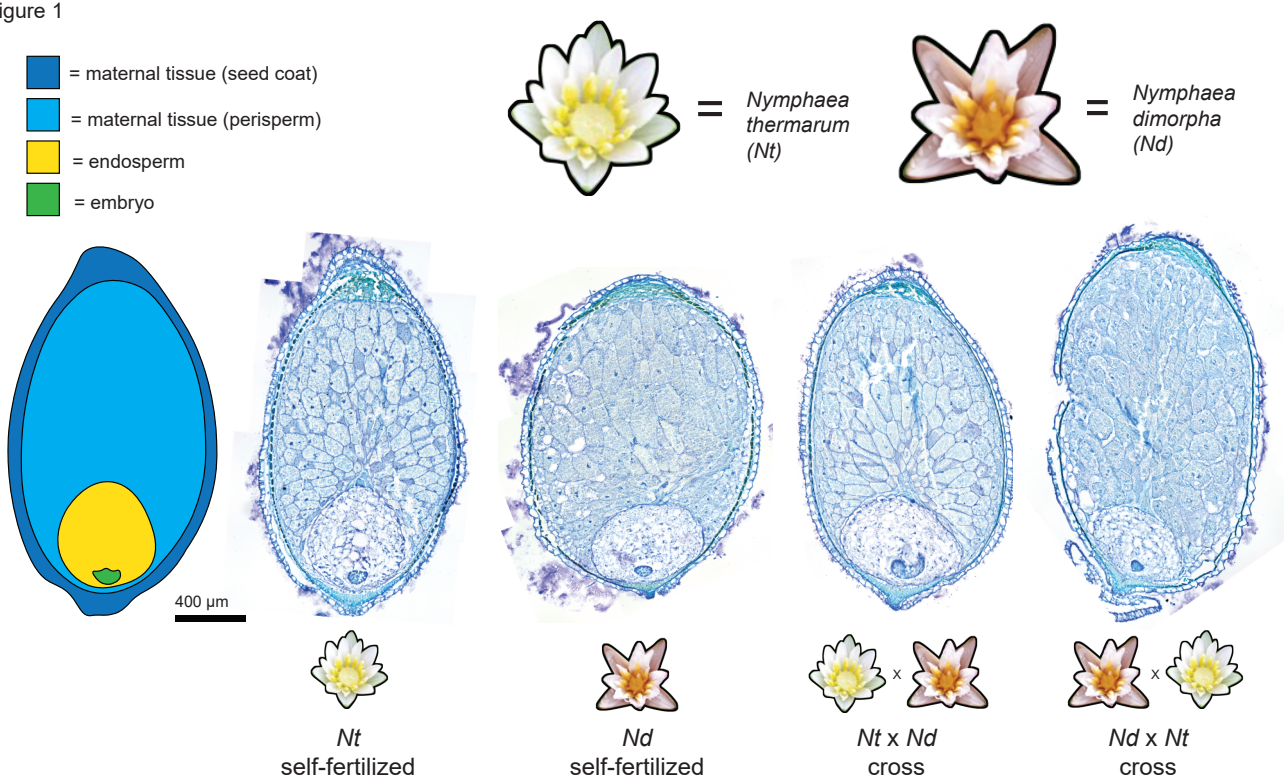
728 Zhang Z, Yu S, Jing L, Zhu Y, Jiang S, Xia H, Zhou Y, Sun D, Liu M, Li C, et al. 2021.
729 Epigenetic modifications potentially controlling the allelic expression of imprinted genes in
730 sunflower endosperm. *BMC Plant Biol.* 21(1): 570. 10.1186/s12870-021-03344-4.

731 Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome
732 assembler. *Bioinformatics* 29(21): 2669-6977. 10.1093/bioinformatics/btt476.

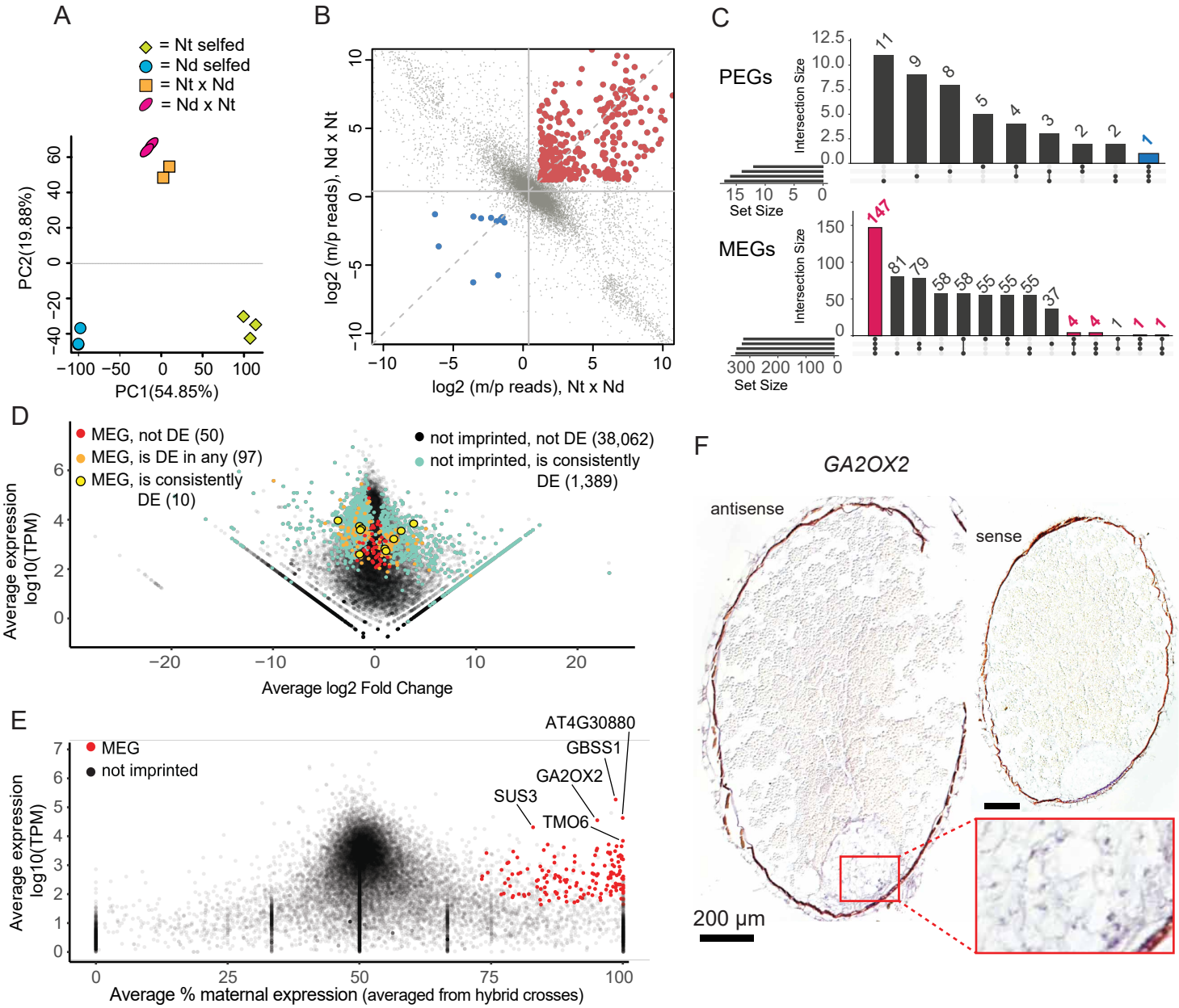
733
734
735 **Figures and Supplementary Figures and Tables**

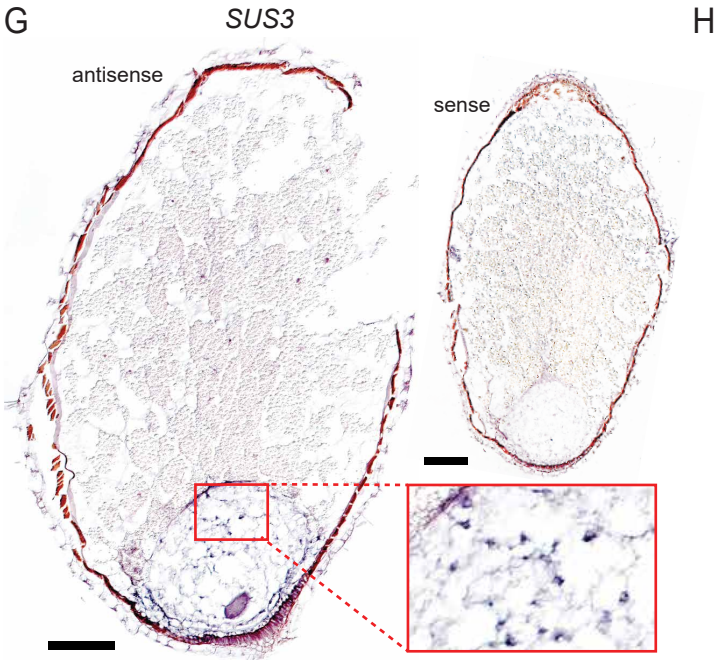
736
737

Figure 1

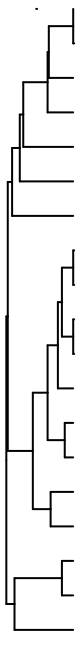

= maternal tissue (seed coat)
= maternal tissue (perisperm)
= endosperm
= embryo
400 µm
= Nymphaea thermarum (Nt)
= Nymphaea dimorpha (Nd)
x
x
Nt self-fertilized
Nd self-fertilized
Nt x Nd cross
Nd x Nt cross

740
741
742 **Figure 1: Seed structure in water lilies and F1 hybrids.** Young seeds (9-10 days after anthesis) of self-
743 fertilized *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*), as well as reciprocal crosses between the two
744 species. In all seeds, young embryos are surrounded by cellular, diploid endosperm, which in turn is
745 surrounded by a maternal nutrient storage tissue (perisperm).
746
747





H



	FDR	Fold	GO Biological Process
Response & development	4.3e−03	2.6	Response to endogenous stimulus
	4.3e−03	2.7	Response to hormone
	2.2e−02	2.6	Response to lipid
	1.0e−02	2.3	Response to O−contain. compound
	2.2e−02	3.7	Response to water deprivation
	2.2e−02	2.5	Plant organ development
	1.3e−02	16	Reactive nitrogen metabolic proc.
Carbohydrate processing	2.2e−02	6.6	Galacturonan metabolic proc.
	2.2e−02	6.7	Pectin metabolic proc.
	2.2e−02	7.8	Cell wall modification
	1.0e−02	6.7	Pectin catabolic proc.
	2.2e−02	6.3	Neg. reg. of catalytic activity
	2.2e−02	4.5	Carbohydrate catabolic proc.
	1.0e−02	7.3	Polysaccharide catabolic proc.
	2.2e−02	2.4	Carbohydrate metabolic proc.
	1.2e−02	3.8	Polysaccharide metabolic proc.
Secondary metabolism	2.2e−02	6.7	2ndary metabolite biosynthetic proc.
	2.4e−02	7.3	Phenylpropanoid biosynthetic proc.
	2.2e−02	19	V. long chain fatty acid biosynth. proc.

750
751 **Figure 2: Genetic imprinting in water lily endosperm.** A) Principal component analysis of RNA-seq data
752 from young endosperm samples isolated from self-fertilized *N. thermaurm* (Nt) and *N. dimorpha* (Nd)
753 seeds, and reciprocal crosses of *N. thermaurm* and *N. dimorpha*. B) Plot of the ratio of maternal to
754 paternal allele reads in endosperm of one reciprocal cross comparison. Female in cross listed first. Genes
755 that pass the stringency cut-offs for being called as imprinted are highlighted in red (MEGs, maternally
756 expressed genes) or blue (PEGs, paternally expressed genes). Gray dots indicate genes that are not
757 called as imprinted. C) Upset plot showing the number and consistency of genes called as MEGs or PEGs
758 across comparisons of different replicates. MEGs and PEGs called in at least 75% of comparisons are
759 highlighted in blue (PEGs) or red (MEGs). D) Average expression and average log2 fold change of genes
760 expressed > 1 TPM, in the comparison of Nt self-fertilized and Nt x Nd hybrid endosperm. Genes are
761 color-coded according to whether they are significantly differentially expressed in no comparisons or
762 consistently (75% or more of comparisons), and/or were identified as a MEG in no samples, in any
763 samples, or consistently (75% or more of samples). Number of genes in each category is noted. Similar
764 graphs for individual comparison types are shown in Supplementary Figure 3. E) Average expression and
765 average percent maternal expression for all expressed genes (TPM > 1). Genes called as MEGs in at least
766 75% of replicate comparisons are shown in red, with putative *Arabidopsis thaliana* homology indicated
767 for some MEGs of interest. F) In situ hybridization of putative homolog of *GA2OX2*. Inset shows
768 magnification of endosperm of sample treated with antisense probe. Scalebars = 200 µm. G) In situ
769 hybridization of putative homolog of *SUS3*. Inset shows magnification of endosperm of sample treated
770 with antisense probe. Scalebars = 200 µm. H) GO-enrichment analysis (for biological process terms) of
771 genes consistently called as MEG in the endosperm, reported with FDR-adjusted p-value and fold-
772 enrichment.
773
774

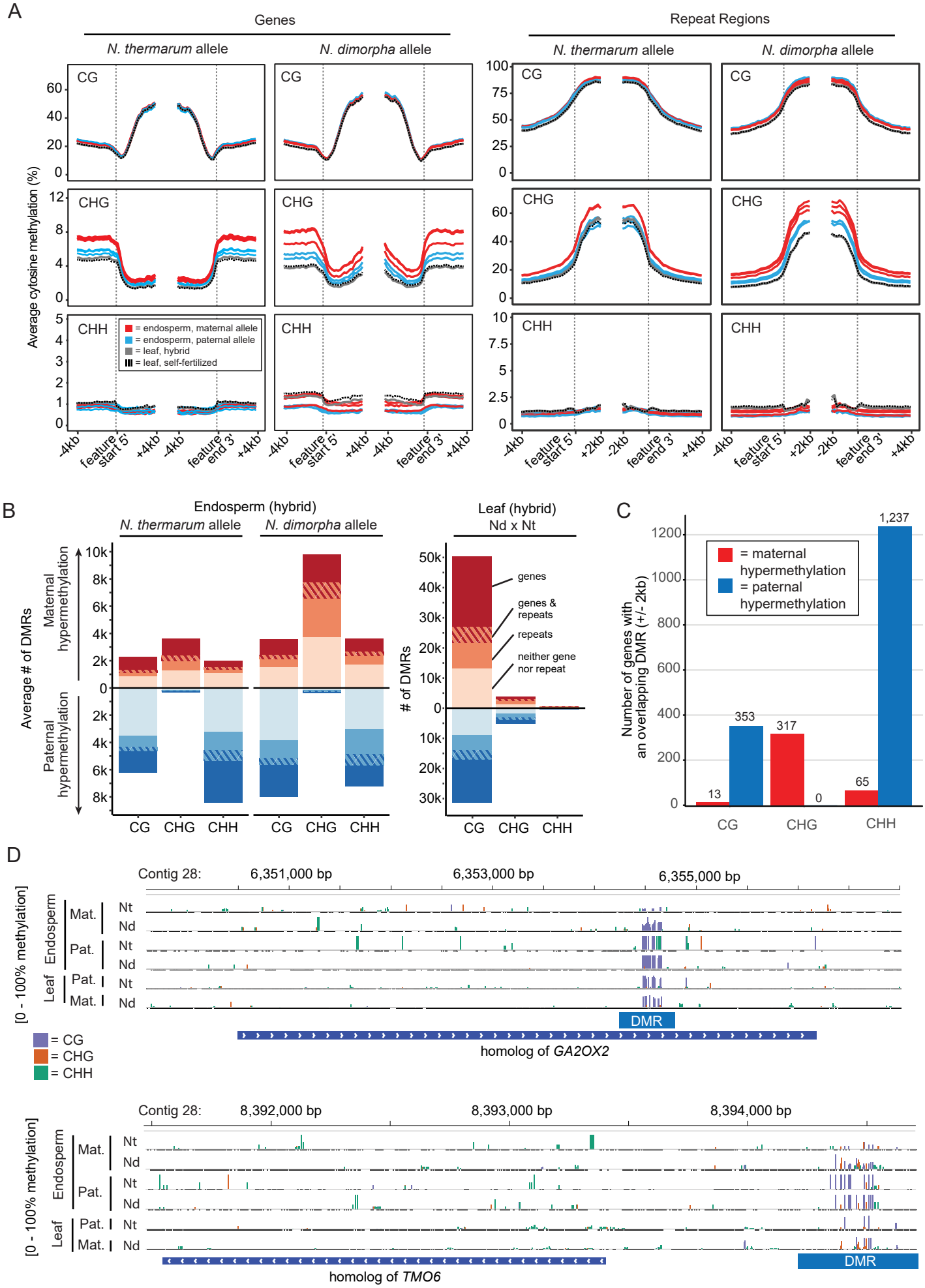
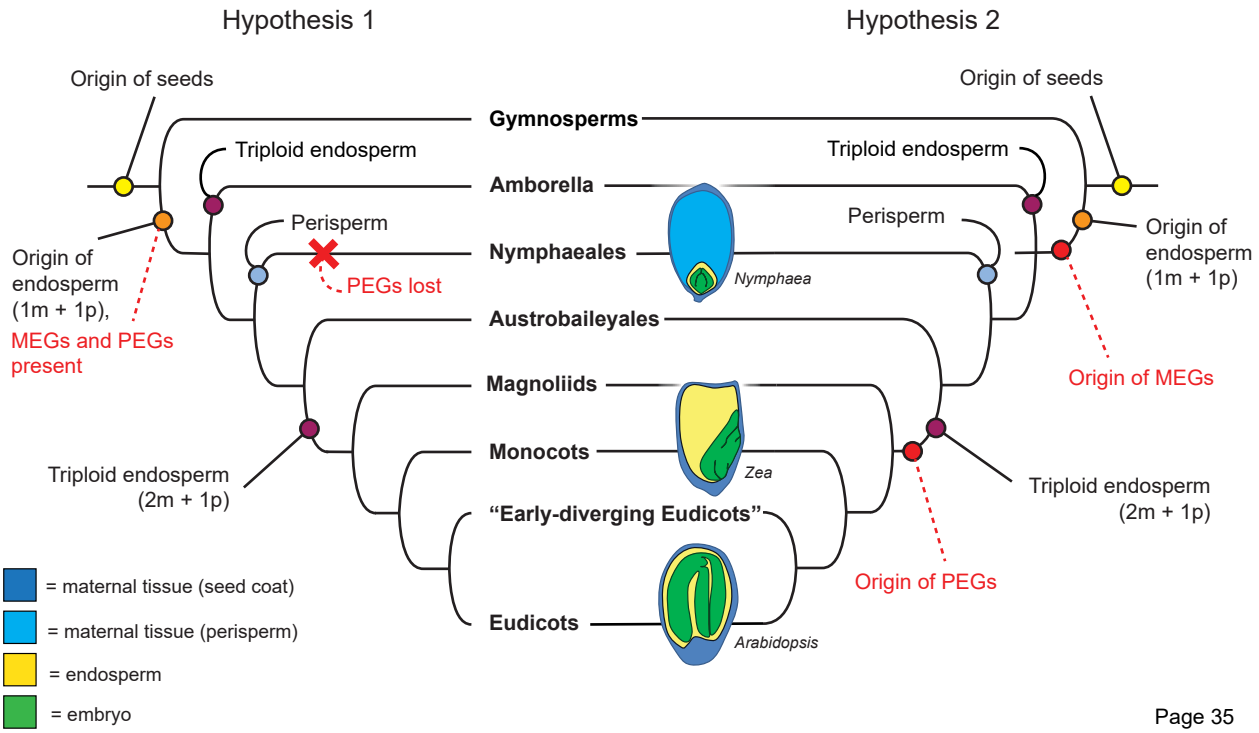

A
Genes
Repeat Regions
N. thermarum allele
N. dimorpha allele
Average cytosine methylation (%)
CG
CHG
CHH
= endosperm, maternal allele
= endosperm, paternal allele
= leaf, hybrid
= leaf, self-fertilized
-4kb
feature start 5'
+4kb
feature end 3'
+2kb
-2kb
B
Endosperm (hybrid)
Leaf (hybrid)
Nd x Nt
Average # of DMRs
Maternal hypermethylation
Paternal hypermethylation
# of DMRs
genes
genes & repeats
repeats
neither gene nor repeat
C
Number of genes with an overlapping DMR (+/- 2kb)
= maternal hypermethylation
= paternal hypermethylation
13
353
317
0
65
1,237
D
Contig 28:
6,351,000 bp
6,353,000 bp
6,355,000 bp
[0 - 100% methylation]
Endosperm
Leaf
Mat.
Pat.
Nt
Nd
DMR
homolog of GA2OX2
= CG
= CHG
= CHH
8,392,000 bp
8,393,000 bp
8,394,000 bp
homolog of TMO6

Figure 3: DNA methylation in water lily endosperm. A) Average parental allele cytosine methylation across genes or repeats in young endosperm of reciprocal crosses between *N. thermarum* and *N. dimorpha*, aligned at either the 5' transcription start site or 3' transcription end site. Each track represents the maternal or paternal alleles of an individual sampled cross (N=3 for *N. thermarum* x *N. dimorpha* crosses, N = 2 for *N. dimorpha* x *N. thermarum* crosses). Tracks are grouped by species and color-coded to indicate whether the track represents the maternally- or paternally-inherited alleles. B) Average number of maternally- or paternally-hypermethylated DMRs that overlap a gene and/or a repeat region, for each species. DMRs were called by comparing the genomes of each species when they were maternally- or paternally-inherited , for hybrid endosperm and leaf tissue (only one cross direction performed for leaf tissue). Red bars (above 0) indicate maternal hypermethylation, blue bars (below 0) indicate paternal hypermethylation. Bars are further color-coded to indicate the proportion of DMRs that overlap a gene, a gene and repeat, a repeat, or neither gene nor repeat. C) Number of genes (+/- 2kb) that consistently overlap maternally- or paternally-hypermethylated DMRs (gene had to have at least one DMR overlap of the indicated type in at least 75% of comparisons, as well as have an overlapping DMR of the opposite type in no more than 25% of comparisons). D) Genome browser snapshots of DNA methylation for homologs of *GA2OX2* and *TMO6*, showing examples of methylation patterning (blue = CG, orange = CHG, and green = CHH) on maternal or paternal genomes, for each species as the maternal and paternal genome, in endosperm and leaf tissue.


Hypothesis 1
Hypothesis 2
Origin of seeds
Origin of seeds
Gymnosperms
Triploid endosperm
Triploid endosperm
Amborella
Perisperm
Perisperm
Origin of endosperm (1m + 1p)
Nymphaeales
PEGs lost
Nymphaea
Origin of endosperm (1m + 1p), MEGs and PEGs present
Austrobaileyales
Origin of MEGs
Magnoliids
Monocots
Zea
Triploid endosperm (2m + 1p)
Triploid endosperm (2m + 1p)
"Early-diverging Eudicots"
Origin of PEGs
Eudicots
Arabidopsis
= maternal tissue (seed coat)
= maternal tissue (perisperm)
= endosperm
= embryo

800
801 **Figure 4: Two hypotheses regarding the evolution of genetic imprinting in endosperm, endosperm**
802 **ploidy, and nutrient storage strategies in angiosperm seeds.** Hypothesis 1 posits that MEGs and PEGs
803 were originally present in endosperm of the last common ancestor of angiosperms, and that PEGs were
804 lost in the water lily lineage in association with the origin of perisperm. Hypothesis 2 posits that MEGs
805 evolved as a maternal response to the addition of a paternal genome complement in a nutrient-
806 mediating tissue, while PEGs evolved as a response to the origin of triploid endosperm, which features
807 the addition of a second maternal genome complement. Dashed lines indicate events hypothesized as a
808 result of this study. Seed diagrams show diversity of mature seed structures, highlighting diversity in
809 developmental origin of the primary site of nutrient storage (*Nymphaea* = maternal perisperm; *Zea* =
810 endosperm; *Arabidopsis* = embryo).
811
812 **(See Supplementary Information PDF) Supplementary Figure 1: Image of individuals of *N. thermarum*,**
813 ***N. dimorpha*, and an F1 hybrid.**
814 Top left = *N. thermarum*, top right = *N. dimorpha*, bottom = F1 hybrid (demonstrating viability of hybrid
815 crosses). Identity of individuals was confirmed by genotyping (EM seq samples s104, 105, s106, see
816 Supplementary Table 1).
817
818 **(See Supplementary Information PDF) Supplementary Figure 2: Genome assemblies for *N. thermarum***
819 **and *N. dimorpha*.** A) Bandage diagrams for the updated *N. thermarum* genome assembly and the novel
820 *N. dimorpha* genome assembly. B) Alignment of the *N. thermarum* genome mapped against the *N.*
821 *dimorpha* genome.
822
823 **(See Supplementary Information PDF) Supplementary Figure 3: Differential gene expression analysis**
824 **between hybrid and parental endosperm with MEGs highlighted.** A) Comparison of fold change and
825 averaged expression (TPM) for *N. dimorpha* self-fertilized endosperm vs. *N. thermarm* x *N. dimorpha*
826 hybrid endosperm. B) Comparison of fold change and averaged expression (TPM) for *N. dimorpha* self-
827 fertilized endosperm vs. *N. dimorpha* x *N. thermarum* hybrid endosperm. C) Comparison of fold change
828 and averaged expression (TPM) for *N. thermarum* self-fertilized endosperm vs. *N. thermarm* x *N*
829 *dimorpha* hybrid endosperm. D) Comparison of fold change and averaged expression (TPM) for *N.*
830 *thermarum* self-fertilized endosperm vs. *N. dimorpha* x *N. thermarum* hybrid endosperm. E) Upset graph
831 showing overlap of MEGs that are significantly DE for the the four different comparison types; the
832 putative *Arabidopsis thaliana* homologs of the 5 consistently DE MEGs is listed.

833 **(See Supplementary Information PDF) Supplementary Figure 4: In situ hybridization positive control**
834 **for perisperm, and *SUS3* expression in seeds of different crosses and self-fertilizations.** A) RNA *in situ*
835 hybridization of putative homolog of a terpene synthase sub-family in *N. thermarum* seeds, showing
836 that detection of gene expression in the perisperm is possible. Results of experiments performed with
837 antisense and sense probes are shown. Scalebars = 200 µm. B) In situ hybridization of putative homolog
838 of *SUS3*, performed in seeds of reciprocal crosses of *N. thermarum* and *N. dimorpha* and in seeds from
839 *N. dimorpha* self-fertilizations. Black arrowheads indicate detection of signal, white arrowheads indicate
840 absence of signal. Scalebars = 200 µm. Results of experiments performed with antisense and sense
841 probes are shown.

842
843 **(See Supplementary Information PDF) Supplementary Figure 5: DNA methylation, including leaf**
844 **tissues from non-hybrid *N, thermarum* and *N. dimorpha* plants**
845 A) Number of DMRs that are hyper- or hypomethylated in the *N. dimorpha* compared to *N. thermarum*
846 genome, that overlap a gene and/or a repeat region, for each species, in leaf tissue. DMRs were called
847 by comparing the genomes of each species within the same sample (hybrid tissue, one sample) or
848 between samples of leaves of each parental species (one sample of each species). B) Genome browser
849 snapshots of DNA methylation for homologs of *GA2OX2* and *TMO6*, showing examples of methylation
850 patterning (blue = CG, red = CHG, and green = CHH) on the genome for each species , in endosperm and
851 leaf tissue, including leaves of non-hybrid *N, thermarum* and *N. dimorpha* plants. Black notches indicate
852 cytosines for which there was sufficient data to include, but were unmethylated.
853
854 **Supplementary Table 1: Summaries for sequencing, mapping, and genome assemblies and**
855 **annotations.**
856
857 **Supplementary Table 2: Results and summaries of imprinting calling tests, report of imprinted loci**
858 **that overlap DMRs, and summary of imprinting status of *Nymphaea* homologs of conserved imprinted**
859 **genes.**
860
861 **Supplementary Table 3: Results and summaries of differential gene expression analysis between**
862 **intraspecies and interspecies crosses**
863
864 **Supplementary Table 4: Results and summaries of imprinting calling tests, when data is adjusted for**
865 **potential maternal tissue contamination of endosperm samples (assuming 50% and 25%**
866 **contamination).**
867
868 **Supplementary Table 5: Report of DMR overlap with genomic features, as well as report and summary**
869 **of genes with DMR overlaps**
870