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

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

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

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

[3]Current affiliation: USDA-ARS, Plant Genetic Resources Unit, Geneva NY 14456.

*Corresponding author

Summary

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 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[1,2,3,4]. Knowledge of endosperm gene imprinting and epigenetic patterning is derived from experiments performed in a phylogenetically-wide array of monocot and eudicot plants[5,6]. However, information from angiosperm lineages whose origins predate the monocot-eudicot divergence (such as Nymphaeales, water lilies) is extremely limited. Additionally, Nymphaeales are an intriguing lineage to investigate seed genetic and epigenetic phenomena, as it is characterized by diploid endosperm and a maternal storage tissue (perisperm), both of which are unusual across angiosperm diversity[7,8,9,10,11,12]. 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 endosperm hypomethylation and maternally-expressed imprinted genes are an ancestral condition for endosperm, and that 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.

Results and Discussion

We sought to determine whether gene imprinting and endosperm DNA hypomethylation, which has been mechanistically linked to gene imprinting, preceded the origin of parental dosage-imbalanced

(triploid) endosperm. *N. thermarum* has been developed as an experimental system for the Nymphaeales[12,13,14], an order which is sister to all other angiosperms except for *Amborella tricopoda*. Assessing parent-of-origin effects at the molecular level requires sequence polymorphisms, of which there are few within the highly-inbred extant populations of *N.* thermarum in cultivation[12]. We therefore assessed parent-of-origin effects by examining F1 tissue from crosses to *N. dimorpha*. We confirmed the internal structure of young seeds of *N. thermarum*, *N. dimorpha*, and F1 reciprocal crosses between these two species and determined that we could ensure consistency in developmental stage between crosses [Figure 1A].

The endosperm of monocot and dicots is hypomethylated compared to the embryo and vegetative tissues due to demethylation of the maternally-inherited genome[15,16,17,18,19,20,21]. This epigenetic asymmetry establishes gene imprinting after fertilization. To determine whether epigenetic reprogramming is a conserved feature of endosperm, we examined DNA methylation in somatic and reproductive tissues of *N. thermarum* [Figure 1B]. We first produced an improved assembly and annotation of the *N. thermarum* genome that included long-read sequencing data [Supplementary Table 1] [Supplementary Figure 1] and then characterized DNA methylation patterning in leaves and seed tissues (embryo, endosperm) during early and later development by bisulfite DNA sequencing [Figure 2A-C]. A PCA analysis of overall cytosine methylation showed that samples clustered according to seed tissue type (PC1, 19.2%) and developmental stage (PCA2, 15.6%) [Figure 2A]. In the CG context, endosperm is globally hypomethylated relative to leaf tissue and to embryos. The difference between offspring tissues increases between early and late seed development, apparently due to a larger increase in embryo methylation over developmental time [Figure 2B]. Hypomethylation of the endosperm compared to the embryo is also seen in the CHG context, but is not observed in the CHH context, where young endosperm appears to be slightly more methylated than embryos. These patterns are also reflected in the number of significantly differentially methylated regions (DMRs) that are either

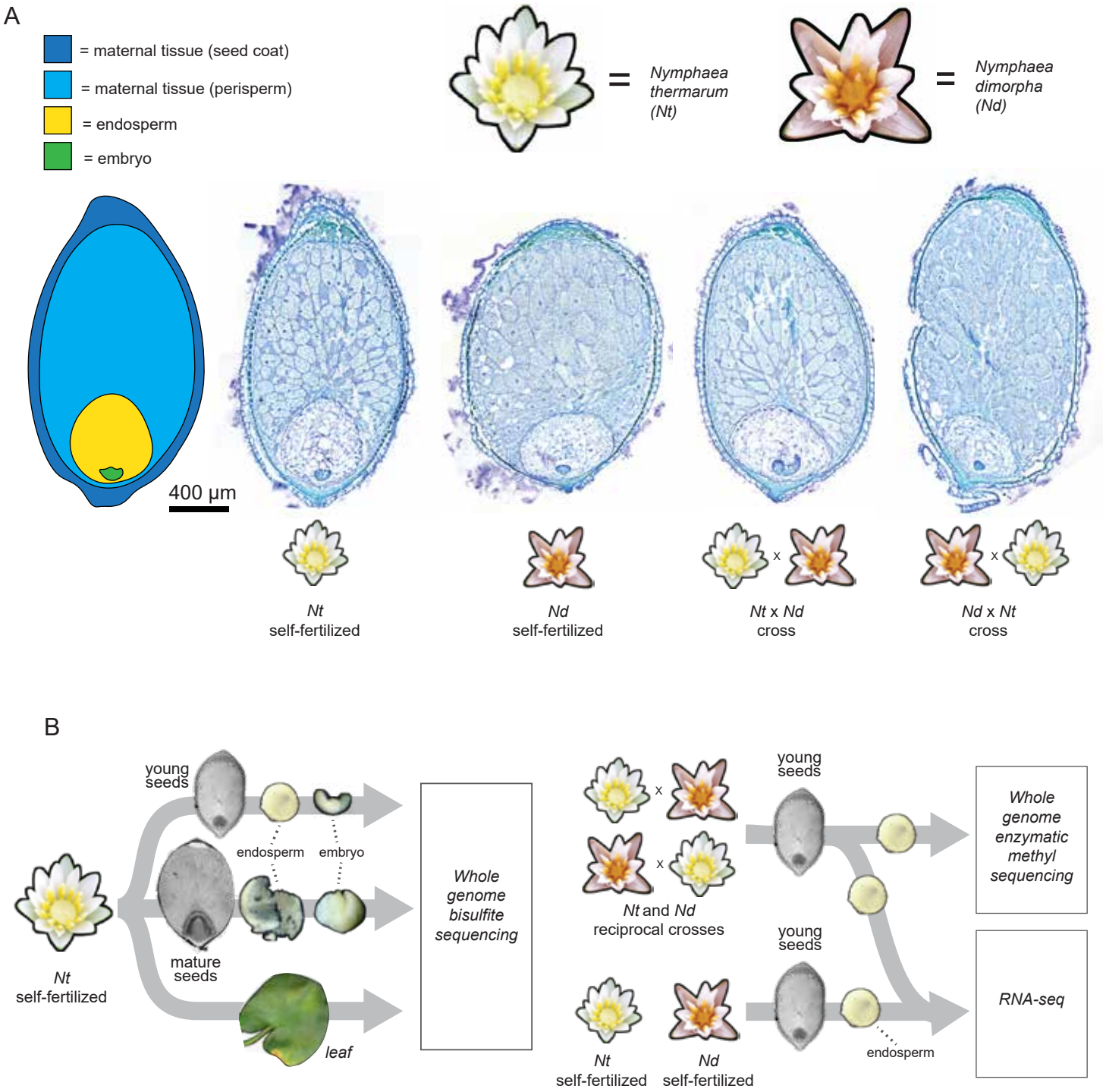


Figure 1: Overview of seed structure in water lilies and experimental workflows. A) Young seeds of self-fertilized *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*), as well as reciprocal crosses between the two species. In all seeds, young embryos are surrounded by cellular, diploid endosperm, which in turn is surrounded by a maternal nutrient storage tissue (perisperm). B) For this study, three main datasets were generated to define seed characteristics. Whole-genome DNA methylation was characterized for embryo and endosperm isolated from young and mature seeds, as well as leaf tissue, of *N. thermarum*. Whole-genome DNA methylation was characterized for young endosperm from seeds of reciprocal crosses of *N. thermarum* and *N. dimorpha*. RNA-seq data was also collected for young endosperm from seeds of reciprocal crosses, as well as self-fertilized *N. thermarum* and *N. dimorpha* seeds.

Figure 2

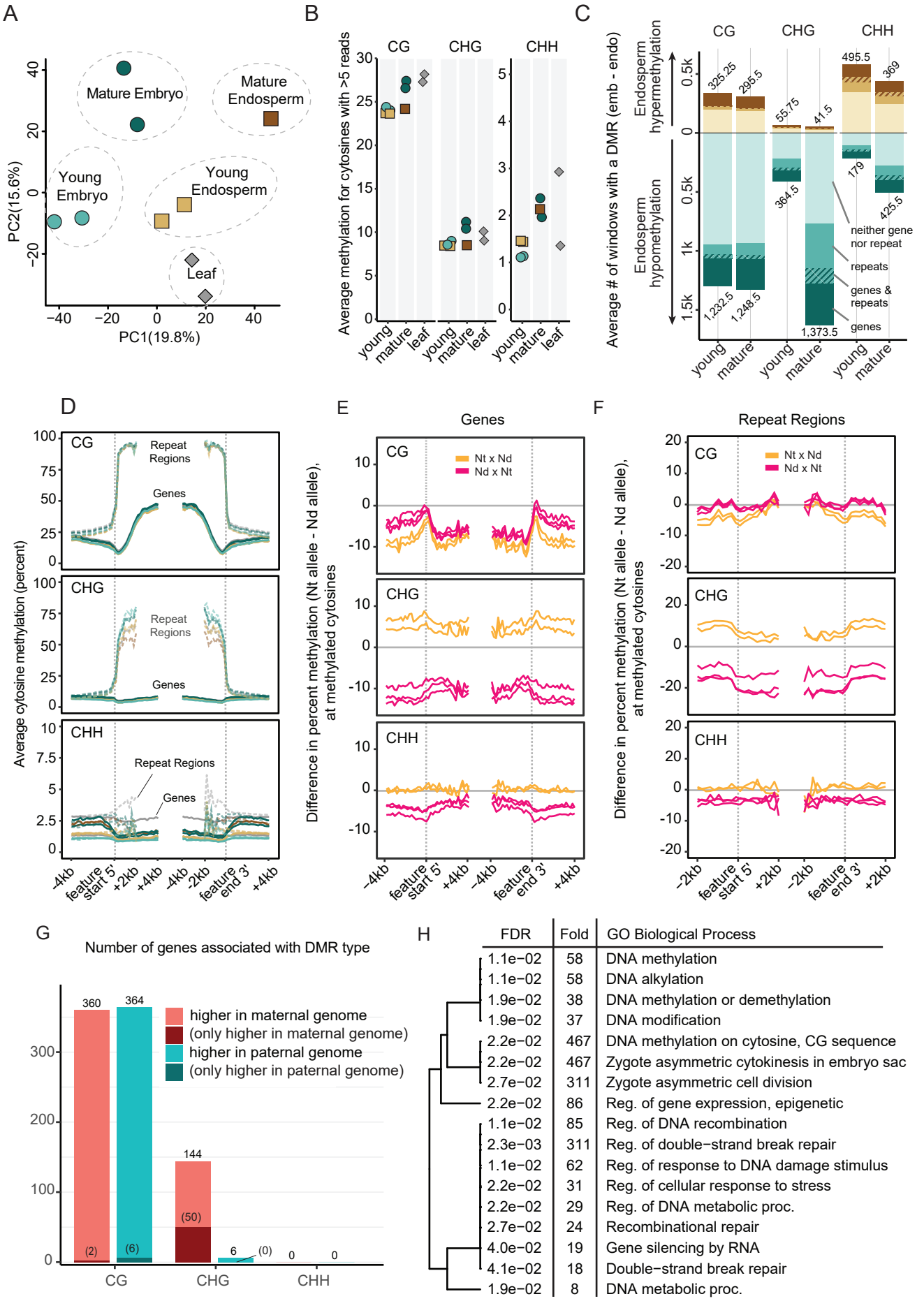

A
Mature Embryo
Mature Endosperm
Young Embryo
Young Endosperm
Leaf
PC2(15.6%)
PC1(19.8%)
B
CG
CHG
CHH
Average methylation for cytosines with >5 reads
young
mature
leaf
C
CG
CHG
CHH
Endosperm hypermethylation
Endosperm hypomethylation
Average # of windows with a DMR (emb - endo)
325.25
295.5
55.75
41.5
495.5
369
179
364.5
425.5
1,232.5
1,248.5
1,373.5
neither gene nor repeat
repeats
genes & repeats
genes
young
mature
D
CG
CHG
CHH
Repeat Regions
Genes
Average cytosine methylation (percent)
-4kb
feature start 5'
+2kb
+4kb
-2kb
feature end 3'
E
Genes
Nt x Nd
Nd x Nt
Difference in percent methylation (Nt allele - Nd allele), at methylated cytosines
F
Repeat Regions
G
Number of genes associated with DMR type
higher in maternal genome
(only higher in maternal genome)
higher in paternal genome
(only higher in paternal genome)
360
364
144
(50)
(2)
(6)
6
(0)
0
0
CG
CHG
CHH
H
FDR | Fold | GO Biological Process
1.1e−02 | 58 | DNA methylation
1.1e−02 | 58 | DNA alkylation
1.9e−02 | 38 | DNA methylation or demethylation
1.9e−02 | 37 | DNA modification
2.2e−02 | 467 | DNA methylation on cytosine, CG sequence
2.2e−02 | 467 | Zygote asymmetric cytokinesis in embryo sac
2.7e−02 | 311 | Zygote asymmetric cell division
2.2e−02 | 86 | Reg. of gene expression, epigenetic
1.1e−02 | 85 | Reg. of DNA recombination
2.3e−03 | 311 | Reg. of double−strand break repair
1.1e−02 | 62 | Reg. of response to DNA damage stimulus
2.2e−02 | 31 | Reg. of cellular response to stress
2.2e−02 | 29 | Reg. of DNA metabolic proc.
2.7e−02 | 24 | Recombinational repair
4.0e−02 | 19 | Gene silencing by RNA
4.1e−02 | 18 | Double−strand break repair
1.9e−02 | 8 | DNA metabolic proc.

Figure 2: DNA methylation in water lily endosperm and other tissues. A) Principal components analysis of cytosine methylation of embryo and endosperm isolated from young and mature seeds, as well as leaf tissue, of *N. thermarum.* B) Average percent methylation across all cytosines (with >5 reads), in the CG, CHG, and CHH contexts). C) Average number of significantly differentially methylated regions (DMRs) between embryo and endosperm samples (averaged across comparisons of biological replicates). The number of DMRs that overlap a gene, repeat region, both, or neither is shown. D) Ends-analysis showing average cytosine methylation across genes (solid lines) and repeat regions (dashed lines), in the different cytosine contexts, for all *N. thermarum* samples. Sample colors indicated as in panel A. E) Ends-analysis showing the difference in parental allele cytosine methylation (*N. thermarum* allele methylation - *N. dimorpha* allele methylation) across genes in young endosperm of reciprocal crosses between *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*) For crosses, female in the cross is listed first. F) Ends-analysis showing the difference in parental allele cytosine methylation (*N. thermarum* allele methylation - *N. dimorpha* allele methylation) across repeat regions in young endosperm of reciprocal crosses between *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*). G) Number of genes (+/- 2kb) that overlap DMRs between maternal and paternal genomes of *N. thermarum* and *N. dimorpha* crosses. Genes were categorized according to whether they overlapped at least one DMR that was more highly methylated in the maternally-inherited genome, only DMR(s) that were higher in the maternally-inherited genome, at least one DMR that was higher in the paternally-inherited genome, or only DMR(s) that were higher in the paternal genome. H) GO-enrichment analysis (biological process terms) of the set of 50 genes (+/- 2kb) that overlapped only with DMRs that were more highly methylated in the maternally-inherited genome.

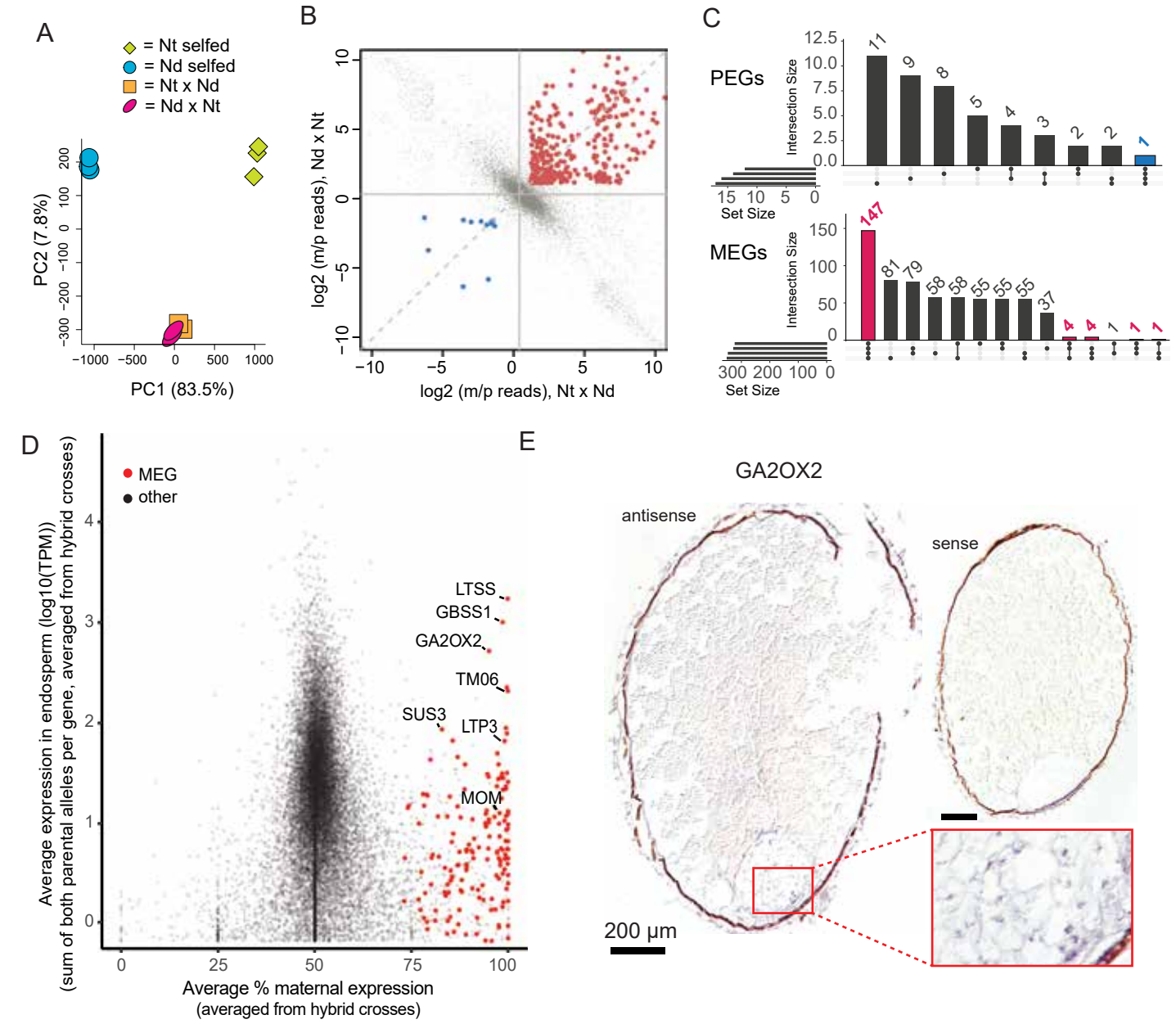
hyper- or hypomethylated in the endosperm, relative to embryos [Figure 2C] [Supplementary Table 2]. When methylation is analyzed separately across genes and repeat elements [Figure 2D], patterns similar to what has been documented in other angiosperms[22] are found: on average repeat regions, including TEs (transposable elements), have higher body methylation relative to genes in all contexts. In the CHG context, endosperm repeat/TE methylation appears to be lower than in leaf and embryo samples. Repeats/TEs are also hypomethylated in Arabidopsis endosperm, although this is not restricted to the CHG context[16,23,24]. Overall these patterns are consistent with descriptions of the embryo and endosperm of other flowering plants including Arabidopsis, maize, and rice[15,16,17,18,19,20]. These findings suggest that endosperm hypomethylation in symmetric contexts is a deeply conserved aspect of the endosperm genomic landscape.

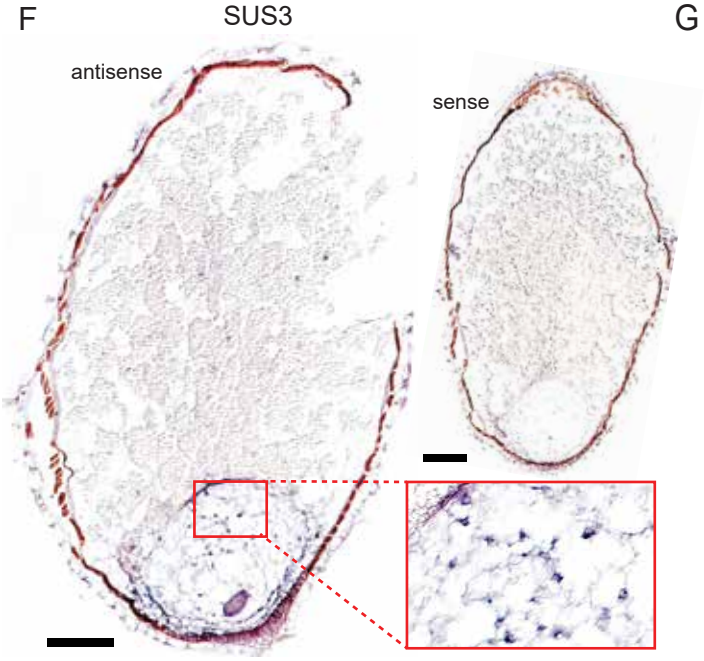
Next, we made use of reciprocal hybrid crosses between *N. thermarum* and *N. dimorpha* to permit allele-specific characterization of DNA methylation patterning in young endosperm [Figures 1B, 2E-F] [Supplementary Table 1]. This work included de novo genome assembly and annotation for *N. dimorpha* (which was formerly known as *N. minuta*) [Supplementary Table 1] [Supplementary Figure 1]. When comparing the difference in average methylation between *N. thermarum* and *N. dimorpha* alleles over all genic [Figure 2E] and repeat [Figure 2F] regions, we found that in the CG context *N. dimorpha* alleles are more highly methylated in both reciprocal cross directions, while in the CHH context the *N. dimorpha* alleles were more highly methylated when *N. dimorpha* was the mother. The higher CG methylation of *N. dimorpha* compared to *N. thermarum* complicates potential analysis of parent-of-origin dependent CG methylation patterns. By contrast, in the CHG context maternal alleles were modestly more highly methylated (5-10%), regardless of cross direction. CHG hypermethylation of maternal alleles has been observed in *A. lyrata* endosperm in gene bodies, but to a much greater extent[25]. The higher methylation of maternal alleles in the CHG context only was also borne out by the number of genes that are associated with a DMR (genic region overlaps with a DMR or has a DMR 2kb 5'

or 3') in the different cytosine contexts [Figure 2G] [Supplementary Table 2]. Whereas in the CG context a similar number of genes were associated with DMRs that were higher in the maternal or paternal genomes, in the CHG context 144 genes were associated with maternal allele hyperDMRs (hypermethylated differentially methylated regions) compared to the 6 genes associated with paternal allele hyperDMRs. Fifty genes were exclusively associated with higher-in-maternal-genome DMRs; these genes were significantly enriched for GO terms that included epigenetic modifications, regulation of gene silencing, and zygotic cell division patterns [Figure 2H]. Intriguingly, one of the Nymphaea homologs of METHYLTRANSFERASE 1 (MET1) is among this set. MET1 is itself a regulator of DNA methylation and has been linked to regulation of genetic imprinting in endosperm of Arabidopsis[26,27].

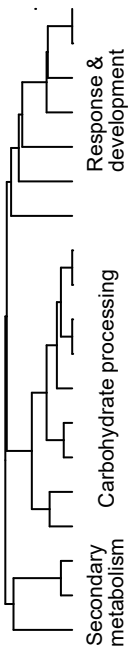
Having found evidence that the endosperm has a DNA methylation state distinct from other tissues, we next investigated whether parent-of-origin effects on gene expression occur. We again made use of reciprocal crosses between *N. thermarum* and *N. dimorpha* as well as self-fertilized seeds, and isolated RNA from young endosperm for 2-3 biological replicates [Figure 1B] [Supplementary Table 1]. A PCA analysis of total gene expression revealed that biological replicates clustered together according to cross type, with hybrid endosperm samples midway along PC1 (83.51%) between endosperm from *N. thermarum* self-fertilized seeds and *N. dimorpha* self-fertilized seeds (Figure 3A). In each possible comparison of a *Nt* x *Nd* and *Nd* x *Nt* cross, we identified transcripts that showed a bias in the number of reads mapping uniquely to either the maternally or paternally-inherited alleles, in both cross directions (Figure 3B) [Supplementary Table 3]. While we found small numbers of paternally-expressed imprinted genes (PEGs) and larger numbers of maternally-expressed imprinted genes (MEGS) in individual cross comparisons, only 1 PEG was consistent in at least 75% of comparisons (Figure 3C); this PEG is a homolog of *CELLULOSE SYNTHASE LIKE G2* (*CLSG2*). In contrast, 157 MEGs were consistent in at least 75% of comparisons [Supplementary Table 3]. Thus, imprinted expression exists in Nymphaea endosperm, but is largely restricted to MEGs.

Figure 3





G



FDR	Fold	GO Biological Process
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.
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.
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.

Figure 3: Genetic imprinting in water lily endosperm. A) Principal component analysis of RNA-seq data from young endosperm samples isolated from self-fertilized *N. thermaurm* (*Nt*) and *N. dimorpha* (*Nd*) seeds, and reciprocal crosses of *N. thermaurm* and *N. dimorpha*. B) Plot of the ratio of maternal or paternal allele reads in endosperm of one reciprocal cross comparison. Female in cross listed first. Genes that pass the stringency cut-offs for being called as imprinted are highlighted in red (MEGs, maternally expressed genes) or blue (PEGs, paternally expressed genes). Gray dots indicate genes that are not called as imprinted. C) Upset plot showing the number and consistency of genes called as MEGs or PEGs across comparisons of different replicates. MEGs and PEGs called in at least 75% of comparisons are highlighted in blue (PEGs) or red (MEGs). D) Average expression and average percent maternal expression for all expressed genes (TPM > 1). Genes called as MEGs in at least 75% of replicate comparisons are shown in red, with putative Arabidopsis thaliana homology indicated for some MEGs of interest. E) In situ hybridization of putative homolog of GA2OX2. Inset shows magnification of endosperm of sample treated with antisense probe. Scalebars = 200 µm. F) In situ hybridization of putative homolog of SUS3. Inset shows magnification of endosperm of sample treated with antisense probe. Scalebars = 200 µm. G) GO-enrichment analysis (for biological process terms) of genes consistently called as MEG in the endosperm.

The identification of MEGs can be influenced by contamination with maternal tissue or transcripts in endosperm samples. Notably, PCA results indicate that hybrid endosperm of both cross directions was more like each other transcriptionally than endosperm of their mother species, suggesting no significant, wide-spread maternal contamination (Figure 3A). We additionally performed RNA in situ hybridizations on seeds of self-fertilized *N. thermarum* for a set of identified MEGs to test whether they showed substantial expression in maternal tissue or lack expression in the endosperm (both of which would indicate maternal tissue contamination). Target genes were selected for high maternal expression bias and high expression (Figure 3D), including a gibberellic acid oxidase homolog (*GA2OX2*) (Figure 3E) and a sucrose synthase 3 homolog (*SUS3*) (Figure 3F). We furthermore confirmed that the high starch and carbohydrate content of the perisperm were not interfering with the in situ hybridization, as we were able to detect expression of a subfamily of terpene synthases in the perisperm, as well as in the endosperm [Supplementary Figure 2]. For both *GA2OX2* and *SUS3*, we detected expression in the endosperm and not in the perisperm, indicating that for these and likely other MEGs, the identified maternal expression bias is unlikely to be caused by contamination with maternal tissue. For *SUS3*, we also performed in situ hybridizations on seeds from reciprocal crosses and from *N. dimorpha* self-fertilizations and found similar expression patterns [Supplementary Figure 2].

Importantly, *SUS3* is a key part of nutrient processing in the endosperm in other species[28]. Overall, MEGs were significantly enriched for GO annotations associated with response and development, carbohydrate processing, and secondary metabolism (Figure 3G). The enrichment for processes integral to development and nutrient dynamics in seeds is similar to what has been found in other species[5,29] and is congruent with genetic imprinting being associated with parental conflict over maternal resource investment[30].

Few genetically imprinted loci were consistently associated (had a DMR that overlapped the gene region or +/- 2kb) with an allele-specific DMRs in all samples [Supplementary Table 2]. Only 4 MEGs

were associated with maternal allele hyperDMRs, and three MEGs were associated with paternal allele hyperDMRs; two of these MEGs were consistently associated with both types of DMRs [Supplementary Table 3]. The one identified PEG was not associated with any significant methylation differences between parental alleles. Notably one MEG, a homologue of the Arabidopsis transcription factor *TARGET OF MONOPTEROS 6* (*TMO6*), was associated with exclusively paternal allele CG hyperDMRs.

In summary, we found that hypomethylation of endosperm in the symmetric CG and CHG contexts and genetic imprinting occur in endosperm of Nymphaea seeds. Both endosperm hypomethylation and genetic imprinting have been suggested to be strategies that can alter the effective maternal or paternal genome/gene dosage. Changes to absolute parental genome dosage (and ploidy) of endosperm has also occurred repeatedly during angiosperm evolution. Endosperm of the Nymphaeales is diploid with a 1:1 maternal:paternal genome ratio, which has been suggested to be the ancestral state for endosperm[7,8,9,10,11,12]. In a common ancestor of monocots and eudicots the addition of an extra maternal genomic complement likely occurred, resulting in the triploid endosperm that is found in most flowering plants, including the vast majority of taxa in which endosperm epigenetic patterning and genetic imprinting have been studied[30]. The ratio of maternal and paternal genomes is important for endosperm viability, as evidenced by the phenomena of parental genome dosage sensitivity[30]. 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[30,31,32]. Endosperm parental genome dosage sensitivity has been documented in *N. thermarum*[13], indicating that parent-of-origin effects on endosperm development predate the evolution of triploid endosperm. Our additional discovery of endosperm hypomethylation and genetic imprinting in Nymphaea suggests that these characters/processes predate the evolution of triploid endosperm and are likely to be have been either

co-opted from preexisting ancestral molecular programmes or are themselves novelties associated with the origin of endosperm itself. In either case, a 2:1 maternal:paternal genome dosage ratio is not a requirement for either endosperm hypomethyation or genetic imprinting. Furthermore, our results suggest that parent-of-origin effects on endosperm development in Nymphaea may be linked to parent-of-origin-specific DNA methylation patterning in the CHG context or maternally-expressed imprinted genes, but not extensively to paternally-expressed imprinted genes.

Another unusual seed character of the Nymphaeales is that nutrients are primarily stored in a perisperm (which is derived from maternal tissue and contains no paternal genome contribution) instead of in offspring tissues[33,12]. Genetic imprinting in the endosperm has been linked to parental conflict over the investment of maternal resources, namely that PEGs are predicted to promote nutrient investment while MEGs are predicted to restrict excessive investment[30,34]. Nutrient dynamics during seed development consists of acquisition, assimilation, storage, mobilization, and transfer of maternal resources[35]. Although the storage function in Nymphaeales occurs in a maternal tissue, perisperm nutrient storage is only initiated after fertilization[12], suggesting influence of offspring tissues on this process.

In the context of parental conflict over the investment of maternal resources in offspring, the evolution of maternal storage tissues and the notable paucity of PEGs in Nymphaea endosperm give rise to two distinct hypotheses about the early evolution of angiosperm seed development and endosperm molecular/genetic processes [Figure 4]. For the first hypothesis, if MEGs and PEGs were both present in the ancestrally diploid endosperm of the earliest flowering plants, then PEGs were largely lost in association with the evolution of perisperm in the Nymphaeales [Figure 4, left]. If genetic imprinting is a tool to establish maternal or paternal control over resource investment in offspring[30], perhaps PEGs are no longer an effective paternal strategy when the mother establishes primary control by storing invested resources in a maternally-derived tissue that does not have a paternal genome contribution. A

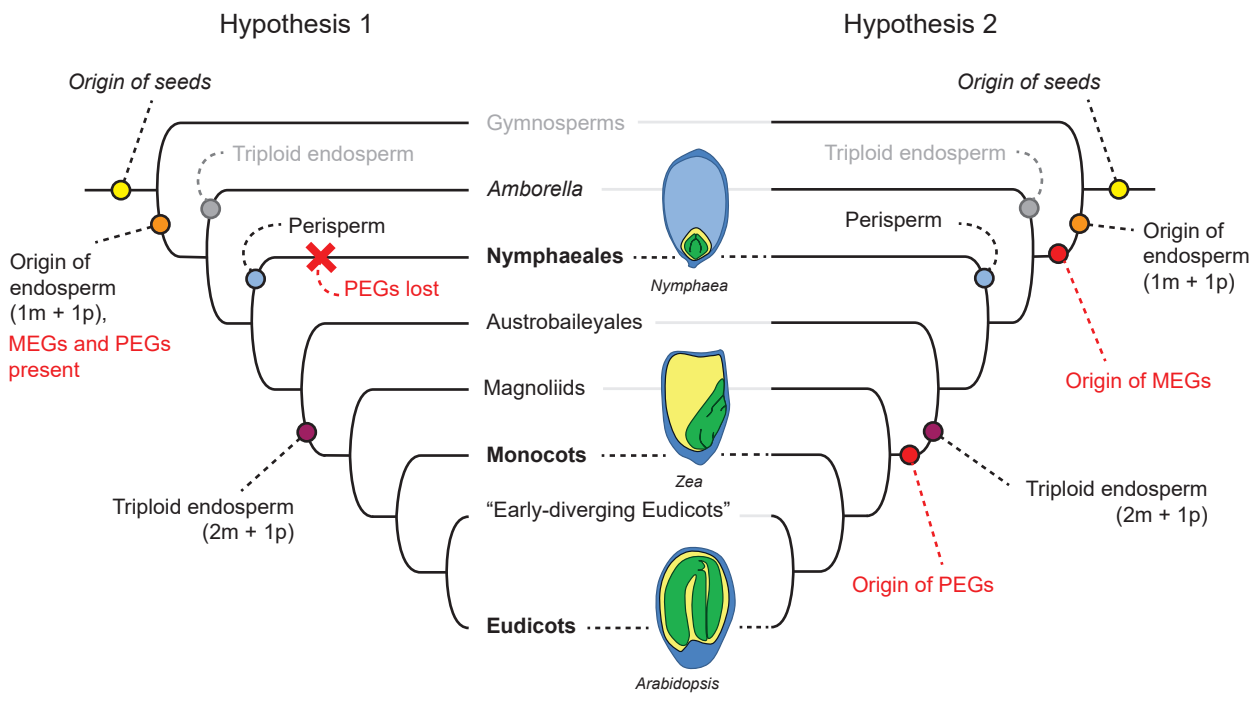


Figure 4: Two hypotheses regarding the evolution of genetic imprinting in endosperm, endosperm ploidy, and nutrient storage strategies in angiosperm seeds. Hypothesis 1 posits that MEGs and PEGs were originally present in endosperm of the last common ancestor of angiosperms, and that PEGs were lost in the water lily lineage in association with the origin of perisperm. Hypothesis 2 posits that MEGs evolved as a maternal response to the addition of a paternal genome complement in a nutrient-mediating tissue, while PEGs evolved as a response to the origin of triploid endosperm, which features the addition of a second maternal genome complement.

second hypothesis is that genetic imprinting has evolved in stepwise fashion along with endosperm ploidy changes [Figure 4, right]. In this case, MEGs may have evolved as a maternal strategy to balance the addition of a paternal genome – and potential for paternal influence on seed development – that resulted in the origin of endosperm. Subsequent addition of a second maternal genome complement with the evolution of triploid endosperm may then be similarly associated with the evolution of PEGs. Studying other members of ANA-grade lineages with diploid endosperm that lack perisperm (such as in Austrobaileyales[36] or taxa with triploid endosperm and perisperm (as can be found in Amaranthaceae[37,38], within the eudicots) would help distinguish between these hypotheses by specifically testing the relationship between maternally-derived storage tissues and the absence of PEGs. Thus, while characterizing parent-of-origin effects on gene expression and epigenetic modifications in Nymphaea endosperm is an important step in understanding the evolution of molecular processes in endosperm, the genetic and developmental diversity across angiosperm seeds deserves further attention.

Acknowledgments

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

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.

Declaration of interests

The authors declare no competing interests.

Figure legends

Figure 1: Overview of seed structure in water lilies and experimental workflows. A) Young seeds of self-fertilized *N. thermarum* (*Nt*) and *N. dimorpha* (*Nd*), as well as reciprocal crosses between the two species. In all seeds, young embryos are surrounded by cellular, diploid endosperm, which in turn is surrounded by a maternal nutrient storage tissue (perisperm). B) For this study, three main datasets were generated to define seed characteristics. Whole-genome DNA methylation was characterized for embryo and endosperm isolated from young and mature seeds, as well as leaf tissue, of *N. thermarum*. Whole-genome DNA methylation was characterized for young endosperm from seeds of reciprocal crosses of *N. thermarum* and *N. dimorpha.* RNA-seq data was also collected for young endosperm from seeds of reciprocal crosses, as well as self-fertilized *N. thermarum* and *N. dimorpha* seeds.

Figure 2: DNA methylation in water lily endosperm and other tissues. A) Principal components analysis of cytosine methylation of embryo and endosperm isolated from young and mature seeds, as well as leaf tissue, of *N. thermarum*. B) Average percent methylation across all cytosines (with >5 reads), in the CG, CHG, and CHH contexts). C) Average number of significantly differentially methylated regions (DMRs) between embryo and endosperm samples (averaged across comparisons of biological replicates). The number of DMRs that overlap a gene, repeat region, both, or neither is shown. D) Ends-analysis showing average cytosine methylation across genes (solid lines) and repeat regions (dashed lines), in the different cytosine contexts, for all *N. thermarum* samples. Sample colors indicated as in panel A. E) Ends-analysis showing the difference in parental allele cytosine methylation (*N. thermaru*m allele methylation - *N. dimorpha* allele methylation) across genes in young endosperm of reciprocal crosses between *N. thermarum* (Nt) and *N. dimorpha* (Nd) For crosses, female in the cross is listed first. F) Ends-analysis showing the difference in parental allele cytosine methylation (*N. thermarum* allele

methylation - *N. dimorpha* allele methylation) across repeat regions in young endosperm of reciprocal crosses between *N. thermarum* (Nt) and *N. dimorpha* (Nd). G) Number of genes (+/- 2kb) that overlap DMRs between maternal and paternal genomes of *N. thermarum* and *N. dimorpha* crosses. Genes were categorized according to whether they overlapped at least one DMR that was more highly methylated in the maternally-inherited genome, only DMR(s) that were higher in the maternally-inherited genome, at least one DMR that was higher in the paternally-inherited genome, or only DMR(s) that were higher in the paternal genome. H) GO-enrichment analysis (biological process terms) of the set of 50 genes (+/- 2kb) that overlapped only with DMRs that were more highly methylated in the maternally-inherited genome.

Figure 3: Genetic imprinting in water lily endosperm. A) Principal component analysis of RNA-seq data from young endosperm samples isolated from self-fertilized *N. thermaurm* (Nt) and *N. dimorpha* (Nd) seeds, and reciprocal crosses of *N. thermaurm* and *N. dimorpha*. B) Plot of the ratio of maternal or paternal allele reads in endosperm of one reciprocal cross comparison. Female in cross listed first. Genes that pass the stringency cut-offs for being called as imprinted are highlighted in red (MEGs, maternally expressed genes) or blue (PEGs, paternally expressed genes). Gray dots indicate genes that are not called as imprinted. C) Upset plot showing the number and consistency of genes called as MEGs or PEGs across comparisons of different replicates. MEGs and PEGs called in at least 75% of comparisons are highlighted in blue (PEGs) or red (MEGs). D) Average expression and average percent maternal expression for all expressed genes (TPM > 1). Genes called as MEGs in at least 75% of replicate comparisons are shown in red, with putative *Arabidopsis thaliana* homology indicated for some MEGs of interest. E) In situ hybridization of putative homolog of GA2OX2. Inset shows magnification of endosperm of sample treated with antisense probe. Scalebars = 200 µm. F) In situ hybridization of putative homolog of SUS3. Inset shows magnification of endosperm of sample treated with antisense

probe. Scalebars = 200 µm. G) GO-enrichment analysis (for biological process terms) of genes consistently called as MEG in the endosperm.

Figure 4: Two hypotheses regarding the evolution of genetic imprinting in endosperm, endosperm ploidy, and nutrient storage strategies in angiosperm seeds. Hypothesis 1 posits that MEGs and PEGs were originally present in endosperm of the last common ancestor of angiosperms, and that PEGs were lost in the water lily lineage in association with the origin of perisperm. Hypothesis 2 posits that MEGs evolved as a maternal response to the addition of a paternal genome complement in a nutrient-mediating tissue, while PEGs evolved as a response to the origin of triploid endosperm, which features the addition of a second maternal genome complement.

Supplementary Figure 1: Genome assemblies for *N. thermarum* and *N. dimorpha*. A) Bandage diagrams for the updated *N. thermarum* genome assembly and the novel *N. dimorpha* genome assembly. B) Alignment of the *N. thermarum* genome mapped against the *N. dimorpha* genome.

Supplementary Figure 2: In situ hybridization positive control for perisperm, and SUS3 expression in seeds of different crosses and self-fertilizations. A) RNA in situ hybridization of putative homolog of a terpene synthase sub-family in *N. thermarum* seeds, showing that detection of gene expression in the perisperm is possible. Results of experiments performed with antisense and sense probes are shown. Scalebars = 200 µm. B) In situ hybridization of putative homolog of *SUS3*, performed in seeds of reciprocal crosses of *N. thermarum* and *N. dimorpha* and in seeds from *N. dimorpha* self-fertilizations. Black arrowheads indicate detection of signal, white arrowheads indicate absence of signal. Scalebars = 200 µm. Results of experiments performed with antisense and sense probes are shown.

References

1. Gehring, M., Huh, J.H., Hsieh, T., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124(3), 495-506. 10.1016/j.cell.2005.12.034.

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

3. Moreno-Romero, J., Del Toro-De León, G., Yadav, V.K., Santos-González, J., and Köhler, C. (2019). Epigenetic signatures associated with imprinted paternally expressed genes in the Arabidopsis endosperm. Genom Biol 20(1), 41. doi: 10.1186/s13059-019-1652-0.

4. Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., Kawashima, T., Voigt, P., Boavida, L., Becker, J., Higashiyama, T., Martienssen, R., and Berger, F. (2020). Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. Nat Cell Biol 22(6), 621-629. 10.1038/s41556-020-0515-y.

5. Picard, C.L., and Gehring, M. (2020). Identification and comparison of imprinted genes across plant species. In Plant Epigenetics and Epigenomics: Methods and Protocols, Methods in Molecular Biology volume 2093, Spillane, C., and McKeown, P., eds. (Springer Science+Business Media, LLC, part of Springer Nature 2020), pp. 173-201. 10.1007/978-1-0716-0179-2_13.

6. Povilus, R.A., and Gehring, M. (2022). Maternal-filial transfer structures in endosperm: a nexus of nutritional dynamics and seed development. Curr Opin Plant Biol 65: 102121. 10.1016/j.pbi.2021.102121.

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

8. Williams, J.H., and Friedman, W.E. (2022). Identification of diploid endosperm in an early angiosperm lineage. Nature 415: 522-526. 10.1038/415522a.

9. Friedman W.E. (2006). Embryological evidence for developmental lability during early angiosperm evolution. Nature 441, 337-340. 10.1038/nature04690.

10. Friedman W.E. (2008). Hydatellaceae are water lilies with gymnospermous tendencies. Nature 453, 94-97. 10.1038/nature06733.

11. Rudall, P.J., Remizowa, M.V., Beer, A.S., Bradshaw, E., Stevenson, D.W., Macfarlane, T.D., Tuckett, R.E., Yadav, S.R., and Sokoloff, D.D. (2008). Comparative ovule and megagametophyte development in Hydatellaceae and water lilies reveal a mosaic of features among the earliest angiosperms. Ann Bot 101(7), 941-956. 10.1093/aob/mcn032.

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

13. Povilus, R.A., Diggle P.K., and Friedman W.E. (2018). Evidence for parent-of-origin effects and interparental conflict in seeds of an ancient flowering plant lineage. Proc Royal Soc B 285(1872): 20172491

14. Povilus, R.A., DaCosta, J.M., Grassa, C., Satyaki, P.R.V., Moeglein, M. Jaenisch, J., Xi, Z., Mathews, S., Gehring, M., Davis, C.C., and Friedman, W.E. (2020). Water lily (Nymphaea thermarum) genome reveals viaraible genomic signatures of ancient vascular cambium losses. PNAS 117(15), 8649-8656. doi.org/10.1073/pnas.1922873117

15. Lauria, M., Rupe, M., Guo, M., Kranz, E., Pirona, R., Viotti, A., and Lund, G. (2004). Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant cell 16(2), 510-522. 10.1105/tpc.017780.

16. Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324(5933), 1447-1451. 10.1126/science.1171609.

17. Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell 23(12), 4221-4233. 10.1105/tpc.111.092668.

18. Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., and Zilberman, D. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337(6100), 1360-1364. 10.1126/science.1224839.

19. Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.J., Sharma, R., Ronald, P.C., Fischer, R.L., and Zilberman D. (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci U S A 110(19), 7934-7939. 10.1073/pnas.1306164110.

20. Zhang, M., Zie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., Li, H., Yang, W., Zhao, H., Wang, G., Chen, Z., Sun, S., Hauch, A., Jin, W., and Lai, J. (2014). Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Res 24(1) 167-176. 10.1101/gr.155879.113.

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

22. Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S.A., Springer, N.M., and Schmitz, R.J. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biol 17, 194. 10.1186/s13059-016-1059-0.

23. Hsieh, T., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science 324(5933), 1451-1454. 10.1126/science.1172417.

24. Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife, 3:3:e03198. 10.7554/eLife.03198.

25. Klosinska, M., Picard, C.L., and Gehring, M. (2016). Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nature Plants 2, 16145. 10.1038/nplants.2016.145.

26. Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). Regulation of Seed Size by Hypomethylation of Maternal and Paternal Genomes. Plant Physiol 142(3): 1160-1168.

27. Hsieh, T., Shin, J., Uzawa, R., and Fischer, R.L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. PNAS 108(5), 1755-1762. 10.1073/pnas.1019273108.

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

29. Xin, M., Yang, R., Li, G., Chen, H., Laurie, J., Ma, C., Wang, D., Yao, Y., Larkins, B.A., Sun, Q., Yadegari, R., Xiangfeng, W., and Ni, Z. (2013). Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. Plant Cell 25(9), 3212-3227. PMID: 24058158

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

31. Birchler, J.A. (1993). Dosage analysis of maize endosperm development. Annu Rev Genet 27, 181-204. 10.1146/annurev.ge.27.120193.001145.

32. Scott, R.J., Spielman, M., Bailey, J., and Dickinson H.G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125(17), 3329-3341. 10.1242/dev.125.17.3329.

33. Lersten N.R. (2004). Endosperm. In Flowering plant embryology. Lersten, N.R., ed. (Blackwell Publishing, Hoboken, NJ), pp. 161–163.

34. Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R., and Wolf, J.B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. Heredity 113, 119-128. 10.1038/hdy.2014.29.

35. Friedman, W.E., Bachelier, J.B., and Hormaza, J.I. (2012). Embryology in Trithuria submersa (Hydatellaceae) and relationships between embryo, endosperm, and perisperm in early-diverging flowering plants. Am J Bot 99(6), 1083-1095. 10.3732/ajb.1200066.

36. Losada, J.M., Bachelier, J.B., and Friedman, W.E. (2017). Prolonged embryogenesis in Austrobaileya scandens (Austrobaileyaceae): its ecological and evolutionary significance. New Phyol 215(2), 851-864. 10.1111/nph.14621.

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

38. López-Fernández, M.P., and Maldonado, S. (2013). Programmed cell death during quinoa perisperm development. J Exp Bot 64(11), 3313-3325. 10.1093/jxb/ert170.

Materials and Methods

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mary Gehring (mgehring@wi.mit.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

1) Data generated as part of this study are being submitted to NCBI under BioProjects PRJNA1085993, PRJNA1086866, PRJNA1086863, PRJNA1085992, and PRJNA1087317, including raw sequence data deposited in SRA, analyzed genomic, epigenetic, and transcriptomic data deposited in GEO, and genome assemblies an annotations deposited in NCBI WGS. Microscopy data reported in this paper will be shared by the lead contact upon request.

2) This paper does not report original code.

3) Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Plant growth and sample collection

Seeds of *N. thermarum* and *N. dimorpha* were sourced from the Arnold Arboretum of Harvard University (Boston, MA, USA) and grown at the Whitehead Institute for Biomedical Research (Cambridge, MA, USA). Plants were grown in individual pots, potted in top soil, and submerged in pH-balanced tap water (pH roughly between 6.5 and 7). *N. thermarum* plants were submerged so that the soil level was 2-5 cm below the water level, while *N. dimorpha* plants were submerged so that the soil level was 15-30 cm below the water level. Plants were grown under a mix of natural and artificial lighting (full spectrum LED), with seasonal lighting supplemented to provide 12 h of light a day. To encourage prolific flowering and fruit set, every 2-4 weeks approximately one half of a CrystalClear Thrive 10-14-8 Aquatic Plant Fertilizer tablet was added to each pot.

For collection of seeds from self-fertilizations, first day of anthesis (time of female receptivity and fertilization[1]) was defined by presence of stigmatic fluid. Fruits were collected at 10 days after anthesis (DAA) (for young seeds) and 30 days after anthesis (for mature seeds) and seeds were immediately removed and dissected with fine forceps in dissection buffer (0.3M sorbitol, 0.3mM MES) to separate the embryo and endosperm; embryo and endosperms were placed in separate droplets of fresh dissection buffer. For each sample, tissue from 10-20 seeds from a single fruit were pooled. Embryos and endosperms were washed by moving to a fresh droplet of dissection buffer via forceps or micropipette; samples were then moved to a third and final droplet of fresh dissection buffer before being moved to a microcentrifuge tube. Excess dissection buffer was drawn off and samples were frozen in liquid nitrogen. Seeds from cross pollinations were collected 9-10 DAA in a similar manner (controlled pollinations were performed as previously described[1,2]). Leaf tissue was collected by selecting young leaves, washing leaf surfaces with sterile water, and freezing in liquid nitrogen after placement in microcentrifuge tubes.

Whole genome sequencing, assembly, and annotation

For long-read and short-read DNA sequencing, high molecular weight DNA was extracted from > 1 g young leaf samples from a single individual plant using a modified CTAB-based protocol. Briefly: frozen leaf tissue was ground in a mortar and pestle and added to a CTAB buffer with 0.25 mM PVP40 (Polyvinylpyrrolidone), 0.005% (v/v) 2-Mercaptoethanol, and 1.3% (v/v) sodium dodecyl sulfate. Samples were incubated for 30 minutes at 65 °C, centrifuged to precipitate debris, and supernatant was removed. Samples were precipitated with addition of potassium acetate (to 0.9M), incubated on ice for 30 minutes and then centrifuged at 4 °C for 20 min. Supernatant was transferred to a new tube and an equal volume of isopropanol was added; samples were incubated on ice for 5 minutes, centrifuged at 4 °C for 20 minutes, and supernatant was removed and discarded. The air-dried DNA pellet was re-suspended in Tris-EDTA buffer; RNAaseA was added and samples incubated for 20 minutes at 37 °C. Samples were transferred to PhaseLock Heavy Gel tubes, and phase-extracted 3 times with 25:24:1, phenol : chloroform : isoamyl alcohol and then 3 times with chloroform (the aqueous phase was transferred to a new PhaseLock gel tube for each step). Samples were then precipitated with sodium acetate (to 0.3 M) and an equal volume of isopropanol; samples were chilled on ice for 5 min and centrifuged at 4 °C for 1 minute. DNA pellets were washed three times with 70% ethanol, air dried and re-suspended in 0.1M Tris-EDTA buffer. Samples were prepared for PacBio sequencing (PacBio Sequel SMRTcell, 20h, v3 chemistry) and were sequenced at the MIT BioMicroCenter. *Nymphaea thermarum* was sequenced in both LR (long-read) and HiFi (high-fidelity) modes; *N. dimorpha* was sequenced in LR mode. *N. dimorpha* short-read data was obtained from the same sample, using 1 lane of an Illumina HiSeq2000 flow-cell (40 bp, paired-end reads) at the MIT BioMicroCenter. Short-read genomic data for *N. thermarum* was downloaded from BioProject PRJNA508901.

Genome assembly for *N. thermarum* and *N. dimorpha* was performed separately using long-reads as input for Canu (version 2.1.1)[3]; short-reads were used to polish the resulting assemblies using POLCA (from MaSuRCA version 3.4.2)[4]. Genome assemblies were visualized with Bandage[5]. The resulting original genome assemblies were separately annotated with MAKER (version 2.31.10)[7] for both species, using an iterative approach to train AUGUSTUS (version 3.3.3)[8] and SNAP (version 2006.07.28-1)[9] gene-model predictors; initial input for all annotation pipelines included the set of transcript and protein sequences from the published *N. thermarum* genome assembly/annotation, the set of all protein sequences from Nymphaeaceae available on NCBI, protein sequences from the *N. colorata* genome assembly and annotation, all basal Magnoliophyta protein sequences on Uniprot, Amborella protein sequences, and TAIR 10 protein sequences. Three rounds of annotation and gene model predictor training were performed for annotation of both species. Repeat identification and masking was performed with RepeatMasker (version 4.0.5)[10] using spermatophyta as the specified query clade and the Embryophyta repeat database.

To create genome assemblies of *N. thermarum* and *N. dimorpha* that shared positional homology, the *N. thermarum* contigs were mapped to *N. dimorpha* contigs using minimap2[6] (the *N. dimorpha* assembly was used as the reference as it had the fewest contigs; genome alignment visualized using D-Genies[11]). The resulting reorganized genomes for both species were separately re-annotated as described above, and the final annotations were resolved using MAKER to the reorganized *N. thermarum* genome, with positional homology used to apply the annotation to the *N. dimorpha* genome assembly. For each species, assembled transcripts were then generated using the resolved annotation and genome assembly of each species, resulting in a set of homologous *N. thermarum* transcripts and a set of *N. dimorpha* transcripts.

For *N. thermarum* and *N. dimorpha* transcripts, homology to *Arabidopsis thaliana* was determined by blastx searching the *N. thermarum* transcripts against the TAIR11 protein set (e-value

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

DNA methylation-sensitive sequencing and analysis

For methylation-sensitive sequencing, DNA was extracted from leaf and embryo samples using Ampure beads; endosperm DNA was extracted using the QIAamp micro kit (Qiagen, cat# 56304). DNA from leaf samples were additionally purified using a Bioneer PCR cleanup kit. For bisulfite sequencing (self-fertilized seed tissue samples and leaves), bisulfite conversion was performed using an Invitrogen MethylCode kit and library preparation was performed using a Zymo Pico MethylSeq kit with additional indexing primers. Sequencing was performed at the Whitehead Genome Technology Core; two additional rounds of PCR were performed on the libraries to remove single stranded DNA and fragments greater than 200 bp were selected using a PippenHT. Libraries were pooled and sequenced across three HiSeq flow cell lanes (100 bp, paired-end reads).

For enzymatic methyl conversion sequencing and library preparation (for hybrid seed tissues samples), an NEBNext Enzymatic Methyl-seq kit was used; one additional AMPure bead clean-up was performed on libraries to remove primer dimer. Sequencing was performed at the Whitehead Genome Technology Core. Libraries were pooled and sequenced across 2 lanes of a NovaSeq SP flowcell (50 bp, paired-end reads).

Bisulfite conversion and enzymatic-methyl sequencing conversion rate was assessed prior to sequencing by performing PCR on converted and unconverted samples using conversion-sensitive primers (forward = 5’ GAGAAGAAAAGAAGGCGTTAAG 3’; reverse = 5’ ATTTCCCTAATTCTTCCCCTTC 3’;

anticipated fragment length = ~275 bp); resulting products were cloned via Invitrogen TOPO TA into pCR4-TOPO vector and sequenced. Conversion rates were calculated using CyMATE[12]. For Bisulfite-converted samples conversation rate of the converted leaf sample was 99.57%, for enzymatic-methyl sequencing converted sample conversion rate averaged 99.85%.

For self-fertilized *N. thermarum* seed tissue and leaf samples, reads from bisulfite-converted samples were trimmed of adaptors and low-quality regions with TrimGalore (version 0.6.7). Each 100 bp read was split into 2 40-base-pair reads to improve mapping efficiency, and mapped as single-end reads to the original *N. thermarum* assembly[13], using BISMARK (version 0.22.3)[14] with deduplication. Analysis of average DNA methylation 5', 3' and interior of features was performed using previously developed custom pipelines[15]. Differentially methylated regions (DMRs) between embryo, endosperm, and leaf samples were identified in the CG, CHG, and CHH contexts using a previously developed pipeline[15]. Briefly, DMRs were defined as 300-bp windows for which 3 or more cytosines with a coverage of 5 or more reads that had a methylation difference of 35% or greater between samples, with a Fisher's exact test with Benjamini-Hochberg correction p-value cutoff of 0.01 to determine significance. DMRs were called between all combinations of biological replicates. For total number of DMRs between embryo and endosperms of a particular developmental stage, the number of DMRs was averaged across all replicate comparisons. Genes and repeat regions were identified as associated with a DMR if the gene or repeat region had a DMR within the annotated region or 2 kb 5' or 3'; a gene or repeat was required to be associated with a DMR in all biological replicates in order to be reported.

Hybrid endosperm samples were analyzed as above, except that reads from bisulfite-converted samples were first mapped to a concatenation of the originally produced *N. thermarum* and *N. dimorpha* assemblies and annotations, using Bismark. The reads that uniquely mapped to either species' genome were sorted into separate sets of *N. thermarum* or *N. dimorpha* reads. *N. thermarum* reads were mapped to the reorganized *N. thermarum* genome annotation and methylation data was extracted

with Bismark. *N. dimorpha* reads were mapped to the reorganized *N. dimorpha* genome annotation and methylation data was extracted with Bismark. Analysis of average DNA methylation features and differentially methylated regions (DMRs) between maternal and paternal genomes were identified as described above.

GO enrichment analysis was performed using ShinyGO 0.77[16]. Putative Arabidopsis homologs of all transcripts were used, and the set of putative Arabidopsis homologs of all transcripts expressed during seed development (TPM > 1)[17] was used as the background set.

RNA sequencing and data analysis

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

For initial analysis of gene expression, reads from all samples were mapped to the concatenated genomes of the originally produced *N. thermarum* and *N. dimorpha* assemblies and annotations, using STAR (version 2.4.1) using default settings, after using TrimGalore to trim reads of adaptors and low-quality regions. Principal components analysis was performed in R using the DESeq2 package[18], including using the varianceStabilizingTransformation() function. For identification of imprinted genes in hybrid samples, the reads that uniquely mapped to either species' genome were sorted into separate sets of *N. thermarum* or *N. dimorpha* reads, and were used for subsequent analysis. *N. thermarum* reads were mapped to the reorganized *N. thermarum* genome annotation via Kallisto (version 0.46.1), using the set of *N. thermarum* transcripts generated from the reorganized genome assembly and annotation. *N. dimorpha* reads were mapped to the reorganized *N. dimorpha* genome annotation via Kallisto, using the

set of *N. dimorpha* transcripts generated from the reorganized genome assembly and annotation. Resulting allele-specific count tables for each transcript were used for calling genetic imprinting. Genetically imprinted genes were called as previously described[15], using a pairwise comparison of all possible combinations of each hybrid cross sample (Nt x Nd, vs. Nd x Nt). While mapping reads, a slight maternal expression bias was noted for both cross directions [Supplemental Table 1]. Therefore, when calling imprinted genes, the expected maternal : paternal expression ratio was adjusted from 1 (the anticipated null ratio for diploid endosperm) to the maximum observed maternal expression bias of 1 .32; to increase stringency minimum allelic-specific read count was set to 50, a minimum imprinting factor was set to 2, and cis-effect factor was set to 15. MEGs were required to have a minimum of 70% maternal allele reads and PEGs were required to a maximum of 30% maternal allele reads in both cross directions. In order for a gene to be considered as consistently imprinted, it had to be called as imprinted in at least 75% of pairwise comparisons.

GO enrichment analysis was performed using ShinyGO 0.77[16], using default settings. Putative Arabidopsis homologs of all transcripts in the test set were used, and the set of putative Arabidopsis homologs of all transcripts expressed during seed development (TPM > 1)[17] was used as the background set.

In situ hybridizations and histology

In situ hybridizations were performed as previously described[19]. Probe primer sequences were as follows: GA2OX2, forward = 5' ATCTAACCCAGCACTCGACC 3' ; GA2OX2, reverse = 5 'CTGTGGTGAAGGTGTCAAGC 3'; SUS3, forward = 5' GCTACTTTATTGGCACATAAGCTG 3'; SUS3, reverse = 5' GAAATCGCATCCCAGTGTTG 3'; Terpene synthase gene subfamily, forward = 5' CTTGTGAGATGTGAGCCTAGG 3'; Terpene synthase gene subfamily, reverse = 5' TGCCACCTTGCCTCAAC

3'. Anticipated probe lengths were as follows: GA2OX2 = ~900 bp; SUS3 = ~1000 bp; terpene synthase gene subfamily = ~1000bp. Preparation of seed samples for histological analysis was performed as previously described for seeds of *A. thaliana*[19] and stained with toluidine blue, with adaptations of incubation times as necessary. All samples were sectioned on a Leica RM 2065 rotary microtome at a thickness of 8 µm and imaged using a Zeiss Axio Imager M2. Image tiling, color and brightness/contrast adjustments and Smart Sharpen were applied to whole images, with particular attention to having an even white-balance across different images (Adobe Photoshop).

References, Materials and Methods

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

2. Povilus, R.A., Diggle P.K., and Friedman W.E. (2018). Evidence for parent-of-origin effects and interparental conflict in seeds of an ancient flowering plant lineage. Proc Royal Soc B 285(1872): 20172491

3. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27, 722-736. 10.1101/gr.215087.116

4. Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics 29(21), 2669-77. 10.1093/bioinformatics/btt476.

5. Wick R.R., Schultz M.B., Zobel J. & Holt K.E. (2015). Bandage: interactive visualisation of de novo genome assemblies. Bioinformatics 31(20), 3350-3352.

6. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100. doi:10.1093/bioinformatics/bty191.

7. Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. Curr Protoc Bioinformatics 48, 4.11.1–4.11.39. 10.1002/0471250953.bi0411s48.

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

9. Korf, I. Gene finding in novel genomes. (2004). BMC Bioinform 5, 59. 10.1186/1471-2105-5-59.

10. Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics, Chapter 4. 1002/0471250953.bi0410s05.

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

12. Hetzl, J., Foerster, A.M., Raidl, G., and Scheid, O.M. (2007). CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. Plant J 51(3), 526-36. 10.1111/j.1365-313X.2007.03152.x

13. Povilus, R.A., DaCosta, J.M., Grassa, C., Satyaki, P.R.V., Moeglein, M., Jaenisch, J., Xi, Z., Mathews, S., Gehring, M., Davis, C.C., and Friedman, W.E. (2020). Water lily (Nymphaea thermarum) genome reveals variable genomic signatures of ancient vascular cambium losses. PNAS 117(5), 8649-8656. 10.1073/pnas.1922873117.

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

15. Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife, 3:3:e03198. 10.7554/eLife.03198.

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

17. Povilus, R.A., and Friedman, W.E. (2022). Transcriptomes across fertilization and seed development in the water lily Nymphaea thermarumL evidences for epigenetic patterning during reproduciton. Plant Reprod 35(5), 161-178. 10.1007/s00497-022-00438-3.

18. Love M.I., Huber W., Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. doi:10.1186/s13059-014-0550-8.

19. Pignatta, D., Novitzky, K., Satyaki, P.R.V., and Gehring, M. (2018). A variably imprinted epiallele impacts seed development. PLoS Genet 14(11), e1007469. 10.1371/journal.pgen.1007469.