

1 **Transcriptomes across fertilization and seed development in the water lily *Nymphaea thermarum***  
2 **(Nymphaeales) reveal dynamic expression of DNA and histone methylation modifiers**

3

4 Rebecca A. Povilus<sup>1</sup> and William E. Friedman<sup>1,2,\*</sup>

5

6 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

7 2) Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, MA 02131, USA

8 \*corresponding author: William E. Friedman (Arnold Arboretum of Harvard University, 1300 Centre

9 Street, Boston, MA 02131, USA; 617.384.7744; [ned@oeb.harvard.edu](mailto:ned@oeb.harvard.edu))

10

11 **Current addresses:**

12 R.A.P.: Whitehead Institute for Biomedical Research, Cambridge MA 02142

13

14 **Abstract**

15 Studies of gene expression during seed development have been performed for a growing  
16 collection of species from a phylogenetically broad sampling of flowering plants (angiosperms).  
17 However, attention has mostly been focused on crop species or a small number of ‘model’ systems.  
18 Information on gene expression during seed development is minimal for those angiosperm lineages  
19 whose origins predate the divergence of monocots and eudicots. In order to provide a new perspective  
20 on the early evolution of seed development in flowering plants, we sequenced transcriptomes of whole  
21 ovules and seeds from three key stages of reproductive development in the waterlily *Nymphaea*  
22 *thermarum*, an experimentally-tractable member of the Nymphaeales. We first explore general patterns  
23 of gene expression, beginning with mature ovules and continuing through fertilization into early- and  
24 mid-seed development. We then examine the expression of genes associated with DNA and histone

25 methylation – processes known to be essential for development in distantly-related and structurally-  
26 divergent monocots and eudicots. Around 60% of transcripts putatively homologous to DNA and histone  
27 methylation modifiers are differentially expressed during seed development in *N. thermarum*,  
28 suggesting that the importance of dynamic epigenetic patterning during seed development dates to the  
29 earliest phases of angiosperm evolution. However, genes involved in establishing, maintaining, and  
30 removing methylation marks associated with genetic imprinting show a mix of conserved and unique  
31 expression patterns between *N. thermarum* and other flowering plants. Our data suggests that the  
32 regulation of imprinting has likely changed throughout angiosperm evolution, and furthermore identifies  
33 genes that merit further characterization in any angiosperm system.

34

35 Keywords:

36 Seed development, Nymphaea, seed transcriptomes, epigenetics

37

## 38 **Introduction**

39 Fertilization and seed development are critical parts of the plant life cycle that involve extensive  
40 transcriptional reprogramming. Seed development in flowering plants (angiosperms) is of particular  
41 interest, as it uniquely involves two separate fertilization events that produce two distinct offspring.  
42 Double fertilization in angiosperms occurs when a pollen tube reaches a mature ovule and delivers two  
43 sperm cells into the female gametophyte. Each sperm cell fuses with one of the two female gametes,  
44 the egg cell and the central cell, to produce (respectively) the embryo and the embryo-nourishing  
45 endosperm. While endosperm does not necessarily persist past seed germination, it surrounds the  
46 embryo throughout seed development and is a crucial mediator of the relationship between an embryo  
47 and its maternal sporophyte.

48           Given that seeds, and specifically endosperm, are the cornerstone of human diets, there has  
49    been much effort to understand the dynamic transcriptional landscape of seed development in a variety  
50    of economically important plants (maize (Li 2014)(Chen 2014), rice (Gao 2013)(Xu 2012), soybean (Jones  
51    2013), peanut (Zhang 2012), camelina (Nguyen 2013), *Brassica* (Gao 2014, Ziegler 2019)) or model  
52    systems (*Arabidopsis* (Girke 2000, Belmonte 2013)). Yet little information exists from within lineages  
53    whose origins predate the divergence of monocots and eudicots, hindering an understanding of the  
54    evolution of developmental processes that contribute to seed development. While this is related to the  
55    difficulty of working with the vast majority of species within these lineages (e.g., *Amborella*,  
56    Nymphaeales, Austrobaileyales, Chloranthales, Ceratophyllales, magnoliids, which are typically long-  
57    lived trees, shrubs, lianas, or aquatic plants), a genetically and experimentally tractable species from  
58    within one of the most-early diverging lineages has been identified (Povilus 2015). *Nymphaea*  
59    *thermarum* (Nymphaeales) is a minute waterlily with a relatively short generation time and a draft  
60    genome assembly and annotation (Povilus 2020) – as such, it is poised to help illuminate questions  
61    about the evolution of flowering plant reproduction.

62           A common thread has emerged from studies of seed development in a wide variety of  
63    angiosperms: epigenetic patterning and imprinting is important for seed development, and particularly  
64    for endosperm development (Haig 1991)(Haig 2013)(Gehring 2017)(Satyaki 2017). Imprinting is a  
65    phenomenon that results in alleles with identical nucleotide sequences that have different expression  
66    patterns, depending on which parent the allele was inherited from (a “parent-of-origin” effect). In  
67    flowering plants, imprinting is largely understood to occur via the establishment of DNA and histone  
68    methylation patterns during gamete and seed development. Because methylation of DNA or histones  
69    can affect how a locus is expressed, different epigenetic patterns established during the development of  
70    male and female gametes can mean that certain loci are expressed preferentially from the maternally-  
71    or paternally-inherited copy of the allele (Zilberman 2006). Imprinting has been noted to be particularly

72 important for the ability of endosperm to function as a nutritional mediator between the embryo and  
73 maternal sporophyte (Haig 1991)(Gehring 2017). However, some of the mechanisms that control DNA or  
74 histone methylation patterns appear to differ between monocots and eudicots (Furihata  
75 2016)(Nalaamilli 2013)(Köhler 2012), leading to the question of when DNA and histone methylation, and  
76 their role in imprinting, became important aspects of seed development in flowering plants.

77 DNA methylation during reproductive development is perhaps best understood in *Arabidopsis*  
78 *and rice*, and involves the coordination of DNA methyltransferases and demethylases, some of which  
79 operate as part of a RNA-dependent DNA methylation mechanism (Satyaki 2017). Members of the DNA  
80 METHYLTRANSFERASE family (MET) establish and maintain CG methylation, while CHROMOMETHYLASE  
81 proteins (CMT) establish or maintain CHG or CHH methylation. METs and CMTs are known to be  
82 expressed both in developing female gametophytes of *Arabidopsis*, as well as in offspring tissues after  
83 fertilization (Jullien 2012)(Köhler 2012). DEMETER (DME) is a DNA glycosylase that removes methylation  
84 established by MET1 and is active during female gametophyte development. DME activity determines  
85 expression of some of the components of the POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) (Hsieh 2011).  
86 The PRC2 complex participates in histone H3K27 methylation, and in doing so regulates the expression  
87 of several genes known to be important for seed development (Hsieh 2011). PRC2 is comprised of  
88 MEDEA (MEA), FERTILIZATION-INDEPENDENT SEED2 (FIS2), FERTILIZATION-INDEPENDENT ENDOSPERM  
89 (FIE), and MULTICOPY SUPPRESSOR OF IRA1 (MSI1), and is active in both the central cell of the female  
90 gametophyte and in endosperm (Furihata 2016). Together, METs, CMTs, DME, and PRC2 regulate  
91 methylation patterns necessary for imprinting of parent-or-origin specific expression patterns.  
92 Methylation-modifying processes not necessarily associated with imprinting, such as RNA-directed DNA  
93 methylation (RdDM), are also active during reproductive development and have been tied to repression  
94 of transposon activity in the egg cell, embryo, or other tissues (Köhler 2012)(Gehring 2017)(Ingouff  
95 2017)(Satyaki 2017).

96           In order to shed light on whether imprinting via DNA and histone methylation could be  
97 responsible for recently discovered parent-of-origin effects on endosperm and embryo development in  
98 *N. thermarum* (Povilus 2018), we examined the expression patterns of genes involved in DNA and  
99 histone methylation, and in particular those known to be important for imprinting. By obtaining libraries  
100 of gene expression during important stages of seed development in the water lily *Nymphaea*  
101 *thermarum*, we provide the first such dataset from within any early-diverging angiosperm lineage. The  
102 three stages sampled represent unique suites of developmental processes (Figure 1)(Povilus 2015). The  
103 first stage of 0 DAA (days after anthesis) consists of whole, unfertilized ovules. The second stage is whole  
104 seeds at 7 DAA, when the endosperm is expanding but the embryo is relatively quiescent. Nutrients are  
105 actively being acquired by, and stored in, a tissue called the perisperm (a maternal sporophyte tissue  
106 derived from the nucellus). The third stage sampled is whole seeds at 15 DAA, when endosperm  
107 expansion and differentiation has nearly been completed. The embryo begins to undergo significant  
108 growth and morphogenesis, displacing space occupied by degenerating endosperm cells. The perisperm  
109 continues to acquire and store nutrient reserves. Thus, while seed components were not spatially  
110 dissected from each other, the selected time points capture important developmental landmarks for the  
111 embryo and endosperm.

112           We compare the expression profiles of genes that regulate DNA and histone methylation in *N.*  
113 *thermarum* with those of their homologs in monocots and *Arabidopsis*. In doing so, we identify  
114 processes that have likely been involved in seed development since the earliest stages of angiosperm  
115 evolution. We also find that that all of the molecular processes known to be involved in imprinting are  
116 indeed expressed before and/or after fertilization – suggesting that imprinting could occur in *N.*  
117 *thermarum*.

118

119 **Materials and methods**

## 120 **Plant Material and Sequencing**

121 The *Nymphaea thermarum* plants sampled for this study were grown in a greenhouse at the  
122 Arnold Arboretum of Harvard University (Boston, MA, United States) according to previously established  
123 protocols (Fischer 2010). Flowers were allowed to self-fertilize. All samples were collected between 10  
124 and 11 am on their respective collection days. Ovules and seeds were quickly dissected from  
125 surrounding carpel or fruit tissue, weighed, then immediately frozen in liquid nitrogen, and stored at -80  
126 C. For 0 DAA, each biological replicate represents material from 3-4 individual flowers from different  
127 plants. For 7 and 15 DAA, each biological replicate includes material from a single fruit.

128 RNA extractions were performed with a modified protocol, originally for use with maize kernels  
129 (Wang 2012)(Supplementary Materials and Methods 3). RNA seq libraries were prepared by the  
130 Whitehead Genome Sequencing Core, according to the manufacturer protocols of the Illumina Standard  
131 mRNA-seq library preparation kit (Illumina) using poly A selection, and were sequenced at the Baur Core  
132 of Harvard University to generate 125-bp, paired-end reads on a Illumina HiSeq Platform. All 12 libraries  
133 were multiplexed and sequenced on 3 lanes.

134

## 135 **Read mapping and differential expression analysis**

136 For each sample, kallisto (Bray 2016) was used to pseudo-align reads to the *Nymphaea*  
137 *thermarum* genome (Povilus 2020) and quantify transcript abundances. 100 mapping bootstraps were  
138 performed, using default parameters for paired-end reads. Kallisto reports both estimated number of  
139 transcript reads per sample (EST), as well as transcript abundance per million reads (TPM, normalized  
140 for transcript length and number of reads per sample)(Supplemental Dataset 1). Transcripts that are  
141 differentially expressed (DE) between time points were identified using sleuth (Pimentel  
142 2017)(Supplemental Dataset 2). The primary DE analysis modeled the effect of time, for all time points,  
143 on transcript abundance. Subsequent DE analysis was conducted for pair-wise comparisons between

144 time points (for this analysis, multiple testing was accounted for by requiring transcripts to be  
145 significantly DE according to the primary DE analysis). Sleuth incorporates information from bootstraps  
146 performed by kallisto to estimate the inferential variance of each transcript; the adjusted variances were  
147 used to determine differential expression for each transcript. Transcripts were considered differentially  
148 expressed if time point (DAA) was a significant factor for transcript abundance, according to both a  
149 conservative likelihood ratio test and a Wald test (multiple-testing corrected p-value < 0.01).

150

### 151 **PCA and Clustering of Biological Replicates**

152 Analyses were performed with R (version 3.4.0, (R Core Team 2017)). To assess similarity of  
153 biological replicates, EST counts for each transcript that was differentially expressed were centered and  
154 scaled (according to transcript means across all samples), using the scale function. Dimensional  
155 expression data was reduced to two dimensions by PCA using the prcomp function. K-means clustering  
156 within PCA space was performed by the kmeans function, with cluster number set to 3 (number chosen  
157 to reflect the number of sampled time points).

158

### 159 **Expression pattern cluster definition and analysis**

160 Analyses were performed with R (version 3.4.0, (R Core Team 2017)). K-means clustering of gene  
161 expression patterns was performed with sample TPM values, using the kmeans function. The cluster  
162 number was set to 9, as the use of higher cluster numbers failed to identify additional unique expression  
163 patterns. Only transcripts that were differentially expressed were used for k-means clustering. The z-  
164 score was calculated for each gene per sample, using the scale function. Only genes whose expression  
165 patterns correlated with the average profile of each cluster (Pearson correlation > 0.9) were used in  
166 further analysis (Supplementary Dataset 3).

167 GO (molecular function) enrichment for each expression pattern cluster was performed with  
168 agriGO (Tian 2017), using *Arabidopsis thaliana* TAIR 10 annotation as the background, with  
169 hypergeometric or chi-squared tests (chi-squared was only performed if the query list had relatively few  
170 intersections with the reference list, and is noted separately), Yekutieli (FDR under dependency)  
171 multiple corrections testing adjustment, and significance level = 0.1. Putative *A. thaliana* homologs of *N.*  
172 *thermarum* transcripts were identified via BLASTX for each *N. thermanum* transcript against a database  
173 of all TAIR 10 *Arabidopsis* transcript amino acid sequences (downloaded from Phytozome (Goodstein  
174 2012)), using the hit with the lowest e-value as the putative homolog match (e-value cutoff =  $1e-15$ ).

175

#### 176 **Identification and analysis of transcription factors**

177 Putative transcription factors (and their respective family type) were identified from the  
178 *Nymphaea thermanum* genome using the 'Prediction' tool available from the Plant Transcription Factor  
179 Database v4.0 (Jin 2017). Enrichment analysis for TF families among the set of DE TFs in each expression  
180 cluster was performed in R with Fisher's exact test; adjusted p-values (FDR) <0.1 and <0.05 are noted.

181

#### 182 **Identification of genes involved in histone and DNA methylation**

183 To comprehensively identify putative homologs of genes known to be involved in regulation of  
184 DNA and chromatin methylation during seed development in other angiosperms, we estimated gene-  
185 family phylogenies for gene families of particular interest (CMT, MET, DME, components of the PCR2  
186 complex). For each gene family, amino acid sequences for *Arabidopsis* members were aligned and used  
187 as the input for HMMER searches (e-value cutoff =  $1e-15$ ) (Eddy 2011) to identify putative homologs  
188 from genomes of *Physcomitrella patens* (*Physcomitrella patens* v3.3, DOE-JGI,  
189 <http://phytozome.jgi.doe.gov/>), *Amborella trichopoda* (Amborella Genome Project 2013), *Nymphaea*  
190 *thermarum* (Povilus 2020), *Aquilegia coerulea* (*Aquilegia coerulea* Genome Sequencing Project,

191 <http://phytozome.jgi.doe.gov/>), *Oryza sativa* (Ouyang 2007), *Zea mays* (Hirsch 2016), *Arabidopsis*  
192 *thaliana* (Lamesch 2012), and *Solanum lycopersicum* (Tomato Genome Consortium 2012). The latest  
193 versions of all annotated genome datasets, except for *N. thermaurm*, were downloaded from  
194 Phytozome (Goodstein 2012). Putative homologs were also identified from within the de-novo  
195 assembled, immature-ovule and non-seed transcriptomes of *N. thermarum* (which includes tissues from  
196 roots, floral buds, leaves, and pre-meiotic ovules)(Povilus 2020). The amino acid sequences for the set of  
197 all putative homologs for a gene family were aligned with MUSCLE (Edgar 2004), alignments were  
198 manually trimmed to represent highly conserved regions, and phylogenetic tree estimation and  
199 bootstrapping (n=100) was performed with RAxML under the PROTGAMMAGTR amino-acid substitution  
200 model (Stamatakis 2014). During further discussion we took a conservative approach, using a relatively  
201 broad definition as to which members were included in particular gene sub-families of interest.

202 All *Arabidopsis* genes annotated as being involved with either DNA methylation (GO:0006306)  
203 and histone methylation (GO:0016571) were collected using QuickGO (<http://www.ebi.ac.uk/>). Putative  
204 *N. thermarum* homologs of *A. thaliana* transcripts were identified via BLASTX for each *N. thermarum*  
205 transcript against a database of all TAIR 10 *Arabidopsis* transcript amino acid sequences (Lamesch 2012),  
206 using the hit with the lowest e-value as the putative homolog match (e-value cutoff = 1e-15).

207

## 208 **Histology and Microscopy**

209 Material collected for microscopy was fixed in 4% v/v acrolein (Polysciences, New Orleans, Louisiana,  
210 USA) in 1X PIPES buffer (50 mmol/L PIPES, 1 mmol/L MgSO<sub>4</sub>, 5 mmol/L EGTA) pH 6.8, for 24 hours. Fixed  
211 material was then rinsed three times (one hour per rinse) with 1X PIPES buffer, dehydrated through a  
212 graded ethanol series, and stored in 70% ethanol. Samples were prepared for confocal microscopy and  
213 imaged according previously established protocols (Povilus 2015). Briefly: tissues were stained according  
214 to the Fielgen method, and then infiltrated with and embedded in JB-4 glycol methacrylate (Electron

215 Microscopy Sciences, Hatfield, PA, USA). Blocks were cut by hand with razor blades to remove  
216 superfluous tissue layers. Samples were mounted in a drop of Immersol 518f (Zeiss, Oberkochen,  
217 Germany) on custom well slides and imaged with a Zeiss LSM700 Confocal Microscope, equipped with  
218 an AxioCam HRc camera (Zeiss, Oberkochen, Germany). Excitation/emission detection settings:  
219 excitation at 405 and 488 nm, emission detection between 400-520 nm (Channel 1) and 520-700  
220 (Channel 2).

221

## 222 **Results**

### 223 **Generation and analysis of RNA-seq Data**

224 Between 66 and 100 million high quality reads were generated for each sample, for a total of  
225 940 million reads. 76.2% of the reads pseudo-mapped to the *Nymphaea thermarum* genome, and  
226 uniquely mapped reads were used to estimate normalized transcript abundance as TPM (transcripts per  
227 million)(Supplementary Table 1). Biological replicates of each time point clustered with each other (and  
228 not with samples of other time points) during PCA and k-means analysis of expression patterns of the  
229 4000 most highly expressed transcripts, except for one sample of 0 DAA seeds that clustered instead  
230 with 7 DAA samples (Figure 2A). This sample was removed from further analysis. When PCA and k-  
231 means clustering were performed with the remaining 11 samples, samples clustered according to  
232 collection time point.

233 In total, 19,412 unique transcripts with at least a minimum abundance of 1 TPM were present  
234 during the sampled time points (Supplemental Dataset 1). This represents 74.4% of the 25,760 genes  
235 identified from the *Nymphaea thermarum* genome (Figure 2B). 16,329 transcripts were present at all  
236 three ovule/seed developmental stages. 0 DAA had the most unique transcripts, and 7 DAA had the  
237 fewest. Among transcripts present in two stages, 0 DAA and 7 DAA shared the most transcripts, while 7  
238 DAA and 15 DAA shared the fewest. The majority of transcript expression levels fell within a similar

239 range across all three stages (Figure 2C). However, the expression levels of the 0.1% most highly  
240 expressed transcripts increased significantly between 7 and 15 DAA (Supplementary Table 2).

241 Besides differences in TPM values among the most highly expressed transcripts, the types of  
242 genes represented by the 10 most highly expressed transcripts varied with time point (Table 1). At 0  
243 DAA, most of the 10 most highly expressed transcripts coded for structural components of histones or  
244 ribosomes. Notably, a putative homolog to an arabinogalactan peptide (AGP16) was the third most  
245 highly expressed transcript at 0 DAA. Arabinogalactans are known to regulate female gametophyte  
246 development and function in pollen tube interactions in other angiosperms (Pereira 2016). At 7 DAA,  
247 ribosome components again featured prominently among the 10 most highly expressed transcripts. A  
248 lipid transfer protein, WAXY starch synthase, and TPS10 terpene synthase were also highly expressed at  
249 7 DAA, likely in relation to the initiation of nutrient import and storage in the seed, and seed coat  
250 differentiation. At 15 DAA, several transcripts involved with terpene synthesis or modification were  
251 among the 10 most highly expressed transcripts, coincident with continued maturation of the seed coat.  
252 A WAXY starch synthase, highly expressed at 7 DAA, continued to be highly expressed at 15 DAA.

253

#### 254 **Analysis of Differential Gene Expression**

255 Out of the 19,412 unique transcripts expressed during seed development in *N. thermarum*,  
256 10,933 were significantly differentially expressed (DE) (Supplementary Dataset 2). The set of DE  
257 transcripts was used to perform hierarchical clustering of all transcripts in all samples (Figure 3A). The  
258 three main 'clades' identified by hierarchical clustering of samples correlated with the three sampling  
259 time points. 7 DAA samples and 15 DAA samples were more closely related to each other, than to 0 DAA  
260 samples.

261 Two sets of differentially expressed (DE) transcripts were considered during further analysis: a  
262 set of all DE transcripts ("all-DE"), and a set of DE transcripts with very high relative changes in

263 expression (“highly-DE”). For the latter set, an absolute b-value more than 2 for the expression  
264 change(s) was used as the filtering criteria. B-values are reported by sleuth (companion to the pseudo-  
265 mapping program kallisto) as part of differential gene expression analysis, and are analogous to fold-  
266 change in what a positive or negative value means for the direction of expression change (Pimentel  
267 2017). However, b-values are derived from the effect size of time point on the log<sub>10</sub>-transformed  
268 transcript abundances – a b-value is therefore not equivalent to the same value fold-change (ie: a b-  
269 value of 2 does not imply a fold-change of 2).

270         Among the set of all DE transcripts (10,933 transcripts), more than three times more transcripts  
271 were differentially expressed between 0-7DAA, than between 7-15 DAA, while the number of transcripts  
272 differentially expressed between 0-7 DAA and 0-15 DAA was more similar (Figure 3B). For each time-  
273 point comparison, between 53 and 58% of the significant changes in expression were due to increases in  
274 expression (as opposed to decreases in expression). The set of highly-DE transcripts was smaller,  
275 consisting of 1,865 unique transcripts. The 7-15 DAA transition had the fewest highly-DE transcripts,  
276 with 0-7 DAA having about 2.5 times as many, and 0-15 DAA having about twice as many as 0-7DAA.  
277 While the proportion of transcripts that increased expression between 0-7DAA and 0-15 DAA was similar  
278 to what was seen in the set of all DE transcripts (between 54-58%), for 7-15DAA the proportion of  
279 highly-DE transcripts that increased expression was much higher (80%, as compared to 58% for the set  
280 of all DE transcripts).

281

## 282 **Analysis of transcripts grouped by expression pattern**

283         The expression patterns of all DE transcripts were associated with 9 expression pattern types (or  
284 ‘clusters’) using a K-means clustering approach (Figure 4A) (Supplementary Table 3). The expression  
285 patterns represented by the 9 clusters include: increased or decreased expression across the entire time  
286 sampled (respectively, Clusters A and B), as well as increased or decreased expression to produce a

287 minimum or maximum at each of the 3 time points (increase and decrease associated with minimum or  
288 maximum at, respectively, 0 DAA = Clusters C,D; 7 DAA = E,F; 15DAA = G,H). The final cluster (cluster I)  
289 represented transcripts that, while differentially expressed, displayed a relatively small magnitude of  
290 change. 10,450 transcripts (96% of DE transcripts) were strongly correlated with the average profile of  
291 their respective clusters (Pearson correlation > 0.9) and were considered for further analysis.

292 In addition to the set of 10,450 DE transcripts represented in the expression pattern clusters  
293 (“all-DE”), a subset of “highly-DE” transcripts ( b-value > 2 for at least one time point transition: 1,783  
294 transcripts) was considered during further analysis of expression clusters (Figure 4B). For the cluster  
295 pairs that represent general change (A,B) or minimum/maximum expression at 0 DAA (C,D), more  
296 transcripts showed expression decreases. In contrast, in the cluster pairs that represent  
297 minimum/maximum expression at 7 DAA (E,F) and 15 DAA (G,H), more transcripts showed increased  
298 expression.

299

### 300 **Functional enrichment of expression pattern clusters**

301 Each cluster was tested for significant enrichment of GO molecular function terms, based on  
302 the TAIR 10 annotations for the putative *A. thaliana* homolog of each *N. thermarum* transcript. All  
303 significantly enriched child terms (ie: the most specialized of a hierarchy) are reported for each cluster,  
304 and terms of particular interest are further discussed (Figure 5).

305 Clusters with general or time-point-specific increases in expression were found to be generally  
306 enriched for functions related to various types of transmembrane transporter activity. For the set of all-  
307 DE transcripts, Cluster G (maximum at 15 DAA) was additionally enriched for functions that appear to be  
308 related to chloroplast activity (chlorophyll-binding, electron-carrier activity). While it seems unlikely that  
309 seeds at this stage (which are enclosed within opaque fruit walls) would be carrying out photosynthesis,  
310 the analogous stage of embryo development in *Arabidopsis* is associated with the formation of

311 chloroplasts within embryo tissues (Mansfield 1991). Among the set of highly-DE transcripts, various  
312 transmembrane transporter activities were again enriched in Clusters A (general increase) and G  
313 (maximum at 15DAA). Cluster C (increase after 0 DAA) was enriched for lipid binding, and Cluster E  
314 (maximum at 7DAA) was enriched for transcription factor activity.

315 Clusters associated with general or time-point-specific decreases in expression prominently  
316 featured significantly enriched terms related to DNA or chromatin binding and modification. For the set  
317 of all-DE transcripts, Cluster B (general decrease) was enriched for terms related to methyltransferase  
318 activity and Cluster D (decrease after 0 DAA) was enriched for transcription factor activity. Among the  
319 set of highly-DE transcripts, Cluster D (decrease after 0 DAA) was enriched for transcription factor  
320 activity.

321

## 322 **TF expression during seed development**

323 The fact that clusters that represent both increases and decreases in expression were enriched  
324 for transcription factors merited further investigation of transcription factor activity. Of the 1,268  
325 putative transcription factors identified from the *N. thermaurum* genome, 1,039 were expressed during  
326 the sampled stages (with a TPM > 1), and 719 were significantly differentially expressed. Among the set  
327 of DE transcription factors, we examined whether expression pattern clusters were significantly  
328 enriched for any of 58 transcription factor families (Figure 6). For the all-DE transcription factor dataset,  
329 Cluster B (general decrease) was enriched for FAR1, Cluster D (decrease after 0 DAA) was enriched for  
330 GRF, and Cluster E (maximum at 7 DAA) was enriched for MYB transcription factors. When only highly-  
331 DE transcription factors were considered ( $b > 2$ ), Cluster C (increase after 0 DAA) was enriched for  
332 WRKY, Cluster D (decrease after 0 DAA) was enriched for ZF-HD and GRF, and cluster E (maximum at 7  
333 DAA) was enriched for MYB activity.

334

335 **Activity of genes associated with imprinting via DNA and histone methylation**

336 Methyltransferase-related terms were enriched in cluster B (consistent decrease in expression),  
337 already hinting at potential for a dynamic DNA and histone methylation landscape during reproductive  
338 development in *N. thermarum*. We constructed gene family phylogenies for genes that are known to be  
339 important regulators of epigenetic patterning during reproduction, with a particular focus on those  
340 involved in gene imprinting (CMT, MET, DME, and the PCR2 components MEA, FIS, FIE, and MSI). Many  
341 of the relationships between gene family members corroborates previous studies (Furihata 2016)  
342 (Bewick 2017).

343 First, we examined genes involved in the establishment or maintenance of imprinting-related  
344 DNA methylation in the CG and CHG contexts: CMT and MET. Although there are 4 MET homologs in  
345 *Arabidopsis*, they appear to be the result of clade-specific gene duplications; the two *N. thermarum* MET  
346 homologs are similarly the result of a clade-specific gene duplication (Figure 7A). Only one *N. thermarum*  
347 CMT homolog was identified, although it's affinity for either of the CMT2 or CMT1/CMT3 clades was  
348 poorly resolved (Figure 7B). All *N. thermarum* homologs of CMT and MET were differentially expressed  
349 during reproductive development in *N. thermarum*, and belonged to expression Cluster D (decrease  
350 after 0DAA) (Figure 9).

351 DME, on the other hand, removes certain types of methylation marks from DNA. We find that  
352 angiosperm DME genes were divided into two poorly-supported clades: one with DME and DML1, and  
353 one with DML2 and DML3 (Figure 7C). The *N. thermarum* DME homologs formed a single well-supported  
354 clade, suggesting clade-specific gene duplication events, that was placed (with poor support) within the  
355 [DME, DML1] clade. Three of the four *N. thermarum* DME homologs were in expression Cluster D  
356 (decrease after 0 DAA). The fourth and most highly expressed DME homolog, while in expression Cluster  
357 H (minimum at 15 DAA), did in fact display a significant increase in expression after 0 DAA (Figure 9).

358 *N. thermarum* homologs were also identified for all components of PRC2, and all were  
359 expressed during reproductive development. Angiosperm homologs of MEA formed two well-supported  
360 clades: one with MEA and SWN, and one with CLF (Figure 8A). Two *N. thermarum* MEA homologs were  
361 identified, with one present in each of the MEA clades. The *N. thermarum* homolog within the  
362 MEA/SWN clade was expressed during reproductive development, but not differentially expressed; the  
363 *N. thermarum* homolog of CLF associated with expression cluster D (decreased expression after ODAA)  
364 (Figure 9). Two *N. thermarum* homologs of MSI1 were identified, and both associated with expression  
365 cluster B (consistent decreased expression) (Figures 8, 9). The FIE gene family appeared to be relatively  
366 simple, with little indication of gene duplications outside of monocots – one copy of FIE was identified  
367 from *N. thermarum* (Figure 8C). Angiosperm FIS2 genes formed two well-supported clades: one  
368 appeared to be specific to *Arabidopsis* (and included VRN2 and FIS2), while the other included EMF2.  
369 Only one homolog was identified from *N. thermarum*, and its placement within the EMF2 clade was  
370 poorly supported (Figure 8D). The FIE and FIS2 homologs in *N. thermarum* were expressed during  
371 reproductive development, but their expression did not significantly change during the sampled time  
372 points (Figure 9).

373

#### 374 **Broader analysis of gene activity associated with DNA and histone methylation**

375 We next used a broader approach to examine the expression of any gene that could be involved  
376 in regulation of DNA or histone methylation patterns (Figure 9). 121 loci in *Arabidopsis thaliana* are  
377 annotating as being involved in DNA or histone methylation; 125 putative homologs were identified  
378 from within the *N. thermarum* genome using pair-wise blast comparisons. 112 of the *Nymphaea*  
379 methyltransferase-related homologs were expressed in mature ovules or during seed development with  
380 a TPM >1, and of those 73 were significantly differentially expressed. Of the 112 putative DNA or histone  
381 methyltransferase-related homologs in *N. thermarum* expressed in mature ovules or developing seeds,

382 20 of them were not present in transcriptomes of root tips, leaves, young floral buds, or young ovules  
383 (Povilus 2020). 11 of the mature ovule/seed-development-specific homologs were differentially  
384 expressed, including putative homologs of RDM12, MTHFD1, MET, FDM1, CYP71, SUVR4, and ATX2; all  
385 were in the expression-pattern clusters that represented either general expression decrease (Cluster B),  
386 or decreased expression after ODAA (Cluster D).

387 Most of the *Nymphaea* DNA and histone methylation-associated transcripts were in expression-  
388 pattern clusters that represented decreased expression at some point (Clusters B,D,H). Only 11  
389 transcripts were among clusters that involved an increase in expression (A,C,E), including homologs of  
390 DTM7, DRM, GEM, CDC73, EFM, VIP3, SUVR3, and APRF1. Among the set of non-DE transcripts, a few  
391 were present with fairly high abundance, including homologs of ZOP1, HTA9, and FIB1.

392

## 393 **Discussion**

### 394 **Overview**

395 We leverage the ability of RNAseq datasets to move beyond a candidate gene approach, to  
396 broadly study seed development in the water lily *Nymphaea thermarum*, and specifically the processes  
397 involved in regulating imprinting-related and non-imprinting-related DNA and histone methylation. We  
398 find that all components of known imprinting mechanisms are expressed during reproductive  
399 development in *N. thermarum*, and that many other DNA or histone methylation regulators are  
400 differentially expressed. This indicates that not only is the epigenetic landscape likely to be dynamic  
401 during reproduction in *Nymphaea*, but that imprinting may also be occurring in this species.  
402 Comparisons with patterns of gene expression during reproductive development in other angiosperms  
403 suggests that the current model for how imprinting is regulated, perhaps best studied in *Arabidopsis*, is  
404 likely a mix of deeply conserved and eudicot-specific processes. Finally, we are able to suggest that the

405 function of several histone-methylation genes merit further investigation during seed development in  
406 not only *N. thermarum*, but any model system.

407

#### 408 **Patterns of gene expression during reproductive development**

409 We find that a large proportion of genes is expressed during reproductive development in *N.*  
410 *thermarum*: 74% of the total number of genes predicted from the *N. thermarum* genome. Furthermore,  
411 56% of the expressed transcripts are differentially expressed. The proportion of genes expressed, either  
412 differentially or not differential, during reproductive development is similar to what has been described  
413 in other species (Chen 2014). The number of unique DE transcripts in *N. thermarum* suggests that the  
414 transitions from female gametophyte maturation, through fertilization, and into mid-seed development  
415 require substantial transcriptional reprogramming (Figure 2B). While female gametophyte and ovule  
416 maturation involve relatively high numbers of unique genes (1,148), the expression of almost as many  
417 genes (916) appears to carry over into early seed development. 7 and 15 DAA shared the expression of  
418 far fewer genes (339; other than those shared by all three stages) – a surprising result given that 7 and  
419 15 DAA are understood to share more developmental processes than 0 and 7 DAA. However, a previous  
420 study provided evidence for a lingering maternal influence on early seed development in *N. thermarum*  
421 (Povilus 2018), which is congruent with the relatively large number of transcripts shared between 0 and  
422 7 DAA. Furthermore, when transcript expression pattern (not just presence/absence of transcripts) is  
423 taken into account, 7 and 15 DAA samples were more similar to each other than either were to 0 DAA  
424 samples (Figure 3A).

425 When DE transcripts are clustered by expression pattern, there are clear similarities among the  
426 enriched putative molecular functions for clusters that represent either increases or decreases in  
427 expression. As could be predicted by the onset of nutrient import and storage after fertilization, most of  
428 the “increased-expression” clusters were enriched for transporter activities, and a homolog of WAXY

429 starch synthase 1 was among the 10 most highly expressed genes at 7 and 15 DAA. However, we also  
430 note patterns of gene expression associated with the onset of embryogenesis and/or endosperm  
431 development: highly-DE transcripts in Cluster E (maximum expression at 7 DAA) were enriched for  
432 transcription factor activity. Furthermore, the set of DE transcription factors in Cluster C and E (both  
433 involve expression increased between 0 and 7 DAA) were enriched for (respectively) WRKY and MYB  
434 genes, which have been associated with embryo and endosperm development in both *Arabidopsis* and  
435 *Zea mays* (Lagacé 2004)(Luo 2005)(Dubos 2010) (Wickramasuriya 2015).

436 Expression pattern clusters that represent a decrease in expression were enriched with an  
437 altogether different set of molecular functions. DNA/chromatin binding, transcription factor activity, and  
438 control of DNA polymerase are featured prominently in both the all-DE and highly-DE datasets.  
439 Intriguingly, expression Cluster D (decrease after 0 DAA) was enriched for GRF and ZF-HD transcription  
440 factors, which are associated with, among other things, cell division and floral development  
441 (Omidbakhshfard 2015). We attribute the pattern of decreased DNA-modification or transcription-  
442 regulation functions to either the cessation of cell proliferation and differentiation associated with ovule  
443 development, and/or the transition from floral development programs to seed development programs.

444

#### 445 **Evidence for dynamic epigenetic landscape during reproductive development**

446 In *N. thermarum*, 125 genes putatively share homology with *Arabidopsis* genes involved in DNA  
447 or histone methylation. A remarkable 89% of these *N. thermarum* homologs are expressed in mature  
448 ovules or during seed development (at TPM > 1), with 58% being differentially expressed, suggesting a  
449 dynamic epigenetic landscape during reproductive development in this species. Among the gene  
450 families known to specifically regulate imprinting-related methylation patterns, MET and CMT homologs  
451 were recovered in expression Cluster D (decreased expression after 0 DAA), as were three-quarters of  
452 the DME homologs. Furthermore, one of the MET homologs appears to be specifically expressed during

453 seed development. The fourth *N. thermarum* DME, while associated with expression cluster H  
454 (decreased expression after 7 DAA), did in fact display a significant increase in expression after 0 DAA.  
455 All components of PRC2 were expressed during reproductive development.

456 Many components of the RdDM pathway were present during the sampled developmental  
457 stages in *N. thermarum*. Most fell into expression-pattern clusters B and D (consistent decrease in  
458 expression, or decreased expression after 0 DAA). Interestingly, most of the DNA or histone  
459 methylation-related homologs expressed only during seed development are components of the RdDM  
460 pathway (RDM12, FDM1), are known to be involved in chromatin remodeling (CYP71), and/or have been  
461 specifically tied to transposon repression (SUVR4, MTHFD1). In addition, DRM, an important component  
462 of the RdDM pathway, showed increased expression after fertilization. Homologs of several genes  
463 involved in histone methylation (GEM, CYP71, EFM, VIP3, SUVR3, APRF1) showed increased expression  
464 after fertilization. Several of these genes have not been previously linked to seed development in any  
465 species; we therefore suggest that their role during sexual reproduction deserves further investigation in  
466 *N. thermaurm* and other angiosperms, such as *Arabidopsis*, rice, and maize.

467 Altogether, our data suggests that DNA methylation patterns are being established, maintained,  
468 and removed before fertilization in *N. thermarum*. After fertilization, gene activity related to DNA  
469 methylation maintenance in the CG and CHG context (CMT, MET) decreases, while the expression of  
470 some genes involved in DNA demethylation (DME) and CHH-context de novo methylation (DRM and  
471 other RdDM components) increases. The components of PRC2, which establish loci-specific H3K27  
472 methylation patterns associated with imprinting, all decrease in expression over time. By the time that  
473 the embryo typically initiates cotyledons at 15DAA, the expression of nearly all DNA and histone  
474 methylation-related genes has decreased in whole seeds, relative to their levels in pre-fertilization  
475 ovules.

476

477 **Comparison of imprinting-related DNA and histone methylation activity with other angiosperms**

478 DNA and histone methylation during sexual reproduction has been studied for a small handful of  
479 distantly-related angiosperms, in particular the eudicot *Arabidopsis* and the monocot *Oryza* (rice)(Köhler  
480 2012). Importantly, DNA and histone methylation have been shown to be dynamic during seed  
481 development in every taxon which has been studied. Although there is wide-spread evidence for parent-  
482 of-origin effects on seed development (Haig 1991), the molecular/genetic evidence for imprinting, which  
483 depends on patterning of DNA and histone methylation, is less consistent (Gleason 2012). It must be  
484 noted, however, that developmental stage sampling is inconsistent in many studies of gene expression  
485 during seed development (due in part to fundamental differences in how seeds develop), so  
486 comparisons should be approached with caution. In addition, complex, lineage-specific histories of gene  
487 duplication and loss can make it difficult to assess specific homology relationships within gene families.

488 A summary of expression patterns for genes related to imprinting in ovules and seeds of  
489 *Nymphaea*, monocots (mostly *Oryza*), and *Arabidopsis* is presented in Figure 10. CMT, and MET  
490 homologs all show decreased expression after fertilization in *Nymphaea*. While decreased expression of  
491 CMTs is similar to what is seen in *Arabidopsis* and rice, the expression pattern of the *Nymphaea* METs is  
492 the opposite of those in *Arabidopsis* and rice (Sharma 2009)(Julien 2012). The comparison for DME is  
493 more complex – one DME copy in *Nymphaea* shares the expression pattern with one barley DME  
494 homolog (Kapazoglou 2013). The expression of the second DME copy in *Nymphaea* is more similar to  
495 most of the rice DME homologs and to the *Arabidopsis* DME (Choi 2002)(Jiang 2016). All copies of PRC2  
496 components decrease in expression after fertilization in *Nymphaea*, while expression patterns of  
497 individual components show more variation in *Arabidopsis* and rice (Baroux 2006)(Anderson  
498 2013)(Nallamilli 2013).

499 Based on the complexity of DNA methylation-related activity, we conclude that DNA  
500 methylation is likely as dynamic during reproductive development in *N. thermarum*, as it is in other

501 angiosperms. If imprinting occurs, however, its regulation is likely different than what is known for  
502 *Arabidopsis* – particularly with respect to the roles of MET and PRC2. Overall, maintenance of CG  
503 methylation by MET may be relatively less important after fertilization. In contrast, de novo CHH  
504 methylation by DRMs and other RdDM components may be relatively more important - whether or not  
505 they are related to imprinted gene expression. Interestingly, there is little evidence that PRC2 as a whole  
506 is a major regulator of seed development in rice (Luo 2009). Yet imprinting occurs in monocots, and  
507 impacts endosperm development – other molecular machinery must be responsible for regulating and  
508 responding to imprinting-related methylation patterns. Until the functions of the *Nymphaea* PRC2  
509 homologs can be determined, we suggest that the role of the PRC2 as a whole in regulating imprinting  
510 may represent a derived condition within eudicots. Importantly, *N. thermarum* homologs of all of the  
511 genes known to be involved in imprinting via DNA or histone methylation are expressed in mature  
512 ovules or developing seeds. While parental-allele-specific RNA expression data is required for  
513 verification, our results indicate it is possible that imprinting via regulation of DNA and histone  
514 methylation may be occurring in this species.  
515

516

517

518 **Declarations**

519 **Funding**

520 We acknowledge support from the National Science Foundation: IOS-0919986 awarded to W.E.F., and

521 DEB-1500963 and IOS-1812116 awarded to R.A.P..

522 **Conflicts of interest/Competing interests**

523 The authors declare no conflicts of interest or competing interests.

524 **Availability of data and material**

525 Raw sequence data and assembled transcriptomes of *N. thermarum* have been submitted to the

526 National Center for Biotechnology Information (NCBI) database under BioProject PRJNA718528.

527 Biological material and all other data are available as Supplemental Data, or from the corresponding

528 authors upon request.

529 **Code availability**

530 Not Applicable

531 **Authors' contributions**

532 R.A.P and W.E.F. conceived of original premise of the project. R.A.P. grew plant samples, performed

533 experiments, and analyzed data. R.A.P. wrote the manuscript with input from W.E.F.

534 ***Additional declarations for articles in life science journals that report the results of studies involving***

535 ***humans and/or animals***

536 Not applicable

537 **Ethics approval (include appropriate approvals or waivers)**

538 Not applicable

539 **Consent to participate (include appropriate statements)**

540 Not applicable

541 **Consent for publication (include appropriate statements)**

542 All authors have given consent to publish this work.

543

544 **Acknowledgements:** We acknowledge support from the National Science Foundation: IOS-0919986

545 awarded to W.E.F., and DEB-1500963 and IOS-1812116 awarded to R.A.P.. We thank the Botanische

546 Gärten der Universität Bonn for providing original plant material for propagation.

547

548 **References**

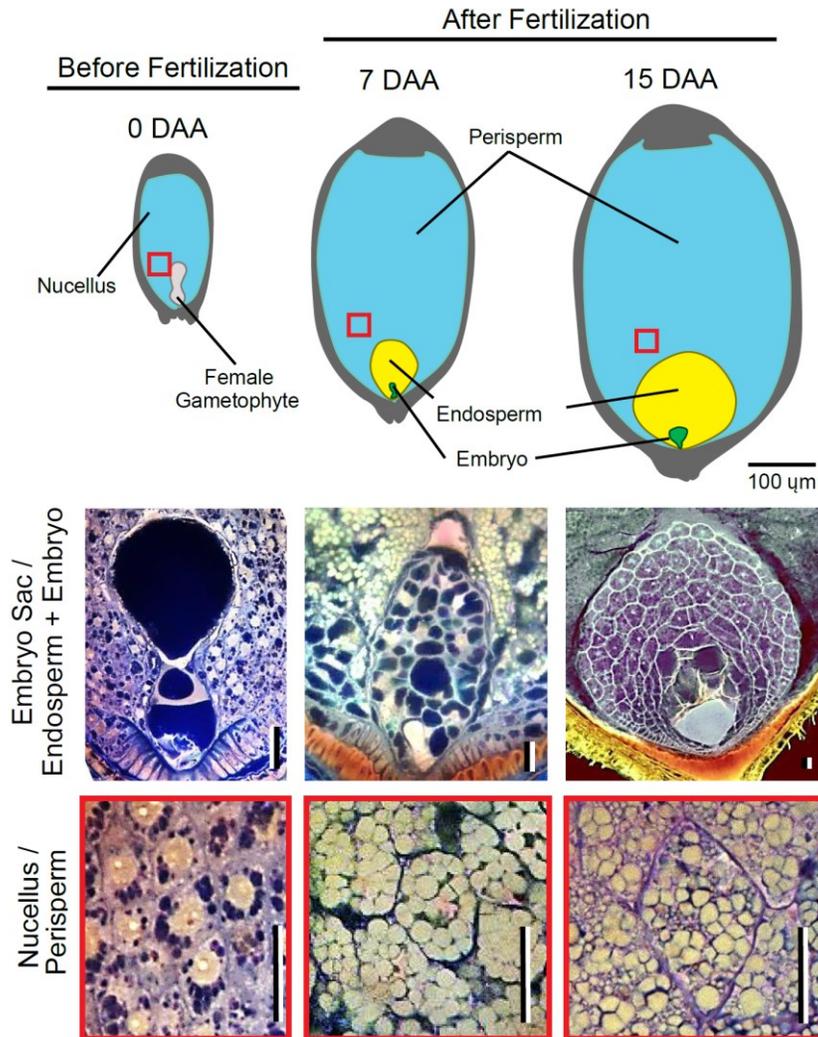
- 549 Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science*  
550 342: 6165
- 551 Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, Sundaresan V (2013) Transcriptomes  
552 of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent  
553 chromatin and epigenetic states before fertilization. *The Plant Journal* 76(5):729-741.
- 554 Baroux C, Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of Polycomb  
555 group genes: MEDEA autoregulation is required for imprinted gene expression in *Arabidopsis*. *Genes Dev*  
556 20(9): 1081-1086.
- 557 Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM,  
558 Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental  
559 profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proceedings of the National*  
560 *Academy of Sciences* 110(5): E435-E444.
- 561 Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ (2017) The evolution of  
562 CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome Biology* 18:65.
- 563 Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification.  
564 *Nature Biotechnology* 34: 525–527.
- 565 Chen J, Zeng M, Xie S, Wang G, Hauck A, Lai J (2014) Dynamic transcriptome landscape of maize embryo  
566 and endosperm development. *Plant Physiology* 166: 252-264.
- 567 Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002)  
568 DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed  
569 viability in *Arabidopsis*. *Cell* 110(11): 33-42.
- 570 Dubos C, Stracke R, Grotewold E, Weisshae B, Martin C, Lepiniec L (2010) MYB transcription factors in  
571 *Arabidopsis*. *Trends in Plant Science* 15(10): 573-581.
- 572 Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology* 7: e1002195
- 573 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*  
574 *Acids Research* 32(5): 1792-97.
- 575 Fischer E, Magdalena-Rodriguez C (2010) *Nymphaea thermarum* (Nymphaeaceae). *Curtis Botanical*  
576 *Magazine* 27: 318–327.
- 577 Furihata HY, Suenaga K, Kawanabe T, Yoshida T, Kawabe A (2016) Gene duplication, silencing, and  
578 expression alteration govern the molecular evolution of PRC2 genes in plants. *Genes and Genetics*  
579 *Systems* 91(2): 85-95.
- 580 Gao Y, Xu H, Shen Y, Wang J (2013) Transcriptome analysis of rice (*Oryza sativa*) developing endosperm  
581 using the RNA-Seq technique. *Plant Molecular Biology* 81(4-5): 363-378.

- 582 Gao J, Yu X, Ma F, Li J (2014) RNA-seq analysis of transcriptome and glucosinolate metabolism in seeds  
583 and sprouts of broccoli (*Brassica oleracea var. italica*). *PLoS ONE* 9(2): e88804.
- 584 Gehring M, Satyaki PR (2017) Endosperm and imprinting, inextricably linked. *Plant Physiology* 173: 143-  
585 154
- 586 Girke T, Todd J, Ruuska S, White J, Benning C, Ohirogge J (2000) Microarray analysis of developing  
587 *Arabidopsis* seeds. *Plant Physiology* 124: 1570-1581.
- 588 Gleason EJ, Kramer EM (2012) Characterization of *Aquilegia* Polycomb Repressive complex 2 homologs  
589 reveals absence of imprinting. *Gene* 507(1): 54–60.
- 590 Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N,  
591 Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*  
592 40(Database issue): D1178-1186.
- 593 Haig D, Westoby M (1991) Genomic imprinting in endosperm: its effects on seed development in crosses  
594 between species and between different ploidies of the same species, and its implications for the  
595 evolution of apomixis. *Philosophical Transactions of the Royal Society B* 333: 1–13.
- 596 Haig D (2013) Kin conflict in seed development: an interdependent but fractious collective. *Annual*  
597 *Review of Cell and Developmental Biology* 29: 189-211.
- 598 Hirsch C, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F,  
599 Hernandez AG, Fields CJ, Wright CL, Koehler K, Springer NM, Buckler ES, Buell CR, de Leon N, Kaeppler  
600 SM, Childs K, Mikel MA (2016) Draft assembly of elite inbred line PH207 provides insights into genomic  
601 and transcriptome diversity in maize. *Plant Cell* 28(11): 2700-2714.
- 602 Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RD, Harada JJ, Ziberman D,  
603 Fischer RL (2012) Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proceedings of the*  
604 *National Academy of Sciences* 108(5): 1755-1762.
- 605 Jiang SY, Ramachandran S (2016) Expansion mechanisms and evolutionary history on genes encoding  
606 DNA glycosylases and their involvement in stress and hormone signaling. *Genome Biology and Evolution*  
607 8(4): 1165-1184.
- 608 Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC and Gao G. (2017). PlantTFDB 4.0: toward a central hub  
609 for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45(D1): D1040-  
610 D1045.
- 611 Jones SI, Vodkin LO (2013) Using RNA-seq to profile soybean seed development from fertilization to  
612 maturity. *PLoS ONE* 8(3): e59270.
- 613 Julien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F (2012) DNA methylation dynamics during  
614 sexual reproduction in *Arabidopsis thaliana*. *Current Biology* 22(19): 1825-1830.
- 615 Kapazoglou A, Drosou V, Argiriou A, Tsaftaris AS (2013) The study of a barley epigenetic regulator,  
616 HvDME, in seed development and under drought. *BMC Plant Biology* 13: 172.

- 617 Köhler C, Wolff P, Spillane C (2012) Epigenetic mechanisms underlying genomic imprinting in plants.  
618 *Annual Review of Plant Biology* 63: 331-352.
- 619 Lagacé M, Matton DP (2004) Characterization of a WRKY transcription factor expressed in late torpedo-  
620 stage embryos of *Solanum chacoense*. *Planta* 219(1): 185-189.
- 621 Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL,  
622 Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012)  
623 The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids*  
624 *research* 40(Database issue): D1202-D1210.
- 625 Li G, Wang D, Yang R, Logan K, Zhang S, Skaggs M, Lloyd A, Burnette WJ, Laurie JD, Hunter BG,  
626 Dannenhoffer JM, Larkins BA, Drews GN, Wang X, Yadegari R (2014) Temporal patterns of gene  
627 expression in developing maize endosperm identified through transcriptome sequencing. *Proceedings of*  
628 *the National Academy of Sciences* 111(21): 7582-7587.
- 629 Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) *MINISEED2 (MINI3)*, a WRKY family gene,  
630 and *HAIKU2 (IKU2)*, a leucine-rich repeat (LRR) *KINASE* gene, are regulators of seed size in *Arabidopsis*.  
631 *Proceedings of the National Academy of Sciences* 102(48): 17531-17536.
- 632 Luo M, Platten D, Chaudhury A, Peacock WJ, Dennis ES (2009) Expression, imprinting, and evolution of  
633 rice homologs of the polycomb group genes. *Molecular Plant* 2: 711–723
- 634 Mansfield SG, Briarty LG (1991) Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo.  
635 *Canadian Journal of Botany* 69(3): 461-476.
- 636 Nallamilli BRR, Zhang J, Mujahid H, Malone BM, Bridges SM, Peng Z (2013) Polycomb group gene OsFIE2  
637 regulates rice (*Oryza sativa*) seed development and grain filling via a mechanism distinct from  
638 *Arabidopsis*. *PLoS Genetics* 9(3): e1003322.
- 639 Nguyen HT, Silva JE, Podicheti R, Macrander J, Yang W, Nazerenus TJ, Nam JW, Jaworski JG, Lu C,  
640 Scheffler BE, Mackaitis K, Cahoon EB (2013) Camelina seed transcriptome: a tool for meal and oil  
641 improvement and translational research. *Plant Biotechnology Journal* 11(6): 759-769.
- 642 Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L,  
643 Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements  
644 and new features. *Nucleic acids research* 35(Database issue): D883-D887.
- 645 Pereira AM, Lopes AL, Coimbra S (2016) Arabinogalactan proteins as interactors along the crosstalk  
646 between the pollen tube and the female tissues. *Frontiers in Plant Science* 7: 1895.
- 647 Pimentel HJ, Bray N, Puente S, Melsted P, Pachter L (2017) Differential analysis of RNA-Seq incorporating  
648 quantification uncertainty. *Nature Methods* 14: 687–690
- 649 Povilus RA, Losada JM, Friedman WE (2015) Floral biology and ovule and seed ontogeny of *Nymphaea*  
650 *thermarum*, a water lily at the brink of extinction with potential as a model system for basal  
651 angiosperms. *Annals of Botany* 115 :211-226

- 652 Povilus RA, Diggle PK, Friedman WE (2018) Evidence for parent-of-origin effects and interparental  
653 conflict in seeds of an ancient flowering plant lineage. *Proceedings of the Royal Society B* 285: 20172491
- 654 Povilus RA., DaCosta JM, Grassa C, Satyaki PRV, Moeglein M, Jaenisch J, Xi Z, Mathews S, Gehring M,  
655 Davis CC, Friedman WE (2020) Water lily (*Nymphaea thermarum*) genome reveals variable genomic  
656 signatures of ancient vascular cambium losses. *Proceedings of the National Academy of Sciences* 17(15):  
657 8649-8656
- 658 R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for  
659 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 660 Satyaki PRV, Gehring M (2017) DNA methylation and imprinting in plants: machinery and mechanisms.  
661 *Critical reviews in biochemistry and molecular biology* 52(2): 163-175
- 662 Sharma R, Mohan Singh RK, Malik G, Deveshwar P, Tyagi AK, Kapoor S, Kapoor M (2009) Rice cytosine  
663 DNA methyltransferases - gene expression profiling during reproductive development and abiotic stress.  
664 *The FEBS Journal* 276(21): 6301-6311.
- 665 Stamatakis A (2014) RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large  
666 phylogenies. *Bioinformatics* 30(9): 1312-1313.
- 667 Tian T, Liu T, Yan H, You Q, Yi X, Du H, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the  
668 agricultural community, 2017 update. *Nucleic Acids Research* 45(W1): W122-W129
- 669 Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit  
670 evolution. *Nature* 485(7400): 635-641.
- 671 Wang G, Wang G, Zhang X, Wang F, Song R (2012) Isolation of high quality RNA from cereal seeds  
672 containing high levels of starch. *Phytochemical Analysis* 23(2): 159-163.
- 673 Wickramasuriya AM, Dunwell JM (2015) Global scale transcriptome analysis of *Arabidopsis*  
674 embryogenesis *in vitro*. *BMC Genomics* 16(1): 301.
- 675 Xu H, Gao Y, Wang J (2014) Transcriptome analysis of rice (*Oryza sativa*) developing embryos using the  
676 RNA-Seq technique. *PLoS ONE* 7(2): e30646.
- 677 Zhang J, Shan L, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y (2011) *De novo* assembly and  
678 characterization of the transcriptome during seed development, and generation of genic-SSR markers in  
679 peanut (*Arachis hypogaea* L.). *BMC Genomics* 13: 90-96.
- 680 Ziegler, DJ, Khan, D, Kalichuk, JL, Becker, MG, Belmonte, MF (2019) Transcriptome landscape of the early  
681 *Brassica napus* seed. *Journal of Integrative Plant Biology* 61: 639– 650.
- 682 Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2006) Genome-wide analysis of *Arabidopsis*  
683 *thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature*  
684 *Genetics* 39: 61-69.
- 685

686 **Figures and Tables**



687

688 **Figure 1: Stages of ovule and seed development in *N. thermarum* sampled for RNA-seq.**

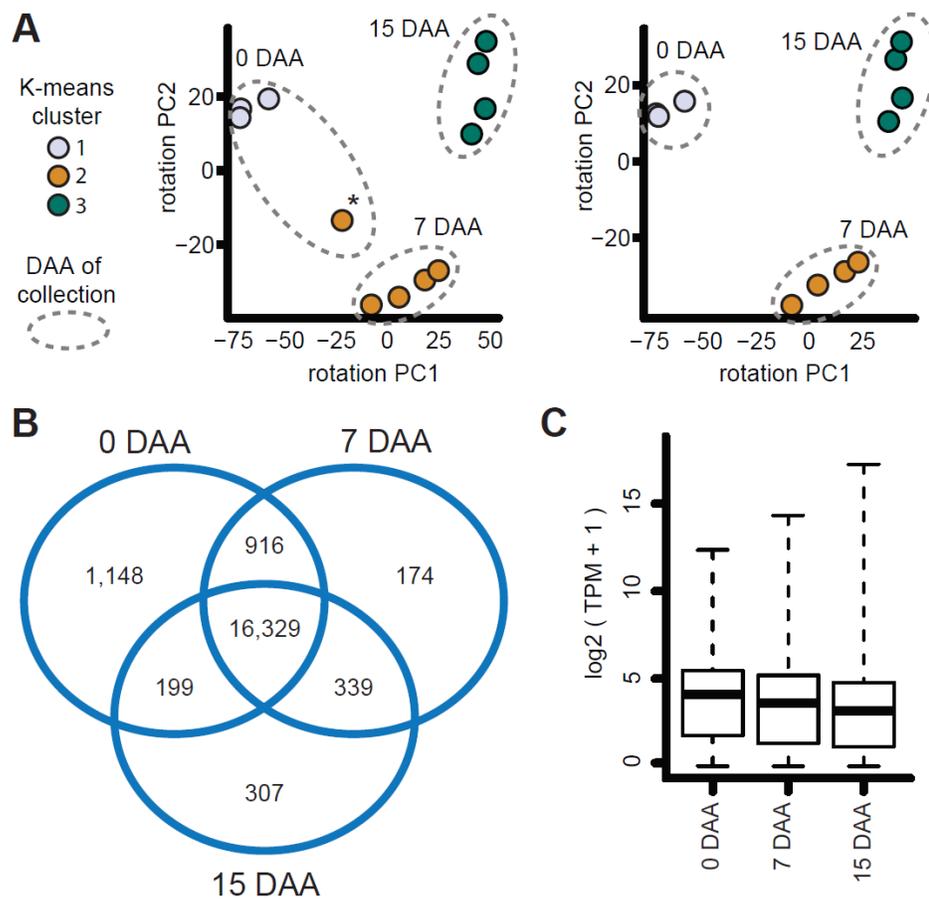
689 Top row: Diagrams of the internal structure of ovules (before fertilization, 0 days before anthesis (DAA))

690 and seeds (after fertilization) at 7 DAA and 15 DAA, with key components labeled. Red boxes indicate

691 location of corresponding image of the nucellus/perisperm (featured in the bottom row). Scale bar = 100

692  $\mu\text{m}$ . Middle and bottom rows: Confocal images of key ovule or seed components at each stage. Scale

693 bars = 20  $\mu\text{m}$ .



694

## 695 **Figure 2: Basic Analysis of Transcriptomes**

696 General characteristics of transcription in all samples. A) PCA of K-means clustering of biological  
697 replicates. Dot color indicates cluster identity, while inclusion within dashed outline indicates which DAA  
698 the sample was collected. Left graph: One 0 DAA sample clustered with 7 DAA samples (\*). This sample  
699 was considered developmentally anomalous and removed from further analysis. Right graph: PCA  
700 without the anomalous sample. B) Venn diagram of unique transcripts with a TPM > 1 at each stage, and  
701 present in multiple stages. C) Distribution of TPM values for all transcripts (TPM >1) at each stage.  
702 Median values are indicated with bold horizontal lines, bottom and top of boxes indicate 25<sup>th</sup> and 75<sup>th</sup>  
703 percentile, and dashed lines indicate minimum and maximum values.

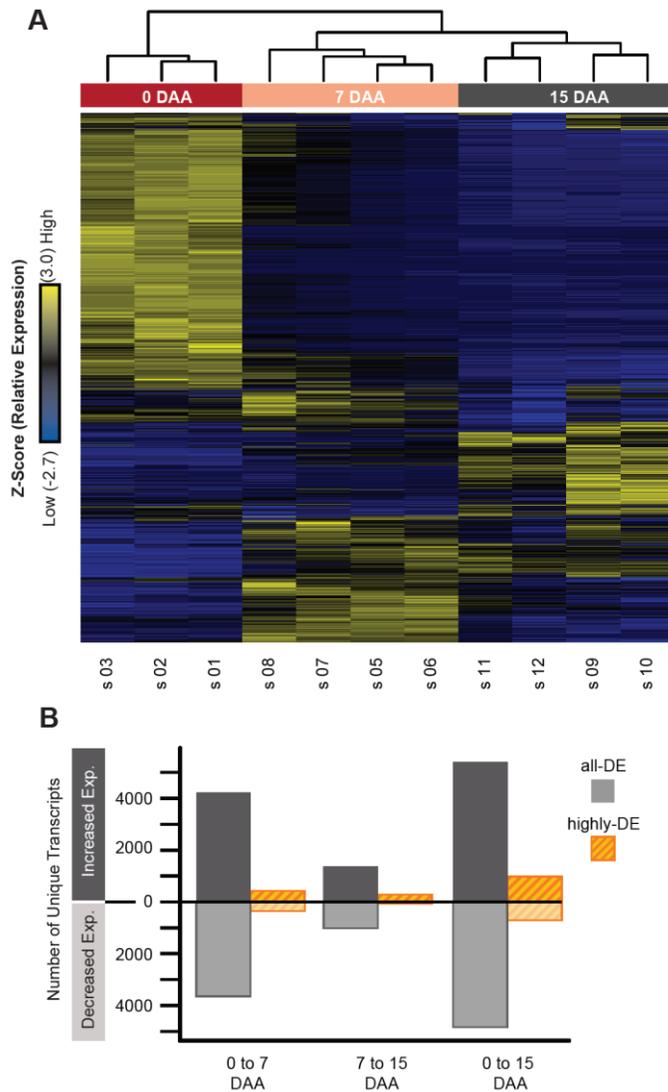
	TPM	Identifier	Putative Homology
0DAA	4966.98	NYTH03818-RA	Ubiquitin ( <i>Triticum aestivum</i> )
	4640.607	NYTH03847-RA	Histone H4 ( <i>Glycine max</i> )
	4466.02	NYTH43515-RA	AGP16 Arabinogalactan peptide 16 ( <i>Arabidopsis thaliana</i> )
	4220.917	NYTH18212-RA	RPS30A 40S ribosomal protein S30 ( <i>Arabidopsis thaliana</i> )
	3675.603	NYTH26983-RA	Os01g0645000 Zinc finger CCCH domain-containing protein 9 ( <i>Oryza sativa subsp. japonica</i> )
	3509.973	NYTH40689-RA	RPL29A 60S ribosomal protein L29-1 ( <i>Arabidopsis thaliana</i> )
	3436.017	NYTH51439-RA	Defensin J1-2 ( <i>Capsicum annuum</i> )
	3241.97	NYTH00189-RA	At4g30220 Probable small nuclear ribonucleoprotein F ( <i>Arabidopsis thaliana</i> )
	3104.7	NYTH38977-RA	Protein of unknown function
	3056.18	NYTH00233-RA	H2B Histone H2B.6 ( <i>Arabidopsis thaliana</i> )
7DAA	19593.65	NYTH51439-RA	Defensin J1-2 ( <i>Capsicum annuum</i> )
	8724.83	NYTH16912-RA	Non-specific lipid-transfer protein 1 ( <i>Lens culinaris</i> )
	8283.093	NYTH03818-RA	Ubiquitin ( <i>Triticum aestivum</i> )
	6436.833	NYTH18212-RA	RPS30A 40S ribosomal protein S30 ( <i>Arabidopsis thaliana</i> )
	6126.765	NYTH40689-RA	RPL29A 60S ribosomal protein L29-1 ( <i>Arabidopsis thaliana</i> )
	6077.737	NYTH45457-RA	HSP22 Small heat shock protein ( <i>Glycine max</i> )
	4974.465	NYTH18464-RA	WAXY Granule-bound starch synthase 1( <i>Antirrhinum majus</i> )
	4805.542	NYTH27985-RA	Protein of unknown function
	4318.512	NYTH59946-RA	TPS10 Terpene synthase 10 ( <i>Ricinus communis</i> )
	4024.218	NYTH38580-RA	RPL37C 60S ribosomal protein L37-3 ( <i>Arabidopsis thaliana</i> )
15DAA	102890.9	NYTH59946-RA	TPS10 Terpene synthase 10 ( <i>Ricinus communis</i> )
	26734.5	NYTH51439-RA	Defensin J1-2 ( <i>Capsicum annuum</i> )
	18853.65	NYTH44750-RA	(-)-alpha-terpineol synthase ( <i>Vitis vinifera</i> )
	18529.12	NYTH18464-RA	WAXY Granule-bound starch synthase 1( <i>Antirrhinum majus</i> )
	13610.54	NYTH45457-RA	HSP22 Small heat shock protein, ( <i>Glycine max</i> )
	9038.013	NYTH16912-RA	Non-specific lipid-transfer protein 1 ( <i>Lens culinaris</i> )
	8806.638	NYTH28556-RA	Protein of unknown function
	8389.802	NYTH59911-RA	Alpha-terpineol synthase, chloroplastic ( <i>Magnolia grandiflora</i> )
	6565.69	NYTH03818-RA	Ubiquitin ( <i>Triticum aestivum</i> )
6242.847	NYTH59528-RA	TPS10 Terpene synthase 10 ( <i>Ricinus communis</i> )	

704

705 **Table 1: TPM, Identifier, and Putative homology of the 10 transcripts with the highest abundances at**

706 **each stage.** Putative homology information was collected from the annotated genome of *N. thermarum*

707 (Povilus 2020).



708

709 **Figure 3: Differential Expression of Transcripts During Seed Development in *N. thermarum*.**

710 Basic analysis of differentially expressed (DE) transcripts. A) Heatmap of the relative expression (Z-scores

711 of the mean TPM of all biological replicates at each stage) for each of 10,933 DE transcripts. Each row

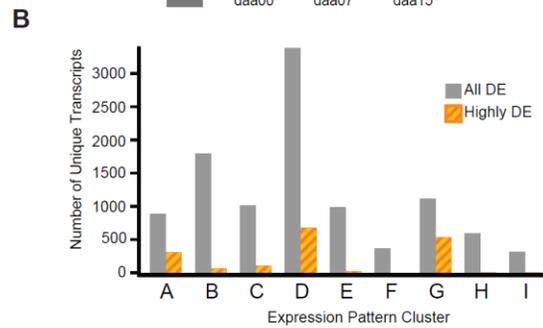
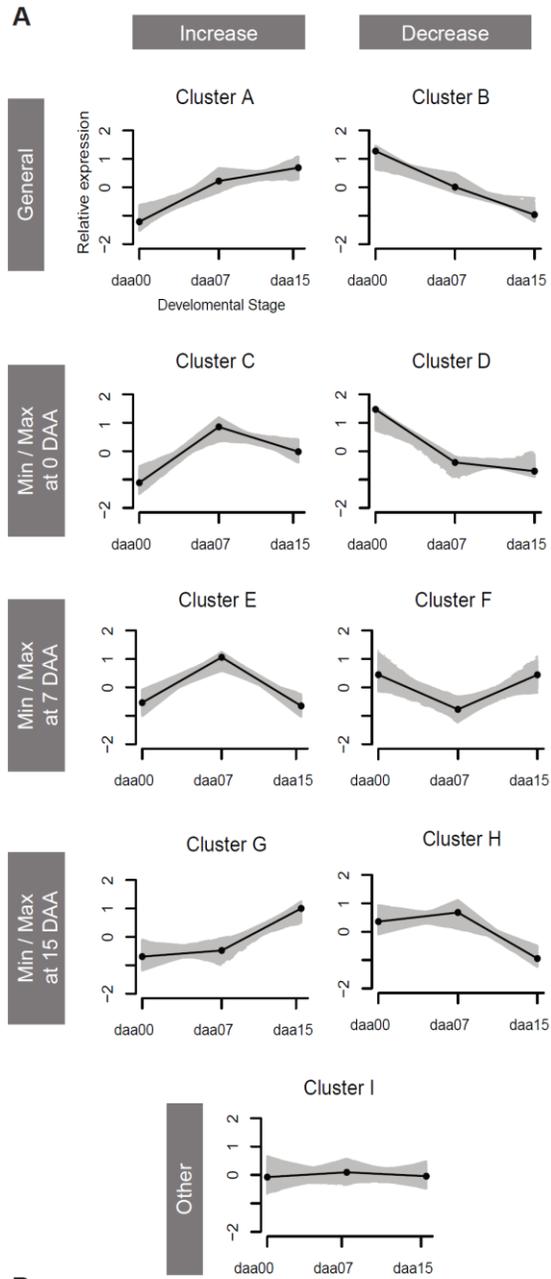
712 represents a single unique transcript; transcript identifies are not included. Rows are hierarchically

713 clustered (dendrogram not included). Each column represents a single sample, and are hierarchically

714 clustered (top dendrogram). B) Number of unique, DE transcripts that showed either an increase or

715 decrease in expression between time points. Results from the set of all-DE or highly-DE transcripts are

716 shown separately.



718 **Figure 4: Expression Pattern Clusters for DE Transcripts**

719 A) 9 expression pattern clusters, which contain 96% of all transcripts differentially expressed during the  
720 sampled stages of reproductive development. Mean TPM values of biological replicates at each stage  
721 were centered and scaled, relative to the mean transcript expression value over all stages. Clusters are  
722 organized by whether they represent initial increases or decreases (columns) to achieve consistent  
723 increased or decreased expression, or minimum or maximum expression at each stage (rows). Grey  
724 areas represent expression of each transcript in a cluster (only includes transcripts whose expression  
725 patterns correlated with the average profile of each cluster (Pearson correlation > 0.9)), while black lines  
726 represent the median expression pattern for each cluster. B) Number of unique transcripts in each  
727 cluster. Results from sets of all-DE and highly-DE transcripts are shown separately.

728

	Cluster	All DE/ highly DE	All DE transcripts	Highly DE transcripts
Increased Expression	A	909	<ul style="list-style-type: none"> <li>Amino acid transmembrane transporter activity</li> <li>Phosphate transmembrane transporter activity</li> <li><b>P-P-bond-hydrolysis protein transmembrane transporter</b></li> <li>Oxidoreductase, acting on CH-CH group of donors</li> <li>Transaminase activity</li> <li>Vitamin binding</li> <li>Fe-S cluster binding</li> <li>Metal ion binding</li> <li>Lyase activity</li> <li>Protein binding</li> </ul>	<ul style="list-style-type: none"> <li>Amino acid transmembrane transporter activity</li> <li>Transferase activity, acyl groups other than amino-acyl</li> <li>Cation binding</li> <li>Oxidoreductase activity</li> </ul>
		316 (35%)		
	C	1062	<ul style="list-style-type: none"> <li>Translation elongation factor activity</li> <li>Cu ion binding</li> <li>Fe-S cluster binding</li> <li>Structural constituent of ribosome</li> <li>Disulfide oxidoreductase activity</li> <li>Hydro-lyase activity</li> <li>Threonine-type endopeptidase activity</li> </ul>	<i>(chi square test)</i> <ul style="list-style-type: none"> <li>Lipid binding</li> </ul>
		105 (6%)		
	E	1036	<ul style="list-style-type: none"> <li>Translation elongation factor activity</li> <li>Structural constituent of ribosome</li> <li>NAD or NADH binding</li> <li>GTP binding</li> <li><b>P-P-bond-hydrolysis protein transmembrane transporter activity</b></li> <li>Nucleobase, nucleoside, nucleotide kinase activity</li> <li>acyltransferase activity</li> <li>Phosphotransferase activity, phosphate group as acceptor</li> <li>GTPase activity</li> <li>Ligase activity, forming C-S bonds</li> <li>Metalloendopeptidase</li> </ul>	<i>(chi square test)</i> <ul style="list-style-type: none"> <li><b>Transcription factor activity</b></li> </ul>
22 (2%)				
G	1129	<ul style="list-style-type: none"> <li>Chlorophyll binding</li> <li>Cu ion binding</li> <li>Heme binding</li> <li>Cofactor binding</li> <li>Lipid binding</li> <li><b>Transition metal ion transmembrane transporter activity</b></li> <li><b>K ion transmembrane transporter activity</b></li> <li><b>ATPase activity, transmembrane movement of substances</b></li> <li><b>symporter activity</b></li> <li>Hydrolase activity</li> <li>NADH dehydrogenase (ubiquinone) activity</li> <li>Lyase activity</li> <li>Electron carrier activity</li> </ul>	<ul style="list-style-type: none"> <li>Chlorophyll binding</li> <li>Oxygen binding</li> <li>Heme binding</li> <li><b>Metal ion transmembrane transporter activity</b></li> <li><b>Hydrogen ion transmembrane transporter activity</b></li> <li><b>Secondary active transmembrane transporter</b></li> <li><b>ATPase, transmembrane movement of substances</b></li> <li>Electron carrier activity</li> <li>Lyase activity</li> <li>Monooxygenase activity</li> <li>NADH dehydrogenase (ubiquinone) activity</li> <li>Peroxidase activity</li> </ul>	
	546 (48%)			
Decreased Expression	B	1879	<ul style="list-style-type: none"> <li>Mismatched DNA binding</li> <li>Damaged DNA binding</li> <li>Zn ion binding</li> <li>ATP binding</li> <li><b>S-adenosylmethionine-dependent methyltransferase</b></li> <li><b>N-methyltransferase activity</b></li> <li>DNA-directed DNA polymerase activity</li> <li>DNA-directed RNA polymerase activity</li> <li>Protein serine/threonine kinase activity</li> <li>Small conjugating protein-specific protease activity</li> <li>3'-5' exonuclease activity</li> <li>Hydrolase activity, C-N (not peptide) bonds</li> <li>ATP-dependent DNA helicase activity</li> </ul>	<i>(chi square test)</i> <ul style="list-style-type: none"> <li>DNA binding</li> </ul>
		69 (4%)		
	D	3537	<ul style="list-style-type: none"> <li>Microtubule motor activity</li> <li>Microtubule binding</li> <li>DNA helicase activity</li> <li>DNA-dependent ATPase activity</li> <li>Hydrolase activity, O-glycosyl compounds</li> <li>UDP-glycosyltransferase activity</li> <li>Transferase activity, hexosyl groups</li> <li>DNA-directed DNA polymerase activity</li> <li>Protein serine-threonine kinase activity</li> <li>Protein tyrosine kinase activity</li> <li>Transmembrane receptor protein kinase activity</li> <li><b>Transcription factor activity</b></li> <li><b>Sequence-specific DNA binding</b></li> <li>Protein self-association</li> <li>Protein homodimerization activity</li> <li>Identical protein binding</li> <li>Chromatin binding</li> <li>ATP binding</li> </ul>	<ul style="list-style-type: none"> <li>Signal transducer activity</li> <li><b>Transcription factor activity</b></li> <li>Protein kinase binding</li> <li>Heme binding</li> <li>Cyclin-dependent protein kinase regulator activity</li> <li>Protein serine-threonine kinase activity</li> <li>Monooxygenase activity</li> <li>ATP binding</li> <li><b>ATPase activity, transmembrane movement of substances</b></li> <li>Oxygen binding</li> <li>Electron carrier activity</li> </ul>
		715 (20%)		
	F	234	<ul style="list-style-type: none"> <li>(none)</li> </ul>	<ul style="list-style-type: none"> <li>(none)</li> </ul>
0 (0%)				
H	605	<ul style="list-style-type: none"> <li>Catalytic activity</li> <li>Nucleotide binding</li> </ul>	<i>(chi square test)</i> <ul style="list-style-type: none"> <li>Hydrolase activity</li> </ul>	
	9 (1%)			
Other	I	59	<i>(chi square test)</i> <ul style="list-style-type: none"> <li>Ligase activity</li> </ul>	<i>(chi square test)</i> <ul style="list-style-type: none"> <li>(none)</li> </ul>
		1 (2%)		

730 **Figure 5: Summary of putative molecular functioned enriched in each expression pattern cluster.**

731 All significantly enriched child terms (ie: the most specialized of a hierarchy) are reported for each

732 cluster. Unless otherwise noted, molecular function enriched was tested with hypergeometric test, using

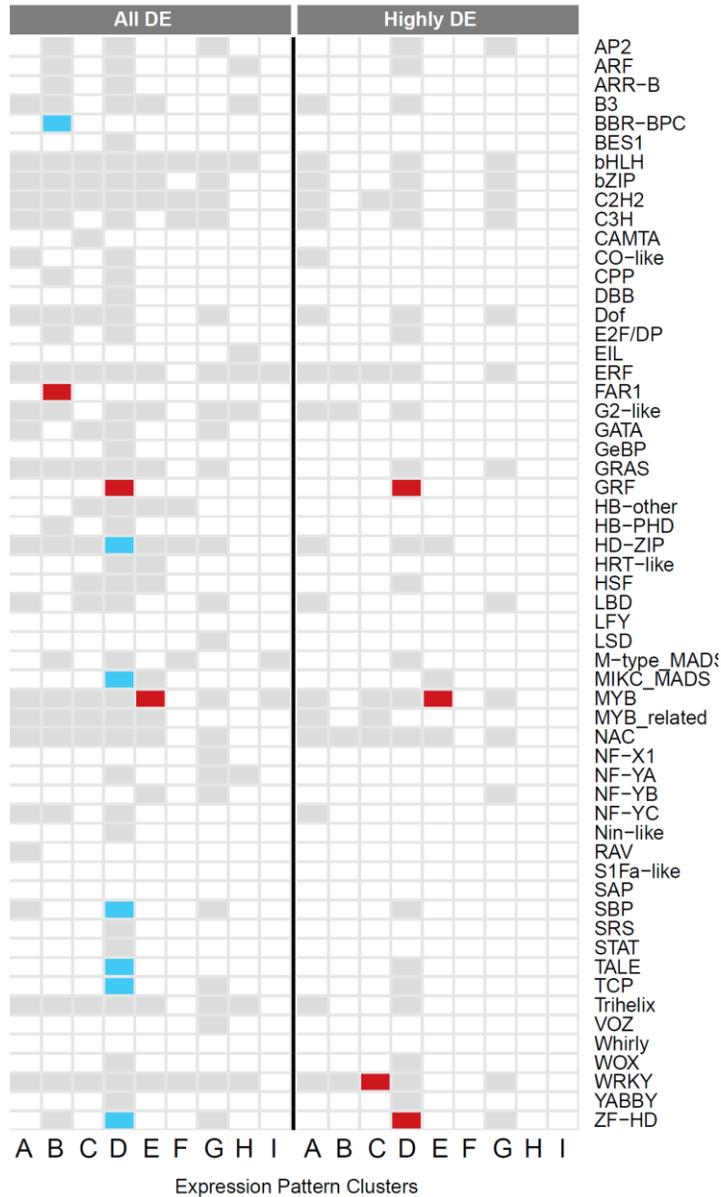
733 Yekutieli (FDR under dependency) multiple corrections testing adjustment, and significance level = 0.1.

734 Molecular function in bold indicate functions of particular interest during discussion.

735

736

737



738

739 **Figure 6: Enrichment of differentially-expressed transcription factor families.**

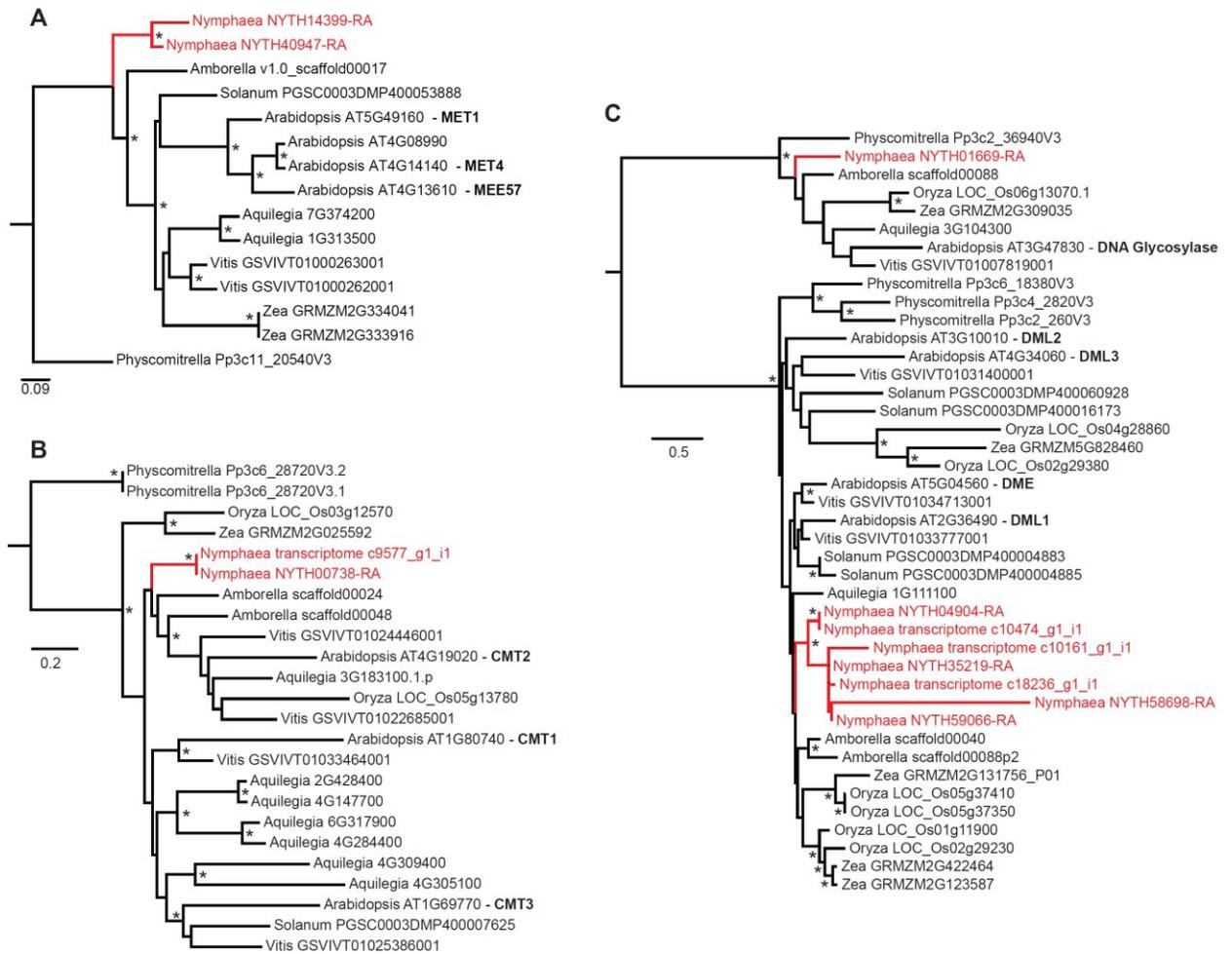
740 Enrichment analysis for TF families among the set of DE TFs in each expression cluster, performed with

741 Fisher's exact test; adjusted p-values (FDR) <0.1 (light blue) and <0.05 (dark red) are noted. Boxes in grey

742 indicate at least one member of a TF family is present in an expression cluster, white indicates that no

743 member of a TF family is present. Results for TFs from the sets of all-DE and highly-DE transcripts are

744 reported separately.



745

746 **Figure 7: Gene family evolution for imprinting-related genes involved in the regulation of DNA**  
 747 **methylation patterns.**

748 All gene family phylogenies were calculated with RAxML, from trimmed amino acid alignments.

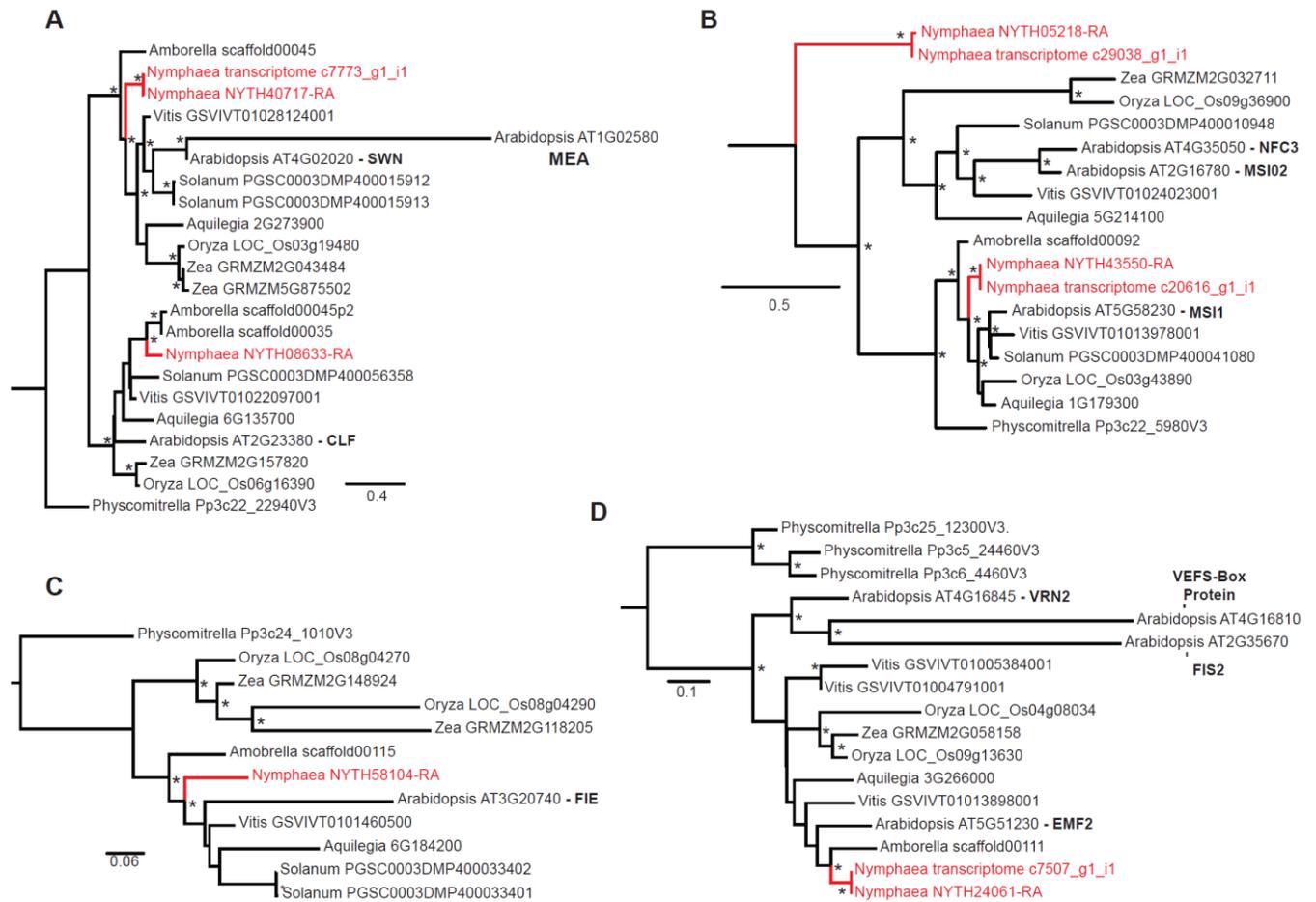
749 Bootstrap support (n=100) > 0.75 indicated by an asterisk. For each included sequence, the organism

750 (genus name) and transcript identifier are noted. The gene names for *Arabidopsis* copies of interest are

751 included in bold text. *Nymphaea* sequences (from transcriptome and genome assemblies) and sequence

752 lineages are colored red. A) MET gene family. B) CMT gene family. C) DME (and DML) gene family.

753



754

755 **Figure 8: Gene family evolution for PRC2 components.**

756 All gene family phylogenies were calculated with RAxML, from trimmed amino acid alignments.

757 Bootstrap support ( $n=100$ ) > 0.75 indicated by an asterisk. For each included sequence, the organism

758 (genus name) and transcript identifier are noted. The gene names for *Arabidopsis* copies of interest are

759 included in bold text. *Nymphaea* sequences and sequence lineages are colored red. A) MEA (and CLF)

760 gene family. B) MSI1 (and MSI02, NFC3) gene family. C) FIE gene family. D) FIS2 (and VRN2 and EMF2)

761 gene family.

762



764 **Figure 9: Expression of *N. thermarum* transcripts putatively involved in DNA and histone methylation.**

765 Transcripts putatively related to imprinting are noted separately from all other transcripts putatively  
766 involved in DNA or histone methylation. For each transcript, the following information is included (from  
767 left to right): gene family the transcript is associated with or the gene name of the most closely-related  
768 *Arabidopsis* homolog; whether the transcript is present in a DE expression pattern cluster; transcript  
769 identifier; expression at each of the sampled stages of reproductive development.

770

	<i>Nymphaea</i>	Monocots (mostly <i>Oryza</i> )	<i>Arabidopsis</i>	Summary
CMT	<b>1 copy</b> • Decreases	<b>2 copies</b> • 1 copy decreases • 1 copy increases, then decreases (Sharma 2006)	<b>3 copies</b> • 1 copy decreases (Sharma 2006)(Julien 2012)	Fairly consistent expression patterns, for the copies expressed during seed development.
MET	<b>2 copies</b> • Both decrease	<b>2 copies</b> (Sharma 2009) • 1 copy constant • 1 copy increases, then decreases	<b>4 copies</b> • 1 copy increases (Sharma 2006)(Julien 2012)	Lineage-specific duplications. <i>Nymphaea</i> copies have opposite expression patterns than what is seen in <i>Arabidopsis</i> and <i>Oryza</i> .
DME*	<b>4 copies</b> • 3 copies decrease • 1 copy increases, then decreases	<b>6 copies?</b> • Most copies decrease (Jiang 2016) • In barely, one MET copy increases (Kapazoglou 2013)	<b>4 copies</b> • 1 copy decreases (Choi 2002)	Lineage-specific duplication in <i>Nymphaea</i> and in monocots. <i>Nymphaea</i> expression resembles certain homologs in <i>Arabidopsis</i> , or in some monocots.
MEA*	<b>2 copies</b> • Both decrease	<b>1 copy</b> • Low in egg cell (Anderson 2013) • Little change in endosperm (Nallamilli 2013)	<b>3 copies</b> • 1 copy changes little (Sharma 2006)	<i>Nymphaea</i> expression decreases, while in <i>Arabidopsis</i> and rice, expression changes little.
FILE	<b>1 copy</b> • Decreases (non-significant)	<b>2 copies</b> • Both copies high in egg cell (Anderson 2013) • 1 copy increases in endosperm (Nallamilli 2013) • 1 copy changes little in endosperm	<b>1 copy</b> • Increases, then decreases (Baroux 2006)	<i>Nymphaea</i> expression decreases (non-significantly), while expression in <i>Arabidopsis</i> and rice increases. Lineage-specific duplication in <i>Oryza</i> may have led to functional divergence.
FIS2*	<b>1 copy</b> • Decreases (non-significant)	<b>2 copies</b> • Both copies high in egg cell (Anderson 2013) • Both copies change little in endosperm (Nallamilli 2013)	<b>4 copies</b> • 1 copy decreases (Baroux 2006)	<i>Nymphaea</i> expression is similar to the one <i>Arabidopsis</i> copy important for seed development.
MSI1*	<b>1 or 2 copies</b> • Decreases	<b>1 copy</b> • High in egg cell (Anderson 2013)	<b>1 copy</b> • Decreases (Baroux 2006)	Fairly consistent expression patterns, for the copies expressed during seed development.

772 **Figure 10: Summary and comparison of imprinting-related methylation regulators in *Nymphaea*,**  
773 **monocots (mostly *Oryza*), and *Arabidopsis*, and their expression before and after fertilization.**

774 For each gene family of interest, the number of copies in each species is reported, as well a brief  
775 summary of their relative expression before and after fertilization. An asterisk next to the gene family  
776 name indicates that a broad definition of gene family was used when assessing copy number (for  
777 example, the MEA\* gene family includes both the MEA and CLF subfamilies).