A variably imprinted epiallele impacts seed development

Daniela Pignatta[1¶,#], Katherine Novitzky[1¶], P.R. V. Satyaki[1], Mary Gehring[1,2*]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, United States of America

[2]Dept of Biology, Massachusetts Institute of Technology, Cambridge, MA, United States of America

*Corresponding author

Email: mgehring@wi.mit.edu (MG)

¶ These authors contributed equally to this work.

Current Address: Indigo Agriculture, Charlestown, MA

Abstract

The contribution of epigenetic variation to phenotypic variation is unclear. Imprinted genes, because of their strong association with epigenetic modifications, represent an opportunity for the discovery of such phenomena. In mammals and flowering plants, a subset of genes are expressed from only one parental allele in a process called gene imprinting. Imprinting is associated with differential DNA methylation and chromatin modifications between parental alleles. In flowering plants imprinting occurs in a seed tissue – endosperm. Proper endosperm development is essential for the production of viable seeds. We previously showed that in *Arabidopsis thaliana* intraspecific imprinting variation is correlated with naturally occurring DNA methylation polymorphisms. Here, we investigated the mechanisms and function of allele-specific imprinting of the class IV homeodomain-Leucine zipper (HD-ZIP) transcription factor *HDG3*. In imprinted strains, *HDG3* is expressed primarily from the methylated paternally inherited allele. We manipulated the methylation state of endogenous *HDG3* in a non-imprinted strain and demonstrated that methylation of a proximal transposable element is sufficient to promote *HDG3* expression and imprinting. Gain of *HDG3* imprinting was associated with earlier endosperm cellularization and changes in seed weight. These results indicate that epigenetic variation alone is sufficient to explain imprinting variation and demonstrate that epialleles can underlie variation in seed development phenotypes.

Author Summary

The contribution of genetic variation to phenotypic variation is well-established. By contrast, it is unknown how frequently epigenetic variation causes differences in organismal phenotypes. Epigenetic information is closely associated with but not encoded in the DNA sequence. In practice, it is challenging to disentangle genetic variation from epigenetic variation, as what appears to be epigenetic variation might have an underlying genetic basis. DNA methylation is one form of epigenetic information. *HDG3* encodes an endosperm specific transcription factor that exists in two states in *A. thaliana* natural populations: methylated and expressed and hypomethylated and repressed. We show that pure epigenetic variation is sufficient to explain expression variation of *HDG3* – a naturally lowly expressed allele can be switched to a higher expressed state by adding DNA methylation. We also show that expression of *HDG3* in strains where it is normally hypomethylated and relatively repressed causes a seed development phenotype. These data indicate that naturally circulating epialleles have consequences for seed phenotypic variation.

54

Introduction

56 DNA methylation is a heritable epigenetic mark that can, on occasion, effect gene transcription and influence development. DNA methylation is a particularly 58 influential regulator of gene expression in endosperm, a triploid extraembryonic seed tissue that supports embryo development. In endosperm, developmentally programmed 60 DNA demethylation causes maternally inherited endosperm genomes to be hypomethylated compared to the paternally inherited endosperm genome (1-3). 62 Methylation differences between maternal and paternal alleles identify their parent-of-origin and establish imprinting, an epigenetic phenomenon in which a gene is expressed 64 primarily from one parental allele (4). Imprinting is theorized to have evolved over conflict between maternally and paternally inherited alleles in offspring over the extent of 66 maternal investment (5,6). Under the kinship theory, silencing of the maternally inherited allele and expression of the paternally inherited allele is predicted to ultimately result for 68 genes where the paternally inherited allele's optimum expression level in offspring is higher than the maternally inherited allele's, (7). Comparison of imprinting between 70 species in the Arabidopsis genus has provided empirical support for this hypothesis (8,9).

72 Recent genomic approaches have revealed extensive natural DNA methylation variation within *Arabidopsis thaliana* (10,11). Whereas the contribution of genetic 74 variation to phenotypic diversity is well-established, the impact of epigenetic variation, or epialleles, on phenotype is only beginning to be understood (12,13). Processes affected 76 by epialleles include patterns of floral development, sex determination, fruit ripening and nutritional content, and senescence, among others (14-19). We previously demonstrated 78 that natural variation in DNA methylation is associated with imprinting variation, with as many as 10% of imprinted genes estimated to be variably imprinted within *A. thaliana*

80 and maize (20,21). Seed development varies extensively among Arabidopsis accessions and has previously been shown to be influenced by parent-of-origin effects (20,22), thus
82 raising the possibility that variation in imprinting could influence seed phenotypes. One of these variably imprinted genes, *HOMEDOMAIN GLABROUS3* (*HDG3*), is a member
84 of the class IV homedomain-Leucine zipper transcription factor (HD-ZIP) family, which regulates diverse aspects of plant patterning and development (23,24). Studies on the
86 function of class IV HD-ZIP genes in trichome differentiation, sepal giant cell formation, and suppression of somatic embryogenesis, among others, have led to the conclusion
88 that class IV HD-ZIP family genes promote endoreduplication and cell differentiation (24-27). Several members of the class IV HD-ZIPs are primarily expressed in endosperm
90 and exhibit imprinted expression patterns, including *FWA/HDG6*, *HDG8, HDG9* and *HDG3* (2,23,28). *FWA, HDG8,* and *HDG9* are maternally expressed imprinted genes
92 (MEGs), whereas *HDG3* is a paternally expressed imprinted gene (PEG) (2,28). The function of the imprinted class IV HD-ZIP genes during seed development, if any, is
94 unknown.

The activity of *HDG3* alleles is correlated with DNA methylation. In endosperm of
96 imprinted strains, the highly expressed paternal *HDG3* allele is methylated and the lowly expressed maternal allele is hypomethylated over a Helitron TE sequence 5' of the
98 transcriptional start site (2). Maternally inherited endosperm alleles are demethylated by the 5-methylcytosine DNA glycosylase gene *DME*; in *dme* mutants, maternal alleles
100 retain their methylation and are also expressed (2,29). Of 927 Arabidopsis accessions with sufficient methylation data (11), 32 (3.5%) have no methylation in the *HDG3* 5'
102 region and 871 (94%) have greater than 50% methylation. When strains where *HDG3* methylation is low, such as Cvi or Kz_9, are the paternal parent in crosses with Col,
104 there is no methylation difference between maternal and paternal alleles in endosperm and *HDG3* is biallelically expressed (20). Together, these data suggest that (1) DNA

106 demethylation promotes repression of the maternally-inherited *HDG3* allele whereas DNA methylation promotes expression (or inhibits repression) of the paternal *HDG3*
108 allele and that (2) imprinting variation is due to *cis* epigenetic variation at *HDG3* (20). However, a *cis* or *trans* genetic contribution to imprinting variation cannot be excluded
110 because of DNA sequence polymorphisms between the strains and alleles that do and do not exhibit imprinting.

112 Here, we show that a naturally occurring epiallele can contribute to variation in seed phenotypes in Arabidopsis. We tested whether *cis* epigenetic variation is sufficient
114 to explain imprinting variation by generating a methylated *HDG3* Cvi allele that mimicked a methylated *HDG3* Col allele. We found that the *HDG3* Cvi allele switched from a
116 hypomethylated, non-imprinted, repressed state to an imprinted, paternally biased, expressed state. Additionally, gain of *HDG3* imprinting altered endosperm development
118 and final seed size. These data indicate that naturally occurring epialleles can have phenotypic consequences in endosperm, a tissue where methylation is dynamic as a
120 programmed part of development.

122 Results

Natural variation in *HDG3* imprinting is associated with gene expression
124 differences

We previously showed that several genes that are imprinted in endosperm when
126 Col is the paternal parent are not imprinted when Cvi is the paternal parent (20). To further examine naturally occurring endosperm gene expression variation, we
128 sequenced the transcriptomes of endosperm from Col x Col and Col x Cvi F_1 seeds. Comparison of these transcriptomes identified 957 genes that were expressed two-fold
130 or higher in Col x Col and 1187 that were expressed two-fold or higher in Col x Cvi

endosperm (Fig 1A; S1 Table). The gene with the lowest expression in Col x Cvi relative
132 to Col is *HDG3*, which is expressed 64-fold lower in Col x Cvi endosperm (Fig 1A).

We previously reported that *HDG3* is a PEG in Cvi x Col crosses but is
134 biallelically expressed in Col x Cvi (20). To further explore the expression variation of *HDG3*, we performed *in situ* hybridization on developing seeds (Fig 1B-C; S1 Fig). In Col
136 x Col seeds, *HDG3* is expressed specifically in the micropylar, peripheral, and chalazal endosperm, with the highest expression at the heart stage of development (Fig 1C). The
138 same pattern was observed in Cvi x Col (Fig 1C). Whereas *HDG3* expression was detected by in situ hybridization in F_1 endosperm when Col was the paternal parent, it
140 was not detected in endosperm when Cvi was the paternal parent (Fig 1C). Additionally, we performed RT-qPCR on biological triplicates of Col, Cvi, and Col-Cvi F_1 endosperm.
142 Expression in Col x Col and Cvi x Col was approximately 10-fold higher than in Cvi x Cvi or Col x Cvi, indicating that *HDG3* expression is higher when it is imprinted (Fig 1D),
144 consistent with the mRNA-seq (Fig 1A) and *in situ* data (Fig 1C). Thus, although expression is from both maternally and paternally inherited alleles in Col x Cvi crosses
146 (and presumably Cvi x Cvi crosses) as detected by mRNA-seq (20), the total expression in those crosses is lower than when *HDG3* is imprinted. As we previously showed that
148 the Cvi allele is naturally hypomethylated (20), together these results suggest that DNA methylation of the *HDG3* 5' region promotes *HDG3* expression (Fig 1E).

150 There is also evidence for imprinting variation of *HDG3* in other species. In *Arabidopsis lyrata*, expression of *HDG3* is also specific to the endosperm but levels
152 differ between two accessions, MN47 and Kar, and their reciprocal crosses (S2 Fig). In endosperm with high *HDG3* expression (Kar x MN47), expression is strongly paternally
154 biased (76% paternal instead of the expected 33%), whereas in the reciprocal cross expression of *HDG3* is much lower and more reflective of the 2:1 maternal:paternal ratio

156 in the endosperm (79% maternal) (S2 Fig) (8). The correlation between high expression of *HDG3* and paternal allele bias in *A. lyrata* thus mirrors *A. thaliana*.

158

Reduced *HDG3* expression affects seed development

160 To examine if *HDG3* influenced endosperm development, we compared seeds from *hdg3* mutant plants and segregating wild-type siblings in the Col background. We 162 confirmed predominantly paternal expression of *HDG3* (2,20) by reciprocal crosses between wild type and *hdg3-1* mutants (Fig 2A). When *hdg3* was crossed as a female to 164 a wild-type sibling male, expression of *HDG3* was detected in endosperm in a similar manner as in Col x Col (Fig 2A). In contrast, when wild-type females were crossed to 166 *hdg3-1* mutant males, the accumulation of *HDG3* transcript in endosperm was dramatically affected, with no transcript detected in most cases, despite the presence of 168 a wild-type maternally inherited allele (Fig 2A). We assessed embryo stage and the extent of endosperm cellularization for sectioned wild-type and *hdg3* seeds at 5 days 170 after pollination. Embryo development was more variable in *hdg3*, although this difference was not statistically significant, but endosperm cellularization was significantly 172 delayed compared to wild-type seeds (Fig 2B, S2 and S3 Table). Reciprocal crosses between wild-type and *hdg3* mutant plants indicated that the endosperm cellularization 174 phenotype was dependent on paternal genotype, consistent with *HDG3* function being primarily supplied from the paternally-inherited allele (S2 and S3 Table). Additionally, 176 the weight and area of *hdg3* seeds was slightly reduced compared to Col, suggesting that in the Col background *HDG3* promotes seed growth or filling (Fig 2C-D). Several 178 PEGs have been shown to influence seed abortion phenotypes in interploidy crosses (30,31), but we found no effect of *hdg3* on this process (S3 Fig). To understand the 180 potential molecular consequences of the loss of *hdg3,* we profiled endosperm gene expression in wild-type Col and *hdg3-1* by RNA-seq at 7 days after pollination (DAP)

182 (Fig 3). 150 genes had at least two-fold higher expression upon loss of *hdg3*, while 238 genes had at least two-fold lower expression in *hdg3* mutant endosperm (Fig 3, S4
184 Table). Differentially expressed genes included developmental regulators such as *Homeobox 3* (*WOX9*) and gibberellin oxidases, which effect the level of a key
186 phytohormone necessary for typical seed development (32) (Fig 3). The loss of *hdg3* also impacted the expression of ten imprinted genes, including the MEG *HDG9* (Fig 3).
188 We hypothesized that the endosperm gene expression phenotypes associated with low expression of *HDG3* from Cvi paternal alleles might in some respects mimic *hdg3*
190 mutants. Indeed, of the 238 genes that are down-regulated in *hdg3* mutants, 100 are also down regulated in Col x Cvi crosses, where *HDG3* expression is also low (Fig 3).
192 This is a highly significant overlap (hypergeometric test in R, p = 6.079e-69) (Fig 3). These data suggest that the Cvi *HDG3* allele, in its hypomethylated and relatively
194 transcriptionally repressed state, could be important for some of the accession-specific developmental traits imparted by Cvi (20,22,33). Thus, to test both the imprinting
196 mechanism and function of *HDG3* further, we introduced methylation at the *HDG3* locus in Cvi.
198

An inverted repeat induces methylation in the region 5' of *HDG3* in Cvi

200 To distinguish the importance of genetic variation from epigenetic variation for *HDG3* expression and imprinting, we generated transgenic lines in which the
202 endogenous *HDG3* Cvi allele gained methylation in the same region that is methylated in Col. Cvi was transformed with a transgene consisting of an inverted repeat (*HDG3* IR) of
204 the 450 bp *HDG3* 5' region from Cvi under the control of the constitutive 35S promoter. Processing of the expressed hairpin RNA into small RNAs is expected to direct
206 methylation to the endogenous *HDG3* Cvi locus. We identified multiple independent transgenic lines in which the *HDG3* 5' region gained methylation in leaves (S4 Fig). DNA

208 methylation was present in the same region as in Col, although non-CG methylation was considerably higher (S4 Fig).

210 To determine whether the Cvi allele remained methylated when paternally inherited in endosperm, Cvi *HDG3* IR plants from three independent transgenic lines

212 were crossed as males to wild type Col females and DNA methylation was evaluated in F_1 endosperm by locus-specific bisulfite-PCR. Although the 35S promoter has been

214 reported to have no activity in syncytial endosperm (34), we detected transcripts from the hairpin RNA in endosperm at 7 DAP (S5 Fig). Bisulfite sequencing showed that the

216 paternally inherited *HDG3* Cvi allele from the IR line was hypermethylated relative to the paternally inherited *HDG3* Cvi allele in Col x Cvi endosperm (Fig 4; S6 Fig). The *HDG3*

218 Cvi allele from Col x Cvi *HDG3* IR endosperm was methylated in both CG and non-CG contexts, indicative of RNA-directed DNA methylation, although at a lower level than in

220 leaves or in F_1 embryos (S6 Fig). Examination of the bisulfite clones indicated some variation in paternal allele methylation, with clones with 0% methylation detected, unlike

222 naturally methylated paternal alleles from Cvi x Col crosses (S6 Fig). This could be due to stochastic silencing of the IR transgene in individual siliques/seeds or ineffective RNA-

224 directed DNA methylation. The maternally inherited Col allele was unaffected in Col x Cvi *HDG3* IR endosperm, remaining hypomethylated like in Col x Cvi endosperm. Thus,

226 we successfully established an alternate epigenetic state specifically for the Cvi *HDG3* allele in endosperm.

228

Methylation of the *HDG3* 5' region is sufficient to promote expression and

230 **imprinting**

Having established a methylated Cvi *HDG3* allele, we tested whether paternal allele

232 methylation was sufficient to switch *HDG3* from a non-imprinted, repressed state to an

imprinted, more active state. In two independent lines, *in situ* hybridization of F_1 seeds
234 from Col x Cvi *HDG3* IR crosses indicated the presence of *HDG3* transcript in
endosperm, in contrast to Col x Cvi endosperm (Fig 5A). Hybridization signal was
236 primarily detected in uncellularized endosperm on the chalazal side of the peripheral
endosperm (Fig 5A). However, the penetrance of Cvi *HDG3* expression was variable,
238 with about half of the seeds exhibiting *HDG3* expression detectable by *in situ* (Fig 5A).
This might be related to the variation in methylation of the *HDG3* Cvi allele in Col x Cvi
240 *HDG3* IR seeds (S6 Fig). Analysis of total *HDG3* transcript abundance by RT-qPCR at
6-7 days after pollination showed that *HDG3* expression was 2-3-fold higher in Col x Cvi
242 *HDG3* IR endosperm compared to Col x Cvi endosperm (Fig 5B). Higher expression of
HDG3 in Col x Cvi *HDG3* IR endosperm is consistent with *HDG3* being more highly
244 expressed when imprinted (Fig 1). Thus, to measure allele-specific expression of *HDG3*,
Col and Cvi alleles were distinguished using TaqMan probes in an RT-qPCR assay. In
246 crosses between Col females and three independent Cvi *HDG3* IR lines, the fraction of
transcript derived from the Cvi allele increased compared to control crosses between Col
248 females and Cvi males. In Col x Cvi, the Cvi allele accounts for 23% of the transcripts by
this assay, in good agreement with prior allele-specific mRNA-seq results (20). In Col x
250 Cvi *HDG3* IR lines, the Cvi fraction was between 50-60%, indicating paternal allele bias
(the expectation for non-imprinted genes is 33% paternal) (Fig 5C). This is slightly less
252 than the fraction of paternal allele expression in Cvi x Col crosses by mRNA-seq (79%)
(20). Together, these data indicate that the naturally occurring methylation variation at
254 *HDG3* is sufficient to explain imprinting variation. We conclude that the methylated Cvi
HDG3 allele in Cvi *HDG3* IR plants is active and the gene is imprinted.
256

Expression of *HDG3* in Cvi promotes endosperm cellularization

258 Does the change of *HDG3* expression and imprinting in Cvi affect seed development? To test the phenotypic consequences of expressing *HDG3* from
260 previously repressed Cvi alleles, we compared the phenotypes of seeds from Col x Cvi (low *HDG3* expression) and Col x Cvi *HDG3* IR (2-4-fold increased *HDG3* expression)
262 seeds by sectioning and staining (Fig 6). In crosses with the *HDG3* IR lines, endosperm cellularization occurred at a significantly earlier stage of embryo development, where it
264 was observed as early as the globular stage of embryogenesis (Fig 6A-B, S2 and S3 Table). Whereas endosperm development appeared accelerated, embryo development
266 was significantly delayed (Fig 6B, S2 and S3 Table). The effect on endosperm cellularization was also observed in Cvi x Cvi *HDG3* IR F_1 seeds, although to a lesser
268 extent (S2, S3 Table). Mature selfed seeds from Cvi *HDG3* IR plants weighed significantly less than selfed seeds from Cvi and had reduced area (Fig 6C-D). This is
270 consistent with known correlations between early endosperm cellularization and the production of smaller seeds (35-37). These observations support the hypothesis that
272 hypomethylation and repression of the Cvi *HDG3* allele is important for Cvi-directed developmental programs and that epiallelic variation contributes to the natural variation
274 in seed development in Arabidopsis.

276 Discussion

The establishment, maintenance, and inheritance of DNA methylation are fairly
278 well understood processes. Disruption of methylation patterns by mutations in DNA methyltransferase enzymes have clear gene expression consequences. However,
280 whether or not methylation is regulatory during development – meaning that dynamic loss or gain of methylation is a normal aspect of gene regulation – is less well
282 understood. An exception to this is in the endosperm, where active DNA demethylation

in the female gamete before fertilization establishes differential DNA methylation after
284 fertilization, a step that is essential for normal seed development (38). We thus
hypothesized that the phenotypic impact of naturally occurring epialleles might be
286 particularly evident in the endosperm, because the differential methylation between
maternal and paternal alleles that is required for gene imprinting could be variable
288 across accessions (20). We have shown that *HDG3* represents a case study of this
proposed phenomenon. By placing a methylation trigger in Cvi (the *HDG3* IR transgene),
290 we were able to convert the Cvi *HDG3* allele from a hypomethylated to a methylated
state. This switch in methylation was sufficient to promote expression of the paternally
292 inherited Cvi *HDG3* allele in endosperm to 3-fold higher levels. Because we altered
methylation at the endogenous *HDG3* Cvi locus, which retains all DNA sequence
294 polymorphisms, we have shown that methylation variation alone is sufficient to cause
expression, and thus imprinting, variation. However, our results also show that it is
296 unlikely that methylation of the proximal TE accounts for all of the expression differences
between paternal Col and Cvi *HDG3* alleles in endosperm. The paternally inherited
298 methylated Cvi allele, while more highly expressed than paternally inherited naturally
hypomethylated Cvi allele, was not as highly expressed as paternally inherited
300 methylated Col alleles in endosperm (Figs 1 and 5). Additional *cis* genetic or *trans*
genetic or epigenetic variation likely also affects *HDG3* expression levels. Finally, it is
302 not possible to determine from the experiments presented here whether the original
difference in methylation between naturally methylated and non-methylated alleles lacks
304 any genetic basis. Cvi lacks the small RNAs associated with the 5' TE that are found in
many other accessions, but the ultimate cause of this difference remains unknown (S4
306 Fig).

Our experiments also shed light on the relative receptiveness of maternal and
308 paternal endosperm genomes to *de novo* methylation. The *HDG3* inverted repeat

transgene should create endosperm small RNAs that are homologous to both Col and
310 Cvi alleles (there are only 4 SNPs and a 3 bp indel between Col and Cvi in the IR target region). Yet, in endosperm from Col x Cvi *HDG3* IR crosses, the paternally inherited Cvi
312 allele had high levels of non-CG methylation, whereas the maternally inherited Col alleles remained hypomethylated despite the presence of the IR transcript (Fig 4, S5 Fig,
314 S6 Fig). In contrast, F_1 embryos from the same crosses were indeed more highly methylated in the non-CG context on maternal Col alleles compared to maternal Col
316 alleles from Col x Cvi crosses (S6 Fig). Thus, maternally inherited *HDG3* alleles in endosperm are refractory to *de novo* methylation even when a methylation trigger is
318 present, in contrast to maternally inherited *HDG3* alleles in embryos. These results further support findings that once a region is actively demethylated on the maternally
320 inherited endosperm genome, it is "protected" from *de novo* methylation even when triggering small RNAs are present (31).

322 Finally, although the direct targets of the *HDG3* transcription factor are still unknown, we have shown that natural variation in *HDG3* expression (expressed in Col,
324 low expression in Cvi) has consequences for seed gene expression programs and development (Figs 2, 3, 6 and 7). Expression of *HDG3* in seeds fathered by Cvi caused
326 dramatically early endosperm cellularization and the seeds were smaller and lighter at maturity (Fig 6). These findings are consistent with class IV HD-ZIP genes inhibiting the
328 cell cycle and promoting cellular differentiation (24,27). However, mutation of *hdg3* in Col, while displaying the predicted opposite effect on endosperm cellularization timing,
330 also resulted in smaller seeds weighing slightly less than wild-type (Figs 2, 6 and 7). Although the effects on final seed size are seemingly contradictory and the physiological
332 basis remains incompletely understood, these results are predicted under the aegis of the kinship theory (7). The theory predicts that PEGs promote maternal investment in
334 offspring, which is consistent with the effects of the *hdg3* mutation in Col (i.e. less

maternal investment results in smaller seeds). Our results suggest that this effect is specific to a Col seed developmental program. In Cvi endosperm, expression of *HDG3* is seemingly maladaptive, leading to the production of smaller seeds. Cvi naturally produces much larger seeds than Col or L*er*, although fewer in number (20,22,33) (Figs 2 and 6). Our results suggest that the loss of *HDG3* expression in Cvi was an important part of the process that resulted in these phenotypic differences.

In summary, we have demonstrated that seed phenotypic differences can be caused by methylation differences at single genes. This study provides further evidence that epigenetic differences underlie developmental adaptations in plants. We have previously shown that the imprinting status of many genes varies between accessions; our current study argues that intraspecific variation in imprinting is an important determinant of seed developmental variation.

Materials and Methods

Plant material

The SALK insertion mutant was obtained from the Arabidopsis Biological Resource Center (39). *hdg3-1* (SALK_033462) was backcrossed to Col-0 three times before experimentation. For experiments comparing or crossing wild-type and *hdg3* mutant plants, plants were F_3 segregants from selfed progeny of *HDG3/hdg3-1*. Plants were grown in a growth chamber or greenhouse with 16-hour days at 22° C. For crosses, flowers were emasculated and then pollinated after 2 days.

In situ hybridization

Controlled floral pollinations were performed for each specified cross. At least two independent *in situ* experiments were performed for each genotype. Siliques were

360 harvested 5 or 6 days after pollination (DAP) and fixed in FAA overnight at 4°C.
Following dehydration and clearing (HistoClear, National Diagnostics), samples were
362 embedded in Paraplast Plus (McCormick Scientific), and sectioned at 9 μM (Leica RM
2065 rotary microtome). Ribbons were mounted with DEPC water on ProbeOn Plus
364 slides (Fisher) at 42°C and dried overnight at 37°C. For probes, a 278 bp region of
HDG3 (S5 Table) and previously published 602 bp probe for *PDF1* (40) were amplified
366 from endosperm cDNA and cloned into P-GEM T vectors (Promega). Plasmids
containing sense and antisense oriented fragments were identified and linear templates
368 were amplified using M13 forward and reverse primers for probe synthesis. Antisense
and sense RNA probes were synthesized in vitro with digoxigenin-UTPs using T7 or SP6
370 polymerase (DIG RNA labeling kit, Roche/Sigma Aldrich). Probes were subsequently
hydrolyzed and dot blots were performed to estimate probe concentration. Pre-
372 hybridization steps were preformed according to (41) except Pronase digestion occurred
for 15 minutes at 37°C. Hybridization and post-hybridizations were performed according
374 to (42), with minor modifications. For higher confidence in directly comparing expression
patterns, slides corresponding to the cross and its reciprocal were processed face to
376 face in the same pairs for hybridization, antibody, and detection steps. Negative controls
consisted of hybridizing sense probes to wild-type tissue and antisense probes to *hdg3*
378 tissue. The sense probe lacked signal (S1 Fig). A probe to *PDF1*, which is expressed in
the L1 embryo layer (43), served as a positive control for successful *in situ* hybridization
380 (S1 Fig). Hybridization was performed overnight at 55°C, slides were then washed twice
in 0.2X SSC for 60 mins each at 55°C, then twice in NTE for 5 min at 37°C and RNaseA
382 treated for 20 min at 37°C, followed by two more 5 min NTE washes. Slides were
incubated at room temperature for 1 hour with Anti-DIG antibody (Roche/Sigma Aldrich)
384 diluted 1:1250 in buffer A (42). Slides were then washed four times for 20 min each at
room temperature with buffer A and once for 5 min with detection buffer (42).

386 Colorimetric detections were performed using NBT/BCIP Ready-To-Use Tablets (Roche/Sigma Aldrich) dissolved in water. Slides were allowed to develop 16-24 hours
388 before stopping. Slides were dehydrated, mounted in Permount (Electron Microscopy Sciences) and imaged using a Zeiss Axio Imager M2. Minor level adjustments and smart
390 sharpen were applied to images to compensate for image transfer from live to digital (Adobe Photoshop).
392

Seed staining

394 Plant material was fixed and embedded as previously described and sectioned at 9 µm. Slides were dewaxed twice in xylenes for 5 minutes, rehydrated through a graded
396 ethanol/0.85% salt series from 100%-30%, 1 minute each, stained in 0.6% Safranin O Solution (Cat# 2016-03, Sigma Aldrich) for 5 minutes, washed with water, stained with a
398 saturated 2.5% Aniline blue (Harleco–EMD Millipore, #128-12) in 2% glacial acetic acid aqueous solution for 3 minutes, washed with water, rapidly dehydrated though graded
400 ethanol/salt series to 100%, 5 seconds for each step, and then twice in xylenes for 5 minutes each. Slides were briefly drained, cover slipped and mounted with Cytoseal™
402 60 (Thermo Scientific) and imaged using a Zeiss Axio Imager M2.

404 **Seed phenotypic analysis**

Previously processed slides from double staining and *in situ* hybridization experiments
406 were re-examined and used for embryo and endosperm developmental analyses. Using previously published endosperm cellularization and embryogenesis stages (34,44),
408 individual seeds at 5 DAP were scored first for embryo stage and then for respective endosperm stage. Endosperm stage was given a numerical score (-3 to +5) depending
410 on the relative stage of endosperm cellularization compared to the expected endosperm cellularization stage given the embryo stage. Individual seeds with matching

412 embryogenesis and endosperm cellularization stages were scored “normal” and ranged from 0-1; seeds that were scored “early” were defined as being +1.5 to +5 stages further
414 along in the cellularization process compared to normal. Seeds that were scored “delayed” were defined as being -1 to -3 stages behind in the cellularization process
416 compared to normal. To determine whether any developmental differences in endosperm cellularization or embryogenesis were statistically significant, we
418 implemented the asymptotic generalized Pearson chi-squared test from the coin package (45) in R with default scoring weights. Developmental stage was treated as an
420 ordinal variable, while cross genotype was treated as a non-ordinal, nominal variable. Pairwise comparisons were carried out with the R function pairwiseOrdinalIndependence
422 from the rcompanion package. For all tests, embryo development data was collapsed into three categories young (pre-globular to globular), middle (late globular to early
424 heart), and older (heart to torpedo) and detailed endosperm cellularization data was collapsed into the categories delayed, normal, and early.

426

Inverted repeat transgene

428 The 450 bp sequence 5' of *HDG3* corresponding to a fragment of AT2TE60490 from Chr2: 13740010-13740460 was amplified from Cvi (S5 Table) and cloned into the
430 directional entry vector pENTR-TOPO-D (Invitrogen). The sequence was then inserted twice in an inverted repeat conformation into the vector pFGCGW (46) with a LR clonase
432 reaction (Invitrogen). Cvi plants were transformed with the inverted repeat transgene by floral dipping and T_1 lines were screened for DNA methylation using bisulfite sequencing.
434 T_3 plants homozygous for the IR transgene and with a methylated *HDG3* 5’ region in leaves, or their T_4 progeny, were identified and used for subsequent experiments.

436

Quantitative RT-PCR

438 RNA was isolated from endosperm dissected from seeds at 6 or 7 DAP as described (47) using RNAqueous Micro Kit (Ambion, Life Technologies Corporation). DNAse I-
440 treated RNA (Invitrogen, Life Technologies Corporation) was used for cDNA synthesis with oligo-dT primer using Superscript II reverse transcriptase (Invitrogen) according to
442 manufacturers' instructions. Quantitative RT-PCR (RT-qPCR) was performed using Fast Sybr-Green mix or TaqMan Mastermix (Applied Biosystems). All reactions were
444 performed in three or four technical replicates using a StepOne Plus Real-Time PCR system (Applied Biosystems). For Sybr-Green based assays, relative expression was
446 calculated using the ddCt method as described (48). The reference gene was AT1G58050 (49). For allele-specific expression in Col-Cvi crosses, a multiplex TaqMan
448 assay was developed by designing primers and PrimeTime® Double-quenched Custom Probes with online tool http://www.idtdna.com/pages/products/gene-expression/custom-
450 qpcr-probes. Cycling conditions were 15 cycles: 95 °C for 15 seconds, 63 °C for 30 seconds, 72 °C for 30 seconds followed by 25 cycles: 95 °C for 15 seconds, 63 °C for 30
452 seconds with touchdown 0.05°C/cycle, and 72 °C for 30 seconds. The relative expression of each allele within each genotype was calculated using a standard curve
454 (R^2 value >0.99) as reference. Primer and probe sequences are available in S5Table.

456 **Bisulfite sequencing**

Genomic DNA was isolated from leaves, endosperm, and embryo at 6 or 7 days after
458 pollination using a CTAB procedure. Bisulfite treatment was performed using the MethylCode Bisulfite Conversion Kit (Invitrogen, Life Technologies Corporation) or
460 BisulFlash DNA Bisulfite Conversion Easy Kit (Epigentek Group Inc.) following the manufacturer's protocols. 2 µl bisulfite treated DNA was used in PCR reactions with 2.5
462 U ExTaq DNA polymerase (Takara) and 0.4 µM primers using the following cycling conditions (95 °C 3 minutes, 40 cycles of [95 °C for 15 seconds, 50 °C for 30 seconds,

464 72 °C for 45 seconds], 72 °C for 5 minutes). PCR products were gel purified, cloned using a TOPO-TA (Invitrogen) or CloneJet (Life Technologies) PCR cloning kit and
466 individual colonies were sequenced. Sequences were aligned using SeqMan and methylation was quantified using CyMate (50).

468

mRNA-seq

470 RNA was isolated from endosperm of Col-0, *hdg3-1* and Col-0 x Cvi seeds at 7 DAP as described above. Three replicates for each cross were obtained. DNAse treated RNA
472 was used as input for the SmartSeq Clontech Ultralow RNA-Seq kit. Libraries were constructed by the Genome Technology Core at Whitehead Institute. Six libraries were
474 multiplexed per lane in a Hi-Seq 2500 Standard mode, 40 base, single read run. Each replicate was sequenced to a depth of between 33 and 41 million reads. Reads were
476 processed with Trim_galore using the command "*trim_galore -q 25 --phred64 --fastqc --stringency 5 --length 18".* Processed reads were aligned to the TAIR10 genome with
478 Tophat2 (51) using the command "*tophat -i 30 -I 3000 --b2-very-sensitive --solexa1.3-quals -p 5 --segment-mismatches 1 --segment-length 18"*. Differential gene expression
480 was detected with Cuffdiff2 (52) and the ARAPORT11 annotation (S1 and S2 Tables). Reads are deposited in GEO GSE118371.

482

Acknowledgements

484 We thank R. Jaenisch and P. Reddien for sharing equipment for *in situ* hybridization, B. Williams for comments on the manuscript, and R. Povilus for assistance with statistical
486 analysis.

488 **Author Contributions**

Conceptualization: MG and DP. Investigation: KN, DP, PRVS, and MG. Writing – Original Draft and Preparation: MG and PRVS. Writing – Reviewing and Editing: KN, DP, PRVS, MG. Funding Acquisition: MG.

References

1. Gehring M, Huh JH, Hsieh T-F, Penterman J, Choi Y, Harada JJ, et al. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell. 2006; 124(3):495–506.

2. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science. 2009; 324(5933):1447–51.

3. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, et al. Genome-wide demethylation of Arabidopsis endosperm. Science. 2009; 324(5933):1451–4.

4. Satyaki PRV, Gehring M. DNA methylation and imprinting in plants: machinery and mechanisms. Crit Rev Biochem Mol Biol. 2017; 52(2):163–75.

5. Haig D. The kinship theory of genomic imprinting. Annu Rev Ecol Syst. 2000; 31: 9-32.

6. Haig D. Kin conflict in seed development: an interdependent but fractious collective. Annu Rev Cell Dev Biol. 2013; 29:189–211.

7. Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R, Wolf JB. The evolution of genomic imprinting: theories, predictions and empirical tests. Heredity. 2014; 113(2):119–28.

8. Klosinska M, Picard CL, Gehring M. Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nat Plants. 2016; 2(10):16145.

9. Patten MM. Epigenetics: Imprinting evolution in Arabidopsis. Nat Plants. 2016; 2(10):16152.

10. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, et al. Patterns of population epigenomic diversity. Nature. 2013; 495(7440):193–8.

11. Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al. Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. Cell. 2016; 166(2):492–505.

12. Weigel D, Colot V. Epialleles in plant evolution. Genome Biol. 2012; 13(10):249.

13. Aller EST, Jagd LM, Kliebenstein DJ, Burow M. Comparison of the Relative Potential for Epigenetic and Genetic Variation To Contribute to Trait Stability. G3 (Bethesda). 2018; 8(5):1733–46.

524 14. Durand S, Bouché N, Perez Strand E, Loudet O, Camilleri C. Rapid Establishment of Genetic Incompatibility through Natural Epigenetic Variation. Current Biology. 2012; 22(4):326–31.

15. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. Nature. 1999; 401(6749):157–61.

16. He L, Wu W, Zinta G, Yang L, Wang D, Liu R, et al. A naturally occurring epiallele associates with leaf senescence and local climate adaptation in Arabidopsis accessions. Nat Comms. 2018; 9(1):460.

17. Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, et al. A transposon-induced epigenetic change leads to sex determination in melon. Nature. 2009; 461(7267):1135–8.

18. Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, et al. Natural occurring epialleles determine vitamin E accumulation in tomato fruits. Nat Comms. 2014; 5:3027–7.

19. Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet. 2006; 38(8):948–52.

20. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. Elife. 2014; 3:e03198.

21. Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, et al. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proc Natl Acad Sci USA. 2013; 110(48):19639–44.

22. Alonso-Blanco C, Blankestijn-De Vries H, Hanhart CJ, Koornneef M. Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. Proc Natl Acad Sci USA. 1999; 96(8):4710–7.

23. Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, et al. Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. Plant Physiol. 2006; 141(4):1363–75.

24. Robinson DO, Roeder AHK. Themes and variations in cell type patterning in the plant epidermis. Current Opinion in Genetics & Development. 2015; 32:55–65.

25. Abe M, Katsumata H, Komeda Y, Takahashi T. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development. 2003; 130(4):635–43.

26. Roeder AHK, Cunha A, Ohno CK, Meyerowitz EM. Cell cycle regulates cell type in the Arabidopsis sepal. Development. 2012; 139(23):4416–27.

27. Horstman A, Fukuoka H, Muino JM, Nitsch L, Guo C, Passarinho P, et al. AIL and HDG proteins act antagonistically to control cell proliferation. Development. 2015; 142(3):454–64.

28. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, et al. One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. Science. 2004; 303(5657):521–3.

29. Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, et al. Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci USA. 2011; 108(5):1755–62.

30. Wolff P, Jiang H, Wang G, Santos-González J, Köhler C. Paternally expressed imprinted genes establish postzygotic hybridization barriers in Arabidopsis thaliana. Elife. 2015; 4:e10074.

31. Erdmann RM, Satyaki PRV, Klosinska M, Gehring M. A Small RNA Pathway Mediates Allelic Dosage in Endosperm. Cell Rep. 2017; 21(12):3364–72.

32. Singh DP, Jermakow AM, Swain SM. Gibberellins are required for seed development and pollen tube growth in Arabidopsis. Plant Cell. 2002; 14(12):3133–47.

33. Herridge RP, Day RC, Baldwin S, Macknight RC. Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. Plant Methods. 2011; 7(1):3.

34. Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, et al. Dynamic analyses of the expression of the HISTONE::YFP fusion protein in arabidopsis show that syncytial endosperm is divided in mitotic domains. Plant Cell. 2001; 13(3):495–509.

35. Scott RJ, Spielman M, Bailey J, Dickinson HG. Parent-of-origin effects on seed development in Arabidopsis thaliana. Development. 1998; 125(17):3329-41.

36. Garcia D, Saingery V, Chambrier P, Mayer U, Jürgens G, Berger F. Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant Physiol. 2003; 131(4):1661–70.

37. Hehenberger E, Kradolfer D, Kohler C. Endosperm cellularization defines an important developmental transition for embryo development. Development. 2012; 139(11):2031–9.

38. Gehring M, Satyaki PR. Endosperm and Imprinting, Inextricably Linked. Plant Physiol. 2017; 173(1):143–54.

39. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science. 2003; 301(5633):653–7.

40. Kunieda T, Mitsuda N, Ohme-Takagi M, Takeda S, Aida M, Tasaka M, et al. NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in Arabidopsis. Plant Cell. 2008; 20(10):2631–42.

41. JACKSON D. In situ hybridization in plants. Molecular plant pathology: a practical approach. Oxford University Press; 1991.

42. Bortiri E, Chuck G, Vollbrecht E, Rocheford T, Martienssen R, Hake S. ramosa2 encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. Plant Cell. 2006; 18(3):574–85.

43. Tanaka H, Watanabe M, Sasabe M, Hiroe T, Tanaka T, Tsukaya H, et al. Novel receptor-like kinase ALE2 controls shoot development by specifying epidermis in Arabidopsis. Development. 2007; 134(9):1643–52.

44. Berger F. Endosperm: the crossroad of seed development. Current Opinion in Plant Biology. 2003; 6(1):42–50.

45. Hothorn T, Hornik K, van de Wiel MA, Zeileis A. A Lego System for Conditional Inference. The American Statistician. 2012; 60(3):257–63.

46. Zhang C, Galbraith DW. RNA interference-mediated gene knockdown within specific cell types. Plant Molecular Biology. 2012; 80(2):169–76.

47. Gehring M, Missirian V, Henikoff S. Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. PLoS ONE. 2011; 6(8):e23687.

48. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25(4):402–8.

49. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. 2005; 139(1):5–17.

50. Hetzl J, Foerster AM, Raidl G, Mittelsten Scheid O. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. Plant J. 2007; 51(3):526–36.

51. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14(4):R36.

52. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013; 31(1):46–53.

Supporting Information

634 **S1 Fig. Negative and positive controls for *in situ* hybridization.**

S2 Fig. *HDG3* expression in *A. lyrata*.

636 **S3 Fig. Tetraploid *hdg3-1* does not rescue interploidy seed lethality caused by paternal genomic excess.**

638 **S4 Fig. Gain of *HDG3* methylation in Cvi *HDG3* IR lines.**

S5 Fig. Accumulation of inverted repeat RNA in Col x Cvi *HDG3* IR endosperm.

640 **S6 Fig. Bisulfite-PCR analysis of *HDG3* methylation in seeds.**

S1 Table. Gene expression difference in Col x Cvi endosperm vs. Col endosperm

642 **at 7 DAP determined using CuffDiff.**

S2 Table. Endosperm cellularization timing and embryogenesis stages at 5 DAP.

644 **S3 Table. Tests for statistical significance for pairwise comparisons of differences in endosperm or embryo development.**

646 **S4 Table. Gene expression differences in *hdg3-1* endosperm vs. Col endosperm at 7 DAP determined using CuffDiff.**

648 **S5 Table. Oligos used in this study.**

650 Figure Legends

Fig 1. Natural variation in imprinting is associated with differences in *HDG3*

652 **expression levels.**

(A) *HDG3* expression is decreased in Col x Cvi endosperm compared to Col, as

654 determined by mRNA-seq. (B) Schematic of the *HDG3* locus. DMR, differentially methylated region. (C) *In situ* hybridization of *HDG3* (purple) in F_1 seeds from the

656 indicated crosses. Female parent written first. In crosses where Col is the male parent, *HDG3* is detected in the micropylar (MCE), peripheral (PEN), and chalazal (CZE)

658 endosperm. Arrowheads indicate nuclear-cytoplasmic domains. Number of seeds with shown pattern out of total seeds assayed is in corner of each image. Scale bars, 50 μm.
660 (D) RT-qPCR analysis of relative *HDG3* transcript abundance in F_1 endosperm. Values are the average of 3 biological replicates, bars represent upper and lower range. (E)
662 Schematic representation of relationship between *HDG3* methylation, expression, and imprinting in endosperm. Thickness of arrows denotes relative expression level.
664 Lollipops represent methylated (filled) and unmethylated (open) cytosines.

666 **Fig 2. Phenotypic effects of mutation of *HDG3* in Col.**
(A) *In situ* hybridization of *HDG3* (purple) in seeds from the indicated crosses. Scale
668 bars, 50 μm. (B) Endosperm cellularization is slightly delayed in *hdg3* compared to wild-type at 5 DAP. For each seed the embryo stage was determined and then the stage of
670 endosperm cellularization was defined as normal, early, or delayed given that embryo stage. (C) Seed weight of wild-type and *hdg3* seeds. Individual data points and mean +/-
672 SD shown. P-value from unpaired two-tailed t-test. (D) Seed area is significantly reduced in *hdg3* seed (n=275) compared to wild-type siblings (n=376) (p=8.51e-11 by Welch's
674 two tailed t-test). Seeds were quantified with ImageJ.

676 **Fig 3. Transcriptional effects associated with low *HDG3* expression.**
(left panel) Genes downregulated in *hdg3* mutant endosperm also have reduced
678 expression in Col x Cvi endosperm compared to Col. The plot shows the expression profile of genes with significantly altered expression in *hdg3-1* endosperm ($p<0.05$).
680 Genes were hierarchically clustered by Euclidean distance and complete linkage using Gene-E. (right panel) A subset of putative developmental regulators with reduced
682 expression in *hdg3* and Col x Cvi endosperm.

684 **Fig 4. Gain of Cvi *HDG3* paternal allele methylation in endosperm.**

Total methylcytosine 5' of *HDG3* in F_1 endosperm, determined by bisulfite-PCR.

686 Maternally inherited Col allele in orange, paternally inherited Cvi allele in blue. Scale to 100%, tick marks below line indicate unmethylated cytosines. Col x Cvi data published in

688 Pignatta et al., *eLife* 2014.

690 **Fig 5. *HDG3* is imprinted in Cvi *HDG3* IR lines.**

HDG3 in situ for indicated genotypes. Arrowheads indicate regions of *in situ* signal. Right

692 panels shows magnification of chalazal region. Scale bars, 50 μm. (B) RT-qPCR of relative *HDG3* transcript abundance in F_1 endosperm at 6-7 DAP. Dashed line separates

694 experiments done at different times. Left, avg of 3 technical replicates. Right, avg of biological duplicates. Bars show upper and lower range. (C) % of *HDG3* from Cvi allele

696 in endosperm by TaqMan RT-qPCR assay.

698 **Fig 6. Effects of *HDG3* imprinting on Cvi seed development.**

(A) Aniline blue and safranin O staining of seed sections at 5 DAP from the indicated F_1

700 seeds. Scale bars, 50 μm. (B) Phenotypic characterization of sectioned seeds, assaying degree of endosperm cellularization relative to embryo stage. (C) Seed weight in selfed

702 Cvi and Cvi *HDG3* IR plants. Individual data points and mean +/- SD shown. P-value from unpaired two tailed t-test. (D) Seed area for self-fertilized Cvi (n=287 seeds), Cvi

704 *HDG3* IR 2-5 (n=496) and Cvi *HDG3* IR 3-4 (n=386). Differences between IR seeds and Cvi are significant at $p < 2.2e-16$ as determined by Welch's two-tailed t-test.

706

Fig 7. Schematic summary of relative seed development at 5 DAP. Shapes represent phenotypic space occupied by the indicated genotypes.

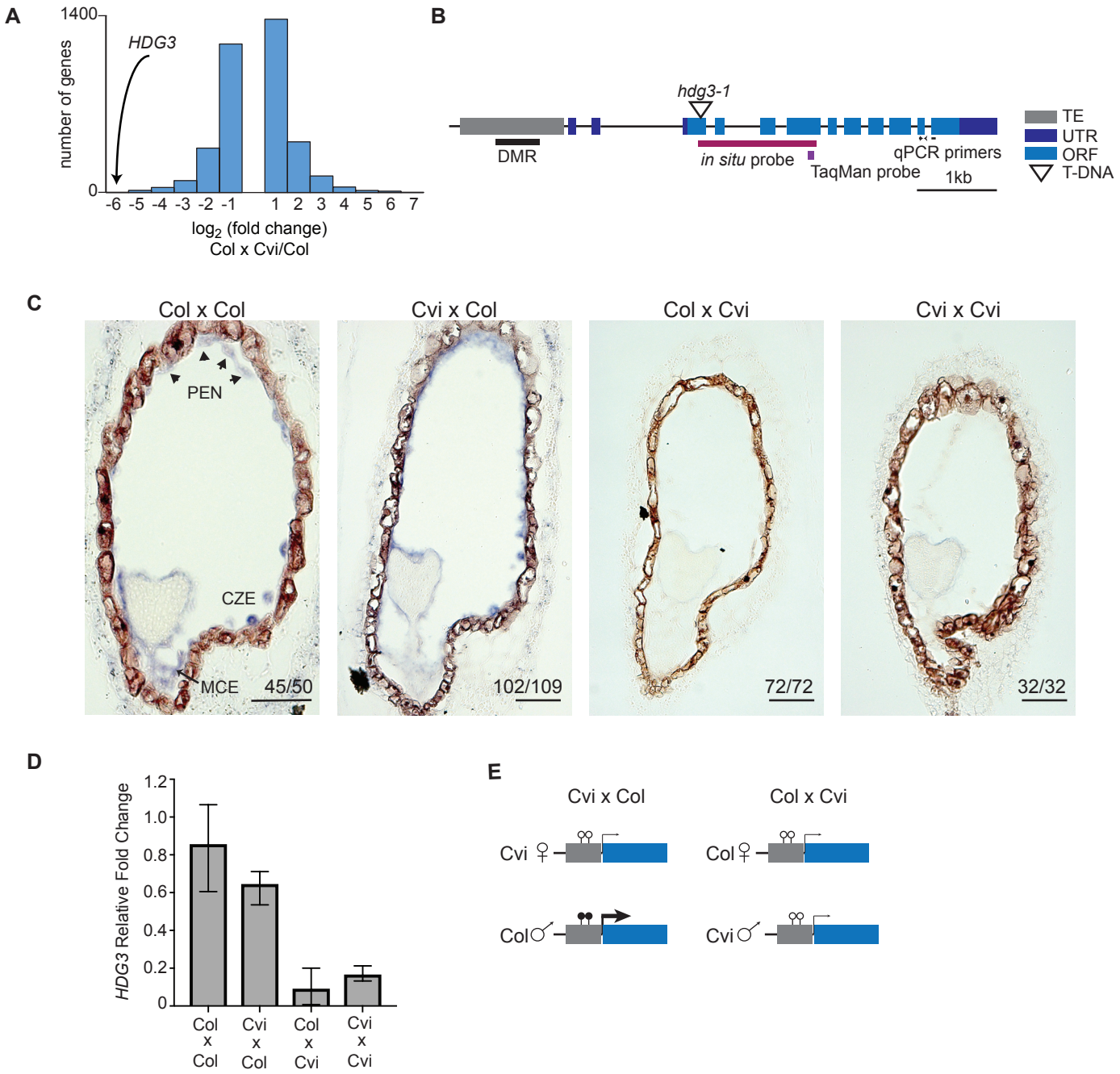

A
1400
0
number of genes
HDG3
-6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7
log2 (fold change)
Col x Cvi/Col
B
hdg3-1
DMR
in situ probe
TaqMan probe
qPCR primers
1kb
TE
UTR
ORF
T-DNA
C
Col x Col
Cvi x Col
Col x Cvi
Cvi x Cvi
PEN
CZE
MCE
45/50
102/109
72/72
32/32
D
HDG3 Relative Fold Change
1.2
1.0
0.8
0.6
0.4
0.2
0
Col x Col
Cvi x Col
Col x Cvi
Cvi x Cvi
E
Cvi x Col
Col x Cvi
Cvi ♀
Col ♂
Col ♀
Cvi ♂

Fig 1.

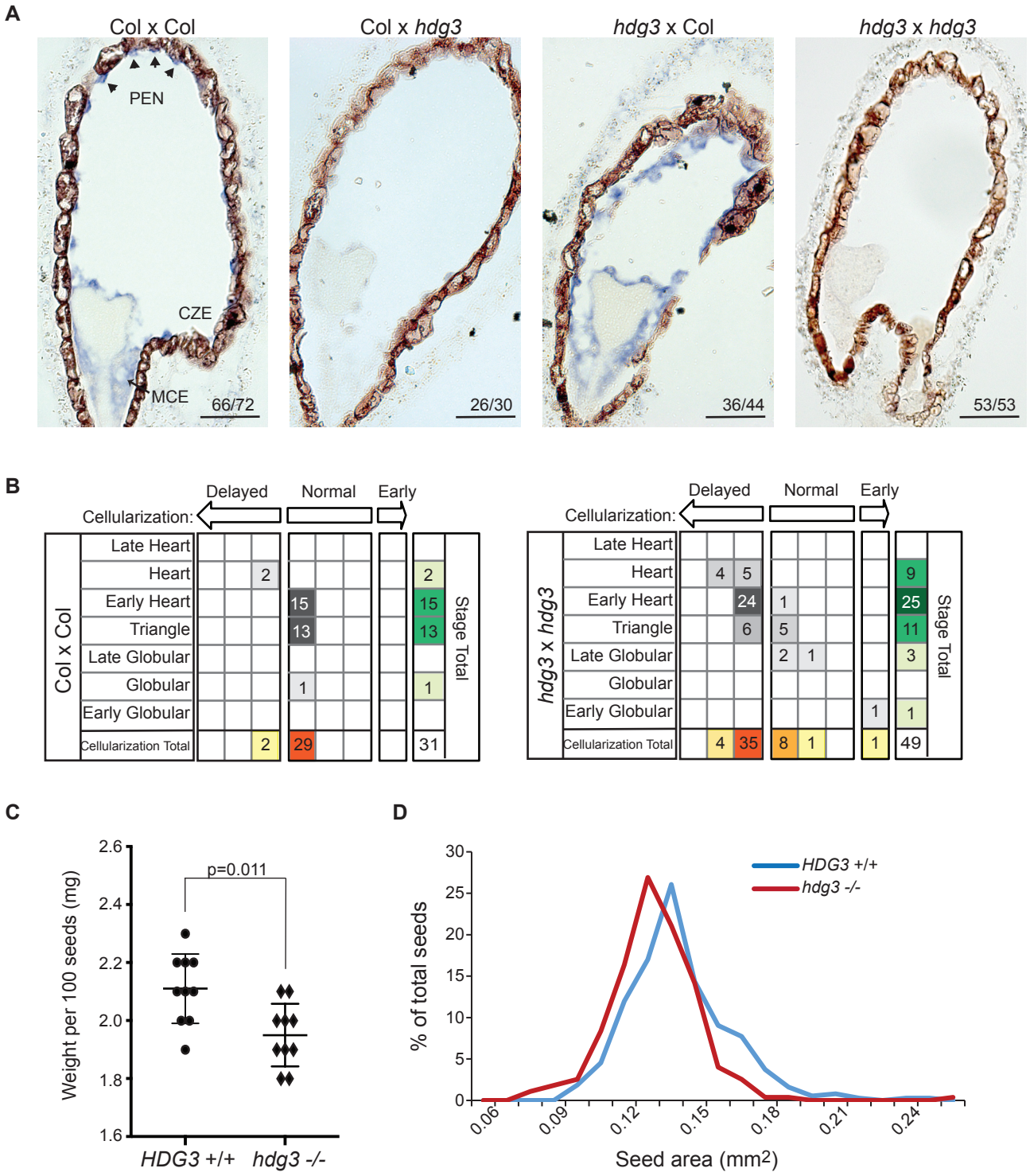

A
Col x Col
Col x hdg3
hdg3 x Col
hdg3 x hdg3
PEN
CZE
MCE
66/72
26/30
36/44
53/53
B
Cellularization:
Delayed
Normal
Early
Col x Col
Late Heart
Heart
Early Heart
Triangle
Late Globular
Globular
Early Globular
Cellularization Total
Stage Total
hdg3 x hdg3
C
p=0.011
Weight per 100 seeds (mg)
HDG3 +/+
hdg3 -/-
D
% of total seeds
Seed area (mm2)
HDG3 +/+
hdg3 -/-

Fig 2.

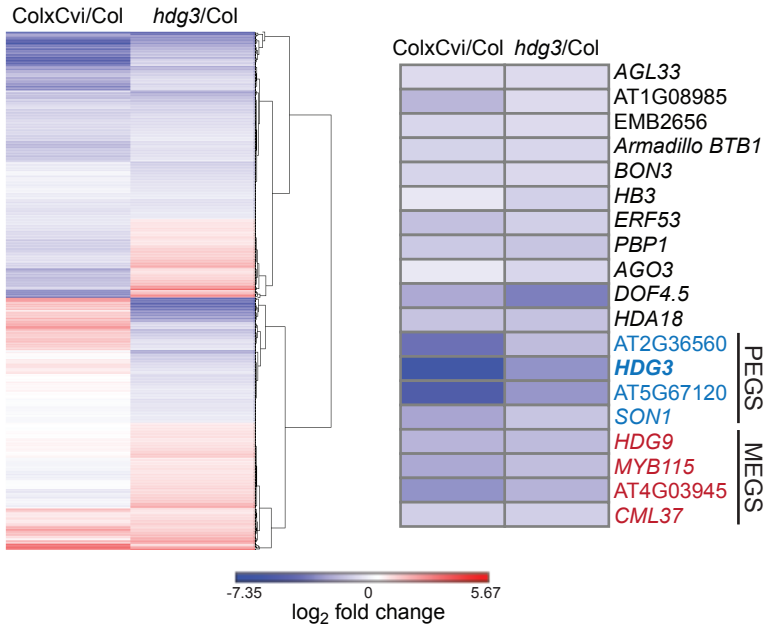

ColxCvi/Col
hdg3/Col
ColxCvi/Col
hdg3/Col
AGL33
AT1G08985
EMB2656
Armadillo BTB1
BON3
HB3
ERF53
PBP1
AGO3
DOF4.5
HDA18
AT2G36560
HDG3
AT5G67120
SON1
HDG9
MYB115
AT4G03945
CML37
PEGS
MEGS
-7.35
0
5.67
log2 fold change

Fig 3.

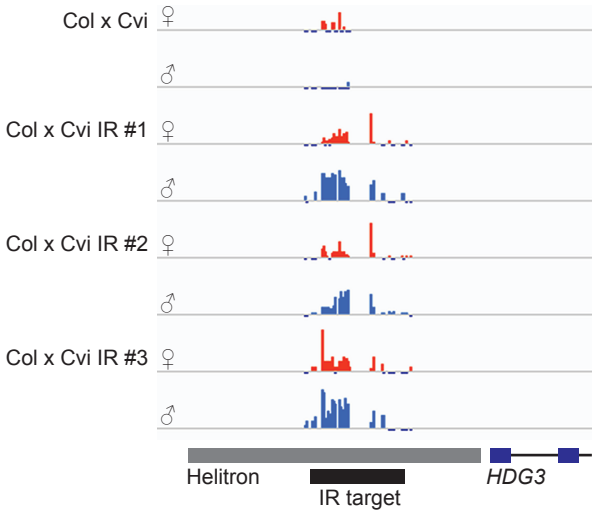

Col x Cvi
♀
♂
Col x Cvi IR #1
♀
♂
Col x Cvi IR #2
♀
♂
Col x Cvi IR #3
♀
♂
Helitron
IR target
HDG3

Fig 4.

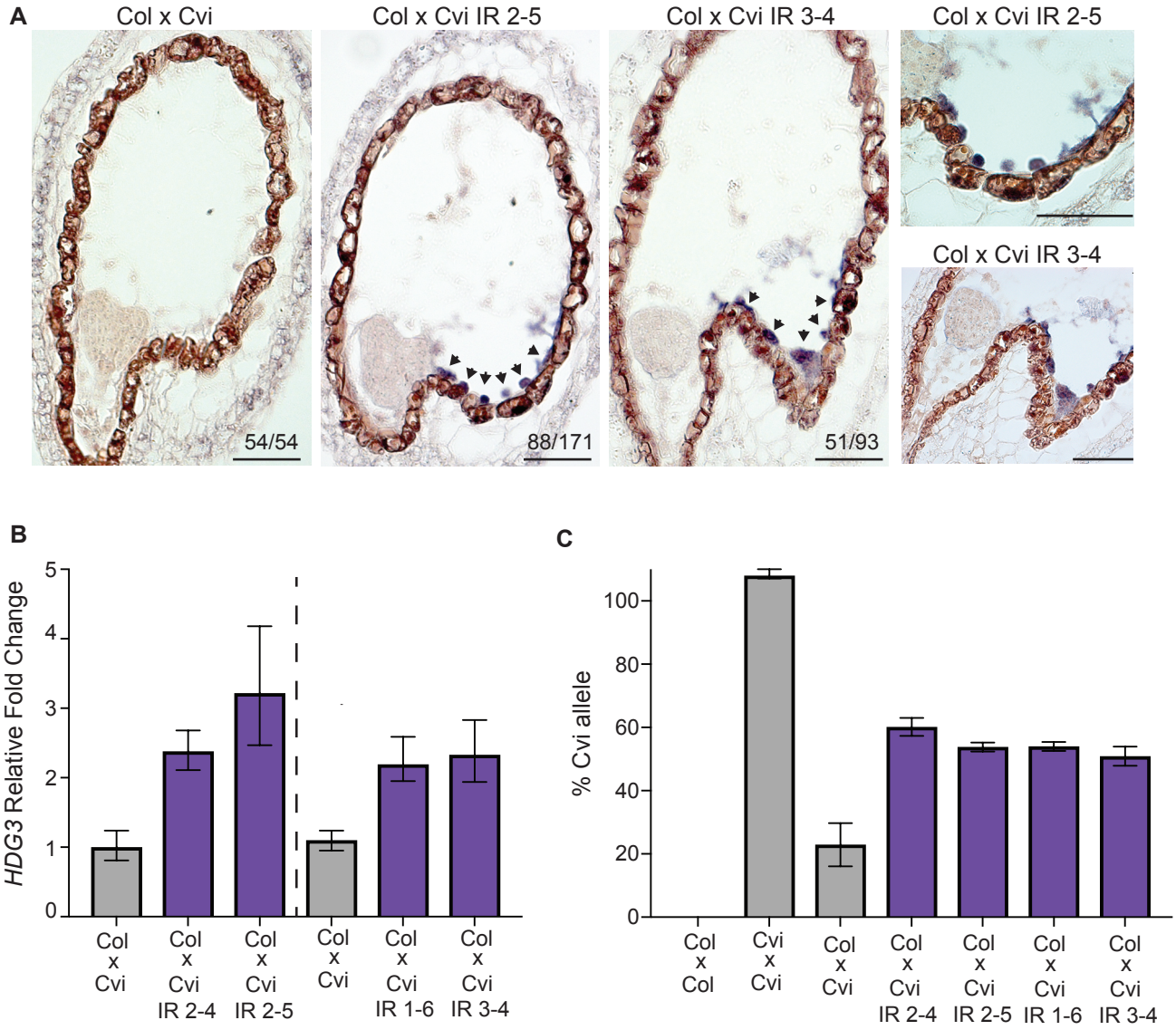

A
Col x Cvi
Col x Cvi IR 2-5
Col x Cvi IR 3-4
Col x Cvi IR 2-5
Col x Cvi IR 3-4
54/54
88/171
51/93
B
HDG3 Relative Fold Change
0
1
2
3
4
5
Col x Cvi
Col x Cvi IR 2-4
Col x Cvi IR 2-5
Col x Cvi
Col x Cvi IR 1-6
Col x Cvi IR 3-4
C
% Cvi allele
0
20
40
60
80
100
Col x Col
Cvi x Cvi
Col x Cvi
Col x Cvi IR 2-4
Col x Cvi IR 2-5
Col x Cvi IR 1-6
Col x Cvi IR 3-4

Fig 5.

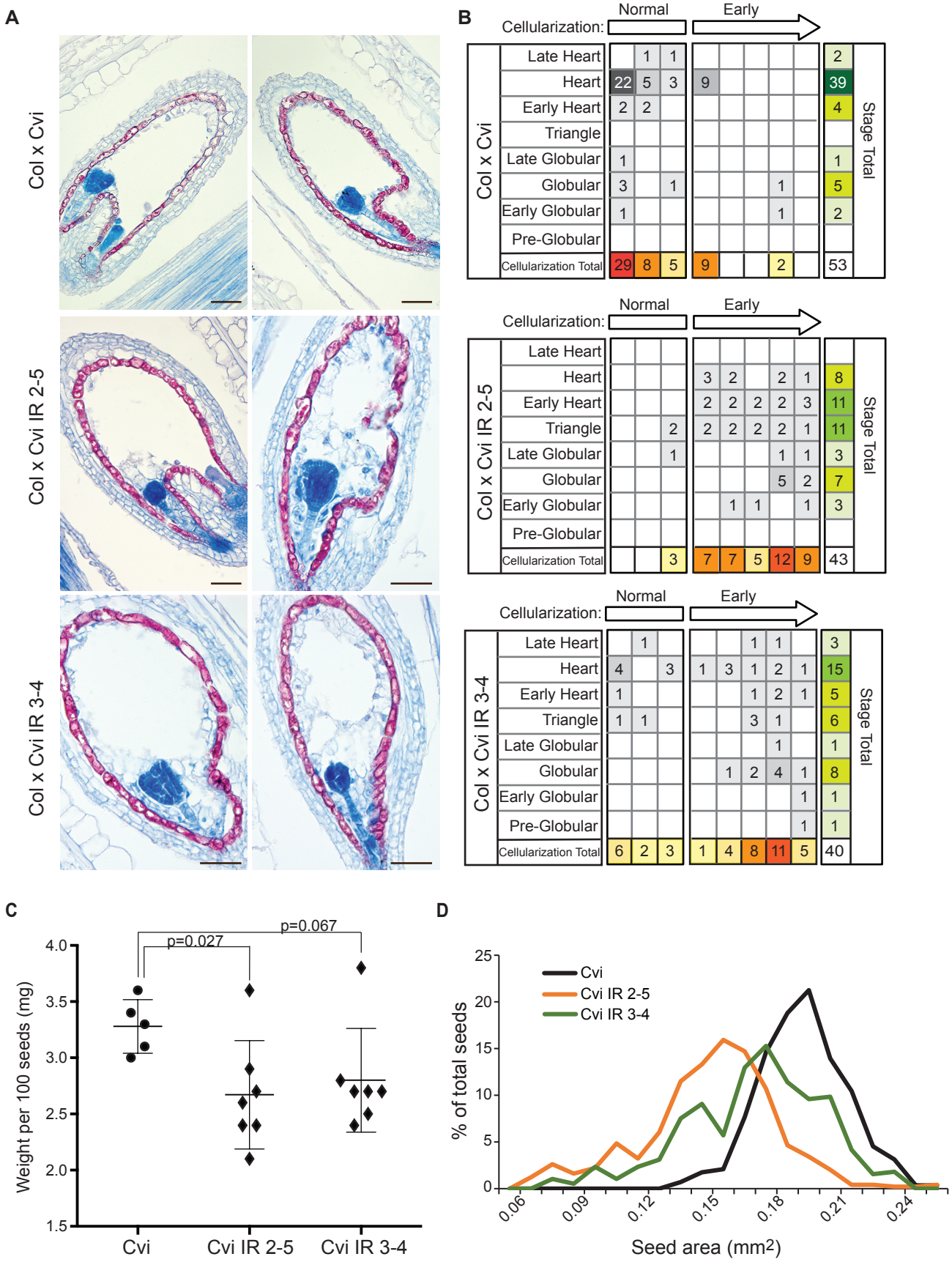

A
Col x Cvi
Col x Cvi IR 2-5
Col x Cvi IR 3-4
B
Cellularization:
Normal
Early
Late Heart
Heart
Early Heart
Triangle
Late Globular
Globular
Early Globular
Pre-Globular
Cellularization Total
Stage Total
Col x Cvi
Col x Cvi IR 2-5
Col x Cvi IR 3-4
C
p=0.027
p=0.067
Weight per 100 seeds (mg)
Cvi
Cvi IR 2-5
Cvi IR 3-4
D
Cvi
Cvi IR 2-5
Cvi IR 3-4
% of total seeds
Seed area (mm2)

Fig 6.

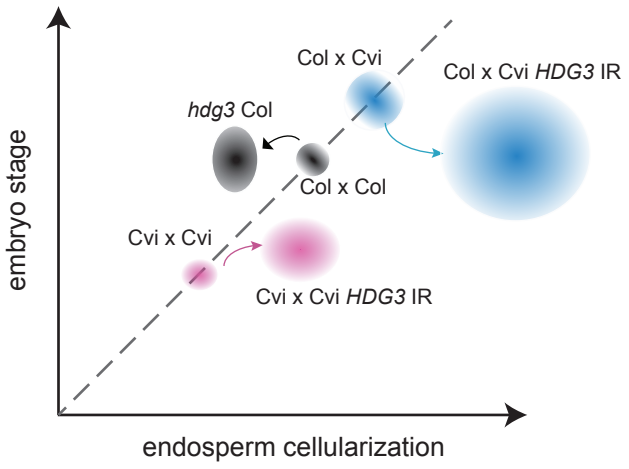
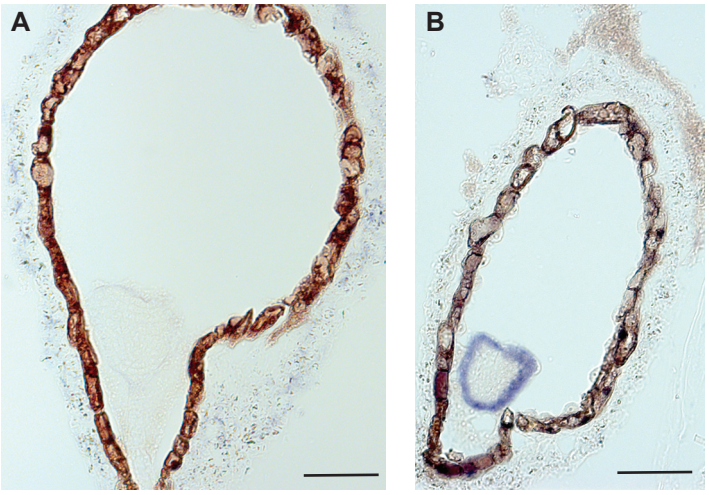
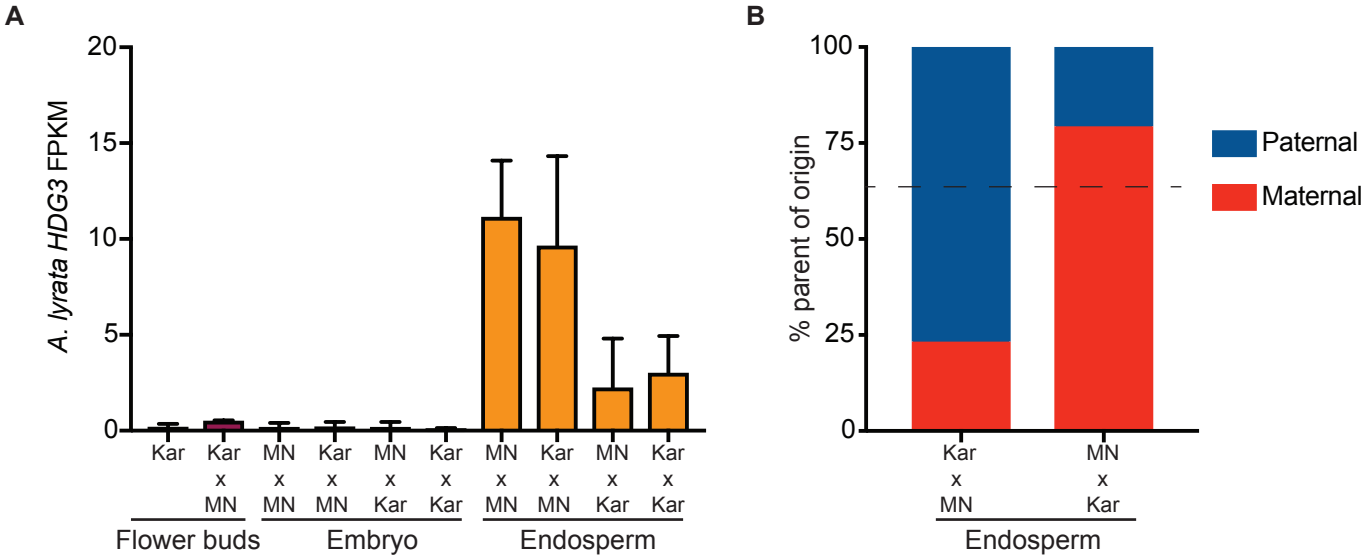

Col x Cvi
Col x Cvi *HDG3* IR
hdg3 Col
Col x Col
Cvi x Cvi
Cvi x Cvi *HDG3* IR
embryo stage
endosperm cellularization

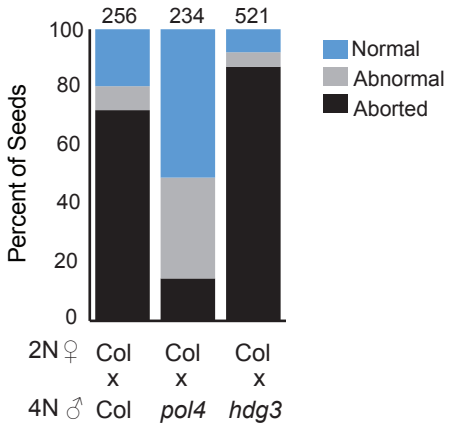
Fig 7.



S1 Fig. Negative and positive controls for *in situ* hybridization. (A) *HDG3* sense probe hybridization for Col x Col seed. (B) *PDF1* antisense probe hybridization for *hdg3-1* x *hdg3-1* seed. Scale bars, 50 μm.

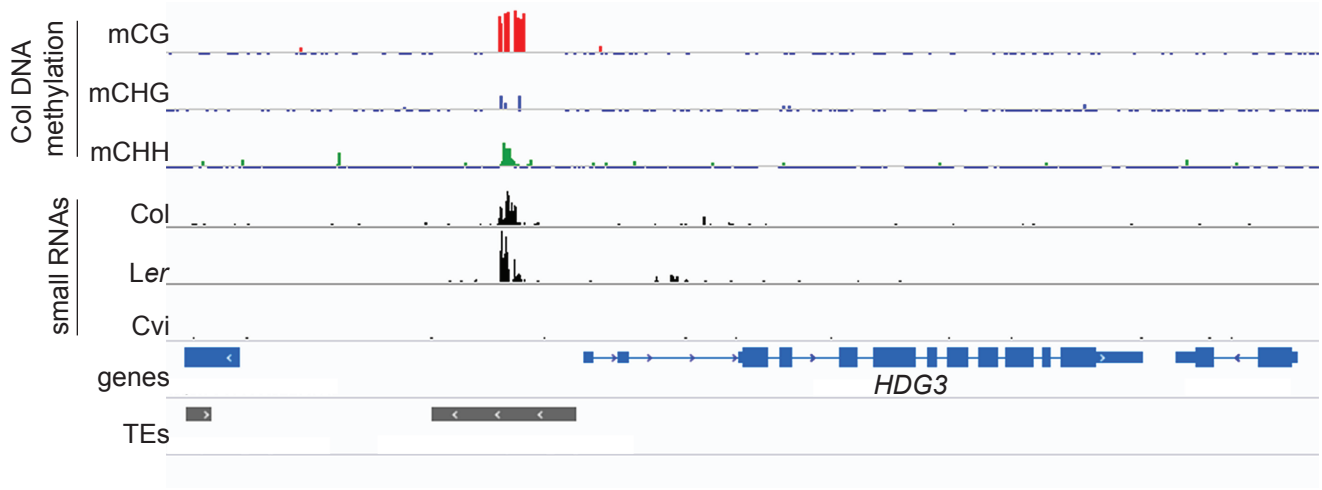


S2 Fig. *HDG3* expression in *A. lyrata*. (A) Expression of *HDG3* (ALAl_scaffold_0004_1698 on scaffold 4:15306009-15309421) is specific to endosperm. Bars show mean FPKM values with std deviation for 2-4 biological replicates per genotype. (B) Percent maternal and paternal allele transcripts for the indicated crosses. All data are culled from Klosinska et al., *Nature Plants* 2016.

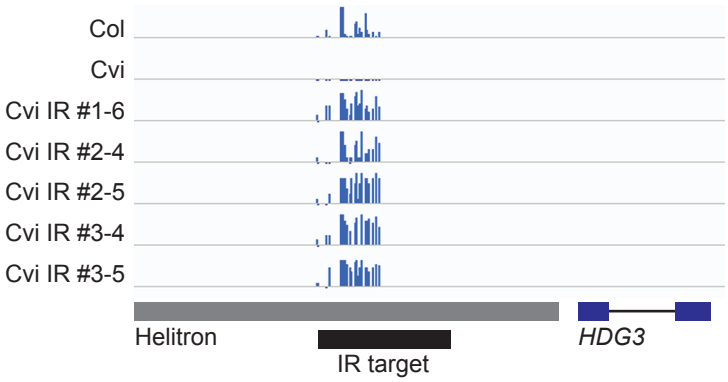


S3 Fig. Tetraploid *hdg3-1* does not rescue interploidy seed lethality caused by paternal genomic excess. Tetraploid plants were created by colchine treatment and confirmed by flow cytometry analysis of DNA content. Tetraploid *pol iv* mutants are a positive control for interploidy seed rescue. Number of seeds analyzed on top of bars.

A



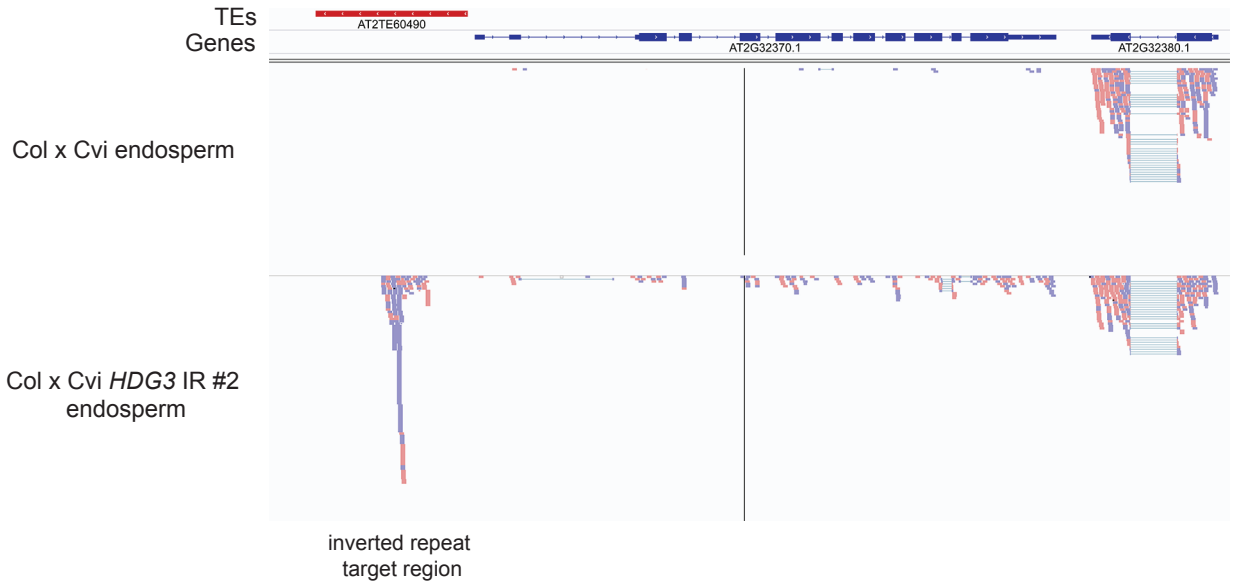
B



C

Tissue	# clones	% CG	% CHG	% CHH
Col leaf	15	80	24	14
Cvi leaf	11	1	0	0
Cvi IR #1-6 leaf	10	94	73	38
Cvi IR #2-4 leaf	7	96	38	21
Cvi IR #2-5 leaf	6	93	72	48
Cvi IR #3-4 leaf	15	99	69	51
Cvi IR #3-5 leaf	8	90	75	46

S4 Fig. Gain of *HDG3* methylation in Cvi *HDG3* IR lines. (A) Cvi naturally lacks small RNAs at the TE 5' of *HDG3*. DNA methylation data is from Col embryos (data from Pignatta et al., *eLife* 2014); small RNA data is from Col, L*er* and Cvi embryos (data from Erdmann et al., *Cell Rep.* 2017). (B) Total DNA methylation in Col, Cvi, and Cvi *HDG3* IR leaves as determined by bisulfite-PCR. Scale is 100%, tick marks below line indicate unmethylated cytosines. (C) Quantification of above data. Col and Cvi leaf data are from Pignatta et al., *eLife* 2014.

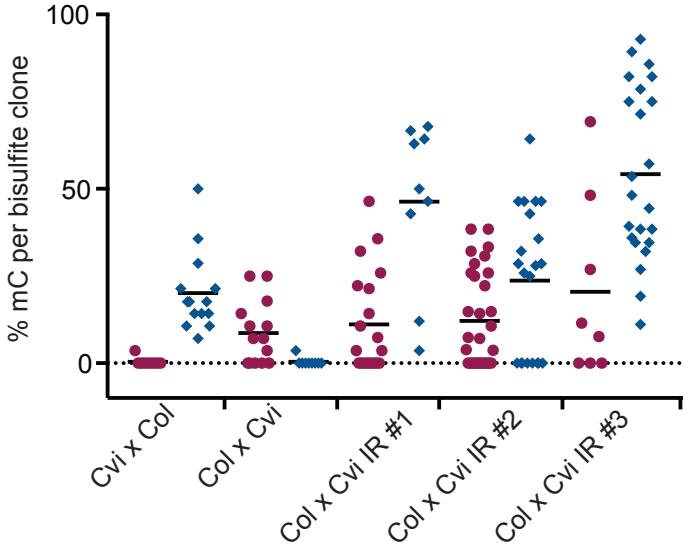


S5 Fig. Accumulation of inverted repeat RNA in Col x Cvi *HDG3* IR endosperm. Mapping of mRNA-seq reads to the *HDG3* (AT2G32370) locus in Col x Cvi and Col x Cvi *HDG3* IR #2 endosperm. Reads that match the inverted repeat target region represent expression of the inverted repeat transgene in endosperm from a different location.

A

Cross	Tissue	Allele	# Clones	% CG	% CHG	% CHH
Cvi x Col	Endosperm	Col	14	77.6	9.5	4.8
		Cvi	9	1.6	0	0
	Embryo	Col	9	74.1	3.7	7
		Cvi	18	2.3	0	2.5
Col x Cvi	Endosperm	Col	14	34.5	9.5	0.4
		Cvi	9	0	0	0.6
	Embryo	Col	16	87.5	14.6	3.6
		Cvi	6	2.4	5.6	0
Col x Cvi IR #1	Endosperm	Col	20	16	13.6	9.5
		Cvi	9	64.5	66.7	35.9
Col x Cvi IR #2	Endosperm	Col	31	31	14.0	5.7
		Cvi	21	18	19.1	28.2
	Embryo	Col	11	100	42.4	35.9
		Cvi	17	92.2	37.3	32
Col x Cvi IR #3	Endosperm	Col	8	30	24.0	17.4
		Cvi	23	86.1	69.6	40.6

B



S6 Fig. Bisulfite-PCR analysis of *HDG3* methylation in seeds. (A) % methylation from the indicated crosses. Female parent listed first. IR lines are independent transgenic events. Cvi x Col and Col x Cvi data are from Pignatta *et al.*, *eLife* 2014. (B) Endosperm total DNA methylation (%) for each bisulfite clone from the above data. Maroon circles, maternal allele clones; blue diamonds, paternal allele clones.

S2 Table: Endosperm Cellularization Timing and Embryogenesis Stages at 5 DAP

	Cellularization:	Delayed			Normal			Early					Stage Total
Col x Col	Torpedo												0
	Late Heart												0
	Heart			2									2
	Early Heart				15								15
	Triangle				13								13
	Late Globular												0
	Globular				1								1
	Early Globular												0
	Pre Globular (4 cell)												0
	Timing Totals	0	0	2	29	0	0	0	0	0	0	0	31
Col x Cvi	Torpedo												0
	Late Heart					1	1						2
	Heart				22	5	3	9					39
	Early Heart				2	2							4
	Triangle												0
	Late Globular				1								1
	Globular				3		1				1		5
	Early Globular				1						1		2
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	29	8	5	9	0	0	2	0	53
Cvi x Cvi	Torpedo												0
	Late Heart												0
	Heart												0
	Early Heart												0
	Triangle				10								10
	Late Globular				23								23
	Globular				222								222
	Early Globular				17								17
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	272	0	0	0	0	0	0	0	272
hdg3-1 x *hdg3-1*	Torpedo												0
	Late Heart												0
	Heart		4	5									9
	Early Heart			24	1								25
	Triangle			6	5								11
	Late Globular				2	1							3
	Globular												0
	Early Globular							1					1
	Pre Globular (4 cell)												0
	Timing Totals	0	4	35	8	1	0	1	0	0	0	0	49
Col x Cvi IR 2-5	Torpedo												0
	Late Heart												0
	Heart							3	2		2	1	8
	Early Heart							2	2	2	2	3	11
	Triangle						2	2	2	2	2	1	11
	Late Globular						1				1	1	3
	Globular										5	2	7
	Early Globular								1	1		1	3
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	0	0	3	7	7	5	12	9	43
Cvi x Cvi IR 2-5	Torpedo												0
	Late Heart						1						1
	Heart					2							2
	Early Heart				3	3							6
	Triangle				4	4	2						10
	Late Globular				3	1							4
	Globular				2	1							3
	Early Globular				2								2
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	14	11	3	0	0	0	0	0	28
Col x *hdg3-1*	Torpedo												0
	Late Heart												0
	Heart			3									3
	Early Heart			9	1								10
	Triangle			8	2								10
	Late Globular				1								1
	Globular				1								1
	Early Globular												0
	Pre Globular (4 cell)												0
	Timing Totals	0	0	20	5	0	0	0	0	0	0	0	25
Col x Cvi IR 3-4	Torpedo												0
	Late Heart					1				1	1		3
	Heart				4		3	1	3	1	2	1	15
	Early Heart				1					1	2	1	5
	Triangle				1	1				3	1		6
	Late Globular										1		1
	Globular								1	2	4	1	8
	Early Globular											1	1
	Pre Globular (4 cell)											1	1
	Timing Totals	0	0	0	6	2	3	1	4	8	11	5	40
Cvi x Cvi IR 3-4	Torpedo												0
	Late Heart												0
	Heart												0
	Early Heart						5						5
	Triangle					3	1						4
	Late Globular					2	1						3
	Globular												0
	Early Globular												0
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	0	5	7	0	0	0	0	0	12
hdg3-1 x Col	Torpedo												0
	Late Heart				1								1
	Heart			1	10	3							14
	Early Heart				21	3							24
	Triangle				25	3	1						29
	Late Globular				8	1							9
	Globular												0
	Early Globular												0
	Pre Globular (4 cell)												0
	Timing Totals	0	0	1	65	10	1	0	0	0	0	0	77
hdg3-1 x Cvi IR 2-5	Torpedo											1	1
	Late Heart												0
	Heart							1	1		3		5
	Early Heart									2	3	1	6
	Triangle						3		1	1		1	6
	Late Globular											3	3
	Globular						1				3	6	10
	Early Globular												0
	Pre Globular (4 cell)												0
	Timing Totals	0	0	0	0	0	4	1	2	3	9	12	31
Cvi x Col	Torpedo												0
	Late Heart												0
	Heart												0
	Early Heart												0
	Triangle												0
	Late Globular				5								5
	Globular				54								54
	Early Globular				24								24
	Pre Globular (4 cell)				5								5
	Timing Totals	0	0	0	88	0	0	0	0	0	0	0	88

S3 Table. Tests of statistical significance for pairwise comparisons of differences in endosperm or embryo development.

Seed Genotype 1	Seed Genotype 2	Endosperm diff?	Embryo diff?
Col x Col	*hdg3* x *hdg3*	***	
Col x Col	Col x *hdg3*	***	
Col x Col	*hdg3* x Col		
hdg3 x Col	Col x *hdg3*	***	
Col x Cvi	Col x Cvi *HDG3* IR 2-5	***	***
Col x Cvi	Col x Cvi *HDG3* IR 3-4	***	*
Col x Cvi	*hdg3* x Cvi *HDG3* IR 2-5	***	***
Cvi x Cvi	Cvi x Cvi *HDG3* IR 2-5		***
Cvi x Cvi	Cvi x Cvi *HDG3* IR 3-4		***
Col x Col	Col x Cvi	**	***
Col x Cvi	Cvi x Col	***	***
Col x Col	Cvi x Cvi	***	***
Col x Cvi	Cvi x Cvi	***	***
Cvi x Col	Cvi x Cvi		
Col x Col	Cvi x Col	*	***

***= difference significant, padj < 0.001; **= difference significant, padj < 0.01; *=difference significant, padj < 0.05; no stars = no significant difference

S5 Table. Oligos used in this study.

Oligo Name	Sequence 5'-3'	Gene	Purpose
Expression			
HDG3 Fwd Set 4	CATCATGGAGTTGGCATTTGG	*HDG3*	TaqMan RT-qPCR
HDG3 Rev Set 4	CAAAGCTAAGCTAGTGCCATTAAA	*HDG3*	TaqMan RT-qPCR
HDG3 Prb Set 4	/56-JOEN/ACT TGA GCC/ZEN/ ATC ACC AAG AGC TCC/3IABKFQ/	*HDG3* – Col	Col probe for TaqMan RT-qPCR
HDG3_Cvi	/56-FAM/ACT TGA GTC /ZEN/ ATC ACC AAG AAC TCC/3IABkFQ/	*HDG3* – Cvi	Cvi probe for Taqman RT-qPCR
MG510	TGCAACGGAGAGATGATGCACAAG	*HDG3*	Sybr Green RT-qPCR
MG511	TGCTCTTGCTAGTGTGTCCATGCC	*HDG3*	
MG446	CCATTCTACTTTTTGGCGGCT	AT1G58050	Sybr Green RT-qPCR
MG447	TCAATGGTAACTGATCCACTCTGATG	AT1G58050	
DNA methylation			
MG422	GTTTAAGGATATTTTGGATAATGTATTGA	*HDG3* 5' TE	BS-PCR
MG423	CTATRCTTTTATTAACTATATARATCRTTATACAC	*HDG3* 5' TE	
MG424	TACATCTCATATCTACAAATARTATTATTAAC	*HDG3* 5' TE	BS-PCR
MG425	TGGTATGAGYYTAGGAGAAATAATGTAAG	*HDG3* 5' TE	
In situ			
KN13	TCTCACCCTTCACCTCCATC	*PDF1*	Amplify 602 bp of *PDF1* coding sequence for in situ probe
KN14	GGGGTTGTGAAAGGGAACTT	*PDF1*	
KN19	GATGGGATCTAAGGGAAATGTCG	*HDG3*	Amplify 278 bp of *HDG3* coding sequence for in situ probe
KN20	ACATTGCCACGAGTGCAGTT	*HDG3*	
IR cloning			
DP215	CACCTACCAAATTATTATCTATTGAT	*HDG3* 5' TE	Amplify IR sequence for cloning
DP216	GAATTTCATAGTGAAATGGACCATC	*HDG3* 5' TE	