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 between parental alleles and in flowering plants occurs in endosperm seed tissue. Proper endosperm development is essential for the production of viable seeds. We previously showed that 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. These results indicate that epigenetic variation alone is sufficient to explain imprinting variation and show 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 not encoded in the DNA sequence, yet it is closely associated. In practice, it is challenging to disentangle genetic variation from epigenetic variation, as what seems like 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

DNA methylation is a heritable epigenetic mark that can, on occasion, effect
56 gene transcription and influence development. DNA methylation is a particularly influential regulator of gene expression in endosperm, a triploid extraembryonic seed
58 tissue that supports embryo development. In endosperm, developmentally programmed DNA demethylation causes maternally inherited endosperm genomes to be
60 hypomethylated compared to the paternally inherited endosperm genome (1-3). Methylation differences between maternal and paternal alleles identify their parent-of-
62 origin and establish imprinting, an epigenetic phenomenon in which a gene is expressed primarily from one parental allele (4).

64 Recent genomic approaches have revealed extensive natural DNA methylation variation within *Arabidopsis thaliana* (5,6), some of which has been shown to have
66 phenotypic consequences (7). Nevertheless, while the contribution of genetic variation to phenotypic diversity is well-established, the impact of epigenetic variation, or epialleles,
68 on phenotype is far less clear (8,9). We previously demonstrated that natural variation in DNA methylation is associated with imprinting variation, with as many as 10% of
70 imprinted genes estimated to be variably imprinted within *A. thaliana* (10). Seed development varies extensively among Arabidopsis accessions and has been previously
72 shown to be influenced by parent-of-origin effects (10,11), thus raising the possibility that variation in imprinting could influence seed phenotypes. One of these variably imprinted
74 genes, *HOMEDOMAIN GLABROUS3* (*HDG3*), is a member of the class IV homedomain-Leucine zipper transcription factor (HD-ZIP) family, which regulates
76 diverse aspects of plant patterning and development (12,13). Studies on the function of class IV HD-ZIP genes in trichome differentiation, sepal giant cell formation, and
78 suppression of somatic embryogenesis, among others, have led to the conclusion that

class IV HD-ZIP family genes promote endoreduplication and cell differentiation (13-16).
80 Several members of the class IV HD-ZIPs are primarily expressed in endosperm and exhibit imprinted expression patterns, including *FWA/HDG6*, *HDG8, HDG9* and *HDG3*
82 (2,12,17). *FWA, HDG8,* and *HDG9* are maternally expressed imprinted genes (MEGs), whereas *HDG3* is a paternally expressed imprinted gene (PEG) (2,17). The function of
84 the imprinted class IV HD-ZIP genes during seed development, if any, is unknown.

The activity of *HDG3* alleles is correlated with DNA methylation. In endosperm of
86 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
88 transcriptional start site (2). Maternally inherited endosperm alleles are demethylated by the 5-methylcytosine DNA glycosylase gene *DME*; in *dme* mutants, maternal alleles
90 retain their methylation and are also expressed (2,18). When strains where *HDG3* methylation is low, such as Cvi, are the paternal parent in crosses with Col, there is no
92 methylation difference between maternal and paternal alleles in endosperm and *HDG3* is biallelically expressed (10). Together, these data suggest that (1) DNA demethylation
94 promotes repression of the maternally-inherited *HDG3* allele whereas DNA methylation promotes expression (or inhibits repression) of the paternal *HDG3* allele and that (2)
96 imprinting variation is due to *cis* epigenetic variation at *HDG3* (10). However, a *cis* or *trans* genetic contribution to imprinting variation cannot be excluded because of DNA
98 sequence polymorphisms between the strains and alleles that do and do not exhibit imprinting.

100 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
102 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
104 hypomethylated, non-imprinted, repressed state to an imprinted, paternally biased,

expressed state. Additionally, gain of *HDG3* imprinting altered endosperm development. These data indicate that naturally occurring epialleles can have phenotypic consequences in endosperm, a tissue where methylation is dynamic as a programmed part of development.

Results

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

We previously showed that several genes that are imprinted in endosperm when Col is the paternal parent are not imprinted when Cvi is the paternal parent (10). To further examine naturally occurring endosperm gene expression variation, we 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 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 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 biallelically expressed in Col x Cvi (10). To further explore the expression variation of *HDG3*, we performed *in situ* hybridization on developing seeds (Fig 1B-C; S1 Fig). In Col 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 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 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.

130 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),
132 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
134 (and presumably Cvi x Cvi crosses) as detected by mRNA-seq (10), the total expression is lower than when *HDG3* is imprinted. These results suggest that DNA methylation of
136 the *HDG3* 5' region promotes *HDG3* expression (Fig 1E).

138 **Reduced *HDG3* expression affects seed development**

To examine if *HDG3* influenced endosperm development, we compared seeds
140 from *hdg3* mutant plants and segregating wild-type siblings in the Col background. We confirmed predominantly paternal expression of *HDG3* (2,10) by reciprocal crosses
142 between wild type and *hdg3-1* mutants (Fig 2A). When *hdg3* was crossed as a female to a wild-type sibling male, expression of *HDG3* was detected in endosperm in a similar
144 manner as in Col x Col (Fig 2A). In contrast, when wild-type females were crossed to *hdg3-1* mutant males, the accumulation of *HDG3* transcript in endosperm was
146 dramatically affected, with no transcript detected in most cases, despite the presence of a wild-type maternally inherited allele (Fig. 2A). We found that seed development was
148 slightly delayed in *hdg3* compared to wild-type at 6 and 8 DAP (Fig 2B; S2 Fig). Several PEGs have been shown to influence seed abortion phenotypes in interploidy crosses
150 (19,20), but we found no effect of *hdg3* on this process (S2 Fig). To understand the potential molecular consequences of the loss of *hdg3,* we profiled endosperm gene
152 expression in wild-type Col and *hdg3-1* by RNA-seq at 7 days after pollination (DAP). 150 genes had at least two-fold higher expression upon loss of *hdg3*, while 238 genes
154 had at least two-fold lower expression in *hdg3* mutant endosperm (Fig 2C, S2 Table). Differentially expressed genes included developmental regulators such as *Homeobox 3*

156 (*WOX9*) and gibberellin oxidases, which effect the level of a key phytohormone necessary for typical seed development (21) (Fig 2D). The loss of *hdg3* also impacted
158 ten imprinted genes, including the MEG *HDG9* (Fig 2D). Thus, we hypothesized that low expression of *HDG3* from Cvi paternal alleles might mimic *hdg3* mutants. Indeed, of the
160 238 genes that are down-regulated in *hdg3* mutants, 100 are also down regulated in Col x Cvi crosses, where *HDG3* expression is also low. This is a highly significant overlap
162 (hypergeometric test in R, p = 6.079e-69) (Fig 2C-D). These data suggest that the Cvi *HDG3* allele, in its hypomethylated and relatively transcriptionally repressed state, could
164 be important for accession-specific developmental traits imparted by Cvi. Thus, to test both the imprinting mechanism and function of *HDG3* further, we introduced methylation
166 at the *HDG3* locus in Cvi.

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

To distinguish the importance of genetic variation from epigenetic variation for
170 *HDG3* expression and imprinting, we generated transgenic lines in which the endogenous *HDG3* Cvi allele gained methylation in the same region that is methylated in
172 Col. Cvi was transformed with a transgene consisting of an inverted repeat (*HDG3* IR) of the 450 bp *HDG3* 5' region from Cvi under the control of the constitutive 35S promoter.
174 Processing of the expressed hairpin RNA into small RNAs is expected to direct methylation to the endogenous *HDG3* Cvi locus. We identified multiple independent
176 transgenic lines in which the *HDG3* 5' region gained methylation in leaves in all sequence contexts (S3 Fig). DNA methylation was present in the same region as in Col,
178 although non-CG methylation was considerably higher (S3 Fig).

To determine whether the Cvi allele remained methylated when paternally
180 inherited in endosperm, Cvi *HDG3* IR plants from three independent transgenic lines were crossed as males to wild type Col females and DNA methylation was evaluated in

182 F_1 endosperm by locus-specific bisulfite-PCR. The paternally inherited *HDG3* Cvi allele from the IR line was hypermethylated relative to the paternally inherited *HDG3* Cvi allele
184 in Col x Cvi endosperm (Fig 3; S4 Fig). The *HDG3* Cvi IR allele was methylated in both CG and non-CG contexts, indicative of RNA-directed DNA methylation, although at a
186 lower level than in leaves or in F_1 embryos (S4 Fig). The maternally inherited Col allele was unaffected in Col x Cvi *HDG3* IR endosperm, remaining hypomethylated like in Col
188 x Cvi endosperm. Thus, we successfully established an alternate epigenetic state specifically for the Cvi *HDG3* allele in endosperm.

190

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

Having established a methylated Cvi *HDG3* allele, we tested whether paternal allele
194 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
196 from Col x Cvi *HDG3* IR crosses indicated the presence of *HDG3* transcript in endosperm, in contrast to Col x Cvi endosperm (Fig 4A-E). Hybridization signal was
198 primarily detected in uncellularized endosperm on the chalazal side of the peripheral endosperm (Fig 4B-E). However, the penetrance of Cvi *HDG3* expression was variable,
200 with about half of the seeds exhibiting *HDG3* expression detectable by in situ (Fig 4B-C.) Analysis of total *HDG3* transcript abundance by RT-qPCR at 6-7 days after pollination
202 showed that *HDG3* expression was 2-3-fold higher in Col x Cvi *HDG3* IR endosperm compared to Col x Cvi endosperm (Fig 4F). Higher expression of *HDG3* in Col x Cvi
204 *HDG3* IR endosperm is consistent with *HDG3* being more highly expressed when imprinted (Fig 1). To measure allele-specific expression of *HDG3*, Col and Cvi alleles
206 were distinguished using TaqMan probes in an RT-qPCR assay. In crosses between Col

females and three independent Cvi *HDG3* IR lines, the fraction of transcript derived from
208 the Cvi allele increased compared to control crosses between Col females and Cvi
males. In Col x Cvi, the Cvi allele accounts for 23% of the transcripts by this assay, in
210 good agreement with prior allele-specific mRNA-seq results (10). In Col x Cvi *HDG3* IR
lines, the Cvi fraction was between 50-60%, indicating paternal allele bias (the
212 expectation for non-imprinted genes is 33% paternal). This is slightly less than the
fraction of paternal allele expression in Cvi x Col crosses by mRNA-seq (79%) (10).
214 These data indicate that the naturally occurring methylation variation at *HDG3* is
sufficient to explain imprinting variation. We conclude that the methylated Cvi *HDG3*
216 allele in Cvi *HDG3* IR plants is active and the gene is imprinted.

Does the change of *HDG3* imprinting in Cvi affect seed development? To test the
218 phenotypic consequences of expressing *HDG3* from previously repressed Cvi alleles,
we compared the phenotypes of seeds from Col x Cvi (low *HDG3* expression) and Col x
220 Cvi *HDG3* IR (2-4-fold increased *HDG3* expression) seeds by sectioning and staining
(Fig. 5). In crosses with the *HDG3* IR lines, endosperm cellularization occurred at an
222 earlier stage of embryo development, where it was observed as early as the globular
stage of embryogenesis. These observations support the hypothesis that
224 hypomethylation and repression of the Cvi *HDG3* allele is important for Cvi-directed
developmental programs and that epiallelic variation contributes to the natural variation
226 in seed development in Arabidopsis.

228 Discussion

The establishment, maintenance, and inheritance of DNA methylation are fairly
230 well understood processes. Disruption of methylation patterns by mutations in DNA
methyltransferases enzymes have clear gene expression consequences. However,

232 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
234 understood. An exception to this is in the endosperm, where active DNA demethylation in the female gamete before fertilization establishes differential DNA methylation after
236 fertilization, a step that is essential for normal seed development (22). We thus hypothesized that the phenotypic impact of naturally occurring epialleles might be most
238 pronounced in the endosperm, because the differential methylation between maternal and paternal alleles that is required for gene imprinting could be variable across
240 accessions (10). We have shown that *HDG3* represents a case study of this proposed phenomenon. By placing a methylation trigger in Cvi (the *HDG3* IR transgene), we were
242 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 inherited Cvi
244 *HDG3* allele in endosperm to 3-fold higher levels. Because we altered methylation at the endogenous *HDG3* Cvi locus, which retains all DNA sequence polymorphisms, we have
246 shown that methylation variation alone, or the act of methylating the DNA, is sufficient to cause imprinting variation. However, our results also show that it is unlikely that
248 methylation of the proximal TE accounts for all of the expression differences between paternal Col and Cvi *HDG3* alleles in endosperm. The paternally inherited methylated
250 Cvi allele, while more highly expressed than paternally inherited hypomethylated Cvi allele, was not as highly expressed as paternally inherited methylated Col alleles in
252 endosperm (Figs 1 and 4). Additional *cis* genetic or *trans* genetic or epigenetic variation likely also affects *HDG3* expression levels.

254 Our experiments also shed light on the relative receptiveness of maternal and paternal endosperm genomes to *de novo* methylation. The *HDG3* inverted repeat
256 transgene should create endosperm small RNAs that are homologous to both Col and Cvi alleles. Yet, in endosperm from Col x Cvi *HDG3* IR crosses, the maternally inherited

258 Col alleles remained hypomethylated, whereas the paternally inherited Cvi allele had high levels of non-CG methylation (Fig 3, S4 Fig). In contrast, F_1 embryos from the same
260 crosses were indeed more highly methylated in the non-CG context on maternal Col alleles compared to maternal Col alleles from Col x Cvi crosses (S4 Fig). Thus,
262 maternally inherited *HDG3* alleles in endosperm are refractory to *de novo* methylation even when a methylation trigger is present, in contrast to maternally inherited *HDG3*
264 alleles in embryos. These results further support findings that once a region is actively demethylated on the maternally inherited endosperm genome, it is "protected" from *de*
266 *novo* methylation even when triggering small RNAs are present (20).

Finally, although the direct targets of the *HDG3* transcription factor are still
268 unknown, we have shown that natural variation in *HDG3* expression (expressed in Col, low expression in Cvi) has consequences for seed gene expression programs and
270 development (Figs 2 and 5). Ectopic expression of *HDG3* in seeds fathered by Cvi caused early endosperm cellularization (Fig 5). These findings are consistent with class
272 IV HD-ZIP genes inhibiting the cell cycle and promoting cellular differentiation (13,16). Thus, seed phenotypic differences can be caused by methylation differences at single
274 genes. This study provides further evidence that epigenetic differences underlie developmental adaptations in plants. We have previously shown that the imprinting
276 status of many genes varies between accessions; our current study argues that intraspecific variation in imprinting is an important determinant of seed developmental
278 variation.

280 Materials and Methods

Plant Material

282 The SALK insertion mutant was obtained from the Arabidopsis Biological Resource Center (23). *hdg3-1* (SALK_033462) was backcrossed to Col-0 three times before
284 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
286 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. For Col and *hdg3-1* seed
288 stage analysis, seeds were collected at 6 or 8 DAP and cleared in chloral hydrate. The evaluator was blind to the seed genotype for experiment #1 (Fig 2, MG evaluated) and
290 experiment #3 (Fig S2, DP evaluated), but not for experiment #2 (Fig S2, MG evaluated). Seed stage was described according to (24).
292

In situ hybridization

294 Controlled floral pollinations were performed for each specified cross. Siliques were harvested 5 or 6 days after pollination (DAP) and fixed in FAA overnight at 4°C.
296 Following dehydration and clearing (HistoClear, National Diagnostics), samples were embedded in Paraplast Plus (McCormick Scientific), and sectioned at 9 μM (Leica RM
298 2065 rotary microtome). Ribbons were mounted with DEPC water on ProbeOn Plus slides (Fisher) at 42°C and dried overnight at 37°C. For probes, a 278 bp region of
300 *HDG3* (S3 Table) and previously published 602 bp probe for *PDF1* (25) were amplified from endosperm cDNA and cloned into P-GEM T vectors (Promega). Plasmids
302 containing sense and antisense oriented fragments were identified and linear templates were amplified using M13 forward and reverse primers for probe synthesis. Antisense
304 and sense RNA probes were synthesized in vitro with digoxigenin-UTPs using T7 or SP6 polymerase (DIG RNA labeling kit; Roche/Sigma Aldrich). Probes were subsequently
306 hydrolyzed and dot blots were performed to estimate probe concentration. Pre-

hybridization steps were preformed according to (26) except Pronase digestion occurred
308 for 15 minutes at 37°C. Hybridization and post-hybridizations were performed according
to (27), with minor modifications. For higher confidence in directly comparing expression
310 patterns, slides corresponding to the cross and its reciprocal were processed face to
face in the same pairs for hybridization, antibody, and detection steps. Negative controls
312 consisted of hybridizing sense probes to wild-type tissue and antisense probes to *hdg3*
tissue. The sense probe lacked signal (S1 Fig). A probe to *PDF1*, which is expressed in
314 the L1 embryo layer (28), served as a positive control for successful *in situ* hybridization
(S1 Fig). Hybridization was performed overnight at 55°C, slides were then washed twice
316 in 0.2X SSC for 60 mins each at 55°C, then twice in NTE for 5 min at 37°C and RNaseA
treated for 20 min at 37°C, followed by two more 5 min NTE washes. Slides were
318 incubated at room temperature for 1 hour with Anti-DIG antibody (Roche/Sigma Aldrich)
diluted 1:1250 in buffer A (27). Slides were then washed four times for 20 min each at
320 room temperature with buffer A and once for 5 min with detection buffer (27).
Colorimetric detections were performed using NBT/BCIP Ready-To-Use Tablets
322 (Roche/Sigma Aldrich) dissolved in water. Slides were allowed to develop 16-24 hours
before stopping. Slides were dehydrated, mounted in Permount (Electron Microscopy
324 Sciences) and imaged using a Zeiss Axio Imager M2. Minor level adjustments and smart
sharpen were applied to images to compensate for image transfer from live to digital
326 (Adobe Photoshop).

328 **Seed Staining**

Plant material was fixed and embedded as previously described and sectioned at 9 µm.
330 Slides were dewaxed twice in xylenes for 5 minutes, rehydrated through a graded
ethanol/0.85% salt series from 100%-30%, 1 minute each, stained in 0.6% Safranin O
332 Solution (Cat# 2016-03, Sigma Aldrich) for 5 minutes, washed with water, stained with a

saturated 2.5% Aniline blue (Harleco–EMD Millipore, #128-12) in 2% glacial acetic acid
334 aqueous solution for 3 minutes, washed with water, rapidly dehydrated though graded ethanol/salt series to 100%, 5 seconds for each step, and then twice in xylenes for 5
336 minutes each. Slides were briefly drained, cover slipped and mounted with Cytoseal™ 60 (Thermo Scientific) and imaged using a Zeiss Axio Imager M2.

338

Inverted repeat transgene

340 The 450 bp sequence 5' of *HDG3* corresponding to a fragment of AT2TE60490 from Chr2: 13740010-13740460 was amplified from Cvi (S3 Table) and cloned into the
342 directional entry vector pENTR-TOPO-D (Invitrogen). The sequence was then inserted twice in an inverted repeat conformation into the vector pFGCGW (29) with a LR clonase
344 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.
346 T_3 plants homozygous for the IR transgene and with a methylated *HDG3* 5' region in leaves were identified and used for subsequent experiments.

348

Quantitative RT-PCR

350 RNA was isolated from endosperm at 6 or 7 DAP as described (30) using RNAqueous Micro Kit (Ambion, Life Technologies Corporation). DNAse I-treated RNA (Invitrogen,
352 Life Technologies Corporation) was used for cDNA synthesis with oligo-dT primer using Superscript II reverse transcriptase (Invitrogen) according to manufacturers' instructions.
354 Quantitative RT-PCR (RT-qPCR) was performed using Fast Sybr-Green mix or TaqMan Mastermix (Applied Biosystems). All reactions were performed in three or four technical
356 replicates using a StepOne Plus Real-Time PCR system (Applied Biosystems). For Sybr-Green based assays, relative expression was calculated using the ddCt method as
358 described (31). The reference gene was AT1G58050 (32). For allele-specific expression

in Col-Cvi crosses, a multiplex TaqMan assay was developed by designing primers and
360 PrimeTime® Double-quenched Custom Probes with online tool
http://www.idtdna.com/pages/products/gene-expression/custom-qpcr-probes. Cycling
362 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 seconds with
364 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 (R^2 value >0.99) as
366 reference. Primer and probe sequences are available in S3 Table.

368 **Bisulfite sequencing**

Genomic DNA was isolated from leaves, endosperm and embryo at 6 or 7 days after
370 pollination using a CTAB procedure. Bisulfite treatment was performed using the
MethylCode Bisulfite Conversion Kit (Invitrogen, Life Technologies Corporation) or
372 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
374 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,
376 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
378 individual colonies were sequenced. Sequenced products were aligned using SeqMan
and methylation was quantified using CyMate (33).

380

mRNA-seq

382 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
384 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 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 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 Tophat2 (34) 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 was detected with Cuffdiff2 (35) and the ARAPORT11 annotation (S1 and S2 Tables). Reads are deposited in GEO GSExxx.

Acknowledgements

We thank R. Jaenisch and P. Reddien for sharing equipment for in situ hybridization and B. Williams for comments on the manuscript.

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. 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.

6. 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.

7. 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.

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

9. 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.

10. 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.

11. 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. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96(8):4710–7.

12. 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.

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

14. Abe M, Katsumata H, Komeda Y, Takahashi T. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development. 2003; 30(4):635–43.

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

16. 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.

448 17. Kinoshita T. One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. Science 2004; 303(5657):521–3.

450 18. Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, et al. Regulation of imprinted gene expression in Arabidopsis endosperm. Proceedings of the
452 National Academy of Sciences of the United States of America. 2011; 108(5):1755–62.

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

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

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

462 22. Gehring M, Satyaki PR. Endosperm and Imprinting, Inextricably Linked. Plant Physiol. 2017;1 73(1):143–54.

464 23. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science. 2003;
466 301(5633):653–7.

24. Bowman J. Arabidopsis: An atlas of morphology and development. Springer-
468 Verlag; 1994.

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

472 26. Jackson D. In situ hybridization in plants. In: Molecular plant pathology : a practical approach. Oxford University Press; 1991.

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

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

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

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

31. 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.

32. 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.

33. 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.

34. 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.

35. 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

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

S2 Fig. Phenotypic effects of low *HDG3* expression.

S3 Fig. Gain of *HDG3* methylation in Cvi IR lines.

S4 Fig. Bisulfite-PCR analysis of *HDG3* methylation in seeds.

S1 Table. Gene expression difference in ColxCvi endosperm vs. Col endosperm at 7 DAP determined using CuffDiff.

S2 Table. Gene expression differences in *hdg3-1* endosperm vs. Col endosperm at 7 DAP determined using CuffDiff.

S3 Table. Oligos used in this study.

Figure Legends

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

514 (A) *HDG3* expression is decreased in Col x Cvi endosperm compared to Col, as determined by mRNA-seq. (B) Schematic of the *HDG3* locus. DMR, differentially
516 methylated region. (C) In situ hybridization of *HDG3* (purple) in F_1 seeds from the indicated crosses. Female parent written first. In crosses where Col is the male parent,
518 *HDG3* is detected in the micropylar (MCE), peripheral (PEN), and chalazal (CZE) endosperm. Arrowheads indicate nuclear-cytoplasmic domains. Number of seeds with
520 shown pattern out of total seeds assayed is in corner of each image. At least two independent in situ experiments were performed for each genotype. Scale bars, 50 μm.
522 (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)
524 Schematic representation of relationship between *HDG3* methylation, expression, and imprinting. Thickness of arrows denotes relative expression level. Lollipops represent
526 methylated (filled) and unmethylated (open) cytosines.

528 **Fig 2. Transcriptional and phenotypic effects associated with low *HDG3* expression.**

530 (A) In situ hybridization of *HDG3* (purple) in seeds from the indicated crosses. Scale bars, 50 μm. (B) Seed development is delayed in *hdg3* (6 DAP, n=290 seeds; 8DAP,
532 n=611) compared to wild-type (6 DAP, n=255; 8 DAP, n=188). (C) Genes downregulated in *hdg3* mutant endosperm also have reduced expression in Col x Cvi endosperm
534 compared to Col. The plot shows the expression profile of genes with significantly altered expression in *hdg3-1* endosperm ($p<0.05$). Genes were hierarchically clustered
536 by Euclidean distance and complete linkage using Gene-E. (D) A subset of putative developmental regulators with reduced expression in *hdg3* and ColxCvi endosperm.
538

Fig 3. Gain of Cvi *HDG3* paternal allele methylation in endosperm.

540 Total methylcytosine 5' of *HDG3* in F_1 endosperm, determined by bisulfite-PCR.
Maternally inherited Col allele in orange, paternally inherited Cvi allele in blue. Scale to
542 100%, tick marks below line indicate unmethylated cytosines. Col x Cvi data published in
Pignatta et al., *eLife* 2014.

544

Fig 4. *HDG3* is imprinted in Cvi *HDG3* IR lines.

546 *HDG3* in situ for (A) Col x Cvi (B) Col x Cvi HDG3 IR #2-5 and (C) Col x Cvi *HDG3* IR
#3-4. Arrowheads indicate regions of in situ signal. (D) Magnification of chalazal region
548 in B. (E) Magnification of chalazal region in C. Scale bars, 50 μm. (F) RT-qPCR of
relative *HDG3* transcript abundance in F_1 endosperm at 6-7 DAP. Dashed line separates
550 experiments done at different times. Left, avg of 3 technical replicates. Right, avg of
biological duplicates. Bars show upper and lower range. (G) % of *HDG3* from Cvi allele
552 in endosperm by TaqMan RT-qPCR assay.

554 **Fig 5. Effects of *HDG3* imprinting on seed development.**
Aniline blue and safranin O staining of seed sections at 5 days after pollination. (A-B)
556 Col x Cvi (C-D) Col x Cvi *HDG3* IR #2-5. (E-F) Col x Cvi *HDG3* IR #3-4. Scale bars, 50
μm.

558

560

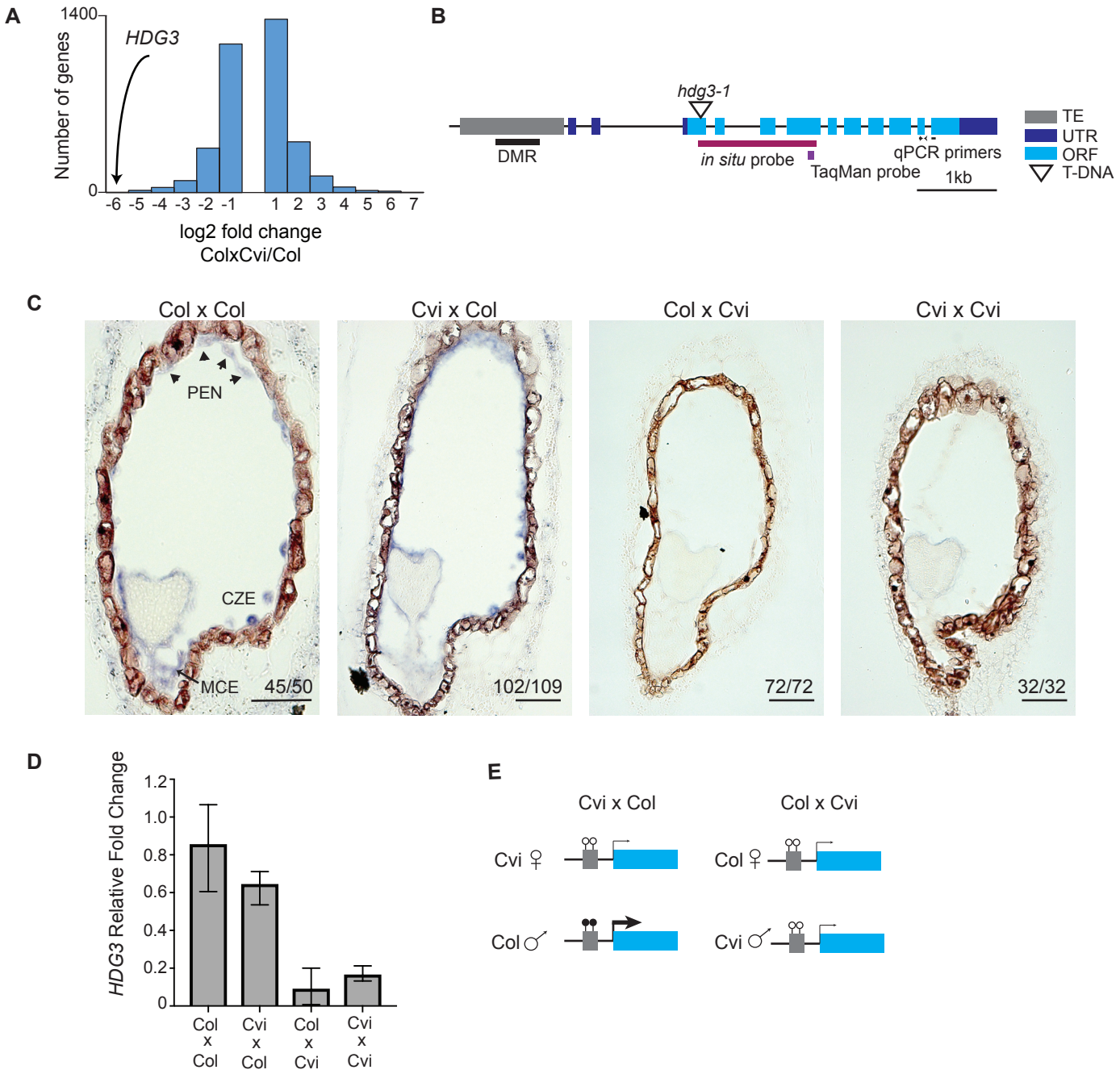


Fig 1. Natural variation in imprinting is associated with differences in *HDG3* expression levels. (A) *HDG3* expression is decreased in Col x Cvi endosperm compared to Col, as determined by mRNA-seq. (B) Schematic of the *HDG3* locus. DMR, differentially methylated region. (C) In situ hybridization of *HDG3* (purple) in F1 seeds from the indicated crosses. Female parent written first. In crosses where Col is the male parent, *HDG3* transcript is detected in the micropylar (MCE), peripheral (PEN), and chalazal (CZE) endosperm. Arrowheads indicate nuclear-cytoplasmic domains. Number of seeds with shown pattern out of total seeds assayed is in corner of each image. At least two independent in situ experiments were performed for each genotype. Scale bars, 50 μm. (D) RT-qPCR analysis of relative *HDG3* transcript abundance in F1 endosperm. Values are average of 3 biological replicates, bars represent upper and lower range. (E) Schematic representation of relationship between *HDG3* methylation, expression, and imprinting in endosperm. Thickness of arrows denotes relative expression level. Lollipops represent methylated (filled) and unmethylated (open) cytosines.

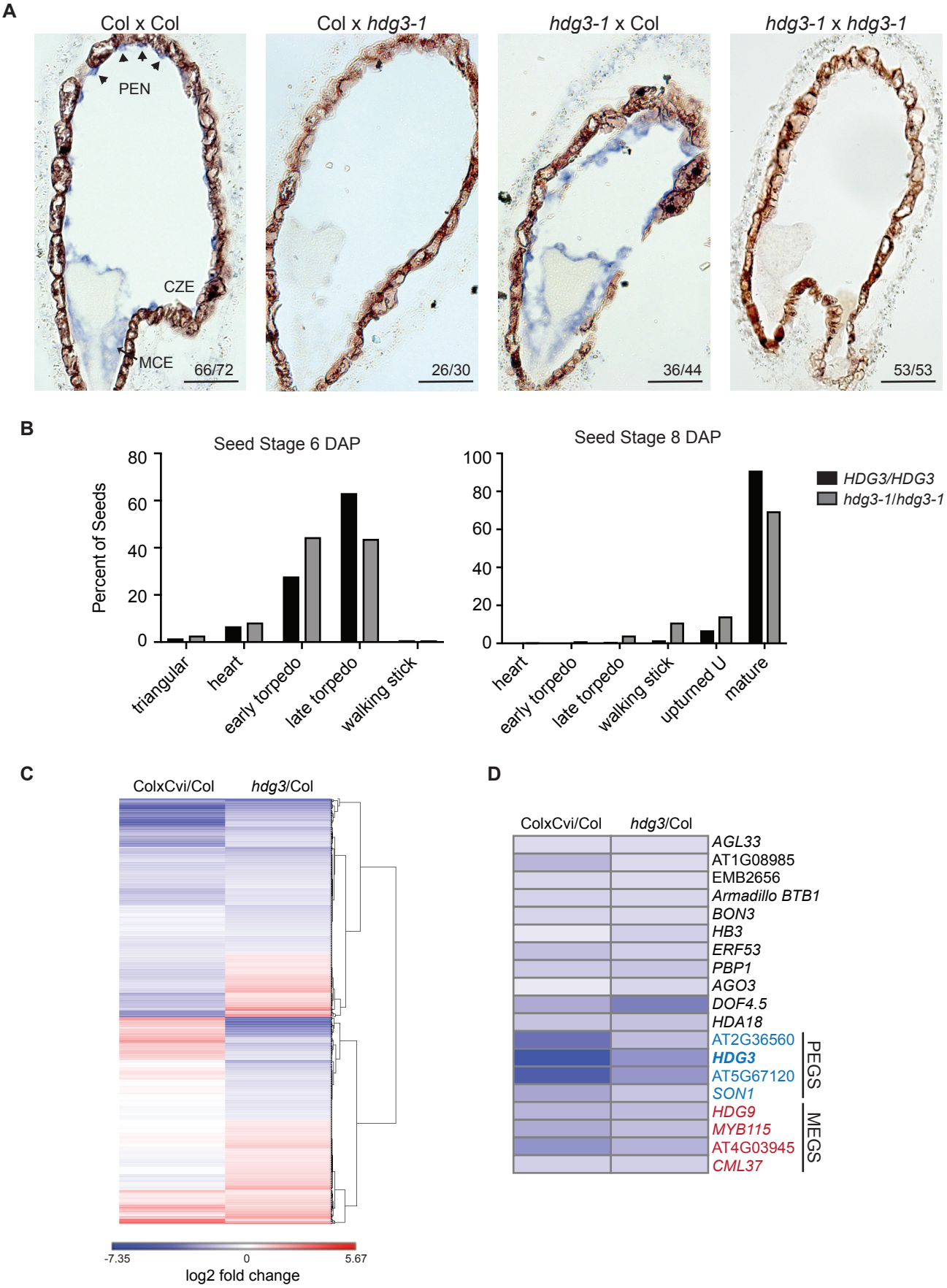


Fig 2. Transcriptional and phenotypic effects associated with low *HDG3* expression. (A) In situ hybridization of *HDG3* (purple) in seeds from the indicated crosses. Scale bars, 50 μm. (B) Seed development is delayed in *hdg3* (6 DAP, n=290 seeds; 8DAP, n=611) compared to wild-type (6 DAP, n=255; 8 DAP, n=188). (C) Genes dowregulated in *hdg3* mutant endosperm also have reduced 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 ($p<0.05$). Genes were hierarchically clustered by Euclidean distance and complete linkage using Gene-E. (D) A subset of putative developmental regulators with reduced expression in *hdg3* and ColxCvi endosperm.

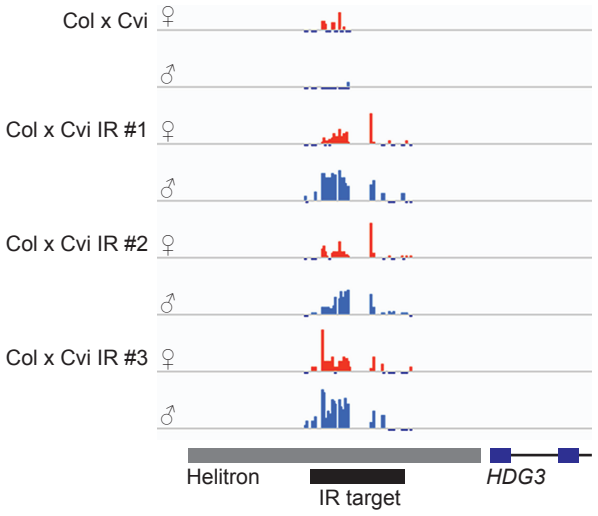


Fig 3. Gain of Cvi *HDG3* paternal allele methylation in endosperm. Total methylcytosine 5' of *HDG3* in F1 endosperm, determined by bisulfite-PCR. Maternally inherited Col allele in orange, paternally inherited Cvi allele in blue. Scale to 100%, tick marks below line indicate unmethylated cytosines. Col x Cvi data published in Pignatta et al., *eLife* 2014.

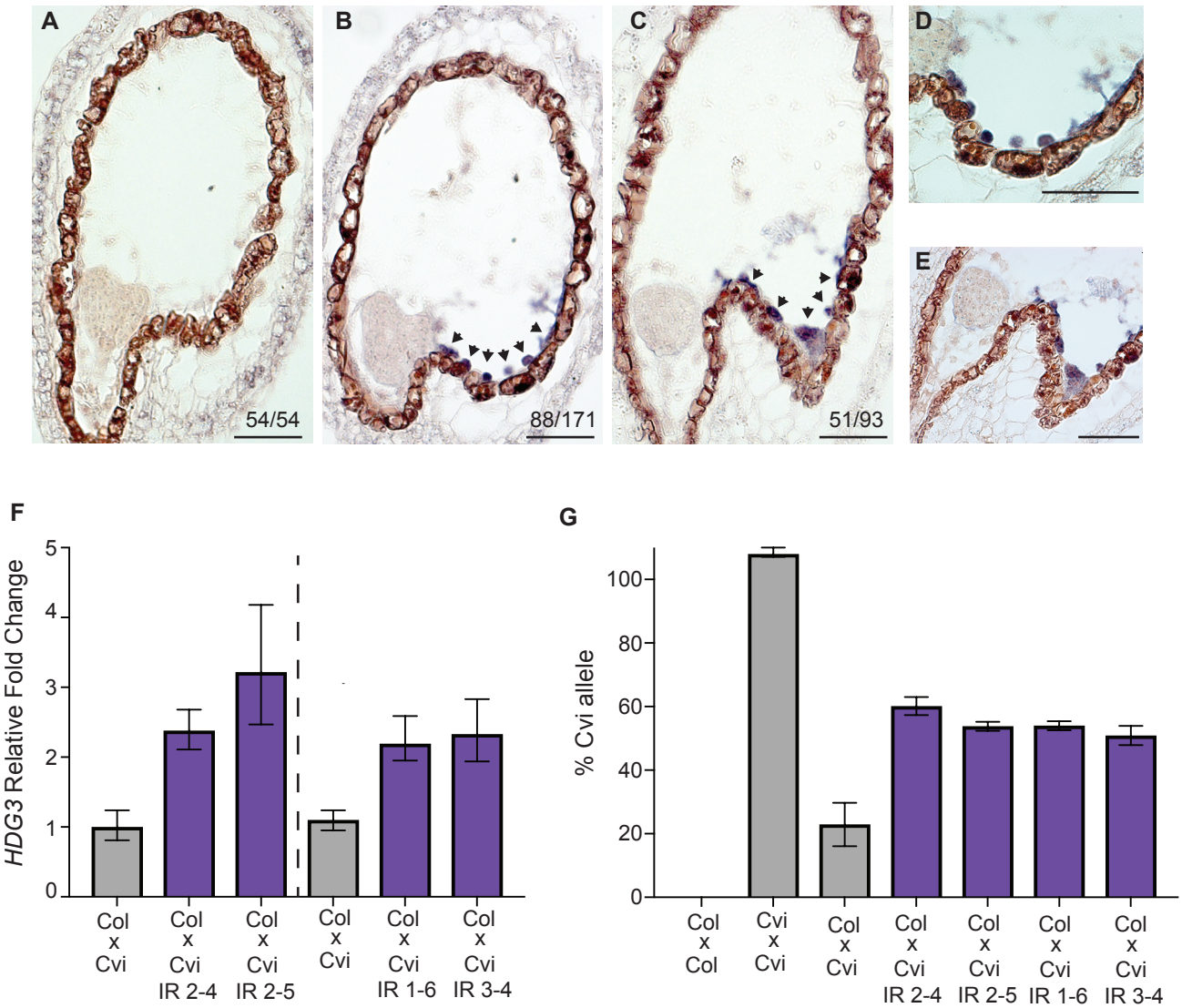


Fig 4. *HDG3* is imprinted in Cvi *HDG3* IR lines. *HDG3* in situ for (A) Col x Cvi (B) Col x Cvi *HDG3* IR #2-5 and (C) Col x Cvi *HDG3* IR #3-4. Arrowheads indicate regions of in situ signal. (D) Magnification of chalazal region in B. (E) Magnification of chalazal region in C. Scale bars, 50 μm. (F) RT-qPCR of relative *HDG3* transcript abundance in F1 endosperm at 6-7 DAP. Dashed line separates experiments done at different times. Left, avg of 3 technical replicates. Right, avg of biological duplicates. Bars show lower and upper range. (G) % of *HDG3* from Cvi allele in endosperm by TaqMan qRT-PCR assay.

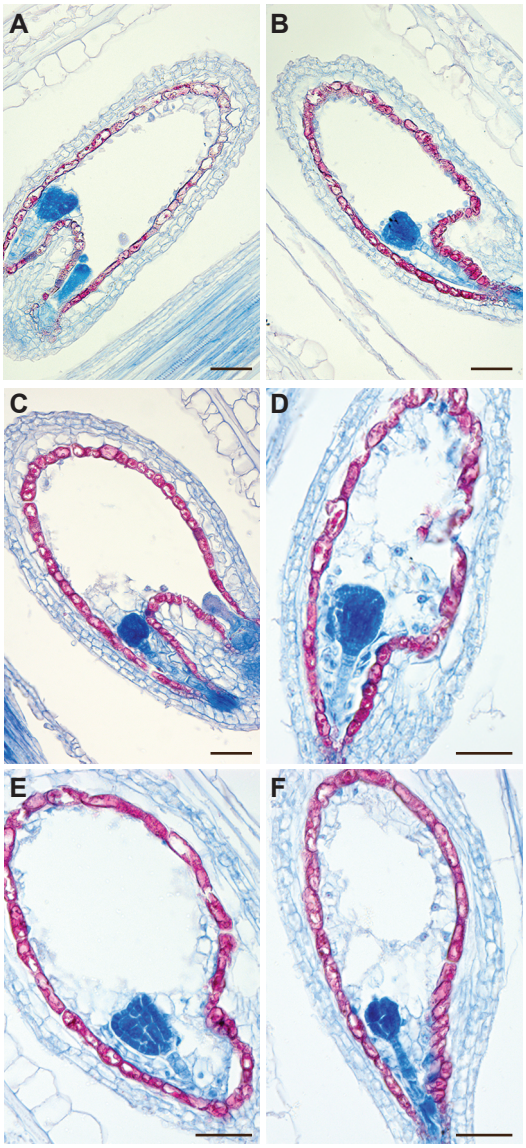
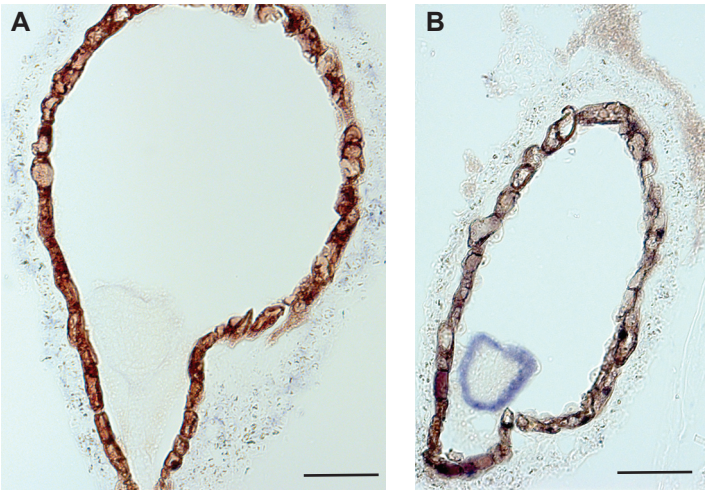


Fig 5. Effects of *HDG3* imprinting on seed development. Aniline blue and safranin O staining of seed sections at 5 DAP. (A-B) Col x Cvi (C-D) Col x Cvi *HDG3* IR #2-5 (E-F) Col x Cvi *HDG3* IR #3-4. Scale bars, 50 μm.



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.

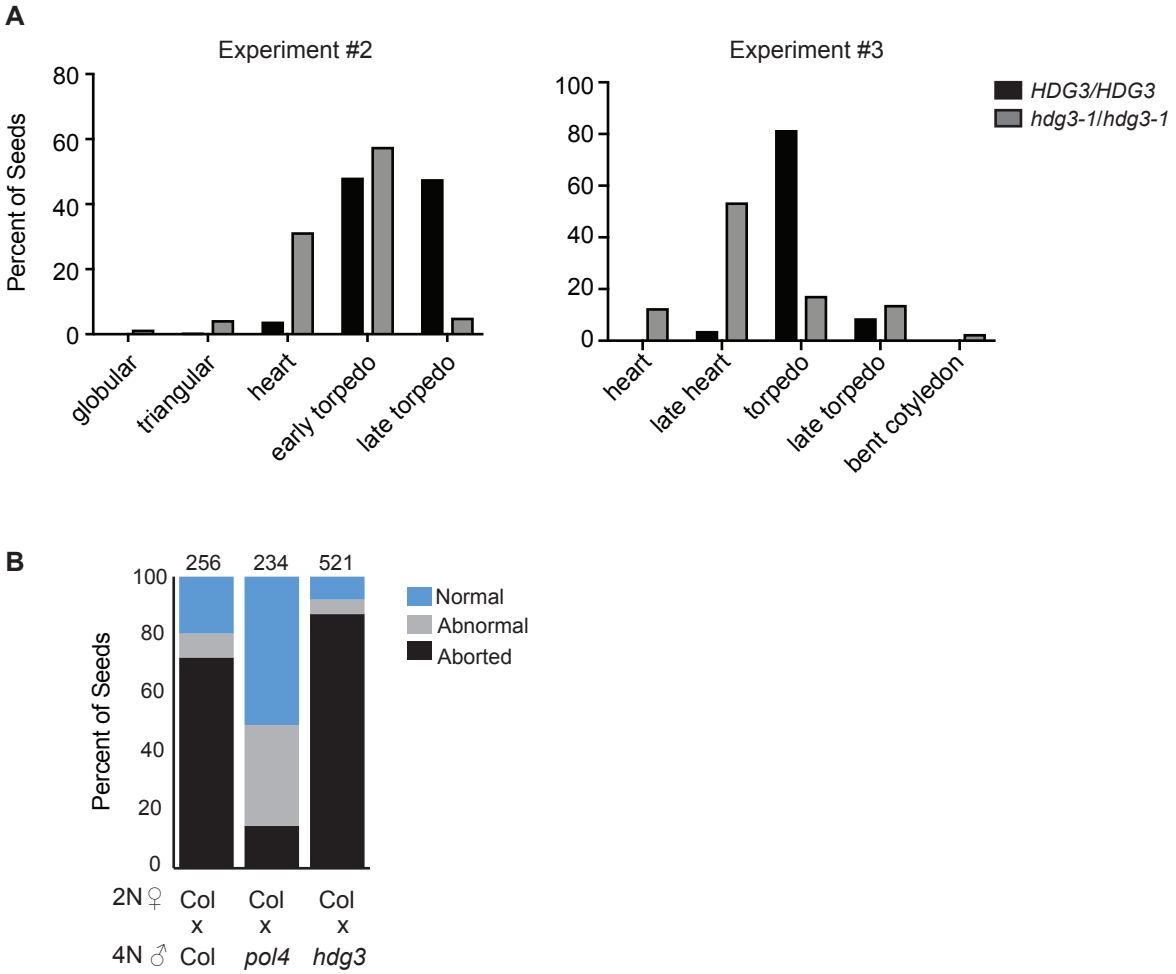
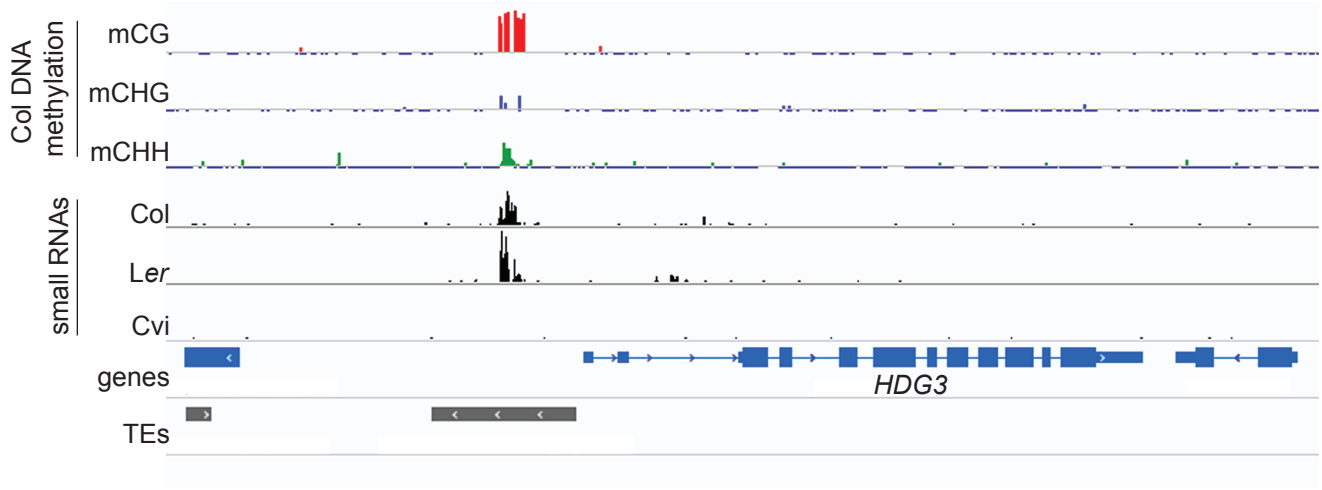
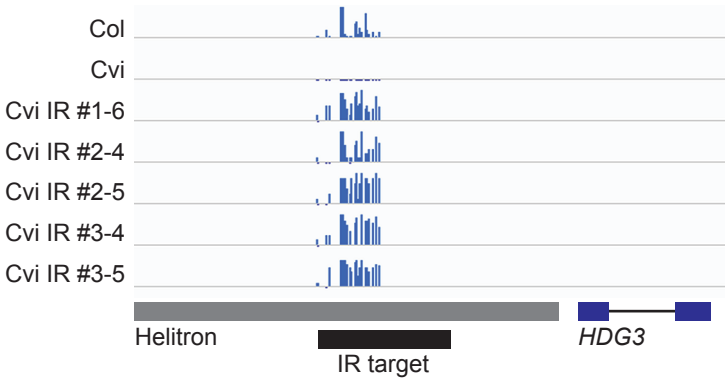


Fig S2. Phenotypic effects of low *HDG3* expression. (A) Additional replicates of *hdg3-1* seed stage phenotyping at 6 DAP. Seeds analyzed for Col are 386 for experiment #2 and 104 for experiment #3, for *hdg3-1* seeds analyzed are 137 for experiment #2 and 230 for experiment #3. (B) Tetraploid *hdg3-1* does not rescue interploidy seed lethality. 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

S3 Fig. Gain of *HDG3* methylation in Cvi 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.

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	7.0
		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	16.0	13.6	9.5
		Cvi	9	64.5	66.7	35.9
Col x Cvi IR #2	Endosperm	Col	31	31.0	14.0	5.7
		Cvi	21	18.0	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	30.0	24.0	17.4
		Cvi	23	86.1	69.6	40.6

S4 Fig. Bisulfite-PCR analysis of *HDG3* methylation in seeds. Female parent in cross listed first. IR lines are independent transgenic events. Cvi x Col and Col x Cvi data are from Pignatta et al., *eLife* 2014.

S3 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	