CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

"Análisis genético y genómico de la expresión de origen parental en el desarrollo embrionario de *Arabidopsis thaliana"*

Tesis que presenta

M en C. GERARDO DEL TORO DE LEON

Para obtener el grado de

Doctorado en Ciencias

En la Especialidad de

Biotecnología de Plantas

Director de la Tesis: Dr. C. Stewart Gillmor III

Irapuato, Guanajuato

CENTER OF RESEARCH AND ADVANCED STUDIES OF THE NATIONAL POLYTECHNIC INSTITUTE

UNIT IRAPUATO

"Genetic and genomic analysis of parent-of-origin gene expression in the *Arabidosis thaliana* embryo development"

Thesis presented by

M. Sc. GERARDO DEL TORO DE LEON

For obtain the degree of

Doctor of Philosophy (PhD.)

Specializing in

Plant Biotechnology

Advisor: C. Stewart Gillmor III, PhD

Irapuato, Guanajuato

TUTORIAL COMMITTEE

Dr. Cei Abreu Goodger

Dr. Luis Rafael Herrera Estrella

Dr. Alexander de Luna Fors

Dr. Juan Riesgo Escovar

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca de doctorado otorgada, sin la cual este proyecto no se hubiera podido llevar a cabo.

Al Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), por haberme dado la oportunidad de integrarme al programa de Doctorado en Ciencias.

Al Laboratorio Nacional de Genómica para la Biodiversidad (Langebio, Unidad de Genómica Avanzada), por haberme permito formar parte de su equipo a lo largo de estos años. Más lugares como Langebio en México.

Al Dr. Stewart Gillmor que me dio la confianza de integrarme a su laboratorio. Por sus todas sus enseñanzas tanto en el ámbito puramente académico como humano; por las discusiones científicas que llevamos a cabo, por su apoyo invaluable a lo largo de estos años. Por las "big-pachangas".

A todos los miembros de mi comité (en orden alfabético): Dr. Cei Abreu, Dr. Alex de Luna, Dr. Luis Herrera Estrella, Dr. Juan Riesgo Escovar. Por las nutridas discusiones acerca del proyecto, por todo su apoyo y orientación.

A todos los miembros del laboratorio de Stewart con los que me tocó convivir, especialmente la Dra. Marcelina García Aguilar, por sus consejos y enseñanzas en el ámbito científico que me acompañaron a lo largo de este tiempo. Por su amistad.
Sin olvidar a compañeros y amigos del Gillmor lab.

Todo esto no habría podido ser sin el apoyo incondicional de mi familia.

TABLE OF CONTENTS

Discussion

Perspectives

Supplementary material

Bibliography

Resumen

Las plantas con flor o angiospermas, generan sus semillas a partir de un proceso conocido como doble fecundación, dando lugar a un embrión diploide y a un endospermo triploide. En animales, las primeras etapas de desarrollo del cigoto se encuentran bajo el control de transcritos y proteínas de origen materno, depositados en la célula huevo antes de la fecundación. Este control materno del desarrollo temprano es importante, ya que asegura la viabilidad del embrión en estas etapas cruciales después de la fecundación, y es vigente hasta que el mando del programa de desarrollo es transferido al genoma cigótico, fenómeno conocido como Transición Maternal-Cigótica. Este proceso consiste en la Activación del Genoma Cigótico, donde el genoma del nuevo organismo adquiere un carácter de transcripción activa, y finalmente desplaza el control materno por medio de la degradación de productos heredados de la célula huevo. En plantas se ha documentado un proceso parecido al que ocurre en animales durante las primeras etapas del desarrollo del embrión. Aunque el tema ha sido ampliamente explorado, las conclusiones de diferentes estudios han sido controversiales. Por medio de estudios genéticos a nivel funcional, aunado a análisis transcriptómicos y de líneas reporteras, el presente trabajo ha aportado nuevos hallazgos que abonan al conocimiento de la activación del genómica cigótico en plantas. Los resultados de esta investigación sugieren que durante los primeros estadios del desarrollo del embrión existe un control mayoritariamente materno a nivel transcripcional y de función. Además, las observaciones demuestran que la dominancia materna durante la embriogénesis temprana, no solo es explicada por la falta de actividad de los alelos paternos, si no que en muchos casos se debe a una actividad restringida en comparación de los alelos maternos. Así mismo, este trabajo aporta nuevas observaciones que demuestran que la falta de consenso en parte puede ser explicado por la gran variación que se observa en este tipo de fenómenos en embriones de origen híbrido, lo que da indicios de un posible papel de este proceso en el establecimiento de fenómenos como la heterosis.

Finalmente, análisis de las regiones promotoras de los genes con expresión mayoritariamente materna durante los primeros estadios de desarrollo han llevado a la identificación de motivos enriquecidos que sugieren ser blanco de actividad represora o promotora de su comportamiento transcripcional, preferencialmente hacia alelos maternos.

Abstract

Fertilization produces a diploid zygote from two distinct haploid genomes, which then function together to direct embryo development. The beginning of large-scale zygotic transcription is referred to as zygotic genome activation (ZGA) (reviewed in Del Toro-De León et al., *COPB*, 2016). Using a functional analysis of early paternal gene activity, we have recently provided an explanation for a long-standing controversy regarding ZGA in plants (Del Toro-De León et al., *Nature*, 2014). We used embryo defective mutants to evaluate the ability of wild-type (wt) paternal alleles to complement phenotypes conferred by maternally inherited mutant alleles. Our experiments showed that in isogenic embryos, most paternal alleles are not fully functional immediately after fertilization, suggesting that the maternal genome makes the dominant contribution to early stages of development. We also demonstrated that hybrid genetic backgrounds have a significant effect on paternal allele activation.

Using a Col x Tsu hybrid transcriptome, we found an overwhelming bias toward maternal transcripts in the first 2 days of embryogenesis, with paternal contributions steadily increasing until the heart stage, when parental contributions are approximately equivalent. We hypothesize that these parent-of-origin effects on transcription in early embryos are the result of two processes: a) transcriptional activation via specific *cis* regulatory motifs (as well as *trans* acting factors), and b) silencing of paternal alleles due to epigenetic marks established in the gametes. Further results with reporter lines show that the mechanism that lead to the parental bias during early embryogenesis is rather complex and more than one genomic mechanism probably acts to regulate maternal and paternal contributions during early seed development.

Introduction

Seed formation in flowering plants begins with the double fertilization of the female gametophyte. In most flowering plants, the female gametophyte harbors two gametic cells: the egg cell (n) which gives rise to the diploid embryo, and the homodiploid central cell which will form the endosperm (3n) (Dresselhaus *et al*, 2016). Embryo development progresses surrounded by the nurturing endosperm tissue, which in coordination with the seed coat will form the mature seed (Figueiredo & Köhler, 2016). The endosperm does not contribute genetically to the following generation, instead it is consumed during embryogenesis, or just after seed germination.

In animals, early embryo development is regulated by stored maternal products in the egg cell, such as mRNAs and proteins. This transient period of maternal control varies in length depending the species and culminates when the zygote or early embryo takes control of its own development (Tadros & Lipshitz, 2009). This phenomenon, known as the maternal-to-zygotic transition (MZT), involves the degradation of stored maternal products and the onset of Zygote Genome Activation (ZGA) (Baroux *et al*, 2009).

In flowering plants, the time of onset of ZGA and the extent of transcripts in the zygote inherited from the gametes are not fully known (Baroux et al., 2001; Grimanelli et al., 2005; Kohler et al., 2005; Pillot et al., 2010; Sørensen et al., 2001; Weijers et al., 2001). This lack of consensus is partly explained by the complexities of transcriptome profiling of the plant embryos, which are embedded in the seed. In addition, transcripts inherited from the gametes are difficult to discriminate from those synthetized *de novo* in the zygote. Consequently, studies of early zygotic transcription have largely made use of reporter lines and RT-PCR studies. Examples of both immediate and delayed paternal activation of marker lines have been documented (Vielle-Calzada *et al*, 2000; Weijers *et al*, 2001; Baroux *et al*, 2001).

The endosperm as well as the embryo exhibit parent-of origin gene expression, such that expression is primarily from the maternal or paternal allele, indicating that the maternal and paternal genomes are not totally equivalent during seed development. We can distinguish at least

two classes of parent-of-origin gene expression in seed development: canonical imprinting (predominantly in the endosperm), and gene expression biases during zygotic genome activation (García-Aguilar & Gillmor, 2015). Imprinting refers to long-term preferential expression of the maternal or paternal allele due to differential epigenetic modifications established during female and male gametogenesis. Depending on the allele bias, the imprinted genes are classified as maternally expressed genes (MEGs) or paternally expressed genes (PEG). A few hundred loci have been categorized as imprinted, most of them in the endosperm (Gehring, 2013). Imprinting also occurs during embryogenesis; a maternally expressed gene has been identified in maize, and eleven genes in Arabidopsis interestingly only one paternally expressed gene was confirmed (Raissig *et al*, 2013; Jahnke & Scholten, 2009). Parent-of-origin gene expression observed during ZGA is a transient bias in the expression towards the maternal genome, where the paternal genome shows limited contributions and is gradually activated. ZGA is a more genome-wide phenomenon that occurs in a short period during early embryogenesis. Imprinting occurs in fewer genes, and extends to later stages of seed development (García-Aguilar & Gillmor, 2015).

Genomic parent-of-origin studies in seeds

Allele Specific Expression (ASE) studies of imprinting and ZGA have taken advantage of Single Nucleotide Polymorphisms (SNPs) that exist between different alleles. SNPs have been the basis to quantify Allele Specific Expression in heterozygotes and hybrid organisms. First conceived to study imprinting (Singer-Sam *et al*, 1992) and then used to asses *cis*-acting regulatory variation (Cowles *et al*, 2002; Yan *et al*, 2002), the use of SNPs has expanded our knowledge of differential allele gene expression in different biological contexts and model organisms. Nevertheless, studies before the genomic era were limited by the scope of the experimental methods used and only a few genes were successfully described in terms of their ASE. Thanks to recent genome-wide methods, allele specific expression studies have advanced rapidly in the last few years.

Genomic studies using SNPs to study ASE have tried to elucidate at least three different biological phenomena regarding to parent-of-origin expression: imprinting, zygotic genome activation (ZGA) and heterosis. Even though each of these three phenomena are related to a different biological question, they share the same concept of allele specific expression.

Genomic imprinting studies in plants have used different hybrid combinations from crosses of a variety of strains. Interestingly, variation has been detected among studies: some genes show imprinted expression in certain genotypes, but not in others. In maize, a genome-wide comparison of reciprocal crosses of 4 highly polymorphic genotypes detected that 12% of the identified imprinted genes in maize, showed allelic variation for imprinting in intraspecific crosses, interestingly the same report compared the conservation of imprinting between maize and rice and found conservation of imprinting of synthetic orthologs in only about 11% of the genes (Waters *et al*, 2013). In *Arabidopsis,* disparity of imprinting due to genotype in intraspecific crosses was calculated to be a maximum 11% of the total number of imprinted genes. This variation was apparently related to Differential Methylated Regions (DMRs). This epiallelic diversity in genomic imprinting was proposed to be a potential source of novel gene expression patterns that might have impacts on phenotype variation in hybrid seeds (Pignatta *et al*, 2014). Previous comparisons of genome wide imprinting studies have shown lower conservation of imprinting among flowering plants, yet this variation may be due to experimental and biological variables. Genes that conserve imprinting expression in divergent species, exhibit functions that suggest important roles for imprinted genes in development (Hatorangan *et al*, 2016).

In the case of ZGA and parent-of origin gene expression in *Arabidopsis*, only two genomic studies have reported ASE profiles during early embryogenesis. These studies reached different conclusions regarding the extent of maternal and paternal contributions to early embryogenesis. The first transcriptome analysis –using the Col x L*er* hybrid combination– showed a clear maternal transcript bias in 2-cell embryos, with paternal transcript contribution increasing by the globular stage (Autran *et al*, 2011). In contrast, the second study, using the Col x CVI combination, showed virtually all genes with perfectly biallelic transcripts at all stages evaluated (only 77 maternally and 45 paternally enriched genes were detected) (Nodine & Bartel, 2012). Although maternal tissue contamination was argued to be the source of the biased maternal expression in the first study (Nodine & Bartel, 2012), this was later ruled out as an explanation (Baroux *et al*, 2013). The Col x L*er* transcriptome lacked a reciprocal cross, so ecotype dominant expression could not be discriminated from parent-of-origin expression. The Col x CVI

transcriptome did have reciprocal crosses, so that ecotype effects could be distinguishing from parental effects.

Mechanisms regulating parent-of-origin expression in early embryogenesis

Imprinting in the endosperm is regulated by at least two mechanisms. One relies on genome-wide methylation in the gametes, followed by demethylation of maternal alleles for specific genes in the central cell, leading to expression of the maternal allele in the endosperm. A second mechanism relies on the POLYCOMB REPRESSIVE 2 (PRC2) complex to silence either the paternal allele (leading to a MEG), or the maternal allele (leading to a PEG) (reviewed in Gehring, 2013). Imprinting of genes in the embryo can occur by PRC2 and non-PRC2 dependent mechanisms (Raissig *et al*, 2013).

Regulation of parental contributions to ZGA has been attributed to maternal repression of the paternal genome, mediated by the RNA-directed DNA methylation (RdDM) pathway (Autran *et al*, 2011). The RdDM pathway is the major mechanism of Transcriptional Gene Silencing (TGS) mediated by small-interfering RNA in plants (Matzke & Mosher, 2014). Initially described to silence transposable elements, RdDM also targets coding regions (Wierzbicki *et al*, 2008). Two plant-specific RNA polymerase enzymes are required for this pathway, making this small RNA-mediated silencing unique to plants. Evidence suggests that these enzymes evolved from an ancient RNA polymerase II (Pol II) (Tucker *et al*, 2010). Canonical RdDM silencing begins with Pol IV transcripts that are acted on by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) to produce double stranded RNAs, which are fragmented by DICER-LIKE 3 (DCL3) into 24-nucleotide siRNAs, followed by methylation at their 3' ends by HUA ENHANCER 1 (HEN1). These 24-nucleotide siRNA are loaded onto ARGONAUTE 4 (AGO4), and target nascent scaffold transcripts synthesized by Pol V by sequence complementarity. This targeting of Pol V transcripts recruits DNA methyltransferases to mediate *de novo* methylation of cytosines in all contexts (CG, CHG and CHH, where H represents A, C or T), triggering transcriptional gene silencing (Matzke & Mosher, 2014).

Interestingly, 24nt sRNAs have been detected in mature ovules (Autran *et al*, 2011), suggesting that active TGS mediated by RdDM methylation could silence the paternal genome after

fertilization. Consistent with this hypothesis, embryos that are maternally mutant for the RdDM pathway show precocious activity of paternal reporter genes (Autran *et al*, 2011). Mutants in the Histone 3 Lysine 9 methyltransferase SUVH4 (also known as KRYPTONITE (KYP)) also show increased paternal transcripts compared with wt (Autran *et al*, 2011). Several lines of evidence have revealed interdependence between non-CG DNA methylation and H3K9 heterochromatin. KYP specifically methylates histone H3 Lys 9 residues, but *KYP* mutants also show a reduction in the CHG methylation, reminiscent of DNA methylation mutants such as CMT3 (Jackson *et al*, 2002; Stroud *et al*, 2013). Structural studies have shown that DNA methyltransferases CMT3 and CMT2 bind to H3K9me marks deposited by KYP (Stroud *et al*, 2014; Du *et al*, 2012). Conversely, KYP binds specifically to non-CG methylated DNA (Du *et al*, 2014). All these together strongly support the model in which histone and non-CG DNA methylation form a feedback loop to reinforce silencing of their targets, suggesting that TGS by RdDM, and histone methylation, function as genomic regulators of parental contributions to ZGA in Arabidopsis.

In addition to repression of paternal transcription, activation of paternal allele expression is promoted by histone variant exchange mediated by the CAF1 complex. Maternal mutants of members of the CAF1 complex (FASCIATA1 (FAS1), FAS2 and MSI) show delayed expression of paternal reporter lines (Autran *et al*, 2011). Since the CAF1 complex acts as an H3/H4-specific chaperone facilitating nucleosome assembly during replication (Hennig *et al*, 2005; Kaya *et al*, 2001), its maternal functions are believed to activate transcription of paternal genome via histone turnover, by incorporating histone variants that promote transcriptional activity.

As in animals, where maternal dominance of early embryo development occurs by the passive transmission of maternal products such as proteins and mRNAs (Tadros & Lipshitz, 2009), a similar process is believed to occur in plants, however this hypothesis has not been successfully tested (Ueda *et al*, 2017).

Biological significance of parent-of-origin expression during ZGA in plants

Female gametophyte mutants defective in post-fertilization processes are not rescued by the paternal (pollen) allele, suggesting a functional requirement for maternal contributions beyond gametogenesis (Pagnussat *et al*, 2005; Ngo *et al*, 2012). Recent reports have demonstrated that

transient maternal dominance of early embryogenesis has important implications for embryo development. Luo *et al.*, showed that in early embryos of tobacco interspecific hybrids, Programed Cell Death (PCD) in the suspensor is regulated by maternal expression of the *NtCYS* cystatin gene. Although both parental transcripts contribute to PCD, the higher expression and subsequent reduction of the maternal allele was necessary for the initiation of PCD in the suspensor and proper progression of embryogenesis (Luo *et al*, 2016). Moreover, He *et al.*, showed that *de novo* transcription of the maternal *NUWA* allele in the zygote and early embryo was essential for mitochondrial function. Interestingly, the paternal *NUWA* allele was first detected only after 16-cell stage (He *et al*, 2017). This observation is in agreement with the hypotheses that after a transient period of maternal transcript enrichment at the zygote and octant stages, most of the maternally biased genes become biallelic during the globular to the heart stage as showed by a recent globular-embryo ASE transcriptome (Alonso-Peral *et al*, 2016). Interestingly, in a recent ASE study of hybrid embryos, from reciprocal C24 x Ler crosses, at dermatogen-globular and heart stage, most genes had equal proportions of transcripts from each parent. This is consistent with observations that in globular embryos most of the parent-of-origin effects have been reprogrammed to biallelic expression (Autran *et al*, 2011; Del Toro-De León *et al*, 2014; Ueda *et al*, 2017; He *et al*, 2017). In addition, recent studies have shown that the asymmetric division of the zygote is maternally regulated. WOX8, a master regulator of early embryogenesis involved in suspensor development, is activated by the maternally derived transcription factors HOMEODOMAIN GLABROUS11/12 (HDG11/12). HDG11/12 activity is only expressed maternally at the zygote stage, thus early embryogenesis development is controlled maternally in *Arabidopsis thaliana*, suggesting a critical importance of maternal factors for early seed development (Ueda *et al*, 2017).

Background

To functionally evaluate paternal genome activation, I performed a genetic test using embryo defective mutants as maternal parents and evaluated the capacity of wt paternal alleles to complement maternally inherited mutations. In most 80% of the cases, paternal alleles failed to immediately complement maternal mutations in early embryos, showing gradual complementation later in development. Nevertheless, for 20% of the tested embryo defective mutants (*emb*), the wt paternal allele was active enough (as early as 2 dap) to complement the mutation inherited by the mother. This functional assay showed that paternal allele activation in early embryogenesis is variable, with the majority initially showing insufficient activity, and only a minority of genes showing enough activity early on to complement the lack of maternal function (Del Toro-De León *et al*, 2014).

Given the fundamental importance of embryo development to many aspects of seed traits in crops, it is critical to understand how parental genomes are programmed to regulate gene expression during embryogenesis. For example, performance of reciprocal crosses show significant differences in offspring (F1) phenotypes despite their identical nuclear parental genomes (Swanson-Wagner *et al*, 2009; Ng *et al*, 2014). This suggests that parent-of-origin gene expression plays an important role in hybrid vigor. Hybrid genomes show allelic interactions that may lead to changes in gene expression (Greaves *et al*, 2012) and variation in parent-of-origin phenomenon (Ng *et al*, 2014).

Here we report that hybrids show high level of variation in the timing of ZGA according with functional allele activation experiments. To make unbiased statements regarding ZGA in plants, it is critical to find the hybrid combination most similar to isogenic Col. By assessing functional complementation kinetics, using paternal alleles derived from several Arabidopsis ecotypes, we identified the Col x Tsu hybrid to be closest to isogenic Col. ASE analysis of embryo development for the Tsu x Col combination shows that ZGA occurs during early embryogenesis, with expression initially biased towards the maternal genome, transitioning to biallelic expression by the globular and heart stages. In addition, analysis of reporter lines provides new insights into mechanisms regulating parental contributions during early embryo development.

7

General Objective:

Investigate the activation of the maternal and paternal genomes by allele specific sequencing, *cis* element analysis, reporter lines, and functional studies.

Specific Aims:

1) Evaluate variation in maternal and paternal functional contributions in isogenic and hybrid embryos.

2) Identify *cis* regulatory regions involved in differential gene expression, by analyzing transcriptomic data from early embryos.

3) Determine the parental contributions during embryo development from Col x Tsu hybrid embryos.

4) Investigate temporal activation of the maternal and paternal genome by functional and reporter-line analysis.

Materials and Methods

Part 1

Plant growth conditions. Plants were grown in growth chambers at $22 \pm 2^{\circ}$C in 16-h-light/8-h-dark cycles at constant humidity rate and 100μ/m2. The substrate used for plant growth is a mixture of vermiculite (GRACE MAN-FIN), perlite (AGROL125) and sunshine mix (PREMEZ FWSS3) (1:1:3 v/v/v). Seeds were sown directly on the soil, and given a 3-day cold treatment to promote germination, before being moved to growth chambers.

Plant materials. The list of EMB mutant lines used in this section are listed in **Table m&m 1**. Ecotype accession numbers are listed in **Table m&m 2**.

Crosses, histology and microscopy (Part 1 and Part 2). Heterozygous plants were identified based on segregation of *emb* phenotypes (Del Toro-De León *et al*, 2014). Crosses were performed using an *emb/+* plant as female parent and wt plants as the male parent (**Figure 2-3**). Heterozygous plants were emasculated, and pollinated 2 days later. The seeds were cleared with a modified Hoyer's solution consisting of 7.5g gum arabic, 100g chloral hydrate, 5ml glycerol, and 30ml water, diluted 1:1 with a dilution solution consisting of 7.5g gum arabic, 5ml glycerol, and 30ml water. Siliques were collected at 2, 3 and 5 dap, and the ovules were cleared and observed as mentioned above. Embryos were observed on a Leica DM6000 B Nomarski microscope. For each cross, at least two biological replicates are reported, meaning crosses were performed with at least two different *emb/+* plants as mothers. Similar procedures were followed for the *nf-yb2* mutant line.

GUS staining analysis. *pCCT::GUS* and *gGCT-GUS* plants were emasculated, and pollinated 2 days later using wt Col pollen. The reciprocal crosses were conducted the same way. Developing siliques at 1, 2, 3, 4 and 5 dap were dissected to expose the ovules, ovules were punctured with forceps, and incubated in permissive GUS staining buffer (0.2% Triton X-100, 2mM potassium ferrocyanide, 2mM potassium ferricyanide, 2mM X-Gluc in N,N-dimethylformamide (DMF) in 50mM sodium phosphate buffer pH7.2) overnight in a humid chamber at 37°C. In the case for *pNF-YB2::GUS* (**Figure 15**), 4 days of staining treatment were employed. The tissue was cleared in 50% glycerol (or Lactic acid-glycerol 25%/25%) and observed on a Leica DM6000 B microscope under Nomarski optics.

Construction of *GCT* and *CCT* GUS reporters. The *gGCT-GUS* translational construct and *pCCT::GUS* transcriptional reporter were made by Stewart Gillmor. For details about the construct, see Del-Toro, García-Aguilar & Gillmor, *Nature*, (2014).

Statistical analysis. To determine whether GUS staining differed between the treatments, or mutant phenotype segregation ratios differed significantly, Fisher's two-tailed Exact Test (www.langsrud.com/fisher.htm) was used (**Supplementary Tables)**. Heat maps were made in R (R Development Core Team) with the heatmap.2 function of the data visualization package gplots 2.11.0; hierarchical clustering of rows was performed using the Ward method, except for **Figures 2**, and **11** where data were ordered according genotype and **Figure 9** where the order is indicated by the tissue. Principal component analysis was performed in the R package using the prcomp function. In order to quantitatively illustrate the relationships among the variables (ecotypes), percentages of the variance explained by each principal component were calculated as shown in **Figure 3**.

Part 2

Col x Tsu transcriptome. The transcriptomic data was provided by Raju Datla's Laboratory. Sequencing data were generated and mapped in the Datla laboratory according to the following methodology: Reciprocal crosses were performed between Col x Tsu ecotypes, zygotes and embryos were washed as in Xiang et al., (2011), developmental stages confirmed by microscopy, and RNA was extracted and amplified from embryos. Libraries were generated for zygote, octant, globular, heart, torpedo, bent and mature embryo stages (**Supplementary Figure 2**). Libraries were sequenced to a depth of 18-48 million 100-bp reads, and mapped to the *Arabidopsis* genome using the Tair9 assembly version. Maternal and paternal alleles were distinguished using SNP profiles of the parental genomes.

Parent-of-origin analysis of the transcriptome. The data were analyzed within R version 3.4.0. For Maternal bias quantification, percentages of maternal reads were calculated for each gene. Statistical differences between maternal and paternal read counts for each gene were calculated by Pearson's Chi-square Test, and the resulting *p*-value was assigned. Categorization of parental bias was performed as indicated in **Table 1**. Graphs were generated with data visualization package gplots 2.11.0; heat maps were generated by hierarchical clustering of rows using the

Ward method. Identical analysis was performed for the data published by Nodine & Bartel (2012).

Gene Ontology Terms analysis. This analysis were performed using the "GOstats" package for R (Falcon & Gentleman, 2007).

Discovery motif analysis. *De novo* motif analysis was performed using the MEME Suite of motif-based sequence analysis tools (http://meme-suite.org/) and sequences corresponding to 2 or 1kb upstream regions from the translation start site (or until the previous gene). The three different models of motif distribution (oops, zoops and anr) were used, however motif analysis showed better performance using 1kb upstream region sequences and the *anr* motif distribution model. A consensus matrix of each overrepresented motif was calculated with all the genes where significant occurrences were detected (**Table m&m 3**). Motif occurrences were calculated by using FIMO from MEME tools using the consensus matrix.

Comparisons with other transcriptomic profiles. The reference for female gametophyte microarray was published by Wuest, et al. (2010). For the seed compartment reference, the microarray published by Belmonte, et al. (2013) was used. Only shared loci between the Col x Tsu and female gametophyte microarray were compared (common universe).

Gene expression. Gene expression analyses were performed using normalized transcriptomic data for Col x Tsu provided by Raju Datla's laboratory.

Contamination test. "The contamination test" was performed as indicated by Schon & Nodine, (2017). This test was created to evaluate the enrichment of genes from maternal tissues such as different sub-regions of the endosperm and seed coat, in the embryo transcriptomes. By using this tool, it is possible to quantify the levels of contamination from other tissues, in the samples of embryos.

Functional complementation test. As described in Part 1.

Part 3

Plant material: The *pNF-YB2::GUS* (CS67020) reporter line was obtained from the Arabidopsis Biological Resource Center (ABRC) (Siefers *et al*, 2009). Insertion lines used in this study are listed in **Table m&m 4**. Growing conditions similar to the **Part 1**.

GUS staining analysis. *pNF-YB2::GUS* reciprocal crosses were performed as in **Part 1**.

Introgression of *pNF-YB2::GUS* in mutant backgrounds. The *pNF-YB2::GUS* reporter line was crossed with homozygous mutant plants (confirmed by PCR) for *kyp-4* and *nrpd1b-11*. The F1 generation was grown and selfed. The F2 generation was analyzed for GUS staining in leaves, and tested by PCR for the corresponding mutation. Homozygous plants for the mutations were identified. Homozygous segregation of the *pNF-YB2::GUS* insertion were confirmed by BASTA resistance segregation in MS plates with 10μg/mL of glyphosinate-ammonium, and by segregation of GUS stained embryos.

Parent-of-origin expression analysis of *pNF-YB2::GUS* in mutant backgrounds. *pNF-YB2::GUS* plants were emasculated, and pollinated 2 days later using *kyp-4*, *nrpd1b-11*, *fie* and *fas2-3* mutant pollen. Developing siliques at 2, 3, 4 and 5 dap were dissected to expose the ovules, ovules were punctured with forceps, and incubated in permissive GUS staining buffer (0.2% Triton X-100, 2mM potassium ferrocyanide, 2mM potassium ferricyanide, 2mM X-Gluc in N,N-dimethylformamide (DMF) in 50mM sodium phosphate buffer pH7.2) for 4 days in humid chamber at 37°C. The tissue was cleared in 50% glycerol (or Lactic acid-glycerol 25%/25%) and observed on a Leica DM6000 B microscope under Nomarski optics. Similarly, GUS staining analysis were performed using the double mutants generated as paternal donors in Col wt maternal plants.

Construction of *EMB2768* tdTomato reporter. Two destination Gateway vectors including nuclear localized tandem-dimer Tomato or 3xVenus reporters were created as indicated in the **Figure m&m 1**. I took as a gene source the binary vectors donated by Wolfgang Lukowitz: pCN-SENB-dBox-nu2T and pCN-SENB-dBox-nu3V. The reporters where excised with restriction enzymes SpeI for 2T and AvrII (NEB) for 3V and MauBI (Thermo Scientific) and ligated to the destination gateway vector pMDC123 using T4 DNA Ligase (Thermo Scientific). Two new gateway destination vectors were obtained: pMDC123-nu3V-NOST and pMDC123-nu2T-NOST Although both plasmids showed very good fluorescence signal in transient tobacco expression with the constitutive 35S promoter (**Figure m&m 2**), only the tdTomato showed good expression in Arabidopsis. Upstream sequences of the genes listed in the **Table m&m 5** were cloned into pMDC123-nu2T-NOST and transformed in Arabidopsis. The floral dip protocol (Clough & Bent, 1998) was employed for plant transformation, using Col wt plants. T1 plants were analyzed and experiments were performed in T2 generations. pMDC123-nu2T-NOST was suitable for transcriptional gene reporter expression.

Parent-of-origin expression of *pEMB2768::tdTomato* reporter. T2 plants were analyzed for embryo expression of the reporter gene. Only heterozygous plants were detected. Heterozygous plants were emasculated and pollinated 2 days after with wt pollen. The reciprocal crosses were conducted the same way. Developing siliques at 1, 2, 3, 4 and 5 dap were dissected, ovules were punctured with forceps and mounted in 50% glycerol glass microscope slides and observed on a Leica DM6000 B microscope under Nomarski optics.

RESULTS

Part 1. Functional genetic analysis shows non-equivalent contributions of parental genomes during embryogenesis and high variation in hybrids

Summary

Sexual reproduction in flowering plants begins with the double fertilization of the female gametophyte, which results in the formation of the endosperm –a nutritive tissue reminiscent to the placenta in mammals- and the embryo. Comprehensive studies have documented that the parental genomes do not contribute equally in terms of gene expression, especially during the first stages of seed development. In endosperm, a parent-of-origin pattern of expression called imprinting has been extensively studied in which several genes show almost monoallelic expression either from the maternal or paternal alleles throughout endosperm development (García-Aguilar & Gillmor, 2015). In the case of embryo development, consensus of the extent of parent-of-origin gene expression has been hard to reach, partly due to the ephemeral developmental window where it is observed and to apparently contradictory results from genomic studies (Del Toro-De León *et al*, 2016). Here we provide genetic and molecular evidence suggesting that during early embryo development, the paternal genome shows a delayed expression and thus, activity. In addition, we demonstrate that parent-of-origin expression during early embryogenesis shows high levels of variation in hybrid embryos –the basic material of ASE sequencing methods– partly explaining the variation observed by previous genomic studies.

Delayed paternal allele activity correlates with expression

Previously we reported that paternal alleles from wt plants of *A. thaliana* were unable to fully complement mutant alleles inherited by the mother during the first stages of embryo development, suggesting a delayed paternal functional activation for ~80% of the *emb* mutant lines analyzed. Although the data indicate that there is a delayed activity of the paternal allele exhibited by the presence of mutant phenotypes, the next question is whether the delayed activity is a consequence of the absence of paternal allele expression. For this purpose, *pCCT::GUS* and *gGCT-GUS* constructs were used to analyze the reporter parent-of-origin expression pattern
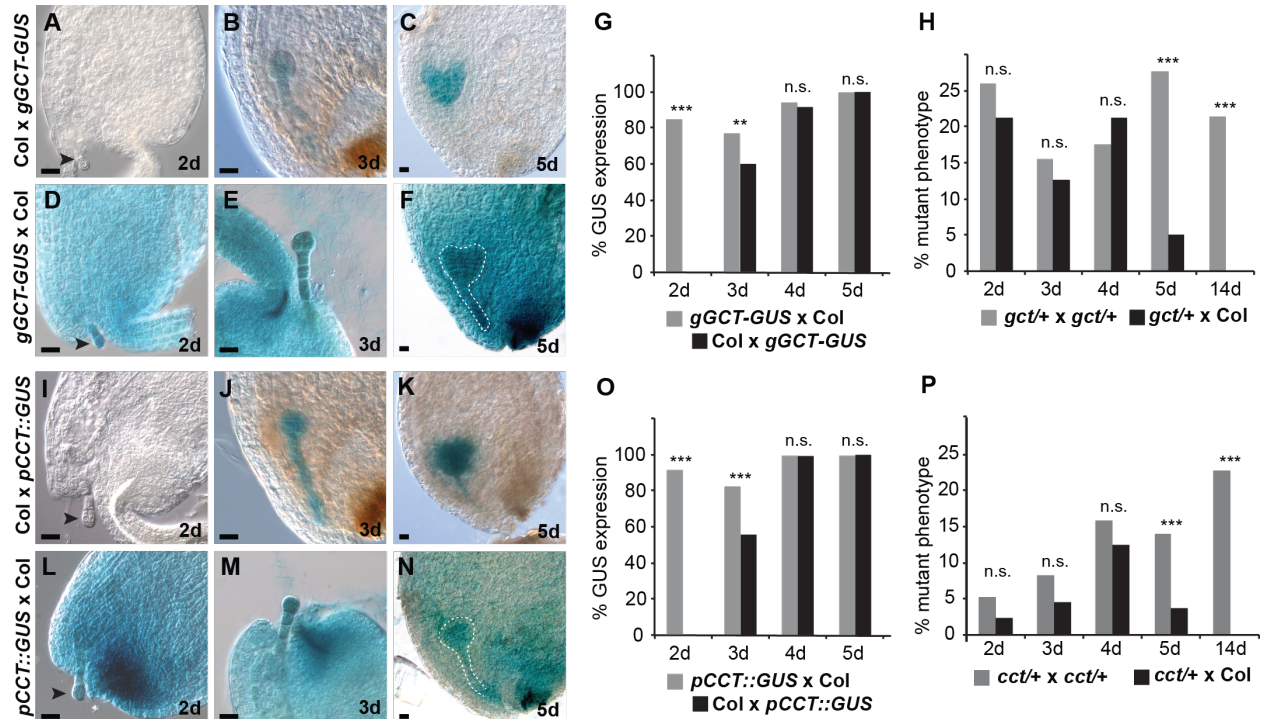
Figure 1. **Delayed expression and functional activity for *GCT* and *CCT* paternal alleles in early embryogenesis**. **A–G**, Expression of paternally and maternally derived *gGCT-GUS* at 2, 3, 4 and 5 days (d) after pollination. **H**, Quantification of phenotypically mutant embryos derived from hand- pollinated selfed gct/1 plants, or crosses with WT pollen (n >90 for all time points). **I–O**, Expression of paternally and maternally derived *pCCT::GUS* at 2, 3, 4 and 5 days after pollination. **P**, Quantification of phenotypically mutant embryos derived from hand-pollinated selfed *cct/+* plants, or from crosses with WT pollen. For all crosses, the maternal genotype is listed first. Embryos in **A**, **D**, **I** and **L** are indicated with arrowheads. Embryos in **F** and **N** indicated with a dotted line. Scale bar, 20 μm. Results of a two tailed Fisher's exact test, performed to determine whether the two samples were statistically different at each time point, are shown above the columns in **G**, **H**, o and **P**: ***p* < 0.001; ***p* < 0.01; NS, no significant difference between the two samples. For each time point, the combined number of embryos examined in at least two biological replicates, as well as exact *p* values for results of two-tailed Fisher's exact test, are listed in Supplementary Table 1. Original figure from Del Toro, García-Aguilar&Gillmor, 2014.

during early embryogenesis. Paternal complementation of both genes was gradual, showing significant proportions of mutant embryos at 2, 3 and 4 dap when a heterozygous plant was pollinated with a wt relative and fully complemented after early heart stage embryo (**Figure 1 h and p**). According with the mutant phenotype of *CCT* and *GCT* during early embryogenesis, self-pollinated plants of the reporter constructs showed *GUS* expression starting at preglobular stages (**Supplementary Figure 1**). Reciprocal crosses were performed between wt plants and the reporter fusions. When the *CCT* and *GCT* reporter lines were used as maternal parents, the resulting F1 embryos showed GUS expression from the 1-cell stage, the same as self-pollinated plants. In contrast, when the reporter constructs were used as male parents, *GUS* expression was not found in any of the F1 seeds at 2dap (**Figure 1 a-f and i-n**), confirming that the paternal allele is not expressed. Paternal expression of the GUS reporter gene is present at 3dap but at considerably lower proportions than the maternal expression, namely some embryos show *GUS*

but in less percentages than wt. Embryos from 4dap onwards, show full activity of paternal GUS expression. Interestingly, the functional analyses show that the mutant phenotypes were almost totally complemented at 5dap for both *cct/+* and *gct/+*, suggesting a correlation between the delayed expression of the paternal alleles with its functional activity (**Figure 1, Supplementary Table 1**). Namely, in our functional assay the delayed complementation of maternally inherited mutations by wt paternal alleles coincides with delayed expression of the paternal alleles. Thus, both, molecular and functional genetic analysis showed a gradual increase of expression and activity of the paternal genome.

Gene activation kinetics differ in hybrid embryos

Our genetic and molecular analyses of parent-of-origin expression during embryogenesis suggest a delayed activation of the paternal genome during early embryogenesis for most of the genes tested. Whether this phenomenon occurs at the genomic scale has been a matter of controversy (Baroux & Grossniklaus, 2015). Since SNPs have been the basis of methods to quantify Allele Specific Expression, hybrids are necessary to evaluate maternal and paternal differential expression. So far, there are two studies of ASE analysis during early embryogenesis, using different hybrid combinations. The first transcriptome analysis published used Col and L*er* hybrid embryos and showed a clear maternal transcript bias in 2-cell stage embryos, with paternal transcripts increasing one day later (in the globular stage embryo) (Autran *et al*, 2011). (Autran *et al*, 2011). A second experiment, also using an SNP based allele specific transcriptome, but in this case with Col and Cvi hybrid embryos, observed an equivalent transcript contribution from the maternal and paternal alleles, regardless of the stage of embryogenesis. Only 77 maternally and 45 paternally enriched genes were identified (Nodine & Bartel, 2012). This last study attributed the maternal transcript bias in Autran, *et al*. 2011, to tissue contamination from the maternal seed coat. However this explanation does not reconcile the comprehensive literature reporting maternal parent-of-origin transcript bias during embryogenesis (Autran *et al*, 2011; Vielle-Calzada *et al*, 2000; Baroux *et al*, 2001). The basis of this variation may reside in part on natural variation between the hybrids used. Similar observations have previously been documented for imprinted genes (Pignatta *et al*, 2014). To test this hypothesis, I decided to evaluate the complementation of maternally inherited mutants using paternal alleles coming from different ecotype backgrounds. If complementation dynamics do

indeed change in hybrids compared to isogenic embryos, different hybrid combinations might lead to different timing of complementation. We used wild type pollen from several ecotypes to complement *emb* lines. I randomly chose eleven insertional lines from our previous study for this experiment (**Table m&m 1**). We used wild type pollen from several ecotypes to complement *emb* lines. I randomly chose eleven insertional lines from our previous study for this experiment (**Table m&m 1**). I emasculated and pollinated the *emb* lines using pollen from five different ecotypes of *A. thaliana*: Columbia (Col), Tsushima (Tsu), Wassilewskija (Ws), and the ecotypes used in the genomic studies, Lansberg *erecta* (L*er*) and Cape Verde Island (CVI) (Autran *et al*, 2011; Nodine & Bartel, 2012). I compared the complementation dynamics of isogenic embryos (Col-Col crosses) with each of the hybrid combinations, by scoring mutant phenotypes at 2, 3 and 5 days after pollination (dap). To control for the possibility of haploinsuficiency, reciprocal crosses using the different ecotypes as mother with the emb/+ lines (in Col) were performed. Overall, the results revealed a large variation in the timing of paternal allele activation between different hybrid combinations, suggesting that the genotype of hybrid embryos affects the kinetics of ZGA compared with an isogenic background. Complementation of mutant phenotypes in the isogenic background showed delayed functional activation as observed previously. Interestingly, CVI showed the highest variation for almost all the genes tested – except for *EMB2421* and *EMB2775*– showing fewer percentages of mutant embryos and faster recovery in the subsequent time points (**Figure 2: complete data shown in Supplementary Table 2**). For example, in the case of *vcl1/+* at 2dap with isogenic pollination, I observed 30.9% of mutant embryos, while pollination with CVI showed 13%. At 3dap the *vcl1/+* x CVI cross showed no mutant phenotypes, whereas the isogenic *vcl1/+* x Col embryos still showed high levels of mutants (**Figure 2: complete data shown in Supplementary Table 2**). Paternal allele rescue from the additional three ecotypes showed similar kinetics, where the complementation was faster for almost all the cases analyzed, but not as drastic as the CVI ecotype. Interestingly, the hybrid embryos with Tsu ecotype showed similar kinetics of paternal allele activation compared with isogenic Col. Pollination with Tsu did not show faster complementation for any of the EMB lines at 2dap, and only in 4 of the 11 cases showed faster functional activation at 3dap (**Fisher's exact test in Supplementary Table 2**), suggesting that the complementation dynamics of the Col-Tsu hybrid embryos is more similar to the observed in isogenic

backgrounds than the other ecotypes. (**Figure 2: complete data shown in Supplementary Table 2**).

Paternal allele rescue from the additional three ecotypes showed a similar kinetics, where the complementation was faster for almost all the cases analyzed, however not as drastic as CVI ecotype. Interestingly the hybrid embryos with Tsu ecotype showed the lower variation compared with isogenic background. Pollination with Tsu did not show faster complementation for any of the EMB lines at 2dap, and only in 4 of the 11 cases showed faster functional activation at 3dap (**Fisher's exact test in Supplementary Table 2**), suggesting that the complementation dynamics of the Col-Tsu hybrid embryos is more similar to the observed in isogenic backgrounds than the other ecotypes. (**Figure 2: complete data shown in Supplementary Table 2**).
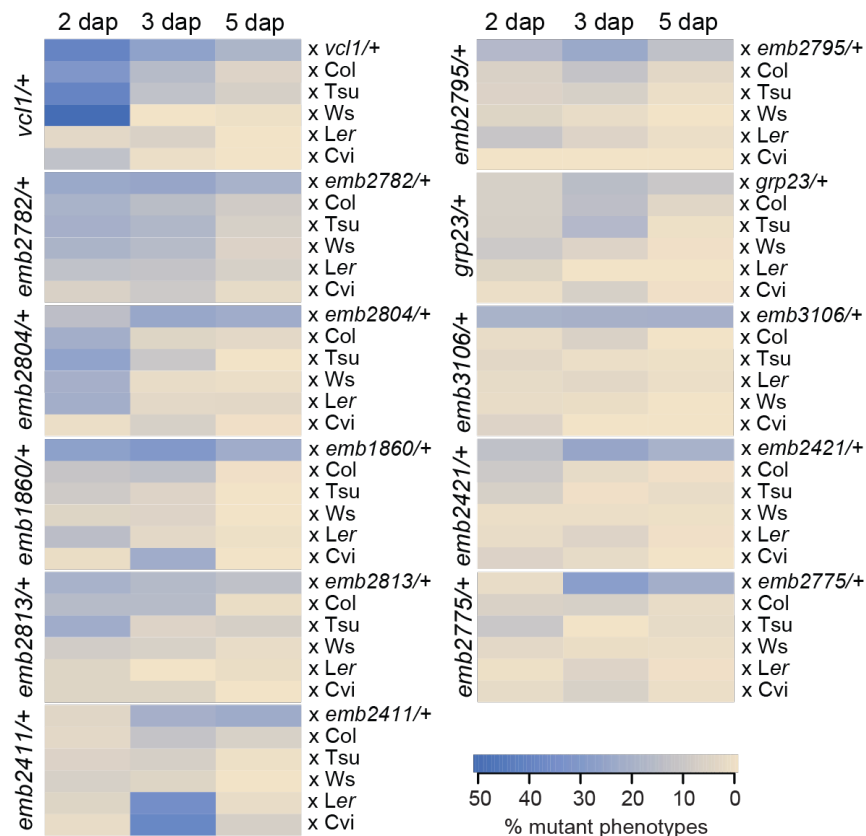
Figure 2. **Paternal allele activation in early Arabidopsis embryogenesis varies between hybrids**. Heat map representations of mutant phenotype percentages observed for 11 different emb/+ mutants for self-crosses (X emb/+), and crosses to Col-0 (X Col), Tsu-1 (X Tsu), Ws (X Ws), Ler (X Ler) and Cvi-0 (X Cvi) pollen. In absence of paternal contribution, the maximum percentage of mutant embryos as consequence of the *emb* defective allele inherited by the mother is 50% (blue) in heterozygous plants, while full paternal allele activity corresponds to 0% observed mutants (beige). Full table data shown in Supplementary Table 2. Original figure from Del Toro, García-Aguilar&Gillmor, 2014.

L*er* and CVI were the ecotypes used in combination with Col to profile Allele Specific expression in the previously reported transcriptomes from early embryos. These studies reported divergent observations, where the Col x L*er* early embryos showed biparental expression with a maternal allele bias for 80% of the genes, whereas the Col x CVI embryos showed equivalent contributions throughout embryogenesis (Autran *et al*, 2011; Nodine & Bartel, 2012). To functionally evaluate whether these specific hybrid combinations show differences in the activation of paternal alleles, I compared our complementation data for L*er* and CVI pollen. I found 6 genes with significant differences between L*er* and CVI: *VCL1*, *EMB2782*, *EMB2804*, *EMB1860*, *EMB2795* and *GRP23*. In five of the six cases the CVI wild type paternal alleles complemented faster than L*er*, showing less percentage of mutant embryos and thus an earlier functional activation of the paternal genome. Interestingly, the differences for all the cases were statistically significant from 2dap, suggesting that the activation of the paternal genome of the Col x CVI embryos occurs before this stage, showing a higher functional complementation at 2dap. L*er* complemented faster than CVI at 3dap only for GRP23 (**Figure 2: complete data shown in Supplementary Table 2**). Thus, our functional genetic data for hybrids correlates with both transcriptome studies, where the Col x L*er* hybrid shows delayed activation of the paternal genome compared with the Col x CVI hybrid. For Col x CVI, I found that functional activation takes place before 2dap, around the same developmental stage of the first time-point evaluated by Nodine & Bartel, 2012. However, both hybrids showed an earlier activation of the paternal allele when compared with isogenic complementation.
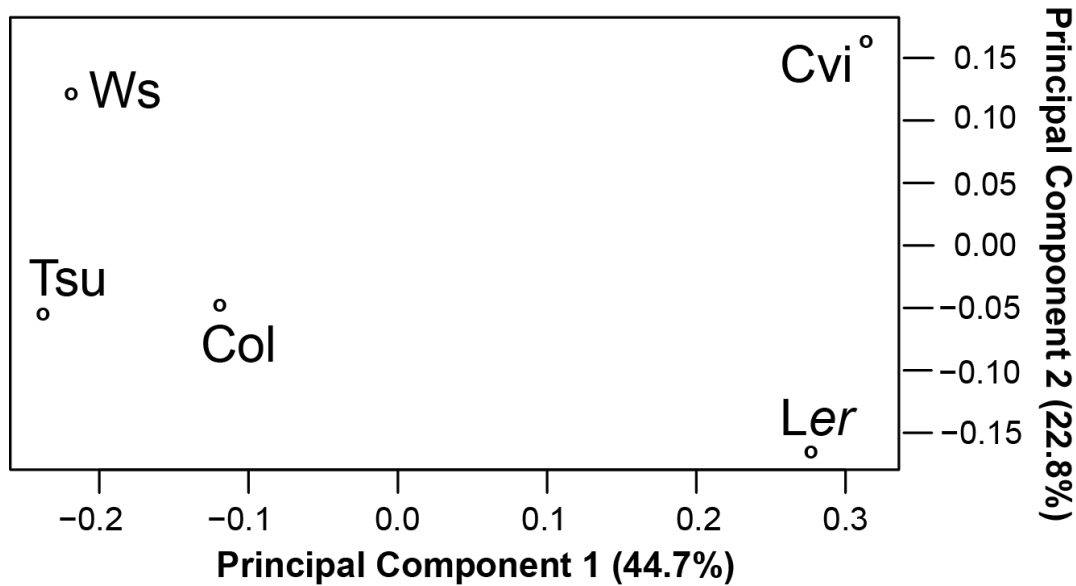
Figure 3. **Principal Component Analysis of the mutant phenotype segregation ratios observed in** ***emb/+*** **crosses to all WT ecotypes**. Representations of mutant phenotype percentages observed for 11 different *emb/+* mutants for self-crosses (X *emb/+*), and crosses to Col-0 (X Col), Tsu-1 (X Tsu), Ws (X Ws), Ler (X Ler) and Cvi-0 (X Cvi) pollen. Full table data shown in Supplementary Table 2. Original figure from Del Toro, García-Aguilar&Gillmor 2014.

In order to analyze all functional data for the hybrid crosses, I performed a Principal Component Analysis of mutant phenotype segregation ratios observed in *emb/+* crosses to all WT ecotypes (**Figure 3**). The graph reproduces the variation observed in the kinetics of the activation of the paternal genome. Interestingly, the isogenic *emb/+* x Col control is closest to the *emb/+* x Tsu hybrid. Both the L*er* and CVI ecotypes are significantly different from the isogenic cross, while CVI was the most distant ecotype, showing the earliest activation of paternal alleles (**Figure 2 and 3**). All these data together suggest that the distinct conclusions in the previous transcriptomic profiles may be partly explained by the variation that occurs in paternal genome activation in hybrid embryos. Since all these ecotypes belong to the same species, new alleles or different epigenetic states may trigger the observed variation in the delay of paternal activation in hybrids.

Discussion

Previous experiments in our laboratory demonstrated delayed functional activation of paternal alleles for 40 out of 49 of genes analyzed (Del Toro-De León *et al*, 2014). *GCT* and *CCT* were among the genes that showed delayed complementation of maternal mutant alleles by WT paternal alleles. Parent-of-origin analysis of *pCCT::GUS* and *gGCT-GUS* reporter lines showed

that maternal allele expression begins as early as the one-cell stage embryo, while the expression of the paternal allele begins gradually at 3dap (**Figure 1**). Full complementation of maternally mutant *gct/+* and *cct/+* embryos by the wt allele does not occur until 5 dap, one day after the *GCT* and *CCT* reporters show full penetrance of expression. Thus, the molecular dosage of wt *GCT* and *CCT* alleles must reach its maximum before the gene product can fully carry out its function.

These data show that maternal and paternal alleles for *CCT* and *GCT* are not equivalently expressed. One possibility is that paternal allele is indeed expressed, but below the thresholds of detection for GUS staining. In either scenario –absence or low expression– the paternal allele is not abundant enough to complement the mutant phenotypes produced by the defective maternal allele. The remaining 38 genes that I previously analyzed, which also showed delayed paternal allele activation, may show similar behavior to *GCT* and *CCT* at the level of transcripts (Del Toro-De León *et al*, 2014). Although our data provide evidence for delayed paternal genome activity at the functional and transcriptional level, further efforts are necessary to show whether this phenomenon occurs at the genome level.

Genome-wide ASE profiles must use SNPs to differentiate between maternal and paternal allele expression in hybrids. Therefore, hybrids are indispensable for testing parent-of-origin expression by RNA sequencing. However, changes in gene expression and histone and DNA methylation have been documented as a consequence of hybridization (Greaves *et al*, 2012; Groszmann *et al*, 2011; Greaves *et al*, 2014). Unlike transcriptomic studies, our previous genetic data for the timing of maternal and paternal contributions was generated in an isogenic background (Del Toro-De León *et al*, 2014). Using the same functional strategy, it is possible to compare isogenic vs hybrid ZGA. The data here reported show that activation kinetics for isogenic paternal alleles differs among genes, and among hybrids tested (**Figures 2 and 3**). Interestingly, the Col x CVI hybrid showed the highest difference with isogenic embryos, as all paternal alleles in this hybrid were activated earlier than in isogenic crosses.

These results agree with the data reported by Nodine & Bartel 2012, where Col x CVI embryos showed equal maternal and paternal contributions. Col x CVI hybrid embryos show epigenetic

differences compared with other combinations, due to the unique features of CVI ecotype (Tessadori *et al*, 2009; Pignatta *et al*, 2014). Interestingly, CVI shows decreased DNA and histone H3K9 methylation, epigenetic marks that play an important role in the RdDM gene silencing pathway (Tessadori *et al*, 2009). As part of the RdDM pathway, the DNA methyltransferase CMT3 binds to H3K9me, a chromatin mark deposited by the SUVH histone methyltransferases, which in turn recognize DNA methylation, forming a feedback loop to reinforce silencing (Johnson *et al*, 2007, 2008). The H3K9 methyltransferase SUVH4 (also known as KRYPTONITE (KYP)) was shown to repress the paternal genome in early embryos, demonstrated by the increased paternal contribution in the *kyp* 2-cell stage embryos. In addition, the levels of 24 nt siRNA that direct methylation in coding regions diminish in hybrids, thus altering gene expression (Groszmann *et al*, 2011). One hypothesis for earlier paternal allele activation in Col x CVI hybrids is that CVI shows less epigenetic (RdDM) silencing at the genome level. Whether variation in RdDM-directed silencing may underlie differences in paternal allele activation in other hybrids remains to be tested.

Part 2. Parent-of-origin expression in a Col x Tsu hybrid embryo transcriptome.

Summary

Previously we showed that, of the hybrids we examined, *emb/+* (Col) x Tsu complementation dynamics were closest to *emb/+* x Col (Part 1, Del Toro-De León et al., 2014). Here I report an Allele Specific Expression analysis of a Col x Tsu hybrid embryo transcriptome. The data show a compendium of different patterns of allele biased expression. At the zygote and octant stages we observed an overwhelming bias toward maternal transcripts, with paternal contributions steadily increasing until the globular-heart stage, when parental contributions are approximately equivalent. I was able to identify *cis* regulatory sequences enriched in maternally biased genes, suggesting that these parent-of-origin effects on transcription in early embryos are the result of the convergent functions of transcriptional activation via specific *cis* regulatory motifs (as well as *trans* acting factors), and silencing of paternal alleles due to epigenetic marks established in the gametes.

Col x Tsu embryo transcriptome shows biparental expression with maternal bias

Given the fundamental importance of embryo development to many applied aspects of seed traits in crops, it is critical to understand how parental genomes are programmed to regulate gene expression during embryogenesis. For example, growth performance of plants generated from reciprocal crosses shows significant differences in the offspring (F1) phenotypes despite their identical nuclear parental genomes. Evidence indicates that some of these differences are established during embryogenesis (Ng *et al*, 2014; Swanson-Wagner *et al*, 2009). This suggests that parent-of-origin gene expression plays an important role in hybrid vigor. Hybrid genomes show allelic interactions that may lead to changes in gene expression (Greaves *et al*, 2012) and variation in parent-of-origin effects (Ng *et al*, 2014). Thus, finding a hybrid combination that behaves similarly to an isogenic line is important to make unbiased statements regarding ZGA in plants. Our previous results suggested that Col x Tsu embryos show paternal allele activation kinetics similar to isogenic embryos. Thus, genomic ASE analysis of this hybrid combination may more faithfully reflect this phenomenon in isogenic backgrounds.

We started a collaboration with the Raju Datla laboratory at the Plant Biotechnology Institute, National Research Council of Canada. Datla's group performed dissection and subsequent transcriptome profiling of F1 embryos that were derived from reciprocal crosses of Col-0 and Tsu-1 for seven developmental stages: zygote, octant, globular, heart, torpedo, bent and mature embryo stages (**Supplemental Figure 1**). These libraries were sequenced to a depth of 18-48 million 100-bp reads, which were mapped to the *Arabidopsis* genome using the Tair10 assembly. To avoid mRNA contamination from the surrounding ovule tissue and endosperm, isolated embryos were washed and the representative embryo stages were confirmed by observation under the microscope. Maternal and paternal alleles were distinguished using SNP profiles of the parental genomes. A total of 17836 expressed genes containing SNPs were detected across the entire *Arabidopsis* genome. This number represents the majority (~90%) of the genes that are estimated to be expressed during *Arabidopsis* embryogenesis (Yilmaz *et al*, 2011). Informative parental genome-specific SNPs were used to quantify maternal and paternal contributions for each gene individually. I then used this data for parent-of-origin analysis. To reduce sampling bias due to low counts, 10 SNP counts per transcript were required as a minimum. Parentally-biased genes were assigned by chi-square tests on results from the reciprocal crosses. A False-discovery rate (FDR) corrected P-value cutoff of 0.05 (see Experimental Procedures) was used to determine the significance of parent-of-origin gene expression for each gene.

During the analysis, I identified a broad group of genes with Reciprocal Maternal Bias expression. For each gene in this group, 60% or more of the SNP counts correspond to the maternal allele, in both reciprocal crosses (FDR ≤ 0.05). Similarly, a group of genes with Reciprocal Paternal Bias expression was identified (reciprocal paternal SNPs ≥ 60%, FDR ≤ 0.05). For biallelic expression I looked for genes with SNP counts between 60% and 40% for each parent; genes with maternal or paternal biased expression but without statistical significance were also considered as biallelic. Notably, not all the genes were classified as reciprocal parental biased or biallelic. I identified a large group of genes with Non-Reciprocal Parental Bias, which showed significant enrichment towards the maternal or paternal allele in one direction of the cross, and biallelic expression in the reciprocal. A final group of transcripts showed preferential expression of one allele, irrespective of whether it was maternally or paternally inherited. I called

this group Allele Dominant. **Table 1** summarizes the different categories and the selection criteria used to classify parent-of-origin gene expression.

Group	Selection criteria
Reciprocally Maternal Bias	reciprocally maternal SNPs ≥ 60%, FDR ≤ 0.05
Maternal bias in Col x Tsu	maternal SNPs ≥ 60%, FDR ≤ 0.05 in Col x Tsu
Maternal bias in Tsu x Col	maternal SNPs ≥ 60%, FDR ≤ 0.05 in Tsu x Col
Reciprocally Paternal Bias	reciprocally paternal SNPs ≥ 60%, FDR ≤ 0.05
Paternal bias in Col x Tsu	paternal SNPs ≥ 60%, FDR ≤ 0.05 in Col x Tsu
Paternal bias in Tsu x Col	paternal SNPs ≥ 60%, FDR ≤ 0.05 in Tsu x Col
Col Allele Dominant	Significant maternal expression bias in one cross and
Tsu Allele Dominant	significant paternal expression bias in the reciprocal cross or vice versa, FDR ≤ 0.05
Biallelic	reciprocally maternal SNPs < 60% and > 40%
Biallelic*	The remainder of the genes

Table 1. **Classification criteria for the genes from Allele Specific Expression analysis of Col x Tsu embryo transcriptome.** These criteria were employed to classify all the genes expressed (with >10 SNP counts in each reciprocal cross) in each developmental stage. *This group of biallelic genes contains genes that did not match any of the other conditionals, for example genes that show maternal bias but without statistical significance (SNPs ≥ 60%, *p*-value ≥ 0.05).

Most maternal and paternal mRNAs were detected in all developmental stages with a parent-of-origin bias, but were not strictly monoallelic or biallelic (**Figure 6**). Interestingly, most of the genes expressed in the zygote and octant stages showed Reciprocal Maternal Bias enrichment (66% and 49% respectively) compared with 2% Biallelic genes (**Figure 4**). This maternal dominance drops dramatically at the globular stage (from 66% in the zygote to 8.5% at globular), when biallelic expression becomes the dominant category. These observations suggest that there is maternal genomic control of embryogenesis at preglobular stages, which is overcome in globular embryos. In contrast to the maternally biased genes, only a handful of transcripts where identified to have a Reciprocal Paternal Bias expression. For instance, at the zygote stage, 42 genes where detected to be paternally enriched, while 4,451 genes were maternally biased (**Figure 4, B**). As embryonic development progresses, reciprocal maternal and paternal biased genes decrease, while biallelic genes show a dramatic increase. Our data also suggest that parent-of-origin effects are not always reciprocal. During the analysis, I found a group of genes with maternal or paternal enrichment in one direction of the cross, but biallelic transcripts in the reciprocal direction. Interestingly we identified more genes with Non-Reciprocal Maternally Biased expression in the Col x Tsu cross. In contrast, Non-Reciprocal Paternally Biased

expression was marginal in the zygote and octant stages. Except for the zygote, octant and globular stages (where there is some variation), this Non-Reciprocal group of genes remains constant in number during embryogenesis. Although these genes show different behavior in one cross vs. the other, I cannot rule out the possibility that the Non-Reciprocal parental enrichment is the result of hybridization. Further research needs to be performed to clarify this observation. One category certainly influenced by ecotype derived effects is the Dominant genes, which show preferential expression of one ecotype (allele). Unlike genes whose parent-of-origin expression changes dramatically over time, allele-dominant genes preferentially remained steady throughout embryogenesis indicative of maintenance of their active expression state (**Figure 4**). The dominant expression of these alleles, regardless of the direction of the cross, is likely due to *cis* regulatory effects.

To address the biological functions of the parentally biased genes and other gene groups at the zygote to globular stages, I conducted gene ontology (GO) analysis for each group. Results showed that at zygote and octant stages, Reciprocal Maternal Bias genes are associated with molecular functions such as structural constitution of ribosome, RNA and protein binding as well as transmembrane transport. Biallelic genes in zygote stage where involved in sterol binding, whereas in octant stages functions associated with nuclei acid binding and translation are observed (**Table 2**). Interestingly, in the globular embryo (where there is the major decrease in maternally biased genes), the functions associated with this category are mainly related with enzymatic activity, although associations with transcription factor activity and sequence-specific DNA binding are still found. Meanwhile, the biallelic gene category in globular embryos is predominated by functions related to active translation and transcription (**Table 2**). Reciprocal Paternal Bias genes were associated with enzymatic functions such as nucleotide dimethylallyltransferase and diphosphate synthase activity. Notably, allele-dominant genes in the zygote, octant and globular stages are involved in the same molecular functions related with translation and signaling, suggesting that this class may play a predominant role in hybrids. These findings suggest that the presence of transcription and translation functions in maternally biased genes is indicative of their central roles in patterning and development during early embryogenesis (Jenik *et al*, 2007). Interestingly these functions seem to be transferred to the biallelic genes in later stages. Transcriptional dynamics of parental genome contributions and

gene ontology function analysis indicate that there is a maternal to zygotic transition in developmental control of embryogenesis, which implies fast reprograming of parental genomes at preglobular stage

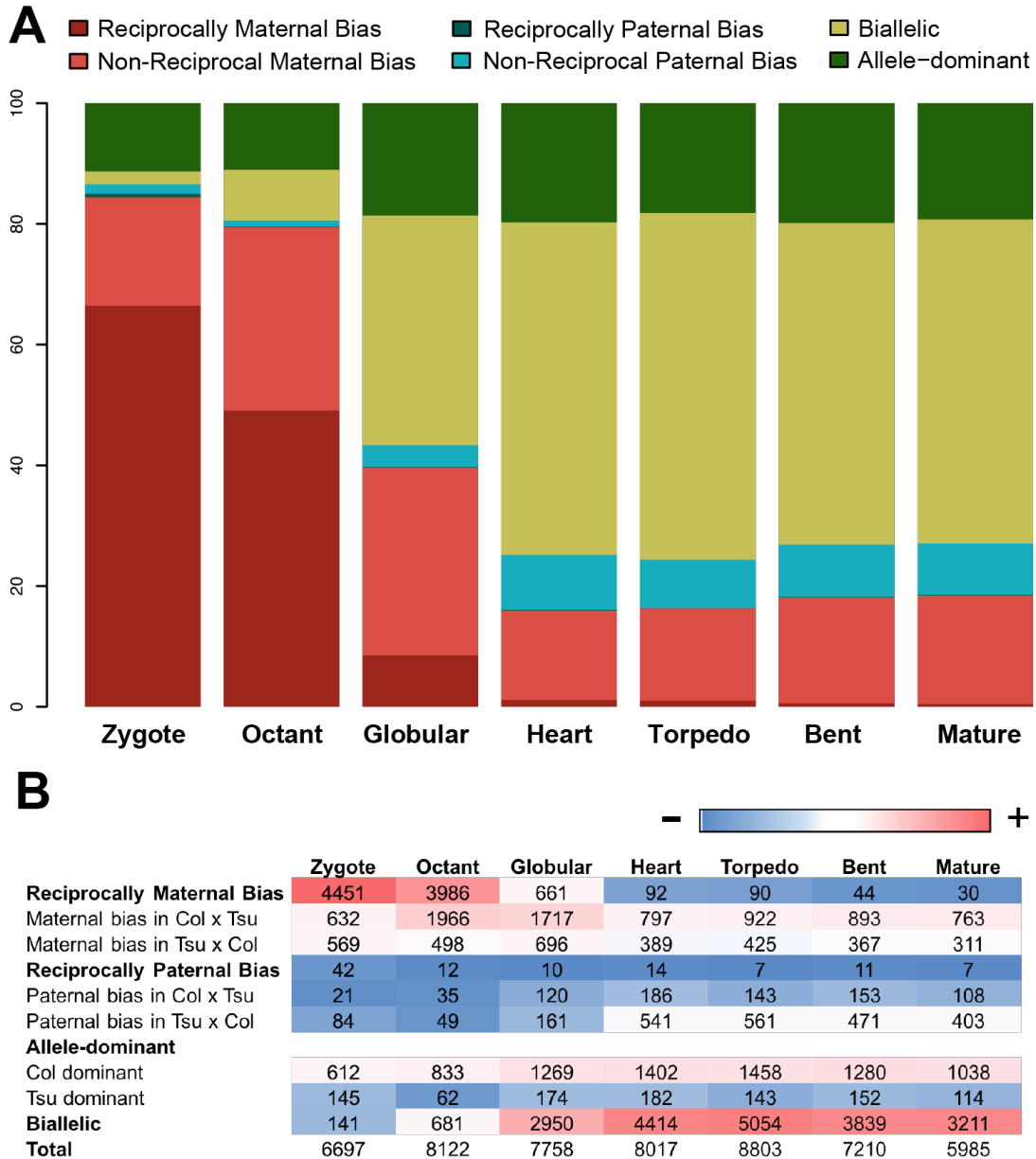**Figure 4**. **Parent-of-origin gene expression during Col x Tsu embryogenesis. A**, Bar graph of the percentages of genes belonging to each category in every developmental stage. **B**, Table with the total number of genes in each category by stage. The color show a gradient to facilitate the identification of trends (major number of genes) with categories containing high (red), medium (white) and low (blue) number of genes.

Group	Zygote	Octant	Globular
Reciprocally Maternal Bias	Structural molecular binding Structural constituent of ribosome	Protein binding Catalytic activity	Catalytic activity Oxidoreductase activity
Maternal bias in Col x Tsu	Iron ion binding	Structural constituent of ribosome Structural molecular activity	Protein binding
Maternal bias in Tsu x Col	Structural constituent of ribosome Structural molecular activity	Structural constituent of ribosome Structural molecular activity	Fatty acid synthase activity Protein binding
Reciprocally Paternal Bias	Ent-copalyl diphosphate synthase activity	ATP, ADP and AMP Dimethylallyltransferase activity	ND
Paternal bias in Col x Tsu	Mannosyl-glycoprotein endo-beta-N-Acetylglucosaminidase activity	transcription factor activity, RNA polymerase II transcription factor binding	Methyltransferase activity
Paternal bias in Tsu x Col	fatty acid alpha-hydroxylase activity	copper ion transmembrane transporter activity	dUTP diphosphatase activity
Col Allele Dominant	Structural molecular binding Structural constituent of ribosome	Structural constituent of ribosome Structural molecular activity	Structural constituent of ribosome Structural molecular activity
Tsu Allele Dominant	CoA-ligase activity	Indole-3-acetate beta-glucosyltransferase activity	Sphingosine N-acyltransferase activity
Biallelic	Oxysterol binding Sterol binding	Structural constituent of ribosome Structural molecular activity	Structural constituent of ribosome Protein binding

Table 2. **Gene Ontology analysis.** The table show the representative Molecular Function category for each group-stage.

Reprogramming of parental genomes during embryo development

To better understand the dynamics of parental genome contributions in preglobular stages, we analyzed changes in parent-of-origin transcripts during embryo development. As shown in **Figure 4**, from the zygote to the globular stages, both Biallelic and Allele-Dominant genes
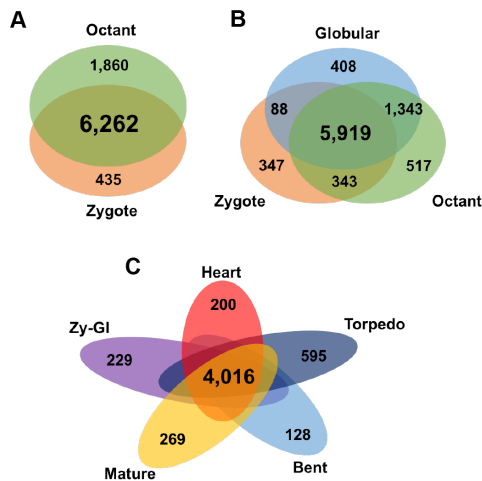
Figure 5. **Number of genes shared between stages.** Venn diagrams showing the number of expressed genes shared between Zygote and Octant (**A**), Zygote, Octant and Globular (**B**) and all the stages of this transcriptome (**C**). In C the total number of genes is 8,642.

rapidly increased, from 2% to 38% and 11% to 18% of the detected genes, respectively, while the Reciprocal Maternal bias genes showed significant reduction, from 66% to 8%. Starting at the heart stage, the Reciprocal Maternal bias genes decreased to less than 1%. In contrast, the Biallelic genes range between 53 and 57%, while Allele-Dominant genes are maintained at ~20% of the expressed genes. Overall, the Allele-Dominant and Biallelic genes showed a strong propensity to remain in their own expression group across all embryo development stages, which explains that this category share the same GO functions. Non-Reciprocal Maternal and Paternal biased genes remain

steady with 16.3% and 8.5% on average respectively. These changes in the proportions during early embryogenesis suggest that the parental genomes undergo a reprograming that allows fully biallelic expression after pregloblar stages; for the Bialleic genes, the highest percentage is observed at heart stage onwards, suggesting that the mechanisms that lead to parental bias expression are no longer in operation and only the Dominant alleles retain their expression bias.

To understand the reprograming of the genome from parentally biased to biallelic, I calculated the percentage of genes that remain in the same category, compared with the previous and the next stage of development (**Table 3**). Interestingly, 77.4% and 93.6% of the total genes expressed in the globular embryo are shared with the zygote and octant stage, respectively (**Figure 5**). Accordingly, most of the genes within the decreasing parentally biased categories were present in the zygote, octant and globular stages, while genes in the growing Biallelic category were present through all embryogenesis (**Table 2**). These observations suggest that during the transition from the zygote to globular stage, parentally biased genes in the preglobular embryo become biallelic, explaining the progressive increase of this category and its predominance beginning in the globular embryo. Indeed, from the Reciprocal Maternal genes in zygote, 35% became Biallelic, 12% allele Dominant, 37% Non-reciprocal maternal bias, 13% remained Reciprocally Maternally biased and just a few were in the paternally biased expressed in the globular embryo (**Table 2**). These observations confirm that there is a progressive reprograming of parentally biased genes from preglobular to the globular embryos and completed by heart stages (**Figure 6**).

To visualize trends in reprograming during embryogenesis, I compared parent-of-origin expression of the 4,016 genes (**Figure 5**) that are shared and expressed throughout embryogenesis (**Figure 6**). The data show maternally biased expression during preglobular stages which is overcome by heart stage (**Figure 6**). Interestingly, the same genes which initially show maternal bias become biparental in later stages. At the bottom of **Figure 6** there is a cluster of genes that show either maternally or paternally biased expression depending the direction of the cross, this observation confirms that the Dominant genes remain Dominant throughout all the embryo development, and that their pattern of expression is independent of parent-of-origin effects. Rather, their dominant expression reflects ecotype-dependent expression. In general, the

data shows that there is neither strict biallelic gene expression (50:50) nor monoallelic, instead the genes show a preferential expression of the maternal alleles in preglobular embryos, which turns biallelic later during development.

Group category	Zygote	Octant		Globular		Heart		Torpedo		Bent		Mature
Reciprocally Maternal Bias	68.2	85.4	16.8	93.0	12.7	69.4	33.8	42.9	19.4	19.4	3.6	3.4
Maternal bias in Col x Tsu	36.3	14.9	30.3	32.7	19.6	42.3	26.1	24.3	31.2	26.6	29.2	26.7
Maternal bias in Tsu x Col	21.7	32.9	21.2	13.7	16.0	31.0	24.1	22.9	29.7	29.0	20.4	20.8
Reciprocally Paternal Bias	13.3	80.0	71.4	55.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Paternal bias in Col x Tsu	0.0	0.0	14.3	4.0	12.5	9.1	9.7	15.0	11.1	7.6	6.7	7.8
Paternal bias in Tsu x Col	4.7	12.0	13.6	4.4	12.7	3.9	16.0	16.2	24.5	25.6	17.8	18.0
Allele-dominant												
Col dominant	53.0	44.5	82.8	54.6	73.9	65.5	75.1	73.1	71.1	75.3	63.5	69.0
Tsu dominant	18.2	48.9	54.4	19.4	43.6	45.0	37.6	49.6	41.9	40.2	26.4	28.6
Biallelic	32.6	12.8	72.9	15.0	79.9	58.6	76.9	76.6	74.4	76.4	72.9	71.5

Table 3. **Percentage of sharing with neighbor stages.** The table shows the percentage that every category share with the same category but from the adjacent stage. Color represents the size of the number (white-red; 0-100).

All this data together demonstrates that early embryogenesis shows a bias towards expression of maternal alleles, which later transitions to biallelic, likely mediated by a reprograming of parent-of-origin gene expression. The mechanisms that lead to this maternally biased expression are still unknown, though they may be mediated by RdDM pathway, as well as the mechanisms of paternal genome activation by histone turnover through CAF1 complex (Autran et al., 2011). In addition, whether maternal transcript bias is due to *de novo* expression in early embryo or a consequence of mRNA carryover from the female gametophyte remains to be tested.
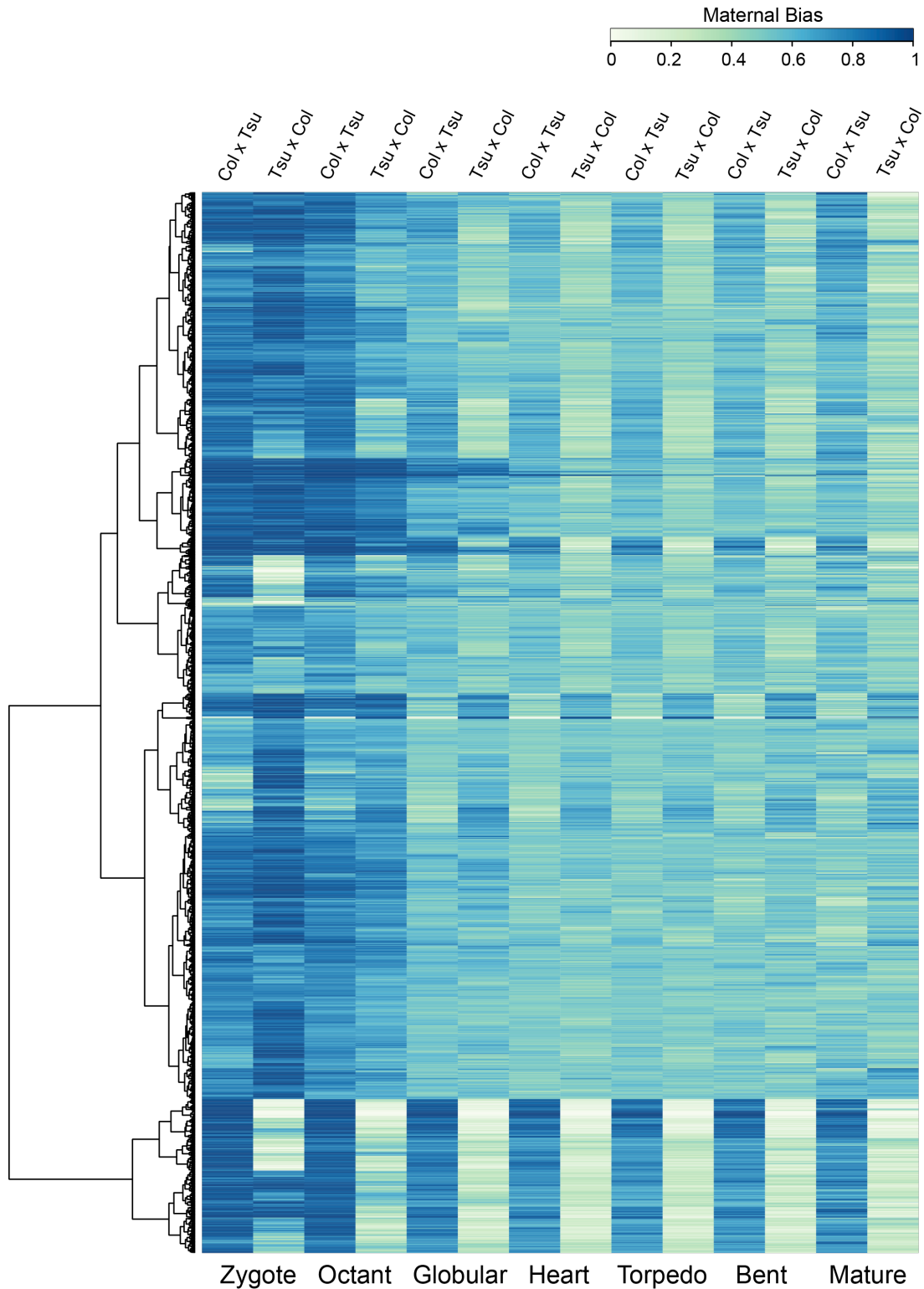
Figure 6. **Reprogramming of parent-of-origin gene expression during Col x Tsu embryogenesis.** Representation of the evolution of the parent-of-origin bias expression for 4,016 genes expressed throughout embryo development. maternal bias percentage was used for this heat map; bluish genes are expressed mainly by the maternal side either Col or Tsu (cross direction indicated at the top of the panel; by convention maternal parent is indicated first).

Maternally biased expression is the result of both *de novo* and carry over expression

The most striking trend observed at zygote and octant stages is the genomic maternal transcript bias. Both reciprocal and non-reciprocal maternal transcript bias predominates among the parent-of-origin categories at preglobular stages (**Figure 4 and 6**). This transcript bias may be the consequence of carry-over from the female gametophyte, by *de novo* expression of maternal alleles, or both. If the genes are *de novo* expressed, this would be reflected by increased or steady transcripts during early embryogenesis; if the transcripts are merely carry-over, the levels would go down after fertilization. For the 2,896 genes that showed Reciprocally Maternal bias, around half did not show any change in the level of expression during zygote-globular stages (**Figure 7B**), suggesting that *de novo* synthesis of maternal mRNAs after fertilization is maintaining the maternal bias. Interestingly, from the 2,576 RMB genes shared between Col x Tsu embryos and a female gametophyte transcriptome (Wuest *et al*, 2010), 47.6% (1,226) are present in the egg cell, suggesting that some of those genes with preferential maternal expression may be influenced by mRNA carryover (**Figure 7A**). Although these observations suggest that maternal transcript bias may be the result of maternal allele specific expression, this may be in addition supported –at least for half of the genes– by carry-over from the female gametophyte. Therefore, carry-over and *de novo* transcription are strongly influencing the transcriptome of the early embryo, suggesting that Zygote Genome Activation takes place just after fertilization. In further support of

Figure 7. **Expression of Reciprocally Maternal genes at Zygote and Octant Stage. A**, Presence of RMB genes from zygote and octant stages in Egg Cell transcriptome. **B**, Expression of 2,896 genes RMB from zygote and octant stages.

this statement, Col x Tsu embryo transcriptome shows 11% of allele-dominant and 2% of biallelic transcripts in the zygote. In addition, 18% of Non-Reciprocal parentally biased genes account for biallelic transcripts in one direction of the cross (**Table 2**). Therefore, the results

suggest that the Arabidopsis zygotic genome is transcriptionally activated soon after fertilization such that it ensures the expression of parent-specific, allele-dominant and biallelic genes in the zygotic genome.

Although the inheritance of maternal products from the female gametophyte has not been successfully tested, it is possible that besides the *de novo* transcription of maternal alleles, carryover from the female gametophyte partly explains the maternal biased expression in early embryogenesis. Interestingly, carry-over may involve not only mRNA, but proteins as well. Preliminary observations in the laboratory have shown that reporter lines specific to the egg cell and central cell (Steffen *et al*, 2007) show signal after fertilization in embryo and endosperm, respectively (**Supplementary Figure 3 and 4, Supplementary Table 11 and 12**). In the absence of detection of the native gene, this reporter expression suggests protein carryover from female gametophyte (Sprunck *et al*, 2012), and reflects what may be happening with other maternal derived mRNA or proteins.

Maternally bias expression shows different degrees of enrichment

The Reciprocal Maternal Bias category includes every gene with maternal transcript bias, where more than 60% of reads come from the maternal allele. However, within this group of genes different degrees of maternal transcript bias may exist. To gain further insights into the level of maternal bias, the RMB category was separated into different groups, according to the degree of maternal enrichment (**Table 4**)

Gene Category	Criteria
Maternal Specific	Reciprocally maternal SNPs ≥ 95%, FDR ≤ 0.05
Highly Maternal	Reciprocally maternal SNPs < 95% and ≥ 80%, FDR ≤ 0.05
Maternal Biased	Reciprocally maternal SNPs < 80% and > 60%, FDR ≤ 0.05

Table 4. **Criteria used in the classification of the level of maternally biased expressed genes.** These criteria were employed to classify the degree of maternal bias expression from all the genes expressed (with >10 SNP counts in each reciprocal cross) in each developmental stage.

Three different subgroups of transcripts were employed: Maternal-Specific (reciprocal maternal SNPs ≥ 95%, FDR ≤ 0.05), High Maternal bias (reciprocal maternal SNPs < 95% and ≥ 80%, FDR ≤ 0.05) and Maternal Bias (reciprocally maternal SNPs < 80% and > 60%, FDR ≤ 0.05). Similar criteria were used for Reciprocal Paternal bias genes. Again, the most striking change in

transcript patterns for the different gene groups occurs from the zygote to the globular stages. In the zygote, ~80% of the RMB genes belong to either Maternal-Specific or High Maternal bias categories (**Figure 8 A**). Interestingly, from the zygote to octant stage both Maternal-Specific and High Maternal bias genes decrease, and in contrast the Maternal Bias category has a slight increase in number. For example, 816 and 679 Maternal-Specific genes were observed in the zygote and octant stages respectively, however only 65 in the globular embryo (**Figure 8 A and F**). Interestingly around 40% of the Maternal Specific genes during Zygote and Octant embryo remain with the same maternal bias during both developmental stages, suggesting an active mechanism promoting Maternal Specific expression in the preglobular embryo (**Figure 8 B**). In contrast, the Highly Maternal and Maternal Biased genes show more variation, probably due to active reprograming during these stages. However, 65% (1,340 genes) of the genes that belong to some subpopulation of maternally enriched category in the zygote, show the same trend in Octant stage when compared with the full RMB category (**Figure 8 E**), suggesting that maternally biased genes are undergoing active reprograming in the zygote and octant stages, resulting in the dramatic change observed in globular embryos (**Figure 8 A**). These observations show that all the maternally biased categories quickly decrease after octant stages, as previously observed.

In comparison with Maternal-Specific genes, only a small number of the Paternal-Specific (13, 3 and 4 in the zygote, octant and globular stages, respectively) and Highly Paternally biased genes (6, 0 and 4 in the zygote, octant and globular stages, respectively) were found during *Arabidopsis* embryogenesis. Interestingly there are no genes with Paternal Specific expression after the globular stage, and only a handful retain Highly Paternally biased and Paternally biased expression in later stages. Previous studies have proposed only a handful of genes to have a paternally biased expression, however only a few have been successfully confirmed and interestingly none of them retain the same bias expression after globular stages, similarly to the observations of this work (Autran *et al*, 2011; Nodine & Bartel, 2012; Raissig *et al*, 2013; Bayer *et al*, 2009).
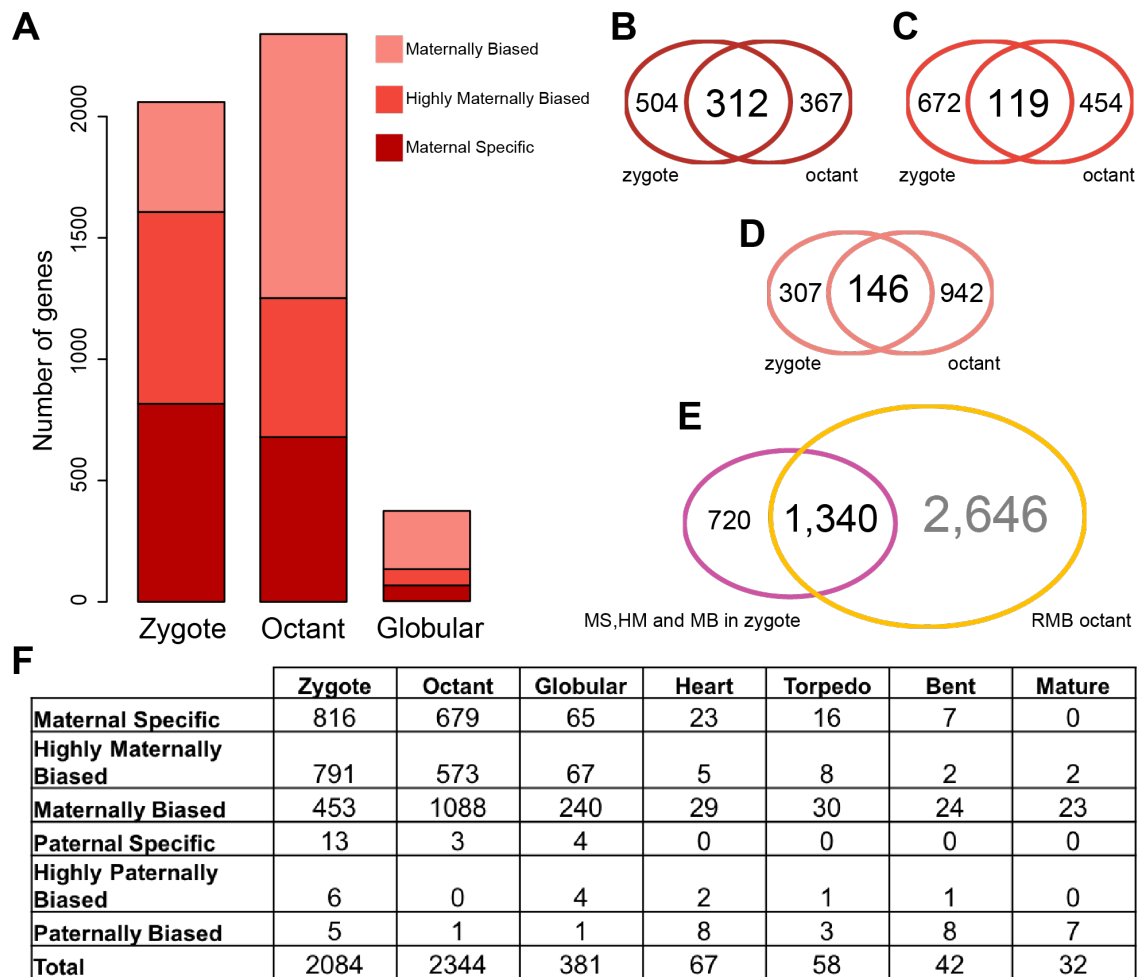
Figure 8. Levels of Maternal Biased expression in Zygote and Octant stages. A, graph showing the number of genes with Maternal Specific (MS), Highly Maternally Biased (HM) and Maternally Biased expression in Zygote, Octant and Globular stages. Number of MS (**B**), HM (**C**) and MB (**D**) genes shared between zygote and octant embryos. **E**, number of genes with MS, HM and MB expression in zygote shared with Reciprocally Maternal Bias group at octant stage (Figure 4 B). **F**, full table data.

Parent-Specific genes represent exclusive expression from either paternal or maternal alleles and may be indicative of imprinted gene expression in embryos. Interestingly, only a handful of Maternal-Specific genes were expressed in late-stage embryos and none were found in the mature stage (**Figure 8 F**). Unlike the genomic parent-of-origin effects during the zygote and octant stages observed in this transcriptome, imprinted expression seems restricted to a handful of genes and involves exclusive expression of maternal or paternal alleles during prolonged developmental periods (García-Aguilar & Gillmor, 2015). Imprinting has been largely observed in the endosperm, with only a few cases reported in plant embryogenesis (Raissig *et al*, 2013; He *et al*, 2017). Interestingly none of those reported imprinted genes show monoallelic expression in adult plants, suggesting that the imprinted mark is removed sometime during late seed

development. Therefore, the Maternal Specific genes observed in Bent Cotyledon and Torpedo stages may be candidates for imprinted genes that reach full reprograming at mature stages (**Figure 8 F**).

Motif analysis points to transcriptional regulation of maternally biased genes

To understand the mechanisms for preferential transcription of the maternal alleles in preglobular stages, I evaluated the presence of common signatures in the promoter regions of groups of genes sharing the same transcriptional profile (**Table 5**). According to previous analysis, Zygote and Octant stages showed a maternal bias (**Figure 4, 6 and 8**), moreover hundreds of genes maintained the same transcriptional maternal bias category, even in postglobular embryos (**Figure 8**). I selected these Maternally biased genes to perform *de novo* motif analysis, focusing on early embryogenesis. In addition to the Col x Tsu data, I also took advantage of another publically available ASE transcriptome profile (Nodine & Bartel, 2012).

Gene Category	Criteria	Source
Maternal Specific	From zygote to heart 1193 genes. 1178 with upstream sequence	ColxTsu transcriptome
Highly Maternal	From zygote to heart 1308 genes. 1293 with upstream sequence	
Maternal Biased	From zygote to heart 1627 genes. 1606 with upstream sequence	
Reciprocally Maternal Bias	From zygote and octant 2896. 2856 with sequence	
Transcriptome Universe	As control, 17590 genes with sequence	
≥4-fold increase	543 genes with ≥4-fold increase in the 8-cell embryos compared to the 1-cell/2-cell stages	Nodine&Bartel2012
≥10-fold increase	280 genes with ≥10-fold increase in the 8-cell embryos compared to the 1-cell/2-cell stages	
Preglobular EP	156 genes. Preglobular embryo proper	Belmonte, et al., 2013
Globular EP	83 genes. Globular embryo proper	

Table 5. **List of groups of genes used for MEME motif discovery analysis**

Previous analyses of the topology of regulatory elements in promoter regions have shown that almost all these elements are located within the first 2 or 3Kb upstream of the translation start site. Expression analysis has shown that these upstream regions are sufficient to recapitulate the pattern of expression of many transcription factors (Lee *et al*, 2006). Thus, I carried out the analysis using 2Kb upstream region for every gene. A preliminary *de novo* motif analysis showed that most motifs localize within the first 500bp of the upstream region (near the TSS (Translation Start Site)), and subsequent analysis showed a better performance of motif survey

using 1Kb upstream of the TSS, therefore final motif enrichment analysis was performed using this region.
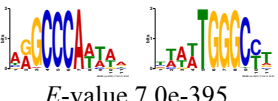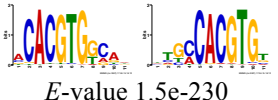
Motif Forward and Reverse Complement	MS	HM	MB	zyocRMB	Universe 17590
E-value 7.0e-395	20.96	**34.26***	**42.83***	**33***	29.04
	247/1178*	443/1293	688/1606	944/2856	5109/17590
	1.19e-9	8.7e-5	4.07e-29	1.6e-5	-
E-value 1.5e-230	**17.74***	**18.17***	15.25	**17.33***	13.8033
	209/1178	235/1293	245/1606	495/2856	2428/17590
	2.6e-4	2.3e-5	0.111	1e-6	-

Table 6. **Motif Occurrences in 1kb upstream regions (%) p<0.0001**
Statistically significant *p* values for Fisher's Two-Tailed Exact Test are shaded in bold to show difference in motif occurrence between the lists of biased expressed genes vs. the universe of genes. *$p \leq$0.001. A consensus matrix for each motif was calculated with meme using the genes with enrichment. *meme* Motif-based sequence analysis tools were employed. * Number of genes with motif occurrences from the total.

Two motifs were overrepresented in the maternally expressed genes. The first motif (TGGGC) was overrepresented in the Highly Maternal, Maternal Bias as well as the Reciprocal Maternal Bias group, compared with the control. The second enriched sequence identified was the CACGTG motif, which was overrepresented in the Maternal Specific group, as well as Highly Maternal group, in the Maternal Bias and the Reciprocal Maternal Bias group compared with the control. The CACGTG motif was the only motif identified to be enriched in the Maternal Specific genes, the TGGGC motif was only found in the Maternally Biased group, while the intermediate category of Highly Maternal biased genes shares both motifs (**Table 6**), reinforcing the idea that this category is a transitory subpopulation of transcripts that are decreasing in its maternal content during the reprogramming of gene expression in early embryogenesis. In accordance, the complete list of genes with Reciprocal Maternal Bias expression shows enrichment for the two motifs identified in this survey (**Table 6**).

To have a better understanding of the genes associated with these motifs, I evaluated their Gene Ontology (GO). The Maternal Specific genes which contained CACGTG motifs have functions

related to fundamental processes during early embryogenesis, such as response to hormone signaling and programmed cell death. Since this is one of the predominant categories during early embryogenesis, this suggests that essential functions during early embryogenesis are under maternal control not only at the level of transcription, but also for function (**Table 7**). Genes from the transitory category of High Maternal Bias show functions related with metabolism, while the Maternally Biased genes containing the TGGGC motif show functions related to translation, gene expression and RNA methylation. Finally, the complete category of Reciprocal Maternal Bias genes shows functions related with transport, metabolism, and response to stimulus.

Computational predictions of DNA sequence preferences for transcription factors (Weirauch *et al*, 2014) showed that the TGGGC motif is a likely target for the TCP transcription factor family Class I. This family of TF has a broad function in flowering, stress response and several other biological functions (Ma *et al*, 2014). 24 members compose this family and they are all expressed during early embryogenesis in *Arabidopsis*. The Class I TCPs are the most highly expressed, and moreover some of them show Maternal Specific expression (**Supplemental Figure 5**). However, the computational target prediction of transcription factors here used (http://cisbp.ccbr.utoronto.ca/TFTools) showed that the *TCP19* (AT5G51910) is the most likely to bind the TGGGC motif (Weirauch *et al*, 2014). The CACGTG motif is the very well-known G-Box motif, which is recognized by a family of basic helix-loop-helix transcription factors (bHLH), one of the largest transcription factor families in Arabidopsis involved in several developmental processes (Toledo-Ortiz *et al*, 2003). The G-Box has not previously been involved in parental biased expression or allele-specific expression The identification of these two motifs may shed some light on transcriptional regulation during early embryogenesis, and future studies should include deletion and swap experiments of the motif in candidate genes to evaluate the presence or absence of maternally biased expression; evaluate the effect of the identified motifs in minimal promoters; binding assay experiments to test the binding of putative transcription factors, and loss of function phenotypes during early embryogenesis.
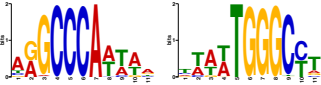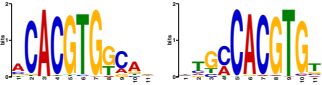
Gene Group	*E*-value 7.0e-395	*E*-value 1.5e-230
Maternal Specific	-	Response to hormones, Response to stress Programmed cell death, Signaling
Highly Maternally Biased	Cofactor metabolic process Generation of precursor metabolites and energy Coenzyme metabolic process	Cofactor metabolic process, Response to stimulus, Single-organism metabolic process
Maternally Biased	Translation, Peptide biosynthetic process Gene expression, RNA methylation	-
Zy-Oc Reciprocally Maternal Bias	Organic substance catabolic process Transport and localization	Single-organism metabolic process Response to stimulus Catabolic, metabolic process

Table 7. Gene Ontology Analysis for Biological Process, *p*-value < 0.01. List of genes with enrichment of the corresponding motif were used for this analysis.

Since we observed at least half of the genes with some degree of maternal bias expression are not reported to be expressed in the egg cell, *de novo* transcription of maternal alleles is necessary to accomplish the levels of maternal allele abundance in early embryogenesis. Identification of motif enriched sequences in the maternal biased genes suggests that this may be the targets of active transcription (or repression), that in coordination with epigenetic marks may result in maternal dominance during early embryogenesis. In addition, GO analysis conferred important functions to the maternally biased transcripts that reinforces the statement that during early embryogenesis there is a maternal control in gene abundance and function.

Maternal Bias expression is not a consequence of seed coat contamination

The plant embryo grows within the embryonic sac, surrounded by the endosperm and the maternal integuments that coordinately develop to form the seed. For this reason, transcriptional profiling of embryos has been challenging, especially during the earliest stages of development. Thus, allele specific transcriptomes of early embryos may be considerably biased by the presence of maternal tissue contamination. To evaluate the impact on maternal tissue contamination in our Col x Tsu transcriptome, I performed the recently reported "Contamination Test" (Schon & Nodine, 2017) which aims to quantify the transcript enrichment of groups of genes that are representative of different compartments from the seed (Belmonte *et al*, 2013). This contamination test used a laser capture miscrodissection microarray of seed development as the

basis for determining transcripts which are enriched in different compartments of the seed (Belmonte *et al*, 2013). To control for any possibility of maternal contamination, I analyzed the parent of origin expression of genes that are specific to the embryo (Belmonte *et al*, 2013). I also analyzed genes that show different degrees of enrichment in embryos, by means of fold change ratio compared with other compartments of the seed. All these groups of genes were selected taking as basis the same reference used by Schon & Nodine (2017) to design the Contamination Test.

The reference for the contamination test is based on a laser-capture microdissection experiment where different compartments of the seed were isolated from developing ovules of preglobular, globular, heart, torpedo and mature green embryo stages, and were hybridized to a full genome Arabidopsis microarray (Belmonte et al., 2013). Since our ColxTsu transcriptome profiled zygote, octant, globular, heart, torpedo, bent cotyledon and mature embryo stages, there is no direct reference to zygote and octant stages (preglobular stages from Belmonte et al. 2013 may include a mix of different stages of developing embryos from the earliest stages). In the absence of corresponding reference, I used "Preglobular" from Belmonte et al., 2013 for both zygote and octant stages. The "Contamination Test" reported by Schon and Nodine (2017), was developed to evaluate the level of contamination by quantifying the enrichment of transcripts from seed compartments other than the embryo. This assumes that the transcript reference, in this case the microarray from Belmonte, et al., (2013) is exempt from contamination. Other limitations of this contamination test correspond to the profiling methodology. Since the reference is a microarray, transcripts that are not detected are automatically ignored and then the contamination test may be biased and imprecise. Thus, I consider this test to be very useful, when its limitations are taken into account.

According to the "Contamination Test", our ColxTsu transcriptome shows high levels of Seed Coat tissues transcript enrichment, especially at the zygote and octant stages. However, at the octant stage, our Col x Tsu transcriptome also shows high enrichment for the embryo proper. At the globular stage, Embryo Proper tissues are enriched in both reciprocal crosses, with weak seed coat contamination in globular embryos. From the heart stage onward, all samples show embryo enrichment, with no seed coat contamination. As mentioned above, there is no appropriate

reference for the zygote stage: zygote and "preglobular" embryos have distinct transcriptomic profiles, since different developmental processes are taking place. It is likely that genes taken as enriched in embryo proper in the reference, are not even expressed in zygote or if they were, the expression may not be equivalent to the preglobular embryo from the reference. Taking this observation into account, the higher enrichment from Seed Coat in this stage may be highly unprecise and is not conclusive. Although there is contamination detected in octant embryos there is enrichment of Embryo Proper tissues as well. Seed coat contamination that is present despite the rigorous washing of the embryos before RNAseq may be due to technical and transcriptional differences between the reference and the ColxTsu transcriptome.
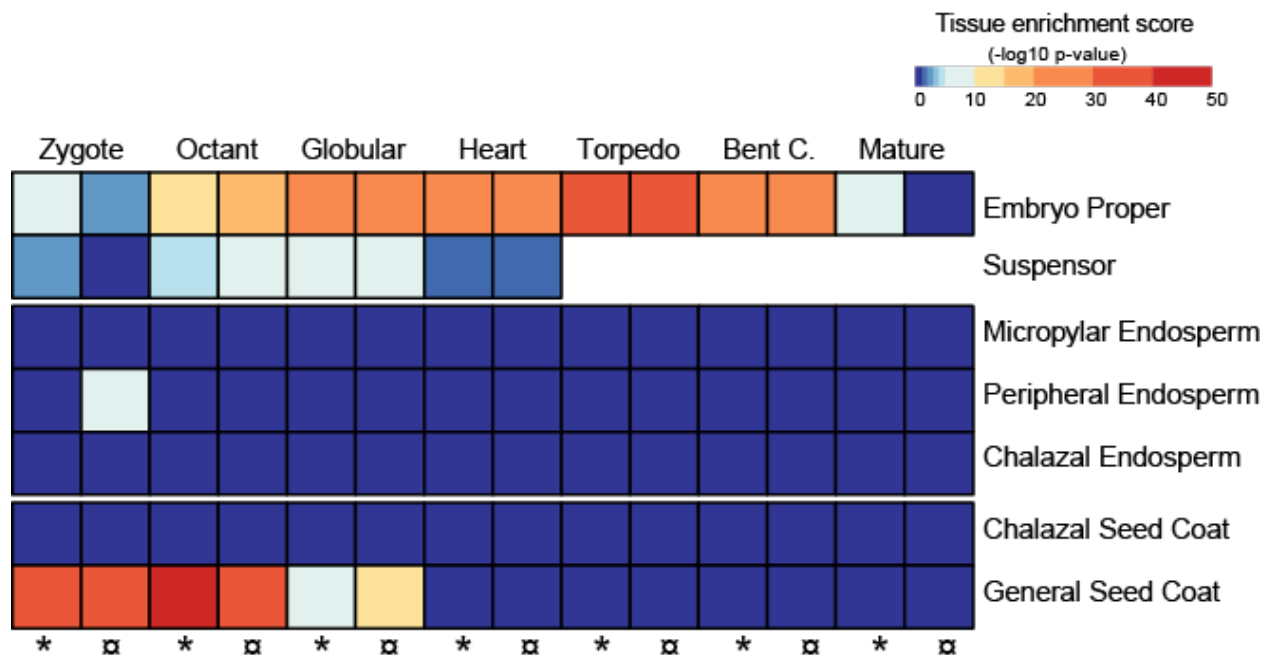
Figure 9. Contamination test from the ColxTsu transcriptome. Tissue enrichment from different sub-regions of the seed based on (Belmonte et al., 2013) within the ColxTsu embryo transcriptome. This analysis was performed based in the test proposed by Schon&Nodine, 2017. * for ColxTsu : ¤ for TsuxCol.

To analyze the impact of maternal contamination in the Col x Tsu embryo transcriptome, I proceeded to analyze the parent-of-origin expression of genes that are subregion-specific to the preglobular and globular embryo (Belmonte *et al*, 2013). According to Belmonte et al., subregion-specific transcripts were defined as those with fivefold or higher level of accumulation in one subregion relative to all others, at a given stage. Therefore, this class of transcripts should more faithfully reflect parent-of-origin expression of early embryos. To reinforce this analysis, 2-, 3-, and 4-Fold Change (FC) of enrichment were included.

Maternal transcript bias was detected for all fold change sets of genes analyzed (**Figure 10**). In the results using the full list of genes, 66% are Reciprocally Maternal Biased (**Figure 4**). In the subpopulation of genes that show 2FC enrichment in the embryo vs other seed tissues, 50% of genes were detected with Reciprocal Maternal Bias. The proportion of maternal transcript bias decreased slightly, in inverse proportion to the increase of the FC enrichment in embryos. However, Reciprocal and Non-reciprocal maternal biased genes still predominate in the zygote and octant stages (**Figure 10 A**). Interestingly, the percentages of biallelic expression were the lowest in all the gene sets analyzed, ranging from 4-11% of the genes in zygote stages (**Figure 4 A**). Moreover, the genes in each group that are expressed from zygote to heart embryos, show the same trend as previously observed where the maternal transcript bias is predominant in zygote and octant and gradually shifts to biallelic (**Figure 5, 10 B**). These observations suggest that maternal contamination detected by the "Contamination Test" does not alter the general conclusions about the predominant maternal bias during early embryogenesis. In addition, the motifs identified to be enriched in the maternally biased genes were overrepresented in the same class of genes at both 2 and 3 fold change (**Figure 10 C**). Although the motifs were not enriched in the other groups, this may be partly because of the small number of genes in other fold change groups.

Maternal contamination in early seed transcriptome profiling is always challenging, especially in ASE analysis in embryos (Baroux & Grossniklaus, 2015). The ColxTsu transcriptome reported here shows levels of contamination that are unlikely to interfere with the general conclusions. Several lines of evidence in the transcriptome point to the reliability of the data. For example, that fact that I could detect allele dominant transcripts is evidence for the small impact of seed coat contamination in early embryogenesis. This class of allele specific expression shows ecotype bias where the allele from a certain ecotype is preferentially expressed regardless of whether it was inherited maternally or paternally. In a scenario of high maternal contamination in preglobular stages, allele dominant genes would increase drastically after the dissipation of maternal contamination. In contrast, the ColxTsu embryo transcriptome shows constant levels of allele dominant expression from zygote stage. In addition, the set of genes with non-reciprocal maternal or paternal bias would also be also undetected in the presence of high levels of maternal contamination. Together, these observations demonstrate that preglobular embryos show a

pronounced bias toward maternal transcripts, which is not a consequence of contamination from maternal tissues.
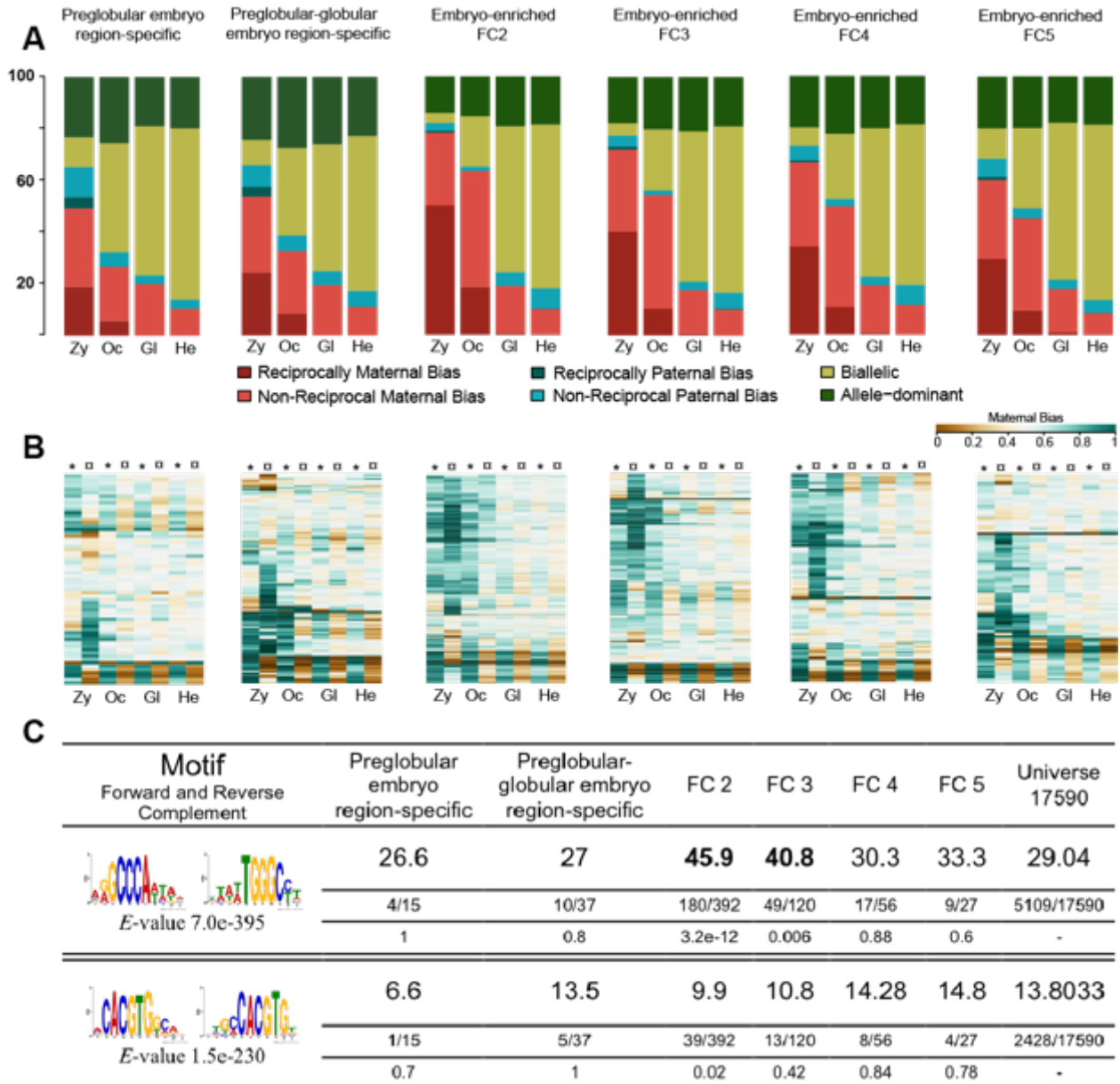
Figure 10. Parent-of-origin and motif analysis from gene groups enriched in Embryo Propter (Belmonte et al., 2013). A, Percentages of genes that show allelic bias expression during early embryogenesis. **B**, Heat map representation of the Maternal bias proportion by gene during early embryogenesis; only those gene expressed through zygote to heart embryo were included in the heat map. **C**, motif enrichment analysis of previously detected sequences overrepresented in maternally biased genes; Motif Occurrences in 1kb upstream regions (%) p<0.0001.
* for ColxTsu : **¤** for TsuxCol

Col x Tsu hybrid embryos show functional activation similar to isogenic Col

Allele Specific Expression studies rely on hybrids between polymorphic parents to discriminate between parental alleles. However, hybridization itself impacts metabolism, development,
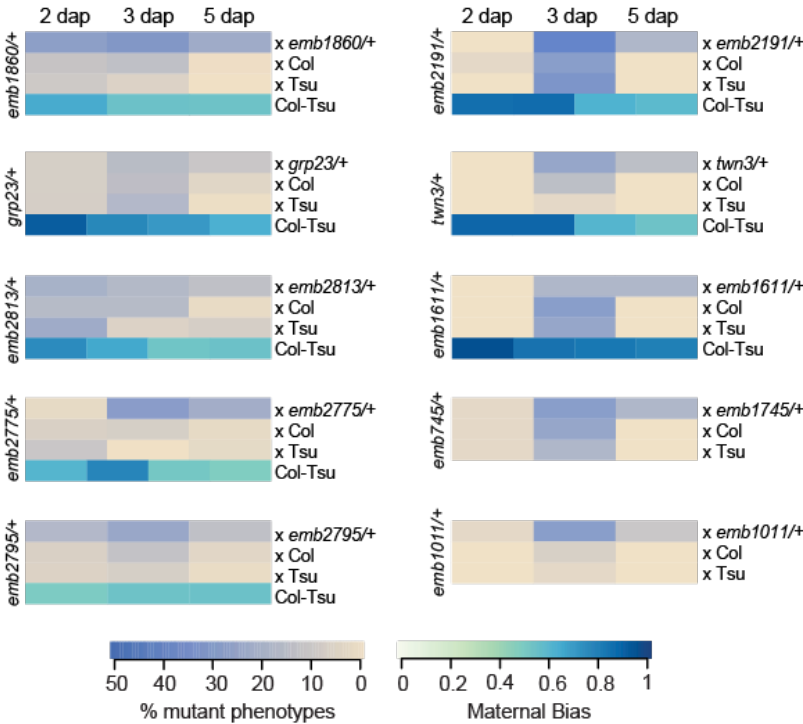
Figure 11. **Functional complementation dynamics of *emb/+* lines in isogenic and ColxTsu hybrid compared with gene expression.** From top to bottom each row represents functional complementation in hand-self heterozygous, pollination with Col, Tsu and equivalent expression data in one cross direction from ColxTsu transcriptome. (Frist column called from **Figure 2**, second column new data.)

epigenetic marks and gene expression (Del Toro-De León *et al*, 2016). Parent-of-origin effects in gene expression vary between Arabidopsis hybrid seeds (Pignatta *et al*, 2014). Two previous reports of ASE in embryos reported opposite conclusions regarding maternal and paternal contributions to early embryogenesis. Autran et al. (2011) studied Col x Ler hybrids in 2-cell stage and globular embryos, and found that most of the genes showed both maternal and paternal transctipts, but with a maternal bias. In contrast, Nodine &

Bartel (2012) used the Col x CVI hybrid combination (with reciprocal crosses) and found biallelic transcripts for almost all genes at all stages analyzed, (including 1-cell/2-cell and 8-cell embryos). In a previous study from our laboratory using functional genetic experiments in isogenic lines, we showed that the many paternal alleles failed to complement a maternally inherited mutation in early embryogenesis, suggesting delayed paternal allele activation (Del Toro-De León et al., 2014). Interestingly, complementation dynamics varied between hybrid and isogenic embryos, suggesting that parent of origin effects are disrupted in some hybrid genomes (Del Toro-De León *et al*, 2016). Different hybrid combinations lead to different complementation dynamics, where some showed hybrids showed earlier paternal genome activation than others. The Col x Tsu combination showed the closest complementation

dynamics compared to isogenic embryos (Del Toro-De León et al., 2014). To corroborate this observation, I conducted more genetic complementation experiments, and compared functional activation of the paternal allele with the Col x Tsu ASE transcriptomic data. I analyzed five *embryo defective (emb)* mutants in the Col background. By using pollen from Col (isogenic) and Tsu ecotype, I could compare complementation dynamic in isogenic and hybrid embryos. The results show that Col x Tsu hybrid embryos show delayed complementation by paternal alleles, similar to isogenic embryos (**Figure 11**). Hybrid and isogenic embryos show equivalent percentages of mutant embryos at 2 and 3dap, until the paternal allele is functional enough to complement the maternal mutant allele at 5dap. Interestingly, for the genes tested, the available Col x Tsu ASE data show initial maternal transcript bias, progressively becoming biallelic (**Figure 11**). Taken together, these observations suggest that the Col x Tsu hybrid shows delayed paternal activation at the function and expression levels, similar to the isogenic Col background. This corroborates the ASE profile of F1 Col x Tsu hybrid embryos as the parent-of-origin study most representative of gene expression in isogenic embryos.

Overall, our data shows maternal transcript bias during early embryogenesis, in agreement with Autran et al. (2011). However, these observations contrast with the Nodine & Bartel (2012) conclusions. The distinct results may be due to the methodologies used, experimental and biological noise, or hybrid effects (as observed previously) (**Figure 2 and 3**). To better understand these discrepancies, I re-analyzed the Nodine & Bartel data using the same methodology I applied for the Tsu x Col profile, and using the same criteria to categorize allele bias. As reported by Nodine & Bartel, most of the genes showed biallelic expression as early as 1-2 cell embryo stage; 33% of the genes in 1-2 cell embryo show biallelic expression, in contrast to the 2.5% that show Reciprocal Maternal or Paternal bias (**Figure 12 A**). Interestingly, all stages show genes with Non-Reciprocal Parental bias (not previously reported by Nodine & Bartel), with similar percentages as seen for the Col x Tsu dataset (about 20% of genes). I also identified a subpopulation of Allele Dominant genes present at the three developmental time points, with more genes being dominant in the Col than in the CVI background (**Figure 12 A, B**).
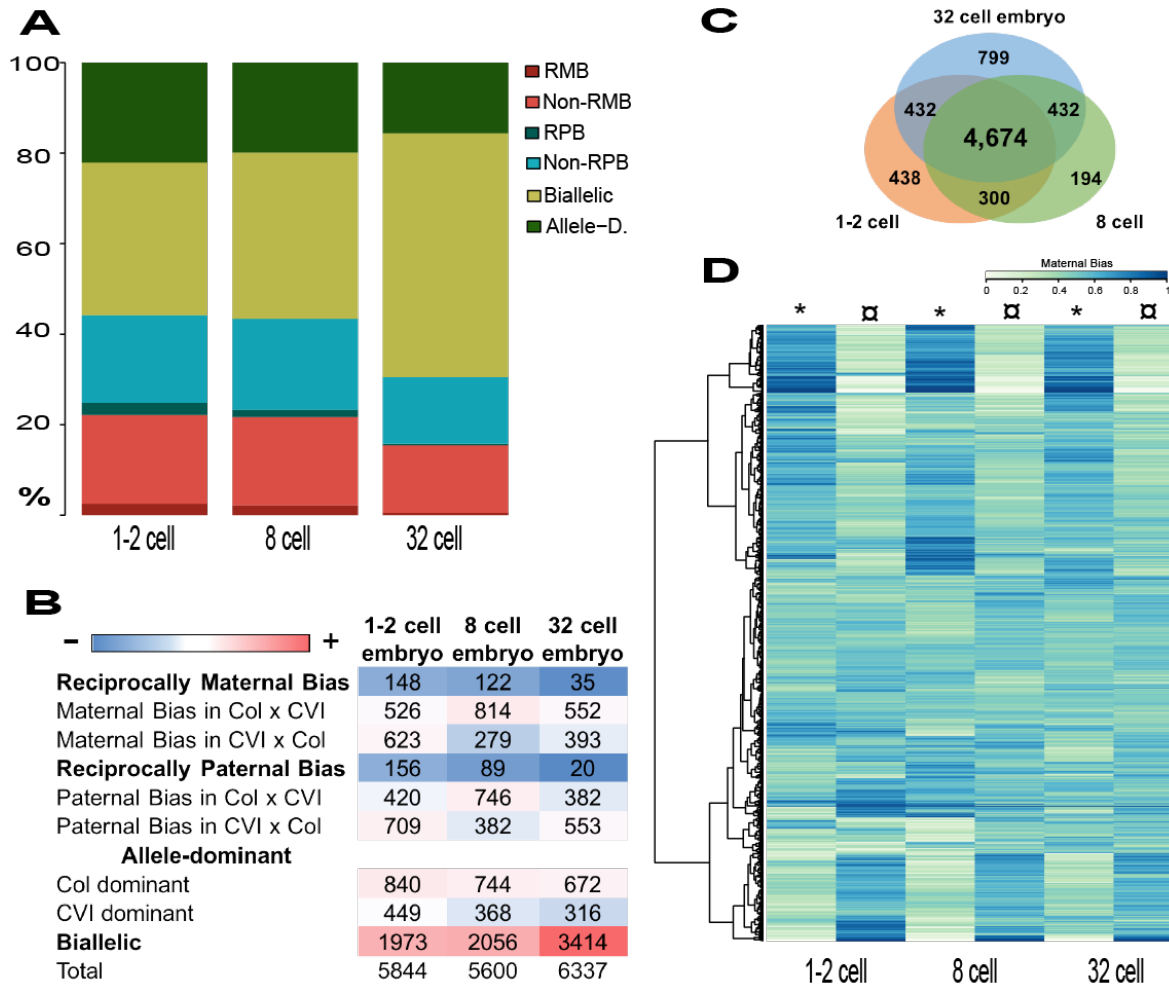
Figure 12. Parent-of-origin and motif analysis form Nodine&Bartel, 2012 data. A, Percentages of genes that show allelic bias expression during early embryogenesis. **B**, Full data of allele bias expression, color code show the accumulation trend with high (red), medium (white) and low (blue) number of genes. **C**, Venn diagram showing the genes that are common between embryo stages. **D**, Heat map of the 4,674 genes shared and expressed in the three profiled embryo stages showing the fluctuation of maternal bias expression. ＊ for ColxTsu : ¤ for TsuxCol

1-cell/2-cell and 8-cell embryos showed very similar patterns of allele-specific expression, suggesting that these preglobular stages share the same parent-of-origin effects on gene expression (**Figure 12 A, B**). Globular embryos show an increase in Biallelic genes (53%, compared to 36% in the 8-cell embryo). Accordingly, Reciprocal Maternal bias genes decrease (2.1% in 8-cell to 0.5% in globular embryo) (**Figure 12 A, B**). These observations suggest that (as described before by Nodine & Bartel), the Col x CVI hybrid embryo does not show a maternal transcript bias in early embryogenesis, as most of the genes are present biallelically as early as 1-2 cell embryo. However, we observed an increase of biallelic genes from preglobular to globular stages suggesting that, although less severe, a reprograming is ongoing even in this

hybrid combination. The increase in biallelic genes in globular stages coincides with a slight decline of Non-Reciprocal Parental Bias genes, Dominant genes, and Reciprocal Parental Bias genes (**Figure 12**). Likewise, the progress of maternal bias expression during development from the genes that are expressed in the three time-points tested (**Figure 12 C**) shows a general bias towards biallelic transcripts, which increase by the globular stages (**Figure 12 D**). Interestingly, the differences between the maternally biased Col x Tsu transcriptome and the more biallelic Col x CVI transcriptome was reproduced when comparing expression and functional data reported in this work (**Supplementary Figure 6**).

These observations were overlooked by Nodine & Bartel (2012), due to the methodology employed to analyze parent of origin bias. Instead of using percentages of maternal and paternal SNP-containing reads, they used proportional expression values (proportion of maternal expression over paternal) and the data obtained from the reciprocal crosses were averaged into a single value. The averaging of the reciprocal crosses reduced or removed all the evidence of non-reciprocal expression and dominant expression. This partial analysis avoided the observation of different kinds of allele biased expression during embryogenesis, and exaggerate the biallelic expression that was reported as the dominant class of expression during embryogenesis.

Discussion

Parent-of-origin effects on gene expression in plants have been widely accepted to occur in the endosperm, due to imprinting. However, genome-wide parent-of-origin effects on gene expression during early embryogenesis has been controversial, despite the comprehensive literature documenting this phenomenon. Expression analysis from reporter lines, single-allele bias quantification, functional analysis, transcriptome studies and epigenetic mechanisms largely described in literature, are evidence of the complex transcriptional regulation that takes place during early embryogenesis (Autran *et al*, 2011; Baroux *et al*, 2001; Del Toro-De León *et al*, 2014; Grimanelli *et al*, 2005; Springer & Stupar, 2007; Vielle-Calzada *et al*, 2000). Therefore, the idea of purely biallelic expression established just after fertilization is not congruent with the current *state of the art* of plant developmental biology.

Part of the controversy is due to the intrinsic difficulty of the methodologies applied, and the natural variation present in this phenomenon. Here, the ASE analysis of reciprocal Col x Tsu F1 embryos demonstrates that Zygote and Octant stages showed an overwhelming maternal transcript bias. Biallelic expression increases progressively with embryo development and predominates from the globular stage onwards, after a major reprograming of parental genomes. Interestingly, despite possible levels of maternal contamination, tissue-specific and embryo enriched genes recapitulate the parent of origin bias observed with the full data. In addition, the high activity of *de novo* expression in Arabidopsis suggests that zygotic genome is activated immediately after fertilization. Identification of motif enrichment in the promoter regions of maternally biased genes suggests a transcriptional regulation mechanism that may be partly involved in the preferencial expression of maternal alleles. Finally, although ASE varies in hybrid genomes, functional data corroborates that the Col x Tsu hybrid combination reproduces parent-of-origin effects similarly to isogenic Col-0 background.

Previous work suggested delayed paternal activation. This work is the first comprehensive transcriptome study that includes functional analysis, showing parent-of-origin effects in the Col x Tsu hybrid that are similar to the isogenic Col. ASE analysis of this data show that 66% and 49% of the expressed genes in zygote and octant stages exhibit Reciprocal Maternal bias, and most of these genes undergo reprogramming to Biallelic by the globular or heart stage. GO analysis of the Reciprocal Maternal bias genes suggests that the maternal genome has prevalent functions in transcription and translation. Interestingly, reprograming is reproduced by GO terms, when these essential functions are transferred to Biallelic genes as embryo development proceeds. These observations underline the importance of the maternal genome during early embryogenesis (Baroux & Grossniklaus, 2015). Genetic screens for female gametophyte mutants that were defective in post-fertilization processes and were not rescued by the paternal (pollen) allele, suggested functional requirements for the maternal contribution beyond gametogenesis (Pagnussat *et al*, 2005; Ngo *et al*, 2012). More recent reports have demonstrated that transient maternal dominance of early embryogenesis has important implications for embryo development. Luo et al., showed that in early embryos of tobacco interspecific hybrids, Programed Cell Death (PCD) in the suspensor is regulated by the predominant maternal allele expression of the *NtCYS* cystatin gene. Although both parental transcripts contribute to PCD, the

higher expression and subsequent reduction of the maternal allele was necessary for the initiation of the PCD in the suspensor and proper progress of the embryogenesis (Luo *et al*, 2016). Moreover, He et al. (2017), showed that *de novo* transcription of the maternal *NUWA* allele in the zygote and early embryo was essential for mitochondrial function. Interestingly, the paternal allele of NUWA gene was first detected only after the 16-cell stage (He *et al*, 2017). These findings correlate with our observations where after a period of maternal transcript enrichment at the zygote and octant stages, most maternally biased genes became biallelic from the globular to the heart stage (Alonso-Peral *et al*, 2016). Interestingly, in a recent ASE study of dermatogen-globular and heart stage C24 x Ler reciprocal hybrid embryos, most genes had equal proportions of transcripts from each parent. This is consistent with the observation that at the globular stage, most parent-of-origin effects have concluded. These data demonstrate that parent biased gene expression is one of the major features of zygote and octant stages of embryo development and undergoes progressive resetting involving significant reprogramming during subsequent stages of embryogenesis.

The mechanisms involved in parental genome reprograming during early embryogenesis are still largely unknown. Maternal dominance of early embryogenesis may be explained by the inheritance of maternal products from the female gametophyte. As in animals, this mechanism implies a passive transmission of mRNA from the egg cell to the zygote, and subsequent diffusion or active degradation (Tadros & Lipshitz, 2009; Yartseva & Giraldez, 2015). Although our understanding of inheritance of maternal products to the early zygote remains is poor, a few examples have demonstrated the importance and feasibility of active communication between reproductive cells (Bayer *et al*, 2009; Costa *et al*, 2014). Preliminary observations in the laboratory suggest that there is passive transmission of reporter expression from central cell (CC) and egg cell (EC) to the early endosperm and embryo (**Supplementary Figure 3 and 4, Supplementary Table 11 and 12**). Reporter lines specifically expressed in the CC and EC, where the endogenous promoter does not show transcriptional activity after fertilization, show reporter signal when the transgenic line is used as maternal donor and crossed with wild-type pollen. Although the signal is transient, this suggests a passive diffusion of proteins and presumably transcripts from the female gametophyte to the early embryo. In fact, peptide signals derived from the fertilized central cell (which generates the endosperm) have been shown to be

important regulators of embryo patterning (Bayer *et al*, 2009; Costa *et al*, 2014). In the present study, ~50% of the genes with Reciprocal Maternal Bias show expression in the egg cell, suggesting that carryover may have a role in the maternal transcript enrichment in the early embryo (**Figure 7**). Interestingly the additional 50% of the genes with maternal bias expression may be the consequence of *de novo* expression of maternal alleles. Previous reports have suggested that RdDM pathway repression may prevent precocious transcription of paternal alleles (Autran et al., 2011), and thus *de novo* expression may be more likely to occur from the maternal genome. However, the true mechanisms for preferential maternal transcription remain unkown. Here I report that maternally biased genes during early embryogenesis show CACGTG and TGGGC motif enrichment of their promoter sequences. These motifs may be the target sites of trans factors that may activate transcription of maternal alleles, or repress transcription of paternal alleles. Further studies are needed to clarify the role of these motifs in transcriptional regulation during embryogenesis, as well as the biological relevance of parent-of-origin expression during early embryogenesis, and its implications on agronomic traits in hybrids.

Part 3. **Paternal gene activation during early embryogenesis**

Summary

Analysis of the Col x Tsu transcriptome gave us a comprehensive scenario of allele specific expression at the genomic level during embryogenesis, and the overwhelming parent-of-origin expression reprograming during the transition from the zygote to globular stage (Part 2). To validate genomic observations on a single gene basis, I started the analysis of candidate genes. Here I show results for *NF-YB2* and *EMB27678*. In addition, I evaluated the effects of previously reported pathways involved in the repression and activation of the paternal genome, using a *pNF-YB2::GUS* reporter line.

pNF-YB2::GUS shows a maternal expression bias in early embryogenesis

Certain *cis* regulatory sequences were enriched in maternally biased genes, suggesting that transcriptional regulation pathways targeting these motifs may have a role in the preferential activation or repression of parental alleles. Fifty percent of the reciprocal maternal bias genes, which contain motif enrichment according with FIMO, are not present in transcriptomic data from egg cell (**Figure 7**), suggesting *de novo* preferential expression of maternal alleles. One hypothesis is that the motifs trigger maternal allele *de novo* expression, in parallel with repressive marks in the paternal genome. To test this hypothesis, I identified several maternally biased genes with motifs enriched in their promoter region and with no reports of expression in the egg cell. Among the candidate genes I found *NF-YB2* (AT5G47640), which belongs to a heterotrimeric transcription factor complex present in nearly all eukaryotes, composed by the subunits NF-YA, NF-YB, and NF-YC (Zhao *et al*, 2017). The combinatorial capacity and the presence of multiple paralogs of these genes in plants, have led to a wide variety of complexes and patterns of expression. Multiple essential roles in plant growth, development, and stress responses have been associated with these transcription factors (Zhao *et al*, 2017; Hou *et al*, 2014; Siefers *et al*, 2009). Transcriptomic data shows that *NF-YB2* gene has maternal specific expression at zygote and octant stage and contains two copies of the CACGTG motif and four of the TGGGC motif within the first 1kb upstream to the start codon (**Figure 13**). To further validate whether *NF-YB2* gene shows parent-of-origin expression, I analyzed it's the expression pattern of the reporter line *pNF-YB2::GUS,* in reciprocal crosses (Siefers *et al*, 2009).
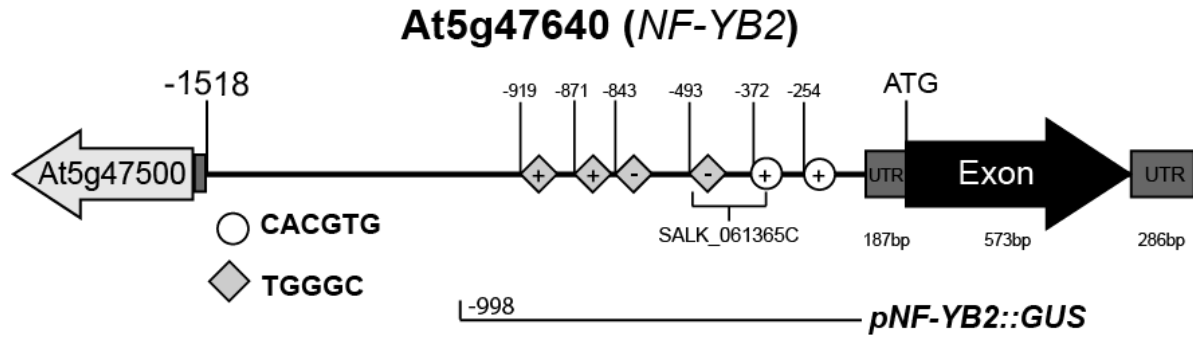
Figure 13. Representation of the structure of the NFYB2 locus. The genomic position of the enriched motifs in maternally biased genes and located in the promoter region of NF-YB2 is represented.; coordinates are estimated by taking the NF-YB2 transcription initiation site as a reference (+1). Putative location of the nf-yb2 insertion is illustrated. The region cloned for the *pNF-YB2::GUS* construct also represented.

Reporter expression was very strongly detected in 1-2 cell embryos, and in endosperm and maternal tissues, when the reporter line was used as maternal donor in an overnight GUS staining treatment (**Supplementary Figure 7**). Maternal expression was very strong in embryos from 1dap, and the expression was present during all the stages of early embryo development. I observed a gradual increase in the expression from 1dap to 3dap, suggesting that the maternal genome also shows gradual *de novo* activation (**Figure 13 A-E**). In addition, the expression in embryos started to decay in intensity and proportion by heart-torpedo stages. In contrast, when the reporter line was used as a paternal donor, the expression was undetected from 1 to 3dap, and triggered by 4dap (**Figure 13 F-J**). The sudden expression at late globular-early heart stage, suggests a quick mechanism of paternal allele activation taking place at this developmental stage. Interestingly *pNF-YB2::GUS* does not show reporter expression in mature female gametophytes (**Supplementary Figure 7**) reinforcing the hypothesis of *de novo* transcription of maternal alleles. Using longer GUS staining treatment (4 days of incubation instead of 1 day), a small percentage of embryos show a less intense paternal signal, suggesting that paternal allele activation may take place gradually starting 2dap. Using longer incubation periods, the most variable time-point was 3dap, suggesting that at this developmental time the activation is in process, or that the repression is gradually being removed (**Figure 13, 14**). These observations suggest that (in agreement with the transcriptomic data) the *NF-YB2* reporter line shows a maternal specific expression during the earliest stages of embryogenesis, which in turn undergoes reprogramming towards biallelic expression, presumably at 3dap and accomplished by 4dap.
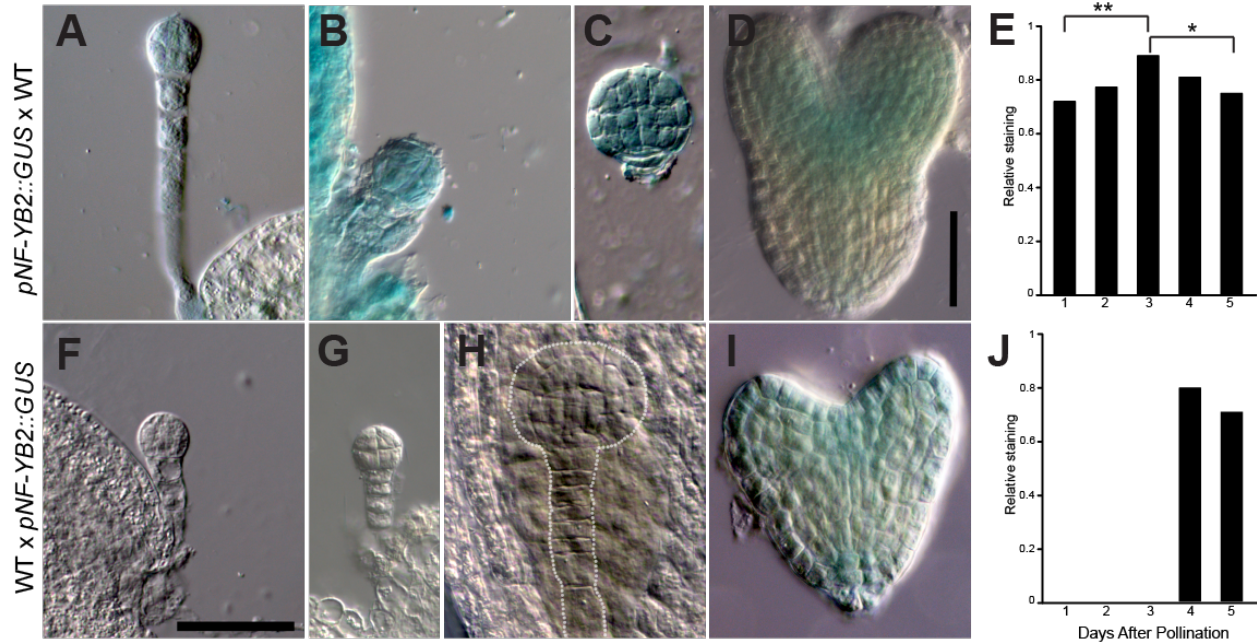
Figure 14. Parent-of-origin expression of the *pNF-YB2::GUS* reporter line. A-E, expression of maternal *pNF-YB2::GUS*. **F-J**, paternal expression of *pNF-YB2::GUS*. Results corresponds to one over-night period with GUS staining solution. Statistically significant *p* values for Fisher's Two-Tailed Exact Test ** $p \leq 0.01$; * $p \leq 0.05$. Full table data shown in Supplementary Table 4.

pNF-YB2::GUS expression is partly affected by canonical pathways

The mechanism regulating parent-of-origin expression during embryogenesis remains mysterious, though some reports indicate that epigenetic pathways that may be involved in this process. The RNA-directed DNA methylation (RdDM) pathway, a small interfering RNAs (siRNAs) mediated epigenetic silencing mechanism, has been proposed to play an important role in parent-of-origin expression during ZGA. Maternal mutant backgrounds of the RdDM pathway have shown precocious expression of paternal reporter lines, suggesting that DNA methylation is involved in paternal genome repression during early embryogenesis (Autran *et al*, 2011). To test whether a similar mechanism is responsible for delayed expression of paternal *pNF-YB2::GUS*, I evaluated paternal allele activation in maternal mutants for important genes involved in the RdDM pathway. I used maternal homozygous null mutants of *NRPD1B* (the largest subunit of nuclear DNA-dependent RNA polymerase V), and of the histone methyltransferase *KRYPTONITE* (*KYP*, responsible for H3K9 methylation, which constitutes a reinforcing loop between histone and DNA methylation) (Law *et al*, 2013; Du *et al*, 2015). Contrary to expectations, the crosses using maternal mutant RdDM pathways, *kyp-4* **x** *pNF-YB2::GUS* and *nrpd1b-11* **x** *pNF-YB2::GUS,* still showed a slight delay of expression of the paternal allele at

2dap. Interestingly, this delay remains in the *nrpd1b-11* cross at 3 and 4dap compared with the wild-type control (**Figure 15 A, Supplementary Figure 8 and Table 5**). These results suggest that p*NF-YB2::GUS* may not be regulated by the maternal RdDM pathway, and that another maternal repressor, normally silenced in the wild-type background, is activated in mutant ovules and slightly silences the activity of the reporter line during early embryogenesis, since its expression is normalized at 5dap in both crosses.

Disruption of paternal DNA-methylation also has been reported to affect imprinted expression of some genes in endosperm, and may disturb parent-of-origin in embryogenesis as well. To test whether the paternal RdDM pathway is involved in the repression of paternal p*NF-YB2::GUS*, I introgressed the marker line into the *kyp-4* and *nrpd1b* backgrounds and evaluated paternal activation when crossed to maternal wild-type plants. In both epigenetic mutant backgrounds, paternal activation of the reporter line shows slight de-repression with higher expression at 3dap. Interestingly, the complete activation of the paternal allele occurs faster than in wild-type (**Figure 15 B, Supplementary Table 7**), suggesting that although the precocious activations, this is not enough to fully activate the paternal allele and in addition maternal silencing extends beyond globular stages. This observation suggests that changes in methylation may activate the paternal alleles earlier than expected when hypomethylated pollen fertilizes wild-type plants, however this may trigger a stronger repression response by the maternal genome which attempts to counter the paternal contribution. This last hypothesis should be tested further.

Polycomb Repressive Complex 2 (PRC2) catalyzes the trimethylation of Histone 3 at Lysine 27 (H3K27me3), which is the hallmark of a repressive chromatin state involved in a variety of developmental processes (Hennig & Derkacheva, 2009). Interestingly maternal PRC2 is involved in the repression of imprinted genes in plant embryos and endosperm (Raissig *et al*, 2013; He *et al*, 2017). To further explore the mechanisms involved in delayed paternal allele expression, I evaluated the paternal expression of p*NF-YB2::GUS* in *fie* maternal homozygous null mutants. *FIE* encodes the unique Extra sex combs (Esc) protein in plants, one of the components of the PRC2 complex. Interestingly, there was no precocious expression of the paternal allele, but instead I observed slower allele activation, similar to the observations in the *kyp-4* maternal background. These data suggest a cooperative mechanism between both K9 and K27H3

methylation to silence an unknown repressor that otherwise represses paternal activation of the p*NF-YB2::GUS* reporter line during early embryogenesis (**Figure 15 A**). Previous reports have shown an interplay between H3K9me2/3 and H3K27 histone repressive marks in human cell lines (Boros *et al*, 2014). Thus, a similar phenomenon may take place in plants. To test whether there is a convergent function for these two chromatin repressive marks, further studies should evaluate the paternal activation of p*NF-YB2::GUS* in *fie* and *kyp-4* double mutant backgrounds. In addition, research should be pursued to discover the unknown repressive factor that may be activated upon chromatin deposition as mentioned above. Experiments such as a genetic screen in the p*NF-YB2::GUS* background, may lead to the identification of mutants with precocious activation of paternal reporter activity. The identification of mutated genes could uncover new regulators of paternal genome contributions.

In addition to repressive mechanisms regulating parent-of-origin expression during embryogenesis, one activation process involved in chromosome assembly has been reported (Autran *et al*, 2011). The CAF1 complex, composed by FASCIATA1 (FAS1), FAS2, and multiple suppressor of IRA1 (MSI1), functions as an H3/H4-specific chaperone facilitating nucleosome assembly during replication (Hennig *et al*, 2005; Kaya *et al*, 2001). The CAF1 complex promotes paternal allele activation, and thus mutants in the CAF1 complex delay paternal allele activation (Autran et al., 2011). To test whether a similar mechanism regulates the expression of *pNF-YB2::GUS*, I tested its expression in a maternal mutant *fas2*. I found a one day delay in onset of expression of the reporter gene, compared to wt. These results support previous observations that CAF1 nucleosome assembly is necessary for the normal activation of the paternal genome (Autran et al., 2011) (**Figure 15 A, Supplementary Table 5**). Since *pNF-YB2::GUS* activation is observed at later stages, other factors must compensate for lack of CAF1 function later in development.
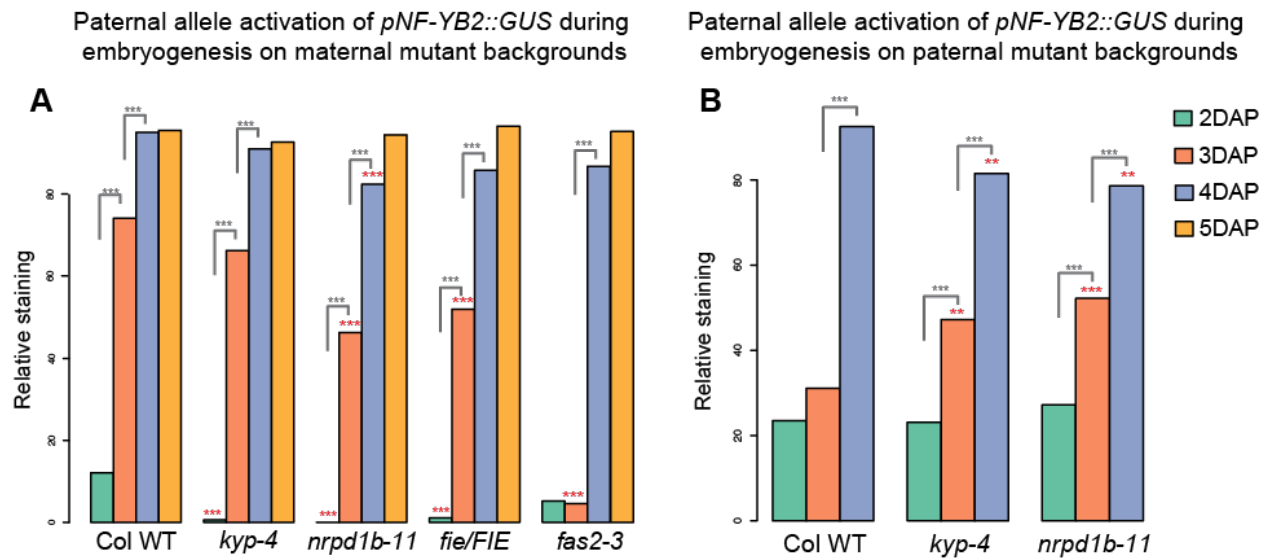
Figure 15. Parent-of-origin expression of *pNF-YB2::GUS* in parental mutant backgrounds. A, paternal *GUS* expression in maternal mutant backgrounds. See Supplementary Figure 8. **B**, paternal *GUS* expression in paternal mutant backgrounds. Mutant plants were confirmed by PCR. Introgressions were confirmed by both PCR and BASTA resistance. Statistically significant *p* values for Fisher's Two-Tailed Exact Test ** $p \leq 0.01$; * $p \leq 0.05$. Grey asterisks refer to comparisons between time-points of the same mutant, red asterisks refer to comparisons with wild-type reference. Full table data shown in Supplementary Table 5 and 7.

Taken together, these data show that the mechanisms for regulation of parent-of-origin expression during embryogenesis are more complex than previously thought. More than one process looks to be involved in the repression of parental contributions. Further studies are needed to identify the additional factors that regulate paternal expression. For example, additional individual reporter gene experiments are required to evaluate the scope of mechanisms of regulation and its functional implications in plant embryogenesis. RNA sequencing experiments will be necessary to test the effect of CAF1 and RdDM mutations on paternal silencing and activation at the genome level.

Parent-of-origin expression of *pEMB2768::tdTomato*

The above data suggest that multiple mechanisms are involved in parent-of-origin expression during early embryo development. To extend the tools to study allele specific expression in embryo development, I generated reporter lines from genes that showed delayed functional complementation of maternal mutations by paternal alleles (Del Toro et al., 2014). Here I report preliminary results for the reporter line *pEMB2768::tdTomato*. *EMB2768* encodes a Tyrosine tRNA Synthetase which functions in translation in mitochondria and chloroplasts. 2,023bp

upstream of the translation start site were cloned and transformed in a binary Gateway vector (See Materials and Methods section).

The wt paternal allele of *EMB2768* is not able to complement maternal loss of function during early embryogenesis (Del Toro et al., 2014), suggesting that in early embryogenesis the paternal allele is not expressed or its levels of expression are not sufficient to complement loss of maternal function. Another possibility is that *emb2768* is haploinsufficient, so both parental copies are necessary to accomplish its function. However, previous observations on reciprocal crosses have shown that this is not the case for *emb2768*, and that parent-of-origin effects are the responsible of these observations (Del Toro et al., 2014). Thus, absence or decreased expression of paternal *EMB2768* most likely explains the lack of paternal complementation of the mutant maternal allele. To test this hypothesis, I analyzed parent-of-origin expression of the *pEMB2768::tdTomato* reporter line.

Reciprocal crosses of *pEMB2768::tdTomato* with wild-type plants show that maternal expression is detected at high intensity from 2dap, and continues throughout embryogenesis (**Figure 16 A-F**). Expression was also strongly detected in maternal seed coat tissues and in the endosperm. Paternal expression was also detected from 2dap, at approximately the same penetrance, but with less intensity than maternal expression at the same stage (compare Figure 16 D and J) (**Figure 16 G-L**, **Supplementary Table 8**). This result supports the hypothesis that low (rather than absent) expression of paternal *EMB2768* is responsible for delayed complementation of the maternal mutant allele (Del Toro et al., 2014). Thus, a different parent-of-origin expression pattern for *pEMB2768::tdTomato* was found than for *pNF-YB2::GUS*. The reporter line for the *NF-YB2* did not show significant levels of paternal expression early in embryogenesis compared with maternal expression, while the reporter for *EMB2768* did (with low signal). Thus, different repressive mechanisms may take place to repress the paternal alleles of these genes.
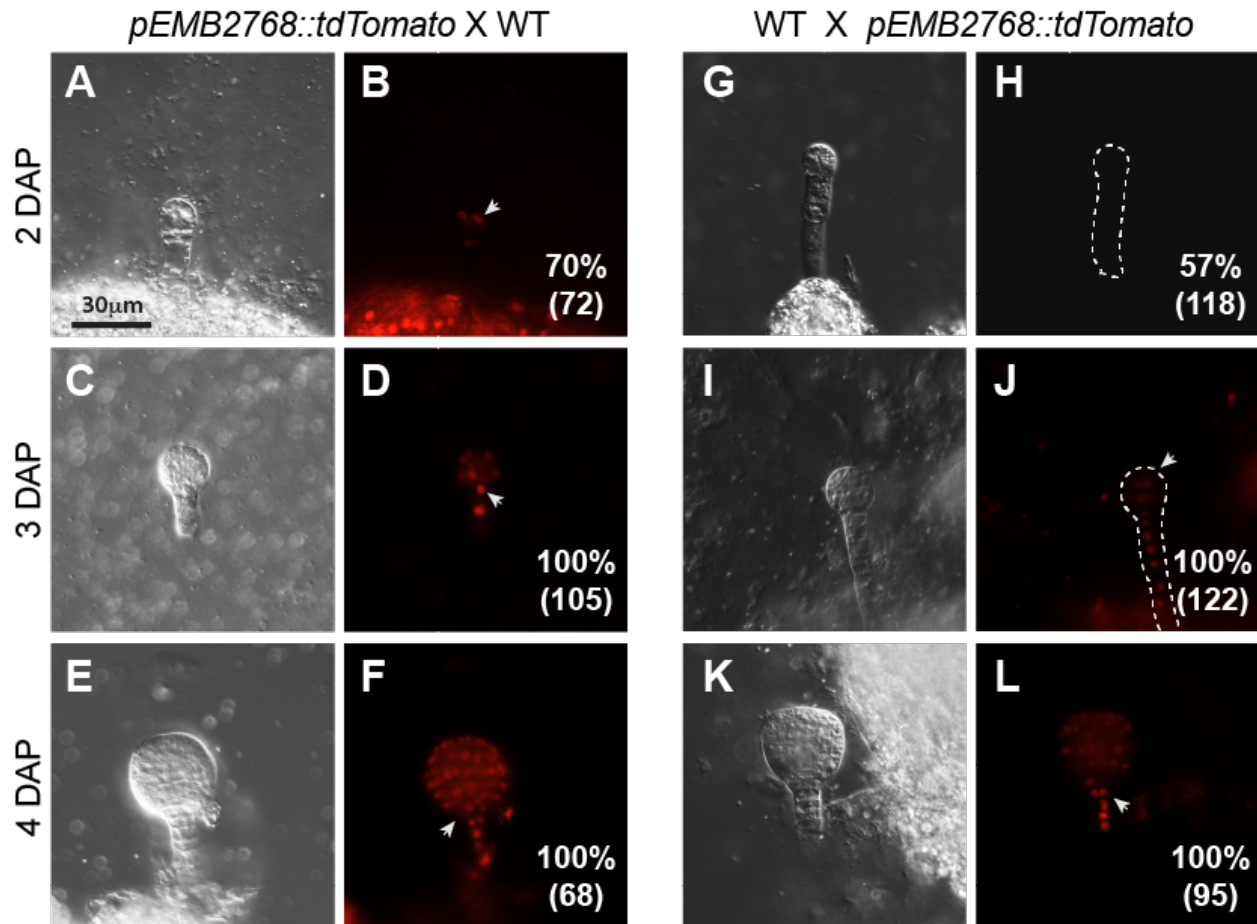
Figure 16. Parent-of-origin expression of *pEMB2768::tdTomato*. A-F, maternal reporter expression when pollinated with wild-type plants. **G-L,** paternal reporter expression in maternal wild-type background. White arrows indicate reporter expression. Dotted lines indicate embryo margin. Full data shown in Supplementary Table 8.

Discussion

Transcriptomic data suggested an overwhelming enrichment of maternal expression during embryogenesis (**Part 2**). However individual gene analysis is necessary to confirm the transcriptional information. Here I report the analysis of two reporter lines, whose results suggest that different mechanisms may be regulating parent-of-origin expression during early embryo development.

The NF-YB2 gene shows maternal specific expression in the ASE Col x Tsu transcriptome. The *pNF-YB2::GUS* transcriptional reporter line shows early maternal expression, while paternal expression is delayed, as predicted by transcriptomic observations (**Figure 14**). Interesting results were found when different GUS solution incubation periods were tested. With one day of

staining, maternal reporter expression was detected beginning at 2dap, while paternal expression was not detected until 4dap. However, 4 day GUS staining allowed detection of weak paternal reporter paternal activity at 2dap. This result suggests that maternal and paternal expression levels are comparable only after globular stages, when the detection threshold by means of GUS staining sensitivity is the same for maternal and paternal alleles. These observations support the hypothesis that the paternal genome is repressed, showing low expression earlier in embryogenesis and then biallelic expression later in development.

The mechanisms regulating parental contributions during embryogenesis are largely unknown. However epigenetic pathways have been proposed to work in the repression of paternal genome (Autran *et al*, 2011; Raissig *et al*, 2013). The *pNF-YB2::GUS* reporter did not show expression in female gametophytes, as predicted by microarray profiles of egg cells in Arabidopsis (**Supplementary Figure 7**) (Wuest *et al*, 2010). In addition, the maternal genome also shows a gradual increase in expression from 1 to 3dap (**Figure 14**). These observations suggest that cooperative mechanisms between epigenetic repression of the paternal activity, with promotion of maternal allele specific transcription, may lead to the monoallelic expression of *NF-YB2*. How this transcriptional activation or repression occurs is not known. One hypothesis is that enriched motifs in the promoter regions may target epigenetic repression of the paternal allele, or activation of the maternal allele. The upstream region of *NF-YB2* includes the enriched motifs identified in Part 2. Further studies should be aimed at elucidating the relevance of these motifs for parent-of-origin expression of this gene (see Perspectives). Interestingly, an insertion line predicted to affect the function of *NF-YB2* shows patterning phenotypes in the suspensor and embryo proper, especially during zygote-globular stages (**Supplementary Figure 9, Supplementary Table 10**). This preliminary observation suggests that maternal expression of this transcription factor may be important for embryo development. Recent reports support the functional implications of maternal dominance during early embryogenesis (Luo *et al*, 2016; He *et al*, 2017; Ueda *et al*, 2017).

Gene silencing pathways previously found to affect parent-of-origin expression during early embryogenesis show unexpected interactions with the *pNF-YB2::GUS* reporter line. In the *kyp-4* maternal mutant background, paternal activation of the *pNF-YB2::GUS* reporter gene is more

delayed than in wt (**Figure 15 A**). Thus, my data suggest that the *NF-YB2* gene is not directly regulated by the RdDM pathway, in contrast to the genes tested by Autran et al. (2011). The PRC2 pathway was previously found not to affect paternal silencing (Autran et al., 2011), also in contrast to my result that maternal *fie* mutation causes even more delayed paternal expression of p*NF-YB2::GUS*. One hypothesis to explain this difference is that an unknown factor is activated because of the disturbance in chromatin silencing in *kyp-4* and *fie*. However, this idea needs to be further explored. Interestingly, the same experiment showed unexpected results for paternal expression of p*NF-YB2::GUS* in the endosperm. Although the focus of my experiments was gene activation in embryos, I noticed precocious paternal expression of p*NF-YB2::GUS* in the endosperm at 3dap in the *nrpd1b-11* maternal background. In contrast, endosperm expression in the *kyp-4* background was decreased, suggesting that embryo and endosperm show different mechanisms of regulation of parental contributions, in particular for the *NF-YB2* gene (**Supplementary Figure 10, Supplementary Table 6**).

Hypomethylated sperm carrying the transcriptional reporter line triggered more gradual activation of the paternal allele, causing extended maternal dominance (**Figure 15 B**). A recent report found similar results in parent-of-origin expression in endosperm, where paternal inheritance of mutations in other RdDM components, in this case Pol IV, causes a maternal transcript bias (Erdmann et al., 2017) These results suggest that, in the case of embryogenesis, after fertilization with hypomethylated sperm, maternal stress responses may be triggered, extending the maternal dominance of embryogenesis beyond late globular stages. This hypothesis may be tested by using double or triple mutants of different methylation pathways (i.e. *met1*, affecting CG methylation) to determine whether increased hypomethylation in paternal genomes triggers greater maternal dominance of embryogenesis.

Different mechanisms seem to be involved in the regulation of parental contributions during early embryogenesis. Observations of paternal activation of the *pEMB2768::tdTomato* reporter line suggest that even with paternal expression, differential levels of expression of maternal and paternal alleles may impact at the functional level. Heterozygous *emb2768/+* showed delayed functional complementation when pollinated using wild-type plants. However, I did not find significant differences in the timing of onset of maternal vs paternal *pEMB2768::tdTomato*

expression, but I did detect differences in the levels of expression (**Figure 16**). Using the reporter line as maternal donor, the signal was strongly detected starting 2dap, and the expression was detected during all the embryo developmental stages. In contrast, paternal expression was very low, judging from the less intense brightness observed in the embryos compared with maternal expression (**Figure 16**). Interestingly paternal expression levels were comparable to the maternal levels after 4dap, suggesting a complete activation of the paternal allele. All these data suggest that ZGA is gradual and not a presence-absence phenomenon, instead, levels of expression need to be evaluated. Further studies aimed at quantification of reporter gene expression are necessary to confirm the differential levels of expression of maternal and paternal genomes.

Discussion

The findings reported here contribute new knowledge regarding activation of the zygotic genome, as well as parental contributions during early embryogenesis in Arabidopsis. Among plant developmental biologists, there has been a long-standing controversy about the differential expression of maternal and paternal genomes during early embryogenesis. The lack of consensus is partly due to the challenging nature of studying plant embryos, which are surrounded by the endosperm and sporophytic maternal tissues, and therefore difficult to isolate. Previous transcriptomic studies of early embryos reported apparently opposite conclusions regarding parental transcript contributions to the early embryo. Although tissue contamination from maternal tissues was argued to be the cause of maternal bias transcript in one of the studies, this conclusion is not in agreement with other reports about the maternal control of embryo development. Since SNPs are the basis of allele-specific profiling, hybrids need to be employed for parent-of-origin studies. My thesis demonstrates that hybrids can disturb parent-of-origin effects, as shown by different complementation kinetics of maternally inherited mutations in hybrid embryos. All hybrid combinations I tested showed faster complementation that in an isogenic background, suggesting that zygotic activation occurs earlier in hybrid embryos. These observations provide a plausible explanation for the opposite conclusions reached by previous transcriptomic analyses, and suggest a role for parent-of-origin gene expression in hybrids, the basis of modern agricultural production.

The allele specific expression analysis of a new hybrid embryo transcriptome is an important contribution to studies of differential allelic expression during seed development. This Col x Tsu hybrid showed the closest functional behavior of parent-of-origin effects to isogenic Col, making this the best existing transcriptome to study parental contributions during early embryogenesis. Despite the maternal contamination detected in this transcriptome, the findings regarding the overwhelming maternal transcript enrichment in zygote and octant stages, are reproducible with different subpopulations of embryo enriched genes. Analyses of maternally biased genes show that early embryogenesis is controlling the developmental program not only because of the abundance of the maternal transcripts, but also by their associated functions. In addition, highly enriched sequences detected in maternally biased genes shed light to the mechanisms that may be

involved either in the maternal transcription or paternal silencing. One hypothesis is that epigenetic mechanisms allow the repression of paternal alleles together with specific activation of maternal alleles, likely mediated by *cis-trans* interactions. As a consequence, there is maternally biased transcript enrichment during zygote and octant stages, after which reprograming of the new genome equilibrates parental contributions.

Recent reports have documented the importance of maternally biased transcription during the earliest stages of plant embryo development. Important developmental processes such as establishment of embryo-endosperm cell identity, programed cell death and mitochondrial development have been reported to be specifically controlled by maternal alleles. In all these cases, biallelic expression is observed later in development.

My observations have demonstrated that maternal transcript bias is regulated by more than one mechanism. Previous studies show that the epigenetic RdDM pathway is involved in the repression of the paternal genome. My results show that another mechanism regulates activation of the paternal allele of the *pNF-YB2::GUS* reporter line, perhaps a repressor which is activated upon changes in methylation derived from loss of the RdDM pathway. My results showed that maternal bias may be due to incomplete repression of paternal transcription, instead of a total absence of paternal transcription. Although paternal activity in reporter lines is observed, it is considerably less than maternal allele activity. These observations suggest that although the repression of the paternal allele is not complete, it is strong enough to negate its function, as suggested by experiments to complement the function of mutant maternal alleles with the wt paternal allele.

Thus, maternal control of embryogenesis is a genomic phenomenon. Further research should be focused on understanding the mechanisms that regulate the process, as well as the relevance of parent-of-origin expression during early embryogenesis for agronomic traits in hybrids.

Perspectives

***cis-trans* regulation of parent-of-origin expression in Arabidopsis**

One of the most striking findings in this study is that parent-of-origin effects are disrupted in hybrids. Functional and transcriptomic data suggest that the source of this variation may reside in alteration of the mechanisms that regulate parental contributions during early embryogenesis. Since enriched motifs were identified in the upstream sequences of maternally biased genes, an interesting experiment to understand the function of *cis* variation in these motifs would be to study promoter sequence variation of genes that show delayed vs immediate functional activation in different hybrid combinations. In addition, cloning and transformation of the divergent upstream sequences from different ecotypes into reporter genes may help to elucidate whether immediate and delayed functional complementation kinetics correspond with their temporal activation. *Cis* variation may regulate parent-of-origin gene expression by three main mechanisms 1) variation of the seed sequence of transcription factors target sites that may cause a reduction in binding affinity; 2) variation in epigenetic levels that may prevent interaction with *trans* factors and 3) divergence in *trans* factors. Epigenetic differences among ecotypes may cause variation in parent-of-origin effects. The TCP binding motif (TBS) TGGGC, and the CACGTG G-box motif, which I found enriched in maternally biased genes, have been associated with parent-of-origin effects on circadian rhythms and growth vigor in hybrids. Heterotic phenotypes in several hybrid combinations of Arabidopsis have shown a non-reciprocal parent-of-origin effect (Miller *et al*, 2012; Ng *et al*, 2014). This phenomenon has been associated with methylation levels in the promoter region of CCA1, a transcriptional repressor involved in circadian regulatory networks (Ng *et al*, 2014). C24 x L*er* hybrids show higher biomass accumulation when the C24 ecotype was the maternal parent. Notable, this phenotype correlates with lower expression of the *CCA1* gene than in the reciprocal cross. C24 x L*er* hybrids show higher CHH methylation levels in its promoter region than in L*er* x C24 hybrids (Ng *et al*, 2014). Interestingly, the promoter regions that showed the most important methylation variation contain the TBS and G-box motifs. As a result of differential methylation in the paternal allele, the transcription factors binding to the TBS are prevented from binding and activating expression (Ng *et al*, 2014). Interestingly, this non-reciprocal methylation between the parental alleles of *CCA1* is mediated by the RdDM pathway during embryogenesis, which correlates with the

observations of non-recriprocal parental bias expression during embryo development and the involvement of RdDM in genome-wide regulation of parent-of-origin effects in early embryogenesis (Ng *et al*, 2014; Autran *et al*, 2011). The next step is to test whether differential methylation of the motifs enriched in maternally biased genes causes differences in parent-of-origin expression. How this phenomenon is established differently between maternal and paternal alleles remains to be explored. A possible explanation may be *trans* chromosomal methylation, where epigenetic differences between alleles from two ecotypes trigger changes in DNA methylation, a phenomenon similar to paramutation (Greaves *et al*, 2014).

Deletion and swap experiments directed to the enriched motifs

Another approach to elucidate the role of the identified motifs in the parent-of-origin expression consists of deletion and swap experiments directed to the motif-containing regions. I have preliminary experimental advances in this regard, with the design of reporter lines containing the different regions associated with the enriched motifs for the *NF-YB2* gene (**Figure m&m 3, Table m&m 6**). Parent-of-origin analysis of these reporter lines will help to elucidate the role of the motifs in the maternally biased expression of this gene. In addition, the NF-YB2 gene, which is maternally expressed, shows mutant phenotypes during early embryogenesis (**Supplementary Figure 9, Supplementary Table 10**). The mutant phenotype shows strong defects in suspensor development, with a short suspensor with aberrant cell divisions, especially observed in the zygote and 1-2 cell stage embryo, suggesting that this gene promotes development and growth of the suspensor. The phenotype is reminiscent of that observed in *ssp-2, yda,* and *wrky2-1* (Ueda *et al*, 2017; Bayer *et al*, 2009), genes that also promote suspensor identity. Further studies like double mutants with these alleles will help to elucidate a possible synergistic role of this transcription factor related to embryo development. Interestingly, recent reports have shown that the maternally derived transcription factors HDG11 and HDG12 regulate suspensor development in embryogenesis (Ueda *et al*, 2017).

Besides the role of the motifs associated with maternally biased genes, and regulation of parent-of-origin expression mediated by the RdDM pathway, our data suggest that other mechanisms are involved in the regulation of parent-of-origin transcription during embryogenesis. *kyp-4* and *nrpd1b-11* showed greater repression of the paternal *NF-YB2* reporter. One hypothesis is that a

repressor of the *NF-YB2* reporter is activated by methylation changes in the *kyp-4* and *nrpd1b-11* by maternal mutations. To test this hypothesis, parent-of-origin expression could be evaluated in a *kyp-4 nrpd1b-11* double mutant, to see whether they participate in the same pathway for the activation of a putative repressor, or they work in different pathways. In addition, new tools for the discovery of mechanisms regulating parental contribution should be generated. For this reason, I was working on generation of new reporter lines, to evaluate whether a similar parent-of-origin expression is observed is other genes (**Supplementary Figure 11**).

Carryover from Female gametophyte and the study of maternal genome activation

One of the limitations for studying zygotic genome activation is the difficulty of differentiating between maternal transcripts inherited from the egg and those transcribed in the zygote. To circumvent this limitation, I was working on the design of new plasmids (**Figure m&m 1 and 4, Table m&m 8**) including reporter plasmids fused to an N-terminal destruction box- containing fragment of CYCLINB1;2 (CYCB1-N) (**Table m&m 7**). Proteins fused to this so-called 'destruction box' (d-box) are degraded during anaphase. Thus, after cell division, only newly synthetized proteins are observed (Colon-Carmona *et al*, 1999; Takada & Jürgens, 2007). Since maternally biased expression may be partly explained by the inheritance of maternal products from female gametophyte, fusion of the d-box to reporters should result in the degradation of proteins inherited from the egg (or sperm).

Although the inheritance of maternal products after fertilization hasn't been tested, one starting point to test this passive transmission is the use of specific reporter lines fused to the destruction box. Preliminary observations of Egg Cell and Central Cell specific reporter lines showed that passive transmission of reporter expression occurs after fertilization (**Supplementary Figure 3 and 4, Supplementary Table 11 and 12**). Fusion of a destruction box to these reporters should allow visualization of exclusively zygotically synthetized proteins. This will serve as a test both for the use of the destruction box to observe newly synthesized proteins, and the passive transmission of maternal products to the endosperm and embryo (**Table m&m 7**). In addition, generation of reporter lines with different fluorescent proteins, with and without the destruction box for the same gene, would allow the co-visualization of maternal paternal and reporter lines in reciprocal crosses.

Turnover of proteins fused to D-box during cell division (control study)

To study the timing of degradation of the proteins fused to the d-box, one approach would be to use a constitutive reporter expressed in embryo or another tissue and follow mitotic division. One good candidate is the reporter line pUBQ14:GFP–TUA6 which functions as a mitotic and meiotic marker (Brownfield *et al*, 2015). Transformation of this reporter line with p35S promoter in the d-box plasmids would allow determining at exactly which part of the cell cycle the d-box confers lability to fluorescent reporters.

Materials and Methods Supplementary Material

Table m&m 1. Embryo defective (EMB) mutant lines used in this study.							
Locus	**Gene Symbol**	**Allele**	**Mutagen**	**ABRC Stock Number**	**Ecotype**	**Predicted Protein Function; Function Details (a)**	**Gene Identity Confirmed? (b)**
At2g38020	*VCL1*	*vcl1-1*	T-DNA	CS3941	Col	Vacuolar Protein Sorting (Vps16); Required for vacuole biogenesis and protein trafficking to the vacuole	Yes
At5g15920	*EMB2782*	*emb2782-2*	T-DNA	SALK_092081	Col	SMC Family Protein (MSS2); Chromosome dynamics (SMC5)	Yes
At3g10220	*EMB2804*	*emb2804-1*	T-DNA	SALK_019471	Col	Tubulin Folding Cofactor B; Microtubule polymerization	Yes
At1g55900	*EMB1860*	*emb1860-1*	T-DNA	CS16035	Col	Inner Mitochondrial Membrane Protein (Tim50); Import of mitochondrial matrix proteins	Yes
At1g67320	*EMB2813*	*emb2813-1*	T-DNA	SALK_045856	Col	DNA Polymerase Alpha (POLA3); DNA replication.	Yes
At1g08840	*EMB2411*	*emb2411-1*	T-DNA	CS24053	Col	Helicase / Nuclease; DNA replication-Okazaki fragment metabolism	Yes
At2g32590	*EMB2795*	*emb2795-1*	T-DNA	SALK_072400C	Col	Non-SMC Condensin Complex Subunit H; Sister chromatid segregation in mitosis	Yes
At1g10270	*GRP23*	*grp23-2*	T-DNA	SALK_128329	Col	Novel PPR Protein; Putative transcriptional regulator	Yes
At1g49880	*EMB3106*	*emb3106-2*	T-DNA	SALK_110883	Col	Mitochondrial Sulfhydryl Oxidase	Yes
At1g24340	*EMB 2421*	*emb2421*	T-DNA	CS16179	Col	Polyketide Hydroxylase Related Monooxygenase	No
At5g27740	*EMB 2775 / EMB161 / RFC3*	*emb2775-1*	T-DNA	SALK_029291C	Col	DNA Replication Factor (RFC3); PCNA loading complex factor	Yes
At1g48175	*EMB 2191*	*emb2191-1*	T-DNA	CS16170	Col	tRNA arginine adenosine deaminase; Catalytic (TAD2) subunit-Modification of wobble position	Yes
At1g74970	*TWN3*	*twn3*	T-DNA	CS24062	Col	Ribosomal Protein S9; Translation in chloroplast	No
At2g34780	*EMB 1611*	*emb1611*	T-DNA	CS16110	Col	Putative role in promoting cell division	Yes
At1g13120	*EMB1745 / AtGLE1*	*emb1745*	T-DNA	CS16088	Col	Involved in polyA mRNA export from nucleus	No
At1g55540	*EMB1011 / LNO1*	*emb1011-1*	T-DNA	CS16007	Col	Nucleoporin: mRNA export	No
a) Predicted Protein Function taken from www.seedgenes.org							
b) Gene identity confirmation from www.seedgenes.org							

Table m&m 2. Ecotypes used in this study.
All the lines were ordered from ABRC

Name	Abbreviated Name	ABRC stock
Columbia	Col-0	CS1092
Tsushima	Tsu-1	CS28782
Wassilewskija	Ws	CS915
Lansberg erecta	L*er*	CS20
Cape Verde Island	CVI-0	CS28198

Table m&m 3. Consensus matrix for the identifies motifs.

Motif	Consensus matrix
TGGGC	0.664319 0.124413 0.070423 0.140845 0.453052 0.000000 0.546948 0.000000 0.023474 0.009390 0.962441 0.004695 0.000000 0.997653 0.000000 0.002347 0.002347 0.988263 0.002347 0.007042 0.004695 0.995305 0.000000 0.000000 1.000000 0.000000 0.000000 0.000000 0.507042 0.023474 0.000000 0.469484 0.312207 0.077465 0.000000 0.610329 0.565728 0.000000 0.093897 0.340376 0.516432 0.143192 0.183099 0.157277
CACGTG	0.580786 0.192140 0.126638 0.100437 0.008734 0.934498 0.030568 0.026201 0.982533 0.000000 0.004367 0.013100 0.013100 0.956332 0.008734 0.021834 0.008734 0.004367 0.960699 0.026201 0.008734 0.026201 0.004367 0.960699 0.013100 0.004367 0.982533 0.000000 0.052402 0.000000 0.646288 0.301310 0.196507 0.668122 0.131004 0.004367 0.611354 0.091703 0.179039 0.117904 0.331878 0.113537 0.266376 0.288210

Table m&m 4. Mutant lines used in this study. All the plants employed were genotyped by PRC to confirm the insertion. All the lines were ordered from ABRC.

Mutant allele name	T-DNA associated Gene locus	Ecotype background	ABRC stock	Primers for genotyping
kyp-4	AT5G13960	Col	Salk_044606C	tttattcgagccaacatttgc agttcggttgacacatttgg
nrpd1b-11	AT2G40030	Col	CS66152 (SALK_029919)	atttcttctttgatgggggag tgtcgtggatatgaccatttg
fie	AT3G20740	Col	Salk_042962	atgtttcactgaggccatttg acaggatctcgttgtccacac
fas2-3	AT5G64630	Col	CS9929 (SALK_118525)	gcccaataatgatccacaatg tttgccctgttgcatttaaac
nf-yb2	AT5G47640	Col	SALK_061365C	atcacctttaatgggccattc ttatcagatgcttctccggtg

Table m&m 5. List of upstream regions cloned from Col wild-type Arabidopsis.

Promoter	Gene ID	Primers	Fragment
pHCS1	AT2G25710	taaactataaagtagccaaatcaatc ggtctcaaggtttgggaag	3092
pEMB1275	AT1G55350	ggcttcttctcccacctacaa cagggaattaatattaacttttgcc	1705
pCCT	AT4G00450	ttctcccaactatatccagcatac ggtgggttgataacaggaca	3766
EMB3106	AT1G49880	cttttttgtttctcggagaaatc tattttgacgttaaatctttgc	295
MINU1	AT3G06010	ggcgggaacttgtattcacc gaaatggtgggtaaaagagc	5754
EMB2768	AT3G02660	aaattttatatgctggaacctttc aatacgtcgacggtgtatttaag	2023
TTN9	AT3G20070	tttctgctcactgtagaagag cgttaatcaagaatcgaccatc	2148
ATR3	AT3G02280	cgccgccactgccaccacc tatctgatctcccttctctctctc	2638
EMB1705	AT5G22370	gcttctgaagacgaagaaga cgaagttataaggttaggtg	821
PTS	AT5G48840	gcttctgaagacgaagaaga cgaagttataaggttaggtg	395
EMB1129	AT1G49400	attttttttgctgattaacaaagt aagtaacctttcaacaataaagag	2177
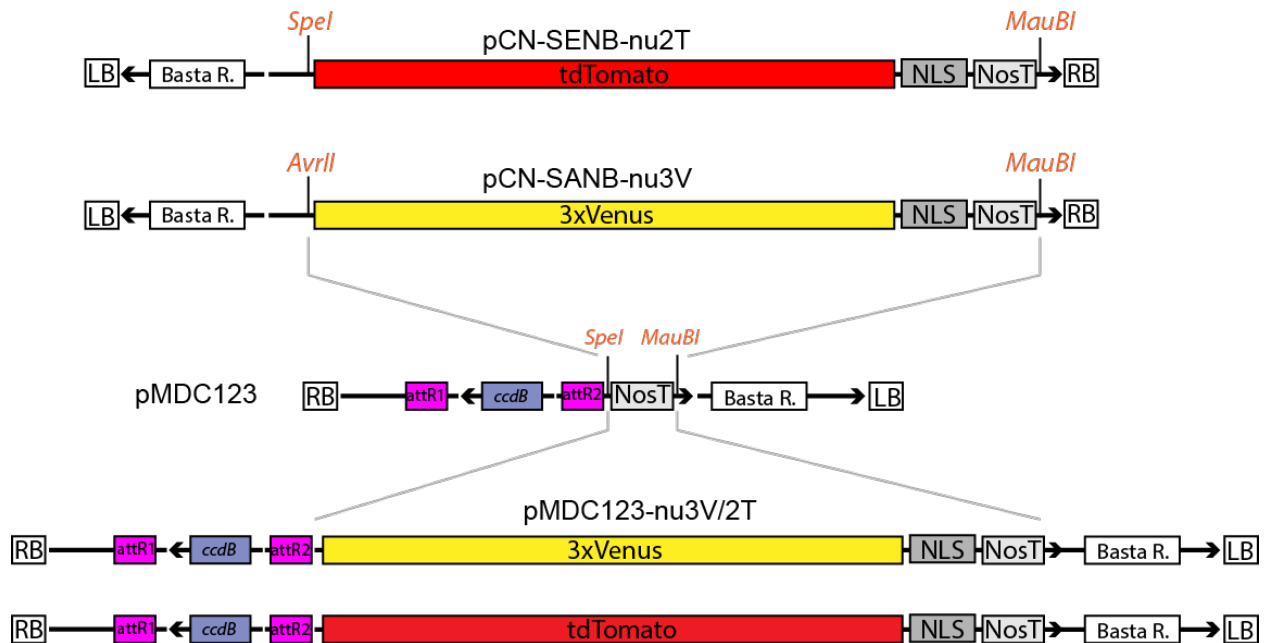
Figure m&m 1. Representation of the construction of gateway destination vectors. See Material and methods section for more details. SG lab plasmid numbers: pSG106, pSG107. Similarly, plasmids were created but using theversions with d-Box (See Figure m&m 4), pCN-SENB-dBoxn2T and pCN-SENB-dBox-n3V plasmids but they only worked in Tobacco (pSG112, pSG113). See m&m 8.

Fluorescence Brightfield Merge
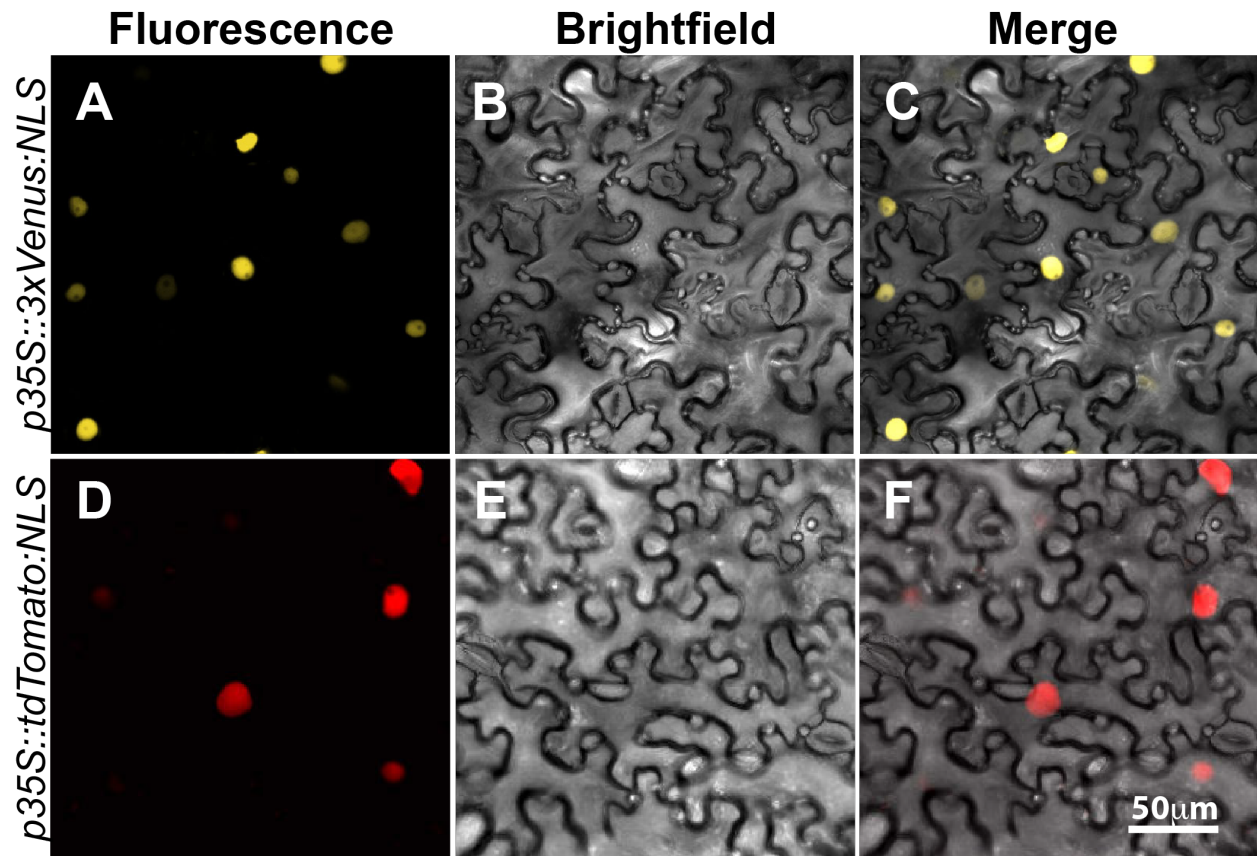
Figure m&m 2. Transient tobacco expression of the plasmids from Fiure m&m 1.

Table m&m 6. Set of primers used to clone NF-YB2 upstream regions containing regions associated to the TGGGC and CACGTG motifs. See Supplementary Figure 11.	
Name of the primer	**Sequence 5' → 3'**
SG953_pNFYB2 NcoI Rev	ccatggCCCCCATttctgaaaat
SG954_pNFYB2 XmaI Fwd-1	cccgggTTCAACATCGTGGGTT
SG955_pNFYB2 XmaI Fwd-2	cccgggCAAGTCCTCATTACAT
SG956_pNFYB2 XmaI Fwd-3	cccgggGATAATAGGTCCATCA
SG957_pNFYB2 XmaI Fwd-4	cccgggTGTTAACTAATCACAC
SG957_pNFYB2 XmaI Fwd-5	cccgggTCGTGACAAAAGTGT

71

Table m&m 7. Primers from supplemental matitial

Primer name	Target gene	Product size (bp)	Sequence
SG734_XMAI_CYCB1-N F SG742_BamHI-CYCB1-N R	CYCB1;2-N AT5G06150	564	cccgggatggcgacgagagcaaac ggatccatcaacatactccaccgca
SG928_pDD45_XmaI_R SG929_pDD45_SpeI_F	pDD45_EC1- 2(At2g21740)	1026	cccgggggttagaagccattattc actagtaaatgttcctcgctgac
SG930_pDD65_XmaI_R SG931_pDD65_SpeI_F	pDD65 (At3g10890)	1275	cccgggcaaacacttcatatcc actagtagtttagtcagcaa

Table m&m 8. Plasmids generated during this work.

Plasmid	SG lab number	Comments
pCAMBIA3301_dBox GUS+	pSG105	Not tested
pCN-SANB-dBox-nu3V-NosT	pSG111	See Figure m&m 4
pCN-SENB-dBox-nu2T-NosT	pSG110	See Figure m&m 4
pMDC123-nu3v-NosT	pSG106	Worked only in Tobacco. Weird expression in Arabidopsis. See Figure m&m 1
pMDC123-nu2T-NosT	pSG107	Worked in Arabidopsis. Bright signal. Good for transcriptional reporter lines. See Figure m&m 1
pMDC123-dBox-nu3v-NosT	pSG112	Worked only in Tobacco. Weird expression in Arabidopsis. See Figure m&m 1 and 4
pMDC123-dBox-nu2T-NosT	pSG113	Worked only in Tobacco. Weird expression in Arabidopsis. See Figure m&m 1 and 4
pMDC123-nu3v-NosT (2)	pSG117	Not tested. See Figure m&m 4
pMDC123-nu2T-NosT (2)	pSG118	Not tested. See Figure m&m 4
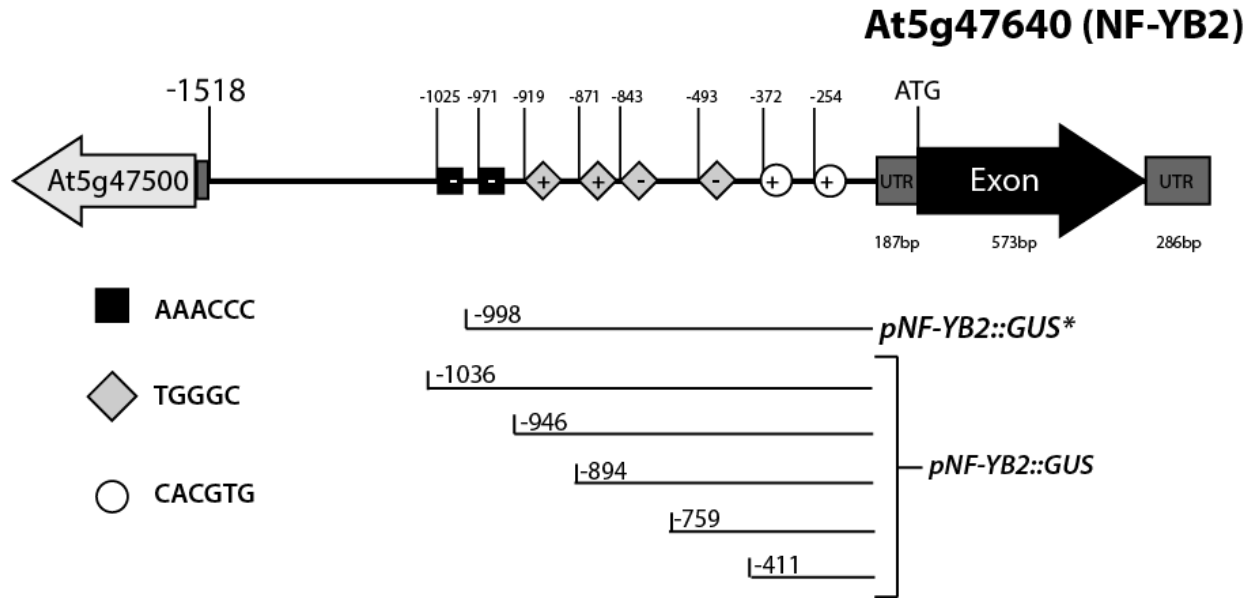
Figure m&m 3. Representation of upstream regions associated to motifs. See Table m&m6.
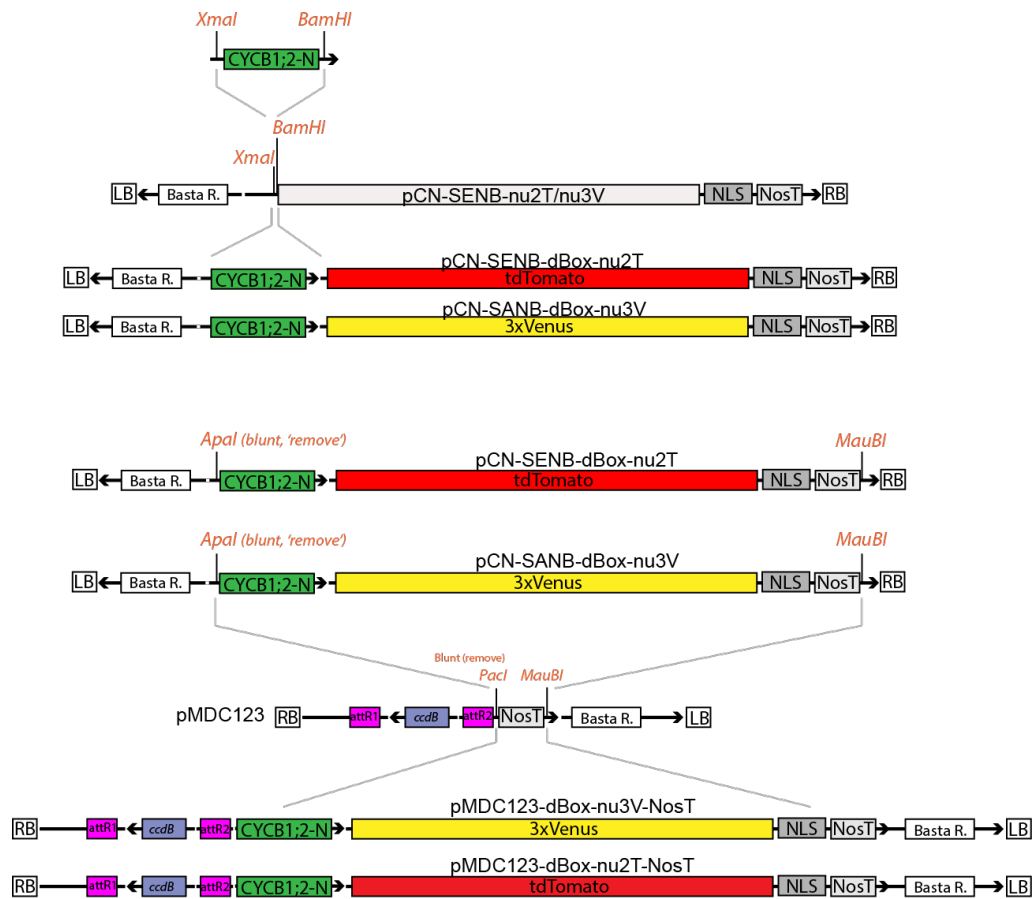
Figure m&m 4. Representation of the construction of gateway destination vectors. See Table m&m 7 and 8. CYCB1;2 N terminal (destruction box) was inserted in the pCN-SENB-nu2T/3V plasmids. Then the fragment inserted in the pMDC123 gateway destination vector. This last gateway versions were created with assistance of Jaime Alaniz. See Table m&m8.

Supplementary Material
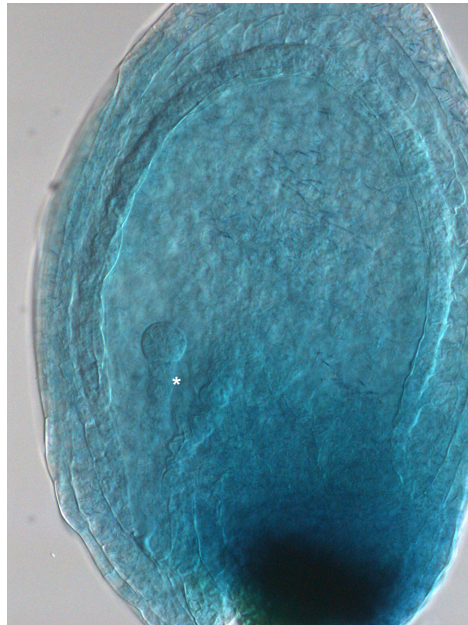
Supplementary Figure 1. Self-pollinated *gGCT-GUS* marker line. The expression of this reporter line is observer ad early as preglobular embryos (around 2dap). Similar observations were made in the *pCCT::GUS* reporter.
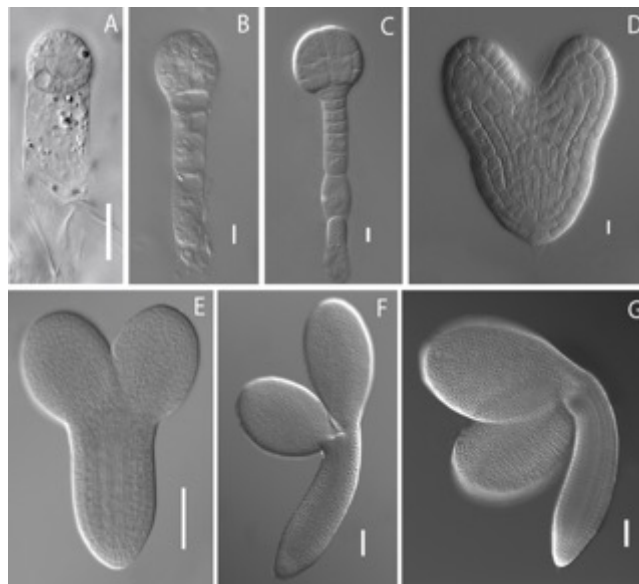
Supplementary Figure 2. Representative images of dissected embryos for ASE profiling. (A-G): zygote (A), octant (B), globular (C), heart (D), torpedo (E), bent (F) and mature embryo (G). Bar=10 µm (A, B, C and D); Bar=100 µm (E, F and G).
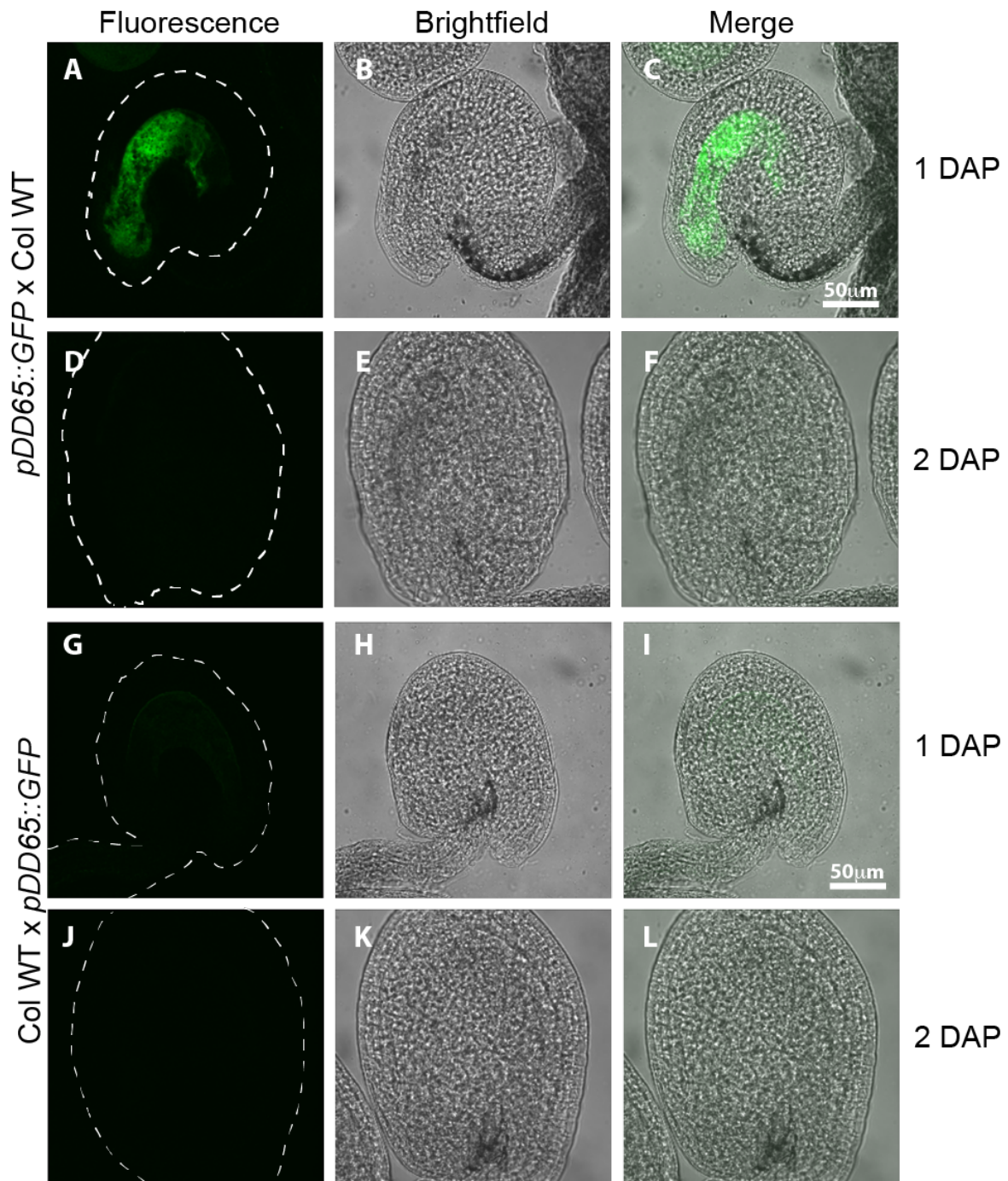
Supplementary Figure 3. Passive transmission of reporter gene GFP expression from central cell to endosperm. Specific reporter line for central cell show transient signal in endosperm. Full data in supplementary Table 12.
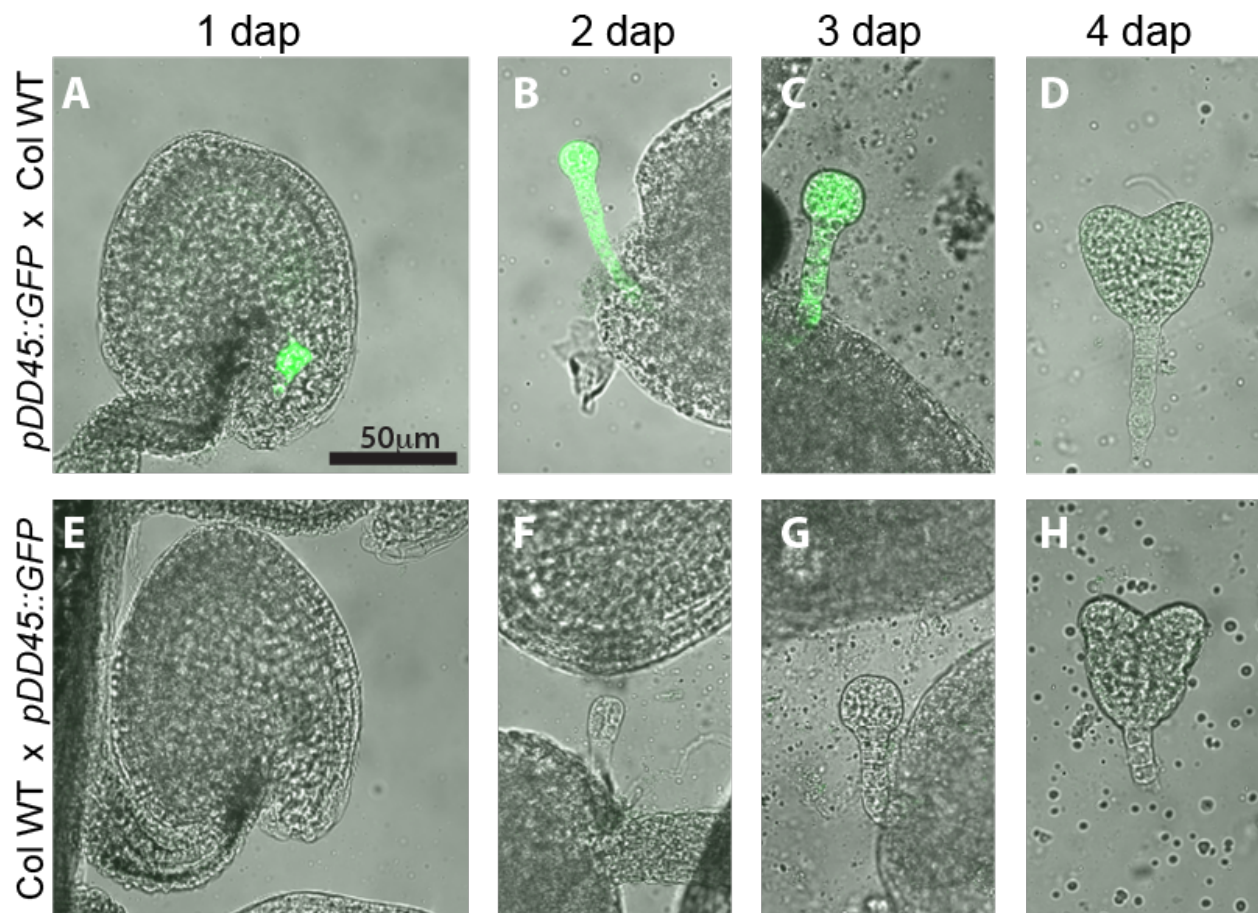
Supplementary Figure 4. Passive transmission of reporter gene GFP expression from egg cell to embryo. Specific reporter line for central cell show transient signal in endosperm. Full data in supplementary Table 11.
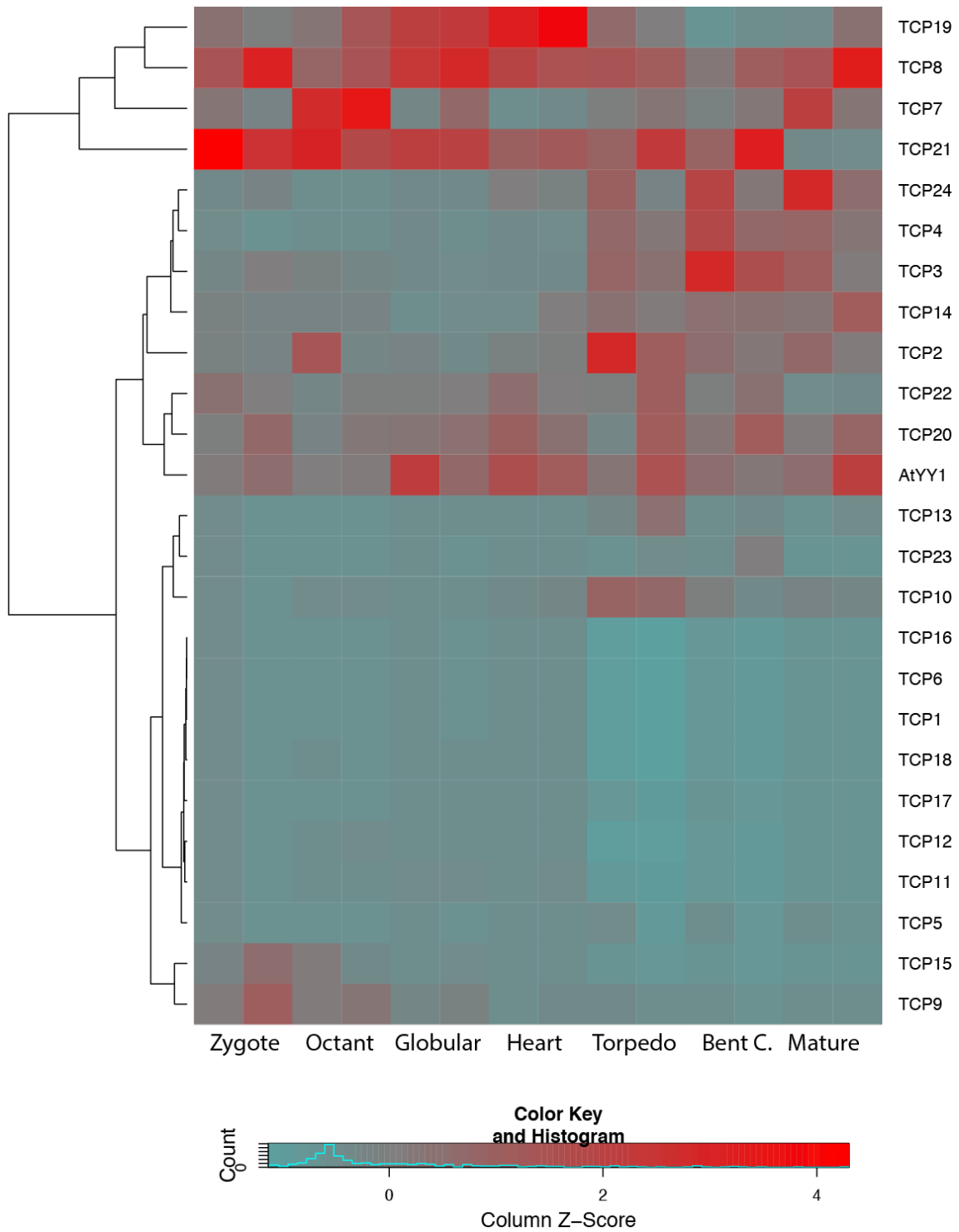
Supplementary Figure 5. Expression of TCP transcription Factors during embryogenesis.
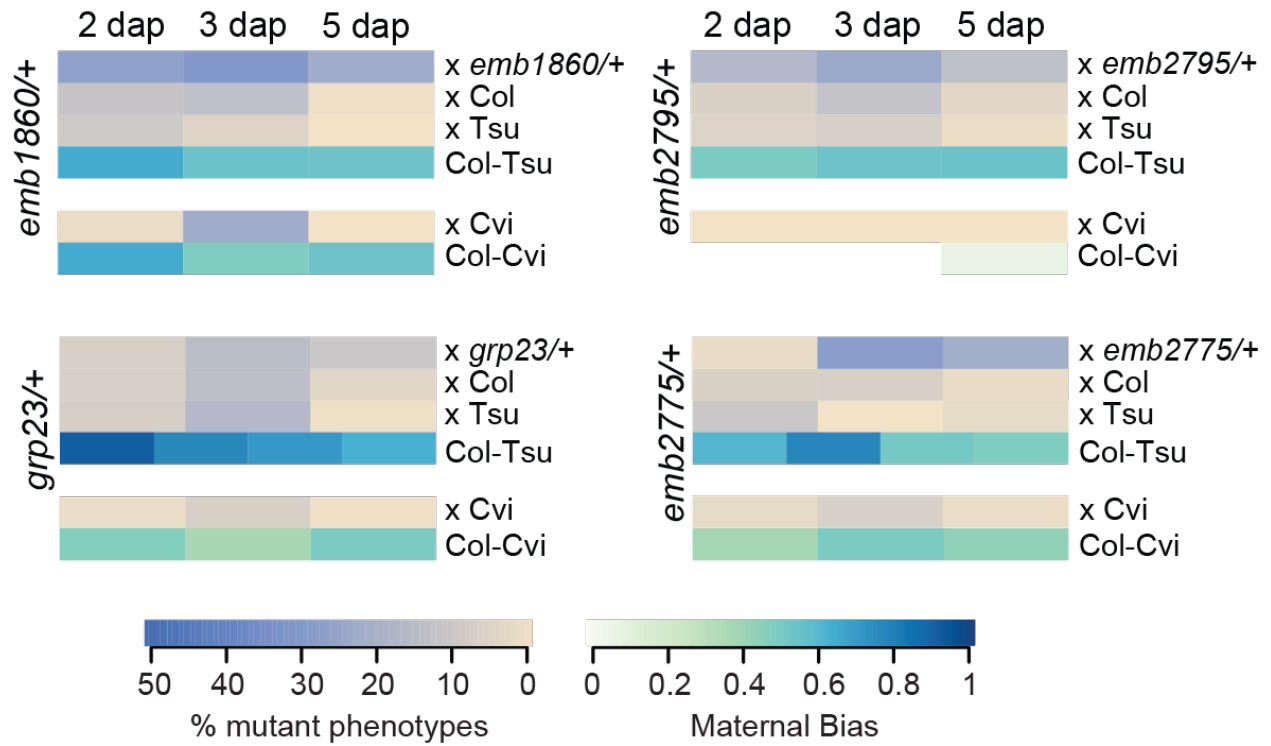
Supplementary Figure 6. Comparison of functional complementation dynamics of maternally inherited mutation by Wil-type pollen from different ecotypes with transcriptomic data from ColxTsu and ColxCVI embryos.
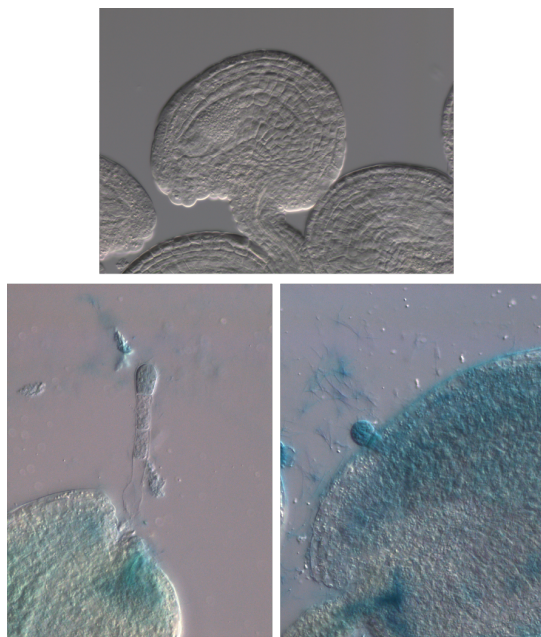
Supplementary Figure 7. Maternal expression of *pNF-YB2::GUS* on mature female gametophytes (top panel) and 1-2 celled embryos (bottom panel)

x *pNF-YB2::GUS*

	2DAP	3DAP	4DAP	5DAP
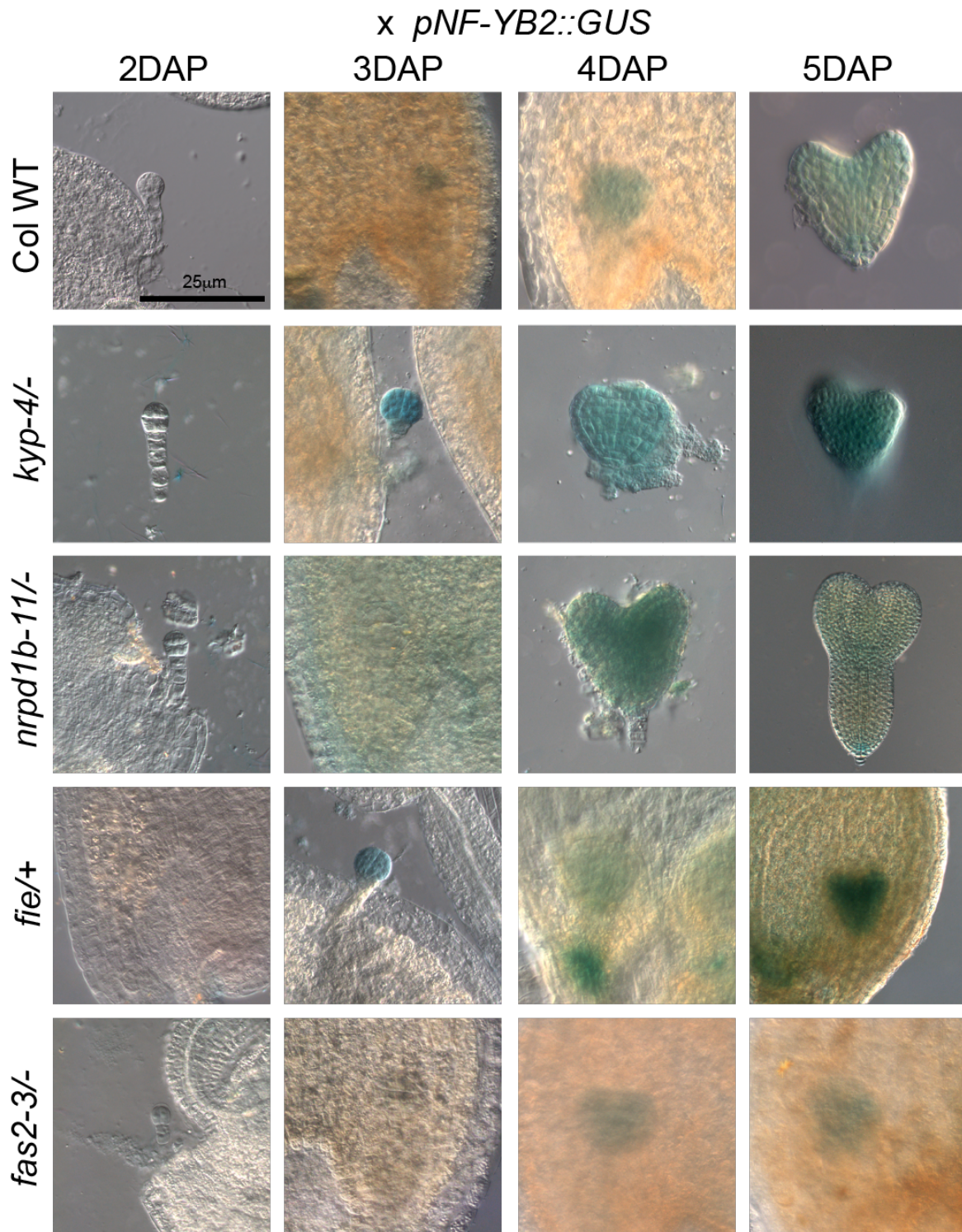
Supplementary Figure 8. pNF-YB2::GUS paternal allele activation images on maternal mutant
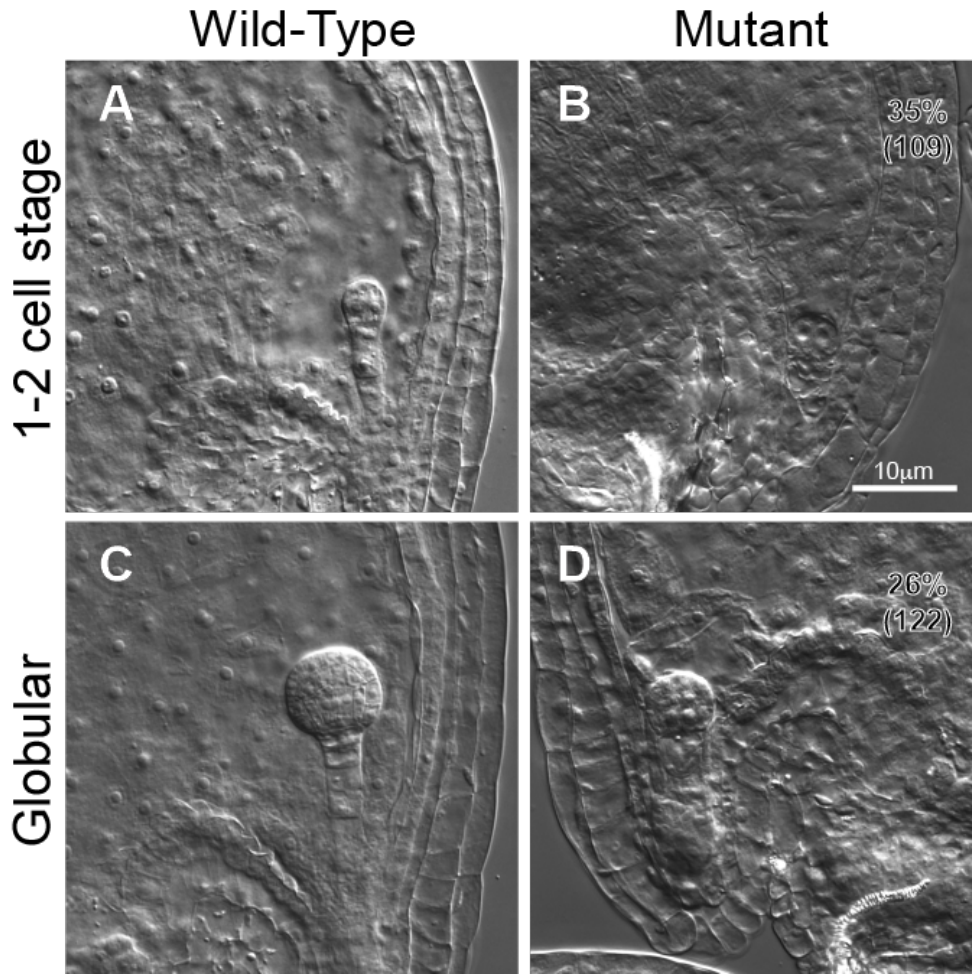
Supplementary Figure 9. Mutant phenotypes observed in the nf-yb2 T-DNA insertion line. Selfed plants were evaluated. See Supplementary Table 10.

Supplementary Figure 10. Activation of paternal *pNF-YB2::GUS* allele in endosperm.
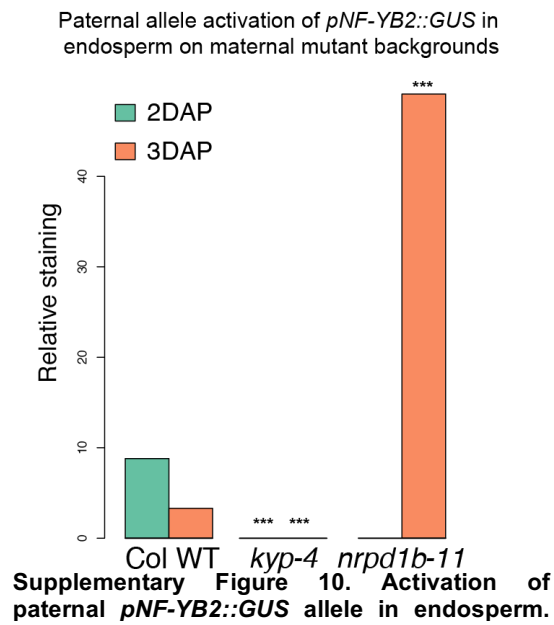
Observations of the last T1 lines
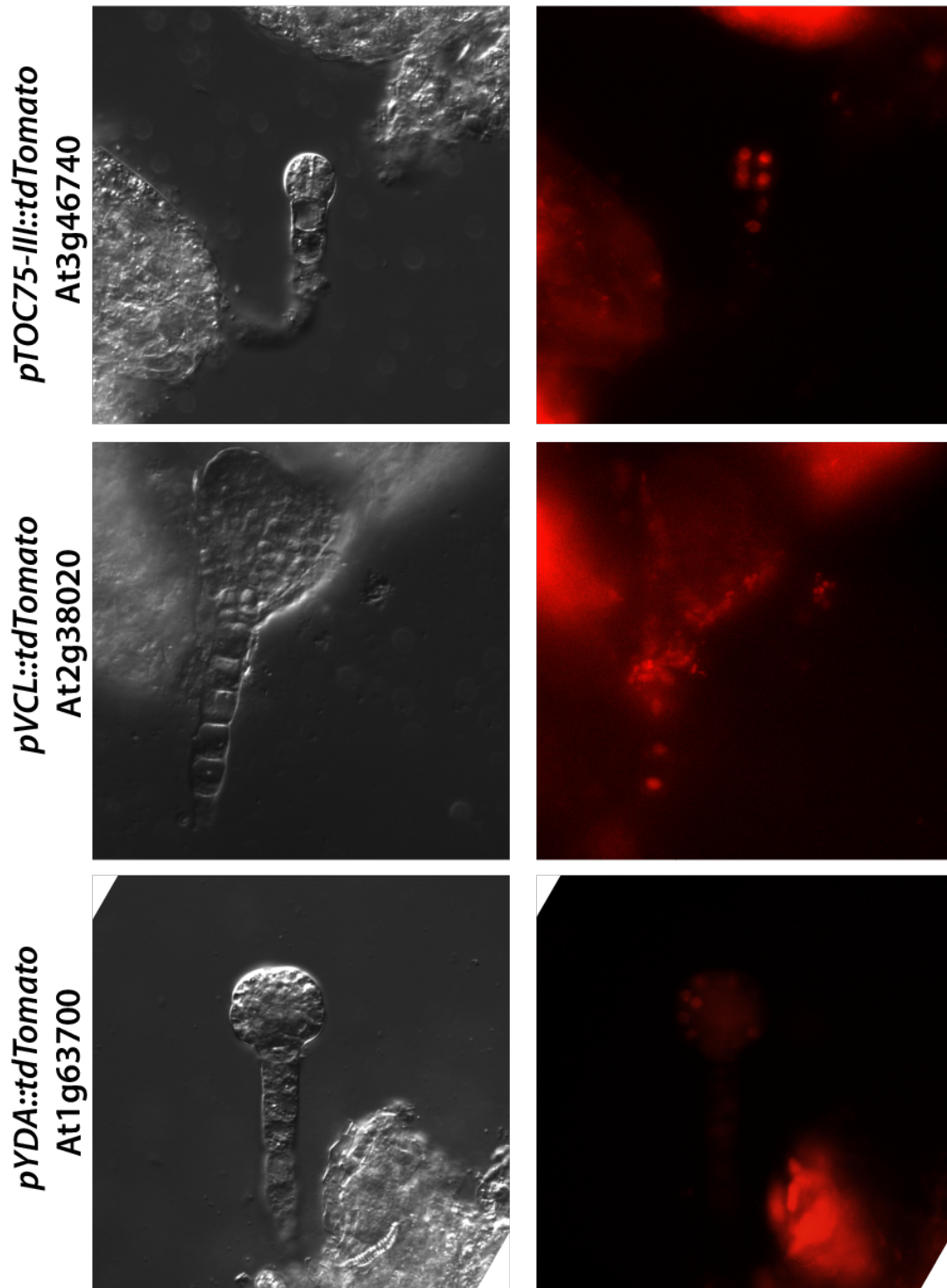
Supplementary Figure 11. New reporter lines with signal in embryos. Lines were created with assistance from Daniel Lepe.

Supplementary Table 1. Two-Tail Fisher's Exact Test to determine significance of *GCT* and *CCT* functional and molecular data from Figure 1

Mutant embryo crosses	2d		3d		4d		5d		14d	
gct-2 At1g55325	*gct/+* x *gct/+*	*gct/+* x Col wt	*gct/+* x *gct/+*	*gct/+* x Col wt	*gct/+* x *gct/+*	*gct/+* x Col wt	*gct/+* x *gct/+*	*gct/+* x Col wt	*gct/+* x *gct/+*	*gct/+* x Col wt
# of mut embryos	33	57	26	33	27	25	27	14	31	0
# of wt embryos	94	211	142	226	126	93	71	267	114	125
p value	0.3063		0.4735		0.5342		1.10E-08		1.02E-09	
significant difference?	**NO**		**NO**		**NO**		**YES**		**YES**	

	2d		3d		4d		5d		14d	
cct-1 At4g00450	*cct/+* x *cct/+*	*cct/+* x Col wt	*cct/+* x *cct/+*	*cct/+* x Col wt	*cct/+* x *cct/+*	*cct/+* x Col wt	*cct/+* x *cct/+*	*cct/+* x Col wt	*cct/+* x *cct/+*	*cct/+* x Col wt
# of mut embryos	8	4	16	9	29	26	33	8	21	0
# of wt embryos	143	170	179	191	154	184	203	209	71	78
p value	0.2374		0.1507		0.3821		0.00012		0.0000012	
significant difference?	**NO**		**NO**		**NO**		**YES**		**YES**	

GUS markers	2d		3d		4d		5d			
gGCT-GUS	gGCT-GUS x Col wt	Col wt x gGCT-GUS	gGCT-GUS x Col wt	Col wt x gGCT-GUS	gGCT-GUS x Col wt	Col wt x gGCT-GUS	gGCT-GUS x Col wt	Col wt x gGCT-GUS		
# of GUS embryos	199	0	93	165	94	109	99	72		
# of no GUS embryos	38	118	28	106	5	9	0	0		
p value	6.28E-61		0.002		0.58		1			
significant difference?	**YES**		**YES**		**NO**		**NO**			

	2d		3d		4d		5d			
pCCT::GUS	pCCT::GUS x Col wt	Col wt x pCCT::GUS	pCCT::GUS x Col wt	Col wt x pCCT::GUS	pCCT::GUS x Col wt	Col wt x pCCT::GUS	pCCT::GUS x Col wt	Col wt x pCCT::GUS		
# of GUS embryos	153	0	121	164	58	83	105	99		
# of no GUS embryos	13	133	26	129	2	1	1	0		
p value	1.60E-70		2.70E-08		0.5708		0.3812			
significant difference?	**YES**		**YES**		**NO**		**NO**			

Two Tail Fisher's Exact Test to determine if there are statistically significant differences between two samples, calculated using http://www.langsrud.com/fisher.htm. For each cross, maternal genotypes are listed first.

Supplementary Table 2. Crosses between *emb/+* plants and Col and Cvi-0, Ler, Ws and Tsu-0 ecotypes to test the effect of hybrid genotypes on phenotypic complementation, and between Col and *emb/+* plants to test the effect of haploinsuficiency on

	2 dap					3 dap					5 dap				
emb phenotypes in isogenic Col embryos. Maternal genotypes listed first. Statistically significant *p* values for Fisher's Two-Tailed Exact Test are shaded in grey. ***p≤0.001; ** p≤0.01; * p≤0.05.															
Cross	% mut	mut	norm	*p* vs Col	*p* Ler vs Cvi	% mut	mut	norm	*p* vs Col	*p* Ler vs Cvi	% mut	mut	norm	*p* vs Col	*p* Ler vs Cvi
vcl1/+ x *vcl1/+*	38.3%	108	174	0.045*		26.0%	67	191	2.22e-5***		18.2%	40	180	2.3e-6***	
vcl1/+ x Col-0	30.9%	142	318			15.8%	85	536			5.8%	24	386		
vcl1/+ x Tsu-0	38.1%	80	130	0.076		13.0%	18	120	1.0		7.8%	4	47	0.53	
vcl1/+ x Ws	48.7%	76	80	9.7e-5***		0.2%	1	58	3.7e-3***		1.36%	1	72	0.15	
vcl1/+ x Ler	3.5%	9	245	7.7e-21***	0.001**	6.8%	20	274	1.8e-3***	0.0053**	0.3%	1	337	5.6e-6***	1
vcl1/+ x Cvi-0	13.0%	26	200	9.9e-9***		1.7%	4	233	6.6e-9***		0%	0	256	1.2e-5***	
Col-0 x *vcl1/+*	0%	0	163	2.2e-22***		0%	0	161	1.8e-9***		0%	0	157	6.2e-4***	
emb2782/+ x *emb2782/+*	23.5%	65	212	0.01**		24.7%	57	174	0.005**		19.8%	44	178	7.2e-4***	
emb2782/+ x Col-0	19.0%	86	453			14.7%	37	215			9.2%	27	266		
emb2782/+ x Tsu-0	20.5%	35	136	0.19		17.5%	25	118	0.47		7.4%	9	112	0.70	
emb2782/+ x Ws	18.8%	32	138	0.40		15.9%	23	122	0.77		6.2%	8	120	0.34	
emb2782/+ x Ler	12.8%	45	260	0.69	0.0052**	12.2%	41	296	0.39	0.50	7.5%	18	221	0.53	0.07
emb2782/+ x Cvi-0	6.8%	15	204	5.8e-4***		10.2%	24	212	0.17		3.0%	4	134	0.01	
Col-0 x *emb2782/+*	0%	0	121	1.4e-8***		0%	0	192	3.5e-10***		0.7%	1	140	2.4e-4***	
emb2804/+ x *emb2804/+*	14.1%	11	67	0.19		23.9%	21	67	1.7e-5***		22.0%	37	131	2.1e-10***	
emb2804/+ x Col-0	21.0%	57	215			4.9%	8	154			3.5%	5	217		
emb2804/+ x Tsu-0	25.6%	21	61	0.36		11.2%	9	71	0.11		0%	0	51	0.58	
emb2804/+ x Ws	20.9%	22	83	1.0		2.9%	3	100	0.53		1.6%	1	59	1.0	
emb2804/+ x Ler	21.8%	17	61	0.87	1.0e-5***	3.6%	3	79	0.75	0.28	4.0%	2	47	0.61	0.19
emb2804/+ x Cvi-0	1.9%	2	105	3.9e-7***		7.6%	14	170	0.37		0.8%	1	122	0.42	
Col-0 x *emb2804/+*	0%	0	94	1.0e-8***		0%	0	113	0.022*		0%	0	126	0.16	
emb1860/+ x *emb1860/+*	26.6%	38	105	2.2e-4***		30.0%	34	79	1.4e-4***		22.4%	32	111	7.2e-13***	
emb1860/+ x Col-0	12.1%	37	267			13.2%	37	243			0.5%	1	207		
emb1860/+ x Tsu-0	9.6%	10	94	0.59		5.5%	7	120	0.024*		0%	0	160	1.0	
emb1860/+ x Ws	5.1%	5	93	0.05*		5.7%	6	100	0.044*		0%	0	106	1.0	
emb1860/+ x Ler	14.6%	21	123	0.54	0.0001***	3.8%	6	151	1.2e-3***	0.51	1.3%	2	147	0.57	0.49
emb1860/+ x Cvi-0	2.1%	3	140	2.6e-4***		2.2%	3	133	1.4e-4***		0%	0	147	1.0	
Col-0 x *emb1860/+*	0%	0	146	3.8e-7***		0%	0	171	1.1e-8***		0%	0	149	1.0	
emb2813/+ x *emb2813/+*	19.8%	25	121	0.77		16.2%	18	93	1.0		13.5%	15	96	3.0e-5***	
emb2813/+ x Col-0	15.8%	36	191			16.0%	40	210			2.0%	5	249		
emb2813/+ x Tsu-0	22%	29	101	0.15		5.6%	9	150	0.001***		8%	11	123	0.005*	
emb2813/+	8.5%	11	118	0.05*		7.3%	8	101	0.028*		2.9%	3	100	0.69	

Cross	%	n	total	p	comb p	%	n	total	p	comb p	%	n	total	p	comb p
x Ws															
emb2813/+ x Ler	5.0%	4	75	0.01**	1	0%	0	77	1.3e-5***	0.064	2.0%	0	131	0.17	1
emb2813/+ x Cvi-0	5.0%	7	133	0.001***		5.3%	5	89	0.007**		0%	0	71	0.58	
Col x emb2813	0%	0	140	1.6e-8***		0%	0	143	7.3e-9***		0%	0	139	0.16	
emb2411/+ x emb2411/+	4.8%	7	139	0.60		20.8%	22	84	0.039*		22.5%	18	80	1.1e-9***	
emb2411/+ x Col-0	3.7%	8	210			12.4%	42	297			7.2%	2	275		
emb2411/+ x Tsu-0	6.1%	11	168	0.34		7.9%	15	175	0.14		1.3%	2	154	0.62	
emb2411/+ x Ws	7.6%	14	171	0.12		5.0%	8	153	0.01**		0%	0	149	0.54	
emb2411/+ x Ler	5.1%	4	75	0.73	0.46	34.2%	40	77	5.4e-7***	0.67	2.7%	3	107	0.14	0.12
emb2411/+ x Cvi-0	2.8%	3	102	1.0		37.3%	41	69	3.4e-8***		8.2%	8	90	4.8e-4***	
Col x emb2411/+	0.9%	1	107	0.28		1.8%	3	164	2.0e-5***		1.3%	2	156	0.62	
emb2795/+ x emb2795/+	16.6%	27	136	0.004**		23.6%	63	104	1.3e-8***		13.5%	15	96	0.005**	
emb2795/+ x Col-0	6.7%	13	181			12.6%	27	187			4.2%	8	184		
emb2795/+ x Tsu-0	5.9%	8	127	0.82		7.7%	11	132	0.16		1.8%	2	112	0.33	
emb2795/+ x Ws	4.9%	7	137	0.64		2.8%	4	137	9.7e-4***		0%	0	162	0.009**	
emb2795/+ x Ler	11.4%	9	70	0.22	4.6e-4***	5.5%	4	68	0.12	0.0176*	1.9%	2	101	0.50	0.50
emb2795/+ x Cvi-0	0%	0	101	0.005**		0%	0	123	3.3e-6***		0%	0	96	0.06	
Col x emb2795/+	0%	0	112	0.003**		0%	0	65	0.001**		0%	0	180	0.007**	
grp23-2/+ x grp23-2/+	7.6%	24	292	1.0		15%	24	136	0.87		11.1%	41	328	0.009**	
grp23-2/+ x Col-0	7.7%	11	132			13.8%	22	138			4.7%	10	201		
grp23-2/+ x Tsu-0	7.8%	16	190	1.0		16.7%	32	160	0.46		1.0%	2	202	0.036*	
grp23-2/+ x Ws	9.9%	15	137	0.54		5.6%	10	167	0.01**		0.6%	1	152	0.028*	
grp23-2/+ x Ler	5.1%	4	75	0.58	0.21	0%	0	77	1.6e-4***	0.0146*	0%	0	131	0.01**	1
grp23-2/+ x Cvi-0	1.7%	3	172	0.01**		7.6%	10	122	0.13		0.7%	1	138	0.05*	
Col x grp23-2/+	0%	0	124	0.001***		1.6%	3	186	9.9e-6***		0%	0	172	0.002**	
emb3106/+ x emb3106/+	19.2%	28	118	2.1e-8***		20.2%	19	75	5.1e-4***		20.6%	14	54	1.5e-11***	
emb3106/+ x Col-0	2.8%	8	278			6.7%	21	293			0%	0	302		
emb3106/+ x Tsu-0	4.0%	7	166	0.58		1.9%	3	158	0.025*		1.2%	2	164	0.12	
emb3106/+ x Ws	3.0%	4	128	1.0		4.1%	6	139	0.39		1.5%	2	133	0.09	
emb3106/+ x Ler	2.9%	4	134	1.0	0.49	2.1%	2	92	0.12	0.50	0%	0	105	1.0	1
emb3106/+ x Cvi-0	5.5%	5	86	0.31		0%	0	87	0.01**		0%	0	105	1.0	
Col x emb3106/+	0%	0	165	0.029*		1.1%	2	186	0.003**		0%	0	177	1.0	
emb2421/+ x emb2421/+	13.3%	14	91	0.36		24.8%	37	112	4.8e-12***		19.5%	33	136	1.0e-13***	
emb2421/+ x Col-0	9.9%	32	291			3.3%	11	320			0.7%	2	294		
emb2421/+	7.6%	11	133	0.49		0.7%	1	146	0.11		2.5%	3	115	0.14	

	%	n	N	p	comb. p	%	n	N	p	comb. p	%	n	N	p	comb. p
x Tsu-0															
emb2421/+ x Ws	1.8%	3	163	6.3e-4***		1.8%	3	160	0.56		1.1%	2	174	0.63	
emb2421/+ x Ler	2.6%	3	110	0.01	0.25	5.6%	8	135	0.30	0.29	0.7%	1	138	1.0	0.45
emb2421/+ x Cvi-0	6.1%	10	154	0.17		3.2%	6	183	1.0		0%	0	167	0.53	
Col x *emb2421/+*	0.5%	1	186	6.0e-6***		0%	0	153	0.02*		3.7%	4	105	0.047*	
emb2775/+ x *emb2775/+*	2.7%	4	145	0.097		27.6%	52	136	2.7e-11***		21.9%	48	171	1.3e-16***	
emb2775/+ x Col-0	6.7%	29	406			7.5%	40	492			2.7%	15	546		
emb2775/+ x Tsu-0	11.2%	18	143	0.085		0%	0	127	2.5e-4***		3.3%	5	148	0.78	
emb2775/+ x Ws	3.8%	6	157	0.23		2.4%	3	126	0.028*		1.8%	3	165	0.77	
emb2775/+ x Ler	1.1%	3	278	2.7e-4***	0.14	5.5%	21	382	0.18	0.41	0.8%	3	364	0.05*	0.4425
emb2775/+ x Cvi-0	3.2%	9	273	0.059		6.9%	19	257	0.77		1.7%	4	235	0.45	
Col-0 x *emb2775/+*	0%	0	143	2.7e-4***		0%	0	131	1.4e-4***		0%	0	155	0.05*	
[#]Col-0 wt x Ler wt	3.9%	5	121			1%	1	100			0%	0	115		
[#]Col-0 wt x Cvi-0 wt	0%	0	91			1.1%	1	92			2.7%	2	71		
[#]Col-0 wt x Ws wt	2.7%	3	106			0%	0	142			0.85%	1	116		
[#]Col-0 wt x Tsu-1 wt	2.3%	3	123			2.5%	4	151			0%	0	159		

[#]For control crosses, embryos with abnormal phenotypes were scored. These phenotypes were distinct from the emb phenotypes.
Col-0 ABRC stock CS1092; Ler ABRC stock CS20; Cvi-0 ABRC stock CS28198; Ws ABRC stock CS915; Tsu-1 ABRC stock CS28782
For each time point, the combined number of embryos examined represents at least two biological replicates.

Supplementary Table 3. Putative transcription factors binding the TCCCG or AAACCC motifs.
⬧ Genes with high identity to the TCP from *Lotus japonicas,* the best match predicted to bind the TGGGC motif. Based in a protein blast analysis.
* Genes with the highest expression in early embryogenesis based in the analysis of normalized expression from Raju's Data.

Transcription Factor type	Gene Locus	Gene Name	Parental Expression	Motif in upstream sequence
TCP Class I subfamily	AT5G41030	*TCP6*	-	TGGGC
	AT5G23280	*TCP7**	Maternal Specific	AAACCC, TGGGC
	AT1G58100	*TCP8**	Maternal Bias	AAACCC, TGGGC
	AT2G45680	*TCP9**⬧	Maternal Specific	TGGGC
	AT2G37000	*TCP11*	-	TGGGC
	AT3G47620	*TCP14*	Maternal Specific	AAACCC
	AT1G69690	*TCP15**	Maternal Specific	AAACCC, TGGGC
	AT3G45150	*TCP16*	-	AAACCC
	AT5G51910	*TCP19**⬧	-	AAACCC, TGGGC
	AT3G27010	*TCP20**	-	AT3G27010
	AT5G08330	*TCP21**	-	AAACCC
	AT1G72010	*TCP22**	Maternal Bias	-
	AT1G35560	*TCP23*	Maternal Specific	AAACCC
TCP Class II subfamily	AT1G67260	*TCP1*	-	AAACCC, TGGGC
	AT4G18390	*TCP2*	-	AAACCC
	AT1G53230	*TCP3*	Maternal Specific	AAACCC
	AT3G15030	*TCP4*	-	AAACCC
	AT5G60970	*TCP5*	-	AAACCC
	AT1G68800	*TCP12*	-	-
	AT3G02150	*TCP13*	-	AAACCC
	AT2G31070	*TCP10*	Maternal Specific	AAACCC
	AT5G08070	*TCP17*	-	AAACCC
	AT3G18550	*TCP18*	-	AAACCC
	AT1G30210	*TCP24*	-	AAACCC
C2H2-TYpe zinc finger Transcription factor	AT4T06634	*YY1**	Maternal Bias	AAACCC

Supplementary Table 4. Paternal activation in **embryos** of *pNF-YB2::GUS* in reciprocal crosses with Wil-Type plants. Statistically significant *p* values for Fisher's Two-Tailed Exact Test ***p≤0.001; ** p≤0.01; * p≤0.05.

			pNF-YB2::GUS x Col WT	Col WT x *pNF-YB2::GUS*	
1DAP	*GUS* expression		49	0	
	No *GUS* expression		19	134	
	Total		68	134	
	p-value		**1.277e-31*** (spanning)**		
2DAP	*GUS* expression	0.53	41	0	1
	No *GUS* expression		12	129	
	Total		53	129	
	p-value		**2.559e-30*** (spanning)**		
3DAP	*GUS* expression	0.09	81	0	1
	No *GUS* expression		10	137	
	Total		91	137	
	p-value		**4.36e-51*** (spanning)**		
4DAP	*GUS* expression	0.19	66	105	**1.3e-49***
	No *GUS* expression		15	27	
	Total		81	132	
	p-value		0.85 (spanning)		
5DAP	*GUS* expression	0.4	106	97	0.12
	No *GUS* expression		34	40	
	Total		140	137	
	p-value		0.41 (spanning)		

Supplementary Table 5. Paternal activation in **embryos** of *pNF-YB2::GUS* reporter expression in maternal mutant backgrounds of regulators of parental genome activation. Statistically significant *p* values for Fisher's Two-Tailed Exact Test between wild-type paternal activation compared with mutant backgrounds.
***p≤0.001; ** p≤0.01; * p≤0.05.

		Col WT x pNF-YB2::GUS	kyp-4/kyp-4 x pNF-YB2::GUS	nrpd1b11/nrpd1b-11 x pNF-YB2::GUS	fie/FIE x pNF-YB2::GUS	fas2-3/fas2-3 x pNF-YB2::GUS
2DAP	*GUS* expression	14	1	0	1	3
	No *GUS* expression	102	155	123	94	54
	Total	116	156	123	95	57
	p-value	-	**3.8e-5***	**2.6e-5***	**0.0019**	0.18
3DAP	*GUS* expression	84	83	58	49	3
	No *GUS* expression	29	42	67	45	61
	Total	113	125	125	94	64
	p-value	-	0.2	**1.2e-5***	**0.001***	**2.98e-21***
4DAP	*GUS* expression	137	168	105	99	47
	No *GUS* expression	7	16	22	16	7
	Total	144	184	127	115	54
	p-value	-	0.19	**0.001***	0.014	0.06
5DAP	*GUS* expression	149	105	160	92	65
	No *GUS* expression	7	8	9	3	3
	Total	156	113	169	95	68
	p-value	-	0.42	0.8	0.74	1

Supplementary Table 6. Paternal activation in **endosperm** of *pNF-YB2::GUS* reporter expression in maternal mutant backgrounds of regulators of parental genome activation. Statistically significant *p* values for Fisher's Two-Tailed Exact Test between wild-type paternal activation compared with mutant backgrounds. ***p≤0.001; ** p≤0.01; * p≤0.05.

		Col WT x pNF-YB2::GUS	*kyp-4/kyp-4* x *pNF-YB2::GUS*	*nrpd1b11/nrpd1b-11* x *pNF-YB2::GUS*
2DAP	*GUS* expression	9	0	0
	No *GUS* expression	93	156	123
	Total	102	156	123
	p-value		**0.0001**	**0.0006**
3DAP	*GUS* expression	4	0	55
	No *GUS* expression	117	125	57
	Total	121	125	112
	p-value		0.05	**4.022e-17***

Supplementary Table 7. Paternal activation in **embryos** of *pNF-YB2::GUS* reporter expression in paternal mutant backgrounds of regulators of parental genome activation. Statistically significant *p* values for Fisher's Two-Tailed Exact Test between wild-type paternal activation compared with mutant backgrounds. ***p≤0.001; ** p≤0.01; * p≤0.05.

		Col WT x pNF-YB2::GUS	Col WT x *pNF-YB2::GUS; kyp-4/kyp-4*	Col WT x *pNF-YB2::GUS; nrpd1b11/nrpd1b-11*
2DAP	*GUS* expression	24	28	39
	No *GUS* expression	78	93	104
	Total	102	121	143
	p-value	-	1	0.55
3DAP	*GUS* expression	38	86	94
	No *GUS* expression	83	96	86
	Total	122	182	180
	p-value		**0.006***	**0.0003***
4DAP	*GUS* expression	112	146	92
	No *GUS* expression	9	33	25
	Total	121	179	117
	p-value	-	**0.006***	**0.002***

			pEMB2768::tdTomato x Col WT	Col WT x *pEMB2768::tdTomato*	
2DAP	*tdTomato* expression		25	34	
	No *tdTomato* expression		47	84	
	Total		72	118	
	p-value		0.421		
3DAP	*tdTomato* expression		54	61	
	No *tdTomato* expression		51	61	
	Total		105	122	
	p-value		0.89		
4DAP	*tdTomato* expression		37	48	
	No *tdTomato* expression		31	47	
	Total		68	95	
	p-value		0.63		

Supplementary Table 8. Paternal activation in **embryos** of *pEMB2768::tdTomato* reporter expression in reciprocal crosses with Col wild-type plants. One T3 heterozygous plant was used. Statistically significant *p* values for Fisher's Two-Tailed Exact Test are shaded in grey. ***p<0.001; ** p<0.01; * p<0.05.

Supplementary Table 9. Paternal activation in **endosperm** of *pEMB2768::tdTomato* reporter expression in reciprocal crosses with Col wild-type plants. One T3 heterozygous plant was used. Statistically significant *p* values for Fisher's Two-Tailed Exact Test are shaded in grey.
***p$\leq$0.001; ** p$\leq$0.01; * p$\leq$0.05.

		pEMB2768::tdTomato x Col WT	Col WT x *pEMB2768::tdTomato*
2DAP	*tdTomato* expression	72	1
	No *tdTomato* expression	0	117
	Total	72	118
	p-value	**2.1631e-52***	
3DAP	*tdTomato* expression	105	7
	No *tdTomato* expression	0	115
	Total	105	122
	p-value	**5.98e-57****	
4DAP	*tdTomato* expression	68	46
	No *tdTomato* expression	0	49
	Total	68	95
	p-value	**4.22e-15***	

Supplementary Table 10. Number of mutant and wild-type embryos quantified in a selfed nf-yb2 homozygous plant.

	Wild-type	Mutant	Total
Preglobular	70	39	109
Globular	90	32	122

		pDD45::GFP x Col WT	Col WT x pDD45::GFP
1DAP	*GFP* expression	122	0
	No *GFP* expression	62	175
	Total	184	175
	% of expression	66.3%	0%
2DAP	*GFP* expression	198	0
	No *GFP* expression	13	264
	Total	211	264
	% of expression	93.8%	0%
3DAP	*GFP* expression	163	0
	No *GFP* expression	15	170
	Total	178	170
	% of expression	91.5%	0%
4DAP	*GFP* expression	116	0
	No *GFP* expression	16	98
	Total	132	98
	% of expression	87.8%	0

Supplementary Table 11. Expression of *pDD45::GFP* Egg Cell specific reporter line in embryo suggesting passive transmission of GFP after fertilization. Although the expression is observed until 4 dap, the intensity is considerably lower (see Supplementary Figure 4).

pDD45::GFP from Steffen et al., 2007

Supplementary Table 12. Expression of *pDD65::GFP* Central Cell specific reporter line in endosperm suggesting passive transmission of GFP after fertilization. The transmission signal is only much lower on intensisty and abundance at 2 dap. (see Supplementary Figure 3).			
		pDD65::GFP x Col WT	Col WT x *pDD65::GFP*
1DAP	*GFP* expression	75	0
	No *GFP* expression	59	177
	Total	134	177
	% of expression	55.9%	0%
2DAP	*GFP* expression	45	0
	No *GFP* expression	73	114
	Total	118	114
	% of expression	38%	0%
pDD65::GFP from Steffen et al., 2007			

Supplementary Table 13. Parent-of-Origin expression of embryo region-specific expressed genes according with Belmonte, et al., 2012. Embryo-enriched genes with different fold-changes also analyzed from Belmonte, et al., 2012.

Preglobular embryo region-specific
161 genes in this list. From those, 147 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 48 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	13	5	0	0
Non-reciprocally Maternal Bias	21	19	18	9
Riciprocally Paternal Bias	3	0	0	0
Non-reciprocally Paternal Bias	8	5	3	3
Allele-dominant	16	23	17	17
Biallelic	8	38	52	57
TOTAL	69	90	90	86

Preglobular-globular embryo region-specific
380 genes in this list. From those, 357 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 94 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	348	158	5	1
Non-reciprocally Maternal Bias	195	384	162	89
Riciprocally Paternal Bias	5	0	0	1
Non-reciprocally Paternal Bias	21	14	47	71
Allele-dominant	98	131	170	167
Biallelic	27	167	499	572
TOTAL	694	854	883	901

Embryo-enriched FC2
1663 genes in this list. From those, 1636 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 849 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	348	158	5	1
Non-reciprocally Maternal Bias	195	384	162	89
Riciprocally Paternal Bias	5	0	0	1
Non-reciprocally Paternal Bias	21	14	47	71
Allele-dominant	98	131	170	167
Biallelic	27	167	499	572
TOTAL	694	854	883	901

Embryo-enriched FC3
680 genes in this list. From those, 672 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 322 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	111	36	2	0
Non-reciprocally Maternal Bias	87	153	61	37
Riciprocally Paternal Bias	3	0	0	1
Non-reciprocally Paternal Bias	12	6	12	23
Allele-dominant	49	70	75	70

| **Biallelic** | 13 | 82 | 209 | 237 |
| TOTAL | 275 | 347 | 359 | 368 |

Embryo-enriched FC4
338 genes in this list. From those, 335 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 132 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	48	19	1	0
Non-reciprocally Maternal Bias	46	70	34	22
Riciprocally Paternal Bias	1	0	0	0
Non-reciprocally Paternal Bias	8	5	6	15
Allele-dominant	28	40	37	36
Biallelic	10	45	106	120
TOTAL	141	179	184	193

Embryo-enriched FC5
196 genes in this list. From those, 196 were also profiled in egg cell microarrays (Wuest et al., 2010) from which 65 are expressed in egg cell.

	Zygote	Octant	Globular	Heart
Riciprocally Maternal Bias	25	10	1	0
Non-reciprocally Maternal Bias	26	38	19	10
Riciprocally Paternal Bias	1	0	0	0
Non-reciprocally Paternal Bias	6	4	4	6
Allele-dominant	17	21	20	22
Biallelic	10	33	68	80
TOTAL	85	106	112	118

Bibliography

Alonso-Peral MM, Trigueros M, Sherman B, Ying H, Taylor JM, James Peacock W & Dennis ES (2016) Patterns of gene expression in developing embryos of Arabidopsis hybrids. *Plant J.* Available at: http://doi.wiley.com/10.1111/tpj.13432 [Accessed November 25, 2016]

Autran D, Baroux C, Raissig MTTT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UCCC, Leblanc O, Vielle-Calzada J-P, Rosenstiel P, Grimanelli D & Grossniklaus U (2011) Maternal epigenetic pathways control parental contributions to arabidopsis early embryogenesis. *Cell* **145:** 707–719 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867411004338

Baroux C, Autran D, Gillmor CS, Grimanelli D & Grossniklaus U (2009) The maternal to zygotic transition in animals and plants. *Cold Spring Harb Symp Quant Biol* **73:** 89–100 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19204068

Baroux C, Autran D, Raissig MT, Grimanelli D & Grossniklaus U (2013) Parental contributions to the transcriptome of early plant embryos. *Curr. Opin. Genet. Dev.* **23:** 72–4 Available at: http://www.sciencedirect.com/science/article/pii/S0959437X13000166 [Accessed June 16, 2015]

Baroux C, Blanvillain R & Gallois P (2001) Paternally inherited transgenes are down-regulated but retain low activity during early embryogenesis in Arabidopsis. *FEBS Lett.* **509:** 11–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11734197 [Accessed June 17, 2012]

Baroux C & Grossniklaus U (2015) The Maternal-to-Zygotic Transition in Flowering Plants: Evidence, Mechanisms, and Plasticity. *Curr Top Dev Biol* **113:** 351–371 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0070215315000332

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T & Lukowitz W (2009) Paternal control of embryonic patterning in Arabidopsis thaliana. *Science (80-.).* **323:** 1485–1488 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19286558

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB & Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the

Arabidopsis seed. *Proc. Natl. Acad. Sci.*: 1222061110- Available at:
http://www.pnas.org/content/early/2013/01/11/1222061110 [Accessed January 15, 2013]

Boros J, Arnoult N, Stroobant V, Collet J-F & Decottignies A (2014) Polycomb Repressive
Complex 2 and H3K27me3 Cooperate with H3K9 Methylation To Maintain
Heterochromatin Protein 1 at Chromatin. *Mol. Cell. Biol.* **34:** 3662–3674 Available at:
http://mcb.asm.org/cgi/doi/10.1128/MCB.00205-14

Brownfield L, Yi J, Jiang H, Minina EA, Twell D & Köhler C (2015) Organelles maintain
spindle position in plant meiosis. *Nat. Commun.* **6:** 6492 Available at:
http://www.nature.com/ncomms/2015/150311/ncomms7492/abs/ncomms7492.html
[Accessed March 31, 2016]

Clough SJ & Bent AF (1998) Floral dip: a simplified method forAgrobacterium-mediated
transformation ofArabidopsis thaliana. *Plant J.* **16:** 735–743 Available at:
http://doi.wiley.com/10.1046/j.1365-313x.1998.00343.x [Accessed July 11, 2017]

Colon-Carmona A, You R, Haimovitch-Gal T & Doerner P (1999) Spatio-temporal analysis of
mitotic activity with a labile cyclin-GUS fusion protein. *Plant J.* **20:** 503–508 Available at:
http://doi.wiley.com/10.1046/j.1365-313x.1999.00620.x [Accessed February 11, 2013]

Costa LM, Marshall E, Tesfaye M, Silverstein K a. T, Mori M, Umetsu Y, Otterbach SL,
Papareddy R, Dickinson HG, Boutiller K, VandenBosch K a., Ohki S & Gutierrez-Marcos
JF (2014) Central Cell-Derived Peptides Regulate Early Embryo Patterning in Flowering
Plants. *Science (80-.).* **344:** 168–172 Available at:
http://www.sciencemag.org/content/344/6180/168.full [Accessed April 11, 2014]

Cowles CR, Hirschhorn JN, Altshuler D & Lander ES (2002) Detection of regulatory variation in
mouse genes. *Nat. Genet.* **32:** 432–7 Available at: http://dx.doi.org/10.1038/ng992
[Accessed August 11, 2015]

Dresselhaus T, Sprunck S & Wessel GM (2016) Fertilization Mechanisms in Flowering Plants.
Curr. Biol. **26:** R125–R139 Available at:
http://linkinghub.elsevier.com/retrieve/pii/S0960982215015584 [Accessed February 7,
2017]

Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, Vashisht AA, Gallego-Bartolome J,
Wohlschlegel JA, Patel DJ & Jacobsen SE (2014) Mechanism of DNA methylation-directed
histone methylation by KRYPTONITE. *Mol. Cell* **55:** 495–504 Available at:

http://www.sciencedirect.com/science/article/pii/S1097276514004912 [Accessed
September 5, 2015]

Du J, Johnson LM, Jacobsen SE & Patel DJ (2015) DNA methylation pathways and their
crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16:** 519–532 Available at:
http://www.nature.com/doifinder/10.1038/nrm4043 [Accessed August 21, 2015]

Du J, Zhong X, Bernatavichute YV V, Stroud H, Feng S, Caro E, Vashisht AAA, Terragni J,
Chin HGG, Tu A, Hetzel J, Wohlschlegel JAA, Pradhan S, Patel DJJ & Jacobsen SEE
(2012) Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes
Directs DNA Methylation in Plants. *Cell* **151:** 167–80 Available at:
http://linkinghub.elsevier.com/retrieve/pii/S009286741201015X [Accessed October 4,
2012]

Falcon S & Gentleman R (2007) Using GOstats to test gene lists for GO term association.
Bioinformatics **23:** 257–258

Figueiredo DD & Köhler C (2016) Bridging the generation gap: communication between
maternal sporophyte, female gametophyte and fertilization products. *Curr. Opin. Plant Biol.*
29: 16–20

García-Aguilar M & Gillmor CS (2015) Zygotic genome activation and imprinting: parent-of-
origin gene regulation in plant embryogenesis. *Curr. Opin. Plant Biol.* **27:** 29–35 Available
at: http://www.sciencedirect.com/science/article/pii/S1369526615000710 [Accessed June
17, 2015]

Gehring M (2013) Genomic imprinting: insights from plants. *Annu. Rev. Genet.* **47:** 187–208
Available at: http://www.annualreviews.org/doi/10.1146/annurev-genet-110711-155527
[Accessed July 14, 2015]

Greaves IK, Groszmann M, Wang A, Peacock WJ & Dennis ES (2014) Inheritance of Trans
Chromosomal Methylation patterns from Arabidopsis F1 hybrids. *Proc. Natl. Acad. Sci. U.
S. A.* **111:** 2017–22 Available at:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3918825&tool=pmcentrez&ren
dertype=abstract

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ & Dennis ES (2012) Trans
Chromosomal Methylation in Arabidopsis hybrids. *Proc. Natl. Acad. Sci.* **109:** 3570–3575
Available at: http://www.pnas.org/content/109/9/3570.full?sid=0aeb0961-26bf-4315-8d7f-

757b536d9da5 [Accessed November 1, 2012]

Grimanelli D, Perotti E, Ramirez J & Leblanc O (2005) Timing of the maternal-to-zygotic transition during early seed development in maize. *Plant Cell* **17:** 1061–1072 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation &list_uids=15749756

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ & Dennis ES (2011) Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. U. S. A.* **108:** 2617–2622 Available at: http://www.pnas.org/content/108/6/2617.full?sid=0aeb0961-26bf-4315-8d7f-757b536d9da5 [Accessed November 1, 2012]

Hatorangan MR, Laenen B, Steige KA, Slotte T & Köhler C (2016) Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. *Plant Cell* **28:** 1815–27 Available at: http://www.plantcell.org/lookup/doi/10.1105/tpc.16.00304 [Accessed July 29, 2016]

He S, Sun Y, Yang Q, Zhang X, Huang Q, Zhao P, Sun M, Liu J, Qian W, Qin G, Gu H & Qu L-J (2017) A Novel Imprinted Gene NUWA Controls Mitochondrial Function in Early Seed Development in Arabidopsis. *PLOS Genet.* **13:** e1006553 Available at: http://dx.plos.org/10.1371/journal.pgen.1006553

Hennig L, Bouveret R & Gruissem W (2005) MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. *Trends Cell Biol.* **15:** 295–302 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15953547 [Accessed July 11, 2017]

Hennig L & Derkacheva M (2009) Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* **25:** 414–23 Available at: http://www.cell.com/trends/genetics/fulltext/S0168-9525(09)00149-8 [Accessed March 20, 2012]

Hou X, Zhou J, Liu C, Liu L, Shen L & Yu H (2014) Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. *Nat. Commun.* **5:** 4601 Available at: http://www.nature.com/ncomms/2014/140808/ncomms5601/full/ncomms5601.html [Accessed September 26, 2015]

Jackson JP, Lindroth AM, Cao X & Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416:** 556–60 Available at:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation &list_uids=11898023 [Accessed March 8, 2012]

Jahnke S & Scholten S (2009) Epigenetic Resetting of a Gene Imprinted in Plant Embryos. *Curr. Biol.* **19:** 1677–1681 Available at: http://dx.doi.org/10.1016/j.cub.2009.08.053 [Accessed March 15, 2012]

Jenik PD, Gillmor CS & Lukowitz W (2007) Embryonic patterning in Arabidopsis thaliana. *Annu Rev Cell Dev Biol* **23:** 207–236 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation &list_uids=17539754

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J & Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17:** 379–84 Available at: http://www.cell.com/current-biology/fulltext/S0960-9822(07)00810-X [Accessed July 20, 2012]

Johnson LM, Law JA, Khattar A, Henderson IR & Jacobsen SE (2008) SRA-Domain Proteins Required for DRM2-Mediated De Novo DNA Methylation. *PLoS Genet.* **4:** e1000280 Available at: http://dx.plos.org/10.1371/journal.pgen.1000280 [Accessed June 11, 2017]

Kaya H, Shibahara K, Taoka K, Iwabuchi M, Stillman B & Araki T (2001) FASCIATA Genes for Chromatin Assembly Factor-1 in Arabidopsis Maintain the Cellular Organization of Apical Meristems. *Cell* **104:** 131–142 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867401001970 [Accessed July 11, 2017]

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ & Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498:** 385–389 Available at: http://www.nature.com/doifinder/10.1038/nature12178 [Accessed May 25, 2013]

Lee J-Y, Colinas J, Wang JY, Mace D, Ohler U & Benfey PN (2006) Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 6055–60 Available at: http://www.pnas.org/content/103/15/6055.full [Accessed February 20, 2014]

Luo A, Zhao P, Zhang L-Y, Sun M-X, Yeung EC, Meinke DW, Goldberg RB, Paiva G de, Yadegari R, Friml J, Jeong S, Bayer M, Lukowitz W, Bozhkov P V., Filonova LH, Suarez MF, Kawashima T, Goldberg RB, Tadros W, Lipshitz HD, et al (2016) Initiation of

programmed cell death in the suspensor is predominantly regulated maternally in a tobacco hybrid. *Sci. Rep.* **6:** 29467 Available at: http://www.nature.com/articles/srep29467 [Accessed July 23, 2016]

Ma J, Wang Q, Sun R, Xie F, Jones DC & Zhang B (2014) Genome-wide identification and expression analysis of TCP transcription factors in Gossypium raimondii. *Sci. Rep.* **4:** 6645 Available at: http://www.nature.com/srep/2014/141016/srep06645/full/srep06645.html [Accessed February 25, 2015]

Matzke MA & Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15:** 394–408 Available at: http://dx.doi.org/10.1038/nrg3683 [Accessed July 10, 2014]

Miller M, Zhang C & Chen ZJ (2012) Ploidy and Hybridity Effects on Growth Vigor and Gene Expression in Arabidopsis thaliana Hybrids and Their Parents. *G3: Genes|Genomes|Genetics* **2:** 505–513 Available at: http://www.g3journal.org/content/2/4/505.full [Accessed July 14, 2015]

Ng DW-K, Miller M, Yu HH, Huang T-Y, Kim E-D, Lu J, Xie Q, McClung CR & Chen ZJ (2014) A Role for CHH Methylation in the Parent-of-Origin Effect on Altered Circadian Rhythms and Biomass Heterosis in Arabidopsis Intraspecific Hybrids. *Plant Cell* **26:** 2430–2440 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24894042 [Accessed July 14, 2015]

Ngo QA, Baroux C, Guthörl D, Mozerov P, Collinge MA, Sundaresan V & Grossniklaus U (2012) The Armadillo repeat gene ZAK IXIK promotes Arabidopsis early embryo and endosperm development through a distinctive gametophytic maternal effect. *Plant Cell* **24:** 4026–43 Available at: http://www.plantcell.org/content/24/10/4026 [Accessed November 18, 2013]

Nodine MD & Bartel DP (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482:** 94–97 Available at: http://dx.doi.org/10.1038/nature10756 [Accessed January 23, 2012]

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D & Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* **132:** 603–14 Available at: http://dev.biologists.org/content/132/3/603.long [Accessed December 17,

2014]

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW & Gehring M (2014) Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife* **3:** e03198 Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4115658&tool=pmcentrez&rendertype=abstract [Accessed May 26, 2015]

Raissig MT, Bemer M, Baroux C & Grossniklaus U (2013) Genomic Imprinting in the Arabidopsis Embryo Is Partly Regulated by PRC2. *PLoS Genet.* **9:** 21–23 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3854695&tool=pmcentrez&rendertype=abstract [Accessed February 25, 2014]

Schon MA & Nodine M (2017) Widespread contamination of Arabidopsis embryo and endosperm transcriptome datasets. *Plant Cell***:** tpc.00845.2016 Available at: http://www.plantcell.org/lookup/doi/10.1105/tpc.16.00845

Siefers N, Dang KK, Kumimoto RW, Bynum WE, Tayrose G & Holt BF (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol.* **149:** 625–41 Available at: http://www.plantphysiol.org/content/149/2/625.full [Accessed October 24, 2015]

Singer-Sam J, Chapman V, LeBon JM & Riggs AD (1992) Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation. *Proc. Natl. Acad. Sci. U. S. A.* **89:** 10469–73 Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=50360&tool=pmcentrez&rendertype=abstract [Accessed August 11, 2015]

Springer NM & Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* **19:** 2391–2402 Available at: http://www.plantcell.org/content/19/8/2391 [Accessed July 14, 2015]

Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U & Dresselhaus T (2012) Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338:** 1093–7 Available at: http://www.sciencemag.org/content/338/6110/1093.abstract [Accessed September 23, 2015]

Steffen JG, Kang I-H, Macfarlane J & Drews GN (2007) Identification of genes expressed in the

Arabidopsis female gametophyte. *Plant J.* **51:** 281–92 Available at:
http://www.ncbi.nlm.nih.gov/pubmed/17559508 [Accessed April 1, 2016]

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ & Jacobsen SE (2014) Non-CG
methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.*
21: 64–72 Available at: http://dx.doi.org/10.1038/nsmb.2735 [Accessed September 5, 2015]

Stroud H, Greenberg MVCC, Feng S, Bernatavichute Y V. & Jacobsen SE (2013)
Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the
Arabidopsis Methylome. *Cell* **152:** 352–64 Available at:
http://www.cell.com/article/S0092867412014304/fulltext [Accessed January 11, 2013]

Swanson-Wagner R a, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D & Schnable PS
(2009) Paternal dominance of trans-eQTL influences gene expression patterns in maize
hybrids. *Science* **326:** 1118–1120 Available at:
http://www.sciencemag.org/content/326/5956/1118.full [Accessed October 1, 2013]

Tadros W & Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts.
Development **136:** 3033–3042 Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation
&list_uids=19700615

Takada S & Jürgens G (2007) Transcriptional regulation of epidermal cell fate in the Arabidopsis
embryo. *Development* **134:** 1141–50 Available at:
http://dev.biologists.org/content/134/6/1141.long [Accessed February 1, 2013]

Tessadori F, van Zanten M, Pavlova P, Clifton R, Pontvianne F, Snoek LB, Millenaar FF,
Schulkes RK, van Driel R, Voesenek LACJ, Spillane C, Pikaard CS, Fransz P & Peeters
AJM (2009) Phytochrome B and histone deacetylase 6 control light-induced chromatin
compaction in Arabidopsis thaliana. *PLoS Genet.* **5:** e1000638 Available at:
http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000638 [Accessed
August 26, 2015]

Toledo-Ortiz G, Huq E & Quail PH (2003) The Arabidopsis Basic/Helix-Loop-Helix
Transcription Factor Family. *Plant Cell Online* **15:** Available at:
http://www.plantcell.org/content/15/8/1749.long [Accessed July 3, 2017]

Del Toro-De León G, García-Aguilar M & Gillmor CS (2014) Non-equivalent contributions of
maternal and paternal genomes to early plant embryogenesis. *Nature* **514:** 624–627

Available at: http://dx.doi.org/10.1038/nature13620 [Accessed September 8, 2014]

Del Toro-De León G, Lepe-Soltero D & Gillmor CS (2016) Zygotic genome activation in isogenic and hybrid plant embryos. *Curr. Opin. Plant Biol.* **29:** 148–153 Available at: http://www.sciencedirect.com/science/article/pii/S1369526615300091 [Accessed January 28, 2016]

Tucker SL, Reece J, Ream TS & Pikaard CS (2010) Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. *Cold Spring Harb. Symp. Quant. Biol.* **75:** 285–97 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21447813 [Accessed July 10, 2017]

Ueda M, Aichinger E, Gong W, Groot E, Verstraeten I, Vu LD, De Smet I, Higashiyama T, Umeda M & Laux T (2017) Transcriptional integration of paternal and maternal factors in the *Arabidopsis* zygote. *Genes Dev.* **31:** 617–627 Available at: http://genesdev.cshlp.org/lookup/doi/10.1101/gad.292409.116 [Accessed April 12, 2017]

Vielle-Calzada JP, Baskar R & Grossniklaus U (2000) Delayed activation of the paternal genome during seed development. *Nature* **404:** 91–94 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10716449

Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M & Springer NM (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 19639–44 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3845156&tool=pmcentrez&rendertype=abstract [Accessed July 14, 2015]

Weijers D, Geldner N, Offringa R & Jurgens G (2001) Seed development: Early paternal gene activity in Arabidopsis. *Nature* **414:** 709–710 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11742384

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano J-C, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, et al (2014) Determination

and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158:** 1431–1443 Available at: http://www.cell.com/article/S0092867414010368/fulltext [Accessed September 11, 2014]

Wierzbicki AT, Haag JR, Pikaard CS, Green PJ, Jacobsen SE, Zilberman D, Jacobsen SE, Carrington JC, Duttagupta R, Ganesh M & al. et (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135:** 635–48 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19013275 [Accessed July 10, 2017]

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C & Grossniklaus U (2010) Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* **20:** 506–12 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20226671 [Accessed September 3, 2013]

Yan H, Yuan W, Velculescu VE, Vogelstein B & Kinzler KW (2002) Allelic variation in human gene expression. *Science* **297:** 1143 Available at: http://www.sciencemag.org/content/297/5584/1143.short#xref-ref-3-1 [Accessed July 28, 2015]

Yartseva V & Giraldez AJ (2015) The Maternal-to-Zygotic Transition During Vertebrate Development : A Model for Reprogramming 1st ed. Elsevier Inc. Available at: http://dx.doi.org/10.1016/bs.ctdb.2015.07.020

Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L & Grotewold E (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Res.* **39:** D1118–D1122 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21059685 [Accessed January 25, 2017]

Zhao H, Wu D, Kong F, Lin K, Zhang H & Li G (2017) The Arabidopsis thaliana Nuclear Factor Y Transcription Factors. *Front. Plant Sci.* **7:** 1–11 Available at: http://journal.frontiersin.org/article/10.3389/fpls.2016.02045/full