,. frontiers
in Plant Science

ORIGINAL RESEARCH
pLbished: 25 Februaiy 2020
doi: 10.3389tfpls.2020.00132

Developmental Analysis of *Mimulus* Seed Transcriptomes Reveals Functional Gene Expression Clusters and Four Imprinted, Endosperm-Expressed Genes

Miguel A. Flores-Vergarai[1], Elen Oneat[2]', Mario Costa[3], Gonzalo Villarino[4], Caitlyn Roberts[5], Maria Angels De Luis Balaguer [1], Silvia Coimbra [3], John Willis [2] and Robert G. Franks[1]*

[1] Department of Plant ard Microbial Biology, North Carolna State Urwersity, Raleigh, NC, United States, [2] Department of Biology, Dti<e University, Durham, NC, U nited States, [3] GreenUPorto, Sustainable Agrifood Prodv:;tion Research Centre, Biology Department, Faculty of Sciences, University of Porto, Porto, Portugal, [4] Biob gy Department, San Diego State University. San Oego, CA, UnitedStates, [5] Department of Biology,BereaCollege, Berea, KY, United States

The double fertilization of the female gametophyte initiates embryogenesis and endosp erm development in seeds *via* the activation of genes involved in cell differentia t ion, organ patterning, and growth. A subset of genes expressed in endosperm exhibit imprinted expression, and the correct balance of gene expression between parental alleles is critical for proper endosperm and seed development. Weusea transcriptionaltime series analysis to identify genes that are associated with key shifts in seed development, including genes associated with secondary cell wall synthesis, mitotic cell cycle, chromatin organization, auxin synthesis, fatty acid metabolism, and seed maturation. We relate these genes to morphological changes in *Mimulus* seeds. We also identify four endosperm-expressed transcripts that display imprinted (paternal) expression bias. The imprinted status of these four genes is conserved in other flowering plants, suggestin g that they are func tionally important in endosperm development. Our study explores gene regulatory dynamics in a species with *ab initio* cellular endosperm development, broadening the taxonomic focus of the literature on gene expression in seeds. Moreover, it is the first to validate genes with imprinted endosperm expression in *Mimu/us guttatus,* and will inform future studies on the genetic causes of seed fail,ure in this model system.

Keywords: seed development, ribonucleic acid sequencing, developmental time course analysis, endospenn, genomic imprinting, K-means clustering, MADS-box genes, *Mimu/us guttatus*

OPEN ACCESS

Edited by:
Cristi1a Ferrard iz, Polytectmic Universityof VaJenda, Spain

Reviewed by:
Raffael a BattagKa, Council for Agrk;ultural and Ecoromics Research, Italy
Dongfang Wa . Spelman College, United States

·Correspondence:
Ben Oneal
e/en.oneaJ@duke.edu
tThese authors have contributed equally to tns work

Specialty section:
This artide was submitted to Plant Oeletopment and EvoDevo, a section of thepu ma/ Frontiers in Plant Science

Received: *23 September 2019*
Accepted: *28 Jamary 2020*
Publishect *25 February 2020*

Citation:
Flores-Vergara MA, Olea/E,Costa M, Vilarino G, Roberts C, De Luis BaJaguerMA, Coimbra S, Wills J and Franks FIG (2020) Developmental Analysis of Miml.ius Seed Transcriptomes Reveals Functional Gene Expression Ouste rs and Foo-Imprinted, Erd osperm-Expressed Genes. Front Plant Sci. 11:132. doi: 10.3389'fpls.2020.00132

INTRODUCTION

Upon their emergence in the Early Cretaceous, seed-bearingplants diversified rapidly, displacing older plant lineages and colonizing nearly every terrestrial habitat {Lidgard and Crane, 1988; Crane and Lidgard, 1989; Magallon and Castillo, 2009). Among the many factors enabling their swift rise to dominance was the emergence of seeds, a major evolutionary and reproductive innovation that frees vascular plants from a dependence on water for gametophytic dispersal and enables the next sporophytic generation to delay germination until conditions are favorable for growth and reproduction. Angiosperm seeds are formed by a unique process called double fertilization, wherein two haploid sperm nuclei contained within a pollen grain act separately to fuse with the haploid egg cell and the homo-diploid central cell of the female megagametophyte to form a diploid embryo and triploid endosperm

Following double fertilization, angiosperm seeds undergo processes of cell differentiation, patterning, and growth. Early embryogenesis establishes the basic shoot -root body plan, after which the embryonic tissue and major organs of the embryo are formed by morphogenesis (West and Harada, 1993; Jurgens et al., 1994). Endosperm development in modern plants can be characterized by three major types: *ab initio* cellular, where each nuclear division isaccompanied by cell division; nuclear, where a syncytial phase offree nuclear division isfollowed bycellular wall formation; and helobial, where an initial division of the primary endosperm cell results in two regions, at least one of which will exhibit free nuclear development (Friedman, 1994; Friedman, 2001). Cellular endosperm development is found in several basal angiosperm lineages (Friedman, 2001) and many diverse groups of asterids, and has likely evolved multiple times independently (Geeta, 2003). As the seed matures the endosperm will accumulate storage reserves for nutritional support of the mature embryo. Once fully formed, the mature embryo will enter a period of developmental arrest in preparation for dormancy.

Coordinated development between seed tissues is critical to ensuring normal development and plays a major role in determining the size of mature seeds (Garcia et al., 2003; Ingouff et al., 2006; Sechet et al., 2018), and recent improvements in transcriptomics have greatly improved our knowledge of the gene expression dynamics involved (Belmonte et al., 2013). In *Arabidopsis,* many genes are seed specific, including transcription factors (TFs) that regulate gene networks involved in cell differentiation and nutrient storage (Le et al., 2010; Chen et al., 2014; Yi et al., 2019). A subset of genes expressed in endosperm exhibit imprinted expression, an epigenetic phenomenon whereby alleles are differentially expressed depending upon their parent-of-origin (Grossniklaus et al., 1998; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2013; Florez-Rueda et al., 2016; Zhang et al., 2016; Lafon-Placette et al., 2018).

To date, most studies on gene expression in seeds have focused on a taxonomically narrow group of species with nuclear endosperm, such as *Arabidopsis* and crops from the family Poaceae. There exists a lack of studies illustrating the transcriptional dynamics of seeds with cellular endosperm development despite the prevalence of this developmental phenotype. We fill this gap by characterizing the gene expression dynamics associated with seed development in *Mimulus,* an asterid which exhibits *ab initio* cellular endosperm development. We perform a time series RNA sequencing experiment to illustrate the major transcriptional events associated with early seed development. Our work has two major goals: first, by focusing on a phylogenetically divergent species with cellular endosperm development it will serve as a data resource enabling comparative studies of gene expression in seed plants. Second, because hybrid seed inviability is a common outcome of hybridization and may be a major cause of speciation in plants, including *Mimulus,* characterizing the transcriptional dynamics of normally developing seeds will provide context for future studies on gene expression changes associated with hybrid seed lethality. Imprinted loci may play key roles in the proper development of endosperm (Grossniklaus et al., 1998) and divergence in the imprinting status of genes has been linked to the rapid emergence of reproductive barriers between even closely related species (Florez-Ruedaet al., 2016; Lafon-Placette et al., 2018) including *Mimulus* (Kinser et al., 2018). Thus, although our study design prohibits a systematic screening for imprinted loci, we examine our data for genes exhibiting paternally imprinted expression in endosperm, as such genes have been implicated in the emergence of hybrid seed inviability. We identify genes whose expression exhibits significant temporal changes associated with key developmental shifts and characterize their patterns of co-expression and biological functions using K- means clustering and gene ontology enrichment analyses. We also identify and validate four genes *(ATRXS, MBD13, DnaJ, BGALl1)* that are paternally imprinted and whose homologues are imprinted in other plants, but which are not linked to hybrid endosperm failure in *Mimulus* {Kinser et al., 2018) or other taxa (Hatorangan et al., 2016; Florez-Rueda et al, 2016; Lafon-Placette et al., 2018). Our study represents an important first step in illustrating the gene regulatory dynamics of *Mimulus* seeds, is the first to identify and validate genes with imprinted endosperm expression in *Mimulus guttatus,* and will inform future studies on the genetic causes of seed failure in this model system.

MATERIALS AND METHODS

We examined hybrid seed from a compatible cross between two members of the M. *guttatus* complex: a serpentine-adapted annual M. *guttatus* and *Mimulus pardalis,* a facultatively selfing annual that is fully interfertile with the outcrossing M. *guttat us.* M. *pardalis* seed was collected in 2005 near the Star-Excelsior Mine in Copperopolis, CA (- 120.85SW, 38.153N). M. *guttatus* seed was collected in 2008 from the Donald and Sylvia McLaughlin Natural Reserve in Lake County, CA (- 122.41SW, 38.861N). Inbred lines from each population were formed by a minimum of five generations of selfing and are named SEC39 (M. *pardalis)* and CSS4 (M. *guttatus),* leading to an expectation of a maximum of 3% genomic-wide heterozygosity in each line. All

individuals used here for morphological analyses as well as DNA and RNA sequencing are from these inbred lines. We chose these two populations because they are interfertile (based on our prior unpublished work and Macnair and Cumbes (1989)) but sufficiently genetically distinct to allow us to distinguish the allelic origin of genie SNPs, and thus search for genes that potentially exhibit *M. guttatus-bias.* All seeds were cold-stratified for 10 days at 4°C before being transferred to a growth chamber at 18-h days, 21°C, and 30% relative humidity. Crosses and self-pollinations were performed as described previously (Oneal et al., 2016).

Seed Histology

We characterized the morphology and developmental ontogeny of selfed and hybrid seeds of *M. pardalis* and *M. guttatus* in order to provide a developmental framework for our time series analysis of gene expression. All fruits were collected from flowers that had been emasculated 1- 3 days prior to fertilization. We used LR-white embedding to visualize and categorize the morphological progression of seed development. Ovaries and fruits were harvested at 0, 2, 4, 6, and 8 days after pollination (DAP) and vacuum-fixed in a solution of 2% paraformaldehyde, 2.5% glutaraldehyde, and 0.001% Tween 20 in 0.025 M PBS (pH 7), then incubated overnight at 4°C. Tissue was washed with 0.025 M PBS for 20 min, then dehydrated in an ethanol series (25%, 35%, 50%, 70%, 80%, 90%, and 3x 100%). LR-white (London Resin Company Ltd, London, UK) impregnation was performed by incubation for 12 h in increasing concentration of resin (10%, 25%, 50%, 75%, 90%, and 2x 100%). Samples were transferred to gelatin capsules and allowed to polymerize for 48 h in a 56°C oven. The 2-micron sections were obtained with a Leica EM UC7 Ultramicrotome, placed on glass slides, and stained with 1% toluidine blue followed by 1% safranin.

Preparation of Libraries and Messenger RNA Sequencing

For the purpose of transcriptomic analysis, total RNA was extracted from ovules and developing hybrid seeds from the SEC39 x CSS4 cross (throughout this paper, the maternal parent is listed first in cross notations). We only prepared a single direction of this cross for transcriptornics analysis due to a significant level of pre-zygotic infertility in the CSS4 x SEC39 crossing direction, which limited the RNA available for

gently shaking into microcentrifuge tubes containing 300 µl of ice-coldisolation buffer [lX First-Strandbuffer Invitrogen, 1 mM dithiothreitol (DTT), and 4% RNaseOUT]. TRizol was then added (3:1 ratio of TRizol:isolation buffer). We used the Pure Link RNA Mini Kit (Invitrogen), modifying it as follows: collected ovules and seeds in solution were ground gently with pellet pestles in Eppendorf tubes for 2-4 min then mixed with 240 µI of chloroform. The solution was shaken vigorously by hand for 15 s, then incubated for 3 min at room temperature followed by centrifugation at 12,000g for 15 min at 4°C to separate phases. Total RNA was isolated according to the manufacturer's instructions, then treated with TURBO DNA-free Dnase I Kit (Ambion) for removal of contaminated DNA. The mixed solution was then spun at 10,000g for 1.5 min to collect the top layer of RNA. We note that this method of hand separation of ovules/seeds from the longitudinal-sectionedfruits may have inadvertently resulted in a small amount of placental-contributed tissue into our samples.

We generated and sequenced a total of 20 RNA sequencing libraries (see **Table 1** for details). We used four maternal plants (SEC39) and five pollen donors (CSS4) to generate hybrid seeds aged 2, 4, 6, and 8 days after pollination (DAP). Because of the difficulty of isolating sufficient RNA for sequencing, ovules were derived from seven maternal plants. Hybrid seed pools generated by one maternal plant constituted one biological replicate. Because our parental plants were highly inbred, allelic differences between replicates should be minimal. This expectation is confirmed by the genomic sequence data (see below), which indicates that in CSS4, 0.115% of covered exonic sites are heterozygous, while in SEC39, 0.04% covered exonic sites are heterozygous. The RNA integrity number (RIN) for all samples > 7.8, as determined by an Agilent 2100 Bioanalyzer. Twenty strand-specific complementary DNA (cDNA) libraries were prepared from approximately 100 ng of total RNA by the North Carolina State University Genomic Science Laboratory using a NEB Ultra Directional Library Prep Kit for Illumina then sequenced on an Illumina NextSeq 500. Average size of the cDNA fragments was approximately 380 bp.

RNA Sequence Alignment Strategy

Known pairwise polymorphism in the *M. guttatus* complex ranges from 0.033 to 0.065 (Brandvain et al., 2014; Gould et al., 2017), raising the prospect of bias from mapping RNA reads to the *M. guttatus* reference genome, which is likely to be different from the lines utilized here. To reduce the number of reads discarded during mapping, we used a pseudo-reference

TABLE 1 | Biological replicates of SEC39 ovules and SEC39 x CSS4 hybrid seeds.

Time point (OAP)	Biological repticates	No. maternal plants	No. pollen donors	Sample names
0	5	8	N.A.	Od_BR1_81, Od_BR4_S10, Od_BR5 _S15 Od_BR6 _S17, Od_BR7_819
2	4	4	5	2d_BR1S11, 2d_BR2_S16
4	3	3	5	2d_BR3_S18 2d_BR4_S20 4d_BR2_S4, 4d_BR3_S7, 4d_BR4_S12
6	4	4	5	6d_BR1S2, 6d_BR2S5, 6d_BR3_S8, 6d_BR4_S13
8	4	4	5	8d_BR1_S3, 8d_BR2_S6, 8d_BR3_S9, 8d_BR4_S14

Altogether; we used eightmaternalplantsandfivepollendonors.Maternalplants #1and 4 wereused to generate two ovulesamples Pd_BR1_S1 and0d_BR4_S10and maternal plEnts 1-4 were used to gmerate **al** *hybrid seeds rolected from 2 to 8 days after pol/nation (Dl>P. Because of the difficulty of isolating sufficimt RNA for sequencing of unfertilized ovules, we poded samples de-ived from mate-na/ plants 2 and 3 withsampes fromthreeadditional maternal plants notusedelsewhere inthisstudy to generate sampes Od_BRS_S15, Od_BR6_S17, End 0d_BR 7 _S19. Thesamefivepo/lendonorswereused to generate* all *hyb-id seeds.*

genome mapping approach, mapping our RNA reads to parental pseudo -reference genomes that we generated from whole-genome DNA sequence data from each parental inbred line. Sequence data was obtained from DNA extracted from bud tissue from one maternal parent and one pollen donor using the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific). Illumina libraries were prepared by the Duke University Genome Sequencing Facility using the lllumina TruSeq DNA Nano library prep kit, then sequenced on an Illumina HiSeq 4000 machine, generating 150 bp paired-end reads. We removed adaptors and low quality reads with Trimmomatic (Bolger et al., 2014) and mapped the remaining reads using bwa mem (http://bio-bwasourceforge.net), retaining only properly paired reads. Mean coverage was 38x and 32x for CSS4 and SEC39, respectively. Variants were called using GATK3.7 and hard-filtered by quality-by-depth (< 2.0), strand bias(> 60.0), mapping quality(< 20.0), mapping quality rank sum(< 12.5),and the read position rank sum test(< - 8.0).We also removed variants with low coverage (<4) and high coverage (> 2 SD from the mean). We generated parental pseudo-reference genomes that incorporated our filtered variants using the package ModTools (Huang et al., 2014). We used bedops (Neph et al., 2012) to calculate pairwise coding sequence divergence between SEC39 and CSS4.

We mapped RNA reads to each parental pseudo-reference using the default settings of the splice-aware aligner STAR, allowing a maximum number of mismatches relative to read length of 0.04 (Dobin et al., 2013). We then used the lapels/suspenders pipeline (Huanget al., 2014; Crowley et al., 2015) to filter reads by mapping position and quality. For each read, lapels determines its mapping location in the maternal and paternal pseudo-reference,compares their mapping qualities in each (as determined by the number of mismatches), and then selects the mapping position with the highest quality. For our study, reads that mapped uniquely to only one parental genome were retained, while reads that mapped to both parental genomes were assigned the coordinates with the highest mapping quality. Only properly paired, uniquely mapped reads were retained for downstream analyses.

Time Series Analysis of Gene Expression Changes With Development

We generated gene counts on our lapels/suspenders filtered alignments using featureCounts (Liao et al., 2014). We performed a principle component analysis (PCA) of normalized gene expression estimates from the 20 RNA libraries using the PlotPCA function of DESeq2, using the 500 most highly variable genes (Love et al., 2014). We took two approaches to discovering genes that were differentially expressed over the course of seed development. First, we combined multiple pairwise comparisons (O *vs.* 2 DAP, 2 *vs.* 4 DAP, 4 *vs.* 6 DAP, and 6 *vs.* 8 DAP) into a generalized linear model in edgeR, performing a likelihood ratio test to determine significance with a false discovery rate (FDR) of 0.01. Count data were trimmed mean of M values (TMM)-normalizedand only genes with greater than 1 count-per-million in at least three samples were retained (Chen et al., 2016). Second, we used Next maSigPro, a program which employs a least-squared polynomial regression and log-likelihoodratio test to detect genes exhibiting significant changes in expression over time (Conesaet al., 2006; Nueda et al., 2014). Next maSigPro selects the best model by goodness-of-fit using a correlational cutoff (i.e., R^2 value) supplied by the user. Such time series models are an improvement over performing multiple pairwise comparisons (Spies et al., 2019), especially when several time points are sampled. Next maSigPro includes batch effects in its regression model; we considered each maternal parent to be a batch, giving us seven total batches: four maternal parents distributed across 0- 8 DAP and three additional batches corresponding to the three pooled samples from O DAP. We performed a polynomial regression with an R^2 of 0.6 and Ben jamini-Hochberg corrected FDR of 0.01. We took the overlap between edgeR and the Next maSigPro time series analysis to be the set of genes that were *differentially expressed* over the course of seed development. All further downstream analyses were performed on these differentiallyexpressed genes (DEGs). We validated our RNA isolation, mapping, and normalization of gene expression by performing quantitative real-time PCR (qRT-PCR) with independently - collected, triplicate RNA samples from unfertilized ovules and whole hybrid seeds from a new set of three maternal plants and three pollen donors on a subset of six genes (see **Figure Sl** for details) and comparing their trends to the transcriptomics analysis. The qRT-PCR temporal expression patterns of all six genes closely matched the RNA-seq based expression patterns.

To uncover trends in the gene ontology (GO) terms ofDEGs, we performed the Hartigan and Wong K-means clustering algorithm (Hartigan and Wong, 1979) on the fragments per kilobase of transcript per million (FPKM)values of all DEGs. We used 20 clusters (as determined by the gap statistic), with 100 random starts and 25 iterations. We performed a GO-term enrichment analysis of resulting clusters in ThaleMine (https://apps.araport.org/thalemine/begin.do) using the most similar *Arabidopsis thaliana* homologues, then used REVIGO (http://revigo.irb.hr/) to filter out redundant GO-terms, using a similarity cutoff index of 0.7. While we used the reference annotation (Phytozome v2.0) to calculate *Mimulus* gene abundances, we established homology of genes relative to *Arabidopsis* using BLASTx (Camachoet al., 2009) with a cutoff of E 0.001, designating the hit with the highest bitscore as the most similar *A. thaliana* homolog. There were few differences between the BLASTx results and the phytozome annotation (Goodstein et al., 2012). Here and elsewhere in this paper, we present the gene name and gene designation for the most similar homologue from *Arabidopsis thaliana* for the *M. guttatus* gene number. We used plantTFDB4.0 (http://planttfdb.cbi.pku.edu.en/) to annotate transcription factors in *M. guttatus,* tomato *(Solanum lycopersicum),A. thaliana,* and rice (*Oryza sativa)* (Jin et al., 2015; Jin et al., 2017). We aligned protein sequences of type I MADS-box genes from these species, and estimated and bootstrapped a neighbor joining (NJ) tree (1000 reps) using clustalW2.l (Larkin et al., 2007).

We compared our expression data with previously published *M. guttatus* transcription data collected from calyx, leaves, petals, and stem tissue (Edger et al., 2017) to determine which genes and transcription factors are expressed exclusively in ovules or seeds. We blasted the transcripts in Edger et al. (2017) to the *M. guttatus* reference genome to obtain their gene identity, selecting the hit with the highest bitscore and using a cutoff of Es; 0.001. We note that both our data and that ofEdgeret al. (2017) include transcripts that do not correspond to the current *M. guttatus* genome annotation; we chose not to pursue these unannotated transcripts further. We considered a gene to be *expressed* in our data if it had a mean FPKM;::: 1.0 in at least one time point. We categorized as *stage-specific*those genes whose expression in our samples had a mean FPKM ;::: 5.0 at that stage but s; 1.0 at all other stages. Finally, we categorize genes as *exclusive* to ovules or seeds if their expression levels as reported in Edger et al. (2017) was an FPKM < LO and they had a minimum mean expression of FPKM;::: 5 in ovules or 2- 8 DAP seeds. We compare our data to seed transcriptomes of *A. thaliana* (Belmonte et al., 2013), maize (Chen et al., 2014; Yi et al., 2019), and domesticated tomato and a near wild relative *(Solanum pimpinellifolium)* (Pattison et al., 2015; Shinozaki et al., 2018) to look for overlapping sets of *seed-exclusive* genes in these taxa

Detecting and Validating Genes With *Mimulus* guttatus-BiasedExpression

We used GATK's ASEReadCounter to assemble *M. guttatus* and *M. pardalis* allele counts at SNP positions distinguishing our inbred lines, CSS4 and SEC39, for each library of 2- 8 DAP seeds. We confined our allele counts to sites that were homozygous in the 5Ginbred lines used to generate the parental pseudo-references. At each time point, we assembled lists of genes that either 1) exhibited no maternal expression at any SNP in any biological replicate, or 2) exhibited significant paternal bias for one or more SNPs within a gene for two or more replicates. For the former we eliminated genes exhibiting either little overall expression (< 2 counts averaged across replicates) or whose expression across replicates within a time point was highly variable (i.e., standard deviation in expression> mean expression). For the second group, ateachSNP wecalculated the ratio*o fM.guttatusallelic*counts toM. *pardalis* allelic counts (Mg/Mp) for each replicate, then calculated the mean Mg/Mp ratio across replicates. We retained only those genes where a majority ofSNPs within the geneexhibited Mg/Mp > 2 and where variance in Mg/Mp was low (i.e., where the standard deviation of Mg/Mp among SNPs was less than the mean across all SNPswithin a gene).Onlygenesexpressed in twoor more replicates were considered for future validation Because our search for *M . guttatus -biased* expression was performed on whole-seed transcriptomes, we cannot apply any *a priori* hypothesis for the ratioofMg/Mp expression at any given gene,since that gene maybe expressed in oneor moretissuetypes(i.e.,seed coat, embryo,and/or endosperm). Thus, our findings do not constitute a screening for imprinted genes in *Mimulus.*

Transcripts that displayan expression biastoward *the M.guttatus* allele could originate from a parent-of-origin bias, an allele-of-origin bias or be false positives. To both validate the expression bias and distinguish between these possibilities we hand-isolated endosperm from both SEC39 x CSS4 and CSS4 x SEC39 Fl seed at 8 DAP. RNA was converted to cDNA and used in semi-quantitative PCR assays. We identified seven candidate genes containing known allele-distinguishing SNPs due to the creation or elimination of a restriction enzyme cutting site. We created pairs of oligos **(Table** SI) to amplify the SNP-containing fragments. Relative expression of maternal and paternal alleles was assayed by agarose-gel electrophoresisand bySanger sequencing.

RESULTS

We generated 20 RNAseq libraries with a total of 3.85 x 10^8 raw reads, averaging 1.93 x 10^7 reads per library (see Table S2 for alignment statistics). Pairwise sequence divergence (1t) between SEC39 and CSS4 in coding regions is 0.028 ± 0.022 S.D. Mapping to the SEC39 pseudoreference produced a mean of 75.3% ±4.25 S.D. properly paired, uniquely mapped reads, while mapping to the CSS4 pseudoreference produced a mean of70.0% ± 2.34 S.D. properly paired, uniquely mapped reads. Mapping to the SEC39 (i.e., maternal) pseudoreference genome was more efficient since our samples contain reads from seed coat (all maternal), embryo (1:1 maternal:paternal), and endosperm (2:1 maternal:paternal). Altogether, our merged alignments have an average of 75.5% ± 3.98 S.D. paired, uniquely mapped reads. This represents a significant improvement over mapping to the IM62 reference genome, which produced 67.7% ± 4.0 S.D. uniquely mapped reads (t_{19}= 68.97, p < 0.0001) for an increased mapping rate of -8%, similar to that found byothers (Huang et al., 2014; Crowley et al., 2015).

Developmental Progression of *Mimulus parda/is* x *Mimulus guttatus* Seeds

As expected, seed set, outer morphology, and germination rates among mature seeds from *M. pardalis* x *M. guttatus* (SEC39 x CSS4) confirms a lack of postzygotic isolation between these inbred lines (Oneal, Munger and Willis, unpublished; Macnair and Cumbes (1989)) (see Figure 1, Figure S2). Selfed *M. pardalis,* selfed *M. guttatus,* and reciprocal *M. pardalis* x *M. guttatus* fruits contained mostly round, viable seeds with a small minority of shriveled or flat seeds (Figure I A, Figure S2), with no significant difference in the distribution of seed types between the selfed fruits and hybrid fruits (MANOVA $F_{9,45}$ = 1.09, p > 0.1). Germination rates were generally high (57.9- 89.74%)(Figure lB).

At 2 DAP, most seeds contain four-cell embryos, progressing to eight-cell and dermatogen embryos by 4 DAP (Figure 2A). By 6 and 8 DAP, seeds contain late globular- and heart-stage embryos, respectively. Endosperm development is *ab initio* cellular (Arekal, 1965; Oneal et al., 2016). There is regular proliferation of endosperm between **4** and 8 DAP, and limited variation in the distribution of embryonic stages at each time point **(Figure 2B),** which may result from differences in the rate of pollen tube growth and timing of fertilization. A PCA of normalized gene expression values from the 20 libraries produced well-defined clusters corresponding to the five different developmental time points assayed **(Figure** 2C), with the exception of some overlap between 2 and **4** DAP, indicating

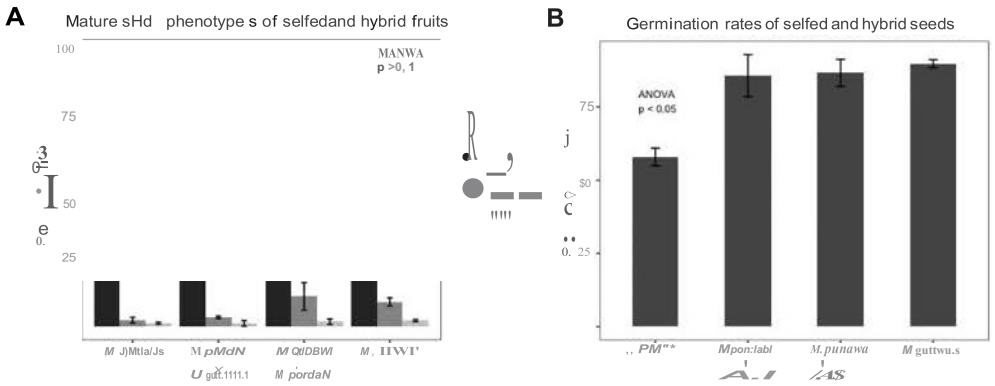

FIGURE 1 | (A) Percentage of mature seed phenotypes (round, slYiveled, and flat) recovered from ftJJy deM3 1oped In. its from selfed *Mimuk.Jspardalis* and *Mim.1k.Jsguttatus* and reciprocal crosses of *M.parda/is* x *M. guttatus*. There was no sigri ficant dif erence in the distribution of seed types between the se edfrutts and hybrid frutts (MJINOVA Fe,45= 1.09, p > 0.1). **(B)** Germination rates of selfed *M. parda/is* and *M. guttatus* seed, as well as *M. pardaKs* x *M. guttatus*, and *M. guttatus* x *M. pardaKs* hybrid seeds. Mp, ***M.****pardalis;* Mg , *M. g.,ttatus.* Bars indicate standard error. G ermination rates were generallyhg, (57.9-89.74%) (N=4 fer each selfing and cross).

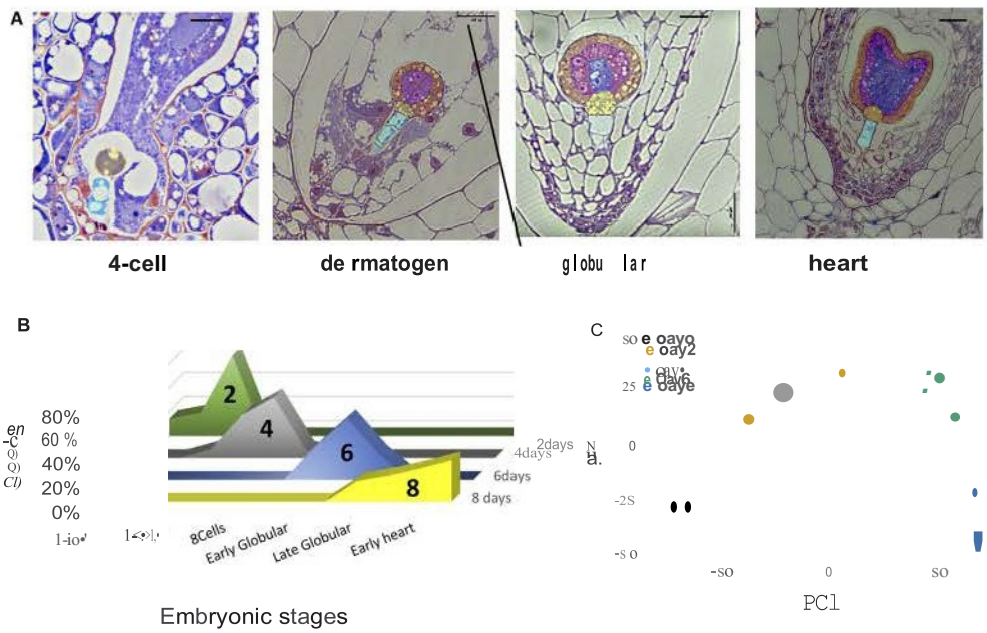

FIGURE 2 | **(A)** Mo rph ology of *MimuAJs parda/is* x *Mimulus guttatus* developing seeds collected 2, 4, 6, and 8 days after pollination (DAP). Sections were obtained from LR-white embedded seeds (see *Methods).* (Bl Frequency distribution of embr)Onic stages of seeds collected at 2, 4, 6, and 8 DAP. (C) Principle component analysis (PCA) plot of RNA libraries.

that these stages exhibit similar transcriptional states. The first axis explains 66% of the variation, and reflects the effect of time, while the second axis explains 22% of the variation.

Correlations Between Gene Expression and the Progression of *Mimulus guttatus* Seed Development

We detected the expression of 18,836 annotated genes in ovules and seeds, representing 67% of all currently annotated M. *guttatus* genes, including 1,011 transcription factors (TFs), of which 40 are MADS-box genes. The number of genes expressed in any stage was highest in ovules (16,681) and lowest in heart-stage seeds (8 DAP: 14,784), but we found no relationship between gene expression diversity and collection time point (ANOVA $F_{4,15} = 2.522$, p = 0.085) (Fig ure S3). A generalized linear model incorporating multiple, progressive pairwise comparisons in edgeR identified 12,384 DEGs (FDR $ 0.01), while a time series analysis in Next maSigPro identified 6,613 DEGs ($R^2 = 0.6$, FDR$ 0.01) (Tables S3 and **S4).** We focus on the overlappingsets ofDEGs between edgeR and Next maSigPro.

These consisted of 6,233 DEGs **(Figure S4),** including 388 transcription factors (TFs) from 49 families, of which 16 are type I MADS and 4 are MIKC-type MADS-box genes.

Clustering Reveals Co-Expression Patterns Enriched for Functional Pathways

To further correlate patterns of gene expression with developmental changes in seeds, we used a Hartigan and Wong K-means clustering algorithm (Hartigan and Wong, 1979) on the 6,233 DEGs identified by edgeR and Next maSigPro. Of 20 K-clusters, 13 were significantly enriched for GO-terms, including several enriched for GO-terms known to play a role in seed development in *Arabidopsis* (see Figure 3, Figure S5, Tables S5 and S6). Several of these clusters are of note and mirror what is known about gene expression patterns in developing seeds in other taxa For example, genes that exhibit peak expression in ovules are enriched for regulation of RNA metabolic process, regulation of gene expression and transcription, and regulation of multicellular organismal development (cluster 5). In maize, genes that are highly expressed in the female gametophyte include transcription factors involved in RNA regulation (Chenet al., 2014). In *Arabidopsis,* epigenetic regulatory pathways established in the egg and central cell of the female gametophyte set the stage for the production of embryo-targeting small interfering RNAs (siRNAs) in the endosperm (Wuest et al, 2010). Development at 6 DAP is characterized by the emergence of the late globular-stage embryo (Figure 2A) and the acceleration of endosperm proliferation (Oneal et al, 2016). A group of co-expressed genes which peaks in both ovules and in globular (6 DAP) seeds is enriched for GO-terms related to cell cycle processes, DNA replication, and chromatin organization (duster 17) as it is in *Arabidopsis* (Le et al., 2010; Belmonte et al, 2013). Another duster that peaks at 6 DAP (duster 2) contains genes involved in cell wall biogenesis and organization. Finally, genes that remain unexpressed until 8 DAP are enriched for carbohydrate and lipid metabolic processes, seed oil body biogenesis, and lipid localization (cluster 18). In *Arabidopsis* and maize these biological processes are enriched in endosperm (Belmonte et al, 2013; Li et al., 2014), suggesting they may perform similar roles in species with *ab initio* cellular endosperm development

We identified a subset of genes whose expression was *stage-specific* in our data (see *Methods),* with 72 genes specific to ovules, and 3, 29, and 77 genes specific to 2, 6, and 8 DAP seeds, respectively **(Table** S7). There were no genes *specific* to 4 DAP seeds from. A comparison of our data with expression data from M. *guttatus* calyx, petal, leaf, and stem tissue (Edger et al., 2017) revealed 30 genes whose expression was exclusive to ovules (i.e., FPKM 5 in ovules and $ 1 in any other tissue), including 4 transcription factors (TFs) **(Table** S7). There were 478 genes *exclusively* expressed in seeds, including 35 TFs from multiple families and 6 type I MADS-box genes encoding transcription factors, with 2, 20, and 48 genes *exclusive* to the 2, 6, and 8 DAP stages of seed development, respectively **(Tables S8** and **S9).** A heatmap of expression of *seed-exclusive* transcription factors in M. *guttatus* shows the majority are not constitutively expressed, but rather, are chiefly expressed in 6 or 8 DAP seeds **(Figure S6).**

Overlap Between *Seed-Exclusive Mimulus guttatus* Genes and *Seed-Exclusive* Genes in Other Taxa

K-means clustering of genes expressed in seeds reveals that a large fraction (N=l20, or 25%) of *seed-exclusive* genes segregate into cluster 18. Another 101 (21.1%) of *seed-exclusive* genes cersegregate into cluster 2. Comparing *seed-exclusive* genes in M. *guttatus* to

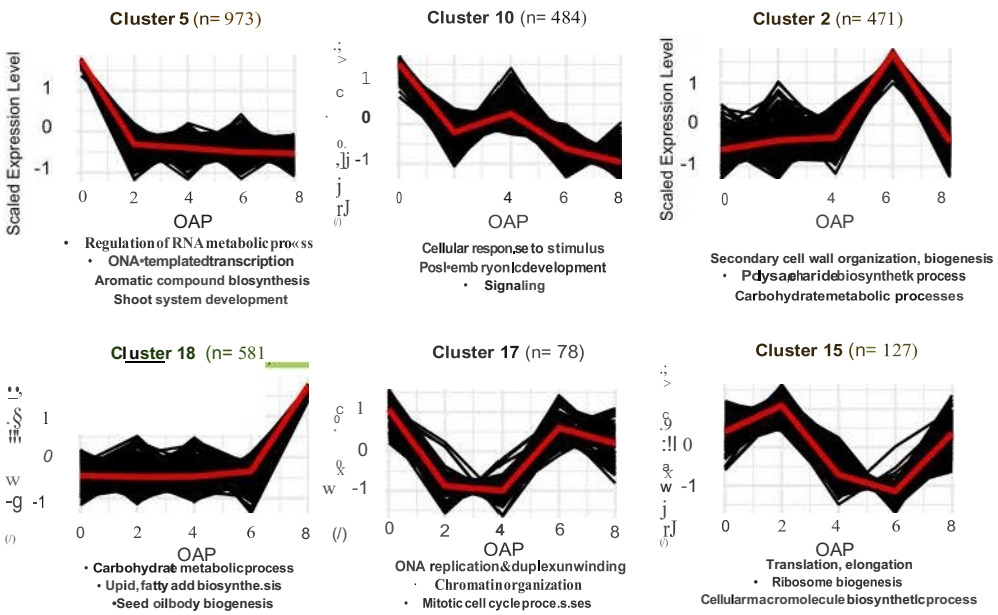

FIG U RE 3 | Exarrples of non-overlappingk-means clusters of co-expressed genes and tre ir significant gene ontology (GO) enrichment terms.

other taxa, including A *thaliana,* maize, and tomato reveals some overlap. In A *thaliana,* there are 43 seed-exclusive genes whose nearest homologs are also seed exclusive in *M. guttatus* (Table S7); one of these genes is enriched in the *A. thaliana* seed coat (At5g39130) (Belmonte et al., 2013). Also among the *M. guttatus seed-exclusive*genes are two that are not exclusive to A *thaliana* seeds but do exhibit enriched expression in *A. thaliana* endosperm: *AtAGL62* (At5G60440) and *AtCYSS* (At5g47550), a putative phytocystatin expressed in seedlings (Song et al., 2017). *AtAGL62* suppresses endosperm cellularization in *A. thaliana* and is a key regulator of endosperm development The *seed-exclusive* expression pattern of 2 *M. guttatus* homologues of *AGL62* (Migut.B00708 and MigutD01476) suggests that they play an important role during *Mimulus* seed development; however, without RNA sequencing of isolated seed tissues and additional functional analyses, we cannot yet determine the role of *M. guttatus AGL62* homologues. Finally, there are 25 genes enriched in A *thaliana* endosperm that are not expressed in *M guttatus* seeds, including three transcription factors *(AtAGL87, AtFIS2,* and *AtFWA).*

In maize, there are 23 seed exclusive genes whose nearest homologs are also *seed exclusive* in *M. guttatus* **(Table** S7), 11 enriched in endosperm, and 2 enriched in the embryo (Chen et al., 2014; Yi et al., 2019). Of the seed-exclusivegenes in *M. guttatus,*54 genes have homologs that are not seed-exclusivebut were found to exhibit enriched expression in the maize embryo, endosperm, or both.relative to other seed tissues (Chen etal., 2014). There are 250 genes enriched in maize endosperm whose homologs are not expressed at all in *M. guttatus* seeds, including 39 transcription factors such as transcription factors activating maize *ABI3 (VPl), AUX/IAA,* and two maize type I MADS genes *(MADS25* and *MADS21).* Another gene enriched in maize endosperm but which is not expressed in *M. guttatus* seeds is the maize *FIEl,* which is maternally imprinted and peaksin expression during the transition from mitotic cell division to endoreduplication in the endosperm (Hermon et al., 2007). Twenty *M. guttatus* genes have nearest homologs that are exclusively expressed in both *A. thaliana* and maize seeds (Table S7). There is, however, no overlap between genes unexpressed in *M. guttatus* seeds but enriched in *A. thaliana* or maizeendosperm.

Finally, in domesticated tomato there are 8 seed-exclusivegenes whose nearest homologs(25genes)are seed exclusive in *M.guttatus* (Zuluaga et al., 2016; Sbinozaki et al , 2018), one of which *(CYSS)* is enriched in the embryo of a wild relative of domesticated tomato (Pattison et al., 2015) (Solyc04g014780). Of the 25 seed-exclusive genes *M. guttatus* with homologs in tomato, 22 do not have seed-exclusive homo logs in *A. thaliana* or maize. These include *M. guttatus* and tomato homologs of *LIPID TRANSFER PROTEINS 4* and *12 (LPT4* and *LPT12:* Migut.M01356/Migut.M01355/ Migut.M01350/Solyc05g053530), *PECTIN METHYLESTERASE INHIBITOR 2 (PMEI2:* Migut.A00l19/Migut.I00331/Solyc09 g072950), and *VACUOLAR INHIBITOR OF FRUCTOSIDASE 1 (VIF I :* Migut.M00087/Solyc09g072950). A GO-term enrichment analysis of these genes finds two overrepresented biological processes: negative regulation of catalytic activity, and negative regulation of molecular function (adjusted p-values 0.003 and 0.004, respectively).

Type I MADS-Box Gene Expression

Because of their known role in female gametophyte and seed development across a range of plant taxa (Colombo et al., 1997; Busi et al., 2003; Kohler et al., 2003; Arora et al., 2007; Colombo et al., 2008; Kang et al., 2008; Steffen et al., 2008; Zhang et al., 2018), we further examined the temporal expression patterns of differentially expressed type I MADS-box genes using a hierarchical clustering analysis. This analysis identified co-expressed sets, or hierarchical expression clusters of type I MADS box genes, a familythat includes *theArabidopsisAGAMOUS-LIKE* genes *AGL26, AGL62,* and *AGLl04.* Two *Mimulus* type I MADS-box DEGs were expressed in ovules but unexpressed after fertilization **(Figure 4,** cluster VI); other MADS-box DEGs were either expressed primarily in four-cell to dermatogen-stageseeds (2-4 DAP, clusters *N* and V), or in globular- and heart-stageseeds (6- 8 DAP, clusters I and II) **(Figure4).** A neighbor joining (NJ)tree of typeI MADSgenesfrom*M.guttatus,S.lycopersicum, A. thaliana,* and 0 . *sativa* suggests that in some cases, genes with similar temporal expression patterns are each other's nearest relatives **(Figure** 5).

Four Imprinted Endosperm-Expressed Genes Display Strong Paternal Expression Bias

We assembled a list of 163 genes that demonstrated expression bias in favor oftheM. *guttatus* allele as described above **(Tables S8** and **S9).** *M. guttatus-bias* could be the result of cis-regulatory allele-specific expression bias or imprinting (Stupar and Springer, 2006; Zhang andBorevitz, 2009); the former would beM.gu ttatus-biased regardless of the crossing direction. We validated 7 of these putatively *M.* guttatus-biased genes as above and all are expressed in endosperm. Four genes were imprinted, exhibiting paternally-biased expression in isolated endosperm regardless of the direction of the cross **(Table 2, Figure 6).** Intriguingly, only one of the validated genes exhibits temporally significant expression changes: Migut.EOll17 is unexpressed until 6 DAP then declines at 8 DAP; the remaining were not identified as DEGs by our time series analysis. Allfour genes haveimprinted homologues in otherspecies **(Table** 2). The remaining three putative *M. guttatus-biased* genes exhibited allele-specific expression (Table 2).

DISCUSSION

Seed development can be partitioned into stages: early embryo patterning and endosperm proliferation, embryo organ initiation and morphogenesis, and the onset of maturation, during which the endosperm accumulates seed storage proteins and the embryo enters dormancy (Agarwal et al., 2011). In *Mimulus,* as in *Arabidopsis,* the majority of genes expressed in ovules and seeds are expressed across all stages of ovule and seed development and comparatively few genes are *stage-specific* with an even smaller number *exclusive* to ovules or seed development from 2 to 8 DAP (Table S7), suggesting that the majority of genes fulfill multiple developmental and biological roles (Le et al., 2010; Belmonte et al., 2013; Chen et al., 2014).

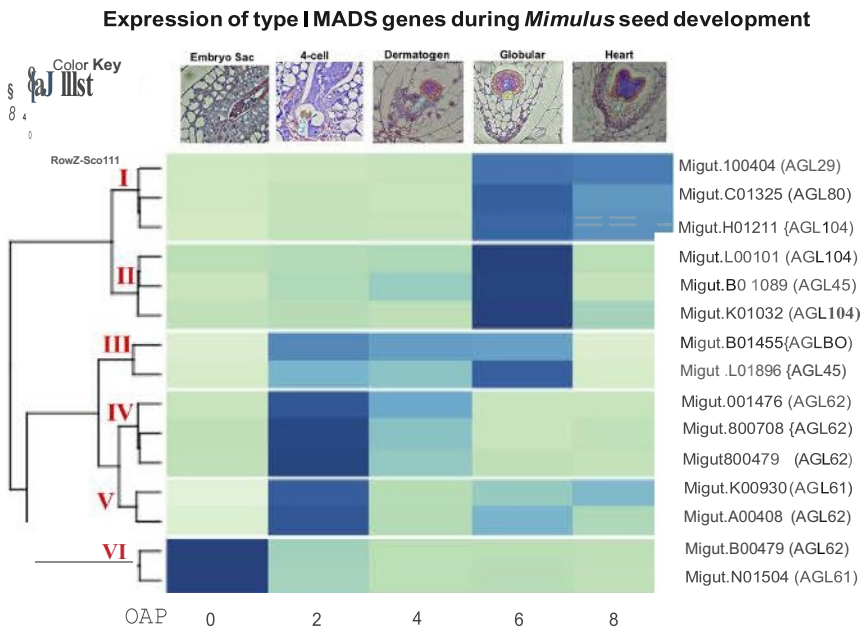

FIGURE 4 | A heatmap of MADS-boxtype I gene expression in ovules and seeds. The nearest *Atabidopsis thafana* homologue for each *Mimuk.Js guttatus* MADS-box gene is indicated in parentheses (as determined by 81..ASTx).

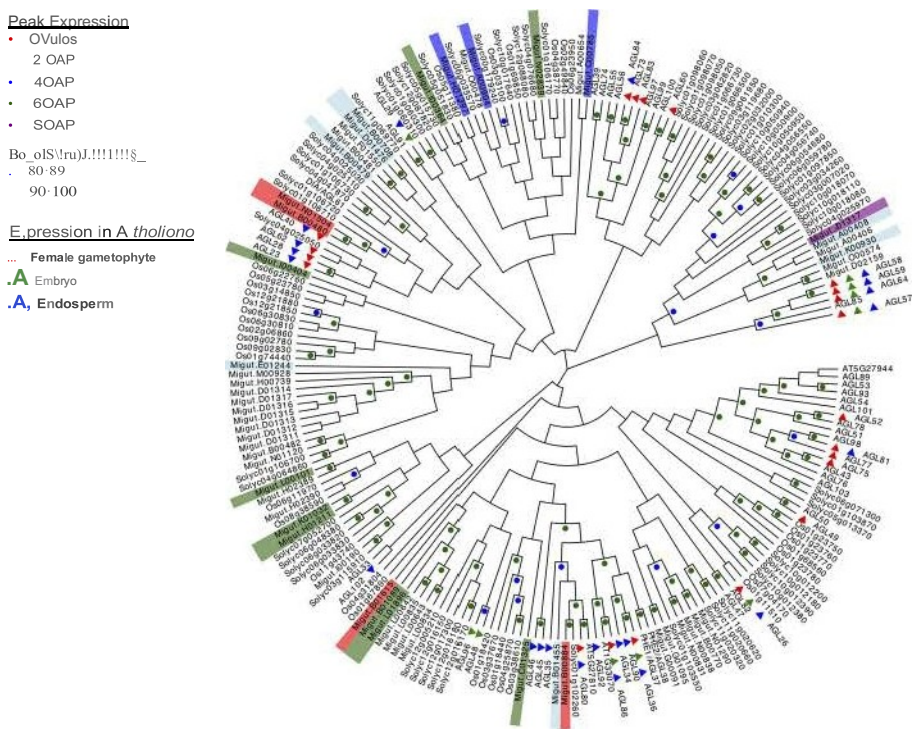

FIGURE 5 | Neighbor joining treeof type I MAffi -boxgenes from rice (Oryza satiVa), *Arabidopsis thafana,* tomato (So/am.m o/CX>Persctm), and*Mimuk.Js guttatus* (reference fineIM62).Bootstrapsupportisrepresented by color(green:;,90, blue80-89). Nodleswh lessthan80% supportarenotmarked. *M. guttatus*genes expressed in our data set are color-coded by time of peak expression. A *thafana* expression profiles are indicated by color-coded triangles.

TABLE 2 | Genes exhibiting *Mmulus guttatus* allele specific bias selected for validation

M. guttatus gene name	*A. thaliana* homologue	Gene name	Description	Imprinted	Imprinted homologues
Migut.E01117	AT5G52230	*MBD13*	Methyl-CpG binding transcription regulator	Yes	*Capse//a.rLbe//a.* (MEG and PEG), maize (PEG), castor (MEG)
Migut.H00744	AT5m 9790	Trtthorax-related 5 *(A7XR5)*	Histone-lysine N-methyltransferase: targets H3K27 for epigenetic silencing of repetttive and transposon elements	Yes	Maize (PEG), *So/anum dlilense* (PEG), and *S. peruvianum* (PEG)
Migut.!00545	AT5G53150	*DnaJ*	DnaJ heat shock N-terminal domain chaperone protein	Yes	Capse//a. *grand/flora* (PEG), *Arabidopsis tha/iam* (PEG), *S. peruvianum, S. dliense* (MEG) and *S. arcamm* (PEG), maize (MEG), sorghum (MEG), castor (PEG)
Migut.!00995	AT1G51730	RWDdomain-containing protein	Ubiquttin-oonjugaitng enzyme family protein	No	NA
Migut.M00083	AT5m 8650	WD-repeat protein	RanBPM protein, function LJ1known	No	NA
Migut.N01317	AT2G16730	Beta-galactosidase 11• related *(fJGAL11)*	Catalysis of the hydrolysis of terminal, non-reducing beta-D-galactoseresidues in beta-D-galactosides.	Yes	Rioe (MEG), *S. crilense* (MEG and PEG)
Migut.l.01896	AT3005860	*AGL45*	MADS-box transcription factor	No	N.A.

Imprinted e>pression was *confrrred byrt-PCR, enzymaticd ;gestion,* and *Sanger s9:1uercing. Refe'ences:Capsella: (HalorEng8!1et al , 2016;Lafon.Placetleet cl., 2018);maize: j\1Vates Elal , 2013); IWJidopsis thaliana: (Pignatta et al , 2014);* Somt.m: *(Fbrez-Rueda et al., 2016; Rothet cl., 2018); sorght.m: (Zhang et al, 2016);*oce*(Lw et al , 2011; C/-en et al , 2018); castnr(Xu et al, 2014.)*ASE, *allele specific expression. We have designated as NA geres that we-enot irrprinted in Mimulus; we did not sea-ch for homoloflJ('S of these genes in other taxa.*

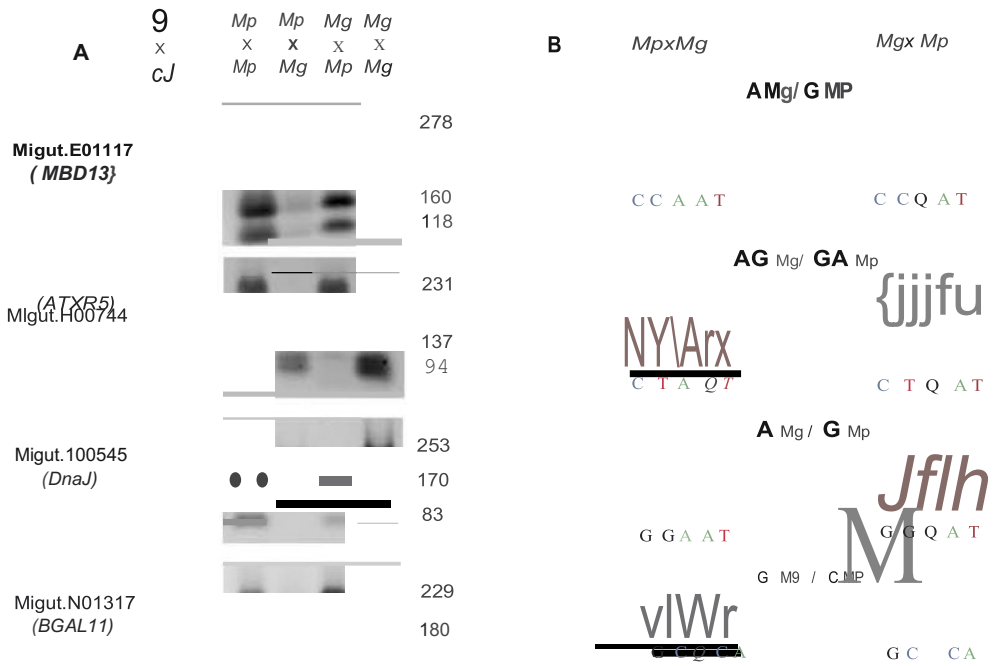

FIGURE 6 | Validation of four paternally expressed imprinted genes: Migut.E01117 *(MB0/3),* Migut.H00744 (ATRX5), Migut.!00545 (f)naJ), and Migut.N01317 *(BGAL11) via* semi-quantttative PCR and Sanger sequencing. **(A)** Allele-specific real-time (RT)-PCR analyses of each gene in the endosperm of selfed and reciprocal crosses of *MimuAJs parda/is* (Mp) and *Mimuus gutta/us* (Mg) is given. For each gene, gel shows the sizes of restriction fragments after RT-PCR amplification and digestion with restriction endonucleases. **(B)** Paternal expression was oonfi!Tll8d by real-time (RT)-PCR sequencing chromatographs at selected SNP regions measuring allele-specific expression in reciprocal crosses.

A question of particular interest is "what is the set of genes that regulate the *ab initio* cellular type of endosperm development and furthermore, are the expression of these key regulators uniquely expressed in cellular endosperms"? We observe some overlap between the set of genes expressed in a *seed-exclusive* manner in M. *guttatus* and *seed-exclusive* genes detected in other species (i.e., A *thaliana,* maize, and tomato). Additionally, there are 8 *seed-exclusive* tomato genes (associated

with 25 *seed-exclusive Mimulus* homologues) that are not expressed in a seed exclusive fashion in maize or *Arabidopsis.* These genes might be good candidates for future functional analyses that aim to identify regulatory differences between *ab initio* cellular and syncytial endosperm developmental strategies. However, it remains possible that the difference between *ab initio* and syncytial endosperm development is determined by differences in structural aspects of encoded proteins or in other changes (e.g., post -transcriptional) that are not reflected in the transcriptional patterns that we have identified in this work. Only future functional analyses can address this question. We do note a lack of overlap between *A. thaliana seed-exclusive* and maize *seed-ex clusive* genes enriched in endosperm but unexpressed in M. *guttatus* seeds. This suggests a significant degree of transcriptional divergence between these two taxa that both display syncytial development and may reflect the fact that these species are highly genetically and evolutionarily divergent and show marked differences in seed development overall *(A. thaliana* is a dicot and maize is a monocot).

Most expressed genes do not exhibit significant changes associated with fertilization and major developmental events within seeds. A substantial portion, however, are differentially expressed and segregate into clusters that are suggestive of their roles in development. Three co-expression gene clusters (clusters 2, 17, and 18) deserve further discussion here due to their associated functional GO-terms and the presence of stage-specific gene expression patterns.

Cluster 17: Genes Enriched for Mitotic Cell Cycle and Chromatin Organization

Genes within cluster 17 are highly expressed in ovules, decline in expression and then increase in globular- and heart-stage seeds **(Figure** 3). Overrepresented GO-terms indicate functional roles in the regulation of chromatin organization, DNA replication and the cell cycle. A *Mimulus* homologue of *TOPOISOMERASE II (TOPI!),* which functions to ensure proper chromosome movement during cell division in *A. thaliana* (Martinez-Garcia et al., 2018) is also contained within cluster 17. Other genes within this cluster have homologues associated with chromatin organization including *AtCMT3* and *AtMETl,* which act to maintain methylation of CHH and CG sites, respectively, during DNA replication (Lindroth et al., 2001; Kankel et al., 2003). In *A. thaliana,* seed development is accompanied by extensive gain of CHH methylation, especially within transposable elements, and *CMT3* and *METl* transcription is largely confined to the embryo (Hsieh et al., 2009; Kawakatsu et al., 2017). By contrast, multiple homologues of *CMT3* are expressed in tomato endosperm at the globular-stage(Roth et al., 2019), raising the possibility that the function of *CMT3* in *Mimulus* is more similar to tomato than *A. thaliana.* Also within this cluster is a *Mimulus* homologue of *AtHMGB3,* a family of proteins with an HMG-box DNA-binding domain that act as global modulators of chromatin structure through the assembly of nucleoprotein assemblies (Pedersen and Grasser, 2010). We also find a homologue of *MINICHROMOSOME MAINTENANCE 7* (MCM7), and the timing of its expression mirrors its expression pattern in *A. thaliana.* AtMCM7 is part of a complex which unwinds DNA prior to replication, and is maternally expressed in the egg and central cells of the ovule, then biparentally expressed the proliferating embryo and endosperm (Herridgeet al., 2014).

Cluster 2: Genes Associated With the Recruitment of a Secondary Cell Wall Biosynthesis Regulatory Module

Cluster 2 is comprised of genes that are minimally expressed from O to **4** DAP, peak dramatically in globular-stage seeds (6 DAP), and then again decline in heart-stage seeds (8 DAP) **(Figure** 3). These genes showa strong overrepresentation of GO-annotations associated with secondary cell wall biogenesis and cell wall remodeling activities (Table S5). Cluster 2 contains *Mimulus* homologues of the *Arabidopsis IRREGULAR XYLEM (IRX)* genes includingIRXl (At4gl8780), *IRX3* (At5gl7420), and *IRXS* (At5g44030), all encoding cellulose synthase catalytic subunits required for secondary wall formation (Tayloret al., 1999; Taylor et al., 2000; Taylor et al., 2004), as well as *Mimulus* homologues of *IRX7* (At2g28110), *IRX9* (At2g37090), *IRXJO* (Atlg27440), *IRX12* (At2g38080), *IRX14* (At5g67230), and *IRX 15-L* (At5g67210), which encode additiona l enzymes involved in secondary cell wall synthesis (Zhong and Ye, 2015). Cluster 2 also contains transcription factors that constitute a key regulatory network of secondary cell wall formation. These include *Mimulus* homologues of *SECONDARY W ALL-ASSOCIATED NAC DOMAIN PROTEIN 2 (SND2)* (At4g28500), *KNATJ* (Atlg62990), and five members of a MYB family of transcription factors that regulate secondary cell wall formation (At4g22680, At3g08500, At4g33450, Atlg73410, Atlg66230). Thus, a large portion of the regulatory machinery underlying secondary cell wall synthesis in *Arabidopsis* appears to be conserved in *Mimulus* and is recruited to mediate some secondary cell wall events specific to this stage of development, which includes rapid proliferation of endosperm (Oneal et al., 2016). In addition to the large number of genes with known functional annotations in *Arabidopsis,* many other genes within the expression cluster would be expected to regulate and function in secondary cell wall biosyn thesis in *Mimulus.* As such, our analysis identifies additional *Mimulus* candidates for future functional studies of secondary cell wall regulation.

Cluster 18: Genes Involved in Auxin Synthesis and the Accumulation of Nutrient Reservoirs in Endosperm

The plant hormone auxin is associated with nearly every aspect of plant growth and development, including seed and fruit development (Zhao, 2010; Li et al., 2016; Figueiredo and Kohler, 2018). We detected the expression of several components of the auxin biosynthesis and signaling pathway in cluster 18 which peaks in 8 DAP seeds **(Figure** 3). Indeed, two such genes expressed in *exclusively* in heart-stage seeds are M. *guttatus* homologues of *TRYPTOPHAN AMINOTRANSFERASE RELATED2 (TAR2;* Migut.D01989) and *YUCCA4 (YUC4;*

MigutA00287) which encode the TAR and YUCCA enzymes in the major auxin biosynthetic pathway (Brumos et al., 2014). Multiple *AUXIN RESPONSE PACIOR (ARP)* genes are also expressed, including three homologues of the transcription factor *ARF19* (Migut.L01587, Migut.J00834, Migut.L01914), which in tomato is highly expressed in seed coat (Pattison et al., 2015) and can play a major role in regulating seed size (Sun et al., 2017).

Arabidopsis homologues of *Mimulus* genes found in this cluster are associated with seed oil biosynthesis, and display increasing expression within maturing *A. thaliana* seeds (At5g52920, At3g02630, and At5g63380) (Hajduch et al., 2010). Thus, we suggest these *Mimulus* genes play a role in the metabolic processes which generate the energy stores required by the mature seed pheno type. One such group of major regulators of seed maturation includes the *LEAFY COTYLEDON (LBC)* group of genes. *FUS3* is a component of this group and acts as a transcriptional activator (Keith et al., 1994; Luerl3en et al., 1998; Yamamoto et al., 2009; Wang and Perry, 2013); *fas3* mutants lose embryo identity and fail to germinate (Harada, 2001). *ABI3* is another transcriptional activator, and *abi3* mutants exhibit impaired embryo maturation and germination (Parcy et al., 1994). Both *FUS3* (Migut. F00019/At3g26790) and *ABI3* (Migut.D0051l/At3g24650) segregate into cluster 18, and are differentially expressed over the course of seed development: *MgFUS3* is expressed solely in heart-stageseeds, while *MgABI3* is expressed in globular- and heart-stage seeds. Similarly, the nearest homolog of *LECl -like* (AtLJL/At5g47670/Migut. M01289) is also contained with this cluster. These results suggest that *MgABI3* and *MgFUS3* may regulate the expression of this cluster of genes in a manner similar to what is seen in *Arabidopsis,* maize, barley, and tomato seeds (McCarty et al., 1991; Bassel et al., 2006; Abraham et al., 2016).

Mimulus MADS-Box Genes Expressed During Seed Development

MADS-box genes are transcription factors that play key roles in regulating developmental transitions involved in germination, root growth, flowering, and reproduction (Zhang and Forde, 1998; Saedler et al., 2001; Bemer et al., 2010). Swveys of type I MADS-box expression in *A. thaliana* have found that members of this gene family are expressed in multiple cell types within the female gametophyte and the developing seed (Bemer et al, 2010; Zhang et al, 2018). The origin of type II MADS-box genes can be attributed largely to whole genome duplication events (Shan et al., 2007), while type I MADS-box genes appear to have evolved *via* bothwhole genome duplication and smaller scale duplication events (Nam et al., 2004). Prior work has found widespread evidence of gene loss, subfunctionalization, and neofunctionalization of this important gene family (Vandenbussche et al., 2003; Kramer et al., 2004; de Martino et al., 2006; Kater et al., 2006; Airoldi and Davies, 2012).

It is intriguing that several temporally co-expressed type I MADS genes are each other's nearest relatives **(Figure** 5). For example, twogenes thatpeakin expression inovules(Migut.B00480 and Migut.N01504) and two that peak in 2 DAP seeds (Migut.B00708 and Migut.D01476) are sister to each other in the NJ tree (Figure5) and fallintothesameexpression clusters (dusters 5 and !,respectively).Together with Migut.B00479, these genes are contained within a clade that includes the MADS genes *AtAGL23* (Atlg65360), *AtAGL28* (Atlg01530), and *AtAGL62* (At5g60440). In *A. thaliana,* these genes function in regulatory networks in the female gametophyte and developing seed (Colombo et al., 2008; Kanget al, 2008; Steffen et al, 2008; Bemer et al., 2010; Figueiredo et al.,2015; Fiumeetal.,2017; Zhang etal.,2018). TwoMADSgenes that are differentiallyexpressed in 6 DAP seeds (Migut.B01089 and Migut.L01896) are contained within a clade that includes *AtAGL%* and *AtAGL48,* both expressed in the globular embryo (Bemeret al., 2010), as is the closely related tomato MADS-box gene Solycl2g016150 (Pattison et al., 2015). Migut.K01032 and Migut.H0121l are also paired and peak in 6 DAP seeds.

Several type I MADS-box genes exhibit temporal patterns of expression in *M. guttatus* that mirror their expression in *A. thaliana.* Four MADS-box genes are expressed in the female gametophyte but not the fertilized seed of *A. thaliana,* including *AtAGL49* (Atlg60040), *AtAGL60* (Atl g72350), *AtAGL73* (At5g38620), and *AtAGL83* (At5g49490) (Bemer et al, 2010). *M. guttatus* homologues of these genes *(AtAGL49:* MigutB01668; *AtAGL60, AGL73,* and *AGL83:* Migut.N01504) are also expressed in *M. guttatus* ovules but not in seeds **(Table** S7). Similarly, *AtAGL102* (Atlg47760), *AtAGL34* (At5g26580), and *AtAGL90* (At5g27960) are expressed in *A. thaliana* seeds but not prior to fertilization (Bemer et al., 2010; Zhang et al., 2018); the *M. guttatus* homologues of these genes *(AtAGL102:* Migut.K00930; *AtAGL34* and *AtAGL90:* Migut.B01455) are also unexpressed in ovules, peaking dramatically in four-cell embryo *M. guttatus* seeds (2 DAP) (TableS7).Moreover, six of the seed-specific genes we identified are type I MADS-box genes, including one that is also seed-specific in *A. thaliana* (AtAGL45/At3g05860: Migut.B01089). Finally, *AtAGL26* (At5g26880), *AtAGL28* (Atlg01530), *AtAGL40* (Atlg01530), and *AtAGL62* (At4g36590) are expressed both in the *A. thaliana* female gametophyte and the seed, as are their *M. guttatus* homologues *(AtAGL26:* Migut.N01481; *AtAGL28, AtAGL40,* and *AtAGL62:* MigutB00480). While the history of duplication, retention and loss of type I MADS-box genes in *Mimulus* remains uncertain, these findings are suggestive of shared roles for these important transcription factors in gametophytic and seed development in A *thaliana* and *M. guttatus.*

Paternally Imprinted Genes in *Mimulus* Endosperm

The prevailing hypothesis for the evolution of imprinted expression is the kinship theory, which posits that biases in the expression of maternally or paternally derived genes have evolved to maximize fitness of mothers *vs.* offspring with respect to offspring provisioning *via* endosperm (Haig and Westoby, 1989; Haig, 2004). Under this scenario, genomic conflict over maternal resources could lead to rapid evolution of imprinted loci, resulting in rapid sequence divergence and/or turnover in the imprinting status of genes among even closely

related species (Kohler et al., 2010; Kohler et al., 2012). In general, surveys of imprinted genes have found little overlap in imprinting status of loci between closely related (Wolff et al., 2011) or distantly related taxa (Luo et al., 2011; Waters et al., 2013; Yoshida et al., 2018), however, some loci are imprinted in several different species. Here we report the first validated imprinted genes in *M. guttatus* [(but see (Kinser et al., 2018)]. We found that all four of our validated paternally imprinted genes (PEGs) have homologues that are imprinted in other species, including maize, sorghum, and rice **(Table** 2). Given our *ad hoc* method for validating candidate PEGs, which included as a criterion the potential for imprinting in other taxa, this is not surprising. Nevertheless, the fact that these genes are imprinted in multiple species, including one with imprinted homologues in six other species (Migut.100545: *Dnaf* heat shock chaperone protein), suggests that proteins encoded by these genes perform similar functions in endosperm across plant taxa. Previous work indicates that many PEGs are involved in epigenetic regulation (Waters et al, 2013; Pignatta et al., 2014) *via* DNA methylation or repressive histone modifications. Two *M. guttatus* genes we identified as exhibiting paternally- biased expression belong to these functional categories: a methyi-CpG binding transcription factor MigutE01117 (AtMBD13/At5g52230) and a chromatin remodeling protein, MigutH00744 *(AtA TXRSI* At5g09790). Methyl-CpG binding transcription factors mediate CpG methylation by coordinating the activities of histone deacetylases and histone methyitransferases (l.emach and Grafi, 2007). Likewise, *AtATXRS* promotes gene silencing at constitutive heterochromatin and repression of TEs by orchestrating mono- methylation at H3K27 Oaoob et al, 2009). Homologues of both genes are imprinted in *Capsella,* maize, and castor *(MBD13)* and in maize and two species of *Solanum* (Waters et al., 2013; Xu et al., 2014; Hatorangan et al, 2016; Roth et al., 2018). Homologues of *Dnaf* heat shock protein (Migut.100545) are imprinted in six other genera, including *Solanum* (Waters et al, 2013; Xu et al., 2014; Hatorangan et al, 2016; Roth et al., 2018), while homologues of *BGALll* (Migut.N01317) are imprinted in rice and *Solanum* (Luo et al, 2011; Chen et al., 2018; Roth et al., 2018).

Disruption in the allelic dosage of paternally and maternally imprinted genes in endosperm can lead to dysfunctional endosperm development and ultimately, seed abortion, and is associated with hybrid seed failure (Haig and Westoby, 1989; Haig and Westoby, 1991; Scott et al, 1998; Kinser et al., 2018). Postzygotic isolation *via* such hybrid seed inviability is widespread in *Solanum* (Baek et al., 2016), and imprinted loci have been implicated in this process (Florez-Rueda et al., 2016; Roth et al., 2018; Roth et al., 2019). It is striking that three of our validated PEGs have homologues that are imprinted in *Solanum* (Table 2), and is suggestive of shared patterns of imprinting in these taxa, which both exhibit a cellular endosperm development (Arekal, 1965; Oneal et al., 2016). Hybrid seed inviability is also common in the *M. guttatus* complex (Vickery, 1966; Vickery, 1978) and is known in some cases to be caused by a failure of endosperm development (Oneal et al., 2016; Coughlan et al., 2018). Moreover, the effects of some genomic loci contributing to hybrid seed inviability is dependent upon whether they were contributed by the maternal or paternal genome (Garner et al., 2016). Future, systematic surveys of gene expression in isolated endosperm of reciprocally-crossed endosperms in *M. guttatus* and in closely related, diploid sister species (e.g., *Mimulus nudatus, Mimulus tilingii,* and *Mimulus decorus),* coupled with characterizations of their expression in incompatible hybrids could demonstrate whether genomic imprinting, which is thought to evolve *via* parental conflict over maternal resources, is a major factor driving hybrid incompatibilityand speciation in this rapidly evolving group.

In conclusion, we used RNA sequencing to profile dynamic gene expression events in seeds across an 8-day time course and used gene ontology enrichment analyses to reveal statistically overrepresented biological functional categories of co-expressed clusters of genes. With respect to the regulation of a number of important biological events during seed development (e.g., secondary cell wall biosynthesis, seed maturation, auxin synthesis) the coexpressed gene clusters reported here provide valuable entry points for the investigation of novel molecular regulatory mechanisms during *Mimulus* seed development. Similarly, our analysis of the *Mimulus* MADS type I transcriptional regulators highlights the subset of the family expected to function during seed development and furthers our understanding of the evolution of this gene family. Furthermore, we detected and validated genes exhibiting paternally imprinted expression. The conservation of the imprinted status of homologous genes within other studied angiosperms strongly indicates they play a functional role during endosperm development in *Mimulus,* an area for future study. Our data broadens the taxonomic framework underlying studies of seed development generally and constitutes a valuable resource to other researchers exploring seed development and hybrid inviability in *Mimulus,* as well as future studies of comparative gene expression.

DATA AVAILABILITY STATEMENT

DNA and RNA sequence data have been deposited in the SRA and GEO database (accessions SRX6435295, SRX6435296 (DNA) and GSE123424 (RNA)). Supporting data is provided in the supplementary tables with this submission and includes all primers used in this study, alignment statistics for RNA libraries, FPKM values for all genes, the results of edgeR, Next maSigPro, clustering, GO-term enrichment analyses, and raw SNP count data for genes exhibiting M. guttatus-biased expression from 2-8 DAP.

AUTHOR CONTRIBUTIONS

MF-V, EO and RF collaboratively designed the study. MF-V managed overall experimental approaches and generated RNA samples and libraries and validated imprinted genes. EO, MF-V, GV and ML analyzed the data.EO managed analysis approaches

and assembled list of genes with M. guttatus-biased expression. MC performed the histological sectioning and analysis. CR generated seed morphology data. MF-V, EO and RF managed data interpretation and wrote the manuscript. SCand *TW* revised and edited the manuscript.

FUNDING

This research was funded by NSF grants IOS-1557692(RF) and IOS-1558113(JW), as well as a grant from the European Union's Horizon2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement (No 690946) to SC and RF. CR's participation in this project was fonded by the Integrated Molecular Plant Systems NSF-REU (Award DBI-1558579).

ACKNOWLEDGMENTS

The authors thank Steffen Heber and Annie Jeong for helpful discussions on RNA sequence analysis, Candace Haigler for helpful discussions on secondary cell wall biogenesis, Jenn Coughlan for feedback on the manuscript, and Danielle Curran, Bianca Clark and Jessica Markowitz for assistance with crossing assays. The authors also thank the Phytotron staff at North Carolina State University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00132/full#supplementary-material

REFERENCES

Abraham, Z., Iglesias-Fernandez, R., Martinez, M., RubierSomoza, I., Diaz, I., Carbonero, P., et al. (2016). A developmental switch of gene expression in the barley seed mediated by HvVPl (Viviparous-I) and HvGAMYB interactions. *Plant Physiol.* 170, 2146-2158. doi: 10.1104/pp.16.00092

Agaiwal, P., Kapoor, S., and Tyagi, A. K (2011). Transcription factors regulating the progression of monocot and dicot seed development *Bioessays* 33, 189- 202. doi: 10.1002/bies.201000107

Airoldi, C. A., and Davies, B. (2012). Gene duplication and the evolution of plant MADS- box transcription factors./. *Genet Genornics* 39,157- 165.doi:10.1016/j.jgg.2012.02.008

Arekal, G.D. (1965). Embryology of Mimulus Ringens. *Bot. Gaz.* 126, 58-66. doi: 10.1086/3362%

Arora, R., Agarwal, P., Ray, S., Singh, AK, Singh, V. P., Tyagi, AK., et al.(2007). MADS- box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genornics* 8,242. doi: 10.1186/1471-2164-8-242

Baek, Y. S, Royer, S. M, Broz, **A. K,** Covey, P. A. Lopez-Casado,G., Nw1ez, R, et al. (2016). lnterspecific reproduc!i\e barriers between sympatric populations of wild tomato species (Solanum section Lycopersicon). *Arn. J. &t.* 103, 1%4-1978. doi: 10.3732/ajb.1600356

Bassel, G. W., Mullen, R. T., and Bewley, J. D. (2006). ABI3 expression ceases following, but not during, germination of tomato and Arabidopsis seeds./. *Exp. Bot.* 57, 1291- 1297. doi: 10.1093/jxb/erjlOl

Belmonte, MF., Kirkbride, RC., Stone, SL, Pelletier, J. M, Bui, A Q., Yeung, E. C., et al (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl Acad. Sci.U.S.* A 110, FA35-E444. doi:10.1073/pnas.1222061110

Bemer, M., Heijmans, K., Airoldi, C., Davies, B., and Angenent, G. C. (2010). An atlasof type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiol.* 154, 287-300. doi: 10.1104/pp.110.160no

Brandvain, Y., Kenney, A M., Flagel, L., Coop, G., and Sweigart, A. L. (2014). Speciation and introgression between Mimulus nasutus and Mimulus guttatus. *PloS Genet.* 10, el 004410. doi: 10.1371/journal.pgen.1004410

Bolger, A M., Lohse, M, and Usadel, B. (2014). Trimmomatic: a flexible trimmer for ilwnina sequence data *Bioinforrnatics*30, 2114-2120. doi:10.lll/ppli2098

Brumos, J., Alonso, J.M., andStepanova, A. N. (2014). Genetic aspects of auxin biosynthesis and its regulation. *Physiol. Plant* 151, 3- 12. doi:10.1111/ppl.12098

Busi, M V., Bustamante, C., D'Angelo,C., HidalgerCuevas, M., Boggio, S. B., Valle, E. M., et al. (2003). MADS- box genes expressed during tomato seed and fruit de\•dopment *Plant Mal. Biol.*52, 801- 815. doi:10.1023/A:1025001402838

Camacho, C, Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K, et al. (2009). BLAST+: architecture and applications. *BMC Bioinf* 10, 421. doi: 10.1186/1471-2105-10-421

Chen,J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A, et al. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol.* 166, 252- 264. doi: 10.1104/pp.114240689

Chen, Y., Luo, A.T. L, and Smyth, G. K (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihoo d pipeline. *FJOO0 Res* 5, 1438. doi: 10.12688/flOOOresearch.89872

Chen, C., Li, T., Zhu,S., Liu, Z.,Shi, Z., Zheng, X., et al. (2018). Characterization of imprinted genes in rice reveals conservation of regulation and in1printing with other plant species. *Plant Physiol.* 177, 1754- 1771. doi: 10.1104/pp.17.01621

Colombo, L, Franken, J., Van der **Krol,** A R, Wittich, P. E., Dons, H. J., and Angenent, G. C. (1997). Downregulation of ovule-specific MADS box genes from petw1ia results in maternally controlled defects in seed development. *Plant Cell* 9, 703- 7 IS. doi: 10.1105/tpc.9.S.703

Colombo, M., Masiero, S., Vanzulli, $., Lardelli, P., Kater, MM, and Colombo, L (2008). AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J.* 54, 1037 - 1048. doi: IO.lilt/j. B65-313X2008.03485.x

Conesa, **A.,** Nueda, M. J., Ferrer, A, and Talon, **M.** (2006). maSigPro: a method to identify signliicantly dilferential expression profiles in time-course mkroarray experiments. *Bioinforrnatics* 22, 10%- 1102. doi:I0.1093/bioinformatics/btl056

Coughlan, J.M., Brown, **M. W.,** and Willis, J. H. (2020). Patterns of hybrid seed inviability in perennials of the *Mirnulus guttatus sp.* complex reveal a potential role of parental conflict in reproductive isolation. *Curr. BioL* 30, 83- 93.eS.doi: 10.1016/j.cub.2019.11.023

Crane, P. R., and Lidgard, S. (1989). Angiosperm diversification and paleolatitudinal gradients in cretaceous floristic diversity. *Science* 246, 675-678. doi: 10.1126/science.246.4930.675

Crowley, J. J., Zhabotynslcy, V.,Sun, W., Huang, S, Pakatci, l K, Kim, Y., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat.Genet.* 47,353- 360. doi:10.1038/ng.3222

de Martino, G., Pan, I., Emmanuel, E., Levy, A, and Irish, V. F. (2006). F,mctional analysesof two tomatoAP£TALA3genes demonstratediversification in theirrolesin regulating floral *development.PlantCell*18,1833- 1845.doi:10.1105/tpc.106.042978

Dobin, A, Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15- 21. doi: 10.1093/bioinformatics/bts635

Edger, P. P., Smith, R., McKain, M. R., Cooley, AM., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150- 2167. doi: 10.1105/tpc.17.00010

Figueiredo, D. 0., and Kohler, C. (2018). Auxin: a molecular trigger of seed development. *Genes Dev.* 32, 479-490. doi: I0.1101/gad.3US46.118

Figueiredo,D. D., Batista, R. A, Roszak, P. J.,and Kohler, C.(2015).Auxin production couples endosperm development to fertilization. *Nat Plants* I, 15184.doi:10.1038/nplants.2015.184

Fiume, E., Coen, 0., Xu, W., Lepiniec, L., and Magnani, E. (2017). Developmental patterning of sub-epidermal cells in the outer integument of Arabidopsis seeds. *PloS One* 12, e0188148. doi: 10.1371/journal.pone.0188148

Florez-Rueda, AM. Paris, M, Schmidt, A., Widmer, A, Grossniklaus, U., and Stadler, T. (2016). Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *MoL BioL EvoL* 33, 2935-2946.doi: 10.1093/molbev/mswl75

Friedman, W. E. (1994). The evolution of embryogeny in seed plants and the de\oelopmental origin and early history of endosperm. *Am.]. Bot* 81, 1468-1486. doi: 10.1002/j.153?2197.1994.tbl5633.x

Friedman, W. E. (2001). Comparative embryology of basal angiosperms. *Curr. Opin. Plant BioL* 4, 14-20. doi: I0.1016/$13695266(00)00129-1

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., and Berger, F. (2003). Arabidopsis haiku mutants reveal new controls of seed size by endosperm. *Plant PhysioL* 131, 1661- 1670. doi: 10.1l04/pp.102.018762

Garner, A. G., Kenney, A. M., Fishman, L, and Sweigart, A L. (2016). Genetic loci with parent-of-origin effects cause hybrid seed lethality in crosses between Mimulus species. *New Phytol.* 211, 319-331. doi: 10.1111/nph.l3897

Geeta, R. (2003). The origin and maintenance of nuclear endosperms: viewing development through a phylogenetic lens. *Proc. R Soc. Lond. Ser. B: BioL Sci.* 270, 29- 35. doi: 10.1098/rspb.20022206

Goodstein, D. M., Shu, S., Howson, **R.,** Neupane, R., Hayes, R 0., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomks. *Nucleic Acids Res.* 40, Dll 78-Dll 86. doi: 10.1093/nar/gkr944

Gould, B. A., Chen, Y.,and Lowry, D. B. (2017). Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive diff"erentiation and reproductive isolation. *Mol. Ecol.* 26, 163- 177. doi: 10.1111/mec.13881

Grossnildaus, U., Vielle-Calzada, J. P., Hoeppner, **M. A.,** and Gagliano, W. B. **(1998).** Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science* 280, 446-450. doi: 10.1126/science.280.5362.446

Haig, D., and Westaby, M. (1989). Parent-Specificgene expression and the triploid endosperm. *Am. Nat.* 134, 147- 155. doi: 10.1086/284971

Haig, D., and Westaby, M (1991). Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between diff"erent ploidies of the same species, and its implications for the evolution of apomixis. *Philos. Trans. R Soc Lond. B BioL Sci.* 333, 1- 13. doi: 10.1098/rstb.19910.057

Haig, D. (2004). Genomic imprinting and kinship: how good is the evidence? *Annu. Rev. Genet.* 38, 553- 585. doi: J0.ll46/annurev.genet.37.110801.l42741

Hajduch, M., Hearne, L.B., Miernyk, J. A., Casteel, J.E., Joshi, T., Agrawal, G. K, et al. (2010). Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. *Plant Physiol.* 152, 2078- 2087. doi: 10.1104/pp.109l52413

Harada, J. J. (2001). Role of Arabidopsis LEAFY COTYLEDON genes in seed development/. *Plant PhysioL* 158, 405-409. doi: I0.1078/0176-1617-00351

Hartigan, J. A., and Wong, M.A. (1979). Algorithm AS136: a K-Means clustering algorithm. /. *R Stat SocSer. CAppl.Stat.* 28, 100-108. doi: 10.2307/2346830

Hatorangan, M. R, Laenen, B., Steige, K. A., $latte, T., and Kohler, C. (2016). Rapid evolution of genomic imprinting in twospecies of the Brassicaceae. *Plant Cell* 28, 1815-1827. doi: 10.1105/tpc.16.00304

Hem1on, P., Srilunchang, K -O., Zou, J., Dresselhaus, T., and Danilevskaya, 0. N. (2007). Activation of the imprinted Polycomb group Fie1 gene in maize endosperm requires demethylation of the maternal allele. *Plant Mol. BioL* 64, 387 - 395. doi: 10.1007/sl ll 03--007 -9160 - 0

Herridge, R. P., Day, R. C., and Macknight, RC. (2014).The role of the MCM2-7 helicase complex during Arabidopsis seed development *Plant Mol. BioL* 86, 69 - 84. doi: 10.1007/sll l03--014--0213 -x

Hsieh, T.-F., Ibarra, C. **A.,** Silva, P., Zemach, A, &b ed-Williams, L., Fischer, R. L, et al. (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science* 324, 1451-1454. doi: 10.1126/science.ll72417

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., *et* al. (2011). Regulation of imprinted geneexpression in Arabidopsis endos perm. *Proc. NatL Acad Sci. U.S. A.* 108, 1755- 1762.doi: 10.1073/pnas.1019273108

Huang,S., Holt, J., Kao, C.-Y., McMillan, L., and Wang, W. (2014). A novel multi-alignment pipeline for high-thr oughput sequencing data. *Database* 2014, 1- 12. doi: 10.1093/database/bau057

lngouff", M, Jullien, P. E., and Berger, F. (2006). The female gan1etophyte and the endosperm control cell proliferation and diff"erentiation of the seed coat in Arabidopsis. *Plant Cell* 18, 3491-3501. doi: 10.1105/tpc.106.047266

Jurgens, G., Torres Ruiz, R. A., and Bedeth, T. (1994). Embryonic pattern formation in flowering plants. *Annu. Rev. Genet.* 28, 351- 371. doi: 10.1146/annurev.ge.28.120194.002031

Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V., Stroud, H., Coleus, S., et al. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat. Struct. Mol. Biol.* 16, 763 - 768. doi: 10.1038/nsmb.1611

Jin, J.,He,K, Tang, X., L Z.Lv, L, Zhao, Y., et al (2015 An arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mo/. BioL Evol.* 32, 1767-1773. doi: 10.1093/molbev/lDS\()58

Jin,J., Ttan, F.,Yang,0 .-C., Meng, Y.-Q., Kong, L, Luo, J.,et al. (2017). PlantTFDB 4.0:toward a central hub for transcription factors and regulatory interactions in plants. *Nudeic Acids Res.* 45, O1040- Dl 045. doi: 10.1093/nar/gkw982

Kohler, C, Henni L.Spillane, C., Pien, S.Grui5san, W ., and Grossniklaus, U.(m3). ThePolyromb-group protein MEDEA regulates seed de\oelopment by controlling expression ofthe MADS-box genePHERESl. *Genes Dev.*17,1 1553.doi:10.1101/gad257403

Kohler, C., M.ittelsten Scheid, 0., and Frilova, A. (2010). The impact of the triploid block on the origin and evolution of polyploid plants. *Trends Genet.* 26, 142- 148. doi: 10.1016/j.tig2009.12.006

Kohler, C., WoUf, P., and Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annu. Rev. Plant BioL* 63, 331 - 352. doi:10.1146/annurev -atp lant--042811-105514

Kang, 1.-H., Steffen, J. G., Portereiko, M F., Lloyd, A, and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell* 20, 635- 647. doi: 10.1105/tpc.107.055137

Kankel, M W., Ramsey, D . E.,Stokes, T. L, Flowers, S. K, Haag, J. R, Jeddeloh, J. A, et al. (2003). Arabidopsis MEf1 cytosine methyltransferase mutants. *Genetics* 163, 1109- ll 22.

Kater, M. M., Dreni, L, and Colombo, L (2006). f'Wlctional conservation of MADS-box factors controlling floral organ identity in rice and Arabidopsis. /. *Exp. Bot.*57, 3433- 3444. doi: l 0.1093/jxb/erl097

Kawakatsu, T., Nery, J. R., Castanon, *R.,* and Ecker, J. R (2017). Dynamic DNA methylation reconfiguration during seed devdopment and germination. *Genome Biol.* 18, 171. doi: 10.1186/s13059--017 - 1251-x

Keith, K, Kraml, M., Dengler, N. G., and McCourt, P. (1994). fusca3: a heterocbronic mutation affecting late embryo development in arabidopsis. *Plant Cell* 6, 589- 6 00. doi: 10.2307/3869865

Kinser, T. J, Smith, R D., Lawrence, A H., Cooley, A. M, Vallejo-Marin, M., Smith, G. D. C., et al. (2018). Mechanisms driving endosperm-based hybrid incompatibilities: insights from hybrid monkeyflowers. *bioRxiv* 461939, 1- 30. doi: 10.ll0l/461939

Kran1er, E. M., Jaramillo, M. A., and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166, 1011- 1023. doi: 10.1534/genetics.166.2.1011

Lafon-Placette, C., Hatorangan, M R., Steige, K A., Camille, A, Lascaux, M, Slotte, T., et al. (2018). Paternally expressed imprinted genes associate with hybridization barriers in Capsella. *Nat. Plants* 4, 352- 357. doi:10.l038/s41477-018--0161 - 6

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A, McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947- 2948. doi: 10.1093/bioinformatics/btm404

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis **of** gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. NatL Acad. Sci. U.* S. *A.* 107, 8063- 8070. doi: 10.1073/pnas.1003530107

Li, G., Wang, 0., Yang, R, Logan, **K,** Chen, H., Zhang, S., *et* al. (2014). Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. *Proc. Natl. Acad. Sci. U.* S. *A.* 111, 7582- 7587. doi: 10.1073/pnas.1406383111

Li, S.-B., Xie, Z.-Z., Hu, C.-G., and Zhang, J.-Z. (2016). A review of Auxin Response Factors (ARFs) in plants. *Front Plant Sci.* 7, 47. doi: 10.3389/fpls2016.00047

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923- 930. doi: 10.1093/bioinformatics/btt656

Lidgard, S., and Crane, P.R.(1988). Quantitative analyses of the early angiosperm radiation. *Nature* 331, 344-346. doi: J0.1038/331344a0

Lindrodi, AM., Cao, X., Jackson, J.P., Zilbemlall, D., McCallwn, C.M., Henikolf, S., et al. (2001}. Requirement of CHROMOME1HYIASE3 for maintenance of CpXpG methylation. *Science* 292, 20n - 2oso. doi: J0.1126/science.1059745

Love, M. I., Huber, W., and Anders, S. (2014}. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/sl3059--014--0550 - 8

Luerilen, H., Kirik, V., Herrmann, P., and Misera, S. (1998}. FUSCA3 encodes a protein with a conserved VP1/ABB -like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana: moleatlar cloning of the FUSCA3 gene of Arabidopsis. *Plant J.* IS, 755-764. doi: 10.1046/j.1365-313X.1998.00259.x

Luo, M., Taylor, J. M, Spriggs, A., Zhang, H., Wu, X., Russell, S., et al. (2011). A genome-wide survey of intprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PloS Genet.* 7, el002125. doi: 10.1371/journal.pgen.1002125

Macnair, M. R., and Cumbes, Q. J.(I 989}.The genetic architecture of interspecific variation in mimulus. *Genetics* 122, 211- 222.

Magallon, S., and Castillo, A. (2009). Angiosperm diversification through time. *Am. J. Bot.* 96, 349- 365. doi: 10.3732/ajb.0800060

Martinez-Garcia, M., Schubert, V., Osman, K., Darbyshire, A., Sanchez- Moran, E., and Franklin, F. C.H.(2018).TOPll and chromosome movement help remove intedocks between entangled chromosomes during meiosis. /. *Cell Biol.* 217, 4070-4079. doi: 10.1083/jcb201803019

McCarty, D.R., Hattori, T., Carson, C. B., Vasil, V., Lazar, M., and Vasil, I. K. (1991). The Viviparous- I developmental gene of maize encodes a novel transcriptional activator. Cell 66, 895 - 905. doi: 10.1016/00928674(91}90436 3

Nam, J., Kim, J., Lee, S., An, G., Ma, H., and Nei, M (2004). Type 1 MADS.box genes have experienced faster birth-and-deathevolution than typeD MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci. U. S.* A IOI, 1910-1915. doi: 10.1073/pnas.0308430100

Neph, S., Kueltn, MS., Reynolds, A. P., Haugen, f., Thurman, Rf., Johnson, AK, et al. (2012}. BEDOPS: high -performance genomic feature operations. *Bioinformatics* 28, 1919- 1920. dot I0.1093/bioinfonnatics/btsZn

Nueda, M. J., Tarazona,,S and Conesa, A. (2014). Next maSigPro: upda ting maSigPro biocondnctor package for RNA-seq time series. *Bioinformatics* 30, 2598-2602. doi: 10.1093/bioinformatics/btu333

Oneal, f., Willis, J. H., and Franks, R. G. (2016}. Disruption of endosperm development is a major cause of hybrid seed inviability between Mintulus guttatus and Munulus nudatus. *New Phytol.* 210, 1107- 1120. doi: 10.1111/nph.13842

Parcy, F., Yalon, C., Rayna!, M., Ga ubier-Come lla, P., Delseny, M., and Giraudat,J. **(1994).** Regulation of gene expression programs duru1g Arabidopsis seed development roles of the ABJ3 locus and of endogenous abscisic acid. *Plant* Cell 6, 1567-1582. doi: 10.1105/tpc.6.ll.1567

Pattison, R. J., Csukasi, F., Zheng, Y., Fei, Z., van der Knaap, E., and Catala, C. (2015). Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during earlytomato fruit development. *Plant PhysioL* 168, 1684- 1701. doi: 10.1104/pp.15.00287

Pedersen, D. S., and Grasser, K. D. (2010). The role of chromosomal *HMGB* proteins in plants. *Biochim. Biophys. Acta* 1799, 171- 174. doi: 10.1016/j.bbagnn.2009.11.004

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

Roth, M., Florez-Rueda, A. M., Paris, M, and Stadler, T. (2018). Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *Plant J.* 95, 1084-1101. doi: 10.1111/tpj.14012

Roth, M, Florez-Rueda, A. M., and Stadler, T. (2019}. Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics* 212, 141- 152. doi: 10.1534/genetics.119.302056

Saedler, H., Becker, A., Winter, K. U., Kirchner, C., and Theissen, G. (2001). MADS- box genes are involved in floral development and evolution. *Acta Biochim. Pol.* 48, 351- 358. doi: 10.18388/abp2001_3920

Scott,R J., Spiehnan,M.,Bailey,J.,and Dickinson,H. G. (1998). Parent-of-origin effects on seed development ht Arabidopsis thaliana. *Development* 125, 3329- 3341.

Sechet, J., Marion-Poll, A, and North, H. M. (2018}. Emerging functions for cell wall polysaccharides accumulated during eudicot seed de\•elopment. *Plants* 7, 1- 15. doi: I0.3390/plants7040081

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and fwtctional diversification during the evolution of the API/SQUA subfamily of plant MADS- box genes. *Mo!. Phyloger1et EvoL* 44, 26-41. doi: 10.1016/j.ympev2007.02.016

Shinozaki, Y., Nicolas, P., Fernandez-Pozo, N., Ma, Q, Evanich, D. J., Shi, Y., et al. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nat. Commun.* 9, 364. doi: 10.1038/s41467--017 - 02782 -9

Song C., Kim,T., Chun W.S., and Lim, C. 0. (2017).The Arabidopsis Phytocystatin AtCYS5 enhances seed germmation and seedling growth under heat stress conditions. *MoL Cells* 40, sn - 586. doi: JO.l4348/molcells2017.0075

Spies, D., Renz, P. F., Beyer, T. A., and Cia udo, C. (2019}. Comparative analysis of differential gene expression tools for RNA sequencing time course data *Brief Bioinform.* 20, 288- 298. doi: 10.1093/bib/bbxllS

Steffen, J. G., Kang, 1.-H., Portereiko, MF., Lloyd, A., and Drews, G. N. (2008). AGl.61 interacts with AGLS0 and is required for central cell development in Arabidopsis. *Plant Physiol.* 148, 259- 268. doi: 10.1104/pp.108.119404

Stupar, R. M., and Springer, N. M. (2006). Cis-transcriptional variation in maize inbred lines B73 and Mol7 leads to additive expression patterns in the Fl hybrid. *Genetics* 173, 2199-2210. doi: 10.1534/genetics.106.060699

Sun, Y., Wang, C., Wang, N., Jiang, X., Mao, H., Zhu, C., et al. (2017). Manipulation of Auxin response factor 19 affects seed size in the woody perennial Jatropha curcas. *Sci. Rep.* 7, 40844. doi: 10.1038/srep40844

Taylor, N.G.,Scheible,W. R, Cutler,S.,Somerville,C.R,andTurner,S.R.(1999).The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell* 11, 769 - 780. doi:10.1105/tpc.11.5.769

Taylor, N. G., Laurie, S., and Turner, S. R (2000}. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. *Plant Cell* 12, 2529 - 2540. doi: 10.1105/tpc.12.122529

Taylor, N. G., Gardiner, J. C, Whiteman, R, and Turner, $. R (2004}. Cellulose synthesisin the Arabidopsissecondary cellwall.*Cellulose* 11,329- 338.cloi:10.1023/B:CELL.0()()()()464()5.11326.a8

Vandenbussche, M., Zethof, J., Souer, E., and Koes, R (2003). Toward the analysis of the petunia MADS box gene family by reverse and fotward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require *SEPALLATA-lik.e* MADS box genes in Petunia. *Plant Cell* 15, 2680- 2693. doi: 10.1105/tpc.017376

Vickery, R. K. (I %6}. Speciation and isolation in sectioo simiolus of the genus min1ulus. *Taxon* 15, 55-63. doi: 10.2307/1217588

Vickery, R. K. (1978). "Case Studies in the Evolution of Species Complexes in Mun ulus," in *Evolutionary Biology.* Eds. M. K. Hecht, W. C. Steere and B. Wallace (Boston, MA: Springer US), 405- 507.

Wang, F., and Perry,S. E. (2013). Identification of directtargets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant Physiol.* 161, 1251-1264. doi: I0.1104/pp.112.212282

Waters, A. J., Bilinski, P., F.ichten, S. R., Vauglm, M. W., Ross-Ibarra, J., Gehring, M, et al (2013}. Compreltenshoe analysis of im printed genes in mai7.e reveals allelic variation for imprinting and limited conservation with oilier species. *Proc. Natl. Acad. Sci. U. S.*A. 110, 19639- 1% 44. doi: J0.1073/pnas.1309182110

West, M, and Harada, J.J. (1993}. Embryogenesis in higher plants: an overview. *Plant Cell* 5, 1361 - 1369. doi: 102307/3869788

Wolif, P., Webiliofer, 1, Seguin, J., Roszak, P., Beisel, C., Donoghue, M T. A., et al. (2011}. High- resolution analysis of parent-of-origin allelic expression in the arabidopsis endospemt. *P/oS Genet.* 7, el002126. doi:10.1371/joumal.pgen.1002126

Wuest, S. E., Vijvemerg, K, Schmidt, A, Weiss, M., Gheyselinck, J., Lohr, M. et al. (2010}. Arabidopsis female gan1etophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol* 20, 506- 512. doi: 10.1016/j.cub2010.0l.051

Xu, W., Dai, M., Li, F., and Liu, A. (2014). Genomic imprinting, methylation and parent-of-origm effects in reciprocal hybrid endosperm of castor bean. *Nucleic Acids Res.* 42, 6987-6998. doi: I0.1093/nar/gku375

Yamamoto, A., Kagaya, Y., Toyoshima, R, Kagaya, M, Takeda, S., and Hattori, T. (2009). Arabidopsis NF-YBsubunitsLECI and LECl -LIKEactivate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* 58, 843- 856. doi: 1 0.ll 11/j.1365-313X2009.038l 7.x

Yi, F., Gu, W., Chen, J., Song, N., Gao, X., Zhang, X., et al. (2019). High Temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell* 31, 974-992. doi: 10.1105/tpc.18.00%1

Yoshida, T., Kawanabe, T., Bo, Y., Fujimoto, R., and Kawabe, A. (2018). Genome-Wide analysis of parent-of-origin allelic eicpression in endospenns of brassicaceaespecies, Brassica rapa. *Plant Cell Physiol* 59, 2590-2601. doi: l0.l0 93/pcp/pcyl78

Zemach, A., and Grafi,G. (2007). Methyl-CpG-binding domain proteins in plants: interpreters of DNA methylation. *Trends Plant Sci.* 12, 80 - 85 . d oi: 10.1016 / j.tplants2006.12.004

Zhang, X., and Borevitz, J.0 . (2009). Global analysis of allele-specific expression in Arabidopsis thaliana. *Genetics* 182, 943- 954. doi: 10.1534/genetics.109.103499

Zhang, H., and Forde, B. G. (1998). An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. *Science Z79,* 407-409. doi: 10.1126/ science279.5349.407

Zhang, M., Li, N., He, W., Zhang, H., Yang, W., and Liu, B. (2016). Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J.* 85 , 424-436. doi: 10.lll 1/tpj.13116

Zhang, S., Wang, D., Zhang, H., Skaggs, M I., Lloyd, A., Ran, D., et al. (2018). Fertilization -independent seed-polycomb repressive complex 2 plays a dual role in regulating type l MADS-box genes in early endosperm development *Plant Physiol.* 177, 285- 299. doi: 10.1104/ pp.17.00534

Zhao, Y. (2010). Auxin biosynthesis and its role in plant development. *Annu . Rev. Plant Biol* 61, 49- 64. doi: 10.1146/ann urev-arplant-042809-112308

Zhong, R., and Ye, Z.- H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 56, 195- 214. doi: I0.l093/pcp/pcul 40

Zuluaga, A., P Vega- Arreguin, J. C., Fe Z., Matas, A J., Patev, S., Fry, W. E., et al. (2016). Analysis of the tomato leaf transcriptome during successive hemibiotrophic stages of a compatible interaction with the oomycete pathogen Phytophthora infestans. *Mol Plant Pathol.* 17, 42- 54. doi:10. ll ll /mpp.12260

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.