ABSTRACT

Title of Dissertation: Genome-wide identification and analysis of imprinted genes in strawberry seed development

Dirk Joldersma, Doctor of Philosophy, 2022

Dissertation directed by: Professor Zhongchi Liu, CBMG

The activation of zygotic gene expression is of fundamental importance to reproductive biology, but its regulation remains poorly understood. Within Angiosperm plants, fertilization occurs simultaneously in two locations, the embryo and its genetic twin, the endosperm, a nutritive tissue that is a defining feature of Angiosperm reproduction. Auxin hormone synthesized in the endosperm is essential to seed and fruit development. In the diploid strawberry *Fragaria vesca*, that auxin synthesis is regulated by FveAGL62, which is expressed specifically after fertilization in endosperm. How fertilization activates *FveAGL62* expression in the endosperm, however, is presently unknown.

I investigated the hypothesis that epigenetically regulated maternally- and paternally expressed genes (MEGs and PEGs, and together, "imprinted genes") regulate the expression of *FveAGL62*. I hybridized two *F. vesca* accessions, isolated the endosperm from the F1 seeds, and sequenced the transcriptome of the F1 endosperm—a result facilitated by strawberry's uniquely accessible seed. To identify imprinted genes within the endosperm, I assembled and annotated the genome of the maternal parent, *F. vesca* accession "Yellow Wonder" (FvYW5AF7), a model for the commercial strawberry. The paternal parent genome was obtained from a collaborator.

809 PEGs and 825 MEGs were identified from RNA sequencing reads that align uniquely to the maternal or paternal genome. MEGs are enriched in genes catabolizing auxin and hence limit seed growth, while PEGs are enriched in genes involved in histone modification, thereby promoting cell differentiation and seed growth. The distinct roles of MEGs and PEGs supports and can be explained by parental conflict and kinship theories, which predict a maternal genome tends to restrict progeny consumption of maternal resources, while a paternal genome will encourage such consumption. In contrast to findings in other species, I find that the endosperm-specific auxin biosynthetic gene *FveYUC10* is maternally expressed, but while its imprinting status has changed, it may still function as a fertilization sensor.

The maternally expressed gene *FveMYB98* contains a binding domain that targets motifs present in *FveAGL62's* promoter and its homolog binds *AtAGL62* promoter in *Arabidopsis.* With collaborators, I showed that overexpression and CRISPR knockout of *FveMYB98* changes seed size. Transient expression, yeast one hybrid and quantitative PCR analyses suggest that *FveMYB98* represses *FveYUC10* expression directly and *FveAGL62* expression indirectly. These results suggest that *FveMYB98* expression is a vehicle for maternal regulation of the level of auxin in the endosperm and thereby endosperm proliferation and seed size.

My dissertation research has produced a new genome assembly of a model strawberry, a transcriptome of strawberry endosperm, and identified imprinted genes at genomic scale. I find *FveMYB98* regulates seed size—a function echoed broadly within MEG and PEG classes—providing supporting evidence for the parental conflict theory within the developing progeny. These results improve our understanding of zygotic expression in developing seeds, addressing a fundamental scientific gap and, more tangibly, may enable future production of fertilization-independent seeds and seedless fruits.

GENOME-WIDE IDENTIFICATION AND ANALYSIS OF IMPRINTED GENES IN STRAWBERRY SEED DEVELOPMENT

by

Dirk Joldersma

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Zhongchi Liu, Chair
Professor Jose Feijo
Professor Eric Haag
Professor Stephen Mount
Professor Lisa Taneyhill

Dedication

To Lisa, with love, hope and gratitude.

Acknowledgements

Collaborations with Patrick Edger of Michigan State University and Winston Timp of Johns Hopkins University and their students and lab members played a critical role in the completion of this dissertation and are gratefully acknowledged.

I would like to thank Dr. Muzi Li for discussion and advice on the assembly and annotation of the "Yellow Wonder" genome.

The University of Maryland (UMD) generously supported my research via:

- A Faculty-Student Research Collaboration Grant.
- Support provided in association with UMD's NIH T32 Molecular and Cell Biology Training Grant.
- A Hockmeyer summer fellowship and two Hockmeyer travel awards.
- A dissertation fellowship provided by the graduate program in Biological Sciences.

I appreciate early training in laboratory techniques received at Georgetown University Medical Center under the direction of Dr. Robert P. Yasuda. I acknowledge with gratitude the teaching of Dr. Harmut Doebel of George Washington University, through whose eyes and words I first noticed that biology could be simultaneously fun and systematic.

Finally, I thank my mentor, Dr. Zhongchi Liu, who has encouraged me to pursue my scientific interests, recognized and cultivated strengths in me known and unknown, and asked consistently and positively that I stretch a little closer to my potential.

Table of Contents

Chapter 1. Introduction.

In 2010, over 50 percent of American calories consumed were derived directly from grain, fruit, plant oils and seeds (Rehkamp, 2016), and that contribution is closer to two-thirds of all calories globally (Powell, 2018). Despite its importance, the mechanism through which the acts of fertilization—cellular, nuclear, and genomic fusion of paternal pollen and the maternal gametophyte—lead development of the seed and fruit that provide this food remains unknown. In this chapter, I review literature relevant to my thesis—which investigates the genetic control of post-fertilization seed development—and introduce the major hypothesis I have investigated.

1.1 Female gametophyte and double fertilization

In Angiosperm plants, seeds arise from the female gametophyte (FG) after fertilization by the male pollen. Seeds consist of three distinct tissues: the embryonic progeny plant that develops from the

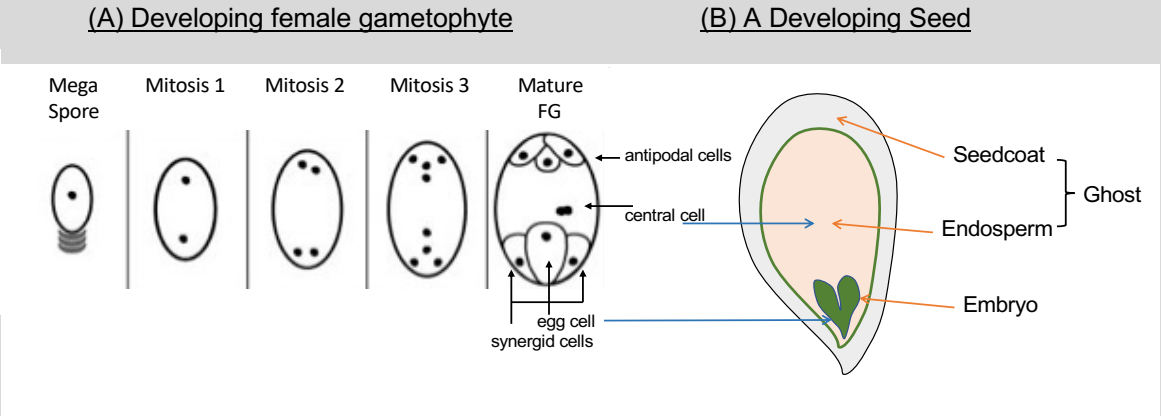


Figure 1.1. Diagrams of female gametophyte development to an early-stage seed. **(A)** Development of haploid megaspore to mature female gamete (FG)**.** Note fission of two nuclei to form diploid central cell; remaining gametic cells are haploid. (B) After fertilization, FG undergoes zygotic development. Fertilized central cell develops into triploid endosperm, while the fertilized egg cell develops into diploid embryo. Seedcoat originates from nonsexual maternal tissues surrounding the FG. (A) adapted from (Yadegari and Drews, 2004)**.**

fertilized egg cell, the seed coat derived from maternal tissues surrounding the FG, and the endosperm—the product of a second fertilization of the central cell (See Figure 1.1).

The complexity of the eight-celled FG is reflected in the distinct functions of the central cell, synergid cells and egg cell. As discussed above, the central cell and egg cell are respectively fertilized by the two sperms in a single pollen—the so-called double fertilization characteristic of Angiosperm plants. While the fertilized egg cell undergoes zygotic development to form an embryo, the fertilized central cell becomes the endosperm, a nutritious tissue for the embryo. The function of the three haploid antipodal cells is largely undescribed. In contrast, the two synergid cells that flank the egg cell are critical to the function of the FG. Laser ablation of synergid cells in the FG caused its failure in attracting pollen (Higashiyama et al., 2001). In Arabidopsis an R2-R3 MYB transcription factor MYB98 is expressed in synergid cells (Figure 1.2A-B) and plants homozygous for mutations in *MYB98* are incapable of successful fertilization despite the

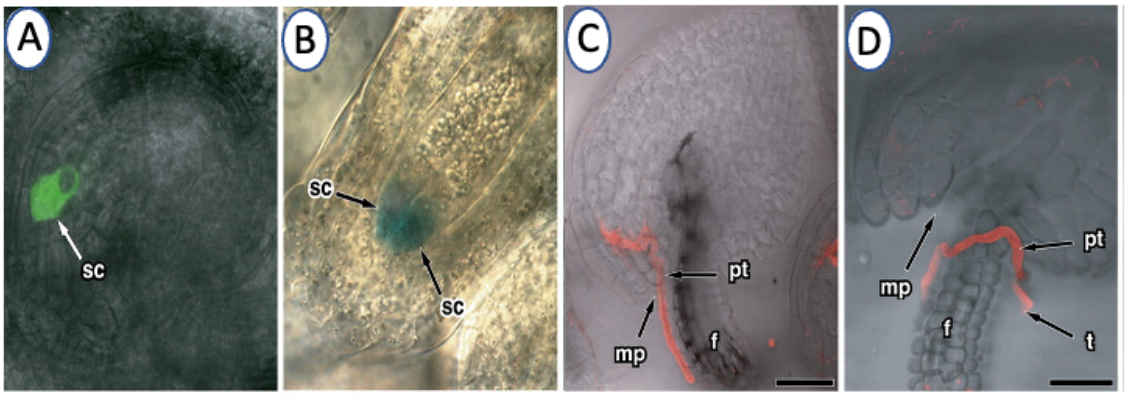


Figure 1.2. MYB98 is necessary for synergid cell function in *A. thaliana.* (A-D) Expression of *proMYB98::GFP* in developing female gametophyte. (C) Pollen tube successfully reaches egg cell in WT ovule. (D) Pollen tube fails to reach an *myb98* gametophyte. sc, synergid cell; mp, micropyle; pt, pollen tube; f, funiculus. (Kasahara et al, 2005)

apparently successful development of the synergid cells themselves—localizing the *myb98* defect to the function of the mature synergid cell (Kasahara et al., 2005). *myb98* ovules do attract pollen tubes, but they are subject to polyspermy and pollen tubes can fail to progress past the synergid cell or release sperm cells near egg cells (Figure 1.2D). While *MYB98* transcripts are observed in the maturing seed, the gene's post-fertilization function remains uncharacterized.

1.2 Endosperm, a post-fertilization tissue that functions as a fertilization sensor and nutrient provider

Sometimes compared to the mammalian placenta, the major function of the triploid endosperm is to mediate the exchange of nutrients and other necessities between the maternal tissue and the developing embryo (Figueiredo et al., 2015). In this capacity, the endosperm acts as a storage organ and is the site of a large reserve of energy in the form of lipids and starches (Miray et al., 2021), largely explaining the importance of the seed to human nutrition, and making manipulation of the endosperm's storage capacities a field of long-standing agricultural and scientific interest (Yan et al., 2014). A fundamental taxonomical division within Angiosperms is the separation of flowering plants into two clades, monocots and dicots—divided by the presence of one or two leaf-like cotyledons upon germination, an embryonic morphology that is present shortly after fertilization (Zhao et al., 2017). The fate of the endosperm differs in the two clades as well. Monocots—wheat, rice, and maize, for example—retain the endosperm through seed development, and thus grains like these contribute the bulk of human calories. In dicots, however, the endosperm frequently is an ephemeral tissue (again, like the placenta) that, after multiplying to several cell layers after fertilization, ultimately is consumed—generally leaving only a single cell layer remaining when a dicot seed is fully matured (Müller et al., 2006). In

both cases, however, the essential commonality is that the fertilized endosperm fails to develop as an independent organism.

As such, the endosperm's contributions to the development of the embryo should be considered an investment in its genetic twin's survival, and the evolution of the endosperm is a defining trait of Angiosperm plants (Friedman, 1995). The necessity of the endosperm's successful function was definitively established by endosperm-specific expression of diphtheria toxin, which killed endosperm cells, leading to the cessation of seed expansion (Weijers et al., 2003). In seeds lacking endosperm, embryo can begin development upon fertilization, but they showed immediate developmental abnormalities, including the loss of normal cell division planarity. No seeds lacking endosperm were capable of germination.

This altruistic function of the endosperm extends beyond nutritive exchange and mechanical stability to include hormone synthesis, the transport and control of which affect the development of both embryo and seedcoat. In *Arabidopsis,* auxin exported from the endosperm is critical to successful seedcoat initiation; auxin accumulates in the integuments after fertilization, defects in endosperm-expressed auxin biosynthesis genes lead to altered seedcoat development and seed abortion, and transgenic lines producing auxin ectopically in the central cell initiate seed coat development without fertilization (Figueiredo et al., 2016). Piskurewicz et al. showed that key effector genes of the gibberellic acid (GA) and abscisic acid (ABA) hormone pathways, *RGL2* and *ABI5*, respectively, are expressed in the endosperm (Piskurewicz et al., 2008). Rgl2 is a transcription factor that catalyzes transcription of GA-responsive genes in the presence of GA. *rgl2* mutants are insensitive to GA and thus fail to germinate—a failure ultimately due to a failure to pass GA signaling from the endosperm to the testa, the successor tissue to the seedcoat. Similarly, ABA-induced, endosperm-expressed *ABI5* also represses germination, intimately

linking the endosperm transcriptome to effects of hormone signaling in the seedcoat (Piskurewicz et al., 2008). As such, the endosperm functions as an essential mediator of nutrient exchange and regulatory information.

1.3 The role of the phytohormone auxin in early seed development

Auxin was the first plant hormone identified, when its action catalyzing organismal growth was found to explain Darwin's observation that grass grows toward a source of light (Enders and Strader, 2015). Auxin's role in other plant tissues was soon investigated. In the 1930's, Gustafson investigated the role of auxin in plant reproductive tissues. Gustafson observed that

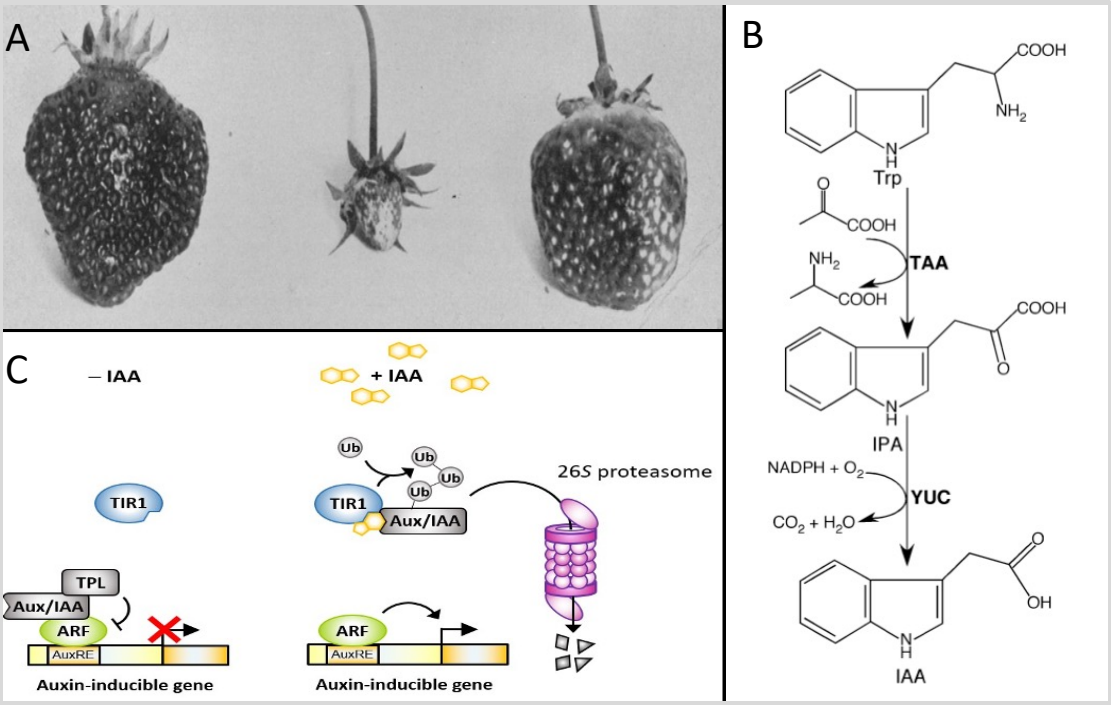


Figure 1.3. Auxin function and signaling cascade.
(A) Normal ripened strawberry (left). A lack of fruit growth due to the removal of seeds (center). A seedless fruit treated with auxin that restored fruit growth (right). (B) The major auxin synthesis pathway. Tryptophan is converted by *TAA* and *YUC* family enzymes into active auxin IAA. (C) Diagrams showing auxin signaling mechanism. In the absence of auxin, AUX/IAA proteins inhibit ARF proteins from interacting with the AuxRE enhance element. When auxin is present, its receptor TIR1 binds and ubiquitinates Aux/IAA proteins, releasing ARF proteins to activate downstream genes. A: (Nitsch, 1950); B: (Zhao, 2012); C: (Kuhn et al, 2020).

some orange fruits are naturally seedless, while others contain seeds. He discovered that the seedless orange fruit develop from floral buds that contain elevated levels of auxin (Gustafson, 1936), indicating that auxin is important to initiating the fruit program of oranges and leading to the hypothesis that seeds may be the source of auxin in the natural, seeded oranges (Gustafson, 1939). This hypothesis was confirmed in the 1950's by Nitsch, who found that the application of auxin to seedless strawberry fruit restored "wild type" fruit growth (Figure 1.3A; Nitsch, 1950), suggesting that auxin was synthesized in the seed post-fertilization and required for initiating fruit and seed development.

Auxin is synthesized primarily from the amino acid tryptophan in two steps by flavin mono-oxygenase enzymes encoded by the *TAA/TAR* and *YUCCA (YUC)* genes (Figure 1.3B; Stepanova et al., 2011; Won et al., 2011). In the presence of auxin, TIR proteins—the receptor for auxin—bind Aux/IAA repressor proteins, which leads to the ubiquitination and proteolysis of Aux/IAA proteins and the derepression of Auxin Response Factor (ARF) transcription factors that activate or inhibit downstream genes (Figure 1.3C). Auxin molecules are actively transported to target cells via PIN efflux carrier proteins—establishing a polar gradient in the tissue or organ (Leyser, 2005, 2010).

The endosperm is the site of the auxin synthesis that catalyzes seed and fruit growth. Early evidence was found in the existence of a class of *A. thaliana* mutants, *Fertilization Independent Seed* (*FIS*) mutants (described in next section), in which seed development could proceed despite the absence of pollination (Ohad et al., 1996). In these mutants, endosperm proliferation and changes in seed coat morphology are observed, despite the lack of fertilization or development of the egg cell (Ohad et al., 1996). This result highlighted that mutant *FIS* proteins belonging to the Polycomb Repressive Complex 2 (PRC2) fail to repress central cells from developing into

endosperms in the absence of fertilization. However, these autonomous seeds are not viable due to a lack of embryo. Microarray data from laser-dissected endosperm tissue indicated auxin biosynthetic genes *YUC10, YUC11,* and *TAA1* are substantially upregulated in endosperm tissues following fertilization in *Arabidopsis thaliana,* and RNA-seq of strawberry ghost tissue (endosperm plus seedcoat) also showed upregulation of auxin biosynthesis genes (Kang et al., 2013; Belmonte et al., 2013). Finally, *FvAGL62*—a gene with endosperm-specific expression and localization in strawberry—recently was shown to regulate auxin biosynthesis, confirming the endosperm as the site of not only auxin biosynthesis but also the site of the regulation of auxin biosynthesis (Guo et al., 2022).

Together, these observations suggest that endosperm-derived auxin biosynthesis can be considered an indicator of fertilization success and the commitment of the fertilized tissues and surrounding maternal tissues to develop into seed and fruit to protect the embryo and disseminate the seeds to ensure reproductive success.

1.4 Autonomous seed development from epigenetic mutations

Autonomous endosperm and seed development in the absence of fertilization was observed in *fis* class mutants. Several *fis* mutations were recovered from mutational screens seeking suppressors of the homeotic *Pistallata* gene, which are male sterile (Chaudhury et al., 1997). The first gene cloned in this class was *MEDEA* (MEA), which encodes a SET domain protein homologous to the *Drosophila* gene, Enhancer of zeste, a member of the Polycomb Repressive Complex 2 (PRC2) that was known to regulate homeotic genes in the fly via trimethylation of histone 3 residue 27 (Grossniklaus et al., 1998; Laugesen et al., 2016). Three further genes with *fis*-like phenotypes were identified; these mutant genes comprised alleles of *MSI1*, *FIS2*, and *FIE* (Guitton et al., 2004; Luo et al., 1999; Ohad et al., 1999). Together, *MEDEA*, *MSI1*, *FIE* and

FIS2 are members of four gene families that code for the four core subunits of Arabidopsis's PRC2 (de Lucas et al., 2016), with the FIS class genes' phenotypic similarity explained by their participation and essentiality to a single protein complex for chromatin regulation.

That role was broadened beyond the PRC2 complex with the description of the gene, *DEMETER*, a DNA glycosolase that removes DNA methylation (Choi et al., 2002; Gehring et al., 2006). *DEMETER* is also a maternal-affect gene and null mutations of the gene leads to seed

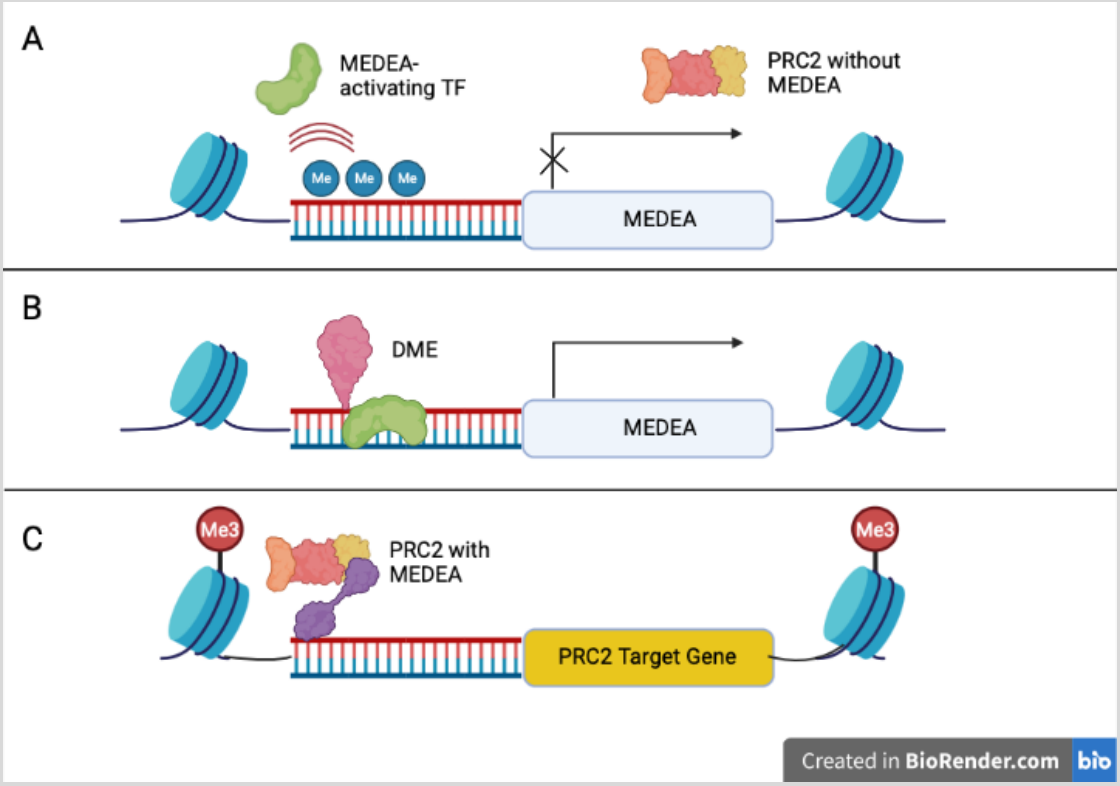


Figure 1.4. Demeter regulates Medea expression in the central cell prior to fertilization. (A) In leaf tissues and in the central cell prior to expression of DME, the promoter of *Medea* is methylated and resistant to transcriptional activation. In central cell, MEDEA protein is necessary for function of PRC2. (B) DME protein, expressed exclusively in the central cell prior to fertilization, demethylates the promoters of maternally inherited *Medea,* enabling expression of MEDEA protein. (C) With addition of MEDEA protein, PRC2 is functional and trimethylates histone tails adjacent to target genes (including male copies of *Medea* present after fertilization). (Cartoon summarizes findings from Choi et al., 2002, and Gehring et al., 2006.)

abortion with 100 percent penetrance. Despite its seed phenotype, *DEMETER* is expressed exclusively prior to fertilization (Choi et al., 2002). *DEMETER's* expression in the central cell is required for expression of *MEDEA* in the central cell and the functioning of the PRC2 complex in the central cell and endosperm (Figure 1.4). Thus, this study showed that gametophyte-specific demethylation of the *MEDEA* locus is required for successful seed and endosperm development, linking histone and DNA methylation to seed and endosperm development. Together, the descriptions of *DEMETER* and the FIS class genes highlight the essential role of parental imprinting and epigenetic regulation in seed development, and the gametophytic expression of these genes also introduces the prominent role of imprinting of maternal gene expression especially.

1.5 Genetic imprinting and kinship theory

Maternal-effect genes can arise from regulation that limits gene expression to the female gamete, as is the case with *DEMETER*. However, maternal-effect genes like *MEDEA* also are expressed zygotically, suggesting that the paternal genome's influence can be limited by other processes. These parent-of-origin effects are called gene imprinting and arise from differential epigenetic regulation of a locus on the maternally- and paternally-inherited genomes (Peters, 2014). Imprinting's relevance to reproductive development was established when Solter and Surani demonstrated that two copies of either maternal or paternal chromosomes were not sufficient to allow embryonic development (McGrath and Solter, 1984; Surani et al., 1984), showing that inherited maternal and paternal chromosomes played functionally distinct and necessary roles. Different expression of the two parental alleles mediates parent-of-origin effects, and as such, genes are characterized as maternally-expressed genes or paternally-expressed genes (MEGs and

PEGs respectively). These differences are established and maintained by several biochemical processes, including methylation of cytosine bases and certain histone residues (Jenuwein and Allis, 2001; Satyaki and Gehring, 2017). Observed most easily in binary, "on-or-off" MEGs or PEGs, it soon came to be appreciated that functional implications also can arise from more subtle changes in the ratio between maternal/paternal expressed transcripts (Patten et al., 2014).

Endosperm tissues are known to be the major site within plants of genetic imprinting, or parent-dependent patterns of gene expression (Gehring and Satyaki, 2017). Imprinted expression patterns have been observed only in Angiosperms and mammals and in the mammalian placenta and the Angiosperm endosperm (Figueiredo et al., 2015; Satyaki and Gehring, 2017). Both endosperm and placenta are fertilized tissues that function as an interface between mother and progeny. Both tissues are terminal tissues, and neither is essential to the future of mother or progeny. Thus each tissue may allow some dysregulation of the genome, including transposons that ordinarily are silenced, which in turn may enable expression of genes that also are silenced in somatic tissues (Köhler and Weinhofer-Molisch, 2010).

The parental conflict theory (Haig and Westoby, 1989) seeks to explain the presence and evolutionary function of imprinting in the endosperm and also is invoked to explain imprinting in the placenta. The theory posits that, while the genetic interest of both the father and progeny is to maximize maternal resource investment in progeny survival, the maternal genome faces a more complex calculation that will seek to moderate the progeny's consumption of maternal resources, as these resources alternatively could be invested in maternal survival. An alternative theory—kinship theory—arises from the observation that Angiosperms double fertilization yields one and only one surviving progeny (Friedman, 1995). Under these circumstances, given that endosperm and embryo are genetically identical, there is a shared interest to outcompete potential half

siblings found in a plant that receives pollens from multiple fathers. Thus, the evolution of the altruistic endosperm maximizes embryonic survival (Friedman, 1995). Such kinship conflict further predicts, given the mother is indifferent among her progeny, that the paternal genome will seek to maximize altruistic behavior of the endosperm (Friedman, 1995). Both parental conflict and kinship conflict theories predict that paternally expressed genes tend to promote metabolism and growth of the progeny, whereas maternally expressed genes tend to restrict growth of the progeny. In humans, evidence for the parental conflict theory is found in the imprinting of the growth factor gene *IGV2*, which is only paternally expressed, and *H19*, which restricts the accumulation of glycogen in the placenta and is maternally expressed (Esquiliano et al., 2009; Giannoukakis et al., 1993).

In plants, auxin biosynthetic genes have been identified as PEGs, which is consistent with the predictions of the paternal conflict theory. In 2011, six genome-wide, RNA-seq based surveys for parental imprinting in endosperm were conducted (3 in Arabidopsis, 2 in maize, and 1 in rice) (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011). *YUC10* was identified as a PEG by each of the three Arabidopsis surveys, and a flavin mono-oxygenase (like *YUC10*) was found as a PEG in rice and one of the two maize studies. *TAA1* was found a PEG in one Arabidopsis study and both maize surveys. These findings raised the possibility that paternally supplied auxin biosynthetic genes not only serve to promote progeny growth but also serve to inform successful fertilization by the sperm and hence act as a fertilization sensor.

1.6 Fertilization sensors

The importance of the auxin signal to seed development and the observation that only male alleles of certain auxin biosynthetic genes are expressed in the zygotic endosperm raises the

possibility that transcription of these paternally expressed auxin synthesis genes is a fertilization sensor, in the sense that it enables a zygotic fate commitment upon the satisfaction of the fertilization requirement. However, many models of fertilization sensing can be hypothesized in other studies.

The simplest genetic model for fertilization sensing is the presence or absence of a single gene product. For example, in rice RNA transcripts of *BabyBoom1* (*BBM1*), an AP2 domain transcription factor, are present in the pollen but not in ovule (Khanday et al., 2019). Immediately post-fertilization, *BBM1* is exclusively paternally expressed (Figure 1.5). Through an apparent autoactivating loop, *BBM1* eventually is expressed biallelically in the zygote (Khanday et al., 2019). Thus, *BBM1* appears to be a fertilization sensor, and indeed, CRISPR-induced loss-of-function of *BBM1* and two closely related homologs led to loss of seed viability

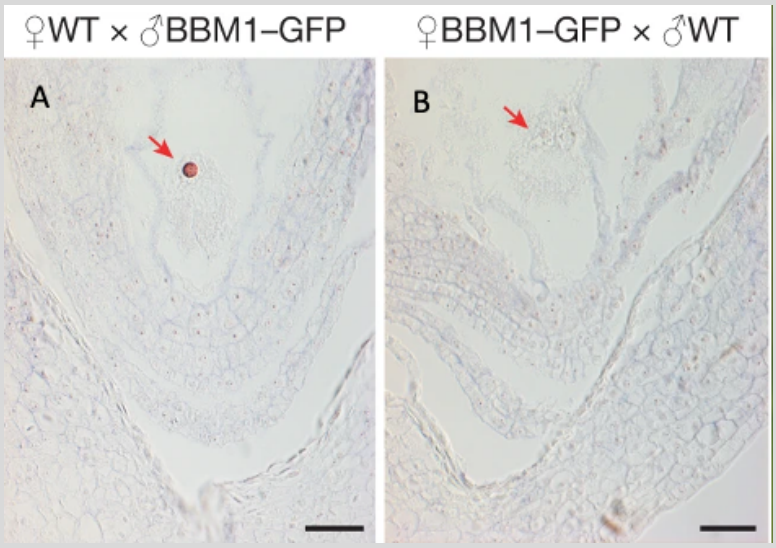


Figure 1.5. Male but not female allele of BBM is expressed after fertilization. Expression of BBM1-GFP detected at 2.5 hours after pollination by GFP-sensitive antibody staining in (A) with paternally-supplied GFP but not in (B), where GFP is maternally supplied. Red arrows point to zygote nuclei. Scale bars: 25 um. (Khanday et al., 2019)

in rice, while ectopic pre-fertilization expression in the egg cell leads to parthenocarpic, or unfertilized, seed development (Khanday et al., 2019).

A more complex system of fertilization sensing could involve two genes, and such systems are present in unicellular *Saccromyces cerevesie* and *Clamydomonas reinhardtii*. In yeast, a product of the MAT-alpha locus (later characterized as MATALPHA2) represses a suite of haploid a-type-specific genes (Miller et al., 1985). However, after "fertilization" by an a-type cell via mating, MATALPHA2 dimerizes with a product of the MAT-a locus to bind to different genetic loci, repressing both a-type and alpha-type (haploid) gene expression and inducing diploid (zygotic) expression (Dranginis, 1990). BLASTP of MATALPHA2 protein sequence against the *Arabidopsis* proteome suggests similarity to BELL1-type homeodomain transcription factors (Supplemental Figure 1.1). In unicellular alga *Chlamydomonas,* GSP1, another BELL1-type homeodomain protein, accumulates only in the cytosol of *plus* cells (Kurvari et al., 1998), while the homeodomain protein GSM1 accumulates exclusively in the cytosol of *minus* gametes

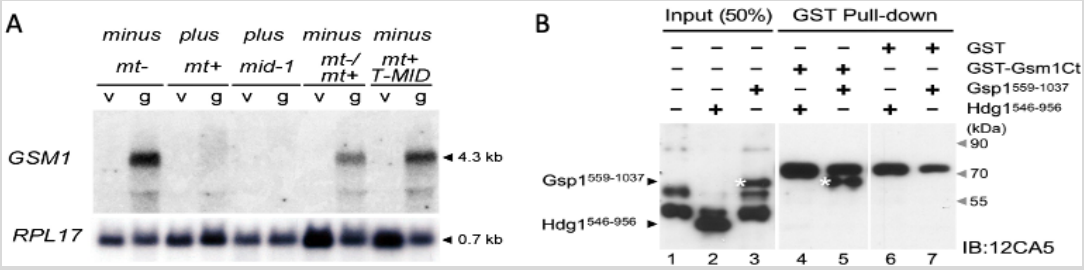


Figure 1.6. In *Clamydomonas reinhardtii* a two-gene fertilization sensing mechanism activates zygotic gene expression. (A) GSM1 is expressed in *minus* gametes. Northern blot of cells that differ in mating specificity (top row, *plus* or *minus*) and/or carry different alleles of *MT* loci (second row), probed with a *GSM1* cDNA fragment. g, gametes; v, vegetative cells. (B) Coprecipitation of GSP1 with Gsm1. In vitro-expressed GSM1$^{512–934}$ attached to glutathione S-transferase (GST-GSM1Ct) was used to pull down in vitro translated HA-tagged HDG1$^{546–956}$ or GSP1$^{559–1037}$. Translated and pulled-down products were analyzed by western blotting with anti-HA antibody. Asterisks indicate GSP1-specific signals. Interacts also shown with three proteins present in the wheat-germ extract (lanes 1–3) and with one protein present in the *E. coli* extract (lanes 4–7). Adapted from Lee et al., 2008.

(Figure 1.6A)(Lee et al., 2008). Upon fertilization, the two proteins heterodimerize (Figure 1.6B), translocate to the nucleus, and activate zygote-specific gene expression (Lee et al., 2008). While GSP1 is a BELL1-type transcription factor, GSM1 is a member of the closely-related KNOX homeodomain transcription factor family (Hisanaga et al., 2021), suggesting that KNOX-BELL interactions may be a conserved pathway to activate zygotic gene expression.

In higher plants, however, KNOX-BELL genes do not appear to regulate the zygotic transition, but rather are associated with meristem maintenance and cell cycle control (Hay and Tsiantis, 2010). The need to regulate a zygotic transition remains even so. How is this transition restrained prior to fertilization, and what fertilization sensor enables the gamete to zygote transition? One possibility is that this function is played by a similar two gene system that consists of genes from outside the KNOX-BELL families. *BabyBoom1*-related findings suggest a second hypothesis that a single gene (and its paralogs) may control this transition, and that redundancy in the gene family has allowed invisibility despite intensive genetic screening. Spatial-temporal restrictions on expression could be established by genetic regulation, but we are reminded such restrictions also can be imposed by genetic imprinting, as seen in the paternal expression of *BBM1* in rice, the imprinted expression of *DEMETER/*FIS genes, and that of auxin biosynthetic genes. However, it is presently unknown the mechanism of this transition in strawberry or other dicot plants.

1.7 *AGL62*: a fertilization reporter and auxin biosynthesis regulator in strawberry and *Arabidopsis.*

AGL62 is a MADS-box transcription factor—a class of transcription factors found in all eukaryotes which have radiated to become a handful of genes in humans, fruit flies and yeast and

more than 100 in higher plants (Gramzow et al., 2010). MADS-box genes are separated into two clades: 1) SRF-like or Type 1 MADS-box proteins, included among which is MCM1 (the 'M' in MADS box), a cofactor required for MATALPHA1-mediated repression of a-type genes (Elble and Tye, 1991), and 2) MEF2- or MIKC-type or Type 2 MADS box genes which have a second, conserved K domain (Gramzow et al., 2010). In *Arabidopsis,* MIKC-type MADS-box genes were extensively studied and shown to function as tetramers, many of which were discovered due to their specification of floral organ identity (Honma and Goto, 2001; Pelaz et al., 2000). In addition to flower development, MIKC-type MADS-box genes have been shown to play diverse functions in plant development, including for example control of vernalization in cereals and fruit ripening in tomato (Trevaskis et al., 2003; Vrebalov et al., 2002). However, Type I MADS box genes are not well studied.

In *A. thaliana,* a Type I MADS box gene *AGL62* functions in early seed development. *AGL62* is expressed in the antipodal cells of the female gametophyte and in the early endosperm and has a seed-lethal phenotype (Kang et al., 2008). Wild type endosperm undergoes early mitotic divisions within a syncytium for approximately five days after fertilization and subsequently undergoes cellularization. In loss-of-function *agl62* mutants, early endosperm cellularization leads to reductions in seed size, as in the aptly named *miniseed* mutants where early cellularization reduces seed size by half (Luo et al., 2005). *agl62* seeds begin cellularization within 24 hours of pollination, which limits endosperm nuclei proliferation (Figueiredo et al., 2016). In *fie* mutants, a defective PRC2 complex led to autonomous endosperm proliferation in the absence of fertilization (Figueiredo et al., 2016). However, the effect of *fie* is suppressed in *fie agl62* double mutants (Figueiredo et al., 2016), suggesting that *AGL62* is required to mediate the autonomous endosperm proliferation phenotype of *fie* mutants. This result suggests that

PRC2 complex may normally repress *AGL62* to prevent autonomous endosperm proliferation without fertilization. However, how *AGL62* promotes endosperm nuclei proliferation and represses precocious cellularization is unknown. Prior studies showed that *AGL62* positively regulates the expression of an ABCB-type transporter in the endosperm, which may be required for auxin transport to the seedcoat to stimulate seed coat development (Figueiredo et al., 2016).

Endosperm cellularization defects are a classic hybridization barrier, and endosperm fertilized by unreduced paternal gametes fail to cellularize (Scott et al., 1998). These endosperm also produce an excess of auxin, while loss-of-function auxin biosynthesis mutants exhibit increased viability despite the presence of unreduced paternal gametes (Batista et al., 2019). These findings linked cellularization defects to dysregulation of auxin biosynthesis, but the regulation of auxin biosynthesis in the endosperm remained unknown until recently, when it was found that *AGL62* and its strawberry homolog *FveAGL62* are a key regulator of fertilization-induced auxin synthesis (Guo et al., 2022).

In diploid strawberry, *Fragaria vesc*a, a model fruit crop studied in our lab, *FveAGL62* is expressed specifically after fertilization (Guo et al., 2022). CRISPR-mediated knockout of *fveagl62* led to nonviable seeds and failed fruit development (Figure 1.7A). *fveagl62* seeds exhibit premature cellularization (Figure 1.7B). Real time qPCR shows that *fveagl62* seeds exhibit reduced transcripts of auxin biosynthetic genes *YUC10, YUC11,* and *TAA1* (Figure 1.7C). Finally, the authors show that the premature cellularization phenotype is rescued by the provision of exogenous auxin (Figure 1.7D), establishing a link between *FveAGL62* and downstream auxin-related phenotypes and suggesting that *FveAGL62* is required for auxin biosynthesis gene expression in seeds. In *Arabidopsis,* the authors show that expression of the *YUC10* auxin biosynthesis gene is reduced and that *agl62* mutant seeds accumulate less auxin than wildtype

seeds, showing that *AGL62* plays a conserved positive role in the regulation of auxin biosynthesis in the seeds of these species (Guo et al., 2022).

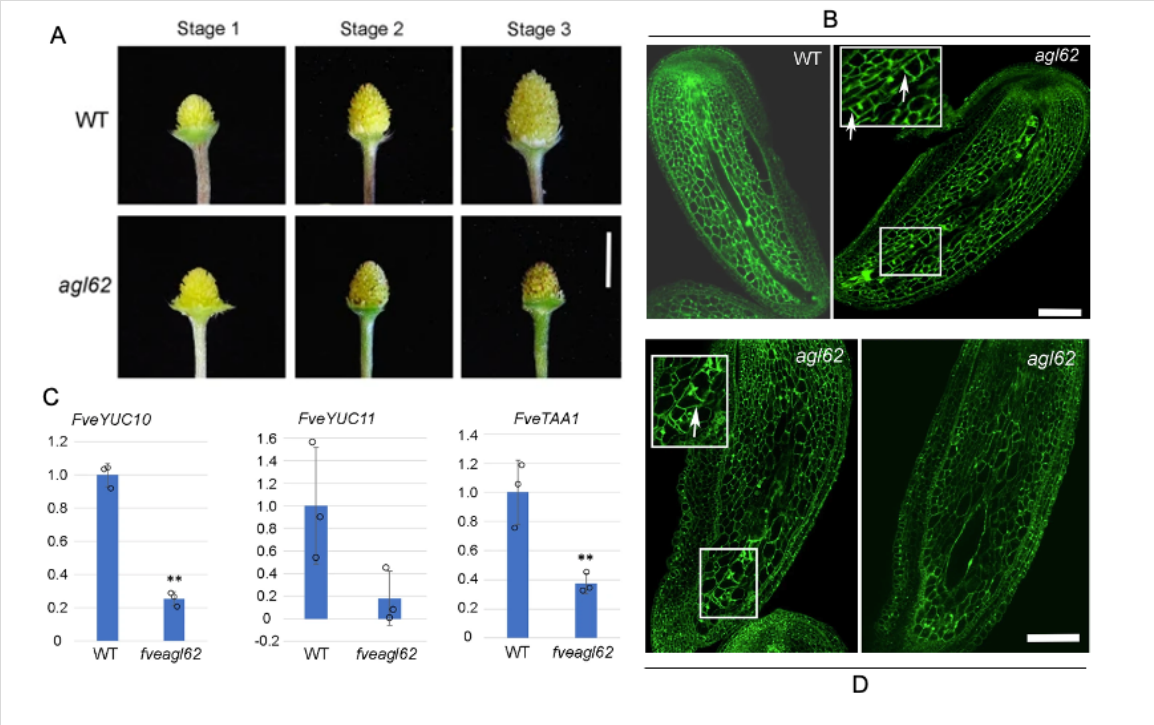


Figure 1.7. *F. vesca agl62* mutant seeds exhibit early cellularization and auxin dysregulation. (A) Receptacle fruit development in WT and *agl62* mutant strawberry at three earliest stages of fruit development. (B) Confocal laser scanning microscopy (CLSM) of WT and *fveagl62* seeds at stage 2 showing precocious cellularization of *fveagl62* endosperm (arrows) (C) RT-qPCR of three auxin biosynthesis genes in WT and *fveagl62* stage 2 seeds. (D) CLSM of *fveagl62* stage 2 seeds mock- or auxin-treated. Scale bars: 1 mm (A), 100 um (B,D). (Guo et al., 2022)

In strawberry seeds, the authors further established that FveAGL62 regulates auxin biosynthesis indirectly via *ATHB* genes, which encode transcription factors with both a zinc-finger and homeodomain. Dual luciferase experiments show that expression of *FveATHB* is reduced in tobacco leaves in the presence of FveAGL62, indicating that FveAGL62 may repress *ATHB* expression (Guo et al., 2022). At the same time, overexpression of the *FveATHB* genes leads to reduced expression of the auxin biosynthetic genes and premature cellularization (Guo et al., 2022).

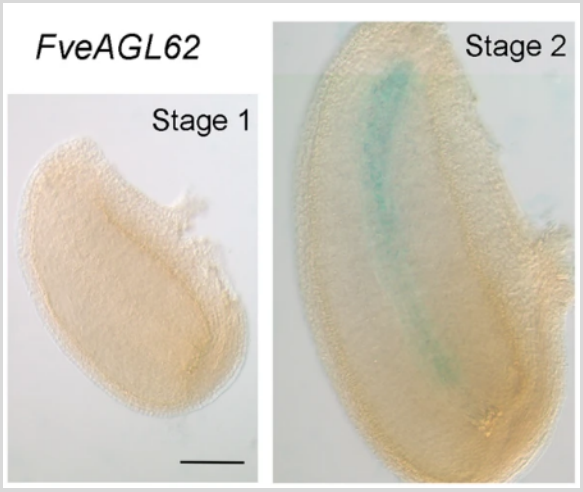


Figure 1.8. FveAGL62 is a fertilization sensor. Expression of reporter gene *ProFveAGL62::FveAGL62-GUS.* Stage 1 is pre-fertilization ovule; stage 2 is the earliest stage post-fertilization. Scale bar:0.12 mm. (Guo et al., 2022)

Establishing FveAGL62 as a regulator of auxin biosynthesis also establishes the gene as a fertilization readout that connects successful fertilization to *FveAGL62* activation. Further, *FveAGL62* is connected to subsequent endosperm proliferation and seed development. Considering the literature on auxin's role in endosperm proliferation and early seed development as well as the imprinting of auxin biosynthesis genes in several species, *FveAGL62* is the gene

acting furthest upstream of the post-fertilization auxin production identified to date, making it a potential candidate fertilization sensor. However, its expression is limited to the initial period after fertilization (Figure 1.8), suggesting that *FveAGL62* is itself regulated and subject to an upstream activator after fertilization.

1.8 Imprinted genes as fertilization sensors and regulators of AGL62-mediated auxin biosynthesis

While we do not know how fertilization induces the development and differentiation of the seed, we have identified a fertilization-induced key regulator of seed development, *FveAGL62*, whose function is conserved in Arabidopsis and strawberry and who can serve as the reporter of fertilization. Strawberry has a history of serving as a model to study how fertilization induces seed and fruit development due to its unique fruit structure in which hundreds of seed-containing ovaries (achenes) develop externally. These achenes are easily accessible for dissection and manipulation. Further, diploid strawberry is genetically tractable (Zhou et al., 2018), extensive resources on seed morphology and transcriptome are currently available (Hollender et al., 2012, 2014; Kang et al., 2013), and recently *FveAGL62* has been established as a fertilization reporter (Guo et al, 2022). As such, *Fragaria vesca* offers an ideal system in which to address the fundamental question of how fertilization is perceived to activate seed development. This dissertation tests the hypothesis that imprinted genes are the primary and initial effectors of fertilization and that *FveAGL62* is the major target of these effectors.

A function for imprinted genes as fertilization sensors is consistent with a gross description of fertilization as the union of parental genomes and explains the necessity of a paternal genomic contribution in the initiation of zygotic life. Thus, one could imagine a simple one gene

fertilization sensing system in strawberry that operates much like rice in which transcription of a paternally imprinted gene regulates expression of *FveAGL62* (Figure 1.9A). An alternative hypothesis could recall the two-gene mating systems of yeast or *Chlamydomonas.* One potential model is that the protein products of a PEG gene and a MEG gene are united upon fertilization, and their heteromerization activates *FvAGL62* (Figure 1.9B). A closely related model could reflect a PEG sequestering a MEG, leading to derepression of *FveAGL62*.

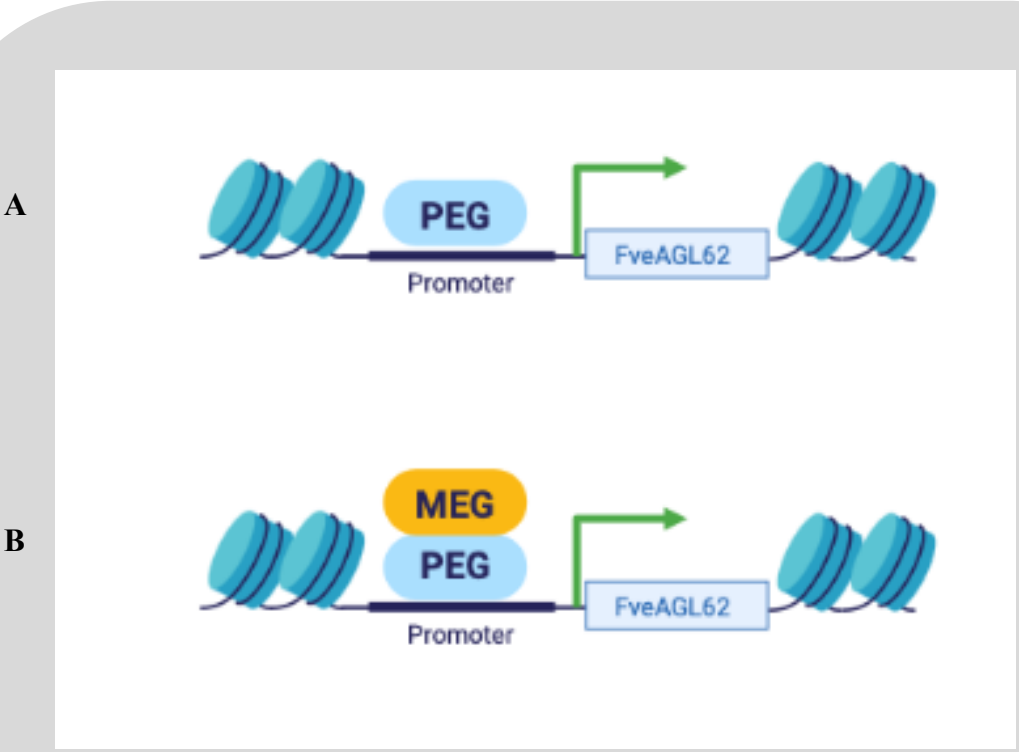


Figure 1.9. Models of fertilization sensing. (A) *A simple, single-gene fertilization sensor:* a PEG initiates transcription of *FveAGL62* in the fertilized endosperm. (B) *A two-gene fertilization sensor:* A PEG transcribed from the paternally inherited chromosome heterodimerizes with a MEG from the paternal genome, and their heterodimer binds *FveAGL62* promoter to activate its expression in the post-fertilization endosperm.

It is entirely possible, of course, that the presence or absence of gene products is governed by conventional genetic regulation, rather than or in conjunction with epigenetic gene imprinting. Thus, a gametophytic or maternal gene product could replace a MEG in the hypotheses above, or a paternal product could be transported with the sperm cells into the fertilized ovule or central cell. A mechanism that does not rely on imprinting seems more likely to explain gametophytic expression of a MEG-like protein prior to fertilization, as the size of the paternal gamete may favor transcriptional activation over pre-fertilization expression of the appropriate PEG-like protein. Thus, the central hypothesis of this thesis is that imprinted genes activate the post-fertilization expression of *FveAGL62*, as illustrated in Figure 1.9 above.

1.9 Conclusion and summary

In CHAPTER ONE, I discussed a central question of fertilization – the activation (or derepression) of zygotic gene expression. I have highlighted the essential role of the Angiosperm endosperm in coordinating between maternal and embryonic tissues. Endosperm-synthesized auxin hormone is essential to seed development, including embryonic development, seedcoat development, and fruit development. I focused on our lab's recent discovery that *FveAGL62* is a conserved activator of auxin biosynthesis in the endosperm soon after fertilization. I proposed several hypotheses regarding how fertilization is sensed and subsequently activates *FveAGL62* expression in the endosperm. First, epigenetic regulation is likely essential to this developmental process as it is to others, and the transient nature of the endosperm increases the likelihood that genetic imprinting may be a component of this epigenetic regulation. Further, I proposed that imprinted MEG and PEG proteins dimerize to activate expression of the fertilization marker gene *FveAGL62*.

In CHAPTER TWO, I report the assembly and annotation of the genome of a new F. vesca accession, “Yellow Wonder” (YW5AF7), a common laboratory model for the octoploid commercial strawberry, as its lack of runner is well suited to growth chamber conditions and its recessive white fruit color provides an excellent marker for genetic crossing. In addition to broader benefits to the research community, this genome assembly was a requirement for the identification of imprinted genes in strawberry as described in Chapter 4.

In CHAPTER THREE, I report the generation and analysis of the first available transcriptome of the *F. vesca* endosperm. This transcriptome was collected from hand-dissected endosperm tissue. By contrasting the RNA-seq results of these endosperm samples with existing resources that are not endosperm-specific, I was able to establish a high confidence endosperm transcriptome. I find that *FveAGL80*, an FveAGL62 cofactor that is known to be expressed exclusively in the endosperm (Guo et al., 2022) is in fact upregulated eightfold in the endosperm transcriptome I established, validating the approach taken to identify endosperm-enriched expression. I also find that genes primarily expressed in the endosperm are enriched on cellular developmental processes, including DNA replication, gene expression, translation and ribosomal assembly, and cell fate specification.

In CHAPTER FOUR, I identify imprinted genes in the endosperm tissue. Through sequence-specific alignment to the paternal (CFRA502) as well as maternal (YW5AF7) genomes of F. vesca, I identified candidate MEGs and PEGs from the transcriptome of the intraspecies hybridization between YW5AF7 and Fragaria vesca bracteata ecotype ‘CFRA502’. This analysis is the first identification of imprinted genes in strawberry and possibly in the commercially important Rosaceae family. Several developmental processes appear to be regulated by imprinted genes. I position these findings in the context of parental and kinship conflict, observing, for

example, that maternally expressed genes seek to reduce auxin levels in the endosperm, while paternally expressed genes simultaneously enhance the transcriptional sensitivity of the tissue to auxin. Intriguingly, I find that the endosperm-specific auxin biosynthetic gene YUC10 is subject to imprinting and is maternally expressed in strawberry, suggesting a functional divergence from other species. I also gather external in silico and in vitro resources with my data on imprinted genes to identify potential imprinted regulators of *FveAGL62* expression, highlighting the MEG MYB98 as one of the three such genes.

In CHAPTER FIVE, I examine the functional role of FveMYB98 in strawberry seed development. Overexpression and knockout of the maternally expressed FvMYB98 leads to changes in seed size that are consistent with the predictions of the parental conflict theory and suggest that FveMYB98 acts as a negative regulator of seed size. qPCR indicates that overexpression of *FveMYB98* reduces expression of *FveAGL62*, suggesting a possible connection to auxin biosynthesis, however, a yeast one hybrid assay indicates that FveMYB98 does not regulate *FveAGL62* directly. Luciferase assays performed by my lab member Lei Guo show that FveMYB98 is a negative regulator of auxin biosynthetic gene YUC10. Together, these findings suggest a mechanism explaining *fvemyb98's* seed size phenotype—a mechanism wherein Myb98 protein represses *FveYUC10*, titrating the level of auxin biosynthesis in the endosperm and thereby reducing *FveAGL62* expression.

Taken together, the results of this dissertation research provide important insights into the regulation of endosperm and seed development in strawberry. The research has provided a new genome assembly of a wild strawberry model, a transcriptome of strawberry endosperm, and genome-scale identification of imprinted genes within that transcriptome, which will hopefully catalyze continued study into the development and function of the strawberry seed development

and fertilization sensors. I have identified three potential fertilization sensors. I have found that FveMYB98 regulates *FveAGL62,* possibly through a newly discovered FveMYB98 function as a repressor of auxin biosynthetic gene *FveYUC10*. The research project contributes to broad literatures on seed and reproductive development, including hormone regulation and gene imprinting. The project has uncovered many examples that support the predictions that interparental and kinship conflicts have shaped the evolution of the transcriptional profile of the endosperm. Finally, and more tangibly, a better understanding of the mechanisms that enable post-fertilization activation of the seed development program may eventually enable commercially viable fertilization-independent seed development, which in turn may allow the production of seedless fruit, the conservation of high value germplasms, and increased resilience of plant reproduction to climate change.

Chapter 2. Assembly and annotation of Fragaria vesca 'Yellow Wonder' genome

2.1 Introduction

Thanks to its small size, diploidy, rapid life cycle, and the availability of genetic tools, woodland strawberry (*Fragaria vesca*) is the preeminent model system for the cultivated strawberry *Fragaria x ananassa*, which is allooctoploid. It also serves as a model for the Rosaceae family of fruit crops, which include apple, pear, peach, raspberry, and others. Several accessions of *F. vesca* have been inbred seven times and developed as models; they are Hawaii 4 (H4), Yellow Wonder, and Rügen (Hawkins et al., 2016; Alger et al., 2021; Slovin et al., 2009). The genome of Hawaii 4 was first published in 2011 (Shulaev et al., 2011) and later in 2018 with a much improved quality in both assembly and annotation (Edger et al., 2018; Tennessen et al., 2014; Li et al., 2019). Recently, a new genome was published of *F. vesca* accession CFRA2339 (Alger et al., 2021), a red-fruited, runnerless variety that is less commonly used in research. The 4th Hawaii 4 annotation and assembly continues to serve as the standard reference genome for genomic studies in strawberry.

However, Hawaii 4 (H4) is just one of the inbred *F. vesca* accessions and may not capture the diversity of F. vesca (Hawkins et al., 2016; Alger et al., 2021; Slovin et al., 2009). Most importantly, Yellow Wonder 5AF7 (FvYW5AF7) has been a preferred research model in many laboratories because of several of its characteristics. First, FvYW5AF7 plants do not form runners, allowing for convenient growth and maintenance in high-density settings like a lab growth chamber (Figure 2.1A-B). This feature also makes it easier in mutagenesis screens without contamination of different mutant individuals by runners. Second, the yellow fruit, the

product of a recessive mutation in the *MYB10* gene (Hawkins et al., 2016), provides a convenient visual marker in distinguishing hybrid F1 from self-fertilized progeny in genetic crosses between FvYW5AF7 female and red-fruited WT accessions (Figure 2.1C, D), and for transient expression assays to study fruit pigment genes and ripening processes (Luo et al., 2018; Hawkins et al., 2016). Third, FvYW5AF7 is ever-flowering, making it convenient to perform genetic crosses and study the development of flower, seed, and fruit at all seasons (Slovin et al., 2009). Finally, FvYW5AF7 was heavily inbred, leading to an improved genetic homozygosity (Hawkins et al., 2016).

Figure 2.1. Plant architecture and fruit phenotype of three *F. vesca* accessions Yellow Wonder (5AF7), Hawaii 4 (H4), and Rügen (Rü). (A-C) Photographs of FVYW, H4, and Rü plants. Note the absence of runners in YW and Rü, but prolific runners (arrows) in H4. (D) Both YW and H4 develop yellow color fruits as shown, but Rü develops red color fruit. Plants are pictured in 4 inch by 4 inch pots.

A survey of literature illustrates the widespread use of FvYW5AF7 in studies ranging from mutagenesis screens and mutant characterization (Caruana et al., 2018; Luo et al., 2018), genome editing (Zhou et al., 2018), transcriptome profiling (Kang et al., 2013; Shahan et al., 2018;

Hawkins et al., 2016), and gene functional characterization (Hawkins et al., 2016; Zhou et al., 2021). However, FvYW5AF7 does not currently have its own reference genome as most studies in *F. vesca* rely on the H4 reference genome. The lack of a true FvYW5AF7 reference increases the difficulties in primer design, gene cloning, genome editing and analysis in FvYW5AF7 due to SNPs and structural variations between FvYW5AF7 and the H4 genomes. Therefore, the lack of a high-quality genome of the Yellow Wonder 5AF7 accession is an impediment to research in this model plant.

I have assembled, annotated, and published the FvYW5AF7 genome as FvYW Version 1.0 (FvYW1.0)(Joldersma et al., 2022). The genome was assembled de novo using Oxford Nanopore long reads, and polished with Illumina reads. To annotate the genome, I lifted over the annotation from H4 v4.0.a2 (Li et al., 2019). FvYW1.0 exhibits high quality, with a BUSCO score of 97%—reflecting the identification of the majority of a conserved set of plant genes—and an N50 of 34MB, indicating chromosome-scale assembly. I further examined and confirmed the molecular lesions in the loci determining the loss of runnering, ever flowering, and yellow fruit color.

This project was completed in a collaboration with Norah Sadowski and Dr. Winston Timp of the Department of Biomedical Engineering at Johns Hopkins University, who extracted DNA from FvYW5AF7 and sequenced the DNA on their Nanopore PromethION.

2.2 Results

2.2.1 Assembly and annotation

To assemble the FvYW5AF7 genome, I combined long-read Oxford Nanopore sequencing and high coverage short read Illumina sequencing. About 2.3 million Nanopore reads, totaling 26

gigabytes (GB) and ~125x coverage of the genome, were provided by Sadowski and Timp. Illumina reads were previously published (Hawkins et al., 2016) and downloaded from NCBI GenBank.

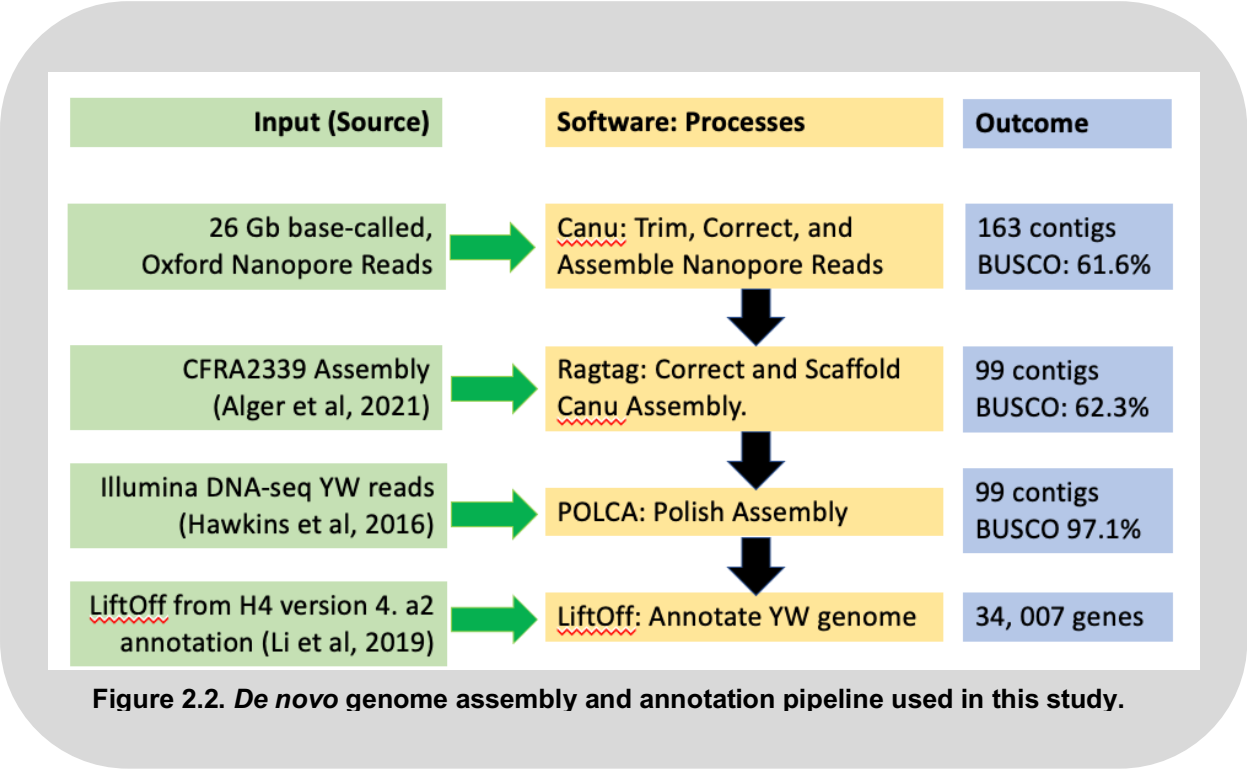


Figure 2.2. *De novo* genome assembly and annotation pipeline used in this study.

To assemble the FvYW5AF7 genome, I followed a pipeline illustrated and presented along with major interim results in Figure 2.2. First, the Canu assembler (Koren et al., 2017) was used to trim and correct the raw Nanopore reads and then assemble a preliminary genome. This initial assembly yielded 163 contigs, with an N50 of 5.6 MB. RagTag was then used to correct and scaffold misassemblies, using homology-based alignments to realign contigs against existing *F. vesca* reference genomes and merge scaffolds into pseudomolecules (Alonge et al., 2019; Alger et al., 2021). I compared the RagTag result of aligning the YW assembly to the H4 (v4.0) reference genome with that of aligning to the CFRA2339 genome (Table 2.1); scaffolding to the

CRFA2339 genome yielded superior statistics in terms of continuity (7kb gap sequence) but slightly reduced N50 (32.8M vs 33.3M). Hence, I selected the CFRA2339 assembly as the final RagTag reference.

Table 2.1. Comparison of RagTag results used to scaffold the initial FvYW5AF7 assembly against two earlier *Fragaria vesca* assemblies.

	YW vs H4 (v4.0)	YW vs CFRA2339
Placed sequences	77	77
Placed bp	212.6Mb	212.7Mb
Unplaced sequences	86	86
Unplaced bp	3.1Mb	2.9Mb
Gap bp	1.4Mb	0.0Mb
Gam sequences	67	67
# of contigs	96	96
Genome size	217.1Mb	215.7Mb
auN	31.1Mb	30.8Mb
N50	33.3Mb	32.8Mb

POLCA, a subprogram of the MASURCA genome assembly software (Zimin and Salzberg, 2020), was used to polish the assembly over three rounds with 11.6 GB of previously-published Illumina DNA-seq data of YW (Hawkins et al., 2016). The final genome assembly of FvYW5AF7 spanned 219.5 MB across 99 contigs with an N50 length of 33.6 MB (Figure 2.2; see also Table 2.2 below). Fragaria vesca has seven chromosomes. In the final assembly, 216 MB of the 219.5 MB total sequence are contained in seven FvYW5AF7 pseudomolecules.

LiftOff (Shumate and Salzberg, 2020) was used to port the high-quality H4 genome annotation (v4.0.a2) (Li et al., 2019) to the FvYW5AF7 genome assembly that resulted in a final annotation that annotates 34,007 genes. The number of FvYW5AF7 genes is very similar to H4 with 34,008 annotated genes. All but 62 genes were localized to the 7 chromosomes, the remaining 62 genes are in the assembly's remaining unanchored 92 contigs. Of the 10 most overrepresented biological process gene ontologies (GOs) related to those 62 genes, six of them contribute to

ribosome assembly and peptide translation (Supplementary Table 1). This result likely reflects the repeated nature of ribosomal genes, which causes difficulty in assigning them to unique loci.

In the annotations, FvYW5AF7 gene names are derived from their LiftOff-identified H4 syntenic homologs. Thus, in the FvYWv1.0.a1 annotation, the MYB10 gene has a gene ID of FvYW_1g22020, which correlates with its homolog in H4, FvH4_1g22020 from the FvH4 annotation v4.0.a2.

2.2.2 Assessment of assembly and annotation quality

Table 2.2. Quality metrics for FvYW_v1.0 and previously published *F. vesca* genomes

Accession	Hawaii-4 (FvH4_v4.0)	CFRA2339	Yellow Wonder (FvYW_v1.0)
Assembly Length (MB)	220	244	220
Total Contigs	29	402	96
auN (MB)	32	31	32
N50 (MB)	34	31	34
L90 (MB)	7	7	7
Number of annotated genes	34,008	30,349	34,007
BUSCO	0.947	0.967	0.961

Table 2.2 summarizes the key characteristics of the YW genome in comparison to the two previously published F. vesca genomes, H4 and CFRA2339. Key indicators such as total contig number, auN, N50, and BUSCO score indicate that FvYWv1.0 is of a quality similar to or slightly better than the previously published genomes. The Benchmarking Universal Single-Copy Orthologs (BUSCO) with the "plantae" database was used to estimate the quality of the assembly and annotation (Simão et al., 2015). The FvYW5AF7 genome was found to have 96% of the core genes in the BUSCO plantae dataset, supporting a high-quality genome assembly and annotation, with 22 duplicated, 10 fragmented and 22 missing BUSCOs of the 956 searched.

2.2.3 Consistency of FvYW1.0 with previous F. vesca genomes

FvYW5AF7 exhibits certain plant architecture and fruit characteristics that have been previously explained to result from DNA variations in specific genes. I examined the new assembly to confirm that these molecular variations underlie the phenotypes of YW (Figure 2.3). Specifically, in the FvYW5AF7 assembly, *MYB10* (FvYW_1g22020) is found to possess the G-to-C SNP that underlies the tryptophan to serine change in the gene's 12th amino acid that leads to production of yellow colored fruit (Figure 2.1D) (Hawkins et al., 2016). Second, the FvYW5AF7 assembly harbors a 9-bp deletion in the *GA20OX4* gene (FvYW_2g35050), which underlies YW's runnerless phenotype (Figure 2.1B) (Tenreira et al., 2017). Finally, the *TFL1* (FvYW_3g24700) aligns with 100 percent sequence identity to FvH4_3g24700, indicating that the gene of FvYW5AF7 harbors the same 2-bp deletion responsible for the perpetual flowering phenotype found in H4 and other previously-identified perpetual flowering varieties (Iwata et al., 2012; Koskela et al., 2012). The recapitulation of these DNA variations in FvYWv1.0 strengthens previous findings and confirms the quality of this assembly.

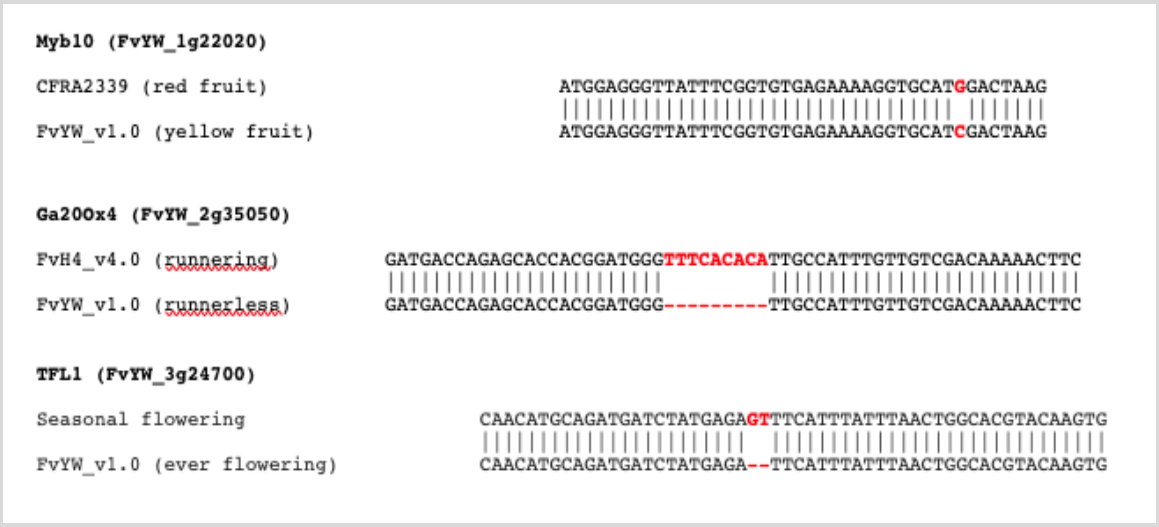


Figure 2.3. FvYW_v1.0 *de novo* assembly corroborates previous phenotype-linked genetic lesions. Blast results showing the alignment in the region of interest. Bold, red nucleotides highlight the genetic variants underlying specific phenotypes.

2.2.4 Comparison between FvYW5AF7 genome and previously published F. vesca genomes

The FvYW5AF7 genome is relatively distinct from previously published F. vesca genomes of H4 and CFRA2339 accessions. I used Sourmash (Brown and Irber, 2016) to calculate the Jaccard similarity coefficients across the three *F. vesca* genomes, using *Fragaria iinumae* as an outgroup (Figure 2.4A). This comparison reveals that the CRFA2339 and H4 genomes are more similar to each other than they are to the FvYW5AF7 genome, although the three F. vesca genomes cluster strongly from the outgroup. This phylogeny also is supported by synteny maps, which show that the FvYW5AF7 accession contains major inversions in three loci in comparison to the genomes of Hawaii 4 or CFRA2339 (Figure 2.4B). The inversions are distal to the centromere in chromosomes 1 and 3 and near the midpoint of the length of chromosome 3. One such inversion (midpoint of chromosome 3) contains the TFL1 (FvYW_3g24700), showing that the SNP

polymorphism responsible for the perpetual flowering phenotype is robust to a genetic inversion. As such, synteny and similarity comparisons demonstrate that FvYW5AF7, at a genetic level, is more distantly related to Hawaii 4 than CFRA2339 (Figure 2.4B).

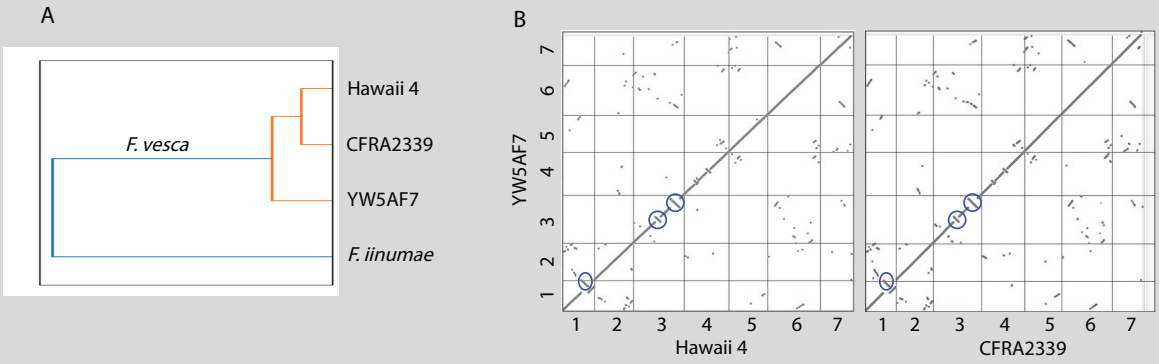


Figure 2.4. FVYW genome structure shows consistent differences from previously published genomes of *Fragaria vesca* accessions. (A) FVYW genome clusters are distinct from previously published *F. vesca* genomes, Hawaii 4 and CFRA2339, as well as *Fragaria iinumae*. This is indicated by Jaccard similarity coefficient calculated by Sourmash. (B) Inversions are observed in chromosomes 1, 3, and 4 between FVYW and the other two *F. vesca* accessions (highlighted in blue circles). Y and X axis show seven chromosomes. Synteny plots generated on COGE.

2.3 Discussion and conclusions

This genome assembly will further improve FvYW5AF7 as a desirable model and facilitate genetic and genomic research in *F. vesca* and commercial strawberry at large. I have assembled the FvYW genome, which is relatively distinct from previously assembled *F. vesca* genomes. Genome quality appears comparable, and the assembly recapitulates previous molecular genetic findings, despite an inversion in the genome in one case.

Chapter 3. Generation and analysis of strawberry endosperm transcriptome

3.1 Introduction

As discussed in Chapter 1, the endosperm is an altruistic product of double fertilization that serves to nurture the developing embryo and coordinate seed development and nutrient exchange with the maternal plant (Friedman, 1995), and thus, the endosperm is essential to seed development to viability (Weijers et al., 2003). Specifically, the endosperm is the site of auxin biosynthesis that is essential to subsequent seed and fruit development, and that biosynthesis is regulated by endosperm-specific transcription of *FveAGL62* (Guo et al., 2022).

As sequencing technology has advanced, the tissue's function has been has been interrogated at the transcriptomic level in *Arabidopsis thaliana* and major food cereals (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; Zhang et al., 2011; Wolff et al., 2011; Belmonte et al., 2013). Unsurprisingly, the endosperm exhibits substantial complexity through time and space, including sub-regionalization of the endosperm that is not attempted in this study (Belmonte et al, 2013). Perhaps the most basic measure of that complexity is that well over half of genomic loci are expressed in this transient tissue (Belmonte et al., 2013; Picard et al., 2021).

Despite the complexity, however, transcriptomic data has served to advance understanding of this critical tissue, both by ratifying previous hypotheses and suggesting unexpected functions. For example, Belmonte et al. supported previous observations regarding the function of endosperm cellularization in seed when they found cytokinesis upregulated in the endosperm, while Picard et al. observed that endosperm nuclei overexpress defense genes, consistent with an immunity-like function being played by a tissue at the convergence of progeny and mother

(Belmonte et al., 2013; Picard et al., 2021). To date, the endosperm-specific transcriptome of strawberry is unavailable, preventing similar insights into seed and fruit development. Kang et al. provided a comprehensive transcriptomic analysis of developing strawberry flower and fruit, publishing 50 unique transcriptomes from 25 tissue and time combinations (Kang et al., 2013), as seen in the eFP for *FveAGL62* reproduced below (Figure 3.1). However, Kang et al. sequenced the ghost tissue of the seed, which consists of the seed coat and endosperm together, as they encountered significant practical difficulty in separating the tissues.

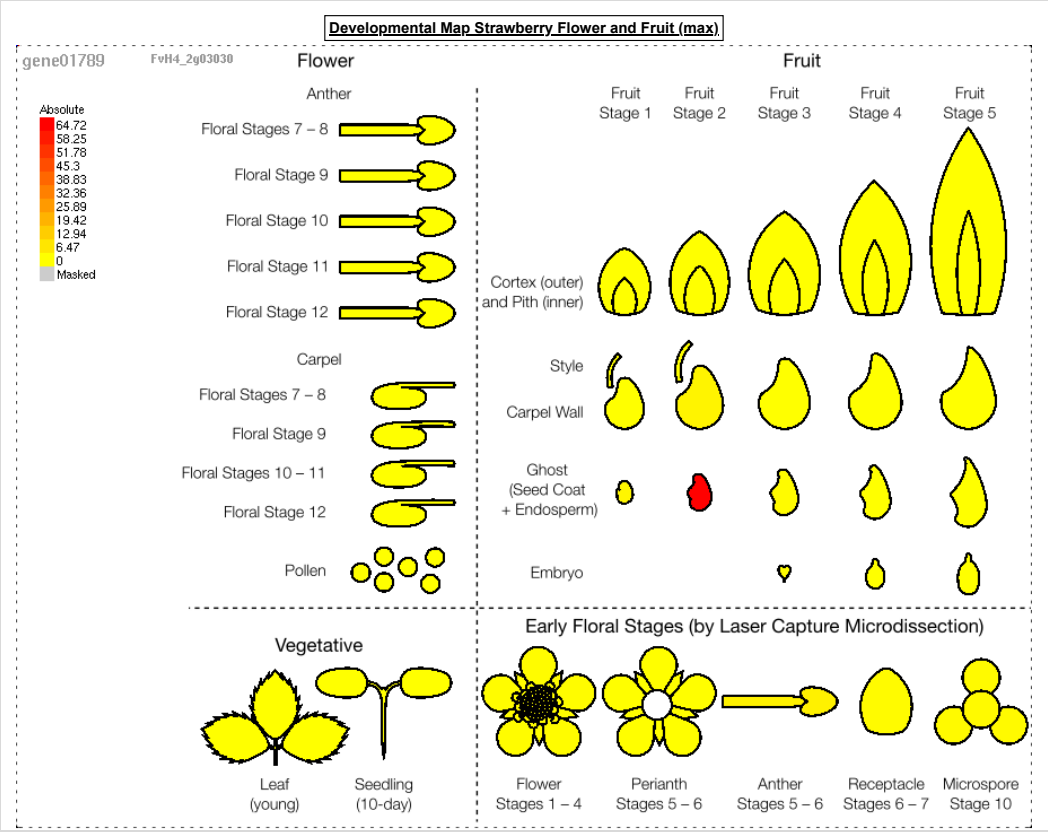


Figure 3.1. RNA-seq shows *FveAGL62* has a precise domain of expression.
eFP result from bar.utoronto.ca.

Given the importance of the endosperm transcriptome to the hypothesis that endosperm-specific expression of imprinted genes serves a fertilization sensor, obtaining an endosperm specific transcriptome is necessary. In this chapter, I report the generation of such a transcriptome. This process leverages careful sample collection with the substantial existing resources to identify endosperm-specific expression. I verify the transcriptome through comparison with existing endosperm marker genes in strawberry and Arabidopsis. I find that the strawberry transcriptome conserves the complexity of seed transcriptomes observed previously and contains gene expression patterns that both recapitulate and diverge from previous work.

3.2 Results

3.2.1 Endosperm tissue collection and RNA-seq

To generate the strawberry endosperm transcriptome, I dissected and isolated endosperm 7-9 days after a cross between the pollen of *Fragaria vesca bracteata* ecotype "CFRA502" and stigma of *Fragaria vesca* 'Yellow Wonder' 5AF7 (FvYW5AF7). The FvYW5AF7 plant chosen harbors a mutation in the strawberry *LEAFY* gene and is male sterile, which eliminates self-pollination (see Figure 3.2). I crossed subspecies to enable later analysis in the parent-of-origin studies described in Chapter 4. From each strawberry "fruit" derived from the cross-pollinated flower, 10 to 30 achenes were hand dissected and their endosperm tissue was isolated by excluding the embryo and seedcoat.

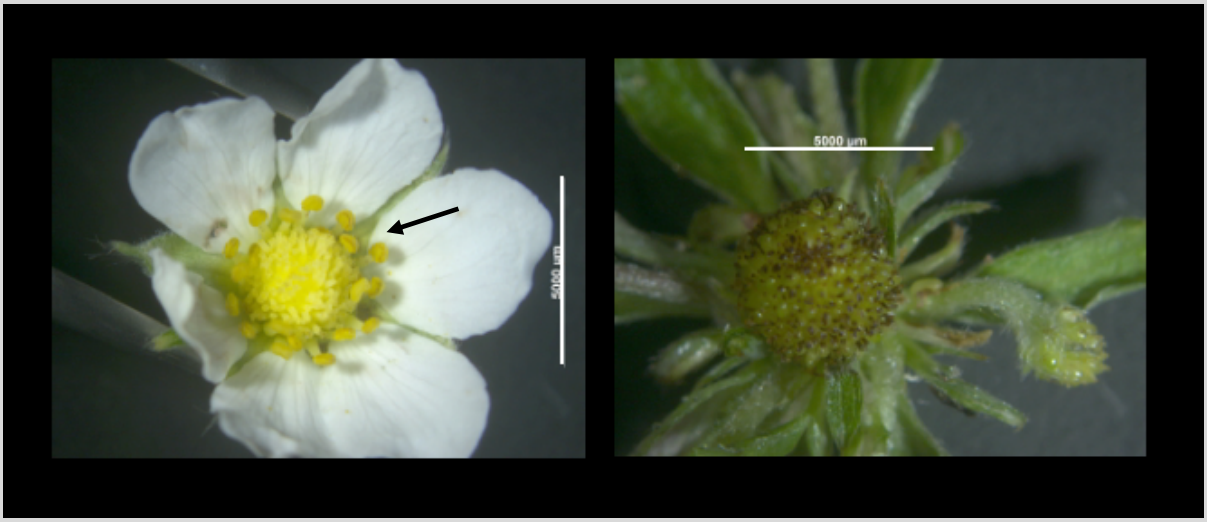


Figure 3.1. *fveleafy* flower is male sterile. Wild type and *fveleafy* FvYW5AF7 flowers. An anther highlighted by arrow in wild type. Note the lack of anthers and floral indeterminacy exhibited by *fveleafy* flower, from which several new florets are growing. Scale bars: 5 mm.

Total RNA was extracted from the endosperms and three biological replicates were sent for RNA-seq using an ultra-low input Illumina platform to generate paired-end, 150 base pair length reads. Between 21.8-36.3 gigabases of sequencing data were obtained per sample, and 90% of nucleotide calls had a quality score more than 30.

These raw reads were trimmed using Fast-P software on default settings (Chen et al., 2018), and I used STAR, a splice-aware aligner (Dobin et al., 2013), to align 77 percent of reads to a unique locus in the reference H4 genome. This revealed a comprehensive putative endosperm transcriptome mapping at least one transcript to 28,531 annotated loci in at least one of the sample tissues (and at least 20 transcripts to 24,719 loci).

3.2.2 Bioinformatic analyses to identify endosperm-enriched genes from strawberry

I leveraged previously-published strawberry seed transcriptomes from pre-fertilization ovule and post-fertilization ghost (endosperm and seedcoat) tissues (Kang et al., 2013) to identify genes predominantly (or specifically) expressed in the endosperm (at the postfertilization stages). See Figure 3.3 for an illustration of the analytical scheme. The analytical steps were:

(1) Because endosperm development takes place after fertilization, I performed differential expression analysis using DESeq2 (Love et al., 2014) to identify 2,969 genes upregulated at least two fold in the post-fertilization stage 3 ghost or post-fertilization stage 2 seed when compared with ovules at the pre-fertilization (stage 1), defining these genes as post-fertilization upregulated genes.

(2) Because endosperm development takes place within the broader context of the ghost, I again applied DESeq2 to identify two-fold upregulated genes in ghost tissues versus non-ghost samples (leaf, achene wall, and fruit cortex and pith), defining these 2,563 genes as ghost-enriched genes.

(3) 1,535 genes are found in the intersection of the post-fertilization and ghost-enriched sets; I define such genes as ghost-enriched, post-fertilization genes.

(4) A set of 3,314 genes were upregulated at least twofold in the endosperm-only samples collected in this project when compared with the stage 3 ghost samples collected previously (stage 3 is comparable to the stages I collected my endosperm samples at 7-9 DPA); these genes I defined as endosperm-enriched genes (Supplemental Data 3.1).

(5) Finally, the intersection between the ghost-enriched, post-fertilization genes and endosperm-enriched genes led to a set of 225 genes as the endosperm-enriched, fertilization-induced genes (Supplemental Data 3.2).

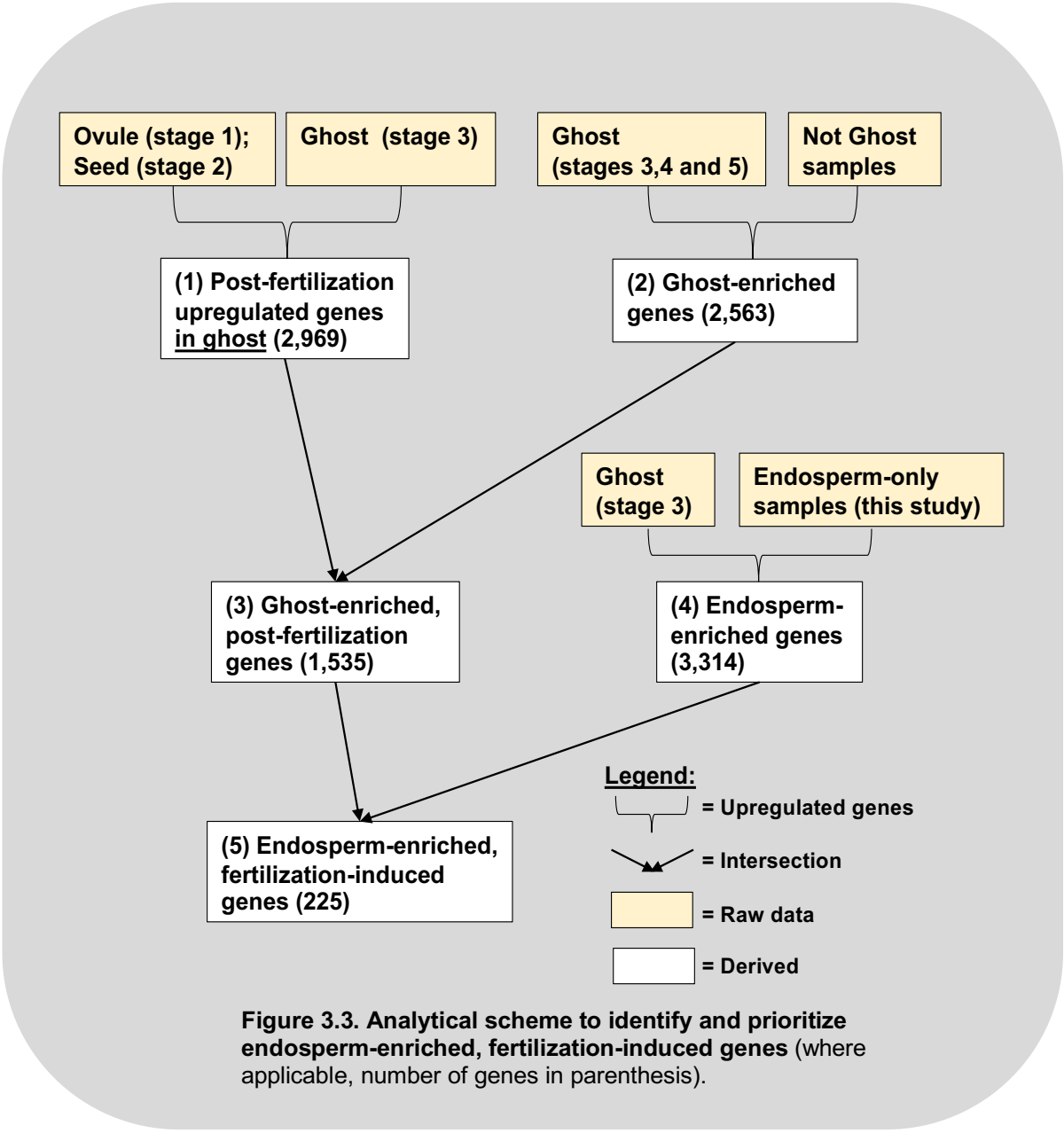


Figure 3.3. Analytical scheme to identify and prioritize endosperm-enriched, fertilization-induced genes (where applicable, number of genes in parenthesis).

3.2.3 Validation of endosperm transcriptome

Two strategies were used to corroborate the identification of endosperm-enriched and fertilization-induced genes (Figure 3.4). First, the endosperm-enriched gene set (Supplemental Data 3.1) was found to contains the strawberry *FveAGL80* (FvH4_6g21170) gene, which is upregulated tenfold in the endosperm-only samples versus the ghost samples that contain seedcoat and endosperm tissues (Figure 3.4A). In both *Arabidopsis* and *Fragaria vesca*, *AGL80* shows endosperm-specific expression (Guo et al., 2022; Picard et al., 2021) (Figure 3.4B). *Arabidopsis thaliana* homologs to *Fragaria vesca* genes were identified as described previously (Supplemental Data 3.3) (Li et al., 2022). Figure 3.4C shows that *Arabidopsis* homologs to 917 genes were found in the endosperm-enriched genes identified here and in *Arabidopsis* endosperm-specific transcriptomes (Picard et al., 2021).Thus, nearly one third of *F. vesca* endosperm-enriched genes are also expressed in the *A. thaliana* endosperm.

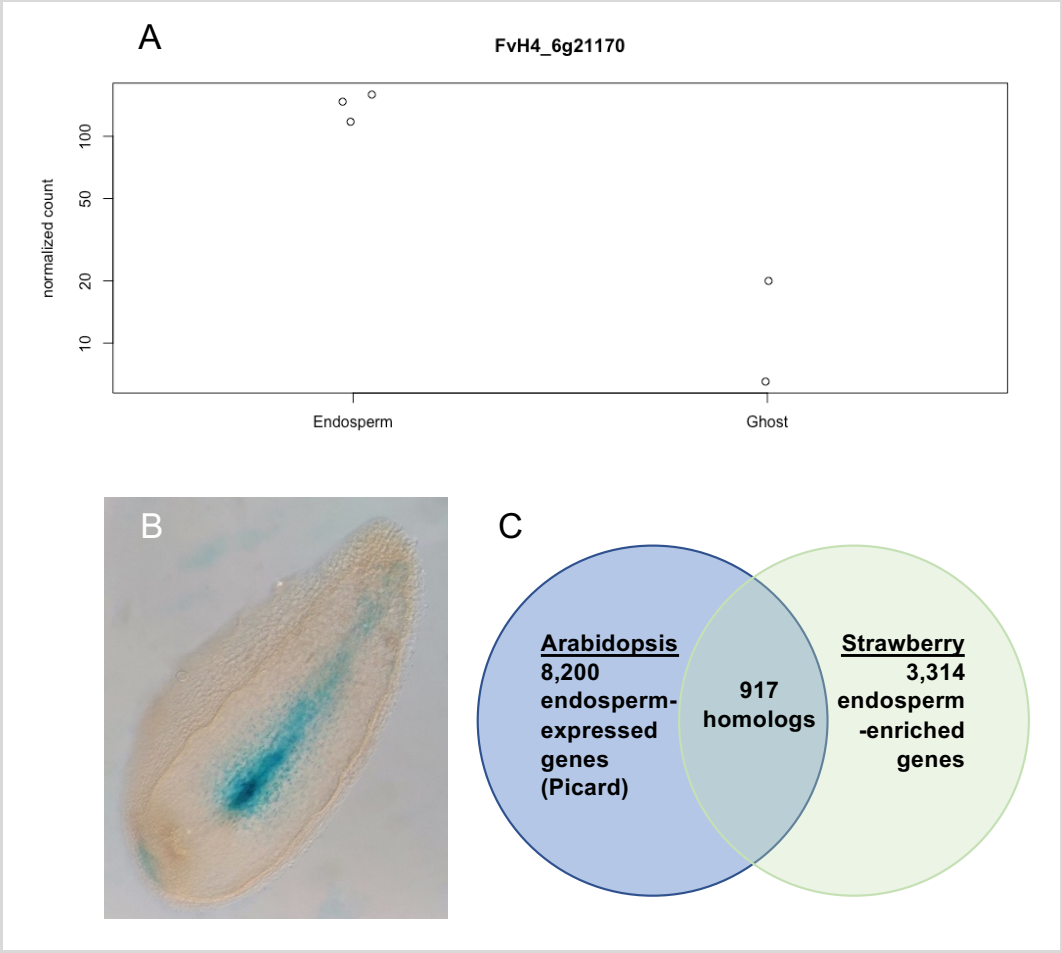


Figure 3.4. Confirmation that endosperm-enriched genes localize their expression to the endosperm. A) Absolute expression of *FveAGL80* in endosperm-only and ghost samples, normalized for each sample's library side. B) *pAGL80::GUS* is expressed exclusively in endosperm. C) 917 homologous genes are expressed in both *A. thaliana* and *F. vesca* endosperms.

3.2.4 Analysis of endosperm and seedcoat transcriptomes

I found endosperm-enriched expression of a number of previously-identified genes of interest, including auxin biosynthetic genes, *FveTAR1* (FvH4_5g05900) and *FveYUC11* (FvH4_4g17890), *FveMyb98* (FvH4_6g5100), and several of the ATHB genes that were recently

identified as downstream targets of *FveAGL62; FveAGL62* represses *ATHB* expression, and *ATHB* genes represses auxin biosynthesis gene expression, all occurring in the strawberry endosperm (Guo et al., 2022).

The three most highly expressed genes in the endosperm are all homologous to a gene that in *Arabidopsis* functions in seed storage potential (FvH4_2g21390, 2g21400 and 3g19080 are homologous to AT5G44120). An oleosin protein is also among the 25 most highly expressed genes; oleosin proteins are thought to stabilize highly energetic oil bodies in the seed (Frandsen et al., 2001). Together the seed storage and oleosin genes manage nutrients derived from maternal tissues. This is a classic endosperm function, further confirming the quality of the process to identify endosperm-specific transcription. While seed storage is not of major interest in strawberry, the importance of this gene family to endosperm function may be more relevance in other species. Also highly expressed is the GA3-oxidase FvH4_2g30020; the Arabidopsis homolog activates gibberellic acid (GA) hormone predecessor molecules (Katyayini et al., 2020). Kang et al. showed that GA plays a role in strawberry fruit development by demonstrating that co-administration of auxin and GA to unpollinated flowers yielded the more enlarged fruit than administration of either hormone alone (Kang et al., 2013). It is not surprising to confirm that GA biosynthesis gene expression is in the endosperm.

To determine which biological processes are enriched in endosperm and seedcoat respectively, I performed GO-enrichment analysis using the TopGO analysis software (Alexa et al., 2006). To identify seedcoat genes, I selected genes downregulated in the endosperm sample when compared with the ghost sample. I found seed GO terms enriched in biological processes in interspecies interactions, defense responses, rejection of self, and cell killing (Figure 3.5A and

Supplemental Data 3.4). Seedcoat expression of genes that provide protection against biotic stress is consistent with those shown in the *Arabidopsis* seedcoat (Belmonte et al., 2013).

In contrast, the overrepresented processes found in endosperm-enriched transcripts appear to be highly focused on internal development, including DNA replication, gene expression, translation and ribosomal assembly, and cell fate specification and commitment (Figure 3.5B and Supplemental Data 3.4). I also find that the endosperm exhibits an upregulation of genes associated with chromatin remodeling. These patterns are consistent with those observed in *Arabidopsis* (Belmonte et al., 2013). Interestingly, I did not find that defense genes are upregulated in endosperm as observed in Arabidopsis (Picard et al., 2021).

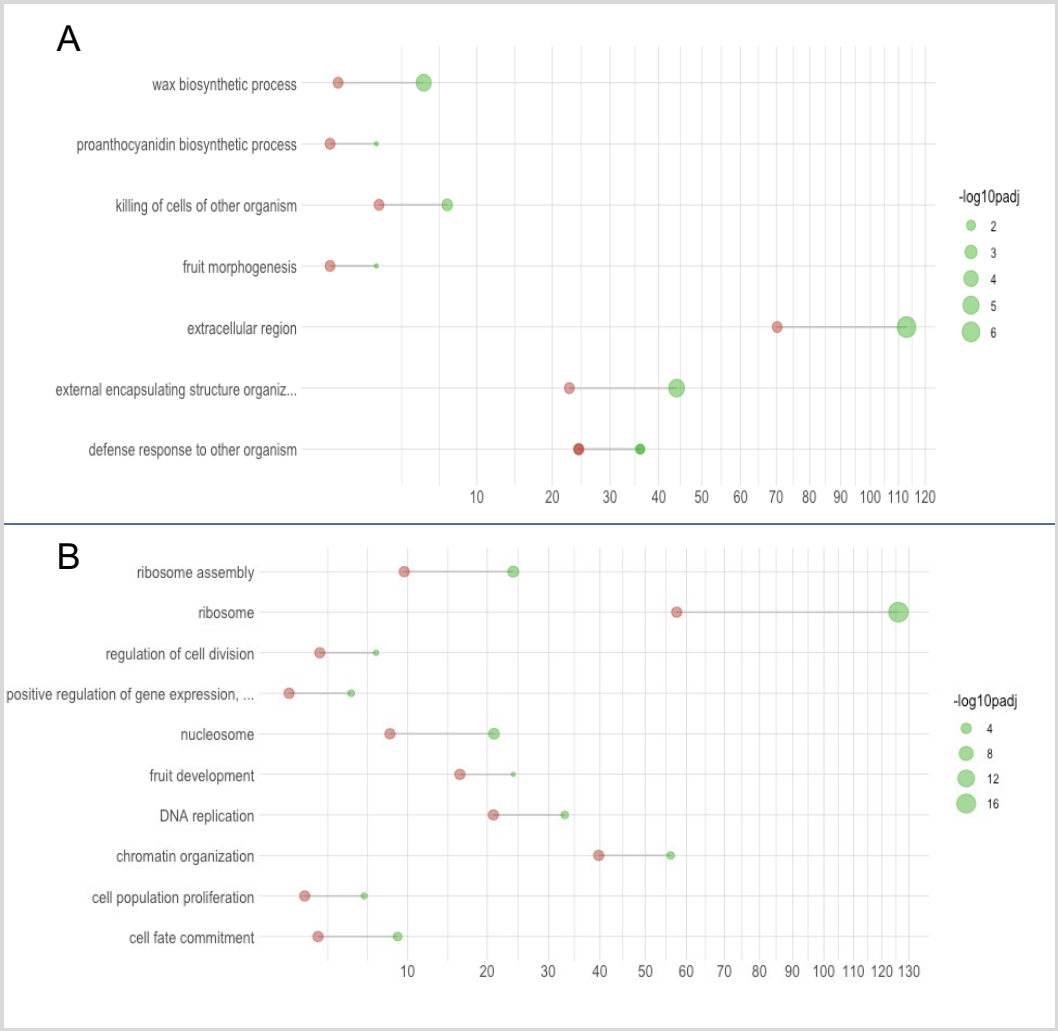


Figure 3.5. Selected gene ontology enrichment results. (A) Enriched GO terms in seedcoat-enriched transcripts from genes downregulated in endosperm-enriched sample versus ghost sample. (B) Enriched GO terms in endosperm transcripts from genes upregulated in the endosperm transcriptome. Red dots are expected gene number, green dots are observed gene number, and the size of the green dot indicates the statistical significance of the enrichment in adjusted p-value per the figures' legends. X-axis: number of genes.

3.3 Discussion

3.3.1 Summary of findings

An endosperm-specific transcriptome has helped to elucidate the distinct role of tissue compartments in the seed in several species, but one has not been previously available in strawberry. To fill that gap, I hand dissected endosperm from ~8 day old developing seeds and extracted bulk RNA, which was sequenced using Illumina 150 bp paired end reads. To identify endosperm-enriched transcription distinguishable from adjacent seedcoat or embryonic tissues, I looked for contrasts between the putative endosperm-specific samples and previously published transcriptomes of ghost (mixed endosperm and seedcoat), embryo, and other strawberry tissue samples (Figure 3.3), reasoning that the difference from those tissues must represent the endosperm-specific transcriptome. I identify a set of 3,314 endosperm-enriched and 225 genes which are upregulated in the post-fertilization, endosperm-enriched transcriptome (Supplemental Data 3.1-3.2); this latter set of gene candidates may be of significant interest to future research for questions relevant to endosperm development, regulation, and function. Quality checks of the endosperm-enriched transcriptome shows overlap with a small number of previously-known endosperm expressed genes in strawberry, including *FveAGL80* (Figure 3.4A,B), and I find that that strawberry endosperm-enriched transcriptome has significant overlap with an endosperm-specific transcriptome gathered at the single-nuclei level in *Arabidopsis* (Figure 3.4C)(Picard et al., 2021). GO analysis of the strawberry endosperm transcripts largely mimic those found in *Arabidopsis*, with the significant bias toward a proliferative state reinforcing the early state of endosperm development in the samples collected (Figure 3.4A-B). I anticipate that this will be the first endosperm-specific strawberry transcriptome to be made publicly available.

3.3.2 Caveats and limitations

The approach taken here reflects the practical limits imposed by work in a nonmodel system. In Arabidopsis thaliana, seed-specific transcriptomes adopt two technologies not employed in this project – cell type specific cell sorting (Del Toro-De León and Köhler, 2019) and single nucleus RNA-seq (Picard et al., 2021). I considered adopting both approaches. However, both rely on knowledge and techniques that are not currently available in strawberry.

The cell sorting approach relies on the presence of a florescent protein driven by a cell-type specific promoter. This approach thus requires both knowledge of that promoter and the ability to transform the florescent transgene into the plant. As I entered into this project, I did not know with certainty that *AGL62* has endosperm-specific expression, as this was demonstrated later (Guo et al., 2022). Further, the temporal domain of *FveAGL62* expression is highly limited to the first few days after pollination, limiting the usefulness of this promoter for studies in subsequent developmental periods. I also rejected this approach for three technical reasons. First, cell sorting requires the removal of cell walls (protoplasting), a technique which is not established in strawberry seeds. Second, I was concerned that seed tissues, with their high sugar and lipid levels, might be "sticky" and have poor sorting characteristics. Finally, even with the knowledge of endosperm-specific expression of *FveAGL62*, transformation of GFP-tagged lines would require at least 9 months.

Single nuclei RNA-seq is more promising, insofar as it does not rely on ex ante knowledge of cell type markers to disaggregate cells into types. However, without established, species-specific marker genes, that disaggregation would not allow the separation of cells into predefined cell types such as seed coat or endosperm. On the technical level, the approach would require the

extraction of nuclei from seed tissues. Nuclei extraction from leaf tissue is well established and indeed was utilized in sequencing the Yellow Wonder genome as described in Chapter 2 and elsewhere (Workman et al.) but had not been described in seed prior to 2021 (Picard et al., 2021). Given that my bulk RNA-seq sample was already prepared for sequencing at that time, and that there was uncertainty adapting the nuclei extraction technique from dry Arabidopsis seeds to strawberry, I moved forward with the bulk RNA-seq approach. Future investigation into the strawberry seed transcriptome might benefit from adopting the single nuclei technique.

Chapter 3 Supplemental Data

Supplemental Data 3.1 Endosperm-enriched transcriptome: GeneIDs.

Supplemental Data 3.2 Post-fertilization, endosperm-enriched transcriptome: GeneIDs.

Supplemental Data 3.3 *Arabidopsis thaliana* homologs of *Fragaria vesca* genes.

Supplemental Data 3.4 GO Enrichment Analysis of the endosperm- and seed-enriched transcriptome.

Chapter 4. Identification and analysis of imprinted gene expression in strawberry endosperm

4.1 Introduction

As discussed in Chapter 1, a major motivating hypothesis of this research program is the possibility—raised by others previously (Gehring and Satyaki, 2017; Figueiredo and Köhler, 2018; Figueiredo et al., 2015)—that expression of paternally expressed genes might act as a fertilization sensor that activates a previously repressed seed and fruit development program and actively seeks to antagonize post-fertilization reoccurrence of that repression. At the same time, the absence of paternally expressed gene products before fertilization is a critical functional quality that prevents precocious seed and fruit development. I further hypothesize that this paternal trigger might require a reciprocal, maternally expressed contribution that similarly limits inappropriate activation of the seed fate in the pollen. Thus, a complete fertilization sensing system may require the union of two genetic components, PEGs, and MEGs, each of which exhibits biased parent-of-origin expression. In this chapter, I identify imprinted genes whose expression shows maternal or paternal bias in wild strawberry from the endosperm transcriptomes derived from F1 seeds of an intraspecies cross. Among these genes, I identify candidate upstream regulators of type I MADS box gene *AGL62* and/or auxin biosynthesis genes.

4.2 Results

4.2.1 Identification of imprinted genes in F. vesca

As described in Chapter 3, I generated RNA-seq data from F1 endosperms that are the product of the cross between *Fragaria vesca vesca* YW5AF7 female plant and *Fragaria vesca bracteata* CFRA502 pollen. Alignment of RNA-seq to the maternal and paternal reference genomes is the standard approach used to identify genes exhibiting maternally or paternally biased expression (Babak et al., 2008; DeVeale et al., 2012; Wang et al., 2008), as single nucleotide polymorphisms (SNPs) that distinguish one parental genome from the other are reported in the RNA-seq read mapping. To facilitate this alignment, I created a "synthetic" genome consisting of the two parental genomes *Fragaria vesca vesca* YW5AF7 genome (Joldersma et al., 2022) to the paternal genome *Fragaria vesca bracteata* CFRA502 (currently unpublished, obtained from Dr. Patrick Edger at the Michigan State University), meaning that the synthetic genome consisted of seven YW5AF7 and seven CFRA502 chromosomes. LiftOff software (Shumate and Salzberg, 2020) was used to identify syntenic maternal and paternal genes in the synthetic genome. I used STAR (Dobin et al., 2013) to align RNA-seq data against the synthetic genome.

Table 4.1. Alignment of RNA-seq reads of endosperm-transcriptome to synthetic *YW5AF7/502* genome

	Millions reads (%)
Total RNA-seq reads	269 (100%)
…of which, reads aligning to synthetic genome (at one or more loci)	248 (82%)
…of which, reads aligned to only one locus in genome	82 (26%)

While over 82 percent of total reads could be aligned to the synthetic genome, in most cases a single read was aligned to homologous maternal and paternal loci that are identical; these "multimapping" reads were discarded as they are not informative of the parent-of-origin

(Table 4.1). I retained only reads that mapped exclusively to the paternal or the maternal genome (25.7 percent of reads; Table 4.1), as these uniquely-mapping reads harbor sequence polymorphisms that distinguish the two parental genomes.

To identify maternally- or paternally-expressed genes (MEGs or PEGs), I counted the reads aligned to the maternal or paternal copy of the syntenic genome using FeatureCounts software (Liao et al., 2014). As sexual endosperm is a triploid tissue with two maternal haploid genomes and one paternal haploid genome, at the genome-wide level maternal transcripts should be twice as abundant as the paternal transcripts. Pooling maternally- and paternally-aligned reads from the three samples, I found 1.7 maternal reads per one paternal read in the three endosperm samples, confirming that genome-wide expression is largely consistent with the 2 to 1 prediction (Supplemental Data 4.1).

At the single gene level, I found no evidence of expression bias at the *FveAGL80* gene locus (Figure 4.1). Alignment of the two parental sequences of the gene shows that there is one SNP, a C to G transversion from FvYW5AF7 to Fv502 (Figure 4.1.A). All 194 reads aligned to the FvYW5AF7 genome possess a C at this position, while 93 of 95 reads aligned to the Fv502 genome contain a G (Figure 4.1.B). I cannot explain the two reads with a C that are aligned to Fv502: there is no other SNP in this gene that should influence this alignment, and the read mapping and nucleotide calls are both made with high confidence. Nonetheless, this error rate is less than 0.7 percent and is unlikely to distort results grossly. 194 of 289 or 0.671 of total reads were aligned to the maternal genome (Figure 4.1.C), indicating that there is nearly 2:1 ratio of maternal to paternal transcripts at the *FveAGL80.*

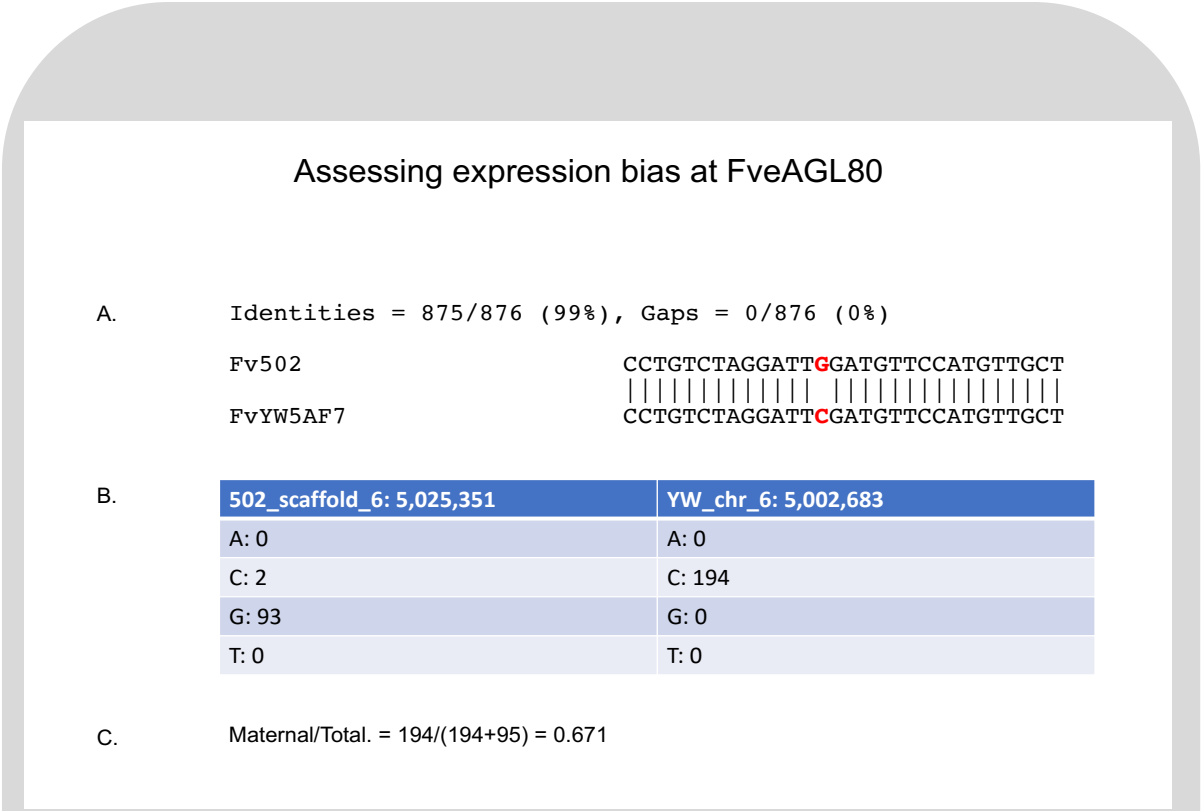


Figure 4.1. Analysis of expression bias at *FveAGL80*. (A) BLASTn result of maternal versus paternal genomic sequence. (B) Read alignment summary at the *FveAGL80* SNP site. (C) Calculation of percentage maternal expression.

A similar process was applied at all syntenic loci in the synthetic genome, and many loci showed a ratio of maternal to paternal reads that deviates dramatically from 2:1, ranging from zero to 100 percent of reads mapped to the maternal FvYW5AF7 genome (Figure 4.2). Figure 4.2 presents 3 histograms, each representing an independent biological sample, of how many genomic loci (y-axis) possess a specific percentage of maternal expression (x-axis), ranging from zero percent to 100 percent maternal expression. The highest peaks around 0.67 reflect the fact that most genes have close to 2:1 maternal:paternal expression. The large tails below 0.1 in blue

and above 0.9 in red represent loci with significant parent-of-origin expression biases in the strawberry endosperm.

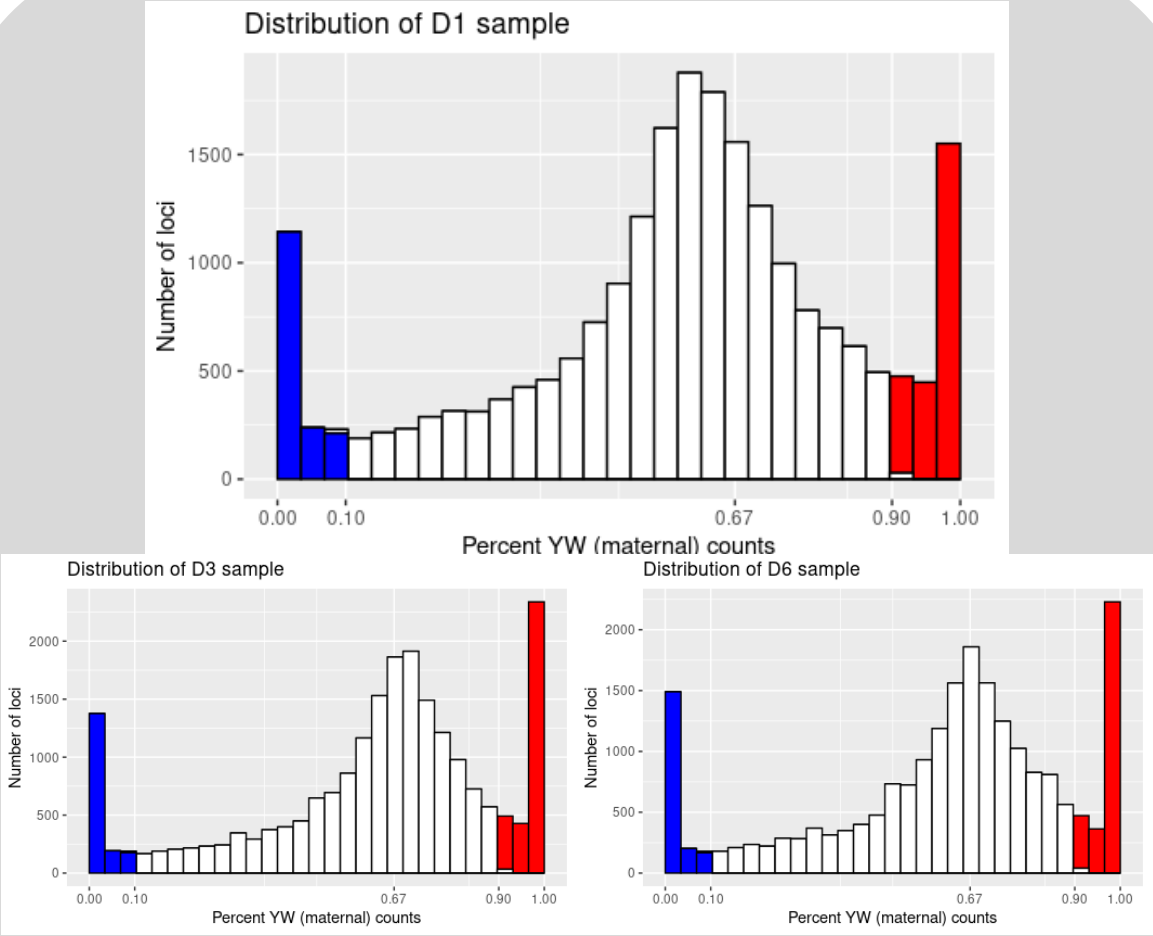
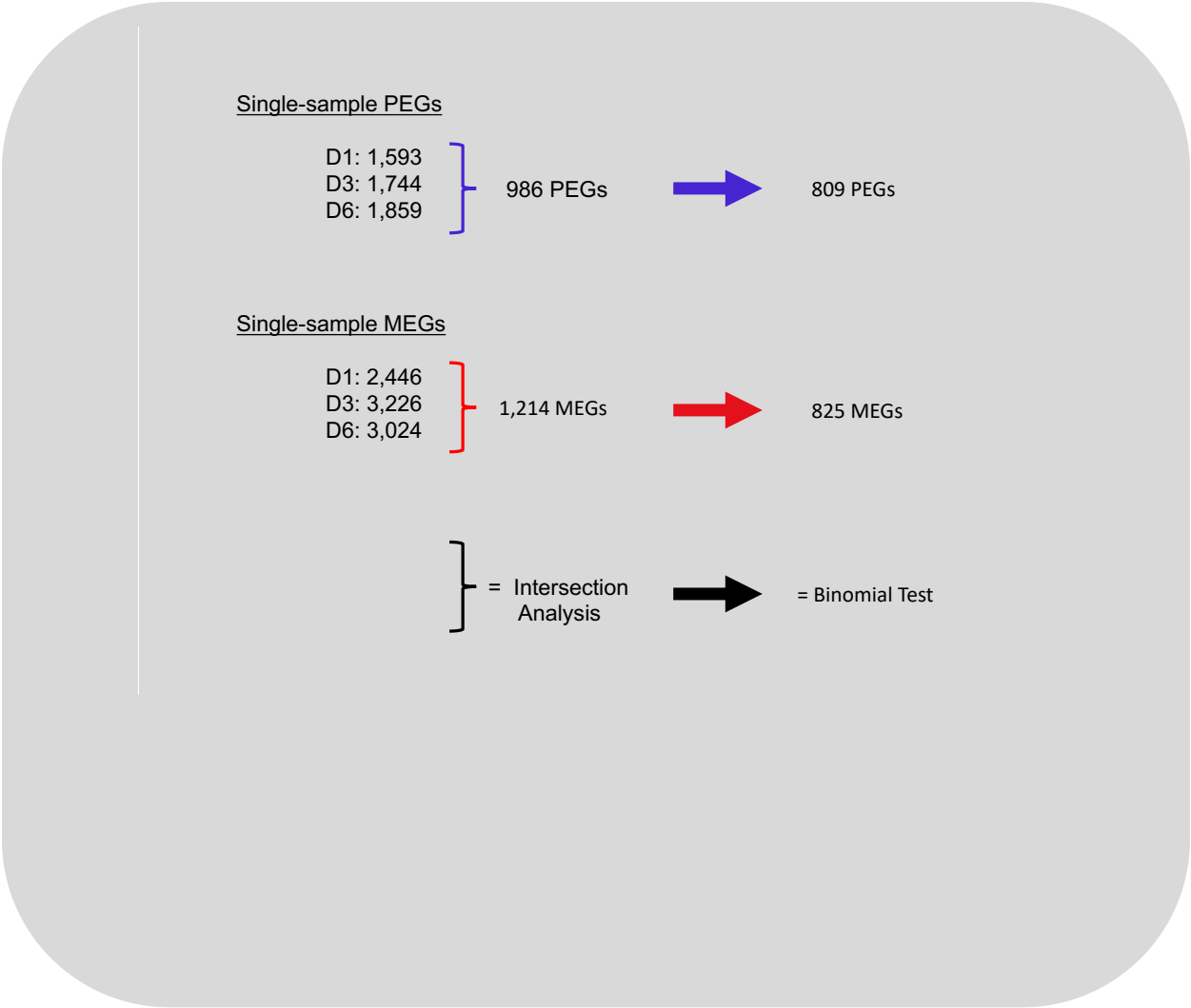


Figure 4.2. Analysis of expression bias at genomic scale.
At each genomic locus, bias is calculated per Figure 4.1. Histograms reflect number of genomic loci (y-axis) exhibiting specified percentage of maternal expression (x-axis). Histograms calculated in R-Studio, plotted in ggplot2. D1, D3, and D6 are three biological replicates of endosperm RNA-seq.

I classified as single-sample MEGs or PEGs genes for which more than 90% of reads were aligned to the maternal or paternal genome (red and blue tails of Figure 4.2). Reasoning that consistent imprinting status is essential to the imprint's functional properties, I conducted

intersection analysis to find genes that retained MEG or PEG status in each of the three samples (Figure 4.3), identifying 986 Preliminary PEGs and 1,214 Preliminary MEGs. I then applied the binomial statistical test to evaluate statistical confidence that the observed ratio at each locus deviated from 0.67, eliminating loci at which the probability of no deviation exceeded 0.001. I identified 825 MEGs and 809 PEGs (Supplemental Data 4.2).



4.2.2 Analysis of strawberry MEGs and PEGs

Several MEGs function in the regulation and homeostasis of auxin (Table 4.2 and Supplemental Data 4.2). Strawberry has four homologs of *DAO1* (*AT1G14130*) (Supplemental Data 3.3), a Fe2+ dioxygenase that is the major enzyme that degrades active auxin hormone to inactive 2-oxoindole-3-acetic acid (Porco et al., 2016). 99 percent of *FveDAO1* reads in these endosperm samples arise from just two genes (FvH4_1g09830 and FvH4_4g14690), and more than 90 percent of the reads at these two loci align to the maternal FvYW5AF7 genome (Table 4.2), indicating that the major two strawberry *DAO1* genes are MEGs. In addition, I found that the strawberry *FveIAGLU* genes *(FvH4_4g13000* and *FvH4_4g13010)*, which are homologous to *Arabidopsis AT4G15550*, are also maternally expressed, although *FvH4_4g13000* is not formally considered a MEG because of insufficient reads and an insufficiently confident pvalue (Table 4.2). In *Arabidopsis*, IAGLU reversibly inactivates active auxin hormone through conjugating auxin to a glucose molecule (Jackson et al., 2001). As auxin is a trigger for seed development, auxin catabolism can be inferred to reduce or slow seed development. These actions by DAO1 and IAGLU proteins to reduce the level of active auxin hormone are consistent with the role of the maternal genome in limiting progeny growth, as predicted in the parental conflict theory.

Table 4.2. Analysis of *FveDAO1* and *FveIAGLU* genes with imprinted status

	D1			D3			D6		
Gene locus	Paternal counts	Maternal counts	Binomial test: P-value	Paternal counts	Maternal counts	Binomial test: P-value	Paternal counts	Maternal counts	Binomial test: P-value
DAO1									
FvH4_1g09830	674	10067	0	6	5176	0	8	600	2.8e-91
FvH4_4g14690	14	442	4.1-58	0	226	4.9e-40	0	62	1.6e-11
FvH4_5g26041	16	7	0.999	2	8	0.307	4	27	0.010
FvH4_5g26050	2	10	0.188	0	19	5.0e-4	18	26	0.897
IAGLU									
FvH4_4g13000	13	94	6.1e-7	0	154	2.2e-16	0	22	1.5e-4
FvH4_4g13010	0	58	8.2e-11	0	32	2.6e-6	0	10	0.018

Homologs of Arabidopsis genes *HAT1* (AT3G54610) and *ASHH3* (AT2G44150) are PEGs (Table 4.3 and Supplemental Data 4.1). HAT1 is a histone acetylase (An et al., 2022) and thus regulates chromatin conformation, typically in a way that enhances transcriptional efficiency (Brownell and Allis, 1996). ASHH3 is a SET-domain containing histone methyltransferase that adds repressive H3K27me3 marks to histone tails (Lee et al., 2015). As regulators of the chromatin state, these proteins could enable the derepression of chromatin, consistent with changing patterns of gene expression in proliferative and differentiating cells.

Table 4.3. Selected MEGs and PEGs

F. vesca MEGs	Arabidopsis Homolog	Arabidopsis Gene ID	Description
FvH4_1g09830	AT1G14130	DAO1	Auxin catabolism
FvH4_4g14690	AT1G14130	DAO1	Auxin catabolism
FvH4_4g13000	AT4G15550	IAGLU	Auxin sequestration
F. vesca PEGs	Arabidopsis Homolog	Arabidopsis Gene ID	Description
FvH4_2g22590	AT3G54610	HAT1	Histone acetylase
FvH4_1g00050	AT2G44150	ASHH3	Histone methyltransferase

I used TopGo to analyze gene ontology enrichment of the 986 Preliminary PEGs (Figure 4.4 and Supplemental Data 4.3). Cell cycle-related genes and derivative processes such as nuclear division or mitotic spindle regulation are overrepresented among PEGs, consistent with the idea that PEGs promote a proliferative state in the developing endosperm. The biosynthesis of flavin-containing compounds is also overrepresented. The key enzymes in the auxin biosynthesis pathway are flavin mono-oxygenases, so this potentially represents a paternal regulatory network to increase and induce auxin production, but this conjecture requires further investigation.

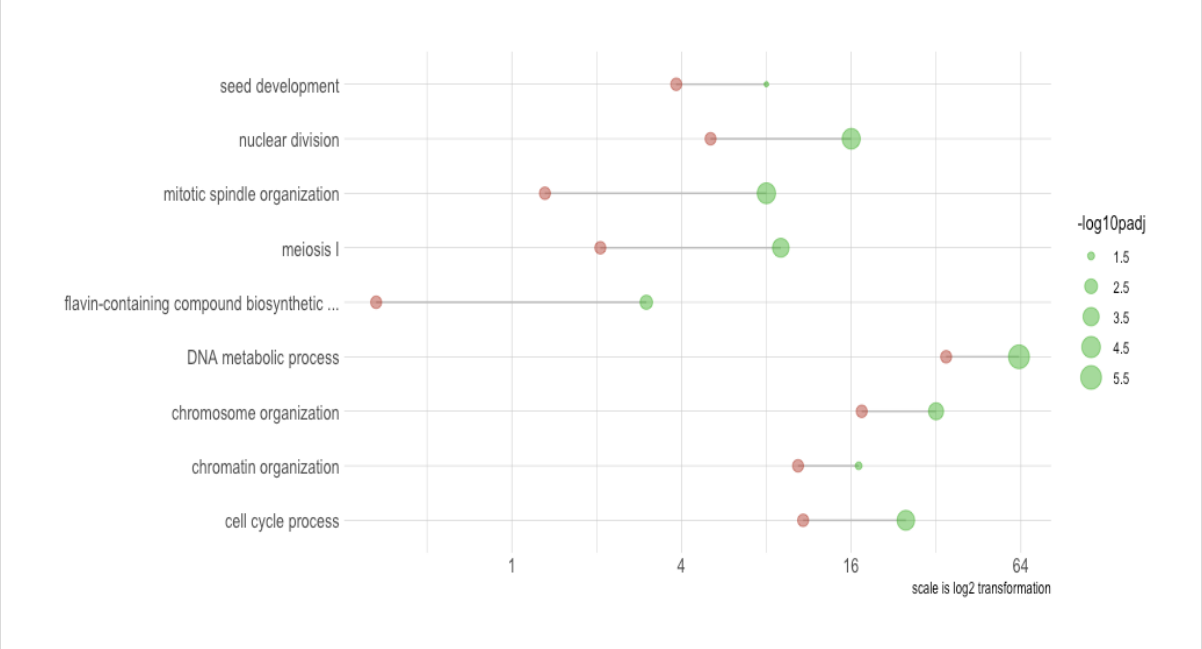


Figure 4.4. Gene ontology enrichment of PEG genes in strawberry transcriptome. Red dots are expected gene number, the green dots indicates actual gene numbers observed, and the size of the green dot indicates the statistical significance of the enrichment in adjusted p-value. X-axis is number of genes, and x-axis scale is log2 transformed.

4.2.3 Inter-species conservation of MEGs and PEGs

Conservation of the imprinting status of homologous loci is thought to be low across species (Rodrigues and Zilberman, 2015). To examine the extent to which strawberry imprinted genes diverge from or are similar to another species, I compared strawberry imprinted genes (Supplemental Data 4.2) to the combined MEGs and PEGs identified in two prior studies in Arabidopsis (Pignatta et al., 2014; Picard et al., 2021)(Supplemental Data 4.4). I find just four PEGs of a possible 180 are conserved across strawberry and Arabidopsis (Table 4.4). In Arabidopsis, *ARF6* encodes a transcription factor that mediates auxin signaling (Guilfoyle and Hagen, 2007) and a positive regulator of flower development (Nagpal et al., 2005). It makes sense that the maternal genome may seek to limit progeny growth by reducing expression of the maternal allele of a key auxin signaling genes.

Table 4.4. Conserved PEGs (*F. vesca* and *A. thaliana*)

At gene ID	At Gene Symbol	FvH4 homolog(s)	Description
AT3G51895	SULTR3; AST12	FvH4_4g00892	sulfate transporter 3;1
AT4G15900	PRL1	FvH4_4g23450	pleiotropic regulatory locus 1
AT4G11400	NA	FvH4_5g02771	ARID/BRIGHT DNA-binding domain; ELM2 and myb-like domain protein
AT1G30330	ARF6	FvH4_6g39140	auxin response factor 6

MEGs are more conserved, as I found 49 conserved MEGs between strawberry and Arabidopsis (Supplemental Data 4.3-4.4), or nearly 8 percent of 604 *Arabidopsis* MEGs. Among these conserved MEGs are *FveAG* (FvH4_3g06720), a homolog to *AGAMOUS (AG) (*AT4G18960)—a MADS-box transcription factor with homeotic function in floral development and that is expressed in maternal tissues in the gynoecium at the time of fertilization (Bowman et al., 1989, 1991)—and *FveKNAT3* (FvH4_3g16431), which is homologous to *KNAT3 (*AT5G25220), a

KNOX transcription factor of the type discussed in Chapter 1 as important to yeast and *Chlamydomonas* zygotic gene activation.

In addition to examples of conservation, I also wish to highlight a counter example—a lack of conservation. Interestingly, I find that *FvYUC10 (*FvH4_2g24750) is a MEG in strawberry (Supplemental Data 4.2). This status differs from its homologs in Arabidopsis and rice, in which *YUC10* was identified as paternally expressed (Luo et al., 2011). At the same time, when considering the four auxin biosynthetic genes expressed in the endosperm collectively *(FveYUC10, FveYUC11* (FvH4_4g17890)*, FveTAA1* (FvH4_4g25850), and *FveTAR1* (FvH4_5g05900)), I find only 20,550 of 44,375 (0.46) reads originate from the maternal genome—indicating a slight bias toward paternal expression. Applying the binomial test to the hypothesis this difference from 0.67 could arise by chance indicates this difference is highly unlikely (2.2e-16).

4.2.4 Identification of PEG and MEGs that may regulate FveAGL62 and auxin biosynthesis

To identify and prioritize among candidate imprinted genes that may regulate *FveAGL62*, I adopted a three-step strategy (Figure 4.5). First, I analyzed the promoter regions (1000 bp upstream of 5' UTR) of *FveAGL62* using PlantPan 3.0 (Chow et al., 2019) to identify candidate transcription factors that might bind to the consensus motif sequences present in the promoter of *FveAGL62*. This analysis identified 567 *F. vesca* transcription factors that may regulate *FveAGL62* (Supplemental Data 4.5). Second, I used a comparative genomics approach, taking advantage of the information on the activity of *A. thaliana* transcription factors found in the DAP-seq database (Bartlett et al., 2017). DAP-seq has identified 162 *A. thaliana* transcription factors that bind to *AtAGL62*, among which I identified 115 *F. vesca* homologs (Supplemental

Data 4.5). I reasoned that if a strawberry transcription factor might bind the promoter of *FvAGL62*, as predicted by PlantPan, and its *A. thaliana* homolog also binds the promoter (or gene body) of *AtAGL62,* as revealed by DAP-seq, such conserved TF-target binding relationship may be evidence of a conserved function. Finally, I examined if any of these candidate transcription factor genes overlap with MEGs or PEGs identified earlier. One PEG and 2 MEGs were identified as they are the only genes that meet the three criteria simultaneously (Table 4.5).

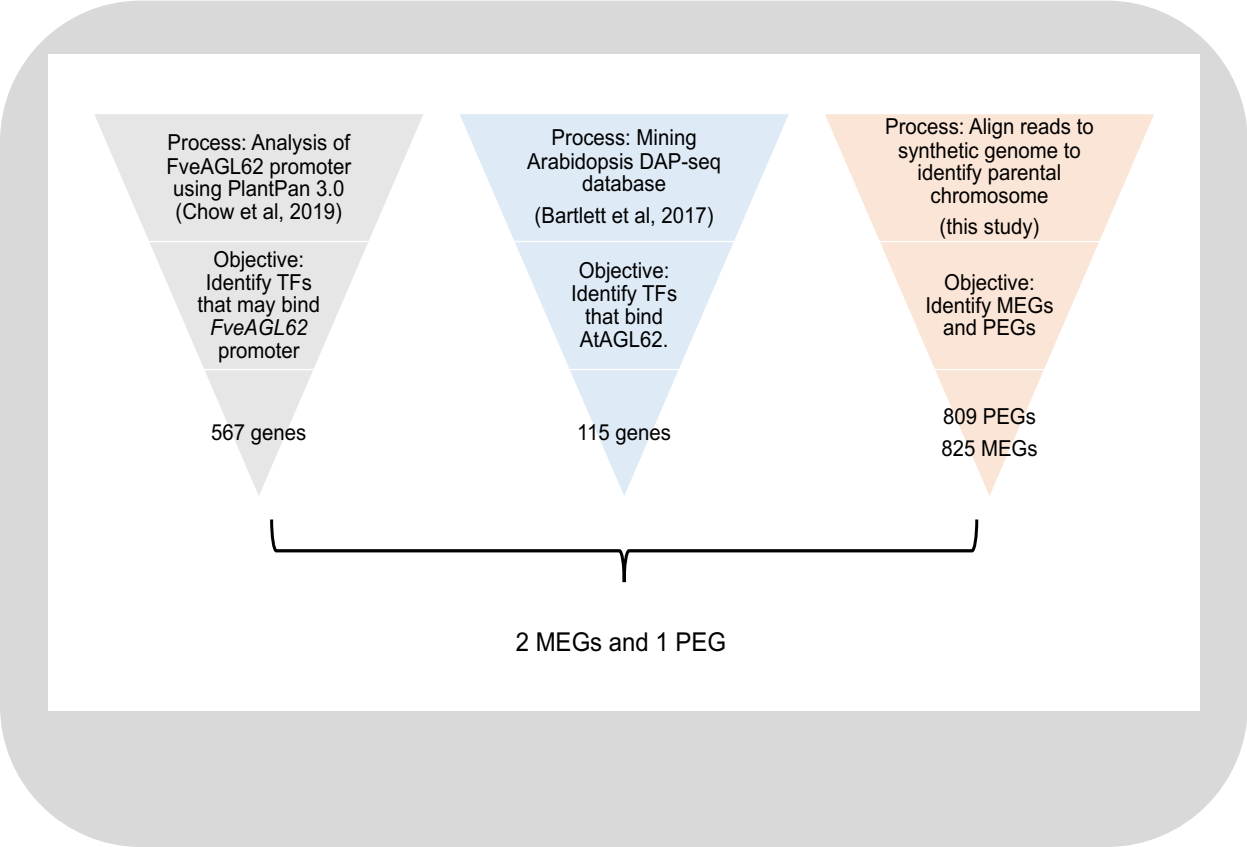


Table 4.5. Putative transcriptional regulators of *FveAGL62*

PEGs

FvH4 gene ID	At homolog(s)	At Gene Symbol	Description
FvH4_4g00820	AT4G14770	TCX2	TESMIN/TSO1-like CXC 2

MEGs

FvH4 gene ID	At homolog(s)	At Gene Symbol	Description
FvH4_6g51000	AT4G18770	MYB98	myb domain protein 98
FvH4_7g32570	AT2G46680	ATHB7	homeobox 7

I repeated the same 3-step pipeline to identify potential regulators of the four auxin biosynthesis genes expressed in the strawberry endosperm, *FveTAA1 (FvH4_4g25850), FveTAR1 (FvH4_5g_05900), FveYUC11 (FvH4_4g17980)* and *FveYUC10.* Specifically, I started with the 1000 bp promoter sequence 5' of each gene, identified candidate binding proteins to the chosen promoter through PantPan 3.0 and Arabidopsis DAP-seq (Supplemental Dataset 4.4), respectively, and find overlapping genes with PEGs and MEGs identified earlier. This analysis led to two PEG genes and six MEG genes that potentially bind the promoter of at least one of the four auxin biosynthetic genes (Table 4.6). Intriguingly, each of the three PEG/MEGs that were identified to regulate *FveAGL62* (Table 4.5) are also included within the set of the eight regulators of auxin biosynthesis genes (Table 4.6), suggesting that *FveAGL62* and auxin biosynthesis genes are perhaps coordinately regulated by the same sets of transcription factors.

Table 4.6. Putative Fragaria vesca Auxin Biosynthesis regulators

PEGs

FvH4 gene ID	Targets	At homolog	At Gene Symbol	Description
FvH4_5g29460	YUC11	AT5G62380	NAC101	NAC-domain protein 101
FvH4_4g00820	TAA1, YUC10	AT4G14770	TCX2	TESMIN/TSO1-like CXC 2

MEGs

FvH4 gene ID	Targets	At homolog		Description
FvH4_7g17540	YUC11	AT2G46770	NAC43	NAC-domain protein 43
FvH4_3g06200	TAA1, TAR1	AT4G18170	WRKY28	WRKY28
FvH4_3g08750	YUC11	AT3G15210	ERF4	ethylene response binding factor 4
FvH4_6g33050	YUC11	AT2G18060	VNAC1	vascular related NAC-domain protein 1
FvH4_6g51000	TAA1	AT4G18770	MYB98	myb domain protein 98
FvH4_7g32570	YUC10, YUC11	AT2G46680	ATHB7	homeobox 7

4.3 Discussion

4.3.1 Caveats and limitations

As discussed in the introduction, the most important limitation to this study arises from the fact that these results are the result of a single-directional cross. I intend to perform similar analysis of a reciprocal cross. Unfortunately, Fv502 is not a developed laboratory strain and there appears to be a partial parent-of-origin specific barrier to the reciprocal cross of paternal FvYW5AF7 into maternal Fv502. Assuming I can collect adequate samples despite this obstacle, I plan to conduct RNA-seq of the F1 endosperm to confirm the PEGs and MEGs using the same analysis pipeline as described above. I will then confirm PEGs and MEGs which retain their status in the reciprocal crosses. While this process will reduce the number of imprinted genes (Supplemental Data 4.2), it will increase confidence of their imprinted status by removing false positives that arise due to genetic—not epigenetic—differences between the maternal and paternal alleles.

It is unlikely, however, that the reciprocal cross will change the major outcomes of this analysis. First, this study was designed to limit tissue samples to endosperm only. This approach reflects the importance of the endosperm tissue generally; it also serves to minimize a major source of allele-specific expression: tissue-specific expression patterns as a result of genetic differences (Tomlinson et al., 2021; Pirinen et al., 2015; Pierre et al., 2022). Second, the study design reflects an unfortunate but pragmatic limit that has not halted similar efforts elsewhere. Imprinting studies in human tissues do not benefit from a reciprocal cross (Baran et al., 2015), and studies in less ethically difficult subjects, like "monkey plant" species in the genus Mimulus, have taken a similar approach to that taken here to identify of PEGs and MEGs (Flores-Vergara et al., 2020).

A second potential area of concern is that I collected sample from 7-9 days after floral anthesis (Chapter 3), in seeds with well-established zygotic transcription. The transcriptional profile of the developing seed changes over time (Belmonte et al., 2013), and thus the transcriptome I obtained did not represent the earliest stage at the onset of zygotic transcription. The collection point reflects the practical difficulty of dissecting endosperm tissue from the seed, as the younger seed is enfolded in a seedcoat that grows more rigid (Haughn and Chaudhury, 2005). In my dissection of endosperm, the older seedcoat was easier to separate from the endosperm than younger seedcoat, and so I collected at day 7-9 post pollination (Chapter 3). The transcriptome I have identified (Chapter 3) has independent value, despite a time point that is later than originally hoped for, and it will provide a point of comparison for future efforts to understand the strawberry transcriptome.

A related concern arises from the timing of the collection of this sample: are results generated from this sample relevant or informative regarding the role of imprinted genes that regulate early

auxin biosynthesis in the seed? In short, the results of this sample are not comprehensive, but they are nonetheless informative. The key facts that buttresses this assertion are the observations that imprinting is maintained through mitosis but the rate of gene imprinting declines through organismal development (Barlow and Bartolomei, 2014; John and Lefebvre, 2011; Ferrón et al., 2011). This work is not sufficient to prove imprinted status at the earlier time points. However, given the patterns of gene imprinting described above, it is likely that imprints detected at this time point existed at earlier time points and relatively unlikely that the imprinting detected in this sample arose at this time point but did not exist at previous time points.

4.3.2 Imprinting versus allele-specific expression

Preferential expression of an allele can arise from a genetic or epigenetic mechanisms (Patten et al., 2014). Examples of genetic change are DNA sequence changes to a promoter region that induce greater or reduced transcriptional activity or a DNA sequence change in the gene body itself that produces a premature stop codon and nonsense-mediated decay (Nurse and Hayles, 2011; Zrimec et al., 2020). Examples of epigenetic changes that regulate gene expression include differential DNA methylation or histone tail modifications that lead to changes to the chromatin state (Allis and Jenuwein, 2016). “Imprinting” generally and in this work refers to bias in expression that arises from epigenetic sources only and in a parent-of-origin specific manner; as such, imprinting is a subset of allele-specific expression (Patten et al., 2014).

The approach taken to identify PEGs and MEGs in section 4.2.1, using RNA-seq reads and alignment to parental genomes, has for more than a decade been the standard approach to identify preferential expression of an allele at genomic scale (DeVeale et al., 2012). The method relies on sequencing to detect differences in the sequences of the two alleles to identify the

parent from which the allele was inherited, and RNA-seq is a direct measure of these genetic differences and gene expression. RNA-seq experiments that aim to measure mRNA expression use poly-A-tail-mediated library construction techniques to eliminate non-mRNAs from the sequencing pool (Zhao et al., 2014). As such, the data from RNA-seq does not reflect differences in promoter regions. This implies two things: first, that RNA-seq cannot detect genetic differences in promoter regions, and second, that promoter-region genetic variation will not be used to assign parent-of-origin identity to a transcript.

Because RNA-seq is insensitive to differences in promoter regions, a single RNA-seq experiment from one-directional cross cannot distinguish between allele-specific expression that arises due to DNA-sequence difference from expression differences that arise from epigenetic imprinting. To verify parent-of-origin expression, most studies measure and compare parental origins of transcripts in products of reciprocal crosses in which maternal and paternal genotypes are reversed (Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011). Therefore, I intend to perform the reciprocal cross and adjust these results before eventual publication of this work.

4.3.3 Statistical robustness of PEG and MEG identification

I analyzed statistical significance of PEG and MEG identifications using three tests—the binomial test, a one-sample t-test and the chi square test. I present the results of using all three tests and each single test in Table 4.7. At any given statistical significance "cutoff" point, I find that application of any single test is nearly as selective as application of all three simultaneously. I further observe that the most selective test varies between the chi square and t-test at different cutoff values, whereas the binomial test is consistently in the middle of the two other tests in

terms of selectivity. As such, I decided to apply only the binomial test to restrict MEGs and PEGs to a statistically significant sample.

Table 4.7 Evaluating Statistical Robustness of MEG and PEG identification

	# of MEGs passing…			
Cutoff value	… ALL tests	… chi square only	… binomial test only	… t-test only
None	1250	1250	1250	1250
0.05	961	1010	965	961
0.01	891	891	897	897
0.001	792	792	825	852

Table 4.8 shows the effect of increasing the stringency of the binomial statistical testing process, as indicated by the number of PEG and MEG genes that pass various p-value cutoff points. I selected 0.001 as my cutoff point because the number of MEGs exceeds the number of PEGs at that cutoff point. Given the prominence of maternal gene expression in seed development, as discussed in Chapter 1, and findings that the number of MEGs exceeds the number of PEGs in other species (Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011; Waters et al., 2011; Luo et al., 2011; Zhang et al., 2011), it seemed reasonable to expect there would be more MEGs than PEGs in strawberry as well.

Table 4.8 Effects of different binomial test p-value cutoff values on number of imprinted genes

Cutoff p-value	# of PEGs	# of MEGs
No cutoff	978	1250
0.05	923	1010
0.01 (1e-2)	859	926
0.001 (1e-3)	809	825
0.0001 (1e-4)	771	769
0.00001 (1e-5)	740	719
1e-10	639	580

4.3.4 Conclusions

In this chapter, I make three chief findings. First, I identified putative 809 PEGs and 825 MEGs in the endosperm transcriptomes of *Fragaria vesca* (Supplemental Data 4.2). Second, gene ontology enrichment analysis of PEGs and MEGs provided several examples consistent with the predictions of parental conflict theory (Supplemental Data 4.2-4.3; Tables 4.2-4.3; Figure 4.4). Third, I identified two short lists (Tables 4.5-4.6) of candidate MEGs and PEGs for further analysis that: (1) contain DNA binding domains for target motifs in the promoters of *FveAGL62* and/or auxin biosynthetic genes expressed in the endosperm, and (2) exhibit a conserved TF-target binding relationship in Arabidopsis.

More speculatively, the identification of *FveYUC10* as a MEG (Supplemental Data 4.2) is in fascinating contrast to its PEG status in Arabidopsis and rice, but is not a complete surprise, as imprinted genes most frequently are not conserved across species (Rodrigues and Zilberman, 2015). A potential interpretation of the maternal expression of *FveYUC10* is that the endosperm's fertilization response is subject to multiple checkpoints that require input from both paternal and maternal genomes for successful release from pre-fertilization restrictions on egg cell development. Thus, while the details of *FveYUC10's* expression pattern may have changed, what may remain unchanged (albeit unconfirmed) is that it continues to serve as a fertilization sensor.

Chapter 4 Supplemental Data

Supplemental Data 4.1. Sample-wide maternal and paternal read distribution

Supplemental Data 4.2. MEGs and PEGs of strawberry endosperm

Supplemental Data 4.3. GO analysis of MEGs and PEGs

Supplemental Data 4.4. Conserved MEG and PEG status

Supplemental Data 4.5. PlantPan 3.0 and DAP-seq data sets.

Chapter 5: Maternally expressed FveMYB98 transcription factor regulates endosperm development

5.1 Introduction

In chapter 4, my analysis led to the identification of strawberry MEGs and PEGs. Among the MEGs, I identified *FveMYB98* (FvH4_6g51000), which encodes an MYB-R2R3 transcription factor. Prior characterization of the *Arabidopsis* homolog *AtMYB98* (AT4G18770) described a gene that is expressed early in female gametophyte differentiation in the synergid, egg and central cells, the last of which is the endosperm predecessor, and is essential to the synergid cell function of pollen tube guidance (Kasahara et al., 2005; Punwani et al., 2008, 2007; Susaki et al., 2021). However, whether *AtMYB98* possesses a function after fertilization is not known.

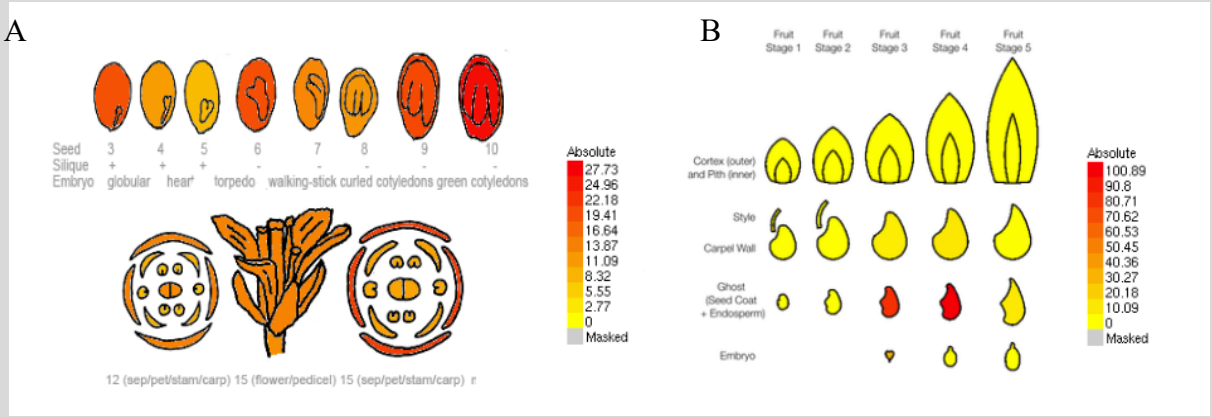


Figure 5.1. eFP showing pre- and post-fertilization expression of MYB98 in Arabidopsis and strawberry. (A) Arabidopsis eFP showing embryo development (top) and flower development (bottom). Scale bar data are normalized by the COS method, TGT value of 100. (B) Strawberry eFP showing only post-fertilization expression in ghost (scale bar in TPM) (Source:

In contrast to *AtMYB98's* pre-fertilization expression and functions in Arabidopsis (Figure 5.1A), the strawberry *FveMYB98* is expressed only after fertilization and predominantly in the ghost tissue consisting of endosperm and seedcoat (Figure 5.1B) (Kang et al., 2013). We further observed that *FveMYB98* is co-expressed with the maternally expressed auxin biosynthesis gene *FveYUC10* based on consensus clustering of co-expression networks in strawberry (Shahan et al, 2018). Given these observations and the bioinformatic analysis of Chapter 4, I hypothesized that FveMYB98 may be a maternally expressed regulator of *FveAGL62* and/or *FveYUC10* in the strawberry endosperm development.

5.2 Results

5.2.1 Description of FveMYB98 overexpression and knockout mutants

To investigate the role of *FveMYB98* in the development of strawberry seeds, I characterized transgenic *F. vesca* that overexpresses *FveMYB98* (*FveMYB98-OE)* via the Arabidopsis ubiquitin promoter as well as CRISPR/Cas9 knockout mutants of *FveMYB98*. Both over-expression and knockout mutants were created by lab member Lei Guo. 15 independent CRISPR knockout lines were obtained, and 20 independent transgenic *FveMYB98-OE* lines were obtained. I examined vegetative and leaf (Figure 5.2) as well as fruit and seed (Figure 5.3) phenotypes of the *fvemyb98* and *FveMYB98-OE* lines.

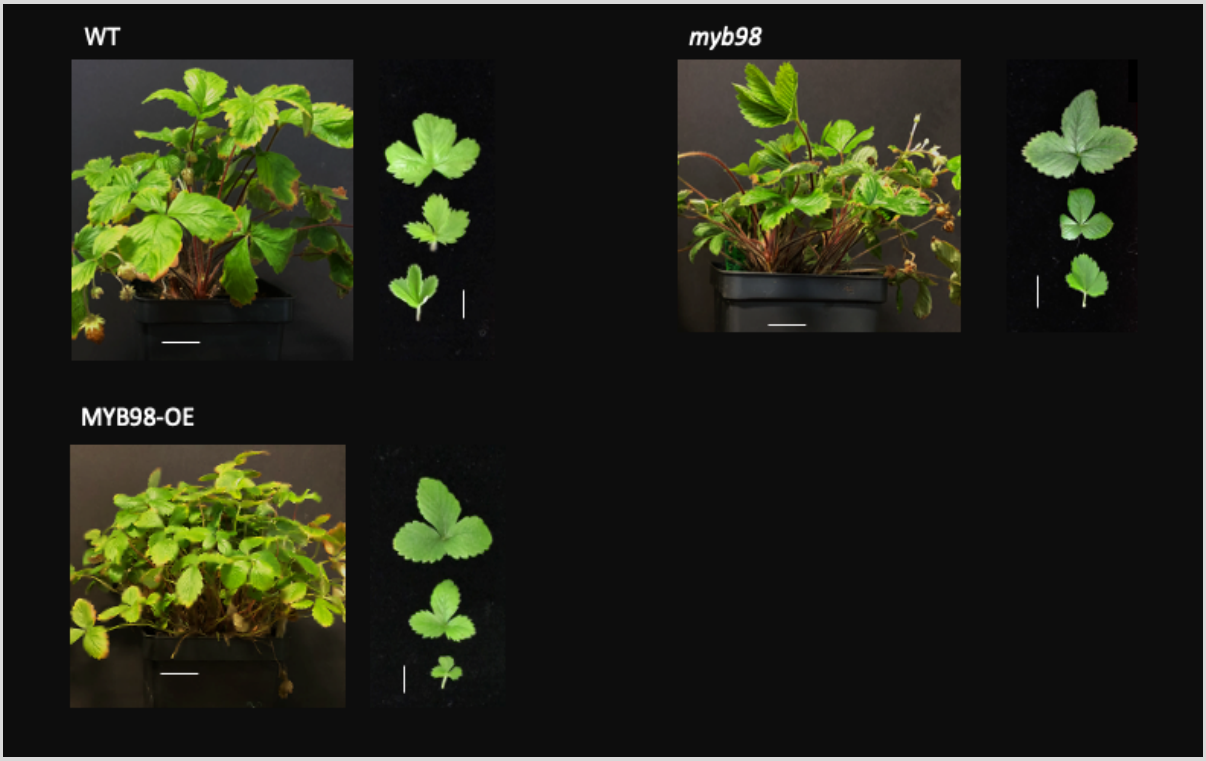


Figure 5.2. MYB98-OE exhibits vigorous vegetative growth. Architecture and leaf closeups of wild type (top left), MYB98-OE (bottom left) and *myb98* (top right) plants. Note the proliferation of shoots from the multiplication of branch crowns and relatively compact shape of MYB98-OE plant. Leaf size of MYB98-OE in general appear smaller than wild type or myb98, although similar size can be obtained (compare top leaf MYB98-OE and WT). Scale bar: 1 inch (potted plant) and 1 cm (leaf).

Examining four plants sourced from three lines, the vegetative growth of *myb98* appears normal (Figure 5.2), consistent with a limited domain of expression in the seed. However, *FveMYB98-OE* lines exhibit a clear vegetative phenotype; the plant grows bushier (Figure 5.2), due to higher number of branch crowns in contrast to wild type or *fvemyb98* and the large number of leaves growing from these branch crowns. The leaf size of *MYB98-OE* appears smaller than WT, but that difference has not been quantified (Figure 5.2).

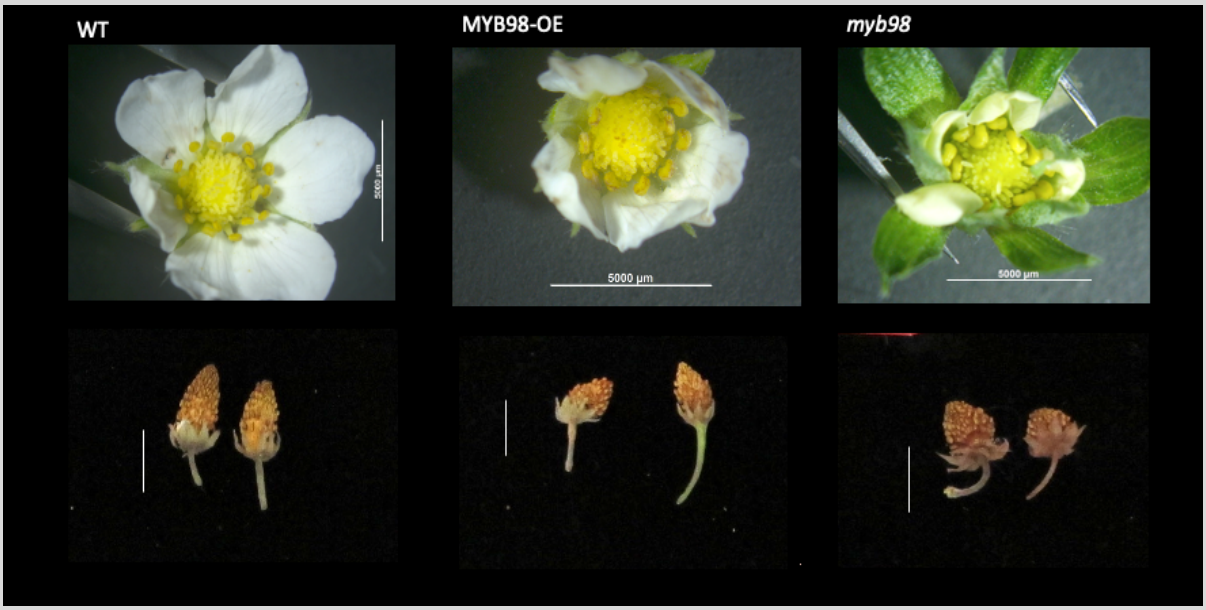


Figure 5.3. Reproductive phenotypes of Fv*MYB98-OE and fvemyb98.*
wild type (left), MYB98-OE (central) and *fvemyb98* (right) flowers are indistinguishable. At a gross level, fruits of *Fv-MYB98* appear smaller than wild type but are not distinguishable from *fvemyb98.* Scale bars: 5 mm (flowers) and 1 cm (fruit).

Flowers in *FveMYB98-OE, fvemyb98* and wild type genetic backgrounds are indistinguishable (Figure 5.3). Fruit length of wild type often appears to exceed that of *FveMYB98-OE* fruit (Figure 5.3), but we have not quantified this difference to date. We have not observed a fruit phenotype in *fvemyb98* fruits (Figure 5.3).

5.2.2 The seed phenotype of FveMYB98 knockout and overexpression mutants

The apparent change in the size of *FveMYB98-OE* versus wild type fruit (Figure 5.2.A-B) raised the possibility that there might be a seed phenotype. To determine if changes in *FveMYB98*

expression leads to any phenotype in seed development, we measured the size (total length) of fully mature, dormant wild type seeds and comparably mature seeds from three *fvemyb98* mutant lines. We found that *fvemyb98* mutant seeds were significantly larger than wild type seeds (Figure 5.4). In contrast, *FveMYB98-OE* plants developed seeds that were significantly smaller than WT (Figure 5.4). Together, these observations suggest the FveMYB98 protein may repress seed development. This function of the maternally inherited *FveMYB98* is consistent with the parental conflict theory, which suggests that the maternally inherited genome will seek to restrict the progeny's consumption of maternal resources.

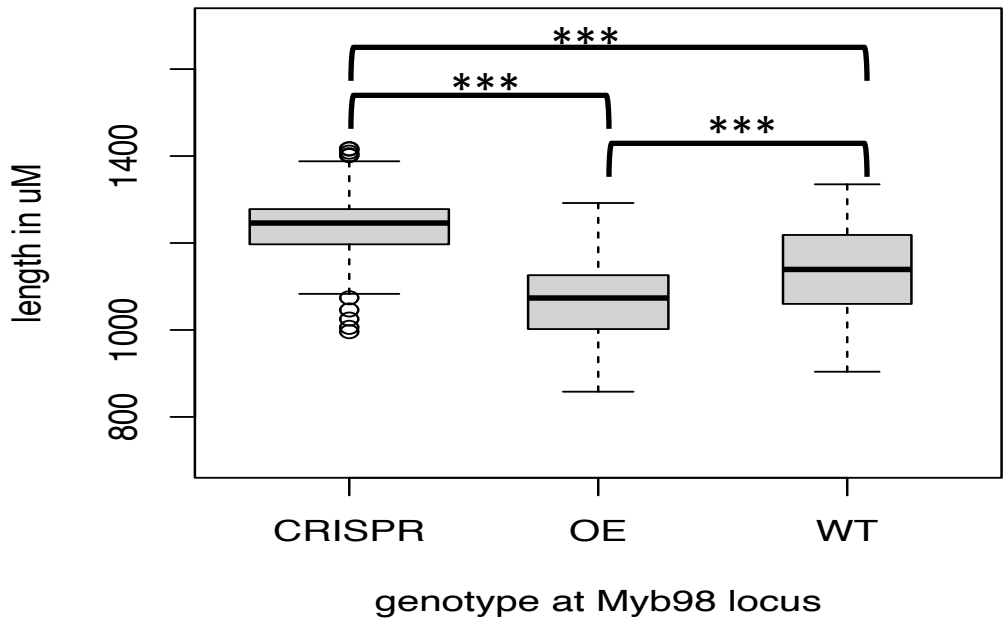


Figure 5.4. Seed size at mature stage in WT, *FveMYB98-OE,* and *fvemyb98* knockout mutants. Seed length (Y-axis) measured by AxioVision imaging software and images were obtained on a Zeiss Stemi stereoscope. N= ~50 of each genotype.

***** TuckeyHSB test indicates padj < 0.0001.**

5.2.3 The mechanism of FveMYB98's repression of seed size

As *FveAGL62* plays a role in auxin biosynthesis in endosperm, we hypothesized that *FveMYB98* may regulate seed size through regulation of *FveAGL62* and/or auxin biosynthetic genes. qPCR of a pooled cDNA of three FveMYB98-OE lines confirmed overexpression of *FveMYB98* in the *FveMYB98-OE* lines versus a wild type of sample (Figure 5.5A). qPCR of cDNA from the same samples also indicated that *FveAGL62* expression is significantly downregulated in the *FveMYB98-OE* cDNA sample (Figure 5.5B). This preliminary result suggests the *FveMYB98* directly or indirectly regulates *FveAGL62* expression; I plan to repeat this experiment using samples from single lines.

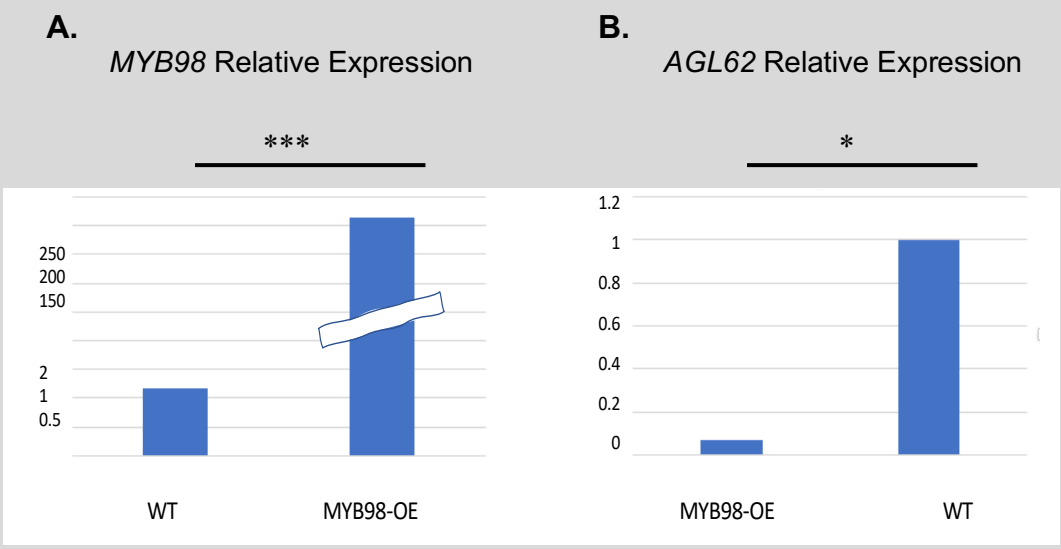


Figure 5.5. Overexpression of *MYB98* reduces expression of *FveAGL62*. RT-qPCR analysis of *FveMYB98* and *FveAGL62* expression in *FveMYB98-OE* and wild type stage 2 seeds. Y-axis indicates the relative expression level against the control gene *FvePP2a* (FvH4_4g27700). Significant difference by two-tailed Student's *t*-test is indicated by *** ($P < 0.002$) or * ($P < 0.1$) between WT and *fveagl62* mutant seed.

In light of the change in expression of *FveAGL62* in *MYB98-OE* plants, and to confirm the PlantPan3.0 prediction that FveMYB98 may bind the promoter of *FveAGL62* (Chapter 4, Table 4.5), we performed a yeast-one hybrid assay. The FveMYB98-GAL4 Activation Domain (GAL4AD) fusion protein was tested in the yeast cells against the antibiotic resistant reporter gene Aureobasidin A (AbA) under the control of *FveAGL62* promoter (1,512 bps upstream from the *FveAGL62* transcription start site). If FveMYB98-GAL4AD can bind the *FveAGL62* promoter, it will activate the expression of the antibiotic resistance gene and enable the yeast cells to survive in a medium containing the antibiotic AbA. I found that yeast cells with a positive control-GAL4AD fusion protein survive antibiotic challenge, but that cells containing the MYB98-GAL4AD and the *FveAGL62* promoter driven reporter failed to survive, a result similar to what was obtained using an empty-vector (AD-only) control (Figure 5.6A). From this experiment, we can infer that an FveMYB98 monomeric protein or FveMYB98 homopolymeric protein complex does not directly bind the promoter of *FveAGL62*. This result does not exclude

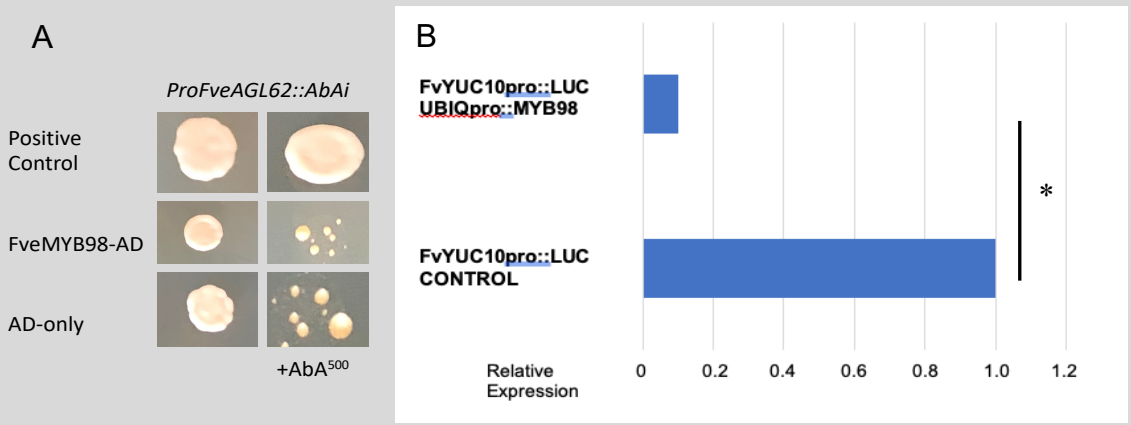


Figure 5.6. FveMYB98 does not bind the promoter of *FveAGL62* but does regulate FveYUC10. (A) Y1H assay. The positive control tests the binding of FveATHB to *ProFveYUC10::AbAi.* Data show yeast cell growth in the SD-Leu-Ura medium with or without 500 ng/ml AbA. AD-only vector is a negative control. (B) Transient repression of *ProFveYUC10::LUC* in tobacco leaves. X-axis indicates the ratio between firefly luciferase (LUC) signal normalized to the Renillia luciferase (REN). * = Statistical significance evaluated with two-tailed t-test; padj < 0.03. Luciferase data are from Lei Guo.

the possibility that FveMYB98 binds the promoter of *FveAGL62 in vivo* in the presence of other partner proteins.

AtYUC10 initiates the conversion of tryptophan to active auxin hormone (Stepanova et al., 2008). To investigate if *FveMYB98* directly regulates the expression of *FveYUC10*, we conducted a transient luciferase assay in tobacco leaves. Twelvefold reduction of LUC driven by the *FveYUC10* promoter (*pFveYUC10::LUC*) was observed when *35S::FveMYB98* was transiently expressed in the tobacco leaves (Figure 5.6B). This *in planta* result provides preliminary support of the hypothesis that FveMYB98 transcription factor might repress the expression of the key auxin biosynthetic gene *FveYUC10* in the strawberry endosperm.

5.3 Discussion and conclusion

In Chapter 5, I investigated the hypothesis that the MEG *FveMYB98* regulates *FveAGL62* or auxin biosynthesis genes. *FveMYB98-OE* plants have bushier branches and smaller leaves, and their fruit appears smaller than wild type (Figures 5.2-5.4). In contrast, the loss-of-function *myb98* mutants have seeds longer than wild type but wild type-like vegetative growth. Seed length is reduced in seeds collected from *FveMYB98-OE* plants, indicating that *FveMYB98* functions to negatively regulate seed size. This is consistent with a reduced *FveAGL62* expression in *MYB98-OE* lines (Figure 5.5). *In vitro* yeast 1 hybrid failed to show direct binding of FveMYB98 protein to the promoter of *FveAGL62* (Figure 5.6A), suggesting an indirect regulation, but binding of MYB98 to the promoter of *FveAGL62* could be further investigated *in vivo* via a chromatin immunoprecipitation-based assay, *in planta* via a luciferase assay, or *in vitro* via electrophoretic mobility shift assay. In contrast, FveMYB98 protein did directly act to repress *FveYUC10* expression in a heterologous tobacco transient expression system (Figure 5.6B).

Together, these experiments suggest that the maternally expressed *FveMYB98* acts to control seed size through negatively regulating the expression of *FveAGL62* and auxin biosynthesis gene *FveYUC10* in the early developing endosperm. One possible model is that expression of *FveAGL62*, which stimulates production of auxin, is itself positively regulated by levels of auxin in the endosperm. (The existence of this feedback loop could be tested by measuring *FveAGL62* expression levels in the presence of elevated or reduced levels of auxin.) Repression of *YUC10* by MYB98 reduces auxin levels and attenuates the proposed *FveAGL62* positive feedback loop.

A broader possibility is that the maternal *FveMYB98* induces endosperm cellularization by reducing auxin synthesis (Figure 5.7). *FveAGL62* and *AtAGL62,* first characterized as a repressor of endosperm cellularization in *Arabidopsis* and later in strawberry (Guo et al., 2022; Kang et al., 2008), activates auxin biosynthesis. Previous work showed that endosperm cellularization is delayed by ectopic expression of auxin biosynthesis genes in crosses with excess paternal contribution to endosperm expression (Batista et al., 2019). FveMYB98 appears to directly regulate the level of auxin biosynthesis in the endosperm via repression of *FveYUC10* and to indirectly regulate expression of the cellularization repressor *FveAGL62* (Figures 5.5, 5.6A), addressing the two major regulators of cellularization identified to date. These functions position *FveMYB98* as a key inducer of cellularization and, as such, a regulator of the transition to a new phase of development of the endosperm. Moreover, as early cellularization is associated with a reduction in seed size (Luo et al., 2005) and thus with a reduction in the seed's need for maternally-supplied nutrients and energy, a potential role for maternally-expressed *FveMYB98's* as an inducer of cellularization is consistent with parental conflict theory.

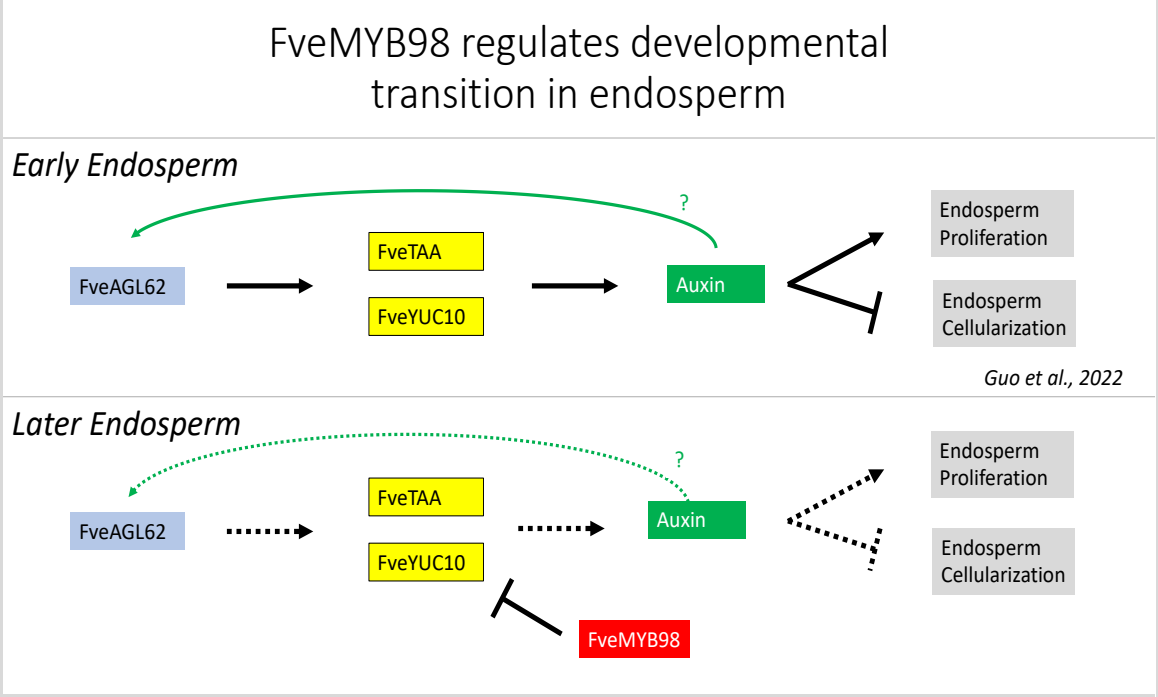


Figure 5.7. A model of FveMYB98 regulation of seed development. In early endosperm, it had been shown by Guo et al. that FveAGL62 induces auxin biosynthesis via regulation of FveTAA1 and FveYUC10, promoting endosperm proliferation and inhibiting endosperm cellularization. In the later endosperm, FveMYB98 inhibits expression of *FveYUC10*, restricting auxin biosynthesis, and inhibits *FveAGL62*, perhaps through the loss of auxin synthesis. That same loss of auxin synthesis may reduce endosperm proliferation and enable endosperm cellularization.

Chapter 6. Conclusions and next steps

This research program was designed to contribute to understanding how imprinted genes regulate the development of the strawberry endosperm. The chief outcomes are:

1. The assembly and annotation of the genome of FvYW5AF7.

2. The generation and analysis of the first available transcriptome of the *F. vesca* endosperm.

3. Identification of the expression of 809 PEGs and 825 MEGs in the endosperm of the F1 progeny of an intraspecies cross of FvYW5AF7 (female) and *F. vesca bracteata* "CFRA502" (male).

4. Description of *FveMYB98* as a regulator of *FveAGL62* and auxin biosynthesis in the developing endosperm.

I enjoyed and was rewarded by analyzing MEGs and PEGs in light of theoretical work on interparental and kinship conflict. My work found examples at the individual gene level and via gene ontology that the predictions of interparental theory apply to the evolution of the endosperm transcriptome of strawberry. The role of the MEG *FveMYB98* as a negative regulator of seed size fits very consistently into this theory.

Loose ends and questions abound, of course. At least six experiments would increase the accuracy or confidence of the findings to date:

1. The reciprocal cross of FvYW5AF7 (male) and *F. vesca bracteata* "CFRA502" (female) represents the major opportunity in front of me. Analysis of the transcriptome of this

cross will reduce false positive identifications of MEGs and PEGs. While RNA-extraction, sequencing, and analysis of the sequencing represents a major time commitment, the accuracy and credibility provided by doing so justifies the effort.

2. qPCR reassessing expression of *FveAGL62* and assessing expression of auxin biosynthetic genes in *MYB98-OE* and wild type background would confirm *in vivo* the findings of the *in planta* luciferase assay. These assays should be conducted in samples generated from a single line with known overexpression as opposed to the mixed sample used to generate the current preliminary data.

3. Y1H assays to assess whether MYB98 binds the promoters of the auxin biosynthetic genes could add further evidence of binding, in the case of *pYUC10*, or new evidence of binding in the case of *pTAA1*. These assays could also take place in the transient luciferase in tobacco, although that assay seems redundant to a Y1H.

4. I observed non-quantitatively that fruit size in *FveMYB98-OE* genetic background appears reduced versus wild type. I would like to quantify that observation.

5. I remain surprised that FveMYB98 did not bind *pFveAGL62* given the presence of binding motif and binding in *Arabidopsis.* I will repeat the Y1H assay under more stringent growth conditions (more antibiotic) and potentially investigate using *in vivo* or *in planta* methods (CHIP-PCR or luciferase, respectively).

6. Seed phenotyping: A key seed phenotype is germination potential. I will measure germination rates of seeds collected from independent lines of wild type, *FveMYB98-OE* and *fvemyb98* plants.

More broadly, there are some findings that raise questions that may be beyond the scope of this work. The first is the maternal expression of *FveYUC10,* which is at odds with the findings in other species. I don't think that change in imprinting status is a major surprise. What does surprise, however, is that *FveYUC10* is a target of maternally expressed MYB98. From the maternal perspective, this makes sense insofar as it establishes redundancy—the maternal genome has two levers (expression of *FveYUC10* or *FveMYB98*) through which it regulates endosperm auxin biosynthesis. But why silence the paternal allele? What is the evolutionary interest of the paternal genome in reducing auxin signaling?

A second set of questions arises around how this study may inform development in other species. The postfertilization role of *FveMYB98's Arabidopsis* homolog is undescribed. Could it play a similar role? Would this be worth investigating in the context of completing my thesis work? Probably not, but I would be very interested in the outcome of that research.

Finally, I note in closing that *FveMYB98* is one of three genes I highlighted in Chapter 4 as candidate post-fertilization regulators of *FveAGL62.* Investigation of *FveMYB98* has been productive. It also benefited massively from the fact that colleagues in the lab had identified *FveMYB98* as a candidate previously, only to later abandon the project. I was thus in position to pick up their handiwork and move things forward. I will not be able to conduct a similarly full functional analysis of the two other genes identified, *FveTCX2* and *FveATHB7* (Table 4.5), but they remain intriguing. To propel the conversation at least a little bit, I hope to complete Y1H assays to see if the proteins bind *pAGL62.*

Materials and methods

YW5AF7 genome assembly and annotation

Seventh-generation inbred Yellow Wonder accession, 5AF7, of *Fragaria vesca* was previously described (Slovin et al., 2009; Hawkins et al., 2016). Genomic DNA was extracted from young leaves by Norah Sadowski at Johns Hopkins University using a method previously described (Workman et al., 2018). Samples were sequenced on PromethION (Oxford Nanopore Technologies, Oxford, UK) for 48 hr in the Winston Timp laboratory, Johns Hopkins University, then raw fast5 data were basecalled with Albacore version 2.1.10 (Oxford Nanopore Technologies, Oxford, UK), yielding 26Gb of raw sequencing data (~125x coverage of the genome). The raw nanopore DNA sequencing data of YW5AF7 have been deposited in NCBI Sequence Read Archive under BioProject ID PRJNA871257.

Canu v1.9 (Koren et al., 2017) was used to correct, trim, and assemble raw Nanopore reads assuming a genome size of 240 MB and under default parameters. I used a conda (version 4.12.0) installation of RagTag (Alonge et al., 2019) in misassembly correction and scaffolding modes to merge the remaining fragmented scaffolds into pseudomolecules, generating a genome guided assembly based on the CRFA2339 reference assembly.

To polished the draft assembly, POLCA, a subprogram of the MASURCA genome assembly software (Zimin and Salzberg, 2020), was used three rounds, using 11.6GB of 50bp, single-ended Illumina reads (Hawkins et al., 2016) aligned to the draft assembly with minimap2 using the "–ax sr" parameter set (Li, 2018). The polished assembly was annotated with Liftoff (2.31.10) in eukaryotic mode using as a reference *F. vesca* H4 annotation v4.0.a2 downloaded

from Genome Database for Rosaceae (GDR), the Rosaceae genomic repository (Shumate and Salzberg, 2020; Li et al., 2019; Jung et al., 2019). Except for the Canu assembly, all computational work was performed using resources provided through the Cyverse iPlant Collaborative (Merchant et al., 2016). mRNA transcripts and protein translations, as well as an index of the assembly, are available at GDR; these were generated using gffread options -w and -y, respectively (Pertea and Pertea, 2020).

GO enrichment analysis

Gene ontology enrichment analysis was conducted using the BioConductor TopGO software package in R Studio, based on data provided in the *F. vesca* Hawaii 4 v4.0.a2 annotation (Alexa et al., 2006; Li et al., 2019). GOs were downloaded from Genome Database for Rosaceae (GDR) (Jung et al., 2019). The GO categories enriched for each gene module in the networks were determined by the Fisher's exact test in R package TopGO v2.34.0 (Alexa et al., 2006). The *P*-value cut-off was set at 0.05. The dot plot was created with R package ggplot2 (Wickham).

Hybrid seed generation, endosperm collection, and RNA-sequencing

Seeds of *F. vesca* accession YW5AF7 were mutagenized as described previously (Caruana et al., 2018), and F2 plants homozygous for a mutation in the *FveLEAFY* gene were identified and confirmed by lab colleague Lei Guo through sequencing. The *fvlfy-5* mutant plants were cultivated in standard long-day conditions for diploid strawberry: 16-hour light at 25 °C followed by an 8-hour darkness at 22 °C with a relative humidity of 50% (Hollender et al., 2012; Guo et al., 2022). *Fragaria vesca* accession "CFRA502" (Fv502) seeds were obtained from the USDA Germplasm Repository and cultivated under similar conditions.

Anthers from Fv502 were used to pollinate the *fveleafy-5* flowers, which lacked stamen and were otherwise sterile. Achenes were collected from the developing fruits derived from the cross at ~8 days after pollination. Endosperm was hand dissected from seeds inside the achenes under a Zeiss Stemi SV 6 microscope equipped with an AxioCam digital camera. The dissected endosperm tissue was stored at -80 degrees *Celsius.* The tissues were eventually pooled into three biological samples.

Total RNA was isolated from the frozen endosperm tissue using a Cetlytrimethylammonium Bromide (CTAB)-based RNA extraction buffer and 1:24 isoamyl alcohol:choroform, as described previously (Zhou et al., 2021). Total RNA was sent to Novogene, Sacramento, California. Three libraries corresponding to the three biological replicates were constructed using Ultra low input mRNA non-directional library preparation and sequenced with an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA).

These raw reads were trimmed using Fast-P software on default settings (Chen et al., 2018), and STAR, a splice-aware aligner (Dobin et al., 2013), was used to align reads to the Hawaii-4 reference genome (FvH4v4.0.a1) downloaded from GDR (Jung et al., 2019).

Genome-scale analysis to identify endosperm-specific genes

Differential gene expression analysis was conducted by DESeq2 v1.22.2 (Love et al., 2014) between the endosperm-derived RNA-seq data and previously published strawberry seed and fruit RNA-seq data (Kang et al. 2013), which was downloaded from NCBI's Sequence Read Archive at NCBI (http://www.ncbi.nlm.nih.gov/sra). The submission code is SRA065786. Genes with low read counts (≤20) were filtered out. The cutoffs $P_{adj} < 0.05$ and $\log_2$FoldChange > 1 were applied to identify DE genes. The built-in function plotPCA in DESeq2 (v1.22.2) was used

for the PCA. Intersection analysis was performed in RStudio using the built-in intersection function, selecting from lists of genes identified via DE analysis as described above.

Homology between Arabidopsis thaliana and Fragaria vesca genes

Arabidopsis thaliana homologs to *Fragaria vesca* genes were identified as described previously (Li et al., 2022). In short, to assign orthologs, Blast, OrthoFinder (v2.3.8), HMMER/Phylogeny, and Synteny analyses were performed (Altschul et al., 1990; Emms and Kelly, 2019; Mistry et al., 2013; Wang et al., 2012), and Blast and OrthoFinder hits were required to ascertain orthology.

Identification of imprinted genes, PEGs, and MEGs, in F. vesca

A synthetic genome of FvYW5AF7 v1.0 (Joldersma et al., 2022) and FvCFRA502 (Fv502)(obtained in collaboration with Patrick Edger, Michigan State) was created by appending the Fv502 genome to the FvYW5AF7 genome using the ">>" operator in unix. LiftOff (Shumate and Salzberg, 2020) was used to lift the FvYW5AF7 v1.0 genome to Fv502 and identify syntenic genes using the "-chroms" option for chromosome-scale lift over , the "-polish" option to correct alignments to coding sequences and "-copies 0.95" option to identify unannotated gene copies with 95% sequence similarity.

STAR (Dobin et al., 2013) was used to align reads to the synthetic genome with the "—outFilterMultimapNmax" option set to "1", requiring that maps that aligned to more than one genomic loci be discarded. Reads were counted using default settings of FeatureCounts software (Liao et al., 2014). MEGs and PEGs were identified using the subset command in R Studio with criteria that total reads from three samples > 20 and that in each of the three samples >90% or

<10% of total reads aligned to the maternal FvYW5AF7 5AF7v1.0 genome for the MEG and PEG assignment, respectively.

Analysis of MEGs and PEGs

To compare strawberry imprinted genes with Arabidopsis imprinted genes, Arabidopsis MEGs and PEGs were sourced from Supplemental Data 4, Tabs 2 and 3, of Picard et al. and Supplemental Data 2, Tabs 2 and 3, of Pignatta et al., which identify imprinted genes from reciprocal crosses using single-nuclei and bulk RNA-seq technologies, respectively (Pignatta et al., 2014; Picard et al., 2021). The union of MEGs and PEGs from the two Arabidopsis studies was calculated using the built-in union function in RStudio. Data frames were created of strawberry PEGs and MEGs and the union of Arabidopsis MEGs and PEGs, with the strawberry data frame including the gene ID of each strawberry gene's closest Arabidopsis homolog. To identify conserved imprints, the Arabidopsis and strawberry data frames were merged on the basis of any shared Arabidopsis gene IDs.

To identify PEGs and MEGs that may regulate *FveAGL62* and/or auxin biosynthesis genes, upstream 1 kb promoter sequence of strawberry PEGs and MEGs were selected from the FvH4 reference genome (v4.0.a1), downloaded from GDR (Jung et al., 2019) and submitted to PlantPan3.0 online portal for promoter analysis (http://plantpan.itps.ncku.edu.tw/promoter.php). The result of this submission was a list of *Arabidopsis thaliana* transcription factors that might bind motifs within the promoter of *F. vesca* PEGs and MEGs. *F. vesca* homologs of the *A. thaliana* transcription factors were identified as described in ortholog identification above.

Arabidopsis DAP-seq results were downloaded from the Plant Cistrome (http://neomorph.salk.edu/dap_web/pages/index.php) (Bartlett et al., 2017; O'Malley et al.,

2016) for *Arabidopsis thaliana* auxin synthesis genes *AT1G21430 (YUC11), AT1G48910 (YUC10), AT1G70560 (TAA1) and AT4G24670 (TAR1)*. Arabidopsis transcription factors with binding peaks in these auxin synthesis genes were identified from the website. *F. vesca* homologs of the *A. thaliana* transcription factors were identified as described above.

Generate Myb98 knockout and overexpression transgenic strawberry plants

The experiments in this section were performed by a Dr. Lei Guo in the Liu lab. For *FveMYB98* CRISPR construct and transgenic plants, gRNA1 (ATGGTGGTACCACCGAGGAGAGG) or gRNA2 (AAACGAAACAGCTGCCGCGGAGG) targeting the coding region of *FveMYB98* were inserted into the JH4 entry vector and then incorporated into the binary vector JH19 via LR reaction (Zhou et al., 2018). Agrobacteria harboring the JH19 vector was transformed into the strawberry strain Hawaii-4 through callus transformation (Guo et al., 2022). To confirm editing in transgenic plants, genomic sequences spanning the target site of respective genes were amplified by PCR and sequenced.

To generate transgenic plants that over-express *FveMYB98* (*FveMYB98-OE*), full length CDS of *FveMYB98* were PCR amplified from cDNA and then assembled into the vector JH23 (Zhou et al., 2021) at the digestion sites of KpnI and PacI through Gibson cloning (Gibson et al., 2009). In this vector, *FveMYB98* were driven by the Arabidopsis Ubiquitin 10 promoter. After transformation into *F. vesca*, more than 20 independent lines of *FveMYB98-OE* were confirmed by PCR genotyping.

Characterize the seed phenotype of FveMyb98 knockout and overexpression lines

To assess seed size, mature seeds were collected from multiple fruits derived from multiple *fvemby98, FveMYB98-OE*, and wild type *F. vesca* plants. The seed size measurement was derived from multiple OE and knockout lines. Seeds were imaged using a Zeiss Stemi SV 6 microscope equipped with an AxioCam digital camera. Seed length was measured along the chalazal to micropyle dimension and quantified using the "Length" function of AxioVision (Release 4.9.1) imaging software.

Gene expression analysis using RT-qPCR

For qPCR, total RNA was extracted from stage 2 fertilized seeds following CTAB RNA extraction methods described earlier and reverse transcribed to cDNA using SuperScript IV VILO Master Mix (Invitrogen, USA). RT-qPCR was performed using BioRad CFX96 Real-time system and SYBR Green PCR MasterMix (Applied Biosystems). The relative expression level was analyzed using a modified $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). For all RT-qPCRs, *FvePP2a* (FvH4_4g27700) was used as the internal control (Caruana et al., 2018). The RT-qPCR experiment is from one representative biological experiment (with three technical repeats). Biological replicates (2 to 3 times) gave similar results.

Yeast one hybrid and LUC transient expression assay

Constructs reported for Y1H and LUC transient assay were made by Dr. Lei Guo. The Y1H constructs and procedures were performed based on the Matchmaker Gold Yeast One-Hybrid Library Screening System User Manual (PT4087-1). Full-length CDS of genes of interest were PCR amplified and ligated into pGADT7 (Clontech Inc.) at the EcoRI/BamHI digestion sites by

Gibson assembly (Gibson et al., 2009). The 1512-bp *FveAGL62* promoter sequence was amplified and integrated into pAbAi (Clontech Inc.) at the KpnI/SalI digestion sites by Gibson assembly (Gibson et al., 2009). Yeast strain Y1HGold was transformed using the LiAc/PEG method (Gietz and Schiestl, 2007).

For the LUC transient assay, 1 kb of the promoter 5' of *FveYUC10* was PCR amplified from genomic DNA. It was first Gibson assembled into a modified gateway entry vector pLAH-R4R3-VP64Ter as described previously (Zhou et al., 2021), and then into the destination vector pLAH-LARm vector (Taylor-Teeples et al., 2015) by LR recombination. The Renilla luciferase gene (*35S:REN*) was used as a control to normalize the LUC readout. The assay was performed as previously described (Zhan et al., 2018). Each experiment was repeated two to three times with one representative result shown in Fig.5.6.

References

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22**: 1600–1607.

Alger, E.I., Platts, A.E., Deb, S.K., Luo, X., Ou, S., Cao, Y., Hummer, K.E., Xiong, Z., Knapp, S.J., Liu, Z., McKain, M.R., and Edger, P.P. (2021). Chromosome-Scale Genome for a Red-Fruited, Perpetual Flowering and Runnerless Woodland Strawberry (Fragaria vesca). Front. Genet. **12**: 1090.

Allis, C.D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. **17**: 487–500.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., and Schatz, M.C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. **20**: 224.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**: 403–410.

An, C., Deng, L., Zhai, H., You, Y., Wu, F., Zhai, Q., Goossens, A., and Li, C. (2022). Regulation of jasmonate signaling by reversible acetylation of TOPLESS in Arabidopsis. Mol. Plant **15**: 1329–1346.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. CB **18**: 1735–1741.

Baran, Y. et al. (2015). The landscape of genomic imprinting across diverse adult human tissues. Genome Res. **25**: 927–936.

Barlow, D.P. and Bartolomei, M.S. (2014). Genomic Imprinting in Mammals. Cold Spring Harb. Perspect. Biol. **6**: a018382.

Bartlett, A., O'Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. **12**: 1659.

Batista, R.A., Figueiredo, D.D., Santos-González, J., and Köhler, C. (2019). Auxin regulates endosperm cellularization in Arabidopsis. Genes Dev. **33**: 466–476.

Belmonte, M.F. et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc. Natl. Acad. Sci. **110**: E435–E444.

Bowman, J.L., Drews, G.N., and Meyerowitz, E.M. (1991). Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. Plant Cell **3**: 749–758.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in Arabidopsis. Plant Cell **1**: 37–52.

Brown, C.T. and Irber, L. (2016). sourmash: a library for MinHash sketching of DNA. J. Open Source Softw. **1**: 27.

Brownell, J.E. and Allis, C.D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. **6**: 176–184.

Caruana, J.C., Sittmann, J.W., Wang, W., and Liu, Z. (2018). Suppressor of Runnerless Encodes a DELLA Protein that Controls Runner Formation for Asexual Reproduction in Strawberry. Mol. Plant **11**: 230–233.

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. **94**: 4223–4228.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**: i884–i890.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell **110**: 33–42.

Chow, C.-N., Lee, T.-Y., Hung, Y.-C., Li, G.-Z., Tseng, K.-C., Liu, Y.-H., Kuo, P.-L., Zheng, H.-Q., and Chang, W.-C. (2019). PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. Nucleic Acids Res. **47**: D1155–D1163.

Del Toro-De León, G. and Köhler, C. (2019). Endosperm-specific transcriptome analysis by applying the INTACT system. Plant Reprod. **32**: 55–61.

DeVeale, B., Kooy, D. van der, and Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. PLOS Genet. **8**: e1002600.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. **29**: 15–21.

Dranginis, A.M. (1990). Binding of yeast a1 and alpha 2 as a heterodimer to the operator DNA of a haploid-specific gene. Nature **347**: 682–685.

Edger, P.P. et al. (2018). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. GigaScience **7**.

Elble, R. and Tye, B.K. (1991). Both activation and repression of a-mating-type-specific genes in yeast require transcription factor Mcm1. Proc. Natl. Acad. Sci. U. S. A. **88**: 10966–10970.

Emms, D.M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**: 238.

Enders, T.A. and Strader, L.C. (2015). Auxin Activity: Past, present, and Future. Am. J. Bot. **102**: 180–196.

Esquiliano, D.R., Guo, W., Liang, L., Dikkes, P., and Lopez, M.F. (2009). Placental glycogen stores are increased in mice with H19 null mutations but not in those with insulin or IGF type 1 receptor mutations. Placenta **30**: 693–699.

Ferrón, S.R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., Fariñas, I., and Ferguson-Smith, A.C. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature **475**: 381–385.

Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L., and Köhler, C. (2016). Auxin production in the endosperm drives seed coat development in Arabidopsis. eLife **5**: e20542.

Figueiredo, D.D., Batista, R.A., Roszak, P.J., and Köhler, C. (2015). Auxin production couples endosperm development to fertilization. Nat. Plants **1**: nplants2015184.

Figueiredo, D.D. and Köhler, C. (2018). Auxin: a molecular trigger of seed development. Genes Dev. **32**: 479–490.

Flores-Vergara, M.A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis Balaguer, M.A., Coimbra, S., Willis, J., and Franks, R.G. (2020). Developmental Analysis of Mimulus Seed Transcriptomes Reveals Functional Gene Expression Clusters and Four Imprinted, Endosperm-Expressed Genes. Front. Plant Sci. **11**.

Frandsen, G.I., Mundy, J., and Tzen, J.T.C. (2001). Oil bodies and their associated proteins, oleosin and caleosin. Physiol. Plant. **112**: 301–307.

Friedman, W.E. (1995). Organismal duplication, inclusive fitness theory, and altruism: understanding the evolution of endosperm and the angiosperm reproductive syndrome. Proc. Natl. Acad. Sci. U. S. A. **92**: 3913–3917.

Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. Cell **124**: 495–506.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PloS One **6**: e23687.

Gehring, M. and Satyaki, P.R. (2017). Endosperm and Imprinting, Inextricably Linked1[OPEN]. Plant Physiol. **173**: 143–154.

Giannoukakis, N., Deal, C., Paquette, J., Goodyer, C.G., and Polychronakos, C. (1993). Parental genomic imprinting of the human IGF2 gene. Nat. Genet. **4**: 98–101.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods **6**: 343–345.

Gietz, R.D. and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. **2**: 31–34.

Gramzow, L., Ritz, M.S., and Theißen, G. (2010). On the origin of MADS-domain transcription factors. Trends Genet. **26**: 149–153.

Grossniklaus, 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.

Guilfoyle, T.J. and Hagen, G. (2007). Auxin response factors. Curr. Opin. Plant Biol. **10**: 453–460.

Guitton, A.-E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.-E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Dev. Camb. Engl. **131**: 2971–2981.

Guo, L., Luo, X., Li, M., Joldersma, D., Plunkert, M., and Liu, Z. (2022). Mechanism of fertilization-induced auxin synthesis in the endosperm for seed and fruit development. Nat. Commun. **13**: 3985.

Gustafson, F.G. (1936). Inducement of Fruit Development by Growth-Promoting Chemicals. Proc. Natl. Acad. Sci. U. S. A. **22**: 628–636.

Gustafson, F.G. (1939). The Cause of Natural Parthenocarpy. Am. J. Bot. **26**: 135–138.

Haig, D. and Westoby, M. (1989). Parent-Specific Gene Expression and the Triploid Endosperm. Am. Nat. **134**: 147–155.

Haughn, G. and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends Plant Sci. **10**: 472–477.

Hawkins, C., Caruana, J., Schiksnis, E., and Liu, Z. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. Sci. Rep. **6**: 29017.

Hay, A. and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. Development **137**: 3153–3165.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima S, null, Kuroiwa, H., and Kuroiwa, T. (2001). Pollen tube attraction by the synergid cell. Science **293**: 1480–1483.

Hisanaga, T., Fujimoto, S., Cui, Y., Sato, K., Sano, R., Yamaoka, S., Kohchi, T., Berger, F., and Nakajima, K. (2021). Deep evolutionary origin of gamete-directed zygote activation by KNOX/BELL transcription factors in green plants. eLife **10**: e57090.

Hollender, C.A., Geretz, A.C., Slovin, J.P., and Liu, Z. (2012). Flower and early fruit development in a diploid strawberry, Fragaria vesca. Planta **235**: 1123–1139.

Hollender, C.A., Kang, C., Darwish, O., Geretz, A., Matthews, B.F., Slovin, J., Alkharouf, N., and Liu, Z. (2014). Floral Transcriptomes in Woodland Strawberry Uncover Developing Receptacle and Anther Gene Networks. Plant Physiol. **165**: 1062–1075.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409**: 525–529.

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. **108**: 1755–1762.

Iwata, H., Gaston, A., Remay, A., Thouroude, T., Jeauffre, J., Kawamura, K., Oyant, L.H.-S., Araki, T., Denoyes, B., and Foucher, F. (2012). The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. Plant J. Cell Mol. Biol. **69**: 116–125.

Jackson, R.G., Lim, E.-K., Li, Y., Kowalczyk, M., Sandberg, G., Hoggett, J., Ashford, D.A., and Bowles, D.J. (2001). Identification and Biochemical Characterization of anArabidopsis Indole-3-acetic Acid Glucosyltransferase *. J. Biol. Chem. **276**: 4350–4356.

Jenuwein, T. and Allis, C.D. (2001). Translating the Histone Code. Science **293**: 1074–1080.

John, R.M. and Lefebvre, L. (2011). Developmental regulation of somatic imprints. Differentiation **81**: 270–280.

Joldersma, D., Sadowski, N., Timp, W., Liu, Z., Joldersma, D., Sadowski, N., Timp, W., and Liu, Z. (2022). Assembly and annotation of *Fragaria vesca* "Yellow Wonder" genome, a model diploid strawberry for molecular genetic research. Fruit Res. **2**: 1–5.

Jung, S. et al. (2019). 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Res. **47**: D1137–D1145.

Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N., and Liu, Z. (2013). Genome-Scale Transcriptomic Insights into Early-Stage Fruit Development in Woodland Strawberry *Fragaria vesca*. Plant Cell Online **25**: 1960–1978.

Kang, I.-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.

Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). MYB98 Is Required for Pollen Tube Guidance and Synergid Cell Differentiation in Arabidopsis. Plant Cell **17**: 2981–2992.

Katyayini, N.U., Rinne, P.L.H., Tarkowská, D., Strnad, M., and van der Schoot, C. (2020). Dual Role of Gibberellin in Perennial Shoot Branching: Inhibition and Activation. Front. Plant Sci. **11**.

Khanday, I., Skinner, D., Yang, B., Mercier, R., and Sundaresan, V. (2019). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature **565**: 91–95.

Köhler, C. and Weinhofer-Molisch, I. (2010). Mechanisms and evolution of genomic imprinting in plants. Heredity **105**: 57–63.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**: 722–736.

Koskela, E.A., Mouhu, K., Albani, M.C., Kurokura, T., Rantanen, M., Sargent, D.J., Battey, N.H., Coupland, G., Elomaa, P., and Hytönen, T. (2012). Mutation in TERMINAL FLOWER1 reverses the photoperiodic requirement for flowering in the wild strawberry Fragaria vesca. Plant Physiol. **159**: 1043–1054.

Kurvari, V., Grishin, N.V., and Snell, W.J. (1998). A Gamete-specific, Sex-limited Homeodomain Protein in Chlamydomonas. J. Cell Biol. **143**: 1971–1980.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2016). Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. Cold Spring Harb. Perspect. Med. **6**: a026575.

Lee, J., Yun, J.-Y., Zhao, W., Shen, W.-H., and Amasino, R.M. (2015). A methyltransferase required for proper timing of the vernalization response in Arabidopsis. Proc. Natl. Acad. Sci. **112**: 2269–2274.

Lee, J.-H., Lin, H., Joo, S., and Goodenough, U. (2008). Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. Cell **133**: 829–840.

Leyser, O. (2005). Auxin distribution and plant pattern formation: how many angels can dance on the point of PIN? Cell **121**: 819–822.

Leyser, O. (2010). The Power of Auxin in Plants. Plant Physiol. **154**: 501–505.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**: 3094–3100.

Li, M., Galimba, K., Xiao, Y., Dardick, C., Mount, S.M., Callahan, A., and Liu, Z. (2022). Comparative transcriptomic analysis of apple and peach fruits: insights into fruit type specification. Plant J. Cell Mol. Biol. **109**: 1614–1629.

Li, Y., Pi, M., Gao, Q., Liu, Z., and Kang, C. (2019). Updated annotation of the wild strawberry Fragaria vesca V4 genome. Hortic. Res. **6**: 1–9.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. **30**: 923–930.

Livak, K.J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif **25**: 402–408.

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.

de Lucas, M., Pu, L., Turco, G., Gaudinier, A., Morao, A.K., Harashima, H., Kim, D., Ron, M., Sugimoto, K., Roudier, F., and Brady, S.M. (2016). Transcriptional Regulation of Arabidopsis Polycomb Repressive Complex 2 Coordinates Cell-Type Proliferation and Differentiation[OPEN]. Plant Cell **28**: 2616–2631.

Luo, H., Dai, C., Li, Y., Feng, J., Liu, Z., and Kang, C. (2018). Reduced Anthocyanins in Petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. J. Exp. Bot. **69**: 2595–2608.

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M. (1999). Genes controlling fertilization-independent seed development in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. **96**: 296–301.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., and Chaudhury, A. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proc. Natl. Acad. Sci. **102**: 17531–17536.

Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. PLoS Genet. **7**.

McGrath, J. and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37**: 179–183.

Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., and Antin, P. (2016). The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. PLOS Biol. **14**: e1002342.

Miller, A.M., MacKay, V.L., and Nasmyth, K.A. (1985). Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature **314**: 598–603.

Miray, R., Kazaz, S., To, A., and Baud, S. (2021). Molecular Control of Oil Metabolism in the Endosperm of Seeds. Int. J. Mol. Sci. **22**: 1621.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. **41**: e121.

Müller, K., Tintelnot, S., and Leubner-Metzger, G. (2006). Endosperm-limited Brassicaceae Seed Germination: Abscisic Acid Inhibits Embryo-induced Endosperm Weakening of Lepidium sativum (cress) and Endosperm Rupture of Cress and Arabidopsis thaliana. Plant Cell Physiol. **47**: 864–877.

Nagpal, P., Ellis, C.M., Weber, H., Ploense, S.E., Barkawi, L.S., Guilfoyle, T.J., Hagen, G., Alonso, J.M., Cohen, J.D., Farmer, E.E., Ecker, J.R., and Reed, J.W. (2005). Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. Dev. Camb. Engl. **132**: 4107–4118.

Nitsch, J.P. (1950). Growth and Morphogenesis of the Strawberry as Related to Auxin. Am. J. Bot. **37**: 211–215.

Nurse, P. and Hayles, J. (2011). The Cell in an Era of Systems Biology. Cell **144**: 850–854.

Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. Proc. Natl. Acad. Sci. U. S. A. **93**: 5319–5324.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell **11**: 407–416.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165**: 1280–1292.

Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R., and Wolf, J.B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. Heredity **113**: 119–128.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature **405**: 200–203.

Pertea, G. and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Research **9**: ISCB Comm J-304.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet. **15**: 517–530.

Picard, C.L., Povilus, R.A., Williams, B.P., and Gehring, M. (2021). Transcriptional and imprinting complexity in Arabidopsis seeds at single-nucleus resolution. Nat. Plants **7**: 730–738.

Pierre, C.L.S., Macias-Velasco, J.F., Wayhart, J.P., Yin, L., Semenkovich, C.F., and Lawson, H.A. (2022). Genetic, epigenetic, and environmental mechanisms govern allele-specific gene expression. Genome Res.: gr.276193.121.

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

Pirinen, M., Lappalainen, T., Zaitlen, N.A., Dermitzakis, E.T., Donnelly, P., McCarthy, M.I., and Rivas, M.A. (2015). Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics **31**: 2497–2504.

Piskurewicz, U., Jikumaru, Y., Kinoshita, N., Nambara, E., Kamiya, Y., and Lopez-Molina, L. (2008). The gibberellic acid signaling repressor RGL2 inhibits Arabidopsis seed germination by stimulating abscisic acid synthesis and ABI5 activity. Plant Cell **20**: 2729–2745.

Porco, S. et al. (2016). Dioxygenase-encoding AtDAO1 gene controls IAA oxidation and homeostasis in Arabidopsis. Proc. Natl. Acad. Sci. **113**: 11016–11021.

Powell, A. (2018). In Harvard studies of plant tug-of-war, mom wins. Harv. Gaz.

Punwani, J.A., Rabiger, D.S., and Drews, G.N. (2007). MYB98 Positively Regulates a Battery of Synergid-Expressed Genes Encoding Filiform Apparatus–Localized Proteins. Plant Cell **19**: 2557–2568.

Punwani, J.A., Rabiger, D.S., Lloyd, A., and Drews, G.N. (2008). The MYB98 subcircuit of the synergid gene regulatory network includes genes directly and indirectly regulated by MYB98. Plant J. **55**: 406–414.

Rehkamp, S. (2016). A Look at Calorie Sources in the American Diet.

Rodrigues, J.A. and Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. Genes Dev. **29**: 2517–2531.

Satyaki, P.R.V. and Gehring, M. (2017). DNA methylation and imprinting in plants: machinery and mechanisms. Crit. Rev. Biochem. Mol. Biol. **52**: 163–175.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Dev. Camb. Engl. **125**: 3329–3341.

Shahan, R., Zawora, C., Wight, H., Sittmann, J., Wang, W., Mount, S.M., and Liu, Z. (2018). Consensus Coexpression Network Analysis Identifies Key Regulators of Flower and Fruit Development in Wild Strawberry. Plant Physiol. **178**: 202–216.

Shulaev, V. et al. (2011). The genome of woodland strawberry (Fragaria vesca). Nat. Genet. **43**: 109–116.

Shumate, A. and Salzberg, S.L. (2020). Liftoff: accurate mapping of gene annotations. Bioinforma. Oxf. Engl.: btaa1016.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210–3212.

Slovin, J.P., Schmitt, K., and Folta, K.M. (2009). An inbred line of the diploid strawberry Fragaria vesca f. semperflorens for genomic and molecular genetic studies in the Rosaceae. Plant Methods **5**: 15.

Stepanova, A.N., Robertson-Hoyt, J., Yun, J., Benavente, L.M., Xie, D.-Y., Doležal, K., Schlereth, A., Jürgens, G., and Alonso, J.M. (2008). TAA1-Mediated Auxin Biosynthesis Is Essential for Hormone Crosstalk and Plant Development. Cell **133**: 177–191.

Stepanova, A.N., Yun, J., Robles, L.M., Novak, O., He, W., Guo, H., Ljung, K., and Alonso, J.M. (2011). The Arabidopsis YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis. Plant Cell **23**: 3961–3973.

Surani, M. a. H., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308**: 548–550.

Susaki, D., Suzuki, T., Maruyama, D., Ueda, M., Higashiyama, T., and Kurihara, D. (2021). Dynamics of the cell fate specifications during female gametophyte development in Arabidopsis. PLOS Biol. **19**: e3001123.

Taylor-Teeples, M. et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature **517**: 571–575.

Tennessen, J.A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. Genome Biol. Evol. **6**: 3295–3313.

Tenreira, T., Lange, M.J.P., Lange, T., Bres, C., Labadie, M., Monfort, A., Hernould, M., Rothan, C., and Denoyes, B. (2017). A Specific Gibberellin 20-Oxidase Dictates the Flowering-Runnering Decision in Diploid Strawberry. Plant Cell **29**: 2168–2182.

Tomlinson, M.J., Polson, S.W., Qiu, J., Lake, J.A., Lee, W., and Abasht, B. (2021). Investigation of allele specific expression in various tissues of broiler chickens using the detection tool VADT. Sci. Rep. **11**: 3968.

Trevaskis, B., Bagnall, D.J., Ellis, M.H., Peacock, W.J., and Dennis, E.S. (2003). MADS box genes control vernalization-induced flowering in cereals. Proc. Natl. Acad. Sci. U. S. A. **100**: 13099–13104.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., and Giovannoni, J. (2002). A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. Science **296**: 343–346.

Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-Wide Identification of Novel Imprinted Genes in Neonatal Mouse Brain. PLOS ONE **3**: e3839.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J.C., and Paterson, A.H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**: e49.

Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.-T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell **23**: 4221–4233.

Weijers, D., van Hamburg, J.-P., van Rijn, E., Hooykaas, P.J.J., and Offringa, R. (2003). Diphtheria Toxin-Mediated Cell Ablation Reveals Interregional Communication during Arabidopsis Seed Development. Plant Physiol. **133**: 1882–1892.

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag: New York, NY).

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLOS Genet. **7**: e1002126.

Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., Kasahara, H., Kamiya, Y., Chory, J., and Zhao, Y. (2011). Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **108**: 18518–18523.

Workman, R., Timp, W., and Fedak, R. High Molecular Weight DNA Extraction from Recalcitrant Plant Species for Third Generation Sequencing.: 6.

Yan, D., Duermeyer, L., Leoveanu, C., and Nambara, E. (2014). The Functions of the Endosperm During Seed Germination. Plant Cell Physiol. **55**: 1521–1533.

Zhan, J., Li, G., Ryu, C.-H., Ma, C., Zhang, S., Lloyd, A., Hunter, B.G., Larkins, B.A., Drews, G.N., Wang, X., and Yadegari, R. (2018). Opaque-2 Regulates a Complex Gene Network Associated with Cell Differentiation and Storage Functions of Maize Endosperm. Plant Cell **30**: 2425–2446.

Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X., Jiao, Y., Song, W., and Lai, J. (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc. Natl. Acad. Sci. **108**: 20042–20047.

Zhao, P., Begcy, K., Dresselhaus, T., and Sun, M.-X. (2017). Does Early Embryogenesis in Eudicots and Monocots Involve the Same Mechanism and Molecular Players?1[OPEN]. Plant Physiol. **173**: 130–142.

Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N., and Perou, C.M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics **15**: 419.

Zhou, J., Sittmann, J., Guo, L., Xiao, Y., Huang, X., Pulapaka, A., and Liu, Z. (2021). Gibberellin and auxin signaling genes RGA1 and ARF8 repress accessory fruit initiation in diploid strawberry. Plant Physiol. **185**: 1059–1075.

Zhou, J., Wang, G., and Liu, Z. (2018). Efficient genome editing of wild strawberry genes, vector development and validation. Plant Biotechnol. J. **16**: 1868–1877.

Zimin, A.V. and Salzberg, S.L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput. Biol. **16**: e1007981.

Zrimec, J., Börlin, C.S., Buric, F., Muhammad, A.S., Chen, R., Siewers, V., Verendel, V., Nielsen, J., Töpel, M., and Zelezniak, A. (2020). Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. Nat. Commun. **11**: 6141.