Plant Physiology®

Plant Physiology, 2024, **196**, 2599–2613
https://doi.org/10.1093/plphys/kiae496
Advance access publication 27 September 2024
Research Article

Identification and analysis of imprinted genes in wild strawberry uncover a regulatory pathway in endosperm development

Dirk Joldersma,[1,†] Lei Guo,[1,†] Elizabeth I. Alger,[2] Christina Ippoliti,[1] Xi Luo,[1] Adrian E. Platts,[2] Patrick P. Edger,[2] Zhongchi Liu[1,*,‡]

[1]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA
[2]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA
*Author for correspondence: zliu@umd.edu (Z.L.)
†Equal contribution.
‡Present address: Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, Shenzhen 518055, China.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Zhongchi Liu (zliu@umd.edu).

Abstract

Fertilization is a fundamental process that triggers seed and fruit development, but the molecular mechanisms underlying fertilization-induced seed development are poorly understood. Previous research has established *AGamous-Like62* (*AGL62*) activation and auxin biosynthesis in the endosperm as key events following fertilization in Arabidopsis (*Arabidopsis thaliana*) and wild strawberry (*Fragaria vesca*). To test the hypothesis that epigenetic mechanisms are critical in mediating the effect of fertilization on the activation of *AGL62* and auxin biosynthesis in the endosperm, we first identified and analyzed imprinted genes from the endosperm of wild strawberries. We isolated endosperm tissues from F1 seeds of 2 wild strawberry *F. vesca* subspecies, generated endosperm-enriched transcriptomes, and identified candidate Maternally Expressed and Paternally Expressed Genes (MEGs and PEGs). Through bioinformatic analyses, we identified 4 imprinted genes that may be involved in regulating the expression of *FveAGL62* and auxin biosynthesis genes. We conducted functional analysis of a maternally expressed gene *FveMYB98* through CRISPR-knockout and over-expression in transgenic strawberries as well as analysis in heterologous systems. FveMYB98 directly repressed *FveAGL62* at stage 3 endosperm, which likely serves to limit auxin synthesis and endosperm proliferation. These results provide an inroad into the regulation of early-stage seed development by imprinted genes in strawberries, suggest the potential function of imprinted genes in parental conflict, and identify *FveMYB98* as a regulator of a key transition point in endosperm development.

Introduction

Fertilization is a fundamental process essential for plant reproduction and agricultural productivity, and yet the molecular events following fertilization that leads to seed and fruit development are poorly understood. One major event induced by fertilization is the seed-synthesis of the phytohormone auxin, which in strawberries is sufficient to induce seed and fruit development (Gustafson 1936, 1939; Nitsch 1950; Kang et al. 2013; Guo et al. 2022). The necessity of auxin and auxin signaling components to successfully induce seed and fruit development is also established across numerous angiosperm plants (Joldersma and Liu 2018), indicating an evolutionarily conserved mechanism that inhibits auxin synthesis in ovules before fertilization.

In *Angiosperm* plants, seeds arise from the female gametophyte after fertilization by the male pollen. The seed consists of 3 distinct tissues: the embryo that develops from the fertilized egg cell, the seed coat that is derived from the maternal tissue surrounding the female gametophyte, and the endosperm, a triploid tissue derived from the fertilization of the diploid central cell in the female gametophyte. Sometimes compared with the mammalian placenta, endosperm's major function is to mediate nutrient exchange between the maternal plant and the developing embryo (Figueiredo et al. 2015). Importantly, the endosperm is also the site of auxin synthesis needed for seed and fruit development. In diploid strawberry *Fragaria vesca*, several auxin biosynthetic genes are substantially upregulated in the endosperm following fertilization (Kang et al. 2013; Feng et al. 2019). In Arabidopsis, triple loss-of-function mutants of auxin biosynthesis genes exhibited severe defects in endosperm proliferation, while increased auxin biosynthesis in the endosperm prevents its cellularization (Figueiredo et al. 2015; Batista et al. 2019). Furthermore, exogenous application of auxin or expression of auxin synthesis genes under a central cell-specific promoter could induce autonomous endosperm proliferation even in the absence of fertilization (Figueiredo et al. 2015). These studies highlighted a central role of auxin in endosperm development but left a major unanswered question how does the union of parental genomes trigger post-fertilization events, including auxin synthesis in the endosperm.

In Arabidopsis, a class of mutants exhibit autonomous seed development in the absence of fertilization. These fertilization-independent mutants were shown to be defective in components of the Polycomb Group Repressive Complex 2 (PRC2) (Ohad et al. 1996; Chaudhury et al. 1997; Kiyosue et al. 1999; Köhler and Lafon-Placette 2015), suggesting involvement of epigenetic mechanisms in preventing auxin synthesis in unfertilized female

Received July 15, 2024. Accepted August 26, 2024.

gametophytes. In a recent study involving wild strawberry (*F. vesca*) and Arabidopsis (*Arabidopsis thaliana*), *AGamous-Like62* (*AGL62*), a MADS-box transcription factor, was shown to be induced by fertilization in the endosperm; subsequently, *AGL62* induces auxin biosynthesis in the endosperm (Guo et al. 2022). Interestingly *AtAGL62* was shown to be a direct target of Fertilization-Independent Seed2 (FIS2)-PRC2 (Hehenberger et al. 2012), suggesting the involvement of epigenetic mechanisms in fertilization-induced activation of *AGL62* in the endosperm.

Genetic imprinting is an epigenetic phenomenon when certain genes are primarily expressed from the maternal or the paternal genome and hence termed Maternally Expressed or Paternally Expressed Genes (MEGs or PEGs). In plants, the endosperm is the major site of genetic imprinting or parent-dependent patterns of gene expression (Wolff et al. 2011; Del Toro-De León and Köhler 2019). Imprinted genes have been identified from a wide variety of plants including tomato (*Solanum lycopersicum*), rapeseed (*Brassica napus*), and monkeyflower (*Mimulus guttatus*) (Liu et al. 2018; Roth et al. 2018; Flores-Vergara et al. 2020). In 2011 alone, 6 genome-wide RNA-seq-based surveys for parental imprinting in endosperm were conducted including 3 in Arabidopsis, 2 in maize (*Zea mays*), and 1 in rice (*Oryza sativa*) (Gehring et al. 2011; Hsieh et al. 2011; Luo et al. 2011; Waters et al. 2011; Wolff et al. 2011; Zhang et al. 2011). Each of these studies identified at least 1 enzyme in the canonical Tryptophan Aminotransferase of Arabidopsis (TAA)-YUCCA (YUC) auxin biosynthetic pathway as a paternally expressed gene. Since increased auxin levels were previously shown to prevent endosperm cellularization (Batista et al. 2019), the paternally biased expression of auxin biosynthesis genes may act to promote progeny growth and delay endosperm cellularization. In a recent study, a cluster of repressive Auxin Response Factors (ARFs) in Arabidopsis were shown to be MEGs and acted to block auxin signaling to enable endosperm cellularization (Butel et al. 2024). The identification of paternally expressed auxin biosynthesis genes and maternally expressed auxin signaling blockers are consistent with the parental conflict theory (Haig and Westoby 1989), which posits that, while the paternal interest is to maximize the investment of maternal resources in his progeny survival, the maternal interest seeks to moderate her progeny's consumption of maternal resources, as these resources alternatively could be invested in maternal survival. This differing interest of parents is also supported by recent studies that show smaller seed size in a mutant of a maize gene *Dosage-Effect Defective1* (*DED1*), a PEG, and larger seeds in mutants of *Ethylene INsensitive 2* (*EIN2*), an MEG in Arabidopsis (Dai et al. 2022; Ando et al. 2023).

Based on these prior studies, we asked if imprinted genes could be the primary effectors of fertilization by regulating *AGL62* and/or auxin biosynthesis genes. We previously assembled the genome of wild strawberry "Yellow Wonder 5AF7" (Joldersma et al. 2022), and here, we assembled a second genome of subspecies *bracteata* CFRA502. With the availability of both genomes, we made F1 hybrids between these 2 *F. vesca* subspecies, isolated RNAs from F1 endosperms, and identified candidate imprinted genes. Through filtering and bioinformatic analysis, we identified 4 candidate regulators of *FveAGL62* and auxin biosynthetic genes. We then focused on characterizing *F. vesca MYB98* (*FveMYB98*), an R2-R3 MYB transcription factor and an MEG. Both CRISPR-knockout and over-expression (OE) of *FveMYB98* affected the expression of *FveAGL62* and auxin biosynthesis genes in the endosperm and thereby seed size. The results are consistent with the parental conflict theory, which predicts that MEGs should seek to dampen the progeny's consumption of maternal resources through repression of auxin biosynthesis. Our results provide an inroad into the regulation of seed development by imprinted genes in strawberry and validated FveMYB98's function in regulating *AGL62* expression and seed size.

Results

Identification of imprinted genes in *F. vesca*

Two *F. vesca* subspecies, *bracteata* CFRA502 and Yellow Wonder (YW5AF7) were chosen for the genetic hybridization to produce F1 hybrid seeds. YW5AF7 served as the female parent by employing a male-sterile mutant (*fvelfy-U600*) (Materials and Methods). The F1 hybrid seeds were harvested at 7 to 9 Days After Pollination (7 to 9 DAP) and manually dissected under a microscope to isolate the hybrid endosperm by excluding embryos and peeling off the seed coat. As hybrid seeds from above cross appeared to develop slower and somewhat asynchronously; the earliest stages we could dissect endosperms were at 7 to 9 DPA. The resulting endosperm was isolated, pooled, and frozen. The reciprocal cross with *bracteata* CFRA502 as female and YW5AF7 as male yielded very few viable seeds (Supplementary Fig. S1), resulting in too little endosperm tissue for RNA-Seq. Therefore, we proceeded with F1 endosperm RNA from YW5AF7 (female) × CRFA502 (male) cross only.

RNA-seq reads were aligned to the maternal and paternal reference genomes based on Single Nucleotide Polymorphisms that distinguish the parental genomes (Babak et al. 2008; Wang et al. 2008; DeVeale et al. 2012). First, we sequenced, assembled, and annotated the *Fragaria vesca bracteata* CFRA502 genome (see Materials and methods), and then, a "synthetic" genome was built that consists of the 2 parental genomes of YW5AF7 (Joldersma et al. 2022) and CFRA502. Over 82% of total reads aligned to the synthetic genome (Fig. 1A). In most cases (74.3% of the mapped reads), each read was aligned to homologous maternal and paternal loci; these "multimapping" reads were discarded (Fig. 1A) as they are not informative of the parent-of-origin. In contrast, the remaining 25.7% of the mapped reads aligned exclusively to the paternal or the maternal genome (i.e. mapped to a single locus in the synthetic genome) and were saved (Fig. 1A).

These "uniquely mapped" reads were counted using the FeatureCounts software (Liao et al. 2014) and categorized as MEGs or PEGs. As summarized in Fig. 1B, the *y*-axis indicates the number of loci that possess a specific percentage of maternal-derived reads shown in the *x*-axis, which ranges from 0 to 100%. The highest number of loci (the peak) is at 0.67, reflecting that most genes have close to a 2:1 maternal: paternal expression ratio (Fig. 1B), consistent with that the endosperm is the product of a fusion between the diploid (2n) maternal central cell and the haploid (1n) paternal sperm. Strikingly, large tails below 0.1 in blue and above 0.9 in red represent loci with significant parent-of-origin expression biases (Fig. 1B); candidate PEGs (0 to 0.1 on *x*-axis) and MEGs (0.9 to 1.0 on *x*-axis) were identified by the loci within the blue and red tails of the graph, respectively (Fig. 1B). The PEG and MEG lists from 3 samples were intersected in a Venn diagram leading to 996 and 1,276 common candidate PEGs and MEGs, respectively (Fig. 1C).

To evaluate the statistical significance of the observed deviation from the expected 0.67 ratio for loci at the 2 tails of Fig. 1B, we used a 1-way binomial test to evaluate the hypothesis that at a maternal locus, the ratio of maternal genome mapping reads to paternal genome mapping reads exceeds 0.666 with a *P*-value cut off at 0.001 to select imprinted genes; we performed the

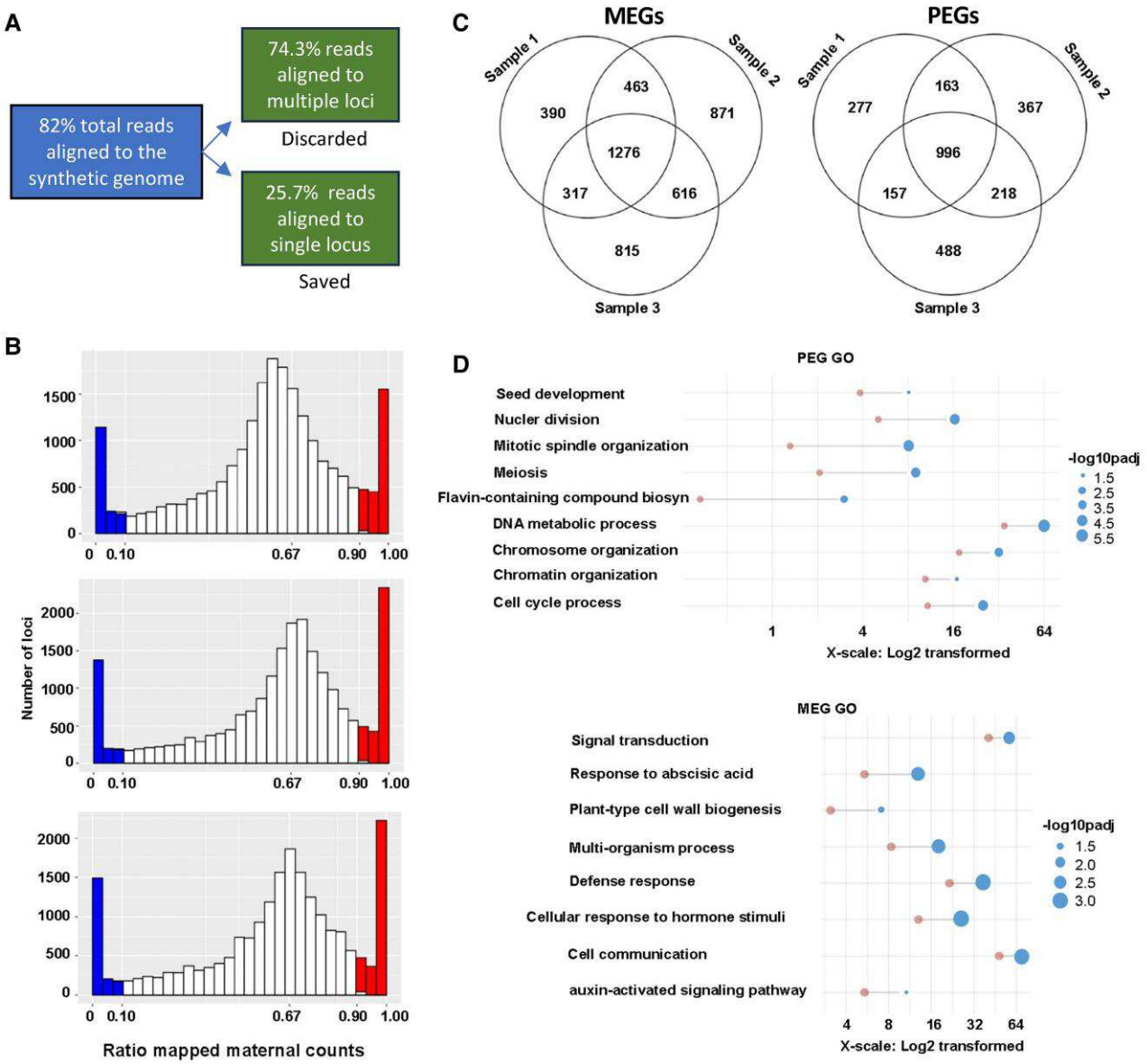

Figure 1. Identification of Maternally Expressed Genes (MEGs) and Paternally Expressed Genes (PEGs). **A)** Analysis pipeline used to identify uniquely mapped reads (either to maternal or paternal genome). **B)** Histograms showing the number of genomic loci (y-axis) exhibiting specified percentage of maternal expression (x-axis). Histograms calculated in R Studio, plotted in ggplot2. Three endosperm samples (replicates) are individually calculated and shown. **C)** Venn diagram showing the number of MEGs and PEGs identified by individual and overlapping endosperm samples. **D)** Gene ontology enrichment of PEGs and MEGs in strawberries. Left dots are expected gene number, right dots indicate actual gene numbers observed, and the size of the dots indicates the statistical significance of the enrichment in adjusted *P*-value. x-Axis is the number of genes, and x-axis scale is log2 transformed.

opposite test (<0.666) at paternal loci. This process identified 825 candidate MEGs and 809 candidate PEGs (Supplementary Table S1; Supplementary Data S1). Due to a lack of reciprocal crosses, genes with accession-biased expression (rather than the parent-of-origin bias) could not be excluded from the list. Hence, we refer to these genes as candidate MEGs and PEGs.

Identification of endosperm-enriched genes in *F. vesca*

Previously, Kang et al. (2013) generated RNA-seq data from "seed coat + endosperm (ghost)" tissue. Differential gene expression (DE) analysis was conducted between our 7 to 9 DAP endosperm transcriptome and the previous ghost transcriptome at similar stages (stage 3), which identified 3,314 endosperm-enriched (upregulated in endosperm by at least 2-fold) and 2,697 seed coat-enriched genes [downregulated in endosperm by at least 2-fold (Supplementary Data Set 2; Fig. S2A)]. *FveAGL80* (FvH4_6g21170), whose expression is increased 10-fold in the *F. vesca* endosperm in comparison to ghost, was shown by a reporter GUS to be specifically expressed in the endosperm of Arabidopsis and *F. vesca* (Picard et al. 2021; Guo et al. 2022). In addition, 917 or nearly one-third of *F. vesca* endosperm-enriched genes overlap with the Arabidopsis endosperm-specific genes (Supplementary Fig. S2B) (Picard et al. 2021), suggesting conservation in expression and regulation. GO enrichment analysis shows that the endosperm-enriched genes are highly focused on translation, DNA replication, chromatin remodeling, and cell proliferation

Table 1. Imprinted genes in auxin, ethylene and chromatin pathways

Category	Locus name	Gene ID	Imprinting status	Paternal counts	Maternal counts	Total counts	Binomial test P-value*
Auxin homeostasis and signaling	DAO1a	FvH4_1g09830	MEG	688	15,843	16,531	0
	DAO1b	FvH4_4g14690	MEG	14	730	744	9.30E−107
	IAGLU1	FvH4_4g13000	MEG	13	270	283	1.70E−32
	IAGLU2	FvH4_4g13010	MEG	0	100	100	2.40E−18
	YUC10	FvH4_2g24750	MEG	8	1125	1133	7.20E−183
	ARF6	FvH4-6g39140	PEG	10,023	340	10,363	0
	ARF17	FvH4_6g15960	PEG	280	2	282	4.80E−130
	ARF16	FvH4_5g3870	PEG	315	10	325	3.00E−134
Ethylene signaling	EIN2N	FvH4_6g16040	MEG	646	6879	7525	0
Chromatin modifiers	HAT1	FvH4_2g22590	PEG	2910	236	3146	0
	ASHH3	FvH4_1g00050	PEG	1236	0	1236	0

*P-value measures the likelihood of a deviation by chance above (in the case of MEGs) or below (PEGs) the predicted ratio of 2 maternal reads per paternal read.

(Supplementary Fig. S2D; Supplementary Data S3). In contrast, seed coat-enriched genes are over-represented in terms of interspecies interactions, defense responses, rejection of self, and cell killing (Supplementary Fig. S2C; Supplementary Data S3). These enriched GO terms are similarly observed in the Arabidopsis endosperm and seed coat, respectively (Belmonte et al. 2013).

Analysis of strawberry MEGs and PEGs

From the MEG gene list (Supplementary Data S1), we identified several genes that function in the regulation or homeostasis of auxin (Table 1). There are 4 *F. vesca* homologs of the Arabidopsis *Dioxygenase for Auxin Oxidation 1* (*DAO1*) (*AT1G14130*), coding for a Fe^{2+} dioxygenase that converts active auxin to inactive 2-oxoindole-3-acetic acid (Porco et al. 2016). However, 2 *FveDAO1* homologs (*FvH4_1g09830* and *FvH4_4g14690*) together contributed to 99% of the *FveDAO1* reads from the endosperm, and more than 90% of these reads aligned to the maternal YW5AF7 genome (Table 1), indicating that these 2 *FveDAO1* genes are likely MEGs. Second, indole-3-acetic acid-glucose synthase (IAGLU) reversibly inactivates auxin through conjugating auxin to glucose (Jackson et al. 2001). Two *FveIAGLU* genes (*FvH4_4g13000* and *FvH4_4g13010*) are maternally expressed (Table 1), although *FvH4_4g13000* has limited reads and an insignificant *P*-value in one of the 3 samples. As increased auxin catabolism could delay seed development, these MEGs may limit progeny growth as predicted by the parental conflict theory. One interesting observation is that *FveYUC10* (*FvH4_2g24750*), identified as a MEG in *F. vesca* (Table 1; Supplementary Data S1), was previously identified as a PEG in Arabidopsis and rice (Luo et al. 2011).

Several *FveARFs* are PEGs (Table 1) including *FvH4_6g39140/ARF6*, *FvH4_6g15960/ARF17*, and *FvH4_5g38770/FveARF16*. *FveARF6* (*FvH4_6g39140*), with ~10,000 mapped reads, is by far the most highly expressed and imprinted *FveARF*. Interestingly, *AtARF6* is also identified as a PEG in Arabidopsis (Pignatta et al. 2014). Knockout mutants of a maternally expressed *mir167*, a cognate microRNA of *ARF6*, exhibited a maternally controlled effect on embryo and endosperm development, and this effect was mediated by increased *ARF6/ARF8* (Yao et al. 2019).

GO Enrichment of the 809 PEGs identified enriched cell cycle-related genes and derivative processes such as nuclear division or mitotic spindle regulation as well as biosynthesis of flavin-containing compounds (Fig. 1D; Supplementary Data S4), consistent with PEGs promoting a proliferative endosperm. On the other hand, MEGs show enriched signal transduction, response to abscisic acid (ABA), a known factor in seed senescence (Jaradat et al. 2013) and cell wall biogenesis (Fig. 1D; Supplementary Data S4).

Identification of candidate PEGs and MEGs with roles in early-stage endosperm development

In the wild type, soon after fertilization (at stage 2), *FveAGL62* is activated leading to auxin synthesis and endosperm nuclear proliferation as a syncytium; at stage 3, *FveAGL62* expression is turned off, and the endosperm nuclei undergo cellularization (Guo et al. 2022). In strawberry *fveagl62* mutants, the reduced auxin in endosperm led to precocious cellularization as early as stage 2, limiting endosperm growth. This precocious cellularization in *fveagl62* can be partially reversed by the application of auxin (Guo et al. 2022). Similar observations were made in Arabidopsis, where increased auxin was shown to delay cellularization and cause larger seeds (Batista et al. 2019), while a cluster of repressive ARFs blocked auxin signaling to enable endosperm cellularization (Butel et al. 2024). Therefore, precise regulation of *AGL62* expression and/or auxin biosynthesis during endosperm development serves to control seed size, and the PEGs and MEGs identified here may shed light on how such regulation is executed.

We adopted a 3-step strategy to identify MEGs and PEGs that may bind the promoters of *FveAGL62* or auxin biosynthesis genes (Supplementary Fig. S3). In step one, we mined PlantPan 3.0 (Chow et al. 2019) using 1000-bp upstream sequence of *FveAGL62*, which identifies 569 *F. vesca* candidate transcription factors that have binding sites in the upstream sequence (Supplementary Data S5). In step 2, we searched the DAP-seq database of Arabidopsis transcription factors (Bartlett et al. 2017) and identified 157 Arabidopsis transcription factors that bind the promoter of the Arabidopsis *AGL62*; from these 157 Arabidopsis genes, we identified 115 orthologous *F. vesca* genes based on a prior ortholog assignment (Li et al. 2022). Finally, the intersection between the 2 lists was compared with the *F. vesca* candidate imprinted gene list. This led to 2 PEGs and 2 MEGs genes (Table 2 in bold). One PEG (*FveTCX2/FvH4_4g00820*) encodes a homolog of the *TESMIN/TSO1-like CXC2/AT4G14770* gene in Arabidopsis (Klepikova et al. 2016), and the other PEG, *FvH4_1g27900*, is a Nam-ATAF1/2-Cuc2 (NAC) transcription factor homologous to the silique-expressed gene *AT5G18270* (Klepikova et al. 2016). The 2 MEGs are *FveMYB98/FvH4_6g51000* and *Fragaria vesca Arabidopsis Thaliana Homeobox7* (*FveATHB7*)/*FvH4_7g32570*, homologs of *AtMYB98/AT4G18770* and *ATHB7/AT2G46680*, respectively (Swarbreck et al. 2008; Berardini et al. 2015).

The same 3-step approach was applied to each of the 4 auxin biosynthesis genes expressed in the *F. vesca* endosperm, *FveTAA1*

Table 2. Summary of candidate TF genes that may regulate *FveAGL62* and/or auxin biosynthesis

PEG/ MEG	FvH4 Gene ID	Gene Symbol[a]	At homolog	Target promoter (*F. vesca*)
PEG	FvH4_5g29460	NAC101	AT5G62380	YUC11
	FvH4_4g00820	**TCX2**	AT4G14770	TAA1, YUC10, AGL62
	FVH4_1G27900	**NAC87**	AT5G18270	YUC11, YUC10, TAR1, AGL62
	FVH4-3G19410	CUC2	AT5G53950	TAR1
MEG	FvH4_7g17540	NAC43	AT2G46770	YUC11
	FvH4_3g06200	WRKY28	AT4G18170	TAA1, TAR1
	FvH4_3g08750	ERF4	AT3G15210	YUC11
	FvH4_6g33050	VNAC1	AT2G18060	YUC11
	FvH4_6g51000	**MYB98**	AT4G18770	TAA1, AGL62
	FvH4_7g32570	**ATHB7**	AT2G46680	YUC10, YUC11, AGL62
	FvH4_2g36350	NAC38	AT2G24430	YUC10, YUC11, TAA1, TAR1
	FvH4_5g511930	MYB62	AT1G68320	YUC10, TAA1
	FvH4_7g17540	NAC43/ NST1	AT2G46770	YUC10, TAR1

[a]Bold names are candidate transcription factors that bind *F. vesca AGL62* promoter.

(*FvH4_4g25850*), *Fragaria vesca Tryptophan Aminotransferase Related1* (*FveTAR1*) (*FvH4_5g05900*), *FveYUC11* (*FvH4_4g17980*), and *FveYUC10* (*FvH4_2g24750*), which led to lists of candidate *F. vesca* MEGs or PEGs that may bind each of the promoters (Supplementary Data S6). Intersection analyses using the pipeline illustrated in Supplementary Fig. S3 led to 4 PEGs and nine MEGs (Table 2). Strikingly, while 7 of these genes appear to regulate multiple auxin biosynthetic genes, 3 of the 4 genes identified as potential *FveAGL62* regulators are among this limited set of 7 genes that regulate multiple auxin biosynthesis genes (Table 2), suggesting that *FveAGL62* and auxin biosynthesis genes may share similar regulators.

Characterizing *FveMYB98*, an MEG

Previous studies of the Arabidopsis *AtMYB98* (*AT4G18770*) gene revealed its ubiquitous expression in flowers and seeds (Klepikova et al. 2016). Further, prior analysis of *Arabidopsis myb98* mutants suggested an essential function for synergid cell function of pollen tube guidance in the female gametophyte (Kasahara et al. 2005; Punwani et al. 2007, 2008; Susaki et al. 2021). However, the post-fertilization function of *AtMYB98* is unknown.

The strawberry homolog of AtMYB98 is FvH4_6g51000 (FveMYB98), which shares a high sequence similarity with AtMYB98 in the R2 and R3 DNA binding domains (Supplementary Fig. S4). Interestingly, in strawberry, *FveMYB98* is expressed only after fertilization and predominantly in the ghost (endosperm + seed coat) at stages 3 and 4 (Supplementary Fig. S5) (Hollender et al. 2012; Kang et al. 2013), which is when *FveAGL62* expression goes away, raising the possibility that *FveMYB98* may act to repress *FveAGL62* expression in the endosperm. Additionally, *FveMYB98* is bioinformatically identified as a candidate binding factor of both *FveAGL62* and *FveTAA1* (Table 2). Therefore, the function of *FveMYB98* in relation to *FveAGL62* and auxin synthesis becomes a primary focus.

We generated CRISPR-knockouts of *FveMYB98* by designing two sgRNAs targeting the coding region of *FveMYB98* (Fig. 2A). The CRISPR vector was transformed into H4 (a standard WT accession that is easier to transform than YW5AF7), and the resulting 15 independent T0 transgenic plant lines were analyzed. Two plants were found to contain 136-bp deletions (−136; −136) between sgRNA1 and sgRNA2, and they were named *fvemyb98-1* (Fig. 2A). Three other plants were homozygous for 1-bp deletion (−1; −1) at the sgRNA1 site; they were named as *fvemyb98-2*. Both alleles likely cause complete loss of the *FveMYB98* function. In addition, we over-expressed *FveMYB98* cDNA in an H4 background using Ubiquitin 10 promoter and obtained 20 independent *FveMYB98* OE lines. RT-qPCR showed that while the *FveMYB98* transcript is reduced in *fvemyb98-1*, perhaps as a result of nonsense-mediated decay, *FveMYB98 OE* lines showed a significantly higher expression than WT (H4; Fig. 2B).

We then examined the phenotype of *fvemyb98-1* and the 3 *FveMYB8 OE* lines (#4, #8, and #15). While *fvemyb98-1* plants appeared wild type (Fig. 2C), all 3 *FveMYB98 OE* lines exhibited a dwarf stature and developed smaller leaves than WT (H4) (Fig. 2C). Since *FveMYB98* is normally expressed only in the ghost, their vegetative phenotype in OE lines may not be directly relevant.

To determine if changes in *FveMYB98* activity lead to abnormal seed development, we dissected achenes and measured the seed size of WT, *fvemyb98-1*, and an *FveMYB98 OE* line (Fig. 2D). *fvemyb98-1* mutant seeds were significantly longer than the wild type (but with similar seed width). In contrast, the *FveMYB98 OE* line developed seeds that were significantly shorter and narrower than WT (Fig. 2D). This function would be consistent with the parental conflict theory, where the maternally derived *FveMYB98* seeks to restrict progeny growth to conserve resources.

To determine if the effect of *FveMYB98* on seed size is primarily determined by the maternal genotype, we performed reciprocal crosses between WT and *fvemyb98-1*, both of which are in the H4 background. When *fvemyb98-1* ovule is fertilized with WT pollen, seed length is significantly longer (average at 1149.98 mm) than the reciprocal cross with WT ovule fertilized by the *fvemyb98-1* pollen (average at 1108.5 mm) or WT ovule fertilized by WT pollen (also at 1103.93 mm; Fig. 2E). The data support a dominant effect of the maternal genotype at *FveMYB98* due to maternally biased expression. Nevertheless, we could not conclude with certainty that *FveMYB98* acted exclusively in the endosperm; RNA-seq data showed normalized average reads of 9,484 in the endosperm and 5,992 in the ghost (endosperm + seed coat), suggesting that seed coat may also express *FveMYB98*, albeit at a lower level than the endosperm.

The mechanism of *FveMYB98* function in seeds

How does an altered *FveMYB98* activity lead to altered seed size? Could it be caused by altered cell proliferation and/or timing of endosperm cellularization? Using confocal microscopy, we examined stage 3 and 4 seeds of WT, *fvemyb98-1*, and *FveMYB98 OE* seeds (Fig. 3). At stage 3 (6 to 7 DPA) characterized by a heart-shape embryo, WT seeds contain 1 to 2 layers of cellularized endosperm surrounding the central vacuole; *FveMYB98 OE* seeds appear similar to the WT. In contrast, many *fvemyb98-1* stage 3 seeds possess an endosperm that fills the entire space without a visible central vacuole (Fig. 3), suggesting more endosperm cells in the seed. At stage 4 (8 to 10 DPA) characterized by a "walking-stick" embryo, WT, *FveMYB98 OE*, and *fvemyb98-1* endosperms appear morphologically similar but exhibit varying degrees of endosperm volume being consumed by the embryo. Overall, our observation supports an over-proliferated endosperm in *fvemyb98-1* stage 3 seeds.

Could the increased endosperm proliferation observed in *fvemyb98-1* seeds be caused by an increased auxin level? The *DR5ver2::β-Glucuronidase (GUS)* reports auxin response level, which

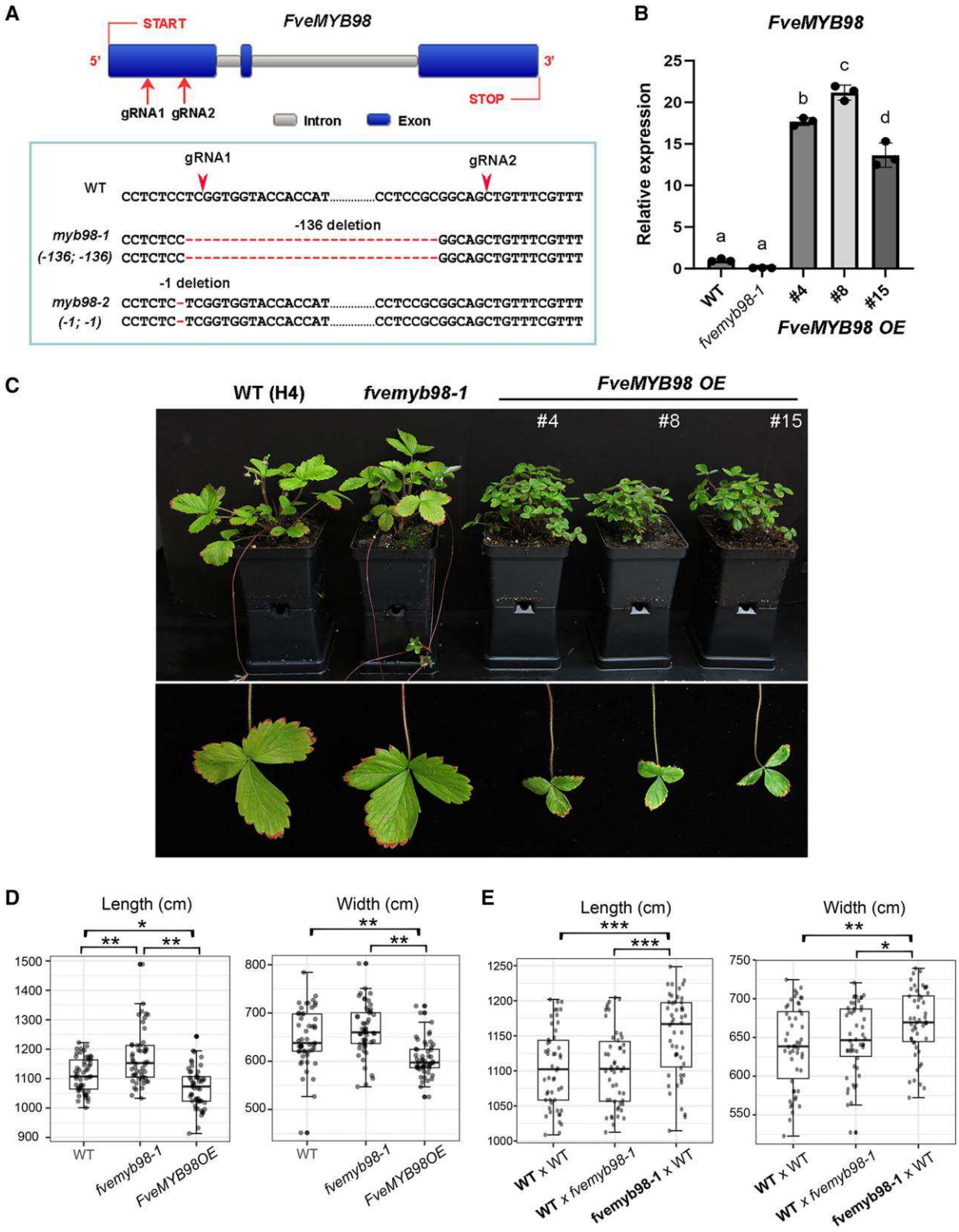

Figure 2. Characterization of *fvemyb98-1* and *FveMYB98 OE* phenotype. **A)** *FveMYB98* gene diagram and the CRISPR-mediated deletion between the 2 sgRNAs. Dash signs indicate deletions. **B)** RT-qPCR measurement of *FveMYB98* in stage 3 seeds ($N = 30$) showing reduced expression of *FveMYB98* in *fvemyb98-1* and increased expression in all 3 *FveMYB98 OE* lines. Statistically significant differences are labeled with letters using 1-way ANOVA followed by Turkey's test, $P < 0.05$. Error bars are used to indicate SD. **C)** Whole plant and leaf phenotypes of *fvemyb98-1* and *FveMYB98 OE* lines. **D)** Stage 5 seed length and width in different genotypes. $N = \sim 50$ seeds of each genotype. **E)** Reciprocal crosses to confirm the dominant effect of the maternal genotype on seed size. Maternal genotype is in bold. In each boxplot, center line indicates median; upper and lower box border indicates first and third quartiles, respectively; whiskers, 1.5× interquartile range; each dot, 1 sample. Significance is indicated by * ($P < 0.05$) and ** ($P < 0.01$) with a 2-tailed Student's *t*-test. All genotypes (WT, *fvemyb98-1*, and *FveMYB98 OE*) are in the H4 background.

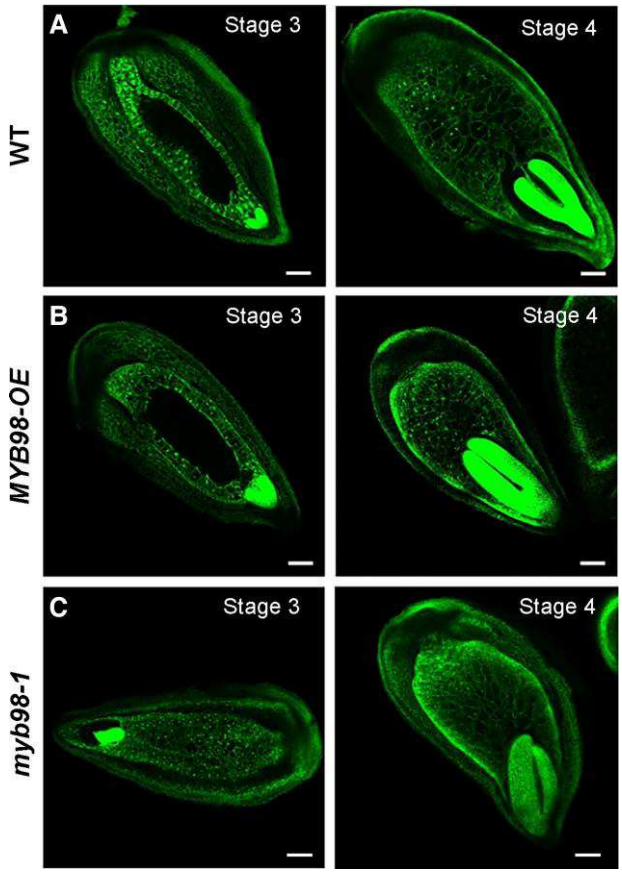

Figure 3. Confocal microscopy reveals a more proliferated endosperm in stage 3 *fvemyb98-1* seeds. **A)** stage 3 (left) and stage 4 (right) WT seeds. **B)** stage 3 (left) and stage 4 (right) *FveMYB98 OE* seeds. **C)** stage 3 (left) and stage 4 (right) *fvemyb98-1* seeds. Staging is based on embryo morphology defined previously (Hollender et al. 2012) with stage 3 (6 to 7 DPA) seeds characterized by possessing a heart-shape embryo and stage 4 (8 to 10 DPA) seeds characterized by a "walking-stick" embryo. All genotypes are in the H4 background. Scale bar: 100 μm.

indirectly reflects the auxin level. We crossed an established *DR5ver2::GUS* reporter line (Feng et al. 2019) into *fvemyb98-1* and compared GUS staining intensity in stage 3 seeds of WT and *fvemyb98-1* sibling plants (Fig. 4A). GUS staining appeared more intense in *fvemyb98-1* seeds than the WT seeds, suggesting that *FveMYB98* may negatively regulate auxin accumulation in stage 3 seeds to limit endosperm proliferation at this stage.

We further hypothesized that *FveMYB98* may repress, directly or indirectly (for example via *FveAGL62*), the expression of auxin biosynthesis genes in the endosperm to reduce auxin accumulation. RT-qPCR showed that *FveTAA1*, *FveYUC10*, and *FveTAR1* transcript levels are respectively 5.16, 2.14, and 7.32-fold higher in *fvemyb98-1* mutant seeds when compared with WT seeds at stage 3 (Fig. 4B), which corroborates with the increased *DR5ver2::GUS* expression in the *fvemyb98-1* mutant seeds (Fig. 4A) and supports a negative regulatory role of *FveMYB98* for auxin synthesis. In the *FveMYB98* OE lines, *FveTAA1* and *FveYUC10* transcripts were reduced (Fig. 4B) but *FveTAR1* expression remained the same or slightly increased. Since OE could interfere with other wild-type gene functions and may not reflect the normal situation in WT, we emphasized the loss-of-function phenotype of *fvemyb98-1*.

Yeast one-hybrid assay was used to determine if FveMYB98 could directly bind the promoters of these auxin synthesis genes. None of the promoters of *FveTAA1*, *FveYUC10*, and *FveTAR1* could be bound by FveMYB98-AD (Fig. 4C), indicating, *inter alia*, that *FveMYB98* may indirectly regulate the expression of auxin synthesis genes or that FveMYB98 may require other plant co-factors to bind these promoters.

Transient luciferase assay was conducted in *Nicotiana benthamiana* leaves. The relative luciferase levels of *pFveTAA1::LUC* and *pFveYUC10::LUC* were reduced by 10-fold or more when *FveMYB98* was transiently expressed in the *N. benthamiana* leaves in comparison to the control GFP (Fig. 4D). Together, these assays indicate that FveMYB98 represses the expression of auxin biosynthesis genes in the seed (in vivo) based on Reverse Transcription quantitative PCR (RT-qPCR; Fig. 4B) but could not determine if the repression is direct or indirect due to limitations of results derived from heterologous tissues and organisms (Fig. 4C, D).

FveMYB98 may directly repress *FveAGL62*

In addition to auxin synthesis genes, *FveAGL62* is a candidate direct target of *FveMYB98* (Table 2), and as noted previously, *FveAGL62* and *FveMYB98* exhibit temporal complementary expression in seeds (Fig. 5A). Specifically and based on RNA-seq data, *FveAGL62* expression is turned on at stage 2 but off at stage 3; on the contrary, *FveMYB98* expression is off at stage 2 but on at stage 3. We used RT-qPCR to test *FveAGL62* expression in stage 2 and stage 3 seeds in *fvemyb98-1*; while *FveAGL62* expression in *fvemyb98-1* mutant is similar to WT at stage 2, it is increased significantly at stage 3, reaching 64.27 fold of WT level at stage 3 (Fig. 5B, left most graph). *FveAGL62* expression was similarly tested in stage 3 seeds of 3 *FveMYB98 OE* lines; as *FveAGL62* expression is already low in stage 3 WT seeds, it is almost undetectable in *FveMYB98 OE* lines (Fig. 5B). The data support a role of *FveMYB98* in turning off *FveAGL62* expression in stage 3 seeds.

To test if FveMYB98 binds to the *FveAGL62* promoter, yeast 1 hybrid was conducted; FveMYB98-AD was able to bind the *FveAGL62* promoter (Fig. 5B, middle panel). Furthermore, in transient luciferase assay, *FveAGL62* promoter activity was reduced by about 3.7-fold in the presence of FveMYB98 when compared with the control (Fig. 5B). Therefore, the data support direct repression of *FveAGL62* expression by FveMYB98 in stage 3 seed.

Discussion

We have generated an endosperm-enriched transcriptome from 7 to 9 DAP seeds of diploid strawberry *F. vesca*. Taking advantage of the available reference genome of YW5AF7 and the assembly of the CFRA502 genome in this study, we produced F1 hybrid seeds from YW5AF7 female crossed with CFRA502 pollen and isolated F1 endosperm tissues. We employed RNA-seq to profile the endosperm transcriptome and subsequently identified 809 and 825 genes that respectively show paternal and maternal-biased expression in the F1 endosperm (Supplementary Data S1). As one-directional cross (YW5AF7 female crossed with CFRA502 male) could not distinguish accession-biased expression from parent-of-origin-biased expression, the genes identified are only candidate PEGs and MEGs and require further confirmation. In humans (Homo sapiens [Baran et al. 2015]) and recently in monkeyflower (Flores-Vergara et al. 2020), imprinting studies were also conducted despite the one-directional cross limitation. Therefore, our study, like those prior, offers valuable insights into paternal- and maternal-biased expression, that form the basis for further genetic and functional validation.

Evidence supporting parental conflict theory

The identities of many of the PEGs and MEGs support the parental conflict theory, as many MEGs possess predicted roles in

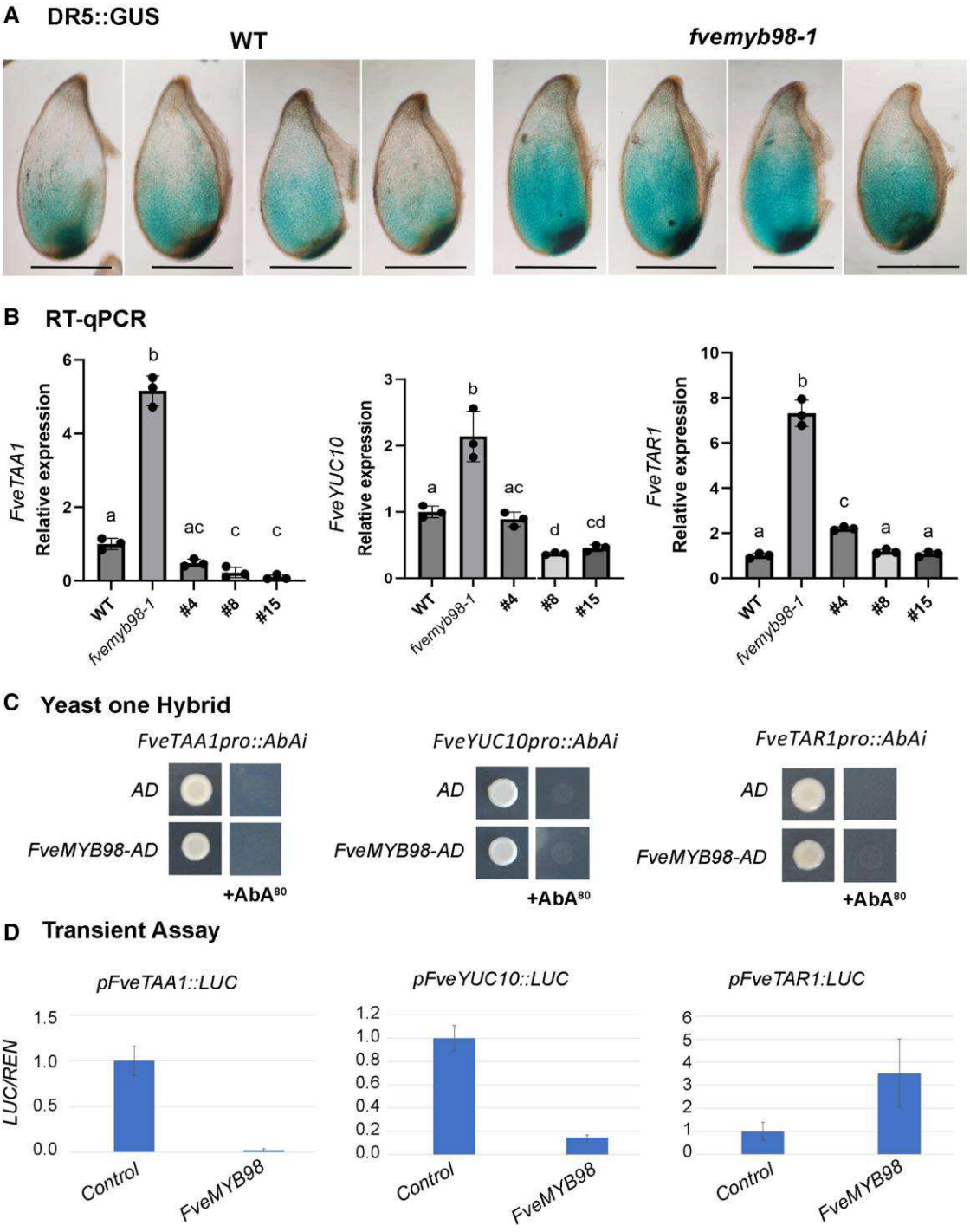

Figure 4. Examining the relationship between *FveMYB98* and auxin. **A)** DR5::GUS staining in WT and *fvemyb98-1* stage 3 seeds. Scale bar, 500 μm. **B)** RT-qPCR measurement of auxin biosynthesis gene expression in stage 3 seeds ($N = 30$) of different genotypes. #4, #8, and #15 are *FveMYB98* OE lines. Statistically significant differences are labeled with letters using 1-way ANOVA followed by Turkey's test, $P < 0.05$. Error bars are used to indicate SD. **C)** Yeast 1-hybrid assay testing the binding of FveMYB98 to the promoters of auxin biosynthesis genes. Absence of growth in the selection medium containing AbA indicates no binding in all tested promoters. **D)** Transient LUC assay showing reduction of *FveTAA1* and *FveYUC10* reporter expression in the presence of *FveMYB98* when compared with the control, which is a vector that expresses GFP. The *y*-axis is the relative luciferase value normalized against the Renilla luciferase (*35S:REN*). Error bars are used to indicate SD.

restricting progeny growth while the PEGs possess predicted functions in promoting progeny growth. For example, a large number of candidate MEGs encode enzymes involved in the degradation or conjugation of auxin (Table 1), effectively reducing auxin accumulation and hence limiting progeny growth. Further, we identified and characterized a maternally expressed *FveMYB98*, which encodes a negative regulator of *FveAGL62* and auxin accumulation.

Interestingly, *YUC10* was previously shown to be a *PEG* in Arabidopsis, maize, and rice, leading to the hypothesis that its induction in the endosperm soon after fertilization could serve as a signal to indicate fertilization success and initiate endosperm

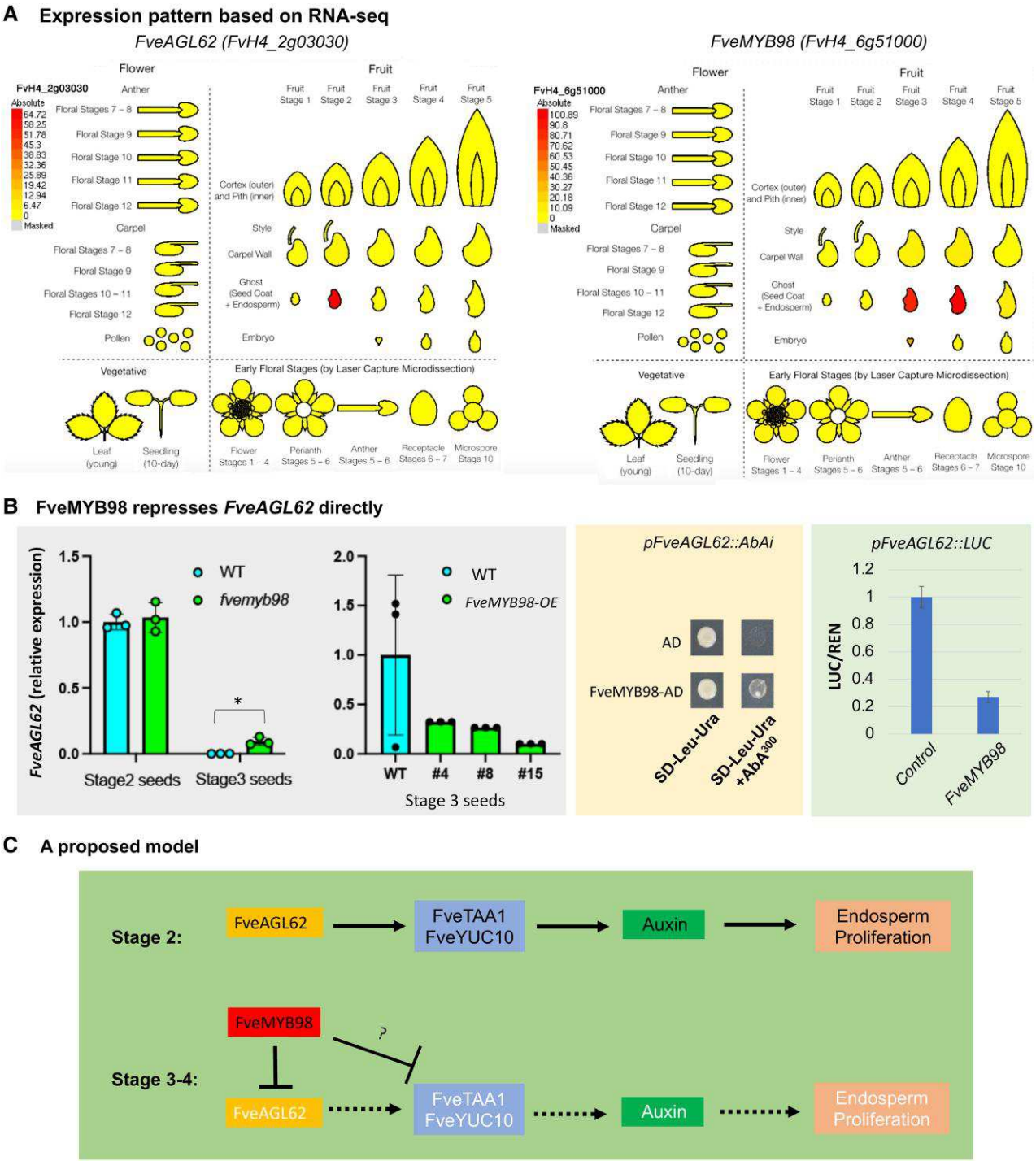

Figure 5. FveMYB98 negatively regulates *FveAGL62* at stage 3 seeds. **A)** RNA-Seq data displayed by the eFP browser, showing highly tissue- and stage-specific expression of *FveAGL62* and *FveMYB98*. Ghost is the dissected seed containing endosperm + seed coat but lacking embryo; data is from Kang et al. (2013) and images are from https://bar.utoronto.ca/efp_strawberry/cgi-bin/efpWeb.cgi. **B)** RT-qPCR showing relative expression of *FveAGL62* in different genotypes and stages, yeast-1-hybrid assay showing direct binding of FveMYB98-AD (AD: activation domain) to the promoter of *FveAGL62*, and transient and relative LUC reporter expression via the *FveAGL62* promoter. Significant difference in relative *FveAGL62* expression (2-tailed Student's *t*-test) is marked with * ($P < 0.05$). Error bars are used to indicate SD. **C)** A proposed model summarizing the data. In the early endosperm (stage 2), *FveAGL62* is induced by fertilization and in turn induces the expression of auxin biosynthesis genes *FveTAA1* and *FveYUC10*. The resulting increase in auxin concentration promotes endosperm proliferation. At stage 3 and 4 endosperm, *FveMYB98* expression is induced, and it represses the expression of *FveAGL62* leading to reduced expression of auxin biosynthesis genes and auxin level, and subsequently reduced endosperm proliferation. Dotted arrows or lines imply a loss of action. Arrowed lines indicate activation or promotion; lines with a perpendicular line at the end indicate repression. The question mark indicates uncertainty.

development (Figueiredo and Köhler 2018). This hypothesis also implies that the silencing of the maternal allele of *YUC10* serves to blockade premature endosperm development in the unfertilized female gametophyte. Here, we identified *FveYUC10* as an MEG. The switch in imprinting status should not be a surprise, as the imprinting status of genes frequently is not conserved across species (Rodrigues and Zilberman 2015). Perhaps, in strawberry, maternal expression of the rate-limiting auxin biosynthetic gene *FveYUC10* is in effect another way that the maternal genome regulates and limits seed development.

FveMYB98 reduces seed size by repressing *FveAGL62* at a key transition point during endosperm development

Previously, we showed that *FveAGL62* expression is induced at stage 2 in the endosperm soon after fertilization. This fertilization-induced *FveAGL62* expression was required for the activation of auxin synthesis in the endosperm and the promotion of endosperm nuclear proliferation (Guo et al. 2022). At stage 3, *FveAGL62* expression is turned off by unknown mechanisms, and subsequently, the endosperm transitions from nuclear proliferation to cellularization. In *fveagl62* loss-of-function mutants with reduced auxin, endosperm enters cellularization precociously even at stage 2 (Guo et al. 2022).

The maternally expressed *FveMYB98* exhibits a temporal expression pattern complementary to that of *FveAGL62* (Fig. 5A); when *FveMYB98* expression is switched on at stage 3, *FveAGL62* expression is switched off. Could *FveMYB98* be the reason for *FveAGL62*'s silencing at stage 3? Functional and molecular analysis of *FveMYB98* revealed a direct and negative role of *FveMYB98* in regulating *FveAGL62* expression. We propose a model (Fig. 5C) that the activation of *FveMYB98* at stage 3 is a key transition point during endosperm development; it turns off *FveAGL62* to bring down auxin and slow down endosperm proliferation. Although our initial intent of the study was to identify the upstream regulatory factors that mediate the fertilization signal to the activation of *FveAGL62*, we instead identified *FveMYB98* as a terminator of *FveAGL62* expression at a key transition point during endosperm development.

Our study also identified auxin biosynthesis genes as downstream regulatory targets of *FveMYB98*, although this regulation could be indirect (Fig. 5C). As strawberry *fveagl62* mutants were previously shown to promote the expression of auxin biosynthesis genes (Guo et al. 2022), the effect of *FveMYB98* on auxin biosynthesis could be mediated by *FveAGL62* (Fig. 5C). Future experiments by constructing *fvemyb98; fveagl62* double mutants would allow us to determine the genetic epistasis between these 2 genes and determine if the effect of *FveMYB98* is partially or completely mediated by *FveAGL62*. The identification of the *FveMYB98-FveAGL62* regulatory relationship laid the foundation for further investigations into this key transition point in endosperm development. As temporal regulation of this transition is directly relevant to seed size regulation, our work could inform strategies for increasing grain size and crop yield.

Materials and methods

F. vesca accessions used in the study

Fragaria vesca bracteata CFRA502 seeds were obtained from Dr. Kim Hummer from the USDA ARS National Clonal Germplasm Repository in Corvallis, Oregon. *Fragaria vesca* accession "Yellow Wonder" (YW5AF7) and Hawaii 4 (H4) were previously described (Slovin et al. 2009; Hawkins et al. 2016). All *F. vesca* plants were cultivated under standard long-day conditions 16-h light at 25 °C followed by 8-h darkness at 22 °C with a relative humidity of 50% (Hollender et al. 2012; Guo et al. 2022).

Identification of a fvelfy-U600 mutant in *F. vesca*

An N-ethyl-N-nitrosourea (ENU) mutagenesis screen in the YW5AF7 background was described previously (Caruana et al. 2018). From the M2 population of this ENU screen, a mutant (U600) was identified and is male sterile. Because of its identical floral phenotype to other *F. vesca fvelfy* mutants (Zhang et al. 2023), we amplified and sequenced the *FveLFY* gene from this U600 mutant and showed that U600 carries a Thr^{245} to Met missense mutation in a highly conserved region. This *fvelfy-U600* allele is identical to the *A. thaliana lfy-3* in nucleotide change that was shown to exhibit an intermediate phenotype (Weigel et al. 1992).

Hybrid seed generation, endosperm collection, and RNA-sequencing

Pollen from CFRA502 was used to pollinate the *fvelfy-U600* stigma. Achenes were collected at 7 to 9 DAP. Endosperm was hand dissected from seeds under a Zeiss Stemi SV 6 microscope equipped with an AxioCam digital camera. The dissected endosperm was stored at −80 °C and was eventually pooled into 3 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. Libraries were constructed using Ultra-low input mRNA nondirectional library preparation and sequenced with an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA). Three biological replicates of paired-end, 150-bp reads were obtained at 21.8 to 36.3 GB per sample; 90% of nucleotide calls had a quality score of more than 30. The 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 77% of the reads to a unique locus in the reference H4 genome [FvH4v4.0.a1 (Edger et al. 2018; Li et al. 2019)] downloaded from the Genome Database for Rosaceae [GDR (Jung et al. 2019)]. At least 1 transcript was mapped to 28,531 annotated loci in at least one of the samples and at least 20 transcripts to 24,719 loci.

Fragaria vesca bracteata CFRA502 genome assembly and annotation

CFRA502 plants were grown in a growth chamber in the Michigan State University growth chamber facility under 16-h light at 23 °C, 8-h dark at 20 °C, and 60% relative humidity. High molecular weight DNA from CFRA502 was extracted from young leaves after 72 h of dark treatment using a published method (Vaillancourt and Buell 2019). RNA was extracted from young leaves, old leaves, dark-treated young leaves, dark-treated old leaves, methyl jasmonate-treated leaves, roots, shoots, flowers, and flowers before anthesis. All tissues, except dark-treated, were collected from CFRA 502 plants at 12 PM (zeitgeber time 6 h) and extracted using MagMAXTM-96 Total RNA Isolation Kit (Thermo Fisher).

The CFRA502 genome was assembled from 11.2GBase (c. 46x) of demultiplexed fastq PacBio HiFi reads [https://www.pacb.com/technology/hifi-sequencing/] using HiFiasm [https://doi.org/10.1038/s41592-020-01056-5] with a default Kmer length of 51. The assembly was further organized using Ragtag [https://doi.org/10.1186/

s13059-022-02823-7] (options correct "-f 10,000 –remove-small" and scaffold) to be structurally homologous to that of the *F. vesca* Hawaii 4 (H4) assembly (Edger et al. 2018). The resulting assembly was 244 MBase in length, comprising primarily seven pseudomolecules (224 MBase) as well as 395 unplaced contigs. Annotation of the assembly utilized Illumina short read transcriptome data, isolated from the aforementioned 10 distinct library types, assembled into preliminary transcripts using Trinity [https://www.nature.com/articles/nprot.2013.084] in reference-guided mode. These models were combined with transcript data from H4 (as altest's), a custom repeat library generated by repeatModeler [https://www.repeatmasker.org/RepeatModeler/] as well as Augustus [https://bioinf.uni-greifswald.de/augustus/], SNAP [https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-5-59], and geneMark [http://opal.biology.gatech.edu/GeneMark/] in-silico models as inputs for Maker [https://www.yandell-lab.org/software/maker.html]. The Maker annotation was further refined with AGAT [https://github.com/NBISweden/AGAT] to correct some GFF3 errors and was then renamed.

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

A synthetic genome of FvYW5AF7 v1.0 (Joldersma et al. 2022) and FvCFRA502 v1.0 described in this study was created by appending the FvCFRA502 genome to the FvYW5AF7 genome using the ">>" operator in unix. LiftOff (Shumate and Salzberg 2020) was used to lift the FvYW5AF7 v1.0 genome annotation to FvCFRA502 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 1 genomic loci be discarded. Reads were counted using the 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 3 samples >20 and that in each of the 3 samples >90% or <10% of total reads aligned to the maternal FvYW5AF7 v1.0 genome for the MEG and PEG assignment, respectively. To evaluate the statistical significance of these assignments, we used the R command binom.test ("X" "Y", $P = 0.6666$, alternative = "less" or "greater"), in which X was the number of reads uniquely mapped to the maternal genome and Y was the sum of X plus the number of reads uniquely mapped to the paternal genome and in which "less" was used to test PEGs and "greater" was used to test MEGs. Genes with a P-value < 0.001 were considered candidate imprinted genes. Genes with zero reads in one or more samples also were not included as PEGs or MEGs, as they were not amenable to statistical testing.

Genome-wide analysis to identify endosperm-specific genes in *F. vesca*

DE analysis was conducted by DESeq2 v1.22.2 (Love et al. 2014) between the endosperm RNA-seq data and the previous seed and fruit RNA-seq data (Kang et al. 2013), which was downloaded from NCBI's Sequence Read Archive (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 principal component analysis. Intersection analysis was performed in R Studio using the built-in intersection function.

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 GDR (Jung et al. 2019). The GO categories enriched for each gene module in the networks were determined by Fisher's exact test in R package TopGO v2.34.0 (Alexa et al. 2006). The P-value cutoff was set at 0.05. The dot plot was created with the R package ggplot2 (Wickham 2009).

Analysis of MEGs and PEGs

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 Hawaii4 reference genome (v4.0.a1) and submitted to PlantPan3.0 online portal for promoter analysis (http://plantpan.itps.ncku.edu.tw/).

Arabidopsis DAP-seq results for Arabidopsis genes *AT1G21430*, *AT1G48910*, *AT1G70560*, and *AT4G24670* were downloaded from the Plant Cistrome (http://neomorph.salk.edu/dap_web/pages/index.php) (O'Malley et al. 2016; Bartlett et al. 2017).

Generate FveMYB98 knockout and OE in transgenic strawberry plants

To generate *FveMYB98* CRISPR construct, gRNA1 (ATGGTGG TACCACCGAGGAGAGG) and gRNA2 (AAACGAAACAGCTGCCG CGGAGG) 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 were transformed into the strawberry H4 accession 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 coding sequence (CDS) of *FveMYB98* was PCR amplified from cDNA using primers (Supplementary Table S2) 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. More than 20 independent *FveMYB98 OE* lines were confirmed by PCR.

Characterize the seed phenotype of FveMYB98 knockout and OE lines

To assess seed size, stage 3 achenes were collected from multiple fruits derived from multiple *fvemyb98*, *FveMYB98 OE*, and wild-type plants; all genotypes are in the H4 background. Seeds were imaged using a Zeiss Stemi SV 6 microscope equipped with an AxioCam digital camera. Seed length and width were quantified using the "Length" function of AxioVision (Release 4.9.1) imaging software.

Confocal imaging was performed as described previously (Guo et al. 2022). A Leica STELLARIS DLS Confocal microscope (Leica Co. USA) with 488-nm excitation was used to image. The autofluorescence signal of the seeds was detected at 498 to 530 nm.

Gene expression analysis using RT-qPCR

Total RNA was extracted from stage 3 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 1 representative biological experiment (with 3 technical repeats). Biological replicates (2 to 3 times) gave similar results.

Yeast one-hybrid and LUC transient expression assay

Yeast one-hybrid assay was conducted based on the Matchmaker Gold Yeast One-Hybrid Library Screening System User Manual (PT4087-1). Full-length CDS of FveMYB98 were PCR amplified and ligated into pGADT7 (Clontech Inc.) at the EcoRI/BamHI digestion sites by Gibson assembly (Gibson et al. 2009). The promoter sequence of each candidate genes (*FveAGL62*: 1512 bp; *FveYUC10*: 1200 bp; *FveTAR1*: 1404 bp; *FveTAA1*: 1604 bp) 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, the promoter sequence of each candidate gene (*FveAGL62*: 1512 bp; *FveYUC5*: 1268 bp; *FveYUC10*: 2614 bp; *FveTAA1*: 2000 bp) 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 2 to 3 times with 1 representative result shown in Fig. 4D.

Accession numbers

Sequence data from this article can be found in the GDR (rosaceae.org) under their gene IDs (F. vesca H4 genome version 4.0 annotation version 2.0) as follows: FveMYB98 (FvH4_6g51000), FveAGL62 (FvH4_2g03030), FveYUC10 (FvH4_2g24750), FveYUC11 (FvH4_4g17980), FveTAR1 (FvH4_5g05900), FveTAA1 (FvH4_4g25850), and FvePP2a (FvH4_4g27700).

Acknowledgments

The authors would like to thank Dr. Michael Hardigan for help in selecting CFRA502 as a parent for the study, Dr. Kim Hummer for CRFA502 seeds, and Dr. Stephen Mount for guidance in data analysis.

Author contributions

D.J., L.G., and Z.L. designed research; D.J., L.G., C.I., and X.L. performed research; E.I.A., A.E.P., and P.P.E. assembled the CFRA502 genome; J.D., L.G., and Z.L. analyzed data; and J.D. and Z.L. wrote the paper.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. Illustration of F1 seeds and achenes derived from a cross between CFRA502 female and YW5AF7 pollen.

Supplementary Figure S2. Identification and analysis of endosperm-enriched genes.

Supplementary Figure S3. Illustration of the 3-step strategy in identifying MEGs or PEGs that may bind to the promoter of *FveAGL62*.

Supplementary Figure S4. Gene structure and protein sequence comparison between AtMYB98 and FveMYB98.

Supplementary Figure S5. Screenshots of eFP browser illustrating the specific expression pattern of *FveMYB98* and its potential regulatory targets.

Supplementary Table S1. Effects of different binomial test *P*-value cutoff on the number of imprinted genes.

Supplementary Table S2. Primers used in this study.

Supplementary Data S1. Candidate *F. vesca* PEGs and MEGs.

Supplementary Data S2. Endosperm- and seed coat-enriched genes.

Supplementary Data S3. Enriched GOs for endosperm- and seed coat-enriched genes.

Supplementary Data S4. Enriched GOs for *F. vesca* PEGs and MEGs.

Supplementary Data S5. Candidate transcriptional regulators of FveAGL62.

Supplementary Data S6. Candidate transcriptional regulator of FveYUC10, FveYUC11, FveTAA1, and FveTAR1.

Funding

This work was supported by the University of Maryland's Faculty-Student Research Collaboration Grant as well as by NSF-IOS 1444987 to Z.L. D.J. was previously supported by the University of Maryland CMNS Dean's Matching Award that is associated with the NIH T32 Molecular and Cell Biology Training Grant, a Hockmeyer summer fellowship, and a dissertation fellowship provided by the University of Maryland's Biological Science Graduate Program. P.P.E., A.P., and E.I.A. were supported by NSF-PGRP 2029959.

Conflict of interest statement. The authors declare no competing interests.

Data availability

The *CFRA502* raw sequence data have been deposited in the Short Read Archive under NCBI BioProject ID PRJNA695578. The *CFRA502* genome assembly, annotations, and other supporting data are available on the Genome Database for Rosaceae (www.rosaceae.org) and the CyVerse CoGe (Haug-Baltzell et al. 2017). The YW5AF7 F1 endosperm RNA-seq data were deposited at the Short Read Archive under NCBI BioProject ID PRJNA871213.

References

Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006:22(13):1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Ando A, Kirkbride RC, Qiao H, Chen ZJ. Endosperm and maternal-specific expression of EIN2 in the endosperm affects endosperm cellularization and seed size in *Arabidopsis*. *Genetics*. 2023:223(2):iyac161. https://doi.org/10.1093/genetics/iyac161

Babak T, Deveale B, Armour C, Raymond C, Cleary MA, van der Kooy D, Johnson JM, Lim LP. Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol*. 2008:18(22):1735–1741. https://doi.org/10.1016/j.cub.2008.09.044

Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, Rivas MA, Pirinen M, Gutierrez-Arcelus M, Smith KS, Kukurba KR, *et al.* The

landscape of genomic imprinting across diverse adult human tissues. *Genome Res*. 2015:25(7):927–936. https://doi.org/10.1101/gr.192278.115

Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc*. 2017:12(8):1659. https://doi.org/10.1038/nprot.2017.055

Batista RA, Figueiredo DD, Santos-González J, Köhler C. Auxin regulates endosperm cellularization in *Arabidopsis*. *Genes Dev*. 2019:33(7–8):466–476. https://doi.org/10.1101/gad.316554.118

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, *et al*. Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc Natl Acad Sci U S A*. 2013:110(5):E435–E444. https://doi.org/10.1073/pnas.1222061110

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. The arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*. 2015:53(8):474–485. https://doi.org/10.1002/dvg.22877

Butel N, Qiu Y, Xu W, Santos-González J, Köhler C. Parental conflict driven regulation of endosperm cellularization by a family of Auxin Response Factors. *Nat. Plants*. 2024:10(6):1018–1026. https://doi.org/10.1038/s41477-024-01706-y

Caruana JC, Sittmann JW, Wang W, Liu Z. Suppressor of runnerless encodes a DELLA protein that controls runner formation for asexual reproduction in strawberry. *Mol Plant*. 2018:11(1):230–233. https://doi.org/10.1016/j.molp.2017.11.001

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ. Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 1997:94(8):4223–4228. https://doi.org/10.1073/pnas.94.8.4223

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chow C-N, Lee T-Y, Hung Y-C, Li G-Z, Tseng K-C, Liu Y-H, Kuo P-L, Zheng H-Q, Chang W-C. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res*. 2019:47(D1):D1155–D1163. https://doi.org/10.1093/nar/gky1081

Dai D, Mudunkothge JS, Galli M, Char SN, Davenport R, Zhou X, Gustin JL, Spielbauer G, Zhang J, Barbazuk WB, *et al*. Paternal imprinting of dosage-effect defective1 contributes to seed weight xenia in maize. *Nat Commun*. 2022:13(1):5366. https://doi.org/10.1038/s41467-022-33055-9

Del Toro-De León G, Köhler C. Endosperm-specific transcriptome analysis by applying the INTACT system. *Plant Reprod*. 2019:32(1):55–61. https://doi.org/10.1007/s00497-018-00356-3

DeVeale B, van der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA–seq: a new perspective. *PLoS Genet*. 2012:8(3):e1002600. https://doi.org/10.1371/journal.pgen.1002600

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, *et al*. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. *GigaScience*. 2018:7(2):1–7. https://doi.org/10.1093/gigascience/gix124

Feng J, Dai C, Luo H, Han Y, Liu Z, Kang C. Reporter gene expression reveals precise auxin synthesis sites during fruit and root development in wild strawberry. *J Exp Bot*. 2019:70(2):563–574. https://doi.org/10.1093/jxb/ery384

Figueiredo DD, Batista RA, Roszak PJ, Köhler C. Auxin production couples endosperm development to fertilization. *Nat Plants*. 2015:1(12):nplants2015184. https://doi.org/10.1038/nplants.2015.184

Figueiredo DD, Köhler C. Auxin: a molecular trigger of seed development. *Genes Dev*. 2018:32(7–8):479–490. https://doi.org/10.1101/gad.312546.118

Flores-Vergara MA, Oneal E, Costa M, Villarino G, Roberts C, De Luis Balaguer MA, Coimbra S, Willis J, Franks RG. Developmental analysis of mimulus seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. *Front Plant Sci*. 2020:11:132. https://doi.org/10.3389/fpls.2020.00132

Gehring M, Missirian V, Henikoff S. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS One*. 2011:6(8):e23687. https://doi.org/10.1371/journal.pone.0023687

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009:6(5):343–345. https://doi.org/10.1038/nmeth.1318

Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*. 2007:2(1):31–34. https://doi.org/10.1038/nprot.2007.13

Guo L, Luo X, Li M, Joldersma D, Plunkert M, Liu Z. Mechanism of fertilization-induced auxin synthesis in the endosperm for seed and fruit development. *Nat Commun*. 2022:13(1):3985. https://doi.org/10.1038/s41467-022-31656-y

Gustafson FG. Inducement of fruit development by growth-promoting chemicals. *Proc Natl Acad Sci U S A*. 1936:22(11):628–636. https://doi.org/10.1073/pnas.22.11.628

Gustafson FG. The cause of natural parthenocarpy. *Am J Bot*. 1939:26(3):135–138. https://doi.org/10.1002/j.1537-2197.1939.tb12880.x

Haig D, Westoby M. Parent-specific gene expression and the triploid endosperm. *Am Nat*. 1989:134(1):147–155. https://doi.org/10.1086/284971

Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*. 2017:33(14):2197–2198. https://doi.org/10.1093/bioinformatics/btx144

Hawkins C, Caruana J, Schiksnis E, Liu Z. Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci Rep*. 2016:6(1):29017. https://doi.org/10.1038/srep29017

Hehenberger E, Kradolfer D, Köhler C. Endosperm cellularization defines an important developmental transition for embryo development. *Development*. 2012:139:2031–2039. https://doi.org/10.1242/dev.077057

Hollender CA, Geretz AC, Slovin JP, Liu Z. Flower and early fruit development in a diploid strawberry, *Fragaria vesca*. *Planta*. 2012:235(6):1123–1139. https://doi.org/10.1007/s00425-011-1562-1

Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, *et al*. Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Natl Acad Sci U S A*. 2011:108(5):1755–1762. https://doi.org/10.1073/pnas.1019273108

Jackson RG, Lim E-K, Li Y, Kowalczyk M, Sandberg G, Hoggett J, Ashford DA, Bowles DJ. Identification and biochemical characterization of anarabidopsis indole-3-acetic acid glucosyltransferase. *J Biol Chem*. 2001:276(6):4350–4356. https://doi.org/10.1074/jbc.M006185200

Jaradat MR, Feurtado JA, Huang D, Lu Y, Cutler AJ. Multiple roles of the transcription factor AtMYBR1/AtMYB44 in ABA signaling, stress responses, and leaf senescence. *BMC Plant Biol*. 2013:13(1): 192. https://doi.org/10.1186/1471-2229-13-192

Joldersma D, Liu Z. The making of virgin fruit: the molecular and genetic basis of parthenocarpy. *J Exp Bot*. 2018:69(5):955–962. https://doi.org/10.1093/jxb/erx446

Joldersma D, Sadowski N, Timp W, Liu Z, Joldersma D, Sadowski N, Timp W, Liu Z. Assembly and annotation of *Fragaria vesca* "Yellow Wonder" genome, a model diploid strawberry for molecular genetic research. *Fruit Res*. 2022:2(1):1–5. https://doi.org/10.48130/FruRes-2022-0013

Jung S, Lee T, Cheng C-H, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, *et al*. 15 years of GDR: new data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res*. 2019:47-(D1):D1137–D1145. https://doi.org/10.1093/nar/gky1000

Kang C, Darwish O, Geretz A, Shahan R, Alkharouf N, Liu Z. Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell*. 2013:25(6):1960–1978. https://doi.org/10.1105/tpc.113.111732

Kasahara RD, Portereiko MF, Sandaklie-Nikolova L, Rabiger DS, Drews GN. MYB98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell*. 2005:17(11): 2981–2992. https://doi.org/10.1105/tpc.105.034603

Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB, *et al*. Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 1999:96(7):4186–4191. https://doi.org/10.1073/pnas.96.7.4186

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-Seq profiling. *Plant J Cell Mol Biol*. 2016:88(6):1058–1070. https://doi.org/10.1111/tpj.13312

Köhler C, Lafon-Placette C. Evolution and function of epigenetic processes in the endosperm. *Front Plant Sci*. 2015:6:130. https://doi.org/10.3389/fpls.2015.00130

Li M, Galimba K, Xiao Y, Dardick C, Mount SM, Callahan A, Liu Z. Comparative transcriptomic analysis of apple and peach fruits: insights into fruit type specification. *Plant J Cell Mol Biol*. 2022:109(6):1614–1629. https://doi.org/10.1111/tpj.15633

Li Y, Pi M, Gao Q, Liu Z, Kang C. Updated annotation of the wild strawberry *Fragaria vesca* V4 genome. *Hortic Res*. 2019:6:61. https://doi.org/10.1038/s41438-019-0142-6

Liao Y, Smyth GK, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu J, Li J, Liu H, Fan S, Singh S, Zhou X-R, Hu Z, Wang H, Hua W. Genome-wide screening and analysis of imprinted genes in rapeseed (*Brassica napus* L.) endosperm. *DNA Res*. 2018:25(6):629–640. https://doi.org/10.1093/dnares/dsy030

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods*. 2001:25(4):402–408. https://doi.org/10.1006/meth.2001.1262

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. 2014:15(12):550. https://doi.org/10.1186/s13059-014-0550-8

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet*. 2011:7(6):e1002125. https://doi.org/10.1371/journal.pgen.1002125

Nitsch JP. Growth and morphogenesis of the strawberry as related to auxin. *Am J Bot*. 1950:37(3):211–215. https://doi.org/10.1002/j.1537-2197.1950.tb12183.x

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL. A mutation that allows endosperm development without fertilization. *Proc Natl Acad Sci U S A*. 1996:93(11):5319–5324. https://doi.org/10.1073/pnas.93.11.5319

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*. 2016:165(5): 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Picard CL, Povilus RA, Williams BP, Gehring M. Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. *Nat Plants*. 2021:7(6):730–738. https://doi.org/10.1038/s41477-021-00922-0

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife*. 2014:3:e03198. https://doi.org/10.7554/eLife.03198

Porco S, Pěnčík A, Rashed A, Voß U, Casanova-Sáez R, Bishopp A, Golebiowska A, Bhosale R, Swarup R, Swarup K, *et al*. Dioxygenase-encoding AtDAO1 gene controls IAA oxidation and homeostasis in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 2016: 113(39):11016–11021. https://doi.org/10.1073/pnas.1604375113

Punwani JA, Rabiger DS, Drews GN. MYB98 positively regulates a battery of synergid-expressed genes encoding filiform apparatus-localized proteins. *Plant Cell*. 2007:19(8):2557–2568. https://doi.org/10.1105/tpc.107.052076

Punwani JA, Rabiger DS, Lloyd A, Drews GN. The MYB98 subcircuit of the synergid gene regulatory network includes genes directly and indirectly regulated by MYB98. *Plant J*. 2008:55(3):406–414. https://doi.org/10.1111/j.1365-313X.2008.03514.x

Rodrigues JA, Zilberman D. Evolution and function of genomic imprinting in plants. *Genes Dev*. 2015:29(24):2517–2531. https://doi.org/10.1101/gad.269902.115

Roth M, Florez-Rueda AM, Paris M, Städler T. Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *Plant J*. 2018:95(6): 1084–1101. https://doi.org/10.1111/tpj.14012

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2020:37(12):1639–1643. https://doi.org/

Slovin JP, Schmitt K, Folta KM. An inbred line of the diploid strawberry *Fragaria vesca* f. semperflorens for genomic and molecular genetic studies in the Rosaceae. *Plant Methods*. 2009:5(1):15. https://doi.org/10.1186/1746-4811-5-15

Susaki D, Suzuki T, Maruyama D, Ueda M, Higashiyama T, Kurihara D. Dynamics of the cell fate specifications during female gametophyte development in Arabidopsis. *PLOS Biol*. 2021:19(3): e3001123. https://doi.org/10.1371/journal.pbio.3001123

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, *et al*. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res*. 2008:36(Database): D1009–D1014. https://doi.org/10.1093/nar/gkm965

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, *et al*. An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature*. 2015:517(7536):571–575. https://doi.org/10.1038/nature14099

Vaillancourt B, Buell CR. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. bioArxiv. https://doi.org/10.1101/783159, 26 September 2019, preprint: not peer reviewed.

Wang X, Sun Q, McGrath SD, Mardis ER, Soloway PD, Clark AG. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One*. 2008:3(12):e3839. https://doi.org/10.1371/journal.pone.0003839

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell*. 2011:23(12):4221–4233. https://doi.org/10.1105/tpc.111.092668

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM. LEAFY controls floral meristem identity in Arabidopsis. *Cell*. 1992:69(5):843–859. https://doi.org/10.1016/0092-86749290295-n

Wickham H. *Ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag; 2009.

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet*. 2011:7(6):e1002126. https://doi.org/10.1371/journal.pgen.1002126

Yao X, Chen J, Zhou J, Yu H, Ge C, Zhang M, Gao X, Dai X, Yang Z-N, Zhao Y. An essential role for miRNA167 in maternal control of embryonic and seed development. *Plant Physiol*. 2019:180(1):453–464. https://doi.org/10.1104/pp.19.00127

Zhan J, Li G, Ryu C-H, Ma C, Zhang S, Lloyd A, Hunter BG, Larkins BA, Drews GN, Wang X, *et al*. Opaque-2 regulates a complex gene network associated with cell differentiation and storage functions of maize endosperm. *Plant Cell*. 2018:30(10):2425–2446. https://doi.org/10.1105/tpc.18.00392

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, *et al*. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A*. 2011:108(50):20042–20047. https://doi.org/10.1073/pnas.1112186108

Zhang Y, Kan L, Hu S, Liu Z, Kang C. Roles and evolution of four LEAFY homologs in floral patterning and leaf development in woodland strawberry. *Plant Physiol*. 2023:192(1):240–255. https://doi.org/10.1093/plphys/kiad067

Zhou J, Sittmann J, Guo L, Xiao Y, Huang X, Pulapaka A, Liu Z. Gibberellin and auxin signaling genes RGA1 and ARF8 repress accessory fruit initiation in diploid strawberry. *Plant Physiol*. 2021:185(3):1059–1075. https://doi.org/10.1093/plphys/kiaa087

Zhou J, Wang G, Liu Z. Efficient genome editing of wild strawberry genes, vector development and validation. *Plant Biotechnol J*. 2018:16(11):1868–1877. https://doi.org/10.1111/pbi.12922