ABSTRACT

Title of Dissertation: A NETWORK APPROACH TO IDENTIFY KEY REGULATORS OF FRUIT DEVELOPMENT IN *FRAGARIA VESCA*, A DIPLOID STRAWBERRY

Rachel Maczis Shahan, Doctor of Philosophy, 2018

Dissertation directed by: Professor Zhongchi Liu, Department of Cell Biology and Molecular Genetics

Post-embryonic organogenesis is a feature unique to plants, an example of which is flower and fruit production. Previous work on strawberry fruit development has focused primarily on the latter stages, including ripening. Comparatively little is known about the molecular events underpinning fruit set, the pivotal stage at which fruit development either proceeds or terminates. This thesis investigates early fruit development using *Fragaria vesca*, a diploid strawberry, as a model.

In collaboration with a bioinformatician, I generated gene co-expression networks from 92 previously generated RNA-Seq libraries profiling multiple tissues and stages of strawberry flower and fruit development. I demonstrate the utility of co-expression networks in illuminating molecular processes underlying fruit development. Experimental validation of the networks includes demonstration of increased iron transport soon after fertilization

and identification of *FveUFO1* as an important regulator of floral meristem determinacy and floral organ identity.

Using the co-expression networks, I discovered the surprising expression of *FvFT1*, a homolog of *FLOWERING LOCUS T* (*FT*), in the fleshy fruit immediately post-fertilization. In many plant species, the FT peptide is a non-cell autonomous signal that initiates flowering in response to inductive photoperiod. I found that *FvFT1* expression is responsive to temperature, but not photoperiod, in strawberry fruit. Further, transcriptional activation is detectable in the vascular bundles connecting the fruit to the seeds, raising the possibility that *FvFT1* may facilitate cross-tissue communication. Signal from an FvFT1-GFP translational fusion protein is visible in seed nuclei despite its localized transcription in the vasculature. However, analysis of *FvFT1 RNAi* plants failed to identify a fruit phenotype, possibly due to redundancy among three *FvFT* paralogs.

Finally, to develop additional research tools for *F. vesca*, I isolated and tested fruit tissue-specific promoters based on genes identified with differential expression analyses. These analyses revealed genes strongly expressed in the receptacle fruit, thereby identifying potential regulators of early fruit development and attractive candidates for future study.

Together, this work advances the systems-level infrastructure for studying molecular regulation of *F. vesca* fruit development, points to a novel role for *FT* distinct from its known function in floral initiation, and provides molecular tools useful to the research community.

A NETWORK APPROACH TO IDENTIFY KEY REGULATORS OF FRUIT DEVELOPMENT IN *FRAGARIA VESCA*, A DIPLOID STRAWBERRY

by

Rachel Maczis Shahan

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
2018

Advisory Committee:
Professor Zhongchi Liu, Chair
Professor Caren Chang
Associate Professor Gary Coleman
Associate Professor Antony Jose
Professor Leslie Pick

Dedication

For my grandmother, Rachel J. Maczis, who told me to approach a challenge with sincerity and a good sense of humor.

Acknowledgements

I first thank my advisor, Dr. Zhongchi Liu, for her thoughtful mentoring. She challenged me to think critically, taught me to present my ideas clearly, and helped me navigate both success and failure. I am especially grateful for the guidance she provided in paper and proposal writing.

I thank my committee members for insightful comments, critiques, and encouraging conversations. I am grateful for strawberry transformation work by Delores Lomberk, Ceil Muller, and Mark Sperry of the USDA ARS lab at Epcot. Dr. Junhui Zhou, Dr. Yuexue Liu, and Dongdong Li provided assistance with CRISPR experiments. Tim Maugel assisted with TEM experiments.

My lab mates, both former and current, made coming to work each day a joy. Drs. Courtney Hollender, Chunying Kang, Dandan Li, Charles Hawkins, Wanpeng (George) Wang, Junhui Zhou, and Julie Caruana patiently taught me techniques and were always happy to help with experimental design. I especially thank Julie and George for being mentors as well as close friends. I thank John Sittmann for always knowing how to make me laugh.

I am grateful for the camaraderie and supportive environment cultivated by the members of the ground floor labs. I especially thank Dr. Heven Sze for mentoring me from my very first semester at UMD. I thank Drs. Caren Chang, José Feijó, Steve Mount, and Bill Snell for encouragement and conversations over lunch. Dr. John Clay and members of the Chang lab were always happy to provide advice and lend materials. Drs. Daniel Santa Cruz

Damineli, Maria Teresa Portes, and Micha Wudick from the Feijó lab provided support as I wrote my thesis.

I am grateful to Dr. Leslie Pick for her work to bring the CMB T32 training grant to UMD. Writing a research proposal and receiving support from the training grant in my second and third years was invaluable to my maturation from a young student to an increasingly confident scientist. I also acknowledge support from a USDA NIFA pre-doctoral fellowship that gave me the freedom to explore ideas and provided resources for presenting my work at several conferences.

Susan Park Ochsner, Sarah Ahlbrand, Cordelia Weiss, and Carey Stuart became my second family; their humor and friendship made grad school fun. I thank my parents, Brenda and Gary Maczis, for their love and for supporting my interest in science from childhood. My paternal grandparents, Lee and Rachel Maczis, allowed me to watch Bill Nye the Science Guy and build vinegar and baking soda volcanoes to my heart's content. My maternal grandparents, Daniel and Frances Clark, nurtured my love of biology and the outdoors. My sister, Rebecca, and cousin, Garrett, know just how to make me laugh at myself and remember not to take everything so seriously.

This thesis would not be possible without the love and support of my husband, Dustin. He has encouraged me through every step of grad school and makes me feel that I can accomplish anything, no matter how great the challenge.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
AG	*AGAMOUS*
an	*anantha*
AP1	*APETALA 1*
AP2	*APETALA 2*
AP3	*APETALA 3*
cDNA	Complementary DNA
CDS	Coding Sequence
CO	*CONSTANS*
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyl Trimethyl Ammonium Bromide
DAB	Diaminobenzidine
DPA	Days Post Anthesis
efo	*extra floral organs*
EMS	Ethyl methanesulfonate
F. vesca	*Fragaria vesca*
FD	FLOWERING LOCUS D
FM	Floral Meristem
FT	*FLOWERING LOCUS T*
FTIP1	*FT INTERACTING PROTEIN 1*
GA	Gibberellic Acid
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GI	*GIGANTEA*
GO	Gene Ontology
GUS	ß-Glucuronidase
H4	Hawaii-4
HD	Hand Dissection
IBA	Indole-3-butyric acid
LCM	Laser Capture Microdissection
LD	Long Day
LFY	*LEAFY*
NAA	1-Naphthaleneacetic acid
NPA	N-1-Naphthylphthalamic acid
PCR	Polymerase Chain Reaction
PI	*PISTILLATA*
PLB	Prolamellar Body
POR	Protochlorophyllide Oxidoreductase
qRT-PCR	Quantitative reverse transcription PCR
RNAi	RNA interference

RPKM	Reads per Kilobase Million
SAM	Shoot Apical Meristem
SD	Short Day
sgRNA	Single guide RNA
SNP	Single Nucleotide Polymorphism
SOCI	*SUPPRESSOR OF CONSTANS I*
SSB	Between Group Sum of Squares
SSE	Sum of Square Error
SUC2	*SUCROSE2*
T-DNA	Transfer DNA
TEM	Transmission Electron Microscopy
TFL1	*TERMINAL FLOWER 1*
TPM	Transcripts Per Million
TSS	Total Sum of Squares
UFO	Unusual Floral Organs
WGCNA	Weighted Gene Co-Expression Network Analysis
WT	Wild Type
YW	Yellow Wonder

Chapter 1: Introduction

A growing human population and global climate change both present major challenges for food security worldwide. By the year 2050, the human population is projected to reach 9-12 billion people, which will demand an increase in food production of 34-70% from today's levels (FAO 2009). Further, the average global temperature has risen by approximately 0.6 degrees Celsius over the past 100 years and is projected to continue ascending (Root et al., 2003). Both an increase in food demand and a less favorable climate for plant growth call for new knowledge and innovations in plant research. In almost all crop plants, successful fertilization, which is achieved after a flower is pollinated, is required for fruit and seed formation. However, pollen viability and normal fertilization processes may be hindered by high temperature (Peet et al., 1998) and a scarcity of pollinators. Without fertilization, flowers yield neither seeds nor fruits (Goetz et al., 2006).

Flowering and fruit set are both required for successful harvests and the subsequent nutrition and health of the human population. Understanding the basic molecular mechanisms underlying flower and fruit development is the first step towards development of new, innovative agricultural techniques to address the challenges of the future. My thesis focuses on investigating molecular mechanisms of fruit development in *Fragaria vesca (F. vesca),* a diploid strawberry.

1.1: *Flowering plants are indispensable to humans*

Angiosperms, colloquially called flowering plants, are interwoven with human civilization as major sources of textiles, timber, medicines, and crops. Angiosperms constitute the

largest of five extant groups of plants with at least 260,000 species grouped into 453 families; nearly 90% of the plant species found on earth today are angiosperms (Soltis and Soltis, 2004; Crepet and Niklas, 2009). Despite their relatively recent appearance ~130 million years ago in the fossil record, the morphological and ecological diversity of angiosperms is such that Darwin famously referred to the evolutionary mechanism behind their rapid radiation as an 'abominable mystery' (discussed by Friedman, 2009). Angiosperms are found in nearly all habitats on earth, excluding only extreme environments like the poles and the deepest oceans (Soltis and Soltis, 2004). Flowering plants include epiphytes that parasitize other plants, freely floating and rooted aquatic species, and many terrestrial species that vary in size, lifespan, and growth habit, including trees, woody shrubs, and herbaceous plants (Thorne, 1992). Two main features unique to the angiosperms are their reproductive structures, flowers (Friedman et al., 2004), and seed-bearing structures, fruits (Palmer et al., 2004; Seymour et al., 2013).

1.2: *Perception of environmental cues is crucial for flower initiation*

Post-embryonic development and the ability to adapt morphological responses to environmental inputs are fundamental differences between plants and animals. Angiosperms have both reproductive and vegetative phases in their life cycles, meaning that they do not flower continuously. Correctly timing the switch from vegetative to reproductive development, i.e., the initiation of flowering, maximizes opportunities for pollination and successful seed set, both of which are crucial for a plant's reproductive success (Bernier et al., 1988). Four main pathways, the photoperiodic, vernalization, autonomous, and gibberellin pathways, integrate endogenous and environmental signals to

govern flower initiation by regulating the developmental fate of the shoot apical meristem (SAM) (Mouradov et al., 2002; Simpson and Dean, 2002).

1.3: *The photoperiodic pathway interprets day length*

The photoperiodic pathway restricts reproductive development to certain seasons by sensing the relative lengths of day and night. Garner and Allard (1920), by growing plants in varying day lengths, found that some species can only initiate flower and fruit production when day length is within a specific range. With this result, they were the first to report that plants can recognize photoperiod, though they did not speculate as to the mechanism. A major question remained: in what organs do plants perceive day length? Knott (1934) determined that leaves are responsible for light perception, despite the fact that flowers are formed in the SAM. Zeevaart (1985) further confirmed the ability of leaves to detect and transmit information about day length using *Perilla crispa*, a species that is known to induce flowering only in response to short day conditions (SDs). By grafting a leaf from a plant grown in SDs, a plant kept in long day conditions (LDs) was induced to flower. Although the SAM contains the stem cells from which all aerial parts of a plant are derived, it is spatially distant from both flowers and leaves in adult plants. How, then, can the message received by a leaf be transferred to the SAM? It was hypothesized that a mobile signal, or florigen, moves from leaves to the SAM to initiate flower formation (Chailakhyan, 1936; Zeevaart, 1976). The identity of the florigen remained elusive for over half a century. Work by King et al. (1968; 1973) indicated that the mobile signal was transmitted from leaves to the SAM through the phloem, though attempts to purify the

substance failed (Corbesier et al., 1998) and reinforced the hypothesis that the florigen was a mixture of substances (Bernier et al., 1993).

During the search for the florigen, the identities of genes in the photoperiodic pathway and their regulatory relationships were first elucidated with a molecular genetics approach in *Arabidopsis* (Figure 1.1). *Arabidopsis* flowers in long day conditions and is an annual plant, meaning it flowers only once before dying. Loss-of-function mutants in the circadian clock-regulated genes *GIGANTEA* (*GI)*, *CONSTANS* (*CO*), and *FLOWERING LOCUS T* (*FT*) all produce a late-flowering phenotype (Koornneef et al, 1998). Conversely, overexpression of the MADS-box transcription factor *SUPPRESSOR OF CONSTANS1* (*SOC1*), *CO*, or *FT* results in early flowering (Kardailsky et al., 1999; Kobayashi et al., 1999; Borner et al., 2000; Lee et al., 2000; Onouchi et al., 2000). *GI* is required for *CO* transcriptional activation (Fowler et al., 1999; Park et al., 1999; Huq et al., 2000; Suarez-Lopez et al., 2001) and *CO,* in turn, transcriptionally regulates *FT* and is required for floral induction in response to conducive day length (Yanovsky and Kay, 2002). An et al. (2004) implicated *CO* as a regulator of the florigen by demonstrating that *CO* acts cell autonomously in the phloem companion cells to trigger a systemic signal capable of crossing graft junctions.

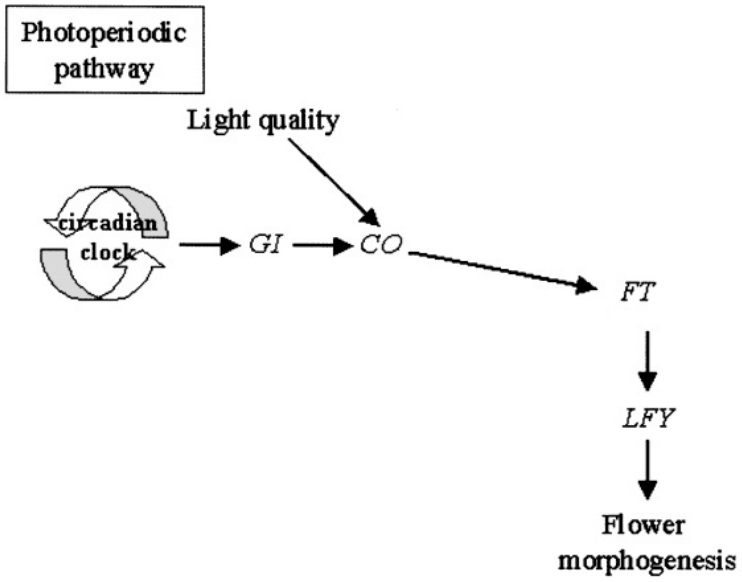


Figure 1.1 Flowering time regulators in the photoperiodic pathway

The photoperiodic pathway promotes flowering under conducive day length, which is specifically long days in *Arabidopsis*. *GI* and *CO* expression are regulated by the circadian clock and light quality controls *CO* protein stability (Valverde et al., 2004). *CO* transcriptionally regulates *FT* which in turn upregulates *LFY* and promotes flowering at the SAM. Figure adapted from Corbesier and Coupland, 2005.

In addition to the genetic evidence implicating its role in photoperiodic control of flowering, *FT* encodes a protein with homology to RAF-kinase inhibitor proteins, which in mammals are known cell signaling factors (Bradley et al., 1997; Kardailsky et al., 1999; Pnueli et al., 2001). *FT* is not expressed at the SAM (Takada and Goto, 2003), though transcripts are detected in shoot apices that contain young leaves (Wigge et al., 2005). Taken together, the data from *Arabidopsis* placed the identity of the florigen within the network of genes that regulate flowering in response to photoperiod and suggested *FT* to be either the mobile flowering signal itself or responsible for the synthesis of the signal.

1.4: *Flowering Locus T (FT), a small peptide, is the florigen*

Lifschitz et al. (2006) demonstrated that the floral stimulus, but no mRNA with similarity to *FT*, crossed the junction between grafted tomato plants. After replicating this result in grafted *Arabidopsis* seedlings, Corbesier et al. (2007) further showed that *FT* transcription in the phloem cells of leaves induces flowering and that *FT* mRNA is only transiently required in the leaf. FT-GFP fusion proteins made specifically in phloem cells travel to the SAM and are mobile across graft junctions of *Arabidopsis* seedlings. With these results, Corbesier et al. showed that the FT protein, a small, 20 kD peptide, is the long-distance signal that induces flowering in *Arabidopsis* in response to a long day photoperiod. Homologs of *FT* have since been implicated as flowering time regulators in multiple diverse species, including strawberry (Mouhu et al., 2013; Koskela et al., 2012; Kurokura et al., 2017).

FT itself has no DNA binding domain but forms a heterodimer with the bZIP transcription factor FLOWERING LOCUS D (FD) in nuclei of SAM cells. The FT-FD complex induces floral meristem (FM) identity by transcriptionally upregulating the FM identity genes *LEAFY* (*LFY*) and *APETALA1* (*AP1*) (Ruiz-Garcia et al., 1997; Abe et al., 2005). *LFY* and *AP1* are two of a group of four FM identity genes with overlapping functions (Irish and Sussex, 1990; Weigel et al., 1992; Mandel and Yanofsky, 1995; reviewed by Kieffer and Davies, 2002). The single, loss-of-function mutants *ap1* or *lfy* show partial defects in FM identity, but the more severe phenotype of the *ap1 lfy* double mutant reveals a partial overlap of their functions (Huala and Sussex, 1992; Bowman et al., 1993). Genetic data also provided the first indication of *FT* function through *AP1*; the phenotype of the *ft lfy*

double mutant closely resembles the *lfy ap1* double mutant (Ruiz-Garcia et al., 1997).

In addition to its role as a regulator of floral initiation, novel functions of *FT* and *FT* paralogs have been documented in several species. I review this information in Chapter 3 and present data suggesting a novel role for the strawberry *FT* ortholog in fruit development.

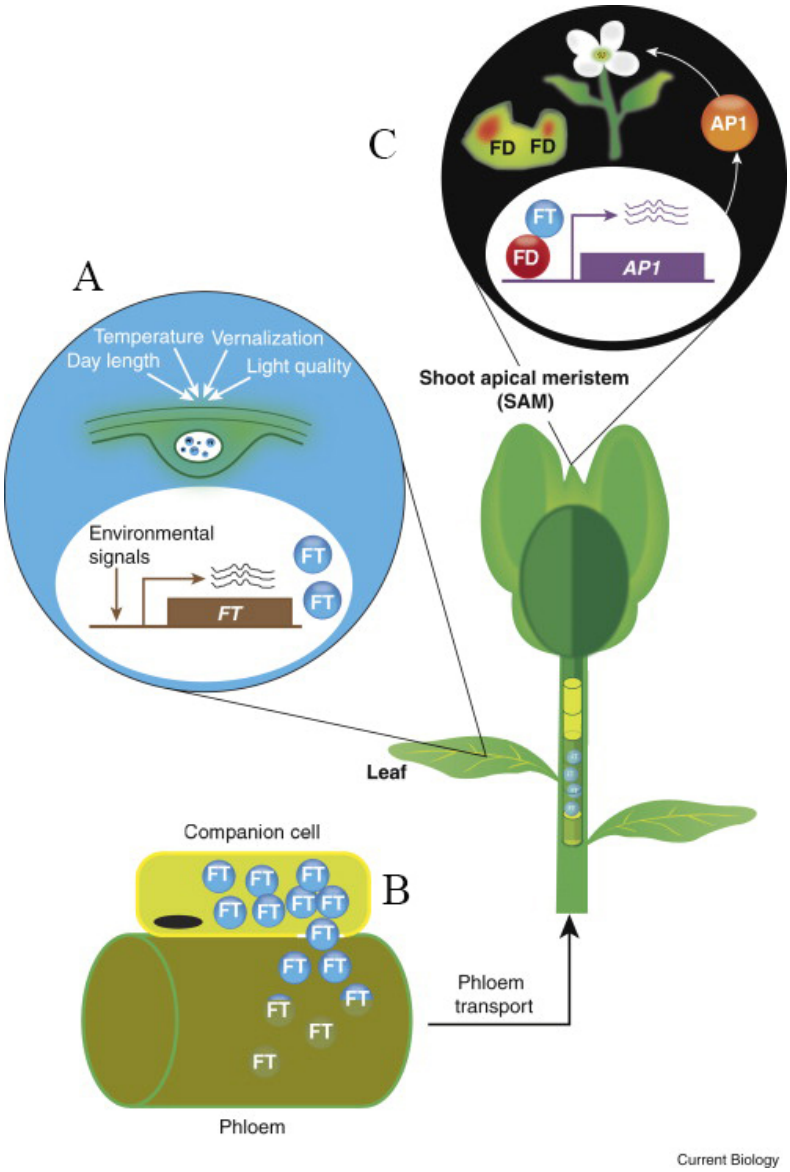


Figure 1.2 Overview of FLOWERING LOCUS T function in floral initiation

(A) In response to inductive environmental signals, *FT* is transcribed in leaves by the transcription factor CONSTANS (CO), specifically in phloem companion cells
(B) After translation, the FT protein is loaded into the phloem and transported to the shoot apical meristem (SAM).
(C) At the SAM, the FT protein is localized to the nucleus where it forms a complex with the bZIP transcription factor FD. The FT-FD complex initiates transcription of the floral meristem identity genes *AP1* and *LFY*. Figure courtesy of Wigge, 2011.

1.5: *ABC model of flower development*

Following induction of floral meristem identity at the SAM mediated, in part, by *FT,* floral organ identity is specified by several well-characterized homeotic genes. Floral architecture is conserved across the angiosperms; flowers consist of four concentric whorls of organs: sepals, petals, stamens (male organs), and carpels (female organs) (Figure 1.3 A-C). The original ABC model of floral development, first presented in a seminal paper by Coen and Meyerowitz (1991), illustrates the interaction and antagonism of homeotic genes that specifies floral architecture (Figure 1.3 C-F). Today, the ABC model is supported by data from many species (Friedman et al, 2004; Irish, 2017), though Coen and Meyerowitz originally synthesized their ideas based on double and triple knockouts of floral identity genes in Arabidopsis and *Antirrhinum majus* (Bowman et al., 1989, 1991; Hill and Lord, 1989; Kunst et al., 1989; Sommer et al., 1990; Yanofsky et al., 1990; Drews et al., 1991; Schwarz-Sommer et al., 1992; Jack et al., 1992). To briefly highlight some of the supporting genetic data from *Arabidopsis*, complete loss of function of the A class genes, *APETALA 1* (*AP1*) and *APETALA 2* (*AP2)*, results in the transformation of sepals to carpels and petals to stamens. In B class mutants, loss of *PISTILLATA* (*PI*) and *APETALA 3* (*AP3*) function causes transformation of petals to sepals and stamens to carpels. Loss of function in Arabidopsis *AGAMOUS* (*AG*), a C class gene, transforms stamens to petals and carpels to sepals. *AG* is also required for floral meristem determinacy; *ag* mutants exhibit a flower-

within-a-flower phenotype in which a new flower develops in place of carpels. Loss of A, B, and C class genes simultaneously in *Arabidopsis* results in floral structures containing only leaf-like organs. Later reverse genetics experiments identified an additional category of homeotic genes: E class (Pelaz et al., 2000; Ditta et al., 2004). Simultaneous loss of function in the four E class genes, *sepallata 1-4* (*sep 1-4*), results in a phenotype similar to simultaneous loss of A, B, and C function (Honma and Goto, 2001).

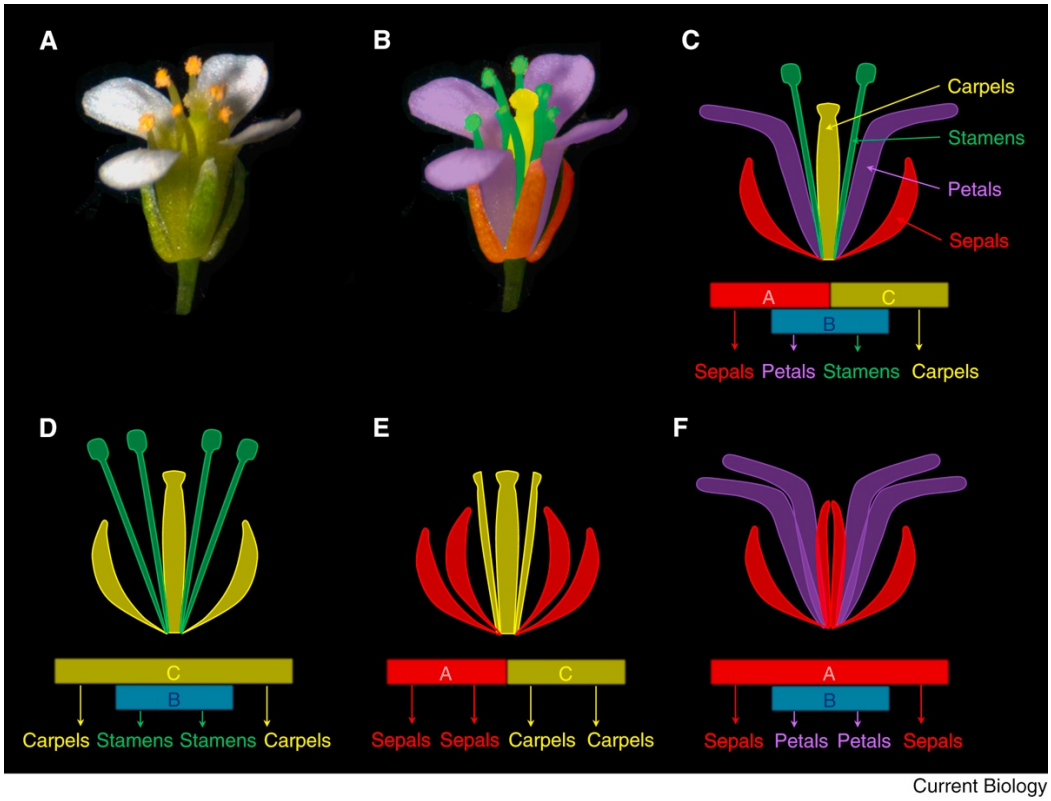


Figure 1.3 The ABC model of flower development

Arabidopsis flower is shown in **(A)** and color coded in **(B)** to indicate the four whorls of florals organs: sepals (red), petals (purple), anthers (green), and carpels (yellow). The flower schematic in **(C)** demonstrates in which domain each class of floral organ identity genes is active. A class genes alone are required for sepal identity. A combination of A and B activity specifies petals while B and C activity are required for stamens. C class gene activity alone is required for carpel formation. **(D)-(F)** show the phenotypes resulting from loss of A, B, and C class function, respectively. Loss of A class activity produces flowers lacking sepals and petals **(D)**. Loss of B class function results in flowers lacking petals and stamens **(E)** and loss of C class gene activity produces flowers without stamens or carpels **(F)**. Figure courtesy of Irish, 2017.

In the ABC model (Figure 1.3), the formation of sepals in the outermost whorl is governed by the activity of A class genes. The formation of the next whorl, the petals, depends on the combined activity of A and B class genes while the whorl of stamens depends on combinatorial B and C class gene activity. The development of the innermost whorl, the carpels, is regulated by C class gene activity alone. Furthermore, A and C class genes negatively regulate each other and prevent an overlap of their activities. The A class proteins repress C class gene activity in the outer whorls and C class proteins repress A class gene activity in the inner whorls. E class proteins function in all four floral whorls and act as partners for the A, B, and C class genes.

Unlike the homeotic genes of animals, e.g., the Hox genes first discovered in *Drosophila* (Bridges and Morgan, 1923), the ABC floral identity genes, except the A class gene *AP2*, are members of the MADS box family of transcription factors. MADS box genes have been identified in many eukaryotic organisms, including mammals, but the family is greatly expanded in plants with at least 107 known members in *Arabidopsis* (Heijmans et al., 2012). The ABC genes regulate transcription of downstream genes to specify development of the correct floral organ in the correct location.

1.6: *Fertilization and fruit set overview*

Proper development and patterning of floral organs is ultimately required for a plant's reproductive success. Ovules, the structures that house female germ cells, are contained within carpel, or ovary, tissue (Soltis and Soltis, 2004). Upon successful pollination and fertilization, ovules develop into seeds (Dumas and Mogensen, 1993; Mascarenhas, 1993). Fertilized ovules trigger development of the floral carpel into a botanical fruit, which

protects seeds and facilitates their dispersal to initiate the next generation of plants (Gaertner, 1788; Ridley, 1930; Seymour et al., 2013).

The developmental switch that initiates transformation of an ovary into a fruit is termed 'fruit set,' or the time point at which a decision is made to either continue or abort fruit development. The development of fruit across all angiosperms is regulated by positive signals, generated either upon pollination or fertilization (Gillaspy et al., 1993). The identities of all of these positive signals have yet to be elucidated, though known regulators include the phytohormones auxin and gibberellic acid (GA) (Nitsch, 1970). Auxin is involved in a broad array of biological processes, including regulation of cell polarity, cell elongation, differential growth, embryo development, and organ formation (Sauer et al., 2013). GA is also involved in many growth and developmental processes throughout the plant life cycle, including the switch from vegetative to reproductive growth, stem elongation, and seed germination (Gupta and Chakrabarty, 2013).

Nitsch (1950) first demonstrated the positive effect of auxin on fleshy fruit development with experiments on strawberry. After removal of developing achenes from the surface of the receptacle, the fleshy fruit fails to develop. However, replacement of achenes with exogenous auxin stimulates receptacle development. Therefore, auxin is required for strawberry receptacle development and the achenes are the source of the auxin (Nitsch, 1950; Dreher and Poovaiah, 1982). Exogenous application of GA also initiates fleshy fruit development in the absence of fertilization in tomato (Gustafson, 1960) and strawberry (Thompson, 1969). In strawberry, simultaneous application of GA and auxin is required to produce receptacles similar in size to fruit resulting from pollination, suggesting that auxin

and GA have dependent as well as independent functions in stimulation of fleshy fruit enlargement (Figure 1.4; Kang et al., 2013). Exogenous GA and auxin application also stimulates strawberry achene and seed enlargement in the absence of fertilization (Figure 1.4 B, C). Work in tomato and *Arabidopsis* suggests that auxin may positively regulate GA (Sastry and Muir, 1963; Serrani et al, 2008; Dorcey et al., 2009; Fuentes et al., 2012), though molecular mechanisms underlying the interplay between biosynthesis and signaling in the context of fruit development are still poorly understood.

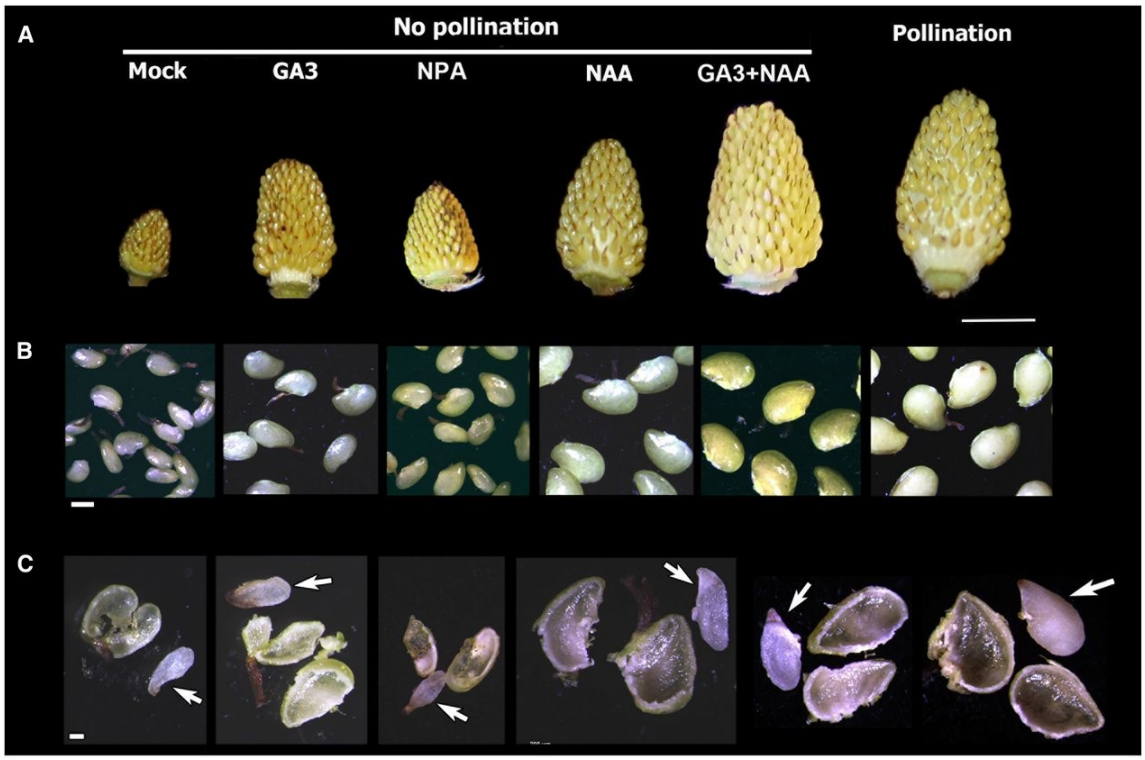


Figure 1.4 Exogenous application of auxin and GA induce *F. vesca* receptacle, achene, and seed enlargement in the absence of fertilization

(A) Receptacle enlargement following application of GA, an auxin inhibitor (NPA), auxin (NAA), and GA + auxin to emasculated flowers. Negative and positive controls are mock treatment and hand pollination, respectively. Scale bar = 5 mm. Photos were taken 12 days after the first hormone treatment.
(B) Achenes following treatments described in **(A)**. Scale bar = 0.5 mm.
(C) Seeds (arrows) dissected from the ovary wall following treatments described in **(A)**. Scale bar = 0.2 mm. Figure courtesy of Kang et al., 2013.

1.7: *Fleshy fruit can develop from diverse floral tissues, even in closely related species*

Although fruit set is regulated by positive fertilization signals in all angiosperms, not all fleshy fruit originates from the floral ovary. The *Rosaceae* family provides a prime example of closely related species with morphologically distinct fruit types, including peach, plum, apple, and strawberry (Xiang et al., 2017). The fleshy fruits of peach and plum are true botanical fruits in that the edible flesh of each is derived from ovary tissue. The fleshy fruit of apple, however, develops from the floral tube, which is a structure of fused sepals. In strawberry, the fleshy fruit is not a true fruit at all, but an accessory fruit due to its development from the floral receptacle, or stem tip. The botanical fruit of strawberry is the dry, ovary-derived achene, over 200 of which dot the surface of the fleshy fruit (Figure 1.5). Despite the major differences in the appearance of their fruit, peach (*Prunus persica*) and strawberry (*Fragaria vesca*) are both diploid species and display extensive genomic synteny (Shulaev et al., 2011), thereby highlighting the variability in fruit morphology even among closely related species. Understanding the molecular underpinnings of these developmental differences is both biologically interesting and economically relevant. As an herbaceous species, *F. vesca* is an attractive model for other members of the *Rosaceae* family, especially the slow-growing tree species.

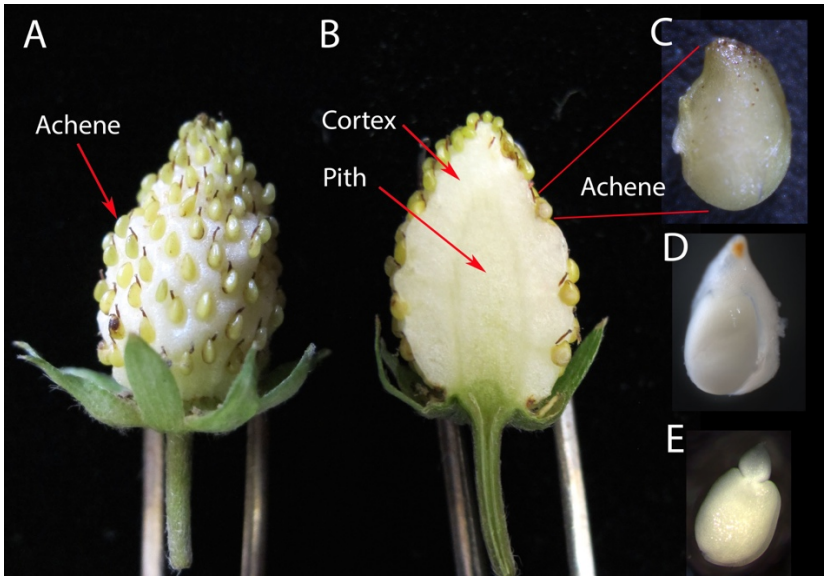


Figure 1.5 Anatomy of strawberry fruit

(A) The botanical fruit of strawberry, the achenes, are externally arranged on the fleshy receptacle.
(B) The fleshy receptacle is made up of two sub-tissues, the cortex and pith. Pith contains mostly vasculature.
(C) The hard, outer shell of the achene is derived from the ovary wall.
(D) Within an achene is a single seed, which houses an embryo **(E)**.

1.8: *Strawberry is a useful model for studying fleshy fruit development*

F. vesca, the alpine or woodland strawberry, has been cultivated since at least the fourteenth century but has only recently been developed as a model system (Darrow, 1966; Slovin et al., 2009; Shulaev et al., 2011). The external seed configuration of strawberry fruit is ideal for studying cross-tissue communication during fruit development, particularly phytohormone biosynthesis and transport. The strawberry botanical fruit, the achene, is derived from the ovary wall and houses a single seed (Figure 1.5). Previous work from our lab indicates that the hormones auxin and GA are synthesized in the achene, particularly in the endosperm and seed coat, and are transported to the underlying receptacle where they initiate fleshy fruit development (Nitsch, 1950; Kang et al., 2013). The primary model systems for studying fruit development have historically been

Arabidopsis (Roeder and Yanofsky, 2006) and tomato (Kimura and Sinha, 2008), the fruits of which are dry and fleshy, respectively, and develop from the ovary wall (Ferrandiz et al., 1999; Gasser and Robinson-Beers, 1993). Studying the strawberry accessory fruit expands our knowledge of general developmental processes.

Specifically, *F. vesca* is a tractable model for the commercial strawberry, the octoploid *F. ananassa,* and other members of the economically important *Rosaceae* family due to its diploidy (2n = 14), small genome size (240 Mb), amenability to transformation, and ease of growth in a lab setting (Slovin et al., 2009; Shulaev et al., 2011). The two accessions used for all experiments described in this thesis, Yellow Wonder 5AF7 (YW 5AF7) and Hawaii-4 (H4), are members of the *semperflorens* subspecies of *Fragaria vesca* and, unlike the commercial strawberry, do not require a short day photoperiod to flower. *F. vesca semperflorens* contains a 2 bp deletion in the floral repressor *TERMINAL FLOWER1* (*FvTFL1*), which leads to a day neutral flowering habit (Koskela et al., 2012). In *Arabidopsis*, TFL1 binds to FD at the SAM and blocks formation of the FT-FD complex, thereby inhibiting floral initiation in non-inductive photoperiods (Hanano and Goto, 2011). The relaxed photoperiodic requirements of the *F. vesca semperflorens* accessions facilitate year-round flowering and enable growth in long day conditions.

As a new model system, *F. vesca* genetic resources are still under development but presently include large quantities of genomic and transcriptomic sequences. Unlike *Arabidopsis*, there is no T-DNA library available for strawberry from which mutants may be ordered. Because all transgenic plants must be generated via tissue culture, a remaining challenge is the extensive time required to generate first-generation transgenics and

propagate plants. Currently, it takes 12 months to transform a construct into strawberry and the time from seed-to-seed is four to five months.

1.9: *Stages of strawberry fruit development*

Following successful pollination and fertilization, the strawberry receptacle transforms from a green, dome-shaped anchor for floral organs into a flavorful and aromatic fleshy fruit. This transformation is marked by three distinct developmental stages: cell proliferation, cell expansion, and ripening (Gillaspy et al., 1993). Traditional, visual markers from the red-fruited, commercial variety *F. ananassa* divide fruit development into the following stages: small green, medium green, big green, white, pink (also called turning), and red (Fait et al., 2008). The entire timeline, from fertilization to a fully ripe fruit, spans approximately 25 days. Although general staging with visual markers is useful for agricultural purposes, developing *F. vesca* into a model system for studying the molecular underpinnings of fruit development requires far more specific markers.

To standardize molecular genetics and genomics experiments in *F. vesca*, a former graduate student in our lab, Courtney Hollender, generated detailed morphological markers corresponding to successive flower and fruit developmental stages in the accession Yellow Wonder 5AF7 (Hollender et al., 2012). Floral development is delimited by 12 stages beginning with the undifferentiated floral meristem (stage 1) and extending to the fully developed flower just prior to anthesis, or opening (stage 12). Early fruit development is marked by five developmental stages, beginning with the open flower at anthesis and extending to 13 days post anthesis (DPA) (Figure 1.6A), which corresponds to the 'big green' stage in cultivated strawberry. Fertilization occurs between stages 1 and 2.

1.10: *Spatial and temporal transcriptomes provide genome-scale insight into strawberry fruit development*

A number of research efforts have focused on the late stages of strawberry fruit development to facilitate study of ripening, flavor, aroma, and nutritional content (Aharoni and O'Connell, 2002; Garcia-Gago et al., 2009), including several transcriptomic studies (Estrada-Johnson et al., 2017; Sanchez-Sevilla et al., 2017). However, knowledge of molecular events underlying fruit set and the early stages of development is equally critical and useful for ensuring consistent crop yield. Based on the morphological markers previously described by Hollender et al. (2012), our lab generated RNA-Seq data for 92 libraries (46 samples x 2 biological replicates) from *F. vesca* Yellow Wonder 5AF7 flower and fruit tissues at multiple early and mid-developmental stages (Kang et al., 2013; Hollender et al., 2014; Hawkins et al., 2017).

The first set of RNA-Seq data (Figure 1.6; Kang et al., 2013) profiles fruit development from stages 1 to 5 for five different tissues: cortex, pith, embryo, ghost (endosperm and seed coat), and ovary wall, plus one developmental stage each for ovule and seed (25 samples x 2 biological replicates = 50 libraries). All of these samples were hand-dissected using a stereomicroscope. The second set of RNA-Seq data (Hollender et al., 2014) profiles 17 floral tissues over twelve stages of development along with leaf and seedling as vegetative controls (19 samples x 2 biological replicates = 38 libraries). The floral tissues include: perianth (petals + sepals), anther, carpel, style, receptacle, microspores (collected from stage 10 anthers), and pollen. Samples from stages 1-7 along with microspores were isolated using laser capture microdissection (LCM) due to their small size. The final RNA-

Seq dataset (Hawkins et al., 2017) includes receptacle fruit tissue harvested at two mid-developmental stages: 15 days post anthesis (DPA) and turning stage (22 DPA), the time point just prior to ripening (2 samples x 2 biological replicates = 4 libraries). This wealth of spatial and temporal transcriptome data provides large-scale insight into biological processes and molecular events underlying early fruit development. Since fruit development across all angiosperms is regulated by positive fertilization signals, insights gained from studying fruit set and the earliest stages of fruit development in strawberry are broadly applicable to other species.

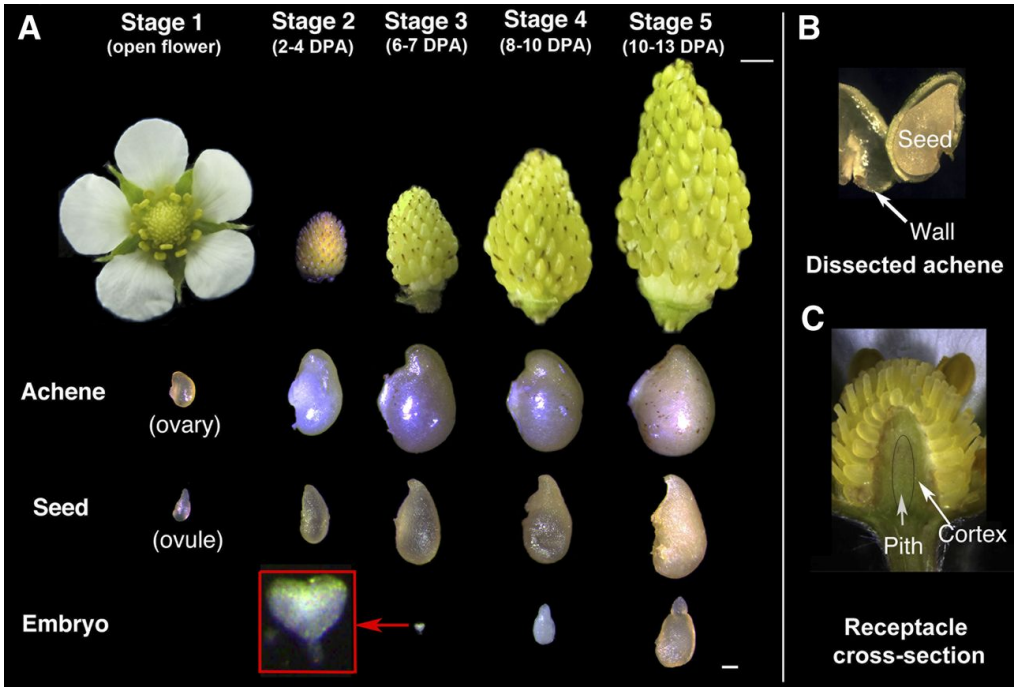


Figure 1.6 Tissues and stages profiled via RNA-Seq for early fruit transcriptome

(A) Five fruit tissues were individually profiled over five stages of development spanning the open flower at stage 1 to stage 5 (top row). Fertilization occurs between stages 1 and 2. The achene and seed are photographed at each stage in rows 2 and 3, respectively. At stage 1, ovary and ovule indicate the pre-fertilized achene and seed, respectively. Embryos (row 5) were collected from stages 3-5 only due to their small size at stages 1 and 2. A larger view of the stage 3 embryo is shown in the inset. Scale bars are 2 mm (row 1) and 0.2 mm (rows 2-4).
(B) Ovary wall (achene) opened to reveal the single seed inside.
(C) Longitudinal section of a strawberry flower. The receptacle is the green stem tip covered in carpels. Figure courtesy of Kang et al., 2013.

1.11: *Co-expression network analysis as a tool to mine transcriptome data*

Recent technological advancements in transcriptome sequencing, coupled with decreasing costs, have created unprecedented opportunities to study non-model and developing model organisms like *F. vesca* (Strickler et al., 2012). RNA-seq experiments can be designed to study a plethora of topics, including comparison of mutant and wild type organisms, development, and abiotic and biotic stress response. Transcriptome data are also highly versatile and can be used to characterize gene expression across space and time (Rowland et al., 2012; Pattison et al., 2015), identify alternative splicing events (Li et al., 2017), identify novel transcripts (Chettoor et al., 2014), and identify key biological processes. Despite the increasing ubiquity of RNA-seq data for strawberry and other developing models, an enduring challenge for the research community, especially biologists lacking advanced computational skills, is how to best visualize and mine massive transcriptome datasets. One solution is to generate gene co-expression networks based on RNA-seq data, a strategy which I present for *F. vesca* in Chapter 2. Co-expression networks group genes that have similar expression profiles across multiple experimental conditions, the idea being that genes with similar expression are more likely to function in the same pathways or regulate the same biological processes (Zhang and Horvath, 2005; Serin et al., 2016). This information can be a useful foundation for hypothesis generation and subsequent experimental design, though an enduring challenge is to validate the accuracy and utility of co-expression networks.

1.12: *Summary of thesis research*

Insight gained from studying strawberry enhances our general knowledge of molecular

events underlying fruit development, the traditional models for which are the ovary-derived *Arabidopsis* and tomato fruits. Using a global transcriptomics approach, our lab has been a primary driver in the development of *F. vesca* as a model system. I used our wealth of comprehensive RNA-Seq data as a foundation on which to design my thesis research. Taken together, my work advances the systems-level infrastructure for studying molecular regulation of *F. vesca* fruit development, points to a novel role for the florigen *FLOWERING LOCUS T* in fleshy fruit, and provides molecular tools useful to both the Liu lab and the *Rosaceae* research community.

In Chapter 2, I present co-expression network analyses generated from 92 RNA-Seq libraries profiling multiple stages of strawberry flower and fruit development. The network, created in collaboration with Chris Zawora and Haley Wight, exemplifies a useful strategy to maximize utilization of a large transcriptome dataset. I demonstrate that co-expression network analyses can illuminate molecular processes underlying developmental events and are useful tools for hypothesis generation and experimental design, especially in a developing model like *F. vesca*. I further explore experimental validation for gene relationships predicted by the consensus networks, including support for a mutation in *FveUFO1* as the cause of altered floral meristem determinacy and floral organ identity in an *F. vesca* mutant. A paper describing this work was submitted to *Plant Physiology*.

In Chapter 3, I expand on the usefulness of the co-expression networks as tools to identify key regulators of flower and fruit development in *F. vesca*. I found that *FvFT1*, a homolog of the florigen *FLOWERING LOCUS T* (*FT*), is surprisingly induced to a high level in the developing strawberry receptacle fruit immediately post-fertilization. In *Arabidopsis*, *FT*

is transcribed in leaves in response to environmental cues and its protein product moves to the SAM where it initiates flower development. My results suggest that *FvFT1* functions in a unique pathway in fruit as compared to leaf and may serve as a channel of communication between the maternal fruit tissue and the progeny seed. In the course of studying *FvFT1 RNAi* plants in an effort to shed light on *FvFT1* function in fruit development, I found that transformation of *F. vesca* plants with RNAi constructs targeting different genes induced an embryo greening phenotype. I later found that result is not specific to *FvFT1* loss of function, though how an RNAi vector can cause embryos to turn green in the absence of light is an intriguing question and discussed further in Appendix 1.

In Chapter 4, I use differential expression analyses of our lab's transcriptome data to identify genes upregulated in the developing receptacle fruit versus all other profiled flower, fruit, and vegetative tissues. This work identified a list of 589 differentially expressed genes, some of which may be important regulators of fruit set and are attractive candidates for future research. Further, I selected a subset of the most strongly differentially expressed genes for the purpose of developing tissue-specific promoters as tools for strawberry fruit development research. I isolated upstream regulatory sequences from eleven genes and generated a *promoter::GUS* reporter for each to confirm transcriptional activity in early stage receptacle fruits. Since flowers and fruit are the last structures to form on a plant, development of tissue-specific regulatory sequences as research tools allows our lab and the strawberry research community to avoid off-target effects of more broadly expressed promoters.

Chapter 2: Consensus co-expression network analysis identifies key regulators of flower and fruit development in wild strawberry

2.1: *Introduction*

Since the publication of the first draft of its genome in 2011, *F. vesca* has been under development as a model for the economically important *Rosaceae* family, including the cultivated, octoploid strawberry *F. ananassa* (Shulaev et al., 2011). As a small, herbaceous perennial, *F. vesca* is much easier to study in a lab setting compared to a number of Rosaceous species, including trees like apple, peach, and plum. In addition to facilitating research within the *Rosaceae* community, insight gained from strawberry fruit development broadens our understanding of general developmental processes. Unlike tomato, the traditional model for fleshy fruit development, strawberry fleshy fruit develops from the floral receptacle and its ripening is non-climacteric. To investigate strawberry fruit development on a genome-wide scale, our lab previously generated spatial and temporal transcriptome data profiling *F. vesca* flower and fruit development pre- and post-fertilization (see Chapter 1, section 1.10). I collaborated with Chris Zawora to generate gene co-expression networks based on this extensive RNA-Seq data for the purpose of facilitating data exploration and hypothesis generation.

Co-expression network analysis is based on correlations between gene expression values. It describes correlation patterns between genes in a pairwise fashion across multiple microarray or RNA-seq samples. Co-expression networks are useful for exploring large, complex datasets and are powerful tools for predicting gene function. Exploration of

neighborhoods of connected genes invokes the 'guilt by association' principle in that genes that are highly connected and have very similar expression patterns are more likely to function in the same pathways or regulate the same biological processes (Ravasz et al., 2002; Spirin and Mirny, 2003; Singer et al., 2005; Wolfe et al., 2005).

Recent analyses have sought to test and optimize gene co-expression networks generated with RNA-Seq data (Iancu et al., 2012; Sekhon et al., 2013; Huang et al., 2017). However, the accuracy of gene functions and relationships predicted by network approaches and the overall utility of co-expression networks is still largely unknown. Computationally, one method to test the stability and robustness of clusters is to use a post-hoc consensus clustering approach (Monti et al., 2003). If clusters accurately represent subpopulations of a larger dataset, the number and composition of clusters should not vary greatly if clustering is repeatedly conducted using different parameters or different subsets of the full dataset. Clusters that are robust in response to sampling variability are more likely to represent true relationships between genes. Therefore, consensus clusters are potentially more reliable for predicting gene functions and interactions, though experimental data is still necessary to assess their accuracy.

In this chapter, I present three sets of co-expression networks generated in collaboration with Chris Zawora using the WGCNA package available in R (Langfelder and Horvath, 2008). The first network incorporates early stage, *F. vesca* floral tissues dissected by laser capture microdissection (LCM). The second includes hand-dissected flower and fruit tissues spanning pre-fertilized flowers to fruit just prior to ripening. The third set concerns receptacle fruit tissues at the ripening stages. The use of multiple, comprehensive RNA-

seq libraries provides extensive gene expression information for network construction and therefore increases the likelihood of identifying genetic correlations (Lee et al., 2004; Wren, 2009; Ballouz et al., 2015). Additionally, Chris Zawora tested a consensus clustering add-on to the WGCNA algorithm and Haley Wight used bootstrapping to test the reliability of clusters generated with both WGCNA alone and with WGCNA plus consensus clustering. I demonstrate that co-expression network analyses can illuminate molecular processes underlying developmental events and are useful tools for hypothesis generation and experimental design. Further, I explore experimental validation for gene relationships predicted by our consensus networks, including support for the identification of a mutation in *FveUFO1* as the basis for a mutation affecting floral meristem determinacy and floral organ development. To increase utility, the networks are presented in a web interface for easy exploration and identification of co-expressed genes. Together, the work reported here illustrates ways to generate robust networks optimized for the mining of large transcriptome datasets, providing a useful resource for hypothesis generation and experimental design in strawberry and related *Rosaceae* fruit crops.

2.2: *Methods*

WGCNA network analysis

Chris Zawora generated gene co-expression networks for *F. vesca* using gene-level TPM (Transcripts Per Million; Wagner et al., 2012) expression measurements from 92 RNA-seq libraries spanning the early developmental stages of plant tissues to ripening fruit. Genes with variance < 0.05 were filtered out and the results were used as input to signed WGCNA network construction (WGCNA v1.60 package in R; Langfelder and Horvath, 2008). In

standard WGCNA networks, power was set to 6, minModuleSize was set to 100, and initial clusters were merged on eigengenes. The mergeCutHeight value was set to 0.25 across all networks. Total connectivity was calculated for all genes in each network.

Consensus network construction

To construct the consensus network, 80% of genes were subsampled 1000 times; paired with each subsampling was a set of randomized parameters standard to the WGCNA. These parameters consisted of: power transformation [1, 2, 4, 8, 12, 16], minModuleSize [40, 60, 90, 120, 150, 180, 210], and merge on eigengenes [true/false]. After 1,000 runs of WGCNA were performed, a weighted adjacency matrix was computed to represent the connection strength between every gene pair. Letting p be the number of genes with count variance > 0.5, adjacency matrix (A) was calculated by:

$$A_{i,j} = \frac{number\ of\ times\ gene\ i\ is\ clustered\ with\ gene\ j}{number\ of\ times\ gene\ \ i\ is\ subsampled\ with\ gene\ j}$$

$$A_{i,j} \in [0,1], \forall i,j \in \{1,,2,\dots p\}$$

The adjacency matrix was then used as a basis for the consensus network, consensus90 network and consensus100 network. The consensus network was constructed by clustering the adjacency matrix using WGCNA with power:6, minModuleSize:100, and no merging on eigengenes. Consensus90 was constructed by translating the weighted adjacency matrix to a graph by thresholding at a value of 0.90. The following clusters were then established by using the connected components function in igraph. Consensus100 was performed similarly with a threshold of 1.

Network Visualization

Module eigengenes were calculated subsequent to network construction using the moduleEigengenes function in WGCNA. This function calculates the 1st principal component of the genes' TPM in a given cluster and can be used as a summary statistic to relate clusters to sample-specific expression levels. Results are then visualized using R's boxplot function. The networks were visualized using Cytoscape _v.3.5.1.

Comparisons between standard WGCNA and consensus networks

Haley Wight compared consensus clusters to a standard WGCNA clustering approach. This comparison was based on gene expression values from the hand-dissected, log transformed TPM values. The standard WGCNA approach was constructed using Pearson's Correlation distance, power set to 6, and minModuleSize set to 100. The Jaccard Index and hypergeometric p-value were determined using the R package GeneOverlap (Shen and Sinai, 2013). The comparative statistics were computed using RS statistic, which is a measure of the variance between clusters to the variance within clusters, calculated by (TSS-SSE)/TSS, where TSS=SSE+SSB (TSS: Total Sum of Squares; SSE: Sum of Square Error; SSB: Between group sum of Squares). By definition, this statistic was based on the Euclidian distance between two genes (Liu et al., 2013). Bootstrap confidence intervals were determined by frequency of co-clustering between gene pairs in 1,000 runs of WGCNA with varied parameters and sampling.

GO enrichment

GO enrichment tests were performed to understand potential functional relationships between co-clustered genes. GO annotations were created using Blast2GO (Conesa and

Gotz, 2008). GO term enrichment p-values were calculated using the Fisher's exact test in the TopGO R package (Alexa and Rahnenfuhrer, 2016).

Web-based application to visualize and download network data

All co-expression network data is available at www.fv.rosaceaefruits.org. The web application was generated with Shiny from RStudio (http://shiny.rstudio.com)

Free iron staining: Perls followed by DAB enhancement

Staining was conducted with the help of John Sittmann. The following protocol is adapted from Roschzttardtz et al. (2009) and Brumbarova et al. (2014). All solutions were made fresh on the day of treatment.

Perls staining: tissues were collected from the wild type *Fragaria vesca (Fv)* accession Yellow Wonder 5AF7. Plants were grown in chambers set to 16 hours of light at 25°C and 8 hours of darkness at 20°C. Ovules and seeds were hand dissected under a stereomicroscope. Tissues were first fixed for 1.5h in a solution containing methanol: chloroform: glacial acetic acid (6:3:1). Then, tissues were vacuum infiltrated for 45 minutes in a solution containing equal volumes of 6% Perls (potassium ferrocyanide) and 4% HCl. Following infiltration, samples were incubated at room temperature under a fume hood for 15 minutes. Next, samples were washed three times with DI water. At this point, receptacle samples were photographed under a stereomicroscope equipped with a Zeiss Axiocam 105 color camera.

Taking advantage of the redox activity of Prussian blue, 3,3'-Diaminobenzidine (DAB) was previously used to intensify Perls staining (Nguyen-Legros et al., 1980; Roschzttardtz et al., 2009). A 0.5% DAB stock solution was prepared by adding 0.05g of DAB to 10ml DI water and letting sit for 5 minutes with occasional vortexing. Next, 300 µl of 37% HCl was added and the solution was left at room temperature for 5 minutes with occasional vortexing. Finally, the DAB solution was filtered through a 0.22um filter. To stain the tissue, plant material previously stained with Perls was incubated for 1 h in the preparation solution (0.01M NaN_3 and 0.3% H_2O_2 in methanol). Following this, samples were washed three times with 0.1M PBS (pH 7.4). DAB intensification was carried out in a 0.1M PBS solution containing 0.0125% DAB, 0.005% H_2O_2 and 0.005% $CoCl_2$. Intensification was carried out for 20 minutes at room temperature. Samples were washed three times in DI water to stop the DAB intensification reaction.

Plant growth condition and *efo* mutant isolation

Wild type Yellow Wonder 5AF7 (YW5AF7), the *efo* mutant in the YW 5AF7 background, and a segregating *efo* mutant population of M3 sister plants were all grown in a growth chamber with 16h light at 25°C followed by 8h dark at 20°C. The *efo* mutant was isolated from an EMS mutagenesis screen of YW 5AF7 (Hollender, 2012).

Bulk segregant mapping-by-sequencing of *F. vesca efo* mutant

The M3 mapping population consisted of three *efo* mutant plants and 22 sister plants with a WT phenotype. Genomic DNA was extracted from young leaves using the NucleoSpin Plant II kit (Machery-Nagel). Equal quantities of gDNA from each mutant plant were combined into one pool (mutant pool). Equal quantities of gDNA from each of the 22 sister

plants were combined into a second pool (WT pool). A total of 2 micrograms gDNA from each pool was sequenced at the Genomics Resources Core Facility at Weill Cornell Medical College on Illumina HiSeq2000. The two libraries were bar coded and each sequenced on one half of one lane. A total of 92,632,150 and 89,565,704, fifty-one bp, single-end reads were generated for the mutant and WT pools, respectively.

Mapping was carried out by Dr. Wanpeng (George) Wang. 51 bp reads were mapped to the *Fragaria vesca* reference genome v1.1 using Bowtie2 with default settings. Variants were called across the two samples using SAMtools. Of the 199,622 total variants called, about 87% were homozygous across both samples. Some of these variants represent differences between YW5AF7 and Hawaii-4, the accession used for the reference genome, and were discarded. Of the remaining variants, 11,242 are G/C to A/T single nucleotide changes, which is the most common mutation induced by EMS treatment. George further filtered these 11,242 variants to select for those that are 1) homozygous in the mutant pool; 2) heterozygous in the WT pool; 3) Located in exons; 4) cause amino acid changes; 5) are nonsense or nonsynonymous mutations; and 6) are present at 13-53% frequency in the WT pool. After filtering, variants in 97 genes remained (Dataset S2.5). Forty-three of these genes had expression of RPKM 10 or higher in stage 1-4 flower buds and only one, gene19967 (*FveUFO1*), is a homolog of an *Arabidopsis* gene known to control floral development (AT1G30950; *UNUSUAL FLORAL ORGANS*; Levin and Meyerowitz, 1995).

Sequence alignment

Amino acid sequences of *UFO* homologs were aligned using Clustal Omega and figures were generated with Boxshade.

qRT-PCR to test transcript levels of *FveUFO1*

Total RNA was isolated from stage 1-4 flower buds using an RNAeasy Plant Mini Kit (Qiagen). Three flower buds isolated from each of three individual plants were pooled into a single biological replicate (9 buds total/replicate). Four biological replicates were analyzed for both WT and *efo*. RNA samples were treated with DNase I (ThermoFisher) to remove contaminating genomic DNA and subsequently re-purified with the NucleoSpin RNA XS kit (Machery-Nagel). cDNA was synthesized from 1 ug of total RNA in a 20 µl solution using the RevertAid First Strand cDNA synthesis kit (ThermoFisher). 1:10 cDNA was used as the template in real-time qPCR. SsoAdvanced Universal SYBR Green Supermix (BioRad) was used to set up real time reactions, which were run and analyzed on a CFX96 Real-Time System. Forty cycles of two-step real time PCR were run as follows: 95 degrees, 30 s. 95 degrees, 15 s. 58 degrees, 30 s followed by melt curve (65 to 95 degrees, 0.5 degree increments). Primer sequences to target *FveUFO1*, gene 19967, are FvUFO1 qPCR F 5' GAACAGGCTGAATGTGCC 3' and FvUFO1 qPCR R 5' GCAAAGCCTTACCAGAACC 3'. Gene03773, which is stably expressed across all profiled stages of receptacle development, was used as a control for normalization. Primer sequences were taken from Lin-Wang et al., 2014: gene03773 control qPCR F: 5' TTTGAAGCGCCTTGCTGAAG 3' and gene03773 control qPCR R: 5' GGCAGATTGCACGCAGAAT 3'. Data was analyzed with the $2^{\Delta Ct}$ method and statistical

significance was calculated with a Student's t-test.

Complementation of the *efo* mutant phenotype

Generation of FveUFO1 complementation construct

A single fragment containing 1,859 bases upstream of the *FveUFO1* start codon, the *FveUFO1* (gene19967) sequence (1,329 bp), and 284 bp of the 3' UTR was PCR amplified from YW 5AF7 genomic DNA and cloned into the Gateway binary vector pMDC99 (Curtis and Grossniklaus, 2003) via Gibson assembly. The web-based assembly tool NEBuilder was used to design the Gibson assembly primers (http://nebuilder.neb.com). Primers to amplify the WT *FveUFO1* gDNA sequence are F: 5' GTTAATTAAGGAATTATCGAGCTTCGTACTAGTTGTAGTGG 3' and R: 5' CCAAGCTTGCATGCCCGACTGTCGATGTCGTGTAATG 3'. Primers to amplify the pMDC99 vector backbone are F: 5' TCGATAATTCCTTAATTAACTAGTTCTAGAGC 3' and R: 5' GGCATGCAAGCTTGGCAC 3'. The construct was subsequently transformed into callus generated from sterilized leaves from two mutant M3 plants. Leaves from M3 mutant plants were sterilized by washing with 70% EtOH for 30 seconds followed by incubation in 25% bleach for 10 minutes. Tissue was then rinsed 5 times with sterile water and placed on MS plates (0.44% w/v MS with Gamborg's vitamins, 2% w/v sucrose, 0.7% w/v Phytoblend agar) for callus induction.

Strawberry transformation

Transformation of the complementation construct into plants was accomplished following a published protocol (Slovin et al., 2009; Chatterjee et al., 2011; Kang et al., 2013). Young leaves from YW 5AF7 plants were infiltrated with GV3101 *Agrobacteria* containing the

construct on an MS salt-based agar medium, pHed to 5.8, and supplemented with 2% sucrose, 3.4 mg/L 6-benzylaminopurine, 0.3 mg/L indole-3-butyric acid, and 0.7% agar. After 3 days of co-cultivation in darkness, leaves were transferred to the same medium plus 250 mg/L timentin and 250 mg/L carbenicillin for 2 weeks. Then, leaves were moved again to the same medium plus 2 mg/L hygromycin. After 2 additional weeks, leaves were moved to the same medium plus 4 mg/L hygromycin. Leaves were subsequently moved to fresh plates containing the medium with 4 mg/L hygromycin every 2 weeks until shoots emerged. Healthy shoots were transferred to rooting media (0.01 mg/L IBA, 2 mg/L hygromycin, 2% glucose, 0.5 MS salts, 0.7% phytoagar, pH 5.8). After the development of roots, plants from 8 independent transgenic lines were moved to soil and genotyped using the following primers to amplify a portion of the hygromycin resistance gene from pMDC99: JC174 F: 5' CAGCTTCGATGTAGGAGGGCGTG 3' and JC79 R: 5' CTATTTCTTTGCCCTCGGACGAGTG 3'.

Supplemental Materials (Files available in Google Drive at this address: https://drive.google.com/drive/folders/13LL1JBgt9IIONyq3rhDgfrVGGw1OSaWk?usp=sharing)

Supplemental tables

Table S2.1. Descriptions for all RNA-Seq samples

Table S2.2. Information for each network: total number of genes, number of clusters, average number of genes per cluster, and notes on creation of each cluster

Supplemental figures

Figure S2.1 Screenshots to illustrate navigation of shiny site

Supplemental datasets

Dataset S2.1. Excel file with each network, the numbers of clusters in each network, and the number of genes per cluster

Dataset S2.2. Excel file with connectivity scores for eigengene of each cluster with all tissues (colored based on connectivity)

Dataset S2.3. Standard_hand dissected (Nc_hd) Clusters 2, 14, and 20 GO and full lists of genes

Dataset S2.4. Genes and lists of transcription factors in consensus_lcm clusters 95 and 100

Dataset S2.5. List of *efo* candidate genes based on filtering

2.3: *Results*

2.3.1: WGCNA network analysis of comprehensive flower and fruit RNA-seq data

RNA-seq data for 46 different tissues/stages (46 tissues x 2 biological replicates) were generated previously using *F. vesca* Yellow Wonder 5AF7 (Supplemental Table S2.1; Kang et al., 2013; Hollender et al., 2014; Hawkins et al., 2017). Most tissues were harvested by hand dissection (HD) under a stereomicroscope and young floral tissues were isolated using laser capture microdissection (LCM). Therefore, LCM and HD data were analyzed separately to avoid variation introduced by different techniques (Hollender et al., 2014; Figure 2.1). Co-expression networks were generated by Chris Zawora using the Weighted Gene Co-Expression Network Analysis (WGCNA) package (Langfelder and Horvath, 2008), in which all co-expressed genes are connected to each other with varying correlation

strengths (Supplemental Dataset S2.1). This is accomplished using soft thresholding, thereby preserving the continuous nature of the dataset and eliminating the need to set an arbitrary correlation score cutoff.

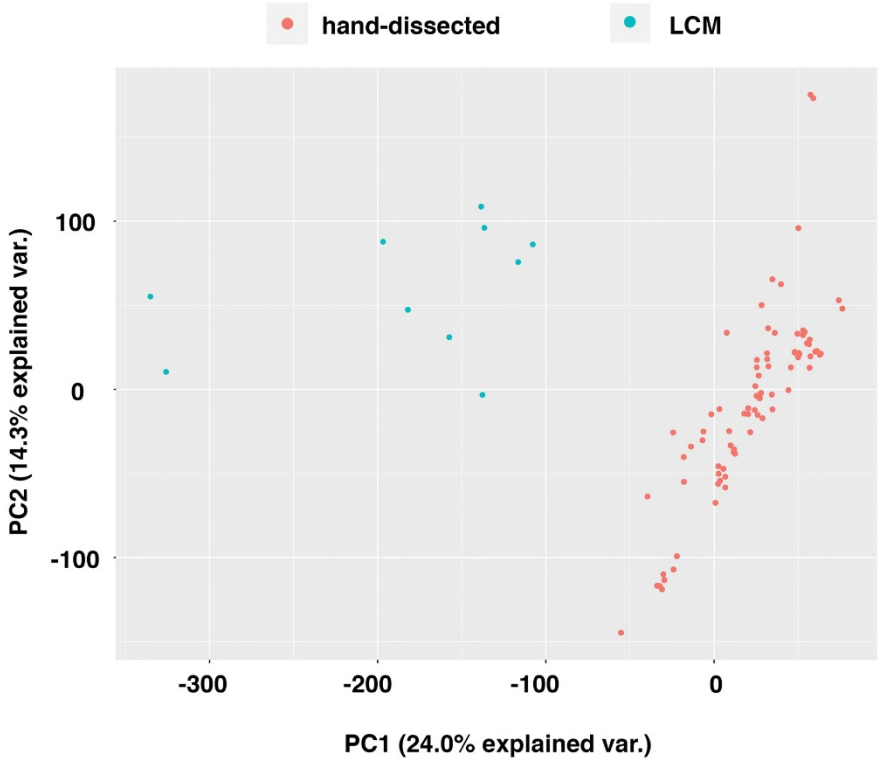


Figure 2.1 Tissue dissection method introduces sample variation
Principle Component Analysis (PCA) plot showing first and second principle components of LCM (blue) and hand-dissected (red) tissues. Tissues separate in the first principle component, indicating variability introduced by harvesting and processing techniques.

When standard parameters (see Methods) are used, the HD tissue network incorporates 33 clusters of co-expressed genes (Figure 2.2A). Eigengenes, the first principal component of a cluster, can be thought of as a representative of a cluster's expression profile. The correlation between each cluster's eigengene and each of the HD tissues was plotted in a heat map (Figure 2.2B; Supplemental Dataset S2.2) to allow for easy visualization of such relationships. For example, cluster 22 correlates with ripening fruits at the turning stage.

Cluster 4 correlates with stage 9-10 anthers. Cluster 10 is more specific to seedlings and leaves. A similar network analysis was applied to the LCM samples and led to 26 clusters. Correlations of clusters with specific tissues are also shown as a heat map (Figure 2.2C; Supplemental Dataset S2.2). The large number of tissues and stages enabled the development of robust co-expression networks, a significant improvement over our lab's previous network analysis utilizing only 16 floral tissues/stages (Hollender et al., 2014).

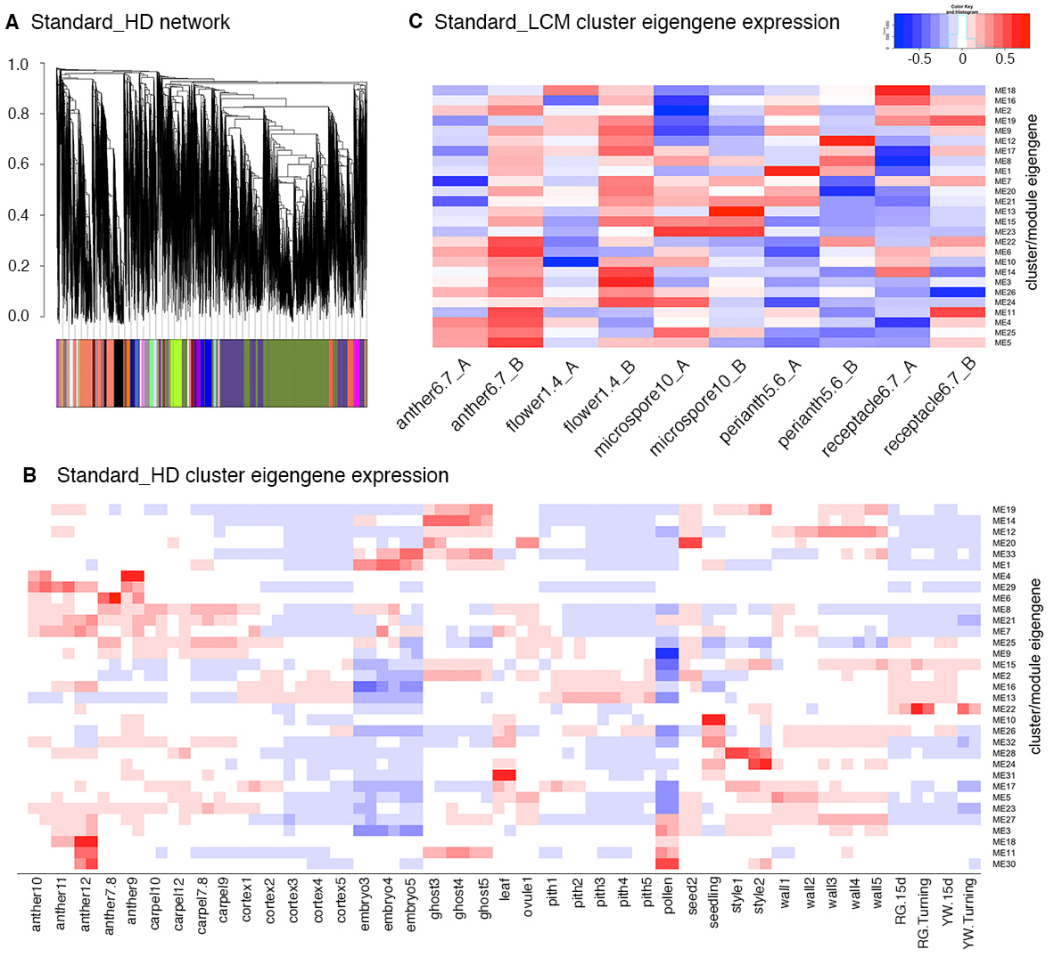


Figure 2.2 WGCNA non-consensus network analyses of 82 hand dissected flower, fruit, and vegetative samples and 10 LCM flower samples

(A) Dendrogram showing co-expression modules identified by WGCNA in standard HD network. Each leaf in the tree is one gene. The major tree branches constitute 33 modules labeled with different colors.

(B) Heatmap showing cluster-tissue association of standard HD network. Each row corresponds to a cluster. Each column corresponds to a specific tissue/stage. The color of

each cell at the row-column intersection indicates the eigengene correlation value on a scale from -1 to 1. Blue color indicates negative correlation and red color indicates positive correlation between the eigengene and the tissue.
(C) Heatmap showing cluster-tissue association of standard LCM network. A total of 26 clusters were identified.

2.3.2: Ghost-associated modules provide insight into iron transport during fruit development

Fertilization initiates the biosynthesis of auxin and GA in seeds; these phytohormones subsequently stimulate fruit set in strawberry (Nitsch 1950; Kang et al., 2013). When a fertilized seed is dissected open to remove the embryo, the remaining seed tissue containing the endosperm and seed coat is referred to as the 'ghost' (Figure 2.3B). Previous transcriptome analysis revealed that auxin and GA biosynthesis genes were transcriptionally induced in the ghost upon fertilization, implicating the importance of the ghost in *F. vesca* fruit set (Kang et al., 2013). Clusters 2, 14, and 20 from the HD network show clear correlation with the ghost (Figure 2.2B and Figure 2.3A). Cluster 20 is correlated with the stage 1 ovule (pre-fertilization) and the stage 2 seed (immediately after fertilization). The top two GO terms in cluster 20 are "Regulation of Fertilization" and "Regulation of Double Fertilization" (Supplemental Dataset S2.3). Abundant MADS Box genes were found among genes in cluster 20 including three annotated as AGL80-like (genes 04949, 15899, and 22916) and four AGL62-like (genes 01789, 07364, 30567, and 07361). Cluster 14 is most strongly associated with ghost stages 3, 4, and 5, all of which are post-fertilization stages. Interestingly, the enriched GO terms are completely distinct from those of cluster 20, suggesting very different molecular events in similar tissues post-fertilization.

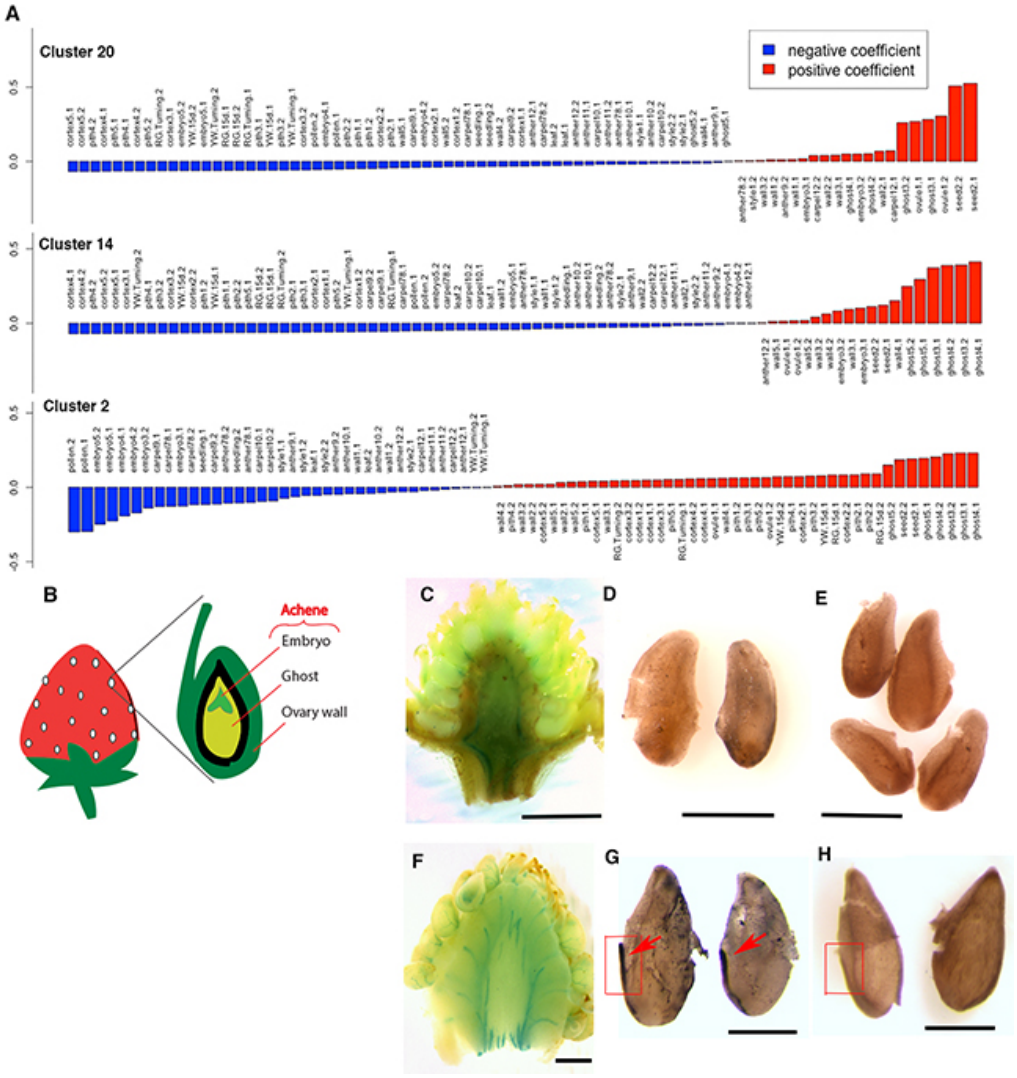


Figure 2.3 Ghost-associated clusters indicate active iron transport after fertilization

Scale bars in C, F: 1mm; D, E, G, H: 0.4mm.
(A) Positive correlation of clusters 20, 14, and 2 with seed and ghost tissues.
(B) Diagram of strawberry receptacle in relation to achene consisting of ovary wall and seed. Each seed consists of ghost and embryo.
(C) Perls staining of the pre-fertilization (stage 1) receptacle.
(D) DAB-enhanced Perls staining of iron in fixed ovules at stage 1 (pre-fertilization). No difference is seen from the DAB-only control in **(E)**.
(E) DAB-only staining of fixed ovules at stage 1.
(F) Perls staining of the receptacle at stage 3 (post-fertilization). Blue lines are stained vascular strands connecting the receptacle to individual achenes.
(G) DAB-enhanced Perls staining of iron in fixed stage 3 seeds (post-fertilization). Strong vascular strand staining is seen along the side of the seeds (arrows). This positive staining contrasts with negative control (rectangle) shown in **(H)**.
(H) DAB-only staining of fixed stage 3 seeds.

Cluster 2 is not only positively correlated with the ghost, but is also negatively correlated with embryos, suggesting endosperm/seed coat-specific molecular events. This cluster is also positively correlated with the cortex and pith tissues of the receptacle. Twenty-eight out of a total of forty enriched GO terms (Biological Process) in cluster 2 are related to iron transport and iron sequestration (Supplemental Dataset S2.3). Cluster 2 contains gene19831, which is annotated as a homolog of the phloem-specific iron transporter *OLIGOPEPTIDE TRANSPORTER 3* (*OPT3*) (Stacey et al., 2002, 2008; Zhai et al., 2014), three metal binding proteins (genes 08918, 10308, and 18489), and two members of the *VACUOLAR IRON TRANSPORTER 1 (VIT1)* family (genes 32625 and 17575) (Kim et al., 2006). The GO terms enriched in cluster 2 suggest that the ghost and receptacle, but not embryo, carry out active iron transport during the earliest stages of strawberry fruit development.

To test the above hypothesis, John Sittmann and I stained free iron in the receptacle and seed using the iron-specific Perls stain (Green and Rogers, 2004; Stacey et al., 2008; Roschzttardtz et al., 2009; Brumbarova et al., 2014). Potassium ferrocyanide, a component of the Perls reagent, reacts with iron to form an insoluble pigment known as Prussian blue. Increased free iron in the vascular tissue of the receptacle was observed post-fertilization as abundant blue strands connecting the receptacle to individual achenes (Figure 2.3C, F). A subsequent intensification reaction with 3,3'-Diaminobenzidine (DAB) was previously shown to enhance Perls staining and produce a dark brown pigment (Nguyen-Legros et al., 1980; Roschzttardtz et al., 2009). Achenes were dissected open to isolate seeds, which were stained with Perls and then DAB. Strongest staining was observed in the strands of vasculature connecting the seed to the subtending receptacle (Figure 2.3G; arrows). The

ovules (precursors of seeds) did not show vascular strand staining (Figure 2.3D). Seeds fixed and treated with only DAB served as negative controls (Figure 2.3E, H) since DAB alone is unable to directly stain iron (Roschzttardtz et al., 2009; see Methods). The significant increase in iron transport from the receptacle to the seed post-fertilization is consistent with increased iron transporter expression in the cluster 2 network. The requirement of an iron cofactor for the GA biosynthetic enzymes GA20ox and GA3ox (Huang et al., 2015; White and Flashman, 2016) exemplifies one of the many post-fertilization molecular events that require iron (Farrow and Facchini, 2014).

2.3.3: Consensus networks provide robust and reproducible clusters

Because of the potential for noise and instability in standard co-expression network analysis, Chris Zawora applied a consensus-clustering approach (Wu et al., 2002; Monti et al., 2003) as an extension to WGCNA. This strategy is independent of parameter selection and ensures module reproducibility. Spurious co-clusterings are reduced by testing the stability of clusters in response to sampling variability. Sampling variability, or perturbations of the dataset, can be simulated with a re-sampling approach. Therefore, 1,000 runs of the WGCNA clustering algorithm were performed with each run resampling 80% of the genes and using randomly generated parameter selections (details in Methods).

Consensus clustering of HD and LCM samples yielded 86 and 123 clusters, respectively (Figure 2.4A, B; Supplemental Dataset S2.1). Haley Wight evaluated the consensus HD clusters against the standard (WGCNA) HD clusters (Figure 2.5). Although both approaches yielded a similar number of clusters of comparable size (Figure 2.5A), the clusters had little overlap as shown by Jaccard Index (Figure 2.5B). The Jaccard Index is a

statistic used for comparing the similarity and diversity of sample sets ranging from 0 to 100% (Bass et al., 2013). The higher the percentage, the more similar the two clusters. While many clusters had a significant overlap (hypergeometric p-value < 0.05) across methods, the Jaccard Index was never higher than 50%.

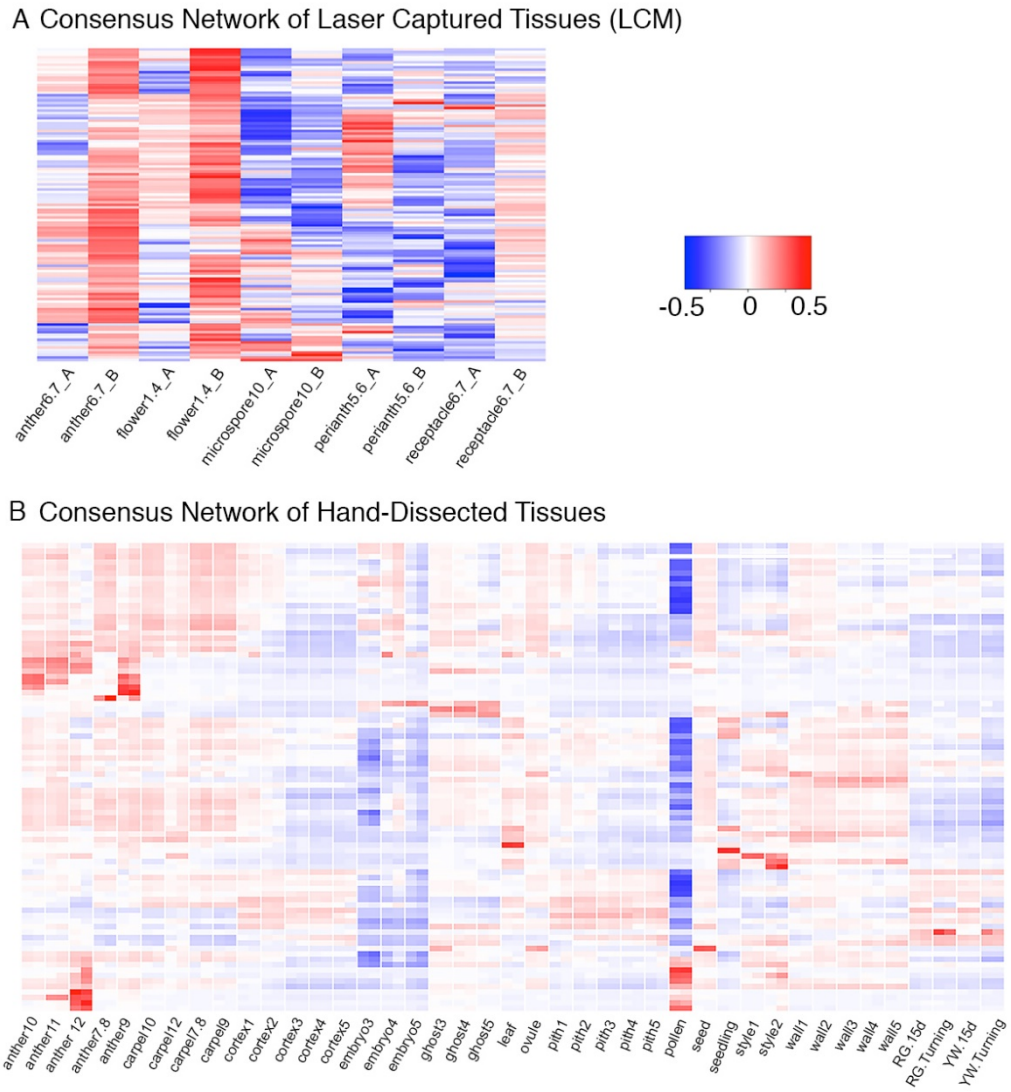


Figure 2.4 Consensus network analyses of LCM and hand-dissected flower, fruit, and vegetative tissues

(A) Heatmap showing cluster-tissue association of consensus LCM network. A total of 123 clusters were identified. Each row corresponds to a cluster. Each column corresponds to a specific tissue/stage. The color of each cell at the row-column intersection indicates the eigengene correlation value on a scale from -1 to 1. Blue color indicates negative

correlation and red color indicates positive correlation between the eigengene and the tissue.
(B) Heatmap showing cluster-tissue association of consensus_hd network. 86 clusters were identified. Two replicates of each tissue are labeled with one name.

A

	Cluster number	Ave. cluster size	Ave. correlation between clustered gene pairs	RS Statistic
Standard	80	324.70	0.5423	0.7672
Consensus	86	329.69	0.5337	0.7453

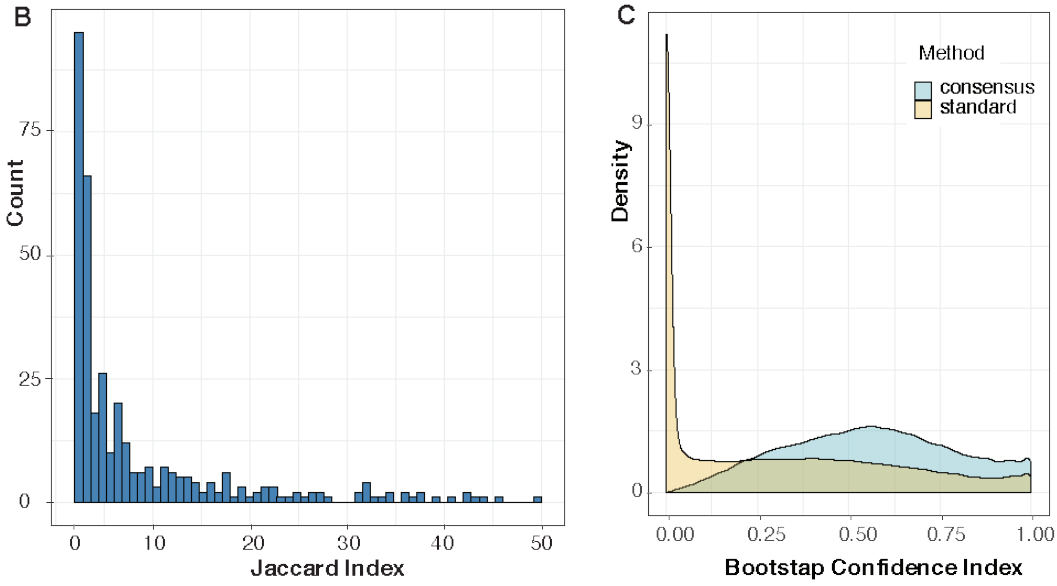


Figure 2.5 Comparison of Standard WGCNA to Consensus Clustering Method
(A) Statistics describing clustering behaviors of both methods.
(B) Similarity between standard and consensus clusters with significant overlap.
(C) Bootstrap confidence intervals of gene pairs within the same cluster.

Since the standard WGCNA and consensus algorithms define clusters that are not known *a priori*, Haley Wight also measured the quality of clustering by internal evaluation criterion. Typical objective functions in clustering aim to attain high intra-cluster similarity and low inter-cluster similarity. The RS value, sometimes referred to as the pseudo-F statistic, is a ratio of the variance between clusters to the variance within clusters, thereby defining the proportion of variation explained by a particular clustering of genes (Sharma,

1995; He et al., 2015) (details in Methods). Both the standard WGCNA and consensus methods performed similarly with regard to RS statistic and intra-cluster correlation (Figure 2.5A). However, bootstrap confidence intervals are significantly higher for consensus clustering (Wilcoxon p-value < 0.05; Figure 2.5C). Bootstrapping allows the assignment of measures of accuracy; a high confidence interval implies that if the entire study were repeated *ad infinitum,* the resulting gene pairs would be the same. This result demonstrates that the consensus clustering method, while preserving the same level of correlation between clustered gene pairs as the standard WGCNA method, also produces clusters with higher reliability.

To maximize the potential of identifying pairs of genes with functional relationships, consensus90 and consensus100 networks were also generated. These networks apply stringent cutoffs of 90% and 100% to the consensus matrix; only genes that cluster together 90% or 100% of the time in the consensus network appear in the consensus90 and consensus100 networks, respectively. The HD and LCM consensus90 networks contain 2,870 and 6,332 clusters, respectively (Supplemental Table S2.2), a significant increase over the number of clusters in the consensus network. Accordingly, each cluster in the consensus90 HD and LCM networks has fewer genes. On average, six genes per cluster in the HD network and five genes per cluster in the LCM network (Supplemental Table S2.2). The HD and LCM consensus100 networks contain 962 and 3,814 clusters, respectively. This decrease in cluster number as compared to the consensus90 networks is due to the decreased total number of genes included in each network as the majority of genes do not cluster with any partners 100% of the time in the consensus matrix. However, genes paired in the consensus100 networks are reliable candidates for functional relationships.

2.3.4: Website offers a user-friendly interface for exploring co-expression networks

To facilitate utilization, exploration, and visualization of the co-expression networks, Chris Zawora and Haley Wight generated the ***Fragaria vesca* gene co-expression network explorer**, a user-friendly web interface, using the Shiny application from R Studio (http://shiny.rstudio.com). The site (www.fv.rosaceaefruits.org) hosts data from the standard, consensus, consensus100, and consensus90 networks for both the HD and LCM datasets as well as ripening fruit tissue-only dataset (4 network types x 3 datasets =12 networks). Users can first choose a specific network to explore (Supp Figure S2.1 A) and subsequently retrieve general information such as the number of clusters in the network or search for a specific strawberry gene to determine in which cluster the gene resides (Supp Figure S2.1 B). Users may also identify a cluster that correlates with a specific tissue type by selecting the 'Clusters' tab (Supp Figure S2.1 C). This generates a list of the top five positively and negatively correlated tissues for the cluster eigengene. This information is also visually displayed on a heatmap under the 'Tissue-Eigengene Expression' tab (such as Figure 2.2 B, C and Figure 2.4 A, B). Users can further obtain detailed information for a specific cluster by choosing a cluster number from the drop-down menu (Supp Figure S2.1 D). Cluster-specific information includes a list of genes in the cluster with annotations and Arabidopsis homologs, a bar graph plotting the cluster's eigengene association with each of the profiled tissues in the network (such as Figure 2.3 A), enriched GO terms, and a plot indicating correlation between clusters (Supp Figure S2.1 E). A downloadable connectivity score file available for each cluster under the 'Downloads' tab can easily be imported into Cytoscape for network visualization as shown in Figure 2.6.

2.3.5: Consensus networks identify potential floral meristem and receptacle meristem regulators

I sought to test the ability of the consensus clusters to predict functional relationships between genes. Based on the consensus_LCM network, cluster 95 (387 genes) appears to correlate more strongly with the young floral meristem, while cluster 100 (244 genes) correlates with floral stage 6-7, the stage at which the receptacle enlarges (Supplemental Dataset S2.2). Transcription factors with the strongest connections (cutoff 0.8 on a scale of 0 to 1) from clusters 95 and 100 were visualized using Cytoscape (Figure 2.6 C, D). Cluster 100 is particularly rich in transcription factors involved in meristem regulation; 15 of the 38 transcription factors encode meristem regulators, including *FvWUS* (gene30464), *FvSTM* (gene19507), and *FvWOX9* (gene28935). Interestingly, seven TALE homeodomain proteins are in cluster100 while cluster 95 has only one. Together, the abundance of meristem regulators in cluster 100 supports previous anatomical and network analyses (Hollender et al., 2012; 2014) suggesting that the receptacle is a floral organ with meristematic activity.

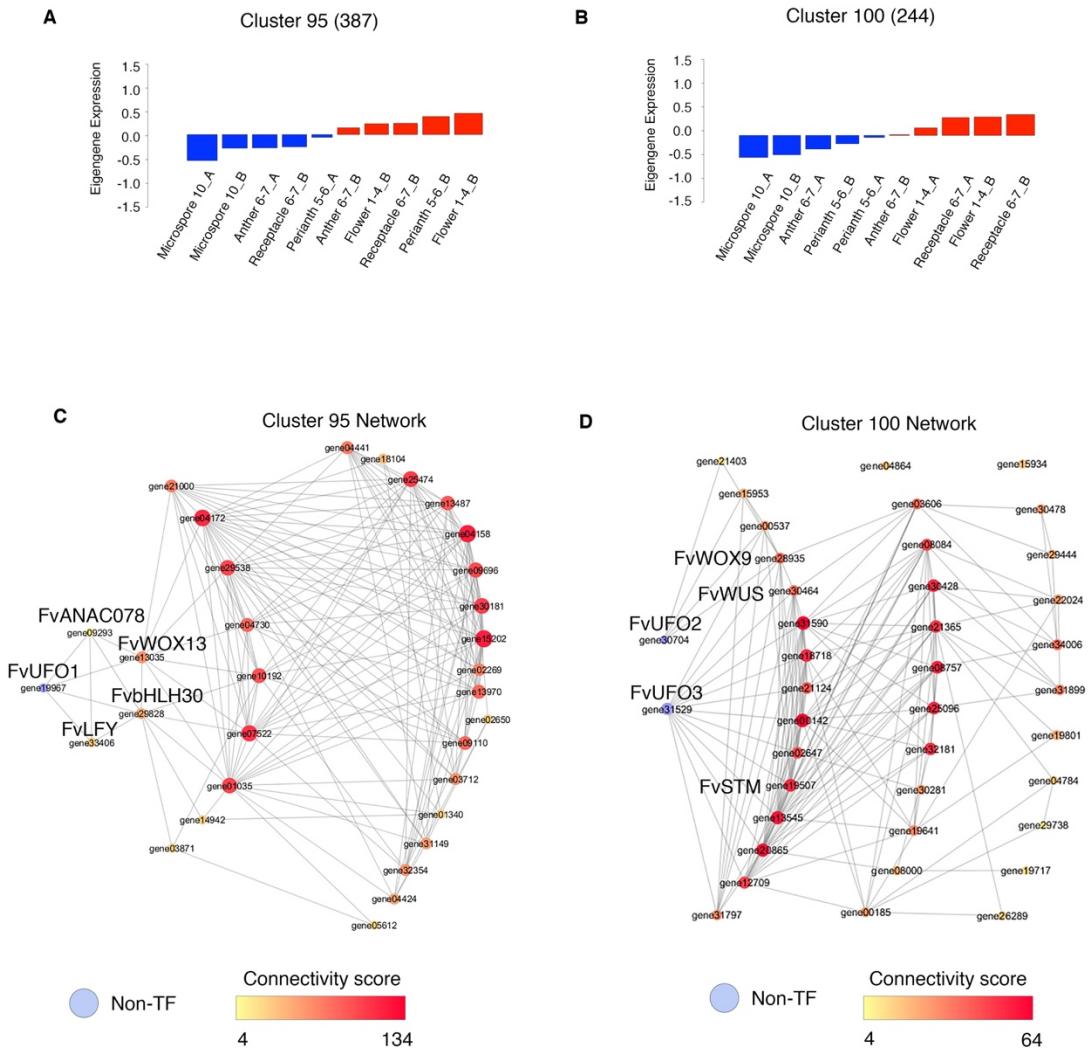


Figure 2.6 Young receptacle-associated clusters and transcription factor networks

(A) Consensus_lcm cluster 95 eigengene expression bar graph
(B) Consensus_lcm cluster 100 eigengene expression bar graph
(C) Network showing connections between *FveUFO1* and transcription factors in cluster 95. Edge cutoff is 0.8. Each colored circle (node) represents one gene. Larger node size and darker red node color indicate greater connectivity within the network.
(D) Network showing connections between *FveUFO2*, *FveUFO3*, and transcription factors in cluster 100. Edge cutoff is 0.8. Larger node size and darker red node color indicate greater connectivity within the network.

Cluster 95 is more closely correlated with the young floral bud (stage 1-4 flower); 31 genes in this cluster are transcription factors, eight of which are meristem regulators.

Interestingly, *FveLFY* is found in this cluster and not in cluster 100, indicating *FveLFY* (gene33406) in promoting the early stage floral meristem development. Gene04172, annotated as a *CO*-like transcription factor, is also in this cluster, suggesting a role for a *CONSTANS* family member in regulating the floral meristem.

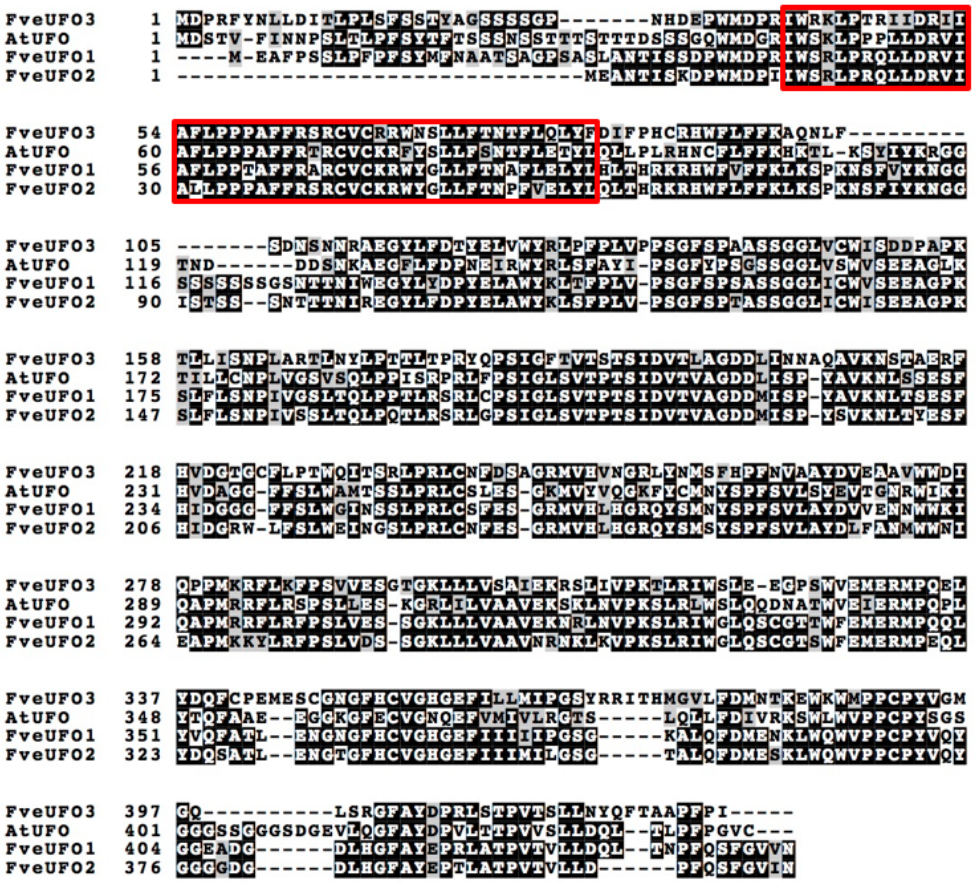


Figure 2.7 Three *F. vesca* genes show sequence homology to UFO

Alignment of amino acid sequences of the three *F. vesca UFO* homologs and the *Arabidopsis UFO* (At1G30950). Red square indicates the conserved F BOX domain.

Although *UFO* does not encode a transcription factor, it was shown in *Arabidopsis* to be an important regulator of *LFY* in flower development. Interaction between *UFO* and *LFY* is required to activate the B class gene *AP3* for petal and stamen identity specification (Lee

et al., 1997; Chae et al., 2008). Three *UFO* homologs in strawberry (Figure 2.7) are found in clusters 95 and 100; *FveUFO1* is in cluster 95 (Figure 2.6C; Supplemental Dataset S2.4) and is strongly correlated with *FveLFY* (edge score of 0.86). In addition to *FveLFY*, the three transcription factors in cluster 95 with the highest correlation to *FveUFO1* are *FvWOX13* (gene13035), *FvbHLH30* (gene29828), and NAC domain-containing transcription factor *FvANAC078* (gene09293). Interestingly, *FveUFO2* (gene30704) and *FveUFO3* (gene31529) are in cluster100 (Figure 2.6D, Supplemental Dataset S2.4), suggesting that perhaps *FveUFO2* and *FveUFO3* act later during flower development and may not be involved in the regulation of *FveLFY,* which is absent from cluster 100. This fine separation between *FveUFO1* and *FveUFO2/3* indicates the possibility that *Fve*UFO1 is not redundant but acts independently of *FveUFO2* and *FveUFO3*. Altered or abolished *FveUFO1* function would hence be likely to produce a floral phenotype.

2.3.6: Identification of a nonsense mutation in FveUFO1

Through an EMS mutagenesis screen of Yellow Wonder 5AF7 (Hollender, 2012), I identified a floral mutant, hereafter called *extra floral organs (efo*), with defects in both floral meristem determinacy and floral organ development (Figure 2.8). Specifically, the floral meristem has shoot meristem characteristics; a single flower can give rise to secondary and tertiary flowers (Figure 2.8 B, C, E). The secondary and tertiary floral buds arise from the axials of sepals or leaf-like organs and resemble the *Arabidopsis ap1* mutants, where new flowers are formed in the axials of sepals (Irish and Sussex, 1990). *efo* also exhibits a repeated sepal-petal-stamen pattern before terminating in an enlarged receptacle topped with supernumerary carpels (Fig. 2.8H), which resembles the weak *Arabidopsis agamous-4 (ag-4)* mutant flower (Sieburth et al., 1995). In addition, the sepals,

petals, and stamens of *efo* often exhibit mosaic organ identity (Figure 2.8I); sepals contain white petal-like patches and petals develop out of the anthers. This mosaic organ phenotype bears resemblance to the *Arabidopsis ufo* mutants (Levin and Meyerowitz, 1995). Hence, *efo* appears to exhibit defects in A, B, and C classes of floral homeotic genes.

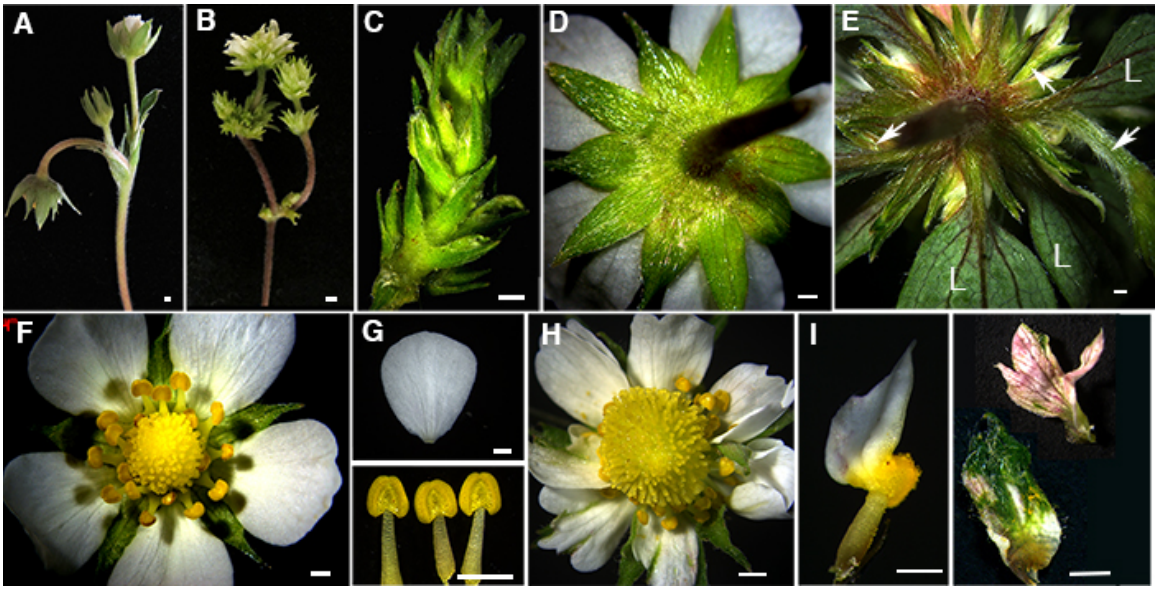


Figure 2.8 Phenotype characterization of *efo/FveUFO1*

(A) A wild type shoot showing the primary flower (bending) and the secondary and tertiary flowers.
(B) *efo* mutant flower showing many more flowers originating from what would be a single flower in wild type.
(C) *efo* flower showing elongated internode between whorls of sepals/leaves and axillary flower buds.
(D) The back of a wild type flower showing 5 bracts in the outermost whorl and 5 sepals in alternating positions.
(E) The back of a mutant flower showing many whorls of leaves (L) or leaf-like organs, in the axil of which many young flower buds reside (arrows).
(F) Wild type flower.
(G) A wild type petal (top) and three wild type stamens (bottom).
(H) *efo* flower showing a larger central receptacle giving rise to more carpels/ovaries than wild type in **(F)**. The central receptacle is flanked by a whorl of stamens and then a whorl of petals, then a whorl of sepal-like organs. This central flower is on top of additional whorls of stamens, petals, and sepals.
(I) Mosaic stamen/petal organs are often seen in the mutant flowers.
(J) Mosaic sepal/petal organs are often seen in the mutant flowers.

Using a bulk segregant mapping-by-sequencing approach (Schneeberger et al., 2009; Cuperus et al, 2010; Hartwig et al., 2012), Dr. Wanpeng (George) Wang and I identified a candidate, recessive mutation in *efo*. Specifically, genomic DNA from 3 mutant and 22 wild type plants, all of which were derived from an M2 parent plant heterozygous for the mutation, was pooled and submitted for whole genome sequencing. An analysis pipeline with a series of filtering steps yielded 97 variants (Supplemental Dataset S2.5), only one of which was in a gene previously known to regulate flower development. Gene19967 encodes *FveUFO1*; a highly conserved W residue in the C-terminus was mutated to a STOP codon in the mutant (Figure 2.9).

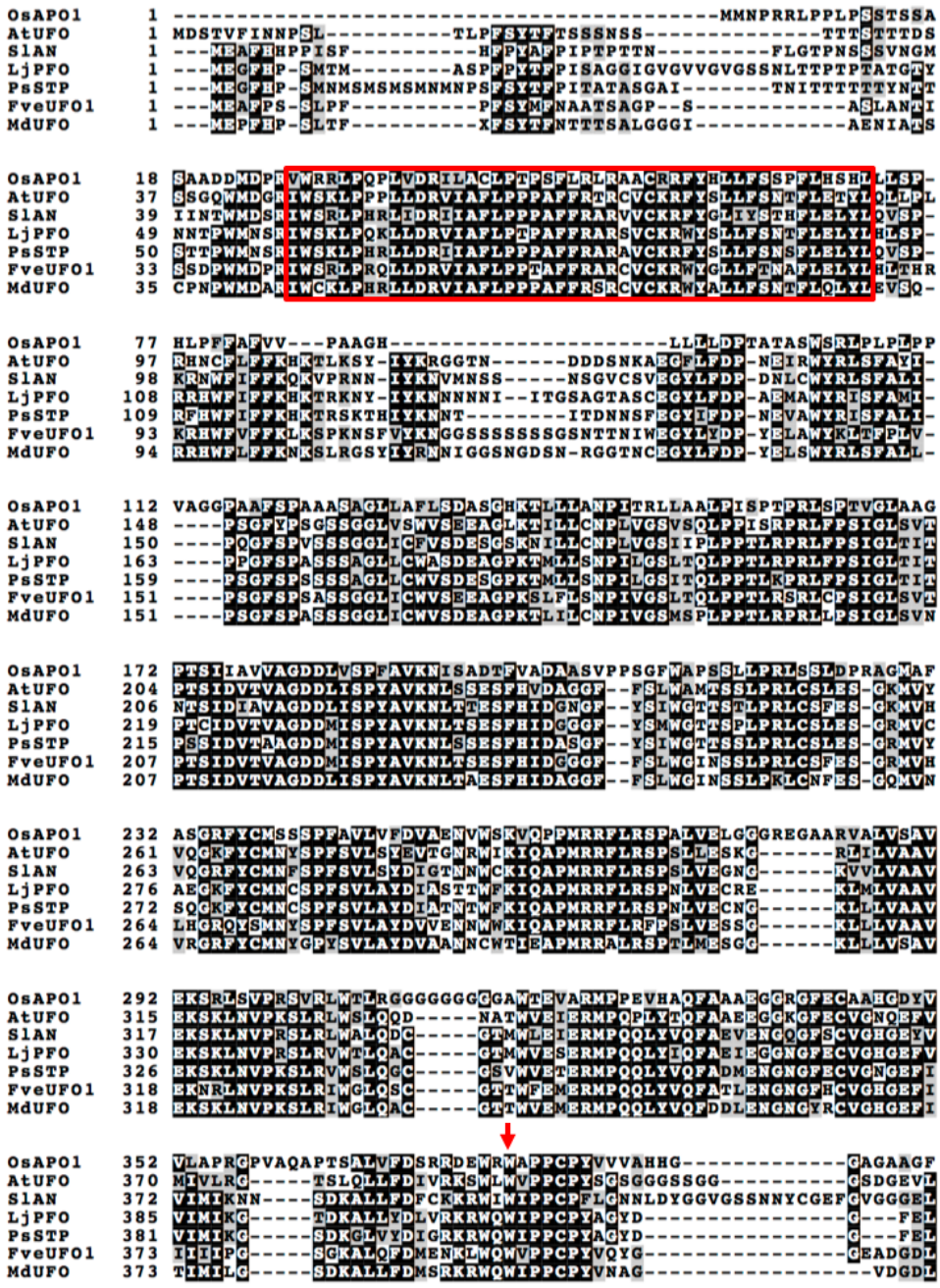


Figure 2.9 *UFO-like* genes are conserved across monocot and dicot species

Alignment of amino acid sequences of characterized UFO homologs from multiple plant species including *Arabidopsis UFO* (At1G30950), *Solanum lycopersicum ANANTHA* (NP_001234215), *Lotus japonicus PFO* (AY156687.1), *Pisum sativum STP* (AY274933.1), and *Oryza sativa APO1* (AB292777.1). The sequence of a *UFO-like* gene from *Malus domestica* (XP_008383315.1), a species closely related to *Fragaria vesca*, is also included. Red square indicates the conserved F BOX domain. Red arrow marks the position of the Trp (W) to STOP mutation of the *efo* mutant plants.

qRT-PCR indicates that the transcript level of *FveUFO1* is increased by 13-fold in stage 1-4 flowers of the mutant versus the wild type (Figure 2.10). This suggests that the mutant phenotype is not a result of the nonsense-mediated decay pathway targeting *UFO1* transcripts for degradation, perhaps due to the nonsense mutation occurring near the C-terminal end. Increased expression in *efo* may result from a different flower tissue composition in the mutant flower or could be a result of a positive feedback loop.

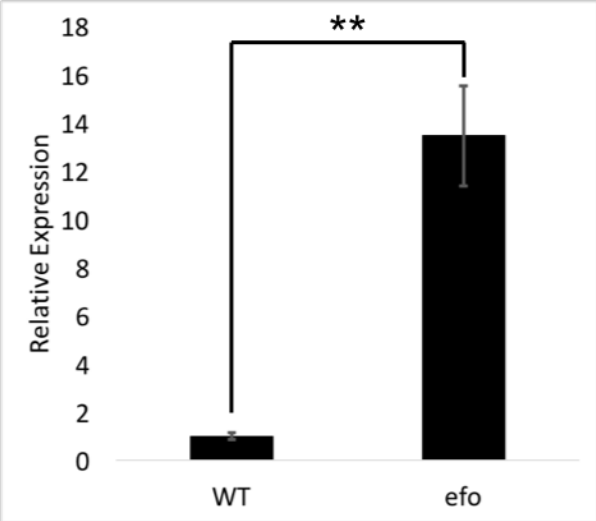


Figure 2.10 *FveUFO1* is more highly expressed in the *efo* mutant compared to WT
qRT-PCR data showing *FveUFO1* expression level in wild type and *efo*. *FveUFO1* is 13-fold higher in *efo* when compared with wild type. Bars represent the average expression +/- standard error relative to WT (set to 1.0) for four biological replicates. Transcript abundance was calculated relative to *FvePP2a*. Statistical significance was calculated with a Student's t-test. ** $p < 0.001$.

2.3.7: First attempt to complement FveUFO1 in efo background was unsuccessful

To definitively show that the floral phenotype of *efo* plants was caused by the identified SNP in *FveUFO1*, I sought to introduce a wild type copy of *FveUFO1* into the *efo* mutant background. If the mutant phenotype was indeed caused by non-functional *fveufo1*, I expected *F. vesca* plants stably expressing a WT copy of *FveUFO1* to show a WT floral

phenotype resulting from complementation. The *efo* mutant proved to be a very difficult genetic background into which to transform the complementation construct; I tried nearly continuously for 2.5 years. After finally generating transgenic plants, I found that their floral phenotype was identical to that of the *efo* mutant. At the same time, a postdoc in our lab, Julie Caruana, also observed that a different EMS-generated *F. vesca* mutant also failed to show complementation when transformed with a construct generated with the same vector stock that I used for my experiment (see Methods). Julie had additional strong data supporting the SNP she identified as the causal mutation underlying her mutant plant's phenotype (Caruana et al., 2017). Combined with Julie's similar experience, I suspect that the failure of *FveUFO1* to complement the *efo* floral phenotype may be due to a problem with the vector stock rather than a result of identifying the wrong causal mutation.

2.4: *Discussion*

This chapter demonstrates how standard co-expression network and consensus network analyses can be used to highlight phenomena like increased iron transport to developing fruit and seeds immediately post-fertilization and the meristem-like nature of the receptacle. This work demonstrates the power of co-expression network analyses in hypothesis building and testing, especially in a developing model system. An intuitive and freely available web interface makes it possible for biologists to explore and mine these networks.

2.4.1: Consensus vs. standard networks

In this study, we generated two types of networks. First, the standard networks were generated by following the published WGCNA analysis pipeline (Langfelder and Horvath,

2008). Second, robust consensus networks were generated by varying parameters and simulating sampling variability over 1,000 runs of the WGCNA clustering algorithm. In total, twelve independent networks were generated; they are standard (non-consensus), consensus, consensus90, and consensus100 networks, respectively for hand dissected tissues, LCM tissues, and ripening fruit tissues (4 network types x 3 datasets =12 networks) (Supplemental Datasets S2.1 and S2.2; Supplemental Tables S2.1 and S2.2).

For the standard networks, Chris Zawora and I chose parameters with the goal of generating a smaller number of clusters. However, correlation scores between pairs of genes are sensitive to the user-selected parameters, including minimum module size, power transformation, and merging on eigengenes. The consensus networks (Wu et al., 2002; Monti et al., 2003) are less influenced by parameter selection due to 1,000 iterations of the WGCNA clustering algorithm with each run using randomly generated algorithm parameters and resampling 80% of the genes. This simulates sampling variability. For most statistics, the standard and consensus methods perform similarly; however, the consensus method greatly increases bootstrapping confidence. As a result, gene pairs that reliably cluster together in a consensus network are more likely to represent true relationships. The consensus clustering approach produces a more robust network without sacrificing the high intra-cluster similarity or low inter-cluster similarity produced by the WGCNA algorithm.

Further, Chris Zawora applied 90% and 100% cutoffs to the consensus matrix and used the remaining genes to generate consensus90 and consensus100 networks, respectively. These networks contain fewer total genes than the consensus networks but many more clusters

with fewer genes per cluster. By adding stringent 90 and 100% cutoffs, we restricted clusters to genes that are the most likely to have functional relationships.

2.4.2: The role of UFO in strawberry flower development

UFO, an F box protein, associates with an SCF complex and targets proteins for degradation via ubiquitination (Samach et al., 1999; Ni et al., 2004). In *Arabidopsis*, UFO was shown to promote LFY transcription factor activity in a positive feedback manner by targeting LFY for degradation (Chae et al., 2008). The primary floral defects reported in *Arabidopsis ufo* mutants include reduced B class gene expression and mosaic organs with unclear boundaries between petals and stamens (Levin and Meyerowitz, 1995).

It is interesting that *FveUFO1* resides in cluster 95, while *FveUFO2* and *FveUFO3* reside in a cluster 100. Hence, FveUFO1 and FveUFO2/3 are likely involved in different temporal processes during development. The network thus predicts that *FveUFO1* may not have a redundant factor in *F. vesca* and acts at the earliest stages of floral meristem formation. The co-expression relationship between *FveUFO1* and *FveLFY* in cluster 95 supports a conserved relationship between *FveUFO1* and *FveLFY* required for regulating strawberry flower development. Indeed, the strawberry *efo* mutant is much stronger in phenotype than the *Arabidopsis ufo* mutant in that it shows abnormal development reminiscent of a combination of class A, B, C, and E mutants. Perhaps FveUFO1 has a more expanded role and is involved in regulating FveLFY in all four whorls of the flower, which in turn promotes the expression of class A (*AP1),* B *(AP3),* C *(AG),* and E (*SEP*) genes. It will be interesting to knock out *LFY* in strawberry and compare the phenotypes to determine if all aspects of the *fveufo1* phenotype are mediated through LFY.

While I have not yet validated with other approaches that the *efo* phenotype is indeed caused by the nonsense mutation in *FveUFO1*, I noticed a striking similarity in the mutant phenotype between *efo* and mutants of *UFO* orthologs in pea (*stp*), *Lotus japonicus (pfo*), tomato (*an*), and *Torenia fournieri* (*Tfufo*), including loss of floral meristem determinacy, proliferating sepals, and in particular the production of ectopic flowers within the primary flowers of pea *stp* mutants (Taylor et al., 2001; Zhang et al., 2003; Lippman et al., 2008; Sasaki et al., 2012). These similarities provide strong support for *FveUFO1* as the gene that underlies the *efo* mutant phenotype in strawberry and suggests a more expanded role of UFO-LFY in regulating all floral whorls in *Solanaceae* and *Rosaceae* species.

Based on a similar experience of another member of our lab, I suspect that the failure of *FveUFO1* to complement the *efo* mutant phenotype may be due to a problem with the vector used to make the complementation construct rather than identification of the wrong causal gene. Since the *efo* mutant background proved to be a difficult genotype to transform, my plan going forward is to generate RNAi or CRISPR constructs targeting *FveUFO1*. Our collaborators at the USDA ARS lab at Epcot will transform these constructs into the WT Yellow Wonder background. A resulting mutant phenotype similar to that of *efo* will provide additional support for the identified SNP in *FveUFO1* as the causal mutation.

2.4.3: Iron transport during fruit set

Fruit set is the process of fruit initiation triggered by fertilization-induced auxin and GA production. Previously, our lab showed that both auxin and GA can be applied exogenously to the receptacle to stimulate its enlargement even in the absence of fertilization (Kang et

al., 2013). In particular, the requirement of an iron co-factor for the GA biosynthetic enzymes GA20ox and GA3ox (reviewed by Yamaguchi, 2008; Huang et al., 2015; White and Flashman, 2016), highlights the importance of iron in addition to GA and auxin for fruit development. However, the requirement for iron is not limited to GA biosynthesis; iron is broadly important in plant metabolism as a co-factor for other 2-oxoglutarate/Fe (II)-dependent dioxygenases. In addition to GA biosynthesis, iron is also required for GA catabolism, ethylene biosynthesis, auxin catabolism, and salicylic acid catabolism (reviewed by Farrow and Facchini, 2014).

Staining of pre- and post-fertilization receptacles with the iron-specific Perls reagent reveals that iron is more abundant post-fertilization in the vasculature connecting the receptacle to the overlying achene and seed indicating active iron transport soon after fertilization. Iron unloading to seeds during development has previously been described (Grusak et al., 1994), though details on the fertilization-induced signals that initiate iron transport to seeds are unknown. DAB-enhanced Perls staining of the vascular strand within the seed indicates the transport route from stem/receptacle to the ghost. Since the embryo is oppositely correlated with cluster 2, my result supports the hypothesis that it is the ghost (endosperm), and not the embryo, that is chiefly involved in auxin and GA biosynthesis in the seed (Kang et al., 2013) and hence the major site of iron transport.

The network analyses presented in chapter 2 provide new insights into the biological processes underlying flower development and fruit set. This work demonstrates how co-expression networks can lead to new hypotheses and guide subsequent experiments. Further, the networks can be more broadly appreciated and utilized by the research

community through a freely available web interface. Anyone with an interest in a specific biological process can easily explore and mine all twelve networks. Therefore, the work reported here sets an example of how co-expression network analyses generated with large-scale RNA-seq data can facilitate research in emerging model systems.

Chapter 3: Investigating the function of *FvFT1*, a homolog of *FLOWERING LOCUS T*, in strawberry fruit development

3.1: *Introduction*

FLOWERING LOCUS T (FT) integrates both environmental and endogenous signals as a central regulator of the switch from vegetative to reproductive development. In *Arabidopsis*, *FT* receives signals from the photoperiod and gibberellin pathways, coordination between which induces flowering (Reeves and Coupland, 2001; Corbesier and Coupland, 2005; Galvao et al., 2012; Porri et al., 2012; Hou et al., 2014; Nguyen et al., 2015). After transcription and translation in the leaf vasculature, the FT peptide travels to the shoot apical meristem (SAM) via the phloem as a long-distance signal. An endoplasmic reticulum membrane protein, *FT INTERACTING PROTEIN 1* (*FTIP1*), has been shown to mediate *FT* peptide transport (Liu et al., 2012). In the leaf, *FTIP1* is required for FT peptide export from companion cells, the site of transcription and translation, into the phloem sieve elements. After arriving at the SAM, FT forms a transcriptional regulatory complex with the transcription factor FLOWERING LOCUS D (FD) and upregulates floral meristem identity genes to initiate flower formation (Abe et al., 2005; Corbesier et al., 2007).

Since its identification as the florigen, homologs of FT have been identified from other angiosperms and implicated as non-cell autonomous, floral initiation signals in monocotyledonous and dicotyledonous species, including annuals (Lifschitz et al., 2006; Lin et al., 2007) and perennials (Endo et al., 2005; Bohlenius et al., 2006; Hsu et al., 2006;

Kotoda et al., 2010). The genome of *Fragaria vesca*, a perennial, encodes three *FT-like* genes. Of the three, *FvFT1* was shown to be a positive regulator of flowering time and a likely ortholog of *FT* (Shulaev et al., 2011; Koskela et al., 2012). Functional characterization of *FvCO* revealed conservation of its role in photoperiodic control of reproductive development in *F. vesca*. *FvCO* upregulates *FvFT1* expression in leaves when plants are exposed to light and, more specifically, controls *FvFT1* diurnal oscillations (Kurokura et al., 2017).

Like strawberry, a number of species have multiple *FT*-like genes, some with unique temporal or spatial expression patterns as well as individual functions. For example, sunflower (*Helianthus anuus*) contains four *FT-like* genes, each with a different expression pattern and unique mutations that have resulted in gains and losses of function. Diversification of *FT* function has been implicated in the variety of photoperiods that induce flowering in different species of the sunflower genus, a trait which was exploited during breeding of cultivated sunflower varieties (Blackman et al., 2010).

Although *FT* function in floral promotion is widely conserved, recent evidence from multiple species indicates regulatory roles for *FT-like* proteins in diverse developmental processes such as vegetative growth, tuberization, and stomatal control (reviewed by Pin and Nilsson, 2012). These novel roles are also attributable to gene duplication events followed by sub- or neo-functionalization. The first reports of *FT* function in a process other than floral initiation came from studies of poplar trees, which contain two paralogous *FT-like* genes: *PtFT1* and *PtFT2.* Both genes contribute to floral initiation, as evidenced by early flowering phenotypes when overexpressed, but one or both also contribute to short

day-induced vegetative growth cessation and bud set (Bohlenius et al., 2006; Hsu et al., 2006). *PtFT1* was subsequently shown to be expressed during late winter while *PtFT2* is expressed during active vegetative growth, suggesting that *PtFT2* is likely to be primarily responsible for regulation of growth cessation and bud set (Pin and Nilsson, 2012).

The potato (*Solanum tuberosum*) genome has four *FT-like* genes, two of which induce early flowering when overexpressed (Navarro et al., 2011). Of the two, *StSP3D* shows the strongest homology to the tomato *SFT*, previously demonstrated to be a mobile floral initiation signal (Lifschitz et al., 2006). *StSP3D RNAi* plants correspondingly have a late flowering phenotype. Intriguingly, plants expressing RNAi against the second paralog, *StSP6A*, show no flowering time defects but instead have defective tuberization. *Solanum tuberosum* plants normally develop tubers only under short day conditions, but plants overexpressing *StSP6A* formed tubers in non-inductive long days (Navarro et al., 2011). Both *StSP3D* and *StSP6A* are expressed in leaves, though *StSP6A* is only induced in response to a short day photoperiod. Navarro et al. (2011) present the exciting possibility that functional divergence after duplication of *FT-like* genes in potato resulted in two independent mobile signals that control differentiation of the floral and tuber meristems.

As the florigen, FT is known for its movement between organs. However, Kinoshita et al. (2011) report a novel, cell-autonomous function in control of *Arabidopsis* stomatal guard cell aperture. In photosynthetic tissues, oxygen and carbon dioxide are exchanged through pores known as stomata. The aperture of each stomata is tightly regulated by two guard cells to minimize water loss. To increase the aperture, water flows into the guard cells and the resulting increase in turgor pressure pulls them apart, thereby widening the stomatal

opening. This osmosis is caused by K^+ uptake, which is mediated by blue light-dependent activation of H^+-ATPase pumps. Kinoshita et al. (2011) demonstrated that overexpression of *FT* transcripts in guard cells results in increased H^+-ATPase activity and subsequent stomatal opening. Conversely, *ft* mutants show reduced light-induced stomatal opening and H^+-ATPase activity. Taken together, these results suggest that *FT* mediates stomatal aperture through regulation of H^+-ATPase activity, though the exact mechanism is unknown.

A common theme from the reports discussed here is that *FT* is a mobile signal that integrates information about light. However, work from a variety of species, including crop plants, illuminates the diversity of *FT* function in multiple developmental processes. These discoveries raise exciting mechanistic questions; do the same interacting partners mediate all *FT* functions or do different functions require unique regulatory pathways? The plant development field is poised to begin generating answers. Uncovering novel roles for *FT* enhances our understanding of molecular mechanisms underlying basic developmental processes across the angiosperms.

In Chapter 3, I expand on the usefulness of the co-expression networks presented in Chapter 2 for identification of key regulators of fruit development. The strawberry *FT* ortholog, *FvFT1*, is surprisingly clustered with genes expressed primarily in the developing receptacle fruit. Homologs of genes previously shown to mediate *FT* function in the leaf and at the SAM in *Arabidopsis* are not co-expressed with *FvFT1* in the fruit. This observation led me to hypothesize that *FvFT1* plays a role in strawberry fruit development and its function is mediated by a novel regulatory pathway. In the receptacle, *FvFT1* is not

differentially expressed in response to photoperiod but is upregulated by cool temperature. *FvFT1* is also transcriptionally upregulated by exogenous application of the phytohormones auxin and GA. Previously generated transcriptome data profiling multiple stages of flower and fruit development, transcriptional reporters, and a GFP translational fusion suggest that the FT protein may move from the developing receptacle fruit to the seed within the achene during the first two weeks post-fertilization. Data from other crop species, including mandarin and apple, indicate that *FT* expression in fruit is not unique to strawberry or to the *Rosaceae* family, thereby pointing to a conserved role for FT in fruit development (Nishikawa et al., 2007; Kotoda et al., 2010).

3.2: *Materials and Methods*

Plant growth conditions

Wild type Yellow Wonder 5AF7 and all transgenic plants in the Yellow Wonder background were grown in growth chambers set to 16 hours of daylight at 25°C and 8 hours of darkness at 20°C. Short day (accession PI551792) *F. vesca* plants were grown at either 18°C with 18 hours of daylight, 22°C with 18 hours of daylight, or 18°C with 12 hours of daylight. The short day *F. vesca* accession was chosen to test *FvFT1* transcript levels in response to photoperiod and temperature because, unlike Yellow Wonder 5AF7, it does not have a mutation in the floral repressor, *TERMINAL FLOWER 1* (*TFL1*), that removes the photoperiod requirement and stimulates continuous flowering (Koskela et al., 2012; reviewed in Chapter 1, section 2.3).

Hormone treatment of flowers

Emasculated flowers were treated with a combination of NAA and GA_3 following the method outlined by Kang et al. (2013). Stock solutions of 50 mM NAA (Sigma) and 100 mM GA_3 (Sigma) were made in ethanol and diluted with water and two drops of Tween prior to application. Final treatment concentrations were 500 µM NAA and 500 µM GA_3. Receptacles of emasculated flowers were coated with 50 µl of each hormone solution every 2d for a total of 6 or 12 days. Mock treatment of emasculated flowers followed the same protocol with a solution of water and Tween.

qRT-PCR

qRT-PCR was performed for three separate experiments. First, *FvFT1* transcript levels were compared between receptacles collected from SD *F. vesca* (accession PI551792) plants grown in short or long days at either 18°C or 22°C (described above in Plant Growth Conditions). The same plants were used throughout the experiment; growth chamber temperature and day length settings were adjusted accordingly for each condition. Plants were given two weeks to acclimate after each change in conditions. Second, *FvFT1* transcript levels were measured in WT YW 5AF7 receptacles that resulted from either hand pollination or exogenous auxin and GA_3 treatment. Third, *FvFT1* transcript levels were compared in stage 4, WT H4 receptacles and stage 4, *FvFT1 RNAi* receptacles.

For all three experiments, receptacles were collected 8-10 hours after dawn and frozen in liquid nitrogen immediately after removal of achenes. Total RNA was isolated from receptacles with achenes removed using the Cetyl trimethyl ammonium bromide (CTAB) method described by Hawkins et al. (2017), which was modified from Gasic et al. (2004).

CTAB was used due to poor RNA yields from receptacle tissue using a column-based isolation kit (Qiagen). At least three receptacles were combined into one biological replicate. Due to their small size, six to seven mock-treated receptacles were combined to form one biological replicate. Three biological replicates were analyzed for each experimental condition and three technical replicates were performed for each biological replicate. Total RNA was treated with DNaseI (ThermoScientific) to remove contaminating gDNA. For the long day versus short day and hormone treatment versus pollination experiments, cDNA was synthesized from 500 ng of total RNA in a 20 µl solution using the iScript cDNA Synthesis kit (BioRad). 1:10 diluted cDNA was used as the template in real-time PCR. SsoFast EvaGreen Supermix (BioRad) was used to set up real time reactions, which were run and analyzed on a CFX96 Real-Time System. Forty cycles of two-step real time PCR were run as follows: 95ºC, 30 s. 95ºC, 5s. 59ºC, 5 s followed by melt curve (65 to 95 ºC, 0.5 degree increments).

For the qRT-PCR experiment to compare *FvFT1* transcript levels in WT H4 versus *FvFT1 RNAi* receptacles, DNase treatment and cDNA synthesis were carried out using the SuperScript IV VILO Master Mix kit (Invitrogen). cDNA was synthesized from 1.5 µg of total RNA in a 20 µl solution. 1:10 diluted cDNA was used as the template in real-time qPCR. SsoAdvanced Universal SYBR Green Supermix (BioRad) was used to set up real time reactions, which were run and analyzed on a CFX96 Real-Time System. Forty cycles of two-step real time PCR were run as follows: 95ºC, 30 s. 95ºC, 15 s. 59ºC, 30s followed by melt curve 65 to 95ºC, 0.5 degree increments).

For all three qRT-PCR experiments, primers to target *FvFT1*, gene 21535, are based on Koskela et al. (2012) with sequences as follows: F: 5’ TGAGCTCAAACCTTCCCAAG

3’ and R: 5’ CAATCTCTTGGCCGAAAACT 3’. *FvPP2a* (gene03773), which is stably expressed across all profiled stages of receptacle development, was used as a control for normalization (Lin-Wang et al., 2014; refer to Chapter 2 Methods for primer sequences).

Vector construction

Transcriptional reporters

A 2,403 bp sequence upstream of *FvFT1* was PCR amplified from YW 5AF7 genomic DNA using Phusion polymerase (NEB, Cat. # M0530S) and primers 21535p-F: 5’ CTG TTC TTT TCG TGT GGC ATG AGC 3’ and 21535p-R: 5’ CCT GTC CCT AGG CAT ATT GAT CC 3’. The resulting fragment was cloned into PCR8/GW/TOPO using a TA cloning kit (Invitrogen, Cat. # K250020). After confirmation by sequencing, the fragment was sub-cloned into the binary vectors pMDC162 (GUS reporter) and pMDC110 (GFP reporter) (Curtis and Grossniklaus, 2003) by Gateway LR reaction (Invitrogen, Cat. #11791-100).

A 2,252 bp sequence upstream of *FvFT2* (gene04680) was amplified from YW 5AF7 genomic DNA using primers F: 5’ TCTAGATCCGGACTCTATATTCCC 3’ and R: 5’ CTAGCCCTCGCCATAGATTTAAGC 3’. A 2,021 bp sequence upstream of *FvFT3* (gene28959) was amplified from YW 5AF7 genomic DNA using primers F: 5’ GGGATTTAGCTTGTTCTGGTTGT 3’ and R: 5’ GATCTCTAGCCTTCGCCATAG 3’. The resulting fragments were individually cloned into PCR8/GW/TOPO and subcloned into pMDC162 as described above.

pFvFT1::FvFT1-GFP translational fusion

A fragment containing 2,403 bp upstream of *FvFT1* together with the *FvFT1* (gene21535) sequence without its stop codon was amplified from YW 5AF7 genomic DNA using Phusion polymerase and primers 21535p-F: 5' CTGTTCTTTTCG TGTGGCATGAGC 3' and FT-R: 5' CGATGATCTTCTCCTTCCG 3'. The resulting 4.8 kb fragment was cloned into pCR8/GW/TOPO with the TA cloning kit. After confirmation by sequencing, the fragment was sub-cloned into the binary vector pMDC110 (Curtis and Grossniklaus, 2003) by Gateway LR reaction.

CRISPR/Cas9 constructs

JH19 FvFT1 FvFT3 CRISPR: a construct simultaneously targeting *FvFT1* (gene21535) and *FvFT3* (gene28959) (all cloning and transformation by Yuexue Liu).

sgRNA sequences targeting *FvFT1* (gene21535) and *FvFT3* (gene28959) were cloned into the entry vector JH4, generated by Junhui Zhou (Zhou et al., 2018 in press). The *FvFT1* sgRNA fragment, GGTGACTTACACTTCTAAGG, was generated using primer sequences FvFT35-F: 5' GCTCGGTGACTTACACTTCTAAGG 3' and FvFT35-R: 5' AAACCCTTAGAAGTGTAAGTCACC 3'. The *FvFT3* sgRNA fragment, GATACTTAAATCTATGGCGA, was generated using primer sequences FvFT59-4F: 5' TCTATGGCGAGTTTTAGAGCTAGAAATAGCAAG 3' and FvFT59-4R: 5' TTTAAGTATCAATCACTACTCGACTCTAG 3'. Forward and reverse primers were annealed to each other to generate the *FvFT1* sgRNA fragment, which was then inserted into JH4 via BsaI digestion and T4 ligation. The *FvFT3* sgRNA fragment was next inserted with the commercial Q5 site-directed mutagenesis kit (NEB). The cassette containing the

two sgRNA sequences-was then LR recombined into JH19, a binary vector created by Junhui Zhou based on the published pMDC99 Gateway vector backbone (Curtis and Grossniklaus, 2003), containing the *Arabidopsis* UBQ promoter driving Cas9 and a 35S::3XGFP marker.

pHSE401 FvFT1 FvFT3: a construct simultaneously targeting *FvFT1* and *FvFT3* (all cloning, transformation, and sequencing by Dongdong Li).

A second CRISPR construct targeting both *FvFT1* and *FvFT3* was generated using the pHSE401 binary vector following a published protocol (Xing et al., 2014). This vector was successfully used for targeted genome editing in *F. vesca* by a former postdoc, Chunying Kang. The *FvFT1* target sequence, GACAGGGACCCCCTCGTTG, was amplified from YW 5AF7 gDNA using primers DT1_FO_FT35 F: 5' TGGACAGGGACCCCCTCGTTGGTTTTAGAGCTAGAAATAGC 3' and DT1_BSF_FT35 R: 5' ATATATGGTCTCGATTGGACAGGGACCCCCTCGTTGGTT 3'. The *FvFT3* target sequence, TTGTCGTCTCGAGAGTGAT, was amplified using primers DT2_RO_FT59 F: 5' AACATCACTCTCGAGACGACAACAATCTCTTAGTCGACTCTAC 3' and DT2_BSR_FT59 R: 5' ATTATTGGTCTCGAAACATCACTCTCGAGACGACAACAA 3'. Amplified fragments were first cloned into the entry vector pCBC-DT1T2 via BSAI digestion and T4 ligation. Fragments were subsequently sub-cloned into the binary vector pHSE401 with a Golden Gate reaction.

35S::FvFT1 overexpression construct

A 531 bp *FvFT1* CDS sequence was PCR amplified from a YW 5AF7 cDNA template using Phusion polymerase and primer sequences FTox F: 5' ATGCCTAGGGACAGG 3' and FTox R: 5' TTACGATGATCTTCTCCTTCCG 3'. The cDNA template was made from RNA isolated from stage 3 pith tissue by Chunying Kang (Kang et al., 2013). The *FvFT1* CDS fragment was cloned into pCR8/GW/TOPO and, after confirmation by sequencing, LR recombined into the Gateway binary vector pMDC32 (Curtis and Grossniklaus, 2003).

Transgenic Hawaii-4 seeds containing the RNAi construct pK7GWIWG2(II) targeting *FvFT1* (gene21535) were kindly provided by Timo Hytönen (Koskela et al., 2012).

Strawberry transformation and genotyping

Constructs were transformed into YW 5AF7 plants following the method described in Chapter 2. Twenty-eight, seven, eleven, twelve, four, nine, seven, and thirty independent transgenic lines were generated for the *pFvFT1::GUS*, *pFvFT1::GFP*, *pFvFT1::FvFT1-GFP*, *pFvFT2::GUS, pFvFT3::GUS, 35S::FvFT1, JH19 FvFT1 FvFT3 CRISPR,* and *pHSE401 FvFT1 FvFT3* CRISPR constructs, respectively. Delores Lomberk and Ceil Muller from the USDA lab at Epcot transformed the *pFvFT1::GFP, pFvFT2::GUS,* and *pFvFT3::GUS* constructs into *F. vesca*. Genomic DNA was extracted from young leaflets using Edwards Buffer (Edwards et al., 1991) and amplified with AccuStart II PCR ToughMix DNA Polymerase (Quantabio). Genotyping primer sequences for *pFvFT1::GUS*, *pFvFT2::GUS,* and *pFvFT3::GUS* plants are: Fv21535p-F: 5' CTG TTC TTT TCG TGT GGC ATG AGC 3'; Fv04680p-F:5'

TCTAGATCCGGACTCTATATTCCC 3'; Fv28959p-F: 5' GGGATTTAGCTTGTTCTGGTTGT 3' and GUS R: 5' AAT GCG AGG TAC GGT AGG AGT 3'. These primers amplify fragments containing the promoter sequences and a portion of the GUS reporter gene. Genotyping primers for *pFvFT1::GFP* and *pFvFT1::FvFT1-GFP* plants are: pFvFT1_sec6F 5' CGTTCACCTGGTCAATTATTGAC 3' and mGFP6 R: 5' ACGACGGGAACTACAAGACAC 3'. Genotyping primers for *35S::FvFT1* are: FTox genotyping F: 5' CCTCCTCGGATTCCATTGC 3' and FTox genotyping R: 5' GGTTGGGATCACTTGGGCT 3'. These primers amplify a 553 bp fragment spanning the 3' end of the *35S* promoter sequence and 5' end of the *FvFT1* CDS sequence. Genotyping primers for CRISPR transgenic plants amplify a portion of Cas9 and sequences are as follows: z-cas9 F: 5' GGGATTCTGCAGACCGTGAA 3' and z-cas9 R: 5' CCCCTCTCAGCCTTTGTGAG 3'.

Identification of CRISPR-induced edits

T_0 transgenic plants were first genotyped to confirm that they contained the CRISPR vectors *JH19 FvFT1 FvFT3* or *pHSE401 FvFT1 FvFT*. Subsequently, PCR fragments spanning the target *FT1* or *FT3* sites were sequenced to determine which individuals, if any, contained CRISPR-induced edits. The target site was amplified using primers FT35 gene21535 CRISPR genotyping Primer F: 5' GGGCAATACCAATACCCTAC 3' and FT35 gene21535 CRISPR genotyping Primer R: 5' CCAAAGAGCTGAGCTAGC 3'. The F genotyping primer was subsequently used for sequencing.

GUS staining and photography

Seeds dissected out of the ovary and longitudinally bisected receptacle fruits from T_0 generation transgenic plants were stored in 100mM sodium phosphate buffer (pH 7.4) during the harvesting process. Next, sodium phosphate buffer was removed and GUS staining solution was added (100mM sodium phosphate buffer, pH 7.4, 1 mg/ml X-glucuronic acid, 0.5 mM potassium ferricyanide, and 0.5mM potassium ferrocyanide). Tissue was vacuum infiltrated for 30 minutes and then incubated overnight at 37°C. Stained tissues were passed through an ethanol series (20%, 35%, 50%) followed by 30 minutes of incubation in FAA (50% ethanol, 5% formaldehyde, 10% acetic acid, water to volume). Tissues were stored in a final solution of 70% ethanol, observed using a stereo microscope, and photographed using a Zeiss Axiocam 105 color camera. The staining procedure was adapted based on published protocols (Takahashi et al., 2013; Sessions et al., 1999).

Fixation of GUS-stained tissue for sectioning

To determine which cells in the receptacle vasculature were GUS stained, representative receptacles at stage 4 (8-10 DPA) were chosen for embedding in wax and sectioning. After storage in 70% ethanol as described above, samples were incubated in 80% and 90% ethanol for 30 minutes each. Then, samples were incubated overnight in 95% ethanol with 0.1% Eosin-Y. On the second day, samples were incubated in 100% ethanol first for 15 minutes and then 30 minutes at room temperature with gentle shaking. Then, samples were incubated in 25% followed by 50% tert-butanol in ethanol at room temperature for 30 minutes each. Finally, samples were incubated overnight at 60°C in 100% tert-butanol. On day three, samples were incubated overnight at 60°C in a solution of 100% tert-butanol

containing 50% paraplast. On day four, the 50% paraplast solution was removed and replaced by 100% liquid paraplast and incubated for 3 hours at 60°C. Paraplast was then removed and samples were incubated for a second time in 100% liquid paraplast, this time overnight at 60°C. On day five, liquid paraplast was changed for a third time and incubated at 60°C for 1-3 hours. Subsequently, the hot paraplast containing the samples were quickly poured into petri dishes on a heated plate. Wax blocks were sectioned at 10 μm. Sections were arranged on glass slides and baked overnight. Slides were de-waxed by incubating twice for 30 minutes each in 100% tert-butanol at 60°C with frequent agitation. Slides were then incubated for 3 minutes at room temperature in 100% ethanol, for 1 minute in 100% xylene, and then for a final 1 minute in 100% xylene before mounting with Permount.

Confocal microscopy

Upon opening, flowers from both *pFvFT1::FT-GFP* transgenic plants were tagged and either hand-pollinated with camel hair paintbrushes or emasculated and treated with a hormone or mock solution (see above). At 6 DPA (stage 3), achenes were removed and seeds dissected using a stereomicroscope. The small size and transparency of seeds at 6 DPA facilitated microscopy. Seeds were mounted in a Nunc Lab-Teck chamber slide in either water or Fluoroshield mounting media containing DAPI (diluted 1:5) (Sigma) and observed with the 63X (1.2 HCXPLAPO CS) water objective of a Leica SP5 X confocal microscope. GFP fluorescent signal was excited at 488 nm with a white light laser and detected at 493-570 nm with time gating set between 0.3 and 3.8 ns. Spectral scanning was initially used to confirm GFP signal and distinguish from the high level of autofluorescence in seeds. The spectral range was set from 505 to 570 nm with detection bandwidth of 10

nm and step size of 5 nm. DAPI fluorescent signal was excited at 405 nm and detected from 415 nm to 505 nm.

Seed germination assay

FT RNAi line 1 (line 2.7.6 in Koskela et al., 2012) and WT H4 plants were grown in growth chambers with 16 hours of daylight at 25°C and 8 hours of darkness at 20°C. Seeds were collected, dried overnight, and subjected to 3 weeks of dry storage in the dark at either 4°C or room temperature. After storage, seeds were surface sterilized in a 20% bleach solution for 8 minutes followed by 6 washes with sterile water. 100-200 seeds per replicate for each condition were plated on media containing 0.22% MS with Gamborg's vitamins, 2% sucrose, and 7 g/L Phytoblend agar. At least three biological replicate seed batches were included for each condition and genotype. Each replicate contained seeds combined from two to three individual plants. Plates were kept in growth chambers with 16 hours of daylight at 25°C and 8 hours of darkness at 20°C. Percent germination was calculated after 10 days.

Supplemental Materials (Files available in Google Drive at this address: https://drive.google.com/drive/folders/13LL1JBgt9IIONyq3rhDgfrVGGw1OSaWk?usp=sharing)

Dataset S3.1: Descriptions for 1,224 genes in cluster 13 from standard, hand-dissected network

Dataset S3.2: GO enrichment analysis for cluster 13

Dataset S3.3: Transcription factors in cluster 13

3.3: *Results*

3.3.1: Network analysis highlights genes co-expressed in the developing receptacle

Cluster 13 from the standard, hand dissected co-expression network analysis presented in Chapter 2 (Fig 2.1B) contains 1,224 genes (Supplemental Dataset S3.1) and is associated with all profiled stages of receptacle development, but is most strongly correlated with the pith after fertilization (Fig 3.1A). Two of the biological process GO terms enriched in the cluster are regulation of hormone levels (GO:0010817), which includes two *PIN* auxin transporters (genes 01267 and 16792) and oxidation reduction (GO:0055114), including a GA20 oxidase (gene19438), reiterating the importance of hormone signaling in fleshy fruit development (Supplemental Dataset S3.2). Like the floral receptacle-associated clusters from the consensus_LCM network presented in chapter 2 (Fig 2.5), examination of the transcription factors in cluster 13 revealed several known meristem regulators in the GRAS, TALE, and WOX families, reflecting the meristem-like attributes of the young receptacle. The MADS box transcription factor family, members of which are known to control ripening in tomato (Vrebalov et al., 2002; Fujisawa et al., 2014), is also represented in cluster 13 (genes 26119, 06301, 07365, 04228, 20134), suggesting a role for MADS box genes in early stage strawberry fruit development (Supplemental Dataset S3.3).

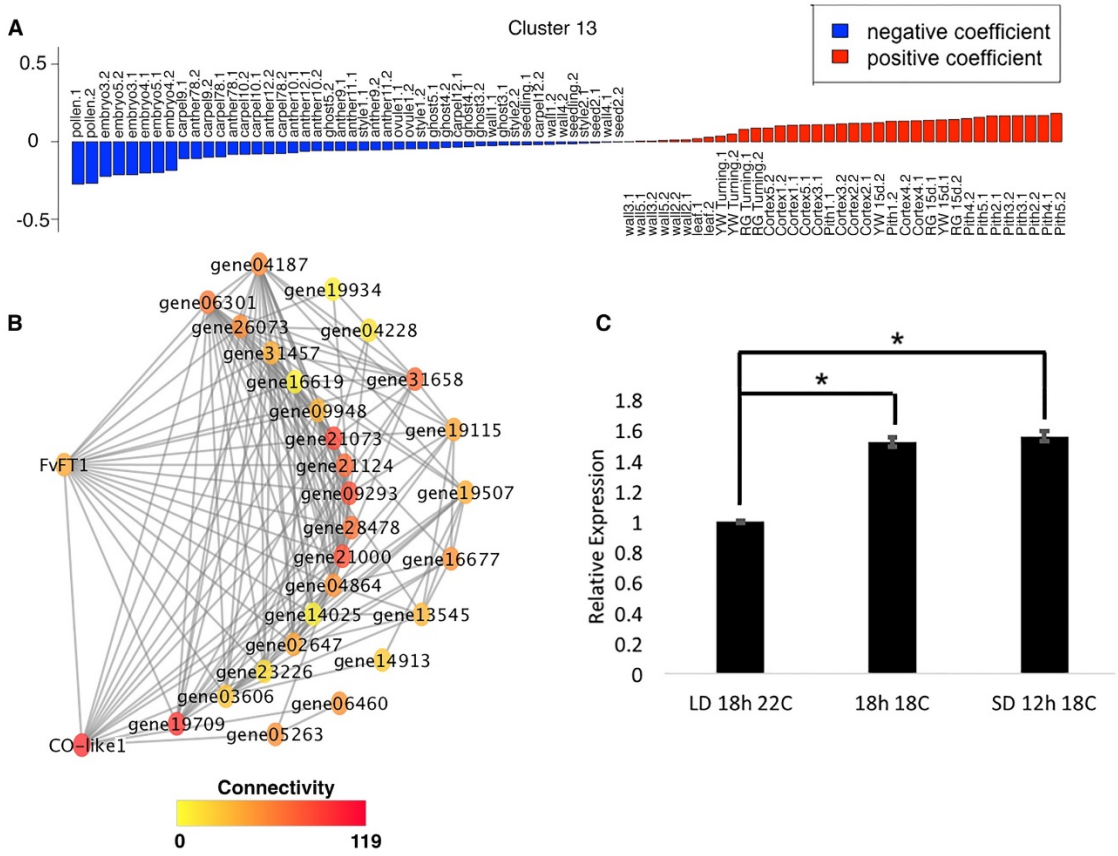


Figure 3.1 *FvFT1* is expressed in the developing receptacle fruit

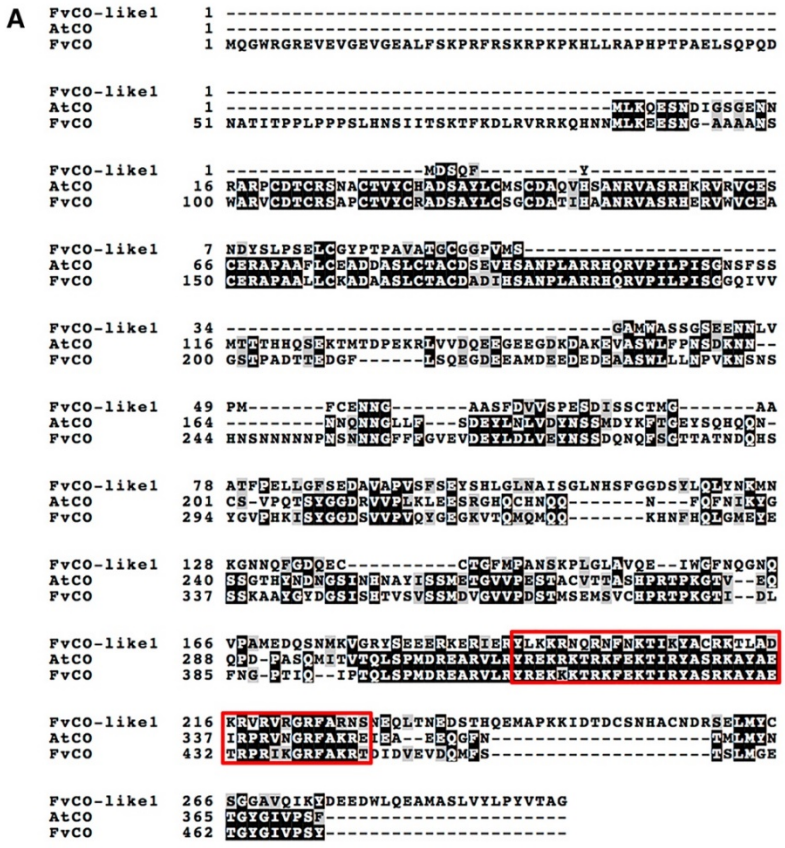
(A) Standard_hand dissected cluster 13 eigengene expression bar graph
(B) Network showing connections between *FvFT1* and transcription factors in cluster 13. Edge cutoff is 0.95. Each colored circle (node) represents one gene. Darker red node color indicates greater connectivity within the network. Image generated with Cytoscape v 3.4.0.
(C) qRT-PCR data showing *FvFT1* expression in 12 DPA receptacles in long day (LD) versus short day (SD) and warm (22°C) versus cool (18°C) temperature conditions. *FvFT1* expression is higher at 18°C than at 22°C, but is not influenced by day length. Bars represent the average expression +/- standard error for three biological replicates. Transcript abundance was calculated relative to *FvePP2A*. Statistical significance was calculated with a one-way ANOVA followed by a Tukey test. * $p < 0.05$.

3.3.2: FvFT1 is differentially expressed in the receptacle in response to temperature, but not photoperiod

Due to its canonical role in flowering time regulation, a surprising member of cluster 13 is the strawberry ortholog of *FT*, *FvFT1* (gene21535; Koskela et al., 2012). Using differential expression analyses, I also identified *FvFT1* in a list of genes more highly expressed in the

young receptacle versus all other reproductive and vegetative tissues profiled with our lab's transcriptome data (differential expression analyses presented in Chapter 4). To investigate whether *FvFT1* is under photoperiodic control in the receptacle as it is in the leaf, I performed qRT-PCR to compare *FvFT1* transcript levels in cool, short days; warm, long days; and cool, long days (Fig 3.1C). *FvFT1* transcript levels were not significantly different in response to 12 hours of daylight versus 18 hours of daylight when plants were grown at 18°C. However, *FvFT1* was more highly expressed in receptacles at 18°C versus 22°C when plants were grown with 18 hours of daylight. *FvFT1* expression, at least in the tested conditions, is sensitive to temperature, but not photoperiod, in the receptacle fruit. This result is in contrast to previous reports of *FT* expression in *Arabidopsis*, where photoperiod plays a significant role in regulating *FT* expression in leaves via *CONSTANS (CO*) (Yanovsky and Kay, 2002; An et al., 2004).

The *F. vesca* ortholog of *CO*, *FvCO* (gene04172; Kurokura et al., 2017) is not expressed in the receptacle but is predictably included in the leaf-associated cluster 26 of the standard, hand dissected network. However, *FvCO-like1 (*gene30045), a putative zinc finger transcription factor with homology to *FvCO,* is strongly co-expressed with *FvFT1* in cluster 13 (edge score = 0.87; Fig 3.1B). Like *FvCO*, *FvCO-like1* has a CCT domain (Fig 3.2A), the motif required for *CO* binding to the *FT* promoter in *Arabidopsis* (Tiwari et al., 2010), thereby making it a candidate transcriptional regulator of *FvFT1* in the receptacle (Fig 3.2B). Further, the strawberry homolog of *FLOWERING LOCUS D*, *FvFD* (gene14556), is not expressed in the developing receptacle fruit or seed (Fig 3.2C), suggesting that FvFT1 function in fruit, unlike in promotion of flowering, is likely independent of FvFD.


A
FvCO-like1
AtCO
FvCO
B
FvCO-like 1
RPKM
26.02
13.0
0
Flower
Anther
Floral stages 7-8
Floral stage 9
Floral stage 10
Floral stage 11
Floral stage 12
Carpel
Floral stages 7-8
Floral stage 9
Floral stages 10-11
Floral stage 12
Pollen
Fruit
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5
Cortex (outer)
Pith (inner)
Style
Carpel wall
Ghost
Embryo
Vegetative
Leaf (young)
Seedling (10 days)
Early Floral Stages (LCM)
Flower st 1-4
Perianth st 5-6
Anther st 5-6
Receptacle st 6-7
Microspore st 10

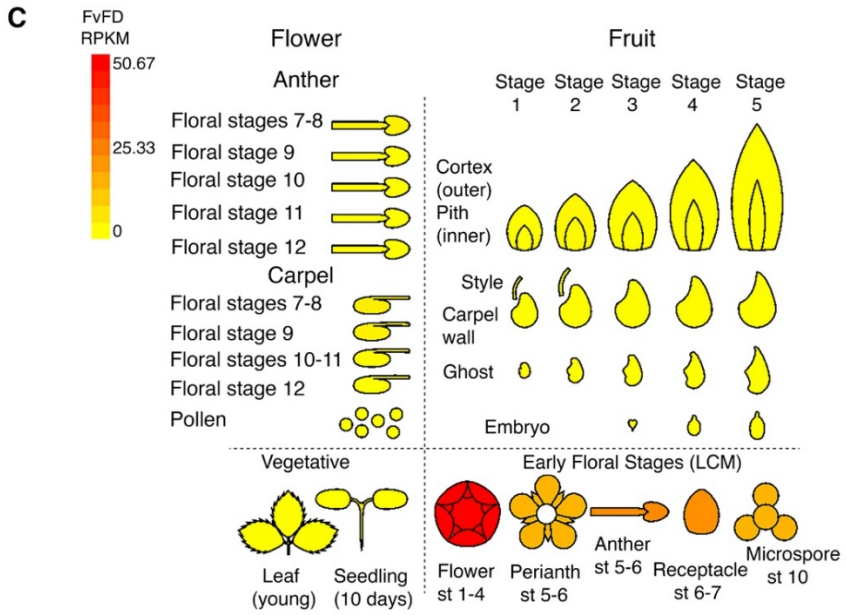


Figure 3.2 *FvFT1* may function in a novel pathway in fruit

(A) Alignment of amino acid sequences of *Arabidopsis CO* (At5G15840), *F. vesca CO* (gene04172), and *FvCO-like1* (gene30045). Red box denotes the CCT domain. Sequences were aligned using Clustal Omega and Boxshade.
(B) eFP (Hawkins et al., 2017) illustrating *FvCO-like 1* (gene30045) expression across multiple stages of fruit and flower development profiled with RNA-Seq.
(C) eFP illustrating *FvFD* (gene14556) expression.

3.3.3: FvFT1 regulatory sequence is active in the vasculature connecting fruit and seed

Previously generated RNA-Seq data (Kang et al., 2013) indicates that *FvFT1* is upregulated post-fertilization at stage 3 and is most highly expressed in both the pith and cortex at stage 4 (8-10 DPA) (Fig 3.3A). In agreement with the transcriptome, a *GUS* transcriptional reporter, *FvFT1promoter::GUS*, confirms that the isolated *FvFT1* regulatory sequence (see Methods) is active in the stage 4 receptacle (Fig. 3.3A) and suggests that *FT* mRNA is transcribed in, and not transported to, the fruit tissue. Further, the *promoter::GUS* reporter revealed strong *FvFT1* transcriptional activation in the vascular bundles of the receptacle (Fig 3.3 B, C). This vasculature connects the receptacle to the overlying achenes (Fig 1.5A)

and GUS staining is visible in the vascular strand extending into the seed, though in agreement with the transcriptome, there is no staining in the seed itself (Fig 3.3 D, E). *FvFT1* expression in the receptacle vascular bundles raises the possibility that FvFT1 is mobile in the strawberry fruit similar to its non-cell autonomous activity as the florigen. The seed is a candidate destination for *FvFT1* due to its proximity to the receptacle and its vascular connection in which *pFvFT1::GUS* signal is detected (Fig 3.3D).

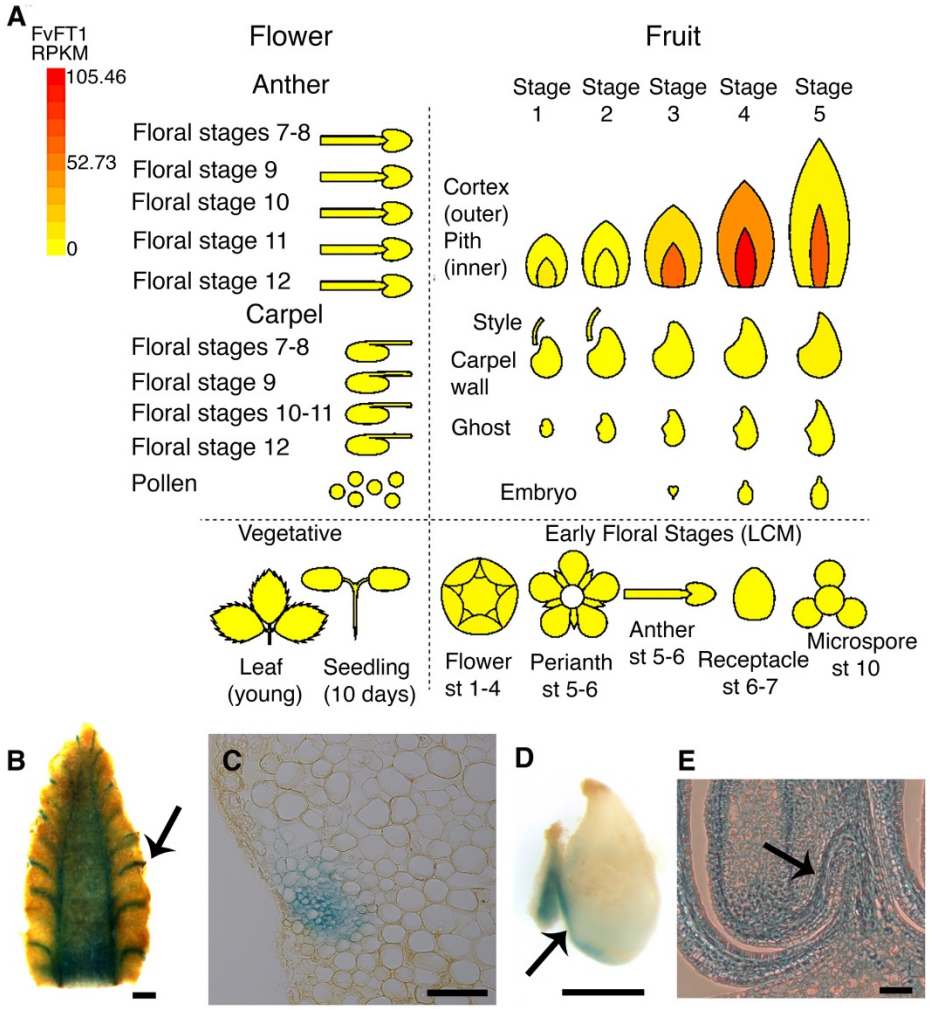


Figure 3.3 GUS transcriptional reporter indicates *FvFT1* expression in vasculature connecting the seed and receptacle

(A) eFP (Hawkins et al., 2017) illustrating *FvFT1* expression across multiple stages of fruit and flower development profiled with RNA-Seq.
(B) Stage 4 *pFvFT1::GUS* receptacle shows GUS staining in the pith and vascular connections between the receptacle and overlying achenes (arrow). Scale bar is 500 µm.
(C) Sections of stage 4 *pFvFT1::GUS* receptacle showing GUS staining localized to the vascular bundle. Scale bar is 100 µm.
(D) Stage 4 *pFvFT1::GUS* seed shows GUS staining in the vascular connection with the underlying receptacle (arrow). Scale bar is 500 µm.
(E) Section of an ovule from a stage 12, open flower clearly shows the vascular strand connecting the seed and receptacle (arrow), which is stained with GUS in (D). Scale bar is 50 µm.

3.3.4: FvFT1 peptide is nuclear localized in the seed

Given the expression pattern of *FvFT1* in the vascular strands connecting the receptacle to the overlaying seeds (Fig 3.3 B-E), I asked if FvFT1 protein is present in seeds. To test native protein localization, I generated a GFP translational reporter (*pFvFT1::FvFT1-GFP*) driven by the same native *FvFT1* regulatory sequence used for the *promoter::GUS* reporter. Previously, Corbesier et al. (2007) showed that an FT-GFP fusion protein expressed specifically in the phloem crossed the graft boundary in *Arabidopsis* seedlings, indicating that tagging with the relatively large GFP protein does not inhibit movement of the small FT peptide. Confocal examination of 6 DPA seeds revealed that a GFP signal is visible at the edges of the seed, in a location near the vascular strand (Figure 3.4 A, B, F). Co-localization of the GFP signal with DAPI indicates that FvFT1-GFP is nuclear localized in the seed (Figure 3.4 C, D), a result consistent with FT's known role as a transcriptional co-factor (Abe et al., 2005).

I sought to confirm that the *FvFT1-GFP* signal visible in seeds reflects localization of the fusion protein and not cleaved GFP. Due to its small size, GFP moves freely between plant cells and GFP in the phloem can be trafficked over long distances to sink tissues such as

flowers and fruit (Imlau et al., 1999). I asked if differences in GFP signal localization could distinguish FvFT1-GFP fusion protein from free GFP. To identify the pattern of fluorescent signal from free GFP, I examined 6 DPA seeds from a *pFT::GFP* transcriptional reporter. Diffuse fluorescent signal is visible in seeds (Figure 3.4F), indicating that the nuclear localized signal from the *pFvFT1::FvFT1-GFP* seeds (Fig 3.4 A-D) is likely from the FvFT1-GFP fusion protein and not simply cleaved GFP unloaded from the phloem.

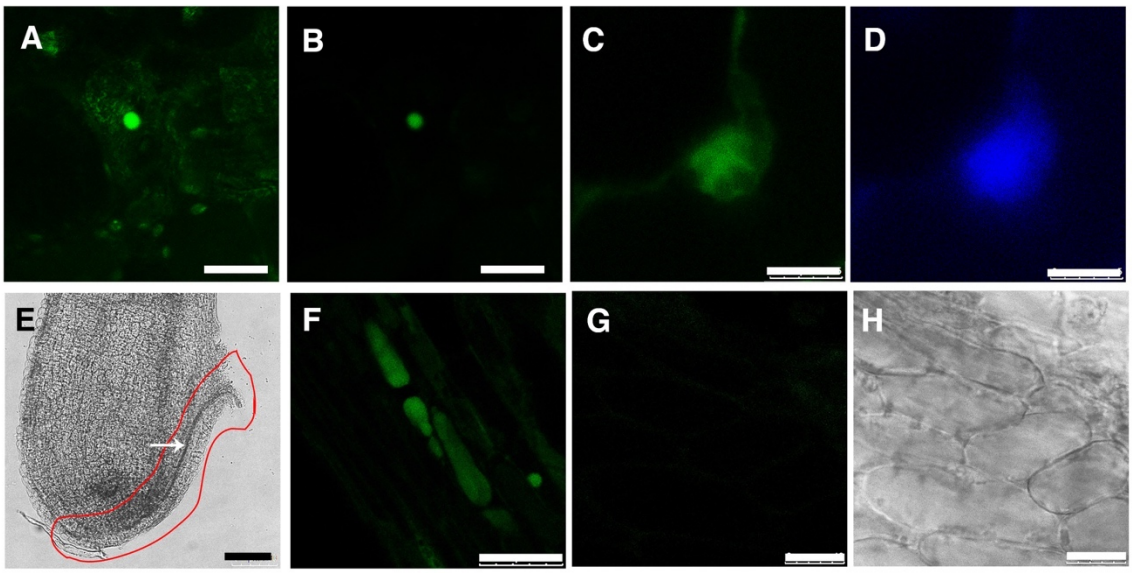


Figure 3.4 FvFT1-GFP translational fusion is nuclear localized in the seed

(A-E) GFP signal from *pFvFT1::FvFT1-GFP* developing seeds at 6 DPA
(A) GFP + autofluorescence showing subcellular fluorescent signal
(B) Same image as A, but without autofluorescence to indicate GFP signal. Scale bar is 7.5 µm.
(C) Co-localization of GFP and DAPI **(D)** confirm FvFT1-GFP in the nucleus. Scale is 5 µm for **(C)** and **(D)**.
(E) BF photo of seed with red outline to indicate location of GFP signal. Vascular strand that connects seed to receptacle is indicated with white arrow. Scale bar is 100 µm.
(F) Diffuse GFP signal from *pFT::GFP* is visible throughout the chalazal end of 6 DPA seeds. Scale bar is 25 µm.
(G) No GFP signal is visible in *pFvFT1::FvFT1-GFP* parthenocarpic seeds from mock-treated, emasculated flowers. Scale bar is 10 µm.
(H) BF photo of **(G)**. Scale bar is 10 µm.

3.3.5: Auxin and/or GA_3 positively regulate FvFT1 in the receptacle

RNA-Seq profiling of the receptacle pre- and post-fertilization indicates that *FvFT1* expression is induced post-fertilization at stage 3 (Fig 3.3A). Based on this observation, I hypothesized that *FvFT1* is regulated by fertilization signals in the receptacle. To test this idea, I performed a qRT-PCR experiment to detect *FvFT1* transcript levels in receptacles that developed from either hand-pollination of flowers or exogenous hormone treatment after floral emasculation. Hereafter, I refer to fruit and seeds developed from hand pollinated flowers as 'fertilized.' Fruit and seeds enlarged as a result of exogenous hormone treatment of emasculated flowers are called 'hormone-induced parthenocarpic.' Fruit and seeds from emasculated flowers treated with water are called 'mock-treated.'

Previously, our lab showed that auxin and GA_3 application are sufficient to induce parthenocarpic receptacle development (Kang et al., 2013; Fig 1.4). I tested two post-fertilization time points: 6 DPA (stage 3) and 12 DPA (stage 5), reasoning that if *FvFT1* is transcriptionally regulated by an unknown, fertilization-induced signal, one would expect to see decreased or abolished *FvFT1* expression in hormone-induced parthenocarpic receptacle fruit. However, *FvFT1* transcript levels are not statistically significantly different between fertilized and hormone-induced parthenocarpic fruit at either 6 or 12 DPA (one-way ANOVA $p = 0.327$; Fig 3.5A) and *FvFT1* is even higher in hormone-induced parthenocarpic fruit at 12 DPA as compared to pollinated fruit. Taken together, this result suggests that *FvFT1* is responsive to either auxin, GA_3, or both. This experiment alone does not indicate whether the response is direct or indirect, though GA was previously shown to indirectly regulate *FT* transcription via mediation by *CO* in the *Arabidopsis* leaf (Xu et al., 2016; Wang et al., 2016).

Although not statistically significantly different, the 12 DPA mock sample shows higher *FvFT1* expression than the 12 DPA pollinated sample, which is unexpected if *FvFT1* expression is upregulated by fertilization-induced signals. I think that collection and testing of additional 12 DPA mock samples would be beneficial to clarify the results.

Strong GUS staining in the vascular strands of stage 4, hormone-induced parthenocarpic, *pFvFT1::GUS* receptacle fruit also supports the responsiveness of *FvFT1* transcription to auxin and/or GA_3 (Fig 3.5B). However, stage 4, hormone-induced parthenocarpic seeds do not show GUS staining in the vascular strand that connects to the receptacle (Fig 3.5C), even though such staining was visible in fertilized seeds (Fig 3.3D). Previously, our lab quantified the enlargement of parthenocarpic seeds dissected from achenes of hormone-induced parthenocarpic fruit (Kang et al., 2013; Fig 1.4C), demonstrating that exogenously applied auxin and GA_3 can influence the ovule. I also observed enlargement of parthenocarpic seeds and achenes in my own hormone application experiments. Taken together, the fertilization-induced development of the seed is necessary for *FvFT1* expression in the portion of the vascular strand that extends into the seed.

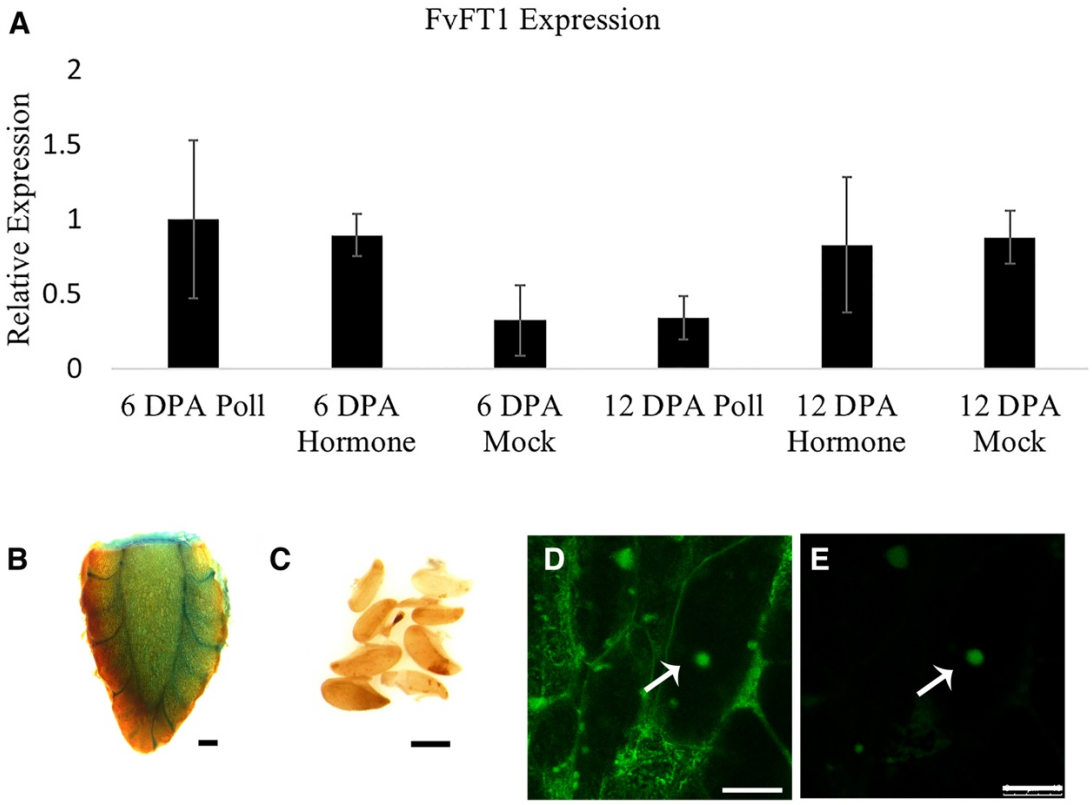


Figure 3.5 Exogenous application of auxin and GA3 induces *FvFT1* expression in the receptacle

(A) qRT-PCR data showing *FvFT1* expression levels in 6 and 12 DPA receptacles developed after either pollination ('Poll'), exogenous auxin and GA_3 treatment of emasculated flowers ('Hormone'), or mock treatment of emasculated flowers ('Mock') (see Methods for details). Bars represent the average expression +/- standard error relative to 6 DPA pollinated (set to 1.0) for three biological replicates. Transcript abundance was calculated relative to *FvePP2a*. Statistical significance was calculated with a one-way ANOVA. p value = 0.327.
(B) GUS staining of stage 4, hormone-induced parthenocarpic, *pFvFT1::GUS* receptacles is similar to staining of stage 4, fertilized receptacles (Fig 2B). However, hormone-induced parthenocarpic seeds **(C)** do not show any GUS staining (compare to fertilized seeds in Fig 3.3D). Scale bars are 500 µm in **(B)** and **(C)**.
(D) GFP + autofluorescence signal (arrow) from parthenocarpic seed dissected out of achene from 6 DPA, hormone-induced parthenocarpic *pFvFT1::FvFT1-GFP* fruit (see Methods for details).
(E) Same as **(D)**, but without autofluorescence to indicate GFP signal (arrow). Scale bars are 10 µm in **(D)** and **(E)**.

3.3.6: Does FvFT1-GFP fusion protein move from receptacle to seed?

The expression of *pFvFT1::GUS* in the vascular strand raises the possibility of *FvFT1* mobility in the fruit. I asked if FvFT1-GFP in the seed originates from the receptacle or from seed-transcribed mRNA in the vascular strand. Taking advantage of the fact that hormone-induced parthenocarpic seeds do not transcribe *FvFT1* in the vascular strand (Fig 3.5C), I tested if the FvFT1 peptide is mobile in the fruit. Hormone-induced parthenocarpic, *pFvFT1::FvFT1-GFP* seeds (Figure 3.5 D, E) show the same nuclear localized GFP signal as fertilized seeds (Fig 3.4), indicating that the FvFT1 peptide is present in a tissue where *FvFT1* is not transcribed. This result suggests that FvFT1 moves into the seed, likely via the phloem from the neighboring receptacle.

3.3.7: Plants expressing FvFT1 RNAi produce seeds with increased germination efficiency in the absence of cold treatment

Given its expression in the receptacle and peptide localization in the seed, I hypothesized that *FvFT1* is a regulator of fruit or seed development. I used *FvFT1* RNAi lines generated by our collaborator, Timo Hytönen, to test for mutant phenotypes in fruit or seeds resulting from loss of *FvFT1* function. Plants stably expressing RNAi against *FvFT1* were previously used to test *FvFT1* function in flowering time regulation and were shown to have a late flowering phenotype (Koskela et al., 2012). *FvFT1* transcript levels tested in one published, independent line of these RNAi plants are significantly reduced in the stage 4 receptacle (Fig 3.6A), but ripe fruits appear wild type (Fig 3.6 B). Since FvFT1-GFP fusion protein is detectable in the seed, I tested the potential function of FvFT1 in seed germination. Germination is a highly regulated process and is generally initiated following

a period of dormancy. Most species require environmental triggers, such as cold temperatures, to break dormancy and initiate germination (Penfield, 2017).

I compared the germination efficiency of *FvFT1* RNAi seeds to WT H4 after 3 weeks of dry storage at either 4°C or room temperature (Fig 3.6C). After 10 days, a greater percentage of WT H4 seeds germinated after experiencing cold treatment compared to those that did not experience cold. Conversely, *FvFT1* RNAi seeds germinated at a similar frequency regardless of cold treatment and at a higher frequency than WT H4 seeds that did not experience cold. *FvFT1* RNAi increases germination efficiency in the absence of cold treatment, suggesting that *FvFT1* functions as a negative regulator of germination.

Upon further examination of *FvFT1* RNAi seeds, I found that embryos dissected from achenes of ripe fruit (25 DPA) are green in color compared to the opaque, white WT embryos of the same developmental stage (Fig 3.6 D-E). This observation led me to hypothesize that loss of *FvFT1* function results in precocious chloroplast maturation and chlorophyll production. However, I found that control plants with RNAi constructs targeting different genes also produce embryos that are green at maturity, suggesting that the phenotype is due to the RNAi vector itself and not loss of *FvFT1* function (further discussed in Appendix 1; Figs. A2 and A3). Future experiments will test whether or not the germination efficiency phenotype is also an artifact of transformation with an RNAi construct.

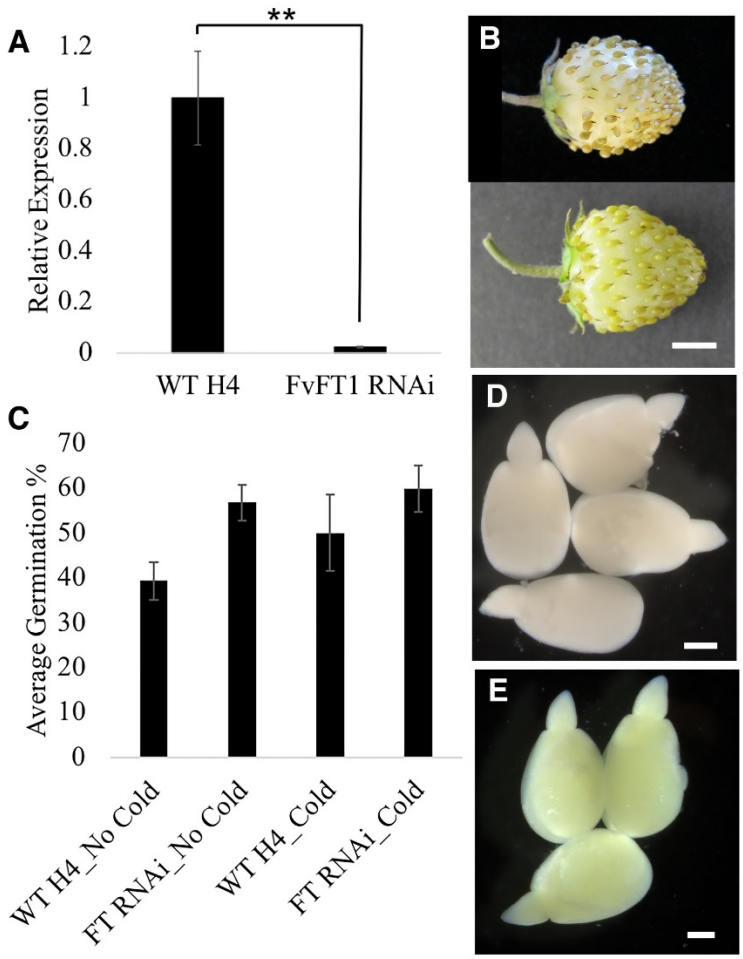


Figure 3.6 *FvFT1* RNAi seeds germinate more efficiently than WT seeds in absence of cold treatment

(A) qRT-PCR data showing *FvFT1* transcript levels in stage 4 receptacles from *FvFT1 RNAi* line 1 and WT H4 plants. Bars represent the average expression +/- standard error relative to WT H4 (set to 1.0) for three biological replicates. Transcript abundance was calculated relative to *FvePP2a*. Statistical significance was calculated with a Student's t-test. ** $p < 0.01$.
(B) Ripe fruit. Top: WT and Bottom: *FvFT1 RNAi*. Scale bar is 500 mm.
(C) Germination of *FvFT1 RNAi* and WT H4 seeds after 3 weeks of dry storage at 4°C (cold) or room temperature (no cold). Data points shown are the average germination percentage +/- standard error of at least three biological replicate seed batches per genotype and condition.
(D) WT H4 embryos dissected from achenes of ripe fruit (25 DPA) are white. Scale bar is 200 µm.
(E) Representative *FvFT1 RNAi* embryos dissected from achenes of ripe fruit (25 DPA) are green. Scale bar is 200 µm.

3.3.8: FvFT paralogs may have functional overlap in fruit

The lack of clear developmental phenotypes in *FvFT1* RNAi fruit and seeds led me to hypothesize that paralogous *FT* genes are functionally redundant. *F. vesca* has two additional genes, *FvFT2* and *FvFT3*, with strong sequence homology to *FT* (Fig 3.7). Both genes have expression patterns similar to *FvFT1* (Fig 3.8), though *FvFT1* has the highest level of expression and is specific to the receptacle fruit. *FvFT2* and *FvFT3* have lower expression levels and are not as receptacle-specific (Fig 3.3 A; Fig 3.8 A, B).

To examine the transcriptional activity of *FvFT2* and *FvFT3* during fruit development in greater detail, I generated *promoter::GUS* reporters. Plants stably expressing *pFvFT2::GUS* indicate that *FvFT2* (gene04680), like *FvFT1*, is expressed in the receptacle and the vascular strand connecting the receptacle to the seed (Fig 3.8C), though the staining pattern is not as specific in the receptacle as that of *FvFT1* (Fig 3.3B). Fruit from plants expressing the *pFvFT3::GUS* reporter showed no positive GUS staining in the receptacle at either stage 1 or stage 4, which is contrary to the transcriptome data (Fig 3.8B). This could be due to elimination of important regulatory elements in the isolated promoter sequence.

To investigate *FT* function in the receptacle and address possible functional overlap of *FvFT* homologs, I collaborated with several lab members to simultaneously target *FvFT1* and *FvFT3* with CRISPR/Cas9 using a single vector approach. Junhui Zhou constructed and tested multiple vectors to pioneer CRISPR genome editing in *F. vesca*. Using vectors from Junhui, Yuexue Liu generated the first constructs (JH19 CRISPR; see Methods) to simultaneously target *FvFT1* and *FvFT3*. Of the four T_0 plants confirmed to contain the

Cas9 transgene, none had CRISPR-induced edits in the target genes. Since CRISPR/Cas9 editing remains active in subsequent generations, a future plan is to generate T_1 plants, check for CRISPR-induced mutations, and examine fruit for phenotypes.

Since genome editing methods are still under development for strawberry, Dongdong Li sought to maximize our chances of generating *FvFT* knockouts by creating a second CRISPR construct simultaneously targeting *FvFT1* and *FvFT3* using the previously published pHSE401 vector (Xing et al., 2014; details in Methods). Four plants had CRISPR-induced edits of *FvFT1,* but all four are polyploid and have not produced fruit, thereby making generation of T1 plants unlikely.

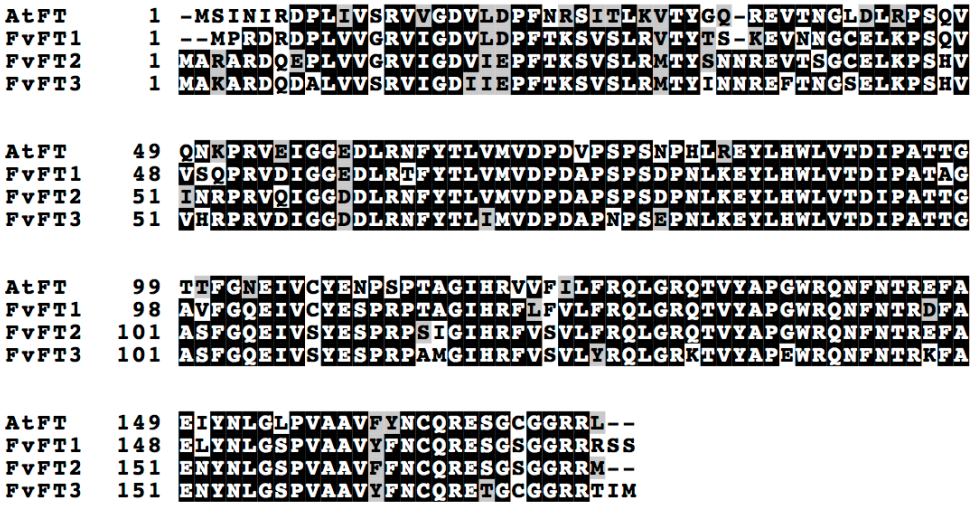


Figure 3.7 *F. vesca* genome includes three *FT-like* genes

Alignment of amino acid sequences of three *F. vesca FT* homologs and the *Arabidopsis FT* (At1G65480). Sequences were aligned using Clustal Omega and Boxshade.

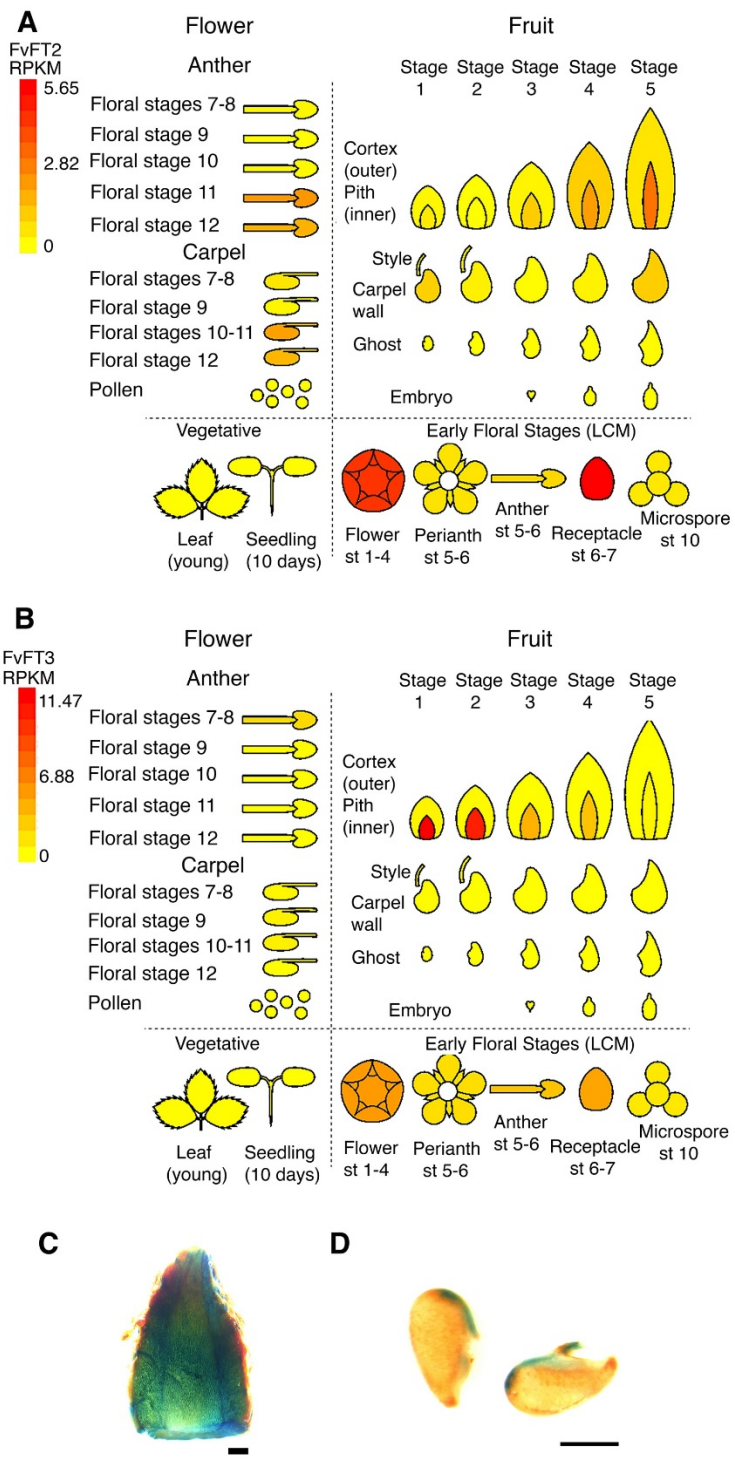


Figure 3.8 *FvFT2* and *FvFT3* are expressed in the developing receptacle

(A) eFP demonstrating *FvFT2* and *FvFT3* **(B)** expression across multiple stages of fruit and flower development profiled with RNA-Seq (Hawkins et al., 2017).
(C) GUS staining pattern of *pFvFT2::GUS* in receptacle and seed **(D)**. Scale bars are 500 µm in **(C)** and 200 µm in **(D)**.

3.3.9: Generation of plants constitutively expressing FvFT1 was unsuccessful

Based on the transcriptome data indicating increased *FvFT1* expression immediately post-fertilization (Fig 3.3A), I originally hypothesized that *FvFT1* may be a fertilization-induced signal involved in triggering fruit set. To test this idea, I generated an overexpression construct with the *35S* constitutive promoter driving the *FvFT1* CDS sequence. My plan was to emasculate flowers from the overexpression plants and test for parthenocarpic fruit development. I reasoned that fruit development resulting from constitutive *FvFT1* expression, even in the absence of fertilization, would suggest a role for *FvFT1* in triggering fruit set. I transformed the *35S::FvFT1* construct into *F. vesca* and generated multiple independent lines, but had difficulty generating reliable genotyping results. Further, none of the plants appeared to have an early flowering phenotype, which I expected based on previous reports from multiple species overexpressing *FT* (Pin and Nilsson, 2012). I emasculated flowers from the transgenic plants, but did not see any evidence of parthenocarpic fruit development. However, since I am unsure of whether or not the *35S::FvFT1* plants are truly transgenic, I cannot draw a solid conclusion. qRT-PCR to test transcript levels of *FvFT1* in the fruit of the *35S::FvFT1* plants compared to WT would confirm whether or not *FvFT1* is actually being overexpressed. If it is not, the *35S::FvFT1* construct will have to be re-transformed into *F. vesca*.

3.4: *Discussion*

A co-expression network based on previously generated RNA-Seq data revealed the surprising, fruit-associated expression pattern of *FvFT1*, which has been shown to regulate floral initiation in *F. vesca* (Koskela et al., 2012). *FvFT1* is a homolog of *Arabidopsis FT*,

which is expressed in leaves and moves to the shoot apical meristem to initiate flowering in response to environmental and endogenous signals (Abe et al., 2005; Corbesier and Coupland, 2005; Corbesier et al., 2007). Coupled with transcriptome data, a *promoter::GUS* reporter showed that *FvFT1* is transcriptionally upregulated in the developing receptacle fruit, particularly in the vasculature connecting the receptacle to the overlying seed. *FvFT1* is differentially expressed in the receptacle in response to temperature, but not photoperiod, and transcriptional upregulation in hormone-induced parthenocarpic fruit is similar to that of fertilized fruit. Nuclear-localized FvFT1-GFP fusion protein is detectable in the seed, even in the absence of *FvFT1* transcriptional activation, suggesting FvFT1 is mobile in the fruit. While WT seeds require cold treatment to germinate at a higher frequency, *FvFT1* RNAi seeds germinate with the same efficiency even without cold treatment. However, *FvFT1*'s influence on germination and embryo greening is still in question due to effects that may be attributable to the RNAi vector itself and independent of loss of *FvFT1* function. The embryo greening phenotype is discussed further in Appendix 1.

3.4.1: FvFT1 likely functions in a novel pathway in fruit

FT mRNA transcribed in the leaf is not detectable in the phloem (Lin et al., 2006; Corbesier et al., 2007). This observation, together with my *p::GUS* transcriptional reporter data, suggests that *FvFT1* mRNA detected in the *F. vesca* receptacle via RNA-Seq is likely transcribed in, and not transported to, the receptacle (Fig 3.3A). Unlike its known photoperiodic regulation as the florigen (Koskela et al., 2012), my result showed that *FvFT1* transcript levels are not affected by photoperiod in the receptacle. Instead, *FvFT1* likely integrates information about temperature (Fig 3.1C). Lack of *FvCO* expression in

the receptacle reinforces the idea that *FvFT1* is not transcriptionally regulated by the same pathway in fruit as it is in leaf. *FvCO-like1*, identified by its strong co-expression with *FvFT1* in cluster 13 (Fig 3.1B), contains the CCT domain known to bind to the *FT* promoter (Tiwari et al., 2010) and I therefore hypothesize that it is a candidate transcriptional regulator of *FvFT1*. A function for an *FvCO-like* gene in fruit is plausible given that, of the 16 *CO-like* genes in *Arabidopsis*, several have roles unrelated to floral promotion (Robson et al., 2001; Griffiths et al., 2003). The co-expression network is a useful tool to generate hypotheses, though direct experimental evidence is still required to confirm transcriptional regulation of *FvFT1* by *FvCO-like1*. One question of interest is whether *FvCO-like1*, like *FvFT1*, is regulated by temperature in the receptacle.

3.4.2: FvFT1 is positively regulated by auxin and/or GA in the receptacle

Fruit set in all angiosperms is triggered by positive, fertilization-induced signals, though Auxin and GA are the only identified signals to date (Nitsch, 1970). Increased mRNA levels in the receptacle at stage 3 suggest that *FvFT1* is upregulated by fertilization (Fig 3.3A; Kang et al., 2013). *FvFT1* transcripts are detectable in the receptacle upon exogenous hormone application to emasculated flowers (Fig 3.5 A-C), suggesting either direct or indirect regulation by auxin and/or GA. Although the experiment reported here cannot differentiate between the effects of auxin and GA, recent work in *Arabidopsis* indicates that DELLAs, proteins that mediate GA-induced transcriptional regulation, physically interact with CO and CO mediates positive regulation of *FT* by GA (Xu et al., 2016; Wang et al., 2016). This raises the possibility that an unknown factor may also mediate GA regulation of *FvFT1* in the receptacle. If so, FvFT1 is a potential integrator of endogenous hormone signals in the developing fruit.

3.4.3: Evidence for FvFT1 mobility in fruit

As in the developing seed, nuclear-localized FvFT1-GFP signal is also visible in parthenocarpic seeds of *pFvFT1::FvFT1-GFP* parthenocarpic fruit, indicating presence of the FvFT1 peptide in a tissue where *FvFT1* is not transcribed (Figure 3.5 C-E). Coupled with its known mobility in floral induction, this result suggests that FvFT1 moves into the seed, likely from the neighboring receptacle via the vascular strand. An experiment in which WT seeds are grafted onto a receptacle expressing the FvFT1-GFP fusion driven by a receptacle-specific promoter would provide the strongest evidence of FvFT1 protein movement from the receptacle into the seed, but to my knowledge, this type of graft has never been successfully conducted. An alternative follow-up experiment could be to block phloem unloading into the seed and determine if FvFT1-GFP fusion protein is still detectable. A previously demonstrated, inducible method to block protein movement between cells in *Arabidopsis* could be adapted for this purpose (Wu and Gallagher, 2014).

Chen et al. (2014) demonstrated FT-GFP translational fusion signal accumulation at the base of the Arabidopsis seed when driven by the *SUCROSE2* (*SUC2*) phloem-specific promoter. This result suggests that the FT peptide in the phloem may move into the seed, although this experiment did not confirm native FT localization in the seed itself. My work shows that an FvFT1-GFP translational fusion driven by the native *FvFT1* promoter, which is a weaker promoter than *pSUC2* (Corbesier et al., 2007), is nuclear localized in the stage 3 seed, though GFP signal in the seed was only readily visible near the vascular strand connecting to the receptacle (Fig 3.4). GFP signal was visible near the surface of the seed, suggesting localization in the seed coat and not the underlying endosperm. However, this observation could also be due to depth constraints of confocal microscopy. Examination of

seeds with a two photon microscope would allow visualization of deeper tissues and stronger evidence to confirm signal from the seed coat. The fact that the FT-GFP signal is present in nuclei indicates that it is not coming from phloem sieve elements (conducting cells), which lack nuclei at maturity. Cells in which the signal is strongest are likely near the phloem; companion cells immediately adjacent to the phloem are one possibility. FT-GFP may be present in cells throughout the seed coat, but fewer proteins in cells further from the vasculature could result in undetectable signal.

3.4.4: FvFT1 may function to maintain seed dormancy in the absence of cold treatment

Many species require dormancy, or a period of quiescence, prior to seed germination. One of the most common forms, physical dormancy, is maintained by germination-inhibiting hormones in the absence of specific environmental triggers. In temperate climates, like those of the regions to which *F. vesca* is native, temperature is the primary environmental signal controlling seed dormancy; many species require a period of cold temperatures to break dormancy and initiate germination (Darrow, 1966; Penfield, 2017).

FvFT1 RNAi reduces the requirement of cold treatment to break dormancy and initiate seed germination (Fig 3.6 C). However, the specificity of this result to reduction of *FvFT1* function must be further investigated due to the green embryo phenotype shared by plants stably expressing an RNAi vector, regardless of the target gene (Fig 3.6 D-E; Fig A2 and A3). If specific to *FvFT1 RNAi*, I will conclude that *FvFT1* maintains dormancy as a negative regulator of seed germination in the absence of cold treatment.

Recent work in *Arabidopsis* suggests that *FT* expressed in the silique communicates maternal temperature experience to progeny and controls seed germination through altered seed coat tannin content. Specifically, *FT* is upregulated in siliques when plants are grown in warmer temperatures and freshly harvested seeds of *ft* mutants germinate less efficiently than WT, suggesting that *FT* positively regulates germination in *Arabidopsis* seeds by conveying temperature information (Chen et al., 2014). In the developing *F. vesca* receptacle fruit, *FvFT1* is differentially expressed in response to cooler temperature (Fig 3.1C) and is upregulated by exogenous auxin and GA application (Fig 3.5), but may instead negatively regulate seed germination in the absence of cold treatment, indicating that *FT*'s role in seed varies between species (Fig 3.6C). Taken together, *FvFT1* may convey information about maternal temperature experience and endogenous hormone levels to the seed, which are relevant for fine-tuning dormancy requirements to region-specific environmental conditions.

3.4.5: FT expression in fruit is not unique to strawberry

My work is the first to characterize *FT* in fleshy fruit development, though qRT-PCR data from apple, mandarin, and kiwi indicate that *FT* expression in fruit is not unique to strawberry or to the *Rosaceae* family (Kotoda et al., 2010; Nishikawa et al., 2007; Varkonyi-Gasic et al., 2013). The apple (*Malus domestica*) genome contains two *FT-like* genes, *MdFT1* and *MdFT2*, with differing expression patterns. *MdFT1* is expressed in apical buds but *MdFT2* is expressed in reproductive organs, including flower buds and young fruit (Kotoda et al., 2010). Conversely, mandarin (*Citrus unshiu*) has three *FT* homologs, two of which are strongly expressed during later stages of fruit development (Nishikawa et al., 2007). The single *FT* homolog identified in kiwi is expressed during fruit

development but declines with the onset of ripening (Varkonyi-Gasic et al., 2013). These observations suggest a conserved role for *FT* in fleshy fruit development, further study of which may have broad agricultural applications.

3.5: *Ongoing Work*

3.5.1: FvFT1 function in fruit

The function of *FvFT1* in *F. vesca* fruit and/or seed development is still unresolved. Junhui Zhou is continuing the effort to generate CRISPR-induced knockouts of *FvFT1*, *FvFT3*, and a simultaneous knockout of *FvFT1* and *FvFT3*. Junhui has been optimizing a CRISPR system in *F. vesca* for several years and his recent successes pave the way for future *FT* knockouts, despite our previous difficulties. A primary goal will be to repeat the germination efficiency experiments discussed previously to determine if loss of *FvFT1* function, or combined loss of *FvFT1* and *FvFT3*, results in reduced dormancy requirements for seed germination. A second goal, particularly for the dual *FvFT1 FvFT3* CRISPR plants, will be to examine fruit and seeds for morphological or developmental phenotypes that may indicate *FT* function.

3.5.2: Does GA mediate FvFT1 expression in the receptacle fruit?

An intriguing result from my thesis work is the detection of *FvFT1* expression in the young receptacle in response to exogenous auxin and GA_3 application to emasculated flowers. This observation leads me to hypothesize that GA, which has previously been shown to indirectly regulate *FT* expression in the *Arabidopsis* leaf via interaction of DELLAs with CO (Wang et al., 2016; Xu et al., 2016), is also indirectly regulating *FvFT1* transcription

in fruit. To test this idea, a future experiment will be to individually treat emasculated flowers with auxin or GA_3 and use qRT-PCR to test *FvFT1* expression in receptacles. This result will separate the effects of auxin and GA on *FvFT1* expression. Further, a postdoc in our lab, Julie Caruana, has recently published a loss-of-function DELLA mutant in *F. vesca* (Caruana et al., 2017). A future plan is to test *FvFT1* expression in receptacles of the DELLA mutant compared to WT. Since previous work in *Arabidopsis* indicates that a DELLA is a repressor of *CO* in the absence of GA (Wang et al., 2016), I hypothesize upregulation of *FvFT1* in the loss-of-function DELLA mutant background. If so, this result would point to DELLA-mediated, and therefore GA-mediated, regulation of *FvFT1* in fruit.

Chapter 4: Differential expression analyses identify receptacle fruit-associated genes and facilitate development of tissue-specific promoters

4.1: *Introduction*

Identification of genes involved in the spatial and temporal regulation of fruit development is useful for both crop improvement and fundamental developmental research. Since fruit and flower tissues are the last organs to form on a plant, the use of broadly expressed promoters may perturb other, non-target parts of the plant. The development of fruit-specific promoters as molecular tools will bypass the problems associated with broadly expressed promoters and open the door for functional analyses of genes involved in fruit development. For example, overexpression or RNAi constructs can be expressed specifically in fruit to test functions of novel genes. Further, genes with receptacle-specific expression patterns may play important roles in regulating receptacle development. Since strawberry fruit is unique in that it is derived from the floral receptacle, characterization of receptacle-specific genes may shed light on novel developmental processes or provide insight into how developmental regulation differs between receptacle-derived fruits and ovary-derived fruits, such as tomato, peach, and *Arabidopsis*.

In chapter 4, I used the comprehensive RNA-Seq datasets previously generated by our lab to identify 589 genes more highly expressed in the developing receptacle than in any of the other profiled flower, fruit, or vegetative tissues (all transcriptome samples are listed in Supplemental Table S2.1). I cloned upstream genomic sequences for a subset of eleven

genes and tested their transcriptional activity in receptacles using a GUS reporter in transgenic strawberry plants. Sequences that successfully activate the reporter are useful tools for the Liu lab and for the wider *F. vesca* research community.

4.2: *Materials and Methods*

Differential expression analyses to identify receptacle fruit-associated genes

Both DESeq2 and edgeR were used in R version 3.3.1.

DESeq2

Following its vignette, DESeq2 version 1.12.14 (Love et al., 2014) was used to identify genes with differential expression between all profiled stages of cortex and pith tissues and all other tissues included in previously generated flower and fruit transcriptome datasets (Kang et al., 2013; Hollender et al., 2014; Hawkins et al., 2017). Read counts mapped against CDS without normalization were used as input. Two groups, 'fruit' and 'other tissues,' were compared to identify differentially expressed genes. The 'fruit' group contained data for cortex stages 1-5 and pith stages 1-5 (two biological replicates each). The 'other tissues' group contained all other tissues profiled in both the flower transcriptome dataset (Hollender et al., 2014) and the fruit transcriptome dataset (Kang et al., 2013). This method does not take into account the different stages profiled for each tissue and is therefore a conservative approach due to the resulting large dispersion estimates. P value was adjusted using the Benjamini-Hochberg method. Cutoff was set at padj < 0.05 and log2 fold change > 2.

edgeR

Following the edgeR user guide, a classic edgeR analysis (edgeR version 3.14.0) (Robinson et al., 2010) was also used to identify genes differentially expressed between *F. vesca* cortex and pith tissues and all other profiled tissues in both the flower and fruit transcriptome datasets. The input and groups compared were the same as detailed for DESeq2 above. Prior to calling differential expression, library sizes were estimated using colSums. After filtering for lowly expressed genes, only genes with at least one read per million in at least three samples were kept. Functions were set at DGEList, calcNormfactors, estimateCommonDisp, estimateTagwiseDisp, exactTest, and topTags. False discovery rate is controlled by the Benjamini-Hochberg method. Cutoff was set at FDR < 0.05 and logFC > 2.

Descriptions for genes identified via both DESeq2 and edgeR were taken from Plaza version 3.0 (Proost et al., 2015). http://bioinformatics.psb.ugent.be/plaza /versions/plaza_v3_dicots/.

Heat maps were generated with Log2 transformed RPKM values using Morpheus, a web interface available from the Broad Institute (https://software.broadinstitute.org/morpheus/).

Generation of GUS transcriptional reporter constructs and strawberry transformation

Sequences upstream of the ATG start codon of genes 03606 (2471 bp), 16792 (2298 bp), 19774 (2318 bp), 07771 (1,592 bp), 25908 (2,115 bp), 21624 (2,394 bp), 09100 (2,338 bp), 02647 (2,459 bp), 06301 (2,490 bp), 28478 (2,441 bp), and 28153 (2,255 bp) were PCR

amplified from YW5AF7 genomic DNA using Phusion polymerase (NEB, Cat. # M0530S) and using primer sequences detailed below in Table 4.1. The resulting fragments were cloned into PCR8/GW/TOPO using a TA cloning kit (Invitrogen, Cat. # K250020). After confirmation by sequencing, promoters were subcloned into the binary vector pMDC162 (Curtis and Grossniklaus, 2003) by Gateway LR reaction (Invitrogen, Cat. #11791-100).

Constructs were transformed into YW 5AF7 plants using the method described in Chapter 2. A subset of the resulting transgenic plants were genotyped using primers GUS-F: 5' ACCGTTTGTGTGAACAACGA 3' and GUS-R: 5' AATGCGAGGTACGGTAGGAGT 3'. In total, six *gene03606p::GUS* lines, seventeen *gene19774p*::GUS lines, six *gene16792p::GUS* lines, one *gene07771p::GUS*, seven *gene25908p::GUS* lines, three *gene21624p::GUS* lines, two *gene09100p::GUS* lines, fifteen *gene02647p::GUS lines*, and eight *gene06301p::GUS* lines were generated. *Gene28478p::GUS* and *gene28153p::GUS* constructs have not yet been successfully transformed into *F. vesca*. Delores Lomberk and Ceil Muller from the USDA ARS lab at Epcot transformed *p::GUS* constructs for genes 07771, 25908, 21624, 09100, 02647, and 06301.

Table 4.1 Primer sequences used to amplify 2 – 2.5 kb regions upstream of fruit-associated genes

Primer Name	Sequence 5' to 3'
03606p-F	AGCCCCTGTGAAATATGAATTGC
03606p-R	CATTTTGCCATGTATGTTTGCTCGG
19774p-F	AGCAAGCTCACTTCCACATCCA
19774p-R	CAACGCCGGCCATTTCTGAAG
16792p-F	GAAACCAAGGCCAGGAGAAATG
16792p-R	GATTTACTATCTAAACAGCTTCAAAGC
07771p-F	TGTGAGCTTGGGAAACACAGCA
07771p-R	CGACATTGTGAGTTTGTGACAAG
25908p-F	CATTTCTGAACTCCAGTTACCTAC
25908p-R	GCCAATGACCAGTGACTTCAACAC

21624p-F	CGGTTGCGTCTCATCTACGTGAAA
21624p-R	CAGCGGCCATATGGGGATTGATAG
09100p-F	ACGCCGTTTTCACCGCTTAAGT
09100p-R	CCTCCCTTGCTCTTCTTCAT
02647p-F	GCAATTGTTAGTAACATGCTCGCT
02647p-R	CCTCCATTTCCAATTCAAAGACTTC
06301p-F	GACGAAAATTGGTGCAGACTTCA
06301p-R	CCTACCCATTTCTCTACTTCTTCTAC
28478p-F	GCCAAAGGGGACTGAATTTGGT
28478p-R	CGAAGCCATGCATACTATATATCTGCAGC
28153p-F	CTGACGAAAACTGGCAGAAATGCG
28153p-R	GGATACCATGGCTTTTCTGTCTC

GUS staining and photography

Longitudinally bisected receptacles from T_0 transgenic plants containing the *promoter::GUS* constructs were stained and photographed according to the method described in Chapter 3. At least 3 fruits from each developmental stage (stages 1-5) from each of at least 3 transgenic lines were GUS stained for all constructs except *pGene09100::GUS*. Only 2 transgenic lines were generated for *pGene09100::GUS* and receptacles from both lines were stained. Representative photos are shown. Dongdong Li, a visiting graduate student, stained fruits at stage 3 and stage 4 with X-Gluc from plants expressing *p::GUS* constructs for genes *09100*, *02647, 06301, 21624,* and *25908*.

GO Term Enrichment Analysis

Analysis was performed using the web-based AgriGO tool (http://systemsbiology.cau.edu.cn/agriGOv2/). GO enrichment was derived using Fisher's exact test and a false discovery rate cutoff < 0.05.

List of Supplemental Files (Files available in Google Drive at this address: https://drive.google.com/drive/folders/13LL1JBgt9IIONyq3rhDgfrVGGw1OSaWk?usp=sharing)

Dataset S4.1: Descriptions of 589 receptacle associated-genes identified via differential expression analyses

Dataset S4.2: RPKM data for 11 genes, the upstream regulatory regions of which were selected for development of cloning promoters

Dataset S4.3: GO Term enrichment analysis

Dataset S4.4: Transcription factors in the list of 589 receptacle associated genes

4.3: *Results*

4.3.1: Identification of receptacle fruit-associated genes using comprehensive flower and fruit transcriptomes

As discussed in chapters 1-3, our lab generated comprehensive RNA-Seq datasets to profile *F. vesca* flower and fruit tissues over multiple developmental stages (Hollender et al., 2012; Kang et al., 2013; Hollender et al., 2014; Hawkins et al., 2017). In total, 92 RNA-Seq libraries representing 46 different tissues and stages were profiled (46 libraries x 2 biological replicates = 92). Twenty-six samples (13 libraries x 2 biological replicates = 26) represent the receptacle fruit and range from floral stage 6-7 to the fruit turning stage, which immediately precedes ripening. Receptacle samples from fruit stages 1-5 were subdivided into pith and cortex (refer to Fig 1.5B). I took advantage of these datasets to ask what key

genes may be involved in stimulating the floral receptacle to develop into a fleshy fruit. To identify genes that are more highly expressed in the receptacle than in any of the other profiled flower, fruit, or vegetative tissues (hereafter called 'receptacle-associated genes'), I performed differential expression analyses using the DESeq2 (Love et al., 2014) and edgeR (Robinson et al., 2010) packages available in R. In total, I identified 589 genes with at least a 4-fold higher expression level in the receptacle versus in all other profiled tissues (Fig 4.1A; Supplemental Dataset S4.1). The majority of genes were identified by both edgeR and DESeq (Fig. 4.1B).

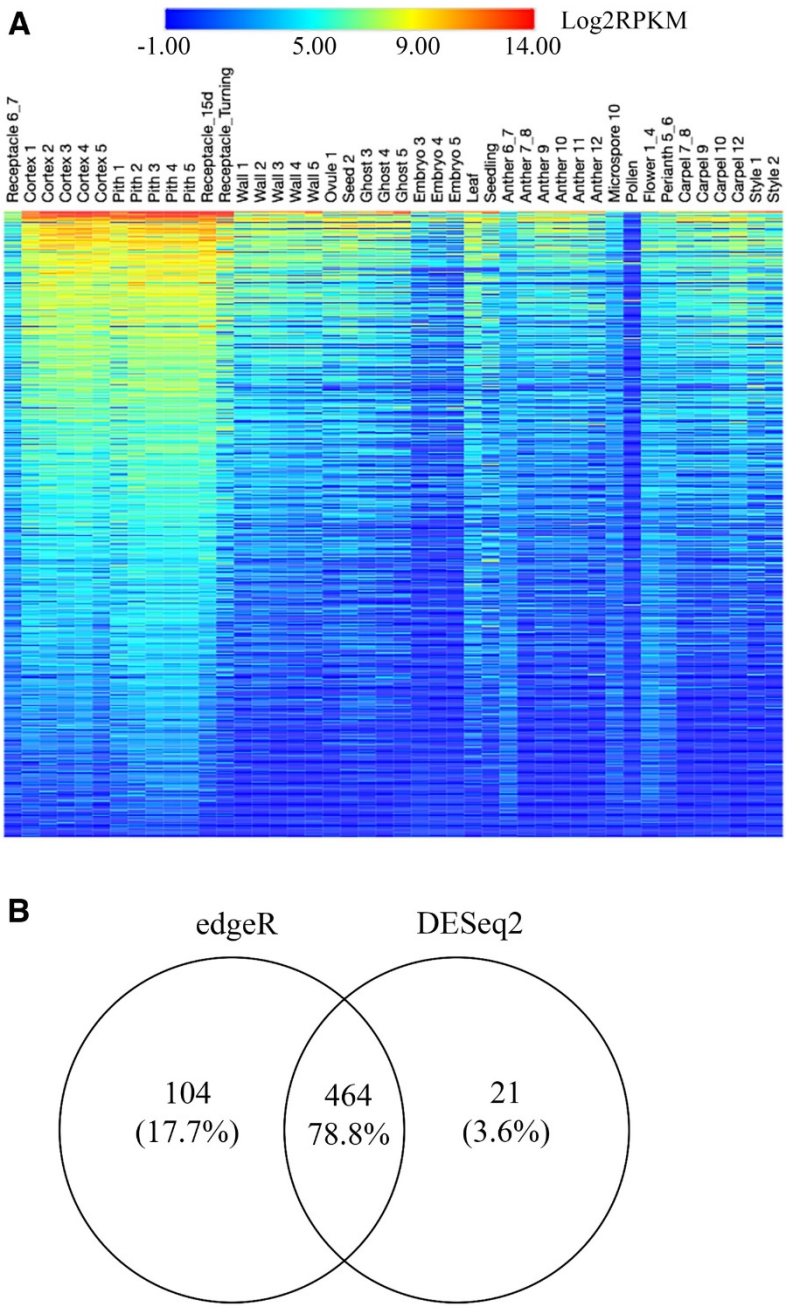


Figure 4.1 Differential expression analyses identified 589 receptacle-associated genes

(A) Heatmap showing expression profiles of 589 receptacle-associated genes across all profiled tissues and stages of development. Each row represents one gene. Warmer colors denote higher expression. Log2 transformed RPKM values were used as input to generate the heatmap. RPKM values from the two biological replicates of each library were averaged.

(B) Venn diagram indicating numbers of differentially expressed genes identified by the edgeR and DESeq2 packages.

4.3.2: Selection of genes for development of tissue-specific promoters

In addition to their potential roles as key regulators of fruit development, genes expressed primarily in the receptacle fruit are useful for the creation of molecular tools to advance *F. vesca* as a model system. I sought to isolate the regulatory sequences of a subset of the receptacle-associated genes and develop them as promoters for cloning. For this purpose, I chose eleven genes from the 589 fruit-associated genes identified via differential expression analysis (Figure 4.2). To develop a variety of promoters, I chose genes that differ in the timing and level of expression as determined by our RNA-Seq data (Supplemental Dataset S4.2). For example, genes 21624, 19774, 02647, 03606, and 09100 are expressed in the floral receptacle at flower stages 6-7 and pre-fertilization at stage 1 in both cortex and pith. Expression of genes 25908 and 06301 is detectable post-fertilization at stage 2 and is highest in cortex and pith from stage 3 through 15 DPA. Further, genes 21624 and 19774 have expression levels that are 100 times greater in the receptacle (RPKM >10,000 in pith post-fertilization) than all other selected genes (Supplemental Dataset S4.2). However, genes 21624 and 19774 also are more highly expressed in other profiled tissues than the other 9 selected genes. Gene02647 is expressed more highly in cortex than in pith and vice versa for gene28153.

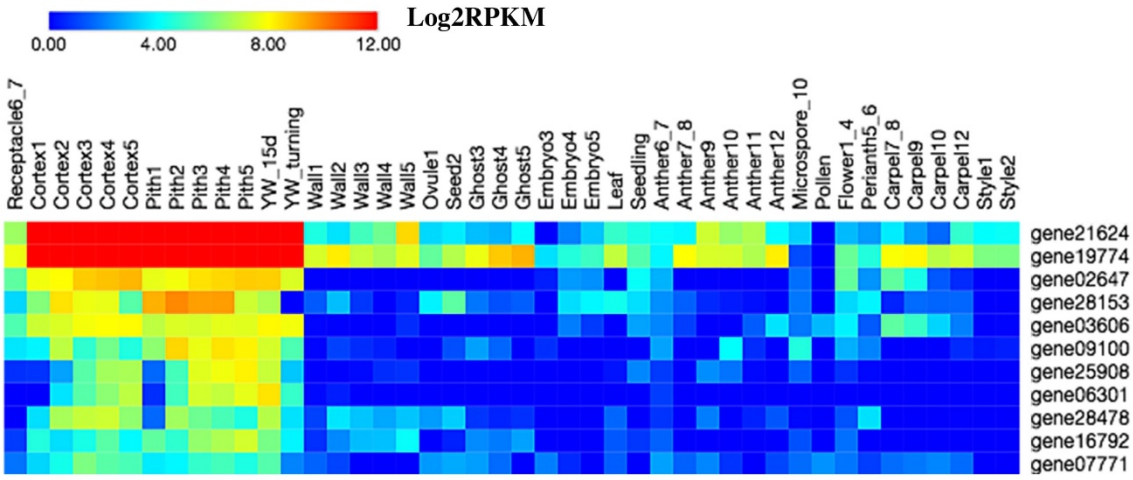


Figure 4.2 Subset of 11 genes chosen for development of tissue-specific cloning promoters

Heatmap showing expression profiles of 11 receptacle-associated genes across all profiled tissues and stages of development. Columns from 'receptacle6_7' to 'YW_turning' all represent receptacle tissues. Each row represents one gene. Warmer colors denote higher expression. Log2 transformed RPKM values were used as input to generate the heatmap. Average RPKM from the two biological replicates of each library are presented.

4.3.3: Confirmation of transcriptional activity of selected regulatory sequences

I isolated 2-2.5 kb genomic sequences ('regulatory sequences') upstream of the ATG start codon of each of the eleven genes and asked if they were sufficient to induce gene expression in the receptacle. To experimentally confirm transcriptional activity, I created constructs containing individual regulatory sequences driving expression of the GUS reporter gene, which generates a blue precipitate upon incubation with the chromogenic substrate X-gluc. Sequences that successfully activate GUS expression in the receptacle may be used as promoters to drive genes of interest in future research.

Thus far, Dongdong Li and I have tested activity of eight out of the eleven isolated regulatory sequences. One of the eleven, *promoterGene07771*, has only been successfully transformed into one independent line to date. The remaining two, *promoterGene28478*

and *promoterGene28153*, have not yet been successfully transformed into *F. vesca*. Of the eight regulatory sequences tested, seven successfully induced expression of the GUS reporter in the receptacle (Figs 4.3 and 4.4). As expected based on the transcriptome data, regulatory sequences upstream of genes 19774, 03606, and 16792 are active from stage 1 to stage 5 of receptacle development (Fig 4.3). Further, and also in agreement with the transcriptome, regulatory sequences of genes 19774 and 16792 induce GUS expression in the seed ('ghost') at stage 4. The isolated sequence upstream of gene03606 induces minimal GUS expression in the seed, visible primarily at the chalazal end (Figure 4.3). The regulatory sequence of gene19774 is particularly strong and GUS staining of both receptacles and seeds was apparent after only a few minutes in the staining buffer.

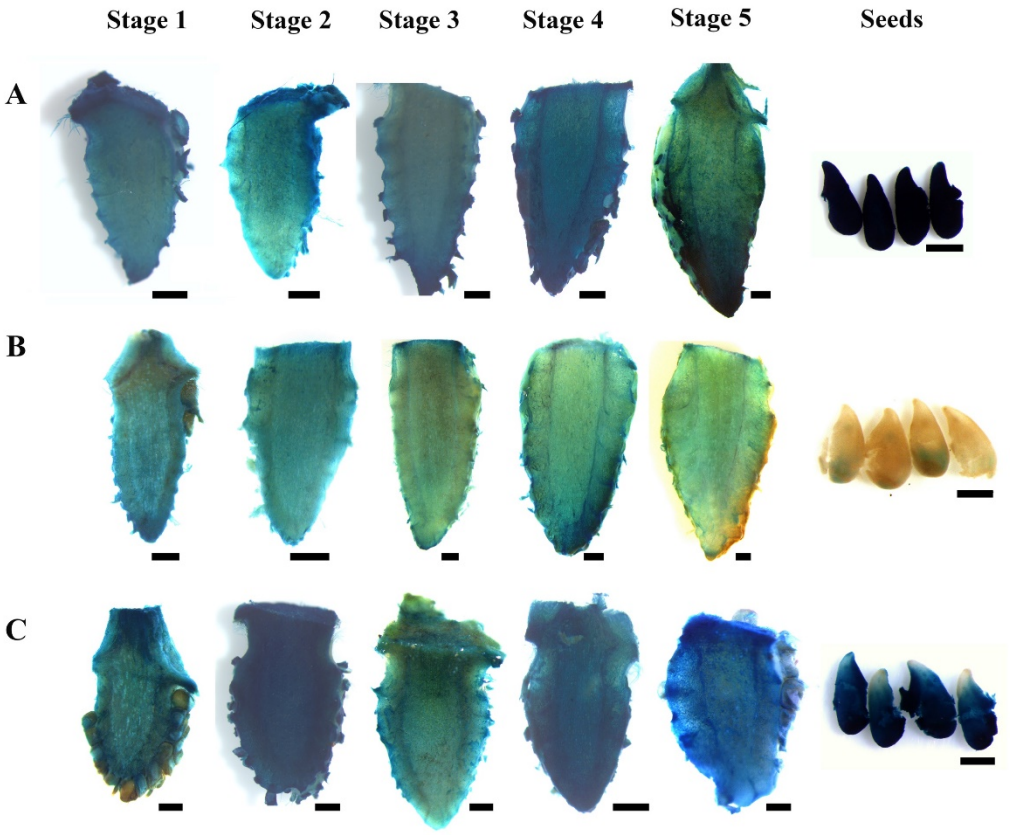


Figure 4.3 Three isolated regulatory sequences are active in receptacle from stages 1 to 5.

Stage 1 to 5 receptacles and stage 4 seeds after GUS staining. All scale bars are 500 µm.

(A) Receptacles and seeds from plants stably expressing *pGene19774::GUS*
(B) *pGene03606::GUS*
(C) *pGene16792::GUS*

Dongdong Li confirmed activity of four regulatory sequences in the receptacle post-fertilization (Figure 4.4). GUS staining is evident in stage three receptacles from plants expressing *pGene21624::GUS, pGene02647::GUS, pGene06301::GUS,* and *pGene25908::GUS* constructs. Further, staining is also evident in stage four receptacles for *pGene21624::GUS* and *pGene02647::GUS.* Receptacles from plants expressing the *pGene09100::GUS* construct did not show GUS staining at any of the tested stages.

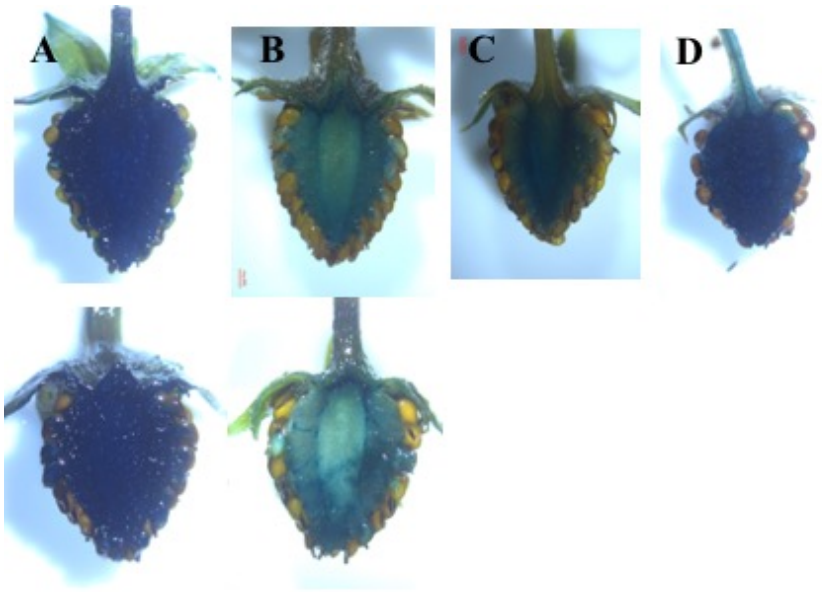


Figure 4.4 Transcriptional activity of an additional four regulatory sequences confirmed in receptacle post-fertilization

Receptacles after GUS staining. Top row: stage 3. Bottom row: stage 4.

(A) Receptacles from plants stably expressing *pGene21624::GUS*
(B) *pGene02647::GUS*
(C) *pGene06301::GUS*
(D) *pGene25908::GUS*

Receptacles expressing *pGene21624::GUS* (Fig 4.4 A) stain more strongly than receptacles expressing either *pGene02647::GUS* or *pGene06301::GUS*, which is expected given the transcriptome data (Figure 4.2; Supplemental Dataset S4.2). Receptacles expressing *pGene02647::GUS*, at least at stage 4 (Fig 4.4 B), seem to show slightly darker staining in the cortex than the pith, which is also expected given the expression profile of Gene02647 (Fig 4.2). However, *pGene25908::GUS* receptacles at stage 3 stain just as darkly as *pGene21624::GUS* receptacles. This is unexpected given the transcriptome data; gene25908 is more lowly expressed than both genes 21624 and 02647 (Fig 4.2). In fact, gene21624 is expressed over 350 times higher in the stage 3 cortex than gene25908 (Supplemental Dataset S4.2).

4.3.4: 'Regulation of Transcription' GO Terms are enriched among the receptacle-associated genes

I asked what biological processes are overrepresented among the 589 receptacle-associated genes, reasoning that key regulators of fruit development will be those that control these processes. Among the sixteen categories of biological process GO terms enriched in the list of receptacle-associated genes, three involve response to abiotic stimuli, including light and radiation (Supplemental Dataset S4.3). Ten categories are metabolic or biosynthetic processes, including regulation of nucleic acid metabolic process, regulation of RNA metabolic process, and regulation of macromolecule biosynthetic process. Three of the enriched GO terms involve transcriptional regulation of gene expression (Supplemental Dataset S4.3). I found that 51 of the 589 fruit-associated genes are annotated as transcription factors (Supplemental Dataset S4.4). Several different transcription factor families are represented, including the Myb, WRKY, and MADS-box families. The MADS-box transcription factors are genes 06301, 30741, and 26119. Gene26119 shares homology with *AP1* from *Arabidopsis,* a known regulator of floral meristem and floral organ identity (Irish and Sussex, 1990).

4.4: *Discussion*

In this chapter, I performed differential expression analyses on previously generated *F. vesca* fruit and flower RNA-Seq data. I identified 589 genes that are more highly expressed in the receptacle versus in all other profiled reproductive and vegetative tissues. I isolated upstream regulatory sequences from 11 of the 589 genes and tested their ability to activate transcription in the receptacle. Thus far, Dongdong Li and I have shown that seven of the

eleven sequences successfully activate expression of the GUS reporter gene in the receptacle. These sequences may now be used as promoters to drive expression of genes of interest in the developing receptacle. Identified receptacle-associated genes, particularly transcription factors, are candidate regulators of early fruit development and further elucidation of their functions in strawberry may provide insight into general developmental processes. Of the 589 identified fruit associated genes, 51 are annotated as transcription factors.

4.4.1: Isolated regulatory sequences are useful to drive expression of genes of interest in the receptacle

In *Arabidopsis*, promoter deletion experiments suggest that most genes have functional promoters within ~1400 bp of the translation start site (Conley et al., 1994; Tjaden et al., 1995; Brown et al., 2003). No similar experiments have thus far been conducted in strawberry, but the *F. vesca* genome size and gene density (240 Mb and 7 kb/gene, respectively; Shulaev et al., 2011) are only slightly larger than those of *Arabidopsis* (135 Mb and 4.9 kb/gene, respectively) (The Arabidopsis Genome Initiative, 2000; Yamamoto et al., 2011). Therefore, in an attempt to capture the entirety of the promoter regions, I amplified and cloned a 2-2.5 kb region upstream of the ATG start codon of each of my selected receptacle-associated genes. It is possible that the cloned regions lack regulatory elements and do not contain full promoter sequences. However, since the goal is to identify promoters for use in future experiments by the strawberry research community, the GUS reporter experiment is enough to confirm that I have identified a regulatory sequence that is sufficient to promote gene expression in the receptacle fruit.

The variation in receptacle specificity and expression strength of the genes from which regulatory sequences were isolated provides a variety of promoters for downstream research. In particular, genes 19774 and 21624 are extremely highly expressed in the receptacle but also strongly expressed in other reproductive tissues (Figs 4.2-4.4). Thus, their regulatory sequences could be useful promoters to overexpress genes of interest in *F. vesca* flowers and fruit. The isolated regulatory sequence for gene03606 is more specific to the receptacle and induces minimal GUS expression in the seed (Fig 4.3), thereby highlighting its usefulness for targeted expression in the receptacle. Isolated regulatory sequences from genes 02647, 06301, 21624, and 25908 successfully induce GUS expression in the receptacle post-fertilization, but I am still testing their specificity in the receptacle vs the seed with the help of Dongdong Li. An additional plan is to GUS stain transgenic flowers and seedlings for all *p::GUS* constructs to test activity of the isolated regulatory sequences in reproductive tissue versus vegetative tissue. The lack of GUS staining in receptacles from each of two lines successfully transformed with *pGene09100::GUS* suggests that the isolated sequence may lack necessary regulatory elements and is thus unable to initiate transcription of the GUS reporter. However, generation of more independent transgenic lines is necessary to confirm.

4.4.2: MADS box transcription factors as candidate regulators of fruit identity

Several of the GO term categories enriched in the list of 589 fruit-associated genes are related to transcriptional regulation. Among the list of receptacle-associated genes are several members of the MADS box transcription factor family, a finding supported by the receptacle-associated cluster 13 from the co-expression network discussed in Chapter 3. Though previous work in tomato revealed a role for the MADS-Box *RIN* gene in ripening

(Vrebalov et al. 2002), genes that establish and maintain fruit identity are largely unknown. The roles of the ABC MADS box genes are well characterized in floral identity specification (see Chapter 1, section 1.5); perhaps members of this family also regulate fruit identity. The *F. vesca AP1-like* gene26119 is highly and specifically expressed in the receptacle (Supplemental Dataset S4.1) and is an attractive candidate for further study based on the known role of *AP1* in *Arabidopsis* floral organ identity (Irish and Sussex, 1990).

Chapter 5: Conclusions and remaining questions

Angiosperms have been bred and cultivated for millennia. Maximizing crop yield and quality for the purposes of human and animal consumption is an historical as well as enduring challenge. The advent of genome-wide analyses, including powerful gene expression data generated via RNA-Seq, has ushered in a new era of plant biology. Economically relevant, yet previously understudied, organisms are increasingly accessible for molecular research. The diploid strawberry, *Fragaria vesca* (*F. vesca*), is a recently developed model for members of the *Rosaceae* family, including the commercial strawberry, apple, peach, and plum. Further, an external seed configuration and the unique derivation of fleshy fruit tissue from the floral receptacle make strawberry an ideal organism with which to study basic mechanisms underpinning fruit development. The work presented in this thesis aims to (1) maximize the utility of previously generated transcriptome data via creation of gene co-expression networks; (2) characterize a novel function for a homolog of *FLOWERING LOCUS T* (*FT*) in strawberry fruit development; and (3) generate molecular tools useful for the *F. vesca* research community.

In Chapter 2, I present standard (WGCNA; Langfelder and Horvath, 2008) and consensus gene co-expression networks generated from 92 libraries of flower and fruit RNA-Seq data with the help of Chris Zawora, a bioinformatician. The aim of the networks is to provide a platform for data exploration and hypothesis generation. All network information is freely available to the research community on a custom web interface. Building the networks from a large number of datasets increases the robustness of gene clusters and is therefore an improvement over a previous network generated with only a small subset of our lab's

transcriptome data (Hollender et al., 2014). Consensus clustering of the standard networks generated with WGCNA aims to further strengthen cluster robustness with a bootstrapping approach to simulate sampling variability. However, to assess the accuracy and overall utility of co-expression network analysis, the biological relevance of predicted genetic relationships must be assessed experimentally. Using the consensus_LCM network, I found that one of the three predicted *F. vesca* homologs of *UFO, FveUFO1*, is clustered separately from the other two homologs. *FveUFO1* is also tightly connected to *FveLFY*. Since UFO and LFY proteins are known to interact in *Arabidopsis,* I hypothesized that *FveUFO1* is not redundant and loss of *FveUFO1* function would cause a mutant phenotype in strawberry. Using a mapping-by-sequencing approach, I identified a candidate mutation in *FveUFO1* as the cause of floral meristem determinacy and floral organ identity defects in an EMS-generated mutant that I named *extra floral organs* (*efo*). This result experimentally supports predictions made based on mining the co-expression networks. Further, the *efo* mutant phenotype supports an expanded role of *UFO* and suggests that *FveUFO1* may regulate A, B, and C class gene activity in strawberry. However, because I was unable to successfully complement the *FveUFO1* mutation, further work is needed to definitively show that the identified mutation in *FveUFO1* is the cause of the *efo* floral phenotypes.

Based on information in the co-expression networks, I hypothesized that iron transport to the ghost (endosperm + seed coat) is increased after fertilization. I used an established staining procedure to experimentally demonstrate increased free iron in the receptacle and seed vasculature post-fertilization, thereby providing initial support for my hypothesis. This work highlights the usefulness of co-expression networks to identify biological

processes underlying developmental phenomena. Further, predicted iron transporters identified by the co-expression network are candidate regulators of iron movement to the ghost (endosperm + seed coat) post-fertilization. GA biosynthesis is a fertilization-induced signal that initiates fleshy fruit development and in strawberry, the likely site of GA biosynthesis is the ghost (Kang et al., 2013). GA biosynthesis enzymes require an iron co-factor, thereby pointing to fertilization-induced iron transport to the ghost as a previously uncharacterized, yet crucial, process for strawberry fleshy fruit development.

In Chapter 3, I used differential expression analyses of our lab's transcriptome data coupled with the gene co-expression networks from Chapter 2 to ask what genes are preferentially expressed in the early stage strawberry receptacle fruit. I reasoned that genes highly expressed in the receptacle soon after fertilization are candidate key regulators of fruit development. I found that expression of *FvFT1*, a homolog of the florigen *FT*, is surprisingly upregulated in the receptacle post-fertilization. This novel observation, coupled with a literature search that revealed evidence for expression of *FT* homologs in mandarin and apple fruit, spurred me to hypothesize that *FvFT1* has a previously uncharacterized role in fleshy fruit development. Further transcriptome analysis also indicated that *FvFT1* likely does not function in the same genetic pathway in fruit as it does in the leaf and at the SAM. However, similar to its expression in leaf vasculature, I showed that *FvFT1* is transcribed in the receptacle vascular bundles, including the vasculature that connects the receptacle to the overlying seeds. A GFP translational fusion driven by the native *FvFT1* promoter indicates that the FvFT1 peptide is localized in seeds and likely originates from the receptacle, thereby indicating FvFT1 mobility in fruit and pointing to a role for *FvFT1* in seed development.

I used *FvFT1 RNAi* plants (Koskela et al., 2012) to interrogate the function of FvFT1 in seeds. Compared to WT, *FvFT1 RNAi* seeds germinate at a higher frequency without cold treatment. However, an embryo greening phenotype observed in *FvFT1 RNAi* seeds is not unique to loss of *FvFT1* function and is observable in *RNAi* plants targeting other genes. An ongoing experiment aims to determine if the seed germination phenotype is unique to *FvFT1 RNAi.* If so, I will conclude that *FvFT1* is a positive regulator of seed dormancy. If the seed germination phenotype is not unique to *FvFT1* and is instead, like the embryo greening phenotype, attributable to transformation with an *RNAi* construct, *FvFT1 RNAi* will have failed to produce a mutant phenotype in fruit. This could be due to functional redundancy of *FvFT1* with its two paralogs, *FvFT2* and *FvFT3*. To address this possibility, experiments are currently underway to simultaneously knock out *FvFT1* and *FvFT3* via CRISPR.

Two recent studies in *Arabidopsis* indicate that DELLAs, proteins that mediate GA-induced transcriptional regulation, physically interact with CONSTANS (CO) and CO mediates positive regulation of *FT* in the leaf by GA (Xu et al., 2016; Wang et al., 2016). In the strawberry receptacle, I showed that auxin and/or GA positively regulates *FvFT1* transcriptional activation. Based on my result and the results of Xu et al. (2016) and Wang et al. (2016), I hypothesized that *FvFT1* transcription is indirectly regulated by GA in the receptacle. To test this idea, an ongoing experiment aims to determine if *FvFT1* expression in the receptacle is upregulated in an *F. vesca* DELLA mutant (Caruana et al., 2017). If the resulting data supports my hypothesis, I will conclude that endogenous GA signals upregulate *FvFT1* in the receptacle and may subsequently influence seed development and/or germination. As indicated above, functional data is necessary to understand what

role *FvFT1* plays in the fruit and seed. If *FT* integrates environmental and endogenous signals in the receptacle as it does in the leaf, I predict that *FvFT1* communicates information from the mother plant to regulate maturation, dormancy, and/or subsequent germination of progeny seeds.

As a new model system, development of molecular tools for *F. vesca* is crucial for future research endeavors. In Chapter 4, I identified 589 receptacle-associated genes via differential expression analyses of our lab's existing transcriptome data. I selected a subset of eleven genes from which I isolated upstream regulatory sequences. The subset of eleven genes vary in expression strength and tissue specificity. I generated transcriptional reporters and have thus far confirmed activity in the receptacle for eight of the eleven sequences. These eight sequences may be used as promoters to drive expression of genes of interest in the receptacle. The promoter of gene03606, which is active in the receptacle but induces minimal transcription in the seed, will be particularly useful for its receptacle specificity.

In conclusion, this thesis advances our understanding of strawberry fruit development and demonstrates the power of large scale transcriptomics, especially in a developing model system, to inspire novel hypotheses and aid experimental design. The work presented here lays the foundation to understand a novel and exciting role of the florigen, *FLOWERING LOCUS T,* in fleshy fruit development. The resources developed increase the utility and accessibility of large scale transcriptomic data, especially for biologists lacking a computational background, and further advance *F. vesca* as a model system, thereby strengthening the foundation for future research and future discoveries.

Appendix 1

Transgenic *F. vesca* plants containing RNAi constructs produce precociously green embryos

Section 1: Brief introduction to chloroplast maturation

Seeds are often covered by soil after dispersal from the mother plant. Upon germination, subterranean seedlings must break the soil surface to receive light exposure. Sunlight is the most important environmental stimulus for chloroplast maturation from precursor proplastids and etioplasts, a process which is required for photosynthesis and subsequent plant growth (Solymosi and Schoefs, 2010; Jarvis and Lopez-Juez, 2013). Prior to chloroplast development and photosynthesis initiation, the embryonic leaves (cotyledons) are storage organs containing proteins, fats, and/or starch. Light exposure initiates a developmental program during which proplastids and etioplasts undergo structural changes as they develop into mature chloroplasts. Specifically, proplastids, etioplasts, and mature chloroplasts are distinguishable by the arrangements of their inner membranes (Fig A1). Proplastids are largely undifferentiated but the inner membranes of etioplasts are arranged in a crystalline structure called the prolamellar body (PLB; Fig A1 A and B) (Solymosi and Schoefs, 2010; Jarvis and Lopez-Juez, 2013). Once a seedling breaks the soil surface and is exposed to light, the PLB arrangement is relaxed and thylakoid membranes are formed as the chloroplast matures (Fig A1 C) (Solymosi and Schoefs, 2010; Pogson and Albrecht, 2011; Jarvis and Lopez-Juez, 2013).

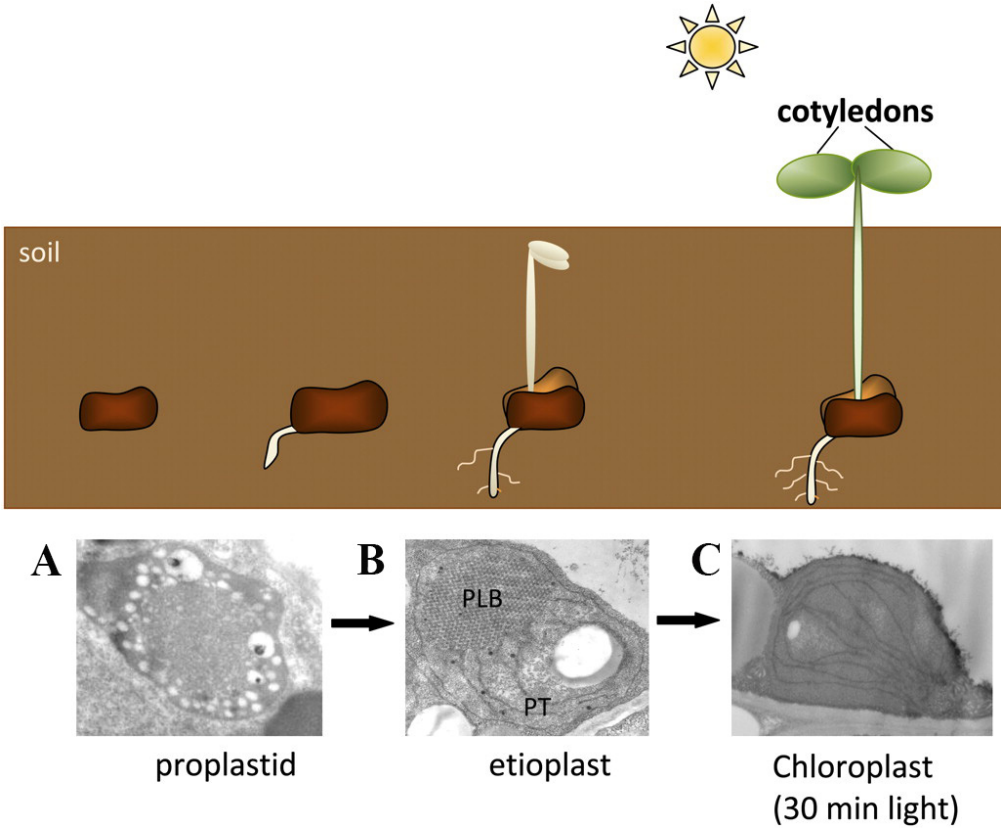


Figure A1 Chloroplast maturation in cotyledons of germinating seedlings

TEM photos showing differences in morphology between proplastid, etioplast, and chloroplast. PLB: prolamellar body. PT: prothylakoids. Figure adapted from Pogson and Albrecht, 2011.

The PLB contains protochlorophyllide, the precursor to chlorophyll, bound to its reducing enzyme protochlorophyllide oxidoreductase (POR). POR converts protochlorophyllide into chlorophyllide *a,* which is then converted into chlorophyll *a* and *b* (Pogson and Albrecht, 2011). In angiosperms, including *F. vesca*, POR enzymatic activity requires light. Thus, seedlings do not produce chlorophyll and do not turn green prior to breaking the soil surface. Unlike angiosperms, gymnosperms also have a light-independent POR enzyme that catalyzes the same reaction (Forreiter and Apel, 1993).

As mentioned in Chapter 3, I observed that *F. vesca* plants stably expressing *FvFT1 RNAi* produce embryos that are green when dissected from seeds of fully ripe fruit. In Appendix

1, I address the specificity of the embryo phenotype to *FvFT1 RNAi* and use transmission electron microscopy to determine if precocious embryo greening may be due to premature chloroplast development in the absence of light.

Section 2: Materials and Methods

Vector construction

Three independent lines of transgenic Hawaii-4 seeds containing the RNAi construct pK7GWIWG2(II) targeting *FvFT1* (gene21535) were provided by Timo Hytönen (Koskela et al., 2012). The *GA20ox* (gene09034) and *PIN5* (gene16792) *RNAi* constructs were generated by Drs. Julie Caruana and Chunying Kang, respectively. Construction of the translational fusion construct *pFvFT1::FvFT1-GFP* is described in Chapter 3.

Strawberry transformation and plant growth conditions

The *GA20ox RNAi, PIN5 RNAi,* and *pFvFT1::FvFT1-GFP* constructs were transformed into either the Hawaii-4 (*GA20ox RNAi*) or Yellow Wonder 5AF7 (*PIN5 RNAi* and *pFvFT1::FvFT1-GFP)* backgrounds via the method described in Chapter 2. Drs. Julie Caruana and Chunying Kang transformed the *GA20ox RNAi* and *PIN5 RNAi* constructs, respectively.

All WT and RNAi plants were grown in growth chambers set to 16 hours of daylight at 25°C and 8 hours of darkness at 20°C.

Tissue Dissection and Photography

Embryos were dissected by hand using a stereomicroscope and photographed with an Axiocam 503 color camera and ZEN Blue software. A KSC 295-814D GFP cube for use with PentaFluor and a Lumen Dynamics X-Cite Series 120 Q light source were used for GFP imaging.

Transmission Electron Microscopy

WT H4 and *FvFT1 RNAi* (line 1 in H4 background) embryo samples were first fixed in 0.12 M Millonig's phosphate buffer (pH 7.4) with 2% glutaraldehyde (Millonig, 1964). After dissection, samples were left in fixative for 1 hour at room temperature and then stored at 4ºC for at least one night. Samples were subsequently washed 3 x 10 minutes in Millonig's buffer to remove excess glutaraldehyde. Then, samples were secondarily fixed in Millonig's buffer with 2% osmium tetroxide at room temperature for 1 hour. Following 3 washes in double distilled water, samples were stained for 1 hour in 2% aqueous uranyl acetate. Samples were then dehydrated through a graded ethanol series (35%, 50%, 70%, 95%, 100%) and embedded in Spurr's resin (Spurr, 1969). Ultrathin (70 to 80 nm) sections were cut on an ultrathin microtome (Reichert Ultracut E). The sections were stained with 0.2% lead citrate for 1.5 minutes and photographed with a lanthanum boride emitter-equipped Hitachi HT7700 transmission electron microscope operated at an accelerating voltage of 80 kV. Two individual embryos were used for each genotype and representative images are shown.

Section 3: Results

As mentioned in Chapter 3, I found that transgenic embryos from plants stably expressing *FvFT1 RNAi* are green when dissected from achenes of ripe fruit. This is in contrast to WT *F. vesca* embryos of the same developmental stage which are opaque and white. I asked at what developmental time point the *FvFT1 RNAi* embryos turn green. Both WT and *FvFT1 RNAi* embryos are translucent from stage 3 to stage 4 (6 and 9 DPA; Fig A2) and opaque beginning at stage 5 (12 DPA). WT embryos remain white through 25 DPA, at which point the fruit is ripe (Fig A2 B). *FvFT1 RNAi* embryos, however, are white at 15 DPA, turn yellow-green at 20 DPA, and remain green at ripening (Fig A2 A). This timeline suggests that *FvFT1 RNAi* embryos are developing similarly to the WT until 20 DPA, which is known as the 'turning stage.' This is the developmental stage at which red fruited varieties turn pink and begin to ripen.

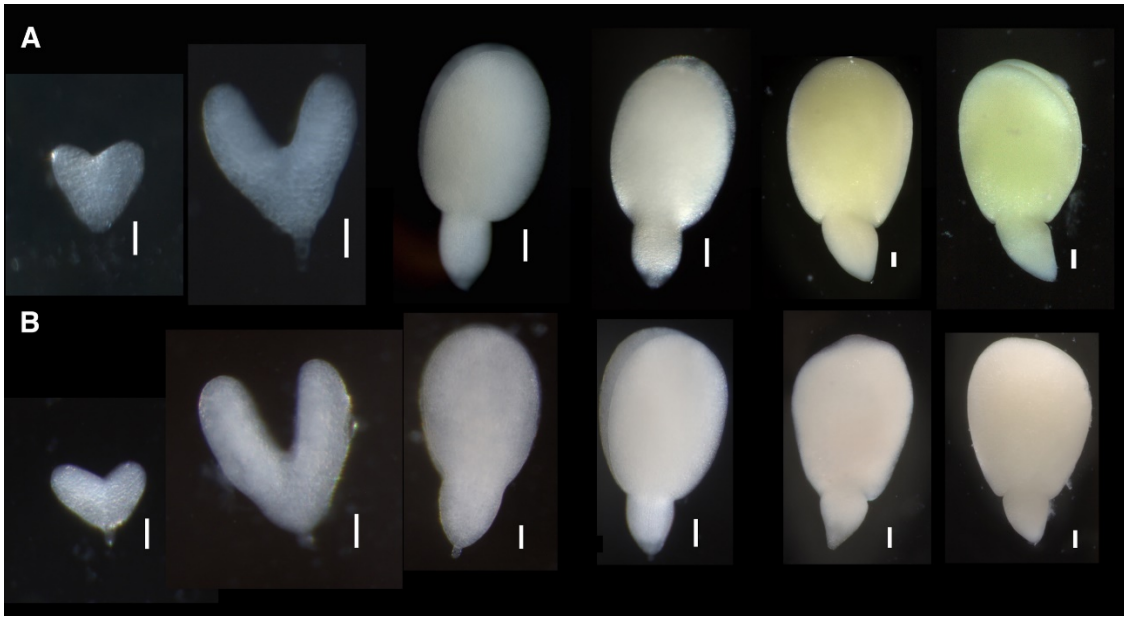


Figure A2 FvFT1 RNAi embryos are green at turning stage (20 DPA)

FvFT1 RNAi line 1 **(A)** and WT H4 **(B)** embryos dissected from achenes at, left to right, 6, 9, 12, 15, 20, and 25 DPA. In A and B, scale bars are 50 μm at 6, 9, and 12 DPA and 100 μm at 15, 20, and 25 DPA.

The precocious greening phenotype is evident in embryos from multiple independent *FvFT1 RNAi* lines (Fig A3 A and B). The phenotype is only visible in transgenic embryos as indicated by a *35S::GFP* marker on the RNAi vector (Fig A3 A, B, D, E). GFP negative, and thus non-transgenic, embryos from plants stably expressing an RNAi vector appear WT in color. Both transgenic and non-transgenic seeds and embryos may be found on the same fruit. Plants stably transformed with a non-RNAi construct, e.g., *pFvFT1::FvFT1-GFP* (Fig A3 C), do not produce green embryos. However, plants transformed with RNAi constructs targeting genes other than *FvFT1* also produce precociously green embryos (Fig A3 D, E), indicating that the phenotype is independent of *FvFT1*. The green embryo phenotype is observable in plants transformed with two different RNAi vectors: pK7GWIWG2(II) (Fig A2 E; Karimi et al., 2002) and pH7GWIWG2-7F2,1 (Fig A3 A, B, D). Both vectors were generated by the Depicker lab at the VIB-UGent Center for Plant Systems Biology.

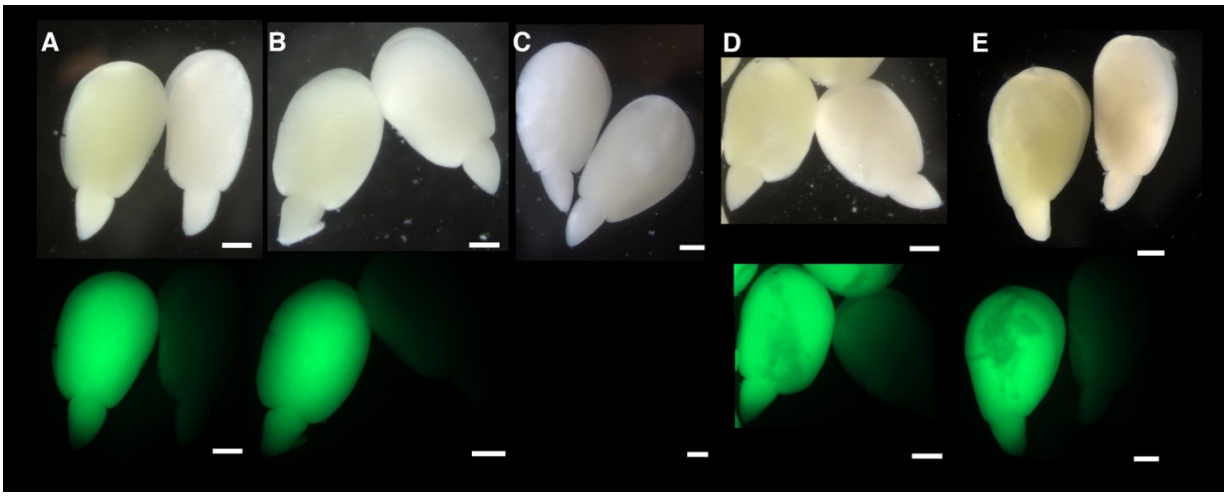


Figure A3 Embryos from plants transformed with RNAi vector, but not other vectors, are precociously green

Top: Transgenic (left) and non-transgenic (right) embryos collected from ripe fruit of *FvFT1 RNAi* line 1 **(A)** *FvFT1 RNAi* line 2 **(B)**, *pFvFT1::FvFT1-GFP* **(C)**, *FvGA20ox RNAi* **(D)**, and *FvPIN5 RNAi* **(E)**. **Bottom:** Seeds from *RNAi* plants are confirmed as transgenic by a *35S::GFP* marker on the vector **(A, B, D, E)**. *pFvFT1::FvFT1-GFP* binary

vector (pMDC110; vector construction discussed in Chapter 3 methods) does not have a GFP marker and nuclear localized GFP signal (discussed in Chapter 3) is not discernible with the fluorescence stereoscope. All scale bars are 200 μm.

I asked if the precocious greening phenotype of *FvFT1 RNAi* embryos results from premature chloroplast development and, thus, premature chlorophyll production. To test this, I sought to examine proplastid, etioplast, and/or chloroplast ultrastructure via transmission electron microscopy (TEM). I reasoned that observation of relaxed prolamellar bodies (PLB) in cotyledon cells of *FvFT1 RNAi* embryos would indicate precocious development of chloroplasts prior to germination and exposure to light. Examination of both 25 DPA embryos from WT H4 (Fig A4 A) and *FvFT1 RNAi* (Fig A4 B and C) revealed what appear to be either proplastids or etioplasts containing crystal-like structures. I observed no relaxed PLBs in the *FvFT1 RNAi* embryo samples (Fig A1 C), suggesting no precocious chloroplast maturation. This result is in line with the fact that the angiosperm POR enzyme requires light for activation and subsequent chlorophyll production (Pogson and Albrecht, 2011). Strawberry seeds are encased in the achene, which has a hard, outer shell and is likely impermeable to light. The cause of the green embryo coloration remains unknown.

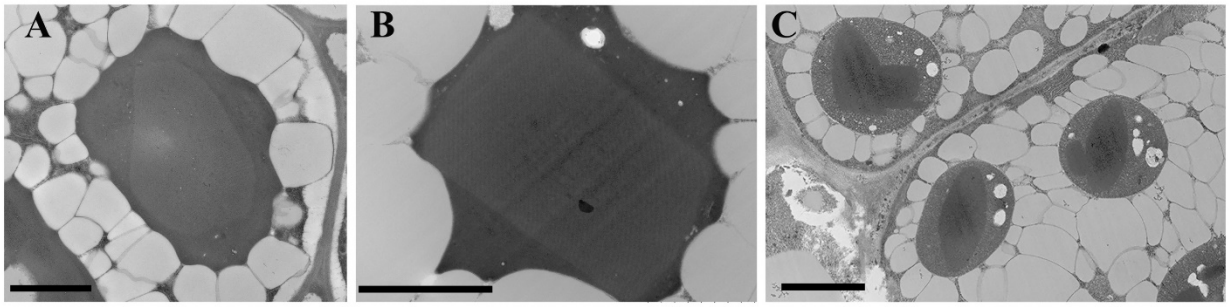


Figure A4 No evidence of developing chloroplasts in cotyledon cells of FvFT1 RNAi embryos

Representative images of proplastid or etioplast ultrastructure in 25 DPA embryos from **(A)** WT H4; **(B)** *FvFT1 RNAi*; and **(C)** *FvFT1 RNAi*. All scale bars are 2 μm.

Section 4: Concluding Remarks

Strawberry embryos are fully formed relatively early in fruit development at 10 DPA (Fig A2; Fait et al., 2008) and have been reported to store protein and fat but not starch (Nitsch, 1950). Later in fruit development, during the third general stage as described by Gillaspy et al. (1993; see Chapter 1, section 2.4), cell division has ceased and fleshy fruit is enlarging due to increases in cell volume. During this period, the embryo undergoes a maturation process during which seed dormancy is initiated. This dormancy is characterized by the accumulation of storage products (protein and fat), suppression of precocious germination, acquisition of desiccation tolerance, and water loss (Bewley and Black, 1994).

In WT *F. vesca*, embryos are translucent early in development until ~12 DPA and subsequently opaquely white through ripening (Fig A2 B). Conversely, embryos dissected from seeds of ripe (25 DPA) *FvFT1 RNAi* fruit are green. Dissection of embryos at six stages throughout development, from 6 to 25 DPA, indicates that the greening phenotype emerges at 20 DPA in *FvFT1 RNAi* embryos, well after embryo development is complete and during what should be the maturation period (Fig A2 A). Additionally, as reported in Chapter 3 (Fig 3.6), *FvFT1 RNAi* seeds do not require cold treatment for germination. Together, my data suggests that *FvFT1 RNAi* embryos develop normally initially but fail to fully mature and initiate dormancy. However, TEM images do not indicate that chloroplasts develop precociously in *FvFT1 RNAi* embryos as compared to WT (Fig A4). This result indicates that the green color of *FvFT1 RNAi* embryos is likely not due to premature chlorophyll production.

It remains to be determined if precocious embryo greening and seed germination in the absence of cold (Fig 3.6C) reflect a common defect induced by *FvFT1 RNAi*. I plan to test *PIN5 RNAi* seeds for increased germination frequency in the absence of cold as compared to WT. This experiment will determine if precociously green embryos and decreased dormancy are observed together regardless of the gene targeted with RNAi. If *PIN5 RNAi* seeds do not germinate at an increased frequency without cold, this result would point to an *FvFT1*-induced phenotype. If loss of *FvFT1* function results in decreased dormancy (i.e., increased germination without cold treatment), I will conclude that *FvFT1* is a negative regulator of germination in the absence of cold.

Transgenic embryos dissected from plants stably expressing non-RNAi constructs are WT in color, suggesting that transformation by tissue culture does not cause the phenotype. However, observation of green embryos from plants independently transformed with RNAi constructs targeting *PIN5* (gene16792), an auxin transport gene, and *GA20ox* (gene09034), a GA biosynthesis gene, indicates that the phenotype is not unique to *FvFT1.* A more robust control experiment would be to transform an empty RNAi vector into *F. vesca* and ask if embryos dissected from seeds of mature fruit are also green. If embryos are green, the result would suggest that transformation with an RNAi vector itself is somehow causing the phenotype. Control plants with an empty pH7GWIWG2-7F2,1 vector will be generated in the near future by our colleagues, Delores Lomberk and Ceil Muller, of the USDA ARS lab at Epcot. An additional plan is to observe embryos from plants with CRISPR-induced loss of *FvFT1* function. This experiment would further determine if it is RNAi itself or loss of *FvFT1* function that is responsible for the embryo phenotype. As discussed in Chapter 3, Dr. Junhui Zhou is currently working to generate *FvFT1* CRISPR plants. Because loss

of *PIN5*, *GA20ox*, and *FvFT1* functions each result in precocious embryo greening, it is likely that the phenotype is more broadly caused by RNAi in general. Although I cannot currently confirm what is causing disrupted embryo maturation in *F. vesca*, it is an interesting result even if it is not attributable to *FvFT1*.

References

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science *309*, 1052–1056.

Aharoni, A., and O'Connell, A.P. (2002). Gene expression analysis of strawberry achene and receptacle maturation using DNA microarrays. J. Exp. Bot. *53*, 2073–2087.

Alexa, A., and Rahnenfuhrer, J. (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0.

An, C.-I., Ichihashi, Y., Peng, J., Sinha, N.R., and Hagiwara, N. (2016). Transcriptome Dynamics and Potential Roles of Sox6 in the Postnatal Heart. PLOS ONE *11*, e0166574.

An, H., Roussot, C., Suárez-López, P., Corbesier, L., Vincent, C., Piñeiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., et al. (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development *131*, 3615–3626.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796–815.

Bernier, G. (1988). The Control of Floral Evocation and Morphogenesis. Annual Review of Plant Physiology and Plant Molecular Biology *39*, 175–219.

Bernier, G., Havelange, A., Houssa, C., Petitjean, A., and Lejeune, P. (1993). Physiological Signals That Induce Flowering. Plant Cell *5*, 1147–1155.

Bewley, J.D., and Black, M. (1994). Seeds: physiology of development and germination. Seeds: Physiology of Development and Germination.

Blackman, B.K., Strasburg, J.L., Raduski, A.R., Michaels, S.D., and Rieseberg, L.H. (2010). The role of recently derived FT paralogs in sunflower domestication. Curr. Biol. *20*, 629–635.

Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H., and Nilsson, O. (2006). CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science *312*, 1040–1043.

Borner, R., Kampmann, G., Chandler, J., Gleissner, R., Wisman, E., Apel, K., and Melzer, S. (2000). A MADS domain gene involved in the transition to flowering in Arabidopsis. Plant J. *24*, 591–599.

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

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. Development *112*, 1–20.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M., and Smyth, D.R. (1993). Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. Development *119*, 721–743.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in Arabidopsis. Science *275*, 80–83.

Bridges, C.B. and Morgan, T.H. (1923). The Third-Chromosome Group of Mutant Characters of Drosophila melanogaster (Washington, DC: Carnegie Institution of Washington).

Brown, R.L., Kazan, K., McGrath, K.C., Maclean, D.J., and Manners, J.M. (2003). A Role for the GCC-Box in Jasmonate-Mediated Activation of the PDF1.2 Gene of Arabidopsis. Plant Physiology *132*, 1020–1032.

Brumbarova, T., and Ivanov, R. (2014). Perls Staining for Histochemical Detection of Iron in Plant Samples. BIO-PROTOCOL *4*.

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. Molecular Plant *11*, 230–233.

Chailakhyan, M. (1936). New facts in support of the hormonal theory of plant development. C. R. Acad. Sci. URSS. *13*, 79–83.

Chatterjee, M., Bermudez-Lozano, C.L., Clancy, M.A., Davis, T.M., and Folta, K.M. (2011). A strawberry KNOX gene regulates leaf, flower and meristem architecture. PLoS ONE *6*, e24752.

Chen, M., MacGregor, D.R., Dave, A., Florance, H., Moore, K., Paszkiewicz, K., Smirnoff, N., Graham, I.A., and Penfield, S. (2014). Maternal temperature history activates Flowering Locus T in fruits to control progeny dormancy according to time of year. PNAS *111*, 18787–18792.

Chettoor, A.M., Givan, S.A., Cole, R.A., Coker, C.T., Unger-Wallace, E., Vejlupkova, Z., Vollbrecht, E., Fowler, J.E., and Evans, M.M. (2014). Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. Genome Biology *15*, 414.

Coen, E.S., and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature *353*, 31.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics *2008*, 619832.

Conley, T.R., Park, S.C., Kwon, H.B., Peng, H.P., and Shih, M.C. (1994). Characterization of cis-acting elements in light regulation of the nuclear gene encoding the A subunit of chloroplast isozymes of glyceraldehyde-3-phosphate dehydrogenase from Arabidopsis thaliana. Mol. Cell. Biol. *14*, 2525–2533.

Corbesier, L., and Coupland, G. (2005). Photoperiodic flowering of Arabidopsis: integrating genetic and physiological approaches to characterization of the floral stimulus. Plant, Cell & Environment *28*, 54–66.

Corbesier, L., Lejeune, P., and Bernier, G. (1998). The role of carbohydrates in the induction of flowering in Arabidopsis thaliana: comparison between the wild type and a starchless mutant. Planta *206*, 131–137.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science *316*, 1030–1033.

Crepet, W.L., and Niklas, K.J. (2009). Darwin's second "abominable mystery": Why are there so many angiosperm species? Am. J. Bot. *96*, 366–381.

Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M., and Carrington, J.C. (2010). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc. Natl. Acad. Sci. U.S.A. *107*, 466–471.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. *133*, 462–469.

Darrow, G.M. (1966). The strawberry: history, breeding, and physiology (Holt, Rinehart and Winston).

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr. Biol. *14*, 1935–1940.

Dorcey, E., Urbez, C., Blázquez, M.A., Carbonell, J., and Perez-Amador, M.A. (2009). Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in Arabidopsis. Plant J. *58*, 318–332.

Dreher, T., and Poovaiah, B.W. (1982). Changes in auxin content during development in strawberry fruits. J. Plant Growth Regul. *1*, 267–276.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. Cell *65*, 991–1002.

Dumas, C., and Mogensen, H. (1993). Gametes and Fertilization: Maize as a Model System for Experimental Embryogenesis in Flowering Plants. Plant Cell *5*, 1337–1348.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res *19*, 1349.

Endo, T., Shimada, T., Fujii, H., Kobayashi, Y., Araki, T., and Omura, M. (2005). Ectopic expression of an FT homolog from citrus confers an early flowering phenotype on trifoliate orange (Poncirus trifoliata L. Raf.). Transgenic Res. *14*, 703–712.

Estrada-Johnson, E., Csukasi, F., Pizarro, C.M., Vallarino, J.G., Kiryakova, Y., Vioque, A., Brumos, J., Medina-Escobar, N., Botella, M.A., Alonso, J.M., et al. (2017). Transcriptomic Analysis in Strawberry Fruits Reveals Active Auxin Biosynthesis and Signaling in the Ripe Receptacle. Front Plant Sci *8*, 889.

Fait, A., Hanhineva, K., Beleggia, R., Dai, N., Rogachev, I., Nikiforova, V.J., Fernie, A.R., and Aharoni, A. (2008). Reconfiguration of the Achene and Receptacle Metabolic Networks during Strawberry Fruit Development. Plant Physiology *148*, 730–750.

FAO (2009). How to feed the world in 2050, high-level expert forum. (Rome: Food and Agriculture Organization of the United Nations).

Farrow, S.C., and Facchini, P.J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. Front Plant Sci *5*.

Ferrándiz, C., Pelaz, S., and Yanofsky, M.F. (1999). Control of carpel and fruit development in Arabidopsis. Annu. Rev. Biochem. *68*, 321–354.

Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M.F. (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. Development *127*, 725–734.

Forreiter, C., and Apel, K. (1993). Light-independent and light-dependent protochlorophyllide-reducing activities and two distinct NADPH-protochlorophyllide oxidoreductase polypeptides in mountain pine (<Emphasis Type="Italic">Pinus mugo</Emphasis>). Planta *190*, 536–545.

Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G., and Putterill, J. (1999). GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. EMBO J. *18*, 4679–4688.

Friedman, W.E. (2009). The meaning of Darwin's "abominable mystery." Am. J. Bot. *96*, 5–21.

Friedman, W.E., Moore, R.C., and Purugganan, M.D. (2004). The evolution of plant development. Am. J. Bot. *91*, 1726–1741.

Fuentes, S., Ljung, K., Sorefan, K., Alvey, E., Harberd, N.P., and Østergaard, L. (2012). Fruit Growth in Arabidopsis Occurs via DELLA-Dependent and DELLA-Independent Gibberellin Responses[W][OA]. Plant Cell *24*, 3982–3996.

Fujisawa, M., Shima, Y., Nakagawa, H., Kitagawa, M., Kimbara, J., Nakano, T., Kasumi, T., and Ito, Y. (2014). Transcriptional Regulation of Fruit Ripening by Tomato FRUITFULL Homologs and Associated MADS Box Proteins. The Plant Cell *26*, 89–101.

Gaertner J. (1788). De fructibus et seminibus plantarum. (Stuttgart: Academiae Carolinae).

Galvão, V.C., Horrer, D., Küttner, F., and Schmid, M. (2012). Spatial control of flowering by DELLA proteins in Arabidopsis thaliana. Development *139*, 4072–4082.

García-Gago, J.A., Posé, S., Muñoz-Blanco, J., Quesada, M.A., and Mercado, J.A. (2009). The polygalacturonase FaPG1 gene plays a key role in strawberry fruit softening. Plant Signal Behav *4*, 766–768.

Garner, W.W., and Allard, H.A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. Mon. Wea. Rev. *48*, 415–415.

Gasic, K., Hernandez, A., and Korban, S.S. (2004). RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. Plant Mol Biol Rep *22*, 437–438.

Gasser, C., and Robinson-Beers, K. (1993). Pistil Development. Plant Cell *5*, 1231–1239.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A Developmental Perspective. Plant Cell *5*, 1439–1451.

Goetz, M., Vivian-Smith, A., Johnson, S.D., and Koltunow, A.M. (2006). AUXIN RESPONSE FACTOR8 Is a Negative Regulator of Fruit Initiation in Arabidopsis. Plant Cell *18*, 1873–1886.

Griffiths, S., Dunford, R.P., Coupland, G., and Laurie, D.A. (2003). The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis. Plant Physiol *131*, 1855–1867.

Grusak, M.A. (1994). Iron Transport to Developing Ovules of Pisum sativum (I. Seed Import Characteristics and Phloem Iron-Loading Capacity of Source Regions). Plant Physiology *104*, 649–655.

Gupta, R., and Chakrabarty, S.K. (2013). Gibberellic acid in plant. Plant Signal Behav *8*.

Gustafson, F.G. (1960). Influence of Gibberellic Acid on Setting and Development of Fruits in Tomato 12. Plant Physiol *35*, 521–523.

Hanano, S., and Goto, K. (2011). Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. Plant Cell *23*, 3172–3184.

Hartwig, B., James, G.V., Konrad, K., Schneeberger, K., and Turck, F. (2012). Fast Isogenic Mapping-by-Sequencing of Ethyl Methanesulfonate-Induced Mutant Bulks. Plant Physiology *160*, 591–600.

Hawkins, C., Caruana, J., Li, J., Zawora, C., Darwish, O., Wu, J., Alkharouf, N., and Liu, Z. (2017). An eFP browser for visualizing strawberry fruit and flower transcriptomes. Hortic Res *4*, 17029.

Heijmans, K., Morel, P., and Vandenbussche, M. (2012). MADS-box Genes and Floral Development: the Dark Side. J Exp Bot *63*, 5397–5404.

Hill, J.P., and Lord, E.M. (1989). Floral development in Arabidopsis thaliana: a comparison of the wild type and the homeotic pistillata mutant. Can. J. Bot. *67*, 2922–2936.

Hollender, C.A. (2012). Molecular and Genetic Analysis of Flower Development in Arabidopsis thaliana and the Diploid Strawberry, Fragaria vesca. PhD thesis, University of Maryland, College Park, MD, USA.

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.

Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L., and Yu, H. (2014). Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. Nat Commun *5*, 4601.

Hsu, C.-Y., Liu, Y., Luthe, D.S., and Yuceer, C. (2006). Poplar FT2 Shortens the Juvenile Phase and Promotes Seasonal Flowering. The Plant Cell *18*, 1846–1861.

Huala, E., and Sussex, I.M. (1992). LEAFY Interacts with Floral Homeotic Genes to Regulate Arabidopsis Floral Development. Plant Cell *4*, 901–913.

Huang, J., Vendramin, S., Shi, L., and McGinnis, K.M. (2017). Construction and Optimization of a Large Gene Coexpression Network in Maize Using RNA-Seq Data. Plant Physiol. *175*, 568–583.

Huq, E., Tepperman, J.M., and Quail, P.H. (2000). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. Proc Natl Acad Sci U S A *97*, 9789–9794.

Iancu, O.D., Kawane, S., Bottomly, D., Searles, R., Hitzemann, R., and McWeeney, S. (2012). Utilizing RNA-Seq data for de novo coexpression network inference. Bioinformatics *28*, 1592–1597.

Imlau, A., Truernit, E., and Sauer, N. (1999). Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. Plant Cell *11*, 309–322.

Irish, V. (2017). The ABC model of floral development. Current Biology *27*, R887–R890.

Irish, V.F., and Sussex, I.M. (1990). Function of the apetala-1 gene during Arabidopsis floral development. The Plant Cell *2*, 741–753.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell *68*, 683–697.

Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. Nat. Rev. Mol. Cell Biol. *14*, 787–802.

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 *25*, 1960–1978.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer FT. Science *286*, 1962–1965.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci. *7*, 193–195.

Kieffer, M., and Davies, B. (2001). Developmental programmes in floral organ formation. Seminars in Cell & Developmental Biology *12*, 373–380.

Kimura, S., and Sinha, N. (2008). Tomato (Solanum lycopersicum): A Model Fruit-Bearing Crop. CSH Protoc *2008*, pdb.emo105.

King, R.W., and Zeevaart, J.A. (1973). Floral stimulus movement in perilla and flower inhibition caused by noninduced leaves. Plant Physiol. *51*, 727–738.

King, R.W., Evans, L.T. and Wardlaw, I.F. (1968). Translocation of the floral stimulus in Pharbitis nil in relation to that of assimilates. Z. Pflanzenphysiol. *59*, 377–388.

Kinoshita, T., Ono, N., Hayashi, Y., Morimoto, S., Nakamura, S., Soda, M., Kato, Y., Ohnishi, M., Nakano, T., Inoue, S., et al. (2011). FLOWERING LOCUS T regulates stomatal opening. Curr. Biol. *21*, 1232–1238.

Knott, J.E. (1934). Effect of a localized photoperiod on spinach. Proc. Am. Soc. Hort. Sci. *31*, 152–154.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. Science *286*, 1960–1962.

Koornneef, M., Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C.J., and Peeters, A.J. (1998). Genetic interactions among late-flowering mutants of Arabidopsis. Genetics *148*, 885–892.

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.

Kotoda, N., Hayashi, H., Suzuki, M., Igarashi, M., Hatsuyama, Y., Kidou, S.-I., Igasaki, T., Nishiguchi, M., Yano, K., Shimizu, T., et al. (2010). Molecular characterization of FLOWERING LOCUS T-like genes of apple (Malus x domestica Borkh.). Plant Cell Physiol. *51*, 561–575.

Kunst, L., Klenz, J.E., Martinez-Zapater, J., and Haughn, G.W. (1989). AP2 Gene Determines the Identity of Perianth Organs in Flowers of Arabidopsis thaliana. Plant Cell *1*, 1195–1208.

Kurokura, T., Samad, S., Koskela, E., Mouhu, K., and Hytönen, T. (2017). Fragaria vesca CONSTANS controls photoperiodic flowering and vegetative development. J. Exp. Bot. *68*, 4839–4850.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lee, H., Suh, S.S., Park, E., Cho, E., Ahn, J.H., Kim, S.G., Lee, J.S., Kwon, Y.M., and Lee, I. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev. *14*, 2366–2376.

Levin, J.Z., and Meyerowitz, E.M. (1995). UFO: an Arabidopsis gene involved in both floral meristem and floral organ development. Plant Cell *7*, 529–548.

Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. Plant J. *90*, 164–176.

Lifschitz, E., Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., Amsellem, Z., Alvarez, J.P., and Eshed, Y. (2006). The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. PNAS *103*, 6398–6403.

Lin, M.-K., Belanger, H., Lee, Y.-J., Varkonyi-Gasic, E., Taoka, K.-I., Miura, E., Xoconostle-Cázares, B., Gendler, K., Jorgensen, R.A., Phinney, B., et al. (2007). FLOWERING LOCUS T Protein May Act as the Long-Distance Florigenic Signal in the Cucurbits. Plant Cell *19*, 1488–1506.

Lin-Wang, K., McGhie, T.K., Wang, M., Liu, Y., Warren, B., Storey, R., Espley, R.V., and Allan, A.C. (2014). Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca). Front Plant Sci *5*.

Liu, L., Liu, C., Hou, X., Xi, W., Shen, L., Tao, Z., Wang, Y., and Yu, H. (2012). FTIP1 Is an Essential Regulator Required for Florigen Transport. PLOS Biology *10*, e1001313.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., and Wu, S. (2013). Understanding and enhancement of internal clustering validation measures. IEEE Trans Cybern *43*, 982–994.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Mandel, M.A., and Yanofsky, M.F. (1995). A gene triggering flower formation in Arabidopsis. Nature *377*, 522–524.

Mascarenhas, J. (1993). Molecular Mechanisms of Pollen Tube Growth and Differentiation. Plant Cell *5*, 1303–1314.

Millonig, G. (1964). Study on the factors which influence preservation of fine structure. In Symposium on Electrong Microscopy, (Rome: Consiglio Nazionale delle Ricerche), p. 347.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning *52*, 91–118.

Mouhu, K., Kurokura, T., Koskela, E.A., Albert, V.A., Elomaa, P., and Hytönen, T. (2013). The Fragaria vesca homolog of suppressor of overexpression of constans1 represses flowering and promotes vegetative growth. Plant Cell *25*, 3296–3310.

Mouradov, A., Cremer, F., and Coupland, G. (2002). Control of Flowering Time. Plant Cell *14*, s111–s130.

Navarro, C., Abelenda, J.A., Cruz-Oró, E., Cuéllar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature *478*, 119–122.

Nguyen, K.T., Park, J., Park, E., Lee, I., and Choi, G. (2015). The Arabidopsis RING Domain Protein BOI Inhibits Flowering via CO-dependent and CO-independent Mechanisms. Mol Plant *8*, 1725–1736.

Nguyen-Legros, J., Bizot, J., Bolesse, M., and Pulicani, J.P. (1980). ["Diaminobenzidine black" as a new histochemical demonstration of exogenous iron (author's transl)]. Histochemistry *66*, 239–244.

Nishikawa, F., Endo, T., Shimada, T., Fujii, H., Shimizu, T., Omura, M., and Ikoma, Y. (2007). Increased CiFT abundance in the stem correlates with floral induction by low temperature in Satsuma mandarin (Citrus unshiu Marc.). J. Exp. Bot. *58*, 3915–3927.

Nitsch, J.P. (1950). Growth and Morphogenesis of the Strawberry as Related to Auxin. American Journal of Botany *37*, 211–215.

Nitsch, J. (1970). Hormonal factors in growth and development. In The Biochemistry of Fruits and Their Products, (London: Academic Press), pp. 427–472.

Onouchi, H., Igeño, M.I., Périlleux, C., Graves, K., and Coupland, G. (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. Plant Cell *12*, 885–900.

Palmer, J.D., Soltis, D.E., and Chase, M.W. (2004). The plant tree of life: an overview and some points of view. Am. J. Bot. *91*, 1437–1445.

Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A., and Nam, H.G. (1999). Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science *285*, 1579–1582.

Pattison, R.J., Csukasi, F., Zheng, Y., Fei, Z., Knaap, E. van der, and Catalá, C. (2015). Comprehensive Tissue-Specific Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit Development. Plant Physiology *168*, 1684–1701.

Peet, M.M., Sato, S., and Gardner, R.G. (1998). Comparing heat stress effects on male-fertile and male-sterile tomatoes. Plant, Cell & Environment *21*, 225–231.

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.

Penfield, S. (2017). Seed dormancy and germination. Current Biology *27*, R874–R878.

Pin, P.A., and Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development. Plant Cell Environ. *35*, 1742–1755.

Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., and Lifschitz, E. (2001). Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. Plant Cell *13*, 2687–2702.

Pogson, B.J., and Albrecht, V. (2011). Genetic Dissection of Chloroplast Biogenesis and Development: An Overview. Plant Physiology *155*, 1545–1551.

Porri, A., Torti, S., Romera-Branchat, M., and Coupland, G. (2012). Spatially distinct regulatory roles for gibberellins in the promotion of flowering of Arabidopsis under long photoperiods. Development *139*, 2198–2209.

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inzé, D., Mueller-Roeber, B., and Vandepoele, K. (2015). PLAZA 3.0: an access point for plant comparative genomics. Nucleic Acids Res *43*, D974–D981.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Reeves, P.H., and Coupland, G. (2001). Analysis of Flowering Time Control in Arabidopsis by Comparison of Double and Triple Mutants. Plant Physiology *126*, 1085–1091.

Ridley, H.N. (1930). The dispersal of plants throughout the world. (Ashford, UK: L. Reeve & Co, Ltd.).

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Robson, F., Costa, M.M., Hepworth, S.R., Vizir, I., Piñeiro, M., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J. *28*, 619–631.

Roeder, A.H.K., and Yanofsky, M.F. (2006). Fruit Development in Arabidopsis. Arabidopsis Book *4*.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C., and Pounds, J.A. (2003). Fingerprints of global warming on wild animals and plants. Nature *421*, 57–60.

Roschzttardtz, H., Conéjéro, G., Curie, C., and Mari, S. (2009). Identification of the endodermal vacuole as the iron storage compartment in the Arabidopsis embryo. Plant Physiol. *151*, 1329–1338.

Rowland, L.J., Alkharouf, N., Darwish, O., Ogden, E.L., Polashock, J.J., Bassil, N.V., and Main, D. (2012). Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. BMC Plant Biology *12*, 46.

Ruiz-García, L., Madueño, F., Wilkinson, M., Haughn, G., Salinas, J., and Martínez-Zapater, J.M. (1997). Different roles of flowering-time genes in the activation of floral initiation genes in Arabidopsis. Plant Cell *9*, 1921–1934.

Sánchez-Sevilla, J.F., Vallarino, J.G., Osorio, S., Bombarely, A., Posé, D., Merchante, C., Botella, M.A., Amaya, I., and Valpuesta, V. (2017). Gene expression atlas of fruit ripening

and transcriptome assembly from RNA-seq data in octoploid strawberry (Fragaria × ananassa). Scientific Reports *7*, 13737.

Sastry, K.K.S., and Muir, R.M. (1963). Gibberellin: Effect on Diffusible Auxin in Fruit Development. Science *140*, 494–495.

Sauer, M., Robert, S., and Kleine-Vehn, J. (2013). Auxin: simply complicated. J Exp Bot *64*, 2565–2577.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.-E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods *6*, 550–551.

Schoefs, B., and Franck, F. (2003). Protochlorophyllide Reduction: Mechanisms and Evolution¶. Photochemistry and Photobiology *78*.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lönnig, W.E., Saedler, H., and Sommer, H. (1992). Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J *11*, 251–263.

Sekhon, R.S., Briskine, R., Hirsch, C.N., Myers, C.L., Springer, N.M., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2013). Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE *8*, e61005.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M., and Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. Front Plant Sci *7*, 444.

Serrani, J.C., Ruiz-Rivero, O., Fos, M., and García-Martínez, J.L. (2008). Auxin-induced fruit-set in tomato is mediated in part by gibberellins. Plant J. *56*, 922–934.

Sessions, A., Weigel, D., and Yanofsky, M.F. (1999). The Arabidopsis thaliana MERISTEM LAYER 1 promoter specifies epidermal expression in meristems and young primordia. Plant J. *20*, 259–263.

Seymour, G.B., Østergaard, L., Chapman, N.H., Knapp, S., and Martin, C. (2013). Fruit Development and Ripening. Annual Review of Plant Biology *64*, 219–241.

Shannon, C.P., Chen, V., Takhar, M., Hollander, Z., Balshaw, R., McManus, B.M., Tebbutt, S.J., Sin, D.D., and Ng, R.T. (2016). SABRE: a method for assessing the stability of gene modules in complex tissues and subject populations. BMC Bioinformatics *17*.

Shen, L, and Sinai, M (2013). GeneOverlap: Test and visualize gene overlaps. R package version 1.14.0.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., et al. (2011). The genome of woodland strawberry (Fragaria vesca). Nat. Genet. *43*, 109–116.

Simpson, G.G., and Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? Science *296*, 285–289.

Singer, G.A.C., Lloyd, A.T., Huminiecki, L.B., and Wolfe, K.H. (2005). Clusters of Co-expressed Genes in Mammalian Genomes Are Conserved by Natural Selection. Mol Biol Evol *22*, 767–775.

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.

Soltis, P.S., and Soltis, D.E. (2004). The origin and diversification of angiosperms. Am. J. Bot. *91*, 1614–1626.

Sommer, H., Beltrán, J.P., Huijser, P., Pape, H., Lönnig, W.E., Saedler, H., and Schwarz-Sommer, Z. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J. *9*, 605–613.

Spirin, V., and Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. U.S.A. *100*, 12123–12128.

Spurr, A.R. (1969). A low-viscosity epoxy resin embedding medium for electron microscopy. J. Ultrastruct. Res. *26*, 31–43.

Strickler, S.R., Bombarely, A., and Mueller, L.A. (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. Am. J. Bot. *99*, 257–266.

Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature *410*, 1116–1120.

Takada, S., and Goto, K. (2003). Terminal flower2, an Arabidopsis homolog of heterochromatin protein1, counteracts the activation of flowering locus T by constans in the vascular tissues of leaves to regulate flowering time. Plant Cell *15*, 2856–2865.

Takahashi, N., Kajihara, T., Okamura, C., Kim, Y., Katagiri, Y., Okushima, Y., Matsunaga, S., Hwang, I., and Umeda, M. (2013). Cytokinins control endocycle onset by promoting the expression of an APC/C activator in Arabidopsis roots. Curr. Biol. *23*, 1812–1817.

Thompson, P.A. (1969). The Effect of Applied Growth Substances on Development of the Strawberry Fruit.: II. Interactions of Auxins and Gibberellins. J Exp Bot *20*, 629–647.

Thorne, R.F. (1992). Classification and geography of the flowering plants. Bot. Rev *58*, 225–327.

Tiwari, S.B., Shen, Y., Chang, H.-C., Hou, Y., Harris, A., Ma, S.F., McPartland, M., Hymus, G.J., Adam, L., Marion, C., et al. (2010). The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. New Phytol. *187*, 57–66.

Tjaden, G., Edwards, J.W., and Coruzzi, G.M. (1995). Cis elements and trans-acting factors affecting regulation of a nonphotosynthetic light-regulated gene for chloroplast glutamine synthetase. Plant Physiol. *108*, 1109–1117.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science *303*, 1003–1006.

Varkonyi-Gasic, E., Moss, S.M.A., Voogd, C., Wang, T., Putterill, J., and Hellens, R.P. (2013). Homologs of FT, CEN and FD respond to developmental and environmental signals affecting growth and flowering in the perennial vine kiwifruit. New Phytol *198*, 732–746.

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.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. *131*, 281–285.

Wang, H., Pan, J., Li, Y., Lou, D., Hu, Y., and Yu, D. (2016). The DELLA-CONSTANS Transcription Factor Cascade Integrates Gibberellic Acid and Photoperiod Signaling to Regulate Flowering. Plant Physiol. *172*, 479–488.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M. (1992). LEAFY controls floral meristem identity in Arabidopsis. Cell *69*, 843–859.

Wigge, P.A. (2011). FT, A Mobile Developmental Signal in Plants. Current Biology *21*, R374–R378.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. Science *309*, 1056–1059.

Wolfe, C.J., Kohane, I.S., and Butte, A.J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics *6*, 227.

Wu, S., and Gallagher, K.L. (2014). The movement of the non-cell-autonomous transcription factor, SHORT-ROOT relies on the endomembrane system. Plant J. *80*, 396–409.

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., and Ma, H. (2017). Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. Mol Biol Evol *34*, 262–281.

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., and Chen, Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biology *14*, 327.

Xu, F., Li, T., Xu, P.-B., Li, L., Du, S.-S., Lian, H.-L., and Yang, H.-Q. (2016). DELLA proteins physically interact with CONSTANS to regulate flowering under long days in Arabidopsis. FEBS Lett. *590*, 541–549.

Yamamoto, Y.Y., Yoshioka, Y., Hyakumachi, M., Maruyama, K., Yamaguchi-Shinozaki, K., Tokizawa, M., and Koyama, H. (2011). Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data. BMC Plant Biol *11*, 39.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., and Meyerowitz, E.M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature *346*, 35–39.

Yanovsky, M.J., and Kay, S.A. (2002). Molecular basis of seasonal time measurement in Arabidopsis. Nature *419*, 308–312.

Zeevaart, J.A.D. (1976). Physiology of Flower Formation. Annual Review of Plant Physiology *27*, 321–348.

Zeevaart, J. (1985). Perilla. In CRC Handbook of Flowering. (Boca Raton, FL: CRC Press).

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.