ABSTRACT

Title of Dissertation:	SHAPING THE DICOT FRUIT: MOLECULAR AND GENOMIC APPROACHES TO FRUIT DEVELOPMENT
	Charles Lynn Presson Hawkins, Doctor of Philosophy, 2016
Dissertation Directed By:	Professor Zhongchi Liu Department of Cell Biology and Molecular Genetics

The fruit is one of the most complex and important structures produced by flowering plants, and understanding the development and maturation process of fruits in different angiosperm species with diverse fruit structures is of immense interest. In the work presented here, molecular genetics and genomic analysis are used to explore the processes that form the fruit in two species: The model organism *Arabidopsis* and the diploid strawberry *Fragaria vesca*.

One important basic question concerns the molecular genetic basis of fruit patterning. A long-standing model of *Arabidopsis* fruit (the gynoecium) patterning holds that auxin produced at the apex diffuses downward, forming a gradient that provides apical-basal positional information to specify different tissue types along the gynoecium's length. The proposed gradient, however, has never been observed and the model appears inconsistent with a number of observations. I present a new, alternative model, wherein auxin acts to establish the adaxial-abaxial domains of the carpel primordia, which then ensures proper development of the final gynoecium.

A second project utilizes genomics to identify genes that regulate fruit color by analyzing the genome sequences of *Fragaria vesca*, a species of wild strawberry. Shared and distinct SNPs among three *F. vesca* accessions were identified, providing a foundation for locating candidate mutations underlying phenotypic variations among different *F. vesca* accessions. Through systematic analysis of relevant SNP variants, a candidate SNP in *FveMYB10* was identified that may underlie the fruit color in the yellow-fruited accessions, which was subsequently confirmed by functional assays.

Our lab has previously generated extensive RNA-sequencing data that depict genome-scale gene expression profiles in *F. vesca* fruit and flower tissues at different developmental stages. To enhance the accessibility of this dataset, the web-based eFP software was adapted for this dataset, allowing visualization of gene expression in any tissues by user-initiated queries.

Together, this thesis work proposes a well-supported new model of fruit patterning in *Arabidopsis* and provides further resources for *F. vesca*, including genome-wide variant lists and the ability to visualize gene expression. This work will facilitate future work linking traits of economic importance to specific genes and gaining novel insights into fruit patterning and development.

SHAPING THE DICOT FRUIT: MOLECULAR AND GENOMIC APPROACHES TO FRUIT DEVELOPMENT

by

Charles Lynn Presson Hawkins

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committe:
Professor Zhongchi Liu, Chair
Professor Caren Chang
Professor Todd Cooke
Professor Eric Haag
Professor Angus Murphy

Acknowledgements

I would like to thank my advisor, Dr. Zhongchi Liu, for her tireless support, invaluable guidance, and for putting up with me for the past six years. She kept me going, kept me on track, and got me through it. I couldn't have done it without her.

I thank my committee for their support and their guidance. Their advice helped shape this dissertation and made it what it is.

For supporting the work that lead to Chapter II, I would like to thank University of Maryland CMNS Dean's Fellowship and MOCB-CA Summer Fellowship to C.H. and NSF grant (MCB0951460) to Z. L. I thank the manuscript reviewers who provided insightful and valuable comments and suggestions on the manuscript that lead to this chapter. I would like also to thank my co-author Zhongchi Liu for her contributions to the ideas behind this chapter and the words to express them.

The work in Chapter III was a collaborative effort, and I would like to thank my fellow collaborators Julie Caruana, Erin Schiksnis, and Zhongchi Liu for their many contributions to the project and the manuscript that lead to the chapter. I thank Ms. Kasuni Wattarantenne for assistance with the PCR in Figure III-4, Dr. Chunying Kang and Ms. Rachel Shahan for comments on the manuscript, and Mr. Wanpeng Wang for discussions of MYB protein structure and function. The work in Chapter III was supported by NSF grants 0923913 and 1444987 to Z.L as well as the Maryland MAES Hatch Project (MD-CBMG-0738) to Z. L. The work also received support from the University of Maryland CBMG Postdoctoral Merit Fellowship (to J. C.) and the MOCB-CA Summer Fellowship (to C. H.).

I thank Chunying Kang for providing the updated transcript mapping data for the flower and fruit transcriptomes that were used directly in creating the eFP-based visualization tool.

Finally, I would like to thank my friends, my family, and my girlfriend for their tireless love, support and encouragement. They got me through the worst times and out the other side.

Table of Contents

Chapter I: Introduction

Fruit Overview

The fruit is the primary reproductive structure of angiosperm plants and represents a key evolutionary innovation in angiosperm. To date, angiosperm is highly successful, diverse and economically important; with more than 223,000 known extant plant species belong to the clade (Scotland and Wortley, 2003). The fruit is key to reproduction of angiosperms both in the wild and under cultivation, and is an economic powerhouse in its own right: The structure and the seeds within are the source not only of fresh fruits, but also the edible parts of corn, grains, and nuts. . In addition, fruit is one of the most morphologically and developmentally-complex structures produced by plants, yet still follows the same basic motifs of plant development illustrated below. For these reasons, understanding the developmental process by which flowers and fruit are formed and how this process determines the final shape, structure, and properties of the fruit is an extremely important field of study from both an economic and a basic science perspective.

Most aerial plant structures follow the pattern of an axis (such as a stem) that grows from the tip, with lateral organ primordia (such as leaves) being initiated at the growing tip. Each of these primordia develops into a lateral organ, which may or may not be externally visible and which has at its axil (the area where it meets the stem from which it initiated) a bud for a new axis (Fig. I-1). Flowers and fruit are no exception to this pattern; the flower as a whole represents an axis surrounded by four whorls of lateral organs, these being the sepals, petals, stamens, and carpels (Coen and Meyerowitz, 1991;

Honma and Goto, 2001; Pelaz et al., 2001; Scutt et al., 2006). The last of these, the carpels, contain the ovules and develop into the (botanical) fruit of the plant upon fertilization.

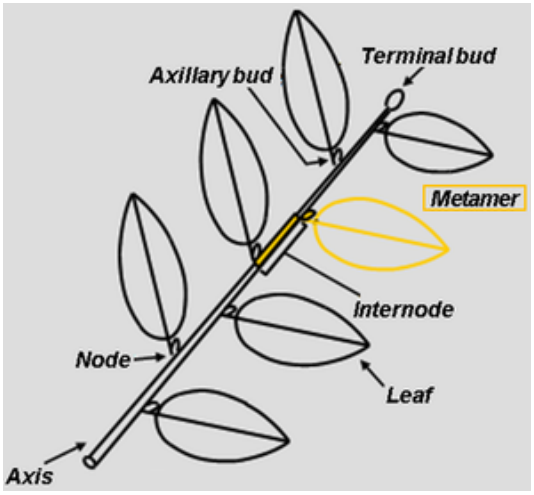

Figure I-1. Leaf axis

Drawing by D. Barthelemy: (www.greenlqb:cirqd:fr)

Growth, via cell division and elongation occurs at the tip, where leaf buds are initiated in a regular, species-dependent pattern. The axil of each leaf is the point where it meets the parent axis, and nestled in each of these is an axial bud that can grow into a new axis (though most will never do so).

Leaf axes and polarity

The structure of lateral organs may be analyzed with reference to three axes of development, most easily illustrated in the case of the leaf (Fig. I-2). These are the apical-basal axis, with the basal direction pointing to the axil of the leaf where it meets the stem and the apical direction pointing to the leaf tip; the medial-lateral axis, with the

medial direction pointing in toward the midrib and the lateral direction pointing out toward the periphery on either side; and the adaxial-abaxial (or dorsal-ventral) axis, with the adaxial direction pointing toward the upper surface of the leaf, and the abaxial direction pointing toward the bottom surface of the leaf.

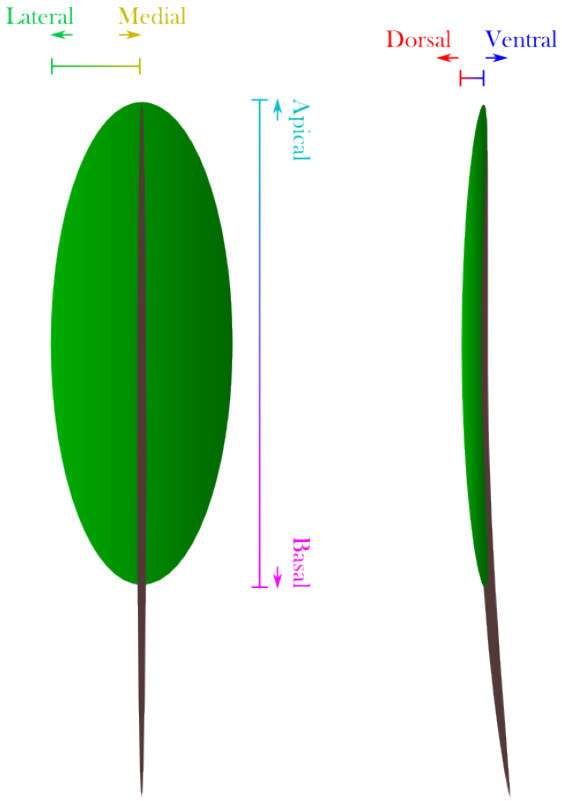

Figure I-2: Axes of organ development

The three axes used to describe form and growth of leaves and other lateral organs. The apical-basal axis runs from the distal tip of the leaf to its base where it meets the stem. The dorsal-ventral (or adaxial-abaxial) axis runs from the upper surface of the leaf to the lower surface. The medial-lateral axis runs from the mid vein of the leaf toward the edge of the leaf blade.

Soon after a leaf primordium is initiated, one of the first signs of patterning appears in the specification of the adaxial (dorsal; AD) and abaxial (ventral; AB) halves of the leaf, each marked by its own constellation of genes, many of which on both sides repress the genes that specify the opposite half, ensuring a clean boundary. This early patterning is believed to happen in response to a signal generated at the apex or shoot apical meristem (Sussex, 1951; reviewed in Husbands et al., 2009). If the path from shoot apex to primordium is blocked, such as by a cut made directly above the incipient primordium, the adaxial-abaxial patterning of the leaf will be disrupted. The identity of this signal is still unknown but auxin remains a possibility (Husbands et al., 2009).

The genes expressed in the abaxial domain include genes in the YABBY and KANADI families, along with the auxin response factor ETTIN (ETT), all transcription factors, while in the adaxial domain members of the HD-ZIPIII family of transcription factors are expressed (reviewed in Kidner and Timmermans, 2007; Liu et al., 2012). Disruption of genes on either side frequently results in no laminar outgrowth and a radially symmetrical, bristle-like leaf (Waites and Hudson, 1995; e.g. Eshed, 2004; Pekker, 2005; Stahle et al., 2009). It was proposed based on these observations that juxtaposition of the AB and AD domains is necessary for lamina out growth that lead to the formation of the flat leaf blade, possibly the result of signaling across the adaxial-abaxial boundary (Husbands et al., 2009).

The Model Systems

Two plants species are the focus of this dissertation. The first, *Arabidopsis thaliana*, has long been the most-studied model organism for higher plants (Somerville and Koornneef, 2002). It is a member of the Brassicaceae (wild mustard) family, a clade it shares with cabbage, broccoli, cauliflower, Brussels sprouts, radishes, and numerous other vegetables. It is desirable as a model organism for its relatively uncomplicated architecture, fecundity (as many as 10,000 seeds per plant), small stature (20 – 25 cm), ease of care, short generation time (six weeks from seed to seed), and the availability of a wide array of molecular genetic tools.

The second plant is the wild strawberry *Fragaria vesca*. A member of the *Rosaceae* family, it is a relative of raspberries and blackberries, of apples and pears, and of peaches, plums, cherries, and almonds. Work has been ongoing for the past several years to develop this wild strawberry into a model organism for the study of both the cultivated strawberry and the *Rosaceae* family. The species has several desirable characteristics in this regard: Like *Arabidopsis*, it is fecund (200 seeds per fruit on average), small in stature, and has a short generation time (3.5 months), particularly in comparison to its larger, arborescent relatives such as pomes and stone fruits. In addition, *F. vesca* is diploid, making it far more tractable for genetic analysis than the octoploid *Fragaria x ananassa* commonly sold in supermarkets. Finally, the genome of *F. vesca* has been sequenced (Shulaev et al., 2011a), making it possible to conduct various genomic studies including genome-wide variant comparisons detailed in Chapter III.

In addition to its economic importance, the *Rosaceae* family is interesting biologically because of the diversity of fruit morphologies among the different genera. The fruits differ greatly in shape and size, but also in structure. In *Fragaria*, for example, the fleshy, edible fruit is derived from the receptacle, the tip of the stem-derived structure.

The structure of the *Arabidopsis* Fruit

The *Arabidopsis* fruit is a small seed pod, dry at maturity, called a silique. It is a simple fruit, composed of a single pistil. The two-fused carpels in the center of the Arabidopsis flower are together called the gynoecium, the precursor to the silique. The gynoecium of *Arabidopsis* is overall a tube-shaped structure rising up from the center of the flower. It may be broadly divided into three regions along the apical-basal axis (Fig. I-3B). The basal-most region is the gynophore, a short stalk connecting the gynoecium to the base of the flower. Corresponding to the petiole of a leaf, it is very short in the wild-type but becomes longer and more visible in some mutants. Above it is the ovary, which accounts for the majority of the length of the organ. This hollow, tubular structure consists of two expanses of valve tissue. Each valve covers half of the tube. At the meeting line between the two valves is the medial axis, where a septum forms and separate the cylinder into two chambers or locules. Along the edge of each carpel is the carpel margin meristem (CMM) (Fig. I-3A), each of which gives rise to a row of ovules. Above the ovary are the style and stigma, providing for pollination.

A number of models have been proposed regarding the molecular mechanism that specifies the fate of tissues along the apical to basal axis of gynoecium. The most influential model is the auxin gradient model, wherein apically-localized auxin biosynthesis combined with diffusion down the length of the structure functions to create an apical to basal (high to low) gradient of auxin, casting auxin in the role of a classical

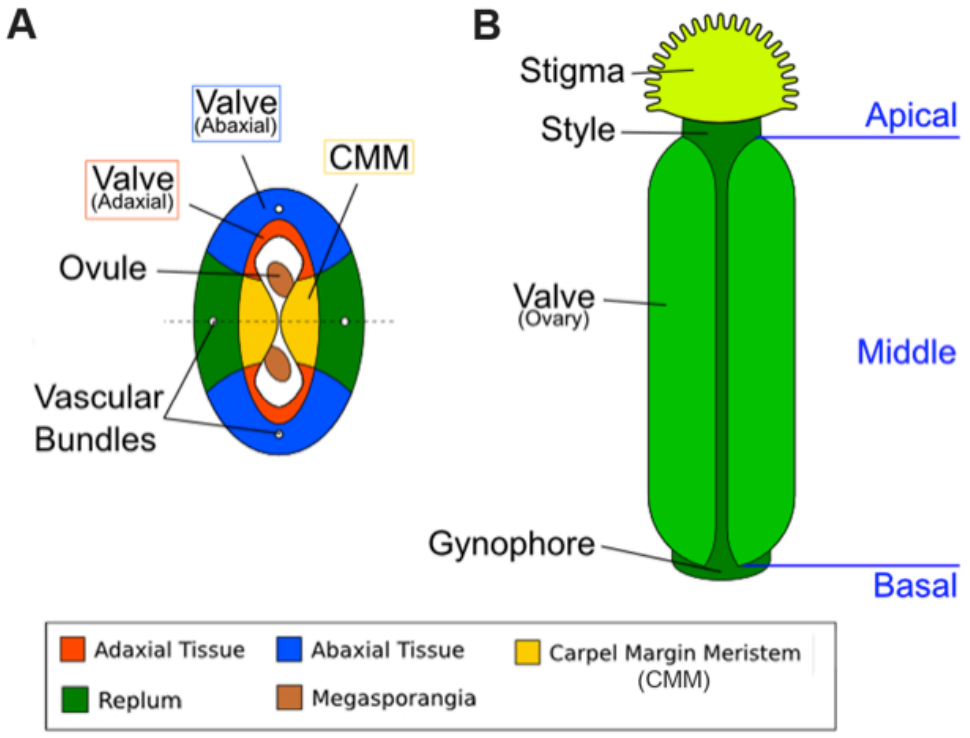

Figure I-3 – The *Arabidopsis* gynoecium.

(A) The cross-sectional view of the *Arabidopsis gynoecium*, consisting of two fused carpels enclosing two locules. Note the vascular bundles. Although there are four rows of ovules, only two ovules are visible in the cross-section since the rows alternate within each locule. **(B)** A diagram of the *Arabidopsis gynoecium*, showing that it consists of three regions along the basal-to-apical axis. The basal section consists of a short stalk, the gynophore, the middle section is the ovary, and the apex consists of style and stigma.

diffusion-gradient morphogen (Nemhauser et al., 2000). The high level of auxin at the apex specifies the stigma/style development, while the low auxin at the base promotes gynophore formation. The midrange level of auxin found in between promotes the formation of ovary wall (valve) tissue. This model was based on the gynoecium phenotypes of auxin biosynthesis and signaling mutants, but despite efforts, no auxin gradient has been observed in the intervening years. In Chapter II, I provide an alternative model based on the evolution of carpels from leaves, and indicate why the gradient model is incorrect.

Molecular Genetics of flower development-the ABCE model

Arabidopsis shoot apical meristem (SAM) initiates leaves in a spiral arrangement along all growing axes. Upon receiving a mobile peptide signal, coded by the *FLOWERING LOCUS T (FT)* gene and expressed in the leaf (Notaguchi et al., 2008), the SAM switches from vegetative to reproductive growth. FT in turn is induced by long day length and regulated by a number of pathways including the autonomous pathway, vernalization pathway, and the plant hormone GA (Turck et al., 2008).

The flower of *Arabidopsis* is typical of insect-pollinated flowers. It consists of four sepals, four white petals, six stamens, and a gynoecium derived from a pair of congenitally-fused carpels in the center of the flower. Post-fertilization, all floral organs except for the carpels are lost, and the gynoecium matures into a dry fruit called a silique.

Because of their shared developmental mechanisms, it is necessary for the floral organs to have a system by which they may distinguish themselves from leaves and from each other, and a system of floral organ identity genes that performs this function was identified using genetic screens. Each gene is assigned a class (A, B, C or E) based on the set of floral organs in which it is expressed and functions (Coen and Meyerowitz, 1991; Fig. I-4). In wild type plants, the four floral organs are arranged in four concentric rings, or whorls, with sepals and petals in whorl 1 and 2 respectively, and stamens and

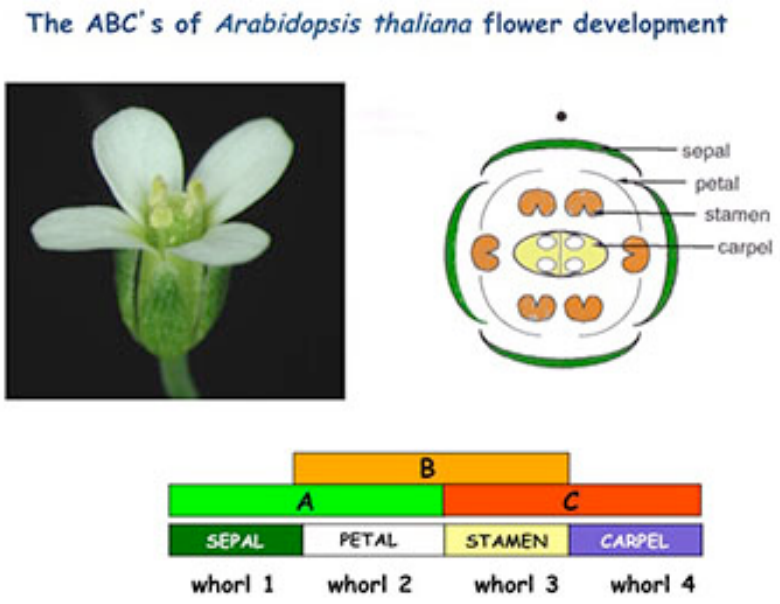

Figure I-4: The ABC model of flower development.

Figure is drawn based on Cohen and Meyerowitz, 1991. Three classes of genes that in combination specify the four classes of floral organ. Expression of A-class genes alone specifies sepals, A and B together specify petals, B and C specify stamens, and C alone specifies carpels. A fourth class, E (not shown) is present in all four whorls and is required to specify that these organs should become floral organs at all, instead of leaves.

carpels in whorl 3 and 4 respectively. Class A genes are expressed in first two whorls and are required for specification of sepal and petal identity. Class B genes are expressed in whorls 2-3 and are required for specification of stamen and carpel identity. Class C genes are expressed in whorls 3-4, specifying stamen and carpels identity. Class E genes, absent from the initial ABC model, were identified later by reverse genetics and are expressed in all four whorls, where they act as functional partners of all A, B, and C genes. In addition, genes of classes A and C are mutually repressive. Thus, the cells in a growing lateral organ will follow the sepal developmental program if class A but not B or C genes are expressed, the petal program if class A and B are expressed, the stamen program if classes B and C are expressed, and the carpel program if class C but not A or B genes are expressed.

This model was deduced via observations of loss of function mutants for each of these gene classes. Class A mutants grow carpels and stamens in their outer two whorls in place of sepals and petals as C-class genes function is extended throughout all four whorls when A function is absent. Class B mutants produce flowers with two whorls of sepals and two whorls of carpels. Class C mutants produce sepals and petals repeating into the interior of the flower as Class C genes are also required for floral determinacy; they normally serve to repress continuous formation of more flowers within a flower. Quadruple mutants of class E genes, which is functionally and phenotypically equivalent to a loss of all ABC genes, result in flowers with leaves replacing the floral organs (Coen and Meyerowitz, 1991). Hence, the lack of all ABC functions result in leaf identify,

suggesting that leaf is the ground state from which floral organs evolved (Coen and Meyerowitz, 1991; Honma and Goto, 2001; Pelaz et al., 2001; Scutt et al., 2006).

These ABCE genes, except the class A gene APETALA2, were later identified as members of the MADS box class of transcription factors. According to the ABCE model, they act as heterotetramers ("quartets") consisting of one or two class E subunits and some combination of class A, B, and C subunits depending upon which are present in the specific floral whorl. These complexes control the transcription of downstream genes to induce development of the proper floral organ (Immink et al., 2009; Theißen, 2001; Theißen and Saedler, 2001).

Figure I-5. A wild strawberry fruit showing the achene (red) and the receptacle (white) at the turning stage of development.

The Structure of the *Fragaria vesca* flower and fruit

The strawberry flower follows a similar overall morphology to that of *Arabidopsis*. The flower is organized into a similar succession of whorls, with five sepals, five petals, 20 stamens (in two whorls of 10 each), and an apocarpous gynoecium consisting of about 250 small, single, non-fused carpels (Darrow and Wallace, 1966; Hollender et al., 2012). Each carpel will develop into an achene, a hard fruit containing a single seed, and these achenes are the "seeds" visible on the exterior of the strawberry (Fig. I-5). The fleshy fruit is derived from the receptacle, the tip of the stem from which the floral organs

Figure I-6. *F. vesca* receptacle in a flower.

The green shape in the center is the receptacle, the tip of the floral axis, which develops into the edible fruit (actually an accessory fruit). Surrounding it are the carpels, each with a single ovule. These develop into the achenes, the hard "seeds" that dot the surface of the strawberry and are the true, botanical fruits of the plant.

initiate, and is thus an accessory fruit (Fig. I-6). The receptacle of the diploid *F. vesca* may be red or yellow in color, depending on the cultivar, and is sweet-tasting.

The diversity of fruit structure within *Rosaceae* is of considerable interest; within the clade the development of the receptacle into an accessory fruit surrounded by dry achenes is unique to strawberry and the closely-related genus *Potentilla*, commonly called "mock strawberries" (Eriksson et al., 2003); even in their closer relatives such a raspberry that exhibit the same aggregate fruit structure, it is the true fruit rather than an accessory fruit that becomes fleshy and sweet during ripening. Study of the development of strawberry's unusual flowers and its fleshy fruit is therefore interesting biologically as well as useful economically.

Strawberry fruit development

The strawberry accessory fruit, upon fertilization of the achenes that surround it, must develop from a small, hard floral structure (the receptacle) with no flavor and no pigmentation into the large, soft, sweet, and red mature fruit. The fruit's development post-fertilization divides into four stages: a cell proliferation stage, a cell expansion stage, and a ripening stage (Gillaspy et al., 1993; Hollender et al., 2012). The fruit will not develop in the absence of successful fertilization, therefore, fertilization serves to trigger fruit initiation—referred to as "fruit set". The strawberry spends eight to ten days growing larger, first via cell proliferation and then by cell elongation. At the end of this period, at stage 5, the fruit ceases growth and begins to ripen, accumulating red pigment in the form of the anthocyanin callistephin, converting cellulose into fructose, and producing the volatile esters and furanones that lend the strawberry its flavor (Bood and Zabetakis, 2002), resulting in the characteristic sweet, soft mature fruit. The strawberry is non-climacteric and the ripening process is not associated with a burst of ethylene as in climacteric fruits.

Our lab has done extensive morphological characterization of the development of the strawberry flower and fruit (Hollender et al., 2012), laying an important foundation for the interpretation of RNA-seq data and mutant phenotypes. Further, two sets of RNA-seq data have been generated to explore the reproductive development of the plant. The first set concerns fruit development: five different fruit tissues (Ovary wall, ghost (endosperm and seedcoat), embryo, receptacle-pith, and receptacle-cortex) at five developmental stages (fruit development stage 1-5) were isolated and RNA-seq data for each of these 25

tissues (five tissues × five stages) was generated (Kang et al., 2013). Two biological replicates were made for each of the 25 tissues. Second, tissues from flower development were isolated and RNA-seq data generated. They include the perianth (petals + sepals), the anthers, the carpels, the receptacle, the microspores, and the pollen, along with leaf and seedling as vegetative controls. The flower tissues were collected from stages 1 – 12 of flower development. The microspores were taken from stage 10 anthers. Because of the small size of young stage floral tissues, Laser capture microdissection (LCM) has been used to isolate the earlier-stage (stages 1–7) floral tissues as well as the microspores. The remaining tissues were isolated via hand dissection. In total, 17 floral tissues were isiolated and RNA-seq data generated. These RNA-seq datasets provide excellent resources for the research community, but the massive database also poses challenges for researchers seeking to mine and visualize the data. The work described in Chapter IV is aimed at solving this obstacle by adapting and deploying a web-based interface allowing that the expression levels of each gene to be visualized in specific tissues in a graphical format.

Regulation of anthocyanin biosynthesis during fruit ripening

The red pigment callistephin produced in the final stages of strawberry ripening is a pelargonidin-based anthocyanin (Crozier et al., 2008). The red color of strawberry fruit significantly affects its market success. In nature, the red fruit attracts birds for seed dispersal (Schaefer et al., 2006). In addition to the visual effect, anthocyanin also serves important functions from protecting plant tissues against UV irradiation to increasing

nutritional benefits to the consumers. In tomato, increased anthocyanin production was shown to double the shelf life of tomato fruit by delaying over-ripening and reducing susceptibility to gray mold (Bassolino et al., 2013; Zhang et al., 2013). Therefore, fruit color research is of significant importance both in basic science and in practical applications.

The anthocyanin biosynthetic pathway is well studied and highly conserved among higher plants. In addition to enzymes that catalyze the synthesis of anthocyanins, transcription factors have been shown to play critical roles in regulating the expression of anthocyanin biosynthesis genes. Chiefly, the MYB, bHLH, and WD-repeat proteins (the MBW complex) were found to regulate the expression of the anthocyanin pathway genes in plant tissues (Albert et al., 2014; Ramsay and Glover, 2005). The most well-studied transcription factors that regulate pigment biosynthesis are R2R3 MYB proteins, which possess two DNA-binding domain repeats (R2R3). Alterations in the expression or function of a single such MYB gene could drastically alter the accumulation of anthocyanins in orange (Butelli et al., 2012), petunia (Schwinn et al., 2006), tomato (Butelli et al., 2008; Cermak et al., 2015), and peach (Tuan et al., 2015). Based on studies in other species, the strawberry homolog of *MYB10*, *FaMYB10,* was identified. Over-expression of *FaMYB10* in the garden strawberry, *F. ananassa,* resulted in plants with elevated anthocyanin levels in roots, foliage, and fruit (Lin-Wang et al., 2010). In *F. vesca*, an RNAi construct against *FvMYB10* converted red fruit into light yellow fruit (Lin-Wang et al., 2014), indicating that *FvMYB10* encodes one important transcription factor that controls strawberry fruit color.

In Fragaria vesca, it is interesting that some alpine and woodland accessions such as Yellow wonder and Hawaii 4 develop yellow colored fruits. However, it is not known which gene(s) and mutation may underlie the yellow fruit color. Two recent studies attempted to identify the gene responsible for the naturally occurring yellow fruit in *F. vesca* by comparing transcripts between red (Rügen) and yellow (Yellow Wonder) varieties (Xu et al., 2014; Zhang et al., 2015); transcript levels of several anthocyanin biosynthesis enzyme genes (C4H, CHS, CHI, F3H, DFR, and ANS) and several MYB genes (MYB1, MYB86 and MYB39) were found to be down-regulated in the Yellow Wonder fruit. However, these differentially expressed genes may reflect downstream effects of the causal mutation. The work described in Chapter III is aimed at identifying the causal mutation that underlies the yellow fruit color in wild strawberries.

The Plant Hormone Auxin and its Role in Fruit Development

Auxin is the first plant hormone ever identified, which controls many processes in plant growth and development. The term "auxin" refers to a family of small organic molecules, but by far the most prevalent and influential is indole-3-acetic acid (IAA). IAA is produced via a two-step biosynthetic pathway from tryptophan, which is converted to indole-3-propionic acid (IPA) by the *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS* (TAA) family of genes, and then to IAA by genes in the YUCCA (YUC) family (Won et al., 2011). The hormone moves from cell to cell apoplastically, as described by the chemiosmotic model (Fig. I-7; Rubery and Sheldrake, 1974). It is biologically active when inside a cell, where the slightly basic pH

ensures that it is deprotonated, in which form it cannot diffuse across the cell membrane. The cell may control its movement by exporting it to the apoplast or the vacuole via transporters in the membrane. Some of these, members of the PIN family, are polarly localized and thus allow export in a particular direction. Others, the ABC transporters, are non-directional (Geisler and Murphy, 2006). Once the auxin is in the apoplast, the lower pH protonates it, allowing it to diffuse into nearby cells. Membrane-bound import proteins also exist that speed the rate of uptake. Auxin is thus a signal that an individual cell can direct toward any of its neighbors, and whose larger patterns of movement the plant can control in order to help coordinate its development.

Auxin signaling is via the AUXIN RESPONSE FACTOR (ARF) family of transcription factors. In the absence of auxin, repressors of the IAA family bind to the ARFs and prevent them from binding to DNA. Auxin binds to and activates receptors of the TIR/AFB family, which then ubiquitinate the IAA-family repressors, triggering their degradation by the cell's proteasomes and freeing the ARFs to bind DNA at auxin response element (AuxRE) sites where they activate or repress downstream genes.

Auxin is produced throughout growing aerial organs, such as growing leaves and shoot tips, and cells along the surface direct the export of their auxin against its concentration gradient toward neighbors that already have a higher concentration of the hormone. As a result, it pools in small areas with a predictable spacing. In these pools, it initiates outgrowth, the nature of which depends on the organ context; pools on shoot tips outgrow into leaves or other lateral organs (leaves, bracts, etc.) (Barton, 2010), while pools on lateral organ edges outgrow into serrations, lobes, or leaflets (Byrne, 2012). A

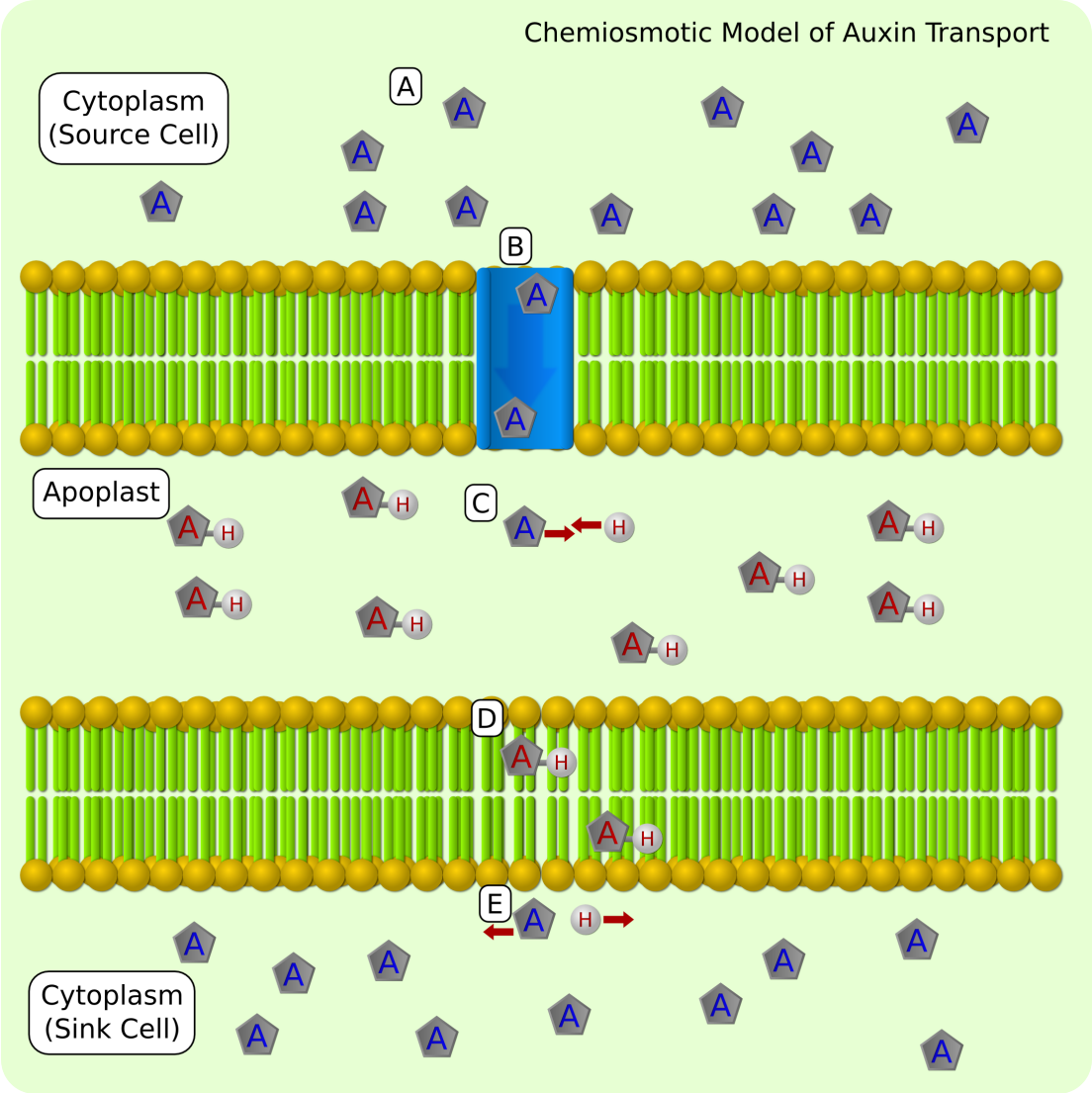

Figure I-7: The Chemiosmotic Model of Auxin Transport

Chemiosmotic Model of Auxin Transport. Auxin in the alkaline cytoplasm of the source cell is deprotonated (A) and cannot diffuse across the membrane. It is actively transported out of the cell by auxin efflux transporters (B). Once outside, the acidic environment of the extracellular space (apoplast) protonates the auxin (C), allowing it to diffuse across the membrane of the sink cell (D). Once there, it is once again deprotonated (E) and cannot re-cross the membrane.

similar pattern is observed in the carpel margin meristem, where the outgrowths are the ovules (Galbiati et al., 2013). Excess auxin is directed downward from the center of the pool, toward the vasculature of the existing organ. Its tendency to be directed against its gradient helps to canalize the flow, and the auxin channels thus formed direct the placement of vasculature in the new organ, ensuring that new vessels connect to existing ones and that growing organs are supplied. Localized auxin production is also important to this process. As the primordium grows into a young organ, auxin production localizes to the base of the organ, and can be visualized via the spatial distribution of YUCCA family of genes that catalyze the final step of the production of auxin from tryptophan (Cheng et al., 2006).

Auxin also plays a critical role in fruit set. In strawberry, removal or non-fertilization of the achenes halts fruit set, but exogenous application of auxin or GA rescues development (Nitsch, 1950; Kang et al., 2013), suggesting that fertilization induces production of auxin by the achenes which travels to the receptacle to activate its growth stage. At the end of the growth phases, the achenes cease to produce auxin and receptacle auxin levels drop, signaling the switch to the ripening stage of development (Archbold and Dennis, 1984; Archbold and Dennis, 1985; Given et al., 1988; Kang et al., 2013).

Gibberellins and fruit development

Gibberellins are a class of hormones that chiefly promote cell growth and elongation and release seeds and dormant tree buds from dormancy (Núñez-Elisea and L. Davenport,

1998; Kucera et al., 2005) They are made via a multistep pathway that converts Geranylgeranyl diphosphate (GGDP) to the intermediate GA_{12}, which is then converted to the bioactive GAs: GA_1, GA_3, GA_4, and GA_7 (Sun, 2011). They act primarily by inducing ubiquitin-mediated degradation of proteins in the DELLA family, derepressing downstream response factors (Sun, 2011). In the shoot apical meristem they appear coincidently with the auxin maxima on the sites of lateral organ initiation, and act downstream of the auxin to induce cell growth, forming the characteristic "bump" of the primordia (Shani et al., 2006).

In strawberry fruit, the gibberellin GA_3 serves an important role in triggering fruit set (Kang et al., 2013). The receptacle will not begin its development into a fruit if the achenes are removed, but exogenous application of GA_3 rescues this phenotype and stimulates fruit set with or without the achenes. This is similar to the effect of auxin on fruit set and indicates an important signaling role for GA_3 in this process. In *Arabidopsis,* too, GA has been shown to induce fruit set and to cause the gynoecium to enlarge into the mature fruit (the silique). In this role the GA functions downstream of auxin produced in the ovules (Dorcey et al., 2009; Galbiati et al., 2013). The auxin acts locally in the ovules to stimulate GA production, and the GA is then transported to the carpel to induce silique fruit growth (Galbiati et al., 2013). In strawberry, a similar process likely occurs, as both auxin and GA produced in the achene induce receptacle fruit set, likely due to a similar regulatory relationship between Auxin and GA as in *Arabidopsis*. Both hormones are transported to the receptacle where response genes for both hormones have been shown to be up-regulated (Kang et al., 2013). This process

similarity between species is interesting given the differences between the two species' fruits: The strawberry and *Arabidopsis* fruits develop from different parts of the flower and consequently differ greatly in structure, yet the same hormones play the same roles in inducing and managing the ripening process.

Summary of my thesis research

Chapter II details a new model for auxin's role in gynoecial patterning. In it, I propose that, contrary to existing models, the defects observed in auxin-related mutants and auxin-disrupting chemical treatments are due to early disruptions in ad/abaxial polarity rather than to the later disruptions in apical-basal patterning described in the auxin gradient model. I base this on the timing of the treatments of auxin transport inhibitor (a major part of the basis for the existing model) relative to various morphogenic milestones and on current models of leaf and lateral organ morphogenesis and the effects of similar treatments and mutations on leaves. Consistent with the accepted idea that carpels are modified leaves, this new model brings carpel development in line with current models of how leaves develop, and places it within the broader framework of how auxin shapes the entire plant. Understanding how plants shape themselves opens up the possibility of modification and control of these processes, of shaping the gynoecium (and other plant structures) to increase yields, adapt to adverse and changing environmental conditions, and adapt plants for more efficient farming methods. This work was published as a theory paper in Frontiers in Plant Science (Hawkins and Liu, 2014).

Chapter III presents an analysis of high-throughput re-sequencing of the genomes of three cultivars of the diploid strawberry *F. vesca*, which identify genome-wide DNA variants and genetic relations among these three cultivars. Through this analysis, I identified a candidate SNP responsible for yellow fruit in two of the *F. vesca* varieties. Further experimental tests (performed by Julie Caruana, a postdoc in our lab) confirmed that the identified SNP indeed caused a complete loss of function of the FvMYB10 gene. Identifying this gene not only helps elucidate the recent evolutionary history of the strawberry, but opens the door for breeding to increase strawberry anthocyanin content, increasing the benefits gained from the compound's antioxidant and antifungal properties. This work has been submitted to Scientific Reports.

The final chapter, Chapter IV, presents a new tool for visualizing the RNA-seq transcriptome data published by Kang et al. (2013) and Hollender et al., (2014). This tool is based on the eFP tool for visualizing *Arabidopsis* microarray data previously developed at the University of Toronto (Winter et al., 2007). I developed this into a new tool that allows the user to search for a specific Strawberry gene and then presents a diagram showing its expression in the different floral and fruit tissues. This tool opens up our transcriptome data to other researchers without the need for complex bioinformatics tools and pipelines, advancing *Fragaria vesca* as a model system and allowing for exploration of the critical processes that regulate flower and fruit development in this economically important fruit crop.

Chapter II: A Model for an Early Role of Auxin in *Arabidopsis* Gynoecium Morphogenesis

Abstract

The female reproductive organ of angiosperms, the gynoecium, often consists of the fusion of multiple ovule-bearing carpels. It serves the important function of producing and protecting ovules as well as mediating pollination. The gynoecium has likely contributed to the tremendous success of angiosperms over their 160 million year history. In addition, being a highly complex plant organ, the gynoecium is well suited to serving as a model system for use in the investigation of plant morphogenesis and development. The longstanding model of gynoecium morphogenesis in *Arabidopsis* holds that apically-localized auxin biosynthesis in the gynoecium results in an apical to basal gradient of auxin that serves to specify along its length the development of style, ovary, and gynophore in a concentration-dependent manner. This model is based primarily on the observed effects of the auxin transport blocker N-1-naphthylphthalamic acid (NPA) as well as analyses of mutants of *Auxin Response Factor 3/ETTIN (ETT).* Both NPA treatment and *ett* mutation disrupt gynoecium morphological patterns along the apical-basal axis. More than a decade after the model's initial proposal, however, the auxin gradient on which the model critically depends remains elusive. Furthermore, multiple observations are inconsistent with such an auxin-gradient model. Chiefly, the timing of gynoecium emergence and patterning occurs at a very early stage when the organ has little-to-no apical-basal dimension. Based on these observations and current models of early leaf patterning, we propose an alternate model for gynoecial patterning. Under this

model, the action of auxin is necessary for the early establishment of adaxial-abaxial patterning of the carpel primordium. In this case, the observed gynoecial phenotypes caused by NPA and *ett* are due to the disruption of this early adaxial-abaxial patterning of the carpel primordia. Here we present the case for this model based on recent literature and current models of leaf development.

The structure of *Arabidopsis* Gynoecium

Angiosperms, plants that produce flowers, are far and away the most diverse division of plants today, with even the most conservative estimates placing the number of known extant species at more than 223,000 (Scotland and Wortley, 2003). In addition to being an incredibly successful group in nature, flowering plants account for the vast majority of plants used and cultivated by humans, both for agricultural and for horticultural purposes. For this reason, there is great promise in the prospect of engineering angiosperm development to increase productivity, fecundity, and survivability. To do that in any systematic way, it is necessary to understand the genetic machinery that drives angiosperm development and that allows these plants to shape themselves into the vast diversity of forms seen in nature.

Evolutionarily, the flower consists of a complex of organs that are derived from leaves growing from a single stem. The idea that the floral organs are modified leaves, first proposed by Goethe in 1790, has been borne out in the intervening centuries by numerous lines of evidence, including morphology, vascular organization, and the organization of cells and tissues (see, e.g., Arber, 1937 for a review of early evidence). In

more recent decades, molecular evidence has emerged that reinforces this conclusion. The absence of a small number of floral organ identity genes is sufficient to convert floral organs into leaves, and their ectopic expression can do the reverse (Coen and Meyerowitz, 1991; Honma and Goto, 2001; Pelaz et al., 2001; Scutt et al., 2006).

A complete flower consists of the stem itself, divided into the pedicel and receptacle, and four different types of leaf-derived floral organs arranged in four concentric whorls around the stem. These are, from outermost to innermost: The sepals, which protect the flower; the petals, which serve as a display to attract pollinators; the stamens, which produce pollen; and the carpels, which contain the ovules that later develop into the seeds when they are fertilized. Carpels are of particular interest and significance as they constitute the angiosperms' defining feature. In many species, the carpels are fused into a single structure called the gynoecium. This structure is of critical economic importance, as it is the source of fruits and of seeds, including nuts, beans, and cereal grains. The interactions of genes and hormones that shape the structure, however, are not completely understood. *Arabidopsis thaliana*, a flowering weed and a model plant, has thus been under intensive investigation to address the underlying molecular mechanisms.

Like the other floral organs, the carpels are widely thought to represent modified leaves or sporophylls (Balanza et al., 2006; Scutt et al., 2006; Vialette-Guiraud and Scutt, 2009; Reyes-Olalde et al., 2013). The ancestral carpel is most likely ascidiate, meaning it represents an invagination of a leaf to form a hollow structure sealed by a secretion (Endress and Igersheim, 2000; Endress and Doyle, 2009; Doyle, 2012). There are a

number of possibilities as to how exactly this occurred, including curled leaf borne on axillary branch or curled leaflets borne along the rachis of a compound leaf (Doyle, 2012). Examples of ascidiate carpels can be found in the basal extant angiosperms such as *Amborella* and water lilies. Most "higher" angiosperms, however, including most monocots and eudicots (*Arabidopsis* among them), instead possess plicate carpels

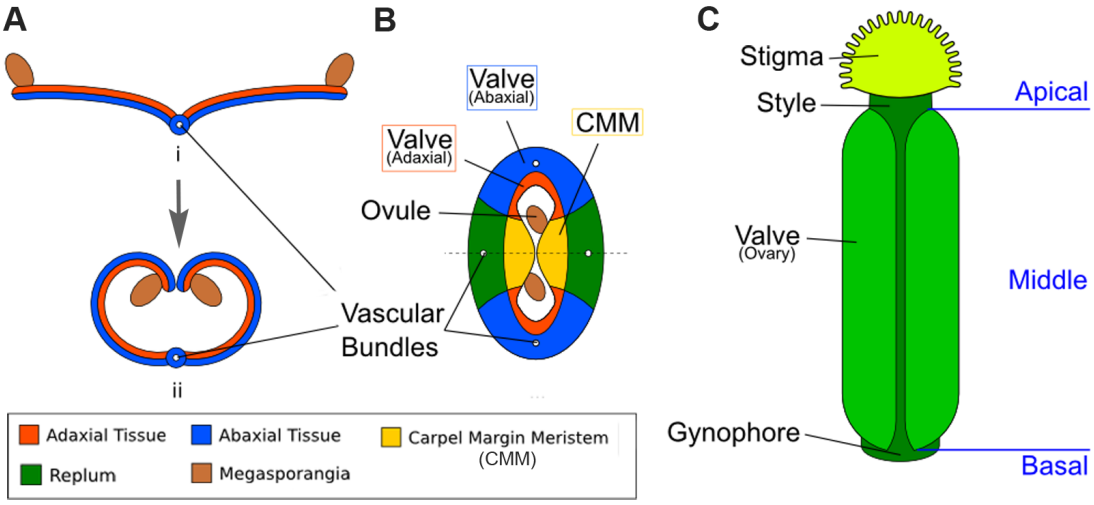

Figure II-1: Diagrams illustrating the homology between modern carpels and ancestral leaves. (A) Hypothetical evolution of a single plicate carpel based on Scagel (1965). (i) A cross section of an ancestral plant's spore-bearing leaf (sporophyll), showing megasporangia at the leaf edge. (ii) Over evolutionary time, inward curling of a megasporangia-bearing leaf and subsequent fusion at the leaf margin led to a one-chamber ovary with two rows of megasporangia on the interior (adaxial side). The actual evolutionary path is more complicated and not fully settled. (B) The cross section view of the *Arabidopsis* gynoecium, consisting of two fused carpels enclosing two locules. Note the vascular bundles. Although there are four rows of ovules, only two ovules are visible in the cross-section since the rows alternate within each locule. (C) A diagram of the *Arabidopsis* gynoecium, showing that it consists of three regions along the basal-to-apical axis. The basal section consists of a short stalk, the gynophore, the middle section is the ovary, and the apex consists of style

(Endress and Doyle, 2009; Doyle, 2012). Rather than being an invagination of the leaf, the plicate carpel is curled or folded along its length into a tube-like or book-like shape, enclosing the ovules within (Fig. II-1A). This type of structure appears to have evolved by elongation of the apical end of the primitive ascidiate carpel. In angiosperms, irrespective of carpel type, the ovule-bearing surface is strictly adaxial (Doyle, 2012).

In *Arabidopsis*, two carpels are fused congenitally to form the gynoecium (Sattler, 1973; Fig. II-1B), and each carpel is homologous to an ancestral spore-bearing leaf (sporophyll) (compare Fig. II-1A with II-1B). The adaxial tissues near the margins of the fused carpels are meristematic and are thus called the carpel margin meristem (CMM) (Fig. II-1B). The CMM is responsible for generating the placenta, ovules, septum, transmitting tract, style, and stigma; these tissues are critical for the reproductive competence of the gynoecium (Wynn et al., 2011; Reyes-Olalde et al., 2013). From the base to the apex of the gynoecium are three morphologically distinct regions (Fig. II-1C). The basal-most region is the gynophore, a short stalk that connects the rest of the gynoecium to the flower. The apical-most region of the gynoecium consists of the style and stigma. In the middle of the gynoecium is the ovary; a cross section of the ovary (Fig. II-1B) shows two valves (also called ovary valves or carpel valves) separated externally by the replum and internally by a septum, dividing the interior into two locules. Each locule protects two rows of ovules initiated along the carpel edges from the CMM.

The homology between carpels and leaf-like lateral organs extends to the resemblance of carpel valves to leaf blades (lamina) and the CMM to the leaf margins. In certain angiosperm species such as *Kalanchoe daigremontiana,* also known as “mother of

thousands", leaf margins produce plantlets and express the meristem marker gene *SHOOT MERISTEMLESS* (*STM*) in a small group of leaf margin cells that were initiating plantlets (Garces et al., 2007), much like the *STM*-expressing placenta along the *Arabidopsis* carpel margins (Long et al., 1996). The possibility of conserved molecular mechanisms that specify the basic organ plan of the leaf and carpel draws support from several prior observations: Firstly, NPA treatment causes the formation of both needle-like leaves without a lamina and of stalk-like gynoecia without valves (Okada et al., 1991a). Further, NPA treated young leaves showed increased density of veins along their margins and multiple parallel midveins, much like NPA-treated gynoecia where the veins linking the gynoecium to the receptacle are increased in number (Nemhauser et al., 2000). Secondly, when one manipulates the expression of A, B, C and E-class floral homeotic genes, floral organs can be turned into leaves or vice versa (reviewed in Goto et al., 2001). Thirdly, single sepals can be readily turned into single, free carpels, such as in *Arabidopsis ap2-2* mutants (Bowman et al., 1989).

Auxin biosynthesis and its role in Arabidopsis gynoecium development

The IAA biosynthetic pathway begins with tryptophan or a tryptophan precursor (Bartel, 1997; Ljung, 2013). Recent reports suggest that auxin biosynthesis in plants involves only a two-step pathway, in which *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1 (TAA1)* and its four homologs *TAR1-4* convert tryptophan to indole-3-propionic acid (IPA). Members of the *YUCCA (YUC*) family of flavin monooxygenases

then catalyze the conversion of IPA to auxin (Mashiguchi et al., 2011; Stepanova et al., 2011; Won et al., 2011; Zhao, 2012).

Analyses of the expression and mutant phenotypes of auxin biosynthesis genes indicate that localized synthesis of auxin is critical to proper gynoecium morphogenesis.

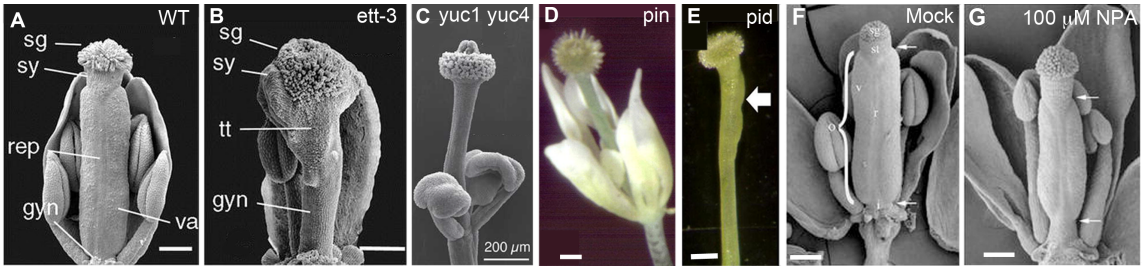

Figure II-2: Gynoecium phenotypes of mutants defective in auxin biosynthesis, transport, or signaling.

(A) Wild-type gynoecium at stage 12 with the parts labeled as stigma (sg), style (sy), replum (rep), valves (va), and gynophore (gyn). (B) *ett-3* gynoecium at stage 12, showing an elongated gynophore, a diminished valve pushed toward the apex, and expanded stigma, style, and transmitting track (tt) tissue. (C) Gynoecium of a *yuc1-1 yuc4-1* double mutant, showing the complete absence of ovary valve and an enlarged apical stigma. (D) A weak *pin* mutant showing a gynoecium without any ovary valve tissue. (E) A *pid* gynoecium with one small ovary valve (arrow). (F and G) NPA-treated wild type *Arabidopsis* gynoecium. The apical and basal boundaries of the ovary are marked by a pair of arrows. The various tissues are indicated with letters: ovary (o), replum (r), valve (v), style (st), and stigma (sg). Images are reproduced from Heisler et al. (2001) (A and B), Cheng et al., (2006) (C); Roeder and Yanofsky (2006) (D-E), and Nemhauser et al., (2000)(F-G) with permissions from Copyright Clearance Center or Creative Commons Attribution-Non-Commercial 4.0 International Liscence. Scale bars: 200 μm (A-C); 250 μm (D-E); 165 μm (F) and 140 μm (G).

Among the 10 *YUC*-family genes, *YUC1* and *YUC4* appear to play important roles in gynoecium development (Cheng et al., 2006) as double *yuc1 yuc4* mutants show a stalk-like gynoecium (Fig. II-2A, C), completely missing the ovary valves. *In situ* hybridization and promoter-*GUS* (β-glucuronidase) fusions have revealed that both *YUC1* and *YUC4* are expressed in inflorescence apices and young floral primordia. Most interestingly, *YUC1* and *YUC4* are expressed at the base of young floral organs including carpel primordial (Cheng et al., 2006). This specific expression pattern at the base of emerging floral organs is likely critical to proper floral organ initiation and apical-basal patterning. In older flowers, *YUC4* expression is concentrated at the apical tip of carpels, stamens and sepals (Cheng et al., 2006) and may be involved in later proper differentiation of floral organs.

Likewise, double mutants of TAA1/TAR family genes exhibit stalk-like gynoecia highly similar to those of *yuc1 yuc4* double mutants (Stepanova et al., 2008). The TAA1-GFP protein is localized in a few cells located at the apex (L1 layer) of young carpel primordia as early as floral stage 2. This localized expression continues to floral stages 4, when a few epidermal cells at the central dome of the carpel primordia express TAA1. As stage 4 is when gynoecium primordia emerge, this localized TAA1 expression may be involved in the apical-basal patterning of the gynoecium. At later floral stages 5-9, TAA1-GFP is prominently expressed in the medial ridge of the gynoecium; this later stage expression maybe relevant to the development of marginal tissues including ovules, styles, and stigma. Based on localized and specific expression patterns of TAA1/TAR, Stepanova et al. (2008) suggested that auxin is synthesized in different regions at

different developmental times and that localized auxin biosynthesis may represent a mechanism redundant to auxin transport in ensuring that robust local auxin maxima are able to form.

Auxin signaling

Auxin signaling consists of a system of the TIR/AFB family of receptors, the IAA family of repressors, and the *Auxin Response Factor* (*ARF*) family of transcription factors. ARFs contain a DNA binding domain but most require homodimerization to bind DNA (Ulmasov et al., 1999). IAA-family repressor proteins bind to ARFs and competitively inhibit their ability to homodimerize. The TIR/AFB family of auxin receptors, when bound by auxin, induces the ubiquitination and degradation of the IAA repressors, thus freeing the ARFs to bind DNA. This may result in transcriptional activation or repression of target genes, depending on the co-factors bound to the ARF (Dharmasiri et al., 2005; Kepinski and Leyser, 2005; Mockaitis and Estelle, 2008; Calderón Villalobos et al., 2012). AUXIN BINDING PROTEIN1 (ABP1) represents a second type of auxin receptor, which acts as part of a system of rapid and local auxin responses on the plasma membrane (Dahlke et al., 2010; Xu et al., 2010; Effendi and Scherer, 2011; Shi and Yang, 2011; Craddock et al., 2012), though this picture has been called into question as of late (Gao et al., 2015). The plasma membrane localized TMK1 receptor-like kinase was recently found to physically associate with ABP1 at the cell surface to regulate ROP GTPase signaling in response to auxin (Xu et al., 2014). In

addition, ABP1 also acts to negatively regulate the SCF (TIR/AFB)-mediated auxin signaling pathway (Tromas et al., 2013).

ETTIN (*ETT*), also known as *AUXIN RESPONSE FACTOR 3* (*ARF3*), is a member of the auxin response factor family. Its closest in-paralog is *ARF4*, from which it appears to have split early in angiosperm evolution (Finet et al., 2010). *ETT* and *ARF4* are also expressed in the abaxial domain of leaves and floral organs, where they are believed to function as abaxialization factors in lateral organ development (Sessions et al., 1997; Pekker, 2005; Hunter, 2006). In the gynoecium, *ett* mutants show diminished or absent carpel valve tissue and an expansion of stigma, stylar, and basal gynophore (Fig. II-2B) (Sessions and Zambryski, 1995; Sessions et al., 1997; Sessions, 1997; Heisler et al., 2001). The severe gynoecium phenotype of *ett* provided one of the earliest clues pointing to auxin as a critical regulator of gynoecium morphogenesis.

Detailed analysis of ETT expression in flower and gynoecium development

ETT/ARF3 is an unusual ARF as it lacks dimerization domains III and IV. These two domains in other ARFs are required for homo- or hetero- dimerization (Ulmasov et al., 1999), and the IAA repressors only bind and inhibit dimerized ARFs. Therefore, auxin-mediated degradation of IAA is unlikely to affect ETT/ARF3 function. This presents a mystery: How is ETT able to respond to auxin without these two domains? Could ETT expression be regulated transcriptionally by auxin? To begin addressing this question, I created a construct of the *ETT* promotor, *pETT*, driving a β-glucuronidase reporter (GUS). If the control mechanism was indeed transcriptional, disruptions to the

synthesis or transport of auxin could be expected to affect the pattern of expression observable with this *pETT::GUS* construct.

A 1.4 kb upstream regulatory sequence of ETT was PCR amplified and fused to the GUS reporter. Transgenic plants harboring such a reporter construct were generated and analyzed (see Methods). The expression of GUS was visualized by the blue staining (pink in dark field) during flower development (Fig II-3). *ETT* promoter is active in the base of sepals in stage 7 young flowers (Fig. II-3A) but is absent in the two carpel primordia at this stage. Starting from stage 9, GUS stain is visible and strong through all the cell layers of the valves, which are the abaxial tissue of the carpel, consistent with ETT as an abaxially expressed gene in leaves (Kidner and Timmermans, 2007). The abaxial expression of ETT is also consistent with ETT's role in specifying carpel valves. In *ett* mutants, the valve tissue is reduced or absent, suggesting either that *ETT* directly promotes the expression of genes involved in valve formation, or *ETT* is involved in setting up the abaxial domain and the proper abaxial domain is required for valve development.

Auxin transport

Auxin travels through the plant via a cell-to-cell, "bucket brigade" style of transport. According to the chemiosmotic model, first proposed by Ruberry and Sheldrake (1974), the acidic environment of the extracellular space (the apoplast) protonates the auxin, allowing IAA to diffuse across the plasma membrane into adjacent cells. Once inside a cell, it is exposed to a more alkaline pH and becomes deprotonated. The resulting anionic IAA^- is unable to cross the lipid bilayer without the help of efflux carriers. There are two

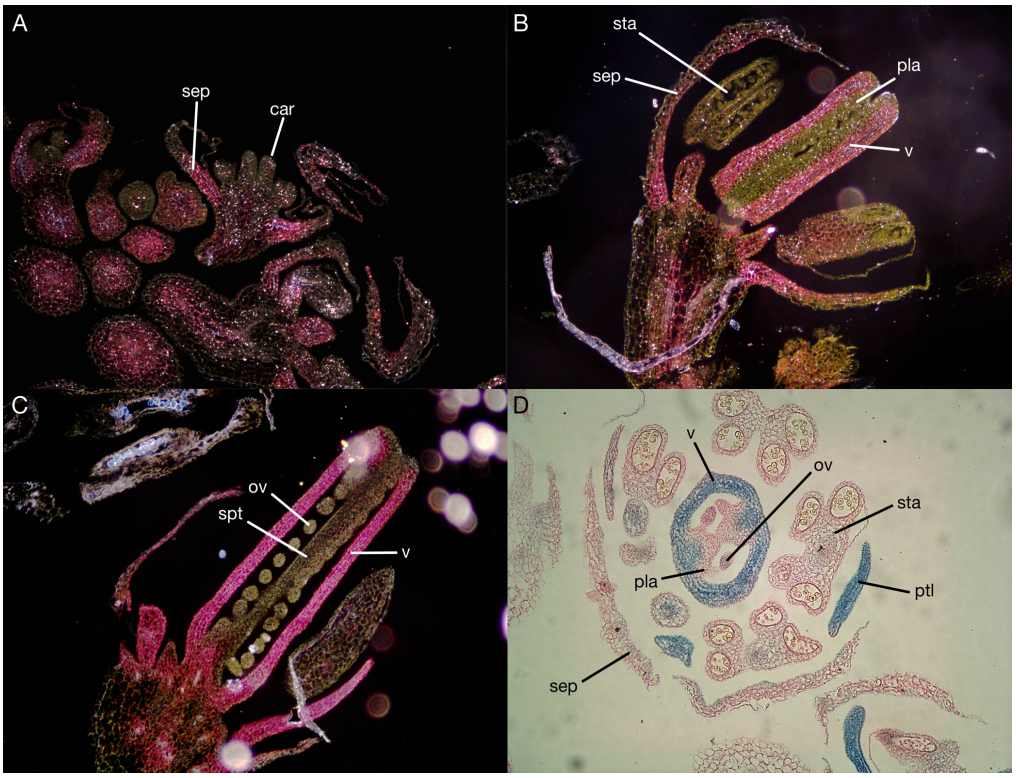

Figure II-3: *pETT::GUS* expression in *Arabidopsis* flowers

Sections of *Arabidopsis* flowers at various stages of development. GUS expression is visualized by the blue or pink (in dark field) stain in flowers of transgenic plants harboring the *pETT::GUS* transgene. A-C are dark field, with GUS expression in pink. D is light-field, with GUS expression in blue. Car: carpel, pla: placenta, ptl: petal, ov: ovule, sep: sepal, spt: septum, sta: stamen, v: valve. (A) An early-stage flower (stage 7-8) showing *ETT* expression at the base of sepals but absent from young carpel primordia. (B) A stage 9-10 flower showing *GUS* expression in the valve (abaxial tissue of carpel). The expression is absent in the adaxial tissues (placenta and ovule primordia) of the gynoecium. (C) A stage 11 flower showing GUS expression in the valves (abaxial tissue) but not in the ovules or septum. Both ovules and septum are adaxial tissues. The entire petal is also expressing GUS. (D) A cross section of a stage 10-11 flower showing *GUS* expression in the valves and in the petal.

different families of efflux transport proteins. The *PIN-FORMED* (*PIN)* family of efflux carriers is localized to a particular pole of the cell, exporting IAA selectively in the direction corresponding to PIN's localization (Wisniewska, 2006; Löfke et al., 2013). The ATP Binding Cassette B (ABCB) transporters represent the second type of auxin

efflux transporters. ABCB and PIN can independently as well as coordinately transport auxin (Titapiwatanakun and Murphy, 2009; Peer et al., 2011). Distinct modes of directional auxin transport operate in different developmental contexts. “Up-the-gradient” PIN1-based transport generates auxin maxima at lateral organ initiation site, while “with-the-flux” PIN1 polarization operates in leaf midvein patterning (Bayer et al., 2009).

A third class of auxin transport proteins is the AUX1/LAX family of auxin uptake symporters. Though IAA is believed to be capable of entering a cell from the apoplast by passing through the membrane on its own (Rubery and Sheldrake, 1974), these auxin uptake symporters are still necessary for a number of developmental processes due to their ability to create sinks for auxin to flow into (reviewed in Titapiwatanakun and Murphy, 2009; Peer et al., 2011). In addition, AUX1 was proposed to play a role in restricting auxin to the epidermis of vegetative meristems by counter-acting the loss of auxin caused by diffusion into the meristem’s inner layers (Reinhardt et al., 2003).

Strong null mutants of *PIN1* produce no lateral organs or axillary shoots, resulting in the bare, pin-like shoot that gives the mutants their name (Okada et al., 1991b; Gälweiler, 1998; Palme and Gälweiler, 1999; Benková et al., 2003). In weak *pin* mutants, lateral organs can develop but the gynoecium is often valveless and topped with stigmatic tissues, which is reminiscent of the abnormal gyoecium of auxin biosynthesis mutants described above (compare Fig. II-2C, D). *PINOID (PID)*, an AGC3-type protein kinase, acts to phosphorylate PIN to regulate PIN’s polar localization in the cell (Friml et al., 2004; Huang et al., 2010). Interestingly a similar gynoecial phenotype was observed in *pid* mutants (Fig. II-2E) (Bennett et al., 1995; Benjamins et al., 2001). The action of *PIN*

proteins in transporting auxin may be blocked via the application of N-1-naphthylphthalamic acid (NPA). Application of NPA to wild type *Arabidopsis* mimics *pin* mutant phenotypes (Okada et al., 1991b; Nemhauser et al., 2000) with pin-like shoots as well as abnormal gynoecia without any valve or with reduced valves (Fig II-2F, G). Taken together, while severe disruption of polar auxin transport abolishes all lateral organ initiation and hence results in the formation of pin-like shoots, milder disruption of polar auxin transport allows lateral organ initiation but blocks proper lateral organ morphogenesis, resulting in stalk-like gynoecia (Fig. II-2D, E). The weaker *pin* and *pid* mutant phenotypes provide strong evidence that polar auxin transport is critical for gynoecium morphogenesis.

The Nemhauser model of gynoecial patterning

Multiple lines of evidence strongly indicate that the action of auxin is critical for proper development and apical to basal patterning of the gynoecium. Mutants of biosynthesis (*yuc* or *taa/tar*) and transport (*pin* and *pid*) genes show the strongest gynoecium phenotype, a phenotype that is nearly identical between them: their valveless gynoecium is basically a thin and round stalk topped with stigmatic tissues (Fig. II-2C–E). Application of the polar auxin transport inhibitor NPA shows a similar but weaker phenotype with reduced ovary valves (Fig. II-2F, G). While mutations in the auxin signaling gene *ett/arf3* cause a similar effect to those of auxin biosynthesis (*yuc/taa/tar*) or transport (*pin/pid*) in reducing ovary valve, *ett/arf3* mutants appear to exhibit more expanded stigma and stylar tissues (Fig. II-2B).

Based on the phenotype of *ett/arf3* and the effect of NPA treatment on wild type and *ett/arf3* gynoecia, Nemhauser et al. (2000) proposed a model wherein auxin

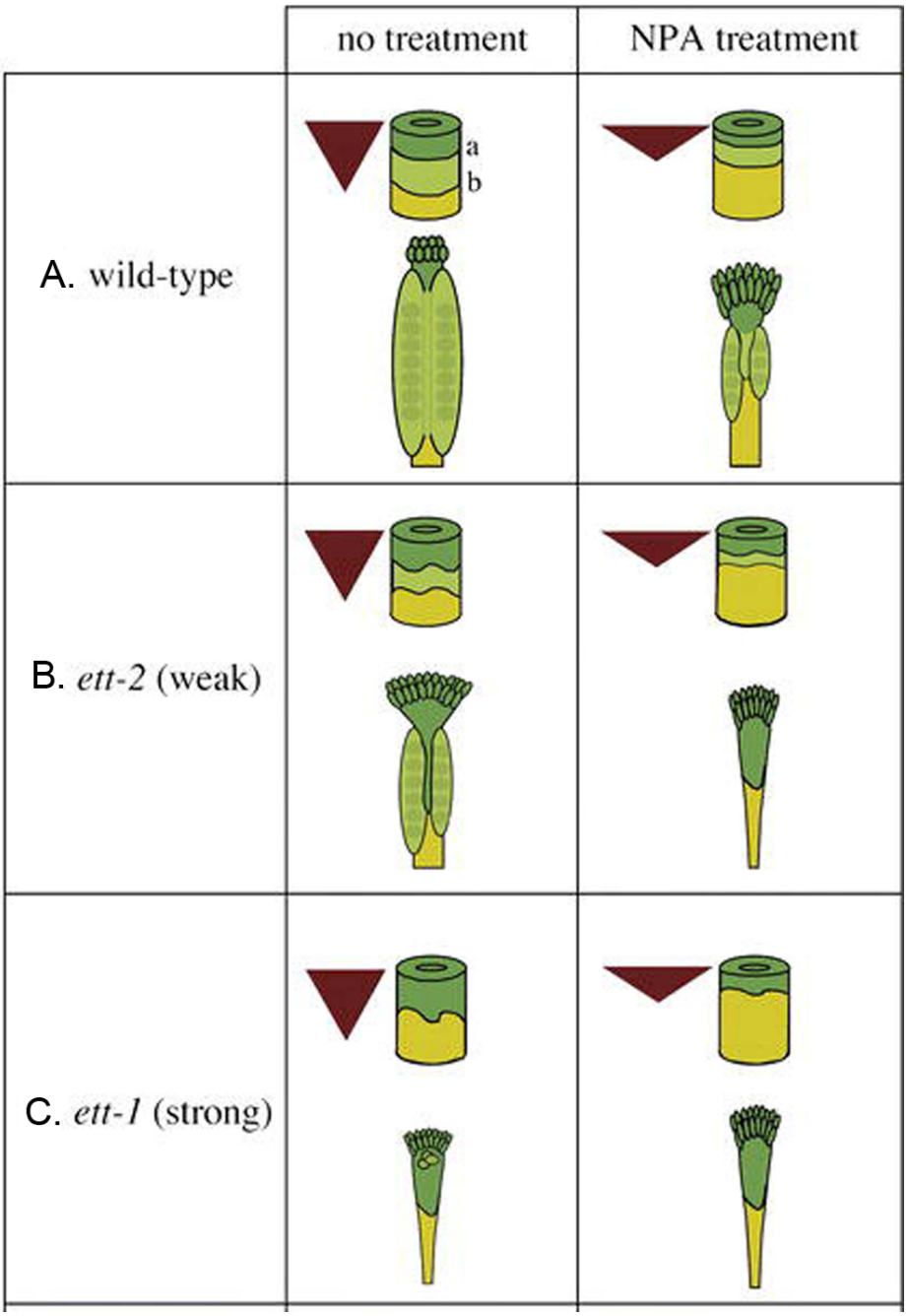

Figure II-4: The auxin gradient model.

Auxin is produced at the apex and transported toward the base, creating a morphogenic gradient that provides positional information, which is interpreted in part by *ETT* to specify ovary valve. The triangle represents the auxin gradient within the gynoecium. The cylinder represents the gynoecium with border marked "a" between the style (dark green) and ovary (light green) and border marked "b" between the ovary and gynophore (yellow). (A) Wild-type gynoecia with and without NPA treatment. (B) Weak *ett-2* mutants with a mild phenotype (left); the phenotype is significantly enhanced when *ett-2* mutants were treated with NPA (right). (C) Strong *ett-1* mutants with a strong phenotype with or without NPA treatment. The figure is reproduced from Nemhauser et al. (2000) with permission from Copyright Clearance Center.

biosynthesized locally at the apex of the gynoecium is transported basipetally, resulting in a gradient of auxin concentration with a maximum at the apex, mid-range level in the middle, and a minimum at the base (Fig. II-4A). The high auxin level at the apex specifies stigma/style, while the mid-range level promotes valve formation. At the base when auxin level is low, gynophore develops. *ETT* is partly responsible for interpreting this gradient, and promotes the formation of valve tissue in the middle region of gynoecium where there is a mid-range level of auxin. Under this model, when the gynoecium is exposed to NPA, the auxin produced at the apex is not transported down as readily, resulting in a steeper and up-shifted gradient (Fig. II-4A–C). This results in the observed phenotype of a smaller amount of valve tissue being formed near the apex of the gynoecium and a "bushier" stigma, which could be explained under this model by pooling and accumulating a higher level of apically-synthesized auxin at the gynoecium

apex. Because of the shift of auxin gradient toward the apex, the basal region, the gynophore, is expanded (Fig. II-4A–C). Mutants of *ETT*, under this model, show a similar phenotype because the job of *ETT* is to interpret the mid-range auxin gradient in the middle segment of the gynoecium to promote valve formation. In the absence of *ETT*, therefore, the auxin gradient is invisible to the plant, and valve fails to form (Fig. II-4C).

Re-evaluating the Auxin Gradient Model

The auxin gradient model was reasonably consistent with the data available at the time. Since then, however, additional information has emerged. The auxin biosynthesis gene *YUC4* is expressed (among other places) in a small region at the tip of multiple lateral organs, including cotyledons, and stamens. However, it does so largely when the organs are close to maturity (Cheng et al., 2006). In the gynoecium, the apical *YUC4* expression is not visible until after the gynoecial apical-to-basal patterning is largely determined (after stage 7-8) (Cheng et al. 2006) and thus is not likely to be responsible for the initial pattern formation of the gynoecium. At earlier stages of floral meristem development (stages 3–7; staging based on Smyth et al., 1990), *YUC4* as well as *YUC1* are expressed at the bases of young floral organ primordia, including the base of young gynoecia. In light of the timing and the dramatic gynoecium phenotype of *yuc1 yuc4* double mutants (Fig. II-2C), the early expression pattern around young floral primordia may be more relevant to gynoecial apical-to-basal patterning than the later-stage *YUC4* expression at the apex. Further, if auxin is made at the apex and responsible for stigma formation, we would expect to see a reduced or diminished stigmatic tissue in *yuc1 yuc4* double mutants. However, *yuc1 yuc4* double mutants as well as *taa/tar* double mutants

produce heads of stigmatic tissue even larger than wild type and their phenotypes are little different from those of plants that fail to transport auxin and therefore supposedly pool the auxin at the apex due to a lack of downward transport (compare Fig. II-2C with II-2D–E; Cheng et al., 2006; Stepanova et al., 2008).

Various attempts have been made to visualize the proposed auxin gradient using the *DR5* reporter. DR5 consists of tandem direct repeats of an 11-bp auxin-responsive element and, when used to drive a reporter gene, serves to report local auxin response (Ulmasov et al., 1997). Larsson et al., (2013) examined auxin distribution during early stage gynoecium development (about stage 7) using the *DR5rev::GFP* reporter. Two weak foci were detected at the apical tips of stage 7 flowers. At later stages (about stage 8), *DR5rev::GFP* expression was expanded into four foci (both medial and lateral domains) and in the pro-vasculature. Throughout the development, no gradient was observed. Other experimental work has also shown localization of auxin only to the apex of gynoecia in flowers at stage 6 or older, without showing a gradient along the apical-to-basal axis at any stage (Benková et al., 2003; Girin et al., 2011; Grieneisen et al., 2013). These data do not support the auxin gradient model. Due to presence of the late apical foci and absence of a clear gradient in the valve, we briefly considered a model wherein there was a "step-wise" gradient, with a high concentration at the top, a midlevel concentration in the middle, and a low concentration at the bottom of the gynoecium, with sharp boundaries between the zones and little-to-no gradients internal to them. This model, however, suffers from the same timing difficulties as the smooth gradient and we

could propose no clear mechanism by which such a stair-step pattern could be achieved without the zones already being defined beforehand. Thus this model was dropped.

Finally, the auxin gradient model proposed that the auxin is transported in a basipetal direction. Yet studies of the polar localization of auxin efflux carrier PIN1 show accumulation in the apical side of the replum cells (Sorefan et al., 2009; Grieneisen et al., 2013), indicating upward transport.

Fourteen years after the proposal of the auxin gradient model, accumulating new data suggest that this model, while highly attractive at the time it was proposed, should be revised or re-evaluated. Alternative models that better interpret and incorporate these new observations should be proposed.

Other alternative models

Prior to the Nemhauser's auxin gradient model, Sessions (1997) proposed a "boundary" model, in which *ETT* was proposed to regulate the two boundary lines that trisect the gynoecium into three regions, with one boundary (the apical line) dividing the ovary from the stylar tissues and the second boundary (the basal line) dividing the gynophore from the ovary above it. Sessions (1997) further proposed that the two boundaries are set as early as stage 6 of flower development, when the effects of *ett* begin to be observed. Based on this model, the effect of *ett* was interpreted as simultaneously lowering the apical boundary line and raising the basal boundary line. These two lines are also proposed in the Nemhauser model (Fig. II-4), which was built upon Sessions' "boundary" model. Since the molecular identify of *ETT* as an auxin response factor was

not published at the time when the "boundary" model was proposed, the connection to auxin was not proposed. Although Sessions (1997) mentioned an adaxial/abaxial boundary located at the distal tip of the carpel primordia, *ETT* was not proposed to regulate the adaxial/abaxial boundary.

Recently, Larsson et al. (2013), unable to detect an auxin gradient along the apical-to-basal axis of early stage gynoecium using the *DR5rev:GFP* reporter described above, pointed out that their data did not strongly support the Nemhauser gradient model. In addition, Larsson et al. (2013) noted the fact that auxin biosynthesis genes are expressed in regions not limited to the gynoecium apex as another inconsistency with the Nemhauser gradient model. They then proposed several alternative ideas/models. One was the proposal of an abaxial domain KANADI (KAN)-ETT complex that regulates PIN activity and localization during positional axis determination in gynoecia. This idea directly links AD/AB polarity with auxin in the determination of the apical-to-basal axis of gynoecia and is similar to what is being proposed below. Another idea put forth by Larsson et al. (2013) was the differential sensitivity or response of the lateral vs. medial tissues of gynoecium to auxin polar transport inhibitors.

Lessons from leaf morphogenesis

Auxin has long been known to play a role in leaf initiation. Auxin is observed to pool in small areas (maxima) on the shoot apical meristem, and the appearance of such an

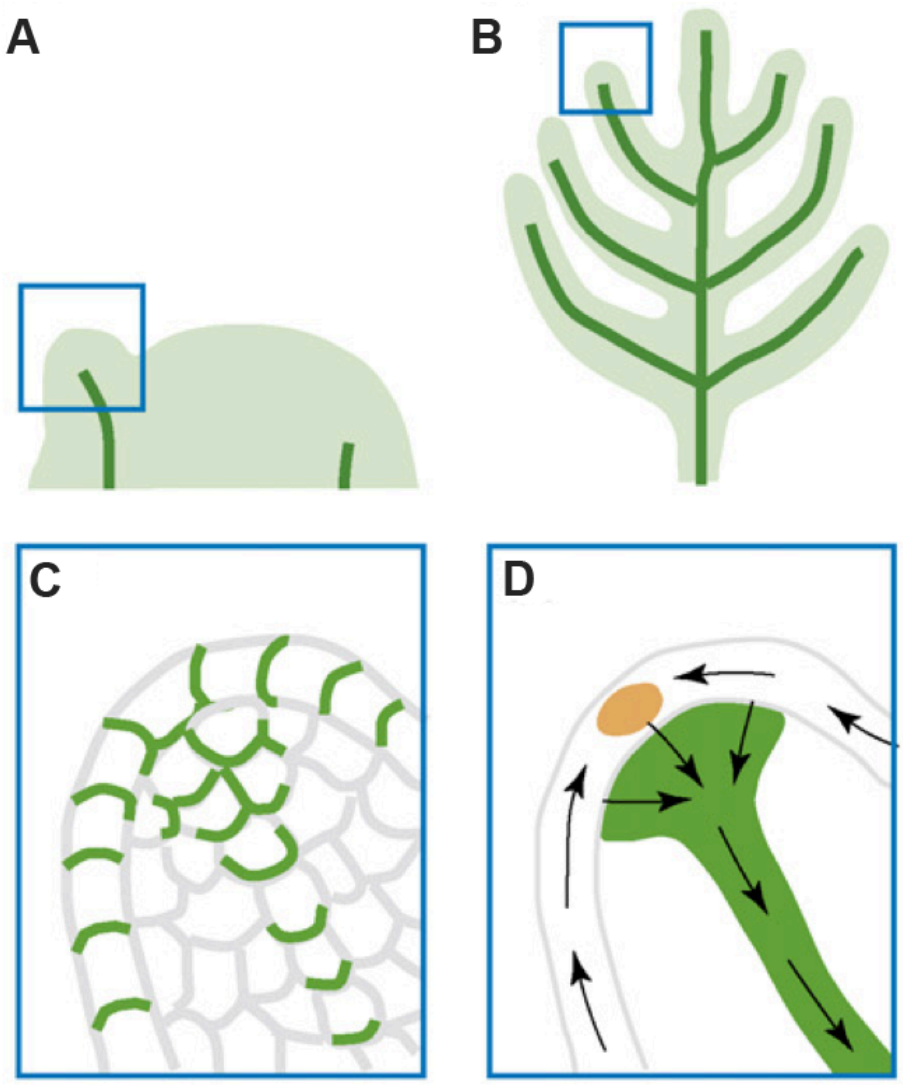

Figure II-5. Illustration of auxin transport during leaf and lateral organ initiation

(A) Leaf primordial initiation. (B) Lateral organ initiation. (C) A zoom-in diagram of the leaf primordium tip showing PIN:GFP (green) polar localization that indicates auxin transport routes. (D) Inferred auxin transport routes (black arrows) based on PIN:GFP localization. The epidermal convergence of two counter-oriented auxin flows results in a change of auxin transport direction toward the internal base of the primordium. This internal flow is responsible for the formation of the midvein. The figure is reproduced from Berleth et al., (2007) with permission from Copyright Clearance Center.

auxin maximum presages the formation of each lateral organ primordium (Reinhardt et al., 2000; Benková et al., 2003; Reinhardt et al., 2003; Heisler et al., 2005; Scarpella et al., 2006; Smith et al., 2006). An auxin maximum in the L1 layer of the meristem is the earliest mark of a new lateral organ primordium. The formation of such auxin maxima correlates with localization of the membrane-associated auxin efflux carrier PIN1, in each epidermal cell, to the side of the cell that faces toward the neighbor with a higher auxin concentration. This "up-the-gradient" transport helps to amplify the localized concentration of auxin. Heisler et al. (2005) showed *pPIN1::PIN-GFP* localization in the L1 layer toward incipient primordia starting at incipient primordium stage 3 (I3) (from youngest to oldest, the stages are I3, I2, I1, budding-primordium1 (P1), P2, etc). The signal intensity of the polarized PIN-GFP toward the auxin maxima increased steadily until primordial stage P1. The PIN1-GFP in the adaxial domain of lateral organ primordia then showed a brief reversal of transport, switching from being directed toward the primordium to being directed away from the primordium. These two waves of auxin transport suggest that auxin may act twice in lateral organ development, first in organ primordium initiation and then possibly in organ growth. If so, the timing and specific context of auxin flow may affect different processes of organ development.

The function of auxin maxima and polar auxin transport in lateral organ initiation and growth was demonstrated by examining the *pin* mutants where auxin maxima as well as lateral organ formation were absent. Further, application of auxin to the peripheral zone of the meristem induces lateral organ formation (Reinhardt et al., 2000; Reinhardt et al., 2003; Smith et al., 2006). However, Smith et al. (2006) showed that short-term NPA

treatment failed to abolish the auxin maxima, suggesting the presence of additional mechanisms that help redistribute auxin within the epidermis of the shoot apical meristem. On reaching their convergence point, the auxin flows switch direction and go basipetally towards the roots (Fig. II-5A–D; Berleth et al., 2007). The internal auxin flows are responsible for the leaf midvein formation and utilize the "with-the-flux" transport mode (Bayer et al., 2009).

Soon after a leaf primordium is initiated, one of the first signs of patterning appears in the specification of the adaxial (upper; AD) and abaxial (lower; AB) halves of the leaf. This early patterning is believed to happen in response to a signal generated at the apex or shoot apical meristem (Sussex, 1951; reviewed in Husbands et al., 2009). If the path from shoot apex to primordium is blocked, such as by a cut made directly above the incipient primordium, the adaxial-abaxial patterning of the leaf will be disrupted. The identity of this signal is still unknown but auxin remains a possibility (Husbands et al., 2009).

The AD and AB domains not only exhibit characteristic cell morphology but also express cohorts of domain-specific genes (reviewed in Kidner and Timmermans, 2007; Liu et al., 2012). These gene cohorts, generally mutually repressive, will remain associated with the AD and AB sides of the leaf as they develop. Therefore, the earliest differentiation of the AD and AB domains in lateral organ primordia can be detected by examining AD- and AB-specific marker genes. As early as stage I1, the adaxial marker *REVOLUTA (REV)* (*pREV::REV-VENUS)* was found to be visibly expressed in the adaxial domain of incipient primordia while the abaxial marker gene *FILAMENTOUS FLOWER (FIL)* (*pFIL::DsRED-N7*) was expressed in the abaxial domain (Heisler et al.,

2005). Further, *pPIN1::PIN1-GFP* expression was found to mark the boundary between AD and AB domains marked respectively by *pREV::REV-VENUS* and *FIL::dsRED-N7* (Heisler et al. 2005). Based on these results, Heisler et al. (2005) proposed that the auxin transport route plays a role in positioning the boundary between adaxial and abaxial cells. Barton (2010) also noted that the AD/AB boundary in a primordium coincides with the point in the primordium on which the epidermal auxin flows from opposite directions converge. If causal, this would indicate that a specific role of auxin transport is to establish the AD/AB boundary in incipient organ primordia.

Proper specification of the AD/AB domains is critical for proper leaf development because it generates the AD/AB boundary and the juxtaposition of the AD and AB domains is essential for leaf blade formation (Waites and Hudson, 1995). Many of these AD/AB polarity genes are required for the leaf to grow a blade (lamina), and disruption of one or more of them often creates needle-like structures, with the lamina absent or severely reduced. Examples of this include single mutants of the adaxialization factor *PHANTASTICA* in *A. majus* (Waites and Hudson, 1995), double or triple mutants of the abaxialization factor family *KANADI* (*KAN*) (Eshed et al., 2004; Pekker et al., 2005), mutants of the HD-ZIPIII adaxially-localized proteins (McConnell and Barton, 1998; Emery et al., 2003), and mutants of YABBY genes (Stahle et al., 2009; Sarojam et al., 2010).

ETT/ARF3 and its paralog *ARF4*, both auxin signaling components, have been suggested as the essential intermediaries for the gradual establishment of abaxial identity in lateral organs initiated by *KANADI (KAN)*. *KAN* encodes a GARP transcription factor

and plays a key role in the abaxial identity specification of leaves, carpels, embryos, and vasculature (Eshed et al., 2001; Kerstetter et al., 2001; Ilegems et al., 2010). Since *KAN* does not regulate *ETT*/*ARF4* transcription, and over-expression of *ETT* or *ARF4* cannot rescue *kan1 kan2* double mutants, they are thought to act cooperatively (Pekker et al., 2005). Interestingly, ETT has been found to physically interact with a KAN family protein, ATS/KAN4 (Kelley et al., 2012). This ETT-KAN complex likely acts in different developmental contexts, embryogenesis, integument development, and leaf lamina growth, by promoting abaxial fate and repressing adaxial fate (Kelley et al., 2012).

Recently it was shown that KAN1 and the adaxial HD-ZIPIII factor, REV, oppositely regulate genes in auxin biosynthesis, transport, and signaling (Merelo et al., 2013; Huang et al., 2014). KAN was shown to regulate PIN1 expression and localization during embryo as well as vascular development (Izhaki and Bowman, 2007; Ilegems et al., 2010). Additionally, the AS1-AS2 nuclear protein complex involved in leaf AD/AB polarity specification was recently shown to directly and negatively regulate *ETT* (Iwasaki et al., 2013). These experiments indicate that proper AD/AB polarity establishment and maintenance in leaves critically depend on proper regulation of auxin synthesis, transport, and signaling. Thus, dynamic auxin regulation and AD/AB polarity specification and maintenance appear to regulate each other in a feedback loop in different tissue and developmental contexts. Any disruption in auxin synthesis, transport, and signaling will affect AD/AB polarity and vice-versa.

A new model: the early action of auxin on gynoecium patterning

The evolutionary derivation of floral organs from leaf-like lateral organs suggests that the basic molecular tenets of the regulation of lateral organ polarity may be conserved. Indeed, carpels, like leaves, express members of the same gene families that control leaf AB/AD polarity. *ETT* and *ARF4* are clearly involved in carpel development and show abaxial domain-specific expression around the outer side of the tube of the developing gynoecium, the side that is equivalent to the underside of the leaf (Pekker et al., 2005). Similarly, the expression of class III HD-ZIP adaxialization factor *PHABULOSA* (*PHB*) and the abaxialization factor *YABBY1* (*YAB1*) are detected in the carpels in an equivalent configuration to that of members of their respective families found in the leaf (Franks et al., 2006; Nole-Wilson et al., 2010).

If an individual carpel primordium develops in an analogous manner to that of a leaf primordium, the AD/AB boundary of the carpels should be set very early in their development, at the incipient carpel primordium stage (approximately at floral stage 3 to 4). Further, auxin should have a major role to play at this stage in specifying the initial AD/AB boundary. The expression of the *YUC1* and *YUC4* genes suggests that auxin production is likely localized to the base of individual floral organ primordia at the very beginning of the primordial initiation (Cheng et al., 2006); this local auxin production and subsequent transport may contribute, at least partly, to the establishment of the AD/AB boundary in developing carpel primordia. As suggested by Stepanova et al. (2008), localized auxin biosynthesis and transport may represent a mechanism redundant to the transport of auxin from elsewhere to ensure robust local auxin maxima at the organ

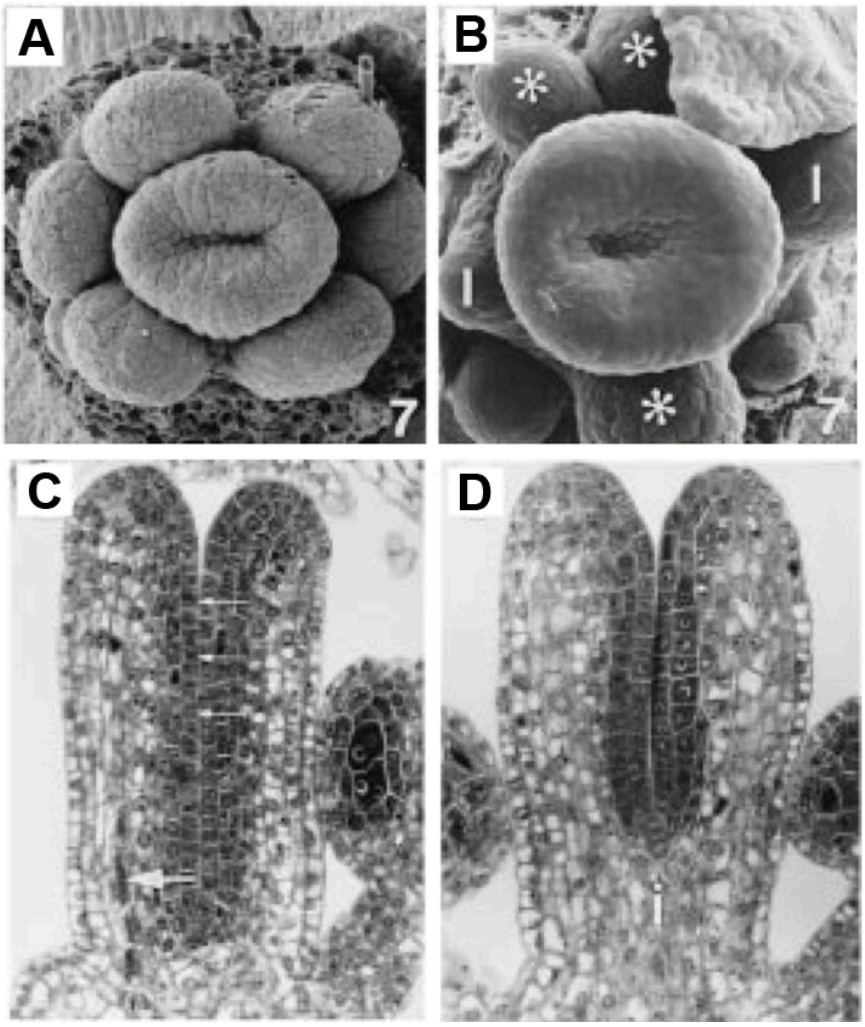

Figure II-6. Early stage wild type and *ett-1* gynoecium development

(A) Stage 7 wild type floral meristem showing upward growth of the gynoecial tube. (B). Stage 7 *ett-1* floral meristem showing a shallower gynoeciual tube. Aberrant stamen is marked with *. Scale bar is 22 uM (A) and 30 um (B), respectively. (C). Section of the medial plane of a stage 8 wild type gynoecium showing inner surface (small arrows) and medial vascular bundle (large arrow). (D). Section in the medial plane of a stage 8 *ett-1* gynoecium showing a shorter tube. The basal gynophore (i) is more prominent. Images reproduced from Session (1997) with permission from American Journal of Botany.

primordia. The site of auxin maximum at the incipient carpel primordium may set the sharp AD/AB boundary, as has been proposed for leaves and lateral organs (Heisler et al., 2005; Barton, 2010).

Based on the ideas put forward by Larsson et al (2013) linking AD/AB polarity to auxin in the determination of the apical-to-basal axis of gynoecia, we further propose that proper AD/AB polarity establishment and boundary juxtaposition in carpels is necessary for the upward growth of the carpel valve, analogous to the requirement of AD/AB boundary juxtaposition in leaf lamina formation. The valveless gynoecia in auxin pathway mutants are therefore much like the bladeless leaves of polarity mutants. Since the two carpels are congenitally fused, their primordia rise as a circular ring (Fig. II-6A; Sessions, 1997). We propose that the AD/AB boundary likely resides at the apical ridge of the ring. The close juxtaposition of AD and AB domains on either side of this boundary causes the ring ridge to grow vertically as a long hollow tube with adaxial tissues facing inward (Fig. II-6C). However, at the base of the gynoecium primordium, the AD/AB boundary is diffuse, resulting in the base of the primordium developing into a single radially symmetric and non-hollow gynophore. If the AD/AB boundary is disrupted, for example in *ett* mutants, the upward growth of the ring ridge fails to occur, or only occurs to limited extent resulting in a shallower tube (Fig. II-6B, D). The elongation of the gynophore may be regulated by a separate mechanism related to the proximal-distal growth similar to the elongation of needle-like leaves in polarity mutants.

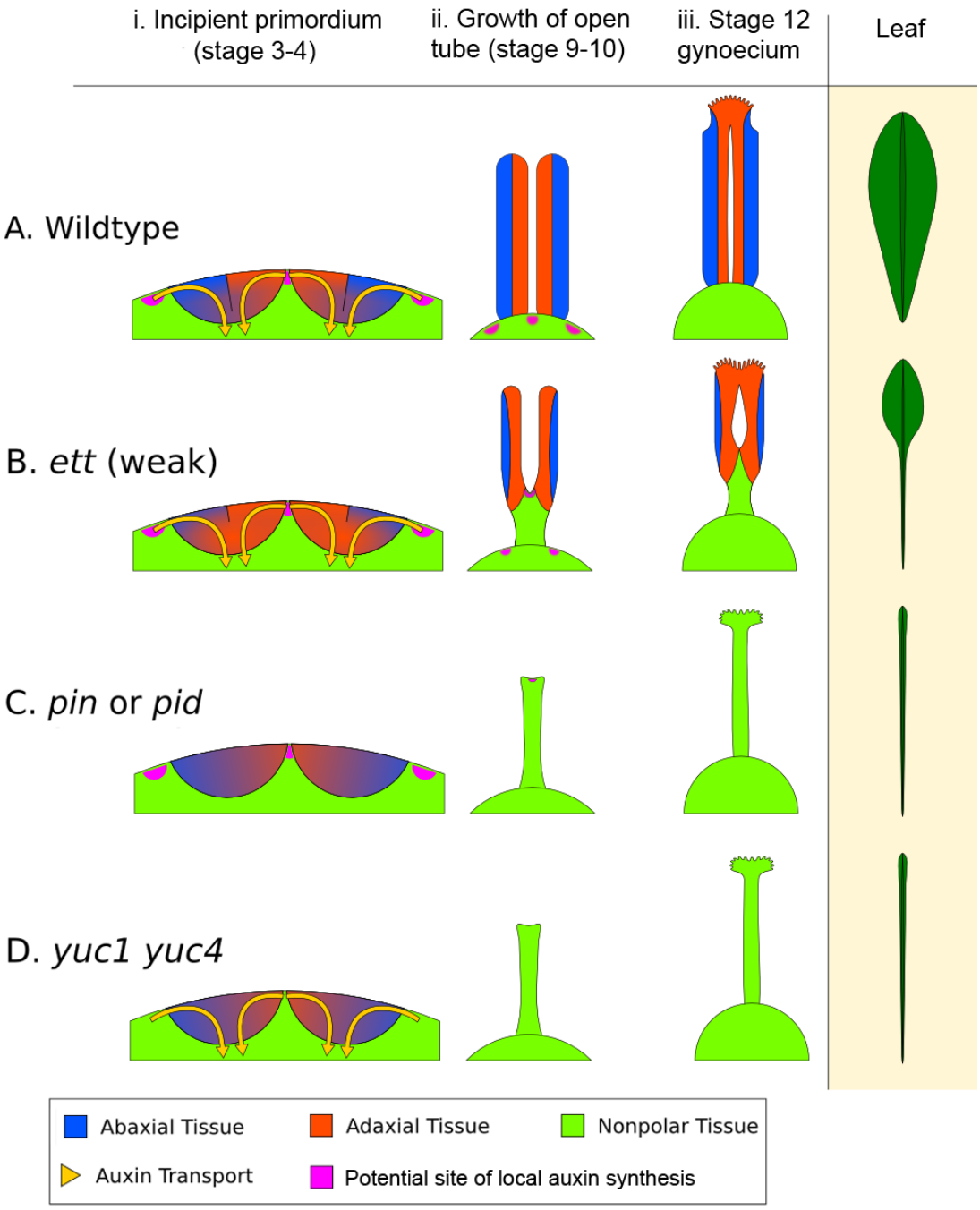
i. Incipient primordium (stage 3-4)
ii. Growth of open tube (stage 9-10)
iii. Stage 12 gynoecium
Leaf
A. Wildtype
B. ett (weak)
C. pin or pid
D. yuc1 yuc4
Abaxial Tissue
Adaxial Tissue
Nonpolar Tissue
Auxin Transport
Potential site of local auxin synthesis

Figure II-7. The early-action model of gynoecium patterning

(A) Wild type (WT) gynoecium development. The diagram in (i) depicts a young floral meristem giving rise to the two incipient carpel primordia, viewed as an enlarged longitudinal section of the floral meristem apex. In WT, opposing auxin flows (indicated by the yellow arrows) converge on the epidermal center of each carpel primordium. The convergence site likely marks the AD/AB boundary, shown as a black line between blue (AB) and orange (AD) domains. The sharp AD/AB boundary ensures upward growth of carpel tube, forming a long tube with AD domain facing interior (ii). Later the cylindrical tube differentiates into stigma/style at the apex and barely visible gynophore at the base (iii). The phenotypic analogy to a normal *Arabidopsis* leaf with lamina along its entire length is shown on the right. (B) In a weak *ett* mutant (*ett-2*), abaxial identity is compromised (but not eliminated entirely), resulting in partial adaxialization of the carpel primordia indicated by expansion of orange color (AD) area (i). As a result, there is diminishing AD/AB boundary, indicated by a shorter boundary line (i). Consequently, only a small area of the carpel primordium near the primordial apex has a clear AD/AB boundary. This shorter (or fuzzier) AD/AB boundary results in limited upward growth and hence a shorter (shallower) tube (ii), and subsequently a reduced ovary valve (iii). This phenotypically resembles leaf polarity mutants (such as double mutants of *KAN*) with a diminished lamina pushed to the leaf tip. (C) In auxin polar transport mutants such as in *pin* or *pid* mutants, the two counter-oriented auxin flows are compromised, resulting in failure to form a sharp AD/AB boundary as well as a lack of clear AD or AB identity, which is indicated by mixed blue-orange color in the primordia (i). Since the AD/AB boundary is required for valve formation, a lack of the AD/AB boundary resulted in only radialized gynophore (ii and iii), which exhibits no AD/AB polarity. (D) In auxin biosynthesis mutants such as in the *yuc1 yuc4* double mutants, a lack of local auxin biosynthesis, and hence a reduced auxin flow, results in little or no AD and AB identity being formed and no AD/AB boundary being established, as indicated by the mixed blue-orange color (i). Without the AD and AB polarity boundary, there is little to no carpel valve growth (ii-iii), analogous to a leaf without lamina (Waites and Hudson, 1995), shown on the right diagram. The pink patches highlight putative local auxin synthesis sites based on Cheng et al. (2006). The medial region expression of *TAA1* in gynoecium at floral stages 5-9 (Stepanova et al., 2008) is not shown.

Figure 7 depicts the early-action model in wild type and different auxin pathway mutants. In wild type (Fig. II-7A), each incipient carpel primordium is divided into AD and AB domains at the site of convergence of the two opposing auxin flows (indicated by the yellow arrows). The sharp AD and AB boundary marked by a black line is located near the apical surface of the incipient primordium and responsible for the upward growth of the hollow tube. Mutants of the auxin signaling component and abaxialization factor *ETT/ARF3* have compromised abaxial identity (Pekker et al., 2005), which may lead to partially adaxialized carpels and hence enlarged adaxial tissues like stigma and style. In weak *ett* mutants (Fig. II-7B), a compromised abaxial domain means a reduced AD/AB boundary at the time of carpel primordium emergence (approximately floral stages 3-4). This is indicated by a short black line (AD/AB boundary) at the apical surface of the incipient primordium (compare Fig. II-7Bi with Fig. II-7Ai) and a shorter gynoecium tube (Fig. II-7Bii). In support of an early role of AD/AB polarity in specifying gynoecium patterning, double mutants of the *KAN* gene family with compromised abaxial identity also exhibit similar gynoecium phenotypes to *ett* mutants (Eshed et al., 2001; Pekker et al., 2005).

Mutants defective in auxin polar transport (in *pin* or *pid* mutants, or by NPA treatment) exhibit weakened or absent auxin flows into the incipient carpel primordium (Fig. II-7Ci-iii), which will lead to a lack of a clear AD/AB boundary in the incipient carpel primordium indicated by a lack of the black line. As a result no valve or a reduced valve will form. Mutants of auxin biosynthesis (in *yuc1 yuc4* or *taa/tar* mutants) likely

have insufficient auxin to be transported toward the incipient primordium, resulting in the absence of AD/AB domains and hence a lack of gynoecium tube (Fig. II-7Di-iii).

In all auxin-pathway mutants (*yuc, taa/tar, pin, pid,* and *ett*), the severity of the defects caused by different alleles negatively correlates the extent to which an AD/AB boundary remains in the primordium. The stronger the defects, the smaller the AD/AB boundary is at the apex, and the smaller the valve. The resulting non-polarized zone at the base of the primordium may lead to a longer gynophore at the base. Gynophore elongation may be regulated by a separate growth mechanism that is related to the proximal-distal growth and independent of the AD/AB polarity.

This early-action model cannot explain why the *yuc1 yuc4* or *pin,* or *pid* mutants are still capable of developing almost normal amount of stigmatic tissues at the apex, other than by proposing that the stigma development may occur later, after the apical to basal patterning of gynocium is established. *STYLISH1/2* and *NGA3* transcription factors are known to activate the late-stage *YUC* gene expression required for stigma development (Sohlberg et al., 2006; Trigueros et al., 2009; Eklund et al., 2010). The fact that *yuc4 yuc1* double mutants still develop stigmatic tissues hints at additional redundancy in sources of auxin for the apex of the gynoecium. This redundancy could be caused by other *YUC* genes such as *YUC2,* which is expressed broadly in floral primordia (Cheng et al., 2006), or by upward transport of auxin via PIN1 localized to the replum cells (Grieneisen et al., 2013). As the replum represents the medial edge of the carpels, this pattern of upward transport is strikingly reminiscent of the Berleth (2007) model of

auxin's movement in aerial organs discussed earlier, which has auxin from the stem being transported up the leaf along its medial edges.

This early-action model could be evaluated experimentally by looking at the expression of genes in the AD/AB cohorts at very early stages of gynoecial development. Under this model, we would expect that *pin1*, *pid*, or *yuc1 yuc4* double mutants fail to show a clear AD/AB boundary in carpel primordia and that *ett* mutants express expanded adaxial-specific molecular markers and shrinking abaxial-specific markers due to adaxialization of carpels. In contrast, the Nemhauser apical gradient model does not imply such a result.

Conclusion

Sixteen years ago, Nemhauser et al. (2000) proposed the auxin gradient model to explain the apical-to-basal morphogenesis of the *Arabidopsis* gynoecium. While it is a highly attractive model, the auxin gradient, on which the Nemhauser model heavily relies, remains elusive and multiple observations made since are inconsistent with aspects of the model. Here, we have proposed an alternative model, the early-action model, based on three observations. One is the timing of the apical-to-basal patterning, which occurs much earlier than the observed auxin biosynthesis at the gynoecium apex. Another is the already-established evolutionary homology between carpel and leaf-like lateral organs. The third is the set of emerging models of auxin's role in leaf and lateral organ development, including the link between auxin transport, synthesis, and signaling and lateral organs' AD/AB boundary establishment. Our model emphasizes auxin's early effects on AD/AB boundary establishment as an explanation for the defects of gynoecium

in apical-basal patterning induced by auxin-disrupting mutations and chemicals. Furthermore, the early-action model unifies the development of carpels with current models of the development of other lateral organs.

Methods

Primers were designed to amplify the 1.4 kb upstream region starting from the *ETT* start codon (Fw: 5'-aaa aat tag atc tag ctc aca aat caa-3'; Rv: 5'-taa aga gag aga aac aga cat aaa gat-3'). The genomic DNA of *Arabidopsis* was extracted and used as a template for PCR with the *pETT* primers. The amplicon was cloned into the pCR8/GW/TOPO vector (Invitrogen) via TA cloning, and from there into the pMDC162 destination vector (Curtis and Grossniklaus, 2003). The construct was transformed into *Agrobacterium,* which was then used to infect *Arabidopsis* Landsberg erecta via floral dip. Transgenic plants were selected on hygromycin and the first generation T1 and second generation T2 plants were stained with X-gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid), the substrate of b-glucuronidase.

Infloresences were incubated in GUS staining buffer (0.2% Triton X-100, 50 mM $NaHPO_4$ (pH 7.2), 2 mM Potassium Ferrocyanide, 2 mM Potassium Ferricyanide), then vacuum infiltrated with 2 mM X-gluc in GUS staining buffer for 20 min. Samples were incubated with the X-gluc overnight (about 16 hrs), then taken through an ethanol series (20%, 35%, 50%, 30 min each), then fixed in FAA (50% ethanol, 5% formaldehyde, 10% acetic acid) for 30 min, then moved to 70% ethanol for 30 min.

The tissue was put through a futher ethanol series (80%, 90%, 30 min each), then stained with 0.1% eosin-y in 95% ethanol overnight. Tissue was incubated 30 min in

100% ethanol twice, then stepped into tert-butanol (12%, 50%, 100%), and then into paraplast, and incubated overnight at 60°C. Samples were sectioned at a thickness of 8 microns on a microtome. Samples were deparafinized with Hemo-De and mounted with Crystal Mount.

Chapter III: Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry

Abstract

Fragaria vesca is a species of diploid strawberry being developed as a model for the octoploid garden strawberry. This work sequenced and compared the genomes of three *F. vesca* accessions: Hawaii 4, Rügen, and Yellow Wonder. Genome-scale analyses of shared and distinct SNPs among these three accessions has revealed that Rügen and Yellow Wonder are more similar to each other than they are to Hawaii 4. Though all three accessions are inbred seven generations, each accession still possesses extensive heterozygosity, highlighting the inherent differences between individual plants even of the same accession. The identification of the impact of each SNP as well as the large number of indel markers provides a foundation for locating candidate mutations underlying phenotypic variations among these *F. vesca* accessions and for mapping new mutations generated through forward genetics screens. Through systematic analysis of SNP variants affecting genes in anthocyanin biosynthesis and regulation, a candidate SNP in *FveMYB10* was identified and then functionally confirmed to be responsible for the yellow color fruits made by many *F. vesca* accessions. As a whole, this study provides further resources for *F. vesca* and establishes a foundation for linking traits of economic importance to specific genes and variants.

Introduction

Fragaria vesca (F. vesca) is a diploid species of wild strawberry that has been cultivated in European gardens for centuries. *F. vesca* is currently being developed as a model species for the garden strawberry (Shulaev et al., 2011b), *Fragaria x anannasa*, as well as the diverse *Rosaceae* family that includes apple, peach, almond and rose. The garden strawberry has a complex octoploid genome and *Rosaceae* fruit trees often need 3 to 7 years of juvenile growth before flowering, making them difficult systems for genetic studies. In contrast, *F. vesca* offers a number of advantages as a research model. First, the genome of a 4th-generation inbred line of *F. vesca*, Hawaii 4x4, has been sequenced and

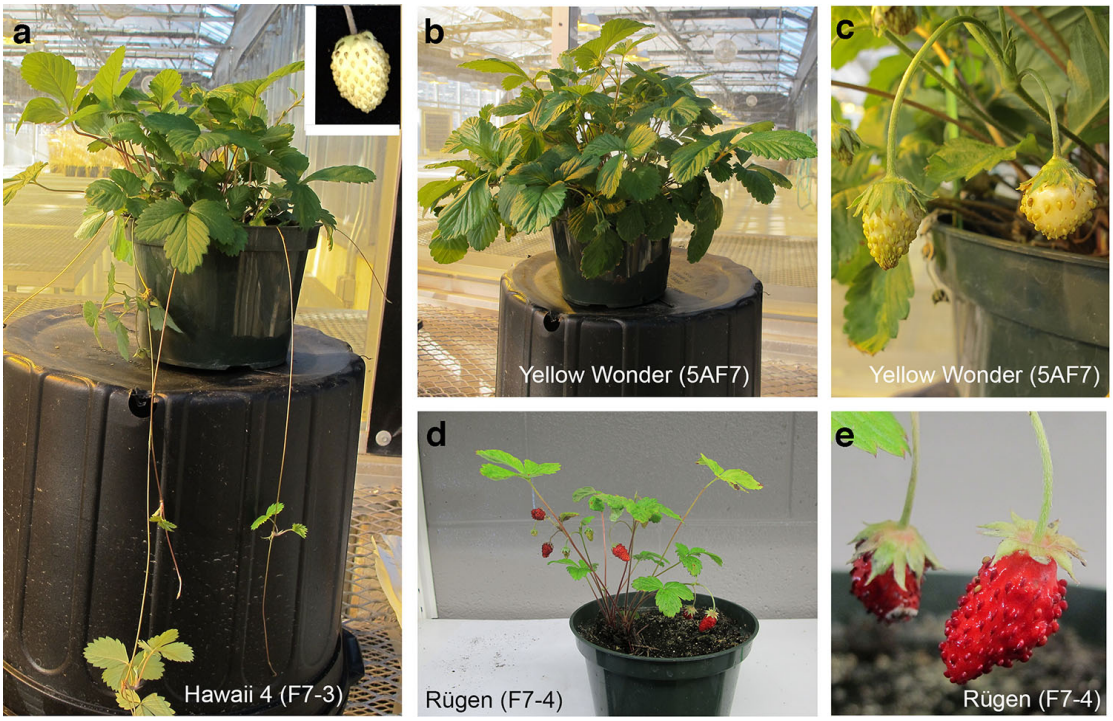

Figure III-1. Three *F. vesca* accessions used in this study.

(A). H4 plant showing runners and yellow fruit (inset). (B). YW plant showing a lack of runners. (C). A YW plant with yellow fruits. (D). Rü plant showing red fruits and a lack of runners. (E) Enlarged image of the red fruit from Rü. Image credit: Zhongchi Liu

served as the reference genome. It consists of 195 MB spread across seven linkage groups (2n = 14) (Shulaev et al., 2011b). Second, *F. vesca* has a short life cycle of 4-6 months, is small in stature, is self-fertile, and is amenable to transformation. Finally, detailed morphological characterization and extensive flower and fruit developmental transcriptomes have been published for *F. vesca*, providing extensive resources to serve as a starting point for further studies (Hollender et al., 2012; Darwish et al., 2013; Kang et al., 2013; Hollender et al., 2014; Darwish et al., 2015).

Three *Fragaria vesca* accessions (or varieties) have been developed for this purpose. They are Hawaii 4 (H4), Rügen (or Ruegen), and Yellow Wonder (YW) (Slovin et al., 2009; Shulaev et al., 2011b; Sun et al., 2014; Xu et al., 2014; Zhang et al., 2014). The three accessions each have a 7^{th}–generation inbred line, named H4 F7-3, Rügen F7-4, and YW5AF7 respectively. Although the three accessions are similar in their day-neutral flowering behavior (due to the *semperflorens* mutation) and in their physical stature and morphology, they also exhibit distinct characteristics. For example, H4 produces runners, a form of asexual reproduction, but YW and Rügen do not (Fig. III-1). In addition, Rügen produces red fruit, while YW5AF7 and H4 produce pale yellow fruit (Fig. III-1). Classical genetic experiments have shown that the runner trait in wild type *F. vesca* is controlled by a single locus (Brown T., 1965). Previous linkage analysis has shown that the runner locus maps to the end of LG2 (Davis and Yu, 1997). The same study, using a red fruited accession (Baron Solemacher) crossed with a yellow-fruited variety (Bush White), revealed that the locus responsible for the color difference is

located near the end of Linkage Group 1 (LG1) (Davis and Yu, 1997). Nevertheless, the identities of the genes responsible for the runner and fruit color traits have remained elusive.

The pigment that causes the characteristic red fruit color in strawberry is callistephin, a pelargonidin-based anthocyanin (Crozier et al., 2008). In addition to the enzymes that catalyze the synthesis of anthocyanins, several transcription factors have been shown to play critical roles in regulating anthocyanin production by controlling the expression of the biosynthesis genes. Chiefly, the MYB, bHLH, and WD-repeat proteins (which form the MBW complex) regulate the expression of the anthocyanin pathway genes in plants (Ramsay and Glover, 2005; Albert et al., 2014). The most well-studied are the R2R3 MYB proteins; alteration in the expression or function of a single such MYB gene can drastically alter the accumulation of anthocyanins in orange (Butelli et al., 2012), petunia (Schwinn et al., 2006), tomato (Butelli et al., 2008; Cermak et al., 2015), and peach (Tuan et al., 2015). The garden strawberry homolog of *MYB10*, *FaMYB10,* was previously identified and shown to play critical roles in regulating the red pigment in the receptacle fruit. *FaMYB10* is specifically expressed during the ripening stage of the receptacle fruit, and its expression was shown to be repressed by auxin and stimulated by ABA (Medina-Puche et al., 2014). Further, RNAi down-regulation of *FaMYB10* resulted in significant reduction of anthocyanin in the receptacle fruit (Medina-Puche et al., 2014). In contrast, over-expression of *FaMYB10* in the garden strawberry resulted in plants with elevated anthocyanin levels in roots, foliage, and fruit (Lin-Wang et al., 2010). In *F.*

vesca, an RNAi construct against *FveMYB10* was also shown to convert red fruit into pale yellow fruit (Lin-Wang et al., 2014), suggesting a similar role of *FveMYB10* in stimulating red pigment in wild strawberry fruits.

In *F. vesca*, both red- and yellow- fruited accessions exist. Zhang et al., (2015) showed that a high level of anthocyanins was present in the red fruit of Rügen from the turning stage to ripening, but not in the yellow fruit of YW5AF7 (Zhang et al., 2015). Further, transcriptome profiling in red (Rügen) and yellow (YW) accessions showed that transcript levels of several anthocyanin biosynthesis enzyme genes (C4H, CHS, CHI, F3H, DFR, and ANS) and several MYB genes (MYB1, MYB86 and MYB39) were reduced in the YW fruit (Xu et al., 2014; Zhang et al., 2015). However, these differentially expressed genes may reflect downstream effects of the causal mutation, and none of the studies conducted functional tests of SNPs that exist between yellow and red accessions.

In this study, we carried out genome-wide identification of variants among three *F. vesca* accessions, H4, YW, and Rügen. Both SNP variants and structural variants that distinguish the three accessions were identified, which enables the development of molecular markers and will aid in gene mapping and gene isolation. Interestingly, although these three accessions all have previously been inbred for seven generations, each accession still possesses a large number of heterozygous loci. To identify the causal mutation that underlies the naturally occurring yellow fruit color in the *F. vesca*

accessions H4 and YW, variants affecting exons of genes involved in anthocyanin biosynthesis and regulation were systematically examined and analyzed. Three SNPs affecting three different *FveMYB* transcription factors in the yellow fruit accessions were identified, and subsequent functional assays indicated that a single SNP (W12S) in the *FveMYB10* gene was responsible for the yellow color in these wild strawberry fruits.

Results

Genome-wide variant analyses reveal Hawaii 4 as genetically more distinct from Yellow Wonder and Rügen

To identify genome-wide variants among the three accessions of *F. vesca*, H4 F7-3, YW5AF7, and Rügen F7-4 (all 7th generation inbred lines), the genomic DNA of each was sequenced using the Illumina HiSeq2000 platform. Single-end, 51-bp reads were obtained for H4 F7-3, YW5AF7, and Rügen F7-4 with 106, 84, and 80 million high quality reads, respectively (Appendix A, Supplemental Table S1). Subsequent sequence analyses (outlined in Appendix A, Supplemental Fig. S1) yielded a genome-wide total variant list with information on each SNP variant and its impact on previously-annotated genes (Appendix A, Supplemental Table S2). Appendix A, Supplemental Table S2 serves as the starting point for further filtering to identify accession-unique variants (Appendix A, Supplemental Table S3), heterozygous variants (Appendix A, Supplemental Table S4), and "high impact" variants in each of the above two categories (Appendix A,

Supplemental Table S2-4). For simplicity, H4, YW, and Rü will be used to refer to the respective 7th-generation inbred lines.

The "accession-unique variants" (Appendix A, Supplemental Table S3) allow us to investigate the relationships among the three accessions. Variants are considered unique to an accession if they are present in one accession but absent from the other two. An accession-unique variant may either exist as homozygous or heterozygous within the accession. We found that H4 has 99,722 unique loci, Rü has 38,404 unique loci, and YW has 42,483 unique loci (Appendix A, Supplemental Fig. S2). Additionally, YW and Rü have more SNPs in common (Appendix A, Supplemental Fig. S2), indicating that YW and Rü are more similar to each other than they are to H4.

We plotted the density of accession-unique loci along each of the seven Linkage Groups (LGs; i.e., chromosomes) (Fig. III-2A). There is considerable variation in the distribution of these loci. LGs 2, 3, and 7 are particularly rich in H4-unique variants, suggesting independent origins of these three LGs in H4. This is consistent with the runner locus being previously mapped to LG2 (Davis and Yu, 1997), as H4 is the only one among the three accessions that exhibits the runner trait. Next, LG1 is the only LG rich in variants unique to Rü. This is consistent with LG1 containing the color locus (Davis and Yu, 1997) for which Rü is the only accession among the three that carries the red allele. Third, LG4 is rich in SNPs unique to YW. All five of the other chromosomes (LG2, 3, 5, 6, 7) are largely shared between Rü and YW (Fig. III-2A). Perhaps introgression of LG1 and LG4, respectively, into a plant ancestral to Rü and YW may underlie the modern day Rügen and Yellow Wonder accessions. Finally, LG5, LG6, and

a roughly 5 Mb stretch of LG7 contain very few loci unique to any accession; the total combined variants (unique and shared) are relatively low in these regions as well (Appendix A, Supplemental Fig. S3), suggesting that these three accessions may have shared history and/or recent crossbreeding.

Analysis of the impact of each variant on protein coding genes

In order to determine the impact of each SNP on protein coding genes, we utilized the snpEff v4.0 program. Of the 366,057 total variants annotated by snpEff v4.0 (Appendix A, Supplemental Table S2), 310,266 were found to be within 5 kb of at least one gene, 106,862 are within introns, 14,220 are silent coding sequence variants, and 26,437 have "high to moderate impacts" on the coding sequence (Appendix A, Supplemental Table S2). Moderate-impact variants are single-residue missense variants that do not affect the start or stop of translation. In contrast, high impact variants are those whose changes to the coding sequence may result in a knockout of the affected gene. They include frameshifts, loss of "start", loss or gain of "stop", and loss or gain of splice sites. 3,030 were categorized as high-impact variants (Appendix A, Supplemental Table S2). High-impact variants unique to each accession were also extracted and plotted across all linkage groups (Fig. III-2B; Appendix A, Supplemental Table S3). Although these high-impact variants are significantly lower in number, their distribution pattern largely mirrors the unique variant distribution (Fig. III-2A) except that the beginning of LG7 is abundant with high impact variants unique to H4.

Gene Ontology (GO) (Conesa et al., 2005) enrichment analysis was performed on the total combined high impact variants using Blast2GO. The enriched GO categories were primarily related to DNA binding and DNA replication (Appendix A, Supplemental Fig. S4), indicating that essential functions in cell survival were preferentially affected by the high impact variants. Other enriched categories appear to be related to RNA-dependent DNA replication, suggesting the potential of rapid evolution of RNA-mediated regulatory mechanisms.

All three varieties exhibit a high level of heterozygosity

A highly homozygous genome is desirable for molecular genetic studies. H4 F7-3, YW5AF7, and Rügen F7-4 are all 7th generation inbred lines (see Methods). While the reference genome derived from H4x4 consists of a single sequence file, we found that multiple alternative bases existed (Appendix A, Supplemental Table S2). In fact, a significant number of loci in the genomes of all three accessions were heterozygous, with H4 showing 90,453 such loci, Rü showing 119,693 heterozygous loci, and YW showing 121,789 heterozygous loci (Appendix A, Supplemental Fig. S2; Appendix A, Supplemental Table S4). The density of these heterozygous SNPs across the genome was plotted (Fig. III-3A). Across most of the genome, heterozygosity is similar among the three accessions. However, H4 shows lower levels of heterozygosity than the other two in some regions, most notably at the end of LG3 and at the beginning of LG2 (Fig. III-3A).

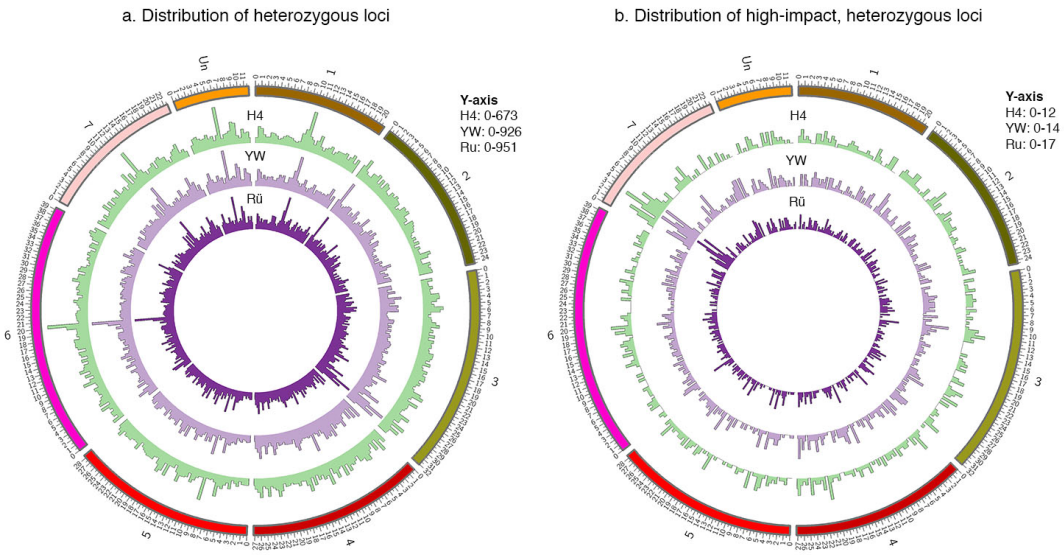

Figure III-3. Genome distribution of heterozygous variants and heterozygous high impact variants.

(A). Circos histogram showing distribution of heterozygous loci in each accession. (B). Circos histogram showing distribution of heterozygous and high impact loci in each accession. Bin size is 500 kb, Y-axis is the number of heterozygous or high-impact/heterozygous variants in that bin. The outermost circle represents the seven *F. vesca* Linkage Groups plus the unanchored scaffolds (Un), drawn to scale in Mbp.

On the other hand, YW shows slightly higher levels of heterozygosity in LG4 than the other two accessions.

One hypothesis for the persistence of these heterozygous loci even after seven generations of single-seed breeding is that heterozygosity may balance high impact SNPs to prevent homozygous lethality. We extracted heterozygous loci from the high impact variant list (Appendix A, Supplemental Table S2) to yield heterozygous high impact variants (Appendix A, Supplemental Table S4). We then plotted them across the seven LGs (Fig. III-3B). Compared with total heterozygous loci (Fig. III-3A), the high impact

heterozygous variants are significantly lower in number but nevertheless are distributed throughout all seven LGs (Fig. III-3B); only LG7 shows a peak region near the beginning of the chromosome. GO term analysis of high impact heterozygous variants could not identify statistically significant categories due to the small number of genes. Hence, we examined the percentage of high impact SNPs among all SNPs vs. the percentage of high impact SNPs among heterozygous loci. We found that 0.87% of all loci are high-impact, while 0.73%, 0.79%, and 0.76% of loci heterozygous in H4, Rügen, and YW5AF7, respectively, are high-impact. Thus, there is no enrichment of high impact SNPs among heterozygous loci, indicating that our prior hypothesis is incorrect.

Identification of Indels and structural variants

In addition to SNP variants, we conducted genome-wide detection of large insertions and deletions (Indels) (Appendix A, Supplemental Table S5) and structural variants (Appendix A, Supplemental Table S6) in the three accessions using CLC Genomics Workbench. Because of the short read length of 51bp and because the reads are unpaired, many large structural variants may be missed. Nevertheless, we developed a set of accession-specific markers based on the Indel information (Appendix A, Supplemental Table S5). Each Indel marker can distinguish one of the three accessions based on a simple PCR reaction (Fig. III-4). Together, this set of Indel markers can easily distinguish each of the three accessions and can be used to confirm and distinguish hybrids among these three accessions. The variant lists (Appendix A, Supplemental Table S3 for SNPs, Appendix A, Supplemental Table S5 for Indels, Appendix A, Supplemental Table S6 for

structural variants) provide useful information for the development of region-specific markers for gene mapping and cloning.

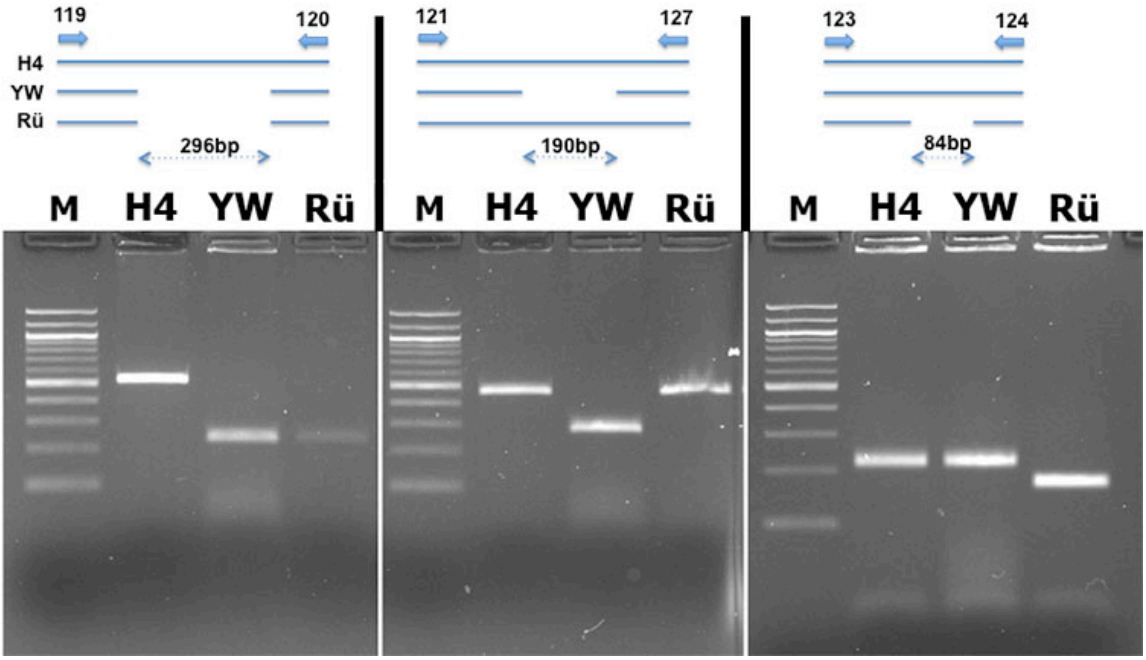

Figure III-4. Indel markers that distinguish the three *F. vesca* accessions.

Three Indel markers have been developed. Corresponding PCR primer pairs 119/120 (LG2), 121/127 (LG4), and 123/124 (LG6) are respectively indicated as small arrows. Dotted arrow indicates deletion. The agarose gel images show variety-specific banding pattern. Genomic DNA from H4, YW, and Rü served as PCR templates. M is the 100 bp ladder.

Identification of candidate SNPs responsible for the yellow fruit color

The red fruit pigment in strawberry is Callistephin (Pelargonidin-3-O-Glucoside), the biosynthetic pathway of which is well established (Fig. III-5A)(Crozier et al., 2008; Li, 2014). We investigated the molecular basis underlying the red vs yellow fleshy fruit in

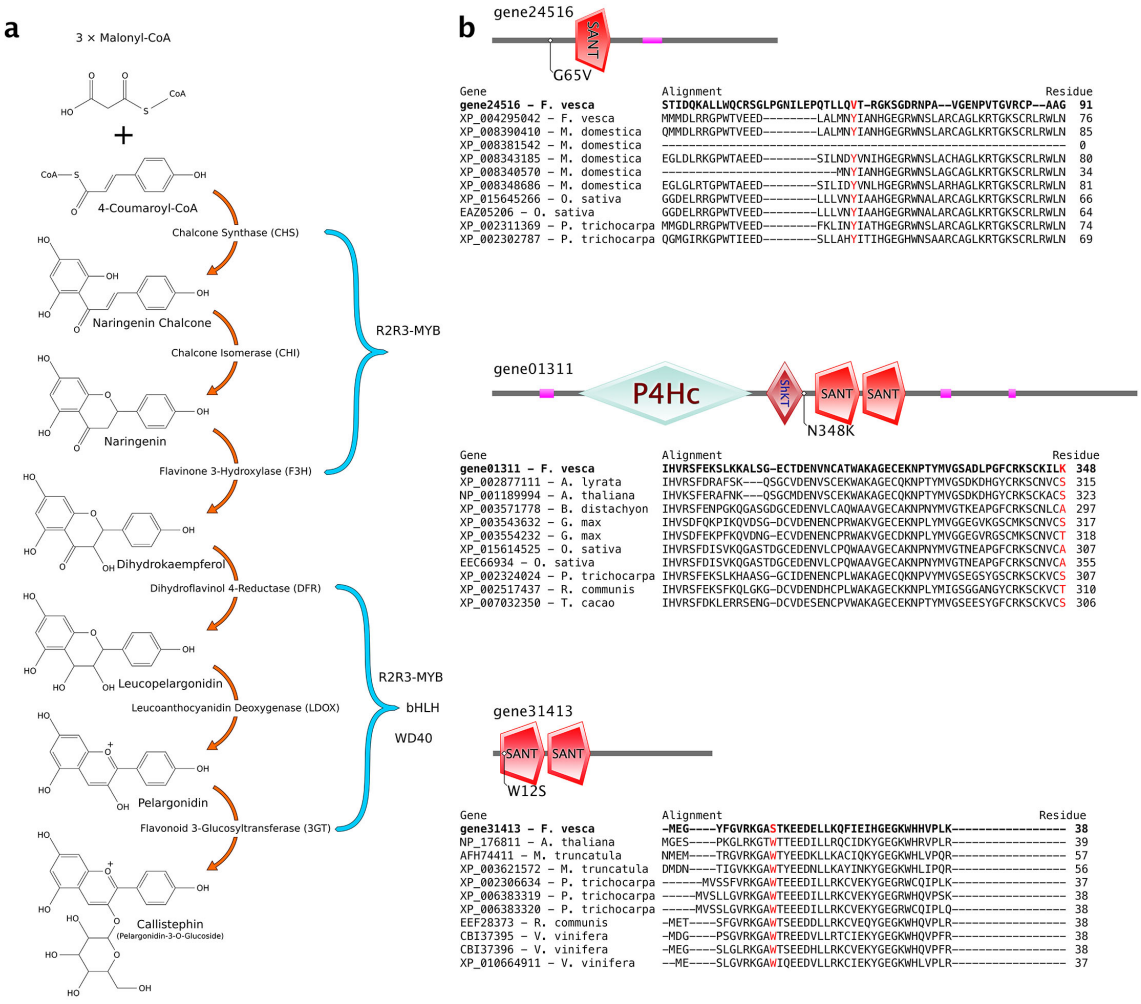

Figure III-5. Canonical pelargonidin biosynthesis pathway and sequence alignment of candidate MYB genes.

(A). Canonical pelargonidin biosynthesis pathway showing enzymes and intermediate pigments in each step. Steps regulated by R2-R3 MYB transcription factors are also indicated. This pathway illustration is based on a published paper (Li, 2014). (B). MYB protein domain structure and the location of the variant in each of the candidate *FveMYB* genes. Sequence alignment of the protein region affected by the concerned variant (red) is shown for each *MYB* gene.

these three accessions. Red-fruited Rü was crossed to yellow-fruited YW, and the F1 progeny produced red-fruit (Fig. III-6A), indicating that the red color is dominant and the yellow color likely results from a loss-of-function mutation. A cross between the two yellow accessions (H4 and YW) resulted in yellow fruit (Fig. III-6A), indicating that H4 and YW are defective in the same gene as they failed to complement each other.

To identify a potential SNP responsible for the yellow fruit color, we first assembled a short list of likely candidate genes based on prior knowledge of fruit pigment biosynthesis and regulation including genes coding for enzymes of pelargonidin biosynthesis as well as the R2R3 MYB-class transcription factors (Fig. III-5A; Appendix A, Supplemental Table S7). Then, we searched for variants that are of high or moderate impacts as well as unique to Rü. The resulting variant list was intersected with the pigment biosynthesis and regulation gene list (Appendix A, Supplemental Table S7). Manual filtering by checking mapped reads and determining homozygosity resulted in eight SNPs (Appendix A, Supplemental Table S8); two reside in 3GT (Flavonoid 3-Glucosyltransferase) genes, two in MYB-like or MYB-related genes, and four reside in the *FveMYB* genes. Among these eight genes, only three *FveMYB* genes possess variants that cause amino acid substitutions from one property type to another (Appendix A, Supplemental Table S8). Specifically, gene01311 has an Asn (N) to Lys (K) substitution (polar to basic) at position 348, gene24516 has a Gly (G) to Val (V) substitution (polar to nonpolar) at position 65, and finally gene31413 exhibits a Trp (W) to Ser (S) substitution

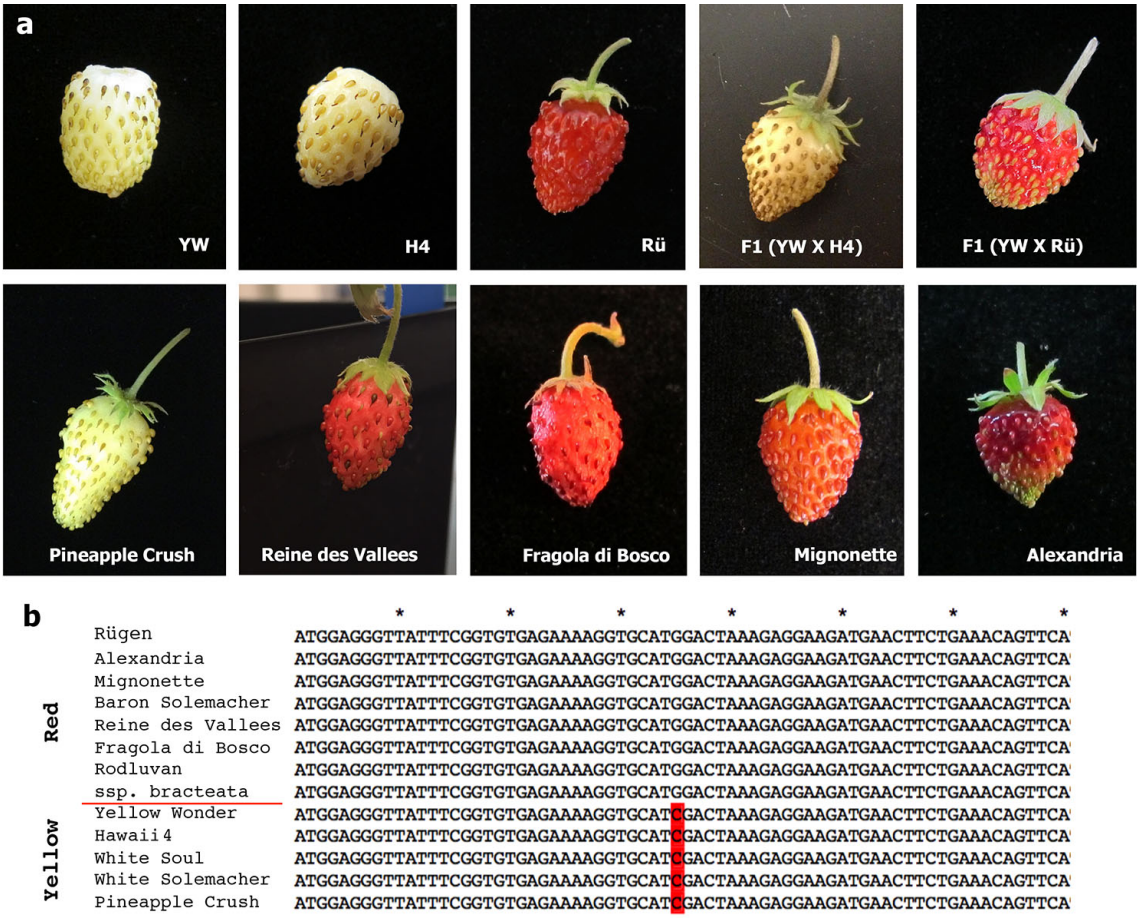

Figure III-6. Genetics of fruit color trait and survey of the *FveMYB10* W12S variant in red and yellow *F. vesca* accessions.

(A). Fruit phenotype from different *F. vesca* accessions including F1 progeny of crosses between red and yellow accessions. (B). Sequence of a segment of *FveMYB10* DNA in red and yellow fruit accessions. The specific SNP (G to C) that converts the W to S at position 12 is highlighted. All red accessions have G, while all yellow accessions have C. Image credit: Julie

(nonpolar to polar) at position 12. To narrow down which SNP is more likely the candidate SNP, we sought to determine if any of the SNPs affects a conserved residue (Fig. III-5B). Neither the SNP in gene01311 nor the SNP in gene24516 affects a conserved residue. In contrast, the SNP in gene31413 affects a highly conserved residue

W within the R2 DNA-binding domain (Fig. III-5B). In addition, gene31413 was previously named as *FveMYB10* and shown by RNAi to cause yellow fleshy fruit in *F. vesca* (Lin-Wang et al., 2014). Therefore, the specific W12S SNP in gene31413 (*FveMYB10*) emerged as the primary candidate for causing the yellow fruit color in the woodland strawberry.

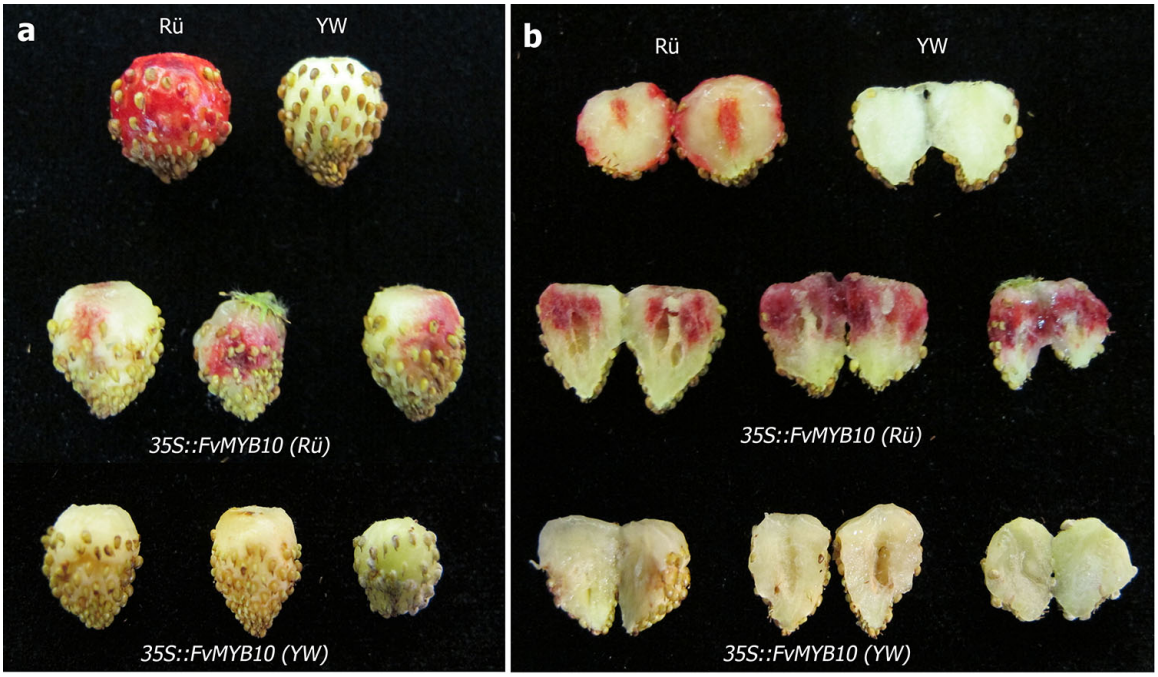

Figure III-7. Functional test of the *FveMYB10* W12S variant in fruit color determination

(A). Images of control fruits (top row) and fruits of YW5AF7 injected with *Agrobacterium* containing the *35S::FveMYB10 (Rü)* (middle row) or *35S::FveMYB10(YW)* (bottom row). Three representative fruits from each construct are shown. (B). Same fruits from (A) were cut in half to reveal their interior color. Image credit: Julie Caruana

Experimental test of the effect of W12S on fruit color

We first tested if the W12S SNP co-segregates with the yellow fruit color in three other yellow-fruited *F. vesca* accessions, White Soul, Pineapple Crush, and White Solemacher. Pair-wise genetic crosses were made between H4 and each of these three yellow fruited accessions. The F1 progeny all produced yellow fruits, indicating that these yellow accessions are all defective in the same gene as H4. PCR amplification and sequencing of the *FveMYB10* gene from these additional yellow-fruited accessions revealed that all of them had the C nucleotide and thus the W12S substitution in *FveMYB10* (Fig. III-6B). Seven additional accessions or subspecies that produced red fruits were similarly analyzed; they were Alexandria, Mignonette, Baron Solemacher, Reine des Vallees, Fragola di Bosco, Rodluvan, and *ssp.* bracteata. All of these red-fruited accessions/subspecies were shown to possess the wild type G nucleotide in *FveMYB10* (Fig. III-6B). Hence, the C (W12S) in *FveMYB10* co-segregates with the yellow fruit color, while the G co-segregates with the red fruit color.

Next we carried out a functional assay to test the ability of *FveMYB10* to restore red color in the YW yellow fruit. Full-length *FveMYB10* cDNAs were isolated from Rü (red) and YW (yellow), respectively. Sequence analysis confirmed that the only difference between these two *FveMYB10* cDNAs is the G in the Rü (red) cDNA and C in the YW (yellow) cDNA. Subsequently, the Rü and YW cDNAs were respectively cloned behind the 35S promoter in the pMDC32 vector. *Agrobacterium tumefaciens* containing each construct was injected into the developing fruit of YW. Out of nineteen yellow fruits injected with the *Agrobacterium* containing *35S::FveMYB10(Rü)*, sixteen developed red

pigmentation at the injection sites (Fig. III-7A, B) and three did not survive. Of the twelve YW fruits injected with *Agrobacterium* containing *35S::FveMYB10(YW)*, 10 survived the treatment but none developed red pigment (Fig. III-7A, B). As the only difference between the Rü cDNA and YW cDNA is the single SNP, we concluded that the G to C change (W12S substitution) renders *FveMYB10* nonfunctional and determines the yellow fruit color in many of the wild *F. vesca* accessions.

Discussion

While the exact history and origin of the three *F. vesca* accessions under study are unclear, all three are perpetual flowering (*F. vesca semperflorens*). The perpetual flowering *F. vesca* was described in 1766 and introduced into gardens all over Europe at that time (Duchesne, 1766). The specific mutation responsible for the perpetual flowering was recently shown to result from a 2 bp deletion in the *FveTFL* gene (Iwata et al., 2012; Koskela et al., 2012). To date, H4, YW, and Rü are frequently used in research due to the availability of the H4 genome sequence and the continuous flowering habit.

Genome-wide identification and analysis of variants among the three varieties provided detailed information on SNP variants (Appendix A, Supplemental Table S2), Indels (Appendix A, Supplemental Table S6), and structural variants (Appendix A, Supplemental Table S7). Combined, they will greatly aid in gene mapping and gene isolation. In addition, determination of total and accession-unique high impact variants (Appendix A, Supplemental Table S2 and S3) provides the starting point to examine genetic changes and the resulting phenotypic differences. The availability of the genome sequence reads of the three accessions submitted to NCBI-SRA (SRP068157) and the

information about their heterozygosity will be of significant interest to researchers wishing to use *F. vesca* to map and isolate genes. Together, the various genomic analysis data reported here help further establish *F. vesca* as a model system.

One surprising finding from this study is the high level of heterozygosity for all three accessions despite each being inbred for seven generations. Specifically, Rügen (F7-4) has 119,693 heterozygous loci, YW5AF7 has 121,789 heterozygous loci, and Hawaii 4 (F7-3) has 90,453 heterozygous loci (Appendix A, Supplemental Fig. S2). The observed heterozygosity may even be an underestimate due to the pooling of only two individual plants per accession for sequencing. Heterozygosity was reported for the soybean reference cultivar Williams 82 after six generations of testcross followed by one generation of self cross (Haun et al., 2011). However, about 1800 heterozygous loci were found in Williams 82. Our unusually high level of heterozygosity in *F. vesca* accessions indicates that the original accessions are likely hybrids of distantly related parents.

Theoretically, every generation of inbreeding should reduce heterozygosity by half; 7^{th} generation inbred plants should have reduced heterozygosity by 128 times compared to the progenitor plant. With current number of approximately 100,000 heterozygous loci spread across the 200 Mb genome (ie. one heterozygous locus every 2 kb), the original strain would be predicted to have one heterozygous variant every 16 bp (2 kb divided by 128), which seems too high. An alternative and more reasonable interpretation is that the inbreeding process has not been efficient and did not cut down heterozygosity by half as predicted. Perhaps, heterozygosity offers significant hybrid vigor, which could be

selected for during the inbreeding process. A second possibility, not supported by our data, is that heterozygosity buffers deleterious effects caused by the high impact SNPs.

The high level of heterozygosity in *F. vesca* accessions raises the question of which nucleotide, for each heterozygous locus, is represented by the reference genome (Shulaev et al., 2011b). As a consequence, the inherent differences between individual plants should always be considered when utilizing the reference genome to design PCR primers, develop molecular markers, conduct sequence comparisons, and perform BSA (Bulk-Segregant-Analysis)-Seq. For example, a mapping population developed from a single F1 individual is advisable when conducting BSA-Seq, and analysis filters should be used to remove heterozygous SNPs before gene mapping during BSA-Seq. To further the usefulness of *F. vesca* as a model plant, it may be necessary to develop double haploid lines, perhaps by applying the CENH3-based haploid inducers (Ravi and Chan, 2010).

Although *MYB10* is known to regulate red pigment synthesis in both diploid and octoploid strawberry fruits (Lin-Wang et al., 2010; Lin-Wang et al., 2014; Medina-Puche et al., 2014), it is not known if the yellow fruit made by *F. vesca* is caused by a defective *FveMYB10* gene. Mutations in genes that regulate *FveMYB10* or genes regulated by *FveMYB10* could also lead to yellow fruit in these *F. vesca* accessions. Therefore, a systematic search of potential candidate genes was necessary, which led to three *FveMYB* genes harboring nonsynonymous mutations (Table S8). However, only the SNP in *FveMYB10* affected the conserved DNA-binding domain (Fig. III-5B) and thus emerged as the primary candidate. Given that RNAi-mediated knockdown of *FveMYB10* was sufficient to cause yellow fruit (Lin-Wang et al., 2014), a demonstration of a non-

functional FveMYB10 (W12S) in YW should provide strong evidence for this variant as the causal mutation leading to the yellow fleshy fruit. Our transient functional assay clearly showed that this W12S in *FveMYB10* abolished *FveMYB10's* ability to restore red pigment in the yellow fruit of YW. Interestingly, we observed that the interior receptacle tissue developed red color after being injected with *35S::FveMYB10(Rü)* (Fig. III-7B) despite that Rü normally does not develop red pigment there. It is likely that *FveMYB10* is normally expressed only in the outer surface of the receptacle, while ectopic *FveMYB10* expression driven by the 35S promoter in the injected portion of the receptacle was responsible for the interior red color.

Additional biochemical and genetic evidence supports a causal role for W12S FveMYB10 in yellow fruit color. In previous biochemical and structural studies of MYB proteins in animals, each of the three conserved tryptophan (W) residues within the R domain was shown to be critical in maintaining the hydrophobic core of the R-domain and hence its DNA-binding function (Kanei-Ishii et al., 1990; Ogata et al., 1992). Therefore, the W (hydrophobic) to S (polar) change in the R2 DNA-binding domain of FveMYB10 likely disrupts the DNA-binding function of R2. Second, previous genetic mapping showed that the fruit color locus in Bush White (another yellow-fruited accession) resides on LG 1 (Davis T, 1997). Based on the newest gene annotation (version 2.0a1), *FveMYB10/*gene31413 resides on LG1 (between 15,405,782bp and 15,407,498bp). The FveMYB10 (W12S) variant most likely arose from a spontaneous mutation in an ancestral *F. vesca* that gave rise to the yellow colored fruits of many *F. vesca* accessions. Given that red fruit is important for attracting birds for seed dispersal, it

is intriguing why the yellow fruited *F. vesca* accessions exist in nature. Our work reported here provides an example of utilizing genomic comparisons to connect traits of economic importance to specific genes and variants.

It has recently been shown that many of the biosynthetic genes that catalyze the callistephin pathway (C4H, CHS, CHI, F3H, DFR, and ANS) are expressed at significantly lower levels in Yellow Wonder than in Rügen (Zhang et al., 2015). This is interesting if the causal mutation is indeed the W12S in FvMYB10, as FvMYB10 most closely resembles Arabidopsis MYB113, an R2R3 MYB that regulates the later steps in the biosynthetic pathway: DFR, LDOX and 3GT (reviewed in Li, 2014). One possible explanation is that FvMYB10, unlike MYB113, regulates both early and late steps of the anthocyanin biosynthesis pathway. Another possibility is that FvMYB10, like MYB113, regulates the later steps of the pathway. However, buildup of intermediates such as dihydrokaempferol may have resulted in the down-regulation of the earlier biosynthetic genes via a negative feedback mechanism. Zhang et al., (2015) also found lowered expression of several other MYB genes that positively regulate the biosynthesis pathway along with heightened expression of MYB 1R1, a homolog of which can suppress accumulation of anthocyanins in tobacco (Nakatsuka et al., 2013). These findings are consistent with the latter hypothesis of negative feedback.

Methods

Plant materials and DNA sequencing

A 7th generation inbred line of Hawaii 4 (H4 F7-3), National Germplasm Repository ID PI664444, was used in this study. Similarly, a previously described 7th generation inbred line of Rügen (Rü F7-4) (Sun et al., 2014; Xu et al., 2014) was used. Yellow Wonder 5AF7 (PI641092) is also a 7th generation inbred line described previously (Slovin et al., 2009; Hollender et al., 2012) . Other strawberry accessions shown in Fig. III-6 were purchased from The Strawberry Store (http://thestrawberrystore.com); these varieties were verified for their fruit color phenotype and were crossed with H4 F7-3 or YW5AF7 for complementation tests on the fruit color trait. For DNA sequencing, young leaves were harvested from two individual plants of each variety, and genomic DNA was isolated using the NucleoSpin® Plant II kit (Macherey-Nagel, Duren, Germany). DNA samples were mailed to the Genomics Research Core Facility at the Weill Cornell Medical College for library preparation and sequencing on the Illumina HiSeq2000 instrument. The resulting single-end, 51-bp read statistics are summarized in Appendix A, Supplemental Table S1.

For sequence analysis of *FveMYB10* from yellow and red strawberry varieties shown in Figure 6, the gene fragment was amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) with primers 5' GCTCAAATATAGGTAACGTCAATACTC 3' (forward) and 5' CAGACGTAAACATACATATAAGCAGC 3' (reverse). PCR fragments were purified

using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) and sequenced at Macrogen (Rockville, MD).

Variant analysis pipeline

The variant analysis pipeline is summarized in Appendix A, Supplemental Figure S1. Briefly, raw sequence reads were mapped to the reference genome of H4, genome-wide SNP variants were called through the GATK program (McKenna et al., 2010), and the resulting SNP variant list was further annotated by the snpEff v4.0 program (Cingolani et al., 2012) to generate an annotated total SNP variant list (Appendix A, Supplemental Table S2).

Specifically, the reference genome of *F. vesca* Hawaii 4 (version 1.1) and the genome annotation file (Fvesca_226_gene.gff3) were downloaded from Phytozome (phytozome.jgi.doe.gov) (Goodstein et al., 2012). Raw sequence reads were mapped to the reference genome using BWA ((Burrows-Wheeler Aligner; Li and Durbin, 2009) after indexing with Picard (http://broadinstitute.github.io/picard/) and SAMtools (Li et al., 2009). The GATK software package ((www.broadinstitute.org/gatk/; McKenna et al., 2010) was used for variant detection. The genome annotation file (Fvesca_226_gene.gff3) representing 32,831 genes needed modification to fix the exon offsets using the custom script "fix-gff.awk" (Appendix A, Supplemental Methods). The modified gene annotation list and the GATK-derived raw variant list were fed into snpEff v4.0 (snpeff.sourceforge.net) (Cingolani et al., 2012), which predicted the impact of

variants on protein-coding genes and yielded the annotated variant list, in which each variant was noted for its impact.

Two filtering steps were then applied to the annotated variant list to eliminate variants with low quality reads as well as variants with reads >50 as regions with abnormally high read abundance may reflect repetitive or questionable regions. To achieve this, BedTools (Quinlan and Hall, 2010) was used to create a map of read depth across the genome, custom script "extractHCRanges.awk" (Appendix A, Supplemental Methods) was used to mark regions with reads >50, and custom script "cutReads" (Appendix A, Supplemental Methods) was used to remove all SNPs located in such regions (reads >50). This filtering step led to the final variant list (Appendix A, Supplemental Table S2). Subsequently, moderate and high impact variants were extracted from the annotated variant list using the "grep" unix command. After the same filtering steps (described above) that eliminate low quality reads and reads >50, the moderate and high impact variants are shown in separate sheets of Appendix A, Supplemental TableS2.

To identify variants unique to each accession, a custom Awk script "uniqueVariants.awk" (Appendix A, Supplemental Methods) was used to filter the annotated variant list, yielding variant unique to each of the three accessions (Appendix A, Supplemental Table S3). ClicO FS (Circular Layout Interactive Converter Free Services), a web-based interface incorporating the Circos circular genome visualization

tool (Krzywinski et al., 2009; Cheong et al., 2015), was used to plot accession-unique SNP variants shown in Figure 2A.

To identify SNP variants heterozygous in each of the accessions, a custom script "heterozygousVariants.awk" (Appendix A, Supplemental Methods) was used that yielded Appendix A, Supplemental Table S4. Since two individual plants (four haploid genomes) of each cultivar were pooled and sequenced, a locus was judged heterozygous in an accession if GATK found more than one allele at that locus within the pool of four haploid genomes. The clicO FS web tool was used to plot heterozygous SNP variants shown in Figure 3A.

Finally, the high impact variants identified above were filtered using custom script "uniqueVariants.awk" to generate high impact variants that is also unique to each accession (Appendix A, Supplemental Table S3 as separate sheets. Similarly, the high impact variants were filtered via script "heterozygousVariants.awk" to generate high impact and heterozygous variants (Appendix A, Supplemental Table S4 as separate sheets). To generate Appendix A, Supplemental Table S3 and S4, we removed SNPs located in regions with reads >50 or reads of low quality as we did with Table S2. The clicO FS web tool was used to plot "high impact unique variants" and "high impact heterozygous variants" in Figure 2B and 3B, respectively.

Several sets of genes were tested for enriched GO categories using Fisher's Exact Test ($p < 0.05$, FDR filtered) performed by the Blast2GO software v3.1.3 (Conesa et al., 2005). Results were reduced to the most specific categories. The set of genes with high-impact variants was tested against the set of all annotated genes in the genome (Appendix A, Supplemental Fig. S3).

Identification of potential variants responsible for the lack of anthocyanin in fruits of YW and H4

Members of anthocyanin biosynthetic genes were identified by their annotation in the Plaza 2.5 database (http://bioinformatics.psb.ugent.be/plaza/) [43]. The *MYB* family transcription factors were extracted from the published list of transcription factor genes in *F. vesca* (Kang et al., 2013). The resulting combined list is Appendix A, Supplemental Table S7.

To identify the candidate mutation responsible for the yellow fruit color in YW and H4, SNPs that are unique in Rü were identified from Appendix A, Supplemental Table S3. Next, SNPs with nonsynonymous changes in protein-coding regions (marked as "MODERATE" or "HIGH" impact by snpEff) were identified from Appendix A, Supplemental Table S2. Subsequently, SNPs fulfilling both of the above requirements and at the same time residing in one of the anthocyanin gene list (Appendix A, Supplemental Table S7) were identified. There were only eight genes affected by such a SNP (Appendix A, Supplemental Table S8). For sequence alignment (Fig. III-5B), each

of the three FveMYB candidate genes served as the query in blasp within Plaza 2.5. Each hit was then identified with an NCBI accession using NCBI BLAST. The top 10 hits that could be identified were aligned with the query using ClustalX. Each gene structure was drawn based on diagrams from SMART (http://smart.embl-heidelberg.de/) [44].

Identification of indels and structural variants and PCR genotyping

To identify indel variants larger than 10 bp in length (the upper limit for GATK), we used the "Indels and Structural Variants" tool in the CLC Genomics Workbench. Illumina reads were first mapped to the *F. vesca* genome (version 2.0) using the following parameters: mismatch cost - 2, insertion cost – 3, deletion cost – 3, length fraction - 0.5, similarity cost - 0.9, no global alignment, and non-specific matches were mapped randomly. Next, the "InDels and Structural Variants" tool was used with default parameters to identify large variants shown in Appendix A, Supplemental Table S5 and S6.

Based on the Indel variant table (Appendix A, Supplemental Table S5), we identified potential indel markers that are unique in at least one accession. Primer sequences and amplicon lengths for each genotype are shown in Appendix A, Supplemental Table S9. PCR was carried out using AccuStart™ II PCR ToughMix (Quanta Biosciences; Gaithersburg, MD) at the following reaction conditions: 94°C for 3 min initial denaturation, 35 cycles of [94°C for 30 sec, 58°C for 30 sec, 72°C for 30sec],

and final extension of 72°C for 5 min. PCR products were visualized on 1% agarose gel using DNA SafeStain (Lamda Biotech, St. Louis MO).

Plasmid construction and functional assays in fruits

RNA was isolated from "white" stage (19 days post-anthesis; Kang et al., 2013) of Rü and YW fruits using the Qiagen RNEasy Plant Mini kit (Qiagen). Approximately 200 ng of total RNA was used to generate cDNA using the iScript cDNA Synthesis kit (Bio-Rad). The *FveMYB10* coding sequence was PCR amplified from both Rü and YW cDNA using primers 5'ATGGAGGGTTATTTCGGTGTGAG 3' (forward) and 5' TACGTAGGAGATGTTGACTAGATCATTGC 3' (reverse) and cloned using the pCR™8/GW/TOPO® TA Cloning Kit (Invitrogen). After sequencing to confirm the correct CDS sequence for both Rü and YW clones, each was independently recombined into the pMDC32 binary vector (Curtis and Grossniklaus, 2003) using the Gateway® LR Clonase® II Enzyme mix (Invitrogen) to produce the overexpression vectors.

The transient expression assay in *F. vesca* fruits was carried out according to the procedure used for *F. x ananassa* (Hoffmann et al., 2006) with slight modifications. *Agrobacterium tumefaciens* strain GV3101 containing the 35S::*FveMYB10* vector was grown overnight to an O.D. of 0.8-1.0, then resuspended at a final O.D. of 0.8 in liquid MS medium containing 2% sucrose. A sterile 1 mL hypodermic syringe was used to evenly inject the suspension into the cortex of the fruits. Fruits ranging from "white" to "turning" (22 days post anthesis) stages were injected once and were left on the plants to

continue growth. Patches of red color appeared 3-14 days after injection, with younger fruits taking longer to show red color. Mature, ripe fruits were not suitable for injection.

Accession numbers

Raw Illumina DNA-sequencing files of Rügen F7-4, YW5AF7, and H4 F7-3 have been submitted to NCBI SRA with the accession number SRP068157.

Chapter IV: An eFP-based Tool for the Visualization of Strawberry Fruit and Flower Transcriptomes

Introduction

Microarray and RNA-seq experiments are generating an ever-increasing body of transcriptome datasets. While most of these are made publicly available, it is desirable to make analysis possible without requiring the use of complicated and specialized bioinformatics tools and pipelines. There is considerable value, therefore, in a simple, web-based interface that allows existing transcriptome datasets to be explored in a visual manner.

The electronic fluorescent pictograph software (eFP) was developed at the University of Toronto in order to facilitate visual exploration of gene expression (Winter et al., 2007). The function of the software is to display a cartoon image depicting various tissues, each colored in with a hue indicating the level of expression of a queried gene. Initially, three instances of this program were made available: a macroscopic overview of *Arabidopsis* showing its organs at various stages in their development, displaying organ-level microarray data; a drawing of a single cell of *Arabidopsis* used to show subcellular localization of a queried protein; and a drawing of various organ systems in mouse (*M. musculus*), with organ-level microarray data (http://bar.utoronto.ca). Since then, numerous other species, including poplar, medicago, soybean, potato, tomato, *Eutrema*, maize, rice, barley, triticale, *Physcomitrella*, and human, encompassing a large number of

tissue sets and experiments, have had instances of eFP created for them (http://bar.utoronto.ca), attesting to the usefulness of the tool.

The eFP software functions as a web app and may be set up on any web server. The code, written in Python, runs on the server and uses a MySQL database as its back-end. Expression data from one or more experiments is placed in the database by the site operator as part of the set-up process, so updates to the data can be easily made. Users can query and visualize the data from anywhere as long as he/she has web access. The original eFP tool for *Arabidopsis* has proven to be highly useful and widely cited. Here I describe my effort to create an eFP web interface for displaying F. vesca flower and fruit RNA_seq data.

Results

The Strawberry flower and fruit transcriptomes

Our lab previously published RNA-seq transcriptomes covering the development of the flower and fruit of the diploid strawberry *F. vesca* (Kang et al., 2013; Hollender et al., 2014). Our fruit transcriptome focuses on the early-stage fruit development, from stage 1 to stage 5. Stage 1 is the pre-fertilization stage and stages 2-5 are post-fertilization stages (Hollender et al., 2012). The accessory fruit, the stem tip which is called the receptacle, can also be divided into the inner pith tissue and the outer cortex tissue (Figure IV-1C) and the botanical fruit, the achene, consists of three tissues: the protective ovary wall, the embryo, and the ghost (the seed coat and endosperm) (Figure IV-1C). RNAs were isolated from each of these fruit tissues at the five stages and are detailed in Appendix B.

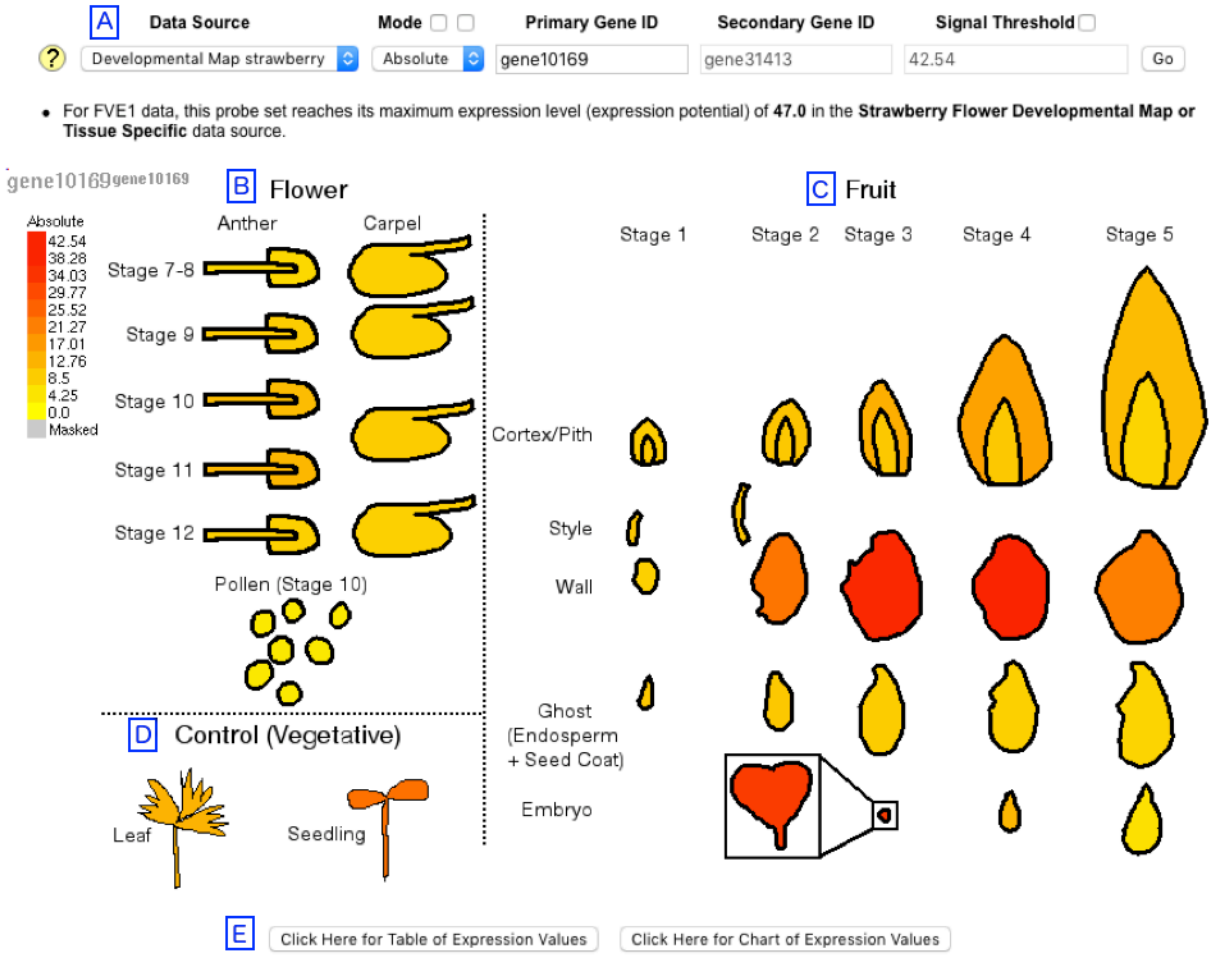

Fig IV-1: eFP User Interface.

Screenshot of Strawberry eFP showing the user interface elements. Gene10169 is shown as an example. Fields along the top allow for different developmental maps to be selected and for a gene ID to be queried **(A)**. The maximum signal value, used to scale the colors of the image, can be adjusted. The image displayed is divided into sections showing the floral stages **(B)**, fruit stages **(C)**, and vegetative control tissue **(D)**. Buttons along the bottom allow for the data to be presented as a table or bar chart **(E)**.

Hollender et al., 2012 has divided the entire process of flower development into 12 stages from floral primordium initiation to the final formation of a complete and open

flower. Because of the small size of the young flowers, Laser capture microdissection (LCM) has been employed to dissect young (stages 1-6) floral tissues. While later floral stages (stages 7-12) were dissected manually. The isolation, processing, and amplification differences between LCM samples and hand dissection samples make it difficult to directly compare gene expression levels between LCM and hand dissection samples (Hollender et al., 2014). Therefore I have focused only on the RNA-seq data from hand-dissected tissues of flowers for eFP-based display (Figure IV-1C). Appendix B lists in detail specific floral samples displayed by eFP: carpels at stages 7-12, stamen at stages 7-12, and pollen taken from open flowers.

Adapting eFP for strawberry transcriptome

While the eFP software was primarily designed to display microarray data, it is readily adaptable for RNA-seq data such as those from strawberry (Kang et al., 2013; Hollender et al., 2014). First, the RNA-seq data were mapped to *F. vesca* whole genome assembly and annotation version 2.0.a1 (Shulaev et al., 2011a; Tennessen et al., 2014) and were kindly provided to me by Chunying Kang at Huazhong Agriculture Univ., China. The mapped reads in FPKM (Fragments Per KB per Million) are in the form of Microsoft Excel files with a column for each sample type and a row for each gene. Second, I drew the flower and fruit stages (Figure IV-1B–D) based on photographs shown in Hollender et al., 2012 and Kang et al., 2014. Third, a Python script, import_transcriptome.py, was written to automatically read an Excel file formatted in this

way and make the appropriate entries in the MySQL database that eFP uses as its back-end.

New samples, whether for new tissue types or different stages, may be added to the database by re-running the script on another Excel file containing the new data. The template image used by the software to generate the diagrams and the developmental map descriptor file must also be updated to show the new samples.

The eFP software is written in Python and will run on an Apache web server that supports CGI. A MySQL database is required, and as eFP doesn't possess the capability to create and configure this database itself, or to input any data into it, these steps must be performed manually during setup. The new import_transcriptome.py script can automate much of the data entry step if the transcriptome data is available in the form of a compatible Excel document. The application, when finalized, will be linked to on the Rosaceae Fruit Project website, http://www.clfs.umd.edu/CBMG/faculty/Liu/lab/Rosaceae/index.shtml.

FveMYB10 as an example in eFP

As described in Chapter III, the fruit color gene identified, FveMYB10, has the gene ID gene31413. The software responds to this query by coloring in its line drawing according to the expression levels (FPKM) for FvMYB10 recorded for each tissue sample (Fig IV-2). The line drawing of each tissue is colored a shade from yellow to red, with yellow indicating zero expression and red indicating maximum expression, a value set by default to the maximum expression of the gene within the dataset but which can be

overridden by the user if desired. Expression of FveMYB10 appears at its highest level in the ghost tissue at stage 5 (Fig. IV-2), reflected in the bright red color of this tissue on the diagram. This dataset does not capture the turning stages wherein the receptacle fruit begins to produce callistephin and turn red, though a new RNA-seq dataset on the turning stage of the receptacle can be incorporated into this site. The diagram however supports the interpretation that FvMYB10 is expressed in a tissue shortly before pigmentation begins to appear, as the achenes turn red first before the receptacle turns red. Because FvMYB10 tightly controls biosynthesis of the red pigment (callistephin), this pattern of expression indicates that red pigment synthesis in achene, the true fruit, and in receptacle the accessary fruit are both controlled by FveMYB10.

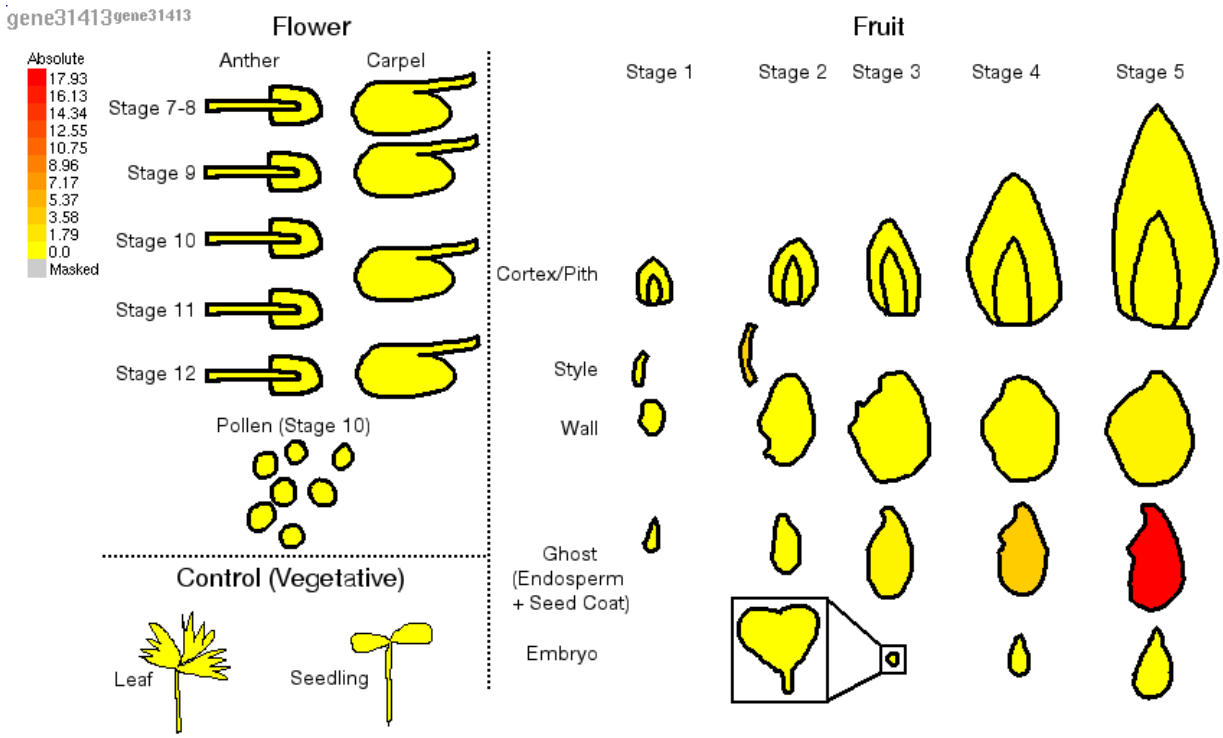

Fig IV-2: Example output for FvMYB10 (gene31413). Expression is seen in late-stage ghost, shortly preceding buildup of anthocyanins.

By default the diagram is scaled from zero to the maximum observed expression for the

queried gene (see scale bar on the left), but the user may enter a different maximum manually. The software may also be used to compare the expression of two genes. Buttons below the diagram may be used to display a table listing exact expression values of the queried gene in each of the samples, or to display a bar chart of those values.

FveYUC10 as an example in eFP and insights into fruit initiation

The ability to analyze gene expression profiles during fruit development will help elucidate developmental processes that lead to the formation of this unusual fruit in

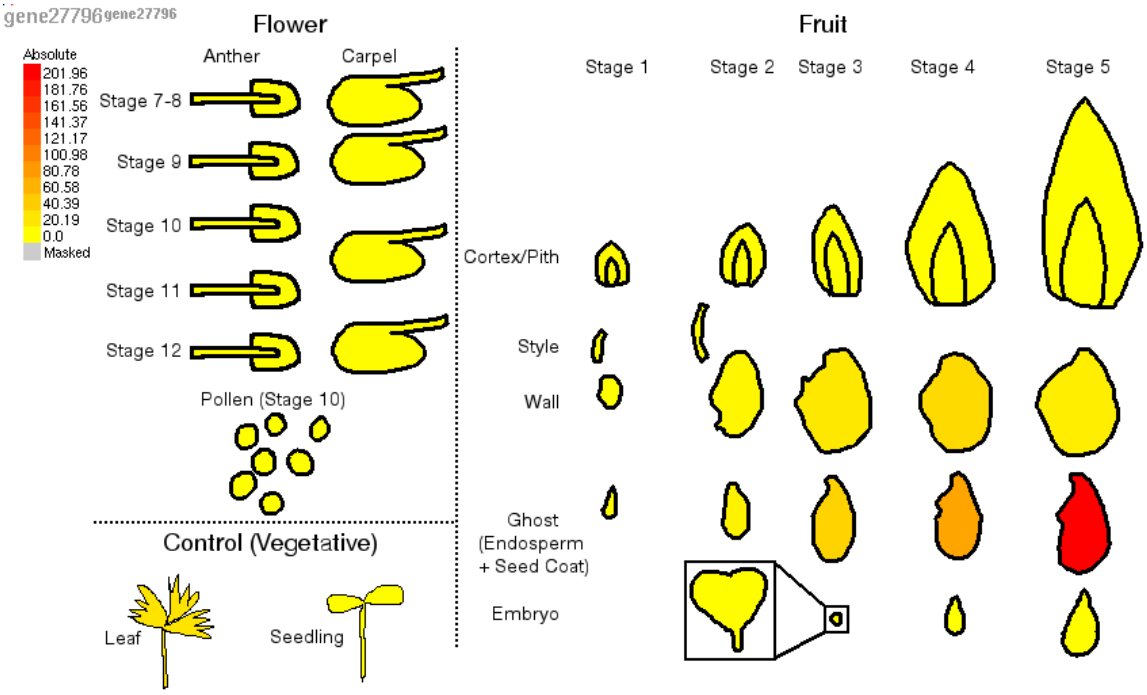

Figure IV-3: Expression of *F. vesca* YUC10.

YUC10 catalyzes the final step of auxin biosynthesis. Its expression in the ghost (endosperm + seed coat) supports the identification of this tissue as the origin of ripening-promoting auxin (Kang et al., 2013)

strawberry and identify candidate genes for functional studies. As discussed in Chapter I, fertilization-induced auxin biosynthesis from the achene is critical to initiation of the receptacle fruit development. Removing achene from the receptacle would block fruit formation, while exogenous spray of the plant hormone auxin can restored receptacle fruit formation even in the absence of achene (Nitch, 1950; Kang et al., 2013). However, achene is a complex organ containing ovary wall and seeds. Inside each seed, there is embryo and endosperm. Exactly where auxin is synthesized upon fertilization?

Here we can examine the expression of auxin biosynthesis genes. YUC10 catalyzes the second step of the auxin biosynthesis. As shown here, YUC10 is found preferentially expressed in the ghost (seed coat + endosperm) post-fertilization (stage stages 3–5) (Figure IV-3). This supports the proposal that auxin biosynthesis post-fertilization occurs mainly in the ghost (Figure IV-4; Kang et al., 2014).

Discussion

Tools for quickly and easily examining biological datasets in a visual manner without the need for complex bioinformatics software and the associated skillset requirements opens up these large, publicly-available datasets to many more researchers and to many more projects. Even where the skillsets to use such technical software are available, doing so to run old analyses on new data may be time-consuming and represent an unnecessary duplication of effort, with the attendant risk of subtle differences in

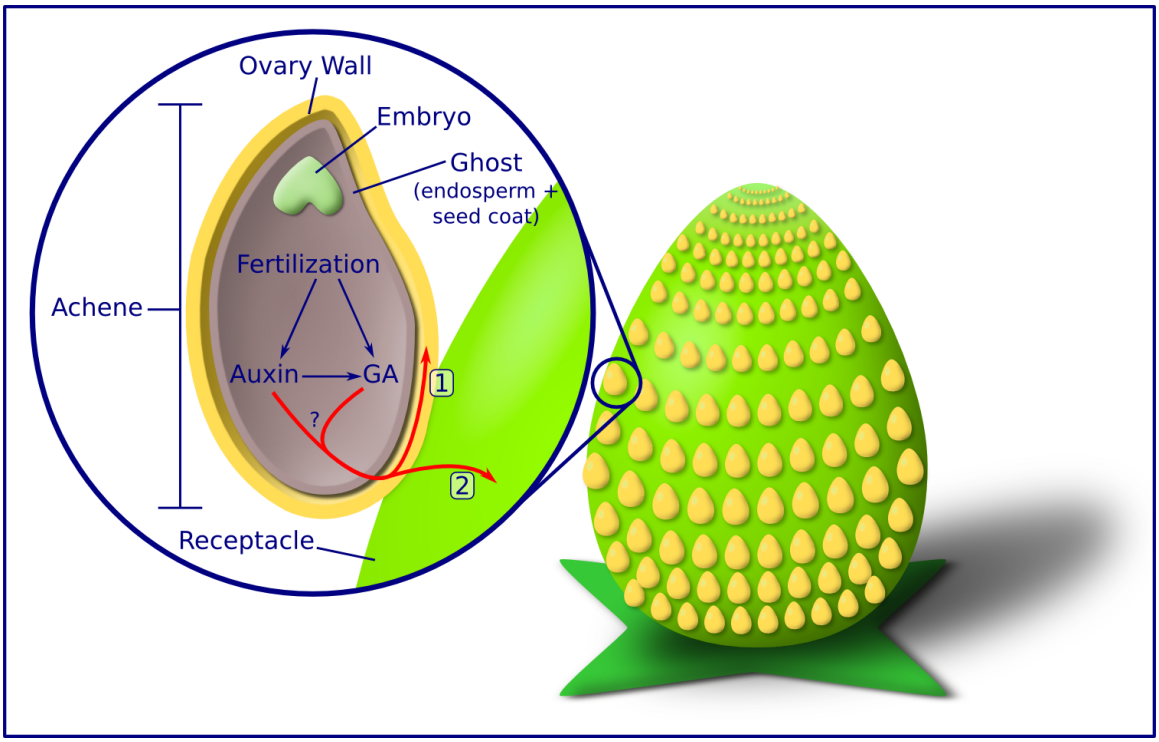

Figure IV-4: Model for the roles of auxin and GA in strawberry fruit ripening.

Under this model, auxin is produced in the ghost (seed coat+ endosperm) where it acts locally to induce GA biosynthesis. The auxin and the GA are both transported to the receptacle where they both activate downstream genes that trigger fruit set and begin the ripening process.

approach that may complicate comparison with existing efforts. Standardized, narrow-purpose software can help alleviate these problems by providing a simple and convenient interface for a specific task. The fact that the same software, eFP, already hosts datasets from species such as *Arabidopsis* and *Medicago* will also facilitate comparisons between strawberry and species with more "conventional" fruit architectures.

The strawberry eFP tool serves to allow easy exploration of the published F. vesca transcriptome datasets, facilitating insights into the genes that control flower and fruit development as well as fruit ripening, and making it easier to compare gene expression among different species. It can be used to compare expression of two genes. Having this tool will also help advance *F. vesca* as a model species. Since *F. vesca* is the model Rosaceae, this work also lays the foundation for the possible extension of this eFP to include other Rosaceae species such as apple and peach, for which abundant RNA-seq data are available (e.g. Zhu et al., 2011; Krost et al., 2012; Wang et al., 2013).

Methods

The diagram was traced by hand from Kang et al. (2013), Fig. 1A using the GIMP image editing software and then exported in TGA format as required by eFP. The eFP software used is version 1.6.0, downloaded from https://sourceforge.net/projects/efpbrowser/, where it is available under the GNU General Public License v 2.0.

Scripts

The new script, import_transcriptome.py, functions to import transcriptome data from Excel spreadsheets to a MySQL database for use with eFP. The Excel spreadsheet must be formatted as a grid of expression values, with each column corresponding to one sample and each row corresponding to one gene. The first row should have the sample names and the first column should have the gene identifiers (these are what will be searched for by the user). The script will read data from only one sheet within the document (the sheet may be specified by name on the command line), but entering data from multiple sheets is possible simply by running the script multiple times, specifying a different sheet each time. The script is written in Python, and accepts the following command-line options (Table IV-1):

Option	**Description**
-h	Show help and usage information
-x file.xlsx	Excel workbook (in xlsx format) containing RPKM values. Columns are samples, rows are genes.
-s sheet	Name of worksheet to use. Default is "RPKM".
-k socket	Socket to connect to the MySQL database.
-u user	Username with which to log into MySQL database. Default is"db_user".
-d database	Which database to push the values into. Default is "strawberry".
-t table	Which table in the database to push the values into. Default is "strawberry_fruit_v2_0".

Table IV-1: Command-line options accepted by import_transcriptome.py. The script is designed to import transcriptomic data from specially-formated Excel documents so a MySQL database for use by eFP.

The script will ask the user to securely type in the MySQL password for the user specified on the command line. The script adds to any existing data in the database; if data needs to be removed first then this must still be done manually.

Thesis Conclusion and Future Directions

The goal of this work was to advance understanding of the molecular mechanisms behind the development and maturation of the fruit, a complex and diverse structure of significant economic and ecological importance. In this work, I have examined the processes that shape the fruit at its earliest stages by presenting a new model for the role of auxin in patterning the structure. I have examined the processes that turn an open flower into a ripe, edible fruit by discovering the locus responsible for naturally-occurring variation in fruit color in strawberry. And I have made our lab's transcriptome data accessible to other researchers by deploying a simple-to-use web-based tool for exploring the fruit and floral transcriptomes, facilitating further discoveries and further exploration of the processes that shape the strawberry fruit.

In the first part of this work, I presented the new model for the role of auxin in the early development of the *Arabidopsis* fruit. The long-standing previous model, based on mutants and chemicals that disrupt auxin transport or response, holds that auxin is produced at the apex and diffuses down the structure, resulting in a diffusion gradient that provides positional information allowing the gynoecium to partition itself into the style, ovary, and gynophore. I showed that this model is not consistent with the information that has emerged since: that the proposed gradient has never been visualized despite efforts to do so, that auxin biosynthesis is not localized to the apex until the partitions are already set, and that patterning must occur at an early stage where the gynoecium has little-to-no apical-basal dimension to support a gradient. I presented an alternative model, "the early action model", based on current models of leaf patterning, where auxin

sets the adaxial-abaxial patterning of the carpels at a very early stage, as it does in leaves. Under this model, disruptions to the normal functioning of the hormone result in disruption of adaxial-abaxial polarity of the carpels, and these are what cause the observed phenotypes, consistent with the phenotypes observed in mutants that disrupt ad/abaxial polarity in both carpels and leaves.

In the second part of this work, I analyzed the re-sequenced genomes of three accessions of the diploid strawberry *Fragaria vesca*: Hawaii 4, Rügen, and Yellow Wonder. I generated genome-wide lists of SNP and small indel variants within and among the three accessions, and predicted their effects on known genes. Using this data, I identified a single candidate SNP likely responsible for the difference in fruit color between Rügen and the other two, a single-base substitution at a highly-conserved site in the DNA binding domain of the transcription factor FvMYB10, already known to regulate the later stages of biosynthesis of the red pigment anthocyanin. Functional analysis confirmed that the Rügen allele is able to induce red coloration in the other varieties. This variant data represents an invaluable resource for identifying the causal polymorphisms behind observed phenotypic variation.

In the third and final part of this work, I adapted the web-based gene expression visualization tool eFP for use with our lab's strawberry fruit and flower transcriptomes. These transcriptomes, covering different tissues in the flower and fruit at different stages of development, were publically available, but exploring the dataset often required the use of complex bioinformatics tools and pipelines. The eFP tool allows for the expression of any gene to be easily visualized without requiring specialized

bioinformatics skills, and facilitates exploration of the genetic circuitry that controls fruit ripening in strawberry.

Expanding the eFP tool for the Rosaceae

There are numerous directions in which to expand this visualization tool. Recently our lab has generated new transcriptome data for the turning stage fruit of *F. vesca*. Addition of this new data to eFP will facilitate exploration of ripening stage fruit development. Fruit and floral transcriptomes from other Rosaceae species such as apple and plum can also be added, and comparison among these related species may yield insights into the genetic mechanisms behind the diverse fruit morphologies found in the Rosaceae family.

The Bio-Analytic Resource for Plant Biology (BAR) at the University of Toronto, the host for most existing eFP tools, also has a tool called the Expression Angler that allows the user to find co-expressed genes or genes that fit a user-defined expression profile. Setting up such a tool for use with the strawberry transcriptomes would facilitate exploration of gene regulatory networks and search for novel genes involved in known processes.

Experimental testing of the early action model

Empirical testing of the early action model is necessary to validate or dispute the model. The early action model predicts that in early stage carpel development, polar auxin transport should direct auxin upward along the adaxial and abaxial surfaces of the carpels and down the interiors where these two domains meet, as is seen in leaves.

Larsson et al. (Larsson et al., 2014) examined polar PIN localization in the *Arabidopsis* gynoecium at floral stage 7 using a GFP fusion protein, and found exactly this pattern; the signal appeared along the acropetal cell surfaces of the cells on the adaxial and abaxial surfaces of the carpels and appeared along the basipetal cell surfaces of the cells in the interior of the carpels, where the adaxial and abaxial domains meet.

An important prediction of the early action model is that altered auxin synthesis, transport, or signaling can abolish or alter the establishment of the abaxial/adaxial domain boundary. Therefore, examining AB and AD domain-specific markers, such as the expression of ETT –an abaxial marker and the expression of PHABULOSA (PHB)-an adaxial marker, in various auxin mutants, yuc1 yuc4 double mutants, pin, and pid, will indicate if AB and AD domains are indeed altered in these mutant background. To achieve this, the AB or AD marker genes can be used in *in situ* hybridization experiments to identify and observe the AB and AD domains in the auxin mutant tissues. Alternatively, the AB and AD promoters can be fused to reporter genes and the reporter gene expression can be monitored during the early stage carpel primordial initiation. Any changes of AB and AD domain in the auxin mutants would support the early action model.

In an attempt to conduct this line of experimental testing, *in situ* probes were prepared to examine the expression of *PHABULOSA (PHB)* and *YABBY (YAB)* genes, which are members of the AD and AB networks respectively. *In situ* hybridization was performed with both probes on wildtype Landsberg erecta, and with *PHB* on *lug-3*. The

reason for examining *lug-3* is that this mutant also shows a similar phenotype in gynoecium development as auxin signaling or transport mutants, and our lab already has *lug-3* tissues embedded for *in situ*. The *in situ* results are shown in Fig V-1 and V-2. The sections did not clearly capture earliest stages of carpel emergence. *PHB* can be seen localized to the adaxial domain of a young flower (Fig. V-1A white arrow). In the *lug-3* mutant the adaxial domain expression of *PHB* is maintained generally (Fig. V-1D white arrow) but is less consistent, with some areas in the adaxial domain missing expression (Fig V-1C, black arrow).

Because of the difficulty in capturing carpel development at the earliest stage as well as the insensitivity of *in situ* hybridization (high background to low signal), we did not pursue this experiment further. It will be much more effective if GFP or other fluorescent reporter constructs can be made to report YABBY and PHB expression. The reporter constructs can be crossed into *yuc, pin, pid,* or *ett* mutant backgrounds and the GFP expression can be monitored to reveal whether or not AB and AD domains are affected.

Conclusion

The work presented here advances understanding of the developmental process of the fruit at all stages, from flower bud to ripe fruit. It employs multiple molecular genetic, genomic, and bioinformatics methods to answer important questions about this complex structure. It advances strawberry as a model organism for the study of rosaceous crops and provides new resources and new tools that lay the foundation for future discoveries.

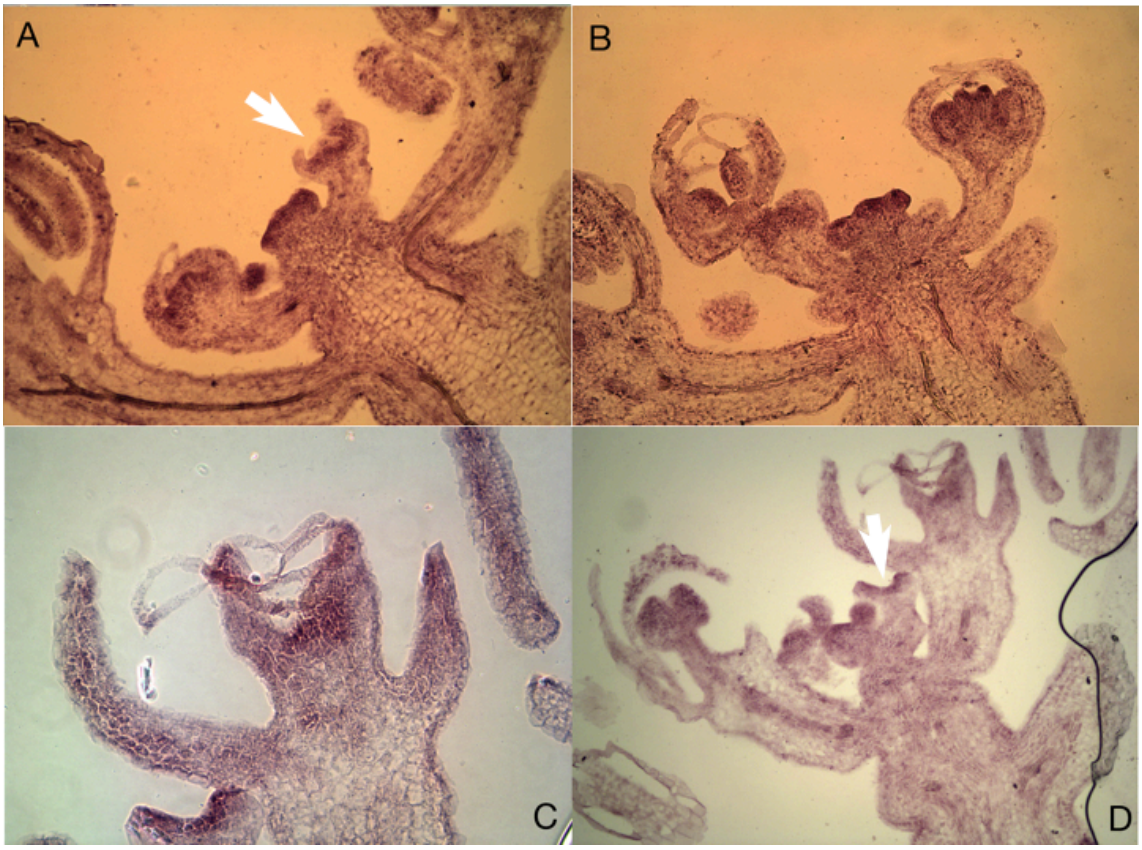

Figure V-1: *PHB in situ* in wild type (Ler) and *lug-1*

(A, B) *PHB in* L*er*. Expression is seen in the inflorescence meristem and the adaxial domains of the carpels. (C, D) *PHB* in *lug-1*. Expression is seen in the adaxial domain of the carpels but inconsistently.

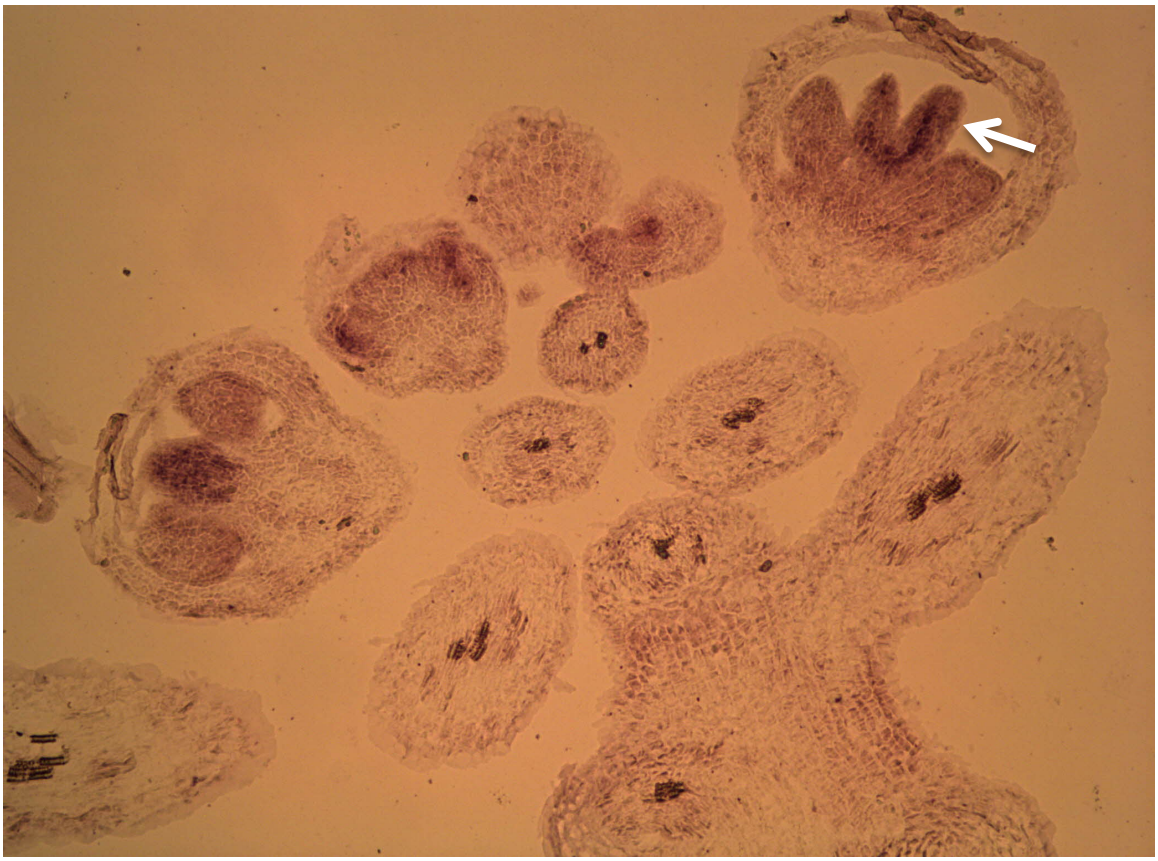

Figure V-1: *YAB in situ* in wild type (Ler)

Expression Is seen in the valve of immature gynoecia (white arrow)

Appendix A: Supplemental Information for Chapter III

Supplemental Figures

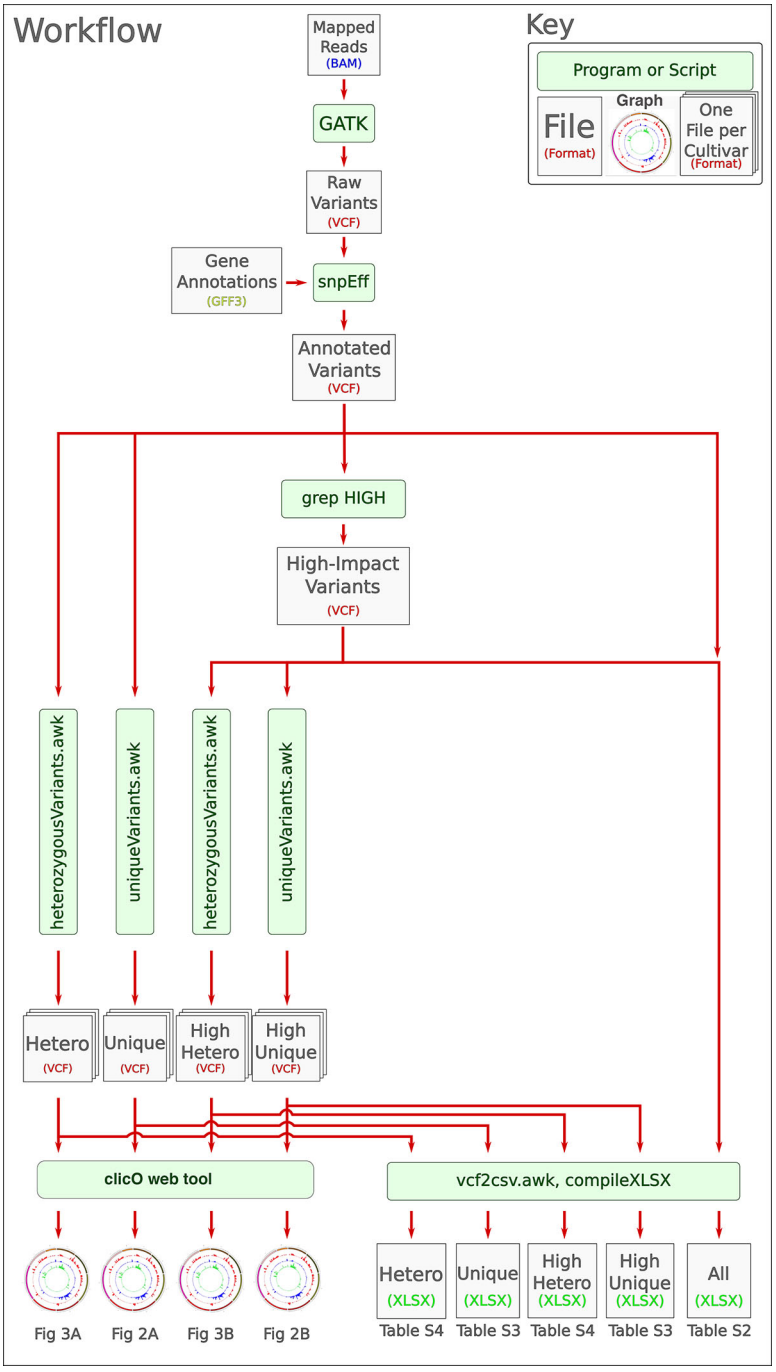

Figure S1. Sequence analysis workflow.

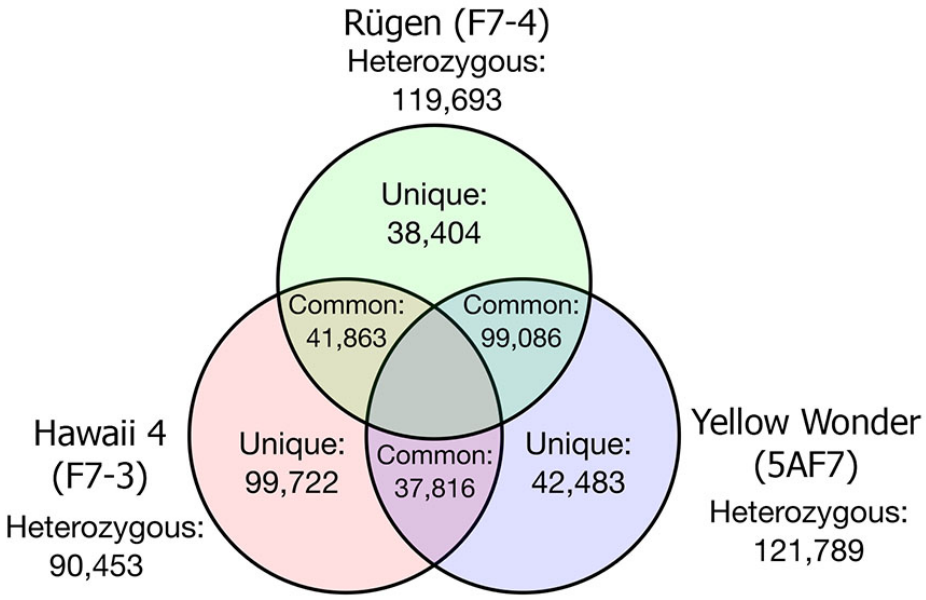

Figure S2. Venn Diagram showing unique and overlapping SNPs found among the three accessions. Number of heterozygous loci in each accession is also indicated.

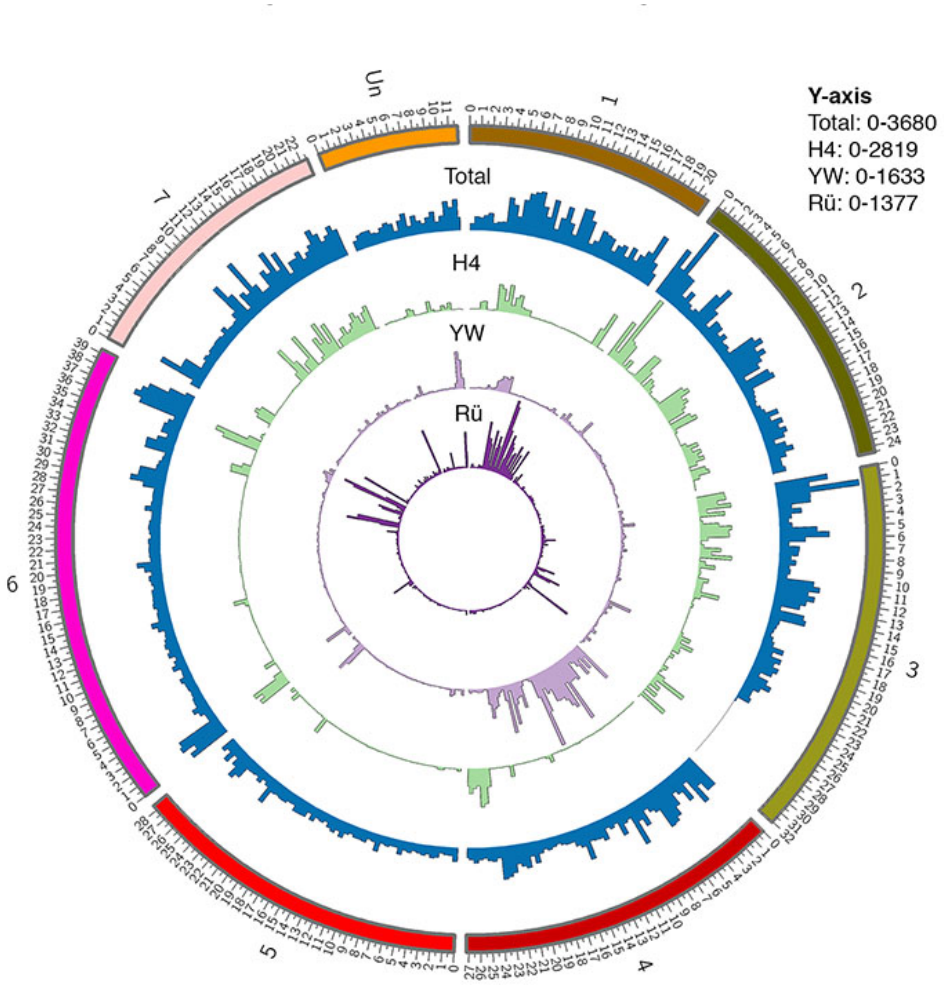

Figure S3. Circos ideogram showing genome-wide distribution of total combined variants in comparison to Accession-Unique variants.

The outermost circle represents the seven *F. vesca* Linkage Groups plus the unanchored scaffolds (Un), drawn to scale in Mbp. The blue track is the histogram showing the distribution of total combined variants (both unique and shared variants) from all three accessions. The inner three tracks are histograms showing distribution of

accession-unique variants, which are identical to Figure 2A. Bin size is 500 kb, Y-axis is the number of total variants or variety-unique variants.

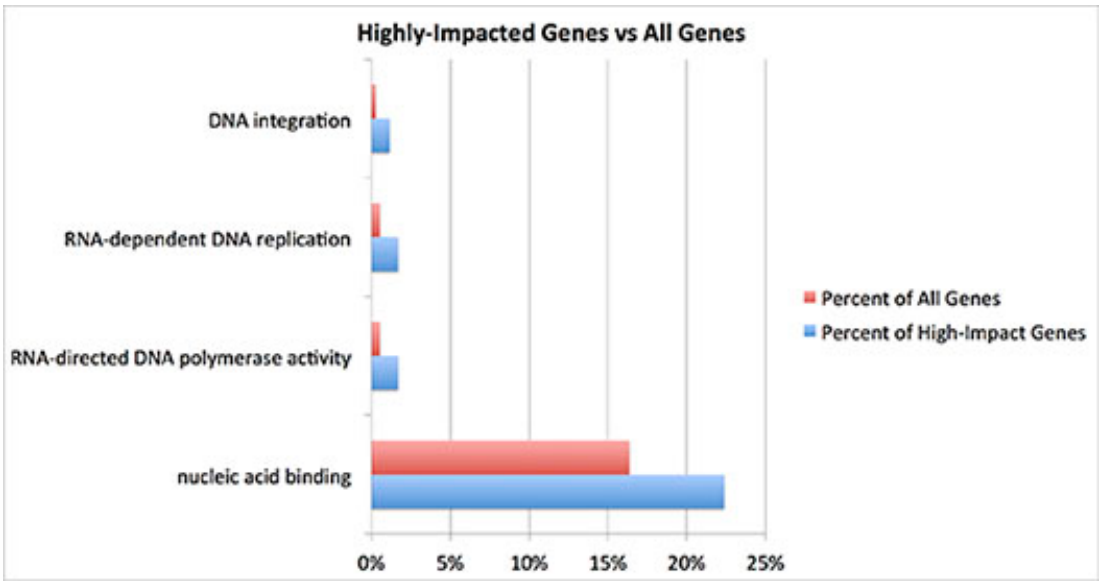

Figure S4. Enriched GO terms among genes (test set) affected by high impact variants in all three accession. The reference set was derived from Phytozome, representing % genes belonging to each GO category.

Supplemental methods

The following is a listing of the various scripts written to perform the data analysis presented in this paper. These scripts were run on Mac OS X for this paper and all should work on most Linux systems without modification provided the necessary tools are installed (Awk, Perl, Zsh, and R). Running them on Windows systems may require modification or the installation of a Unix environment.

VCF Filtering Scripts

The scripts in this category will take .vcf files and extract out only the lines matching the script's criteria. Most of them operate under the assumption that there are exactly three cultivars being analyzed and may not work for a different number of samples. They all use awk. Arguments are generally given to awk scripts using -v variable=value. These are all able to work with .eff.vcf files from snpEff. Most of these scripts take one or two sample numbers (from 0 to 2) to indicate which cultivar(s) you want them to act on. In our .vcf files the cultivars are numbered as follows:

0 = Hawaii 4

1 = Rügen

2 = Yellow Wonder

uniqueVariants.awk

This script extracts .vcf lines denoting loci where the given cultivar has no allele that is also found in either of the other two varieties. The variety to analyze may be

specified by setting the "sample" variable.

```
./uniqueVariants.awk -v sample=0 infile.vcf >
```

The above will take the lines from infile.vcf where there is an allele unique to Hawaii 4 (sample 0) and save them in outfile.vcf.

heterozygousVariants.awk

This script extracts .vcf lines denoting loci where the given variety is heterozygous (has two or more alleles). The variety to analyze may be specified by setting the "first" variable.

```
./heterozygousVariants.awk -v first=2
infile.vcf > outfile.vcf
```

The above will take the lines from infile.vcf where Yellow Wonder (sample 2) is heterozygous and save them in outfile.vcf.

homozygousVariants.awk

This script extracts .vcf lines denoting loci where the given variety is homozygous

(has only one allele). The variety to analyze may be specified by setting the "first" variable.

```
    ./homozygousVariants.awk -v first=0 infile.vcf
> outfile.vcf
```

The above will take the lines from infile.vcf where Hawaii 4 (sample 0) is homozygous and save them in outfile.vcf.

differingVariants.awk

This script extracts .vcf lines denoting loci where the two given varieties differ (do not share any alleles). The varieties to analyze may be specified by setting the "first" and “second” variables.

```
    ./differingVariants.awk -v first=2 -v second=1
infile.vcf > outfile.vcf
```

The above will take the lines from infile.vcf where Yellow Wonder (sample 2) and Rügen (sample 1) do not share any alleles and save them in outfile.vcf.

commonVariants.awk

This script extracts .vcf lines denoting loci where the two given cultivars are the same, meaning that they are both homozogous for the same allele and do not share this allele with the remaining cultivar. The varieties to analyze may be specified by setting the "first" and “second” variables.

```
./commonVariants.awk -v first=0 -v second=1
infile.vcf > outfile.vcf
```

The above will take the lines from infile.vcf where Hawaii 4 (sample 2) and Rügen (sample 1) are homozygous for the same allele and this allele is not found in Yellow Wonder. It will save these lines in outfile.vcf.

VCF to CSV

The following scripts both extract information from a .vcf file and save it to a .csv file.

vcf2csv.awk

This extracts the chromosome, location, and variant list for each cultivar for each variant and saves them to a .csv file that can be opened in a spreadsheet program. Each

cultivar has a column in the .csv file that lists the sequences of the alleles it has, separated with a vertical bar ("|") if more than one is present.

```
./vcf2csv infile.vcf > outfile.csv
```

This script was used in the generation of the supplemental Excel files in the paper.

extractLocs

This script is similar to the above but will only extract the chromosome and location of each .vcf line. Its output is intended for use in generating the histograms using the R scripts below. It requires Zsh to be installed (which it is by default in Mac OS but may not be in some Linux systems).

```
./extractLocs infile.vcf > outfile.csv
```

Fixing a GFF file for use with snpEff

fix-gff.awk

This script will fix a Phytozome .gff3 file so that snpEff can read it. The

Fvesca_226_genes.gff3 file taken from Phytozome had the frame offset field for each exon calculated in a way different from what snpEff expects, and so snpEff did not interpret it correctly. This script was written to process the .gff3 file so as to correct this incompatibility. It is not guaranteed to work with other .gff files. The input and output should not be the same file.

```
./fix-gff.awk infile.gff > outfile.gff
```

Removing variants in high-read regions

The scripts in this category are used to remove variants that are in regions where the reads are too high.

extractHCRanges.awk

This script is the first stage of the process of cutting out variants that are in regions where the read count is too high (any such region likely represents a sequencing anomaly and all variants found there should therefore be considered unreliable). The task of this first script is to mark the regions of the genome within which variants should be excluded. It takes as input a coverage file produced by the genomeCoverageBed utility found in BedTools (http://bedtools.readthedocs.org). This utility should be run on the final .bam file, the one that is also fed into the variant calling software. Run it with the -

bg and -ibam options, as in this example for Hawaii 4:

```
   genomeCoverageBed -bg -ibam Hawaii4Recal.bam >
Hawaii4.bg
```

This will produce the file Hawaii4.bg which can be fed into extractHCRanges.awk. A read cutoff may be specified to extractHCRanges.awk with -v cutoff=50 (50 was the cutoff used for the analysis described in this paper). To continue the same Hawaii 4 example:

```
   extractHCRanges.awk -v cutoff=50 Hawaii4.bg >
Hawaii4.HC50.csv
```

This will produce a .csv file to be fed into cutReads (described below) to specify which regions the latter script should cut. Each row in this file represents a region wherein the coverage exceeds the cutoff, and has three fields: the chromosome, the start of the region in bp, and the end of the region in bp. The latter two are inclusive (i.e. the region LG1,1,10 would include bases 1 and 10 on the chromosome LG1)

cutReads

This script will take a file that lists variants and will remove those entries that fall within any region marked in the csv file produced by extractHCRanges.awk. It can cut both VCF files and variant-containing CSV files of the kinds produced by vcf2csv.awk and extractLocs (described above). It requires the Perl libraries Getopt::Long and List::Util 1.33, both available on CPAN. The command-line options are as follows:

- --zones | -z: Specify the zones file, the CSV file produced by extractHCRanges.awk
- --varsfile | -v: Specify the file containing the variants to be cut using the zones file
- --filetype | -f: Specify as "csv" if the variants file is a CSV file or "vcf" if it is a VCF file. Default is "csv"

Continuing our Hawaii 4 example, we can cut a VCF file with:

```
cutReads -v H4-Het.vcf -z Hawaii4.HC50.csv -f
vcf > Hawaii4.Het.HC50.vcf
```

Or we could cut a CSV with:

```
cutReads -v H4-Het.csv -z Hawaii4.HC50.csv >
Hawaii4.Het.HC50.csv
```

Converting CSV to XLSX

compileXlsx

This Perl script converts and compiles a set of .csv files into a set of multi-sheet Excel (.xlsx) documents. It requires the Perl libraries Excel::Writer::XLSX, Text::CSV, and Switch. A configuration file, itself in CSV format, is used to specify the names of the Excel workbooks, the name of each sheet in each workbook, and the name of the .csv file whose contents should be placed in each sheet. The configuration file should consist of single-column rows with excel file names, each followed by one or more two-column rows wherein the first column is the name to be given to a sheet and the second column is the name of the .csv file whose contents should be placed in that sheet. The first row of each .csv file is presumed to be column labels and is therefore styled bold in the Excel file. The name of the configuration file should be the sole argument given to compileXlsx.

```
compileXlsx config.csv
```

Appendix B: Sample Descriptions for Chapter IV

Sample	Description
cortex1-1	cortex from just open flower, replicate 1
cortex1-2	cortex from just open flower, replicate 2
cortex2-1	cortex from the flowers which have been pollinated for about 3 days, replicate 1
cortex2-2	cortex from the flowers which have been pollinated for about 3 days, replicate 2
cortex3-1	cortex at about 6 DPA, same age as embryo-3, replicate 1
cortex3-2	cortex at about 6 DPA, same age as embryo-3, replicate 2
cortex4-1	cortex at about 9 DPA, same age as embryo-4, replicate 1
cortex4-2	cortex at about 9 DPA, same age as embryo-4, replicate 2
cortex5-1	cortex at about 12 DPA, same age as embryo-5, replicate 1
cortex5-2	cortex at about 12 DPA, same age as embryo-5, replicate 2
embryo3-1	heart stage embryos,at about 6 DPA, replicate 1
embryo3-2	heart stage embryos,at about 6 DPA, replicate 2
embryo4-1	immature cotyledon stage embryos,at about 9 DPA, replicate 1
Embryo4-2	immature cotyledon stage embryos,at about 9 DPA, replicate 2
embryo5-1	mature embryos which fill up entire ovules,at about 12 DPA, replicate 1
embryo5-2	mature embryos which fill up entire ovules,at about 12 DPA, replicate 2
ghost3-1	seeds without embryos inside at about 6 DPA, same age as embryo-3, replicate 1
ghost3-2	seeds without embryos inside at about 6 DPA, same age as embryo-3, replicate 2
ghost4-1	seeds without embryos inside at about 9 DPA, same age as embryo-4, replicate 1
Ghost4-2	seeds without embryos inside at about 9 DPA, same age as embryo-4, replicate 2
ghost5-1	seeds without embryos inside at about 12 DPA, same age as embryo-5, replicate 1
Ghost5-2	seeds without embryos inside at about 12 DPA, same age as embryo-5, replicate 2
leaf-1	young trifoliate leaves, replicate 1
leaf-2	young trifoliate leaves, replicate 2
Ovule1-1	ovules from just open flower, replicate 1
Ovule-1-2	ovules from just open flower, replicate 2
Ovule2-1	seeds from the flowers which have been pollinated for about 3 days, replicate 1
Ovule2-2	seeds from the flowers which have been pollinated for about 3 days,

	replicate 2
pitch1-1	pith from just open flower, replicate 1
pitch1-2	pith from just open flower, replicate 2
pitch2-1	pith from the flowers which have been pollinated for about 3 days, replicate 1
pitch2-2	pith from the flowers which have been pollinated for about 3 days, replicate 2
pith3-1	pith at about 6 DPA, same age as embryo-3, replicate 1
pith3-2	pith at about 6 DPA, same age as embryo-3, replicate 2
pith4-1	pith at about 9 DPA, same age as embryo-4, replicate 1
pith4-2	pith at about 9 DPA, same age as embryo-4, replicate 2
pith5-1	pith at about 12 DPA, same age as embryo-5, replicate 1
pitch5-2	pith at about 12 DPA, same age as embryo-5, replicate 2
seedling-1	whole seedlings at 10 days post germination growing in MS medium, replicate 1
seedling-2	whole seedlings at 10 days post germination growing in MS medium, replicate 2
Style-1-1	styles and stigmas from just open flowers, replicate 1
Style-1-2	styles and stigmas from just open flowers, replicate 2
style2-1	style from the flowers which have been pollinated for about 3 days, replicate 1
style2-2	style from the flowers which have been pollinated for about 3 days, replicate 2
wall1-1	carpel walls or ovary walls from just open flower, replicate 1
wall1-2	carpel walls or ovary walls from just open flower, replicate 2
wall2-1	carpel walls or ovary walls from the flowers which have been pollinated for about 3 days, replicate 1
wall2-2	carpel walls or ovary walls from the flowers which have been pollinated for about 3 days, replicate 2
wall3-1	carpel walls or ovary walls at about 6 DPA, same age as embryo-3, replicate 1
wall3-2	carpel walls or ovary walls at about 6 DPA, same age as embryo-3, replicate 2
wall4-1	carpel walls or ovary walls at about 9 DPA, same age as embryo-4, replicate 1
Wall4-2	carpel walls or ovary walls at about 9 DPA, same age as embryo-4, replicate 2
Wall5-1	carpel walls or ovary walls at about 12 DPA, same age as embryo-5, replicate 1
Wall5-2	carpel walls or ovary walls at about 12 DPA, same age as embryo-5,

	replicate 2
Anther_7-8_A	Anthers from stage 7 or 8 flowers, replicate 1, Arcturis kit
Anther_7-8_B	Anthers from stage 7 or 8 flowers, replicate 2, Arcturis kit
Anther_9_A	Anthers from stage 9 flowers, replicate 1
Anther_9_B	Anthers from stage 9 flowers, replicate 2
Anther_10_A	Anthers from stage 10 flowers, replicate 1
Anther_10_B	Anthers from stage 10 flowers, replicate 2
Anther_11_A	Anthers from stage 11 flowers, replicate 1
Anther_11_B	Anthers from stage 11 flowers, replicate 2
Anther_12_A	Anthers from stage 12 flowers, replicate 1
Anther_12_B	Anthers from stage 12 flowers, replicate 2
Carpel_7-8_A	Carpels from stage 7 or 8 flowers, replicate 1, Arcturis kit
Carpel_7-8_B	Carpels from stage 7 or 8 flowers, replicate 2, Arcturis kit
Carpel_9_A	Carpels from stage 9 flowers, replicate 1, Arcturis kit
Carpel_9_B	Carpels from stage 9 flowers, replicate 2, Arcturis kit
Carpel_10_A	Carpels from stage 10 flowers, replicate 1
Carpel_10_B	Carpels from stage 10 flowers, replicate 2
Carpel_12_A	Carpels from stage 12 flowers, replicate 1, Arcturis kit
Carpel_12_B	Carpels from stage 12 flowers, replicate 2, Arcturis kit
Pollen_A	Pollen collected from open flowers, replicate 1
Pollen_B	Pollen collected from open flowers, replicate 2

References

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE (2014) A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell **26**: 962–80

Arber A (1937) The Interpretation of the Flower: A Study of Some Aspects of Morphological Thought. Biol Rev **12**: 157–184

Archbold DD, Dennis FG (1984) Quantification of free ABA and free and conjugated IAA in strawberry achene and receptacle tissue during fruit development. J Am Soc Hortic Sci **109**: 330–335

Archbold DD, Dennis FG (1985) Strawberry receptacle growth and endogenous IAA content as affected by growth regulator application and achene removal. J Am Soc Hortic Sci **110**: 816–820

Balanza V, Navarrete M, Trigueros M, Ferrandiz C (2006) Patterning the female side of Arabidopsis: the importance of hormones. J Exp Bot **57**: 3457–3469

Bartel B (1997) Auxin Biosynthesis. Annu Rev Plant Physiol Plant Mol Biol **48**: 51–66

Barton MK (2010) Twenty years on: The inner workings of the shoot apical meristem, a developmental dynamo. Dev Biol **341**: 95–113

Bassolino L, Zhang Y, Schoonbeek HJ, Kiferle C, Perata P, Martin C (2013) Accumulation of anthocyanins in tomato skin extends shelf life. New Phytol **200**: 650–5

Bayer EM, Smith RS, Mandel T, Nakayama N, Sauer M, Prusinkiewicz P, Kuhlemeier C (2009) Integration of transport-based models for phyllotaxis and midvein formation. Genes Dev **23**: 373–384

Benjamins R, Quint A, Weijers D, Hooykaas P, Offringa R (2001) The PINOID protein kinase regulates organ development in Arabidopsis by enhancing polar auxin transport. Development **128**: 4057–4067

Benková E, Michniewicz M, Sauer M, Teichmann T, Seifertová D, Jürgens G, Friml J (2003) Local, Efflux-Dependent Auxin Gradients as a Common Module for Plant Organ Formation. Cell **115**: 591–602

Bennett SRM, Alvarez J, Bossinger G, Smyth DR (1995) Morphogenesis in pinoid mutants of Arabidopsis thaliana. Plant J **8**: 505–520

Berleth T, Scarpella E, Prusinkiewicz P (2007) Towards the systems biology of auxin-transport-mediated patterning. Trends Plant Sci **12**: 151–159

Bood K g., Zabetakis I (2002) The Biosynthesis of Strawberry Flavor (II): Biosynthetic and Molecular Biology Studies. J Food Sci **67**: 2–8

Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in Arabidopsis. Plant Cell **1**: 37–52

Brown T. WP (1965) Genetical control of everbearing habit and 3 other characters in varieties of Fragaria vesca. Euphytica **14**: 97–112

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell **24**: 1242–55

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, et al (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol **26**: 1301–8

Byrne ME (2012) Making leaves. Curr Opin Plant Biol **15**: 24–30

Calderón Villalobos LIA, Lee S, De Oliveira C, Ivetac A, Brandt W, Armitage L, Sheard LB, Tan X, Parry G, Mao H, et al (2012) A combinatorial TIR1/AFB–Aux/IAA co-receptor system for differential sensing of auxin. Nat Chem Biol **8**: 477–485

Cermak T, Baltes NJ, Cegan R, Zhang Y, Voytas DF (2015) High-frequency, precise modification of the tomato genome. Genome Biol **16**: 232

Cheng Y, Dai X, Zhao Y (2006) Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. Genes Dev **20**: 1790–1799

Cheong W-H, Tan Y-C, Yap S-J, Ng K-P (2015) ClicO FS: an interactive web-based service of Circos. Bioinforma Oxf Engl **31**: 3685–3687

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) **6**: 80–92

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature **353**: 31–37

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinforma Oxf Engl **21**: 3674–3676

Craddock C, Lavagi I, Yang Z (2012) New insights into Rho signaling from plant ROP/Rac GTPases. Trends Cell Biol **22**: 492–501

Crozier A, Yokota T, Jaganath IB, Marks SC, Saltmarsh M, Clifford MN (2008) Secondary metabolites in fruits, vegetables, beverages and other plant based dietary components. Plant Second. Metab. Occur. Struct. Role Hum. Diet

Curtis MD, Grossniklaus U (2003) A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. Plant Physiol **133**: 462–469

Dahlke RI, Luethen H, Steffens B (2010) Abp1. Plant Signal Behav **5**: 1–3

Darrow GM, Wallace HA (1966) The Strawberry: History, Breeding, and Physiology, 1st edition. Holt, Rinehart and Winston

Darwish O, Shahan R, Liu Z, Slovin JP, Alkharouf NW (2015) Re-annotation of the woodland strawberry (Fragaria vesca) genome. BMC Genomics **16**: 29

Darwish O, Slovin JP, Kang C, Hollender CA, Geretz A, Houston S, Liu Z, Alkharouf NW (2013) SGR: an online genomic resource for the woodland strawberry. BMC Plant Biol **13**: 223

Davis TM, Yu H (1997) A Linkage Map of the Diploid Strawberry, Fragaria vesca. J Hered **88**: 215–221

Davis T YH (1997) A linkage map of the diploid strawberry, Fragaria vesca. J Hered **88**: 215–221

Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, Hobbie L, Ehrismann JS, Jürgens G, Estelle M (2005) Plant Development Is Regulated by a Family of Auxin Receptor F Box Proteins. Dev Cell **9**: 109–119

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

Doyle JA (2012) Molecular and Fossil Evidence on the Origin of Angiosperms. Annu Rev Earth Planet Sci **40**: 301–326

Duchesne AN (1766) Histoire naturelle des fraisiers. Didot le jeune

Effendi Y, Scherer GFE (2011) AUXIN BINDING-PROTEIN1 (ABP1), a receptor to regulate auxin transport and early auxin genes in an interlocking system with PIN proteins and the receptor TIR1. Plant Signal Behav **6**: 1101–1103

Eklund DM, Staldal V, Valsecchi I, Cierlik I, Eriksson C, Hiratsu K, Ohme-Takagi M, Sundstrom JF, Thelander M, Ezcurra I, et al (2010) The Arabidopsis thaliana STYLISH1 Protein Acts as a Transcriptional Activator Regulating Auxin Biosynthesis. Plant Cell **22**: 349–363

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL (2003) Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Curr Biol **13**: 1768–1774

Endress PK, Doyle JA (2009) Reconstructing the ancestral angiosperm flower and its initial specializations. Am J Bot **96**: 22–66

Endress PK, Igersheim A (2000) Gynoecium Structure and Evolution in Basal Angiosperms. Int J Plant Sci **161**: S211–S223

Eriksson T, Hibbs MS, Yoder AD, Delwiche CF, Donoghue MJ (2003) The Phylogeny of Rosoideae (Rosaceae) Based on Sequences of the Internal Transcribed Spacers (ITS) of Nuclear Ribosomal DNA and the trnL/F Region of Chloroplast DNA. Int J Plant Sci **164**: 197–211

Eshed Y (2004) Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. Development **131**: 2997–3006

Eshed Y, Baum SF, Perea JV, Bowman JL (2001) Establishment of polarity in lateral organs of plants. Curr Biol **11**: 1251–1260

Eshed Y, Izhaki A, Baum SF, Floyd SK, Bowman JL (2004) Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. Development **131**: 2997–3006

Finet C, Fourquin C, Vinauger M, Berne-Dedieu A, Chambrier P, Paindavoine S, Scutt CP (2010) Parallel structural evolution of auxin response factors in the angiosperms: ARF evolution in angiosperms. Plant J **63**: 952–959

Franks RG, Liu Z, Fischer RL (2006) SEUSS and LEUNIG regulate cell proliferation, vascular development and organ polarity in Arabidopsis petals. Planta **224**: 801–811

Friml J, Yang X, Michniewicz M, Weijers D, Quint A, Tietz O, Benjamins R, Ouwerkerk PBF, Ljung K, Sandberg G, et al (2004) A PINOID-Dependent Binary Switch in Apical-Basal PIN Polar Targeting Directs Auxin Efflux. Science **306**: 862–865

Galbiati F, Sinha Roy D, Simonini S, Cucinotta M, Ceccato L, Cuesta C, Simaskova M, Benkova E, Kamiuchi Y, Aida M, et al (2013) An integrative model of the control of ovule primordia formation. Plant J **76**: 446–455

Gälweiler L (1998) Regulation of Polar Auxin Transport by AtPIN1 in Arabidopsis Vascular Tissue. Science **282**: 2226–2230

Gao Y, Zhang Y, Zhang D, Dai X, Estelle M, Zhao Y (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. Proc Natl Acad Sci **112**: 2275–2280

Garces HMP, Champagne CEM, Townsley BT, Park S, Malho R, Pedroso MC, Harada JJ, Sinha NR (2007) Evolution of asexual reproduction in leaves of the genus Kalanchoe. Proc Natl Acad Sci **104**: 15578–15583

Geisler M, Murphy AS (2006) The ABC of auxin transport: The role of p-glycoproteins in plant development. FEBS Lett **580**: 1094–1102

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

Girin T, Paicu T, Stephenson P, Fuentes S, Korner E, O'Brien M, Sorefan K, Wood TA, Balanza V, Ferrandiz C, et al (2011) INDEHISCENT and SPATULA Interact to Specify Carpel and Valve Margin Tissue and Thus Promote Seed Dispersal in Arabidopsis. Plant Cell **23**: 3641–3653

Given NK, Venis MA, Gierson D (1988) Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. Planta **174**: 402–406

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40**: D1178–D1186

Goto K, Kyozuka J, Bowman JL (2001) Turning floral organs into leaves, leaves into floral organs. Curr Opin Genet Dev **11**: 449–456

Grieneisen VA, Marée AFM, Østergaard L (2013) Juicy Stories on Female Reproductive Tissue Development: Coordinating the Hormone Flows: Fruit Patterning. J Integr Plant Biol **55**: 847–863

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, et al (2011) The Composition and Origins of Genomic Variation among Individuals of the Soybean Reference Cultivar Williams 82. Plant Physiol **155**: 645–655

Hawkins C, Liu Z (2014) A model for an early role of auxin in Arabidopsis gynoecium morphogenesis. Plant Evol Dev **5**: 327

Heisler MG, Atkinson A, Bylstra YH, Walsh R, Smyth DR (2001) SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. Dev Camb Engl **128**: 1089–1098

Heisler MG, Ohno C, Das P, Sieber P, Reddy GV, Long JA, Meyerowitz EM (2005) Patterns of Auxin Transport and Gene Expression during Primordium Development Revealed by Live Imaging of the Arabidopsis Inflorescence Meristem. Curr Biol **15**: 1899–1911

Hoffmann T, Kalinowski G, Schwab W (2006) RNAi-induced silencing of gene expression in strawberry fruit (Fragaria × ananassa) by agroinfiltration: a rapid assay for gene function analysis. Plant J **48**: 818–826

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

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

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

Huang F, Kemel Zago M, Abas L, van Marion A, Galvan-Ampudia CS, Offringa R (2010) Phosphorylation of Conserved PIN Motifs Directs Arabidopsis PIN1 Polarity and Auxin Transport. Plant Cell **22**: 1129–1142

Huang T, Harrar Y, Lin C, Reinhart B, Newell NR, Talavera-Rauh F, Hokin SA, Barton MK, Kerstetter RA (2014) Arabidopsis KANADI1 Acts as a Transcriptional Repressor by Interacting with a Specific cis-Element and Regulates Auxin Biosynthesis, Transport, and Signaling in Opposition to HD-ZIPIII Factors. Plant Cell **26**: 246–262

Hunter C (2006) Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in Arabidopsis. Development **133**: 2973–2981

Husbands AY, Chitwood DH, Plavskin Y, Timmermans MCP (2009) Signals and prepatterns: new insights into organ polarity in plants. Genes Dev **23**: 1986–1997

Ilegems M, Douet V, Meylan-Bettex M, Uyttewaal M, Brand L, Bowman JL, Stieger PA (2010) Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. Development **137**: 975–984

Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, Borst JW, Angenent GC (2009) SEPALLATA3: the "glue" for MADS box transcription factor complex formation. Genome Biol **10**: R24

Iwasaki M, Takahashi H, Iwakawa H, Nakagawa A, Ishikawa T, Tanaka H, Matsumura Y, Pekker I, Eshed Y, Vial-Pradel S, et al (2013) Dual regulation of ETTIN (ARF3) gene expression by AS1-AS2, which maintains the DNA methylation level, is involved in stabilization of leaf adaxial-abaxial partitioning in Arabidopsis. Development **140**: 1958–1969

Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LH, Araki T, Denoyes B, Foucher F (2012) The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. Plant J **69**: 116–25

Izhaki A, Bowman JL (2007) KANADI and Class III HD-Zip Gene Families Regulate Embryo Patterning and Modulate Auxin Flow during Embryogenesis in Arabidopsis. PLANT CELL ONLINE **19**: 495–508

Kanei-Ishii C, Sarai A, Sawazaki T, Nakagoshi H, He DN, Ogata K, Nishimura Y, Ishii S (1990) The tryptophan cluster: a hypothetical structure of the DNA-

binding domain of the myb protooncogene product. J Biol Chem **265**: 19990–19995

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

Kelley DR, Arreola A, Gallagher TL, Gasser CS (2012) ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in Arabidopsis. Development **139**: 1105–1109

Kepinski S, Leyser O (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature **435**: 446–451

Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS (2001) KANADI regulates organ polarity in Arabidopsis. Nature **411**: 706–709

Kidner C, Timmermans M (2007) Mixing and matching pathways in leaf polarity. Curr Opin Plant Biol **10**: 13–20

Koskela EA, Mouhu K, Albani MC, Kurokura T, Rantanen M, Sargent DJ, Battey NH, Coupland G, Elomaa P, Hytonen T (2012) Mutation in TERMINAL FLOWER1 reverses the photoperiodic requirement for flowering in the wild strawberry Fragaria vesca. Plant Physiol **159**: 1043–54

Krost C, Petersen R, Schmidt ER (2012) The transcriptomes of columnar and standard type apple trees (Malus x domestica) — A comparative study. Gene **498**: 223–230

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Res **19**: 1639–1645

Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. Seed Sci Res **15**: 281–307

Larsson E, Franks RG, Sundberg E (2013) Auxin and the Arabidopsis thaliana gynoecium. J Exp Bot **64**: 2619–2627

Larsson E, Roberts CJ, Claes AR, Franks RG, Sundberg E (2014) Polar auxin transport is essential for medial versus lateral tissue specification and vascular-mediated valve outgrowth in Arabidopsis gynoecia. Plant Physiol **166**: 1998–2012

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**: 1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The

Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl **25**: 2078–2079

Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol **10**: 50

Lin-Wang K, McGhie TK, Wang M, Liu Y, Warren B, Storey R, Espley RV, Allan AC (2014) Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca). Front Plant Sci **5**: 651

Li S (2014) Transcriptional control of flavonoid biosynthesis. Plant Signal Behav **9**: e27522

Liu T, Reinhart BJ, Magnani E, Huang T, Kerstetter R, Barton MK (2012) Of Blades and Branches: Understanding and Expanding the Arabidopsis Ad/Abaxial Regulatory Network through Target Gene Identification. Cold Spring Harb Symp Quant Biol **77**: 31–45

Ljung K (2013) Auxin metabolism and homeostasis during plant development. Development **140**: 943–950

Löfke C, Luschnig C, Kleine-Vehn J (2013) Posttranslational modification and trafficking of PIN auxin efflux carriers. Mech Dev **130**: 82–94

Long JA, Moan EI, Medford JI, Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. Nature **379**: 66–69

Mashiguchi K, Tanaka K, Sakai T, Sugawara S, Kawaide H, Natsume M, Hanada A, Yaeno T, Shirasu K, Yao H, et al (2011) The main auxin biosynthesis pathway in Arabidopsis. Proc Natl Acad Sci **108**: 18512–18517

McConnell JR, Barton MK (1998) Leaf polarity and meristem formation in Arabidopsis. Development **125**: 2935–2942

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20**: 1297–1303

Medina-Puche L, Cumplido-Laso G, Amil-Ruiz F, Hoffmann T, Ring L, Rodríguez-Franco A, Caballero JL, Schwab W, Muñoz-Blanco J, Blanco-Portales R (2014) MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria × ananassa fruits. J Exp Bot **65**: 401–417

Merelo P, Xie Y, Brand L, Ott F, Weigel D, Bowman JL, Heisler MG, Wenkel S (2013) Genome-Wide Identification of KANADI1 Target Genes. PLoS ONE **8**: e77341

Mockaitis K, Estelle M (2008) Auxin Receptors and Plant Development: A New Signaling Paradigm. Annu Rev Cell Dev Biol **24**: 55–80

Nakatsuka T, Yamada E, Saito M, Fujita K, Nishihara M (2013) Heterologous expression of gentian MYB1R transcription factors suppresses anthocyanin pigmentation in tobacco flowers. Plant Cell Rep **32**: 1925–1937

Nemhauser JL, Feldman LJ, Zambryski PC (2000) Auxin and ETTIN in Arabidopsis gynoecium morphogenesis. Development **127**: 3877–3888

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

Nole-Wilson S, Azhakanandam S, Franks RG (2010) Polar auxin transport together with AINTEGUMENTA and REVOLUTA coordinate early Arabidopsis gynoecium development. Dev Biol **346**: 181–195

Notaguchi M, Abe M, Kimura T, Daimon Y, Kobayashi T, Yamaguchi A, Tomita Y, Dohi K, Mori M, Araki T (2008) Long-Distance, Graft-Transmissible Action of Arabidopsis FLOWERING LOCUS T Protein to Promote Flowering. Plant Cell Physiol **49**: 1645–1658

Núñez-Elisea R, L. Davenport T (1998) Gibberellin and temperature effects on dormancy release and shoot morphogenesis of mango (Mangifera indica L.). Sci Hortic **77**: 11–21

Ogata K, Hojo H, Aimoto S, Nakai T, Nakamura H, Sarai A, Ishii S, Nishimura Y (1992) Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc Natl Acad Sci **89**: 6428–6432

Okada K, Ueda J, Komaki M, Bell C, Shimura Y (1991a) Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. Plant Cell **3**: 677–684

Okada K, Ueda J, Komaki MK, Bell CJ, Shimura Y (1991b) Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. Plant Cell **3**: 677–684

Palme K, Gälweiler L (1999) PIN-pointing the molecular basis of auxin transport. Curr Opin Plant Biol **2**: 375–381

Peer WA, Blakeslee JJ, Yang H, Murphy AS (2011) Seven Things We Think We Know about Auxin Transport. Mol Plant **4**: 487–504

Pekker I (2005) Auxin Response Factors Mediate Arabidopsis Organ Asymmetry via Modulation of KANADI Activity. PLANT CELL ONLINE **17**: 2899–2910

Pekker I, Alvarez JP, Eshed Y (2005) Auxin Response Factors Mediate Arabidopsis Organ Asymmetry via Modulation of KANADI Activity. Plant Cell **17**: 2899–2910

Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF (2001) APETALA1 and SEPALLATA3 interact to promote flower development. Plant J **26**: 385–394

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842

Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci **10**: 63–70

Ravi M, Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. Nature **464**: 615–8

Reinhardt D, Mandel T, Kuhlemeier C (2000) Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs. Plant Cell **12**: 507–518

Reinhardt D, Pesce E-R, Stieger P, Mandel T, Baltensperger K, Bennett M, Traas J, Friml J, Kuhlemeier C (2003) Regulation of phyllotaxis by polar auxin transport. Nature **426**: 255–260

Reyes-Olalde JI, Zuñiga-Mayo VM, Chávez Montes RA, Marsch-Martínez N, de Folter S (2013) Inside the gynoecium: at the carpel margin. Trends Plant Sci **18**: 644–655

Rubery PH, Sheldrake AR (1974) Carrier-mediated auxin transport. Planta **118**: 101–121

Sarojam R, Sappl PG, Goldshmidt A, Efroni I, Floyd SK, Eshed Y, Bowman JL (2010) Differentiating Arabidopsis Shoots from Leaves by Combined YABBY Activities. Plant Cell **22**: 2113–2130

Sattler R (1973) Organogenesis of flowers;: A photographic text-atlas. University of Toronto Press, Toronto, Buffalo

Scarpella E, Marcos D, Friml J, Berleth T (2006) Control of leaf vascular patterning by polar auxin transport. Genes Dev **20**: 1015–1027

Schaefer HM, Levey DJ, Schaefer V, Avery ML (2006) The role of chromatic and achromatic signals for fruit detection by birds. Behav Ecol **17**: 784–789

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proc Natl Acad Sci **95**: 5857–5864

Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus Antirrhinum. Plant Cell **18**: 831–51

Scotland RW, Wortley AH (2003) How many species of seed plants are there? Taxon **52**: 101–104

Scutt CP, Vinauger-Douard M, Fourquin C, Finet C, Dumas C (2006) An evolutionary perspective on the regulation of carpel development. J Exp Bot **57**: 2143–2152

Sessions A, Nemhauser JL, McColl A, Roe JL, Feldmann KA, Zambryski PC (1997) ETTIN patterns the Arabidopsis floral meristem and reproductive organs. Development **124**: 4481–4491

Sessions R (1997) Arabidopsis (Brassicaceae) flower development and gynoecium patterning in wild type and Ettin mutants. Am J Bot **84**: 1179–1179

Sessions RA, Zambryski PC (1995) Arabidopsis gynoecium structure in the wild and in ettin mutants. Development **121**: 1519–1532

Shani E, Yanai O, Ori N (2006) The role of hormones in shoot apical meristem function. Curr Opin Plant Biol **9**: 484–489

Shi J-H, Yang Z-B (2011) Is ABP1 an Auxin Receptor Yet? Mol Plant **4**: 635–640

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, et al (2011a) The genome of woodland strawberry (Fragaria vesca). Nat Genet **43**: 109–116

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, et al (2011b) The genome of woodland strawberry (Fragaria vesca). Nat Genet **43**: 109–16

Slovin JP, Schmitt K, Folta KM (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

Smith RS, Guyomarc'h S, Mandel T, Reinhardt D, Kuhlemeier C, Prusinkiewicz P (2006) A plausible model of phyllotaxis. Proc Natl Acad Sci **103**: 1301–1306

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in Arabidopsis. Plant Cell **2**: 755–767

Sohlberg JJ, Myrenås M, Kuusk S, Lagercrantz U, Kowalczyk M, Sandberg G, Sundberg E (2006) *STY1* regulates auxin homeostasis and affects apical-basal patterning of the Arabidopsis gynoecium. Plant J **47**: 112–123

Somerville C, Koornneef M (2002) A fortunate choice: the history of Arabidopsis as a model plant. Nat Rev Genet **3**: 883–889

Sorefan K, Girin T, Liljegren SJ, Ljung K, Robles P, Galván-Ampudia CS, Offringa R, Friml J, Yanofsky MF, Østergaard L (2009) A regulated auxin minimum is required for seed dispersal in Arabidopsis. Nature **459**: 583–586

Stahle MI, Kuehlich J, Staron L, von Arnim AG, Golz JF (2009) YABBYs and the Transcriptional Corepressors LEUNIG and LEUNIG_HOMOLOG Maintain Leaf Polarity and Meristem Activity in Arabidopsis. Plant Cell **21**: 3105–3118

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

Stepanova AN, Yun J, Robles LM, Novak O, He W, Guo H, Ljung K, Alonso JM (2011) The Arabidopsis YUCCA1 Flavin Monooxygenase Functions in the Indole-3-Pyruvic Acid Branch of Auxin Biosynthesis. Plant Cell **23**: 3961–3973

Sun J, Liu X, Yang T, Slovin J, Chen P (2014) Profiling polyphenols of two diploid strawberry (Fragaria vesca) inbred lines using UHPLC-HRMS(n.). Food Chem **146**: 289–98

Sun T (2011) The Molecular Mechanism and Evolution of the GA–GID1–DELLA Signaling Module in Plants. Curr Biol **21**: R338–R345

Sussex IM (1951) Experiments on the Cause of Dorsiventrality in Leaves. Nature **167**: 651–652

Tennessen JA, Govindarajulu R, Ashman T-L, Liston A (2014) Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps. Genome Biol Evol **6**: 3295–3313

Theißen G (2001) Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol **4**: 75–85

Theißen G, Saedler H (2001) Plant biology: Floral quartets. Nature **409**: 469–471

Titapiwatanakun B, Murphy AS (2009) Post-transcriptional regulation of auxin transport proteins: cellular trafficking, protein phosphorylation, protein maturation, ubiquitination, and membrane composition. J Exp Bot **60**: 1093–1107

Trigueros M, Navarrete-Gomez M, Sato S, Christensen SK, Pelaz S, Weigel D, Yanofsky MF, Ferrandiz C (2009) The NGATHA Genes Direct Style Development in the Arabidopsis Gynoecium. PLANT CELL ONLINE **21**: 1394–1409

Tromas A, Paque S, Stierlé V, Quettier A-L, Muller P, Lechner E, Genschik P, Perrot-Rechenmann C (2013) Auxin-Binding Protein 1 is a negative regulator of the SCFTIR1/AFB pathway. Nat Commun. doi: 10.1038/ncomms3496

Tuan PA, Bai S, Yaegaki H, Tamura T, Hihara S, Moriguchi T, Oda K (2015) The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biol **15**: 280

Turck F, Fornara F, Coupland G (2008) Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. Annu Rev Plant Biol **59**: 573–594

Ulmasov T, Hagen G, Guilfoyle TJ (1999) Dimerization and DNA binding of auxin response factors. Plant J **19**: 309–319

Ulmasov T, Hagen G, Guilfoyle TJ (1997) ARF1, a Transcription Factor That Binds to Auxin Response Elements. Science **276**: 1865–1868

Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K (2012) Dissecting plant genomes with the PLAZA comparative genomics platform. Plant Physiol **158**: 590–600

Vialette-Guiraud ACM, Scutt CP (2009) Carpel Evolution. *In* L stergaard, ed, Fruit Dev. Seed Dispersal. Wiley-Blackwell, Oxford, UK, pp 1–34

Waites R, Hudson A (1995) phantastica: a gene required for dorsoventrality of leaves in Antirrhinum majus. Development **121**: 2143–2154

Wang L, Zhao S, Gu C, Zhou Y, Zhou H, Ma J, Cheng J, Han Y (2013) Deep RNA-Seq uncovers the peach transcriptome landscape. Plant Mol Biol **83**: 365–377

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE **2**: e718

Wisniewska J (2006) Polar PIN Localization Directs Auxin Flow in Plants. Science **312**: 883–883

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

Wynn AN, Rueschhoff EE, Franks RG (2011) Transcriptomic Characterization of a Synergistic Genetic Interaction during Carpel Margin Meristem Development in Arabidopsis thaliana. PLoS ONE **6**: e26231

Xu T, Wen M, Nagawa S, Fu Y, Chen J-G, Wu M-J, Perrot-Rechenmann C, Friml J, Jones AM, Yang Z (2010) Cell Surface- and Rho GTPase-Based Auxin Signaling Controls Cellular Interdigitation in Arabidopsis. Cell **143**: 99–110

Xu W, Peng H, Yang T, Whitaker B, Huang L, Sun J, Chen P (2014) Effect of calcium on strawberry fruit flavonoid pathway gene expression and anthocyanin accumulation. Plant Physiol Biochem **82**: 289–98

Zhang Q, Folta KM, Davis TM (2014) Somatic embryogenesis, tetraploidy, and variant leaf morphology in transgenic diploid strawberry (Fragaria vesca subspecies vesca “Hawaii 4”). BMC Plant Biol **14**: 23

Zhang Y, Butelli E, De Stefano R, Schoonbeek HJ, Magusin A, Pagliarani C, Wellner N, Hill L, Orzaez D, Granell A, et al (2013) Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. Curr Biol **23**: 1094–100

Zhang Y, Li W, Dou Y, Zhang J, Jiang G, Miao L, Han G, Liu Y, Li H, Zhang Z (2015) Transcript Quantification by RNA-Seq Reveals Differentially Expressed Genes in the Red and Yellow Fruits of Fragaria vesca. PLoS ONE. doi: 10.1371/journal.pone.0144356

Zhao Y (2012) Auxin Biosynthesis: A Simple Two-Step Pathway Converts Tryptophan to Indole-3-Acetic Acid in Plants. Mol Plant **5**: 334–338

Zhu H, Dardick CD, Beers EP, Callanhan AM, Xia R, Yuan R (2011) Transcriptomics of shading-induced and NAA-induced abscission in apple (Malus domestica) reveals a shared pathway involving reduced photosynthesis, alterations in carbohydrate transport and signaling and hormone crosstalk. BMC Plant Biol **11**: 138