ABSTRACT

Title of Dissertation:	INVESTIGATING MOLECULAR MECHANISMS SPECIFYING DIVERSE ROSACEAE FRUIT TYPES THROUGH COMPARATIVE TRANSCRIPTOMIC ANALYSIS
	Muzi Li, Doctor of Philosophy, 2022
Dissertation directed by:	Professor, Zhongchi Liu, Department of Cell Biology and Molecular Genetics Associate Professor, Stephen Mount, Department of Cell Biology and Molecular Genetics

Rosaceae is a plant family with over 3,000 species including a number of economically important fruit-bearing species. Although plants in Rosaceae family have similar basic flower structure, their fruit flesh comes from distinct floral tissues. In drupe fruit, such as peach and plum, the ovary wall becomes enlarged and fleshy. In pome fruit, such as apple and pear, the fruit fleshy is mainly derived from the hypanthium that encases the ovary. In drupetum fruit, such as raspberry, numerous unfused ovaries each grow into a fleshy drupelet. In achenetum fruit, such as strawberry, the numerous unfused ovaries eventually dry up, but the receptacle, the stem tip that supports these ovaries, instead develops into the fruit flesh. By investigating and comparing the transcriptomes from these four Rosaceae fruits, peach (*Prunus persica*), apple (*Malus x domestica*), strawberry (*Fragaria vesca*), and raspberry (*Rubus idaeus*), at the earliest stages of fruit development, we gain important insights into the genetic mechanisms underlying

fleshy fruit diversity. The expression of B class MADS-box genes, *PISTILLATA*, *APETALA3* and *TM6*, shows negative correlation with the ability to form fleshy fruit tissues. Based on RNA transcript and phylogenetic analysis, *FBP9*, a MADS-box gene related to the E class, appears to be necessary but insufficient for flesh formation. In addition to the regulatory roles MADS-box genes play in fruit identity specification, extensive lignification of the strawberry ovary wall may contribute to the inability of strawberry ovary to become fleshy. Finally, a database (ROsaceae Fruit Transcriptome database, ROFT) is established for researchers to query for orthologous genes and their expression patterns during fruit development in the four species as well as to query for the tissue-specific and tissue- and stage-specific genes. Together, these findings provide the framework for functional investigations of fruit type specification and insights into the evolution of diverse fruit types in the Rosaceae family. The knowledge gained will advance our understanding in the evolution of fleshy fruits, a defining feature of angiosperm, and enable the creation of new fruit types for consumers.

INVESTIGATING MOLECULAR MECHANISMS SPECIFYING DIVERSE ROSACEAE FRUIT TYPES THROUGH COMPARATIVE TRANSCRIPTOMIC ANALYSIS

by

Muzi Li

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

Advisory Committee:
Professor Zhongchi Liu, Chair
Associate Professor Stephen Mount, Co-chair
Assistant Professor Philip Johnson
Professor Carlos Machado
Associate Professor Rob Patro

Acknowledgements

First, I would like to thank my advisors, Dr. Zhongchi Liu and Dr. Stephen Mount. Dr. Zhongchi Liu helped me to better understand and conduct my projects from biological perspectives, and taught me a lot about plant science and molecular genetics. As a female scientist, she totally understands the challenges I face during my graduate life. She gives me courage and shows me how to be a strong woman. Dr. Stephen Mount guided my computational analyses and enriched my knowledge of computational biology by lending me books and discussing the cutting-edge research and technology with me. He has been a great listener that I can always turn to when I need help. Both Dr. Zhongchi Liu and Dr. Stephen Mount provide me unlimited support and guidance, and they always encourage me to think outside the box.

I also want to acknowledge my committee members for their constructive feedback, which greatly helped me improve my work. I learnt how to use R to perform different statistical analyses by taking Dr. Philip Johnson's course, Statistics & Modeling for Biologists. I also learnt the underlying algorithms of the bioinformatics tools that I used a lot in my research from Dr. Rob Patro in his class, Algorithms, Data Structure and Inference for High-throughput Genomics. And I really appreciate that Dr. Carlos Machado agreed to serve as the Dean's representative and to provide expert advice and assistance.

I'm also grateful for the helps from my collaborators. Drs. Kelsey Galimba, Chris Dardick, and Ann Callahan at the Appalachian Fruit Research Station of USDA

collected the RNA-Seq data from early developing peach and apple fruits. Dr. Junhui Zhou, a previous postdoc in the Liu lab, characterized the early fruit development of red raspberry, sequenced RNA isolated from dissected fruit tissues at multiple early developmental stages, and applied plant hormones to the emasculated flowers to investigate their impacts on fruit growth. The wet-lab experiments conducted by my collaborators laid the foundation for my projects. I also want to thank Yuwei Xiao, a graduate student in the Liu lab, for retrieving the developmental programmed cell death genes in Arabidopsis, examining the presence of the important type II MADS-box genes in the eudicots, interpreting the results of my RNA-Seq analyses with Dr. Zhongchi Liu and me, and performing the experiments to validate our hypotheses. In addition, I'm thankful to Dr. Yongping Li at Chinese University of Hong Kong for his efforts in the red raspberry genome assembly and annotation. And I thank the undergraduate student, Andrew Tong, for his help with the database construction.

The people in the Liu lab were always happy to answer the biology questions I had, which allowed me to come up with new ideas. Meanwhile, I have learned a lot of novel computational approaches from the people in the Mount lab, and they also gave me useful suggestions when I had technical issues.

I feel lucky to have many friends, especially Dr. Fuxi Wang, Yuija Li, Bixuan Wang and Dr. Yile Huang. I would like to thank Dr. Fuxi Wang who worked in the Liu lab and obtained her Ph.D. degree a few months ago for supporting me a lot during the pandemic. Because of her accompany, I never felt lonely when I needed to work late.

I also appreciate the help from my roommate, Yujia Li. Her words of wisdom always inspired me and comforted me when I was upset. I loved to chat with Bixuan Wang, a member of the Mount lab, because she always brought a lot of fun to me. And I have known Dr. Yile Huang for over ten years. She was my roommate when I was an undergraduate student. No matter how far we are, we are still each other's best friend.

I'd like to thank the NSF (National Science Foundation) grant (IOS 1444987) for supporting the transcriptomic comparative analysis of the Rosaceae fruits. And I'm also grateful that my research was supported in part by NSF award DGE-1632976 from COMBINE (Computation and Mathematics for Biological Networks).

Finally, I'd like to thank my family, especially my parents, for their unconditional love and support. They always have the confidence in me, and respect my choices. My parents treat me like a friend and understand my feelings, therefore I'm willing to share my joys and pains with them. I really want to thank them for being the best parents in the world.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGL, AGAMOUS-LIKE
AP3, APETALA 3
ARF, Auxin Response Factor
AUX1, AUXIN-RESISTANT1
BUSCO, Benchmarking Universal Single-Copy Orthologs
CPS, ent-copalyl diphosphate synthase
CRISPR-CAS9, Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9
DAA, Days After Anthesis
DAB, Days After Bloom
DE, Differentially Expressed or Differential Expression
DEG, Differentially Expressed Genes
dPCD, developmental Programmed Cell Death
DPA, Days Post Anthesis
eFP, electronic Fluorescent Pictograph
EST, Expressed Sequence Tag
GA, Gibberellic Acid
GA2ox, GA 2-oxidase
GA3ox, GA 3-oxidase
GA20ox, GA 20-oxidase
GDR, Genome Database for Rosaceae
GEO, Gene Expression Omnibus
GO, Gene Ontology
IAA, INDOLE-3-ACETIC ACID INDUCIBLE
KAO, ent-kaurenoic acid oxidase
KO, ent-kaurene oxidase
KS, ent-kaurene synthase
LAX, LIKE AUX1
NCBI, National Center for Biotechnology Information
PASA, Program to Assembled Spliced Alignments
PCA, Principal Component Analysis
PI, PISTILLATA
PIN, PIN-FORMED
PlantTFDB, Plant Transcription Factor Database
PlantRegMap, Plant Transcriptional Regulatory Map
PMN, Plant Metabolic Network
RNA-Seq, RNA Sequencing
ROFT, ROsaceae Fruit Transcriptome database
SAUR, Small Auxin Up-regulated RNA
SEP, SEPALLATA
SGR, Strawberry Genomic Resources
SMRT, Single Molecule Real-Time
SOG, Syntenic OrthoGroup
TAA1, TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1
TAR, TRYPTOPHAN AMINOTRANSFERASE RELATED

TF, Transcription Factor
WGD, Whole Genome Duplication
YUC, YUCCA (flavin monooxygenase)

Chapter 1: Introduction

1.1 Rosaceae Family

Rosaceae, a large plant family of more than 3,000 species, consists of many economically important fruit and ornamental crops, such as peach, apple, strawberry, raspberry, cherry, and rose. These horticultural crops are not only important economic drivers in many regions of the world, but also major sources of human nutrition. Additionally, due to the diversity of fruit types in Rosaceae, this plant family offers excellent opportunities for investigations into fleshy fruit diversity, evolution, and development. By comparing the early fruit development of the Rosaceae species that produce different fruit types, the genetic mechanisms underlying the diversity of fleshy fruit types, specifically the mechanism of fruit identity specification will be revealed, which enables future manipulation of fruit types and engineering fruit formation without the need for pollination and fertilization, and improvement of fruit quality.

Recently, 125 Rosaceae species' transcriptomes were sequenced and used to construct a well-resolved phylogeny of Rosaceae (Xiang et al., 2017). This phylogenetic study included 124 Rosaceae species and 24 other angiosperm species. The Rosaceae species were grouped into three subfamilies (Amygdaloideae, Rosoideae, and Dryadoideae) and 16 tribes (**Figure 1.1**). Apple and pear are from the tribe Maleae and are very close to each other in the phylogenetic tree. The *Prunus* species, including plum, apricot, almond, peach, and cherry, belong to the tribe Amygdaleae. The tribes Maleae and Amygdaleae are both in the subfamily Amygdaloideae. Rosoideae, another subfamily of Rosaceae, includes three important tribes, Potentilleae (strawberry), Roseae (rose), and Rubeae (blackberry and raspberry). And compared with Rubeae, Roseae is more closely related to Potentilleae.

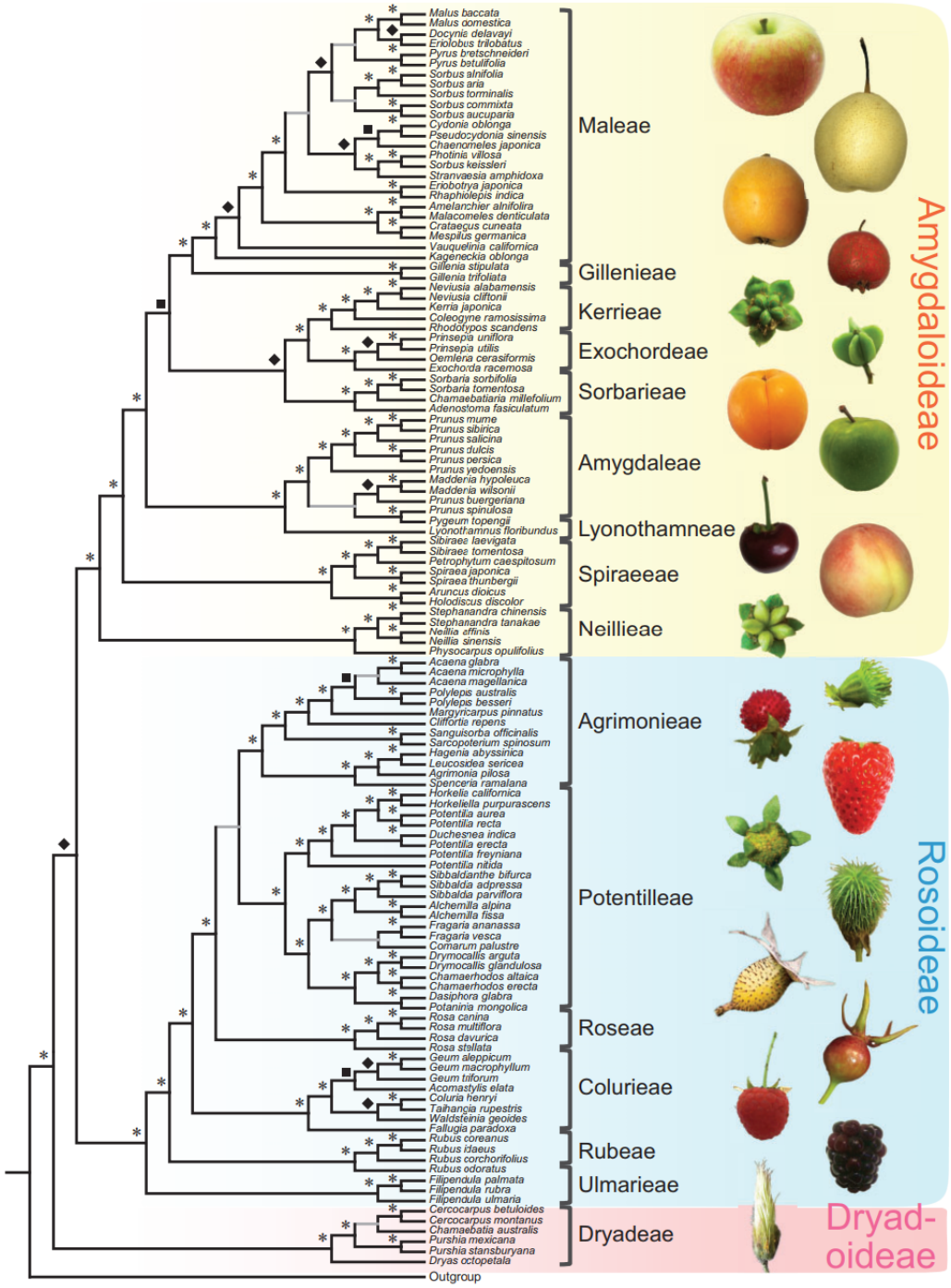

Figure 1.1 Rosaceae phylogeny

The figure was adapted from Xiang et al., 2017.

Plants in Rosaceae family have similar basic flower structure consisting of five sepals, five petals, and numerous stamens. However, the structure of their female organ, the carpel, varies from species to species. A carpel is usually composed of stigma, style and ovary. While the stigma and style are involved in receiving the pollen, the ovary consists of an ovary wall encasing one or more ovules inside. There could be multiple or single carpels that can be fused or independent. In peach, each flower has one single carpel that simply attaches to the base of the hypanthium, a cup-like green tissue surrounding the ovary (Figure 1.2A) (Z. Liu et al., 2020). However, in apple, the flower has connate carpels that fuse to the hypanthium (Figure 1.2B) (Z. Liu et al., 2020). Moreover, in drupetum and achenetum fruits, there are numerous unfused carpels sitting on the dome-like receptacle (the stem tip) (Figure 1.2C, D) (Z. Liu et al., 2020). The various fruit types are sometimes contributed by differences in the floral structure.

Upon successful fertilization, the ovary wall further develops to become the fruit that facilitates seed dispersal. The ovary wall derived fruits are more prevalent in plants and often referred to as “botanical fruit”. However, fruit flesh could also develop from non-ovary tissues, in which case, they are referred to as “accessory fruit”. In Rosaceae, some species develop the botanical fruits while others develop into accessory fruits, therefore, Rosaceae is a great system to investigate mechanisms of fruit type diversity.

In drupe fruit, such as peach and plum (**Figure 1.2A**), post fertilization, the middle layer of the ovary wall (mesocarp) enlarges and becomes the fruit flesh, the outermost layer (exocarp) of the ovary wall forms the protective skin of the fruit, and the innermost layer (endocarp) lignifies to form a hard stone encasing the seeds (Z. Liu et al., 2020). In pome fruit, a type of accessory fruit, such as apple and pear (**Figure 1.2B**), the hypanthium eventually develops

into the fruit flesh. In the drupetum fruit, such as raspberry (**Figure 1.2C**), every individual carpel grows into a fleshy fruit (drupelet) (Z. Liu et al., 2020). And in each drupelet, stone is formed to protect the seed like a drupe fruit. However, in achenetum fruit, such as strawberry (**Figure 1.2D**), the fruit flesh is derived from the receptacle, another example of accessory fruit, while the carpels dry up on the fruit surface (Z. Liu et al., 2020).

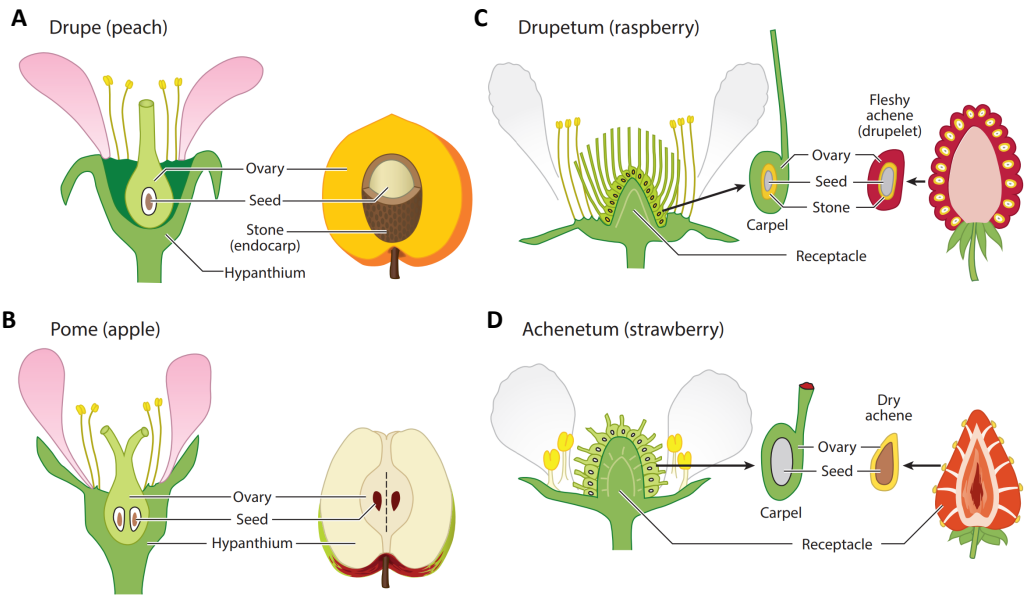

Figure 1.2 Four typical fleshy fruit types in Rosaceae

Flower and fruit structure in peach **(A)**, apple **(B)**, raspberry **(C)**, and strawberry **(D)**. The figure was adapted from Z. Liu et al., 2020.

The Rosaceae phylogeny mentioned above allows for a better understanding of fruit type evolution in Rosaceae (**Figure 1.3**) (Z. Liu et al., 2020; Xiang et al., 2017). It was hypothesized that dry fruit achenetum was the ancestral fruit type for Rosoideae and could develop into three different fleshy fruit types through distinct evolutionary paths (**Figure 1.3B**). In the first path, the outer part of each achene became thick and fleshy resulting in drupetum-type fruit (raspberry). In the second path, strawberry was derived from the enlarged receptacle with achenes that underwent few changes. In the third path giving rise to rose, the

hypanthium with increased fleshiness grew upwards and trapped the achenes inside it. In Amygdaloideae, the hypothetical ancestor, dry fruit follicetum, could evolve into drupe and pome fruits (**Figure 1.3c**). The carpels were either reduced in number leading to drupe fruit with fleshy ovary wall or fused together producing pome fruit with fleshy hypanthium.

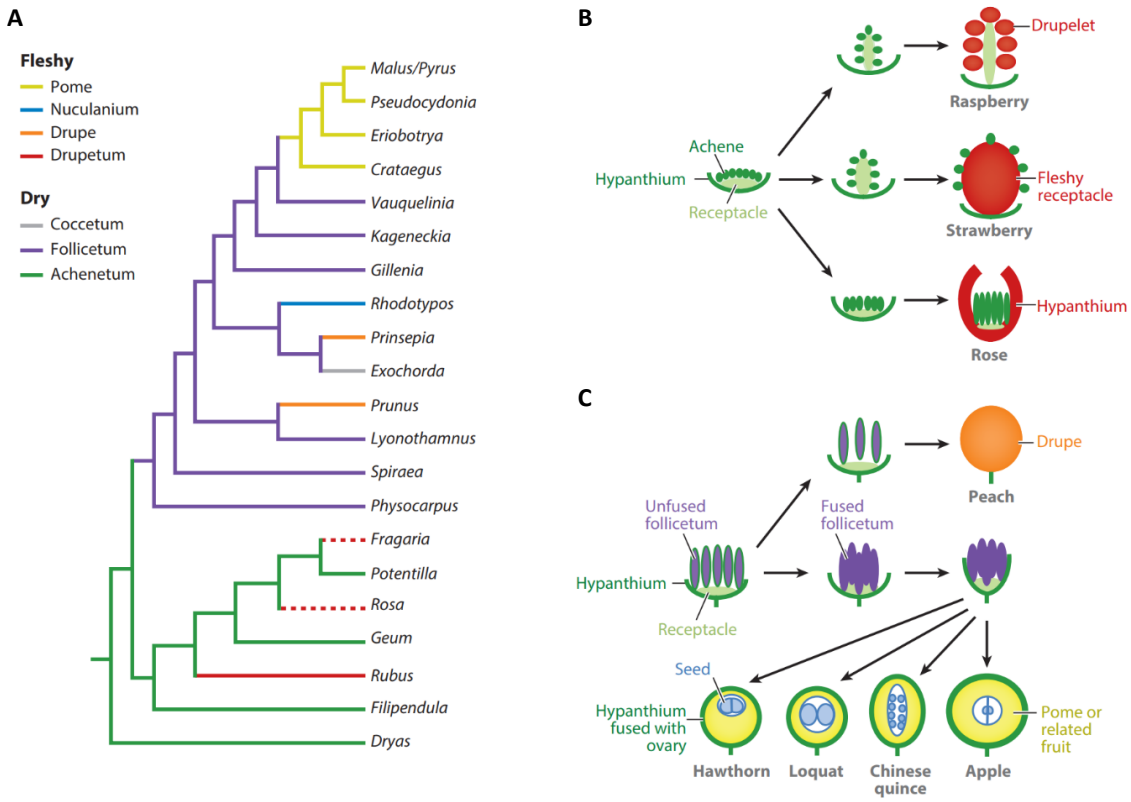

Figure 1.3 Evolution of different fruit types of Rosaceae

(A) Rosaceae phylogeny. The tip labels are Rosaceae genera. The branch colors represent different fruit types. The red dashed lines specify the genera with achenetum-type fruit. **(B)** Proposed evolutionary history of three fleshy fruit types in Rosoideae. **(C)** Proposed evolutionary history of two fleshy fruit types in Amygdaloideae. The figure was adapted from Z. Liu et al., 2020.

Although the Rosaceae phylogeny uncovers the possible evolutionary histories of different fleshy fruit types, a number of fundamental questions have remained unanswered. In this study, I asked the question what genetic factors cause distinct floral organs to develop into fruit flesh that result in diverse fruit types in the Rosaceae family. To investigate the fruit

diversity of Rosaceae, peach, apple, strawberry and raspberry were chosen as the representatives of different fleshy fruit types. All the four species are widely grown fruit crops and extensively studied as models for tree fruits. Further, all of them have publicly available genomes (Daccord et al., 2017; Edger et al., 2018; Verde et al., 2017; Wight et al., 2019). The high-throughput RNA-Seq data collected from comparable tissues of these four fruit types in the Liu laboratory allowed me to examine and compare the transcriptome profiling of the four Rosaceae fruits and identify candidate genes and pathways that may provide the instructions for the development of a fruit.

1.2 Flower and fruit development

1.2.1 ABC model of flower development

Flowering plants, angiosperms, are the most successful land plants on earth owing to their ability to form the complex reproductive structure, the flower. How do flowers evolve and develop? The ABC model of flower development established in the 1990s provided key insights to this question. The model was first published in 1991 by analyzing single, double and triple homeotic mutants of Arabidopsis (Bowman et al., 1991, 2012). The details of the model were further improved by a series of molecular genetic experiments (Colombo et al., 1995; Ditta et al., 2004; Egea-Cortines et al., 1999; Favaro et al., 2003; Honma & Goto, 2001; Pelaz et al., 2000; Theißen, 2001). It was found that the identity of each floral organ, sepal, petal, stamen, or carpel, was determined by protein tetramers formed by the ABCDE genes. The extended model includes A class gene (AP1), B class genes (AP3 and PI), C class gene (AG), D class genes (SHP and STK), and E class genes (SEP1, SEP2, SEP3, and SEP4). A and E class genes control sepal identity, A, B and E class genes determine petals, B, C and E class genes specify stamens, C and E class genes are required for carpel development, and C, D, and E class genes act together to produce ovules (**Figure 1.4**) (Theißen et al., 2016).

Interestingly, the ABCDE genes almost all encode type II MADS box transcription factors with a highly conserved MADS box DNA-binding domain (Theißen et al., 2016).

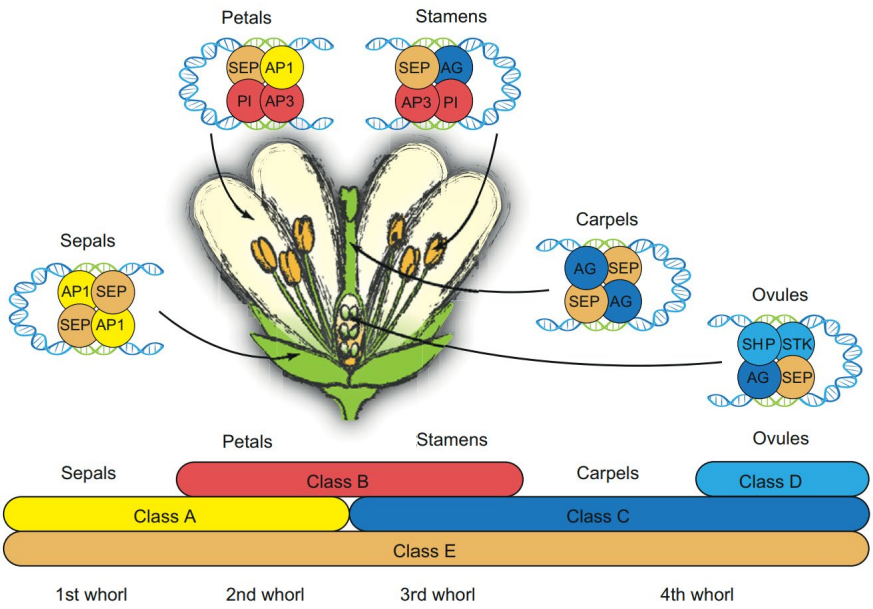

Figure 1.4 The ABC model of flower development with D and E class genes added later in research

Different combinations of tetrameric MADS box proteins specify different floral organs in *Arabidopsis thaliana*. The figure was adapted from Theißen et al., 2016.

Many of the genetic experiments that revealed the above findings were conducted using *Arabidopsis thaliana*. However, the flowers are morphologically distinct from one another in different plant species. Will the ABC model apply to different types of flowers? In daisies, even though the whole flowers, flower head, look different from *Arabidopsis* flowers, the development of each small flower in the central disc is still controlled by the ABC genes (**Figure 1.5A**) (Irish, 2017). In the cultivated roses, the extra petals are resulted from the restricted expression of C class gene that causes a shifted A/C boundary (**Figure 1.5B**) (Dubois et al., 2010; Irish, 2017). In tulips, the extended expression of B class genes leads to the formation of tepals (**Figure 1.5C**) (Irish, 2017; Kanno et al., 2003). Despite the modifications of the ABC model, the fundamental rules underlying the ABC program appear

to be conserved among the flowering plants (Irish, 2017). The fruit is developed from the flower, therefore specific genes, specifically MADS box genes, could work together in a conserved manner to determine fruit identity, i.e., the ability of a floral organ for develop into a fruit tissue post-fertilization.

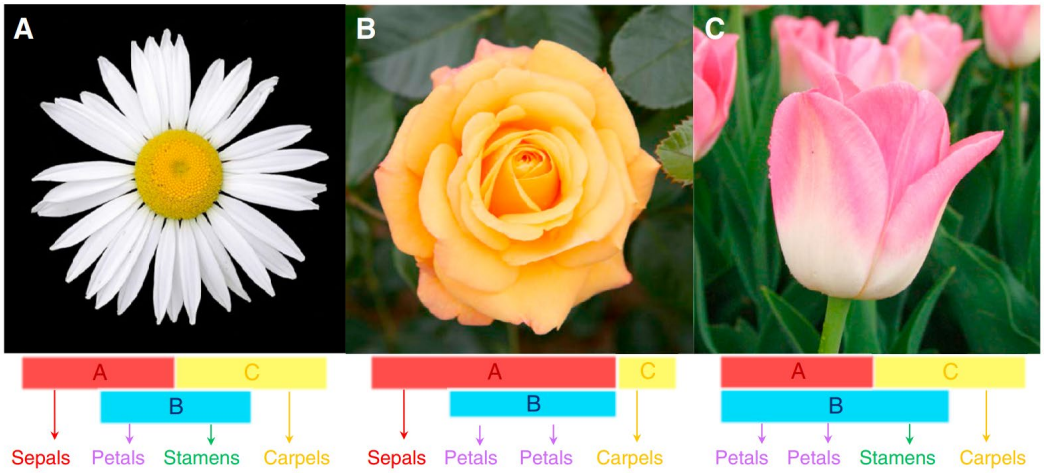

Figure 1.5 Variations in flower morphology and ABC model

Changes in the expression domains of the ABC genes may explain the diversity of floral forms as shown in daisy **(A)**, rose **(B)**, and tulip **(C)**. The figure was adapted from Irish, 2017.

1.2.2 Rosaceae fruit development

In many plants, fruit development is broadly composed of four phases (Gillaspy et al., 1993; Mayorga-Gómez & Nambeesan, 2020). In the first phase (fruit set), the plant decides whether to abort or proceed with fruit development depending on whether pollination and fertilization are successful. In the second phase, cell division is activated that leads to fruit growth. In the third phase, cell expansion begins and helps increase fruit size rapidly. In the fourth phase, fruit ripens and generally becomes softer and tastier. Below, the specific development processes of the major Rosaceae fruits will be described in detail.

In peach (*Prunus persica* 'Dixiland'), the fruit development is divided into five stages (**Figure 1.6**) (Lombardo et al., 2011). Stage E, roughly from 0 DAB (days after bloom) to 22

DAB, is the early period of fruit development. In stage S1 (about 23-37 DAB), there is an obvious increase in the fruit size. However, in the following stage S2 (about 38-66 DAB), the fruit grows extremely slowly and the stone starts to form. In stage S3 (about 67-94 DAB), the fruit growth becomes rapid again. And eventually, the fruit gradually ripens during stage S4 that approximately starts from 95 DAB.

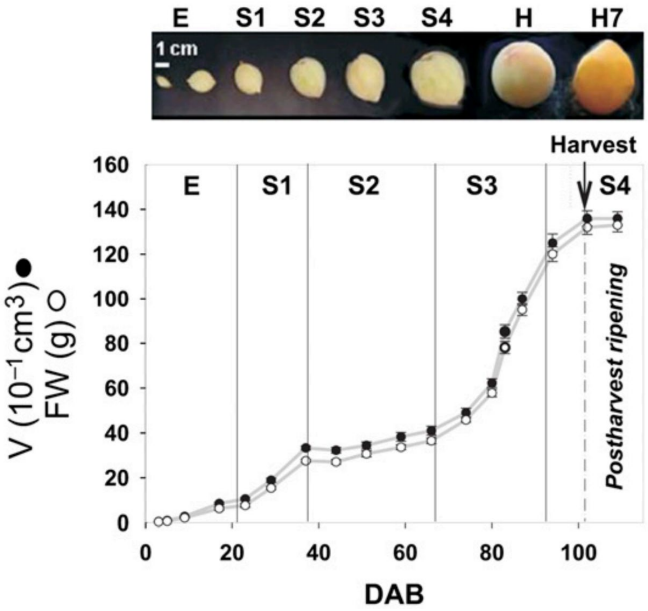

Figure 1.6 Peach fruit development

The top of the figure shows the peach fruits collected at different developmental stages. H indicates that the peach fruits were harvested at 102 DAB, and H7 indicates that the fruits continued to ripen for 7 days at 20 ºC after harvest. The data points on the two growth curves represent the volume (black) and fruit weight (white) of the peach fruits at stage E (3, 5, 9, and 17 DAB), stage S1 (23, 29, and 37 DAB), stage S2 (44, 51, 59, and 66 DAB), stag S3 (74, 80, 83, and 87 DAB), and stage S4 (94 and 102 DAB, and 7 DAH (days after harvest)). The figure was adapted from Lombardo et al., 2011.

Different from peach, apple fruit (*Malus domestica* 'Royal Gala') doesn't undergo lignin formation. The apple fruit cells expand at an increased rate and start to accumulate starch immediately after cell division at 35 DAA (days after anthesis) (**Figure 1.7**) (Janssen et al., 2008). Moreover, the rate of cell expansion and starch accumulation reaches its peak at 60 DAA. Afterwards, the apple fruit continues to grow but at a decreased rate until it ripens. And

the starch levels begin to decline after 87 DAA. By 146 DAA, the fruit finally becomes a fully developed apple with appealing color and flavor.

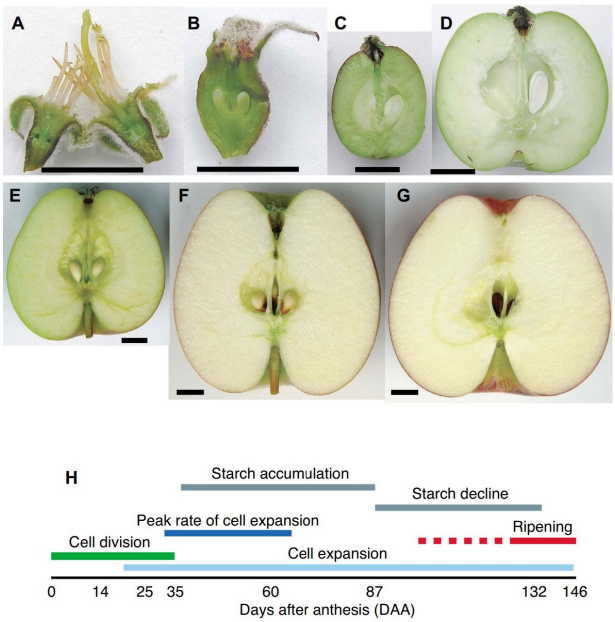

Figure 1.7 Apple fruit development

(A-G) Apple fruits collected at 0, 14, 35, 60, 87, 132 and 146 DAA. Scale bar: 1cm. **(H)** The timeline of important biological events during apple fruit development. The dashed and solid red lines stand for the pre-climacteric and climacteric stages, respectively. The figure was adapted from Janssen et al., 2008.

Strawberry fruit development (Fragaria vesca 'Yellow Wonder 5AF7') is composed of 12 stages containing seven early stages (S1-S7) and five ripening stages (RS1-RS5) **(Figure 1.8)** (Liao et al., 2018). And similar to apple, strawberry fruit doesn't form stone neither. The strawberry fruit gradually grows larger from S1 to S5, which is followed by a dramatic increase in the fruit weight between S5 and RS1. Subsequently, the fruit grows at a declined rate during fruit ripening.

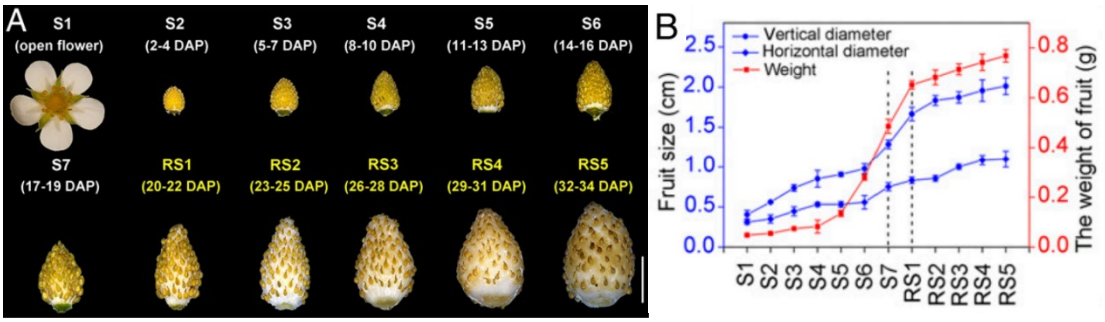

Figure 1.8 Strawberry fruit development

(A) 12 developmental stages of strawberry fruit growth. DAP: days after pollination. Scale bar: 1cm. **(B)** Changes in fruit length, width and weight over fruit development. The figure was adapted from Liao et al., 2018.

Raspberry (*Rubus chingii* 'L7') fruit development comprises eight stages that include small green (SG, 7 DPA (days post anthesis)), medium green (MG, 14 DPA), big green I (BGI, 21 DPA), big green II (BGII, 28 DPA), big green III (BGIII, 35 DPA), green-to-yellow (GY, 42 DPA), yellow-to-orange (YO, 48 DPA), and red (Re, 54DPA) (**Figure 1.9**) (Z. Chen et al., 2021). Raspberry fruit follows a similar growth pattern to peach fruit. Between two stages of fast fruit growth (7-21 DPA and 35-54 DPA), there is a period (21-35 DPA) with few changes in fruit size.

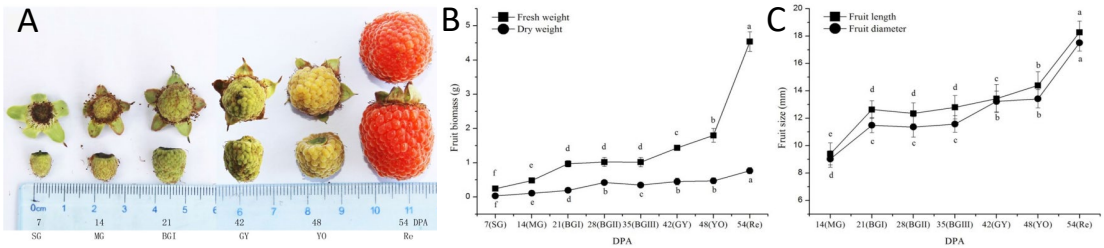

Figure 1.9 Raspberry fruit development

(A) Raspberry fruits derived from important stages of fruit development. **(B)** Fruit biomass of fleshy fruit and dry fruit during fruit development. **(C)** Fruit length and diameter during fruit development. The figure was adapted from Chen et al., 2021.

In this study, we are interested in identifying the potential important genes that can specify the fruit identity. Since the fruit identity is usually determined early in fruit development, we collected data from early developing fruits and mainly focused on the changes happening before and after pollination and fertilization.

1.3 Plant hormones: auxin and gibberellic acid (GA)

1.3.1 Auxin and GA pathways

It was known that flowers can enter the phase for fruit development only when pollination and fertilization are successful. The decision made by the flower to enter fruit development is called fruit set. It implies that pollination or fertilization can trigger certain floral compartment(s) to initiate development into a fruit. However, seedless (parthenocarpic) fruits can be achieved by applying plant hormones, such as auxin and GA, to unpollinated flowers in the Rosaceae species, which indicates that the hormonal signals positively control the Rosaceae fruit development and bypass the need for pollination/fertilization (Cong et al., 2019; Crane et al., 1960; Galimba et al., 2019; Kang et al., 2013). For instance, GA (GA3), synthetic auxin (NAA), and both (GA3+NAA) were applied to unpollinated woodland strawberry, respectively (Kang et al., 2013). The GA3-treated and NAA-treated strawberry both become enlarged but have smaller fruit size than the hand-pollinated strawberry (**Figure 1.10**). Only the strawberry treated with both GA3 and NAA is able to grow as large as the pollinated control (**Figure 1.10**). It suggests that the auxin-GA interaction also plays an important role in strawberry fruit development. Therefore, pollination and fertilization serve to trigger the biosynthesis of phytohormones, whose production subsequently leads to fruit initiation.

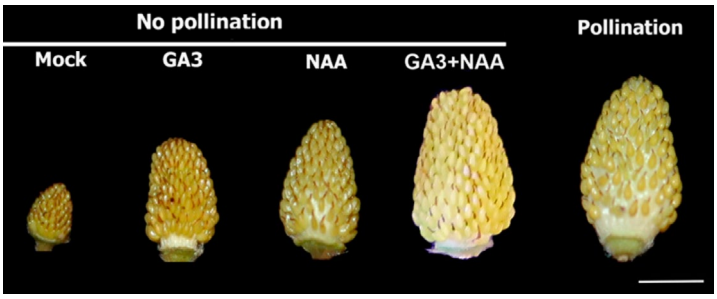

Figure 1.10 GA and auxin (NAA)-treated strawberry fruit without pollination still enlarges

The figure was adapted from Kang et al., 2013. Mock: mock-treated, GA3: GA-treated, NAA: auxin-treated, GA3+NAA: GA- and auxin-treated, Pollination: hand-pollinated. Woodland strawberry cultivar: 'Yellow Wonder 5AF7'. Scale bar: 5mm.

Auxin and GA appear to be most well studied regarding their roles in initiating Rosaceae fruit development. Genes in the auxin and GA pathways (**Table 1.1**) are described in further detail below. Auxin (indole-3-acetic acid), a small amino-acid-like molecule, is synthesized from the amino acid tryptophan by TAA1/TAR (TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1/TRYPTOPHAN AMINOTRANSFERASE RELATED) and YUC (YUCCA, flavin monooxygenase) proteins (**Figure 1.11A**) (Hofmann, 2011; Zhao, 2014). Other genes are involved in auxin homeostasis including GH3 proteins that catalyze auxin conjugation to temporarily inactivate the auxin or irreversibly degrade it (Chapman & Estelle, 2009; Ludwig-Müller, 2016) and auxin transporters such as PIN (PIN-FORMED), an auxin efflux carrier, and AUX1/LAX (AUXIN-RESISTANT1/LIKE AUX1), and auxin influx carriers (Chapman & Estelle, 2009; Geisler, 2021). Auxin signaling involves two major signaling proteins, ARF (auxin response factor) and IAA (INDOLE-3-ACETIC ACID INDUCIBLE) (**Figure 1.11B**). In the absence of auxin, IAA proteins interact with ARF transcription factors and prevent ARF proteins from regulating the expression of the downstream genes (He & Yamamuro, 2022; Leyser, 2018). In the presence of auxin, IAA proteins are degraded by the ubiquitin/26s proteasome pathway, which enables ARF proteins to activate or repress their target genes (He & Yamamuro, 2022; Leyser, 2018).

Proteins	Functions	Reference
TAA1, TAR, YUCCA	Auxin biosynthetic proteins	Hofmann, 2011; Zhao, 2014
GH3	Auxin homeostasis regulators	Chapman & Estelle, 2009; Ludwig-Müller, 2016
PIN	Auxin efflux carriers	Chapman & Estelle, 2009; Geisler, 2021
AUX1/LAX	Auxin influx carriers	Chapman & Estelle, 2009; Geisler, 2021
IAA, ARF	Auxin responsive proteins	He & Yamamuro, 2022; Leyser, 2018
CPS, KS, KO, KAO, GA20ox, GA3ox	GA biosynthetic proteins	NAKAYAMA et al., 2005
GA2ox	GA degradation enzymes	Yamaguchi & Kamiya, 2000
RGA, GAI, RGL1, RGL2, RGL3	DELLA proteins, GA responsive proteins	Davière & Achard, 2016; He & Yamamuro, 2022

Table 1.1 Important genes that participate in auxin and GA pathways

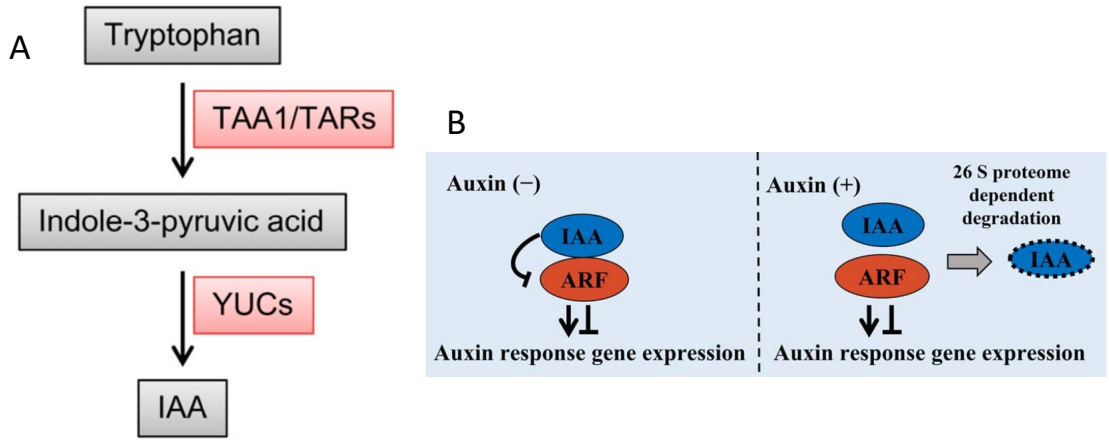

Figure 1.11 Auxin biosynthesis and signaling pathways

(A) Auxin biosynthesis pathway. TAA1/TAR1 and YUC are the two key enzymes. **(B)** Auxin signaling mechanism mediated by IAA and ARF transcription factors. A and B are adapted from Hofmann, 2011 and He & Yamamuro, 2022, respectively.

The biosynthesis of bioactive GA requires multiple steps catalyzed by enzymes including CPS (ent-copalyl diphosphate synthase), KS (ent-kaurene synthase), KO (ent-kaurene oxidase), KAO (ent-kaurenoic acid oxidase), GA20ox (GA 20-oxidase) and GA3ox (GA 3-oxidase) (**Figure 1.12A**) (NAKAYAMA et al., 2005). The bioactive GA can be deactivated by GA2ox (GA 2-oxidase) via 2β-hydroxylation (Yamaguchi & Kamiya, 2000). Analogous

to auxin signaling, in the absence of GA, DELLA proteins (RGA, GAI, RGL1, RGL2, and RGL3) act as repressors that bind to transcription factors (TFs) and inhibit their activities (Davière & Achard, 2016; He & Yamamuro, 2022). When GA exists, DELLA proteins are degraded by the 26S proteosome-mediated pathway; TFs are released from DELLA inhibition and able to fulfil their functions (**Figure 1.12B**) (Davière & Achard, 2016; He & Yamamuro, 2022).

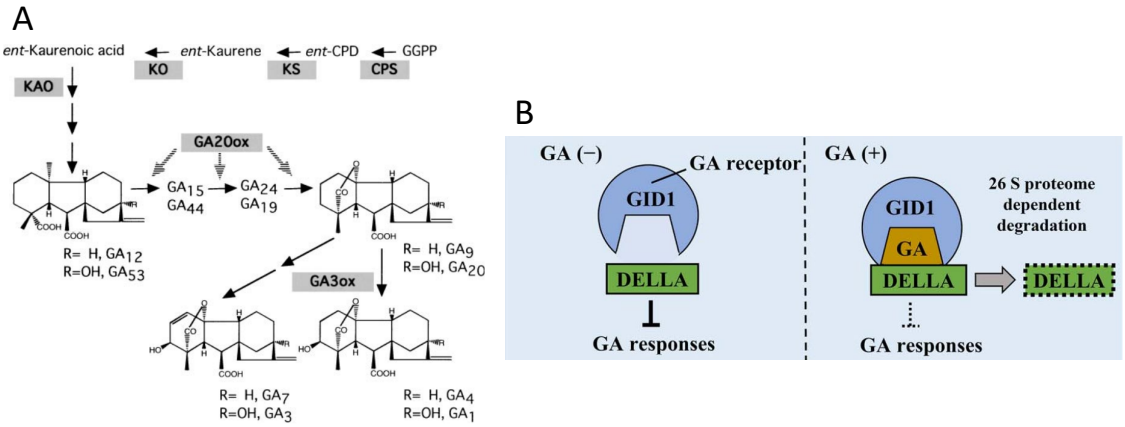

Figure 1.12 GA biosynthesis and signaling pathways

(A) GA biosynthesis pathway, adapted from NAKAYAMA et al., 2005. **(B)** GA signaling mechanism, adapted from He & Yamamuro, 2022.

1.3.2 Function of auxin and GA pathway genes in fruit development

A considerable amount of research was conducted to investigate the functions of auxin and GA pathway genes in different fruits. In tomato, overexpression of *SlGA20ox* (García-Hurtado et al., 2012), mutations in *SlARF8* (Goetz et al., 2007), and silencing of *SlARF5* (S. Liu et al., 2018), *SlARF7* (De Jong et al., 2009) and *SlIAA9* (H. Wang et al., 2005) all led to parthenocarpic fruits (He & Yamamuro, 2022), suggesting their influence on fruit set. Furthermore, the ARF7/IAA9 complex and a DELLA protein were reported to regulate tomato fruit initiation together (He & Yamamuro, 2022; Hu et al., 2018). It indicates that there is a crosstalk between auxin and GA that is important for fruit development. In grape,

reduced expression of *VvARF7* and *VvIAA9* was observed after application of GA at pre-bloom, which also supports the interplay between auxin and GA during fruit set (He & Yamamuro, 2022; C. J. Jung et al., 2014).

Compared to tomato, fewer studies have been carried out to investigate the impact of auxin and GA pathways genes on Rosaceae fruit development. Woodland strawberry is one of the Rosaceae species whose fruit development processes have been better characterized. The experiments were performed in woodland strawberry to uncover the possible genetic mechanisms underlying the crosstalk between auxin and GA (Zhou et al., 2021). CRISPR-knockout of *FveRGA1*, a DELLA protein-encoding gene, produced parthenocarpic strawberry fruits, suggesting FveRGA1/DELLA a negative regulator of strawberry fruit set. FveARF8 was also shown to negatively regulate fruit growth. Though *fvearf8* loss-of-function mutants failed to produce parthenocarpic fruit, they formed larger fruits than wide-type woodland strawberry after fertilization. Additionally, the interactions between FveRGA1 and FveARF8, FveIAA4 and FveARF8, and FveARF8 and FveGID1c were demonstrated in the same study. Based on the above findings, it was proposed that FveRGA1 and FveIAA4 inhibit FveARF8 and other TFs from regulating the downstream targets in the absence of auxin and GA (**Figure 1.13**). After the transport of auxin and GA from the achenes to the receptacle following fertilization, the activator TFs (to be identified) are allowed to stimulate the expression of the fruit growth-related genes, such as *FveCYCD2;1, FveSAUR1,* and *FveEXPL_B1* (**Figure 1.13**). And FveARF8 can also be released to suppress *FveGID1c* encoding the GA receptor (**Figure 1.13**).

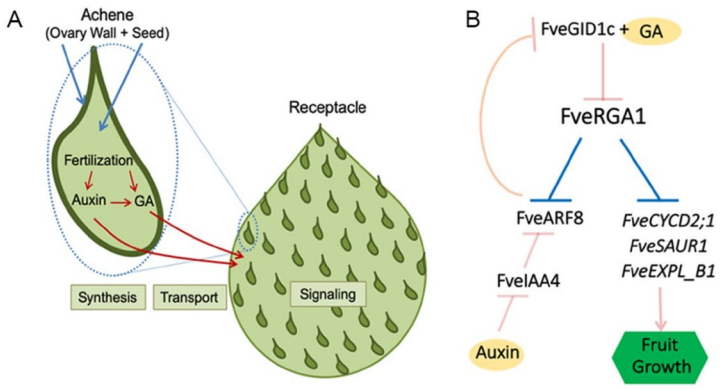

Figure 1.13 Proposed models of strawberry fruit development

(A) Auxin and GA are likely to be synthesized in the achene and then be transported to the receptacle to initiate the fleshy fruit enlargement. **(B)** The hypothesized genetic pathways that regulate strawberry fruit growth. The figure was adapted from Zhou et al., 2021.

Since auxin and GA pathway genes greatly influence the fruit development in various species including Rosaceae fruits, we are eager to understand whether specific tissues have different responses to the hormone signals in different species, which may result in diverse fruit types. Therefore, we utilized the transcriptome data to compare how the hormone pathway genes work in distinct tissues across the four Rosaceae early fruits.

1.4 Type II MADS-box transcription factors

MADS-box family genes are divided into two types, type I (SRF-like) and type II (MEF2-like) MADS-box genes, based on the dissimilarity of the MADS-box sequence (**Figure 1.14**)(Gramzow & Theissen, 2010). Most of the ABCDE class genes belong to the type II MADS-box transcription factor family, which, in addition to possessing the MADS domain at the N-terminus, contains a relatively conserved keratin-like (K) domain flanked by two less conserved domains, intervening (I) and C-terminal (C) domains (Gramzow & Theissen, 2010; Ng & Yanofsky, 2001). The I and K domains contribute to protein-protein interaction (PPI), and the C domain is involved in transactivation and PPI stabilization (Gramzow & Theissen, 2010; Ng & Yanofsky, 2001). Therefore, the type II MADS-box genes are also known as

MIKC-type genes. While they were originally discovered to play important roles in specifying flower organ identity, these MADS box genes appear to play important role in specifying fruit identity.

Figure 1.14 Diagrams of Type I and type II MADS-box proteins

The figure was adapted from Gramzow & Theissen, 2010. Orange: DNA-binding, blue: protein-protein interaction, purple: transactivation.

Until now, a number of studies have demonstrated that type II MADS-box genes are involved in Rosaceae fruit development. An apple tree with a loss-of-function gene in *MdPI* (B class) was found to give rise to parthenocarpic fruit (**Figure 1.15A**) (Yao et al., 2001), indicating that MdPI acts as a repressor of fruit set. Recently, overexpressing *MdPI* in apple produced flat fruit with smaller size, which further suggests PI's negative impact on fruit formation (**Figure 1.15B**) (Yao et al., 2018). Besides *PI*, *SEP* genes (E class) were also shown to regulate fruit development in Rosaceae species. Suppression of *MdMADS8/9* (*SEP1/2-like* genes) reduced the size of apple fruit flesh (**Figure 1.16A**) (Ireland et al., 2013). In the same transgenic apple, the expression of a related E class gene *MdMADS6/7 (FBP9)* also decreased, but only after pollination (Ireland et al., 2013). Therefore, SEP1/2 and FBP9 are likely to be positive regulators of fruit growth. In contrast, SEP3 was found to be a potential repressor of fruit development in woodland strawberry. An EMS-induced point mutation in *FveSEP3* caused one amino acid change (G27E) in its MADS-box domain and produced parthenocarpic fruits that developed slowly (**Figure 1.16B**) (Pi et al., 2021). And the *fvesep3* knockout mutants generated by CRISPR/Cas9 exhibited similar phenotypes (Pi et al., 2021). In my dissertation work, I sought for the type II MADS-box genes that appeared to be a conserved regulator of fruit development by comparing their expression among the four

Rosaceae species. I would like to see whether the type II MADS-box genes show specific expression patterns that are positively or negatively correlated with fruit formation in order to identify their function during fruit development.

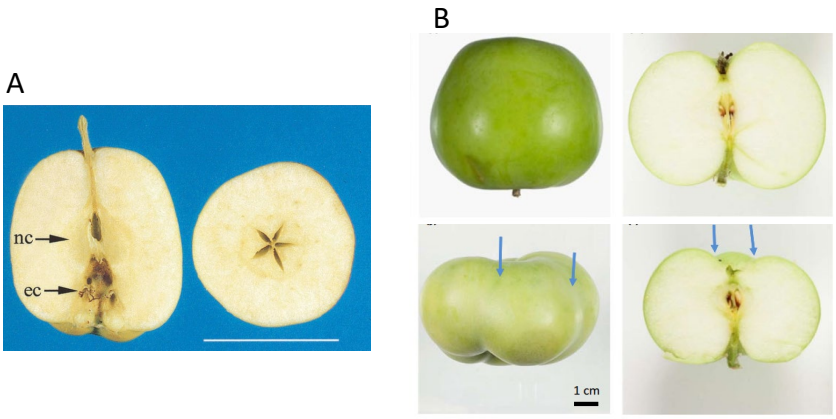

Figure 1.15 Alteration of PI function or expression influences apple fruit

(A) A parthenocarpic apple fruit developed from a loss-of-function *MdPI* mutant. A lack of seeds supports that the fruit is parthenocarpic. Apple mutant: 'Rae Ime. nc: normal carpels, ec: ectopic carpels. Scale bar: 5cm. **(B)** Upper panel: wide-type apple at 132 DAP (days after pollination), lower panel: an apple with overexpressed *MdPI* at 132 DAP. The fruit is flattened. Apple cultivar: 'Bolero'. The blue arrows point to the longitudinal grooves. (A) and (B) are respectively adapted from Yao et al., 2001 and Yao et al., 2018.

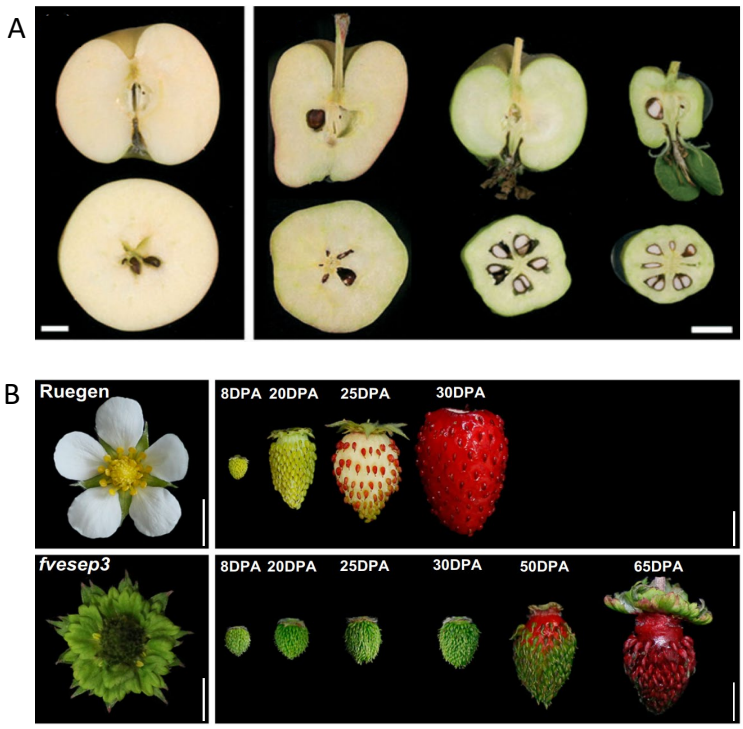

Figure 1.16 Class E genes *SEPALLATA* (*SEP*) are important regulators of Rosaceae fruit development

(A) Left panel: wide-type apple; right panel: apple from three *MdMADS8 (SEP1/2-like)* antisense transgenic lines. Apple cultivar: 'Royal Gala'. Scale bar: 1cm. **(B)** Upper panel: wide-type woodland strawberry, lower panel: a loss of function woodland strawberry mutant in *FveSPE3*. Fruit still enlarges despite a back of seed development. Woodland strawberry cultivar: Ruegen. (A) and (B) are respectively adapted from Ireland et al., 2013 and Pi et al., 2021.

1.5 Rosaceae genomic resources

My comparative transcriptomic analyses of the four Rosaceae species will not be possible in the absence of their high-quality reference genomes. Below summarizes the improvements of the genome sequencing resources over time for the four species.

1.5.1 New technologies for improving genome assembly and annotation

The rapid development of sequencing and related technologies, such as PacBio single molecule real-time (SMRT) sequencing, PacBio HiFi sequencing, Oxford Nanopore sequencing, Hi-C, and BioNano optical mapping over the past 10 years have greatly facilitated genome assembly and annotation, and revolutionized plant research. The increasing number of sequenced plant genomes and higher quality genomes make molecular research, genome editing, and marker-assisted breeding possible in species previously recalcitrant to molecular genetic research. PacBio and Nanopore both belong to the third-generation (single-molecule and real-time) sequencing technology. Their long-read DNA-seq helps overcome challenges of genome assembly caused by repetitive regions (Jiao & Schneeberger, 2017; Lu et al., 2016; Rhoads & Au, 2015) and facilitates splicing isoform prediction and genome annotation (Rhoads & Au, 2015). Different from traditional PacBio sequencing, HiFi sequencing produces highly accurate long reads by repeatedly sequencing the circularized templates and generating consensus reads (Hon et al., 2020). Hi-C and BioNano optical mapping are two scaffolding technologies that help to construct

chromosome-level scaffolds from contigs by providing long-range genomic information (Jiao & Schneeberger, 2017; Korbel & Lee, 2013; Tang, Lyons, et al., 2015). Many important crop species' genomes have benefitted from several rounds of genome assembly and annotation whenever a new technology was applied (**Table A.1**, see Appendices). Here, I will summarize the updates in genome assemblies and annotations of major Rosaceae species.

1.5.1.1 Several updates in peach genome assembly and annotation

Sanger sequencing reads were used to construct the first peach (*Prunus persica* 'Lovell') genome (Verde et al., 2013). And the genome was further annotated based on gene models predicted by FGENESH+ (Salamov & Solovyev, 2000) and GenomeScan (Yeh et al., 2001), transcripts assembled from ESTs by PASA (Haas et al., 2003), and protein homology. In 2017, an improved peach genome was reported (Verde et al., 2017). High-density linkage maps and Illumina resequencing data led to a chromosome-level high-quality peach genome assembly. Moreover, RNA-Seq data collected from various tissues were utilized to update the genome annotation. And about 20,000 isoforms were newly identified in the upgraded annotation.

Following the publication of the peach genome and subsequent improvement (Verde et al., 2013, 2017), the pan-genome of peach (*P. persica*) is a much-welcomed next step (Cao et al., 2020). A pan-genome consists of the entire set of genes and genetic variations within a species, and the portion of the pan-genome common to all cultivars in the species forms the core genome. A pan-genome identifies genetic variations among cultivars, provides valuable resources, and supports evolutionary studies. In this study, 100 *P. persica* accessions were sequenced, giving rise to 27,796 genes in the pangenome. Furthermore, the genomes of four wild peaches (*Prunus mira*, *Prunus kansuensis*, *Prunus davidiana*, *Prunus ferganensis*) were

assembled *de novo*, and the core genome shared by peach and its four wild relatives consists of 15,216 gene families. The analysis reveals dramatic variation in gene content between congeneric species and suggests that *P. mira* is the primitive ancestor of the cultivated peach.

1.5.1.2 Several updates in apple genome assembly and annotation

The progressive improvement of apple genome assemblies nicely illustrates the application of newer technologies. The first genome of apple (*Malus domestica* 'Golden Delicious') was published in 2010 using traditional Sanger sequencing and 454 next-generation sequencing (Velasco et al., 2010). Six years later, an improved apple genome of 'Golden Delicious' was assembled based on a combination of Illumina short reads and PacBio long reads (X. Li et al., 2016). Accordingly, the contig N50 of the apple genome was 111,619 bp, almost seven times the previous N50 (16,171 bp). In 2017, another *de novo* genome assembly of double haploid 'Golden Delicious' (GDDH13) was published (Daccord et al., 2017). In addition to the Illumina and PacBio data, a BioNano optical mapping was used in scaffolding. As a result, the scaffold N50 was increased to 5,558 kb.

In 2019, Illumina, PacBio, BioNano, and Hi-C technologies were integrated to construct a high-quality genome assembly of 'Hanfu' (HFTH1) apple, a *Malus domestica* cultivar grown in northern China (L. Zhang et al., 2019). The scaffold N50 was increased to 6,988 kb. Compared with the HFTH1 genome, the three published 'Golden Delicious' genomes shared 11,502 deletions and 6,590 insertions with an average length of 508 bp and 519 bp, respectively (Daccord et al., 2017; X. Li et al., 2016; Velasco et al., 2010). The average density of shared SNPs with the 'Golden Delicious' genomes is 2.15/kb. The HFTH1 genome was utilized to completely fill 488 gaps in the GDDH13 genome; the average length of the filled gaps is 78,864 bp (L. Zhang et al., 2019).

In 2020, it was reported that the diploid and haploid genomes of *Malus domestica* 'Gala' and its two wide progenitors *Malus sieversii* and *Malus sylvestris* were assembled using Illumina paired-end and mate-pair reads, 10x Genomics linked reads, and PacBio HiFi reads (Sun et al., 2020). The scaffold N50s of the diploid and haploid assemblies were 3.3-4.3 Mb and 16.8-35.7 Mb, respectively. 2,392,029 SNPs, 413,350 indels, and 324 inversions were identified by comparing Gala with GDDH13 (Daccord et al., 2017; Sun et al., 2020). And 3,260,700 SNPs, 464,386 indels, and 459 inversions were detected between Gala and HFTH1 (Sun et al., 2020; L. Zhang et al., 2019). Because of the genetic variations between 'Gala', 'Hanfu' and 'Golden Delicious', it is beneficial to use the genome assembly of the cultivar most closely related to the cultivars under one's study as a reference.

Besides the diploid and haploid genome assemblies, Sun et al. (2020) constructed three pan-genomes for *Malus domestica* Gala, *Malus sieversii,* and *Malus sylvestris* using the resequencing data from 36 *M. domestica*, 36 *M. sieversii*, and 11 *M. sylvestris* accessions. 81.3-87.3% of the genes constituted the core genomes of the three species. Moreover, 69,411 orthologous groups were identified in the three pan-genomes. And the presence of the orthologous groups was examined in each accession. Only a few orthologous groups were shown to be unique to *M. domestic*a, indicating that domestication may have low influence on gene selection.

In addition to genome assembly, high-quality genome annotations are essential to enhance the utility of the genome. In the first 'Golden Delicious' genome published in 2010, the genome annotation was based on the gene prediction programs and ESTs from Genbank (Birney et al., 2004; Korf et al., 2001; Majoros et al., 2004; Sayers et al., 2020; Solovyev et al., 2006; Velasco et al., 2010). In 2014, an improved apple reference transcriptome was constructed

using RNA-Seq data generated from 'Golden Delicious' fruits at 14 time points during development (Bai et al., 2014). In 2016, the *de novo* "Golden Delicious" genome assembly was supplemented by annotations based on RNA-Seq data from three distinct tissues (leaves, flowers, and stems) as well as *ab initio* and protein homology-based predictions (X. Li et al., 2016). To annotate the latest 'Golden Delicious' GDDH13 genome, mRNA was extracted and sequenced from more tissues, including leaves, roots, fruits, apex, stems, and flowers (Daccord et al., 2017). The GDDH13 genome annotation has the lowest number of protein-coding genes at 42,140 (Daccord et al., 2017) compared with 53,922 (X. Li et al., 2016) and 63,141 (Velasco et al., 2010). However, GDDH13 possesses the highest BUSCO completeness at 94.9% (Daccord et al., 2017) compared with 51.5% (X. Li et al., 2016) and 86.7% (Velasco et al., 2010).

1.5.1.3 Several updates in strawberry genome assembly and annotation

As with apples, the diploid woodland strawberry (*Fragaria vesca* ssp. *vesca* 'Hawaii4') genome assembly and annotation went through several rounds of updates. The first woodland strawberry genome became available at the end of 2010, and its genome annotation (v1.0) was generated by GeneMark-ES+ (Lomsadze et al., 2005), which integrated *ab initio* gene prediction and EST evidence (Shulaev et al., 2011). An updated assembly (v1.1) with improved contig arrangement was constructed in 2011. And the gene models were mapped from the previous genome (v1.0) to the new assembly (v1.1). In 2015, a new annotation (v1.1.a2) was created that combined different evidence, such as *de novo* and genome-guided transcriptome assembly from RNA-Seq reads, *ab initio* gene models, and plant protein sequences from UniProt (Darwish et al., 2015). More than 2000 new genes were added in the v1.1.a2 annotation. In 2014, dense linkage maps were leveraged to construct an improved woodland strawberry genome assembly (v2.0.a1) (Tennessen et al., 2014). In 2017, based on

PacBio long reads and Illumina short reads from *F. vesca* fruit receptacles as well as prior short-read RNA-Seq data, a new annotation (v2.0.a2) was generated (Y. Li et al., 2018). Although the total number of protein-coding genes decreased slightly, 13,168 protein-coding genes were updated in their gene structures, alternatively spliced (AS) isoforms were identified for 7,370 genes, and the BUSCO completeness score was increased to 95.7% from the prior version (88.9%). At the end of 2017, a high-quality woodland strawberry genome (v4.0.a1) was assembled using PacBio long reads, Illumina short reads, and BioNano optical mapping (Edger et al., 2018). This version uses a different gene-naming system, moving from the geneXXXXX to FvH4XgXXXXX format. Y. Li et al. (2019) included a supplementary table in their publication that correlates the *F. vesca* gene names between the old and new naming systems. In addition, a new annotation (v4.0.a2) was created based on comprehensive short- and long-read RNA-Seq data (Y. Li et al., 2019), adding 5,419 new protein-coding genes, improving the BUSCO completeness score to 98.1% from the prior 91.1%, and adding AS isoforms detected for about 30% of the genes. In 2021, the chromosome-level genomes of the five diploid *Fragaria* species (*F. mandschurica*, *F. viridis*, *F. daltoniana*, *F. nilgerrensis*, and *F. pentaphylla*) were assembled using a combination of second- and third-generation technologies (Qiao et al., 2021). Additionally, a pan-genome was constructed for strawberry species including *F. iinumae*, *F. mandschurica*, *F. vesca*, *F. viridis*, *F. daltoniana*, and *F. nilgerrensis*. And the percentage of the core genes in the pan-genome is around 42.9%.

In 2013, the first draft octoploid garden strawberry genome (*Fragaria x ananassa* 'Reikou') was reported (Hirakawa et al., 2014). Homoeologous sequences of the allo-octoploid strawberry are integrated into a haploid genome named FANhybrid_r1.2 with an N50 of 5.14 kb. Gene prediction was done *ab initio* using Augustus (Stanke et al., 2006). In the same study, the genomes of several Fragaria species were sequenced and assembled, including *F.*

orientalis, *F. iinumae*, *F. nipponica*, and *F. bucharica*, and *F. bucharica* (USDA accession CFRA522) was originally misidentified as *Fragaria nubicola* (Tennessen et al., 2014). In 2019, a near-complete chromosome-scale assembly of the *Fragaria x ananassa* 'Camarosa' was constructed with a contig N50 of about 79.97 kb, taking advantage of Illumina, 10X Genomics, and PacBio long reads (Edger et al., 2019). This chromosome-scaled genome consists of A, B, C, and D subgenomes, and the genome annotation (v1.0.a1) utilized RNA-Seq data from diverse tissue types (108,087 protein-coding genes) (Edger et al., 2019). In the same year, a garden strawberry reference transcriptome was constructed using PacBio sequencing (Yuan et al., 2019). The PacBio data in this study, together with other publicly available Illumina RNA-Seq data were recently utilized to improve the annotation of the *Fragaria x ananassa* 'Camarosa' genome (v1.0.a1) (T. Liu et al., 2021). Compared with *Fragaria x ananassa* v1.0.a1, the new annotation v1.0.a2 had a slight increase in the number of protein-coding genes (108,447). Importantly, the new annotation (v1.0.a2) for *Fragaria x ananassa* 'Camarosa' includes AS isoforms for 11,044 genes and adds 5′ and 3′ UTR information to a large proportion of the protein-coding genes (v1.0.a1: 38.93%, v1.0.a2: 73.61%).

The complete genome sequencing of *Fragaria x ananassa* 'Camarosa' allowed the identification of diploid progenitors, which has long been a mystery and recently a topic of intense debate. Based on the tree-searching algorithm (PhyDS), Edger et al. (2019) proposed four diploid species (*F. vesca*, *F. iinumae*, *F. viridis*, and *F. nipponica*) as the four progenitors of the octoploid and suggest the hexaploidy *F. moschata* as an intermediate species (Edger et al., 2019). However, Liston et al. (2020) reanalyzed the four subgenomes in a phylogenomic context and found support for *F. vesca* and *F. iinumae* but disputed *F. viridis*, *F. nipponica*, and *F. moschata* as progenitors (Liston et al., 2020). In response, a new

chromosome-scale genome of *F. iinumae* was subsequently assembled, and a reanalysis using PhyDS supports their original proposal regarding the four diploid species as the progenitors (Edger et al., 2020). A third group recently sequenced and assembled the genomes of three wild diploid species, *F. nilgerrensis*, *F. nubicola*, and *F. viridis* (Feng et al., 2021). Combining these three genomes with the previously sequenced *F. vesca* and *F. iinumae* genomes, the group utilized sppIDer (Langdon et al., 2018) to map short-read sequencing data of *F. x ananassa* to a composite reference genome, and the result supports that *F. vesca* and *F. iinumae*, but not others, are the progenitor species of the cultivated garden strawberry (Feng et al., 2021).

1.5.1.4 Several updates in raspberry genome assembly and annotation

The first genome assembly of black raspberry (Rubus occidentalis 'ORUS 4115-3') was published in 2016 (VanBuren et al., 2016). Illumina paired-end reads were generated to build a draft genome assembly, and then a high-density linkage map was constructed to anchor the scaffolds onto the seven pseudo-chromosomes. The reference-guided and de novo transcriptome assemblies were produced using the RNA-Seq data collected from a variety of tissues including leaf, root, fruit and cane. The assembled transcripts were further imported to MAKER (Cantarel et al., 2008) to annotate the genome. Moreover, the amino acids translated from the assembled transcripts, as well as the protein sequences from different species, served as the protein homology evidence for genome annotation. And SNAPhmm (Korf, 2004), Augustus (Stanke et al., 2006), and GeneMarkHMM (Lomsadze et al., 2005) were applied to predict gene models within the MAKER pipeline. The genome assembly was later improved by Hi-C in February 2018 (Jibran et al., 2018). About 97.2% of the sequence was able to be anchored onto the chromosomes. It was reported in August 2018 that PacBio reads were yielded to generate a near-complete genome assembly of 'ORUS 4115-3' (VanBuren et al.,

2018). Compared with the first black raspberry genome (v1.0.a1), the scaffold N50 of the new genome (v3.0) was increased from 0.35 Mb to 41.1 Mb. And the RNA-Seq data obtained from black raspberry fruit, flower, cane, root, leave and methyl jasmonate-treated leave were used for genome annotation by MAKER-P (Campbell, Law, et al., 2014). Consequently, a higher number of protein-coding genes were predicted in the new genome (v1.0.a1: 28,005, v3: 34,545).

In 2019, PacBio and Illumina reads were utilized to assemble the first draft genome of red raspberry (*Rubus Idaeus* 'Joan J') with scaffold N50 of 0.64 Mb (Wight et al., 2019). And the genome was further annotated using the RNA-Seq data collected from three different fruit tissues (receptacle, ovary wall, and ovule/seed) at two developmental stages (0 and 12 DPA), producing 35,566 predicted protein-coding genes. In 2021, a chromosome-level genome of *Rubus chingii* was constructed using Nanopore, Illumina and Hi-C technologies (L. Wang et al., 2021). And RNA extracted from leaf, stem, flower, root, and fruit tissues was sequenced for genome annotation. The resulting high-quality genome had scaffold N50 of 8.2 Mb, and 33,130 protein-coding genes were identified in the genome. In 2022, a new genome of *Rubus Idaeus* ('Anitra') was reported (Davik et al., 2022). PacBio, Illumina, and Hi-C sequencing were combined to generate a chromosome-scale assembly. The scaffold N50 of the genome was 34.5 Mb, and 39,448 protein-coding genes were predicted in the genome by Augustus (Hoff & Stanke, 2019). In this dissertation work, I was involved in the genome assembly and annotation of the genome of red raspberry *Rubus idaeus* version 1.0 and version 2.0. Chapter 3 describes the effort in the version 2.0 genome. My dissertation work has contributed to the genomic resources of the economically important red raspberry.

1.5.2 Computational databases for Rosaceae species

The establishment of various online databases provides easy access and interaction with the genomic data. These databases help organize genomic resources, facilitate data sharing, and enable genome comparison across different species. Below, I will highlight some useful databases that are of particular importance to Rosaceae research (**Table A.2**, see Appendices). In my dissertation Chapter 4, I describe my contribution to this area in establishing a new comparative genomic database for Rosaceae.

1.5.2.1 Rosaceae genome databases

Genome Database for Rosaceae (GDR) (https://www.rosaceae.org/) (S. Jung et al., 2019) is, by far, the best resource hub for Rosaceae research. It hosts the most comprehensive and up-to-date collection of genome assembly and annotation versions for widely studied genera, *Fragaria*, *Malus*, *Prunus*, *Potentilla*, *Pyrus*, *Rosa*, and *Rubus*. For instance, GDR hosts *Fragaria vesca* genome assemblies of v1.0, v1.1 (an improved pseudochromosome assembly of v1.0), v2.0.a1, and v4.0.a1. Moreover, it incorporates corresponding updated annotations v1.1.a2, v2.0.a2, and v4.0.a2. In addition, GDR serves as the database of record for Rosaceae gene names; standardized gene-naming guideline should be followed to ensure uniformity and clarity (S. Jung et al., 2015). Besides the genes and genomes, GDR provides genetic maps, markers, germplasm, and trait information as well as an impressive set of tools. For example, the search tools of GDR enable users to search for specific gene sequence, maps, and markers; its MegaSearch tool allows downloading different data types in bulk. With the GDRCyc tool, users can search, visualize, and overlay pathway data. With the Synteny Viewer tool, one can select specific Rosaceae species for comparison, visualize syntenic blocks, and obtain information on syntenic genes.

The NCBI Genome (https://www.ncbi.nlm.nih.gov/genome) (Tatusova et al., 1999) on the other hand collects genomes from a broader range of Rosaceae species, including lesser-known species, such as Drummond's Mountain avens (*Dryas drummondii*), wood avens (*Geum urbanum*), and bitterbrush (*Purshia tridentate*) (Griesmann et al., 2018; Jordan et al., 2018).

1.5.2.2 *Rosaceae* species-specific databases

Many genome sequencing or annotation papers of Rosaceae species are accompanied by species-specific websites that provide tools, including BLAST searches for genes of interest. For instance, the genomes of Yoshino cherry (*Cerasus* x *yedoensis*) and sweet cherry (*Prunus avium*) are both deposited in DBcherry (http://cherry.kazusa.or.jp/) (Shirasawa et al., 2017, 2019). The built-in BLAST enables users to search their sequences of interest against the cherry genomes, and JBrowse is embedded in the database for visualizing the genomic regions. The genomes of garden strawberry (*Fragaria x ananassa*) and multiflora rose (*Rosa multiflora*) are available in Strawberry GARDEN (http://strawberry-garden.kazusa.or.jp/) and Rosa multiflora DB (http://rosa.kazusa.or.jp/), respectively (Hirakawa et al., 2014; Nakamura et al., 2018). These two websites as well as the database for cherry are all supported by the Kazusa DNA Research Institute.

Several Rosaceae species have developed species-specific databases with multiple analysis tools and resources. Strawberry Genomic Resources (SGR, http://bioinformatics.towson.edu/strawberry/default.aspx) is a website that integrates different types of woodland strawberry (*Fragaria vesca*) genomic data (Darwish et al., 2013). It allows users to access the transcriptome analysis of the woodland strawberry early fruit development (Kang et al., 2013). Users can acquire differentially expressed genes between distinct tissues

and stages by searching the database and use the eFP browser to visualize RNA-Seq data across tissues and stages for genes of interests (Hawkins et al., 2017). An updated *F. vesca* eFP browser is hosted at the BAR (http://bar.utoronto.ca/efp_strawberry/cgi-bin/efpWeb.cgi).

A reference transcriptome of Chinese pear (*Pyrus pyrifolia*) was constructed by utilizing PacBio, 454, and Sanger sequencing, and it is stored in the database TRANSNAP (http://plantomics.mind.meiji.ac.jp/nashi/) (Koshimizu et al., 2019). The database also includes gene functional annotation performed by BLASTP (Altschul et al., 1990), KAAS (Moriya et al., 2007), and InterProScan (Jones et al., 2014). Users can examine gene-expression patterns generated from GEO (https://www.ncbi.nlm.nih.gov/geo/) microarray data.

The *Fragaria vesca* gene co-expression network explorer (http://159.203.72.198:3838/fvesca/) was developed to host the non-consensus and consensus co-expression networks generated using RNA-Seq data from flower and fruit tissues of the woodland strawberry (Shahan et al., 2018). Users are able to search for genes of interest and the transcriptional co-expression clusters to which they belong, obtain network statistics, visualize cluster eigengene expression, examine enriched GO terms in the cluster of interest, and download the cluster graphml structure.

1.6.2.3 Useful plant databases for comparative genomics, metabolic networks, and others

Although the summary above focuses on Rosaceae databases, many plant databases are also highly useful for Rosaceae research, for example, Plant Transcription Factor Database (http://planttfdb.gao-lab.org/), Plant Transcriptional Regulatory Map (http://plantregmap.gao-

lab.org/), and CANTATAdb (http://cantata.amu.edu.pl/) for plant lncRNAs. Below, we highlight four such databases.

PLAZA (https://bioinformatics.psb.ugent.be/plaza/) (Van Bel et al., 2022) and Phytozome (https://phytozome-next.jgi.doe.gov/) (Goodstein et al., 2012) are databases for plant genome comparisons. Currently, Dicots PLAZA 5.0 has integrated genomic resources from 100 species, including five Rosaceae species, garden strawberry (*Fragaria* x *ananassa*), woodland strawberry (*Fragaria vesca*), China rose (*Rosa chinensis*), apple (*Malus domestica*), and peach (*Prunus persica*). Phytozome v13 has gathered 261 annotated genomes, including three Rosaceae species, woodland strawberry, garden strawberry, apple, and peach.

Plant Metabolic Network (PMN, https://plantcyc.org/) (Schläpfer et al., 2017) and Plant Reactome (https://plantreactome.gramene.org/index.php?lang=en) (Naithani et al., 2020) are both databases for plant pathways. Plant Metabolic Network is focused on metabolic pathways and hosts the database PlantCyc that contains shared pathways among more than 500 plant species. Additionally, a single species database was also constructed in PMN, which allows users to access pathways and enzymes for individual species. PpersicaCyc, SweetcherryCyc, MdomesticaCyc, EuropeanpearCyc, Fvesca_VescaCyc, RmultifloraCyc, and RchinensisCyc are developed for Rosaceae family members. Besides the metabolic pathways, Plant Reactome hosts different types of pathways, including gene regulatory pathways, hormone signaling pathways, and others. Users can view and interact with the pathways in the browser and identify chemical compounds and proteins involved in the processes. The database encompasses multiple Rosaceae species, such as woodland strawberry, apple, sweet cherry, and peach. Furthermore, the database enables researchers to

perform pathway enrichment analysis and species comparison between pathways of rice and those of selected species.

1.6 Summary of the dissertation

In this dissertation, we focus our study on the four Rosaceae species (peach, apple, strawberry, and raspberry) that represent four main fruit types (drupe, pome, achenetum, and drupetum). Although the basic flower structure of the four species is similar, their fruit flesh can be derived from different flower tissues. We investigated the questions of what mechanisms make the flowers of the four Rosaceae species grow into distinct fruit types and which genetic factors determine whether a tissue is able to eventually become fruit flesh or not.

Since the fleshy fruit identity is usually determined early in the fruit development, to answer the above questions, we collected and dissected the fruits of three Rosaceae species (peach, apple, and raspberry) starting at the pre-pollination stage and then several stages after pollination. The fourth species, woodland strawberry, was already dissected and characterized previously (Kang et al., 2013). And we examined the morphologic changes in distinct fruit tissues over time to better understand the early fruit development in the three Rosaceae species, peach, apple, and raspberry. We then sequenced and analyzed RNA transcripts extracted from the dissected tissues of the three Rosaceae species at early fruit stages, taking advantages of high-throughput RNA sequencing technologies and abundant publicly available genomic resources. I conducted pairwise transcriptomic comparisons between peach and apple (Chapter 2) and between raspberry and strawberry (Chapter 3) because peach and apple belong to the same subfamily of Rosaceae (Amygdaloideae) and raspberry and strawberry are both in the Rosoideae subfamily.

A large number of open-source computational tools were utilized in my comparative analyses including principal component analysis, grade of membership analysis, differential gene expression analysis, consensus co-expression analysis, tissue-specific gene analysis, and others. Additionally, type II MADS-box genes and auxin and GA pathway genes were analyzed in more depth due to their known function in fruit development. I compared their expression trends across Rosaceae species to uncover their potentially conserved function in fruit identify specification and their potential ability in promoting "fleshiness" of the fruit.

Lastly, while there are many species-specific databases for Rosaceae family, databases for comparative analysis are lacking. In Chapter 4, I constructed the ROsaceae Fruit Transcriptome database (ROFT), a comparative analysis database for the four Rosaceae fruits. The searchable function allows one to search and obtain information of orthologs in all four species. Further, an eFP browser in the database allows one to simultaneously and visually compare the expression of genes of interests across the four species, providing a foundation for hypothesis generation and gene function prediction.

My dissertation work benefited from collaborations with several fruit researchers. For Chapter 2, Drs. Kelsey Galimba, Chris Dardick, and Ann Callahan at the Appalachian Fruit Research Station of USDA collected the RNAs from peach and apple early fruits. Yuwei Xiao, a graduate student in the Liu lab helped retrieve the developmental programmed cell death genes in Arabidopsis and examine the presence of MADS box genes FBP9 and TM6 in eudicot species. For Chapter 3, the wet-lab experiments including raspberry fruit dissection, tissue sectioning, and hormone treatment were conducted by Dr. Junhui Zhou in the Liu lab. The initial genome assembly and annotation were performed by Dr. Yongping Li at the Chinese University of Hong Kong, China. For Chapter 4, an undergraduate student, Andrew

Tong, helped with the implementation of hyperlinks using R shiny for Rosaceae fruit transcriptome database (ROFT).

Chapter 2: Comparative analysis of peach and apple transcriptomes from early developing fruits

2.1 Introduction

Rosaceae is a plant family that includes many economically important fruit-bearing species. Peach (*Prunus persica*) and apple (*Malus x domestica*) are two species from the same Rosaceae subfamily (Amygdaloideae) (Xiang et al., 2017), and they both have publicly available high-quality genomes (Daccord et al., 2017; Verde et al., 2017). Peach has a relatively small genome with 8 haploid chromosomes while apple underwent a recent whole genome duplication (WGD) resulting in 17 haploid chromosomes (Velasco et al., 2010). Peach and apple share a basic flower structure, but their fruit flesh is developed from different floral organ after fertilization. In peach, the middle layer of the ovary wall (mesocarp) develops into the fruit flesh, and the hypanthium senesces (**Figure 2.1A**), but in apple, the hypanthium becomes fleshy (**Figure 2.1B**) (Z. Liu et al., 2020).

In most plants, fruit growth is usually dependent on successful pollination and fertilization. However, the application of auxin and/or gibberellic acid to unpollinated flowers can substitute for fertilization and produce parthenocarpic (seedless) fruits (Cong et al., 2019; Crane et al., 1960; Galimba et al., 2019; Kang et al., 2013). It suggests that auxin and GA production may be induced by the action of pollination and/or fertilization, and they in turn promote fruit development. Auxin and GA synthesis and signaling pathways have been well characterized in existing research (Chapter 1), and the genes involved in auxin and GA pathways, such as *FveRGA1* and *FveARF8*, were found to regulate fruit growth in Rosaceae species (Zhou et al., 2021).

In addition to phytohormone auxin and GA, Type II MADS box genes, such as *PI* and *SEP*, have also been shown to play a role in fruit set and fruit growth (Chapter 1). Type II MADS-box genes were originally discovered due to their roles in determining floral organ identity and were classified as ABCDE classes (Bowman et al., 1991, 2012; Colombo et al., 1995; Ditta et al., 2004; Egea-Cortines et al., 1999; Favaro et al., 2003; Honma & Goto, 2001; Pelaz et al., 2000; Theißen, 2001). Later, it was revealed that the type II MADS-box genes not only control flower development, but also influence the fruit set, fruit growth, and fruit fleshiness. In apple, the B class gene, *MdPI*, acted as a negative regulator of fruit set and fruit growth (Yao et al., 2001, 2018) while the E class genes, *MdSEP1/2* and *MdFBP9*, potentially positively regulate fruit flesh formation (Ireland et al., 2013).

In this study, RNA-Seq data were collected from hypanthium, ovary wall, and ovule/seed at four early stages of peach and apple fruit development (0, 5/6, 12, 18/20 days post anthesis). These peach and apple early fruit transcriptomes were compared to identify the genetic factors that may determine the fruit identity and explain the distinct fleshy fruit type between the drupe fruit type of peach and the pome fruit type of apple. Auxin and GA pathway genes and type II MADS-box genes were specifically examined to determine their possible roles in early fruit development and fruit type specification. This chapter is part of the published work (M. Li et al., 2022).

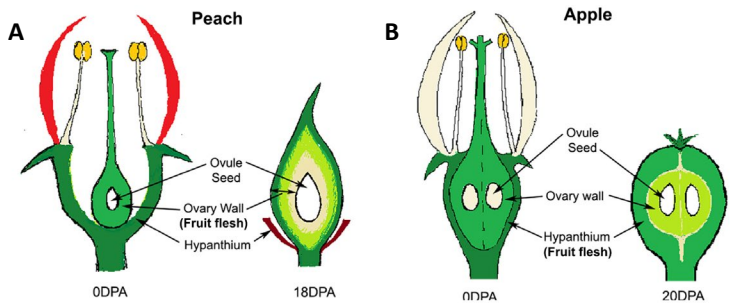

Figure 2.1 Peach (A) and apple (B) flower and fruit structures

The diagrams were drawn by Dr. Zhongchi Liu.

2.2 Material and methods

Plant sample collection and RNA sequencing

Our collaborators from USDA (Dr. Kelsey Galimba, Dr. Chris Dardick, and Dr. Ann Callahan) collected, dissected and sequenced three specific fruit tissues (hypanthium, ovary wall, and ovule/seed) at four critical early stages of fruit development from peach and apple in triplicate. Peach fruits were collected at 0, 5, 12, and 18 DPA (days post anthesis), and apple fruits were collected at 0, 6, 12, and 20 DPA. RNA was extracted and sequenced using Illumina HiSeq 2000. The details of the experiments were described in M. Li et al., 2022.

Gene symbols and IDs of the important type II MADS-box genes

The gene IDs for B class genes are *AJ291490* (*MdPI*), *Md15G1250200* (*MdTM6b*), *Md02G1136500* (*MdTM6a*), *Md08G1021300* (*MdAP3*), *Pp1G489400* (*PpPI*), *Pp7G164100* (*PpTM6*) and *Pp1G371300* (*PpAP3*). Gene IDs for E class are *Md13G1121500* (*MdMADS18|MdSEP3a*), *Md16G1121800* (*MdMADS118|MdSEP3b*), *Md17G1065400* (*MdMADS1|MdMADS8|MdSEP1,2b*), *Md09G1073900* (*MdMADS9|MdSEP1,2a*), *Md14G1215600* (*MdMADS3|MdMADS7|MdFBP9c*), *Md06G1204300* (*MdMADS6|MdFBP9b*), *Md06G1204100* (*MdFBP9a*), *Md16G1058600* (*MdMADS104|MdSEP4b*), *Md13G1059300* (*MdMADS4|MdSEP4a*), *Pp1G223600* (*PpSEP3*), *Pp3G249400* (*PpMADS7|PpSEP1,2*), *Pp5G208400* (*PpMADS2|PpFBP9*) and *Pp1G290500* (*PpSEP4*).

Global and differential gene expression analysis

Salmon (v0.11.2) (Patro et al., 2017) was used to quantify transcripts in reference transcriptome (Daccord et al., 2017; Verde et al., 2017) downloaded from Genome Database for Rosaceae (GDR) (S. Jung et al., 2019). The transcript sequence of *MdPI* (*AJ291490*) (Yao

et al., 2001) was added to the apple transcriptome because the *M. domestica* GDDH13 genome assembly lacks this gene. The transcriptome indices were built by Salmon with k-mer size set to 31. –keepDuplicates command was passed to the indexer to keep the identical transcripts in the reference transcriptomes. The –seqBias flag was used for sequence-specific bias correction. The transcript abundance (TPM) for each gene was summarized by tximport (v1.10.1) (Soneson et al., 2016).

To identify peach and apple homologs of Arabidopsis genes, peach (*P. persica*) and apple (*M. domestica*) protein sequences were retrieved from GDR (Prunus_persica_v2.0.a1.allTrs.pep.fa and GDDH13_1-1_prot.fasta, respectively), and the longest protein isoforms were used to search against TAIR10 protein database (file name: TAIR10_pep_20101214) (Lamesch et al., 2012) by local BLASTP (v2.5.0) (Altschul et al., 1990). Any peach or apple protein whose best blast hit to Arabidopsis protein had an *E*-value $<10^{-5}$ was selected.

To assign orthologs, Blast, OrthoFinder (v2.3.8), HMMER/Phylogeny, and Synteny analyses were performed (Altschul et al., 1990; Emms & Kelly, 2015; Mistry et al., 2013; Tang et al., 2008); different methods sometimes led to different results. In most cases, we used both Blast and OrthoFinder to ascertain orthology. When orthology could not be confirmed with OrthoFinder, the Blast result was used but the gene name ends with 'L (Like)'. For type II MADS-box genes, we assigned orthology based on phylogenetic tree analysis (see below).

Arabidopsis auxin and GA pathway genes were obtained from RIKEN Plant Hormone Research Network (http://hormones.psc.riken.jp/pathway_hormones.html) and KEGG PATHWAY (https://www.genome.jp/dbget-bin/www_bget?ath04075), and then manually curated. Yuwei Xiao, a graduate student in the Liu lab retrieved the Arabidopsis dPCD genes

from tables S3 and S5 of a prior publication (Olvera-Carrillo et al., 2015). Arabidopsis TFs were downloaded from PlantTFDB (v4.0) (Jin et al., 2017).

Differential gene expression analysis was conducted by DESeq2 (v1.22.2) (Love et al., 2014). For each developmental stage, pairwise comparisons were performed among the three flower/fruit tissues. For each tissue, adjacent stages were compared. The two factors, tissue and stage, were combined into a single factor in the design formula. The genes with low read counts (≤ 36) were filtered out. The cutoffs $P_{adj} < 0.05$ and $|\log_2 FoldChange| > 1$ were applied to identify DE genes. Average $\log_2(TPM + 1)$ values were used to construct heatmaps, which were generated by R package pheatmap (v1.0.12) (https://rdrr.io/cran/pheatmap/). The built-in function plotPCA in DESeq2 (v1.22.2) was used for the PCA.

Identification of peach and apple type II MADS-box (MIKC) genes

Full Pfam (El-Gebali et al., 2019) alignment files (PF00319 and PF01486), respectively, of MADS-box and K-box protein domains were used to search the peach and apple proteins by HMMER (v3.3) (Mistry et al., 2013). The majority of the K-box domain containing proteins also contain a predicted MADS-box domain. Arabidopsis Type II MADS-box proteins were obtained from TAIR (https://www.arabidopsis.org/browse/genefamily/mads_tffamily.jsp). Full-length sequences of these peach, apple and Arabidopsis type II MADS-box proteins were aligned by MAFFT (v7.458) (Katoh & Standley, 2013). Based on the alignment, a type II MADS-box phylogenetic tree was constructed by RAXML (v8.2.12) (Stamatakis, 2014) using an automatically determined model and a combination of ML search and rapid bootstrapping (100 bootstrap replicates). The tree was visualized by FigTree (v1.4.3) and shown in **Figure 2.11**.

In the most recent *M. domestica* genome assembly, 'Golden Delicious' doubled-haploid apple (GDDH13) (Daccord et al., 2017), the apple *PI* gene (AJ291490) is absent even though it was previously isolated and characterized (Yao et al., 2001). Furthermore, sequence reads corresponding to the *PI* gene are present in the raw 'Golden Delicious' GDDH18 genome reads (accession SRX2648079) (Daccord et al., 2017). An earlier draft version of the 'Golden Delicious' genome also contains *MdPI* (*MDP0000286643*) as a single copy gene (Velasco et al., 2010). Therefore, *AJ291490* (*MdPI*) sequence was added to the most recent GDDH13 genome before performing the RNA-seq analysis.

To build a phylogenetic gene tree of B and E class MADS-box genes (**Figure 2.12**), tomato B and E class genes were extracted from a previous publication (Z. Zhang et al., 2018), and their protein sequences were obtained from Sol Genomics Network (https://solgenomics.net/) (Bombarely et al., 2011). The phylogenetic analyses of B and E class proteins from peach, apple, tomato and Arabidopsis were performed by PhyML (v3.0) (Guindon et al., 2010) using automatic model selection, 100 bootstrap replicates, and SPR as type of tree improvement. Protein sequence alignments were produced by MAFFT (v7.458).

Consensus co-expression network analysis and GO analysis

Consensus clustering approach (Monti et al., 2003) was applied to build robust signed co-expression networks with $\log_2$(TPM +1) as the gene expression measurement. Eighty percent of the genes were subsampled and clustered by WGCNA (v1.68) (Langfelder & Horvath, 2008) 1000 times using the biweight mid-correlation method and other randomized parameters (Shahan et al., 2018). The genes with little variance (≤ 0.05) and zero median absolute deviation were filtered out. A weighted adjacency matrix was computed by dividing the number of times genes were clustered by the number of times genes were subsampled together, which was then used to construct the consensus network by WGCNA (v1.68) with

power 6 and minModuleSize 100. The networks were visualized by Cytoscape (v3.7.1) (Shannon et al., 2003).

Peach and apple GOs were downloaded from GDR (Prunus_persica_v2.0.a1_gene_functions.txt and Malus_x_domestica_GDDH13_v1.1_interpro.txt, respectively). The GO categories enriched for each gene module in the networks were determined by the Fisher's exact test in R package TopGO (v2.34.0) (Alexa & Rahnenfuhrer, 2022). The *P*-value cut-off was set at 0.05.

The clusters were selected if one (or more) of the top five GOs are directly associated with the biological processes of interests with a corresponding P-value < 0.001. Apple P41 and P46 were selected due to involvement in cell division even though they did not meet the requirements. The Sankey plot showing the number of common GOs between the peach and apple clusters was generated with R package networkD3 (Allaire et al., 2017). The dot plot was created with R package ggplot2 (Wickham, 2016).

SOG (syntenic orthogroup) identification and analysis

MCScan (jcvi v0.8.12) (Tang et al., 2008) was used to conduct synteny search between peach and apple to identify SOGs. Using the differential gene expression analysis data described earlier, SOG IDs were transferred to corresponding genes, and the number of DE SOGs in each comparison was identified and presented in **Figure 2.5** using the UpSetR (v1.3.3) (Conway & Gehlenborg, 2019).

Investigation of *TM6* and *FBP9* in dry fruits

The built-in BLASTP tool in PLAZA 4.0 Dicots (Van Bel et al., 2018) was used to blast tomato FBP9 and TM6 protein sequences against the PLAZA protein database. Most of the blast hits for TM6 were from the gene families ORTHO04D003666 and ORTHO04D005076. All of the proteins in the two families were aligned by MAFFT (v7.458) (Katoh & Standley, 2013), and a phylogenetic tree was built by RAXML (v8.2.12) (Stamatakis, 2014). The same approach was applied for a phylogenetic analysis on FBP9 using ORTHO04D000581. Yuwei Xiao, a graduate student in the Liu laboratory, checked the presence of *FBP9* and *TM6* in the eudicot species in PLAZA Dicot 4.0 database by examining the phylogenetic tree and performing BLAST searches against EST or RefSeq databases in NCBI.

2.3 Results

2.3.1 Global view of early fruit RNA-Seq data

Peach and apple fruits were harvested before and after pollination and fertilization at four early developmental stages (peach: 0, 5, 12, 18 DPA, apple: 0, 6, 12, 20 DPA) (**Figure 2.2A, B**). Pollination occurs between 0 DPA and 5/6 DPA in peach and apple (M. Li et al., 2022). However, fertilization happens at different stages in peach and apple. In peach, fertilization occurs at 5-12 DPA while in apple, fertilization occurs at 12-20 DPA (M. Li et al., 2022). And three flower/fruit tissues (hypanthium, ovary wall, and ovule/seed) were hand-dissected at each stage (**Figure 2.2A, B**). Eventually, a total of 72 RNA-Seq libraries including three biological replicates per sample type were generated and sequenced. It was observed that peach ovary wall gradually grows larger from 0 DPA to 18 DPA, and its hypanthium deteriorates during fruit development (**Figure 2.2A**). Unlike peach, apple hypanthium and ovary wall develop coordinately at early stages (**Figure 2.2B**).

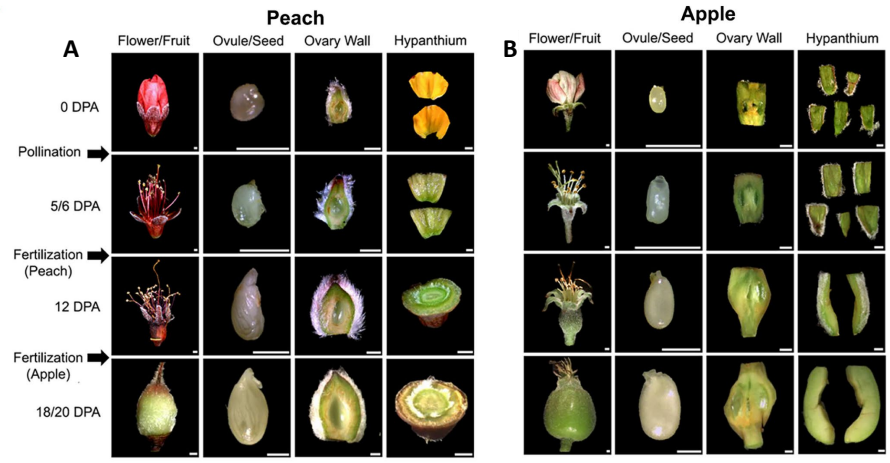

Figure 2.2 Dissection of peach (A) and apple (B) flower and fruit tissues

Dr. Kelsey Galimba dissected the flower and fruit tissues and took the photos. Scale bar: 1mm.

The principal component analysis (PCA) was performed based on the normalized and transformed gene expression data. It reveals that peach hypanthium is distinct from the other two tissues, which is consistent with the fact that peach hypanthium senesces during fruit growth and fails to be part of the fruit (**Figure 2.3A**). In contrast, apple hypanthium and ovary wall are more similar to each other supporting the observation that apple hypanthium and ovary wall both become enlarged in early development (**Figure 2.3B**). Additionally, dramatic gene expression changes occur between 0 DPA and 5 DPA in peach hypanthium (**Figure 2.3A**). And the huge gene expression differences are also found in apple hypanthium and ovary wall after pollination (0-6 DPA) (**Figure 2.3B**). It indicates that pollination may be an important trigger in both peach and apple fruit development.

Subsequently, the differentially expressed genes (DEGs) were identified between adjacent stages and distinct tissues. And the results agree with what have been found in PCA. In peach, relatively fewer DEGs were detected between ovary wall and ovule/seed while in apple, ovary wall and hypanthium show fewer differences (**Figure 2.4B, D**). Moreover, a

large number of DEGs were discovered between 0 and 5/6 DPA in peach hypanthium and apple hypanthium and ovary wall (**Figure 2.4A, C**).

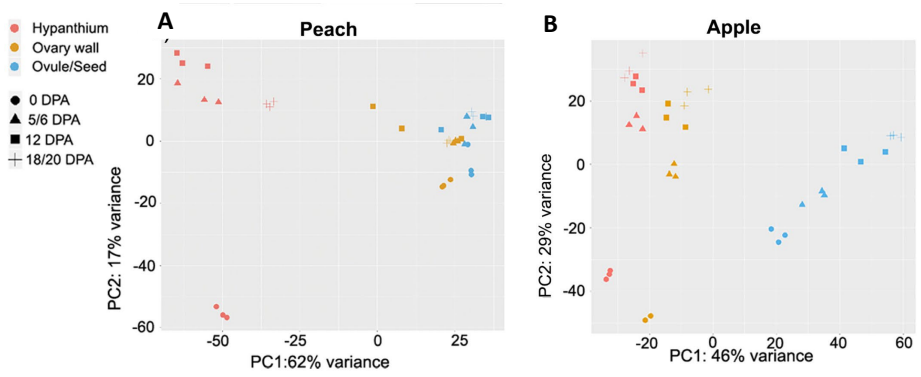

Figure 2.3 PCA plots of peach (A) and apple (B) samples

Different colors represent different tissues, and different shapes stand for different stages.

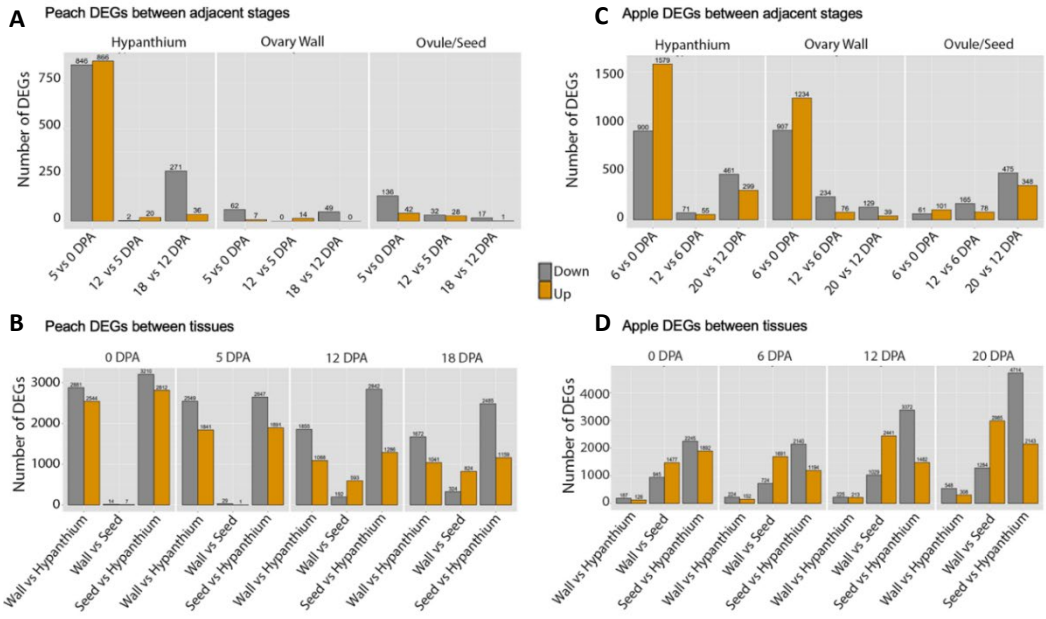

Figure 2.4 Differential expression analyses in peach and apple

(A, C) The number of differentially expressed genes (DEGs) between adjacent stages in each tissue in peach (A) and apple (C). **(B, D)** The number of DEGs between distinct tissues at each stage in peach (B) and apple (D).

2.3.2 Comparative analysis of differentially expressed syntenic orthogroups (SOGs) between peach and apple

The syntenic orthogroups (SOGs) between peach and apple were detected by MCScan (Tang et al., 2008). Firstly, the differentially expressed (DE) SOGs between adjacent stages in each

tissue were identified in peach and apple, separately. And then, for individual tissue, the DE SOGs were compared between peach and apple to uncover the overlapping and unique DE SOGs (**Figure 2.5** and **Figure 2.6A, D**). GO enrichment analysis was conducted to examine the biological processes unique to hypanthium soon after fertilization (**Figure 2.6**) as peach and apple hypanthium were shown to have more differentially expressed genes and syntenic orthogroups between pre- (0 DPA) and post-fertilization (5-6 DPA) (**Figure 2.4A, C and Figure 2.5**). It was found that the GO term 'response to hormone' is enriched among the SOGs that are uniquely down-regulated at 5 DPA, in comparison to 0DPA, in peach hypanthium, including ARF2, ARF8L, ARF9, IAA3, IAA4, IAA11, IAA14, and IAA19 (**Figure 2.6B**). The weakened response to auxin in peach hypanthium may explain its senescence after pollination. Additionally, a number of uniquely up-regulated SOGs at 6 DPA in apple hypanthium are associated with 'microtubule-based process' and 'cell cycle' (**Figure 2.6F**). It suggests that apple hypanthium has high cell division activities and is capable of growing larger. Therefore, the comparative analysis of DE SOGs enables us to better understand why peach and apple hypanthium have totally different fates in fruit development and apparently this starts soon after pollination.

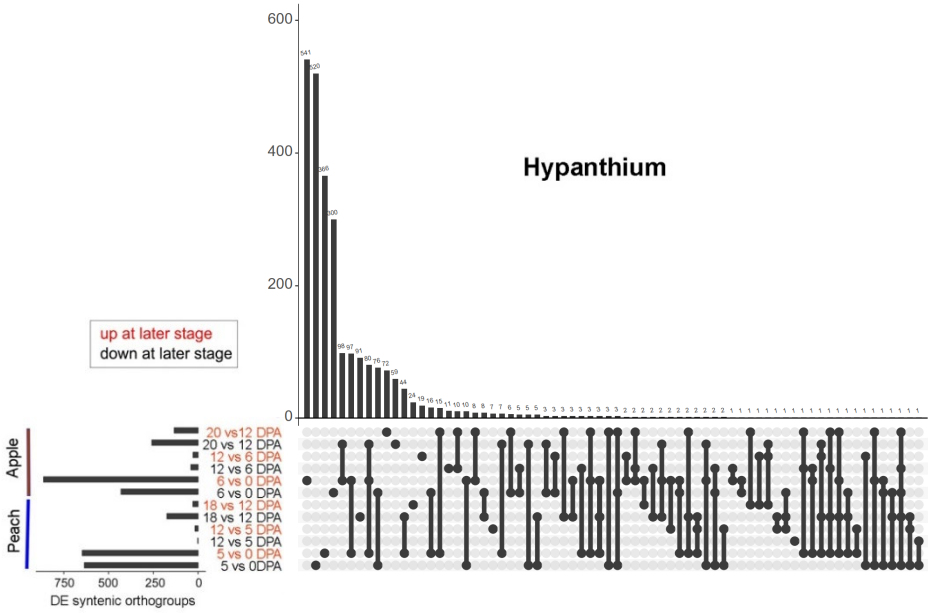

Figure 2.5 The UpSetR plot exhibiting the intersections of the differentially expressed (DE) syntenic orthogroups (SOGs) between stages in peach and apple hypanthium

The bar graph at the bottom left shows the number of DE SOGs between stages in peach and apple hypanthium. The connected dots at the bottom right indicate the intersection between the DE SOGs. And the bars at the top show the specific number of overlapping DE SOGs for each intersection. The red labels mean that the SOGs are up-regulated at the later stage while the black labels mean that the SOGs are down-regulated at the later stage.

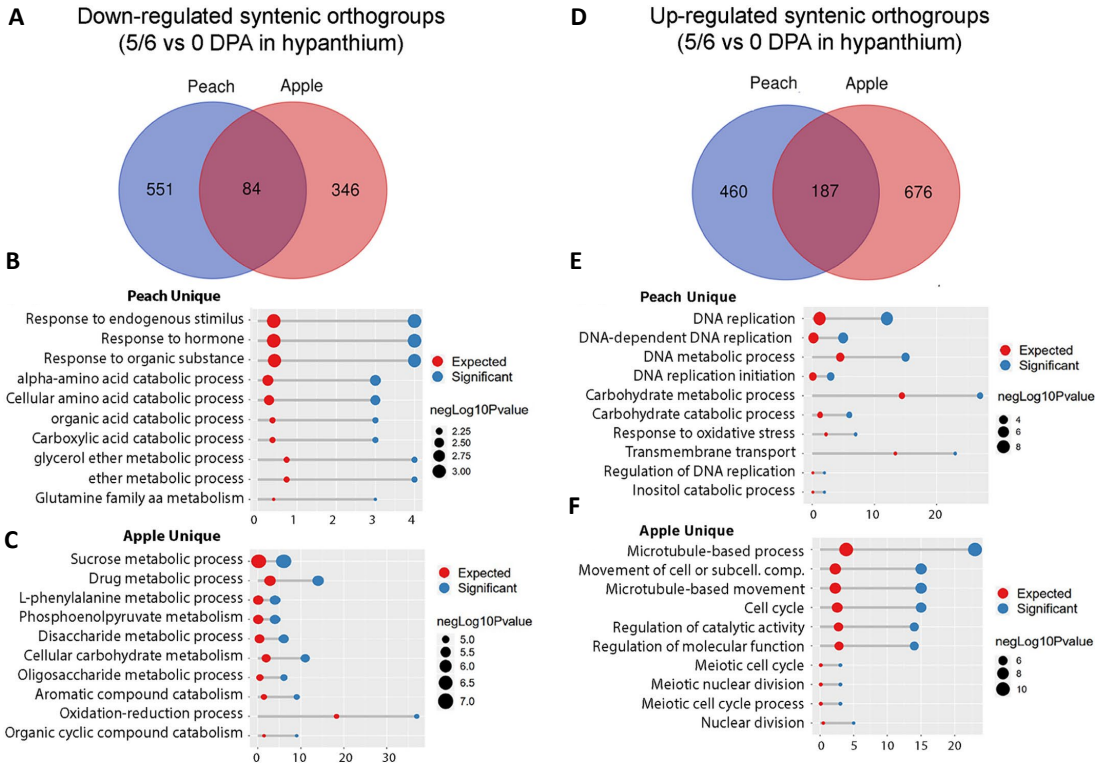

Figure 2.6 Comparing the differentially expressed (DE) syntenic orthogroups (SOGs) between 5/6 DPA and 0 DPA in peach and apple hypanthium

(A, D) The Venn diagram showing the number of SOGs that are uniquely down-regulated (A)/up-regulated (D) at 5/6 DPA in peach or apple hypanthium, and SOGs that are down-regulated (A)/up-regulated (D) in both. **(B, E)** GO enrichment of the SOGs that are uniquely down-regulated (B)/up-regulated (E) at 5 DPA in peach hypanthium. **(C, F)** GO enrichment of the SOGs that are uniquely down-regulated (C)/up-regulated (F) at 6 DPA in apple hypanthium.

2.3.3 Identification and comparison of functionally similar clusters in peach and apple consensus co-expression networks

Co-expressed genes across a large number of tissues and stages may fulfil similar functions. I constructed consensus co-expression networks of peach and apple fruit RNA-seq data based a

published pipeline (Shahan et al., 2018). Consequently, the peach and apple genes were grouped into 49 and 59 modules/clusters, respectively (**Figure 2.7**). And the enriched GOs were identified for each cluster. In the study, we are specifically interested in the clusters whose top five enriched GOs are directly related to cell division (group 1), photosynthesis (group 2), signaling (group 3), and response to endogenous stimuli (group 4) (**Figure 2.8A**). The functionally linked peach and apple clusters were shown to share a great number of GO terms (**Figure 2.8B**). And then, the expression patterns of the functionally similar clusters in peach and apple were examined. The 'cell division' clusters in apple exhibit induced expression in the hypanthium and ovary wall while the 'cell division' clusters in peach are relatively lowly expressed in the hypanthium (**Figure 2.8C**). It further supports the finding that apple hypanthium and ovary wall can both become fleshy at early stages, but peach hypanthium fails to grow into fruit.

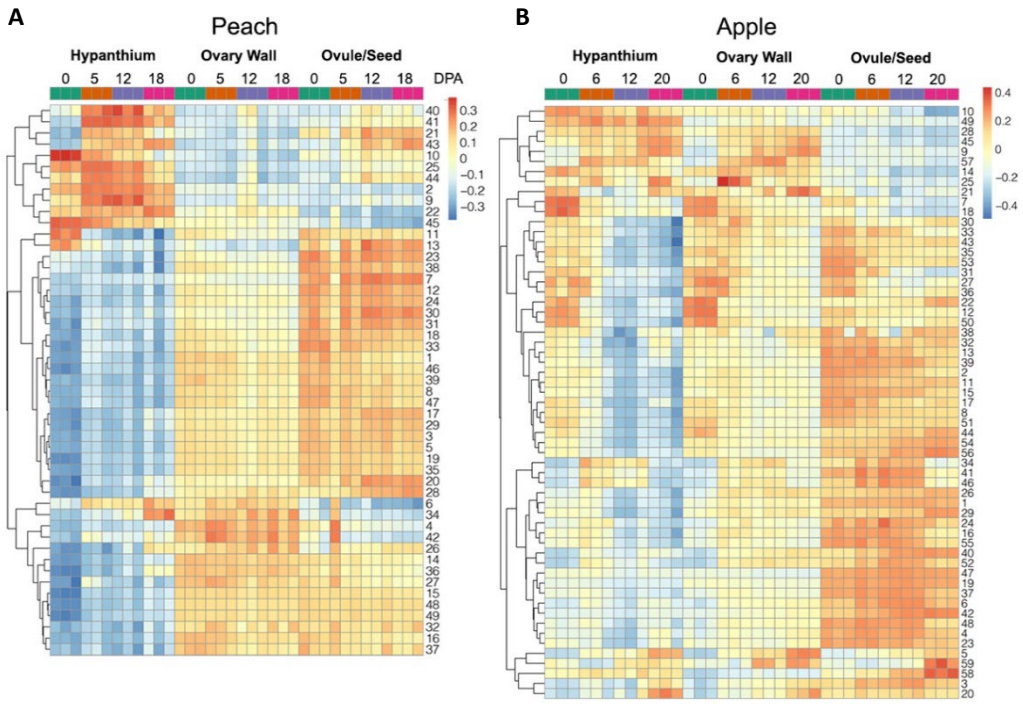

Figure 2.7 The heatmaps showing the eigengenes of each cluster in peach (A) and apple (B)

The eigengenes represent the expression levels of a cluster in different tissues at different stages.

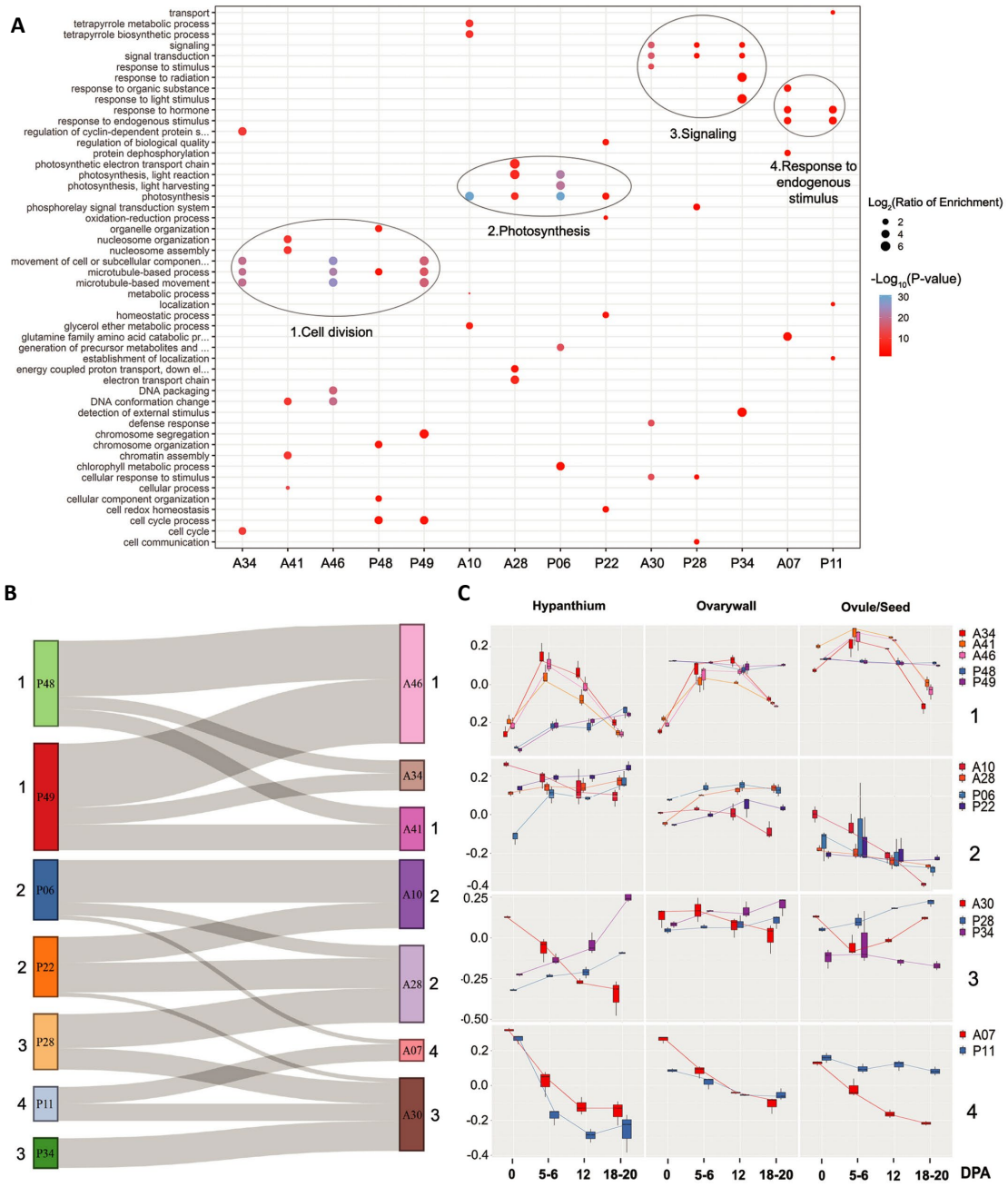

Figure 2.8 Functionally similar clusters in peach and apple

(A) The dot plot shows the 'cell division' (group 1), 'photosynthesis' (group 2), 'signaling' (group 3), and 'response to endogenous stimulus' (group 4) clusters in peach and apple. The clusters labeled with 'A' are apple clusters while the clusters labeled with 'P' are peach clusters. Only the top 5 enriched GO terms are displayed for each cluster. **(B)** The Sankey diagram shows the shared GO terms between peach and apple clusters. The diagram was generated using the top 20 enriched GOs of each cluster. **(C)** For each functional group, the expression levels (eigengenes) of the peach and apple clusters are shown in the box plots. The apple clusters are presented in warm colors, and the peach clusters are presented in cold colors.

2.3.4 Auxin and GA biosynthetic genes are expressed in the fruit flesh-forming tissues

Auxin and GA are important plant hormones that can promote fruit development. The differentially expressed auxin and GA pathway genes were specifically examined to reveal their impacts on peach and apple fleshy fruit formation. It was found that multiple auxin biosynthetic genes, such as *PpTAR2* (*Pp5G168300*), *PpYUC4* (*Pp1G468500*), *PpYUC6* (*Pp1G453400*), and *PpYUC10L* (*Pp8G252500*), are relatively highly expressed in peach ovary wall (**Figure 2.9A**). In apple, the two *MdYUC6* genes that may be resulted from the recent whole genome duplication show distinct expression patterns. The expression of *MdYUC6a* (*Md08G1119300*) was induced in apple hypanthium and ovary wall after pollination (**Figure 2.9B**). But *MdYUC6b* (*Md15G1098700*), as well as *MdYUC8* (*Md10G1172800*), has increased expression after fertilization (**Figure 2.9B**). The data shows that the fruit flesh-forming tissues in peach and apple (peach ovary wall and apple hypanthium and ovary wall) may be able to synthesize auxin in early fruit development and induced by pollination.

Similar to auxin biosynthetic genes, the GA biosynthetic genes can also be expressed in the fruit flesh-forming tissues. Peach ovary wall has a relatively high expression of several GA biosynthetic genes, including *PpGA2* (*Pp4G128500*), *PpGA3* (*Pp1G388500*), *PpKAO1* (*Pp2G109700*), and *PpGA20ox2* (*Pp2G286800*) (**Figure 2.9A**). And the expression of *MdGA3ox1L* (*Md07G1054800*) goes up in apple hypanthium and ovary wall after pollination (**Figure 2.9B**). It indicates that auxin and GA biosynthesis may happen in peach ovary wall and apple hypanthium and ovary wall, which may facilitate the development of the tissues into fruit flesh.

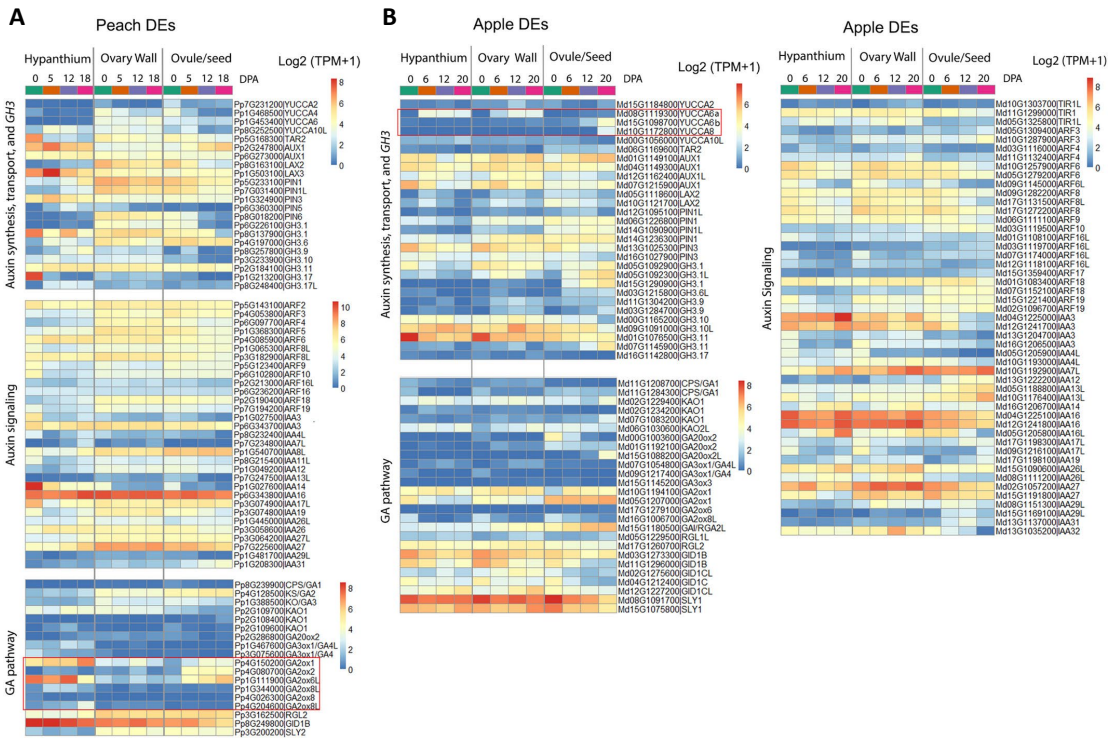

Figure 2.9 Differentially expressed auxin and GA pathway genes in peach (A) and apple (B)

The gene expression was averaged by the three biological replicates and represented by $\log_2$(TPM +1). The gene name was assigned primarily based on BLAST result. If the OrthoFinder and BLAST results were inconsistent, 'L (Like)' would be added to the end of the gene name.

2.3.5 The senescence of peach hypanthium may result from the high expression of *GA2ox* and dPCD genes

GA2ox genes are known to deactivate the bioactive GA (Yamaguchi & Kamiya, 2000). By looking at the expression of the differentially expressed GA pathway genes in peach and apple, we found that multiple *GA2ox* genes, namely *PpGA2ox1* (*Pp4G150200*), *PpGA2ox6L* (*Pp1G1119000*), and *PpGA2ox8L* (*Pp1G344000* and *Pp4G204600*), are highly expressed in peach hypanthium, which was not observed in apple hypanthium (**Figure 2.9**). It indicates that peach hypanthium may not be able to accumulate bioactive GA due to the activity of GA2ox, and thus fail to develop into fruit.

In addition, the expression of the developmental programmed cell death (dPCD) genes was examined in peach and apple. Plentiful dPCD genes were shown to have induced expression after pollination in peach hypanthium (**Figure 2.10**), but they do not exhibit similar expression patterns in apple (**Figure 2.10**). The increased expression of dPCD genes in peach hypanthium may also leads to its senescence.

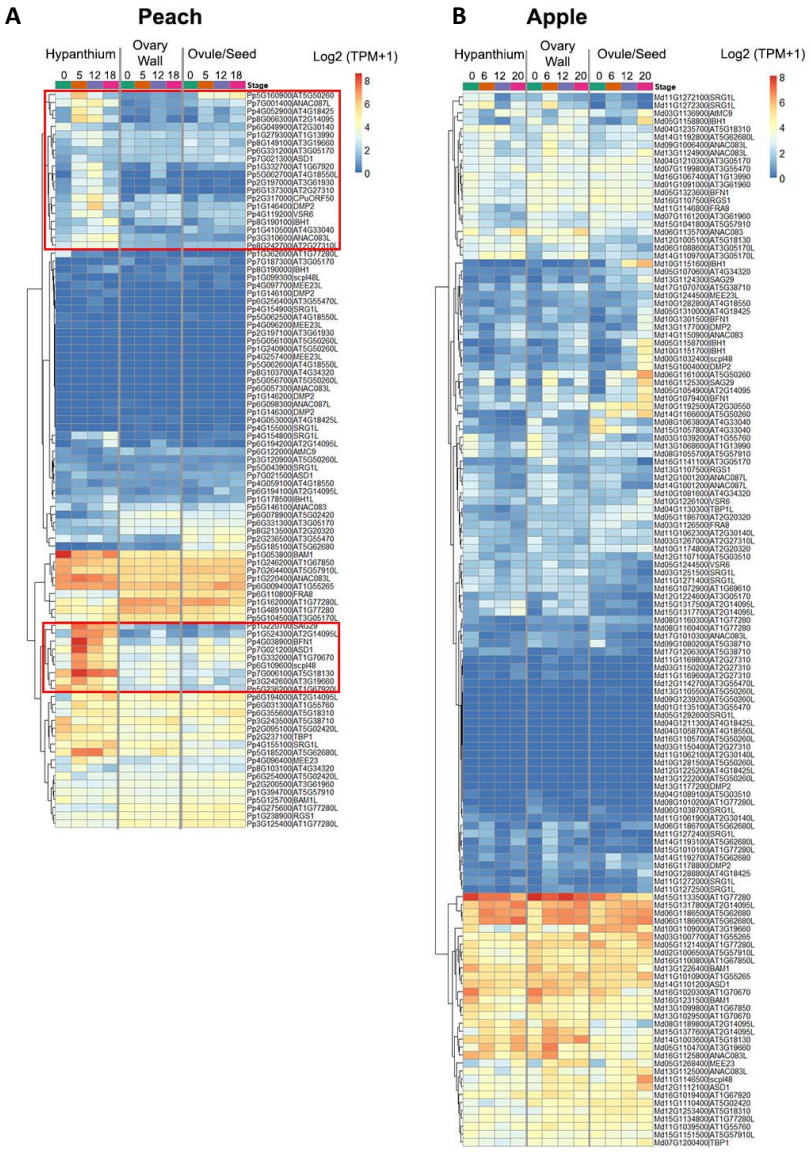

Figure 2.10 Expression of dPCD genes in peach (A) and apple (B) fruit development

The gene expression values are average $\log_2$(TPM +1) of the three biological replicates. The gene name is mainly based on the BLAST result. 'L (Like)' is appended to the end of the gene name when the BLAST and OrthoFinder results are different. The genes that are induced by pollination are highlighted in the red boxes.

2.3.6 Type II MADS-box genes *PI*, *TM6*, and *SEP4* appear to negatively impact fleshy fruit growth

Phylogenetic trees were constructed to identify the specific type II MADS-box genes (MIKC genes) (**Figure 2.11** and **Figure 2.12A, B**) in peach and apple. The expression of differentially expressed MIKC genes was further investigated, which revealed their potential association with fruit development (**Figure 2.12D, E**). Peach *PpPI* (*Pp1G489400*) is expressed at a high level from 0 DPA to 12 DPA in the hypanthium, but almost has no expression in the ovary wall and ovule/seed (**Figure 2.12C**). But in apple, the expression of *MdPI* (*AJ291490*) dramatically decreases after pollination in the hypanthium and ovary wall. As a result, *PI* is likely to be a conserved negative regulator of fruit development. Its high expression in peach hypanthium may inhibit the hypanthium form growing into a fruit, and its reduced expression in apple hypanthium and ovary wall may allow the two tissues to become fruit. *TM6* is also a B class gene, but is absent in Arabidopsis. The *TM6* genes (*Pp7G164100*, *Md02G1136500*, and *Md15G1250200*) show similar expression patterns to *PI* genes in both peach and apple (**Figure 2.12C**). *TM6* is known to be able to form heterodimers with *PI* in other species (Broholm et al., 2010; Vandenbussche et al., 2004) and may work together with *PI* to repress fruit development.

Besides the B class genes, there are other type II MADS-box genes exhibiting interesting expression trends. Similar to *PI* and *TM6*, *PpSEP4* (*Pp1G290500*) is highly expressed in peach hypanthium, but the expression of *MdSEP4b* (*Md16G1058600*) decreases after pollination in apple hypanthium and ovary wall, suggesting *SEP's* negative influence on fruit growth (**Figure 2.12D, E**). Moreover, *SOC1* also appears to be a repressor of fruit development. In peach, *PpSOC1* (*Pp2G151200*) has increased expression in the hypanthium, but decreased expression in the ovary wall after pollination (**Figure 2.12D**). In apple,

pollination seems to reduce the expression of *MdSOC1b* (*Md07G1123600*) in both hypanthium and ovary wall (**Figure 2.12E**).

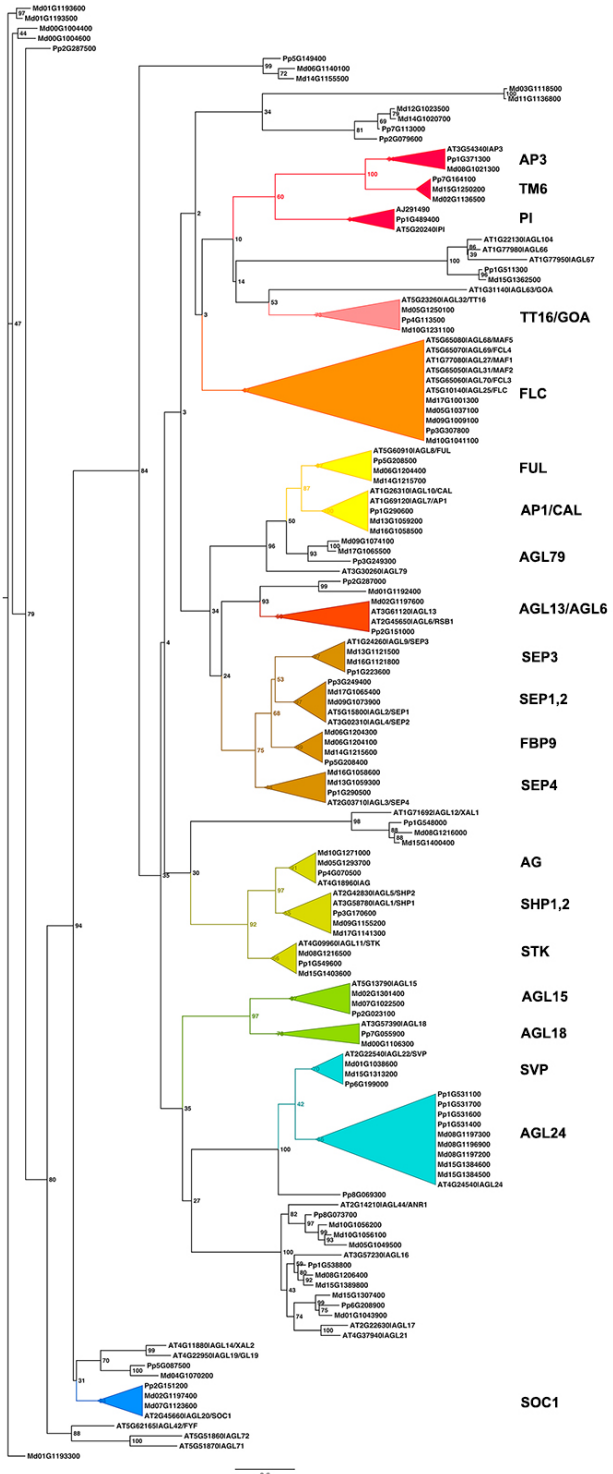

Figure 2.11 Phylogenetic tree of type II MADS-box genes

The phylogenetic tree was built using the Arabidopsis, peach, and apple proteins.

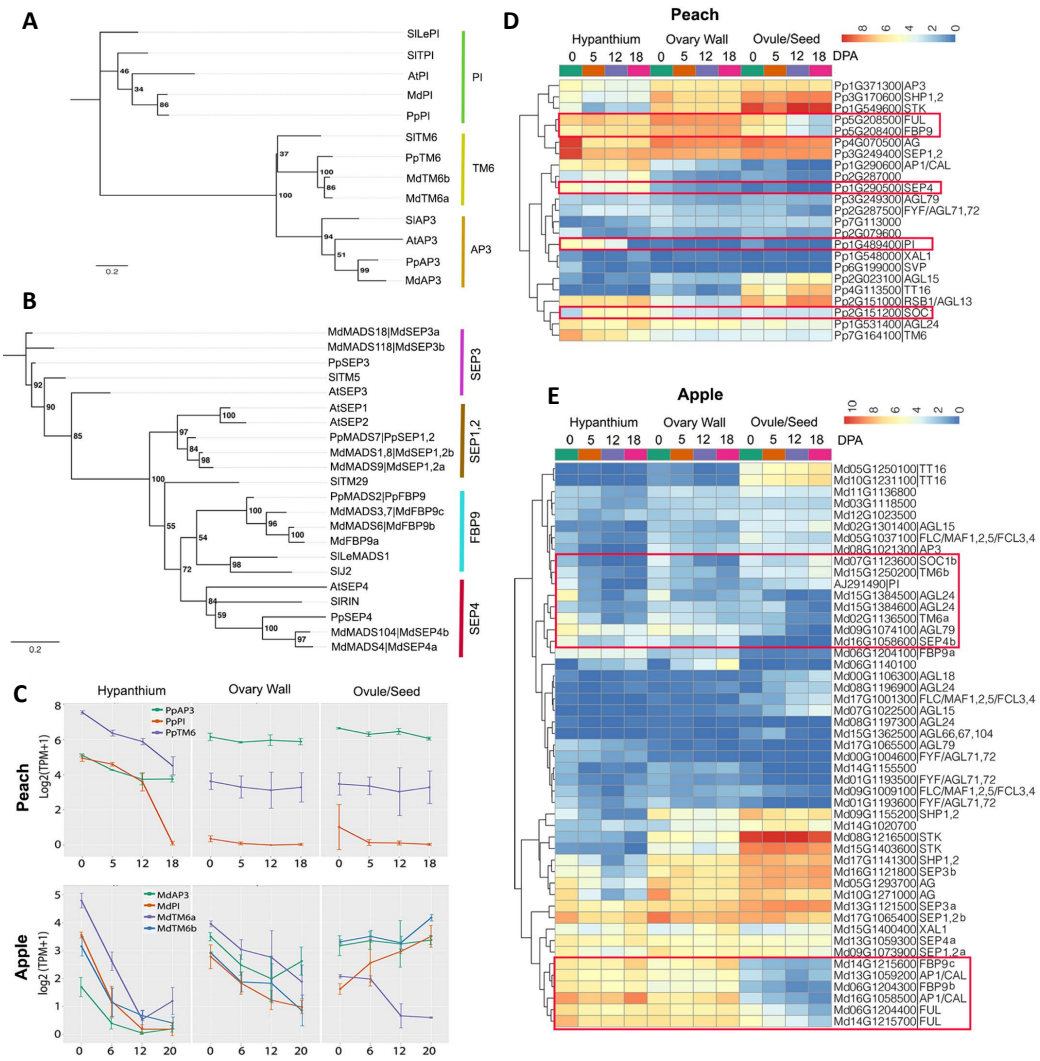

Figure 2.12 Interesting type II MADS-box genes that may regulate fruit development

(A, B) Phylogenetic tree of B (A) and E (B) class genes. At: Arabidopsis, Sl: tomato, Pp: peach, Md: apple. The specific gene IDs that correspond to the gene symbols can be found in the 'Material and method'. **(C)** The expression patterns of B class genes in peach and apple. **(D, E)** Differentially expressed type II MADS-box genes in peach (D) and apple (E). The scale bars show the expression levels represented by average $\log_2$(TPM +1). The gene name assignment is based on the phylogenetic trees. And type II MADS-box genes without a given gene name are the ones that do not have clear Arabidopsis orthologs. The genes in the red boxes are the genes that show particularly interesting expression patterns.

2.3.7 Investigation of type II MADS-box gene *FBP9*'s role in fruit development

It was reported that *FBP* may be positively associated with apple fruit development (Ireland et al., 2013). Therefore, we examined the expression of *FBP9* using our RNA-Seq data generated from early-stage peach and apple fruits. *PpFBP9* (*Pp5G208400*) has slightly higher

expression in peach ovary wall (**Figure 2.12D**). In apple, three *FBP9* genes were identified. *MdFBP9b* (*Md06G1204300*) and *MdFBP9c* (*Md14G1215600*) are both highly expressed in the hypanthium and ovary wall while *MdFBP9a* (*Md06G1204100*) is more highly expressed in the hypanthium (**Figure 2.12E**). *FBP9* genes tend to be preferentially expressed in the fruit flesh-forming tissues.

We then looked into the co-expressed genes with *FBP9* in peach and apple. In peach consensus network, *PpFBP9* is assigned to cluster 6 and is co-expressed with *PpFUL* (*Pp5G208500*), another type II MADS-box gene (**Figure 2.13A**). In apple consensus network, *MdFBP9a* is grouped to cluster 10, and *MdFBP9b* and *MdFBP9c* are both in cluster 28 (**Figure 2.13C**). Moreover, *MdFBP9b* and *MdFBP9c* are co-expressed with two *MdAP1/CAL* genes (*Md13G1059200* and *Md16G1058500*) that are sisters of *FUL* (**Figure 2.11** and **Figure 2.13C**). Since MDAS-box proteins tend to dimerize, FUL and AP1/CAL could be candidate heterodimer candidates of FBP9. Interestingly, 'photosynthesis' is among the top most significant enriched GO terms for both peach cluster 6 and apple clusters 10 and 28 (**Figure 2.8A** and **Figure 2.13B, D**). It suggests that *FBP9*, *FUL*, and *AP1/CAL* genes may promote fleshy fruit growth via photosynthesis.

Since *FBP9* is absent from the genome of Arabidopsis that bears dry fruits, we further investigated the presence or absence of *FBP9* in the genomes of 45 eudicot species available in PLAZA Dicot 4.0 database (Van Bel et al., 2018) to see whether there is an association between presence of an *FBP9* gene and fleshy fruit formation (**Figure 2.14**). Among the 17 fleshy fruit-bearing species, 15 species have *FBP9*. The two exceptions are *Ziziphus jujube* and *Coffea canephora*. However, *FBP9* (*JZ475661.1*) was eventually found in *Z. jujube* by searching against EST or RefSeq databases via BLAST in NCBI, which was later confirmed

by phylogenetic analysis (**Figure 2.15**). Although *FBP9* was not identified in *C. canephora* using similar approach, coffee fruit is mostly composed of seed (coffee bean) and contains very little flesh. In dry fruit forming species, many harbor *FBP9* including *Cajanus cajan* whose *FBP9* was identified by BLAST searches (Figure 2.12). In contrast to fleshy fruit forming species, *FBP9* is lost in 12 of 28 eudicot species that develop dry fruits (**Figure 2.15**). Therefore, dry-fruit forming species more easily lose *FBP9* in their genomes but fleshy fruit species do not lose *FBP9* (with the exception of coffee). This indicates that *FBP9* may be a necessary but insufficient factor for fleshy fruit development.

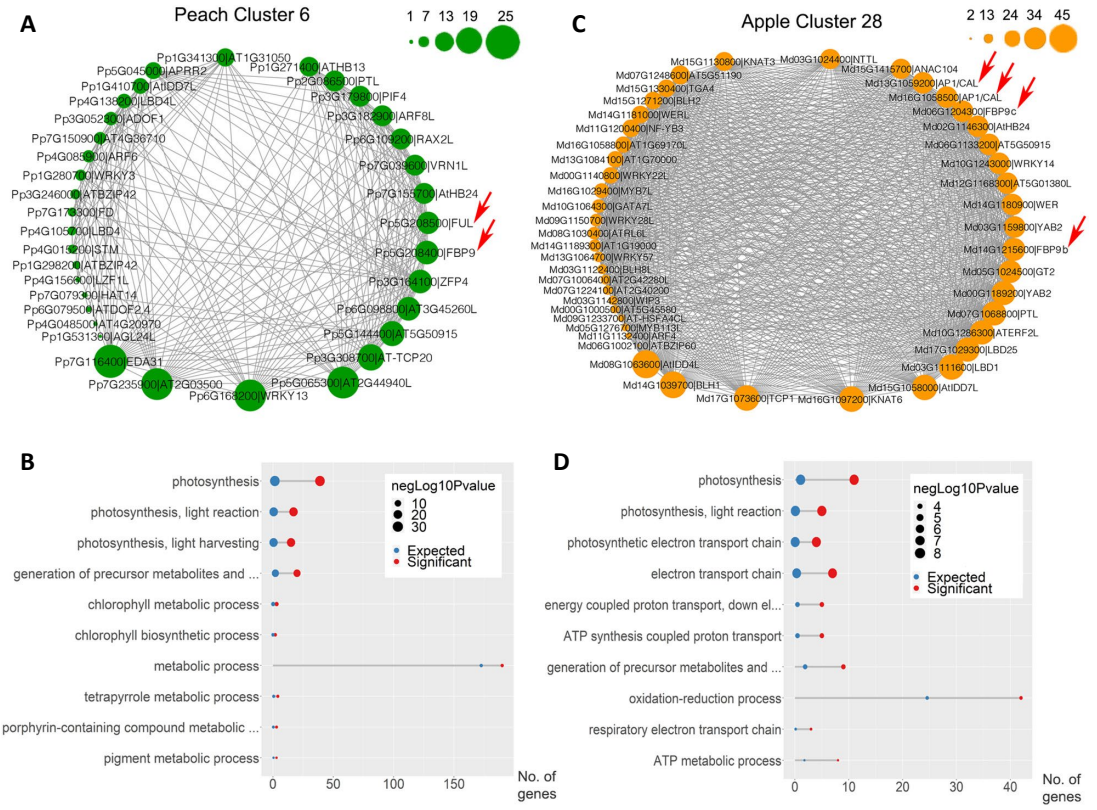

Figure 2.13 The peach and apple clusters possessing *FBP9*

(A, C) Connected transcription factors in peach cluster 6 (A) and apple cluster 28 (C). Each node is a transcription factor. The edge between two nodes indicates that their correlation frequency is no less than 0.5. The size of a node represents its degree. The red arrows point to the interesting type II MADS-box genes. **(B, D)** The top 10 most significant enriched GO terms for peach cluster 6 (B) and apple cluster 28 (D). The red and blue dots show the number of genes with corresponding GO observed and expected in the gene list, respectively.

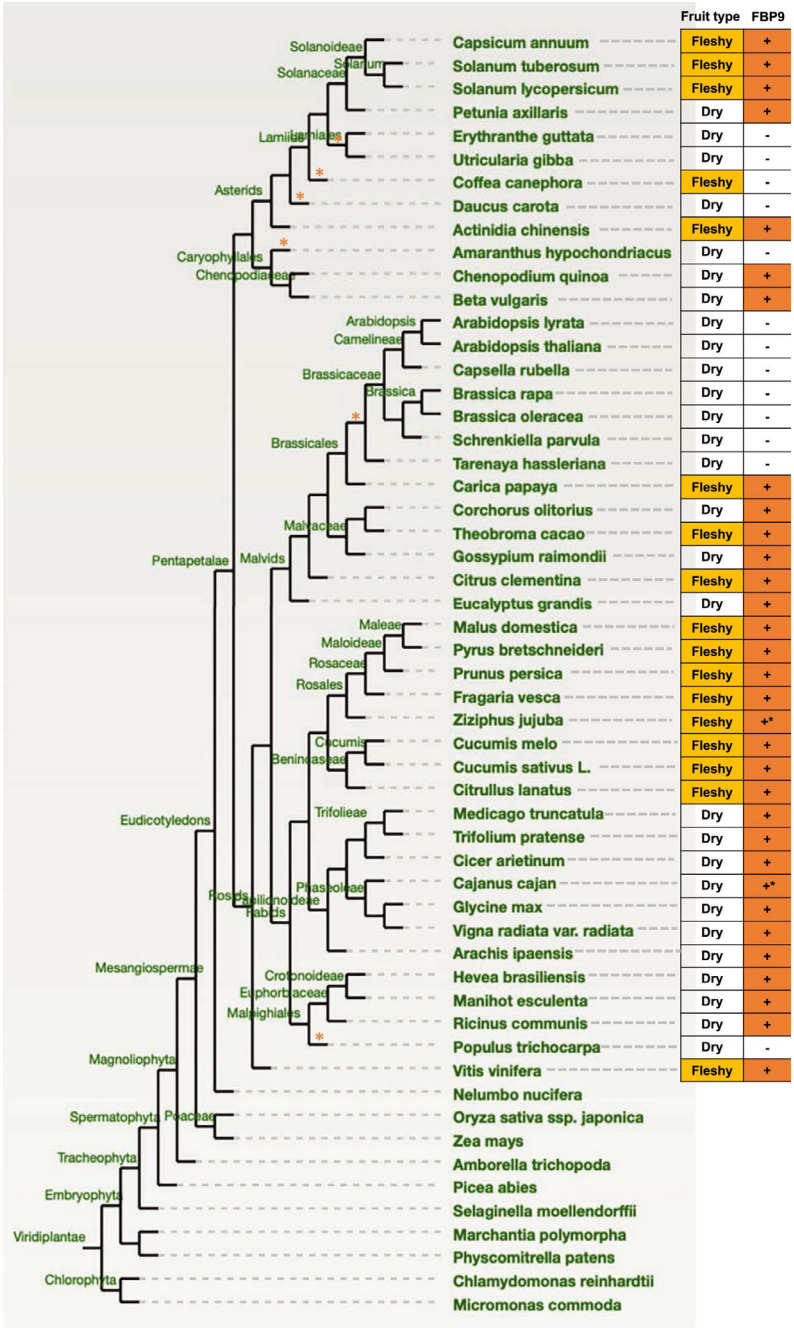

Figure 2.14 Examination of the presence of *FBP9* in the 45 eudicot species in PLAZA Dicot 4.0 database

The phylogenetic tree was adapted from the PLAZA web page (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/). The orange stars (*) point out the losses of *FBP9*. '+' indicates the presence of *FBP9* in the species while '-' indicates the absence of *FBP9* in the species. '+*' highlights the *FBP9* genes that were identified by BLAST searches and confirmed by phylogenetic analysis. The graduate student Yuwei Xiao in the Liu laboratory helped with the identification of *FBP9* in the eudicot species.

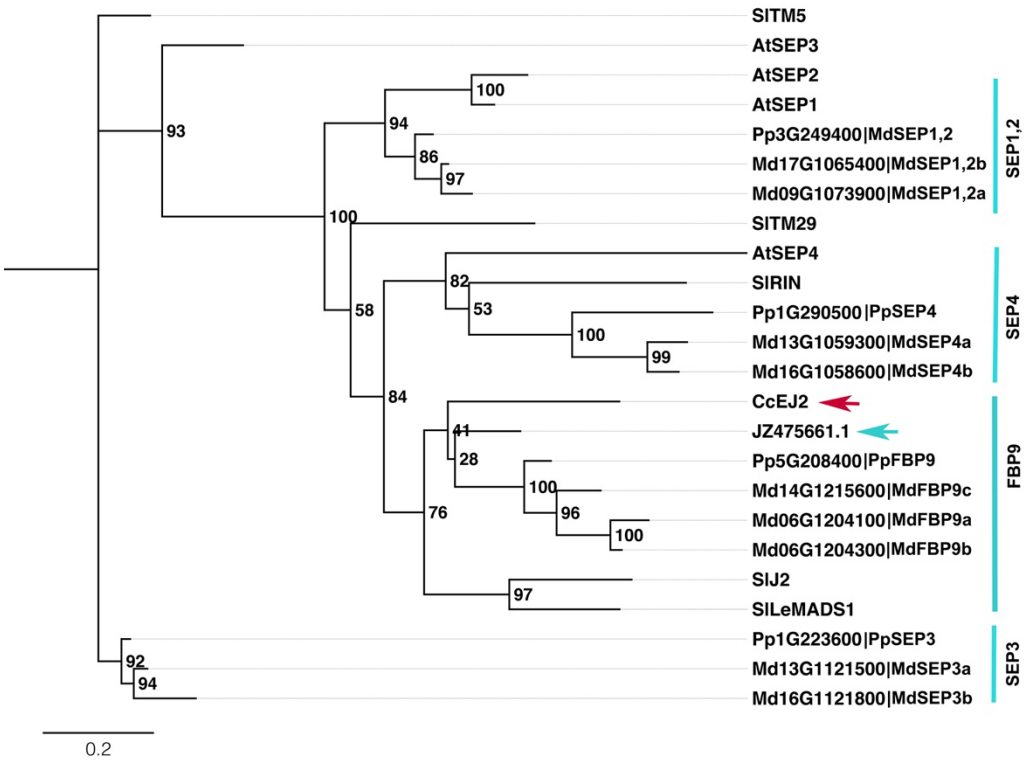

Figure 2.15 Confirmation of the presence of *FBP9* in *Ziziphus jujube* and *Cajanus cajan*

The red arrow points to the *FBP9* in *Cajanus cajan*, and the blue arrow points to the *FBP* in *Ziziphus jujube*. Yuwei Xiao, a graduate student in the Liu laboratory, identified the *FBP* genes in *Cajanus cajan* and *Ziziphus jujube* by performing BLAST searches against EST or RefSeq databases in NCBI.

2.4 Discussion

According to the above analyses, we proposed a model to explain the fruit type specification in four steps (**Figure 2.16**). At the first step, the flower identity is determined by ABCDE genes. At step two, the expression of *FBP9* makes the specific tissues competent to become fleshy. FUL or AP1/CAL may also work together with FBP9 in this process. At step three, pollination/fertilization decreases the expression of *PI* and *TM6* that are two potential negative regulators of fruit development, and stimulates auxin and GA biosynthesis. At step four, the tissues with high expression of *FBP9* and low expression of *PI* can respond to the auxin and GA, and gradually develop into fleshy fruit.

Furthermore, our analyses reveal that pollination itself can also have great influence on fruit development. In peach, pollination and fertilization occur at 0-5 DPA and 5-12 DPA, respectively (M. Li et al., 2022). In apple, the time of pollination is 0-6 DPA that is similar to peach, but fertilization happens later at 12-20 DPA (M. Li et al., 2022). In both peach and apple, there is a short interval between pollination and fertilization. And morphological changes, such as fruit enlargement, can be observed during this period, especially in apple. It suggests a potential correlation between pollination and early fruit growth. Moreover, RNA-Seq data analyses also reveal the general gene expression changes induced by pollination. Both principal component analysis and differential gene expression analysis show dramatic differences between 0 DPA and 5/6 DPA in peach hypanthium and apple hypanthium and ovary wall (**Figure 2.3A, B** and **Figure 2.4A, C**). One important conclusion from this study is that pollination alone plays an important role in initiating fruit development. In terms of specific genes, a great number of dPCD genes are induced by pollination in peach hypanthium, which is likely to lead to the senescence of the hypanthium (**Figure 2.10**).

Our analysis result and proposed model are supported by prior experimental data where *PI*, was shown to encode a fruit repressor in apple (Yao et al., 2001, 2018). A loss-of-function mutation in *PI* can induce parthenocarpic apple fruits (Yao et al., 2001), but overexpression of *PI* had very minor effect by producing apple fruits with reduced size (Yao et al., 2018). One hypothesis for the lack of a severe fruit phenotype in *PI* overexpression is that formation of heterodimers between PI and TM6 may be required to fully inhibit fruit growth. As such a loss of *PI* is sufficient to release the brake on fruit set, but overexpressing *PI* alone is not enough to completely repress fruit development. As *TM6* has similar expression patterns as *PI* in both peach and apple (**Figure 2.12D, E**), it's likely to encode a heterodimer partner of PI in regulating fruit formation. Simultaneous over-expression of *PI* and *TM6* may be more

likely to cause a complete suppression of fruit enlargement. In addition, *PI* may serve as a conserved repressor of fruit development in both peach and apple suggested by our observation that *PI* exhibits deceased expression after pollination in both apple hypanthium and ovary wall, which allow the growth of these two tissues (**Figure 2.12E**). Since multiple type II MADS-box genes, such *TM6*, *SEP4*, *SOC1*, and *AGL24*, are also down-regulated after pollination in apple hypanthium and ovary wall (**Figure 2.12E**), they may also serve as potential partners of *PI*.

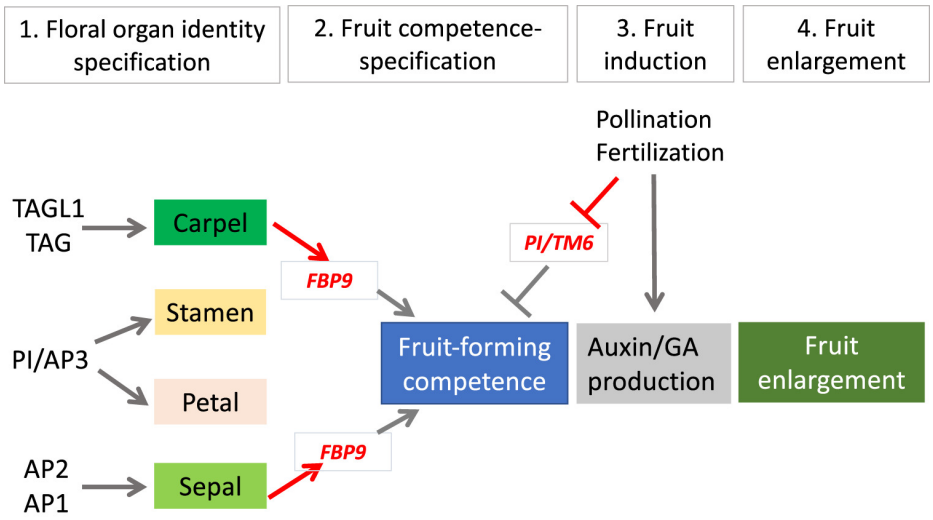

Figure 2.16 A hypothetical model of fleshy fruit formation

Positive correlation was denoted by arrows, and negative correlation was denoted by bars.

Chapter 3: Comparative analysis of raspberry and strawberry transcriptomes from early developing fruit

3.1 Introduction

Raspberry and strawberry are two closely related species that belong to the Rosaceae family, an economically important plant family. Raspberry and strawberry exhibit a similar flower structure with numerous individual carpels sitting on the receptacle, the dome-like stem tip. However, their fruit fleshy comes from distinct floral tissues. In raspberry, each carpel is capable of becoming a fleshy drupelet, while in strawberry, the carpels dry up on the fruit surface, and the receptacle grows into the fruit flesh (**Figure 3.1A**). Given their relatedness, belonging to the same Rosoideae subfamily within Rosaceae, comparative transcriptomic analyses of raspberry and strawberry early-stage fruits would shed light upon how raspberry and strawberry fruit development diverge from each other.

Woodland strawberry (*Fragaria vesca*) is a diploid strawberry that has been widely studied. Chromosome-level genome assembly and comprehensive genome annotation of woodland strawberry have been established (Shulaev. et al., 2011; Edger et al, 2018; Li et al., 2019), which considerably facilitates the transcriptomic analyses of woodland strawberry. A large amount of RNA-Seq data was generated from woodland strawberry fruits about a decade ago to globally investigate the strawberry fruit development (Kang et al., 2013). Moreover, a great number of experiments have been conducted in woodland strawberry to uncover the specific genetic factors that influence strawberry fruit growth, such as MADS box genes, and auxin and GA pathway genes (Pi et al., 2021; Zhou et al., 2021).

Comparatively speaking, much fewer genetic and genomic resources are available for raspberry, which can be divided into black raspberry (*Rubus occidentalis*) and red raspberry

(*Rubus idaeus*). The red raspberry is one of the most economically important small fruits and the subject of this study. To better understand raspberry fruit development, we dissected red raspberry fruit tissues at important early developing stages to observe their morphological changes over time. And plant hormones, such as auxin and GA, were applied to the raspberry fruits to reveal the impacts of these hormones on its fruit growth. Moreover, we generated a high-quality genome of red raspberry (Rubus idaeus 'Joan J.') with thorough annotation using RNA-seq data from distinct fruit tissues at specific stages. Therefore, we were able to compare our newly generated red raspberry RNA-Seq data to the publicly available woodland strawberry RNA-Seq data (Kang et al., 2013), and proposed a model to potentially explain why similar flower tissues show distinct capability to develop into fruits in raspberry and strawberry.

3.2 Material and methods

Data collection

Genomic DNA was isolated from young leaves of *Rubus idaeus* 'Joan J.' for genome assembly. The 'Joan J.' plants were obtained from Appalachian Fruit Research Station of USDA ARS. The PacBio and Illumina DNA sequencing was conducted at the Genomics Resource Center of the University of Maryland School of Medicine's Institute of Genome Sciences. And the Nanopore DNA reads were produced in the Wimp lab at the Johns Hopkins University.

Raspberry fruits were collected from 'Joan J.' at 0, 2, 4, 6, 9, 12 DPA. And the fruits were further dissected into three tissues, receptacle, ovary wall, and ovule/seed. RNA was extracted from the four biological replicates that were generated for each tissue at every

single stage. And the RNA sequencing data were yielded by Weill Cornell Medical College, New York.

Dr. Junhui Zhou worked on the experimental procedures, including fruit tissue dissection and section, sequencing sample preparation, and hormone application.

Genome assembly

Nanopore and PacBio raw reads were initially corrected by Canu (v2.0) (Koren et al., 2017) that was further used to construct Nanopore reads into draft contigs. The contigs were first polished by Racon (v1.4.3) (Vaser et al., 2017) for two rounds using the corrected Nanopore reads and then another round with the corrected PacBio reads. Subsequently, the accurate Illumina short reads were utilized by Pilon (v1.23) (Walker et al., 2014) to further correct the contigs. The size of the resulting genome assembly was 464 Mb. However, the genome size estimated by flow cytometry was around 293Mb (Graham & Woodhead, 2009), which suggested significant polymorphism and heterozygosity in the genome. A coverage analysis performed by the Purge Haplotigs pipeline (v1.0.4) (Roach et al., 2018) revealed a large proportion of uncollapsed haplotype contigs in the assembly. And a BUSCO (v3.0.1) (Simão et al., 2015) analysis further supported it by showing a high duplication score of 43.0% (complete: 95.5%). Therefore, the Purge Haplotigs pipeline was applied to generated a haplotype-fused assembly with size of 286Mb using the 100X Nanopore raw reads. Furthermore, super scaffolds were constructed by SSPACE-LongRead (v1-1) (Boetzer & Pirovano, 2014) using the corrected Nanopore reads that were later used by GapFinisher (v0.1) (Kammonen et al., 2019) for gap filling. Eventually, two rounds of genome assembly correction were performed by Pilon with Illumina reads.

To anchor the assembly to genetic linkage maps of Heritage and Tulameen, BLAT (Kent, 2002) was used to map the genetic markers from the previous published genetic maps (Ward et al., 2013) to the assembled scaffolds. ALLMAPS (JCVI utility libraries v1.0.1) (Tang, Zhang, et al., 2015) was employed to anchor the scaffolds into pseudochromosomes with default parameters. Chimeric scaffolds were manually broken at positions with low coverage.

For Repeat annotation, *De novo* repeat library was built by LTR-retriever (v2.9.0) (Ou & Jiang, 2018) and RepeatModeler (v1.0.11) (https://www.repeatmasker.org/RepeatModeler/). The library was then fed into RepeatMasker (v4.1.0) (https://www.repeatmasker.org/) to annotate and mask the repetitive elements in the genome.

Genome annotation

A combination of ab initio gene models, transcript evidence derived from RNA-Seq data, and protein homology-based evidence was used to predict the protein-coding genes. Three gene predictors, MAKER (v2.53) (Campbell, Holt, et al., 2014), AUGUSTUS (v2.7) (Stanke et al., 2006) and BRAKER2 (v2.1.4) (Hoff et al., 2019), were used to identify the potential gene models in the repeat-masked genome. The gene predictors could also make use of the transcript and protein evidence for better genome annotation. The Illumina RNA-Seq reads were trimmed by fastp (v0.20.0) (S. Chen et al., 2018) and mapped to the genome using STAR (v2.7.4a) (Dobin et al., 2013). StringTie (v2.0.1) (Pertea et al., 2015) was further used to assemble the mapped reads into transcripts with default settings except that the lowly expressed isoforms were excluded by setting the minimum isoform fraction to 0.2. A de novo transcriptome assembly was generated by Trinity (v2.11.0) (Grabherr et al., 2011), and GMAP (v2015-07-03) (Wu & Watanabe, 2005) was used to map the transcripts to the genome. Moreover, the plant protein sequences from black raspberry (VanBuren et al., 2018),

strawberry (Y. Li et al., 2019), *Arabidopsis* (https://www.arabidopsis.org/index.jsp), and UniProt (https://www.uniprot.org/) were aligned to the genome with Exonerate (v2.2.0) (Slater & Birney, 2005).

Next, EVidenceModeler (EVM v1.1.1) (Haas et al., 2008) was used to identify the confident consensus gene models based on the gene models produced by the three predictors, and the transcript and protein evidence using nonstochastic weight values. Afterwards, PASA (v2.4.1) (Haas et al., 2003) further modified the gene structures and added the UTRs and alternatively spliced isoforms to the EVM gene models. Finally, the gene models were manually curated using Apollo (v2.6.1) (Lewis et al., 2002) based on the mapped RNA-seq reads.

To further improve genome assembly, the misassembled scaffolds were further broken based on the ALLMAPS output. Subsequently, the genetic markers were mapped to the resulting scaffolds with BLAT (Kent, 2002). And the scaffolds were assembled into a chromosome-level genome using ALLMAPS (Tang, Zhang, et al., 2015) again. The genome annotation was then lifted over to the new assembly by Liftoff (v1.6.1) (Shumate & Salzberg, 2021). QUAST (v5.0.2) (Gurevich et al., 2013) and esl-seqstat from Easel (v0.48) (https://github.com/EddyRivasLab/easel) were used to evaluate the quality of the genome assemblies of Rubus idaeus 'Joan J.' and 'Anitra' (Davik et al., 2022). And the genome and transcriptome completeness were assessed by BUSCO (v5.3.1) (Simão et al., 2015) using dataset embryophyta_odb10. OmicsBox (v1.2.4) (Götz et al., 2008) was utilized to functionally annotate the longest peptides of red raspberry and woodland strawberry (*Fragaria vesca* Genome v4.0.a2) (Y. Li et al., 2019), respectively.

The initial genome assembly and annotation were performed by Dr. Yongping Li. And the improved genome and lifted over annotation were generated by me. The downstream analyses were conducted by me as well.

Ortholog Identification

OrthoFinder (v2.3.8) (Emms & Kelly, 2015) was applied to identify the orthologs among the six plant species including Arabidopsis (TAIR10_pep_20101214), peach (*Prunus persica* Genome v2.0.a1) (Verde et al., 2017), apple (*Malus x domestica* GDDH13 Whole Genome v1.1) (Daccord et al., 2017), woodland strawberry (*Fragaria vesca* Genome v4.0.a2) (Y. Li et al., 2019), black raspberry (*Rubus occidentalis* whole genome assembly v3.0) (VanBuren et al., 2018), and red raspberry. Due to lack of MdPI gene in the apple GDDH13 genome, MdPI (AJ291490) sequences were manually added to the GDDH13 reference transcripts and proteins.

BLAST (v2.5.0) (Altschul et al., 1990) was used to search the red raspberry and woodland strawberry (*Fragaria vesca* Genome v4.0.a2) (Y. Li et al., 2019) proteins against Arabidopsis database (TAIR10_pep_20101214). The cut-off *E*-value was 10^{-5}.

Proteins with K-domain were detected by HMMER (v3.3) (Mistry et al., 2013) in five Rosaceae species (peach, apple, woodland strawberry, black raspberry, and red raspberry) using the alignment file (PF01486) downloaded from Pfam (El-Gebali et al., 2019). The type II MADS-box proteins in Arabidopsis were derived from TAIR (https://www.arabidopsis.org/browse/genefamily/mads_tffamily.jsp). MAFFT (v7.458) (Katoh & Standley, 2013)was further used to align the full protein sequences of the type II MADS-box genes from the six species. Based on the alignment, the phylogenetic tree was

eventually constructed by RAXML (v8.2.12) (Stamatakis, 2014) and visualized by FigTree (v1.4.3) (http://tree.bio.ed.ac.uk/software/figtree/).

Only the longest peptides were used for the above analyses. Moreover, the gene names assigned to the woodland strawberry and red raspberry genes mainly depend on the BLAST and OrthoFinder results. If the BLAST and OrthoFinder results of a gene disagree with each other, the gene name will be the BLAST result plus 'L' ('LIKE'). And the gene names of the type II MADS-box genes in woodland strawberry and red raspberry will be given according to the phylogenetic tree.

Differential gene expression analysis

The low-quality bases with the scores less than 25 were trimmed by Cutadapt (v2.8) (M. Martin, 2011) from the 3' ends of the red raspberry RNA-Seq reads. And only the reads with a length no less than 36 bp were remained for the downstream analyses. The woodland strawberry RNA-Seq data were obtained from SRA (PRJNA187983) (Kang et al., 2013). The woodland strawberry and red raspberry RNA-Seq reads were mapped to their corresponding genomes by Salmon (v0.11.2) (Patro et al., 2017), respectively. And the transcript abundances were summarized by tximport (v1.10.1) (Soneson et al., 2016) for gene-level analyses.

In both woodland strawberry and red raspberry, the differentially expressed genes (DEGs) between adjacent stages in each tissue and the DEGs between different tissues at each stage were identified by DESeq2 (v1.22.2) (Love et al., 2014). The lowly expressed genes (raspberry: read counts ≤ 72, strawberry: read counts ≤ 46) were excluded in the DE analysis. The genes with $P_{adj} < 0.05$ and $|\log_2 FoldChange| > 1$ were considered as differentially

expressed genes. And the PCA plots were generated by the DESeq2 built-in function plotPCA and ggplot2 (Wickham, 2016).

Grade of membership (GoM) analysis

CountClust (v1.10.1) (Dey et al., 2017) was utilized to perform the grade of membership analysis using TPM values for red raspberry and strawberry. Different number of clusters (K=10, 15, 20, 30) were tested to fit the GoM model to the gene expression data.

Identification of tissue- and stage-specific genes

The tissue- and stage-specific genes were identified by TissueEnrich (v1.2.1) (Jain & Tuteja, 2019) using the TPM values averaged over biological replicates. A specific tissue that is collected at a certain stage is considered as a specific sample type. Tissue- and stage-specific genes are the genes whose TPM is no less than one in a specific sample type, and their expression in this sample type are at least five-fold higher than that in all other sample types. The GO enrichment analysis was performed for the tissue- and stage-specific genes by topGO (v2.34.0) (Alexa & Rahnenfuhrer, 2022). And the *P*-value cut-off was 0.05.

3.3 Results

3.2.1 Early fruit development of raspberry

The red raspberry fruits were collected and dissected at critical stages of early fruit development from 0 DPA (days post anthesis) to 12 DPA. At 0 DPA, long and white styles are attached to the top of the ovaries (**Figure 3.1B**). At 4 DPA, fertilization has occurred (**Figure 3.1C**). The styles become dry, and the ovaries and the receptacle dramatically enlarge (**Figure 3.1B**). At 6 DPA, the ovaries continue to grow, but the ovary wall is still thin (**Figure 3.1B**). Moreover, globular embryos appear in the transparent seeds, and soft stone

are formed with jagged edges (**Figure 3.1C**). At 9-10 DPA, the seeds turn opaque and the embryos inside grow into a heart shape (**Figure 3.1B**). Additionally, the stone becomes yellow and a little bit hardened (**Figure 3.1B**). At 12 DPA, the seeds are entirely white, and the stone is completely hardened (**Figure 3.1B**). Despite a lot of changes happening in other tissues, raspberry receptacle hardly shows any differences in its size from 6 DPA to 12 DPA (**Figure 3.1B**). The observation of the morphological changes in the raspberry fruit tissues not only allows us to better understand the fruit development processes in raspberry, but also benefits the downstream RNA-Seq analyses by identifying key stages and corresponding changes.

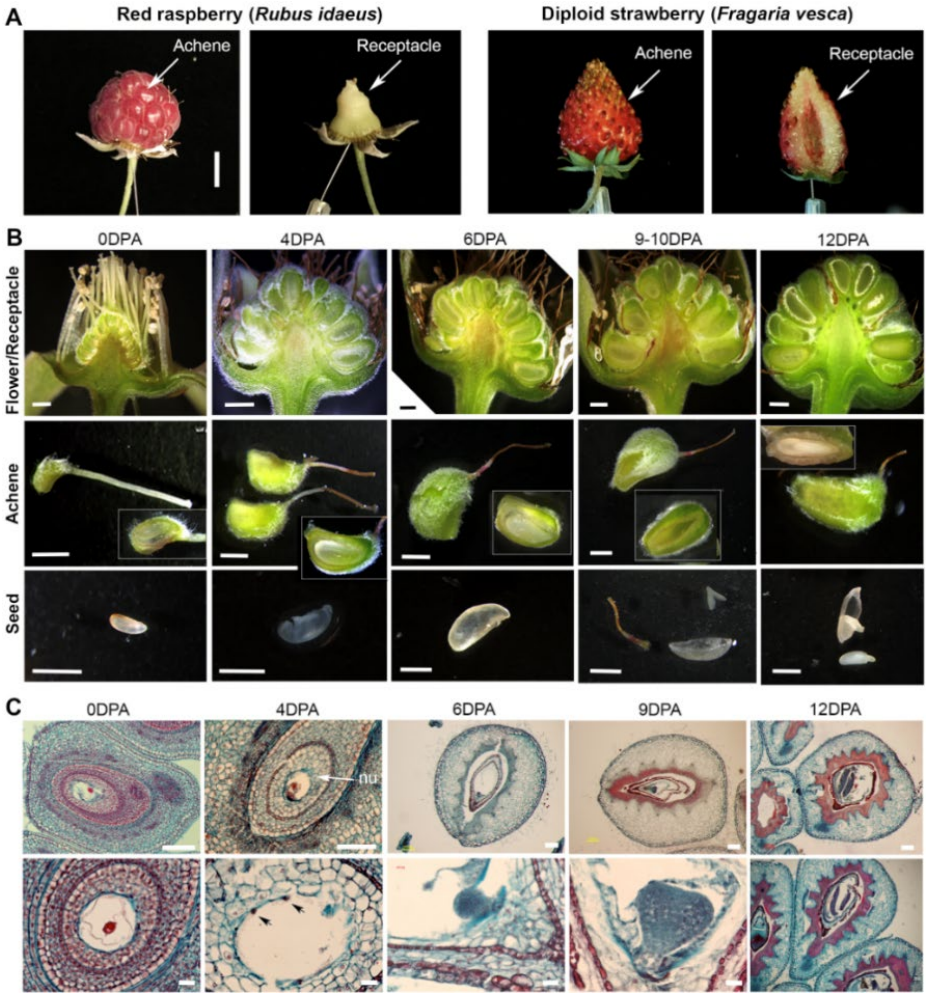

Figure 3.1 Early fruit development of raspberry

The fruit tissue dissection and section were conducted by Dr. Junhui Zhou. **(A)** Red raspberry fruit compared to the woodland strawberry fruit. Scale bar: 1cm. **(B)** Dissected flower/fruit tissues at 0, 4, 6, 9-10, and 12 DPA. Scale bar: 1mm. **(C)** Tissue sections at 0, 4, 6, 9, and 12 DPA. Black arrows point to the divided endosperm nuclei indicating successful fertilization by 4 DPA. The lignin deposition in the stone was stained red from 9 DPA. nu: nucellus. Scale bar (top): 100um. Scale bar (bottom): 20um.

3.2.2 Parthenocarpic raspberry induced by auxin and GA treatments together

Plant hormones, auxin (NAA) and GA (GA3), were applied to emasculated raspberry flowers separately or together to reveal their impacts on fruit development (**Figure 3.2A**). And the negative and positive controls are mock-treated emasculated flower and self-pollinated flower, respectively (**Figure 3.2A**). It was found that only when the emasculated flowers are treated with NAA and GA3 together, they can develop into the parthenocarpic fruits (**Figure 3.2A**). The parthenocarpic drupelets were cut open showing the absence of the seed and normal lignin formation (**Figure 3.2B**). Different from raspberry, application of single hormone (NAA or GA3) is sufficient to produce fruits, albeit at a reduced size, in strawberry, and application of both NAA and GA3 leads to full fruit enlargement (Kang et al., 2013). It suggests there is difference in the auxin and GA cross-talks between strawberry and raspberry during fruit initiation.

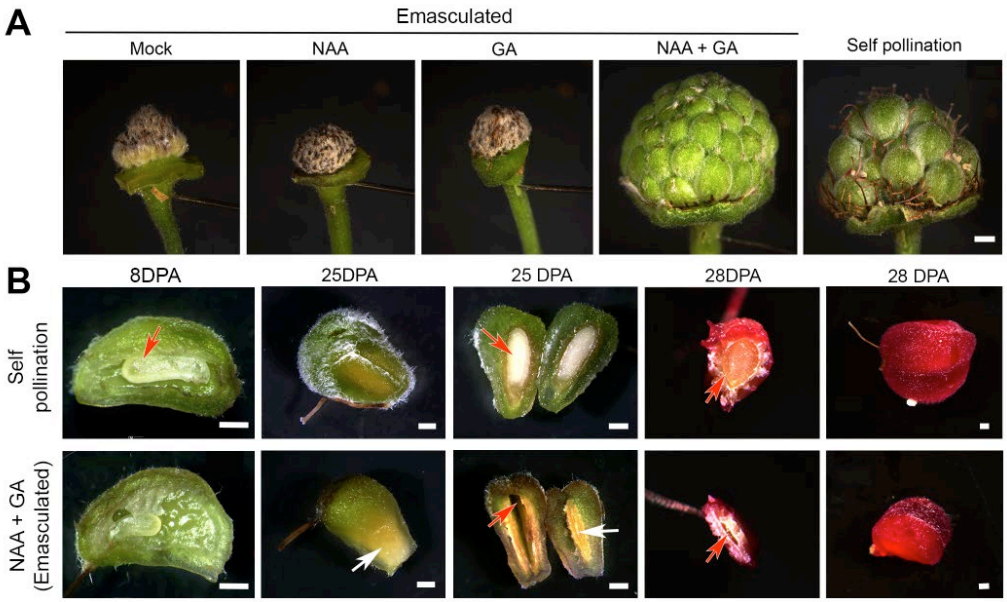

Figure 3.2 Investigation of plant hormone's impact on raspberry fruit development

The hormone treatment and fruit dissection were conducted by Dr. Junhui Zhou. **(A)** Hormone-treated raspberry fruits. Left to right: Mock-treated (negative control), NAA-treated, GA3-treated, and NAA- and GA-treated emasculated raspberry fruits, and self-pollinated raspberry fruit (positive control). Scale bar: 1000 um. **(B)** Dissected achenes from self-pollinated raspberry (top) and NAA- and GA-treated raspberry (bottom). Red arrows point to the seed, and the white arrows point to the stone. Scale bar: 500 um.

3.2.3 Construction of *Rubus idaeus* 'Joan J.' genome

The first *Rubus idaeus* ('Joan J.') genome (version 1.0) was constructed using PacBio long reads and Illumina short reads (Wight et al., 2019). However, the draft genome has a low scaffold N50 of 638 kb, and the size of its largest scaffold is 4.46 Mb (**Table 3.1**). Recently, the genome of *Rubus idaeus* 'Anitra' has been sequenced (Davik et al., 2022). The scaffold N50 of the genome assembly is 34.49 Mb, but the BUSCO completeness score of the transcriptome is 81.4% (**Table 3.1**). Here, Nanopore, PacBio and Illumina sequencing data, as well as the linkage maps, were utilized to assemble a high-quality genome of the red raspberry *Rubus idaeus* 'Joan J.' (version 2.0) (**Figure 3.3A and Figure 3.4**). The genome assembly has a total length of 297 Mb and a scaffold N50 equal to 33.74 Mb (**Table 3.1**). Moreover, the genome was further annotated using the fruit RNA-Seq data (**Figure 3.3B**). As a result, 33,865 protein-coding genes and 57,192 isoforms were predicted in the genome, and the BUSCO completeness score of the transcriptome is 96.5% (**Table 3.1**). The new genome assembly and annotation allow us to estimate the gene expression more accurately, which is beneficial for the downstream RNA-Seq analyses.

	'Joan J.' version 2.0 (This work)	'Joan J.' version 1.0 (Wight et al., 2019)	'Anitra' version 1.0 (Davik et al., 2022)
Total genome Length (bp)	297,436,202	300,259,977	293,523,434
Scaffold N50 (bp)	33,744,647	638,152	34,491,998
Largest Scaffold (bp)	44,668,142	4,458,320	43,293,752
Smallest Scaffold (bp)	4,032	501	1,000
N's	870,445	174,429	333,000
Sequence GC's	37.96%	37.9%	37.86%
BUSCO Completeness Score (Genome)	C:98.4% [S:93.7%, D:4.7%], F:0.8%, M:0.8%, n:1614	C:98.4% [S:89.7%, D:8.7%], F:0.7%, M:0.9%, n:1614	C:98.6% [S:92.4%, D:6.2%], F:0.7%, M:0.7%, n:1614
Number of Protein Coding Genes	33,865	35,566	39,448
Number of Transcripts	57,192	35,566	48,176
BUSCO Completeness Score (Transcriptome)	C:96.5% [S:46.7%, D:49.8%], F:2.3%, M:1.2%, n:1614	C:99.2% [S:88.5%, D:10.7%], F:0.2%, M:0.6%, n:1614	C:81.4% [S:74.3%, D:7.1%], F:5.1%, M:13.5%, n:1614

Table 3.1 Comparison of three Rubus idaeus genome assemblies.

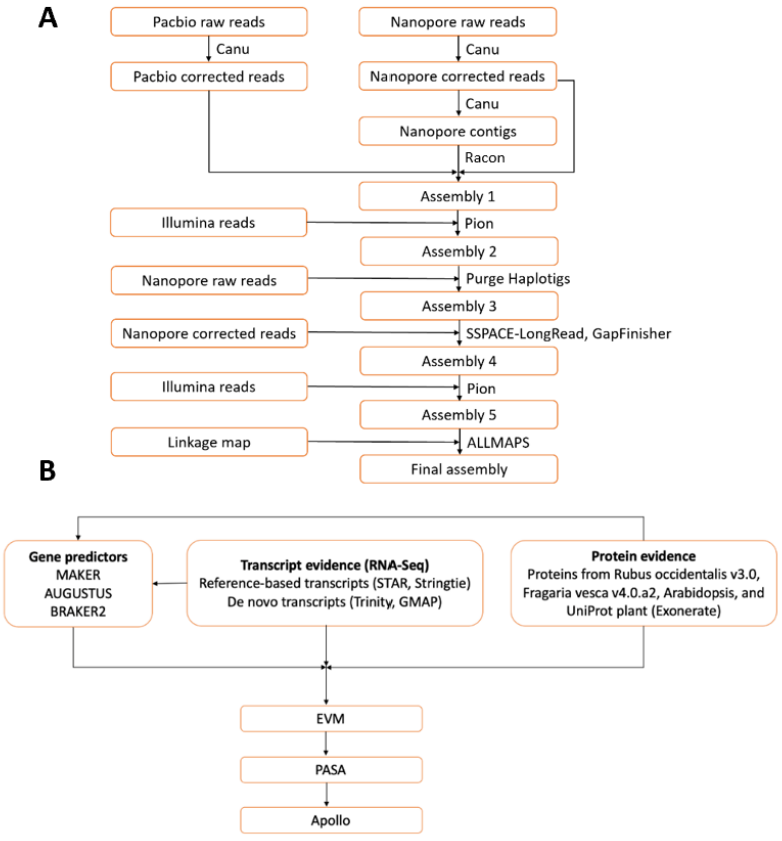

Figure 3.3 Genome assembly (A) and annotation (B) pipelines

Initial genome assembly and annotation were conducted by Dr. Yongping Li. The further genome improvement, annotation lift over, and the downstream analyses was performed by me.

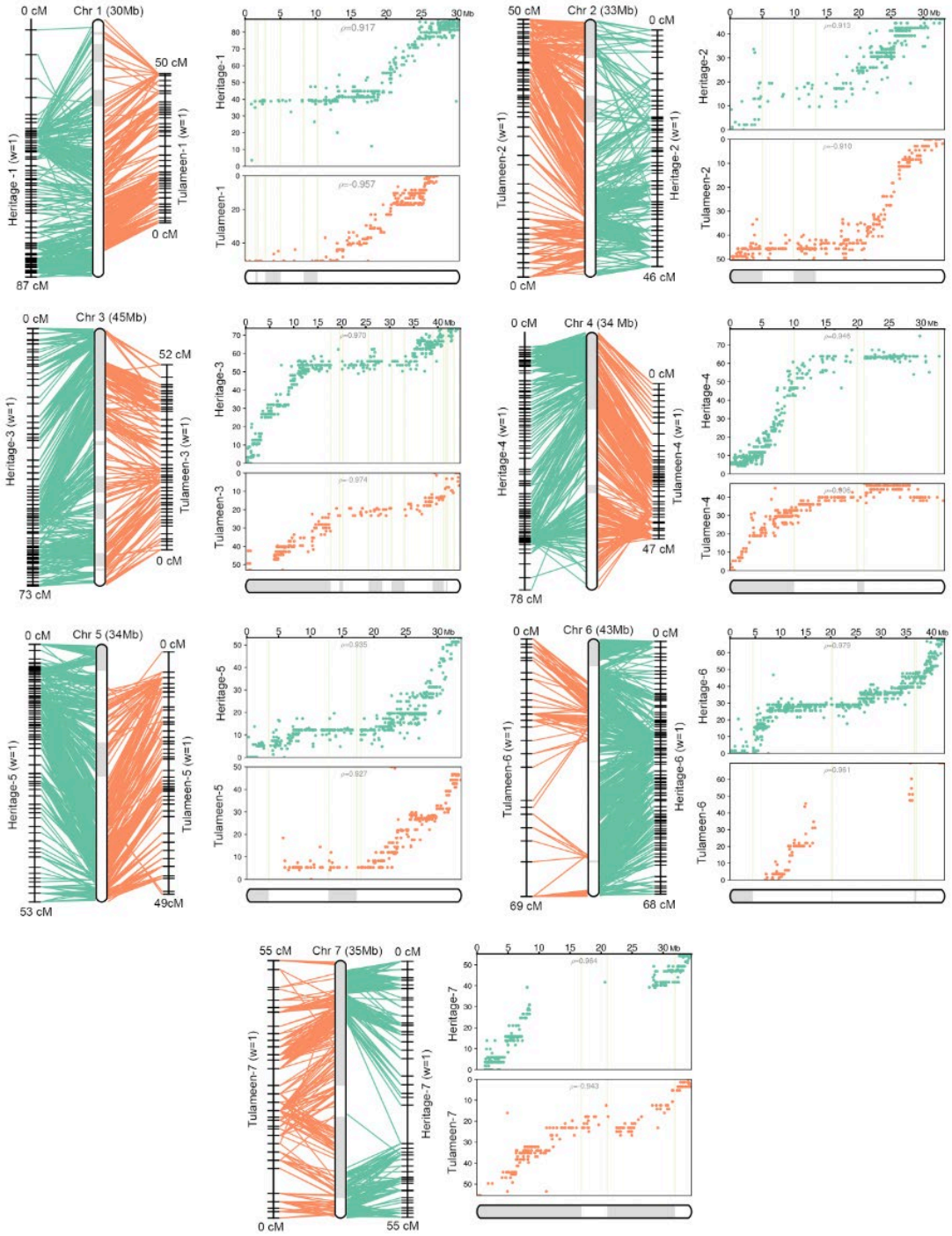

Figure 3.4 Seven pseudomolecules assembled using the linkage maps of 'Heritage' and 'Tulameen'

253 Mb (85.1%) of sequences were anchored onto the seven chromosomes. The left panel displays the connections between the constructed chromosomes and the linkage maps. The Rho (ρ) values shown at the top of the scatter plots (right panel) are the Pearson correlation coefficients between the physical positions on the chromosome (x axis) and map positions (y axis).

3.2.4 Global view of early fruit RNA-Seq data

The raspberry early fruit RNA-Seq data were collected from three distinct tissues (receptacle, ovary wall, and ovule/seed) at six early developmental stages (0, 2, 4, 6, 9, 12 DPA), and were further compared with the previously published woodland strawberry RNA-Seq data from early fruit development (Kang et al., 2013). The principal component analysis revealed that compared to the ovary wall and the seed, the receptacle undergoes relatively fewer changes during early fruit development of both raspberry and strawberry (**Figure 3.5A**). In raspberry receptacle, there are obvious gene expression differences between 0 and 2 DPA, which may explain the enlargement of the receptacle at the beginning of the fruit development (**Figure 3.5B**). However, raspberry receptacle samples from later stages are grouped together in the PCA plot supporting that no dramatic morphological changes are observed at later stages in the receptacle (**Figure 3.5B**). Different from raspberry receptacle, strawberry receptacle, especially cortex, shows clear gene expression changes over time, which is consistent with the fact that strawberry receptacle gradually enlarges throughout the whole early fruit development (**Figure 3.5B**). Additionally, in raspberry, the ovary wall and seed start to grow from 4 DPA (**Figure 3.5A**). And there are remarkable differences between 6 DPA and 9 DPA in raspberry ovary wall and seed (**Figure 3.5A**).

Additionally, grade of membership analysis estimated the cluster membership proportions in each sample (**Figure 3.5C**). Gene cluster 3 that constitutes a significant proportion of raspberry receptacle was found to participate in ion transport, auxin signaling, and response to biotic and abiotic stresses based on GO enrichment analysis (**Figure 3.5C**). And in strawberry, the cortex and pith, two tissues in the receptacle, are greatly represented by gene cluster 1 that has enriched GO terms mainly in ion or zinc transport (**Figure 3.5C**). The receptacle vasculature supports the finding of potential active ion transport in red raspberry

and strawberry receptacle. Moreover, the enriched GO terms of the gene cluster 3 that characterizes the strawberry ovary wall are associated with photosynthesis and cell wall synthesis (**Figure 3.5C**). In raspberry ovary wall, the representative gene clusters 1 and 8 have enriched GO terms related to peptide synthesis and photosynthesis, respectively (**Figure 3.5C**). Therefore, the clusters characterizing the raspberry and strawberry ovary wall have different enriched GO terms, which may help explain the distinct developmental fate of the ovary wall in each species.

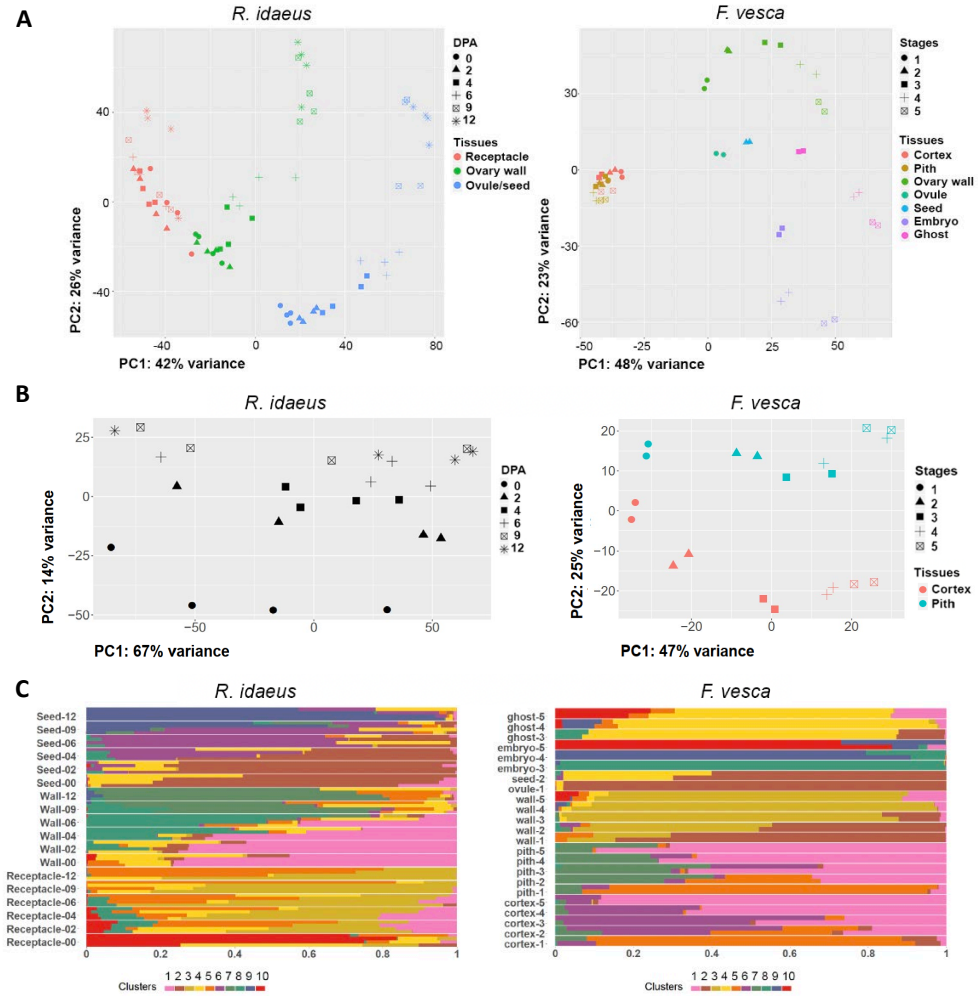

Figure 3.5 Global view of raspberry and strawberry RNA-Seq data from early-stage fruits

(A) PCA plots of red raspberry (*Rubus idaeus*, left) and woodland strawberry (*Fragaria vesca*, right). **(B)** PCA plots of red raspberry (left) and strawberry (right) receptacle. **(C)** Structure plots of raspberry (left) and strawberry (right) generated by grade of membership analyses showing the proportions of gene clusters in each sample.

3.2.5 Identification of tissue- and stage-specific genes

The tissue- and stage-specific genes were identified by TissueEnrich (Jain & Tuteja, 2019) for raspberry and strawberry respectively. Many genes were found to be specifically expressed in the raspberry receptacle at 0 DPA (**Figure 3.6A**). The most significant enriched GO term of this set of genes is 'response to auxin' (**Figure 3.6B**), and a number of *SAUR* (small auxin up-regulated RNA) genes were included in the 128 receptacle- and 0 DPA-specific genes. Therefore, the expression of *SAUR* genes was further examined in raspberry and strawberry (**Figure 3.7A, B**). Compared to strawberry, raspberry has more *SAUR* genes that exhibit dramatically decreased expression after 0 DPA in the receptacle (**Figure 3.7A, B**). The sharply reduced expression of *SAUR* genes may suggest repressed acid-growth in the raspberry receptacle (Stortenbeker & Bemer, 2019).

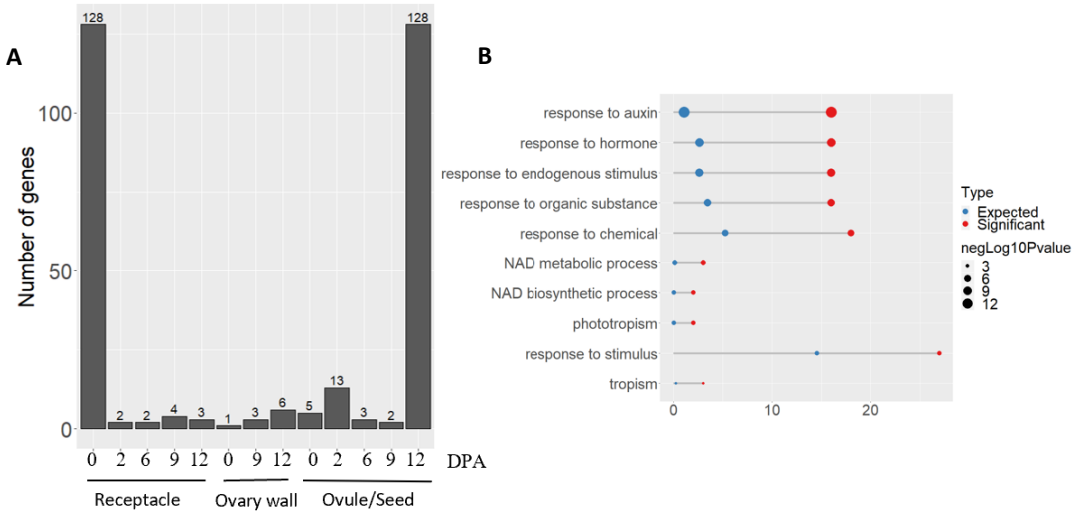

Figure 3.6 The tissue- and stage-specific genes in raspberry

(A) The number of genes that are specifically expressed in a particular tissue at a certain stage. **(B)** The enriched GO terms of the genes that are specifically expressed in the raspberry receptacle at 0 DPA. The number of expected (blue) and observed (red) genes with a specific GO in the gene list was depicted in the lollipop plot. -$\log_{10}p$-value is represented by the size of the dots.

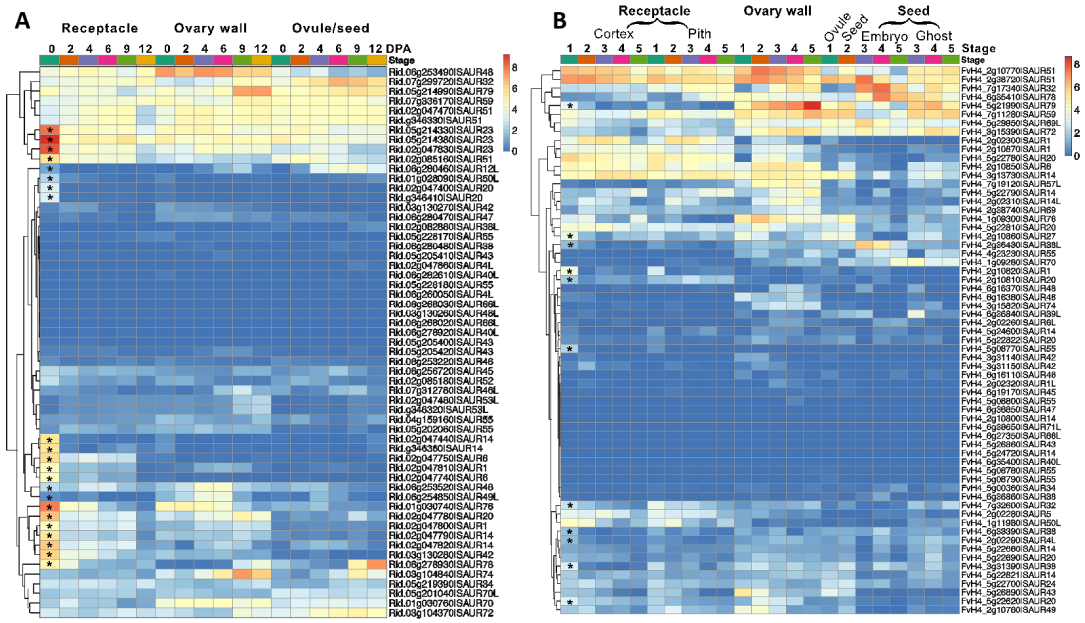

Figure 3.7 Heatmaps showing the expression of *SAUR* genes in raspberry (A) and strawberry (B)

The stars (*) highlight the interesting *SAUR* genes that exhibit reduced expression after 0 DPA in raspberry (A) and strawberry (B). The expression levels are represented by average $\log_2$(TPM+1).

3.2.6 Early lignification of strawberry mesocarp may be a factor that limits its growth

Red raspberry is a drupetum fruit that has fleshy mesocarp and lignified endocarp. The formation of stone, lignified endocarp, is characteristic of drupe fruit (peach) and drupetum (raspberry). Strawberry, an achenetum fruit, however doesn't form stones. Therefore, we examined the expression of the lignin biosynthetic genes and the transcription regulators of lignin biosynthesis genes in raspberry and strawberry (Dardick et al., 2010; Q. Liu et al., 2018; Xie et al., 2018). Most of the lignin biosynthesis-related genes, including multiple *MYB* transcription factors (*MYB46, MYB83, MYB83L, MYB85*), exhibit induced expression from 6 DPA to 9 DPA in the raspberry ovary wall (**Figure 3.8A**), which is consistent with the observation that lignin (red stain) is detected in the raspberry endocarp at 9 DPA and onward (**Figure 3.8C**). And the conserved positive role of *MYB* genes in lignin formation was also supported by the analysis.

Although stone is not formed in strawberry fruit, the lignin biosynthesis-related genes are expressed at a high level starting from stage 1 (0 DPA) in the strawberry ovary wall (**Figure 3.8B**). It agrees with the fact that the lignin deposition (red stain) is observed in the strawberry mesocarp and endocarp from 0 DPA (**Figure 3.8C**). And the lignin deposition spreads out from the tip to the base of the achene over time (**Figure 3.8C**). The findings above suggest that the early lignification of the strawberry ovary wall, especially the mesocarp, may prevent the ovary wall from becoming fruit flesh by constraining its cell division or expansion.

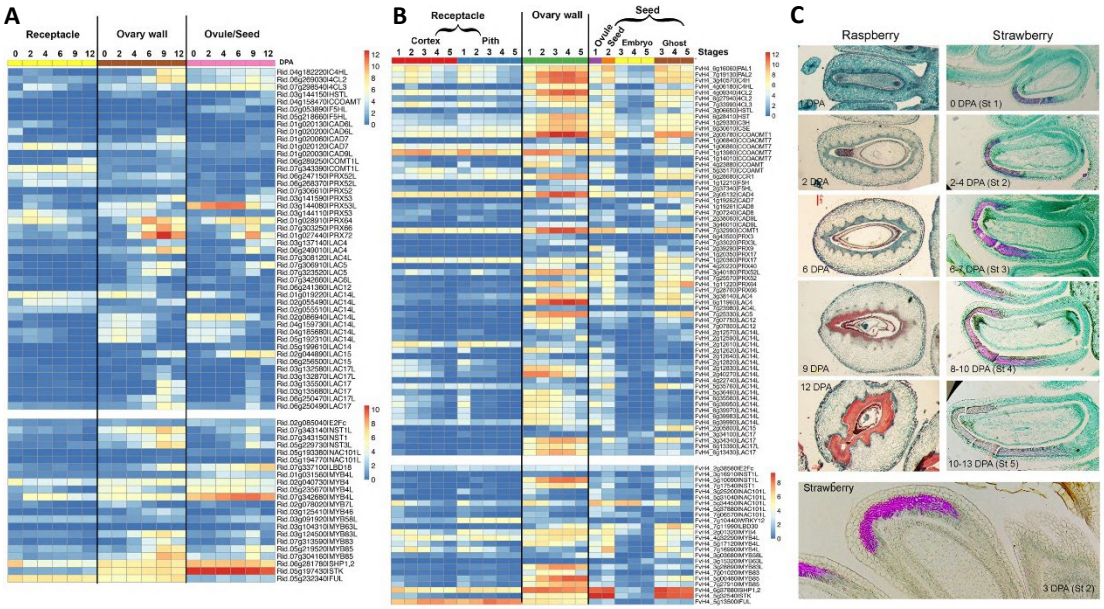

Figure 3.8 Lignin formation in raspberry and strawberry ovary wall

(A, B) Expression of lignin biosynthesis-related genes in raspberry (A) and strawberry (B). The expression levels are represented by average $\log_2$(TPM+1). **(C)** Lignin staining of raspberry and strawberry achene by phloroglucinol-HCL.

3.2.7 *PI*, a type II MADS-box gene, potentially inhibit the development of raspberry receptacle

B class type II MADS-box genes (*PI*, *TM6*, and *AP3*) are known to negatively regulate fruit growth in multiple species, such as apple (Yao et al., 2001, 2018) and tomato (de Martino et

al., 2006; Mazzucato et al., 2008). My comparative analysis in apple and peach also suggests *PI* and *TM6* as repressors of fruit formation in peach hypanthium (Chapter 2; M. Li et al., 2022). Therefore, we identified the B class genes in raspberry and strawberry based on the phylogenetic analysis and then examined the expression patterns of the B class genes in the two species (**Figure 3.9**). In raspberry, the two *RiPI* genes and *RiAP3* gene are persistently highly expressed in the receptacle from 0 DPA to 12 DPA (**Figure 3.9**). And *RiTM6* has shown decreased expression trend after 0 DPA in the receptacle (**Figure 3.9**). Thus, the high expression levels of *PI* and *AP3* genes in the raspberry receptacle make it possible to form a B class heterodimer PI/AP3 and to repress fruit formation. In the ovary wall, the fruit flesh-forming tissue of raspberry, only *RiAP3* is expressed at a high level (**Figure 3.9**). Such a single B class gene expression may be insufficient to inhibit fruit growth in the ovary wall and allow the development of raspberry ovary wall into the fruit tissue. Different from raspberry, strawberry receptacle has either reduced or low levels of B class gene expression, which allows the receptacle to grow into fruit flesh (**Figure 3.9**).

The expression of B class genes in peach and apple is also examined and compared in the newly established eFP browser (Chapter 4). In peach, *PpPI* and *PpTM6* are highly expressed in the hypanthium while *PpAP3* is highly expressed in the ovary wall (**Figure 3.9**). PpPI and PpTM6 may function together to inhibit the hypanthium from forming fruits and indirectly cause the senescence of the peach hypanthium. But the presence of a single B gene *PpAP3* by itself in the peach ovary wall fails to inhibit the growth of the peach ovary wall, therefore the peach ovary wall eventually develops into fruit flesh. Moreover, in apple, the expression of all four B class genes is reduced after 0 DPA in the hypanthium, which may explain why peach hypanthium senesces but apple hypanthium becomes enlarged and fleshy during fruit development (**Figure 3.9**). The comparative analysis of B class gene expression across all

four Rosaceae fruit types supports a negative regulatory role of B class heterodimers for fruit development and indicates B class MADS-box genes key regulators of fruit type.

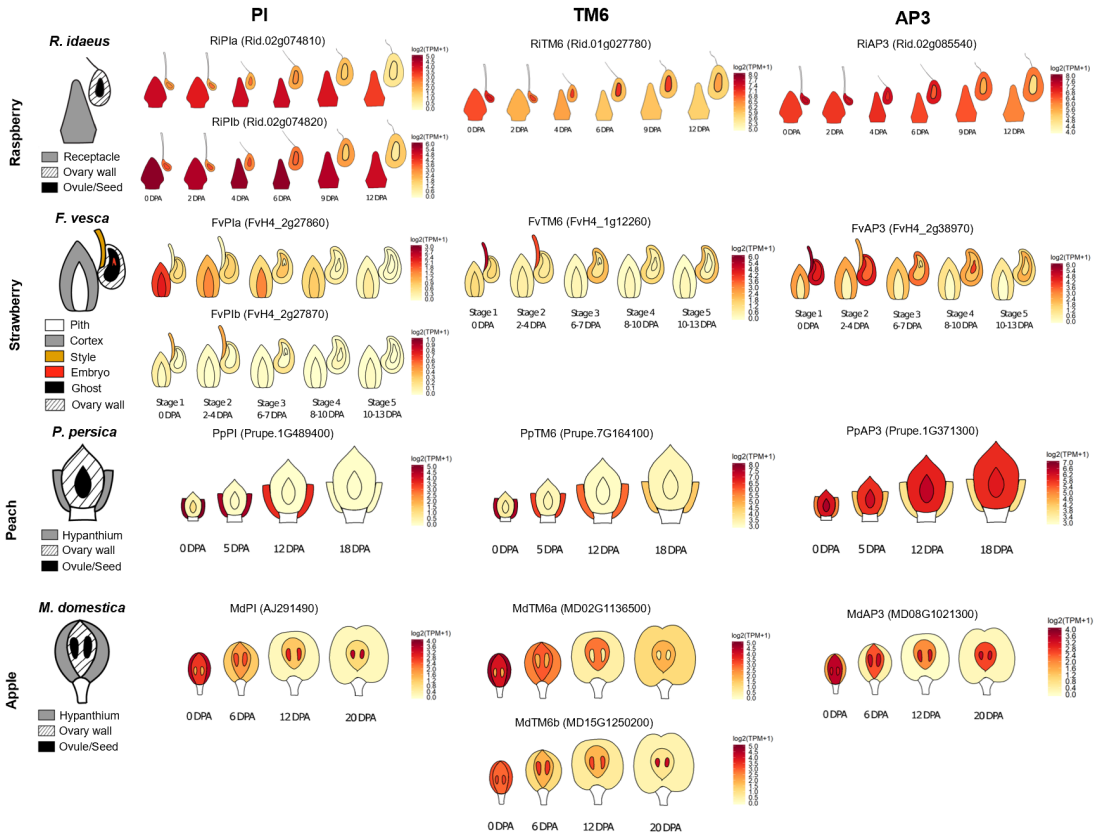

Figure 3.9 The expression patterns of the B class genes in raspberry, strawberry, peach, and apple (top to bottom) shown in the eFP (electronic Fluorescent Pictograph) browser

The R package ggefp (https://github.com/hredestig/ggefp) was utilized to visualize the expression of the B class genes in distinct tissues at different stages.

2.4 Discussion

This is a collaborative project with experimental contributions by Dr. Junhui Zhou. Specifically, Dr. Junhui Zhou performed raspberry fruit dissection, harvests, and morphological characterizations (**Figure 3.1**) as well as tissue sectioning and staining (**Figure 3.1C and 3.8C**). Furthermore, Dr. Junhui Zhou treated the raspberry fruits with different plant hormones to show that both auxin and GA were required to induce

parthenocarpic fruits in raspberry (**Figure 3.2**), which emphasizes the irreplaceable role of auxin and GA in raspberry fruit development.

Additionally, high-quality genome assembly and annotation of *Rubus idaeus* 'Joan J.' were generated in the study. The initial genome assembly and annotation were constructed by Dr. Yongping Li. And the further genome improvement and annotation lift over were conducted by me. These new genomic resources of Rubus idaeus 'Joan J.' enable me to perform the downstream analyses, including the functional annotation, ortholog identification, phylogenic analysis of type II MADS-box genes, RNA-Seq analyses and so on.

In raspberry, the middle layer of the ovary wall (mesocarp) becomes enlarged and fleshy, and the inner most layer of the ovary wall (endocarp) lignifies and forms the stone. But in strawberry, the receptacle develops into the fruit flesh, and the achenes dry out on the fruit surface. Our study has uncovered a new mechanism that explains why strawberry achenes could not develop into fruit due to the lignification of the ovary mesocarp. Based on our transcriptomic comparisons between raspberry and strawberry early developing fruits, a model was proposed to explain why raspberry and strawberry fruits are derived from distinct flower tissues (**Figure 3.10**). Firstly, B class genes (*PI* and *AP3*) were continuously highly expressed in the raspberry receptacle, which suggests that PI and AP3 may work together to inhibit the growth of the receptacle in raspberry. But B class genes show low expression in the strawberry receptacle, which may allow the strawberry receptacle to become fleshy fruit. Secondly, the *SAUR* genes (auxin responsive genes) show more obvious decreased expression in raspberry receptacle after 0 DPA. Because *SAUR* genes are correlated with cell elongation (Stortenbeker & Bemer, 2019), the reduced expression of the *SAUR* genes may lead to the inhibited growth of the raspberry receptacle. Lastly, the lignin biosynthesis-related

genes are expressed in the raspberry ovary wall at 6-9 DPA, and the stone formation is observed in the raspberry endocarp from 9 DPA. However, in strawberry, the lignin biosynthesis-related genes are highly expressed in the ovary wall at very beginning of the fruit development (0 DPA). And the lignin deposition in the mesocarp and endocarp of strawberry starting at 0 DPA and continuously onward may have prevented it from growing into fruit flesh.

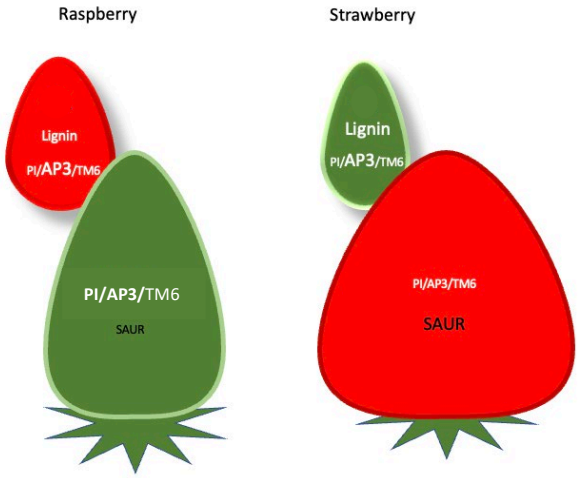

Figure 3.10 A proposed model showing the action of identified genetic factors that may explain the distinct fruit types in raspberry and strawberry

The red and green colors highlight the fleshy and non-fleshy parts of the fruit, respectively. Larger font size represents higher expression. White texts indicate negative regulation while black texts indicate positive regulation.

Chapter 4: Rosaceae Fruit Transcriptome Database (ROFT) – a useful genomic resource for four economically important fruits, apple, peach, strawberry, and raspberry

4.1 Introduction

Rosaceae is a large plant family consisting of over 3000 species. Many economically important fruits belong to the family including peach, apple, strawberry, raspberry, and others. *Rosaceae* family plants also form a large number of diverse fruit types including four main fleshy fruit types including drupe (peach), pome (apple), drupetum (raspberry), and achenetum (strawberry). Hence Rosaceae is an ideal family for investigations of molecular mechanisms underlying the diversity of fruit type (Z. Liu et al., 2020; Xiang et al., 2017). In drupe fruit, such as peach, the middle layer of the ovary wall (mesocarp) grows into fruit flesh, and the innermost layer of the ovary wall (endocarp) forms a hard stone encasing a single seed. In pome fruit, for example, apple, the fruit flesh is derived from the hypanthium, a cup-like structure outside the ovary wall. In drupetum and achenetum fruits, there are numerous individual carpels crowding the dome-like receptacle. In drupetum fruit like raspberry, each carpel turns into a fleshy drupelet. However, in achenetum fruit, namely strawberry, the receptacle becomes fleshy while the carpels dry up to become the achenes dotting the fruit surface.

Early fruit development is an important stage, during which fruit tissue identity is determined and subsequently becomes enlarged. This early stage is usually composed of three phases, fruit set, cell division, and cell expansion (Gillaspy et al., 1993). In the first phase, successful pollination and fertilization trigger the development of a specific floral tissue to become a fruit. In the second phase, cell division is activated and fruit starts to enlarge. In the third phase, fruit size increases rapidly as a result of cell expansion and cell wall loosening.

With the development of advanced high-throughput sequencing technologies, such as Illumina, PacBio and Oxford Nanopore sequencing, more and more high-quality sequencing data with deep depth become available for many Rosaceae species, which gives rise to a dramatic increase in the number of chromosome-scale genome assemblies and comprehensive genome annotations (M. Li et al., 2021). The abundant genomic resources make it possible to investigate plant growth globally and thoroughly, and thus help stimulate novel ideas for molecular research.

The genomic data explosion also contributes to a greater number of databases established for economically important Rosaceae species (M. Li et al., 2021). For instance, the Genome Database for Rosaceae (GDR) (S. Jung et al., 2019) contains a large collection of Rosaceae genomes and integrates various tools for genetic and genomic analyses. *Fragaria vesca* co-expression network explorer (http://159.203.72.198:3838/fvesca) (Shahan et al., 2018) offers the consensus networks constructed to predict gene-gene relationships in strawberry. Furthermore, TRANSNAP (Koshimizu et al., 2019) and strawberry eFP browser (Hawkins et al., 2017) (http://bar.utoronto.ca/efp_strawberry/cgi-bin/efpWeb.cgi) are developed for exploring the pear and strawberry transcriptome data, respectively. Although there is a variety of databases presenting different types of data from various Rosaceae species, few databases are built for comparative work among different Rosaceae species.

In this study, we developed a comparative transcriptome database, ROFT, using the published RNA-Seq data from four early Rosaceae fruits, peach (*Prunus persica*) (M. Li et al., 2022), apple (*Malus x domestica*) (M. Li et al., 2022), strawberry (*Fragaria vesca*) (Kang et al., 2013), and raspberry (*Rubus Idaeus*) (Chapter 3). Each of them represents one of the four main fruit types. The comparative eFP browser in the database allows the users to compare gene expression patterns between orthologs in different species harboring the four fruit types.

In addition, the co-expressed genes and tissue-specific genes are provided for each of the four Rosaceae species. As a result, the common and distinct molecular features among the four fruit types can be identified, which may lead to a better understanding of early-stage fruit development and potential regulatory mechanisms underlying diverse fruit types.

4.2 Material and methods

Data sources and processing

The RNA-Seq data of peach and apple early fruit development were generated from three specific fruit tissues (hypanthium, ovary wall and ovule/seed) at four critical stages (0 DPA, 5/6 DPA, 12 DPA, and 18/20 DPA) in triplicates. Tissue morphology, staging, and the RNA-Seq data with three replicates were described previously (M. Li et al., 2022). The data was deposited at SRA with the accession number PRJNA661345.

The strawberry early fruit development was divided into five stages, stage 1 (0 DPA), stage 2 (2-4 DPA), stage 3 (6-7 DPA), stage 4 (8-10 DPA), and stage 5 (10-13 DPA) (Hollender et al., 2012). The strawberry fruit tissues including style, pith, cortex, ovary wall, and ovule/seed (ghost and embryo) were dissected and harvested at their corresponding stages (Kang et al., 2013). Two biological replicates were prepared for RNA-Seq, which was deposited at SRA with accession number PRJNA187983.

Three fruit tissues (receptacle, ovary wall, and ovule/seed) were dissected at six early stages of red raspberry fruit development (0 DPA, 2 DPA, 4 DPA, 6 DPA, 9 DPA, and 12 DPA) (Chapter 3). And the RNA-Seq data were collected from four biological replicates.

Cutadapt (v2.8) (M. Martin, 2011) was used to trim the low-qualify bases (cutoff: 25) from the 3' end of the red raspberry reads. Only the reads with a minimum length of 36 bp are remained for the downstream analyses.

Table A.3 (see Appendices) summarizes the information of all the samples used for RNA-Seq analyses.

Salmon (v0.11.2) (Patro et al., 2017) was applied to quantify the transcript expression for the four Rosaceae species. The peach, apple, and strawberry reference transcripts were retrieved from GDR (*Prunus persica* Genome v2.0.a1, *Malus x domestica* GDDH13 Whole Genome v1.1, and *Fragaria vesca* Genome v4.0.a2) (Daccord et al., 2017; Edger et al., 2018; Y. Li et al., 2019; Verde et al., 2017). For index construction, k-mer size was set to 31, and --keepDuplicates was specified to keep identical sequences in the reference transcripts. And --seqBias was passed to the quantifier to correct the sequence-specific bias. Tximport (v1.10.1) (Soneson et al., 2016) was further utilized to summarize the transcript abundance into gene level.

Ortholog detection

BLAST (v2.5.0) (Altschul et al., 1990) was employed to search the longest protein isoforms of the four Rosaceae species against Arabidopsis protein database generated using the longest peptides in TAIR10_pep_20101214 (https://www.arabidopsis.org/download_files/Proteins/TAIR10_protein_lists/TAIR10_pep_20101214). The protein sequences of the four Rosaceae species were derived from the same genome annotations used for transcript quantification. Only the best Arabidopsis BLAST hits with *E*-value less than 10^{-5} were presented in the database.

The genes in peach, apple, strawberry, red raspberry, black raspberry (GDR: *Rubus occidentalis* whole genome assembly v3.0) (VanBuren et al., 2018), and Arabidopsis were assigned to different orthogroups by OrthoFinder (v2.3.8) (Emms & Kelly, 2015). The longest protein isoforms from each species were fed into OrthoFinder for the ortholog identification.

Development of comparative eFP browser

The R package ggefp (https://github.com/hredestig/ggefp) was used to visualized the gene expression data ($\log_2$(TPM+1)) of the four Rosaceae species. The fruit diagrams shown in the comparative eFP browser was first created by Adobe Illustrator. The line drawings (PS format) were further transformed to ggproto objects by ggplot2 (Wickham, 2016) and saved in RDA files that would be used by ggefp as exhibits for data visualization (https://github.com/hredestig/ggefp/blob/master/etc/trace.R).

Consensus network construction

A robust signed co-expression network was constructed using the consensus clustering approach for each Rosaceae species (Shahan et al., 2018). WGCNA (peach and apple: v1.68, strawberry and red raspberry: v1.70.3) (Langfelder & Horvath, 2008) was run 1000 times with subsampled genes and randomized parameters. The gene expression was presented in $\log_2$(TPM+1). The genes with little variance (≤ 0.05) and zero median absolute deviation were excluded. And biweight midcorrelation was chosen to measure the similarity between the expression values of each pair of genes. A weighted adjacency matrix was produced by dividing the number of times genes were clustered together by the number of times genes were subsampled together. The ultimate consensus network was built based on the adjacency matrix using WGCNA with power 6 and minModuleSize 10.

The GO terms associated with peach and apple genes were obtained from GDR (Prunus_persica_v2.0.a1_gene_functions.txt and Malus_x_domestica_GDDH13_v1.1_interpro.txt) while the GO annotations of strawberry and red raspberry were conducted by OmicsBox (v1.2.4) (Götz et al., 2008). The R package topGO (v2.34.0) (Alexa & Rahnenfuhrer, 2022) was applied to perform the GO enrichment analysis for each gene module. The significance of the enriched GO categories was determined by Fisher's exact test. The *p*-value threshold was 0.05.

Identification of tissue-specific genes

A combination of a tissue and a stage was regarded as a condition. Average TPM values of biological replicates were calculated to show the gene expression in each tissue at each stage (at each condition), which were later used by TissueEnrich (v1.2.1) (Jain & Tuteja, 2019) to identify the tissue- and stage-specific genes (condition-enriched genes), as well as the tissue-specific genes (multiple-condition-enriched, but within the same tissue). A gene is considered to be a tissue- and stage-enriched gene if its TPM is no less than one, and its expression at a particular condition is at least two/five-fold higher than that at other conditions. Moreover, a gene is considered as a tissue-specific gene if its TPM is no less than one, its expression at a group of conditions (within the same tissue) is at least two/five-fold higher than that at other conditions, and the gene is not counted as a tissue- and stage-specific gene.

Database implementation

The R package mongolite (v2.1.0) (Ooms, 2014) was used to import the data into or fetch the data from MongoDB, a NoSQL database platform. And the user interface of the database was built by the R package Shiny (v1.3.2) (Chang et al., 2019). An undergraduate student, Andrew Tong, helped to implement hyperlinks using R Shiny. The system hosting the

database was Ubuntu (v18.04.5). And the BLAST tool (v2.5.0) (Altschul et al., 1990) embedded in the database was installed locally.

4.3 Usage and access

ROsaceae **F**ruit **T**ranscriptome database (ROFT) mainly consists of eight functional modules, Gene, Comparative eFP Browser, Co-expression Network, Tissue-specific Genes, BLAST, Retrieve Data, Download, and Help (**Figure 4.1**). They are described in each of the section below.

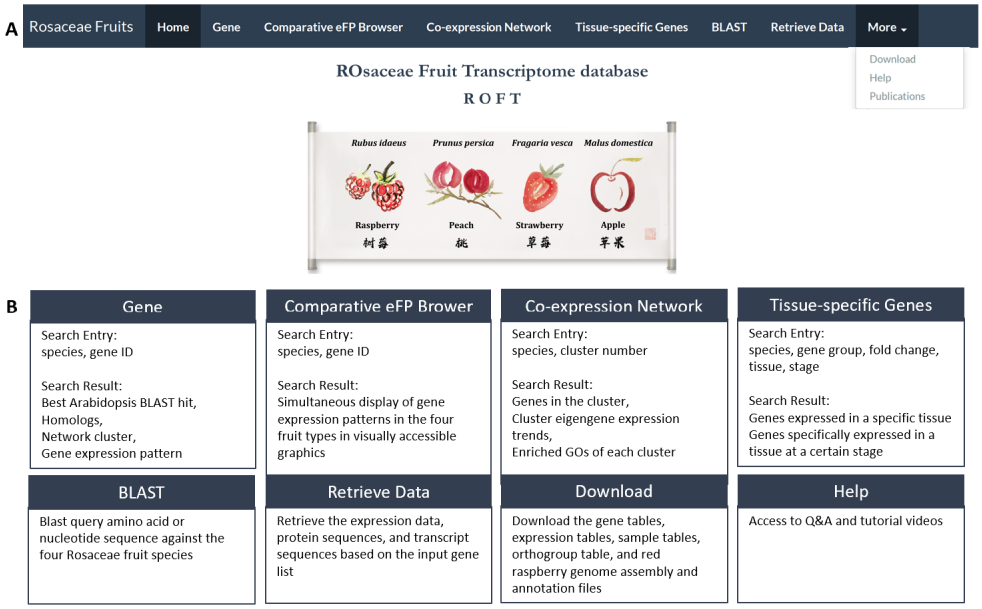

Figure 4.1 Summary of the ROFT Database

(A) Home page with tabs. **(B)** Summary of the key functions provided by each of the eight tabs.

4.3.1 Gene

In the 'Gene' section, users can search for information on a gene from any of the four Rosaceae species, peach, apple, raspberry and strawberry, by entering in the search box a

specific gene ID (**Figure 4.2A**). If such a gene ID is not known, one can use nucleotide or protein sequence of the gene to blast using this database (ROFT)'s BLAST function, or blast in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or GDR (https://www.rosaceae.org/blast) to obtain the gene ID of the species of interest. The search returns results including the gene's best BLAST hit in Arabidopsis (Arabidopsis gene ID, gene symbol, and gene description) (**Figure 4.2C**). In addition, the search will return information on the query's orthologs (specific gene IDs). We use orthologs here to refer to those genes belonging to the same orthogroup identified by OrthoFinder (Emms & Kelly, 2015). If genes from different species originated from the same gene in the last common ancestor of those species, the genes will be grouped into one orthogroup (Emms & Kelly, 2015).

In addition, the search will return the specific consensus co-expression cluster the query belongs to, and its expression values in different fruit-related tissues at different stages as shown in a box plot (**Figure 4.2C**). Further, hyperlinks are provided for each ortholog ID and co-expression cluster so that users can learn more about them by simply clicking the links (**Figure 4.2C, D**).

If one starts by simply asking how an ortholog of an Arabidopsis gene functions in any of the four Rosaceae species, one can directly enter in the search box the Arabidopsis gene ID (**Figure 4.2B**), which can be obtained by searching TAIR (https://www.arabidopsis.org/). The search in ROFT using the Arabidopsis gene ID will return its Rosaceae orthologs (**Figure 4.2B**). And clicking one of the ortholog IDs will lead the users to the 'Gene' page of the Rosaceae ortholog (**Figure 4.2 B, C**).

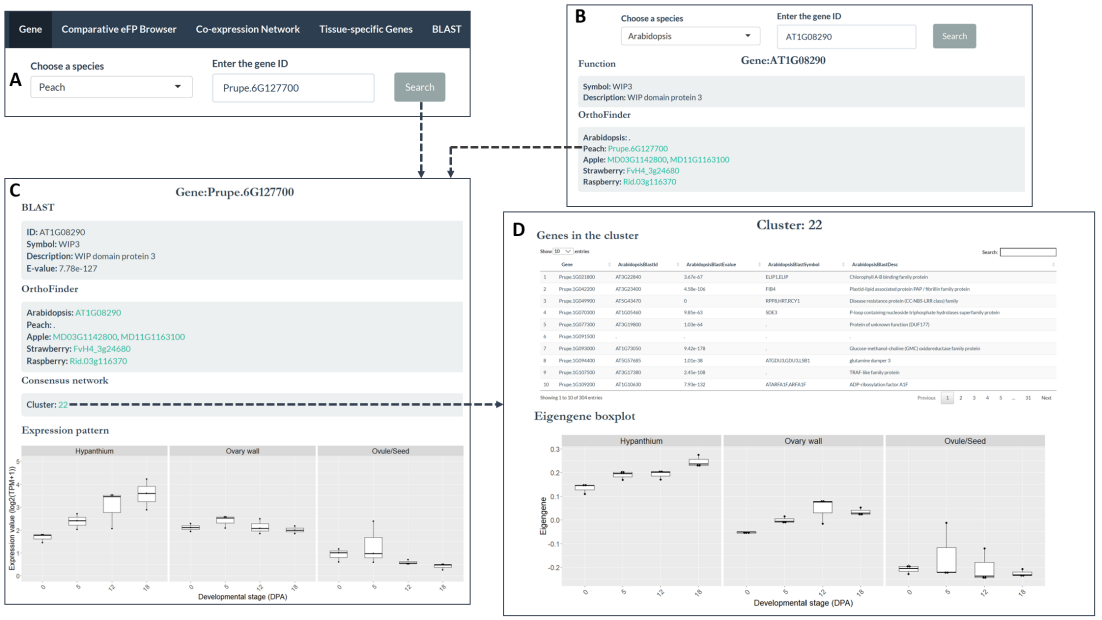

Figure 4.2 Illustration of the information derived by searching in the Gene tab

(A) Searching using the peach gene ID. **(B)** Searching using the Arabidopsis gene ID. *Prupe.6G127700* is the peach ortholog of the Arabidopsis gene *AT1G08290*. Clicking 'Prupe.6G127700' in B will load the 'Gene' page of the peach ortholog in C. **(C)** The result from searching the peach gene ID. The information about the peach gene *Prupe.6G127700* is shown, including its Arabidopsis, apple, strawberry, and raspberry orthologs, the consensus network cluster this gene belongs to, and the expression pattern in the three tissues, hypanthium, ovary wall, and ovule/seed at four early fruit developmental stages (days post anthesis, DPA). **(D)** Information about the Cluster 22 of the consensus network by clicking '22' in C. It includes information about all 304 genes in this cluster as well as the eigengene expression trend of the cluster. The top 20 enriched GO terms for the cluster are also provided in the database but not shown here. The undergraduate student, Andrew Tong, helped to create the hyperlinks in the ROFT database.

4.3.2 Comparative eFP (electronic Fluorescent Pictograph) Browser

The 'Comparative eFP Browser' section displays the expression patterns of all orthologs of the input gene in apple, peach, strawberry, and raspberry. If there are more than one ortholog, as is often the case for apple, The users can easily visualize and compare the gene expression patterns in similar and unique fruit tissues in the four Rosaceae species (**Figure 4.3**). Users enter the gene ID of one of the species in the search box, and the search returns not only the expression pattern of the input gene but also expression patterns of all orthologs in the other three species. This comparative expression visualization makes it easier to compare and

analyze gene expression differences between species and enable users to better explain and illustrate the expression patterns and formulate novel hypothesis. A graphic key about the fruit tissues is provided at the top of the page (**Figure 3**) and **Table A.3** (see Appendices) provides details about the tissue samples from which RNAs were extracted; the information provides users accurate information about the specific tissues (**Figure 4.3**).

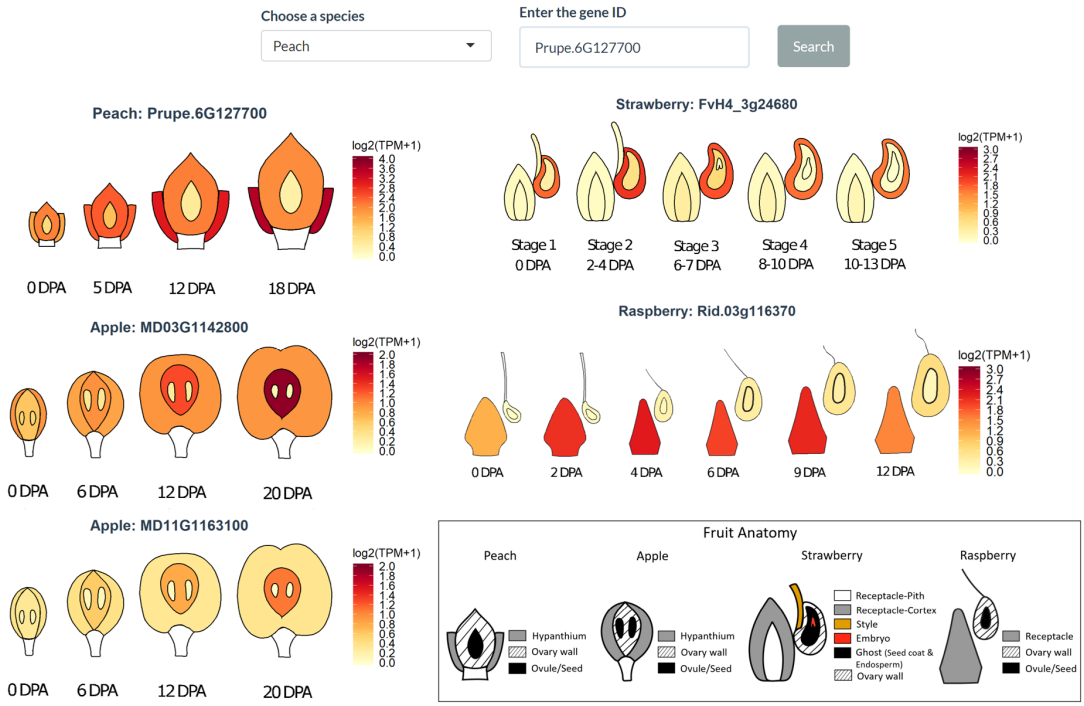

Figure 4.3 Comparative eFP browser showing the expression patterns of *WIP3* in all four Rosaceae species

In the search box, one enters the gene ID of a specific gene in one of the four Rosaceae species. The peach *WIP3* gene (gene ID *Prupe.6G127700*) is entered here as a search example. The search result is displayed as electronic Fluorescent Pictographs (eFPs) showing gene expression levels of orthologous genes in peach, apple, strawberry, and raspberry fruits. There are two apple orthologs, and their expressions are both displayed. The scale bars are RNA-seq read levels expressed as $\log_2$(TPM+1). The keys to the fruit tissue types are drawn in the box and specific fruit developmental stages (DPA) are shown beneath each stage. Sample description is detailed in Table A.3 (see Appendices).

4.3.3 Co-expression Network

The ‘Co-expression Network’ section includes two sub-tabs, ‘Summary’ and ‘Network’. The ‘Summary’ page provides basic statistics of the consensus co-expression networks, the

number of clusters in each species, and brief methods of constructing the networks (**Figure 4.4A**). Additionally, each species' network clusters are presented as heatmaps showing the eigengene value of each cluster (**Figure 4.4B**). This provides users an overview of each cluster's expression trend, based on which the users could choose a cluster of a specific expression pattern for further exploration.

In the 'Network' page, one can search the specific cluster number, it will return with a list of genes (gene ID, Arabidopsis ortholog, ortholog description) in the cluster. It will also provide the cluster eigengene boxplot showing the cluster eigengene expression level in different samples (**Figure 4.5**), as well as the top 20 enriched GOs in the cluster (**Figure 4.5**).

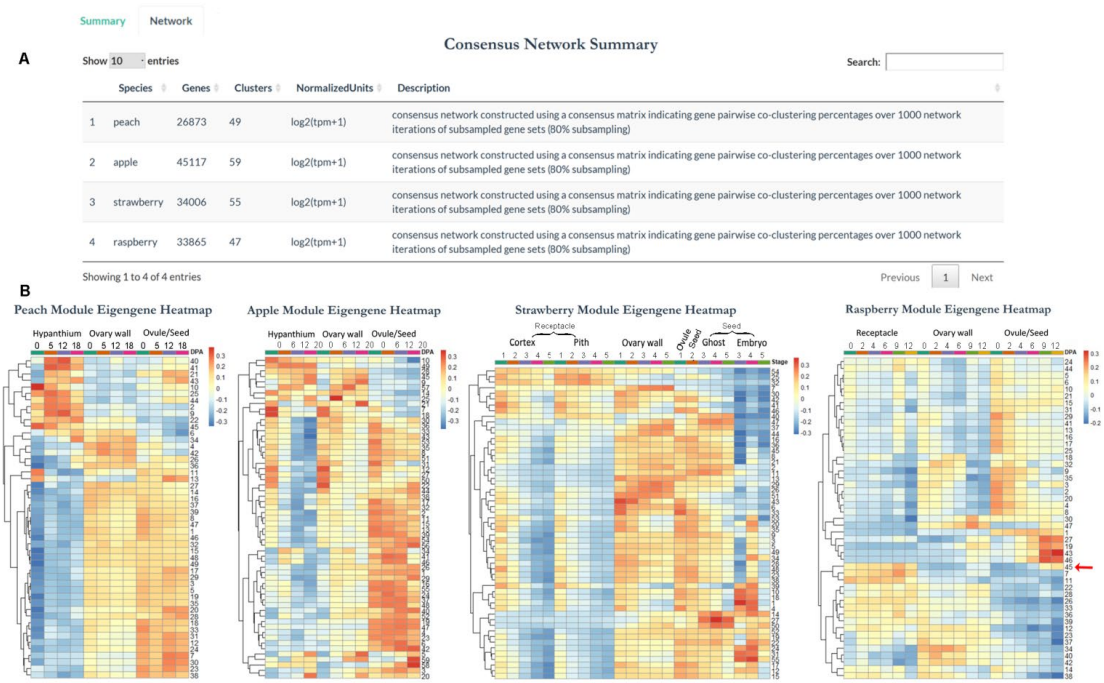

Figure 4.4 Illustration of the 'Summary' tab on the consensus co-expression network information

(A) Summary of consensus co-expression network analysis result showing the number of clusters in each species. **(B)** Heatmap of cluster eigengene which provides the general expression trend of each cluster. Red arrow points to cluster 45 of raspberry.

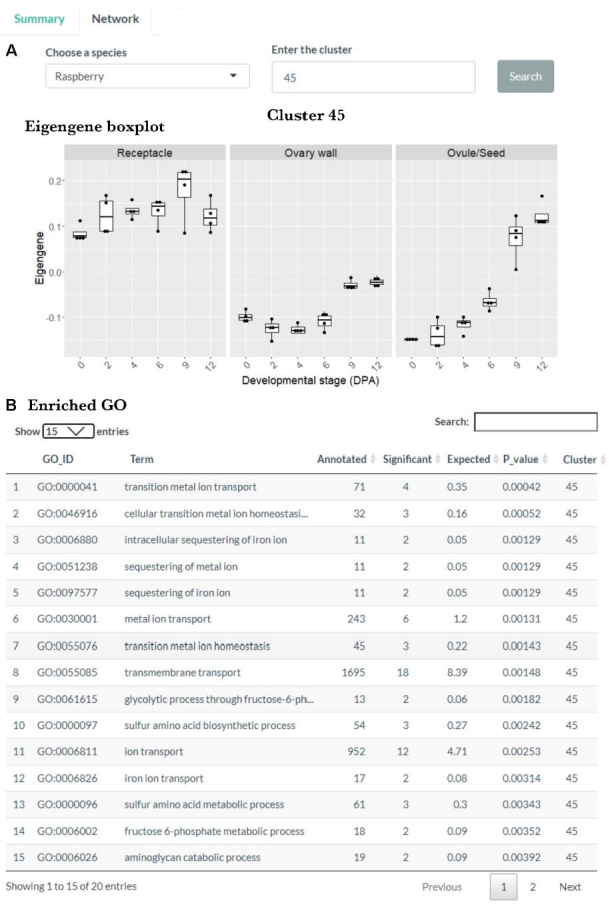

Figure 4.5 Illustration of the 'Network' tab on the consensus 'Co-expression Network' information

(A) By searching for cluster 45 in the search box, one can obtain information on eigengene expression pattern of cluster 45 shown as boxplots. **(B)** Enriched GO terms for cluster 45 are also listed. Genes in the cluster can be obtained as well (not shown here).

4.3.4 Tissue-specific Genes

The 'Tissue-specific Genes' section, available for all four species, allows one to search for genes that are specifically expressed in a tissue at a certain stage. To do that, users will first select "Tissue&Stage-Enriched" under Gene Group, and then select a specific tissue under "Tissue" and a specific stage under "Stage". In the example shown (Figure 6A), "Ghost" (seedcoat and endosperm) and "stage 3" (6-7 DPA) are entered in the selection box. In addition, users are given an option of selecting 2-fold or 5-fold enrichment under "Fold Change" (**Figure 4.6A**). The example shown chooses 2-fold that is less stringent than the 5-

fold and hence yields more genes. The search resulted in a list of 314 genes that are enriched by 2-fold in the ghost tissue at stage 3, which is a stage soon after fertilization (**Figure 4.6B**). Several Type I MADS-box genes including *AGL62* and *AGL80* are among the list of genes. A download button is at the bottom of the gene list table (**Figure 4.6B**) that allows users to obtain the entire list of 314 genes (**Figure 4.6B**). Alternatively, if one is only interested in tissue-specific genes, one can search for tissue-enriched genes (that can be enriched at two or more stages in the same tissue) by choosing "Tissue-Enriched" under the Gene Group (**Figure 4.6A**).

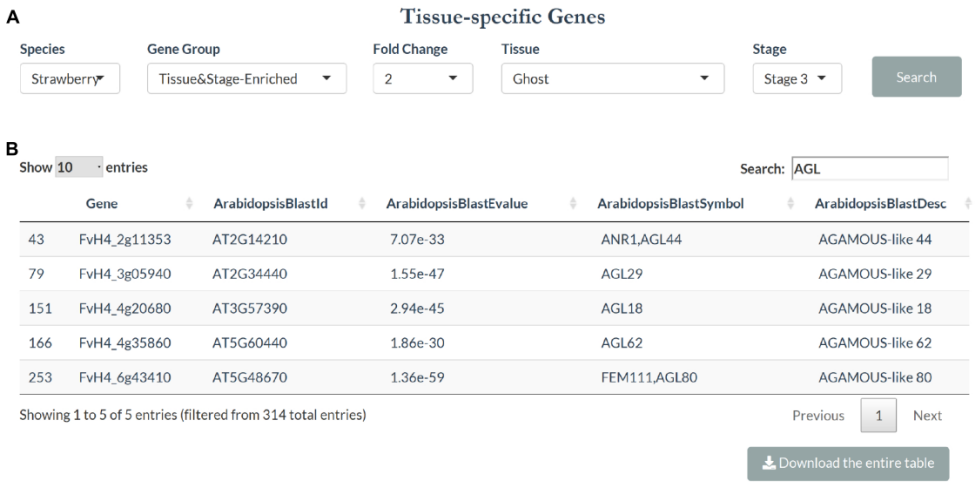

Figure 4.6 Demonstration of the Tissue-Specific Genes function

(A) The example search is in strawberry for genes specifically expressed in the 'Ghost (seed coat and endosperm)' at stage 3 with a minimum of 2-fold enrichment. **(B)** The enriched gene list, in strawberry ghost tissue at stage 3, includes several Type I MADS-box genes shown. Note the download button beneath the table.

4.3.5 BLAST

A BLAST function is included in the ROFT database, which allows the users to identify the Rosaceae orthologs of the input sequence. Four BLAST programs, BLASTP, BLASTN, BLASTX, and TBLASTN, are available in the 'BLAST' section of ROFT.

4.3.6 Retrieve Data and Download

The RNA-Seq expression data (in TPM) in different tissues and stages, and the transcript/protein sequences of genes of interest can be retrieved by providing the gene IDs in the 'Retrieve Data' section. In the 'Download' section, users can download the gene tables, expression tables, and sample tables (**Table A.3**, see Appendices) for each *Rosaceae* species. The gene tables consist of the results from BLAST search, ortholog identification in the four species as well as their corresponding Arabidopsis ortholog, and consensus network analysis. The expression tables contain the gene expression levels (TPM) across different floral/fruit tissues during fruit development. And the sample tables (**Table A.3**, see Appendices) describe the details of the samples and stages from which the RNAs were isolated. An orthogroup table that purely summarizes the orthologs among the Rosaceae species in comparison to Arabidopsis is also accessible in the 'Download' section. In addition, users can obtain the red raspberry genome assembly and annotation files.

4.3.7 Help

A 'Help' page is provided to assist the users to take full advantage of the database. The 'Help' section is composed of two components, 'Q&A' and 'How-To Videos'. The 'Q&A' mainly answers anticipated technical questions while the 'How-To Videos' part is a screen recording tutorial video to demonstrate how one can explore various functions of the database.

4.4 Case studies

Following are several case studies to demonstrates the biological insights gained from mining the ROFT database.

4.4.1 *WIP3*, a likely conserved repressor of fleshy fruit development in Rosaceae species

WIP proteins are transcription factors that are known to inhibit plant organ development in different species including *Cucumis melo* (A. Martin et al., 2009), *Gerbera hybrida* (Ren et al., 2018), and Arabidopsis (Roldan et al., 2020). The expression of the *WIP* genes were examined in the four *Rosaceae* species by mining the ROFT. First, the Rosaceae *WIP3* genes were identified by searching in the 'Gene' section for the Arabidopsis *WIP3* gene (*AT1G08290*) (**Figure 4.2B**). The search result lists the *WIP3* orthologs with gene ID in Arabidopsis, peach, apple, strawberry, and raspberry under OrthoFinder (**Figure 4.2B**). It shows two *WIP3* orthologs in apple, and one *WIP3* each in strawberry, raspberry, and peach respectively.

Second, using the comparative eFP browser in ROFT, we discovered that the *WIP3* genes show extremely interesting expression patterns in the fruits of all four species. Specifically, *WIP3* appears always expressed at a low level in the fruit tissue and a higher level in the non-fruit tissues (**Figure 4.3B**). For example, the *WIP3* gene is relatively highly expressed in peach hypanthium, apple ovary wall, red raspberry receptacle, and strawberry ovary wall, tissues that are not forming fruit. Therefore, the *WIP3* gene could encode a potentially conserved repressor that inhibits floral tissue from becoming fruit tissues.

4.4.2 Metal ion transport appears to be very active in red raspberry receptacle and post-fertilization seed

Previously, we showed that iron can travel from the receptacle to the ghost after fertilization in strawberry (Shahan et al., 2018). The iron transported to the ghost may serve as the cofactor for GA biosynthetic enzymes, GA20ox and GA3ox, which lead to GA synthesis required for strawberry receptacle fruit enlargement (Kang et al., 2013). Therefore, we

explored the red raspberry consensus network in ROFT to determine if such iron transport activity may be conserved in the red raspberry. First, through the Co-expression Network's 'Summary' page of ROFT, we identified the cluster 45 of raspberry that exhibits receptacle enriched expression as well as fertilization-induced expression in seeds (**Figure 4.4B**). Further exploration of cluster 45 in the 'Network' page revealed that the top-ranking enriched GO terms in cluster 45 are associated with metal ion transport and homeostasis (**Figure 4.5B**). Hence, similar to strawberry, red raspberry receptacle also appears to experience active iron transport.

4.4.3 Multiple type I MADS-box transcription factors encoded by *AGL*s (*AGAMOUS-LIKE*) may regulate strawberry seed development immediately post-fertilization

Each strawberry seed could be manually dissected and separated into embryo and ghost (seed coat and endosperm) starting from stage 3, which allows for the identification of the ghost-specific genes involved in seed development. To investigate what genes might be induced by fertilization in the endosperm-an important tissue for auxin and GA synthesis to stimulate post-fertilization programs, we mined the ROFT database in the 'Tissue-specific Genes' section. 314 genes were identified that fulfill the search criteria: Species- Strawberry, Gene Group-Tissue&Stage-Enriched, Fold Change-2, Tissue-Ghost, Stage-Stage 3 (**Figure 4.6**). Among the 314 genes expressed in the stage 3 ghost are five type I MADS-box genes including *AGL44* (*FvH4_2g11353*), *AGL29* (*FvH4_3g05940*), *AGL18* (*FvH4_4g20680*), *AGL62* (*FvH4_4g35860*), and *AGL80* (*FvH4_6g43410*) (**Figure 4.6B**). The functions of these genes in early seed development are being tested. Interestingly, an interaction between AGL62 (FvH4_2g03030) and AGL80 (FvH4_6g08460) was shown in strawberry, and their function in promoting auxin biosynthesis in the strawberry seed was also validated (Lei et al.,

2022). Therefore, the ROFT database provides an important bioinformatic resource for users to identify candidate genes and formulate hypotheses for future functional tests.

Chapter 5: Conclusions and future directions

By pairwise comparison of the RNA-Seq data between peach and apple and between strawberry and raspberry during early-stage fruit development, we identified a number of genetic factors and potential mechanisms that may explain different fruit types in the Rosaceae family. First, *PI,* a B class type II MADS-box gene, is highly expressed in non-fruit forming tissues, such as peach hypanthium and raspberry receptacle, but exhibits reduced or low expression in fruit-forming tissues, such as apple hypanthium and ovary wall and strawberry receptacle. As a result, *PI* functions as a potential conserved repressor of fruit development across the four Rosaceae species. Since PI functions by dimerizing with other B class proteins (AP3 and TM6), these other B class genes may also participate in inhibiting fruit set and the fruit growth as dimerization partners of PI. *FBP9*, a E class type II MADS-box gene, is expressed at a high level in fruit tissues that are capable of becoming fleshy. Additionally, a survey of eudicot genomes of 45 species available in the PLASA Dicot 4.0 database shows that the *FBP9* gene is presents in all but one fleshy fruit forming eudicot species, but is absent in a larger number of dry fruited eudicot species, suggesting that *FBP9* is likely to be necessary for fleshy fruit formation. Since *FBP9* is present in some of the dry fruit eudicot species, its mere presence is necessary but insufficient for fleshy fruit formation.

Besides the type II MADS-box genes, other genetic factors that relate to fruit type specification have also been identified by our comparative transcriptomic analyses. For example, *GA2OX* genes, encoding the GA degradation enzyme, are preferentially expressed in the peach hypanthium, which is likely to reduce the GA accumulation in the peach hypanthium and thus limits the growth of the hypanthium. Moreover, the induced expression of developmental programmed cell death (dPCD) genes in the peach hypanthium after

pollination may also contribute to the senescence of the peach hypanthium. In raspberry, the dramatically decreased expression of the *SAUR* genes (auxin responsive genes) in raspberry receptacle after 0 DPA may constrain the cell expansion in raspberry receptacle and inhibit the receptacle growth. Additionally, the early lignification of strawberry mesocarp and endocarp, the middle and innermost layers of the ovary wall, potentially restricts the enlargement of the strawberry ovary wall. Together these analysis results provide the direction for future functional tests through CRISPP gene editing or transgenic over-expression of the genetic factors identified.

Most of our studies are based on gene-level analyses. But alternative splicing (AS), a key step of post-transcriptional control, has also been observed in fruit development. In tomato, it has been uncovered that ARF genes are alternatively spliced in distinct tissues of the early developing fruit and the splice variants exert different functions in individual tissues (K. Wang et al., 2016). Moreover, the dynamics of AS has also been characterized in strawberry (Y. Li et al., 2017). Intron retention events are abundant before fertilization in strawberry, but the retained introns are spliced out soon after fertilization, allowing rapid release of mature mRNA for translation needed for cell division and fruit expansion (Y. Li et al., 2017). Therefore, to better understand the fruit type specification, we can further examine the conserved AS events that occur at the early-stage fruit development in specific fruit-forming tissues of the four Rosaceae species in the future.

RNA has been extracted and sequenced at different stages from dissected fruit tissues of hormone-treated apple (Galimba et al., 2019). However, this type of data is not available for other Rosaceae species. Analyzing the RNA-Seq data collected from the fruits that are treated with plant hormones, such as auxin and GA, will allow us to understand whether distinct

tissues show different genetic responses to the hormones over time and how auxin and GA pathway genes are genetically linked.

Furthermore, two other Rosaceae species with special fruit morphology, *Potentilla micrantha* and *Rosa rugosa*, can be included in the comparative analyses of the Rosaceae early fruit development. The two species are phylogenetically close to the *Fragaria* species. However, *Potentilla micrantha* doesn't produce fleshy fruits, and *Rosa rugosa* fruit is derived from the hypanthium with numerous seeds trapped inside. We can dissect the fruits of the two Rosaceae species at different developmental stages and observe the morphological changes in the fruit tissues over time. Since the two species both have high-quality genomes (Buti et al., 2018; F. Chen et al., 2021), the RNA-Seq data can be collected from the early developing fruits of the two species and analyzed using these available genomic resources. Expanding the comparisons to other species of the Rosaceae with various fruit types will test and validate the function of the factors identified in my analysis, uncover new factors or mechanisms, and determine if a mechanism is conserved broadly or specific to a particular phylogenetic branch. Overall, my research work has played an important role in uncovering the genetic mechanisms that determine different fruit types and laid the foundation for future investigations into the evolution of fruit type in flowering plants.

Appendices

Table A.1 List of selective Rosaceae species and corresponding genome resources.

Species	Ploidy	Common name	Variety	GDR	Reference
Prunus davidiana	2n = 2x = 16	Chinese wild peach	Zhou Xing Shan Tao 1#	*Prunus davidiana* Genome v1.0	Cao et al., 2020
Prunus ferganensis	2n = 2x = 16		Ka Shi 1#	*Prunus ferganensis* Genome v1.0	Cao et al., 2020
Prunus kansuensis	2n = 2x = 16		Hong Gen Gan Su Tao 1#	*Prunus kansuensis* Genome v1.0	Cao et al., 2020
Prunus mira	2n = 2x = 16		2010-138	*Prunus mira* Genome v1.0	Cao et al., 2020
Prunus persica	2n = 2x = 16	Peach	Lovell doubled haploid	*Prunus persica* Genome v1.0	Verde et al., 2013
Prunus persica	2n = 2x = 16	Peach	Lovell doubled haploid	*Prunus persica* Genome v2.0.a1	Verde et al., 2017
Malus domestica	2n = 2x = 34	Apple	Golden Delicious	*Malus* x *domestica* Genome v1.0	Velasco et al., 2010
Malus domestica	2n = 2x = 34	Apple	Golden Delicious		X. Li et al., 2016
Malus domestica	2n = 2x = 34	Apple	Golden Delicious doubled haploid (GDDH13)	*Malus* x *domestica* GDDH13 Whole Genome v1.1	Daccord et al., 2017
Malus domestica	2n = 2x = 34	Apple	Hanfu anther-derived homozygous line (HFTH1)	*Malus* x *domestica* HFTH1 Whole Genome v1.0	Zhang et al., 2019
Malus domestica	2n = 2x = 34	Apple	Gala	*Malus* x *domestica* Gala haploid/diploid v1.0 genome	Sun et al., 2020
Malus sieversii	2n = 2x = 34			*Malus sieversii* haploid/diploid v1.0 genome	Sun et al., 2020
Malus sylvestris	2n = 2x = 34	European crab apple		*Malus sylvestris* haploid/diploid v1.0 genome	Sun et al., 2020
Fragaria x *ananassa*	2n = 8x = 56	Garden strawberry	Reikou	*Fragaria* x *ananassa* Reference Genome v1.0 (FANhybrid_r1.2)	Hirakawa et al., 2014
Fragaria x *ananassa*	2n = 8x = 56	Garden strawberry	Camarosa	*Fragaria* x *ananassa* Camarosa Genome v1.0.a1 & v1.0.a2	Edger et al., 2019; Liu et al., 2021
Fragaria bucharica	2n = 2x = 14			*Fragaria bucharica* Genome v1.0 (FBU_r1.1)	Hirakawa et al., 2014; Tennessen et al., 2014
Fragaria daltoniana	2n = 2x = 14			*Fragaria daltoniana* YNU Genome v1.0	Qiao et al., 2021
Fragaria iinumae	2n = 2x = 14			*Fragaria iinumae* Genome v1.0 (FII_r1.1)	Hirakawa et al., 2014
Fragaria iinumae	2n = 2x = 14			*Fragaria iinumae* Genome v1.0	Edger et al., 2020
Fragaria mandschurica	2n = 2x = 14			*Fragaria mandschurica* YNU Genome v1.0	Qiao et al., 2021
Fragaria nilgerrensis	2n = 2x = 14			*Fragaria nilgerrensis* SCBG Genome v1.0	Feng et al., 2021
Fragaria nilgerrensis	2n = 2x = 14			*Fragaria nilgerrensis* YNU Genome v1.0	Qiao et al., 2021
Fragaria nipponica	2n = 2x = 14			*Fragaria nipponica* Genome v1.0 (FNI_r1.1)	Hirakawa et al., 2014

Fragaria nubicola	2n = 2x = 14			*Fragaria nubicola* SCBG Genome v1.0	Feng et al., 2021
Fragaria orientalis	2n = 4x = 28			*Fragaria orientalis* Genome v1.0 (FOR_r1.1)	Hirakawa et al., 2014
Fragaria pentaphylla	2n = 2x = 14			*Fragaria pentaphylla* YNU Genome v1.0	Qiao et al., 2021
Fragaria viridis	2n = 2x = 14			*Fragaria viridis* SCBG Genome v1.0	Feng et al., 2021
Fragaria viridis	2n = 2x = 14			*Fragaria viridis* YNU Genome v1.0	Qiao et al., 2021
Fragaria vesca	2n = 2x = 14	Woodland strawberry	Hawaii 4	*Fragaria vesca* Genome v1.0 & v1.1.a2	Darwish et al., 2015; Shulaev et al., 2011
Fragaria vesca	2n = 2x = 14	Woodland strawberry		*Fragaria vesca* Genome v2.0.a1 & v2.0.a2	Y. Li et al., 2018; Tennessen et al., 2014
Fragaria vesca	2n = 2x = 14	Woodland strawberry	Hawaii 4	*Fragaria vesca* Genome v4.0.a1 & v4.0.a2	Edger et al., 2018; Y. Li et al., 2019
Rubus chingii	2n = 2x = 14	Chinese raspberry (Fu-Pen-Zi)		*Rubus chingii* Hu whole genome v1.0	Wang et al., 2021
Rubus idaeus	2n = 2x = 14	Red raspberry	Joan J.		Wight et al., 2019
Rubus idaeus	2n = 2x = 14	Red raspberry	Anitra		Davik et al., 2022
Rubus occidentalis	2n = 2x = 14	Black raspberry	ORUS 4115-3	*Rubus occidentalis* whole genome assembly v1.0.a1	VanBuren et al., 2016
Rubus occidentalis	2n = 2x = 14	Black raspberry	ORUS 4115-3	*Rubus occidentalis* whole genome assembly v1.1	Jibran et al., 2018
Rubus occidentalis	2n = 2x = 14	Black raspberry	ORUS 4115-3	*Rubus occidentalis* whole genome assembly v3.0	VanBuren et al., 2018

Table A.2 List of websites/databases useful for Rosaceae research.

Database name	URL	Species	Information provided	Reference
Rosaceae databases				
GDR (Genome Database for Rosaceae)	https://www.rosaceae.org/	Rosaceae	Genome	Jung et al., 2019
DBcherry	http://cherry.kazusa.or.jp/	*Cerasus* x *yedoensis* *Prunus avium*	Genome	Shirasawa et al., 2017, 2019
Strawberry GARDEN	http://strawberry-garden.kazusa.or.jp/	*Fragaria x ananassa*	Genome	Hirakawa et al., 2014
Rosa multiflora DB	http://rosa.kazusa.or.jp/	*Rosa multiflora*	Genome	Nakamura et al., 2018
THE APPLE GENOME AND EPIGENOME	https://iris.angers.inra.fr/gddh13/	*Malus domestica*	Genome Epigenome	Daccord et al., 2017
RchiOBHm-V2	https://lipm-browsers.toulouse.inra.fr/pub/RchiOBHm-V2/	*Rosa chinensis*	Genome	Raymond et al., 2018
SGR (Strawberry Genome Resources)	http://bioinformatics.towson.edu/strawberry/default.aspx	*Fragaria vesca*	Transcriptome	Darwish et al., 2013
Strawberry eFP Browser	http://bar.utoronto.ca/efp_strawberry/cgi-bin/efpWeb.cgi	*Fragaria vesca*	Transcriptome	Hawkins et al., 2017
TRANSNAP	http://plantomics.mind.meiji.ac.jp/nashi/	*Pyrus pyrifolia*	Transcriptome	Koshimizu et al., 2019
Fragaria vesca co-expression network explorer	http://159.203.72.198:3838/fvesca/	*Fragaria vesca*	Co-expression	Shahan et al., 2018
General plant databases				
NCBI Genome	https://www.ncbi.nlm.nih.gov/genome	General	Genome	Tatusova et al., 1999
PLAZA	https://bioinformatics.psb.ugent.be/plaza/	Plant	Genome	Van Bel et al., 2022
Phytozome	https://phytozome-next.jgi.doe.gov/	Plant	Genome	Goodstein et al., 2012
EnsemblPlants	https://plants.ensembl.org/index.html	Plant	Genome	Bolser et al., 2017
PMN (Plant Metabolic Network)	https://plantcyc.org/	Plant	Pathway	Schläpfer et al., 2017
Plant Reactome	https://plantreactome.gramene.org/index.php?lang=en	Plant	Pathway	Naithani et al., 2020
PlantTFDB (Transcription Factor Database)	http://planttfdb.gao-lab.org/	Plant	TF regulation	Jin et al., 2017
PlantRegMap (Plant Transcriptional Regulatory Map)	http://plantregmap.gao-lab.org/	Plant	TF regulation	Tian et al., 2020
Next-Gen Sequence Databases and sRNA Tools	https://mpss.meyerslab.org/	Plant	small RNA	Nakano et al., 2020
CANTATAdb	http://cantata.amu.edu.pl/	Plant	lncRNA	Szcześniak et al., 2016

Table A.3 RNA-Seq samples generated from four Rosaceae species (strawberry, raspberry, peach and apple)

Strawberry (*F. vesca*) hand dissected early-stage fruit tissues (Kang et al., 2013)	
TissueStage-Replicate No.	**Sample description**
Cortex1-1	Cortex of receptacle from just open flower, replicate 1
Cortex1-2	Cortex of receptacle from just open flower, replicate 2
Cortex2-1	Cortex of receptacle from the flowers which have been pollinated for about 3 days, replicate 1
Cortex2-2	Cortex of receptacle from the flowers which have been pollinated for about 3 days, replicate 2
Cortex3-1	Cortex of receptacle at about 6 DPA, same age as embryo-3, replicate 1
Cortex3-2	Cortex of receptacle at about 6 DPA, same age as embryo-3, replicate 2
Cortex4-1	Cortex of receptacle at about 9 DPA, same age as embryo-4, replicate 1
Cortex4-2	Cortex of receptacle at about 9 DPA, same age as embryo-4, replicate 2
Cortex5-1	Cortex of receptacle at about 12 DPA, same age as embryo-5, replicate 1
Cortex5-2	Cortex of receptacle at about 12 DPA, same age as embryo-5, replicate 2
Pith1-1	Pith of receptacle from just open flower, replicate 1
Pith1-2	Pith of receptacle from just open flower, replicate 2
Pith2-1	Pith of receptacle from the flowers which have been pollinated for about 3 days, replicate 1
Pith2-2	Pith of receptacle from the flowers which have been pollinated for about 3 days, replicate 2
Pith3-1	Pith of receptacle at about 6 DPA, same age as embryo-3, replicate 1
Pith3-2	Pith of receptacle at about 6 DPA, same age as embryo-3, replicate 2
Pith4-1	Pith of receptacle at about 9 DPA, same age as embryo-4, replicate 1
Pith4-2	Pith of receptacle at about 9 DPA, same age as embryo-4, replicate 2
Pith5-1	Pith of receptacle at about 12 DPA, same age as embryo-5, replicate 1
Pith5-2	Pith of receptacle at about 12 DPA, same age as embryo-5, replicate 2
Ovary wall1-1	Carpel walls (achene walls) from just open flower, replicate 1
Ovary wall1-2	Carpel walls (achene walls) from just open flower, replicate 2
Ovary wall2-1	Carpel walls (achene walls) from the flowers which have been pollinated for about 3 days, replicate 1
Ovary wall2-2	Carpel walls (achene walls) from the flowers which have been pollinated for about 3 days, replicate 2
Ovary wall3-1	Carpel walls (achene walls) at about 6 DPA, same age as embryo-3, replicate 1
Ovary wall3-2	Carpel walls (achene walls) at about 6 DPA, same age as embryo-3, replicate 2
Ovary wall4-1	Carpel walls (achene walls) at about 9 DPA, same age as embryo-4, replicate 1
Ovary wall4-2	Carpel walls (achene walls) at about 9 DPA, same age as embryo-4, replicate 2
Ovary wall5-1	Carpel walls (achene walls) at about 12 DPA, same age as embryo-5, replicate 1
Ovary wall5-2	Carpel walls (achene walls) at about 12 DPA, same age as embryo-5, replicate 2
Ovule1-1	Unfertilized Ovules from just open flower, replicate 1
Ovule1-2	Unfertilized Ovules from just open flower, replicate 2
Seed2-1	Seeds from the flowers which have been pollinated for about 3 days, replicate 1
Seed2-2	Seeds from the flowers which have been pollinated for about 3 days, 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 seed, at about 12 DPA, replicate 1
Embryo5-2	Mature embryos which fill up entire seed, at about 12 DPA, replicate 2
Ghost3-1	Seeds without embryos inside (ie. endosperm + embryo) at about 6 DPA, same age as embryo-3, replicate 1
Ghost3-2	Seeds without embryos inside (ie. endosperm + embryo) at about 6 DPA, same age as embryo-3, replicate 2
Ghost4-1	Seeds without embryos inside (ie. endosperm + embryo) at about 9 DPA, same age as embryo-4, replicate 1
Ghost4-2	Seeds without embryos inside (ie. endosperm + embryo) at about 9 DPA, same age as embryo-4, replicate 2
Ghost5-1	Seeds without embryos inside (ie. endosperm + embryo) at about 12 DPA, same age as embryo-5, replicate 1
Ghost5-2	Seeds without embryos inside (ie. endosperm + embryo) at about 12 DPA, same age as embryo-5, replicate 2
Red Raspberry (*Rubus idaeus*) Hand-dissected early-stage fruit tissues (Chapter 3)	
Tissue (DPA)-sample No.	**Sample description**
Receptacle (0DPA)-17	Entire receptacle with achenes removed at the day when flower just opens, replicate 1
Receptacle (0DPA)-27	Entire receptacle with achenes removed at the day when flower just opens, replicate 2
Receptacle (0DPA)-41	Entire receptacle with achenes removed at the day when flower just opens, replicate 3
Receptacle (0DPA)-S1	Entire receptacle with achenes removed at the day when flower just opens, replicate 4
Receptacle (2DPA)-1	Entire receptacle with achenes removed at the day 2 after pollination, replicate 1
Receptacle (2DPA)-17	Entire receptacle with achenes removed at the day 2 after pollination, replicate 2
Receptacle (2DPA)-4	Entire receptacle with achenes removed at the day 2 after pollination, replicate 3
Receptacle (2DPA)-S20	Entire receptacle with achenes removed at the day 2 after pollination, replicate 4
Receptacle (4DPA)-1	Entire receptacle with achenes removed at the day 4 after pollination, replicate 1
Receptacle (4DPA)-11	Entire receptacle with achenes removed at the day 4 after pollination, replicate 2
Receptacle (4DPA)-7	Entire receptacle with achenes removed at the day 4 after pollination, replicate 3
Receptacle (4DPA)-S25	Entire receptacle with achenes removed at the day 4 after pollination, replicate 4
Receptacle (6DPA)-1	Entire receptacle with achenes removed at the day 6 after pollination, replicate 1
Receptacle (6DPA)-10	Entire receptacle with achenes removed at the day 6 after pollination, replicate 2
Receptacle (6DPA)-7	Entire receptacle with achenes removed at the day 6 after pollination, replicate 3
Receptacle (6DPA)-S23	Entire receptacle with achenes removed at the day 6 after pollination, replicate 4
Receptacle (9DPA)-1	Entire receptacle with achenes removed at the day 9 after pollination, replicate 1
Receptacle (9DPA)-11	Entire receptacle with achenes removed at the day 9 after pollination, replicate 2
Receptacle (9DPA)-7	Entire receptacle with achenes removed at the day 9 after pollination, replicate 3
Receptacle (9DPA)-S28	Entire receptacle with achenes removed at the day 9 after pollination, replicate 4
Receptacle (12DPA)-1	Entire receptacle with achenes removed at the day 12 after pollination, replicate 1
Receptacle (12DPA)-13	Entire receptacle with achenes removed at the day 12 after pollination, replicate 2
Receptacle (12DPA)-4	Entire receptacle with achenes removed at the day 12 after pollination, replicate 3
Receptacle (12DPA)-S4	Entire receptacle with achenes removed at the day 12 after pollination, replicate 4
Ovary wall (0DPA)-17	Carpel walls (achene walls) from the flower which just opens, ovules have been removed, replicate 1
Ovary wall (0DPA)-24	Carpel walls (achene walls) from the flower which just opens, ovules have been removed, replicate 2
Ovary wall (0DPA)-7	Carpel walls (achene walls) from the flower which just opens, ovules have been removed, replicate 3
Ovary wall (0DPA)-S2	Carpel walls (achene walls) from the flower which just opens, ovules have been removed, replicate 4

Ovary wall (2DPA)-1	Carpel walls (achene walls) at about 2 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (2DPA)-17	Carpel walls (achene walls) at about 2 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (2DPA)-4	Carpel walls (achene walls) at about 2 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (2DPA)-S7	Carpel walls (achene walls) at about 2 DPA (days post anthesis), seeds have been removed, replicate 4
Ovary wall (4DPA)-1	Carpel walls (achene walls) at about 4 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (4DPA)-11	Carpel walls (achene walls) at about 4 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (4DPA)-7	Carpel walls (achene walls) at about 4 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (4DPA)-S24	Carpel walls (achene walls) at about 4 DPA (days post anthesis), seeds have been removed, replicate 4
Ovary wall (6DPA)-1	Carpel walls (achene walls) at about 6 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (6DPA)-10	Carpel walls (achene walls) at about 6 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (6DPA)-7	Carpel walls (achene walls) at about 6 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (6DPA)-S21	Carpel walls (achene walls) at about 6 DPA (days post anthesis), seeds have been removed, replicate 4
Ovary wall (9DPA)-1	Carpel walls (achene walls) at about 9 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (9DPA)-11	Carpel walls (achene walls) at about 9 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (9DPA)-7	Carpel walls (achene walls) at about 9 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (9DPA)-S26	Carpel walls (achene walls) at about 9 DPA (days post anthesis), seeds have been removed replicate 4
Ovary wall (12DPA)-1	Carpel walls (achene walls) at about 12 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (12DPA)-13	Carpel walls (achene walls) at about 12 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (12DPA)-4	Carpel walls (achene walls) at about 12 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (12DPA)-S5	Carpel walls (achene walls) at about 12 DPA (days post anthesis), seeds have been removed, replicate 4
Ovule (0DPA)-26	Unfertilized ovules dissected out of the achenes in flowers that just open, replicate 1
Ovule (0DPA)-41	Unfertilized ovules dissected out of the achenes in flowers that just open, replicate 2
Ovule (0DPA)-7	Unfertilized ovules dissected out of the achenes in flowers that just open, replicate 3
Ovule (0DPA)-S3	Unfertilized ovules dissected out of the achenes in flowers that just open, replicate 4
Seed (2DPA) -1	Seeds dissected out of the achenes at about 2 DPA (days post anthesis), replicate 1
Seed (2DPA)-17	Seeds dissected out of the achenes at about 2 DPA (days post anthesis), replicate 2
Seed (2DPA)-4	Seeds dissected out of the achenes at about 2 DPA (days post anthesis), replicate 3
Seed (2DPA)-S19	Seeds dissected out of the achenes at about 2 DPA (days post anthesis), replicate 4
Seed (4DPA) -1	Seeds dissected out of the achenes at about 4 DPA (days post anthesis), replicate 1
Seed (4DPA)-11	Seeds dissected out of the achenes at about 4 DPA (days post anthesis), replicate 2
Seed (4DPA)-7	Seeds dissected out of the achenes at about 4 DPA (days post anthesis), replicate 3
Seed (4DPA)-S8	Seeds dissected out of the achenes at about 4 DPA (days post anthesis), replicate 4
Seed (6DPA) -1	Seeds dissected out of the achenes at about 6 DPA (days post anthesis), replicate 1
Seed (6DPA)-10	Seeds dissected out of the achenes at about 6 DPA (days post anthesis), replicate 2
Seed (6DPA)-7	Seeds dissected out of the achenes at about 6 DPA (days post anthesis), replicate 3
Seed (6DPA)-S22	Seeds dissected out of the achenes at about 6 DPA (days post anthesis), replicate 4
Seed (9DPA) -1	Seeds dissected out of the achenes at about 9 DPA (days post anthesis), replicate 1
Seed (9DPA)-11	Seeds dissected out of the achenes at about 9 DPA (days post anthesis), replicate 2
Seed (9DPA)-7	Seeds dissected out of the achenes at about 9 DPA (days post anthesis), replicate 3
Seed (9DPA)-S27	Seeds dissected out of the achenes at about 9 DPA (days post anthesis), replicate 4
Seed (12DPA) -1	Seeds dissected out of the achenes at about 12 DPA (days post anthesis), replicate 1

Seed (12DPA)-13	Seeds dissected out of the achenes at about 12 DPA (days post anthesis), replicate 2
Seed (12DPA)-7	Seeds dissected out of the achenes at about 12 DPA (days post anthesis), replicate 3
Seed (12DPA)-S6	Seeds dissected out of the achenes at about 12 DPA (days post anthesis), replicate 4
Peach (*Prunus persica*) Hand dissected early-stage fruit tissues (M. Li et al., 2022)	
Tissue (DPA)-sample No.	**Sample description**
Hypanthium (0DPA)-9	Hypanthium dissected from flowers that just open, replicate 1
Hypanthium (0DPA)-28	Hypanthium dissected from flowers that just open, replicate 2
Hypanthium (0DPA)-44	Hypanthium dissected from flowers that just open, replicate 3
Hypanthium (5DPA)-9	Hypanthium dissected from pollinated flowers at 5 DPA (days post anthesis), replicate 1
Hypanthium (5DPA)-28	Hypanthium dissected from pollinated flowers at 5 DPA (days post anthesis), replicate 2
Hypanthium (5DPA)-44	Hypanthium dissected from pollinated flowers at 5 DPA (days post anthesis), replicate 3
Hypanthium (12DPA)-9	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 1
Hypanthium (12DPA)-28	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 2
Hypanthium (12DPA)-44	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 3
Hypanthium (18DPA)-9	Hypanthium dissected from pollinated flowers at 18 DPA (days post anthesis), replicate 1
Hypanthium (18DPA)-28	Hypanthium dissected from pollinated flowers at 18 DPA (days post anthesis), replicate 2
Hypanthium (18DPA)-44	Hypanthium dissected from pollinated flowers at 18 DPA (days post anthesis), replicate 3
Ovary wall (0DPA)-9	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 1
Ovary wall (0DPA)-28	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 2
Ovary wall (0DPA)-44	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 3
Ovary wall (5DPA)-9	Ovary wall dissected from pollinated flowers at 5 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (5DPA)-28	Ovary wall dissected from pollinated flowers at 5 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (5DPA)-44	Ovary wall dissected from pollinated flowers at 5 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (12DPA)-9	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (12DPA)-28	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (12DPA)-44	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (18DPA)-9	Ovary wall dissected from pollinated flowers at 18 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (18DPA)-28	Ovary wall dissected from pollinated flowers at 18 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (18DPA)-44	Ovary wall dissected from pollinated flowers at 18 DPA (days post anthesis), seeds have been removed, replicate 3
Ovule (0DPA)-9	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 1
Ovule (0DPA)-28	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 2
Ovule (0DPA)-44	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 3
Seed (5DPA)-9	Seeds dissected out of the ovaries at about 5 DPA (days post anthesis), replicate 1
Seed (5DPA)-28	Seeds dissected out of the ovaries at about 5 DPA (days post anthesis), replicate 2
Seed (5DPA)-44	Seeds dissected out of the ovaries at about 5 DPA (days post anthesis), replicate 3
Seed (12DPA)-9	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 1
Seed (12DPA)-28	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 2
Seed (12DPA)-44	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 3
Seed (18DPA)-9	Seeds dissected out of the ovaries at about 18 DPA (days post anthesis), replicate 1
Seed (18DPA)-28	Seeds dissected out of the ovaries at about 18 DPA (days post anthesis), replicate 2

Seed (18DPA)-44	Seeds dissected out of the ovaries at about 18 DPA (days post anthesis), replicate 3
Apple (*Malus domestica*) Hand dissected early-stage fruit tissues (M. Li et al., 2022)	
Tissue (DPA)-replicate No.	**Sample description**
Hypanthium (0DPA)-1	Hypanthium dissected from flowers that just open, replicate 1
Hypanthium (0DPA)-2	Hypanthium dissected from flowers that just open, replicate 2
Hypanthium (0DPA)-3	Hypanthium dissected from flowers that just open, replicate 3
Hypanthium (6DPA)-1	Hypanthium dissected from pollinated flowers at 6 DPA (days post anthesis), replicate 1
Hypanthium (6DPA)-2	Hypanthium dissected from pollinated flowers at 6 DPA (days post anthesis), replicate 2
Hypanthium (6DPA)-3	Hypanthium dissected from pollinated flowers at 6 DPA (days post anthesis), replicate 3
Hypanthium (12DPA)-1	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 1
Hypanthium (12DPA)-2	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 2
Hypanthium (12DPA)-3	Hypanthium dissected from pollinated flowers at 12 DPA (days post anthesis), replicate 3
Hypanthium (20DPA)-1	Hypanthium dissected from pollinated flowers at 20 DPA (days post anthesis), replicate 1
Hypanthium (20DPA)-2	Hypanthium dissected from pollinated flowers at 20 DPA (days post anthesis), replicate 2
Hypanthium (20DPA)-3	Hypanthium dissected from pollinated flowers at 20 DPA (days post anthesis), replicate 3
Ovary wall (0DPA)-1	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 1
Ovary wall (0DPA)-2	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 2
Ovary wall (0DPA)-3	Ovary wall dissected from flowers that just open, ovule has been removed, replicate 3
Ovary wall (6DPA)-1	Ovary wall dissected from pollinated flowers at 6 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (6DPA)-2	Ovary wall dissected from pollinated flowers at 6 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (6DPA)-3	Ovary wall dissected from pollinated flowers at 6 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (12DPA)-1	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (12DPA)-2	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (12DPA)-3	Ovary wall dissected from pollinated flowers at 12 DPA (days post anthesis), seeds have been removed, replicate 3
Ovary wall (20DPA)-1	Ovary wall dissected from pollinated flowers at 20 DPA (days post anthesis), seeds have been removed, replicate 1
Ovary wall (20DPA)-2	Ovary wall dissected from pollinated flowers at 20 DPA (days post anthesis), seeds have been removed, replicate 2
Ovary wall (20DPA)-3	Ovary wall dissected from pollinated flowers at 20 DPA (days post anthesis), seeds have been removed, replicate 3
Ovule (0DPA)-1	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 1
Ovule (0DPA)-2	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 2
Ovule (0DPA)-3	Unfertilized ovules dissected out of the ovaries in flowers that just open, replicate 3
Seed (6DPA)-1	Seeds dissected out of the ovaries at about 6 DPA (days post anthesis), replicate 1
Seed (6DPA)-2	Seeds dissected out of the ovaries at about 6 DPA (days post anthesis), replicate 2
Seed (6DPA)-3	Seeds dissected out of the ovaries at about 6 DPA (days post anthesis), replicate 3
Seed (12DPA)-1	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 1
Seed (12DPA)-2	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 2
Seed (12DPA)-3	Seeds dissected out of the ovaries at about 12 DPA (days post anthesis), replicate 3
Seed (20DPA)-1	Seeds dissected out of the ovaries at about 20 DPA (days post anthesis), replicate 1
Seed (20DPA)-2	Seeds dissected out of the ovaries at about 20 DPA (days post anthesis), replicate 2
Seed (20DPA)-3	Seeds dissected out of the ovaries at about 20 DPA (days post anthesis), replicate 3

Bibliography

Alexa, A., & Rahnenfuhrer, J. (2022). *topGO: Enrichment Analysis for Gene Ontology* (2.48.0) [Computer software]. Bioconductor version: Release (3.15). https://doi.org/10.18129/B9.bioc.topGO

Allaire, J. J., Ellis, P., Gandrud, C., Kuo, K., Lewis, B. W., Owen, J., Russell, K., Rogers, J., Sese, C., & Yetman, C. J. (2017). *networkD3: D3 JavaScript Network Graphs from R* (0.4) [Computer software]. https://CRAN.R-project.org/package=networkD3

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bai, Y., Dougherty, L., & Xu, K. (2014). Towards an improved apple reference transcriptome using RNA-seq. *Molecular Genetics and Genomics*, *289*(3), 427–438. https://doi.org/10.1007/s00438-014-0819-3

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, *14*(5), 988–995. https://doi.org/10.1101/gr.1865504

Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, *15*(1), 211. https://doi.org/10.1186/1471-2105-15-211

Bolser, D. M., Staines, D. M., Perry, E., & Kersey, P. J. (2017). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomic Data. In A. D. J. van Dijk (Ed.), *Plant Genomics Databases: Methods and Protocols* (pp. 1–31). Springer. https://doi.org/10.1007/978-1-4939-6658-5_1

Bombarely, A., Menda, N., Tecle, I. Y., Buels, R. M., Strickler, S., Fischer-York, T., Pujar, A., Leto, J., Gosselin, J., & Mueller, L. A. (2011). The Sol Genomics Network (solgenomics.net): Growing tomatoes using Perl. *Nucleic Acids Research*, *39*(suppl_1), D1149–D1155. https://doi.org/10.1093/nar/gkq866

Bowman, J. L., Smyth, D. R., & Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development*, *112*(1), 1–20. https://doi.org/10.1242/dev.112.1.1

Bowman, J. L., Smyth, D. R., & Meyerowitz, E. M. (2012). The ABC model of flower development: Then and now. *Development*, *139*(22), 4095–4098. https://doi.org/10.1242/dev.083972

Broholm, S. K., Pöllänen, E., Ruokolainen, S., Tähtiharju, S., Kotilainen, M., Albert, V. A., Elomaa, P., & Teeri, T. H. (2010). Functional characterization of B class MADS-box transcription factors in Gerbera hybrida. *Journal of Experimental Botany*, *61*(1), 75–85. https://doi.org/10.1093/jxb/erp279

Buti, M., Moretto, M., Barghini, E., Mascagni, F., Natali, L., Brilli, M., Lomsadze, A., Sonego, P., Giongo, L., Alonge, M., Velasco, R., Varotto, C., Šurbanovski, N., Borodovsky, M., Ward, J. A., Engelen, K., Cavallini, A., Cestaro, A., & Sargent, D. J. (2018). The genome sequence and transcriptome of Potentilla micrantha and their comparison to Fragaria vesca (the woodland strawberry). *GigaScience*, *7*(4). https://doi.org/10.1093/gigascience/giy010

Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in*

Bioinformatics, *48*(1), 4.11.1-4.11.39. https://doi.org/10.1002/0471250953.bi0411s48

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C. J., Ware, D., Shiu, S.-H., Childs, K. L., Sun, Y., Jiang, N., & Yandell, M. (2014). MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiology*, *164*(2), 513–524. https://doi.org/10.1104/pp.113.230144

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. https://doi.org/10.1101/gr.6743907

Cao, K., Peng, Z., Zhao, X., Li, Y., Liu, K., Arus, P., Zhu, G., Deng, S., Fang, W., Chen, C., Wang, X., Wu, J., Fei, Z., & Wang, L. (2020). Pan-genome analyses of peach and its wild relatives provide insights into the genetics of disease resistance and species adaptation. *BioRxiv*, 2020.07.13.200204. https://doi.org/10.1101/2020.07.13.200204

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2019). shiny: Web Application Framework for R. *R*.

Chapman, E. J., & Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annual Review of Genetics*, *43*(1), 265–285. https://doi.org/10.1146/annurev-genet-102108-134148

Chen, F., Su, L., Hu, S., Xue, J.-Y., Liu, H., Liu, G., Jiang, Y., Du, J., Qiao, Y., Fan, Y., Liu, H., Yang, Q., Lu, W., Shao, Z.-Q., Zhang, J., Zhang, L., Chen, F., & Cheng, Z.-M. (Max). (2021). A chromosome-level genome assembly of rugged rose (Rosa rugosa) provides insights into its evolution, ecology, and floral characteristics. *Horticulture Research*, *8*, 141. https://doi.org/10.1038/s41438-021-00594-z

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

Chen, Z., Jiang, J., Shu, L., Li, X., Huang, J., Qian, B., Wang, X., Li, X., Chen, J., & Xu, H. (2021). Combined transcriptomic and metabolic analyses reveal potential mechanism for fruit development and quality control of Chinese raspberry (Rubus chingii Hu). *Plant Cell Reports*, *40*(10), 1923–1946. https://doi.org/10.1007/s00299-021-02758-6

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C., & van Tunen, A. J. (1995). The petunia MADS box gene FBP11 determines ovule identity. *The Plant Cell*, *7*(11), 1859–1868. https://doi.org/10.1105/tpc.7.11.1859

Cong, L., Yue, R., Wang, H., Liu, J., Zhai, R., Yang, J., Wu, M., Si, M., Zhang, H., Yang, C., Xu, L., & Wang, Z. (2019). 2,4-D-induced parthenocarpy in pear is mediated by enhancement of GA4 biosynthesis. *Physiologia Plantarum*, *166*(3), 812–820. https://doi.org/10.1111/ppl.12835

Conway, J., & Gehlenborg, N. (2019). *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets* (1.4.0) [Computer software]. https://CRAN.R-project.org/package=UpSetR

Crane, J. C., Primer, P. E., & Campbell, R. C. (1960). Gibberellin induced parthenocarpy in Prunus. *Proceedings. American Society for Horticultural Science*, *75*, 129–137.

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., … Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, *49*(7), 1099–1106. https://doi.org/10.1038/ng.3886

Dardick, C. D., Callahan, A. M., Chiozzotto, R., Schaffer, R. J., Piagnani, M. C., & Scorza, R. (2010). Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to Arabidopsisdehiscence. *BMC Biology*, *8*(1), 13. https://doi.org/10.1186/1741-7007-8-13

Darwish, O., Shahan, R., Liu, Z., Slovin, J. P., & Alkharouf, N. W. (2015). Re-annotation of the woodland strawberry (Fragaria vesca) genome. *BMC Genomics*, *16*(1), 29. https://doi.org/10.1186/s12864-015-1221-1

Darwish, O., Slovin, J. P., Kang, C., Hollender, C. A., Geretz, A., Houston, S., Liu, Z., & Alkharouf, N. W. (2013). SGR: An online genomic resource for the woodland strawberry. *BMC Plant Biology*, *13*(1), 223. https://doi.org/10.1186/1471-2229-13-223

Davière, J.-M., & Achard, P. (2016). A Pivotal Role of DELLAs in Regulating Multiple Hormone Signals. *Molecular Plant*, *9*(1), 10–20. https://doi.org/10.1016/j.molp.2015.09.011

Davik, J., Røen, D., Lysøe, E., Buti, M., Rossman, S., Alsheikh, M., Aiden, E. L., Dudchenko, O., & Sargent, D. J. (2022). A chromosome-level genome sequence assembly of the red raspberry (Rubus idaeus L.). *PLOS ONE*, *17*(3), e0265096. https://doi.org/10.1371/journal.pone.0265096

De Jong, M., Wolters-Arts, M., Feron, R., Mariani, C., & Vriezen, W. H. (2009). The Solanum lycopersicum auxin response factor 7 (SlARF7) regulates auxin signaling during tomato fruit set and development. *The Plant Journal*, *57*(1), 160–170. https://doi.org/10.1111/j.1365-313X.2008.03671.x

de Martino, G., Pan, I., Emmanuel, E., Levy, A., & Irish, V. F. (2006). Functional Analyses of Two Tomato APETALA3 Genes Demonstrate Diversification in Their Roles in Regulating Floral Development. *The Plant Cell*, *18*(8), 1833–1845. https://doi.org/10.1105/tpc.106.042978

Dey, K. K., Hsiao, C. J., & Stephens, M. (2017). Visualizing the structure of RNA-seq expression data using grade of membership models. *PLOS Genetics*, *13*(3), e1006599. https://doi.org/10.1371/journal.pgen.1006599

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., & Yanofsky, M. F. (2004). The SEP4 Gene of Arabidopsis thaliana Functions in Floral Organ and Meristem Identity. *Current Biology*, *14*(21), 1935–1940. https://doi.org/10.1016/j.cub.2004.10.028

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

Dubois, A., Raymond, O., Maene, M., Baudino, S., Langlade, N. B., Boltz, V., Vergne, P., & Bendahmane, M. (2010). Tinkering with the C-Function: A Molecular Frame for the Selection of Double Flowers in Cultivated Roses. *PLOS ONE*, *5*(2), e9288. https://doi.org/10.1371/journal.pone.0009288

Edger, P. P., McKain, M. R., Yocca, A. E., Knapp, S. J., Qiao, Q., & Zhang, T. (2020). Reply to: Revisiting the origin of octoploid strawberry. *Nature Genetics*, *52*(1), 5–7. https://doi.org/10.1038/s41588-019-0544-2

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., Smith, R. D., Teresi, S. J., Nelson, A. D. L., Wai, C. M., Alger, E. I., Bird, K. A., Yocca, A. E., Pumplin, N., Ou, S., Ben-Zvi, G., Brodt, A., Baruch, K., Swale, T., … Knapp, S. J. (2019). Origin and evolution of the octoploid strawberry genome. *Nature Genetics*, *51*(3), 541–547. https://doi.org/10.1038/s41588-019-0356-4

Edger, P. P., VanBuren, R., Colle, M., Poorten, T. J., Wai, C. M., Niederhuth, C. E., Alger, E. I., Ou, S., Acharya, C. B., Wang, J., Callow, P., McKain, M. R., Shi, J., Collier, C., Xiong, Z., Mower, J. P., Slovin, J. P., Hytönen, T., Jiang, N., … Knapp, S. J. (2018). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. *GigaScience*, *7*(2). https://doi.org/10.1093/gigascience/gix124

Egea-Cortines, M., Saedler, H., & Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. *The EMBO Journal*, *18*(19), 5370–5379. https://doi.org/10.1093/emboj/18.19.5370

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., & Colombo, L. (2003). MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *The Plant Cell*, *15*(11), 2603–2611. https://doi.org/10.1105/tpc.015123

Feng, C., Wang, J., Harris, A. J., Folta, K. M., Zhao, M., & Kang, M. (2021). Tracing the Diploid Ancestry of the Cultivated Octoploid Strawberry. *Molecular Biology and Evolution*, *38*(2), 478–485. https://doi.org/10.1093/molbev/msaa238

Galimba, K. D., Bullock, D. G., Dardick, C., Liu, Z., & Callahan, A. M. (2019). Gibberellic acid induced parthenocarpic 'Honeycrisp' apples (Malus domestica) exhibit reduced ovary width and lower acidity. *Horticulture Research*, *6*, 41. https://doi.org/10.1038/s41438-019-0124-8

García-Hurtado, N., Carrera, E., Ruiz-Rivero, O., López-Gresa, M. P., Hedden, P., Gong, F., & García-Martínez, J. L. (2012). The characterization of transgenic tomato overexpressing gibberellin 20-oxidase reveals induction of parthenocarpic fruit growth, higher yield, and alteration of the gibberellin biosynthetic pathway. *Journal of Experimental Botany*, *63*(16), 5803–5813. https://doi.org/10.1093/jxb/ers229

Geisler, M. M. (2021). A Retro-Perspective on Auxin Transport. *Frontiers in Plant Science*, *12*. https://www.frontiersin.org/article/10.3389/fpls.2021.756968

Gillaspy, G., Ben-David, H., & Gruissem, W. (1993). Fruits: A Developmental Perspective. *The Plant Cell*, *5*(10), 1439–1451.

Goetz, M., Hooper, L. C., Johnson, S. D., Rodrigues, J. C. M., Vivian-Smith, A., & Koltunow, A. M. (2007). Expression of Aberrant Forms of AUXIN RESPONSE FACTOR8 Stimulates Parthenocarpy in Arabidopsis and Tomato. *Plant Physiology*, *145*(2), 351–366. https://doi.org/10.1104/pp.107.104174

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Graham, J., & Woodhead, M. (2009). Raspberries and Blackberries: The Genomics of Rubus. In K. M. Folta & S. E. Gardiner (Eds.), *Genetics and Genomics of Rosaceae* (pp. 507–524). Springer. https://doi.org/10.1007/978-0-387-77491-6_24

Gramzow, L., & Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biology*, *11*(6), 214. https://doi.org/10.1186/gb-2010-11-6-214

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y. P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M. G., Hocher, V., … Cheng, S. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*, *361*(6398). https://doi.org/10.1126/science.aat1743

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Hawkins, C., Caruana, J., Li, J., Zawora, C., Darwish, O., Wu, J., Alkharouf, N., & Liu, Z. (2017). An eFP browser for visualizing strawberry fruit and flower transcriptomes. *Horticulture Research*, *4*(1), 1–8. https://doi.org/10.1038/hortres.2017.29

He, H., & Yamamuro, C. (2022). Interplays between auxin and GA signaling coordinate early fruit development. *Horticulture Research*, *9*, uhab078. https://doi.org/10.1093/hr/uhab078

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., Kohara, M., Watanabe, A., Kishida, Y., Fujishiro, T., Tsuruoka, H., Minami, C., Sasamoto, S., Kato, M., Nanri, K., Komaki, A., Yanagi, T., Guoxin, Q., Maeda, F., … Isobe, S. N. (2014). Dissection of the Octoploid Strawberry Genome by Deep Sequencing of the Genomes of Fragaria Species. *DNA Research*, *21*(2), 169–181. https://doi.org/10.1093/dnares/dst049

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-Genome Annotation with BRAKER. In M. Kollmar (Ed.), *Gene Prediction: Methods and Protocols* (pp. 65–95). Springer. https://doi.org/10.1007/978-1-4939-9173-0_5

Hoff, K. J., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*(1), e57. https://doi.org/10.1002/cpbi.57

Hofmann, N. R. (2011). YUC and TAA1/TAR Proteins Function in the Same Pathway for Auxin Biosynthesis. *The Plant Cell*, *23*(11), 3869. https://doi.org/10.1105/tpc.111.231112

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

Hon, T., Mars, K., Young, G., Tsai, Y.-C., Karalius, J. W., Landolin, J. M., Maurer, N., Kudrna, D., Hardigan, M. A., Steiner, C. C., Knapp, S. J., Ware, D., Shapiro, B., Peluso, P., & Rank, D. R. (2020). Highly accurate long-read HiFi

sequencing data for five complex genomes. *Scientific Data*, *7*(1), 399. https://doi.org/10.1038/s41597-020-00743-4

Honma, T., & Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, *409*(6819), 525–529. https://doi.org/10.1038/35054083

Hu, J., Israeli, A., Ori, N., & Sun, T. (2018). The Interaction between DELLA and ARF/IAA Mediates Crosstalk between Gibberellin and Auxin Signaling to Control Fruit Initiation in Tomato. *The Plant Cell*, *30*(8), 1710–1728. https://doi.org/10.1105/tpc.18.00363

Ireland, H. S., Yao, J.-L., Tomes, S., Sutherland, P. W., Nieuwenhuizen, N., Gunaseelan, K., Winz, R. A., David, K. M., & Schaffer, R. J. (2013). Apple SEPALLATA1/2-like genes control fruit flesh development and ripening. *The Plant Journal*, *73*(6), 1044–1056. https://doi.org/10.1111/tpj.12094

Irish, V. (2017). The ABC model of floral development. *Current Biology*, *27*(17), R887–R890. https://doi.org/10.1016/j.cub.2017.03.045

Jain, A., & Tuteja, G. (2019). TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics*, *35*(11), 1966–1967. https://doi.org/10.1093/bioinformatics/bty890

Janssen, B. J., Thodey, K., Schaffer, R. J., Alba, R., Balakrishnan, L., Bishop, R., Bowen, J. H., Crowhurst, R. N., Gleave, A. P., Ledger, S., McArtney, S., Pichler, F. B., Snowden, K. C., & Ward, S. (2008). Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. *BMC Plant Biology*, *8*(1), 16. https://doi.org/10.1186/1471-2229-8-16

Jiao, W.-B., & Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology*, *36*, 64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jibran, R., Dzierzon, H., Bassil, N., Bushakra, J. M., Edger, P. P., Sullivan, S., Finn, C. E., Dossett, M., Vining, K. J., VanBuren, R., Mockler, T. C., Liachko, I., Davies, K. M., Foster, T. M., & Chagné, D. (2018). Chromosome-scale scaffolding of the black raspberry (Rubus occidentalis L.) genome based on chromatin interaction data. *Horticulture Research*, *5*(1), 1–11. https://doi.org/10.1038/s41438-017-0013-y

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045. https://doi.org/10.1093/nar/gkw982

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jordan, C. Y., Lohse, K., Turner, F., Thomson, M., Gharbi, K., & Ennos, R. A. (2018). Maintaining their genetic distance: Little evidence for introgression between widely hybridizing species of Geum with contrasting mating systems. *Molecular Ecology*, *27*(5), 1214–1228. https://doi.org/10.1111/mec.14426

Jung, C. J., Hur, Y. Y., Yu, H.-J., Noh, J.-H., Park, K.-S., & Lee, H. J. (2014). Gibberellin Application at Pre-Bloom in Grapevines Down-Regulates the Expressions of VvIAA9 and VvARF7, Negative Regulators of Fruit Set Initiation, during Parthenocarpic Fruit Development. *PLOS ONE*, *9*(4), e95634. https://doi.org/10.1371/journal.pone.0095634

Jung, S., Bassett, C., Bielenberg, D. G., Cheng, C.-H., Dardick, C., Main, D., Meisel, L., Slovin, J., Troggio, M., & Schaffer, R. J. (2015). A standard nomenclature for gene designation in the Rosaceae. *Tree Genetics & Genomes*, *11*(5), 108. https://doi.org/10.1007/s11295-015-0931-5

Jung, S., Lee, T., Cheng, C.-H., Buble, K., Zheng, P., Yu, J., Humann, J., Ficklin, S. P., Gasic, K., Scott, K., Frank, M., Ru, S., Hough, H., Evans, K., Peace, C., Olmstead, M., DeVetter, L. W., McFerson, J., Coe, M., … Main, D. (2019). 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Research*, *47*(D1), D1137–D1145. https://doi.org/10.1093/nar/gky1000

Kammonen, J. I., Smolander, O.-P., Paulin, L., Pereira, P. A. B., Laine, P., Koskinen, P., Jernvall, J., & Auvinen, P. (2019). gapFinisher: A reliable gap filling pipeline for SSPACE-LongRead scaffolder output. *PLOS ONE*, *14*(9), e0216885. https://doi.org/10.1371/journal.pone.0216885

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. *The Plant Cell*, *25*(6), 1960–1978. https://doi.org/10.1105/tpc.113.111732

Kanno, A., Saeki, H., Kameya, T., Saedler, H., & Theissen, G. (2003). Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (Tulipa gesneriana). *Plant Molecular Biology*, *52*(4), 831–841. https://doi.org/10.1023/A:1025070827979

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, *12*(4), 656–664. https://doi.org/10.1101/gr.229202

Korbel, J. O., & Lee, C. (2013). Genome assembly and haplotyping with Hi-C. *Nature Biotechnology*, *31*(12), 1099–1101. https://doi.org/10.1038/nbt.2764

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*(1), 59. https://doi.org/10.1186/1471-2105-5-59

Korf, I., Flicek, P., Duan, D., & Brent, M. R. (2001). Integrating genomic homology into gene structure prediction. *Bioinformatics*, *17*(suppl_1), S140–S148. https://doi.org/10.1093/bioinformatics/17.suppl_1.S140

Koshimizu, S., Nakamura, Y., Nishitani, C., Kobayashi, M., Ohyanagi, H., Yamamoto, T., & Yano, K. (2019). TRANSNAP: A web database providing

comprehensive information on Japanese pear transcriptome. *Scientific Reports*, *9*(1), 18922. https://doi.org/10.1038/s41598-019-55287-4

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., & Huala, E. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, *40*(D1), D1202–D1210. https://doi.org/10.1093/nar/gkr1090

Langdon, Q. K., Peris, D., Kyle, B., & Hittinger, C. T. (2018). sppIDer: A Species Identification Tool to Investigate Hybrid Genomes with High-Throughput Sequencing. *Molecular Biology and Evolution*, *35*(11), 2835–2849. https://doi.org/10.1093/molbev/msy166

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Lewis, S., Searle, S., Harris, N., Gibson, M., Iyer, V., Richter, J., Wiel, C., Bayraktaroglu, L., Birney, E., Crosby, M., Kaminker, J., Matthews, B., Prochnik, S., Smith, C., Tupy, J., Rubin, G., Misra, S., Mungall, C., & Clamp, M. (2002). Apollo: A sequence annotation editor. *Genome Biology*, *3*(12), research0082.1. https://doi.org/10.1186/gb-2002-3-12-research0082

Leyser, O. (2018). Auxin Signaling. *Plant Physiology*, *176*(1), 465–479. https://doi.org/10.1104/pp.17.00765

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

Li, M., Xiao, Y., Mount, S., & Liu, Z. (2021). An Atlas of Genomic Resources for Studying Rosaceae Fruits and Ornamentals. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.644881

Li, X., Kui, L., Zhang, J., Xie, Y., Wang, L., Yan, Y., Wang, N., Xu, J., Li, C., Wang, W., van Nocker, S., Dong, Y., Ma, F., & Guan, Q. (2016). Improved hybrid de novo genome assembly of domesticated apple (Malus x domestica). *GigaScience*, *5*(1). https://doi.org/10.1186/s13742-016-0139-0

Li, Y., Dai, C., Hu, C., Liu, Z., & Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *The Plant Journal*, *90*(1), 164–176. https://doi.org/10.1111/tpj.13462

Li, Y., Pi, M., Gao, Q., Liu, Z., & Kang, C. (2019). Updated annotation of the wild strawberry Fragaria vesca V4 genome. *Horticulture Research*, *6*(1), 1–9. https://doi.org/10.1038/s41438-019-0142-6

Li, Y., Wei, W., Feng, J., Luo, H., Pi, M., Liu, Z., & Kang, C. (2018). Genome re-annotation of the wild strawberry Fragaria vesca using extensive Illumina- and SMRT-based RNA-seq datasets. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *25*(1), 61–70. https://doi.org/10.1093/dnares/dsx038

Liao, X., Li, M., Liu, B., Yan, M., Yu, X., Zi, H., Liu, R., & Yamamuro, C. (2018). Interlinked regulatory loops of ABA catabolism and biosynthesis coordinate fruit growth and ripening in woodland strawberry. *Proceedings of the National Academy of Sciences*, *115*(49), E11542–E11550. https://doi.org/10.1073/pnas.1812575115

Liston, A., Wei, N., Tennessen, J. A., Li, J., Dong, M., & Ashman, T.-L. (2020). Revisiting the origin of octoploid strawberry. *Nature Genetics*, *52*(1), 2–4. https://doi.org/10.1038/s41588-019-0543-3

Liu, Q., Luo, L., & Zheng, L. (2018). Lignins: Biosynthesis and Biological Functions in Plants. *International Journal of Molecular Sciences*, *19*(2), 335. https://doi.org/10.3390/ijms19020335

Liu, S., Zhang, Y., Feng, Q., Qin, L., Pan, C., Lamin-Samu, A. T., & Lu, G. (2018). Tomato AUXIN RESPONSE FACTOR 5 regulates fruit set and development via the mediation of auxin and gibberellin signaling. *Scientific Reports*, *8*(1), 2971. https://doi.org/10.1038/s41598-018-21315-y

Liu, T., Li, M., Liu, Z., Ai, X., & Li, Y. (2021). Reannotation of the cultivated strawberry genome and establishment of a strawberry genome database. *Horticulture Research*, *8*(1), 1–9. https://doi.org/10.1038/s41438-021-00476-4

Liu, Z., Ma, H., Jung, S., Main, D., & Guo, L. (2020). Developmental Mechanisms of Fleshy Fruit Diversity in Rosaceae. *Annual Review of Plant Biology*, *71*(1), 547–573. https://doi.org/10.1146/annurev-arplant-111119-021700

Lombardo, V. A., Osorio, S., Borsani, J., Lauxmann, M. A., Bustamante, C. A., Budde, C. O., Andreo, C. S., Lara, M. V., Fernie, A. R., & Drincovich, M. F. (2011). Metabolic Profiling during Peach Fruit Development and Ripening Reveals the Metabolic Networks That Underpin Each Developmental Stage. *Plant Physiology*, *157*(4), 1696–1710. https://doi.org/10.1104/pp.111.186064

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506. https://doi.org/10.1093/nar/gki937

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

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Ludwig-Müller, J. (2016). Auxin conjugation: Growing out of the shade. *Nature Plants*, *2*(4), 1–2. https://doi.org/10.1038/nplants.2016.44

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics*, *20*(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., Pitrat, M., Dogimont, C., & Bendahmane, A. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature*, *461*(7267), 1135–1138. https://doi.org/10.1038/nature08498

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mayorga-Gómez, A., & Nambeesan, S. U. (2020). Temporal expression patterns of fruit-specific α- EXPANSINS during cell expansion in bell pepper (Capsicum annuum L.). *BMC Plant Biology*, *20*(1), 241. https://doi.org/10.1186/s12870-020-02452-x

Mazzucato, A., Olimpieri, I., Siligato, F., Picarella, M. E., & Soressi, G. P. (2008). Characterization of genes controlling stamen identity and development in a parthenocarpic tomato mutant indicates a role for the DEFICIENS ortholog in the control of fruit set. *Physiologia Plantarum*, *132*(4), 526–537. https://doi.org/10.1111/j.1399-3054.2007.01035.x

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, *41*(12), e121. https://doi.org/10.1093/nar/gkt263

Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning*, *52*(1), 91–118. https://doi.org/10.1023/A:1023949509487

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(suppl_2), W182–W185. https://doi.org/10.1093/nar/gkm321

Naithani, S., Gupta, P., Preece, J., D'Eustachio, P., Elser, J. L., Garg, P., Dikeman, D. A., Kiff, J., Cook, J., Olson, A., Wei, S., Tello-Ruiz, M. K., Mundo, A. F., Munoz-Pomer, A., Mohammed, S., Cheng, T., Bolton, E., Papatheodorou, I., Stein, L., … Jaiswal, P. (2020). Plant Reactome: A knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Research*, *48*(D1), D1093–D1103. https://doi.org/10.1093/nar/gkz996

Nakamura, N., Hirakawa, H., Sato, S., Otagaki, S., Matsumoto, S., Tabata, S., & Tanaka, Y. (2018). Genome structure of Rosa multiflora, a wild ancestor of cultivated roses. *DNA Research*, *25*(2), 113–121. https://doi.org/10.1093/dnares/dsx042

Nakano, M., McCormick, K., Demirci, C., Demirci, F., Gurazada, S. G. R., Ramachandruni, D., Dusia, A., Rothhaupt, J. A., & Meyers, B. C. (2020). Next-Generation Sequence Databases: RNA and Genomic Informatics Resources for Plants. *Plant Physiology*, *182*(1), 136–146. https://doi.org/10.1104/pp.19.00957

NAKAYAMA, A., NAKAJIMA, M., & YAMAGUCHI, I. (2005). Distribution of Gibberellins and Expressional Analysis of GA 20-oxidase Genes of Morning Glory during Fruit Maturation. *Bioscience, Biotechnology, and Biochemistry*, *69*(2), 334–342. https://doi.org/10.1271/bbb.69.334

Ng, M., & Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics*, *2*(3), 186–195. https://doi.org/10.1038/35056041

Olvera-Carrillo, Y., Van Bel, M., Van Hautegem, T., Fendrych, M., Huysmans, M., Simaskova, M., van Durme, M., Buscaill, P., Rivas, S., S. Coll, N., Coppens, F., Maere, S., & Nowack, M. K. (2015). A Conserved Core of Programmed Cell Death Indicator Genes Discriminates Developmentally and Environmentally Induced Programmed Cell Death in Plants. *Plant Physiology*, *169*(4), 2684–2699. https://doi.org/10.1104/pp.15.00769

Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *ArXiv:1403.2805 [Cs, Stat]*. http://arxiv.org/abs/1403.2805

Ou, S., & Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiology*, *176*(2), 1410–1422. https://doi.org/10.1104/pp.17.01310

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, *405*(6783), 200–203. https://doi.org/10.1038/35012103

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Pi, M., Hu, S., Cheng, L., Zhong, R., Cai, Z., Liu, Z., Yao, J.-L., & Kang, C. (2021). The MADS-box gene FveSEP3 plays essential roles in flower organogenesis and fruit development in woodland strawberry. *Horticulture Research*, *8*, 247. https://doi.org/10.1038/s41438-021-00673-1

Qiao, Q., Edger, P. P., Xue, L., Qiong, L., Lu, J., Zhang, Y., Cao, Q., Yocca, A. E., Platts, A. E., Knapp, S. J., Van Montagu, M., Van de Peer, Y., Lei, J., & Zhang, T. (2021). Evolutionary history and pan-genome dynamics of strawberry (Fragaria spp.). *Proceedings of the National Academy of Sciences*, *118*(45), e2105431118. https://doi.org/10.1073/pnas.2105431118

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., Vergne, P., Moja, S., Choisne, N., Pont, C., Carrère, S., Caissard, J.-C., Couloux, A., Cottret, L., Aury, J.-M., Szécsi, J., Latrasse, D., Madoui, M.-A., François, L., … Bendahmane, M. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nature Genetics*, *50*(6), 772–777. https://doi.org/10.1038/s41588-018-0110-3

Ren, G., Li, L., Huang, Y., Wang, Y., Zhang, W., Zheng, R., Zhong, C., & Wang, X. (2018). GhWIP2, a WIP zinc finger protein, suppresses cell expansion in Gerbera hybrida by mediating crosstalk between gibberellin, abscisic acid, and auxin. *New Phytologist*, *219*(2), 728–742. https://doi.org/10.1111/nph.15175

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(1), 460. https://doi.org/10.1186/s12859-018-2485-7

Roldan, M. V. G., Izhaq, F., Verdenaud, M., Eleblu, J., Haraghi, A., Sommard, V., Chambrier, P., Latrasse, D., Jégu, T., Benhamed, M., Szécsi, J., Bendahmane, M., Boualem, A., & Bendahmane, A. (2020). Integrative genome-wide analysis reveals the role of WIP proteins in inhibition of growth and development. *Communications Biology*, *3*(1), 1–12. https://doi.org/10.1038/s42003-020-0969-2

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio Gene Finding in Drosophila Genomic DNA. *Genome Research*, *10*(4), 516–522. https://doi.org/10.1101/gr.10.4.516

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2020). GenBank. *Nucleic Acids Research*, *48*(D1), D84–D86. https://doi.org/10.1093/nar/gkz956

Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., Dreher, K., Chavali, A. K., Nilo-Poyanco, R., Bernard, T., Kahn, D., & Rhee, S. Y. (2017). Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiology*, *173*(4), 2041–2059. https://doi.org/10.1104/pp.16.01942

Shahan, R., Zawora, C., Wight, H., Sittmann, J., Wang, W., Mount, S. M., & Liu, Z. (2018). Consensus Coexpression Network Analysis Identifies Key Regulators of Flower and Fruit Development in Wild Strawberry. *Plant Physiology*, *178*(1), 202–216. https://doi.org/10.1104/pp.18.00086

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shirasawa, K., Esumi, T., Hirakawa, H., Tanaka, H., Itai, A., Ghelfi, A., Nagasaki, H., & Isobe, S. (2019). Phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (Cerasus × yedoensis). *DNA Research*, *26*(5), 379–389. https://doi.org/10.1093/dnares/dsz016

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., & Isobe, S. (2017). The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding. *DNA Research*, *24*(5), 499–508. https://doi.org/10.1093/dnares/dsx020

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S. P., Burns, P., Davis, T. M., Slovin, J. P., Bassil, N., Hellens, R. P., Evans, C., Harkins, T., Kodira, C., Desany, B., … Folta, K. M. (2011). The genome of woodland strawberry (Fragaria vesca). *Nature Genetics*, *43*(2), 109–116. https://doi.org/10.1038/ng.740

Shumate, A., & Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics*, *37*(12), 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), 31. https://doi.org/10.1186/1471-2105-6-31

Solovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology*, *7*(1), S10. https://doi.org/10.1186/gb-2006-7-s1-s10

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(suppl_2), W435–W439. https://doi.org/10.1093/nar/gkl200

Stortenbeker, N., & Bemer, M. (2019). The SAUR gene family: The plant's toolbox for adaptation of growth and development. *Journal of Experimental Botany*, *70*(1), 17–27. https://doi.org/10.1093/jxb/ery332

Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., Cheng, L., Zhong, G.-Y., & Fei, Z. (2020). Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics*, *52*(12), 1423–1432. https://doi.org/10.1038/s41588-020-00723-9

Szcześniak, M. W., Rosikiewicz, W., & Makałowska, I. (2016). CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant and Cell Physiology*, *57*(1), e8–e8. https://doi.org/10.1093/pcp/pcv201

Tang, H., Lyons, E., & Town, C. D. (2015). Optical mapping in plant comparative genomics. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0044-y

Tang, H., Wang, X., Bowers, J. E., Ming, R., Alam, M., & Paterson, A. H. (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research*, *18*(12), 1944–1954. https://doi.org/10.1101/gr.080978.108

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., & Lu, J. (2015). ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biology*, *16*(1), 3. https://doi.org/10.1186/s13059-014-0573-1

Tatusova, T. A., Karsch-Mizrachi, I., & Ostell, J. A. (1999). Complete genomes in WWW Entrez: Data representation and analysis. *Bioinformatics*, *15*(7), 536–543. https://doi.org/10.1093/bioinformatics/15.7.536

Tennessen, J. A., Govindarajulu, R., Ashman, T.-L., & Liston, A. (2014). Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes

Revealed by Dense Targeted Capture Linkage Maps. *Genome Biology and Evolution*, *6*(12), 3295–3313. https://doi.org/10.1093/gbe/evu261
Theißen, G. (2001). Development of floral organ identity: Stories from the MADS house. *Current Opinion in Plant Biology*, *4*(1), 75–85. https://doi.org/10.1016/S1369-5266(00)00139-4
Theißen, G., Melzer, R., & Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Development*, *143*(18), 3259–3271. https://doi.org/10.1242/dev.134080
Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., & Gao, G. (2020). PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Research*, *48*(D1), D1104–D1113. https://doi.org/10.1093/nar/gkz1020
Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., & Vandepoele, K. (2018). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, *46*(D1), D1190–D1196. https://doi.org/10.1093/nar/gkx1002
Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, L., Botzki, A., Coppens, F., & Vandepoele, K. (2022). PLAZA 5.0: Extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Research*, *50*(D1), D1468–D1474. https://doi.org/10.1093/nar/gkab1024
VanBuren, R., Bryant, D., Bushakra, J. M., Vining, K. J., Edger, P. P., Rowley, E. R., Priest, H. D., Michael, T. P., Lyons, E., Filichkin, S. A., Dossett, M., Finn, C. E., Bassil, N. V., & Mockler, T. C. (2016). The genome of black raspberry (Rubus occidentalis). *The Plant Journal*, *87*(6), 535–547. https://doi.org/10.1111/tpj.13215
VanBuren, R., Wai, C. M., Colle, M., Wang, J., Sullivan, S., Bushakra, J. M., Liachko, I., Vining, K. J., Dossett, M., Finn, C. E., Jibran, R., Chagné, D., Childs, K., Edger, P. P., Mockler, T. C., & Bassil, N. V. (2018). A near complete, chromosome-scale assembly of the black raspberry (Rubus occidentalis) genome. *GigaScience*, *7*(8). https://doi.org/10.1093/gigascience/giy094
Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K., & Gerats, T. (2004). The Duplicated B-Class Heterodimer Model: Whorl-Specific Effects and Complex Genetic Interactions in Petunia hybrida Flower Development. *The Plant Cell*, *16*(3), 741–754. https://doi.org/10.1105/tpc.019166
Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116
Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S. K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nature Genetics*, *42*(10), 833–839. https://doi.org/10.1038/ng.654
Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M. T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L.,

Jenkins, J., Vendramin, E., Meisel, L. A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., … Rokhsar, D. S. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, *45*(5), 487–494. https://doi.org/10.1038/ng.2586

Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarani, G., Vendramin, E., Paris, R., Aramini, V., Gazza, L., Rossini, L., Bassi, D., Troggio, M., Shu, S., Grimwood, J., Tartarini, S., Dettori, M. T., & Schmutz, J. (2017). The Peach v2.0 release: High-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics*, *18*(1), 225. https://doi.org/10.1186/s12864-017-3606-9

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, H., Jones, B., Li, Z., Frasse, P., Delalande, C., Regad, F., Chaabouni, S., Latché, A., Pech, J.-C., & Bouzayen, M. (2005). The Tomato Aux/IAA Transcription Factor IAA9 Is Involved in Fruit Development and Leaf Morphogenesis. *The Plant Cell*, *17*(10), 2676–2692. https://doi.org/10.1105/tpc.105.033415

Wang, K., Jiao, Z., Xu, M., Wang, Y., Li, R., Cui, X., Gu, L., & Zhang, S. (2016). Landscape and Fruit Developmental Regulation of Alternative Splicing in Tomato by Genome-Wide Analysis. *Horticultural Plant Journal*, *2*(6), 338–350. https://doi.org/10.1016/j.hpj.2017.01.007

Wang, L., Lei, T., Han, G., Yue, J., Zhang, X., Yang, Q., Ruan, H., Gu, C., Zhang, Q., Qian, T., Zhang, N., Qian, W., Wang, Q., Pang, X., Shu, Y., Gao, L., & Wang, Y. (2021). The chromosome-scale reference genome of Rubus chingii Hu provides insight into the biosynthetic pathway of hydrolyzable tannins. *The Plant Journal*, *107*(5), 1466–1477. https://doi.org/10.1111/tpj.15394

Ward, J. A., Bhangoo, J., Fernández-Fernández, F., Moore, P., Swanson, J., Viola, R., Velasco, R., Bassil, N., Weber, C. A., & Sargent, D. J. (2013). Saturated linkage map construction in Rubus idaeususing genotyping by sequencing and genome-independent imputation. *BMC Genomics*, *14*(1), 2. https://doi.org/10.1186/1471-2164-14-2

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing.

Wight, H., Zhou, J., Li, M., Hannenhalli, S., Mount, S. M., & Liu, Z. (2019). Draft Genome Assembly and Annotation of Red Raspberry Rubus Idaeus. *BioRxiv*, 546135. https://doi.org/10.1101/546135

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., & Ma, H. (2017). Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in

the Context of Geological Times and Genome Duplication. *Molecular Biology and Evolution*, *34*(2), 262–281. https://doi.org/10.1093/molbev/msw242

Xie, M., Zhang, J., Tschaplinski, T. J., Tuskan, G. A., Chen, J.-G., & Muchero, W. (2018). Regulation of Lignin Biosynthesis and Its Role in Growth-Defense Tradeoffs. *Frontiers in Plant Science*, *9*. https://www.frontiersin.org/article/10.3389/fpls.2018.01427

Yamaguchi, S., & Kamiya, Y. (2000). Gibberellin Biosynthesis: Its Regulation by Endogenous and Environmental Signals. *Plant and Cell Physiology*, *41*(3), 251–257. https://doi.org/10.1093/pcp/41.3.251

Yao, J.-L., Dong, Y.-H., & Morris, B. A. M. (2001). Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proceedings of the National Academy of Sciences*, *98*(3), 1306–1311. https://doi.org/10.1073/pnas.98.3.1306

Yao, J.-L., Xu, J., Tomes, S., Cui, W., Luo, Z., Deng, C., Ireland, H. S., Schaffer, R. J., & Gleave, A. P. (2018). Ectopic expression of the PISTILLATA homologous MdPI inhibits fruit tissue growth and changes fruit shape in apple. *Plant Direct*, *2*(4), e00051. https://doi.org/10.1002/pld3.51

Yeh, R.-F., Lim, L. P., & Burge, C. B. (2001). Computational Inference of Homologous Gene Structures in the Human Genome. *Genome Research*, *11*(5), 803–816. https://doi.org/10.1101/gr.175701

Yuan, H., Yu, H., Huang, T., Shen, X., Xia, J., Pang, F., Wang, J., & Zhao, M. (2019). The complexity of the Fragaria x ananassa (octoploid) transcriptome by single-molecule long-read sequencing. *Horticulture Research*, *6*(1), 1–14. https://doi.org/10.1038/s41438-019-0126-6

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., Zhang, C., Tian, Y., Liu, G., Gul, H., Wang, D., Tian, Y., Yang, C., Meng, M., Yuan, G., Kang, G., Wu, Y., Wang, K., Zhang, H., … Cong, P. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications*, *10*(1), 1494. https://doi.org/10.1038/s41467-019-09518-x

Zhang, Z., Coenen, H., Ruelens, P., Hazarika, R. R., Al Hindi, T., Oguis, G. K., Vandeperre, A., van Noort, V., & Geuten, K. (2018). Resurrected Protein Interaction Networks Reveal the Innovation Potential of Ancient Whole-Genome Duplication. *The Plant Cell*, *30*(11), 2741–2760. https://doi.org/10.1105/tpc.18.00409

Zhao, Y. (2014). Auxin Biosynthesis. *The Arabidopsis Book / American Society of Plant Biologists*, *12*, e0173. https://doi.org/10.1199/tab.0173

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