UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Evolution of Reproductive Development in Angiosperms

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Dinusha C. Maheepala Mudalige

September 2019

Dissertation Committee:
Dr. Amy Litt, Chairperson
Dr. Patricia Springer
Dr. Jason Stajich

The Dissertation of Dinusha C. Maheepala Mudalige is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

After decades of walking by, over, under and into plants, I succeeded in dedicating several years to studying them. Needless to say I enjoyed doing the research on their evolutionary development, documented in this dissertation, very much. My foremost gratitude in this effort goes to my PhD adviser, Dr. Amy Litt who has had an immense positive impact on my scientific career. Her lab members, Dr. Elizabeth McCarthy, Alex Rajewski, Yi Huang and Glen Morrison, are a joyful bunch who were always there for me.

I received a wealth of knowledge and encouragement during this period from Dr. Patricia Springer who served on all my committees including chairing the qualifying exam, Dr. Jason Stajich who was a member of my dissertation committee, Dr. Jaimie Van Norman, Dr. Carolyn Rasmussen, Dr. Joel Sachs and Dr. Erich Reck who were all part of my qualifying exam committee and Dr. Carol Lovatt who was a member of my guidance committee.

My graduate student advisers, Dr. Linda Walling and Dr. Norman Ellstrand were dedicated to getting me started in the Plant Biology program. In addition to Dr. Walling who ensured my progress until I completed the degree, the graduate student services advisors Laura McGeehan, Jammy Yang and Deidra Kornfeld were instrumental in keeping me on track.

I value the support I received from Dr. Christopher Emerling (Department of Biology, Whittier College, Whittier, CA) with the *FRUITFULL* (*FUL*) gene selection

pressure analyses and Dr. John Heraty (Department of Entomology, UCR) with the gene tree generation in chapter I, Dr. Thomas Girke (Institute for Integrative Genome Biology (IIGB), UCR) with the tomato and *Solanum pimpinellifolium* transcriptome analyses in chapter II, and Dr. Stephen Wright (Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada), Dr. Susan Kalisz and Alannie-Grace Grant (Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN) with the *Collinsia* transcriptome analyses in Chapter III. I am grateful to Dr. Isgouhi Kaloshian (UCR) from whom I learnt the fundamentals of molecular biology and associated techniques. I also appreciate the technical help on numerous occasions by the IIGB faculty and staff including Dr. David Carter, Dr. Glenn Hicks, John Weger, Holly Eckelhoefer, and Clay Clark.

I consider myself fortunate to have been able to mentor a group of promising young scientists who provided me with exceptional help in both the laboratory and the greenhouse. Among them, Jenna Macon, Arman Baghaei, Allen Le and Victor Herrera worked on sequencing Solanaceae *FUL* gene candidates (Chapter I) and Janette Cardenas Bernal, Alexandra Johnson and Ethan Miles worked on generating CRISPR/Cas9 constructs and transformants to isolate *FUL* gene knockouts in tomato and *S. pimpinellifolium*. (The latter project is not part of this dissertation.)

I am indebted to the Botany and Plant Sciences administrative staff including Mariella Valdivia, Deborah Terao, April Meinzer, Henry Gutierrez, Jodie Messin, Jessica Perez, Amy Ricks, Summer Willis and Maria Sedillo for their support throughout my graduate student life at UCR.

The past few years in academia would have been mundane without the friendship and joviality of my fellow graduate students. Last but not least, I am extremely grateful to my family in whom I consistently find comfort, stability and sanity that I can always count on.

For Kasia, *with love*.

ABSTRACT OF THE DISSERTATION

The Evolution of Reproductive Development in Angiosperms

by

Dinusha C. Maheepala Mudalige

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, September 2019
Dr. Amy Litt, Chairperson

Evolutionary shifts in angiosperms have facilitated their dispersal and establishment throughout the world. It is believed that angiosperms have undergone extensive coevolution with the animal pollinator/ dispersal agents. As such, the evolutionary shift to fleshy edible fruit from dry dehiscent ones has occurred numerous times. Remarkably, despite the coevolutionary interdependence between plants and animals, the shift from outcrossing to self-mating has also been common. However, the molecular mechanisms that may underlie either of these shifts has not been established. In the Solanaceae (nightshades) and the Plantaginaceae, fleshy fruits and self-mating, respectively, have evolved multiple times. We investigated the potential molecular underpinning of these shifts using comparative sequence and expression analyses between pre- and post-transition taxa. *FRUITFULL* (*euFUL*) transcription factors have different roles in dry and fleshy fruit development. Our findings suggest that the coding sequence in some Solanaceae *euFUL* gene clades are evolving faster compared to their sister clades. In addition,

we found evidence indicating a potential pseudogenization event in one of these clades. However, we were not able to detect any change in amino acid sequence associated with the transition to fleshy fruit. In the genus *Collinsia* (Plantaginaceae), multiple sister species pairs consist of an outcrosser and a selfer. A change in the developmental timing of the reproductive whorls underlies these evolutionary transitions to selfing. However, the molecular basis of this developmental phenomenon is unknown. We compared expression data across the entire floral development in the two sister taxa, the outcrossing *C. linearis* and the seling *C. rattanii*. Our data revealed there might be an association between putative metal ion binding proteins and the change in the developmental timing in *C. rattanii*. In addition, in agreement with a previous report, our results suggest that putative genes involved in pollen development and pollinator attraction are downregulated in *C. rattanii*.

Table of Contents

Table of Contents (Continued)

Page

List of Figures

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Figures (Continued)

List of Tables

List of Tables (Continued)

List of Tables (Continued)

Introduction

The evolution of reproductive development in angiosperms

The reproductive efficiency of a population is a major factor determining its survival and dispersal. A plethora of reproductive traits have evolved in angiosperms, facilitating their establishment in a diverse range of environments. As angiosperm diversification has been largely driven by their coevolution with animal pollinators or seed dispersers, a majority of these reproductive traits are associated with attracting these agents (Barrett and Willis, 2001; Hu et al., 2008). These include, among the characters related to pollination, a large corolla with numerous color patterns composed of different pigments, fragrant volatile concoctions and nectar, and among those involved in seed dispersal, fruit that have a fleshy edible pericarp, which attracts frugivores and pericarps that develop hooked surfaces or secrete viscous substances, promoting their attachment to animal coats (Du et al., 2009; Johnson and Steiner, 2000; Rosas-Guerrero et al., 2014; Schemske and Bradshaw, 1999; Stewart and Cole, 2005).

Since pollinators (e.g., insects, hummingbirds, etc.) tend to choose the energy sources that require a minimum effort to reach, plant species with large flowers are favored as these organs are easily detectable from a distance, which reduces the foraging time (Spaethe et al., 2001). Various petal colors and patterns created by plant pigments also facilitate pollinator attraction. These pigments consist of carotenoids, anthocyanins and betalains (Grotewold, 2006). Carotenoids are lipid-soluble, yellow to red color pigments derived from the isoprenoid pathway while

the water-soluble anthocyanins and betalains are orange/red to violet/blue and yellow to red pigments, and are derived from flavonoides and tyrosine, respectively (Tanaka et al., 2008). Although both carotenoids and anthocyanins have a wide phylogenetic presence, betalains are only found in the Caryophyllalles (Strack et al., 2003). Floral fragrances may consist of numerous volatiles depending on the species (Dobson, 2006; Levin et al., 2001). More than a thousand such volatile compounds have been identified, indicating the immense potential for diverse interactions that exists in the angiosperms (Knudsen et al., 2006) Floral nectar is almost entirely composed of water, sucrose, fructose and glucose, although their ratios may vary, in addition to minor amounts of amino acids (Heil, 2011; Pyke, 2016). In some species, secondary compounds such as catalpol, nectarin and gelsemine may function as repellents against any nectar robbers, which do not reciprocate by pollinating the flowers (Adler, 2000; Carter and Thornburg, 2004; Pyke, 2016; Stephenson, 1982).

Fleshy fruits (discussed below) contain a pulpy, nutritious pericarp that attracts many species of frugivores and omnivores, which in turn disperse (zoochory) the seeds of these fruit to potentially new environments (Gosper et al., 2005; Kollmann, 2000; Rey and Alcántara, 2000). Zoochory is also achieved via the protrusions such as barbs, spines or hooks that develop on the outer surface of fruits or viscous exudates in certain species (Gorb and Gorb, 2002; Pijl and van der Pijl, 1969).

The frequency of independent evolutionary events that have led to the same repro-

ductive trait in angiosperms may be a measure of its selective advantage. Evidence suggests that fleshy fruits have evolved from dry dehiscent capsules on numerous occasions during the evolution of angiosperms (Bolmgren and Eriksson, 2010). Interestingly, despite the extent of the coevolutionary interdependence between angiosperms and animals, transitions from outcrossing to self-mating in plants, which eliminates any requirement for pollinators, are considered to be common (Barrett, 2002; Sicard and Lenhard, 2011). Thus, angiosperms have displayed a great amount of plasticity in the evolution of reproductive development, which may have been a major force behind their successful establishment across the planet. Exemplifying this immense plasticity, I further investigated the two disparate topics on the transition to fleshy fruit and to self-mating.

The evolution and development of fleshy fruit

Fleshy fruits have evolved in multiple plant orders and likely, more than once within the same lineage from plesiomorphic dry fruit (see Bolmgren and Eriksson (2010) for a comprehensive list). For example, in Solanaceae (nightshades), fleshy fruits have independently evolved in the subfamily Solanoideae as well as the genera *Duboisia* and *Cestrum* (Knapp, 2002). It has been hypothesized that there is an association between the increase in seed mass, which is positively correlated with the probability of embryo survival, and the emergence of fleshy fruit (Bolmgren and Eriksson, 2010; Moles and Westoby, 2002).

Some have suggested that herbivorous dinosaurs during the Cretaceous might have been the first dispersal agents of fleshy fruit (Chang et al., 2002; Llorente

et al., 2016). Still others have predicted that other contemporary animal groups such as early mammals as the likely dispersers (Barrett and Willis, 2001). These hypotheses as well as the current discourses on the preferences of dispersers such as birds is based upon the potential ability of these agents to distinguish the orange/red or purple/black color pigments in ripe fruit (Valenta et al., 2018; Willson and Whelan, 1990).

The pulp of fleshy fruit contains all major components of a diet including carbohydrates, proteins and lipids (Schaefer et al., 2014). However, in some species, there are unpleasant-tasting secondary compounds that may selectively deter seed predators that offer no selective advantage to the plant (Tewksbury and Nabhan, 2001). In contrast, these deterrents might not have any effect on more efficient dispersers. For example, birds are not sensitive to capsaicin, the pungent deterrent found in chili pepper (*Capsicum annuum*) (Tewksbury and Nabhan, 2001).

The molecular mechanisms involved in the evolutionary transitions to fleshy fruit have not been elucidated. However, empirical data exists on some of the genes that have functions in fleshy fruit development. FRUITFULL (FUL), a MADS-box transcription factor has a role in patterning the dehiscence zone in the dry siliques in *Arabidopsis thaliana* (Gu et al., 1998). Evidence suggests a similar function for a *FUL* ortholog in tobacco (*Nicotiana*, Solanaceae) (Smykal et al., 2007). However, in tomato (*Solanum lycopersicum*), a model species for fleshy fruit development, the orthologs of *FUL*, *SlFUL1* and *SlFUL2* have functions in ripening associated carotenoid pigment accumulation (Bemer et al., 2012; Wang et al., 2019). In addi-

tion, *SlFUL1* is involved in the biosynthesis of ethylene, a hormone important for ripening while *SlFUL2* may also have a role in patterning the pericarp during early stages of tomato development (Wang et al., 2019). Therefore, this data on *FUL* and its orthologs in dry and fleshy fruit development suggest that these genes have undergone a change in the evolutionary transition to fleshy fruit in Solanaceae.

FUL limits the expression of *SHATTERPROOF1/2* (*SHP1/2*), which encodes a MADS-box transcription factor, in *A. thaliana* to the valve margins where they are involved in patterning the dehiscence zone (Colombo et al., 2010). However, the orthologs of *SHP1/2*, *TOMATO-AGAMOUS-LIKE1* (*TAGL1*), are expressed throughout the pericarp and are involved in fruit expansion and ethylene induced ripening (Vrebalov et al., 2009). Thus, similar to *SlFUL1/2*, the role of *TAGL1* in fleshy fruit development compared to *SHP1/2* in dry fruit suggests a change in function for these genes in the transition to fleshy fruit.

A number of genes with functions in tomato ripening have been identified. *Colourless non-ripening* (*Cnr*) encodes a SQUAMOSA promoter binding protein-like transcription factor and has a role in the changes in pigmentation and cell adhesion (Chen et al., 2015; Manning et al., 2006). *Cnr* is thought to act upstream of all genes that have a role in ripening discussed here (Bemer et al., 2012; Chen et al., 2015; Karlova et al., 2011). *NONRIPENING* (*NOR*), *RIPENING-INHIBITOR* (*RIN*) and *NEVER-RIPE* (*NR/ETR3*) encode a NAC-domain transcription factor, a MADS-box transcription factor and an ethylene receptor (ETR) family protein, respectively. These control the fruit ripening through the mediation of ethylene (Cantu et al.,

2009; Hackett et al., 2000; Ito et al., 2015; Karlova et al., 2011; Ma et al., 2018; Osorio et al., 2011). *APETALA2a* (*AP2a*) encodes an ethylene responsive factor (ERF) family protein that represses ethylene production while simultaneously inducing carotenoid biosynthesis (Chung et al., 2010). However, currently there no empirical studies on the potential functions of these genes in dry fruit development.

A lack of comparative data for these genes in dry and fleshy fruited species has been a limitation for investigating the molecular mechanisms that may underlie the shift to fleshy fruit. In the Solanaceae, there have been multiple shifts to fleshy fruit as well as a reversal to dry fruit (Knapp, 2002). This plus the availability of multiple sequenced genomes and protocols for genetic manipulation makes this family amenable to elucidating the genetic underpinning of fleshy fruit evolution (Bombarely et al., 2016; Consortium and The Potato Genome Sequencing Consortium, 2011; Tomato Genome Consortium, 2012). I generated sequence data for *FUL* orthologs across the Solanaceae phylogeny to characterize sequence evolution that might be correlated with the transition to fleshy fruit in Solanoideae (Maheepala et al., 2019). We also created transcriptome data for tomato, *S. pimpinellifolium*, the closest wild relative of the cultivated tomato, and desert tobacco (*Nicotiana obtusifolia*) across the entirety of fruit development to identify any differences in molecular mechanisms between the dry and fleshy fruit types. In addition, we conducted comparative transcriptome analyses between cultivated and wild tomato to search for any molecular signatures associated with artificial selection

— domestication has altered the flesh of the tomato by a great degree. Thus, our work encompases data related to the molecular traits integral to fleshy fruit development as well as those with some plasticity. This work is described in chapters 1 and 2 of this dissertation.

The molecular mechanisms underlying the evolutionary transition to self-mating

Flowering plants and pollinator species have undergone extensive coevolution since angiosperm diversification in the Cretaceous (Hu et al., 2008). Despite this, there have been numerous evolutionary transitions from outcrossing to self-mating systems in plants (Barrett, 2002). Changes in the developmental timing of the reproductive whorls underlie these evolutionary shifts to selfing. However, the molecular mechanisms involved in such changes are largely unknown.

Collinsia (Plantaginaceae) is a mixed-mating genus with pairs of sister taxa composed of a predominantly outcrossing and a selfing species (Randle et al., 2009). Thus, these independent evolutionary transitions in the genus, with the advent of affordable genome sequencing, provide an opportunity to investigate the molecular mechanisms involved in the shift to selfing. In the self-mating *Collinsia* species, changes in the developmental timing of the reproductive whorls have resulted in reductions in the spatial separation between the anthers and the stigma (i.e., reduced herkogamy) and the temporal separation between the maturation of the stamens and the pistil (i.e., reduced dichogamy). Studies have suggested that

multiple genetic loci might be involved in the evolutionary shift to selfing (Holtsford and Ellstrand, 1992; Shore and Barrett, 1990). Still others have hypothesized that as few as two loci might be enough when those loci are pleiotropic or tightly linked (Fishman et al., 2002; Fishman and Stratton, 2004). The locus *se2.1* on tomato chromosome 2 consists of the five genes that influence the style length (*STYLE2.1*), stamen length (*STAMEN2.1*, *STAMEN2.2*, and *STAMEN2.3*) and stamen architecture (*DEHISCENCE2.1*) (Chen and Tanksley, 2004; Pan et al., 2017). However, the genes that might be involved in similar functions in *Collinsia* have not been identified.

In addition to the traits directly underlying the evolutionary shift to selfing, a number of phenotypes that emerge post-transition, collectively called the "selfing syndrome," have been reported (Sicard and Lenhard, 2011; Vos et al., 2014). These include the breakdown of biochemical self-incompatibility (SI), and reductions in the pollen-to-ovule ratio and the floral size. Three different biochemical mechanisms related to the breakdown of SI in three different plant families have been reported (Fujii et al., 2016). In the Solanaceae, a style encoded glycoprotein S-RNase and multiple pollen encoded F-box proteins disrupt self-pollen tube growth via ribonuclease/detoxification activity (Goldraij et al., 2006; Kubo et al., 2010; Lee et al., 1994; McClure et al., 1990, 2011; Murfett et al., 1994). In the Brassicaceae, the pollen coat encoded S-locus protein 11 (SP11) and the stigma encoded S-locus receptor kinase (SRK) causes the rejection of self-pollen (Kachroo et al., 2001; Shimosato et al., 2007; Takayama et al., 2001). In the Papaveraceae,

when *Papaver rhoeas* female and male style S (PrsS) proteins from the same haplotype interact, potential cytoplasmic calcium increase and reactive oxygen species induction result in the breakdown of microtubules and the fragmentation of DNA, ultimately leading to apoptosis (de Graaf et al., 2006; Thomas and Franklin-Tong, 2004; Wheeler et al., 2010; Wilkins et al., 2015). This suggests that a diverse set of biochemical mechanism regarding the breakdown of SI have evolved in plant families.

I generated RNAseq libraries spanning the entire floral development of the two *Collinsia* sister species, the predominantly outcrossing *C. linearis* and selfing *C. rattanii*, and searched for any molecular signatures that might coincide with the evolutionary transition to selfing. This work is described in chapter 3 of this dissertation.

Adler, L. S. (2000). The ecological significance of toxic nectar. Oikos 91, 409–420. doi:10.1034/j.1600-0706.2000.910301.x.

Auld, J. R. (2010). The effects of predation risk on mating system expression in a freshwater snail. Evolution 64, 3476–3494. doi:10.11 11/j.1558-5646.2010.01079.x.

Barrett, P. M., and Willis, K. J. (2001). Did dinosaurs invent flowers? Dinosaur–angiosperm coevolution revisited. Biological Reviews of the Cambridge Philosophical Society 76, 411–447. doi:10.1017/s1464793101005735.

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. Nature Reviews Genetics 3, 274–284. doi:10.1038/nrg776.

Bemer, M., Karlova, R., Ballester, A. R., Tikunov, Y. M., Bovy, A. G., Wolters-Arts, M., et al. (2012a). The tomato *FRUITFULL* homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. Plant Cell 24, 4437–4451.

Bolmgren, K., and Eriksson, O. (2010). Seed mass and the evolution of fleshy fruits in angiosperms. Oikos 119, 707–718. doi:10.1111/j.1600-0706.200 9.17944.x.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat Plants 2, 16074.

Brys, R., de Crop, E., Hoffmann, M., and Jacquemyn, H. (2011). Importance of autonomous selfing is inversely related to population size and pollinator availability in a monocarpic plant. American Journal of Botany 98, 1834–1840. doi:10.3732/ajb.1100154.

Cantu, D., Blanco-Ulate, B., Yang, L., Labavitch, J. M., Bennett, A. B., and Powell, A. L. T. (2009). Ripening-regulated susceptibility of tomato fruit to *Botrytis cinerea* requires *NOR* but not *RIN* or ethylene. Plant Physiol. 150, 1434–1449.

Carter, C., and Thornburg, R. W. (2004). Is the nectar redox cycle a floral defense against microbial attack? Trends Plant Sci. 9, 320–324.

Chang, B. S. W., Jönsson, K., Kazmi, M. A., Donoghue, M. J., and Sakmar, T. P. (2002). Recreating a functional ancestral archosaur visual pigment. Mol. Biol. Evol. 19, 1483–1489.

Chen, K.-Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of *se2.1*: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. Genetics 168, 1563–1573.

Chen, W., Kong, J., Lai, T., Manning, K., Wu, C., Wang, Y., et al. (2015). Tuning *LeSPL-CNR* expression by SlymiR157 affects tomato fruit ripening. Sci. Rep. 5, 7852.

Chung, M.Y., Vrebalov, J., Alba, R., Lee, J., McQuinn, R., Chung, J.D., et al. (2010). A tomato (*Solanum lycopersicum*) *APETALA2/ERF* gene, *SlAP2a*, is a negative regulator of fruit ripening. Plant J. 64, 936–947.

Colombo, M., Brambilla, V., Marcheselli, R., Caporali, E., Kater, M. M., and Colomb o, L. (2010). A new role for the *SHATTERPROOF* genes during *Arabidopsis* gynoecium development. Dev. Biol. 337, 294–302.

Consortium, T. P. G. S., and The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature 475, 189–195. doi:10.1038/nature10158.

de Graaf, B. H. J., Rudd, J. J., Wheeler, M. J., Perry, R. M., Bell, E. M., Osman, K., et al. (2006). Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. Nature 444, 490–493.

Dobson, H. (2006). Relationship between floral fragrance composition and type of pollinator. Biology of Floral Scent, 147–198. doi:10.1201/9781420 004007.sec4.

Du, Y., Mi, X., Liu, X., Chen, L., and Ma, K. (2009). Seed dispersal phenology and dispersal syndromes in a subtropical broad-leaved forest of China. For-est Ecology and Management 258, 1147–1152. doi:10.1016/j.foreco.20 09.06.004.

Fishman, L., Kelly, A. J., and Willis, J. H. (2002). Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. Evolution 56, 2138–2155. doi:10.1111/j.001 4-3820.2002.tb00139.x.

Fishman, L., and Stratton, D. A. (2004). The genetics of floral divergence and postzygotic barriers between outcrossing and selfing populations of *Are-naria uniflora* (Caryophyllaceae). Evolution 58, 296–307.

Fujii, S., Kubo, K.-I., and Takayama, S. (2016). Non-self- and self-recognition models in plant self-incompatibility. Nat Plants 2, 16130.

Goldraij, A., Kondo, K., Lee, C. B., Hancock, C. N., Sivaguru, M., Vazquez-Santana, S., et al. (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439, 805–810.

Gorb, E., and Gorb, S. (2002). Contact separation force of the fruit burrs in four plant species adapted to dispersal by mechanical interlocking. Plant Phys-iology and Biochemistry 40, 373–381. doi:10.1016/s0981-9428(02)01381-5.

Gosper, C. R., Stansbury, C. D., and Vivian-Smith, G. (2005). Seed dispersal of fleshy-fruited invasive plants by birds: contributing factors and management options. Diversity Distributions 11, 549–558. doi:10.1111/j.1366-9516.2005.00195.x.

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. Annu. Rev. Plant Biol. 57, 761–780.

Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125, 1509–1517.

Hackett, R. M., Ho, C. W., Lin, Z., Foote, H. C., Fray, R. G., and Grierson, D. (2000). Antisense inhibition of the *Nr* gene restores normal ripening to the tomato *Never-ripe* mutant, consistent with the ethylene receptor-inhibition model. Plant Physiol. 124, 1079–1086.

Heil, M. (2011). Nectar: generation, regulation and ecological functions. Trends Plant Sci. 16, 191–200.

Holtsford, T. P., and Ellstrand, N. C. (1992). Genetic and environmental variation in floral traits affecting outcrossing rate in *Clarkia tembloriensis* (Onagraceae). Evolution 46, 216–225. doi:10.11 11/j.1558-5646.1992.tb019 96.x.

Hu, S., Dilcher, D. L., Jarzen, D. M., and Winship Taylor, D. (2008). Early steps of angiosperm pollinator coevolution. Proc. Natl. Acad. Sci. U. S. A. 105, 240–245.

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., and Toki, S. (2015). CRISPR/Cas9-mediated mutagenesis of the *RIN* locus that regulates tomato fruit ripening. Biochem. Biophys. Res. Commun. 467, 76–82.

Johnson, S. D., and Steiner, K. E. (2000). Generalization versus specialization in plant pollination systems. Trends Ecol. Evol. 15, 140–143.

Jorgensen, R., and Arathi, H. S. (2013). Floral longevity and autonomous selfing are altered by pollination and water availability in *Collinsia heterophylla*. Ann. Bot. 112, 821–828.

Kachroo, A., Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (2001). Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293, 1824–1826.

Karlova, R., Rosin, F. M., Busscher-Lange, J., Parapunova, V., Do, P. T., Fernie, A. R., et al. (2011). Transcriptome and metabolite profiling show that *APETALA2a* is a major regulator of tomato fruit ripening. Plant Cell 23, 923–941.

Knapp, S. (2002). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J. Exp. Bot. 53, 2001–2022.

Knudsen, J. T., Eriksson, R., Gershenzon, J., and Ståhl, B. (2006). Diversity and distribution of floral scent. The Botanical Review 72, 1–120. doi:10.1663/0006-8101(2006)72[1:dadofs]2.0.co;2.

Kollmann, J. (2000). Dispersal of fleshy-fruited species: a matter of spatial scale? Perspectives in Plant Ecology, Evolution and Systematics 3, 29–51. doi:10.1078/1433-8319-00003.

Kubo, K.I., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., et al. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science 330, 796–799.

Lafuma, L., and Maurice, S. (2007). Increase in mate availability without loss of self-incompatibility in the invasive species *Senecio inaequidens* (Asteraceae). Oikos 116, 201–208. doi:10.1111/j.2006.0030-1299.15220.x.

Lee, H. S., Huang, S., and Kao, T. (1994). S proteins control rejection of incompatible pollen in *Petunia inflata*. Nature 367, 560–563.

Levin, R. A., Raguso, R. A., and McDade, L. A. (2001). Fragrance chemistry and pollinator affinities in Nyctaginaceae. Phytochemistry 58, 429–440. doi:10.1016/s0031-9422(01)00257-6.

Llorente, B., D'Andrea, L., and Rodríguez-Concepción, M. (2016). Evolutionary recycling of light signaling components in fleshy fruits: New insights on the role of pigments to monitor ripening. Front. Plant Sci. 7, 263.

Lloyd, D. G. (1992). Self- and cross-fertilization in plants. II. The selection of self- fertilization. International Journal of Plant Sciences 153, 370–380. doi:10.1086/297041.

Maheepala, D. C., Emerling, C. A., Rajewski, A., Macon, J., Strahl, M., Pabón-Mora, N., et al. (2019). Evolution and diversification of *FRUITFULL* genes in Solanaceae. Frontiers in Plant Science 10. doi:10.3389/fpls.2019.00043.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. 38, 948–952.

Ma, X., Balazadeh, S., and Mueller-Roeber, B. (2018). Tomato fruit ripening factor NOR controls leaf senescence. doi:10.1101/436899.

McClure, B. A., Gray, J. E., Anderson, M. A., and Clarke, A. E. (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature 347, 757–760. doi:10.1038/347757a0.

McClure, B., Cruz-García, F., and Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. Ann. Bot. 108, 647–658.

Moles, A. T., and Westoby, M. (2002). Seed addition experiments are more likely to increase recruitment in larger-seeded species. Oikos 99, 241–248. doi:10.1034/j.1600-0706.2002.990204.x.

Murfett, J., Atherton, T. L., Mou, B., Gasser, C. S., and McClure, B. A. (1994). S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. Nature 367, 563–566.

Osorio, S., Alba, R., Damasceno, C. M. B., Lopez-Casado, G., Lohse, M., Zanor, M. I., et al. (2011). Systems biology of tomato fruit development: combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. Plant Physiol. 157, 405–425.

Pan, C., Ye, L., Zheng, Y., Wang, Y., Yang, D., Liu, X., et al. (2017). Identification and expression profiling of microRNAs involved in the stigma exsertion under high-temperature stress in tomato. BMC Genomics 18, 843.

Pijl, L. van der, and van der Pijl, L. (1969). Principles of dispersal in higher plants. doi:10.1007/978-3-662-00799-0.

Pyke, G. H. (2016). Floral nectar: Pollinator attraction or manipulation? Trends in Ecology & Evolution 31, 339–341. doi:10.1016/j.tree.2016.02.013.

Randle, A. M., Slyder, J. B., and Kalisz, S. (2009). Can differences in autonomous selfing ability explain differences in range size among sister-taxa pairs of *Collinsia* (Plantaginaceae)? An extension of Baker's Law. New Phytol. 183, 618–629.

Rey, P. J., and Alcántara, J. M. (2000). Recruitment dynamics of a fleshy-fruited plant (*Olea europaea*): connecting patterns of seed dispersal to seedling establishment. Journal of Ecology 88, 622–633. doi:10.1046/j.13 65-2745.2000.00472.x.

Rosas-Guerrero, V., Aguilar, R., Martén-Rodríguez, S., Ashworth, L., Lopezaraiza-Mikel, M., Bastida, J. M., et al. (2014). A quantitative review of pollination syndromes: do floral traits predict effective pollinators? Ecol. Lett. 17, 388–400.

Schaefer, H. M., Valido, A., and Jordano, P. (2014). Birds see the true colours of fruits to live off the fat of the land. Proc. Biol. Sci. 281, 20132516.

Schemske, D. W., and Bradshaw, H. D., Jr (1999). Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). Proc. Natl. Acad. Sci. U. S. A. 96, 11910–11915.

Shimosato, H., Yokota, N., Shiba, H., Iwano, M., Entani, T., Che, F.S., et al. (2007). Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in *Brassica* self-incompatibility. Plant Cell 19, 107–117.

Shore, J. S., and Barrett, S. C. H. (1990). Quantitative genetics of floral characters in homostylous *Turnera ulmifolia* var. *angustifolia* Willd. (Turneraceae). Heredity 64, 105–112. doi:10.1038/hdy.1990.13.

Sicard, A., Kappel, C., Lee, Y. W., Woźniak, N. J., Marona, C., Stinchcombe, J. R., et al. (2016). Standing genetic variation in a tissue-specific enhancer underlies selfing-syndrome evolution in *Capsella*. Proceedings of the National Academy of Sciences 113, 13911–13916. doi:10.1073/pnas.161339 4113.

Sicard, A., and Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Ann. Bot. 107, 1433–1443.

Smykal, P., Gennen, J., De Bodt, S., Ranganath, V., and Melzer, S. (2007). Flowering of strict photoperiodic *Nicotiana* varieties in non-inductive conditions by transgenic approaches. Plant Mol. Biol. 65, 233–242.

Spaethe, J., Tautz, J., and Chittka, L. (2001). Visual constraints in foraging bumblebees: flower size and color affect search time and flight behavior. Proc. Natl. Acad. Sci. U. S. A. 98, 3898–3903.

Stephenson, A. G. (1982). Iridoid glycosides in the nectar of *Catalpa speciosa* are unpalatable to nectar thieves. Journal of Chemical Ecology 8, 1025–1034. doi:10.1007/bf00987883.

Stewart, K. M., and Cole, D. (2005). The commercial harvest of devil's claw (*Harpago phytum* spp.) in southern Africa: The devil's in the details. Journal of Ethnopharmacology 100, 225–236. doi:10.1016/j.jep.2005.07.004.

Strack, D., Vogt, T., and Schliemann, W. (2003). Recent advances in betalain research. Phytochemistry 62, 247–269.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., et al. (2001). Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility. Nature 413, 534–538. doi:10.1038/35097104.

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 54, 733–749.

Tewksbury, J. J., and Nabhan, G. P. (2001). Directed deterrence by capsaicin in chillies. Nature 412, 403–404. doi:10.1038/35086653.

Thomas, S. G., and Franklin-Tong, V. E. (2004). Self-incompatibility triggers programmed cell death in *Papaver* pollen. Nature 429, 305–309.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641.

Valenta, K., Nevo, O., and Chapman, C. A. (2018). Primate fruit color: Useful concept or alluring myth? International Journal of Primatology 39, 321–337. doi:10.1007/s10764-018-0025-y.

Vos, J. M. de, de Vos, J. M., Wüest, R. O., and Conti, E. (2014). Small and ugly? Phylogenetic analyses of the "selfing syndrome" reveal complex evolutionary fates of monomorphic primrose flowers. Evolution 68, 1042–1057. doi:10.1111/evo.12331.

Vrebalov, J., Pan, I. L., Arroyo, A. J. M., McQuinn, R., Chung, M., Poole, M., et al. (2009). Fleshy fruit expansion and ripening are regulated by the tomato *SHATTERPROOF* gene *TAGL1*. Plant Cell 21, 3041–3062.

Wang, R., Tavano, E. C. da R., Lammers, M., Martinelli, A. P., Angenent, G. C., and de Maagd, R. A. (2019). Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Sci. Rep. 9, 1696.

Wheeler, M. J., Vatovec, S., and Franklin-Tong, V. E. (2010). The pollen S-determina nt in *Papaver*: comparisons with known plant receptors and protein ligand partners. J. Exp. Bot. 61, 2015–2025.

Wilkins, K. A., Bosch, M., Haque, T., Teng, N., Poulter, N. S., and Franklin-Tong, V. E. (2015). Self-incompatibility-induced programmed cell death in field poppy pollen involves dramatic acidification of the incompatible pollen tube cytosol. Plant Physiol. 167, 766–779.

Willson, M. F., and Whelan, C. J. (1990). The evolution of fruit color in fleshy-fruited plants. The American Naturalist 136, 790–809. doi:10.1086/285132.

Chapter I:

Evolution and Diversification of *FRUITFULL* Genes in Solanaceae

Abstract

Ecologically and economically important fleshy edible fruits have evolved from dry fruit numerous times during angiosperm diversification. However, the molecular mechanisms that underlie these shifts are unknown. In the Solanaceae there has been a major shift to fleshy fruits in the subfamily Solanoideae. Evidence suggests that an ortholog of *FRUITFULL* (*FUL*), a transcription factor that regulates cell proliferation and limits the dehiscence zone in the silique of *Arabidopsis*, plays a similar role in dry-fruited Solanaceae. However, studies have shown that *FUL* orthologs have taken on new functions in fleshy fruit development, including regulating elements of tomato ripening such as pigment accumulation. *FUL* belongs to the core eudicot *euFUL* clade of the angiosperm *AP1/FUL* gene lineage. The *euFUL* genes fall into two paralogous clades, *euFULI* and *euFULII*. While most core eudicots have one gene in each clade, Solanaceae have two: *FUL1* and *FUL2* in the former, and *MBP10* and *MBP20* in the latter. We characterized the evolution of the *euFUL* genes to identify changes that might be correlated with the origin of fleshy fruit in Solanaceae. Our analyses revealed that the Solanaceae *FUL1* and *FUL2* clades probably originated through an early whole genome multiplication event. By contrast, the data suggest that the *MBP10* and *MBP20* clades are the result of a later tandem duplication event. *MBP10* is expressed at weak to moderate levels, and its atypical short first intron lacks putative transcription factor binding

sites, indicating possible pseudogenization. Consistent with this, our analyses show that *MBP10* is evolving at a faster rate compared to *MBP20*. Our analyses found that Solanaceae *euFUL* gene duplications, evolutionary rates, and changes in protein residues and expression patterns are not correlated with the shift in fruit type. This suggests deeper analyses are needed to identify the mechanism underlying the change in *FUL* ortholog function.

Introduction

Fleshy fruits are agriculturally and economically important plant organs that have evolved from dry fruits many times during angiosperm evolution. However, the genetic changes that are required for this shift to occur are as yet unknown (Bolmgren and Eriksson, 2010). In the agriculturally, pharmacologically, and horticulturally important plant family Solanaceae (nightshades), there was a shift to fleshy fruit in the subfamily Solanoideae from plesiomorphic dry fruit (Figure 1.1) (Knapp, 2002). In the family two independent transitions to fleshy fruits have also occurred in the genera *Duboisia* (subfamily Anthocercideae) and *Cestrum* (subfamily Cestroideae), as well as a reversal to dry fruit in the genus *Datura* (subfamily Solanoideae) (Knapp, 2002).

Evidence from tomato (*Solanum lycopersicum*, subfamily Solanoideae) indicates that FRUITFULL (FUL) transcription factors (TFs) have novel functions in fleshy fruit development compared to *Arabidopsis* (Brassicaceae) and *Nicotiana* (Solanaceae, subfamily Nicotianoideae) (Gu et al., 1998; Smykal et al., 2007; Bemer et al., 2012; Shima et al., 2013, 2014; Wang et al., 2014). FUL is a MADS-box TF

that plays pleiotropic roles in both reproductive and vegetative development in the model plant Arabidopsis thaliana (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000; Rajani and Sundaresan, 2001; Melzer et al., 2008). *FUL* controls cell proliferation in the fruit valves and spatially limits the formation of the dehiscence zone in the dry silique of *A. thaliana*, enabling the mature fruits to dehisce (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000, 2004; Rajani and Sundaresan, 2001). Overexpression of a *Nicotiana tabacum FUL* ortholog in woodland tobacco (*Nicotiana sylvestris*) resulted in indehiscent fruits with reduced lignification at the dehiscence zones, suggesting a role similar to that observed in silique development in *A. thaliana* (Smykal et al., 2007). Several groups have examined the function of *euFUL* genes, the core-eudicot clade to which *FUL belongs*, in tomato (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). All studies showed defects in fruit pigmentation during ripening when *FUL* ortholog expression was downregulated, and some studies also suggested roles in ethylene production and pericarp and cuticle thickness (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). These data indicate that *euFUL* genes are controlling different processes in dry and fleshy fruits in the Solanaceae.

Early in the diversification of core-eudicots, there was a duplication in the *euFUL gene* clade, which resulted in the *euFULI* and *euFULII* clades (Litt and Irish, 2003; Shan et al., 2007). The *A. thaliana FUL* gene belongs to the *euFULI* clade while its paralog, *AGL79* which plays a role in lateral root development, branching, leaf morphology, and transition to flowering, belongs to the *euFULII* clade (Gao et

al., 2018). The *euFULI* clade has duplicated in Solanaceae resulting in two subclades, designated here as *FUL1* and *FUL2*; likewise the *euFULII* clade has two Solanaceae-specific subclades, here designated *MBP10* and *MBP20* (Hileman et al., 2006; Bemer et al., 2012; The Tomato Genome Consortium, 2012). We studied the evolution of *euFUL* genes in Solanaceae to characterize patterns of selection, duplication, and sequence evolution to identify changes that might be correlated with the shift to fleshy fruit. We tested the following hypotheses: (1) following the duplication of *euFUL* genes, there was a relaxation of selection in some or all of the resulting clades that resulted in sequence diversification; (2) changes in amino acid sequences are correlated with the origin of fleshy fruit. Although we found several sites showing changes in amino acid residues that might have resulted in changes in protein function, none of these were associated with the evolution of fleshy fruit. Consistent with our hypothesis, we found that the *FUL1* and *MBP10* genes are evolving at significantly faster rates in comparison to *FUL2* and *MBP20*. In combination with the relatively weak expression of *MBP10* and loss of potential regulatory elements, our data suggest that the *MBP10* lineage may be undergoing pseudogenization.

Materials and Methods

Plant Material for Sequencing

Sources of plant and tissue material for sequencing are listed in Table 1.3. Plants were grown in temperature controlled glasshouses at University of California, Riverside (UCR), The New York Botanical Garden, NY (NYBG), and The University of

Antioquia, Colombia (UdeA) or collected from the grounds at UCR and the Universidad de Antioquia or the field at Parque Arvi, Vereda Santa Elena, El Tambo, Colombia.

For ease of reference and to simplify language, throughout the paper, members of Solanoideae, including the dry-fruited *Datura*, will be referred to as "fleshy-fruited species" (rather than "fleshy-fruited species and *Datura*"). Likewise non-Solanoideae, including the fleshy-fruited *Cestrum* and *Duboisia*, will be referred to as "dry-fruited species" (rather than "dry-fruited species and *Cestrum* and *Duboisia*").

RNA Isolation, cDNA Synthesis/Library Preparation, and Sequencing

RNA was extracted from fruit, floral/inflorescence or leaf tissue using RNeasy Plant Mini Kits (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. For *Grabowskia glauca*, *Dunalia spinosa*, *Fabiana viscosa*, and *Salpiglossis sinuata* RNA extractions, lysis buffer RLC was used instead of RLT and 2.5% (w/v) polyvinylpyrrolidone (PVP) was added. The RLT buffer was used for extracting RNA from all other species. RNA quality was checked using a BioSpectrometer Basic (Eppendorf, Hamburg, Germany) and stored at –80°C. cDNA was synthesized using SuperScript III Reverse Transcriptase (Thermo Fisher, San Diego, CA, United States) according to the manufacturer's protocol and the product was checked by amplifying *ACTIN*. Clade-specific degenerate primers were designed to target specific *euFUL* gene homologs based on conserved regions in Solanaceae *euFUL gene* alignments (Table 1.7). PCR was run for two initial cycles with an anneal-

ing temperature between 40 and 45°C followed by 30 cycles at 55°C annealing temperature. The PCR products were visualized on a 1% agarose gel. If multiple amplicon band sizes were present, the annealing temperature of the first two cycles was increased until only one product size was achieved.

PCR products were purified using QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's protocol. The purified product was then cloned using TOPO TA Cloning Kit (Life Technologies, Carlsbad, CA, United States) according to the manufacturer's protocol, and the ligated plasmids were transformed into chemically competent TOP10 strain of *Escherichia coli*. Transformants were plated on LB plates with kanamycin selection (50 μg/mL) coated with 40 μL of 25 mg/mL X-Gal and IPTG, and incubated at 37°C overnight. Individual positive (white) colonies were used as templates in amplification with M13F and M13R primers (Life Technologies, Carlsbad, CA, United States) to identify those colonies with inserts of the expected size between 500 bp and 1 kb. These were grown overnight in 5 mL liquid LB medium supplemented with kanamycin (50 μg/mL) in an incubator-shaker at 250 RPM and 37°C. Plasmids were extracted from the liquid cultures using Plasmid Miniprep Kit (QIAGEN) according to the manufacturer's protocol, and sequenced using M13 reverse primer at the Institute for Integrative Genome Biology (IIGB) at UCR or Eton Bioscience, Inc. (San Diego, CA, United States).

For library preparation, RNA quality was checked using a Bioanalyzer (Agilent, Santa Clara, CA, United States). RNAseq library preparation was done accord-

ing to the manufacturer's Poly(A) mRNA Magnetic Isolation Module protocol for NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England BioLabs, Ipswich, MA, United States). *Cestrum diurnum*, *C. nocturnum*, and *Schizanthus grahamii* libraries were sequenced on an Illumina NextSeq v2 platform with high-output runs of 75 bp paired-end reads while *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis sinuata* libraries were sequenced on an Illumina NextSeq v2 platform with high-output runs of both 75 bp paired-end reads and 150 bp single-end reads at IIGB, UCR. *Nicotiana obtusifolia* libraries were generated at NYBG and sequenced at the Beijing Genomics Institute (Shenzhen, China), and *Brunfelsia australis* and *Streptosolen jamesonii* libraries (Ortiz-Ramírez et al., 2018) were generated at UdeA and sequenced at Macrogen (Korea). All resulting *euFUL* sequences from both degenerate primer PCRs and transcriptomes are listed in Table 1.3. Individual sequences from PCR-based methods have been deposited in the GenBank (for accession numbers, see Table 1.3) and transcriptome data for *N. obtusifolia*, *C. diurnum*, *C. nocturnum*, *D. spinosa*, *F. viscosa*, *G. glauca*, and *S. grahamii* have been deposited on the SolGenomics network (ftp://ftp.solgenomics.net/manuscripts/Litt_2018).

Mining *euFUL* Sequences From de novo Transcriptome Assembly and Databases

For transcriptome assembly, raw paired-end reads and single-end reads from Illumina sequencing were first quality trimmed using Trimmomatic v0.36 (Bolger et al., 2014) or TrimGalore (Krueger, 2017) and de novo assembled on the UCR

High Performance Computing Cluster (HPCC) using the default settings of Trinity v2.4.0 (Grabherr et al., 2011). *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis sinuata* libraries were assembled by combining both 75 bp paired-end and 150 bp single-end reads. Each assembled transcriptome was then used to create a custom Basic Local Alignment Tool (BLAST) (Altschul et al., 1990) database. The BLAST database for each species was queried on the HPCC with both blastn and tblastx using all available sequences in our *euFUL* sequence file using a UNIX command line that sequentially matched each sequence in our query file against the database (BLAST® Command Line Applications User Manual, 2008). BLAST analyses were also conducted on the NCBI (https://www.ncbi.nlm.nih.gov/blast) (NCBI Resource Coordinators, 2017) and oneKP (https://db.cngb.org/blast4onekp) (Matasci et al., 2014) databases using *A. thaliana FUL* and various Solanaceae *FUL* homologs as query. Matching output sequences (Table 1.3) from both transcriptomes assemblies and database mining were further confirmed by compiling a gene tree as described below. We confirmed the accuracy of our sequences using gene specific primers and Sanger sequencing. Unless specified otherwise, all sequences referred to in this manuscript are the full or partial mRNA sequences.

Gene-Tree Generation

The Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) tool was used to align *euFUL* sequences (Table 1.8). The appropriate model for tree building, GTR+G, was determined with jModelTest 2.0 (Darriba et al., 2012).

Ten independent maximum likelihood (ML) analyses starting with random trees were performed using GARLI v2.1 (Genetic Algorithm for Rapid Likelihood Inference) (Bazinet et al., 2014). *euFUL* genes from Convolvulaceae (Convolvulus, *Cuscuta* and *Ipomoea species*), which were retrieved from the oneKP database (https://sites.google.com/a/ualberta.ca/onekp), were designated as the outgroup in each analysis, which meant these sequences were automatically excluded from the ingroup clades. Each ML run was set to terminate when there was no significantly better scoring topology for 20,000 consecutive generations. The ten resulting trees were checked for agreement by calculating the pairwise Robinson–Foulds distance using 'ape' and 'phangorn' packages on R (Robinson and Foulds, 1981; Paradis et al., 2004; Schliep, 2010; R Core Team, 2018). The tree with the largest ML value was chosen as the starting tree in a bootstrap analysis involving 1,000 replicates. The results of the replicates were summarized and bootstrap values were calculated using SumTrees tool of DendroPy package on Python ver. 2.7 (Python Language Reference, 2010; Sukumaran and Holder, 2010) or Geneious 10.2 (Darling et al., 2010; Kearse et al., 2012).

Any sequences that did not group with any of the subclades were aligned with the paralogs to investigate whether these may have been splice isoforms. Any such isoform was expected to have large insertions/deletions at splice junctions. None were noted.

Selection Pressure Analysis

The CODEML program within the Phylogenetic Analysis by Maximum Likelihood (PAML) (Yang, 1997) v 1.3 (http://abacus.gene.ucl.ac.uk/software/paml.html) software package was run on the HPCC at UCR to analyze the selection pressure acting on *euFUL genes*. These analyses were performed to test if different gene lineages as well as sub-groups within those lineages were evolving at significantly different rates. Further scenarios were considered in which each gene, the transition branches from dry to fleshy fruit trait, or specific sites in the sequences were tested for significantly different rates of evolution. Model 0 (M0) was used to estimate a single evolutionary rate for all genes when the clades being analyzed encompassed the entire dataset. Model 2 (M2) was used when two groups encompassing the entire data set have different rates or when two groups that are being compared do not encompass the entire data set. In the latter case, the two clades being compared were grouped together to obtain a single evolutionary rate in comparison to the rate for the remaining data (background). This single rate for the two clades grouped together was then compared to the rates for each clade separately to determine if the separate rates were significantly different from the combined rate. The test statistic, $2\Delta L$ (twice the difference of the resulting log-likelihood values), and the degrees of freedom (df), were then used in chi-squared tests to check for statistical significance. In any comparison where the P-value was less than 0.05, the second hypothesis was considered to have the better fit than the first, implying there is statistical power to support that the gene clades are evolving at different rates. Since Solanaceae has a well-supported phylogeny (Olmstead et al.,

2008; Särkinen et al., 2013), for PAML analyses, the branches of the gene-tree described above were adjusted to match the phylogenetic relationships of the species included in the analysis. In the *euFUL gene* groups that are evolving faster, sites undergoing positive selection were analyzed using mixed effects model of evolution (MEME; http://datamonkey.org/meme) (Murrell et al., 2012).

The gene alignments for the *euFUL* subclades that are evolving at statistically significantly faster than the other subclades were translated using AliView (Larsson, 2014). In these protein alignments, the sites that changed from hydrophilic to hydrophobic or vice versa were identified manually. Those changes that might have been functionally deleterious versus those that might have been neutral were identified using the PROVEAN Protein tool (http://provean.jcvi.org) (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015).

MADS (M), intervening/interacting (I) and keratin-like (K) domains of the proteins were identified using a published MADS-box protein model (Kaufmann et al., 2005).

The structure of M, I, and K domains of tomato FUL1 and MBP10 were predicted using PHYRE2 server (http://www.sbg.bio.ic.ac.uk/∼phyre2) (Kelley et al., 2015).

***MBP10/MBP20* Synteny and Intron Analyses**

One-million-base-pair regions surrounding tomato *MBP10* and *MBP20* were analyzed for synteny using the progressive Mauve alignment tool on Geneious 10.2 (https://www.geneious.com) (Darling et al., 2010; Kearse et al., 2012).

Putative TF binding site searches for *MBP10* and *MBP20* first introns were done using PROMO 3.0 (http://alggen.lsi.upc.es/recerca/frame-recerca.html) at a maximum matrix dissimilarity rate of zero (Messeguer et al., 2002; Farré et al., 2003).

Solanaceae *euFUL* Expression Analysis

The expression patterns of *euFUL* genes were analyzed using RT-PCR data for *Solanum pimpinellifolium* organs, and transcriptome data from this study for five stages of fruit development in *S. pimpinellifolium* and tomato following stages identified by Gillaspy et al. (1993) and Tanksley (2004). Additional expression data were obtained from the eFP browser (http://bar.utoronto.ca) for tomato, *S. pimpinellifolium*, potato (*S. tuberosum*) (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012) and from the Gene Expression Atlas (http://benthgenome.qut.edu.au) for *Nicotiana benthamiana* (Nakasugi et al., 2014), and other publications (Hileman et al., 2006; Burko et al., 2013).

The TF binding sites for the 2 kb and 5 kb regions upstream of the *euFUL* gene transcription start sites of tomato (GCF_000188115.4) (The Tomato Genome Consortium, 2012), potato (GCF_000226075.1) (Potato Genome Sequencing Consortium et al., 2011) and *N. sylvestris* (GCA_000393655.1) (Sierro et al., 2013) were predicted using PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw) (Chang et al., 2008). Due to the limitations of available contig length, the longest promoter region used for *N. sylvestris MBP10* was 3.3 kb.

Results

Solanaceae have four clades of *euFUL* genes

Our analysis consisted of 106 sequences from 45 species in 26 genera obtained from direct amplification, transcriptomes, and online genomic databases (Table 1.3). Of these, 64 sequences belonged to species from the Solanoideae, characterized by the derived fleshy fruit, whereas the other 42 sequences were from species with the ancestral dry-fruit trait. We designated *euFUL* genes from Convolvulaceae, the sister-group of Solanaceae, as the outgroup (Stefanović et al., 2003). For many species in the analysis, we have an incomplete set of paralogs; however, we had substantial and diverse representation from across the phylogeny, which allows us to test hypotheses regarding the evolution of this gene lineage in Solanaceae.

We used maximum likelihood methods (Garli v2.1) (Bazinet et al., 2014) to reconstruct the relationships of Solanaceae *euFUL* genes (Figure 1.2). The resulting tree shows two major lineages of *euFUL* genes, with 80% and 100% bootstrap support, respectively, that correspond to the previously identified core eudicot *euFULI* and *euFULII* lineages (Litt and Irish, 2003; Shan et al., 2007). A Solanaceae whole-genome triplication has been proposed (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016), which would suggest that all Solanaceae should have three *euFULI* and three *euFULII* genes. However, others have suggested a duplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012). Our data and other studies, as

well as searches of the tomato genome have shown that tomato has four *euFUL* genes: two *euFULI* and two *euFULII* (Hileman et al., 2006; Bemer et al., 2012; The Tomato Genome Consortium, 2012) instead of the six predicted by a triplication. Additional genome sequencing (e.g., potato, *Capsicum annuum*) (Potato Genome Sequencing Consortium et al., 2011; Hulse-Kemp et al., 2018), transcriptome sequencing, and PCR-based analyses (this study) have also found two *euFULI* and two *euFULII* genes. This suggests the loss of one paralog from each of the *euFULI* and *euFULII* clades following a whole-genome triplication (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016) or, alternatively one or more duplication events (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012).

For the purposes of this paper, we will refer to the *euFULI* and *euFULII* subclades by the name currently used for the tomato gene in each subclade (Hileman et al., 2006; Bemer et al., 2012). Thus, the two *euFULI* subclades will be referred to as the *FUL1* and *FUL2* clades, and the *euFULII* subclades will be referred to as the *MPB10* and *MBP20* subclades (Figure 1.2). In our gene tree, while the *FUL2*, *MBP10*, and *MBP20* clades had high bootstrap support of 83, 99 and 89%, respectively, the *FUL1* clade had only 53% support (Figure 1.2). A single gene from *Streptosolen* grouped sister to the *FUL1* and *FUL2* clades, while a gene from *Schizanthus*, one of the earliest diverging genera (Olmstead et al., 2008; Särkinen et al., 2013), grouped as sister to the *euFULII* clade. To confirm the above were not artifacts, we re-assembled the *Streptosolen* transcriptome while searching for

reads supporting the gene contig, and amplified the *Schizanthus* sequence using gene-specific primers.

The presence of both *FUL1* and *FUL2* genes in species from across the phylogeny is consistent with the event that produced these two clades being part of a family-wide, whole-genome duplication or triplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012; The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016). However, we did not find a *FUL2* ortholog in *Schizanthus*, using transcriptome data, or *Goetzia*, using PCR. These two genera are among the earliest diverging in the family (Olmstead et al., 2008; Särkinen et al., 2013), and are the earliest that we sampled. This raises the possibility that the *FUL1*/*FUL2* clades resulted from a duplication that occurred following the diversification of *Schizanthus* and *Goetzia*. In addition, although we obtained *MBP10* sequences from *Nicotiana* and most of the genera that diversified subsequently (Figures 1.1 and 1.6), we did not find members of the *MBP10* clade in genera that diverged prior to *Brunfelsia*. This suggests that the *MBP10* and *MBP20* subclades were produced by a duplication that occurred later in Solanaceae diversification, after the *euFULI* duplication and any proposed family-wide whole-genome events.

The *euFULII* clades are the result of a tandem gene duplication

To investigate the nature of the *MBP10*/*MBP20* duplication, we mapped the location of the four *euFUL* paralogs to the genome of cultivated tomato. *FUL1* and *FUL2* are located on chromosomes 6 and 3, respectively, consistent with their

origin from a whole genome multiplication. By contrast, *MBP10* and *MBP20* are both located on chromosome 2, about 14.3 million base pairs apart (Figure 1.3). The location of both *euFULII* genes on the same chromosome, and the presence of only one ortholog in early diverging species, support the hypothesis that these paralogs may be the result of a tandem gene duplication. Moreover, comparing a 1-million-base-pair region surrounding both *MBP10* and *MBP20* shows synteny, further supporting a tandem duplication (Figure 1.3). Annotations indicate that these syntenic zones contain 17 homologous regions. The regions that show homology are located on the opposite sides of *MBP10* and *MBP20*, suggesting an inversion of the tandemly duplicated region.

Although we recovered an *MBP10*-clade member in *Brunfelsia* australis using transcriptome analysis, we were unable to amplify this gene from leaf or floral tissue of *Fabiana* or *Plowmania*, genera that are most closely related to *Brunfelsia* (Figures 1.1 and 1.6). In addition, *Petunia* is also a member of the clade that includes *Brunfelsia*, and searches of the published *Petunia* genomes (Bombarely et al., 2016) also failed to turn up an *MBP10*-clade member. However, the *Brunfelsia* sequence in our analysis, obtained from transcriptome data, falls in the expected place in the phylogeny, and we confirmed the presence of *MBP10* transcript in *Brunfelsia floribunda* floral RNA. This suggests that the *MBP10*/*MBP20* duplication occurred before the divergence of the *Brunfelsia*/*Fabiana*/*Petunia*/*Plowmania* clade but the *MBP10* paralog was lost in *Fabiana*, *Petunia* and *Plowmania*.

MBP10 has a short first intron with no TF binding sites

A long first intron ranging from 1 to 10 kb, with multiple potential TF binding sites, is a general feature of *FUL* homologs (Table 1.2) (Takumi et al., 2011). By contrast, *MBP10* has a short first intron of about 80 bp in both cultivated tomato and its closest wild relative, *S. pimpinellifolium*, and about 110 bp in *Nicotiana obtusifolia* (Table 1.2). The expression of most *euFUL* genes is strong across nearly all vegetative and reproductive organs (Ferrándiz et al., 2000; Shchennikova et al., 2004; Kim et al., 2005; Hileman et al., 2006; Bemer et al., 2012; Pabón-Mora et al., 2012, 2013; Scorza et al., 2017); however, diverse analyses using both quantitative and non-quantitative methods indicate that *MBP10* expression is relatively weak in tomato, *S. pimpinellifolium*, and *N. obtusifolia* in most organs (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012; Nakasugi et al., 2014), however, some studies have suggested moderate expression in leaves (Figure 1.4). To determine if the short first intron lacks putative TF binding sites, we searched the first intron of *MBP10* and *MBP20* in tomato (Promo v3.0) (Messeguer et al., 2002; Farré et al., 2003). We found that the first intron of *MBP10* contains no putative TF binding sites, while that of *MBP20* contains 88 putative TF binding sites for eight different TFs. These TFs belong to five main families (Figure 1.7): MYB (MYB2, C1), HSF (HSF1), Dof (Dof1, MNB1a, PBF), WRKY (SPF1) and MADS-box (SQUA). A similar situation was observed for *Nicotiana obtusifolia*, which had 133 putative binding sites in the first intron of *MBP20* for a similar array of TFs, while *MBP10* had only four such sites. In addition, we searched the first intron of *AGL79*, the *euFULII*

paralog of *FUL* in *A. thaliana*, and found 49 putative binding sites, also for similar TFs and TF families. This suggests a loss of regulatory motifs in *MBP10*.

FUL1* and *MBP10* are evolving at a faster rate than *FUL2* and *MBP20

Using the Solanaceae *euFUL* sequence data (Table 1.3), we conducted selection pressure analyses (PAML v1.3) (Yang, 1997) to investigate if there was a shift in evolutionary rate following the *FUL1/FUL2* or *MBP10/MBP20* duplication. Selection pressure (ω) acting upon different *euFUL* gene subclades was calculated as the ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions (dN/dS) (Yang, 1997; Yang and Nielsen, 2000). An ω value of less than 1 means the coding regions are under purifying selection and that protein function is conserved. By contrast, an ω of more than 1 means that the coding regions are under diversifying selection (Yang and Nielsen, 2000). This is interpreted as allowing potential divergence in protein function (Torgerson et al., 2002; Almeida and Desalle, 2009). The nucleotide alignments we used in these analyses excluded the C-termini for all sequences except for those in the FUL2 clade, due to the high variability of this region, which prevents reliable alignment.

Our results indicate that all Solanaceae *euFUL* gene clades are undergoing purifying selection ($\omega \leq 0.20$; Tables 1.1 and 1.5), suggesting conservation of function. The two main lineages, *euFULI* (ω = 0.13) and *euFULII* (ω = 0.16) are evolving at statistically indistinguishable rates. However, within the *euFULI* clade, genes of the *FUL1* clade are evolving at a significantly higher rate (ω = 0.17) compared to those of the *FUL2* clade (ω = 0.11). Within the *euFULII* clade, *MBP10* genes

are also evolving at a significantly higher rate (ω = 0.19) compared to *MBP20* (ω = 0.15). Comparing each clade against all other clades showed that *FUL2* ortholog sequences are the most conserved while *MBP10* ortholog sequences have the weakest purifying selection rates, followed by *FUL1*, implying the possibility of diversifying functions in the latter two subclades (Tables 1.1 and 1.5). None of the gene groups showed a change in evolutionary rates in comparisons between dry- and fleshy-fruited species (Table 1.5).

The rapidly evolving sites are in the regions responsible for protein complex formation

We further analyzed the sequences to identify changes at individual amino acid sites, specifically those that involved a change between polar/charged and non-polar, that might have resulted in a change in protein conformation and function and that were correlated with the change from dry to fleshy fruit. The eu-FUL proteins belong to the Type II MADS-domain containing proteins, which are characterized by a MADS (M) domain, which functions in DNA binding and DNA-protein dimer specificity, an intervening/interacting (I) domain that also has a role in dimer specificity, a keratin-like (K) domain important for protein–protein interactions, and a C-terminal (C) domain, implicated in protein-multimerization, transcription activation, and additional functions (Cho et al., 1999; Heijmans et al., 2012). The C-termini were excluded from this analysis. We selected comparisons in which our results showed two gene groups evolving at significantly different rates (e.g., *FUL1* vs. *FUL2*; Tables 1.1 and 1.5). In the faster evolving group, we

searched for sites in the M, I, and K regions that are undergoing diversifying selection (>1) using mixed effects model of evolution (MEME) (http://datamonkey.org/meme) (Murrell et al., 2012). The results (Figure 1.8) suggest that sites undergoing diversifying selection are located mainly between amino acids 90 and 180 (out of $\sim$210 amino acids in the protein). This region corresponds to the K domain ($\sim$90 to $\sim$180 amino acids) (Kaufmann et al., 2005). In comparison, the M ($\sim$1 to $\sim$60 amino acids) and the I domains ($\sim$60 to $\sim$90 amino acids) had relatively few sites undergoing diversifying selection. Since these TFs function in complexes with other MADS-domain proteins as well as other proteins, novel interactions made possible by amino acid changes in this region might lead to changes in transcriptional activity.

The K domain had 14 sites undergoing diversifying selection in the FUL1 proteins and four of those showed a change in polarity (Figure 1.8). Of those four, a site that corresponds to the 153rd residue in the tomato protein had negatively charged glutamate (E) in most of the non-Solanoideae (mainly dry-fruited) species (11 out of 15 sequences) while all Solanoideae (mainly fleshy-fruited) species had a non-polar residue: valine (V; 13 species) or methionine (M; 1 species) (Figure 1.9). This change was due to a single nucleotide change from an A to T in the former and G to A in the latter. All other changes in FUL1 proteins that result in a change in charge appeared to be reversible, and none were correlated with the phylogeny nor with phenotypic changes. We used the PROVEAN tool on all four K-domain sites that showed a change in charge to predict whether these transitions were likely to

be deleterious or neutral (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015). Two of these sites, one with a histidine (H) to glutamine/asparagine (Q/N) shift at the 95th residue, and one with a lysine (K) to glutamine/threonine (Q/T) shift at the 157th residue (Figure 1.9), were predicted to be functionally deleterious while the other two sites, including the 153rd residue with E to V change, were predicted to be neutral. There were five rapidly changing sites in the M domain and six sites undergoing positive selection in the I domain of FUL1. None of the sites in the M domain showed a change in polarity. Only one site in the I domain showed a change in polarity, but this site was predicted to be neutral functionally. MBP10 proteins had 20 sites undergoing diversifying selection in the K domain, only 1 such site in the M domain and 3 in the I domain (Figures 1.8 and 1.9). Of these, only three sites in the I domain showed a change in charge, all of which were also predicted not to have a negative effect on function.

The Solanaceae *euFULI* and *euFULII* homologs may have experienced distinct mechanisms of cis-regulatory evolution

We compared *euFUL* expression data for the cultivated and wild tomato species, potato and *Nicotiana benthamiana* to identify any patterns that might be the result of changes in the regulatory regions following the duplications of these genes. Not all data from online sources were comparable across species, as different studies included different organs and developmental stages in their analyses, limiting cross-species comparisons. The analysis shows similar spatial expression patterns for *FUL1* and *FUL2* (Figures 1.4 and 1.5). These two paralogs are broadly ex-

pressed in leaves, flowers and fruits of tomato, potato, and tobacco. Although the eFP browser data (Figure 1.4) shows no expression for *FUL1* and *FUL2* in tomato leaves, our RT-PCR data (Figure 1.5) and previous publications (Hileman et al., 2006; Burko et al., 2013) show expression of all four *euFUL* homologs in these organs. Both *euFULI* genes are expressed relatively weakly in the roots of tomato, potato, and tobacco (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012; Nakasugi et al., 2014) (Figures 1.4 and 1.5). Although spatial domains of expression are similar for the *euFULI* genes, they differ in temporal expression over the course of fruit developmental stages in tomato. Although both *FUL1* and *FUL2* are expressed in the fruits of all species, in tomato *FUL2* is highly expressed during the early stages of fruit development and then tapers off, whereas *FUL1* expression increases with time (Figures 1.4 and 1.5).

In comparison to the *euFULI* genes, the two *euFULII* paralogs show more striking differences in spatial expression at the organ level (Figures 1.4 and 1.5), and also between species. In all species for which expression is reported, *MBP10*, alone among the *euFUL* genes in Solanaceae, is not expressed in fruits, or is expressed at barely detectable levels. In tomato, *MBP20* is expressed strongly in roots while *MBP10* is not. By contrast, in potato tubers, *MBP10* expression is high and *MBP20* is not expressed (Figure 1.4). The online sources and our RT-PCR data also show subtle intra-specific differences in expression between *MBP10* and *MBP20* in flowers (Figures 1.4 and 1.5). In addition, our RT-PCR data show that *MBP10*

is expressed relatively weakly in petals and stamens in tomato while *MBP20* is expressed throughout the flower (Figure 1.5). However, these differences seem to be a matter of expression intensity in comparison to the more striking contrasts seen in roots, tubers, and fruits.

The types of differences in expression between *FUL1* and *FUL2* versus *MBP10* and *MBP20* might be due to differences in the regulatory environment as a result of the different ways in which these duplicates arose. A tandem duplication and inversion may have disrupted regulatory regions in ways that would not be associated with a whole genome duplication or triplication (Tanimoto et al., 1999; Kmita et al., 2000; Vogel et al., 2009; Lupiáñez et al., 2015; Puig et al., 2015). To investigate this, we searched for putative TF binding sites in the promoter regions (2 and 5 kb upstream from the transcription start site) of *euFUL* genes in tomato, potato, and woodland tobacco to compare the differences between the pairs of paralogs (Table 1.6). Woodland tobacco was used rather than *N. benthamiana* since relatively longer promoter sequence lengths for *euFUL* genes were available for this genome assembly (Sierro et al., 2013). Despite this, the maximum available promoter length for *NsMBP10* was about 3.3 kb. We found that the differences in types and numbers of predicted TF binding sites between *FUL1* and *FUL2* were comparable to the differences between *MBP10* and *MBP20* (Table 1.6). Nonetheless we did find some differences that may underlie observed differences in expression between paralogs. Some of these differences were presence/absence of binding sites for a particular TF, and some were in the number and distribution of sites. Putative

binding sites for AUXIN RESPONSE FACTORS (ARF) were absent from the tomato *FUL2* promoter while they were present in the promoters of all other *euFUL* genes in all species examined. Only *FUL2* in tomato, *FUL1* in potato, and *MBP10* in woodland tobacco contained binding sites for STOREKEEPER (STK). ETHYLENE INSENSITIVE 3 (EIN3) has three sites in tomato *FUL1* and five in tomato *FUL2*, but the distribution of the sites differs. In FUL1, there are no sites within 2 kb of the coding sequence, and three within 5 kb, whereas in FUL2 there is one site in the 2 kb region and four in the full 5 kb region. In woodland tobacco, there are three EIN3 sites in *FUL1*, all of which are within the 2 kb region, and only one in *FUL2*, which is located between 2 and 5 kb. These types of differences may underlie observed differences in expression.

Discussion

Solanaceae *euFUL* gene tree shows the history of duplications in this lineage

In Solanaceae, there has been a major shift to fleshy fruit in the Solanoideae (Knapp, 2002). However, we do not know the molecular basis of this economically and ecologically important evolutionary event. *FUL* negatively regulates lignification in the dehiscence zone in the dry silique of *A. thaliana*, and functions in cauline leaf development, the transition to flowering and determinacy (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000, 2004; Rajani and Sundaresan, 2001; Melzer et al., 2008). Studies of *FUL* ortholog function across the angiosperms have shown that it is labile, and orthologs have acquired diverse roles over evolutionary time. *VEGETATIVE 1* (*VEG1*), an ortholog of *FUL* in pea

(*Pisum sativum*), is involved in secondary inflorescence meristem identity (Berbel et al., 2012). *AGAMOUS-like 79* (*AGL79*), the *A. thaliana euFULII* paralog of *FUL*, is mainly expressed in the root and has functions in lateral root development and may also play a role function in leaf shape, leaf number, branching, and time to flowering (Gao et al., 2018). However, the overexpression of an *AGL79* ortholog from snapdragon (*Antirrhinum majus*) in *A. thaliana* resulted in indehiscent siliques, suggesting a role more similar to *A. thaliana FUL* (Müller et al., 2001). Evidence suggests that in tomato, one of the *AGL79* orthologs, *MBP20*, plays a role in leaf development (Burko et al., 2013). *VERNALIZATION 1* (*VRN1*) genes, which are *FUL-like* orthologs in grass species such as wheat (*Triticum* spp.) and barley (*Hordeum vulgare*), function in the vernalization response (Preston and Kellogg, 2008). Evidence to date, therefore, suggests that *euFUL* function is labile, and has changed substantially in different plant lineages during the course of angiosperm evolution. Thus it is not surprising to find a change in function of *euFUL* orthologs in Solanaceae.

There is evidence to suggest that Solanaceae *euFUL* orthologs play a role similar to that of *A. thaliana FUL* in the development of dry dehiscent fruits (Smykal et al., 2007). However, studies suggest that in the fleshy fruit of Solanoideae, *FUL* orthologs play roles in pigmentation as well as ethylene response, cell wall modification, glutamic acid degradation, volatile production, and pericarp and cuticle thickness (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). To determine if we could identify changes in *euFUL* sequences or selection that might shed light

on this change in function, we analyzed *euFUL* gene evolution in Solanaceae.

We performed a maximum likelihood phylogenetic analysis (Garli v2.1) (Bazinet et al., 2014) on a data set that consisted of 106 Solanaceae members of the *euFUL* gene lineage (Litt and Irish, 2003; Shan et al., 2007), which we obtained through amplification and sequencing (37 sequences), generating transcriptome sequence data (29 sequences), or mining databases (40 sequences). As outgroup we used 10 *euFUL* genes from Convolvulaceae, the sister family to Solanaceae (Figure 1.2 and Table 1.3) (Stefanović et al., 2003). The resulting tree shows the two major clades of core-eudicot *euFUL* genes, the *euFULI* and *euFULII* lineages (Litt and Irish, 2003; Shan et al., 2007). Within each of these clades there is evidence of a Solanaceae-specific duplication, resulting in two subclades in each lineage. Within each subclade, the order of branches correlates well with the topology of the Solanaceae phylogeny (Olmstead et al., 2008; Särkinen et al., 2013); discrepancies at the genus level are likely due to the short length of some sequences and sequence divergence in some taxa. Each of the subclades includes orthologs from both fleshy- and dry-fruited species, indicating that the subclade duplications preceded the origin of fleshy fruit.

Although duplications in these genes are common (Litt and Irish, 2003; Preston and Kellogg, 2007; Pabón-Mora et al., 2013), we did not find significant evidence of taxon-specific duplications. We did, however, find two genes that did not fall into a specific subclade. A third *Streptosolen* gene grouped sister to the rest of the *euFULI* clade (76% identity among the three *Streptosolen* genes), potentially the

result of a taxon-specific duplication followed by sequence divergence. In addition, a *Schizanthus* gene grouped sister to the *euFULII* clade (77% pairwise identity with *Schizanthus MBP20*). This may also be a divergent genus-specific paralog, but since *Schizanthus* is one of the earliest diverging genera (Olmstead et al., 2008; Särkinen et al., 2013), it is also possible this gene might be a remaining paralog from the reported whole genome duplication/triplication that occurred early in Solanaceae diversification (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012; The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Bombarely et al., 2016). Examination of sequences showed that these are not likely to be splice isoforms. We also found potential evidence of loss – not every Solanaceae species we studied had a copy of each *euFUL* gene. We did not, for example, find *FUL2* genes in *Iochroma*, *Fabiana*, *Solandra*, *Juanulloa*, *Schizanthus*, or *Goetzia*, even though these all had genes in the *FUL1* clade (see Table 1.3 for a complete list). However, although this may represent paralog loss, it is possible we did not recover all gene copies due to PCR primer mismatches, low expression levels, or the absence of transcript in the sampled tissue.

In addition to the major shift to fleshy fruit in the Solanoideae subfamily, fleshy fruits have independently evolved in *Cestrum* and *Duboisia*, and there has also been a reversal to a dry fruit in *Datura* (Knapp, 2002). Our analysis does not include genes from *Duboisia*, but the *euFUL* genes from *Cestrum* and *Datura* grouped in positions in the tree that were expected based on their phylogenetic position, and did not show any notable differences in sequence from the *euFUL* genes of

their close relatives.

The *euFULI* and *euFULII* clade duplicates have experienced different levels of purifying selection

We compared dN/dS ratios between and among Solanaceae *euFULI* and *euFULII* lineages, as well as between sequences before and after the transition to fleshy fruit, to investigate if any changes in selection might be correlated with sequence diversification. All ω values from our analyses are closer to 0 than to 1 (Tables 1.1 and 1.5), which indicates that all *euFUL* gene clades are under strong purifying selection (Yang, 1997; Yang and Nielsen, 2000; Torgerson et al., 2002; Almeida and Desalle, 2009). Studies suggest that this is the norm for most protein coding genes, and that under such stringent evolutionary constraints, slight differences in evolutionary rates may result in functional diversification (Yu et al., 2017). Our data show a weakening of purifying selection in *FUL1* genes relative to *FUL2* genes (ω = 0.17 vs. 0.11, $p < 0.0005$) and in *MBP10* genes relative to *MBP20* genes (ω = 0.19 vs. 0.15, $p < 0.01$). Immediately after the *euFULI* duplication, the *FUL1* and *FUL2* lineage genes would have been fully redundant, which might have allowed the reduction in purifying selection on the *FUL1* genes resulting in potential functional divergence. Similarly, the duplication that resulted in the two *euFULII* gene clades would have resulted in redundancy in the *MBP10* and *MBP20* lineages, possibly allowing the more rapid diversification of *MBP10* genes.

Although studies indicate that the *euFULI* genes of tomato have novel functions compared to those in dry fruit (Gu et al., 1998; Smykal et al., 2007; Bemer et al.,

2012; Shima et al., 2014; Wang et al., 2014), it remains unclear whether the new functions are the result of changes in coding sequences, regulatory regions, or downstream gene targets. Our analysis shows that *euFUL* genes in both dry- and fleshy-fruited species are evolving at similar rates (Table 1.5). This suggests conservation of the coding sequences in both fleshy- and dry-fruited species despite the central roles in the development of these distinct fruit morphologies.

Sixty-four of the sequences in our analysis were from fleshy-fruited species whereas only 42 were from dry-fruited species. Although, we had broad representation across the dry grade, it is possible with additional representation from dry fruited species, more evolutionary patterns would be revealed (Anisimova et al., 2001; Domazet-Loso, 2003; Nielsen et al., 2005).

The FUL1 and MBP10 proteins show amino acid changes in conserved functional domains

An analysis of selection across an entire sequence may indicate different types of selection for the whole gene, but this overlooks the fact that key residues may be undergoing rapid evolution that may result in functional changes (Ota and Nei, 1994; Nei et al., 1997; Yang and Bielawski, 2000; Piontkivska et al., 2002; Martinez-Castilla and Alvarez-Buylla, 2003; Jeffares et al., 2006). Other empirical studies have further described functional changes due to a change in a single amino acid residue (Ingram, 1957; Hanzawa et al., 2005; Hichri et al., 2011; Zhao et al., 2012; Fourquin et al., 2013; Dai et al., 2016; Sakuma et al., 2017) specifically associated with changes in polarity (Schröfelbauer et al., 2004; Hoekstra et

al., 2006) or conformation (Aseev et al., 2012). Studies in A. thaliana, show that a single amino acid mutation in GLABRA1 (GL1) results in the inhibition of trichome formation (Dai et al., 2016) and a change of a single residue is sufficient to convert the function of TERMINAL FLOWER 1 (TFL1), which inhibits flower formation, to that of the closely related FLOWERING LOCUS T (FT), which promotes flowering (Hanzawa et al., 2005). Three-dimensional modeling has also shown that a single amino acid change in a highly conserved domains may lead to changes in protein–protein interactions (Teng et al., 2009; David et al., 2012; Li et al., 2014). We searched for individual sites in the predicted amino acid sequences that showed evidence of positive selection within the gene groups that, although under purifying selection, were found to have statistically significantly accelerated evolutionary rates (i.e., the *FUL1* and *MBP10* clades) to determine if any amino acid changes at these sites had the potential to result in a change in protein function.

Our findings show that more residues are rapidly changing in the K domain compared to the M and I domains (Figure 1.8). The K domain is predicted to have an α-helix structure that facilitates protein–protein interactions (Figure 1.9) (Yang et al., 2003a,b; Kaufmann et al., 2005; Immink et al., 2010). The α-helix structure depends on conserved hydrophobic residues spaced through the domain (Eisenberg et al., 1982). Therefore, changes to protein residues that alter charge and/or conformation in this region can lead to changes in such interactions. Most of the rapidly evolving sites did not show an amino acid change specifically associated with the shift to fleshy fruit, but rather showed changes and reversals over the

course of gene evolution. Interestingly, in the FUL1 proteins, we found one site in the K domain, corresponding to the 153rd residue in the tomato protein (Slugina et al., 2018), at which 11 out of 15 sequences from dry-fruited species have a negatively charged glutamate (E) residue. In comparison, 100% of the fleshy clade contains a non-polar residue: valine (V) (13 species) or methionine (1 species). However, since the remaining four FUL1 sequences from dry-fruited species have non-polar glutamine (Q) or V at this site, the change from charged to non-polar is not associated with the shift to fleshy fruit. In addition, a PROVEAN analysis predicted the changes at this site to be neutral with regards to function.

Two other sites in the FUL1 K domain show changes that are predicted to have functionally deleterious consequences according to our PROVEAN analysis (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015). These include a charged histidine (H) to a non-polar glutamine/asparagine (Q/N) transition at the 95th residue and a charged lysine (K) to non-polar glutamine/threonine (Q/T) transition at the 157th residue (Figure 1.9). Polar residues are important for protein–protein interactions of the K domain α-helix (Sheinerman et al., 2000; Curran and Engelman, 2003; Ma et al., 2003; Zhou et al., 2018) and changes might disrupt interactions with other proteins (Liu et al., 2014). However, since these changes are not correlated with the fruit type, it seems unlikely that any alteration to protein function affects fruit morphology. It is also plausible that any negative effect at these sites is masked by the FUL2 paralog, which is likely to be functionally redundant (Bemer et al., 2012; Wang et al., 2014). This is consistent with FUL1 evolving relatively faster

(Tables 1.1 and 1.5), thus enabling divergence compared to FUL2, which appears to be more highly functionally conserved based on stricter sequence conservation.

None of the sites undergoing positive change in the K domain of MBP10 showed a change in charge, suggesting these changes are not likely to affect protein function. We also observed residues in the M domain that are under diversifying selection in both the FUL1 and MBP10 clades. These residues are located not in the α-helix region that directly binds to DNA, but in the β-sheet region of the MADS domain (Figure 1.8) (Immink et al., 2010). β-sheets are important for protein arrangement in three dimensional space. Therefore, any changes in this region might change protein conformation, influencing DNA binding of the α-helix as well as the ability of the euFUL proteins to form higher order complexes (Pellegrini et al., 1995). However, these shifts were reversible, with no phylogenetic pattern or change in charge, and there was no correlation with the fruit type. Therefore it is unlikely that these shifts have significant functional impact.

A previous report that investigated the evolution of MADS-box genes in *A. thaliana* also found rapidly evolving sites in the M and K domains of Type II MADS-box proteins, which might have been involved in the functional diversification of this group, but did not report changes in the I domain (Martinez-Castilla and Alvarez-Buylla, 2003). Residues in this domain that are directly involved in forming an α-helix structure are expected to be highly conserved, whereas the remaining residues may not be under such constraints (Yang et al., 2003a,b; Kaufmann et al., 2005). We found residues in the conserved region of the I domain that are

undergoing diversifying selection in both FUL1 and MBP10 clades. Of these, one site in FUL1 and three sites in MBP10 had undergone changes in charge but none were predicted to negatively affect the function (Figures 1.8 and 1.9). In addition, as with the sites in the M and K domains, none of these was correlated with the Solanaceae phylogeny or changes in fruit morphology. It has been reported that higher rates of substitution in lineages that show weakened purifying selection or even diversifying selection may be occurring at residues of minimal functional importance (Jacobsen et al., 2016). This might explain the apparent ease of reversibility and lack of phylogenetic signal among the rapidly changing sites we observed.

The *MBP10* and *MBP20* clades are the result of a tandem duplication event

The *FUL1* and *FUL2* genes of tomato are located on different chromosomes (6 and 3, respectively), which is consistent with the proposed Solanaceae whole genome duplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012) or triplication (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016) followed by loss of one paralog. The lack of a *FUL2* ortholog in our dataset from *Goetzia* or *Schizanthus* (Figure 1.2), the two earliest diverging genera that we included in our analyses, raises the possibility that the *FUL1*/*FUL2* clades originated via a duplication that occurred after the diversification of these genera, and not as a result of a whole genome event that preceded the diversification of the family. Whole genome sequences from multiple early diverging lineages will be needed to determine the timing and

nature of these early events.

We did not recover an *MBP10* ortholog from any of the genera that diverged prior to *Brunfelsia* (Figures 1.2 and 1.6). Our investigation revealed that in tomato, both *MBP10* and *MBP20* are located on chromosome 2, about 14.3 million base pairs apart. The 1 million base-pair region surrounding each gene shows synteny, but the order of the homologous regions is reversed (Figure 1.3). Together, this suggests that the *MBP10*/*MBP20* clades are the result of a tandem duplication accompanied by an inversion (Purugganan et al., 1995; Vision et al., 2000; Achaz et al., 2001; Prince and Pickett, 2002). Supporting this, a previous report that investigated genomic duplication events in tomato also found evidence for large-scale intra-chromosomal duplications in chromosome 2 (Song et al., 2012). Although the authors suggest this event was concurrent with a whole genome duplication at the origin of the family, they give a large window, 36–82 million years ago (MYA), for the timing of this event. The stem age of the family is predicted to be approximately 49 MYA (Särkinen et al., 2013), indicating that this duplication might have happened later in Solanaceae diversification. Our data suggest that this duplication event is independent of the reported whole genome events, occurring prior to the diversification of the *Brunfelsia* clade but after the event that produced the *FUL1* and *FUL2* clades (Figures 1.2 and 1.6).

The expected topology for the *euFULII* clade, based on a duplication prior to the divergence of the *Brunfelsia* clade, would be a paraphyletic grade of pre-duplication *euFULII* genes, from species that diversified prior to *Brunfelsia*, and nested *MBP10*

and *MBP20* clades that would include post-duplication genes from all species that diversified subsequent to the duplication. However, in our tree, the pre-duplication genes do not form such a basal grade (Figure 1.2). Rather, they form a clade with the post-duplication *MBP20* genes. The results of our PAML analyses indicate that the *MBP20*-clade genes show less sequence divergence than *MBP10* genes; this higher degree of similarity among pre-duplication sequences and post-duplication *MBP20* genes may underlie their grouping into one clade (Pegueroles et al., 2013).

Our results indicate that the *euFULII* duplication occurred prior to the origin of the clade containing *Brunfelsia*. We would therefore expect to find both an *MBP10* and an *MBP20* in all species of that clade. However, we did not find an *MBP10* ortholog in members of this clade other than *Brunfelsia*. *MBP10* appears to have been lost from the genome of *Petunia*, based on analyses of multiple fully sequenced genomes (Bombarely et al., 2016), and potentially from *Plowmania* and *Fabiana*. We were able to recover *MBP10* orthologs from *Nicotiana* and most other later-diverging genera. However, our analysis includes fewer species from the dry grade of the Solanaceae phylogeny than the fleshy-fruited Solanoideae clade (17 out of 45) and even fewer species that diverged prior to *Brunfelsia* (7). In the *MBP10* clade in particular, our analysis includes 13 orthologs from species in the fleshy-fruited clade but just four from the dry-fruited species, and our analysis only includes sequence data from four genera that diverged prior to the origin of the *Brunfelsia* clade (*Streptosolen*, *Cestrum*, *Goetzia*, *Schizanthus*) (Figures 1.1 and 1.2). Thus there may be genera that originated prior to *Brunfelsia* that contain *MBP10* that

our sampling did not include. Floral and fruit transcriptomes, which provided *MBP10* orthologs from later diverging species, yielded no *MBP10* sequences from *Cestrum* and *Schizanthus*; nonetheless, whole genome sequences of early diverging species are needed to determine the timing of the *MBP10*/*MBP20* duplication.

The *euFULII* expression divergence may be associated with cis-regulatory re-coupling

Our analysis of Solanaceae *euFUL* homologs show that *FUL1* and *FUL2* are broadly expressed in leaves, flowers, and fruit (Figures 1.4 and 1.5). This overall similarity in expression may indicate a conservation of cis-regulatory elements in gene copies following duplication (Haberer et al., 2004). Supporting this, our investigation into the number of putative TF binding sites in the promoter region of *euFULI* homologs did not reveal statistically significant differences (Table 1.6). In tomato fruit development, *FUL1* expression increases with time, whereas *FUL2* expression reaches a maximum at early stages and then decreases over later stages (Figures 1.4 and 1.5). This variation in expression associated with the developmental stages might be due to changes in cis-elements as a result of the accumulation of random mutations over time (Force et al., 1999; Haberer et al., 2004).

Our analysis did find differences in the number and location of predicted binding sites for specific TFs or families, for instance for ARF, STK, and EIN3 TFs, which may account for the types of differences in expression seen between *euFUL* paralogs. The 5 kb region upstream of the *FUL1* transcription start site in tomato contains three putative ARF binding sites but the corresponding region of *FUL2*

in tomato contains no such motifs (Table 1.6). ARF TFs, important in tomato fruit development, are activated in response to auxin and may upregulate or repress downstream genes (de Jong et al., 2010, 2015; Liu et al., 2018); the absence of binding sites from the *FUL2* promoter is the type of factor that might underlie differences in expression observed between *FUL1* and *FUL2*. Predicted STK binding sites are only found in the promoters of potato *FUL1*, tomato *FUL2* and woodland tobacco *MBP10*. STK and STK-like proteins appear to function in storage protein synthesis, glucose reception, and vegetative and reproductive development (Zourelidou et al., 2002; Curaba et al., 2003; Bömer et al., 2011; Chung et al., 2016; Nietzsche et al., 2017). Meanwhile, the 2 kb upstream region of *FUL2* contains a putative site for EIN3. This protein is involved in the development of tomato in response to ripening-associated ethylene production (Tieman et al., 2001). No such motifs are found in the corresponding region of *FUL1*. In contrast, the 2–5 kb region in *FUL2* contains four putative sites for EIN3 while the corresponding region in *FUL1* contains three such sites (Table 1.6). Such variation in number and location of TF binding sites has been shown to be associated with the temporal differences in gene expression (Lebrecht et al., 2005; Liu et al., 2006; Giorgetti et al., 2010; Guertin and Lis, 2010; White et al., 2013; Ezer et al., 2014; Levo et al., 2015; Payne and Wagner, 2015).

Whereas the *euFULI* members largely overlap in spatial expression with some variation associated with developmental stages, the *euFULII* homologs show less consistent spatial expression patterns. Only *MBP20* is expressed in tomato roots and

potato fruit while only *MBP10* is expressed in potato tubers (Figure 1.4). However, these "on" or "off" expression patterns cannot be explained by the presence or absence of any putative TF binding sites (Table 1.6). These two paralogs, which appear to be the result of a tandem duplication and inversion, are located approximately 14.3 Mbp apart (Figure 1.3) on chromosome 2. Although gene clusters resulting from tandem duplications are often coexpressed, this is not the case when there are large physical distances between the genes (Lercher et al., 2003). An investigation into the expression of human transgenes in mice also found changes in expression as a consequence of an inversion, possibly through disrupting enhancer activity or changes to chromatin structure (Tanimoto et al., 1999; Vogel et al., 2009; Puig et al., 2015). Chromosomal rearrangements such as inversions may also result in novel connections between coding regions and other promoters or long distance regulatory motifs while disrupting the original regulatory mechanisms (Kmita et al., 2000; Lupiáñez et al., 2015). This sort of re-coupling of one of the two paralogs might lead to the types of contrasting expression patterns observed for *MBP10* and *MBP20*. However, the expression patterns are not consistent across species (Figures 1.4 and 1.5) and this might be due to additional changes following the inversion (Cosner et al., 1997; Lupski, 1998; Haberle et al., 2008; Chiang et al., 2012). An in-depth analysis of the entire loci and their genomic environment for all paralogs in multiple species would be necessary to determine if the tandem duplication and inversion are associated with changes in proximity to heterochromatin, additional rearrangements, or other phenomena that might have altered gene expression.

MBP10 shows signs of pseudogenization

The first intron of some MADS-box genes contains cis-elements important for the regulation of expression (Gazzani et al., 2003; Michaels et al., 2003; Schauer et al., 2009). Studies have found that deletions in the first intron of a *FUL-like* gene in *Aegilops tauschii* alters expression and results in the loss of the vernalization requirement (Fu et al., 2005; Takumi et al., 2011). Consistent with this, the first introns of angiosperm *euFUL* orthologs are generally in the range of 1–10 kb (Table 1.2) (Takumi et al., 2011). In contrast, tomato *MBP10* has a short first intron of 80 bp. We compared the putative TF binding sites in the first introns of *MBP10* and *MBP20* in tomato to characterize potential loss of such sites, which might suggest reduced gene regulation. The first intron of MBP10 is predicted to have no TF binding sites, while the first intron of MBP20 is predicted to contain 88 TF binding sites (Figure 1.7). These included binding sites for MYB, HSF, Dof, WRKY, and MADS-box TFs. Specific TFs predicted to bind to these sites include MYB2 and C1 (MYB), which play roles in anthocyanin accumulation and lignin biosynthesis, PBF (Dof), which plays a role in endosperm storage protein accumulation, and SPF1 (WRKY), thought to function in fruit ripening (Bovy et al., 2002; Fei et al., 2004; Hwang et al., 2004; Lee et al., 2010; Xu et al., 2014; Jun et al., 2015). A similar pattern was found in analysis of the first intron of *MBP10* in *Nicotiana obtusifolia*, which is 110 bp (Table 1.2). This analysis found three putative TF binding sites for MYB2 and one for PBF. By contrast, the first intron of *N. obtusifolia MBP20* is predicted to have 133 TF binding sites and include a repertoire similar to those found for tomato *MBP20*. To determine whether the difference in TF binding site

number between the paralogs represented a gain of sites in the *MBP20* genes or a loss in the *MBP10* genes, we also searched for TF binding sites in the first intron of *AGL79*, the single *euFULII* ortholog in *A. thaliana* (Gao et al., 2018). We found that it contains 49 predicted TF binding sites for five different TFs in four families: MYB (MYB2, GAMYB), HSF (HSF1), WRKY (SPF1), and GT-box (GT-1). Although this number is substantially smaller than the number of sites predicted in the first introns of the Solanaceae *MBP20* genes, the results suggest that there has been a loss of TF binding sites in *MBP10*.

Core-eudicot *euFUL* and basal-eudicot *FUL-like* genes frequently have broad expression patterns and are generally expressed in fruit (Ferrándiz et al., 2000; Shchennikova et al., 2004; Kim et al., 2005; Hileman et al., 2006; Bemer et al., 2012; Pabón-Mora et al., 2012, 2013; Scorza et al., 2017). Therefore, the absence or extremely weak expression of *MBP10* in fruits of all species, and its weak expression in most organs of tomato and potato is notable (Figures 1.4 and 1.5). This relatively weak expression may at least in part be due to the loss of TF binding sites in the first intron and suggests a potentially reduced role in regulating fruit-related developmental processes. Importantly, the loss of putative TF binding sites and low expression, combined with the faster evolutionary rate, suggest *MBP10* might be in the process of becoming a pseudogene. Further support for this hypothesis comes from an examination of the *MBP10* sequences, which suggests that at least two of the sequences in our study (from *N. sylvestris* and *Dunalia spinosa*) show a frameshift that would result in an premature stop codon.

Conclusion

Our results suggest that there was a weakening in purifying selection following the *euFUL* gene duplications in Solanaceae, resulting in coding sequence diversification in *FUL1* and *MBP10* clades relative to *FUL2* and *MBP20*. Expression of the *euFULI* genes is broad, while the *euFULII* genes have contrasting patterns at the organ level, potentially resulting from cis-regulatory changes associated with the inversion event. We also found evidence to suggest that the *MBP10* clade is becoming a pseudogene. Although at least some clades of Solanaceae *euFUL* genes took on new functions associated with the development of fleshy fruit we did not find any amino acid shifts that were correlated with the change in fruit type. It is also possible that the novel functions are a consequence of downstream changes, perhaps as the result of changes in binding partners or targets. Therefore, the mechanism underlying the shift in *euFUL* function from dry to fleshy fruit in Solanaceae awaits additional analyses.

Achaz, G., Netter, P., and Coissac, E. (2001). Study of intrachromosomal duplications among the eukaryote genomes. Mol. Biol. Evol. 18, 2280–2288.

Albert, V. A., and Chang, T.-H. (2014). Evolution of a hot genome. Proc. Natl. Acad. Sci. U. S. A. 111, 5069.

Almeida, F. C., and Desalle, R. (2009). Orthology, function and evolution of accessory gland proteins in the *Drosophila repleta* group. Genetics 181, 235–245.

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

Anisimova, M., Bielawski, J. P., and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. 18, 1585–1592.

Aseev, L. V., Chugunov, A. O., Efremov, R. G., and Boni, I. V. (2012). A single missense mutation in a coiled-coil domain of *Escherichia coli* ribosomal protein S2 confers a thermosensitive phenotype that can be suppressed by ribosomal protein S1. J. Bacteriol. 195, 95–104.

Bazinet, A. L., Zwickl, D. J., and Cummings, M. P. (2014). A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63, 812–818.

Bemer, M., Karlova, R., Ballester, A. R., Tikunov, Y. M., Bovy, A. G., Wolters-Arts, M., et al. (2012). The tomato *FRUITFULL* homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. Plant Cell 24, 4437–4451.

Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O. S., Sussmilch, F. C., et al. (2012). *VEGETATIVE1* is essential for development of the compound inflorescence in pea. Nat. Commun. 3. doi:10.1038/ncomms1801.

Blanc, G., and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667–1678.

BLAST® Command line applications user manual Bethesda (MD): National Center for Biotechnology Information (US); 2008. Available at: https://www.ncbi.nlm.nih.gov/books/NBK279690.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bolmgren, K., and Eriksson, O. (2010). Seed mass and the evolution of fleshy fruits in angiosperms. Oikos 119, 707–718.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat Plants 2, 16074.

Bömer, M., Uhrig, J., Jach, G., and Müller, K. (2011). Increased vegetative development and sturdiness of *STOREKEEPER*-transgenic tobacco. Open Life Sciences 6. doi:10.2478/s11535-011-0009-9.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., et al. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes *LC* and *C1*. Plant Cell 14, 2509–2526.

Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for APETALA1/fruitfull transcription factors in tomato leaf development. Plant Cell 25, 2070–2083.

Chang, W.-C., Lee, T.-Y., Huang, H.-D., Huang, H.-Y., and Pan, R.-L. (2008). PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics 9, 561.

Chiang, C., Jacobsen, J. C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat. Genet. 44, 390–7, S1.

Choi, Y. (2012). A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein.' Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine-BCB, Orlando, FL. doi: 10.1145/2382936.2382989.

Choi, Y., and Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One 7, e46688.

Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y. H., An, G., et al. (1999). Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. Plant Mol. Biol. 40, 419–429.

Chung, M.S., Lee, S., Min, J.H., Huang, P., Ju, H.W., and Kim, C. S. (2016). Regulation of *Arabidopsis thaliana plasma membrane glucose-responsive regulator* (*AtPGR*) expression by *A. thaliana* storekeeper-like transcription factor, AtSTKL, modulates glucose response in *Arabidopsis*. Plant Physiol. Biochem. 104, 155–164.

Cosner, M. E., Jansen, R. K., Palmer, J. D., and Downie, S. R. (1997). The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr. Gen et. 31, 419–429.

Curaba, J., Herzog, M., and Vachon, G. (2003). *GeBP*, the first member of a new gene family in *Arabidopsis*, encodes a nuclear protein with DNA-binding activity and is regulated by KNAT1. Plant J. 33, 305–317.

Curran, A. R., and Engelman, D. M. (2003). Sequence motifs, polar interactions and conformational changes in helical membrane proteins. Curr. Opin. Struct. Biol. 13, 412–417.

Dai, X., Zhou, L., Zhang, W., Cai, L., Guo, H., Tian, H., et al. (2016). A single amino acid substitution in the R3 domain of GLABRA1 leads to inhibition of trichome formation in *Arabidopsis* without affecting its interaction with GLABRA3. Plant Cell Environ. 39, 897–907.

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5, e11147.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

David, A., Razali, R., Wass, M. N., and Sternberg, M. J. E. (2012). Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum. Mutat. 33, 359–363.

de Jong, M., Wolters-Arts, M., García-Martínez, J. L., Mariani, C., and Vriezen, W. H. (2010). The *Solanum lycopersicum AUXIN RESPONSE FACTOR 7* (*SlARF7*) mediates cross-talk between auxin and gibberellin signalling during tomato fruit set and development. J. Exp. Bot. 62, 617–626.

de Jong, M., Wolters-Arts, M., Schimmel, B. C. J., Stultiens, C. L. M., de Groot, P. F. M., Powers, S. J., et al. (2015). *Solanum lycopersicum AUXIN RESPONSE FACTOR 9* regulates cell division activity during early tomato fruit development. J. Exp. Bot. 66, 3405–3416.

Domazet-Loso, T. (2003). An evolutionary analysis of orphan genes in *Drosophila*. Genome Res. 13, 2213–2219.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 299, 371–374.

Ezer, D., Zabet, N. R., and Adryan, B. (2014). Homotypic clusters of transcription factor binding sites: A model system for understanding the physical mechanics of gene expression. Comput. Struct. Biotechnol. J. 10, 63–69.

Farré, D., Roset, R., Huerta, M., Adsuara, J. E., Roselló, L., Albà, M. M., et al. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res. 31, 3651–3653.

Fei, Z., Tang, X., Alba, R. M., White, J. A., Ronning, C. M., Martin, G. B., et al. (2004). Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J. 40, 47–59.

Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M. F. (2000). Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETAL A1* and *CAULIFLOWER*. Development 127, 725–734.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.

Fourquin, C., del Cerro, C., Victoria, F. C., Vialette-Guiraud, A., de Oliveira, A. C., and Ferrándiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. Plant Physiol. 162, 907–917.

Fu, D., Szücs, P., Yan, L., Helguera, M., Skinner, J. S., von Zitzewitz, J., et al. (2005). Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol. Genet. Genomics 274, 442–443.

Gao, R., Wang, Y., Gruber, M. Y., and Hannoufa, A. (2018). miR156/SPL10 modulates lateral root development, branching and leaf morphology in *Arabidopsis* by silencing *AGAMOUS-LIKE 79*. Front. Plant Sci. 8. doi:10.3389/fpls.2017.02226.

Gazzani, S., Gendall, A. R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiol. 132, 1107–1114.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A developmental perspective. Plant Cell 5, 1439.

Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., et al. (2010). Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. Mol. Cell 37, 418–428.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Guertin, M. J., and Lis, J. T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. 6, e1001114.

Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125, 1509–1517.

Haberer, G., Hindemitt, T., Meyers, B. C., and Mayer, K. F. X. (2004). Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. Plant Physiol. 136, 3009–3022.

Haberle, R. C., Fourcade, H. M., Boore, J. L., and Jansen, R. K. (2008). Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J. Mol. Evol. 66, 350–361.

Hanzawa, Y., Money, T., and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. Proceedings of the National Academy of Sciences 102, 7748–7753.

Heijmans, K., Morel, P., and Vandenbussche, M. (2012). MADS-box genes and floral development: the dark side. J. Exp. Bot. 63, 5397–5404.

Hichri, I., Deluc, L., Barrieu, F., Bogs, J., Mahjoub, A., Regad, F., et al. (2011). A single amino acid change within the R2 domain of the VvMYB5b transcription factor modulates affinity for protein partners and target promoters selectivity. BMC Plant Biol. 11, 117.

Hileman, L. C., Sundstrom, J. F., Litt, A., Chen, M., Shumba, T., and Irish, V. F. (2006). Molecular and phylogenetic analyses of the MADS-box gene family in tomato. Mol. Biol. Evol. 23, 2245–2258.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., and Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313, 101–104.

Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. Horticulture Research 5. doi:10.1038/s41438-017-0011-0.

Hwang, Y.-S., Ciceri, P., Parsons, R. L., Moose, S. P., Schmidt, R. J., and Huang, N. (2004). The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. Plant Cell Physiol. 45, 1509–1518.

Immink, R. G. H., Kaufmann, K., and Angenent, G. C. (2010). The "ABC" of MADS domain protein behaviour and interactions. Semin. Cell Dev. Biol. 21, 87–93.

Ingram, V. M. (1957). Gene mutations in human hæmoglobin: The chemical difference between normal and sickle cell hæmoglobin. Nature 180, 326–328.

Jacobsen, M. W., da Fonseca, R. R., Bernatchez, L., and Hansen, M. M. (2016). Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the *NADH2* gene in whitefish (*Coregonus* ssp.). Mol. Phylogenet. Evol. 95, 161–170.

Jeffares, D. C., Pain, A., Berry, A., Cox, A. V., Stalker, J., Ingle, C. E., et al. (2006). Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. Nat. Genet. 39, 120–125.

Jun, J. H., Liu, C., Xiao, X., and Dixon, R. A. (2015). The transcriptional repressor MYB2 regulates both spatial and temporal patterns of proanthocyandin and anthocyanin pigmentation in *Medicago truncatula*. Plant Cell 27, 2860–2879.

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347, 183–198.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.

Kim, S., Koh, J., Yoo, M.-J., Kong, H., Hu, Y., Ma, H., et al. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J. 43, 724–744.

Kmita, M., Kondo, T., and Duboule, D. (2000). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. Nat. Genet. 26, 451–454.

Knapp, S. (2002). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J. Exp. Bot. 53, 2001–2022.

Krueger, F. (2017). Trim Galore!. Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278.

Lebrecht, D., Foehr, M., Smith, E., Lopes, F. J. P., Vanario-Alonso, C. E., Reinitz, J., et al. (2005). Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. Proc. Natl. Acad. Sci. U. S. A. 102, 13176–13181.

Lee, S., Chung, E.J., Joung, Y.H., and Choi, D. (2010). Non-climacteric fruit ripening in pepper: increased transcription of *EIL-like* genes normally regulated by ethylene. Funct. Integr. Genomics 10, 135–146.

Lercher, M. J., Blumenthal, T., and Hurst, L. D. (2003). Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. Genome Res. 13, 238–243.

Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A. C., Kalma, Y., Lotam-Pompan, M., et al. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. 25, 1018–1029.

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404, 766–770.

Liljegren, S. J., Roeder, A. H. K., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., et al. (2004). Control of fruit patterning in *Arabidopsis* by *INDEHISCENT*. Cell 116, 843–853.

Li, M., Petukh, M., Alexov, E., and Panchenko, A. R. (2014). Predicting the Impact of Missense Mutations on protein–protein binding affinity. J. Chem. Theory Comput. 10, 1770–1780.

Litt, A., and Irish, V. F. (2003). Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics 165, 821–833.

Liu, M., Watson, L. T., and Zhang, L. (2014). Quantitative prediction of the effect of genetic variation using hidden Markov models. BMC Bioinformatics 15, 5.

Liu, S., Zhang, Y., Feng, Q., Qin, L., Pan, C., Lamin-Samu, A. T., et al. (2018). Tomato *AUXIN RESPONSE FACTOR 5* regulates fruit set and development via the mediation of auxin and gibberellin signaling. Sci. Rep. 8, 2971.

Liu, X., Lee, C.K., Granek, J. A., Clarke, N. D., and Lieb, J. D. (2006). Whole-genome comparison of Leu3 binding *in vitro* and *in vivo* reveals the importance of nucleosome occupancy in target site selection. Genome Res. 16, 1517–1528.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025.

Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet. 14, 417–422.

Ma, B., Elkayam, T., Wolfson, H., and Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc. Natl. Acad. Sci. U. S. A. 100, 5772–5777.

Martinez-Castilla, L. P., and Alvarez-Buylla, E. R. (2003). Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. Proc. Natl. Acad. Sci. U. S. A. 100, 13407–13412.

Massa, A. N., Childs, K. L., Lin, H., Bryan, G. J., Giuliano, G., and Buell, C. R. (2011). The transcriptome of the reference potato genome *Solanum tuberosum* group Phureja clone DM1-3 516R44. PLoS One 6, e26801.

Matasci, N., Hung, L.H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 Plants (1KP) project. Gigascience 3, 17.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat. Genet. 40, 1489–1492.

Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albà, M. M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18, 333–334.

Michaels, S. D., He, Y., Scortecci, K. C., and Amasino, R. M. (2003). Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. 100, 10102–10107.

Müller, B. M., Saedler, H., and Zachgo, S. (2001). The MADS-box gene *DEFH28* from *Antirrhinum* is involved in the regulation of floral meristem identity and fruit development. Plant J. 28, 169–179.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.

Nakasugi, K., Crowhurst, R., Bally, J., and Waterhouse, P. (2014). Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. PLoS One 9, e91776.

NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45, D12–D17.

Nei, M., Gu, X., and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. U. S. A. 94, 7799–7806.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3, e170.

Nietzsche, M., Guerra, T., Alseekh, S., Wiermer, M., Sonnewald, S., Fernie, A. R., et al. (2017). STOREKEEPER RELATED1/G-Element Binding Protein (STKR1) Interacts with Protein Kinase SnRK1. Plant Physiol. 176, 1773–1792.

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. Taxon 57, 1159–1181.

Ortiz-Ramírez, C. I., Plata-Arboleda, S., and Pabón-Mora, N. (2018). Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae. Ann. Bot. 121, 1211–1230.

Ota, T., and Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin *VH* gene family. Mol. Biol. Evol. 11, 469–482.

Pabón-Mora, N., Ambrose, B. A., and Litt, A. (2012). Poppy *APETALA1/FRUITFULL* orthologs control flowering time, branching, perianth identity, and fruit development. Plant Physiol. 158, 1685–1704.

Pabón-Mora, N., Sharma, B., Holappa, L. D., Kramer, E. M., and Litt, A. (2013). The *Aquilegia FRUITFULL-like* genes play key roles in leaf morphogenesis and inflorescence development. Plant J. 74, 197–212.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290.

Payne, J. L., and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. Front. Genet. 6, 322.

Pegueroles, C., Laurie, S., and Albà, M. M. (2013). Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. Mol. Biol. Evol. 30, 1830–1842.

Pellegrini, L., Tan, S., and Richmond, T. J. (1995). Structure of serum response factor core bound to DNA. Nature 376, 490–498.

Piontkivska, H., Rooney, A. P., and Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone *H4* gene family. Mol. Biol. Evol. 19, 689–697.

Potato Genome Sequencing Consortium, Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., et al. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475, 189–195.

Preston, J. C., and Kellogg, E. A. (2007). Conservation and divergence of *APETALA1 /FRUITFULL-like* gene function in grasses: evidence from gene expression analyses. Plant J. 52, 69–81.

Preston, J. C., and Kellogg, E. A. (2008). Discrete developmental roles for temperate cereal grass *VERNALIZATION1/FRUITFULL-like* genes in flowering competency and the transition to flowering. Plant Physiol. 146, 265–276.

Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet. 3, 827–837.

Puig, M., Casillas, S., Villatoro, S., and Cáceres, M. (2015). Human inversions and their functional consequences. Brief. Funct. Genomics 14, 369–379.

Purugganan, M. D., Rounsley, S. D., Schmidt, R. J., and Yanofsky, M. F. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 140, 345–356.

Python Language Reference (2010). Python Software Foundation Available at: http://www.python.org.

Rajani, S., and Sundaresan, V. (2001). The *Arabidopsis* myc/bHLH gene *ALCATRAZ* enables cell separation in fruit dehiscence. Curr. Biol. 11, 1914–1922.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: URL https://www.R-project.org.

Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.

Sakuma, S., Lundqvist, U., Kakei, Y., Thirulogachandar, V., Suzuki, T., Hori, K., et al. (2017). Extreme suppression of lateral floret development by a single amino acid change in the VRS1 transcription factor. Plant Physiol. 175, 1720–1731.

Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. 13, 214.

Schauer, S. E., Schlüter, P. M., Baskar, R., Gheyselinck, J., Bolaños, A., Curtis, M. D., et al. (2009). Intronic regulatory elements determine the divergent expression patterns of *AGAMOUS-LIKE6* subfamily members in *Arabidopsis*. Plant J. 59, 987–1000.

Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.

Schlueter, J. A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J. J., et al. (2004). Mining EST databases to resolve evolutionary events in major crop species. Genome 47, 868–876.

Schröfelbauer, B., Chen, D., and Landau, N. R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc. Natl. Acad. Sci. U. S. A. 101, 3927–3932.

Scorza, L. C. T., Hernandes-Lopes, J., Melo-de-Pinna, G. F. A., and Dornelas, M. C. (2017). Expression patterns of *Passiflora edulis APETALA1/FRUITFULL* homologues shed light onto tendril and corona identities. Evodevo 8. doi:10.1186/s13227-017-0066-x.

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and functional diversification during the evolution of the *AP1/SQUA* subfamily of plant MADS-box genes. Mol. Phylogenet. Evol. 44, 26–41.

Shchennikova, A. V., Shulga, O. A., Immink, R., Skryabin, K. G., and Angenent, G. C. (2004). Identification and characterization of four chrysanthemum MADS-box genes, belonging to the *APETALA1/FRUITFULL* and *SEPALLATA3* subfamilies. Plant Physiol. 134, 1632–1641.

Sheinerman, F. B., Norel, R., and Honig, B. (2000). Electrostatic aspects of protein–protein interactions. Curr. Opin. Struct. Biol. 10, 153–159.

Shima, Y., Fujisawa, M., Kitagawa, M., Nakano, T., Kimbara, J., Nakamura, N., et al. (2014). Tomato *FRUITFULL* homologs regulate fruit ripening via ethylene biosynthesis. Biosci. Biotechnol. Biochem. 78, 231–237.

Shima, Y., Kitagawa, M., Fujisawa, M., Nakano, T., Kato, H., Kimbara, J., et al. (2013). Tomato FRUITFULL homologues act in fruit ripening via forming MADS-box transcription factor complexes with RIN. Plant Mol. Biol. 82, 427–438.

Sierro, N., Battey, J. N. D., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., et al. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol. 14, R60.

Slugina, M. A., Shchennikova, A. V., Pishnaya, O. N., and Kochieva, E. Z. (2018). Assessment of the fruit-ripening-related *FUL2* gene diversity in morpho-physiologically contrasted cultivated and wild tomato species. Mol. Breed. 38. doi:10.1007/s11032-018-0842-x.

Smykal, P., Gennen, J., De Bodt, S., Ranganath, V., and Melzer, S. (2007). Flowering of strict photoperiodic *Nicotiana* varieties in non-inductive conditions by transgenic approaches. Plant Mol. Biol. 65, 233–242.

Song, C., Guo, J., Sun, W., and Wang, Y. (2012). Whole genome duplication of intra- and inter-chromosomes in the tomato genome. J. Genet. Genomics 39, 361–368.

Spence, J., Vercher, Y., Gates, P., and Harris, N. (1996). "Pod shatter" in *Arabidopsis thaliana*, *Brassica napus* and *B. juncea*. J. Microsc. 181, 195–203.

Stefanović, S., Austin, D. F., and Olmstead, R. G. (2003). Classification of Convolvulaceae: A phylogenetic approach. Syst. Bot. 28, 791–806.

Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571.

Takumi, S., Koyama, K., Fujiwara, K., and Kobayashi, F. (2011). Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. Genes Genet. Syst. 86, 183–195.

Tanimoto, K., Liu, Q., Bungert, J., and Engel, J. D. (1999). Effects of altered gene order or orientation of the locus control region on human β*-globin* gene expression in mice. Nature 398, 344–348.

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16 Suppl, S181–9.

Teng, S., Madej, T., Panchenko, A., and Alexov, E. (2009). Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. Biophys. J. 96, 2178–2188.

Tieman, D. M., Ciardi, J. A., Taylor, M. G., and Klee, H. J. (2001). Members of the tomato *LeEIL* (*EIN3-like*) gene family are functionally redundant and regulate ethylene responses throughout plant development. Plant J. 26, 47–58.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641.

Torgerson, D. G., Kulathinal, R. J., and Singh, R. S. (2002). Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol. Biol. Evol. 19, 1973–1980.

Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369, 20130353.

Vision, T. J., Brown, D. G., and Tanksley, S. D. (2000). The origins of genomic duplications in *Arabidopsis*. Science 290, 2114–2117.

Vogel, M. J., Pagie, L., Talhout, W., Nieuwland, M., Kerkhoven, R. M., and van Steensel, B. (2009). High-resolution mapping of heterochromatin redistribution in a *Drosophila* position-effect variegation model. Epigenetics Chromatin 2, 1.

Wang, S., Lu, G., Hou, Z., Luo, Z., Wang, T., Li, H., et al. (2014). Members of the tomato *FRUITFULL* MADS-box family regulate style abscission and fruit ripening. J. Exp. Bot. 65, 3005–3014.

White, M. A., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proc. Natl. Acad. Sci. U. S. A. 110, 11952–11957.

Xu, Q., Yin, X.R., Zeng, J.K., Ge, H., Song, M., Xu, C.J., et al. (2014). Activator- and repressor-type MYB transcription factors are involved in chilling injury induced flesh lignification in loquat via their interactions with the phenylpropanoid pathway. J. Exp. Bot. 65, 4349–4359.

Yang, Y., Fanning, L., and Jack, T. (2003a). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. 33, 47–59.

Yang, Y., Xiang, H., and Jack, T. (2003b). *pistillata-5*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. Plant J. 33, 177–188.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17, 32–43.

Yu, J., Wang, L., Guo, H., Liao, B., King, G., and Zhang, X. (2017). Genome evolutionary dynamics followed by diversifying selection explains the complexity of the *Sesamum indicum* genome. BMC Genomics 18, 257.

Zhao, H., Wang, X., Zhu, D., Cui, S., Li, X., Cao, Y., et al. (2012). A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in *Arabidopsis*. J. Biol. Chem. 287, 14109–14121.

Zhou, P., Hou, S., Bai, Z., Li, Z., Wang, H., Chen, Z., et al. (2018). Disrupting the intramolecular interaction between proto-oncogene c-Src SH3 domain and its self-binding peptide PPII with rationally designed peptide ligands. Artif. Cells Nanomed. Biotechnol. 46, 1122–1131.

Zourelidou, M., de Torres-Zabala, M., Smith, C., and Bevan, M. W. (2002). Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of *patatin* expression. Plant J. 30, 489–497.

Figure 1.1. Solanaceae phylogeny with fruit type (dry vs. fleshy) mapped (adapted from Knapp, 2002; Olmstead et al., 2008). The shift to fleshy fruit in the subfamily Solanoideae is indicated with the star. The capsule represents the ancestral fruit-type while the berry represents the generic fruit-type following this shift. The reversal to dry fruit and the independent evolutionary origins of fleshy fruit are highlighted in magenta and blue, respectively. Black circles mark the genera referred to in the text.

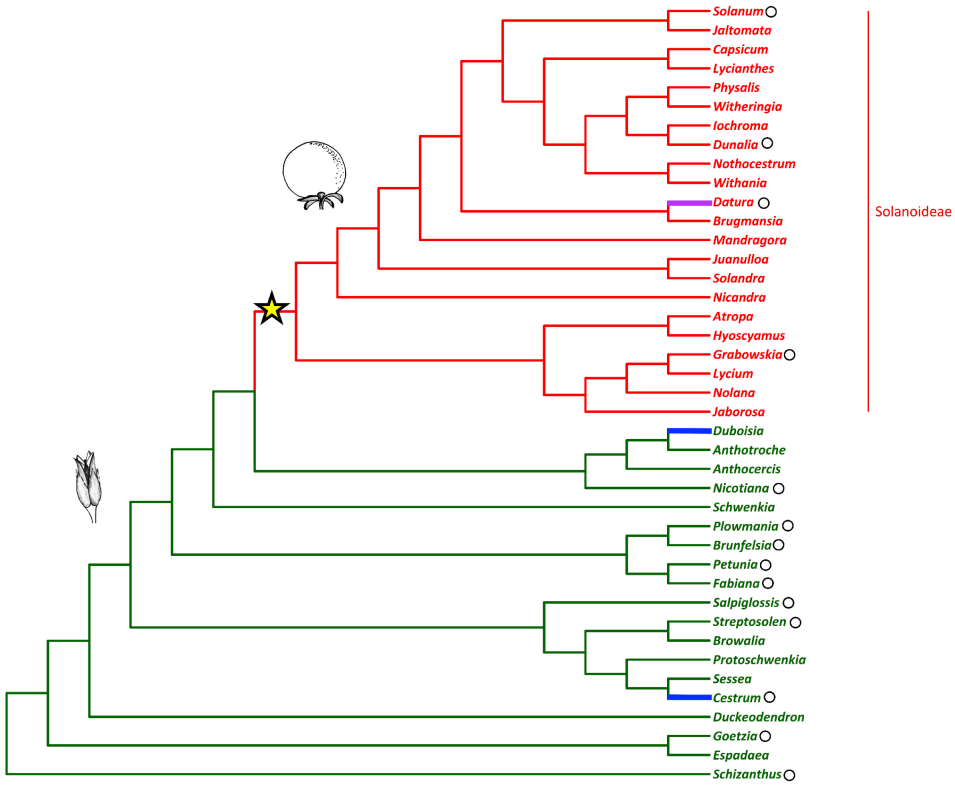


Figure 1.2. Solanaceae *euFUL* Maximum Likelihood gene tree. *FUL1*, *FUL2*, *MBP10*, and *MBP20* clades are colored in blue, green, red, and orange, respectively. A hexagon is placed next to the *Streptosolen* gene that is sister to *FUL1* and *FUL2* clades, and a star is placed next to the *Schizanthus* gene that is sister to the *euFULII* clade. The Convolvulaceae outgroup is highlighted in yellow. The numbers on the branches indicate the bootstrap support.

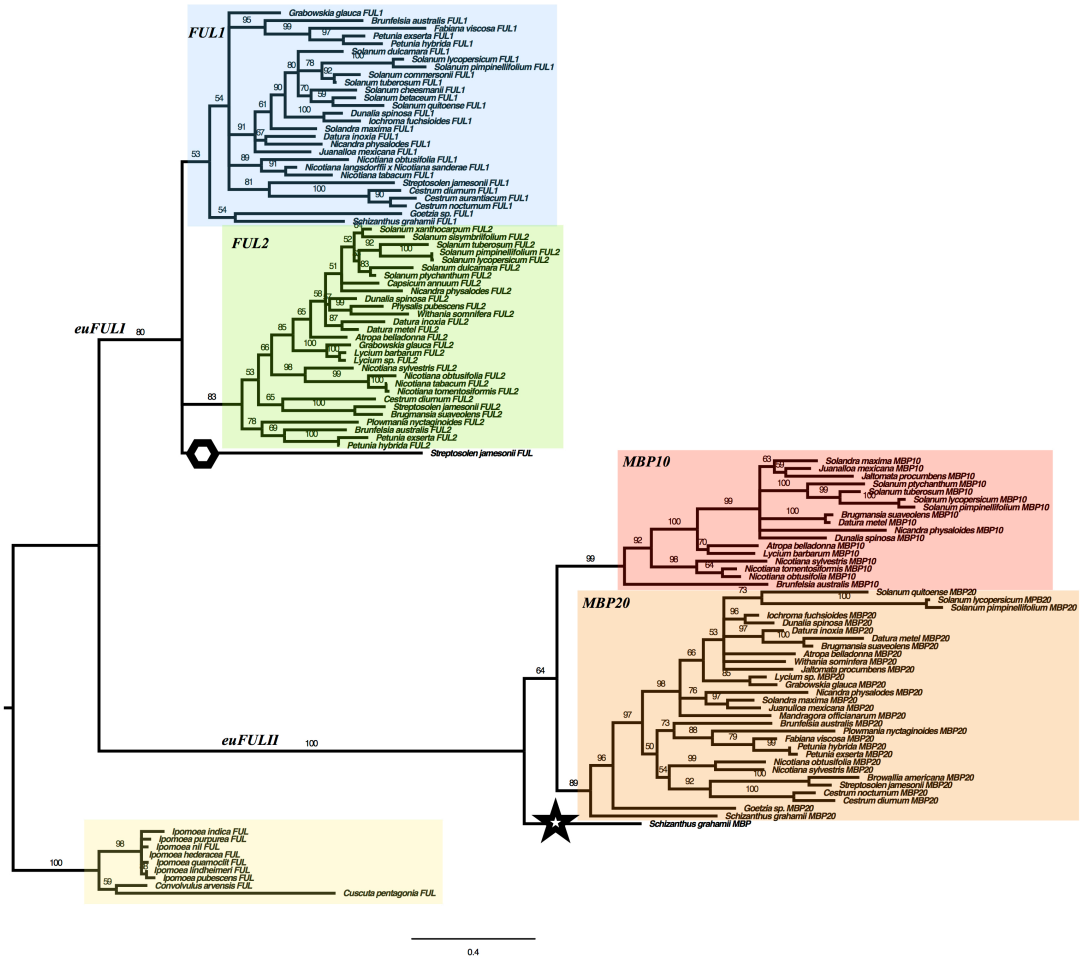


Figure 1.3. Reverse synteny of the regions surrounding *MBP10* and *MBP20* on tomato chromosome 2. The gray block at the top contains the 1 Mbp region surrounding *MBP10* and the white block at the bottom contains the 1 Mbp region surrounding *MBP20*. A colored box in one block is homologous to a box with the identical color in the other block. *MBP10* and *MBP20* genomic sequences are in the center homologous region of the respective block. In *MBP20*, the boxed regions below the red horizontal line are in reverse orientation to the corresponding homologous regions in *MBP10*.

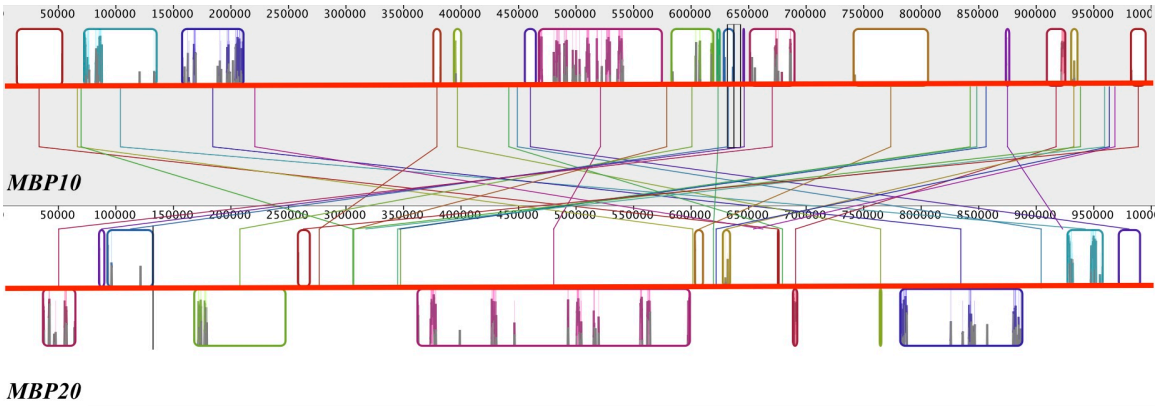


Figure 1.4. The *euFUL* expression profiles in *Solanum lycopersicum*, *S. pimpinellifolium*, *S. tuberosum* from eFP browser
(http://bar.utoronto.ca), and *Nicotiana benthamiana* from the Gene Expression Atlas
(http://benthgenome.qut.edu.au) data.

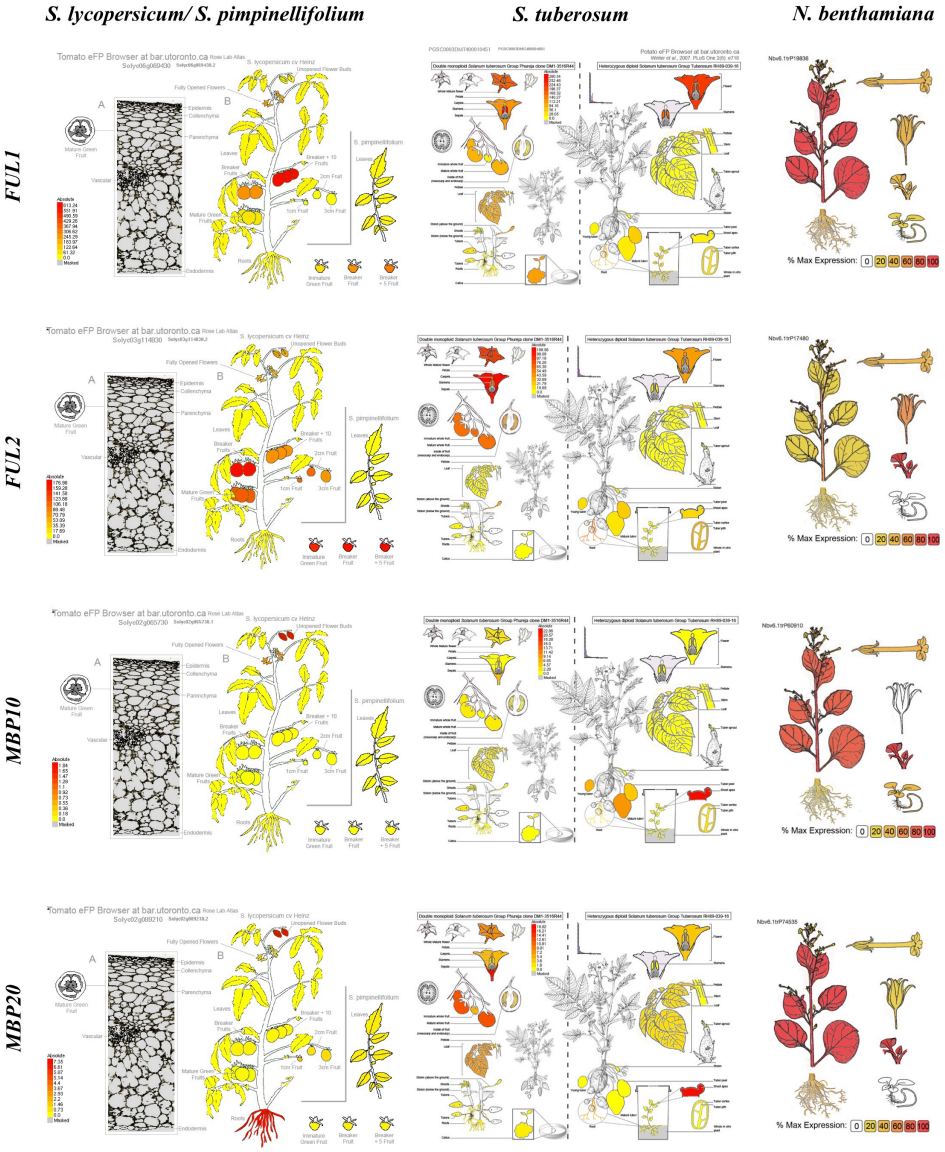
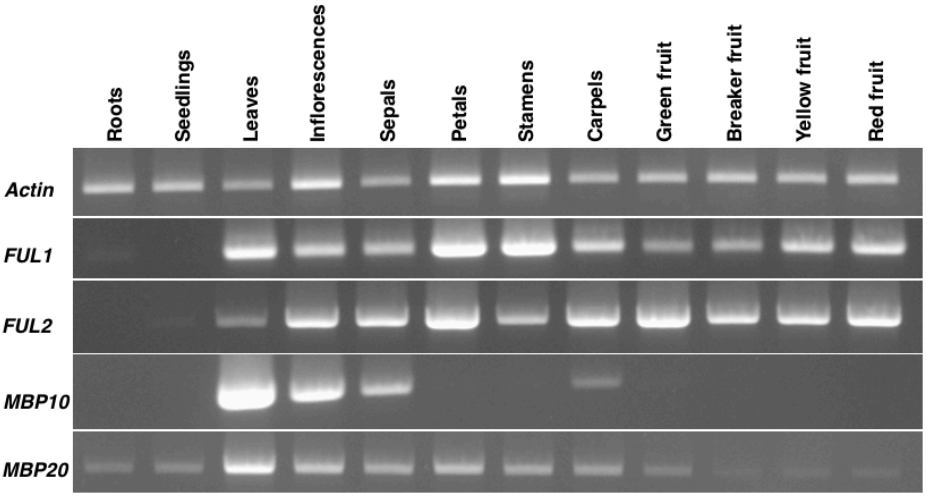


Figure 1.5. *euFUL* expression in *S. lycopersicum* and *S. pimpinellifolium*. (A) A composite of gel images of RT-PCR for *FUL1* (35 cycles), *FUL2* (30 cycles), *MBP10* (30 cycles), *MBP20* (35 cycles) and *ACTIN* (28 cycles) in *S. pimpinellifolium*. The same cDNA was used for all five amplifications of a given tissue. (B) Transcript numbers of *euFUL* genes converted to log counts per million (LogCPM) from RNAseq libraries (unpublished data) of *S. lycopersicum* var. Ailsa Craig (AC) and *S. pimpinellifolium* (PIMP) fruit. We compiled libraries from five different stages of fruit development (Gillaspy, 1993;Tanksley, 2004) in each of the two species. Pre-anth: 1 day pre-anthesis; DPA: days post-anthesis.

(A)



(B)

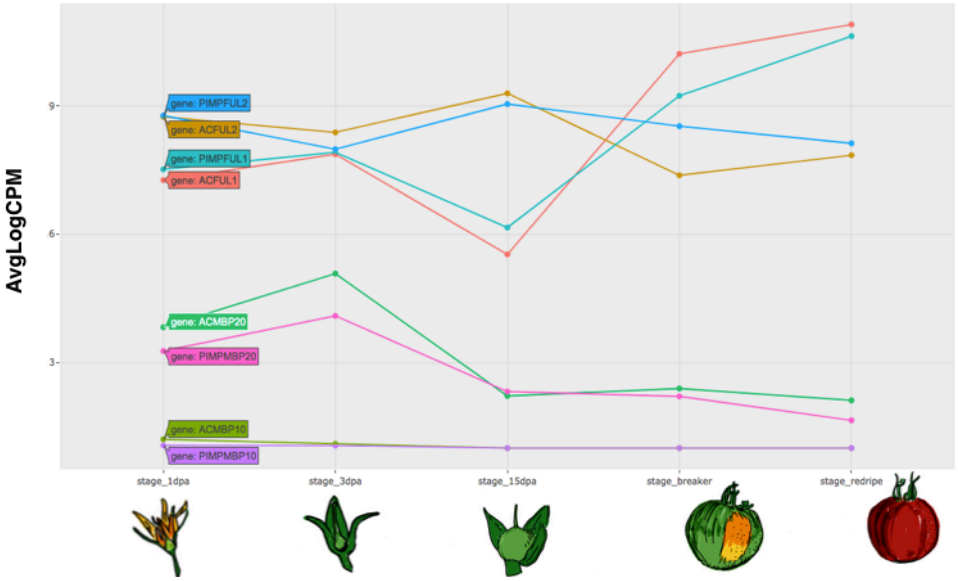


Figure 1.6. The presence/absence of *MBP10*/*MBP20*. The star indicates where in the phylogeny we have evidence for a tandem gene duplication related to the origin of the *MBP10*/*MBP20* clades. The cyan squares and the red dots represent the taxa that we have included in our analysis and those in which we have have found *MBP10*, respectively.

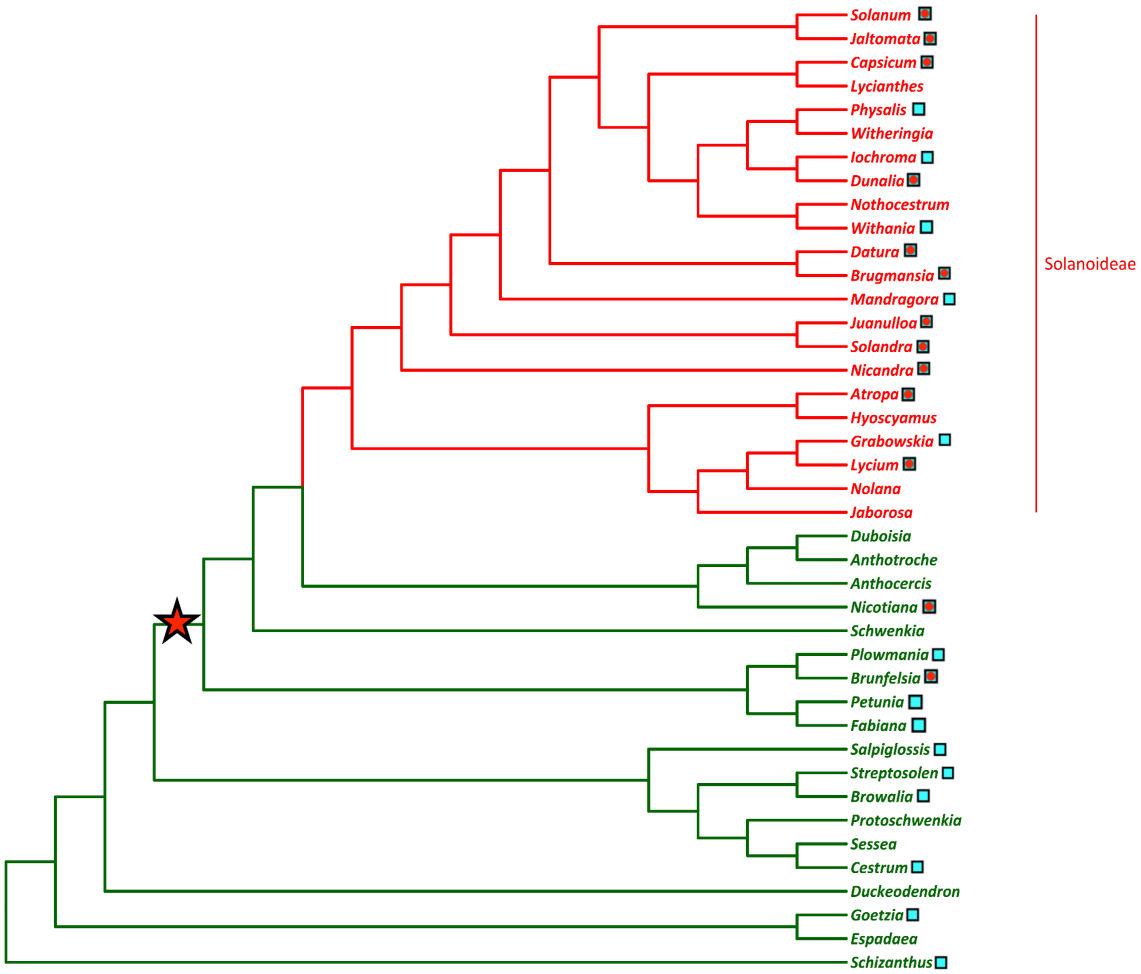


Figure 1.7. Putative transcription factor binding sites for tomato *MBP20* first intron.

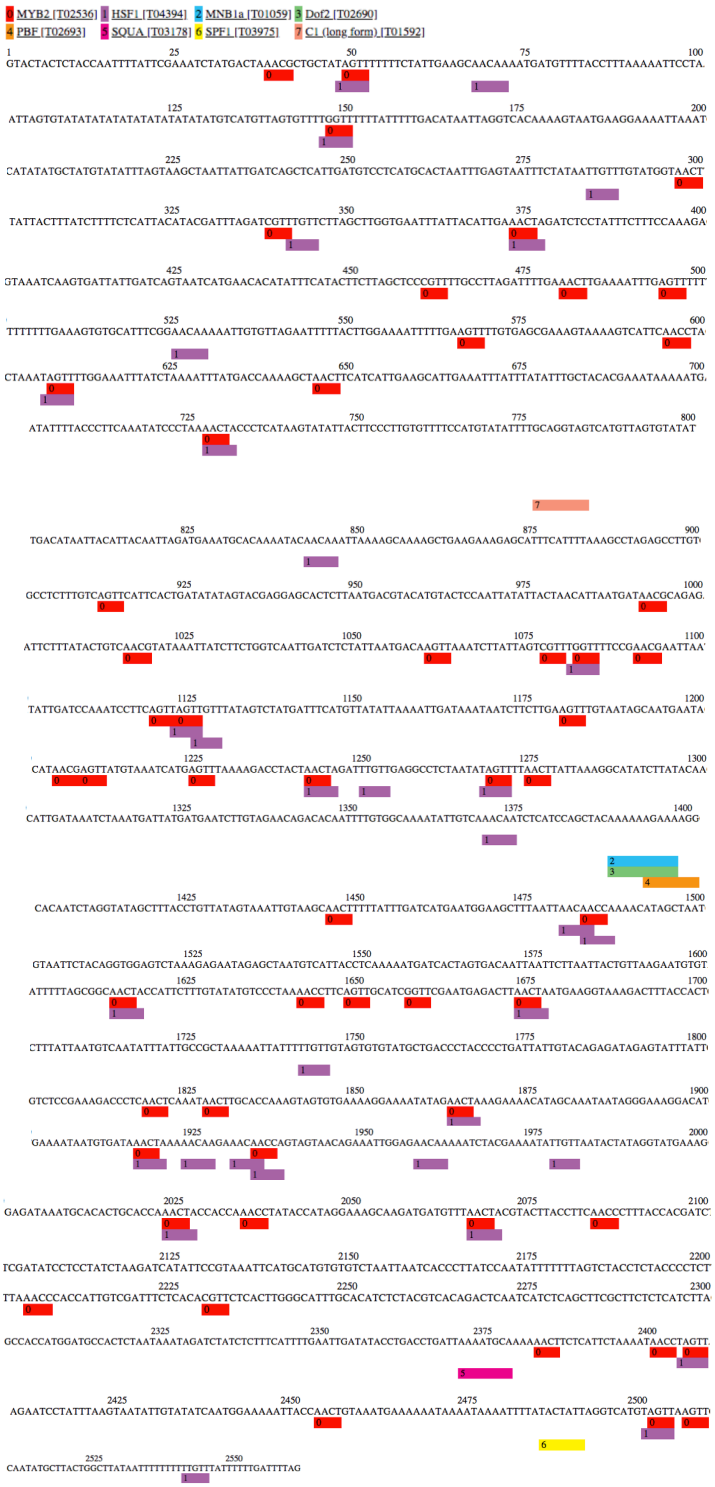


Figure 1.8. Individual sites in euFUL proteins are undergoing rapid evolution. Number of branches under positive selection in (A) FUL1, and (B) MBP10 proteins.

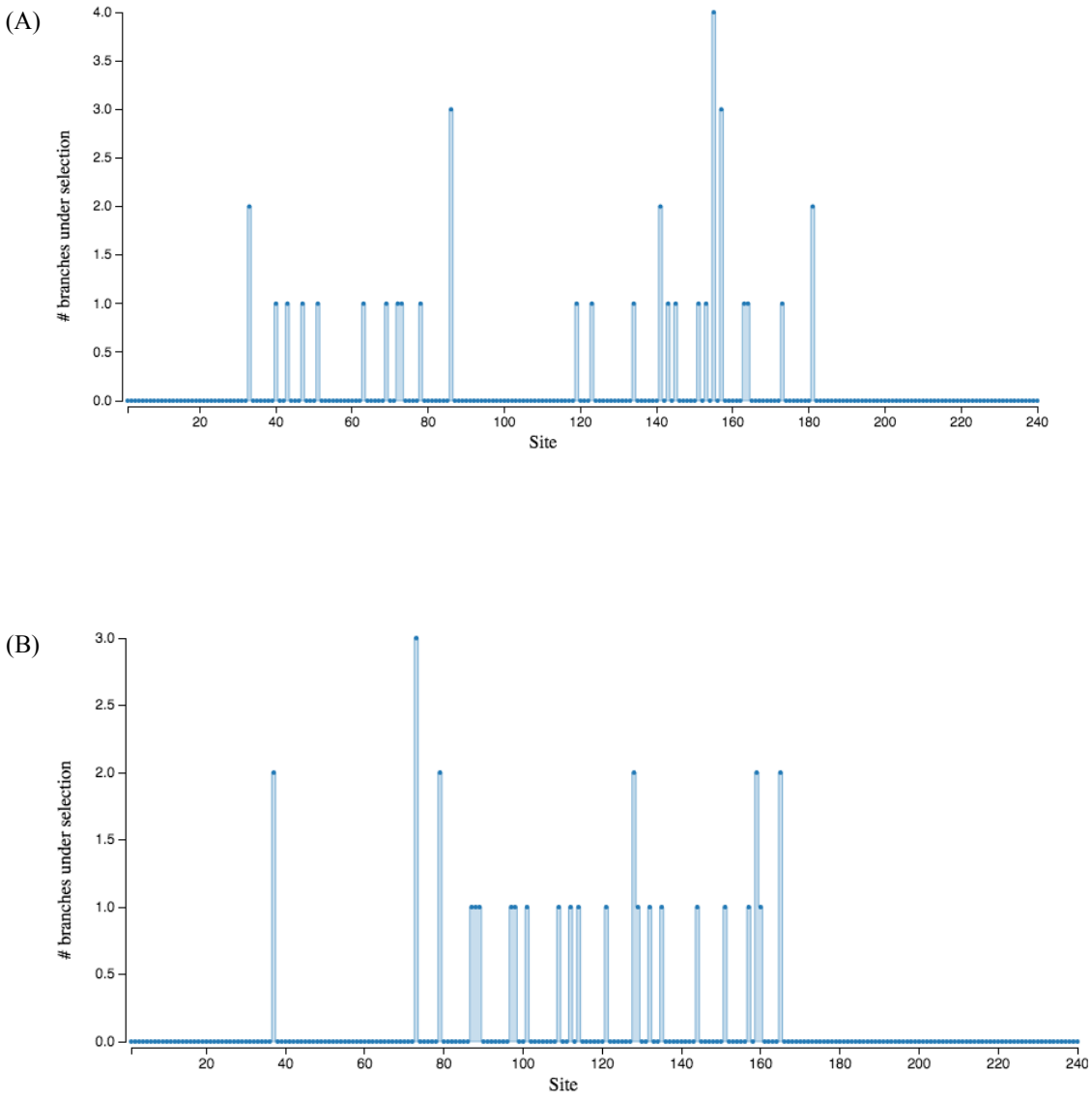


Figure 1.9. Rapidly evolving sites that show a change in charge in FUL1 (A) MADS, (B) I, and (C) K domains and MBP10 (D) MADS, (E) I, and (F) K domains plotted on the predicted structures of the relevant ortholog in *Solanum lycopersicum*. Green helix: α-helix; blue arrow: β-sheet; red ellipse: sites with potentially deleterious changes in function.

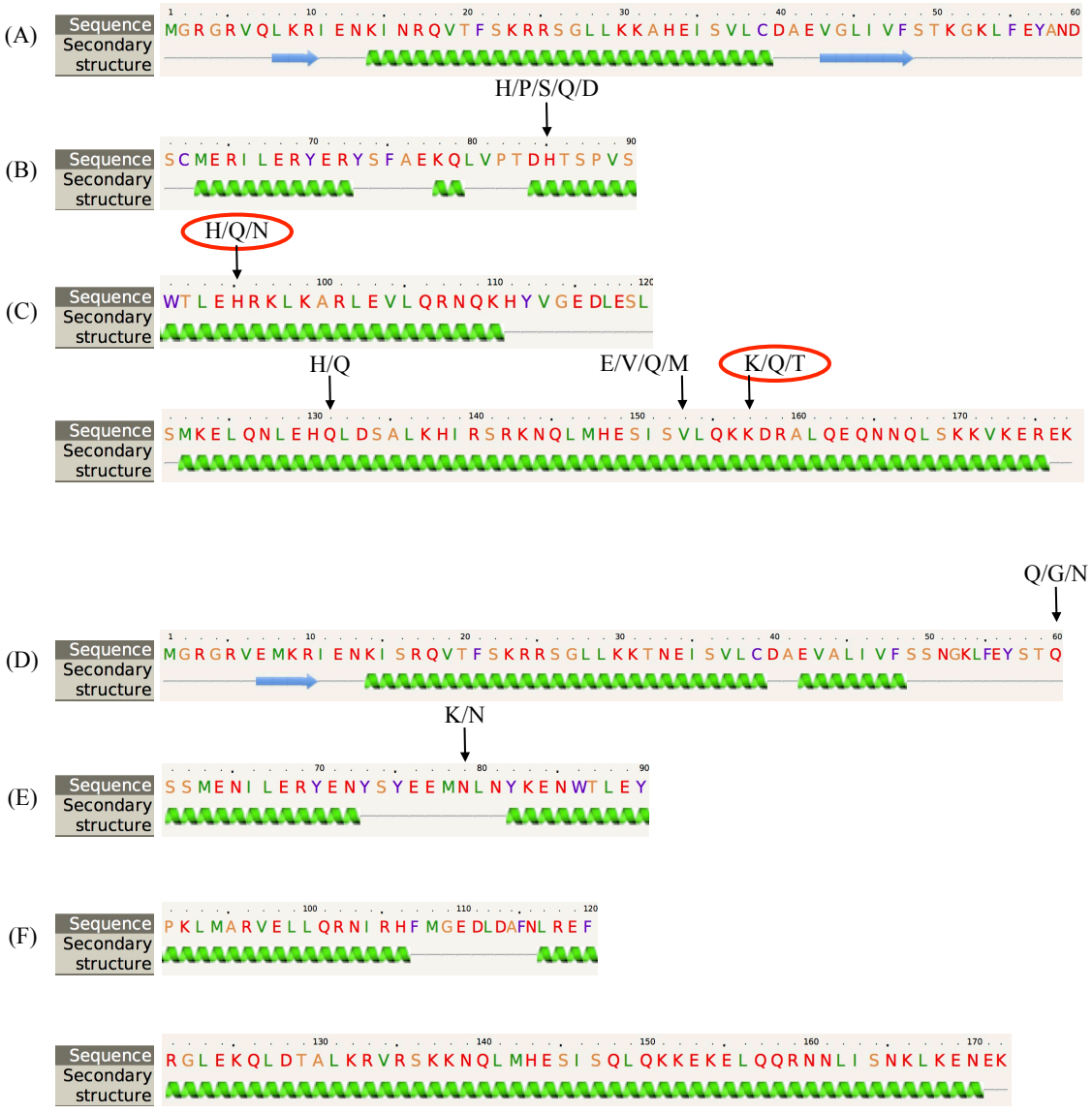


Table 1.1. Evolutionary rates of *euFUL* gene clades that are evolving at statistically different rates.

Comparison	Model	ω_0	ω_1	ω_2	2ΔL	*df*	*P*-value
FUL1 vs. *FUL2*	$M2_A$ (ω_0: background; ω_1: *FUL1* and *FUL2*)	0.1577	0.1311	_	15.5040	1	0.0001
	$M2_B$ (ω_0: background; ω_1: *FUL1*; ω_2: *FUL2*)	0.1577	0.1710	0.1064			
MBP10 vs. *MBP20*	$M2_A$ (ω_0: background; ω_1: *MBP10* and *MBP20*)	0.1311	0.1577	_	7.0291	1	0.0080
	$M2_B$ (ω_0: background; ω_1: *MBP10*; ω_2: *MBP20*)	0.1279	0.1939	0.1514			
FUL1 vs. other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	5.3906	1	0.0001
	M2 (ω_0: FUL_1; ω_1: other euFUL)	0.1706	0.1344	_			
FUL2 vs. other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	19.3663	1	0.0000
	M2 (ω_0: *FUL2*; ω_1: other *euFUL*)	0.1065	0.1622	_			
MBP10 vs. other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	8.7258	1	0.0031
	M2 (ω_0: *MBP10*; ω_1: other *euFUL*)	0.1943	0.1348	_			

2ΔL, test statistic (two times the difference of log-likelihood values); df, degrees of freedom. When the groups being compared did not encompass the entire data set, Model 2 required a two-step analysis (see the section "Materials and Methods"). In this case subscript A and B were used to distinguish the analyses.

Table 1.2. Approximate lengths of the first introns of several *FUL* homologs.

Gene	Length (bp)
Solanum lycopersicum FUL1	5, 000
S. lycopersicum FUL2	4, 400
S. lycopersicum MBP10	80
S. lycopersicum MBP20	2, 500
S. pimpinellifolium MBP10	80
Nicotiana obtusifolia FUL1	5, 300
N. obtusifolia FUL2	3, 800
N. obtusifolia MBP10	110
N. obtusifolia MBP20	3, 000
Arabidopsis thaliana FUL	900
A. thaliana AGL79	1, 700
Aegilops tauschii VRN-D1 (Takumi et al., 2011)	8, 600

Table 1.3. Sources and accession numbers of sequence data. In the third column, "PCR" and "transcriptomes" refer to sequences generated for this project at NYBG (The New York Botanical Garden), UdeA (University of Antioquia, Colombia) or UCR (University of California, Riverside). For sequences obtained from public databases, the database is named in this column.

Species	Gene clade	Source	Source of seed/ tissue
Solanaceae			
Atropa belladonna	*FUL2*	oneKP (BOLZ_scaffold_2025221)	NA
Atropa belladonna	*MBP10*	oneKP (BOLZ_scaffold_2003062)	NA
Atropa belladonna	*MBP20*	oneKP (BOLZ_scaffold_2003065)	NA
Browallia americana	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931101)	NYBG
Brugmansia suaveolens	*FUL2*	Degenerate primer PCR (NYBG; GenBank MH931102)	NYBG
Brugmansia suaveolens	*MBP10*	Degenerate primer PCR (UCR; GenBank MH931103)	NYBG
Brugmansia suaveolens	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931104)	NYBG
Brunfelsia australis	*FUL1*	Transcriptomes (UdeA; GenBank MH931105)	UdeA
Brunfelsia australis	*FUL2*	Transcriptomes (UdeA; GenBank MH931106)	UdeA
Brunfelsia australis	*MBP10*	Transcriptomes (UdeA; GenBank MH931107)	UdeA
Brunfelsia australis	*MBP20*	Transcriptomes (UdeA; GenBank MH931108)	UdeA
Capsicum annuum	*FUL2*	GenBank (NM_001324623.1)	NA
Cestrum aurantiacum	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931109)	NYBG
Cestrum diurnum	*FUL1*	Transcriptomes (UCR; GenBank MH931110)	Chileflora.com
Cestrum diurnum	*FUL2*	Transcriptomes (UCR; GenBank MH931111)	Chileflora.com
Cestrum diurnum	*MBP20*	Transcriptomes (UCR; GenBank MH931112)	Chileflora.com
Cestrum nocturnum	*FUL1*	Transcriptomes (UCR; GenBank MH931113)	Chileflora.com
Cestrum nocturnum	*MBP20*	Transcriptomes (UCR; GenBank MH931114)	Chileflora.com
Datura inoxia	*FUL1*	Degenerate primer PCR (UCR; GenBank MH931117)	NYBG
Datura inoxia	*FUL2*	Degenerate primer PCR (NYBG; GenBank MH931115)	NYBG
Datura inoxia	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931116)	NYBG
Datura metel	*FUL2*	oneKP (NVS_scaffold_2038250)	NA
Datura metel	*MBP10*	oneKP (JNVS_scaffold_2043321)	NA
Datura metel	*MBP20*	oneKP (JNVS_scaffold_2040412)	NA
Dunalia spinosa	*FUL1*	Transcriptomes (UCR; GenBank MH931118)	Chileflora.com
Dunalia spinosa	*FUL2*	Transcriptomes (UCR; GenBank MH931119)	Chileflora.com
Dunalia spinosa	*MBP10*	Transcriptomes (UCR; GenBank MH931120)	Chileflora.com
Dunalia spinosa	*MBP20*	Transcriptomes (UCR; GenBank MH931121)	Chileflora.com
Fabiana viscosa	*FUL1*	Transcriptomes (UCR; GenBank MH931122)	Chileflora.com
Fabiana viscosa	*MBP20*	Transcriptomes (UCR; GenBank MH931123)	Chileflora.com
Goetzia sp.	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931124)	Fairchild Tropical Garden
Goetzia sp.	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931125)	Fairchild Tropical Garden
Grabowskia glauca	*FUL1*	Transcriptomes (UCR; GenBank MH931126)	Chileflora.com
Grabowskia glauca	*FUL2*	Transcriptomes (UCR; GenBank MH931127)	Chileflora.com
Grabowskia glauca	*MBP20*	Transcriptomes (UCR; GenBank MH931128)	Chileflora.com
Iochroma fuchsioides	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931129)	NYBG
Iochroma fuchsioides	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931130)	NYBG
Jaltomata procumbens	*MBP10*	Degenerate primer PCR (NYBG; GenBank MH931131)	Chileflora.com
Jaltomata procumbens	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931132)	Chileflora.com
Juanalloa mexicana	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931133)	Chileflora.com
Juanalloa mexicana	*MBP10*	Degenerate primer PCR (NYBG; GenBank MH931134)	Chileflora.com
Juanulloa mexicana	*MBP20*	Degenerate primer PCR (UCR; GenBank MH931135)	Chileflora.com
Lycium barbarum	*FUL2*	oneKP (LWCK_scaffold_2017804)	NA
Lycium barbarum	*MBP10*	oneKP (LWCK_scaffold_2003151)	NA
Lycium sp.	*FUL2*	oneKP (OSMU_scaffold_2017828)	NA
Lycium sp.	*MBP20*	oneKP (OSMU_scaffold_2017069)	NA
Mandragora officianarum	*MBP20*	Degenerate primer PCR (UCR; GenBank MH931136)	alchemy-works.com
Nicandra physalodes	*FUL1*	Degenerate primer PCR (UCR; GenBank MH931137)	NYBG
Nicandra physalodes	*FUL2*	Degenerate primer PCR (NYBG; GenBank MH931138)	NYBG
Nicandra physalodes	*MBP10*	Degenerate primer PCR (NYBG; GenBank MH931140)	NYBG
Nicandra physaloides	*MBP20*	Degenerate primer PCR (UCR; GenBank MH931139)	NYBG
Nicotiana langsdorffii x Nicotiana sanderae	*FUL1*	GenBank (DQ471787.1)	NA

Continuation of Table 1.3.

Species	Gene clade	Source	Source of seed/ tissue
Nicotiana obtusifolia	*FUL1*	Transcriptomes (NYBG; GenBank MH931141)	US Nicotiana Germplasm Collection
Nicotiana obtusifolia	*FUL2*	Transcriptomes (NYBG; GenBank MH931142)	US Nicotiana Germplasm Collection
Nicotiana obtusifolia	*MBP10*	Transcriptomes (NYBG; GenBank MH931143)	US Nicotiana Germplasm Collection
Nicotiana obtusifolia	*MBP20*	Transcriptomes (NYBG; GenBank MH931100)	US Nicotiana Germplasm Collection
Nicotiana sylvestris	*FUL2*	GenBank (NM_001302579.1)	NA
Nicotiana sylvestris	*MBP10*	GenBank (XM_009763875.1)	NA
Nicotiana sylvestris	*MBP20*	GenBank (XM_009776014.1)	NA
Nicotiana tabacum	*FUL1*	GenBank (DQ534202.1)	NA
Nicotiana tabacum	*FUL2*	GenBank (NM_001325205.1)	NA
Nicotiana tomentosiformis	*FUL2*	GenBank (XM_009627559.2)	NA
Nicotiana tomentosiformis	*MBP10*	GenBank (XM_009618497.2)	NA
Petunia exserta	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931144)	NYBG
Petunia exserta	*FUL2*	Degenerate primer PCR (UCR; GenBank MH931145)	NYBG
Petunia exserta	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931146)	NYBG
Petunia hybrida	*FUL1*	GenBank (AF176782.1)	NA
Petunia hybrida	*FUL2*	GenBank (AF176783.1)	NA
Petunia hybrida	*MBP20*	GenBank (AF335245.1)	NA
Physalis pubescens	*FUL2*	Degenerate primer PCR (UCR; GenBank MH931147)	NYBG
Plowmania nyctaginoides	*FUL2*	Degenerate primer PCR (NYBG; GenBank MH931149)	NYBG
Plowmania nyctaginoides	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931150)	NYBG
Salpiglossis sinuata	*FUL1*	Transcriptomes (UCR; GenBank MH931151)	Chileflora.com
Salpiglossis sinuata	*FUL2*	Transcriptomes (UCR; GenBank MH931152)	Chileflora.com
Salpiglossis sinuata	*MBP20*	Transcriptomes (UCR; GenBank MH931153)	Chileflora.com
Schizanthus grahamii	*FUL1*	Transcriptomes (UCR; GenBank MH931154)	Chileflora.com
Schizanthus grahamii	*MBP*	Transcriptomes (UCR; GenBank MH931155)	Chileflora.com
Schizanthus grahamii	*MBP20*	Transcriptomes (UCR; GenBank MH931156)	Chileflora.com
Solandra maxima	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931157)	NYBG
Solandra maxima	*MBP10*	Degenerate primer PCR (UCR; GenBank MH931158)	NYBG
Solandra maxima	*MBP20*	Degenerate primer PCR (UCR; GenBank MH931159)	NYBG
Solanum betaceum	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931160)	NYBG
Solanum cheesmanii	*FUL1*	oneKP (UGJI_scaffold_2125762)	NA
Solanum commersonii	*FUL1*	GenBank (AF002666.1)	NA
Solanum dulcamara	*FUL1*	oneKP (GHLP_scaffold_2055028)	NA
Solanum dulcamara	*FUL2*	oneKP (GHLP_scaffold_2043858)	NA
Solanum lycopersicum	*FUL1*	GenBank (X60757.1, NC_015443.2)	NA
Solanum lycopersicum	*FUL2*	GenBank (AK327202.1, NC_015440.2)	NA
Solanum lycopersicum	*MBP10*	GenBank (XM_004233345.3, NC_015439.2)	NA
Solanum lycopersicum	*MBP20*	GenBank (XM_010317904.2, NC_015439.2)	NA
Solanum pimpinellifolium	*FUL1*	SolGenomics (Sopim06g069430.0.1, contig:unspecified:1090932:1:2183:1)	NA
Solanum pimpinellifolium	*FUL2*	SolGenomics (Sopim03g114830.0.1, contig:unspecified:5836421:1:4775:1)	NA
Solanum pimpinellifolium	*MBP10*	SolGenomics (Sopim02g065730.0.1, contig:unspecified:6626854:1:18349:1)	NA
Solanum pimpinellifolium	*MBP20*	SolGenomics (Sopim02g089210.0.1, contig:unspecified:1205759:1:362:1)	NA
Solanum ptychanthum	*FUL2*	oneKP (DLJZ_scaffold_2010261)	NA
Solanum ptychanthum	*MBP10*	oneKP (DLJZ_scaffold_2053583)	NA
Solanum quitoense	*FUL1*	Degenerate primer PCR (NYBG; GenBank MH931161)	NYBG
Solanum quitoense	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931162)	NYBG
Solanum sisymbriifolium	*FUL2*	Degenerate primer PCR (UCR; GenBank MH931163)	NYBG
Solanum tuberosum	*FUL1*	GenBank (NM_001288213.1)	NA
Solanum tuberosum	*FUL2*	GenBank (XM_006345039.2)	NA
Solanum tuberosum	*MBP10*	GenBank (XM_006365593.2)	NA
Solanum xanthocarpum	*FUL2*	oneKP (LQJY_scaffold_2015692)	NA
Streptosolen jamesonii	*FUL*	Transcriptomes (UdeA; GenBank MH931164)	Parque Arvi, Vereda Santa Elena, El Tambo, Colombia
Streptosolen jamesonii	*FUL1*	Transcriptomes (UdeA; GenBank MH931165)	Parque Arvi, Vereda Santa Elena, El Tambo, Colombia

Continuation of Table 1.3.

Species	Gene clade	Source	Source of seed/ tissue
Streptosolen jamesonii	*FUL2*	Transcriptomes (UdeA; GenBank MH931166)	Parque Arvi, Vereda Santa Elena, El Tambo, Colombia
Streptosolen jamesonii	*MBP20*	Transcriptomes (UdeA; GenBank MH931167)	Parque Arvi, Vereda Santa Elena, El Tambo, Colombia
Withania somnifera	*FUL2*	Degenerate primer PCR (UCR; GenBank MH931169)	alchemy-works.com
Withania somnifera	*MBP20*	Degenerate primer PCR (NYBG; GenBank MH931168)	alchemy-works.com
Arabidopsis thaliana			
Arabidopsis thaliana	*FUL*	GenBank (NM_125484.4)	NA
Arabidopsis thaliana	*AGL79*	GenBank (NM_113925.3, NC_003074.8)	NA
Convolvulaceae			
Convolvulus arvensis	*FUL*	oneKP (CPOC_scaffold_2010291)	NA
Cuscuta pentagonia	*FUL*	oneKP (AHRN_scaffold_2082598)	NA
Ipomoea coccinea	*FUL*	oneKP (ERWT_scaffold_2042911)	NA
Ipomoea hederacea	*FUL*	oneKP (QSLH_scaffold_2053329)	NA
Ipomoea indica	*FUL*	oneKP (OQBM_scaffold_2015411)	NA
Ipomoea lindheimeri	*FUL*	oneKP (NAUM_scaffold_2053058)	NA
Ipomoea nil	*FUL*	oneKP (NHAG_scaffold_2046547)	NA
Ipomoea pubescens	*FUL*	oneKP (EMBR_scaffold_2056425)	NA
Ipomoea purpurea	*FUL*	oneKP (VXKB_scaffold_2010684)	NA
Ipomoea quamoclit	*FUL*	oneKP (ALUC_scaffold_2003652)	NA

Table 1.4. Sampled tissue and repository for data generated in this study.

Species	Data repository	Sampled tissue
Browallia americana	GenBank	Leaves
Brugmansia suaveolens	GenBank	Leaves
Brunfelsia australis	GenBank	Vegetative and reproductive meristems, floral buds, leaves or fruits
Cestrum aurantiacum	GenBank	Leaves
Cestrum diurnum	GenBank, SolGenomics	Fruits, leaves
Cestrum nocturnum	GenBank, SolGenomics	Inflorescences
Datura inoxia	GenBank	Leaves
Dunalia spinosa	GenBank, SolGenomics	Leaves
Fabiana viscosa	GenBank, SolGenomics	Leaves
Goetzia sp.	GenBank	Leaves
Grabowskia glauca	GenBank, SolGenomics	Leaves
Iochroma fuchsioides	GenBank	Leaves
Jaltomata procumbens	GenBank	Leaves
Juanalloa mexicana	GenBank	Leaves
Mandragora officianarum	GenBank	Leaves
Nicandra physalodes	GenBank	Leaves
Nicotiana obtusifolia	GenBank, SolGenomics	Leaves
Petunia exserta	GenBank	Leaves, flowers
Physalis pubescens	GenBank	Leaves
Plowmania nyctaginoides	GenBank	Leaves
Salpiglossis sinuata	GenBank	Leaves
Schizanthus grahamii	GenBank, SolGenomics	Inflorescences, leaves
Solandra maxima	GenBank	Leaves
Solanum betaceum	GenBank	Leaves
Solanum quitoense	GenBank	Leaves
Solanum sisymbriifolium	GenBank	Leaves
Streptosolen jamesonii	GenBank	Vegetative and reproductive meristems, floral buds, leaves or fruits
Withania somnifera	GenBank	Leaves

Table 1.5. Evolutionary rates of *euFUL* gene clades. Comparisons in which gene clades are evolving at statistically different rates ($p < 0.05$) are highlighted in red.

Comparison	Model	ω0	ω1	ω2	2ΔL	df	P-value
euFULI vs *euFULII*	M0 (ω_0: all branches)	0.1423	_	_	4.4050	2	0.1105
	M2 (ω_0: *euFULI*; ω_1: *euFULII*)	0.1311	0.1577	_			
FUL1 vs *FUL2*	$M2_A$ (ω_0: background; ω_1: *FUL1* and *FUL2*)	0.1577	0.1311	_	15.5040	1	0.0001
	$M2_B$ (ω_0: background; ω_1: *FUL1*; ω_2: *FUL2*)	0.1577	0.1710	0.1064			
MBP10 vs *MBP20*	$M2_A$ (ω_0: background; ω_1: *MBP10* and *MBP20*)	0.1311	0.1577	_	7.0291	1	0.0080
	$M2_B$ (ω_0: background; ω_1: *MBP10*; ω_2: *MBP20*)	0.1279	0.1939	0.1514			
FUL1 vs other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	5.3906	1	0.0001
	M2 (ω_0: *FUL1*; ω_1: other *euFUL*)	0.1706	0.1344	_			
FUL2 vs other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	19.3663	1	0.0000
	M2 (ω_0: *FUL2*; ω_1: other *euFUL*)	0.1065	0.1622	_			
MBP10 vs other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	8.7258	1	0.0031
	M2 (ω_0: *MBP10*; ω_1: other *euFUL*)	0.1943	0.1348	_			
MBP20 vs other *euFUL*	M0 (ω_0: all branches)	0.1423	_	_	0.7890	1	0.3743
	M2 (ω_0: *MBP20*; ω_1: other *euFUL*)	0.1519	0.1390	_			
All dry vs all fleshy	M0 (ω_0: all branches)	0.1423	_	_	3.1106	1	0.0777
	M2 (ω_0: all dry; ω_1: all fleshy)	0.1309	0.1530	_			
*FUL1*dry vs *FUL1*fleshy	$M2_A$ (ω_0: background; ω_1: all *FUL1*)	0.1344	0.1706	_	0.0011	1	0.9735
	$M2_B$ (ω_0: background; ω_1: *FUL1*dry; ω_2: *FUL1*fleshy)	0.1344	0.1712	0.1701			
*FUL2*dry vs *FUL2*fleshy	$M2_A$ (ω_0: background; ω_1: all *FUL2*)	0.1622	0.1065	_	0.1109	1	0.5283
	$M2_B$ (ω_0: background; ω_1: *FUL2*dry; ω_2: *FUL2*fleshy)	0.1622	0.0999	0.1109			
*MBP10*dry vs *MBP10*fleshy	$M2_A$ (ω_0: background; ω_1: all *MBP10*)	0.1348	0.1943	_	0.2540	1	0.6142
	$M2_B$ (ω_0: background; ω_1: *MBP10*dry; ω_2: *MBP10*fleshy)	0.1348	0.1753	0.2009			
*MBP20*dry vs *MBP20*fleshy	$M2_A$ (ω_0: background; ω_1: all *MBP20*)	0.1390	0.1518	_	2.9476	1	0.0860
	$M2_B$ (ω_0: background; ω_1: MBP20dry; ω_2: *MBP20*fleshy)	0.1391	0.1318	0.1766			

Table 1.6. Putative transcription factor (TF) binding sites in the 2/5kb promoter regions of tomato (Sl), potato (St) and woodland tobacco (Ns). Cells highlighted in red had zero predicted transcription factor binding sites.

TF family	# Putative binding sites in the promoter (2/5kb upstream of ATG)																							
	SlFUL 1		SlFUL 2		SlMBP10		SlMBP20		StFUL 1		StFUL 2		StMBP10		StMBP20		NsFUL 1		NSFUL 2		NsMBP10		NsMBP20	
	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	5kb	2kb	3.3kb	2kb	5kb
MYB	41	72	30	48	46	76	31	51	45	78	47	62	38	79	31	56	50	71	40	63	52	60	43	82
NAC	9	13	3	19	5	7	8	9	5	8	5	13	4	8	8	15	6	18	15	17	2	2	16	19
Dof	8	15	11	13	15	16	14	14	9	15	13	16	11	16	9	13	12	15	5	10	13	15	12	14
GATA	2	17	11	17	13	14	7	7	5	17	12	14	2	13	3	9	2	9	12	13	12	13	5	12
MADS	5	12	7	11	3	4	6	11	4	5	5	11	11	11	4	18	6	12	2	10	3	7	1	4
AP2	14	28	6	22	14	20	4	9	12	22	5	21	16	17	21	30	14	27	25	44	5	19	27	38
EIN3		3	1	5	2	5	4	4	2	2	4	5	4	5	3	6	3	3		1	4	4	3	4
WOX	4	5	5	5	5	5	5	5	5	5	5	5	2	2	5	5	4	5	3	4	2	3	5	5
ARF	2	3			1	1	1	2	2	3	2	2	2	3	2	2	2	3	3	4	1	1	1	3
bHLH	4	6	3	9	6	7	7	8	6	9	3	40	39	39	36	36	6	8	8	13	3	5	40	40
WRKY	2	25	10	24	10	16	10	25	2	26	27	29	28	29	28	28	11	19	22	35	9	10	29	31
SBP	21	21	20	21	16	16	15	15	21	21	19	20	3	5	14	14	18	19	20	21	4	15	3	13
HD-ZIP	27	32	24	28	28	31	32	32	31	32	29	2	25	28	29	32	23	30	19	26	20	28	23	27
bZIP	19	20	7	15	25	29	24	27	19	33	10	40	52	53	48	51	9	27	12	41	7	10	30	31
Trihelix	5	5	3	4	2	5	4	6	5	5	2	5	2	6	4	6	5	5	5	5	2	2	4	5
YABBY	2	4	2	5	3	4	2	4	2	4	3	4		4	3	4	4	4	1	3	3	4	1	2
HSF		1	1	1	1	1	1	1	1	1	3	3	1	1		1	1	3	1	1	1	1		2
TCP	7	19	2	5	19	20	2	9	6	20		12	8	8	7	8	2	5	4	4	3	9	6	8
TBP	8	8	8	8	7	8	8	8	8	8	8	8	6	8	8	8	8	8	7	7	7	8	6	8
C2H2	7	10	6	8	7	7	7	9	6	9	4	8	7	12	5	8	3	7	2	7	7	8	7	11
LBD				1						1				1					1	1				2
BES1	1	1			1	1				1		2	2	2	2	2							2	2
TALE	1	3	1	1	3	4	2	2	1	4	1	1	4	4	3	3	1	4	3	4	1	1	5	5
CAMTA				2			3	3					3	3	3	3		3	3	3	3	3		1
SRS	1	2	1	1	2	2	2	2	2	2	2	2		3	1	2	1	1	2	3	2	2	2	3
NF-Y/ CBF	1	1	1	2	1	2	1	2	1	1	1	1	2	2	1	1	1	1	1	2	1	1	1	1
CSD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1
DHN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
STK			1	1						1												2		
FAR1				1			2	2					2	2	2	2						1		1
GRAS	1	1					1	1	1	1	1	1			1	1	1	1	1	1				
TCX/ CPP									4	5	5	5	3	5	5	5	4	5	1	2	5	5	1	3
GRF	1	1								1		1					1	1	1	1	1	1	1	1
Totals	195	330	166	279	237	303	205	270	207	342	218	335	279	371	288	371	200	316	220	347	175	242	276	380

Table 1.7. Primers sequences used for PCR and cloning in this study.

Primer Name	Target	Sequence
AN221 (An221-Actin_121_Fwd)	*ACTIN*	GATGGATCCTCCAATCCAGACACTGTA
AN222 (An222-Actin_122_Rev)	*ACTIN*	GTATTGTGTTGGACTCTGGTGATGGTGT
AN104 (Litt_Fwd_MADS_3)	MADS-Box	GTNCARYTNARRMGNATNGARAAYAAGAT
AN105 (Oligo_dT_1228)	MADS-Box	GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGTTTTTTTTTTTTTTTTTTTTTTTT
AN106F (Litt_Fwd_nested_AP1MDS1)	MADS-Box	GCICWTGARMTNTCNRTNYTNTGYGATGC
AN108 (Litt_Rev_nested_AGL8R)	MADS-Box	AGRTGRYKAASCATCCAIKGIGGCA
UC1_MBP20R_1	*MBP20*	BTHNTTGCTCCAAATGGTCC
UC2_MBP20R_2	*MBP20*	BTHNTTCCTCCAAAAGSCCC
UC3_MBP10R_1	*MBP10*	GKTTGCTKCTTCTCATTTYCTT
UC4_FUL1R_1	*FUL1*	TGTTGAAAAATAAATGAAGGTGA
UC5_FUL2R_1	*FUL2*	GGSGGCATCACAGAAGYGTT
UC6_FUL2R_2	*FUL2*	CATGGCGGCATYACAGTGTT
UC19_FUL2R_3	*FUL2*	KAASYRTYYRKKGNGGCATBACAG
UC20_FUL2R_4	*FUL2*	KAASCATCCAKGGNGGCATBACAG
UC21_FUL1R_2	*FUL1*	CAYCCAKKGKGGCATYRMWVTATTA
UC22_FUL1R_3	*FUL1*	CATCCAKKGKGGCATCACAGTATTA
UC23_MBP10R_2	*MBP10*	WRTTARMACYADKRCGAKTTTGMCC
UC24_MBP10R_3	*MBP10*	WATTAGAACCADGRCGAKTTTGMCC
UC25_MBP20R_3	*MBP20*	CCAAWHKTHARRTYWKRAVBHYRNH
UC26_MBP20R_4	*MBP20*	CCAATTGTTAGGTTAGGAAGTTGGT
UC27_M13F	TOPO vector	GTA AAA CGA CGG CCA G
UC28_M13R	TOPO vector	CAG GAA ACA GCT ATG AC

Table 1.8. Solanaceae *euFUL* gene alignment.

Streptosolen FUL1 ----------
GAGATCTCTGTGCTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGCAAACTCTTT
GAGTTT---
TCCTCTGATTCCTGCATGGAAAAGATCCTTGAAAGATATGAAAGATACTCATTTGCTGAGCGGCA
GCTGGTTAATACTGATCATAGCTCC---
CCGGGAGGCTGGACTCTGGAACATGCAAAGCTTAAGGCCAGAATTGAGGTTCTGCAGAGAAACG
AAAGGCACTATATGGGAGAAGAATTGGATTCGTTGAGTATGAAGGAACTTCAGAATGTGGAGCAC
CAGCTTGATTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAACTGATGCACCAGTCCATTTC
TGAGCTTCAGAAGAAGGACAAAGCATTGCAGGAGCAAAATAAACAGCTTTCGAAGAAGGTGAA
GGAAAGGGAGAAAGAGGTG----------

Schizanthus grahamii FUL1

GATGCTGAGGTTGGTTTGATTGTTTTCTCAACTAAGGGAAAACTCTACGAGTAT---
GCCACCGATTCCTGCATGGAAAGGATTCTCGAAAGGCACGAACGATACTCGTATGCTGAGAGGC
ACCTTGTGGCTACTGATCATAGCTCC---
ACGGGAAGCTGGACTCTGGAACATGCCAAACTTAAGGCCAGAGTTGAGGTTTTGCAGAGAAACC
AAAGGCATTACATGGGAGAAGACTTGGACACGCTAAGTCTGAAAGAGCTTCAGAATCTAGAGCA
CCAGCTGGATTCTGCTCTTAAACACATTCGGTCAAGAAAGAACCAACTGATGCATGAATCCATAT
CTGAGCTGCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACAACCAGCTTTCTAAGAAGGTTA
AGGAGAGGGAGAAAGAGCAG----------

Petunia exserta FUL1 ----------
GCGCTTGAAATTTCNGTGTTGTGTGATGCTGAAGTTGGTTTAACTGTTTTTTCTACTAAAGGCAAA
CTCTTTGAGTAT---
GCTACTGATTCTTGCATGGAGAGGATTCTTGAAAGATATGAAAGATACTCATATGCTGAGAGGCAG
CTTGTTTCTACTGATCATAGCTCC---
CCGGGAAGCTGGAATCTGGAACATGCAAAACTTAAGGCCAGAATTGAGGTTTTGCAGAGAAACC
AAAGGCATTATATGGGAGAAGATTTGGACTCGTTAAGTATGAAAGAACTTCAGAATTTGGAACAA
CAACTAGATTCTTCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GAGCTTCAAAAAAAGGACAAATCATTGCAAGAGCAAAACAACCTTCTTTCAAAGAAGGTGAAG
GAGAGGGAGAAAGAGTTG----------

Petunia hybrida FUL1
ATGGGAAGAGGAAGAGTGCAGATGAAGAGAATTGAGAATAAAATTAATAGACAAGTTACTTTTTC
AAAACGTCGATCTGGATTATTGAAGAAAGCTCATGAAATCTCTGTGCTTTGTGATGCTGAAGTTG
GTTTAATTGTTTTTTCTACTAAAGGCAAACTCTTTGAGTAT---
GCTACTGATTCTTGCATGGAGAGGATTCTTGAAAGATATGAAAGATACTCATATGCTGAGAGGCAG
CTTGTTTCTACTGATCATAGCTCC---
CCGGGAAGCTGGAATCTGGAACATGCAAAACTTAAGGCCAGAATTGAGGTTGTGCAGAGAAACC
AAAGGCATTATATGGGAGAAGATTTGGACTCGTTAAGTATGAAAGACCTTCAGAATTTAGAACAA
CAGCTGGATTCTTCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GAGCTTCAAAAAAAGGACAAATCATTGCAAGAGCAAAACAACCTTCTTTCAAAGAAGGTGAAG
GAGAGGGAGAAAGAGTTG----------

Solandra maxima FUL1 ----------
TGCGCATGAGATGTCAATCTTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCTACTAAAGGCAA
ACTCTTTGAATAT---
GCCACTGATTCTTGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCT
GCTTGTTCCTCCTGATCATAGCTCC---
CCGGGAAGCTGGACTCTGGAACAGGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAAC
CAAAAGCATTATGTGGGAGAAGATTTGGACTCGTTAAATATGAAAGAACTTCAGAAACTGGAGC
AACAGCTTGATTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTT
CTATGCTTCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACCACCAGCTTTCCAAGAAGGTGA
AGGAGAGGGAGAAAGAGCTG----------

Continuation of Table 1.8.

Goetzia sp. FUL1 ----------
GCGCTTGAACTTTCGGTGTTCTGCGATGCTGATGTTGGTTTAACCGTTTTCTCTACTAAAGGCAAA
CTCTACGAGTAT---
GCCTCTGACTCTTGCATGGAAAAGATTGTTGAAAGGTACGAAAGATATTCATATGCTGGGAGAGA
GCTTGTTGCGACTGATAGTAGCTCA---
CCGCGGAACTGGACTCTGGGACATGCCAAGCTTAAGGCAAGACTTGAGGTTTTGCAGAGAAACC
AAAGGCATTATATGGGAGAAGACTTGAACTCTTTAAGCATGAAAGACCTTCAGAACTTAGAGCAC
CAGCTCGATTCTGCTCTTAAACACATTCGATCAAGAGAGAACCAATTGATGCATGAGTGTATATCT
CAACTGCAGAAAAAGGGCAAAGCATTGCAGGAGCAAAACAACCAGCTATCAAAGAAGGCGAAG

AAGGAGAAAGAGCCG----------

Fabiana FUL1
ATGGGAAGAGGAAGAGTGCAGATGAAGAGAATTGAGAACAAAATTAATAGACAAGTTACTTTTT
CAAAACGTCGATCTGGATTATTGAAGAAAGCTCATGAAATCTCTGTTCTTTGTGATGCTGAAGTTG
GTTTAATTGTTTTTTTCTACTAAAGGCAAACTCTGTGAGTAT---
GCTACTGATTCTTGCATGGCGAGGATTCTTGAAAGATATGAAAGATACTCATATGCTGAGAGGCAG
CTTGATTCTACTGATCATAGCTCC---
CCGGGAAGCTGGAATCTGGAACATGCAAAACTTAAGGCAAGAATTGAGTTTTTGCAGAGAAACC
AAAGGCATTATATGGGAGAAGACTTGGACTCGTTAAGTATGAAAGAACTTCAGAATTTGGAACAA
CAGTTAGATTCTGCTCTTAAACGCATTCGTGCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GAGCTTCAAAAAAAGGACAAAGCATTGCGAGAGCAAAACAACCTTCTTGCAAAGAAGGTGAAG
GAGAGGGAGAAAGAGTTG----------

Brunfelsia FUL1

ATGGAGAGGATTCTTGAAAGATATGAAAGATACTCATATGCTGAGAGGCAGCTTGTTCCTACTGA
AGATAGCTCC---
CAGGGAGACTGGAATCTGGAACATGCAAAACTTAAGGCCAGAATTGAGATTTTGCAGAGAAACC
AAAGGCATTATATGGGAGAAGACTTGGACTCATTAAGTATGAAGGAACTCCAGAATTTGGAGCAC
CAGCAAGATTCTGCTCTTAAACACATTCGCTCAAGAAAGAACCAATTGATGAATGAGTCCATTTC
TGAGCTTAAAAAAAAGGACAAAGAATTGCAGGAGCAAAACAACCAGCTTTTGAAGAAGGTGAA
GGAGAGGGAGAAAGAGCTG----------

Solanum lycopersicum FUL1
ATGGGAAGAGGAAGAGTCCAGTTGAAGCGAATAGAGAACAAAATTAACCGTCAAGTTACCTTCT
CGAAACGTCGATCTGGTTTGCTGAAGAAAGCCCATGAGATCTCTGTGCTTTGTGATGCTGAGGTT
GGTTTGATTGTTTTTTCTACTAAAGGAAAACTCTTTGAATAT---
GCCAACGATTCCTGCATGGAGAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAAACA
GCTTGTTCCTACTGATCATACCTCC---
CCGGTAAGCTGGACCCTTGAACATCGAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAAGATTTGGAGTCTTTAAGTATGAAGGAACTTCAGAATCTGGAGCAC
CAGCTTGATTCAGCTCTTAAACACATTCGATCAAGAAAGAATCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAGAGCATTGCAGGAGCAAAACAACCAGCTTTCGAAGAAGGTGAAG
GAGAGGGAGAAG----------

Cestrum diurnum FUL1 ----------
TGCGCATGAGATGTCAGTTTTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCTACGAAAGGCAA
ACTCTTTGAGTAT---
GCCACTGATTCTTGCATGGAAAAGACCCTTGAAAGATATGAAAGATACTCATATGTTGAGCGCCA
ACTTGTTGCTACTGATCCTGCCTCT---
CTGGGAAGCTGTACTTTGGAGCATGCTAAACTTAAGGCCAGACTTGAGGTTCTCCAGAGAAACC
AAAAGCATTATATGGGAGAAGATTTGAATTCTTTAAGTATGAAAGAACTTCAGAATGTTGAGCAC
CAGCTTGATTCTTCTCTTAAACACATTCGATCAAGGAAGAACCAATTGATGCATGAGTCTATTTCT
GAGCTTCAAAAGAAGGACAAGGCATTGCAGGAGCAAAATAACCAGCTTTTGAAGAAGATGAGG
GAAAGGGAGAAAGAGCTA----------

Continuation of Table 1.8.

Cestrum aurantiacum FUL1 ----------
GCGCTTGAAATGTCGATGCTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCTACGAAAGGCAAA
CTCTTTGAGTAT---
GCACCTGATTCTTGCATGGAAAAGATCCTTGAAAGATATGAAAGATACTCATATGCTGAGCGCCA
ACTTGTTGCTACTGATCCTGCCTCT---
CCGGGAAGATGGACTTTGGAGCATGCGAAACTTAAGGCCAAACTTGAGGTTCTCCAGAAAAACC
AAAAGCATCATATGGGAGAAGATTTGGATTCTTTAAGTATAAAAGAACTTCAGAATGTTGAGCAC
CAGCTTGATTCTGCTCTTAAACACGTTCGATCAAGGAAGAATCAATTGATGCATGAGTCTATTTCT
GAGCTTCAAGAGAAGGACAAGGCATTGCAGGAGAAAAATAACCAGCTTTCGAAGAAGATGAAG
GAAAGGGAGAAAGAGCTA----------

Cestrum nocturnum FUL1
ATGGGAAGAGGAAGGGTGCAGTTGAAAAGAATAGAGAACAAAATAAACCGGCAAGTGACTTTC
TCTAAAAGACGATCTGGTTTGCTCAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGT
TGGTTTGATTGTTTTTTCTACGAAAGGCAAACTCTTTGAGTAT---
GCCACTGATTCTTGCATGGAAAAGATCCTTGAAAGATATGAAAGATACTCATTTGCTGAGCGCCA
ACTTGTTGCTACTGATCCTGCCTCT---
CCGGGAAGATGGACTTTGGAGCATGCGAAACTTAAGGCCAGACTTGAGGTTCTCCAGAAAAACC
AAAAGCATTATATGGGAGAAGATTTGGATTCTTTAAGTATGAAAGAACTTCAGAATGTTGAGCAC
CAGCTTGATTCTGCTCTTAAACACGTTCGATCAAGGAAGAATCAATTGATGCATGAGTCTATTTCT
GAGCTTCAAAAGAAGGACAAGGCATTGCAGGAGAAAAATAACCAGCTTTCGAAGAAGATGAAG
GAAAGGGAGAAAGAGCTA----------

Nicotiana obtusifolia FUL1

----------TTCTACAAAAGGCAAACTCTTTGAATAT---
GCCACTGATTCTTGCATGGAGAGGATCCTTGAAAGATACGAAAGATACTCATATGCTGAGAGGAA
GCTTGTTACTACTGATCATAGCTCC---
CCGGGAAGCTGGAACCTGGAACATGCAAAACTTAAGGCTAGAGTTGAGGTTTTACAGAGAAACC
AAAGGCATTATATGGGAGAAGATTTGGACTCGTTAAGTACGAAAGAACTTCAGAATTTGGAGCAG
CAGCTGGATTCTGCTCTTAAACTCATTCGCTCAAGAAAGAATCAATTGATGCATGAGTCTATTTCT
GAGCTTCAAAAAAAGGACAAAGCACTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAA
GGAGAGGGAGAAAGAGCTG----------

Grabowskia FUL1
ATGGGGAGAGGAAGAGTGCAGCTGAAGAGAATAGAGAACAAAATTAATCGACAAGTGACTTTCT
CTAAACGTCGATCTGGTTTGTTGAAGAAAGCCAATGAGATCTCTGTGCTTTGTGATGCTGAGGTT
GGTTTGATTGTTTTTTCTACTAAAGGCAAACTCTTTGAATAT---
GCTACTGATTCTTGCATGGAAAGGGTGCTTGAAAGATATGAAAGATACTCATACGCTGAGAGGCA
GCTTGTTCCTACTGATCCTACCTCC---
CCGGGAAGCTGGACTCTGGAACATGCAAAACTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATATGGGAGAAGACTTGGACTTATTAAGTATGAAAGAACTTCAGAATGTGGAGCAC
CAGCTTGATTCTGCTCTTAAACACATTCGCTCAAGAAAGAGCCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAAG
GCAAAGGAGAA----------

Nicotiana langsdorffii sanderae FUL1

TCAAGAAAGAACCAATTGATGCATGAGTCCATTTCTGAGCTTCAAAAAAAGGACAAAGCACTGC
AGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAAGGAGAGGGAGAAAGAGCTG----------
----

Nicotiana tabacum FUL1
ATGGGAAGAGGAAGGGTGCAGTTGAAGAGAATTGAGAACAAAATTAATAGGCAAGTTACTTTCT
CAAAACGTCGATCTGGTTTGCTTAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGTT

Continuation of Table 1.8.

GGTTTGATTGTTTTTTCTACAAAAGGCAAACTCTTTGAATAT---
GCCACTGATTCTTGCATGGAGAGGATCCTTGAAAGATACGAAAGATACTCATATGCTGAGAGGCA
ACTTGTTACTACTGATCATAGCTGC---
CCGGGAAGCTGGACCCTGGAACATGCAAAACTTAAGGCTAGACTTGAGGTTTTGCAGAGAAACC
AAAGGCATTATACGGGAGAAGATTTGGACTCGTTAAGTACGAAGGAACTTCAGAATTTGGAACA
CCAGCTGGATTCTGCTCTTAAACACATTCGCTCAAGCAAGAACCAATTGATGCATGAGTCTATTTC
TGAGCTTCAAAAAAAGGACAAAGCACTGCAGGAGCAAAACAACCAGCTTTGCAAGAAGGTGAA
GGAGAGGGAGAAAGAGTTG----------

Juanalloa mexicana FUL1 ----------
TGCGCTTGAGATATCGGTTCTGTGCGATGCTGAGGTTGGTTTGATTGTTTTTTCTAGTAAAGGCAA
ACTCTTTGAATAT---
GCCACTGAATCATGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGACA
GCTTGTTCCTACTGATCATAGCTCC---
CCGGGAAGCTGGACTCTGGAACAGGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAAC
CGAAAGCATTATGTGGGAGAAGATTTGGACTCGTTAACTATGAAAGAACTTCAGAATCTGGAGCA
CCAGCTTGATTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTC
TGTGCTTCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACAACCTGCTTTCCAAGAAGGTGAA
GGAGAGGGAGAAAGAGCTG----------

Datura inoxia FUL1 ----------
TGAGCTGTCGGTGCTATGCGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGCAAACTCTT
TGAATAC---
GCTACAGATTCTTGCATGGAAAGGATACTGGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GGTTGCTCCTACTGATCATACCTCC---
CCGAGAAGCTGGATTCTGGAACAGGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGGGGAAC
CAAAAGCATTATGTTGGAGAAGATTTGGAGTCATTAAATATGAAAGAACTTCAGAATCTGGAACA
CCAGCTTGATTCTGCTCTCAAACACATAAGATCAAGAAAGAACCAATTGATGCATGAGTCCATTT
CTGTGCTTCAAAAAAAGGACAAAGCACTGCAGGACCAAAACAACCAGCTTTCCAAGAAGGTGA
AGGAGAGGGAGAAAGAGTTG----------

Nicandra physalodes FUL1 ----------
ATGAAATTTCGGTGCTGTGTGATGCCGAGGTTGGTTTGATTGTTTTCTCAACTAAAGGGAAACTCT
TTGAATAT---
GCTACCGATTCTTGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GCTTGCTCCTACTGATCATAGCACC---
CCGGGAAGTTGGACTCTGGAACACGCAAAACTTAAGGCCAGACTTGAGGTTCTCCAGAGGAAC
CAAAAGCATTATGTGGGAGAAGATTTGGACTCGTTAAATATGAAAGAACTTCAGAATCTGGAACA
TCAGCTTGATTCTGCTCTTAAACATATTCGATCAAGAAAGAACCAATTGATGCATGAGTCAATTTC
TGTGCTTCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAA
GGAGAGGGAGAAAGAAATG----------

Dunalia FUL1

----------TCTACTAAAGGCAAACTTTTTGAATAT---
GCCAATGATTCTAGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATATGCTGAGAGGCA
GCTTGTTCCTACTGATCATTCCTCC---
CCGGAAAGCTGGACTCTGGAGCATGCAAAACTTAAGGCCAGACTTGAGGTTCTACAGAGGAACC
AAAAGCATTACGTGGGAGAAGATTTGGAGTCGTTAAATATGAAAGAACTTCAGAATCTGGAGCA
CCAGCTTGATTCTGCTCTTAAACACATTCGATCAAAGAAGAACCAATTGATGCATGAGTCCATTTC
TGTGCTTCGAAAAAAGGACAAAGCATTGGCGGAGCAAAACAACCAACTTTCCAAGAAGGTGAA
GGAGAGGGAGAAAGAGCTG----------

Iochroma fuchsiodeas FUL1 ----------
GCGCTTGAGCTTTCGGTTTTTTGTGATGCTGGGGTTGGTTTGATTGTTTTTTCTACTAAAGGCAAA
CTCTTTGAATAT---GCCAATGATTCT----------
NAGGGAAGCTGGACTCTGGAGCATGCAAAACTTATGGCCAGACTTGAGGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAAGATTTGGAGTTGTTAAATATGAAAGAACTTCAGAATCTGGAGCAC

Continuation of Table 1.8.

CAGCTTCATTCTGCTCTTAAACACATTCGATCAAAGAAGAACCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAAAGCATTGGCGGAGCAAAACAACCAACTTTCCAAGAAGGTGAAG
GAGAGGGAGAAAGAGCTG----------

Solanum cheesmanii FUL1
ATGGGAAGAGGAAGAGTGCAGCTGAAGAGAATAGAGAACAAAATAAATCGACAAGTTACTTTCT
CAAAACGTCGATCTGGTTTGTTGAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGTT
GGTTTGATTGTTTTTTCAAATAAAGGAAAACTCTTTGAATAC---
GCCAATGATTCCTGCATGGAAAGGACACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GCTTGTCCCTGCTGATCAAACCTCC---
CCGGGAAGCTGGACTCTGGAACATGCAAAACTTAAGGCCAGACTTGAAGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAGGATTTGGATTCGTTAAATATGAAAGAACTTCAGAATCTGGAGCAT
CAACTTGATTCTGCTCTTAAACATATGAGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAAAGCATTGCAGGATCAAAACAACCAGCTTTCCAAGAAGGTGAAG
GAGAAGGAGAAAGAGGTG----------

Solanum pimpinellifolium FUL1
ATGGGAAGAGGAAGAGTCCAGTTGAAGCGAATAGAGAACAAAATTAACCGTCAAGTTACCTTCT
CGAAACGTCGATCTGGTTTGCTGAAGAAAGCCCATGAGATCTCTGTGCTTTGTGATGCTGAGGTT
GGTTTGATTGTTTTTTCTACTAAAGGAAAACTCTTTGAATAT---
GCCAACGATTCCTGCATGGAGAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAAACA
GCTTGTTCCTACTGATCATACCTCC---
CCGGTAAGCTGGACCCTTGAACATGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAAGATTTGGAGTTCTTAAGTATGAAGGAACTTCAGAATCTGGAGCAC
CAGCTTGATTCAGCTCTTAAACACATTCGATCAAGAAAGAATCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAGAGCATTGCAGGAGCAAAACAACCAGCTTTCGAAGAAGGT---------
---

Solanum dulcamara FUL1 ----------
TGAGATCTCTGTGCTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGAAAACTCTT
TCAATAT---
ACCAATGATTCCTGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GCTTGTTCCTACTGATCATACCTCC---
CCGGGAAGCTGGACTCTGGAACATGCAAAACTTAAGGCGAGACTTGAGGTTCTGCAGAGGAAC
CAAAAGCATTACGTAGGAGAAGATTTGGAATCGTTAAATATGAAAGAACTTCAGAAACTGGAAC
AACAACTTGATTCTTCTCTTAAACACATTCGATCAAGAAAGAACCAACTGATGCATGAGTCTATTT
CTGTGCTTCGAAAAAAGGACAAAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGA
AGGAGAGGGAGAAAGAAGGG----------

Solanum commersonii FUL1

TGAGATCTCTGTGCTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGAAAACTCTT
TGAATAT---
GCAACTGATTCATGCATGGAGAGGTTACTTGAAAGATATGAAAGATACTCATTTGCTGAGAAGCA
GCTTGTTCCTACTGATCATACATCC---
CCGGGAAGCTGGACTCTTGAAAATGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAACG
AAAAGCTTTATGTGGGAGAAGATTTGGAGTCGTTAAATATGAAAGAACTTCAGAATCTTGAACAC
CAGCTTGCTTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAACAGGACAGAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAAG
GAGAGGGAGAAAGAGGTG----------

Solanum tuberosum FUL1 ----------
AGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGAAA
ACTCTTTGAATAT---
GCCAATGATTCATGCATGGAGAGGCTACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GCTTGTTCCTACTGATCATACATCC---
CCGGGAAGCTGGACTCTGGAACATGCAAAACTTAAGGCCAGACTTGAGGTTCTTCAGAGGAACC
AAAAGCATTATGTGGGAGAAGATTTGGAGTCGTTAAATATGAAAGAACTTCAGAATCTTGAACAC

Continuation of Table 1.8.

CAGCTTGATTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAACAGGACAGAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAAG
GAGAGGGAGAAAGAGGTG----------

Solanum betaceum FUL1 ----------
TGCGCTTGAACTGTCTATGCTCTGCGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGAAA
ACTCTTTGAATAT---
GCGAATGATTCCTGCATGGAAAGGATACTTGAAAGATATGAAAGATACTCATTTGCTGAGAGGCA
GTTTGTTCCTACTGATCATACCTCC---
CCGGGAAGCTGGACTCTGGAACATGCAAAACTTAAGGCCAGACTTGAAGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAGGATTTGGAGTCATTAAATATGAAAGAACTTCAGAATCTGGAGCAC
CAACTTGATTCTGCTCTTAAACACATTCGATCAAGAAAGAACCAATTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAACGGACAAAGCATTGCAGGAGCAAAACAACCAGCTTTCCAAGAAGGTGAAG
GAGAGGGAGAAAGAGGTG----------

Solanum quitoense FUL1 ----------
GCGCTTGAGATGTCTGTCTTTTGCGATGCTGAGGTTGGTTTGATTGTCTTTTCAAATAAAGGAAAA
CTCTTTGAATAT---
GCCAATGATTCCTGCATGGAAAGGATACTCGAAAGATATGAAAGATACTCATTTGCTGAGAGGAA
GCTTGTTCCTACTGACCATACCTCG---
TCGGGAAGCTGGACTCTGAAACATGCAAAACTTAAGGCTAGACTTGAGGTTCTGCAGAGGAACC
AAAAGCATTATGTGGGAGAGGATTTGGAGTTATTAAATATGAAAGAACTTCAGAATCTGGAGCAC
CAACTTGATTCTGCTCTTAAACACATAAGATCTAGAAAGAACCAAGTGATGCATGAGTCCATTTCT
GTGCTTCAAAAAAAGGACAAAGCATTGCAGGAGCAAAACAATCAGCTTTCCAAGAAGATGAAA
GAGAGGGAGAAAGAGGTG----------

Streptosolen FUL2
ATGGGGAGAGGAAGAGTGCAAATGAAGAGAATTGAGAACAAGATCAATAGGCAAGTTACTTTCT
CGAAGAGGAGAAGTGGGTTGCTGAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGT
TGGTTTGATTGTTTTTTCCACTAAAGGAAAACTCTTTGAGTAC---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTACGAACGATACTCATATGCTGAGAGGCA
GCTCAATCCTGCTGATCAGGACTCC---
CCGGCTAGCTGGACTCTGGAGCATGCTAAGCTTAAGGCTAGAATTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCTGGAGAAGAACTGGACTCTCTAAGTATGAAAGAACTTCAGAATCTGGAGCAT
CAGCTCGATTCTGCTGTCAAACACATTCGATCAAGAAAGAATCAATTGATGCATGAATCTATTTCT
GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCACGAAGCAGGTTAAG
GAAAGAGAAAAAGAGATTGCTCAGCAGAATCAGTGGGAGCAACAAAACCATGATCATCTCAACT
CATCTTCATTTGTGTTGTCACAGCCTATGAACTCTCTTCACATTGGGGAAGCATACCCGGCTGCAG
GAGACAATGGAGAAATTGAAGGATCTTCGCGGCATCAACCACCTAACGTGATGCCGCCATGGATG
Petunia exserta FUL2

GTCGGTGCTCTGCGATGCTGAAGTTGGACTAATTGTTTTCTCCACTAAAGGCAAACTCTTTGAGTA
T---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGATACTCATATGCTGAGAGGCAG
CTTAGTGCCACTGATAATGATACT---
CCGGGGAGCTGGACTCTGGAACATGCTAAGCTTAAGGCCAGGCTTGAAGTATTGCAAAGAAACC
AAAAGCATTATGCGGGAGAAGACTTGGATTCCTTAAGCATGAAAGAGCTTCAAAATTTGGAGCA
GCAGCTCGATTCTGCTCTTAAACAGATTCGATCAAGAAAGAACCAATTGATGCATGAGTCTATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGCGA
AGGAAAGGGAGAAAGAGCTAGCCCAGCAGAGTCAGTGGGAACCACAGAGTCATGAT---
CTCAACTCATCTTCATTCGTTTTGTCACAGCCCTTGAACTCTCTTCACCTTGGGGAAGCATACCCT
AGTGCAGGAGACAATGGAGAAGTTGAAGGGTCTTCAAGGCAGCAACCACCAAACGTGATGCCC
CCCTGGATG
Petunia hybrida FUL2
ATGGGGAGAGGAAGAGTGCAACTCAAGAGAATTGAAAACAAAATCAATCGACAAGTTACGTTTT
CGAAACGACGATCTGGGTTGTTGAGGAAAGCTCATGAGATTTCTGTGCTTTGTGATGCTGAAGTT
GGACTAATTGTTTTCTCCACTAAAGGCAAACTCTTTGAGTAT---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGATACTCATATGCTGAGAGGCAG

Continuation of Table 1.8.

CTTAGTGCCACTGATAATGATACT---
CCGGGGAGCTGGACTCTGGAACATGCTAAGCTTAAGGCCAGGCTTGAAGTTTTGCAAAGAAACC
AAAAGCATTATGCGGGAGAAGACTTGGATTCCTTAAGCATGAAAGAGCTTCAAAATTTGGAGCA
GCAGCTCGATTCTGCTCTTAAACAGATTCGATCAAGAAAGAACCAATTGATGCATGAGTCTATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGTGAA
GGAAAGGGAGAAAGAGCTAGCCCAGCAGAGTCAGTGGGAACCACAGAGTCACGAT---
CTCAACTCATCTTCATTCGTTTTGTCACAGCCCTTGAACTCTCTTCACCTTGGGGAAGCATACCCT
AGTGCAGGAGACAATGGAGAAGTTGAAGGGTCTTCAAGGCAGCAACCACCAAACGTGATGCCC
CCATGGATG
Plowmania nyctaginoides FUL2

GCGCTTGAAATATCGGTTCTTTGTGATGCTGAAGTTGGTTTAATTGTTTTTTCTACTAAAGGCAAA
CTCTTTGAGTAC---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAGAGGTATGAAAGATATTCATATGCTGAGAGGCAG
CTTAGTACTACTGATCAAGACACC---
CCGGGAAGCTGGACACTGGAACATGCTAAGCTTAAGGCCAGACTTGAGATTTTGCAAAGAAACC
AAAATCATTACGCGGGAGAGGATTTGGACACATTGAGTATGAAAGAGCTTCAAAATCTGGAGCA
CCAGCTTGATTCTGCTCTCAAACACGTTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGTTAA
GGNAAGGGAGAAAGAGCTGGCACAGCTGAATCAGTGGGAGCAACAGCACCATGAT---
TTCAATTCATCTTCATTCATTTTGTCACAGCCCTTGAACTCTCTTCACCTTGGTGAAGCATACCCAA
CTGCAGGAGACAATGGAGAAGTTGAAGGATCTTCGCGGCAGCAGCACCCGCCAGTGATGCCCCC
CTGGATG
Withania somnifera FUL2 ----------
TGAGATGTCGATGTTGTGCGATGCTGAAGTTGGTTTGATTGTTTTCTCAAATAAAGGCAAACTATT
TGAGTAT---
TCTACTGATTCTTGCATGGAAAGAATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTACTGCTACTGATGTTGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAACTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCGGGAGAGGACTTGGACTCGTTGAGTATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTTGATTCTGCTATTAAGCATATTAGATCAAGAAAGAACCAATTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAAGCATTGCAAGAACAAAACAACAATCTTTCAAAGAAGGTGAAA
GAAAGGGAGAAAGAGTTGGCCCAACGGACTCCATGGGAGCAACAGAGCCATGATCATCTCAACT
CATCTTCATTCCTTTTGCCACACCCCTTGAACGATCTTCATCTAGGGGAAGCATACCCAACTGCCG
GAGACAATGGAGAAGTTGAAGGATCATCGCAGCAGCAGCCACAAAACGTGATGCCCCCCTGGAT
G
Physalis pubescens FUL2 ----------
TGAGATCTCCGTGCTTTGTGATGCTGAAGTTGGTTTGATCGTTTTCTCAAATAAAGGCAAACTATT
TGAGTAT---
TCTACTGATTCTTGCATGGAAAGAATTCTTGAAAGGTATGAGAGGTACTCATATGCTGAGAGGCA
GCTTAATGCTACTGATATCGAAACC---
CCGGGAAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCTAAGAAACC
AAAGGCATTATGCGGGAGAGGACTTGGACTCATTGGGTATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTTGATTCTGCTCTTAAGCATATTAGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAGGAACAAAACAACAATCTTACGAAGAAGGTGAA
GGAAAGGGAGAAAGAGTTGGCCCAGCGGACTCCGTGGGAGCAGCAGAGCCATGATCATCTCAA
CTCATCTTCATTCGTTTTGCCACACCCCTTGAACAGCCACCACCTTGGGGAAGCATACCCAACTG
CAGGAGACAATGGAGAAGTTGAAGGATCCTCGCAGCAGCAGCAGCAACAAGTGATGCCGCCAT
GGATG
Nicotiana sylvestris FUL2
ATGGGGAGAGGAAGAGTGCAACTGAAGAGAATTGAGAACAAGATCAATCGACAAGTCACCTTC
TCAAAAAGAGCATCTGGTTTGCTTAAGAAAGCTCATGAAATCTCTGTGCTTTGTGATGCTGAGGT
TGGTTTAATTGTTTTTTCTACTAAAGGGAAACTCTTTGAGTAT---
TCCACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTACTGCTACTGATGATGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCAGGAGAAGATTTGGACTCATTAAGTATGAAAGAGCTTCAGAATCTTGAGCAC
CAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAATCAATTGATGCATGAATCCATTTCT

Continuation of Table 1.8.

GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAATCTCTCAAAGCAGGTGAAA
GAAAGGGAGAAAGAGCTAGCTCAGCAGACTCAATGGGAGCAACAGAGCCATGATCATCTCAACT
CATCTTCATTCGTTTTAACACAGCCCTTGAGCTCTCTTCACCTCGGGGAAGCGTACCCGACTGCA
GGAGACAACGGAGAAGTGGAAGGATCATCGCGGCAACAACAACAAAACGTGATGCCGCCATGG
ATG

Nicotiana obtusifolia FUL2

ATGGGGAGAGGAAGAGTGCAACTGAAGAGAATTGAGAACAAGATCAATCGACAAGTCACCTTC
TCAAAAAGGCGATCTGGGTTGCTCAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAGGT
TGGTTTAATTGTTTTTTCTACAAAAGGCAAACTCTTTGAGTAT---
TCCACCGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGATACTCATATGCCGAGAGGCA
GCTTACTGCTACTGATCATGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCAAGACTTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCAGGAGAAGATTTGGACACATTAAGTATGAAAGAGCTGCAGAATCTTGAGCAC
CAGCTCGATTCTGCTTTAAAGCACATTCGATCAAGAAAGAATCACTTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGTGAA
AGAAAGGGAGAAAGAGATGGCTCAGCAGACTCAGTGGGAGCAACAGAGCCATGATCATCTCAA
CTCATCTTCATTCGTTTTGTCACAGCCCTTGAGCTCTCTTCACCTTGGGGAAGCGTACCCGACTGC
AGGAGACAACGGAGAAGTTGAAGGATCATCGCGGCAACAACAACAGAACGTGATGCCGCCATG
GATG

Nicotiana tabacum FUL2

ATGGGGAGAGGAAGAGTGCAACTGAAGAGAATTGAGAACAAGATCAATCGACAAGTCACCTTC
TCAAAAAGACGATCTGGTTTGCTCAAGAAAGCTCATGAGATCTCTGTACTTTGTGATGCTGAGGT
TGGTTTAATTGTTTTTTCTACAAAAGGCAAACTCTTTGAGTAT---
TCCACCGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTACTGCTACTGATCATGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCAAGATTTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCAGGAGAAGATTTGGACTCATTAAGTATGAAAGAGCTGCAGAATCTTGAGCAC
CAGGTCGATTCTGCTTTAAAGCACATTCGATCAAGAAAGAATCAATTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGTGAA
AGAAAGGGAGAAAGAGCTGGCTCAGCAGACTCAGTGGGAGCAACAGAGCCATGATCATCTCAA
CTCATCTACATTCGTTTTGTCACAGCCCTTGAGCTCTCTTCACCTTGGGGAAGCGTACTCAACTGC
AGGAGACAACGGAGAAGTTGAAGGATCATCGCGGCAACAACAACAGAACGTAATGCCGCCATG
GATG

Nicotiana tomentosiformis FUL2

TGAGATCTCTGTACTTTGTGATGCTGAGGTTGGTTTAATTGTTTTTTCTACAAAAGGCAAACTCTT
TGAGTAT---
TCCACCGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTACTACTACTGATCATGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCAAGAATTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCAGGAGAAGATTTGGACTCATTAAGTATGAAAGAGCTGCAGAATCTTGAGCAC
CAGGTCGATTCTGCTTTAAAGCACATTCGATCAAGAAAGAATCAATTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTCTCGAAGCAGGTGAA
AGAAAGGGAGAAAGAGCTGGCTCAGCAGACTCAGTGGGAGCAACAGAGCCATGATCATCTCAA
CTCATCTACATTCGTTTTGTCACAGCCCTTGAGCTCTCTTCACCTTGGGGAAGCGTACTCAACTGC
AGGAGACAACGGAGAAGTTGAAGGATCATCGCGGCAACAACAACAGAAC----------

Solanum pimpinellifolium FUL2

ATGGGTAGAGGAAGAGTACAATTGAAGAGAATTGAGAACAAAATTAATCGTCAAGTTACTTTTTC
AAAGAGGCGATCTGGTTTGCTTAAAAAAGCTCATGAGATCTCTGTGCTTTGCGATGCTGAAGTTG
GACTCATTGTTTTCTCAACTAAAGGAAAACTCTTTGAGTAT---
TCTACTGACTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAAAGGCA
GCTTAATGCTACTGATATTATAACC---
CCGGGTAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGAAGAGTTGGACACATTGAGTATGAAAGAGCTTCAGAATCTGGAACA
CCAGCTCGATTCTGCTCTTAAGCACATTCGATCTAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTTCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTTAA
GGAAAGGGAGAAAGAGATGGCCCAACAGACTCCGTGGGAGCAACAGAGTCATGATCATCTCAAT
TCATCTTCGTTTGTTTTGCCACACCCCTTTAACAATCTTCACATAGGGGAAGCATACCCAAATGCA

Continuation of Table 1.8.

GGAGACAATGGAGAAGTAGAAGGATCATCGCGGCAACAACAACAAAACGTGATGCCTCCATGGA
TG
Solanum tuberosum FUL2 ----------
TGAGATCTCTGTGCTTTGCGATGCTGAAGTTGGACTCATTGTTTTTTCAACTAAAGGAAAACTCTT
TGAGTAT---
TCCACCGACTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAAAGGCA
GCTTAATGCTACTGATATTGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGAAGAGTTGAACACATTGAGTATGAAAGAGTTTCAGAATCTGGAACA
CCAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGCGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAATAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGAGATGTCCCAACAGACTCCGTGGGAGCAACAGAGTCATGATCATCTCAAT
TCATCTTCGTTTGTTTTGCCACACCCCTTTAACAACCTTCACATGGGGGAAGCATACCCAACTGCA
GGAGACAATGGAGAAGTTGAAGGATCATCGCGGCAGCAACAACAAAAC----------
Solanum sisymbriifolium FUL2

GCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCAGCTTAATGCTACT
GATATCGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGAAGAGTTGGACTCATTGAGTATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGAGCTGGCCCAGCAAACTCCATGGGAGCAACAGAGTCATGACCATCTCAA
TTCATCTTCATTCGTTTTGCAACACCCCTTTAACAACCTTCACTTAGGGGAAGCATACCCAACTGC
AGGAGACAATGGGGAAATTGAAGGATCATCGAGGCAGCAACAACAAAACGTGATGCCGCCATG
GATG
Solanum xanthocarpum FUL2
ATGGGAAGAGGAAGAGTACAGCTTAAGAGAATTGAAAACAAAATCAATCGTCAAGTCACTTTTT
CAAAGAGGCGATCTGGTTTACTCAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAAGTT
GGACTCATTGTTTTTTCAACTAAAGGAAAATTATTTGAGTAT---
TCAACAGACTCATGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTAATGCTACTGATATCGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGCAGAGTTGGACTCATTGAGTATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGAGCTGGCCCATCAAACTCCTTGGGAGCAACAGAGTCATGATCATCTCAAT
TCATCTTCGTTCGTTTTGCCACACCCCTTTAACAACCTTCACTTAGGGGAAGCATACCCAACTGCA
GGAGACAACGGAGAAATTGAAGGATCATCGAGGCAGCAACAACAAAACGTGATGCCGCCATGG
ATG
Solanum dulcamara FUL2

CC---
CCGGGGAGCTGGACTTTGGAATATGCTAAACTTAAGGCCAGACTTGACGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGAAGAGTTGGACTCATTGAGTATGAAAGAGCTTCAAAATCTGGAACA
CCAGCTCGATTCTTCTCTTAAGCATATTCGATCGCGAAAGAACCAATTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAGGCACTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTGAAG
GAAAGGGAGAAAGAGATGGCCCAGCAGACTCCGTGGGAGCAACA----------

Solanum ptychanthum FUL2
ATGGGGAGAGGAAGAGTACAACTTAAGAGAATTGAAAACAAAATTAATCGTCAAGTAACTTTTT
CAAAGAGACGATCTGGTTTACTTAAGAAAGCTCATGAGATCTCTGTGCTTTGCGATGCGGAAGTT
GGACTCATTGTTTTTTCAACTAAAGGAAAACTCTTTGAGTAT---
TCCACTGACTCTTGCATGGAAAGGATACTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
ACTTAATGCTACTGATATCGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAACTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC

Continuation of Table 1.8.

AAAAGCATTATGCAGGAGAAGAGTTGGACACATTGAGTATGAAAGAACTTCAGAATCTGGAGCA
CCAGCTCGATTCTGCTCTTAAGCATATTCGATCACGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGAGATGGCCCAGCAGACTCCGTGGGAGCAACAGAGTCAGGATCATCTCAA
TTCATCTTCATTCATTTTGCCACACCCTTTTAACAACCTTCACCTAGGGGAAGCATACCCAACTGC
AGGAGACAATGGAGAAGTTGAAGGATCGTCACGGCAGCAACAACAAAACGTGATGCCACCATG
GATG

Brunfelsia FUL2

ATGGGAAGAGGAAGAGTTCAGCTGAAGAGAATTGAGAACAAAATCAATCGACAAGTTACGTTTT
CGAAACGTCGATCTGGTTTGTTGAAAAAAGCTCATGAAATTTCAGTACTTTGTGACGCTGAAGTT
GGATTAATTGTTTTTTCCACTAAAGGCAAACTCTTTGAGTAT---
TCAAATGATTCTTGCATGGAAAGGATTCTTGAGAGGTATGAAAGATACTCATATGCTGAGAGGCA
GCTTAATGCTACTGATCATGACACC---
CCGGGGAGCTGGACACTGGAACATGCTAAGCTTAAGGCCAGACTTGAAGTTTTGCAAAGAAATC
ATAAGCACTATGCGGGGGAACACTTGGACTCATTAAGTATGAAAGAGCTTCAGAATCTGGAGCAT
CAGCTCGATTCTGCTCTTAAACAAGTTCGATCAAGAAAGAACCAATTGATGCACGAATCCATTAC
TGAGCTACAAAAGAAGGACAAGGCGTTGCAAGAACAAAACAACAAGCTCTCTAAGCAGGTGAA
GGAAAGGGAGAAAGAGCTAGCCCAGCAGAGTCAGTGGGAGCAACAAAGCCATGAT---
CTCAACTCATCTTCATTCGTTCTGACACAGCCCTTGAACTCTCTTCACATTGGTGAAGCATACCCA
ACAACAGGAGACAATGGAGAAGTTGAAGGATATTCGCGGCAACAACCTCAAAACGTGATGCCCC
CATGGATG

Brugmansia suaveolens FUL2 ----------

TGCGCTTGAGATCTCTGTGTTGTGTGATGCTGAGGTTGGTTTGATTGTTTTTTCCACTAAAGGAAA
ACTCTTTGAGTAC---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTACGAACGATACTCATATGCTGAGAGGCA
GCTCAATCCTACTGATCAGGACTCC---
CCGGCGAGCTGGACTCTGGAGCATGCTAAGCTTAAGGCTCGAATTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCGGGGGAAGACCTGGACCCTCTAAGTATGAAAGAGCTTCAGAATCTGGAGCA
TCAGCTCGATTCTGCTCTCAAACATATTCGATCAAGAAAGAACCAATTGATGCATGAACCTATTTC
TGAGCTGCAAANGAAGGACAAGGCATTGCAAGAGCAAAACAACAAGCTTTCGAAGCAGGTAAA
GGAAAGAGAAAAAGAGATTGCTCAGCAGAATCAGTGGGAGCAACAAAACCATGATCATCTCAA
CTCATCTTCATTTGTGTTGTCACAGCCTATGAACTCTCTTCACATTGGGGAAGCATACCCGACTGC
TGGAGACAATGGAGAAATTGAAAGATCTTCGCGGCAACAACCACCAAAC----------

Cestrum diurnum FUL2

ATCACTGTGCTTTGTGATGCTGAAGTTGGTTTGATTGTTTTTTCCACTAAAGGCAAACTCTTTGAG
TAC---
TCTACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCTTATGCTGAGAGGCA
GCTCAATCCTACTCAT---GACACC---
CCGGGTAGCTGGATTCTGGAACATGCTAAGCTTAAGGCTAGACTTGAGGTTTTGCAAAGAAACCA
AAGGCATTATGCCGGAGAAGACTTGGACTCATTAAGTACGAAGGAACTTCAGAATCTGGAGCAC
CAACTTGATTCTGCTCTCAAACACATTCGATCAAGAAAGAACCAATCGATGCATGAATCAATCTC
AGAACTGCAAAAAAAGGAGAAGGCATTGCAAGAGCAAAACAACAAACTCTCAAAACAGGTAA
AGGAAAGGGAGAAAGAGCTGGCTCAGCAGAATC----------

Atropa belladonna FUL2

ATGGGGAGAGGAAGAGTACAGTTGAAGAGGATTGAGAACAAAATTAATCGGCAAGTGACCTTCT
CGAAAAGGCGATCTGGGTTGTTGAAGAAAGCKCWTGARMTSTCKGTSCTWTGTGATGCTGAAG
TTGGTTTAATTGNTTTTTCAACTAAAGGCAAACTCTTTGAGTAT---
TCCACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTAATGCTACTGCTATCGAAACC---
CCGGGGAGCTGGACTCTGGAACRTGCTAAGCTTAAGGCCAGACTTGAGGTCTTGCAAAGAAACC
AAAGGCATTATGCGGGAGAAGACTTGGACTCGTYGAGTATGAAAGAGCTTCAGAATTTGGAGCA
CCAACTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTACAAGAGCAAAACAACAATCTCTCAARGCAGGTGAA
GGAAAGGGAGAATGAGATAGCCCAGCAGAATCAGTGGGAGCAACAAAGCCATGATCATCTCAAC
TCATCTTCATTCGTTATGTCACACCCCTTGAACAACCTTCACCTAGAGGAAGCATACCCGACTGCA

Continuation of Table 1.8.

GGAGACAATGGAGAAGTTGAAGGATCGTCGCGACAGCAACAACAAAACGTGATGCCCCCCTGG
ATG
Solanum lycopersicum FUL2
ATGGGTAGAGGAAGAGTACAATTGAAGAGAATTGAGAACAAAATTAATCGTCAAGTTACTTTTTC
AAAGAGGCGATCTGGTTTGCTTAAAAAAGCTCATGAGATCTCTGTGCTTTGCGATGCTGAAGTTG
GACTCATTGTTTTCTCAACTAAAGGAAAACTCTTTGAGTAT---
TCTACTGACTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAAAGGCA
GCTTAATGCTACTGATATTATAACC---
CCGGGTAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTATGCAGGAGAAGAGTTGGACACATTGAGTATGAAAGAGCTTCAGAATCTGGAACA
CCAGCTCGATTCTGCTCTTAAGCACATTCGCTCTAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTTCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTTAA
GGAAAGGGAGAAAGAGATGGCCCAACAGACTCCGTGGGAGCAACAGAGTCATGATCATCTCAAT
TCATCTTCGTTTGTTTTGCCACACCCCTTTAACAATCTTCACATAGGGGAAGCATACCCAAATGCA
GGAGACAATGGAGAAGTAGAAGGATCATCGCGGCAACAACAACAAAACGTGATGCCTCCATGGA
TG
Capsicum annuum FUL2
ATGGGAAGAGGAAGAGTTCAATTGAGGAGGATTGAAAATAAGATAAATAGGCAAGTGACTTTTT
CGAAGAGGCGATCTGGTTTGTTGAAGAAAGCTCATGAGATCTCTGTCCTTTGTGATGCTGAAGTT
GGCTTGATTGTTTTTTCTTCTAAAGGGAAACTATTTGAGTAT---
TCTACTGACTCTTGCATGGAAAGGATTCTTGAGAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTAATGCAACTGATGTCGAAACC---
CCGGGGAGTTGGACTTTGGAACATGCTAAGCTTAAGGCCAGGCTTGAGGTTTTGCAAAGAAACC
AAAGGCATTATGCGGGAGAAGACTTGGACTCATTGAGTATGAAAGAGCTTCAGAATCTGGAGCA
GCAACTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGATGAA
GGAAAGGGAGAAACAGCTGGCCCAGCAGACTCCGTGGGAGCAACAGAACCATGACCATCTCAA
CTCATCTTCATTTGGTCTGCCACATCCCTTTAACAACAATCACCTAGGGGAAGTATATCCAACTGC
AGGAGACAATGGAGAAGTTGAAGGATCATCGCGGCAGCAACAACAAAACGTGATGCCGCCATG
GATG
Dunalia FUL2
ATGGGGAGAGGAAGAGTTCAGCTGAAGAGGATTGAGAACAAAATCAATAGGCAAGTCACTTTCT
CCAAGAGGCGATCTGGTTTGCTAAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAAGTT
GGTTTGATTGTTTTCTCAACTAAAGGCAAATTATTTGAGTAT---
TCCACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTAATGCTACTGATGTCGAAACC---
CCGGACAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGATGTTTTGCAAAGAAACC
AAAGGCATTATGCGGGAGAAGACTTAGACTCATTGAGTATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTC
TCAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGAGCTGGTCCAGCAGACTCCGTGGGAGCAACAGAGCCATGATCATCTCAA
TTCATCCTCATTCGTTTTGCCACACCCCTTGAACAACCTTCACCTAGGGGAAGCATACCCAACTGC
AGGAGGCAATGGAGAAGTTGAAGGATCATCGCAGCAGCACCAACAAAACGTGATGCCGCCATG
GATG
Datura inoxia FUL2 ----------
TTCGCTTGAAATTTCGGTGCTTTGTGATGCTGAAGTTGGTTTGATTGTTTTCTCATCTAAAGGCAA
ACTCTTTGAGTAT---
TCCACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGACA
GCTTAATGCTACTGAT---GAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAGGTTTTGCAAAGAAACC
AAAAGCATTACGCAGGAGAAGACTTGGAATCATTGAGCATGAAAGAGCTTCAGAATCTGGAGCA
CCAGCTTGATTCTGCTCTTAAGCACATTAGATCAAGAAGGAATCAATTGATGCATGAATCAATTTC
TGAGCTGCAAAAGAAGGACAAGGCATTACAAGAACAAAACAACAATCTTTCAAAGCAGGTGAA
GGAAAGGGAGAAAGGACTGGCCCAGCAGACTCAGTGGGAGCAACAGAGCCATGATCATCTCAA
CTCATCTTCATTCATTTTGCCACACCCCTTGAACAACCTTCACCTTGGGGAAGCATACCCAACTGC
AGGAGACAATGGAGAAGTTGAAGGATCGTCGAGGCAGCAACAACACAACGTGATGCCCCCCTG
GATG

Continuation of Table 1.8.

Lycium barbarum FUL2
ATGGGGAGAGGAAAATTGCAACTGAAGAGGATTGAGAATAAAATAAATCGGCAAGTGACGTTCT
CTAAGAGGCGATCTGGGTTGCTTAAGAAAGCTCAGGAGATCTCTGTGCTTTGTGATGCTGAAGTT
GGGTTGATTGTTTTTTCAACTAAAGGCAAACTCTTTGAGTAT---
TCCACCGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCG
GCATAATCCTACTGATCAGGAAACC---
CCGGGGAGCTGGACTCTGGAATATGCTAAGCTTAAGGCCAGACTTGAAGTTTTGCAAAGAAACC
AAAGGCATTATGTGGGAGAAGACTTGGAGTCGTTAAGTATGAAGGAGCTTCAGAATCTGGAGCA
CCAGCTTGATTCGGCTCTGAAGCACATCCGATCAAGAAAGAACCAATTGATGCATGAATCCATTT
CTGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAATCTCTCAAAGCAGGTGA
AGGAAAGGGAGAAAGAGATAGCCCGGCAGAGTCAGTGGGAGCAACAGAGCCATGATCATCTCA
ACTCATCTTCATTCGTTTTGTCACACCCCTTGAACAACCTTCACCTAGGGGAAGCATACCCGGATG
CAGGAAACCAGGGAGAAGTTGAAGGATCATCGCGGCACCAACCACAAAACGTCATGCCGCCAT
GGATG
Lycium sp. FUL2
ATGGGGAGAGGAAAAGTGCAACTGAAGAGGATTGAGAATAAAATAAATCGGCAAGTGACGTTCT
CTAAGAGGCGATCTGGGTTGCTTAAGAAAGCTCAGGAGATCTCTGTGCTTTGTGATGCTGAAGTT
GGGTTGATTGTTTTTTCAACTAAAGGCAAACTCTTTGAGTAT---
TCCACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCATAATCCTACTGATCAGGAAACC---
CCGGGGAGCTGGACTCTAGAATATGCTAAGCTTAAGGCCAGACTTGAAGTTTTGCAAAGAAACC
AAAGGCATTATGTGGGAGAAGACTTGGAGTCGTCAAATATGAAGGAGCTTCAGAATCTGGAGCA
CCAGCTTGATTCGGCTCTGAAGCACATCCGATCAAGAAAGAACCAATTGATGCATGAATCCATTT
CTGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAATCTCTCAAAGCAGGTGA
AGGAAAGGGAGAAAGAGATAGCCCAGCAGAGTCAGTGGGAGCAACAGAGCCATGATCATCTCA
ATTCATCTTCATTCGTTTTGTCACACCCCTTGAACAACCTTCACCTAGGGGAAGCATACCCGGATG
CAGGAAACCATGGAGAAGTTGAAGGATCATCGCGGCACCAATCACAAAACGTCATGCCACCATG
GATG
Grabowskia FUL2
ATGGGGAGAGGAAAAGTGCAACTGAAGAGGATTGAGAATAAAATAAATCGGCAAGTGACGTTCT
CTAAGAGGCGATCTGGGTTGCTTAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCTGAAGTT
GGGTTAATTGTTTTTTCAACTAAAGGCAAACTCTTTGAGTAT---
TCAACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATACACTGAGAGGCA
GCTTAATCCTACTGATCAGGAAACC---
TTGGGGAGCTGGACTCTGGAATATTCTAAGCTTAAGGCCAGACTTGAAGTTTTGCAAAGAAACCA
AAGGCATTATGCGGGAGAAGATTTGGAGTCATTAAGTATGAAGGAGCTTCAGAATCTGGAGCACC
AGCTCGATTCTGCTGTGAAGCACATCCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTCT
GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAAC----------

Datura metel FUL2

GCATGAATCAATTTCTGAGCTGCAAAAGAAGGACAAGGCATTGCAAGAACAAAACAACAATCTT
TCAAAGCAGGTAAAGGAAAGGGAGAAAGGGCTGGCTCAGCAGACTCAGTGGGAGCAACAGAG
CCATGATCATCTCAACTCTTCTTCGTTCGTTTTGCCACACCCCTTGAACAACCTTCACCTTGGGGA
AGCATACCCGACTGCAGGAGATAATGGAGAAGTTGAAGGATCGTTGCGGCAGCAACAACACAAC
GTGATGCCGCCATGGATG
Nicandra physalodes FUL2 ----------
TGCGCATGAAATTTCGGTGCTGTGTGATGCTGAAGTTGGACTTATTGTTTTTTCTACTAAAGGAAA
ACTATTTGAGTAT---
TCAACTGATTCTTGCATGGAAAGGATTCTTGAAAGGTATGAAAGGTACTCATATGCTGAGAGGCA
GCTTAATGCTACTGAGCTCGAAACC---
CCGGGGAGCTGGACTTTGGAACATGCTAAGCTTAAGGCCAGACTTGAAGTTCTTCAAAGAAACC
AAAGGCATTATGCGGGAGAAGATTTGGATTCATTAAGTATGAAAGAGCTTCAGAATCTGGAGCAC
CAGCTCGATTCTGCTCTTAAGCACATTCGATCAAGAAAGAACCAATTGATGCATGAATCCATTTCT

Continuation of Table 1.8.

GAGCTGCAAAAGAAGGACAAGGCATTGCAAGAGCAAAACAACAATCTTTCAAAGCAGGTGAAG
GAAAGGGAGAAAGAGATGGCCCAGCAGAGTCAATGGGAGCAACAGAGTCATGATCATCTCAATT
CATCTTCATTCGCTTTGTCACACCCCTTGAATAACCTTCACCTAGGAGAAGCATACCCACCTGCAG
GAGACAATGGAGAAATCGAAGGATCGTCAAGGCAGCAACAACAAAACGTGATGCCCCCCTGGAT
G

Nicotiana obtusifolia MBP10

CCACCGGATCCAGCATGGAAAGTATCCTCGAAAGATACGAAAGTTATTCATATGCTGAGAGGAAG
TTGAATGCAAATGACTCTGAACCT---
AAGGAAAACTGGACTCTGGAGTACCCAAAGCTCATGTCAAGGATTGAACTTCTCCAAAGAAATAT
AAGGCATTATATGGGAGAGGATTTGGGTACCTTCGGTCTGCGAGAGTTTGATGGTTTGGAGCAAC
AACTCGATACAGCTTTGAAGCGAATACGCACCAGGAAGAACCAACTGATGCATGAGTCCATTTCC
CAGCTACGGAAAAAGGAAAAAGAGCTGCAAGAGCAAAACCACTTAATGTCGAAGAAGCTGAAA
GGAAATGAGAAG----------

Nicotiana tomentosiformis MBP10

CCACCGAATCCAGCATGGAAAGTATCCTCGAAAGATACGAAAGTTATTCATATGCTGAGAGGAAG
TTGAATGCAAATGACTCTGAACCT---
AAGGAAAACTGGACTCTGGAGTACCCAAAGCTCATGTCAAGGATAGAACTTCTGCAAAGAAATA
TAAGGCATTATATGGGAGAGGATCTGGATTCCTTCGGTCTGCGGGAGTTTCATGGTTTAGAGCAAC
AGCTTGATACAGCTTTGAAGCGAATACGAACTAGGAAGAATCAACTGATGCATGAGTCCATTTCC
CAACTGCAGAAAAAGGAAAAAGAGTTGCAAGAGCAAAACCACTTAATGTCGAAGAAGCTGAAA
GGAAATGAAAAG----------

Brunfelsia MBP10

ATGGGAAGGGGTAAGGTTCAATTGAAGAGGATCGAAAACAAGATTAGCAGGCAAGTTACTTTCT
CAAAGAGACGCTCCGGTTTGTTGAAGAAAGCTCATGAGATCTCAGTCTTGTGTGATGCGGATGTT
GCTTTGATTGTCTTCTCTGCAAAAGACAAGCTCTTTGAGTAC---
TCCACTGAATCTGGCATGGAAAATATCCTGGAAAGATACGAAACATACTCATACGCCGAGAGGAA
GCTGAATGCGAATGACTCTGAACCTAATGAGGTAAACTGGAATCTTCAGTACCAAAAGCTCATGG
CAAGGAATGAACTTCTGCAAAAAAATATAAGGCATTATATTGGAGAGGATTTGGATTCCCTCGGTA
TGCGAGAGTTTCAAGGTTTAGAGCAACAGCTCGATACAGCTTTGAAGCGAATACGAACAAGGAA
GAACCAACTGATGCATGATTCCATTTCCCAGCTGCAGAAAAAGGAAAAAGAGCTGCAAGAGCAA
AAGAACTTGATGTCGAAGAAGCTGAAAGAAAATGAGAAA----------

-

Solanum ptychanthum MBP10

ATCCAGTATGGAAAATATACTGGAAAGATATGAAAGTTACTCATATGCGGAGAGGAACTTGAAT----
----
TATAAGGAAAAGTGGAGTCTCGAGTACCCAAAGCTCACGGCTAGGGTTGAACTTCTGCAAAGAA
ATATAAGGCATTTTATGGGAGAAGATCTGGACGCCTTTAATTTGCGTGAGTTTCAGGGTTTAGAGC
AACAGCTCGATACCGCTCTGAAGCGAGTACGAACTAAGAAGAATCAACTGATGCATGAGTCCATT
TCCCAGCTGCAGAAAAAGGAAAAAGAACTGCAAGAGCGAAACAACTTAATTTCCAAAAAGCTT
AAAGAAAATGAGAAG----------

Solanum tuberosum MBP10

TGTGATGCTGACGTGGCATTAATTGTCTTCTCTTCAAATGGCAAGCTCTTTGAGTAC---
TCCACTCAATCCAGCATGGAAAATATATTGGAAAGATATGAAAGTTACTCATCTGCGGAAAGGAA
CTTGAAT----------
TATAAGGAAAACTGGACTCTCGAGTACCCAAAGCTCATGGCAAGAGTTGAACTTCTGCAAAGAA
ATATAAGGCATTTTATGGGAGAAGATCTGGATGCCTTTAATCTGCGTGAATTTCAGGGTTTAGAGC
AACAACTCGATACAGCTCTGAAACGAGTGCGATCTAGGAAGAATCAACTGATGCATGAGTCCATT

Continuation of Table 1.8.

TCCCAGCTGCAGAAAAAGGAAAAAGAACTGCAAGAGCGAAACAACTTAATTTCTAAGAAGCTTA
AAGAAAATGAGAAG----------

Solanum lycopersicum MBP10
ATGGGGCGGGGTAGGGTGGAGATGAAGCGTATCGAAAATAAAATAAGCAGACAAGTTACATTCT
CAAAGAGACGATCCGGTTTGTTGAAGAAAACCAACGAGATCTCTGTGCTATGTGATGCTGAGGT
GGCATTAATTGTTTTCTCTTCAAATGGAAAACTATTTGAGTAC---
TCTACTCAATCAAGCATGGAAAATATATTGGAAAGATATGAAAATTACTCATACGAGGAGATGAAC
TTGAAT----------
TATAAGGAAAATTGGACTCTTGAGTACCCAAAGCTCATGGCAAGAGTTGAACTTCTGCAAAGAAA
TATAAGGCATTTTATGGGAGAAGATCTGGACGCCTTTAATCTGCGTGAATTTCGGGGTTTAGAGAA
ACAGCTCGATACAGCTCTAAAGCGAGTGCGATCTAAGAAGAACCAACTGATGCACGAGTCCATTT
CCCAGCTGCAGAAAAAGGAAAAAGAACTGCAACAGCGAAACAACTTAATTTCTAACAAGCTTAA
AGAAAATGAGAAG----------

Solanum pimpinellifolium MBP10
ATGGGGCGGGGTAGGGTGGAGATGAAGCGTATCGAAAATAAAATAAGCAGACAAGTTACATTCT
CAAAGAGACGATCCGGTTTGTTGAAGAAAACCGACGAGATCTCTGTGCTATGTGATGCTGAGGT
GGCATTAATTGTTTTCTCTTCAAATGGAAAACTATTTGAGTAC---
TCTACTCAATCAAGCATGGAAAATATATTGGAAAGATATGAAAATTACTCATACGAGGAGATGAAC
TTGAAT----------
TATAAGGAAAATTGGACTCTTGAGTACCCAAAGCTCATGGCAAGAGTTGAACTTCTACAAAGAAA
TATAAGGCATTTTATGGGAGAAGATCTGGACGCCTTTAATCTGCGTGAATTTCGGGGTTTAGAGCA
ACAGCTCGATACAGCTCTAAAGCGAGTGCGATCTAAGAAGAACCAACTGATGCACGAGTCCATTT
CCCAGCTGCAGAAAAAGGTAAAAGAACTGCAACAGCGAAACAACTTAATTTCTAACAAGCTTAA
AGAAAATGAGAAG----------

Jaltomata procumbens MBP10 ----------
TGCGCTTGAACTTTCGATACTTTGTGATGCTGAGGTGGCATTGATTGTTTTCTCCCCCAATGGAAA
GCTCTTTGAGTAC---
TCCACTGAATCCAGCATGGAAAATATACTGGAAAGATACGAAAATTACTCATATGCGGAGAGGAA
GTTGAATGGAAATGATTCTCAAACTTATAAGGAAAACTGGACTCTAGAGTACCCAAAGCTCATGG
CAAGGGTTGAACTTCTTCAAAGAAATATAAGGCATTTTATGGGAGAGGATCTGGATGCCTTCAATC
TGCGAGAGTTTCAGGGTTTAGAGCAACAACTCGATACAGCTCTCAAGCGAATACGAACCAGGAA
GAATCAACTGATGCATGCGTCCATTTCCCTGCTGCAGAAAACGGAAAAAGAACTGCAAGAGCGA
AACAACTTAATTTCCAAGAAGCTTAAAGAAAATGAGAAG----------

Juanalloa mexicana MBP10 ----------
TGCGCATGAACTGTCTGTGCTGTGTGATGCTGAGGTGGCATTGATTGTCTTCTCCCCCAAGGGCA
AGCTCTTTGAGTAC---
TCCACTGAATCCAGCATGGAAAATATACTGGAAAGATACGAAAGTTACTCATATGCAGAGAGGAA
GTTGAATACAAATGACTCTCAAACTTATAAGGAAAACTGGACGCTAGAGTACCCAAAGCTCCTGG
CAAGGGTTGAACTTCTGCAAAAAAATATAAGGCATTTTATGGGAGAGGATCTGGATGCCTTCAAT
CTGCGTGGGTTTCAGGGTTTAGAGCAACAGCTCGATACAGCTCTGAAGCGAATACGAACCAGGA
AGAACCAACTGATGCATGAGTCCATTTCCCTGCTGCACAAAAAGGAAAAAGAACTGCAAGAGCG
AAACAACTTAATTTCCAAGAAGCTTAAAGAAAATGAGAAG----------

--
Nicandra physaloides MBP10 ----------
TGCGCTTGAACTGTCAATTTTGTGTGATGCTGATGTGGCATTGATTGTCTTCTCCTCCAATGGCAA
GCTCTTTGAGTAC---
TCCACTCAGTCCAGCATGGAAAGTATCCTGGAAAGATATGAAAGTTACTCGCATGGGGAGAGGA
AATTGAATGCAAATGATTCTCAAACTTATAAGCCAAACTGGGCGCTCGAGTTCCCAAAGCTCATG
TCGAGGGTTGAACTTCTCCAAAGAAATATAAGGCATTTTATGGGAGAGGATTTGGATGCCTTCAAT
CTGCGTGAGTTTCAGAGTTTAGAGCAACAAATTGATACAGCTCTGAAACGAATACGAACCAAGA
AGAACCAACTGATGCATGAGTCCATTTCCCTACTGCAGAAGAGGGAAAAAGAACTGCAAGAGAG
GAACAACTTAATTTCCAAGAAGCTTAAAGAAAATGAGAAG----------

Continuation of Table 1.8.

--
Brugmansia suaveolens MBP10

AGATGTCGGTGTTGTGTGATGCTGAGGTGGCATTGATTGTCTTCTCCCCTAAAGGCAAGCTCTTTG
AGTAC---
TCCACTCAATCCAGCATGGAAAATATGCTGGAAAGATACGAAAGTTACTCCTATGCG----------
----
CAAACTTCTAAGGAAAACAGGACGCTGGAGTACCAAAAGCTCACGGCAAGGGTTGAACTTCTG
CAAAGAAATATAAGGCATTTTATGGGAGAGGATCTGGATGCCTTCAATCTGCGAGAGTTTCAGGG
TTTAGAGCAACAGATTAATACAGCTCTGAAGCGAATACGAACCAGGAAGAACCAACTGATGCTT
GAGTCCATTTCCCTGCTGCAGAGAAAGGAAAAAAGACTGCAAGAGCAAAACAACTTAATTTCCA
AGAAGCTTAAAGAATATGAGAAG----------

Solandra maxima MBP10

TCTGTGTTGTGTGATGCTGAGGTGGCATTGATTGTCTTCTCCCCCAAAGGCAAGCCCTTTGAGTAC

TCCACTGAATCAAGCATGGAAAATATACTGGAAAGATACGAAAGTTACTCATATGCGGAGAAGAA
GTTGAATGCTAATGACTCTCAAACTTATAAGGAAAACTGGACACCAGAGTACCCAAAGCTCATGG
CAAGGGTTGAACTTCTGCAAAAAAATATAAGTCATTTTATGGGAGAGGATCTGGATGCCTTCAATC
TGCGTGAGTTTCAGGATTTAGAGCAACAGCTCGATACAGCTCTGAAGCGAATACGAACCAGGAA
GAACCAACTGATGCATGAGTCCATTTCTCTGCTGCAGAAAAAGGAAAAAGAACTGCGAGAGCGA
AACAACCTAATTTCCAAGAAGCTTAAAGAAAACGAGAAG----------

Datura metel MBP10

-
CTAAGGAAAACAGGACGCTGGAGTACCAAAAGCTCACGGCAAGGGTTGAACTTCTGCAAAGAA
ATATAAGGCATTTTATGGGAGAGGATCTGGATGCCTTCAATCTGCGAGAGTTTCAGGGTTTAGAGC
AACAGATTAATACAGCTCTGAAGCGAATACGAACCAGGAAGAACCAACTGATGCTTGAGTCCATT
TCCCTGCTGCAGAGAAAGGAAAAAAGACTGCAAGAGCAAAACAACTTAATTTCCAAGAAGCTT
AAAGAATATGAGAAG----------

Atropa belladonna MBP10

CCACTGAGTCCAGCATGGAAAGTATCCTGGAAAGATATGAAAGTTACTCACATGCGGAGAGGAA
GTTGAATGCAAATGACTCTCAAACTTATAAGGAAAATTGGACTCTCGAGTACCCAAAGCTCATGG
CAAGGACTGAACTTCTGCAAAGAAATATAAGGCATTTTATGGGAGAGGATCTGGATTCCTTCAAT
CTGCGAGAGTTCCAGGGTTTAGAACAACAGCTCGATACAGCTCTGAAGCGAATACGAACCAGGA
AGAACCAACTGATACATGAGTCCATTTCCCAGCTGCAGAAAAAGGAAAAAGAGCTGCATGAGCG
AAACCACTTAATTTCCAAAAAGCTGAAAGAAAATGAGAAG----------

--
Lycium barbarum MBP10

AGGAAAATTGGACTCTCGAGTACCCAAAGCTCAGGGCAAGGACTGAACTTCTGCAAAGAAATAT
AAGGCATTTTATGGGAGAGGATCTGGATACCTTCAATCTGCGAGAATTTCAGGGTTTAGAGCAAC
AGCTCGATACAGCTCTCAAGCGAATACGAACCAGGAAGAACCAACTGATGCATGAGTCCATTTCC
CAGCTGCAGAAAAAGGAAAAAGAGCTGCAGGACCGAAACAACTTAATTTCCAAGAAGCTGAAA
GAAAATGAGAAG----------

Dunalia MBP10 ----------
GTTGAGATGAAGCGGATCGAGAACAAAATAAGCAGGCAAGTGACTTTCTCGAAGAGACGATCCG

Continuation of Table 1.8.

GTTTGTTGAAGAAGACTCATGAGATCTCCGTGTTGTGTGATGCTGAAGTGGCATTGATTGTCTTCT
CCTCCAGTGGCAAGCTCTTTGAGTAC---
CCTACTCAATCCAGCATGGAAAGTATCCTGGAAAGGTACGAAAATTACTCATATGCGGAGAGGAA
GTTGAATGCAAATGACACCGAAACTAATAAGGAGAACTGGACGCTCGAGTACCCAAAGCTCATG
GCAAGGGTGGAACTTCTGCAAAGAAATATAAGGCATTNTATGGGAGAGGATCTGGATGCCTTCAA
CCTGCGTGAGTTTCAGAGTTTAGAGCAACAGCTCGATACAGCTCTCAAGCGAATACGAACCAGG
AAGAACCAACTGATGTTCGAGTCCATTTCCCTGCTGCAGAAAAAGGAAAAAGAAATGCAAGAGC
GAAACAACTTAATTTCCAAGAAGCTTAAAGAAAATGAGAAG----------

Nicotiana sylvestris MBP10

CCACTGAATCCAGCATGGAAAGTATCCTCGAAAGATACGAAAGTTACTCATATGCTGAGAGGAAG
TTGAATGCAAATGACTCTGAACCT----------
GAAAAACCAAAGCTCATGTCAAGGATTGAACTTCTACAAAGAAATATAAGGCATTATATGGGAGA
GGATCTGGATTCCTTCTGTCTGCGAGAGTTTCATGGTTTAGAGCAACAACTTGATACAGCTNTGA
AGCGAATACGCGCCAGGAAGAACCAACTGATGCATGAGTCCATTTCCCAG----------

Withania sominfera MBP20 ----------
GCGCTTGAGATGTCAGTTTTCTGTGATGCTGATGTTGCTTTGATTGTTTTCTCTACCAAAGGCAAG
CTCTTTGAGTTCTCT---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATATGAAAGATACTCATATGCAGATAGAAAGATG
AATGCAAATGACATTGATCCC---
AAGGAAAATTGGAATGTGGAGTATCCGAAACTCATGTCAAGGATTGAACTCTTACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTTGACCCTCTCAGTTTGCGAGAGATCCAGAGCTTAGAGCAAC
AGATTGATACTTCATTAAAGAGAATAAGAAGCAGGAAGAACCAGCTGATGCATGAGTCCATCTCT
GAGCTGCAGAAAAAGGAGAAAGCGGTACAAGAACAAAATAACTTGATAACTAAGAAGCTGAAA
GAAAAGGAGAAG----------

Iochroma fuchsiodeas MBP20 ----------
TGCGCTTGAGCTGTCGGTGCTATGCGATGCTGATGTTGCTTTGATTGTTTTCTCTACCAAAGGCAA
GCTCTTTGAGTACTCC---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATATGAAAGATACTCACATGCAGAGAGAAAGAT
GAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGAATGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTTTATGGGTCAGGATCTTGACCCTCTCAGTTTGCGAGAGCTCCAGAGTTTAGAGCAAC
AGATTGATACTTCATTAAAGCGAATAAGAAGCAGGAAGAACCAGCTGATGCATGAATCCATTTCT
GAGATGCAGAAAAAGGAGAAAGCGATGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAAGGAGAAG----------

Nicandra physalodes MBP20

TGAGCTGTCGATGTTGTGCGATGCTGATGTTGCTTTGATTGTTTTTTCTACCAAAGGCAAACTCTT
TGAGTACTCCTCCACTGAATCCAGCATGGAAAGTATTCTGGAAAGATACGAAAGATATTCATATGC
AGAGAGAATGTTGAATGCAAATGACGACGATCCA---
AAGGATAATTGGAGTGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTCTTACAAAGAAATAT
AAGGCATTATATGGGTCAGGATTTGGATCCTCTCAGTTTGCGAGAGCTCCAGAGTTTAGAGCAAC
AGATTGATACTTCATTGAAGCGAATACGAAGCAGAAAGAACCAACTAATGCATGAGTCTATTTCG
GAGCTGCAGAAAAAGGAGAAAACACTGCAAGAACAAAACAGCTTGATAAGTAAAAAGCTGAAA
GAACACAACAAG----------

Dunalia MBP20
ATGGGAAGAGGGAGGGTAGAGTTGAAGCGGATCGAGAACAAGATAAGCAGACAAGTGACATTC
TCAAAGAGACGATCTGGATTGTTGAAGAAAGCTAATGAGATCTCCGTTCTCTGTGATGCTGATGT
TGCTTTGATTGTTTTCTCTACCAAAGGCAAGCTCTTTGAGTACTCC---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATATGAAAGATACTCATATGCAGAGAGAAAGAT

Continuation of Table 1.8.

GAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGAATGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTTTATGGGTCAGGATCTTGACCCTCTCAGTTTGCGAGAGCTCCAGAGTTTAGAGCAAC
AGATTGATACTTCATTAAAGCGAATAAGAAGCAGGAAGAACCAGCTGATGCATGAATCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGATGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAAGGAGAAG----------

Schizanthus grahamii MBP20

TGCGATGCTGAGGTTGCTTTGGTCGTCTTCTCCACTAAAGGAAAGCTCTTTGAGTACTCC---
ACTGACTCCAGAATGGAAAGGATTATGGAAAGATATGAAAGATACTCATATGCTGAGAGAAAGTT
GAATGCAGATGACTCTGAACCC----------
TGGACTCTGGAGTACCCCAAGCTCACGGCAAGGATGGAACTTCTACAAAGAAACATGAGGAATT
ATATGGGTGAGGATCTGGACCCTCTCAGTTTGCGAGAGTTTCAGAGTTTAGAGCAACAACTTGAT
ACGGCTTTGAAACGAATACGAACCAGGAAGAATCAACTGATGCGTGAGTCCATCTCTGAACTGC
AGAAAAAGGAGAAAACGCTGCAAGAACAAAACAACTTTATGACTAAGAAGCTCAAAGAAGATG
AGAAG----------

Streptosolen MBP20
ATGGGAAGGGGTAGGGTTGAGCTGAAGCGGATCGAGAACAAAATAAGCAGGCAAGTGACTTTCT
CGAAAAGGCGTAGCGGATTGTTGAAGAAAGCACATGAGATCTCAGTTCTGTGTGAAGCTGAGGT
TGGTTTGATTGTTTTCTCCACTAAAGGCAAGCTCTTTGAGTACTCC---
ACTGAATCCAGCATGGAAAATATTCTGGAACGATACGAAAGATACTCATATGCAGAAAGGAAGTT
GAATGGAAATGACTCTGATCCC---
AAGGAAAATTGGAGTTTGGAGTACCCGAAGCTTATGTCAAGGGTTGAACTTATACAAAGAAATAT
GAGGCATTATATGGGTCAGGATCTGGACCCTCTCAGTTTGCGGGAGCTGCAGAGTTTGGAGCAAC
AGGTTGATACTGCTTTGAAGCGAATACGCACCAGGAAGAACCAAGTGATGCACGAGTCCATATCT
GAGCTGCAGAAAAAGGAGAAAGCACTGCATGAACAAAACAACCTGATGACTAAGAAGTTGAAC
GAAAAGGAGAAG----------

Goetzia sp. MBP20 ----------
GCGCTTGAAATATCGGTTTTTTGTGATGCTGAAGTTGCTTTGATCGTATTCTCTTCCAAAGGCAAG
CTCTTTGAGTACTCC---
ACTGAATCCAGCATGGAAAGTATTCTGGAGAGATACGAAAGATACTCATATGCTCAGAGAAAGCA
CAATGCTAATGATTCTGATCCC---
GGGGAAAATTGGACCATGGAGTACCCGAAGCTCATGTCAAGGATTGAACTTCTACAAAGAAATAT
AAGGCATTATATGGGTGAGGATCTGGACCCTCTCAGTTTGCGAGAGATTCAGAGTTTAGAGCAAC
AAATTGATACAGCTTTGAAGCGAATACGAAGCAGGAAGAACCAACTTATGCACGAGACCATTTCT
GAACTACAGAAAAAGGAGAAAGCGCTGCAAGAGCAAAACAACTTAATAACCAAGAAGCTGAAA
GAAAAAGAAAAG----------

Mandragora officianarum MBP20

GTCGGTGCTGTGTGATGCTGACGTTGCTTTGATTGTTTTCTCTACCAAAGGCAAGCTCTTTGAGTA
CTCT---
ACTGACTTCAGTATGGAAAGTATTCTGGAAAGGTGCGAAAGATACTCATATGCAGAAAGAAAGTT
GAACGGAAATGAGTCTGATCCC---
AAGGAAAATTGGAGTGTGGAGTATCCCAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAAAAT
AAGGCATTATATGGGTCAGGATCTGGACCCTCTCAGTTTGCGAGAGCTCCAGAGTTTAGAGCAAC
GGATTGATACTTCATTAAAGCGAATAAGAAGCAGGAAGAACCATCTGATGCACGAGTTCATTTCT
GAGTTGCAGAAAAAGGAGAAGGCGCTGCACGAACAAAACAACTTGATAACTAAGAAGCTGATT
GAAAGCGAGAAG----------

Plowmania nyctaginoides MBP20

TGCGCTTGAACTTTCCATGTTTTGTGATGCTGATGTTGCTTTAATTGTTTTCTCAACTAAAGGCAA
GCTATTTGAGTACTCC---
TCTGAGTCCAGTATGGAAAGCATTCTGGAAAGGTATGAAAGATACTCATATGCAGAGAGAAAGGT

Continuation of Table 1.8.

GAATCCCAATGACTCTAATCCC---
CAGGAAAATTGGACATTGGAGTACCCTAAGCTTATGTCAAGGATTGAACTTGTACAAAGAAATAT
AAGGCATTATATGGGTCAGGACCTGGACCCTCTCAGTTTGCGAGAGCTGCAAAATCTAGAGCAAC
AGATTGACACTGCATTGAAGCGAATACGCAGCAGGAAGAATCAACTGATGCACGAGTCCATTTCT
GAGCTGCATAAAAAGGAGAAAGCATTGCAAGAACAAAATAACTTGATGACTAAGA----------
---

Nicotiana obtusifolia MBP20
ATGGGAAGAGGTAGGGTTCAGCTGAAGCGGATCGAGAACAAGATCAGCAGGCAAGTCACCTTCT
CAAAGAGGCGTTCTGGATTGTTGAAGAAAGCAAATGAGATCTCCGTTTTGTGTGATGCTGATGTT
GCCTTGATTGTTTTCTCCACCAAAGGCAAGCTCTTTGAGTACTCA---
TCCGAGTCCAGCATGGAAAGTATTCTGGAAAGATATGAAAGATACTCATATGCAGAGAGAAAGTC
AAATGCAAATGACTCTGATCCC---
ATGGAAAATTGGACTCTGGAGTACCCGAAACTCATGTCAAGGATTGAACTTATACAAAGAAACAT
AAGGCATTATACGGGCCAGGATCTGGACCCTCTTAGTTTGAGAGAGCTACAGAGTTTAGAGCAAC
AGATTGATACAGCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACAAGTCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCACTGCAAGAACAGAACAACTCGATGACTAAGAAGCTGAA
AGACAAAGAGAAG----------

Nicotiana sylvestris MBP20

----------CC---
TCCGAGTCCAGCATGGAAAGTATTCTGGAAAGATACGAAAGGTACTCATATGCAGAGAGAAAGTT
GAATGCCAATGACGTTGATCCC---
ATGGAAAATTGGACTCTGGAGTACCCGAAGCTCATGTCAAGGATTGAACTTATACAAAGAAACAT
AAGGCATTATACGGGCCAGGATCTGGACCCTCTTAGTTTGCGAGAGCTACAGAGTTTAGAGCAAC
AGATGGATACAGCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCCATTTC
TGAGCTGCAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTCGATGACTAAGAAGCTGAA
AGACGAAGAGAAG----------

Solandra maxima MBP20 ----------
GTGTGCTGGAATTCGCCCTTGCGCTTGAGCTGTCGGTGCTGTGTGATGCTGACGTTGCTTTGATTG
TTTTCTCTACCAAAGGCAAGCTCTTTGAGTACTCC---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATACGAAAGATACTCATATGCAGAGAGAAAGAT
GAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGAGTGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAC
AAGGCAATATATGGGTCAGGATCTGGACCCTCTCAGTCTGCGAGATCTGCAGAGTTTAGAGCAAC
TGATTCATACATCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCTATTTCG
GAGCTGCAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGATAACAAAGAAGATGAA
AGAAAACGAGAAG----------

Juanulloa mexicana MBP20

GGTGTTTTGTGATGCTGACGTTGCTTTAATTGTTTTCTCTACCAAAGGCAAGCTCTTTGAGTATTC
C---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATACGAAAGATACTCATATGCAGAGAGAAAGAT
GAATGCAAATGACTCCGATCCG---
AAGGAAAATTGTAGTGTGGAGTATCCGAAGCTCATGTCAAGAATTGAACTTTTACAAAGAAATAC
AAGGCAATATATGGGTCAGGATCTGGACGCTCTCAGTTTGCGAGATCTGGAGAGTTTAGAGCAAC
AAATTGATACATCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCTATTTCG
GAGATGCAGAAGAAAGAGAAAGCGCTGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAACGAGAAG----------

Petunia exserta MBP20 ----------
TGCGCTTGAAATTTCTGTTCTGTGTGATGCTGATGTTGCTTTAATAGTTTTTTCTACCAAAGGCAA
GTTATTTGAGTACTCC---
ACTGAGCCCAGCATGGAAAGTATACTGGAAAGGTACGAAAGATACTCATATGCAGAGAGAAAGC
TGAATGCTAATGACTCTGATCCC---

Continuation of Table 1.8.

AAGGAAAATTGGACACTGGAGTACCCGAAGCTCATGTCAAGAATTGAACTTATACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTGGACCCTCTCAGTTTGCGAGAGCTGCAGAGTTTAGAGCAAC
AAATTGACACAGCATTAAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCCATTTC
TGAGCTGCACAAAAAGGAGAGAGCGCTGCAAGAACAAAATAACTTGATGACTAAGAAGCTGAA
AGAAAATGAGAAG----------

Petunia hybrida MBP20

TCCAGCATGGAAAGTATACTGGAAAGGTACGAAAGATACTCATATGCAGAGAGAAAGCTGAATG
CTAATGACTCTGATCCC---
AAGGAAAATTGGACACTGGAGTACCCGAAGCTCATGTCAAGAATTGAACTTATACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTGGACCCTCTCAGTTTGCGAGAGCTGCAGAGTTTAGAGCAAC
AAATTGACACAGCATTAAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCCATTTC
TGAGCTGCACAAAAAGGAGAGAGCGCTGCAAGAACAAAATAACTTGATGACTAAGAAGCTGAA
AGAAAATGA----------

Brunfelsia MBP20

ATGGGAAGGGGTAGGGTTCAGTTGAAACGAATCGAAAACAAGATCAGCAGGCAAGTCACCTTTT
CCAAGAGGCGCTCAGGATTGTTGAAGAAAGCACATGAGATCTCAGTTTTATGTGATGCTGAGGTT
GCCTTGATCATTTTCTCTACTAAAGGCAAGTTATTTGAGTACTCC---
ACTGAGTCCAGCATGGAAAGTATCCTGGAAAGGTACGAAAGATACTCCTACGCAGAGAGAAGGT
TGAATAGAGATGACTCTGATCCC---
AAGGAAAATTGGACCCTGGAGTACCCGAAGCTCATGTCAAGGATTGAAATTATACAAAGAAATAT
AAGGCATTATACGGGTCAGGATTTGGACCCTCTCAATTTGCGAGAGCTGCAAAGTTTAGAGCAAC
AGATTGATACTGCATTGAAGCGAATAAGAAGCAGGAAGAACCAACTGATGCAGGAGACCATTTC
TGAGCTGCATAAAAAGGAGAAATTTCTGCAAGAGCAAAACAACTTGATGACCAAGAAGCTGAA
AGAAAATGAGAGG----------

Browallia americana MBP20 ----------

GCGCTTGAGATTTCTATCCTTTGCGATGCTGAAGTTGGTTTGATTGTTTTCTCCACTAAAGGCAAG
CTCTTTGAGTACTCC---
ACTGAATCCAGCATGGAAAATATTCTGGAACGATACGAAAGATACTCATATGCAGAAAGGAAGTT
GAATGGAAATGACTCTGATCCC---
AAGGAAAATTGGAGCTTGGAGTACCCAAAGCTTATGTCAAGGGTTGAACTTATACAAAGAAATAT
GAGGCATTATATGGGTCAGGATCTGGACCCTCTCAGTTTGCGGGAGCTGCAGAGTTTGGAGCAAC
AGATTGATACTGCTTTGAAGCGAATACGCACCAGGAAGAATCAAATGATGCACGAGTCCATCTCC
GAGCTGCAGAAAAAGGAGAAAGCACTGCATGAACAAAACAACCTGATGACTAAGAAGTTGAAC
GAAAAGGAGAAG----------

Datura inoxia MBP20 ----------

GCGCTTGARATKTCKGTGTTTTGTGATGCTGACGTTGCTTTGATTGTTTTCTCTACCAAAGGCAAG
CTCTTTGAGTACTCC---
ACTGACTCCAGTATGGAAAGTATTCTGGAAAGATATGAAAGATACTCATGCGCAGAGAGAAAGAT
GAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGAGTGTGGAGTATCCAAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTGGACCCTCTCGGTTTACGAGAGCTGCAGAGTTTAGAGCAAC
AGATTGATACTTCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCTATTTCT
GAGCTGCAGAGAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAACGAGAAG----------

Brugmansia suaveolens MBP20 ----------

GCGCTTGAACTATCCGTTTTTTGTGATGCTGACGTTGCTTTGATTGTTTTCTCTACCAAAGGCAAG
CTATTTGAGTACTCC---
AATGACTCCAGTATGGAAAGTATTCTGGAAAGATACGAAAGATACTCATATGCAGAGAGAAAGAT
GAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGAGTGTGGAGTATCCAATGCTAACGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTGGATCCTCTTAGTTTACGAGAGCTGCAGAGTTTAGAGCAAC

Continuation of Table 1.8.

AGATTGATACTTCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACAAGTCTATTTCG
GAGCTGCAGAAAAAGGAGAAAGCGATGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAACGACAAG----------

Datura metel MBP20

CCTCTTAGTTTACGAGAGCTGCAGAGTTTAGAGCAACAGATTGATACTTCATTGAAGCGAATACG
AAGCAGGAAGAACCAACTGATGCACAAGTCTATTTCGGAGCTGCAGAAAAAGGAGAAAGCGAT
GCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAAGAAAACGACAAG----------

Atropa belladonna MBP20

ATGGGAAGAGGTAGGGTAGAGTTGAAGCGGATAGAGAACAAGATAAGCAGGCAAGTGACTTTCT
CAAAGAGACGATCTGGATTGTTGAAGAAAGCAAATGAGATCTCCGTTTTATGTGATGCTGATGTT
GCTTTGATTGTTTTCTCTACAAAAGGCAAGCTCTTTGAGTACTCT---
ACCGACTCAAGTATGGAAAGCATTCTGGAAAGATACGAAAGATACTCATATGCAGAGAGAAAGC
TGAATGCAAATGACTCTGATCCC---
AAGGAAAATTGGACAGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGAATTATATGGGTCAGGATCTAGACCCTCTCAGTTTGCGAGAGCTCCAGAGTTTAGAGCTCC
AGATTGATACTTCATTGAAACGAATACGAAGCAGGAAGAACCAACTGATGCACGAATCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGATAAGTAAGAAGCTGAAA
GAAAATGACAAG----------

Grabowskia MBP20

ATGCACGAGTCCATTTCTGAGCTACAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGA
TAAGTAAGAAGCTGAAAGAAAAGGAGAAG----------

Lycium sp. MBP20

TCCAGCATGGAAAGTATTCTGGAAAGATACGAAAGATACTCATATGCAGAGAGAGAGTTGAATCC
TAATGACTCTGATCCC---
AAGGAAAATTGGAGTGTGGAGCGTCCGAAGCTAATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTATATGGGTCAGGATCTGGACCCTCTCAATTTGCGGGAGCTCCAAAGTTTAGAACAAC
AGATTGATACATCATTGAAGCGAATACGAACCAGGAAGAACCAACTGATGCACGAGTCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGATAAGTAAGAAGCTGAAA
GAAAACGAGAAG----------

Solanum pimpinellifolium MBP20

ATGGGAAGAGGTAGGGTAGAGTTGAAACGGATCGAGAACAAAATAAGCAGACAAGTAACATTCT
CAAAGAGACGATCTGGATTATTGAAGAAAGCTAATGAGATCTCAGTATTATGTGATGCTGATGTTG
CATTGATTGTGTTTTCTACCAAAGGCAAACTTTTCGAGTATTCCTCAAATGACTCAAGTATGGAAA
GTATTCTTGAAAGATATGAAAGATGCTCATATGCAGAGAGACAGATGAATGCTAATGATTCTGATC
CC---
AAGGAAAATTGGAGTGTGGAGTATCCGAAGCTCATGTCAAGAATTGAACTTTTACAAAGAAATAT
AAGGCATTACATGGGTCAGGATCTGGACCCTCTCAGTTTGCGTGAGCTCCAGAGTATAGAGCAAC
AGATTGACACTTCATTAAAGAGAATTAGAAGCAGGAAGAATCAACTGATGCACGAGTCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGCTCCAAGAACAAAACAACTTGATTACTAAGAAGCTAAAA
GAAAATGAGAAG----------

Jaltomata procumbens MBP20 ----------
TGCGCTTGAGCTGTCGATGCTTTGTGATGCTGATGTTGCTTTGATTGTTTTCTCTACAAAAGGCAA

Continuation of Table 1.8.

GCTCTTTGAGTACTGCTCAACTGACTCCAGTATTGAAAGTATTCAGGAAAGATACGAAAGATGCT
CATTTGCAGAGAGAAAGATGAATGCAAATGACGCTAATCCC---
AAGGAAAATTGGAGTGTGGAGTATCCGAAGCTCATGTCAAGGATTGAACTTTTACAAAGAAATAT
AAGGCATTATATGGGTCAAGATCTGGACCCTCTCAGTTTACGAGAGCTCCGGAGTTTAGAGCAAC
AAATTGATACTTCATTGAAGCGAATACGAAGCAGGAAGAACCAACTGATGCACGAGTCTATTTCG
GAGCTGCAGAAAAAGGAGAAAGCGTTGCAAGACCAAAACAACTTGATGACTAAGAAGCTGAAA
GAAAAGGAGAAG----------

Solanum quitoense MBP20 ----------
TGCGCTTGAACTTTCTGTTTTTTGTGATGCTGATGTTGCTTTGATTATTTTTTCTACTAAAGGAAAG
CTATCTGAGTATGCCTCCACTGACTCCAGTATGGAAAGTATTCTGGAAAGATACGAAAGATACTCA
TATGCAGAGAGAGATATGAACGCAAATGATTCTGATCCC---
AAGGAAAATTGGAGTGTGGAATGTCCGAAGCTCATGTCAAGGATTGAACTTTTACAGAAAAATAT
AACGCATTACATGGGTCATGATCTAGACCCTCTCAGTTTACGTGAGCTCCAGAGTTTAGAGCAAC
AGATTGATACTTCATTAAAGAGAATTAGAAGCAGGAAGAACCAACTGATGCACGAGTCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGCTGCAAGAACAAAACAACTTGATAACTAAGAAGCTGAAA
GAAAATGAGAAG----------

Cestrum diurnum MBP20 ----------
TGCGCTTGAGCTCTCGGTCTTTTGTGATGCTGAAGTTGCTTTGATTGTTTTCTCCACCAAAGGCAA
GGTCTTTGAGTACTCGTCCACTGAATCCAGCATGGAAAGTATTCTGGAAAGATATGAAAGATACT
CATACGCAGAGAAGAAGTTGAACGCCAATGACTCTGATCCC---
AAGGAAAATTGGAGTCTGGAGTGCTCGAAGCTTATGTCAAGGATTGAACTTATACAAAGAAACAT
GAGGCACTACACGGGTCAAGATCTGGATCCCCTCGGTTTGAAAGAGCTGCAGAGCTTAGAGCAG
CAGATTGATACTGCATTGAAGCGAATACGAAGCAGGAAGAACCAAATGATGCACCAGTCCATTTC
TGAGCTCCAGAAAAAGGAGAAAGCGCTGCACGAACAAAACAACCTGATGACTAAGAAGTTGAA
AGAATATGAGAAG----------

Cestrum nocturnum MBP20 ---
GGAAGAGGTAGGGTTCAGTTGAAGCGGATCGAGAACAAGATCAGCAGGCAAGTTACCTTCTCTA
AGAGGCGTTCTGGATTGTTGAAGAAAGCACATGAGATCTCAGTTTTGTGTGATGCTGAAGTTGCT
TTGATTGTTTTCTCCACCAAAGGCAAGCTCTTTGAGTACTCGTCCACTGAATCCAGCATGGAAAG
TATTCTGGAAAGATATGAAAGATACTCATACGCAGAGAAAAATTTGAACGCCAATCACTCTGATCC
C---
AAGGAAAATTGGAGTCTGGAGTACTGGAAGCTTATGTCAAGGATTGAACTTATACAAAGAAACAT
GAGGCACTATACGGGTCAAGATCTGGATCCCCTCGGTTTGAAAGAGCTGCAGAGTTTAGAGCAG
CAGATTGATACTGCATTGAAGCGAATACGAAGCAGGAAGAACCAAATGATGCACCAGTCCATTTC
TGAGCTCCAGAAAAAGGAGAAAGCGCTGCACGAACAAAACAACCTGATGACTAAGAAGTTGAA
AGAAAATGAGAAG----------

Solanum lycopersicum MBP20
ATGGGAAGAGGTAGGGTAGAGTTGAAACGGATCGAGAACAAAATAAGCAGACAAGTAACATTCT
CAAAGAGACGATCTGGATTATTGAAGAAAGCTAATGAGATCTCAGTATTATGTGATGCTGATGTTG
CATTGATTGTGTTTTCTACCAAAGGCAAACTTTTCGAGTATTCCTCAAATGACTCAAGTATGGAAA
GTATTCTTGAAAGATATGAAAGATGCTCATATGCAGAGAGACAGATGAATGCTAATGATTCTGATC
CC---
AAGGAAAATTGGAGTGTGGAGTATCCGAAGCTCATGTCAAGAATTGAACTTTTACAAAGAAATAT
AAGGCATTACATGGGTCAGGATCTGGACCCTCTCAGTTTGCGTGAGCTCCAGAGTATAGAGCAAC
AGATTGATACTTCATTAAAGAGAATTAGAAGCAGGAAGAATCAACTGATGCACGAGTCCATTTCT
GAGCTGCAGAAAAAGGAGAAAGCGCTCCAAGAACAAAACAACTTGATTACTAAGAAGCTAAAA
GAAAATGAGAAG----------

Acknowledgement of Previous Publication and Author Contributions

The material in chapter I was reformatted from the original research article "Evolution and Diversification of *FRUITFULL* Genes in Solanaceae" published in Frontiers in Plant Science on 21 February 2019
(https://doi.org/10.3389/fpls.2019.00043).

The contributing authors (in the published order of authorship) were Dinusha C. Maheepala (DM), Christopher A. Emerling (CE), Alex Rajewski (AR), Jenna Macon (JM), Maya Strahl (MS), Natalia Pabón-Mora (NP-M) and Amy Litt (AL).

AL designed and supervised research, and assisted in writing the paper. DM contributed to the design of the study, generated *Cestrum diurnum*, *C. nocturnum*, and *Schizanthus grahamii* transcriptome libraries, retrieved sequences from PCR-based methods and database mining, analyzed the data, and wrote the paper. CE assisted with PAML analysis, contributed suggestions for analyses, and made suggestions on the paper. AR generated *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis sinuata* transcriptome libraries, contributed suggestions for analyses, contributed in recording the associated protocols, and commented on the paper. JM retrieved sequences from PCR-based methods. MS generated the *Nicotiana obtusifolia* transcriptome libraries and additional sequences using PCR-based methods. NP-M generated *Brunfelsia australis* and *Streptosolen jamesonii* transcriptome libraries, contributed suggestions for analyses, and made suggestions on this paper.

Chapter II:

Comparative transcriptome analyses of fleshy and dry fruit development

Abstract

Although fleshy fruits have evolved from dry fruit on numerous occasions during angiosperm evolution, we do not know the molecular mechanisms that underlie these shifts. In the nightshades, fleshy berries have evolved from dry dehiscent capsules. As part of a larger project that aims to characterize the molecular basis of fleshy and dry fruit development, I generated transcriptomes for five stages of fruit development in both cultivated (*Solanum lycopersicum*) and wild (*S. pimpinellifolium*) tomato, which have fleshy fruit. As cultivated tomato has undergone extensive artificial selection, I analyzed the two transcriptomes for any molecular traits associated with domestication. The results included several gene ontology categories that might be associated with the larger fruit size in domesticated tomato. Co-expression cluster analysis for the expression data identified groups of genes upregulated in just one of the species. However, we could not clearly connect these clusters to developmental processes affected by the domestication process. In addition, I extracted information about the expression dynamics of nine genes that have been shown to play key roles in fruit development, to investigate any interactions between them.

Introduction

Fleshy fruits are of great economic and ecological value, and these plant organs have evolved multiple times from dry fruit during the evolution of angiosperms. Although an association between increased seed mass and evolutionary shift to fleshy fruit has been suggested (Bolmgren and Eriksson, 2010), the molecular mechanisms that underlie these shifts are unknown. In Solanaceae (nightshades), there has been a major shift to fleshy edible fruits in the subfamily Solanoideae from the ancestral dry dehiscent capsules (Knapp, 2002). Fleshy fruits have also independently evolved in the solanaceous genera *Duboisia* and *Cestrum*, in addition to a reversal to dry fruit in *Datura* (see Fig.1, Chapter 1) (Knapp, 2002). These evolutionary events plus the availability of multiple sequenced genomes and the ability to manipulate gene function in this family offer opportunities to understand the mechanisms related to fleshy fruit evolution and development (Albert and Chang, 2014; Tomato Genome Consortium, 2012).

Fleshy fruit have pericarps consisting of multiple layers of cells that expand, change color and accumulate nutrients at ripening. These features help attract frugivores that consume the flesh and in the process, disperse the seeds away from the parent plant. Compared to fleshy fruit, dry fruit have pericarps with a relatively few cell layers, which do not expand, but become woody and dehisce at maturity. Wind, water or the coats of animals may act as passive seed dispersal agents in this case. Four major stages of development have been identified for both fleshy and dry fruits based on anatomical/ physiological data (Gillaspy et al., 1993; Pabón-

Mora and Litt, 2011; Tanksley, 2004). The first stage consists of ovary development; at this early stage no consistent morphological differences between the two fruit types have been described. In stage 2, which immediately follows fertilization, there is increased anticlinal cell division (in the inner and outer epidermal layers in Solanaceae) in dry fruits and an increase in anticlinal as well as periclinal cell divisions throughout the pericarp in fleshy fruit. During stage 3, the pericarp cell walls in dry fruit become lignified, and the pericarp cells of fleshy fruits undergo endoreduplication and cell expansion. In the fourth and the final stage of development, dry fruit become dehydrated, which generates tension between the cell layers in the dehiscence zone, resulting in the splitting open of these fruit and facilitating seed scatter. In contrast, fleshy fruit ripen, which involves a change in color via the accumulation of pigments such as carotenoids and accumulation of sugars and other nutrients, while the cell walls of the pericarp as well as the internal tissue such as placenta degrade (i.e., liquefaction). Our goal is to elucidate the differences in gene expression between dry and fleshy fruit development, from these early stages to maturity. Therefore, as part of a project investigating differences in molecular mechanisms associated with dry and fleshy fruit development, we generated RNAseq libraries for all four stages of the dry-fruited desert-tobacco (*Nicotiana obtusifolia*) and the fleshy tomato (*Solanum lycopersicum*). I have limited the contents of this chapter to my contribution to this project, which was on tomato transcriptome analysis.

A number of studies have used transcriptome analyses to address questions on

tomato development. Most of the previous analyses have focused on ripening (Barry and Giovannoni, 2006; Guo et al., 2012; Karlova et al., 2011; Lee et al., 2012; Pandey et al., 2015; Shinozaki et al., 2018; Tieman et al., 2000; Wang et al., 2016; Ye et al., 2015). As a result, we have a more thorough understanding of phenomena such as ripening associated pigment accumulation and ethylene induction than we do of the processes associated with earlier stages. Pattison et al. (2015) compared gene expression at anthesis and stage 2 using *S. pimpinellifolium*, the closest wild-relative of cultivated tomato, and found an increase in auxin- and stress-related gene expression but a decrease in gene expression related to the induction of ethylene production in fruit at 4 days post anthesis. Zhang et al. (2016) explored patterns of gene expression prior to anthesis in ovules and the ovary/fruit wall and found that genes that function in determining fruit size have specific expression domains; one such gene is expressed only in the pericarp while the other is only expressed in the ovules. This indicates the ovules, as well as the pericarp, have roles in the development of domestication associated traits such as fruit size. In addition, some studies have focused on expression differences between tomato and *S. pimpinellifolium* that might have resulted from domestication. Koenig et al. (2013) investigated sequence and expression divergence related to domestication using RNAseq data from cultivated tomato and five wild tomato species. They found evidence for changes in expression in only a few loci that might be associated with domestication whereas they found changes in many genes that might have been involved in the adaptation of wild tomato species to new environments.

We investigated the dynamics of gene expression over the course of tomato fruit development by generating transcriptomes that represent all four stages. Our expression data for stage four includes RNAseq libraries from both breaker stage, which is the onset of ripening as marked by color change, and red ripe fruit. We also investigated the potential differences between cultivated and wild tomato due to domestication. Therefore, in addition to the transcriptomes of the cultivated tomato (*S. lycopersicum* cv. Ailsa Craig), I generated expression data for the corresponding stages in *S. pimpinellifolium*. We hypothesized that any dissimilarities in gene expression between the cultivated and wild tomato species may be related to artificial selection.

To identify groups of genes that may be expressed differently between cultivated and wild tomato, I extracted co-expression modules specific to each of the two species and also generated hypothetical interacting gene networks using this data. Our early explorations failed to turn up guidelines for these analyses; after these analyses were completed we received information that our sample size is insufficient for reliable results. I am, therefore, interpreting our results in the context of this chapter, but they cannot be considered informative outside of this context.

Materials and Methods

Plant material

I acquired the seeds of *S. lycopersicum* cv. Ailsa Craig (accession #LA2838A) and *S. pimpinellifolium* (accession #LA 2547) from the UC Davis Tomato Genetics Resource Center. I germinated these seeds directly on soil and cultivated the plants

at 26°C in temperature controlled glasshouses at University of California, Riverside (UCR).

Fixing tissue and staining

To confirm the timing of the stages of fruit development in *S. lycopersicum* cv. Ailsa Craig and *S. pimpinellifolium* according to the descriptions of Gillaspy and Tanksley (Gillaspy et al., 1993; Tanksley, 2004), I fixed, sectioned and stained ovaries and fruit harvested at one day pre-anthesis, two, three, five and 15 days post-anthesis.

1. Using the protocol provided by Darleen DeMason (Professor Emerita, Botany and Plant Sciences, UCR), I dehydrated the ovaries and fruits using an ethanol series with ethanol concentrations of 10%, 35%, 50%, 70%, 85%, 95%, 100% for two hours each and 100% overnight.

2. I incrementally replaced the ethanol with Citra-Solv (Citra Solv, LLC, Danbury, CT) by immersing the organs for two hours first in equal volumes Citra-Solv and ethanol, followed by 1 part ethanol to 3 parts Citra-Solv and finally in pure Citra-Solv overnight.

3. I saturated the solvent with paraffin by gradually dissolving Paraplast Plus paraffin chips (Sigma-Aldrich, St. Louis, MO) in the Citra-Solv and afterwards moved the organs twice into clean batches of melted paraffin.

4. I embedded multiple ovaries/fruits in round, flat-bottomed aluminium weighing boats (7cm diameter) containing melted paraffin.

5. After the paraffin solidified, I used a scalpel to cut the block into smaller cubes that contained an individual ovary or fruit.

6. I used a rotary manual microtome with Tissue-Tek Accu-Edge disposable blades (VWR, Radnor, PA) to section the specimens at a thickness of 10 μm.

7. I floated the sections on droplets of water on ProbeOn Plus microscope slides (Fisher Scientific, Hampton, NH) and placed the latter on a slide warmer to dry at 40°C.

8. I removed the paraffin from the sections by dipping the slides in Citra-Solv and then, to rehydrate the specimens, immersed them in an ethanol series of decreasing concentration (3 minutes each in 100%, 95%, 70%, 50%, 25% and pure water).

9. I stained the sections in a 1% aqueous solution of safranin for 20 minutes and briefly rinsed the slides in water, and then dehydrated the sections again using several short immersions in 25%, 50%, 70% and 95% ethanol.

10. I counterstained with 0.5% fast-green in 95% ethanol for 40 seconds and rinsed the slides using several brief immersions in ethanol followed by xylene and let them dry in the airflow of a fume hood.

11. I applied coverslips with Histomount mounting medium (National Diagnostics, Atlanta, GA) , and let the slides dry on a slide warmer at 40°C.

Tissue collection

I harvested one-day-pre-anthesis ovaries (stage 1) from flower buds with some yellowing on the petals, just prior to opening. To harvest three-day (stage 2), and fifteen-day (stage 3) post-anthesis fruit, I tagged flowers at anthesis. I collected breaker stage fruit at the first appearance of yellow or orange patches in the pericarp and mature red-ripe fruit just after the entire pericarp turned red.

I separated the stage 1 ovary wall and stage 2 pericarp tissue from the ovules/seeds with the aid of a stereoscope (Leica M165 MC, Wetzlar, Germany) under 100x magnification. For this, I floated the organs in a sterile dish of deionized water and used a sterile scalpel and a dissecting tenaculum. The remaining stages of fruit were large enough that I was able separate the pericarp tissue with the unaided eye. After dissecting out the ovules/seeds, I kept harvested organs stored at -80°C.

RNA isolation and library preparation

I used Qiagen RNeasy Plant Mini Kits (QIAGEN, Hilden, Germany) to extract RNA from the ovary wall/pericarp tissue according to the manufacturer's protocol. The RNA quality was checked using a Bioanalyzer (Agilent, CA, USA) by the staff at the Institute for Integrative Genome Biology (IIGB) UCR. I stored the RNA at -80°C until further use.

I used an NEBNext Ultra Directional RNA Library Prep Kit for Illumina and Protocol for use with NEBNext Poly(A) mRNA Magnetic Isolation Module protocol (New England BioLabs, MA, USA) for RNAseq library generation according to the manufacturer's protocol.

Sequencing, cleanup of the raw sequencing reads and mapping of the reads to the tomato genome

The libraries were sequenced on an Illumina NextSeq v2 platform with high-output runs of 75bp paired-end reads at IIGB. Then I quality trimmed the raw paired-end reads using TrimGalore (Krueger, 2017). I mapped the libraries for both *S. lycopersicum* and *S. pimpinellifolium* to the tomato (*S. lycopersicum*) reference genome since an annotated genome was not available for *S. pimpinellifolium* (SL3.0/ ITAG 3.2 release; http://solgenomics.net) with Star (Dobin et al., 2013) on the UCR High Performance Computing Cluster (HPCC) using the general settings (twopassMode= Basic; sjdbOverhang= 199; outSAMtype= BAM Unsorted). Following that, I counted the mapped reads using HTSeq (Anders et al., 2015).

Differential gene expression analysis

I analyzed differential gene expression using the DESeq2 package on R (Love et al., 2014; R Core Team, 2018). Genes were considered differentially expressed (DE) if the adjusted p value (false discovery rate) was < 0.01 and the log2foldchange was > 2. I used the the rlog-transformed counts from DESeq2 and the ggplot2 package in R to generate a PCA plot to visualize any patterns of similarities and differences in expression between stages and species (R Core Team, 2018; Wickham, 2016). To generate the heatmaps of DE gene expression, I used pheatmap and RColorBrewer packages on R (Kolde, 2012; Neuwirth and Brewer, 2014; R Core Team, 2018).

Co-expression cluster (WGCNA) analysis

To generate hierarchical co-expression clusters, I analyzed the log-transformed normalized counts using the step-by-step network construction and module detection of the Weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). Initially, I chose two soft threshold powers since a scale independence plot (Fig. 2.45; Table 2.2), which generally contains an individual plateau used to determine the threshold, had two such plateaus. These values included the soft threshold powers of 7, which approaches the scale-free topology fit index value of 0.8, and 15, which approaches the scale-free topology fit index value of 0.9 (Table 2.2). Following that, I used the subsequent scripts in the step-by-step protocol to generate a signed adjacency matrix (type = "signed", corFnc = "bicor", corOptions = "use = 'p', maxPOutliers = 0.1"). I used this matrix to generate co-expression modules using parameters commonly used in other publications (Lin et al., 2018; Liu et al., 2017a; Pei et al., 2017; Takahagi et al., 2018) with 30 being the minimum number of genes per module (minModuleSize = 30), and a module separation threshold (mergeCutHeight) of 0.25 for both soft-threshold values separately. I next compared the modules generated for soft-threshold=7 (Table 2.3) and soft-threshold=15 (Table 2.4) to investigate the similarity between the modules. However, a threshold that provides a stringent output is generally recommended (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html). Compared to soft-threshold=7 where all genes were grouped into one of the clusters, some genes did not get assigned to clusters at the more stringent soft-threshold=15. Although WGCNA groups such

outliers into a “grey” cluster (Table 2.4), this might also include genes misassigned by the program as not co-expressed with any clusters based on the initial set of parameters (Greenfest-Allen et al., 2017; Reinhold et al., 2017). However, as documented in the “powerEstimate” column of Table 2.2a, the pickSoftThreshold function on WGCNA also output 15 as the recommended threshold for our expression data. Therefore, to study the genes contained within the modules more closely in relation to fruit development, I chose the output for soft-threshold=15.

Gene ontology analysis

To extract the categories of biological processes enriched among the DEGs, I analyzed the International Tomato Annotation Group (ITAG) 3.2 IDs on the PANTHER Gene List Analysis webtool (http://www.pantherdb.org) (Ashburner et al., 2000; Mi et al., 2017; The Gene Ontology Consortium, 2017). For clarity in interpreting the resulting GO terms, I chose the subcategories of biological processes with a p value < 0.05 (Bonferroni adjusted). I then visualized these GO categories in relation to the DEG number and fold enrichment using dot plots generated by the ggplot2 package on R (R Core Team, 2018; Wickham, 2016). The R script for dot plots (Bonnot T, et al. unpublished) was created by Titouan Bonnot in the Dawn Nagel lab at UCR.

Gene network analysis

I generated hypothetical interaction networks of genes previously reported to be involved in fruit development using the two different methods below:

1. I exported the gene networks for the co-expression modules generated by WGCNA in Cytoscape format (Shannon et al., 2003) as described on https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/FemaleLiver-06-ExportNetwork.pdf.

   Since I grouped both *S. lycopersicum* and *S. pimpinellifolium* transcriptomes together to identify shared or unique co-expression clusters using WGCNA, the network generated from the output of this analysis was based on the expression data from both species combined. Therefore, to generate hypothetical gene interaction networks specific for each species, I also performed WGCNA analyses for each species separately. For *S. lycopersicum* and *S. pimpinellifolium*, I used soft-thresholds of 12 and 13, respectively, as indicated by the "powerEstimate" columns of Table 2.2b and 2.2c.

   In addition, as part of a larger project elucidating the function of *FRUITFULL* (*FUL*) MADS-box transcription factors in tomato development, I used the two species-specific WGCNA results to generate hypothetical networks between the *FUL* orthologs and their 20 potentially closest interacting partners.

2. I generated an adjacency matrix based on Spearman's correlation, independent of the WGCNA analysis. The script for this step was generated by Ya-

sunori Ichihashi, Jie Peng and Hokuto Nakayama in Dr. Neelima Sinha's (nrsinha@ucdavis.edu) lab at UC Davis.

In each case, I visualized the predicted gene interaction networks with the help of the igraph package in R (Csardi and Nepusz, 2006; R Core Team, 2018).

Results

Differential gene expression analysis reveals intra-specific differences in fleshy fruit development but relatively little inter-specific variation

To identify genes that are differentially expressed within species between consecutive stages and between species at a given stage, I used the DESeq2 package in R (Love et al., 2014; R Core Team, 2018). Following published analyses (Yang et al., 2018) the genes with a p value < 0.01 and a log2foldchange > 2 were considered to be differentially expressed (DE) (Figs. 2.1 and 2.2). The numbers of DE genes are shown in Table 2.1.

In all intraspecific comparisons between two consecutive stages other than stage 1 vs stage 2 in cultivated tomato (*S. lycopersicum* cv. Ailsa Craig; AC), the later stage had more downregulated genes in comparison to the earlier stage (Fig. 2.1 and Table 2.1). In both AC and *S. pimpinellifolium* (Sp), the largest DE gene counts (>1000) are observed for the comparisons between stages 2 and 3, and 3 and breaker. The large number of DEGs between stages 2 and 3 is harder to explain since there is overlap between these two stages. In addition, although having a large number of DEGs is consistent with considerable physiological differences

between stages 3 and breaker (Picton et al., 1993a; Shinozaki et al., 2018), it seems surprising to have more downregulated genes than upregulated ones in stage 4 given the tremendous changes that occur during ripening.

In the comparisons between species at each stage, there are relatively more genes downregulated in AC compared to Sp, with the breaker stage being the exception (Fig. 2.2 and Table 2.1). The overall numbers of DE genes are larger in the comparisons between developmental stages than in the comparisons between species, which reflects the close evolutionary relationship between the two species (Tomato Genome Consortium, 2012).

A principal-component analysis (PCA) plot for the expression data showed that the samples for both species for a given stage clustered together, also suggesting an overall similarity in fruit development between AC and Sp. Stage 1 and 2 samples formed one cluster, and the breaker stage and red ripe samples formed another cluster, with all stage 3 (15 DPA) samples in a single cluster (Fig. 2.3 and 2.4). These clusters lined up along the first principle component (PC1= 56%) according to developmental time. Meanwhile, both clusters containing stages 1 and 2, and breaker and 4, are grouped separately from stage 3 samples along the second principle component (PC2= 13%). This reflects the large numbers of differentially expressed genes in the comparisons between stage 2 vs 3 and stage 3 vs breaker.

The overall gene expression profiles are consistent with the observed developmental traits of fleshy fruit development

To investigate the molecular changes associated with specific stages of fleshy fruit development, I extracted the over-represented gene ontology (GO) categories for biological processes by comparing each of the two consecutive stages within each species separately. The molecular profiles in Sp are comparable to our results for AC. However, only in AC, in a comparison of stage 2 to stage 1, our data show >35 fold enrichment of downregulated genes associated with secondary cell wall biogenesis (Fig. 2.5 and Fig. 2.6). Although the pericarp is mainly made up of parenchyma and collenchyma cells with only a primary cell wall, it does contain vascular bundles, which include xylem and fiber cells that have secondary cell walls (Schaffer and Petreikov, 1997). However, there are no published examples of mechanisms related to secondary cell wall biogenesis acting differently between the two species that might explain this finding. In addition, there is > 9 fold enrichment in AC and > 6 fold enrichment in Sp of upregulated genes associated with stress in stage 2 compared to stage 1 (Fig. 2.7), which has been documented previously (Pattison et al., 2015). In stage 3 compared to stage 2, stress-related genes are downregulated in both species (Fig. 2.8 and Fig. 2.9) (Fig. 2.44). Meanwhile, genes with functions in cell wall processes, sugar metabolism and photosynthesis are upregulated in stage 3 compared to stage 2 in AC (Fig. 2.10). A comparison of the same stages in Sp shows > 7 fold enrichment of upregulated genes associated with cell wall processes and > 10 fold enrichment of upregulated genes in the photosynthetic pathway >, as well as 22 fold enrichment of upregulated genes related

to DNA replication, (Fig. 2.22). This suggests that the cell wall is consistently being adjusted in the processes related to fruit enlargement, facilitated by increased photosynthesis, which might be providing energy as well as wall components as others have documented (Faurobert et al., 2007; Zhang et al., 2016).

Compared to stage 3, the breaker stage shows downregulation of genes related to cell wall and cytoskeletal organization, auxin signaling, and water transport in AC, and cell division/ expansion- and photosynthesis in Sp (Fig. 2.11, 2.12, 2.23 and 2.24). This stage marks the onset of fruit ripening (Picton et al., 1993a; Shinozaki et al., 2018). Since auxin is important for the cell cycle (David et al., 2007), it is possible that the reduction in auxin signaling-related gene expression is associated with the end of the vigorous cell divisions/expansions of stages 2 and 3. In addition, in the breaker stage for both species, consistent with the onset of ripening, our data show ∼20 fold enrichment of upregulated genes related to ripening, which has also been reported by others (Picton et al., 1993a; Shinozaki et al., 2018).

In the breaker stage of AC compared to stage 3, there is >8 fold enrichment of downregulated genes related to water transport (Vandeleur et al., 2009). (Koenig et al., 2013), This change in expression might be due to the relatively thicker cuticle in the breaker stage compared to stage 3, which would reduce water loss, thereby reducing the need to transport water within the fruit. However, we found somewhat contradictory results since, in this same comparison of breaker stage to stage 3, there is also > 7.5 fold enrichment of upregulated genes involved in re-

sponse to water deprivation such as those encoding dehydrins (Liu et al., 2017b), which facilitate stress tolerance, suggesting a water deficit in the fruit (Fig. 2.13). This appears to be inconsistent with the expectation that ripening fruit would be accumulating water as it becomes juicy. One possible explanation is that there was a lag time between harvesting the fruits and storing them in the freezer because of the time associated with separating the pericarp from the seeds. Since the picked fruits are potentially metabolically active, the upregulation of water deficit response genes might conceivably be related to picking the fruit and thereby separating it from its water supply.

Relative to the breaker stage, when the fruit is red ripe there is $>$ 25 fold enrichment of downregulated genes related to photosynthesis and sugar metabolism (Fig. 2.14 and Fig. 2.15). In red ripe fruit, there are only chromoplasts, which contain carotenoids, unlike the stages prior to ripening that contain chloroplasts (Egea et al., 2011). Thus photosynthesis has come to a halt by the red ripe stage. There are only 21 genes that are upregulated in red ripe relative to the breaker stage in AC and none of them can be categorized into any functional group with high confidence (Fig. 2.16). Similarly in Sp, none of the down- (33) or upregulated (37) genes in red ripe compared to breaker stage have been assigned to any currently recognized GO category (Fig. 2.26, 2.27 and 2.28). The annotations of these genes are not associated with functions regarded to be ripening-associated in general. Therefore, it is likely that the ripening-related gene expression, which begins around the breaker stage (Shinozaki et al., 2018) is maintained at a similar

level well into the red ripe stage.

Differences in gene expression between AC and Sp reflect the larger ovary and fruit size of domesticated tomato

The gene expression patterns between AC and Sp are broadly similar, reflecting the overall similarity in molecular developmental mechanisms between the two species. Nonetheless, I searched for any differences in expression profiles between the corresponding stages of the two species that might reveal some of the genetic changes associated with domestication. We found a number of differences that are consistent with the larger fruit size in AC that has resulted from domestication. In AC compared to Sp in stage 1, our data show >8 fold enrichment of downregulated genes involved in hormone metabolism (Fig. 2.29 and Fig. 2.30). These include gibberellin 2-oxidase, which otherwise inactivates gibberellins (Heuvel et al., 2001; Lo et al., 2008; Voegele et al., 2011), and thus, may be related to the larger cell size and greater number of cells in AC. This comparison also shows a $>$ 10 fold enrichment of upregulated genes involved in the cell cycle, which also is consistent with the relatively larger ovary of AC (Fig. 2.31). In addition, in AC stage 1, there is also $\sim$100 fold enrichment of upregulated genes involved in fumarate metabolism. Fumarate is important for photosynthesis (Nunes-Nesi et al., 2007) and its increased metabolism in AC, leading to more carbon assimilation, might also be associated with the larger ovary size.

None of the downregulated genes in stage 2 in AC compared to Sp belonged to any GO category with high confidence (Fig. 2.32 and Fig. 2.33). Similar to stage 1, our

data for stage 2 show ∼50 fold enrichment of upregulated genes (Fig. 2.34) involved in DNA replication and ∼100 fold enrichment of upregulated genes involved in fumarate metabolism, corresponding to more cell division and photosynthesis-related carbon assimilation, respectively, in the larger fruited AC (Nunes-Nesi et al., 2007).

At stage 3, we observed downregulation of genes related to secondary metabolite biosynthesis in AC but not Sp, which may be associated with domestication efforts aimed at achieving a larger fruit through allocating more resources towards development. Secondary metabolites have roles in responses to biotic and abiotic stresses, and resources that may otherwise be used for development are allocated for their biosynthesis (Campos et al., 2016; Chen et al., 2006; Huot et al., 2014). The downregulated genes in the GO category for hormone metabolism encode a castasterone 26-hydroxylase and two gibberellin oxidase-3's. These genes influence development through deactivating brassinosteroids (Ohnishi et al., 2006) and gibberellin (Lo et al., 2008), respectively, and the comparatively low transcription of the related genes in AC corresponds with its larger size. At this stage there is once more ∼100 fold enrichment of upregulated genes with roles in fumarate metabolism (Nunes-Nesi et al., 2007) in AC (Fig. 2.37).

In breaker stage in AC compared to Sp, there is $>$ 10 fold enrichment of upregulated genes involved in photosynthesis, consistent with other data suggesting increased carbon assimilation through photosynthesis, associated with the larger fruit size (Fig. 2.40). In addition, in red ripe AC compared to Sp, there is also ∼100 fold

enrichment of upregulated genes related to fumarate metabolism (Fig. 2.43).

There were several intriguing results for which we could not find a direct connection with domestication. Among these, in breaker stage fruit of AC relative to Sp, there is > 3 fold enrichment of downregulated genes in the GO category for lipid metabolism (Fig. 2.38 and Fig. 2.39). Although this suggests downregulation of lipid metabolism, the processes described in this stage would be expected to involve upregulation. These processes include lipid metabolic processes related to the breakdown of the plasma membrane (Thompson et al., 1987, 1998) and the synthesis of some of the ripening-associated volatiles, which are also derived from lipids (Ties and Barringer, 2012). However, it is unclear whether a potentially decreased lipid metabolism in cultivated tomato has any association with domestication. In addition, at the red ripe stage in Sp compared to AC there is $\sim$30 fold enrichment of upregulated genes with roles in DNA replication, which is indicative of some nuclear activity (Fig. 2.41 and Fig. 2.42). This might be associated with the reported occurrence of endoreduplication well into ripening (Teyssier et al., 2008). However, we did not observe this result at breaker stage and as the cells in Sp are smaller than AC, the significance of our finding is unclear.

The co-expression modules containing a set of genes involved in fruit development are associated with stages consistently with gene function

Genes often function in clusters in a given biosynthetic pathway (Weber et al., 2015). We can predict such clusters by grouping genes based on shared expression patterns. Therefore, to further investigate the molecular mechanisms of fleshy

fruit development, I performed a hierarchical expression cluster analysis using WGCNA (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). However, our results may not be reliable since the number of RNAseq libraries we generated is not enough for this type of analysis.

One of the important steps in WGCNA analysis is choosing a proper soft-threshold value to achieve a scale-free topology for the expression data, where only a minority of genes are considered to be extensively connected (hubs) while the majority of the expressed genes have very few connections (Arita, 2005; Del Genio et al., 2011; Lopes et al., 2014). Initially, I chose the soft-threshold values of 7 and 15, which approached scale-free topology index values of 0.8 and 0.9, respectively (Table 2.2 and Fig. 2.45). These particular thresholds were selected since they approach plateaus above a scale-free topology index of 0.8 on a scale-independence plot (Fig. 2.45) as generally recommended (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). I performed a separate WGCNA analysis at each of these two soft-threshold values. This included generating adjacency matrices that indicate how connected a given gene is to another, followed by grouping the genes into co-expression clusters based on similarity in expression (i.e. adjacency) (Tables 2.2 and 2.3). At both soft-threshold values, there are more co-expression modules that are upregulated than those that are downregulated ($p < 0.05$). The up- or downregulation of co-expression modules depends on the synchronous up- or downregulation of their constituent genes (Shahan et al., 2018). Thus, in our expression data, although the genes in most modules are simultaneously upreg-

ulated, their downregulation is not as synchronous.

The soft-threshold of 7 is less stringent and therefore, all expressed genes are grouped into modules, whereas for the soft-threshold of 15, there is a "grey" module (Table 2.4) that contains genes that did not group into any module with high confidence (Tables 2.2 and 2.3). To explore the individual genes in the co-expression modules in association with different developmental stages and species, I chose the more stringent soft-threshold value of 15 as recommended by WGCNA protocols (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). In addition, soft-threshold=15 is the value recommended by the WGCNA pickSoftThreshold function (Fig. 2.45 and Table 2.2), which further supported our choice.

In WGCNA, genes are grouped into modules based on similar expression patterns (Langfelder and Horvath, 2008). In turn, a co-expression module is considered to be associated with a given stage if there is a statistically significant correlation between that stage and the module eigengene, which is the magnitude of up- or down-regulation of the constituent genes. Such an association between a module and a stage indicates that the genes in that cluster might act in a specific biological pathway during that stage. Our results included 23 co-expression modules (Table 2.4) that are associated with one or more fruit developmental stages with high statistical confidence ($p < 0.05$).

Our preliminary analyses based on grouping both species together show that in both the cultivated and wild tomato species, the "red" co-expression module (1432 genes) is upregulated in stages 3, breaker and red ripe (Table 2.4), although the

association between module and stage is not statistically significant. This module includes two genes, *Colourless non-ripening* (*Cnr*) and *TOMATO-AGAMOUS-LIKE1* (*TAGL1*) that function in ripening (Chen et al., 2015; Manning et al., 2006; Vrebalov et al., 2009). *Cnr* encodes a SQUAMOSA promoter binding protein-like transcription factor and is considered a major hub for other known fruit-ripening related genes (Chen et al., 2015; Manning et al., 2006). Studies have implicated *Cnr* in fruit ripening associated changes in pigmentation and cell adhesion (Chen et al., 2015; Manning et al., 2006). In addition, *Cnr* is thought to act upstream of all ripening-related genes discussed in this chapter (Bemer et al., 2012; Chen et al., 2015; Karlova et al., 2011). *TAGL1* is an ortholog of *SHATTERPROOF1/2* (*SHP1/2*) in *Arabidopsis thaliana* and encodes a MADS-box transcription factor (Colombo et al., 2010; Vrebalov et al., 2009). In comparison to *SHP1/2* which are involved in the differentiation of the dehiscence zone and the lignification of the valve margins in the dry silique of *A. thaliana*, *TAGL1* has roles in fleshy fruit expansion as well as ethylene-induced ripening via the induction of *ACC Synthase 2* in tomato (Vrebalov et al., 2009). The known functions of *Cnr* and *TAGL1* are consistent with upregulation during the later stages of fruit development. The "red" module also contained 10 lyases including pectin-lyases, which might be involved in cell wall expansion and breakdown during cell expansion and ripening.

There are no co-expression modules that are upregulated in both species in stage 1. In stage 2, the "brown" module (3366 genes) is upregulated in both AC and Sp (Table 2.4). This module includes *MADS-box Protein 10* (*MBP10*) and *MADS-*

box Protein 20 (*MBP20*), which are paralogs of tomato *FRUITFULL1* (*SlFUL1*) and *FRUITFULL2* (*SlFUL2*) (Hileman et al., 2006; Litt and Irish, 2003). A function for *MBP10* and *MBP20* in fleshy fruit development has not been established, although the latter gene has been implicated in tomato leaf development (Burko et al., 2013). Despite having low expression levels throughout fruit development compared to *SlFUL1* and *SlFUL2*, the highest expression of *MBP10* and *MBP20* is observed during stages 1 and 2 (fig. 2.46). This might suggest some function for *MBP10/20* in early fruit development. In addition, there are 8 jasmonic- and 4 salicylic-acid related genes in the "brown" module. This is in line with the upregulation of stress-related genes in stage 2 reported in this study and by Pattison et al. (2015). The module also contained 18 auxin- and 7 cytokinin-related genes, and 8 *SUN-like* genes, which may be involved in the extensive cell divisions and the development of fruit shape during stage 2.

The co-expression modules associated with stage 3 or the breaker stage in both species are all upregulated; no downregulated modules are associated with these stages. The "black" module (1336 genes) is upregulated during stage 3 in both species (Table 2.4). This module includes *SlFUL2*, which has been previously reported to play a role in ripening (Bemer et al., 2012; Shima et al., 2013; Wang et al., 2014, 2019). In addition, *SlFUL2* has potential roles in cuticle and pericarp development prior to ripening (Wang et al., 2019). In our data, the inclusion of *SlFUL2* in a co-expression module that is upregulated during stage 3 supports its putative roles prior to ripening (Fig. 2.46). However, although *SlFUL2* has roles in

ripening, the co-expression module it belongs to is only upregulated during stage 3, which is also when this gene is at its peak expression, and is downregulated in red ripe. In addition, the "black" module included 5 expansins, 3 gibberellin-related genes and 8 *SUN-like* genes, all of which have roles in cell wall modification associated with cell enlargement, a dominant process during stage 3 (Huang et al., 2013; Jiang et al., 2009; Lo et al., 2008; Lu et al., 2016).

In the red ripe stage, the "blue" module (4028 genes) is upregulated in both species (Table 2.4). This module included *SlFUL1*, *NONRIPENING* (*NOR*), which encodes a NAC-domain transcription factor, *RIPENING-INHIBITOR* (*RIN*), which encodes a MADS-box transcription factor, *NEVER-RIPE* (*NR*/*ETR3*), which encodes an ethylene receptor (ETR) family protein and *APETALA2a* (*AP2a*), which encodes a protein that belongs to the ethylene responsive factor (ERF) family. Wang et al. (2019) reported that *SlFUL1*, similar to *SlFUL2*, functions in ripening. *NOR* and *RIN*, *NR* and *AP2a* also function in ripening (Cantu et al., 2009; Chung et al., 2010; Hackett et al., 2000; Ito et al., 2015; Karlova et al., 2011; Ma et al., 2018; Osorio et al., 2011). Along with the upregulation of the module containing *Cnr* and *TALG1* at this stage, these data support the empirical evidence for *Cnr*, *TAGL1*, *RIN*, *NOR*, *NR*, *AP2a* and *SlFUL1* in fruit ripening. The "blue" module also contained 20 ethylene-response genes that may be involved in the climacteric burst in ethylene production (Alexander and Grierson, 2002; Su et al., 2015), and five carotenoid biosynthetic genes that might have a role in ripening-induced pigmentation (Su et al., 2015).

Each module contained various numbers of methyltransferases which selectively methylate DNA, leading to transcriptional silencing (Bewick and Schmitz, 2017; Gallusci et al., 2016; Matzke and Mosher, 2014; Zhu et al., 2015) and may be involved in various biological processes associated with each stage. The "pink" module (5495 genes), which is downregulated in both species at red ripe (Table 2.4), contained 93 methyltransferases. Since this module is also upregulated during the immature stages (Table 2.4), although not at a significant p value, the set of methyltransferases contained within this module might have functions in preventing premature ripening (Gallusci et al., 2016; Zhong et al., 2013). In addition, the "pink" module contained the *fruit weight 2.2* (*fw2.2*), a QTL locus, which has a role in increased fruit size in domesticated tomato (Frary et al., 2000). The only gene that has been assigned to *fw2.2* locus, *ORFX*, influences fruit size through negative regulation of cell division (Liu, 2003). This gene functions prior to anthesis and is downregulated later consistent with the pattern seen in the "pink" module. In addition, there are also 72 chlorophyll-related genes among the genes downregulated in this module, consistent with the end of photosynthetic carbon assimilation at the red ripe stage.

Co-expression clusters associated with domestication

Next, I analyzed the WGCNA modules associated with only one of the two tomato species as these might indicate the suites of genes that differ in expression due to domestication. Our analyses show that of the 23 co-expression modules, 18 are associated with only one of the species with high confidence ($p < 0.05$; Table

2.4, 2.6 and Fig. 2.51). However, none of the GO categories for these modules appeared to be related to domestication traits such as increase in fruit size, prolonged shelf-life, etc. (Table 2.6 and 2.7). In addition, we investigated the *SUN-like*, auxin-related and methyltransferase genes in our co-expression modules as these genes have functions in determining fruit shape, cell division and regulating gene expression, respectively (Chuikov et al., 2004; Jiang et al., 2009; Mambro et al., 2017). However, the modules specific to each species included a number of these genes (Table 2.7) and these seemed to have no clear association with domestication.

Networks of genes implicated in fruit development

Since *in silico* generated top-down networks are based on indirect data, they may not represent the true *in vivo* gene interactions and therefore, less reliable compared to bottom-up networks generated using gene knockouts (Gaiteri et al., 2015; Pezzulo and Levin, 2016; Rodriguez-Zas et al., 2008). However, top-down networks provide hypotheses for future work that may test these interactions. I generated hypothetical interaction networks among 9 genes that have been implicated in fruit development (*Cnr*, *RIN*, *NR*, *NOR*, *AP2a*, *TAGL1*, *fw2.2*, *SlFUL1*, and *SlFUL2*) and two additional *FUL* genes (*MBP10* and *MBP20*), using the WGCNA adjacency matrix calculated for the expression data based on biweight midcorrelation as well as an independent method from Sinha Lab (College of Biological Sciences, UC Davis) to create an adjacency matrix based on Pearson correlation (see Materials and Methods; Fig. 2.47 and Fig. 2.48). Although we subsequently learned we did not

have enough samples for the results of this analysis to be reliable, these analyses proved an informative exercise.

In the WGCNA-generated gene networks for both species, *Cnr*, *TAGL1*, *SlFUL1*, *RIN*, *NR*, *AP2a* and *NOR* ("core ripening cluster") are connected to each other (Fig. 2.47a and 2.47c). In AC, *SlFUL2*, *MBP10*, *MBP20* and *fw2.2* formed a network cluster separate from these core-ripening genes (Fig. 2.47a). However, in Sp, while *MBP10*, *MBP20* and *fw2.2* formed a separate cluster as in AC, *SpFUL2* joined directly with the core ripening cluster (Fig. 2.47c). To investigate how the separate network clusters were connected in each species, I queried the program for genes that formed these connections. In AC (Fig. 2.47b), both *MBP10* and *MBP20* are connected to *SlFUL1* through *Solyc03g083500.3*, which encodes a protein of unknown function (DUF1442). *In silico* motif annotation has suggested some methyltransferase activity for this protein in *Medicago truncatula* (https://www.uniprot.org/uniprot/G7JLB7). In addition, *SlFUL2* and *fw2.2* connected with *Cnr* through *Solyc02g091900.3*, which encodes a cysteine desulfurase-like protein. *MBP10* and *MBP20* connected to *SpFUL1* via *Solyc02g093880.3*, which encodes a GTE8 transcription factor. However, the potential interactions represented by these genes as intermediates between ripening-related transcription factors are unclear. For instance, while some transcription factors might directly activate methyltransferases (Kryczek et al., 2014), there is no evidence that transcription factors interact with cysteine desulfurases or GTE8s. Thus it is possible that at least some these intermediate genes were an artifact due to the low number of samples we have and/

or limiting the analysis to specific genes, which might have eliminated the likeliest intermediate pathway genes.

The networks generated using the Sinha Lab method, which creates an adjacency matrix independently of the WGCNA program, connected *Cnr*, *TAGL1*, *NOR*, *RIN*, *NR*, *AP2a*, *SlFUL1*, *SlFUL2*, *MBP10* and *MBP20* without any intermediates in both species (Fig. 2.48a and 2.48b). In AC (Fig. 2.48a), *TAGL1* is only connected to *AP2a* while in Sp, it is connected to *NR*, *RIN* and *SpFUL2* in addition to *AP2a*. However, evidence suggests *TAGL1* might directly interact with *Cnr*, *RIN* and *SlFUL1/2* in cultivated tomato (Bemer et al. 2012), which seems to be in agreement with the WGCNA based network for AC in comparison to the second method (Fig. 2.47).

As we are currently working on elucidating the function of all four *FUL* genes in fruit development, I used our WGCNA output to create hypothetical gene networks centered on each gene in both species (Fig. 2.49). For this, I extracted the 10 genes that each *FUL* gene might have the strongest connection with. Most of the hypothetical interacting genes had been annotated with a general functional category (eg. CBS domain-containing protein-like, GTP-binding family protein, transmembrane 9 superfamily member, receptor-like kinase) identified by sequence similarity. These have broad functions in redox homeostasis, cellular signal transduction, cytoskeletal organization and immunity (Bertoni, 2011; Goff and Ramonell, 2007; Wu and Zhou, 2013; Ye et al., 2017). However, we also found a few genes with reported functions in fruit development among the predicted interactors of *Sl/SpFUL1* and *Sl/SpFUL2*. Our data suggest that in AC, *SlFUL1* is connected

to *Solyc10g080900.2* (*ripening related X72730*) (Fig. 2.49a), which is also called *ETHYLENE-RELATED10* (*ERT10*) (Giovannoni et al., 1999; Picton et al., 1993b). *SlFUL1* is reported by Wang et al. (2019) to upregulate ripening-related ethylene production. Therefore, *ERT10* may be a candidate gene that might act downstream of *SlFUL1* in ethylene biosynthesis. *SpFUL1* is predicted to be connected to *TERPENE-SYNTHASE18* (*TPS18*) (Fig. 2.49b). The latter is involved in the biosynthesis of defense compounds and is expressed in the immature fruit but with no or low level of transcription in the ripe fruit (Falara et al., 2011).

Our network analysis suggests *SpFUL2* might be connected to *Solyc11g008820.2*, which encodes an endoglucanase, and *Solyc05g010180.3*, which encodes a carotenoid isomerase (Fig. 2.49d). Endoglucanases might be involved in the cleaving of cell wall components (Qin et al., 2003). These enzymes are synthesized during processes such as ripening (Marín-Rodríguez et al., 2002). It is possible that the gene encoding the carotenoid isomerase (Isaacson et al., 2002) might be upregulated by *SlFUL2*, which has confirmed roles in the change in fruit pigmentation (Bemer et al. 2012) during ripening.

AC *MBP10* is shown to be potentially connected to *Solyc03g044085.1*, encoding a glucan endo-13-beta-glucosidase and, *Solyc02g068600.3*, encoding an ankyrin repeat-containing protein (Fig. 2.49e). It has been suggested that glucan endo-13-beta-glucosidases have roles in cell wall modification (Muñoz-Espinoza et al., 2016) and that ankyrin repeat-containing proteins have roles in fruit development in tomato (Yuan et al., 2013). *SpMBP10* is predicted to be connected to

Solyc02g093950.3, which encodes another ankyrin repeat protein (Fig. 2.49g). However, as no data exist on the function of *MBP10*, the functional implications of these predicted interactions are not clear.

Similarly, no study thus far has implicated *MBP20* in fruit development. Our data for AC *MBP20* (Fig. 2.49h) predict it might be connected to *Solyc10g084150.2* encoding a cytokinin riboside 5'-monophosphate phosphoribohydrolase, which biologically activates cytokinins. AC *MBP20* is also predicted to be connected to *Solyc02g068600.3* (a gene that is predicted to be connected to AC *MBP10*), which encodes an ankyrin protein. In addition, Sp *MBP20* is predicted (Fig. 2.49i) to be connected to *Solyc07g061740.3*, encoding yet another ankyrin protein. There is evidence to suggest that transcription factors may directly interact with ankyrin repeats (Wilson-Rawls et al., 1999). Thus, our data suggest direct interactions between *FUL* and ankyrin repeats, which according to Yuan et al. (2013) might have roles in fruit development.

Discussion

Due to the insufficient number of samples we employed in the comparative WGCNA analysis of co-expression clusters between cultivated and wild tomato, the results reported here are only intended as preliminary data.

Patterns of gene expression set stage 3 apart from other stages

Our data revealed that the patterns of gene expression in stage 3 are considerably different from all other stages in both AC and Sp (Fig. 2.3). There are increased

numbers of DEGs (> 1000) that are up- and downregulated in stage 3 relative to stage 2 as well as breaker stage, which is not the case for comparisons between stages 1 and 2, and breaker and red ripe (Fig. 2.1). Since stages 1 and 2 involve growth, cell division may be the dominant feature in both these stages, which might account for the similarities in gene expression we observed between them. Likewise, ripening begins at breaker and continues into red ripe, thus we expect, and find, highly similar gene expression at those two stages. We do not, necessarily, expect significant differences between stages 2 and 3, as these stages overlap. Some regions in the pericarp are still undergoing cell division characteristic of stage 2 while simultaneously in other regions cells are expanding, a feature of stage 3. Furthermore, both stages 2 and 3 involve nuclear division. Therefore, we expect significant overlap in gene expression between stages 2 and 3. However, our data indicate that the molecular mechanisms that occur during stage 3 are markedly distinct from stage 2. In contrast, stage 4 has typically been thought of as distinct from stage 3, in that endoreduplication has been described as ceasing prior to the onset of the major physiological changes associated with ripening (Tanksley reference). This is consistent with our data, which show strongly significant changes in gene expression from stage 3 to stage 4. However, evidence indicates that endoreduplication continues after ripening has been initiated (Teyssier, Cheniclet references), suggesting overlap of these stages as well. Nonetheless, the large number of genes that are associated with ripening appears to be enough to strongly differentiate stage 4 from stage 3.

One explanation for the difference between stages 2 and 3 might be a global increase in gene expression due to endoreduplication (Bourdon et al., 2012), which transforms the tomato genome from diploid to polyploid status. However, the conventional methods of DEG analyses such as the one I employed are not sensitive to changes in ploidy and only regard a very small number of genes as differentially expressed (Pirrello et al., 2018). Therefore, the expression patterns we observe for stage 3 appear not to be a direct result of change in gene copy number. Although there is some overlap between stages 2 and 3, it is possible that our sampling included timepoints early enough during stage 2 and late enough during stage 3 to avoid the overlap and to give relatively clean separation between the stages. In comparison, stages 3 and 4 do not overlap to the same extent; although endoreduplication appears to continue into ripening (Teyssier, Cheniclet citations), a large number of novel processes related to ripening are initiated at stage 4, consistent with the large numbers of DEGs we observed.

Gene expression patterns shared by wild and cultivated tomato

Our expression data is largely congruent with expectations based on the developmental processes observed in tomato. Anatomical data for both wild and cultivated tomato show increased cell division in stage 2 followed by cell expansion and endoreduplication during stage 3 (Chevalier et al., 2011, 2014; Czerednik et al., 2015; Gillaspy et al., 1993; Tanksley, 2004). In accordance with this, our data show genes that have functions in the GO category for cell wall organization are upregulated in these stages while the genes with functions in the GO category

for cell differentiation are downregulated. A comparison with previous studies of fruit development at these stages indicates these studies did not report on the expression of these genes, but our results are consistent with known processes occurring at that time. In addition, photosynthesis related genes are upregulated in stage 3, which is also reported by Zhang et al. (2016). As previously reported by Pattison et al. (2015) for Sp, we detected upregulation of stress related genes in stage 2 in AC and Sp. Pattison et al. suggest this increase might be due to the location of the pericarp at the boundary between the external and internal environments. Compared to the ovary in the prior stage, during which it is surrounded by the calyx, the stage 2 ovary may not be as protected from the external environment. Similarly, the stage 2 fruit lacks a thick waxy cuticle, which aids in defense. Therefore, chemical defense might be the main means of protection available during stage 2. In agreement with Shinozaki et al. (2018), who reported that the molecular processes associated with ripening are initiated in advance of ripening, our RNAseq data show that ripening-related genes are upregulated in breaker stage in both species.

Our co-expression analysis indicate that several important fruit ripening-related genes (*Cnr*, *TAGL1*, *NR*, *NOR*, *RIN*, *AP2a*, *SlFUL1* and *SlFUL2*) are upregulated at the same stage in both AC and Sp. In both species, the module that contained *Cnr*, which acts upstream of genes with roles in ripening (Chen et al., 2015; Manning et al., 2006), is not upregulated in stages 1 and 2, but is relatively upregulated in stage 3 and stage 4 (breaker and red ripe in our analysis) (Table 2.4). This is in

agreement with the finding that the *Cnr* promoter is hypermethylated in immature fruit while becoming demethylated towards the breaker stage, leading to its maximum expression level (Ecker, 2013; Manning et al., 2006; Zhong et al., 2013). The modules with *TAGL1*, *SlFUL1*, *RIN*, *NR*, *AP2a* and *NOR*, all of which act downstream of *Cnr*, are also upregulated in stage 4 in accordance with their reported functions in later stages. *TAGL1* has roles in fruit expansion and ethylene-induced ripening, and its ectopic expression results in fleshy sepals that accumulate lycopene (Vrebalov et al., 2009). *NR* encodes an ethylene receptor important for ripening (Hackett et al., 2000; Ito et al., 2015; Vrebalov et al., 2002) while *NOR* and *RIN* also function in ripening (Ito et al., 2015; Vrebalov et al., 2002). However, new data (Ito et al., 2017) have revealed that *RIN* is not essential for ripening but is required for normal levels of ethylene biosynthesis and lycopene accumulation. In addition, *AP2a* represses ethylene production while simultaneously inducing carotenoid biosynthesis (Chung et al., 2010).

Molecular changes associated with tomato domestication

Endoreduplication has been described as being limited to stage 3 (Tanksley, 2004). While there is some overlap between stages 2 and 3, previous descriptions have suggested that endoreduplication associated with stage 3 stops at the onset of stage 4 (Tanksley, 2004). However, studies (Cheniclet et al., 2005; Teyssier et al., 2008) have reported that endoreduplication continues into the ripening stage. Our data for stage 4 show that the GO category for DNA-replication, which may be associated with endoreduplication, is only enriched in Sp. However, this is

inconsistent with studies by Cheniclet et al. (2005) and Teyssier et al. (2008), which documented a continuous increase in genome size into the ripening stage due to endoreduplication in AC as well as Sp. It is also possible that the the genes involved in endoreduplication are not expressed at high enough levels to register as relatively upregulated.

Our analyses suggest little overall gene expression difference between AC and Sp despite extensive artificial selection efforts. Although we found co-expression clusters upregulated in just one of the two species (Table 2.6), the GO categories for these gene clusters are not associated with any described characteristics that differ as a result of domestication. However, the genes involved in domestication might be few in number, as suggested by Koenig et al. (2013), and methods such as GO category extraction that group multiple genes together might not be sensitive enough to detect these few genes.

Reports have shown that epigenetic factors affect fruit size (Liu et al., 2011; Tilly et al., 1998; van der Knaap et al., 2014) and shape (Sauvage et al., 2017), which are traits initiated in early fruit development (Frary et al., 2000; Ku et al., 2000; van der Knaap and Tanksley, 2001) and that have also been the focus of domestication. Some methyltransferases have epigenetic functions in DNA or histone methylation (Hayashi et al., 2005; Lyko and Brown, 2005). Our preliminary hierarchical cluster analysis on the genes associated with specific stages revealed that the modules upregulated in stage 1 in AC contained 104 methyltransferases (Bewick and Schmitz, 2017; Gallusci et al., 2016; Matzke and Mosher, 2014; Zhu et al., 2015). However,

the corresponding stage in Sp had only 6 methyltransferases. Because I used the annotated tomato genome to map both AC and Sp transcripts, it is possible that the difference in the number of methyltransferases for the two species is an artifact of sequence variation between the annotated genome of cultivated tomato and Sp. However, this is unlikely since the overall numbers of differentially expressed genes between the corresponding stages between AC and Sp (Table 2.1) including the number of methyltransferases for the remaining stages, are comparable. Thus, the explanation for this difference is unclear.

The expression of *FUL* genes and their hypothetical genetic interactors

As part of a larger project that investigates the function of *FUL* genes in fruit development, I searched for any patterns in our expression data that might provide hypotheses on the functions of the four *FUL* copies in AC and their orthologs in Sp. Our expression data show that *SlFUL1* and *SlFUL2* have reverse expression patterns in stages 3 and 4; *SlFUL2* expression is the highest in stage 3 when *SlFUL1* expression is at the lowest, while the opposite is true for breaker and red ripe stages (Fig. 2.46). Wang et al. (2019) document in *slful2*, but not in *slful1*, light colored superficial stripes that develop at the distal end of early green fruit. This coincides with our expression data which show elevated expression of *SlFUL2* in stage 3 (Fig. 2.46), suggesting a function in fruit development prior to ripening.

Compared to other *FUL* paralogs (*MBP10/20*), *SlFUL1/2* expression is relatively high during all stages of fruit development (Fig. 2.46). *SlFUL1* and *SlFUL2* also have high sequence similarity (80%) and both RNAi downregulation (Bemer et al.,

2012) and CRISPR/Cas9 knockout (Wang et al., 2019) studies have implicated them in fruit ripening-related processes. Wang et al. (2019) report that *slful1* mutants, but not *slful2*, produce significantly decreased levels of ethylene compared to wildtype, which refines previous findings that suggested ripening-associated ethylene induction by *FUL* orthologs (Shima et a., 2014; Wang et a., 2014). Consistent with these findings, our hypothetical gene networks suggest that *SlFUL1* might be connected with *ERT10*, a gene involved in ethylene induction (Giovannoni et al., 1999; Picton et al., 1993b) while the double mutant pericarp remained orange even in stage 4, the single mutants turned red similar to wildtype fruit. Therefore, both *SlFUL1* and *SlFUL2* seem to have redundant roles in ripening-associated carotenoid pigment accumulation, which is also suggested by Bemer et al. (2012). Consistent with this, our hypothetical networks suggest *SpFUL2* might be closely connected to *Solyc05g010180.3*, a gene that encodes a carotenoid isomerase. However, *Solyc05g010180.3* is not a component of our hypothetical network for AC, while *ERT10* is not present in the network for Sp. These differences cast doubt on the results and suggest they may be due to the noise expected in constructing networks with too few samples.

Gene networks reveal molecular interactions of fleshy fruit development in cultivated and wild tomato

I generated top-down interaction networks for *Cnr*, *TAGL1*, *RIN*, *NR*, *AP2a*, *NOR*, *SlFUL1*, *SlFUL2*, *MBP10* and *MBP20* using our expression data. Some data regarding the interactions of these genes or their protein products already exist. *Cnr* is

considered a major hub as it is highly connected, either directly or indirectly, with many genes with known roles in fruit ripening (Bemer et al., 2012; Chen et al., 2015; Karlova et al., 2011). Evidence suggests it upregulates the expression of *SlFUL1* and *SlFUL2* (Bemer et al., 2012). *Cnr* is required for the binding of *RIN* to its targets (Bemer et al., 2012), which suggests protein complex formation by the products of these two genes (Martel et al., 2011). *Cnr*, *RIN* and *NOR* upregulate the expression of *AP2a*, which in turn downregulates *Cnr* expression in the pericarp (Karlova et al., 2011). Yeast two-hybrid assays have provided plausible evidence that RIN might form tetramers with TAGL1, SlFUL1 and SlFUL2 *in vivo* (Fujisawa et al., 2014; Leseberg et al., 2008; Martel et al., 2011; Shima et al., 2013). There is evidence to suggest that *TAGL1* might also form heterodimers with *SlFUL1/2* (Wang et al., 2014). In addition, it has been hypothesized that *NOR* might act upstream of *NR*, which encodes an ethylene receptor, while both these genes might function upstream of *SlFUL1* in the ripening-associated ethylene response (Osorio et al., 2011). Fujisawa et al. (2014), using chromatin immunoprecipitation coupled with microarray analysis (ChIP-chip), showed that *SlFUL1*, *SlFUL2* and *RIN* bind to the promoter regions of *AP2a*, *RIN*, *NOR* and *Cnr*. However, we still lack a complete understanding of how all these genes interact during fruit development. In addition, although no functional data exists for *MBP10* and *MBP20* with regard to fruit development, they are expressed in the fruit (Fig. 2.46). Therefore, our preliminary networks involving these genes might provide hypotheses for an improved understanding of their interactions in fruit development.

In hierarchical networks such as the ones I generated, each expressed gene is connected to all others. Thus, it is important to choose an appropriate threshold to remove the unlikeliest of interactions. I chose 0.02 as the threshold when generating networks (Fig. 2.47) using WGCNA, since this is the standard used in publications (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). In the networks for AC and Sp, *Cnr*, *TAGL1*, *SlFUL1*, *RIN*, *NOR*, *NR* and *AP2a* are connected (Fig. 2.47), suggesting interaction among these genes important for ripening. However, while *SpFUL2* connected with *Cnr* directly, in the AC network, *SlFUL2* connected with *Cnr* via *Solyc02g091900.3*, which encodes a cysteine desulfurase-like protein. The latter has a potential role in reducing chlorophyll accumulation (Ahn et al., 2005; Du et al., 2016; Fu et al., 2016). Therefore, it is possible that *Solyc02g091900.3* might function in the ripening-associated change in pigmentation, a role attributed to both *Cnr* and *SlFUL2* (Bemer et al., 2012; Fraser et al., 2001; Wang et al., 2019). In AC, *fw2.2* also connected to *Cnr* through *Solyc02g091900.3*. However, in Sp, *fw2.2* connected to *Cnr* via *Solyc01g006000.3*, encoding a glycosylphosphatidylinositol (GPI) mannosyltransferase (Greene et al., 2015), a group of these that have roles in cell wall formation (Borner et al., 2002, 2003; Eisenhaber et al., 2003; Gillmor et al., 2005). In addition, in the AC network, *MBP10* and *MBP20* connected to *SlFUL1* through *Solyc03g083500.3*, which encodes a DUF1442 family protein with potential methyltransferase activity (https://pfam.xfam.org/family/PF07279), while in the Sp network, this connection is through *Solyc02g093880.3*, which encodes a GTE8 transcription factor with kinase activity (https://www.arabidopsis.org/servlets/TairObject?type=locus&name=at3g2

7260). There is no evidence for the involvement of any of these intermediate kinase or methyltransferase in fruit development. However, *VERNALIZATION1* (*TaVRN1*), a *FUL* gene ortholog in wheat (*Triticum aestivum*, Poaceae) that has a role in the vernalization response is activated through a combination of histone methylation and demethylation, which might be mediated by methyltransferases (Chuikov et al., 2004; Oliver et al., 2009; Preston and Kellogg, 2007). In addition, kinase-induced phosphorylation is important for the stabilization of ACC SYNTHASE2, a rate-limiting enzyme in the ripening-associated ethylene biosynthesis that involves *SlFUL1* (Gapper et al., 2013; Tatsuki and Mori, 2001; Wang et al., 2019).

To investigate if other network generation methods would produce results similar to our WGCNA networks, which used the biweight midcorrelation method, I applied the Sinha Lab method with Spearman's correlation. I used the standard threshold of 0.99 to remove the unlikeliest connections between genes. In these networks, there are no intermediate genes as in the WGCNA networks (Fig. 2.48). In addition, while in the network for Sp, *TAGL1* is connected to *AP2a*, *NR*, *RIN* and *SlFUL2*, it is only connected to *AP2a* in AC (Fig. 2.48). However, evidence suggests interaction between *TAGL1* and *Cnr*, *RIN* and *SlFUL1/2* in cultivated tomato (Bemer et al., 2012).

According to our data (Fig. 2.46), previous publications (Palumbo et al., 2014) and the Tomato Expression Atlas (http://tea.solgenomics.net/) (Fernandez-Pozo et al., 2017; Pattison et al., 2015; Shinozaki et al., 2018), *MBP10* and *MBP20* have their highest expression during the immature stages of fruit development while *Sl-*

FUL1, *SlFUL2*, *RIN*, *NOR*, *NR*, *Cnr*, *TAGL1* and *AP2a* have their peak expressions in stages 3 and 4. The association between the WGCNA co-expression modules that contained these genes and the developmental stages coincided with this timing of highest expression. The overlap of peak expression levels of constituent genes in a co-expression module at a given stage produces a larger module eigengene, which represents a strong association between that module and the stage. Therefore, this type of network analyses might represent *in vivo* interactions more closely when a particular set of genes function in a given pathway during their peak expression levels. However, genes such as those that encode transcription factors might still be influential at low expression levels although these might be overlooked in this type of analyses.

Potential weaknesses of the analyses

In our preliminary analyses that I report in this chapter, we used five stages with three replicates each in both AC and Sp to generate co-expression clusters as well as networks. We later discovered that we did not have enough samples to get reliable results for these analyses. On the order of 50 samples are needed (with biological replicates not counting as separate samples) to have enough data to distinguish signal from noise (Z. Fei, personal communication). Therefore, any future work will require additional samples to test any hypotheses in this chapter.

Due to the lack of an annotated Sp genome, I used the genome annotation for cultivated tomato to map the Sp RNA-seq libraries. This is based on 99.4% sequence similarity (Pattison et al., 2015; Tomato Genome Consortium, 2012). However, in

cases where there is sequence divergence, alignment might not have been successful, leading to reduced Sp transcript counts.

Our initial analysis step in WGCNA produced two potential soft-threshold values. Although I chose the recommended higher value, which provides a more stringent output, this may have resulted in different associations of co-expression modules with developmental stages. For instance, at soft-threshold=15, in both species, there is one down-regulated module associated with stages 1, 2 and 4, one up-regulated module associated with stages 2 and 3, and two upregulated modules associated with stage 4. However, at soft-threshold=7, in both species, there is one downregulated module associated with stages 2, 3 and 4, and one upregulated module associated with stages 3, breaker and 4. Only this threshold also included a co-expression module that is upregulated in AC breaker and stage 4 but is down-regulated in the same stages in Sp. In addition, at soft-threshold=15, there are 104 methyltransferases in modules associated with AC stage 1 and only 6 methyltransferases in those associated with Sp stage 1. However, at soft-threshold=7, there are 155 methyltransferases in the modules associated with AC stage 1 but only 3 methyltransferases in the modules that are associated with Sp stage 1. Therefore, although the overall pattern in the number of methyltransferases for stage 1 in the two species is similar, the numbers are different.

A general shortcoming of predicting top-down *in silico* networks (Gaiteri et al., 2015; Pezzulo and Levin, 2016; Rodriguez-Zas et al., 2008) based on RNAseq data is that a gene might also have functions during stages when its expression is not

high. Therefore, the connectivity between the genes in the predicted network might be different from the *in vivo* mechanisms. In addition, I performed network analyses using two different methods, and obtained two different connection patterns between the targeted genes. WGCNA separates gene clusters based on their peak expression prior to network generation. Based on our data, due to pre-clustering, WGCNA is outputting weak connections for genes based on their grouping in different clusters as opposed to finding strong connections for genes within the same cluster whereas the Sinha lab method does not employ such pre-clustering. In addition, two different methods of statistical correlation (biweight midcorrelation in WGCNA and Spearman's correlation in the Sinha lab method) are used in these analyses, which might also have contributed to the discrepancies in our networks.

To generate these preliminary networks, I only used specific genes with known functions in fruit development (with *MBP10* and *MBP20* being an exception) to limit the scope to a tractable size. The only exception was when a gene did not directly connect with the core-ripening genes (e.g., *Cnr* and *SlFUL1*, *SlFUL2* and *MBP20*, *SlFUL2* and *MBP10/20*), I also included any other gene that formed the required link although the latter may have no known functions in fruit development. Therefore, our networks do not include the complete genetic architecture of tomato development. In addition, these as well as the networks for the 10 putative genes most closely connected with *FUL* orthologs produced different results for the two species. However, given their close evolutionary relatedness, it is doubtful the basic molecular mechanisms of fruit development will considerably vary between

AC and Sp. Thus, these discrepancies may be attributed to not having enough transcriptome samples in our analyses.

Conclusion

Our differential gene expression analyses revealed intraspecific differences in gene expression over the course of fruit development in cultivated tomato and *S. pimpinellifolium*. However, in line with previous research, we found only minor variation between these two closely related species. Our preliminary co-expression cluster analysis produced gene groups with concerted high or low expression patterns during a given stage in both species or only in one species, potentially associated with domestication. We also generated networks for a group of genes involved in fruit development using two different methods and found considerable differences between the output. The number of samples we employed did not satisfy that generally required for co-expression cluster analysis or network generation, which may have contributed to this discrepancy in results.

Ahn, C. S., Lee, J. H., and Pai, H.-S. (2005). Silencing of *NbNAP1* encoding a plastidic SufB-like protein affects chloroplast development in *Nicotiana benthamiana*. Mol. Cells 20, 112–118.

Albert, V. A., and Chang, T.-H. (2014). Evolution of a hot genome. Proc. Natl. Acad. Sci. U. S. A. 111, 5069.

Alexander, L., and Grierson, D. (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. J. Exp. Bot. 53, 2039–2055.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Arita, M. (2005). Scale-freeness and biological networks. J. Biochem. 138, 1–4.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

Barry, C. S., and Giovannoni, J. J. (2006). Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. Proc. Natl. Acad. Sci. U. S. A. 103, 7923–7928.

Bemer, M., Karlova, R., Ballester, A. R., Tikunov, Y. M., Bovy, A. G., Wolters-Arts, M., et al. (2012). The tomato *FRUITFULL* homologs *TDR4/FUL1* and *MBP7/FUL2* regulate ethylene-independent aspects of fruit ripening. Plant Cell 24, 4437–4451.

Bertoni, G. (2011). CBS domain proteins regulate redox homeostasis. Plant Cell 23, 3562.

Bewick, A. J., and Schmitz, R. J. (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. 36, 103–110.

Bolmgren, K., and Eriksson, O. (2010). Seed mass and the evolution of fleshy fruits in angiosperms. Oikos 119, 707–718.

Borner, G. H. H., Lilley, K. S., Stevens, T. J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. Plant Physiol. 132, 568–577.

Borner, G. H. H., Sherrier, D. J., Stevens, T. J., Arkin, I. T., and Dupree, P. (2002). Prediction of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A genomic analysis. Plant Physiol. 129, 486–499.

Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Brown, S., et al. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development 139, 3817–3826. doi:10.1242/dev.084053.

Burko, Y., Shleizer-Burko, S., Yanai, O., Shwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for APETALA1/FRUITFULL transcription factors in tomato leaf development. Plant Cell 25, 2070–2083.

Campos, M. L., Yoshida, Y., Major, I. T., de Oliveira Ferreira, D., Weraduwage, S. M., Froehlich, J. E., et al. (2016). Rewiring of jasmonate and phytochrome B signalling uncouples plant growth-defense tradeoffs. Nat. Commun. 7, 12570.

Cantu, D., Blanco-Ulate, B., Yang, L., Labavitch, J. M., Bennett, A. B., and Powell, A. L. T. (2009). Ripening-regulated susceptibility of tomato fruit to *Botrytis cinerea* requires *NOR* but not *RIN* or ethylene. Plant Physiol. 150, 1434–1449.

Chen, H., Jones, A. D., and Howe, G. A. (2006). Constitutive activation of the jasmonate signaling pathway enhances the production of secondary metabolites in tomato. FEBS Lett. 580, 2540–2546.

Cheniclet, C., Rong, W. Y., Causse, M., Frangne, N., Bolling, L., Carde, J.-P., et al. (2005). Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. Plant Physiol. 139, 1984–1994.

Chen, W., Kong, J., Lai, T., Manning, K., Wu, C., Wang, Y., et al. (2015). Tuning *LeSPL-CNR* expression by SlymiR157 affects tomato fruit ripening. Sci. Rep. 5, 7852.

Chevalier, C., Bourdon, M., Pirrello, J., Cheniclet, C., Gévaudant, F., and Frangne, N. (2014). Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J. Exp. Bot. 65, 2731–2746.

Chevalier, C., Nafati, M., Mathieu-Rivet, E., Bourdon, M., Frangne, N., Cheniclet, C., et al. (2011). Elucidating the functional role of endoreduplication in tomato fruit development. Ann. Bot. 107, 1159–1169.

Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., et al. (2004). Regulation of p53 activity through lysine methylation. Nature 432, 353–360. doi:10.1038/nature03117.

Chung, M.-Y., Vrebalov, J., Alba, R., Lee, J., McQuinn, R., Chung, J.-D., et al. (2010). A tomato (*Solanum lycopersicum*) *APETALA2/ERF* gene, SlAP2a is a negative regulator of fruit ripening. Plant J. 64, 936–947.

Colombo, M., Brambilla, V., Marcheselli, R., Caporali, E., Kater, M. M., and Colomb o, L. (2010). A new role for the *SHATTERPROOF* genes during *Arabidopsis* gynoecium development. Dev. Biol. 337, 294–302.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems 1695, 1–9.

Czerednik, A., Busscher, M., Angenent, G. C., and de Maagd, R. A. (2015). The cell size distribution of tomato fruit can be changed by overexpression of *CDKA1*. Plant Biotechnol. J. 13, 259–268.

David, K. M., Couch, D., Braun, N., Brown, S., Grosclaude, J., and Perrot-Rechenm ann, C. (2007). The auxin-binding protein 1 is essential for the control of cell cycle. Plant J. 50, 197–206.

Del Genio, C. I., Gross, T., and Bassler, K. E. (2011). All Scale-Free Networks Are Sparse. Phys. Rev. Lett. 107. doi:10.1103/physrevlett.107.178701.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Du, L., Song, J., Forney, C., Palmer, L. C., Fillmore, S., and Zhang, Z. (2016). Proteome changes in banana fruit peel tissue in response to ethylene and high-temperature treatments. Hortic Res 3, 16012.

Ecker, J. R. (2013). Epigenetic trigger for tomato ripening. Nat. Biotechnol. 31, 119–120.

Egea, I., Bian, W., Barsan, C., Jauneau, A., Pech, J.-C., Latché, A., et al. (2011). Chloroplast to chromoplast transition in tomato fruit: spectral confocal microscopy analyses of carotenoids and chlorophylls in isolated plastids and time-lapse recording on intact live tissue. Ann. Bot. 108, 291–297.

Eisenhaber, B., Wildpaner, M., Schultz, C. J., Borner, G. H. H., Dupree, P., and Eisenhaber, F. (2003). Glycosylphosphatidylinositol lipid anchoring of plan t proteins. Sensitive prediction from sequence- and genome-wide studies for *Arabidopsis* and rice. Plant Physiol. 133, 1691–1701.

Falara, V., Akhtar, T. A., Nguyen, T. T. H., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., et al. (2011). The tomato terpene synthase gene family. Plant Physiol. 157, 770–789.

Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T., Negroni, L., Sommerer, N., et al. (2007). Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol. 143, 1327–1346.

Fernandez-Pozo, N., Zheng, Y., Snyder, S. I., Nicolas, P., Shinozaki, Y., Fei, Z., et al. (2017). The Tomato Expression Atlas. Bioinformatics 33, 2397–2398.

Frary, A., Clint Nesbitt, T., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., et al. (2000). *fw2.2*: A auantitative trait locus key to the evolution of tomato fruit size. Science 289, 85–88.

Fraser, P. D., Bramley, P., and Seymour, G. B. (2001). Effect of the *Cnr* mutation on carotenoid formation during tomato fruit ripening. Phytochemistry 58, 75–79.

Fu, D.-Q., Meng, L.-H., Zhu, B.-Z., Zhu, H.-L., Yan, H.-X., and Luo, Y.-B. (2016). Silencing of the *SlNAP7* gene influences plastid development and lycopene accumulation in tomato. Sci. Rep. 6, 38664.

Fujisawa, M., Shima, Y., Nakagawa, H., Kitagawa, M., Kimbara, J., Nakano, T., et al. (2014). Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. Plant Cell 26, 89–101.

Gaiteri, C., Chen, M., Szymanski, B., Kuzmin, K., Xie, J., Lee, C., et al. (2015). Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. Sci. Rep. 5, 16361.

Gallusci, P., Hodgman, C., Teyssier, E., and Seymour, G. B. (2016). DNA methylation and chromatin regulation during fleshy fruit development and ripening. Front. Plant Sci. 7, 807.

Gapper, N. E., McQuinn, R. P., and Giovannoni, J. J. (2013). Molecular and genetic regulation of fruit ripening. Plant Molecular Biology 82, 575–591. doi:10.1007/s11103-013-0050-3.

Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: A developmental perspective. Plant Cell 5, 1439.

Gillmor, C. S., Lukowitz, W., Brininstool, G., Sedbrook, J. C., Hamann, T., Poindexter, P., et al. (2005). Glycosylphosphatidylinositol-anchored proteins are required for cell wall synthesis and morphogenesis in *Arabidopsis*. Plant Cell 17, 1128–1140.

Giovannoni, J., Yen, H., Shelton, B., Miller, S., Vrebalov, J., Kannan, P., et al. (1999). Genetic mapping of ripening and ethylene-related loci in tomato. Theoretical and Applied Genetics 98, 1005–1013. doi:10.1007/s00122005 1161.

Goff, K. E., and Ramonell, K. M. (2007). The role and regulation of receptor-like kinases in plant defense. Gene Regulation and Systems Biology 1, 117762500700100. doi:10.1177/117762500700100015.

Greene, W. K., Macnish, M. G., Rice, K. L., and Thompson, R. C. A. (2015). Identification of genes associated with blood feeding in the cat flea, *Ctenocephalides felis*. Parasit. Vectors 8, 368.

Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M. A., and Stoeckert, C. J. (2017). iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. doi:10.1101/234062.

Guo, F., Zhou, W., Zhang, J., Xu, Q., and Deng, X. (2012). Effect of the citrus lycopene β-cyclase transgene on carotenoid metabolism in transgenic tomato fruits. PLoS One 7, e32221.

Hackett, R. M., Ho, C. W., Lin, Z., Foote, H. C., Fray, R. G., and Grierson, D. (2000). Antisense inhibition of the *Nr* gene restores normal ripening to the tomato *Never-ripe* mutant, consistent with the ethylene receptor-inhibition model. Plant Physiol. 124, 1079–1086.

Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438, 374–378.

Heuvel, K. J. P. T., K J P, Barendse, G. W. M., and Wullems, G. J. (2001). Effect of gibberellic acid on cell division and cell elongation in anthers of the gibberellin deficient *gib-1* mutant of tomato. Plant Biol. 3, 124–131.

Hileman, L. C., Sundstrom, J. F., Litt, A., Chen, M., Shumba, T., and Irish, V. F. (2006). Molecular and phylogenetic analyses of the MADS-box gene family in tomato. Mol. Biol. Evol. 23, 2245–2258.

Hiwasa-Tanase, K. (2016). Fruit ripening in tomato and its modification by molecular breeding techniques. Biotechnology in Agriculture and Forestry, 155–174.

Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., and van der Knaap, E. (2013). Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFP* and *YABBY* gene family in tomato. Mol. Genet. Genomics 288, 111–129.

Huot, B., Yao, J., Montgomery, B. L., and He, S. Y. (2014). Growth–defense tradeoffs in plants: A balancing act to optimize fitness. Mol. Plant 7, 1267–1287.

Hyodo, H., Yamakawa, S., Takeda, Y., Tsuduki, M., Yokota, A., Nishitani, K., et al. (2003). Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, *XTH9*, in inflorescence apices is related to cell elongation in *Arabidopsis thaliana*. Plant Mol. Biol. 52, 473–482.

Isaacson, T., Ronen, G., Zamir, D., and Hirschberg, J. (2002). Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. Plant Cell 14, 333–342.

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., Shima, Y., Nakamura, N., et al. (2017). Re-evaluation of the *rin* mutation and the role of *RIN* in the induction of tomato ripening. Nat Plants 3, 866–874.

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., and Toki, S. (2015). CRISPR/Cas 9-mediated mutagenesis of the *RIN* locus that regulates tomato fruit ripening. Biochem. Biophys. Res. Commun. 467, 76–82.

Jiang, N., Gao, D., Xiao, H., and van der Knaap, E. (2009). Genome organization of the tomato sun locus and characterization of the unusual retrotransposon Rider. Plant J. 60, 181–193.

Karlova, R., Rosin, F. M., Busscher-Lange, J., Parapunova, V., Do, P. T., Fernie, A. R., et al. (2011). Transcriptome and metabolite profiling show that *APETALA2a* is a major regulator of tomato fruit ripening. Plant Cell 23, 923–941.

Kimbara, J., Ohyama, A., Chikano, H., Ito, H., Hosoi, K., Negoro, S., et al. (2018). QTL mapping of fruit nutritional and flavor components in tomato (*Solanum lycopersicum*) using genome-wide SSR markers and recombinant inbred lines (RILs) from an intra-specific cross. Euphytica 214. doi:10.1007/s10681-018-2295-z.

Klee, H. J. (2010). Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. New Phytol. 187, 44–56.

Knapp, S. (2002). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J. Exp. Bot. 53, 2001–2022.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version 61.

Krueger, F. (2017). Trim Galore!. Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

Kryczek, I., Lin, Y., Nagarsheth, N., Peng, D., Zhao, L., Zhao, E., et al. (2014). IL-22(+)CD4(+) T cells promote colorectal cancer stemness via STAT3 transcription factor activation and induction of the methyltransferase DOT1L. Immunity 40, 772–784.

Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., et al. (2009). Functional analyses of LONELY GUY cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. Plant Cell 21, 3152–3169.

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

Lee, J. M., Joung, J.-G., McQuinn, R., Chung, M.-Y., Fei, Z., Tieman, D., et al. (2012). Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. Plant J. 70, 191–204.

Leseberg, C. H., Eissler, C. L., Wang, X., Johns, M. A., Duvall, M. R., and Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. J. Exp. Bot. 59, 2253–2265.

Lin, X., Li, J., Zhao, Q., Feng, J.R., Gao, Q., and Nie, J.Y. (2018). WGCNA reveals key roles of *IL8* and *MMP-9* in progression of involvement area in colon of patients with ulcerative colitis. Curr Med Sci 38, 252–258.

Litt, A., and Irish, V. F. (2003). Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics 165, 821–833.

Liu, J. (2003). Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantitative trait locus *fw2.2* controls fruit size. Plant Physiol. 132, 292–299.

Liu, W., Li, L., Ye, H., Chen, H., Shen, W., Zhong, Y., et al. (2017a). From *Saccharomyces cerevisiae* to human: The important gene co-expression modules. Biomedical Reports 7, 153–158.

Liu, X., Kim, Y. J., Müller, R., Yumul, R. E., Liu, C., Pan, Y., et al. (2011). *AGAMOUS* terminates floral stem cell maintenance in *Arabidopsis* by directly repressing *WUSCHEL* through recruitment of Polycomb Group proteins. Plant Cell 23, 3654–3670.

Liu, Y., Wang, L., Zhang, T., Yang, X., and Li, D. (2017b). Functional characterization of KS-type dehydrin ZmDHN13 and its related conserved domains under oxidative stress. Sci. Rep. 7, 7361.

Lopes, F. M., Martins, D. C., Barrera, J., and Cesar, R. M. (2014). A feature selection technique for inference of graphs from their known topological properties: Revealing scale-free gene regulatory networks. Inf. Sci. 272, 1–15.

Lo, S.F., Yang, S.Y., Chen, K.T., Hsing, Y.I., Zeevaart, J. A. D., Chen, L.J., et al. (2008). A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell 20, 2603–2618.

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

Lu, Y., Liu, L., Wang, X., Han, Z., Ouyang, B., Zhang, J., et al. (2016). Genome-wide identification and expression analysis of the expansin gene family in tomato. Mol. Genet. Genomics 291, 597–608.

Lyko, F., and Brown, R. (2005). DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. JNCI: Journal of the National Cancer Institute 97, 1498–1506. doi:10.1093/jnci/dji311.

Mambro, R. D., Di Mambro, R., De Ruvo, M., Pacifici, E., Salvi, E., Sozzani, R., et al. (2017). Auxin minimum triggers the developmental switch from cell division to cell differentiation in the *Arabidopsis* root. Proceedings of the National Academy of Sciences 114, E7641–E7649. doi:10.1073/pnas.1705833114.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. 38, 948–952.

Marín-Rodríguez, M. C., Orchard, J., and Seymour, G. B. (2002). Pectate lyases, cell wall degradation and fruit softening. J. Exp. Bot. 53, 2115–2119.

Martel, C., Vrebalov, J., Tafelmeyer, P., and Giovannoni, J. J. (2011). The tomato MADS-Box transcription factor RIPENING INHIBITOR Interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-Dependent Manner. Plant Physiol. 157, 1568–1579.

Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. 15, 394–408.

Ma, X., Balazadeh, S., and Mueller-Roeber, B. (2018). Tomato fruit ripening factor NOR controls leaf senescence. doi:10.1101/436899.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Mouritsen, O., and Styrb_k, K. (2014). Umami: Unlocking the secrets of the fifth taste. Columbia University Press.

Muñoz-Espinoza, C., Di Genova, A., Correa, J., Silva, R., Maass, A., González-Agüero, M., et al. (2016). Transcriptome profiling of grapevine seedless segregants during berry development reveals candidate genes associated with berry weight. BMC Plant Biol. 16, 104.

Neuwirth, E., and Brewer, R. C. (2014). ColorBrewer palettes. R package version, 1–1.

Nunes-Nesi, A., Carrari, F., Gibon, Y., Sulpice, R., Lytovchenko, A., Fisahn, J., et al. (2007). Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. Plant J. 50, 1093–1106.

Ohnishi, T., Nomura, T., Watanabe, B., Ohta, D., Yokota, T., Miyagawa, H., et al. (2006). Tomato cytochrome P450 CYP734A7 functions in brassinosteroid catabolism. Phytochemistry 67, 1895–1906.

Oliver, S. N., Finnegan, E. J., Dennis, E. S., Peacock, W. J., and Trevaskis, B. (2009). Vernalization-induced flowering in cereals is associated with chang es in histone methylation at the *VERNALIZATION1* gene. Proc. Natl. Acad. Sci. U. S. A. 106, 8386–8391.

Osorio, S., Alba, R., Damasceno, C. M. B., Lopez-Casado, G., Lohse, M., Zanor, M. I., et al. (2011). Systems biology of tomato fruit development: combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. Plant Physiol. 157, 405–425.

Pabón-Mora, N., and Litt, A. (2011). Comparative anatomical and developmental analysis of dry and fleshy fruits of Solanaceae. Am. J. Bot. 98, 1415–1436.

Palumbo, M. C., Zenoni, S., Fasoli, M., Massonnet, M., Farina, L., Castiglione, F., et al. (2014). Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. Plant Cell 26, 4617–4635.

Pandey, A., Misra, P., Choudhary, D., Yadav, R., Goel, R., Bhambhani, S., et al. (2015). *AtMYB12* expression in tomato leads to large scale differential modulation in transcriptome and flavonoid content in leaf and fruit tissues. Sci. Rep. 5, 12412.

Pattison, R. J., Csukasi, F., Zheng, Y., Fei, Z., van der Knaap, E., and Catalá, C. (2015). Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. Plant Physiol. 168, 1684–1701.

Pei, G., Chen, L., and Zhang, W. (2017). WGCNA application to proteomic and metabolomic data analysis. Methods in Enzymology, 135–158.

Pezzulo, G., and Levin, M. (2016). Top-down models in biology: explanation and control of complex living systems above the molecular level. J. R. Soc. Interface 13. doi:10.1098/rsif.2016.0555.

Picton, S., Barton, S. L., Bouzayen, M., Hamilton, A. J., and Grierson, D. (1993a). Altered fruit ripening and leaf senescence in tomatoes expressing an antisense ethylene-forming enzyme transgene. Plant J. 3, 469–481.

Picton, S., Gray, J., Barton, S., AbuBakar, U., Lowe, A., and Grierson, D. (1993b). cDNA cloning and characterisation of novel ripening-related mRNAs with altered patterns of accumulation in the *ripening inhibitor* (*rin*) tomato ripening mutant. Plant Molecular Biology 23, 193–207. doi:10.1007/bf00021431.

Pirrello, J., Deluche, C., Frangne, N., Gévaudant, F., Maza, E., Djari, A., et al. (2018). Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. Plant J. 93, 387–398.

Preston, J. C., and Kellogg, E. A. (2007). Conservation and divergence of *APETALA1/FRUITFULL-like* gene function in grasses: evidence from gene expression analyses. Plant J. 52, 69–81.

Qin, Q., Bergmann, C. W., Rose, J. K. C., Saladie, M., Kolli, V. S. K., Albersheim, P., et al. (2003). Characterization of a tomato protein that inhibits a xyloglucan-specific endoglucanase. Plant J. 34, 327–338.

Qin, Y.M., Hu, C.Y., Pang, Y., Kastaniotis, A. J., Hiltunen, J. K., and Zhu, Y.X. (2007). Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. Plant Cell 19, 3692–3704.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: URL https://www.R-project.org/.

Reinhold, D., Morrow, J. D., Jacobson, S., Hu, J., Ringel, B., Seibold, M. A., et al. (2017). Meta-analysis of peripheral blood gene expression modules for COPD phenotypes. PLoS One 12, e0185682.

Ren, J., Sun, L., Wu, J., Zhao, S., Wang, C., Wang, Y., et al. (2010). Cloning and expression analysis of cDNAs for *ABA 8'-hydroxylase* during sweet cherry fruit maturation and under stress conditions. J. Plant Physiol. 167, 1486–1493.

Rodriguez-Zas, S. L., Ko, Y., Adams, H. A., and Southey, B. R. (2008). Advancing the understanding of the embryo transcriptome co-regulation using meta-, functional, and gene network analysis tools. Reproduction 135, 213–224.

Sauvage, C., Rau, A., Aichholz, C., Chadoeuf, J., Sarah, G., Ruiz, M., et al. (2017). Domestication rewired gene expression and nucleotide diversity patterns in tomato. Plant J. 91, 631–645.

Schaffer, A. A., and Petreikov, M. (1997). Sucrose-to-starch metabolism in tomato fruit undergoing transient starch accumulation. Plant Physiol. 113, 739–746.

Shahan, R., Zawora, C., Wight, H., Sittmann, J., Wang, W., Mount, S. M., et al. (2018). Consensus coexpression network analysis identifies key regulators of flower and fruit development in wild strawberry. Plant Physiol. 178, 202–216.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Shima, Y., Kitagawa, M., Fujisawa, M., Nakano, T., Kato, H., Kimbara, J., et al. (2013). Tomato FRUITFULL homologues act in fruit ripening via forming MADS-box transcription factor complexes with RIN. Plant Mol. Biol. 82, 427–438.

Shinozaki, Y., Nicolas, P., Fernandez-Pozo, N., Ma, Q., Evanich, D. J., Shi, Y., et al. (2018). High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nat. Commun. 9, 364.

Su, L., Diretto, G., Purgatto, E., Danoun, S., Zouine, M., Li, Z., et al. (2015). Carotenoid accumulation during tomato fruit ripening is modulated by the auxin-ethylene balance. BMC Plant Biol. 15, 114.

Takahagi, K., Inoue, K., and Mochida, K. (2018). Gene co-expression network analysis suggests the existence of transcriptional modules containing a high proportion of transcriptionally differentiated homoeologs in hexaploid wheat. Front. Plant Sci. 9, 1163.

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. The Plant Cell 16, S181–S189.

Tatsuki, M., and Mori, H. (2001). Phosphorylation of tomato 1-aminocyclopropane-1-carboxylic acid synthase, LE-ACS2, at the C-terminal region. J. Biol. Chem. 276, 28051–28057.

Teyssier, E., Bernacchia, G., Maury, S., How Kit, A., Stammitti-Bert, L., Rolin, D., et al. (2008). Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. Planta 228, 391–399.

The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. Nucleic Acids Res. 45, D331–D338.

Thompson, J. E., Froese, C. D., Madey, E., Smith, M. D., and Hong, Y. (1998). Lipid metabolism during plant senescence. Prog. Lipid Res. 37, 119–141.

Thompson, J. E., Legge, R. L., and Barber, R. F. (1987). The role of free radicals in senescence and wounding. New Phytol. 105, 317–344.

Tieman, D. M., Taylor, M. G., Ciardi, J. A., and Klee, H. J. (2000). The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. Proc. Natl. Acad. Sci. U. S. A. 97, 5663–5668.

Ties, P., and Barringer, S. (2012). Influence of lipid content and lipoxygenase on flavor volatiles in the tomato peel and flesh. J. Food Sci. 77, C830–7.

Tilly, J. J., Allen, D. W., and Jack, T. (1998). The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects. Development 125, 1647–1657.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641.

Umezawa, T., Okamoto, M., Kushiro, T., Nambara, E., Oono, Y., Seki, M., et al. (2006). CYP707A3, a major ABA 8'-hydroxylase involved in dehydration and rehydration response in *Arabidopsis thaliana*. Plant J. 46, 171–182.

Vandeleur, R. K., Mayo, G., Shelden, M. C., Gilliham, M., Kaiser, B. N., and Tyerman, S. D. (2009). The role of plasma membrane intrinsic protein aquaporins in water transport through roots: diurnal and drought stress responses reveal different strategies between isohydric and anisohydric cultivars of grapevine. Plant Physiol. 149, 445–460.

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., et al. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. Front. Plant Sci. 5, 227.

Viquez Zamora, A. M. (2015). Exploiting wild relatives of *S. lycopersicum* for quality traits. Available at: http://edepot.wur.nl/353426.

Voegele, A., Linkies, A., Müller, K., and Leubner-Metzger, G. (2011). Members of the gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. J. Exp. Bot. 62, 5131–5147.

Vrebalov, J., Pan, I. L., Arroyo, A. J. M., McQuinn, R., Chung, M., Poole, M., et al. (2009). Fleshy fruit expansion and ripening are regulated by the tomato *SHATTERPROOF* Gene *TAGL1*. Plant Cell 21, 3041–3062.

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

Wang, R.H., Yuan, X.Y., Meng, L.H., Zhu, B.Z., Zhu, H.L., Luo, Y.B., et al. (2016). Transcriptome analysis provides a preliminary regulation route of the ethylene signal transduction component, *SlEIN2*, during tomato ripening. PLoS One 11, e0168287.

Wang, R., Tavano, E. C. da R., Lammers, M., Martinelli, A. P., Angenent, G. C., and de Maagd, R. A. (2019). Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Sci. Rep. 9, 1696.

Wang, S., Lu, G., Hou, Z., Luo, Z., Wang, T., Li, H., et al. (2014). Members of the tomato *FRUITFULL* MADS-box family regulate style abscission and fruit ripening. J. Exp. Bot. 65, 3005–3014.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0–a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243.

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

Wilson-Rawls, J., Molkentin, J. D., Black, B. L., and Olson, E. N. (1999). Activated notch inhibits myogenic activity of the MADS-Box transcription factor myocyte enhancer factor 2C. Mol. Cell. Biol. 19, 2853–2862.

Wu, Y., and Zhou, J.-M. (2013). Receptor-like kinases in plant innate immunity. Journal of Integrative Plant Biology 55, 1271–1286. doi:10.1111/jipb.12123.

Yang, Y., Bao, S., Zhou, X., Liu, J., and Zhuang, Y. (2018). The key genes and pathways related to male sterility of eggplant revealed by comparative transcriptome analysis. BMC Plant Biol. 18, 209.

Ye, J., Hu, T., Yang, C., Li, H., Yang, M., Ijaz, R., et al. (2015). Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. PLoS One 10, e0130885.

Ye, L.-S., Zhang, Q., Pan, H., Huang, C., Yang, Z.-N., and Yu, Q.-B. (2017). *EMB2738*, which encodes a putative plastid-targeted GTP-binding protein, is essential for embryogenesis and chloroplast development in higher plants. Physiol. Plant. 161, 414–430.

Yuan, X., Zhang, S., Qing, X., Sun, M., Liu, S., Su, H., et al. (2013). Superfamily of ankyrin repeat proteins in tomato. Gene 523, 126–136.

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

Zhang, S., Xu, M., Qiu, Z., Wang, K., Du, Y., Gu, L., et al. (2016). Spatiotemporal transcriptome provides insights into early fruit development of tomato (*Solanum lycopersicum*). Sci. Rep. 6, 23173.

Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epig enome modifications associated with ripening. Nat. Biotechnol. 31, 154–159.

Zhu, M., Li, Y., Chen, G., Ren, L., Xie, Q., Zhao, Z., et al. (2015). Silencing *SlELP2L*, a tomato *Elongator complex protein 2-like* gene, inhibits leaf growth, accelerates leaf, sepal senescence, and produces dark-green fruit. Sci. Rep. 5, 7693.

Figure 2.1. The number of differentially expressed genes between two consecutive stages. In each comparison below, the terms up- or downregulated refers to the genes of the later stage. AC: Ailsa Craig; PIMP: *S. pimpinellifolium*; DPRE: days pre-anthesis; DPA: days post-anthesis; 1DPRE: stage 1; 3DPA: stage2; 15DPA: stage3; redripe: stage 4.

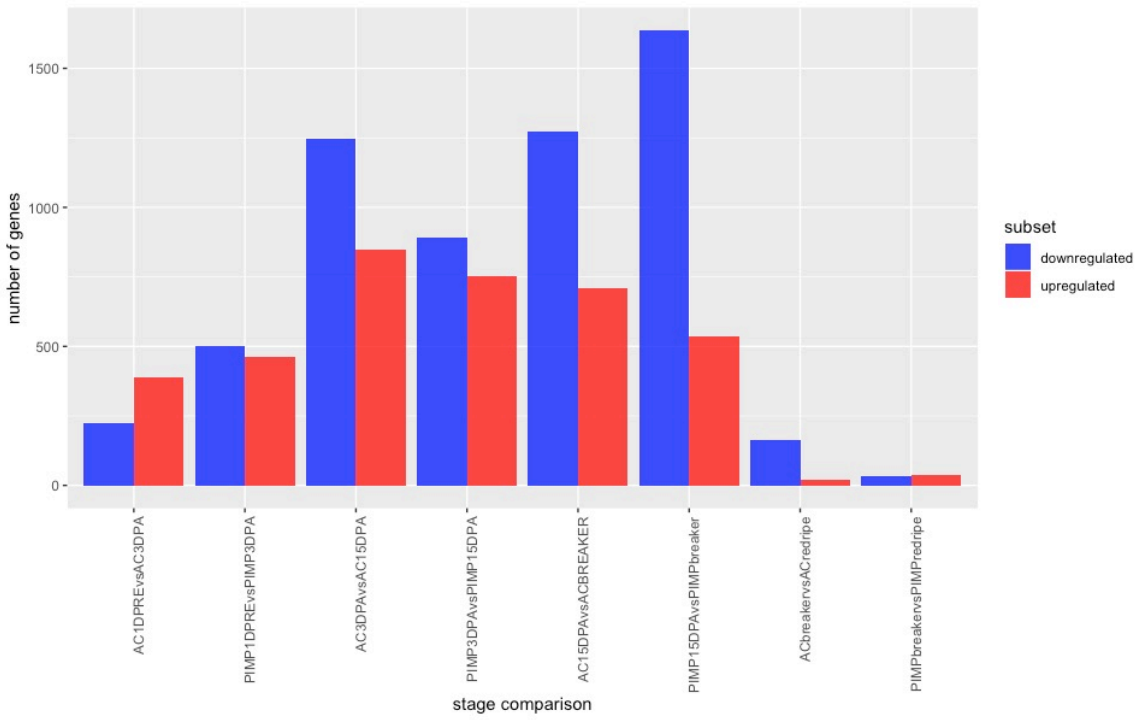


Figure 2.2. The number of differentially expressed genes between the corresponding stages of the two species. In each comparison below, the terms up- or downregulated refers to the genes in *S. lycopersicum* cv. Ailsa Craig in relation to *S. pimpinellifolium*. AC: Ailsa Craig; PIMP: *S. pimpinellifolium*; DPRE: days pre-anthesis; DPA: days post-anthesis; 1DPRE: stage 1; 3DPA: stage2; 15DPA: stage3; redripe: stage 4.

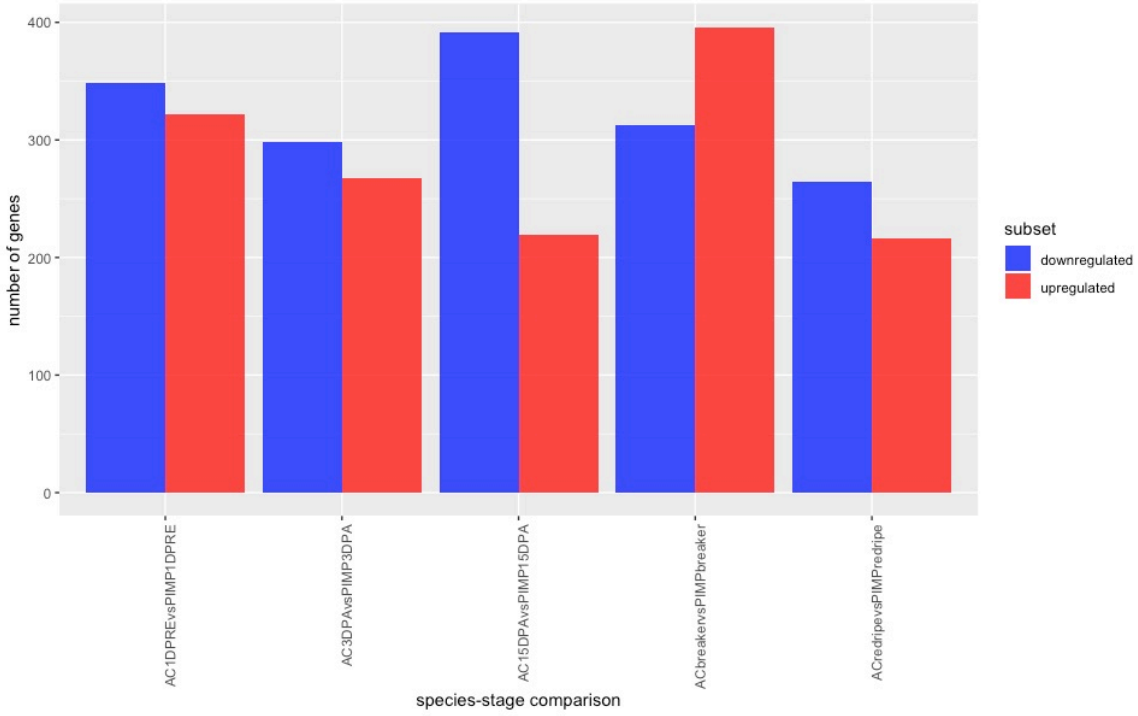


Figure 2.3. A PCA plot showing the variation among the RNAseq libraries. DPRE: days pre-anthesis; DPA: days post-anthesis; 1DPRE: stage 1; 3DPA: stage2; 15DPA: stage3; redripe: stage 4. The dotted circles represent AC while the circles with solid colors represent Sp. The dots were added manually.

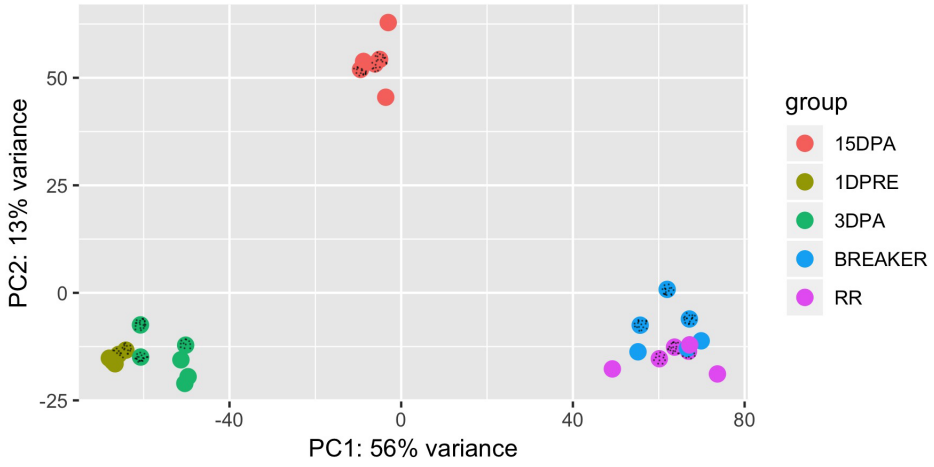


Figure 2.4. The approximate timing of the breaker stage and stage 4 of fleshy fruit development in *S. lycopersicum* cv Ailsa Craig and *S. pimpinellifolium*.

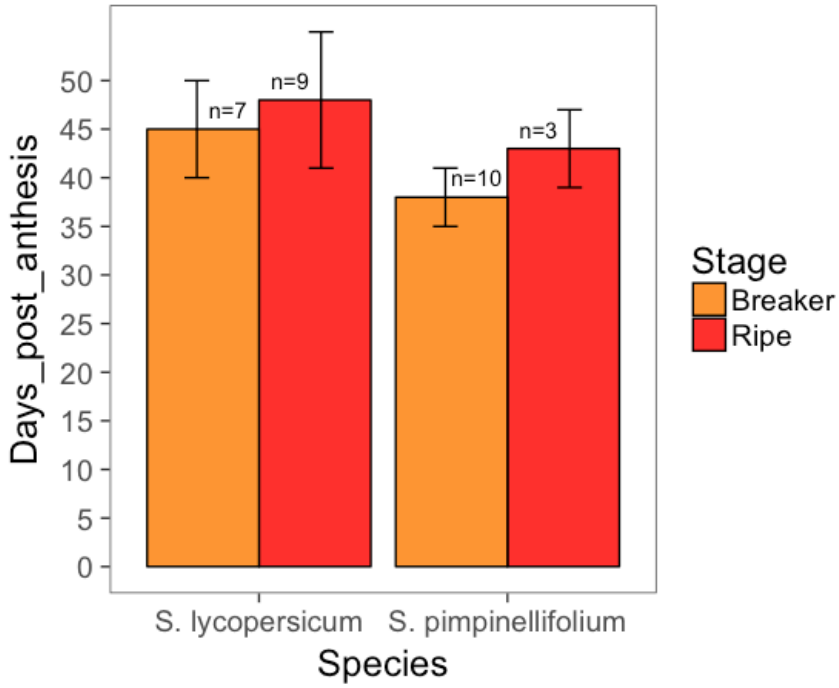


Figure 2.5. The expression heatmap for the differentially expressed genes in stage 2 vs stage 1 in *S. lycopersicum* cv Ailsa Craig.

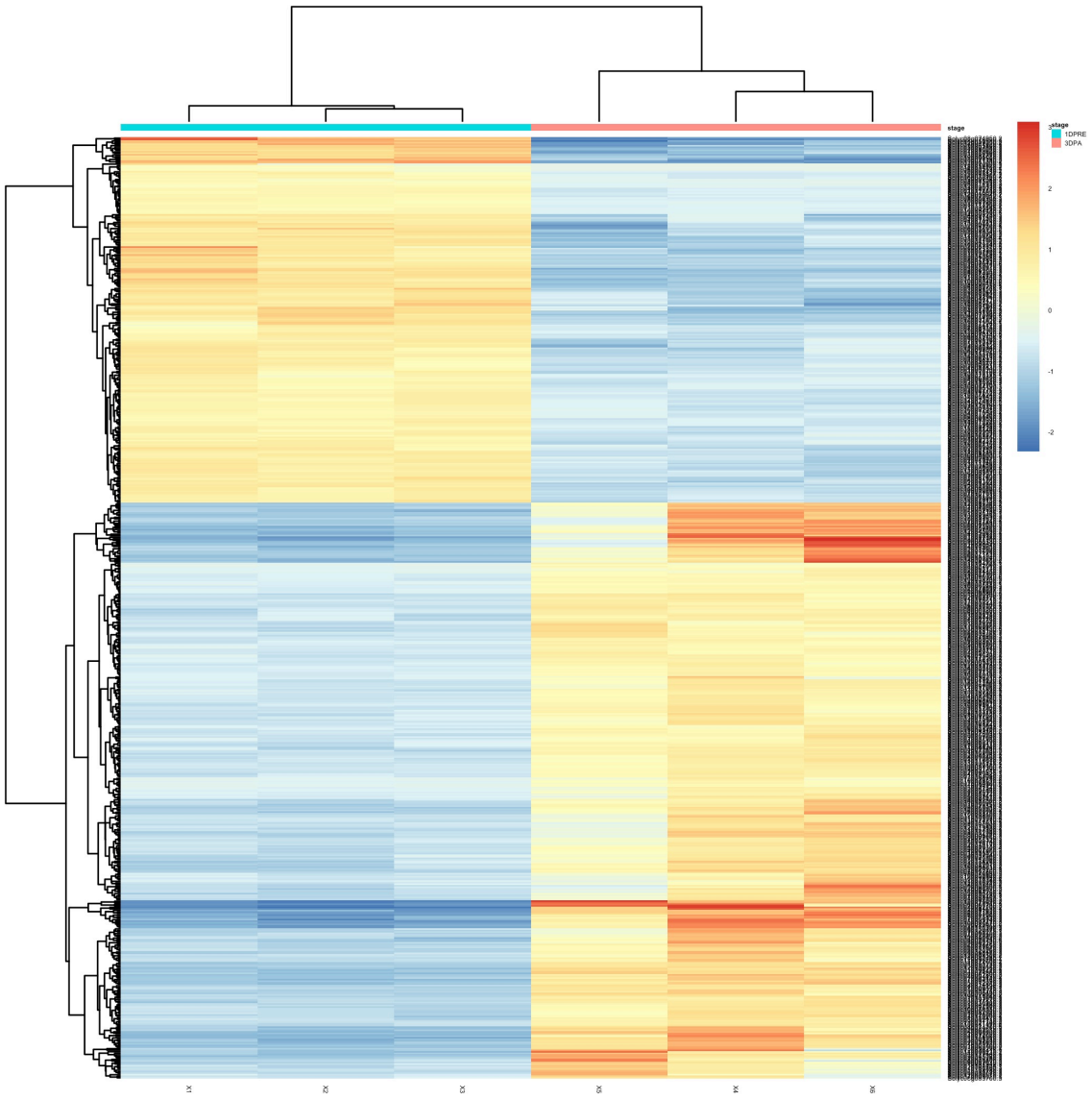


Figure 2.6. The enrichment of gene ontology terms for the downregulated genes in stage 2 vs stage 1 in *S. lycopersicum* cv Ailsa Craig.

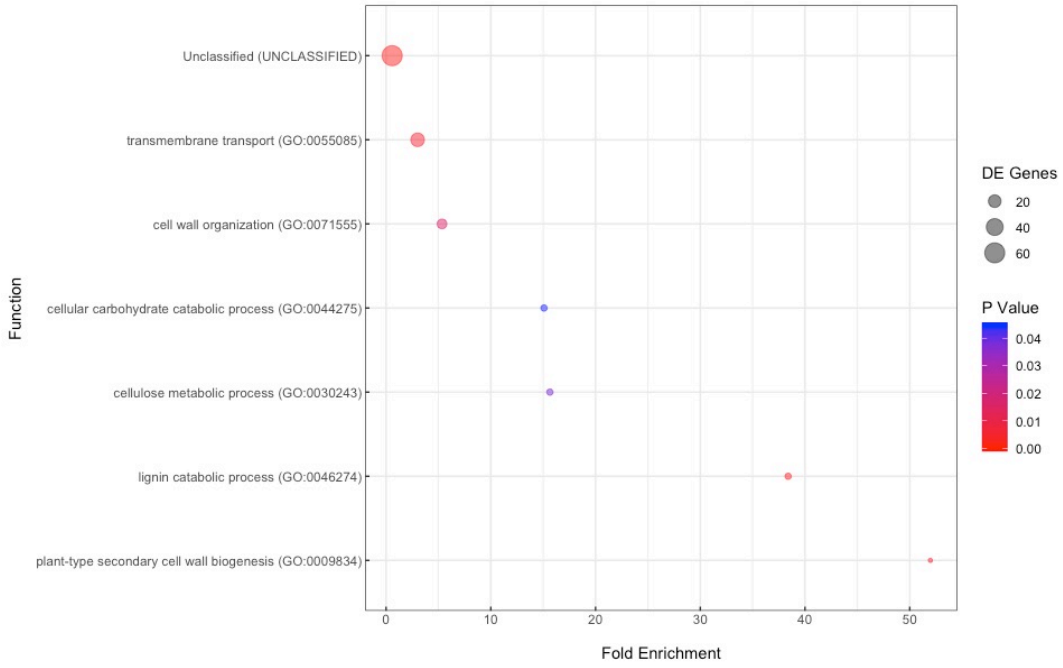


Figure 2.7. The enrichment of gene ontology terms for the upregulated genes in stage 2 vs stage 1 in *S. lycopersicum* cv Ailsa Craig.

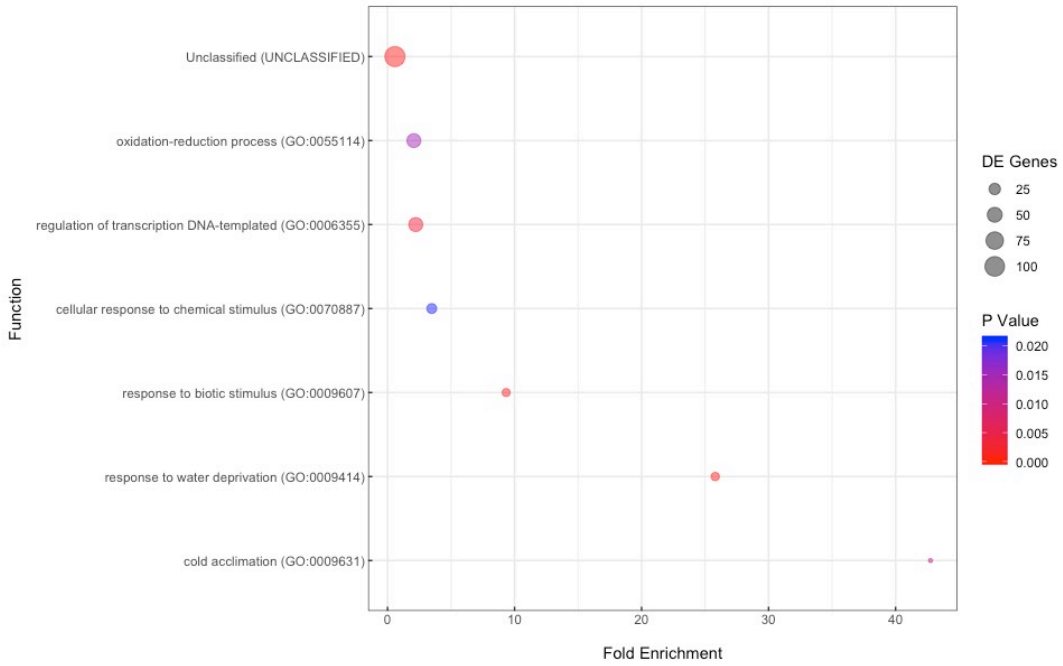


Figure 2.8. The expression heatmap for the differentially expressed genes in stage 3 vs stage 2 in *S. lycopersicum* cv Ailsa Craig.

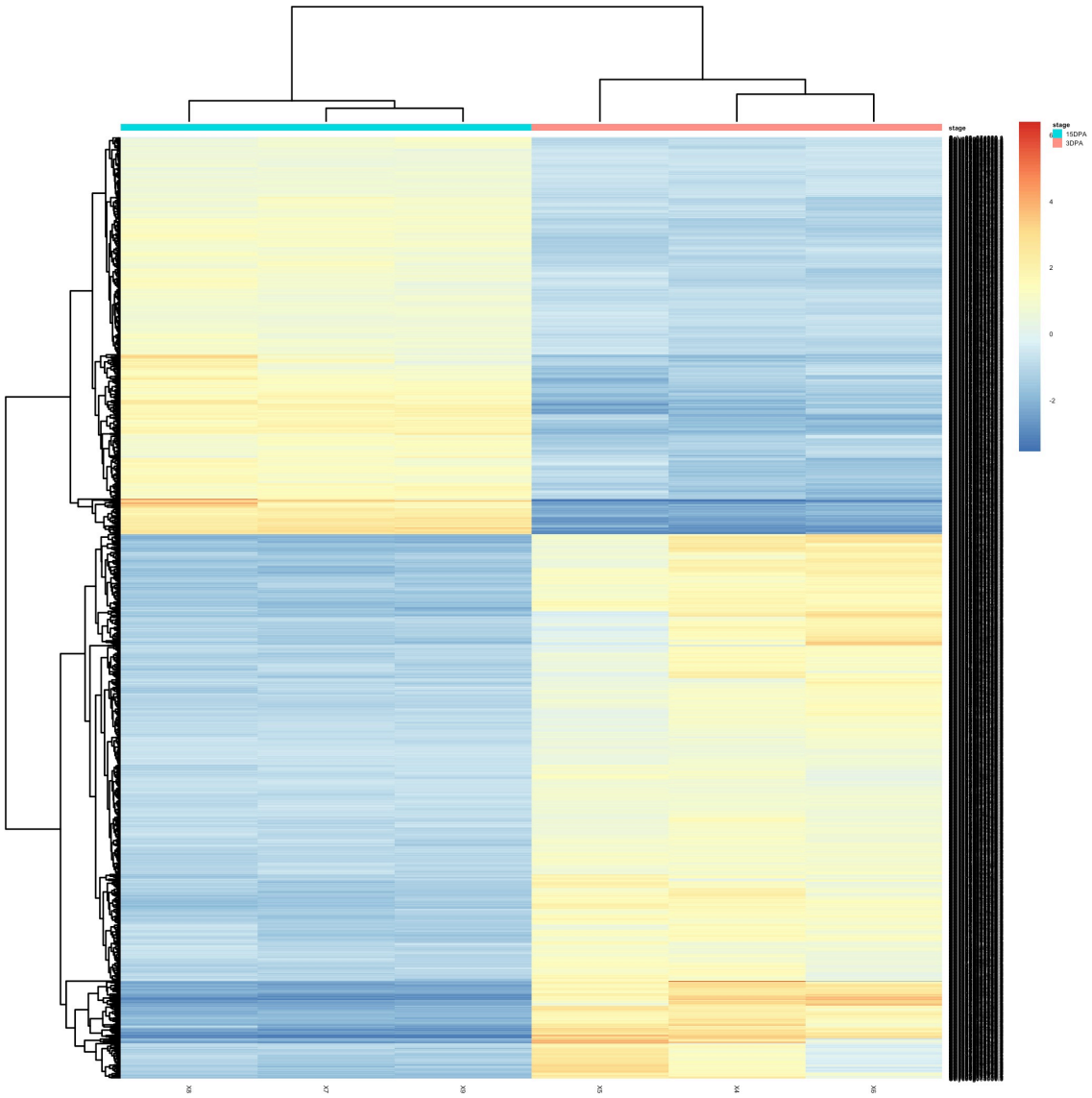


Figure 2.9. The enrichment of gene ontology terms for the downregulated genes in stage 3 vs stage 2 in *S. lycopersicum* cv Ailsa Craig.

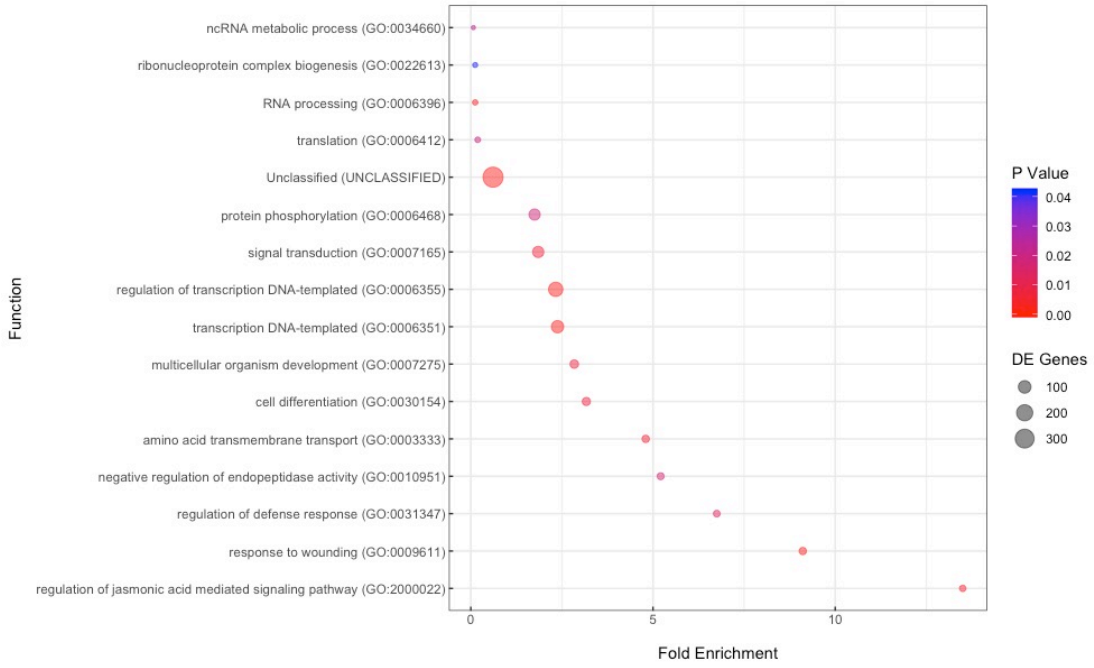


Figure 2.10. The enrichment of gene ontology terms for the upregulated genes in stage 3 vs stage 2 in *S. lycopersicum* cv Ailsa Craig.

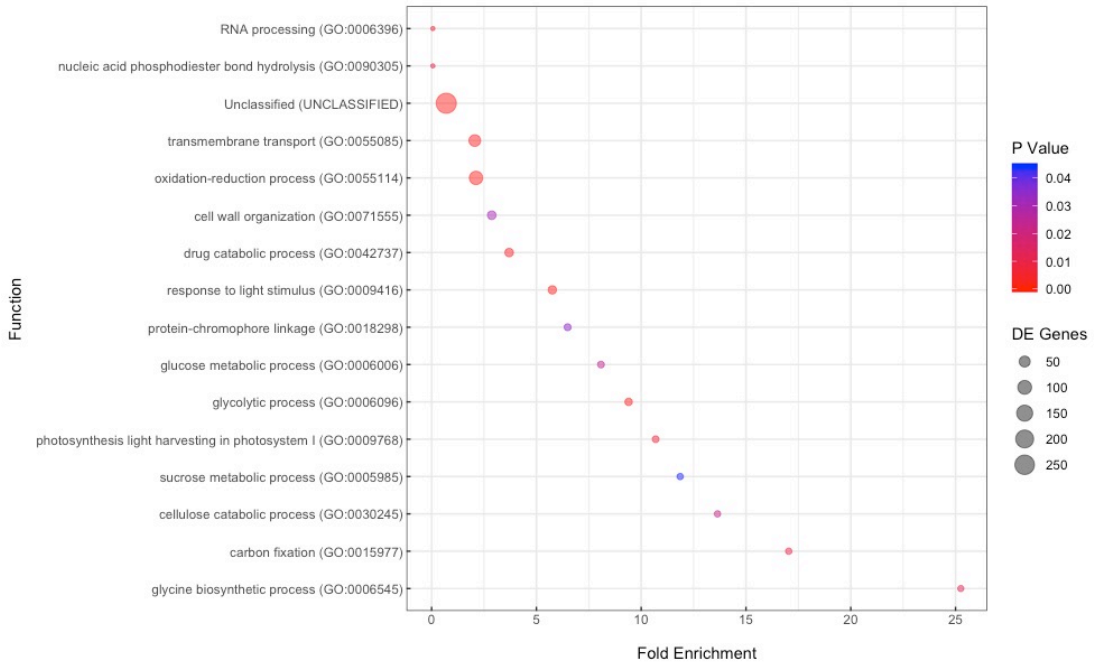


Figure 2.11. The expression heatmap for the differentially expressed genes in breaker stage vs stage 3 in *S. lycopersicum* cv Ailsa Craig.

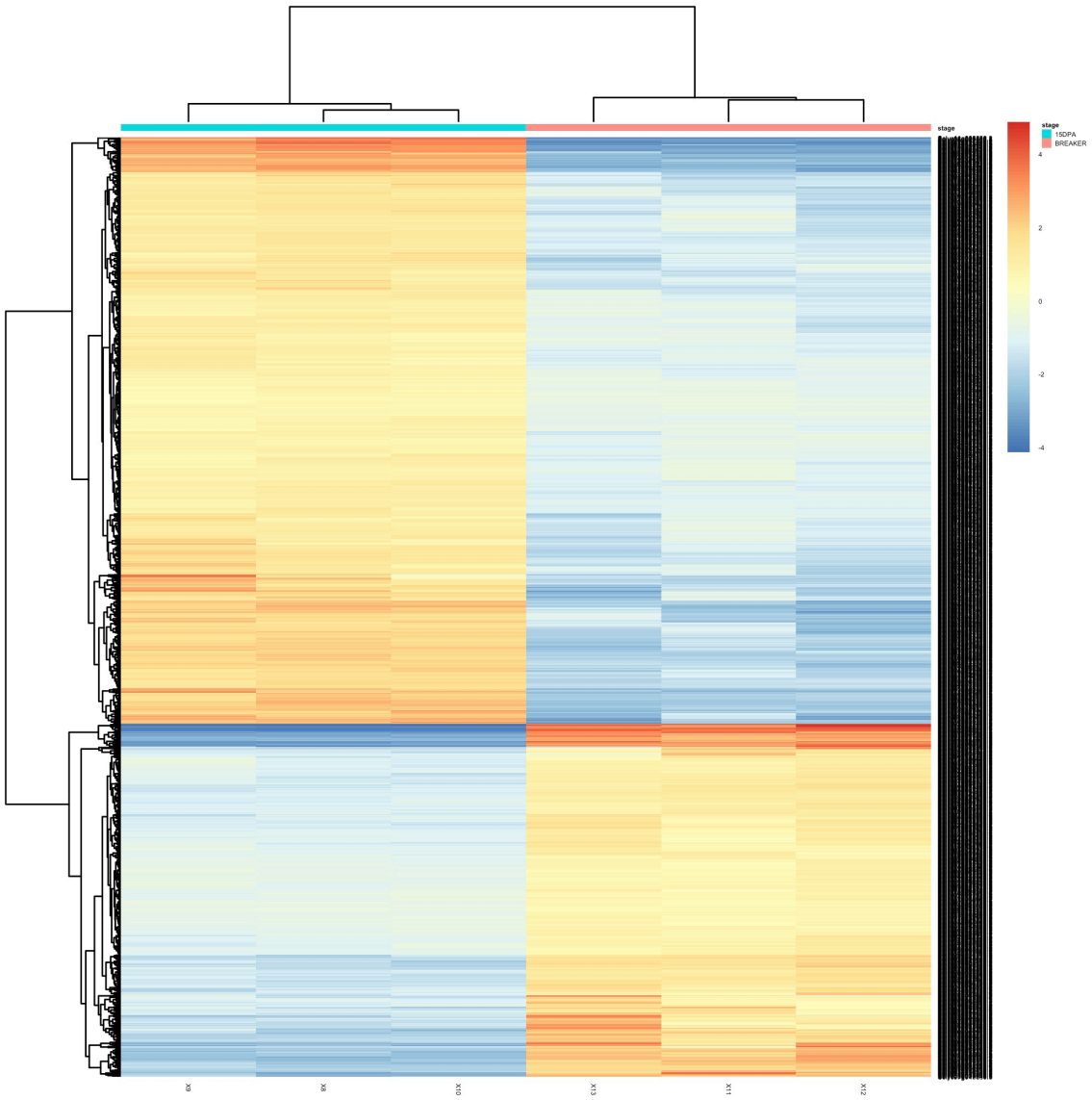


Figure 2.12. The enrichment of gene ontology terms for the downregulated genes in breaker stage vs stage 3 in *S. lycopersicum* cv Ailsa Craig.

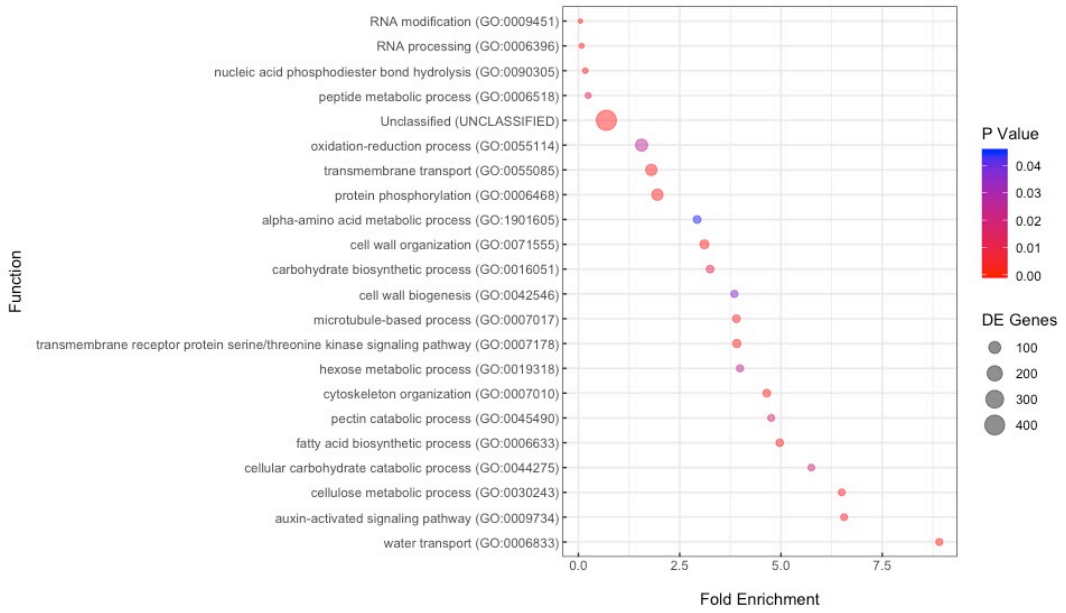


Figure 2.13. The enrichment of gene ontology terms for the upregulated genes in breaker stage vs stage 3 in *S. lycopersicum* cv Ailsa Craig.

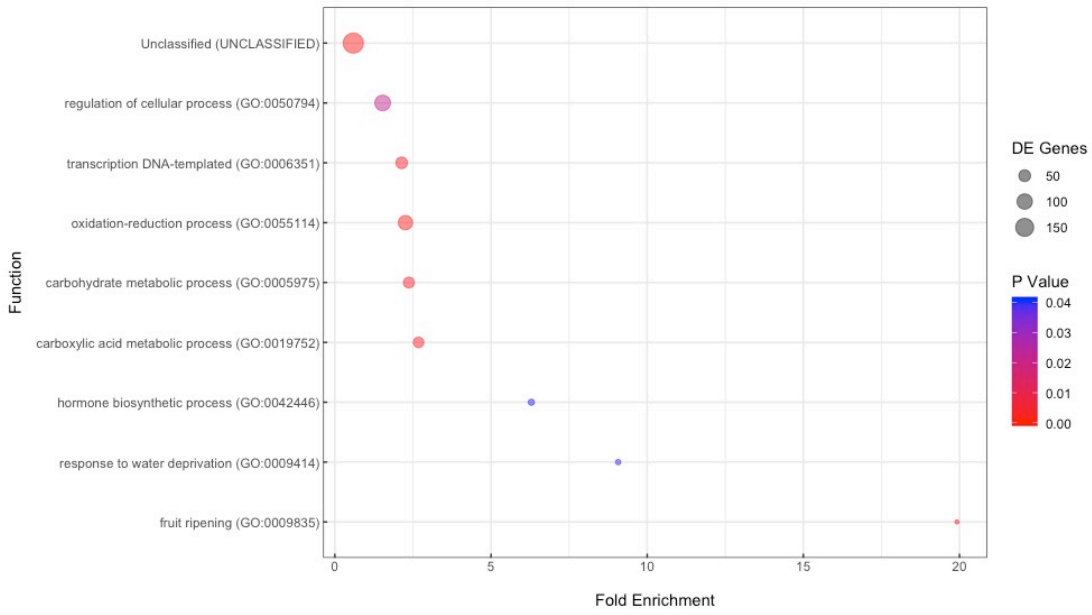


Figure 2.14. The expression heatmap for the differentially expressed genes in stage 4 vs breaker stage in *S. lycopersicum* cv Ailsa Craig.

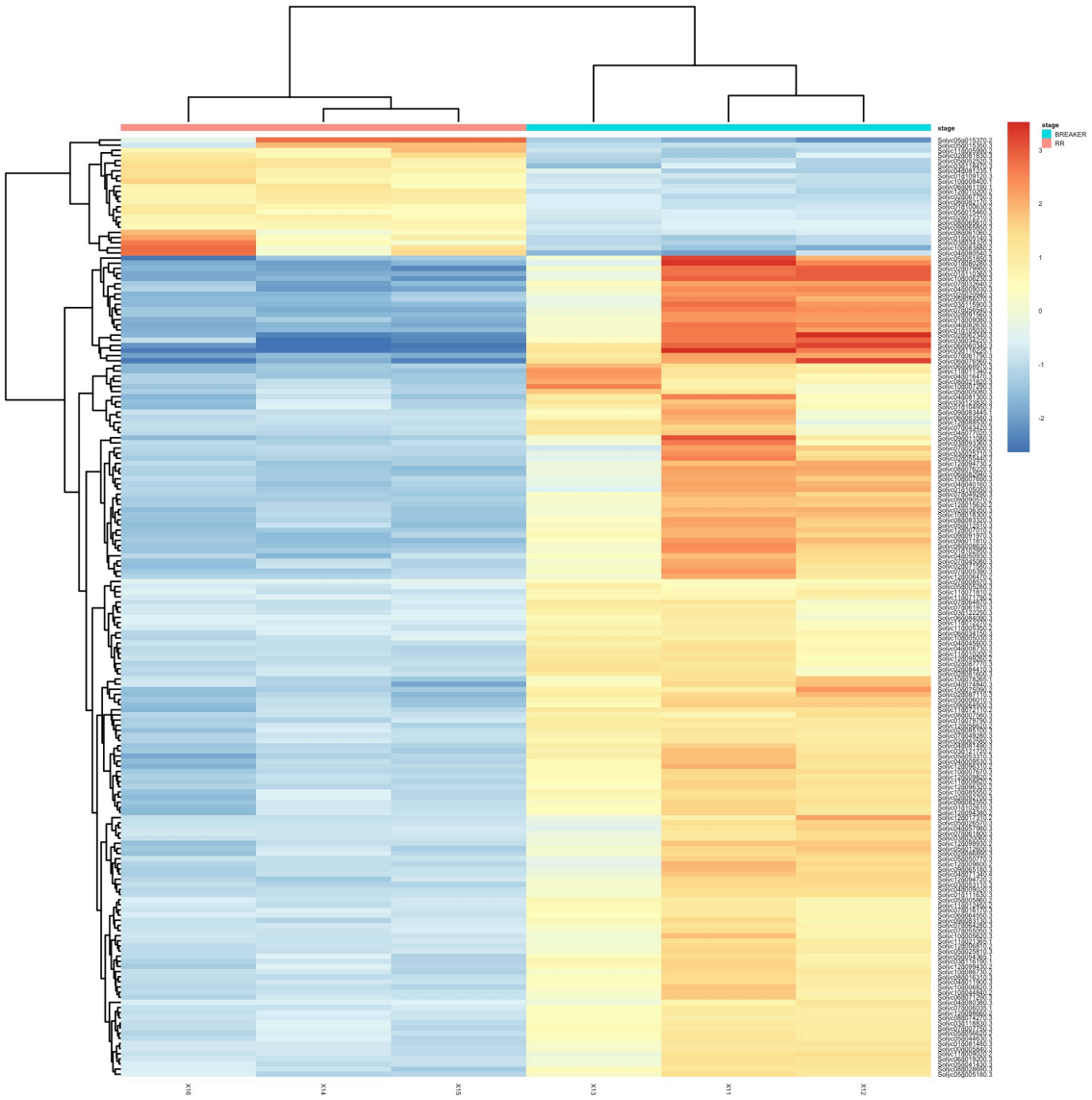


Figure 2.15. The enrichment of gene ontology terms for the downregulated genes in stage 4 vs breaker stage in *S. lycopersicum* cv Ailsa Craig.

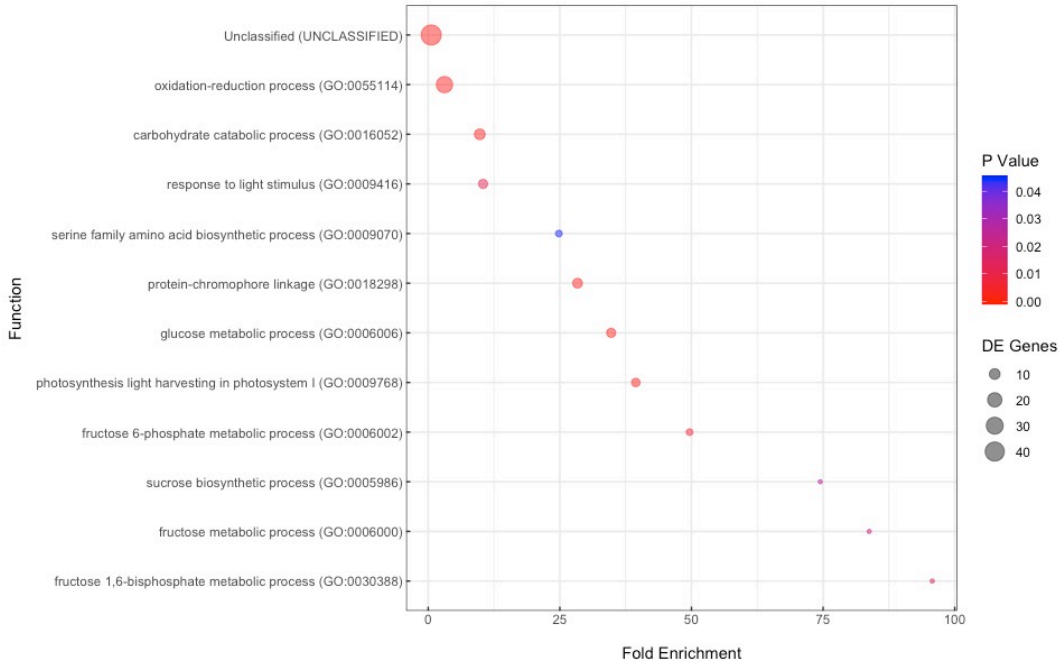


Figure 2.16. The enrichment of gene ontology terms for the upregulated genes in stage 4 vs breaker stage in *S. lycopersicum* cv Ailsa Craig.

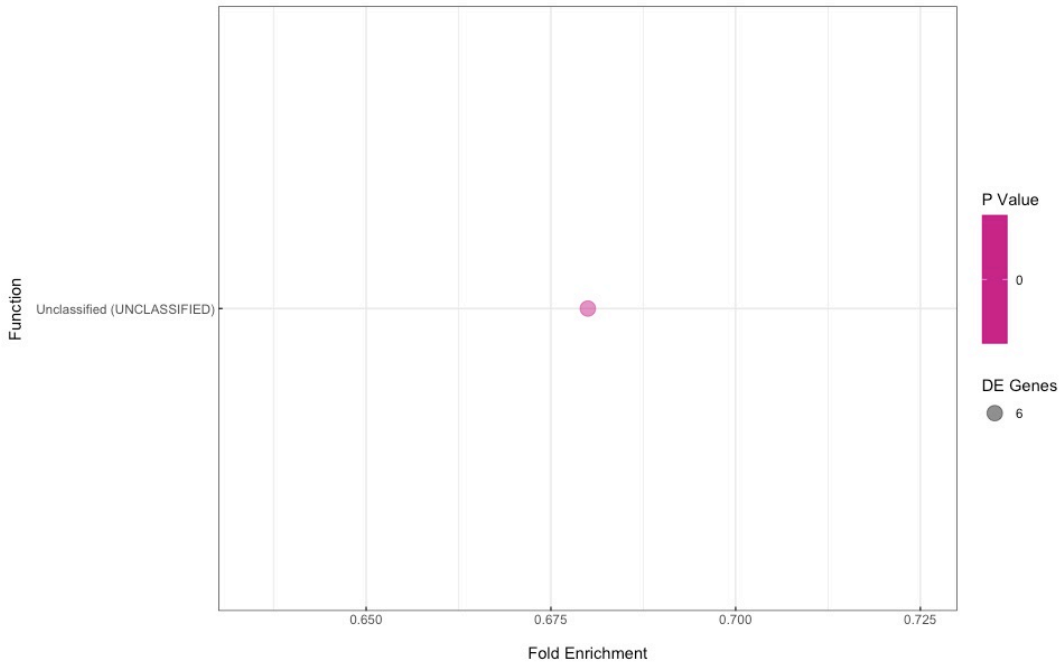


Figure 2.17. The expression heatmap for the differentially expressed genes in stage 2 vs stage 1 in *S. pimpinellifolium*.

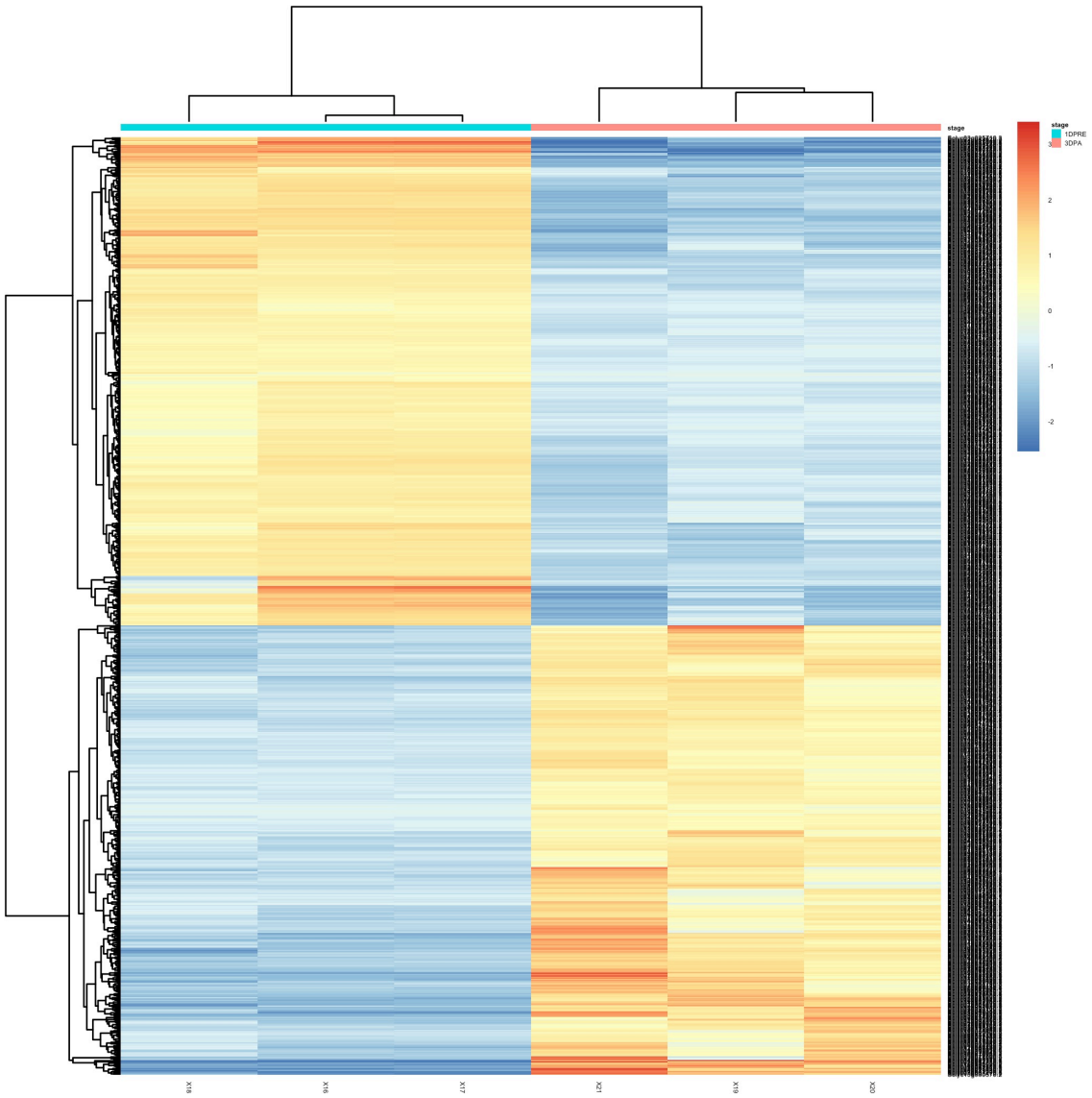


Figure 2.18. The enrichment of gene ontology terms for the downregulated genes in stage 2 vs stage 1 in *S. pimpinellifolium*.

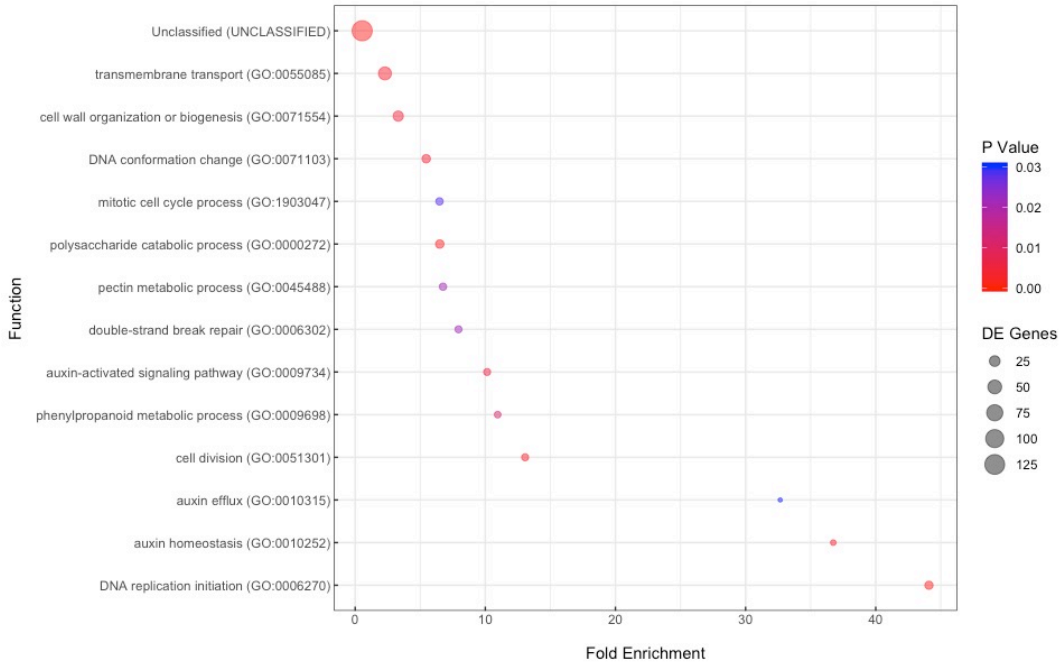


Figure 2.19. The enrichment of gene ontology terms for the upregulated genes in stage 2 vs stage 1 in *S. pimpinellifolium*.

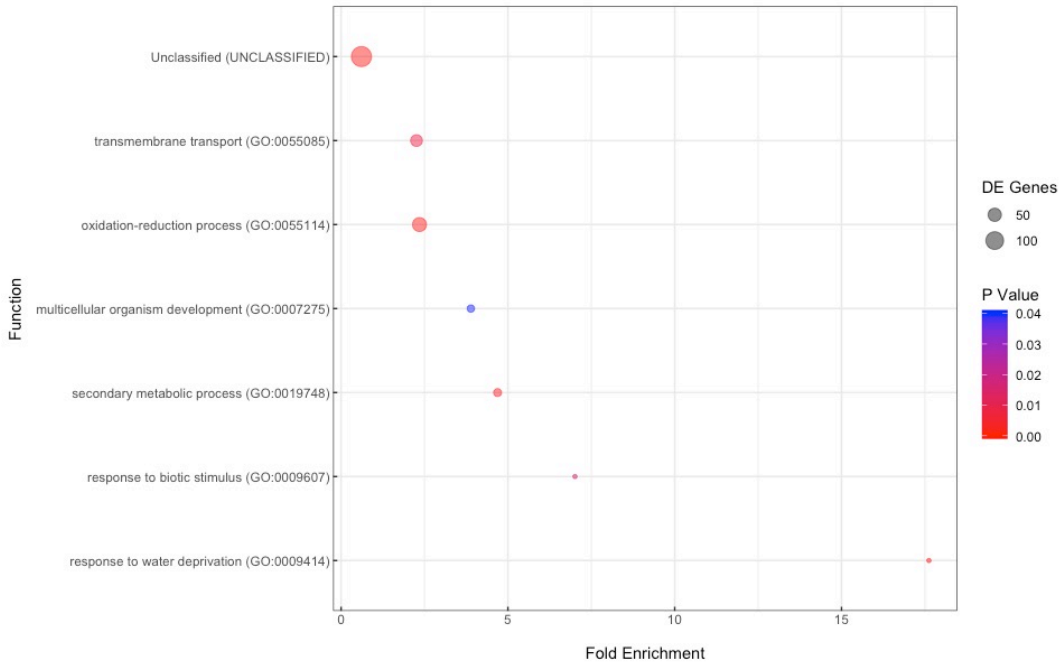


Figure 2.20. The expression heatmap for the differentially expressed genes in stage 3 vs stage 2 in *S. pimpinellifolium*.

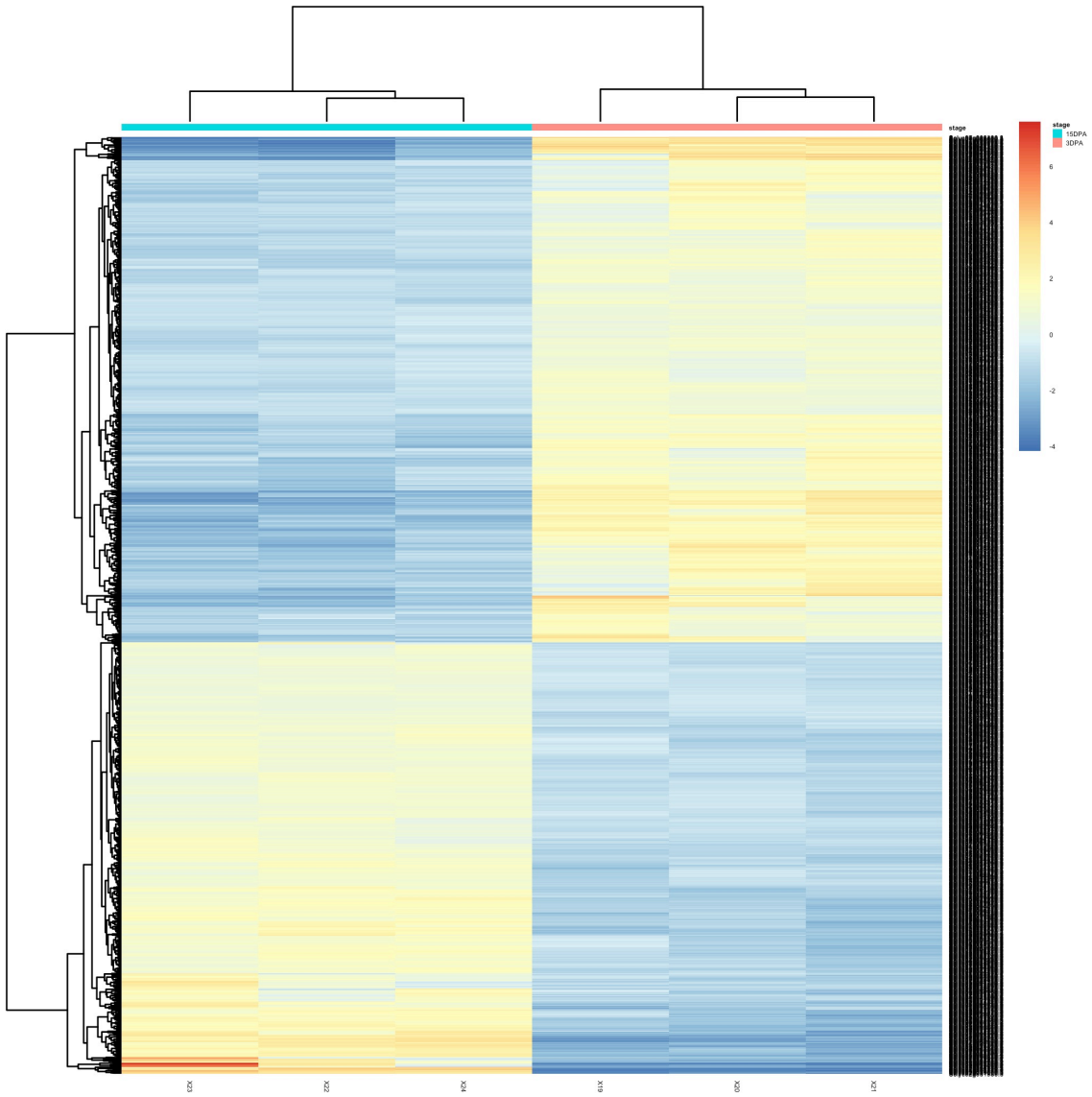


Figure 2.21. The enrichment of gene ontology terms for the downregulated genes in stage 3 vs stage 2 in *S. pimpinellifolium*.

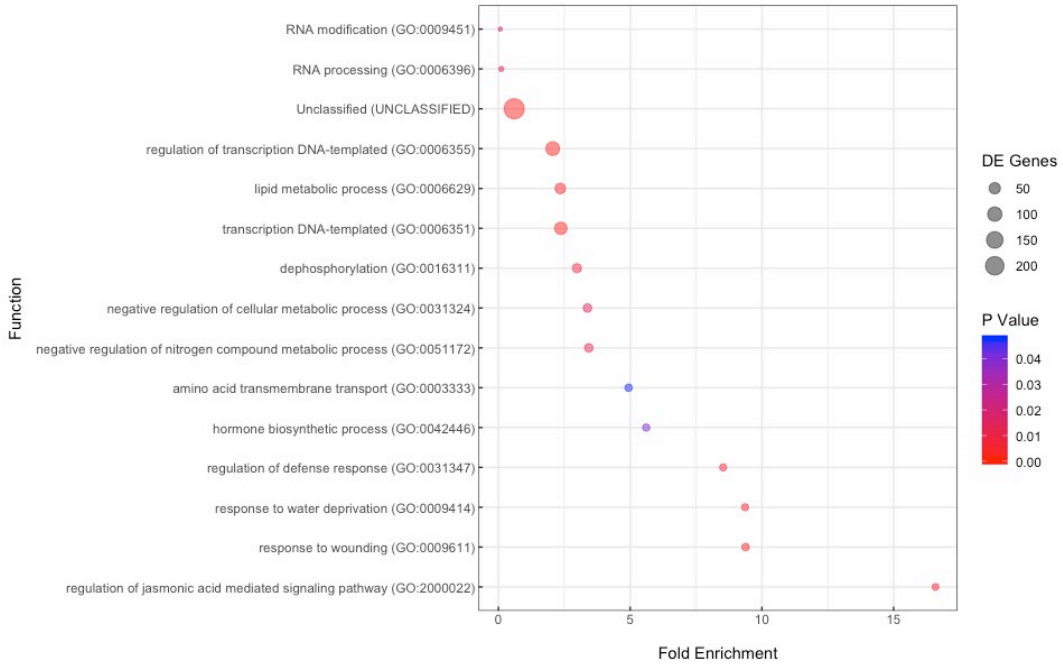


Figure 2.22. The enrichment of gene ontology terms for the upregulated genes in stage 3 vs stage 2 in *S. pimpinellifolium*.

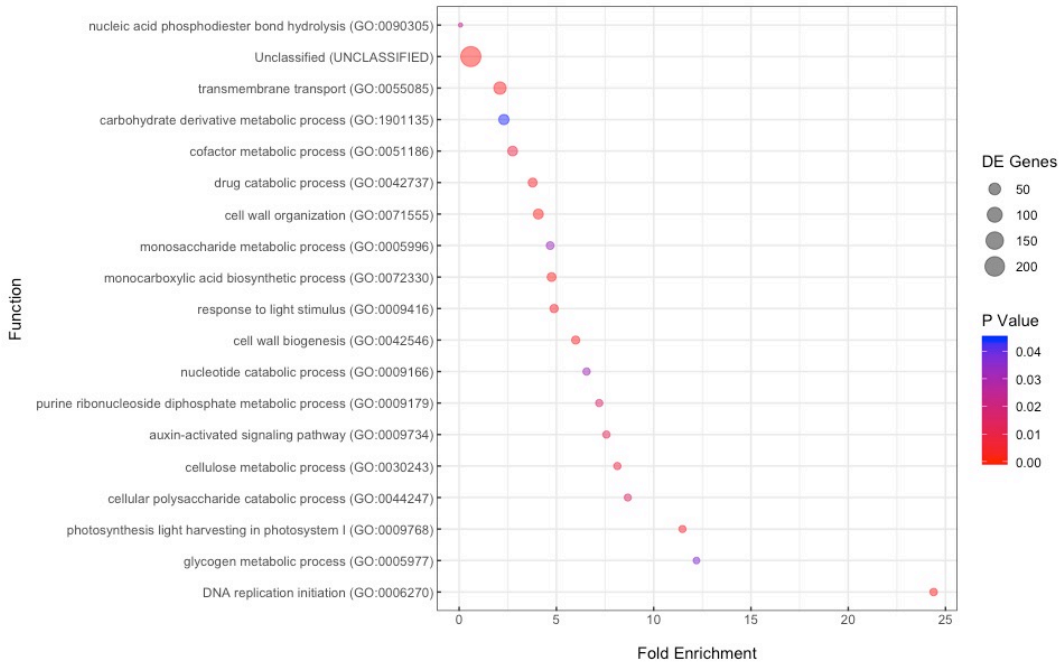


Figure 2.23. The expression heatmap for the differentially expressed genes in breaker stage vs stage 3 in *S. pimpinellifolium*.

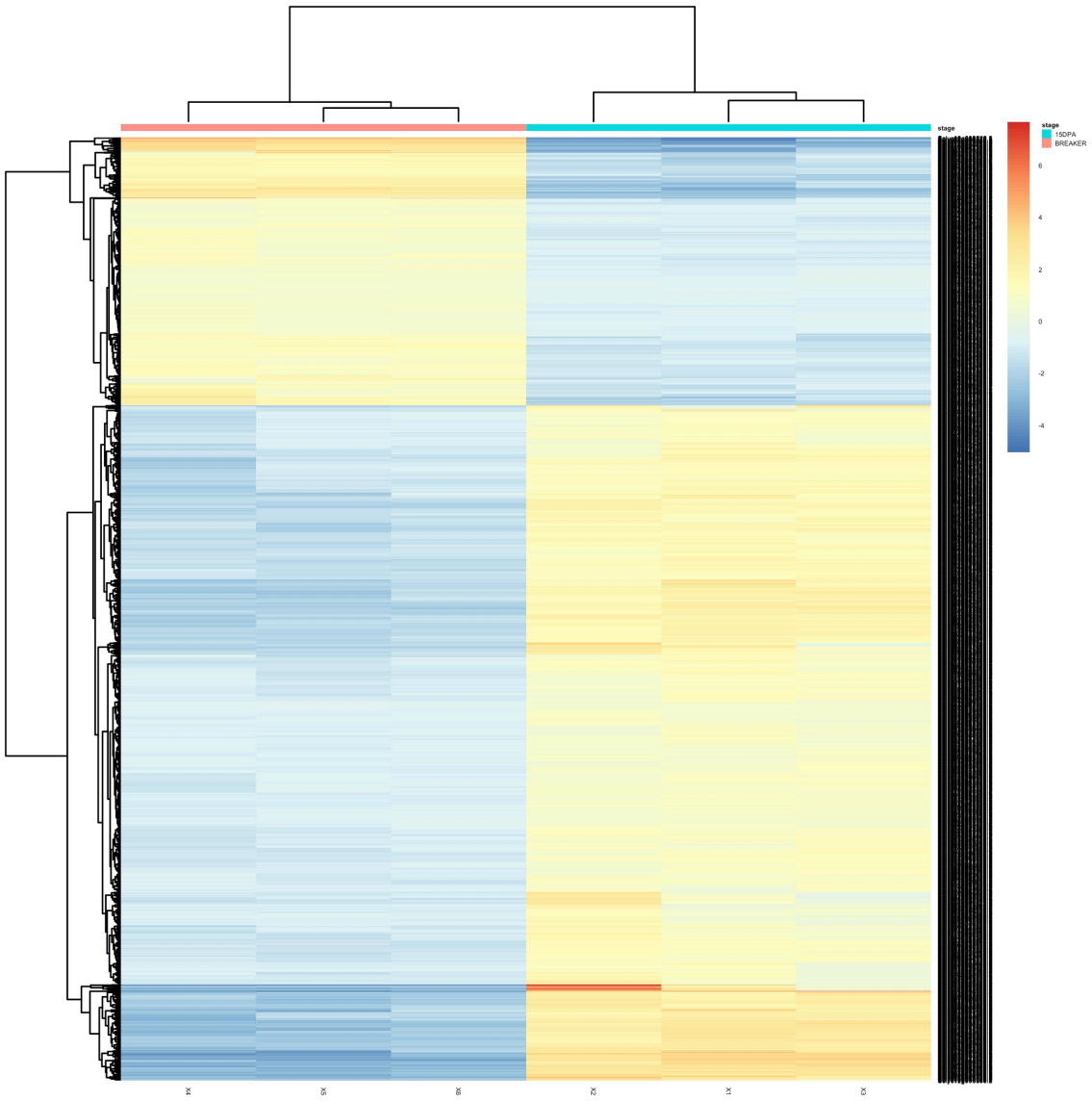


Figure 2.24. The enrichment of gene ontology terms for the downregulated genes in breaker stage vs stage 3 in *S. pimpinellifolium*.

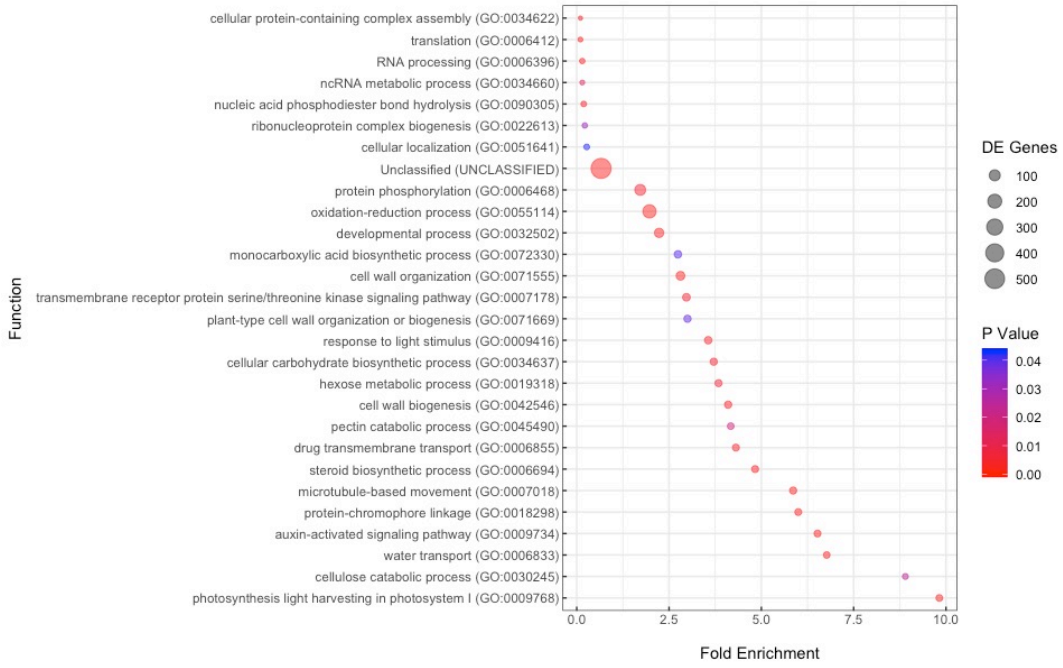


Figure 2.25. The enrichment of gene ontology terms for the upregulated genes in breaker stage vs stage 3 in *S. pimpinellifolium*.

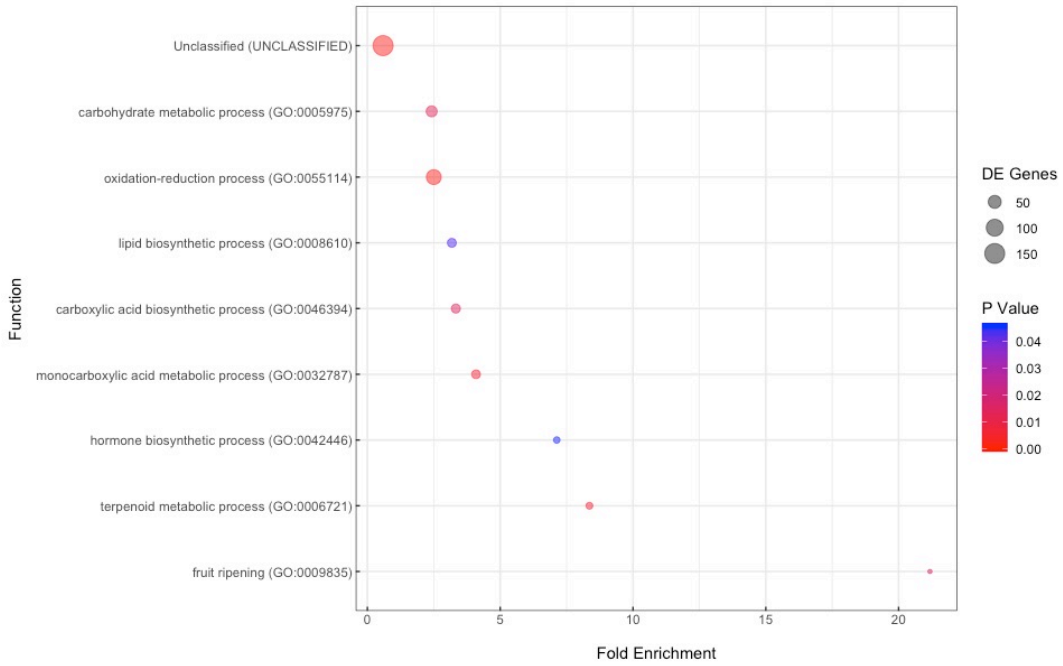


Figure 2.26. The expression heatmap for the differentially expressed genes in stage 4 vs breaker stage in *S. pimpinellifolium*.

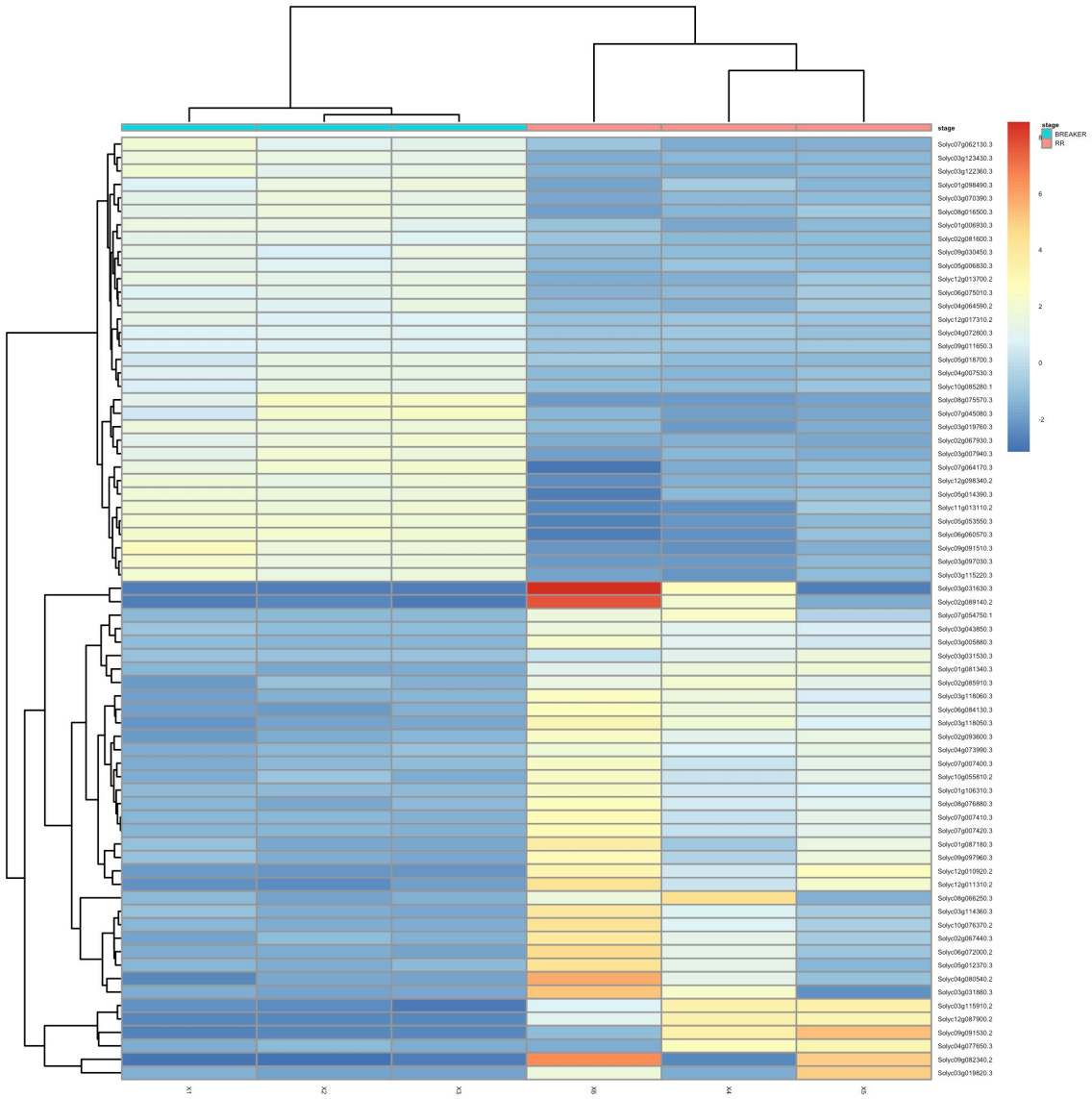


Figure 2.27. The enrichment of gene ontology terms for the downregulated genes in stage 4 vs breaker stage in *S. pimpinellifolium*.

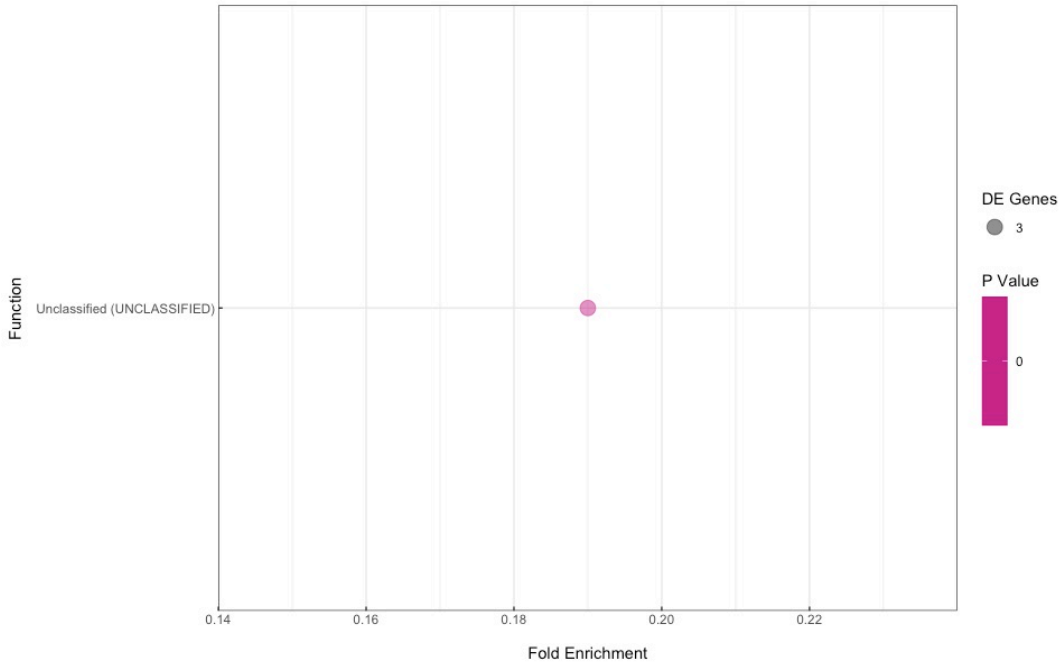


Figure 2.28. The enrichment of gene ontology terms for the upregulated genes in stage 4 vs breaker stage in *S. pimpinellifolium*.

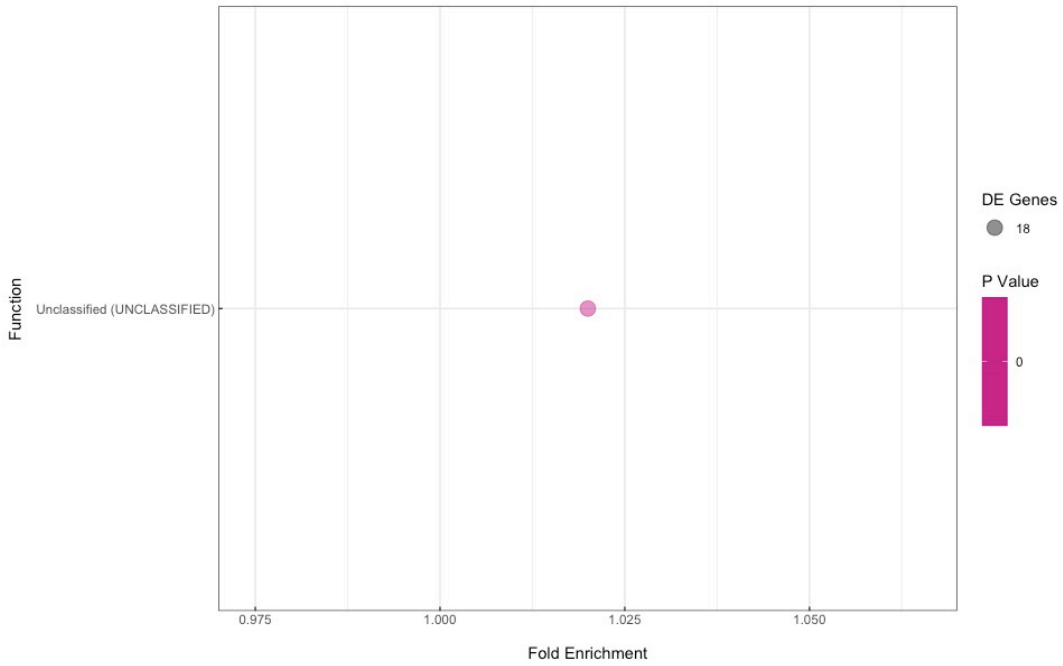


Figure 2.29. The expression heatmap for the differentially expressed genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 1.

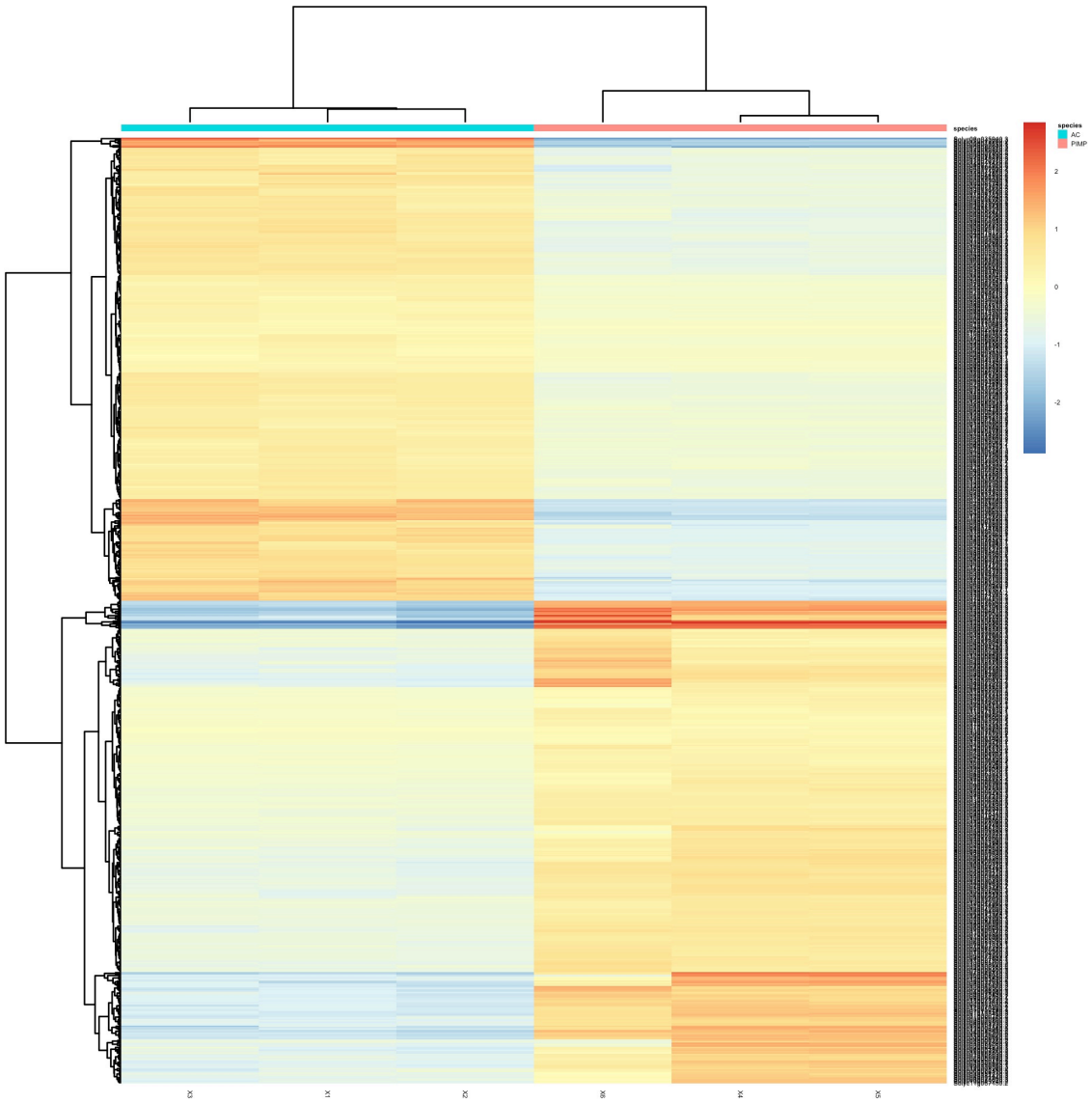


Figure 2.30. The enrichment of gene ontology terms for the downregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 1.

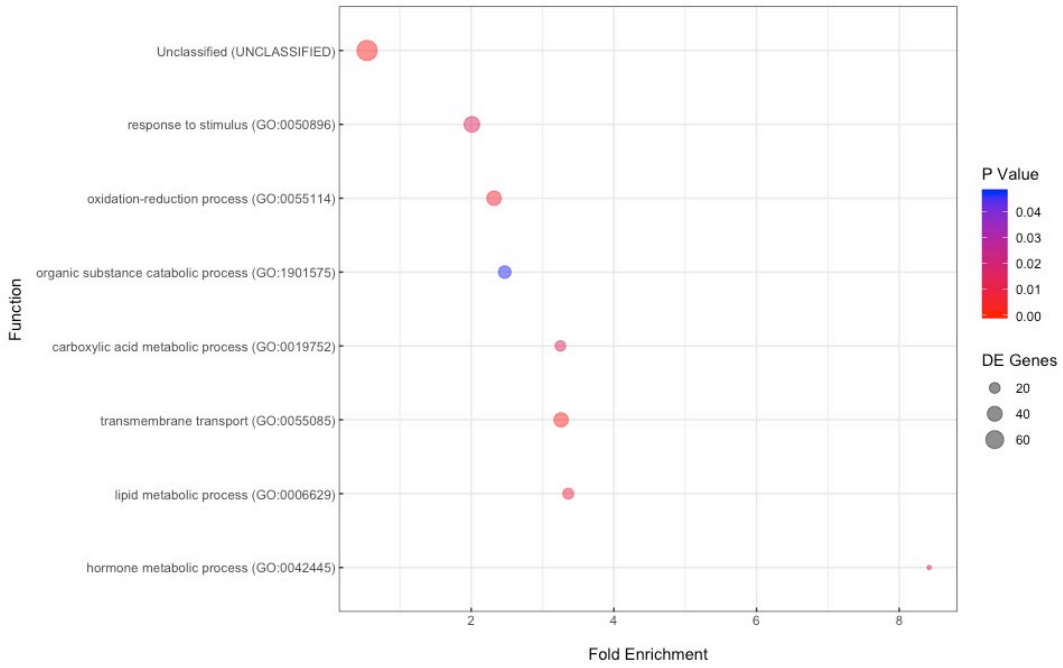


Figure 2.31. The enrichment of gene ontology terms for the upregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 1.

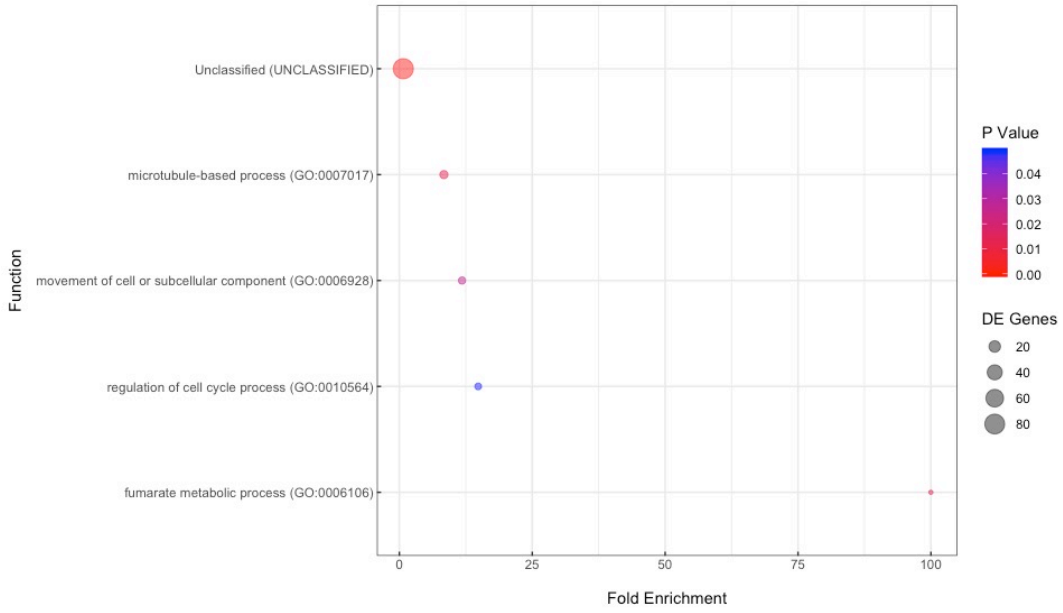


Figure 2.32. The expression heatmap for the differentially expressed genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 2.

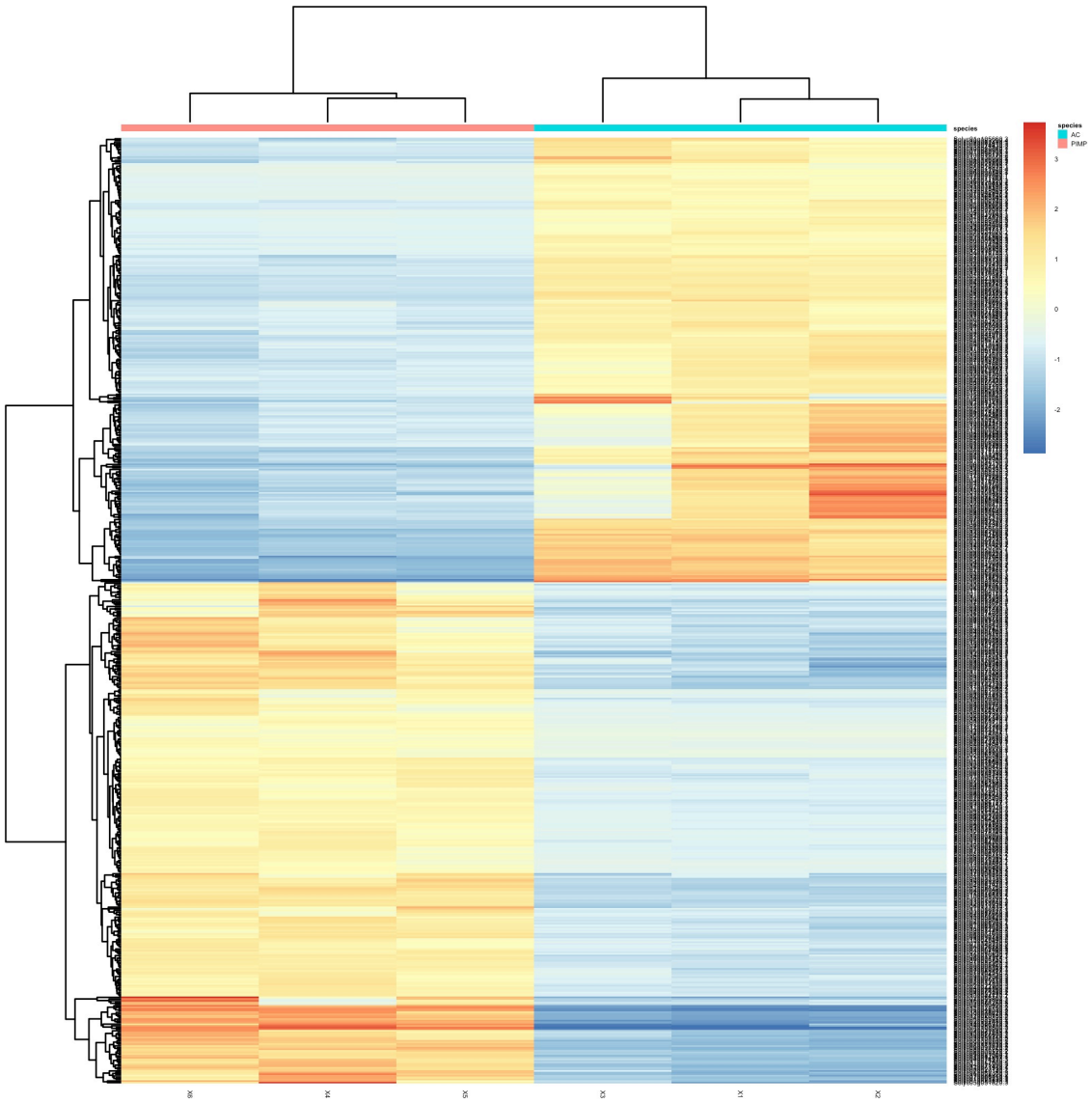


Figure 2.33. The enrichment of gene ontology terms for the downregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 2.

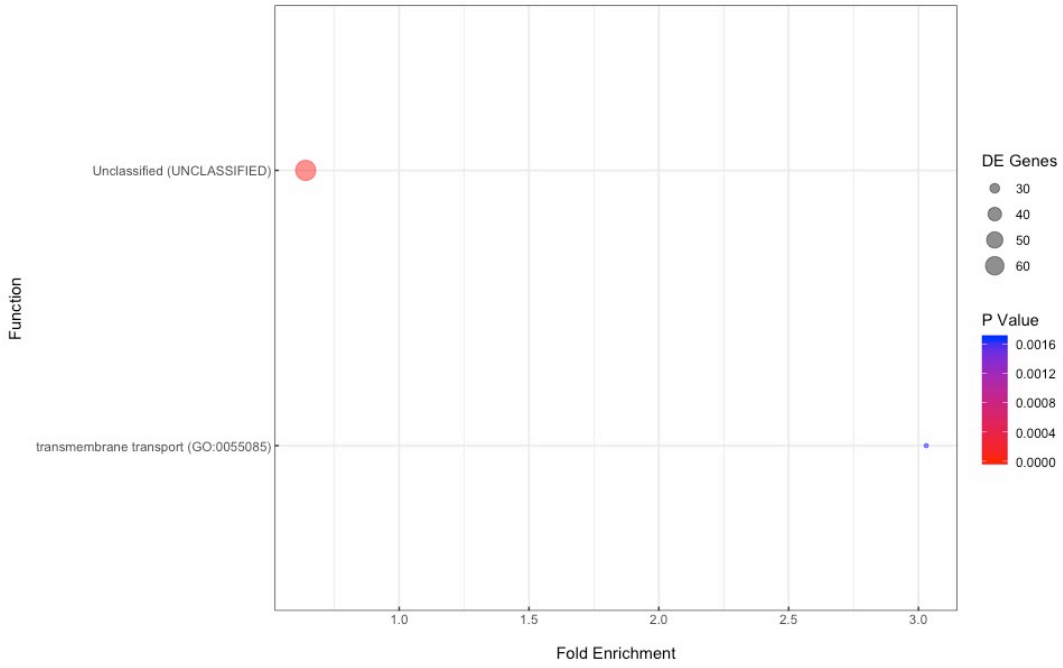


Figure 2.34. The enrichment of gene ontology terms for the upregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 2.

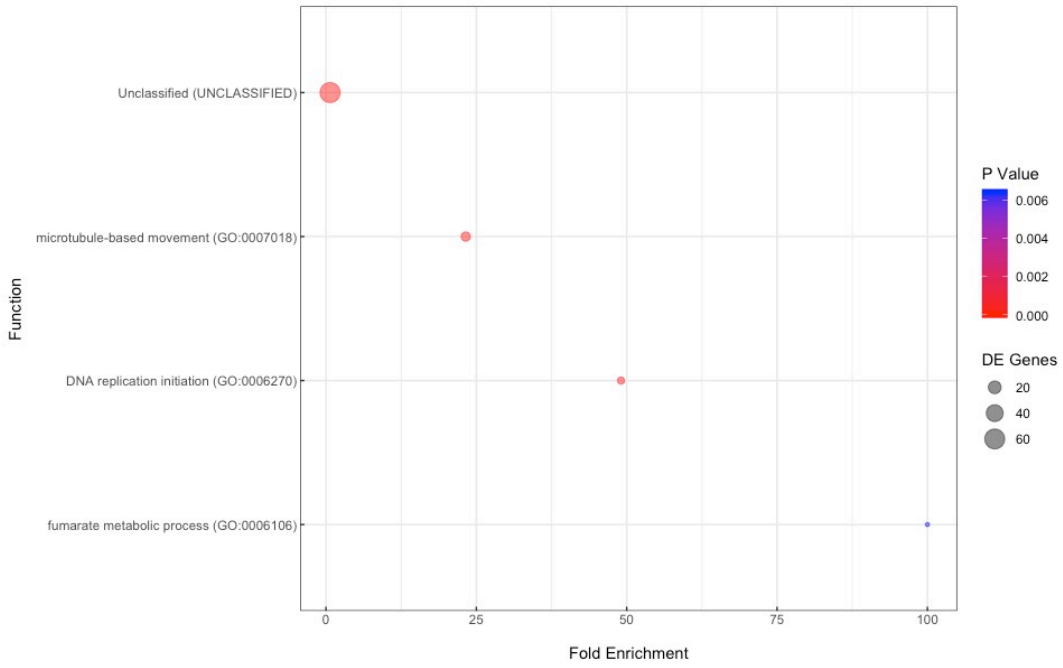


Figure 2.35. The expression heatmap for the differentially expressed genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 3.

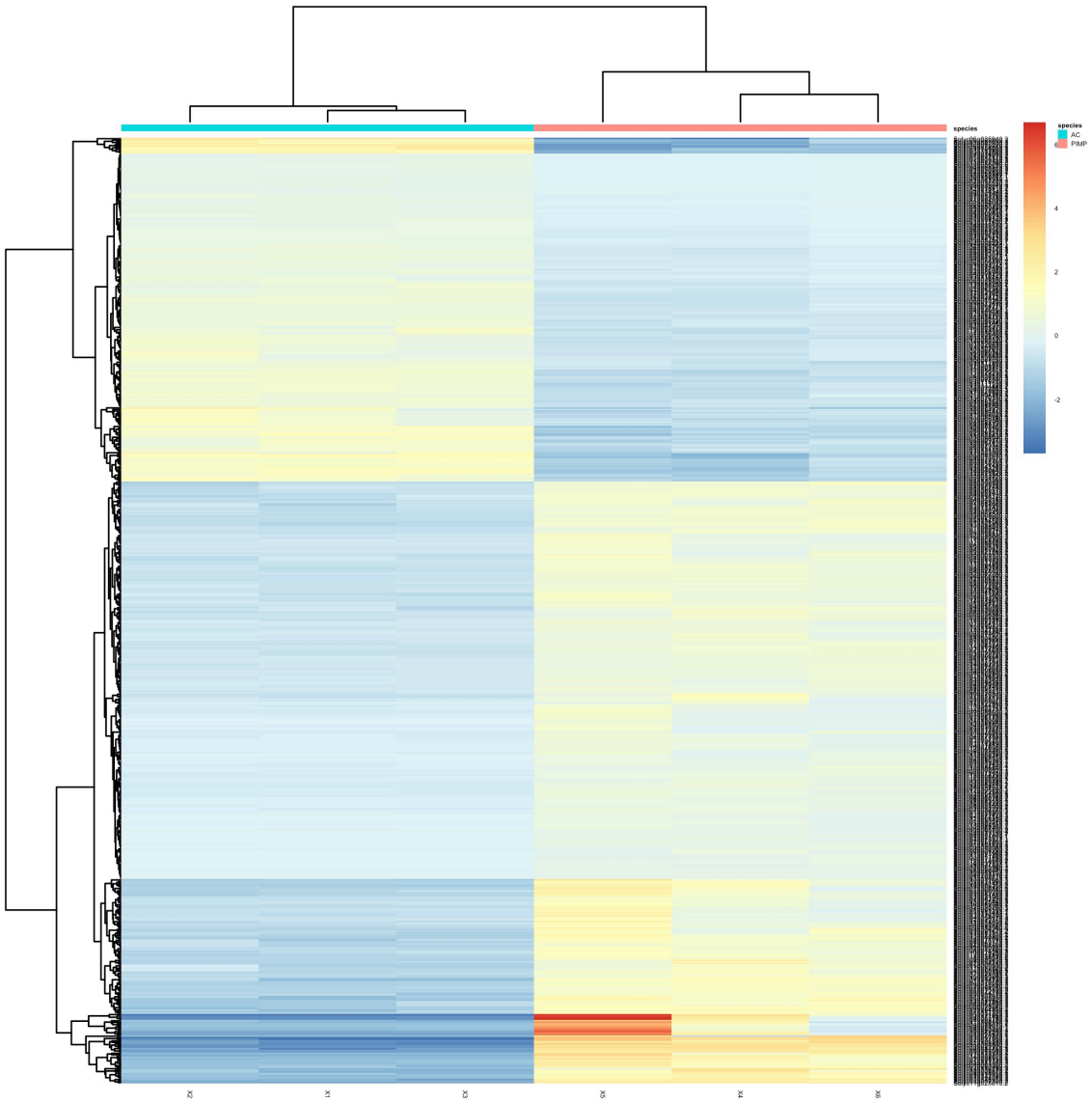


Figure 2.36. The enrichment of gene ontology terms for the downregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 3.

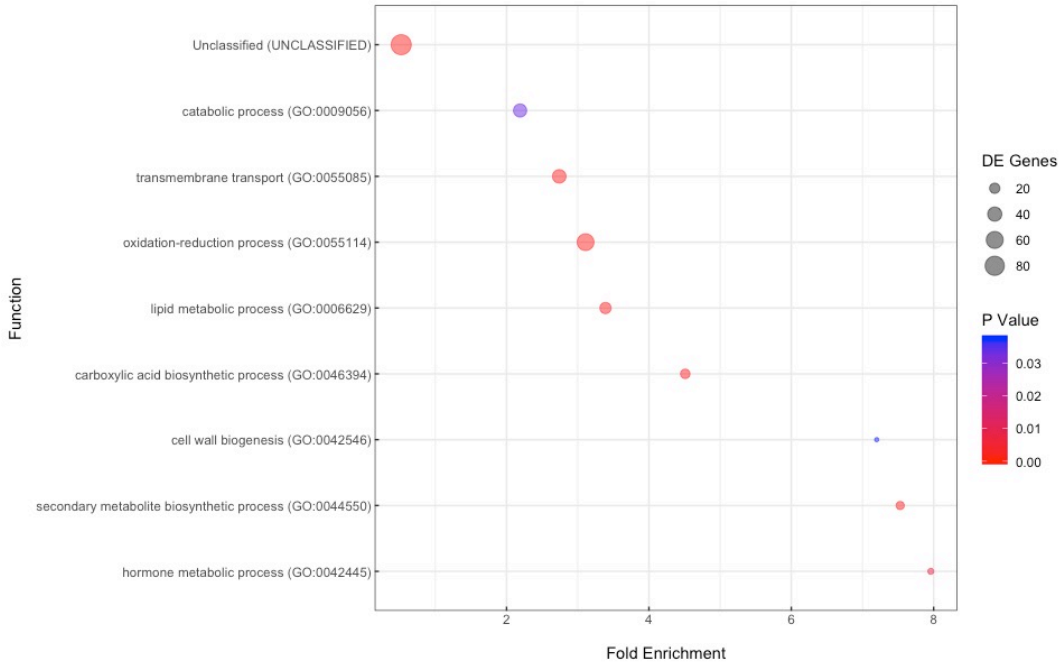


Figure 2.37. The enrichment of gene ontology terms for the upregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in stage 3.

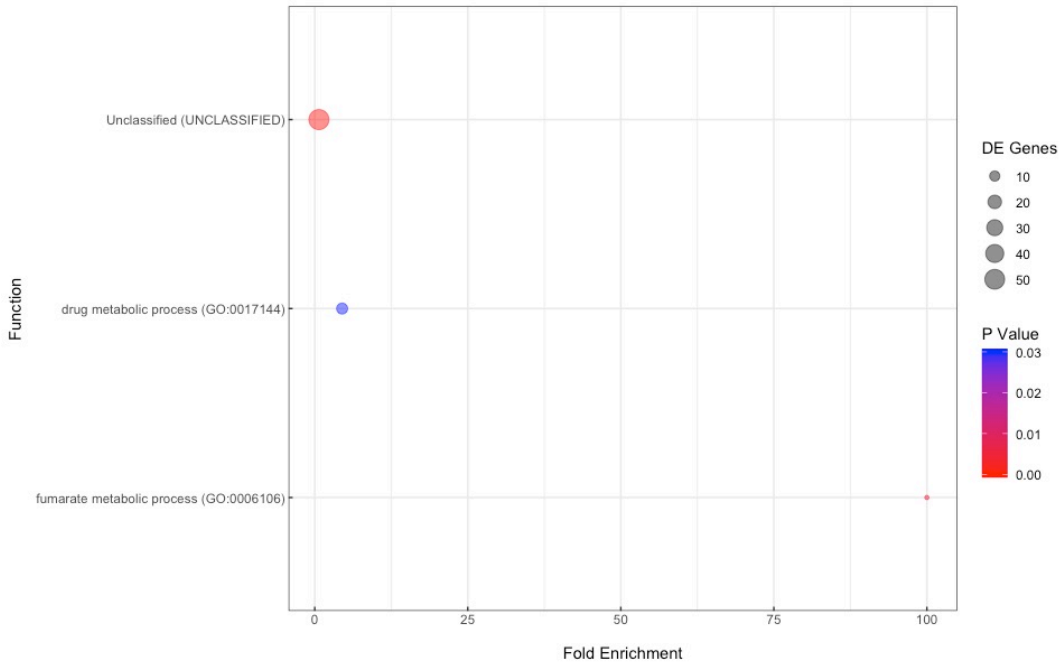


Figure 2.38. The expression heatmap for the differentially expressed genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in breaker fruit.

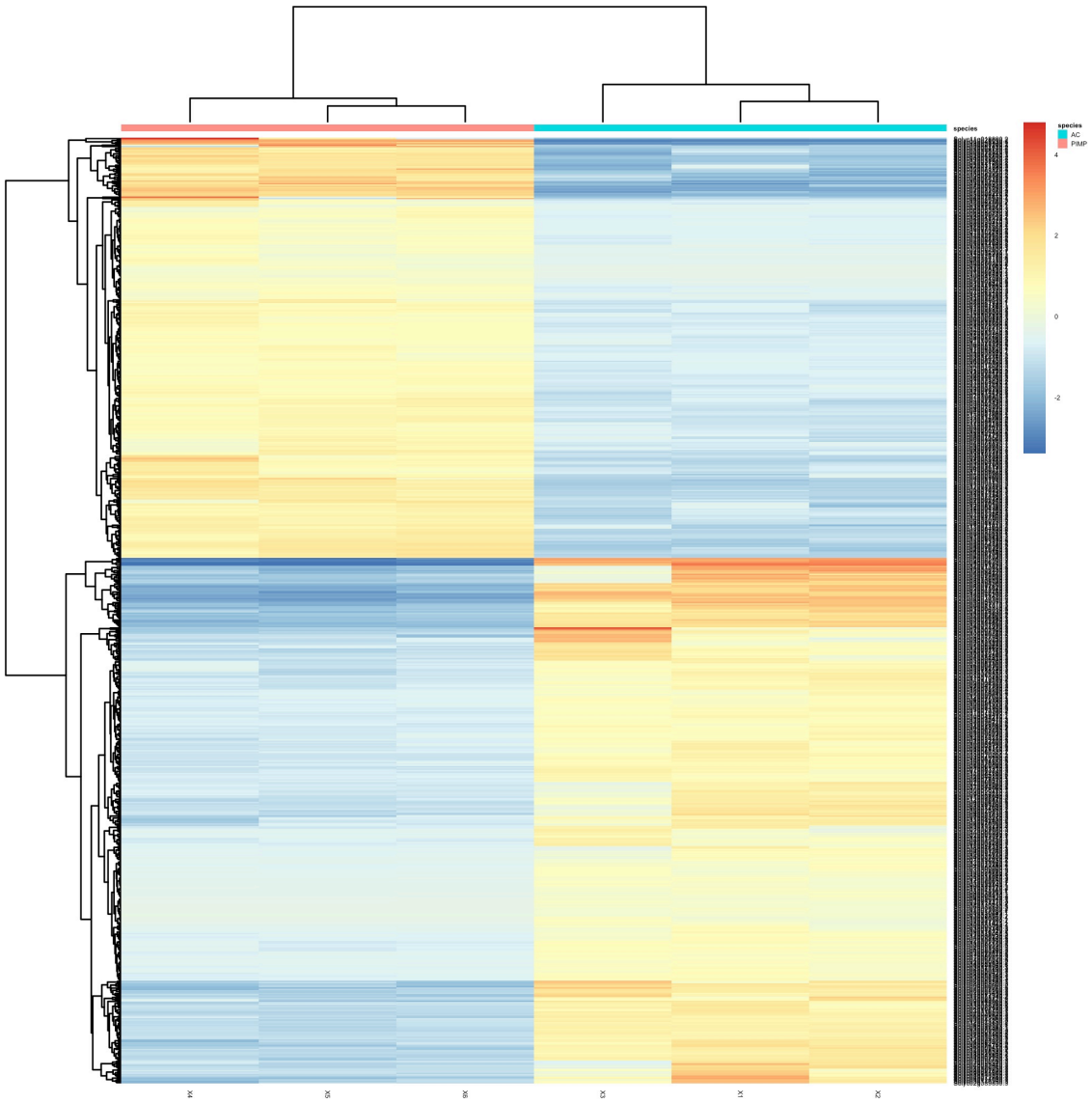


Figure 2.39. The enrichment of gene ontology terms for the downregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in breaker fruit.

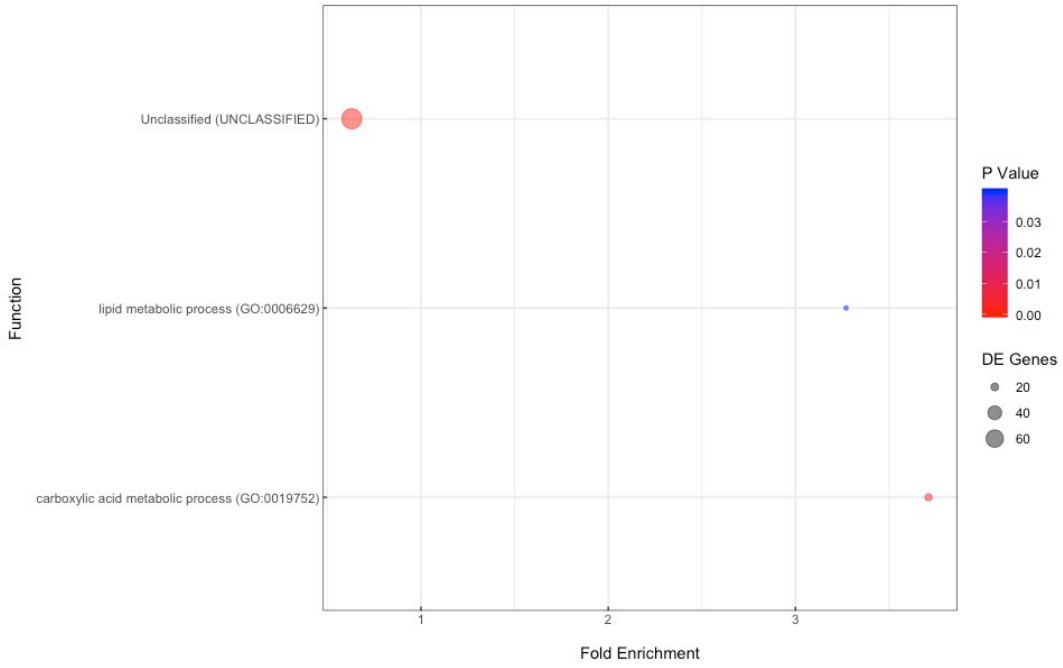


Figure 2.40. The enrichment of gene ontology terms for the upregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in breaker fruit.

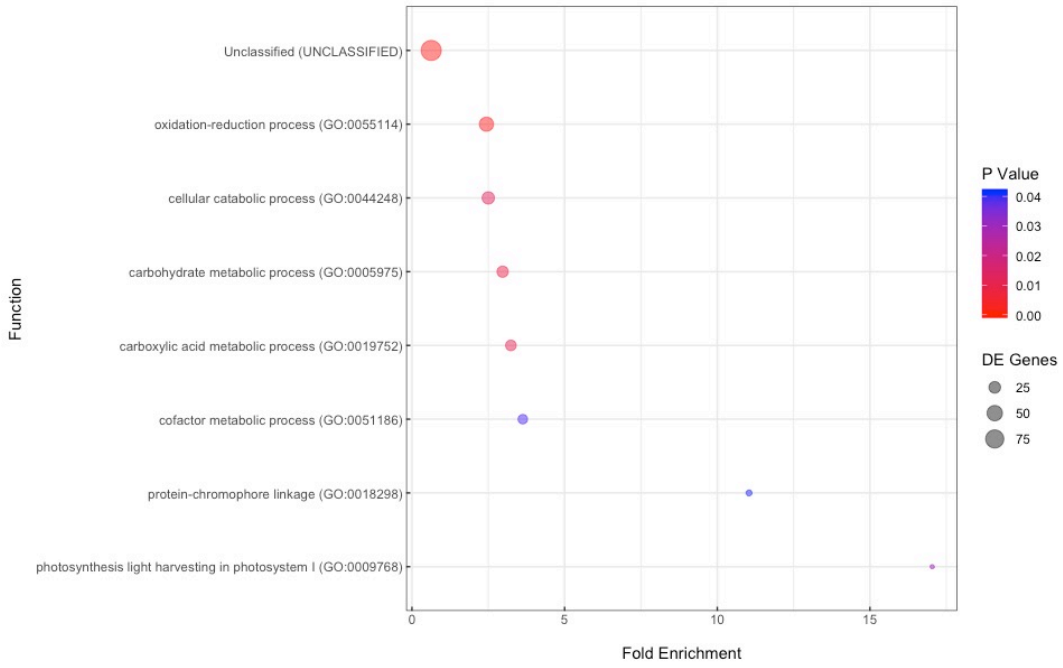


Figure 2.41. The expression heatmap for the differentially expressed genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in red ripe fruit.

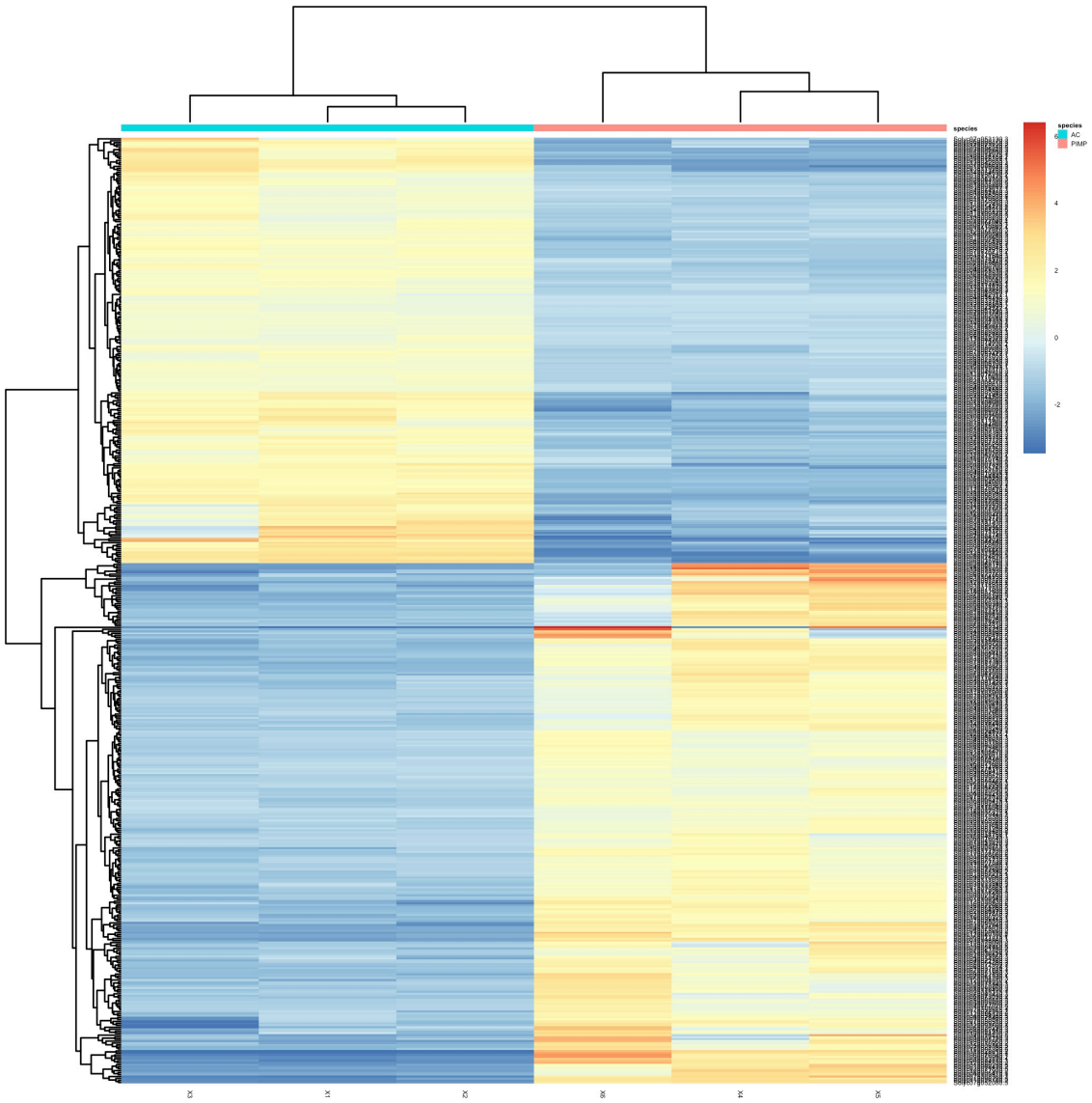


Figure 2.42. The enrichment of gene ontology terms for the downregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in red ripe fruit.

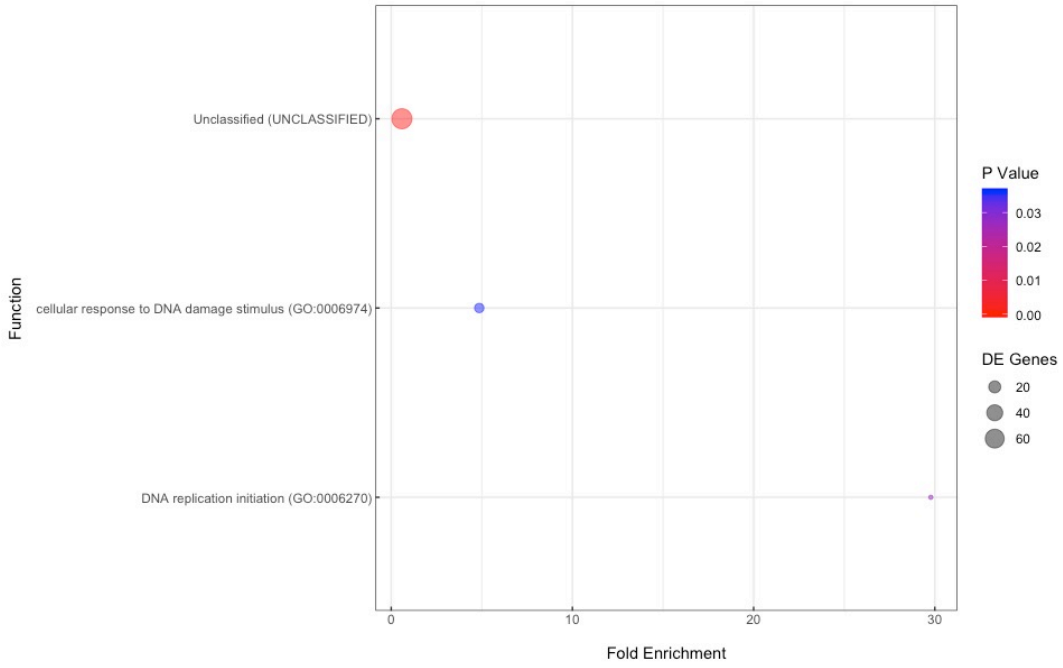


Figure 2.43. The enrichment of gene ontology terms for the upregulated genes in *S. lycopersicum* cv Ailsa Craig vs *S. pimpinellifolium* in red ripe fruit.

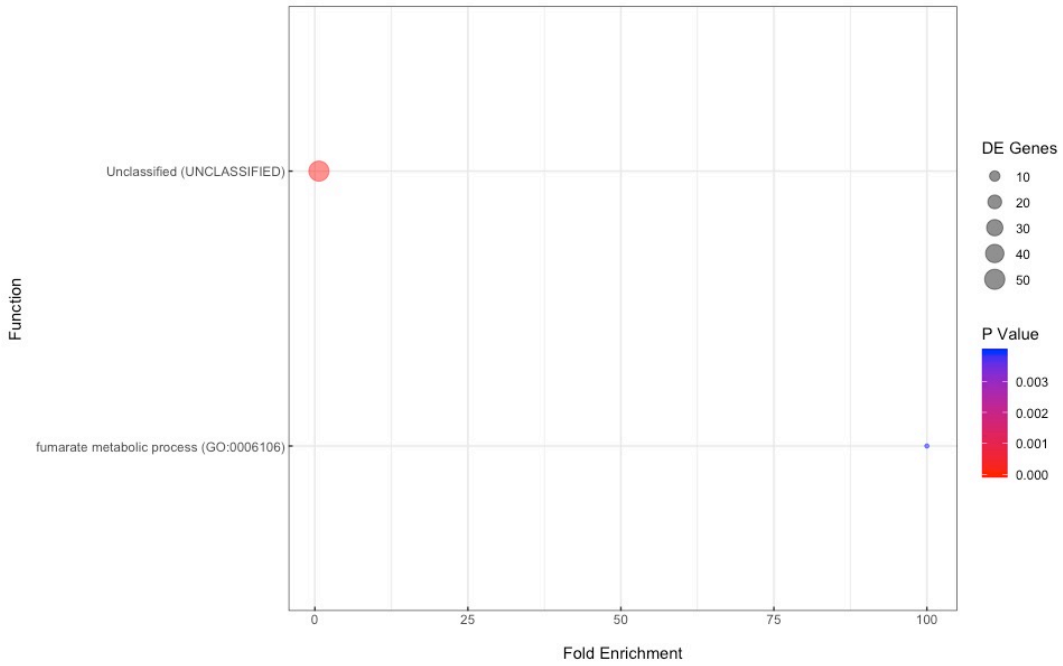


Figure 2.44. Stained cross-sections of stages 1, 2 and 3 of ovaries and fruit in *S. lycopersicum* cv Ailsa Craig (top row) and *S. pimpinellifolium* (bottom row).

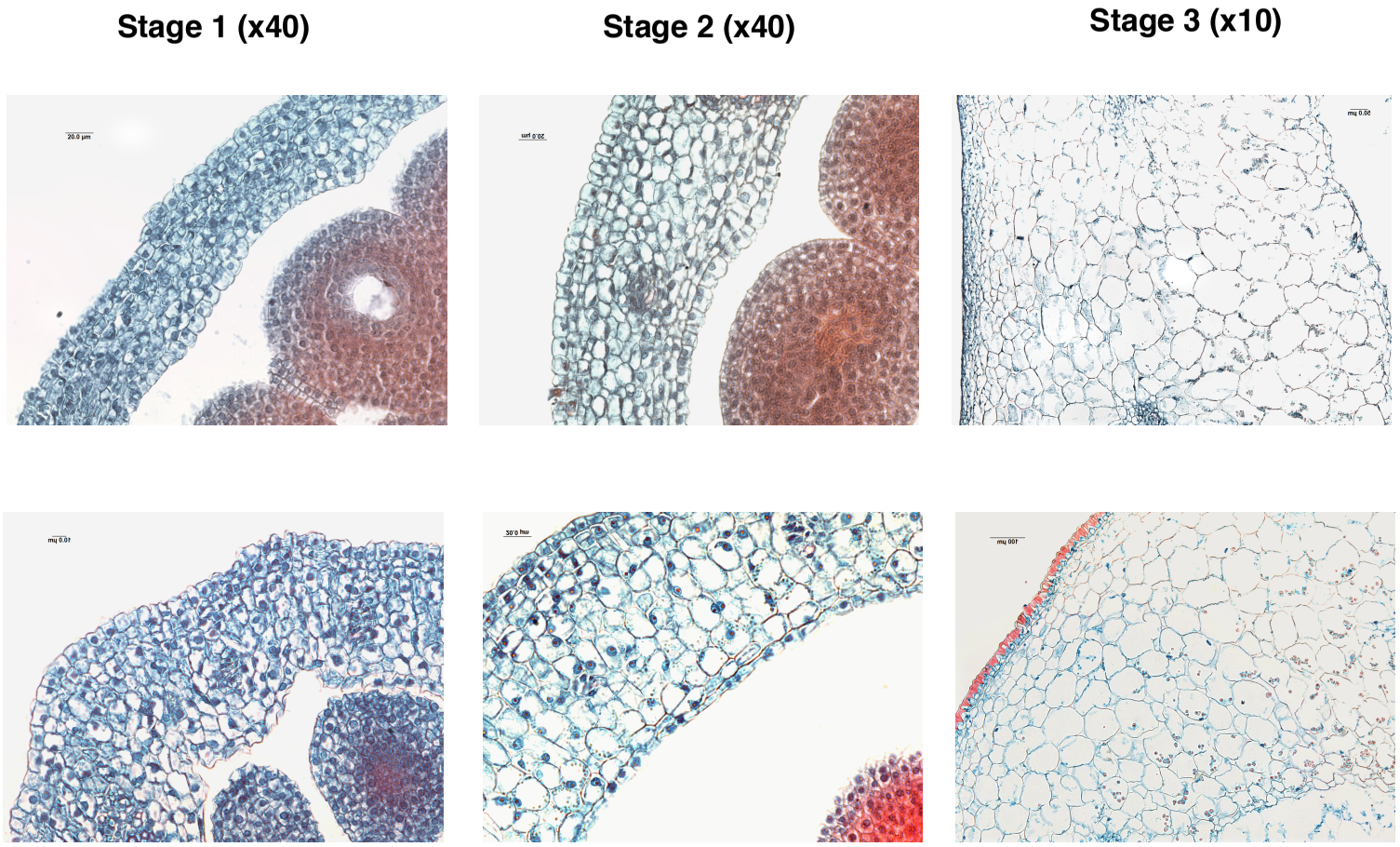
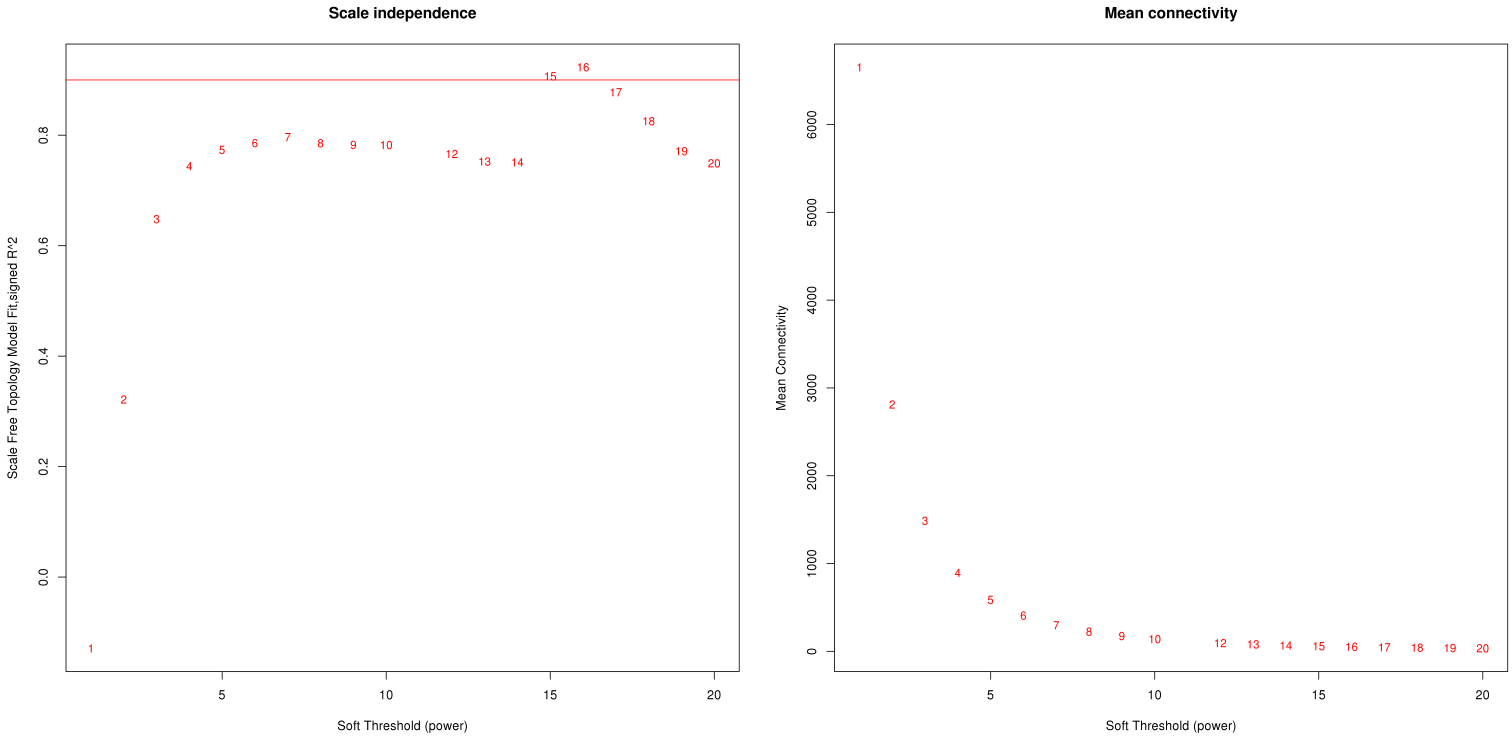


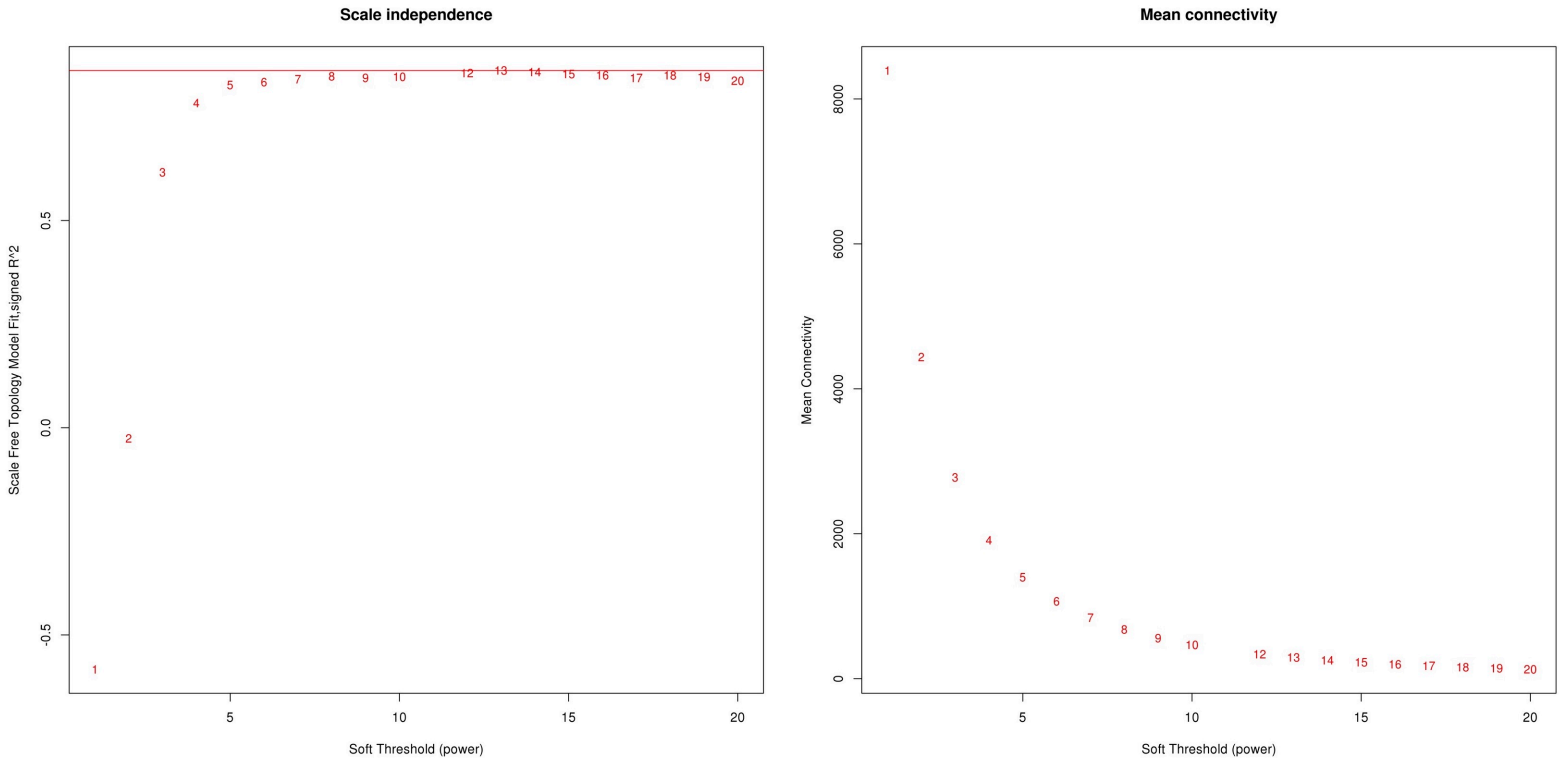
Figure 2.45. Scale independence and mean connectivity plots from WGCNA analysis. (a) All data combined (b) Only *S. lycopersicum* cv. Ailsa Craig (c) *Only S. pimpinellifolium*.

(a)



Continuation of Figure 2.45.

(b)



Continuation of Figure 2.45.

(c)

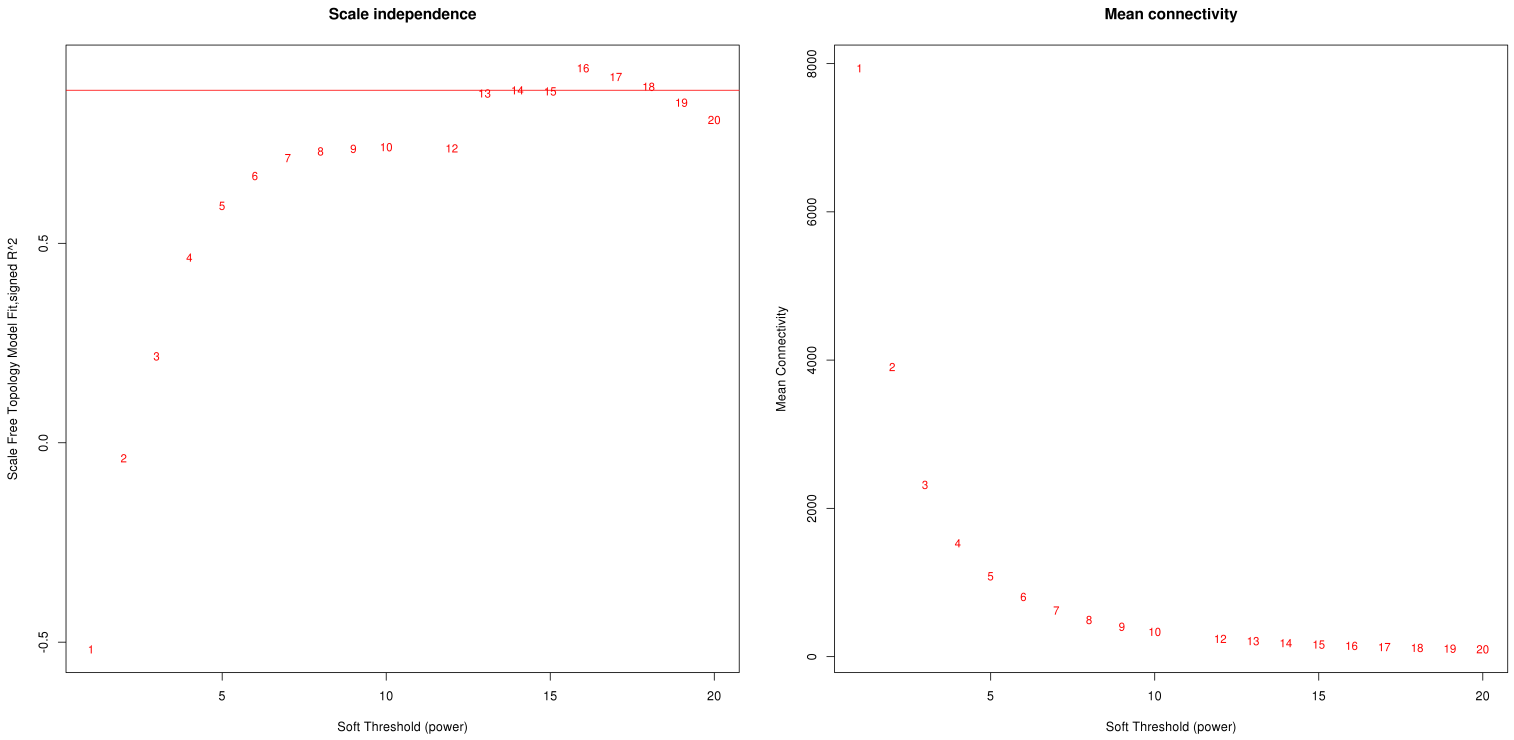


Figure 2.46. The expression patterns of *FRUITFULL* gene homologs in *S. lycopersicum* cv Ailsa Craig (top) and *S. pimpinellifolium* (bottom).

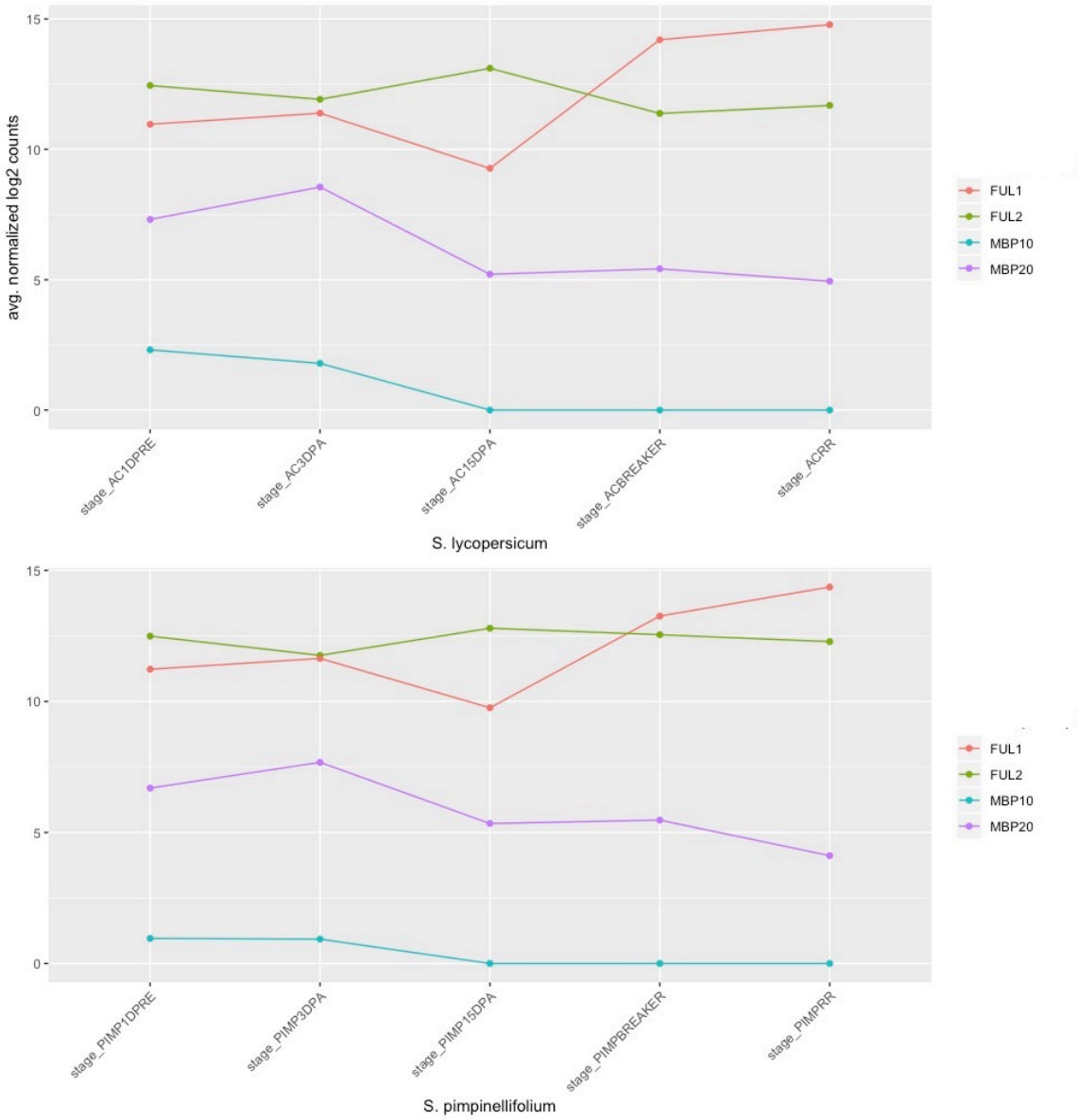
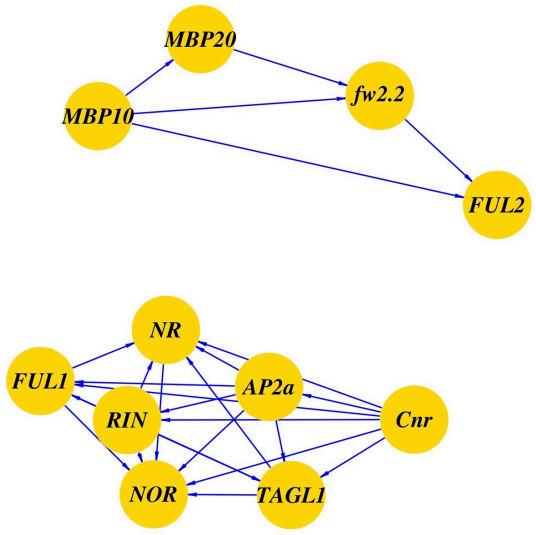
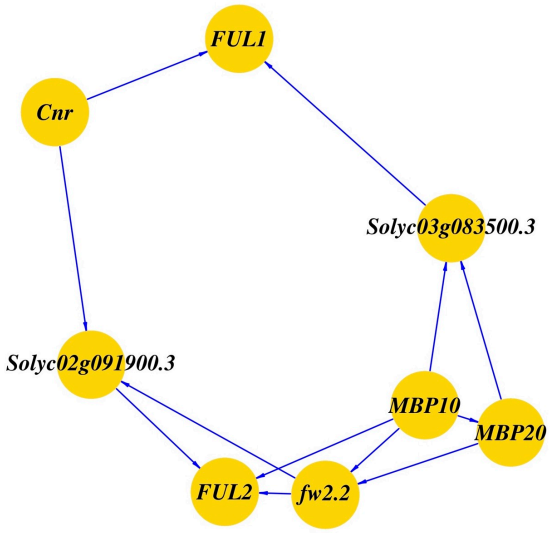


Figure 2.47. Hypothetical WGCNA co-expression networks of genes with established roles in fruit development. (a) The network for *S. lycopersicum* cv. Ailsa Craig without intermediary genes (b) The network for *S. lycopersicum* cv. Ailsa Craig with intermediary genes (only *Cnr* and *SlFUL1* are shown from the core-ripening module) (c) The network for *S. pimpinellifolium* without intermediary genes (d) The network for *S. pimpinellifolium* with intermediary genes (only *Cnr* and *SlFUL1* are shown from the core-ripening module).

(a)

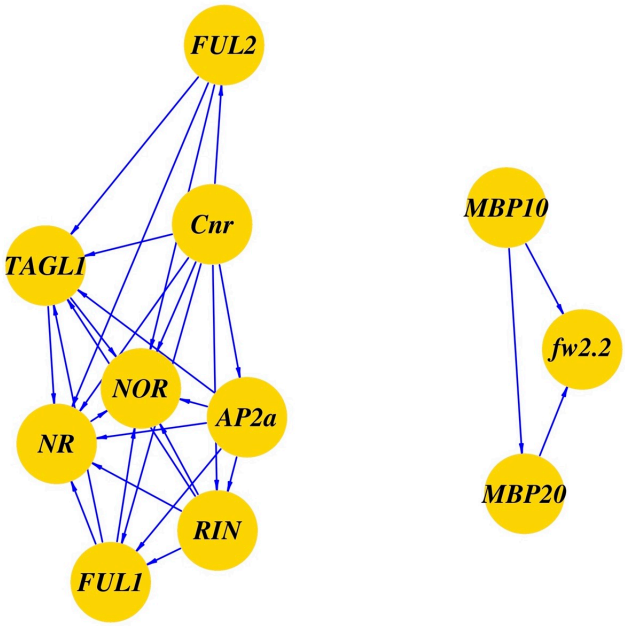


(b)



Continuation of Figure 2.47.

(c)



(d)

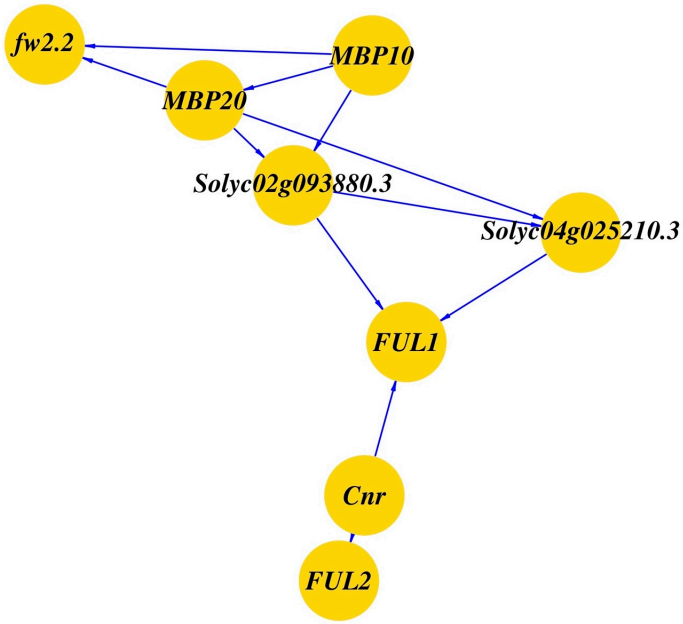
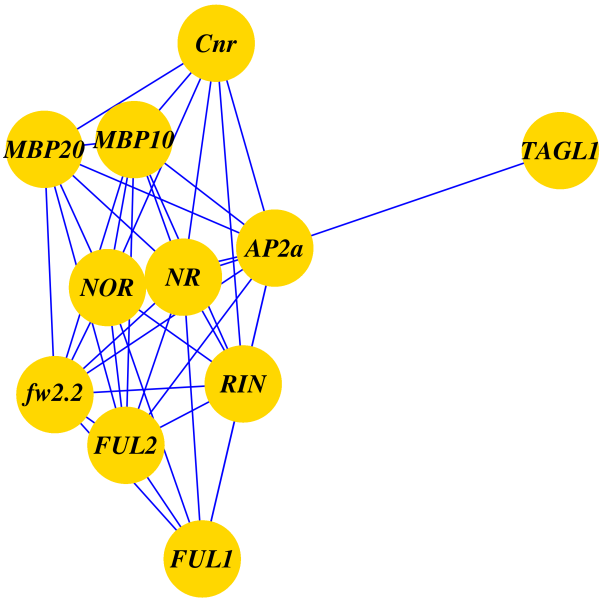


Figure 2.48. Hypothetical co-expression networks of genes with established roles in fruit development using the Sinha Lab method. (a) The network for *S. lycopersicum* cv. Ailsa Craig (b) The network for *S. pimpinellifolium*.

(a)



(b)

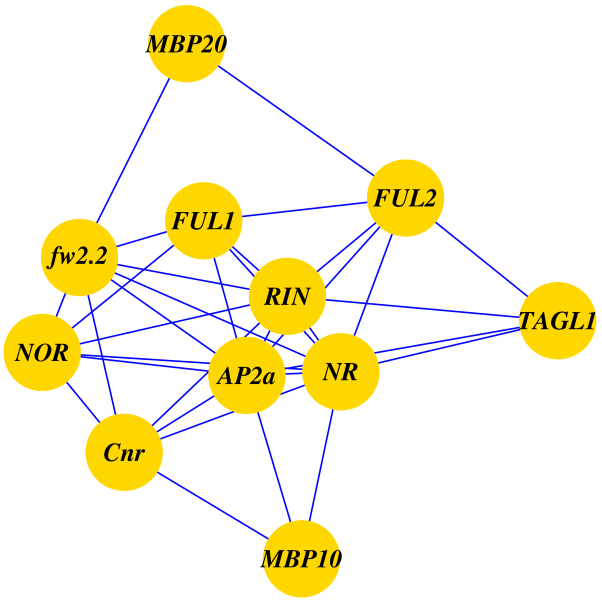
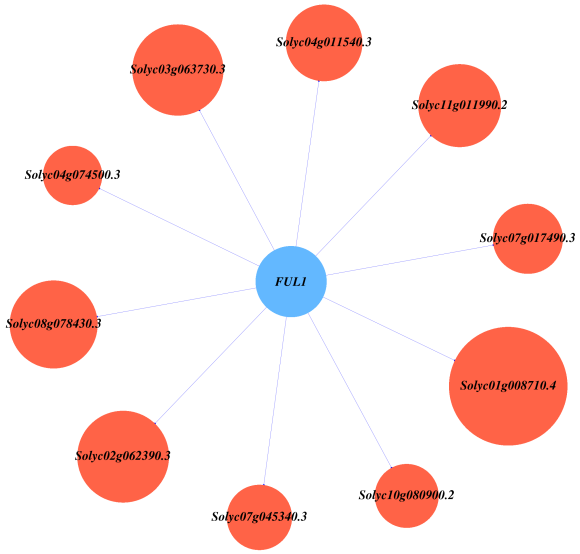
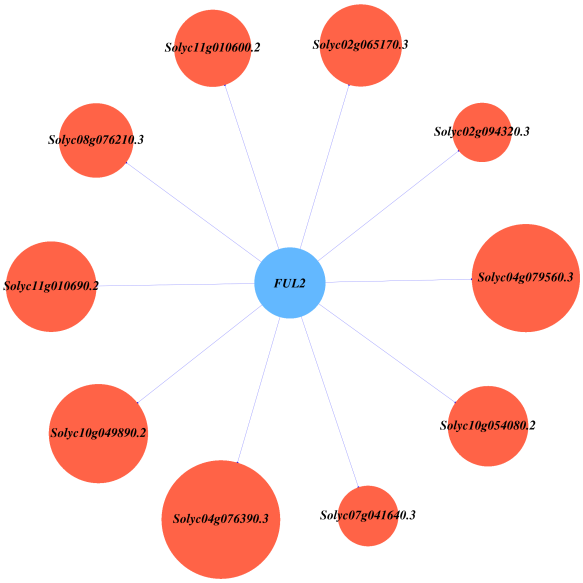


Figure 2.49. Hypothetical WGCNA networks of *FUL* orthologs and their 20 most closely connected genes in *S. lycopersicum* cv. Ailsa Craig and *S. pimpinellifolium*. AC= *S. lycopersicum* cv. Ailsa Craig; Sp= *S. pimpinellifolium*. (a) AC *SlFUL1* (b) AC *SlFUL2* (c) AC *MBP10* (d) AC *MBP20* (e) *SpFUL1* (f) *SpFUL2* (g) Sp *MBP10* (h) Sp *MBP20*

(a)

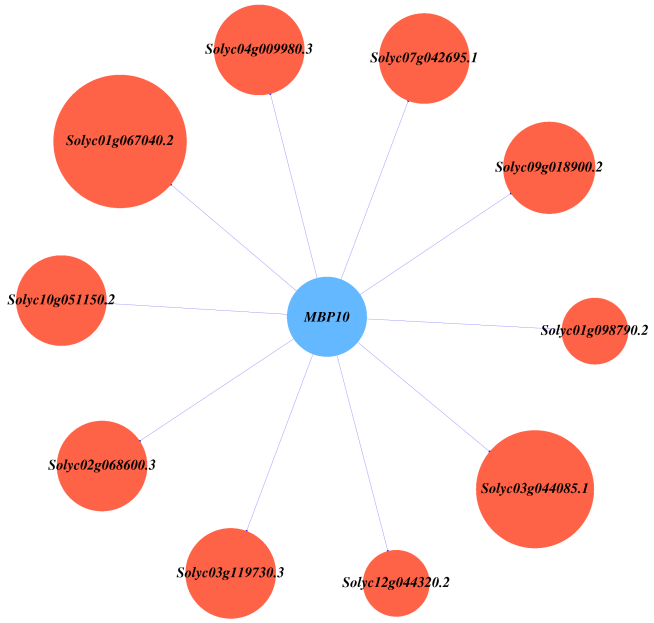


(b)

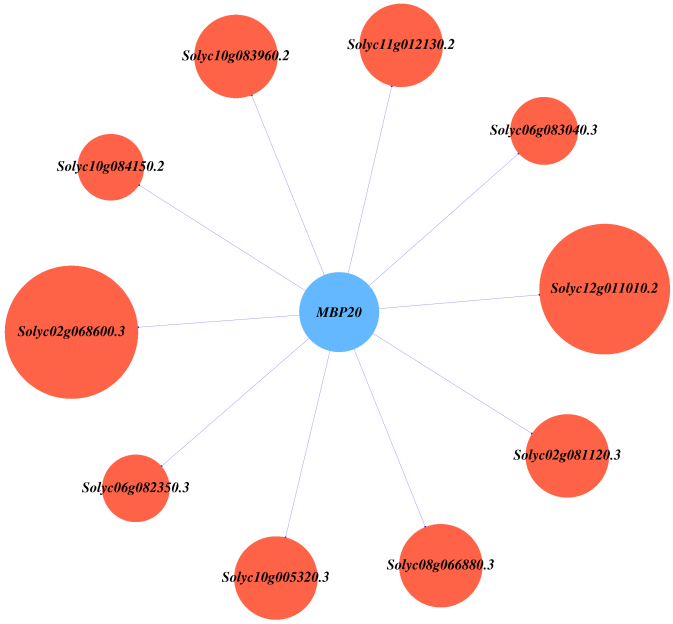


Continuation of Figure 2.49.

(c)

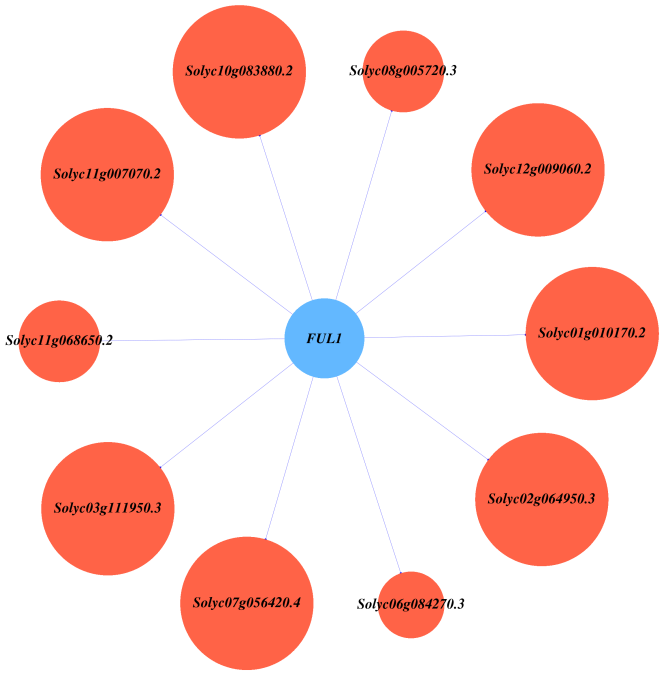


(d)

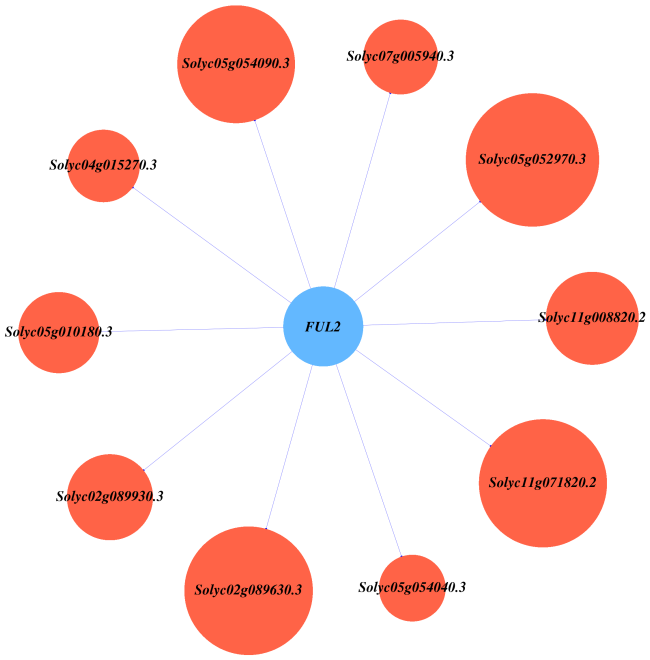


Continuation of Figure 2.49.

(e)



(f)



Continuation of Figure 2.49.

(g)

(h)

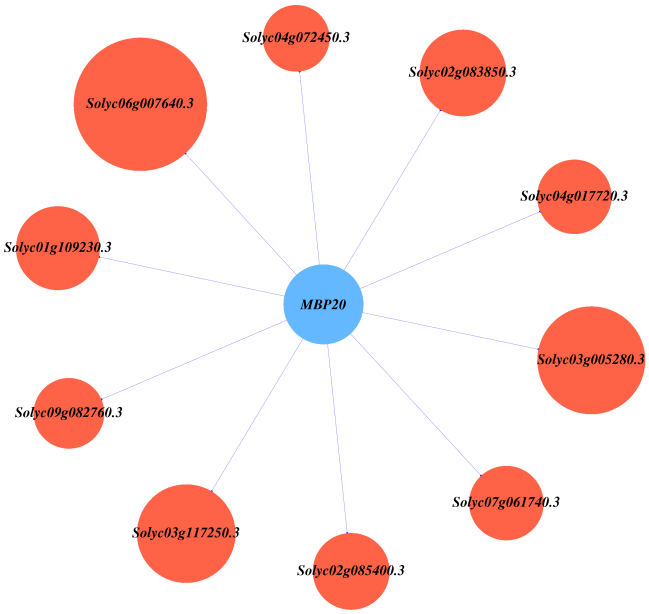
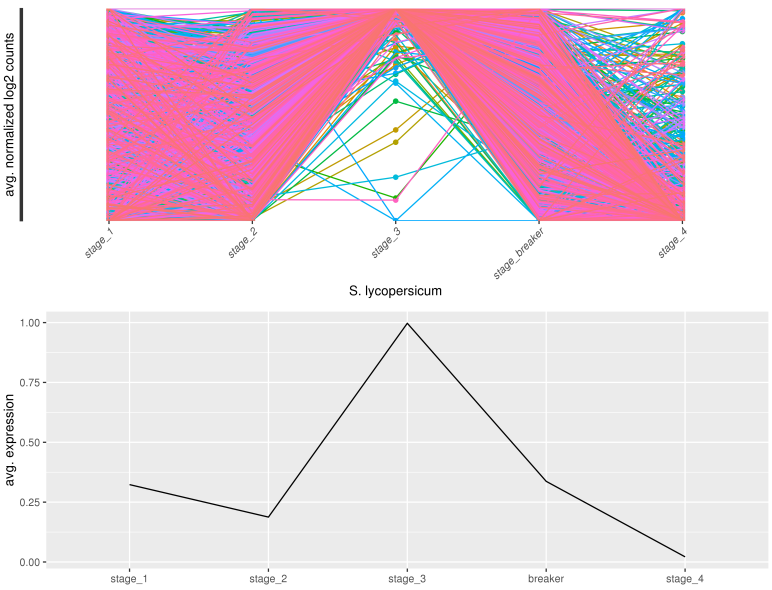
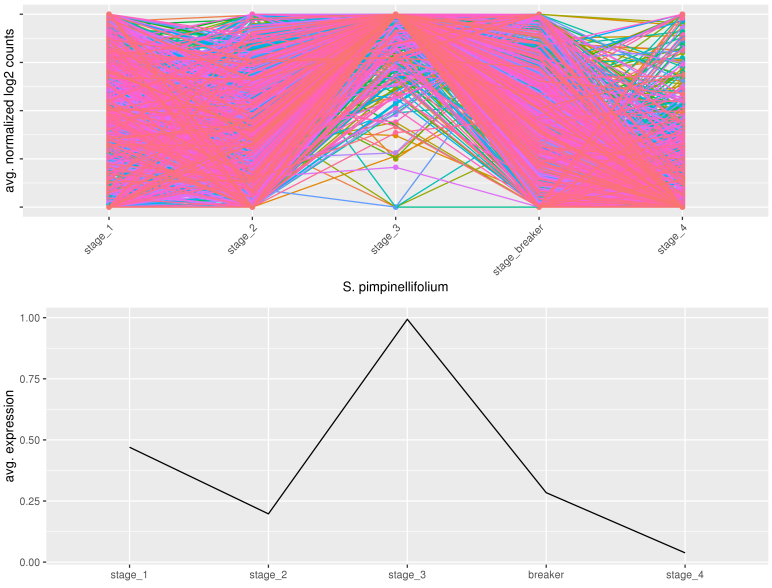


Figure 2.50. Expression patterns of the WGCNA co-expression modules that contain the fruit development-related genes discussed in this chapter. AC= *S. lycopersicum* cv. Ailsa Craig; Sp= *S. pimpinellifolium* (a) AC black (b) Sp black (c) AC blue (d) Sp blue (e) AC brown (f) Sp brown (g) AC pink (h) Sp pink (i) AC red (j) Sp red.

(a)

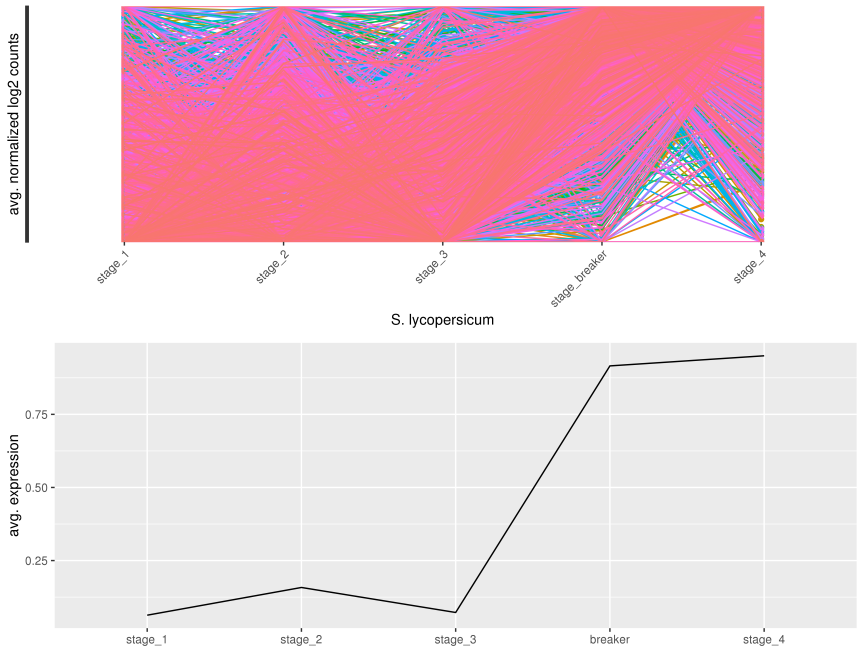


(b)

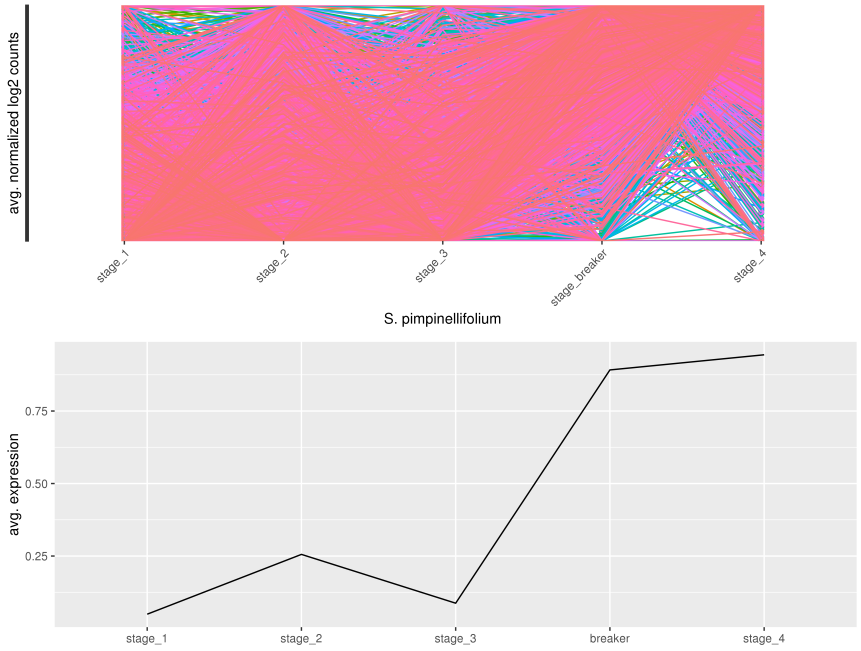


Continuation of Figure 2.50.

(c)

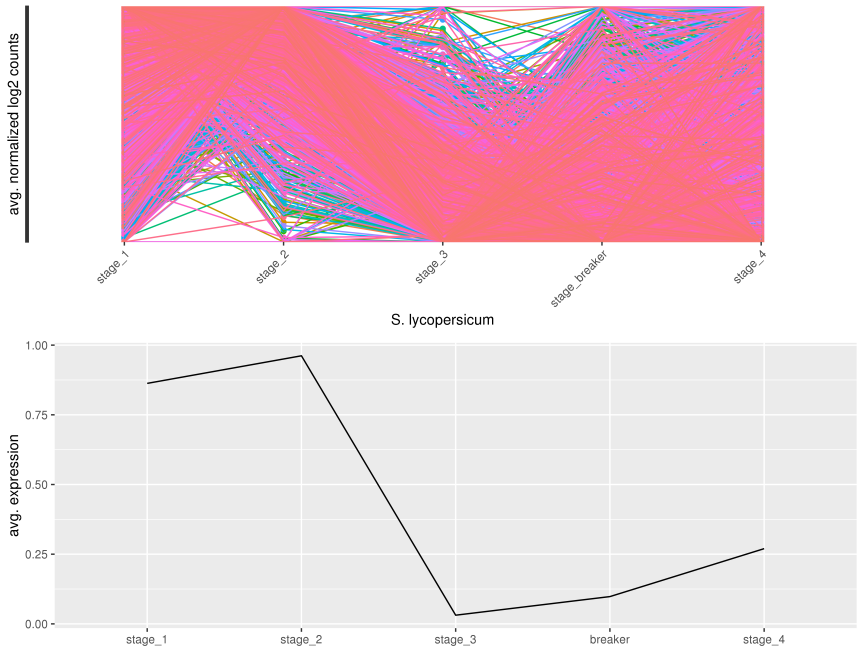


(d)

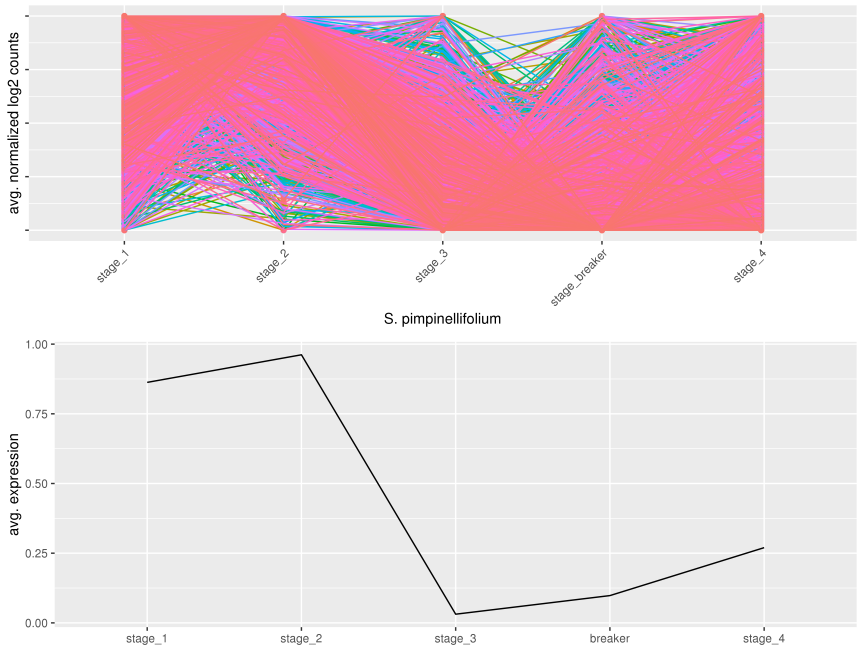


Continuation of Figure 2.50.

(e)

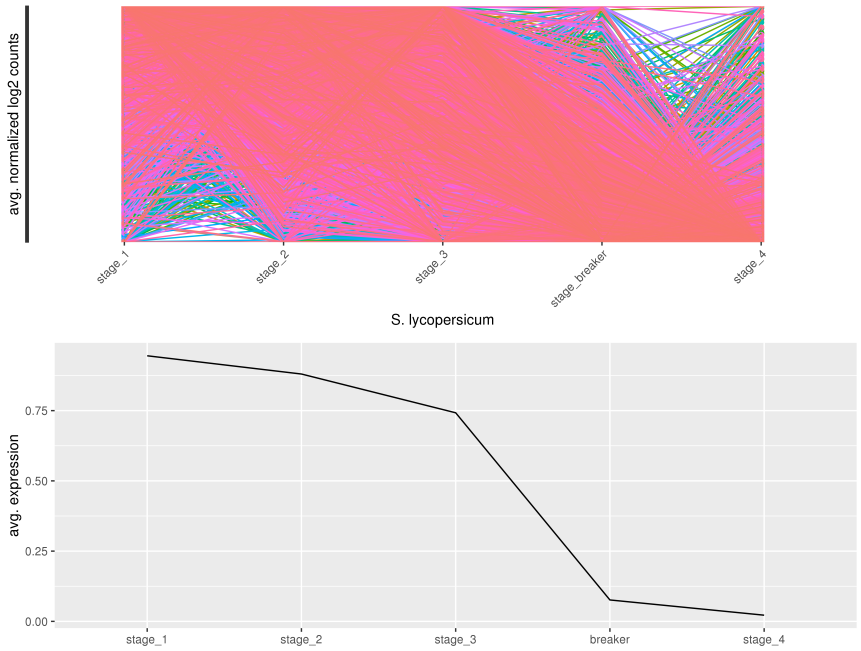


(f)

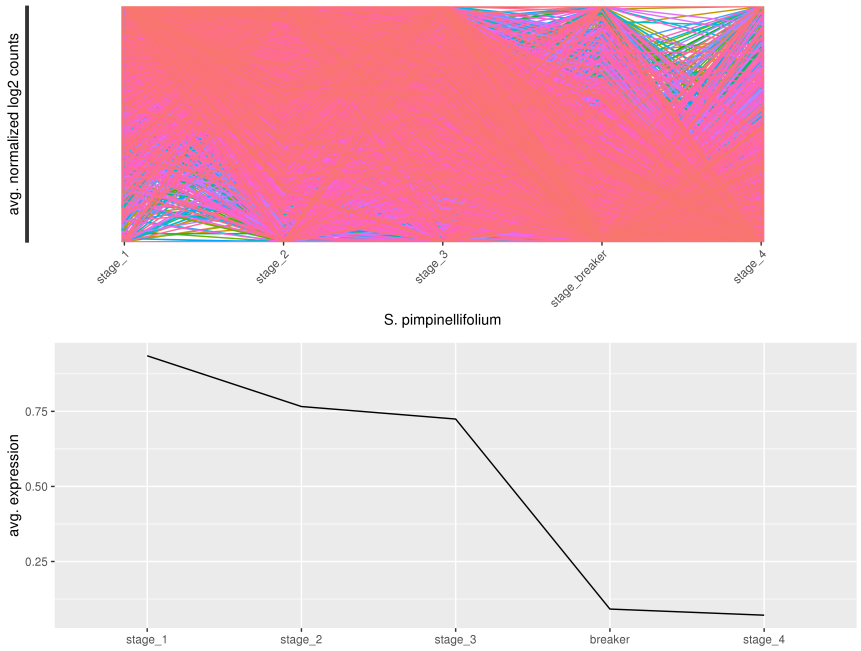


Continuation of Figure 2.50.

(g)

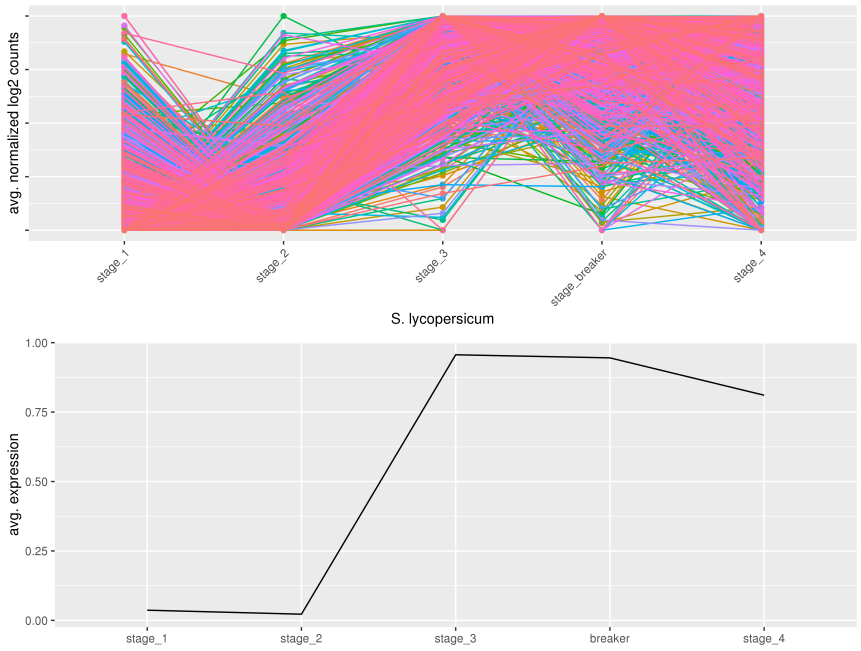


(h)



Continuation of Figure 2.50.

(i)



(j)

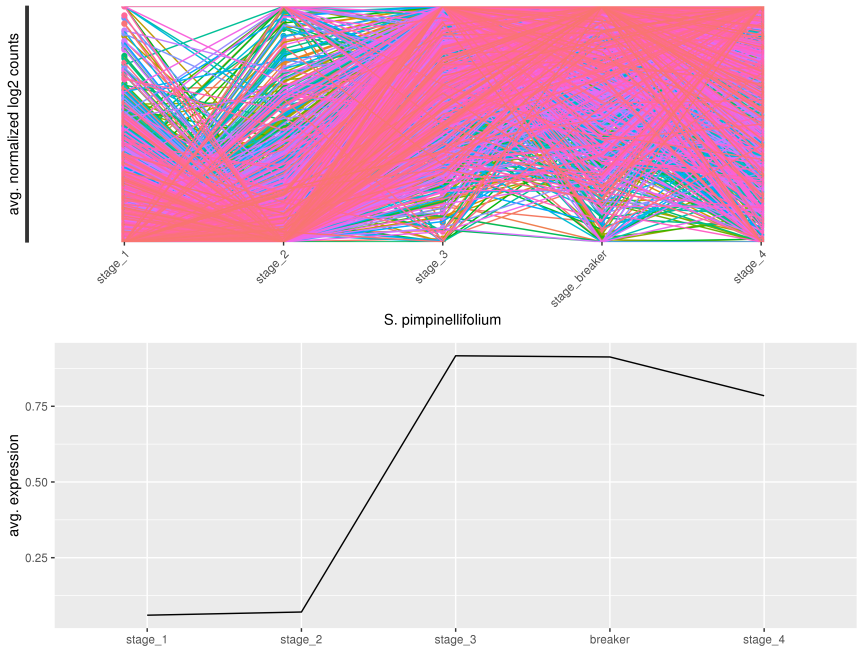
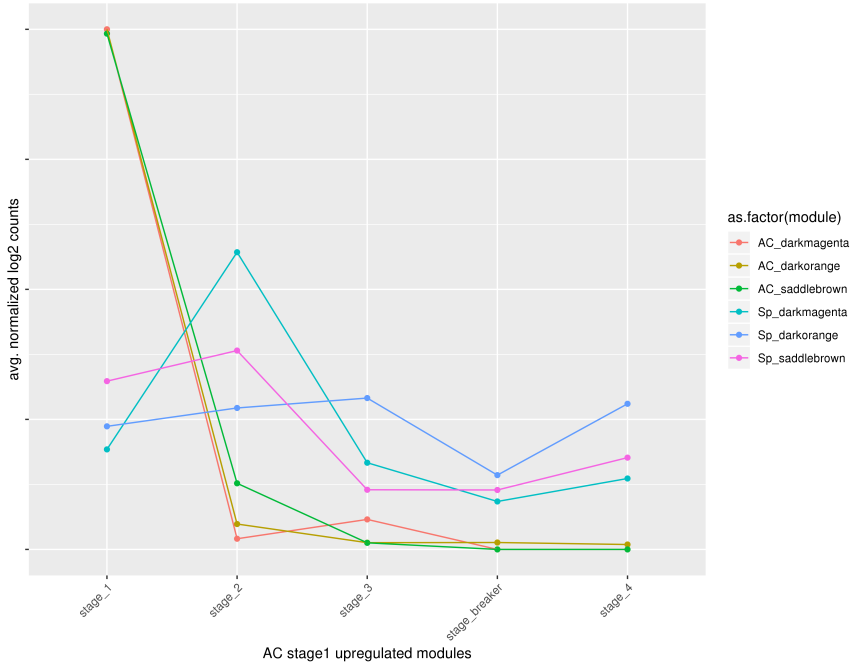


Figure 2.51. Expression line graphs for modules upregulated in only in (a) *S. lycopersicum* cv. Ailsa Craig stage 1 (b) *S. pimpinellifolium* stage 1.

(a)



(b)

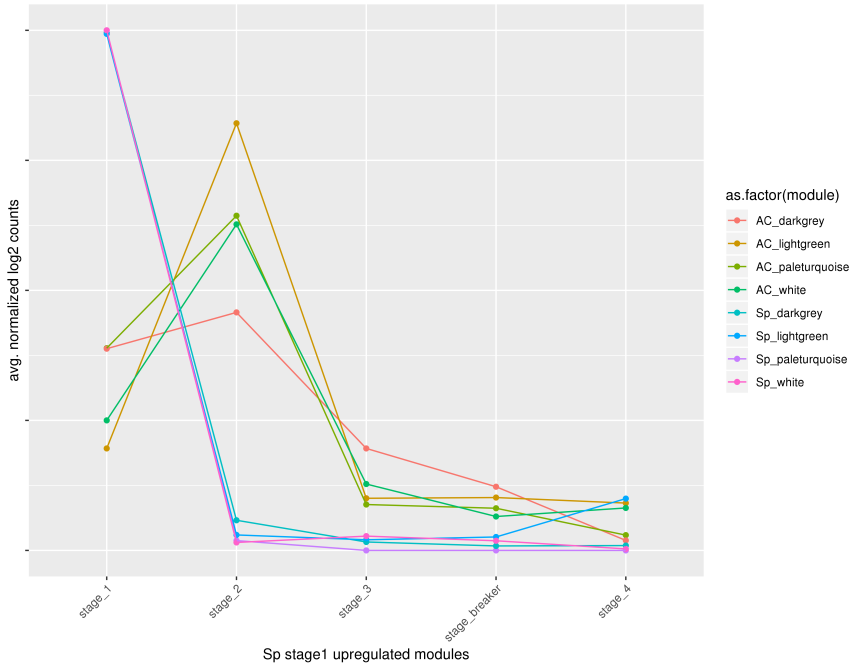


Table 2.1. The differentially expressed genes numbers. AC= *S. lycopersicum* cv. Ailsa Craig; Sp= *S. pimpinellifolium*. The numbers in the "Comparison" column refer to the stages of fruit development.

Comparison	**# Upregulated genes**	**# Downregulated genes**
AC2 vs AC1	389	224
Sp2 vs Sp1	463	503
AC3 vs AC2	846	1246
Sp3 vs Sp2	754	893
AC-breaker vs AC3	711	1273
Sp-breaker vs Sp3	537	1636
AC4 vs AC-breaker	21	162
Sp4 vs Sp-breaker	37	33
AC1 vs Sp1	322	348
AC2 vs Sp2	268	298
AC3 vs Sp3	218	392
AC-breaker vs Sp-breaker	395	313
AC4 vs Sp4	216	264

Table 2.2. Soft-threshold estimates for WGCNA analysis. The "powerEstimate" column includes the recommended soft-threshold value for a signed network analysis.
(a) All data combined. (b) Only *S. lycopersicum* cv. Ailsa Craig (c) Only *S. pimpinellifolium*.

(a)

	powerEstimate	fitIndices.Power	fitIndices.SFT.R.sq	fitIndices.slope	fitIndices.truncated.R.sq	fitIndices.mean.k.	fitIndices.median.k.	fitIndices.max.k.
1	15	1	0.128995171796262	0.476059120281209	0.84647036163281	6651.88763112044	6632.63295358302	10665.0131618786
2	15	2	0.321608639370275	-0.523066180646806	0.86804536433822	2811.67583150959	2610.34815399761	6365.82279998365
3	15	3	0.648087322759821	-0.921714103885271	0.894258716486687	1486.1744402579	1226.29761855388	4387.89461426097
4	15	4	0.74421210098504	-1.15488114964794	0.910328955226495	894.435517159907	681.266986816259	3274.26681917516
5	15	5	0.773453747920376	-1.31216816317997	0.913715046391792	585.830962605659	396.811067656868	2555.55119761775
6	15	6	0.785564214192593	-1.41174357585291	0.913442519275939	407.425130127904	245.771980874177	2055.63069690895
7	15	7	0.796595779786181	-1.48128167957993	0.920417438563811	296.483888641447	156.921895868483	1690.06218650476
8	15	8	0.785407101354496	-1.53807575868497	0.906361860158932	223.640526027209	107.785515593821	1412.98742178028
9	15	9	0.782478129136987	-1.56405067652659	0.905603811647153	173.750678070492	77.6423383139481	1197.22045914697
10	15	10	0.782245619762624	-1.57035308559598	0.898708322253236	138.406573714424	57.2967770483667	1025.57784598823
11	15	12	0.76615655765653	-1.57670745903359	0.858962145169948	93.4665762519515	31.6783517739957	772.592885294482
12	15	13	0.752317201416167	-1.51605196214749	0.814856984229619	78.8670541737225	24.4506893025003	677.795484842771
13	15	14	0.751092805445948	-1.41856979312047	0.755244718452445	67.5697990639698	18.7258876924008	598.183665539117
14	15	15	0.906508979334641	-1.23795417780139	0.880978237387908	58.6947620373175	14.471279581323	530.715127922103
15	15	16	0.922926903767138	-1.19627439955948	0.904772561569184	51.6286496579487	11.2938478893374	488.955228418914
16	15	17	0.877829278859497	-1.24332974085913	0.846655632611881	45.9350517437182	8.870265426627	482.061742156259
17	15	18	0.825316157594357	-1.29333944426577	0.779171813102373	41.2976540164165	6.99970255290011	475.71248501108
18	15	19	0.771076383495245	-1.34341712607343	0.709788394914394	37.4834019956927	5.59469298967491	469.833048927744
19	15	20	0.748983203790414	-1.37042669816025	0.684618324550312	34.3180458936345	4.52749537986234	464.362546951893

Continuation of Table 2.2.

(b)

	powerEstimate	"fitIndices.Power"	"fitIndices.SFT.R.sq"	"fitIndices.slope"	"fitIndices.truncated.R.sq"	"fitIndices.mean.k."	"fitIndices.median.k."	"fitIndices.max.k."
1	12	1	0.582983117473941	1.47910136053109	0.814586166572041	8391.47632201889	8435.03957720753	12319.8871120358
2	12	2	0.0258263602172499	0.0856611183509898	0.404485157037976	4438.15008280942	4232.17191436608	8451.38733280729
3	12	3	0.616285710574434	-0.406732606105019	0.755343812314056	2776.42565139946	2443.08586878582	6441.14717890795
4	12	4	0.783654298756432	-0.630208474942242	0.829469001577718	1910.84736988295	1531.4329550823	5189.30476070654
5	12	5	0.826913432871496	-0.765902775304181	0.859824596006869	1398.03580153326	1019.03245036417	4325.9058603899
6	12	6	0.833881529178519	-0.873837357750367	0.861714616229669	1067.20922387106	707.266485422277	3689.22910166206
7	12	7	0.840996701535251	-0.956096770937395	0.867861320606878	840.557503452162	507.132777439686	3198.87099501406
8	12	8	0.847904557355498	-1.02562337745974	0.876420180004978	678.199184563878	374.055612315089	2809.85619558257
9	12	9	0.843938156272318	-1.07542987831362	0.878420768641413	557.834775334504	298.194210005289	2493.53978376917
10	12	10	0.846766305575824	-1.11034282843698	0.889689683512942	466.136013231858	237.823261730626	2231.51956098563
11	12	12	0.855961184854573	-1.19016230787397	0.903818925664044	338.029882271879	167.114021788287	1823.58736522208
12	12	13	0.861799447411722	-1.21901028432246	0.912216436018414	292.349862427432	135.755712467793	1662.15472893476
13	12	14	0.858452833439311	-1.22509391415713	0.923273364865346	255.036104062964	113.909396708868	1521.97122475826
14	12	15	0.853326517331673	-1.24866133500939	0.925363701406745	224.19915516534	94.7396947076743	1399.2713432541
15	12	16	0.850317647043434	-1.27368098854676	0.9280891688674	198.452807626106	80.0115517476133	1291.12279936945
16	12	17	0.844117408775002	-1.3018549429271	0.926780398420449	176.761057385194	68.1979759882282	1195.20771771663
17	12	18	0.85009506597554	-1.31782184666527	0.935822968926801	158.336941290342	58.4618166723473	1109.66971306441
18	12	19	0.846272253393946	-1.33403414334025	0.935411039739741	142.574003469628	52.6157286546705	1033.00465946505
19	12	20	0.83697473817614	-1.34854276905947	0.929082859223371	128.998850799181	45.1951129860086	963.981163928268

Continuation of Table 2.2.

(c)

	powerEstimate	"fitIndices.Power"	"fitIndices.SFT.R.sq"	"fitIndices.slope"	"fitIndices.truncated.R.sq"	"fitIndices.mean.k."	"fitIndices.median.k."	"fitIndices.max.k."
1	13	1	0.517917681248729	1.93894317642443	0.887363328241091	7936.32767222419	8149.81892731095	11360.7273781629
2	13	2	0.0388647064940137	0.177672084822429	0.812755328023981	3906.23150458181	3942.8484566525	7158.34068118311
3	13	3	0.216501425641982	-0.368288583533498	0.783945806532479	2316.29280118238	2225.73265042292	5114.65576251904
4	13	4	0.464366177615013	-0.630012969386777	0.828332616045768	1529.47381645358	1369.23437790885	3912.893326099
5	13	5	0.59392414077823	-0.808350479657368	0.857232947465571	1083.07946194528	904.332661377024	3124.70929036918
6	13	6	0.668675023270598	-0.920442766026085	0.871168830907657	805.74352366366	667.819369793263	2569.73313871828
7	13	7	0.713825654094025	-1.00645979419641	0.882252763345598	621.956511349877	481.043244168476	2159.12012956793
8	13	8	0.73040840846127	-1.07691039975406	0.875775355896475	494.153700512312	356.619855195699	1848.9138801409
9	13	9	0.7368224676721	-1.14279930114503	0.865410646565094	401.899112607642	268.831250851753	1608.06303136517
10	13	10	0.740928094158893	-1.15310536717836	0.841781937718491	333.286732026562	214.87089834527	1413.25182605971
11	13	12	0.738018293268597	-1.19678211851784	0.794765613974607	240.317671124833	139.034425529522	1119.02636831194
12	13	13	0.875675298712945	-1.09542139131736	0.892965634625592	208.122290961657	111.973594084894	1005.86104921108
13	13	14	0.883831404777989	-1.08747124820459	0.878117685464075	182.256047120769	93.9336997296179	909.20428864778
14	13	15	0.880785198729816	-1.09306803516552	0.86234576800307	161.202168311874	80.437337374684	833.366808097952
15	13	16	0.938856488877245	-1.0915823328887	0.92433965741787	143.867589339848	67.5525092381685	798.247252675586
16	13	17	0.917014216848305	-1.12379527012285	0.89541587126661	129.449091527857	57.0454823840237	768.066195346196
17	13	18	0.892420604416513	-1.15329764242729	0.862171432613389	117.3462320206	48.0573364854038	741.933620917426
18	13	19	0.852643976388375	-1.19422411767785	0.810549217546487	107.103264456736	40.7462665214865	719.152859517294
19	13	20	0.809109193181513	-1.23429045485969	0.75497381995902	98.3695269195095	36.0251529674505	699.171892580824

Table 2.3. Co-expression clusters from WGCNA analysis depicting module-trait relationships for a soft-threshold value of 7. An individual box for each module contains an eigenvalue associated with the extent of up- or downregulation at a given stage and the p value (within parenthesis).

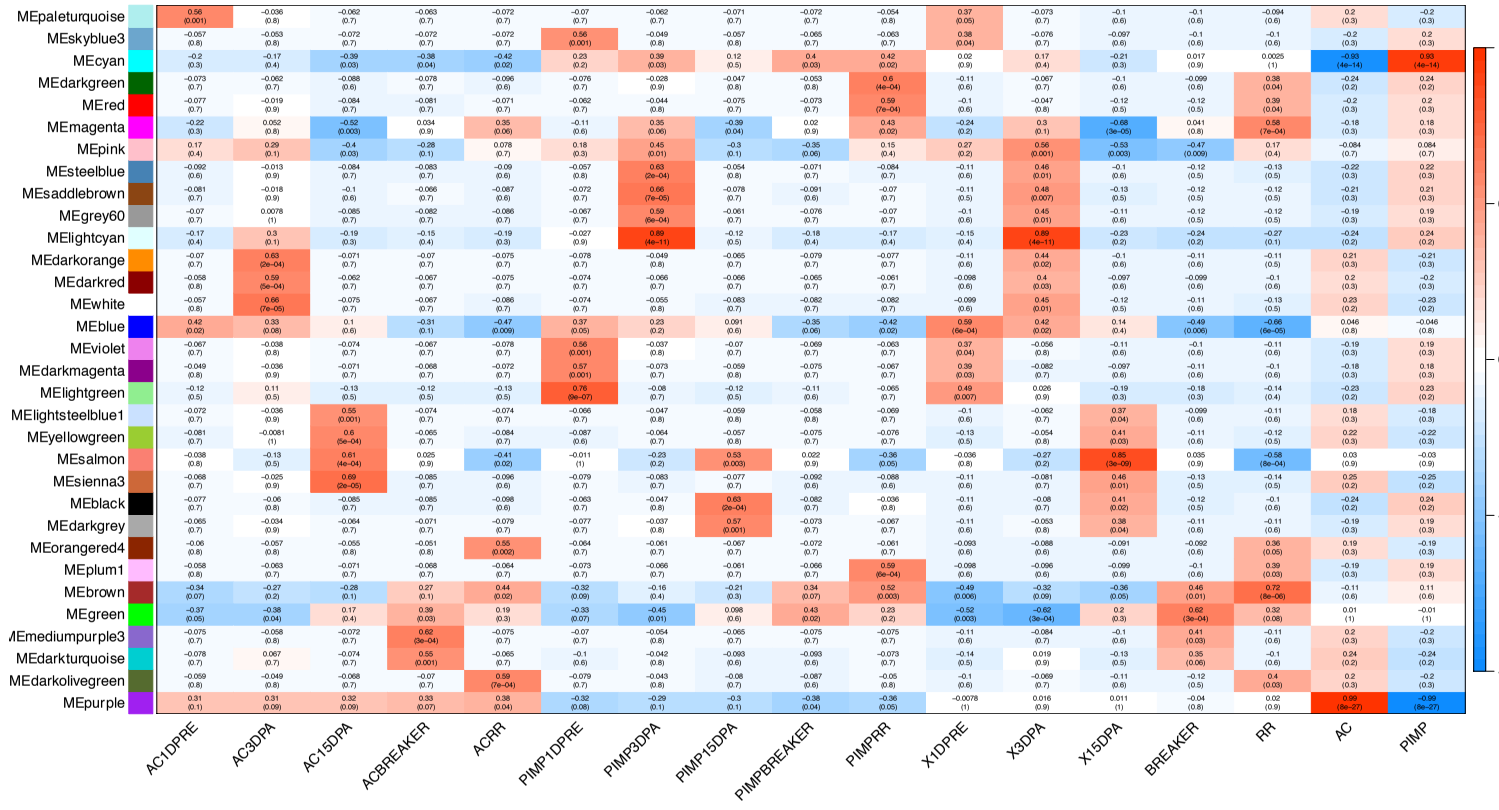


Table 2.4. Co-expression clusters from WGCNA analysis depicting module-trait relationships for a soft-threshold value of 15. An individual box for each module contains an eigenvalue associated with the extent of up- or downregulation at a given stage and the p value (within parenthesis).

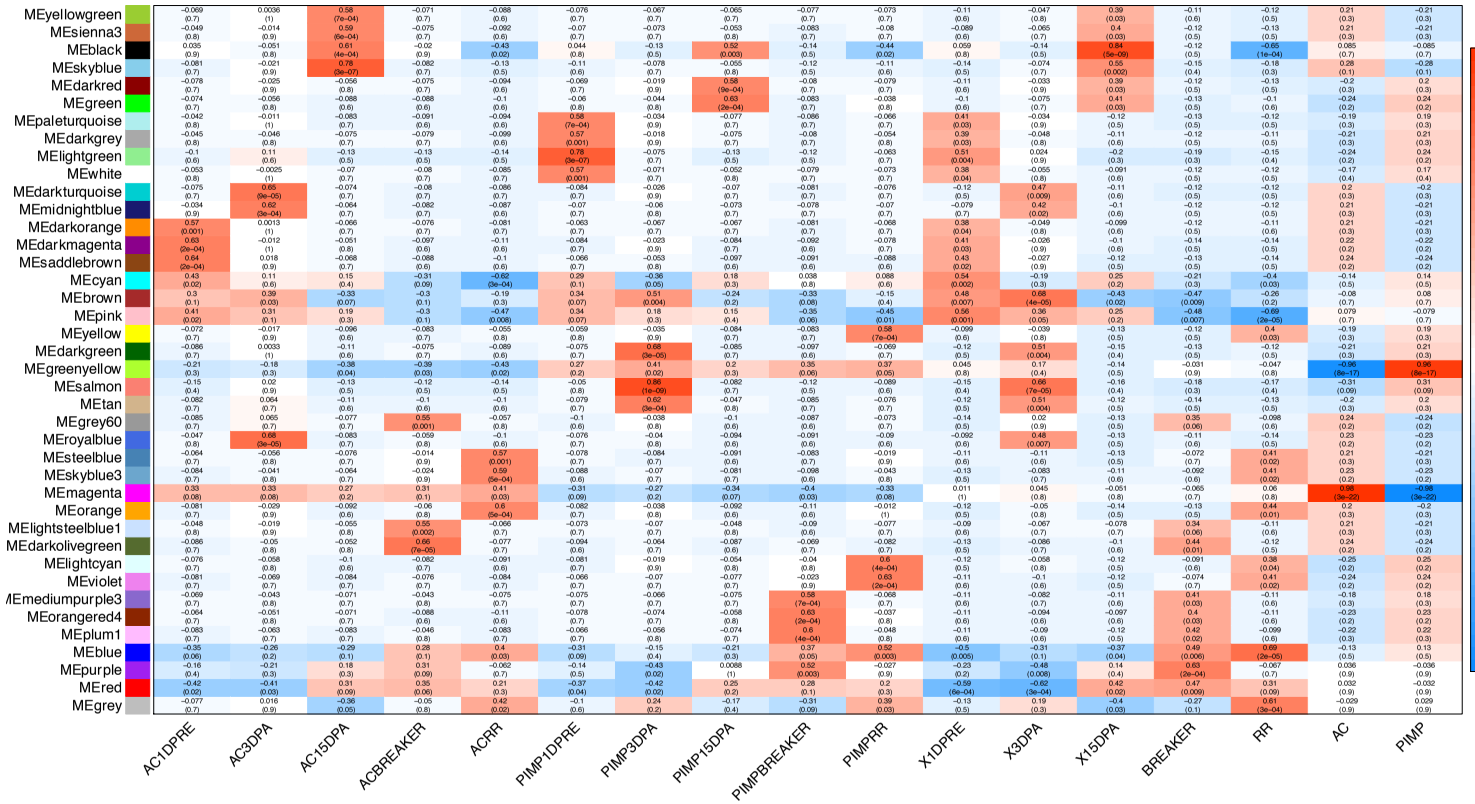


Table 2.5. *FUL* ortholog interactors in hypothetical WGCNA networks. (a) AC *SlFUL1* and AC *SlFUL2* (b) AC *MBP10* and AC *MBP20* (c) *SpFUL1* and *SpFUL2* (d) Sp *MBP10* and Sp *MBP20*.

(a)

***FUL* ortholog**	**Direction of interaction**	**Interactors (in descending order of contact strength)**
AC *SlFUL1*	From	1) *Solyc08g078430.3* (U4/U6 small nuclear ribonucleoprotein Prp31, 2) *Solyc11g011990.2* (plastid terminal oxidase), 3) *Solyc07g017490.3* (Protein Red - AHRD V3.3 *** A0A0B0MQ48_GOSAR), 4) *Solyc07g045340.3* (RPM1-interacting protein 4 family protein - AHRD V3.3 *** A0A061DN91_THECC), 5) *Solyc10g080900.2* (ripening related X72730), 6) *Solyc12g011180.2* (Transcription factor-related family protein - AHRD V3.3 *** B9IJ55_POPTR), 7) *Solyc08g074420.3* (Coiled-coil domain-containing proteinputative - AHRD V3.3 *** B9RY29_RICCO), 8) *Solyc07g026930.2* (Major facilitator superfamily protein - AHRD V3.3 *** AT5G42210.1), 9) *Solyc12g096420.2* (RING/U-box superfamily protein - AHRD V3.3 *** AT5G10650.2), 10) *Solyc09g008020.3* (OTU domain-containing protein - AHRD V3.3 *** A0A199VP67_ANACO)
	To	1) *Solyc01g008710.4* (inactive endo-beta-mannanase), 2) *Solyc02g062390.3* (Dehydrin (AHRD V3.3 *** E7BXD9_JATCU), 3) *Solyc03g063730.3* (Protein DETOXIFICATION -AHRD V3.3 *** M1AVC8_SOLTU), 4) *Solyc04g011540.3* (Calcium-dependent lipid-binding CaLB domain family protein AHRD V3.3 *** AT3G55470.1), 5) *Solyc04g074500.3* (Protein kinase AHRD V3.3 *** Q40264_MESCR), 6) *Solyc02g071980.3* (RING/FYVE/PHD zinc finger-containing protein AHRD V3.3 *** F4I346_ARATH), 7) *Solyc01g108880.3* (2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein AHRD V3.3 *** AT3G19000.1), 8) *Solyc06g009070.3* (Zinc finger transcription factor 41), 9) *Solyc01g101010.3* (F2P16.20 proteinputative isoform 1 -AHRD V3.3 *** A0A061GXF8_THECC), 10) *Solyc01g088030.3* (Structure-specific endonuclease subunit SLX1-like protein AHRD V3.3 *-* A0A072TLC3_MEDTR)
AC *SlFUL2*	From	1) *Solyc04g076390.3* (Cysteinyl-tRNA synthetaseputative AHRD V3.3 *** B9RP10_RICCO), 2) *Solyc04g079560.3* (Amino acid transporter family protein AHRD V3.3 *** B9MVJ1_POPTR), 3) *Solyc10g049890.2* (D-3-phosphoglycerate dehydrogenase (AHRD V3.3 *** K4D0E5_SOLLC), 4) *Solyc11g010690.2* (Nucleobase-ascorbate transporter-like protein AHRD V3.3 *-* A0A072TRH3_MEDTR), 5) *Solyc10g054080.2* (Kinesin-related protein (AHRD V3.3 *** D3YBF5_TRIRP), 6) *Solyc11g010600.2* (O-fucosyltransferase family protein AHRD V3.3 *** AT5G65470.1), 7) *Solyc08g076210.3* (Secretory carrier-associated membrane protein AHRD V3.3 *** G7K9I4_MEDTR), 8) *Solyc07g041640.3* (Growth-regulating factor-like protein AHRD V3.3 *** A0A072US13_MEDTR), 9) *Solyc06g005050.3* (F-box family protein AHRD V3.3 *** B9HWR1_POPTR), 10) *Solyc07g063230.3* (Choline transporter-related family protein AHRD V3.3 *** B9HZW2_POPTR)
	To	1) *Solyc02g065170.3* (L-ascorbate oxidase like AHRD V3.3 *** A0A0B2RUS1_GLYSO), 2) *Solyc02g094320.3* (Actin (AHRD V3.3 *** M4QSQ0_9ERIC), 3) *Solyc03g006100.3* (Receptor-kinaseputative AHRD V3.3 *** B9RVA8_RICCO), 4) *Solyc03g112790.3* (Wall-associated receptor kinase 3-like protein AHRD V3.3 *** A0A0B0PX41_GOSAR), 5) *Solyc01g067640.3* (Protein kinase superfamily protein AHRD V3.3 *** AT1G54610.3), 6) *Solyc01g058270.3* (O-acetyltransferase family protein AHRD V3.3 *** AT3G06550.1), 7) *Solyc01g111210.3* (Amidaseputative AHRD V3.3 *** B9SQJ8_RICCO), 8) *Solyc02g070790.3* (3-oxoacyl-[acyl-carrier-protein] synthase AHRD V3.3 *** A0A0V0IG10_SOLCH), 9) *Solyc01g094700.3* (Glycerol-3-phosphate acyltransferase AHRD V3.3 *** G7LBQ7_MEDTR), 10) *Solyc03g026120.3* (S-adenosyl-L-methionine-dependent methyltransferases superfamily protein AHRD V3.3 *** AT4G00750.1)

Continuation of Table 2.5.

(b)

FUL ortholog	Direction of interaction	Interactors (in descending order of contact strength)
AC *MBP10*	From	1) *Solyc03g044085.1* (Glucan endo-13-beta-glucosidase), 2) *Solyc09g018900.2* (SCP1-like small phosphatase 4 AHRD V3.3 --* AT5G46410.3), 3) *Solyc10g051150.2* (DNA (Cytosine-5)-methyltransferase 1replication foci domain-containing protein AHRD V3.3 *-* A0A124SE69_CYNCS), 4) *Solyc03g119730.3* (Histone deacetylase AHRD V3.3 *** A0A023IUN7_ARAHY), 5) *Solyc02g068600.3* (Ankyrin repeat-containing protein AHRD V3.3 *-* A0A103XBI3_CYNCS), 6) *Solyc07g042695.1* (Major facilitator superfamily protein AHRD V3.3 *** AT3G13050.4), 7) *Solyc04g009980.3* (LIGHT-DEPENDENT SHORT HYPOCOTYLS-like protein (DUF640) AHRD V3.3 *** AT1G07090.1), 8) *Solyc12g044320.2* (Transducin/WD40 repeat-like superfamily protein AHRD V3.3 *** AT3G50390.1), 9) *Solyc09g075900.3* (Spermidine synthase AHRD V3.3 *** F8WLB6_CITUN), 10) *Solyc02g085490.2* (DUF620 family protein AHRD V3.3 *** G7ZW25_MEDTR)
	To	1) *Solyc01g067040.2* (PWWP domain-containing family protein AHRD V3.3 *-* U5GDF3_POPTR), 2) *Solyc01g098790.2* (HSP20-like chaperones superfamily protein AHRD V3.3 *** AT1G20870.2), 3) *Solyc01g107650.3* (Ovary receptor kinase 1 AHRD V3.3 *** S4WIP5_SOLCH), 4) *Solyc01g091010.3* (CRABS CLAW-like protein1a), 5) *Solyc01g111310.3* (SlLAX2), 6) *Solyc01g091540.2* (Growth-regulating factor AHRD V3.3 *** M4T4B6_ARAHY), 7) *Solyc02g024070.3* (Homeobox leucine-zipper protein AHRD V3.3 *** Q8H963_ZINVI), 8) *Solyc00g007260.3* (Ulp1 protease family proteinputative AHRD V3.3 *-* Q6L4B6_SOLDE), 9) *Solyc01g066450.3* (Epoxide hydrolase AHRD V3.3 *** A0A072UTA2_MEDTR), 10) *Solyc01g109610.3* (Erythronate-4-phosphate dehydrogenase family protein AHRD V3.3 *** G7JG61_MEDTR)
AC *MBP20*	From	1) *Solyc12g011010.2* (protodermal factor 1 AHRD V3.3 *-* AT2G42840.1), 2) *Solyc08g066880.3* (Bark storage protein A AHRD V3.3 *** A0A151TDP1_CAJCA), 3) *Solyc10g005320.3* (Tryptophan synthase AHRD V3.3 *** K4CX65_SOLLC), 4) *Solyc11g012130.2* (Blue copper protein AHRD V3.3 *** A0A0B2RLQ3_GLYSO), 5) *Solyc10g083960.2* (Phosphoadenosine phosphosulfate reductase family protein AHRD V3.3 *** G7L3W8_MEDTR), 6) *Solyc06g083040.3* (wound-inducible carboxypeptidase), 7) *Solyc06g082350.3* (Biotin carboxyl carrier acetyl-CoA carboxylase AHRD V3.3 --* G7KRK6_MEDTR), 8) *Solyc10g084150.2* (Cytokinin riboside 5'-monophosphate phosphoribohydrolase AHRD V3.3 *** I0IUQ2_SOLLC), 9) *Solyc04g014650.3* (Non-specific serine/threonine protein kinase AHRD V3.3 *-* M1A662_SOLTU), 10) *Solyc04g009980.3* (LIGHT-DEPENDENT SHORT HYPOCOTYLS-like protein (DUF640) AHRD V3.3 *** AT1G07090.1)
	To	1) *Solyc02g068600.3* (Ankyrin repeat-containing protein AHRD V3.3 *-* A0A103XBI3_CYNCS), 2) *Solyc02g081120.3* (class 1 knotted-like homeodomain protein), 3) *Solyc01g009565.1* (Leucine-rich repeat receptor-like protein kinase AHRD V3.3 *-* Q9XGZ2_ARATH), 4) *Solyc02g071700.3* (LipaseGDSL AHRD V3.3 *** A0A103XTV4_CYNCS), 5) *Solyc02g068610.2* (Ankyrin repeat family protein AHRD V3.3 *-* AT5G35830.1), 6) *Solyc01g067040.2* (PWWP domain-containing family protein AHRD V3.3 *-* U5GDF3_POPTR), 7) *Solyc01g090740.3* (Zinc finger protein AHRD V3.3 *** A0A0B0MVF4_GOSAR), 8) *Solyc01g059760.2* (Sentrin-specific protease 1 AHRD V3.3 *-* W9RLZ4_9ROSA), 9) *Solyc02g083520.2* (LOW QUALITY:BZIP transcription factor AHRD V3.3 *** G7LB63_MEDTR), 10) *Solyc01g099870.2* (Bidirectional sugar transporter SWEET AHRD V3.3 *-* M1CB29_SOLTU)

Continuation of Table 2.5.

(c)

FUL ortholog	Direction of interaction	Interactors (in descending order of contact strength)
SpFUL1	From	1) *Solyc12g009060.2* (SNF7 family protein), 2) *Solyc07g056420.4* (glutathione S-transferase AY082341), 3) *Solyc11g007070.2* (mitochondrial substrate carrier family protein), 4) *Solyc10g083880.2* (tonoplast intrinsic protein 1.3), 5) *Solyc11g068650.2* (Mitochondrial fission protein ELM1 AHRD V3.3 *** ELM1_ARATH), 6) *Solyc08g005720.3* (Terpene synthase 18 AHRD V3.3 *-* R9R6F3_SOLLC), 7) *Solyc06g084270.3* (GTP-binding family protein AHRD V3.3 *** AT4G02790.1), 8) *Solyc08g014130.3* (Isopropylmalate synthase AHRD V3.3 *** K4CJ46_SOLLC), 9) *Solyc07g056510.3* (Glutathione S-transferase AHRD V1 **** D3Y4H6_9ROSI), 10) *Solyc07g039200.3* (WD-repeat proteinputative AHRD V3.3 *** B9SP08_RICCO)
	To	1) *Solyc02g064950.3* (CBS domain-containing protein-like AHRD V3.3 *** B8AP73_ORYSI), 2) *Solyc03g111950.3* (Cytochrome P450 AHRD V3.3 *** A0A103XWH5_CYNCS), 3) *Solyc01g010170.2* (Beta-glucosidaseputative AHRD V3.3 --* B9REF8_RICCO), 4) *Solyc04g015140.2* (Low PSII Accumulation 3 AHRD V3.3 --* AT1G73060.1), 5) *Solyc01g086740.3* (DnaJ-like protein AHRD V3.3 *** C4T7Z2_TOBAC), 6) *Solyc02g078840.3* (Secretory carrier membrane protein AHRD V3.3 *** B3TLU6_ELAGV), 7) *Solyc02g038750.3* (transducin family protein / WD-40 repeat family protein AHRD V3.3 *** AT2G34260.1), 8) *Solyc01g108860.3* (2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein AHRD V3.3 *** A0A061E0N9_THECC), 9) *RIN* (Ontology_term=GO:0003700GO:0046982), 10) *Solyc06g065825.1* (dipeptide transport ATP-binding protein AHRD V3.3 *** AT3G05570.1)
SpFUL2	From	1) *Solyc05g052970.3* (S-adenosyl-L-methionine-dependent methyltransferases superfamily protein AHRD V3.3 *** AT2G39750.1), 2) *Solyc11g071820.2* (Kinase family protein AHRD V3.3 *** D7M0T2_ARALL), 3) *Solyc05g054090.3* (UDP-3-O-acyl N-acetylglycosamine deacetylase family protein AHRD V3.3 --* AT1G25210.2), 4) *Solyc11g008820.2* (Endoglucanase AHRD V3.3 *** A0A0V0II53_SOLCH), 5) *Solyc05g010180.3* (Carotenoid isomerase AHRD V3.3 *** A0A077EPD5_TOBAC), 6) *Solyc07g005940.3* (Vacuolar ATPase subunit H protein AHRD V3.3 *** B6SIW6_MAIZE), 7) *Solyc04g015270.3* (Hexosyltransferase AHRD V3.3 *** M1CLK3_SOLTU), 8) *Solyc05g054040.3* (Transmembrane 9 superfamily member AHRD V3.3 *** K4C283_SOLLC), 9) *Solyc08g082170.3* (Pectin lyase-like superfamily protein AHRD V3.3 *** A0A097PQY8_SOLLC), 10) *Solyc06g066310.3* (Phosphomevalonate kinase AHRD V3.3 *** A0A0X9ZV04_9ROSI)
	To	1) *Solyc02g089630.3* (Proline dehydrogenase AHRD V3.3 *** A0A0H4CNX3_BETPL), 2) *Solyc02g089930.3* (Protein DA1-related 1 AHRD V3.3 *** A0A199UXH9_ANACO), 3) *Solyc01g105540.3* (2-oxoglutarate/malate translocatorchloroplastic AHRD V3.3 *** A0A0B0P714_GOSAR), 4) *Solyc02g083670.3* (histidine kinase 1 AHRD V3.3 *** AT2G17820.1), 5) *Solyc01g007150.3* (Coatomer subunit beta AHRD V3.3 *** K4AST1_SOLLC), 6) *Solyc02g014150.3* (Photosystem II stability/assembly factor HCF136 AHRD V3.3 *** G7JPH8_MEDTR), 7) *Solyc02g083810.3* (Ferredoxin--NADP reductase AHRD V3.3 *** A0A0V0I3W8_SOLCH), 8) *Solyc03g025910.3* (transmembrane protein AHRD V3.3 *** AT4G00585.1), 9) *Solyc02g092530.3* (Acetamidase/Formamidase family protein AHRD V3.3 *** AT4G37560.1), 10) *Solyc02g086330.3* (LisH domain and HEAT repeat-containing protein KIAA1468-like protein AHRD V3.3 *** A0A0B2SQV2_GLYSO)

Continuation of Table 2.5.

(d)

***FUL* ortholog**	**Direction of interaction**	**Interactors (in descending order of contact strength)**
Sp *MBP10*	From	1) *Solyc06g007640.3* (Protein FRIGIDA-like protein AHRD V3.3 *** A0A0B0PA17_GOSAR), 2) *Solyc08g080830.3* (Receptor kinaseputative AHRD V3.3 *** B9RC93_RICCO), 3) *Solyc02g093950.3* (Ankyrin repeat AHRD V3.3 *** A0A061EKD2_THECC), 4) *Solyc11g010170.2* (LanC-like protein 2 AHRD V3.3 *** A0A0B0NZR1_GOSAR), 5) *Solyc12g099560.2* (Protein arginine n-methyltransferaseputative AHRD V3.3 *** B9RT51_RICCO), 6) *Solyc04g072450.3* (Zinc finger CCCH domain-containing protein 54 AHRD V3.3 --* C3H54_ORYSJ), 7) *Solyc12g035710.2* (DUF21 domain-containing protein AHRD V3.3 *** A0A0B2P5X7_GLYSO), 8) *Solyc10g008450.1* (F-box associated interaction domain-containing protein AHRD V3.3 *** A0A103XHT5_CYNCS), 9) *Solyc07g006620.3* (Receptor-like kinase AHRD V3.3 *** G7JD52_MEDTR), 10) *Solyc02g078450.3* (Tetraspanin family protein AHRD V3.3 *** AT2G20740.1)
	To	1) *Solyc00g164680.2* (ABC transporter G family member AHRD V3.3 *-* A0A0K9NWI5_ZOSMR), 2) *Solyc01g099880.3* (Bidirectional sugar transporter SWEET AHRD V3.3 *** K4B122_SOLLC), 3) *Solyc01g109230.3* (Cysteine/Histidine-rich C1 domain family protein AHRD V3.3 *** A0A061E4Q7_THECC), 4) *Solyc01g087200.3* (Disease resistance protein AHRD V3.3 *** A0A118JXS4_CYNCS), 5) *Solyc00g007200.3* (MLO-like protein AHRD V3.3 *** A0A0V0IJD9_SOLCH), 6) *Solyc01g007955.1* (alpha/beta-Hydrolases superfamily protein AHRD V3.3 *** AT1G10740.4), 7) *Solyc01g111610.3* (SBP (S-ribonuclease binding protein) family protein AHRD V3.3 *** AT4G35070.1), 8) *Solyc01g100040.3* (Protein phosphatase 2C family protein AHRD V3.3 *** AT5G27930.2), 9) *Solyc01g109460.3* (Polyol monosaccharide transporter 4), 10) *Solyc01g066490.2* (Unknown protein AHRD V3.3)
Sp *MBP20*	From	1) *Solyc06g007640.3* (Protein FRIGIDA-like protein AHRD V3.3 *** A0A0B0PA17_GOSAR), 2) *Solyc03g005280.3* (Aspartic proteinase AHRD V3.3 *** A0A151SSN8_CAJCA), 3) *Solyc03g117250.3* (SPFH/Band 7/PHB domain-containing membrane-associated protein family AHRD V3.3 *** AT5G25260.1), 4) *Solyc07g061740.3* (Ankyrin repeat-containing-like protein AHRD V3.3 *** A0A061FUD1_THECC), 5) *Solyc04g017720.3* (Gibberellin-regulated family protein AHRD V3.3 *** A0A061EVH9_THECC), 6) *Solyc09g082760.3* (Aspartic proteinase AHRD V3.3 *** A0A0B2PY44_GLYSO), 7) Solyc04g072450.3 (Zinc finger CCCH domain-containing protein 54 AHRD V3.3 --* C3H54_ORYSJ), 8) *Solyc04g071890.3* (Peroxidase AHRD V3.3 *** K4BTH6_SOLLC), 9) *Solyc12g005300.2* (Chlorophyllase AHRD V3.3 *** A0A165YG74_POAPR), 10) *Solyc12g049400.2* (Jasmonate-zim-domain protein AHRD V3.3 *** A0A167V6B0_CAMSI)
	To	1) *Solyc02g083850.3* (Calcium-dependent protein kinase AHRD V3.3 *** C6KGT3_SOLLC), 2) *Solyc01g109230.3* (Cysteine/Histidine-rich C1 domain family protein AHRD V3.3 *** A0A061E4Q7_THECC) 3) *Solyc02g085400.3* (Sulfite exporter TauE/SafE family protein AHRD V3.3 *** AT2G25737.1), 4) *Solyc01g109460.3* (Polyol monosaccharide transporter 4), 5) *Solyc02g084410.3* (Lactoylglutathione lyase / glyoxalase I family protein AHRD V3.3 *** A0A061DGQ7_THECC), 6) *Solyc02g084880.3* (basic helix-loop-helix (bHLH) DNA-binding superfamily protein AHRD V3.3 *-* AT1G27660.1), 7) *Solyc02g068510.2* (PLATZ transcription factor family protein AHRD V3.3 *** A0A061DWT2_THECC), 8) *Solyc02g078450.3* (Tetraspanin family protein AHRD V3.3 *** AT2G20740.1), 9) *Solyc02g067800.3* (LOB domain-containing proteinputative AHRD V3.3 *** B9S640_RICCO), 10) *Solyc02g080410.3* (Heat-shock proteinputative AHRD V3.3 *** B9RXN2_RICCO)

Table 2.6. List of WGCNA modules (soft-threshold=15) specifically upregulated in only one species.

Stage	Species	Module names	GO categories
1	AC	darkorange, darkmagenta, saddlebrown	unclassified
	Sp	paleturquiose, darkgrey, lightgreen, white	cell wall modification, pectin catabolic processes
2	AC	darkturquiose, midnightblue, royalblue	metabolic process, detoxification
	Sp	darkgreen	unclassified
3	AC	yellowgreen, sienna3, skyblue	response to light stimulus, protein-chromophore linkage, photosynthesis light harvesting in photosystem I
	Sp	darkred, green	transcription (DNA-templated), cell wall organization/biogenesis, secondary metabolite biosynthetic process, hormone metabolic process, defense response to other organisms, histone deacetylation, negative regulation of transcription by RNA polymerase II
Breaker	AC	grey60, lightsteelblue1, darkolivegreen	protein phosphorylation, secondary metabolic processes, recognition of pollen
	Sp	mediumpurple3, orangered4, plum1	unclassified
4	AC	steelblue, skyblue3, orange	unclassified
	Sp	lightcyan, yellow, violet	nucleic acid phosphodiester bond hydrolysis, RNA modification, peptidyl-threonine phosphorylation
All	AC	magenta	cellular protein metabolic process, macromolecule modification, defense response
	Sp	greenyellow	carbohydrate derived biosynthetic processes, nucleotide biosynthetic processes, double strand break repair via homologous recombination

Table 2.7. The number of methyltransferases in WGCNA modules specific to one of the species.

Stage	Species	
	AC	**Sp**
1	104	6
2	5	38
3	21	28
Breaker	21	26
4	52	30

Chapter III:

Molecular mechanisms in the shift to selfing in *Collinsia*

Abstract

The evolutionary transition from outcrossing to self-mating species is considered a common natural phenomenon. A multitude of molecular mechanisms that may underlie some of these transitions have been proposed. In the genus *Collinsia* (Plantaginaceae), multiple pairs of sister species consist of an outcrossing species and a self-mating one. We analysed the floral transcriptomes of the sister pair, *C. linearis* and *C. rattanii* to investigate the potential molecular basis associated with the shift to selfing in the latter species. Our intraspecific comparisons between consecutive stages of floral development indicate genes encoding putative metal ion binding proteins might be associated with the change in developmental timing of the reproductive whorls in the transition to selfing. In addition, in agreement with previous findings, our data suggest low expression of genes with putative roles in pollen development and pollinator attraction in *C. rattanii*.

Introduction

The exchange of genetic material between organisms via outcrossing results in progeny with diverse allelic content. Such maintenance of genetic diversity increases population fitness, thereby, raises the probability of species survival. In angiosperms, despite the increased fitness via outcrossing, there have been numerous evolutionary shifts to selfing, which has the potential for the accumulation of deleterious alleles, thus, decreased fitness (Barrett, 2002). Studies have

suggested the shortages of pollinator or mate availability as the most likely ecological scenario that enforces the evolution of selfing (Brys et al., 2011; Lafuma and Maurice, 2007; Lloyd, 1992). In addition, as selfing generally requires a relatively shorter duration to complete compared to outcrossing, the former might be selected for in habitats with short-term resources or increased predation (Auld, 2010; Jorgensen and Arathi, 2013). However, there is currently no consensus on the molecular mechanisms that may underlie these shifts.

In the self-compatible genus *Collinsia* (Plantaginaceae), multiple pairs of sister-taxa consist of a predominantly outcrossing and a selfing species (Randle et al., 2009). These multiple evolutionary transitions in the genus, with the expansion of available genomic resources, provide an opportunity to investigate the mechanisms involved in the shift to selfing. The developmental phenomena that directly underlie the transition to selfing are the changes in the developmental timing of the reproductive whorls which cause reductions in 1) spatial separation between the anthers and the stigma (i.e., reduced herkogamy); 2) temporal separation between the maturation of the stamens and the pistil (i.e., reduced dichogamy). In the predominantly outcrossing *C. linearis*, stamen filaments elongate prior to the initiation of stigmatic receptivity, which sets apart the anthers from the stigma and, thus, prevents early self-pollination (Kalisz et al., 2012). It is only late in the reproductive cycle that the style elongates and the stigma reaches the dehisced anthers, which provides a potential opportunity for selfing (Kalisz et al., 2012). In comparison, in the transition to selfing in *C. rattanii*, the sister species of *C.*

linearis, there has been a reduction in the developmental timing of the reproductive whorls resulting in the elimination of herkogamy and dichogamy. Thus, the anther dehiscence and the initiation of stigmatic receptivity occur simultaneously in *C. rattanii* and there is no noticeable lapse between the anther filament and style elongation that delays selfing. However, the molecular mechanisms associated with this reduction in the developmental timing in *C. rattanii* remain to be elucidated.

Several groups have investigated the genetic architecture underlying the factors that influence herkogamy and dichogamy in other plant species but reached different conclusions. The findings on *Clarkia tembloriensis* (Onagraceae) and *Turnera ulmifolia* (Passifloraceae) suggest that a large number of loci may be involved in determining the developmental timing of the reproductive whorls (Holtsford and Ellstrand, 1992; Shore and Barrett, 1990). Meanwhile, other findings on *Mimulus* spp. (Phrymaceae) and *Arenaria uniflora* (Caryophyllaceae) point towards as few as two loci with potentially pleiotropic effects or linkage (Fishman et al., 2002; Fishman and Stratton, 2004). To date, only few genes associated with developmental timing of reproductive whorls have been identified. These include five genes in tomato (*Solanum lycopersicum*) that affect the style length (*STYLE2.1*), stamen length (*STAMEN2.1*, *STAMEN2.2*, and *STAMEN2.3*) and stamen architecture (*DEHISCENCE2.1*), which are all located in the quantitative trait locus (QTL) *SE2.1* (Chen and Tanksley, 2004; Pan et al., 2017). *STYLE2.1*, which encodes a basic helix-loop-helix (bHLH) protein, is hypothesized to be functioning downstream of

AUXIN RESPONSE FACTORS (*ARF*) *6* and *8* and is downregulated by microRNA 167a (Liu et al., 2014; Wang et al., 2015). However, no data currently exist on the mechanisms of *STAMEN2.1*, *STAMEN2.2*, *STAMEN2.3* or *DEHISCENCE2.1*.

A suite of biological traits that emerge following the shift to selfing, termed the "selfing syndrome," has also been identified (Sicard and Lenhard, 2011; Vos et al., 2014). This includes the breakdown of biochemical self-incompatibility (SI), and the general reductions in the pollen-to-ovule ratio and floral size. Among these, the molecular basis of SI is relatively better understood as the causative mechanisms have been elucidated in three plant families (Fujii et al., 2016). In the Brassicaceae, the S-locus protein 11 (SP11) in the pollen coat and the S-locus receptor kinase (SRK) in the stigma from the same haplotype results in the rejection of self-pollen (Kachroo et al., 2001; Shimosato et al., 2007; Takayama et al., 2001). In the Papaveraceae, the interaction between the female and male *Papaver rhoeas* style S (PrsS) proteins from the same haplotype leads to programmed cell death potentially via an increase in cytoplasmic calcium and reactive oxygen species, the breakdown of microtubules and the fragmentation of DNA (de Graaf et al., 2006; Thomas and Franklin-Tong, 2004; Wheeler et al., 2010; Wilkins et al., 2015). In the Solanaceae, a glycoprotein S-RNase produced in the style and multiple S-locus genes that encode F-box proteins in the pollen inhibit the growth of self-pollen tubes through ribonuclease/detoxification activity (Goldraij et al., 2006; Kubo et al., 2010; Lee et al., 1994; McClure et al., 1990, 2011; Murfett et al., 1994).

To investigate the molecular mechanisms associated with the evolutionary shift to selfing in *Collinsia*, we generated floral transcriptomes of the outcrossing *C. linearis* and its selfing sister species, *C. rattanii*. We hypothesized some of the differentially expressed genes may underlie the change in developmental timing of the reproductive whorls associated with the shift to selfing in the latter species. A previous study by Hazzouri et al. (2013) compared the transcriptomes of early flower buds in these two species. This group reported downregulation of genes with potential roles in pollen development and pollination in *C. rattanii* compared to *C. linearis*, which might be indicative of the low investment associated with the "selfing syndrome". In comparison to Hazzouri et al. (2013), we identified three stages spanning the entirety of floral development (see methods) and generated RNAseq libraries that represented these stages. My analyses described in this chapter consist of the intraspecific comparisons of our expression data between consecutive stages.

Materials and Methods

Plant material

Pre-anthesis plants and seeds of *C. linearis* and *C. rattanii* were acquired from Dr. Susan Kalisz at the College of Arts and Sciences, University Tennessee. The plants were grown in a glasshouse at University of California, Riverside (UCR) according to temperature and day-length conditions described by Hazzouri et al. (2013).

Tissue collection

The corresponding stages of floral development in *C. linearis* and *C. rattanii* were identified as follows:

Stage 1: all organs developed but green petals, green (immature) anthers, green styles in both species.

Stage 2: white petals, brown (mature, undehisced) anthers, green styles in *C. linearis* and white petals, yellow or brown (mature, dehisced/undehisced) anthers, white styles in *C. rattanii*.

Stage 3: white petals, yellow (dehisced) anthers, white styles in *C. linearis* and violet petals, yellow (dehisced) anthers, purple styles in *C. rattanii*.

For collecting tissue, the lengths of the buds at each stage were noted:

Stage 1: <1.5mm in *C. linearis* and <2mm in *C. rattanii*.

Stage 2: 3-4mm in both species.

Stage 3: 5mm or larger in both species.

Using the bud lengths noted for each stage, whole buds for three replicates were collected by Alannie-Grace Grant, then a graduate student from University of Tennessee and Knoxville, and myself.

RNA isolation and library preparation

RNA for one of the stage 2 *C. linearis* samples was extracted by Ms. Grant. I extracted RNA from the remaining samples. We used Qiagen RNeasy Plant Mini Kits (QIAGEN, Hilden, Germany) to extract RNA from the buds according to the manufacturer's protocol. The RNA quality was checked using a Bioanalyzer (Agilent,

CA, USA) by the staff at the Institute for Integrative Genome Biology (IIGB) UCR. We stored the RNA at –80°C until further use.

I used an NEBNext Ultra Directional RNA Library Prep Kit for Illumina and Protocol for use with NEBNext Poly(A) mRNA Magnetic Isolation Module protocol (New England BioLabs, MA, USA) for RNAseq library generation according to the manufacturer's protocol.

Sequencing, cleanup of the raw-sequencing-reads, and mapping of the reads to the *C. rattanii* genome

The libraries were sequenced on an Illumina HiSeq 2500 platform, with 125bp paired-end reads, at the Genome Innovation Center, McGill University, Quebec, Canada. I quality trimmed the raw sequence reads using TrimGalore (Krueger, 2017). To map the reads, I first assembled the *C. rattanii* genome provided by Dr. Stephen Wright at the Department of Ecology and Evolutionary Biology, University of Toronto, Canada using the FGENESH suite (Softberry, Inc., NY, USA). I then mapped the cleaned sequence reads to the assembled *C. rattanii* genome using Star (Dobin et al., 2013) as described in chapter 2. To assemble the genome and map the reads, I used the computer cluster at the Department of Ecology and Evolutionary Biology, University of Toronto.

Differential gene expression analysis

I performed intraspecific differential gene expression analyses comparing 2 consecutive stages using the DESeq2 package (Love et al., 2014; R Core Team, 2018). Genes were considered differentially expressed (DE) if the adjusted p value (false

discovery rate) was < 0.01 and the log2foldchange was > 2. I used the rlog-transformed counts from DESeq2 and the ggplot2 package in R to generate a PCA plot to visualize any patterns of similarities and differences in expression between stages and species (R Core Team, 2018; Wickham, 2016). To generate the heatmaps of DE gene expression, I used pheatmap and RColorBrewer packages on R (Kolde, 2012; Neuwirth and Brewer, 2014; R Core Team, 2018).

BLAST analysis of differentially expressed genes

For the 10 most strongly up- and downregulated genes in each set of differentially expressed (DE) genes, I retrieved the best matching sequence annotations by performing protein BLAST analyses on the High Performance Computer Cluster (HPCC) at UCR. If there were fewer than 10 genes, I used all of them. For this, I downloaded the National Centre for Biotechnology Information (NCBI) protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) and searched for the 10 best matches for each DE gene using the following command:

ncbi_refseqprotein/ncbi-blast-2.8.1+/bin/blastp -query query.txt -db refseq_protei
n -max_target_seqs 10 -num_threads 32 -outfmt "6 qseqid stitle" » output.txt

To interpret the results of our DE gene analyses, I used the most common BLAST hit (out of 10 hits) for each query sequence.

Gene ontology analysis

To extract the gene ontology (GO) categories enriched among the DEGs, I used the Gene Ontology Functional Enrichment Annotation Tool (GO FEAT; http://computa

tionalbiology.ufpa.br/gofeat) with the general settings (e-value=10) (Araujo et al., 2018). The output GO categories belonged to one of the three groups: "biological process," "cellular component" and "molecular function." I combined all these GO categories to detect any molecular markers that might be associated with the shift to self-mating.

Results

The PCA plots for the RNAseq data showed that *C. linearis* and *C. rattanii* samples clustered separately, suggesting the gene expression patterns were distinct between the two species a (Fig. 3.1). However, the plots for intraspecific expression data (Fig. 3.2 and 3.3) depict substantial variation among replicates. This might have been due to any errors in sample collection as the difference in bud lengths between two consecutive stages was minor (see methods). Under such circumstances, the outlying replicates might have belonged to the developmental boundary between two stages or a completely different stage. Therefore, the latter scenario has the potential to undermine the outcomes of the analyses that I report here.

I conducted differential expression (DE) analyses between two consecutive stages in *C. linearis* and *C. rattanii* separately to identify any genes that might be associated with the evolutionary transition to selfing in the latter species. The intraspecific comparisons for *C. linearis* resulted in more DE genes than those for *C. rattanii* (Fig. 3.4). In stage 1 vs 2 comparisons, there were 93 DE genes in *C.*

linearis and only 7 in *C. rattanii* (Tables 3.1 and 3.3, respectively). In stage 2 vs 3 comparisons, there were 147 DE genes in *C. linearis* but only 14 in *C. rattanii* (Tables 3.2 and 3.4, respectively). These considerable differences in DE gene numbers between the two species might represent a bias in read-mapping. However, I used the *C. rattanii* genome to map the expression data from both species. Therefore, it seems unlikely that a mapping-bias led to the lower DE gene count in *C. rattanii* but not in *C. linearis*, which is likely to have undergone some sequence divergence compared to its sister species. In selfing *C. rattanii*, reproductive maturity occurs relatively faster compared to the outcrossing *C. linearis* (Hazzouri et al., 2013). Thus, it is possible there was overlap in the processes between consecutive stages in *C. rattanii*, which might account for the low DE gene counts, whereas the drawn out development of *C. linearis* may allow for greater separation of processes.

The relative expression levels of all DE genes in each comparison are depicted in Figures 3.5 through 3.8. Out of these genes, I selected the ten most up- and down-regulated ones in each comparison and performed BLAST analyses to hypothesize their putative functions to investigate whether any of these might be associated with the change in developmental timing in the shift to selfing (Table 3.5).

DE gene analysis for *C. linearis* developmental stages

In *C. linearis* stage 2 compared to stage 1, a gene encoding a putative flavonoid 3',5'-hydroxylase was upregulated (Table 3.5). This has also been reported by a previous investigation that performed interspecific transcriptome analyses of *C. linearis* and *C. rattanii* flower buds (Hazzouri et al., 2013). Flavonoid 3',5'-hydroxylase is

involved in the production of blue/ purple anthocyanin pigmentation and, thus, it may have a potential role in pollinator attraction starting at stage 2 when the stigma becomes receptive in contrast to the unreceptive stage 1 flowers (Seitz et al., 2006). A gene encoding a putative isoform of tetraketide alpha-pyrone reductase 1, was downregulated in *C. linearis* stage 2 compared to stage 1 (Table 3.5). This enzyme has a function in the biosynthesis of sporopollenin, which is a major biopolymer that makes up the outer pollen wall (exine) (Grienenberger et al., 2010; Tian et al., 2017). It has been reported that sporopollenin is integrated into the exine during early stages of development (Blackmore et al., 2007; Dickinson and Potter, 1976) and, thus, the process might be largely completed by stage 2. I was not able to find any potential relationships between the remainder of the DE genes (in stage 2 compared to stage 1) and floral development.

In *C. linearis* stage 3 compared to stage 2, two putative alpha-farnesene synthase-like genes were upregulated (Table 3.5). Alpha-farnesene is a known component of floral fragrances which facilitate pollinator attraction (Azuma et al., 2002; Jürgens et al., 2002; Pott et al., 2002). Since *C. linearis* is predominantly an outcrosser, it is possible this chemical helps guide pollinators to mature flowers. A gene encoding a putative serine racemase isoform was downregulated in stage 3 compared to stage 2 (Table 3.5). Serine racemase has been reported to play a role in guiding the growth of pollen tubes into the ovules (Michard et al., 2011). Thus, the significance of the gene encoding a putative serine racemase being downregulated in stage 3, during the maximum activity of pollen tube growth in the style, is unclear. One

possible explanation is that our samples for stage 3 might have been from flowers that may have reached the end of fertilization and pollen tube growth had already been completed.

DE gene analysis for *C. rattanii* developmental stages

In *C. rattanii* stage 2 compared to stage 1, none of the upregulated genes had functions specifically associated with floral development and fertilization. However, a gene encoding a putative polyol transporter 6-like (PMT6-like) was downregulated in stage 2 compared to stage 1 (Table 3.5). Six polyol transporters have been identified in *Arabidopsis thaliana* (*AtPMT1-6*) (Klepek et al., 2010). Out of these, *AtPMT1* and *AtPMT2* are reported to have potential roles in mature pollen and growing pollen tubes (Klepek et al., 2010). In contrast, *AtPMT5* is not expressed in pollen and, thus, does not seem to have a role in pollen development (Klepek et al., 2005). The function or the expression domains of *AtPMT3/4/6* have not been elucidated. However, the downregulation of *PMT6-like* in *C. rattanii* stage 2 when pollen grains are more mature compared to stage 1 suggests a role that is potentially different from those of *AtPMT1* and *AtPMT2*, which function in mature pollen.

In *C. rattanii* stage 3 compared to stage 2, none of the downregulated genes had functions specifically associated with floral development and fertilization. A gene that encodes a putative flavanone 3-hydroxylase-like was upregulated in stage 3 compared to stage 2 (Table 3.5). As this enzyme has a role in the synthesis of floral pigmentation (Tan et al., 2013), this is in agreement with the transformation

of white petals in stage 2 to violet in stage 3.

GO categories associated with the evolutionary transition to selfing

The BLAST hits for some of the top DE genes had roles associated with phenotypic traits that are frequently found in selfers. However, none of the putative functions for these genes were directly related to the changes in the developmental timing of the reproductive organs that directly underlie the evolutionary transition to self-mating. To search for such molecular signatures, I analyzed the enriched functional terms for all DE genes between two consecutive stages in each species separately.

Our data indicate that the number of both up- and downregulated genes with potential functions in metal ion binding (Tables 3.6 and 3.7) increased over the course of floral development in *C. linearis*. Evidence suggests that metal ion (e.g. Ca^{2+}) binding proteins might have potential roles in stamen filament and style elongation through cellular expansion (Chaiwongsar et al., 2009; O'Brien et al., 2002). In *C. linearis* stage 2 compared to stage 1, two putative genes in the metal ion binding category were upregulated while six genes in the same category were downregulated (Table 3.6). In the predominantly outcrossing *Collinsia* species such as *C. linearis*, the stamen filaments start elongating in stage 2, thereby positioning the anthers away from the stigma, inhibiting early self-mating (Kalisz et al., 2012). Thus, the putative metal ion binding category might be involved in filament elongation during stage 2 in *C. linearis* (Chaiwongsar et al., 2009). However, there is not enough evidence in other species to suggest whether the up- and downregu-

lated metal ion binding proteins may have identical or antagonistic roles in this process.

In *C. linearis* stage 3 compared to stage 2, 16 putative genes in the metal ion binding category were upregulated while 13 genes in the same category were downregulated (Table 3.7). During stage 3 in this species, as the stamen filament elongation continues and the style elongation initiates, some of the metal ion binding protein encoding genes might be involved in these developmental processes (Chaiwongsar et al., 2009; O'Brien et al., 2002). Meanwhile, the stigma also becomes receptive during stage 3 (Kalisz et al., 2012). Although not associated with developmental timing, metal ion (e.g. Ca^{2+}, heme) binding proteins may also have functions in pollen germination and pollen tube growth (Chaiwongsar et al., 2009; Ge et al., 2009; Guyon et al., 2000; Rato et al., 2004; Wood, 2017). Therefore, it is possible that some of the differentially expressed genes in the putative metal ion binding category might be involved in such post-fertilization processes. However, we do not know if any of our stage 3 *C. linearis* flowers had been fertilized as these plants were grown in a glasshouse that excluded any access to pollinators.

Although the number of DE genes in the putative metal ion binding category also increased over the course of *C. rattanii* floral development, it was minor compared to *C. linearis* and none of these genes were shared between the two species (Tables 3.6 through 3.9). In *C. rattanii* stage 2 compared to stage 1, one of the two upregulated genes belonged to the metal ion binding category (Table 3.8). In this species, as the time gap between flower maturity and fertilization is relatively short com-

pared to *C. linearis* and there is no noticeable stamen filament/ style elongation in the interim, it is unlikely that potential genes in the metal ion binding category are associated with developmental timing. However, in contrast to *C. linearis*, the stigma of self-mating *C. rattanii* becomes receptive and some of the mature anthers dehisce during stage 2, resulting in self-pollination. Thus it is possible that the upregulated gene in the metal ion binding category in stage 2 might have a role in pollen germination and pollen tube growth (Chaiwongsar et al., 2009; Ge et al., 2009; Guyon et al., 2000; Rato et al., 2004; Wood, 2017). None of the downregulated genes in *C. rattanii* stage 2 compared to stage 1 were identified in the metal ion binding category (Table 3.8).

In *C. rattanii* stage 3 compared to stage 2, only four of the upregulated genes and one of the downregulated genes were in the metal ion binding category (Table 3.9). As the remainder of the anthers dehisce during stage 3, it is possible these genes also are involved in pollen germination and pollen tube growth (Chaiwongsar et al., 2009; Ge et al., 2009; Guyon et al., 2000; Rato et al., 2004; Wood, 2017). In contrast to the metal ion binding category, which might have roles in developmental timing of the reproductive whorls, I was not able to find any association between the other GO categories and floral development (Tables 3.6 through 3.9).

Discussion

Although we used the *C. rattanii* genome to map the sequence reads, our differential expression analysis identified a considerably low number of genes between the consecutive stages of *C. rattanii* compared to those in *C. linearis*. Thus these results are the opposite of what may be observed under a potential mapping bias. It is possible that, since *C. rattanii* has the smallest flowers between the two species, which makes the difference in bud length between the stages miniscule, we might have collected samples that were in the transition zone between two stages. If so, this would explain the almost identical expression profiles of *C. rattanii* samples that we had marked as belonging to separate stages. Alternatively, the rapid development of flowers in selfing species such as *C. rattanii* might be achieved through overlapping gene expression patterns between stages. In this scenario, there is an overlap between the stages – the processes characteristic of stage 2 in *C. linearis* begin in *C. rattanii* before stage 1 is completed whereas those of stage 3 in *C. linearis* begin in *C. rattanii* stage 2. This may shorten the length of time that might have otherwise been spent in inducing complex genetic pathways and reaching optimal protein levels. However, I was not able to find any evidence in these two species that support the latter hypothesis.

Genes encoding metal ion binding proteins are associated with the shift to selfing

Our data indicate a relatively higher number of DE genes in the metal ion binding GO category in *C. linearis* compared to *C. rattanii*: 29 vs. 6, respectively (Tables

3.6 through 3.9). The metal ion binding related genes in *C. linearis* might have potential functions in establishing herkogamy via stamen filament/ style elongation (Chaiwongsar et al., 2009; O'Brien et al., 2002). Thus, we hypothesize these genes might be associated with the shift to selfing in *C. rattanii* via the breakdown of herkogamy.

During stage 2 in *C. linearis*, the stamen filaments elongate placing the anthers away from the stigma, thereby preventing early self pollination (Kalisz et al., 2012). While the remaining stamen filaments elongate during stage 3, the style also starts to elongate, ultimately reaching the level of the dehisced anthers, which might result in delayed self-pollination in the absence of pollinator action (Kalisz et al., 2012). The results of our GO analyses included Ca^{2+}, Zn^{2+}, K^{+}, Mg^{2+} binding proteins (Tables 3.6 through 3.9). Evidence suggests that such metal ion binding proteins are involved in both anther filament and style elongation. Chaiwongsar et al. (2009) reported short stamen filaments in *A. thaliana* due to defects in cell elongation when calmodulin binding cyclic nucleotide-gated channel 2 (CNGC2) was mutated, which disrupted Ca^{2+} signaling. O'Brien et al. (2002) reported a correlation between the strong expression of *PIP2a*, which encodes a type of aquaporin that potentially binds to Cd^{2+} and Ca^{2+} (Nyblom et al., 2009), and style elongation in *Solanum chacoense*. Among the other metal ion binding proteins that might have roles in stamen filament and style extension via cell elongation/expansion include class III peroxidases that contain heme binding motifs (Cosio et al., 2009; Francoz et al., 2015), plasma membrane ATPases through the interaction with Mg–

ATP (Hager, 2003; Pitann et al., 2009), zinc finger proteins that have Zn^{2+} binding motifs and respond to gibberellic acid (Lin et al., 2011), and high-affinity potassium (HKT) transporters with K^{+} binding sites (Gierth and Mäser, 2007; Mäser et al., 2002).

There has been a change in the developmental timing of the reproductive whorls in the shift to selfing in *C. rattanii*. In this species, there is no noticeable elongation of stamen filaments followed by the style as in *C. linearis* and both reproductive whorls mature simultaneously. Thus, the few DE genes in the metal ion binding category in *C. rattanii* might not have any functions associated with the developmental timing of the stamens and pistils. In contrast, these metal ion binding proteins might be involved in pollen germination and pollen tube growth through processes such as Ca^{2+} homeostasis (Chaiwongsar et al., 2009; Ge et al., 2009;Rato et al., 2004;Wood, 2017) and Fe^{2+} binding (Guyon et al., 2000).

Although pollination occurs in *C. linearis* during stage 3, interestingly, none of the putative metal ion binding genes we hypothesized to be involved in pollen germination and pollen tube growth in *C. rattanii* were among the DE genes in the former species. It is unlikely that different repertoires of genes will be expressed in the same biological process in two such closely related species. As we grew our plants in the greenhouse in the absence of a pollinator, it may be that *C. linearis* stigma did not reach self-pollen until late in stage 3 when the petals start falling off (Kalisz et al., 1999). However, we did not collect such late stage flowers, which suggests a lack of germinated pollen in our *C. linearis* samples. This might explain

why none of the putative metal binding related genes from *C. rattanii* were among the DE genes in *C. linearis*.

Expression data reveals reduced investments in pollen development following the transition to selfing

Our expression data for outcrossing *C. linearis* included a DE gene that encodes a putative tetraketide alpha-pyrone reductase 1, which has a potential role in pollen development. However, this gene was not differentially expressed between any of the consecutive stages in the self-mating *C. rattanii*. Tetraketide alpha-pyrone reductase 1 is involved in the biosynthesis of sporopollenin that makes up the exine (Grienenberger et al., 2010; Tian et al., 2017). Sporopollenin is able to withstand harsh environmental conditions (Brooks and Shaw, 1978; Domínguez et al., 1999; Yule et al., 2000), enabling the pollen from outcrossing species to survive unpredictable conditions while being transported to a potential mate. The lack of any *tetraketide alpha-pyrone reductase 1* upregulation in *C. rattanii* might imply the exine in this species is not as enforced with sporopollenin as in *C. linearis*. This is in agreement with the reported reduced investment in the male function following the shift to selfing (Cruden, 1977; Sicard and Lenhard, 2011); as the resources that may be allocated to reproduction are limited, selection pressure on the mechanisms for lengthening the span of pollen survival or transmission, which are traits relevant to outcrossers, may be relaxed in selfers as suggested by Hazzouri et al. (2013).

A low investment in pollen wall formation proteins might, in addition to any im-

plications on the reduced hardiness of individual pollen grains, also suggest a reduction in their overall numbers. It has been reported that the pollen-to-ovule ratio tends to decrease owing to a reduction in the pollen count during the reallocation of resources in the transition to selfing (Cruden, 2000; Lozada-Gobilard et al., 2019). Since the probability of fertilization is high for the pollen in a self-compatible flower, the number of pollen grains can decrease without affecting reproductive assurance. In addition, a study has reported a positive relationship between the pollen-to-ovule ratio and herkogamy, which is the spatial separation between the anthers and the stigma, in *Melochia* (Malvaceae) (Faife-Cabrera et al., 2018). This implies a reduced pollen-to-ovule ratio when the distance between the anthers and the stigma are reduced. Therefore, our expression data provides the theoretical background for a scenario where *C. rattanii*, which is a selfer with reduced herkogamy compared to its sister species, might also have a reduced pollen-to-ovule ratio. However, no comparative analyses on pollen counts between these two *Collinsia* species are available at present.

Molecular signatures associated with the loss of pollinator attraction following the transition to selfing

According to our data, genes encoding a putative flavonoid 3',5'-hydroxylase and two putative alpha-farnesene synthases are upregulated only in *C. linearis*. Hazzouri et al. (2013) also reported the upregulation of a flavonoid 3',5'-hydroxylase in *C. linearis* compared to *C. rattanii* and suggest a function for this enzyme in attracting pollinators. In addition, the alpha-farnesene synthases are involved

in the production of volatiles emitted by flowers that attract pollinators (Azuma et al., 2002; Jürgens et al., 2002; Pott et al., 2002). Although flavonoid 3',5'-hydroxylase and the alpha-farnesene synthases are not differentially expressed between any two consecutive stages in *C. rattanii*, these genes are still expressed in this species. This suggests there has been a divergence in the regulatory mechanisms of these orthologs between the two species, which might be associated with the re-programming of pollination mechanisms.

Significant correlations between changes in the epigenome and the mating system have been reported for *Kryptolebias marmoratus* (fish species) (Ellison et al., 2015), *A. thaliana–lyrata* and *Capsella rubella–grandiflora* hybrids (Nasrallah et al., 2007). Therefore, this raises the possibility that the interspecific differences in the expression of the orthologs potentially involved in pollinator attraction in *C. linearis* might be due to any epigenetic changes in the transition to selfing in *C. rattanii*. In addition, changes to regulatory regions have also been identified in correlation with the transition to selfing in *Capsella rubella* (Steige et al., 2015) and *Eichhornia paniculata* populations (Arunkumar et al., 2016). Thus, it is also possible that in comparison to the coding sequences, the regulatory regions of these pollination associated genes in *C. rattanii* might have undergone divergence in the evolutionary shift to selfing.

Conclusion

Our analyses were not sensitive enough to identify any molecular markers that may underlie the breakdown of dichogamy in the evolutionary shift to selfing in *C. rattanii*. However, our gene ontology category analyses suggest an association between genes in the metal ion binding group and the establishment of herkogamy in the predominantly outcrossing *C. linearis*. Thus, we hypothesized this gene set might be involved in the transition to selfing in *C. rattanii*, which lacks any selfing inhibition through herkogamy. Our differential gene expression data for this species also indicate reduced expression of genes with potential functions in pollen development and pollinator attraction, which are traits that typically follow the transition to selfing.

Araujo, F. A., Barh, D., Silva, A., Guimarães, L., and Ramos, R. T. J. (2018). GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Sci. Rep. 8, 1794.

Arunkumar, R., Maddison, T. I., Barrett, S. C. H., and Wright, S. I. (2016). Recent mating-system evolution in *Eichhornia* is accompanied by cis-regulatory divergence. New Phytol. 211, 697–707.

Auld, J. R. (2010). The effects of predation risk on mating system expression in a freshwater snail. Evolution 64, 3476–3494. doi:10.11 11/j.1558-5646.2010.01079.x.

Azuma, H., Toyota, M., Asakawa, Y., Takaso, T., and Tobe, H. (2002). Floral scent chemistry of mangrove plants. J. Plant Res. 115, 47–53.

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. Nature Reviews Genetics 3, 274–284. doi:10.1038/nrg776.

Blackmore, S., Wortley, A. H., Skvarla, J. J., and Rowley, J. R. (2007). Pollen wall development in flowering plants. New Phytol. 174, 483–498.

Brooks, J., and Shaw, G. (1978). Sporopollenin: A review of its chemistry, palaeochemistry and geochemistry. Grana 17, 91–97. doi:10.1080/0017313780 9428858.

Brys, R., de Crop, E., Hoffmann, M., and Jacquemyn, H. (2011). Importance of autonomous selfing is inversely related to population size and pollinator availability in a monocarpic plant. American Journal of Botany 98, 1834–1840. doi:10.3732/ajb.1100154.

Chaiwongsar, S., Strohm, A. K., Roe, J. R., Godiwalla, R. Y., and Chan, C. W. M. (2009). A cyclic nucleotide-gated channel is necessary for optimum fertility in high-calcium environments. New Phytol. 183, 76–87.

Chen, K.Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of *se2.1*: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus Lycopersicon. Genetics 168, 1563–1573.

Cosio, C., Vuillemin, L., De Meyer, M., Kevers, C., Penel, C., and Dunand, C. (2009). An anionic class III peroxidase from zucchini may regulate hypocoty l elongation through its auxin oxidase activity. Planta 229, 823–836.

Cruden, R. W. (1977). Pollen-ovule ratios: A conservative indicator of breeding systems in flowering plants. Evolution 31, 32–46.

Cruden, R. W. (2000). Pollen grains: why so many? Pollen and Pollination, 143–165. doi:10.1007/978-3-7091-6306-1_8.

de Graaf, B. H. J., Rudd, J. J., Wheeler, M. J., Perry, R. M., Bell, E. M., Osman, K., et al. (2006). Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. Nature 444, 490–493.

Dickinson, H. G., and Potter, U. (1976). The development of patterning in the alveolar sexine of *Cosmos bipinnatus*. New Phytologist 76, 543–550. doi:10.1111/j.1469-8137.1976.tb01490.x.

Dietrich, P., Anschütz, U., Kugler, A., and Becker, D. (2010). Physiology and biophysics of plant ligand-gated ion channels. Plant Biol. 12 Suppl 1, 80–93.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Domínguez, E., Mercado, J. A., Quesada, M. A., and Heredia, A. (1999). Pollen sporopollenin: degradation and structural elucidation. Sexual Plant Reproduction 12, 171–178. doi:10.1007/s004970050189.

Ellison, A., Rodríguez López, C. M., Moran, P., Breen, J., Swain, M., Megias, M., et al. (2015). Epigenetic regulation of sex ratios may explain natural variation in self-fertilization rates. Proc. Biol. Sci. 282. doi:10.1098/rspb.2015.1900.

Faife-Cabrera, M., Ferrero, V., and Navarro, L. (2018). Relationship between herkogamy, incompatibility and reciprocity with pollen–ovule ratios in *Melochia* (Malvaceae). Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 152, 80–89. doi:10.1080/11263504.2016.1255266.

Fishman, L., Kelly, A. J., and Willis, J. H. (2002). Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. Evolution 56, 2138–2155. doi:10.1111/j.001 4-3820.2002.tb00139.x.

Fishman, L., and Stratton, D. A. (2004). The genetics of floral divergence and postzygotic barriers between outcrossing and selfing populations of *Arenaria uniflora* (Caryophyllaceae). Evolution 58, 296–307.

Francoz, E., Ranocha, P., Nguyen-Kim, H., Jamet, E., Burlat, V., and Dunand, C. (2015). Roles of cell wall peroxidases in plant development. Phytochemistry 112, 15–21. doi:10.1016/j.phytochem.2014.07.020.

Fujii, S., Kubo, K.-I., and Takayama, S. (2016). Non-self- and self-recognition models in plant self-incompatibility. Nat Plants 2, 16130.

Ge, L. L., Xie, C. T., Tian, H. Q., and Russell, S. D. (2009). Distribution of calcium in the stigma and style of tobacco during pollen germination and tube elongation. Sex. Plant Reprod. 22, 87–96.

Gierth, M., and Mäser, P. (2007). Potassium transporters in plants - Involvement in K acquisition, redistribution and homeostasis. FEBS Letters 581, 2348–2356. doi:10.1016/j.febslet.2007.03.035.

Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. Biochem. J 405, 199–221.

Goldraij, A., Kondo, K., Lee, C. B., Hancock, C. N., Sivaguru, M., Vazquez-Santana, S., et al. (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439, 805–810.

Grienenberger, E., Kim, S. S., Lallemand, B., Geoffroy, P., Heintz, D., Souza, C. de A., et al. (2010). Analysis of TETRAKETIDE α-PYRONE REDUCTASE function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. Plant Cell 22, 4067–4083.

Guyon, V. N., Astwood, J. D., Garner, E. C., Dunker, A. K., and Taylor, L. P. (2000). Isolation and characterization of cDNAs expressed in the early stages of flavonol-induced pollen germination in petunia. Plant Physiol. 123, 699–710.

Hager, A. (2003). Role of the plasma membrane H^+-ATPase in auxin-induced elongation growth: historical and new aspects. J. Plant Res. 116, 483–505.

Hazzouri, K. M., Escobar, J. S., Ness, R. W., Killian Newman, L., Randle, A. M., Kalisz, S., et al. (2013). Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection, and evolution of biased gene conversion with an ongoing mating system shift. Evolution 67, 1263–1278.

Holtsford, T. P., and Ellstrand, N. C. (1992). Genetic and environmental variation in floral traits affecting outcrossing rate in *Clarkia tembloriensis* (Onagraceae). Evolution 46, 216–225. doi:10.11 11/j.1558-5646.1992.tb0199 6.x.

Jorgensen, R., and Arathi, H. S. (2013). Floral longevity and autonomous selfing are altered by pollination and water availability in *Collinsia heterophylla*. Ann. Bot. 112, 821–828.

Jürgens, A., Witt, T., and Gottsberger, G. (2002). Flower scent composition in night-flowering *Silene* species (Caryophyllaceae). Biochemical Systematics and Ecology 30, 383–397. doi:10.1016/s0305-1978(01)00106-5.

Kachroo, A., Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (2001). Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 293, 1824–1826.

Kalisz, S., Randle, A., Chaiffetz, D., Faigeles, M., Butera, A., and Beight, C. (2012). Dichogamy correlates with outcrossing rate and defines the selfing syndrome in the mixed-mating genus *Collinsia*. Ann. Bot. 109, 571–582.

Kalisz, S., Vogler, D., Fails, B., Finer, M., Shepard, E., Herman, T., et al. (1999). The mechanism of delayed selfing in *Collinsia verna* (Scrophulariaceae). Am. J. Bot. 86, 1239–1247.

Klepek, Y.-S., Geiger, D., Stadler, R., Klebl, F., Landouar-Arsivaud, L., Lemoine, R., et al. (2005). *Arabidopsis* POLYOL TRANSPORTER5, a new member of the monosaccharide transporter-like superfamily, mediates H^+-symport of numerous substrates, including myo-inositol, glycerol, and ribose. Plant Cell 17, 204–218.

Klepek, Y.-S., -S. Klepek, Y., Volke, M., Konrad, K. R., Wippel, K., Hoth, S., et al. (2010). *Arabidopsis thaliana* POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H symporters in pollen and young xylem cells. Journal of Experimental Botany 61, 537–550. doi:10.1093/jxb/erp322.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version 61.

Kölling, M., Kumari, P., and Bürstenbinder, K. (2019). Calcium- and calmodulin-regulated microtubule-associated proteins as signal-integration hubs at the plasma membrane–cytoskeleton nexus. Journal of Experimental Botany 70, 387–396. doi:10.1093/jxb/ery397.

Krueger, F. (2017). Trim Galore!. Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

Kubo, K.I., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., et al. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. Science 330, 796–799.

Lafuma, L., and Maurice, S. (2007). Increase in mate availability without loss of self-incompatibility in the invasive species *Senecio inaequidens* (Asteraceae). Oikos 116, 201–208. doi:10.1111/j.2006.0030-1299.15220.x.

Lee, H. S., Huang, S., and Kao, T. (1994). S proteins control rejection of incompatible pollen in *Petunia inflata*. Nature 367, 560–563.

Lin, P.-C., Pomeranz, M. C., Jikumaru, Y., Kang, S. G., Hah, C., Fujioka, S., et al. (2011). The *Arabidopsis* tandem zinc finger protein AtTZF1 affects ABA- and GA-mediated growth, stress and gene expression responses. Plant J. 65, 253–268.

Liu, N., Wu, S., Van Houten, J., Wang, Y., Ding, B., Fei, Z., et al. (2014). Down-regulation of *AUXIN RESPONSE FACTORS 6* and *8* by microRNA 167 leads to floral development defects and female sterility in tomato. J. Exp. Bot. 65, 2507–2520.

Lloyd, D. G. (1992). Self- and Cross-Fertilization in Plants. II. The Selection of Self- Fertilization. International Journal of Plant Sciences 153, 370–380. doi:10.1086/297041.

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

Lozada-Gobilard, S., Weigend, M., Fischer, E., Janssens, S. B., Ackermann, M., and Abrahamczyk, S. (2019). Breeding systems in Balsaminaceae in relation to pollen/ovule ratio, pollination syndromes, life history and climate zone. Plant Biol. 21, 157–166.

Mäser, P., Hosoo, Y., Goshima, S., Horie, T., Eckelman, B., Yamada, K., et al. (2002). Glycine residues in potassium channel-like selectivity filters determine potassium selectivity in four-loop-per-subunit HKT transporters from plants. Proc. Natl. Acad. Sci. U. S. A. 99, 6428–6433.

McClure, B. A., Gray, J. E., Anderson, M. A., and Clarke, A. E. (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature 347, 757–760. doi:10.1038/347757a0.

McClure, B., Cruz-García, F., and Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. Ann. Bot. 108, 647–658.

Michard, E., Lima, P. T., Borges, F., Silva, A. C., Portes, M. T., Carvalho, J. E., et al. (2011). Glutamate receptor-like genes form Ca^{2+} channels in pollen tubes and are regulated by pistil D-serine. Science 332, 434–437.

Murfett, J., Atherton, T. L., Mou, B., Gasser, C. S., and McClure, B. A. (1994). S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. Nature 367, 563–566.

Nasrallah, J. B., Liu, P., Sherman-Broyles, S., Schmidt, R., and Nasrallah, M. E. (2007). Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. Genetics 175, 1965–1973.

Neuwirth, E., and Brewer, R. C. (2014). ColorBrewer palettes. R package version, 1–1.

Nyblom, M., Frick, A., Wang, Y., Ekvall, M., Hallgren, K., Hedfalk, K., et al. (2009). Structural and functional analysis of SoPIP2;1 mutants adds insight into plant aquaporin gating. J. Mol. Biol. 387, 653–668.

O'Brien, M., Bertrand, C., and Matton, D. P. (2002). Characterization of a fertilization-induced and developmentally regulated plasma-membrane aquaporin expressed in reproductive tissues, in the wild potato *Solanum chacoense* Bitt. Planta 215, 485–493.

Pan, C., Ye, L., Zheng, Y., Wang, Y., Yang, D., Liu, X., et al. (2017). Identification and expression profiling of microRNAs involved in the stigma exsertion under high-temperature stress in tomato. BMC Genomics 18, 843.

Pitann, B., Schubert, S., and Mühling, K. H. (2009). Decline in leaf growth under salt stress is due to an inhibition of H -pumping activity and increase in apoplastic pH of maize leaves. Journal of Plant Nutrition and Soil Science 172, 535–543. doi:10.1002/jpln.200800349.

Pott, M. B., Pichersky, E., and Piechulla, B. (2002). Evening specific oscillations of scent emission, SAMT enzyme activity, and SAMT mRNA in flowers of *Stephanotis floribunda*. Journal of Plant Physiology 159, 925–934. doi:10.1078/0176-1617-00699.

Randle, A. M., Slyder, J. B., and Kalisz, S. (2009). Can differences in autonomous selfing ability explain differences in range size among sister-taxa pairs of *Collinsia* (Plantaginaceae)? An extension of Baker's Law. New Phytol. 183, 618–629.

Rato, C., Monteiro, D., Hepler, P. K., and Malhó, R. (2004). Calmodulin activity and cAMP signalling modulate growth and apical secretion in pollen tubes. Plant J. 38, 887–897.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: URL https://www.R-project.org.

Seitz, C., Eder, C., Deiml, B., Kellner, S., Martens, S., and Forkmann, G. (2006). Cloning, Functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the Asteraceae family. Plant Molecular Biology 61, 365–381. doi:10.1007/s11103-006-0012-0.

Shimosato, H., Yokota, N., Shiba, H., Iwano, M., Entani, T., Che, F.S., et al. (2007). Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in *Brassica* self-incompatibility. Plant Cell 19, 107–117.

Shore, J. S., and Barrett, S. C. H. (1990). Quantitative genetics of floral characters in homostylous *Turnera ulmifolia* var. *angustifolia* Willd. (Turneraceae). Heredity 64, 105–112. doi:10.1038/hdy.1990.13.

Sicard, A., and Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Annals of Botany 107, 1433–1443. doi:10.1093/aob/mcr023.

Steige, K. A., Reimegård, J., Koenig, D., Scofield, D. G., and Slotte, T. (2015). Cis-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. Mol. Biol. Evol. 32, 2501–2514.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., et al. (2001). Direct ligand–receptor complex interaction controls *Brassica* self-incompatibility. Nature 413, 534–538. doi:10.1038/35097104.

Tan, J., Wang, M., Tu, L., Nie, Y., Lin, Y., and Zhang, X. (2013). The flavonoid pathway regulates the petal colors of cotton flower. PLoS One 8, e72364.

Thomas, S. G., and Franklin-Tong, V. E. (2004). Self-incompatibility triggers programmed cell death in *Papaver* pollen. Nature 429, 305–309.

Tian, Y., Xiao, S., Liu, J., Somaratne, Y., Zhang, H., Wang, M., et al. (2017). MALE STERILE6021 (MS6021) is required for the development of anther cuticle and pollen exine in maize. Sci. Rep. 7, 16736.

Vos, J. M. de, de Vos, J. M., Wüest, R. O., and Conti, E. (2014). Small and ugly? Phylogenetic analyses of the "selfing syndrome" reveal complex evolutionary fates of monomorphic primrose flowers. Evolution 68, 1042–1057. doi:10.1111/evo.12331.

Wang, J., Hu, Z., Zhao, T., Yang, Y., Chen, T., Yang, M., et al. (2015). Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (*Solanum lycopersicum*). BMC Genomics 16, 39.

Wheeler, M. J., Vatovec, S., and Franklin-Tong, V. E. (2010). The pollen S-determina nt in *Papaver*: comparisons with known plant receptors and protein ligand partners. J. Exp. Bot. 61, 2015–2025.

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

Wilkins, K. A., Bosch, M., Haque, T., Teng, N., Poulter, N. S., and Franklin-Tong, V. E. (2015). Self-incompatibility-induced programmed cell death in field poppy pollen involves dramatic acidification of the incompatible pollen tube cytosol. Plant Physiol. 167, 766–779.

Wood, B. W. (2017). Flavonoids, alkali earth, and rare earth elements affect pecan pollen germination. HortScience 52, 85–88. doi:10.21273/hortsci1 1426-16.

Yule, B. L., Roberts, S., and Marshall, J. E. A. (2000). The thermal evolution of sporopollenin. Organic Geochemistry 31, 859–870. doi:10.1016/s0146-6380(00)00058-9.

Figure 3.1. PCA plot of expression patterns between *C. linearis* and *C. rattanii*.

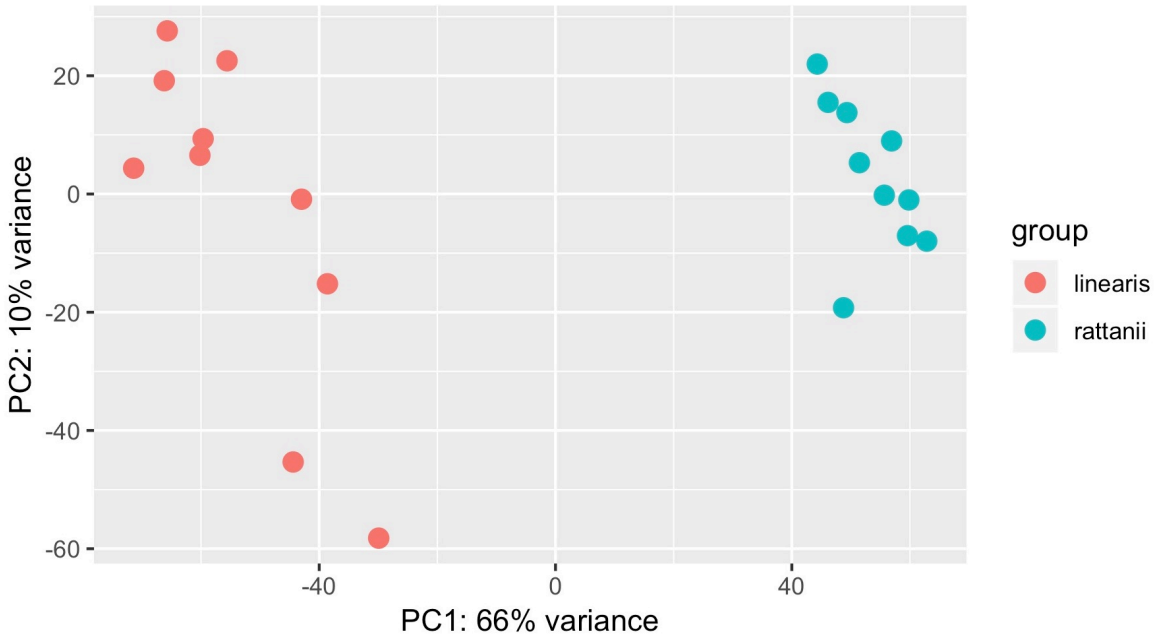
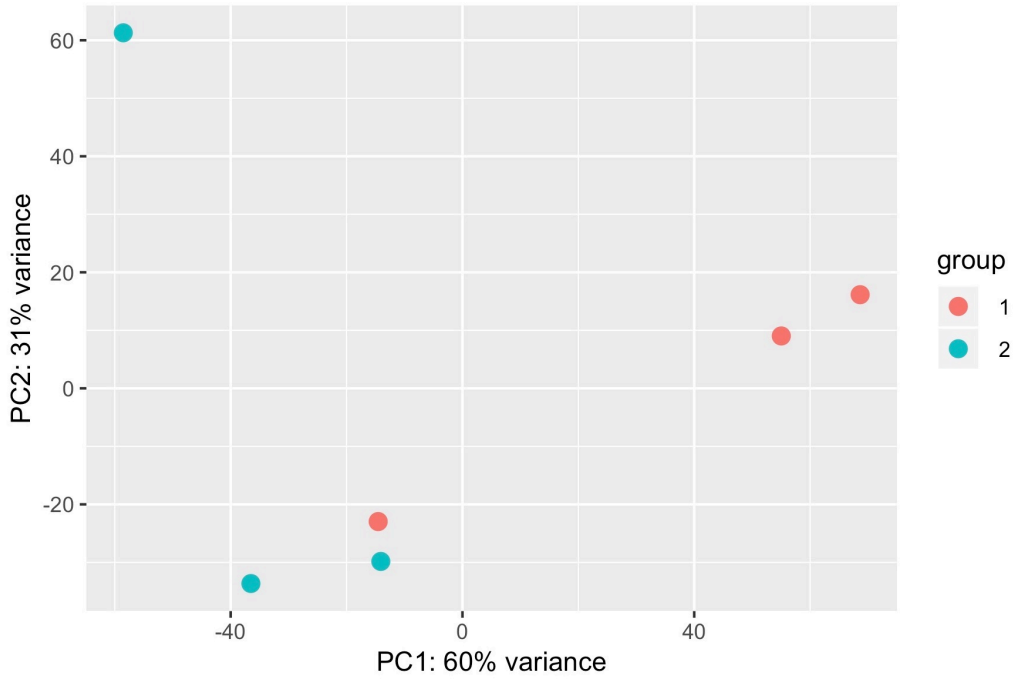


Figure 3.2. PCA plot of expression patterns for intraspecific comparisons in *C. linearis* (a) Stage 1 vs stage 2 (b) Stage 2 vs stage 3.

(a)



(b)

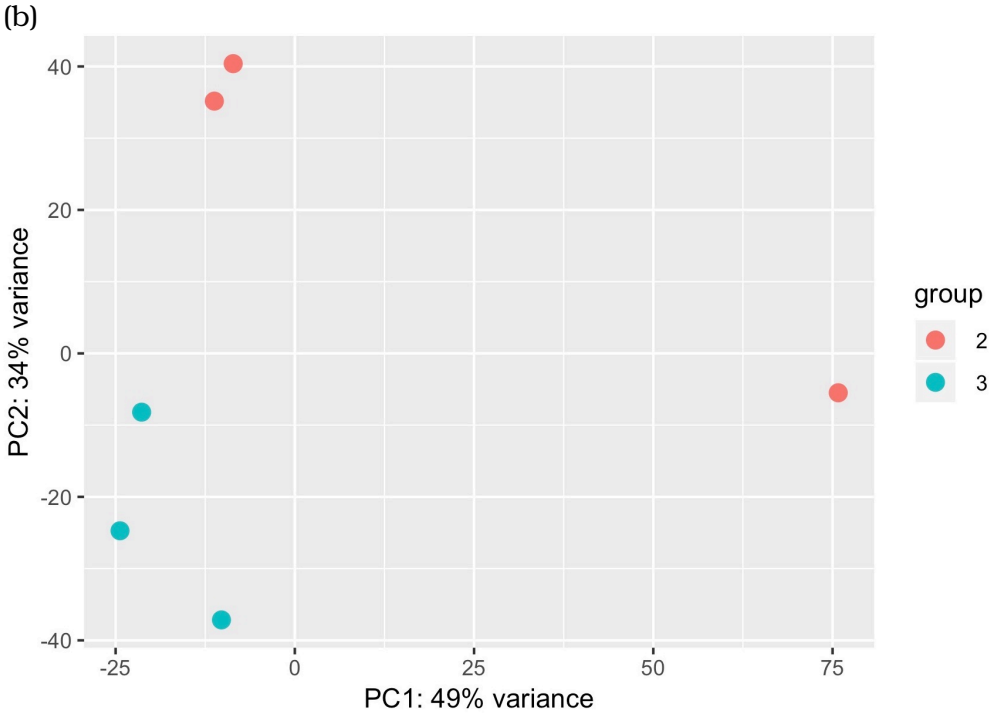
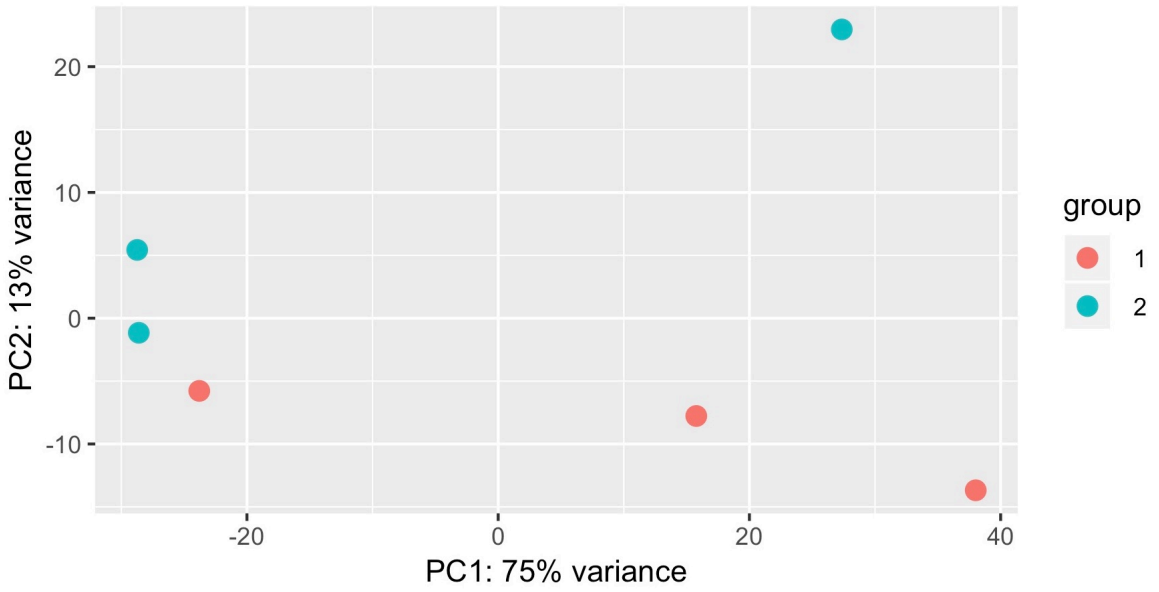


Figure 3.3. PCA plot of expression patterns for intraspecific comparisons in *C. rattanii* (a) Stage 1 vs stage 2 (b) Stage 2 vs stage 3.

(a)



(b)

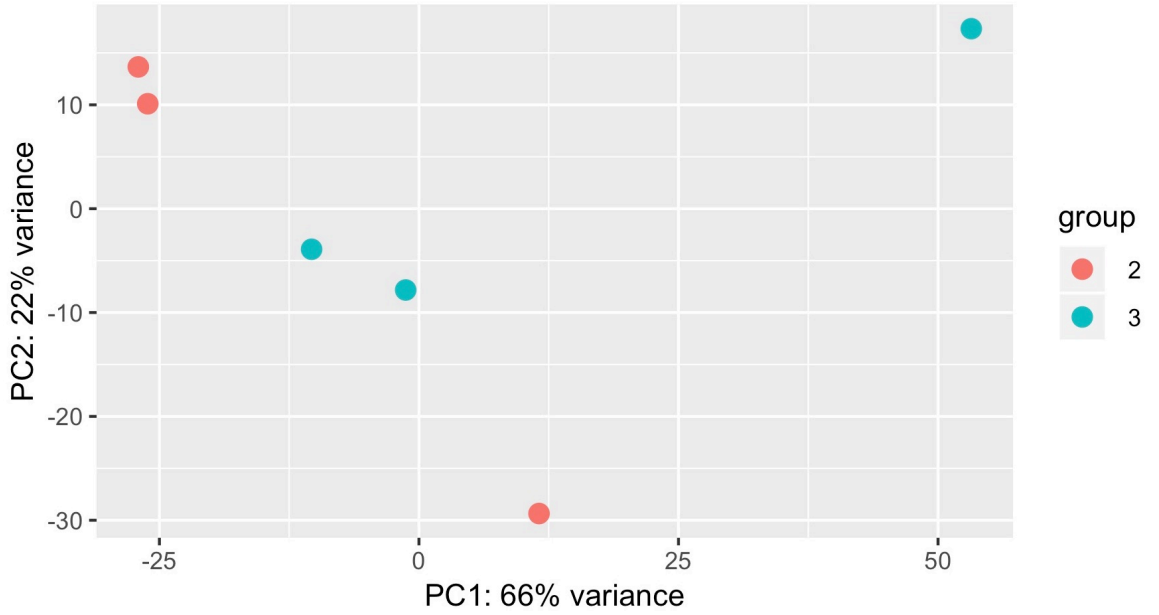


Figure 3.4. The numbers of differentially expressed genes in intraspecific comparisons.

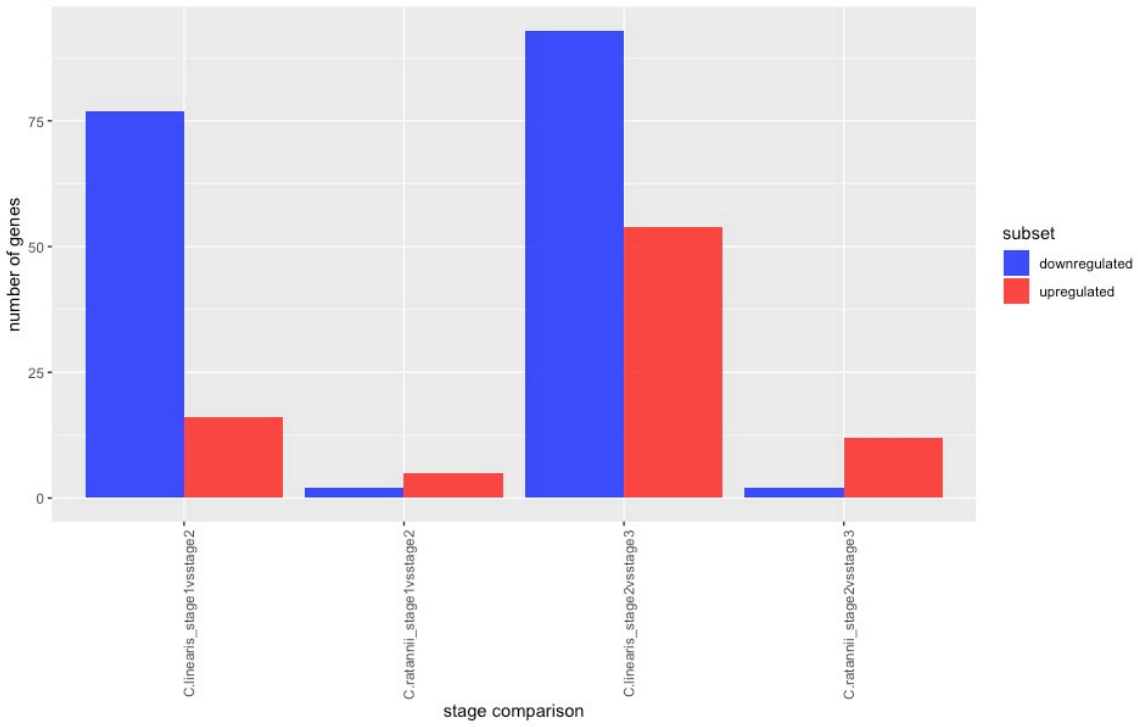


Figure 3.5. Heatmap of DE genes in *C. linearis* stage 1 vs stage 2.

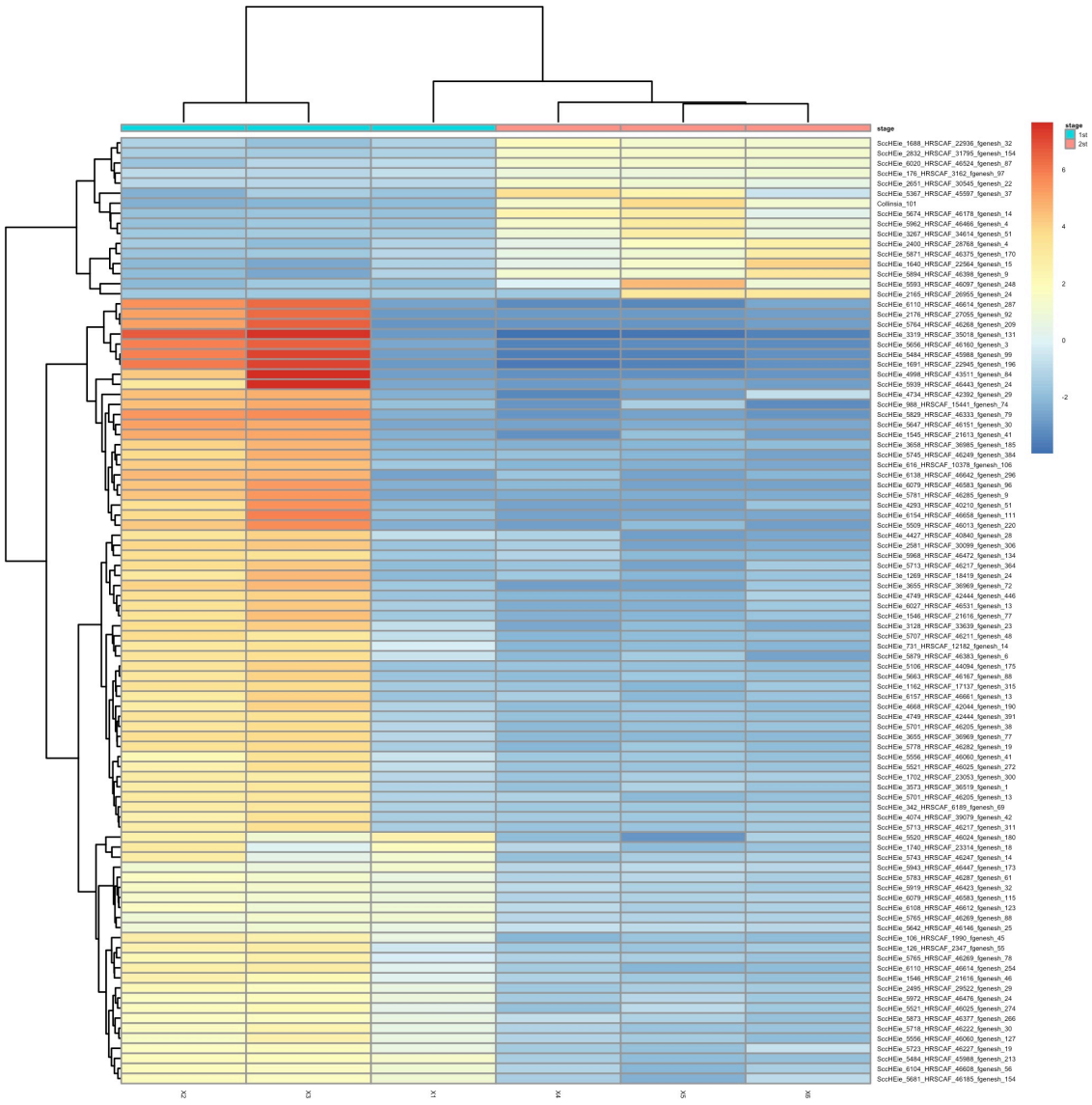


Figure 3.6. Heatmap of DE genes in *C. linearis* stage 2 vs stage 3.

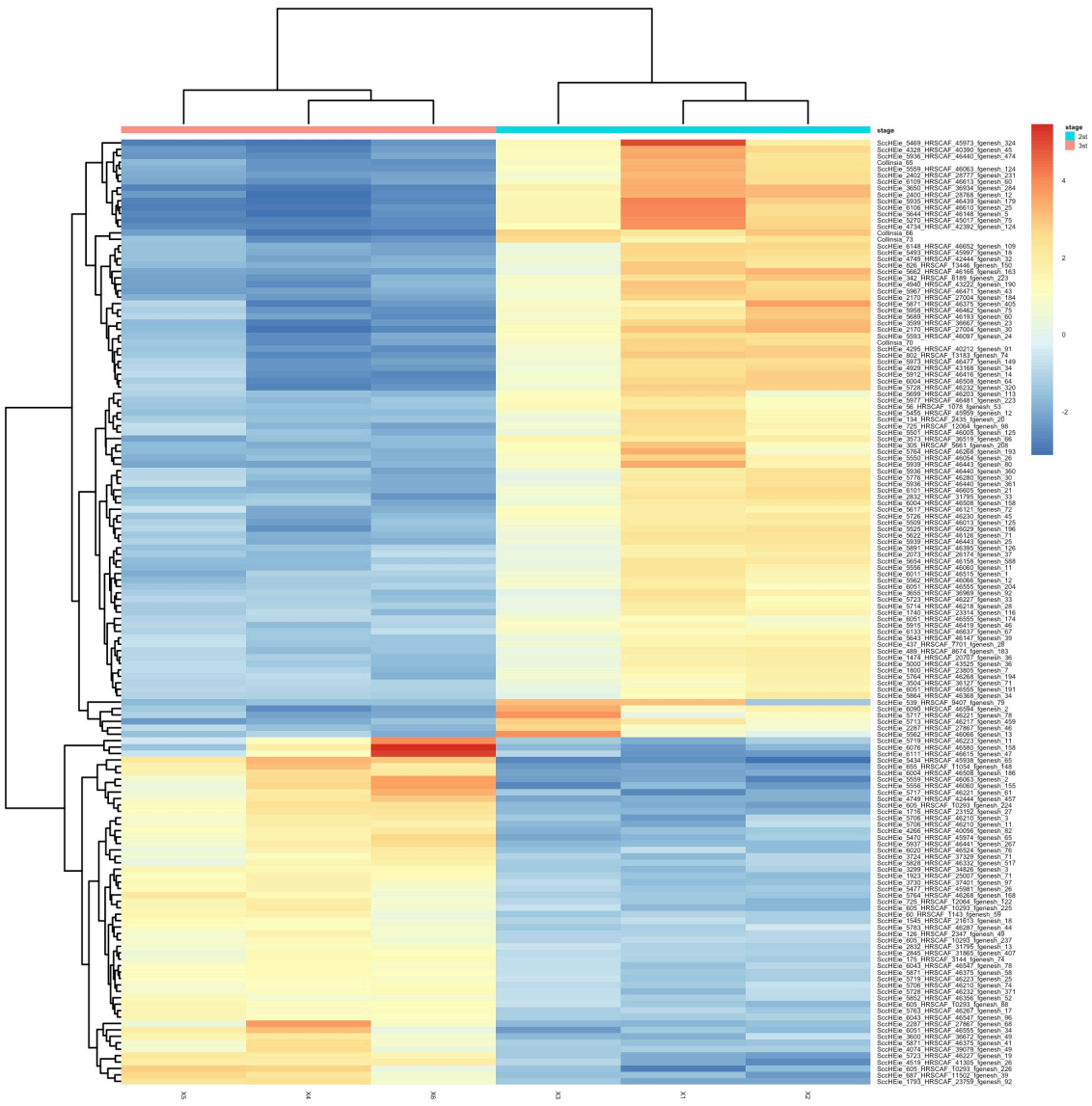


Figure 3.7. Heatmap of DE genes in *C. rattanii* stage 1 vs stage 2.

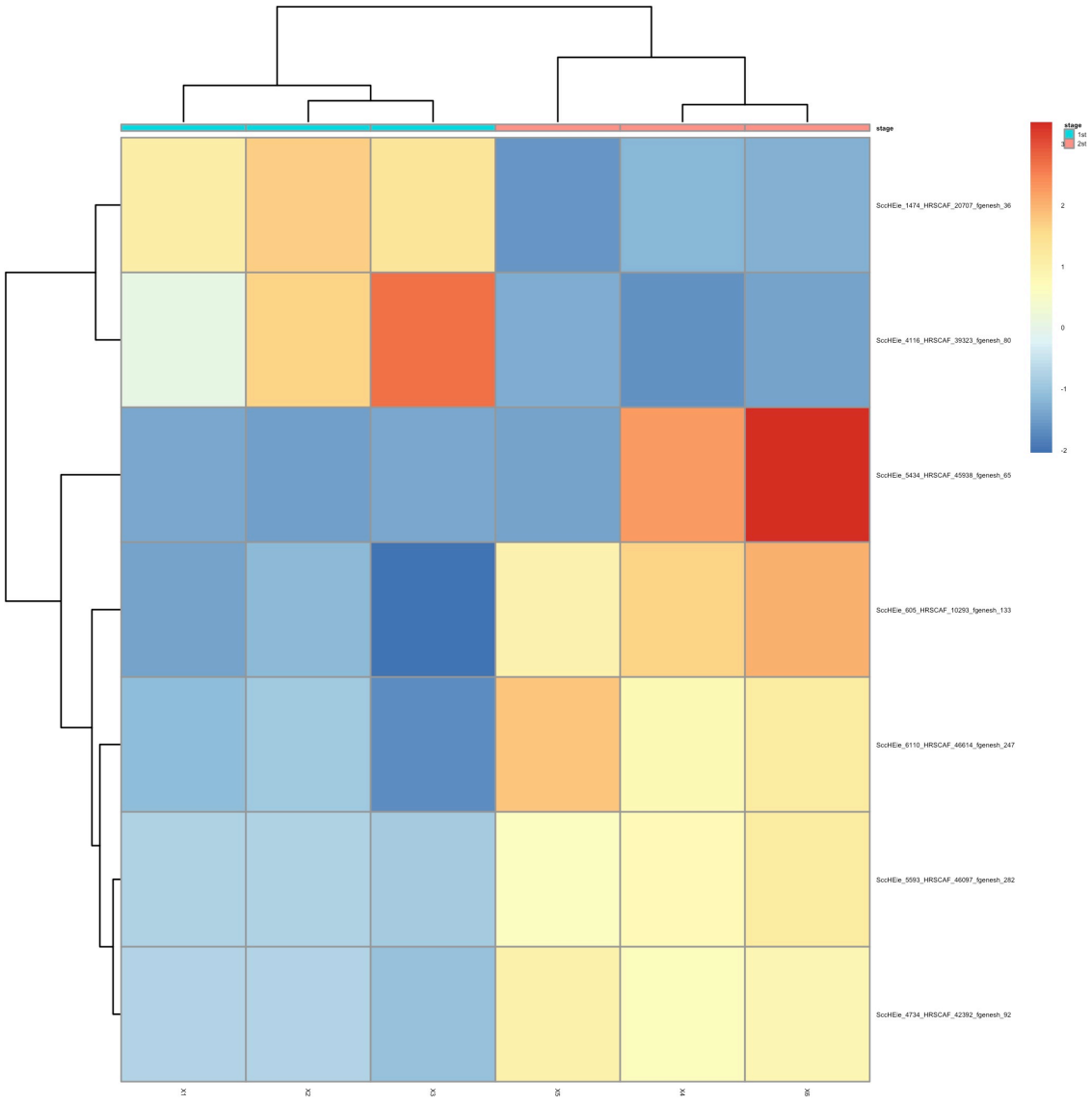


Figure 3.8. Heatmap of DE genes in *C. rattanii* stage 2 vs stage 3.

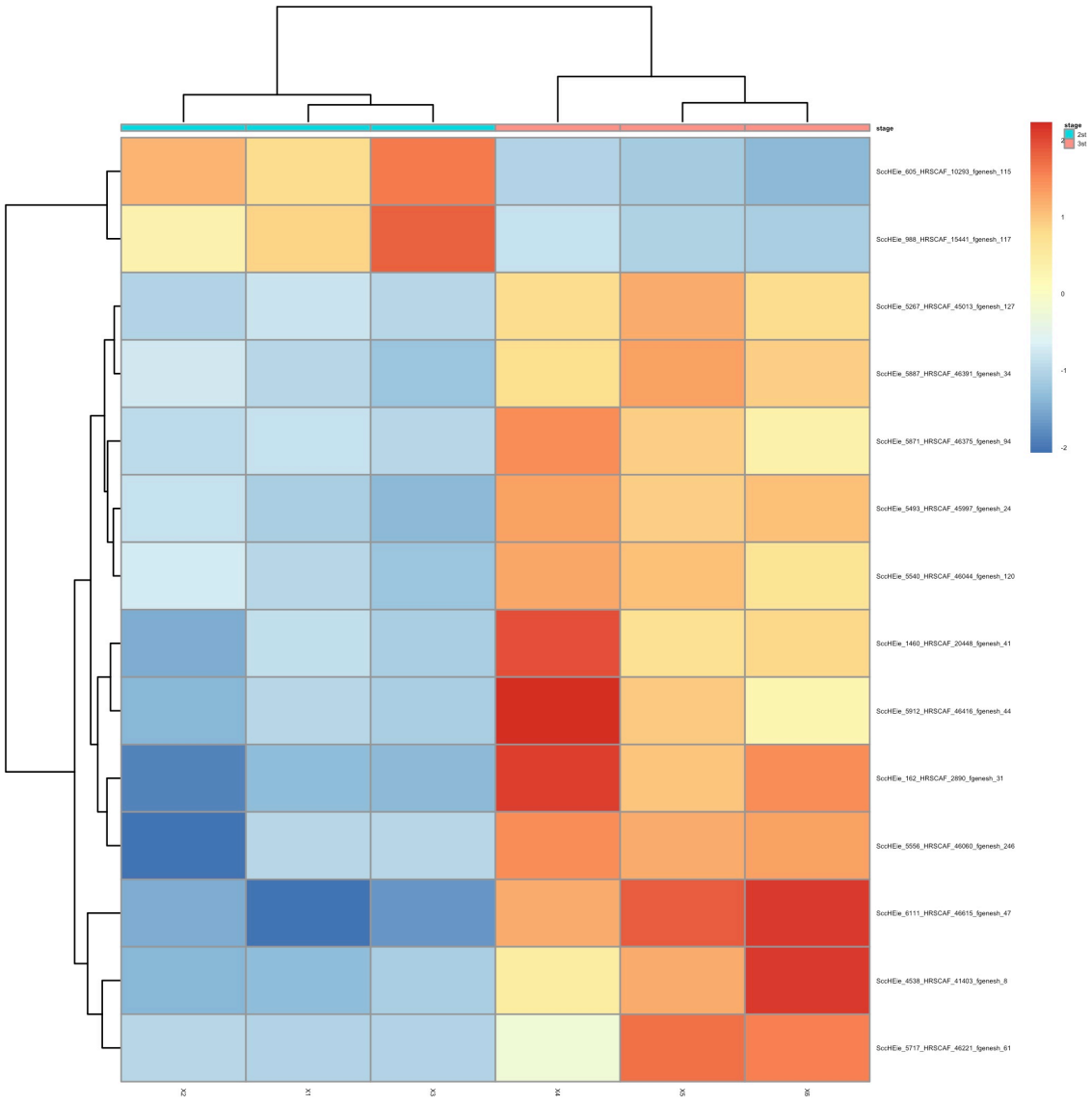


Table 3.1. Up- (red) and downregulated (blue) genes in *C. linearis* stage 2 compared to stage 1.

Gene	log2FoldChange
Collinsia_101	9.65616532696977
SccHEie_5674_HRSCAF_46178_fgenesh_14	9.42114305074839
SccHEie_5593_HRSCAF_46097_fgenesh_248	9.29989399978433
SccHEie_5962_HRSCAF_46466_fgenesh_4	9.12701917781224
SccHEie_3267_HRSCAF_34614_fgenesh_51	8.65067972681036
SccHEie_2651_HRSCAF_30545_fgenesh_22	7.28866146697443
SccHEie_2165_HRSCAF_26955_fgenesh_24	7.074144105485
SccHEie_5367_HRSCAF_45597_fgenesh_37	7.06133818943285
SccHEie_2832_HRSCAF_31795_fgenesh_154	6.50666433500444
SccHEie_1640_HRSCAF_22564_fgenesh_15	6.48445885534821
SccHEie_5894_HRSCAF_46398_fgenesh_9	6.07318387569094
SccHEie_2400_HRSCAF_28768_fgenesh_4	5.71507174532578
SccHEie_5871_HRSCAF_46375_fgenesh_170	4.71638087883586
SccHEie_1688_HRSCAF_22936_fgenesh_32	4.34254983857597
SccHEie_6020_HRSCAF_46524_fgenesh_87	3.35550700146595
SccHEie_176_HRSCAF_3162_fgenesh_97	2.73183020642493
SccHEie_5642_HRSCAF_46146_fgenesh_25	-2.52946091369603
SccHEie_5783_HRSCAF_46287_fgenesh_61	-3.67421395956235
SccHEie_5723_HRSCAF_46227_fgenesh_19	-3.79697539328439
SccHEie_5681_HRSCAF_46185_fgenesh_154	-4.13013213353225
SccHEie_5873_HRSCAF_46377_fgenesh_266	-4.1313269753876
SccHEie_5919_HRSCAF_46423_fgenesh_32	-4.31547425418567
SccHEie_6108_HRSCAF_46612_fgenesh_123	-4.43019048476856
SccHEie_6104_HRSCAF_46608_fgenesh_56	-4.66426666852325
SccHEie_6079_HRSCAF_46583_fgenesh_115	-4.69696476785922
SccHEie_1546_HRSCAF_21616_fgenesh_46	-4.9474581575352
SccHEie_1740_HRSCAF_23314_fgenesh_18	-5.00892988966544
SccHEie_5521_HRSCAF_46025_fgenesh_274	-5.23461496809292
SccHEie_5765_HRSCAF_46269_fgenesh_78	-5.33668507565675
SccHEie_5718_HRSCAF_46222_fgenesh_30	-5.43141728268757
SccHEie_5972_HRSCAF_46476_fgenesh_24	-5.46048659190605
SccHEie_4074_HRSCAF_39079_fgenesh_42	-5.70918554460386
SccHEie_5520_HRSCAF_46024_fgenesh_180	-5.71329515409889
SccHEie_1702_HRSCAF_23053_fgenesh_300	-5.74582971449708
SccHEie_5556_HRSCAF_46060_fgenesh_127	-5.84291953247281
SccHEie_5484_HRSCAF_45988_fgenesh_213	-5.98972878908639
SccHEie_126_HRSCAF_2347_fgenesh_55	-6.25442654697014
SccHEie_5743_HRSCAF_46247_fgenesh_14	-6.27751070522186
SccHEie_5968_HRSCAF_46472_fgenesh_134	-6.38769294182651
SccHEie_4749_HRSCAF_42444_fgenesh_391	-6.47277286300601
SccHEie_6157_HRSCAF_46661_fgenesh_13	-6.53018833899535

Continuation of Table 3.1

Gene	log2FoldChange
SccHEie_5765_HRSCAF_46269_fgenesh_88	-6.70471294059602
SccHEie_5556_HRSCAF_46060_fgenesh_41	-6.70678992063142
SccHEie_1162_HRSCAF_17137_fgenesh_315	-6.74761509715331
SccHEie_6110_HRSCAF_46614_fgenesh_254	-6.79715366577542
SccHEie_5701_HRSCAF_46205_fgenesh_13	-6.85783981772952
SccHEie_5106_HRSCAF_44094_fgenesh_175	-6.88628016688879
SccHEie_4749_HRSCAF_42444_fgenesh_446	-7.16180686497434
SccHEie_4427_HRSCAF_40840_fgenesh_28	-7.17204123019594
SccHEie_1269_HRSCAF_18419_fgenesh_24	-7.23570913790791
SccHEie_5943_HRSCAF_46447_fgenesh_173	-7.24522237460064
SccHEie_5879_HRSCAF_46383_fgenesh_6	-7.25399160635653
SccHEie_5713_HRSCAF_46217_fgenesh_364	-7.38568926732982
SccHEie_6027_HRSCAF_46531_fgenesh_13	-7.57054749396397
SccHEie_5713_HRSCAF_46217_fgenesh_311	-7.70098821367812
SccHEie_2581_HRSCAF_30099_fgenesh_306	-7.73035554871062
SccHEie_4734_HRSCAF_42392_fgenesh_29	-7.90849306373285
SccHEie_342_HRSCAF_6189_fgenesh_69	-7.91249525603482
SccHEie_1546_HRSCAF_21616_fgenesh_77	-8.08152396154913
SccHEie_106_HRSCAF_1990_fgenesh_45	-8.09688701427072
SccHEie_5521_HRSCAF_46025_fgenesh_272	-8.21822164464137
SccHEie_5663_HRSCAF_46167_fgenesh_88	-8.34418027574783
SccHEie_5745_HRSCAF_46249_fgenesh_384	-8.36298581873649
SccHEie_3658_HRSCAF_36985_fgenesh_185	-8.37839383374182
SccHEie_2495_HRSCAF_29522_fgenesh_29	-8.41698690432295
SccHEie_3573_HRSCAF_36519_fgenesh_1	-8.45614446748825
SccHEie_988_HRSCAF_15441_fgenesh_74	-8.81222681976341
SccHEie_3655_HRSCAF_36969_fgenesh_72	-8.85609746248862
SccHEie_6138_HRSCAF_46642_fgenesh_296	-8.87554645512301
SccHEie_5778_HRSCAF_46282_fgenesh_19	-8.92789230702018
SccHEie_4293_HRSCAF_40210_fgenesh_51	-9.0612641040647
SccHEie_5509_HRSCAF_46013_fgenesh_220	-9.26350831492687
SccHEie_3655_HRSCAF_36969_fgenesh_77	-9.31910045615398
SccHEie_5701_HRSCAF_46205_fgenesh_38	-9.39255511851596
SccHEie_5707_HRSCAF_46211_fgenesh_48	-9.62067183347039
SccHEie_6079_HRSCAF_46583_fgenesh_96	-9.85966324410449
SccHEie_4668_HRSCAF_42044_fgenesh_190	-10.2100392237863
SccHEie_6154_HRSCAF_46658_fgenesh_111	-10.2299222776485
SccHEie_2176_HRSCAF_27055_fgenesh_92	-10.4139702878482
SccHEie_731_HRSCAF_12182_fgenesh_14	-10.4398288890367
SccHEie_616_HRSCAF_10378_fgenesh_106	-10.5329037479975
SccHEie_1545_HRSCAF_21613_fgenesh_41	-10.7679542029632
SccHEie_5781_HRSCAF_46285_fgenesh_9	-10.8173454286118
SccHEie_3128_HRSCAF_33639_fgenesh_23	-10.8213497366139

Continuation of Table 3.1

Gene	log2FoldChange
SccHEie_6110_HRSCAF_46614_fgenesh_287	-10.9829392859345
SccHEie_5647_HRSCAF_46151_fgenesh_30	-11.1665887847855
SccHEie_5764_HRSCAF_46268_fgenesh_209	-11.3247611162853
SccHEie_5939_HRSCAF_46443_fgenesh_24	-11.9580750029435
SccHEie_4998_HRSCAF_43511_fgenesh_84	-12.0915417827414
SccHEie_1691_HRSCAF_22945_fgenesh_196	-13.3776711801085
SccHEie_5484_HRSCAF_45988_fgenesh_99	-13.5662714992973
SccHEie_5829_HRSCAF_46333_fgenesh_79	-13.9223094774217
SccHEie_3319_HRSCAF_35018_fgenesh_131	-14.0163925897725
SccHEie_5656_HRSCAF_46160_fgenesh_3	-14.3459664067061

Table 3.2. Up- (red) and downregulated (blue) genes in *C. linearis* stage 3 compared to stage 2.

Gene	log2FoldChange
SccHEie_6004_HRSCAF_46508_fgenesh_186	10.6073474554594
SccHEie_2287_HRSCAF_27867_fgenesh_68	9.91020641727335
SccHEie_4749_HRSCAF_42444_fgenesh_457	9.45534416004861
SccHEie_5434_HRSCAF_45938_fgenesh_65	9.14098822712935
SccHEie_6076_HRSCAF_46580_fgenesh_158	8.80198130154678
SccHEie_655_HRSCAF_11054_fgenesh_148	8.73309274337281
SccHEie_5937_HRSCAF_46441_fgenesh_267	8.72689770709103
SccHEie_6043_HRSCAF_46547_fgenesh_96	8.19245130932355
SccHEie_5556_HRSCAF_46060_fgenesh_155	7.78726430297833
SccHEie_5559_HRSCAF_46063_fgenesh_2	7.7697501673682
SccHEie_5828_HRSCAF_46332_fgenesh_517	7.32556113193233
SccHEie_5871_HRSCAF_46375_fgenesh_41	7.26340689305052
SccHEie_1716_HRSCAF_23152_fgenesh_27	7.0947229696293
SccHEie_6111_HRSCAF_46615_fgenesh_47	7.00578286718433
SccHEie_1793_HRSCAF_23759_fgenesh_92	6.74818442337482
SccHEie_605_HRSCAF_10293_fgenesh_224	6.64796979340165
SccHEie_5717_HRSCAF_46221_fgenesh_61	6.56571941958479
SccHEie_5470_HRSCAF_45974_fgenesh_65	6.26757229773376
SccHEie_605_HRSCAF_10293_fgenesh_226	6.15919569678103
SccHEie_5719_HRSCAF_46223_fgenesh_11	5.94658661406661
SccHEie_687_HRSCAF_11502_fgenesh_39	5.92052078298304
SccHEie_6051_HRSCAF_46555_fgenesh_34	5.83951921281384
SccHEie_5723_HRSCAF_46227_fgenesh_19	5.42661439780109
SccHEie_4074_HRSCAF_39079_fgenesh_49	5.35633077455762
SccHEie_4519_HRSCAF_41305_fgenesh_26	5.09762790060573
SccHEie_4266_HRSCAF_40056_fgenesh_82	4.74685611297113
SccHEie_5706_HRSCAF_46210_fgenesh_3	4.57595973740826
SccHEie_5477_HRSCAF_45981_fgenesh_26	4.5404947727614
SccHEie_5764_HRSCAF_46268_fgenesh_168	4.43254273980421
SccHEie_3600_HRSCAF_36672_fgenesh_49	4.2511085476946
SccHEie_5706_HRSCAF_46210_fgenesh_11	4.19719750142762
SccHEie_605_HRSCAF_10293_fgenesh_225	4.161658585878
SccHEie_3299_HRSCAF_34826_fgenesh_3	4.11374513289792
SccHEie_1923_HRSCAF_25007_fgenesh_71	4.07570118381615
SccHEie_175_HRSCAF_3144_fgenesh_74	4.01238926872708
SccHEie_725_HRSCAF_12064_fgenesh_122	3.99240452610997
SccHEie_6020_HRSCAF_46524_fgenesh_76	3.94318583889918
SccHEie_3730_HRSCAF_37401_fgenesh_97	3.9374041747971
SccHEie_60_HRSCAF_1143_fgenesh_59	3.65115501418349
SccHEie_3724_HRSCAF_37329_fgenesh_71	3.64676124078553
SccHEie_605_HRSCAF_10293_fgenesh_88	3.49645282556966

Continuation of Table 3.2

Gene	log2FoldChange
SccHEie_5763_HRSCAF_46267_fgenesh_17	3.35784100458406
SccHEie_5728_HRSCAF_46232_fgenesh_371	3.32587572607709
SccHEie_5871_HRSCAF_46375_fgenesh_58	3.22245355836023
SccHEie_1545_HRSCAF_21613_fgenesh_18	3.21896494319414
SccHEie_6043_HRSCAF_46547_fgenesh_78	3.10776898314454
SccHEie_126_HRSCAF_2347_fgenesh_49	3.09527193289576
SccHEie_5719_HRSCAF_46223_fgenesh_25	3.02663272103815
SccHEie_605_HRSCAF_10293_fgenesh_237	2.97353343679793
SccHEie_2845_HRSCAF_31865_fgenesh_407	2.96663263261529
SccHEie_5783_HRSCAF_46287_fgenesh_44	2.95075934715508
SccHEie_5852_HRSCAF_46356_fgenesh_52	2.79499385960705
SccHEie_2832_HRSCAF_31795_fgenesh_13	2.74823088299998
SccHEie_5706_HRSCAF_46210_fgenesh_74	2.70307378414422
SccHEie_3504_HRSCAF_36127_fgenesh_71	-3.1573661312281
SccHEie_6133_HRSCAF_46637_fgenesh_67	-3.23967398662957
SccHEie_5915_HRSCAF_46419_fgenesh_46	-3.39509383644191
SccHEie_5643_HRSCAF_46147_fgenesh_39	-3.61804249958008
SccHEie_5764_HRSCAF_46268_fgenesh_194	-3.63669490062081
SccHEie_1800_HRSCAF_23805_fgenesh_7	-3.68601020207221
SccHEie_2073_HRSCAF_26174_fgenesh_37	-3.78697360347084
SccHEie_1474_HRSCAF_20707_fgenesh_36	-3.94703492742472
SccHEie_5714_HRSCAF_46218_fgenesh_28	-3.96212296774949
SccHEie_5000_HRSCAF_43525_fgenesh_36	-4.00056574044992
SccHEie_489_HRSCAF_8674_fgenesh_183	-4.10751923593145
SccHEie_437_HRSCAF_7701_fgenesh_28	-4.11015053964474
SccHEie_5891_HRSCAF_46395_fgenesh_126	-4.15369467331683
SccHEie_5617_HRSCAF_46121_fgenesh_72	-4.31381911274486
SccHEie_5562_HRSCAF_46066_fgenesh_12	-4.32057441485385
SccHEie_6011_HRSCAF_46515_fgenesh_1	-4.33045140581648
SccHEie_5723_HRSCAF_46227_fgenesh_33	-4.3389724681649
SccHEie_6051_HRSCAF_46555_fgenesh_174	-4.38324809566009
SccHEie_5501_HRSCAF_46005_fgenesh_125	-4.4028384399176
SccHEie_5977_HRSCAF_46481_fgenesh_223	-4.41426794799187
SccHEie_3655_HRSCAF_36969_fgenesh_92	-4.45240325498099
SccHEie_5556_HRSCAF_46060_fgenesh_11	-4.47365337456424
SccHEie_2287_HRSCAF_27867_fgenesh_46	-4.55676500393548
SccHEie_725_HRSCAF_12064_fgenesh_98	-4.5787521419055
SccHEie_5936_HRSCAF_46440_fgenesh_361	-4.64455661106825
SccHEie_1740_HRSCAF_23314_fgenesh_116	-4.6636728886648
SccHEie_6004_HRSCAF_46508_fgenesh_158	-4.68264444442204
SccHEie_5936_HRSCAF_46440_fgenesh_360	-4.73091073660784
SccHEie_5509_HRSCAF_46013_fgenesh_125	-4.78013236887185
SccHEie_5699_HRSCAF_46203_fgenesh_113	-4.81016562085483

Continuation of Table 3.2

Gene	log2FoldChange
SccHEie_5654_HRSCAF_46158_fgenesh_588	-4.88883181144894
SccHEie_5525_HRSCAF_46029_fgenesh_196	-4.98986436789968
SccHEie_5776_HRSCAF_46280_fgenesh_30	-5.04676672546846
SccHEie_5939_HRSCAF_46443_fgenesh_25	-5.13019002640145
SccHEie_134_HRSCAF_2435_fgenesh_20	-5.22673152689246
SccHEie_5726_HRSCAF_46230_fgenesh_45	-5.27349379351872
SccHEie_2832_HRSCAF_31795_fgenesh_33	-5.39666403773701
SccHEie_5689_HRSCAF_46193_fgenesh_60	-5.51830734662087
SccHEie_6051_HRSCAF_46555_fgenesh_204	-5.52966816564778
SccHEie_5550_HRSCAF_46054_fgenesh_26	-5.56562969857931
SccHEie_5973_HRSCAF_46477_fgenesh_149	-5.70028800205098
SccHEie_5912_HRSCAF_46416_fgenesh_14	-5.73563765026643
SccHEie_4929_HRSCAF_43168_fgenesh_34	-5.73938065170265
SccHEie_4749_HRSCAF_42444_fgenesh_32	-5.8442726667123
SccHEie_3573_HRSCAF_36519_fgenesh_66	-5.85309830485236
Collinsia_70	-6.07421021100478
SccHEie_5728_HRSCAF_46232_fgenesh_320	-6.07783367587882
SccHEie_6148_HRSCAF_46652_fgenesh_109	-6.12693056481345
SccHEie_5622_HRSCAF_46126_fgenesh_71	-6.18906202491109
SccHEie_5958_HRSCAF_46462_fgenesh_75	-6.20603225052701
SccHEie_5562_HRSCAF_46066_fgenesh_13	-6.23780971149568
SccHEie_6004_HRSCAF_46508_fgenesh_64	-6.25783540119862
SccHEie_5871_HRSCAF_46375_fgenesh_405	-6.30429955571724
SccHEie_5593_HRSCAF_46097_fgenesh_24	-6.34482986374958
Collinsia_73	-6.47659964174269
SccHEie_6090_HRSCAF_46594_fgenesh_2	-6.55756505322551
SccHEie_4295_HRSCAF_40212_fgenesh_91	-6.63478660090164
SccHEie_2170_HRSCAF_27004_fgenesh_184	-6.64209406176259
SccHEie_5713_HRSCAF_46217_fgenesh_459	-6.64394393050717
SccHEie_305_HRSCAF_5661_fgenesh_208	-6.64928508244864
SccHEie_5455_HRSCAF_45959_fgenesh_12	-6.74893539878093
SccHEie_342_HRSCAF_6189_fgenesh_223	-6.78029633656871
SccHEie_5967_HRSCAF_46471_fgenesh_43	-6.84304811941605
SccHEie_802_HRSCAF_13183_fgenesh_74	-6.91700327788121
SccHEie_5764_HRSCAF_46268_fgenesh_193	-6.9287537565951
SccHEie_4940_HRSCAF_43222_fgenesh_190	-6.9591615890514
SccHEie_6101_HRSCAF_46605_fgenesh_21	-6.96945359029661
SccHEie_826_HRSCAF_13446_fgenesh_150	-7.04682416770321
SccHEie_5717_HRSCAF_46221_fgenesh_78	-7.13262227074328
Collinsia_65	-7.17565283281654
SccHEie_3599_HRSCAF_36667_fgenesh_23	-7.26747392077257
SccHEie_2402_HRSCAF_28777_fgenesh_231	-7.37358749371261
SccHEie_6051_HRSCAF_46555_fgenesh_191	-7.47005489502444

Continuation of Table 3.2

Gene	log2FoldChange
SccHEie_6109_HRSCAF_46613_fgenesh_60	-7.54855561430133
SccHEie_2170_HRSCAF_27004_fgenesh_30	-7.600038818973
SccHEie_5939_HRSCAF_46443_fgenesh_80	-7.71169915283026
SccHEie_56_HRSCAF_1078_fgenesh_53	-7.71350695552517
SccHEie_5864_HRSCAF_46368_fgenesh_34	-7.72705248057935
SccHEie_539_HRSCAF_9407_fgenesh_79	-7.85350959559626
SccHEie_4328_HRSCAF_40390_fgenesh_45	-8.01141186862197
SccHEie_5493_HRSCAF_45997_fgenesh_18	-8.27383468831632
SccHEie_5936_HRSCAF_46440_fgenesh_474	-8.36547329281818
Collinsia_66	-8.45899845071041
SccHEie_5559_HRSCAF_46063_fgenesh_124	-8.72863017268904
SccHEie_2400_HRSCAF_28768_fgenesh_12	-8.87153847121628
SccHEie_5644_HRSCAF_46148_fgenesh_5	-9.59140568922709
SccHEie_6106_HRSCAF_46610_fgenesh_25	-9.80531600837956
SccHEie_3650_HRSCAF_36934_fgenesh_284	-9.97732081161831
SccHEie_5935_HRSCAF_46439_fgenesh_179	-10.1703258026581
SccHEie_5469_HRSCAF_45973_fgenesh_324	-10.2401446408493
SccHEie_5270_HRSCAF_45017_fgenesh_75	-10.5860806308374
SccHEie_5662_HRSCAF_46166_fgenesh_163	-10.9361799621522
SccHEie_4734_HRSCAF_42392_fgenesh_124	-11.7986879925781

Table 3.3. Up- (red) and downregulated (blue) genes in *C. rattanii* stage 2 compared to stage 1.

Gene	log2FoldChange
SccHEie_5434_HRSCAF_45938_fgenesh_65	7.05484160670367
SccHEie_605_HRSCAF_10293_fgenesh_133	5.2715681118212
SccHEie_6110_HRSCAF_46614_fgenesh_247	4.2739941436559
SccHEie_4734_HRSCAF_42392_fgenesh_92	3.07704402971413
SccHEie_5593_HRSCAF_46097_fgenesh_282	2.74276990338169
SccHEie_1474_HRSCAF_20707_fgenesh_36	-4.82818548144399
SccHEie_4116_HRSCAF_39323_fgenesh_80	-6.45138032047223

Table 3.4. Up- (red) and downregulated (blue) genes in *C. rattanii* stage 3 compared to stage 2.

Gene	log2FoldChange
SccHEie_5717_HRSCAF_46221_fgenesh_61	8.38436225638882
SccHEie_6111_HRSCAF_46615_fgenesh_47	5.7852468686438
SccHEie_162_HRSCAF_2890_fgenesh_31	5.23371060765546
SccHEie_5912_HRSCAF_46416_fgenesh_44	4.55881584379902
SccHEie_4538_HRSCAF_41403_fgenesh_8	4.45741537622483
SccHEie_1460_HRSCAF_20448_fgenesh_41	4.18757678894658
SccHEie_5493_HRSCAF_45997_fgenesh_24	4.06937992829676
SccHEie_5556_HRSCAF_46060_fgenesh_246	4.02596148188628
SccHEie_5267_HRSCAF_45013_fgenesh_127	3.62614306591631
SccHEie_5887_HRSCAF_46391_fgenesh_34	3.40224031742287
SccHEie_5871_HRSCAF_46375_fgenesh_94	3.26043291613767
SccHEie_5540_HRSCAF_46044_fgenesh_120	3.05603048649875
SccHEie_605_HRSCAF_10293_fgenesh_115	-3.85665324978671
SccHEie_988_HRSCAF_15441_fgenesh_117	-3.86684269722287

Table 3.5. Blast hits for top 10 up- (red) and downregulated (blue) genes in the later stage in each comparison.(Some candidates did not have any matching BLAST hits.)

Species	Comparison	Blast hit
C. linearis	Stage 1 vs 2	flavonoid 3',5'-hydroxylase, defensin-like protein (23), high affinity nitrate transporter 2.4-like, auxin-responsive protein SAUR71-like, dammarenediol II synthase-like, cytochrome P4t50 76A1-like
		protein ECERIFERUM 26-like, 3-ketoacyl-CoA synthase 15-like, acyltransferase-like protein At1g54570, chloroplastic isoform X, endochitinase EP3-like, fatty acyl-CoA reductase 2-like, ABC transporter G family member 26 isoform X, tetraketide alpha-pyrone reductase 1 isoform X, ABC transporter G family member 9-like
C. linearis	Stage 2 vs 3	alpha-farnesene synthase-like, cytochrome P450 83B1-like, gamma-cadinene synthase-like, polyphenol oxidase I, chloroplastic-like, acyl-acyl carrier protein thioesterase ATL3 -chloroplastic-like, tropinone reductase homolog At5g06060, palmitoyl-acyl carrier protein thioesterase, chloroplastic-like, auxin efflux carrier component 2-like, auxin efflux carrier component 2
		serine racemase isoform, probable indole-3-acetic acid-amido synthetase GH3.1, subtilisin-like protease SBT4.15, glutamate dehydrogenase A, plasma membrane ATPase-like, L-ascorbate peroxidase, cytosolic-like, GDSL esterase/lipase At2g40250-like, phosphoenolpyruvate-carboxykinase (ATP)-like, TetR/ AcrR family transcriptional regulator

Continuation of Table 3.5

Species	Comparison	Blast hit
C. rattanii	Stage 1 vs 2	xyloglucan endotransglucosylase/ hydrolase 2-like, gibberellin 20 oxidase 2-like, BURP domain-containing protein BNM2A-like, trans-resveratrol di-O-methyltransferase -like, acyl-acyl carrier protein thioesterase ATL3, chloroplastic-like
		probable polyol transporter 6
C. rattanii	Stage 2 vs 3	flavonol synthase/flavanone 3-hydroxylase -like, tetratricopeptide repeat protein 28-like, WAT1-related protein At5g07050-like, protein DETOXIFICATION 27- like, peroxidase 64-like, pathogen-related protein-like, transcription factor ORG2-like isoform, shikimate O-hydroxycinnamoyltransferase -like, cytochrome P450 84A1-like, protein DOWNY MILDEW RESISTANCE 6, senescence-specific cysteine protease SAG39-like, random slug protein 5-like
		trans-cinnamate 4-monooxygenase, basic leucine zipper 61

Table 3.6. The GO categories representing the up- (red) and downregulated (blue) genes in *C. linearis* stage 2 compared to stage 1. Within each up- or down-regulated gene group the GO categories are organized in the descending order of the number of hits. The categories discussed in the text are highlighted with a darker shade of the respective color.

GO Term	Hits	Gene Name
integral component of membrane	6	Collinsia 101;SccHEie 5367 HRSCAF 45597 fgenesh 37;SccHEie 5593 HRSCAF 46097 fgenesh 248;SccHEie 1640 HRSCAF 22564 fgenesh 15;SccHEie 2400 HRSCAF 28768 fgenesh 4;SccHEie 176 HRSCAF 3162 fgenesh 97
metal ion binding (heme binding, iron ion binding)	2	Collinsia 101;SccHEie 1640 HRSCAF 22564 fgenesh 15
monooxygenase activity	2	Collinsia 101;SccHEie 1640 HRSCAF 22564 fgenesh 15
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2	Collinsia 101;SccHEie 1640 HRSCAF 22564 fgenesh 15
ATP binding	1	SccHEie 2400 HRSCAF 28768 fgenesh 4
beta-amyrin synthase activity	1	SccHEie 5593 HRSCAF 46097 fgenesh 248
lanosterol synthase activity	1	SccHEie 5593 HRSCAF 46097 fgenesh 248
lipid binding	1	SccHEie 6020 HRSCAF 46524 fgenesh 87
nucleic acid binding	1	SccHEie 3267 HRSCAF 34614 fgenesh 51
protein dimerization activity	1	SccHEie 5894 HRSCAF 46398 fgenesh 9

Continuation of Table 3.6

GO Term	Hits	Gene Name
protein serine/threonine kinase activity	1	SccHEie 2400 HRSCAF 28768 fgenesh 4
RNA-DNA hybrid ribonuclease activity	1	SccHEie 3267 HRSCAF 34614 fgenesh 51
defense response to fungus	1	SccHEie 2832 HRSCAF 31795 fgenesh 154
DNA integration	1	SccHEie 3267 HRSCAF 34614 fgenesh 51
lipid transport	1	SccHEie 6020 HRSCAF 46524 fgenesh 87
response to auxin	1	SccHEie 2165 HRSCAF 26955 fgenesh 24
transmembrane transport	1	SccHEie 5367 HRSCAF 45597 fgenesh 37
triterpenoid biosynthetic process	1	SccHEie 5593 HRSCAF 46097 fgenesh 248
extracellular region	1	SccHEie 2832 HRSCAF 31795 fgenesh 154
lipid droplet	1	SccHEie 5593 HRSCAF 46097 fgenesh 248
integral component of membrane	20	SccHEie 5718 HRSCAF 46222 fgenesh 30;SccHEie 126 HRSCAF 2347 fgenesh 55;SccHEie 4749 HRSCAF 42444 fgenesh 391;SccHEie 4749 HRSCAF 42444 fgenesh 446

Continuation of Table 3.6

GO Term	Hits	Gene Name
integral component of membrane (continued)	20	SccHEie 1269 HRSCAF 18419 fgenesh 24;SccHEie 5879 HRSCAF 46383 fgenesh 6; SccHEie 5713 HRSCAF 46217 fgenesh 364;SccHEie 6027 HRSCAF 46531 fgenesh 13;SccHEie 4734 HRSCAF 42392 fgenesh 29;SccHEie 106 HRSCAF 1990 fgenesh 45;SccHEie 988 HRSCAF 15441 fgenesh 74;SccHEie 3655 HRSCAF 36969 fgenesh 72;SccHEie 5778 HRSCAF 46282 fgenesh 19;SccHEie 6154 HRSCAF 46658 fgenesh 111;SccHEie 731 HRSCAF 12182 fgenesh 14;SccHEie 616 HRSCAF 10378 fgenesh 106;SccHEie 6110 HRSCAF 46614 fgenesh 287;SccHEie 1545 HRSCAF 21613 fgenesh 41;SccHEie 5939 HRSCAF 46443 fgenesh 24;SccHEie 5829 HRSCAF 46333 fgenesh 79
ATP binding	9	SccHEie 5765 HRSCAF 46269 fgenesh 78;SccHEie 5718 HRSCAF 46222 fgenesh 30;SccHEie 5556 HRSCAF 46060 fgenesh 127;SccHEie 5556 HRSCAF 46060 fgenesh 41

Continuation of Table 3.6

GO Term	Hits	Gene Name
ATP binding (continued)	9	SccHEie 1269 HRSCAF 18419 fgenesh 24;SccHEie 5943 HRSCAF 46447 fgenesh 173;SccHEie 4668 HRSCAF 42044 fgenesh 190;SccHEie 6110 HRSCAF 46614 fgenesh 287;SccHEie 5939 HRSCAF 46443 fgenesh 24
metal ion binding (heme binding, iron ion binding, 4 iron - 4 sulfur cluster binding)	6	SccHEie 126 HRSCAF 2347 fgenesh 55;SccHEie 5713 HRSCAF 46217 fgenesh 364;SccHEie 3655 HRSCAF 36969 fgenesh 72;SccHEie 3655 HRSCAF 36969 fgenesh 77;SccHEie 6104 HRSCAF 46608 fgenesh 56;SccHEie 6079 HRSCAF 46583 fgenesh 115
transferase activity, transferring acyl groups other than amino-acyl groups	5	SccHEie 5106 HRSCAF 44094 fgenesh 175;SccHEie 2176 HRSCAF 27055 fgenesh 92;SccHEie 5484 HRSCAF 45988 fgenesh 99;SccHEie 5829 HRSCAF 46333 fgenesh 79;SccHEie 5656 HRSCAF 46160 fgenesh 3
carbohydrate metabolic process	5	SccHEie 4074 HRSCAF 39079 fgenesh 42;SccHEie 6157 HRSCAF 46661 fgenesh 13;SccHEie 1162 HRSCAF 17137 fgenesh 315
carbohydrate metabolic process (continued)	5	SccHEie 4427 HRSCAF 40840 fgenesh 28;SccHEie 1691 HRSCAF 22945 fgenesh 196

Continuation of Table 3.6

GO Term	Hits	Gene Name
ATPase activity	4	SccHEie 1269 HRSCAF 18419 fgenesh 24;ScHEie 6110 HRSCAF 46614 fgenesh 287;ScHEie 5939 HRSCAF 46443 fgenesh 24;ScHEie 4998 HRSCAF 43511 fgenesh 84
catalytic activity	4	SccHEie 6104 HRSCAF 46608 fgenesh 56;ScHEie 6079 HRSCAF 46583 fgenesh 115;ScHEie 5707 HRSCAF 46211 fgenesh 48;ScHEie 5764 HRSCAF 46268 fgenesh 209
monooxygenase activity	4	SccHEie 126 HRSCAF 2347 fgenesh 55;ScHEie 5713 HRSCAF 46217 fgenesh 364;ScHEie 3655 HRSCAF 36969 fgenesh 72;ScHEie 3655 HRSCAF 36969 fgenesh 77
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	4	SccHEie 126 HRSCAF 2347 fgenesh 55;ScHEie 5713 HRSCAF 46217 fgenesh 364;ScHEie 3655 HRSCAF 36969 fgenesh 72;ScHEie 3655 HRSCAF 36969 fgenesh 77
protein kinase activity	4	SccHEie 5718 HRSCAF 46222 fgenesh 30;ScHEie 5556 HRSCAF 46060 fgenesh 127; SccHEie 5556 HRSCAF 46060 fgenesh 41;ScHEie 4668 HRSCAF 42044 fgenesh 190; SccHEie 5783 HRSCAF 46287 fgenesh 61

Continuation of Table 3.6

GO Term	Hits	Gene Name
regulation of transcription, DNA-templated	4	SccHEie 5873 HRSCAF 46377 fgenesh 266;ScHEie 6108 HRSCAF 46612 fgenesh 123;ScHEie 3573 HRSCAF 36519 fgenesh 1
DNA binding	3	SccHEie 5783 HRSCAF 46287 fgenesh 61;ScHEie 5873 HRSCAF 46377 fgenesh 266;ScHEie 3573 HRSCAF 36519 fgenesh 1
hydrolase activity	3	SccHEie 5642 HRSCAF 46146 fgenesh 25;ScHEie 5521 HRSCAF 46025 fgenesh 274;ScHEie 5521 HRSCAF 46025 fgenesh 272
hydrolase activity, hydrolyzing O-glycosyl compounds	3	SccHEie 4074 HRSCAF 39079 fgenesh 42;ScHEie 5520 HRSCAF 46024 fgenesh 180;ScHEie 1162 HRSCAF 17137 fgenesh 315
transmembrane transporter activity	3	SccHEie 4749 HRSCAF 42444 fgenesh 391;ScHEie 106 HRSCAF 1990 fgenesh 45;ScHEie 616 HRSCAF 10378 fgenesh 106
biosynthetic process	3	SccHEie 5106 HRSCAF 44094 fgenesh 175;ScHEie 6027 HRSCAF 46531 fgenesh 13;ScHEie 2176 HRSCAF 27055 fgenesh 92
riboflavin biosynthetic process	3	SccHEie 5521 HRSCAF 46025 fgenesh 274;ScHEie 5521 HRSCAF 46025 fgenesh 272;ScHEie 3658 HRSCAF 36985 fgenesh 185

Continuation of Table 3.6

GO Term	Hits	Gene Name
protein dimerization activity	2	SccHEie 1546 HRSCAF 21616 fgenesh 46;SccHEie 2581 HRSCAF 30099 fgenesh 306
riboflavin kinase activity	2	SccHEie 5521 HRSCAF 46025 fgenesh 274;SccHEie 5521 HRSCAF 46025 fgenesh 272
base-excision repair	2	SccHEie 6104 HRSCAF 46608 fgenesh 56;SccHEie 6079 HRSCAF 46583 fgenesh 115
mitochondrial matrix	2	SccHEie 5743 HRSCAF 46247 fgenesh 14;SccHEie 3128 HRSCAF 33639 fgenesh 23
nucleus	2	SccHEie 5873 HRSCAF 46377 fgenesh 266;SccHEie 3573 HRSCAF 36519 fgenesh 1
beta-amyrin synthase activity	1	SccHEie 1545 HRSCAF 21613 fgenesh 41
beta-galactosidase activity	1	SccHEie 4427 HRSCAF 40840 fgenesh 28
carbohydrate binding	1	SccHEie 4427 HRSCAF 40840 fgenesh 28
channel activity	1	SccHEie 988 HRSCAF 15441 fgenesh 74
chitin binding	1	SccHEie 1691 HRSCAF 22945 fgenesh 196
chitinase activity	1	SccHEie 1691 HRSCAF 22945 fgenesh 196
coenzyme binding	1	SccHEie 5764 HRSCAF 46268 fgenesh 209
electron transfer activity	1	SccHEie 1740 HRSCAF 23314 fgenesh 18

Continuation of Table 3.6

GO Term	Hits	Gene Name
fatty acid binding	1	SccHEie 4293 HRSCAF 40210 fgenesh 51
fatty-acyl-CoA reductase (alcohol-forming) activity	1	SccHEie 4998 HRSCAF 43511 fgenesh 84
flavin adenine dinucleotide binding	1	SccHEie 4734 HRSCAF 42392 fgenesh 29
galactosyltransferase activity	1	SccHEie 4749 HRSCAF 42444 fgenesh 446
glucan endo-1,3-beta-D-glucosidase activity	1	SccHEie 6157 HRSCAF 46661 fgenesh 13
lanosterol synthase activity	1	SccHEie 1545 HRSCAF 21613 fgenesh 41
methyltransferase activity	1	SccHEie 5968 HRSCAF 46472 fgenesh 134
nucleic acid binding	1	SccHEie 5701 HRSCAF 46205 fgenesh 13
serine-type endopeptidase activity	1	SccHEie 6079 HRSCAF 46583 fgenesh 96
squalene monooxygenase activity	1	SccHEie 4734 HRSCAF 42392 fgenesh 29
strictosidine synthase activity	1	SccHEie 6027 HRSCAF 46531 fgenesh 13
structural constituent of nuclear pore	1	SccHEie 5701 HRSCAF 46205 fgenesh 38
tetrahydrofolylpolyglutamate synthase activity	1	SccHEie 5943 HRSCAF 46447 fgenesh 173
transferase activity, transferring hexosyl groups	1	SccHEie 5509 HRSCAF 46013 fgenesh 220
xanthoxin dehydrogenase activity	1	SccHEie 5781 HRSCAF 46285 fgenesh 9

Continuation of Table 3.6

GO Term	Hits	Gene Name
xyloglucan:xyloglucosyl transferase activity	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
auxin-activated signaling pathway	1	SccHEie 5873 HRSCAF 46377 fgenesh 266
carbohydrate transport	1	SccHEie 616 HRSCAF 10378 fgenesh 106
cell wall biogenesis	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
cell wall macromolecule catabolic process	1	SccHEie 1691 HRSCAF 22945 fgenesh 196
cell wall organization	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
chitin catabolic process	1	SccHEie 1691 HRSCAF 22945 fgenesh 196
cytokinin biosynthetic process	1	SccHEie 5642 HRSCAF 46146 fgenesh 25
DNA integration	1	SccHEie 5701 HRSCAF 46205 fgenesh 13
fatty acid biosynthetic process	1	SccHEie 5829 HRSCAF 46333 fgenesh 79
intracellular protein transport	1	SccHEie 731 HRSCAF 12182 fgenesh 14
lipid metabolic process	1	SccHEie 4998 HRSCAF 43511 fgenesh 84
protein glycosylation	1	SccHEie 4749 HRSCAF 42444 fgenesh 446
signal transduction	1	SccHEie 3319 HRSCAF 35018 fgenesh 131
sporopollenin biosynthetic process	1	SccHEie 5764 HRSCAF 46268 fgenesh 209
sterol biosynthetic process	1	SccHEie 4734 HRSCAF 42392 fgenesh 29

Continuation of Table 3.6

GO Term	Hits	Gene Name
systemic acquired resistance	1	SccHEie 4293 HRSCAF 40210 fgenesh 51
transmembrane transport	1	SccHEie 6154 HRSCAF 46658 fgenesh 111
triterpenoid biosynthetic process	1	SccHEie 1545 HRSCAF 21613 fgenesh 41
xyloglucan metabolic process	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
apoplast	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
cell wall	1	SccHEie 5520 HRSCAF 46024 fgenesh 180
endoplasmic reticulum	1	SccHEie 731 HRSCAF 12182 fgenesh 14
Golgi membrane	1	SccHEie 4749 HRSCAF 42444 fgenesh 446
lipid droplet	1	SccHEie 1545 HRSCAF 21613 fgenesh 41
nuclear pore	1	SccHEie 5701 HRSCAF 46205 fgenesh 38
nucleolus	1	SccHEie 5663 HRSCAF 46167 fgenesh 88

Table 3.7. The GO categories representing the up- (red) and downregulated (blue) genes in *C. linearis* stage 3 compared to stage 2. Within each up- or down-regulated gene group the GO categories are organized in the descending order of the number of hits. The categories discussed in the text are highlighted with a darker shade of the respective color.

GO Term	Hits	Gene Name
metal ion binding (heme binding, iron ion binding, magnesium ion binding, zinc ion binding, voltage-gated potassium channel activity)	16	SccHEie 5937 HRSCAF 46441 fgenesh 267;ScHEie 5828 HRSCAF 46332 fgenesh 517;ScHEie 5706 HRSCAF 46210 fgenesh 3;ScHEie 5706 HRSCAF 46210 fgenesh 11;ScHEie 6043 HRSCAF 46547 fgenesh 78;ScHEie 126 HRSCAF 2347 fgenesh 49;ScHEie 605 HRSCAF 10293 fgenesh 237; SccHEie 655 HRSCAF 11054 fgenesh 148;ScHEie 5556 HRSCAF 46060 fgenesh 155;ScHEie 5559 HRSCAF 46063 fgenesh 2; ScHEie 6043 HRSCAF 46547 fgenesh 96;ScHEie 3600 HRSCAF 36672 fgenesh 49;ScHEie 3724 HRSCAF 37329 fgenesh 71; ScHEie 6076 HRSCAF 46580 fgenesh 158;ScHEie 1716 HRSCAF 23152 fgenesh 27; ScHEie 1545 HRSCAF 21613 fgenesh 18
integral component of membrane	9	SccHEie 6004 HRSCAF 46508 fgenesh 186;ScHEie 5477 HRSCAF 45981 fgenesh 26

Continuation of Table 3.7

GO Term	Hits	Gene Name
integral component of membrane (continued)	9	SccHEie 175 HRSCAF 3144 fgenesh 74;SccHEie 725 HRSCAF 12064 fgenesh 122; SccHEie 1545 HRSCAF 21613 fgenesh 18;SccHEie 6043 HRSCAF 46547 fgenesh 78; SccHEie 126 HRSCAF 2347 fgenesh 49;SccHEie 605 HRSCAF 10293 fgenesh 237;SccHEie 5783 HRSCAF 46287 fgenesh 44
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7	SccHEie 5937 HRSCAF 46441 fgenesh 267;SccHEie 5828 HRSCAF 46332 fgenesh 517;SccHEie 5706 HRSCAF 46210 fgenesh 3;SccHEie 5706 HRSCAF 46210 fgenesh 11;SccHEie 6043 HRSCAF 46547 fgenesh 78;SccHEie 126 HRSCAF 2347 fgenesh 49;SccHEie 605 HRSCAF 10293 fgenesh 237
cell wall organization	6	SccHEie 5871 HRSCAF 46375 fgenesh 41;SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 5719 HRSCAF 46223 fgenesh 25;SccHEie 2832 HRSCAF 31795 fgenesh 13

Continuation of Table 3.7

GO Term	Hits	Gene Name
monooxygenase activity	5	SccHEie 5937 HRSCAF 46441 fgenesh 267;SccHEie 5828 HRSCAF 46332 fgenesh 517;SccHEie 6043 HRSCAF 46547 fgenesh 78;SccHEie 126 HRSCAF 2347 fgenesh 49;SccHEie 605 HRSCAF 10293 fgenesh 237
cell wall	5	SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 5719 HRSCAF 46223 fgenesh 25;SccHEie 2832 HRSCAF 31795 fgenesh 13
hydrolase activity, hydrolyzing O-glycosyl compounds	4	SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 2832 HRSCAF 31795 fgenesh 13
xyloglucan:xyloglucosyl transferase activity	4	SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 2832 HRSCAF 31795 fgenesh 13
cell wall biogenesis	4	SccHEie 6051 HRSCAF 46555 fgenesh 34

Continuation of Table 3.7

GO Term	Hits	Gene Name
cell wall biogenesis (continued)	4	SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 2832 HRSCAF 31795 fgenesh 13
xyloglucan metabolic process	4	SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 2832 HRSCAF 31795 fgenesh 13
apoplast	4	SccHEie 6051 HRSCAF 46555 fgenesh 34;SccHEie 1923 HRSCAF 25007 fgenesh 71;SccHEie 5763 HRSCAF 46267 fgenesh 17;SccHEie 2832 HRSCAF 31795 fgenesh 13
diacylglycerol O-acyltransferase activity	3	SccHEie 605 HRSCAF 10293 fgenesh 224;SccHEie 605 HRSCAF 10293 fgenesh 226;SccHEie 605 HRSCAF 10293 fgenesh 225
enzyme inhibitor activity	3	SccHEie 5477 HRSCAF 45981 fgenesh 26;SccHEie 725 HRSCAF 12064 fgenesh 122;SccHEie 5852 HRSCAF 46356 fgenesh 52
pectinesterase activity	3	SccHEie 5477 HRSCAF 45981 fgenesh 26;SccHEie 725 HRSCAF 12064 fgenesh 122
pectinesterase activity (continued)	3	SccHEie 5852 HRSCAF 46356 fgenesh 52

Continuation of Table 3.7

GO Term	Hits	Gene Name
terpene synthase activity	3	SccHEie 655 HRSCAF 11054 fgenesh 148;SccHEie 5556 HRSCAF 46060 fgenesh 155;SccHEie 5559 HRSCAF 46063 fgenesh 2
glycerolipid biosynthetic process	3	SccHEie 605 HRSCAF 10293 fgenesh 224;SccHEie 605 HRSCAF 10293 fgenesh 226;SccHEie 605 HRSCAF 10293 fgenesh 225
pectin catabolic process	3	SccHEie 5477 HRSCAF 45981 fgenesh 26;SccHEie 725 HRSCAF 12064 fgenesh 122;SccHEie 6020 HRSCAF 46524 fgenesh 76
nucleus	3	SccHEie 687 HRSCAF 11502 fgenesh 39;SccHEie 3600 HRSCAF 36672 fgenesh 49;SccHEie 5871 HRSCAF 46375 fgenesh 58
aspartyl esterase activity	2	SccHEie 5477 HRSCAF 45981 fgenesh 26;SccHEie 725 HRSCAF 12064 fgenesh 122
DNA binding	2	SccHEie 687 HRSCAF 11502 fgenesh 39;SccHEie 5871 HRSCAF 46375 fgenesh 58
hydrolase activity	2	SccHEie 5719 HRSCAF 46223 fgenesh 25;SccHEie 5706 HRSCAF 46210 fgenesh 74
oxidoreductase activity	2	SccHEie 4749 HRSCAF 42444 fgenesh 457;SccHEie 3724 HRSCAF 37329 fgenesh 71

Continuation of Table 3.7

GO Term	Hits	Gene Name
protein dimerization activity	2	SccHEie 4266 HRSCAF 40056 fgenesh 82;SccHEie 3299 HRSCAF 34826 fgenesh 3
cell wall modification	2	SccHEie 5477 HRSCAF 45981 fgenesh 26;SccHEie 725 HRSCAF 12064 fgenesh 122
extracellular region	2	SccHEie 5871 HRSCAF 46375 fgenesh 41;SccHEie 5719 HRSCAF 46223 fgenesh 25
catechol oxidase activity	1	SccHEie 6076 HRSCAF 46580 fgenesh 158
channel activity	1	SccHEie 5783 HRSCAF 46287 fgenesh 44
cysteine desulfurase activity	1	SccHEie 5470 HRSCAF 45974 fgenesh 65
cysteine-type peptidase activity	1	SccHEie 6111 HRSCAF 46615 fgenesh 47
fatty-acyl-CoA reductase (alcohol-forming) activity	1	SccHEie 6020 HRSCAF 46524 fgenesh 76
isopentenyl-diphosphate delta-isomerase activity	1	SccHEie 5706 HRSCAF 46210 fgenesh 74
lyase activity	1	SccHEie 5871 HRSCAF 46375 fgenesh 41
nucleic acid binding	1	SccHEie 4074 HRSCAF 39079 fgenesh 49
O-methyltransferase activity	1	SccHEie 4266 HRSCAF 40056 fgenesh 82
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	1	SccHEie 1716 HRSCAF 23152 fgenesh 27

Continuation of Table 3.7

GO Term	Hits	Gene Name
polygalacturonase activity	1	SccHEie 5871 HRSCAF 46375 fgenesh 41
protein phosphatase regulator activity	1	SccHEie 5764 HRSCAF 46268 fgenesh 168
RNA-DNA hybrid ribonuclease activity	1	SccHEie 4074 HRSCAF 39079 fgenesh 49
S-adenosylmethionine-dependent methyltransferase activity	1	SccHEie 4266 HRSCAF 40056 fgenesh 82
thiolester hydrolase activity	1	SccHEie 2287 HRSCAF 27867 fgenesh 68
transferase activity	1	SccHEie 60 HRSCAF 1143 fgenesh 59
transferase activity, transferring acyl groups other than amino-acyl groups	1	SccHEie 5719 HRSCAF 46223 fgenesh 11
triglyceride lipase activity	1	SccHEie 4519 HRSCAF 41305 fgenesh 26
aromatic compound biosynthetic process	1	SccHEie 4266 HRSCAF 40056 fgenesh 82
carbohydrate metabolic process	1	SccHEie 5871 HRSCAF 46375 fgenesh 41
fatty acid biosynthetic process	1	SccHEie 2287 HRSCAF 27867 fgenesh 68
isoprenoid biosynthetic process	1	SccHEie 5706 HRSCAF 46210 fgenesh 74
lipid metabolic process	1	SccHEie 6020 HRSCAF 46524 fgenesh 76
oxidation-reduction process	1	SccHEie 2845 HRSCAF 31865 fgenesh 407
oxylipin biosynthetic process	1	SccHEie 1716 HRSCAF 23152 fgenesh 27

Continuation of Table 3.7

GO Term	Hits	Gene Name
pigment biosynthetic process	1	SccHEie 6076 HRSCAF 46580 fgenesh 158
signal transduction	1	SccHEie 5764 HRSCAF 46268 fgenesh 168
transmembrane transport	1	SccHEie 6004 HRSCAF 46508 fgenesh 186
chloroplast	1	SccHEie 2287 HRSCAF 27867 fgenesh 68
cytosol	1	SccHEie 4266 HRSCAF 40056 fgenesh 82
plasma membrane	1	SccHEie 175 HRSCAF 3144 fgenesh 74
protein phosphatase type 2A complex	1	SccHEie 5764 HRSCAF 46268 fgenesh 168

Continuation of Table 3.7

GO Term	Hits	Gene Name
integral component of membrane	30	SccHEie 3504 HRSCAF 36127 fgenesh 71;SccHEie 5764 HRSCAF 46268 fgenesh 194;SccHEie 2073 HRSCAF 26174 fgenesh 37;SccHEie 5714 HRSCAF 46218 fgenesh 28;SccHEie 5000 HRSCAF 43525 fgenesh 36;SccHEie 489 HRSCAF 8674 fgenesh 183;SccHEie 5501 HRSCAF 46005 fgenesh 125;SccHEie 3655 HRSCAF 36969 fgenesh 92;SccHEie 5556 HRSCAF 46060 fgenesh 11;SccHEie 1740 HRSCAF 23314 fgenesh 116;SccHEie 5509 HRSCAF 46013 fgenesh 125;SccHEie 5699 HRSCAF 46203 fgenesh 113; SccHEie 5939 HRSCAF 46443 fgenesh 25;SccHEie 5726 HRSCAF 46230 fgenesh 45;SccHEie 5689 HRSCAF 46193 fgenesh 60;SccHEie 2832 HRSCAF 31795 fgenesh 33;SccHEie 5912 HRSCAF 46416 fgenesh 14; SccHEie 5728 HRSCAF 46232 fgenesh 320;SccHEie 6148 HRSCAF 46652 fgenesh 109;SccHEie 5958 HRSCAF 46462 fgenesh 75;SccHEie 4295 HRSCAF 40212 fgenesh 91

Continuation of Table 3.7

GO Term	Hits	Gene Name
integral component of membrane (continued)	30	SccHEie 305 HRSCAF 5661 fgenesh 208;SccHEie 802 HRSCAF 13183 fgenesh 74;SccHEie 5764 HRSCAF 46268 fgenesh 193;SccHEie 5717 HRSCAF 46221 fgenesh 78;SccHEie 2170 HRSCAF 27004 fgenesh 30;SccHEie 5939 HRSCAF 46443 fgenesh 80;SccHEie 4328 HRSCAF 40390 fgenesh 45;SccHEie 5493 HRSCAF 45997 fgenesh 18;SccHEie 5935 HRSCAF 46439 fgenesh 179
metal ion binding (zinc ion binding, calcium ion/calmodulin binding, iron ion binding, heme binding, calcium transmembrane transporter activity/ phosphorylative mechanism, molybdate ion transmembrane transporter activity, voltage-gated potassium channel activity, regulation of ion transmembrane transport, 3 iron - 4 sulfur cluster binding)	13	SccHEie 437 HRSCAF 7701 fgenesh 28;SccHEie 5936 HRSCAF 46440 fgenesh 361;SccHEie 5936 HRSCAF 46440 fgenesh 360;SccHEie 5654 HRSCAF 46158 fgenesh 588; SccHEie 2170 HRSCAF 27004 fgenesh 30;SccHEie 725 HRSCAF 12064 fgenesh 98; SccHEie 5562 HRSCAF 46066 fgenesh 13;SccHEie 5713 HRSCAF 46217 fgenesh 459;SccHEie 6051 HRSCAF 46555 fgenesh 191; SccHEie 3650 HRSCAF 36934 fgenesh 284; SccHEie 2832 HRSCAF 31795 fgenesh 33

Continuation of Table 3.7

GO Term	Hits	Gene Name
metal ion binding (continued)	13	SccHEie 489 HRSCAF 8674 fgenesh 183; ScHEie 5714 HRSCAF 46218 fgenesh 28
nucleus	8	SccHEie 5617 HRSCAF 46121 fgenesh 72;ScHEie 5723 HRSCAF 46227 fgenesh 33;ScHEie 5936 HRSCAF 46440 fgenesh 361;ScHEie 5654 HRSCAF 46158 fgenesh 588;ScHEie 5713 HRSCAF 46217 fgenesh 459;ScHEie 5455 HRSCAF 45959 fgenesh 12;ScHEie 6051 HRSCAF 46555 fgenesh 191;ScHEie 6109 HRSCAF 46613 fgenesh 60
sequence-specific DNA binding	6	SccHEie 2287 HRSCAF 27867 fgenesh 46;ScHEie 5936 HRSCAF 46440 fgenesh 361;ScHEie 5936 HRSCAF 46440 fgenesh 360;ScHEie 5654 HRSCAF 46158 fgenesh 588;ScHEie 5455 HRSCAF 45959 fgenesh 12;ScHEie 6109 HRSCAF 46613 fgenesh 60
ATP binding	6	SccHEie 3655 HRSCAF 36969 fgenesh 92;ScHEie 2832 HRSCAF 31795 fgenesh 33;ScHEie 5967 HRSCAF 46471 fgenesh 43

Continuation of Table 3.7

GO Term	Hits	Gene Name
ATP binding (continued)	6	SccHEie 4940 HRSCAF 43222 fgenesh 190;ScHEie 5644 HRSCAF 46148 fgenesh 5;ScHEie 5935 HRSCAF 46439 fgenesh 179
transmembrane transport	5	SccHEie 1740 HRSCAF 23314 fgenesh 116;ScHEie 5726 HRSCAF 46230 fgenesh 45;ScHEie 2170 HRSCAF 27004 fgenesh 184;ScHEie 2170 HRSCAF 27004 fgenesh 184;ScHEie 4295 HRSCAF 40212 fgenesh 91
DNA binding	4	SccHEie 5617 HRSCAF 46121 fgenesh 72;ScHEie 5723 HRSCAF 46227 fgenesh 33;ScHEie 5593 HRSCAF 46097 fgenesh 24;ScHEie 6051 HRSCAF 46555 fgenesh 191
carbohydrate metabolic process	4	SccHEie 5915 HRSCAF 46419 fgenesh 46;ScHEie 5550 HRSCAF 46054 fgenesh 26;ScHEie 5871 HRSCAF 46375 fgenesh 405;ScHEie 5967 HRSCAF 46471 fgenesh 43
transcription, DNA-templated	3	SccHEie 5723 HRSCAF 46227 fgenesh 33;ScHEie 2287 HRSCAF 27867 fgenesh 46;ScHEie 5593 HRSCAF 46097 fgenesh 24

Continuation of Table 3.7

GO Term	Hits	Gene Name
hydrolase activity, hydrolyzing O-glycosyl compounds	3	SccHEie 5915 HRSCAF 46419 fgenesh 46;SccHEie 6004 HRSCAF 46508 fgenesh 158;SccHEie 134 HRSCAF 2435 fgenesh 20
DNA-binding transcription factor activity	3	SccHEie 2287 HRSCAF 27867 fgenesh 46;SccHEie 5455 HRSCAF 45959 fgenesh 12;SccHEie 6109 HRSCAF 46613 fgenesh 60
cellular amino acid metabolic process	3	SccHEie 1800 HRSCAF 23805 fgenesh 7;SccHEie 5469 HRSCAF 45973 fgenesh 324;SccHEie 4734 HRSCAF 42392 fgenesh 124
cell wall	3	SccHEie 6004 HRSCAF 46508 fgenesh 158;SccHEie 134 HRSCAF 2435 fgenesh 20;SccHEie 539 HRSCAF 9407 fgenesh 79
xyloglucan:xyloglucosyl transferase activity	2	SccHEie 6004 HRSCAF 46508 fgenesh 158;SccHEie 134 HRSCAF 2435 fgenesh 20
tricarboxylic acid cycle	2	SccHEie 5550 HRSCAF 46054 fgenesh 26;SccHEie 5967 HRSCAF 46471 fgenesh 43
transmembrane transporter activity	2	SccHEie 5000 HRSCAF 43525 fgenesh 36;SccHEie 5501 HRSCAF 46005 fgenesh 125
transferase activity, transferring glycosyl groups	2	SccHEie 5764 HRSCAF 46268 fgenesh 194;SccHEie 5764 HRSCAF 46268 fgenesh 193

Continuation of Table 3.7

GO Term	Hits	Gene Name
RNA binding	2	SccHEie 6011 HRSCAF 46515 fgenesh 1;SccHEie 4749 HRSCAF 42444 fgenesh 32
pyridoxal phosphate binding	2	SccHEie 1800 HRSCAF 23805 fgenesh 7;SccHEie 4734 HRSCAF 42392 fgenesh 124
proton-exporting ATPase activity, phosphorylative mechanism	2	SccHEie 5935 HRSCAF 46439 fgenesh 179;SccHEie 5935 HRSCAF 46439 fgenesh 179
positive regulation of transcription, DNA-templated	2	SccHEie 5936 HRSCAF 46440 fgenesh 361;SccHEie 5654 HRSCAF 46158 fgenesh 588
plasma membrane	2	SccHEie 4295 HRSCAF 40212 fgenesh 91;SccHEie 5935 HRSCAF 46439 fgenesh 179
peroxidase activity	2	SccHEie 6090 HRSCAF 46594 fgenesh 2;SccHEie 3650 HRSCAF 36934 fgenesh 284
nutrient reservoir activity	2	SccHEie 6004 HRSCAF 46508 fgenesh 64;Collinsia 66
malate metabolic process	2	SccHEie 5550 HRSCAF 46054 fgenesh 26;SccHEie 5967 HRSCAF 46471 fgenesh 43
L-malate dehydrogenase activity	2	SccHEie 5550 HRSCAF 46054 fgenesh 26;SccHEie 5967 HRSCAF 46471 fgenesh 43
hydrolase activity	2	SccHEie 3504 HRSCAF 36127 fgenesh 71;SccHEie 2832 HRSCAF 31795 fgenesh 33
fucose metabolic process	2	SccHEie 5764 HRSCAF 46268 fgenesh 194;SccHEie 5764 HRSCAF 46268 fgenesh 193

Continuation of Table 3.7

GO Term	Hits	Gene Name
electron transfer activity	2	SccHEie 5776 HRSCAF 46280 fgenesh 30;SccHEie 5939 HRSCAF 46443 fgenesh 80
DNA-directed 5'-3' RNA polymerase activity	2	SccHEie 5723 HRSCAF 46227 fgenesh 33;SccHEie 5593 HRSCAF 46097 fgenesh 24
cell redox homeostasis	2	SccHEie 5556 HRSCAF 46060 fgenesh 11;SccHEie 6090 HRSCAF 46594 fgenesh 2
cell	2	SccHEie 5556 HRSCAF 46060 fgenesh 11;SccHEie 6090 HRSCAF 46594 fgenesh 2
apoplast	2	SccHEie 6004 HRSCAF 46508 fgenesh 158;SccHEie 134 HRSCAF 2435 fgenesh 20
xyloglucan metabolic process	1	SccHEie 134 HRSCAF 2435 fgenesh 20
tRNA processing	1	SccHEie 5967 HRSCAF 46471 fgenesh 43
translation	1	SccHEie 6011 HRSCAF 46515 fgenesh 1
transferase activity, transferring acyl groups	1	SccHEie 5912 HRSCAF 46416 fgenesh 14
transferase activity	1	SccHEie 5967 HRSCAF 46471 fgenesh 43
synaptic vesicle	1	SccHEie 2400 HRSCAF 28768 fgenesh 12
structural constituent of ribosome	1	SccHEie 6011 HRSCAF 46515 fgenesh 1
starch biosynthetic process	1	SccHEie 4940 HRSCAF 43222 fgenesh 190

Continuation of Table 3.7

GO Term	Hits	Gene Name
starch binding	1	SccHEie 2170 HRSCAF 27004 fgenesh 184
serine-type endopeptidase activity	1	SccHEie 5270 HRSCAF 45017 fgenesh 75
serine racemase activity	1	SccHEie 4734 HRSCAF 42392 fgenesh 124
ribonucleoside binding	1	SccHEie 5593 HRSCAF 46097 fgenesh 24
retrotransposon nucleocapsid	1	SccHEie 6051 HRSCAF 46555 fgenesh 174
response to oxidative stress	1	SccHEie 3650 HRSCAF 36934 fgenesh 284
regulation of transcription, DNA-templated	1	SccHEie 5936 HRSCAF 46440 fgenesh 360
protein-containing complex assembly	1	SccHEie 2400 HRSCAF 28768 fgenesh 12
protein kinase activity	1	SccHEie 3655 HRSCAF 36969 fgenesh 92
plant-type cell wall organization	1	SccHEie 539 HRSCAF 9407 fgenesh 79
plant seed peroxidase activity	1	SccHEie 5562 HRSCAF 46066 fgenesh 13
phosphoenolpyruvate carboxykinase (ATP) activity	1	SccHEie 5644 HRSCAF 46148 fgenesh 5
peroxiredoxin activity	1	SccHEie 6090 HRSCAF 46594 fgenesh 2
pentose-phosphate shunt	1	SccHEie 5891 HRSCAF 46395 fgenesh 126
oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor	1	SccHEie 5469 HRSCAF 45973 fgenesh 324

Continuation of Table 3.7

GO Term	Hits	Gene Name
		Collinsia 73
oxidoreductase activity	1	
organelle membrane	1	SccHEie 5562 HRSCAF 46066 fgenesh 13
nucleotide-sugar metabolic process	1	SccHEie 342 HRSCAF 6189 fgenesh 223
nucleotide binding	1	SccHEie 5469 HRSCAF 45973 fgenesh 324
NADP binding	1	SccHEie 5891 HRSCAF 46395 fgenesh 126
NAD-dependent histone deacetylase activity (H3-K14 specific)	1	SccHEie 5713 HRSCAF 46217 fgenesh 459
methyltransferase activity	1	SccHEie 5689 HRSCAF 46193 fgenesh 60
membrane	1	SccHEie 539 HRSCAF 9407 fgenesh 79
malate transport	1	SccHEie 802 HRSCAF 13183 fgenesh 74
lipid droplet	1	SccHEie 5562 HRSCAF 46066 fgenesh 13
ligase activity	1	SccHEie 5525 HRSCAF 46029 fgenesh 196
large ribosomal subunit	1	SccHEie 6011 HRSCAF 46515 fgenesh 1
L-aspartate:2-oxoglutarate aminotransferase activity	1	SccHEie 1800 HRSCAF 23805 fgenesh 7
L-arabinose metabolic process	1	SccHEie 6133 HRSCAF 46637 fgenesh 67
hydrolase activity, acting on ester bonds	1	SccHEie 6106 HRSCAF 46610 fgenesh 25

Continuation of Table 3.7

GO Term	Hits	Gene Name
histone acetyltransferase activity	1	SccHEie 6051 HRSCAF 46555 fgenesh 191
glycogen biosynthetic process	1	SccHEie 4940 HRSCAF 43222 fgenesh 190
glutamate synthase (NADH) activity	1	SccHEie 725 HRSCAF 12064 fgenesh 98
glutamate biosynthetic process	1	SccHEie 725 HRSCAF 12064 fgenesh 98
glucose-6-phosphate dehydrogenase activity	1	SccHEie 5891 HRSCAF 46395 fgenesh 126
glucose-1-phosphate adenylyltransferase activity	1	SccHEie 4940 HRSCAF 43222 fgenesh 190
glucose metabolic process	1	SccHEie 5891 HRSCAF 46395 fgenesh 126
gluconeogenesis	1	SccHEie 5644 HRSCAF 46148 fgenesh 5
Flavin mono nucleotide (FMN) binding	1	SccHEie 725 HRSCAF 12064 fgenesh 98
flavin adenine dinucleotide binding	1	SccHEie 725 HRSCAF 12064 fgenesh 98
extracellular region	1	SccHEie 539 HRSCAF 9407 fgenesh 79
endoplasmic reticulum	1	SccHEie 5562 HRSCAF 46066 fgenesh 13
dTDP-glucose 4,6-dehydratase activity	1	SccHEie 342 HRSCAF 6189 fgenesh 223
chloroplast	1	SccHEie 4940 HRSCAF 43222 fgenesh 190
chemical synaptic transmission	1	SccHEie 2400 HRSCAF 28768 fgenesh 12

Continuation of Table 3.7

GO Term	Hits	Gene Name
cellulose catabolic process	1	SccHEie 5493 HRSCAF 45997 fgenesh 18
cellulase activity	1	SccHEie 5493 HRSCAF 45997 fgenesh 18
cellular glucan metabolic process	1	SccHEie 6004 HRSCAF 46508 fgenesh 158
cell wall organization	1	SccHEie 134 HRSCAF 2435 fgenesh 20
cell wall biogenesis	1	SccHEie 134 HRSCAF 2435 fgenesh 20
carbohydrate binding	1	SccHEie 5871 HRSCAF 46375 fgenesh 405
biosynthetic process	1	SccHEie 1800 HRSCAF 23805 fgenesh 7
beta-galactosidase activity	1	SccHEie 5871 HRSCAF 46375 fgenesh 405
AT DNA binding	1	SccHEie 56 HRSCAF 1078 fgenesh 53
amyloid-beta binding	1	SccHEie 2400 HRSCAF 28768 fgenesh 12
alpha-L-arabinofuranosidase activity	1	SccHEie 6133 HRSCAF 46637 fgenesh 67

Table 3.8. The GO categories representing the up- (red) and downregulated (blue) genes in *C. rattanii* stage 2 compared to stage 1. Within each up- or down-regulated gene group the GO categories are organized in the descending order of the number of hits. The categories discussed in the text are highlighted with a darker shade of the respective color.

GO Term	Hits	Gene Name
protein dimerization activity	2	SccHEie 605 HRSCAF 10293 fgenesh 133;SccHEie 605 HRSCAF 10293 fgenesh 133
metal ion binding	1	SccHEie 4734 HRSCAF 42392 fgenesh 92
cell wall (biogenesis, organization)	1	SccHEie 5593 HRSCAF 46097 fgenesh 282
hydrolase activity, hydrolyzing O-glycosyl compounds	1	SccHEie 5593 HRSCAF 46097 fgenesh 282
O-methyltransferase activity	1	SccHEie 605 HRSCAF 10293 fgenesh 133
oxidoreductase activity	1	SccHEie 4734 HRSCAF 42392 fgenesh 92
xyloglucan:xyloglucosyl transferase activity	1	SccHEie 5593 HRSCAF 46097 fgenesh 282
xyloglucan metabolic process	1	SccHEie 5593 HRSCAF 46097 fgenesh 282
apoplast	1	SccHEie 5593 HRSCAF 46097 fgenesh 282
transmembrane transporter activity	1	SccHEie 4116 HRSCAF 39323 fgenesh 80
integral component of membrane	1	SccHEie 4116 HRSCAF 39323 fgenesh 80

Table 3.9. The GO categories representing the up- (red) and downregulated (blue) genes in *C. rattanii* stage 3 compared to stage 2. Within each up- or down-regulated gene group the GO categories are organized in the descending order of the number of hits. The categories discussed in the text are highlighted with a darker shade of the respective color.

GO Term	Hits	Gene Name
metal ion binding (heme binding, iron ion binding)	4	SccHEie 5540 HRSCAF 46044 fgenesh 120;SccHEie 5556 HRSCAF 46060 fgenesh 246;SccHEie 162 HRSCAF 2890 fgenesh 31 ;SccHEie 5912 HRSCAF 46416 fgenesh 44
integral component of membrane	3	SccHEie 5887 HRSCAF 46391 fgenesh 34;SccHEie 5267 HRSCAF 45013 fgenesh 127;SccHEie 5912 HRSCAF 46416 fgenesh 44
alcohol O-acetyltransferase activity	1	SccHEie 4538 HRSCAF 41403 fgenesh 8
antiporter activity	1	SccHEie 5267 HRSCAF 45013 fgenesh 127
cysteine-type peptidase activity	1	SccHEie 6111 HRSCAF 46615 fgenesh 47
DNA binding	1	SccHEie 1460 HRSCAF 20448 fgenesh 41
monooxygenase activity	1	SccHEie 5912 HRSCAF 46416 fgenesh 44
naringenin 3-dioxygenase activity	1	SccHEie 162 HRSCAF 2890 fgenesh 31
oxidoreductase activity	1	SccHEie 5540 HRSCAF 46044 fgenesh 120
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	1	SccHEie 5912 HRSCAF 46416 fgenesh 44

Continuation of Table 3.9

GO Term	Hits	Gene Name
peroxidase activity/ hydrogen peroxide catabolic process/ response to oxidative stress	1	SccHEie 5556 HRSCAF 46060 fgenesh 246
protein dimerization activity	1	SccHEie 1460 HRSCAF 20448 fgenesh 41
transmembrane transporter activity	1	SccHEie 5887 HRSCAF 46391 fgenesh 34
xenobiotic transmembrane transporter activity	1	SccHEie 5267 HRSCAF 45013 fgenesh 127
regulation of transcription by RNA polymerase II	1	SccHEie 1460 HRSCAF 20448 fgenesh 41
extracellular region/ plasmodesma/ plant-type cell wal	1	SccHEie 5556 HRSCAF 46060 fgenesh 246
metal ion binding (heme binding, iron ion binding)	1	SccHEie 988 HRSCAF 15441 fgenesh 117
DNA-binding transcription factor activity	1	SccHEie 605 HRSCAF 10293 fgenesh 115
monooxygenase activity/ oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	1	SccHEie 988 HRSCAF 15441 fgenesh 117

General Conclusions

In angiosperms, the evolution of reproductive development is highly plastic. Traits such as fleshy fruit and self-mating systems have evolved multiple times from dry fruit and out-crossing systems, respectively (Barrett, 2002; Bolmgren and Eriksson, 2010). However, the molecular mechanisms that underlie these shifts remain to be elucidated.

The Solanaceae (nightshades) provides opportunities to investigate the molecular basis associated with the evolution of fleshy fruit as there have been multiple evolutionary transitions to fleshy fruit as well as a reversal to dry fruit in this family (Knapp, 2002). In addition, this family, in general, is amenable to genetic manipulation and multiple sequenced genomes are available (Bombarely et al., 2016; Consortium and The Potato Genome Sequencing Consortium, 2011; Tomato Genome Consortium, 2012). *FRUITFULL* (*FUL*) functions in patterning the dehiscence zone in the dry silique of *Arabidopsis thaliana* (Gu et al., 1998). The *FUL* ortholog has a similar role in the capsule of *Nicotiana* (Smykal et al., 2007). In contrast, in the fleshy tomato (*Solanum lycopersicum*), CIRSPR/Cas9 knockouts of *FUL* orthologs, *SlFUL1* and *SlFUL2* have revealed a function for these genes in the ripening-related change in coloration, ethylene production as well as some role in early fruit development (Wang et al., 2019). This suggests a change in gene function for these genes in the shift to fleshy fruit in the Solanaceae.

I characterized the evolution of *FUL* orthologs across the Solanaceae phylogeny (Maheepala et al., 2019). In addition to *FUL1* and *FUL2*, Solanaceae has two other

FUL orthologs, *MBP10* and *MBP20* (Litt and Irish, 2003; Shan et al., 2007). Our analyses suggest while *FUL1* and *FUL2* might be a result of a whole genome multiplication event, *MBP10* and *MBP20* clades were probably a result of a later occurring tandem gene duplication event. Our evolutionary rate analyses indicate *FUL1* and *MBP10* coding sequences are evolving faster in relation to *FUL2* and *MBP20*, respectively. In addition, the overall weak expression levels and an atypically short first intron suggest *MBP10* might be becoming a pseudogene.

However, we did not find evidence for any prevalent shifts in amino acid sequence associated with the fruit type. Thus, it might be that the diversification of *FUL* ortholog function is due to altered roles of downstream genes. Therefore, it would be beneficial to investigate the functional diversification of the genes that interact with *FUL* orthologs. Although the function of *FUL1* and *FUL2* have been elucidated with regard to tomato development (Wang et al., 2019), their function remains to be confirmed in dry fruit development. In addition, no functional data for *MBP10/20* exists on fruit development. Therefore, CRISPR/Cas9 knockout studies that elucidate the function of all four *FUL* orthologs in both dry and fleshy fruit development would further our understanding of the genetic architecture of fleshy fruit evolution.

I also generated transcriptome data for all stages of fruit development in tomato (*S. lycopersicum*), *S. pimpinellifolium*, which is the closest wild relative of the cultivated tomato, and desert tobacco (*N. obtusifolia*). My contribution to this project consisted of generating the tomato expression data and their comparative analysis

to identify any molecular changes associated with the extensive artificial selection that the cultivated tomato has undergone. The results of our gene ontology category analyses between cultivated tomato and *S. pimpinellifolium* coincide with the larger fruit size in the former species. I also analyzed co-expression gene clusters and the potential interactions between fruit development-related genes in our tomato transcriptomes. However, the results were inconclusive probably due to a limitation in the number of samples we had.

In the genus *Collinsia* (Plantaginaceae), multiple pairs of sister taxa consist of a predominantly outcrossing and a selfing species indicating the transition to selfmating has occurred multiple times in this genus (Randle et al., 2009). These evolutionary shifts are due to changes in the developmental timing of the reproductive whorls; in general, anthers and pistils develop and mature at different rates in outcrossers while these events are synchronous in selfers. Considering that the available sequence data for *Collinsia* are also on the rise, this genus provides an opportunity to investigate the molecular mechanisms that may underlie such changes in the developmental timing. We generated representative floral developmental transcriptomes for the two sister species, the outcrossing *C. linearis* and the selfing *C. rattanii* and conducted intraspecific comparative analyses between consecutive stages of reproductive development. Our gene ontology analyses suggest that metal ion binding proteins might be associated with the difference in the developmental timing of the reproductive whorls in these two species. In addition, we found that putative genes related to pollen development and pollinator

attraction are downregulated in the selfing species, which has also been reported by a another group (Hazzouri et al., 2013).

Due to the minor phenotypic variation between the floral developmental phases and rapid maturation in these two species, especially the selfing *C. rattanii* which matures earlier compared to *C. linearis*, some of our expression data might have overlapped the boundary zone between two consecutive developmental stages. Future studies aimed at identifying the molecular changes in the transition to selfing in this genus will benefit from extensive expression data representing clearly demarcated developmental stages in all sister taxa. In addition, investigating the potential modes of gene regulation (i.e., changes in the regulatory regions or the epigenome) between the two mating systems may provide helpful insights.

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. Nature Reviews Genetics 3, 274–284. doi:10.1038/nrg776.

Bolmgren, K., and Eriksson, O. (2010). Seed mass and the evolution of fleshy fruits in angiosperms. Oikos 119, 707–718. doi:10.1111/j.1600-0706.2009.17944.x.

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat Plants 2, 16074.

Consortium, T. P. G. S., and The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature 475, 189–195. doi:10.1038/nature10158.

Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The *FRUITFULL MADS-box* gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125, 1509–1517.

Hazzouri, K. M., Escobar, J. S., Ness, R. W., Killian Newman, L., Randle, A. M., Kalisz, S., et al. (2013). Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection, and evolution of biased gene conversion with an ongoing mating system shift. Evolution 67, 1263–1278.

Knapp, S. (2002). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J. Exp. Bot. 53, 2001–2022.

Litt, A., and Irish, V. F. (2003). Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics 165, 821–833.

Maheepala, D. C., Emerling, C. A., Rajewski, A., Macon, J., Strahl, M., Pabón-Mora, N., et al. (2019). Evolution and diversification of *FRUITFULL* genes in Solanaceae. Frontiers in Plant Science 10. doi:10.3389/fpls.2019.00043.

Randle, A. M., Slyder, J. B., and Kalisz, S. (2009). Can differences in autonomous selfing ability explain differences in range size among sister-taxa pairs of *Collinsia* (Plantaginaceae)? An extension of Baker's Law. New Phytol. 183, 618–629.

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and functional diversification during the evolution of the *AP1/SQUA* subfamily of plant MADS-box genes. Mol. Phylogenet. Evol. 44, 26–41.

Smykal, P., Gennen, J., De Bodt, S., Ranganath, V., and Melzer, S. (2007). Flowering of strict photoperiodic *Nicotiana* varieties in non-inductive conditions by transgenic approaches. Plant Mol. Biol. 65, 233–242.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641.

Wang, R., Tavano, E. C. da R., Lammers, M., Martinelli, A. P., Angenent, G. C., and de Maagd, R. A. (2019). Re-evaluation of transcription factor function in tomato fruit development and ripening with CRISPR/Cas9-mutagenesis. Sci. Rep. 9, 1696.