AN ABSTRACT OF THE DISSERTATION OF

Peter S. Boches for the degree of Doctor of Philosophy in Horticulture presented on February 25, 2009.

Title: Breeding Tomato for Increased Fruit Phenolics

Abstract approved:

_____

James R. Myers

The purpose of this research was to develop tools and germplasm for the production of tomatoes (*Solanum lycopersicum* L.) with high fruit phenolics without the use of genetic engineering. A candidate gene analysis was undertaken to identify genes regulating anthocyanin biosynthesis in tomato fruit (*Aft, Abg, atv*, and *Purple Smudge*). The genes *Aft* and *Purple Smudge* were associated with a DNA polymorphism in a tomato MYB transcription factor (*SlAn2*) similar to *Petunia An2*. The *SlAn2* DNA polymorphism co-segregated with the *Aft* trait in 183 F_2 individuals from an *Aft* x 'Legend' population. Expression analysis of the *SlAn2* gene using semi-quantitative reverse transcription PCR showed a close correlation between transcript levels of *SlAn2*, anthocyanin expression, and the anthocyanin biosynthetic gene dihydroflavonol 4-reductase (*DFR*). Expression analysis of the related MYB transcription factor *SlAnt1* showed poor correlation with anthocyanin expression in the fruit. The *SlAn2* gene was sequenced in several genotypes. Phylogenetic analysis of the predicted amino acid sequences of tomato *SlAnt1* and *SlAn2* indicated that they are orthologous to potato *StAn1* and *StAn2*, respectively. To create a high flavonol tomato line, an elite high anthocyanin line with the genes *Aft* and *atv* was crossed to a tomato line with the *aw* gene, which encodes a non-functional *DFR*. *AftAft/atvatv/awaw* F_3 segregants had fruit with levels of rutin and chlorogenic acid that were significantly higher than the cultivar 'Legend' fruit under greenhouse conditions. Total phenolics levels in fruit of *AftAft/atvatv/awaw* F_3 segregants were not significantly different from

an elite *AftAft/atvatv* line under field conditions. However, small amounts of anthocyanin were unexpectedly found in fruit. In order to identify and introgress additional genes that would increase total phenolics or result in novel phenolic compounds in tomato fruit, we screened a core collection of *Solanum lycopersicum* var. *cerasiforme*, and identified several accessions with good horticultural quality and fruit with significantly higher levels of total phenolics and polyphenolic compounds not found in cultivated tomatoes.

Breeding Tomato for Increased Fruit Phenolics

by
Peter S. Boches

A DISSERTATION

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented February 25, 2009
Commencement June 2009

Doctor of Philosophy dissertation of Peter S. Boches
presented on February 25, 2009

APPROVED:

_____

Major Professor, representing Horticulture

_____

Head of the Department of Horticulture

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Peter S. Boches, Author

ACKNOWLEDGEMENTS

I sincerely thank my wife, Charlotte, and my daughters, Althea and Mithril, all of whom are a constant inspiration to me. I thank my mother, Peggy, for all her help with everything. Without their affection and help this would have been a much colder and lonelier endeavor. My mentor, Dr. Jim Myers has been my benefactor, taskmaster, and sounding board. I thank him for the opportunity to conduct this research. The other members of my committee, Ken Johnson, Hiro Nonogaki, Jennifer Kling, and Isabel Vales, have been extremely professional and always done their best to assist my academic endeavors. Steve Tanksley, Ilan Levin, Boris Igić, Ingrid Phillips, Walter De Jong, and Rune Slimestad all provided valuable communications and shared data openly with me. Without their help this work would not have been realized. Brooke Peterschmidt assisted with this research as part of her undergraduate thesis and HHMI internship, she was an enormous help and motivation. I thank Deborah Kean and Randy Hopson for their dedication to the Vegetable Farm, excellent humor, and good spirits. I extend hearty thanks to the Department of Horticulture office staff. Gina Hashagen and LeeAnn Julson have handled more logistics on my behalf than I care to remember. Penny Pinard and Clara Knutson worked hard to ensure that I had the materials necessary for this research. Viki Meink has always been extremely quick and effective in helping me to deal with any administrative business I needed to finish. Special thanks are due to Patti Skinkis and Fred Stevens, who graciously allowed me to use their labs for HPLC analysis. Bob Durst and Jeff Morre provided technical assistance with the phenolics analysis. I really appreciate Anita Azarenko, the head of the Department of Horticulture, for showing sincere care for me and all of the graduate students in the Horticulture Department. I also really appreciate the efforts of Stephanie Duckett and the Student Parent Advisory Board for helping student parents like me. I am very thankful for the companionship provided by students in the Horticulture department, Paul Kusolwa, Erron Haggard, Kahraman Gurcan, Vidyasagar Sathuvalli, Shawna Zimmerman, Miles Barrett, Levi Frederickson, Mike Dossett, Mike Russel; without y'all it would have been very boring indeed.

CONTRIBUTION OF AUTHORS

Dr. Jim Myers assisted with experimental design, analysis, and writing of Chapters 2, 3, and 4. Brooke Peterschmidt performed phenolics extractions for the summer of 2007, assisted with HPLC analysis of this material, and contributed to data analysis and writing of chapter 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

DEDICATION

This thesis is dedicated in loving memory of my father, who always encouraged my interests in science.

CHAPTER 1

Introduction

Peter S. Boches

Tomato breeding research at Oregon State University

Tomato (*S. lycopersicum* L.) breeding at Oregon State University (OSU) began with vegetable breeder Dr. Tex Frazier and was continued by his successors Dr. Jim Baggett and Dr. Jim Myers. Early efforts were aimed at producing tomato varieties that would produce good yields of high quality tomatoes in a typical Oregon growing season, which tends to be cold and short for tomato production. These efforts produced early maturing tomatoes such as 'Oregon Spring', 'Siletz', 'Santiam', and 'Willamette', which are still popular among gardeners today. Another trait that was discovered and incorporated into OSU tomatoes was parthenocarpy (seedlessness). Parthenocarpy may also improve earliness since parthenocarpic tomatoes do not need fertilization to set fruit. 'Legend' and 'Gold Nugget' are both parthenocarpic OSU varieties. 'Gold Nugget' is also notable for its early maturity and because it is a plum tomato that has a bush (determinate) habit. 'Legend', released by both Dr. Baggett and Dr. Myers, is parthenocarpic as well and characterized by late blight resistance and a large set of early maturing tomatoes that are nestled within the foliage to protect them from sunburn. 'Legend' is probably the most popular tomato variety released by OSU today. Dr. Jim Myers has continued to work on developing new tomato varieties at OSU with efforts to generate a tomato with better late blight resistance than 'Legend' and a tomato with enhanced nutritional value.

Under the direction of Jim Myers, graduate student Carl Jones investigated the effects of several genes capable of altering the nutritional profiles of tomato fruit (Jones, 2000). Examples include genes that result in enhanced lycopene production (*og^c)*, enhanced beta carotene production (*B*), production of anthocyanins (*Abg, Aft, atv*), and enhanced production of all pigments (*hp-1)*. Lycopene, beta carotene, and anthocyanin are all thought to contribute to an antioxidant or healthful effect in the fruit. Peter Mes continued this work by examining the effects of combining these genes. Jones, Mes and Myers published an article describing the inheritance of *Aft* as well as *Abg* and *atv* (Jones et al, 2003). Mes found that many combinations of these genes were additive in effect. At the outset of this research in 2005, these lines were approaching homozygozity (F_4). For example, combining the anthocyanin producing genes *Abg* and *atv* or *Aft* and *atv* resulted in a tomato with higher anthocyanin

production than the *Abg* or *Aft* parent lines (Mes, 2005; Mes et al., 2008). Similarly, the addition of the *hp-1* gene caused higher anthocyanin production in an anthocyanin producing line (*Aft*). Anthocyanin producing genes could also be combined with carotenoid altering genes without decreasing total carotenoids. In addition, Mes found preliminary evidence that high anthocyanin tomatoes had enhanced resistance to microbial pathogens (Mes, 2005).

Health benefits of tomato metabolites

Increasingly, consumption of fruits and vegetables is recognized as having health benefits such as a lowered risk of cardiovascular disease and cancer, due, in part, to the bioactive properties of plant derived compounds like polyphenolics and carotenoids (Kris-Etherton et al., 2002). Flavonoids, a class of polyphenolic that includes the anthocyanins and flavonols, have been widely recognized as having anti-tumor activity both in vitro and in vivo (Kandaswami et al., 2005; Li et al., 2007; Ren et al., 2003). Numerous potential mechanisms for the bioactivity of flavonoids exist, including an antioxidant effect, an antiproliferative effect, inhibition of cell cycle progression, and interaction with epidermal growth factor receptors, mitogen-activated protein kinases, phosphatodylinositol-3kinases, or androgen/estrogen receptors (Lin and Weng, 2006). Flavonoids have also been shown to interact with numerous mammalian enzyme systems, suppress viral reverse transcriptases, and modulate the immune and inflammatory responses (Middleton and Kandaswami, 1993). Recent research indicates that the bioactivity of polyphenolics *in vivo* is due to properties other than their antioxidant effect (Halliwell et al., 2005; Lotito and Frei, 2006; Pietta, 2000; Scalbert et al., 2005). Interest in the health benefits of bioactive metabolites has resulted in extensive research quantifying the levels of polyphenolics and other antioxidants in different tomato genotypes and products (Bino et al., 2005; Capanoglu et al., 2008; Frusciante et al., 2007; Hernandez et al., 2007; Lenucci et al., 2006; Long et al., 2006; Minoggio et al., 2003; Overy et al., 2005; Peng et al., 2008; Rousseaux et al., 2005; Stewart et al., 2000; Tokusoglu et al., 2003; Torres et al., 2005).

Regardless of the mechanism, the health benefits of consuming an adequate amount of plant derived polyphenolics are clear. Although tomatoes are a major

dietary source of phenolics because they are consumed in large quantities (Economic Research Service, 2008), they are relatively low in phenolics compared to other vegetables (Vinson et al., 1998) with most of the phenolics being concentrated in the skin (Kushad et al., 2003). This has resulted in numerous efforts to increase the levels of polyphenolics in tomato fruit using wild species (Willits et al., 2005) and transgenic approaches (Bovy et al, 2007; Schijlen et al., 2004; Schijlen et al., 2006; van der Rest, et al. 2006), with limited success. Although Willits et al. (2005) identified tomato accessions with high expression levels of polyphenol biosynthetic genes, the tomato lines derived from crosses to these accessions were infertile and the research was discontinued. In all likelihood, the genes responsible for flavonoid accumulation would have been linked to unfavorable alleles as well. Early transgenic approaches targeted only structural genes or utilized transcription factors from maize and generated flavonoids only in tomato peel, where they are normally expressed (Bovy et al., 2007). Recently, transgenic approaches using transcription factors from snapdragon (*Antirrhinum majus*) or *Arabidopsis* were more successful in generating high levels of anthocyanins (Butelli et al., 2008) or flavonols (Luo et al., 2008) throughout the tomato fruit. However, due to safety concerns over genetically modified foods, the benefits of this research to the consumer may not be realized for some time.

The popularity of these bioactive health compounds is rather faddish. For example, marketers in the blueberry industry assure me that blueberry sales have received a huge boost due to articles in the popular press about the health benefits of the blueberry phenolic compounds. Without a doubt, many horticultural crops besides blueberries contain beneficial phenolics; and most nutritionists would probably agree that Americans would be better off if they just ate more fruits and vegetables, period. Similarly, current television ads for energy drinks and granola bars boast about added antioxidants, when in fact the antioxidant value of plant bioactive compounds *in vivo* has come under fire in the medical research community. So, are increased levels of phenolics (or other bioactive compounds) in crop species a worthwhile breeding target?

On the one hand, the actual mechanism of action for these compounds is usually not known. In some cases the mechanism may be very specific, as in the case of proanthocyanidins and urinary tract infections. Recent research has demonstrated that action of cranberry proanthocyanidins with a particular stereochemistry (A linkages) in cranberries prevent the adhesion of bacteria to bladder cells, while other proanthocyanidins (those with B type linkages) from cranberry and other plants have much less of an effect (Howell, 2005). In other cases, the action may due to synergistic effects of several types of compounds. Without knowing more about which compounds are beneficial, the actual target is rather vague and plant breeders may be putting the cart before the horse. Another concern is that more may not necessarily be better, as the bioavailability of many of these compounds may be low (Espín et al., 2007). In reality, some of these compounds may have been bred out of crops during the domestication process because they negatively affect quality. In an attempt to increase the levels of kaempferol and quercetin in modern potato cultivars to the levels found in Andean land-races, Rommens et al. (2008) found that the resulting product also had increased browning (probably due to higher levels of chlorogenic acid).

On the other hand, the levels of phenolics attainable through traditional methods are unlikely to be toxic (Harwood et al., 2007). And despite the fact that the exact mechanism is not yet known, many of them do have established benefits. Though a golden era of functional foods when compounds with specific, established health benefits are produced at high levels in crop plants in a targeted fashion is not yet a reality, increasing the consumption of a variety of phenolics in the meantime has more benefits than risks. Because the trend towards convenient food only seems to be increasing, increasing the levels of these healthful compounds, especially in processed foods, is likely to have a positive impact on people's lives. In a model species like tomato, creating varieties with different phenolic profiles also provides a valuable research tool for studying the effects of these compounds, since their effects are likely to be synergistic and dependent on the matrix they are present in. Finally, many of these compounds are pigments that can contribute novel colors to crops, making them more fun to eat, and increasing consumption of fruits and vegetables. And eating more

fruits and vegetables is exactly what your nutritionist (or hopefully your grandma) would tell you to do.

The objective of the present research was to continue Mes' (2005) research in breeding a tomato cultivar with anthocyanin in the fruit, and to extend it by developing tomatoes high in other polyphenolic compounds, using traditional (non-transgenic) approaches.

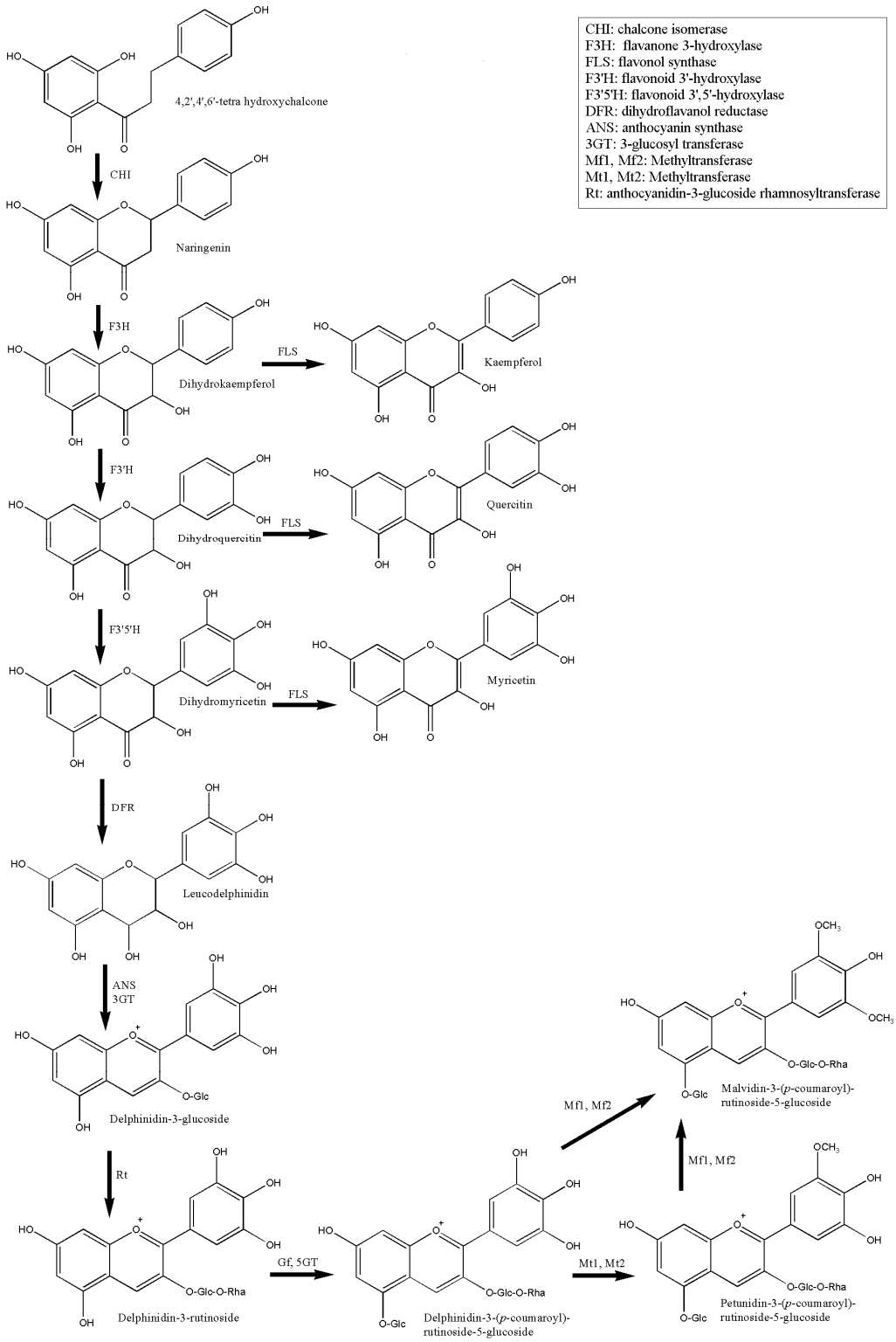


Figure 1.1. Tomato anthocyanin biosynthetic pathway, from Mes (2005).

Genes controlling anthocyanin biosynthesis

A hypothetical anthocyanin biosynthetic pathway for tomato is depicted in Figure 1.1. Regulation of anthocyanin biosynthesis has become a model system for complex gene regulation since the phenotype (pigmented tissue) is so easily observed. Genetic evidence indicates that the activity of anthocyanin biosynthetic genes is regulated at the transcriptional level by a complex of three interacting proteins, the MYB-bHLH-WD40 protein complex (Broun 2005). Figure 1.2 depicts a model of this interaction.

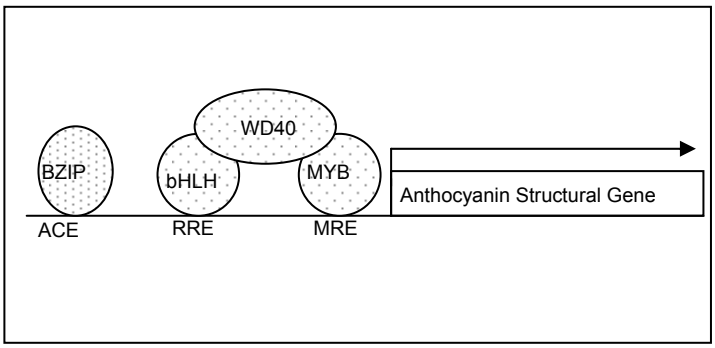


Figure 1.2. A model of the MYB-bHLH-WD40 complex based on Hartmann et al. (2005). Abbreviations: ACE, ACGT-containing element; bHLH, basic helix-loop-helix; BZIP, basic region/leucine zipper; MRE, MYB Responsive element; RRE, R-responsive element.

MYB and basic helix-loop-helix (bHLH) proteins are DNA binding transcription factors and WD40 proteins are known to stabilize protein-protein interactions. The WD40 protein most likely has the function of promoting the transient interaction of the two types of transcription factors (Van Nocker and Ludwig 2003). The MYB regulators appear to have greater functional specificity than those of their bHLH partners (Broun, 2005). The bHLH factors appear to activate a range of overlapping biosynthetic pathways, while the MYB factors are required to activate the bHLH factors in specific tissues or only in certain branches of a pathway (Ramsay and Glover, 2005). MYBs in plants have either one repeat of the MYB domain (R3 MYBs) or two repeats of the MYB domain (R2R3 MYBs) (Kranz et al., 2000). The

R3 MYBs typically act as negative regulators, while the R2R3 MYBs typically act as transcriptional activators. Specific binding sites for the bHLH and MYB proteins (the R-responsive element and the MYB-responsive element, respectively) have been identified in the promoters of anthocyanin structural genes in *Arabidopsis* (Hartmann et al., 2005). An ACGT-element (ACE) binding site necessary for the light responsiveness of anthocyanin structural genes has also been discovered (Hartmann et al., 2005). Basic region/ leucine zipper (BZIP) transcription factors are known to bind ACE elements, but anthocyanin mutants for this type of protein have not yet been found (Hartmann et al., 2005).

Most of the regulators of anthocyanin biosynthesis discovered to date are activators of transcription. Several candidates for negative regulators of anthocyanin biosynthesis have been advanced, including genes that regulate protein turnover of the MYB via the ubiquitination pathway, R2R3 MYBs that inhibit transcription, and R3 (single MYB repeat) MYBs that compete with the presence of activator type MYBs in the MYB-bHLH-WD40 complex but do not activate transcription (Quattrocchio et al., 2006).The *Arabidopsis* gene *ICX1* has been identified as a negative regulator of flavonoid production and epidermal cell fates in response to a broad range of environmental stimuli (Wade et al., 2003). The strawberry *FaMYB1* transcription factor is one example of a R2R3 MYB that acts as a negative regulator (Aharoni et al., 2001). The R3 MYBs have been demonstrated to compete with the presence of R2R3 MYBs in the MYB-bHLH-WD40 complex, inactivating it. R3 MYBs are also capable of movement into neighboring cells, while the R2R3 MYBs are not (Zhao et al., 2008). Furthermore, the R3 MYBs are direct targets for transcriptional activation by R2R3 MYBs (Zhao et al., 2008). The most current model of root hair and trichome patterning thus explains the zone of lateral inhibition observed surrounding cells designated to become root hairs or trichomes (Savage et al., 2008). Recently, a single repeat MYB that acts as negative regulator anthocyanin biosynthesis in *Arabidopsis* was discovered (Dubos et al., 2008; Matsui et al., 2008).

MYB genes are obvious candidates for the primary anthocyanin regulatory genes in crop species. Searches to identify the genetic loci responsible for controlling anthocyanin production in crop species have identified MYB genes in maize (Paz-

Ares et al., 1987), apple (Takos et al. 2006), petunia (Quattrochio et al., 1999), pepper (Borovsky et al., 2004), potato (De Jong et al., 2005), morning glory (Morita et al., 2006), snapdragon (Schwinn et al., 2006), *Gerbera* (Laitenin et al., 2008), and grape (Lijavetsky et al., 2006).

Continuing OSU research on production of polyphenolics in tomato fruit

At the outset of this research, some basic questions about the anthocyanin genes remained in light of the work of Jones (2003) and Mes (2005). For one, it was not clear whether *Abg* and *Aft* are alleles of one locus or separate loci because of the difficulty of distinguishing their phenotypes in the field. Another problem was that *Abg* could not be fixed in a homozygous state because *Abg* lines segregated for sterility even in advanced generations. Whether this sterility is due to linkage drag of a recessive lethal, the *S. lycopersicoides* chromosome 10 inversion (Ji and Chetelat, 2003), or persistent incompatibility-induced sterility from combining *S. lycopersicoides* with *S. lycopersicum* was not determined.

The question of whether *Aft* and *Abg* are allellic or separate genes would be normally be determined with an allelism test by growing a large number of F_2 plants in the field to see if they segregate 15 anthocyanin fruited : 1 normal. The situation is complicated by the fact that *Abg* cannot be maintained in a homozygous state, making it necessary to select an *Abg*/*Aft* F_1 visually, since half of the F_1's would be expected to be *Aft*/_. Furthermore, the F_1 would most likely still segregate for sterility. Given the state of knowledge of the genes regulating anthocyanin biosynthesis, attempting to identify the underlying genes at the molecular level was an appropriate alternative to classical allelism tests. This approach has many additional benefits, such as the development of molecular markers for the gene itself. This research forms the basis of Chapter 2.

The research of Mes (2005), Mes et al. (2008), Sapir et al. (2008), and van Tuinen et al. (2006) has indicated that in addition to containing anthocyanins, *Aft* and *AftAft/atvatv* tomatoes have higher levels of total phenolics than normal tomatoes. It was therefore important to characterize the full spectrum of phenolic compounds that were present in *AftAft/atvatv* tomatoes. Some of these compounds might have health

benefits equal to those of the anthocyanins, and tomatoes containing them might be better received by consumers, who have developed preferences for red tomatoes. Numerous recessive mutants that fail to express anthocyanin have been discovered in tomato, some of which have been mapped to particular biosynthetic genes. By combining regulatory genes that increase flux through the flavonoid pathway with genes blocking anthocyanin biosynthesis (but not flavonol biosynthesis), it might be possible to redirect the flux in the pathway towards flavonols, creating a non-transgenic high flavonoid tomato. Such a line might have some of the health benefits of a high anthocyanin tomato and greater consumer appeal because of its normal color. Another advantage might be faster and more consistent germination, which has been documented for anthocyaninless mutants in tomato (Atanassova and Shtereva, 1995). This research forms the content of Chapter 3.

In addition, we searched for germplasm that could be used to increase the level of total phenolics in cultivated tomatoes, and introduce novel polyphenolics as well. The highest levels of total phenolics in tomato have been found in the small fruited tomato species *Solanum pimpinellifolium* (B. Juss.) Miller and *Solanum lycopersicum var. cerasiforme* Dunal (Hanson et al., 2004). Steven Tanksley and collaborators have designated a core set of 100 *S. l. var. cerasiforme* lines that provide an ideal entry point for evaluating small fruited tomatoes. These lines will supposedly be extensively characterized using DNA and phenotypic analysis, which will facilitate the identification of correlation between high phenolics and DNA markers or phenotypic traits. This research is summarized in Chapter 4.

References

Aharoni, A., C.H. Ric De Vos, M. Wein, Z. Sun, R. Greco, A. Kroon, J.N.M. Mol, and A.P. O'Connell. 2001. The strawberry FAMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J. 28:319-332.

Atanassova, B., and L. Shtereva. 1995. Anthocyaninless mutants- potential for improvement of germination ability in *Lycopersicon esculentum* Mill. Tomato Genet. Coop. Rpt. 45: 7-9.

Bino, R.J., C.H. Ric de Vos, M. Lieberman, R.D. Hall, A. Bovy, H.H. Jonker, Y. Tikunov, A. Lommen, S. Moco, and I. Levin. 2005. The light hyperresponsive high pigment-2dg mutation of tomato: alterations in the fruit metabolome. New Phytol. 166:427-438.

Borovsky, Y., M. Oren-Shamir, R. Ovadia, W. De Jong, I. Paran. 2004. The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to *Anthocyanin2* of *Petunia.* Theor. Appl. Genet. 109:23-29

Bovy, A., E. Schijlen, and R.D. Hall. 2007. Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. Metabolomics 3:399-412.

Broun, P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. Curr. Opin. Plant Biol. 8:272-279.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

Capanoglou, E., J. Beekwilder, D. Boyacioglu, R. Hall, and R. de Vos. 2008. Changes in antioxidant and metabolite profiles during the production of tomato paste. J. Agric. Food Chem. 56:964-973.

De Jong, W.S., N.T. Eannatta, D.M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor. Appl. Genet. 108:423-432.

Dubos, C., J. Le Gourrierec, A. Baudry, G. Huep, E. Lanet, I. Debeaujon, J.-M.. Routaboul, A. Alboresi, B. Weisshar, and L. Lepinec. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. Plant J. 55:940-953.

Economic Research Service, USDA. 2008. Vegetables and melons situation and outlook yearbook. VGS-2008. 30 October 2008. <http://www.ers.usda.gov/Publications/VGS/2008/05May/VGS2008.pdf>.

Espin, J.C., M.I. Garcia-Conesa, and F.A. Tomas Barbaran. 2007. Nutraceuticals: facts and fiction. Phytochemistry 68:2986-3008.

Frusciante, L., P. Carli, M.R. Ercolano, R. Pernice, A. Di Matteo, V. Fogliano, and N. Pelligrini. 2007. Antioxidant nutritional quality of tomato. Mol. Nutr. Food. Res. 51:609-617.

Halliwell, B., J. Rafter, and A. Jenner. 2005. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? Am. J. Clin. Nutr. 81(suppl):268S-276S.

Hanson, P.M., Yang, R., Wu, J., Chen, J., Ledesma, D., and Tsou S.C.S. 2004. Variation for antioxidant activity and antioxidants in tomato. J. Amer. Soc. Hort. Sci. 129:704-711.

Hartmann, U., M. Sagasser, F. Mehrtens, R. Stracke, and B. Weisshar B. 2005. Differential combinatorial interactions of *cis*-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol. Biol. 57:155-171

Harwood, M., B. Danielewska-Nikiel, J.F. Borzelleca, G.W. Flamm, G.M. Williams, and T.C. Lines. A critical review of the data related to the safety of quercetin and lack of evidence of in vivo toxicity, including lack of genotoxic/carcinogenic properties. 2007. Food Chem. Toxicol. 45:2179-2205.

Hernandez, M., E. Rodriguez, and C. Diaz. 2007. Free hydroxycinnamic acids, lycopene, and color parameters in tomato cultivars. J. Agric. Food Chem. 55:8604-8615.

Howell, A.B., J.D. Reed, C.G. Krueger, R. Winterbottom, D.G. Cunningham, and M. Leahy. 2005. A-type cranberry proanthocyanidins and uropathogenic bacterial anti-adhesion activity. Phytochemistry 66:2281-2291.

Ji, Y. and R.T. Chetelat. 2003. Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic and substitution lines in tomato. Theor. Appl. Genet. 106:979-989.

Jones, C.M. 2000. Evaluation of carotenoids and anthocyanins in high pigment, processing, heirloom and anthocyanin fruit tomatoes. Ph.D. Dissertation, Oregon State University, Corvallis.

Jones CM, Mes P, and J.R. Myers. 2003. Characterization and inheritance of the *Anthocyanin fruit* (Aft) tomato. J. Heredity 94:449-456

Kandaswami, C., L.T. Lee, P.P. Lee, J.J. Hwang, F.C. Ke, Y.T. Huang, and M.T. Lee. 2005. The antitumor activities of flavonoids. In Vivo. 5:895-909.

Kris-Etherton, P.M., K.D. Hecker, A. Bonanome, S.M. Coval, A.E. Binkoski, K.F. Hilpert, A.E. Griel, and T.D. Etherton. 2002. Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. Am. J. Med. Suppl. 9B:71S-88S.

Kranz, H., K. Scholz, and B. Weisshar. 2000. C-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. Plant J. 21:231-235.

Kushad M.M., J. Masiunas, M.A.L. Smith, W. Kalt, and K. Eastman. 2003. Health promoting phytochemicals in vegetables. Hort. Rev. 28:125-185.

Laitenin, R.A.E, M. Ainasoja, S.V. Broholm, T.H. Teeri, and P. Elomaa. 2008. Identification of target genes for a MYB-type anthocyanin regulator in *Gerbera hybrida*. J. Exp. Bot. 59:3691-3703.

Lenucci, M.S., D. Cadinu, M. Taurino, G. Piro, and G. Dallessandro. 2006. Antioxidant composition in cherry and high-pigment tomato cultivars. J. Agric. Food Chem. 54: 2606-2613.

Li, Y., H. Fang, and W. Xu. 2007. Recent advances in research of flavonoids as anticancer agents. Mini Rev. Med. Chem. 7:663-78.

Lijavetsky D., L. Ruiz-Garcia, J.A. Cabezas, M. T. De Andrés, G. Bravo, A. Ibáñez, J. Carreño, F. Cabello, J. Ibáñez, and J. Martínez-Zapater. 2006. Molecular genetics of berry colour variation in table grape. Mol. Gen. Genomics 276:427-435.

Lin, J.-K., and M.-H. Weng. 2006. Flavonoids as nutraceuticals, p.213-238. In: Grotewold, E. (ed). The Science of Flavonoids. Springer, NY.

Long, M., D.J. Millar, Y. Kimura, G. Donovan, J. Rees, P.D. Fraser, P.M. Bramley, and G.P. Bolwell. 2006. Metabolite profiling of carotenoid and phenolic pathways in carotenoid and transgenic pathways in mutant and transgenic lines of tomato: Identification of a high antioxidant fruit line. Phytochemistry 67:1750-1757.

Lotito, S.B. and B. Frei. 2006. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? Free Radical Biol. & Med. 46:1727-1746.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56: 316-326.

Matsui, K., Y. Umemura, and M. Ohme-Takagi. 2008. AtMYBL2, a protein with a single myb domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. Plant J. 55: 954-967.

Mes, P. 2005. Breeding tomatoes for improved antioxidant activity. Ph.D. Dissertation, Oregon State University, Corvallis.

Mes, P.J., P. Boches, J.R. Myers, and R. Durst. 2008. Characterization of tomatoes expressing anthocyanin in the fruit. J. Amer. Soc. Hort. Sci. 133:262-269.

Middleton, E. and C. Kandaswami. 1993. The impact of flavonoids on mammalian biology: implications for immunity, inflammation and cancer. p. 589-618. In: J.B. Harborne (ed.). The flavonoids: advances in research since 1986. Chapman and Hall, London.

Minoggio, M., Bramati, L., Simonetti, P., Gardana, C., Lemoli, L., Santangelo, E., Mauri, P.L., Spigno, P., Soressi, G.P., and Pietta, P.G. 2003. Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. Ann. Nutr. Metab. 47:64-69.

Morita, Y., M. Saitoh, A. Hoshino, E. Nitasaka, S. Iida. 2006. Isolation of cDNAs for R2R3-MYB, bHLH, and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese morning glory. Plant Cell Phys. 47:457-470.

Overy, S.A., H.J. Walker, S. Malone, T.P. Howard, C.J. Baxter, L.J. Sweetlove, S.A. Hill, and W.P. Quick. 2005. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. J. Exp. Bot. 56:287-296.

Peng, Y., Y. Zhang, and J. Ye. 2008. Determination of phenolic compounds and ascorbic acid in different fractions of tomato by capillary electrophoresis with electrochemical detection. J. Agric. Chem. 56:1838-1844.

Paz-Ares, J., D. Ghosal., U. Wienard., P.A. Peterson, and H. Saedler. 1987. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. 6:3553-3558.

Pietta, P.-G. 2000. Flavonoids as antioxidants. J. Nat. Prod. 63:1035-1042.

Quattrocchio, F., A. Baudry, L. Lepiniec, and E. Grotewold. 2006. The regulation of flavonoid biosynthesis, p.97-122. In: Grotewold, E. (ed). The Science of Flavonoids. Springer, NY.

Quattrocchio, F., J. Wing, K. Van der Woude, E. Souer, N. de Vetten, J. Mol, and R. Koes. 1999. Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. Plant Cell 11:1433-1444.

Ramsay, N.A., B.J. Glover. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10:62-70.

Ren, W., Z. Qiao, H. Wang, L. Zhu, and L. Zhang. 2003. Flavonoids: promising anticancer agents. Med. Res. Rev. 23:519-534.

Rousseaux, M.C., C.M. Jones, D. Adams, R. Chetelat, A. Bennet, and A. Powell. 2005. QTL analysis of fruit antioxidants in tomato using *Lycopersion pennellii* introgression lines. Theor. Appl. Genet. 111: 1396-1408.

Rommens, C.M., Richeal, C.M., Yan, H., Navarre, D.A., Ye, J., Krucker, M., and K. Swords. 2008. Engineered native pathways for high kaempferol and caffeoylquinate production in potato. Plant Biotechnol. 6:870-886.

Sapir, M., M. Oren-Shamir, R. Ovadia, M. Reuveni, D. Evenor, Y. Tadmor, S. Nahon, H. Shlomo, L. Chen, A. Meir, and I. Levin. 2008. Molecular aspects of *Anthocyanin fruit* tomato in relation to *high pigment-1*. J Hered. 99:292-303.

Savage, N.S., T. Walker, Y. Wieckowski, J. Scheifelbein, L. Dolan, and N.A.M Monk. 2008. A mutual support mechanism through intercellular movement of CAPRICE and GLABRA3 can pattern the *Arabidopsis* root epidermis. PLoS Biol. 6:e235.

Scalbert, A., I.T. Johnson, and M. Saltmarsh. 2005. Polyphenols: antioxidants and beyond. Am. J. Clin. Nutr. 81(suppl):215S-217S.

Schijlen, E., C.H. Ric de Vos, H. Jonker, H. van den Broeck, J. Molthoff, A. van Tunen, S. Martens, and A. Bovy. 2006. Pathway engineering for phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol. J. 4:433-444.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648.

Schwinn, K., J. Venail, Y. Shang, S. Mackay, V. Alm, E. Butelli, R. Oyama, P. Bailey, K. Davies, and C. Martin. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell 18:831-851.

Stewart, A.J., S. Bozonnet, W. Mullen, G.I. Jenkins, M.E.J. Lean, and A. Crozier. 2000. Occurrence of flavonols in tomatoes and tomato-based products. J. Agric. Food Chem. 48:2663-2669.

Takos, A.M., Jaffe, F.W., Jacob, S.R., Bogs, J., Robinson, S.P., Walker, A.R. 2006. Light induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol. 142:1216-1232.

Tokusoglu, O., M.K. Unal, and Z. Yildirim. 2003. HPLC-UV and GC-MS characterization of the flavonol aglycons quercetin, kaempferol, and myricitin in tomato pastes and other tomato based products. Acta Chromatographica 13:196-207.

Torres, C.A., N.M. Davies, J.A. Yanez, and P.K. Andrews. 2005. Disposition of selected flavonoids in fruit tissues of various tomato (*Lycopersicon esculentum* Mill.) genotypes. J. Agric. Food. Chem. 53:9536-9543.

van der Rest, B., S. Danoun, A.M. Boudet, and S.F. Rochange. 2006. Down-regulation of cinnamoyl-coA reductase in tomato (*Solanum lycopersicum* L.) induces dramatic changes in soluble phenolic pools. J. Exp. Bot. 57:1399-1411.

Van Nocker, A., and P. Ludwig. 2003. The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. BMC Genomics 4:50-61.

van Tuinen, A., C.H Ric de Vos, R.D. Hall, L.H.W. van der Plas, C. Bowler, and R.J. Bino. 2006. Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from matural light hyperresponsive mutants, p. 339-356. In: P.K. Jaiwal (ed.) Plant Genetic Engineering Vol. 7: metabolic engineering and molecular farming.

Vinson J, Y. Hao, X. Su, and L. Zubik. 1998. Phenol antioxidant quantity and quality in foods: vegetables. J. Agric. Food Chem. 46:3630-3634.

Wade, H.K., A.K. Sohal, and G.I. Jenkins. 2003. *Arabidopsis* ICX1 is a negative regulator of several pathways regulating flavonoid biosynthesis genes. Plant Physiol. 131:707-715.

Willits, M.G., C.M. Kramer, R.T.N Prata, V. De Luca, B.G. Potter, J.C. Steffens, and G. Graser. 2005. Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. J. Agric. Food Chem. 53:1231-1236.

Zhao, M., K. Morohashi, G. Hatlestad, E. Grotewold, and A. Loyd. 2008. The TTG1-bHLH-MYB complex controls cell fate and and patterning through direct targeting of regulatory loci. Development. 135:1991-1999.

CHAPTER 2

Associations Between MYB Transcription Factors and Anthocyanin Genes in Tomato

Peter S. Boches and James R. Myers

Abstract

PCR primers were designed to amplify genes involved in anthocyanin production in tomato (*Solanum lycopersicum* L.) using publicly available tomato, potato (*Solanum tuberosum* L.), and petunia (*Petunia x hybrida* Juss.) DNA sequences. The resulting PCR amplification products from normal and purple fruited tomatoes were examined for polymorphism associated with purple fruit color. An initial screening indicated that cultivated tomatoes had a 1,100 bp allele of a MYB transcription factor (*SlAn2*), whereas anthocyanin fruited tomatoes with the *Anthocyanin fruit* (*Aft*) gene in their pedigree had a 700 bp allele (*ScAn2*, from *Solanum chilense* [Dunal] Reiche). The co-segregation of the 700 bp allele with the *Aft* gene was verified in a 'Legend' x 'LA 1996' (*Aft)* F_2 population. Out of the 43 individuals with the *anthocyanin gainer* (*ag*) phenotype from an *ag* x *Aft* F_2 population, zero had the 700 bp allele of *ScAn2* from the *Aft* parent. This indicated tight linkage between *ag* and *SlAn2*. The *An2* gene from anthocyanin fruited and cultivated tomato was sequenced. The 1100 bp allele found in cultivated tomato resulted from the insertion of a Miniature Inverted-repeat Transposable Element (MITE) into intron 2. Expression analysis of *An2* and the previously described *Ant1* gene in normal and purple fruited tomatoes indicated that transcription of *An2* but not *Ant1* was strongly correlated with anthocyanin and *dihydroflavonol 4-reductase* (*DFR*) expression in fruit. Sequence analysis indicated that *SlAnt1* is homologous to *S. tuberosum An1* (*StAn1*) and *SlAn2* is homologous to *S. tuberosum An2* (*StAn2*). Using 3'RACE, additional *SlAn2* and *SlAnt1* homologs were isolated.

Introduction

Tomato is an important food in the human diet, being the second most widely consumed vegetable in the United States (Economic Research Service, 2008). Anthocyanins are part of a class of plant derived compounds called polyphenolics with a wide range of reported health benefits (Kushad et al., 2003) and biological activities (Middleton and Kandaswami, 1993). Although tomatoes are a major dietary source of phenolics because they are consumed in large quantities, they are relatively low in phenolics compared to other vegetables (Vinson et al., 1998) with most of the

phenolics being concentrated in the skin (Kushad et al., 2003). One of the current objectives of the OSU tomato breeding program is to release tomato varieties containing anthocyanins, which impart health benefits and a novel purple or black color. Taxonomically, cultivated tomato is a member of a clade of red fruited tomato species that, along with *Solanum pimpinellifolium* L. and *Solanum cheesmaniae* L. Riley (Fosberg), lack significant anthocyanin production in the fruit (Moyle, 2008). The ancestors of red fruited tomato species have green fruit and often produce anthocyanin in the fruit in a striped or spotted pattern (Moyle, 2008). The restoration of anthocyanin production in the fruit of *S. lycopersicum* has been accomplished by introgression of genes from green fruited tomato species such as *Anthocyanin fruit (Aft*, from *S. chilense;* Jones et al., 2003), *Aubergine* (*Abg,* from *Solanum lycopersicoides* Dunal; Rick et al., 1994) and *Purple Smudge* (from *S. peruvianum* L.; Young, 1954). The gene *atv (atroviolaceum*), introgressed from *S. cheesmaniae*, causes the production anthocyanin mainly in the foliage of cultivated tomato fruits, but when combined with *Aft* or *Abg* results in a large of the amount of anthocyanin in fruit (Mes et al., 2008; van Tuinen et al., 2006).

Anthocyanin formation is regulated at the transcriptional level by a complex of three interacting proteins, the MYB-bHLH-WD40 protein complex (Broun, 2005). MYB and basic helix-loop-helix (bHLH) proteins are DNA binding transcription factors and WD40 proteins are known to stabilize protein-protein interactions. The bHLH factors appear to activate a range of overlapping biosynthetic pathways, while the MYB factors are required to activate the bHLH factors in specific tissues or only in certain branches of a pathway (Ramsay and Glover, 2005). Over-expression of MYB and bHLH transcription factors from other species in tomatoes has resulted in the accumulation of large amounts of anthocyanin in tomato fruit (Butelli et al., 2008).

Using activation tagging, Matthews et al. (2003) isolated a MYB transcription factor from tomato, *SlAnt1* (formerly *LeAnt1*), that induced anthocyanin in leaves (and sometimes fruit) when ectopically expressed. Mathews et al. (2003) did not detect expression of *SlAnt1* in normal tomatoes, however. Recently, Sapir et al. (2008) reported on linkage between *Aft* and *ScAnt1* and amino acid sequence polymorphisms between *Sc*ANT1 (from *Aft*) and *Sl*ANT1 that could potentially affect the function of

the protein. Sapir et al. (2008) detected expression of *Ant1* in *S. lycopersicum* and *Aft* at similar levels in tomato fruit.

In the present research, a candidate gene approach was used to investigate whether anthocyanin genes such as *Aft* are encoded by homologs of known DNA sequences, such as the MYB encoding *An2* of petunia (*PhAn2*). Previously, De Jong et al. (2004) had mapped many anthocyanin related genes in tomato, and Borovsky et al. (2004) had shown that the *A* locus controlling anthocyanin accumulation in pepper fruit is a MYB homologous to *PhAn2*. The corresponding locus in tomato (*SlAn2)* was mapped using *PhAn2* as an RFLP probe (De Jong et al., 2004). *SlAn2* was present in two copies, tightly linked ~1 cM north of RFLP marker TG233, on the long arm of chromosome 10. (One of these copies may represent *SlAnt1*, which is also a *PhAn2* homolog). This is very close to the previously mapped location of *ag*, which co-segregated with TG233 (Tanksley et al., 1992). The *ag* mutation is characterized by the absence of anthocyanin in hypocotyls, and limited anthocyanin expression on the underside of cotyledons and leaves. *Abg* has also been reported to map to chromosome 10 (Rick et al., 1994) although its exact placement has been problematic because of a large pericentric inversion (Ji and Chetelat, 2003). Here, we report the existence of a tomato gene (*SlAn2*) with sequence similarity to *PhAn2* and *StAn2*, distinct from *SlAnt1*, and highly expressed in fruit of *Aft*, *Purple Smudge,* and wild *S. chilense*.

Materials and Methods

Plant material. Seed of 'Legend' is maintained in the Oregon State University vegetable breeding program, seed of 'Purple Smudge Goldman' and 'Purple Smudge Orange Flesh' were obtained from Amy Goldman (pers. comm.), seed of 'Green Grape' was obtained from Seed Savers Exchange (Decorah, IA). All other seeds were obtained from the Tomato Genetics Resource Center (Davis, CA) (Table 2.1). Tomato seeds were sown in Sunshine SB40 professional growing mix (Sun Gro Horticulture, Bellevue, WA) in 5-cm-diameter plastic transplant cells in the greenhouse.

Table 2.1. Tomato genotypes used in the study of candidate anthocyanin genes in 2006, 2007, and 2008.

Name[1]	Anthocyanin Genes	Background	Description
Green Grape	NA[2]	NA	Released by Tater Mater seed in 1986. Developed from a 'Yellow Pear' x 'Evergreen' cross.
IL 10-3	NA	M82	*Solanum pennelii* Correll introgression line for lower arm of chromosome 10.
LA0797	*atv (atroviolaceum)*	VF36	Contains an anthocyanin inducing gene introgressed from *S. cheesmaniae*
LA1963	NA	NA	*S. chilense* wild germplasm
LA1996	*Aft (Anthocyanin fruit)*	Unknown (possibly cv. Vigoroz)	Introgression line, contains an anthocyanin inducing gene from *S. chilense*
LA2378	*Purple Smudge*	NA	Derived from a cross between *S. peruvianum* and *S. lycopersicum*
LA3163	*ag (anthocyanin gainer)*	'Ailsa Craig'	A mutation that results in delayed anthocyanin expression in leaves
LA3668	*Abg (Aubergine)*	Unknown	Introgression line, contains an anthocyanin inducing gene from *S. lycopersicoides*
LA3736	*atv (atroviolaceum)*	'Ailsa Craig'	Contains an anthocyanin inducing gene introgressed from *S. cheesmaniae*
'Legend'	wild type	NA	Most recent tomato variety release from Oregon State University
PI 290858	*Purple Smudge*	NA	Derived from a cross between *S. peruvianum* and *S. lycopersicum*
Purple Smudge Goldman	*Purple Smudge*	NA	Purple Smudge variety obtained from Amy Goldman
Purple Smudge Orange Flesh	*Purple Smudge*	NA	Purple Smudge variety obtained from Amy Goldman
VF36	wild type	NA	Commonly used genetic stock.

[1]Accessions obtained from the Tomato Genetic Resource Center except where alternative source noted. [2]NA= not applicable

RNA expression was carried out using greenhouse grown material. Seedlings were transplanted into 3.8-L pots filled with Special Blend potting mix and 20 g of 14N–6.1P–11.6K slow-release fertilizer (Simplot 14–14–14) after 3 weeks. Plants were given 10 g of supplemental fertilizer after 4 weeks in the pots. Greenhouse temperature was set to 18°C night and 25°C day. Supplemental lighting was provided for 16 hours per day by 400-W metal halide and 400-W high-pressure sodium Sun System 3 high-intensity discharge lamps (Sunlight Supply, Woodland, WA).

In order to verify co-segregation of the *Aft* and *ag* genes with the *SlAn2* DNA sequence, LA1996 (*Aft*) x 'Legend' and LA1996 x LA3163 (*ag*) were crossed using standard greenhouse crossing techniques (Rick, 1980). The LA1996 x 'Legend' F_2 population was phenotyped for the *Aft* trait at the Vegetable Research Farm using 123 plants in the summer of 2006 and 60 plants in the summer of 2007. Plants were transplanted in late May into rows 1.8 m apart with 60 cm within-row spacing. Fertilizer (505 $kg \cdot ha^{-1}$ of 12N-29P-10K-4S) was banded before transplanting. Plants were irrigated at weekly intervals until mid-August when water was withheld. Phenotyping for the *Aft* trait was carried out at the mature green stage (when anthocyanin pigmentation is prominent) and verified at the ripe fruit stage. Due to background effects such as leaf cover, some F_2 plants had very little fruit anthocyanin. In 2007, 176 LA1996 x LA3163 F_2 individuals were phenotyped for *ag* in the greenhouse at the seedling stage. Twenty five LA1996 x LA3163 F_2 plants with the *ag* phenotype were transplanted to the field in summer 2007 and phenotyped for *Aft* as was done for the LA1996 x 'Legend' population (above).

For screening tomato genetic stocks, DNA was extracted from two or more greenhouse grown individuals. In 2006, DNA of LA1996 x 'Legend' F_2's was extracted from greenhouse grown material before transplanting to the field. In 2007, DNA of LA1996 x 'Legend' F_2's was extracted from field grown material at the time of phenotyping. DNA extraction of 95 LA1996 x LA3163 F_2 individuals was carried out using greenhouse grown material at the first true leaf stage, at the time of phenotyping. RNA was extracted from greenhouse grown plants (three or more biological replicates per tissue). Leaf tissue was harvested for RNA extraction at two

developmental stages: young (third set of newly expanded true leaves) and old (fruit at mature green stage). Fruit for RNA extraction was harvested at 10 days after pollination (DAP), the immature green stage (30 DAP), and mature green stage (45 DAP). Anthers were harvested for RNA extraction slightly before pollen shed.

The distribution of anthocyanin cells in fruit of TGRC accessions was characterized in hand sections of greenhouse grown fruit at the mature green to breaker stage, at 10X or 40X magnification.

Nucleic acid extraction. DNA extraction was carried out according to the method of Acotto et al. (2000) with modifications, or by the method of Fulton et al. (1995). Tissue for RNA extraction was frozen in liquid nitrogen immediately after harvest. RNA was extracted with the RNEASY kit (Qiagen, Valencia, CA).

Polymerase Chain Reaction (PCR). DNA sequences (from tomato or closely related species) for candidate genes were obtained from Genbank or SolGenes websites. SolGenes intron finder software (accessed at the Solgenes Website) was used to predict intron locations and primers were designed to flank introns when possible in order to increase the chances of amplifying a polymorphic PCR product. Primers were designed using default settings on Primer3 (Rozen and Skaletsky, 2000). Primer were synthesized by Operon (Huntsville, AL). Candidate gene primers were screened in PCR reactions using template DNA from the tomato lines in Table 2.2. Primers for existing markers predicted to flank the *An2* gene on chromosome 10 were obtained from the SolGenes website (Table 2.3).

Table 2.2. Candidate genes examined for polymorphism in normal tomatoes and anthocyanin mutants. Unless indicated otherwise (see footnotes), primers were designed using the Primer3 software to amplify the specified sequences from the Genbank or SolGenes websites (Sequence ID). The *An2* and *Ant1* loci were sequenced using two sets of overlapping primers, the forward and reverse primers that amplify the 5' end of the gene are listed first (from top). The ANT1REV primer was used as an alternative to Ant1_594_1013R with higher specificity to the Sl*Ant1* locus.

Gene	Description	Origin	Sequence ID	Primer sequences (5' to 3')
SlAn2	MYB protein	Tomato	SGN-U228086 (SolGenes)	1F:GCATCGTTGGGAGTTAGGAA, 1R:AACGAGGACGAGAATGAGGA; 2F:GGAAGGACAGCAAACGATGT, 2R:AGAGAGAGGTTGAGTTTGACACAC
SlAnt1	MYB protein	Tomato	EF433416 (GenBank)	LeANTF:TTCATTGGGAGTGAGAAAAGGTTC[1], LeANTR:CATTGTGCTTGAGAAATACTTGCG[1]; ANT1_594_1013F:ACGACGCAAGTATTTCTCAAGCAC[2], ANT1_594_1013R:GGACTAGTTTAATCAAGTAGATTCCATAAGTCA[2], ANT1REV:GATTCCATAAGTCAATTTCAGCAG
AN11	WD40 protein	Tomato	SGN-U215976 (SolGenes)	F:GTCTTATCTCCCAGCGAACG, R:TCCATCAGCAGAAACAGACG
JAF13	bHLH protein	Tomato	SGN-E235386 (SolGenes)	F:TCAGGGGATCACTACCGAAC, R:TCCCATCAAGGTTGGAAGAC
CHS	chalcone synthase	Potato	U47738 (GenBank)	F:GCGACTCCTTCGAACTGTG[3], R:AAGTTTTTCGGGCTTTAGGC[3]
DFR	dihydroflavanol 4-reductase	Tomato	Z18277.1 (GenBank)	F:CATCTACGTTCCCACCAAGC, R:TCCAAGATGATTACAAAAGATGG
EF-1-α	housekeeping gene, RT-PCR control	Potato	AB061263 (GenBank)	F:ATTGGAAACGGATATGCTCCA[4], R:TCCTTACCTGAACGCCTGTCA[4]

[1]Primers from Matthews et al. 2003; [2]primers used in Sapir et al (2008) obtained from I. Levin, pers. comm.; [3]Primers from De Jong et al., 2004; [4]Primers from Nicot et al. 2005.

Table 2.3. Marker loci flanking *SlAn2* on the tomato linkage map. For each marker, the type, map location (chromosome, location in centimorgans), assay method (type of electrophoresis and/or restriction enzyme), expected sizes, sequence source, and the primer sequences are listed. Abbreviations: SSR, Simple Sequence Repeat; COSII, Conserved Orthologous Set II; CAPS, Cleaved Amplified Polymorphic Sequence.

Name	Type	Map	Assay	Sizes	Sequence	Primer Sequence (5' to 3')
SSR74	SSR	10, 74	capillary electrophoresis	203+199	SGN-C142969 (SolGenes)	F:ACTCACCATGGCTGCTTCTT, R:TTTCTTGAAGGGTCTTTCCC
SSR223	SSR	10, 75	2% agarose gel	191+190	SGN-M896 (SolGenes)	F:TGGCTGCCTCTTCTCTGTTT, R:TTTCTTGAAGGGTCTTTCCC
U221455	COSII	10, 81	2% agarose, polyacrilamide gel	750-870	SGN-M9542 (SolGenes)	F:AGGCGCTTCTTATTATCTTTCTC, F:ACCACAAGCAATCATTTCTACACC
TG233	CAPS	10, 86	cut w/ *AluI*, 2% agarose; sequenced PCR product	500/430 (*AluI*)	SGN-M37 (SolGenes)	F:CATGCCTTTTTCTTGGGATG, R:TGGAACCCCTTTAACTGTGC

PCR reactions were carried out in 15 µl of 1X Biolase NH4 reaction buffer, 2 mM $MgCl_2$, 200 µM each of dATP, dCTP, dGTP, and dTTP, 0.3 µM each of forward and reverse primers, 0.25 units of Biolase Taq DNA polymerase (Bioline Inc., Randolph, MA), and 2.5 ng of template DNA. The PCR program consisted of an initial denaturation at 94° C for 3 minutes, 35 cycles of 94° C for 30 s, 60° C for 30 s, 72° C for 30 s, followed by a 72° extension for 5 minutes. PCR products of candidate gene primers were separated on 2% agarose gel and visualized under UV light after staining with ethidium bromide. One hundred base pair DNA ladders from Promega (Madison, WI) or Invitrogen (Carlsbad, CA) were used as size standards on 2% agarose gels. PCR products from the DNA marker loci in Table 2.3 were separated on agarose gels as well. In addition, U221455 and TG233 products were separated on polyacrilamide gels. SSR74 PCR products were generated with a 6-carboxyfluorescein (6-FAM) fluorescently labeled forward primer (Operon, Huntsville, AL) and separated by capillary electrophoresis on an ABI3100 sequencer (Applied Biosystems, Foster City, CA). The ABI Genescan® and Genotyper® software (Applied Biosystems, Foster City, CA) were used to acquire and analyze capillary electrophoresis results. Polymorphism at the *Ant1* and TG233 loci was assayed using a Cleaved Amplified Polymorphic Sequence (CAPS) marker. The CAPS polymorphism was obtained by digesting 50 µL of PCR product with a restriction enzyme. For the *Ant1* locus, Ant1_594-1013F+R PCR products were cut with *NcoI* (Sapir et al 2008); TG233F+R PCR products were cut with *AluI*. The digested PCR product was separated on 2% agarose. All PCR results reported were verified in a duplicate reaction.

Semi Quantitative RT-PCR. cDNA was reverse transcribed from 1 µg of total RNA using the SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA). RNA samples were treated with DNAse (Promega, Madison, WI) prior to cDNA synthesis to remove DNA contamination. The absence of DNA contamination was verified with primers that had different sized PCR products from the genomic vs the cDNA templates (*Actin*, DFR, and *An2* 1F+R primers). A sample of LA1996 fruit mRNA with no reverse transcriptase added was used as an additional negative control.

Genomic DNA was used as a positive control. The cDNA was treated with RNAse cocktail (ISC Bioexpress) before second strand PCR reactions were carried out to improve priming. The gene specific primers in Table 2.1 were used for second strand reactions. For *An2*, the 1F+R primers were used. For *Ant1*, leANTF+R, Ant1_594_1013F+R, and Ant1_594_1013F + ANT1REV were used. Second strand PCR reactions were carried out in 15 µl of 1X Biolase NH4 reaction buffer, 2 mM $MgCl_2$, 200 µM each of dATP, dCTP, dGTP, and dTTP, 0.3 µM each of forward and reverse primers, 0.25 units of Biolase Taq DNA polymerase (Bioline Inc., Randolph, MA), and 1 ul of cDNA as template. The PCR program consisted of of an initial denaturation at 94° C for 3 minutes, followed by either 30, 32, or 35 cycles of 94° C for 30 s, 60° C for 30 s, 72° C for 60 s; followed by a 72° C extension for 5 minutes. Expression levels were normalized to *EF-1α* (Nicot et al., 2005) or *actin2* (Liu et al. 2005). All results reported were verified in at least one biological replicate. For leaf and fruit tissue of 'Legend' and LA1996, semi-quantitative RT-PCR was repeated using three or more biological replicates and the PCR products were quantitated from gel images with Scion Image software (Scion, Frederick, Maryland) using the gel2 macro. Optical density curves were calibrated using 25, 50, 100, 150, 250, and 500 ng of calf thymus DNA (Sigma, St. Louis, MO).

DNA sequencing. Alleles of the *An2*, *Ant1*, and TG233 loci were sequenced directly from PCR products purified with the Promega (Madison, WI) Wizard® SV gel and PCR clean-up system. When multiple PCR products were present the PCR products were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and the colony PCR products were sequenced. PCR products were sequenced on an Applied Biosystems ABI 3100 (Foster City, CA) by the Center for Genome Research and Biocomputing core facility at Oregon State University. The tomato cDNA clone cTOA-12-C2 (corresponding to the unigene SGN-U228086) was obtained from the SolGenes website and used as a positive control for sequencing. Sequence analysis was conducted with the BioEdit Software (Hall, 1999). Forward and reverse sequences were compiled into contigs using the CAP3 contig assembly program in BioEdit. DNA and protein sequences were aligned using ClustalW in BioEdit with default

alignment parameters. The intron splice sites and amino acid sequences were assembled manually from six frame translations and comparison to reference sequences, and the predicted protein sequence was aligned with previously described MYB protein sequences from the Solanaceae.

The 3' portion of the *SlAn2* locus, as well as loci similar to *SlAn2* and *SlAnt1*, were obtained by 3' Rapid Amplification of CDNA Ends (3' RACE) using the Bioline 3'RACE kit (Bioline, Randolph, MA). The gene specific primer used was: 3'GSP3 (GGAAGGACAGCAAACGATGT), designed from the 'Legend' sequence for *SlAn2*. Four separate mRNA samples were used for 3'RACE (newly expanded 'Legend' leaves, mature green 'Legend' fruit, newly expanded LA1996 leaves, or mature green LA1996 fruit).

Statistical analysis, phylogenetic analysis, and bioinformatics software. SAS software (Cary, NC) was used for analysis of variance, means separations using Fisher's Least Significant Difference (LSD), and calculation of Pearson's correlation coefficient (R). CLUSTALW was used to align DNA sequences and amino acid translations of *An2* and *Ant1*. Maximum likelihood trees were generated from amino acid alignments using the ProML program in BioEdit. DNA folding of the Legend and LA1963 MITEs was predicted using the UNAFOLD software (accessed online at the Rensselaer Bioinformatics Server Website; see also Zuker, 2003).

Results

Co-segregation of An2 *with* Aft*,* Ant1*, and* ag. During initial screening of candidate gene primers (Table 2.2) on 2% agarose gels, size polymorphisms were observed only for the *SlAn2* locus. Normal tomato cultivars such as 'Legend' and 'VF36' had an 1100 bp allele of *An2*. A 700 bp allele of *An2* was present only in the wild tomato species *S. pennellii*, *S. chilense*, and tomato genotypes with the *Aft* gene (Figure 2.1).

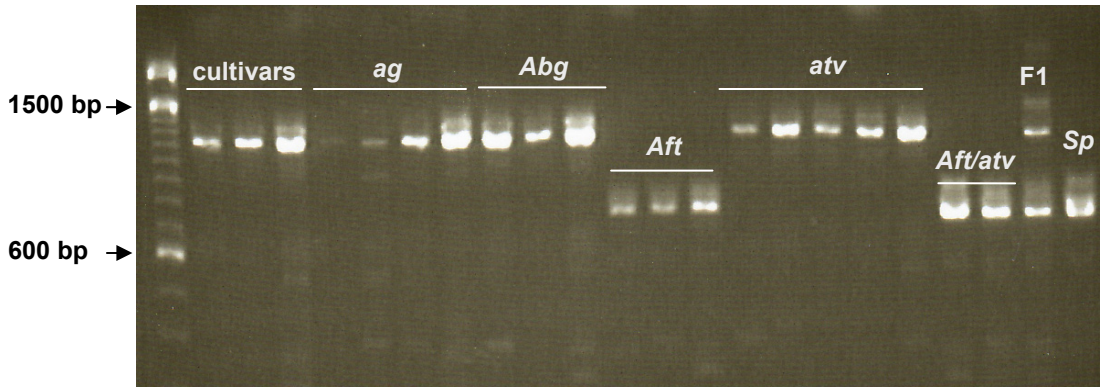


Figure 2.1. *An2* PCR polymorphism on 2% agarose gel stained with ethidium bromide. DNA from wild type tomatoes (cultivars 'Legend', 'VF36', and 'Green Grape'), anthocyanin deficient mutants (LA 3163 *ag*), single anthocyanin fruit mutants (LA 3668 *Abg*, LA1996 *Aft*, LA3736 *atv*), double anthocyanin fruit mutants (*Aft*/*atv*), an anthocyanin fruit mutant heterozygote (F_1, *Aft*-/*atv*-), and wild species (*Sp*, *Solanum pennellii* LA 1926). Far left, 100 bp DNA ladde (Invitrogen); right lane (-) is negative control (water template).

Tomato accessions and varieties with the *Purple Smudge* anthocyanin fruit gene had a slightly smaller (650 bp) allele of *SpAn2* (not shown). The *S. chilense* accession LA 1963 had individuals with a *ScAn2* allele that was slightly smaller than the 'Legend' *SlAn2* allele (~1050 bp) in addition to a 700 bp allele. Individuals of LA1963 with the 1050 bp allele had fruit containing anthocyanin (as is normal for this species). The 700 bp allele of *An2* also co-segregated with a CAPS polymorphism in the *ScAnt1* gene in 45 *Aft* x 'Legend' F_2 individuals (Figure 2.2a). The 700 bp allele of *ScAn2* completely co-segregated with the *Aft* phenotype (purple fruit color) in a total of 183 LA1996 x 'Legend' F_2 individuals (Figure 2.2b).

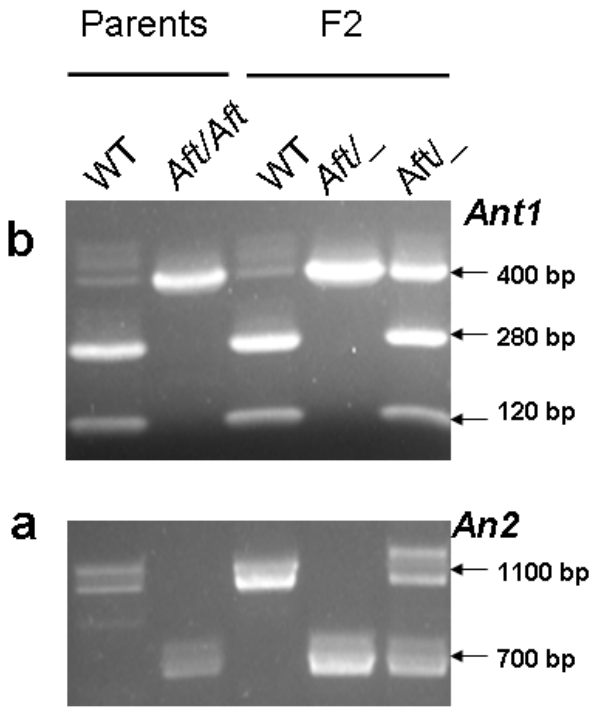


Figure 2.2. Co-segregation of *Ant1*, *An2* and *Aft*. Gel image of *Ant1* PCR products digested with *NcoI* restriction enzyme and *An2* PCR products separated on 2% agarose gels stained with ethidium bromide; from parental lines and 'Legend' x LA1996 F_2 genotypes. Parental genotypes are 'Legend' (WT) and LA1996 (*Aft*/*Aft*). The phenotypes of F_2 individuals are indicated on top of the gel, the WT lane was from a normal tomato fruit, the two *Aft*/_ lanes were from individuals displaying the *Aft* phenotype. (a) The *Ant1* polymorphism produced by the *Ant1* CAPS marker, (b) the *An2* polymorphism produced by the *An2* 1F+R primers. The 280 bp and 120 bp bands are *NcoI* restriction fragments.

When *Aft* (LA1996) was crossed to *ag* (LA3163), the progeny segregated 125 WT: 51 *ag*, which did not deviate significantly from the expected 3:1 ratio (χ^2=1.485, P=0.22). Twenty five *agag* segregants were grown to maturity, none of which displayed the *Aft* trait. This would be a very unlikely segregation ratio if the two genes are unlinked and *ag* is not epistatic to *Aft* (χ^2=75, P<0.001). It is theoretically possible that *ag* is not epistatic to *Aft*, since *ag* plants have the ability to produce anthocyanin on the undersides of older leaves. However, if *ag* is epistatic to *Aft*, it is imposible to determine whether *Aft* and *ag* are linked or unlinked based on phenotypic data. To

ascertain whether *ag* and *Aft* are linked or unlinked, we extracted DNA from 95 *Aft* x *ag* F_2 progeny and genotyped them at the *An2* locus. The 700 bp allele of *ScAn2* from *Aft* segregated in complete repulsion with the *ag* phenotype, indicating that the two loci are tightly linked or allelic.

Map location of Aft. Four DNA markers flanking the previously mapped location of *SlAn2* (Table 2.2) were examined for polymorphism in VF36, *Aft*, *Abg*, and *atv* genotypes. No size polymorphism was detected after separation on 2% agarose gels (all markers), polyacrilamide gels (U221455 and TG233), or using capillary electrophoresis (SSR74). No polymorphism was detected for the TG233 *AluI* CAPS marker, either. Sequencing of the TG233 locus in 'Legend' and LA1996 revealed a total of 20 single nucleotide polymorphisms (SNPs) in comparison to the reference sequence (SGN-M37) (Figure 2.3). Of these, 13 SNP alleles were shared between the reference sequence and the 'Legend' sequence, while 7 SNP alleles were shared between the *Aft* and 'Legend' sequence. The seven SNP alleles shared between *Aft* and 'Legend' are potentially sequencing errors in the reference sequence, since the TG233 sequence represents a single read and these SNPs occur near the 3' end. In contrast to the reference sequence, the *Aft* and 'Legend' sequences represent the contig of forward and reverse sequencing reads. Based on the four markers assayed, the *S. chilense* chromosomal introgression in LA1996 appears to be limited to the area between the U221455 and TG233 markers (the very tip of chromosome 10 in the current tomato linkage map).

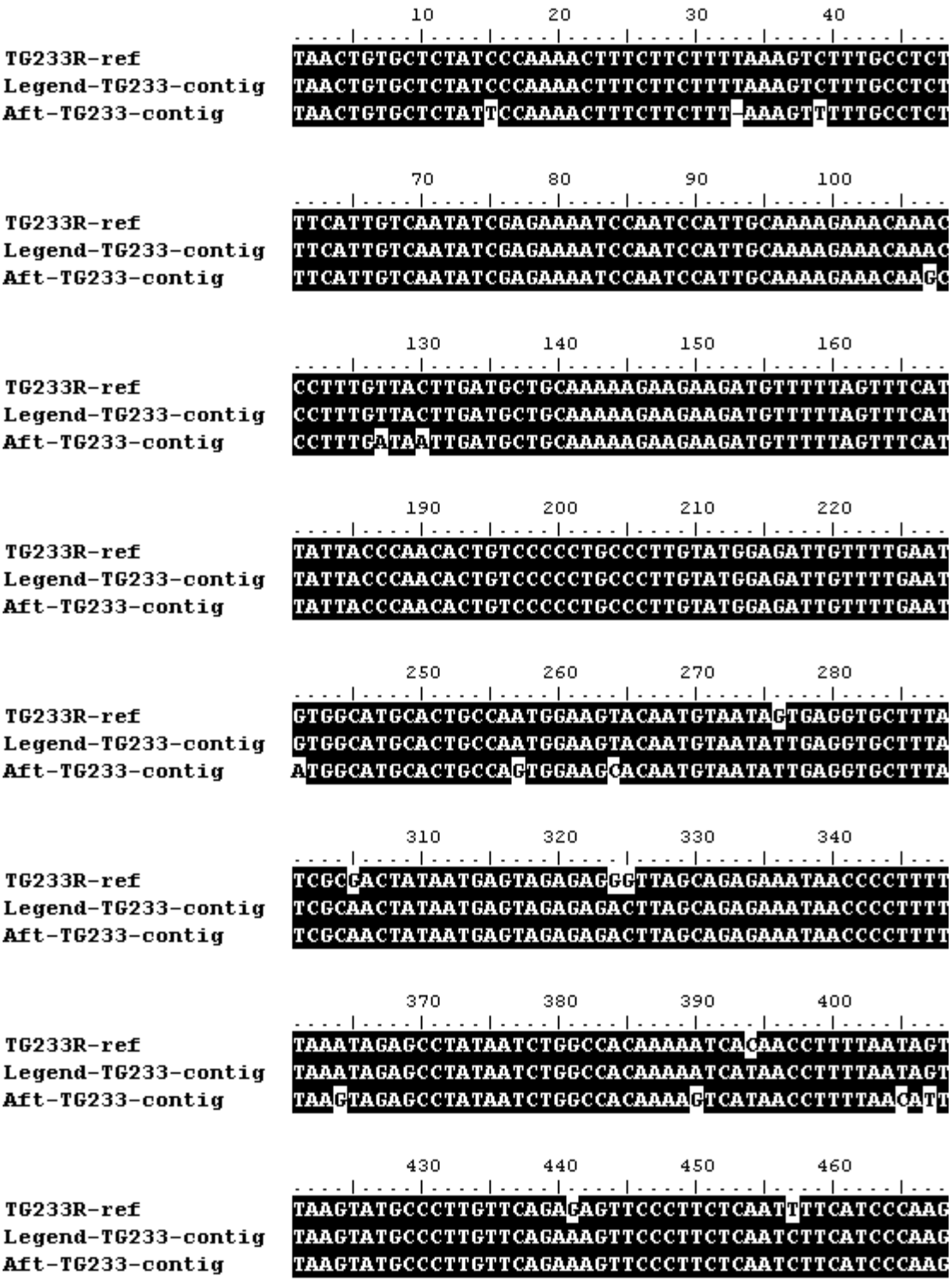


Figure 2.3. Alignment of TG233 sequences from *Aft* (Aft-TG233-contig) and 'Legend' (Legend-TG233-contig) tomatoes in comparison to the reference sequence (TG233R-ref). Identical nucleotides are shaded.

Semi-quantitative RT-PCR. Expression of the *An2* and *Ant1* genes was examined in leaf, fruit, and anther tissue of LA1996 (Aft), cv. Legend, LA 1963 (*S. chilense*), and PI290858 (*Purple Smudge*) (Figure 2.4). *An2* was strongly expressed in anthocyanin producing tissues such as leaves and the fruit of the genotypes LA1996, LA1963, and PI290858, detectable after 30 rounds of PCR amplification. Some *An2* expression was detected at low levels in non-anthocyanin producing tissues as well, such as fruit of 'Legend', but only after 35 cycles of PCR amplification. Levels of *An2* expression were lower in the fruit of PI290858 and LA1963 relative to LA1996. LA1996 had lower levels of *An2* in leaves relative to 'Legend' or LA1963 (LA1996 also had lower levels of *Ant1*, CHS, and DFR in leaves). *Ant1* expression was barely detectable (in leaf tissue from 'Legend' only) after 30 rounds of PCR amplification (not shown). After 35 cycles of PCR amplification *Ant1* expression was detected at low levels in all samples, the highest levels were found in leaf tissue of 'Legend' and LA1963.

The expression levels of *CHS*, *DFR, AN11*, and *JAF13* were examined in the same tissues. *CHS* expression was detectable in all samples after 30 cycles of PCR (Figure 2.4D), strongly so after 35 cycles of PCR (not shown). *DFR* expression was detected only in anthocyanin containing tissues (Figure 2.4E and F). The highest levels of *DFR* expression were found in older leaves of LA1996, older leaves of 'Legend', and fruit of LA1996. Expression of *DFR* was barely detectable in anthers of PI290858 after 35 cycles of PCR. The *AN11* and *JAF13* genes were strongly expressed in all genotypes and tissues examined (not shown).

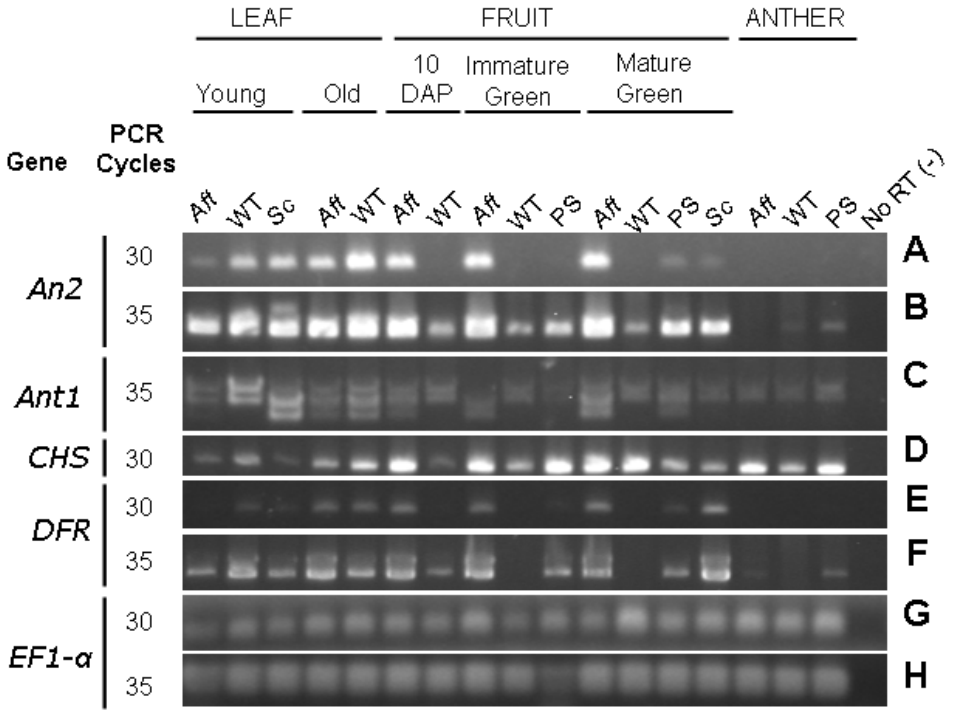


Figure 2.4. Gel images of semi-quantitative RT-PCR analysis of selected tissues of tomato and related species. Gene-specific primers for *An2* (A and B*), Ant1* (C)*, CHS* (D), *DFR* (E and F), and the housekeeping gene *EF1*-α (E and H) were used to amplify each cDNA source. The PCR products were separated on 2% agarose gels and stained with ethidium bromide. PCR was run for 30 cycles (A, D, E, and G) or 35 cycles (B, C, F, and H) (indicated in the column to the left of the gel image). The tissues for each cDNA (indicated above the images) are as follows: young leaf (third set of true leaves, newly expanded), old leaf (harvested from plants with mature fruit), fruit ten days after pollination (DAP), immature green fruit, mature green fruit, and anther. The following genotypes were sampled (abbreviations in parentheses): LA1996 '*Aft*' (*Aft*), 'Legend' (WT), LA 1963 *S. chilense* (*Sc*), and PI290858 'Purple Smudge' (PS). A total RNA sample from *Aft* fruit with no reverse transcriptase added was used as a negative control (No RT (-), far right lane). The primer sequences used are listed in Table 2.2; the 1F+R primers were used for *An2* and the Ant1_594_1014F+R primers were used for *Ant1*.

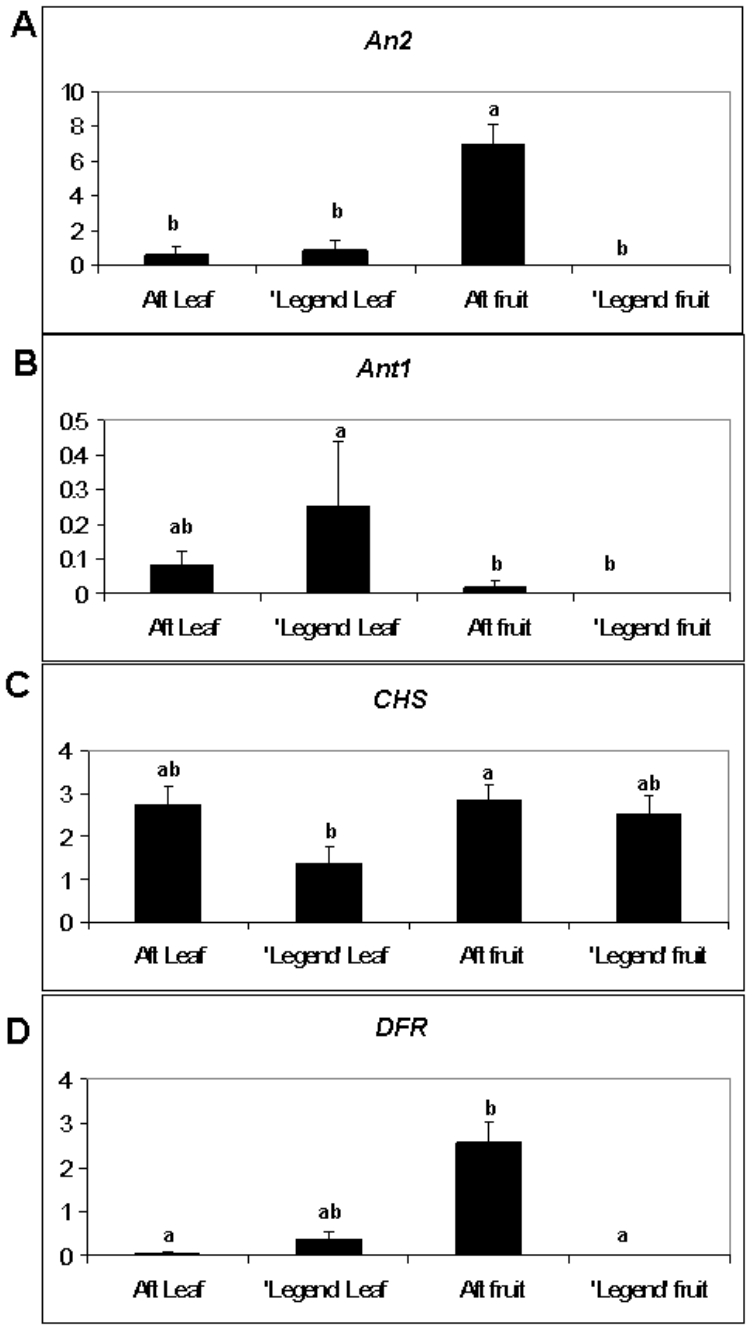


Figure 2.5. Semi-quantitative RT-PCR analysis of *Aft* (LA1996) and 'Legend tomatoes. Gene-specific primers for *An2* (A), *Ant1* (B), *CHS* (C), and *DFR* (D) were used to amplify each cDNA source (young leaf or mature green fruit total RNA). PCR products were separated on agarose gels, photographed and quantitated using Scion image analysis software. Expression of each gene was quantified as arbitrary units relative to *EF1*-α. Values represent the mean of 3 or more biological replicates, error bars = one standard error. Means not sharing a letter were significantly different using Fisher's Least Significant Difference ($P \leq 0.05$). Gene specific primers were 1F+R for *An2* and Ant1_594_1013F + AN1REV for *Ant1*, PCR was run for 32 cycles.

Expression levels were significantly different between treatments for *An2* and *DFR* but not *Ant1* or CHS. *An2* and *DFR* expression levels were significantly higher in *Aft* fruit (Figure 2.5A + D), while the levels of *Ant1* in leaf and fruit were not significantly different between genotypes (Figure 2.5B). *Ant1* levels were significantly higher in leaves than fruit (Figure 2.5B). For *An2*, the reverse was true (for *Aft* fruit only, levels of *An2* in 'Legend' fruit were low and not significantly different from leaves) (Figure 2.5A). *An2* expression was significantly correlated with DFR expression in both fruit and leaf tissues (R=0.75, p=.0003). *Ant1* expression, however, was not significantly correlated with either DFR expression (R=0.25, p=0.3192) or *An2* expression (R = 0.23, p=0.3546).

On cDNA templates, the Ant1_594_1013F+R primers sometimes produced two bands (~350 and 400 bp in size) (Fig. 2.4C). We tested the Ant1_593-1013F+R primers on a variety of cDNA and DNA templates (Fig 2.6), including the cTOA-12-C2 cDNA clone from which the *An2* primers were designed. The Ant1_594_1013F+R primers amplified the 350 bp product from the cTOA-12-C2 template as well as a 400 bp product from most DNA and cDNA templates (Fig. 2.6A). The Ant1_594_1013F+R primers also amplified the 350 bp products from some cDNA templates (Fig. 2.6A), especially those with high levels of *An2* expression (such as *Aft* fruit). We designed the AN1REV primer to be more specific to *Ant1* based on alignments of the *Ant1* and *An2* sequences. The Ant1_594_1013F + AN1REV primers failed to amplify any product from the cTOA-12-C2 template (Figure 2.6B). The AN1REV primer was used for quantitating expression levels of *Ant1* in Figure 2.5B.

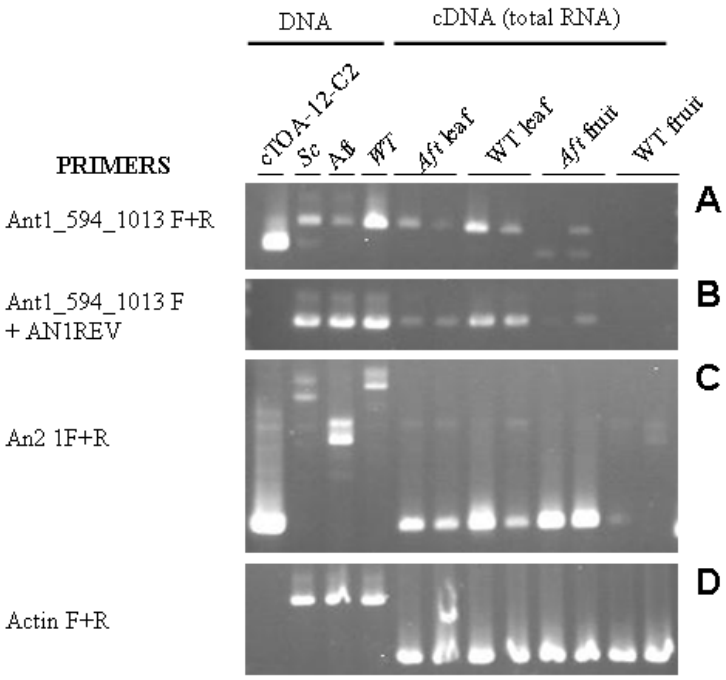


Figure 2.6. Gel image showing variable locus specificity of *Ant1* and *An2* primers in tomato and related species. A, B, and C: *Ant1* primers (Ant1_594-1012F+R and AN1REV), the *An2* primers (An2 1F+R) were used on DNA and cDNA templates and the PCR products were separated on a 2% agarose gel and stained with ethidium bromide. Template abbreviations: cTOA-12-C2, type cDNA clone of *An2* EST (SGN-U228086); *Sc, S. chilense* LA1963 (allele 2); *Aft*, LA1996; WT, 'Legend'. Actin primers (D) were used as a positive control. The distortion of the PCR product in the second *Aft* sample for D was caused by a bubble in the gel. Total RNA for cDNA templates of leaves was extracted from young (first fully expanded) leaves. Total RNA for cDNA templates of fruit was extracted from mature green fruit.

DNA Sequencing. All DNA sequences obtained in this study are listed in Table 2.3. In this article, numbers in parentheses refer to sequence numbers (column 1) in Table 2.3. The 3' sequence of the *An2* locus was obtained by 3'RACE. 3' RACE sequences with amino acid translations nearly identical to SGN-U228086 (*SlAn2*) were isolated from 3' RACE adapted cDNA of 'Legend' leaf, *Aft* fruit, and *Aft* leaf. The *Aft* 3'RACE sequence (18) was 895 bp long, the 'Legend' 3'RACE sequence (19) was 1620 bp. The two sequences (18 and 19) have high similarity for the first 200 nucleotides of the 3' UTR, but very little thereafter. The 'Legend' 3'UTR sequence

continues for an additional 726 nt, and contains a fragment of the 'Legend' *SlAn2* gene spanning part of exon 2, intron 2, and part of exon 3. We did not investigate whether this represented a PCR artifact or a transcribed pseudogene, but we did obtain other cDNA clones of this size with 3' RACE.

The 3' RACE sequence of *SlAn2* from 'Legend' was used to design a reverse primer (Table 2.2, *An2,* primer 2R) to amplify the *An2* gene. The *Aft* and 'Legend' *An2* sequences were then obtained by creating contigs (sequences 1 and 2) from two sets of overlapping *An2* primers (Table 2.2, 1F+R and 2F+R). The *An2* sequences for *Aft* and 'Legend' both contained uninterrupted reading frames with correct splicing. The first intron is in phase 1 (spliced between after the second nucleotide in the codon) and the second is phase 2 (spliced after the third nucleotide in the codon). The difference in size between the *Aft* and 'Legend' alleles of *An2* is due mainly to a 382 bp insertion in intron 2 of the Legend allele. An alignment of the predicted AN2 amino acid sequence for 'Legend' and *Aft* in comparison to unigene SGN-U228086 and sequences of ANT1 from 'Ailsa Craig' and LA1996 is shown in Figure 2.7. The *Sl*AN2 amino acid sequence of 'Legend' is identical to that of SGN-U228086 and VF36. The AN2 amino acid sequence of *Aft* contains four variant residues (134, 187, 258, and 260) (Figure 2.7) relative to 'Legend.

Table 2.4. *Ant1* and *An2* DNA sequences from tomato and related species. PCR products were sequenced directly (numbers 1-13,16 and17) or from cloned 3'RACE products (14 and 15). PCR products were generated using the primers indicated in Table 2.1 (*An2* and *Ant1*), or Table 2.2 (TG233).

No	GenBank	Gene	Species	Genotype	Description[1]
1	FJ705319	*SlAn2*	*S. lycopersicum*	cv. 'Legend'	5' clipped gene, gDNA
2	FJ705320	*ScAn2*	*S. lycopersicum*[2]	LA1996 (*Aft*)	5' clipped gene, gDNA
3	FJ705321	*SlAn2*	*S. lycopersicum*	LA3163 (*ag)*	5' and 3' clipped gene, gDNA
4	FJ705322	*SpAn2*	*S. lycopersicum*	Purple Smudge Goldman	5' and 3' clipped gene, gDNA
5	FJ705323	*SpAn2*	*S. pennellii*	LA1926	5' and 3' clipped gene, gDNA
6	FJ705324	*ScAn2*	*S. chilense*	LA1963	5' and 3' clipped gene, gDNA
7	FJ705325	*ScAn2*	*S. chilense*	LA1963	5' and 3' clipped gene, gDNA
8	FJ705326	*SlAnt1*	*S. lycopersicum*	cv. 'Legend'	5' and 3' clipped gene, gDNA
9	FJ705327	*Ant1*	*S. lycopersicum*	LA3668 (*Abg*)	5' and 3' clipped gene, gDNA
10	FJ705328	*SlAnt1*	*S. lycopersicum*	LA3736 (*atv*)	5' and 3' clipped gene, gDNA
11	FJ705329	*SpAnt1*	*S. lycopersicum*	LA2378 (*Purple Smudge)*	5' and 3' clipped gene, gDNA
12	FJ705330	*SpAnt1*	*S. lycopersicum*	PI290858 (*Purple Smudge*)	5' and 3' clipped gene, gDNA
13	FJ705331	*SpAnt1*	*S. lycopersicum*	Purple Smudge Orange Flesh	5' and 3' clipped gene, gDNA
14	FJ705332	*SlAnt1* like	*S. lycopersicum*	cv. 'Legend'	3'RACE (leaf), mRNA
15	FJ705333	*ScAn2* like	*S. lycopersicum*	LA1996 (*Aft*)	3'RACE (fruit), mRNA
16	FJ705334	TG233	*S. lycopersicum*	LA1996 (*Aft*)	RFLP locus, gDNA
17	FJ705335	TG233	*S. lycopersicum*	cv. 'Legend'	RFLP locus, gDNA
18	FJ744760	*ScAn2*	*S. lycopersicum*	LA1996 (*Aft*)	3'RACE (fruit), mRNA
19	FJ744761	*SlAn2*	*S. lycopersicum*	cv. 'Legend'	3'RACE (leaf), mRNA

[1]Clipped gene indicates a partial gene sequence, gDNA=origin from genomic DNA, mRNA = origin from 3' RACE adapted cDNA. [2]Sequences from LA1996 and *Purple Smudge* genotypes probably represent alleles from *S. chilense* or *S. peruvianum*, respectively.

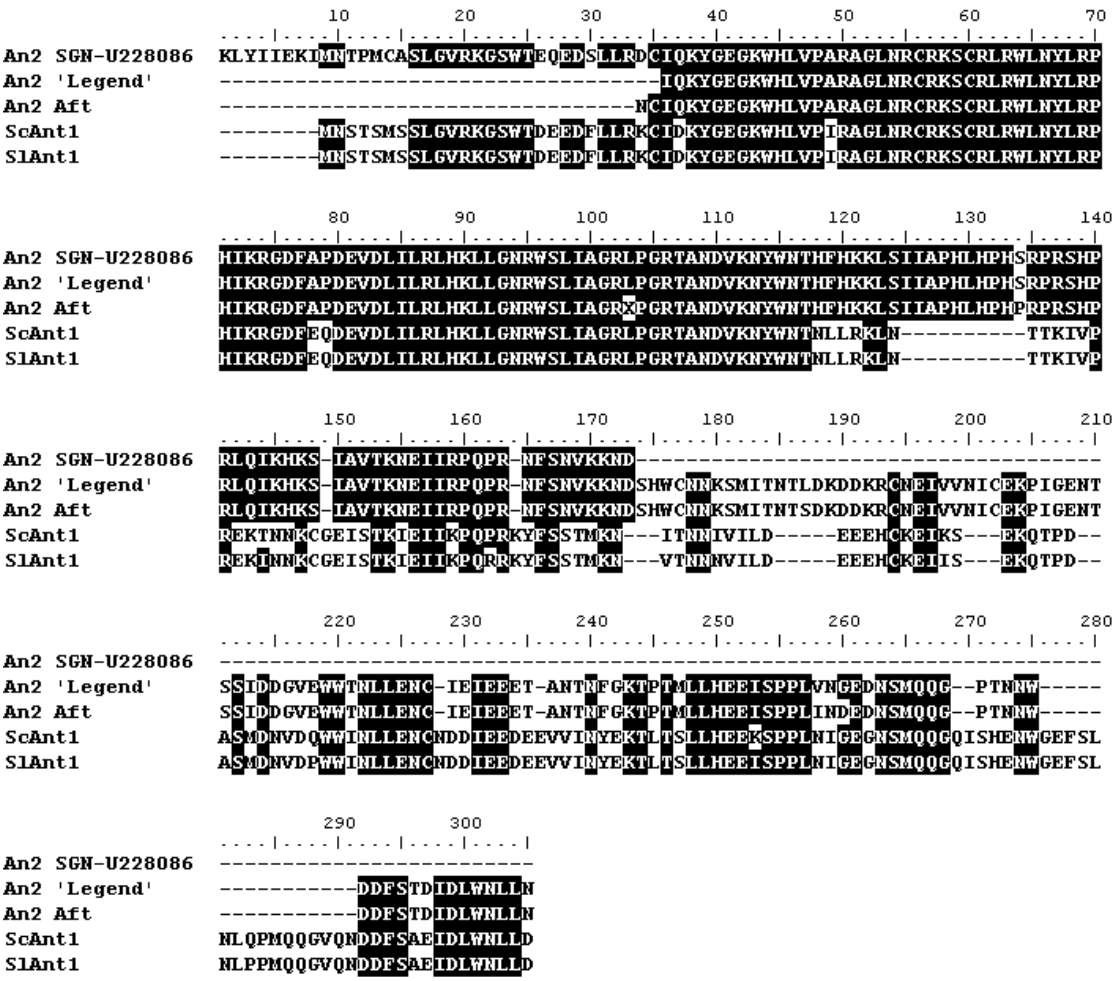


Figure 2.7. Amino acid sequence alignment of *Ant1* and *An2* from cultivated tomato and LA1996 (*Aft*). Amino acid sequences were predicted from genomic DNA sequence and aligned (using CLUSTALW) with the predicted amino acid sequence of SolGenes unigene SGN-U228086, *ScAnt1* from LA1996 (EF433417), and *SlAnt1* from 'Ailsa Craig' (EF433416). Identical amino acids are shaded in black.

The *An2* gene is distinct from the *Ant1* gene (Figure 2.7). The R2R3 domains of the two proteins are well conserved, but the C-terminal portions of the two proteins are highly divergent. The amino acid identity between ANT1 and AN2 sequence alignments in *Aft* was 55.5% overall, 88.6 % for the R2R3 domain, and 39.0% for the C-terminal domain.

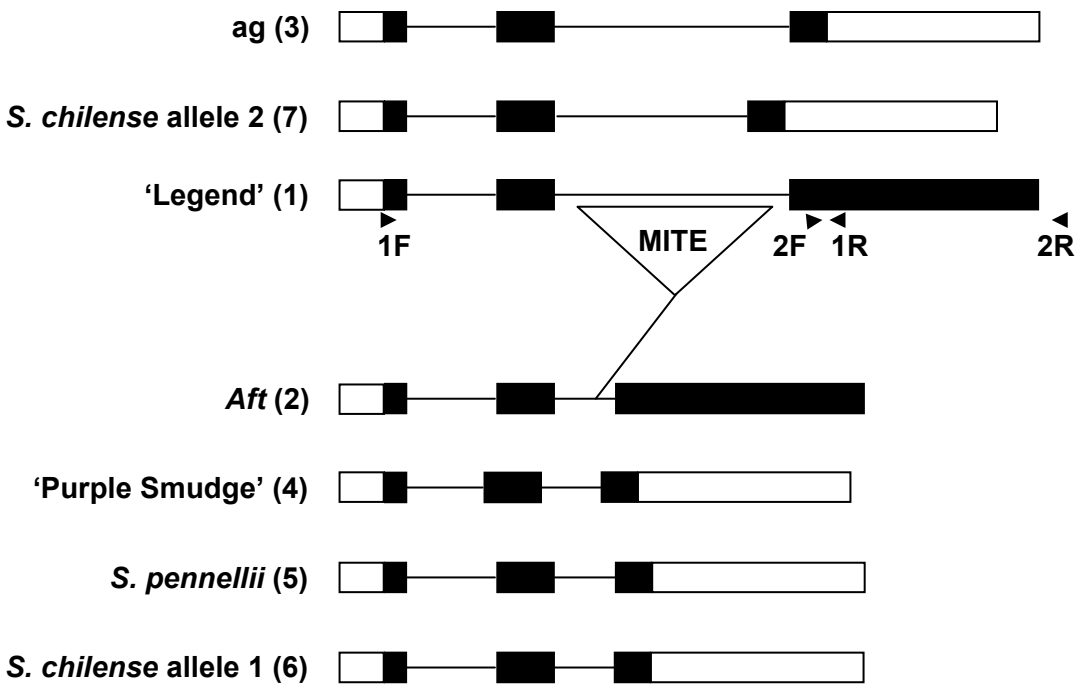


Figure 2.8. Structure of the *An2* MYB gene in tomato and related species deduced from partial genomic sequences. Exons are indicated by boxes, introns are indicated by lines. Sequenced portions of the exons are indicated in black, white boxes indicate an exon predicted from reference sequences. Numbers in parentheses after a sequence name refer to sequence numbers in Table 2.4 (column 1). The positions of the primers used for sequencing are indicated by small triangles next to the 'Legend' sequence. The putative minature inverted-repeat transposable element (MITE) found in intron 2 of the 'Legend' allele is indicated by a triangle.

A portion of the *An2* gene spanning exons 1, 2 and 3 was sequenced (using the *An2* primers 1F+R) in LA3163 (3), 'Purple Smudge Goldman' (4), LA1926 (5), and LA 1963 (6 and 7). A schematic representation of the *An2* gene in all genotypes is shown in Figure 2.8. *SlAn2* alleles from LA3163 (*ag*) and 'Legend' differed at 3 nucleotide positions, none of which were predicted to alter the amino acid sequence of the protein. Both 'Legend' and LA3163 alleles shared the large (~460 bp) insertion located in intron 2. The large allele of *ScAn2* from LA1963 (7) also contained a large insertion located in intron 2. The inserted sequence in the LA1963 allele was different from that in 'Legend' and LA3163 (49.4 % sequence identity). The sequence insertions in intron 2 show characteristics of a MITE (see below). Alleles of *An2* from LA1996 (2) and 'Purple Smudge Amy Goldman' (4) were similar (90.6 % nucleotide identity). The LA1996 allele contained two small (~10 bp) insertions relative to

'Purple Smudge' (Amy Goldman) located in intron 1. The *ScAn2* allele from LA1996 (1) and the small allele from *S. chilense* LA1963 (6) were almost identical (98.4% nucleotide identity). The *SpAn2* allele from *S. pennelli* LA1963 (5) was also very similar to that of LA1996 (95.8% nucleotide identity).

The inserted sequence in Legend has following characteristics of a MITE: it is flanked on either side by a 12 bp direct repeat ('AATTATTTCATT'), contains terminal inverted repeats, and has an AT rich nucleotide composition (72% AT). The sequence found in the 'Legend' intron was BLASTED against the GenBank nucleotide database and 17 sequences with e-values $\geq 1e^{-68}$ and maximum similarity values ≥ were found in the tomato genome. BLAST hits were found in untranscribed regions near a TAPG2 polygalacturonase gene (GenBank accession no. AF001001), an expansin gene (AF311953.1) and a CBF transcription factor (AY497899).

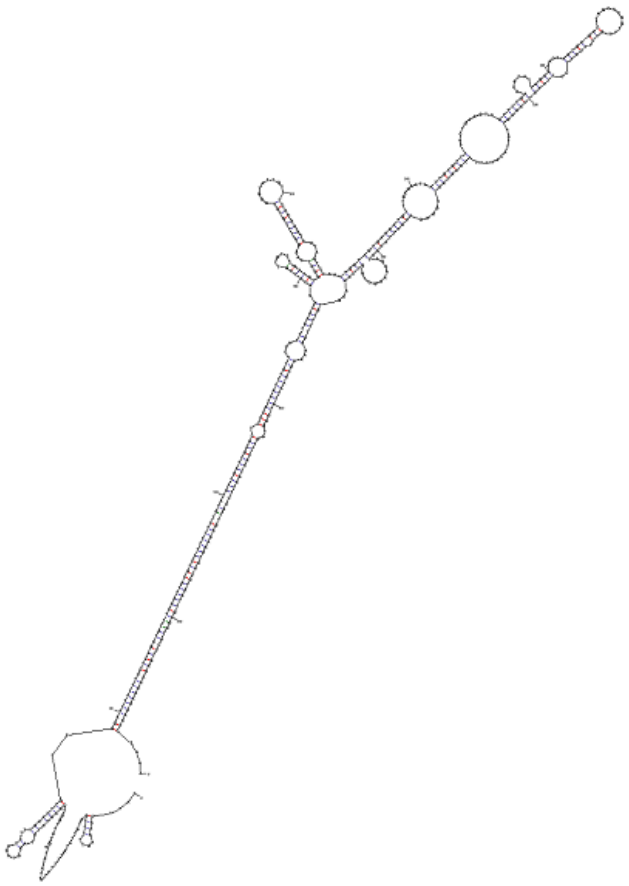


Figure 2.9. DNA folding structure prediction for the MITE sequence in 'Legend'. The sequence shows hallmarks of a MITE, including terminal inverted repeats, visible as the fold-back structure at the 5' and 3' ends of the sequence. The UNAFOLD software was used to predict DNA folding.

DNA sequence folding structure predictions (Figure 2.9) for the portion of the sequence flanked by the 12 bp direct repeats indicated that the sequence was capable of forming hairpin structures with a high free energy (ΔG=69.54 Kcal/molecule). The LA1963 MITE was flanked by a similar direct repeat (AATTATTTTATC), was predicted to fold into a hairpin structure (predicted free energy of folding: ΔG= -78.1 Kcal/molecule), with five BLAST hits in the tomato genome at an e-value $\geq 3e^{-68}$. The number one BLAST hit for the LA1963 MITE was located in between the NBS-LRR gene *Mi-1A* and the NBS-LRR pseudogene *Mi-1B* (DQ863289).

Portions of the *Ant1* gene were also sequenced in several genotypes (sequences 7-13, Table 2.2). The R2R3 portion of the *Ant1* gene from LA3668 (*Abg*) (9) was sequenced using the LeANTF+R primers and was identical to the corresponding sequence from 'Legend' (8) or 'Ailsa Craig' (not shown). The C-terminal domain of *SpAnt1* was sequenced in several 'Purple Smudge' genotypes, including LA2378 (11), PI290858 (12), and 'Purple Smudge Orange Flesh' (13) using the Ant1_594_1013F+R primers. The *Ant1* sequence from PI290858 was identical to 'Ailsa Craig'. The corresponding sequence from LA2378 and 'Purple Smudge Orange Flesh' differed from 'Ailsa Craig' at several nucleotides (two in LA2378 and four in 'Purple Smudge Orange Flesh'). None were predicted to alter the amino acid sequence of the ANT1 protein. The *SlAnt1* gene was sequenced in LA3736 (*atv*) using both the LeANTF+R and the Ant1_594_1013F+R primers. The resulting sequence (10) was identical to 'Ailsa Craig'.

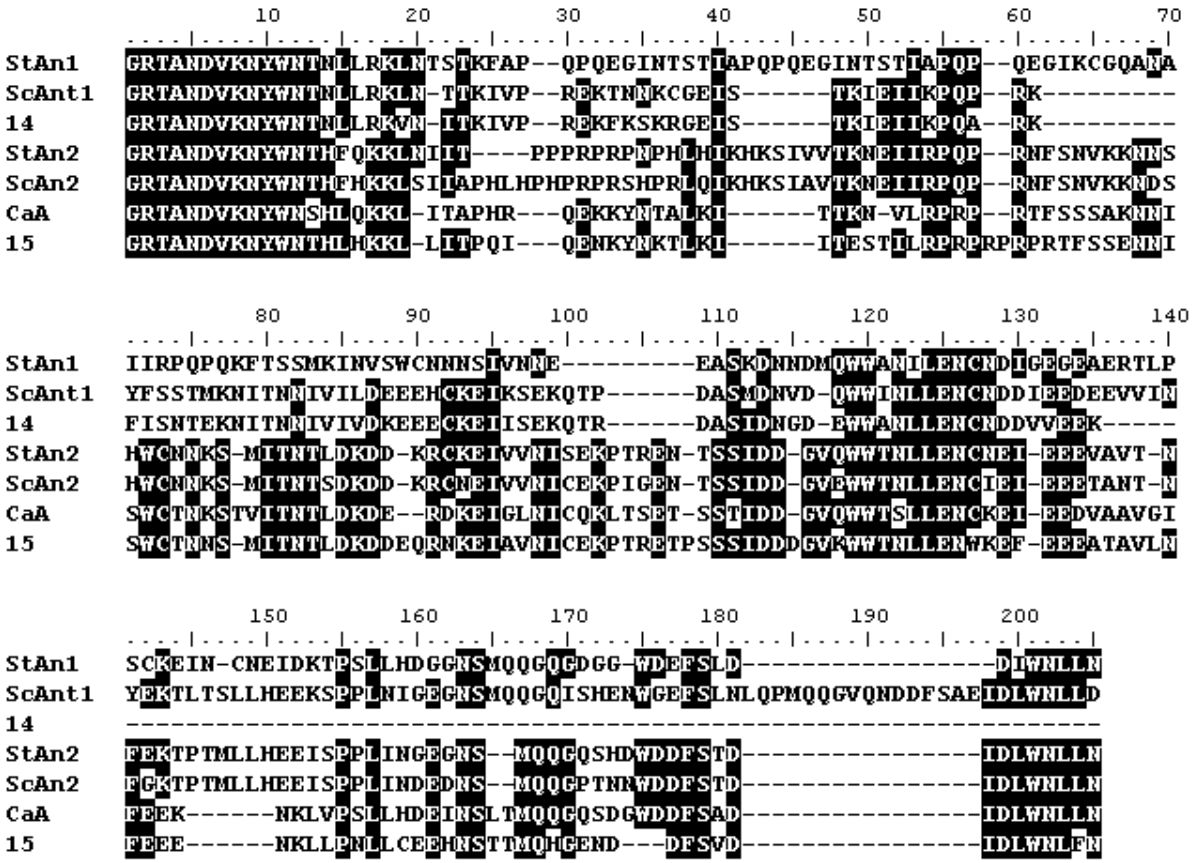


Figure 2.10. ClustalW alignment of predicted amino acid sequences from DNA sequences 14 and 15, *Ant1* and *SlAn2* like genes isolated using 3'RACE. Sequence 14 (isolated from 'Legend' leaf cDNA) is most similar to *Sl*ANT1 but truncated. Sequence 15 (isolated from *Aft* fruit cDNA) is more similar to pepper A than *SlAnt1* or *SlAn2*. Abbreviations (GenBank accession numbers in parentheses): StAn1, *Solanum tuberosum An1* (AAX53092); *ScAnt1*, *Solanum chilense Ant1*, (EF433417); *StAn2*, *Solanum tuberosum AN2* (AAX53092), *ScAn2*, *Solanum chilense* (LA1996) *An2* (FJ705320); *CaA, Capsicum annuum A*, (CAE75745).

Additional *SlAnt1* and *SlAn2* like sequences (14 and 15) were found using 3'RACE. Sequence 14, isolated from 3' RACE adapted cDNA of 'Legend' leaves, is similar to a truncated *Sl*ANT1 (Figure 2.10). Sequence 14 is 76.8% identical to SlANT1 (excluding the portion of the *Sl*ANT1 sequence that is truncated in sequence 14). Sequence 15 is *Sl*AN2 like, but actually bears more sequence similarity to the pepper *A* MYB (64.0%) than to the *Sl*AN2 MYB (54.0%) (Figure 2.10).

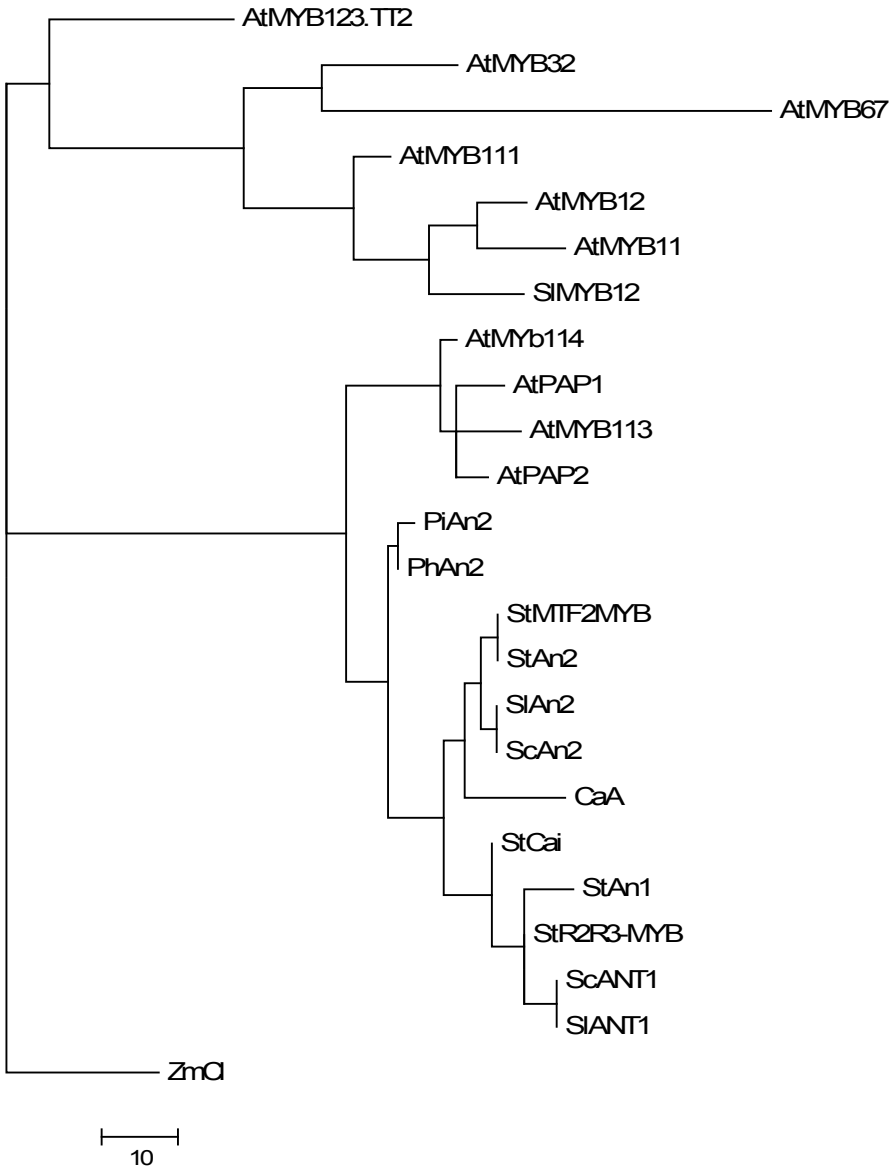


Figure 2.11. Maximum likelihood tree based on amino acid sequence alignment of R2R3 domains from flavonoid regulatory MYB proteins. The following amino sequences were used: ZmCl (P10290), AtMYB123.TT2 (Q9FJA2), AtMYB32(NP_195225), AtMYB67(NP_566434), AtMYB111(NP_199744), AtMYB12(NP_182268), AtMYB11(NP_191820), SlMYB12(ACB46530), AtMYB114(NP_176812), AtPAP1(NP_176057), AtMYB113(NP_176811), AtPAP2(NP_176813), PiAn2(ABO21073), PhAn2(ABO21074), StMTF2MYB(ABY40371), StAn2(AAX53093), SlAn2(FJ705319), ScAn2(FJ705320), CaA(CAE75745), StCai(ABY40370), StAn1(AAX53092), StR2R3-MYB(ABK29436), ScANT1(ABO26065), SlANT1(ABO26064). Species abbreviations: Zm, *Zea mays*; At, *Arabidopsis thaliana*; Pi, *Petunia integrifolia*; Ph, *Petunia hybrida*; St, *Solanum tuberosum*; Sl, *Solanum lycopersicum*; Sc, *Solanum chilense*; Ca, *Capsicum annuum*.

A maximum likelihood tree based on amino acid alignments from the R2R3 domain of AN2 from 'Legend' and *Aft* and other flavonoid regulating MYBs shows that *Sl*AN2 has more similarity to pepper *A* and *St*AN2 whereas *Sl*ANT1 is more similar to *St*AN1 (Figure 2.11). Both *Sl*ANT1 and *Sl*AN2 are more closely related to the *Arabidopsis* PAP1/PAP2 gene family than the *Arabidopsis* MYB11/12/111 gene family (Figure 2.11). A similar tree based on amino acid alignments from the C-terminal domain (Figure 2.12) that included the 3'RACE sequences 14 and 15 shows that 14 is very similar to *Sl*ANT1, whereas 15 has similarity to the entire *Solanum* ANT1/AN2 gene family and especially pepper *A*.

Phenotypic characterization of tomato fruit pigmentation patterns. The distribution and morphology of anthocyanin containing cells was variable among anthocyanin fruited tomatoes. Since some of these appear to be alleles of the same gene, we characterized their phenotypes in more detail. We examined the *S. pennellii* introgression line IL 10-3 for the presence of anthocyanin cells since it contains the chromosomal segment from *S. pennellii* containing *Aft*. We observed small anthocyanin containing cells in IL10-3 (Figure 2.12, A and B), though to our knowledge this has not yet been reported. Anthocyanin containing cells observed in IL10-3 were small, spherical cells located near the epidermis in the 1st or 2nd layer of pericarp cells. LA1996 (*Aft*) anthocyanin cells are located in the first or second layer of pericarp cells as well (Figure 2.13F) but were larger. *Aft* anthocyanin cells located in the 1st layer of pericarp cells were often flattened as well. 'Purple Smudge' (PI290858) anthocyanin cells were similar to *Aft,* being located in the first or second layer of the pericarp (Figure 2.13B and D). In early maturing fruit they tended to be small but in late maturing fruit large, flat anthocyanin cells were observed in 'Purple Smudge' fruit. LA3668 (*Abg*) anthocyanin cells, in contrast, were located several cell layers deeper (4-6 cells deep) (Figure 2.13E). We also observed small numbers of anthocyanin cells in LA3736 (*atv*) fruit. Anthocyanin cells in *atv* fruit were intermediate between *Abg* and *Aft*. Anthocyanin cells in *atv* were located approximately 3 cell layers under the epidermis (not shown).

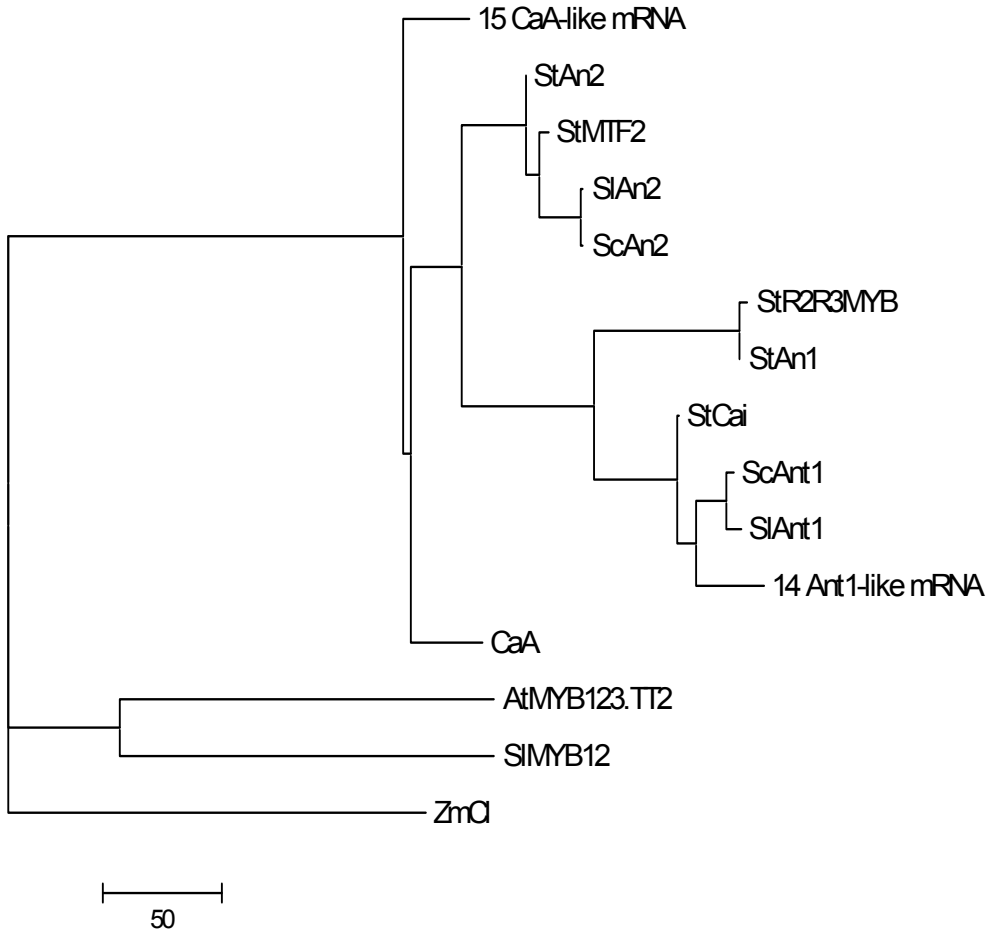


Figure 2.12. Maximum likelihood tree based on amino acid sequence alignment of C-terminal domains from flavonoid regulatory MYB proteins with 3' RACE sequences. The following amino sequences were used: ZmCl (P10290), AtMYB123.TT2 (Q9FJA2), SlMYB12 (ACB46530), CaA (CAE75745), 15 CaA-like mRNA (FJ705333), StMTF2MYB (ABY40371), StAn2 (AAX53093), SlAn2 (FJ705319), ScAn2 (FJ705320), CaA (CAE75745), StCai (ABY40370), StAn1 (AAX53092), StR2R3-MYB (ABK29436), ScANT1 (ABO26065), SlANT1 (ABO26064), 15Ant1-like mRNA (FJ705332). Species abbreviation: Zm, *Zea mays*; At, *Arabidopsis thaliana*; St, *Solanum tuberosum*; Sl, *Solanum lycopersicum*; Sc, *Solanum chilense*; Ca, *Capsicum annuum*.

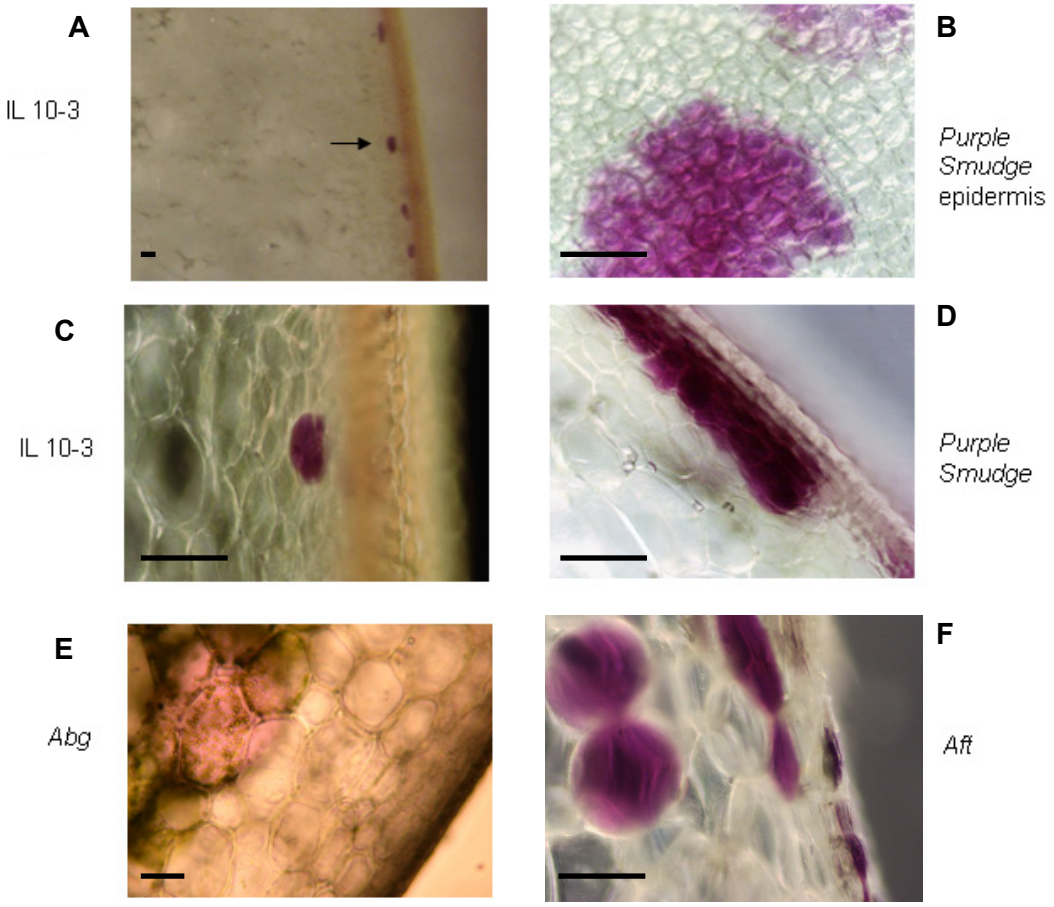


Figure 2.13. Morphology and distribution of anthocyanin cells in tomato fruit. (A) Cross section of breaker fruit from IL 10-3. Arrow indicates an anthocyanin cell shown in detail in (C). (B) Epidermis of 'Purple Smudge' (PI290858) unripe fruit, viewed head on. (C) Close-up of a cross section of IL 10-3 fruit in (A), showing location of an anthocyanin cell relative to cuticle. (D) Cross section of 'Purple Smudge' (PI290858) unripe fruit, showing size and location of anthocyanin cells relative to cuticle. (E) Cross section of *Abg* (LA3668) unripe fruit. (F) Cross section of *Aft* (LA1996) fruit. Scale bar is approximately 50 µm.

Discussion

Associations between MYB *genes and anthocyanin mutants in tomato.* Our PCR and sequencing results as well as those of Sapir et al. (2008) indicate that *Aft*, *Purple Smudge*, *S. pennellii*, and *S. chilense* genotypes have different alleles of *Ant1* and *An2*. PCR results indicated that *Abg* had the same allele of *An2* as tomato cultivars. This was surprising because *Abg* also maps to chromosome 10, and originates from *S. lycopersicoides*, which is more distantly related to tomato than *S. pennellii* or *S. chilense*. We did not confirm this result with sequencing, but the partial sequence of

Ant1 for *Abg* was also identical to that of cultivated tomato, unlike the corresponding sequence in *S. chilense*. Because *Abg* cannot be maintained in a homozygous state, it is possible that our PCR primers are only amplifying the *S. lycopersicum* allele for the *Ant1* and *An2* loci. This question could be resolved by using these primers on DNA of wild *S. lycopersicum* accessions. The identical result was also obtained for *ag*, which maps very close to *An2*. However, since *ag* occurred as a natural mutation in cultivated tomato, much less polymorphism would be expected.

Our semi-quantitative RT-PCR results indicate that the expression of the *An2* transcript is closely correlated with the presence of anthocyanin in tomato fruit and *DFR* transcript levels in tomato fruit and leaves. The expression pattern of *Ant1* was more cryptic: it was most highly expressed in leaf or fruit tissues with high amounts of anthocyanin, but it was expressed at low levels even in tissues with no detectable anthocyanin or *DFR* expression. *Ant1* was more consistently expressed in young leaf tissue than any other tissue. Sapir et al. (2008) reported very low levels of *Ant1* transcripts in fruit of both 'Moneymaker', LA1996, and their F_1 hybrid. They reported no significant difference in the transcript levels between the genotypes. This led them to suggest that an amino acid polymorphism between *S. chilense* and *S. lycopersicum* in *Ant1* was responsible for the difference in fruit phenotypes between LA1996 and cultivated tomato. The *SlAnt1* gene was originally discovered by activation tagging, which can activate genes that are not normally transcribed (Mathews et al. 2003). In originally describing *SlAnt1*, Mathews et al. (2003) detected no expression of the *SlAnt1* gene in wild type ('Micro-Tom') plants using RT-PCR. Also, the phenotypes of transgenic *SlAnt1* overexpressing plants always included elevated anthocyanin in the leaves and hypocotyls, but only expressed anthocyanin in the fruit in some (extreme) cases.

Given our results, we feel that *An2* is a more likely candidate for *Aft* and *Purple Smudge* than *Ant1*, but confirmation of this will require tests using functional genetics approaches. Linkage data implies that *ag* is a mutated allele of an *SlAnt1* or *SlAn2* MYB homolog with an expression domain that includes hypocotyls and the upper leaf surface. The possibility that *ag* rather than *Aft* corresponds to *SlAnt1* warrants further investigation. It is worth mentioning that several other classical

anthocyanin mutants which are differentially expressed in the cotyledons and primary leaves (*al-2*), veins (*vio, vio-2*), or stems (*al, ai*) may correspond to a member of the *Ant1*/*An2* gene family. Studying these genes is complicated by the existence of at least two, potentially several, tightly linked homologs with overlapping functions. The sequencing of chromosome 10 will hopefully identify the complete structure of the locus. Ectopic expression of *An2* and related sequences in normal tomatoes would provide valuable information as to the target genes and metabolites for *An2*, but probably won't prove whether *Ant1* or *An2* is the sequence encoding *Aft*, *ag*, or *Purple Smudge*, since ectopic expression of MYBs from other species results in high anthocyanin production due to functional conservation (Butelli et al., 2008). Silencing the various members of the *Ant1* and *An2* family would be an alternative technique. Using virus induced gene silencing (VIGS) allowed Jablonska et al. (2006) to distinguish among multiple *Mi* homologs when faced with a similar problem, since VIGS uses very short (23 nt.) sequences to silence genes with a high degree of specificity.

Presence of an Ant1/An2 *MYB gene family in tomato*. We report here the existence of a small gene family that represents the tomato homologs of *StAn1* and *StAn2*. *SlAnt1* has been previously characterized (Mathews et al. 2003; Sapir et al., 2008), but *SlAn2* has not. The functional specialization of *SlAnt1* and *SlAn2* apparently traces back at least as far as the last common ancestor of tomato and potato. We also describe an additional two tomato mRNA sequences with similarity to *SlAnt1* and *SlAn2* that may map to the same region, including a *SlAn2*-like gene that is most similar to pepper 'A'; and a tomato *SlAnt1*-like gene with a truncated C-terminal domain.

Maximum likelihood analysis of the R2R3 domain from members of the *Ant1* and *An2* gene family indicated that *SlAn2* and *SlAnt1* are more similar to the *Arabidopsis* PAP1/PAP2 gene family than the *Arabidopsis* MYB11/12/111 gene family (Figure 2.11). PAP1 (*Production of Anthocyanin Pigment 1*) over-expressing plants are characterized by upregulation of parts of the phenylpropanoid pathway leading specifically to the production of cyanidin derivatives but not other flavonoids (e.g. quercetin and kaempferol) (Tohge et al., 2005). In contrast, the *At*MYB11/12/111

family has been demonstrated to specifically upregulate the production of flavonoid glycosides in a tissue specific, additive fashion (Stracke et al. 2007). A tomato homolog of *At*MYB12 (*Sl*MYB12*)* was recently discovered (Luo et al., 2008), and grouped with the AtMYB11/12/111 family in our analysis as well (Figure 2.11).

Although PAP1, PAP2, *At*MYB113, and *At*MYB114 have all been demonstrated to cause anthocyanin production and upregulate structural genes in the phenylpropanoid pathway (Gonzalez et al. 2008), the tissue specificities and distinction between the four genes is not yet known. Members of the *At*MYB11/12/111 family are differentially expressed in various tissues and lead to different flavonol accumulation patterns (Stracke et al., 2007). The final flavonol accumulation pattern is determined by the combination of *At*MYB11/12/111 expression (Stracke et al. 2007). Triple mutants lacking all three members of this MYB subfamily (*myb11 myb12 myb111*) produce no flavonols at all, while anthocyanin expression is unaffected. We propose that the members of the *Ant1*/*An2* gene family in the Solanaceae act in a similar fashion, each member being expressed in a tissue specific fashion and acting additively to produce different anthocyanin profiles in various tissues.

The existence of a second MYB gene near the *SlAnt1* locus was not unexpected. Sapir et al. (2008) recently reported linkage between *Aft* and *ScAnt1* as a single locus. However, De Jong et al. (2004) reported two closely linked *PhAn2* homologs in tomato separated by one recombination event in a population of 83 individuals. We observed complete linkage between the *SlAn2* and *SlAnt1* loci in tomato, indicating that they represent the two *SlAn2* homologs observed by DeJong et al. (2004). *StAn1* and *StAn2* are tightly linked in potato as well (W.S. De Jong, pers. comm.). The presence of additional RNA species similar to these loci may indicate the presence of two gene clusters, as observed for resistance gene loci in tomato such as *Cf* (resistance to *Cladosporium fulvans*) and *Mi* (resistance to *Meloidogyne incognita*) (Jablonska et al., 2006).

Numerous examples of duplicated MYB genes exist. In *Arabidopsis*, all four genes in the PAP1/PAP2 family are located on *Arabidopsis* chromosome 1. PAP2, *At*MYB113, and *At*MYB114 are very closely linked in head to tail fashion over a ~15

Kb stretch of DNA (TAIR website). In all likelihood, PAP1 and the PAP2 gene cluster represent one round of gene duplication, and the three genes in the PAP2 cluster represent a more recent round of gene duplications. Indeed, tightly linked clusters of MYB orthologs regulating anthocyanin biosynthesis have been reported in maize (Dias et al., 2003), sorghum (Boddu et al., 2006), grape (Matus et al., 2008), and snapdragon (Schwinn et al., 2006). An amplification of a variety of MYB gene families that predates the monocot-dicot split has also been documented (Rabinowicz et al., 1999).

The regulation of flavonoid biosynthesis is a model system for studying complex loci across species (Koes et al., 2005). The function of the WD40-bHLH-MYB complex regulating anthocyanin biosythesis is conserved between monocots and dicots (Quattrochio et al., 1993). The regulatory specificity of the complex is determined by a combination of: different associations between the bHLH and MYB partners, promoter sequences of target genes, and regulation of the MYB and/or bHLH proteins. Alternative splicing (Grotewold et al., 1991; Li et al., 2006), upstream enhancers (Zhang and Peterson, 2005), and epigenetic phenomena (Cocciolone et al., 2001) have been implicated in regulation of the MYBs responsible for regulating flavonoid biosynthesis. MYB genes regulating other biological processes such as leaf development and light response have been shown to be regulated by RNA silencing (Ori et al., 2007) and control of protein turnover (Hong et al., 2009), respectively. Paralogous gene recombination (Zhang and Peterson, 2005), segmental gene duplication (Zhang et al., 2000), and transposon induced homologous recombination (Athma and Peterson, 1991) have been reported to create alleles with altered tissue specificity in maize. The tissue specificity of MYB proteins has also been shown to depend on the C-terminal domain of the MYB protein interactions (Chopra et al., 1996). The ability of a MYB protein to activate transcription has also been shown to depend on interactions with the bHLH partner (Hernandez et al., 2004), which is due to mainly the structure of the R3 domain (Zimmerman, 2004; Hernandez et al., 2004).

In trichome and root hair cell patterning, the activity of the WD40-bHLH-MYB complex is regulated by competition between single repeat R3 MYBs, which do not function as transcriptional activators, and R2R3 MYBs (Zhao et al., 2008). The

R2R3 MYBs have been shown to directly activate the R3 MYBs, which can move to adjacent cells while the R2R3 MYB cannot (Zhao et al., 2008). The specification of cell fate is determined more by a threshold level of the activator (R2R3 MYB) relative to the repressor (R3 MYB) rather than the presence or absence of the activator, with the result that cells designated to be trichomes or root hairs laterally inhibit this cell fate in neighboring cells. A single repeat MYB inhibitor of anthocyanin biosynthesis, *At*MYBL2, was recently described in *Arabidopsis* (Matsui et al., 2008; Dubos et al., 2008). This model helps explain the distribution of anthocyanin pigments seen in tomato fruit (Figure 2.12), which form in isolated cells in response to light. If *ScAn2* corresponds to *Aft*, this model explains why *An2* expression is strongly correlated with *DFR* expression but also found at low levels in non-anthocyanin expressing tissues. Presumably, *An2* only activates downstream genes such as *DFR* when a critical threshold concentration of the protein has been reached.

Inactivation of MYB transcription factors by transposons has played a significant role in the evolution of different colored grape cultivars (Lijavetzky et al., 2006). In *Petunia*, a white flowered sub-species of *P. axillaris* has apparently arisen multiple times as a result of the loss of function of *PaAn2* due to the insertion and excision of a transposable element (Quattrocchio et al., 1999). Here, we report independent insertions of MITE-like sequences into the same location in intron 2 of the *An2* in cultivated tomato and *S. chilense* LA1963. In maize, a 743 bp fragment flanked by a 10 bp direct repeat was found in intron 2 of the *P1-rr* and *P1-wr* alleles but not in *P1-rw* and *p2* alleles (Jiang et al., 2004).

MITEs are small Transposable Elements (TEs) that have been documented in plants, vertebrates, insects and nematodes. MITEs are characterized by TIRs, small size, no coding potential, AT richness, and the potential to form stable secondary structures (Wessler et al., 1995). Structurally, MITEs resemble degraded Class II TEs that transpose directly as DNA and are present in low copy number throughout the genome. However, the occurrence of identical MITE sequences throughout plant genomes (Menzel et al., 2006) suggest that they are amplified by an as-yet unidentified mechanism, similar to type I TEs that transpose via an RNA intermediate (Cascuberta and Santiago, 2003). MITES have been implied as having widespread

influence in the evolution of plant genomes (Cascuberta and Santiago, 2003; Wessler et al., 1995). We did not find any direct evidence that the MITE found in *SlAn2* has disrupted the function of the gene, but hairpin structures in introns have been shown to be a source of a novel class of small interfering RNAs in animals (Okamura et al., 2007) and probably plants as well (Zhu, 2008).

Tomato is a model species for the study of fruit ripening as well as ecological and evolutionary genetics (Moyle, 2008), but MYB genes in tomato have been studied less than in related species such as petunia and pepper because cultivated tomatoes do not normally produce anthocyanin in their fruit. Further study of the *SlAnt1* and *SlAn2* gene family could provide important information about these genes in tomato and other Solanaceous species, as well as a model system for the study of complex loci.

References

Accotto, G.P., J. Navas-Castillo, E. Norris, E. Moriones, D. Louro. 2000 Typing of tomato leaf curl viruses in Europe. European J. Plant Pathol. 106:179-186.

Athma, P. and T. Peterson. 1991. *Ac* induces homologous recombination at the maize *P* locus. Genetics 128:163-173.

Baudry, A., M. Caboche, and L. Lepiniec. 2006. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J. 46:768-779.

Boddu, J., C. Jiang, V. Sangar, T. Olson, T. Peterson, and S. Chopra. 2006. Comparative structural and functional characterization of sorghum and maize duplications containing orthologous Myb transcription regulators of 3-deoxyflavonoid biosynthesis. Plant Mol. Biol. 60:185-199.

Borovsky, Y., M. Oren-Shamir, R. Ovadia, W. De Jong, and I. Paran. 2004. The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to *Anthocyanin2* of *Petunia.* Theor. Appl. Genet. 109:23-29.

Broun, P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. Curr. Opin. Plant Biol. 8:272-279.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

Chopra, S., P. Athma, and T. Peterson. 1996. Alleles of the maize *P* gene with distinct tissue specificities encode Myb-homologous proteins with C-terminal replacements. Plant Cell 8:1149-1158.

Cocciolone, S.M., S. Chopra, S.A. Flint-Garcia, M.D. McMullen, and T. Peterson. 2001. Tissue-specific patterns of a maize *Myb* transcription factor are epigenetically regulated. Plant J. 27:467-478.

Cascuberta, J.M., and N. Santiago. 2004. Plant LTR-retrotransposons and MITES: control of transposition and impact on the evolution of plant genes and genomes. Gene. 311:1-11.

De Jong, W.S., N.T. Eannatta, D.M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor. Appl. Genet. 108:423-432.

Dias, A.P., E.L. Braun, M.D. McMullen, and E. Grotewold. 2003. Recently duplicated maize R2R3 genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. Plant Phys. 131:610-620.

Dubos, C., J. Le Gourrierec, A. Baudry, G. Huep, E. Lanet, I. Debeaujon, J.-M.. Routaboul, A. Alboresi, B. Weisshar, and L. Lepinec. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. Plant J. 55: 940-953.

Economic Research Service, USDA. 2008. Vegetables and melons situation and outlook yearbook. VGS-2008. 30 October 2008.
<http://www.ers.usda.gov/Publications/VGS/2008/05May/VGS2008.pdf.>.

Fulton T.M., J. Chugwongse, S.D. and Tanksley. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rptr. 13:207-209.

Genbank Website. 13 January 2009. <www.ncbi.gov>.

Gonzalez, A., M. Zhao, J.M. Leavitt, A.M. Lloyd. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J. 53:814-827.

Grotewold, E., P. Athma, and T. Peterson. 1991. Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA-binding domain of *myb*-like transcription factors. Proc. Natl. Acad. Sci. USA 88:4587-4591.

Grotewold E., M.B. Sainz, L. Tagliani, J.M. Hernandez, B. Bowen, V.L. Chandler. 2000. Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proc. Nat. Acad. Sci. 97:13579-13584.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Hernandez, J.M., G.F. Heines, N.G. Irani, A. Feller, M-G. Kim, T. Matulnik, V.L. Chandler, and E. Grotewold. 2004. Different mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor C1*. J. Biotechnol. 12:48205-48213.

Hong, S.H., H.J. Kim, J.S. Ryu, H. Choi, S. Jeong, J. Shin, G. Choi, and H.G. Nam. 2008. CRY1 inhibits COP1-mediated degradation of BIT1, a MYB transcription factor, to activate blue light-dependent gene expression in *Arabidopsis*. Plant J. 55:361-71.

Jablonska, B., J.S.S. Amiraju, K.K. Bhattarai, S. Mantelin, O.M de Ilarduya, P.A. Roberts, and I. Kaloshian. 2006. The *Mi-9* gene from *Solanum arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of *Mi-1*. Plant Physiol. 143:1044-1054.

Ji, Y. and R.T. Chetelat. 2003. Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic and substitution lines in tomato. Theor. Appl. Genet. 106:979-989.

Jiang C., X. Gu, and T. Peterson. 2004. Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. Genome Biol. 5:R46.

Jones, C.M., P. Mes, and J.R. Myers. 2003. Characterization and inheritance of the *Anthocyanin fruit* (*Aft*) tomato. J. Hered. 94:449-456.

Koes, R., W. Verweij, and F. Quattrocchio. 2005. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. 10:236-42.

Kushad M.M., J. Masiunas, M.A.L. Smith, W. Kalt, and K. Eastman. 2003. Health promoting phytochemicals in vegetables. Hort. Rev. 28:125-185.

Li, J., X. Li, L. Guo, F. Lu, X. Feng, K. He, L. Wei, Zhangliang Chen, L.J. Qu, and H. Gu. 2006. A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. J. Exp. Bot. 57: 1263-1273.

Lijavetsky D., L. Ruiz-Garcia, J.A. Cabezas, M. T. De Andrés, G. Bravo, A. Ibáñez, J. Carreño, F. Cabello, J. Ibáñez, and J. Martínez-Zapater. 2006. Molecular genetics of berry colour variation in table grape. Mol. Gen. Genomics 276:427-435.

Liu, P-.P., N. Koizuka, R.C. Martin, and H. Nonogaki. 2005. The *BME3* (*Blue Micropylar End 3*) GATA zinc finger transcription factor is a positive regulator of *Arabidopsis* seed germination. Plant J. 44: 960-971.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56: 316-326.

Mathews H., S.K. Klendennen, C.G. Caldwell, X. Liang Liu, K. Connors, N. Matheis, D.. Schuster, D.J. Menasco, W. Wagoner, J. Lightner, and D.R. Wagner. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689-1703.

Matsui, K., Y. Umemura, and M. Ohme-Takagi. 2008. AtMYBL2, a protein with a single myb domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. Plant J. 55: 954-967.

Matus, J.T., F. Aquea, and P. Arce-Johnson. 2008. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. BMC Plant Biol. 8:83.

Menzel, G., D. Dechyeva, H. Keller, C. Lange, H. Himmelbauer, and T. Schmidt. 2006. Mobilization and evolutionary history of minature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. Chromosome Res.14:831-844.

Mes, P.J., P. Boches, J.R. Myers, and R. Durst. 2008. Characterization of tomatoes expressing anthocyanin in the fruit. J. Amer. Soc. Hort. Sci. 133:262-269.

Middleton, E. and C. Kandaswami. 1993. The impact of flavonoids on mammalian biology: implications for immunity, inflammation and cancer, p. 589-618. In: J.B. Harborne (ed.). The flavonoids: advances in research since 1986. Chapman and Hall, London.

Moyle, L.C. 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). Evolution. 62: 2995-3013.

Nicot, N., J.-F. Hausman, L. Hoffman, and D. Evers. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J. Exp. Bot. 56:2907-2914.

Okamura, K., J.W. Hagen, H. Duan, D.M. Tyler, and E.C. Lai. 2007. The mirtron pathway generates microRNA-class regulatory class RNAs in Drosophila. Cell. 130:89-100.

Ori, N., A. R. Cohen, A. Etzioni, A. Brand, O. Yanai, S. Shleizer, N. Menda, Z. Amsellam, I. Efroni, I. Pekker, J.P. Alvarez, E. Blum, D. Zamir, and Y. Eshed. 2007. Regulation of LANCEOLATE by *miR319* is required for compound-leaf development in tomato. Nat. Genet. 39:787-791.

Paz-Ares, J., D. Ghosal., U. Wienard., P.A. Peterson, and H. Saedler. 1987. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural simlarities to transcriptional activators. EMBO J. 6:3553-3558.

Quattrochio, F., J.F. Wing, H.T.C. Leppen, J.N.M. Mol, and R.E. Koes. 1993. Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell. 5:1497-1512.

Quattrocchio, F., J. Wing, K. Van der Woude, E. Souer, N. de Vetten, J. Mol, and R. Koes. 1999. Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. Plant Cell 11:1433-1444.

Rabinowicz, P.D., E.L. Braun, A. D. Wolfe, B. Bowen, and E. Grotewold. 1999. Maize R2R3 Myb genes: sequence analysis reveals amplification in the higher plants. Genet. 153:427-444.

Ramsay, N.A., and B.J. Glover. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10:62-70.

Rensselaer Bioinformatics Website. 8 February 2009. <http://mfold.bioinfo.rpi.edu/>.

Rick, C.M., P. Cisneros, R.T. Chetelat, and J.W. DeVerna. 1994. *Abg*– A gene on chromosome 10 for purple fruit derived from *S. lycopersicoides*. Tomato Genet. Coop. Rpt. 44:29–30.

Rick, C.M. 1980. Tomato, p. 669–680. In: Fehr, W.R. and H. Hadley (eds.). Hybridization of crop plants. Amer. Soc. Agron., Crop Sci. Soc. Amer., Madison, WI.

Rozen S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ.

Sapir, M., M. Oren-Shamir, R. Ovadia, M. Reuveni, D. Evenor, Y. Tadmor, S. Nahon, H. Shlomo, L. Chen, A. Meir, and I. Levin. 2008. Molecular aspects of *Anthocyanin fruit* tomato in relation to *high pigment-1*. J Hered 99:292-303

Schwinn, K., J. Venail, Y. Shang, S. Mackay, V. Alm, E. Butelli, R. Oyama, P. Bailey, K. Davies, and C. Martin. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell 18:831-851.

SOL Genomics Network website. 20 March 2007. <http://www.sgn.cornell.edu/>.

Stracke R., H. Ishihara, G. Huep, A. Barsch, F. Mehrtens, K. Niehaus, B. Weisshar. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 50:660-677.

TAIR, the *Arabidopsis* Information Resource website. 1 October 2008. <http://www.arabidopsis.org/>.

Tanksley S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G.B. Martin, R. Messeguer, J.C. Miller, L. Miller, A.H. Paterson, O. Pineda, M.S. Roder, R.A Wing, W. Wu, and N.D. Young. 1992. High density maps of the tomato and potato genomes. Genet. 132:1141-1160.

Tohge, T., Y. Nishiyama, M. Yokota Hirai, M. Yano, J.-I. Nakajima, M. Awazuhara, E. Inoue, H. Takahashi, D.B. Goodenowe, M. Kitoyama, M. Noji, M. Yamazaki, and K. Saito. 2005. Functional genomics analysis by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J. 42:218-235.

van Tuinen, A., C.H Ric de Vos, R.D. Hall, L.H.W. van der Plas, C. Bowler, and R.J. Bino. 2006. Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from matural light hyperresponsive mutants, p. 339-356. In: P.K. Jaiwal (ed.) Plant Genetic Engineering Vol. 7: Metabolic engineering and molecular farming. Studium Press, Houston, TX.

Vinson, J.A., Y. Hao, X. Su, L. Zubik. 1998. Phenol antioxidant quantity and quality in foods: vegetables. J. Agric. Food Chem. 46:3630-3634.

Wessler, S.R., T.E. Bureau, and S.E. White. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5:814-821.

Young, P.A. 1954. Purple tops on green tomato fruits. Tomato Genet. Coop. Rpt. 4:18.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406-15.

Zhang, F. and T. Peterson. 2005. Comparisons of maize *pericarp color1* alleles reveal paralogous gene recombination and an organ-specific enhancer region. Plant Cell. 17:903-914.

Zhang, F., S. Chopra, T. Peterson. 2000. A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. Plant Cell. 12:2311-2322.

Zhao, M., K. Morohashi, G. Hatlestad, E. Grotewold, and A. Loyd. 2008. The TTG1-bHLH-MYB complex controls cell fate and and patterning through direct targeting of regulatory loci. Development. 135: 1991-1999.

Zhu, Q.H., A. Spriggs, L. Matthews, L. Fan, G. Kennedy, F. Gubler, and C. Helliwell. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. 18: 1456-1465.

Zimmerman, I.M., M.A. Heim, B. Weisshar, and J.F. Uhrig. 2004. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J. 40:22-34.

CHAPTER 3

Manipulation of the Phenylpropanoid Pathway with Naturally Occurring Flavonoid Mutants to Produce a Non-Transgenic High Phenolic Tomato

Peter S. Boches and James R. Myers

Abstract

The anthocyanin regulatory genes *atroviolaceum* (*atv*) and *Anthocyanin fruit* (*Aft*), cause increased anthocyanin production in tomato (*Solanum lycopersicum* L.) fruit when combined. *Aft* and *atv* were pyramided with the anthocyanin null mutants *anthocyanin absent* (*aa*) or *anthoyanin without* (*aw*), which probably encode non-functional enzymes from the anthocyanin biosynthetic pathway. Marker assisted selection for the *Aft* gene and selection based on the Folin-Ciocalteau (F-C) assay for total phenolics content was used to select putative *AftAft/atvatv/awaw* or *AftAft/atvatv/aaaa* homozygotes. Putative *AftAft/atvatv/awaw* or *AftAft/atvatv/aaaa* homozygotes unexpectedly produced anthocyanin, in the fruit only. Greenhouse grown fruit of *AftAft/atvatv/awaw* F_3 individuals had elevated levels of chlorogenic acid and rutin relative to tomato cultivars, based on high performance liquid chromatography (HPLC) and F-C analysis. Field grown fruit of *AftAft/atvatv/awaw* F_3's had levels of total phenolics that were not significantly different from an *AftAft/atvatv* line. Feruloyl-caffeoylquinic acid (FCQA), not previously reported in tomato fruit, was identified in extracts of *AftAft/atvatv/awaw* and *AftAft/atvatv* fruit.

Introduction

Tomatoes are one of the most widely consumed vegetables in the world and a major source of phenolics, including flavonoids, for many people. Because of the health benefits of flavonoids in the human diet, increasing the flavonoid content of tomatoes has garnered increasing attention in recent years. Other phenolics such as chlorogenic acid have been recognized as having health benefits as well (Robbins, 2003). Attempts to raise the levels of phenolic compounds in tomato fruit have included searching for genes from wild species that increase the expression of flavonoid biosynthetic genes (Willits et al., 2005) as well as transgenically increasing the expression of genes involved in flavonoid production, both regulatory (Schijlen et al., 2004; 2006) and structural (Muir et al., 2001). Both of these approaches have some drawbacks. The use of wild species genes was hampered by interspecific incompatibility (and potentially by linkage drag of unfavorable alleles linked to the increased flavonoid trait); whereas the use of transgenic technology is limited by regulatory expenses and difficulties in

the public perception of genetically modified foods. Here we describe the results of an approach aimed at producing a high flavonoid tomato using the well characterized anthocyanin mutants *Aft*, *atv*, *aw*, and *aa.*

The tomato genes *Aft* and *atv* are introgressed from *Solanum chilense* Dunal (Reiche) and *Solanum cheesemaniae* (L. Riley) Fosberg, respectively (Mes et al., 2008). *Aft* causes increased anthocyanin expression in the fruit (Jones et al. 2003), while *atv* causes increased anthocyanin expression in the leaves (Mes et al., 2008). Because they appear to act upon multiple genes in the pathway, both genes appear to have some regulatory function (Kerckhoffs et al., 1997; Mes et al., 2008). Recently, an association between *Aft* and a small gene family of MYB transcription factors named *An2* and *Ant1* involved in anthocyanin regulation has been confirmed in tomato (Boches and Myers, 2007; Sapir et al., 2008). When *atv* is combined with *Aft*, the result is a dramatic increase in the level of anthocyanin and total phenolics in tomato fruit (van Tuinen et al., 2006; Mes et al., 2008).

Table 3.1. Putative associations between mapped genes and classical loci in tomato influencing anthocyanin biosynthesis, based on De Jong et al. (2004).

Gene	Classical Locus	Gene Symbol
chalcone isomerase (CHI)	*anthocyanin free*	*af*
flavanone 3-hydroxylase (F3H)	*anthocyanin reduced*	*are*
flavanoid 3'5'-hydroxylase (F3'5'H)	*anthocyaninless*	*a*
dihydroflavanol 4-reductase (DFR)	*anthocyanin without*	*aw*
anthocyanidin synthase (ANS)	*entirely anthocyaninless*	*ae*

Numerous recessive mutants that fail to express anthocyanin have been discovered in tomato. Some of these loci (*af, are, aa, a, aw,*and *ae*) were associated with RFLP markers for structural genes in the anthocyanin biosynthetic pathway (De Jong et al. 2004) (Table 3.1).The association of *a* with *F3'5'H* and *are* with *F3H* was confirmed by sequencing (De Jong et al., 2004). Tong and Yoder (1995) also found a close association between *aw* and *DFR*. This was confirmed by transformation experiments when a tomato line with the *aw* gene was restored to wild type by transformation with a functional *DFR* gene (Tong and Yoder, 1995, Goldsbrough et al, 1994). We hypothesized that by combining *AftAft/atvatv* double mutants with

increased flux through the anthocyanin pathway with mutants blocked in anthocyanin synthesis (*aaaa* or *awaw*), the result would be an accumulation of flavonols or caffeoylquinic acids. These individuals may be of use in breeding new tomato cultivars because they are expected to contain high amounts of polyphenolics, while retaining more of the (color) characteristics of a traditional tomato cultivar.

Another advantage might be faster and more consistent germination, which has been documented for anthocyaninless mutants in tomato (Atanassova and Shtereva, 1995). However, anthocyanin lacking plants may also suffer from increased susceptibility to both biotic and abiotic stresses, since anthocyanins are known to confer tolerance to both. The strategy of combining a MYB regulator of the anthocyanin pathway with a gene that blocks anthocyanin production has recently been used to engineer a potato that produces high amounts of kaempferol and chlorogenic acid in potato (Rommens et al., 2008). These authors used a genetic engineering approach to manipulate these genes, however.

Materials and Methods

Plant hybridization and selection. Seeds of 3-121 (*aw*), LA1996 (*Aft*), LA3736 (*atv*), and LA1194 (*aa*) were obtained from the Tomato Genetics Resources Center (Davis, CA). 3-121 is a chemically induced provisional *aw* (aw^{prov3}) mutant in the VF36 background. LA3736 is highly inbred in the 'Ailsa Craig' background, LA1996 is highly inbred in an undeclared background. LA1194 is the original '*aa*' mutant that arose spontaneously in cv. Marmande. Lines P19-2, P20-2, and P20-7 are stable high anthocyanin lines selected from an (+/*Abg* x *atvatv*) x (*AftAft* x *atvatv*) cross as described for line P20-3 in Mes et al. (2008). The genotype is inferred to be ++/*AftAft/atvatv*, however, since we and other authors have been unable to produce stable *Abg/Abg* homozygotes, presumably because *Abg* resides in the inversion on the long arm of *S. lycopersicoides* chromosome 10 relative to *S. esculentum* (Ji and Chetelat, 2003). *Abg* and *Aft* also have different alleles at the *An2* locus (see PCR below). *Abg* has the same allele size as cultivated tomato, while P19-2, P20-2, and P20-7 are all homozygous for the *Aft* allele. However, since *Abg* is heterozygous, it is possible that *Abg* has a null allele for *An2* from *S. lycopersicoides*.

The high phenolics populations P172 (*AftAft/atvatv x awaw*), P173 (*AftAft/atvatv* x *aaaa*), and P187 (*aaaa* x *AftAft/atvatv*) were created from P19-2 x 3-121, P20-2 x LA1194, and LA1194 x P20-7 crosses, respectively. P19-2 is a large (slicer type) high anthocyanin line with good flavor characteristics, P20-2 and P20-7 are smaller fruited (saladette type) high anthocyanin lines similar to P19-2 but with poorer flavor than P19-2. Crosses were made in the greenhouse using standard crossing and emasculation techniques (Rick, 1980). Since the *aw* and *aa* phenotypes were expected to mask (be epistatic to) those of *Aft* and *atv*, selections in segregating populations were made as follows. In the F_2 generation, green hypocotyl (*awaw*) individuals were selected at the seedling stage and subjected to marker assisted selection for the *Aft* gene using the *An2* PCR marker (see DNA Extraction and PCR below). Putative homozygous *AftAftawaw* individuals were planted in the field in 2007 and selected for high total phenolics using the F-C assay (see Phenolics extraction and chemicals below). F_3 families from F_2 selections were grown to maturity in the greenhouse during winter 2008, phenolics were extracted from fruit, and phenolics content was evaluated using HPLC (see below). F_3 families (five plants each) from F_2 selections were grown again in the field during summer 2008 and evaluated for horticultural quality. In summer 2008, phenolics were extracted from fruit of F_3 individuals evaluated using the F-C and pH differential assays.

Plant Growth. Tomato seeds were sown in Sunshine SB40 professional growing mix (Sun Gro Horticulture, Bellevue, WA) in 5-cm-diameter plastic transplant cells. F_2 populations were grown during the summer of 2007 at the Oregon State University (OSU) Vegetable Farm. Plants were transplanted May 21 into rows 0.9 m apart with 60 cm within-row spacing. 'Gold Nugget', 'Siletz', and LA 1996 were included as standards for low, medium, and high phenolics tomato lines, respectively. Fertilizer (505 $kg \cdot ha^{-1}$ of 12N-29P-10K-4S) was banded before transplanting. Plants were irrigated at weekly intervals until mid-August when water was withheld. Plots were maintained using standard cultural practices for fresh market production with application of copper hydroxide fungicide (Kocide; Griffin L.L.C. Valdosta, GA) at

label rates late in the growing season when climatic conditions favored development of late blight [*Phytopthora infestans* (Mont.) de Bary].

During winter 2007, F_3 individuals from F_2 selections were started in the greenhouse in 5-cm-diameter plastic transplant cells and transplanted into 3.8-L pots filled with Special Blend (Sun Gro Horticulture, Bellevue, WA) potting mix and 20 g of 14N–6.1P–11.6K slow-release fertilizer (Simplot 14–14–14) after 3 weeks. Plants were given 10 g of supplemental fertilizer after 4 weeks in the pots. Greenhouse temperature was set to 18°C night and 25°C day. Supplemental lighting during was provided for 16 hours per day by 400-W metal halide and 400-W high-pressure sodium Sun System 3 high-intensity discharge lamps (Sunlight Supply, Woodland, WA). Plants were grown in three replicates in a completely randomized design. 'Legend', LA 1996, and P20-3 (a sister line to P20-2) were included as standards for greenhouse evaluations. In summer 2008, F_3 families were grown in the field at the OSU Vegetable Farm as for summer 2007. P20-3 and 'Legend' were included as standards in summer 2007.

DNA extraction and Polymerase Chain Reaction (PCR). DNA extraction was carried out according to the method of Fulton et al. (1995). Primers designed from a tomato EST (*SlAn2,* SGN-U228086) with similarity to Petunia *An2* (5' to 3', An2F:GCATCGTTGGGAGTTAGGAA, AnR:AACGAGGACGAGAATGAGGA) were used to genotype F_2 individuals at the *SlAn2* locus (Boches and Myers, 2007). These primers produce a 700 bp product in *Aft* and an 1100 bp product in cultivated tomatoes, in a co-dominant fashion. PCR reactions were carried out in 15 µl of 1X Biolase NH4 reaction buffer, 2 mM $MgCl_2$, 200 µM each of dATP, dCTP, dGTP, and dTTP, 0.3 µM each of forward and reverse primers, 0.25 units of Biolase Taq DNA polymerase (Bioline Inc., Randolph, MA), and 2.5 ng of template DNA. The PCR program consisted of 35 cycles of 94° C for 30 s, 60° C for 30 s, 72° C for 30 s, followed by a 72° extension for 5 minutes. PCR products were separated on 2% agarose gel and visualized under UV light after staining with ethidium bromide.

Phenolics extraction and chemicals. Combined flesh and peel extracts were created from fruit cut in half lengthwise, with seeds, placenta, and columella tissue removed. Peel extracts were created from strips of peel (epidermis and upper pericarp cells) created by scoring tomato fruit lengthwise and removing the peel in strips. The peel or peel and flesh tissue was frozen with liquid nitrogen and ground to a fine powder using a mortar and pestle. Weighed portions of this powder were used for extractions. For field grown samples, 70-150 mg of the tomato powder was extracted overnight in 300 µL of acidified methanol (1% HCl) at 4° C in 1.5 mL Eppendorf tubes. Extractions were brought to 500 µL with nanopure H_20, vortexed with 500 µL chloroform, and centrifuged at 13,000 rpm for ten minutes. The aqueous supernatant was removed and stored at -20° C for further analysis. The greenhouse grown samples were processed using a scaled up version of this protocol (1 g of powdered tissue was extracted in 5 mL 60:40 1% HCl methanol: water and 5 mL chloroform in 15 mL glass tubes). Gallic acid, rutin, 5-caffeoylquinic acid, and naringenin purchased from Sigma Chemical Corp (St Louis, MO) were used as standards for F-C and HPLC.

F-C *and pH Differential.* F-C assays were performed according to the method of Singleton and Rossi (1965) and Waterhouse (2005), with modifications. F-C reactions were performed on a micro-scale: each reaction contained 198 µL water, 12.5 µL F-C reagent (Sigma), 2.5 µL of each sample or gallic acid standard, and 37.5 µL sodium carbonate solution. Reactions were incubated for 1 hour at room temperature, and absorbance at 765 nm was read using a SPECTRAmax microplate spectrophotometer (Molecular Devices, Sunnyvale, CA, USA). F-C values were expressed as gallic acid equivalents (GAE) based on a gallic standard curve in the range of 0-500 $mg \cdot L^{-1}$. Monomeric anthocyanin content was measured using the pH differential method as described by Giusti and Wrolstad (2001). For the pH differential method, the samples were diluted 1:5 with the pH buffer and absorbance was read at an observed $\lambda_{vis\text{-}max}$ of 540 nm. The predominant anthocyanin in tomato fruit has previously been reported to be petunidin-3-(p-coumaryl)-rutinoside-5-glucoside (Mes et al., 2008). For the calculation of the monomeric anthocyanin pigment content, a molecular weight of 934

and a molar absorptivity (ε) of 17000 was used, corresponding to petunidin-3-(p-coumaryl)-rutinoside-5-glucoside in acidified methanol (Price and Wrolstad, 1995).

HPLC *and Mass Spectrometry (MS) analysis.* A Hewlett-Packard (Palo Alto, CA) 1050 series auto sampler and pump coupled to a HP 1040M diode array detector detection system equipped with an ES Industries (West Berlin, NJ) reversed phase LiChrospher RP18 endcapped column (25 cm x 4.6 mm, 5 µm particle size) and a guard column of the same material was used. The separation gradient and buffers followed that of the tomato metabolite database (Moco et al., 2006). Solvent A was formic acid: water (1: 1000). Solvent B was formic acid: acetonitrile (1: 1000). The flow rate was 1 $mL \cdot min^{-1}$, and the injection volume was 40 µL. The elution conditions were as follows: 0-45 min, linear gradient from 5% B to 35% B; 45-50 minutes, isocratic at 85% B to wash the column; 50-60 minutes, isocratic at 5% B to equilibrate the column. Peak spectra were monitored at 280 nm. UV spectra were recorded from 200 to 600 nm in 0.5 nm intervals. HPLC-PDA data were analyzed using Agilent ChemStation Rev.A.09.01 (Santa Clara, CA). Compounds were quantified as equivalents to naringenin, chlorogenic acid, or rutin using a standard curve and summarized according to compound classes. Standards were used in the range of 25-250 $mg \cdot L^{-1}$ for rutin, 5-50 $mg \cdot L^{-1}$ for naringenin, and 25-500 $mg \cdot L^{-1}$ for chlorogenic acid.

HPLC-MS was performed using the same column, buffers, and elution conditions on a Shimadzu LC-20-AD chromatography system (Shimadzu, Columbia, MD) coupled to an ABI 3200 QTRAP LC/MS/MS (Applied Biosystems, Foster City, CA) triple quadropole mass spectrometer equipped with a Turbo V ionization source. HPLC-MS was performed on skin extracts of field grown P20-3 (*AftAft*/*atvatv*) and an F_3 P172 selection (*AftAft*/*atvatv*/*awaw*). Combined flesh and peel extracts from *Aft*, 'Legend', and a P172 F_3 individual (*AftAft*/*atvatv*/*awaw*) were also analyed by HPLC-MS. Acquisition was performed in Q1 MS scan mode monitoring *m/z* from 200 to 1000, 1 scan 2 s^{-1}. Turbospray conditions were as follows: curtain gas=20 psi, ion spray voltage=-4500 V, temperature=650°C, ion source gas 1=40 psi, ion source gas 2=40 psi, declustering potential=-45 V, entrance potential = -10 V. Data analysis was

carried out using Applied Biosystems Analyst software V1.4.2. HPLC-MS^2 analysis of peel extract of P20-3 was performed on the same system, with collision energy=-27 V.

Statistical Analysis. SAS software (Cary, NC) was used for analysis of variance and means separations using Fisher's Least Significant Difference (LSD).

Results

Plant Hybridization and Selection. The P172 cross (*AftAft/atvatv* x *awaw*) segregated 354 purple: 86 green, in the expected ratio of 3 purple: 1 green (χ^2= 0.094, p=0.759). An additional 20 green seedlings (not included in the above segregation ratio) were included for a total of 106 *awaw* seedlings. The P173 cross (*AftAft/atvatv* x *aaaa*) segregated 195 purple: 47 green (χ^2 for 3 purple: 1 green= 4.016, p=0.045) and the P187 cross (*aaaa* x *AftAft/atvatv*) segregated 189 purple: 56 green (χ^2 for 3 purple: 1 green= 0.6, p=0.439). Although the P173 cross deviated significantly from the expected 3:1 ratio, a 3:1 ratio was still the best fit for the segregation observed. Green (homozygous *aw* or *aa*) seedlings were genotyped for the presence or absence of the *Aft* allele at *An2* locus using the *SlAn2* PCR primers. We selected F_2 individuals that had only the 700 bp *Aft* allele of *An2* (Figure 3.1). Some *Aft*/+ heterozygotes may have been included in the selected populations, however, due to a PCR amplification preference for the *S. chilense Aft* allele when both alleles are present. Out of 106 green (*awaw*) *AftAft/atvatv* x *awaw* F_2's, we selected 40 putative *AftAft/awaw* individuals. Out of 95 green *aaaa* x *AftAft/atvatv* or *AftAft/atvatv* x *aaaa* F_2's, we selected 29 putative *AftAft/aaaa* individuals.

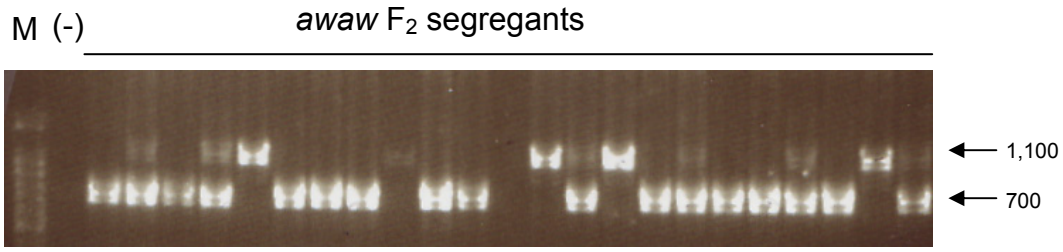


Figure 3.1. Marker assisted selection of *AftAft/atvatv* x *awaw* F_2 segregants (P172). Gel image of An2 PCR products from DNA of F_2 segregants with green hypocotyls (*awaw*) separated on a 2% gel and visualized with ethidium bromide under UV light. Arrows indicate the 1,100 bp allele from the *aw* parent and the 700 bp allele from the *Aft* parent. Individuals lacking the 1,100 bp allele were selected. Abbreviations: M, 100 bp ladder (Promega).

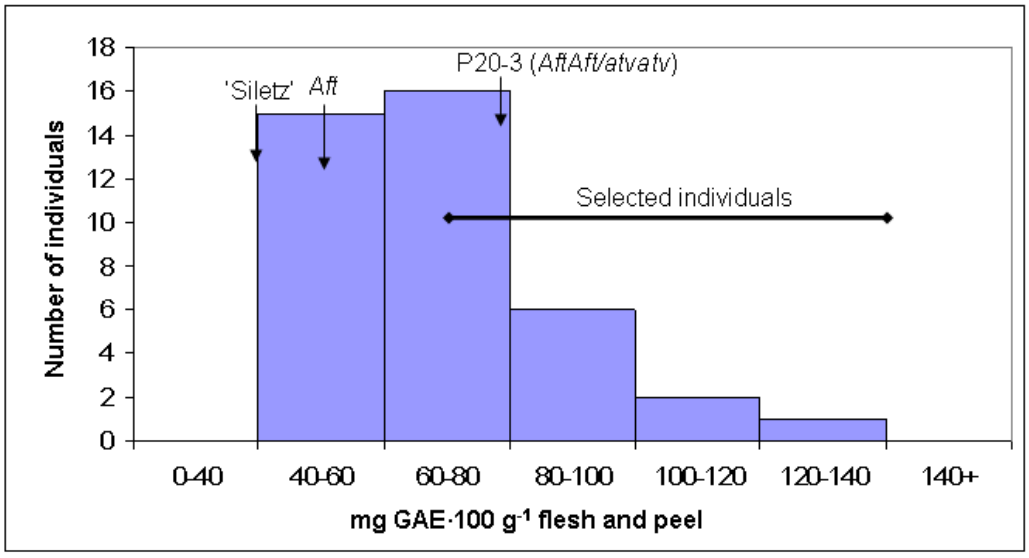


Figure 3.2. Histogram of F-C values from fruit flesh and peel extracts of P172 *AftAftatvatv* x *awaw* F_2 individuals. F_2 individuals with total phenolics ≥ 68 mg·100 g^{-1} FW (the lowest observed score for a P20-3 individual) were selected. Average F-C values for the standard genotypes 'Siletz' (40 mg·100 g^{-1} FW), LA1996 *Aft* (49 mg·100 g^{-1} FW), and P20-3 (*AftAft*/*atvatv,* 78 mg·100 g^{-1} FW) are indicated by arrows. Abbreviations: FW, fresh weight.

The distribution of total phenolics scores for the *AftAft/atvatv x awaw* population is shown in Figure 3.2. Fruit flesh and peel of the 40 putative *AftAft/awaw* individuals had a mean total phenolics content of 64.13 mg·100 g^{-1} FW GAE (high =167.26, low = 24.19) as evaluated by the F-C assay. This is slightly higher than the

value obtained for *Aft* (mean=48.97, standard error=8.11). A selection cut-off value for total fruit phenolics content ≥ 68 mg·100 g^{-1}FW (the lowest value observed for an *AftAft/atvatv* line) was applied. Twenty of the putative *AftAft/awaw* F_2's exceeded the selection cutoff and were evaluated further evaluation as potential *AftAft/atvatv/awaw* high polyphenolic lines. The mean total phenolics content of fruit flesh and peel in the *aaaa* x *AftAft/atvatv* and *AftAft/atvatv x aaaa* population was only 37.55 mg·100 g^{-1} FW GAE (high= 57.76, low=22.48). Due to the low values observed for the *aa* segregants, we selected only five individuals from the *aaaa* x *AftAft/atvatv* and *AftAft/atvatv x aaaa* populations, for observational purposes only.

Unexpectedly, small amounts of anthocyanin were observed in the fruit of some segregants for both the *aw* and *aa* crosses (Figure 3.3). We noticed anthocyanin in approximately 5 of 20 selected *AftAft/atvatv* x *awaw* F_2 individuals. Anthocyanin was observed in the majority (19 of 20) *AftAft/atvatv* x *awaw* F_3 families from the 20 selected F_2's (five plants per F_3 family). The *awaw* plants producing anthocyanin in the fruit did not produce any noticeable anthocyanin in the leaves and otherwise had the green hypocotyl phenotype typical of *aa* or *aw* plants. The *AftAft/atvatv* x *awaw* population also segregated for dark green foliage color. Plants with dark green foliage seemed to have intensified anthocyanin accumulation in the fruit; however, some plants with light green foliage and anthocyanin in the fruit were observed. The population also segregated for uniform ripening (*U*). Uniform ripening fruit (*uu)* did not produce as much anthocyanin as fruit with green shoulder (U^+), however, some *uu* fruit expressing anthocyanin in the fruit were observed. The color of the anthocyanin in the *AftAft/atvatv/awaw* and *AftAft/atvatv/aa* F_2 and F_3 segregants appeared to be somewhat muddy or dusky compared to *Aft*/*atv* fruit. When we soaked slices of unripe fruit in 10% HCl to visualize anthocyanin pigments, *AftAft/atvatv* fruit peel turned a bright pink color, while *AftAft/atvatv/awaw* fruit peel turned a brick red color, indicating a different composition of anthocyanin pigments in *Aft*/*atv*/*aw* fruit (Figure 3.4).

Figure 3.3. Light and dark fruited segregants of *aaaa* x *AftAft/atvatv* (left) and *AftAft/atvatv* x *awaw* (right) F_3 families. For both photos the light fruited segregant is shown at left, middle fruit is dark fruited segregant (*AftAft/atvatv/aaaa* or *AftAft/atvatv/awaw*), and right fruit is *AftAft/atvatv* parent. Top row, ripe fruit; bottom row, unripe fruit.

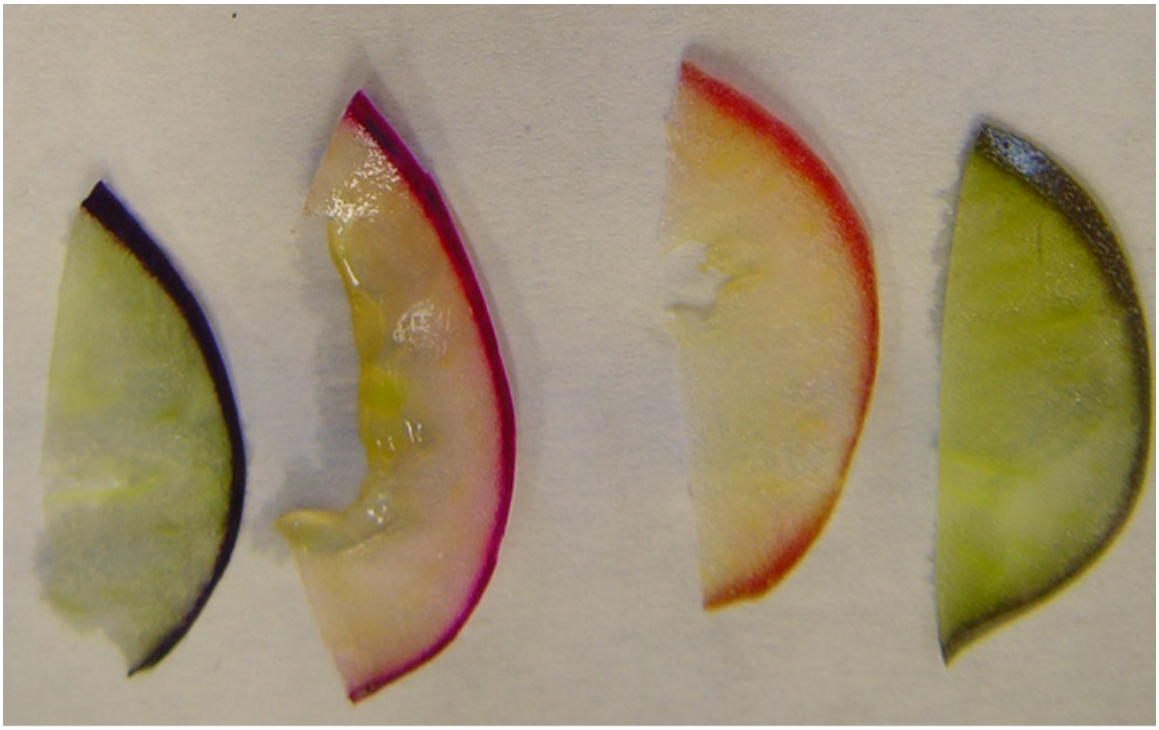


Figure 3.4. Slices of unripe fruit from *AftAft/atvatv* and *AftAft/atvatv/awaw* genotypes treated with 10% HCl to visualize anthocyanin pigments. Left to right: *AftAft/atvatv* untreated slice, *AftAft/atvatv* treated slice, *AftAft/atvatv/awaw* treated slice, *AftAft/atvatv/awaw* untreated slice. Note the brick red color of the *AftAft/atvatv/awaw* treated fruit, an indication of different anthocyanin pigments.

Phenolics Analysis. We grew F_3 individuals from the *AftAft/atvatv* x *awaw* population under greenhouse conditions and examined extracts of fruit flesh and peel using HPLC-PDA to examine the phenolics profiles of putative *AftAft*/*atvatv*/*awaw* F_3 segregants (Figure 3.5). Thirty two peaks were putatively identified to a compound or class of compounds using a combination of UV- and accurate mass spectra (Table 3.2). Fruit extracts of a dark fruited (fruit anthocyanin producing) *AftAft/atvatv* x *awaw* F_3 segregant (P172-1) had a different phenolics profile than a light fruited (non-fruit anthocyanin producing) *AftAft/atvatv* x *awaw* F_3 segregant (P172-2) (Figure 3.5).

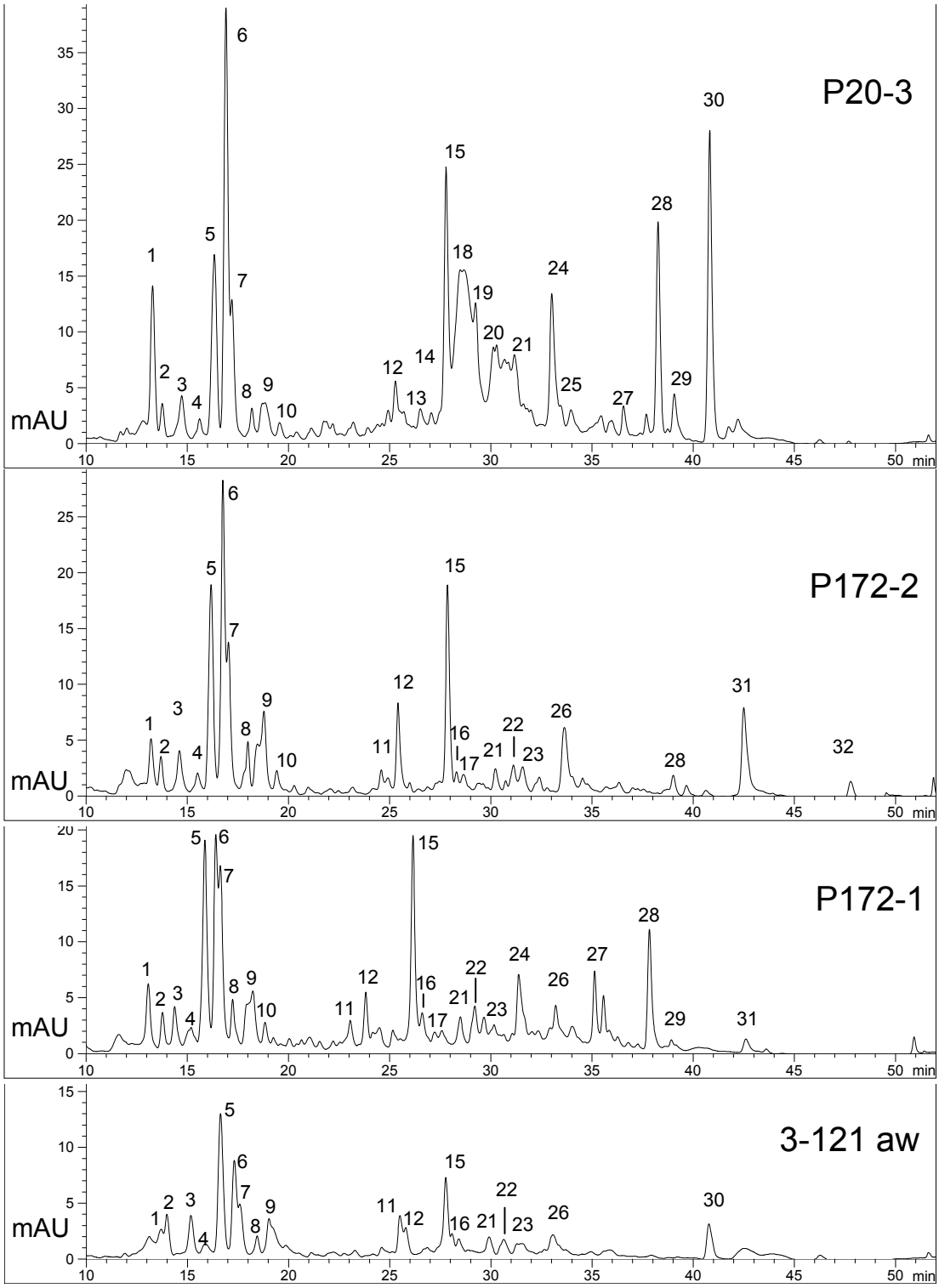


Figure 3.5. HPLC-PDA chromatograms (280 nm) of flesh and peel extracts from *AftAft/atvatv* x *awaw* F_3 individuals (P172-1 and P172-2, middle), the *AftAft/atvatv* parent (P20-3, top), and the *awaw* parent (3-121, bottom). Peak numbers refer to metabolite information in Table 3.2.

Table 3.2. Metabolites found in flesh and peel extracts of tomato fruit. Peak numbers (no.) refer to Figure 3.5. RT= retention time in minutes, UV-Vis = UV-Visible spectrum (sh=shoulder), [M-H]- = detected mass of peak identified at a similar RT in HPLC-MS analysis, Type=compound class used for summarizing compounds (see Figures 3.6 and 3.7B). Putative identifications (ID) and references (see footnotes) are provided.

No.	RT	UV-VIS	[M-H] -	Type	Putative ID
1	13.41	294		CAH	Cinnamic acid
2	13.85	268		CAH	Cinnamic acid
3	14.88	322, 290sh	341.0	CAH	Caffeic acid hexose[1]
4	15.47	292, 318		CAH	Cinnamic acid
5	16.41	294	341, 325	CAH	Caffeic acid hexose, coumaric acid hexose[1,8]
6	17.01	326, 300sh	353.3	CQA	3-caffeoylquinic acid (chlorogenic acid)[1]
7	17.29	320, 290sh	353.3	CQA	5-caffeoylquinic acid[1]
8	18.13	256, 344	771.3	QUE	Quercetin 3-O-Rutinoside-7-O-Glucoside[1,2]
9	18.46	280, 308sh		CHA	Chalcone derivative
10	18.64	328	353.0	CQA	4-caffeoylquinic acid[1]
11	24.95	328	367.0	FQA	Feruloylquinic acid[1]
12	25.34	256, 352	741.1	QUE	Pentosyl rutin, (Quercetin 3-(2"Apiosyl-6" rhamnosylglucoside)[1,3,4]
13	25.96	312		FQA	
14	26.40	312		FQA	
15	27.55	256, 354	609.1	RUT	Rutin[1]
16	28.25	286, 330sh	597.2	CHA	Phloretin 3',5'-Di-C-glucoside[3]
17	28.65	318, 296sh			
18	27.90	292, 534	931.5, 949.6	-	Anthocyanins
19	28.43	264, 344	593.2	KAE	Kaempferol 3-(6"Rhamnosylglucoside)[2,3,5]
20	30.05	316, 290sh		-	Anthocyanin
21	30.22	322		DI	Diacylquinic acid
22	32.06	326, 292sh	515.1	DI	Dicaffeoylquinic acid isomer[1,6]
23	33.21	316		DI	Diacylquinic acid
24	32.40	328, 292sh	515.1	DI	Dicaffeoylquinic acid isomer[1,6]
25	33.20	low		-	
26	33.56	326, 292sh	515.3	DI	Dicaffeoylquinic acid isomer[1,6]
27	35.81	316		QUE	Quercetin derivative[7]
28	37.48	258, 314, 360sh	887.6, 771.4	QUE	Quercetin hexose-deoxyhexose-pentose-p-coumaric acid, [2,5]; Myrecitin derivative[7]
29	38.36	268, 314, 354sh		QUE	Quercetin derivative[7]
30	40.70	324		TRI	Tricaffeoylquinic acid isomer[1]
31	42.45	328		TRI	Tricaffeoylquinic acid isomer[1]
32	47.80	366	271.2	NAR	Naringenin chalcone[3]

[1]Moco et al., 2006; [2]Ijiima et al., 2008; [3]Slimestad et al., 2008; [4]Giuntini et al., 2008; [5]Moco et al., 2007; [6]Clifford et al., 2006, [7]Butelli et al., [8]2008; Luo et al., 2008.

Aft/*atv*/*aw* F_3 segregant (P172-1) (Figure 3.5).

Fruit extracts of *AftAft/atvatv* x *awaw* F_3 segregants contained several compounds that were not detected or detected in only trace quantities from cultivars (e.g. peaks 10, 24, 26, 27, 28, and 29). Based on peak identifications in Table 3.2, phenolic content was summarized by compound class (assignment of peaks to a compound class follows the 'Type' column of Table 3.2) and quantitated using a standard curve from a similar compound. Standard curves were linear over the range sampled ($R^2 \geq 0.9816$).

Figures 3.6 and 3.7 summarize replicated phenolics analysis of greenhouse grown fruit from *AftAft/atvatv* x *awaw* F_3 segregants. P20-3 (*AftAft*/*atvatv*) consistently had the highest levels of most phenolic compounds, with the exception of a putative cinnamic acid hexose (peak 1 in Table 3.2), chalcone derivatives, and naringenin chalcone. Peak 1 was the predominant phenolic compound in flesh and peel of LA1996 (mean=7.61 mg·100 g^{-1} FW, a level significantly higher than in any other genotype examined) (Figure 3.7B). The highest levels of chalcone derivatives were found in P172-2. Naringenin chalcone was relatively more abundant in cultivated tomatoes (Figure 3.7B). The dark fruited *AftAft/atvatv* x *awaw* F_3 segregant (P172-2) and P20-3 had significantly higher ($P \geq 0.05$) levels of caffeoylquinic acids, diacylquinic acids, triacaffeoylquinic acids, rutin, chalcone derivatives, and total phenolics (F-C) than 'Legend' or LA1996 (Figures 3.6 and 3.7). Anthocyanin (measured by pH differential) was only present at detectable levels in P20-3 (though P172-2 was seen to produce anthocyanin containing cells in fruit). The light fruited *AftAft/atvatv* x *awaw* F_3 segregant (P172-1) had measurably higher levels of many phenolic compounds than 'Legend' or LA1996, but the differences were only significant ($P \geq 0.05$) in the case of caffeoylquinic acids, diacylquinic acids, and rutin (Figures 3.6 and 3.7).

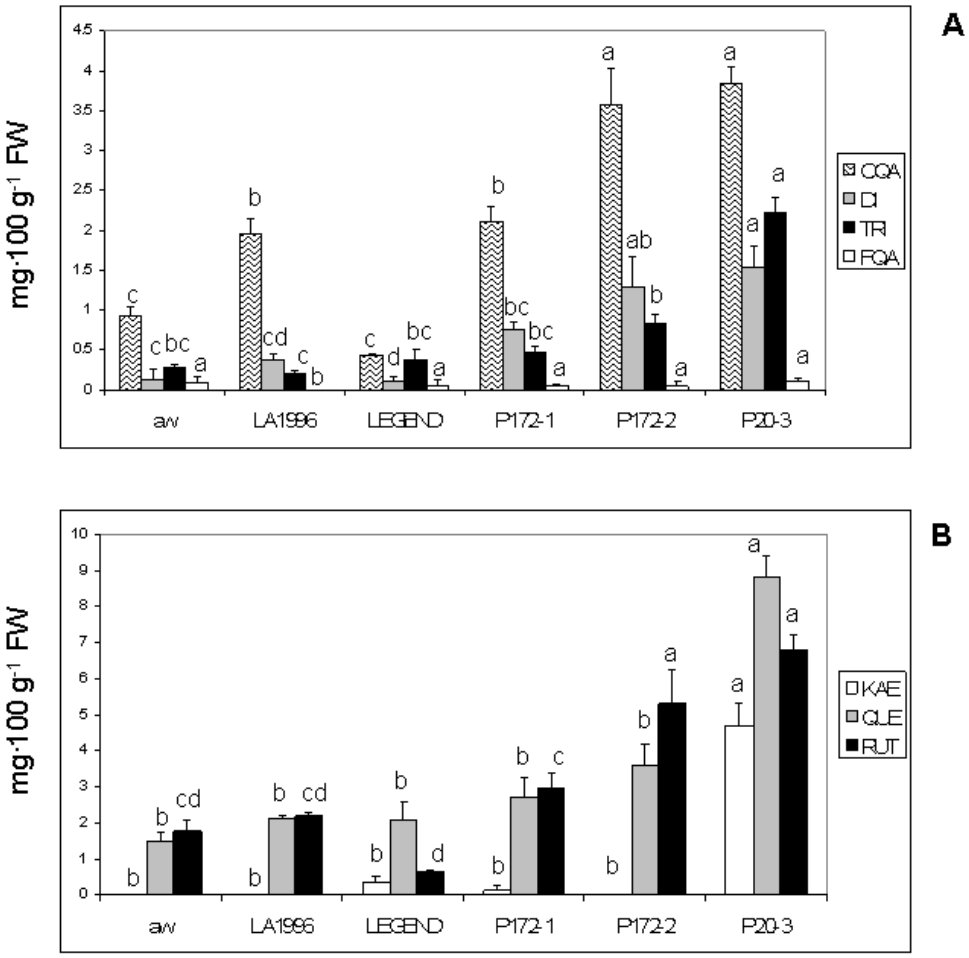


Figure 3.6. Phenolics content of flesh and peel extracts from greenhouse grown tomato fruit. (A) Summary of HPLC results for caffeoylquinic acids (CQA), diacylquinic acids (DI), tricaffeoylquinic acids (TRI), and feruloylquinic acids (FQA). (B) Summary of HPLC results for kaempferol derivatives (KAE), quercetin derivatives excluding rutin (QUE), and rutin (RUT). Peaks included in each class of compounds are indicated in the 'Type' column of Table 3.2. Values represent the mean of 3 or more biological replicates, except for *aw* which was analyzed in 2 replicates. Error bars = one standard error. Means not sharing a letter are significantly different (within compound classes), determined by Fisher's LSD, $P \geq 0.05$.

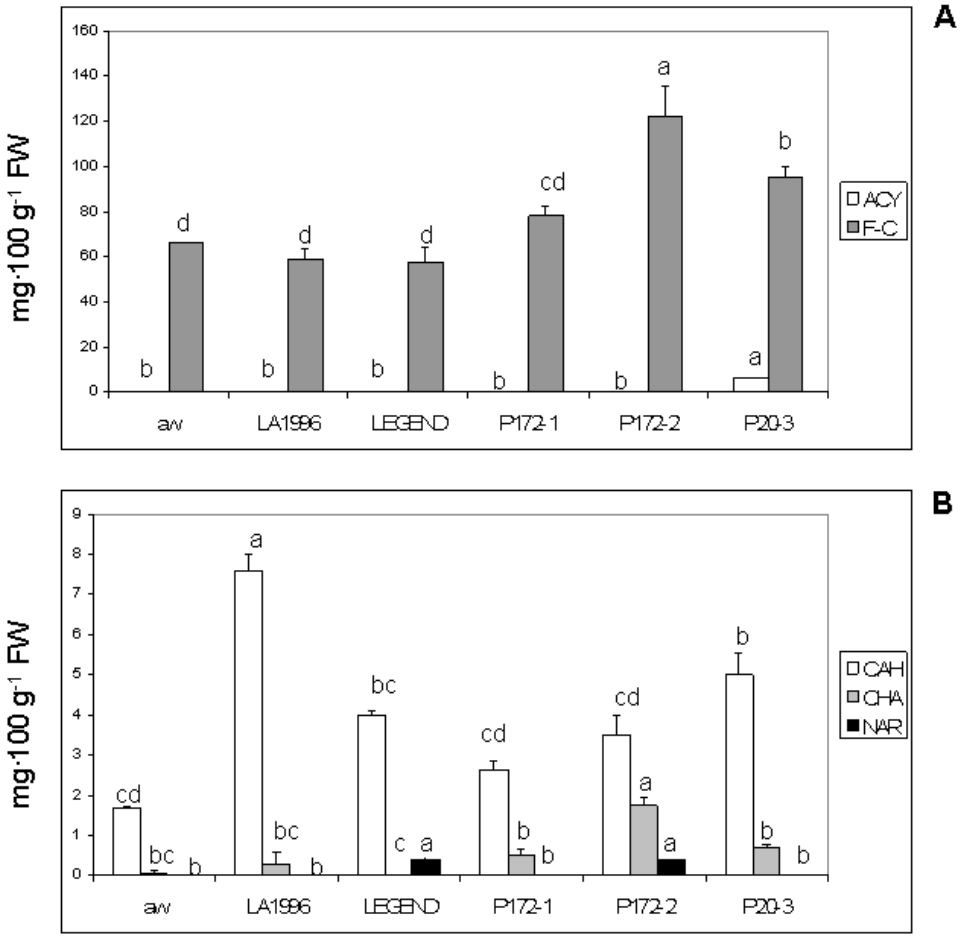


Figure 3.7. Phenolics content of flesh and peel extracts from greenhouse grown tomato fruit. (A) Anthocyanin content and total phenolics measured by pH differential and F-C assays, respectively. (B) Summary of HPLC results for cinnamic acid hexoses (CAH), chalcone derivatives (CHA), and naringenin chalcone (NAR). Peaks included in each class of compounds are indicated in the 'Type' column of Table 3.2. Values represent the mean of 3 or more biological replicates, except for *aw* which was analyzed in 2 replicates. Error bars = one standard error. Means not sharing a letter are significantly different (within compound classes), determined by Fisher's LSD, $P \geq 0.05$.

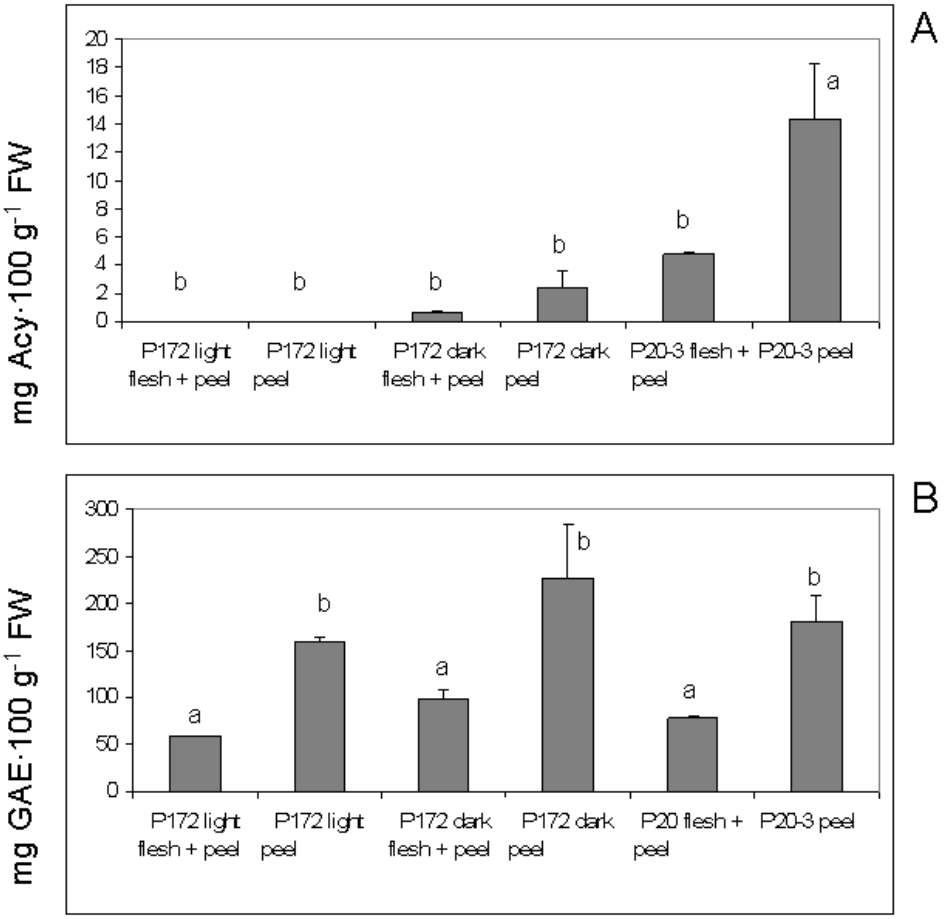


Figure 3.8. Anthocyanin (A) and total phenolics (B) content of field grown P20-3 (*AftAft/atvatv*) and P172 (*AftAft/atvatv* x *awaw*) tomato fruit (2008). Fruit anthocyanin producing (P172 dark) and non-fruit anthocyanin producing (P172 light) P172 *AftAft/atvatv* x *awaw* F_3 individuals were sampled separately. Samples were taken from flesh and peel, and peel only. Total phenolics was measured using the F-C assay, anthocyanin was measured by the pH differential method.Values represent the mean of 4 or more biological replicates, error bars = one standard error. Means not sharing a letter are significantly different, Fisher's LSD, P≥0.05.

Fruit extracts from both peel and flesh and peel alone were taken from field grown *AftAft/atvatv* x *awaw* F_3 segregants (dark and light fruited) and P20-3 to verify the levels of phenolics in P172 *AftAft/atvatv* x *awaw* F_3 segregants relative to P20-3 (Figure 3.8). Total phenolics and anthocyanins were measured using the F-C assay and pH differential method, respectively. Total phenolics and anthocyanins (when present) were significantly higher in peel extracts than flesh and peel extracts among all genotypes. The levels of total phenolics were not significantly different between light or dark fruited P172 individuals and P20-3 (within a tissue type). Anthocyanin levels were measurable in dark fruited P172 individuals (mean=2.43 mg·100 g^{-1} FW peel, 0.58 mg·100 g^{-1} FW flesh and peel). The levels of anthocyanins in P20-3 peel (mean=14.39 mg·100 g^{-1} FW) were significantly higher than the levels of anthocyanins found in P20-3 flesh and peel (mean=4.72 mg·100 g^{-1} FW) or P172 dark fruited segregants.

Peel extracts of field grown P20-3 and a P172 dark fruited individual were analyzed using HPLC-MS to identify differences in phenolics profiles (Figure 3.9 and Table 3.3). Total ion chromatograms (Figure 3.9) indicated several peaks (1, 2, 3, 4, 8, 9, 10, 11, 12, 18, 19, 21, 25, 28, and 29) found at different levels between the two genotypes. Of note are peaks 16, 17, 19, and 21. Peaks 16 and 17 were present in P20-3 but not P172, eluted at the same time as P20-3 anthocyanins (slightly after rutin), and have masses corresponding to petunidin-3-(coumaroyl)rutinoside-5-glucoside (*m/z* 931.5) and petunidin-3-(caffeoyl)rutinoside-5-glucoside (*m/z* 949.6). Selective fragment-targeted MS^2 experiments by collision induced dissociation (CID-MS/MS) of peak 17 (*m/z* 949.6) produced a fragment ion at with *m/z* 787.8, corresponding to petunidin+caffeic acid+rutinoside. Peaks 19 and 21 were present in P172, but absent or present at much lower levels in P20-3, and had masses corresponding to delphinidin-3(caffeoyl)rutinoside-5-glucoside (*m/z* 933.4) and delphinidin-3-rutinoside-5-glucoside (*m/z* 771.4). Peak 25 was also found at higher levels in P172 than P20-3, and had a mass corresponding to a quercetin hexose-deoxyhexose-pentose-p-coumaric acid (*m/z* 887.4). MS^2 experiments targeting the mass of this peak

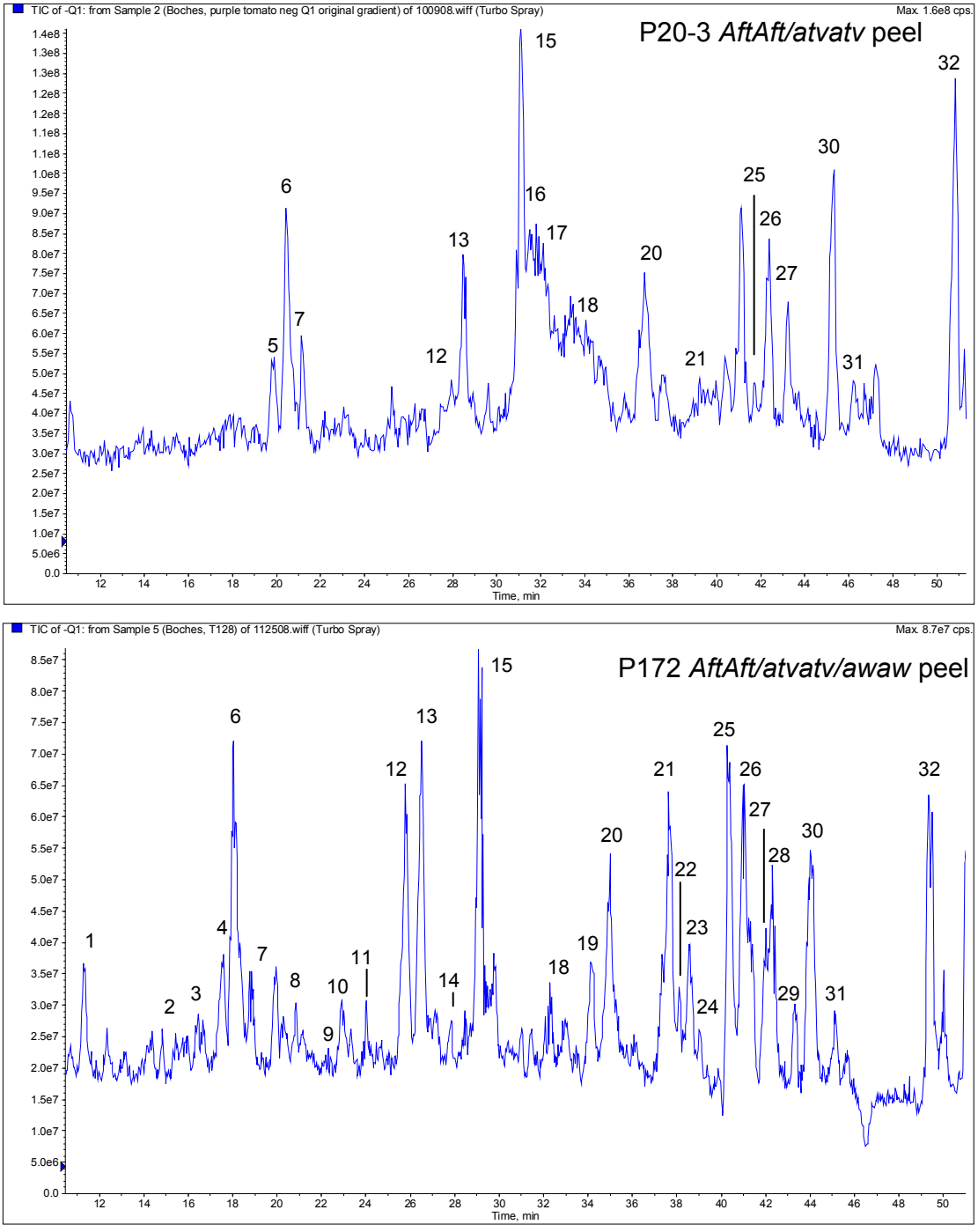


Figure 3.9. Total ion chromatograms of tomato peel extracts from P20-3 (top) and a dark fruited P172 individual (bottom). Accurate masses and putative identities of numbered peaks are listed in Table 3.3. Note that peak numbers do not necessarily correspond to those of Figure 3.5. Units on Y axis = ion counts per second.

Table 3.3. Accurate masses of peaks detected by HPLC-MS in peel extracts of P20-3 and P172 dark fruited segregants. Abbreviations: No., peak number; RT, retention time; $[M-H]^-$, accurate mass detected in negative ionization mode; ID, putative identification (see footnotes). Peak numbers (no.) refer to Figure 3.9, note that peak numbers do not necessarily correspond to those in Figure 3.5 (HPLC-PDA analysis).

No.	RT	$[M-H]^-$	ID
1	12.12	341.0	Caffeic acid hexose[1,5]
2	14.95	341.2	Caffeic acid hexose[1,5]
2	14.95	325.0	Coumaric acid hexose[1,5]
3	16.74	341.0	Caffeic acid hexose[1,5]
4	18.34	325.0	Coumaric acid hexose[1,5]
5	18.34	341.1	Caffeic acid hexose[1,5]
6	18.82	353.2	Caffeoylquinic acid[1,5]
7	19.70	771.3	Quercetin 3-o-rutinoside-7-o-glucoside
8	21.60	353.2	Caffeoylquinic acid[1,5]
9	22.00	755.2	Kaempferol dihexose-deoxyhexose[2]
10	23.65	787.3	?
11	24.70	367.2	Feruloylquinic acid[1,5,6]
12	26.48	625.2	Myricetin rhamnoglucoside[7]
13	27.28	367.2	Feruloylquinic acid[1,5,6]
13	27.28	741.2	Rutin pentoside[1,2,3,4]
14	29.20	367.1	Feruloylquinic acid
15	29.86	609.2	Rutin[1,2,3,4,5]
16	31.20	931.5	Petunidin-3-(coumaroyl)rutinoside-5-glucoside[9,10,11]
17	31.20	949.3	Petunidin-3-(caffeoyl)rutinoside-5-glucoside[9,10,11]
18	33.07	593.1	Kaempferol 3-(6"rhamnosylglucoside)
19	34.94	933.4	Delphinidin-3(caffeoyl)rutinoside-5-glucoside[9,10,11]
20	35.75	515.3	Dicaffeoylquinic acid[1,6,8]
21	38.49	771.4	Delphinidin-3-rutinoside-5-glucoside[9,10,11]
22	39.00	771.2	Delphinidin-3-rutinoside-5-glucoside[9,10,11]
23	39.40	801.3	Malvidin -3-rutinoside-5-glucoside[9,10,11]
24	39.80	887.4	Quercetin hexose-deoxyhexose-pentose-p-coumaric acid[2]
25	41.16	785.3	Patuletin+rutinoside+rhamnoside[12]
26	41.80	755.3	Quercetin derivative[7]
27	42.80	755.2	Quercetin derivative[7]
27	42.80	785.3	Patuletin+rutinoside+rhamnoside[12]
28	43.20	529.3	Caffeoyl-feruloylquinic acid[6,13]
29	44.10	799.3	Malvidin-3-rutinoside-5-glucoside[9,10,11]
30	44.70	677.3	Tricaffeoylquinic acid[1,6,8]
30	44.87	769.3	Petunidin+rutinoside+coumaric acid[10]
31	45.80	799.3	Malvidin-3-rutinoside-5-glucoside[9,10,11]
32	49.93	271.1	Naringenin/Naringenin chalcone[1,2,3,5]

[1]Moco et al., 2006; [2]Ijiima et al., 2008; [3]Slimestad et al., 2008; [4]Giuntini et al., 2008; [5]Moco et al., 2007; [6]Clifford et al., 2006, [7]Butelli et al., 2008; [8]Luo et al., 2008; [9]Ando et al., 1999; [10]Mes et al., 2008; [11]Mathews et al. 2003, [12]Rommens et al. 2008, [13]Kammerer et al., 2004.

produced an ion with *m/z* 741.0, corresponding to quercetin + hexose + deoxyhexose + pentose. Peak 28, present at higher levels in P172-1 than P20-3 (1.4e7 count·s^{-1} vs. 4.0e6 count·s^{-1}) had a mass corresponding to an FCQA isomer (*m/z* 529.3). MS^2 experiments targeting this mass produced ions with *m/z* 367.0 [feruloylquinic acid-$H^+]^-$, *m/z* 349.0 [feruloylquinic acid -H_2O-$H^+]^-$; *m/z* 178.8 [caffeic acid-$H^+]^-$ and *m/z* 135.0 [caffeic acid-CO_2-$H^+]^-$ (Figure 3.10).

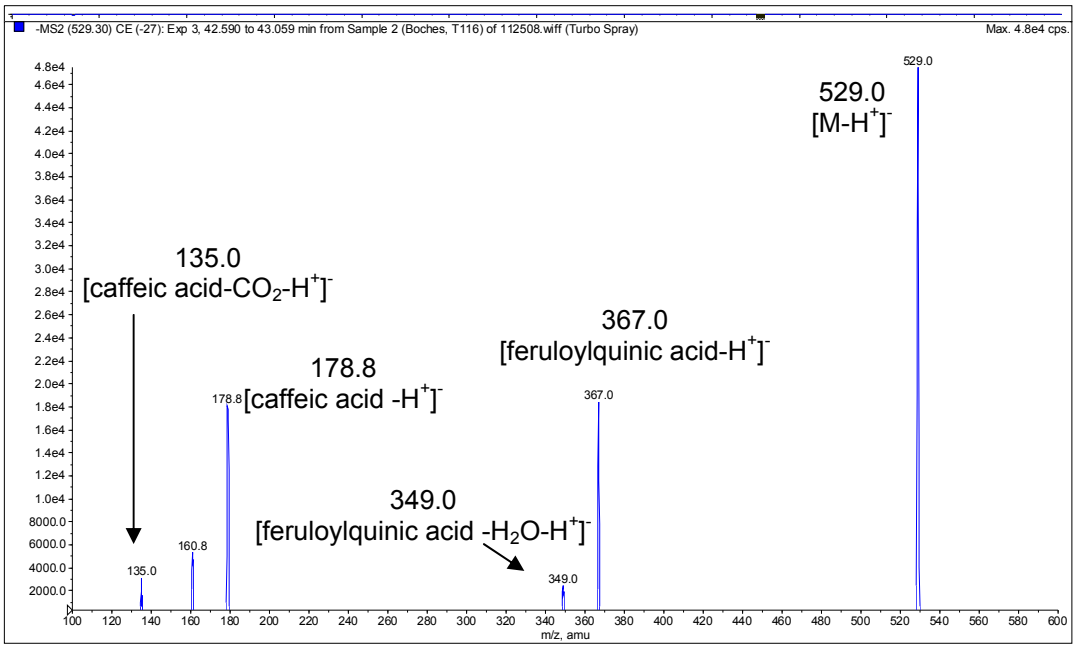


Figure 3.10. MS^2 spectra of putative FCQA (peak 29 in Figure 3.9 and Table 3.3) from peel extract of a P20-3 (*AftAft/atvatv*) individual.

Discussion

Combining regulatory genes that increase flux through the phenylpropanoid pathway leading to anthocyanins (*Aft* and *atv*) with a structural gene encoding a non-functional dihydroflavonol 4-reductase gene (*aw*) resulted in tomatoes with increased levels of chlorogenic acids and quercetin derivatives (i.e. caffeoylquinic acid and rutin) relative to tomato cultivars. In addition to producing anthocyanin, *Aft* tomatoes have been reported to be higher in quercetin and kaempferol than 'Moneymaker' (Sapir et al., 2008). *AftAft/atvatv* double mutants have been shown to be significantly higher in

anthocyanin and total phenolics than *Aft* tomatoes (Mes et al., 2008). The combination of *Aft* with the photomorphogenic *high pigment* mutant *hp1*^{w}, results in increased levels of flavonoids normally found in tomatoes (rutin) as well as the production of numerous polyphenolics not found at significant levels in either parent (van Tuinen et al., 2006). To date, the polyphenolics in *AftAft/atvatv* fruit other than anthocyanins have been minimally characterized. Here we demonstrate that *AftAft/atvatv* tomato fruit contain quercetin derivatives, kaempferol derivatives, and chlorogenic acid derivatives not found at significant levels in normal tomatoes or *Aft*. The addition of the *aw* gene to the *AftAft/atvatv* may have resulted in additional perturbations in the tomato fruit phenolics profile, including an increase in the level of the FCQA, which has previously been reported in coffee (Clifford et al., 2006) and black carrot (Kammerer et al., 2004) but not tomato.

The occurrence of anthocyanin in the fruit of *aa* and *aw* genotypes was unexpected. One possible explanation is that the *aw* and *aa* mutants are leaky alleles of *F3H* and *DFR*, respectively. In the P73 cross (*aaaa* x *AftAft/atvatv*) we observed more purple individuals than expected, (195 vs 181.5), but the significance of this is questionable (p=0.045) and would need to be confirmed in additional segregating populations. No anthocyanin was observed in the leaves or stems of any *aa* or *aw* individuals in the populations, even on individuals producing anthocyanin in the fruit. Since some *AftAft/atvatv* segregants normally produce large amounts of anthocyanin in the leaves, the leaky allele hypothesis seems somewhat unlikely. Another possible explanation is that *F3H* or *DFR* is present in multiple copies which are expressed in a tissue specific fashion. Both *F3H* and *DFR* are reported as single copy in tomato (De Jong et al., 2004, Gouldsbrough et al., 1994). Two *DFR* transcripts are present in tomato however, due to alternate polyadenylation sites (Mathews et al., 2003, Bongue-Bartelsman et al. 1994). The significance of the presence of two *DFR* transcripts is unkown, but might be involved here. Another possibility (which we consider the most likely) is that *ANS* or other enzymes in the phenylpropanoid pathway possess partial *DFR* activity. In *Arabidopsis*, plants carrying TT6 (*AtF3H*) alleles (even ones shown to eliminate enzyme activity *in vivo*) gradually develop light testa color due to low levels of flavonols and condensed tannins (Owens et al., 2008). Owens et al. (2008)

proposed that *ANS* or *FLS* are able to partially substitute for *F3H in vivo*, given that *ANS* and *FLS* have been shown to carry out the *F3H* reaction *in vitro*.

The *AftAft/atvatv/awaw* tomatoes that contained anthocyanin appear to have a different anthocyanin composition than *AftAft/atvatv* tomatoes. The predominant anthocyanin in *Aft* tomatoes is petunidin-3-(p-coumaryl)-rutinoside-5-glucoside, based on HPLC-PDA analysis (Mes et al., 2008). Butelli et al. (2008) identified delphinidin-3-(p-coumaryl)-rutinoside-5-glucoside and petunidin-3-(p-coumaryl)-rutinoside-5-glucoside as the major anthocyanins in transgenic tomatoes expressing anthocyanin in the fruit due to overexpression of snapdragon anthocyanin regulatory genes. Our HPLC-MS results were inconclusive with regard to the anthocyanins found in *AftAft/atvatv* and *AftAft/atvatv/awaw* genotypes. The HPLC-MS and HPLC-PDA conditions we used were optimized for flavonoids and chlorogenic acids, not anthocyanins. When we tried to run HPLC-MS on fruit extracts in positive ionization mode (which is optimal for anthocyanins) there was excessive noise in total ion chromatograms in the form of polymers differing by 44 daltons. In negative ionization mode, peaks with *m/z* 931.5 and 949.3 eluted at the same retention times as anthocyanins (Table 3.2). An *m/z* = 931.5 would correspond to petunidin-3-(coumaroyl)rutinoside-5-glucoside in negative ionization mode, in agreement with the results of both Mes et al. (2008) and Butelli et al. (2008). An *m/z* =949.3 in negative ionization mode doesn't correspond to any anthocyanin masses reported in tomato, but would correspond to petunidin-3-(caffeoyl)rutinoside-5-glucoside (Mes et al., 2008) or delphinidin-3-(feruloyl)rutinoside-5-glucoside (Butelli et al., 2008) in positive ionization mode. When we ran MS^2 on this mass (in negative ionization mode), we detected a fragment with *m/z* 787.8, corresponding to petunidin + caffeic acid + rutinoside or dephinidin + ferulic acid + rutinoside. In HPLC-MS of peel tissue, several peaks with masses corresponding to delphinidin derivatives were more prominent in *AftAft/atvatv/awaw* tomatoes than *AftAft/atvatv* tomatoes (Figure 3.9, Table 3.3). These peaks had elution times later than reported in Ando et al. (1999) or Butelli et al. (2008), however. *Petunia* species with predominantly delphinidin pigments are reddish-orange rather than the red-purple typical of species with mainly petunidin derivatives. If *AftAft/atvatv/awaw* tomatoes do turn out to contain

predominantly dephinidin, it would explain the red color observed in HCl treated fruit of *AftAft/atvatv/awaw* compared to the pink color observed in *AftAft/atvatv* fruit.

It isn't entirely clear why the *AftAft/atvatv* x *aaaa* and *aaaa* x *AftAft/atvatv* populations did not produce very much usable material. One possibility is that it is due to differences in the genetic background of 3-121 (*aw*) vs. LA1194 (*aa*). The *AftAft/atvatv/awaw* segregants had smaller and later maturing fruit than *AftAft/atvatv/aaaa* segregants, and this could have factored into the levels of total phenolics observed. Another possibility is that it is due to differences in the effect of the *aw* and *aa* genes themselves. Whereas the association between *aw* and *DFR* has been demonstrated rigorously, the association between *aa* and *F3H* is tentative. However, if *aa* does corresopond to *F3H* or another biosynthetic gene upstream of *DFR* in the flavonoid pathway, the accumulation of its substrate could be leading to a feedback inhibition effect not observed in *DFR* null genotypes. Under field conditions in 2007 and 2008, *AftAft/atvatv/awaw* segregants appeared vigorous and healthy. They did not appear to be particularly susceptible to diseases or environmental stresses under field conditions, however, the possibility that resistance to stresses was compromised by the loss of anthocyanin production in this material warrants further investigation. Because naringenin derivatives can give bitter flavors in citrus fruits (Frydman et al., 2004), we were concerned that the *AftAft/atvatv/awaw* gene combination might negatively affect tomato flavor. Fruit of *AftAft/atvatv/awaw* F_3 segregants actually had a somewhat bland flavor, but we expect that the flavor could be improved through selection in the *AftAft/atvatv/awaw* background.

References

Aaby, K., Skrede, G., Wrolstad, R.E. 2005. Phenolic composition and antioxidant activities in flesh and achenes of strawberries (*Fragraria ananassa*). J. Agr. Food Chem. 53:4032-4040.

Ando, T., N. Saito, F. Tatsuzawa, T. Kakefuda, K. Yamagake, E. Ohtani, M. Koshi-ishi, Y. Matsusake, H. Kokubun, H. Watanabe, T. Tsukamoto, Y. Ueda, G. Hashimoto, E. Marchesi, K. Asakura, R. Hara, H. Seki. 1999. Floral anthocyanins in wild taxa of *Petunia* (Solanaceae). Biochem. Systematics and Ecol. 27:623-650.

Atanassova, B., and Shtereva, L. 1995. Anthocyaninless mutants- potential for improvement of germination ability in *Lycopersicon esculentum* Mill. Tomato Genet. Coop. Rpt. 45:7-9.

Boches, P. and J.R. Myers. 2007. The *Anthocyanin fruit* tomato gene (*Aft*) is associated with a DNA polymorphism in a MYB transcription factor. HortScience 42:856 (abstr.).

Bongue-Bartelsman M., S.D. Oneill, Y. Tong, and J.I. Yoder. 1994. Characterization of the gene encoding dihydroflavonol-4-reductase in tomato. Gene. 19:327-336.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

De Jong, W.S., N.T. Eannatta, D.M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor. Appl. Genet. 108:423-432.

Fulton, T.M., J. Chugwongse, S.D. and Tanksley. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rptr. 13:207-209.

Giuntini, D., V. Lazzeri, V. Calvenzani, C.D. Asta, G. Galaverna, C. Tonelli, K. Petroni, and A. Ranieri. 2008. Flavonoid profiling and biosynthetic gene expression in flesh and peel of two tomato genotypes grown under UV-B-Depleted conditions during ripening. J. Agric. Food Chem. 56:5905-5915.

Giusti M.M. and Wrolstad R.E., 2001. Unit F1.2.1-13. Characterization and measurement of anthocyanins by UV-visible spectroscopy. In: Current Protocols in Food Analytical Chemistry. R.E. Wrolstad (ed). John Wiley & Sons, NY.

Goldsbrough, A., F. Belzile, and J.I. Yoder. 1994. Complementation of the tomato *anthocyanin without* (*aw*) mutant using the dihydroflavonol 4-reductase gene. Plant Physiol. 105:491-496.

Frydman, A., O. Weisshaus, M. Bar-Peled, D.V. Huhman, L.W. Sumner, F.R. Marin, E. Lewinsohn, R. Fluhr, J. Gressel, and Y. Eval. 2004. Citrus fruit bitter flavors: isolation and functional characterization of the gene Cm1,2RhaT encoding a rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of the citrus. Plant J. 40:88-100.

Ijiima, Y., Y. Nakamura, Y. Ogata, K. Tanaka, S. Sakurai, K. Suda, T Suzuki, H. Suzuki, K. Okazaki, M. Kitayama, S. Kanaya, K. Aoki, D. Shibata. 2008. Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949-962.

Ji, Y. and R. Chetelat. 2003. Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic addition and substitution lines of tomato. Theor. Appl. Genet. 106:979-989.

Jones C.M., P. Mes, J.R. Myers. 2003. Characterization and inheritance of the *Anthocyanin fruit* (*Aft*) tomato. J. Heredity 94:449-456.

Kammerer, D., R. Carle, and A. Schieber. 2004. Characterization of phenolic acids in black carrots (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.) by high-performance liquid chromatography/electrospray ionization mass spectrometry. Rapid Comm. Mass Spec. 18:1331-1340.

Kerckhoffs, L.H.J., N.A.M.A. De Groot, A. Van Tuinen, M.E.L. Schreuder, A. Nagatini, M. Koornneef, and R.E. Kendrick. 1997. Physiological characterization of exaggerated-photoresponse mutants of tomato. J. Plant Physiol. 150:578-587.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56:316-326.

Mathews H., S.K. Klendennen, C.G. Caldwell, X. Liang Liu, K. Connors, N. Matheis, D.. Schuster, D.J. Menasco, W. Wagoner, J. Lightner, and D.R. Wagner. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell. 15:1689-1703.

Mes, P. 2005. Breeding tomatoes for improved antioxidant activity. Ph.D. Dissertation, Oregon State University, Corvallis.

Mes P.J., Boches P., and Myers J.R. 2008. Characterization of tomatoes expressing anthocyanin in the fruit. J. Am. Soc. Hort. Sci. 133:262-269.

Moco, S., R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, and C.H.R. de Vos. 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141:1205-1218.

Moco, S., E. Capanoglu, Y. Tikunov, R.J. Bino, D. Boyacioglu, R.D. Hall, J. Vervoort, and R.C.H. De Vos. 2007. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J. Exp. Bot. 58:4131-4146.

Muir, S.R., Collins, G.J., Robinson, S., Hughes, S., Bovy, A., de Vos, C.H.R., van Tunen, A.J., Verhoeyen, M.E. 2001. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nat. Biotechnol. 19:470-474.

Owens D.K., K.C. Crosby, J. Runac, B.A. Howard, and B.S.J. Winkel. 2008. Biochemical and genetic characterization of *Arabidopsis* flavanone 3β-hydroxylase. Plant Physiol. Biochem. 46:833-843.

Price C.L., and Wrolstad R.E., 1995. Anthocyanin pigments of Royal Okanogan Huckleberry juice. J. Food Sci.60:369-374.

Rick, C.M. 1980. Tomato, p. 669–680. In: Fehr, W.R. and H. Hadley (eds.). Hybridization of crop plants. Amer. Soc. Agron., Crop Sci. Soc. Amer., Madison, WI.

Robbins, R. 2003. Phenolic acids in foods: an overview of analytical methodology. J. Agric. Food Chem. 51:2866-2887.

Rommens, C.M., Richeal, C.M., Yan, H., Navarre, D.A., Ye, J., Krucker, M., Swords, K. 2008. Engineered native pathways for high kaempferol and caffeoylquinate production in potato. Plant Biotechnol. 6:870-886.

Sapir, M., M. Oren-Shamir, R. Ovadia, M. Reuveni, D. Evenor, Y. Tadmor, S. Nahon, H. Shlomo, L. Chen, A. Meir, and I. Levin. 2008. Molecular aspects of *Anthocyanin fruit* tomato in relation to *high pigment-1*. J. Hered. 99:292-303.

Schijlen, E., C.H. Ric de Vos, H. Jonker, H. van den Broeck, J. Molthoff, A. van Tunen, S. Martens, and A. Bovy. 2006. Pathway engineering for phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol. J. 4:433-444.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry. 65:2631-2648.

Singleton, V.L., and Rossi, J.A. 1965. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Amer. J. Enol. Viticult. 16:144-158.

Slimestad, R., T. Fossen, and M.J. Verheul. 2008. The flavonoids of tomatoes. J. Agric. Food Chem. 56:2436-2442.

Tong, Y., and J.Yoder. 1995. The tomato anthocyanin without mutation (*aw*) co-segregates with the dihydroflavonol 4-reductase gene. Tomato Genet. Coop. Rpt. 45:48.

van Tuinen, A., C.H Ric de Vos, R.D. Hall, L.H.W. van der Plas, C. Bowler, and R.J. Bino. 2006. Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from matural light hyperresponsive mutants, p. 339-356. In: P.K. Jaiwal (ed.) Plant Genetic Engineering Vol. 7: Metabolic engineering and molecular farming.

Vinson J, Y. Hao, X. Su, and L. Zubik. 1996. Phenol antioxidant quantity and quality in foods: vegetables. J. Agric. Food Chem. 46:3630-3634.

Waterhouse, A. 2005. Determination of total phenolics. p.I.1.1.1-I.1.1.8 in: R.E. Wrolstad, T.E. Acree, E.A. Decker, M.H. Penner, D.S. Reid, S.J. Schwartz, C.F. Shoemaker, D. Smith, an P. Sporns (eds.). Handbook of Food Analytical Chemistry: v.2. John Wiley & Sons, Hoboken, NJ.

Willits, M.G., Kramer, C.M., Prata, R.T.N., De Luca, V., Potter, B.G., Steffens, J.C., Graser, G. 2005. Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. J. Agric. Food Chem. 53:1231-1236.

CHAPTER 4

Evaluation of a *Solanum lycopersicum* var. *cerasiforme* Core Collection for Phenolics Content

Peter S. Boches, Brooke C. Peterschmidt, and James R. Myers

Abstract

One hundred accessions of *Solanum lycopersicum* var. *cerasiforme* Dunal from the 'Tanksley' designated core collection were evaluated for horticultural quality under greenhouse conditions. Fourteen selected accessions were grown under field conditions in a replicated trial to evaluate leaf and fruit phenolic content. Phenolics content of fruit and leaf tissue was assessed using the Folin-Ciocalteau (F-C) assay and high-performance liquid chromatography (HPLC) combined with photodiode array (HPLC-PDA) and mass spectrometry (HPLC-MS). Total fruit phenolics ranged from 40.05 to 81.85 mg·100 g^{-1} GAE fresh weight (FW) as measured by F-C, and 12.57 to 107.86 mg·100 g^{-1} FW as measured by HPLC-PDA. Five accessions had significantly higher total F-C determined phenolics than check cultivars ($p \leq 0.05$). Marker-trait associations between previously characterized DNA markers and phenolics traits were identified. Five accessions selected for high phenolic content under open field conditions were grown again under greenhouse conditions and re-evaluated for phenolics content. These lines were crossed to elite high anthocyanin and high flavonol tomato lines for the development of high phenolics tomato cultivars.

Introduction

Tomato (*S. lycopersicum* L.) is an important food crop and a significant constituent in the human diet, being the second most consumed vegetable in the U.S. after potatoes (Economic Research Service, USDA, 2008). Because they comprise a large portion of the average vegetable intake, tomatoes are an important source of phenolics in the human diet. Phenolics are a group of secondary plant metabolites that include phenolic acids, anthocyanins, and flavonoids. These compounds are produced only by plants, and exhibit a broad range of health promoting effects when consumed in foods of plant origin. Phenolics have been correlated with anti-inflammatory, anti-viral, and anti-allergenic effects (Kushad et al., 2003; Middleton and Kandaswami, 1993). Phenolics are also strong antioxidants, but recent research suggests that phenolics do not act as antioxidants *in vivo* but rather, exert their health-beneficial effects through other means (Lotito and Frei, 2006; Halliwell et al., 2005; Scalbert et al., 2005).

Compared to many vegetables, tomato phenolic levels are relatively low, ranking 15th in total phenolics of 23 vegetables (Vinson et al., 1998). The highest concentrations of these compounds are found in the skin of the fruit, with lesser amounts found in the flesh (Kushad et al., 2003; Stewart et al., 2000). Considerable effort has been made to increase the levels of phenolics in tomato fruit through the use of genetic engineering (Schijlen et al. 2006, Schijlen et al. 2004) and introgression of genes from wild species (Willits et al., 2005). However, the use of genetic engineering is limited by public concerns over the safety of genetically modified foods, whereas the introgression of novel traits from wild species is often hindered by linkage of undesirable traits.

The feral or wild tomato subspecies *S. l.* var. *cerasiforme* ranges throughout much of South and Central America and represents the closest wild gene pool for the introgression of novel traits into cultivated tomato. Many *S. l. var. cerasiforme* accessions are similar to cultivated tomato, however, this subspecies has previously been reported to contain high levels of total phenolics (Hanson et al., 2004). Recently a set of 100 *S. l.* var. *cerasiforme* accessions were assembled in a core collection with to maximizing sampling of geographical and morphological diversity within the species (S. Tanksley, pers. comm.). These accessions will also be extensively genotyped using molecular markers and the information made publicly available on SolGenes (S. Tanksley, pers. comm.) and selected accessions will be genotyped even more extensively through the Solanaceae CAP project. These accessions could provide a source of high phenolics traits from wild species that would facilitate introgression of these genes with minimal linkage to negative traits. The objective of this research was to evaluate tomato accessions for phenolic content, with the aim of selecting accessions that could be crossed to elite lines for the development of high phenolics tomato cultivars.

Materials & Methods

Plant material. One hundred accessions of *S. l.* var. *cerasiforme* (Table 4.1) were obtained from Steve Tanksley (pers.comm.). Since some of these accessions originate from cross-pollinated, heterozygous populations; Tanksley proposed selfing them to

homozygosity to provide a more permanent genetic resource amenable to genetic analysis. All one hundred accessions were evaluated under greenhouse conditions for horticultural qualities such as earliness and yield in the S_1 generation. Tomato seeds were sown in Sunshine SB40 professional growing mix (Sun Gro Horticulture, Bellevue, WA) in 5-cm-diameter plastic transplant cells. Seedlings were transplanted into 3.8-L pots filled with Special Blend potting mix and 20 g of 14N–6.1P–11.6K slow-release fertilizer (Simplot 14–14–14) after 3 weeks. Plants were given 10 g of supplemental fertilizer after 4 weeks in the pots. Greenhouse temperature was set to 18°C night and 25°C day. Supplemental lighting during was provided for 16 hours per day by 400-W metal halide and 400-W high-pressure sodium Sun System 3 high-intensity discharge lamps (Sunlight Supply, Woodland, WA). Fourteen accessions with good horticultural qualities were grown at the OSU Vegetable Research Farm, Corvallis, Oregon during the summer of 2007 in the S_2 generation. 'Gold Nugget' (Baggett et al., 1985), 'Siletz' (Baggett et al., 1998), and LA 1996, a high flavonoid introgression line that expresses the *Anthocyanin fruit (Aft)* gene (Jones et al., 2003); were included as standards for low, medium, and high phenolics tomato lines, respectively. Three replications of five plants each were arranged in a randomized complete block design. Plants were transplanted May 21 into rows 1.8 m apart with 60 cm within-row spacing. Five-hundred five $kg \cdot ha^{-1}$ of 12N-29P-10K-4S fertilizer was banded before transplanting. Plants were irrigated at weekly intervals until mid-August when water was withheld. All leaf samples (one sample per plot) for phenolics analysis were harvested on June 22, from the uppermost fully expanded leaf on a randomly selected plant. Fruit samples (~500 g per plot) were harvested as fruit reached maturity in mid August. Based on the results of the field trial, the five accessions with the highest phenolics content and horticultural quality were grown under greenhouse conditions in the S_3 generation. Three or more replicates of one plant each were grown under the same greenhouse conditions as before and re-evaluated again for phenolics content (one sample per plant). 'Legend', 'Gold Nugget', and LA 1996 were used as standards in the second greenhouse evaluation.

Table 4.1. *S. l.* var. *cerasiforme* germplasm screened for horticultural quality under greenhouse conditions. Accession numbers from the Tomato Germplasm Resource Center (TGRC) in Davis, CA were obtained as S_1 selfs from the Tanksley lab.

TGRC Accession	Tanksley ID	Origin
LA0292	05T859	Galapagos Islands, Ecuador
LA1203	05T801	Ciudad Vieja, Guatemala
LA1204	05T860	Quetzaltenango, Guatemala
LA1205	05T861	Copan, Honduras
LA1206	05T862	Copan Ruins, Honduras
LA1208	05T802	Sierra Nevada, Columbia
LA1226	05T863	Morona-Santiago, Columbia
LA1228	05T864	Morona-Santiago, Columbia
LA1231	05T865	Napo, Ecuador
LA1247	05T803	Loja, Ecuador
LA1268	05T866	Lima, Peru
LA1286	05T867	Junin, Peru
LA1290	05T804	Junin, Peru
LA1307	05T868	Ayacucho, Peru
LA1312	05T869	Cusco, Peru
LA1314	05T805	Cusco, Peru
LA1320	05T870	Apurimac, Peru
LA1323	05T871	Cusco, Peru
LA1324	05T872	Cusco, Peru
LA1334	05T873	Arequipa, Peru
LA1338	05T806	Napo, Ecuador
LA1372	05T874	Lima, Peru
LA1385	05T807	Cusco, Peru
LA1386	05T808	Amazonas, Peru
LA1388	05T875	Junin, Peru
LA1420	05T876	Napo, Ecuador
LA1423	05T809	Pichincha, Ecuador
LA1425	05T877	Cauca, Colombia
LA1429	05T878	Manabi, Ecuador
LA1455	05T879	Nuevo Leon, Mexico
LA1456	05T810	Vera Cruz, Mexico
LA1457	05T811	Puebla, Mexico
LA1464	05T812	El Progreso, Yoro, Honduras
LA1511	05T813	Minas Gerais, Brazil
LA1512	05T814	Lago do Llopango, El Salvador
LA1542	05T880	Turrialba, Costa Rica
LA1543	05T815	Upper Parana, Brazil
LA1545	05T881	Campeche, Mexico
LA1549	05T882	Pasco, Peru
LA1569	05T816	Vera Cruz, Mexico
LA1574	05T883	Lima, Peru
LA1620	05T817	Bahia, Brazil
LA1621	05T885	Hidalgo, Mexico
LA1623	05T818	Yucatan, Mexico
LA1632	05T886	Madre de Dios, Peru
LA1654	05T819	San Martin, Peru
LA1668	05T820	Guerrero, Mexico

Table 4.1 continued…

TGRC Accession	Tanksley ID	Origin
LA1701	05T821	La Libertad, Peru
LA1703	05T822	Tamaulipas, Mexico
LA1711	05T887	Zamorano, Honduras
LA1712	05T888	Pejibaye, Costa Rica
LA1909	05T823	Cusco, Peru
LA1953	05T824	Arequipa, Peru
LA2076	05T825	Naranjitos, Bolivia
LA2077	05T826	La Paz, Bolivia
LA2078	05T827	Rio Grande de Sol, Brazil
LA2095	05T889	Loja Ecuador
LA2121	05T828	Zamora-Chinchipe, Ecuador
LA2126	05T829	Zamora-Chinchipe, Ecuador
LA2131	05T890	Zamora-Chinchipe, Ecuador
LA2135	05T830	Santiago-Morona, Ecuador
LA2136	05T831	Santiago-Morona, Ecuador
LA2137	05T832	Santiago-Morona, Ecuador
LA2177	05T833	Cajamarca, Peru
LA2308	05T891	San Martin, Peru
LA2312	05T834	Amazonas, Peru
LA2393	05T835	Guanacaste, Costa Rica
LA2394	05T836	Guanacaste, Costa Rica
LA2402	05T837	Santa Catarina, Brazil
LA2411	05T838	Puno, Peru
LA2616	05T892	Huanuco, Peru
LA2619	05T893	Loreto, Peru
LA2626	05T839	Cusco, Peru
LA2628	05T840	Cusco, Peru
LA2632	05T841	Cusco, Peru
LA2633	05T842	Cusco, Peru
LA2640	05T843	Apurimac, Peru
LA2660	05T844	Beni, Bolivia
LA2664	05T894	Puno, Peru
LA2670	05T895	Puno, Peru
LA2675	05T896	Puno, Peru
LA2683	05T845	Cusco, Peru
LA2688	05T897	Madre de Dios, Peru
LA2690	05T846	Cusco, Peru
LA2697	05T847	Valle, Colombia
LA2698	05T848	Valle, Colombia
LA2710	05T849	Porto Firme, Colombia
LA2783	05T850	Antioquia, Colombia
LA2792	05T851	Antioquia, Colombia
LA2793	05T852	Antioquia, Colombia

Table 4.1 continued…

TGRC Accession	Tanksley ID	Origin
LA2844	05T853	San Martin, Peru
LA2845	05T898	San Martin, Peru
LA2871	05T899	Sud Yungas, Bolivia
LA3136	05T854	Holguin, Cuba
LA3137	05T855	Holguin, Cuba
LA3141	05T856	La Habana, Cuba
LA3158	05T857	Sinaloa, Mexico
LA3623	05T858	Manabi, Ecuador
LA3652	05T900	Apurimac, Peru

Samples and Chemicals. Leaf phenolics were extracted from three weighed leaf discs (approximately 40-60 mg fresh weight) by grinding the fresh leaves in acidified methanol with a micro-pestle. Fruit were cut in half lengthwise, and seeds placenta, and columella tissue were removed. The remaining pericarp and epidermal tissue was frozen with liquid nitrogen and ground to a fine powder using a mortar and pestle. Weighed portions of this powder were used for extractions. For field samples, 70-150 mg of the tomato powder was extracted overnight in 300 µL of acidified methanol (1% HCl) overnight at 4° C in 1.5 mL Eppendorf tubes. Extractions were brought to 500 µL with nanopure H_20, vortexed with 500 µL chloroform, and centrifuged at 13,000 rpm for ten minutes. The aqueous supernatant was removed and stored at -20° C for further analysis. The greenhouse grown samples were processed using a scaled up version of this protocol (1 g of powdered tissue was extracted in 5 mL 60:40 1% HCl methanol: water and 5 mL chloroform in 15 mL glass tubes). Gallic acid, rutin, 5-caffeoylquinic acid, and naringenin purchased from Sigma Chemical Corp (St Louis, MO) were used as standards for F-C asssys and HPLC.

F-C analysis. F-C assays were performed according to the method of Singleton and Rossi (1965) and Waterhouse (2005), with modifications. F-C reactions were performed on a micro-scale: each reaction contained 198 µL water, 12.5 µL F-C reagent (Sigma), 2.5 µL of each sample or gallic acid standard, and 37.5 µL sodium carbonate solution. Reactions were incubated for 1 hour at room temperature, and absorbance at 765 nm was read using a SPECTRAmax microplate spectrophotometer

(Molecular Devices, Sunnyvale, CA, USA). F-C values were expressed as gallic acid equivalents (GAE) based on a gallic acid standard curve in the range of 0-500 $mg \cdot L^{-1}$.

HPLC analysis. For HPLC analysis of the field samples collected in 2007, the three samples from different plots of each line were consolidated into a single sample. For HPLC coupled to photodiode array detection (HPLC-PDA) a Hewlett-Packard (Palo Alto, CA) 1050 series auto sampler and pump coupled to a HP 1040M diode array detector detection system was used. An ES Industries (West Berlin, NJ) reversed phase LiChrospher RP18 endcapped column (25 cm x 4.6 mm, 5 μm particle size) was used with a guard column of the same material. The separation gradient and buffers followed that of the tomato metabolite database (Moco et al., 2006). Solvent A was formic acid:water (1:1000). Solvent B was formic acid:acetonitrile (1:1000). The flow rate was 1 $mL \cdot min^{-1}$, and the injection volume was 40 μL. The elution conditions were as follows: 0-45 min, linear gradient from 5% B to 35% B; 45-50 minutes, isocratic at 85% B to wash the column; 50-60 minutes, isocratic at 5% B to equilibrate the column. Peak spectra were monitored at 280 nm. UV spectra were recorded from 200 to 600 nm in 0.5 nm intervals. HPLC data were analyzed using Agilent ChemStation Rev.A.09.01 (Santa Clara, CA). Compounds were quantified as equivalents to naringenin, chlorogenic acid, or rutin using a standard curve and summarized according to compound classes. Standards were used in the range of 25-250 $mg \cdot L^{-1}$ for rutin, 5-50 $mg \cdot L^{-1}$ for naringenin, and 25-500 $mg \cdot L^{-1}$ for chlorogenic acid. HPLC-Mass Spectrometry (HPLC-MS) was performed using the same column, buffers, and elution conditions on a Shimadzu LC-20-AD chromatography system (Columbia, MD) coupled to an ABI 3200 QTRAP LC/MS/MS (Foster City, CA) triple quadropole mass spectrometer equipped with a Turbo V ionization source and connected to Analyst software V1.4.2. Acquisition was performed in Q1 MS scan mode monitoring *m/z* from 200 to 1000, 1 scan 2 s^{-1}. Nitrogen was used for the ion source gas and collision gas for MS/MS. The source and nebulizer temperature was set to 650°C, capillary charge = -4500 V, cone voltage = 45 V, entrance potential = -10 V.

Statistical Analysis. SAS software (Cary, NC) was used for analysis of variance and means separations using Fisher's Least Significant Difference (LSD). PowerMarker (Liu and Muse, 2005) and TASSEL (Bradbury et al., 2007) software were used for association mapping using single locus F-tests.

Results

A wide range of fruit and plant types was observed among *S. l.* var *cerasiforme* accessions. Since the main objective of this study was to select *S. l.* var. *cerasiforme* accessions for use in breeding a high phenolics or high anthocyanin cherry tomato cultivar, the choice of which accessions to evaluate for phenolics content was initially based on several horticultural characteristics. The vast majority of *S. l.* var *cerasiforme* accessions evaluated had extremely late maturity (>120 days) or low yield (measured as the quantity of fruit relative to vegetation) relative to cultivated tomatoes. Other undesirable qualities for our objectives included inflorescences that resumed vegetative growth, mealy fruit, seedy fruit, thick skinned fruit, pale fruit color, and larger fruit. Accessions were evaluated for these characters (Table 4.2) and the 14 best accessions were selected for further evaluation in a replicated field trial. In many cases, early maturing accessions were low yielding, or high yielding accessions had poor fruit quality, so the final decision was made subjectively.

The 14 selected accessions were evaluated further for horticultural quality (including flavor), in the field trial (Table 4.3). Fungal diseases including powdery mildew (*Oidium lycopersicum*) and gray mold (*Botrytis cinerea*) were a problem in some accessions and strong genotype dependent responses were observed. Some of the best accessions in terms of horticultural quality (for a cherry tomato) were LA1620, LA1549, LA 1668, LA1512, LA 2633 and LA1455.

Table 4.2. Greenhouse notes for *S. l.* var *cerasiforme* accessions with maturity <120 days. Bold indicates the accession was selected for further evaluation of phenolics content in replicated field trials. Scale used : Maturity, 1= early, 5= late; Yield, 1=low, 5=high.

TGRC Accession	Maturity	Yield	Notes
LA1208	4	3	Small red plum.
LA1290	**4**	**3**	**Large cherry, good fruit quality.**
LA1314	4	3	
LA1338	**4**	**2**	**Shiny fruit.**
LA1455	**4**	**1**	**Currant.**
LA1457	5	2	Small cherry, inflorescence indeterminate, poor fruit quality.
LA1464	4	2.5	Blocky plum, green shoulder.
LA1512	**5**	**2**	**Very small currant.**
LA1542	4	2	
LA1545	**4**	**2**	
LA1549	5	2	Fruit small cherry, round. Inflorescence indeterminate, continues to flower after fruit ripen. Seedy.
LA1569	2	2.5	
LA1620	**2**	**2.5**	**Shiny fruit.**
LA1621	4	2	
LA1623	4	2	Small cherry, poor fruit quality.
LA1632	4	3.5	Large cherry, good fruit quality.
LA1654	4	3	
LA1668	**5**	**2**	**Small cherry, good fruit quality and firmness.**
LA1701	4	3	Plum.
LA1703	4	1	Yellow fruit, poor fruit quality.
LA1712	**4**	**1**	**Small cherry, good fruit quality.**
LA1953	4	2.5	
LA2076	**2**	**5**	**Large cherry.**
LA2077	2	5	Large cherry, some 3 loculed fruit, indeterminate inflorescence.
LA2078	4	4	
LA2131	4	5	'Brandywine' fruit shape.
LA2135	**4**	**3**	**Cherry.**
LA2137	5	2	
LA2308	**4**	**4**	**Large cherry.**
LA2312	4	3	Plum.
LA2393	4	2	Small cherry.
LA2411	4	3	
LA2626	**5**	**2**	**Very small, pubescent fruit.**
LA2632	**4**	**2**	**Shiny, pubescent fruit.**
LA2633	**3**	**4**	**Shiny, small cherry with pubescent fruit.**
LA2660	4	4	'Marmande' shape.
LA2664	4	3	Bad fruit quality, compact growth habit, very firm fruit.
LA2675	4	2	
LA2792	4	2	Small cherry, bright crimson fruit.
LA3136	4	1	Poor quality.
LA3137	4	2	

Table 4.3. Field notes for selected *S. l.* var *cerasiforme* accessions. A complex of fungal diseases including powdery mildew (*Oidium lycopersicum*) and gray mold (*Botrytis cinerea*) was observed in many accessions. Fruit weight is measured as the average of 10 fruit.

TGRC Accession	Field Notes	Average fruit weight (g)
LA1290	Larger fruit, watery flavor. Good yield. Variable fruit shape.	14.2
LA1338	Early, very dark red fruit. Very poor fungal disease resistance. Horrible flavor (possibly due to fungal infection?).	7.1
LA1455	Sweet fruit. Upright growth habit. Thick skin. Good yield and yield architecture. Flavor ok. Fruit with glandular trichomes.	3.3
LA1512	Glandular trichomes on fruit, thin skin. Late maturity.	3.8
LA1549	Good yield and fungal disease resistance. Firm fruit, flavor ok.	4.3
LA1620	Good yield and fungal disease resistance. Relatively good fruit quality, excessive gel on seeds, earlier than LA1549.	5.9
LA1668	High yield. Upright growth habit. Sweet, good quality fruit. Thin skin.	4
LA1712	Very acidic. Poor fungal disease resistance. Low to medium yield. Late maturity.	5
LA2076		17.4
LA2135	Early, poor resistance to fungal diseases. Medium to poor fruit quality. Medium size, lower yielding.	13.3
LA2308	Large fruit. Medium to good yield. Bitter flavor, very astringent.	23.6
LA2626	Bad taste (astringent or causing a prickly sensation in throat). Poor fungal disease resistance. Good yield, early maturing.	6
LA2632	Very small pubescent fruit. Poor flavor (bland). Firm pericarp.	2.1
LA2633	Hard skin, very sweet, high solids. Later maturity, good yield, long fruit trusses.	2.3

Six basic patterns of UV-spectra were found in HPLC-PDA chromatograms from fruit extracts of field grown cultivars and *S. l.* var. *cerasiforme* accessions (Figure 4.1). Pattern A was typical of chlorogenic acids with a peak from 322-328 nm and a shoulder around 290 nm. Patterns B and C had peaks or shoulders around 290 nm and 310 nm as has been reported for coumaric, caffeic, or ferulic acid derivatives (Moco et al., 2007; Alonso-Salces et al., 2009). Pattern D had peaks around 256 nm and 354 nm, typical of quercetin and kaempferol derivatives. Pattern E had a peak around 286 nm and a shoulder around 330 nm as has been reported for phloretin and other naringenin derivatives (Slimestad et al. 2008; Iijima et al., 2008). Pattern F was typical of naringenin chalcone (Slimestad et al. 2008).

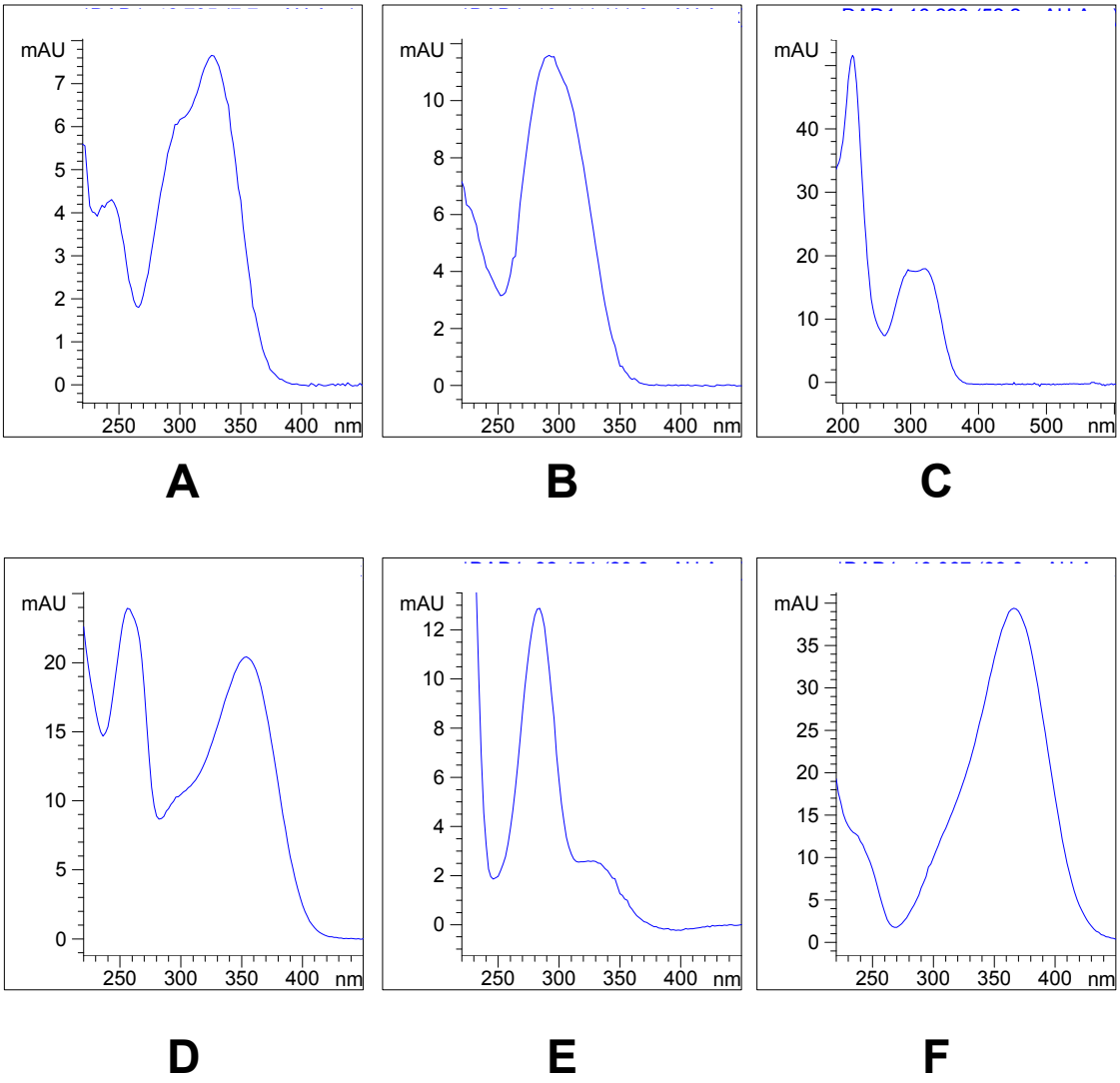


Figure 4.1. UV spectra typical of phenolic compounds identified in fruit extracts of *S. l.* var *cerasiforme* accessions and tomato cultivars. A, caffeic acid hexoses, mono-, di-, and tri- acyl quinic acids; B, coumaric acid containing peaks; C, unidentified hydroxycinnamic acids; D, quercetin and kaempferol derivatives; E, chalcone derivatives; F, naringenin chalcone.

Peak assignments (Figures 4.2 & 4.3, Table 4.4) were made using a combination peak retention times, UV-Spectra, and mass spectra. Mass spectra and UV-spectra were acquired on separate instruments. Mass spectral information was only acquired for the genotypes 'Legend' and LA2633. Retention times for major tomato phenolic compounds (e.g. chlorogenic acid, rutin, and naringenin chalcone) were approximately 2 minutes later on the HPLC-MS system than on the HPLC-PDA

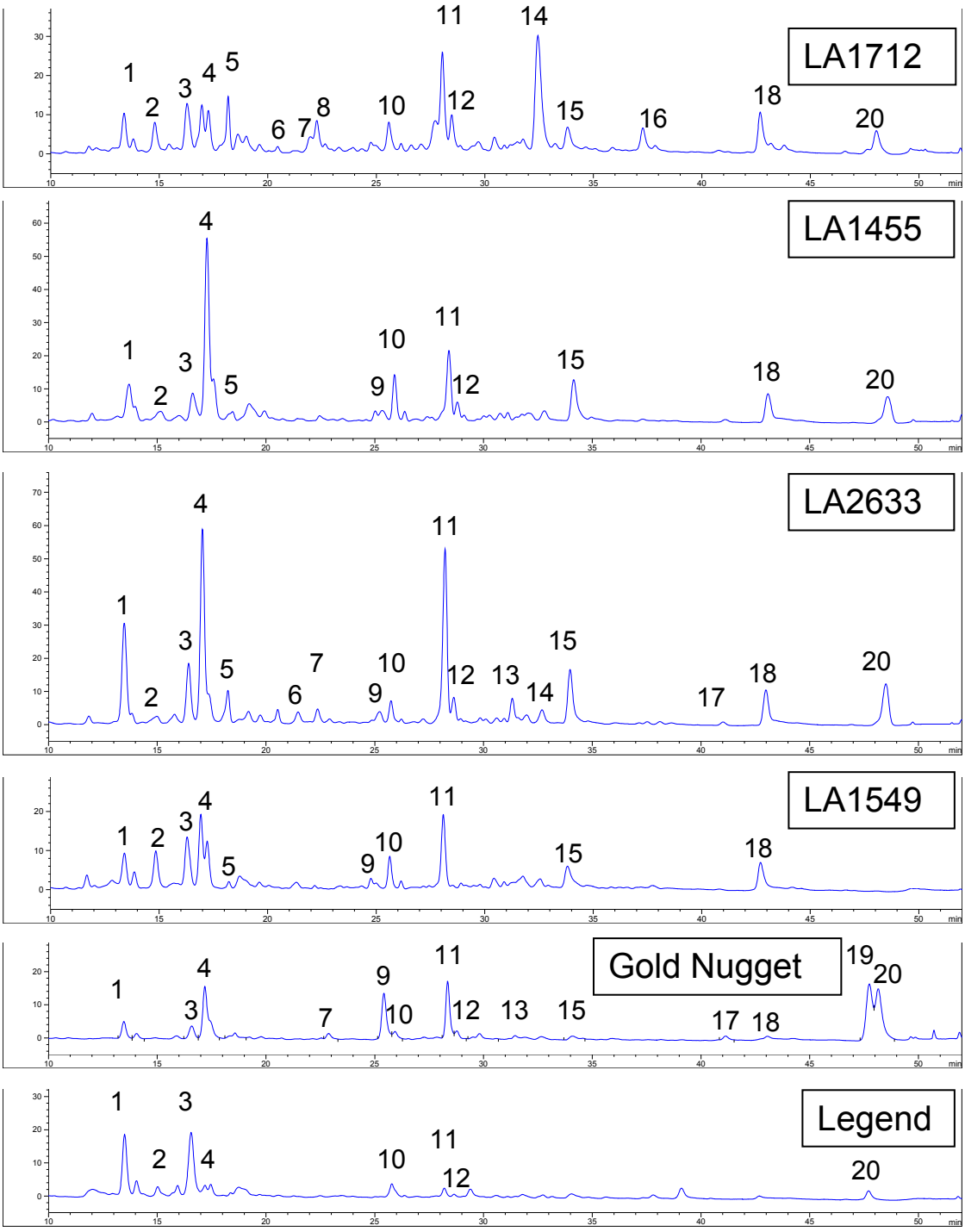


Figure 4.2. HPLC-PDA chromatograms (280 nm) of selected *S. l.* var. *cerasiforme* accessions from greenhouse grown fruit. For peak assignments, see Table 4.4. All chromatograms to same scale. Units on Y axis: mAu at 280nm.

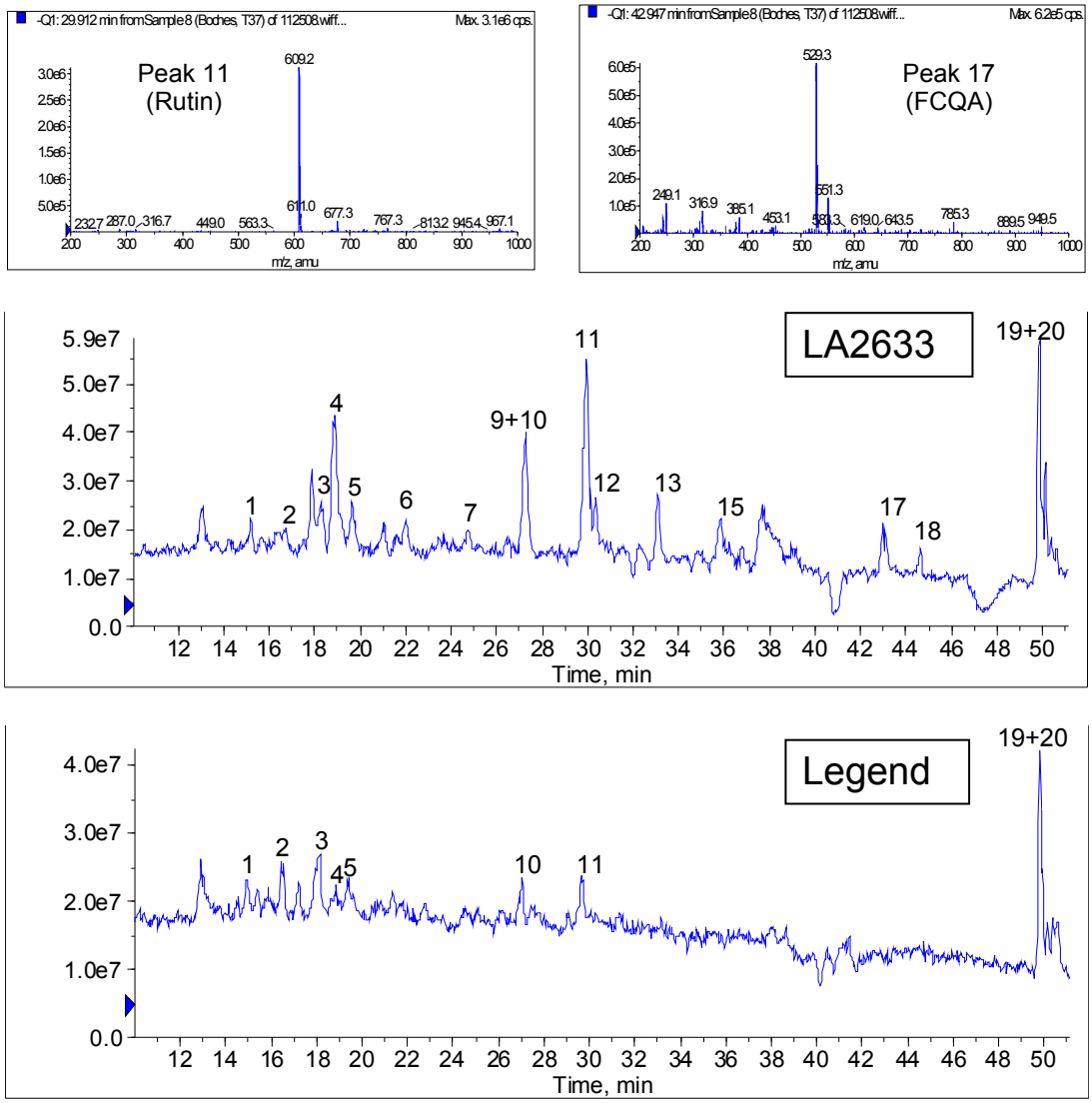


Figure 4.3. Total ion chromatograms of LA2633 and 'Legend' (lower two panels) and accurate mass spectra of peaks 11 and 17 (upper two panels) from LC-MS analysis. Peak assignments are as for Figure 4.2 and Table 4.4. (Peaks corresponding to 8,14, and 16 were not detected by LC-MS).

Table 4.4. Tomato metabolites putatively identified by HPLC-PDA and HPLC-MS. Peaks (identified in Figures 4.2 and 4.3), retention time (RT) under HPLC-PDA conditions, Mass ($[M-H]^-$), putative ID (footnotes refer to literature references), observed UV_{max}, and references from the literature. Abbreviations: na, not available; nd, uv absorbance not detected; sh, shoulder of UV_{max}.

Peak	RT	Mass	Putative ID	UV_{max}
1	13.5	341.0, 325.0	Caffeic acid hexose/ coumaric acid hexose[1]	292, 310sh
2	15.17	341.0	Caffeic acid hexose[1]	328, 298sh
3	16.45	341.0, 325.0	Caffeic acid hexose/ coumaric acid hexose[1]	292
4	17.08	353.5	4- and 5- caffeoylquinic acid[1]	326, 298sh; 320, 290sh
5	18.25	771.3	Quercetin dihexose-deoxyhexose (Quercetin 3-O-Rutinoside-7-O-Glucoside)[1,2]	256, 352
6	20.53	755.4	Kaempferol dihexose-deoxyhexose[1,2]	264, 344
7	21.48	367.0	Feruloylquinic acid[1,3,4]	328, 298sh
8	22.26	na	Chalcone derivative	286, 350sh
9	25.23	367.0	Feruloylquinic acid [1,3]	326
10	25.74	741.3	Quercetin 3-(2"Apiosyl-6"rhamnosylglucoside), rutin pentoside, Quercetin hexose-deoxyhexose pentose[2,5,6]	256, 350
11	28.23	609.2	Rutin[1,2,5,6]	256, 354
12	28.6	597.2	Phloretin 3',5'-Di-C-glucoside[6]	286, 330sh
13	31.31	593.2	Kaempferol 3-(6"Rhamnosylglucoside)[2,6,7]	264, 344
14	32.44	na	Chalcone derivative	284, 326sh
15	33.98	515.3	Dicaffeoylquinic acid[1,3]	326, 298sh
16	37.28	na	Chalcone derivative	284, 326sh
17	41.07	529.3	Feruloylcaffeoylquinic Acid (FCQA)[3,4]	n.d.
18	42.99	677.3	Tricaffeoylquinic acid[1,3]	330, 294sh
19	47.74	na	Naringenin[1,2,6,7]	288, 344sh
20	48.5	271.0	Naringenin chalcone[1,2,6,7]	366

[1]Moco et al., 2006; [2]Ijiima et al., 2008; [3]Clifford et al., 2006; [4]Kammerer et al., 2004; [5]Giuntini et al., 2008; [6]Slimestad et al., 2008; [7]Moco et al. 2007;

system, but the chromatograms showed relatively close alignment (Figures 4.2 & 4.3). Twenty peaks were putatively identified using a combination of UV- and accurate mass spectra (Table 4.4). In addition, numerous minor peaks could be identified only to a general class of compounds based on UV-spectra. HPLC-PDA chromatograms (Figure 4.2) revealed that *S. l.* var. *cerasforme* accessions contained many compounds that were not detected or detected in only trace quantities in cultivars (e.g. peaks 6, 8, 13, 14, 15, 16, 17, and 18). Based on peak identifications in Table 4.4 and/or peak assignments to compound class, phenolic content was summarized by compound class (coumaric and caffeic acid hexoses, peaks 1, 2, and 3; caffeoylquinic acids, peak 4; feruloylquinic acids, peaks 7 and 9; diacylquinic acids, peaks 15 and 17, tricaffeoylquinic acids, peak 18; unkown chlorogenic acids, various un-assigned

peaks; kaempferol derivatives, peaks 6 and 13; quercetin derivatives, peaks 5 and 10; rutin, peak 11; chalcone derivatives, peaks 8, 12, 14 and 16 plus unassigned minor peaks; naringenin and naringenin chalcone, peaks 19 and 20) and quantitated using a standard curve from a similar compound (Table 4.5 and Figure 4.4).

Standard curves were linear over the range sampled ($R^2 \geq 0.9816$). HPLC-PDA data for field grown material (Table 4.5) and greenhouse grown material (Figure 4.4) are presented separately. Total phenolics content of fruit and leaf tissue for field grown material are presented in Table 4.5. Five accessions (LA1455, LA1668, LA1712, LA2632, and LA2633) had significantly higher fruit phenolics than cultivars as measured by the F-C assay (total phenolics F-C) (Table 4.5). Only LA1455, LA1712, and LA2633 had significantly higher total phenolics than LA1996 (Table 4.5). The five accessions with the highest total phenolics as determined by HPLC (total phenolics HPLC) were LA 2633, LA1996, LA1455, LA1712, and LA1620. There was a modest but statistically significant negative correlation between total phenolics (F-C) and fruit weight ($r = -0.49$, $P = 0.046$). Total phenolics (F-C) in fruit tissue and total phenolics (F-C) in leaf tissue were correlated, but was not statistically significant ($R = 0.67$, $P = 0.065$).

Five accessions (LA1455, LA1549, LA1620, LA1712, LA2633) were selected for crossing to elite tomato lines and further evaluation in the greenhouse. The decision of which accessions to carry forward was based on a combination of their F-C total phenolics scores (Table 4.5) and their horticultural quality (Table 4.3).

Table 4.5. Phenolics content of *S. l.* var *cerasiforme* accessions and cultivated tomatoes leaf and fruit extracts from 2007 field trial as measured by HPLC-PDA and F-C assay. Coumaric and caffeic acid hexoses (Cou./Caf.), caffeoylquinic acids (CQA), feruloylquinic acids (FQA), Diacylquinic acids (Di.), tricaffeoylquinic acids (Tri.), and unidentifed cinnamic acids (Unk.) are expressed as chlorogenic acid equivalents (Eq.). Kaempferol derivatives (Kaem.), miscellaneous quercetin derivatives (Quer.), and rutin are expressed as rutin equivalents. Chalcone (Chalc.) derivatives are expressed as equivalents to naringenin (Nar.). Total phenolics as measured by F-C assays are expressed as gallic acid equivalents and presented as the average of three biological replicates ± standard error (SE). Extracts from biological replicates were consoidated for HPLC.

	HPLC-PDA											Folin-Ciocalteau				
	Chlorogenic Acid Eq.						Rutin Eq.			Nar. Eq.		Gallic Acid Eq.				
	Cou./ Caf.	CQA	FQA	Di.	Tri.	Unk.	Kaem.	Quer.	Rutin	Chalc.	total HPLC	Mean GAE (Fruit)	SE		Mean GAE (Leaf)	SE
	mg·100g^{-1} Fresh Weight															
Gold Nugget	2.23	2.91	2.34	2.00	0.00	0.58	0.00	6.31	16.52	1.07	33.96	42.31 fg[z]	±	10.63	587.73 bc	± 50.51
LA1290	4.00	1.27	0.91	0.00	0.00	0.91	0.00	1.67	4.00	0.57	13.32	45.43 defg	±	0.84	468.77 bcd	± 37.46
LA1338	17.74	1.55	0.75	1.68	0.66	0.38	0.00	5.86	7.20	0.00	35.81	57.18 bcdefg	±	17.35	472.17 bcd	± 39.11
LA1455	8.27	6.19	3.71	0.90	0.71	5.68	0.00	13.55	17.80	1.19	58.00	75.53 ab	±	8.67	654.53 bc	± 88.38
LA1512	8.47	1.90	1.15	0.30	0.00	0.00	0.00	3.24	6.70	0.62	22.37	51.62 cdefg	±	4.80	482.46 bcd	± 25.58
LA1549	6.30	6.37	4.01	0.39	0.40	0.00	0.00	9.38	17.39	0.00	44.25	63.21 abcdefg	±	2.98	632.70 bc	± 57.06
LA1620	12.31	5.56	3.45	1.31	0.52	1.12	0.00	6.21	12.45	1.38	44.33	63.64 abcdef	±	6.50	482.93 bcd	± 52.25
LA1668	5.34	6.39	3.50	1.11	0.43	0.33	0.00	6.74	14.73	1.15	39.72	68.21 abcd	±	3.97	619.33 bc	± 47.25
LA1712	5.56	5.45	3.04	1.17	0.49	1.34	0.00	10.18	16.49	1.26	44.99	81.85 a	±	15.42	539.00 bcd	± 47.27
LA1996	40.15	6.57	1.89	1.17	0.00	3.78	0.00	3.40	10.48	0.00	67.43	48.97 defg	±	8.11	684.26 b	± 134.73
LA2076	2.81	3.24	2.09	0.00	0.00	0.00	0.00	4.42	7.68	0.44	20.67	48.91 defg	±	3.17	440.50 cd	± 13.44
LA2135	9.66	2.96	1.75	0.00	0.00	0.00	0.00	4.05	5.29	0.36	24.08	60.37 abcdefg	±	9.02	545.70 bcd	± 85.72
LA2308	7.82	3.41	1.79	0.00	0.00	1.16	0.00	4.57	3.43	0.00	22.18	44.20 efg	±	1.69	322.76 d	± 246.71
LA2626	4.01	2.28	1.54	0.00	0.00	0.00	0.00	6.95	15.94	0.87	31.59	51.90 cdefg	±	7.24	979.80 a	± 125.99
LA2632	3.85	6.98	4.58	0.71	0.00	0.00	0.00	9.67	10.78	1.30	37.88	67.53 abcde	±	4.37	602.20 bc	± 51.35
LA2633	6.74	8.31	5.41	1.53	0.00	5.58	5.99	41.36	30.87	2.07	107.86	73.99 abc	±	3.30	674.74 b	± 61.29
Siletz	6.84	1.31	1.79	0.00	0.00	0.00	0.00	1.27	1.36	0.00	12.57	40.05 g	±	5.18	639.36 bc	± 69.21

[z]Any two means in a column not followed by the same letter are significantly different at $p<0.05$, Fisher's Least Significant Difference.

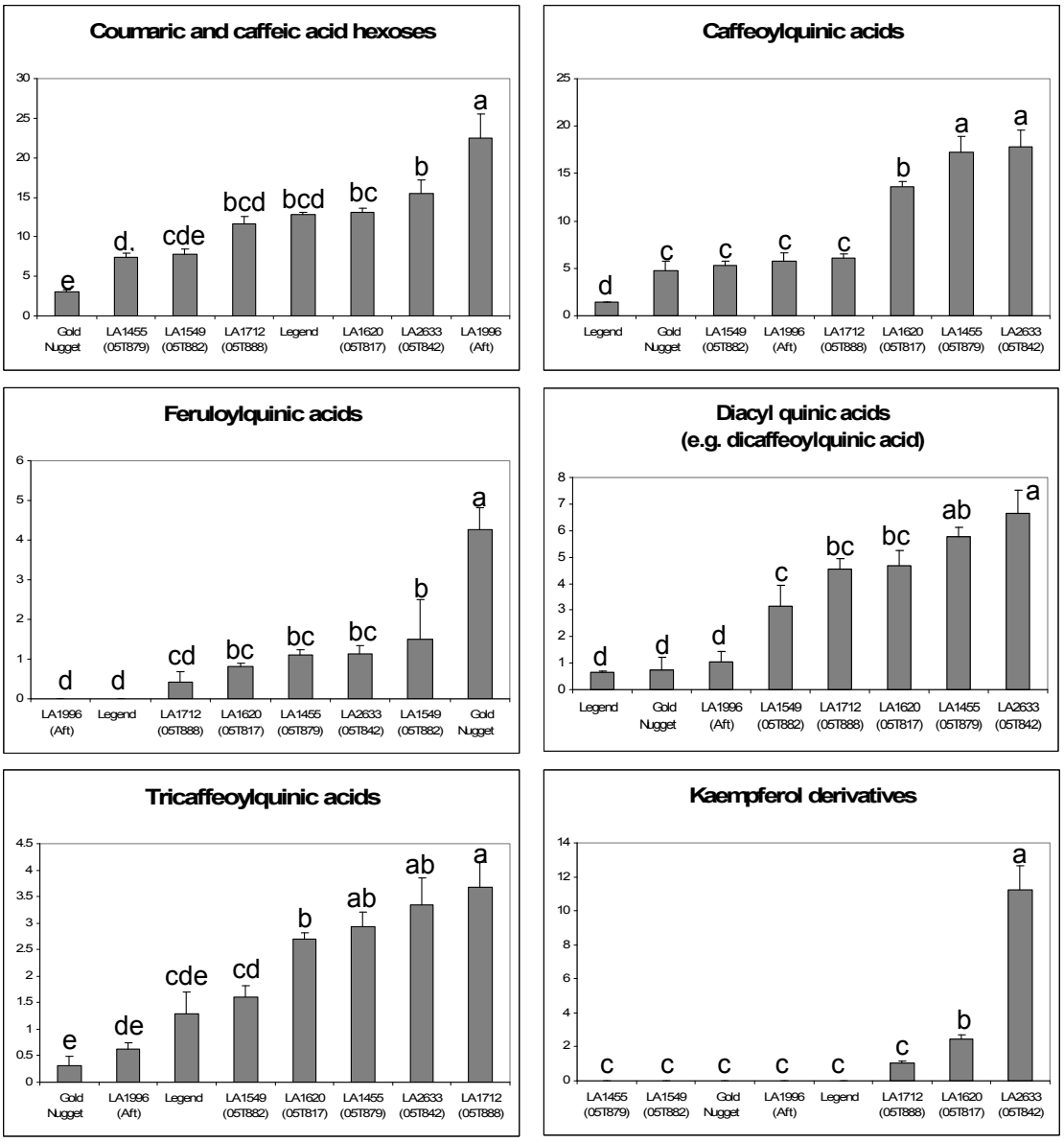


Figure 4.4. Phenolic content (mg·100 g^{-1} FW, measured by HPLC-PDA) of *S. l.* var *cerasiforme* accessions, as verified in second greenhouse trial, following field trials. Values are the mean of three or more biological replicates, error bars = one standard error. Means sharing the same letter(s) are not significantly different as determined by Fisher's Least Significant Difference, $P \leq 0.05$.

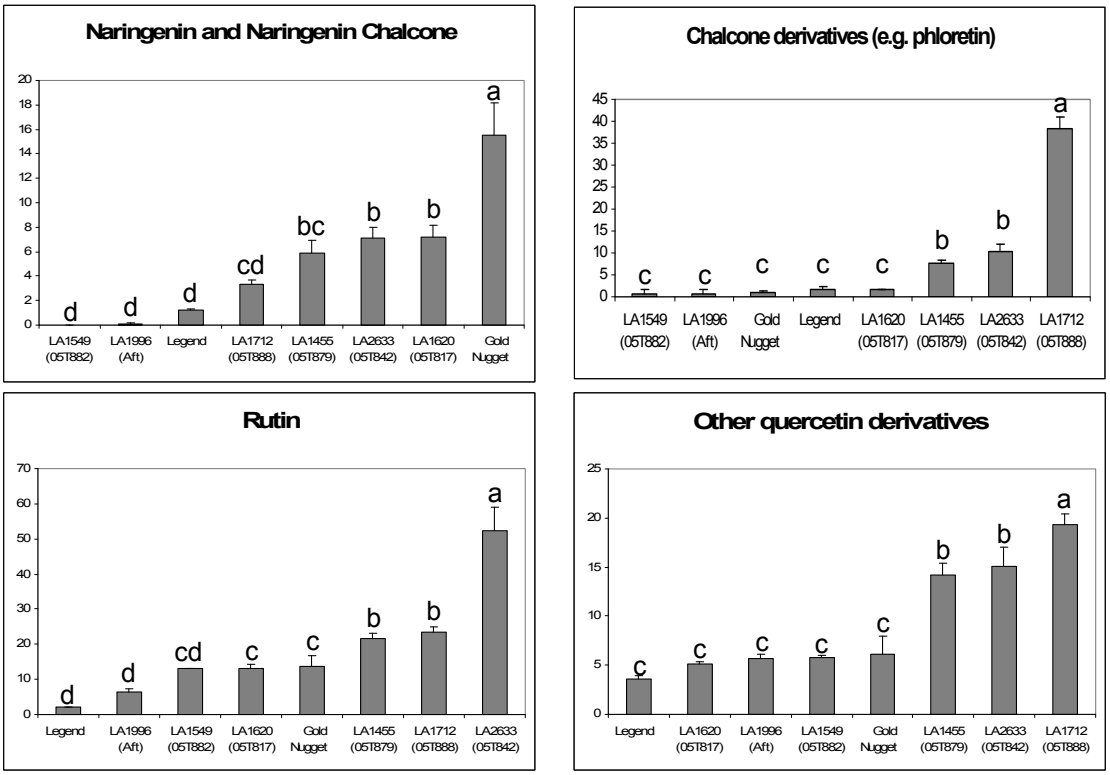


Figure 4.4. Continued…

HPLC-PDA information for greenhouse grown material (Figure 4.4) was based on three or more biological replicates. LA2633 was notable for having levels of quercetin and kaempferol derivatives that were significantly higher than cultivars and many other *S. l.* var. *cerasforme* accessions. Similarly, LA1712 had significantly higher levels of chalcone derivatives than cultivars and other *S. l.* var. *cerasforme* accessions. Most *S. l.* var. *cerasforme* accessions had significantly higher levels of mono-, di-, and tri- acylquinic acids than cultivars.

Phenolics data from field grown material (Table 4.5) and marker genotype information for 87 markers spanning all 12 chromosomes from the Tanksley lab (Tanksley, pers. comm.) was used to look for marker-trait associations with single locus F-tests using both PowerMarker and TASSEL. Marker information TA209, a cultivated tomato, was substituted for 'Gold Nugget' and 'Siletz' given the extremely low levels of polymorphism for most markers observed in cultivated tomato (Labate et al., 2009). The analysis was run with and without the inclusion of 'Gold Nugget' and 'Siletz' without any change in the markers considered significant at $P \leq 0.05$. The

results obtained using PowerMarker were identical to those from TASSEL. Significant marker-trait associations were found and marker haplotypes for all wild accessions are displayed in Table 4.6. In order to filter out false positives among significant markers, the F-statistic for each single locus F-test was plotted against chromosomal location (Figure 4.5) to visualize chromosomal regions that contained multiple marker-trait associations.

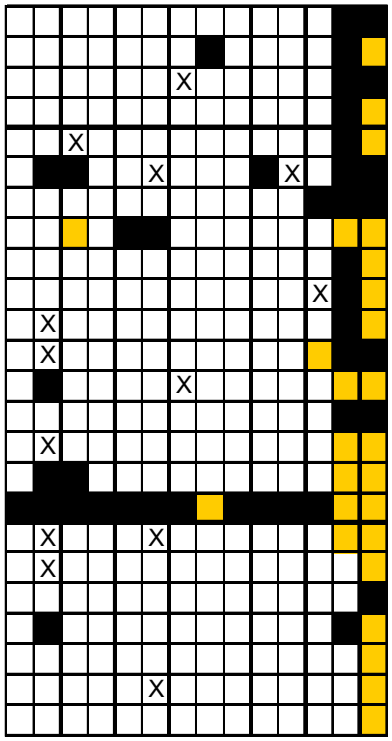
Discussion

Tomato is a model species for the study of fruit development and there has been a recent explosion in metabolomics research related to analysis of phenolic compounds in tomato (Moco et al. 2006, Moco et al. 2007, Iijima et al. 2008, Mintz-Oron et al. 2008). By using information from metabolite databases and standardized chromatography procedures (Moco et al. 2006), we were able to identify a number of well known tomato phenolics (e.g. chlorogenic acid, rutin, naringenin) with confidence. We also have strong confidence in our identification of several lesser known tomato phenolics which have only recently been described, such as quercetin 3-(2"apiosyl-6"rhamnosylglucoside) (Slimestad et al. 2008; Butelli et al. 2008), a quercetin and kaempferol dihexose-deoxyhexose (Moco et al., 2007; Iijima et al. 2008), feruloylquinic acid (Moco et al. 2006), dicaffeoylquinic acids (Moco et al. 2006; Luo et al., 2008), and tricaffeoylquinic acids (Moco et al. 2006; Luo et al., 2008). We also report the occurrence of a FCQA isomer in *S. l.* var *cerasiforme* ($[M-H]^-$=529.3). This compound has been reported in coffee (Clifford et al. 2006) and carrot (Kammerer et al. 2004), but to our knowledge this is the first report of this compound in a wild tomato species. We also found two putative isomers of feruloylquinic acid ($[M-H]^-$=367.0) where only one has been previously reported in tomato (Moco et al. 2006).

Table 4.6. Significant marker-trait associations for phenolic compounds in 15 *S.l.* var *cerasiforme* accessions. Abbreviatons: coumaric and caffeic acid hexoses (CC), caffeoylquinic acids (CQA), feruloylquinic acids (FQA), dicaffeoylquinic acids and feruloyl-caffeoylquinic acids (DI), tricaffeoylquinic acids (TRI), kaempferol derivatives (K), miscellaneous quercetin derivatives (Q), rutin (R), chalcone derivatives (CHA), total phenolics as measured by HPLC (HPLC), and total phenolics (F-C). Phenolics data is from the 2007 field trial. For each marker, the chromosome (Ch.), location (in cM) and p-values for association tests (F-test from one way ANOVA) are listed. Bold text indicates a significant p-value. Genotype codes (for haplotypes): A='LA1290, B=LA1338, C=LA1455, D=LA1512, E=LA1549, F=LA1620, G=LA1668, H=LA1712, I=LA2076, J=LA2135, K=LA2308, L=LA2626, M=LA2632, N=LA2633. Shading indicates allelic state: dark=homozygous for wild allele, light=heterozygous for wild species allele, no shading= homozygous cultivated allele, and X=missing genotype data. Phenylpropanoid related genes mapping nearby a marker are listed under candidate genes.

			p-value (1 way ANOVA F-test)										
Marker	Ch.	Pos.	CC	CQA	FQA	DI	TRI	K	Q	R	CHA	HPLC	F-C
5g13030	1	15.3	0.53	**0.01**	**0.00**	0.37	0.29	**0.00**	**0.00**	0.07	**0.02**	**0.01**	0.15
5g13700	1	46.5	0.70	**0.04**	**0.02**	0.40	0.77	1.00	**<0.00**	**0.01**	**0.02**	**<0.00**	**0.04**
SSR222	1	97.5	0.51	**0.01**	**0.00**	0.35	0.33	**0.01**	**0.00**	0.07	**0.02**	**0.01**	0.14
dTG608	2	2.0	0.73	0.05	**0.02**	0.50	0.58	1.00	**<0.00**	**0.02**	**0.04**	**<0.00**	0.35
1g78690	2	32.0	0.76	**0.04**	**0.01**	0.51	0.63	1.00	**<0.00**	**0.02**	**0.04**	**<0.00**	0.26
SSR331	2	78.0	**0.05**	0.21	0.24	0.31	0.43	0.19	0.09	0.48	0.36	0.08	0.10
5g51110	3	74.0	0.52	0.15	**0.08**	0.41	0.41	1.00	**<0.00**	0.01	0.04	**<0.00**	0.47
5g60160	3	83.0	0.69	0.00	**0.00**	0.58	0.34	0.11	**0.01**	**0.06**	**0.05**	**0.01**	0.06
2g45730	4	129.5	0.73	0.05	**0.02**	0.50	0.58	1.00	**<0.00**	**0.02**	**0.04**	**<0.00**	0.35
1g07080	6	3.5	0.70	0.07	0.02	0.54	0.55	1.00	**<0.00**	**0.02**	**0.05**	**0.00**	0.40
T834	6	32.0	0.69	0.06	**0.02**	0.40	0.64	1.00	**<0.00**	**0.03**	**0.05**	**0.00**	0.38
4g10030	6	44.0	0.60	**0.05**	**0.02**	0.40	0.47	**0.02**	**0.01**	0.18	0.08	0.06	0.38
1g19140	7	23.5	**0.01**	**0.02**	**0.01**	0.19	0.15	**0.02**	**0.01**	0.18	**0.03**	0.05	0.34
2g20860	7	43.3	0.53	**0.01**	**0.00**	0.37	0.29	**0.00**	**<0.00**	0.07	**0.02**	**0.01**	0.15
5g27390	8	20.5	0.60	**0.02**	**0.00**	0.28	0.33	**0.01**	**<0.00**	0.08	**0.02**	**0.01**	0.17
TG302	8	37.0	**0.02**	**0.05**	**0.02**	0.19	**0.00**	**0.02**	**<0.00**	0.20	**0.05**	**0.02**	0.18
SSR335	8	50.0	0.54	0.21	0.39	0.08	0.52	0.23	0.11	0.39	0.36	0.12	**0.02**
U233990	8	68.5	0.66	**0.02**	**0.01**	0.28	0.37	**0.01**	**<0.00**	**0.09**	0.02	**0.02**	0.20
TG348	9	62.0	0.87	0.06	**0.03**	0.17	0.51	1.00	**<0.00**	**0.01**	**0.03**	**<0.00**	0.24
T0156	9	101.0	0.89	0.07	0.04	0.16	0.56	1.00	**<0.00**	**0.01**	**0.02**	**<0.00**	0.15
4g22260	11	38.3	0.36	0.17	0.10	0.24	0.65	1.00	**<0.00**	**0.02**	0.08	**<0.00**	0.42
3g44890	11	62.0	0.72	0.08	**0.03**	0.45	0.61	1.00	**<0.00**	**0.03**	0.05	**0.00**	0.44
T302	11	89.0	0.73	0.05	**0.02**	0.50	0.58	1.00	**<0.00**	**0.02**	**0.04**	**<0.00**	0.35
TG68	12	21.0	0.91	**0.04**	**0.02**	0.24	0.45	1.00	**<0.00**	**0.01**	**0.02**	**<0.00**	0.21

Haplotype (by Genotype)



Candidate Genes

hp-2[1]

F3H (*are*)[2]

AN11 (WD-40)[2]

JAF13 (bHLH)[2]

AN1 (bHLH)[2]

FLS[2]

F3'5'H[2]

[1]Van Tuinen et al., 1997; [2]De Jong et al., 2004

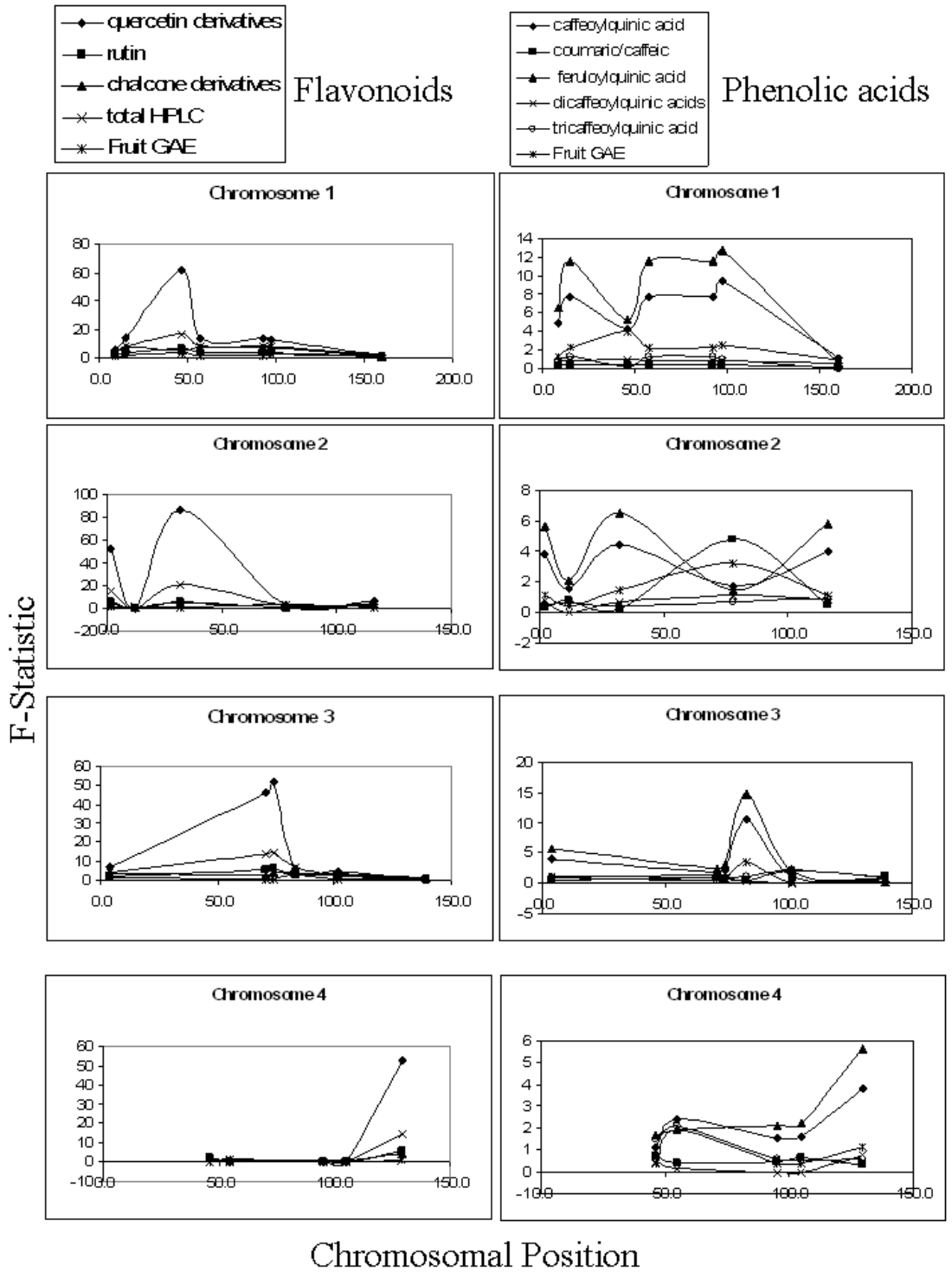


Figure 4.5. Marker-trait association analysis for phenolic compounds using phenotypic data from field trial and prior genotype data from *S. l.* var. *cerasiforme* accessions. Single locus F-tests were conducted using PowerMarker and TASSEL, and F-statistics were plotted against marker chromosomal position in cM. Flavonoid compounds and total phenolics as determined by HPLC (left side) and phenolic acids (right side) were plotted separately. For a list of significant markers, see Table 6.

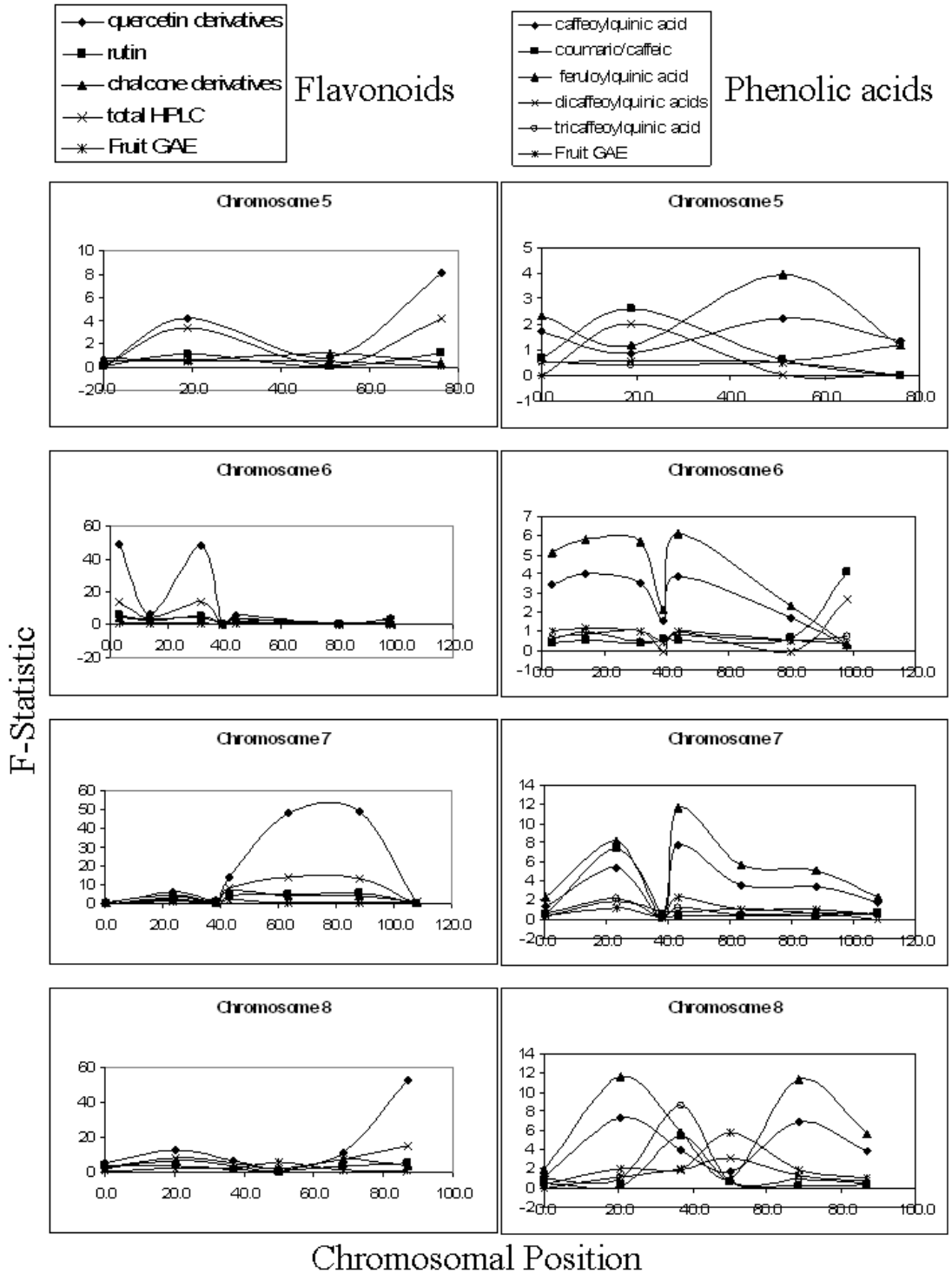


Figure 4.5. Continued.

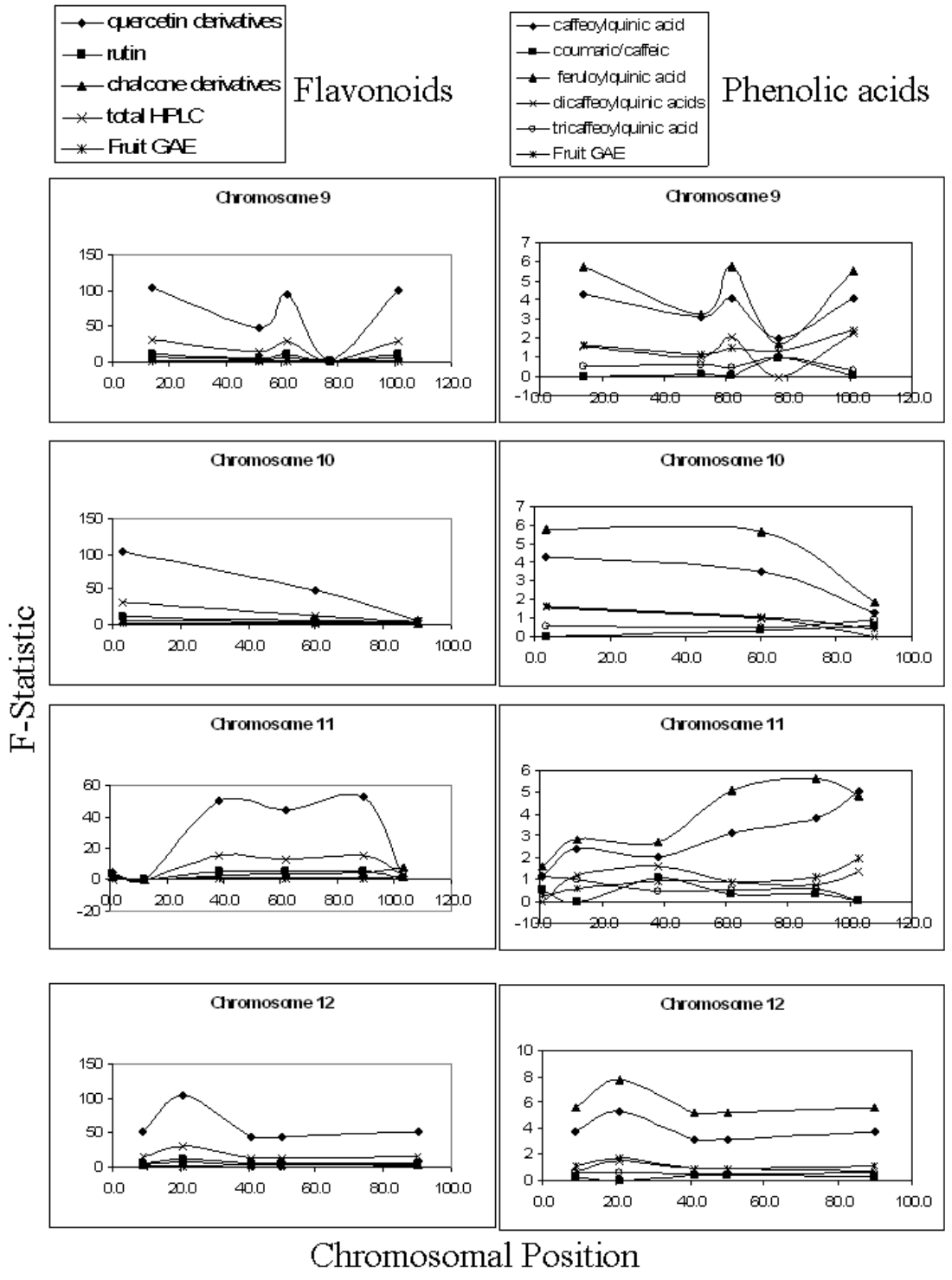


Figure 4.5. Continued.

Two of the UV-spectra (Figure 4.1, B and C) observed for early eluting peaks (e.g peaks 1 and 3) were not a good match for any previously reported tomato phenolics. Kammerer et al. (2004) reported very similar UV-spectra for hydroxycinnamic acid conjugates with amino acids, but at much later retention times than those observed. Total ion chromatograms (Figure 4.3) contained several peaks in this region with $[M-H]^-$=341.0 and 325.0, some of which co-eluted, consistent with reports of several caffeic and coumaric acid hexose isomers in tomato (Moco et al., 2006). We propose that the anomalous UV-spectra observed in peaks 1 and 3 may be due to co-elution of caffeic and coumaric acid hexoses, but our results are not conclusive. In some cases, peaks 1 and 3 made a significant contribution to the levels of total phenolics as determined by HPLC. For example, in LA1996 (*Aft*), the levels of compounds 1 and 3 were up to twice as high as in any other accession studied and were the predominant phenolic compounds. This may have contributed to the poor correlation between scores for total phenolics as measured by HPLC and F-C analysis. When standards are not available, there may be poor correlation between the concentration of a phenolic compound and UV-absorbance at 280 nm.

Another compound for which the identification is uncertain is phloretin 3',5'-di-c-glucoside (peak 12). Slimestad et al. (2008) report the occurrence of this chalcone derivative eluting shortly after rutin and kaempferol 3-(6”rhamnosylglucoside) with a mass of 598, UV-max at 286 nm and a shoulder at 330 nm. Peak 11 fits these criteria except that it elutes before a compound with mass and UV-spectra that match kaempferol 3-(6”rhamnosylglucoside). The retention time for phloretin 3',5'-di-c-glucoside in Slimestad et al. (2008) is in error (R. Slimestad, pers. comm.) but Slimestad is confident that kaempferol 3-(6”rhamnosylglucoside) elutes very shortly after rutin (R. Slimestad, pers. comm.). In the absence of further MS/MS data for this compound, our identification is tentative.

HPLC-PDA chromatograms of *S. l.* var. *cerasiforme* accessions revealed significant differences in their phenolics profiles (Figure 4.2). LA 1712 (Pejibaye, Costa Rica) was notable for having large amounts of putative chalcone derivatives (peaks 8, 12, 14, and 16), quercetin derivatives (peaks 5, 10, and 11) and di- and tri-

caffeoylquinic acids (peaks 15 and 18). Horticulturally, LA1712 had poor fungal disease resistance and very acidic flavor. Among DNA markers associated with phenolic traits, LA1712 (haplotype 'H' in Table 4.6) had wild species alleles for only 5g13700 and SSR335. LA1455 (Nuevo Leon, Mexico) had very high levels of 4-caffeoylquinic acid (chlorogenic acid). LA1455 also had upright plant architecture, high yield; and small, sweet, fruit with surface trichomes. LA1455 (haplotype 'C' in Table 4.6) had wild species alleles for the markers SSR331, 5g60160, TG302, and SSR335. LA2633 (Cusco, Peru) had high levels of chorogenic acid (peak 4), kaempferol and quercetin derivatives (peaks 5, 6, 10, 11, and 13) and di- and tri-acylquinic acids (peaks 15, 17, and 18). LA2633 (haplotype 'N' in Table 4.6) had wild species alleles for every locus with significant marker-trait associations. LA2633 had hard skin, sweet fruit with high solids, good yield, and long fruit trusses. LA1549 (Pasco, Peru) had moderately high levels of chorogenic acid (peak 4), quercetin derivatives (peaks 5, 10, and 11) and di- and tri- caffeoylquinic acids (peaks 15 and 18). LA1549 (haplotype 'E' in Table 4.6) had wild species alleles for the loci 5g60160 and SSR335, a similar haplotype to LA1455 and LA1620 which had similar phenolics profiles.

The levels of phenolic compounds found in tomatoes or other fruits vary greatly depending on the sample preparation method, tissue(s) sampled, genotype, and growing conditions (Robbins, 2003). Among cultivated tomatoes, tissue type and maturity stage have major effects on the quantity and profile of compounds (Moco et al., 2007). According to the USDA Database for the Flavonoid Content of Selected Foods, Release 2.1 (2007) the year-round average naringenin and quercetin content of red tomatoes is 0.68 and 0.58 mg·100 g^{-1} FW, respectively; while for cherry tomatoes these values are 3.19 and 2.76 mg·100 g^{-1} FW, respectively. These values are based on the entire edible portion of the fruit. In this study, seeds, gel, and columellar tissue were excluded because the inclusion of liquid nitrogen ground seeds in the extraction might affect the fruit phenolics profile in ways that are biologically unrealistic as far as human consumption is concerned. Another reason for excluding these tissues was to avoid the dilution of minor compounds below the limit of detection without concentration of the samples, which can cause degradation. We report values 4-8

times those of the USDA for cultivated tomatoes in the tissue sampled here (pericarp and epidermis). This value is reasonable given that the seeds, columella, gel and placental tissues constitute a significant portion of the weight of a fresh tomato, and the phenolics are concentrated in the outer layers of the pericarp and epidermis. Recently, flavonoid concentrations of up 200 mg·100 g^{-1} FW have been reported for tomato peel (Giuntini et al., 2008; Torres et al., 2005). Thus, we believe that the high concentration of quercetin and other phenolics reported here for some small fruited, wild accessions (up to 70 mg·100 g^{-1} FW quercetin in LA2633 under greenhouse conditions) is relatively accurate.

In this study, the largest source of error in quantifying many compounds by HPLC may be the lack of available standards for many compounds, which were quantified as equivalents to rutin, chlorogenic acid, or naringenin. For rutin and chlorogenic acid, standards were available so the standard curves were directly related to the amount of the compound present. These compounds were quantified separately from other compounds for this reason. For chlorogenic acid, LA1455, LA1549, LA1620, LA1668, LA1712, LA1996, LA2632, and LA2633 had two or more times as much chlorogenic acid as cultivars in the field trial (Table 4.5). LA1620, LA1455, and LA2633 had significantly more chlorogenic acid than LA1996 in the greenhouse evaluation ($P \leq 0.05$, Figure 4.4). In the field trial, only LA2633 had noticeably more rutin than 'Gold Nugget' (Table 4.5), though because the replicate samples were consolidated for HPLC analysis, statistical analysis was not possible. In the greenhouse evaluation, similar results were obtained but the difference in the amount of rutin between 'Gold Nugget' and LA1455, LA1712, and LA2633 was significant ($P \leq 0.05$, Figure 4.4).

Hanson et al. (2004) reported total phenolics as measured by the Folin-Ciocalteau method in tomato in the range of 60-90 mg·100 g^{-1} FW for cultivars and 150 mg·100 g^{-1} FW for small fruited *S. pimpinellifolium* accessions. Hernandez et al. (2007) reported lower values in the range of 10-30 mg·100 g^{-1} FW for cultivated tomatoes. These differences could be due to the sample preparation method (fresh vs. frozen puree, respectively) but differences in the standard curves are another possible

explanation. We report values intermediate between the two studies (40 mg·100 g^{-1} FW).

Smaller tomato fruit also have higher phenolics concentrations due to their increased surface area: volume ratio. There were several indications that the high phenolic content observed in *S. l.* var. *cerasiforme* accessions was not due to small fruit size alone, however. We found a significant correlation between fruit size and phenolics content, but fruit size was not strictly correlated with phenolics content (e.g. LA1512 had small fruit but only modest phenolics levels). Most importantly, the occurrence of unique phenolic compounds and profiles among *S. l.* var. *cerasiforme* accessions strongly suggests that some of these accessions contain unique structural or regulatory genes related to phenypropanoid biosynthesis.

The absence of all but trace amounts of naringenin or naringenin chalcone in samples from field grown plants was unexpected. The possibility that this is an artifact that occurred during extraction or storage is possible. However, recovery of naringenin from spiked samples was over 95% using this extraction protocol (data not shown) and naringenin chalcone was observed in greenhouse grown samples. We have not observed significant degradation of other samples over similar time periods at -20° C (data not shown), in accordance with results of Moco et al. (2006). Another possibility is that differences between greenhouse and field grown fruit are due to environmental differences such as lighting. Giuntini et al. (2008) found large differences in the amount of naringenin at various tissues and fruit ripening stages depending on cultivar and exposure to UV-B light. For example, in the cultivar 'Esparanza' there were large amounts of naringenin in the flesh of UV-B shielded fruits at the red-ripe stage, while in the cultivar 'DRW 5981' naringenin was found mainly in flesh of UV-B shielded fruits at the turning (breaker) stage and in the peel of fruits exposed to full sunlight at the red-ripe stage. Since studies using greenhouse grown fruit and/or hydrolyzed flavonoids for analysis are over-represented in the literature, the relative amount of naringenin *in vivo* under field growing conditions may be over-estimated at present.

We have observed that in high phenolics tomato genotypes such as *Aft/atv* and *hp-1* homozygotes, naringenin is present in levels equal to or less than those found in normal tomato fruit, while the levels of compounds downstream of naringenin such as

rutin or anthocyanins reach levels up to an order of magnitude greater than those found in a normal tomato (data not shown). Luo et al. (2008) reported an identical phenomenon when tomato was transformed with an *Arabidopsis* MYB transcription factor (AtMYB12) to increase flavonoid production in the fruit, resulting in a 30-40 fold increase of quercetin and kaempferol levels with no increase in the level of naringenin chalcone. Thus, the low levels of chalcone isomerase transcription that result in the accumulation of naringenin chalcone (Schijlen et al. 2006, Schijlen et al. 2004) in cultivated tomato may not be the normal state in wild tomatoes or genotypes with increased phenolics.

Nearly all of the high phenolics *S. l.* var. *cerasiforme* accessions had levels of di- and tri- acylquinic acids that were significantly higher than cultivated tomatoes (Figure 4.4), making this *S. l.* var. *cerasiforme* a good source for this trait. MicroTom cultivars transformed with *At*MYB12 produced di- and tri- acylquinic acids at a concentration of 4-6 $mg \cdot g^{-1}$ DW (Luo et al., 2008), a level similar to that observed for high phenolics *S. l.* var. *cerasiforme* accessions in this study (4-7 $mg \cdot 100\ g^{-1}$ FW; approximately equal to 4-7 $mg\ g^{-1}$ DW). The genetic basis of this trait therefore seems more likely to be due to a difference in regulation of the pathway leading to these compounds rather than a lack of the necessary structural genes in the cultivated tomato. The only marker significantly associated with di- or tri- caffeoylquinic acid production was TG302 (Table 4.6), which maps to chromosome 8 at an offset of 37 cm, which is not near any genes known to be involved in phenylpropanoid biosynthesis.

Some markers that were significantly associated with phenolics production did map nearby to previously mapped phenylpropanoid metabolism genes. For example, marker 5g13700, which was associated with chlorogenic acid, quercetin derivatives, chalcone derivatives, and total phenolics (F-C) (Table 4.6) is located on chromosome 1 nearby marker TG125 which co-segregated with the *hp-2* and phytochrome B genes (Van Tuinen et al., 1997). Other genes known to be involved in phenylpropanoid metabolism mapping nearby significant markers are listed in Table 4.6 under 'Candidate Genes'.

This study provides the basis for a more extensive candidate gene analysis using classical mutants, mapped ESTs, and the rapidly developing tomato genome sequence. Many of the markers used in this study are Conserved Ortholog Set II (COSII) markers derived from EST sequences (Fulton et al., 2002; Wu et al., 2006). Numerous other COSII markers also map nearby (<5cM) most significant markers, providing additional candidate genes. Gene sequences that map nearby markers associated with phenolics production could be tested further for association with phenolics content in segregating populations.

During the second greenhouse evaluation, S_3 plants of LA1455, LA1549, LA1620, LA1712, and LA2633 were crossed to high anthocyanin (*AftAft*/*atvatv*) and high flavonol (*AftAft*/*atvatv/awaw*) tomato lines with the objective of developing tomato lines that have elevated fruit phenolic levels or unique phenolic profiles. Some of these populations could also be used for mapping phenylpropanoid related genes. Since LA2633 exhibited a high degree of DNA marker polymorphism (Table 4.6) as well as elevated levels of a wide range of polyphenolics, it would make an ideal candidate for developing mapping populations. In contrast, *S. l.* var. *cerasiforme* accessions with high phenolics and a lower proportion of wild species alleles such as LA1712 (Table 4.6) might make better candidates for the introducing the trait into cultivars, in order to reduce linkage drag to unfavorable alleles for other traits.

In conclusion, screening of a genotypically characterized *S. l.* var. *cerasiforme* core collection for accessions with (relatively) good horticultural quality and high phenolics content identified several accessions with elevated phenolics and unique phenolics profiles. These accessions represent a useful resource for developing high phenolics tomato germplasm or identifying genes associated with phenylpropanoid metabolism.

References

Alonso-Salces, R.M., C. Guillou, and L.A. Berrueta. 2009. Liquid chromatography coupled with ultraviolet absorbance detection, electrospray ionization, collision induced association and tandem mass spectrometry on a triple quadropole for the online characterization of polyphenols and methylxanthines in green coffee beans. Rapid Commun. Mass Spectrom. 29:363-383.

Baggett J. R. ; Kean D. 1985. Gold Nugget tomato. HortScience 20:957-958.

Baggett, J. R.; Kean, D.; Mansour, N. S. 1998. Siletz parthenocarpic tomato. HortScience. 32:1299-1300.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-5.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

Clifford, M.N., S. Knight, B. Surucu, and N. Kuhnert. 2006. Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: dimethoxycinnamic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoylquinic acids, and feruloyl-dimethoxycinnamoylquinic acids. J. Agric. Food Chem. 54:1957-1969.

De Jong, W.S., N.T. Eannetta, D.M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the *Solanaceae*. Theor. Appl. Genet. 108:423-432.

Economic Research Service, USDA. 2008. Vegetables and melons situation and outlook yearbook. VGS-2008. 30 October 2008.
http://www.ers.usda.gov/Publications/VGS/2008/05May/VGS2008.pdf.

Giuntini, D., V. Lazzeri, V. Calvenzani, C. Dall'Asta, G. Galaverna, C. Tonelli, K. Petroni, and A. Ranieri. 2008. Flavonoid profiling and biosynthetic gene expression in flesh and peel of two tomato genotypes grown under UV-B-depleted conditions during ripening. J. Agric. Food Chem. 56:5905-5915.

Fulton, T.M., R. Van der Hoeven, N.T. Eannetta, and S.D. Tanksley. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457-67.

Halliwell, B., J. Rafter, and A. Jenner. 2005. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? Am. J. Clin. Nutr. 81(suppl):268S-276S.

Hanson, P.M., R. Yang, J. Wu, J. Chen, D. Ledesma, S.C.S. Tsou, and T.C. Lee. 2004. Variation for antioxidant activity and antioxidants in tomato. J. Amer. Soc. Hort. Sci. 129:704-711.

Hernández, M., E. Rodríguez, and C. Díaz. 2007. Free hydroxycinnamic acids, lycopene, and color parameters in tomato cultivars. J. Agric. Food Chem. 55:8604-8616.

Ijiima, Y., Y. Nakamura, Y. Ogata, K. Tanaka, S. Sakurai, K. Suda, T Suzuki, H. Suzuki, K. Okazaki, M. Kitayama, S. Kanaya, K. Aoki, D. Shibata. 2008. Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949-962.

Jones, C.M., P. Mes, and J.R. Myers. 2003. Characterization and identification of the *Anthocyanin fruit* (*Aft*) tomato. J. Hered. 94:449-456.

Kammerer, D., R. Carle, and A. Schieber. 2004. Characterization of phenolic acids in black carrots (*Daucus carota* ssp. *Sativus* var. *atrorubens* Alef.) by high-performance liquid chromatography/electrospray ionization mass spectrometry. Rapid Comm. Mass Spec. 18:1331-1340.

Kushad M.M., J. Masiunas, M.A.L. Smith, W. Kalt, and K. Eastman. 2003. Health promoting phytochemicals in vegetables. Hort. Rev. 28:125-185.

Labate, J.A., L.D. Robertson, F. Wu, S.D. Tanksley, and A.M. Baldo. 2009. EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (*Solanum lycopersicum* L.). Theor. Appl. Genet. Published online (doi:10.1007/s00122-008-0957-2) 20 January 2009.

Liu, K. and Muse, S.V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-9.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56:316-326.

Lotito, S.B. and B. Frei. 2006. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? Free Rad. Biol. & Med. 46:1727-1746.

Middleton, E. and C. Kandaswami. 1993. The impact of flavonoids on mammalian biology: implications for immunity, inflammation and cancer. p. 589-618. In: J.B. Harborne (ed.). The flavonoids: advances in research since 1986. Chapman and Hall, London.

Mintz-Oron, S., T. Mandel, I. Rogachev, L. Feldberg, O. Lotan, M. Yativ, Z. Wang, R. Jetter, I. Venger, A. Adato, and A. Aharoni. 2008. Gene expression and metabolism in tomato fruit surfaces. Plant Physiol. 147:823-851.

Moco, S., R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, and C.H.R. de Vos. 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141:1205-1218.

Moco, S., E. Capanoglu, Y. Tikunov, R.J. Bino, D. Boyacioglu, R.D. Hall, J. Vervoort, and R.C.H. De Vos. 2007. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J. Exp. Bot. 58:4131-4146.

Robbins, R. 2003. Phenolic acids in foods: an overview of analytical methodology. J. Agric. Food Chem. 51:2866-2887.

Scalbert, A., I.T. Johnson, and M. Saltmarsh. 2005. Polyphenols: antioxidants and beyond. Am. J. Clin. Nutr. 81(suppl):215S-217S.

Schijlen, E., C.H. Ric de Vos, H. Jonker, H. van den Broeck, J. Molthoff, A. van Tunen, S. Martens, and A. Bovy. 2006. Pathway engineering for phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol. J. 4:433-444.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648.

Singleton, V.L., and J.A. Rossi. 1965. Colorimetry of total phenolics with phospomolybdic-phosphotungstic acid reagents. Amer. J. Enol. Viticult. 16:144-158.

Slimestad, R., T. Fossen, and M.J. Verheul. 2008. The flavonoids of tomatoes. J. Agric. Food Chem. 56:2436-2442.

Stewart, A.J., S. Bozonnet, W. Mullen, G.I. Jenkins, M.E.J. Lean, and A. Crozier. 2000. Occurrence of flavonols in tomatoes and tomato-based products. J. Agric. Food Chem. 48:2663-2669.

Torres, C.A., N.M Davies, J.A. Yanez, and P.K. Andrews. 2005. Disposition of selected flavonoids in fruit tissues of various tomato (*Lycopersicon esculentum* Mill.) genotypes. J. Agric. Food. Chem. 53:9536-9543.

United States Department of Agriculture. 2007. USDA Database for the Flavonoid Content of Selected Foods, Release 2.1. http://www.nal.usda.gov/fnic/foodcomp/Data/Flav/Flav02-1.pdf, last accessed 26 January, 2009.

Vinson, J.A., Y. Hao, X. Su, L. Zubik. 1998. Phenol antioxidant quantity and quality in foods: vegetables. J. Agric. Food Chem. 46:3630-3634.

van Tuinen, A., M.M. Cordonnier-Pratt, L.H. Pratt, R.Verkerk, P. Zabel, and M. Koornneef. 1997. The mapping of phytochrome genes and photomorphogenic mutants of tomato. Theor. Appl. Genet. 94:115-122.

Waterhouse, A. 2005. Determination of total phenolics. p.I.1.1.1-I.1.1.8 in: R.E. Wrolstad, T.E. Acree, E.A. Decker, M.H. Penner, D.S. Reid, S.J. Schwartz, C.F. Shoemaker, D. Smith, an P. Sporns (eds.). Handbook of Food Analytical Chemistry: v.2. John Wiley & Sons, Hoboken, NJ.

Willits, M.G., C.M. Kramer, R.T.N Prata, V. De Luca, B.G. Potter, J.C. Steffens, and G. Graser. 2005. Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. J. Agric. Food Chem. 53:1231-1236.

Wu, F., L.A. Mueller, D. Crouzillat, V. Petíard, and S.D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407-20.

CHAPTER 5

Conclusions

Peter S. Boches

The interest in developing high flavonoid tomatoes has never been greater than at present. The recent development of transgenic tomatoes with high levels of flavonol (Luo et al., 2008) and anthocyanins (Butelli et al., 2008) throughout the fruit using targeted expression of MYB and bHLH transcription factors shows the potential for translating knowledge from model systems such as *Arabidopsis* into crop plants. It also demonstrates the high degree of specificity for particular sets of flavonoid metabolites that certain MYBs have. For example, the MYB used to produce the high flavonol tomato by Luo et al. (2008) specifically activates the production of quercetin glycosides but not anthocyanins in *Arabidopsis* (Stracke et al., 2007) and tomato (Luo et al., 2008). Other genetic engineering experiments with biosynthetic genes have resulted in tomatoes that produce resveratrol or isoflavones (Bovy et al., 2007; Schijlen et al., 2006) by transforming tomato with the appropriate genes from grape or legumes, respectively. Using combinations of engineered transcription factors and biosynthetic genes, the potential to produce large amounts of a wide variety of phenolic compounds in tomato fruit is now technically feasible.

Concerns over the safety of genetically modified organisms currently limit the extent to which the consumer will have access to these products (EFSA GMO panel working group on animal feeding trials, 2008). Though acceptance that the benefits of genetically modified crops with altered nutrient profiles outweigh their risks seems inevitable to this author, in the interim the modification of flavonoid biosynthesis in crop plants will remain in the domain of traditional breeding. Hopefully the research herein will make a useful contribution to this effort. Even in an era when genetic engineering is commonly used, the importance of conserving crop species germplasm and tradional plant breeding knowledge cannot be overemphasized. Natural variation for metabolites and genes will continue to provide valuable information for use in genetic engineering for years to come. Genetic engineering is capable of modifying traits in ways that are impossible with traditional plant breeding techniques. Traditional plant breeding techniques may be more appropriate for combining multiple traits or selecting for subtle phenotypes, however.

Tomato is a model organism for fruit development (White, 2002), molecular breeding (Foolad, 2007), and ecological and evolutionary genomics (Moyle, 2008). The flavonol biosynthetic pathway has been well characterized in tomato (Schijlen et al., 2006) but the genes regulating it, in particular the MYB genes, are only partly characterized at present. The discovery of an undescribed MYB ortholog (*An2*) in this work is an exciting contribution, and further study of this gene and its relatives could shed light on the evolution of complex loci in nature.

Despite extensive study, the phenolic metabolites of tomato have only begun to be characterized. Only recently, there was the recognition that what was previously described as the primary flavonoid of tomato, naringenin (Schijlen et al., 2004), is an artifact created from the spontaneous isomerization of naringenin chalcone during storage (Slimestad et al., 2008). Early studies of tomato frequently analyzed only the hydrolyzed flavonol aglycones for the sake of simplicity (Tokusoglu et al., 2003) or disregard numerous compounds that could not be identified (Hernandez et al., 2007; van Tuinen et al., 2006). Even recent studies contain obvious discrepancies with the literature as a whole, for example, in the work of Luo et al. (2008), a peak with a retention time slightly before rutin is identified as naringenin, when all other reports identify naringenin as eluting at nearly the same time as naringenin chalcone, the last flavonoid to elute (Moco et al., 2006; Slimestad et al., 2008). Similarly, a prominent quercetin derivative of tomato in our analyses (a quercetin hexose-deoxyhexose-pentose), has only recently been reported (Moco et al., 2006; Slimestad et al., 2008, Giuntini et al., 2008). The amounts of flavonoids reported in tomato vary as well, with some authors reporting values that differ by up to an order of magnitude from the rest of the literature (Torres et al., 2005; Giuntini et al., 2008). The structures of the major flavonoids of tomato have only begun to be characterized by NMR (Slimestad et al., 2008). Notably, some researchers have begun systematically studying how phenolics profiles vary in response to particular environmental conditions (Giuntini et al., 2008; Clé et al., 2008). Studies tracking metabolites and gene expression simultaneously such as the work of Carrari et al. (2006) and Mintz-Oron et al. (2008) will help to elucidate gene-metabolite networks.

Although the regulation of anthocyanin biosynthesis is a model system for the understanding of complex gene regulation, many new and exciting discoveries undoubtably remain. The negative regulators of anthocyanin biosynthesis, such as the newly discovered AtMYBL2 (Matsui et al., 2008; Dubos et al., 2008) are an area of active research in model organisms that will have large implications for crop species as well. These negative regulators would make excellent candidate genes for *atv*. Another area that merits further attention is the possibility that the MYB genes controlling anthocyanin biosynthesis are themselves regulated by micro RNAs, as has been demonstrated for the *lanceolate* gene, an R3 MYB which patterns leaf morphology (Ori et al., 2007).

We are entering the era of metabolomics with the widespread availability of technologies like LC-MS (Dixon et al., 2006; Sumner et al., 2002). The work of Moco et al. (2006) and Iijima et al. (2008) identifies literally hundreds of phenolic compounds in tomato fruit using HPLC-MS. The relative amounts of these metabolites that are present in tomato fruit, or how they vary with tissue and environment, is only just beginning to be known (Moco et al., 2007). The discovery of a previously unreported diacylquinic acid (FCQA) in this work in the fruit of anthocyanin mutants (Chapter 3) and wild species (Chapter 4) shows that many more compounds could be expected to be found in wild tomato species. Determining the structures and biosynthetic pathways for such compounds is new territory to be explored. Using a core set of genotypically characterized lines for such metabolomics research, as in Chapter 4, will enhance its usefulness to the scientific community.

The complexity of the tomato metabolite profile revealed by metabolomics approaches complicates the task of determining the relative health benefits of these many compounds, at a time when there is still debate about the antioxidant vs. other biological activities of phenolic compounds (Lotito et al., 2006). Research in animal models typically progresses faster than research in plant models. Ironically, it would seem that by the time that medical researchers have determined exactly what the bioactivity of specific phenolic compounds is, plant scientists will have the technologies to produce them in hand.

References

Bovy, A., E. Schijlen, and R.D. Hall. 2007. Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. Metabolomics 3:399-412.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

Carrari, F., C. Baxter, B. Usadel, E. Urbanczyk-Wochniak, M.I. Zanor, A. Nunes-Nesi, V. Nikiforova, D. Centero, A. Ratzka, M. Pauly, L.J. Sweetlove, and A.R. Fernie. 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiol. 142:1380-1396.

Clé, C., L.M. Hill, R. Niggeweg, C.R. Martin, Y. Guisez, E. Prinsen, M.A.K. Jansen. 2008. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochemistry 56:3391-3397.

Clifford, M.N., S. Knight, B. Surucu, and N. Kuhnert. 2006. Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: dimethoxycinnamic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoylquinic acids, and feruloyl-dimethoxycinnamoylquinic acids. J. Agric. Food Chem. 54:1957-1969.

Dixon, R.A., D.R. Gang, A.J. Charlton, O. Fiehn, H.A. Kuiper, T.L. Reynolds, R.S. Tjeerdema, E.H. Jeffery, J.B. German, W.P. Ridley, and J.N. Seiber. 2006. Applications of metabolomics in agriculture. J. Agric. Food. Chem. 54:8984-8994.

Dubos, C., J. Le Gourrierec, A. Baudry, G. Huep, E. Lanet, I. Debeaujon, J.-M.. Routaboul, A. Alboresi, B. Weisshar, and L. Lepinec. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. Plant J. 55:940-953.

EFSA GMO panel working group on animal feeding trials. 2008. Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. Food Chem. Toxicol. 46(Suppl):S2-S70.

Foolad, M.R. 2007. Genome mapping and molecular breeding of tomato. Int. J. Plant Genomics 2007:64538.

Giuntini, D., V. Lazzeri, V. Calvenzani, C. Dall'Asta, G. Galaverna, C. Tonelli, K. Petroni, and A. Ranieri. 2008. Flavonoid profiling and biosynthetic gene expression in flesh and peel of two tomato genotypes grown under UV-B-depleted conditions during ripening. J. Agric. Food Chem. 56:5905-5915.

Hernandez, M., E. Rodriguez, and C. Diaz. 2007. Free hydroxycinnamic acids, lycopene, and color parameters in tomato cultivars. J. Agric. Food Chem. 55:8604-8615.

Ijiima, Y., Y. Nakamura, Y. Ogata, K. Tanaka, S. Sakurai, K. Suda, T Suzuki, H. Suzuki, K. Okazaki, M. Kitayama, S. Kanaya, K. Aoki, D. Shibata. 2008. Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949-962.

Lotito, S.B. and B. Frei. 2006. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? Free Rad. Biol. & Med. 46:1727-1746.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56:316-326.

Matsui, K., Y. Umemura, and M. Ohme-Takagi. 2008. AtMYBL2, a protein with a single myb domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. Plant J. 55:954-967.

Mintz-Oron, S., T. Mandel, I. Rogachev, L. Feldberg, O. Lotan, M. Yativ, Z. Wang, R. Jetter, I. Venger, A. Adato, and A. Aharoni. 2008. Gene expression and metabolism in tomato fruit surfaces. Plant Physiol. 147:823-851.

Moco, S., R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, and C.H.R. de Vos. 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141:1205-1218.

Moco, S., E. Capanoglu, Y. Tikunov, R.J. Bino, D. Boyacioglu, R.D. Hall, J. Vervoort, and R.C.H. De Vos. 2007. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J. Exp. Bot. 58:4131-4146.

Moyle, L.C. 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). Evolution 62:2995-3013.

Ori, N., A. R. Cohen, A. Etzioni, A. Brand, O. Yanai, S. Shleizer, N. Menda, Z. Amsellam, I. Efroni, I. Pekker, J.P. Alvarez, E. Blum, D. Zamir, and Y. Eshed. 2007. Regulation of LANCEOLATE by *miR319* is required for compound-leaf development in tomato. Nat. Genet. 39: 787-791.

Schijlen, E., C.H. Ric de Vos, H. Jonker, H. van den Broeck, J. Molthoff, A. van Tunen, S. Martens, and A. Bovy. 2006. Pathway engineering for phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol. J. 4:433-444.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry

Slimestad, R., T. Fossen, and M.J. Verheul. 2008. The flavonoids of tomatoes. J. Agric. Food Chem. 56:2436-2442.

Stracke R., H. Ishihara, G. Huep, A. Barsch, F. Mehrtens, K. Niehaus, B. Weisshar. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 50:660-677.

Sumner, L.W., P. Mendes, R.A. Dixon. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62: 817-836.

Tokusoglu, O., M.K. Unal, and Z. Yildirim. 2003. HPLC-UV and GC-MS characterization of the flavonol aglycons quercetin, kaempferol, and myricitin in tomato pastes and other tomato based products. Acta Chromatographica 13:196-207.

Torres, C.A., N.M Davies, J.A. Yanez, and P.K. Andrews. 2005. Disposition of selected flavonoids in fruit tissues of various tomato (*Lycopersicon esculentum* Mill.) genotypes. J. Agric. Food. Chem. 53:9536-9543.

van Tuinen, A., C.H Ric de Vos, R.D. Hall, L.H.W. van der Plas, C. Bowler, and R.J. Bino. 2006. Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from matural light hyperresponsive mutants, p. 339-356. In: P.K. Jaiwal (ed.) Plant Genetic Engineering Vol. 7: Metabolic engineering and molecular farming.

White, P.J. 2002. Recent advances in fruit development and ripening: an overview. J. Exp. Bot. 53:1995-2000.

BIBLIOGRAPHY

Aaby, K., Skrede, G., Wrolstad, R.E. 2005. Phenolic composition and antioxidant activities in flesh and achenes of strawberries (*Fragraria ananassa*). J. Agr. Food Chem. 53:4032-4040.

Accotto, G.P., J. Navas-Castillo, E. Norris, E. Moriones, D. Louro. 2000 Typing of tomato leaf curl viruses in Europe. European J. Plant Pathol. 106:179-186.

Aharoni, A., C.H. Ric De Vos, M. Wein, Z. Sun, R. Greco, A. Kroon, J.N.M. Mol, and A.P. O'Connell. 2001. The strawberry FAMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J. 28:319-332.

Alonso-Salces, R.M., C. Guillou, and L.A. Berrueta. 2009. Liquid chromatography coupled with ultraviolet absorbance detection, electrospray ionization, collision induced association and tandem mass spectrometry on a triple quadropole for the online characterization of polyphenols and methylxanthines in green coffee beans. Rapid Commun. Mass Spectrom. 29:363-383.

Ando, T., N. Saito, F. Tatsuzawa, T. Kakefuda, K. Yamagake, E. Ohtani, M. Koshi-ishi, Y. Matsusake, H. Kokubun, H. Watanabe, T. Tsukamoto, Y. Ueda, G. Hashimoto, E. Marchesi, K. Asakura, R. Hara, H. Seki. 1999. Floral anthocyanins in wild taxa of *Petunia* (Solanaceae). Biochem. Systematics Ecol. 27:623-650.

Atanassova, B., and Shtereva, L. 1995. Anthocyaninless mutants- potential for improvement of germination ability in *Lycopersicon esculentum* Mill. Tomato Genet. Coop. Rpt. 45:7-9.

Athma, P. and T. Peterson. 1991. *Ac* induces homologous recombination at the maize *P* locus. Genetics. 128:163-173.

Baggett J. R., and D. Kean. 1985. Gold Nugget tomato. HortScience 20:957-958.

Baggett, J. R., D. Kean, and N.S. Mansour. 1998. Siletz parthenocarpic tomato. HortScience 32:1299-1300.

Baudry, A., M. Caboche, and L. Lepiniec. 2006. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J. 46:768-779.

Bino, R.J., C.H. Ric de Vos, M. Lieberman, R.D. Hall, A. Bovy, H.H. Jonker, Y. Tikunov, A. Lommen, S. Moco, and I. Levin. 2005. The light hyperresponsive high pigment-2dg mutation of tomato: alterations in the fruit metabolome. New Phytol. 166:427-438.

Boches, P. and J.R. Myers. 2007. The *Anthocyanin fruit* tomato gene (*Aft*) is associated with a DNA polymorphism in a MYB transcription factor. HortScience 42:856 (abstr.).

Boddu, J., C. Jiang, V. Sangar, T. Olson, T. Peterson, and S. Chopra. 2006. Comparative structural and functional characterization of sorghum and maize duplications containing orthologous Myb transcription regulators of 3-deoxyflavonoid biosynthesis. Plant Mol. Bio. 60:185-199.

Bongue-Bartelsman M., S.D. Oneill, Y. Tong, and J.I. Yoder. 1994. Characterization of the gene encoding dihydroflavonol-4-reductase in tomato. Gene 19:327-336.

Borovsky, Y., M. Oren-Shamir, R. Ovadia, W. De Jong, I. Paran. 2004. The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to *Anthocyanin2* of *Petunia.* Theor. Appl. Genet. 109:23-29.

Bovy, A., E. Schijlen, and R.D. Hall. 2007. Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. Metabolomics 3:399-412.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-5.

Broun, P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. Curr. Opin. Plant Biol. 8:272-279.

Butelli, E., L. Titta, M. Giorgio, H.-P. Mock, A. Matros, S. Peterek, E.G.W.M. Schijlen, R.D. Hall, A.G. Bovy, J. Luo, C. Martin. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature Biotechnol. 26:1301-1308.

Capanoglou, E., J. Beekwilder, D. Boyacioglu, R. Hall, and R. de Vos. 2008. Changes in antioxidant and metabolite profiles during the production of tomato paste. J. Agric. Food Chem. 56:964-973.

Cascuberta, J.M., and N. Santiago. 2004. Plant LTR-retrotransposons and MITES: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1-11.

Carrari, F., C. Baxter, B. Usadel, E. Urbanczyk-Wochniak, M.I. Zanor, A. Nunes-Nesi, V. Nikiforova, D. Centero, A. Ratzka, M. Pauly, L.J. Sweetlove, and A.R. Fernie. 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Pant Physiol. 142: 1380-1396.

Clé, C., L.M. Hill, R. Niggeweg, C.R. Martin, Y. Guisez, E. Prinsen, M.A.K. Jansen. 2008. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochemistry 56:3391-3397.

Cocciolone, S.M., S. Chopra, S.A. Flint-Garcia, M.D. McMullen, and T. Peterson. 2001. Tissue-specific patterns of a maize *Myb* transcription factor are epigenetically regulated. Plant J. 27:467-478.

De Jong, W.S., N.T. Eannetta, D.M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the *Solanaceae*. Theor. Appl. Genet. 108:423-432.

Dias, A.P., E.L. Braun, M.D. McMullen, and E. Grotewold. 2003. Recently duplicated maize R2R3 genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. Plant Phys. 131:610-620.

Dixon, R.A., D.R. Gang, A.J. Charlton, O. Fiehn, H.A. Kuiper, T.L. Reynolds, R.S. Tjeerdema, E.H. Jeffery, J.B. German, W.P. Ridley, and J.N. Seiber. 2006. Applications of metabolomics in agriculture. J. Agric. Food. Chem. 54:8984-8994.

Dubos, C., J. Le Gourrierec, A. Baudry, G. Huep, E. Lanet, I. Debeaujon, J.-M.. Routaboul, A. Alboresi, B. Weisshar, and L. Lepinec. 2008. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. Plant J. 55:940-953.

Economic Research Service, USDA. 2008. Vegetables and melons situation and outlook yearbook. VGS-2008. 30 October 2008. http://www.ers.usda.gov/Publications/VGS/2008/05May/VGS2008.pdf.

EFSA GMO panel working group on animal feeding trials. 2008. Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. Food Chem. Toxicol. 46(Suppl):S2-S70.

Espin, J.C., M.I. Garcia-Conesa, and F.A. Tomas Barbaran. 2007. Nutraceuticals: facts and fiction. Phytochemistry 68:2986-3008.

Foolad, M.R. 2007. Genome mapping and molecular breeding of tomato. Int. J. Plant Genomics 2007:64538.

Frydman, A., O. Weisshaus, M. Bar-Peled, D.V. Huhman, L.W. Sumner, F.R. Marin, E. Lewinsohn, R. Fluhr, J. Gressel, and Y. Eval. 2004. Citrus fruit bitter flavors: isolation and functional characterization of the gene Cm1,2RhaT encoding a rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of the citrus. Plant J. 40:88-100.

Fulton, T.M., J. Chugwongse, S.D. and Tanksley. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Rptr. 13:207-209

Fulton, T.M., R. Van der Hoeven, N.T. Eannetta, and S.D. Tanksley. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457-67.

Frusciante, L., P. Carli, M.R. Ercolano, R. Pernice, A. Di Matteo, V. Fogliano, and N. Pelligrini. 2007. Antioxidant nutritional quality of tomato. Mol. Nutr. Food. Res. 51:609-617.

Genbank Website. 13 January 2009. <www.ncbi.gov>.

Giuntini, D., V. Lazzeri, V. Calvenzani, C. Dall'Asta, G. Galaverna, C. Tonelli, K. Petroni, and A. Ranieri. 2008. Flavonoid profiling and biosynthetic gene expression in flesh and peel of two tomato genotypes grown under UV-B-depleted conditions during ripening. J. Agric. Food Chem. 56:5905-5915.

Giusti M.M. and Wrolstad R.E., 2001. Unit F1.2.1-13. Characterization and measurement of anthocyanins by UV-visible spectroscopy. In: Current Protocols in Food Analytical Chemistry. R.E. Wrolstad (ed). John Wiley & Sons, NY.

Gonzalez, A., M. Zhao, J.M. Leavitt, A.M. Lloyd. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J. 53:814-827.

Goldsbrough, A., F. Belzile, and J.I. Yoder. 1994. Complementation of the tomato *anthocyanin without* (*aw*) mutant using the dihydroflavonol 4-reductase gene. Plant Physiol. 105:491-496.

Grotewold, E., P. Athma, and T. Peterson. 1991. Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA-binding domain of *myb*-like transcription factors. Proc. Natl. Acad. Sci. USA 88: 4587-4591.

Grotewold E., M.B. Sainz, L. Tagliani, J.M. Hernandez, B. Bowen, V.L. Chandler. 2000. Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proc. Natl. Acad. Sci. 97:13579-13584.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Halliwell, B., J. Rafter, and A. Jenner. 2005. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? Am. J. Clin. Nutr. 81(suppl):268S-276S.

Hanson, P.M., R. Yang, J. Wu, J. Chen, D. Ledesma, S.C.S. Tsou, and T.C. Lee. 2004. Variation for antioxidant activity and antioxidants in tomato. J. Amer. Soc. Hort. Sci. 129:704-711.

Hartmann U., M. Sagasser, F. Mehrtens, R. Stracke, and B. Weisshar. 2005. Differential combinatorial interactions of *cis*-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol. Biol. 57:155-171.

Harwood, M., B. Danielewska-Nikiel, J.F. Borzelleca, G.W. Flamm, G.M. Williams, and T.C. Lines. 2007. A critical review of the data related to the safety of quercetin and lack of evidence of in vivo toxicity, including lack of genotoxic/carcinogenic properties. Food Chem. Toxicol. 45:2179-2205.

Hernandez, M., E. Rodriguez, and C. Diaz. 2007. Free hydroxycinnamic acids, lycopene, and color parameters in tomato cultivars. J. Agric. Food Chem. 55:8604-8615.

Hong, S.H., H.J. Kim, J.S. Ryu, H. Choi, S. Jeong, J. Shin, G. Choi, and H.G. Nam. 2008. CRY1 inhibits COP1-mediated degradation of BIT1, a MYB transcription factor, to activate blue light-dependent gene expression in *Arabidopsis*. Plant J. 55:361-71.

Howell, A.B., J.D. Reed, C.G. Krueger, R. Winterbottom, D.G. Cunningham, and M. Leahy. 2005. A-type cranberry proanthocyanidins and uropathogenic bacterial anti-adhesion activity. Phytochemistry 66:2281-2291.

Ijiima, Y., Y. Nakamura, Y. Ogata, K. Tanaka, S. Sakurai, K. Suda, T Suzuki, H. Suzuki, K. Okazaki, M. Kitayama, S. Kanaya, K. Aoki, D. Shibata. 2008. Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949-962.

Jablonska, B., J.S.S. Amiraju, K.K. Bhattarai, S. Mantelin, O.M de Ilarduya, P.A. Roberts, and I. Kaloshian. 2006. The *Mi-9* gene from *Solanum Arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of *Mi-1*. Plant Phys. 143:1044-1054.

Ji, Y. and R. Chetelat. 2003. Homoeologous pairing and recombination in *Solanum lycopersicoides* monosomic addition and substitution lines of tomato. Theor. Appl. Genet. 106:979-989.

Jiang C., X. Gu, and T. Peterson. 2004. Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. Genome Biol. 5:R46.

Jones, C.M., P. Mes, and J.R. Myers. 2003. Characterization and identification of the *Anthocyanin fruit* (*Aft*) tomato. J. Hered. 94:449-456.

Kammerer, D., R. Carle, and A. Schieber. 2004. Characterization of phenolic acids in black carrots (*Daucus carota* ssp. *Sativus* var. *atrorubens* Alef.) by high-performance liquid chromatography/electrospray ionization mass spectrometry. Rapid Comm. Mass Spec. 18:1331-1340.

Kandaswami, C., L.T. Lee, P.P. Lee, J.J. Hwang, F.C. Ke, Y.T. Huang, and M.T. Lee. 2005. The antitumor activities of flavonoids. In Vivo: 5:895-909.

Kerckhoffs, L.H.J., N.A.M.A. De Groot, A. Van Tuinen, M.E.L. Schreuder, A. Nagatini, M. Koornneef, and R.E. Kendrick. 1997. Physiological characterization of exaggerated-photoresponse mutants of tomato. J. Plant Physiol. 150:578-587.

Koes, R, W. Verweij, and F. Quattrocchio. 2005. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. 10:236-42.

Kris-Etherton, P.M., K.D. Hecker, A. Bonanome, S.M. Coval, A.E. Binkoski, K.F. Hilpert, A.E. Griel, and T.D. Etherton. 2002. Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. Am. J. Med. Suppl. 9B:71S-88S.

Kranz, H., K. Scholz, and B. Weisshar. 2000. C-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. Plant J. 21:231-235.

Kushad M.M., J. Masiunas, M.A.L. Smith, W. Kalt, and K. Eastman. 2003. Health promoting phytochemicals in vegetables. Hort. Rev. 28:125-185.

Labate, J.A., L.D. Robertson, F. Wu, S.D. Tanksley, and A.M. Baldo. 2009. EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (*Solanum lycopersicum* L.). Theor. Appl. Genet. Published online (doi:10.1007/s00122-008-0957-2) 20 January 2009.

Laitenin, R.A.E, M. Ainasoja, S.V. Broholm, T.H. Teeri, and P. Elomaa. Identification of target genes for a MYB-type anthocyanin regulator in *Gerbera hybrida*. J. Exp. Botany 59:3691-3703.

Lenucci, M.S., D. Cadinu, M. Taurino, G. Piro, and G. Dallessandro. 2006. Antioxidant composition in cherry and high-pigment tomato cultivars. J. Agric. Food Chem. 54:2606-2613.

Li, Y., H. Fang, and W. Xu. 2007. Recent advances in research of flavonoids as anticancer agents. Mini Rev. Med. Chem. 7:663-78.

Li, J., X. Li, L. Guo, F. Lu, X. Feng, K. He, L. Wei, Zhangliang Chen, L.J. Qu, and H. Gu. 2006. A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. J. Exp. Bot. 57:1263-1273.

Lijavetsky D., L. Ruiz-Garcia, J.A. Cabezas, M. T. De Andrés, G. Bravo, A. Ibáñez, J. Carreño, F. Cabello, J. Ibáñez, and J. Martínez-Zapater. 2006. Molecular genetics of berry colour variation in table grape. Mol. Gen. Genomics 276:427-435.

Lin, J.-K., and M.-H. Weng. 2006. Flavonoids as nutraceuticals. P. 213-238 in: Grotewold, E. (ed). The Science of Flavonoids. Springer, NY.

Liu, K. and Muse, S.V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-9.

Liu, P-.P., N. Koizuka, R.C. Martin, and H. Nonogaki. 2005. The *BME3* (*Blue Micropylar End 3*) GATA zinc finger transcription factor is a positive regulator of *Arabidopsis* seed germination. Plant J. 44: 960-971.

Long, M., D.J. Millar, Y. Kimura, G. Donovan, J. Rees, P.D. Fraser, P.M. Bramley, and G.P. Bolwell. 2006. Metabolite profiling of carotenoid and phenolic pathways in carotenoid and transgenic pathways in mutant and transgenic lines of tomato: identification of a high antioxidant fruit line. Phytochemistry 67:1750-1757.

Lotito, S.B. and B. Frei. 2006. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? Free Rad. Biol. & Med. 46:1727-1746.

Luo, J., E. Butelli, L. Hill, A. Parr, R. Nigeweg, P. Bailey, B. Weisshaar, and C. Martin. 2008. AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. 56: 316-326.

Mathews H., S.K. Klendennen, C.G. Caldwell, X. Liang Liu, K. Connors, N. Matheis, D.. Schuster, D.J. Menasco, W. Wagoner, J. Lightner, and D.R. Wagner. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689-1703.

Matsui, K., Y. Umemura, and M. Ohme-Takagi. 2008. AtMYBL2, a protein with a single myb domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. Plant J. 55: 954-967.

Matus, J.T., F. Aquea, and P. Arce-Johnson. 2008. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. BMC Plant Biol. 8:83.

Menzel, G., D. Dechyeva, H. Keller, C. Lange, H. Himmelbauer, and T. Schmidt. 2006. Mobilization and evolutionary history of minature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. Chromosome Res. 14:831-844.

Mes, P. 2005. Breeding tomatoes for improved antioxidant activity. Ph.D. Dissertation, Oregon State University, Corvallis.

Mes P.J., Boches P., and Myers J.R. 2008. Characterization of tomatoes expressing anthocyanin in the fruit. J. Am. Soc. Hort. Sci. 133:262-269.

Middleton, E. and C. Kandaswami. 1993. The impact of flavonoids on mammalian biology: implications for immunity, inflammation and cancer. p. 589-618. In: J.B. Harborne (ed.). The flavonoids: advances in research since 1986. Chapman and Hall, London.

Mintz-Oron, S., T. Mandel, I. Rogachev, L. Feldberg, O. Lotan, M. Yativ, Z. Wang, R. Jetter, I. Venger, A. Adato, and A. Aharoni. 2008. Gene expression and metabolism in tomato fruit surfaces. Plant Phys. 147:823-851.

Minoggio, M., Bramati, L., Simonetti, P., Gardana, C., Lemoli, L., Santangelo, E., Mauri, P.L., Spigno, P., Soressi, G.P., and Pietta, P.G. 2003. Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. Ann. Nutr. Metab. 47:64-69.

Moco, S., R.J. Bino, O. Vorst, H.A. Verhoeven, J. de Groot, T.A. van Beek, J. Vervoort, and C.H.R. de Vos. 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141:1205-1218.

Moco, S., E. Capanoglu, Y. Tikunov, R.J. Bino, D. Boyacioglu, R.D. Hall, J. Vervoort, and R.C.H. De Vos. 2007. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J. Exp. Bot. 58:4131-4146.

Morita, Y., M. Saitoh, A. Hoshino, E. Nitasaka, S. Iida. 2006. Isolation of cDNAs for R2R3-MYB, bHLH, and WDR transcriptional regulators and identification of c and ca mutations conferring white flowers in the Japanese morning glory. Plant Cell Phys. 47:457-470.

Moyle, L.C. 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). Evolution 62:2995-3013.

Muir, S.R., Collins, G.J., Robinson, S., Hughes, S., Bovy, A., de Vos, C.H.R., van Tunen, A.J., Verhoeyen, M.E. 2001. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nat. Biotechnol. 19:470-474.

Nicot, N., J.-F. Hausman, L. Hoffman, and D. Evers. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J. Exp. Bot. 56:2907-2914.

Okamura, K., J.W. Hagen, H. Duan, D.M. Tyler, and E.C. Lai. 2007. The mirtron pathway generates microRNA-class regulatory class RNAs in *Drosophila*. Cell 130:89-100.

Ori, N., A. R. Cohen, A. Etzioni, A. Brand, O. Yanai, S. Shleizer, N. Menda, Z. Amsellam, I. Efroni, I. Pekker, J.P. Alvarez, E. Blum, D. Zamir, and Y. Eshed. 2007. Regulation of LANCEOLATE by *miR319* is required for compound-leaf development in tomato. Nat. Genet. 39:787-791.

Overy, S.A., H.J. Walker, S. Malone, T.P. Howard, C.J. Baxter, L.J. Sweetlove, S.A. Hill, and W.P. Quick. 2005. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. J. Exp. Bot. 56:287-296.

Owens DK, Crosby KC, Runac J, Howard BA, Winkel BSJ (2008) Biochemical and genetic characterization of *Arabidopsis* flavanone 3β-hydroxylase. Plant Phys. Biochem. 46:833-843.

Paz-Ares, J., D. Ghosal., U. Wienard., P.A. Peterson, and H. Saedler. 1987. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural simlarities to transcriptional activators. EMBO J. 6:3553-3558.

Peng, Y., Y. Zhang, and J. Ye. 2008. Determination of phenolic compounds and ascorbic acid in different fractions of tomato by capillary electrophoresis with electrochemical detection. J. Agric. Chem. 56: 1838-1844.

Pietta, P.-G. 2000. Flavonoids as antioxidants. J. Nat. Prod. 63:1035-1042.

Price C.L., and Wrolstad R.E., 1995. Anthocyanin pigments of Royal Okanogan Huckleberry juice. J. Food Sci. 60: 369-374.

Quattrocchio, F., A. Baudry, L. Lepiniec, and E. Grotewold. 2006. The regulation of flavonoid biosynthesis, p. 97-122. In: Grotewold, E. (ed). The Science of Flavonoids. Springer, NY.

Quattrochio, F., J.F. Wing, H.T.C. Leppen, J.N.M. Mol, and R.E. Koes. 1993. Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5:1497-1512.

Quattrocchio, F., J. Wing, K. Van der Woude, E. Souer, N. de Vetten, J. Mol, and R. Koes. 1999. Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. Plant Cell 11:1433-1444.

Rabinowicz, P.D., E.L. Braun, A. D. Wolfe, B. Bowen, and E. Grotewold. 1999. Maize R2R3 Myb genes: sequence analysis reveals amplification in the higher plants. Genetics 153:427-444.

Ramsay, N.A., and B.J. Glover. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10:62-70.

Ren, W., Z. Qiao, H. Wang, L. Zhu, and L. Zhang. 2003. Flavonoids: promising anticancer agents. Med. Res. Rev. 23:519-534.

Rensselaer Bioinformatics Website. 8 February, 2009. <http://mfold.bioinfo.rpi.edu/.>

Rick, C.M. 1980. Tomato, p. 669–680. In: Fehr, W.R. and H. Hadley (eds.). Hybridization of crop plants. Amer. Soc. Agron., Crop Sci. Soc. Amer., Madison, WI.

Rick, C.M., P. Cisneros, R.T. Chetelat, and J.W. DeVerna. 1994. *Abg*– A gene on chromosome 10 for purple fruit derived from *S. lycopersicoides*. Tomato Genet. Coop. Rpt. 44:29–30.

Robbins, R. 2003. Phenolic acids in foods: an overview of analytical methodology. J. Agric. Food Chem. 51:2866-2887.

Rommens, C.M., Richeal, C.M., Yan, H., Navarre, D.A., Ye, J., Krucker, M., Swords, K. 2008. Engineered native pathways for high kaempferol and caffeoylquinate production in potato. Plant Biotechnol. 6:870-886.

Rousseaux, M.C., C.M. Jones, D. Adams, R. Chetelat, A. Bennet, and A. Powell. 2005. QTL analysis of fruit analysis in tomato using *Lycopersion pennellii* introgression lines. Theor. Appl. Genet. 111:1396-1408.

Rozen S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ.

Sapir, M., M. Oren-Shamir, R. Ovadia, M. Reuveni, D. Evenor, Y. Tadmor, S. Nahon, H. Shlomo, L. Chen, A. Meir, and I. Levin. 2008. Molecular aspects of *Anthocyanin fruit* tomato in relation to *high pigment-1*. J Hered 99:292-303.

Savage, N.S., T. Walker, Y. Wieckowski, J. Scheifelbein, L. Dolan, and N.A.M Monk. 2008. A mutual support mechanism through intercellular movement of CAPRICE and GLABRA3 can pattern the *Arabidopsis* root epidermis. PLoS Biol. 6:e235.

Scalbert, A., I.T. Johnson, and M. Saltmarsh. 2005. Polyphenols: antioxidants and beyond. Am. J. Clin. Nutr. 81(suppl):215S-217S.

Schijlen, E., C.H. Ric de Vos, H. Jonker, H. van den Broeck, J. Molthoff, A. van Tunen, S. Martens, and A. Bovy. 2006. Pathway engineering for phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol. J. 4:433-444.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648.

Schwinn, K., J. Venail, Y. Shang, S. Mackay, V. Alm, E. Butelli, R. Oyama, P. Bailey, K. Davies, and C. Martin. 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. Plant Cell 18:831-851.

Singleton, V.L., and J.A. Rossi. 1965. Colorimetry of total phenolics with phospomolybdic-phosphotungstic acid reagents. Amer. J. Enol. Viticult. 16:144-158.

Slimestad, R., T. Fossen, and M.J. Verheul. 2008. The flavonoids of tomatoes. J. Agric. Food Chem. 56:2436-2442.

SOL Genomics Network website. 20 March 20, 2007. <http://www.sgn.cornell.edu/>.

Stewart, A.J., S. Bozonnet, W. Mullen, G.I. Jenkins, M.E.J. Lean, and A. Crozier. 2000. Occurrence of flavonols in tomatoes and tomato-based products. J. Agric. Food Chem. 48:2663-2669.

Stracke R., H. Ishihara, G. Huep, A. Barsch, F. Mehrtens, K. Niehaus, B. Weisshar. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J 50:660-677.

Sumner, L.W., P. Mendes, R.A. Dixon. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817-836.

Takos, A.M., Jaffe, F.W., Jacob, S.R., Bogs, J., Robinson, S.P., Walker, A.R. 2006. Light induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol. 142:1216-1232.

TAIR, the *Arabidopsis* Information Resource website. 1 October 2008. <http://www.arabidopsis.org/>.

Tanksley S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G.B. Martin, R. Messeguer, J.C. Miller, L. Miller, A.H. Paterson, O. Pineda, M.S. Roder, R.A Wing, W. Wu, and N.D. Young. 1992. High density maps of the tomato and potato genomes. Genetics 132:1141-1160.

Tokusoglu, O., M.K. Unal, and Z. Yildirim. 2003. HPLC-UV and GC-MS characterization of the flavonol aglycons quercetin, kaempferol, and myricitin in tomato pastes and other tomato based products. Acta Chromatographica 13:196-207.

Tohge, T., Y. Nishiyama, M. Yokota Hirai, M. Yano, J.-I. Nakajima, M. Awazuhara, E. Inoue, H. Takahashi, D.B. Goodenowe, M. Kitoyama, M. Noji, M. Yamazaki, and K. Saito. 2005. Functional genomics analysis by integrated analysis of metabolome and transcripome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J 42:218-235.

Tong, Y., and J. Yoder. 1995. The tomato anthocyanin without mutation (*aw*) co-segregates with the dihydroflavonol 4-reductase gene. Tomato Genet. Coop. Rpt. 45:48.

Torres, C.A., N.M. Davies, J.A. Yanez, and P.K. Andrews. 2005. Disposition of selected flavonoids in fruit tissues of various tomato (*Lycopersicon esculentum* Mill.) genotypes. J. Agric. Food. Chem. 53:9536-9543.

United States Department of Agriculture. 2007. USDA Database for the Flavonoid Content of Selected Foods, Release 2.1. 26 January, 3009. <http://www.nal.usda.gov/fnic/foodcomp/Data/Flav/Flav02-1.pdf>.

van der Rest, B., S. Danoun, A.M. Boudet, and S.F. Rochange. 2006. Down-regulation of cinnamoyl-coA reductase in tomato (*Solanum lycopersicum* L.) induces dramatic changes in soluble phenolic pools. J. Exp. Bot. 57:1399-1411.

Van Nocker, A., and P. Ludwig. 2003. The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. BMC Genomics 4:50-61

van Tuinen, A., M.M. Cordonnier-Pratt, L.H. Pratt, R.Verkerk, P. Zabel, and M. Koornneef. 1997. The mapping of phytochrome genes and photomorphogenic mutants of tomato. Theor. Appl. Genet. 94:115-122.

van Tuinen, A., C.H Ric de Vos, R.D. Hall, L.H.W. van der Plas, C. Bowler, and R.J. Bino. 2006. Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from matural light hyperresponsive mutants, p. 339-356. In: P.K. Jaiwal (ed.) Plant Genetic Engineering Vol. 7: Metabolic engineering and molecular farming.

Vinson, J.A., Y. Hao, X. Su, L. Zubik. 1998. Phenol antioxidant quantity and quality in foods: vegetables. J. Agric. Food Chem. 46:3630-3634.

Wade, H.K., A.K. Sohal, and G.I. Jenkins. 2003. *Arabidopsis* ICX1 is a negative regulator of several pathways regulating flavonoid biosynthesis genes. Plant Phys. 131:707-715.

Waterhouse, A. 2005. Determination of total phenolics. p.I.1.1.1-I.1.1.8 in: R.E. Wrolstad, T.E. Acree, E.A. Decker, M.H. Penner, D.S. Reid, S.J. Schwartz, C.F. Shoemaker, D. Smith, an P. Sporns (eds.). Handbook of Food Analytical Chemistry: v.2. John Wiley & Sons, Hoboken, NJ.

Wessler, S.R., T.E. Bureau, and S.E. White. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. in Genet. Devel. 5:814-821.

White, P.J. 2002. Recent advances in fruit development and ripening: an overview. J. Exp. Bot. 53:1995-2000.

Willits, M.G., C.M. Kramer, R.T.N Prata, V. De Luca, B.G. Potter, J.C. Steffens, and G. Graser. 2005. Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. J. Agric. Food Chem. 53:1231-1236.

Wu, F., L.A. Mueller, D. Crouzillat, V. Petíard, and S.D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174: 1407-20.

Young, P.A. 1954. Purple tops on green tomato fruits. Tomato Genet. Coop. Rpt. 4:18.

Zhang, F., S. Chopra, T. Peterson. 2000. A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. Plant Cell 12:2311-2322.

Zhang, F. and T. Peterson. 2005. Comparisons of maize *pericarp color1* alleles reveal paralogous gene recombination and an organ-specific enhancer region. Plant Cell: 17:903-914.

Zhao, M., K. Morohashi, G. Hatlestad, E. Grotewold, and A. Loyd. 2008. The TTG1-bHLH-MYB complex controls cell fate and and patterning through direct targeting of regulatory loci. Development 135:1991-1999.

Zhu, Q.H., A. Spriggs, L. Matthews, L. Fan, G. Kennedy, F. Gubler, and C. Helliwell. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. 18:1456-1465.

Zimmerman, I.M., M.A. Heim, B. Weisshar, and J.F. Uhrig. 2004. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J. 40:22-34.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406-15.