October 2018

Variation and Evolution of Fruit Ripening Traits in Tomato Species

Ian M. Gillis

Variation and Evolution of Fruit Ripening Traits in Tomato Species

A Dissertation Presented

By

IAN M. GILLIS

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 2018

Plant Biology

Variation and Evolution of Fruit Ripening Traits in Tomato Species

A Dissertation Presented

By

IAN M. GILLIS

Approved as to style and content by:

_____
Ana Caicedo, Chair

_____
Sam Hazen, Member

_____
Lynn Adler, Member

_____
Adam Porter, Member

_____
Li-Jun Ma
Graduate Program Director
Plant Biology Graduate Program

DEDICATION

To my friends and family for providing seemingly unlimited help and support

ACKNOWLEDGMENTS

I would like to thank my parents and sisters for their constant support and encouragement through all my endeavors, and for their role in shaping me. I would also like to thank my extended family, who took over the family roles for me during my time in Massachusetts.

I would like to thank my advisor Ana Caicedo, who helped guide me from the time I came to this school and who provided me with incredible opportunities. Her constant support offered a reassuring backbone that kept me going through all the difficulties I faced and helped me find the path through graduate school that was best for me.

I would like to thank the faculty of UMass in general and of the Plant Biology program specifically and my fellow PB grad students, for being so willing to help, collaborate and lend guidance. The supportive environment that exists here contributed greatly to my ability to succeed.

I would like to thank my committee members, Lynn, Sam and Adam who provided vital direction in my research, ensuring this dissertation would become a reality.

I would like to thank Susan Capistran for her absolutely indispensable help with navigating the bureaucratic side of graduate school. The countless times I found myself completely adrift she got me back to land.

Finally, I would like to thank my PB housemates through all the years, who provided me with a family away from home. Their innumerable varieties of help with graduate life and life outside of graduate school kept me grounded and got me moving again whenever I got stuck.

ABSTRACT

VARIATION AND EVOLUTION OF FRUIT RIPENING TRAITS IN TOMATO SPECIES

SEPTEMBER 2018

IAN M. GILLIS, B.S. UNIVERSITY OF WASHINGTON

Ph.D. UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Ana L. Caicedo

As angiosperm seeds mature within their ovaries, ovary tissue tends to grow and transform itself into fruit, which aids the success of the seeds. Fruits that are fleshy provide numerous ways to aid in the protection and the dispersal of seeds. First, they keep seeds hidden, encased in hard walls, surrounded by poisons and unpalatable compounds, and second, they undergo developmental changes that facilitate seeds' release. Tomatoes, a model fleshy fruit, have all these protective traits, and over the course of ripening they become the familiar fruit that is a staple crop around the world. The wild relatives of cultivated tomatoes, however, have substantial variation in ripening habits. I characterized several fruit traits and their change during ripening in wild tomato species to get a better understanding of the phenotypic variation that exists in fruits. Acquiring this background for the clade enables further investigation of genes behind these variable traits and inferences of how the traits have evolved. To associate fruit traits with genes and genomic regions for further analysis I grew introgression lines (ILs) stemming from introgressions of small portions of the genome of the tomato clade

outgroup *Solanum lycopersicoides*, in the background of the cultivated tomato, *S. lycopersicum*. With these lines, I found regions of the genome that are associated with change of fruit firmness during ripening, providing data for further investigation of the genetics behind this trait. I also investigated the genetic basis of ripe fruit color variation by characterizing the gene *CYC-B*, which produces the enzyme responsible for turning red lycopene into the orange β-carotene, across the tomato clade. My results suggest that regulation of *CYC-B* has been key to the evolution of different fruit colors across the clade, and that the promoter region of the gene is involved in differentiating a β-carotene accumulating plant from a lycopene accumulating plant. The research performed here enhances our understanding of phenotypic and genotypic variation in an understudied angiosperm organ that can alter how plant species interact with animals around them, contributing to our knowledge of how fruit traits evolve and how they can enable plant success.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

CHARACTERIZING THE WILD TOMATO CLADE FOR VARIABLE FRUIT TRAITS

Introduction

Fruits and flowers are structures that are exclusive to angiosperms, and are thus largely correlated with the success of this speciose group. Flowers provide novel means of pollinator attraction and specificity, which help ensure successful pollination with subsequent fertilization and embryo formation (Taiz et al., 2015). Fruits are formed from the ovary surrounding the fertilized ovule, so they house the developing seeds through maturity (Taiz et al., 2015). Angiosperm fruits can take many forms, and can provide many novel ways of protecting developing seeds, and of enabling dispersal of mature seeds. Dry fruits generally enable wind dispersal, such as winged maple seeds or seeds with a pappus acting as a parachute as in dandelions; some dry fruits can enable animal dispersal, such as burrs which attach to animal fur. Fleshy fruits allow a way for a plant to specifically attract types of animal dispersers, and increase the success of seed dispersal and the distance over which seeds can be dispersed.

Fleshy fruits may serve two purposes over the course of their growth. They protect the developing seeds, and they attract animals that can efficiently disperse the mature seeds. This dual role is possible because of the process of ripening, where fruits develop in numerous ways that increase their palatability; developmental changes include alterations in firmness, texture, taste, scent, and color. However, there is abundant variation among species in palatability traits at the end of the ripening process. Whether fleshy fruit traits are determined primarily by selective pressure for seed dispersal, selection for seed protection, or are byproducts of plant metabolic processes is an active

area of inquiry (Bolmgren and Eriksson, 2010; Eriksson et al., 2000; Iseki et al., 2011; Mack, 2000; Willson and Whelan, 1990; Willson et al., 1989). Thus, why fleshy fruit phenotypic diversity exists, what genes are responsible for this diversity, and what processes drive fruit evolution are currently not well understood.

In the past decades, cultivated tomato (*Solanum lycopersicum*), has been used as a model organism for the study of fleshy fruit development (Sato et al., 2012). This is due to the small stature and short life cycle of some tomato cultivars. However, the clade of species containing cultivated tomato also contains much variation in fleshy fruit traits, making it an ideal system in which to study evolution of fleshy fruits. Tomatoes belong to the very large *Solanum* genus which contains c.1,500 species including other major crops like potato (*S. tuberosum*) and eggplant (*S. melongena*), as well as many regionally important crops such as lulo (*S. quitoense*) and tamarillo (*S. betaceum*) (Carrillo-Perdomo et al., 2015; Ministerio de Agricultura y Desarrollo Rural, B.C., 2011). The genus has worldwide distribution with biodiversity hotspots on several continents including around the Andes mountains in South America where the tomato group, *Solanum* section *lycopersicon*, a fairly young clade (~2.5myo (Pease et al., 2016a)), is endemic. The ~13 wild tomato species (Fig. 1) are spread across the widely varying habitats making up the western coast of South America from Ecuador to Chile, including deserts, jungles, coasts, mountains and islands (Moyle, 2008; Pease et al., 2016a).

Species of the tomato clade have adapted to diverse habitats and exhibit various phenotypic differences, including fruit traits and the extent to which these are altered by the process of ripening. Most wild species in the clade do not change color when they ripen, but softening of the fruit is very prevalent (Grumet et al., 1981). Aside from these

traits, the limited phenotyping on wild tomato fruits has been confined to fully mature fruits (Asano et al., 1996; Furui et al., 1997; Iijima et al., 2013). This biases studies towards the potential roles of fruit traits in seed dispersal, while ignoring fruit's role as protector of young seeds. A fruit that maintains high sugar throughout its development is different from, and likely under different selective pressures than, a fruit that only accumulates high levels of sugar while ripening. The tomato clade presents a good opportunity to study fruit diversification and evolution in the traits of unripe and ripe fruits to see the balance of protection and dispersal qualities.

From the beginning of fertilization and seed development, tomato fruits serve to protect the seeds in several ways. In general, immature tomatoes are firm and thick, to varying degrees, which creates a physical barrier a distance from the seeds. They also contain chloroplasts, which, aside from feeding the growing fruit (Hetherington et al., 1998; Powell et al., 2012), help the fruits blend in with the rest of the plant's vegetative growth. Solanaceous plants are known for their toxic defense compounds, such as deadly nightshade, which famously contains atropine and scopolamine. Tomatoes contain the less toxic steroidal glycoalkaloid α-tomatine, which can kill fungi (Ökmen et al., 2013; Zaccardelli et al., 2011) and insects (Eigenbrode and Trumble, 1994; Güntner et al., 1997) that grow/eat into the fruits, preventing pests from getting deep enough to damage seeds. The taste of α-tomatine can also be quite unpleasant, from a human viewpoint at any rate, which could dissuade potential dispersers from eating the fruits before the seeds are ready, though other research suggests that some frugivorous insects may gain protective benefits by consuming α-tomatine as well as other compounds found in tomatoes (Traugott and Stamp, 1997).

In cultivated tomatoes the ripening process removes these protective functions to yield the edible fruit we are familiar with. The fruit softens and all the flesh that had served as a barrier can now become palatable food, also increasing the fruit's vulnerability (Cantu et al., 2008). Chloroplasts are converted to chromoplasts, eliminating all the photosynthetic machinery and accumulating the bright red pigment lycopene. The α-tomatine is broken down to undetectable levels (Kozukue and Friedman, 2003), which also gets rid of its bitter taste. The taste is further improved by the accumulation of sugars, aromatic compounds and many other molecules that affect the overall flavor and aroma of ripe tomatoes (Bennett, 2012; Tieman et al., 2012).

Some of these fruit ripening traits seen in cultivated tomatoes exist in wild tomatoes, but there exists much variation that may have evolved due to varying species habitats and perhaps dispersal mechanisms. Fruits of *S. galapagense* and *S. cheesmaniae,* species endemic to the Galapagos Islands, are thought to be eaten and dispersed by birds and tortoises on the Galapagos Islands (Rick and Bowman, 1961). Fruit in these species vary from bright orange to pale yellow, and are soft when ripe. It is possible that these traits may have been selected for by dispersers, though islands can easily become limited for food and thus the selective pressure may not be linked to food choice among dispersers. Some accessions of the species *S. habrochaites*, which is native to western Ecuador and Peru, have been reported to undergo no outward sign of ripening before dehiscence (Grumet et al., 1981). These fruits may not be attractive for animal consumption, but round hard fruits falling from a plant can bounce and roll which may be a suitable method for seed dispersal, especially as these tomatoes are commonly found growing on Andean mountainsides. These rocky areas may also be a food limited habitat

in which animal dispersers will eat these fruits, despite lack of traits that humans would consider palatable. The seeds of many wild tomato species require extensive seed coat erosion before they will germinate (Rick and TGRC, 2013) which in nature can be achieved by passing through a digestive tract (Rick and Bowman, 1961) or by remaining in a fallen, uneaten fruit (from personal experience). The lack of knowledge of the seed dispersers and predators of wild tomato species limits understanding of the forces driving fruit evolution, as does an incomplete understanding of trait variation in the group.

Studies in animal preference and diet provide prior information that can inform how traits may be shaped by reliance on particular animal dispersers. Sugar content is often associated with the type of seed disperser that species interact with in the wild. Birds largely cannot taste or digest disaccharides (Baldwin et al., 2014; Del Rio and Stevens, 1989), though hummingbirds are a notable exception, so plants that rely on bird frugivores as seed dispersers tend to have high glucose and fructose levels (Baker et al., 1998). Mammals, on the other hand, have a greater affinity for sucrose and fructose than for glucose, so plants with mammalian seed dispersers tend to have fruits higher in these sugars (Baker et al., 1998; Floerchinger et al., 2010; Ramirez, 1990). Fruit firmness and change of firmness over the course of ripening can be informative about the protection of seeds and their dispersal. Small fruits are often swallowed whole by birds and do not need to soften but usually have visual cues to their ripeness (Wang and Schaefer, 2014), while larger fruits are less likely to be swallowed whole so softening of these fruits could aid in their consumption. Genes underlying traits associated with increased seed dispersal may show signs of this selection, allowing a gene-centric view into selective pressures faced by different species.

Tomato domestication began thousands of years ago and tomatoes have become a staple crop for much of the world (Statista, 2013) providing an important source of dietary carotenoids, which are required for human health (Berman et al., 2014; Fraser and Bramley, 2004). This domestication and further crop improvement has caused a loss of much variation in cultivated tomatoes, resulting in fixation of 25% of the total genome (Lin et al., 2014). The variation found in wild tomatoes, particularly *S. pennellii*, has been used many times for crop improvement by breeding beneficial alleles for biotic and abiotic resistance traits as well as fruit quality traits into cultivated tomato (Hajjar and Hodgkin, 2007; Menda et al., 2014; Fridman et al., 2001). Despite the need and proven use for wild variation in crop improvement, few species have received as much study as *S. pennellii* and few traits have been phenotyped in wild species as thoroughly as resistance traits have. Further documentation of the variation that exists in wild tomato species could strengthen the foundation for future attempts of finding beneficial traits and associated alleles to target for use in crop improvement.

Aside from potential utility, the biodiversity of this group provides an opportunity to understand the evolution of fleshy fruit traits. Since tomato is a model organism, many resources are available that can aid in the work with wild species, making the process easier than with many wild plant species. Three species, cultivated *S. lycopersicum* as well as the wild species *S. pimpinellifolium* and *S. pennellii*, have fully sequenced genomes completed and available online, via solgenomics.net, providing a resource for finding the genes underlying these traits. This system allows the study of evolutionary processes affecting fleshy fruits, a structure that has had an important role in the success of angiosperms as well as in human welfare.

The purpose of this study is to carry out an in-depth characterization of phenotypic diversity in certain key fruit traits across the tomato clade. Due to the dual role fruits play in a plant's reproductive success, changes to ripening characteristics can have a large effect. A more complete characterization will lead to insights on the evolution of these traits as well as helping to develop the tomato clade for further study of fleshy fruit evolution.

Materials and Methods

Plant materials and growth conditions

Seeds were obtained from the UC Davis C.M. Rick Tomato Genetics Resource Center, maintained by the Department of Plant Sciences, University of California, Davis, CA. Cultivated tomato, *S. lycopersicum*, the 'feral' admixed variety (Ranc et al., 2008) *S. lycopersicum var. cerasiforme,* and eleven wild tomato species from *Solanum* section *lycopersicon* were used in this study (Fig. 1); *S. chilense* was not included in our project due to poor fruit set. Three accessions of each species were chosen based on maximizing diversity from across the species range; a single individual of each accession was grown. Plants were grown in Conviron growth chambers with 12-hour day length. As pollination is required for fruit set, the self-compatible species, *S. galapagense*, *S. cheesmaniae*, *S. pimpinellifolium, S. neorickii* and *S. chmielewskii*, were self-pollinated, while the self-incompatible species were pollinated with pollen pooled from all grown individuals of the same species.

Fruits from all accessions were collected at the mature green stage, determined by the point where fruits have reached their full size but have not begun to ripen (Sargent

and Moretti, 2014); this stage is referred to in this chapter as unripe. The mature green stage was identified by observing various fruits on the same plant and on the same raceme (the collection of separate flowers along one stem) and observing unripe fruits that were the same size as ripe fruits. Along racemes the more basal fruits ripen first, so fruits further down the raceme can be collected as the earlier fruits ripen. Fruits from all accessions were also collected at the fully ripe stage. From personal observation I've found this stage is associated with browning of the calyx in all species. In addition, fruits of most species abscise when ripe and four species change color (Grumet et al., 1981). All measurements were made on fruits at both ripeness stages to monitor how these traits change over the course of ripening.

Fruit collection and firmness measurements

At the time of collection, firmness was measured by pressing into each fruit with a 0.5mm^2 blunt metal stick attached to a digital force gauge (Imada, Northbrook, IL). The fruit was pierced from four different directions to the columella of the fruit, with two piercing through septa and two through locules, and the maximum force used in each case recorded. The four measurements were averaged to give the fruit firmness. Fruits were then halved and seeds were removed; the remaining tissues were frozen in liquid nitrogen. Frozen tissue in quantities of 1-3 halves depending on fruit size and availability was ground in a Retsch ball mill with a single 1cm stainless steel ball, to form a fine powder that was stored at -80°C for further analyses. Firmness measurements include more fruits than other traits, as all grown fruits were immediately measured for firmness, but only a portion were used in other trait analyses.

Sugar content measurements

A target sample of frozen powder samples from three fruits at each stage of ripeness from three accessions of each species was set for sugar measurements. In some cases, low fruit set led to different sample sizes for some accessions of the same species. Additionally, some species had more successful fruit set, and fruits from more accessions of those species were sampled. To measure sugar content of tomato fruits, the ground and frozen tissue was thawed and the solid and liquid phases were separated using a centrifuge at 8,000 rpm for 5 minutes. A total of 2µl from the liquid phase was used directly for sugar measurements with the R-Biopharm sugar kit (Roche Yellow line Sucrose/D-Glucose/D-Fructose enzymatic kit), using the modified method by (Velterop and Vos, 2001) to accommodate the small volume of liquid available from wild tomato fruits. Briefly, the kit allows for measuring of sugar levels indirectly, based on the amount of NADP formed during the enzymatic reduction of each sugar. Levels of glucose, fructose, and sucrose were measured at both stages of ripeness for all samples.

α-Tomatine measurements

A target sample of frozen powder samples from three fruits at each stage of ripeness from three accessions of each species was set for α-tomatine measurements. In some cases, low fruit set led to different sample sizes for some accessions of the same species. Additionally, some species had more successful fruit set and fruits from more accessions of those species were sampled. For each sample, approximately 0.020 grams of powdered fruit tissue was used, to account for lack of precision and varying availability of tissue absorbance area was normalized by the milligrams of fruit used for extraction. α-Tomatine was extracted using solid phase extraction (SPE) Oasis HLB 1 cc

vac cartridges [WAT094225] from Waters (Milford, MA). Tomatine was measured using high performance liquid chromatography (HPLC) via a Waters 1515 HPLC pump with a Waters 2489 UV/Visible light detector set at 208nm. We used the XBridge BEH C18 2.5µm 3.0x100mm XP column with the matching 20mm guard column, heated to 40°C for the runs. The mobile phase used was 24:76 acetonitrile (ACN):0.02M potassium phosphate (KH2PO4) adjusted to pH 7, though the final run used a mobile phase with the exact makeup as the running solvent. The running solvent used to dissolve the extract was 5:3:2 tetrahydrofuran:acetonitrile:0.02M KH2PO4.

Analyses of fruit trait variation

Analyses of fruit traits was done in R (R Core Team, 2015). For trait analysis, the lme4 package (Bates et al., 2015) was used to create a linear mixed model, using the lmer function, which can account for many sources of random variation. The lme4 package readily handles multiple sources of random variation, which can be caused by imbalance that comes from differing fruit production between individuals and species. Our models accounted for fruits by individual, block by individual, and individual by species. The linear mixed models were then used as input for ANOVA and Tukey's HSD tests, which were carried out with the anova function of the stats package and the glht function of the multcomp package (Hothorn et al., 2008) respectively. Analyses were done comparing all species at both ripeness stages for each trait.

Phylogenetically independent contrasts

A newick tree file from (Pease et al., 2016a) was used for phylogenetic analysis. The species *S. lycopersicum, S. sitiens* and *S. chilense* were removed from the tree since

they were not in our phylogenetic analyses. For species with multiple representative individuals in the tree, we selected the shortest branch length to avoid overemphasizing species. A single value to represent the trait for each species at each ripeness stage was taken from the linear mixed effects models created above and used for further analysis. Phylogenetically independent contrasts (PICs) were run using the R package 'phytools' and the command phylosig, with the test methods λ and K. Pagel's λ and Blomberg's K are statistics comparing the variation in a trait as measured to what would be expected if the trait were evolving under pure Brownian Motion (BM), given the known tree structure for the species measured. Under a Brownian Motion model, the amount of phenotypic differences between species should be proportional to the time of divergence between them, and this is considered to be analogous to evolution of neutral genes under drift (Blomberg et al., 2003). Pagel's λ is a measure of whether closely related species are similar to each other for the trait as expected under BM ($\lambda \sim 1$) or if their relatedness does not predict similar trait values ($\lambda \sim 0$). Blomberg's K is a measure of whether the variance in the trait occurs within clades ($K<1$) or between clades ($K>1$) or if it occurs as expected under BM ($K=1$). These both measure phylogenetic signal in different ways. They can be thought of as measures of how much closely related species covary (λ), and whether the overall variance occurs mostly within or between clades (K).

Results

Groupings of tomato species based on ancestry and fruit color

We verified the distribution of ripe fruit color among species. Four of the sampled species have fruits that change color during ripening, *S. cheesmaniae*, *S. galapagense*, *S. lycopersicum*, and *S. pimpinellifolium*, and multiple sources of evidence show that these species have a monophyletic origin (Marshall et al., 2001; Mohan et al., 2016; Spooner et al., 2005) (Fig. 1). In this chapter, this group of species will be referred to as the "color" group. Remaining species have fruit that is green at maturity, and in this study include *S. chmielewskii*, *S. arcanum*, *S. neorickii*, *S. huaylasense*, *S. peruvianum*, *S. corneliomulleri*, *S. habrochaites*, and *S. pennellii* (Fig. 1). These species will be referred to as the "green" group throughout the manuscript. The outgroup, *S. lycopersicoides*, is not considered a part of the tomato clade, and its fruit also remains green through ripening. Colored species can be split into species with red ripe fruits, *S. lycopersicum* and *S. pimpinellifolium*, and species with orange ripe fruits, *S. cheesmaniae* and *S. galapagense*. Among green species, *S. chmielewskii*, *S. arcanum*, and *S. neorickii*, here referred to as the CAN group, are distinguished by their sister clade status to the color group. *S. pennellii* is another green fruited species to highlight, as it is sometimes phenotypically similar to the color group, although phylogenetically quite distant.

Sugar content and concentration

There is no one dominant sugar type in the unripe fruit between species. Unripe color species are higher in glucose than most of the green species, with 8-12 mg/ml versus 4-10 mg/ml; however, *S. pennellii*, a green-fruited species, also has a very high

glucose concentration at 20 mg/ml unripe (Fig. 2a; Fig. 3a, Table 1). In contrast, fructose concentrations of unripe fruits vary substantially between green species, with the CAN species having the lowest levels of fructose with concentrations of 5-8 mg/ml in unripe fruits while the remainder of the green species show large variance in fructose concentrations from 9-17 mg/ml (Fig. 2a, Fig. 3c, Table 1). The color species and *S. pennellii* are consistently at the high end of the glucose content variance at 10-15 mg/ml (Fig. 2a, Table 1). Sucrose concentrations vary substantially within some species (Fig. 3e), but are generally highest in the green species with a large range from 5-24 mg/ml, while the lowest concentrations come from the fairly uniform color group with 1-3 mg/ml sucrose in unripe fruits. Overall sugar content in unripe tomato fruits is generally low with little variation, though several species do not follow this trend (Fig. 2a, Fig. 3g, Table 1). *Solanum neorickii* has the lowest overall sugar concentration of 16.4 mg/ml and *S. corneliomulleri* had the highest overall unripe sugar concentration at 50.2 mg/ml (Fig. 3g).

In ripe fruits, a general trend is that glucose is very low in most green fruited species, ranging between 2-5 mg/ml, compared with the color species, which have large variability in concentration at 10-32 mg/ml (Fig. 2b, Fig. 3b, Table 1). *Solanum pennellii* stands out as an atypical green-fruited species due to its high levels of glucose in ripe fruits. Fructose levels are lowest in *S. lycopersicoides* and the CAN species with a concentration range of 4-8 mg/ml, whereas the remainder of the green species as well as the color group have fructose concentration ranging from 10-20 mg/ml in their ripe fruit (Fig. 2b, Fig. 3d, Table 1). Sucrose is another sugar where a stark difference is seen between color and green-fruited species (Fig. 2b, Fig. 3f, Table 1). Sucrose has high

concentrations in most green species' fruits though there is a large concentration range, from 12.9-45.8 mg/ml. Sucrose is present in very low levels in color species' fruits and *S. lycopersicoides*, with a range of 1-8.5 mg/ml. Sucrose is the main sugar in ripe fruits of the green group and glucose makes up the smallest proportion of total sugars (Table 2). In ripe fruits of the color group the predominant sugar is glucose, aside from *S. cheesmaniae* fruits, which have more fructose, and sucrose makes up the smallest proportion of total sugars. *S. pennellii* stands out again with no sugar making up more than 40% of the total and roughly even proportions of fructose and sucrose (Table 2). Total sugar concentrations cover a somewhat larger range than seen for unripe fruits, with concentrations ranging from 20.6 mg/ml in *S. lycopersicoides* up to 68.3 mg/ml in *S. corneliomulleri* (Fig. 2b, Fig. 3h).

All species increased the quantity of total sugar in fruits over the course of ripening. The sugars that were preferentially accumulated in fruits varied between species, as did the concentrations (Fig. 2, Table 3). The outgroup, *S. lycopersicoides*, shows the least accumulation of sugars during ripening and has the lowest amount of total sugar at ripeness. The green-fruited species, *S. neorickii*, also stands out in having very low levels of total sugars at ripeness, though it shows more accumulation than the outgroup (Fig. 3g,h). At the other end of the extreme, *S. galapagense* more than doubles its total sugar concentration during ripening and *S. habrochaites* also doubles total sugar concentration (Table 3). *S. corneliomulleri* stands out as the tomato species with the highest sugar content both at the unripe stage and at maturity (Fig. 3g,h). Glucose levels decreased in most green species during ripening, while showing mostly modest gains in color species, except in *S. pimpinellifolium*, which more than doubled its content,

reaching 32.2 mg/ml, and in *S. galapagense* which tripled in glucose concentration during ripening to 30 mg/ml (Fig. 3a,b). Fructose concentration in *S. pimpinellifolium* shows almost no change over the course of ripening while the rest of the color species gain fructose, as does *S. pennellii*, which has the largest increase (Table 3). The CAN species, which start with low levels of fructose, show little change or some decrease in fructose concentration during ripening. The remainder of the green species show no pattern during ripening with small to moderate increases in fructose (Table 3, Fig. 3c,d). Sucrose changes show little pattern in color species, with some gain and some loss in concentration, though *S. galapagense* and *S. pimpinellifolium* both more than double in concentration. Green species all accumulate sucrose during ripening. Several species double sucrose concentration, while *S. huaylasense* triples and *S. habrochaites* quadruples sucrose during ripening (Table 3, Fig. 3e,f).

Sugar phylogenetic analysis

A superficial examination of species phylogeny and species sugar accumulation suggests that these are correlated, as there are sizeable differences in sucrose and glucose accumulation between colored and green-fruited species. This is somewhat reflected by PIC results, supporting a model of evolution following Brownian Motion, though there are also some more subtle differences supporting this model of evolution as well. Glucose levels in unripe fruits support a BM model of evolution, with a significant λ very close to 1. Glucose is accumulated in ripe fruits of the color species while green species, aside from *S. pennellii*, do not have much glucose in their ripe fruits. However, the closely related Galapagos species, *S. cheesmaniae* and *S. galapagense*, differ substantially in glucose levels compared to the near uniformity seen among green species, this causes the

ripe glucose levels to break from the BM model. Fructose levels in ripe fruits vary in a way that follows the phylogeny, i.e. the closer the relationship between species the more similar their fructose levels are, with well-supported K and λ values close to 1 (Table 4). The change in sucrose levels over the course of ripening is related to phylogeny though following a BM model of evolution less than fructose levels, with highly supported K and λ values around 0.85. A λ value slightly less than 1 suggests that close species trait values will be a little less predictable than under a BM model, and K value slightly less than 1 suggests trait variance is more concentrated to within clades than a BM model would suggest. Total sugar levels vary substantially between species, though the lack of significance for these λ and K values suggests the variation does not correlate with the phylogeny (Table 4).

Firmness

Unripe fruits vary in firmness among species throughout the tomato clade from 157 to 354 grams of force, although species range values overlap substantially and do not differ greatly from the outgroup species (Table 5, Fig. 4a). Two species stand out as outliers from this middling firmness, and these are sister species at the base of the tomato clade: *S. habrochaites* and *S. pennellii*, which are the firmest and softest species respectively (Fig. 4a). At the ripe stage, color species are marginally softer than the green species, but this division is not significant (Fig. 4b). *Solanum arcanum* has the firmest ripe fruits in the clade, second to the fruit firmness of the outgroup, *S. lycopersicoides*. *Solanum pennellii* fruits continue to be the softest through ripening, with little overlap with the firmness range of other species (Table 5, Fig. 4b).

At the ripe stage, fruit firmness ranges from 63 to 209 grams of force. All species show some amount of softening over the course of ripening, though the outgroup, *S. lycopersicoides*, shows the least softening and the fruits of some individuals of this species do not soften at all during ripening (Table 5). *Solanum pennellii* has the softest fruits at both ripe and unripe stages and softens by the greatest percentage (Table 5). The color species, aside from the cultivated *S. lycopersicum*, soften by more than 50% while most of the green species show less than 50% softening during ripening.

Firmness phylogenetic analysis

Firmness of unripe fruits does not show correlation with the phylogeny (Table 6) due to a general uniform firmness across the entire clade, aside from the sister species *S. habrochaites* and *S. pennellii* which are, respectively, the firmest and the softest species at the unripe stage (Fig. 2a). Ripe fruit firmness is less uniform, with color species having softer ripe fruits than green species, but PIC analyses do not support this trend (Fig. 2b; Table 6). Percent change in fruit firmness during ripening does show a strong phylogenetic signal, with significant K and λ values (Table 6). This is consistent with observations of color species softening by more than 50% and the green species softening by less than 50%. However, it does not capture the outliers: green-fruited *S. pennellii* has the softest fruits and softens the most, with a 60% change in value during ripening, and the outgroup has the hardest ripe fruits and softens the least at 27.7% change during ripening (Table 5).

α-Tomatine

Measurement of α-tomatine presented many difficulties throughout the course of the project. Using mass spectrometry appended to our HPLC setup we determined that dehydrotomatine wasn't removed by our extraction method and it eluted at the same time as α-tomatine for many measurements. Dehydrotomatine occurs at lower quantities in fruits than α-tomatine (Friedman and Levin, 1998; Iijima et al., 2013) but has stronger absorbance due to an additional double bond; thus, overlap in HPLC elution time prevented accurate measurement of the α-tomatine in many samples. Additionally, even though the standards provided linear absorbance curves as expected, quantification efforts using these curves routinely overestimated the amount of α-tomatine present. Results from this method suggested that by weight some of our fruits were 120% α-tomatine; the lowest levels calculated were 5% by weight α-tomatine though it has been reported in other publications that α-tomatine only comprises up to 0.08% of fresh fruit weight in cultivated tomato (Friedman and Levin, 1998; Iijima et al., 2013; Rick et al., 1994a). These results show that the method used here is unable to accurately quantify α-tomatine levels in fruits. Instead, relative amounts of α-tomatine as measured by absorbance area per milligram of fruit used can be discussed (Table 7), though only for samples for which we could resolve the peak for α-tomatine from dehydrotomatine. The number of usable individual measurements from each species are given in Table 8.

Due to the above problems, the α-tomatine dataset is less complete than for other traits and only internal comparisons with the data are possible (Fig. 5). In other studies of cultivated tomato, α-tomatine is present in unripe fruits at varying levels but uniformly decreases to undetectable or near undetectable levels during ripening (Asano et al., 1996;

Friedman and Levin, 1998); however our data did not show a decrease. As expected, *S. lycopersicum* did have the lowest levels of α-tomatine out of all ripe fruit measured, though we found the concentration basically unchanged from unripe fruits – in fact, there was a slight increase. Instead of reflecting true levels of α-tomatine, these measurements may represent the lower limit of our ability to detect α-tomatine with this method. However, increases in tomatine absorbance values with ripening were observed in many other species, suggesting this continued increase through development and ripening is the ancestral state.

Solanum lycopersicoides had the highest levels of α-tomatine in unripe fruits, followed by *S. habrochaites* and *S. chmielewskii*. *Solanum chmielewskii* was the only species decreasing in α-tomatine during ripening, but the unripe value is based on a single fruit and may not be representative. All other species showed an increase in α-tomatine concentration during ripening, which was unexpected as the maturing seeds should need less protection, though fungi and insects are most strongly affected by α-tomatine and they are never dispersers so there may be no selection to shut down α-tomatine production. *Solanum lycopersicoides* had the highest levels of α-tomatine in ripe fruit, though it showed only a small increase during ripening. *S. pennellii* reached the second highest α-tomatine levels, increasing over 200% during ripening. *S. arcanum* also showed a 200% increase of α-tomatine during ripening. *S. corneliomulleri* showed an increase of over 600% though the unripe value is based on a single fruit and may not be representative.

α-Tomatine phylogenetic analysis

Though PIC tests cannot be used due to the limited dataset, some patterns can be seen. Unripe fruits do not have a clear pattern; although green species' α-tomatine levels are lower in general than color species', outliers, such as *S. galapagense* and *S. corneliomulleri*, disrupt this pattern. In ripe fruits, the color species have lower α-tomatine levels compared to most green species, but these differences were not significant. Moreover, variances in tomatine content were larger in many green species compared to color species. At both stages of ripeness, cultivated tomatoes and their feral relative had the lowest tomatine levels of all.

Discussion

Fruit trait variation can be seen throughout the tomato clade, at both stages of ripeness and in how much traits are affected by the ripening process. Variation in unripe fruit traits is believed to relate to seed protection needed by each species. Ripe fruit traits show how much and in what ways the fruits aid in the dispersal of a plant's seeds. The effects of ripening and degree to which traits change over the course of this process can be related to how specialized a fruit is for one task or if it plays a role in both seed protection and dispersal. These traits, measured across the clade, can reveal many aspects of the evolutionary history of tomato and their fruits.

Green-fruited tomato species show considerable variation in many of the traits measured, with *S. pennellii* being an exception to the rest of this group for most traits. *Solanum pennellii* is native to the high elevations and extremely arid conditions of the Andes Mountains and nearing the Atacama Desert, a very harsh environment (Bolger et

al., 2014). Its unripe fruits contrast from other green-fruited species in that they have high levels of all sugars and they are much softer than all other fruits tested, and during ripening further soften and accumulate more sugar, mostly sucrose and fructose. The lack of preferential accumulation of sucrose makes this species' sugar profile completely unlike all other green-fruited species. Some of the ripe fruits lose their internal structure so that, once pierced, the skin yields to a thick fluid matrix suspending the seeds rather than discrete seed containing locules like those seen in cultivated tomato. The softness of unripe fruits suggests that the young seeds do not need protection or that there is some other mechanism to protect the seeds. It may also be part of a different overall protection scheme than is commonly seen, as every part of these plants is covered in sticky glaborous trichomes, and all leaves, flowers, and fruits are easily detached from the plant at any time (personal observation). These traits are unique to *S. pennellii*, in the tomato clade. High glucose and soft fruits may be adaptations to survival in the desert, perhaps attracting hungry dispersers without the use of colorful visual cues.

Solanum habrochaites is the sister species to *S. pennellii*, together making a basal clade long separated from the rest of the tomato clade (Rodriguez et al., 2009). Interestingly, *S. habrochaites* has the hardest unripe fruits though it softens by 45% during ripening, which brings it into the firmness range typical of ripe green fruits. In general, all plants from all species sampled show some level of fruit softening during ripening, the only exception being some individuals of the *S. lycopersicoides* outgroup. These results suggest that softening of the fruit during ripening is a trait specific to the tomato clade, and continuously firm fruits may be a basal trait.

Solanum habrochaites has low concentrations of sugar when unripe but over the course of ripening accumulates sucrose and fructose. This again makes *S. habrochaites* quite distinct from its sister species and more like the other green-fruited species. Sugar profiles high in sucrose are associated with mammalian dispersal (Baker et al., 1998; Ramirez, 1990). Although the animals that eat *S. habrochaites* fruits are not known, this species' habitat covers mountainous areas of Ecuador and Peru (Moyle, 2008), which harbors many potential mammalian and avian dispersers.

The next most basal clade of tomato species is made up of *S. peruvianum, S. corneliomulleri*, and *S. huaylasense* which are found in the desert habitat along the coast of Peru, but in environments that receive more rainfall than that of *S. pennelli* (Moyle, 2008). *Solanum chilense* is also a part of this clade but was not included in our analyses. Recent molecular evidence suggests that *S. huaylasense* may not be a true species, but rather a hybrid of *S. peruvianum* and *S. corneliomulleri* (Pease et al., 2016a). *Solanum peruvianum* and *S. corneliomulleri* have very high sucrose levels, though these species also show considerable variation, with some fruits containing the same amount or less sucrose as found in fruits from *S. huaylasense*. *Solanum huaylasense* has sugar levels more similar to *S. habrochaites* than to either of its putative parent species, both in accumulating lower levels of sugars and in reduced variability. *Solanum corneliomulleri* has the highest fructose levels in its ripe fruits of all species measured, but again with high variability like in its sucrose concentrations. Fructose is generally associated more with avian dispersers (Baker et al., 1998; Ramirez, 1990), and combined with the high levels of sucrose may suggest both mammalian and avian dispersers for this species. *Solanum huaylasense* and *S. peruvianum* are similar in firmness throughout ripening

while *S. corneliomulleri* softens a bit more, and ends with some of the softest fruits of green-fruited species at ripeness. However, none of the three species stands out in terms of its firmness from the bulk of the green-fruited species.

The last sub-clade of green-fruited tomatoes is largely capable of self-fertilization, like all the color-fruited species, but unlike the rest of the tomato clade (Rick, 1988). However, *S. arcanum* is not self-fertile (Li and Chetelat, 2014) and stands out in this study as it has the hardest ripe fruits measured, aside from the outgroup, and its fruits have very high levels of sucrose at the unripe stage and then accumulate more sucrose, ending with high sugar levels like *S. corneliomulleri* and *S. peruvianum*, and showing similarly high variance amongst fruits as well. The other members of this clade, *S. chmielewskii* and *S. neorickii*, have hard ripe fruits more similar to the other green-fruited species but accumulate less sugar than most species, *S. neorickii* has the lowest levels of sugar in ripe fruits out of the entire clade. All of these species are found in non-desert mountainous regions of Peru, and *S. neorickii* has the largest range of these species, which extends north into Ecuador (Moyle, 2008). Ripe *S. chmielewskii* fruits contain a bit more sugar, though only reach levels seen in *S. habrochaites*, and both species showed little variation in ripe sugar levels. *S. neorickii* fruits have several unique characteristics, in particular, when the fruit ruptures or is pierced, seeds will shoot out as though the inside of the fruit is pressurized. It is common for the fruits of many tomato species to rupture in times of high water availability, though in other species seeds typically do not escape the fruit at all in these instances (personal observation). I also noticed that *S. neorickii* fruits have a pungent putrid smell that was unlike the odor of any other fruit of this clade. These traits, together with the low sugar levels may suggest that this species is

not reliant on animal dispersers; it has a suite of traits that make the fruits unpalatable but may also enable the plants to spread their seeds via rupturing fruits, which would likely be linked to times of heavy rainfall that occur in its Andean range.

The color fruited group contains species that are separated by a body of water, with *S. pimpinellifolium* occurring on the mainland of South America and *S. cheesmaniae* and *S. galapagense* growing exclusively on the Galapagos Islands. Colored fruit is associated with animal seed dispersal and red fruits are more specifically associated with avian dispersal (Willson and Whelan, 1990). This provides a plausible hypothesis of how the progenitors of *S. cheesmaniae* and *S. galapagense* could have crossed 600 miles of ocean to colonize the Galapagos Islands. The colored species all shared multiple traits that are also observed in cultivated tomato. They all soften during ripening by more than 50%, while all green-fruited species soften by less than 50%, aside from *S. pennellii*, making this a trait with a strong phylogenetic correlation. Ripe fruits of color species are fairly high in fructose compared to other species, showing small increases through ripening, and they especially stand out because they accumulate large amounts of glucose, particularly *S. pimpinellifolium* and *S. galapagense*. Combined with very low sucrose levels, fruits of these species reflect the sugar palate preferred by birds (Baker et al., 1998). *S. cheesmaniae* stands out in that it has much less change in sugar levels during ripening than the other color-fruited species. This may be due to a loss of selective pressure on the fruits of this species to attract animal dispersers. This somewhat mirrors the colors of these species' fruits, where *S. pimpinellifolium* and *S. galapagense* fruits gain vibrant colors while *S. cheesmaniae* fruits show much more color variation with some just having a faint yellow color and some a cream color (see chapter 3).

Tomato species show a variety of ways in which fruits can fulfill the dual role of aiding both seed protection and seed dispersal. Throughout the clade, unripe fruits are camouflaged in green and most provide physical protection in the form of tough tissues, yet *S. pennellii* does not provide a mechanical defense, as even its unripe fruits are quite soft. These plants are covered in glaborous trichomes that can trap insects (personal observation) and may provide a way to present defense chemicals on the exterior of the fruit, deterring/killing predators without necessitating the skin of the fruit to be breached. Many more avenues are followed in helping the dispersal of seeds, along with fruit softening and sugar accumulation, which occur throughout the clade. The color species eponymously and uniquely have fruits that gain red and orange color as they ripen, removing the green camouflage and possibly creating strong contrast. Visual attraction is not the only tool for drawing in animal seed dispersers, since mammals are more driven by scents (Van der Pijl, 1969; Willson et al., 1989). Moreover, the increases in sucrose may be enough of a reward for dispersers to begin to seek out these green fruits.

The role that α-tomatine plays in the various tomato species is less clear; it has been shown to harm fungi and insects (Mulatu et al., 2006; Ökmen et al., 2013) but there are no studies suggesting that mammals or birds are affected strongly by its presence (Friedman et al., 1996). It appears that there is no strong selection for the destruction of α-tomatine in ripening fruits and this trait in modern tomatoes may have been selected for in domestication.

This study of phenotypic diversity in wild-tomato fruit traits lays groundwork for future research. Ecological studies can build off this when looking for correlations between fruit traits and natural dispersers or different abiotic conditions the plants

experience in their habitats. This information can also serve genetic studies by demonstrating the range of species variation, which fuels gene discovery. There is still little understanding as to what drives fruit trait diversity, but the characterization of these ripening traits in closely related wild species enables further understanding of the evolution of these traits.

Tables

Table 1 Sugar content in fresh fruit extract broken down by sugar type for unripe and ripe fruits. Representative values, in mg/ml, are from lme4 models for species based on all measurements taken. Each column is color coded on a spectrum with the largest values in blue and the smallest values in red.

	Unripe				Ripe			
Species	**glucose**	**fructose**	**sucrose**	**total**	**glucose**	**fructose**	**sucrose**	**total**
Sc	8.71	14.67	3.84	27.22	12	17.68	4.57	34.25
Sg	9.83	10.64	3.03	23.49	30.97	15.4	7.59	53.96
Sl	11.96	13.43	1.77	27.17	16.27	17.57	1.7	35.54
Slc	12.37	15.1	2.96	30.44	17.4	19.93	0.94	38.27
Spm	13.25	14.47	1.6	29.32	32.19	15.16	4.3	51.64
Sn	4.23	6.39	5.76	16.38	3.66	4.96	12.97	21.58
Sa	5.22	7.85	27.3	40.38	3.11	7.52	45.83	56.46
Scm	7.81	7.89	8.06	23.76	4.35	8.17	21.59	34.11
Scr	10.48	17.2	22.58	50.25	5.07	19.23	44.02	68.32
Shy	6.15	12.07	7.07	25.29	2.81	12.12	26.2	41.13
Spv	4.56	10.13	23.61	38.3	3.81	12.04	44.6	60.45
Sh	6.48	8.79	5.37	20.64	4.56	12.69	24.16	41.41
Spe	21.39	11.01	7.58	39.98	23.67	18.65	18.35	60.68
Sly	11.81	2.63	3.74	18.18	9.29	4.86	6.5	20.64

Table 2 Proportion of each sugar in ripe fruits of each species based on the predicted values in Table 1. Cells are color coded within rows to signify the main sugar (in green) the median sugar (in yellow) and the minor sugar (in red) for each species.

	% of total		
Species	**glucose**	**fructose**	**sucrose**
Sc	35%	52%	13%
Sg	57%	29%	14%
Sl	46%	49%	5%
Slc	45%	52%	2%
Spm	62%	29%	8%
Sn	17%	23%	60%
Sa	6%	13%	81%
Scm	13%	24%	63%
Scr	7%	28%	64%
Shy	7%	29%	64%
Spv	6%	20%	74%
Sh	11%	31%	58%
Spe	39%	31%	30%
Sly	45%	24%	31%

Table 3 Change in sugar content during ripening broken down by sugar type. Values calculated using predicted values in Table 1. Each column is color coded on a spectrum, with the largest values in blue and the smallest values in red for each sugar type.

	Change during ripening				Percent change			
Species	**glucose**	**fructose**	**sucrose**	**total**	**glucose**	**fructose**	**sucrose**	**total**
Sc	3.29	3.01	0.73	7.02	38%	21%	19%	26%
Sg	21.15	4.76	4.56	30.47	215%	45%	151%	130%
Sl	4.31	4.14	-0.08	8.37	36%	31%	-5%	31%
Slc	5.03	4.83	-2.03	7.83	41%	32%	-69%	26%
Spm	18.93	0.68	2.7	22.31	143%	5%	169%	76%
Sn	-0.57	-1.44	7.21	5.2	-13%	-23%	125%	32%
Sa	-2.11	-0.33	18.53	16.08	-40%	-4%	68%	40%
Scm	-3.46	0.28	13.52	10.34	-44%	4%	168%	44%
Scr	-5.41	2.04	21.44	18.07	-52%	12%	95%	36%
Shy	-3.34	0.05	19.13	15.84	-54%	0%	271%	63%
Spv	-0.75	1.91	20.99	22.15	-16%	19%	89%	58%
Sh	-1.92	3.89	18.79	20.77	-30%	44%	350%	101%
Spe	2.28	7.64	10.77	20.7	11%	69%	142%	52%
Sly	-2.53	2.23	2.76	2.46	-21%	85%	74%	14%

Table 4 Results from PIC analysis of sugars based on lme4 model output, including change in the sugar concentrations during ripening and the percentage change from unripe value during ripening. Sl and Slc are not included for PIC analysis as these are known to have undergone artificial selection.

Sugar	**Stage**	**K**	**λ**
Glucose	unripe	0.719	0.999*
	ripe	0.217	0.539
Fructose	unripe	0.475	0.734
	ripe	0.907*	0.960*
Sucrose	unripe	0.410	0.000
	ripe	0.585	0.570
Total	unripe	0.409	0.000
	ripe	0.264	0.000
ΔGlucose		0.174	0.444
ΔFructose		0.654	0.864
ΔSucrose		0.855*	0.867*
ΔTotal		0.094	0.000
ΔGluc%		0.171	0.476
ΔFruc%		0.734*	0.822
ΔSuc%		0.258	0.000
ΔTotal%		0.077	0.000

*indicates p-value < .05

Table 5 Firmness values in grams of force for species from lme4 model output, including change in firmness during ripening and the percentage change from unripe value during ripening. Each column is color coded on a spectrum with the largest values in red and the smallest values in blue.

Species	Unripe Firmness	Ripe Firmness	Change	Percent Change
S. lycopersicum	250.9299	139.0766	-111.853	-44.6%
S. lycopersicum var. cerasiforme	299.7839	145.3827	-154.401	-51.5%
S. cheesmaniae	301.7047	127.2839	-174.421	-57.8%
S. galapagense	250.775	104.9666	-145.808	-58.1%
S. pimpinellifolium	314.5426	135.4662	-179.076	-56.9%
S. neorickii	303.7335	174.7285	-129.005	-42.5%
S. arcanum	309.5283	209.1379	-100.39	-32.4%
S. chmielewskii	272.249	169.2674	-102.982	-37.8%
S. corneliomulleri	256.8578	149.0352	-107.823	-42.0%
S. huaylasense	268.6337	179.1026	-89.5311	-33.3%
S. peruvianum	253.322	168.8899	-84.4321	-33.3%
S. habrochaites	354.591	192.8753	-161.716	-45.6%
S. pennellii	156.9274	62.78298	-94.1444	-60.0%
S. lycopersicoides	284.2864	205.5621	-78.7243	-27.7%

Table 6 Results from PIC analysis of firmness data from table 5. Sl and Slc are not included for PIC analysis as these are known to have undergone artificial selection.

Stage	K	λ
Unripe	0.385	0.810
Ripe	0.645	0.951
Change	0.604	0.832
Percent change	1.084**	1.000**

** indicates p-value < .01

Table 7 α-Tomatine values (absorbance per mg fresh weight) for species from lme4 model output, including change in α-tomatine during ripening and percent change from unripe value during ripening. Each column is color coded on a spectrum with the largest values in red and the smallest values in blue.

Species	Unripe	Ripe	Change	Percent change
S. lycopersicum	7367	8095	728	10%
S. lycopersicum var. cerasiforme	8137	8954	817	10%
S. cheesmaniae	25194	42678	17484	69%
S. galapagense	46267	58319	12052	26%
S. pimpinellifolium	34366	42121	7755	23%
S. neorickii	16157			
S. arcanum	19021	60917	41896	220%
S. chmielewskii	70457	23509	-46947	-67%
S. corneliomulleri	13210	98764	85554	648%
S. huaylasense	18602			
S. peruvianum	42574	95290	52716	124%
S. habrochaites	85084			
S. pennellii	36698	126814	90116	246%
S. lycopersicoides	119871	135761	15889	13%

Table 8 Number of individuals used for α-tomatine measurements at ripe and unripe stages.

Stage	Sl	Slc	Sc	Sg	Spm	Sn	Sa	Scm	Scr	Shy	Spv	Sh	Spe	Sly
Unripe	2	5	6	6	3	13	9	1	4	4	5	7	7	7
Ripe	2	3	5	7	7	0	5	6	5	0	2	0	2	4

Figures

Figure 1

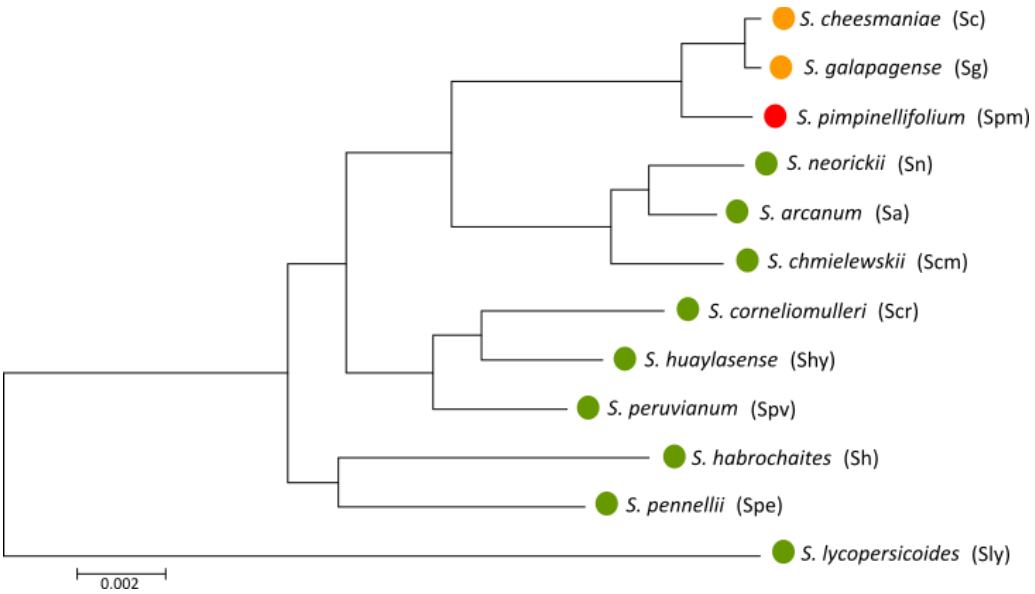


Figure 1 Phylogeny created from Pease et al. (2016) data showing the wild tomato clade with an outgroup. Ripe fruit color is indicated by accompanying colored dots; species codes are used in other figures; and areas of the text follow the species name in parentheses. *Solanum chilense*, a sister species of *S. peruvianum*, was left off this tree as it was excluded from our analyses. *Solanum lycopersicum* (Sl) and *S. lycopersicum var. cerasiforme* (Slc), sister species to *S. pimpinellifolium*, were not included in the phylogeny as they are the result of artificial selection.

Figure 2

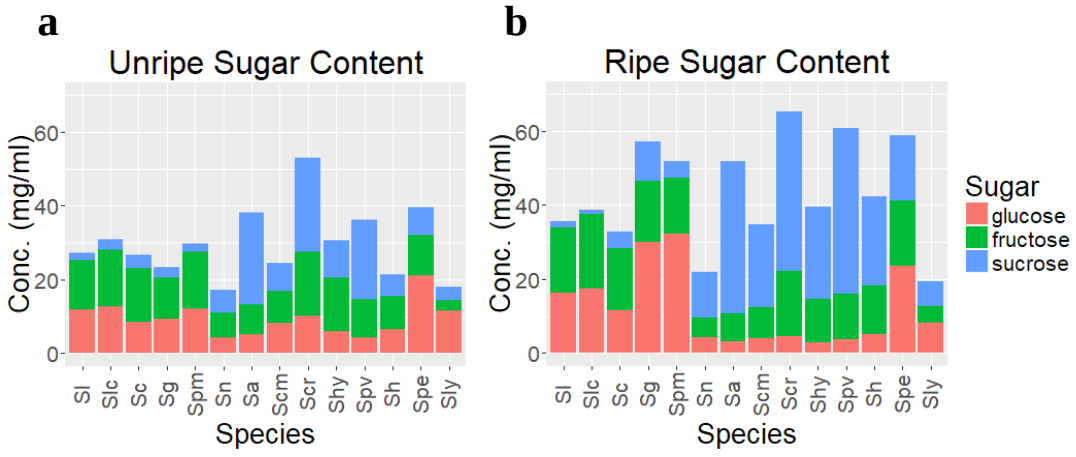
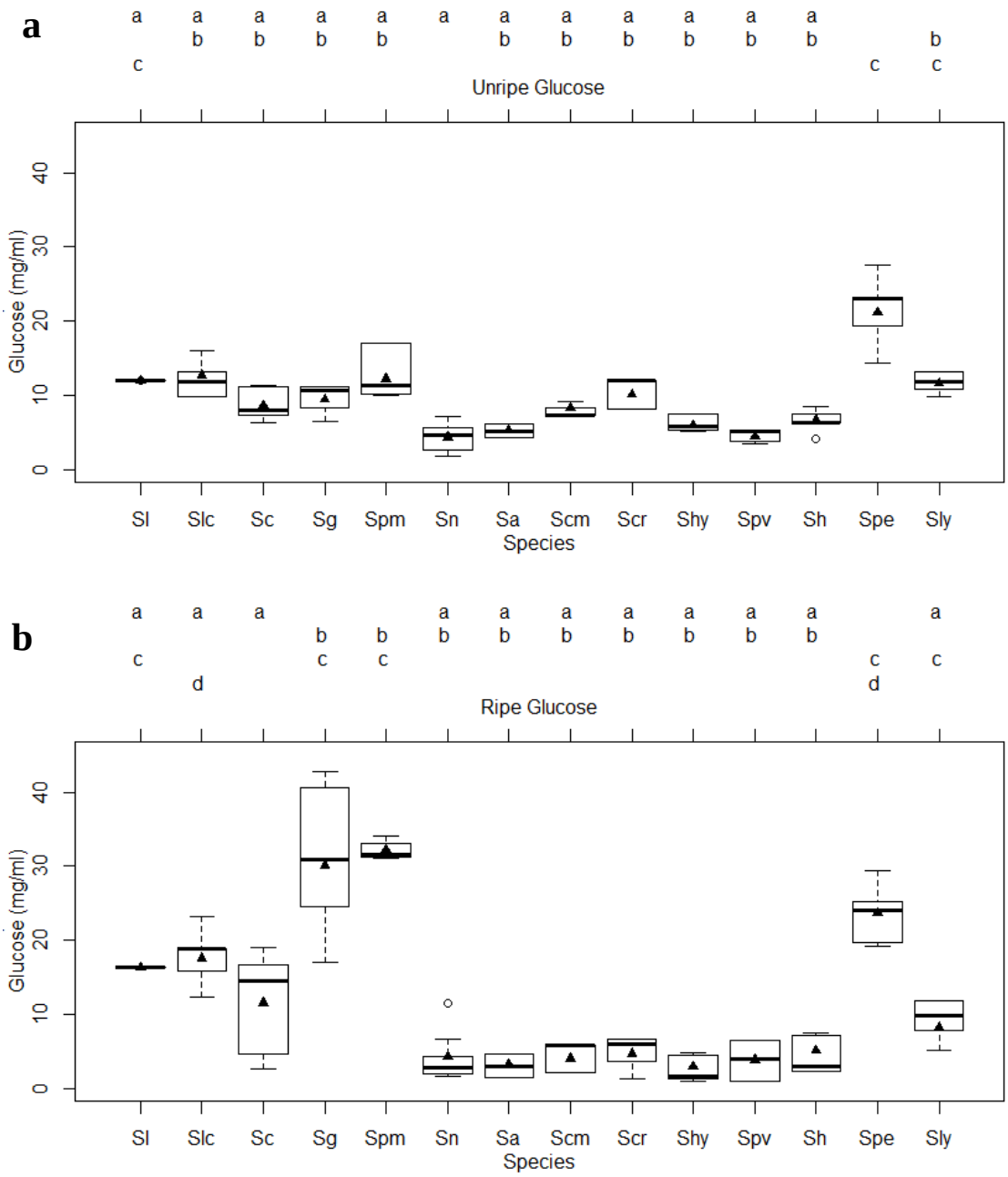
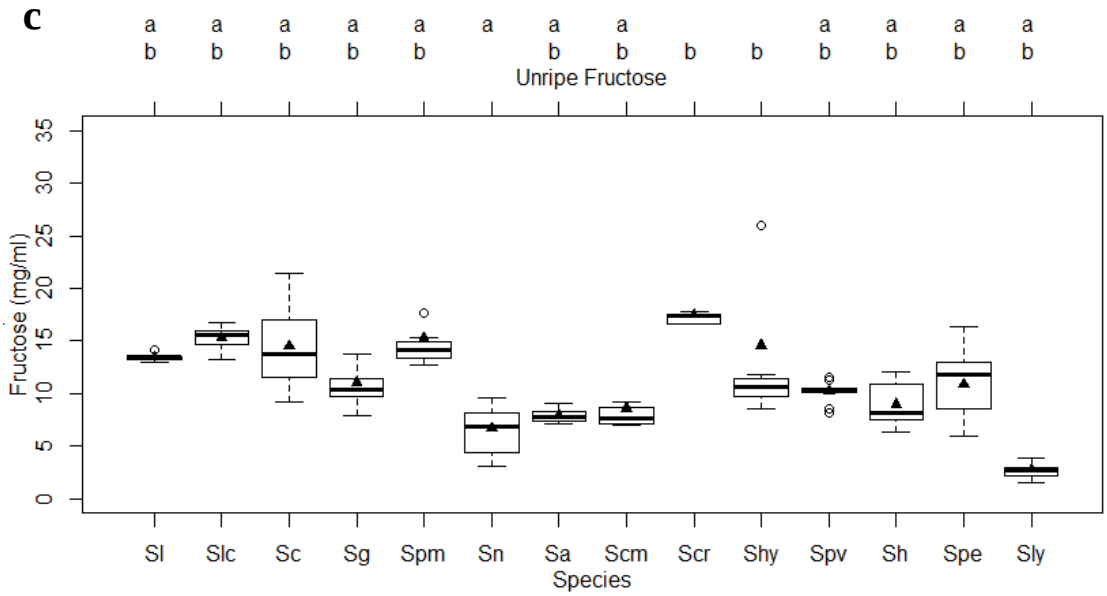


Figure 2 Sugar concentrations based on lme4 model predicted values for each species at unripe (a) and ripe (b) stages. Species codes are shown in figure 1.

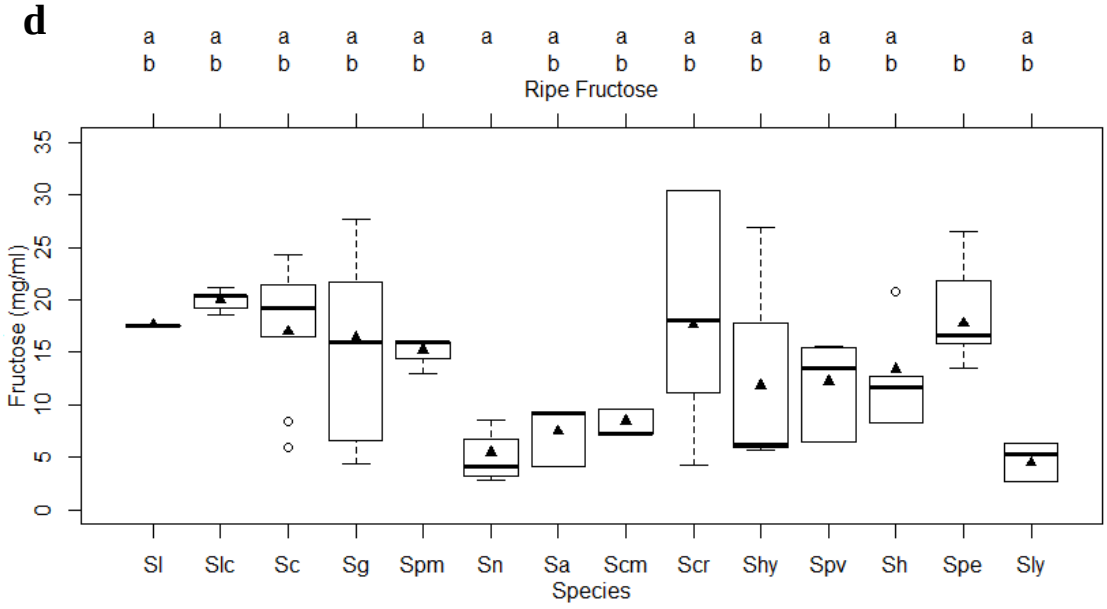
Figure 3


a
Unripe Glucose
Glucose (mg/ml)
0
10
20
30
40
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species
b
Ripe Glucose
Glucose (mg/ml)
0
10
20
30
40
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species

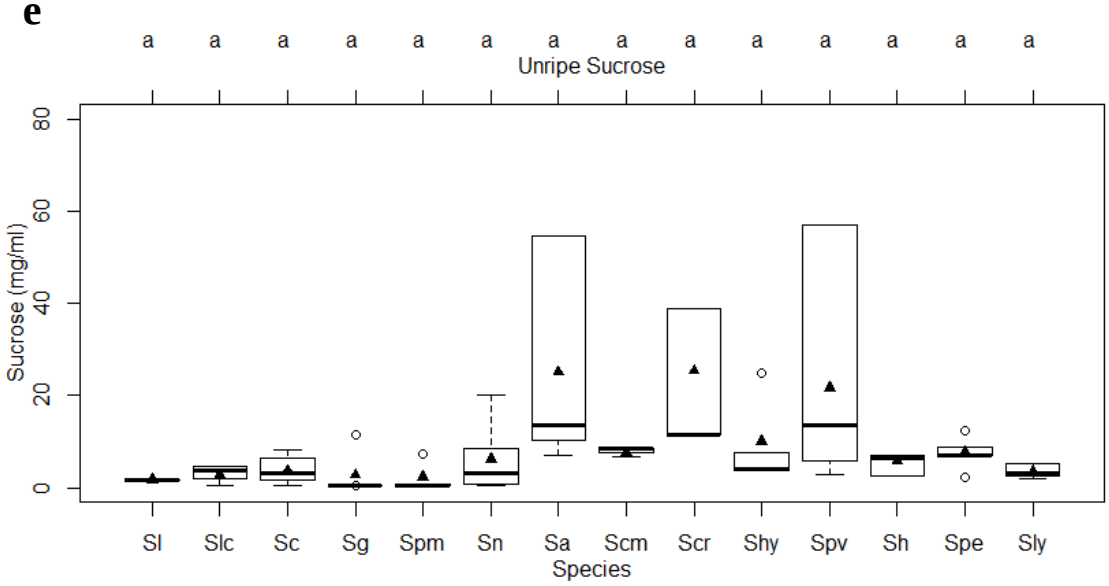
c


a a a a a a a a b b b b b b b b
Unripe Fructose
Fructose (mg/ml)
0 5 10 15 20 25 30 35
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species

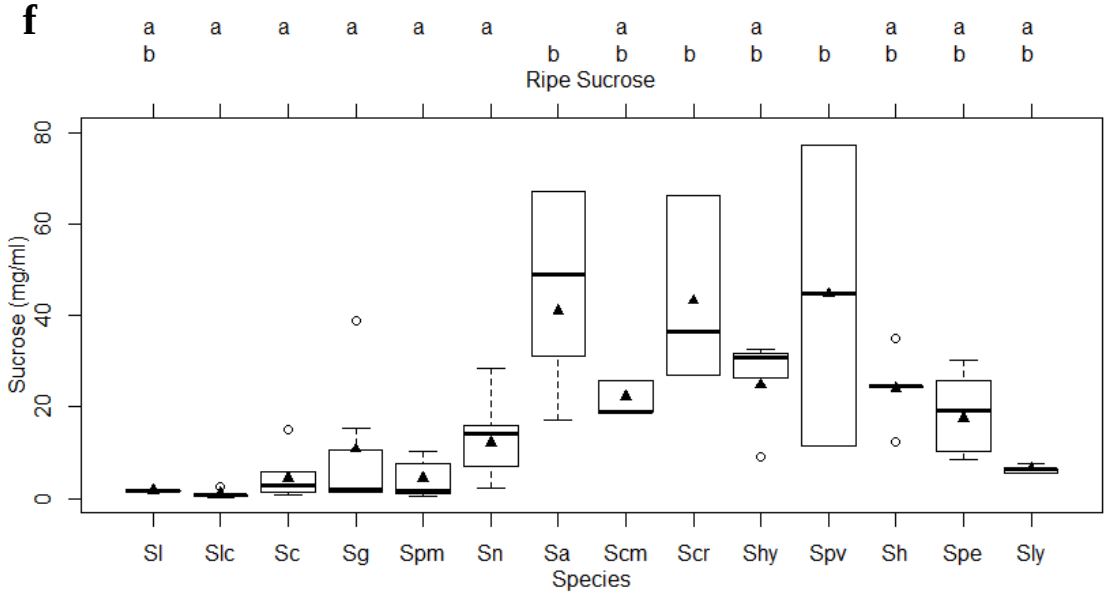
d


a a a a a a a a a a a a a a a a
Ripe Fructose
Fructose (mg/ml)
0 5 10 15 20 25 30 35
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species

e


a a a a a a a a a a a a a a
Unripe Sucrose
Sucrose (mg/ml)
0 20 40 60 80
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species

f


a a a a a a a a a a a a a a
b b b b b b b
Ripe Sucrose
Sucrose (mg/ml)
0 20 40 60 80
Sl Slc Sc Sg Spm Sn Sa Scm Scr Shy Spv Sh Spe Sly
Species

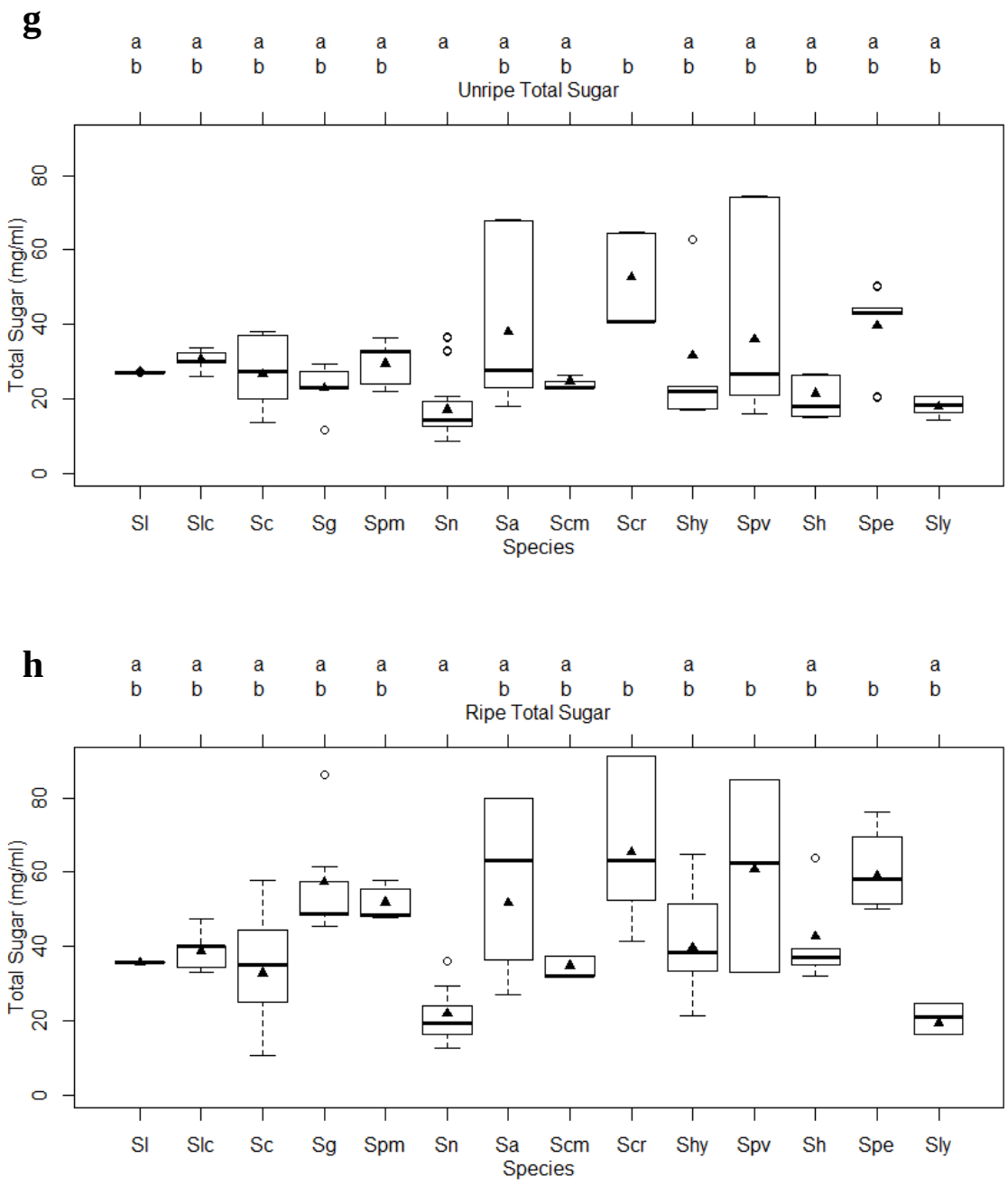


Figure 3 Box and whisker plots showing all sugar data of unripe and ripe fruits for each sugar separately (a-f) and total sugars(g-h). Graphs are scaled the same for the stages of each sugar, but not across sugars. Results from Tukey's HSD are listed along the top using the compact letter display, species with non-overlapping letters are significantly different. Predicted fruit values from lme4 model are indicated with filled triangles.

Figure 4

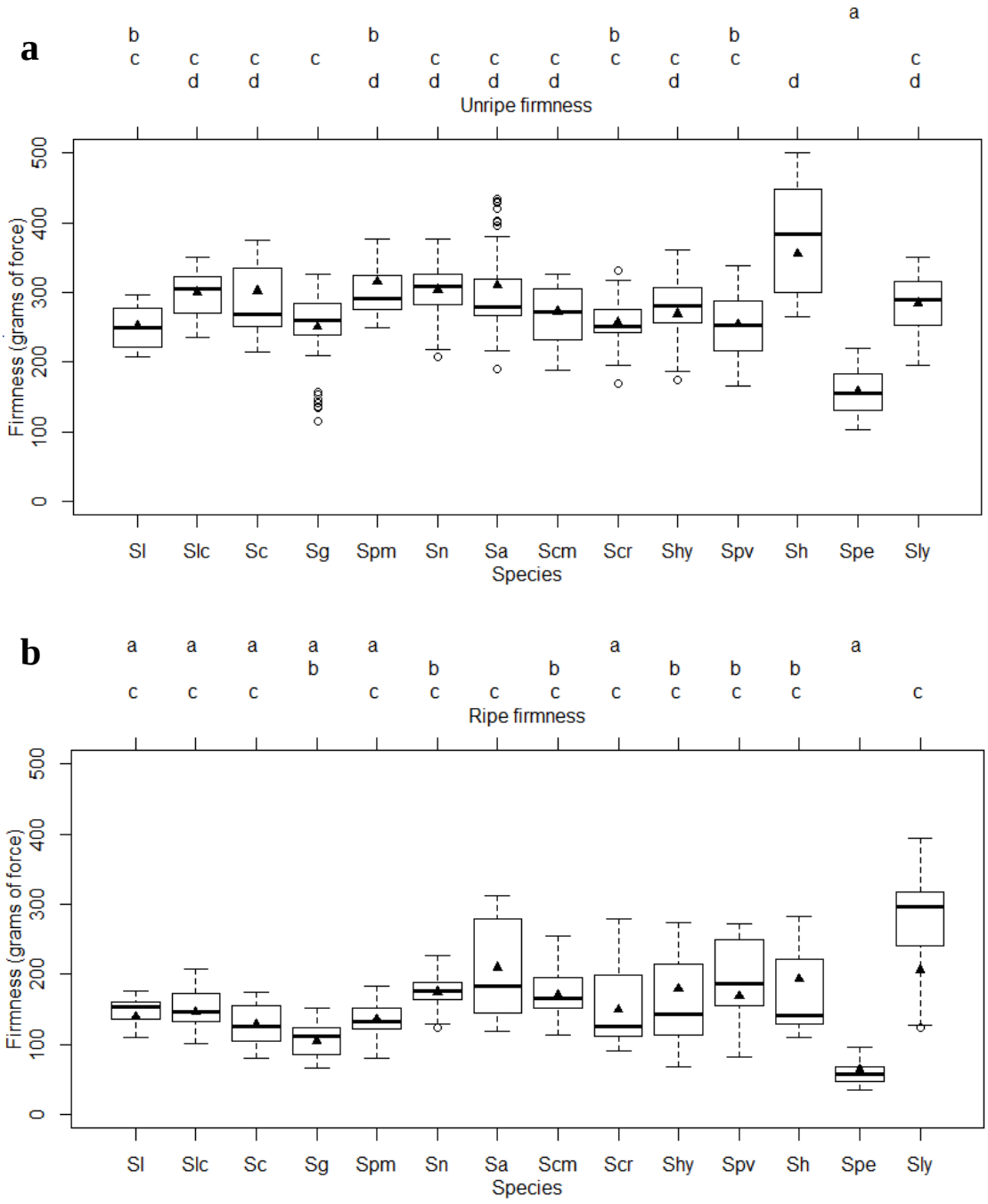


Figure 4 Box and whisker plots showing all firmness data of unripe (a) and ripe (b) fruits. Results from Tukey's HSD are listed along the top using the compact letter display, species with non-overlapping letters are significantly different. Predicted fruit values from lme4 model are indicated with filled triangles.

Figure 5

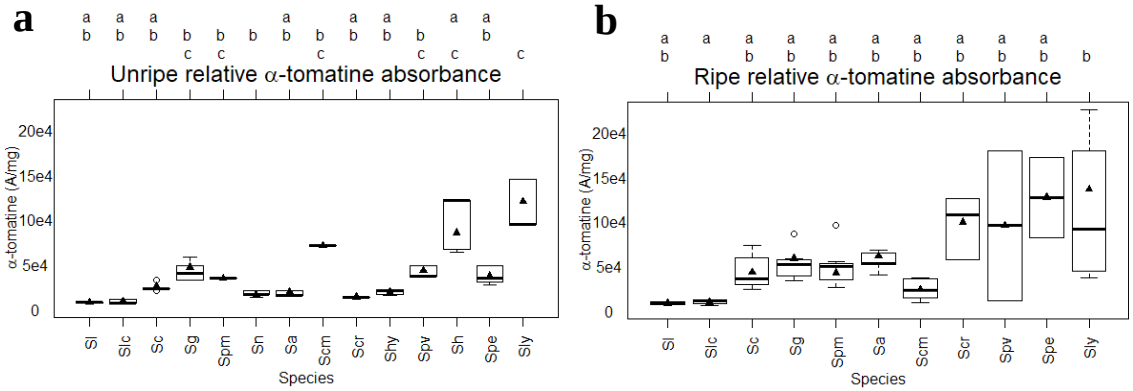


Figure 5 Box and whisker plots showing all α-tomatine data of unripe (a) and ripe (b) fruits. Species codes are shown in figure 1. Results from Tukey's HSD are listed along the top using the compact letter display; species with non-overlapping letters are significantly different. Predicted fruit values from lme4 model are indicated with filled triangles

CHAPTER 2

MAPPING OF FRUIT SUGAR CONTENT AND FIRMNESS TRAITS USING INTROGRESSION LINES

Introduction

Fleshy fruits provide a structure that can help protect vulnerable seeds and attract dispersers for mature seeds. The firmness of the fruit tissue plays a large role in both of these functions. Firm fruit tissue can be difficult to get through for seed predators, while soft fruits can be easier to eat and more appealing to vertebrate seed dispersers, especially when sugar accumulation coincides. However, to function as both protector and attractant the fruit firmness and sugar content need to change along with the maturity of the seeds, in the familiar process known as fruit ripening. In the wild, the firmness and flavor of fruits at either stage, as well as the change during ripening may be subjected to selection, as these traits can have a strong effect on survival and seed dispersal and thus the reproductive success of plants.

The wild tomato clade in the *Solanum* genus shows a great deal of variation in firmness and softening, as explored in chapter 1. This includes species that tend to stay hard throughout development like *S. arcanum* and the tomato clade's outgroup, *S. lycopersicoides*, species that are remarkably soft throughout development, such as *S. pennellii*, as well as species with very firm unripe fruits but soft ripe fruits, such as *S. pimpinellifolium* and *S. galapagense*. Regardless of the softness of the fruits, all species undergo some amount of softening as the fruits ripen, but the genes controlling this phenotype and whether this variation is related to expression of softening genes or functionality of the proteins are all unknown.

Similarly, the variation in total sugar accumulation as well as which sugars wild tomato species accumulate was documented in chapter 1. Within this clade there are some species, such as *S. neorickii* and the outgroup, *S. lycopersicoides*, that have very little sugar when unripe and do not accumulate much during ripening; some species, such as *S. corneliomulleri*, have large amounts of sugar and continue to accumulate during ripening; and some that have a typical amount of sugar when unripe, but then accumulate large amounts, such as *S. galapagense*. There is also variation in the type of sugar accumulated, with color-fruited species accumulating primarily glucose, while green-fruited species typically accumulate primarily sucrose. Additionally, fructose is depleted in some species through ripening while it is accumulated in others, though fructose does not dominate in the ripe fruits of any species. As an outlier, *S. pennellii* seems to not favor any particular sugar with an almost even split of the three upon maturity. Though many of the pathways involved in sugar production and use in cultivated tomato are known (Bianchetti et al., 2017; Chetelat et al., 1993; Klann et al., 1993; Miron et al., 2002), the different genes and pathways used by wild fruits to produce these varying sugar profiles remains an area for study.

In the study of fleshy fruit development, domesticated tomato has been used as a model organism due to the small stature and short life cycle of some tomato cultivars. Since other species in the clade encompassing cultivated tomato also show variation in fruit firmness and softening, tomato is also an ideal system in which to study the evolution and genetics of softening and sugar accumulation. Tomatoes belong to the very large *Solanum* genus which contains approximately 1,500 species, including other major crops like potato (*S. tuberosum*) and eggplant (*S. melongena*), as well as many regionally

important crops such as naranjilla (*S. quitoense*) and nakati (*S. aethiopicum*). The genus exists worldwide and one of its biodiversity hotspots is found in western South America where the ~2.5 myo tomato clade is endemic (Pease et al., 2016a).

Tomato fruit flesh can become softer due to action of several different classes of enzymes involved in the degradation of various cell wall components (Carey et al., 2001; Tomassen et al., 2007; Wang et al., 2009). Many genes involved in ripening-related softening have been studied in cultivated tomatoes as this is an important trait for the modern world, requiring fruits strong enough to withstand shipping while still having a pleasant mouth feel. This research has shown that softening in tomatoes is a quantitative trait, where no gene is alone responsible for the phenotype. This provides numerous sites of potential genetic variation for natural selection to act on in the evolution of the phenotypic variation in this clade.

The accumulation of different sugars in different amounts could be achieved in many ways, including transport of sugar into fruits (Reuscher et al., 2016), production of sugars in the fruits (Powell et al., 2012), converting between sugar types of those already present (Klann et al., 1993), or combinations of these methods (Miron et al., 2002). Sugar composition of tomato cultivars has been well studied since it affects the taste profile of tomatoes and products made from them, and allows different cultivars to be used for different purposes, with some intended to be cooked and sauced and others meant to be eaten raw. However in the wild the sugar profiles can have a dramatic effect on frugivore preference (Baker et al., 1998; Floerchinger et al., 2010; Ramirez, 1990), which can determine the successful dispersion and germination of their seeds.

Using a set of introgression lines (ILs) created from phenotypically differing species can allow us to find genomic regions that may be associated with evolution of fruit softening and sugar accumulation during ripening. Introgression lines are a collection of plants that have been crossed and then backcrossed to one parent until only a fraction of a line's total genome comes from the outgroup. The introgression lines used in this study are from a cross between cultivated tomato, *S. lycopersicum*, and the outgroup to the wild tomato clade, *S. lycopersicoides*, backcrossed into *S. lycopersicum* (Canady et al., 2005; Rick et al., 1988). The parental species show very different proclivity for softening through ripening, with fruits from *S. lycopersicoides* softening little, if at all, as seen in chapter 1, while cultivated tomato is well known for its softening. We also saw in chapter 1 that ripe *S. lycopersicum* fruits have very little sucrose with roughly equal glucose and fructose, while ripe *S. lycopersicoides* fruits have roughly equal amounts of all three sugars but accumulate only about half as much total sugar. For any trait, variation in phenotype among ILs from the parental values can be tentatively attributed to genes in the specific genomic regions that the introgressed line possesses. The full set of introgression lines gives complete coverage of the *S. lycopersicoides* genome as can be seen in Table S1; growing all lines together permits identification of lines with fruits that soften and accumulate sugars differently from the parents, and directly connect those differences with the introgressed genomic regions. This highlights promising areas to investigate further to find genes that may have functional or regulatory variances causing the differences in these phenotypes. The phenotypic differences along with the phylogenetic distance between these parents makes

this an excellent resource for finding regions of the genome associated with various phenotypes.

The aim of this study is to identify candidate genes potentially involved in fruit softening and sugar accumulation variation within this clade. Both phenotypes are the result of complex processes involving regulators, transcription factors and enzymes working together but in various different ways. Understanding which genes may be regulated differently or have different function between *S. lycopersicoides* and *S. lycopersicum* can inform future studies on fruit softening, sugar accumulation and the evolution of fruit traits in the wild.

Materials and Methods

Plants and growth conditions

Seeds of all 56 plants of the primary *S. lycopersicum* x *S. lycopersicoides* introgression lines (Rick et al., 1988) as well as the parental genotypes, LA2951 and LA0490, were obtained from the UC Davis C.M. Rick Tomato Genetics Resource Center, maintained by the Department of Plant Sciences, University of California, Davis, CA. For the introgressed parent, *S. lycopersicoides*, we grew the same accession used for the cross but were unable to get any fruits, so the parental value is represented by values from other individuals of the same species (accessions LA1966, LA2407, LA4322). For *S. lycopersicum* the parental accession, LA0490, was grown. A single replicate of each line was grown in each of three blocks in the College of Natural Sciences Greenhouse at the University of Massachusetts Amherst campus. There were two growing cycles starting in February of 2013 and 2014 and grown through the summer. Lines were

randomized within each block. The temperature was set to 26°C during the day and 22°C at night, humidity was kept below 60% and day length was kept at 12 hours. Plants were grown in two-gallon pots, weighted with gravel in the bottom and filled with Sunshine LB 2 soil. Starting after one month of growth, plants were fertilized every week alternating between 15-5-15 high calcium-magnesium fertilizer at 300 ppm nitrogen to promote healthy plant and fruit growth and 10-30-20 fertilizer at 200 ppm nitrogen to promote flowering.

Fruit collection and measuring firmness

Fruits were collected at the mature green stage, referred to here as unripe, and at the fully ripe stage (https://www.ams.usda.gov/grades-standards/tomato-grades-and-standards). We attempted to collect five fruits from each plant at each maturity stage; however, certain lines had difficulty growing, flowering and setting fruits. This difficulty, due to the unusual genetic background of the lines, has been reported elsewhere (Canady et al., 2005; Chetelat et al., 1989). A second growing season was required to grow these individuals to get better fruit set, though some lines still failed to produce usable fruits. In total, we were able to collect fruits at the unripe and fully ripe stage in all but one line at each stage; IL5 is not represented in unripe measures and IL 49 is not represented in ripe measures. Up to 16 fruits were collected for any one line; exact numbers used per line per stage are shown in Table S2.

Firmness was measured at time of collection using an Imada DS2-110 digital force gauge with a blunt 0.5mm^2 thin metal probe attached. The probe was then used to pierce to the center of the fruit from four different directions; the peak force from these

four measurements were averaged to give a single value for the firmness of the fruit. Fruits were then halved and seeds were removed; the remaining tissues were frozen. Frozen tissue was ground using a ball mill to a fine powder for use in analyses of α-tomatine content and sugar content. However, α-tomatine measurements were later discarded due to problems with consistency of the technique.

Sugar content measurements

A target sample size of one fruit at each stage of ripeness from plants of every line in each of three replicate blocks was set for sugar measurements. Unfortunately, due to poor fruit set and substantial loss of collected fruits, and because ground tissue was divided among tomatine and sugar measurements, some lines had no fruits measured, several lines had only one ripeness stage represented, and most lines had only one total fruit measured at each stage (Table S2). Additionally, the fruits that were measured for sugars came from the various replicate blocks, as no single block provided enough fruits to obtain measurements of every line. Due to these issues, the statistical measures are more suggestive than representative.

To measure sugar content of tomato fruits, the ground frozen tissue was thawed in a sonicating water bath to ensure any clumps that may have formed were broken up. The solid and liquid phases were separated using a centrifuge at 8,000 rpm for 5 minutes. A total of 2µl from the liquid phase was used directly for sugar measurements with the R-Biopharm sugar kit (Roche Yellow line Sucrose/D-Glucose/D-Fructose enzymatic kit), using the modified method by (Velterop and Vos, 2001) to accommodate the small volume of liquid available from wild tomato fruits. Briefly, the kit allows for measuring

of sugar levels indirectly, based on the amount of NADP formed during the enzymatic reduction of each sugar. Levels of glucose, fructose, and sucrose were measured at both stages of ripeness for all samples available.

Data analysis

Analysis of the IL fruit traits was done in R (R Core Team, 2015). For trait analysis the lme4 package (Bates et al., 2015) was used to create a linear mixed model, using the lmer function which can account for many sources of random variation in the input data. The lme4 package readily handles multiple sources of random variation; Our models accounted for individuals of the same line, block by individual, and fruits of a line by block. These models provide a value representative of the trait for each introgression line based on the values measured from all individuals and fruits of that line; these values were used as input for genome scanning. Analyses were done comparing all lines at both ripeness stages for each trait.

Mapping was done in R using the 'qtl' package (Broman et al., 2003). The genotype map with phenotype data was read in using the read.cross function. The data was then processed with the function calc.genoprob to run a hidden Markov model, the output of which can be used for genome scans. The genome scan was done with the function scanone with the default EM maximum likelihood algorithm to get marker $\log_{10}$ of odds (LOD) scores. To get a genome-wide LOD significance threshold for each dataset, the scanone function was run with 1000 permutations. Potential genes of interest were gathered from reports in the literature of their role in fruit softening or sugar accumulation, or if they were thought to have products involved in cell wall

reorganization or sugar use/synthesis, these lists were pared down focusing on regions of interest once those were known.

Results

Firmness in introgression lines

The fruits of the parental introgressed outgroup species, *S. lycopersicoides*, lose 26% of their firmness during ripening, and their ripe fruits are the firmest measured in this study (Fig. 6, Table S3). The *S. lycopersicum* parent has fairly firm unripe fruits which lose nearly 45% of their firmness, resulting in fairly soft ripe fruits. The unripe firmness variation of the introgression lines (ILs) was outside the range of the parents (Fig. 6a); the firmest unripe fruits were seen in IL25, which required 343 grams of force to pierce, and the softest were seen in IL50, whose unripe fruits required only 167 grams of force to pierce, making this fruit softer than some of the ripe fruits measured. The ripe firmness of the ILs did not exceed *S. lycopersicoides'* 205 grams of force (Fig. 6b). IL16 had the hardest ripe fruits of the ILs at 181 grams of force (Table S3). IL4 had the softest ripe fruits, requiring only 77 grams of force to pierce, which is nearly half the firmness of the *S. lycopersicum* parent's ripe fruits, though only a single fruit was grown to ripeness in this line. The softest well-represented line is IL22 at 99 grams of force. There is only a slight, though significant, correlation between unripe and ripe fruit firmness, with Pearson's correlation coefficient r of 0.3525 (Fig. 7), suggesting that unripe and ripe fruit firmness are under different genetic controls. Percentage softening variation also exceeded the parental range, from only 20% loss of firmness in IL16 up to 69% in IL4 (Table S3).

Firmness genome scan

In genome scans looking at fruit firmness, no association between unripe firmness and introgressed region was found (Fig. 8a), although three peaks stood out. The first is a region at the end of chromosome 2, from marker TG151 to the end of the chromosome, covering approximately 5Mb. The tallest peak is in the middle of chromosome 5, including markers TG358 and TG23, and covering 15Mb after the centromere. The final unripe peak is on chromosome 9 around marker CT143, which covers approximately 1.4Mb near the beginning of that chromosome.

For ripe fruits, only chromosome 3 showed significant association with firmness (Fig. 8b). A 6Mb portion including markers TG284 and TG152 on chromosome 3 was linked with firmer ripe fruits (Fig. 9a). One end of chromosome 5 from marker TG69 to the end of the chromosome, representing approximately 3.5Mb, came close to significance in the genome scan with a LOD of 4.41 where the p-value < 0.05 cutoff was at 4.47 (Fig. 9b).

Sugars in introgression lines

The IL parent species are quite different in sugar accumulation: *S. lycopersicoides* has low sucrose levels, like color-fruited species, low glucose, like green-fruited species, and the lowest fructose of all tomato species; *S. lycopersicum* accumulates primarily fructose and glucose as is typical for all color-fruited species, as seen in chapter 1. *Solanum lycopersicoides* has glucose concentrations at 12.2 mg/ml in unripe fruit and 8.8 mg/ml in ripe fruit, which is low compared to the ILs' average (Fig. 10a-d, Table S4); sucrose at 4 mg/ml in unripe fruit and 7.1 mg/ml in ripe fruit, which is higher than the IL

average; and it consistently has the absolute lowest fructose concentration at 2.6 mg/ml in unripe fruit and 5.1 mg/ml in ripe fruit, despite showing the largest percent increase in fructose during ripening at 93%. This low fructose also contributed to it having some of the lowest total sugar concentrations at 19.1 mg/ml in unripe fruit and 21.6 mg/ml in ripe fruit. The other parental species, *S. lycopersicum*, had concentrations for all sugars that were a bit under the ILs' average for unripe fruit sugar levels though its ripe sugar concentrations were very close to average (Fig. 10e-h) for most sugars due to slightly above average percentage increases. Sucrose decreased in *S. lycopersicum* during ripening, resulting in one of the lowest concentrations, though many ILs had undetectable sucrose concentrations.

In unripe fruits, the highest glucose concentration measured was in IL53 at 20.4 mg/ml (Table S4) and the lowest were in ILs 43, 37, 35, and 44 with roughly half the concentration of IL53. The highest fructose concentrations for unripe fruits were in ILs 36, 31, 29, 51 and 8 at 17.7-17 mg/ml, the lowest fructose concentration in the ILs was 10.1 mg/ml in IL2 which was still well above the 2.6 mg/ml found in the parental species *S. lycopersicoides* (Fig. 10b). Sucrose levels were very low in the ILs, with many lines having undetectable concentrations; this makes sense given the *S. lycopersicum* background and its low sucrose levels, but since several of these were measured as having negative concentrations for sucrose, low levels could also be due to errors in the measurements or the way sucrose concentration is calculated by subtracting glucose and fructose concentrations from total sugar concentration. Despite this there were also some measurements of very high concentrations in ILs, with IL37 containing 11.3 mg/ml and IL1 9.5 mg/ml. The highest concentration of total sugar in unripe fruits was found in

IL52, which did not have high levels of any particular sugar, but moderate concentrations of each sugar giving an overall concentration of 37.4 mg/ml. The lowest total sugar concentration for unripe IL fruits was in IL35 at 21.4 mg/ml, which was just slightly above the *S. lycopersicoides* concentration (Fig. 10d).

In ripe fruits, IL53 still had the highest glucose concentration at 27.7 mg/ml (Table S4), but IL48 and IL52 were close behind. The lowest levels were found in IL51 at only 2.4 mg/ml of glucose, and this IL also had the lowest concentration for all sugars at the ripe stage. For fructose, IL23 had the highest concentration at 25.8 mg/ml but ILs 33, 30 and 28 also had concentrations over 24 mg/ml. IL51 again had the lowest concentration of fructose in the ILs at 5.2 mg/ml, just above the 5.1 mg/ml found in *S. lycopersicoides* (Fig. 10f). For sucrose in ripe fruits, both IL30 and IL9 had concentrations over 15 mg/ml, but many more lines showed no sucrose present at the unripe stage, which is consistent with most lines presenting the typical *S. lycopersicum* phenotype of breaking down sucrose during ripening. The highest overall sugar concentrations were found in IL9 at 52.1 mg/ml and IL30 at 51.6 mg/ml, both of which had very high fructose and sucrose levels but low glucose levels. The lowest total sugar in ripe fruits was found in IL51, which had the lowest measures of all sugars, at 8.2 mg/ml; this was less than half the total sugars of the next lowest line, IL22 which had a concentration of 18.9 mg/ml of total sugar, but did have measurable sucrose concentration. The Pearson correlation coefficients comparing the various sugars and stages are collected in Table S5. Fructose concentrations across stages had significant correlation, as did sucrose concentrations (Table S5a), though there is still variation in the amount of change during ripening in these traits. Glucose and fructose were correlated

significantly with total sugar at all stages (Table S5b-e), and change in concentration during ripening was correlated between glucose and fructose.

Sugars genome scan

The LOD scores for sugar concentrations in unripe fruits are shown in Figures 11a-d. The only measure that had significant LOD scores in the unripe fruits was that of total sugar (Fig. 11d). This measure had a significant region on chromosome 3 for the marker TG479 (Fig. 12a), which covers 1.3Mb on the end of the chromosome, and on chromosome 8 from the beginning of the chromosome to marker TG41 (Fig. 12b) covering an entire chromosome arm, approximately 59Mb. There was also a large, though not significant peak for sucrose concentration on a different region of chromosome 8 at marker TG510, representing a region covering approximately 3.3Mb. There was an unusual result for the unripe fructose measure, where most of the genome came back with very high LOD scores, though the significance limit was also quite high such that no regions reached significance for this trait.

The LOD scores for sugar concentration in ripe fruits are shown in Figures 11e-h. In ripe fruits, there were significant LOD scores for measures of fructose concentration (Fig. 11f) as well as total sugar concentration (Fig. 11h). Fructose concentration had significant regions on chromosome 5 at marker TG432 (Fig. 12c), covering 3Mb, and on chromosome 11 from marker TG46 to the end of the chromosome (Fig. 12d), covering over 20Mb. The total sugar concentration in ripe fruit was also significant for the same marker on chromosome 5 (Fig. 12e) and had a large though non-significant peak corresponding to the same region of chromosome 11, suggesting that significance of

LOD scores for fructose levels is driving those of the overall sugar level. There was a peak in ripe sucrose concentration but the LOD score is less than half the significance cutoff; the peak stands out only because so many of the fruits measured had no sucrose and thus there was a lack of association throughout most of the genome.

Discussion

The introgression lines in this study showed quite a bit of transgressive segregation, so rather than the parental species defining the extremes, trait values frequently were exceeded by the ILs on either side of the distribution range. This is an occurrence well studied in tomato (deVicente and Tanksley, 1993; Rick and Smith, 1953), with several proposed mechanisms. Those most applicable to our IL population, due to how related the lines are and how they were created, likely are 1) complementary alleles from the parental species that drive a trait toward the same extreme and 2) the potential for "overdominant" alleles, where a gene in a new background may be freed from epistatic controls allowing the gene to drive the trait without constraint. Further work would be needed to determine which mechanism was involved with the traits tested here and loci identified. In a previous study the second mechanism was found to occur for only 7% of traits and loci for other tomato traits (Rick and Smith, 1953), though nothing can be assumed for our traits. The first mechanism can occur in numerous ways; the newly introduced region may bring stronger regulators, alleles more responsive to regulation, or alleles with similar action being brought together. It is also important to keep in mind that the value used here for the fruits of *S. lycopersicoides* does not come from the parental accession but is a representative value for the species.

We attempted to determine if known candidate genes occurred in any of the regions we identified as significantly associated with the traits studied. Firmness of unripe fruits did not have any regions with significant LOD scores, and the three that came close to significance do not contain any known firmness genes (Table S11). Ripe fruit firmness did have a significant and nearly significant marker LOD score due to some lines with particularly firm fruits (Fig. 8b, Fig. 9). The introgression line with the firmest ripe fruits, IL16, requiring 181 grams of force to pierce and the least softening at 20% change from unripe, is responsible for the markers found significant in the genome scan. IL16 is homozygous for two introgressed segments from *S. lycopersicoides*, one on chromosome 3 and the other on chromosome 5 (Table S1). The region of chromosome 3 represented in IL16 is heterozygous in IL15 and this line has ripe fruits of a more typical firmness which may suggest that the introgressed alleles are recessive. For protein coding regions, having one allele that is expressed or one allele that produces a fully functional enzyme may be enough to recover most of the softening phenotype.

IL16 has the end of chromosome 3 introgressed, but the very last marker is also homozygous for introgression in IL14. IL14 fruits have a ripe firmness of 125 grams of force which suggests the end of the chromosome may not contain the genes responsible for the firmness in IL16. The beginning of chromosome 3 is also present in IL14, but this region is heterozygous in this line and is not covered well by other lines so it is difficult to determine if this region affects the firmness. This leaves the region on chromosome 3 between the markers TG284 and TG152 as likely linked to ripe fruit firmness. There are two known softening genes between these markers with a third just outside this range (Table S6).

Polygalacturonase, *XPG1* (Sitrit et al., 1999; Tomassen et al., 2007), and auxin response factor 2A, *ARF2* (Breitel et al., 2016; Hao et al., 2015), fall between the markers on chromosome 3, and beta-galactosidase 3, *TBG3* (Smith and Gross, 2000), is 2Mb away from the significant marker TG152 and 1 Mb away from the non-significant marker from IL14, TG244 (Table S6). *XPG1* produces a pectin hydrolase and *TBG3* produces a glycoside hydrolase, both of which can take part in cell wall rearrangement processes in softening. The product of *ARF2* is a regulator in the ethylene ripening response and may coordinate several ripening processes, which could affect degree or timing of response.

The introgressed segment from chromosome 5 found in IL16 is also present homozygously in IL25, which is also quite firm at 152 grams of force, though it has the firmest measured unripe fruits and shows the most softening. This line includes a slightly longer segment of chromosome 5 than IL16 and it is also homozygous for a small region in chromosome 2. This segment of chromosome 2 is shared with IL13 which has fairly typical ripe firmness of 130 grams of force. This is the only segment introgressed in IL13, which suggests that it likely does not play a role in the firmness of IL25. This end region of chromosome 5 does not contain genes known to contribute to tomato softening but could be a good candidate region for finding new genes (Table S6).

The beginning portion of chromosome 5 is introgressed in the softest well-represented line, IL22, the fruit of which is pierced with 95 grams of force. IL21 and IL23 also contain portions of this segment and have fairly soft fruits at 114 and 117 grams of force respectively. The master regulator, *rin*, is very close to this region of chromosome 5 (Table S6), though the ILs with introgressed regions inclusive of *rin* have more typical fruit firmness, at 127 grams of force. A portion of chromosome 5 that is

introgressed in IL22 is shared with IL51, whose fruits are soft at 105 grams of force, though this introgressed region stops before the marker near *rin*.

IL51 has a second introgressed genome segment that is on chromosome 11 from marker TG46 to marker TG393; this genome region is not covered in any of the other lines grown. Within this introgressed segment of chromosome 11 there is another beta-galactosidase gene, *TBG5* (Table S6), producing an enzyme that could soften fruits through cell wall remodeling (Ishimaru et al., 2009; Smith and Gross, 2000).

For fruit sugar concentrations, many more regions with significant LOD scores were identified in the genome scans, however, fewer lines were evaluated and much fewer biological replicates were measured, so these regions bear further examination. We are still able to identify regions of the genome that may house genes involved in fruit sugar concentration and accumulation to be better explored in future studies.

In unripe fruits, the total sugars are significant in two genomic regions (Fig. 11d). Chromosome 3 is significant at the first marker on the chromosome, TG479 (Fig. 12a), due to the low sugar concentration in IL2, which is the only line measured that is homozygous for the outgroup's B allele at this marker (Table S1). There are three known glucose transporter protein genes (*SlSTP7 SlSTP10* and *SlSTP13*) on this portion of chromosome 3 (Table S7). *SlSTP7* has been found to only be expressed in roots, though this could still potentially affect sugar levels in the fruits by altering the overall sugar content of vascular tissues.

Chromosome 8 also has a significant region for unripe total sugar concentrations involving three markers at one end of the chromosome: markers TG176, TG45 and TG41

(Fig. 12b). The significance for these markers is due to the low sugar concentration found in IL35, which is the only line homozygous for the *S. lycopersicoides* alleles at these markers, though IL36 is heterozygous for all these markers and IL37 is heterozygous for TG41 (Table S1) and both these lines have much higher sugar concentrations (Table S4). For IL37 this is due purely to sucrose concentration, which is more similar to that of the introgressed species. IL35 is also homozygous for a region in chromosome 6 (Table S1), though this is also shared with IL29 which has higher sugar levels. The region on chromosome 8 contains many sugar related genes (Table S7). The genes here include several UDP-glucose glucosyltransferases and several other classes of enzymes that act upon sugars; these enzymes can act by altering sugars or diverting sugars to be used for other processes, which could lead to sugar depletion instead of accumulation. There is also a marker that was nearly significant in the sucrose concentration genome scan on the other end of this chromosome, but there are no known sugar genes in that region.

Ripe fruits have a genomic region that is significant for total sugar accumulation (Fig. 11h) as well as two regions that show significance for fructose accumulation (Fig. 11f). One marker shows significance in both sugars, marker TG432 on chromosome 5 (Figs. 12c & 12e), which is homozygous in IL21, IL22 and IL51 and heterozygous in IL23 (Table S1). IL51 has the lowest values for these sugar concentrations followed by IL22, while IL21 and IL23 have much higher concentrations (Table S4). IL22 and IL21 only have this region of chromosome 5 introgressed, while IL51 also has a portion of chromosome 11 introgressed, a region not shared by any other ILs (Table S1) and is the other genomic region with a significant LOD score (Fig. 12d).

The significant region of chromosome 5 contains two genes coding for glucose interacting enzymes, and is near but non-inclusive of *rin,* which codes for a major transcription factor involved in tomato fruit ripening (Table S7). The significant region of chromosome 11 covers three markers and over 24 Mb, but only has three known sugar interacting genes, which code for a glucose/ribose dehydrogenase, a neutral invertase and a dehydrogenase/reductase SDR family member (Table S7). These enzymes can be involved in metabolism and invertases are involved in breaking disaccharides into their monosaccharide constituents, altering the overall amount of sugars and the specific types that are accumulated. There is also a region of interest on chromosome 1 because of the peak in the sucrose LOD score, though it is not significant, which contains a sugar transporter encoding gene known to be expressed in late stage fruit development (Reuscher et al., 2014).

The traits studied here, sugar accumulation and fruit softening, are known to be quantitative, so many combinations of enzyme coding genes, their promoters, transcription factors and other regulatory elements can have effects building up to the observed phenotypes. We found that many introgression lines had phenotypes between the parental phenotypes but many also transgressed their parents' bounds. This highlights that there are many possible ways for these traits to evolve, and even just a reshuffling of existing alleles can create new phenotypes that could affect frugivore or pest behavior and impact the plant's reproductive success. Due to broad interfertility between wild tomato species, this raises the possibility of introgression between species in the wild as a means for trait evolution. More study is needed to determine the genes that affect these

traits and to build a better picture of the regulatory processes involved in the development of these traits.

Numerous genomic regions that include genes known to be involved in ripening traits, as well as others that may be involved but are not well studied, have been identified in this paper. Future studies can target these areas for sequencing to find allelic differences in the parental species and look into the potential functional differences between the alleles. Population studies can target these genes for expression and sequence variation analysis in wild tomato species to see how trait variation is brought about in different species and look for signs of selection. This will improve understanding not just of traits of interest in crop plants, but also of the evolutionary paths that have brought them about much like we explored for the *CYC-B* gene in chapter 3.

Figures

Figure 6

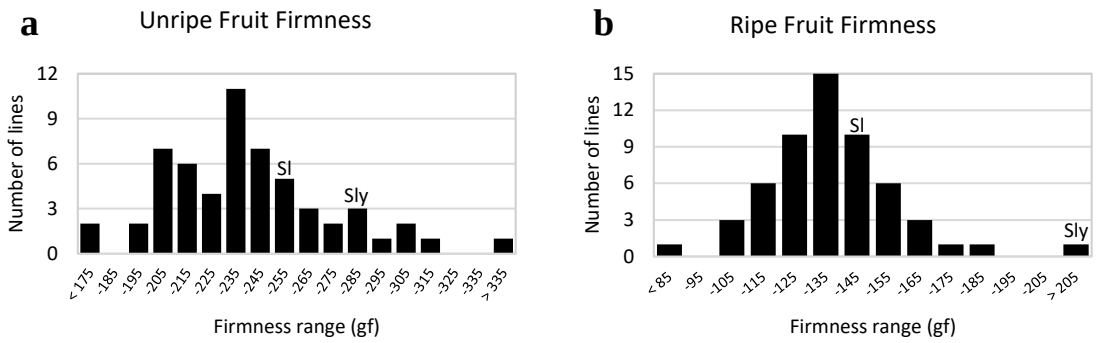


Figure 6 Frequency distribution of firmness values for ILs with both parental species values marked using abbreviated notation for *S. lycopersicum*, Sl, and *S. lycopersicoides*, Sly. Values shown are the predicted line values from the lme4 model in grams of force.

Figure 7

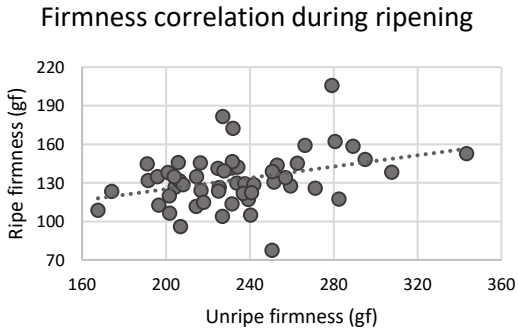


Figure 7 Unripe and ripe firmness values, in grams of force, graphed against each other to show degree of correlation between the two values in the IL fruits. Pearson's correlation coefficient r = 0.3525, p = 0.007.

Figure 8

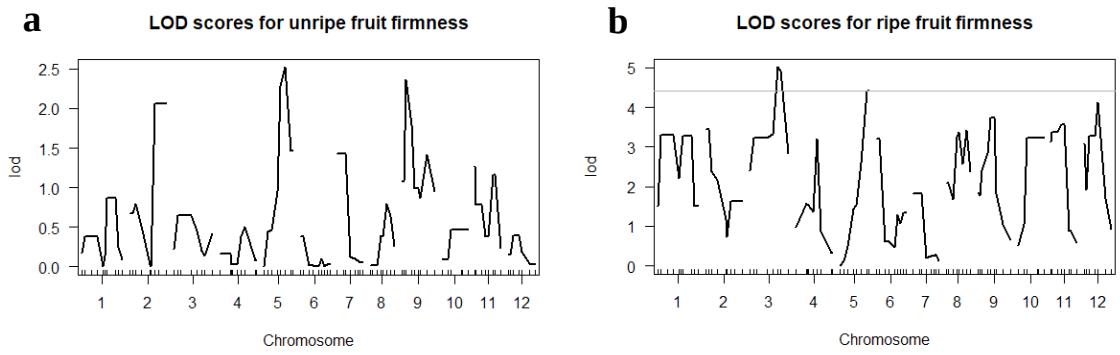


Figure 8 Linkage analysis associating gene regions with firmness values of (a) unripe and (b) ripe tomato fruits of plants from the ILs. Higher $\log_{10}$ of odds (LOD) scores correspond to increased likelihood that the trait and gene region are associated. Hashmarks along the X-axis indicate the location of markers across the chromosomes. For unripe firmness no LOD scores reached the significance cutoff at LOD = 3.72, but for ripe fruits the LOD cutoff for p-value < 0.05 is indicated by a gray horizontal line at LOD = 4.47.

Figure 9

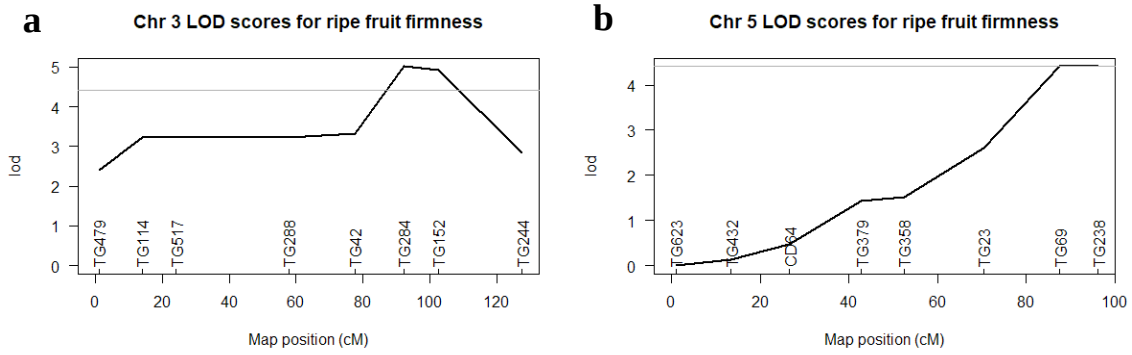


Figure 9 Detail of figure 8b showing LOD curves of (a) chromosome 3 and (b) chromosome 5 for ripe fruit firmness. Higher $\log_{10}$ of odds (LOD) scores correspond to increased likelihood that the trait and gene region are associated. Hashmarks along the X-axis indicate the location of markers across the chromosomes. The LOD cutoff for p-value < 0.05 is indicated by a gray horizontal line at LOD = 4.47.

Figure 10

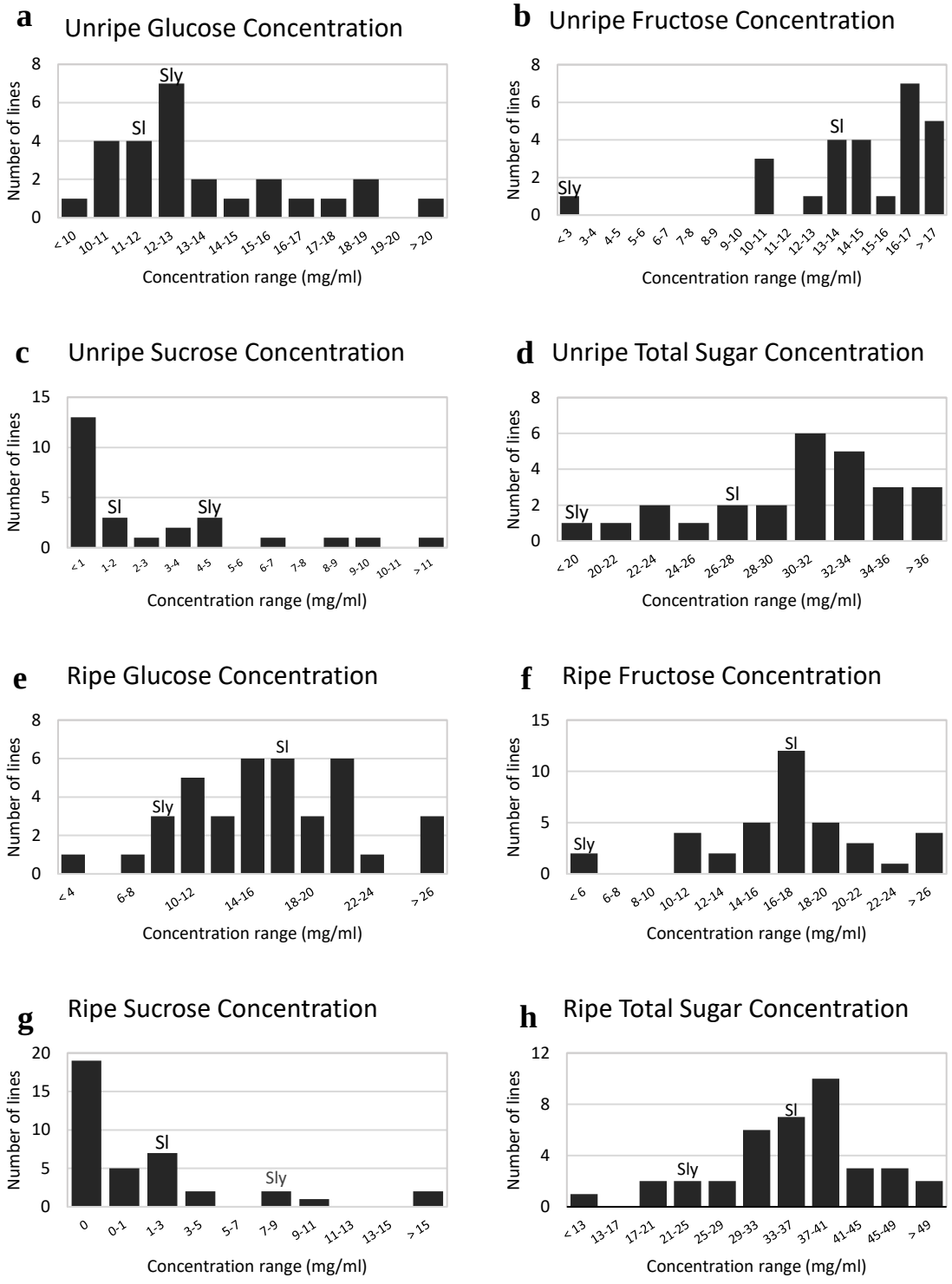


Figure 10 Frequency distribution of sugar concentration values (mg/ml) in unripe (a-d) and ripe (e-h) fruits for ILs with both parental species values marked using abbreviated notation for *S. lycopersicum*, Sl, and *S. lycopersicoides*, Sly. Values shown are the predicted line values from the lme4 model.

Figure 11

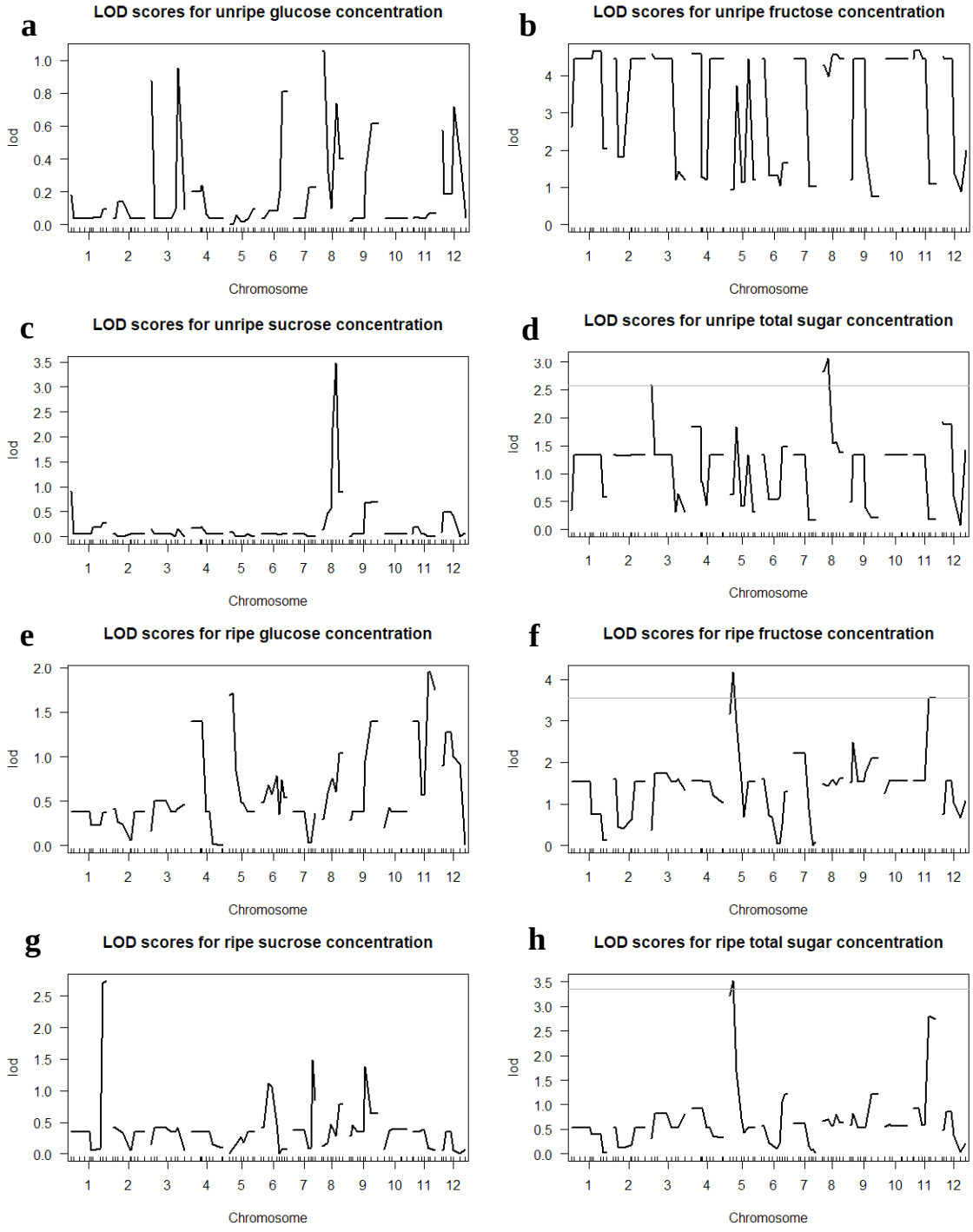


Figure 11 Linkage analysis associating gene regions with sugar concentrations of (a-d) unripe and (e-h) ripe tomato fruits of plants from the ILs. Higher $\log_{10}$ of odds (LOD) scores correspond to increased likelihood that the trait and gene region are associated. Hashmarks along the X-axis indicate the location of markers across the chromosomes. The LOD cutoff for p-value < 0.05 is indicated by a gray horizontal line, graphs with no horizontal grey line did not have any significant LOD scores but the cutoffs for all are listed here, (a) 2.85 (b) 5.32 (c) 4.05 (d) 2.58 (e) 2.59 (f) 3.54 (g) 7.98 (h) 3.35.

Figure 12

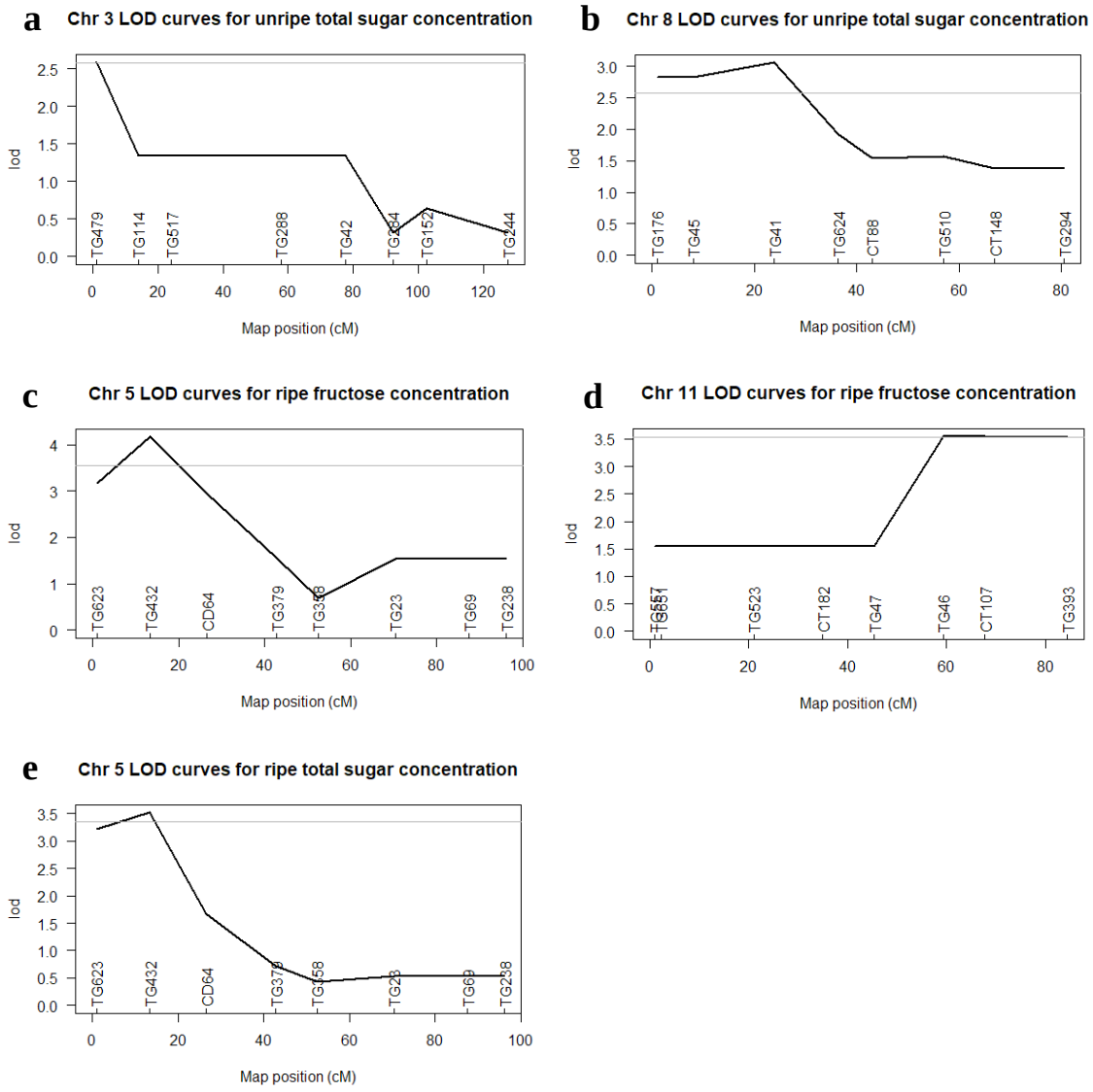


Figure 12 Close-up from analyses in figure 11 of chromosomes with significant LOD scores for sugar concentration. Higher LOD scores correspond to increased likelihood that the trait and gene region are associated. Hashmarks along the X-axis indicate the location of markers across the chromosomes. The LOD cutoff for p-value < 0.05 is indicated by a gray horizontal line.

CHAPTER 3

ANALYSIS OF FRUIT COLOR AND CYC-B GENETIC VARIATION IN WILD TOMATO SPECIES

Introduction

Angiosperm seeds grow within ovary tissue, which can continue to grow and develop along with the maturing seeds. This ovary tissue becomes the fruit, a structure that can help the survival of the seeds within them or can aid in attracting potential dispersers for those seeds. For fleshy fruits, the developmental changes occurring as seeds mature are known as ripening. Ripening concludes once seeds have fully developed and are capable of germination. Because seeds have different needs at different points of their developmental process, the role and appearance of fruit usually changes during this time too. During early seed development, there is a need to protect the developing embryo, and in fleshy fruits this protection is provided by thick strong walls and other structural or chemical defenses. Mature seeds usually benefit in some manner from escaping the fruit's tissue, and ideally dispersing away from the mother plant. The ripe stage of fruits provides means of seed release, either through mechanical devices or by attracting animal dispersers that can consume the fruit flesh and free the seed.

Animals can be efficient seed dispersers, but taking advantage of their services requires changes in the fruits to make them appealing and to signal that a desirable food source now exists. Color is a common ripening indicator in fleshy fruits, but far from the only option. Mammals tend to forage more by smell than sight, so plants that rely on mammalian dispersal commonly have duller colors, or may even remain green while going through dramatic aromatic changes (Eriksson et al., 2000; Van der Pijl, 1969). Birds, on the other hand, are visual foragers, attracted to brightly colored fruits (Van der

Pijl, 1969); red and black are the colors most commonly associated with avian dispersers (Willson and Whelan, 1990). The ripening process for fruits that change color involves conversion of the cells' energy-producing chloroplasts into pigment-storing chromoplasts; production and storage of large amounts of pigment compounds gives color to the fruit tissue.

Fruit color can directly affect the survival and dispersal of a plant's seeds, and thus evolutionary changes to fruit color could be under strong selection. For instance, a switch from primarily avian to primarily mammalian dispersers may put the seeds at jeopardy of being ground by teeth, or the new disperser may help by extending the species' range to new areas. The selective pressures on fruits to have or maintain a certain color is not straightforward, though. Birds will often eat fruits of various colors, and mammals will often eat red and black fruits that are commonly associated with avian dispersal (Willson and Whelan, 1990). Thus, very little is known about the exact evolutionary processes leading to differences in fruit color among species.

Cultivated tomatoes (*Solanum lycopersicum*) have long been a model plant for studying the development of fleshy fruits (Kimura and Sinha, 2008; Knapp et al., 2004; Rick and Chetelat, 1995; Sato et al., 2012). The tomato clade also contains a range of fruit colors amongst closely related species, making wild tomatoes an excellent candidate system for studying the evolution of fruit color. There are 13 species of wild tomato in South America (Fig. 1), which have undergone two color transitions through the history of the clade. Basal species have green fruits when ripe and maintain chloroplasts throughout their ripening, while three species have colored fruits: *Solanum pimpinellifolium* is red-fruited, *S. galapagense* is orange-fruited and *S. cheesmaniae* has

fruit colors ranging from yellow to orange. *Solanum pimpinellifolium* is the wild tomato species that gave rise to cultivated tomato, *S. lycopersicum* (Nesbitt and Tanksley, 2002), as well as *S. cheesmaniae* and its daughter species *S. galapagense* (Pailles et al., 2017), both of which are endemic to the Galapagos Islands (Darwin et al., 2003).

To date, nothing is known about the evolutionary processes that have driven the changes in fruit color of tomato species, though it is notable that only red fruits are found on the mainland of South America with all of the green-fruited species, while the small Galapagos Islands harbor two species with a range of yellow and orange fruit colors. Both of these island species are noted for their tolerance to salt and osmotic stress (Albaladejo et al., 2015). The Galapagos Islands are known to go through drought conditions during La Niña, which creates bottlenecks for the plant populations (Restrepo et al., 2012), perhaps removing the selective pressures on traits that maintain fruit color in deference to traits enabling the plants to survive through harsh conditions. Alternatively, there could be selective pressure against bird endozoochory, with reptiles and mammals potentially depositing a greater portion of seeds into suitable habitats on these isolated islands, rather than into ocean waters.

Species in the tomato clade with colored fruits owe their color to two of the most well-known carotenoids, lycopene and β-carotene (Stommel and Haynes, 1994). Carotenoids are a class of pigments commonly found in nature, providing yellow, orange, pink or red colors, depending both on the specific carotenoid and its concentration. There are over 700 known carotenoids (Tanaka et al., 2008), which range in color from pale yellow to bright red. They are one of the most common classes of plant pigments, along with anthocyanins and the much less common betalains (Tanaka et al., 2008).

Carotenoids, such as β-carotene and xanthophyll, are ubiquitous in chloroplasts as part of the photosynthetic apparatus, providing photoprotective function to protect structures in chloroplasts, allowing more efficient photosynthesis (Niyogi, 2000). Most yellow flowers owe their color to carotenoids, and carotenoid pigments can also produce orange flowers. Carotenoids also commonly co-occur with anthocyanins in floral tissue, producing novel colors that would not be possible otherwise (Forkmann, 1991). The color of many common fruits, such as squash, guava, orange, cantaloupe, mango, watermelon, papaya and chili peppers, are provided by carotenoid pigments.

Both carotenoids and anthocyanins are universally found in plants, and thus there exists little phylogenetic pattern to their use for pigmentation. For instance, plants of the *Solanum* genus variously produce either pigment class for coloring their fruits, and even within the tomato clade, where orange, red, and yellow fruits are colored by carotenoids, there are some species which have anthocyanin stripes on their fruits (Rick et al., 1994b). Within the *Solanum* genus the ancestral state is having green-mature fruit with a firm texture (Wang et al., 2015) while derived fruits of the genus come in many colors, sizes, and textures, suggesting fruit traits may be quick to change. Since the regulatory pathway to produce wide varieties of pigments exist in all Solanaceous plants, they are capable of gaining new color from single mutations (Wang et al., 2015).

We sought to more thoroughly document wild variation in fruit color in the tomato clade, to aid development of this group as a model for fruit color evolution. Additionally, we sought to identify possible genes associated with fruit color differences, document genetic diversity in these genes, and identify what might be drivers for evolution in these genes. This helps to improve the use of tomato as a model system by

providing more knowledge on gene function and the wild alleles that exist in close relatives. Additionally, it provides a chance to build the more general knowledge base about fruit color evolution with a case of several changes in closely related species.

Materials and Methods

Plants and growth conditions

Seeds from multiple accessions of 13 *Solanum* species, including *S. galapagense, S. cheesmaniae, S. pimpinellifolium, S. neorickii, S. cornelliomuelleri, S. chmielewskii, S. chilense, S. pennellii, S. habrochaites, S. peruvianum, S. lycopersicum, S. lycopersicum var. cerasiforme,* and the outgroup to the tomato clade, *S. lycopersicoides,* were obtained from the UC Davis C.M. Rick Tomato Genetics Resource Center, maintained by the Department of Plant Sciences, University of California, Davis, CA. Plants were grown in the Morrill greenhouses on the University of Massachusetts Amherst campus with supplementary lighting and heating to maintain 12 hour day length and temperatures above 26°C during the day and 22°C at night. Young leaf material was collected for DNA extraction from all plants. Four species, which included the color-fruited species *S. galapagense, S. cheesmaniae, S. pimpinellifolium,* and the closely related green-fruited species *S. neorickii,* were grown until maturity and fruit set. Since these species are self-compatible, new flowers were regularly agitated to promote self-pollination. Fruits were collected at the ripe stage or near the beginning of ripening, at the breaker stage, as needed for expression analysis. Fruits were halved, seeds were removed and the remaining tissue was flash frozen in liquid nitrogen.

Phenotyping

Ripe fruit color was phenotyped by photographing each fruit with a Spydercube (Datacolor, Lawrenceville, NJ) to enable reproducible white balance between images. White balancing was carried out with Photoshop (Adobe, San Jose, CA), which was also used to measure the color of each fruit, giving numerical values to denote color. Measurements were taken from three fruits per plant and averaged to give a single value representative of fruit color. All color values were measured in the L*a*b* color spectrum which uses three values to represent a color, where the L* value denotes lightness with 0 being black and 100 being white, and the values of a* and b* are given on a scale from -128 to 128, with maximal saturation of a color occurring at each extreme. The values of a* measure the amount of green in the negative direction and amount of red in the positive direction, values of b* measure the amount of blue in the negative direction and amount of yellow in the positive direction.

Quantification of carotenoid content

Frozen mature tomato fruit tissue from various accessions belonging to four species was powdered using a Retsch (Haan, Germany) MM400 ball mill with a 25ml milling jar at the University of Massachusetts Amherst. A total of 0.2 grams was collected from each sample for carotenoid extraction. Fruits were chosen that represented the variation in fruit color of each species. Accession Spm1, naming follows that used in Figure 1 with numbers to differentiate accessions, has red fruit typical of *S. pimpinellifolium* mature fruit; accession Sg2 has dark orange colored fruit, and Sg6 and Sg7 have bright orange fruits typical of *S. galapagense*; the flesh of Sg14 fruits is a typical orange although a dark pigment in the skin of these fruits makes them appear very

dark; accession Sc19 fruits are similar in color to Sg6 and Sg7, and common for *S. cheesmaniae* though more color variation exists in this species than the other two, while Sc3 has a pale orange/yellow colored fruits and Sc11 fruits are very pale with little noticeable coloration; accession Sn7 is representative of green-ripe, non-chromoplast forming species such as *S. neorickii,* which is expected to show a baseline amount for carotenoids that are used in chloroplasts. Further analysis was carried out in the Giovannoni lab at the Boyce Thompson Institute at Cornell. Carotenoids were extracted following their lab procedure as previously described (Alba et al., 2005). The sample extracts were then measured for content of the carotenoids phytoene, ζ-carotene, β-carotene, lycopene, and lutein. Quantification of carotenoids was carried out with a Waters (www.Waters.com) HPLC system equipped with a 4.6x250mm Waters reverse phase YMC Carotenoid S-5 column and a Dionex (www.Dionex.com) PDA-100 photodiode array detector, which allows simultaneous detection of multiple carotenoids.

Expression profiling

Gene expression was measured with semi-quantitative RT-PCR of RNA extracted from fully mature fruits, which was determined by browning of the calyx in species of all colors. Ripe fruits were used because this allows the inclusion of green-colored species, for which developmental stages are difficult to determine except at maturity. RNA was extracted using a Qiagen (Hilden, Germany) RNeasy plant mini kit from 2-3 fruit halves from the same plant. RNA was converted to cDNA using the Ambion (Thermo Fisher Scientific, Waltham, MA) RETROscript kit with oligo d(T) primers following manufacturer's instructions. Primers were designed using the online tool Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012), and the primer sequences used

can be found in Table 9. Gene expression was measured for the five genes encoding the enzymes of the carotenoid synthesis pathway (Fig. 13) starting from *Psy1*, which acts on the general molecule geranylgeranyl diphosphate, required for diversion into the carotenoid pathway, and ending with *CrtR-b2*, which continues the pathway past the carotenoids and on to the xanthins. *Actin* was used as an expression standard. PCR were run 25 cycles for all genes to get semi-quantitative results showing approximate relative expression levels.

Microarray expression profiling

To measure expression levels, microarrays were carried out with the Affymetrix (Santa Clara, CA) GeneChip Tomato Genome Array with four RNA samples: two replicate extractions each from *S. cheesmaniae* accession LA0421 and *S. pimpinellifolium* accession LA0373 at the turning stage of ripening, to contrast orange and red fruits. The "turning stage" occurs shortly after the fruit stops growing and begins color changes associated with full maturation. This stage was chosen because the enzymes involved in color change are being actively produced so gene expression can be measured. The Affymetrix microarray contains 11 probe pairs per sequence covering over 9,200 tomato gene transcripts (Affymetrix, 2011), providing significant coverage and increasing the ability for homologous genes in closely related species to be identified by the chip. The microarray chip hybridization was done by the University of Massachusetts Medical School Genomics Core Facility. Data was analyzed using Bioconductor (Gentleman et al., 2004) in the R statistical computing environment (R Core Team, 2015). The affy package (Gautier et al., 2004) was used for interpretation of the intensity values to numerical values and the snm package (Mecham et al., 2010) for supervised data

normalization. A ratio between two values is calculated by taking the difference of those values, as the values are logarithmic. Replicate values were averaged and differences between the species with a value greater than two were considered to be significant.

DNA sequencing

To allow sequencing, DNA was extracted from leaf tissues using a modified CTAB miniprep DNA extraction procedure (Hillis et al., 1996). Between four and nine accessions of species selected for their color and phylogenetic placement were included in this study. The final number included per species also depended on success in amplifying the desired genes. Included were: nine accessions each of *S. pimpinellifolium* and *S. cheesmaniae*; eight accessions of *S. galapagense*; four accessions each of *S. lycopersicum, S. lycopersicum* var. *cerasiforme, S. chmielewskii* and *S. neorickii*; two accessions each of *S. chilense, S. habrochaites, S. corneliomulleri, S. peruvianum* and *S. pennellii*; and a single accession of *S. lycopersicoides*.

Portions of five loci were targeted for amplification. The first locus was *CYC-B*, the gene found to have the greatest difference between color morphs in expression out of the genes in the carotenoid synthesis pathway (Fig. 13). *CYC-B* lies on chromosome 6, and primers to amplify approximately 985 bases of promoter and 1348 bases of the coding region were designed using Primer 3 (Caicedo and Schaal, 2004) (Table 9). Four DNA segments not involved in carotenoid synthesis, denoted as *ct066, ct093, ct179* and *vac*, were chosen to serve as comparison for *CYC-B. Ct066* is an approximately 654 base section from the coding sequence of the argenine decarboxylase gene on chromosome 10. *Ct093* is an approximately 575 base section from the coding sequence of the S-

adenosylmethionine decarboxylase gene on chromosome 5. *Ct179* is an approximately 985 base section from the coding sequence, with two introns, of the δ-tonoplast intrinsic protein on chromosome 3. Primers for all these loci were obtained from (Zuriaga et al., 2008). *Vac* is an approximately 742 base section from the second intron within a vacuolar invertase gene on chromosome 3, and primers for this locus were obtained from (Caicedo and Schaal, 2004). Sanger sequencing was carried out by Beckman Coulter (Pasadena, CA). Sequences were aligned and cleaned using BioEdit (Hall, 2013).

Population level analyses

Natural selection tests were carried out using the Hudson-Kreitman-Aguade (HKA) test, which assesses whether loci display an excess of interspecific divergence or intraspecific polymorphism compared to genomic averages. Test were carried out using the HKA software developed and made publicly available online by the Jody Hey lab at Temple University (https://bio.cst.temple.edu/~hey/software/software.htm) along with the maximum-likelihood HKA (MLHKA) software developed by Stephen I. Wright (Wright and Charlesworth, 2004). For all tests, sequence from *S. lycopersicoides* was used as an outgroup. Signatures of selection were tested at all sequenced loci for *S. pimpinellifolium, S. cheesmaniae* and *S. galapagense,* the latter of which grouped together as they form the orange-fruited clade and have little genetic diversity between them, and *S. neorickii* and *S. chmielewskii* also grouped together as they represent the green-fruited sister clade to the color-fruited clade. Since *S. cheesmaniae, S. galapagense,* and *S. neorickii* showed no heterozygous sites they were treated as haploid. *S. pimpinellifolium* had numerous heterozygous sites and *S. chmielewskii* had some heterozygous sites, so both were treated as diploid.

The HKA software was used to produce an overall assessment of whether the sequence data indicated selection in a species for the loci tested. Time and θ estimates from the HKA output were used for the MLHKA infile. The polymorphism and divergence statistics π and Tajima's D statistic (Tajima, 1989), also needed to run the program, were calculated using DnaSP software (Librado and Rozas, 2009). The MLHKA software allows testing of individual loci for selection, providing a more detailed view of selection in each group and allowing for determination of the likelihood of a model assuming no loci are under selection and a model assuming one or more loci are under selection. This likelihood ratio was tested for significance using a χ^2 test.

Phylogenetic analyses

Phylogenetic trees were created for each sequenced locus using RAxML-HPC2 (Stamatakis, 2014) with the GTRGAMMA bootstrapping model. This was done on XSEDE through the CIPRES science gateway at www.phylo.org (Miller et al., 2010). The resultant newick file was used to create phylogenetic trees with the MEGA software (Tamura et al., 2013). The final trees used in figures for this paper were refined using Inkscape vector graphics software (www.inkscape.org).

Analysis of promoter binding sites

Promoter sequences of *CYC-B* in red and orange fruited species were analyzed using the online tool PlantPAN (Chang et al., 2008), to find potential transcription factor (TF) binding sites that could be affected by fixed sequence differences between the color-fruited species. Transcription factor binding motifs from *Arabidopsis thaliana* and *Glycine max* were used, as these are the most closely related plants to the tomato clade

with well documented transcription factor binding motifs available for use. These results were then edited to remove TF families not found in *Solanum* to give a list of possible TF binding sites that could account for differences in orange and red fruits.

Results

Variation in fruit color

This analysis provided a way to quantify the observed visible color variation in the fruits of this clade. Ripe fruits from the *Solanum* clade span a large range on the red-green and yellow areas of the L*a*b* color space (Fig. 14), and the size of the color space area occupied varied greatly between species. Red fruits, all of which belong to *S. pimpinellifolium* or its derived domesticated or semi-domesticated species, *S. lycopersicum* and *S. lycopersicum* var. *cerasiforme*, were very similar in L*a*b* color values. The orange fruits of *S. galapagense* and *S. cheesmaniae,* on the other hand, showed large variation (Fig. 14). *S. galapagense* individuals showed variation in the amount of yellow, primarily due to accession Sg14, which stands out due to dark pigments found in the skin of its fruits throughout development and is unrelated to ripening color changes. *S. cheesmaniae* individuals showed wide variation along the red-green and blue-yellow axes, with Sc19 and Sc20 being rich orange, similar in color to the fruits of *S. galapagense*, while Sc1 and Sc11 have very pale fruits that can appear creamy or somewhat green (Fig. 14). Most individual plants belonging to the same accession produced fruits that were fairly uniform in color, so only one measurement from each accession provides a representative color value, however, fruits from Sc11 were quite variable and so values from a slightly green and slightly yellow fruit were added along

with the representative Sc11 value of the more typical cream color. Green-fruited species do not accumulate carotenoids during ripening and a small sample of ripe green fruits from two species were included to show the differences in the L*a*b* color space between green-ripe fruits and the colored carotenoid accumulating fruits. As expected, green fruits clustered toward more green and less yellow values in the color space.

Carotenoid content

Both type and quantity of carotenoid pigments differed considerably among tomato species. The red fruits of *S. pimpinellifolium* were the only fruits to contain any lycopene, which is also the main carotenoid they accumulate (Fig. 15a). Red fruits also contained an order of magnitude more total carotenoids than any of the other fruits (Fig. 15b). Green colored *S. neorickii* fruits contained mostly lutein, a light-colored pigment that is found in chloroplasts, and this pigment also predominated in our very light-colored, somewhat greenish *S. cheesmaniae*, Sc11. The fruits of *S. galapagense* plants varied in carotenoid content despite their visual similarity, accumulating primarily β-carotene and phytoene, a colorless carotenoid precursor. Sg2 had the deepest orange fruits, nearly red, but these were unable to be regrown for the color measurements above. Fruits of *S. cheesmaniae* plants accumulated very little carotenoids overall, most of which was phytoene with small amounts of β-carotene, though no β-carotene was detected in Sc19 despite its orange color (Fig. 14). Table 10 shows the technical replicate results from β-carotene quantification. There is high variability which, along with the surprising Sc19 results, highlights that there is too little replication here to draw strong conclusions about carotenoid levels and correlation with fruit colors.

Carotenoid synthesis gene expression

We initially examined expression of carotenoid genes in an orange-fruited *S. cheesmaniae* and a red-fruited *S. pimpinellifolium* at the turning stage of ripening using a microarray. Green-fruited species were not included, due to difficulties in identifying the homologous developmental stage. Significant differential expression between color morphs was observed for *CYC-B* (Table 11); in orange turning-stage fruits, *CYC-B* expression was over three times the level seen in red fruits, suggesting active synthesis of β-carotene at the expense of lycopene (Fig. 13). Out of all the probes on the array, *CYC-B* was the 12th most differentially expressed. Though not significant, the expression of *CrtR-b2* was almost two times lower in orange fruits compared to red and is the 101st most differentially expressed gene on the chip, out of over 9,000 genes covered. Higher expression of *CrtR-b2* could facilitate depletion of any β-carotene synthesized in red fruits, and lower expression could facilitate a greater ability to accumulate β-carotene in orange fruits (Fig. 13). Expression of other genes in the pathway differed by less than 70%. Full microarray results can be found in Table S8, the genes showing an expression ratio greater than 2 between the species are shown in Table S9.

To complement these results, we used RT-PCR to examine expression of carotenoid synthesis genes at the fully ripe stage in a larger panel of accessions, including one green-fruited sample, *S. habrochaites* (Fig. 16). From these results, genes in the middle of the carotenoid synthesis pathway seem to be the most active across accessions of different colors, with varying expression levels between individuals (Fig. 16). Since *Psy1* is only showing expression in Sg13, it is possible that the beginning of the pathway is being shut down at this stage of development, as it would have to have been active at

some point. The consistent difference between red fruit and fruits of other colors was a large down-regulation of *CYC-B* gene expression (Fig. 16), which would allow these red fruits to accumulate lycopene, while orange fruits are producing the enzymes to deplete lycopene, allowing the accumulation of β-carotene instead.

CYC-B and non-carotenoid related gene trees

Because *CYC-B* is the carotenoid synthesis gene that most consistently showed expression differences between tomatoes of two colors (orange and red), we examined DNA sequences for this gene and its promoter region (Fig. 17), and compared phylogenetic relationships to those observed for four other genes unrelated to carotenoid synthesis (Fig. 18). For all genes, strong differentiation was seen between green and colored-fruited species, with colored-fruited species forming strongly supported monophyletic groups, albeit with the occasional green-fruited allele falling in the colored-fruited clade. This is consistent with known tomato phylogeny and the monophyly of colored species (Pease et al., 2016b; Spooner et al., 2005). Orange and red-fruited species were not consistently differentiated in any gene tree, except that of the *CYC-B* promoter, for which the red-fruited clade had a high bootstrap support (Fig. 17b). Thus, although the promoter and coding regions of *CYC-B* have unique single nucleotide polymorphisms (SNPs) and indels that group the color-fruited species together, only the promoter region could differentiate between the red and orange-fruited species. The yellow-orange species, *S. cheesmaniae* and *S. galapagense*, could not be differentiated by the sequence of any of the genes examined in this project.

We examined more closely the clade differentiation in the promoter and coding region of *CYC-B*. The coding region in color-fruited species is defined by two fixed non-synonymous mutations and two fixed synonymous mutations that differentiate them from green-fruited species, as listed in Table 12. In addition to these alleles unique to the color-fruited species, six synonymous and two non-synonymous segregating sites were found to be fixed in the color clade while both alleles still segregated in green-fruited relatives (Table 12).

Within the *CYC-B* coding region, cultivated tomatoes (*S. lycopersicum* and *S. lycopersicum* var. *cerasiforme*) formed a well-supported clade that is separated from other colored species. These domesticated species contain a fixed G to T SNP at 868 bp from the beginning of the coding region. This SNP results in an exchange of an asparagine (Asn) for aspartic acid (Asp). While this change is from a polar neutral amino acid to an acidic negatively charged one, these two amino acids can serve similar functions, and are frequently referred to interchangeably as Asx for structures and motifs that are functionally similar. Thus, it is unlikely that this fixed SNP leads to functional changes in the lycopene β-cyclase enzyme.

The promoter region of *CYC-B* had multiple polymorphisms that differentiated red-fruited accessions from orange-fruited accessions, as well as polymorphisms that differentiated colored accessions from green-fruited accessions (Table 13). In our data, for the promoter region we sequenced, ten SNPs were fixed in color-fruited species and differentiated from green-fruited species. Additionally, there were two single base pair insertions and three multi-nucleotide deletions that are fixed and specific to the color-fruit

clade. There were also twenty segregating sites that have become fixed for a single allele in the color-fruited species, but still segregate in green-fruited species (Table 13).

Red-fruited and orange-fruited species were differentiated by one SNP 883 bp upstream of the start codon that segregates in green-fruited species, but is differentially fixed between *S. pimpinellifolium* and orange-fruited species (Table 13). Additionally, there were three point mutations (sites -527, -412 and -78) for which a novel derived allele is fixed exclusively in the red clade, and two point mutations (sites -640 and -206) for which the orange clade has become fixed for a novel derived allele.

Genetic diversity and natural selection measures

Levels of genetic diversity were examined for key tomato groups for all sequenced loci. In general, the *CYC-B* coding region tended to have among the lowest levels of nucleotide diversity compared to other loci (Table 14). This was particularly pronounced in *S. pimpinellifolium* and the combined *S. neorickii-S. chmielewskii* group, and less so in the *S. cheesmaniae-S. galapagense* group. *Solanum cheesmaniae* and *S. galapagense* tend to have low levels of diversity across the genome, likely due to a bottleneck upon island colonization and high levels of self-fertilization, which impacts all loci. The promoter region of *CYC-B* was also low in diversity in colored species, but not as notably so in the green-fruited species. In general, levels of nucleotide diversity in all loci were comparable for *S. pimpinellifolium* and the *S. neorickii-S. chmielewskii* group, and lower for the Galapagos group, *S. cheesmaniae* and *S. galapagense*. For the most part, Tajima's D values for *CYC-B* were unremarkable. The orange-fruited clade was the only group that showed an extreme Tajima's D for the *CYC-B* promoter, which was

different than seen on the loci not involved in carotenoid synthesis (Table 14). In this clade, there is a negative Tajima's D for *CYC-B* promoter while the coding region shows no polymorphism at all, which could suggest purifying selection against new mutations in this gene and its promoter, or recent positive selection on the locus. A negative Tajima's D value is consistent with an excess of rare mutations, which can occur after positive selection, but can also be attributed to an expanding population.

To test for possible selection on the *CYC-B* locus in any tomato group, we used HKA tests (Hudson et al., 1987; Wright and Charlesworth, 2004). HKA tests take divergence into account along with within-group polymorphism by comparing the species of interest against an outgroup for each locus, and assessing if either quantity is different from patterns across the genome. χ^2 statistics were computed for maximum likelihood values from comparing a model assuming no loci are under selection to one assuming that the listed locus or loci are under selection (Table 15).

Surprisingly, the assumed neutral locus *ct066* was the only locus with significant χ^2 in more than one group, both in *S. cheesmaniae-S. galapagense* and *S. neorickii-S. chmielewskii*. The *ct066* locus covers a portion of the coding region of an arginine decarboxylase gene on chromosome 10. This locus shows lower divergence than most other loci but higher levels of polymorphism, causing the significant result in both cases (Table 16). In these two groups, no other locus or combination of loci showed significance. False positives in *S. neorickii-S. chmielewskii* could occur in areas where there are divergent SNPs between the species of that group that in turn may imply a higher frequency of rare alleles when assuming them to be one species, however this is not likely to have been a large issue as this group only had a significant χ^2 at one locus.

In *S. pimpinellifolium* the assumed neutral *vac* locus has a significant χ^2, suggesting this locus is under selection in the group, with much higher levels of polymorphism than the other loci examined in this species (Table 16). The *vac* locus is within the second intron of a vacuolar invertase gene, which could be involved in fruit sugar levels, though the locus sequenced is non-coding, and it is not clear why diversifying selection would act on a sugar-related gene. Curiously, when the *CYC-B* coding region and promoter were tested separately for signals of selection, neither showed significance in any group. However, if the *CYC-B* gene is tested as a whole, a model suggesting selection is strongly supported in *S. pimpinellifolium* (Table 15). This likely stems from the high levels of divergence relative to low polymorphism when looking at these loci together (Table 16), which is a common signal of positive selection. *S. cheesmaniae-S. galapagense* have similar divergence to polymorphism ratio for this gene but do not show significant χ^2 because this group has low polymorphism at all loci.

Transcription factor binding sites

Despite only limited signals of selection for the *CYC-B* gene in orange vs. red fruited species, the differentiation in the promoter region suggest that there may be functional consequences related to polymorphism in the promoter. Because polymorphisms that differentiated red and green fruited species and their potential effect on transcription factor (TF) binding have already been described in (Mohan et al., 2016), here we focus on polymorphism that may possibly lead to differences in regulation of *CYC-B* in red and orange-fruited species. It is worth noting that binding sites for the major ripening induced regulator *RIN* do not vary between red and orange-fruited species but do vary between green and color-fruited species (Table 17). In the promoter region

sequenced, we found six SNPs that are fixed between orange and red-fruited species as shown in Table S10, which we refer to here by their position from the start codon in our alignment. Three of these SNPs have novel alleles in the red clade and two have novel alleles in the orange clade. All of these SNPs lead to changes in possible TF binding sites, with the fixed SNP in orange-fruited species at -206 creating novel binding possibilities and fixed SNP in red-fruited species at -78 which eliminated existing potential TF binding sites (Table S10).

The novel T allele fixed in the red clade at position -883 created a new motif that may be recognized by NAC and bZIP family promoters, though it reduces the similarity to Myb core TFs found in cyclin promoters. The orange fixed allele at -640 slightly lessens similarity to a SAUR auxin responsive motif while increasing similarity to a SURE sulfur responsive motif and ruining similarity to a SORLIP light induced motif. The derived G allele fixed in the red clade at position -527 somewhat reduces motif recognition; though both alleles can be recognized by bZIP and HD-ZIP family TFs the red clade loses potential recognition by AP2, AT-Hook and ERF TF families. The derived C allele fixed in the red clade at position -412 retains potential Dof family recognition while gaining recognition by TFs in the HSF and SBP families, though losing AT-Hook TF family recognition. In the orange clade, the derived T allele fixed at position -206 creates a new potential TF binding site recognized by Dof family TFs as well as gaining similarity to a SORLREP light repressed motif. The derived G allele fixed in the red clade at position -78 eliminates potential recognition by Dof and Homeodomain TFs as well as eliminating similarity to a root specific motif that is unlikely to be active in the fruits.

Discussion

The tomato clade has only three wild species that have colored ripe fruits (Grumet et al., 1981), but these three species contain quite a bit of visual variety. *Solanum pimpinellifolium* has bright red deeply colored fruits, while the species from the Galapagos Islands are orange fruited. *Solanum galapagense*'s fruits are intensely orange colored, while *S. cheesmaniae*'s fruits range in color from deep orange to pale orange to yellow to cream.

This pattern of visual consistency/variation was supported in the measured color values but less so in the measured carotenoid contents. *Solanum pimpinellifolium* contained very high levels of the red pigment lycopene. This was expected since their fruits are a deep red color and cultivated tomato is well known for containing lycopene. However, the concentrations are orders of magnitude higher than what was seen for carotenoids of other color fruits. The orange fruits had dramatically less carotenoids overall and contained no lycopene, accumulating β-carotene instead. The fruits of *S. galapagense* varied quite a bit in β-carotene content; Sg2 was the darkest orange fruit with hints of red and also had the highest β-carotene content. Sg6 and Sg7 have visually similar fruits that vary slightly in their L*a*b* values, but Sg6 contains almost no β-carotene. Sg7 is also visually similar and very close in L*a*b* values to Sc19 which contains no β-carotene and barely any carotenoids at all. If these values are correct it suggests that much less carotenoid investment is required to produce rich colors with β-carotene than with lycopene, though the ability to attract avian dispersers could justify the cost.

The gene expression in ripening fruits highlighted the division between red/orange fruits and their lycopene/β-carotene accumulation. Red fruits showed very low expression of *CYC-B*, limiting production of lycopene beta-cyclase, allowing the accumulation of lycopene, a molecule acted upon by this enzyme. Orange *S. cheesmaniae* fruits, conversely, express *CYC-B*, allowing the conversion of lycopene to the β-carotene that these fruits accumulate, while showing a slight reduction in expression of *CrtR-b2*, a gene producing β-carotene hydroxylase, an enzyme which acts on β-carotene. This confirmation of the genetic control of color production suggests that, to achieve red or orange fruit colors, the expression of genes that produce enzymes which modify colored carotenoids are being affected, and not their function.

The *CYC-B* gene was sequenced to look at whether differences had evolved between red-fruited, orange-fruited and green-fruited species. This gene contains many differences specific to carotenoid accumulating species in both promoter and coding region sequence. The promoter region has many SNPs, some larger indels as well as loss of allele diversity at other SNP sites differentiating color-fruited species from the green-fruited species. As previously reported, none of these lie in binding sites associated with the ripening associated transcription factor RIN (Mohan et al., 2016). There are other TF recognized sequences that are affected by these SNPs which may play a role in the overall expression, though how this regulation may occur is less understood than RIN associated expression. Within the coding region of *CYC-B* the color-fruited species have SNPs resulting in synonymous and non-synonymous changes. Two of these non-synonymous SNPs are unique to the color-fruited clade and two others result from a loss of diversity (Table 12). The division between orange and red-fruited species, however, is

limited to the promoter region, where both groups have unique SNPs. These SNPs do not affect RIN binding sites, but they do all fall inside putative TF binding motifs, which could lead to differential expression of this carotenoid acting enzyme to produce different colored fruits.

Some loci that were included in this study as neutrally evolving background loci, uninvolved in the color phenotype, actually showed highly significant likelihood to have been evolving under selective pressures (Table 15). Neither the *S. cheesmaniae-S. galapagense* nor *S. neorickii-S. chmielewskii* groups showed evidence of *CYC-B* evolving under selection. However, there was evidence that *ct066,* a portion of the coding region from an arginine decarboxylase gene, may be evolving under diversifying selection in both groups. In *S. pimpinellifolium, vac*, the second intron within a vacuolar invertase gene, might also be under diversifying selection (Table 15), which may indicate a diversity of sugar accumulation profiles or sugar use profiles in the species. The promoter and coding sequence of *CYC-B* in *S. pimpinellifolium* show likelihood of evolution under positive selection, but only when taken together, as neither sequence on its own was significant. However, it is only the promoter region of *CYC-B* that shows red specific sequence. With these results, it is likely that *CYC-B* has been under more selective pressures in red-fruited *S. pimpinellifolium* to decrease expression during fruit ripening, than it has in increasing fruit ripening expression in the orange-fruited Galapagos tomatoes.

Despite having many phenotypic differences, *S. cheesmaniae* and *S. galapagense* have high levels of genetic similarity, so much so that for a long time *S. galapagense* was classified as a subspecies of *S. cheesmaniae*. Sequences from three of the regions used in

this study had no differences at all between the two species and, overall the sequences from these species had very little variation compared to variation found in the sequences of the same regions in other species. It was thus unsurprising that no region sequenced could be used to differentiate these species.

CYC-B expression is tightly linked to tomato fruit color in determining red versus orange, since for the accumulation of lycopene little to no lycopene β-cyclase being present is required. The likelihood of this gene evolving under selection is only high in *S. pimpinellifolium*, suggesting maintenance of red coloration and lycopene accumulation is important for this species. The orange fruits of the Galapagos Islands' species have a less direct story. The colors of these fruits vary and the results from carotenoid profiling are unlikely to be robust as they represent unlikely or impossible situations, where some orange fruits apparently contain no β-carotene, leaving much room for a more thorough carotenoid profiling of these fruits. There is also no clear reason for the observed differences in color between orange fruits; an explanation may require gene expression profiling of a wider array of orange fruits. Due to high sequence similarity in the promoter of *CYC-B* in these orange fruits, an understanding of the TFs active in these fruits may also be needed to understand how this color variation is brought about. Given what is currently known, it is possible that fruit color variation in *S. cheesmaniae* and *S. galapagense* is not under selective pressure on the Galapagos Islands, producing the observed variation through drift.

Tables

Table 9 Primers used for carotenoid synthesis pathway genes and actin standard, followed by gene regions assumed to be evolving separately from fruit color.

Gene	Forward primer	Reverse primer
Psy1	5'-GCGTTTGATGTCACTTTTGCTGA-3'	5'-AGCAGGATTTCACAACACGGCTA-3'
PDS	5'-GGTAGTGCAATCGAGGGAGATGC-3'	5'-GCGCCTTCCATTGAAGCCAAGTAT-3'
ZDS	5'-ATGGGTCACCTGGATTCTTGGTT-3'	5'-TGTAAGGGTGCTCCAACTGGAAA-3'
CYC-B	5'-TCCCTCTTTTCTAAGTCCCACCA-3'	5'-TGTCCGAAAAGACACAAGCTGAG-3'
CrtR-b2	5'-TGCTGTAATTTAATGCTGTGGTCCT-3'	5'-TGAAATCCCCAGACAGCAGAATC-3'
actin	5'-TTGCTGACCGTATGAGCAAG-3' 5'-GTTGGAATGGGTCAGAAAGATGC-3'	5'-GGACAATGGATGGACCAGAC-3' 5'-GACTCACACCATCACCAGAGTCC-3'
ct066	5'-CAATCAGGACAGGTTCGTTGTTG-3'	5'-AATTGCTCTGCCACTTTCGCTAC-3'
ct093	5'-GGAAATGGACTTGCCAGTTTCTG-3'	5'-ATGTGAGCAGCCGAACTTTCTTC-3'
ct179	5'-CGAATTCATCTCCACACTCA-3'	5'-TAAGACCAGCCAAACTACCAC-3'
vac	5'-GGATTCTGATTGGATGCT-3'	5'-GTATGACCCACATAACGTG-3' 5'-GGCCCAACTATTGGTATTATT-3'

Table 10 β-carotene individual measurements showing high variability between technical replicates (indicated by -1 and -2). Species codes are shown in figure 1 with a number representing the accession used.

β-Carotene	replicate content (µg)	average content (µg)	stdev
Sg14-1	0.262746	0.192225	0.099732
Sg14-2	0.121704		
Sc11-1	n.a.	0	0
Sc11-2	n.a.		
Sc3-1	0.105072	0.097746	0.010361
Sc3-2	0.09042		
Sc19-1	n.a.	0	0
Sc19-2	n.a.		
Sn7-1	n.a.	0	0
Sn7-2	n.a.		
Spm1-1	0.416394	0.781935	0.516953
Spm1-2	1.147476		
Sg6-1	4.57215	5.600496	1.454301
Sg6-2	6.628842		
Sg7-1	0.526416	0.354882	0.242586
Sg7-2	0.183348		

Table 11 Expression levels of genes involved in the carotenoid pathway as determined by microarray analysis. The ratio gives the difference between the average species value for each gene's expression.

Gene Name	Symbol	Sc rep 1	Sc rep 2	Spm rep 1	Spm rep 2	ratio
Phytoene Synthetase	*Psy1*	6.952828	6.959477	6.590111	6.596759	0.362718
Phytoene Desaturase	*PDS*	8.455209	8.223653	8.431434	8.199878	0.023775
ζ-Carotene Desaturase	*ZDS*	6.511347	6.796353	7.112425	7.39743	-0.60108
Lycopene β-Cyclase	*CYC-B*	9.698078	9.11466	6.370464	5.787046	3.327613
β-Carotene Hydroxylase	*CrtR-b2*	6.102995	6.144434	7.940746	7.982184	-1.83775

Table 12 Mutations present in *CYC-B* coding sequence differentiating the color-fruited species (C) from the green-fruited species (G). Site number is counted from the beginning of the start codon in my alignment. For nonsynonymous mutations, the main amino acid (AA) effect difference from the green clade allele to the fixed color clade allele are given in the final column.

Site	Mutation type	G Allele	G AA	C Allele	C AA	AA Change Effect
59	Nonsyn	G/A	Lysine/Arginine	G	Arginine	Increase Hydrophilicity
60	Syn	G/A		G		
66	Nonsyn	T	Phenylalanine	G	Valine	Increase Hydrophobicity
125	Nonsyn	C/T	Proline/Leucine	T	Leucine	Hydrophilic to Hydrophobic
232	Nonsyn	G	Aspartic acid	T	Asparagine	Acidic to Neutral
249	Syn	C/T		C		
261	Syn	C/T		C		
459	Syn	G/A		A		
462	Syn	C/T		T		
795	Syn	A		G		
798	Syn	C/T		T		
912	Syn	G		T		

Table 13 SNP and indel alleles, found in the promoter region of *CYC-B*, differentiating green, red and orange-fruited species. For sites that are variable within a group, both alleles are listed. The site of the polymorphism is given in distance to the start codon in my alignment, for multiple base polymorphisms only the start site location is given.

Site	Green	Red	Orange
-1003	A/G	G	G
-956	T	C	C
-883	C/T	T	C
-793	C/T	T	T
-776	C/T	C	C
-774	A	T	T
-771	del	T	T
-753	T	del	del
-743	C/T	C	C
-742	C/T	C	C
-735	A/G	G	G
-686	AA	del	del
-679	T/G	G	G
-671	A/T	T	T
-664	A	T	T
-661	C/G	C	C
-652	C	T	T
-640	C	C	T
-527	A	G	A
-495	CGAAGTAT	del	del
-467	A/T	T	T
-440	A/C	C	C
-415	A	G	G
-412	T	C	T
-394	T/C	C	T/C
-368	G	A	A
-364	T/C	C	C
-362	A/G	G	G
-347	C	T	T
-344	A/G	A	A
-328	A	G	G
-317	CCAAATAT	del	del
-301	C/T	C/T	T
-299	G	A	A
-206	A	A	T
-142	del	T	T
-90	A/G	G	G
-78	T	G	T
-24	T	del	del
-14	del	A	A
-7	T	C	C

Table 14 π, θ and Tajima's D diversity statistics for loci sequenced, separated by group, species codes are shown in figure 1.

Statistic	*ct066*	*ct093*	*ct179*	*vac*	*CYC-B* pro	*CYC-B* orf
Spm						
π	0.00129	0.00126	0.00157	0.00793	0.00046	0.00021
θ	0.00133	0.00164	0.00172	0.00678	0.00035	0.00027
Tajima's D	-0.09241	-0.7077	-0.28831	0.64919	0.64998	0.20063
ScSg						
π	0.00069	0	0	0.00043	0.00019	0
θ	0.00102	0	0	0.00042	0.00037	0
Tajima's D	-0.9092	---	---	0.08512	-1.14053	---
SnScm						
π	0.00586	0.00057	0.00159	0.0086	0.0013	0.00245
θ	0.00457	0.0006	0.00203	0.00667	0.00157	0.00259
Tajima's D	1.14124	-0.10001	-0.82229	1.32247	-0.68914	-0.43284

Table 15 HKA results for each gene locus, species listed are compared against the outgroup *S. lycopersicoides*, species codes are shown in figure 1. Maximum likelihood (ML) values are from ML-HKA tests assuming the locus was under selection, 'No selection' ML value given for reference. Significance was determined from χ^2 distribution using one degree of freedom, except '*CYC-B* pro + orf' which had 2 degrees of freedom, significance is indicated with * for $p < 0.05$ and ** for $p < 0.01$.

	Spm		ScSg		SnScm	
Locus	**ML**	**χ^2**	**ML**	**χ^2**	**ML**	**χ^2**
No selection	-30.85		-22.4427		-32.6451	
ct066	-30.5987	0.5026	-20.3187	4.2480*	-28.9961	7.2980**
ct093	-30.1017	1.4966	-22.2438	0.3978	-32.5251	0.2400
ct179	-30.5936	0.5128	-21.2891	2.3072	-31.8547	1.5808
vac	-28.6275	4.4450*	-22.4395	0.0064	-32.4895	0.3112
CYC-B pro	-29.407	2.8860	-22.3736	0.1382	-30.9326	3.4250
CYC-B orf	-29.1158	3.4684	-21.7695	1.3464	-32.6096	0.0710
CYC-B pro + orf	-27.2	7.3000*	-21.7	1.4854	-30.9	3.4902

Table 14 Collected values of polymorphic (polym) and divergent (diver) sites for each locus used in population level analyses. Species codes are shown in figure 1. Length measurements are based on overall length in my alignments.

		Spm		ScSg		SnScm	
locus ID	**length (bp)**	**Polym**	**Diver**	**Polym**	**Diver**	**Polym**	**Diver**
ct066	654	3	10	2	11	9	7
ct093	576	3	7	0	9	1	7
ct179	961	5	45	0	45	4	35
vac	755	16	41	1	42	12	36
CYC-B pro	1005	1	26	1	28	2	33
CYC-B orf	1348	1	29	0	29	8	29

Table 15 Binding sites in the promoter region of *CYC-B* for the overall ripening regulator *RIN*. Start site given in distance to start codon in my alignment. Species containing the motif variant are given using the species codes used in figure 1.

RIN binding motifs	Start Site	Species
CCTTTATGGG	-653	Spe
CCTTTATAGG	-653	Sch, Scm, Sh, Sn, Spe, Sly
CTTTTATAGG	-653	Sc, Sg, Sl, Slc, Spm
CCTTTTTTG	-146	Spe
CTTTTTTTG	-146	Sch, Scm, Sh, Sn, Spe, Sly
CTTTTTTTTG	-146	Sc, Sg, Sl, Slc, Spm
CAATATTTTG	-104	All

Figures

Figure 13

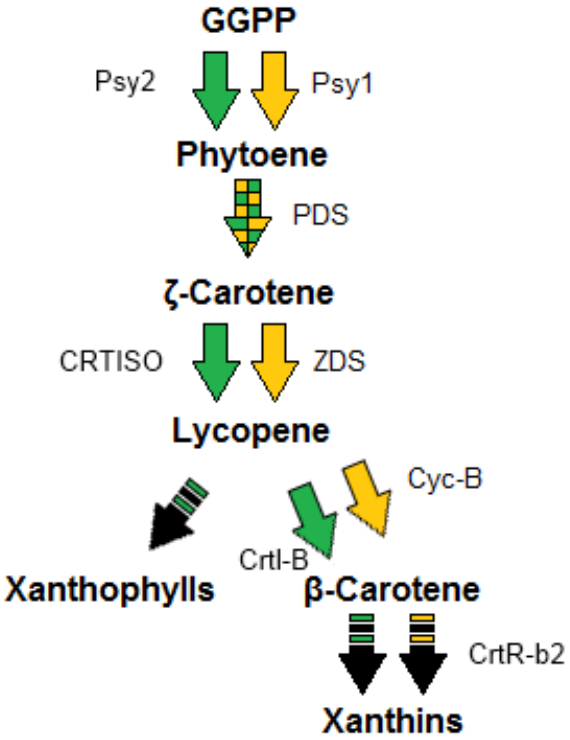


Figure 13 Carotenoid synthesis pathway, showing enzymes active in each step. Green arrows represent enzymes active in leaves, yellow arrows show enzymes active in flowers and fruits, checkered arrows are shared enzymes, and dashed arrows indicate that multiple unlisted enzymes exist between the two products (modified from Galpaz, 2006).

Figure 14

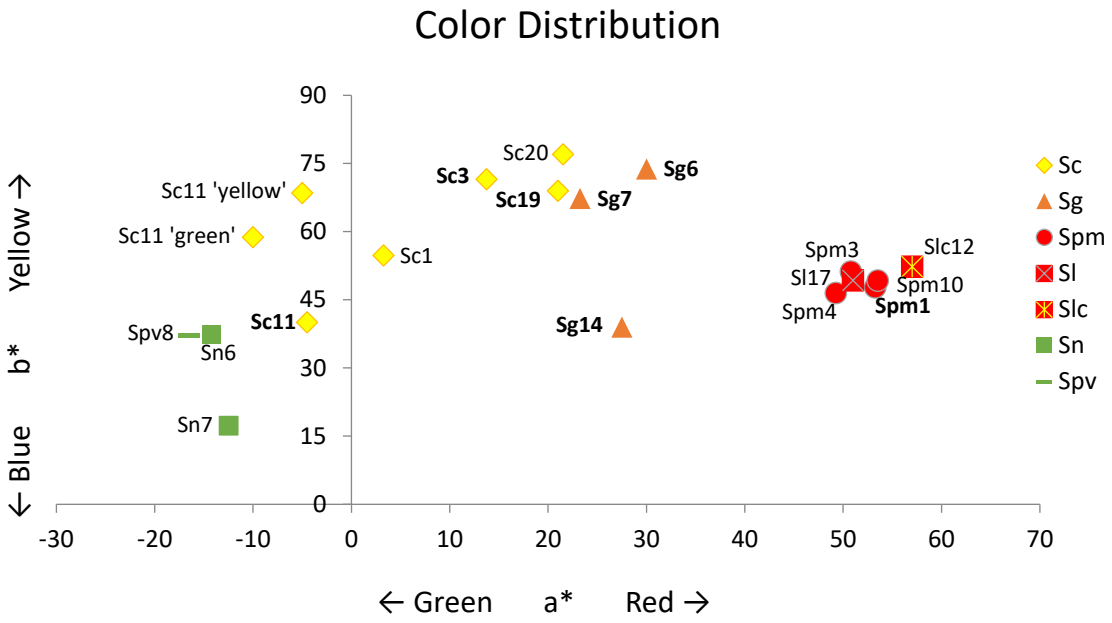
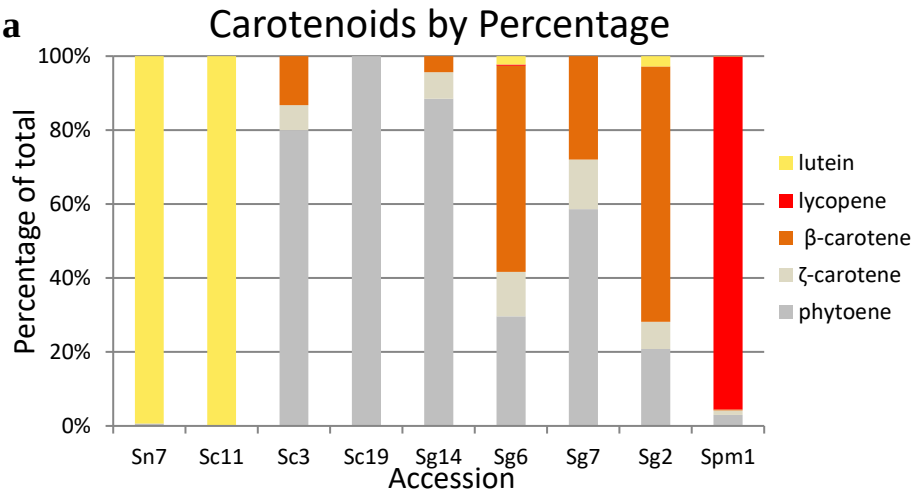


Figure 14 Color distribution chart showing the a* and b* aspects of L*a*b* color space values for fruits used in carotenoid measurements. L* measures the lightness/darkness, so it was not included here. Chart shows just the area of the color space in which these fruit colors occur; both a* and b* extend from -256 to 256. Plant names in bold represent the same fruits that were used in carotenoid HPLC. Multiple accessions from each species were used. Species codes are shown in figure 1 with a number representing the accession used.

Figure 15



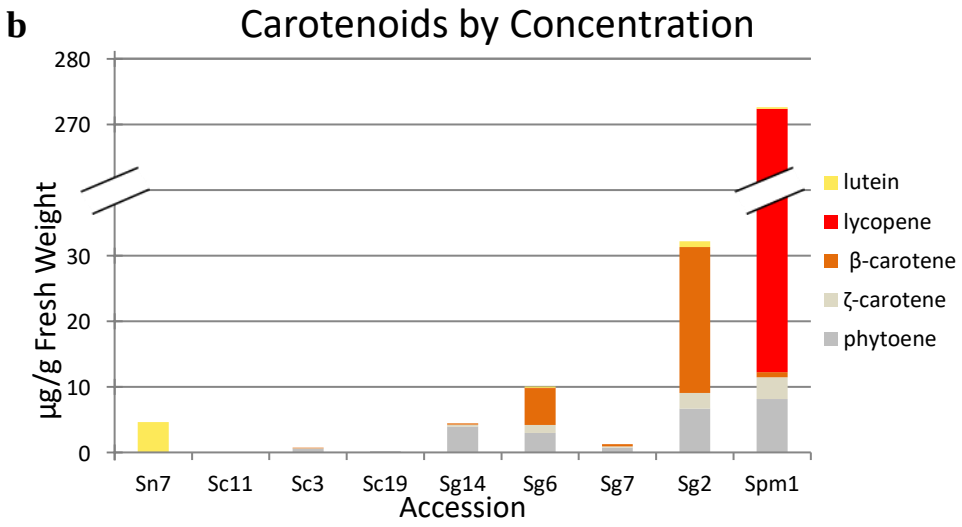


Figure 15 Carotenoid HPLC results showing (a) percentage of different carotenoids present and (b) total carotenoid concentration by weight. Species codes are shown in figure 1 followed by a number representing the accession used.

Figure 16

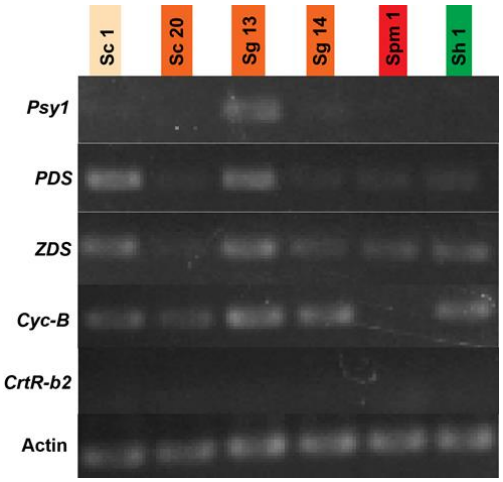


Figure 16 RT-PCR results for genes involved in the carotenoid synthesis pathway and actin standard. RNA was extracted from fruits at the turning stage, near the beginning of the ripening process. Species codes are shown in figure 1 followed by a number representing the accession used.

Figure 17

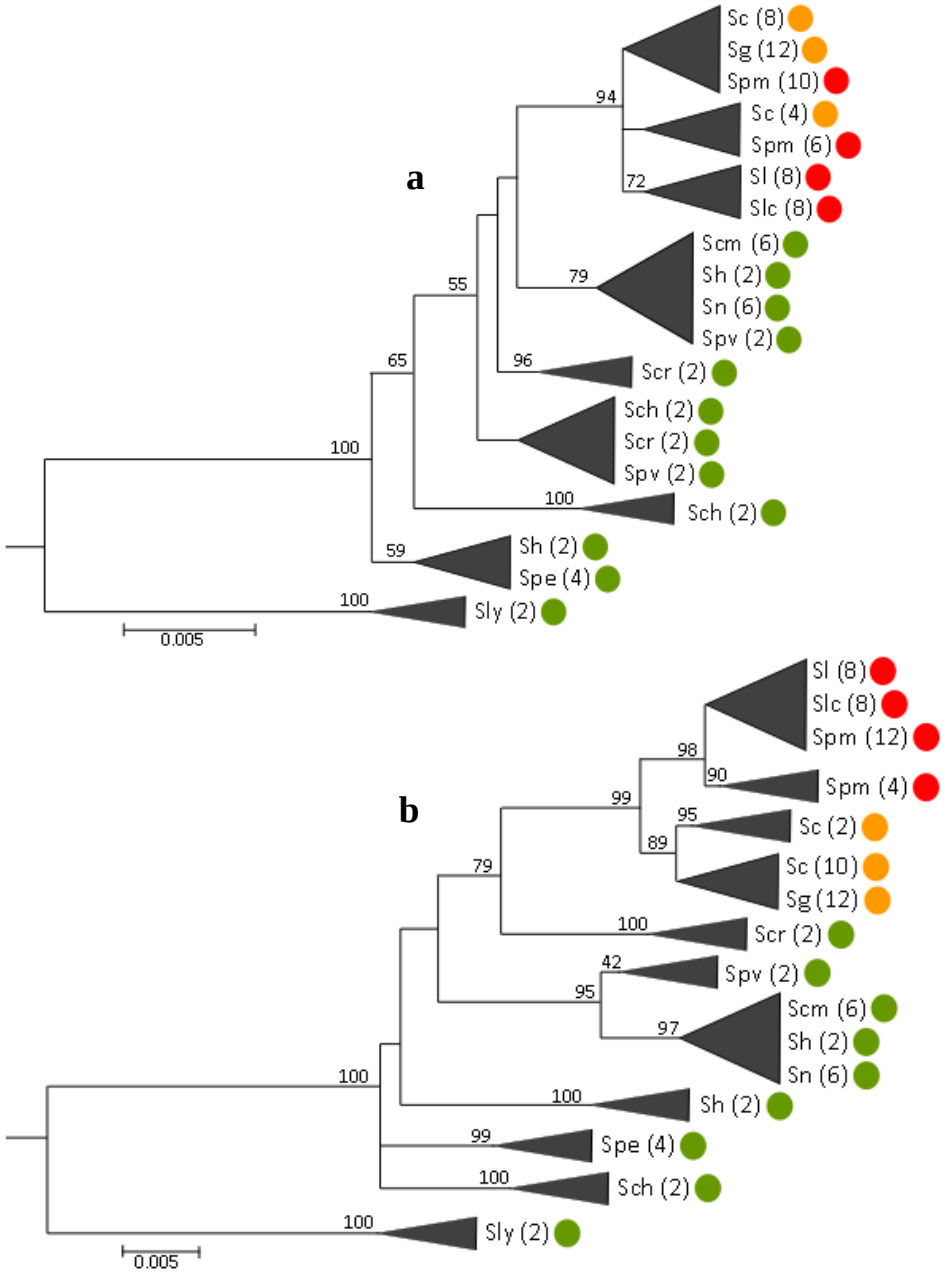
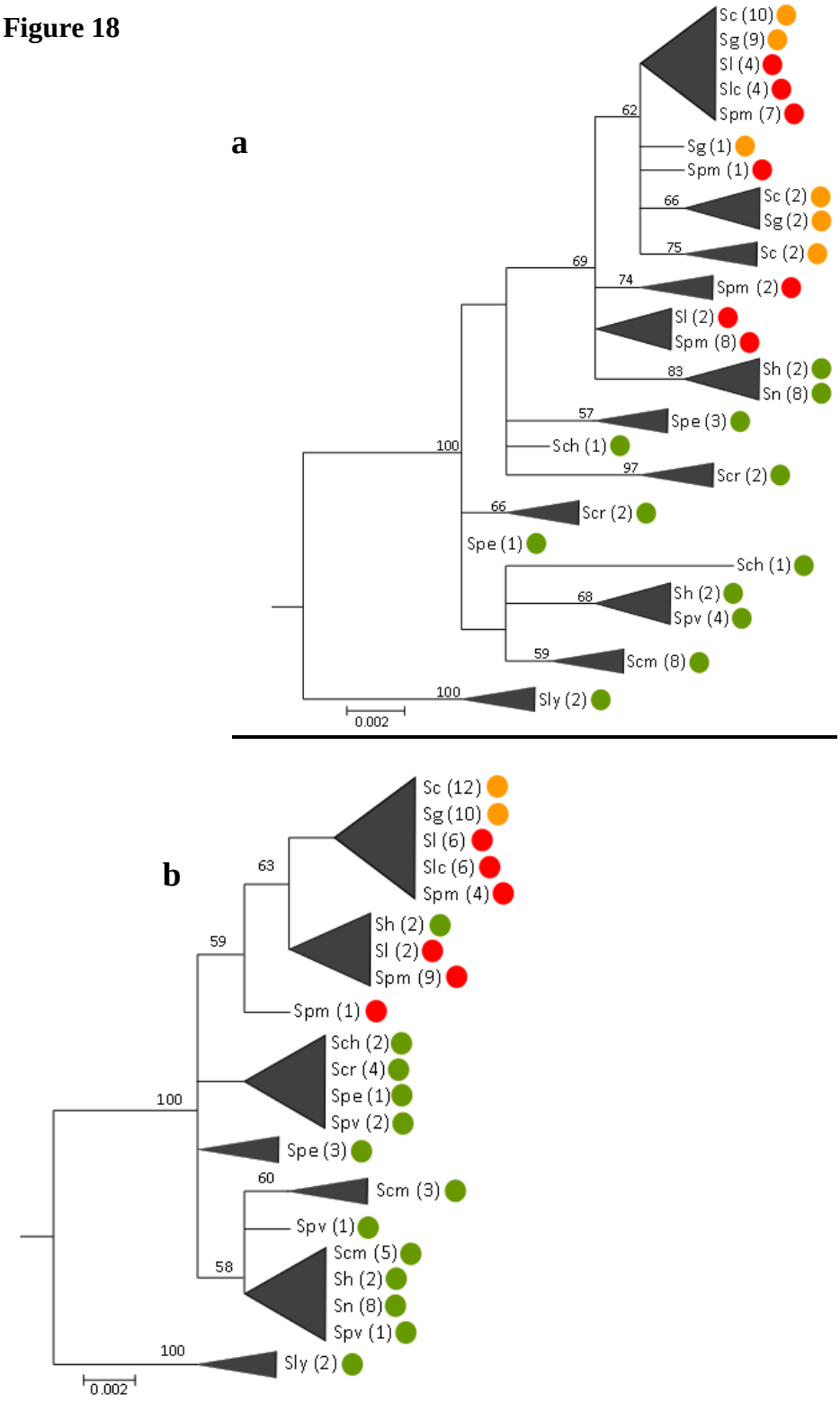


Figure 17 Trees created from sequence data of the *CYC-B* gene. (a) is created from the coding region of the gene, and (b) is from the promoter sequence. Species codes are shown in figure 1, the number in parentheses represents the number of sequences from each species in the collapsed branch, dot color indicates the color of mature fruits from the species. Bootstrap values are from 100 trees; values under 50 are not shown.

Figure 18


a
Sc (10)
Sg (9)
Sl (4)
Slc (4)
Spm (7)
62
Sg (1)
Spm (1)
66
Sc (2)
Sg (2)
75
Sc (2)
69
74
Spm (2)
Sl (2)
Spm (8)
83
Sh (2)
Sn (8)
57
Spe (3)
100
Sch (1)
97
Scr (2)
66
Scr (2)
Spe (1)
Sch (1)
68
Sh (2)
Spv (4)
59
Scm (8)
100
Sly (2)
0.002
b
Sc (12)
Sg (10)
Sl (6)
Slc (6)
Spm (4)
63
Sh (2)
Sl (2)
Spm (9)
59
Spm (1)
Sch (2)
Scr (4)
Spe (1)
Spv (2)
100
Spe (3)
60
Scm (3)
Spv (1)
Scm (5)
58
Sh (2)
Sn (8)
Spv (1)
100
Sly (2)
0.002

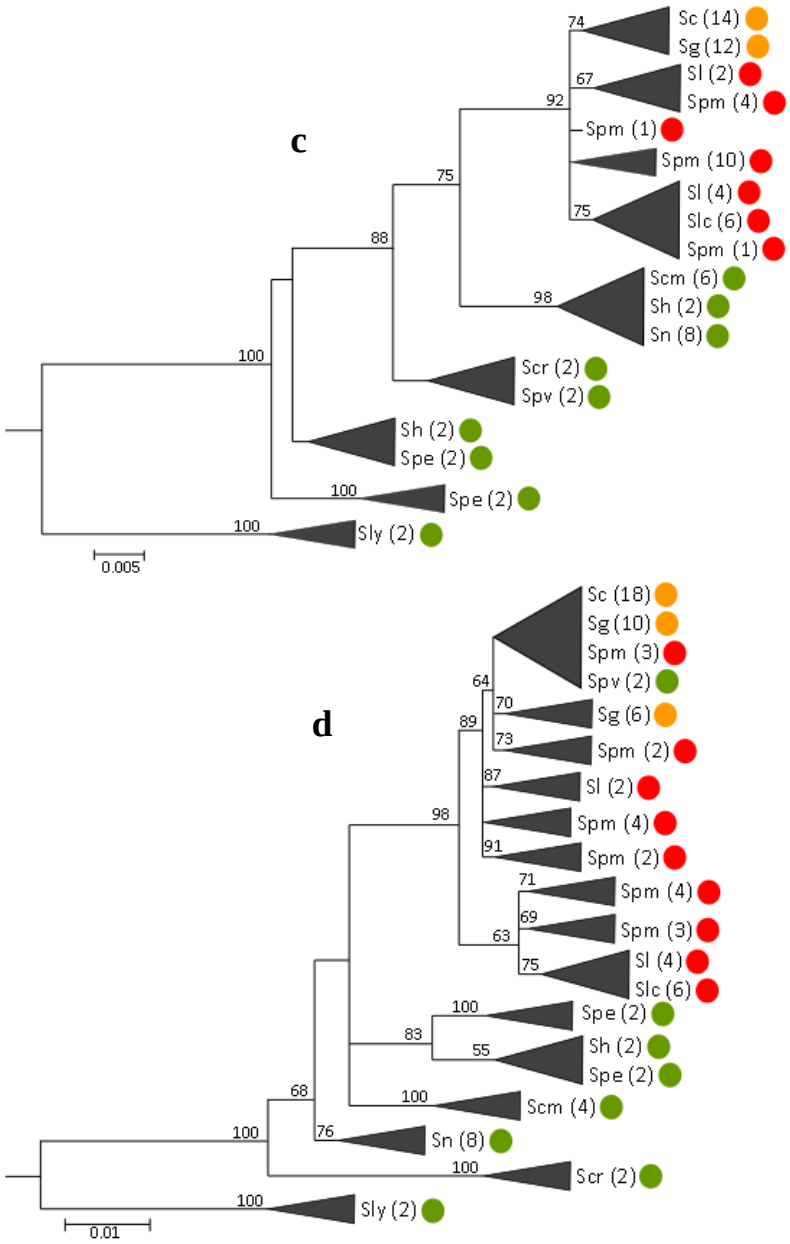


Figure 18 Additional trees created from (a) *ct066*, (b) *ct093*, (c) *ct179* and (d) *vac* regions, unrelated to carotenoid development. Species codes are shown in figure 1, the number in parentheses represents the number of sequences of each species in the collapsed branch, dot color indicates the color of mature fruits from the species. Bootstrap values are from 100 trees, values under 50 are not shown.

BIBLIOGRAPHY

Affymetrix (2011). GeneChip Tomato Genome Array | Affymetrix.

Alba, R., Payton, P., Fei, Z., McQuinn, R., Debbie, P., Martin, G.B., Tanksley, S.D., and Giovannoni, J.J. (2005). Transcriptome and Selected Metabolite Analyses Reveal Multiple Points of Ethylene Control During Tomato Fruit Development. Plant Cell *17*, 2954–2965.

Albaladejo, I., Plasencia, F.A., Meco, V., Egea, M.I., Bolarin, M.C., and Flores, F.B. (2015). Different Strategies Used by Domesticated Tomato and Wild-related Species to Confront Salt Stress. Procedia Environmental Sciences *29*, 91–92.

Asano, M., Shirota, K., Anan, T., Yamashoji, S., and Isshiki, K. (1996). Measurement of Tomatine Content in Tomatoes with Bioassay Procedure. Journal of the Japanese Society for Food Science and Technology-Nippon *43*, 275–280.

Baker, H.G., Baker, I., and Hodges, S.A. (1998). Sugar Composition of Nectars and Fruits Consumed by Birds and Bats in the Tropics and Subtropics. Biotropica *30*, 559–586.

Baldwin, M.W., Toda, Y., Nakagita, T., O'Connell, M.J., Klasing, K.C., Misaka, T., Edwards, S.V., and Liberles, S.D. (2014). volution of Sweet Taste Perception in Hummingbirds by Transformation of the Ancestral Umami Receptor. Science *345*, 929–933.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software *67*, 1–48.

Bennett, A.B. (2012). Taste: Unraveling Tomato Flavor. Current Biology *22*, R443–R444.

Berman, J., Zorrilla-López, U., Farré, G., Zhu, C., Sandmann, G., Twyman, R.M., Capell, T., and Christou, P. (2014). Nutritionally Important Carotenoids as Consumer Products. Phytochemistry Reviews *14*, 727–743.

Bianchetti, R.E., Cruz, A.B., Oliveira, B.S., Demarco, D., Purgatto, E., Peres, L.E.P., Rossi, M., and Freschi, L. (2017). Phytochromobilin Deficiency Impairs Sugar Metabolism Through the Regulation of Cytokinin and Auxin Signaling in Tomato Fruits. Scientific Reports *7*, 7822.

Blomberg, S.P., Garland, T., and Ives, A.R. (2003). Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits are More Labile. Evolution *57*, 717–745.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., et al. (2014). The Genome of the Stress-Tolerant Wild Tomato Species *Solanum pennellii*. Nature Genetics *46*, 1034–1038.

Bolmgren, K., and Eriksson, O. (2010). Seed Mass and the Evolution of Fleshy Fruits in Angiosperms. OIKOS *119*, 707–718.

Breitel, D.A., Chappell-Maor, L., Meir, S., Panizel, I., Puig, C.P., Hao, Y., Yifhar, T., Yasuor, H., Zouine, M., Bouzayen, M., et al. (2016). AUXIN RESPONSE FACTOR 2 Intersects Hormonal Signals in the Regulation of Tomato Fruit Ripening. Plos Genetics *12*, e1005903.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL Mapping in Experimental Crosses. Bioinformatics *19*, 889–890.

Caicedo, A.L., and Schaal, B.A. (2004). Population Structure and Phylogeography of *Solanum pimpinellifolium* Inferred from a Nuclear Gene. Molecular Ecology *13*, 1871–1882.

Canady, M.A., Meglic, V., and Chetelat, R.T. (2005). A Library of *Solanum lycopersicoides* Introgression Lines in Cultivated Tomato. Genome *48*, 685–697.

Cantu, D., Vicente, A.R., Greve, L.C., Dewey, F.M., Bennett, A.B., Labavitch, J.M., and Powell, A.L.T. (2008). The Intersection Between Cell Wall Disassembly, Ripening, and Fruit Susceptibility to *Botrytis cinerea*. Proceedings of the National Academy of Sciences *105*, 859–864.

Carey, A.T., Smith, D.L., Harrison, E., Bird, C.R., Gross, K.C., Seymour, G.B., and Tucker, G.A. (2001). Down-regulation of a Ripening-related beta-Galactosidase gene (TBG1) in Transgenic Tomato Fruits. Journal of Experimental Botany *52*, 663–668.

Carrillo-Perdomo, E., Aller, A., Cruz-Quintana, S.M., Giampieri, F., and Alvarez-Suarez, J.M. (2015). Andean Berries from Ecuador: A Review on Botany, Agronomy, Chemistry and Health Potential. Journal of Berry Research *5*, 49–69.

Chang, W.-C., Lee, T.-Y., Huang, H.-D., Huang, H.-Y., and Pan, R.-L. (2008). PlantPAN: Plant Promoter Analysis Navigator, for Identifying Combinatorial Cis-Regulatory Elements with Distance Constraint in Plant Gene Groups. BMC Genomics *9*, 561.

Chetelat, R., Rick, C., and Deverna, J. (1989). Isozyme Analysis, Chromosome-Pairing, and Fertility of *Lycopersicon esculentum* x *Solanum lycopersicoides* Diploid Backcross Hybrids. Genome *32*, 783–790.

Chetelat, R.T., Klann, E., DeVerna, J.W., Yelle, S., and Bennett, A.B. (1993). Inheritance and Genetic Mapping of Fruit Sucrose Accumulation in *Lycopersicon chmielewskii*. The Plant Journal *4*, 643–650.

Darwin, S.C., Knapp, S., and Peralta, I.E. (2003). Taxonomy of Tomatoes in the Galápagos Islands: Native and Introduced Species of *Solanum* section *Lycopersicon* (Solanaceae). Systematics and Biodiversity *1*, 29–53.

Del Rio, C.M., and Stevens, B.R. (1989). Physiological Constraint on Feeding Behavior: Intestinal Membrane Disaccharidases of the Starling. Science *243*, 794–796.

deVicente, M.C., and Tanksley, S.D. (1993). Qtl Analysis of Transgressive Segregation in an Interspecific Tomato Cross. Genetics *134*, 585–596.

Eigenbrode, S.D., and Trumble, J.T. (1994). Fruit-Based Tolerance to Damage by Beet Armyworm (Lepidoptera: Noctuidae) in Tomato. Environmental Entomology *23*, 937–942.

Eriksson, O., Friis, E., and Löfgren, P. (2000). Seed Size, Fruit Size, and Dispersal Systems in Angiosperms from the Early Cretaceous to the Late Tertiary. The American Naturalist *156*, 47–58.

Floerchinger, M., Braun, J., Boehning-Gaese, K., and Schaefer, H.M. (2010). Fruit Size, Crop Mass, and Plant Height Explain Differential Fruit Choice of Primates and Birds. Oecologia (Berlin) *164*, 151–161.

Forkmann, G. (1991). Flavonoids as Flower Pigments: The Formation of the Natural Spectrum and its Extension by Genetic Engineering. Plant Breeding *106*, 1–26.

Fraser, P.D., and Bramley, P.M. (2004). The Biosynthesis and Nutritional Uses of Carotenoids. Progress in Lipid Research *43*, 228–265.

Fridman, E., Amit, G., and Zamir, D. (2001). Phenotyping, Fine Mapping and Cloning of a Wild Species QTL That Increases Horticultural Yield in Tomatoes (Mackay: Australian Society of Sugar Cane Technologists).

Friedman, M., and Levin, C.E. (1998). Dehydrotomatine Content in Tomatoes. Journal of Agricultural and Food Chemistry *46*, 4571–4576.

Friedman, M., Henika, P.R., and Mackey, B.E. (1996). Feeding of Potato, Tomato and Eggplant Alkaloids Affects Food Consumption and Body and Liver Weights in Mice. The Journal of Nutrition *126*, 989–999.

Furui, H., Inakuma, T., Ishiguro, Y., and Kiso, M. (1997). Absorptiometric Measurement of Tomatine in Tomatoes. Nippon Nogeikagaku Kaishi-Journal of The Japan Society for Bioscience *71*, 777–782.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy---Analysis of Affymetrix GeneChip Data at the Probe Level. Bioinformatics *20*, 307–315.

Gentleman, R.C., Carey, V.J., Bates, D.M., and Others (2004). Bioconductor: Open Software Development for Computational Biology and Bioinformatics. Genome Biology *5*, R80.

Grumet, R., Fobes, J.F., and Herner, R.C. (1981). Ripening Behavior of Wild Tomato Species. Plant Physiology *68*, 1428–1432.

Güntner, C., González, A., Reis, R.D., González, G., Vázquez, A., Ferreira, F., and Moyna, P. (1997). Effect of Solanum Glycoalkaloids on Potato Aphid, *Macrosiphum euphorbiae*. Journal of Chemical Ecology *23*, 1651–1659.

Hajjar, R., and Hodgkin, T. (2007). The Use of Wild Relatives in Crop Improvement: A Survey of Developments Over the Last 20 Years. Euphytica *156*, 1–13.

Hall, T. (2013). BioEdit (Ibis Biosciences).

Hao, Y., Hu, G., Breitel, D., Liu, M., Mila, I., Frasse, P., Fu, Y., Aharoni, A., Bouzayen, M., and Zouine, M. (2015). Auxin Response Factor SlARF2 Is an Essential Component of the Regulatory Mechanism Controlling Fruit Ripening in Tomato. Plos Genetics *11*, e1005649.

Hetherington, S.E., Smillie, R.M., and Davies, W.J. (1998). Photosynthetic Activities of Vegetative and Fruiting Tissues of Tomato. Journal of Experimental Botany *49*, 1173–1181.

Hillis, D.M., Maple, B.K., Larson, A., Davis, S.K., and Zimmer, E.A. (1996). Molecular Systematics (Sinauer, Sunderland, MA).

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal *50*, 346–363.

Hudson, R.R., Kreitman, M., and Aguade, M. (1987). A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics *116*, 153–159.

Iijima, Y., Watanabe, B., Sasaki, R., Takenaka, M., Ono, H., Sakurai, N., Umemoto, N., Suzuki, H., Shibata, D., and Aoki, K. (2013). Steroidal Glycoalkaloid Profiling and Structures of Glycoalkaloids in Wild Tomato Fruit. Phytochemistry *95*, 145–157.

Iseki, N., Sasaki, A., and Toju, H. (2011). Arms Race Between Weevil Rostrum Length and Camellia Pericarp Thickness: Geographical Cline and Theory. Journal of Theoretical Biology *285*, 1–9.

Ishimaru, M., Smith, D.L., Mort, A.J., and Gross, K.C. (2009). Enzymatic Activity and Substrate Specificity of Recombinant Tomato Beta-Galactosidases 4 and 5. Planta *229*, 447–456.

Kimura, S., and Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A Model Fruit-Bearing Crop. Cold Spring Harbor Protocols *2008*, pdb.emo105.

Klann, E., Chetelat, R., and Bennett, A. (1993). Expression of Acid Invertase Gene Controls Sugar Composition in Tomato (*Lycopersicon*) Fruit. Plant Physiology *103*, 863–870.

Knapp, S., Bohs, L., Nee, M., and Spooner, D.M. (2004). Solanaceae—A Model for Linking Genomics with Biodiversity. Comparative and Functional Genomics *5*, 285.

Koressaar, T., and Remm, M. (2007). Enhancements and Modifications of Primer Design Program Primer3. Bioinformatics *23*, 1289–1291.

Kozukue, N., and Friedman, M. (2003). Tomatine, Chlorophyll, Beta-Carotene and Lycopene Content in Tomatoes During Growth and Maturation. Journal of the Science of Food and Agriculture *83*, 195–200.

Li, W., and Chetelat, R.T. (2014). The Role of a Pollen-Expressed Cullin1 Protein in Gametophytic Self-Incompatibility in *Solanum*. Genetics *196*, 439–442.

Librado, P., and Rozas, J. (2009). DnaSP v5: A Software for Comprehensive Analysis of DNA Polymorphism Data. Bioinformatics *25*, 1451–1452.

Lin, T., Huang, S., Mazzucato, A., Causse, M., Fei, Z., Giovannoni, J.J., Chetelat, R.T., Zamir, D., Städler, T., Zhu, G., et al. (2014). Genomic Analyses Provide Insights into the History of Tomato Breeding. Nature Genetics *46*, 1220–1226.

Mack, A. (2000). Did Fleshy Fruit Pulp Evolve as a Defence Against Seed Loss Rather Than as a Dispersal Mechanism? Journal of Biosciences *25*, 93–97.

Marshall, J.A., Knapp, S., Davey, M.R., Power, J.B., Cocking, E.C., Bennett, M.D., and Cox, A.V. (2001). Molecular Systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) Using the Nuclear ITS rDNA Region. Theoretical and Applied Genetics *103*, 1216–1222.

Mecham, B.H., Nelson, P.S., and Storey, J.D. (2010). Supervised Normalization of Microarrays. Bioinformatics *26*, 1308–1315.

Menda, N., Strickler, S.R., Edwards, J.D., Bombarely, A., Dunham, D.M., Martin, G.B., Mejia, L., Hutton, S.F., Havey, M.J., Maxwell, D.P., et al. (2014). Analysis of Wild-Species Introgressions in Tomato Inbreds Uncovers Ancestral Origins. Bmc Plant Biology *14*, 287.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In Proceedings of the Gateway Computing Environments Workshop, (New Orleans, LA), pp. 1–8.

Ministerio de Agricultura y Desarrollo Rural, B.C. (2011). Anuario Estadístico del Sector Agropecuario Resultados Evaluaciones Agropecuarias Municipales 2011. Documentación E Información, Datos Estadísticos Y Sistemas de Información *November 2012 ed.*

Miron, D., Petreikov, M., Carmi, N., Shen, S., Levin, I., Granot, D., Zamski, E., and Schaffer, A.A. (2002). Sucrose Uptake, Invertase Localization and Gene Expression in Developing Fruit of *Lycopersicon esculentum* and the Sucrose-Accumulating *Lycopersicon hirsutum*. Physiologia Plantarum *115*, 35–47.

Mohan, V., Pandey, A., Sreelakshmi, Y., and Sharma, R. (2016). Neofunctionalization of Chromoplast Specific Lycopene Beta Cyclase Gene (CYC-B) in Tomato Clade. PLOS ONE *11*, e0153333.

Moyle, L.C. (2008). Ecological and Evolutionary Genomics in the Wild Tomatoes (*Solanum*sect. *Lycopersicon*). Evolution *62*, 2995–3013.

Mulatu, B., Applebaum, S., Kerem, Z., and Coll, M. (2006). Tomato Fruit Size, Maturity and alpha-Tomatine Content Influence the Performance of Larvae of Potato Tuber Moth *Phthorimaea operculella* (Lepidoptera : Gelechiidae). Bulletin of Entomological Research *96*, 173–178.

Nesbitt, T.C., and Tanksley, S.D. (2002). Comparative Sequencing in the Genus *Lycopersicon*: Implications for the Evolution of Fruit Size in the Domestication of Cultivated Tomatoes. Genetics *162*, 365–379.

Niyogi, K.K. (2000). Safety Valves for Photosynthesis. Current Opinion in Plant Biology *3*, 455–460.

Ökmen, B., Etalo, D.W., Joosten, M.H.A.J., Bouwmeester, H.J., de Vos, R.C.H., Collemare, J., and de Wit, P.J.G.M. (2013). Detoxification of α-Tomatine by *Cladosporium fulvum* is Required for Full Virulence on Tomato. New Phytologist *198*, 1203–1214.

Pailles, Y., Ho, S., Pires, I.S., Tester, M., Negrão, S., and Schmöckel, S.M. (2017). Genetic Diversity and Population Structure of Two Tomato Species from the Galapagos Islands. Frontiers in Plant Science *8*.

Pease, J.B., Haak, D.C., Hahn, M.W., and Moyle, L.C. (2016a). Data from: Phylogenomics Reveals Three Sources of Adaptive Variation During a Rapid Radiation.

Pease, J.B., Haak, D.C., Hahn, M.W., and Moyle, L.C. (2016b). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLOS Biology *14*, e1002379.

Powell, A.L.T., Nguyen, C.V., Hill, T., Cheng, K.L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernández-Muñoz, R., Vicente, A., et al. (2012). Uniform Ripening Encodes a Golden 2-like Transcription Factor Regulating Tomato Fruit Chloroplast Development. Science *336*, 1711–1715.

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing *Vienna, Austria*.

Ramirez, I. (1990). Why do Sugars Taste Good? Neuroscience & Biobehavioral Reviews *14*, 125–134.

Ranc, N., Muños, S., Santoni, S., and Causse, M. (2008). A Clarified Position for *Solanum lycopersicum var. cerasiforme* in The Evolutionary History of Tomatoes (Solanaceae). BMC Plant Biology *8*, 130.

Restrepo, A., Colinvaux, P., Bush, M., Correa-Metrio, A., Conroy, J., Gardener, M.R., Jaramillo, P., Steinitz-Kannan, M., and Overpeck, J. (2012). Impacts of Climate Variability and Human Colonization on the Vegetation of the Galápagos Islands. Ecology *93*, 1853–1866.

Reuscher, S., Akiyama, M., Yasuda, T., Makino, H., Aoki, K., Shibata, D., and Shiratake, K. (2014). The Sugar Transporter Inventory of Tomato: Genome-Wide Identification and Expression Analysis. Plant and Cell Physiology *55*, 1123–1141.

Reuscher, S., Fukao, Y., Morimoto, R., Otagaki, S., Oikawa, A., Isuzugawa, K., and Shiratake, K. (2016). Quantitative Proteomics-Based Reconstruction and Identification of Metabolic Pathways and Membrane Transport Proteins Related to Sugar Accumulation in Developing Fruits of Pear (*Pyrus communis*). Plant and Cell Physiology *57*, 505–518.

Rick, C.M. (1988). Evolution of Mating Systems in Cultivated Plants. In Plant Evolutionary Biology, L.D. Gottlieb, and S.K. Jain, eds. (Dordrecht: Springer Netherlands), pp. 133–147.

Rick, C.M., and Bowman, R.I. (1961). Galápagos Tomatoes and Tortoises. Evolution *15*, 407–417.

Rick, C.M., and Chetelat, R.T. (1995). Utilization of Related Wild Species for Tomato Improvement. Acta Horticulturae 21–38.

Rick, C.M., and Smith, P.G. (1953). Novel Variation in Tomato Species Hybrids. The American Naturalist *87*, 359–373.

Rick, C.M., and TGRC (2013). Recommendations for Flowering and Reproducing Wild Tomato Species.

Rick, C., Uhlig, J., and Jones, A. (1994a). High alpha-Tomatine Content in Ripe Fruit of Andean *Lycopersicon esculentum var cerasiforme* - Developmental and Genetic Apsects. Proceedings of The National Academy of Sciences of the United States *91*, 12877–12881.

Rick, C.M., Chetelat, R.T., and DeVerna, J.W. (1988). Recombination in Sesquidiploid Hybrids of *Lycopersicon esculentum* x *Solanum lycopersicoides* and Derivatives. Theoretical and Applied Genetics *76*, 647–655.

Rick, C.M., Cisneros, P., Chetelat, R.T., and DeVerna, J.W. (1994b). Abg—A Gene on Chromosome 10 for Purple Fruit Derived from *S. lycopersicoides*. Report of the Tomato Genetics Coop *44*, 29–30.

Rodriguez, F., Wu, F., Ané, C., Tanksley, S., and Spooner, D.M. (2009). Do Potatoes and Tomatoes Have a Single Evolutionary History, and What Proportion of the Genome Supports This History? BMC Evolutionary Biology *9*, 191.

Sargent, S.A., and Moretti, C.L. (2014). Tomato Commodity Summary. USDA, ARS The Commercial Storage of Fruits, Vegetables, and Florist and Nursery Stocks.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko, T., Nakamura, Y., Shibata, D., Aoki, K., et al. (2012). The Tomato Genome Sequence Provides Insights into Fleshy Fruit Evolution. Nature *485*, 635–641.

Sitrit, Y., Hadfield, K.A., Bennett, A.B., Bradford, K.J., and Downie, A.B. (1999). Expression of a Polygalacturonase Associated with Tomato Seed Germination. Plant Physiology *121*, 419–428.

Smith, D.L., and Gross, K.C. (2000). A Family of at Least Seven β-Galactosidase Genes Is Expressed during Tomato Fruit Development. Plant Physiology *123*, 1173–1184.

Spooner, D., Peralta, I., and Knapp, S. (2005). Comparison of AFLPs with other Markers for Phylogenetic Inference in Wild Tomatoes (*Solanum* section *Lycopersicon*). TAXON *54*, 43–61.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics btu033.

Statista (2013). World Vegetable Production by Type 2013 | Statistic.

Stommel, J.R., and Haynes, K.G. (1994). Inheritance of Beta-Carotene Content in the Wild Tomato Species *Lycopersicon cheesmanii*. Journal of Heredity *85*, 401–404.

Taiz, L., Zeiger, E., Møller, I.M., and Murphy, A.S. (2015). Plant Physiology and Development (Sunderland, Massachusetts: Sinauer Associates, Inc., Publishers).

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics *123*, 585–595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution *30*, 2725–2729.

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of Plant Pigments: Anthocyanins, Betalains and Carotenoids. The Plant Journal *54*, 733–749.

Tieman, D., Bliss, P., McIntyre, L.M., Blandon-Ubeda, A., Bies, D., Odabasi, A.Z., Rodríguez, G.R., van der Knaap, E., Taylor, M.G., Goulet, C., et al. (2012). The Chemical Interactions Underlying Tomato Flavor Preferences. Current Biology *22*, 1035–1039.

Tomassen, M.M.M., Barrett, D.M., van der Valk, H.C.P.M., and Woltering, E.J. (2007). Isolation and Characterization of a Tomato Non-Specific Lipid Transfer Protein Involved in Polygalacturonase-Mediated Pectin Degradation. Journal of Experimental Botany *58*, 1151–1160.

Traugott, M.S., and Stamp, N.E. (1997). Effects of Chlorogenic Acid- and Tomatine-Fed Caterpillars on Performance of an Insect Predator. Oecologia *109*, 265–272.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., and Rozen, S. (2012). Primer3 - New Capabilities and Interfaces. Nucleic Acids Research *40*, e115.

Van der Pijl, L. (1969). Principles of Dispersal in Higher Plants (Springer, Berlin Heidelberg New York).

Velterop, J.S., and Vos, F. (2001). A Rapid and Inexpensive Microplate Assay for the Enzymatic Determination of Glucose, Fructose, Sucrose, L-Malate and Citrate in Tomato (*Lycopersicon esculentum*) Extracts and in Orange Juice. Phytochemical Analysis *12*, 299–304.

Wang, Z., and Schaefer, H.M. (2014). Limits of Selection Against Cheaters: Birds Prioritise Visual Fruit Advertisement Over Taste. Oecologia *174*, 1293–1300.

Wang, A., Li, J., Zhang, B., Xu, X., and Bewley, J.D. (2009). Expression and Location of Endo-Beta-Mannanase During the Ripening of Tomato Fruit, and the Relationship Between its Activity and Softening. Journal of Plant Physiology *166*, 1672–1684.

Wang, L., Li, J., Zhao, J., and He, C. (2015). Evolutionary Developmental Genetics of Fruit Morphological Variation Within the Solanaceae. Frontiers in Plant Science *6*.

Willson, M.F., and Whelan, C.J. (1990). The Evolution of Fruit Color in Fleshy-Fruited Plants. The American Naturalist *136*, 790–809.

Willson, M., Irvine, A., and Walsh, N. (1989). Vertebrate Dispersal Syndromes in Some Australian and New Zealand Plant Communities, with Geographic Comparisons. Biotropica *21*, 133–147.

Wright, S.I., and Charlesworth, B. (2004). The HKA Test Revisited: A Maximum-Likelihood-Ratio Test of the Standard Neutral Model. Genetics *168*, 1071–1076.

Zaccardelli, M., Campanile, F., Cammareri, M., and Grandillo, S. (2011). Agronomical use of α-Tomatine and Crude Extracts of *Solanum* Spp. to Control Phytopathogenic Fungi. Acta Horticulturae 401–404.

Zuriaga, E., Blanca, J., and Nuez, F. (2008). Classification and Phylogenetic Relationships in *Solanum* Section *Lycopersicon* Based on AFLP and Two Nuclear Gene Sequences. Genetic Resources and Crop Evolution *56*, 663–678.