

Solanum galapagense-derived purple tomato fruit color is conferred by novel alleles of the *anthocyanin fruit* and *atroviolacium* loci

Sean Fenstemaker¹  | Leah Sim¹ | Jessica Cooperstone^{2,3}  | David Francis¹ 

¹Department of Horticulture and Crop Science, The Ohio State University, Wooster, Ohio, USA

²Department of Food Science and Technology, The Ohio State University, Columbus, Ohio, USA

³Department of Horticulture and Crop Science, The Ohio State University, Columbus, Ohio, USA

Correspondence

David Francis, Department of Horticulture and Crop Science, The Ohio State University, Wooster, OH 44691, USA.
Email: francis.77@osu.edu

Funding information

Hatch, Grant/Award Number: OHO014405; USDA AFRI Specialty Crops Research Initiative, Grant/Award Number: 2016-51181-25404

Abstract

One hypothesis for the origin of endemic species of tomato on the Galápagos islands postulates a hybridization of *Solanum pimpinellifolium* and *Solanum habrochaites*. *Solanum galapagense* accession LA1141 has purple fruit pigmentation, previously described in green-fruited wild tomatoes such as *S. habrochaites* or *Solanum chilense*. Characterization of LA1141 derived purple pigmentation provides a test of the hybridization hypothesis. Purple pigmentation was recovered in progenies derived from LA1141, and the anthocyanins malvidin 3-(coumaroyl)rutinoside-5-glucoside, petunidin 3-(coumaroyl) rutinoside-5-glucoside, and petunidin 3-(caffeoyl)rutinoside-5-glucoside were abundant. Fruit color was evaluated in an introgression population, and three quantitative trait loci (QTLs) were mapped and validated in subsequent populations. The loci *atroviolacium* on chromosome 7, *Anthocyanin fruit* on chromosome 10, and *uniform ripening* also on chromosome 10 underly these QTLs. Sequence analysis suggested that the LA1141 alleles of *Aft* and *atv* are unique relative to those previously described from *S. chilense* accession LA0458 and *Solanum cheesmaniae* accession LA0434, respectively. Phylogenetic analysis of the LA1141 *Aft* genomic sequence did not support a green-fruited origin, and the locus clustered with members of the red-fruited tomato clade. The LA1141 allele of *Aft* is not the result of an ancient introgression from the green-fruited clade and underlies a gain of anthocyanin pigmentation in the red-fruited clade.

Highlight

Anthocyanin fruit and *atroviolacium* confer purple pigmentation in *Solanum galapagense* LA1141, confirming a mechanism described for green-fruited tomatoes. LA1141 alleles cluster with red-fruited homologs suggesting an independent gain of pigmentation.

KEYWORDS

anthocyanin fruit, *atroviolacium*, Galápagos Islands, inbred backcross (IBC), LA1141, phylogenetics, purple, quantitative trait loci (QTL), *S. galapagense*, tomato

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Plant Direct* published by American Society of Plant Biologists and the Society for Experimental Biology and John Wiley & Sons Ltd.



1 | INTRODUCTION

Rick (1961) hypothesized that species of tomato endemic to the Galápagos, *Lycopersicon cheesmanii* f. *minor*, now classified as *Solanum galapagense*, might have resulted from the hybridization of *Solanum pimpinellifolium* and *Solanum habrochaites* progenitors. This hypothesis was based on three unique traits found in both *S. habrochaites* and *S. galapagense*. Both species have alleles of *B* capable of conferring high β -carotene (Lincoln & Porter, 1950; Tomes et al., 1954). *S. galapagense* also possesses an accrescent calyx and pubescence reminiscent of *S. habrochaites* (Rick, 1961). *S. galapagense* accession LA1141 has purple pigmentation in immature fruit, similar to species in the green-fruited tomato clade, including *S. habrochaites*. The presence of this fourth trait common to *S. galapagense* and *S. habrochaites* suggested that characterizing the chemical and genetic basis of purple fruit derived from LA1141 could provide a test of Rick's, 1961 hypothesis.

The endemic Galápagos tomatoes possess morphological and physiological traits that distinguish them from other wild species. These traits include orange fruit color at maturity, yellow-green foliage, tiny seed size, seed dormancy, and affinity for dry conditions (Darwin et al., 2003; Rick, 1961). These species can hybridize easily with cultivated tomatoes, making them useful donors of novel alleles (Rick, 1961). Several genes from Galápagos tomatoes have been used in breeding contemporary varieties. An allele of *uniform ripening* (*u*) from *Solanum cheesmaniae* accession LA0428 is responsible for the uniform distribution of light green pigmentation in immature fruits (Rick, 1967). Alleles conferring *jointless* (j^2) pedicel (Rick, 1956) and *arthritic articulation* (j^{2in}) (Joubert, 1962, available at: <https://tgc.ifas.ufl.edu/vol12/Volume12.pdf>) have enabled mechanical harvest. *S. cheesmaniae* accession LA0422 has a recessive allele, *anthocyanin gainer* (ag^2), which results in fruit and foliage lacking anthocyanin at early developmental stages (De Jong et al., 2004; Rick, 1967). Alleles of the *Beta* (*B*) locus found in all *S. galapagense* and *S. cheesmaniae* accessions confer high β -carotene and the characteristic orange fruit (Orchard et al., 2021). Alleles of *B* from LA0317 and LA0166 have been introgressed into cultivated tomatoes (Stommel, 2001). Anthocyanin-mediated purple pigmentation in both the fruit and foliage was described in *S. cheesmaniae* accession LA0434, the donor of the *atviolacea* (*atv*) locus (Rick, 1956; Rick, 1961; Rick, 1967). Additionally, an accession of *S. cheesmaniae* (LA0428) was described as having immature fruits with a purple color that resemble *Solanum peruvianum* (Rick, 1967). Identification and analysis of loci that confer purple fruit color may shed light on broader questions about the evolutionary history of the Galápagos tomatoes.

Water-soluble vacuolar pigments called anthocyanins cause purple fruit pigmentation in species of *Solanum* (Chaves-Silva et al., 2018; Mes et al., 2008; Timberlake & Bridle, 1982). Within the taxonomic framework of tomato (*Solanum* sect. *Lycopersicon*) (Peralta et al., 2008), the red-fruited clade corresponds to the group *Esclulentum* (Pease et al., 2016), which generally lack anthocyanins in the fruit. The green-fruited clade is grouped into *Arcanum*, *Peruvianum*, and *Hirsutum* based on whole-transcriptome concatenated molecular clock phylogeny (Pease et al., 2016). Purple pigmentation is a

characteristic found throughout the green-fruited tomato clade. For example, *S. habrochaites* accession LA1777 has pronounced anthocyanin spots in its fruit (Dal Cin et al., 2009). Additionally, *S. peruvianum* fruit is purple-tinged with pigment in lines and blotches (Muller, 1940). The chemical basis of purple fruit derived from tomato species in the green-fruited clade is attributed to the anthocyanins petunidin and malvidin (Jones et al., 2003; Mathews et al., 2003; Ooe et al., 2016). Two loci affect the regulation of anthocyanin accumulation in tomato fruit, one on chromosome 7 and a second on chromosome 10. A nonfunctional R3 MYB repressor on chromosome 7 underlies the *atv* locus (Cao et al., 2017). On chromosome 10, functional R2R3 MYB-encoding activator genes underly the *Anthocyanin fruit* (*Aft*) locus described in the donor parent *Solanum chilense* accession LA0458 (Georgiev, 1972, available at: <https://tgc.ifas.ufl.edu/vol22/vol22.pdf>; Jones et al., 2003; Mes et al., 2008). Additionally, the *aubergine* (*Abg*) locus from *Solanum lycopersicoides* accession LA2408 results in dark purple fruit (Rick et al., 1994). The *Abg* locus also maps to chromosome 10 and is possibly allelic to *Aft* (Rick et al., 1994). The synergistic interaction between a nonfunctional R3 MYB repressor *atv* and a functional R2R3 MYB activator at *Aft* elevates anthocyanin levels in tomato fruit and imparts purple color (Colanero, Tagliani, et al., 2020; Povero et al., 2011; Yan et al., 2020).

We conducted experiments to describe the chemical and genetic basis of purple pigmentation in fruit derived from LA1141. Our results suggest that the regulatory mechanism described for accessions from the green-fruited tomato clade also confers pigmentation in LA1141. However, the LA1141 alleles of *Aft* and *atv* are distinct from those previously characterized. Phylogenetic analysis of *Aft* sequence does not support a green-fruited origin of the LA1141 locus, which suggests that purple fruit pigmentation in this accession results from a loss of function that disrupts *atv* and a gain of function that restores *Aft*. Mutations in *Aft* may have arisen spontaneously in *S. galapagense* or be derived from an earlier, red-fruited ancestor, plausibly *S. pimpinellifolium*. These findings fail to support Rick's (1961) hypothesis on the origin of the Galápagos tomatoes.

2 | MATERIALS AND METHODS

2.1 | Plant materials and growing conditions

An inbred backcross (IBC) population was initiated in 2014 for the simultaneous introduction and characterization of purple pigmented fruit. The IBC population was derived from an initial hybridization of *S. galapagense* S.C. Darwin and Peralta (formerly *Lycopersicon cheesmaniae* f. *minor*) (Hook. f) C.H.Mull. accession LA1141 as the female parent and *Solanum lycopersicum* L. (formerly *Lycopersicon esculentum* Mill) OH8245 as the male parent. Accession LA1141 was acquired from the C.M. Rick Tomato Genetic Resources Center, University of California, Davis, CA, USA. The processing tomato variety OH8245 was described previously (Berry et al., 1991). A single LA1141 \times OH8245 F_1 plant was backcrossed as the female parent to OH8245. BC_1 individuals were then separately backcrossed again

with OH8245 as the pollen donor. BC₂ plants were then self-pollinated with single seed descent in alternating greenhouse and field production cycles to create a BC₂S₃ IBC population composed of 160 inbred backcross lines (IBLs). During these studies, the IBC population was further inbred to BC₂S₅. The BC₂S₃ IBLs SG18-124 (Figure 1c) and SG18-200 (Figure 1b) were selected based on purple pigmentation in the fruit to generate populations for validation of quantitative trait loci (QTLs). The IBLs SG18-124 and SG18-200 were again crossed to OH8245, and the self-pollination of the resulting F₁ plants created populations with F₂ segregation for specific LA1141 chromosomal regions.

Seedling care for greenhouse and field trials followed the same protocol. The 160 BC₂S₃ IBLs and the SG18-124 and SG18-200 derived F₂ progenies were sown in 288 cell trays with a cell volume of 13 ml. Greenhouse temperatures were set to 27°C during the day and 25°C at night with a 16-h photoperiod. Photosynthetically active radiation (PAR) was supplied by natural sunlight, 1,000-W metal-halide lamps (Multi-Vapor® GE Lighting, East Cleveland, OH, USA), and 1,000-W high-pressure sodium lamps (Ultra Sun® Sunlight Supply, Vancouver, WA) with a target radiation threshold of 250 W m⁻² natural light or approximately 113 μmol m⁻² s⁻¹ PAR. Fertilization was applied using a 20-20-20 fertilizer (20% N, 20% P₂O₅, and 20% K₂O) (Jack's professional All-Purpose Fertilizer, JR Peters INC., Allentown,

PA, USA) delivered at a concentration of 1,000 mg L⁻¹ (200 mg L⁻¹ N) twice per week. Plants were irrigated once or twice per day as needed.

IBC and F₂ progenies were evaluated in field trials as single plants. The BC₂S₃ IBC population was evaluated with 60-cm spacing and 164 plants, including controls. Progenies derived from SG18-124 and SG18-200 were transplanted for greenhouse and field evaluations of pigmentation in the fruit. Plants with three to five expanded leaves were transferred to 3.78-L containers (Hummert, EARTH City, MO) and spaced 30 cm apart on the greenhouse bench. There were 36 F₂ plants evaluated in the greenhouse. The remaining SG18-124 and SG18-200 derived F₂ progenies were evaluated in field trials with 60 cm spacing with a total of 145 plants harvested.

2.2 | Tomato fruit color measurements

Three mature green (MG) fruits were randomly selected from each plot and measured at the midpoint between the fruit's shoulder and the blossom end. Fruit Color was measured with a colorimeter (chromameter CR-300; Minolta Camera Co., Ltd., Ramsey, NJ, USA). Values of the red, green, yellow, and blue components of fruit were obtained using the "L*a*b*" CIELAB color space (Commission Internationale de



FIGURE 1 Heritable fruit pigmentation from *Solanum galapagense* accession LA1141. We determined a role for several candidate genes underlying the *Anthocyanin fruit* (*Aft*), *atrorivulacea* (*atv*), and *uniform ripening* (*u*) loci derived from LA1141. Homozygous LA1141 *Aft* is designated as *Aft*, homozygous LA1141 *atv* is designated as *atv*, and homozygous LA1141 *u* is designated as *u*⁺. Allele notation follows the rules for nomenclature in tomato genetics (Barton et al., 1955). (a) LA1141 mature green fruit (*Aft*; *atv*; *u*⁺). (b) Inbred backcross line (IBL) SG18-200 (*Aft*; *atv*⁺; *u*). (c) IBL SG18-124 (*Aft*; *atv*; *u*⁺). (d) IBL SG18-251 (*Aft*⁺; *atv*⁺; *u*⁺)

l'Eclairage, 1978). The L^* coordinate represented a measure of the darkness or lightness. Coordinates, a^* and b^* , are measured color along the axis of a color wheel with $+a^*$ in the red direction, and $-a^*$ in the green direction, $+b^*$ in the yellow direction, and $-b^*$ in the blue direction (Kabelka et al., 2004). Chroma and hue were derived from measurements of a^* and b^* . Chroma was calculated as $(a^{*2} + b^{*2})^{1/2}$ and used to measure how bright or dull the color was. Hue was calculated using the equation $(180/\pi) [\cos^{-1} (a^*/\text{chroma})]$ for positive values of a^* . For negative values of a^* , we calculated hue using the equation $360-(180/\pi) [\cos^{-1} (a^*/\text{chroma})]$ (Darrigues et al., 2008; Kabelka et al., 2004). The average hue, chroma, and L^* values were used as the response variable for our genetic studies.

2.3 | Anthocyanidin extraction, analysis, and identification

Tomato fruit samples at different stages of maturity from SG18-124 \times OH8245 derived F_2 plants were blended, and 3.5 g of juice was extracted with 4 ml 1% HCl in MeOH. The extracts were dried under nitrogen gas. Anthocyanins were separated using an C18 column (HSS T3, 2.1 \times 100 mm, 1.8 μ m, Agilent Technologies) and a gradient of water (A) and acetonitrile (B), both with 5% formic acid. The gradient was as follows: isocratic with 0% B from 0–2 min, linear gradient to 30% B from 2–8 min, linear gradient to 100% from 8–12 min, hold at 100% B for 1 min, and return to initial conditions. Samples were run on an Agilent 1290 ultra high performance liquid chromatography (UHPLC) with photodiode array (PDA) detection, coupled to a high resolution 6545 quadrupole time-of-flight mass spectrometer (QTOF-MS) (Agilent, Santa Clara, CA, USA). The MS was run using electrospray ionization and operated in both positive and negative modes using reference masses for accurate mass determination.

2.4 | DNA isolation and genotyping

Genomic DNA was isolated from fresh, young leaf tissue from the 160 BC_2S_3 progenies, 96 of each F_2 population, and parental lines using a modified CTAB method consistent with previous studies (Sim et al., 2015). Single-nucleotide polymorphisms (SNPs) between OH8245 and LA1141 were identified using a 384-marker panel (Bernal et al., 2020). Genotyping of the BC_2S_3 progenies was performed using the PlexSeq™ platform as a service (Agriplex Genomics, Cleveland, OH, USA) to detect specific SNPs through a pooled amplicon sequencing strategy.

2.5 | Marker development for candidate genes

Selected SNP markers and candidate genes were converted to polymerase chain reaction (PCR) based insertion/deletion (INDEL) markers for visualization on agarose gels. These markers, when appropriate,

were added to the linkage maps described below. The genetic notation used in this study followed the rules for nomenclature as recommended (Barton et al., 1955). A summary containing forward and reverse primers, genome location, and expected polymorphism for these markers is available at <https://doi.org/10.5281/zenodo.5650150> (Fenstemaker et al., 2021a). Candidate genes included MYB transcription factor sequences corresponding to *atv* (MF197509, NC_015447.3; Cao et al., 2017), *Aft* (EF433416; EF433417; MN433086; MN433087; FJ705319; NC_015447.3; Cao et al., 2017; Mes et al., 2008; Sapir et al., 2008), *GOLDEN2-LIKE (GLK2)* transcription factor sequences corresponding to *u* (JX163897; JQ316459; NC_015447.3; Powell et al., 2012), and *Lycopene β -cyclase (Cyc-B)* sequences corresponding to *B* (KP233161; Orchard, 2014). These sequences were targeted as candidate genes based on initial QTL mapping and because of their previously known association with tomato fruit color. The INDEL and cleaved amplified polymorphism sequences (CAPS) markers were developed using a sequence comparison approach between *S. lycopersicum* variety Heinz 1706, *S. cheesmaniae* (L. Riley) Fosberg, 1987 in [Fosberg FR (1987b)] formerly *Lycopersicon cheesmaniae* L. Riley, 1925 in [Riley LAM (1925c)] accession LA0483, *S. galapagense* accession LA1401 and LA1044, and *S. lycopersicum* variety Indigo Rose. Primers were designed using Primer3 (v..4.0) (Untergasser et al., 2012). These primers were used to genotype LA1141, OH8245, the BC_2S_3 IBC population, and the subsequent F_2 progenies derived from IBL selections SG18-124 and SG18-200.

PCR was carried out with an initial incubation at 94°C for 3 min, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 60 s. After completing the cycles, a final elongation step at 72°C was carried out for 7 min. The PCR products for markers detected as CAPS were digested with a restriction enzyme (Fenstemaker et al., 2021a) for 2 h at 37°C. The PCR products were separated on a 2.5% agarose gel.

2.6 | Linkage map construction

A genetic linkage map was constructed based on the IBC population. The R/qtl package version 1.47-9 was used in the R statistical software environment version 4.03 (Broman et al., 2003; R Core Team, 2020). We used the “read.cross” function from BC_sF_t tools to read in data, with $s = 2$ and $t = 3$ (Shannon et al., 2013). Of the 384 SNPs in the marker panel, 157 were polymorphic in the IBC population, and no markers were removed. A summary of the 157 polymorphic SNPs is available at <https://doi.org/10.5281/zenodo.5650152> (Fenstemaker et al., 2021b). The genetic map was constructed by using the “est.map” function in R/qtl. Markers were placed in the same linkage group if they had a logarithm of the odds (LOD) score greater than 1.8 and an estimated recombination fraction lower than .45. The Kosambi map function was used for map construction and to convert recombination frequency to genetic distance (Kosambi, 1944). The marker order in each linkage group was estimated with the functions “orderMarkers” and “ripple” in R/qtl. Changes in

chromosome length and loglikelihood were investigated, dropping one marker at a time with the function “dropmarker” in R/qtl. Marker order was compared with the physical position in the Tomato Genome version SL4.0 (Hosmani et al., 2019) using both linear (adjusted correlation coefficient R^2) and rank regression ($\rho(p)$) to assess linkage map quality.

2.7 | QTL analysis in BC₂S₃ IBLs

Composite interval mapping (CIM) was used for QTL detection (Zeng, 1994) using the “cim” function in the R/qtl package (Broman et al., 2003). Analysis was performed using a 2-cM step, one marker selected as a cofactor, and a 40-cM window with cofactor and window selected due to limited recombination and expected skewed segregation in the BC₂S₃ population. Haley Knott regression (Haley & Knott, 1992) (hk) was chosen as the solution-generating algorithm. Significance thresholds were generated by using the permutation test ($\alpha = .05$, $n = 1,000$; Churchill & Doerge, 1994). The resampled LOD cutoffs used were LOD = 6.8 for hue, LOD = 4.5 for chroma, and LOD = 3.65 for L*. Genetic effects were evaluated as differences between phenotype averages expressed as regression coefficients using the “fitqtl” function with the argument “get.ests = TRUE” and “dropone = FALSE” in R/qtl. Additionally, the percent phenotypic variance explained was estimated by the “fitqtl” function with the argument “dropone = TRUE” in R/qtl.

2.8 | QTL validation

The IBLs SG18-124 and SG18-200 were chosen because of pigmentation in their fruit (Figure 2b,c). Seedlings were also grown as previously described. Segregating SG18-124 × OH8245 F₂ and SG18-200 × OH8256 F₂ progenies were transplanted to the field and greenhouse, and fruit pigmentation was measured using the Minolta CR300 colorimeter as described above. Marker data were scored on 91 progenies derived from the SG18-124 × OH8245 F₂ population and 90 from the SG18-200 × OH8245 F₂ population. Genetic effects and allele substitutions were evaluated using linear model analysis of variance (ANOVA) as implemented by the “lm” function in the R core package (R Core Team, 2020). The linear model was $Y = \mu x + M + E$, where Y was the color trait value, μx was equal to the population mean, M was the effect of each marker allele, and E was the associated error, equivalent to genotype (marker). We compared the marker-locus genotypic classes of LA1141 and OH8245. The markers Ant1_1, An2-like_exon2_intron2, atv_ex4, u_gal_3, and BetaRSAcorresponding to candidate genes of interest were tested (Fenstemaker et al., 2021a). Marker evaluations were conducted in both F₂ populations independently. We used F -tests as previously described to determine if hue, chroma, and L* were associated with significant differences in marker-locus genotypic classes and used the mean phenotypic differences to estimate the effect of allele substitutions.

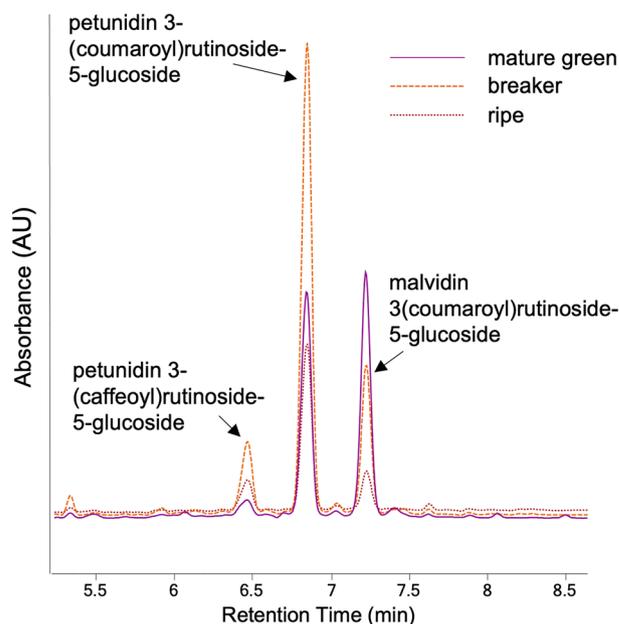


FIGURE 2 Predominant pigments in the fruit of LA1141 derived lines. The chromatograms show ultrahigh performance liquid chromatography separation and photodiode array (UHPLC-PDA) absorbance at 520 nm for fruit from mature green, breaker, and ripe fruit. The predominant peaks were identified as anthocyanins and are labeled above

Additionally, we tested possible allelic combinations of *Aft* and *atv* in the IBC and F₂ populations. We used the linear model $Y = \mu x + M_1 + M_2 + M_1:M_2 + E$, where Y was the color trait value, μx was equal to the population mean, M_1 and M_2 were effects of individual marker alleles, $M_1:M_2$ was the interaction between marker alleles, and E was the associated error, equivalent to genotype (marker) to test for significant markers interactions. We conducted a linear model ANOVA using the “lm” function in the R core package (R Core Team, 2020) to test the pairwise combination of chromosome 7 (*atv*) and chromosome 10 (*Aft*). When marker classes were significantly different ($p < .05$), we used Tukey’s Honest Significant Difference test, with the “HSD.test” function in the R package Agricolae (De Mendiburu, 2017) to compare means.

2.9 | Sequence alignment and phylogeny

A PCR amplification strategy was used for sequencing the *Aft* candidate sequences derived from LA1141 and OH8245. Amplified products were purified by precipitation using a 9:1 ethanol: sodium acetate (3 M), pH 5.2 mixture. Sequencing was performed at the Molecular and Cellular Imaging Center in Wooster, Ohio, using a dideoxy Sanger procedure on an ABI Prism Sequencer 3100x1 (Grand Island, NY, USA). The DNA sequence was generated in forward and reverse directions for each amplicon. All sequence data were quality-checked and trimmed before alignment. We used the UGENE v. 37 software package (Okonechnikov et al., 2012) to create contigs



from the forward and reverse sequence generated sequence from LA1141 and OH8245 corresponding to MYB-encoding genes underlying *atv* and *Aft*. The genomic sequences corresponding to MYB-encoding genes underlying *atv* were aligned using default settings with ClustalW (Larkin et al., 2007). Alignments using MUSCLE (version 3.8.31) (Edgar, 2004) were used for genes corresponding to *Aft* because of MUSCLE's ability to handle INDELS greater than 100 nucleotides.

2.10 | Bioinformatics pipeline

Genomes from 84 unique cultivated and wild tomato accessions published as part of the 100 Tomato Genome Sequencing Consortium (The 100 Tomato Genome Sequencing Consortium et al., 2014) and a reference-quality sequence for OH8245 generated as part of a collaboration between the Tomato Pan Genome Consortium and NRGene (Ness-Ziona, Israel; see <https://www.nrgene.com/solutions/consortia/tomato/>) were stored on the Ohio Supercomputer Center (OSC) (Ohio Supercomputer Center, 1987) computing environment, and a nucleotide BLAST database was created using the function “makeblastdb” in the Basic Local Alignment Search Tool (BLAST) version/2020-04 (Altschul et al., 1990) program. Our workflow parsed through sequence matches, identified the highest quality match, and created a FASTA output file. Parsing was facilitated by “SearchIO”, “Seq”, and “SeqIO”, functions in BioPerl (Stajich et al., 2002) following implementation of the “blastn” function in the BLAST core package. The steps in the pipeline were automated using the Bash scripting language (Gnu, 2007) in a Unix shell on the OSC.

Passport data for all accessions and a summary of sequences, including genomic and coding sequence (CDS) length, are available at <https://doi.org/10.5281/zenodo.5650141> (Fenstermaker et al., 2021c). As described above, the genomic sequences and CDS were retrieved from regions corresponding to the tomato *Aft* locus from LA1141, OH8245, Heinz 1706. CDS corresponding to MYB-encoding genes were determined by comparing genomic sequences to the Heinz reference Tomato Genome CDS (ITAG release 4.0) available from the Solanaceae Genome Network (SGN) (available at https://solgenomics.net/organism/Solanum_lycopersicum/genome). Additional CDS sequences were retrieved from the National Center for Biotechnology Information (NCBI) from the following Genbank records: Indigo Rose (MN433087; Yan et al., 2020), *S. lycopersicum* accession LA1996 (MN242011.1, EF433417.1; Colanero, Tagliani, et al., 2020; Sapir et al., 2008), *S. chilense* (Dunal) Reiche (formerly *Lycopersicon chilense* Dunal), and accession LA1930 (MN242012.1; Colanero, Tagliani, et al., 2020).

Orthologous sequences corresponding to tomato *Aft* were retrieved from *S. lycopersicoides* Dunal accession LA2951 genome (v0.6) made available by The *S. lycopersicoides* Genome Consortium (Powell et al., 2020). *Solanum tuberosum* L. Group Phureja clone DM1-3 genome (PGSC DM v4.03 Pseudomolecules) was made available by the Potato Genome Sequence Consortium (Potato Genome

Sequencing Consortium, 2011) and *Capsicum annum* L., 1753 in [Linnaeus C (1753c)] cv. CM334 genome (*C. annum* cv CM334 genome chromosome release 1.55, Hulse-Kemp et al., 2018). These corresponding sequences were retrieved using the BLAST tool (<https://solgenomics.net/tools/blast/>). Syntenic chromosomal regions were compared using known marker positions of tomato, potato, and pepper and the comparative map viewer (available at <https://solgenomics.net/cview>) on chromosome 10. Orthologous sequences corresponding to *Salvia miltiorrhiza* Bunge, 1833 (KF059503.1; Li & Lu, 2014), *Arabidopsis thaliana* (L.) Heynh, 1842 (Arabidopsis), (NM_105308.2, NM_105310.4; Berardini et al., 2015; Cominelli et al., 2008; Teng et al., 2005) were chosen based on homology and gene annotation that described positive R2R3 MYB regulation of anthocyanins.

The CDS corresponding to the *Aft* genes were retrieved from the CDS reference genomes available from the Sol Genomics Network (SGN): Tomato Genome CDS (ITAG release 4.0), Potato PGSC DM v3.4 CDS sequences, *C. annum* cv CM334 Genome CDS (release 1.55) or from NCBI available online (<https://www.ncbi.nlm.nih.gov>). To retrieve CDS from NCBI, we accessed the “RefSeq” section of the Genbank records mentioned above. The CDS were extracted from the “features” section of the Genbank records and exported as FASTA files. The UGENE v. 37 software package (Okonechnikov et al., 2012) was used for sequence trimming prior to alignment using MUSCLE (version 3.8.31) (Edgar, 2004) in the OSC Unix command line.

Phylogenetic trees were constructed using the phangorn R package (Schliep, 2011) for the R2R3 MYB-encoding genes *Ant1* and *Ant2-like*. Genomic sequence files were combined from the MYB-encoding genes *Ant2-like* and *Ant1* to create a single *Aft* locus contig, aligned in MUSCLE and imported into phangorn. We constructed maximum likelihood (ML) trees based on the nucleotide alignment using the general time reversible model with the rate variation among sites described by a gamma distribution and the proportion of invariable sites (a.k.a. GTR + G + I model). The “optim.pml” function was used to optimize model parameters with a stochastic search algorithm to compute the likelihood of the phylogenetic trees (Nguyen et al., 2015). This methodology was used for both genomic and CDS sequences. Clade support was estimated with 1,000 bootstrap replicates using the function “bootstrap.pml.” Phylogenetic studies that used genomic sequence were rooted using *S. lycopersicoides* as an outgroup. Studies that used CDS were rooted using *Arabidopsis* as an outgroup. Trees were drawn and annotated using the Interactive Tree Of Life (ITOL) (Letunic & Bork, 2021; available at <https://itol.embl.de/>) software.

3 | RESULTS

3.1 | Accession LA1141 phenotypic description

We observed purple pigmentation in the MG fruit of LA1141 (Figure 1a), and we were able to recover purple fruit in BC₂S₃



progenies (Figure 1b,c). Purple pigmentation occurs in the skin and the pericarp tissues beneath the skin. Pigmentation was visible at all fruit maturity stages but most apparent at the MG stage. The interior of the fruit did not contain visible purple pigment. Fruit hue values in the IBC progenies ranged from 231.27° to 283.35° with a mean of 240.75° for the population. Hue values greater than 250° were designated as “deep purple” (Figure 1c). Progenies with hue values that ranged between 245° and 250° also had visible spotting or speckling of purple pigment. We designated progenies in this range of hue as “light purple” (Figure 1b). All hue values measured on fruit below 245° were green (Figure 1d). IBLs with purple pigmentation in the fruit had hue values greater than 245°, L^* values ranging from 44.40 to 64.29 units, and chroma values ranging from 7.91 to 35.22 units. We expected the LA1141 × OH8245 BC₂S₃ IBC population to be roughly 87.5% recurrent parent (OH8245), with the remaining 12.5% representing random introgressions from the LA1141 donor parent. The observed segregation of individuals with deep purple phenotypes approximated the expected genotypic percentages for two unlinked loci ($\chi^2 = .339$, $p = .843$). Plants with deep purple fruit (Figure 1c) also display darker green leaves with purple veins and purple pigmentation in the stems. In contrast, a single introgression could explain plants with light purple phenotypes (Figure 1b) ($\chi^2 = 2.053$, $p = .358$).

3.2 | Chemical analysis of LA1141 × OH8245 BC₂S₃ derived purple tomatoes

We used UHPLC-PDA-QTOF-MS to identify compounds that absorb light at 520 nm, which is characteristic of anthocyanins. The peaks in the chromatogram (Figure 2) indicate that the predominant anthocyanidins were petunidin and malvidin based on accurate masses previously published (Mathews et al., 2003; Ooe et al., 2016). Petunidin-3-(caffeoyl)rutinoside-5-glucoside (C₄₃H₄₉O₂₄⁺) was identified at a retention time of 6.46 min and had an observed mass [M^+] of 949.2623 (1-ppm mass error), petunidin-(coumaroyl)rutinoside 5-glucoside (C₄₃H₄₉O₂₃⁺) at a retention time of 6.85 min with a mass [M^+] of 933.2686 (2-ppm mass error), and malvidin-3 (coumaroyl)rutinoside-5-glucoside (C₄₄H₅₁O₂₃⁺) at a retention time of 7.22 min with a mass [M^+] of 947.2834 (1.3-ppm mass error) (Figure 2). These anthocyanins are present in all fruit maturity stages. We see a change in the predominant anthocyanins from the MG to breaker fruit stage (Figure 2). The anthocyanins petunidin-(coumaroyl)rutinoside 5-glucoside and anthocyanin malvidin 3(coumaroyl)rutinoside-5-glucoside are of similar intensity at MG (Figure 2). The anthocyanin petunidin-(coumaroyl)rutinoside 5-glucoside was the predominant anthocyanin at breaker and ripe stages (Figure 2). Additionally, we observed changes in individual anthocyanin abundance over ripening (Figure 2). The anthocyanin malvidin 3(coumaroyl)rutinoside-5-glucoside was most abundant at the MG stage (Figure 2). The anthocyanins petunidin-(coumaroyl)rutinoside 5-glucoside and petunidin-3-(caffeoyl)rutinoside-5-glucoside are most abundant at the breaker stage (Figure 2).

3.3 | LA1141 × OH8245 BC₂S₃ linkage map quality

Linkage maps were constructed based on marker data from the BC₂S₃ IBC population and defined 13 linkage groups corresponding to each tomato chromosome. We split chromosome 1 into two linkage groups (1a and 1b) because of a recombination fraction greater than .45 between adjacent markers. The total number of markers in each linkage group ranged between 2 and 27, and linkage group 4 had the most markers at 27 (Table 1). The centimorgan (cM) length per linkage group ranged between 25.8 and 121.6 cM (Table 1). The average cM distance between markers was 8.1, and the largest distance in cM between markers was 41.8 (Table 1). SNP marker physical position using the tomato SL4.0 physical map (Hosmani et al., 2019) agreed with the estimated genetic position using both linear correlation and rank correlation (Table 1). As previously demonstrated, correlations are not perfectly linear due to reduced recombination in the pericentromeric region (Sim et al., 2012). Linear correlations ranged from .28-.99, while rank correlations ranged from .96 to 1 (Table 1).

3.4 | QTL analysis of tomato color in the LA1141 × OH8245 BC₂S₃ population

We identified three putative QTLs that explained between 8.24% and 35.53% of the total phenotypic variation for hue, chroma and L^* (Figure 3 and Table 2). All QTLs that contribute to the purple color are derived from the LA1141 donor parent, with purple pigmentation defined by an increase in hue and a decrease in both chroma and L^* (Table 2). A region on the distal arm of chromosome 10 explained between 22.63% and 24.04% of the total phenotypic variation and increased hue between 6.74° and 7.5° (Figure 3 and Table 2). Two QTLs, one on the proximal arm and one on the distal arm of chromosome 10, were associated with chroma and explained between 18.02% and 28.53% (proximal arm) and between 15.95% and 23.08% (distal arm) of the total phenotypic variance (Figure 3 and Table 2). These QTLs decreased chroma between 3.96 and 17.53 units (proximal arm) and 6.26 and 8.42 units (distal arm) (Figure 3 and Table 2). Two QTLs were associated with L^* , one on chromosome 6 and one on chromosome 10 (proximal) (Figure 3 and Table 2). These QTLs explained between 8.24% and 35.53% of the total phenotypic variation (Table 2). The QTL on chromosome 10 explained between 22.03% and 35.53% of the phenotypic variation and reduced L^* by 9.23 units (Table 2). The QTL on chromosome 6 explained between 8.24% and 10.13% of the phenotypic variation and reduced L^* between 4.53 and 5.05 units (Table 2).

3.5 | Candidate genes

Candidate genes were selected because of their previously characterized role in regulating tomato fruit pigmentation and because of their locations within the physical intervals of our QTLs (Table 2). The

TABLE 1 Genetic map quality for the inbred backcross population (OH8245 × LA1141 BC₂S₃)

Genetic map					Genetic map versus physical map (SI4.0) correlation		
Linkage group	Number of markers	Chromosome length (cM)	Average distance between markers (cM)	Largest distance between markers (cM)	<i>p</i> value ^a	<i>R</i> ^{2b}	Rho (ρ) ^c
1a	8	42.1	6	33	.0001	.9024	1.000
1b	6	28.4	5.7	20.4	.0003	.9909	1.000
2	9	74.2	9.3	18.2	.0000	.9789	1.000
3	14	121.6	9.4	38.9	.0000	.8900	.986
4	27	96.2	3.7	32.9	.0000	.6554	.965
5	8	63.7	9.1	32.6	.0271	.5151	1.000
6	12	57.1	5.2	15.1	.0050	.6017	1.000
7	9	64.3	8	27.5	.0131	.6416	1.000
8	6	35.5	7.1	17.5	.0168	.7445	1.000
9	17	113.7	7.1	34.3	.0000	.8510	1.000
10	22	121.4	5.8	41.9	.0000	.8229	1.000
11	2	25.8	25.8	25.8	NA	NA	NA
12	22	78.4	3.7	35.6	.0058	.2888	1.000

^aThe *p* value was derived from the regression equation (Genetic position ~ Physical position) based on markers physical position according to the *Solanum lycopersicum* (tomato) genome version 4.0 (Hosmani et al., 2019) and genetic distances calculated in the OH8245 × LA1141 BC₂S₃ genetic map.

^bAdjusted correlation coefficient (*R*²) was derived from the regression equation (Genetic position ~ Physical position) based on markers physical position according to the *Solanum lycopersicum* (tomato) genome version 4.0 and genetic distances calculated in the OH8245 × LA1141 BC₂S₃ population.

^cRank correlation or rho (ρ) is the rank order correlation derived from the regression equation (Genetic position ~ Physical position) based on markers physical position according to the *Solanum lycopersicum* (tomato) genome version 4.0 and genetic distances calculated in the OH8245 × LA1141 BC₂S₃ population.

R2R3 MYB-encoding candidate genes *Ant1* (Sapir et al., 2008) and *An2-like* (Qiu et al., 2019; Yan et al., 2020) are located within the QTL interval on the distal arm of chromosome 10 (Table 2). The MYB-encoding genes *Ant1* and *An2-like* are members of the multigene MYB family associated with the *Aft* locus (Yan et al., 2020). The transcription factor *Golden2-like 2 (u)* (Powell et al., 2012) maps to the proximal arm of chromosome 10 within the QTL regions identified for L* and chroma (Table 2). Additionally, we chose the fruit-specific *Cyc-B* gene (*B*) to investigate the QTL on chromosome 6 because accession LA1141 has the characteristic ripe orange fruit associated with the *Beta* locus (Orchard et al., 2021). We chose The R3 MYB repressor *atv (atroviolacea)* on chromosome 7 (Cao et al., 2017; Colanero et al., 2018) because of its previously described synergistic interaction with *Aft* which results in a purple phenotype similar to what we observe in our deep purple accession (Figure 1c). We added these markers to the linkage maps described above and used them in our QTL analysis.

3.6 | QTL mapping using candidate genes in the IBC population

Genetic evidence supports a role for *Aft*, *atv*, and *u* conferring purple pigmentation in the fruit of LA1141. The markers corresponding to the MYB-encoding genes *Ant1* and *An2-like* are physically near one another (Hosmani et al., 2019) (Table 2) and have less than 1%

recombination frequency. For measurements of hue, the markers *Ant1* and *An2-like* (LOD = 9.4) exceeded our resampled LOD cutoff (LOD = 6.8), explained 24.04% of the phenotypic variation, and increased hue by 7.05° (Table 2). The markers *B* (LOD = 2.74), *atv* (LOD = 2.6), and *u* (LOD = 2.65) did not exceed our resampled LOD cutoffs for hue (Table 2).

The markers *Ant1* and *An2-like* (LOD = 14.24) exceeded our resampled LOD cutoffs for chroma (LOD = 4.5), explained 23.08% of the total phenotypic variance, and reduced chroma by 8.24 units (Table 2). The marker *u* (LOD = 12) also exceeded our resampled LOD cutoff, accounted for 28.53% of the total phenotypic variation, and reduced chroma by 17.53 units (Table 2). The markers *B* (LOD = 2.74) and *atv* (LOD = 2.61) did not exceed our resampled LOD cutoff for chroma (Table 2).

Regions on chromosome 6 and the proximal arm of chromosome 10 were targeted for measurements of L*. The marker *u* (LOD = 15.25) exceeded our resampled LOD cutoff (LOD = 3.65), explained 35.53% of the total phenotypic variance, and reduced L* by 9.32 units (Table 2). The marker *B* (LOD = 1.26) did not exceed our resampled LOD cutoff for L*, and our QTL analysis failed to support a role for *B* as a candidate gene on chromosome 6. Additionally, the *Aft* markers (LOD = 1.13) and *atv* (LOD = 1.48) did not appear to be associated with L* (Table 2).

Although the marker *atv* did not exceed our LOD significance thresholds for hue, chroma, or L* (Table 2), segregation ratios for the

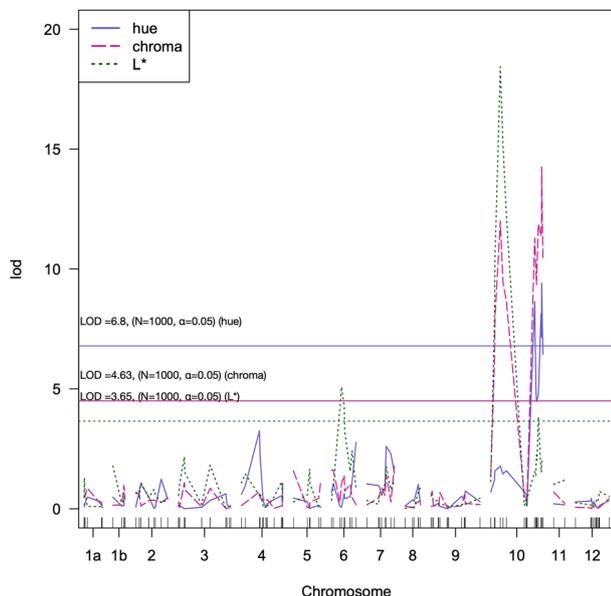


FIGURE 3 Composite interval mapping (CIM) of fruit color measured as hue (violet), chroma (pink), and L^* (green) in the LA1141 \times OH8245 BC₂S₃ inbred backcross population. The y-axis is the logarithm of the odds (LOD). The horizontal lines are the resampled LOD significance cutoff ($\alpha = .05$, $N = 1,000$ permutations) for hue (violet), chroma (pink), and L^* (green). The x-axis represents the 12 chromosomes in tomato, and chromosome distance in centimorgans was calculated using the Kosambi function to correct for multiple crossovers

deep purple phenotype in the BC₂S₃ progenies suggested two unlinked loci were responsible. The known regulatory mechanism involving MYB-encoding genes underlying *atv* and *Aft* led us to test for interaction effects for loci on chromosome 7 and chromosome 10. The interaction between LA1141 *Aft* and the LA1141 *atv* was significant ($p < 2.2e-16$) (Figure 4). We compared the hue values of BC₂S₃ IBL progenies with *Aft* and *atv* derived from LA1141 (*Aft*;*atv*) to those derived from OH8245 (*Aft*⁺;*atv*⁺) (Figure 4a). The BC₂S₃ IBLs with both the LA1141 *Aft* and *atv* had higher hue values than the *Aft*⁺;*atv*⁺ genotypes (Figure 4a). Additionally, we compared all possible allele combinations, including the genotypes *Aft*;*atv*⁺ and *Aft*⁺;*atv*. The *Aft*;*atv* LA1141 genotype had higher hue values than all other combinations (Figure 4a). However, the *Aft*;*atv*⁺ genotype had higher hue values than the *Aft*⁺;*atv* and *Aft*⁺;*atv*⁺ genotypes (Figure 4a).

3.7 | Confirmation of QTLs in the F₂ validation populations

We evaluated F₂ populations originating from the selected IBL progenies SG18-124 (Figure 1c) and SG18-200 (Figure 1b) for measurements of hue, chroma, and L^* to validate the QTLs identified in the BC₂S₃ generation. The IBL SG18-124 had deep purple fruit (Figure 1c). The mean hue value of the SG18-124 derived F₂ population was 238.5° and ranged from 227.24° to 284.4°. The mean

chroma value was 24.8 units and ranged from 5.7 to 39 units. The mean L^* value was 46.3 units and ranged from 30.3 to 67.1 units. The IBL SG18-200 had light purple fruit (Figure 1b). The mean hue value in the SG18-200 derived F₂ population was 239.7° and ranged from 234.8 to 264°. The mean chroma value was 29.1 and ranged from 13.7 to 33.79 units. The mean L^* value was 52.2 and ranged from 42.3 to 60.2 units.

In the SG18-124 derived F₂ population, the markers Ant1 ($p = 1.513e-09$) and An2-like ($p = 2.118e-09$) were significantly associated with hue, explained 37% of the phenotypic variation, and increased hue by 19.45° and 22.05°, respectively (Table 3). The marker *atv* ($p = .022$) was also significantly associated with hue, explained 9% of the phenotypic variation, and increased hue by 11.99° (Table 3). The marker *u* ($p = .901$) was not significant for hue (Table 3). However, the marker *u* ($p = 2.071e-05$) was significantly associated with chroma, explained 23% of the phenotypic variation, and decreased chroma by 10.67 units (Table 3). The markers An2-like ($p = 4.051e-04$) and Ant1 were significantly associated with chroma, explained 14% and 27% of the total phenotypic variation, and decreased chroma by 10.80 and 12.23 units (Table 3). The marker *u* ($p = 3.181e-04$) was significantly associated with L^* , explained 17% of the phenotypic variation, and decreased L^* by 10.38 units (Table 3). The marker *B* was not significantly associated with hue ($p = .103$), chroma ($p = .842$), or L^* ($p = .715$) in the SG18-124 derived F₂ population (Table 3).

In the SG18-200 derived F₂ population, the markers *atv* and *u* were monomorphic (Table 3). Therefore, we did not test the estimated effects of allele substitutions and associations in this population. The markers Ant1 ($p = 5.702e-04$) and An2-like ($p = 3.691e-05$) were significantly associated with hue, explained 17% and 23% of the phenotypic variation, and increased hue by 4.36 and 5.03° (Table 3). Although marker *B* was not significantly associated with hue in the SG18-124 derived F₂ population described above, it was significantly associated with the SG18-200 population ($p = .001$) (Table 3). Marker *B* explained 14% of the total phenotypic variation and increased hue by 3.23° (Table 3). The markers Ant1 ($p = 1.475e-09$) and An2-like ($p = 7.13e-11$) were significantly associated with chroma, explained 48% and 52% of the phenotypic variation, and decreased chroma by 9.04 and 9.15 units (Table 3). The marker *B* ($p = .06$) was marginally non-significant for chroma (Table 3). The markers Ant1 ($p = 7.042e-05$) and An2-like ($p = 2.296e-05$) were significantly associated with L^* , explained 24% and 25% of the total phenotypic variation, and decreased L^* by 6.15 and 5.75 units (Table 3).

We validated the interaction between homozygous LA1141 *Aft* and homozygous LA1141 *atv* in the F₂ progeny (Figure 4b). Our results confirm that an interaction between *Aft* and *atv* is needed for the deep purple fruit phenotype (Figure 1c), and a single introgression of LA1141 *Aft* confers purple pigmentation, designated as a light purple phenotype (Figure 1b). Progeny homozygous for LA1141 *Aft* and *atv* genotypes had higher hue values compared with all other marker-locus classes (Figure 4b). Progeny homozygous for *Aft* and heterozygous for *atv* (*Aft*;*atv*/+) also had higher hue values than other marker-locus classes, except for the homozygous LA1141 *Aft*;*atv* genotype

TABLE 2 Markers associated with tomato fruit color

LA1141 × OH8245 BC ₂ S ₃							
Trait ^a	Marker	LOD ^b	Donor allele	Allele substitution effect ^c	Percent phenotypic variance explained ^d	Chromosome	Physical position ^e
Hue	B	2.74 (ns)	LA1141	4.88	7.63	6	43,562,526
	atv	2.61 (ns)	LA1141	3.94	7.30	7	61,003,154
	u	2.65 (ns)	LA1141	18.42	7.41	10	2,293,088
	solcap_snp_sl_100691	7.15	LA1141	6.74	22.63	10	64,276,927
	Ant1	9.4	LA1141	7.50	24.04	10	64,287,679
	An2-like	9.4	LA1141	7.50	24.04	10	64,317,522
	solcap_snp_sl_8787	6.45 (ns)	LA1141	3.13	17.04	10	64,366,981
Chroma	B	.173 (ns)	LA1141	-1.57	.73	6	43,562,526
	atv	1.43 (ns)	LA1141	-3.97	3.42	7	61,003,154
	solcap_snp_sl_46386	8	LA1141	-3.96	18.02	10	1,610,355
	u	12	LA1141	-17.53	28.53	10	2,293,088
	solcap_snp_sl_34373	9.45	LA1141	-3.90	20.5	10	3,783,034
	solcap_snp_sl_100691	11.98	LA1141	-7.05	15.95	10	64,276,927
	Ant1	14.24	LA1141	-8.24	23.08	10	64,287,679
	An2-like	14.24	LA1141	-8.42	23.08	10	64,317,522
L*	solcap_snp_sl_14458	4.18	LA1141	-4.53	8.87	6	36,520,866
	solcap_snp_sl_1337	5.08	LA1141	-5.05	10.13	6	37,305,722
	solcap_snp_sl_12757	4.25	LA1141	-5.04	8.24	6	38,186,675
	B	1.26 (ns)	LA1141	-5.62	3.57	6	43,562,526
	atv	1.48 (ns)	LA1141	-5.08	4.17	7	6,112,941
	solcap_snp_sl_46386	8.64	LA1141	-5.14	22.03	10	1,610,355
	u	15.25	LA1141	-9.32	35.53	10	2,293,088
	solcap_snp_sl_34373	12.43	LA1141	-5.27	30.08	10	3,783,034
	Ant1	1.13 (ns)	LA1141	-4.21	3.2	10	64,287,679
	An2-like	1.13 (ns)	LA1141	-4.21	3.2	10	64,317,522

^aColor was measured as hue, chroma, and L* in the OH8245 × LA1141 BC₂S₃ population.

^bLogarithm of the odd (LOD) significance cutoffs were determined by a resampling of the data ($\alpha = .05$, $N = 1,000$ permutations). LOD cutoffs for traits were hue (LOD = 6.8), chroma (LOD = 4.5) and L* (LOD = 3.65).

^cGenetic effects were evaluated as differences between phenotype averages expressed as regression coefficients.

^dPercent variance explained was estimated by $1 - 10^{-2 \text{ LOD}/n}$, where n is the sample size and LOD is the LOD score.

^ePhysical position in base pairs corresponds to the Tomato Genome version SL4.0 (Hosmani et al., 2019).

(Figure 4b). These results suggested that the *Aft;atv/+* genotype can accumulate enough anthocyanins to measure differences in hue. The *Aft;atv⁺* and *Aft/+;atv* genotypes had higher hue than the *Aft⁺;atv⁺*, *Aft⁺;atv*, *Aft⁺;atv/+*, and *Aft/+;atv/+* genotypes (Figure 4b). They had lower hue values than the *Aft;atv* and *Aft;atv/+* genotypes (Figure 4b).

3.8 | Sequence analysis of candidate genes

Sequence reads for *atv* covered 1,353 bps from the first putative start codon. Sequence analysis suggested that the LA1141 *atv* may be non-functional compared with the cultivated accessions OH8245 and

Heinz 1706. There is an 18-bp INDEL in the first intron of the LA1141 *atv* sequence and two G to A SNPs in the coding region of the second exon (Figure 5). These G to A SNPs in the coding region may result in the loss of a functional R3/bHLH binding domain (Figure 5). The LA1141 *atv* sequence is distinct from the allele previously described in Indigo Rose derived from *S. cheesmaniae* accession LA0434 (Figure 5).

Contigs assembled from sequencing reads of the LA1141 and OH8245 of R2R3 MYB-encoding gene *An2-like* covered approximately 1,363 out of 1,356 base pairs (bps) from the putative start codon. FASTA files corresponding to sequences for tomato accessions used in this study are available at <https://doi.org/10.5281/zenodo.5649546> for *An2-like* and <https://doi.org/10.5281/zenodo.5649996>

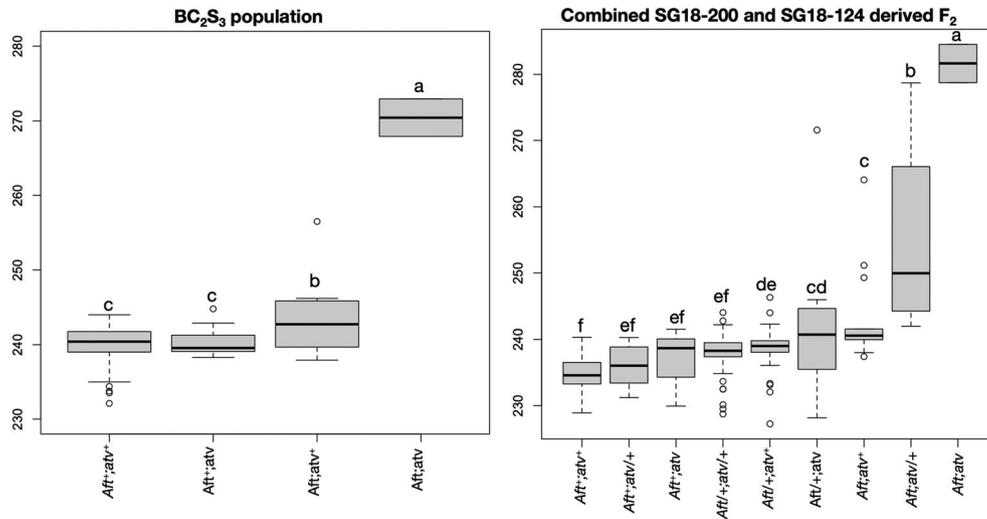


FIGURE 4 Box plots represent interactions between the *Anthocyanin fruit* (*Aft*) and *atroviolacium* (*atv*) loci. The x-axis is marker-locus genotypic class, and the y-axis is degrees of hue. (a) The interaction is shown in the BC₂S₃ population and (b) the combined F₂ validation populations. Homozygous LA1141 alleles are notated as *Aft*⁺; *atv*⁻, heterozygous alleles as *Aft*^{+/+}; *atv*^{+/+}, and homozygous OH8245 alleles as *Aft*⁺; *atv*⁺. Allele notation follows the rules for nomenclature in tomato genetics (Barton et al., 1955). Different letters indicate statistically significant differences among groups (Tukey's honest significant difference, $p < .05$)

for *Ant1* (Fenstemaker et al., 2021d, 2021e). There were several unique SNPs and INDELS in the LA1141 *An2-like* sequence, but none of them were in the conserved R2R3 domains (Fenstemaker et al., 2021d). However, LA1141 possesses the previously characterized G to A SNP in the 5' splice site of the second intron (Sun et al., 2020; Yan et al., 2020; Fenstemaker et al., 2021d). Sequencing reads covered 1,182 of 1,012 of LA1141 and 1,012 out of 1,012 bps of OH8245 from the first putative start codon in the R2R3 MYB-encoding gene *Ant1*. In the third exon of the LA1141 *Ant1* sequence, there is 170-bp insertion/deletion (INDEL), which contained MYB core type 1 and type 2 cis-regulatory elements, an AC-rich sequence type 2 cis-regulatory element (Fenstemaker et al., 2021e). Sequence analysis suggests that LA1141 may have a functional R2R3 MYB activator at *Aft* and the R3 MYB repressor corresponding to *atv* is likely nonfunctional. Additional characterization of transcripts, proteins, and protein interactions is needed for *An2-like*, *Ant*, and *atv* to confirm functional changes.

3.9 | Phylogenetic analysis of *Aft*

We combined the genomic sequences from LA1141, OH8245, accession LA1996, *S. lycopersicoides*, and 84 resequenced accessions representative of the Esculentum, Arcanum, Peruvianum, and Hirsutum groups (Pease et al., 2016). The red-fruited, Esculentum, clade is represented by commercial, landrace, and heirloom tomato varieties, as well as *S. lycopersicum cerasiforme*. This clade also includes *S. pimpinellifolium* and the orange-fruited Galápagos species *S. cheesmaniae* and *S. galapagense*. The green-fruited clade is divided into Arcanum, represented by *Solanum arcanum* and *Solanum chmielewskii*, Hirsutum represented by *S. habrochaites* and *Solanum pennellii*, and

Peruvianum represented by *Solanum huaylasense*, *Solanum neorickii*, *S. chilense*, and *S. peruvianum*. Genomic sequence corresponding to *Ant1* ranged from 1,023 to 1,993 bps, and genomic sequences corresponding to *An2-like* ranged from 2,292 to 2,547 bps (Fenstemaker et al., 2021c). The differences in contig length correspond to insertions and deletions within the sequences as contigs matched at the 5' and 3' ends.

A maximum likelihood (ML) model phylogeny was constructed for the R2R3 MYBs representing *Aft* from 89 sequences (Fenstemaker et al., 2021f). *S. lycopersicoides* was used as the outgroup to root the tree, and clustering resolved major tomato clades and groups (Figure 6). Accessions were grouped into their expected clades with 60.5% bootstrap support for the separation of red-fruited species and green-fruited species (Figure 6). The purple fruited accession LA1996 with an *Aft* introgression from *S. chilense* (LA0458) clusters with other members of the green-fruited clade and is closer to *S. habrochaites* accession LA1777 which has green fruit with distinct purple stripes/spots (Figure 6). Our purple accession *S. galapagense* LA1141 clusters with other members of endemic Galápagos tomatoes with 80.3% bootstrap support (Figure 6). The four *S. pimpinellifolium* accessions displayed the most variation within the Esculentum group, with LA1578 clustering closest to the Galápagos accessions with 38.3% support. Broadly, the Esculentum group clustering of *Aft* is supported at 75.9%. LA1141 does not cluster with members of the green-fruited clade based on sequence homology within the MYB-encoding genes underlying *Aft* (Figure 6).

Additionally, we clustered the CDS corresponding to the *Ant1* and *An2-like* MYB genes underlying *Aft* from LA1141, OH8245, and 84 resequenced accessions with outgroup sequences from *A. thaliana*, *Salvia miltorrhiza*, *S. tuberosum*, *S. lycopersicoides*, *C. annuum*, accession LA1996, *S. chilense* accession LA1930, and *S. lycopersicum* variety

TABLE 3 Candidate gene associations validated in subsequent F₂ populations

SG18–124 IBL derived F ₂ validation population							
Trait ^a	Marker	p value ^b	Parent allele	Allele substitution effect ^c	R ^{2d}	Chromosome	Physical position ^e
Hue	B	.103 (ns)	LA1141	−.63	.04	6	43,562,526
	atv	.022	LA1141	11.99	.09	7	61,003,154
	u	.901 (ns)	LA1141	.4026	−.02	10	2,293,088
	Ant1	<.000	LA1141	19.45	.37	10	64,287,679
	An2-like	<.000	LA1141	22.05	.37	10	64,366,981
Chroma	B	.842 (ns)	LA1141	−1.47	−.02	6	43,562,526
	atv	.06 (ns)	LA1141	−6.62	.06	7	61,003,154
	u	<.000	LA1141	−10.67	.23	10	2,293,088
	Ant1	<.000	LA1141	−12.23	.27	10	64,287,679
	An2-like	<.000	LA1141	−10.80	.14	10	64,366,981
L*	B	.715 (ns)	LA1141	−2.41	−.02	6	43,562,526
	atv	.452 (ns)	LA1141	−3.65	−.01	7	61,003,154
	u	<.000	LA1141	−10.38	.17	10	2,293,088
	Ant1	.136 (ns)	LA1141	−4.24	.02	10	64,287,679
	An2-like	.1876 (ns)	LA1141	−4.55	.01	10	64,366,981
SG18–200 IBL derived F ₂ validation population							
Hue	B	.001	LA1141	3.23	.14	6	43,562,526
	atv	NA	OH8245	NA	NA	7	61,003,154
	u	NA	OH8245	NA	NA	10	2,293,088
	Ant1	<.000	LA1141	4.36	.17	10	64,287,679
	An2-like	<.000	LA1141	5.03	.23	10	64,366,981
Chroma	B	.06 (ns)	LA1141	−2.53	.04	6	43,562,526
	atv	NA	OH8245	NA	NA	7	61,003,154
	u	NA	OH8245	NA	NA	10	2,293,088
	Ant1	<.000	LA1141	−9.04	.48	10	64,287,679
	An2-like	<.000	LA1141	−9.15	.52	10	64,366,981
L*	B	.186 (ns)	LA1141	−1.97	.01	6	43,562,526
	atv	NA	OH8245	NA	NA	7	61,003,154
	u	NA	OH8245	NA	NA	10	2,293,088
	Ant1	<.000	LA1141	−6.15	.24	10	64,287,679
	An2-like	<.000	LA1141	−5.75	.25	10	64,366,981

^aFruit color was measured as hue, chroma, and L* in the BC₂S₃ IBL derived F₂ populations.

^bANOVAs were conducted, and *F*-tests were used to determine if significant variation in hue, chroma, and L* was associated with differences in marker-locus genotypic classes. If NA, the marker was not segregating in the population and therefore could not be tested for differences in marker-locus genotypic classes.

^c*F*-tests to determine if hue, chroma, and L* were associated with significant differences in marker-locus genotypic classes and used the line mean differences to estimate the effect of allele substitutions.

^dAdjusted correlation coefficient (R²) calculated from linear model analysis of variance (ANOVA) is the percent of total phenotypic variance explained.

^ePhysical position in base pairs corresponds to the Tomato Genome version SL4.0 (Hosmani et al., 2019).

Indigo Rose (Fenstemaker et al., 2021g). The CDS corresponding to *A. thaliana* MYB genes that were determined to be homologous to *Solanum Aft* sequence were used as an outgroup to root the tree (supporting information Figure S1). The ML phylogeny separated *Ant1* and *An2-like* CDS with 98.4% bootstrap support (supporting information Figure S1). *A. thaliana* and *S. miltiorrhiza* clustered closer together

compared with accessions of *Solanum* for both *Ant1* and *An2-like* (supporting information Figure S1). For *Ant1* CDS, accession LA1141 clustered with members of the red-fruited clade with 81.3% bootstrap support. For *An2-like* CDS, accession LA1141 clustered with members of the red-fruited clade with 49.7% bootstrap support (supporting information Figure S1).

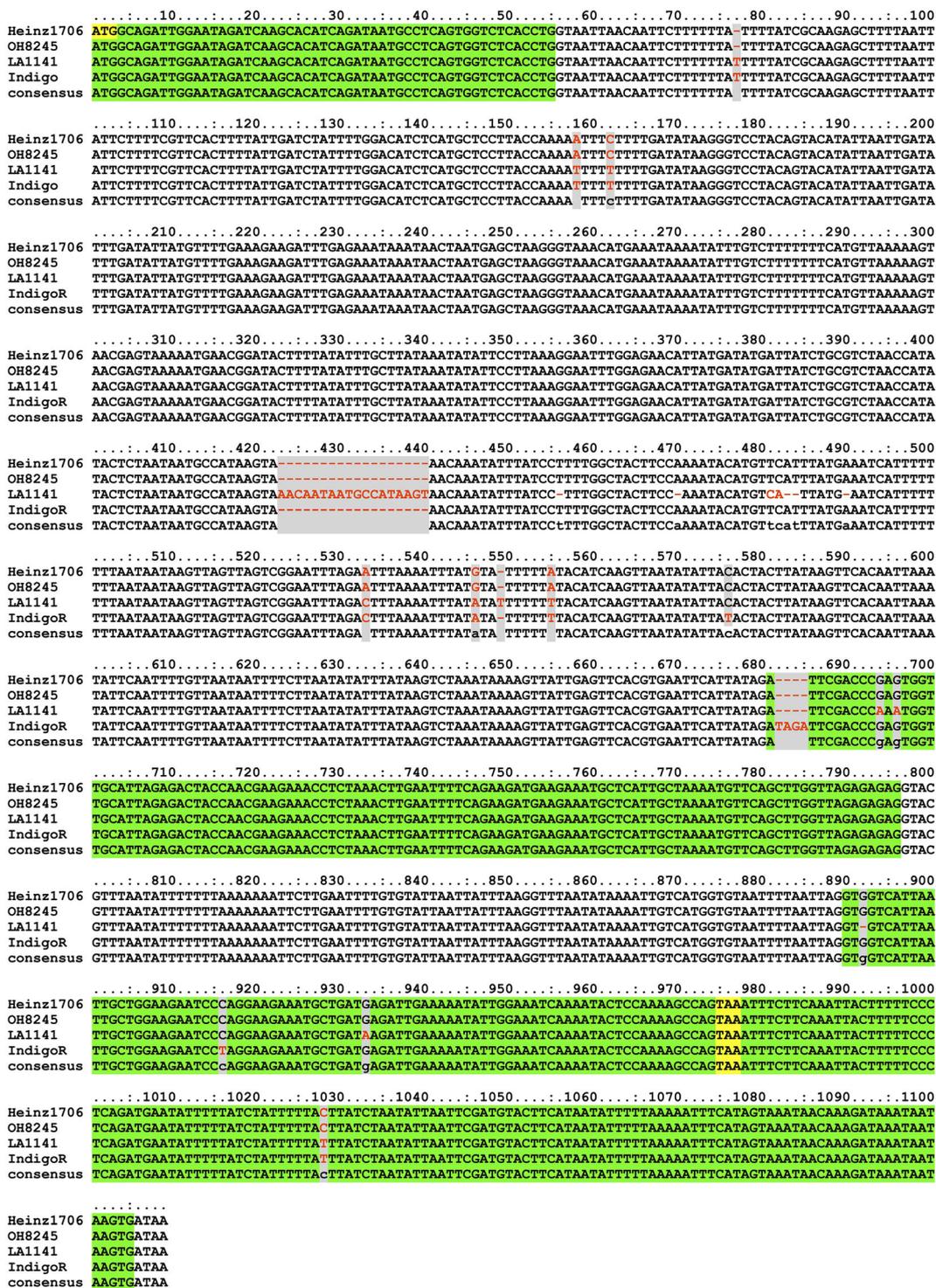


FIGURE 5 Sequence polymorphisms found in genomic regions of *Solyc07g052490 (SIMyBtv)* underlying the *atroviolacium (atv)* locus. Exons are highlighted green, and introns are unhighlighted black text. Exon and intron boundaries are consistent with previous studies (Cao et al., 2017; Colanero et al., 2018). Start and stop codons are highlighted yellow. Single-nucleotide polymorphisms (SNPs) and other sequence variants are highlighted gray with red text. The causal 4 base pair (bp) INDEL (*smybatv*) previously characterized in the tomato cultivar Indigo Rose (IndigoR) (Cao et al., 2017) is shown at bp position 681. Two G to A SNPs located at bp positions 694 and 696 are present in a region identified by CRISPR/CAS9 as important for the function of the conserved R3 domain (Yan et al., 2020). Sequences were aligned with ClustalW (Larkin et al., 2007) using default settings. The Heinz reference sequence (Heinz 1706), OH8245, LA1141, and IndigoR genomic *atv* sequences are represented

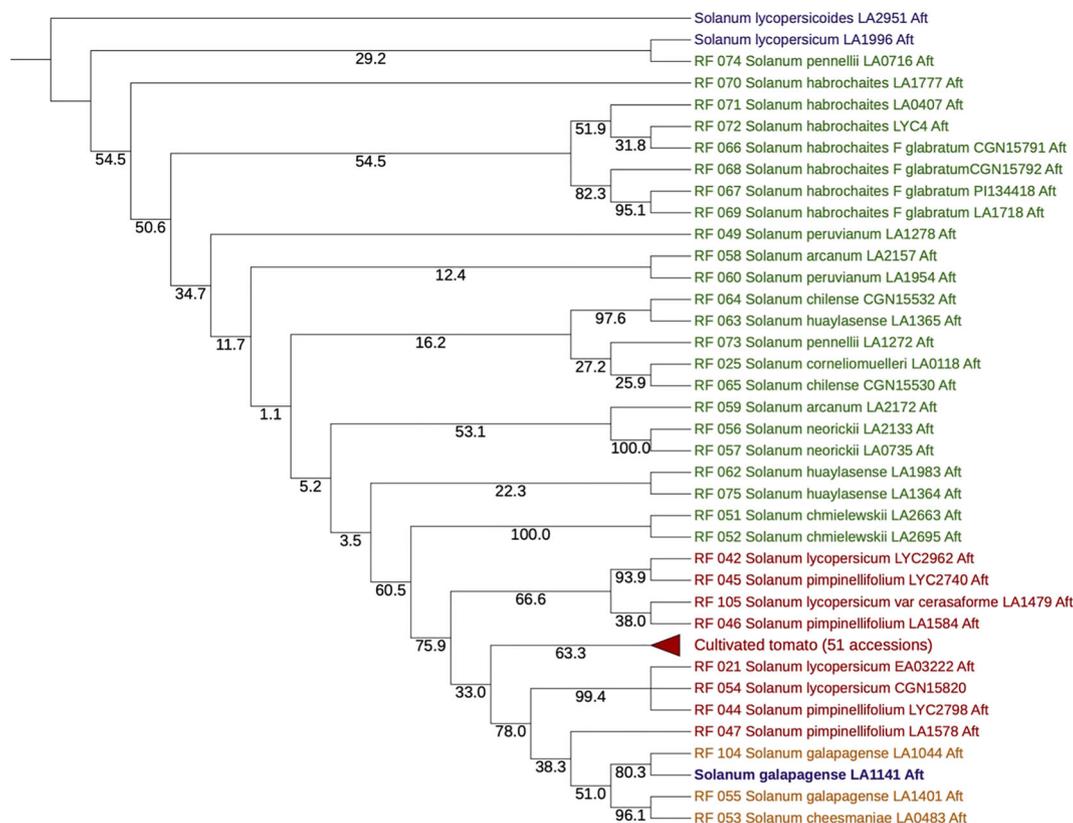


FIGURE 6 Phylogenetic tree showing the relatedness of MYB transcription factors underlying the *Aft* locus. The tree represents the clustering of genomic sequences underlying the *Anthocyanin fruit (Aft)* locus. *Solanum galapagense* LA1141 (purple, bold), *Solanum lycopersicum* OH8245, *S. lycopersicum* LA1996 (purple), 84 unique tomato accessions from the 100 tomato genome sequencing consortium, and *Solanum lycopersicoides* LA2951 (purple) are clustered. A maximum likelihood tree was constructed in the phangorn R package (Schliep, 2011) using the general time reversible model with the rate variation among sites described by a gamma distribution and the proportion of invariable sites (a.k.a. GTR + G + I model). Data resampling using 1,000 bootstrap replications was performed using the bootstrap.pml function, and bootstrap values are given for each branch. The tree was outgroup rooted at *S. lycopersicoides*. Identical *S. lycopersicum* sequences are condensed (red triangle)

4 | DISCUSSION

4.1 | Measuring tomato fruit pigmentation with quantitative methods

Tomato color depends on the type and quantity of pigments synthesized in the fruits. Anthocyanins are responsible for the purple coloration of immature LA1141 fruit. Delphinidin-3-rutinoside and petunidin-3-(p-coumaroyl-rutinoside)-5-glucoside were the major anthocyanins identified. As fruit ripened, the predominant anthocyanin changed from petunidin 3-(coumaroyl)rutinoside-5-glucoside in the MG stage to malvidin 3(coumaroyl)rutinoside-5-glucoside in the breaker stage (Figure 2). The chemical basis of pigmentation in progenies derived from LA1141 is consistent with those identified in introgression lines containing alleles from the green-fruited wild relatives (Jones et al., 2003). Phenotyping with quantitative measurements of color allowed us to distinguish classes of fruit that were useful for later genetic analysis. Cao et al. (2017) reported that it was difficult to distinguish marker classes of *atv* with qualitative phenotyping, but we were able to detect differences in hue values between homozygous

and heterozygous genotypes (Figure 4b). Additionally, our linkage analysis using quantitative measurements distinguished classes and showed that *Aft* is necessary to recover light purple color in progenies (Figure 1b). However, two unlinked loci are needed to recover the deep purple phenotype characteristic of IBL selection SG18-124 (Figure 1c). Inheritance of the light purple pigmentation in the progenies derived from LA1141 is consistent with patterns inherited from wild relatives in the green-fruited clade, and deep purple pigmentation is characteristic of a nonfunctional MYB repressor and a functional MYB activator.

4.2 | Three putative QTL affect LA1141 fruit color

Fruit color was associated with QTLs on chromosomes 7 and 10, and candidate genes were identified. The MYB-encoding gene family underlying the *Aft* locus mapped to the distal arm of chromosome 10 and was associated with higher hue values. Two QTLs, one on the proximal arm and one on the distal arm of chromosome 10, were associated with chroma. The *Golden 2-like* transcription factor



underlying the *uniform ripening (u)* locus maps to the proximal arm and mediates the brightness or dullness of the color. Accession LA1141 has a functional *Golden 2-like* allele underlying the *u* locus. The *u* locus increases chromoplast number, chlorophyll accumulation, and changing chromoplast distribution (Powell et al., 2012). This chlorophyll accumulation causes immature fruit to have patches of darker green color, especially on the fruit shoulders (Figure 1d). Sequence analysis of MYB-encoding genes underlying the *Aft* locus suggested that LA1141 may have a functional R2R3 MYB activator which could explain its purple pigmentation in early stages of fruit development, as measured by hue, chroma, and L^* . A nonfunctional allele of *atv* on chromosome 7 was detected based on interactions with *Aft* that increased pigmentation measured as hue (Figure 4a). The QTLs and the interaction between chromosomes 7 and 10 were also validated in the subsequent IBL derived F_2 generations (Figure 4b).

Two QTLs were associated with L^* , one on chromosome 6 and one on the proximal arm of chromosome 10. Only the QTL on chromosome 10 was validated in subsequent generations (Table 3). The region on chromosome 10 mapped to *u*. The *u* locus is likely affecting measurements of fruit darkness for similar reasons mentioned above. We expected the QTL on chromosome 6 to be associated with the *Beta* locus. However, mapping *B* failed to support this locus as a candidate (Table 2). We could not identify a candidate for the QTL on chromosome 6 corresponding to L^* in the IBC population. However, the QTL on chromosome 6 only explained 10% of the phenotypic variance compared with 35% of the variance explained by *u* (Table 2). Additionally, when we mapped *B* in the subsequent F_2 populations, we could detect association in only one of the populations (Table 3). In the SG18-200 derived F_2 population, *B* was associated with hue but not with chroma or L^* (Table 3). We believe that our ability to detect *B* in this population is attributed to the monomorphic alleles for *atv* and *u* reducing the range of hue (Table 3).

The wild tomato relative *S. chilense* is the source of *Aft* in cultivated tomato varieties such as Indigo Rose (Jones et al., 2003; Mes et al., 2008). Introgression of *Aft* from accession LA1141 offers an alternative source of this trait and may have advantages such as greater recombination and potentially less linkage drag (Hamlin et al., 2020). Deep purple pigmentation will be enhanced by the addition of *atv* and may be further increased in nonuniform (*u*) ripening backgrounds.

4.3 | The primary regulatory mechanism for anthocyanin accumulation is conserved in LA1141

The interaction between chromosome 7 (*atv*) and chromosome 10 (*Aft*) in the LA1141 \times OH8245 IBC population results in deep purple fruit (Figure 1c). This interaction suggests that the role of synergistic MYB regulatory genes underlying loci on 7 and 10 is conserved between LA1141 and the green-fruited species. A complex of interacting MYB transcription factors, basic helix-loop-helix transcription factors (bHLH), and WD40 repeat domains (WDR), known as the MYB-bHLH-WDR (MBW), modulate anthocyanin accumulation in

plants (Colanero, Perata, & Gonzali, 2020). The R2R3 MYB activators compete with the R3 MYB repressors for interaction with the bHLH transcription factor in the MBW complex (Colanero, Perata, & Gonzali, 2020). A CRISPR/Cas9 mediated silencing of MYB genes underlying the *Aft* locus suggested that only *An2-like* is needed for purple pigmentation in the peel of the tomato variety Indigo Rose (Yan et al., 2020). The same study showed that restoring function of *atv* in Indigo Rose reverts the coloration back to the light purple phenotype we observed in SG18-200 (Figure 1b) (Yan et al., 2020). Additionally, *atv* sequence targeted using CRISPR in the coding region of the second exon, where we observed the G to A SNP in LA1141, resulted in a loss of function of the R3/bHLH binding domain in LA1996 (Yan et al., 2020). This targeted mutation caused a purple phenotype similar to what we observed in our deep purple accession (Figure 1c).

4.4 | The potential origin of *Aft* and *atv* mutations in accession LA1141

In the red-fruited clade/Esculentum group, the structure of the *Aft* phylogeny places *S. galapagense* accessions closer to *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiformae*, and cultivated tomatoes, which is consistent with previously published *Solanum* phylogeny (Pease et al., 2016). Additionally, results from the outgroup rooted tree using CDS from distantly related species suggest that the green-fruited clade may be ancestral as proposed previously (Peralta & Spooner, 2001). Pigmentation in the tomato clade of *Solanum* is considered a phylogenetic signal with the corresponding to green and red-fruited clades with low expression of carotenoids and high expression of anthocyanins found in green-fruited groups (Gonzali & Perata, 2021). There are several possibilities that may explain the origin of *Aft* and purple pigment in accession LA1141. One possibility is that the LA1141 mutation in *Aft* may have arisen de novo. Alternatively, *Aft* may have been introduced into *S. galapagense* by introgression from *S. pimpinellifolium*. The mutation may also represent an ancestral variant present in the Esculentum group prior to the migration of tomatoes to the Galapagos. A spontaneous mutation hypothesis is consistent with the observation that *Aft* is not widespread in accessions of *S. galapagense* nor has a functional *Aft* been described in accessions of *S. pimpinellifolium*. However, additional investigation of *Aft* and *atv* alleles in *S. pimpinellifolium*, the likely ancestor of *S. galapagense* (Strickler et al., 2015), is needed for confirmation. The gain of function at *Aft* in LA1141 appears to originate in the red-fruited clade and is not likely an ancient introgression from a green-fruited progenitor.

It is interesting to speculate about how LA1141 acquired its purple fruit and how selective forces could maintain this pigmentation. The duplication of MYB transcription factors in flowering plants in general and the locus of linked family members on chromosome 10, specifically, provide opportunities for diversification and selection (Pickersgill, 2018). Our data do not suggest how this mutation is maintained. The occurrence of purple fruit in LA1141 could be indicative of convergent evolution with tomatoes in the green-fruited clade

due to unidentified selective pressure. Alternatively, the mutation could be maintained simply by chance. The flower morphology of *S. galapagense* is indicative of a strong inbreeding tendency (Rick & Fobes, 1975), and previous studies have suggested that there is limited genetic diversity within the species (Pailles et al., 2017). Consequently, high and low anthocyanin traits may have become fixed in different populations of Galápagos tomatoes because of genetic drift combined with inbreeding. It is also possible that purple pigmentation serves an adaptive role related to the synthesis of protective anthocyanins under high irradiance or the enticement of organisms that disperse seed (Grotewold, 2006). As an example, orange fruit is postulated to have a selective advantage on the Galápagos Islands due to seed disperser color preferences (Gibson et al., 2021). An investigation of known seed disperser preferences on the Galápagos Islands and LA1141 fruit could elucidate a possible evolutionary mechanism. However, the complete absence of anthocyanins in some populations of Galápagos tomatoes, such as accessions of *S. cheesmaniae* from the San Cristobal Island (Rick, 1967), is incompatible with a theory involving selective forces acting on purple color.

Evolutionary hypothesis explaining the maintenance of purple pigment in LA1141 must also take into account the role of the *atv* loss of function mutation on chromosome 7. The sequence data presented here suggest that multiple loss of function alleles at *atv* occurs on the Galápagos Islands. It is possible that purple pigmentation is not costly for the plant, and there is a lack of selective pressure on the Galápagos Islands to maintain the function of *atv*: similar to the loss of function of *phytoene synthase-1* at the *yellow flesh* (*r*) locus resulting in yellow fruit with low carotenoids (Gibson et al., 2021). A lack of selective pressure may explain why *atv* alleles may be more widespread in Galápagos tomatoes. Hypothesis related to adaptive maintenance, as discussed above, may offer an alternative explanation.

5 | CONCLUSION

This work describes the chemical and genetic basis of purple pigmentation in the fruit of accession of *S. galapagense* LA1141. The anthocyanins malvidin and petunidin are responsible for this color. Genes underlying the *atv*, *Aft*, and *u* loci are implicated as candidates for major QTLs. The loci *atv* and *Aft* interact, suggesting that the same mechanism producing anthocyanins in other purple fruited tomato accessions is responsible for pigment patterns in LA1141 fruit. *Aft* was previously known only from wild accessions in the green-fruited clade, and we probed Rick's hypothesis about an ancient hybridization event between progenitors of *S. galapagense* using genomic sequence from the *Aft* locus. Our phylogenetic analysis concluded that a functional allele of *Aft* in LA1141 is not derived from introgression from a green-fruited relative. The LA1141 *Aft* clusters with other Galápagos accessions and *S. pimpinellifolium* accessions from the South American mainland. Our findings guide us toward a better understanding of purple color found in the endemic Galápagos tomatoes and provide additional resources for characterizing anthocyanin biosynthesis in wild tomato relatives.

ACKNOWLEDGMENTS

We thank Jihuen Cho and the farm crews from the Ohio Agricultural Research and Development Center (OARDC) Wooster for assistance with management of the research. We thank Marcela Carvalho Andrade, Regis de Castro Carvalho, and Wilson Roberto Maluf from The Federal University of Lavras, 37200-000 Lavras, Brazil, for assistance with the LA1141 IBC population development. We also thank Dr. Jonathan Fresnedo Ramirez, Dr. Christine Sprunger, Dr. Chieri Kubota, and the anonymous reviewers for their constructive comments and insight which has helped improve the manuscript. Salaries and research support were provided by state and federal funds appropriated to The Ohio State University, OARDC, Hatch project OHO01405, and grant funds from USDA Specialty Crops Research Initiative Award number 2016-51181-25404. The Cooperstone lab was supported by Foods for Health, a focus area of the Discovery Themes Initiative at The Ohio State University and The Lisa and Dan Wampler Endowed Fellowship for Foods.

CONFLICT OF INTEREST

The authors have no conflict of interests to declare.

AUTHOR CONTRIBUTIONS

SF and DF: conceptualization. SF: phenotyping. JC: chemical analyses. SF: linkage map construction. SF: QTL mapping. SF and LS: marker development and sequencing. SF: bioinformatics and sequence analysis. SF: phylogenetic analysis. SF: writing. and DF: contribution to writing.

DATA AVAILABILITY STATEMENT

All data supporting the findings of this study are available within the paper. Additionally, pertinent supporting information tables and FASTA files are available in Zenodo at the following:

- i. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021a. Summary of PCR based markers used in this study (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650150>
- ii. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021b. LA1141 × OH8245 IBC single-nucleotide polymorphism (SNP) markers for genetic studies (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650152>
- iii. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021c. Accession passport and sequence data (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650141>
- iv. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021d. FASTA file containing the MYB-encoding gene An2-like genomic sequences corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5649546>
- v. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021e. FASTA file containing the MYB-encoding gene Ant1 genomic sequences corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5649996>



- vi. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021 f. FASTA file containing the MYB-encoding genes at the Aft locus with genomic sequences corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650058>
- vii. Fenstemaker, S., Sim, L., Cooperstone, J., Francis, D., 2021 g. FASTA file containing the MYB-encoding gene An2-like and Ant1 CDS corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650072>

ORCID

Sean Fenstemaker  <https://orcid.org/0000-0001-5689-075X>

Jessica Cooperstone  <https://orcid.org/0000-0001-7920-0088>

David Francis  <https://orcid.org/0000-0003-2016-1357>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Barton, D. W., Butler, L., Jenkins, J. A., Rick, C. M., & Young, P. A. (1955). Rules for nomenclature in tomato genetics: Including a list of known genes. *Journal of Heredity*, 46(1), 22–26. <https://doi.org/10.1093/oxfordjournals.jhered.a106504>
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: Making and mining the. “Gold Standard” Annotated Reference Plant Genome. *Genesis*, 53(8), 474–485. <https://doi.org/10.1002/dvg.22877>
- Bernal, E., Liabeuf, D., & Francis, D. M. (2020). Evaluating quantitative trait locus resistance in tomato to multiple *Xanthomonas* spp. *Plant Disease*, 104(2), 423–429. <https://doi.org/10.1094/PDIS-03-19-0669-RE>
- Berry, S. Z., Gould, W. A., & Wiese, K. L. (1991). Ohio 8245 processing tomato. *HortScience*, 26(8), 1093–1093. <https://doi.org/10.21273/HORTSCI.26.8.1093>
- Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7), 889–890. <https://doi.org/10.1093/bioinformatics/btg112>
- Cao, X., Qiu, Z., Wang, X., Van Giang, T., Liu, X., Wang, J., Wang, X., Gao, J., Guo, Y., Du, Y., & Wang, G. (2017). A putative R3 MYB repressor is the candidate gene underlying atroviolacium, a locus for anthocyanin pigmentation in tomato fruit. *Journal of Experimental Botany*, 68(21–22), 5745–5758. <https://doi.org/10.1093/jxb/erx382>
- Chaves-Silva, S., Dos Santos, A. L., Chalfun-Júnior, A., Zhao, J., Peres, L. E., & Benedito, V. A. (2018). Understanding the genetic regulation of anthocyanin biosynthesis in plants—tools for breeding purple varieties of fruits and vegetables. *Phytochemistry*, 153, 11–27. <https://doi.org/10.1016/j.phytochem.2018.05.013>
- Churchill, G. A., & Doerge, R. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3), 963–971. <https://doi.org/10.1093/genetics/138.3.963>
- Colanero, S., Perata, P., & Gonzali, S. (2018). The atroviolacea gene encodes an R3-MYB protein repressing anthocyanin synthesis in tomato plants. *Frontiers in Plant Science*, 9, 830. <https://doi.org/10.3389/fpls.2018.00830>
- Colanero, S., Perata, P., & Gonzali, S. (2020). Whats behind purple tomatoes? Insight into the mechanisms of anthocyanin synthesis in tomato fruits. *Plant Physiology*, 182(4), 1841–1853. <https://doi.org/10.1104/pp.19.01530>
- Colanero, S., Tagliani, A., Perata, P., & Gonzali, S. (2020). Alternative splicing in the anthocyanin fruit gene encoding an R2R3 MYB transcription factor affects anthocyanin biosynthesis in tomato fruits. *Plant Communications*, 1(1), 100006. <https://doi.org/10.1016/j.xplc.2019.100006>
- Cominelli, E., Gusmaroli, G., Allegra, D., Galbiati, M., Wade, H. K., Jenkins, G. I., & Tonelli, C. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in *Arabidopsis thaliana*. *Journal of Plant Physiology*, 165(8), 886–894. <https://doi.org/10.1016/j.jplph.2007.06.010>
- Commission Internationale de L Eclairage. (1978). *Recommendations on Uniform Color Spaces, Color Difference and Psychometric Color Terms*. 15(Suppl. 2). Paris, France: CIE.
- Consortium, T. G. S., Aflitos, S., Schijlen, E., de Jong, H., de Ridder, D., Smit, S., Finkers, R., Wang, J., Zhang, G., Li, N., & Mao, L. (2014). Exploring genetic variation in the tomato (*Solanum section Lycopersicon*) clade by whole-genome sequencing. *The Plant Journal*, 80(1), 136–148. <https://doi.org/10.1111/tpj.12616>
- Dal Cin, V., Kevany, B., Fei, Z., & Klee, H. J. (2009). Identification of *Solanum habrochaites* loci that quantitatively influence tomato fruit ripening-associated ethylene emissions. *Theoretical and Applied Genetics*, 119(7), 1183–1192. <https://doi.org/10.1007/s00122-009-1119-x>
- Darrigues, A., Hall, J., van der Knaap, E., Francis, D. M., Dujmovic, N., & Gray, S. (2008). Tomato analyzer-color test: A new tool for efficient digital phenotyping. *Journal of the American Society for Horticultural Science*, 133(4), 579–586. <https://doi.org/10.21273/JASHS.133.4.579>
- Darwin, S. C., Knapp, S., & Peralta, I. E. (2003). Taxonomy of tomatoes in the Galápagos Islands: Native and introduced species of *Solanum section Lycopersicon* (Solanaceae). *Systematics and Biodiversity*, 1(1), 29–53. <https://doi.org/10.1017/S1477200003001026>
- De Jong, W. S., Eannetta, N. T., De Jong, D. M., & Bodis, M. (2004). Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. *Theoretical and Applied Genetics*, 108(3), 423–432. <https://doi.org/10.1007/s00122-003-1455-1>
- De Mendiburu, F. 2017. Package agricolae: Statistical Procedures for Agricultural Research. R package version 1.2–8. <https://cran.r-project.org/web/packages/agricolae/agricolae.pdf>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2021a). Summary of PCR based markers used in this study (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650150>
- Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2021b). LA1141 × OH8245 inbred backcross (IBC) single nucleotide polymorphism (SNP) markers for genetic studies (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650152>
- Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2021c). Accession passport and sequence data (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5650141>
- Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2021d). FASTA file containing the MYB encoding gene An2-like genomic sequences corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5649546>
- Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2021e). FASTA file containing the MYB encoding gene Ant1 genomic sequences corresponding to wild and cultivated tomato accessions (Version 1) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5649996>
- Georgiev, C. (1972). Anthocyanin fruit (Af). *Rep Tomato Genet Coop*, 22, 10.
- Gibson, M. J., de Lourdes Torres, M., Brandvain, Y., & Moyle, L. C. (2021). Introgression shapes fruit color convergence in invasive Galapagos tomato. *eLife*, 10, e64165. <https://doi.org/10.7554/eLife.64165>



- Gnu. (2007). Free Software Foundation. Bash. (3.2.48) [Unix shell program]. <https://www.gnu.org/software/bash/>
- Gonzali, S., & Perata, P. (2021). Fruit colour and novel mechanisms of genetic regulation of pigment production in tomato fruits. *Horticulturae*, 7(8), 259. <https://doi.org/10.3390/horticulturae7080259>
- Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology*, 57(1), 761–780. <https://doi.org/10.1146/annurev.arplant.57.032905.105248>
- Haley, C. S., & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4), 315–324. <https://doi.org/10.1038/hdy.1992.131>
- Hamlin, J. A., Hibbins, M. S., & Moyle, L. C. (2020). Assessing biological factors affecting postspeciation introgression. *Evolution Letters*, 4(2), 137–154. <https://doi.org/10.1002/evl3.159>
- Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., van Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., & Fei, Z. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, hi-C proximity ligation and optical maps. *BioRxiv*. <https://doi.org/10.1101/767764>
- Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., Weisenfeld, N., Ramakrishnan, S., Kumar, V., Shah, P., Schatz, M. C., Church, D. M., & Van Deynze, A. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Horticulture Research*, 5(1), 4. <https://doi.org/10.1038/s41438-017-0011-0>
- Jones, C. M., Mes, P., & Myers, J. R. (2003). Characterization and inheritance of the anthocyanin fruit (aft) tomato. *Journal of Heredity*, 94(6), 449–456. <https://doi.org/10.1093/jhered/esg093>
- Joubert, T. L. G. (1962). A new jointless gene, $j-2^{in}$. *TGC Rep.*, 12, 29–30. <https://tgc.ifas.ufl.edu/vol12/Volume12.pdf>
- Kabelka, E., Yang, W., & Francis, D. M. (2004). Improved tomato fruit color within an inbred backcross line derived from *Lycopersicon esculentum* and *L. hirsutum* involves the interaction of loci. *Journal of the American Society for Horticultural Science*, 129(2), 250–257. <https://doi.org/10.21273/JASHS.129.2.0250>
- Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., Giuliano, G., Quattrocchio, F., Koes, R., Perata, P., & Gonzali, S. (2015). Tomato R2R3-MYB proteins SIAN1 and SIAN2: Same protein activity, different roles. *PLoS ONE*, 10(8), e0136365. <https://doi.org/10.1371/journal.pone.0136365>
- Kosambi, D. D. (1944). The estimation of map distance. *Annals of Eugenics*, 12, 505–525. <https://doi.org/10.1111/j.1469-1809.1943.tb02321.x>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49, W293–W296. <https://itol.embl.de/10.1093/nar/gkab301>
- Li, C., & Lu, S. (2014). Genome-wide characterization and comparative analysis of R2R3-MYB transcription factors shows the complexity of MYB-associated regulatory networks in *Salvia miltiorrhiza*. *BMC Genomics*, 15(1), 1–12. <https://doi.org/10.1186/1471-2164-15-277>
- Lincoln, R. E., & Porter, J. W. (1950). Inheritance of beta-carotene in tomatoes. *Genetics*, 35(2), 206–211. <https://doi.org/10.1093/genetics/35.2.206>
- Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., Schuster, D. K., Menasco, D. J., Wagoner, W., Lightner, J., & Wagner, D. R. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *The Plant Cell*, 15(8), 1689–1703. <https://doi.org/10.1105/tpc.012963>
- Mes, P. J., Boches, P., Myers, J. R., & Durst, R. (2008). Characterization of tomatoes expressing anthocyanin in the fruit. *Journal of the American Society for Horticultural Science*, 133(2), 262–269. <https://doi.org/10.21273/JASHS.133.2.262>
- Muller, C. H. (1940). *A Revision of the Genus Lycopersicon* (Vol. 382). US Department of Agriculture.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ohio Supercomputer Center, (1987). Ohio supercomputer center. <https://www.osc.edu/sites/osc.edu/files/documentation/osc-overview-au15.pdf>
- Okonechnikov, K., Golosova, O., Fursov, M., & Team, U. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, 28(8), 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>
- Ooe, E., Ogawa, K., Horiuchi, T., Tada, H., Murase, H., Tsuruma, K., Shimazawa, M., & Hara, H. (2016). Analysis and characterization of anthocyanins and carotenoids in Japanese blue tomato. *Bioscience, Biotechnology, and Biochemistry*, 80(2), 341–349. <https://doi.org/10.1080/09168451.2015.1091715>
- Orchard, C. J. (2014). *Naturally Occurring Variation in the Promoter of the Chromoplast-Specific Cyc-B Gene in Tomato Can Be Used to Modulate Levels of β -Carotene in Ripe Tomato Fruit*. Master's Thesis. The Ohio State University. http://rave.ohiolink.edu/etdc/view?acc_num=osu1416926782. Accessed September 2021
- Orchard, C. J., Cooperstone, J. L., Gas-Pascual, E., Andrade, M. C., Abud, G., Schwartz, S. J., & Francis, D. M. (2021). Identification and assessment of alleles in the promoter of the Cyc-B gene that modulate levels of β -carotene in ripe tomato fruit. *The Plant Genome*, 14(1), e20085. <https://doi.org/10.1002/tpg2.20085>
- Pailles, Y., Ho, S., Pires, I. S., Tester, M., Negrão, S., & Schmöckel, S. M. (2017). Genetic diversity and population structure of two tomato species from the Galapagos Islands. *Frontiers in Plant Science*, 8, 138. <https://doi.org/10.3389/fpls.2017.00138>
- Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology*, 14(2), e1002379. <https://doi.org/10.1371/journal.pbio.1002379>
- Peralta, I. E., Spooner, D. M., & Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect). *Lycopersicoides*, Sect. *Juglandifolia*, Sect. *Lycopersicon*; *Solanaceae*. *Systematic Botany Monographs* (Vol. 84, pp. 186–186). University of Michigan Herbarium.
- Peralta, I. E., & Spooner, D. M. (2001). Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). *American Journal of Botany*, 88(10), 1888–1902. <https://doi.org/10.2307/3558365>
- Pickersgill, B. (2018). Parallel vs. convergent evolution in domestication and diversification of crops in the Americas. *Frontiers in Ecology and Evolution*, 6, 56. <https://doi.org/10.3389/fevo.2018.00056>
- Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, 475(7355), 189–195. <https://doi.org/10.1038/nature10158>
- Povero, G., Gonzali, S., Bassolino, L., Mazzucato, A., & Perata, P. (2011). Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of aft and atv genes. *Journal of Plant Physiology*, 168(3), 270–279. <https://doi.org/10.1016/j.jplph.2010.07.022>
- Powell, A. F., Courtney, L. E., Schmidt, M. H. W., Feder, A., Vogel, A., Xu, Y., Lyon, D. A., Dumschott, K., McHale, M., Sulpice, R., & Bao, K. (2020). A *Solanum lycopersicoides* reference genome facilitates biological discovery in tomato. *BiorXiv*. <https://doi.org/10.1101/2020.04.16.039636>



- Powell, A. L., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernández-Muñoz, R., Vicente, A., & Lopez-Baltazar, J. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science*, 336(6089), 1711–1715. <https://doi.org/10.1126/science.1222218>
- Qiu, Z., Wang, H., Li, D., Yu, B., Hui, Q., Yan, S., Huang, Z., Cui, X., & Cao, B. (2019). Identification of candidate HY5-dependent and-independent regulators of anthocyanin biosynthesis in tomato. *Plant and Cell Physiology*, 60(3), 643–656. <https://doi.org/10.1093/pcp/pcy236>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for statistical Computing. <https://www.R-project.org>
- Rick, C. M. (1956). Genetic and systematic studies on accessions of *Lycopersicon* from the Galapagos Islands. *American Journal of Botany*, 43(9), 687–696. <https://doi.org/10.1002/j.1537-2197.1956.tb14433.x>
- Rick, C. M. (1961). Biosystematic studies on Galápagos Tomatoes. *Occasional Papers of the California Academy of Sciences*, 44, 59–77.
- Rick, C. M. (1967). Fruit and pedicel characters derived from Galápagos Tomatoes. *Economic Botany*, 21(2), 171–184. <https://doi.org/10.1007/BF02897867>
- Rick, C. M., Cisneros, P., Chetelat, R. T., & DeVerna, J. W. (1994). Abg—A gene on chromosome 10 for purple fruit derived from *S. lycopersicoides*. *Rep. Tomato Genet. Coop*, 44, 29–30.
- Rick, C. M., & Fobes, J. F. (1975). Allozymes of Galapagos tomatoes: Polymorphism, geographic distribution, and affinities. *Evolution*, 29(3), 443–457. <https://doi.org/10.1111/j.1558-5646.1975.tb00834.x>
- Sapir, M., Oren-Shamir, M., Ovadia, R., Reuveni, M., Evenor, D., Tadmor, Y., Nahon, S., Shlomo, H., Chen, L., Meir, A., & Levin, I. (2008). Molecular aspects of anthocyanin fruit tomato in relation to high pigment-1. *Journal of Heredity*, 99(3), 292–303. <https://doi.org/10.1093/jhered/esm128>
- Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evolutionary Biology*, 13(1), 1–15. <https://doi.org/10.1186/1471-2148-13-214>
- Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593. <https://cran.r-project.org/web/packages/phangorn/phangorn.pdf>, <https://doi.org/10.1093/bioinformatics/btq706>
- Schreiber, G., Reuveni, M., Evenor, D., Oren-Shamir, M., Ovadia, R., Sapir-Mir, M., Bootbool-Man, A., Nahon, S., Shlomo, H., Chen, L., & Levin, I. (2012). ANTHOCYANIN1 from *Solanum chilense* is more efficient in accumulating ANTHOCYANIN metabolites than its *Solanum lycopersicum* counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. *Theoretical and Applied Genetics*, 124(2), 295–307. <https://doi.org/10.1007/s00122-011-1705-6>
- Shannon, L.M., Yandell, B.S. & Broman, K. (2013). Users guide for new BCsFt tools for R/qlt. <http://cran.fhrc.org/web/packages/qlt/vignettes/bcsft.pdf>
- Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganai, M. W., Van Deynze, A., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., & Francis, D. M. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, 7(7), e40563. <https://doi.org/10.1371/journal.pone.0040563>
- Sim, S. C., Robbins, M. D., Wijeratne, S., Wang, H., Yang, W., & Francis, D. M. (2015). Association Analysis for Bacterial Spot Resistance in a Directionally Selected Complex Breeding Population of Tomato. *Phytopathology*, 105(11), 1437–1445. <https://doi.org/10.1094/phyto-02-15-0051-r>
- Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., & Lehväslaiho, H. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, 12(10), 1611–1618. <https://doi.org/10.1101/gr.361602>
- Strickler, S. R., Bombarely, A., Munkvold, J. D., York, T., Menda, N., Martin, G. B., & Mueller, L. A. (2015). Comparative genomics and phylogenetic discordance of cultivated tomato and close wild relatives. *PeerJ*, 3, e793. <https://doi.org/10.7717/peerj.793>
- Stommel, J. R. (2001). USDA 97L63, 97L66, and 97L97: Tomato breeding lines with high fruit beta-carotene content. *HortScience*, 36(2), 387–388. <https://doi.org/10.21273/HORTSCI.36.2.387>
- Sun, C., Deng, L., Du, M., Zhao, J., Chen, Q., Huang, T., Jiang, H., Li, C. B., & Li, C. (2020). A Transcriptional Network Promotes Anthocyanin Biosynthesis in Tomato Flesh. *Molecular Plant*, 13(1), 42–58. <https://doi.org/10.1016/j.molp.2019.10.010>
- Teng, S., Keurentjes, J., Bentsink, L., Koornneef, M., & Smeekens, S. (2005). Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. *Plant Physiology*, 139(4), 1840–1852. <https://doi.org/10.1104/pp.105.066688>
- Timberlake, C. F., & Bridle, P. (1982). *Distribution of Anthocyanins in Food Plants* (Vol. 125). Academic Press. <https://doi.org/10.1016/B978-0-12-472550-8.50009-3>
- Tomes, M. L., Quackenbush, F. W., & McQuistan, M. (1954). Modification and dominance of the gene governing formation of high concentrations of beta-carotene in the tomato. *Genetics*, 39(6), 810–817. <https://doi.org/10.1093/genetics/39.6.810>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115–e115. <https://doi.org/10.1093/nar/gks596>
- Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., Liu, X., Cao, B., & Qiu, Z. (2020). Anthocyanin fruit encodes an R2R3-MYB transcription factor, SIAN2-like, activating the transcription of SIMYBATV to fine-tune anthocyanin content in tomato fruit. *New Phytologist*, 225(5), 2048–2063. <https://doi.org/10.1111/nph.16272>
- Zeng, Z. B. (1994, August). A composite interval mapping method for locating multiple QTLs. In *Proceedings, 5th World Congress on Genetics Applied to Livestock Production* (Vol. 7). University of Guelph.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Fenstemaker, S., Sim, L., Cooperstone, J., & Francis, D. (2022). *Solanum galapagense*-derived purple tomato fruit color is conferred by novel alleles of the anthocyanin fruit and atrovialacium loci. *Plant Direct*, 6(4), e394. <https://doi.org/10.1002/pld3.394>