Allelic Variation of MYB10 Is the Major Force Controlling Natural Variation in Skin and Flesh Color in Strawberry (Fragaria spp.) Fruit

Cristina Castillejo, a,b Verónica Waurich, c,d Henning Wagner, c,d Rubén Ramos, a,b Nicolás Oiza, a,b Pilar Muñoz, a,b Juan C. Triviño, a Julie Caruana, f Zhongchi Liu, g Nicolás Cobo, a,b Michael A. Hardigan, h Steven J. Knapp, i José G. Vallarino, i,b Sonia Osorio, i,b Carmen Martín-Pizarro, i,b David Posé, i,b Tuomas Toivainen, j Timo Hytönen, j,k,l Youngjae Oh, m Christopher R. Barbey, m Vance M. Whitaker, m Seonghee Lee, m Klaus Olbricht, c José F. Sánchez-Sevilla, a,b and Iraida Amaya a,b,1

a Laboratorio de Genómica y Biotecnología, Instituto Andaluz de Investigación y Formación Agraria y Pesquera (IFAPA) Centro de Málaga, 29140 Málaga, Spain
b Unidad Asociada de I + D + i IFAPA-Consejo Superior de Investigaciones Científicas-Universidad de Málaga (IFAPA-IHSM Biotecnología y Mejora en Fresa, Málaga 29071, Spain

c Hansabred GmbH & Co. KG, 01108 Dresden, Germany
d Instituto de Botánica, Universidad de Huelva, 21013 Huelva, Spain

INTRODUCTION

The characteristic red color of strawberry (Fragaria spp) fruit is due to the accumulation of anthocyanins. These water-soluble pigments are synthesized through the flavonoid pathway (Almeida et al., 2007; Tohge et al., 2017). The initial substrate, 4-coumaroyl-CoA (CoA), is produced from phenylalanine through the sequential activity of the general phenylpropanoid pathway enzymes phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate:CoA ligase. The flavonoid pathway begins with

the condensation of one molecule of 4-coumaroyl-CoA and three molecules of malonyl-CoA by chalcone synthase (CHS). Chalcone isomerase (CHI) subsequently converts naringenin chalcone into naringenin. The subsequent steps involving flavonoid 3-hydroxylase (F3H) and flavonoid 3',5'-hydroxylase (F3'5'H) generate dihydroflavonols. The genes encoding the enzymes involved in these steps are known as early biosynthetic genes. Downstream genes of the pathway are known as late biosynthetic genes. These genes encode dihydroflavonol 4-reductase (DFR), which generate colorless leucoanthocyanidins, anthocyanidin synthase (ANS), which produces anthocyanidins, and several glycosyltransferases (UFGTs) that attach sugar molecules to anthocyanidins to generate the first stable anthocyanins.

Anthocyanins are synthesized at the endoplasmic reticulum and are later transported into the vacuole for storage via different types of mechanisms, including transport via glutathione S-transferases (GSTs; Luo et al., 2018). Anthocyanin biosynthesis is tuned through the transcriptional regulation of structural genes by transcription factors including MYB, bHLH, and WD-repeat proteins, which form the MBW ternary complex (Jaakola, 2013; Zhang et al., 2014; Xu et al., 2015). R2R3 MYB transcription factors are often the major determinants of natural variation in anthocyanin biosynthesis (Allan et al., 2008). In strawberry, the R2R3 MYB10 TF is considered to be the primary activator of structural genes in the anthocyanin pathway during fruit development (Lin-Wang et al., 2010; Medina-Puche et al., 2014), whereas another R2R3 MYB transcription factor from ripening strawberry, FaMYB1, represses the transcription of anthocyanin-related genes (Aharoni et al., 2001; Salvatierra et al., 2013).

Fruit color is a key quality trait for fruit breeders. Fruit color in the genus Fragaria varies widely from completely white fruits (WFs) to dark red, with wide variation in the internal concentration and distribution of anthocyanins throughout the fruit (Hancock, 1999; Hancock et al., 2003). Gaining insight into the genetic factors affecting natural variation in external (skin) and internal (flesh) fruit color is crucial for the efficient modification of this trait in new cultivars and will facilitate the rapid development of fruits with increased or reduced levels of anthocyanins. Besides contributing to fruit color, anthocyanins possess anti-oxidative properties, with several health-promoting effects and positive impacts on cardiovascular disorders and degenerative diseases (He and Giusti, 2010; Del Rio et al., 2013; Forbes-Hernandez et al., 2016).

The genus Fragaria belongs to the Rosaceae family and comprises 24 species, including the worldwide cultivated strawberry (Fragaria × ananassa; Staudt, 2009; Liston et al., 2014). This genus displays a series of ploidy levels, ranging from diploid species such as Fragaria vesca (2n = 2x = 14) to decaploid species such as some accessions of Fragaria iturupensis (2n = 10x = 70). The cultivated strawberry, Fragaria × ananassa, originated in France nearly 300 years ago via hybridization between two wild octoploid species, Fragaria chiloensis and Fragaria virginiana, which were introduced from South and North America, respectively.
(Darrow, 1966; Hancock, 1999). Cultivated strawberry and its wild progenitor species are allo-polyplpoids with $2n = 8x = 56$ chromosomes. Recent studies agree that one of the four subgenomes of the octoploid Fragaria species originated from a F. vesca ancestor and another from a F. iinumae ancestor, while the origin of the remaining two subgenomes, possibly related to the extant species Fragaria viridis and Fragaria nipponica, is still under investigation (Tennen et al., 2014; Edger et al., 2019, 2020; Liston et al., 2020).

Due to its simpler diploid genome, F. vesca is a model for genetic and genomic studies. Efficient genomic resources have been generated for this species, including transcriptomes (Hollender et al., 2012; 2014; Li et al., 2019) and a recently improved near-complete genome sequence (Shulaev et al., 2011; Edger et al., 2018). F. vesca accessions with WFs, including the sequenced Hawai4 accession, have been described and stored in multiple germplasm repositories. Despite having red skin color, F. vesca accessions, such as ‘Reine des Vallées’ or ‘Ruegen,’ are all characterized by white or pale-yellow flesh (NCGR, Corvallis repository). Red versus white external fruit color in F. vesca is governed by a single locus named C (Brown and Wareing, 1965). The C locus from the ‘Yellow Wonder’ cultivar was subsequently mapped to the bottom of linkage group (LG) 1 (Williamson et al., 1995; Deng and Davis, 2001). A recent genome-scale variant analysis showed that a single nucleotide polymorphism (SNP; G35C) causing an amino acid change (W12S) in the FvMYB10 gene was responsible for the loss of anthocyanins and the pale color of ‘Yellow Wonder’ fruits (Hawkins et al., 2016).

Several studies have focused on characterizing most of the structural and regulatory genes necessary for color development in the genus Fragaria (Moyano et al., 1998; Lunkenbein et al., 2006; Griesser et al., 2008; Carbone et al., 2009; Salvatierra et al., 2010; 2013; Thill et al., 2013; Fischer et al., 2014; Lin-Wang et al., 2014; Medina-Puche et al., 2014; Miosic et al., 2014; Duan et al., 2017; Hossain et al., 2018; and references therein). The identification of genes or genomic regions governing fruit color variation in cultivated strawberry is crucial for hastening the development of new cultivars with desired characteristics. However, the octoploid nature of this species complicates genetic analyses because up to eight alleles can be found if each homoeologous locus from the four subgenomes were conserved after polyploidization (Edger et al., 2019). A number of studies have detected quantitative trait loci (QTL) contributing to the external color intensity or anthocyanin content of strawberry fruits (Zorrilla-Fontanesi et al., 2011; Lerceteau-Köhler et al., 2012; Castro and Lewers, 2016). Each of these three studies identified three to six genomic regions that contribute to color-related traits, but the phenotypic variation explained by each QTL was relatively low, with only a few QTLs explaining more than 15% of phenotypic variation.

More recently, an 8-bp insertion in the coding region of FaMYB10 was associated with the loss of anthocyanins in fruits of cv Snow Princess, an octoploid strawberry cultivar with completely WFs (Wang et al., 2020), but its subgenome location was not described. Three full-length homoeologous FaMYB10 genes were annotated in the octoploid cv Camarosa genome (Edger et al., 2019). The patterns of expression and roles of these genes in determining fruit color variation are not fully known. Are all three genes mutated in white octoploid strawberry? How many homoeologs are expressed in developing fruit? Natural strawberry fruit color mutants affected in genes other than MYB10 have not been found. However, strawberry fruit color variants have been obtained by transient downregulation of the anthocyanin biosynthesis repressor FcMYB1, resulting in increased concentrations of anthocyanins. In addition, transient knock-down of RAP, encoding a GST operating as carrier protein of anthocyanins, led to reduced fruit coloration in cultivated strawberry (Salvatierra et al., 2013; Gao et al., 2020). Are any of these genes, apart from MYB10, behind fruit color variations found in nature?

In this study, we performed an extensive phenotypic, genetic, and molecular analysis of Fragaria genetic resources to identify genetic determinants that contribute to natural fruit color variation in strawberry. First, we screened available diversity for fruit color in the diploid F. vesca and identified new mutations for the loss of anthocyanins and associated red color. We then extended the analysis to seven octoploid accessions with fruit with either white flesh/red skin or white flesh/white skin compared to fruits with red flesh/red skin. One of our objectives was to identify specific alleles in cultivated strawberry that could be targeted for marker development that would aid breeders in the efficient development of new improved cultivars with desired fruit color. Strikingly, our results show that all analyzed color variants in the genus Fragaria are caused by independent mutations in the same gene, MYB10. Different genetic lesions are responsible for distinct flesh and skin color phenotypes, and therefore, allele-specific markers will need to be developed to track specific traits. We further show that independent mutations in only one of three MYB10 homoeologs (MYB10-2) cause skin- and flesh-color variation in octoploid strawberry. Red-flesh color was linked to a CACTA-like transposon insertion in the FaMYB10 promoter, while white-flesh mutant alleles lacking this insertion were inherited from white-fleshed F. chiloensis donors. We developed a high-resolution melting (HRM) marker able to predict WF skin color derived from a mutation in the coding region of FaMYB10. Subsequently, we developed two additional DNA markers, including one based on PCR and agarose gel and another on KASP (Kompetitive allele-specific PCR) that predict internal fruit color in diverse octoploid strawberry germplasm. These markers represent useful tools for the selection of fruit color, particularly when F. chiloensis accessions are used in breeding programs.

RESULTS AND DISCUSSION

Identification of Loci Controlling External Fruit Color in F. vesca

To identify genetic factors affecting fruit color in F. vesca, we developed an F2 mapping population of 145 lines derived from cv Reine des Vallées (RV660) and the IFAPA white-fruited accession ESP138.596 (WVS96). As previously described for other backgrounds (Brown and Wareing, 1965), red versus WF color in the RV660 × WVS96 F2 population segregated as expected for a single mutation (1:3; χ² test, P = 0.922). To fast-map the locus controlling red color in this F. vesca population, we prepared two pools and subjected them to a QTL-Seq approach (Takagi et al., 2013) that combines bulk-segregant analysis (Michelmore et al.,
with NGS technologies to locate candidate genomic regions more rapidly than traditional linkage mapping. We selected 34 F$_2$ plants with red fruits (RFs) and 32 F$_2$ plants with WFs to generate the two DNA pools referred to as RF and WF, respectively. Next, we performed Illumina high-throughput sequencing of each pool to a 50× genome coverage and aligned short reads to the F. vesca reference genome v4.0.a1 (Edger et al., 2018). Frequencies for SNPs in each pool were calculated, and average SNP-indexes were computed in 3-Mb intervals. The difference in SNP allele frequencies between the RF and WF pools (ΔSNP-index) were plotted against the F. vesca genome (Figure 1A). At the 97% confidence level, significant SNPs (in blue) were only detected on chromosome 1 (Fvb1) at intervals 7,970,000 to 14,800,000 and 23,450,000 to 23,870,000 bp.

We extended the analysis to small and large insertions/deletions (INDELs) in addition to SNP variants. Within significant intervals, we searched for genes with nonsynonymous variants and ΔSNP-indexes >0.5 and homozygous (SNP-index = 0 or 1) in the WF pool. The target intervals we identified harbored 105 candidate genes (Supplemental Data Set 1), including FvH4_1g22020 (FvMYB10). Bioinformatic analysis of large insertions detected a 52-bp insertion in FvMYB10 in the WF pool compared to the RF pool and the reference F. vesca genome. We focused our analysis on this 52-bp INDEL located at the beginning of the third

Figure 1. QTL-Seq Identifies a Significant Region on Fvb1 Controlling Fruit Color in RV660 × WV596.

(A) ΔSNP-index plot (Y-axis) over the seven F. vesca chromosomes. Line represents a sliding window of 3 Mb at 10-kb intervals. Significant SNPs are highlighted in blue (P < 0.03).

(B) FvMYB10 PCR amplification using primers R/W-F and R/W-R designed flanking the 52 bp INDEL (c.278_279ins). In RV660, the expected fragment size was 181 bp according to the reference Hawaii-4 genome. See scheme in (C).

(C) Schematic representation of FvMYB10-gypsy, the Long Terminal Repeat Transposable Element (LTR-TE) identified in the WV596 FvMYB10 coding sequence. In the scheme, primers used for population genotyping are highlighted (R/W-F, R/W-R, and ccm1, the latter at the 5’ end of the LTR-TE; Supplemental Figure 1).
exon of FvMYB10, the strongest candidate mutation underlying the WF phenotype.

The FvMYB10 Coding Region in White-Fruited WV596 Carries an Long Terminal Repeat-Transposable Element Insertion that Impairs Anthocyanin Accumulation

We designed primers flanking the 52-bp insertion and used them to amplify the corresponding genomic fragment from WF accession WV596 and the RF cv Reine des Vallées (RV660). Standard PCR generated the expected 181-bp product from RV660 but failed to generate a product in WV596. Reanalysis using the long-range 5-Prime PCR Extender System amplified a 10-kb product from WV596 (Figure 1B). Cloning and Sanger sequencing of the 10-kb genomic fragment from WV596 allowed us to confirm the insertion at position 278 of FvMYB10 cDNA, but it was larger than initially predicted. Further sequencing by primer walking of the entire 10-kb fragment revealed that the insertion located in the third exon of WV596 FvMYB10 corresponds to a Long Terminal Repeat Transposable Element (LTR-TE), 9509 bp long, belonging to the Gypsy subfamily. This new FvMYB10 allele was designated fvmbyb10-2, and the LTR-TE was named FvMYB10-gypsy. Upon insertion, FvMYB10-gypsy generated the 5 bp AAGAA target site duplication. Two almost identical (one nucleotide substitution) 1.7-kb Long Terminal Repeats (LTRs) were identified flanking the open reading frames (ORFs) necessary for TE replication and transposition (Figure 1C). To determine how tightly linked fvmbyb10-2 was with the WF phenotype in the RV660 × WV596 F2 population, we used a reverse primer that binds to the FvMYB10-gypsy 5′ terminal repeat to genotype the population in combination with primers flanking the insertion point (Figure 1C). The fvmbyb10-2 allele cosegregated with the white phenotype in the entire population (Supplemental Figure 1) following the expected 1:2:1 ratio for a single-gene codominant trait.

FvMYB10 transcript accumulation was not altered in the WF pool upstream of the insertion site but, as expected, it was abolished downstream of the insertion site (Figure 2A). In addition, the LTR-TE sequence introduced a series of premature stop codons in the FvMYB10 coding sequence. The predicted truncated protein lacks the C-terminal-conserved motif KPRPR[S/T]F for Arabidopsis (Arabidopsis thaliana) anthocyanin-promoting MYBs (Stracke et al., 2001); this motif is also found in Rosaceous MYB10 and known anthocyanin MYB regulators from other species (Lin-Wang et al., 2010). As a result, fvmbyb10-2 is expected to be nonfunctional and unable to induce the expression of FvMYB10 target genes from the anthocyanin pathway. Therefore, we tested the expression of some representative FvMYB10 targets in the fvmbyb10-2 background using RT-qPCR in ripe fruits from the RF and WF pools. As shown in Figure 2A, transcript accumulation from the downstream structural genes CHI, F3H, DFR, ANS/LDOX, and UFGT was significantly reduced in ripe WF, confirming the notion that fvmbyb10-2 is a loss-of-function allele.

To identify which secondary metabolites were affected by the downregulation of FvMYB10, we analyzed ripe fruits in the WF and RF pools by ultra performance liquid chromatography coupled to tandem mass spectrometry (UPLC-Orbitrap-MS/MS). The use of the two bulked pools of F2 lines instead of parental lines RV660 and WV596 allowed us to identify secondary metabolites affected only by FvMYB10 downregulation, excluding the metabolites that differ in the two accessions due to variations in other genes. A total of 88 metabolites were identified, including 30 ellagitannins, 41 flavonoids, 11 hydroxycinnamic acid derivatives, three hydroxybenzoic acid derivatives, and three terpenoids (Supplemental Data Set 2). Significant differences were detected for only 14 metabolites, and the levels of all but one metabolite were increased by FvMYB10 (Figure 2B). As expected, the levels of cyanidin hexose and pelargonidin hexose, the two main anthocyanins, were 625-and 1,950-fold higher in the RF pool, respectively. Mutation of FvMYB10 reduced the level of one terpenoid (sesquiterpenoid hexose) and one benzoic derivative (hydroxybenzoic acid-hexose) 2.2- and 9-fold, respectively. Interestingly, the mutation in FvMYB10 reduced the level of ellagic acid hexose 3.4-fold, while that of galloyl-bis (HHDP)-Glc increased 58%. Finally, the levels of one quercetin derivative and six hydroxycinnamic acid derivatives were reduced in the WF pool.

Effects in the same direction on quercetin derivatives and coumaric acid derivatives were observed when FaMYB1 was transiently downregulated in octoploid strawberry (Medina-Puche et al., 2014) or when FvMYB10 was engineered in F. vesca (Lin-Wang et al., 2014). Other studies that compared secondary metabolites in the fruits of cultivars with different fruit color also demonstrated that quercetin derivatives, coumaric and cinnamic hexoses, and ellagic acid were also associated with MYB10 function (Härtl et al., 2017; Wang et al., 2020). These results indicate that a very limited number of metabolites are affected by MYB10 regulation, suggesting that MYB10 has a primary influence on anthocyanin biosynthesis and effects on a few specific flavonols, ellagitannins, and hydroxycinnamic acids. The reduction in the levels of these metabolites did not affect the total antioxidant capacity of fruits (Supplemental Figure 2). Similarly, no significant differences in total soluble solids content (SSC), titratable acidity, or ascorbic acid content were observed between the RF and WF pools (Supplemental Figure 2). Therefore, fruits that are white due to the downregulation of MYB10, like red fruits, are rich sources of nutritional compounds other than anthocyanins.

Finally, we functionally validated fvmbyb10-2 as the causal agent of the lack of anthocyanin accumulation by transiently overexpressing FvMYB10 in WV596 fruits. Two constructs were generated to test the complementation: 3S pro:RV660 FvMYB10 and 3S pro:WV596 fvmbyb10-2. RV660 FvMYB10, the allele from the RF cv Reine des Vallées, was able to induce anthocyanin accumulation in these fruits, but WV596 fvmbyb10-2 was not (Figure 2C), further confirming the notion that fvmbyb10-2 is the causal gene of the WF phenotype. Previous studies using different F. vesca accessions identified an independent polymorphism (G35S) in FvMYB10 as the underlying cause of anthocyanin-less fruits (Zhang et al., 2015; Hawkins et al., 2016). This specific SNP is translated into a W12S amino acid substitution at a conserved residue within the R2 DNA binding domain and has been identified in five WF F. vesca accessions: Hawaii-4, cv Yellow Wonder, cv Pineapple Crush, cv White Soul, and cv White Solemacher. All of these accessions have the C nucleotide and thus the W12S substitution in FvMYB10 (Hawkins et al., 2016). This polymorphism, named fvmbyb10-1 in this study, was not found in WV596, which has the wild-type G nucleotide.
Independent FvMYB10 Mutations Explain the Lack of Anthocyanins in Diverse F. vesca Ecotypes

To assess the incidence of fvmyb10-1 or fvmyb10-2 alleles in F. vesca accessions with WF color, we broadened our analysis to include accessions with WFs from different geographic origins: Hawaii-4, cv Yellow Wonder, cv Pineapple Crush, cv White Soul, cv White Solemacher, GER1, GER2, cv South Queensferry, UK13, SE100, and FIN12 (Supplemental Data Set 3). PCR using our newly designed marker revealed that FvMYB10-gypsy insertion...
(fvmyb10-2 allele) was not present in any of the accessions analyzed besides WV596 (Figure 3A). In FIN12, we did not detect the FvMYB10-gypsy associated band or the endogenous FvMYB10 gene (Supplemental Figure 3A), pointing to a putative large rearrangement in the surrounding chromosomal region.

**Figure 3.** Multiple FvMYB10 Variants Were Detected in a Panel of 11 White-Fruited *F. vesca* Ecotypes. 
(A) FvMYB10-gypsy insertion (fvmyb10-2) was not detected by PCR analysis in FvMYB10 from other white-fruited *F. vesca* ecotypes. 
(B) Schematic representation of FvMYB10 showing the three alleles identified in white *F. vesca* ecotypes. Fvmyb10-2 and Fvmyb10-3 were identified in this study, whereas Fvmyb10-1 was previously described by Zhang et al. (2015) and Hawkins et al. (2016). 
(C) Outline of the protein products encoded by FvMYB10 alleles represented in (B) and list of the accessions in which they were found. The two repeats (R2R3) of the structurally conserved DNA binding domain are shaded in gray. The conserved motif KPRPR[S/T]F found in anthocyanin MYB regulators is shown in red. 
(D) FIN12 has a large deletion in Fvb1 resulting in the loss of 7 predicted genes (Supplemental Figure 3). FvMYB10 is among them and is highlighted in the scheme. Primers ccms2 and ccms49 were designed flanking the region in Hawaii-4. Both primers are 100 Kb apart, resulting in no PCR product. In FIN12, a >10 kb band was obtained. The picture on the right shows FvMYB10 CDS amplification with primers FvMYB10-F and FvMYB10-R only in Hawaii-4 but not in FIN12. 
(E) FvMYB10 transient overexpression assays showing complementation of GER2 and FIN12 white fruit. Two representative fruits are shown.
White-fruited ecotypes Hawaii-4, cv Yellow Wonder, cv Pineapple Crush, cv White Soul, and cv White Soremacher are known to carry the *fvmyb10-1* allele (Hawkins et al., 2016). We cloned the *FvMYB10* CDS from the remaining accessions, GER1, GER2, cv South Queensferry, UK13, SE100, and FIN12, and subjected them to Sanger sequencing. Sequence analysis revealed a third allele variant previously not described, *fvmyb10-3*, in the accessions from Northern Europe: GER1, GER2, and SE100. The *fvmyb10-3* allele had an A nucleotide insertion at position 329 of the cDNA (c. 329_330insA). The predicted protein contains the Asp111Gly substitution and a frameshift generating a premature stop codon 14 residues downstream from the insertion site (Figures 3B and 3C). Similar to *fvmyb10-2*, the resulting 123 amino acid protein transcribed from *fvmyb10-3* lacks the C-terminal domain found in anthocyanin-related MYB transcription factors (Stracke et al., 2001; Lin-Wang et al., 2010). The British accessions cv South Queensferry and UK13 both carried the previously published *fvmyb10-1* allele with the G35C SNP. In the accession FIN12, we were not able to amplify the *FvMYB10* coding region, but whole genome resequencing of FIN12 revealed an extensive region (~100 kb) in FIN12 chromosome 1 (Fvb1) with extremely low sequence coverage (Supplemental Figure 3B), suggesting that a large deletion affects this chromosomal fragment. The uncovered region of FIN12 roughly corresponds to positions 13,890,000 to 13,990,000 bp from the reference Hawaii-4 genome, which contains seven predicted coding sequences, including *FvMYB10* (Figure 3D). Primers designed in the predicted flanking regions detected a ~10-kb band in FIN12 but not in Hawaii-4 (Figure 3D), confirming the large deletion from FIN12 Fvb1.

To demonstrate that (1) *fvmyb10-3* in GER1, GER2, and SE100 and (2) the FIN12 large deletion from Fvb1 were the causal mutations leading to WFs, we tested whether a functional *FvMYB10* copy was able to restore fruit pigmentation when transiently expressed in fruits from these accessions. The same constructs described for *fvmyb10-2* complementation in WV596 were used in this assay. Once again, the expression of the full-length RV660 *FvMYB10* allele was sufficient to induce anthocyanin accumulation in all WF accessions (Figure 3E). Interestingly, anthocyanin accumulation was observed in the fruit epidermis as well as the inner receptacle, where red color is not observed in wild-type fruits. This phenomenon was also observed in WV596 in this study and by Hawkins et al. (2016) when complementing cv Yellow Wonder fruits. This probably resulted from the use of a constitutive 3SS promoter, suggesting that *FvMYB10* might not normally be expressed in the internal tissues.

In summary, we identified three *FvMYB10* mutations, in addition to the previously described *fvmyb10-1* allele (Hawkins et al., 2016), which explain the lack of pigmentation in the skin of several white-fruited *F. vesca* accessions: (1) An LTR-TE insertion at the third exon of *FvMYB10* (fvmyb10-2); (2) a single nucleotide (A) insertion at position 329 of *FvMYB10* cDNA (*fvmyb10-3*); and (3) a large deletion in chromosome 1 that removed seven predicted genes, including *FvMYB10*. The different allelic variants affecting fruit color described here are summarized in Supplemental Data Set 3. Notably, none of the WF F. *vesca* accessions we studied carried a functional *FvMYB10* gene. This indicates that the WF phenotype in *F. vesca* arose through different independent mutations in the same MYB gene, illustrating a convergent/parallel evolutionary mechanism.

A white-fruited strawberry mutant targeting a gene other than *MYB10* was artificially generated in a previous study. This mutant, the *reduced anthocyanins in petioles (rap)* mutant, was identified in a mutagenized *F. vesca* population (Luo et al., 2018). The *RAP* gene encodes a GST that binds anthocyanins to facilitate their transport from the cytosol to the vacuole. Similar WF and white stem phenotypes were observed in cultivated strawberry after *RAP* knockout using clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) (Gao et al., 2020). Neither rap mutants nor other mutants in different genes have been identified in nature. We therefore speculate that mutations in genes that result in a general lack of anthocyanins are negatively selected due to the role of anthocyanins in protecting plants against a wide range of abiotic stresses (Tohge et al., 2017), while the lack of anthocyanins only in fruits might not be as detrimental.

**Detection of QTLs Controlling Fruit Color in Octoploid Strawberry**

The improvement of cultivated strawberry is more challenging than the improvement of the diploid *F. vesca*, not only because of its octoploid genome but also because frequent homoeologous exchanges have occurred following polyploidization, which replaced substantial portions of some subgenomes with sequences derived from ancestrally related chromosomes (Edger et al., 2019). These exchanges are biased toward the *F. vesca*-like subgenome, although they are not completely unidirectional (Edger et al., 2019). It is therefore crucial to characterize the genetic control of color variation within the octoploid *Fragaria* species and to answer questions such as the following: How many genes are involved in controlling this trait? Are all possible homoeologous copies present in the genome? Are they being expressed? Are previously identified mutations in *FvMYB10* also present in octoploid strawberry?

To accomplish this objective, we studied two diverse octoploid populations characterized by broad variation in fruit skin (University of Florida strawberry breeding population 17.66) and flesh color (Hansabred SS × FcL population). The 17.66 population is an F₃ derived from biparental cross between two University of Florida (UF) advanced selections (FL 13.65-160 and FL 14.29-1) that segregates for white skin color (Supplemental Figure 4). Within this population, seedlings with white or light pink skin are also characterized by white internal flesh. Using the second-generation high-density 50K SNP Array (FanaSNP; Hardigan et al., 2020), we conducted a genome-wide association study (GWAS) to identify the chromosomal regions and specific SNPs associated with skin color in this breeding population. We performed association analyses of 43,422 SNP markers with 95 genotypes to detect marker-trait associations using three different analysis models [a general linear model (GLM), MLM, and multiple-locus mixed model (MLMM)]. SNPs strongly associated with WF skin color were located on chromosome Fvb1-2 (Figure 4A). The most significant SNP marker in all three models was probe AX-184080167, which is adjacent to a *MYB10* homoeolog (Fvb1-2 cv Camarosa, 15,395,876 bp).
Figure 4. SNPs and QTLs Associated with Fruit Color Traits in Octoploid Strawberry.

(A) Manhattan plots of a GLM, MLM, and MLMM for white fruit in UF breeding population 17.66. Plots for different chromosomes are shown in different colors, which follow the order: chromosome 1-1 to chromosome 7-4. Green line represents the significance threshold.
The second population used in this study, SS×FcL, is an F2 derived from the interspecific cross between F. ×ananassa cv Senga Sengana and F. chiloensis ssp. lucida USA2 that segregates for skin and flesh color in fruits. While all 105 F2 individuals accumulated various levels of anthocyanins in the skin, the variation in fruit flesh color was somewhat qualitative, with a number of F2 individuals displaying white flesh (Supplemental Figure 5). The population was phenotyped for skin and flesh color over three seasons and genotyped using DArTseq markers previously developed for octoploid strawberry (Sánchez-Sevilla et al., 2015). Using these data, we generated a linkage map comprising 2991 SNPs and covering a total length of 2377.21 cM (Supplemental Figure 6). Three QTLs (qSkinCol-1-2, qSkinCol-3-1, and qSkinCol-2-3) were detected for fruit skin color. The first QTL was detected in all three years, while the other two were only detected in one or two years (Supplemental Data Set 4). The phenotypic variance contributed by each QTL ranged from 12.2% to 25.9%.

Previous studies have detected QTLs for color traits in LGs belonging to Homoeology group (HG) 1 (Lerceteau-Köhler et al., 2012), 2 (Lerceteau-Köhler et al., 2012; Castro and Lewers, 2016; Zorrilla-Fontanesi et al., 2011), 3 (Lerceteau-Köhler et al., 2012; Zorrilla-Fontanesi et al., 2011), 5 (Zorrilla-Fontanesi et al., 2011; Castro and Lewers, 2016), and 6 (Castro and Lewers, 2016; Lerceteau-Köhler et al., 2012), although the different types of markers used in each study make it difficult to compare LGs and positions. Interestingly, a major QTL for flesh color, qFleshCol-1-2, was detected on LG 1-2 in the SS×FcL population in all three years (Figures 4B and 4C). This QTL explained a high proportion of the phenotypic variation (68.2% to 68.7%) and colocalized with qCoreCol-1-2, a QTL controlling 40.1% to 42.4% of variation in fruit core color (Figure 4B; Supplemental Data Set 4). The same gene likely affects fruit flesh and core color (internal color), and to a lesser extent skin color, strawberry. The 2-logarithm of the odds (LOD) confidence interval of qFleshCol-1-2 spanned a region on LG 1-2 from 53.2 to 55.0 cM (Supplemental Data Set 4) that corresponds to the region from 13.75 to 15.35 Mb on F. vesca chromosome 1 (v4.0.a1; Edger et al., 2018). The 1.6-Mb internal color QTL confidence interval region was found to contain 171 annotated genes. Among these, once again, MYB10 was the most likely candidate as the gene underlying the QTL.

FaMYB10-2 Is the Dominant Homoeolog in Octoploid Strawberry

As both analysis of UF population 17.66 and SS×FcL studies identified MYB10 from chromosome Fvb1-2 as a putative causal locus, we further investigated this homoeolog within an octoploid reference genome sequence. The F. ×ananassa cv Camarosa reference genome (Edger et al., 2019) contains four FaMYB10 homoeologs. Chromosomes Fvb1-1 and Fvb1-2 carry one FaMYB10 homoeolog each: FaMYB10-1 (maker-Fvb1-1-snap-gene-139.18 or FxaC_4g15020) and FaMYB10-2 (maker-Fvb1-2-snap-gene-157.15 or FxaC_2g30690), respectively. Two FaMYB10 genes were found on chromosome Fvb1-3, which were designated FaMYB10-3A (maker-Fvb1-3-augustus-gene-143.29 or FxaC_3g25620) and FaMYB10-3B (maker-Fvb1-3-augustus-gene-144.30 or FxaC_3g25830), but only FaMYB10-3B has a full-length ORF. FaMYB10-3A cannot be functional in activating anthocyanin biosynthesis, at least in cv Camarosa, as MYB10-3A CDS is interrupted by a Ty1-copia retrotransposon insertion at the end of the second intron. As a result, a truncated MYB10 protein lacking the 152 C-terminal residues is predicted. Finally, no FaMYB10 allele was found on chromosome Fvb1-4.

Therefore, besides FaMYB10-2, which is located at the same position as the identified QTL in both octoploid populations, other MYB10 copies could potentially be functional, as they encode full-length MYB proteins. However, alignment of transcriptomic sequences from a previous RNA-seq study (Sánchez-Sevilla et al., 2017) to the chromosome-scale cv Camarosa genome (Edger et al., 2019) allowed us to obtain subgenome-specific global expression profiles. This subgenome-specific expression analysis revealed that FaMYB10-2, in the F. inunae-derived subgenome, is the dominant homoeolog throughout fruit development in strawberry receptacle and achene tissues (Supplemental Figure 7). In pink (tuning) and red receptacles, for example, where FaMYB10 expression peaks, the expression of FaMYB10-2 represents 97% of total FaMYB10 expression in the respective ripening stage. By contrast, transcript accumulation from the other two full-length homoeologs from chromosomes Fvb1-1 and Fvb1-3, MYB10-1 and MYB10-3B, was barely detectable, accounting for only 0.6% of total FaMYB10 expression (Supplemental Figure 7). Similarly, a re-examination of MYB10 expression from a previous study using 61 strawberry lines (Barbey et al., 2020) demonstrated that FaMYB10-2 is the dominantly expressed homoeolog compared with those on Fvb1-1 and Fvb1-3. Thus, we expect that a nonfunctional FaMYB10-2 would be sufficient to abolish the induction of the anthocyanin pathway in octoploid strawberry.

Polyploidy, or whole genome duplication, is an important contributor to speciation in flowering plants. The formation of an allopolyploid involves the merger of genomes with separate evolutionary histories and often brings along different mechanisms to compensate for the increased gene dosage, including subgenome dominance. One important factor for subgenome dominance is a lower abundance of methylated TEs relative to other subgenomes (Alger and Edger, 2020). Strawberry is a complex allopolyploid that exhibit dominance in the F. vesca-derived subgenome, which contains the lowest TE density (Edger...
et al., 2019). In particular, chromosomal derived from the *F. vesca* subgenome are responsible for the expression of 88% of structural genes of the anthocyanin biosynthesis pathway (Edger et al., 2019). However, in this study, we demonstrated that anthocyanin biosynthesis is activated predominantly by MYB10-2, the homoeolog in the *F. iinumae*-derived subgenome.

**MYB10 Allelic Variants in White-Fruited Octoploid Accessions**

To identify sequence variation in MYB10 between white- and red-skinned fruits from UF breeding population 17.66, we performed NGS-based bulked-segregant analysis. FaMYB10 cDNAs amplified from pools of white- and red-fruited accessions were sequenced using the Illumina MiSeq platform. A total of 242,900 and 236,178 reads were generated in the WF and RF pools, 93.85% and 88.24% of which were mapped in the FaMYB10-2 gene, respectively. In WF accessions, we found an 8-bp insertion (ACTTATAC) at position 491 of the FaMYB10-2 ORF generating a premature stop codon and a predicted 179 amino acid protein instead of the 233 amino acid wild-type FaMYB10-2 (Supplemental Figure 8). Deletion of the 54 C-terminal residues may render FaMYB10 inactive. In fact, this same polymorphism was recently shown to be associated with WFs in cv Snow Princess, and the restoration of anthocyanin biosynthesis was only possible when wild-type MYB10 was overexpressed (Wang et al., 2020). The presence of the same mutation, hereafter named famyb10-1, suggests that the white-fruited selection FL 14.29-1 and cv Snow Princess may share a common pedigree.

Unlike in *F. vesca* accessions or in *F. xananassa* line FL 14.29-1, a comparison of the MYB10 coding sequence between cv Senga Sengana and USA2, the parents of the SS × FcL F2 population, revealed no polymorphisms potentially affecting protein stability/functionality. In fact, anthocyanins accumulated in fruit skin in all F2 individuals, suggesting that a functional MYB10 gene was expressed in this tissue. To examine the presence of structural variation in the cv Senga Sengana and USA2 MYB10 promoters (MYB10pro), we designed primers based on the Hawaii-4 *F. vesca* reference genome (Shulaev et al., 2011) intended to amplify a 1-kb product, from −941 to +55 relative to the FaMYB10 translational start codon (Figure 5A), and used them for PCR in two red-cv Camarosa and cv Senga Sengana) and three white-fleshed genotypes (USA1, USA2, and FC157). All accessions tested showed the expected 1-kb amplicon together with an unexpected 1.6-kb band. In accessions with white-fleshed fruits, we detected an extra 2.1-kb band in USA1 and USA2 and a 2.8-kb band in FC157 (Figure 5B). We cloned and sequenced PCR products from cv Senga Sengana, USA2, and FC157, confirming they were all MYB10pro alleles.

To assign each MYB10pro variant to its respective subgenome, we aligned them with the upstream regulatory sequences of the four cv Camarosa FaMYB10 homoeologs. Binding sequences of the primers used for MYB10pro amplification were localized in all four cv Camarosa homoeologs. The sequence between them was retrieved and used for alignment and homology tree construction together with the different sequenced alleles from cv Senga Sengana, USA2, and FC157 (Figure 5C). The alignment revealed that the region upstream of MYB10 is extremely polymorphic, containing a high density of SNPs, INDELs, and transposon-derived sequences. The initially expected ~1-kb product corresponds to the 938 bp long FaMYB10-3Apro, All MYB10-3Apro alleles from the different backgrounds were grouped in the same clade. All of them were very similar in length (924 to 938 bp) and presented a high degree of sequence conservation (97% to 99% identity). The 1.6-kb allele was common to MYB10-1 on chromosome Fvb1-1 and MYB10-3B on chromosome Fvb1-3. Both MYB10-1pro and MYB10-3Bpro alleles from cv Camarosa were identical and shared 90% identity in the overlapping region with MYB10-3Apro. However, MYB10-3Bpro and MYB10-1pro had a 710-bp insertion at position −276 from the initial ATG (green segment in Figure 5C). Interestingly, MYB10-3A pro shares higher homology with F. vesca FvMYB10pro than with MYB10-3Bpro. This relationship might reflect a chromosome translocation event from Fvb1-4, which did not retain the copy of MYB10, rather than a duplication of MYB10-3B (or MYB10-3A), as according to (Edger et al., 2019), the octoploid chromosome Fvb1-4 originated from the diploid *F. vesca* progenitor. As observed for MYB10-3Apro, MYB10-1pro, and MYB10-3Bpro alleles from the different accessions were almost identical in length (1641 to 1649 bp) and sequence (96% to 97% identity) and grouped together in the same clade of the homology tree.

Finally, the fourth FaMYB10pro allele from the reference cv Camarosa Fvb1-2 chromosome, FaMYB10-2pro, turned out to be much longer than the other three homoeologs, that is, almost 23 kb long. This size prevented us from amplifying FaMYB10-2pro from cv Senga Sengana and cv Camarosa under routine PCR conditions. Nonetheless, we were able to amplify MYB10-2pro alleles from white-fleshed accessions, which were significantly shorter than 23 kb: 2.1 kb in USA1 and USA2, and 2.8 kb in FC157. USA2 MYB10-2pro was highly similar to the USA2 MYB10-1pro and MYB10-3Bpro 1.6 kb alleles (94% identity) but it had a tandem duplication of a 471-bp sequence. The first unit of the tandem duplication was a MYB10-1pro / MYB10-3Bpro specific sequence (represented in light green in Figure 5C), and the second unit is colored in blue in the same scheme, highlighting that it is Fvb1-2-specific. On the other hand, FC157 MYB10-2pro was almost identical (98% identity) to USA2 MYB10-2pro, but FC157 had an additional 660-bp insertion disrupting one unit of the tandem duplicated sequence (dark blue segment in Figure 5C). Despite their tremendous size differences, all MYB10-2pro variants were clustered in the same tree branch, as occurred with the other MYB10pro homoeologs. Thus, we can conclude that MYB10pro alleles from the same homoeolog from different accessions are phylogenetically closer to each other than the four alleles of a given background. Notably, MYB10-2pro is the allele that shows higher polymorphism among the different accessions, and the shorter MYB10-2pro alleles found in white-fleshed accessions are strong candidates to be the underlying polymorphisms at the qFleshCol-1-2 QTL and the cause for white flesh in strawberry fruits.

**A Large Transposon Insertion in the MYB10-2 Promoter Is Associated with the Red-Flesh Phenotype.**

Sequence comparison among the 23-kb cv Camarosa MYB10-2pro sequence and the corresponding alleles from the white-fleshed accessions USA2 and FC157 revealed that the
common region of cv Camarosa MYB10-2pro shares 96% identity with alleles from both white-flushed accessions. The substantial size differences were due to the presence of four large INDELs of 4797, 1496, 454, and 14,064 bp (Figure 5C). These INDELs are located 17.7, 16, 15.4, and 986 upstream of the ATG initiation codon of MYB10-2, respectively. With use of the Repbase and the software tool Censor (Kohany et al., 2006), the 4797- and 14,064-bp insertions were identified as class II (DNA-type) TEs belonging to the CACTA family based on sequence similarity with the F. vesca EnSpm-1_FV element (Jurka, 2013; Shulaev et al., 2011). We designated them FaEnSpm-1 (4797 bp) and FaEnSpm-2 (14,064 bp). Both elements are bordered by identical 12-bp
terminal inverted repeat sequences 5‘-CACTACCAGAAA-3‘, and several subterminal direct and inverted repeats were also identified. Upon insertion, FaEnSpm-1 generated the 3-bp target site duplication TCG, whereas FaEnSpm-2 is flanked by the target site duplication CAA. FaEnSpm-2 and FaEnSpm-1 share important sequence similarity in their internal regions, but FaEnSpm-1 is thought to be a defective deletion derivative, as most of the sequences have been lost, including two transposase_21 domains (pfam02992; Marchler-Bauer et al., 2015).

The presence of FaEnSpm-2 at close proximity (<1 kb) away to the MYB10-2 coding region only in the red-fleshed accessions Camarosa and Senga Sengana prompted us to examine whether its loss correlates with the lack of anthocyanin accumulation in fruit flesh in the SS × Flc mapping population. We developed a codominant Fvb1-2-specific PCR marker intended to be predictive for internal fruit color (IFC-1 marker) using a combination of three primers depicted in Figure 6A. When assayed in the entire population, this marked was able to predict the phenotype in >95% of the F2 individuals (Figure 6A). Depending on the season, the phenotypic scores of 4 to 5 out of 105 F2 individuals did not match with their genotypes. This subtle deviation might be explained by the observed variability among fruits from the same line, which could lead to mis-scoring of the phenotype. Alternatively, due to the quantitative nature of this trait, these individuals might represent the natural genetic variation not associated with the large-effect qFleshCol-1-2 QTL.

The applicability of a given marker in breeding programs depends on its significance beyond a single cross. Therefore, we tested the relevance of our IFC-1 marker and its ability to predict internal fruit color in a wider set of white-fruited F. chiloensis accessions. To maximize genetic diversity, we selected accessions from ssp. chiloensis, ssp. lucida, Malus × domestica, and Capsicum annuum, and Brassica oleracea (Butelli et al., 2012; Jung et al., 2019; Yan et al., 2019; Zhang et al., 2019). Remarkably, the TE that enhances the expression of the host MYB gene in B. oleracea and pepper is (like FaEnSpm-2) a CACTA element (Jung et al., 2019; Yan et al., 2019). To investigate the association of FaEnSpm-2 with red flesh color, we performed RNA-seq to profile global gene expression in fruits from cv Senga Sengana and the six white-fleshed accessions previously genotyped in this study except for USA2. USA2 is a male plant and does not set fruits. Instead, we took advantage of its female sister line USA1, which presents the same phenotype segregating in the SS × Flc F2 population (Supplemental Data Set 3). The Fragments per kilobase of transcript per million fragments mapped (FPKM) values for each MYB10 homoeolog are presented in Figure 6C. As shown for cv Camarosa (Supplemental Figure 7), in cv Senga Sengana, there is an obvious expression level dominance biased toward MYB10-2. Interestingly, MYB10-2 was upregulated in cv Senga Sengana compared to all white-fleshed accessions. Total MYB10 expression was lower in all white-fleshed accessions too, although transcript accumulation from MYB10-1 was higher in the four F. chiloensis ssp. chiloensis accessions compared to cv Senga Sengana, USA1, and FC285, pointing to possible transcriptional compensation.

Comparative analysis of RNA-seq reads from the four MYB10 homoeologs allowed us to identify SNPs and INDELs among all transcripts relative to the reference cv Camarosa sequence (Supplemental Data Set 5). A total of 15 polymorphic nucleotides were identified in MYB10 coding regions from the seven samples analyzed. Among these, six were silent mutations, eight were missense mutations at nonconserved residues, and one was a nonsense mutation. None of the previously described polymorphisms in FvMYB10 or the FaMYB10 ACTTATAC insertion were present in the analyzed octoploid accessions. The nonsense mutation was only identified in F. chiloensis ssp. chiloensis accessions (FC156, FC157, FC160, and FC187) and consisted of a G to T transversion at position 478 of MYB10 ORF. The substitution predicts a premature stop codon at amino acid position 160, generating a truncated protein lacking 74 residues from the C-terminal end (Figure 6D). This new MYB10 allele derived from F. chiloensis was designated fcmyb10-1. This allele was found in three full-length MYB10 homoeologs from the F. chiloensis ssp. chiloensis accessions at different allelic dosage (Supplemental Data Set 5). We speculate that the presence of this premature stop codon might be triggering a genetic compensation response (El-Brolisy et al., 2019; Ma et al., 2019) by inducing MYB10-1 expression (Figure 6C). Remarkably, fruits from three of the F. chiloensis ssp. chiloensis accessions (FC156, FC160, and FC187) were homozygous for the fcmyb10-1 allele in Fvb1-2 and are completely white, with no anthocyanins present in the flesh or skin. By contrast, the fourth F. chiloensis ssp. chiloensis accession, FC157, was heterozygous for the G-to-T SNP at nucleotide 478 in all MYB10 homoeologs and has light pink epidermis, indicating once more that the MYB10 C-terminal end is required to induce anthocyanin biosynthesis (Figure 6D). Notably, no color developed in FC157 fruit flesh, suggesting that MYB10 might not be expressed in this tissue.

We used RNA-seq data to analyze the expression levels of the main structural genes of the anthocyanin biosynthesis pathway, finding that most anthocyanin biosynthesis pathway genes were
Different mutations in the coding sequence lead to completely white fruits lacking anthocyanins in the epidermis. (A) FaEnSpm-2 cosegregates with red flesh color in the SS × FcL F₂ population. The Fvb1-2-specific codominant IFC-1 marker was developed using a combination of three primers. Primer ccm107 binds to the junction sequence of FaEnSpm-2 and together with ccm109 amplify a 317-bp band associated with the red-flesh phenotype, whereas the 645 bp product (no TE insertion) associates with the white-flesh phenotype. Colored circles indicate flesh color: cream color = 1, orange = 2, and red = 3. Circles with two colors represent different scores from different seasons (Supplemental Figure 5). (B) IFC-1 marker in F. chiloensis accessions. Only the red-fleshed FC154 carried the FaEnSpm-2 TE at pMYB10-2. Among white-fleshed F. chiloensis, two different genotypes were found: (1) F. chiloensis ssp. lucida (USA1 and USA2) and pacifica (FC285) accessions both contain the smaller promoter allele version corresponding to the 2.1-kb pMYB10 variant cloned from USA2, whereas F. chiloensis ssp. chiloensis accessions (FC156, FC157, FC160, and FC187) harbor the larger 2.8-kb promoter allele identified in FC157. (C) Data from RNA-seq showing the expression level (FPKM value) of each MYB10 homoeolog in cv Senga Sengana and six white-fleshed F. chiloensis accessions. Expression level dominance is biased toward MYB10-2 in all accessions tested.
significantly downregulated in all white-fleshed accessions (Figure 7; Supplemental Data Set 6). Transcripts from genes whose products are involved in the early steps of this pathway, including PAL, C4H, and 4-coumarate:CoA ligase from the common phenylpropanoid pathway and the early biosynthetic gous for the putative nonfunctional expression levels were found in accessions that were homozy-

- **CHS**, **CHI**, and **F3H** were the most strongly affected. However, **UFGT** (a late biosynthetic gene) and GST were also notably downregulated in the accessions with the **fcmyb10-1** allele compared to the red-fleshed cv Senga Sengana. **DFR** and **ANS** did not appear to be under the transcriptional control of MYB10: even though they were downregulated in some white-fleshed or completely white accessions, this could be interpreted as a background effect. The most dramatically reduced gene expression levels were found in accessions that were homozy-

ous for the putative nonfunctional **fcmyb10-1** allele on chromosome Fvb1-2: FC156, FC160, and FC187. In these accessions, the expression of **CHS**, **F3H**, **UFGT**, and **GST** was practically abolished, in agreement with the total white phenotype of their fruits, thus confirming that **fcmyb10-1** is a loss-of-function allele.

In USA1 and FC285, the downregulation of some structural genes of the anthocyanin biosynthesis pathway was more moderate, but it should be noted that fruits from these accessions have red epidermis and therefore accumulate anthocyanins. Similar to our results, the downregulation of FaMYB10 did not affect expression in octoploid strawberry (Medina-Puche et al., 2014), while lower **ANS** expression in WFs than RFs was observed in *F. vesca* (Figure 2; Lin-Wang et al., 2014; Händt et al., 2017). Therefore, FaMYB10 and FvMYB10 may differ in the regulation of **ANS**, as previously suggested (Lin-Wang et al., 2014). Finally, we compared **MYB1** expression levels among the different accessions but found no correlation between fruit color or **MYB10** transcript levels and **MYB1** transcript accumulation (Supplemental Data Set 6), supporting the finding by Medina-

Puche et al. (2014) that FaMYB1 is not under FaMYB10 transcriptional control.

**Mining Putative Regulatory Elements in FaEnSpm-2**

Next, we investigated whether the FaEnSpm-2 element could be responsible for the activation of **MYB10-2** in red-fleshed fruits by providing novel cis-regulatory sites that behave as enhancers, or that respond to different stimuli and/or provide flesh-specific expression. We used the PlantPAN 3.0 database (Chow et al., 2019) to interrogate the presence of putative transcription factor binding sites (TFBSs) in **MYB10-2** upstream regulatory regions from cv Camarosa and the white-fleshed accessions USA2 and FC157. The cv Camarosa **FaMYB10-2** sequence analyzed spanned 3986 bp upstream of the ATG start codon, including the proximal 3 kb of **FaEnSpm-2** element and 986 bp of promoter sequence downstream of the **FaEnSpm-2** insertion site. For USA2 and FC157 **MYB10-2** sequences, we surveyed the sequences 2069 and 2729 bp upstream of the initial ATG. All three fragments shared almost 1 kb of the most proximal promoter region downstream of the FaEnSpm-2 insertion point. In the overlapping region, 36 SNPs and 4 small INDELS were found, not considering the large 660-bp insertion in FC157 **MYB10-2**. The results were filtered to a list of 84 putative cis-regulatory elements (at 156 positions) found exclusively at cv Camarosa **FaMYB10-2** (Supplemental Data Set 7).

We then focused our analysis on the motifs that are potentially relevant for fruit ripening and anthocyanin biosynthesis. Among these, hormone-responsive elements such as abscisic acid (ABA)- and methyl jasmonate (MeJA)-responsive elements were significantly enriched, with a total of 13 and 4 different putative motifs, respectively (Figure 8A; Supplemental Data Set 8). Additionally, three different MYB binding motifs were identified. These elements might be of special significance, as they could provide a feed-forward mechanism resulting in **MYB10** upregulation. In particular, the BOXLCOREDCPAL (ACCWWCCT) motif, a variant of the conserved MYBPLANT (MACCWAMC) and MYBPZM (CCWACC) elements, is likely to be bound by MYB10, as other R2R3-MYBs involved in phenylpropanoid and anthocyanin bio-

synthesis bind to this motif (Grotewold et al., 1994; Sablowski et al., 1994; Jian et al., 2019). Notably, the majority of red-flesh associated TFBSs were located within the FaEnSpm-2 TE, while only 6 elements (at 6 positions) specific to cv Camarosa **FaMYB10-2** were detected in the −1-kb promoter fragment shared by the three accessions (Supplemental Data Set 7).

Cultivated strawberry fruits accumulate anthocyanins in both their flesh and skin, while it is more common to find fruits and vegetables that accumulate anthocyanins only in their skin (Jaakola, 2013; Chaves-Silva et al., 2018). A wide variation in skin and flesh color has also been reported in apple, another Rosaceae, and color phenotypes are associated with different alleles of the apple **MYB10** ortholog (Espley et al., 2009; Chagné et al., 2013; Zhang et al., 2019). By contrast, the R1 and R6 motifs found in apple **MYB10** promoters and shown to enhance **MYB10** expression in different species (Espley et al., 2009; Brendolise et al., 2017) were not detected in any of the strawberry promoter se-

quences analyzed.

Light intensity and quality are important enhancers of antho-

cyanin biosynthesis, especially in fruit skin (reviewed by Jaakola, 2013). Furthermore, **MYB10** positively regulates light-controlled anthocyanin biosynthesis in apple and strawberry (Lin-Wang et al., 2010; Li et al., 2012; Kadomura-Ishikawa et al., 2015). Independently from light, ABA also promotes FaMYB10 expression, resulting in the induction of anthocyanin biosynthesis (Medina-

Puche et al., 2014; Kadomura-Ishikawa et al., 2015). Strawberry is a nonclimacteric fruit, and ABA plays a crucial role in the ripening process (Chai et al., 2011; Jia et al., 2011). Additionally, Suc can

Figure 6. (continued).

(D) Scheme representing the **fcmyb10-1** allele, which contains a G/T SNP in the coding sequence leading to a predicted truncated protein. Accessions homozygous for this variant in Fvb1-2, that is, FC187, are completely white. FC157 is heterozygous **fcmyb10-1/FcMYB10** in chromosome Fvb1-2 (Supplemental Data Set 5) and has a light pink epidermis. The two repeats (R2R3) of the structurally conserved DNA binding domain are shaded in gray. The conserved motif KPRPR[S/T]F found in anthocyanin MYB regulators is shown in red.
induce ABA accumulation and promote strawberry fruit ripening, including anthocyanin accumulation, via ABA-dependent and -independent mechanisms (Jia et al., 2013, 2016; Luo et al. 2019). Finally, MeJA treatment also increases anthocyanin accumulation in strawberry fruit (Concha et al., 2013). Further work will be required to better understand the significance of the predicted putative cis-elements and the molecular mechanisms driving higher \( \text{MYB10} \) expression in red-fleshed accessions. However, it is tempting to speculate that in the interior of the receptacle, where light quality or intensity might not be as effective in inducing \( \text{MYB10} \) expression, other endogenous signals, such as ABA, Suc, or MeJA, would be required to accumulate enough \( \text{MYB10} \) transcript to induce anthocyanin biosynthesis in a light-independent manner. In this scenario, promoters lacking the \( \text{FaEnSpm-2} \) element would not contain the putative regulatory sequences able to recruit the transcription factors that respond to those stimuli and would fail to induce anthocyanin biosynthesis in the receptacle flesh.

Known \( \text{F. vesca} \) accessions are characterized by red skin and white flesh, while cultivated strawberry cultivars display a characteristic and preferred red interior. Like in the white-fleshed \( \text{F. chiloensis} \) accessions analyzed in this study, \( \text{F. vesca MYB10}_{\text{pro}} \) lacks a \( \text{FaEnSpm-2} \)-like TE, as it is not predicted in the \( \text{F. vesca} \) Hawaii4 reference genome (Shulaev et al., 2011) and was not detected by PCR in accessions analyzed in this study such as RV660 or WV596. When putative TFBS were predicted in the homologous \( \text{MYB10}_{\text{pro}} \) region from \( \text{F. vesca} \) (941-bp long), only 5 of the 84 elements exclusively found at cv Camarosa \( \text{FaMYB10-2}_{\text{pro}} \) were detected (Supplemental Data Set 9). None of them were among the selected cis-elements (Supplemental Data Set 8) that are potentially relevant for anthocyanin production in fruit.
Figure 8. Prediction of Putative Regulatory cis-Elements in the Promoter Region of cv Camarosa FaMYB10-2 Not Found in MYB10-2pri from the White-Fleshed Accessions USA2 or FC157.

Functional validation of MYB10 as the causal agent under the internal fruit color QTL.
Transient Overexpression of MYB10 Overcomes White Flesh and Skin Phenotypes in Strawberry Fruit

Finally, if reduced MYB10 expression in the interior of the receptacle is the underlying cause of white-fleshed phenotype, we reasoned that increasing MYB10 expression should lead to phenotype complementation. Thus, we performed functional validation in fruits from CS-52, a white-fleshed F$_1$ line from the SS×FvL population, and the remaining white-fleshed F. chiloensis accessions, including FC156, FC157, FC160, FC187, and FC285. In all of these lines, transient expression of RV660 FvMYB10 under the control of the cauliflower mosaic virus 35S promoter was able to promote anthocyanin accumulation in both fruit flesh and epidermis, but transient expression of the truncated fmyb10-2 from WV596 (used as a control) was not (Figure 8B).

Taken together, these results confirm that MYB10-2, the MYB10 homoeolog from the F. iinumae-derived subgenome, is the dominant homoeolog in octoploid strawberry and, furthermore, the causal locus responsible for natural variation in internal and external fruit color. Alleles from this locus bear the FaEnSpm-2 CACTA element in the upstream regulatory region are associated with enhanced MYB10 expression, which results in anthocyanin accumulation in the inner receptacle. By contrast, strawberry accessions lacking the TE were characterized by white flesh, while heterozygous lines displayed an intermediate phenotype, indicating incomplete dominance, as also shown by QTL analysis, where the dominance effect of qFleshCol-1.2 QTL was estimated at 0.24-0.36, depending on the year (Supplemental Data Set 4). We postulate that the increase in FaMYB10 expression might be due to an expansion of its expression domain into fruit flesh. Additional analysis of FaEnSpm-2 indicated the presence of putative promoter motifs involved in ABA, MeJA, and sugar responses, as well as predicted MYB binding sites potentially involved in a positive feedback mechanism. Along with the promoter polymorphism, a number of F. chiloensis ssp. chiloensis accessions with white flesh and skin carry the new fcmyb10-1 allele at all three full-length homoeologous copies of FcMYB10, although at different doses (Supplemental Data Set 5). The predicted fcmyb10-1 is a truncated protein lacking 74 residues from the end portion of the activation domain. We have shown fcmyb10-1 fails to activate the expression of downstream anthocyanin structural genes PAL, C4H, CHS, F3H, and GT, leading to completely WFs or pink fruits, depending on the allelic dosage (Figure 7). We also found an independent missense mutation, famyb10-1, in lines with white skin color from UF breeding population 17.66. This mutation is the same as a previously identified INDEL (Wang et al., 2020). Our study precisely located this mutation to the dominant homoeologous allele on FvB1-2 and showed that it controls fruit skin color. Whereas polymorphisms found in the coding region appear to be more specific to a subset of accessions, the promoter polymorphism described in this study is common to a taxonomically diverse selection of white-fleshed accessions.

Development of Predictive High-Throughput Markers for Fruit Flesh and Skin Color

The development of high-throughput DNA tests with direct applicability for strawberry breeders has lagged behind that of other crops due to the complexity of the octoploid strawberry genome and the lack of quality subgenome-scale sequence information. Nevertheless, progress in that direction is expected due to the recent release of the first high-quality chromosome-scale octoploid strawberry genome (Edger et al., 2019). Assays for SNP detection such as KASP (Semagn et al., 2013) and high-resolution melting (Wittwer et al., 2003) have become tests of choice for breeders due to their accuracy, ease of scoring, and suitability for polyploid species such as strawberry (Whitaker et al., 2020).

We first developed an HRM marker (WS_CID_01) to predict the presence of the famyb10-1 allele for the 8-bp INDEL in MYB10 using the 17.66 population (102 individuals). Marker WS_CID_01 perfectly predicted white and red skin color (Figure 9A; Supplemental Data Set 10). For flesh color prediction, the IFC-2 marker that we developed accurately predicted MYB10-2 alleles and flesh color phenotypes. However, to develop a more high-throughput assay for marker-assisted selection of white- or red-fleshed strawberries, we designed the KASP marker IFC-2. To identify homoeolog-specific primers that were not expected to amplify MYB10 homoeologs in other subgenomes or other off-target DNA sequences, we aligned the promoter fragments described here to those of cv Camarosa and other octoploid accessions from Hardigan et al. (2020). We identified an A/G SNP 20 bp upstream of the initial ATG and 966 bp downstream of the FaEnSpm-2 insertion (Supplemental Figure 9). The A allele for the IFC-2 marker was exclusively observed in white-fleshed individuals and accessions. We used the IFC-2 KASP marker to genotype the SS×FvL mapping population (n = 108), in addition to two red-fleshed F. ×ananassa cvs (Camarosa and Candonga), the red-fleshed F. chiloensis accession FC154, and the 6 white-fleshed F. chiloensis accessions described earlier. The IFC-2 KASP marker produced codominant genotypic clusters and was 99% predictive of white- and red-fleshed phenotypes (Figure 9B). The red-fleshed allele (G) was observed in the white-fleshed accession FC285, which was the only disparity in our study. The region targeted with this marker is highly polymorphic and even though two common primers accounting for an additional SNP were employed, it might not work in some specific backgrounds. Still, this marker worked in >99% of the genotypes tested, making it a valuable diagnostic tool for breeders. Both the WS_CID_01 and IFC-2 markers should facilitate the rapid introgression of
target alleles into elite backgrounds and accelerate the development of white or red-flushed cultivars, respectively.

Since cultivated strawberry breeding began in the 1800s using a small number of lines (Darrow, 1966), the diversity of *F. × ananassa* has increased and its genome reshaped by repeated interspecific hybridizations with phylogenetically diverse *F. chiloensis* and *F. virginiana* accessions (Darrow, 1966; Bringhurst and Voth, 1984; Hancock et al., 2001; 2002; Gil-Ariza et al., 2009; Hancock, 1999; Hardigan et al., 2018; Liston et al., 2014). The introgression of beneficial alleles from these species has also resulted in the accumulation of unfavorable alleles in cultivated strawberry. One unwanted trait in most strawberry breeding...
programs aimed at developing varieties with red skin is white internal flesh. On the other hand, strawberries with both white skin and white flesh are becoming popular in some countries such as Japan. Furthermore, current breeding programs are particularly directed toward increasing fruit quality and pathogen resistance (Whitaker et al., 2020), and both F. chiloensis and F. virginiana represent reservoirs of interesting alleles for the future improvement of biotic and abiotic stress tolerance and fruit flavor and aroma (Aharoni et al., 2004; Johnson et al., 2014; Hancock, 1989).

Our findings show that MYB10 exerts simple genetic control over strawberry skin and flesh color, making it a good target for molecular breeding. Therefore, we anticipate that the markers developed in this study will enable efficient advances in breeding, particularly when wild relatives are used as parental lines. Alternatively, CRISPR-Cas9 technology could be used to speed up the breeding process, allowing the direct manipulation of the color trait in any genetic background. CRISPR/Cas9 gene editing has been successfully utilized in both woodland (Zhou et al., 2018) and cultivated strawberries (Martín-Pizarro and Posé, 2018), highlighting the potential of this technique for modifying strawberry fruit color in a more efficient and precise manner.

METHODS

Plant Materials and Phenotypic Evaluation

**Diploid Fragaria vesca Germplasm**

An F₂ mapping population of 145 lines between the everbearing, non-runnering Fragaria vesca cv Reine des Vallées (ESP138,660; RV660) and the WF F. vesca ESP138,589 (WVS96) was developed from one F₁ plant (F₁-14) that was able to produce runners and had RFs. The population was grown in a greenhouse in Málaga, Spain and phenotyped for two years. The two traits, runnering and red/white fruits, were found to segregate independently as two single mutations. The remaining F. vesca accessions were grown under the same conditions and are described in Supplemental Data Set 3 and include cv South Queensferry, GER1, and GER2 from the Günter Staudt collection (Dresden, Germany) and UK13, SE100, and FIN12 from the T. Hytönen and D. Posé collection.

**Octoploid Fragaria Germplasm**

The University of Florida breeding population 17.66 was derived from a cross between FL 13.65-160 (red) and FL 14.29-1 (white) selections (Supplemental Figure 4; Supplemental Data Sets 3 and 8). The population (102 individuals) was grown under open field conditions with two plants per plot. Fruit skin color was assayed three times from December 2018 to February 2019 at the UF/IFAS Gulf Coast Research and Education Center (Balm, Florida).

To generate the SS×Fcl F₂ population of 105 progenies, a cross between Fragaria ×ananassa cv Senga Sengana and F. chiloensis ssp. lucida USA2 was performed at Hansabred, Germany in 2008 (Supplemental Figure 5). An F₁ seedling, cloned under the number P-90999, was selected among the progeny based on key breeding traits such as tolerance to two spotted spider mite (*Tetranychus urticae* Koch), yield, and fruit aroma and color. The selected F₁ individual was self-pollinated to obtain the F₂ mapping population. The population was grown under field conditions at Hansabred and scored for fruit skin, flesh, and core color in three seasons (2014, 2016, and 2019). Skin color was evaluated using a five-score scale from completely white to dark red. Flesh and core color was evaluated using a three-score scale: 1, white; 2, light red; and 3, red (Supplemental Figure 10). The parental line F. chiloensis ssp. lucida USA2 is a male individual that does not set fruit, and therefore USA1, a sister female plant collected from the same area, was included in phenotypic and molecular analyses, as were other F. chiloensis accessions from diverse origins (Supplemental Data Set 3).

**DNA Extraction and QTL-Seq Analysis of Diploid Fragaria vesca**

DNA was extracted from young leaves of parental plants, F₁ plants, and each F₂ line using the CTAB method (Doyle and Doyle, 1990) with minor modifications. For rapid mapping of the fruit color mutation, we performed whole-genome resequencing of DNA from two bulked populations as previously described by Takagi et al. (2013). The RF and WF pools were produced by mixing equal amounts of DNA from 34 RF and 32 WF F₂ lines, respectively.

Pair-end sequencing libraries with insert sizes of ~350 bp were prepared, and 100-bp sequences were produced with 50× genome coverage using an Illumina HiSeq 2000. The reads from the RF and WF pools were mapped independently against the latest version of the Fragaria vesca reference genome, F. vesca H₄ V4.1 (Edger et al., 2018). The low-quality reads (<Q20 Phred scale) were filtered using the samtools method (Li et al., 2009). Duplicates that originated from PCR were eliminated using the picard-tools program.

For the single nucleotide variants and small INDELS calling process, the gatk algorithm (McKenna et al., 2010) was applied. The large INDELS were identified using the Manta algorithm (Chen et al., 2016). The variants with coverage less than 20 reads in both samples were not considered in downstream analyses.

SNP frequencies in each pool were calculated as the proportion of reads harboring the SNP different from the reference genome. Thus, SNP frequency = 0 if all short reads match the reference sequence and 1 if they correspond to the alternative allele. SNP positions with read depth <7 and frequencies <0.3 in both pools were excluded, as they may represent spurious SNPs called due to sequencing or alignment errors. The parameter ΔSNP-index was calculated as the absolute difference between SNP frequencies in the RF and WF pool samples. To detect significant genomic regions, the genome was split into 3-Mb sliding windows with 10-kb increments. For each window, the average ΔSNP-index was calculated and used for the sliding window plot. To identify statistically significant regions, a Wilcoxon test was applied using a confidence level of 0.03.

**Whole Genome Resequencing of FIN12 and Data Analysis**

Whole genome resequencing of the FIN12 accession was performed at DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Finland using an Illumina NextSeq 500 sequencer. Library preparation, pair-end sequencing (150 + 150 bp), and the alignment of the sequencing reads on the F. vesca H₄ reference genome v4.1 (Edger et al., 2018) were performed as described by Koskela et al. (2017).

**GWAS Analysis of F. × ananassa Breeding Germplasm**

A total of 95 individuals from UF family 17.66 were used for GWAS. Total genomic DNA was isolated from leaf tissues using a modified CTAB method described by Noh et al. (2018). Whole-genome SNP Genotyping was performed using the FanaSNP 50K Array (Hardigan et al., 2020). A GLM, mixed linear model (MLM), and MLM were used for association analyses. For rapid mapping of the fruit color mutation, we performed downstream analyses.

**Linkage Mapping and Detection of QTL in Octoploid Fragaria**

DNA was extracted from young leaves of parental plants, F₁ plants, and the 105 F₂ lines using the same CTAB method described for diploid Fragaria. A
total of 16,070 SNP markers were produced using the strawberry DArTseq platform (Sánchez-Sevilla et al., 2015). SNP markers that were mono-

morph in the progeny were removed, and markers that did not fit the expected 3:1 segregation ratio using the χ² test (P = 0.05). Next, markers with rowSums <400 were also removed. Finally, markers with more than 5% missing scores (more than five progeny lines) were excluded, resulting in a total of 9005 dominantly scored SNP markers. For 5856 markers, the reference and the alternative allele segregated at a 1:2:1 ratio (as expected for a F₂ population) and were transformed into 2523 codominant markers (COD-SNPs). The 2523 COD-SNPs were used together with 2446 dominant

SNP markers (DOM-SNPs) for mapping using JoinMap 4.1 (van Oojen, 2006). First, JoinMap software was used (coding the markers as a CP population) to infer the phases of markers heterozygous in both parental lines and thus of unknown origin in the F₂ line. Phase information was then used to assign phases and to code the SNP markers as a F₂ population. Grouping was performed using independence LOD and the default settings in JoinMap 4.1, and LGs were chosen at a LOD of 9 for all 28 groups obtained. Map construction was performed using the maximum likelihood mapping algorithm and the following parameters: chain length 5,000, initial accep-
tance probability 0.250, cooling control parameter 0.001, stop after 30,000 chains without improvement, length of burn-in chain 10,000, number of Monte Carlo EM cycles 4, chain length per Monte Carlo EM cycle 2,000, and sampling period for recombination frequency matrix samples 5. A total of 422 identical loci and 517 loci with similarity > 0.99 were removed. The seven HGs were named HG 1 to 7 based on the corresponding LGs in the diploid Fragaria reference map. BLAST analysis was performed using the SNP marker sequences as queries against the recently published Camarosa genome and assigned to the best matching chromosome. LGs within homoeologous groups were then named 1 to 4 according to the corresponding Camarosa chromosome number (Edger et al., 2019).

QTL analyses were performed using MapQTL 6 (van Oojen, 2009). The raw relative data were first analyzed using a nonparametric Kruskal-Wallis rank-sum test. A stringent significance level of P ≤ 0.005 was used as a threshold to identify markers linked to QTLs. Second, transferred data sets for nonnormally distributed traits were used to identify and locate mQTLs using Interval Mapping with a step size of 1 cm and a maximum of five neighboring markers. Significance LOD thresholds were estimated with a 1,000-permutation test for each trait. The most significant markers were then used as cofactors for restricted multiple QTL mapping analysis.

mQTLs with LOD scores greater than the genome-wide threshold at P ≤ 0.05 were declared significant. Linkage maps and QTLs were drawn using MapChart 2.2 (Voorrips, 2002).

**RNA Extraction and RT-quantitative PCR**

Total RNA was extracted from three independent pools (20 to 25 ripe fruits each) of each WF and RF phenotype. Fruits for these biological triplicates were collected from the same F₂ lines of the ‘RV660’ × ‘WW596’ population used for QTL-seq. We followed a CTAB method described by Gambino et al. (2008), with minor modifications. Briefly, 300 mg of powdered fruit sample was used as starting material and RNA was resuspended in 30 μL sterile water. Following digestion with a TURBO DNA-free Kit (Thermo Fisher Scientific AM1907), cDNA was synthesized from 1 μg of RNA using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific 4,368,814). RT-qPCR analysis was performed with IQ SYBR Green Supermix in an iCycler iQ5 system (Bio-Rad) following the manufacturer’s instructions. Three technical replicates (within each experiment) were performed for each biological replicate. The relative expression level of a gene of interest in each sample was calculated as Eₐ₀⁻¹ and normalized by the Eₐ₀⁻¹ value for GAPDH as the internal reference gene. The se-

quences of primers used in this study are listed in Supplemental Data Set 11.

**Semi-polar Metabolite Analysis Using UPLC-Orbitrap-MS/MS**

To investigate the effect of FvMYB10 mutation, three independent pools of 20 to 25 ripe fruits from each phenotype (RF and WF) were analyzed. Fruits were collected from the same F₂ lines used for QTL-seq and RT-qPCR. Fruit tissue from each pool was ground in liquid nitrogen and stored at −80°C. Secondary metabolite profiles of RF and WF pools were obtained as described by Vallarino et al. (2018) using a Waters Acquity UPLC system. Full documentation of metabolite profiling data acquisition and interpretation is provided in Supplemental Data Set 2.

**Quantification of Fruit Quality Parameters**

To determine SSC, TA, Trolox equivalent antioxidant capacity (TEAC), and ascorbic acid levels, frozen fruit powder from the same samples (in bi-

ological triplicate) used for expression and metabolic studies were used. For SSC, ~0.5 g of sample was defrosted to room temperature and the puree deposited onto the lens of a refractometer (Atago PR32, Japan). Titratable acidity was evaluated with an automatic titrator (Titroline easy, Schott North America) as previously described by Zorrilla-Fontanesi et al. (2011). Polyphenols were extracted in 80% (v/v) methanol, and the anti-

oxidant capacity of fruit samples was measured based on the ability of antioxidant molecules to quench the ABTS⁺ radical cation [2,2′-azino-

bis(3-ethylbenzothiazoline-6-sulfonate); Sigma–Aldrich] compared with the antioxidant activity of standard amounts of Trolox. TEAC assays were performed as described (Fellegrini et al., 1999; Pe et al., 1999) using 2 μL of sample and 250 μL of radical reagent. The absorbance at 734 nm was measured after 5 min at 25°C in a spectrophotometer. Results were ex-

pressed as μmoles of Trolox equivalents per gram of fresh weight (μmol TE/ g FW).

L-Ascorbic acid (L-AA) was measured by HPLC (Davey et al., 2006; Zorrilla-Fontanesi et al., 2011). In brief, 0.25 g of fruit powder was ho-

mogenized in 1 mL cold extraction buffer (metaphosphoric acid 2% [w/v], EDTA 2 mM) and incubated at 0°C for 20 min. The samples were cen-

trifuged at 14,000 rpm for 20 min at 4°C, and the supernatant was filtered (0.45-mm membrane) and transferred to HPLC vials kept on ice. Finally, a 5-

mL sample was injected in a Rx-C18 reverse-phase HPLC column (4.6 × 100 mm, 3.5 mM, Agilent Technologies) and detection was performed at 254 nm in a photodiode array detector (G1315D, Agilent Technologies). The mobile phase consisted of a filtered and degassed solution of 0.1 M NaH₂PO₄, 0.2 mM EDTA, pH 3.1 (Harapanhalli et al., 1993), and the flow rate was 0.7 mL/min. L-AA content was calculated by comparison with values obtained from a standard curve and were expressed as mg L-AA per 100 g of fresh weight.

**Amplicon Sequence Analysis**

Amplicon sequencing was performed using cDNA from WF and RF accessions from UF family 17.66. The fruits from each white-skinned and red-skinned accession were collected in the field, and 2-mm-thick fruit skin samples were collected for RNA extraction using a surgical blade. Total RNA was extracted from each sample with a RNeasy Mini Kit (Qiagen, Germany), and cDNA synthesis was performed using LunaScript RT SuperMix Kit (New England Biolabs, USA) following the manufacturer’s instructions. For amplicon sequencing, primer sets were developed for the coding region of MYB10 (cv Camarosa, maker-Fvb1-2-snap-gene-157.15-mRNA-1) using IDT’s PrimerQuest Software (Supplemental Data Set 11). The PCR products were checked in 3% (w/v) agarose gels and further purified for amplicon sequencing at GENEWIZ (South Plainfield, NJ). The sequencing reads from each WF and RF pool (10 samples per pool) were mapped to the MYB10 gene (maker-Fvb1-2-snap-gene-157.15-

mRNA-1) with Geneious v.11.0.5 using default values. The consensus sequences of the MYB10 coding region of white and red strawberry pools were aligned using T-coffee (Notredame et al., 2000).
Gene and Promoter Cloning and Sequence Homology Analysis

Gene and promoter amplification were performed using MyFi™ DNA Polymerase (Bioline). FvMYB10-gypsy and FvB1 FIN12 genomic fragments flanking the deleted region were amplified with the 5Prime PCR Extender System (5Prime) using the provided 10× Tuning Buffer following the manufacturer instructions. PCR products <8 kb were purified with a FavorPrep GEL/PCR Purification Kit (Favorgen). PCR products >8 kb were precipitated (30% [v/v] PEG8000: 30mM MgCl₂). Purified products were cloned into the pGEM-T Easy Vector Systems (Promega) for Sanger sequencing.

Multiple sequence alignment of MYB10 promoter fragments was performed using the Geneious algorithm (Gap open penalty = -30; Gap extension penalty = 0; 50 refinement iterations) and used for homology tree construction (Genetic distance model: Jukes-Cantor; Tree build method: Neighbor-Joining; Resampling method: Bootstrap 1000 replicates), both in Geneious 7.1.9 (Kearse et al., 2012).

F. vesca MYB10_sequence was chosen as outgroup as it belongs to a different species. Promoter sequence alignment and machine-readable tree files are provided as Supplemental Data Sets 12 and 13.

RNA-Seq Analysis, Differential Gene Expression, and Variant Calling

Total RNA was extracted from two biological replicates of ripe fruits from the seven selected octoploid accessions as described above. Each replicate consisted of a pool of 3 to 5 fruits collected throughout different days at the same time of day (Zeitgeber Time 7). The fruits were immediately frozen in liquid N₂ and stored at -80°C until processing. RNA quantity, quality, and integrity were determined based on absorbance ratios at 260 nm/280 and 260 nm/230 nm using a NanoDrop spectrophotometer (ND-1000 V3.5, NanoDrop Technologies, Inc.) and agarose gel electrophoresis.

Libraries were produced following standard procedures (Trapnell et al., 2012). For the reference genome (Edger et al., 2019), we used the latest assembled genome (F. vesca FvMYB10mo sequence was chosen as outgroup as it belongs to a different species. Promoter sequence alignment and machine-readable tree files are provided as Supplemental Data Sets 12 and 13.

The primer set targeting the 8-bp mutation was designed using IDT’s PrimerQuest Software and ordered from IDT (San Jose, California; Supplemental Data Set 11). PCR amplification was performed in a 5 µL reaction containing 0.5 µM of each primer set, 2× AccuStart™ II PCR ToughMix (Quantabo, Beverly, Massachusetts), 1× LC Green Plus HRM dye (BioFire, Salt Lake City, Utah), and 0.5 µL of DNA. The PCR conditions were as follows: an initial denaturation at 95°C for 5 min; 45 cycles of denaturation at 95°C for 10 s, annealing at 62°C for 10 s, and extension at 72°C for 20 s. After PCR amplification, the samples were heated to 95°C for 1 min and cooled to 40°C for 1 min. Melting curves were obtained by melting over the desired range (60° to 95°C) at a rate of 50 acquisitions per 1°C. The HRM assay was performed with the LightCycler 480 System II (Roche Life Science, Germany). The HRM data were analyzed using Melt Curve Genotyping and Gene Scanning Software (Roche Life Science, Germany). Analysis of HRM variants was based on differences in the shapes of the melting curves and Tm values.

SNP Genotyping using KASP

An SNP in FvMYB10-2 from position 20 from the initial ATG associated with white/red flesh color in the sequenced accessions was targeted for the KASP assay. To produce a subgenome-specific assay, a combination of four primers was used in each reaction: two common forward primers (ccm90 and ccm91) to account for a background-specific G/A SNP, and two allele-specific primers (ccm92 and ccm93), each carrying the standard FAM or HEX dye tails and the targeted SNP at the 3’ end. A single common primer is generally used for standard KASP assays, but the high degree of polymorphism in the targeted region made it necessary to include an extra common forward primer to account for a different, background-associated G/A SNP not linked to the white-fleshed trait. Reaction volumes of 5 µL for 384-well plates were used; 2.5 µL of MasterMix, 0.07 µL of assay/primer mix, and 2.5 µL of DNA (12.5 ng total) were used for genotyping. The final primer concentrations in the reaction mix were 210 nM for ccm90 and ccm91; 150 nM for ccm92; and 190 nM for ccm93. PCR was conducted in a BioRad CFX-384 instrument using the following protocol: An initial denaturation step at 94°C for 15 min, 12 touchdown cycles (94°C for 20 s, 65°C for 80 s, dropping 0.6 °C per cycle), 29 cycles (94°C for 20 s, 58°C for 80 s), and a final recycling for 2 cycles (94°C for 20 s, 60°C for 80 s). BioRad Software was used to estimate the final results.
Accession Numbers

F. vesca FIN12 accession sequencing data are stored at National Center for Biotechnology Information Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA655237.

The raw reads data from (1) illumina high-throughput sequencing of RF and WF DNA pools from F2 lines of the F. vesca RV660 × WV596 population and (2) RNA-seq from ripe fruits of white and red octoploid accessions were stored at the Sequence Read Archive (SRA) of the European Nucleotide Archive (ENA) under project reference number PRJEB38128. In the same project, Sanger sequences were stored under the following specific codes: (1) fmyb10-3 mutant alleles sequences from SE100, GER1, and GER2: LR862245, LR862246, and LR862247; (2) MYB10 promoter sequences from F. chiloensis ssp. chiloensis FC157 MYB10-1_pro, MYB10-3B_pro, MYB10-2_pro, and MYB10-3A_pro (LR862237-LR862239), F. chiloensis ssp. lucida USA2 MYB10-3A_pro, MYB10-1_pro, MYB10-3B_pro, and MYB10-2_pro (LR862240-LR862242), Fragaria × ananassa cv Senga Sengana MYB10-3A_pro and MYB10-1_pro/MYB10-3B_pro (LR862243-LR862244); and (3) The FvMYB10-gypsy LTR-TE sequence from fmyb10-2: LR8622681.

Supplemental Data

Supplemental Figure 1. The LTR-TE insertion in FvMYB10 cosegregates with the white fruit phenotype in the entire RV660xWF596 population.

Supplemental Figure 2. TEAC, soluble solid content (SSC), titratable acidity (TA), and L-ascorbic acid content of RF and WF F2 pools of the RV660 x WV596 population.

Supplemental Figure 3. F. vesca accession FIN12 has a large deletion affecting an ~100 kb region in Fvb1 that results in the loss of seven predicted genes including FvMYB10 (FvH4_1g22020).

Supplemental Figure 4. Fruit skin color in UF population 17.66 (FL 13.65-160 × FL 14.29-1).

Supplemental Figure 5. Fruit flesh and skin color phenotypes in the Hansabred SSxFcL population.

Supplemental Figure 6. Senga Sengana × F. chiloensis ssp. lucida linkage map and comparison to the physical positions in the F. vesca reference genome.

Supplemental Figure 7. Expression of cv Camarosa FaMYB10 homoeologs in different tissues and during fruit ripening.

Supplemental Figure 8. Sequence alignment of WT and mutant MYB10-2 in UF breeding population 17.66.

Supplemental Figure 9. Alignment of the MYB10 upstream region showing the SNP targeted for KASP genotyping.

Supplemental Figure 10. Fruit color scale used to phenotype octoploid accessions.

Supplemental Data Set 1. List of significant SNPs and INDELS in the RV660 × WV596 population.

Supplemental Data Set 2. Complete metabolite data.

Supplemental Data Set 3. List of Fragaria ssp. included in this study and description of mutations detected in MYB10.

Supplemental Data Set 4. QTLs detected for fruit color traits in the cv Senga Sengana × Fragaria chiloensis ssp. lucida F2 population by Kruskal-Wallis (KW) and restricted multiple QTL mapping (rMOM) analysis.

Supplemental Data Set 5. Polymorphisms at MYB10 homoeologs from F. chiloensis accessions and cv Senga Sengana vs. the cv Camarosa reference genome.


Supplemental Data Set 7. Predicted putative cis-elements specific to cv Camarosa FaMYB10-2_pro.

Supplemental Data Set 8. Predicted putative cis-elements specific to cv Camarosa FaMYB10-2_pro potentially significant in the context of fruit ripening and anthocyanin production.

Supplemental Data Set 9. Putative cis-elements predicted in FvMYB10_pro overlapping with those exclusively predicted in cv Camarosa FaMYB10-2_pro.


Supplemental Data Set 11. List of primers.

Supplemental Data Set 12. Sequence alignment file in fasta format of MYB10_pro variants.


ACKNOWLEDGMENTS

We thankfully acknowledge the computer resources and the technical support provided by the Plataforma Andaluza de Bioinformática of the University of Málaga. We thank Francisco J. Durán at IFAPA and the UF/IFAS strawberry breeding team for his excellent care of the strawberry germplasm. This work was supported by the Spanish Ministry of Science and Innovation and FEDER (grants RFPI2015-00011-00-00 and PID2019-111496RR-I00), IFAPA, FEDER funds (PRAVA.AVA2019.034), EC | European Research Council (ERC) (ERC-2014-STG 638134), and by the European Union’s Horizon 2020 research and innovation programme (GoodBerry; grant 679303). This study is also part of the joint research network SPIRED, which was funded by the German Federal Ministry of Education and Research (BMBF, grants FKZ 031A216A and B).

AUTHOR CONTRIBUTIONS

C.C. and I.A. planned and designed the experiments, analyzed the data, and wrote the article with input from all other authors.; C.C., performed molecular analyses, phylogenies, expression analysis by RT-qPCR, and transient overexpression by agro-infiltration of fruits; V.W., H.W., and K.O. developed the octoploid SSxFcL mapping population and performed the phenotypic analysis; R.R. and I.A. performed the genetic mapping and QTL analysis in the octoploid SSxFcL population; N.O. and P.M. performed fruit quality assessments; J.C.T., J.C., Z.L., and I.A. performed the genetic mapping in diploid F. vesca; N.C., M.A.H., and S.J.K. provided genomic data for the cultivated and wild octoploid accessions and participated in KASP marker development; J.G.V. and S.O. conducted the metabolomics data for the cultivated and wild octoploid accessions and participated in KASP marker development; J.G.V. and S.O. conducted the metabolomics analysis; C.M.-P., D.P., T.T., and T.H. provided genomic data for the wild F. vesca accessions; Y.O., C.R.B., and V.M.W. developed the UF breeding population 17.66 (FL 13.65-160 × FL 14.29-1).

Received June 22, 2020; revised August 13, 2020; accepted September 25, 2020; published September 30, 2020.
REFERENCES


Li, Y.-Y., Mao, K., Zhao, C., Zhao, X.-Y., Zhang, H.-L., Shu, H.-R., and Hao, Y.-J. (2012). MdCOPI ubiquitin E3 ligases interact with...


Allelic Variation of MYB10 Is the Major Force Controlling Natural Variation in Skin and Flesh Color in Strawberry (Fragaria spp.) Fruit

Cristina Castillejo, Veronika Waurich, Henning Wagner, Rubén Ramos, Nicolás Oiza, Pilar Muñoz, Juan C. Triviño, Julie Caruana, Zhongchi Liu, Nicolás Cobo, Michael A. Hardigan, Steven J. Knapp, José G. Vallarino, Sonia Osorio, Carmen Martín-Pizarro, David Posé, Tuomas Toivainen, Timo Hytönen, Youngjae Oh, Christopher R. Barbey, Vance M. Whitaker, Seonghee Lee, Klaus Olbricht, José F. Sánchez-Sevilla and Iraida Amaya

Plant Cell 2020;32;3723-3749; originally published online September 30, 2020;
DOI 10.1105/tpc.20.00474

This information is current as of March 7, 2021

Supplemental Data
/content/suppl/2020/10/01/tpc.20.00474.DC1.html
/content/suppl/2020/10/21/tpc.20.00474.DC2.html

References
This article cites 114 articles, 15 of which can be accessed free at:
/content/32/12/3723.full.html#ref-list-1

Permissions

eTOCs
Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts
Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information
Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY