Universitat Autònoma de Barcelona

Genetic characterization of the Japanese plum LG3-MYB10 region and development of molecular markers for fruit color

Arnau Fiol Garví

Doctoral thesis

Universitat Autònoma de Barcelona
Facultat de Biociències
Departament de Biologia Animal, Biologia Vegetal i Ecologia

Genetic characterization of the Japanese plum LG3-MYB10 region and development of molecular markers for fruit color

Dissertation presented by Arnau Fiol Garví for the degree of Doctor in Plant Biology and Biotechnology by Universitat Autònoma de Barcelona

Thesis director and tutor

Dr. Maria José Aranzana Civit

PhD candidate

Arnau Fiol Garví

Barcelona, February 2022

This thesis has been performed in the research group of Rosaceae Genetics and Genomics, from the research program of Plant and Animal Genomics at the Centre for Research in Agricultural Genomics (CRAG) - CSIC-IRTA-UAB-UB consortium. The PhD student, Arnau Fiol Garví, has been supported by an FPI grant (reference BES-2016-079060) funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future", which also funded an international stage at the Instituto de Investigaciones Agropecuarias (INIA), department of Biotechnology in the INIA-La Platina center (Santiago, Chile).

Dedicat al meu pare

Acknowledgments

Inclús abans de començar aquesta tesi, ja tenia ben clar que la bona ciència no és producte de l'esforç d'una sola persona sinó de moltes. Arribat al moment d'escriure aquestes paraules, reconec que no sabia aleshores fins a quin punt això era cert. Des de les persones que han col·laborat activament en aquest projecte, a les persones que m'han ofert la seva ajuda al laboratori i anàlisis, fins a les persones que em treien un somriure i m'animaven a seguir treballant. Per totes aquestes persones, el meu més sincer agraïment i les següents línies.

Voldria agrair primerament a la meva directora, Txosse, haver confiat en mi per aquesta tesi, però sobretot, per haver prioritzat sempre la meva formació com a científic, entrenant-me en la presa de decisions, dissenys experimentals, autonomia, esperit crític, comunicació oral i meticulosa escriptura científica. Gràcies per exercir no només de directora d'un projecte, sinó també de mentora científica. També agraïments al Pere, Amparo, Jordi, Montse i Marta per tots els consells, correccions i donar diferent punts de vista al moment d'exposar o voler presentar un paper. Agraeixo especialment al Ibo, perquè a més de lo anterior he entrat moltes (masses) vegades al seu despatx a preguntar-li coses del color i sempre m'ha rebut amb un somriure, al Werner perquè més allà de la seva participació al projecte també m'ha ensenyat molt en el disseny, ús i optimització de marcadors moleculars per la millora vegetal, i al Mourad perquè apart de ser qui em va ensenyar a fer extracció en placa quan jo era un pipiolo, és de les persones que em feia més amè el dia a dia (recorda que us esperem a Girona quan vulgueu venir amb la família!). Aprofito també per agrair a col·laboradors externs que també han ajudat molt: Igor, porqué aparte me recibiste muy bien en Chile; y Elena, porqué en el congreso de Praga me enseñaste lo que podía lograr el CRISPR-Cas9 enrichment y a la vista está el buen resultado que salió de todo esto! Gracias por tanto.

Voldria agrair a tots els tècnics que he tingut el plaer de compartir lab. A més de fer la feina més divertida, sou uns grans professionals i realment l'únic mur que separa l'ordre del caos absolut al lab. L'Àngel solucionador de tots els problemes (i el "dealer"), la Fuensi que sempre és la gran esperança per quan falla qualsevol experiment, la Vanessa i la Sara que sempre tenien protocols per tot per molt raro que fos... Gràcies! Vull agrair especialment a l'Esteve, perquè tot i estar amagat a la cantonada és un crack que m'ha ensenyat molt més del que pensa a ser eficient a la feina. I sobretot, sobretot a la Elena, per ser la meva mestra a *in vitro*, i per ser la meva (nostra) mare al laboratori, perquè sempre ajuda a tothom mostrant el seu afecte (i renya si ho mereixem). Gràcies

per tot! Y evidentemente tampoco me puedo olvidar de agradecer a mis otros muchísimos maestros del INIA: Humberto, Daniela, Ricardo, Marisol, Blanca, Caro, Philippe,… ¡Gracias por enseñarme tanto sobre un tema tan interesante como complicado, los transgénicos en *Prunus*! Lo que hacéis es pura magia. En especial agradecer a Carlos (Dr. P.), Pablo, Felipe y Micco por, además, recibirme tan bien los meses que pasé allá en Chile.

Agraïments també amb qui he tingut el plaer de treballar al dia a dia. Bea per la teva ajuda i interès, Federico pels teus anàlisis i tot el que m'has ensenyat de bioinformàtica (perdón por tantos cambios de gráfica). Sé que arribareu molt lluny al món de la ciència, ho valeu i mostreu cada dia. Al meu compipa paraguayo què dir, el meu germà (gran). Tot i treballar en projectes diferents hem tingut una sinergia molt productiva, espero poder seguir col·laborant amb tu en el futur. I també tenir més asados, jocs de taula i piques a la consola. També agrair als estudiants que han passat pel projecte: Sergio, Laura, Alejandro i Laia. Espero que l'experiència us servís tant com m'ha servit a mi. Agrair també els que han compartit amb mi l'experiència de ser estudiant, Núria i Naveen, que vàrem començar el doctorat a la vegada i inclòs vam viure junts, quina experiència! Miguel, Nathalia, Laura, Carlos, Lorena, Felipe, Lorena, Ari... i els que ja han arribat a la meta, Yu, Pol, Neus, Octavio... Gràcies per cada moment. Als estudiants del màster de quan començava al 2017/2018... Paula, Jose, molt grans! Carles, hem tingut la sort de compartir molt més camí i ets un gran amic, perdó per cada broma molestosa. Jorge, ets els verdader rei de la prunera, gràcies per l'ajuda amb el material vegetal quan més em feia falta, espero poder anar aviat a veure't a tu i els camps. Gràcies també als professionals de PLANASA, perquè gràcies al seu material i a un fenotipat acurat hem pogut trobar resultats aplicables al seu i molts altres programes de millora. També gracies als companys del IMIDA, CEBAS, IDV del Baix Llobregat i Joan París per l'ajuda prestada.

Finalment, voldria agrair a tots els amics que, tot i no estar en el meu dia a dia al IRTA-CRAG, han tingut a veure amb aquest treball. Als amics de sempre que s'han interessat per com m'anava la tesis i m'han perdonat les múltiples absències, com la Bistec Band (ho sento, espero poder-ho compensar) i els amics de la uni de Girona (amb el Dr. Olallo com el meu mirall i veu de l'experiència). I a la meva família, els més importants a la meva vida, pare, mare, germana... Gràcies pel vostre interès i amor, per ensenyar-me el valor de l'esforç i la feina ben feta. I sobretot gràcies a la Carla, la que més m'ha aguantat durant anys al dia a dia, compartit i celebrat amb mi cada bon resultat, animat quan he defallit i que sé que mai, mai em fallarà. A tots: moltes gràcies per ser el suport amb el qual en baso i recolzo dia a dia i en el futur. Arribo a escriure aquestes línies gràcies a vosaltres.

Index of contents

Index of figures, tables and data

Abbreviations

ABL: Advanced breeding lines

AFLP: Amplified fragment length polymorphism

bp: Base pair

CaMV: Cauliflower Mosaic Virus

Cas9: CRISPR associated protein 9

cDNA: Complementary DNA

Chr: Chromosome

cM: Centimorgan

CNV: Copy number variation

CRISPR: Clustered regularly interspaced short palindromic repeat

crRNA: CRISPR-RNA

CV: Commercial varieties

DNA: Deoxyribonucleic acid

dNTP: Deoxynucleotide triphosphate

EST: Expressed sequence tag

GBS: Genotyping by sequencing

GDR: Genome Database for Rosaceae

gRNA: Guide RNA

GWAS: Genome wide association studies

InDel: Insertion/Deletion

kb: Kilobase

LG: Linkage group

LOD: Logarithm of the odds

LTR: Long terminal repeat

MAB: Marker-assisted breeding

MAS: Marker-assisted selection

Mb: Megabase

MBW: MYB, bHLH and WD-repeat complex

mRNA: Messenger RNA

NB: Native Barcode

NCBI: National Center for Biotechnology Information

ONT: Oxford Nanopore Technologies

Pa: *Prunus avium* gene identifier

PAM: Protospacer adjacent motif

PAV: Presence/absence variation

PCR: Polymerase chain reaction

Pp: *Prunus persica* gene identifier

Ps: *Prunus salicina* gene identifier

QTL: Quantitative trait locus

QTLs: Quantitative trait loci

RAPD: Randomly amplified polymorphic DNA

RFLP: Restriction fragment length polymorphism

RNA: Ribonucleic acid

ROI: Region of interest

RT-PCR: Reverse transcription PCR

SLAF-Seq: Specific locus amplified fragment sequencing

SNP: Single nucleotide polymorphism

SSR: Simple sequence repeat

SV: Structural variant

Ta: Temperature of annealing

TF: Transcription factor

tracrRNA: Transactivating crRNA

VIGS: Virus-induced gene silencing

ABSTRACT

Abstract

Japanese plum (*P. salicina* and its hybrids) is a highly heterozygous fruit tree originated by the interspecific hybridization of more than eleven diploid plum species. This event might explain that, within the Rosaceae family, it is one of the crops with the widest fruit color variability. Despite the economic importance of Japanese plum fruits in the fresh market, its marker-assisted breeding is still in its infancy. This applies for the fruit color trait, which influences the consumer acceptance and, consequently, is included within the key breeding objectives. The red to black hues are caused by the accumulation of anthocyanins, pigments synthesized from the flavonoid pathway that have nutraceutical properties. The *MYB10* genes are candidates for anthocyanin accumulation in plant organs, and QTLs for red fruit color in *Prunus* have been mapped in linkage group 3 (LG3) in a region containing three gene copies in peach: *PpMYB10.1*, *PpMYB10.2* and *PpMYB10.3*. The objective of this thesis is to study the role of the LG3-MYB10 genes in the regulation of fruit color in Japanese plum and provide efficient molecular markers for marker-assisted breeding.

To study the allelic variability of the Japanese plum LG3-*MYB10* genes (*PsMYB10*), we used a set of primers designed in conserved sites of the peach *PpMYB10* genes by using its genome assembly (Chapter 1). We found one allele (a356) highly associated with the fruit skin color, but none for the flesh. By means of progeny segregation and allele cloning we assigned the alleles into haplotypes and identified a *PsMYB10.1* triplication (Chapter 2). Allele a356 was a *PpMYB10.1* homolog, with allele a470 segregating as an alternative allele, recessive for the locus (*PsMYB10.1a*). To sequence the LG3-*PsMYB10* region, we used a novel long-read targeted sequencing technology, the CRISPR-Cas9 enrichment followed by Oxford Nanopore Technology sequencing (Chapter 3). Despite the target region was large, highly variable and with several duplicated genes, with our strategy we could extract polymorphisms from the genome alignments against reference genomes and from *de novo* aligned sequences. All these previous results served to find association of the LG3-*MYB10* region with the red flesh color (Chapter 4), identifying a retrotransposon in the *PsMYB10.2* gene correlating with its expression. The gene function was validated by setting up a protocol for its transient overexpression in Japanese plum fruits.

The main outputs of this thesis are 1) the development of efficient molecular markers for the fruit skin and flesh colors, 2) the validation of a strategy for the CRISPR-Cas9 enrichment of large and highly variable regions with gene duplications in a pool of samples, applicable in other plant and animal studies, and 3) a protocol for the transient gene overexpression in Japanese plum fruits, applicable for the validation of other candidate genes in the crop.

Resumen

El ciruelo japonés (*P. salicina* e híbridos) es un árbol frutal altamente heterocigoto, originado a partir de la hibridación interespecífica de más de once especies diploides de ciruelo. Este evento explicaría que, dentro de la familia de las Rosáceas, es uno de los cultivos con más variabilidad en el color de los frutos. A pesar de la importancia de sus frutos en el mercado, la mejora asistida por marcadores en ciruelo japonés está todavía en fase inicial. Esto se aplica al color de fruto, que influye en la aceptación del consumidor y, en consecuencia, es uno de los objetivos clave de mejora. Las tonalidades del rojo al negro son debidas a la acumulación de antocianos, pigmentos sintetizados desde la ruta de los flavonoides que tienen propiedades nutracéuticas. Los genes *MYB10* son candidatos para su acumulación en los órganos de las plantas, y QTLs para el color de fruto en *Prunus* se han mapeado en el grupo de ligamiento 3 (LG3) en una región que contiene tres copias génicas en melocotonero: *PpMYB10.1*, *PpMYB10.2* y *PpMYB10.3*. El objetivo de esta tesis es estudiar el rol de los genes LG3-*MYB10* en la regulación del color de fruto del ciruelo japonés y proporcionar marcadores moleculares eficientes para su mejora asistida por marcadores.

Para estudiar la variabilidad alélica de los genes LG3-*MYB10* (*PsMYB10*) en ciruelo japonés, usamos un conjunto de cebadores diseñados en sitios conservados de los genes *PpMYB10* de melocotonero usando su genoma (Capítulo 1). Encontramos un alelo (a356) altamente asociado con el color de la piel, pero ninguno para la carne. A partir de la segregación y clonaje de los alelos, pudimos asignarlos en haplotipos e identificar una triplicación del gen *PsMYB10.1* (Capítulo 2). El alelo a356 era homólogo del *PpMYB10.1* y el a470 segregaba como alternativo, recesivo para el locus (*PsMYB10.1a*). Para secuenciar la región LG3-*PsMYB10* usamos una reciente tecnología para la secuenciación de regiones concretas de genoma con lecturas largas: el enriquecimiento de secuencias por CRISPR-Cas9 seguido de la secuenciación con tecnología Oxford Nanopore (Capítulo 3). A pesar que la región objetivo era larga, altamente variable y con duplicaciones de varios genes, la estrategia permitió extraer polimorfismos a partir de los alineamientos con las secuencias genómicas de referencia y el alineamiento de las secuencias *de novo*. Estos resultados sirvieron para encontrar asociación de la región LG3-*MYB10* con el color rojo de la carne (Capítulo 4), identificando un retrotransposón en el gen *PsMYB10.2* correlacionado con su expresión. Su función se validó poniendo a punto un protocolo para su sobreexpresión en ciruelas japonesas.

Esta tesis proporciona: 1) marcadores moleculares eficientes para el color de fruto en la piel y carne, 2) la validación de una estrategia para el enriquecimiento de secuencias por CRISPR-Cas9 en regiones altamente variables con duplicaciones génicas en un conjunto de muestras, aplicable a otros estudios en plantas y animales, y 3) un protocolo para la sobreexpresión transitoria de genes en frutos de ciruelo japonés, aplicable para la validación de otros genes candidatos en el cultivo.

Resum

La prunera japonesa (*P. salicina* i híbrids) és un arbre fruiter altament heterozigot originat a partir de la hibridació interespecífica de més d'onze espècies diploides de pruna. Aquest esdeveniment explicaria que, dintre de la família de les Rosàcies, és un dels cultius amb més variabilitat en el color dels seus fruits. Malgrat la importància dels seus fruits en el mercat, la millora assistida per marcadors en prunera japonesa està encara en fase inicial. Això s'aplica al color del fruit, que té influència en l'acceptació del consumidor i, en conseqüència, és un dels objectius clau de millora. Les tonalitats del vermell al negre es deuen a l'acumulació d'antocians, pigments sintetitzats des de la ruta dels flavonoides que tenen propietats nutracèutiques. Els gens *MYB10* són candidats per la seva acumulació en els òrgans de les plantes, i QTLs pel color de fruit en *Prunus* s'han mapat en el grup de lligament 3 (LG3) en una regió que conté tres còpies gèniques en presseguer: *PpMYB10.1*, *PpMYB10.2* i *PpMYB10.3*. L'objectiu d'aquesta tesis és estudiar el rol dels gens LG3-*MYB10* en la regulació del color en fruits de prunera japonesa i proporcionar marcadors molecular eficients per la seva millora assistida per marcadors.

Per estudiar la variabilitat al·lèlica dels gens LG3-*MYB10* (*PsMYB10*) en prunera japonesa, vàrem fer servir un conjunt d'encebadors dissenyats en llocs conservats dels gens *PpMYB10* de presseguer fent servir el seu genoma (Capítol 1). Vàrem trobar un al·lel (a356) altament associat al color de pell, però cap per la carn. A partir de la segregació i clonatge dels al·lels, vam poder assignar-los a haplotips i identificar una triplicació del gen *PsMYB10.1* (Capítol 2). L'al·lel a356 era homòleg del *PpMYB10.1* i el a470 segregava com alternatiu, recessiu pel locus (*PsMYB10.1a*). Per seqüenciar la regió LG3-*MYB10* vàrem fer servir una tecnologia recent per la seqüenciació de regions concretes del genoma amb lectures llargues: L'enriquiment de seqüències per CRISPR-Cas9 seguit de la seqüenciació per tecnologia Oxford Nanopore (Capítol 3). Malgrat que la regió objectiu era llarga, altament variable i amb duplicacions de varis gens, amb aquesta estratègia vam podem extraure polimorfismes a partir dels alineaments amb les seqüències genòmiques de referència i l'alineament de les seqüències *de novo*. Tots aquests resultats van servir per trobar associació de la regió LG3-*MYB10* amb el color vermell de la carn (Capítol 4), identificant un retrotransposó en el gen *PsMYB10.2* correlacionat amb la seva expressió. La seva funció es va validar posant a punt un protocol per la seva sobreexpressió en fruits de prunera japonesa.

Aquesta tesis proporciona: 1) marcadors moleculars eficients pel color de fruit de la pell i la carn, 2) la validació d'una estratègia per l'enriquiment de seqüències per CRISPR-Cas9 a regions altament variables amb duplicacions gèniques en un conjunt de mostres, aplicable a altres estudis a plantes i animals, i 3) un protocol per la sobreexpressió transitòria de gens en fruits de prunera japonesa, aplicable per la validació d'altres gen candidats al cultiu.

MAIN INTRODUCTION

SECTION 1: JAPANESE PLUM

1.1 Taxonomy

Japanese plum (*Prunus salicina* Lindley) is a diploid fruit tree species that belongs to the *Prunus* genus within the Rosaceae family, widely cultivated for the fresh consumption of its fruits. Other common names are Chinese plum or Asian plum. It is classified as follows:

Kingdom: Plantae

Phylum: Tracheophyta

Order: Rosales

Family: Rosaceae

Subfamily: Amygdaloideae

Genus: *Prunus L.*

Subgenus: *Prunophora*

Section: *Euprunus*

Species: *Prunus salicina*

The Rosaceae is a family in the order of Rosales with around 3,000 species, some of them are crops of high economic importance with improved qualities after extensive human domestication such as apples, pears, strawberries, peaches, plums, cherries and raspberries; and some others have ornamental value such as roses, crabapples, rowans and Japanese apricots (Xiang et al., 2017). The Rosaceae members can be classified into three subfamilies (Potter et al., 2007): Rosoideae (around 2,000 species), Amygdaloideae (around 1,000) and Dryadoideae (less than 30).

The *Prunus* genus belongs to the Amygdaloideae subfamily and includes more than 200 species of deciduous and evergreen trees and shrubs, some of them widely cultivated for either their edible fruits or for ornamental purposes. Fruits from *Prunus* are commonly known as stone fruits due to the presence of a pit. The name of the genus comes from the Latin name of plum, which was the

first species in it (Okie, 2008). The taxonomy of the genus has been controversial specially on its delimitation (Lee and Wen, 2001), being the most widely accepted classification the proposed by Rehder (1940) and complemented by Ingram (1948) consisting in six subgenera: *Amygdalus* (peaches and almonds), *Prunophora* (plums and apricots), *Cerasus* (cherries), *Laurocerasus* (evergreen laurel-cherries), *Padus* (deciduous bird-cherries) and *Lithocerasus* (flowering sand cherries).

The *Prunophora* subgenus is divided into sections (Topp et al., 2012) including:

- *Euprunus*: Eurasian plums, includes Japanese plum (*P. salicina*), European plum (*P. domestica*), cherry plum (*P. cerasifera*), Simon plum (*P. simonii*) and Blackthorn (*P. spinosa*).
- *Prunocerasus*: Plums originated in America, includes American plum (*P. americana*), Pacific plum (*P. subcordata*) and Chickasaw plum (*P. angustifolia*).
- *Armeniaca*: Includes apricots, such as the widely cultivated apricot (*P. armeniaca*), Japanese apricot (*P. mume*) and Siberian apricot (*P. sibirica*).

1.2 Origin and distribution of Japanese plum

Prunus salicina was originated around the 300 BC in the Yangtze River basin (Yoshida, 1987) in China, where wild populations are still found in Hubei and Yunnan areas (Okie and Weinberger, 1996). It was introduced early into Japan, where fruit stones have been found dating back to 200 BC and its cultivation was first mentioned in 500 AD (Yoshida, 1987). However, they were mainly used in gardens and improvement efforts were not made until the last few centuries (Okie and Hancock, 2008). It was in 1870 when some of these improved Japanese varieties were imported to California (USA) and some of them were widely distributed, such as 'Kelsey', which is still cultivated nowadays (Hartmann and Neumüller, 2009). This import event conferred the common name to the Japanese plum. Few years later, Luther Burbank imported and incorporated Japanese cultivars and others of Chinese origin such as *P. simonii* to his breeding program, and then intercrossed them including plums of American origin to generate cultivars with enhanced fruit quality and adaptability. In these hybridizations, *P. salicina* positively contributed to the fruit size, flavor and keeping ability, *P. simonii* to the flesh firmness and acidity, and American species provided disease resistance, tough skin and aromatic quality (Okie and Hancock, 2008). The genetic relationship of many cultivars was stablished with RAPD markers and revealed that a specie from Europe, *P. cerasifera*, also had an important

contribution on their gene pool (Boonprakob et al., 2001). The fruits from these improved cultivars were more suitable for long distance shipping than most cultivars from the time, and became the basis of the plum shipping industry worldwide (Okie and Ramming, 1999). These cultivars, most of them multispecies hybrids, were cloned and tested around the world and some are still grown nowadays; others that not performed so well as in California were further crossed with local diploid plums to enhance their adaptability, increasing the hybridization degree of the crop (Okie and Hancock, 2008).

Today, Japanese plum is distributed in temperate regions worldwide, being the most cultivated varieties the obtained by Burbank or others derived from them (Okie and Ramming, 1999). Burbank used at least 11 plum species in the breeding of the crop (Karp, 2015), and so now Japanese plum does not refer to a single species, but to all the hybrid cultivars derived from the interspecific crosses in which the genome and botanical characteristics of *P. salicina* predominate.

1.3 Botanical characteristics

Japanese plums are small to medium sized trees with rounded or columnar crowns that can reach 8 to 10 meters in height (Okie, 2008). The size of the leaves greatly varies depending on the vigor of the tree and ranges between 5 to 10 cm in length and 3 to 6 cm in width, with crenate or serrate margins and a shape between elliptic to obovate, ending with an acute to acuminate tip (Okie, 2008). Their leaves are deciduous and fall in Autumn, and to be released from dormancy and stimulate leafing and flowering the chilling and heat requirements of the tree have to be fulfilled. The chilling requirement greatly varies among Japanese plum cultivars (334 to 987 chilling units) and is determinant for the differences observed in their flowering time (Ruiz et al., 2018). Auxiliary buds grow on summer at the base of the leaves (Okie, 2008).

Typically, one or two flower buds grow flanking a vegetative bud, each of them with 3 to 6 flowers depending on the cultivar, creating large blooms of white color which can fade to pink before falling. The flowers have five sepals and five petals which, once open, form a cup-shaped corolla. The pistil has a glabrous ovary and, when the flowers are at full bloom, the stigma is slightly beyond the 20 to 30 anthers (Okie, 2008). The percentage of flowers that develop into fruits (5-14%) is low compared with other *Prunus* species (>18%), but the high number of flowers ensures a good fruit yield per tree (Guerra and Rodrigo, 2015). Due to a long juvenile stage, the seedlings take 3 to 4 years to begin

fruiting. Most cultivars exhibit gametophytic self-incompatibility and to achieve an optimal production they require a pollen donor: a cultivar with different self-incompatibility alleles, synchronized with the flowering time of the receptor.

More than 15 different *Prunus* species have been reported in the hybridization of Japanese plum, which have conferred the high genetic diversity of the crop (Guerrero et al., 2021). In general, *Prunus salicina* has shown a high degree of cross-compatibility with members of the *Prunophora* subgenus. It has often been crossed with diploid members of *Euprunus* and *Prunocerasus* sections and generated hybrids that were fertile and productive; hybrids generated with *Armeniaca* members have generally presented low fruit setting, while hybrids generated with members outside the subgenus, such as peaches, were also reported but were usually sterile (Okie and Ramming, 1999; Okie, 2008). The European plums (*P. domestica*, hexaploid) are hybrids derived from an interspecific cross between a myrobalan plum (*P. cerasifera*, diploid) and a blackthorn plum (*P. spinosa*, tetraploid, also known as sloe) that might have been itself an interspecific hybrid originated from a myrobalan plum (Zhebentyayeva et al., 2019). Both the European plum and the blackthorn plum are not cross-compatible with Japanese plum, despite being from the same taxonomic section, due to their different ploidy level.

Like in all *Prunus*, the fruits are drupes due to the presence of a hardened endocarp (the pit) surrounding a single seed. Japanese plum fruits present a fleshy mesocarp usually attached to the stone (clingstone), a waxy bloom over the skin, a suture line and a pit flatter that those in cherries (Okie, 2008). Most varieties produce climacteric fruits, a few are categorized as suppressed-climacteric (limited ethylene synthesis but sensitive to an exogenous source), and one bud-sport mutant was described as non-climacteric (Minas et al., 2015). The fruit color shows a wide range of tonalities: the skin can have black, purple, red, orange, yellow or green hues; the flesh can have red, yellow, orange or green colors (Topp et al., 2012). The wax bloom, a powdery white cover over the fruit, can alter the perceived color on the skin, as well as the presence of a transparent yellow skin over red flesh (Okie, 2008).

1.4 Plum production worldwide and nutraceutical value

The edible parts of the plum fruit are the fleshy mesocarp and the epicarp tissues such as in peaches, cherries and apricots; in other *Prunus* members such as almonds the seed is the part used for

consumption. Peaches lead the *Prunus* fruit world production with 26 million tonnes (Mt), followed by plums with 13 Mt and apricots with 4 Mt; however, plums lead the total area harvested with 2.7 million hectares (Mha), followed by almonds (2.1 Mha) and peaches (1.5 Mha) (FAOSTAT, 2019). Their production shows an increasing tendency over the years, especially for peach and plums which have increased 5 and 2-fold their total production, respectively, from the first record on the year 1961 (**Figure I.1A**).

The plum production is focused on Asia (65%), followed by Europe (23%), America (8.5%), Africa (3.4%) and Oceania (0.1%) (FAOSTAT, 2019). The main producer country is China with 55.5% of the global production, which is 10 times more than the second producer country, Romania with 5.5%. The top 15 plum world producers concentrate 88.2% of the global production (**Figure I.1B**).

The former statistics for plums include not only Japanese plums but all plums and sloes in the *Euprunus* and *Prunocerasus* sections. The most cultivated crops on this group are the Japanese and the European plums. Almost all of the plum production from China, the top world producer, is from Japanese plums. Other countries with an important Japanese plum production are USA (mostly in California), Chile, Argentina and Mediterranean bordering countries such as Italy, France and Spain (Okie and Hancock, 2008; FAOSTAT, 2019). European plums are generally processed through drying, canning or used to produce juices, jams, jellies and alcoholic beverages; while Japanese plums are mainly used for fresh consumption and, less frequently, dried (Topp et al., 2012). The Japanese plum production is between May and October in the northern hemisphere (Okie and Ramming, 1999).

Japanese plum fruits have a mean of 15% sugar content mainly due to glucose, 1.2% of dietary fiber and an energy value of 255 kJ in 100 g of flesh weight (Lozano et al., 2009). They are rich in antioxidants such as phenolic compounds, vitamin C and carotenoids. These are more concentrated in the skin than in the flesh, being the total phenolic compounds the most correlated to the total antioxidant activity (Gil et al., 2002). Antioxidants are of nutraceutical interest, as their incorporation to the diet has been associated to health-promoting effects such as the prevention of

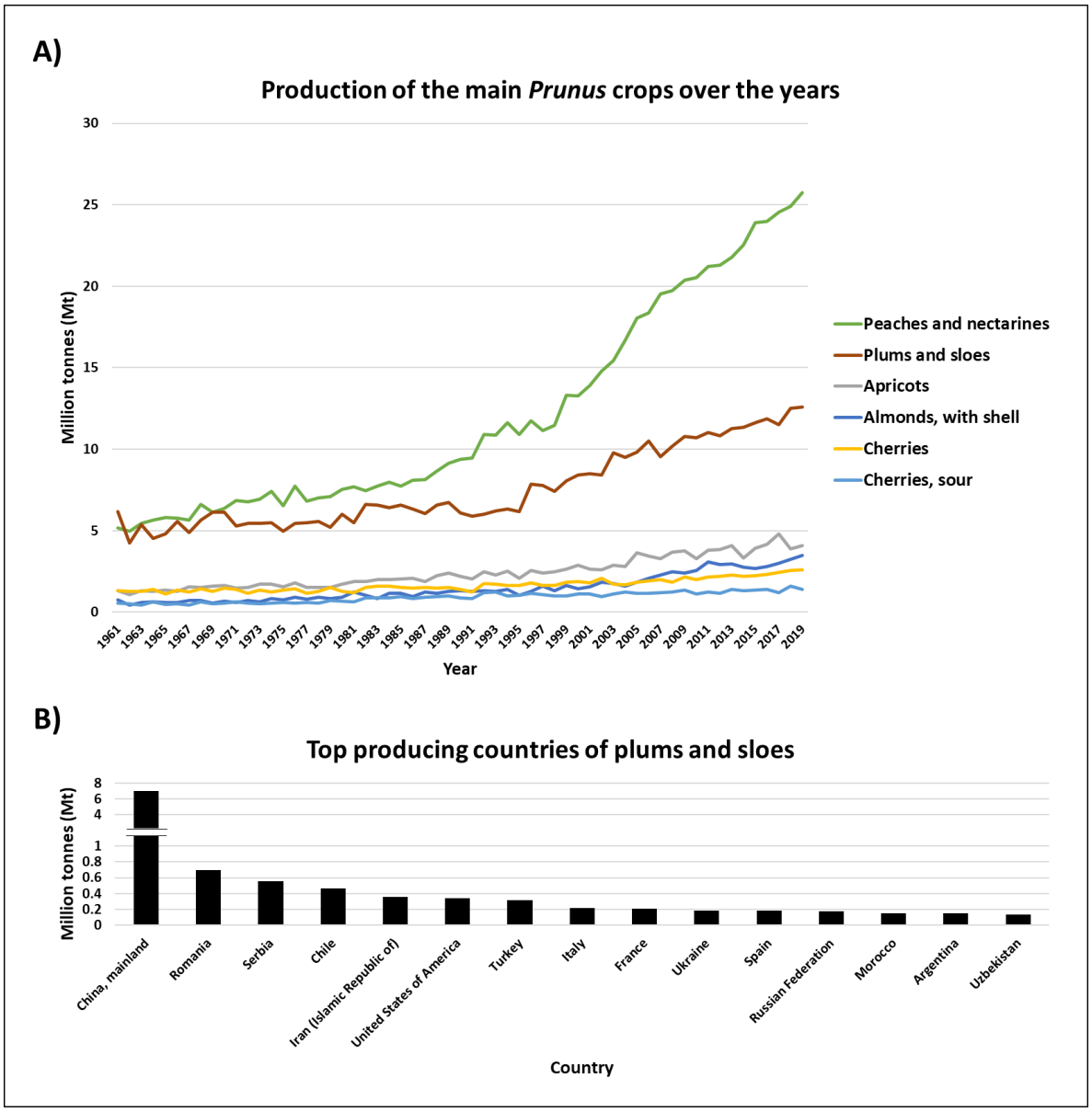


Figure I.1. A) Production of the main *Prunus* crops using all year reports from FAOSTAT (1961-2019). **B)** The 15 top producing countries of plums and sloes, appearing in descending order by their total production. The y-axis changes of scale after the first million tonne (FAOSTAT, 2019).

cardiovascular diseases, certain types of cancer, diabetes, obesity and other benefits such as the improvement of visual health and antimicrobial and neurological protection (Gil et al., 2002; Khoo et al., 2017). Anthocyanins are phenolic compounds that, aside from contributing to this list of health-promoting effects, contribute to the fruit appearance and to the consumer acceptance because they provide the diversity of red to purple or almost black hues of the fruits. The breeding of several successful varieties highly productive and with dark fruit skin that concealed the bruises on it provided that, nowadays, the most cultivated varieties on Occident have dark or red skin (Okie

and Ramming, 1999; Topp et al., 2012), which favors the intake of anthocyanin health beneficial compounds.

1.5 Genetic and genomic resources

Japanese plum has a diploid genome with 8 chromosomes like most species in the *Prunus* genus such as peach, apricot, sweet cherry and almond (2n = 2x = 16). The most genetically characterized species in the genus is the peach, which serves as the model for genetic and genomic studies in *Prunus* species (Byrne et al., 2012).

The breeding history of the Japanese plum crop is characterized by three main genetic bottlenecks that occurred along its domestication (Topp et al., 2012). Despite this, the outcrossing with other species and the self-incompatibility trait has resulted in a high genetic variability and heterozygosity of the crop, much higher than those values obtained from other main crops of the genus (Byrne, 1990; Carrasco et al., 2012; Guerrero et al., 2021). This heterozygosity also means that large seedling families are required for cultivar development (Topp et al., 2012), which can be aided by marker-assisted selection of traits of interest (detailed in Section 3: Molecular Breeding).

Japanese plum genetics have been less studied compared to other crops, mainly due to a reduced number of breeding programs and to the self-incompatibility trait that prevents selfing and determines the parents used in the development of backcross populations (Okie and Hancock, 2008). To date, only few genetic maps exist for the crop, all generated from F1 populations (**Table I.1**). Genetic maps are useful genetic resources to understand genome conformation and allow evolutionary and comparative studies, and also for the genome mapping of QTLs (quantitative trait loci) linked to traits of agronomic interest. Only the maps from '98–99' × 'Angeleno' have been successfully used for QTL mapping, identifying genomic regions involved in fruit quality, post-harvest and phenology parameters (Salazar et al., 2017, 2020; Valderrama-Soto et al., 2021).

Table I.1. Summary of the available genetic maps in Japanese plum.

Population (individuals)	Genotyping method	Map	Markers	Map length [markers*] (cM)	LGs	QTLs mapped	Reference
'Chatard' × 'Santa Rosa' (88)	AFLP	Chatard	56 AFLPs	905.5 [16.2]	11	No	Vieira et al. (2005)
		Santa Rosa	84 AFLPs	1349.6 [16.1]	14	No	
'98–99' × 'Angeleno' (153)	GBS	98–99	479 SNPs	688.8 [1.5]	8	Yes	Salazar et al. (2017)
		Angeleno	502 SNPs	647.0 [1.3]	8	Yes	
'Angeleno' × 'Aurora' (137)	GBS	Angeleno	714 SNPs	578 [0.81]	8	No	Carrasco et al. (2018)
		Aurora	320 SNPs	472 [1.48]	8	No	
		Consensus	732 SNPs	617 [0.96]	8	No	
'98–99' × 'Angeleno' (151)	GBS and SSRs	98–99	554 SNPs 19 SSRs	557.32 [0.97]	8	Yes	Salazar et al. (2020)
		Angeleno	654 SNPs 21 SSRs	576.51 [0.85]	8	Yes	
'09-06' × 'Fortune' (114)	SLAF-Seq	Consensus	720 SNPs	869.9 [1.21]	8	No	Zhang et al. (2020)

**Mean distance between the markers mapped, in cM.*

Contrarily to other Rosaceae crops such as apple, pear, peach and cherries that count with high density SNP (single nucleotide polymorphism) arrays testing thousands of polymorphic positions (Chagné et al., 2012; Peace et al., 2012; Verde et al., 2012; Li et al., 2019), Japanese plum lacks a specific high-throughput genotyping platform. Thus, in modern studies the genetic characterization in Japanese plum varieties and progenies is usually done with SSR (simple sequence repeat, also known as microsatellite) markers from *Prunus* (Carrasco et al., 2012; González et al., 2016b; Acuña et al., 2019; Ruiz et al., 2019; Guerrero et al., 2021), which are highly transferable between members of the genus (Mnejja et al., 2010), or by the sequencing of genomic libraries with Illumina technology. For example, the '98–99' × 'Angeleno' map from Salazar et al. (2017) was improved by adding SSRs markers and SNPs derived from mapping the GBS (genotyping by sequencing) Illumina reads to the newest version of the peach genome (Verde et al., 2017) previous to the SNP call.

The lack of a reference genome for the crop and the high synteny between *Prunus* species (Aranzana et al., 2019) has forced the utilization of the peach high-quality genome for Japanese plum genetic

and genomic studies. Examples can be found in the mapping studies relying in GBS and SLAF-Seq (specific locus amplified fragment sequencing) genotyping included in **Table I.1**, as well as in other studies (Marti et al., 2018; Salazar et al., 2019; Wei et al., 2021). For RNA-seq experiments, the *de novo* transcriptome assembly method has been preferred over the alignment to the peach reference genome (Jo et al., 2015; Fang et al., 2016; González et al., 2016a).

It has been only recently that three genome assemblies were generated and made available to the scientific community (**Table I.2**). Currently, the *Prunus salicina* Sanyueli Genome v1.0 (Liu et al., 2020) has a second version (v2.0), in which the only improvement consists in the renaming of the scaffolds to match the name of the syntenic chromosomes from *Prunus*. A genome assembly from the same cultivar has been developed by another research group (Fang et al., 2022). The other available genome assembly is for 'Zhongli No. 6' cultivar (Huang et al., 2021).

Table I.2. Genome assemblies for Japanese plum, available in the Genome Database for Rosaceae (GDR) portal (https://www.rosaceae.org).

Genome assembly	Sequencing method	Genome size [N50] (Mb)	Sequences scaffolded in the 8 LGs	Reference
Prunus salicina Sanyueli Genome v1.0[1]	Illumina, PacBio, Hi-C	284.2 [1.8]	96.56%	Liu et al. (2020)
Prunus salicina Zhongli No. 6 Genome v1.0	Nanopore, Hi-C	318.6 [2.3]	99%	Huang et al. (2021)
Prunus salicina Sanyueli FAAS Genome v1.0	Illumina, PacBio	282.4 [1.37]	99%	Fang et al. (2022)

[1]*Version v2.0 differs from v1.0 in the renaming of the scaffolds in consonance with other Prunus species*.

SECTION 2: CRISPR-CAS9 ENRICHMENT SEQUENCING

The CRISPR-Cas9 enrichment is a novel DNA sequencing methodology that allows, by the use of CRISPR-Cas9 technology, the target sequencing of specific genome regions with third-generation sequencing platforms such as MinION (Oxford Nanopore Technologies) or PacBio SMRT (Pacific Biosciences).

The third-generation sequencing can produce long-reads with 13 kb to 20 kb of average (Tyson et al., 2018), and even single reads over 2 Mb have been obtained (Payne et al., 2019). Compared to the short-reads generated by the second-generation platforms (up to 600 bp), long-reads have an improved detection of structural variants (SVs), mapping certainty, identification of transcript

isoforms and the generation of *de novo* assemblies (Amarasinghe et al., 2020). The main drawbacks are that long-reads require higher computational costs and that they are less accurate in terms of sequencing errors, although there has been an improvement over the years and the error rate has been reduced to <5% (Amarasinghe et al., 2020).

Target sequencing methods have been extensively used in second-generation sequencing platforms to obtain, with a reduced sequencing and data management cost, greater sequencing depths in selected genomic regions at the expenses of neglecting data from the rest of the genome (Bewicke-Copley et al., 2019). Until recently, no effective protocol for the long-read targeted sequencing of regions of interest existed.

The CRISPR-Cas9 enrichment sequencing was first described by Gilpatrick et al. (2020), accomplishing the preferential ligation of the sequencing adaptors to the region of interest (ROI) by using CRISPR (clustered regularly interspaced short palindromic repeat) technology. To construct the sequencing library, the methodology requires the extracted DNA to be dephosphorylated (**Figure I.2**). Then, specific CRISPR RNAs (crRNAs) and transactivating crRNAs (tracrRNA) are assembled into guide RNAs (gRNAs), which direct the Cas9 enzyme to specifically cleave the ROI by aiming the two flanking protospacer adjacent motifs (PAM). This expose 5' phosphorylated ends at the cut sites, which later with the d(A)-tailing enables the sequencing adaptors to be preferentially ligated to the fresh cleaved ends. Then it proceeds to the sequencing of the DNA library, which will be sequence-enriched for the ROI without requiring amplification steps.

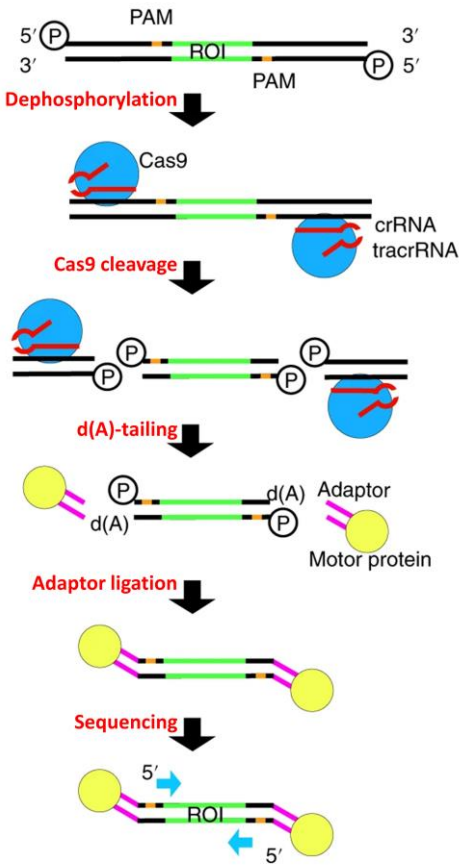


Figure I.2. The schematic representation of the DNA library preparation required for the CRISPR-Cas9 enrichment sequencing, to allow the long-read targeted sequencing of regions of interest (ROI). Figure edited from Gilpatrick et al. (2020).

Gilpatrick et al. (2020) showed that this procedure allowed the pooling of several crRNAs for the multiplex sequencing of 10 loci, of between 12 kb to 24 kb, on the human genome. The high depth obtained could overcome the error rates of the MinION sequencing and confidently identify SNPs, SVs and, as it sequences the native strands without amplification steps, identify differential methylation levels in the regions tested.

This methodology has already been validated in plants, allowing the sequencing of an 8 kb region containing the apple *MYB10* gene and a polymorphism that was difficult to identify with short-read sequencing methods (López-Girona et al., 2020), consisting in a minisatellite motif associated to the increased gene expression in apples with red flesh.

SECTION 3: MOLECULAR BREEDING

3.1 Plant breeding and objectives in plum

Plant breeding consist in the obtention of new cultivars that have a combination of several desired characteristics, obtained either by their segregation from selected progenitors, somatic mutations, induced mutagenesis or genetic transformation.

One of the main determinants for consumer choice regarding fruit and vegetables is their freshness (Massaglia et al., 2019), which firstly takes into account the visual inspection and later the organoleptic qualities. Visual traits have an effect on the decision to purchase the fruit, while optimal organoleptic qualities can drive consumers to purchase the fruit again. Other factors such as nutraceutical properties can also be important (Topp et al., 2012). The breeding of new cultivars seeks to improve the quality of the fruits to meet the consumers' demands, but also the crop yield, adaptability, in-field management, post-harvest behavior and the period of fruit seasonality; all which turns into a more profitable inversion for the producers and the market chain.

In plums there are twelve main fruit characters of interest, listed not by their importance but in the order in which they are usually evaluated by the breeders: Maturity date, fruit size and shape, productivity, fruit color (skin, ground and flesh), fruit firmness, freeness of the stone, fruit texture perception, fruit organoleptic quality and tree and fruit disease resistance (Okie and Hancock, 2008). All of them are regulated at the gene level in some degree and have high variability between genotypes. Their relevance is detailed below:

- *Maturity date*

  A single cultivar at the same location produces fruits that are usually harvested over 1-2 weeks, meaning that to ensure production for a 6 month window, 12 to 24 cultivars with different maturation times are required (Topp et al., 2012). This number can be reduced by an optimal post-harvest behavior of the fruit, which is desired to keep the recollected fruits for longer time. This also allows the export of fruits to opposite hemispheres, for they differ in the climatic season. An optimal post-harvest behavior is especially interesting in the latest varieties, because they can extend the season of fruit availability in the market. Early blooming cultivars can suffer spring frosts and diminish their production (Okie and Hancock, 2008), which has to be taken into account when selecting the cultivar to plant in the field. The maturity date is quantitatively inherited (Weinberger and Thompson, 1962), and some cultivars originated from bud sport mutations have been obtained showing delayed ripening times compared to their source cultivar (Minas et al., 2015).

- *Fruit size and shape*

  Large symmetrical fruits are often preferred to reduce costs derived from harvesting and packaging, to avoid their damage in the process (Topp et al., 2012) and for their increased market price. The preferred shapes are round, oblate, ovate or heart-shaped, without pointed tips that can be easily damaged. The fruit size is quantitatively inherited, and for the shape multiple factors are involved and neither the preferred round or ovate shapes are inherited in a dominant way (Weinberger and Thompson, 1962).

- *Productivity*

  High fruit productivity is essential for the successfulness of a new plant variety and takes into account the yield (weight/area) and the fruiting regularity (Goldschmidt, 2013). Genetic but also environmental and field management factors are involved in Japanese plum productivity: Tree vigor, impact of diseases, fruit hardiness, tree architecture, pruning and others (Okie and Hancock, 2008). Self-incompatible cultivars require the nearby planting of trees from a cultivar with compatible pollen to ensure fruit productivity in the orchard (Guerra and Rodrigo, 2015), and while the achievement of self-fruiting cultivars is advantageous to overcome this limitation, in few cases these still require or greatly benefit from cross-pollinators (LaRue and Norton, 1989).

- *Fruit color*

  Although the consumers' color preference generally varies from country to country, in general, the green fruit color is preferred in Asia and the blue-like skinned fruit in Middle Europe (Hartmann and Neumüller, 2009). Producers have preferred the dark skin over the light colored because it hides imperfections and bruises, which both have a negative effect on the consumer choice. However, the dark skin appears weeks before reaching the consumption maturation stage and complicates the identification of the right harvest time, which has a negative sales impact if consumers purchase unripen fruits. Therefore, due in part to the overproduction of plums with dark skin color and their lowering price, the market demands varieties with a wider range of fruit colors (Okie and Hancock, 2008). In general, homogenously colored fruits have been preferred (Lozano et al., 2009) but mottled varieties are also valued in the market. The development of cultivars producing fruits with high antioxidant properties is of interest due to their enhanced nutraceutical value (Netzel et al., 2012; Valderrama-Soto et al., 2021). These fruits are often accompanied by an increase in their anthocyanin pigments, which consumers seeking this added value can easily identify them by the presence of red coloration. The intensity of anthocyanin coloration is regulated by multiple genes, but the yellow skin color appears to be controlled by a single gene recessive over the red, purple or black, which are quantitatively inherited; and the red flesh color is dominant over the yellow and is inherited by a single factor (Weinberger and Thompson, 1962).

- *Firmness and texture perception*

  An appropriate fruit firmness is one of the most important traits to keep the fruits at their optimal conditions during the harvesting, handling, storage and marketing; and is essential for far-distance shipping because firm fruits are often accompanied by a better postharvest behavior (Byrne, 2012). Fruit firmness is highly inherited and was one of the main targets in the early hybridizations performed by Burbank in the late 19th century. This, together with the importance of the trait, has resulted in the production of firm fruits from the most commercialized *P. salicina* hybrids, and some genotypes have been bred with outstanding results in this aspect (Okie and Hancock, 2008). The fruit firmness is influenced by the texture, which can be from very fine to fibrous. Some juiciness is preferred for fresh consumption, while the mealy, dry texture is avoided (Hartmann and Neumüller, 2009).

- *Freeness of the stone*

  Fruits with the flesh not attached to the pit (freestone) are desired over those that are clingstone. The freestone phenotype is especially important when the fruits are not for fresh consumption, because it reduces the cost of processing (Hartmann and Neumüller, 2009). The freestone character is recessive over the clingstone, and intermediate phenotypes such as semi-clingstone are possible. The fruit ripening stage and the flesh firmness can affect the degree of stone adhesion (Okie and Weinberger, 1996). Generally, the cultivars have cling to semi-clingstone, but some are completely freestone. In some cases, freestone plums are described as “air-free” when an air pocket surrounds the stone (Okie and Hancock, 2008).

- *Organoleptic quality*

  The flavor is a combination of taste and aroma. The trait is hereditable, although with a high environmental component. The climatic conditions, the crop management and fruit ripening stage at harvest will determine the organoleptic quality of the fruit. Measuring the soluble solids concentration (SSC) and the titratable acid (TA) levels of the fruits allow the objective comparison between genotypes (Topp et al., 2012). In Japanese fruits these two factors, as in many other fruit crops, heavily determine consumer acceptance: High SSC levels (>12%) were preferred independently of the TA levels studied, while below these SSC levels higher TA levels negatively influenced the consumer acceptability (Crisosto et al., 2004). The breeding aim is to produce cultivars with high SSC levels that will still have adequate levels even if they are harvested before their full maturity (Topp et al., 2012). In some cases, a moderate acidity in the skin is desired to contrast with the sweetness of the flesh. Most Japanese plums are poor in volatile compounds (Lozano et al., 2009) compared to other related species, and can be improved by crosses with apricots to incorporate their rich aromatic profiles (Gómez and Ledbetter, 1993).

- *Disease resistance*

  There are described more than 60 diseases that can affect plums, being the causal agents 6 viruses, 19 fungi, 4 bacteria, 4 nematodes and 36 insects (Okie, 2008). Some of them cause several diseases that can decrease the crop production and fruit quality, even leading to the death of the trees (Ramming and Cociu, 1991). Breeding to incorporate a resistance or tolerance

to a specific disease requires the crosses with genotypes carrying the specific causal gene/s, but which might not have overall desired characteristics for fruit production. For this reason, obtaining a new cultivar resistant for a specific disease with high-quality fruits is a hard and long process, and breeders should take into account whether it compensates for the application of pest controls (if possible) given the incidence of the disease in their region (Topp et al., 2012). The plum pox virus (PPV) causes the Sharka disease, responsible for the most severe impact on plum productivity and is very difficult to control once it has spread (Okie, 2008). Breeding for PPV resistance would be a great advantage, and two resistant cultivars have been identified (Ramming and Cociu, 1991). As an alternative, resistance to the virus can also be incorporated by generating transgenic lines (Rubio et al., 2019) if it is allowed by the regulatory framework. Among the diseases caused by bacteria, the bacterial canker (*Pseudomonas syringae* pv. *syringae*), the bacterial leaf spot (*Xanthomonas campestris* pv. *pruni*) and the plum leaf scald (*Xylella fastidiosa*) are the most widespread, and resistant or tolerant cultivars have been identified for all of them. For a high number of diseases caused by fungi, sources of resistance (black knot, silver leaf, fruit spot, stem cankers, others) or tolerance (powdery mildew, leaf scald) have been found, while for others (peach scab, rose mildew, bladder plum) none has been described (Okie and Hancock, 2008). Resistant or tolerant genotypes have not been found yet for the main insect pests. The resistance to root-affecting nematodes is especially useful for rootstocks, and resistant genotypes exist that and are compatible with Japanese plum scions.

3.2 Marker-assisted selection in breeding

Conventional breeding is based on the selection of genotypes through their phenotype, acquired after the visual evaluation or quantification of the traits of interest. On the other hand, marker-assisted selection (MAS) consists in the use of markers to select for traits or genomic regions of interest, seeking to increase the efficiency of the breeding programs by overcoming the limitations of conventional breeding. The main applications of molecular markers in marker-assisted breeding (MAB) are (Xu and Crouch, 2008):

1) The pyramiding of multiple traits, even multigenic and/or recessive
2) The selection of traits that manifest only under certain environments or development stages
3) The selection of traits that are difficult to manage (difficult or expensive to phenotype, have low penetrance or a complex inheritance)

4) Maintenance of recessive alleles and speeding up backcrosses

The high heterozygosity of Japanese plum results in a great phenotypic variance in seedlings (**Figure I.3**), requiring the generation of large populations to increase the chances of pyramiding a high number of desired traits in a single genotype (Topp et al., 2012). Most of the desired traits refer to fruit characteristics, which means that in conventional breeding, all the seedlings have to be maintained until overcome their juvenility period and begin fruiting. MAS can overcome this handicap since the fruit trait can be predicted early in the plant life cycle. This is particularly useful in species with long juvenile periods such as most *Prunus* crops (Byrne et al., 2012). This early selection allows to generate larger populations, increasing the chances of pyramiding the desired traits while reducing the field or greenhouse maintenance costs. Another advantage of MAB is the possibility to estimate the phenotypic segregation expected from crossing two lines and, therefore, plan the crosses according to the breeding objectives. This is of special interest for the identification of recessive alleles which may be in heterozygosity in the parents.

The first markers used were morphological, which relied in different phenotype variations conferred by single genes (major genes) and their possible co-segregation due to their genetic linkage. However, their use was imprecise, tedious and limited to a few model crops (Arús, 2017). Other markers used were the cytological, based on chromosome staining and visualization (Nadeem et al., 2018).

The use of molecular markers for plant breeding was popularized when it was shown that they could accelerate the introgression of monogenic traits from exotic germplasm into elite cultivars (Xu and Crouch, 2008). Isozyme markers (biochemical markers) were the first and their use represented a huge step because they showed codominance and, as it interrogated protein variations, did not rely in the environment. However, their main disadvantage was that only a few of them existed for each species; and by the year 1980 they started to be replaced by DNA molecular markers, which their number was virtually unlimited because they could interrogate all the genome (Arús, 2017).

Figure I.3. Several fruit traits segregating from a F1 population developed by PLANASA breeding company. **A)** Fruits segregating for skin color (purple, mottled, yellow), shape (round, ovoid, heart-shaped) and skin texture (smooth, rugous). **B)** Fruits segregating for flesh color (red, yellow, white) on the same population.

The first DNA markers were RFLPs (restriction fragment length polymorphism), that despite being expensive and time consuming allowed many uses for crop improvement. These included the marker-assisted backcrossing and the construction of saturated genetic maps covering the whole genome, which boosted QTLs analysis and discovery (Beckmann and Soller, 1986; Arús, 2017). Then by the decade of the 1990s these markers were replaced by others mostly based on PCR (polymerase chain reaction), such as RAPDs (randomly amplified polymorphic DNA) and AFLP (amplified fragment length polymorphisms), until reaching the two types of molecular markers most used nowadays: SSRs and SNPs (Arús, 2017; Nadeem et al., 2018). The advantages of SSRs and SNPs

over the earliest molecular markers are their co-dominance, high reproducibility, wide distribution over the genome and easier interpretation of the results (Nadeem et al., 2018).

With the evolution of the sequencing technologies, more molecular markers could be identified all over the genome, conferring tools for the generation of more saturated genetic maps and of GWAS (genome wide association studies) (Nadeem et al., 2018), which consists in the search of genotype-phenotype associations in large germplasm collections using their whole-genome variability. This greatly enhanced the identification of QTLs and the fine-mapping of major genes. With the easiness to detect polymorphisms by sequencing, the markers could be more targeted to a gene or genomic region of interest (Arús, 2017), improving the trait association and the efficiency of MAS.

In some cases, a molecular marker can be simplified by a single PCR reaction and the results visualized in agarose gels, a technology available in most molecular laboratories at a reduced marker usage cost. However, the lack of scalability turns it unsuitable when a large number of markers have to be screened, and the development of efficient, low-cost high-throughput genotyping systems would be preferable for large-scale MAS (Xu and Crouch, 2008).

3.3 Marked-assisted selection in Japanese plum

Few breeding programs exist dedicated at obtaining new Japanese plum varieties, compared to other crops that are more relevant in the fruit market. The scarcity of genetic and genomic studies related to QTL identification (detailed in section 1.5, Genetic and genomic resources) has hindered the development of molecular markers associated to agronomic traits of interest in the crop (Okie and Hancock, 2008).

The only validated molecular marker to date is for the self-incompatibility trait, used in breeding programs to assess the pollen-pistil compatibility between germplasm and, therefore, plan the crosses accordingly to the marker results. The trait is genetically regulated in the *S* locus by two close genes: The pistil determinant is a ribonuclease (*S*-RNase) that inhibits the pollen tube growth from pollen of incompatible genotypes, the pollen determinant is a F-box protein (SFB) that might be involved in the degradation of the *S*-RNases (Guerra and Rodrigo, 2015). The *S*-genotyping consist in the size differences of an intron in the *S-RNase* gene, and the PCR reaction has been optimized to precisely determine its length by capillary electrophoresis (Guerra et al., 2012). Despite the

usefulness of the molecular marker, the self-compatibility is still not accurately predicted due to the existence of self-incompatible cultivars genotyped with *S*-alleles associated with the self-compatibility (Guerra and Rodrigo, 2015).

With the objective to provide molecular markers for the fruit skin color, a set of expressed sequence tag SSRs (EST-SSRs) markers from key genes of the anthocyanin biosynthetic pathway were designed (González et al., 2016b). The genotypic data obtained in 29 cultivars with three EST-SSRs, corresponding to the *PsMYB10*, *PsMYB1* and the *PsbHLH35* genes, clustered the 6 yellow-skin cultivars in a single group which also included four of the red skin cultivars. These markers should be validated in a bigger panel of cultivars to assess their effectivity for the prediction of red/purple/black skin coloration in MAS. Despite the markers could not accurately identify the recessive phenotypes (yellow/green skin), the study was relevant because it was the first approximation of the use of molecular markers selecting for fruit color in Japanese plum cultivars.

3.4 Major genes and QTLs mapped in the *Prunus* genome

The design of the molecular markers detailed above relied in the conservation of genetic mechanisms within the genus. Due to the low number of genetic studies specific for Japanese plum and the high synteny between *Prunus* species, the major genes and QTLs mapped in the genome of other *Prunus* are a valuable resource for the development of molecular markers for this crop.

The saturated linkage map of the F2 population from 'Texas' (almond) x 'Early Gold' (peach) serves as the *Prunus* reference map (abbreviated T×E). The map was initially constructed with 226 RFLPs and 11 isozyme markers (Joobeur et al., 1998) and has been improved over the years with more RFLPs and the addition of SSRs markers (Joobeur et al., 2000) and later reconstructed using only SNPs and SSRs (Donoso et al., 2015). As a result of the high transferability of the T×E markers developed and the high synteny between *Prunus* species, several maps constructed with markers in common were interconnected and the position of 28 major genes was integrated into an interspecific consensus reference map (Dirlewanger et al., 2004a). Among these 28 mapped major genes, 19 were mapped in peach progenies, 6 in almond or almond × peach, 2 in apricot and 1 in a myrobalan plum (**Figure I.4**).

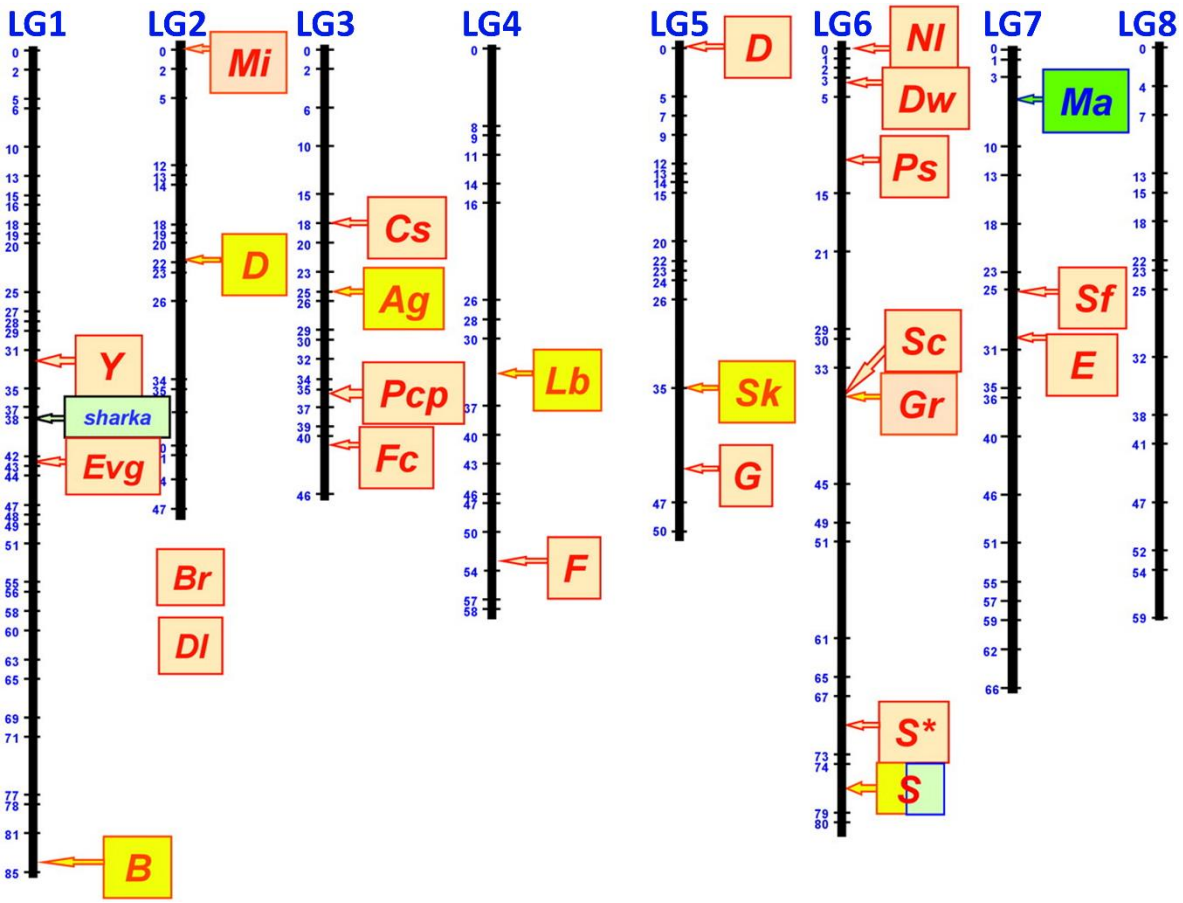


Figure I.4. The approximate position of 28 major genes mapped in different populations of peach (orange background), almond and almond × peach (yellow), apricot (blue) and myrobalan plum (green) integrated in a single *Prunus* reference map. The major genes correspond to peach white/yellow flesh color (*Y*), PPV resistance (*sharka*), evergrowing (*Evg*), almond/peach petal color (*B*), nematode resistance (*Mi*), almond shell hardness (*D*), broomy plant habit (*Br*), double flower (*Dl*), flesh color around the stone (*Cs*), anther color (*Ag*), polycarpel (*Pcp*), flower color (*Fc*), blooming date (*Lb*), flesh stone adherence (*F*), non-acid fruit in peach (*D*), bitter kernel (*Sk*), fruit skin pubescence (*G*), leaf shape (*Nl*), dwarf plant (*Dw*), male sterility (*Ps*), fruit skin color (*Sc*), leaf color (*Gr*), fruit shape (*S**), apricot and almond self-incompatibility (*S*), myrobalan plum root-knot nematode resistance (*Ma*), powdery mildew resistance (*Sf*), leaf gland shape (*E*). The approximate position of *Br* and *Dl* is unknown. Figure from Dirlewanger et al. (2004).

The genetic variation in these 28 major genes lead to great phenotypic diversity useful for breeding, as they mostly regulate traits of agronomic interest following a Mendelian inheritance (mendelian trait loci, MTL). Because of this, they are excellent candidates to design markers for MAS. Some are already incorporated in MAB of peach for the white/yellow flesh color (*Y*), the subacid fruit (*D*), the flat fruit shape (*S**), the non-glabrous/glabrous nectarine skin (*G*), the slow ripening (*Sr*), the skin

blush (*Rf*), the red skin color suppression (*H*) and the stone adherence and melting flesh (*F-M*) (Gu et al., 2016; Vanderzande et al., 2018). In sweet cherry, MAS exists for the fruit color (Sandefur et al., 2016) and the self-compatibility/self-incompatibility (Wünsch and Hormaza, 2004); markers for this last trait (*S*) are also applied in almond, apricot and Japanese plum (Channuntapipat et al., 2003; Romero et al., 2004; Guerra et al., 2012). The MAS of nematode resistance (*Mi*) and root-knot nematode resistance (*Ma*) genes has facilitated the pyramiding of the traits for rootstock breeding (Dirlewanger et al., 2004b).

Markers associated to QTLs are more difficult to identify and to incorporate into MAB than those targeting major genes, especially those QTLs regulated by many genes and with a high environmental effect. For example, although the resistance to Sharka (plum pox virus) in apricot is controlled by a major gene, in peach 6 genomic regions involved in the PPV resistance have been identified in a *P. davidiana* clone (Marandel et al., 2009), which hinders the introgression and pyramiding of the resistance into elite peach susceptible cultivars. Several QTLs in *Prunus* have been identified, a few of them involved in fruit production and quality are shown in **Table I.3**.

Table I.3. Some QTLs for fruit color, soluble solid content, fruit yield and resistance to diseases, identified from several *Prunus* populations and mapping in more than one linkage group (LG).

Trait	LG	Species	Reference
For fruit color			
Red flesh coloration	1	Peach/*P. davidiana*	Quilot et al. (2004)
Skin color	2	Peach/*P. ferganensis*	Verde et al. (2002)
Skin color	3	Japanese plum	Salazar et al. (2017)
External color	3	Peach	Eduardo et al. (2011)
Blood flesh	3	Peach/almond	Donoso et al. (2016)
Percentage skin color	3	Peach/almond	Donoso et al. (2016)
Red flesh coloration	3	Peach/*P. davidiana*	Quilot et al. (2004)
Fruit skin color	3	Apricot	Ruiz et al. (2010)
Skin color	3	Sweet cherry	Sooriyapathirana et al. (2010)
Flesh color	3	Sweet cherry	Sooriyapathirana et al. (2010)
Skin color	4	Japanese plum	Salazar et al. (2017)
Percentage skin color	4	Peach/almond	Donoso et al. (2016)
Skin color	5	Peach/*P. davidiana*	Quilot et al. (2004)
External color	6	Peach	Eduardo et al. (2011)
Skin color	6	Sweet cherry	Sooriyapathirana et al. (2010)

(Table continues)

Table I.3 (continued)

Trait	LG	Species	Reference
Skin color	6	Peach/*P. ferganensis*	Verde et al. (2002)
External color	7	Peach	Eduardo et al. (2011)
Flesh color	8	Sweet cherry	Sooriyapathirana et al. (2010)
For soluble solids content			
Soluble solids content	2	Peach/*P. ferganensis*	Verde et al. (2002)
Soluble solids content	4	Japanese plum	Salazar et al. (2020)
Soluble solids content	4	Peach	Eduardo et al. (2011)
Soluble solids content	4	Peach/*P. davidiana*	Quilot et al. (2004)
Soluble solids content	5	Peach/*P. davidiana*	Quilot et al. (2004)
Soluble solids content	5	Japanese plum	Salazar et al. (2020)
Soluble solids content	6	Peach/*P. ferganensis*	Verde et al. (2002)
For fruit yield			
Fruit cheek diameter	1	Peach/*P. davidiana*	Quilot et al. (2004)
Fruit suture diameter	1	Peach/*P. davidiana*	Quilot et al. (2004)
Fruit weight	2	Japanese plum	Salazar et al. (2020)
Fruit weight	2	Sweet cherry	Zhang et al. (2010)
Fruit weight	3	Apricot	Ruiz et al. (2010)
Fruit dimensions	4	Peach/almond	Donoso et al. (2016)
Fruit mass	4	Peach/*P. davidiana*	Quilot et al. (2004)
Fruit cheek diameter	4	Peach/*P. davidiana*	Quilot et al. (2004)
Fruit weight	4	Apricot	Ruiz et al. (2010)
Fruit production	6	Peach/almond	Donoso et al. (2016)
Fruit weight	6	Peach/almond	Donoso et al. (2016)
Fruit weight	6	Sweet cherry	Zhang et al. (2010)
Fruit weight	8	Apricot	Ruiz et al. (2010)
For disease resistance			
PPV resistance	1	*P. davidiana*	Marandel et al. (2009)
PPV resistance	4	*P. davidiana*	Marandel et al. (2009)
PPV resistance	5	*P. davidiana*	Marandel et al. (2009)
PPV resistance	6	*P. davidiana*	Marandel et al. (2009)
PPV resistance	7	*P. davidiana*	Marandel et al. (2009)
Resistance to mildew	7	Peach/*P. ferganensis*	Verde et al. (2002)
Resistance to mildew	8	Peach/*P. ferganensis*	Verde et al. (2002)

3.5 Gene function validation

Once the association between a genomic region and a trait is found through QTL or GWAS mapping, the identification of candidate gene/s and/or possible causal genetic variants is addressed to increase the efficiency of MAS. This process is known as fine-mapping and involves several strategies such as narrowing down the region by using molecular markers in recombinant genotypes, DNA sequencing and gene expression analysis; all aiming at reducing the list of candidates to a few genes and causal polymorphisms (Schaid et al., 2018). The polymorphism(s) must also correlate with the studied phenotype in other germplasm collections, and the function of the candidate gene(s) validated by knock-out or overexpression methods (Liu and Yan, 2019).

Most *Prunus* species are recalcitrant for genetic transformation, mostly because of poor regeneration rates. However, the genetic transformation of Japanese plum has been achieved, although with reduced regeneration rates compared to those obtained for genetic transformation of the European plum (Urtubia et al., 2008; Rubio et al., 2019). Due to the tremendous effort and long time required to obtain transgenic plantlets, a transient overexpression in a heterologous system, such as the leaves of *Nicotiana* plants, is preferred for gene functional validation studies (Lin-Wang et al., 2010; Fang et al., 2021b). The transient gene expression in fruits allows for a fast functional validation of genes in the species of interest whenever the traits are expressed in the fruit and their expression is observable and/or quantifiable. This strategy has been applied in some rosaceous crops such as peach, apricot, apple, pear and strawberry (Spolaore et al., 2001; Xi et al., 2019), while in Japanese plum the only approach used consisted in the transient silencing of a gene through virus-induced gene silencing (VIGS) (Fang et al., 2021a).

SECTION 4: THE RED-TO-PURPLE FRUIT COLORATION

4.1 Fruit color in *Prunus*

Breeding for high quality fruits is one of the main objectives in breeding programs. Fruit color highly determines its appearance and the attractiveness for the consumer, and is mainly caused by the accumulation of anthocyanins (conferring red to purple fruit colors), carotenoids (yellow to red) and chlorophylls (green) pigments (Lancaster et al., 1994; Khoo et al., 2011, 2017; Sinecen et al., 2015). During the fruit maturation process the fruit progressively changes its coloration by the degradation of chlorophylls and the synthesis of anthocyanins and carotenoids. In some cultivars the green color

is still visible in the mature fruit, and post-harvest treatments reducing the activity of the chlorophyllase enzyme have been shown to delay the loss of green color (Luo et al., 2009).

One of the major genes mapped in the *Prunus* genome is for the white/yellow flesh coloration (*Y*) in LG1, in which the white color is dominant over the yellow. The candidate gene is a carotenoid cleavage dioxygenase 4 (*CCD4*) responsible for the degradation of the pigment, explaining the decreased carotenoid levels observed in white-fleshed peaches over those that are yellow (Falchi et al., 2013).

The coloration conferred by anthocyanins overlies the produced by chlorophylls and carotenoids when they are in the same tissue. In Japanese plum the yellow skin is transparent, and if accompanied by red flesh it is observable as a reddish ground color in a mottled skin named "bronze" (Okie, 2008). Several QTLs for the red fruit coloration have been identified along the eight *Prunus* linkage groups, most involved in the coloring of single tissues rather than the whole fruit (**Table I.3** in page 32). Some loci can produce a similar phenotype on the ripe fruit, while can differ in their inheritance and in their expression points along the fruit maturation process (Werner et al., 1998; Zhou et al., 2015). Moreover, environmental factors such as light exposure, temperature and post-harvest treatments can modulate the anthocyanin-based coloration (Minas et al., 2013; Fang et al., 2021b).

The final coloration in the mature fruit is the result of a complex regulation network, which together with the high heterozygosity of Japanese plum has given rise to the wide fruit color variation observed in the crop (**Figure I.5**).

Figure I.5. Fruits from six different Japanese plum cultivars, showing the wide variability existing in the crop for the skin and flesh colors. The left figure in the bottom row corresponds to a mottled-type fruit with the bronze skin defined by Okie (2008).

4.2 Anthocyanins. Structure, synthesis and properties

Anthocyanins are secondary metabolites whose name derive from the Greek words *anthos* (flower) and *kyanos* (dark blue), although they are responsible for a wider spectrum of colors (pale yellow, orange, red, magenta, violet and blue) in more plant tissues such as organs, leaves, stems and storage structures (He and Giusti, 2010). These plant pigments are phenolic compounds classified as flavonoids, formed by the glycosylation of an anthocyanidin molecule that contains a C-15 backbone structure, organized in two C-6 benzyl rings and a heterocyclic C-3 ring carrying a positive charge (Bueno et al., 2012) (**Figure I.6**).

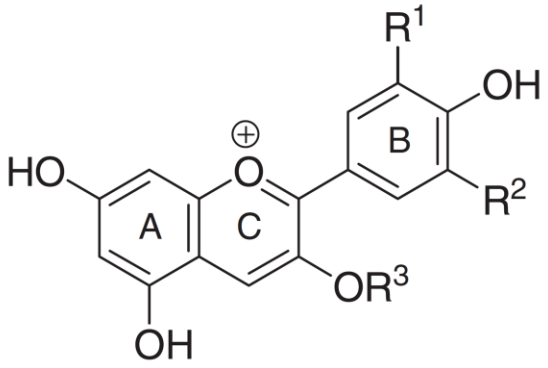


Figure I.6. Structure of the main anthocyanins-3-O-glucoside found in fruits. A and B correspond to the benzyl rings and C to the heterocyclic ring carrying a positive charge. Substitutes in R1, R2 and R3 define the six main anthocyanidins detailed in Table I.4 (Figure from Bueno et al., 2012).

The glycosylation stabilizes the molecule and also increases its solubility in water. Contrarily, its acylation produces the aglycone form which is rarely found in nature due to its low stability (He and Giusti, 2010). Over 600 anthocyanins have been identified mainly differing in the number, position and type of sugars attached to their main backbone, which in 94% of the cases is based in only 6 types of anthocyanidins that greatly determine their coloration (**Table I.4**) (Bueno et al., 2012; Celli et al., 2018).

The anthocyanin molecules rearrange their structure depending of the pH, being the most stable form the conformed in acidic conditions. Anthocyanins are stored in the cell vacuole, which are mildly acidic in most plant cells but hyperacidified in specific cells such as flower petals, with pH differences in the vacuole determining the anthocyanic coloration (Faraco et al., 2014). The stability of anthocyanins is also determined by the temperature, oxygen, light, water activity and the presence of other compounds such as metal ions and copigments (like flavonols) (Yoshida et al., 2009; Celli et al., 2018).

Table I.4. The major anthocyanin-3-O-glucosides found in fruits according to their R1, R2 and R3 substitutes, the colors conferred and their maximum visible length of absorption when solubilized in methanol-hydrochloric acid (0.01%). Table from Bueno et al. (2012).

Anthocyanidin	Abbreviation	R1	R2	λmax (nm)		Colors
				R3=H	R3=Glucose	
Delphinidin	Dp	OH	OH	546	541	Purple, mauve and blue
Petunidin	Pt	OH	OCH3	543	540	Purple
Malvidin	Mv	OCH3	OCH3	542	538	Purple
Cyanidin	Cy	OH	H	535	530	Magenta and crimson
Peonidin	Pn	OCH3	H	532	528	Magenta
Pelargonidin	Pg	H	H	520	516	Orange salmon

The main anthocyanins found in Japanese plum fruits are cyanidin-3-O-glucoside and cyanidin-3-O-rutinoside (Yu et al., 2020; Valderrama-Soto et al., 2021). Anthocyanins are synthesized from the phenylpropanoid pathway, which branches to the flavonoid pathway by the action of the chalcone synthase (CHS) enzyme that generates a tetrahydroxychalcone molecule from 4-coumaroyl-CoA and malonyl-CoA, which by the action of several enzymes is transformed into anthocyanidins (**Figure I.7**). These are then glycosylated by the UDP-glucose:flavonoid 3-O-glucosyl transferase (UFGT) enzyme to generate the anthocyanin pigments, which are later transported to the vacuole by the action of glutathione-S-transferase (GST) enzymes (Zhao et al., 2020). The pathway can branch from the dihydroflavonols to produce flavonols, and from the leucoanthocyanidins and anthocyanidins to proanthocyanidins and produce condensed tannins.

The most recognized property of anthocyanins is their antioxidant activity, in which they efficiently scavenge free radicals and reactive oxygen species (ROS). In the plant, anthocyanins are important for the protection of chloroplasts from the photoinhibitory and photooxidative stresses, caused when the photosynthetic system receive an excess of light energy causing an efficiency decrease and the generation of ROS. This scavenging property also aids in the recovery of mechanical injures (Gould, 2004).

The colors provided by anthocyanins in flowers and fruits enhances the pollination and seed dispersal mediated by animals by increasing their visual attractiveness.

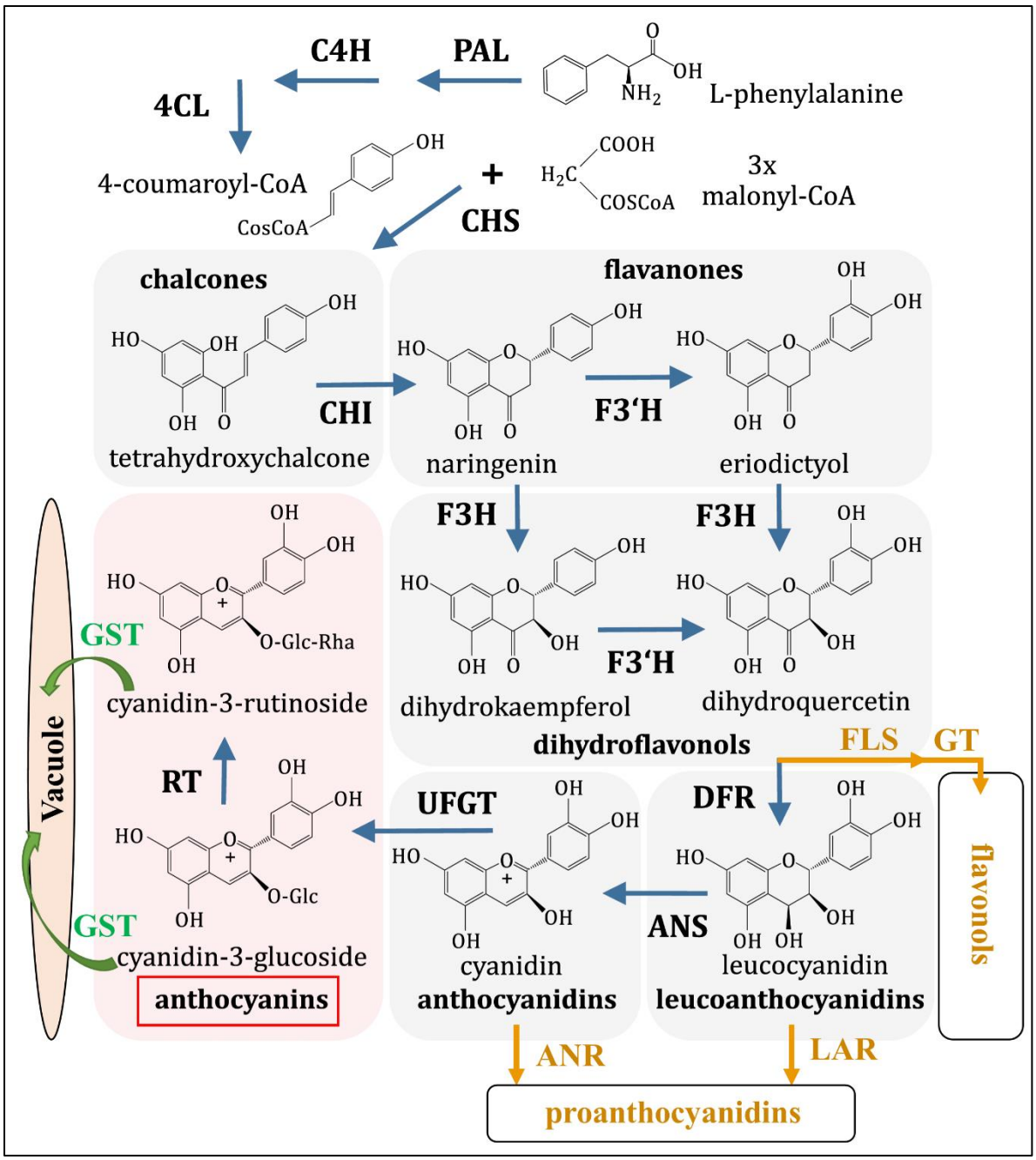


Figure I.7. The anthocyanin biosynthesis pathway showing, in black capital letters, the abbreviations of the enzymes involved: phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3'-hydroxylase (F3'H), flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR), anthocyanidin synthase (ANS, also referred as LDOX from leucoanthocyanidin dioxygenase), UDP-glucose:flavonoid 3-O-glucosyl transferase (UFGT) and rhamnosyltransferase (RT). In green capital letters, the enzyme transporting the anthocyanins to the vacuole: glutathione S-transferase (GST). In orange capital letters, the enzymes diverging to flavonol synthesis: flavonol synthase (FLS), glycosyl transferase (GT); and to proanthocyanidin synthesis: anthocyanidin reductase (ANR) and leucoanthocyanidin reductase (LAR). Figure edited from Starkevič et al. (2015).

Anthocyanin presence in leaves, pollen and meristems provides UV photoprotection to avoid DNA damage, while also prevents the catabolism of photolabile defense compounds (Winkel-Shirley, 2002). They also contribute to the biotic resistance by protecting the plants from herbivores such as insects, which lack red light receptors and red leaves can result unattractive to them. Anthocyanin ingestion can be toxic for insects, as well as for some bacteria and fungi, acting as defense compounds (Kong et al., 2003; Gould, 2004). Anthocyanins have also been associated to abiotic stresses such as resistance to desiccation, chilling and freezing and heavy metal contamination (Gould, 2004).

When incorporated to the human diet, anthocyanins have a vast list of health-promoting effects, mostly related to their antioxidant properties. In humans, ROS production is important for many body functions such as cell signaling and the immune system, but an overproduction is linked to inflammation, cardiovascular diseases, cancer and accelerated aging, which can be lessened by the ROS scavenging activity of anthocyanins. As such, anthocyanin intake has been associated with anticarcinogenic and anti-inflammatory activity, neuroprotection, the prevention of cardiovascular diseases and obesity, the control of diabetes, an enhancement in the visual health and finally, the antimicrobial effect against intestinal pathogens without altering the beneficial bacteria (He and Giusti, 2010; Khoo et al., 2017). The intake of anthocyanins can come from its natural source by the ingestion of fruits and vegetables, as dietary supplements or by its use as natural colorants in processed foods and beverages (Khoo et al., 2017).

4.3 The role of *MYB10* genes in fruit color

The biosynthesis of anthocyanins is regulated by transcriptional factors, which can activate or repress several structural genes of the flavonoid pathway to modulate, under specific development and environmental signals, the class of pigment produced and the tissue where it is synthesized (Gonzalez et al., 2008). These transcription factors include proteins from the MYB, the bHLH and the WD-repeat families, which bind to form a transcriptional complex named MBW, a mechanism conserved in all plant species studied (Gonzalez et al., 2008). The MBW is able to repress or promote several branches of the phenylpropanoid pathway, binding directly to the promoters of the structural genes and activating or repressing their transcription (Koes et al., 2005).

The MYB family is represented in all eucaryotes and is one of the largest families of plant transcription factors, with their members being functionally diverse. MYB proteins are recognized by a conserved DNA-binding protein domain formed by up to three imperfect repeats of about 53 amino acids creating a helix-turn-helix structure, which include three regularly spaced tryptophan residues. These three repeat domains (R1, R2 and R3) give name to the protein subfamilies. The R2R3-MYB is its largest subfamily in plants (Stracke et al., 2001), and the role of some of its members in the regulation of anthocyanin synthesis was early discovered when the first plant transcription factor was cloned: the *colorless1* (*C1*) MYB gene from maize (Cone et al., 1986). However, members of this subfamily are not functionally limited to anthocyanin regulation and have a wide range of functions. R2R3-MYB proteins mostly regulate plant-specific processes, which might be explained because they are only found in the plant kingdom (Stracke et al., 2001).

In Arabidopsis, the R2R3-MYB phylogenetic subgroup 6 (SG6) is composed by four genes that are positive regulators of anthocyanin biosynthesis: PAP1 (*AtMYB75*), PAP2 (*AtMYB90*), *AtMYB113* and *AtMYB114*, the last three in tandem conformation (Stracke et al., 2001; Gonzalez et al., 2008). In other species, orthologs of these genes conserve the ability to promote the synthesis of anthocyanins by the conserved MBW mechanism, such as in apple. Different research groups isolated three SG6 orthologous genes from this crop, which were associated with the coloration of the apple skin and flesh fruit tissues and could promote anthocyanin synthesis in both homologous and heterologous systems (Takos et al., 2006; Ban et al., 2007; Espley et al., 2007). These genes were named *MdMYB1*, *MdMYBA* and *MdMYB10*, but further homology and linkage studies identified them as alleles of the same gene, which was finally named *MdMYB10* (Lin-Wang et al., 2010). By using conserved PCR primers, the *MYB10* gene was isolated from several crops in the Rosaceae family such as peach, plum, apricot, sweet cherry, pear, strawberry and others. Conserved amino acids were identified between their protein sequences, and it was shown that the isolated genes could promote the synthesis of anthocyanins when overexpressed in *Nicotiana* (Lin-Wang et al., 2010). Therefore, the *MYB10* constituted a candidate gene for promoting the fruit coloration in the whole Rosaceae family.

In the peach reference genome, a cluster of three *MYB10* genes was identified in LG3 colocalizing with the anther color (*Ag*) trait (Rahim et al., 2014) (**Figure I.4** in page 31) and also with QTLs involved in the fruit coloration in other *Prunus* species, including Japanese plum skin color (**Table I.3** in page

32). In peach these duplicated genes locate within an 80 kb region and are named *MYB10.1*, *MYB10.2* and *MYB10.3* (**Figure I.8**).

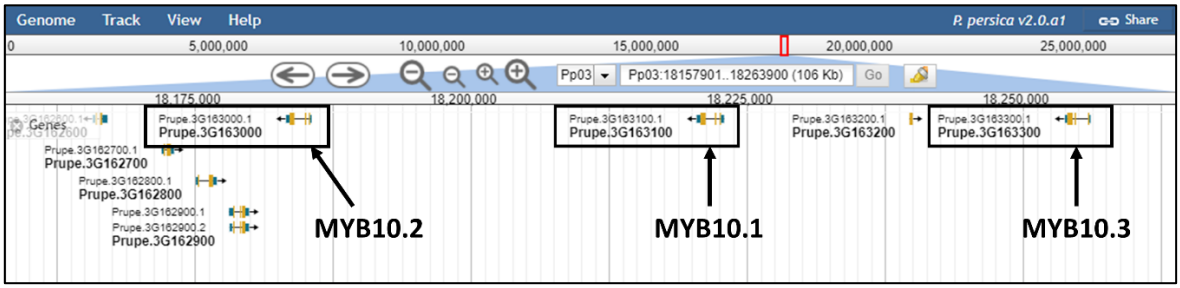


Figure I.8. The three MYB10 genes located within a region of 80 kb in the *Prunus persica* genome v2.0.a1 (Verde et al., 2017), visualized using JBrowse in the GDR portal (http://rosaceae.org). The three genes are in reverse orientation and consist of three exons, with a full gene length sequence between 2.2 and 2.5 kb. In the peach reference genome, the *MYB10.1* corresponds to *Prupe.3G163100* (former *ppa026640* in the *P. persica* v1.0 genome), the *MYB10.2* to *Prupe.3G163000* (former *ppa016711*) and the *MYB10.3* to *Prupe.3G163300* (former *ppa020385*) annotated genes.

In peach, the expression of these genes and their role in the acquisition of skin and flesh coloration varies between cultivars and fruit tissues (Ravaglia et al., 2013; Rahim et al., 2014; Tuan et al., 2015), but it was finally established that the alleles of the *MYB10.1* gene determined the skin color. Markers for the *MYB10.1* alleles allowed the MAS of red/non-red skin color in peach (Tuan et al., 2015). Later it was observed that the *MYB10.1* allelic variation also cosegregated with the highlighter (*h*) phenotype, in which the anthocyanin synthesis in the whole fruit tissues, including the skin blushing, is suppressed (Bretó et al., 2017). Variations in *MYB10.1* were also correlated with the red flesh color around the stone in some genotypes (Guo et al., 2020). In other studies, the segregation of the blood flesh color was caused by the ability of a NAC gene to activate the *MYB10.1* expression (Zhou et al., 2015).

In other *Prunus* species, the *MYB10* genes are named according to their closest homology with the peach genes. In sweet cherry, the *MYB10.1* homolog also correlated with the red fruit coloration (Starkevič et al., 2015; Jin et al., 2016). In a F1 population from Japanese plums segregating red to purple skin colors, the SNP most associated with the trait variation was found close to the *MYB10.1* gene (Salazar et al., 2017). In apricot, a *MYB10* gene was also responsible for the red blushed skin (Xi et al., 2019).

Aside from the *MYB10* genes in LG3, in peach three other *MYB10* homologous genes were identified in LG6 (named *MYB10.4*, *MYB10.5* and *MYB10.6*) also in tandem conformation, co-localizing within the leaf color trait (*Gr*) involved in the anthocyanin accumulation in vegetative organs (Rahim et al., 2014; Zhou et al., 2014). The function of *MYB10.4* was validated and associated with the red leaf coloration in peach (Zhou et al., 2014), while in *P. cerasifera* var. *atropurpurea* (red foliage plum) the leaves and fruit tissues were colored due to the *MYB10.6* constitutive expression (Gu et al., 2015).

The MYB10 protein, when assembled into the MBW complex, has been found to activate late structural genes from the anthocyanin biosynthesis pathway such as *UFGT*, *ANS* or *DFR*; and also the *GST*, that is responsible for the transport of anthocyanins to the vacuole (Ravaglia et al., 2013; Jiang et al., 2019b). Other MYBs repress the synthesis of anthocyanins by redirecting the pathway to other branches, such as the MYBPA1, that activates the promoters of *ANS* and *DFR* but not the *UFGT* to promote the synthesis of proanthocyanidins (Ravaglia et al., 2013). Other MYBs lacking functional domains can repress the phenylpropanoid pathway by competing with the functional MYBs in the formation of the MBW complex (Ramsay and Glover, 2005).

The multiple QTLs identified for fruit coloration indicate a multigenic inheritance, which is in concordance with the complex regulatory network required to synthesize the anthocyanins. Aside from the R2R3-MYB gene, the other two members of the MBW complex have a role in the DNA binding and formation of the complex (Ramsay and Glover, 2005) and therefore their variability also determines the synthesis of anthocyanins, as well as variations in regulatory or structural genes from the pathway (Wu et al., 2017; Zhao et al., 2020; Lu et al., 2021) or genes that directly or indirectly activate or repress the MYB10 genes (Zhou et al., 2015).

However, both the ectopic activation of the *MYB10* genes in transgenic plants and their allelic variation have a huge impact on the acquisition of color, indicating that these genes are dominant factors in the regulation of anthocyanin pigments (Ban et al., 2007). Therefore, the *MYB10* genes variability is worth the study, which could aid the development of molecular markers linked to fruit color traits in crops of interest.

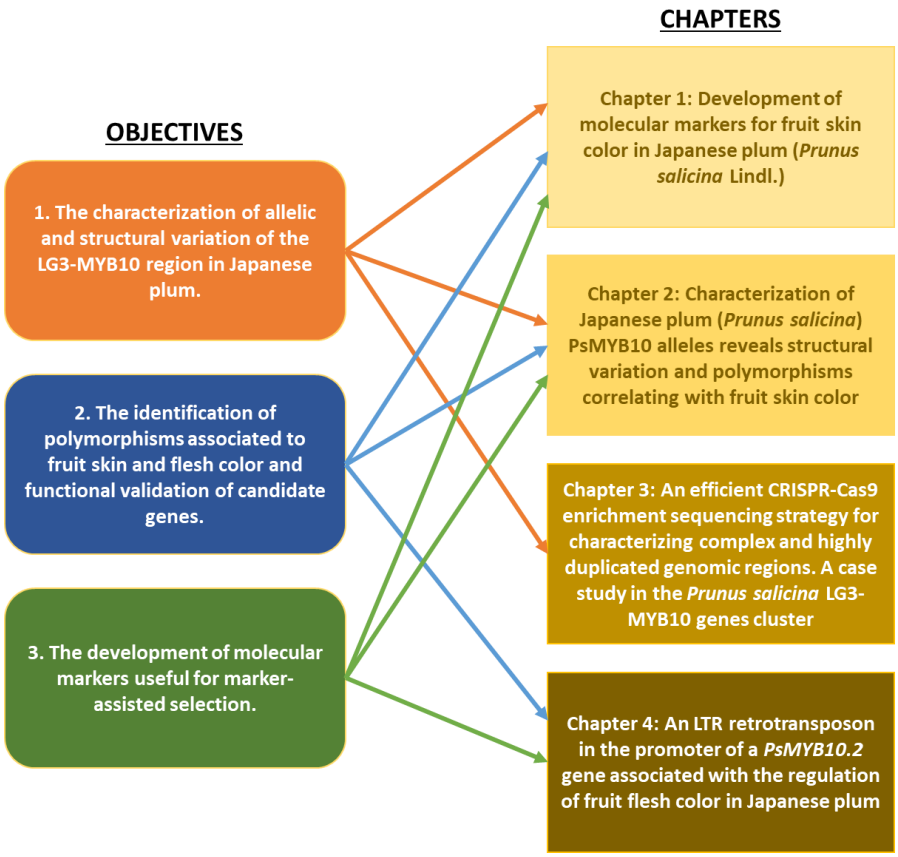
OBJECTIVES

The main objective of this thesis is to unravel the genetic variability regulating anthocyanin fruit color in Japanese plum and to aid the marker-assisted breeding in the crop. The efforts were focused in the LG3-MYB10 genomic region, which is candidate for anthocyanin fruit color in *Prunus* species.

To achieve such main goal we proposed three sub-objectives:

1. The characterization of allelic and structural variation of the LG3-MYB10 region in Japanese plum.
2. The identification of polymorphisms associated to fruit skin and flesh color and functional validation of candidate genes.
3. The development of molecular markers useful for marker-assisted selection.

This thesis is structured in four chapters corresponding to four manuscripts published or under review in peer-review journals, three of them (Chapter 2, 3 and 4) included in the SCI Journal. The correspondence between objectives and chapters is represented in the following chart.



Chapter 1

Development of molecular markers for fruit skin color in Japanese plum (*Prunus salicina* Lindl.)

A. Fiol[1], W. Howad[2], A. Surya[1] and M.J. Aranzana[2]

[1]CRAG, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain.

[2]IRTA, Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain.

This chapter has been published in Acta Horticulturae (March 2021)

DOI: 10.17660/ActaHortic.2021.1307.34

ABSTRACT

Japanese plum is a diploid fruit tree species, member of the Rosaceae family, generated by hybridization of *Prunus salicina* with diverse *Prunus* species. Japanese plum cultivars show great variability for fruit skin and flesh color, which are both major objectives in plum breeding. Subsequently, molecular markers for early selection of these traits in breeding programs are highly desirable. Despite candidate genes for fruit color have been identified in several Rosaceae species, no validated markers have been described for Japanese plum yet. In Rosaceae family, the MYB10 transcription factor has been described as the main gene determining anthocyanin pigment accumulation in fruits, which is responsible for red, purple and black coloration. In order to design a useful marker for marker-assisted selection (MAS), we have explored the variability of the *MYB10* gene group in Japanese plum and its association with fruit color. A set of primers were designed targeting conserved MYB10 domains using sequences from peach genome. Primer combinations were tested using a reduced set of *P. salicina* accessions. The primer pair amplifying more polymorphic alleles was selected and used to genotype a collection of 78 Japanese plum cultivars. One allele, present in all 51 skin-colored accessions and absent in the rest, was found associated with anthocyanin accumulation. This dominant marker can effectively predict the presence or absence of fruit skin coloration caused by anthocyanin pigments.

Keywords: MYB10, fruit color, anthocyanins, marker-assisted selection

INTRODUCTION

Japanese plum (*Prunus salicina* Lindl.) is a fruit tree species member of the Rosaceae family, together with other also important fruit trees as peach, almond, apricot and sweet cherry. Unlike the European plum, which is hexaploid, Japanese plum is diploid (2n=2x=16). It originated in China, and its modern breeding history involves interspecific crosses with *P. simonii*, *P. americana*, *P. cerasifera* and other diploid plums (Okie and Ramming, 1999). These complex hybridizations have allowed great genotypic and phenotypic variability.

One of the most valuable traits for consumers is fruit color, which in Japanese plum can range from yellow to red, purple or black. Such variability depends on the accumulation of the water soluble pigment anthocyanin, described to have broad health promoting effects (He and Giusti, 2010). Anthocyanins are synthetized from the flavonoid pathway. The structural genes from this pathway are highly regulated at the transcriptomic level by the MYB-bHLH-WD40 complex (Ramsay and

Glover, 2005; Gonzalez et al., 2008). In the Rosaceae family, R2R3-MYB transcription factors have been described as main determinants of anthocyanin accumulation in fruit, foliage, flower and roots (Lin-Wang et al., 2010). In peach (*Prunus persica* L. Batsch), a region in LG3 clusters 3 MYB genes, named *MYB10.1*, *MYB10.2* and *MYB10.3*, and was found associated with anthocyanin fruit color (Rahim, Busatto and Trainotti, 2014). *MYB10.1* has been proposed as the candidate gene for skin coloration in peach (Tuan et al., 2015; Bretó et al., 2017) and sweet cherry (Starkevič et al., 2015; Jin et al., 2016), and molecular markers have successfully been designed for marker assisted selection (MAS).

In Japanese plum, however, no markers to select for skin color have been validated yet. González et al. (2016) designed a combination of EST-SSR markers targeting *MYB10*, *MYB1* and *bHLH35* genes, however its discrimination power was limited. Salazar et al. (2017) identified a QTL for the trait in LG3 in a F1 population genotyped by sequencing (GBS). Within this region, the SNP better explaining phenotypic variability was close to *MYB10.1*, giving further proofs of the relevance of this region in fruit coloration.

The main goal of this research is to explore the MYB10 genetic variability in Japanese plum to find molecular markers associated with fruit color that are efficient for MAS.

MATERIALS AND METHODS

Plant material and DNA extraction

In this study we have analyzed a collection of 78 Japanese plum accessions and breeding lines. Fruits were phenotyped visually for their color at maturity stage. Genomic DNA was extracted from young leaves using Doyle CTAB method (Doyle and Doyle, 1990).

Primer design on MYB10 genes

In order to identify possible *MYB10* polymorphisms associated with fruit color, we designed a set of primers targeting conserved sites of *PpMYB10.1* (*Prupe.3G163100*) and *PpMYB10.2* (*Prupe.3G163000*) genes (**Figure 1.1**), using the peach Genome v2.0.1 as reference. In total, 3 fluorescently labeled forward and 3 reverse primers were designed (**Table 1.1**), and used in Japanese plum cultivars. Each forward-reverse combination was tested in a reduced panel of accessions with contrasting fruit color. PCR amplification reaction consisted in 10x NH_4 Reaction Buffer (1x), $MgCl_2$ Solution (1.5 mM) and 1U of BIOTAQ DNA Polymerase (BioLine), with 0.2 mM

dNTPs, 0.2 µM of each primer combination (IDT) and MilliQ water to amplify 40 ng of extracted DNA in a 10 µL reaction. Thermocycler conditions were as follows: 1 minute at 94ºC, 35 cycles of 94ºC for 30 seconds, 30 seconds of annealing at 55ºC and 30 seconds of extension at 72ºC, and a last extension step at 72ºC for 5 minutes. Amplified fragments were separated by capillary electrophoresis in a ABI 3130xl Genetic Analyzer, and amplicons length precisely scored, using 500 LIZ™ (GeneScan) dye standard, with GeneMapper Software 5 (Applied Biosystems).

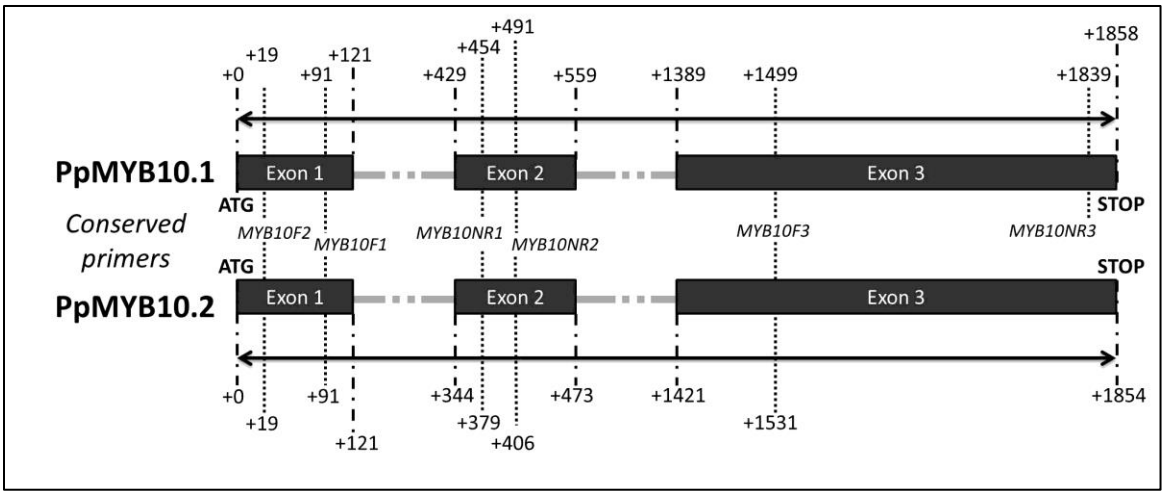


Figure 1.1. Position of each primer designed in both *PpMYB10.1* and *PpMYB10.2* genes. Numbers represent the number of bases from the ATG codon (+0) of each gene and the primer. Dots in introns indicate a much larger length than represented.

Table 1.1. Primer sequences designed in *PpMYB10.1* and *PpMYB10.2* conserved sites and tested in *P. salicina* accessions.

Forward/reverse	Primer name	Sequence (5'-3')	Label
Forward	MYB10F1	GAAAGTGGCACCAAGTTCCT	PET
	MYB10F2	GTGTGAGAAAAGGAGCTT	FAM
	MYB10F3	CCCAAGAAACAATAAAGACC	NED
Reverse	MYB10NR1	CCTTAGTCTACAGCTCTTCC	-
	MYB10NR2	GATATTTGGCTTCAAATAGTTC	-
	MYB10NR3	ACGCTAAAAGAGAAATCACC	-

MYB10 genotyping and association with fruit color

MYB10F2 and MYB10NR2 primer combination was used to genotype the collection of 78 Japanese plum cultivars. Genotyping was performed using the same procedure described above when testing the primers. To detect possible associations between genotypic data and fruit skin and flesh coloration, phenotypes were simplified as a binary trait of presence (red, purple or black colorations) and absence (yellow, green or white colorations) of anthocyanin. For both skin and flesh color, associations between phenotypes and genotypes were done with a Chi-square test.

RESULTS AND DISCUSSION

Primers on MYB10 loci amplify polymorphic alleles

The nine primer combinations used in this work were designed to amplify, in a single PCR reaction, alleles of the two genes, *PsMYB10.1* and *PsMYB10.2*. Each of the primer combinations used in a small panel of accessions amplified between 1 to 7 different polymorphic fragments (**Table 1.2**), with an average number per genotype ranging from 1.29 to 5. The maximum number of alleles per genotype was yielded by the primer pair MYB10F2/MYB10NR2 (7 fragments) suggesting the amplification of more than the 2 genes. This could indicate the amplification of other *MYB10* off-target genes or a duplication of one of the *MYB10* genes in LG3 cluster. LG3 *MYB10* genes duplication has been observed in *Prunus avium* as suggested initially by Starkevič et al. (2015) and confirmed later by Shirasawa et al. (2017) with the genome assembly. Therefore, although further work is required, we may consider a possible duplication of *MYB10.1* in *P. salicina*.

Table 1.2. Primers in *MYB10* genes amplified polymorphic fragments that could be used to search for association with color phenotype. Primer combinations with no amplification and alleles above 500 bp are not listed. MYB10F2 and MYB10NR2 combination amplified more polymorphic alleles and were selected for association studies.

Primer combination		Number of polymorphic alleles per genotype			Total number of alleles	Allele range (bp)
		min	max	Mean		
MYB10F1	MYB10NR1	2	5	3.57	7	252-377
	MYB10NR2	2	5	4.29	7	279-403
	MYB10NR3	1	4	1.29	4	105-411
MYB10F2	MYB10NR1	2	5	3.29	7	323-447
	MYB10NR2	3	7	5.00	9	350-495
MYB10F3	MYB10NR3	0	0	0	1	340

Despite differences in allele sizes, the information provided by most of the alleles amplified by each of the primer combinations was redundant. Among all, MYB10F2/MYB10NR2 was the most informative primer combination, and therefor used for further analysis in the large panel of varieties.

A dominant allele is associated with fruit skin color

We analyzed a panel of 78 Japanese plums phenotyped for skin and flesh color. Regarding skin color, 27 were yellow or green (absence of anthocyanins) and 51 colored; for the flesh color, 50 were yellow, white or green and 28 colored.

The number of polymorphic alleles amplified with the MYB10F2/MYB10NR2 primer combination in the full collection increased to 15. Among them, 10 showed frequency higher than 0.20 and were used to find association with the traits. Chi-square test detected one allele, of 356 base pairs, dominantly associated with anthocyanin accumulation in fruit skin (p-value = 1.03×10^{-18}). Its association was clear, since it was present in all 51 color-skinned cultivars and absent in the remaining 27 varieties of the collection. Two other alleles were found statistically associated to skin coloration (p-values of 1.61×10^{-8} and 4.45×10^{-8}), but in less degree since they could not predict all skin phenotypes. No clear association with absence of anthocyanin pigments in the skin could be found. None of the alleles could explain variability in fruit flesh coloration.

CONCLUSIONS

Allele 356 bp is dominantly associated with skin coloration caused by anthocyanins, and effectively predicts presence of coloration in the *P. salicina* collection analyzed. Further work will be focused on allele cloning and segregation analyses to identify alleles of the same locus that could be associated to absence of coloration, with the objective of designing a single locus codominant marker to select for fruit skin color in Japanese plum breeding programs.

ACKNOWLEDGEMENTS

This work has been funded by FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación/Project Reference (AGL2015-68329-R and RTI2018-100795-B-I00). This work was supported by grant from Ministerio de Ciencia, Innovación y Universidades (Reference BES-2016-079060F) and by the CERCA Programme/Generalitat de Catalunya. We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo

Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533). We acknowledge the Spanish breeding company PLANASA for its support to this research.

Literature cited

Bretó, M.P., Cantín, C.M., Iglesias, I., Arús, P., and Eduardo, I. (2017). Mapping a major gene for red skin color suppression (highlighter) in peach. Euphytica 213, 14. http://doi.org/10.1007/s10681-016-1812-1

Doyle, J. J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Gonzalez, A., Zhao, M., Leavitt, J.M., and Lloyd, A.M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant J 53, 814–827. http://doi.org/10.1111/j.1365-313X.2007.03373.x

González, M., Salazar, E., Castillo, J., Morales, P., Mura-Jornet, I., Maldonado, J., Silva, H., and Carrasco, B. (2016). Genetic structure based on EST–SSR: a putative tool for fruit color selection in Japanese plum (*Prunus salicina* L.) breeding programs. Mol Breeding 36, 68. http://doi.org/10.1007/s11032-016-0491-x

He, J., and Giusti, M.M. (2010). Anthocyanins: Natural Colorants with Health-Promoting Properties. Annu. Rev. Food Sci. Technol. 1, 163–187. http://doi.org/10.1146/annurev.food.080708.100754

Jin, W., Wang, H., Li, M., Wang, J., Yang, Y., Zhang, X., Yan, G., Zhang, H., Liu, J., and Zhang, K. (2016). The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). Plant Biotechnol J 14, 2120–2133. http://doi.org/10.1111/pbi.12568

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T.K., Espley, R.V., Hellens, R.P., and Allan, A.C. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol 10, 50. http://doi.org/10.1186/1471-2229-10-50

Okie, W.R., and Ramming, D.W. (1999). Plum breeding worldwide. HortTechnology 9, 162–176.

Rahim, Md.A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. Planta 240, 913–929. http://doi.org/10.1007/s00425-014-2078-2

Ramsay, N.A., and Glover, B.J. (2005). MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science 10, 63–70. http://doi.org/10.1016/j.tplants.2004.12.011

Salazar, J.A., Pacheco, I., Shinya, P., Zapata, P., Silva, C., Aradhya, M., Velasco, D., Ruiz, D., Martínez-Gómez, P., and Infante, R. (2017). Genotyping by Sequencing for SNP-Based Linkage Analysis and Identification of QTLs Linked to Fruit Quality Traits in Japanese Plum (*Prunus salicina* Lindl.). Front. Plant Sci. 8. http://doi.org/10.3389/fpls.2017.00476

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., and Isobe, S. (2017). The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res. 24, 499–508. http://doi.org/10.1093/dnares/dsx020

Starkevič, P., Paukštytė, J., Kazanavičiūtė, V., Denkovskienė, E., Stanys, V., Bendokas, V., Šikšnianas, T., Ražanskienė, A., and Ražanskas, R. (2015). Expression and Anthocyanin Biosynthesis-Modulating Potential of Sweet Cherry (*Prunus avium* L.) MYB10 and bHLH Genes. PLoS ONE 10, e0126991. http://doi.org/10.1371/journal.pone.0126991

Tuan, P.A., Bai, S., Yaegaki, H., Tamura, T., Hihara, S., Moriguchi, T., and Oda, K. (2015). The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biol 15, 280. http://doi.org/10.1186/s12870-015-0664-5

Chapter 2

Characterization of Japanese plum (*Prunus salicina*) *PsMYB10* alleles reveals structural variation and polymorphisms correlating with fruit skin color

Arnau Fiol[1], Beatriz E. García-Gómez[1], Federico Jurado-Ruiz[1], Konstantinos Alexiou[1,2], Werner Howad[1,2] and Maria José Aranzana[1,2]

[1]Centre for Research in Agricultural Genomics, CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain

[2]Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain

This chapter has been published in Frontiers in Plant Science (June 2021)

DOI: 10.3389/fpls.2021.655267

ABSTRACT

The red to blue hue of plant organs is due to anthocyanins, which are water-soluble flavonoid pigments. The accumulation of these pigments is regulated by a complex of R2R3-MYB transcription factors (TFs), basic-helix-loop-helix (bHLH), and WD-repeat (WDR) proteins (MBW complex). In Rosaceae species, R2R3-MYBs, particularly MYB10 genes, are responsible for part of the natural variation in anthocyanin colors. Japanese plum cultivars, which are hybrids of *Prunus salicina*, have high variability in the color hue and pattern, going from yellow-green to red and purple-blue, probably as a result of the interspecific hybridization origin of the crop. Because of such variability, Japanese plum can be considered as an excellent model to study the color determination in Rosaceae fruit tree species. Here, we cloned and characterized the alleles of the *PsMYB10* genes in the linkage group LG3 region where quantitative trait loci (QTLs) for the organ color have been mapped to other *Prunus* species. Allele segregation in biparental populations as well as in a panel of varieties, combined with the whole-genome sequence of two varieties with contrasting fruit color, allowed the organization of the *MYB10* alleles into haplotypes. With the help of this strategy, alleles were assigned to genes and at least three copies of *PsMYB10.1* were identified in some varieties. In total, we observed six haplotypes, which were able to characterize 91.36% of the cultivars. In addition, two alleles of *PsMYB10.1* were found to be highly associated with anthocyanin and anthocyanin-less skin. Their expression during the fruit development confirms their role in the fruit skin coloration. Here, we provide a highly efficient molecular marker for the early selection of colored or non-colored fruits in Japanese plum breeding programs.

Keywords: MYB10, anthocyanins, fruit color, structural variability, marker-assisted selection, Japanese plum, Asian plum

INTRODUCTION

Anthocyanins are water-soluble flavonoid pigments that confer the purple-blue hue to plant organs. In flowers and fruits, these pigments enhance flower pollination and seed dispersion mediated by animals; in vegetative tissues, they provide tolerance to abiotic stresses including photoprotection after exposure to long-term light stress, resistance to chilling and desiccation, and recovery from mechanical injury (Winkel-Shirley, 2002; Kong et al., 2003; Gould, 2004). The antioxidant activity of anthocyanins has been further linked to human health-promoting effects when incorporated in the diet. The list of beneficial effects includes anticarcinogenic and anti-inflammatory activity and the prevention of cardiovascular diseases, diabetes, and obesity (Khoo et al., 2017).

Anthocyanins are synthesized through a complex regulatory mechanism for the final appropriate pigment levels to meet the demands of plant development and environmental responses (Albert et al., 2014a,b). In eudicots, biosynthetic genes of anthocyanins are mostly transcriptionally activated by a complex consisting of R2R3-MYB transcription factors (TFs), basic-helix-loop-helix (bHLH), and WD-repeat (WDR) proteins (MBW complex), that induces the expression of several genes in the anthocyanin biosynthesis pathway by binding directly to their promoters (Koes et al., 2005; Gonzalez et al., 2008). Allelic variants of R2R3-MYB genes, including changes in the promoter and coding regions, or even a variation of methylation levels in the promoter and gene bodies, may produce significant changes in the biosynthesis of anthocyanins, causing a variation in fruit coloring (Espley et al., 2007; Telias et al., 2011; Wang et al., 2013; Lü et al., 2018) indicating that R2R3-MYBs are the keys to specify the action of the MBW complex. R2R3-MYB TFs are activated by developmental or environmental signals (Feng et al., 2013; Vimolmangkang et al., 2014; An et al., 2020a). Likewise, the TFs involved in light and hormone signaling pathways may regulate the biosynthesis of anthocyanins directly or indirectly (Shin et al., 2013; Wang et al., 2018; Zhang et al., 2018; An et al., 2020b; Li et al., 2020). MYB TFs can also act as a negative regulator of anthocyanin levels by repressing the expression of anthocyanin genes, possibly by recruiting MBW complexes (Albert et al., 2009; Zhou et al., 2019) or by retaining the ability to form an MBW complex and target DNA motifs when the MYB alleles are truncated (Paz-Ares et al., 1990; Gonzalez et al., 2008; Velten et al., 2010).

Rosaceous fruits such as apples, pears, peaches, apricots, plums, cherries, and strawberries are broadly considered as a source of anthocyanins. In these fruits, such pigments are present in the skin and the flesh, overlying chlorophylls and carotenoids from the fruit development till ripening. External factors such as sunlight, cold temperatures, and crop management are the keys in color development and patterning. Fruit color has an important impact on the choice of consumer, but, while the consumers are interested in this and other traditional quality traits related to flavor, they also demand attributes including nutritional quality. Skin fruit color is a readily observable trait for variety identification. In some fruits, red color is associated with ripeness and better taste and flavor, and red fruits are also highly valued for their content of healthy compounds. Therefore, there is considerable interest in breeding these crops to obtain new varieties bearing fruits with diverse colors, hues, and patterns, while enhancing fruit nutritional quality (Ogah et al., 2014; Panche et al., 2016; García-Gómez et al., 2021). Such breeding efforts have been accompanied by a growing

scientific interest in understanding the molecular mechanisms underlying the biosynthesis and accumulation of anthocyanins.

The developmental regulatory network and specific regulators of the biosynthesis of anthocyanins have been studied in most of the major rosaceous fruit species. Structural genes involved in the early and late anthocyanin biosynthetic path have been isolated (Honda et al., 2002; Ubi et al., 2006; Fischer et al., 2007; García-Gómez et al., 2020) and the MYB genes responsible for the observed variation in the color and pattern in plant organs. Apple (*Malus x domestica*) is a rosaceous crop where the anthocyanin regulation has been more studied, where three MYB genes (*MdMYB1*, *MdMYBA*, and *MdMYB10*) controlling the biosynthesis of anthocyanins in the skin and/or the fruit cortex have been isolated and their function is validated (Takos et al., 2006; Ban et al., 2007; Espley et al., 2007). Their high amino acid sequence identity and same mapping position in the linkage group LG9 (Chagné et al., 2007; Kumar and Pandey, 2013) suggest that all three are alleles of a single gene (Lin-Wang et al., 2010; Telias et al., 2011). Two additional genes, *MdMYB110a* [paralog of *MdMYB10* (Chagné et al., 2013)] and *MdMYB114* (Jiang et al., 2019b), have been found to be highly expressed in the cortex and the fruit skin, respectively, in correlation with anthocyanin levels. The genomic organization of flavonoid genes is comparable between *Malus* and its related genera *Pyrus*, which has allowed the elucidation of the main flavonoid pathway in *Pyrus* using a homology-based cloning approach (Fischer et al., 2007). In the transcriptional level, *Pyrus sp. PyMYB10*, *PyMYB114*, *Pyrus communis PcMYB10*, and *Pyrus bretschneideri PbMYB10b* and *PbMYB9* act as activators of the anthocyanin pathway (Feng et al., 2010; Zhai et al., 2016; Yao et al., 2017; Zhang et al., 2020). Several studies on both apple and pear sport mutants differing in fruit skin color intensity or pattern have identified hypermethylated regions in the promoter or gene bodies of some of these MYB TFs, most likely preventing their expression, causing a variation in fruit color (Telias et al., 2011; Xu et al., 2012; Wang et al., 2013; El-Sharkawy et al., 2015).

In apples and pears, the *MYB10* mapping region in LG9 is collinear with a region in *Prunus* LG3 (Illa et al., 2011) where some traits related to the plant organ color have been mapped (Espley et al., 2007; Sooriyapathirana et al., 2010; Socquet-Juglard et al., 2013; Donoso et al., 2016). This region contains a cluster of MYB10 genes, although their number differs between the *Prunus* species. In peach (*Prunus persica*) and almond (*Prunus dulcis*), there are three MYB10 genes (*PpMYB10.1*, *PpMYB10.2*, and *PpMYB10.3*) in this cluster (Verde et al., 2017; Alioto et al., 2020), while this number rises to five annotated MYBs in the syntenic region of sweet cherry (*Prunus avium*)

(Shirasawa et al., 2017). In these species, MYB10 TFs have been described as positive regulators of the biosynthesis of anthocyanins, and allelic variants have been found to be highly correlated with a variation in fruit color (Tuan et al., 2015; Jin et al., 2016; Bretó et al., 2017; Guo et al., 2020). Similarly, Wang et al. (2020) identified an *MYB10* colorless specific allele in strawberry, a rosaceous crop with a different fruit type, indicating that this mechanism is conserved within the Rosaceae family.

Within the rosaceous crops, Japanese plum is among those with the highest fruit color variation, in both tonality and pattern, especially in the skin, where it can range from anthocyanin-less green and yellow to red, purple, or blue hues. In addition, the skin pigmentation does not necessarily fully cover the fruit, forming patterns with the visible fruit flesh color as a background. Molecular markers suitable for marker-assisted selection (MAS) would be of great value for breeders in order to predict at the seedling stage the color of the fruit that will be produced in 3–4 years. Despite several efforts to map and identify the markers associated with the Japanese plum fruit color, reliable markers for their use in breeding programs are still lacking. González et al. (2016c) designed three EST-SSR markers targeting *PsMYB10*, *PsMYB1*, and *PsbHLH35* genes to determine the genetic structure of 29 Japanese plum cultivars with different skin colorations, finding that all the yellow-skinned cultivars grouped in a cluster together with some red-skinned cultivars. Later, Salazar et al. (2017) applied GBS for the quantitative trait locus (QTL) mapping of several fruit quality traits in a Japanese plum F1 population, mostly with red or purple skin. The red/purple skin color traits were mapped to LG3 and LG4. One single-nucleotide polymorphism (SNP), close to the *PsMYB10.1* gene, largely explained the tendency for purple fruit skin. All together suggest the role of the LG3 MYB10 region in the determination of the fruit color in Japanese plum.

Japanese plum is a species with a diploid genome derived from the hybridization of *Prunus salicina* with several other species of the *Prunophora* subgenus, which increases the complexity of the analysis of the genes responsible for color variation in the absence of a reference genome. To date, databases include a number of *MYB10* read sequences obtained from Rosaceae crops, including Japanese plum. Their alignment and phylogenetic analysis have shed some light on MYB variability within the family (Lin-Wang et al., 2010); however, a high homology between the MYB factors, their genome localization in a cluster, and high variability especially in non-coding regions hamper its separation into genes or alleles of the same gene.

The main objective of this study is to analyze the genetic variability of the LG3 MYB10 genes in the Japanese plum germplasm and to find possible correlations with fruit anthocyanin color. Here, we have identified and assigned the allelic variants to different MYB10 genes by means of whole-genome sequencing, MYB10 allele cloning, phylogeny, and phasing through progeny analysis. We have found high genetic variability in the Japanese plum LG3 MYB10 gene cluster, which contains three copies of the *MYB10.1* gene in at least some Japanese cultivars. Through the allele and haplotype association analysis, we have identified haplotypes that can predict the fruit skin color with high efficiency in seedlings. These haplotypes can be easily obtained with a single PCR reaction, being highly useful for MAS. Possible reasons and implications of different copies of *MYB10.1* genes and allele combinations are discussed.

MATERIALS AND METHODS

Plant Material, Phenotyping, and Nucleic Acid Extraction

The plant material analyzed was from accessions and progenies. The accessions were 81 Japanese plum advanced breeding lines (ABL) from a commercial breeding program and 31 commercial cultivars. The progenies were six F1 biparental families obtained after crossing 10 of the advanced breeding cultivars: C9 × C6 (83 individuals, 49 with fruits phenotyped; from now on referred to as P1), C19 × C26 (111 individuals, 79 of them phenotyped; P2), C14 × C4 (48 individuals, 37 with fruits phenotyped; P3), C11 × C8 (43 individuals, 30 with fruits phenotyped; P4), C8 × C5 (64 individuals, 10 with fruits phenotyped; P5), and C11 × C66 (33 individuals, 14 with fruits phenotyped; P6). The parental lines and progenies were grown open field, in the warm weather conditions of Huelva (Spain), and phenotyped for the fruit skin and flesh color at maturity, while the color information for commercial varieties (CV) was obtained from descriptors of breeder. Skin color descriptors were yellow, green, mottled (yellow or green skin mottled over colored background, **Supplementary Figure 2.1**), pale red, red, purple, and black. Flesh color descriptors were white, green, yellow, orange, red, and purple. The fruit color of all materials is presented in **Supplementary Data 2.1 and 2.2**.

The DNA of accessions and progenies was isolated from young, fresh, or frozen leaves ground with liquid nitrogen by using the Doyle CTAB method (Doyle and Doyle, 1987). RNA was extracted from the skin of two commercial cultivars with contrasting fruit coloration: ‘Black Gold’ (BG, dark red skin) and ‘Golden Japan’ (GJ, yellow skin). Fruits were collected in triplicates at three different stages: S1 (an immature fruit with green skin), S2 (an intermediate-mature fruit, initial appearance of

coloration), and S3 (a maturity stage fruit, full coloration). The fruit skin was separated from the flesh, immediately frozen in liquid nitrogen, and kept at −80°C. Samples were ground in liquid nitrogen, and the RNA was extracted by using the Maxwell RSC Plant RNA Kit (Promega, Madison, WI, USA), and then further DNase-treated by using the TURBO DNA-free Kit (Invitrogen, Carlsbad, CA, USA). Quality and quantity of DNA and RNA were checked in a NanoDrop ND-1000 Spectrophotometer and in 0.8% agarose gels.

MYB10 Allele Amplification

A unique primer pair designed in the conserved regions of the *P. persica MYB10.1* (*Prupe.3G163100*) and *MYB10.2* (*Prupe.3G163000*) genes was used to genotype all plant materials. Forward (MYB10F2:GTGTGAGAAAAGGAGCTT) and reverse (MYB10NR2:GATATTTGGCTTCAAATAGTTC) primers hybridize in exon 1 (20 bases downstream the start codon of both genes) and in exon 2 (40 bases downstream its start position), respectively (**Supplementary Figure 2.2**).

Each PCR reaction contained 1 × NH4 and 1.5 mM MgCl2 buffers for 1U BioTaq Polymerase (Bioline), 0.2 mM dNTP mix, 0.2 µM of each MYB10F2 (fluorescently labeled) and MYB10NR2 primer, 40 ng of DNA, and MilliQ water to a 10 µl total reaction. PCR conditions were: 94°C for 1 min, 30 cycles of 15 s at 94°C, 15 s at 55°C, and 30 s at 72°C, with a final elongation step of 5 min at 72°C. Amplified fragments were separated by capillary electrophoresis using the ABI Prism 3130xl Genetic Analyzer. The GeneMapper 5.0 software was used to visualize the amplified fragments and precisely size them by correlation to the GeneScan™ 500 LIZ™ size standard.

SSR Genotyping, Linkage Mapping, and Identification of Haplotypes

Families P1, P2, and P3 were genotyped with LG3 SSR markers (MA039a, UDAp-496, PaCITA10, EPPCU0532, BPPCT007, and BPPCT039). PCR reaction, thermocycler conditions, and capillary electrophoresis were as before, with the annealing temperature adapted to each marker (**Supplementary Table 2.1**).

LG3 genetic maps of the pollen and seed parents were constructed by using JoinMap 5.0 (Stam, 1993). For the analysis, each of the multiple bands amplified by MYB10F2/MYB10NR2 was entered as an independent dominant allele. Groups were established with a logarithm of the odds (LOD) score of 3.0, and map distances were calculated with the mapping function of Kosambi (Kosambi, 2016).

To phase the *MYB10* polymorphic alleles into allelic combinations, and therefore into haplotypes (H), we inspected their segregation in the six progenies. Haplotypes for the ABL and CV were inferred with the PHASE v2.1 software (Stephens and Donnelly, 2003; Scheet and Stephens, 2006).

Amplicon Cloning and Sequencing

All PCR fragments with frequencies higher than 5% in a panel of parental lines were cloned. PCR reactions were purified by using the High Pure PCR Product Purification Kit (Roche, Basel, Switzerland). Amplification was checked in 2% agarose TAE gel, and the amplified fragments were ligated into the Promega pGEM®-T Easy vector and transformed with JM109 *Escherichia coli* cells according to the kit instructions. Colonies were screened by direct colony-PCR using MYB10F2/MYB10NR2 genotyping and capillary electrophoresis. For colonies carrying one fragment of interest, colony-PCR was repeated with common vector T7 and SP6 primers, using the same PCR reaction and thermocycler conditions. PCR products were purified, checked on agarose gel, and sequenced with T7 and SP6 primers by the fluorescent dye termination detection in the ABI 3730 DNA Analyzer.

Diversity and Phylogeny Analysis

Sequences of the fragments cloned were aligned, vector trimmed, and manually edited by using the Sequencher 5.0 software (Gene Codes Corporation, Ann Arbor, MI, USA). Aligned sequences were imported into Jalview v2 (Waterhouse et al., 2009) for the alignment visualization. Nucleotide basic local alignment search tool (BLASTn) software (Altschul et al., 1990) was run on the individual sequences to find its closest homologous R2R3 MYB TF in the NCBI GenBank database. The same software was used in the GDR webpage (Jung et al., 2019) by using Peach Genome v2.0.a1 as the database to find the homologous *P. persica* genomic region. Annotated genes were visualized in JBrowse (Buels et al., 2016). Gene sequences of the closest matches were imported into the Sequencher 5.0 file, with mismatches manually edited. The full contig was exported to MEGA X (Kumar et al., 2018) for the phylogeny analysis. Sequences were clustered by using the UPGMA method (Sneath and Sokal, 1973) using the Tamura–Nei model (Tamura and Nei, 1993) and 500 bootstrap (Felsenstein, 1985) replications. For exon and intron discrimination, cloned alleles were aligned to the *R2R3-MYB10* mRNA sequences of *P. persica* (*Prupe.3G163100.1*), *P. dulcis* (EU155159.1), *P. avium* (GU938689.1), *P. salicina* (KX349091.1), *Prunus domestica* (EU153580.1), and *P. domestica subsp. insititia* (EU153579.1). DNAsp software v6 (Rozas et al., 2017) was used for the nucleotide diversity analysis, considering gaps. Exonic sequences were translated *in silico* by

using ExPASy Translate (Artimo et al., 2012). The resulting amino acid sequences were aligned and compared by using Clustal Omega (Sievers et al., 2011) and visualized in Jalview v2.

Whole-Genome Sequencing, Gene Cloning, and Analysis of Variability

High-quality DNA isolated from the leaves of two varieties with a contrasting fruit color phenotype was sent to CNAG (Centre Nacional d'Anàlisi Genòmica, Barcelona) for the paired-end library preparation and processed with the Illumina HiSeq 2000 Sequencer. Adapter removal and quality-based trimming of the raw resequencing data were done with Trimmomatic version 0.36 (Bolger et al., 2014). FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for read quality control before and after trimming. High-quality reads were mapped to the *P. persica* genome v2.0.a1 (Verde et al., 2017), *P. dulcis* Texas Genome v2.0 (Alioto et al., 2020), and *P. avium* Genome v1.0 (Shirasawa et al., 2017) by using the Burrows–Wheeler Aligner (BWA) (Li and Durbin, 2009), and the resulting alignment files were sorted and filtered by discarding multi-mapped reads. General statistics, depth, and breadth of coverage of sequencing libraries were obtained by using Samtools (Li et al., 2009). Normalized sequencing depth with respect to the mean number of reads of the whole-genome and MYB region was visualized with Circos version 0.69.9 (Krzywinski et al., 2009). The MYB10 region was considered from the first base of 5' untranslated region (UTR) of the *MYB10.2* gene to the end of *MYB10.3* 3' UTR. (LG3:18183274–18256025 in peach, chr3: 12147156–12215700 in sweet cherry, and Pd03:15372325–15392346 in almond genomes). The alignment against the *P. persica* and *P. avium* genomes was used to design PCR primers to extend the allelic bands corresponding to the *PsMYB10.1* gene and upstream region. Allele a356 sequence was obtained from the accessions C46 (with H1) and C6 (with H3); a470 from C57 (H2) and C20 (H4), and a243 from C50 (H_i10/H_i11). All primers were manually designed based on the sequence read coverage, primer optimal properties, and specificity using the BLASTN function. Primer sequences and annealing temperature used are listed in **Supplementary Table 2.2**. Purified PCR products from red and yellow accessions were ligated to a pCR™2.1-TOPO® vector. Ligation reactions were as above with a higher PCR extension time (2 min) for longer fragments. From the whole-gene sequencing, *in silico* translated proteins were aligned and compared for polymorphism identification and to search for the signature motifs conserved in Rosaceae anthocyanin-promoting R2R3-MYB genes. Upstream sequences were aligned and their *cis*-regulatory motifs were explored by using the PLACE database (Higo et al., 1999).

Association of MYB10 Alleles and Haplotypes With Fruit Color and Validation in a Panel of CV

The association of the alleles and haplotypes obtained in the 81 ABL with the skin and flesh fruit color was evaluated through a Chi-squared test. Phenotypes were binary classified into presence (for red, purple, and black colors) or absence (for mottled, white, green, yellow, and orange) of anthocyanin-based color on the fruit skin and flesh. The MYB10 primer pair was tested in a panel of 31 commercially cultivated Japanese plum varieties to validate the results (**Supplementary Data 2.2**).

Gene Expression Analysis

The RNA from the fruit skin of two cultivars was reverse transcribed with oligo(dt)20 primer by using the PrimeScript RT-PCR Kit (TaKaRa, Dalian, China). Based on the results of the previous full-gene sequencing, primers M101_RT_F and M101_RT_R were designed to amplify and sequence the whole coding sequence of the alleles a356 and a470 (**Supplementary Table 2.2**). The specificity of the primers was tested by PCR using genomic DNA from the two cultivars. Later, PCR with complementary DNA (cDNA) as a template was used to check the expression of a356 and a470. Expression of the *MON* reference gene was used as a positive control (Kim et al., 2015). Amplification products were purified and sequenced by using the same primer pair. The results were imported into Sequencher 5.0 and aligned to full-gene sequences of a356 and a470 for their identification. Sequencing results were incorporated into the previously created contig containing messenger RNA (mRNA) sequences.

RESULTS

MYB10 Homology Analysis and Allele Mapping

To amplify and isolate *PsMYB10* allelic bands, we used the PCR primers designed in the conserved regions of *PpMYB10.1*, *PpMYB10.2*, and *PpMYB10.3*. In a panel of 81 Japanese plum ABL, we identified 16 alleles with size ranging from 243 bp to 500 bp, with only (a466) being monomorphic. In all, the minimum number of alleles observed per genotype was three and the maximum was eight, with a mean of 5.7 alleles, indicating the amplification of more than three *MYB10* loci for a diploid species.

The alleles with a frequency above 5% (12 in total including the monomorphic a466) were cloned and sequenced. Allele sequence alignment identified SNPs as well as InDels. Nucleotide diversity

was π = 0.204, with most variations in the intronic region (πExon = 0.071 vs. πIntron = 0.244) (**Figure 2.1A**, **Supplementary Data 2.3**).

The BLAST analysis of these sequences against *P. persica* and the other *Prunus* sequences available in the NCBI database showed the best hit of all of them with MYB10 genes (**Supplementary Table 2.3**). Seven (a243, a350, a356, a454, a462, a470, and a473) were homologous to the peach *PpMYB10.1* gene sequence (*Prupe.3G163100*), with the closest homology to the *P. domestica subsp. insititia MYB10* (*PiMYB10*) gene sequence EU153579.1 (**Supplementary Table 2.3**). Only the monomorphic a466 allele was homologous to *PpMYB10.2* (*Prupe.3G163000*), with high homology to the *P. domestica MYB10* (*PdmMYB10*) gene sequence EU153580.1. The four remaining alleles (a443, a477, a492, and a495) were homologous to *PpMYB10.3* (*Prupe.3G163300*), with the closest homology to the *P. avium MYB10* (*PaMYBA1*) gene sequence EU153580.1.

Consistent with the BLAST results, the phylogeny analysis of the amplified alleles together with orthologous *Prunus* MYB10 genes identified three groups in agreement with the peach *PpMYB10.1*, *PpMYB10.2*, and *PpMYB10.3* gene organization (**Figure 2.1B**). The seven alleles homologous to *PpMYB10.1* were clustered with the *PiMYB10* gene sequence EU153579.1 and with the *PaMYB10* gene sequence GU938680.1. The monomorphic allele homologous to *PpMYB10.2* was clustered with the *MYB10* sequences of *P. dulcis* (EU155159.1), *P. domestica* (EU153580.1), *Prunus cerasifera* (EU153583.1), and *P. salicina* (EU155161.1). However, the four alleles homologous to *PpMYB10.3* were clustered only with this peach gene, in contrast to what was expected from the BLAST analysis where best hits were found with *P. avium*. This discrepancy is attributable to the low query coverage of these alleles in the BLAST analysis (≤62%) (**Supplementary Table 2.3**).

Nucleotide diversity was higher across the *MYB10.1* allele than the *MYB10.3* allele (π = 0.066 and π = 0.007, respectively) (**Figure 2.1A**, **Supplementary Data 2.3**). The deduced amino acid sequence revealed that all alleles encoded a peptide, which included a characteristic R2 domain (**Figure 2.2**). Changes were observed in 19 out of 54 amino acid positions, including deletions and conservative substitutions. A large number of variants in the exons were synonymous, with average Ka/Ks = 0.319 for the 66 allele pair comparisons. Amino acid sequences of PiMYB10, a350, a356, a454, a470, and a473, were identical and shared 98.15% identity to PpMYB10.1, while the other MYB10.1 alleles, a243 and a462, shared 98 and 90.38% sequence identity, respectively, with the previous group (**Supplementary Data 2.3**). The amino acid sequence of a466, homologous to PpMYB10.2, was

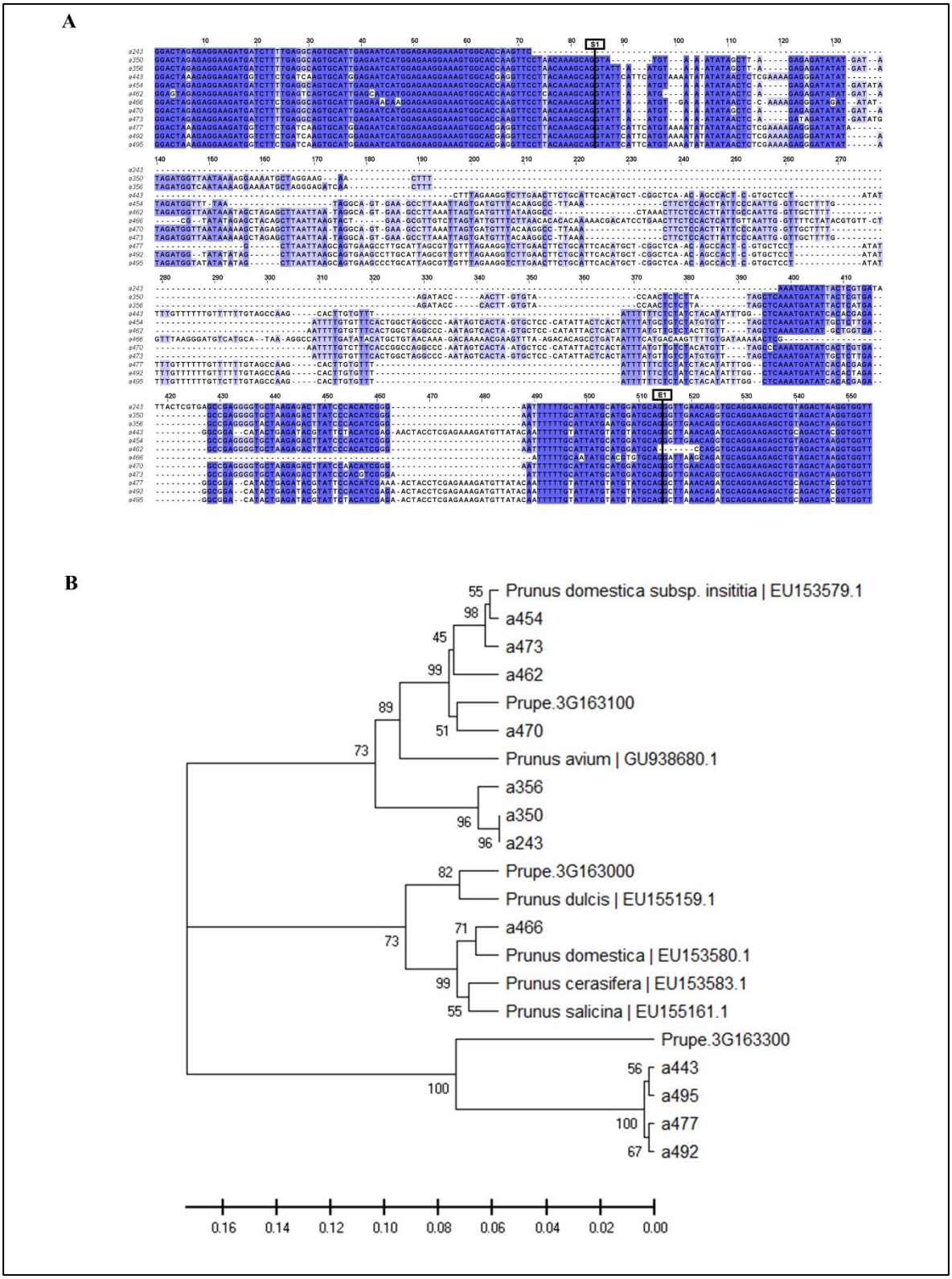


Figure 2.1. **(A)** View of the alignment of the 12 cloned allele sequences (a243 to a495). Intron starts at position 85 (S1) and ends at position 514 (E1). **(B)** Phylogeny tree grouping the cloned sequences with other MYB10 *Prunus* sequences from the NCBI database and the genes of peach *MYB10.1* (*Prupe.3G163100*), *MYB10.2* (*Prupe.3G163000*), and *MYB10.3* (*Prupe.3G163300*).

closer to *P. domestica* (98.15% identity) than to the peach (94.44%) while the remaining alleles shared 100% protein identity between them and 92.59% compared to the PpMYB10.3.

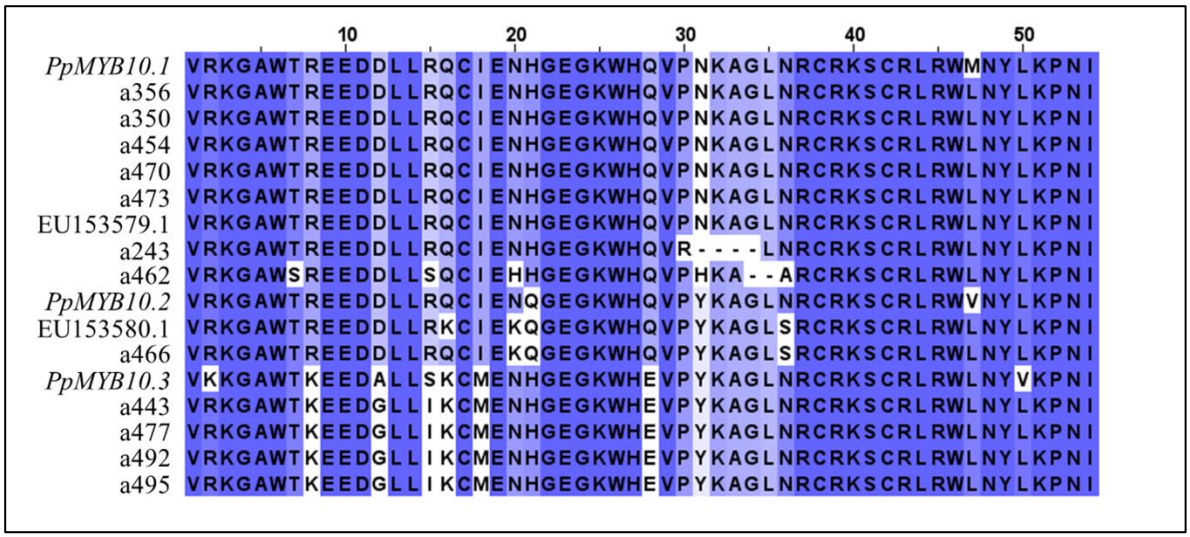


Figure 2.2. Deduced amino acid sequence of the cloned alleles (a243 to a495) compared to the peach PpMYB10, *Prunus domestica subsp. insititia* (EU153579.1) and *P. domestica* (EU153580.1) R2R3-MYB10 translated sequence.

To confirm that the alleles corresponded to the LG3 MYB10 genes, three families (P1, P2, and P3) were genotyped with the MYB10 primer pair plus five LG3 SSR markers. All alleles mapped between the markers MA039a and/or UDAp-496 and PaCITA10 (**Supplementary Figure 2.3**), which are in agreement with the position of the peach MYB10 gene cluster.

Assignation of MYB Alleles Into Haplotypes

Considering that two alleles segregating in phase cannot be allelic of the same gene, we evaluated the segregation of the alleles in 382 individuals of 6 biparental families. We identified six allele combinations segregating in phase, i.e., determining haplotypes (H). Sorting the alleles by their similarity with the *PpMYB10.1-3* genes and size, these were H1 with the alleles a350-a356-a454-a466-a492, shared by C4, C5, C8, C9, and C26; H2 with the alleles a470-a466-a492, shared by C4, C8, C19, C26, and C66; H3 with a356-a462-a473-a466-a495, shared by C6 in homozygosis and by C5, C11, C14, and C66; H4 with alleles a462-a470-a466-a443, shared by C14 and C19; H5 with a470-a466-a477, in C9; and H6 with a462-a466-a477, in C11 (**Figure 2.3A**, **Supplementary Figure 2.4**).

These six haplotypes explained the genotype of 74 (91.36%) of the ABL, while five additional inferred haplotypes (H_i7–11) were required to explain the genotype of the seven remaining ABL (**Figure 2.3B**). In six of them (7.41%), the H_i was found to be in combination with an observed H, while C50

was unique with the two haplotypes inferred. All H_i contained one or several of the alleles with a frequency ≤5%, and were, therefore, discarded for cloning, sequencing, and homology analysis (haplotype combinations for all ABL are shown in **Supplementary Data 2.1**).

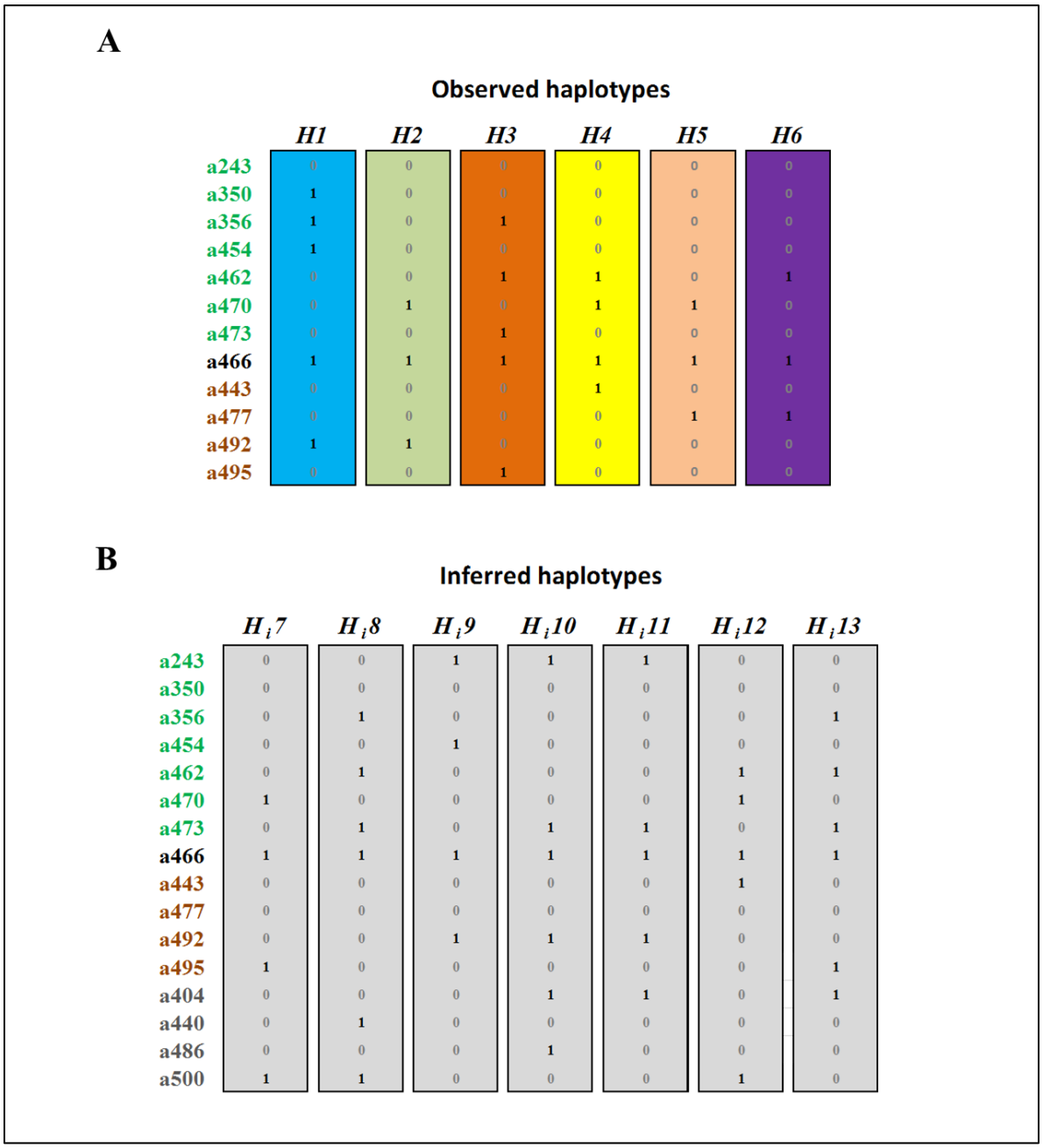


Figure 2.3. Haplotypes with (1) or without (0) each MYB10 allele, grouped and colored according to their homology with *PpMYB10.1*, *PpMYB10.2*, and *PpMYB10.3*. **(A)** Haplotypes identified in six F1 progenies. **(B)** Haplotypes inferred from the collections of advanced breeding lines (ABL) (H_i7–H_i12) and commercial varieties (CV) (H_i12, H_i13).

The combination of haplotype and allele homology data revealed that the haplotypes H1, H3, and H_i8 contain three *MYB10.1* alleles; haplotypes H4, H_i9, H_i10, and H_i11 contain two alleles, and H2, H5, H6, and H_i7 contain only one allele. This indicates multiple copies of *MYB10.1*, at least in some Japanese plum cultivars. By contrast, only one *MYB10.2* and one *MYB10.3* allele per haplotype were identified.

High-Throughput Sequence Analysis of the MYB10 Region

The accessions C20, with yellow skin and flesh and C46, with red skin and flesh, were selected for whole-genome sequencing. The number of reads obtained was 34.671 M for C20 and 117.115 M for C46. Reads were mapped against the reference genomes of peach (Verde et al., 2017), almond (Alioto et al., 2020), and sweet cherry (Shirasawa et al., 2017). For each of the two cultivars, the whole genome as well as along LG3, the depth was similar for all three alignments (**Supplementary Table 2.4**, **Figure 2.4**); however, it varied considerably along the MYB10 region depending on the species genome. Such differences were observed in both genic and intergenic regions. Depth in the MYB10 region was higher than that of the observed genome width when aligned against a peach and an almond, mainly due to a high number of reads aligned in the genic regions (1.7 times higher in C20 and 1.8 times in C46). In contrast, depth in the MYB10 region was lower, both in the genic and the intergenic regions, when aligned against a sweet cherry. The breadth of coverage was also higher in the MYB10 region than genome width in the peach and almond alignments, while the opposite was observed when aligned against a sweet cherry. This data (higher depth in a high breadth coverage scenario) are consistent with a higher number of copies of the MYB10 genes in C20 and C46 compared with the peach and almond reference genomes.

Gene cloning

Whole-genome sequence alignments against the peach and almond reference sequences were used to design primers to fully amplify *PsMYB10.1*, *PsMYB10.2*, and *PsMYB10.3* alleles in the accessions selected according to their haplotypes. As a result, we obtained the whole-gene sequence corresponding to the a356 allele in two haplotypes (a356-H1 and a356-H3), to the a470 allele also in two haplotypes (a470-H2 and a470-H4), as well as the a466-H1 and a492-H1 alleles. The remaining alleles could not be isolated with the primers designed for *PsMYB10.1*.

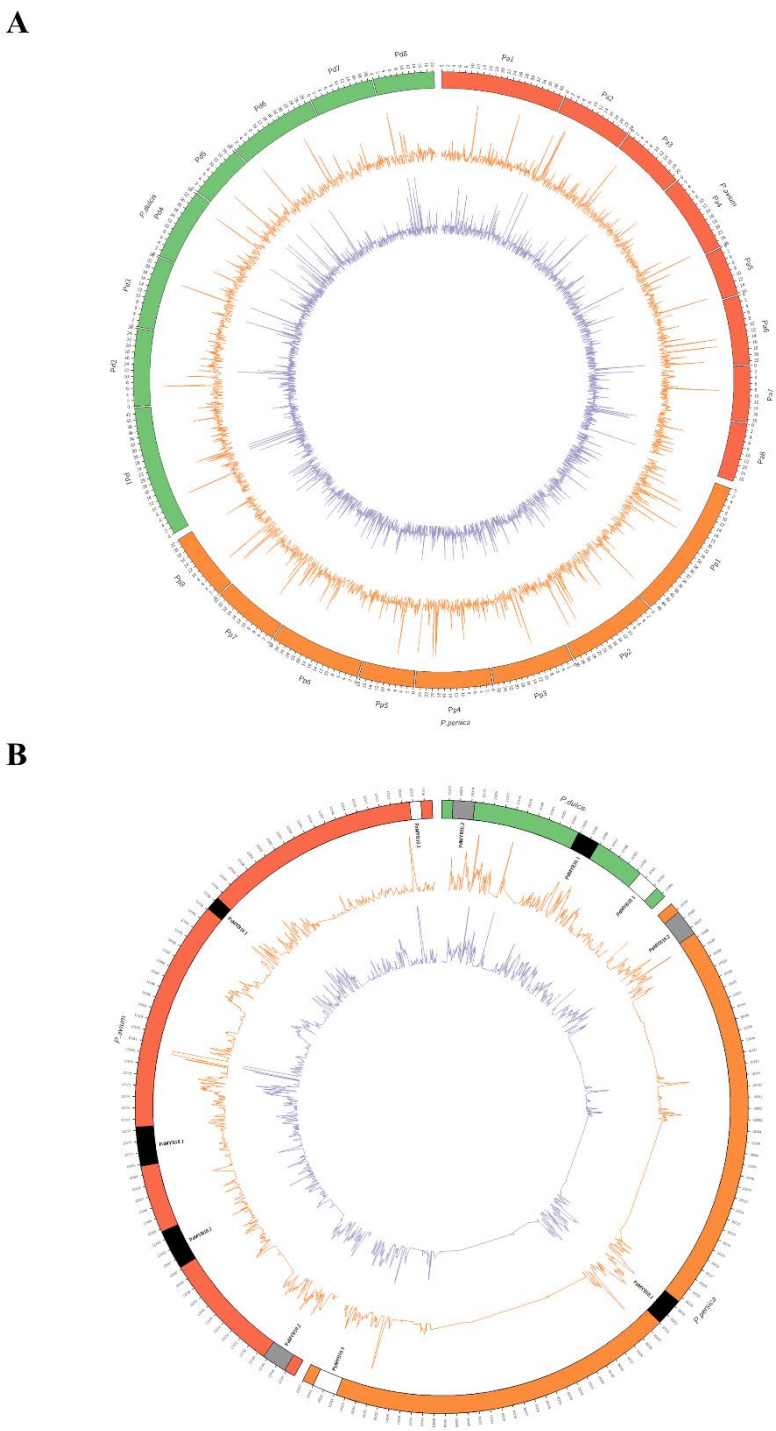


Figure 2.4. Circular plots of the sequencing depth of C20 (orange line) and C46 (purple line) alignments against the genomic reference sequences of *Prunus avium* (Shirasawa et al., 2017), *Prunus persica* (Verde et al., 2017), and *Prunus dulcis* (Alioto et al., 2020). **(A)** Whole-genome alignment (genomic units in megabases) and **(B)** alignment of the chromosome 3 MYB10 region (genomic units in kilobases). MYB10 genes (*MYB10.1* in black, *MYB10.2* in gray, and *MYB10.3* in white) are represented in their corresponding genomic position for each reference species. Depth was normalized for a better visualization and sequence comparison. The image with full resolution can be found online: https://doi.org/10.3389/fpls.2021.655267

Amplicon size of a356-H1, a470-H2, and 470-H4 was 1.5 kb while that of a356-H3 was ca. 3 kb due to a larger intron 2 (**Supplementary Data 2.4**). Their CDS encoded a 239 amino acid protein. Amino acid changes were observed at eight positions, with two at R3 but none at R2 (**Figure 2.5A**). While both a470 alleles were identical, six amino acid substitutions were found between the two a356 alleles, four with residues with biologic similarity. Motifs and amino acids conserved among Rosaceae anthocyanin-promoting R2R3-MYB genes, described in Lin-Wang et al. (2010) and Zhou et al. (2018), were detected in all a356 and a470 sequences. Amplicon size of a466 and a492 was 2 and 3 kb and encoded the proteins of 243 and 160 amino acids, respectively.

We took advantage of a high homology between the Japanese plum and a sweet cherry at the intergenic MYB10 regions to amplify and sequence the upstream regulatory region of the *PsMYB10.1* gene. We obtained up to 1,850 bp upstream alleles a243-H$_i$10, a356-H1, a356-H3, a470-H2, and a470-H4. The sequences revealed additional polymorphisms (**Figure 2.5B**), among them, an insertion of 8 bp in exon 1 of a243 before the hybridization site of the primer initially used to obtain the short *PsMYB10* allelic bands (MYB10F2) leading to a STOP codon 4 bp after the insertion. In addition, we identified, in a356 and a243 contigs, a 44 bp insertion containing a G-box binding motif 1,208 bp upstream of the gene start codon, and two polymorphisms in the R1-motif (described in Espley et al., 2009) of a470 with an SNP and a 5 bp polymorphism.

We identified the polymorphism described by Salazar et al. (2017) (SNP S3_12879559) 128 bp upstream of the start codon of *MYB10.1*. In this position, all the alleles except a356-H1 carried the nucleotide T.

Haplotypes Associated With Skin Color

In the ABL collection, 65.4 and 38.2% of the accessions produced fruits with anthocyanin colors in the skin or flesh, respectively (**Supplementary Figure 2.5**). The 11 polymorphic alleles with a frequency higher than 5% were tested for association with skin and flesh color traits in this collection through a Chi-squared test. The allele showing the highest association with skin color was a356 ($p = 1.96 \times 10^{-18}$) (**Supplementary Table 2.5**). This allele was absent in all varieties with yellow or green skin but present in 52 out of 53 skin-colored ones. No alleles were found to be associated with fruit flesh color.

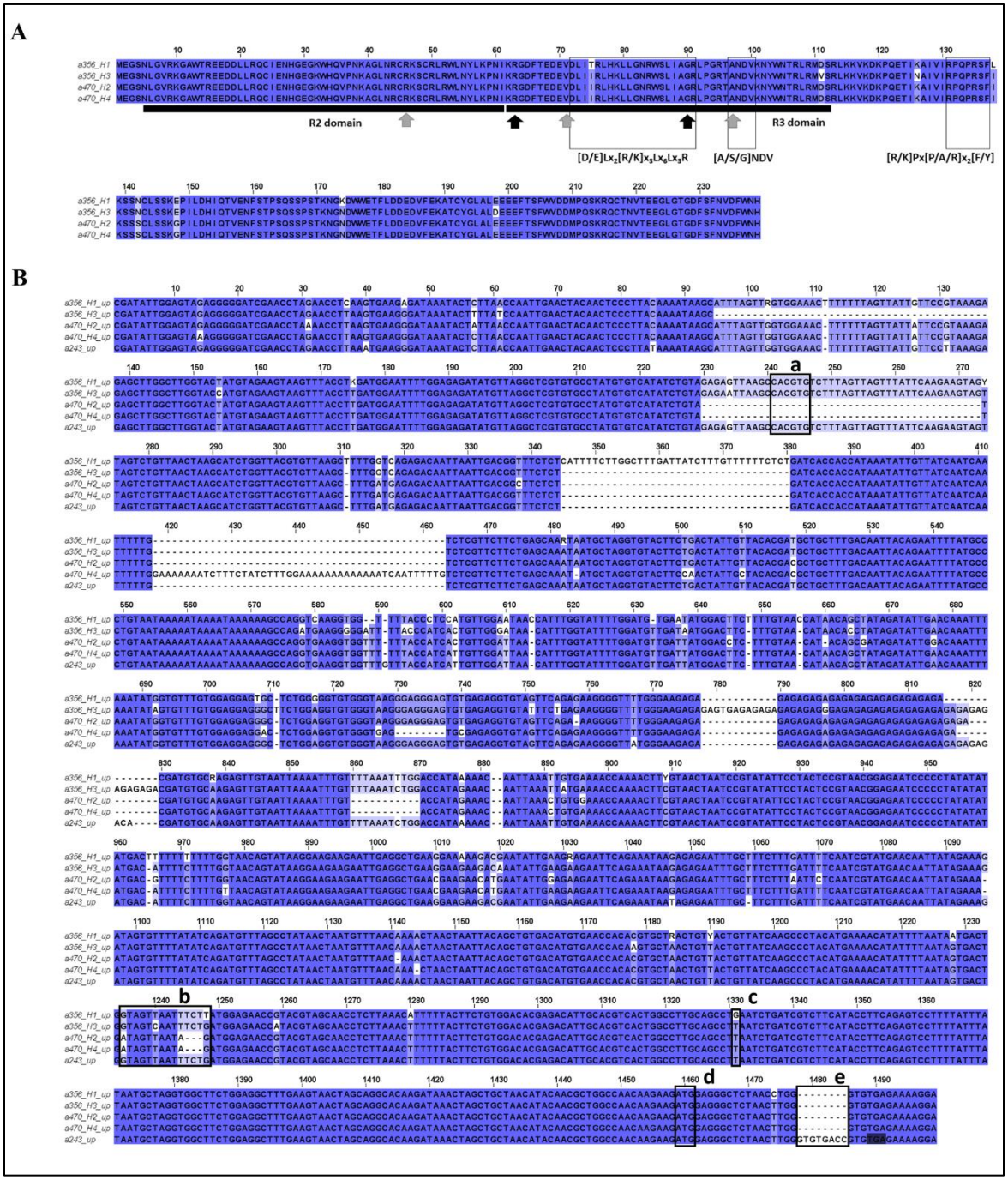


Figure 2.5. **(A)** Alignment of *PsMYB10.1 in silico* translated proteins from a whole-gene sequence. Conserved amino acids (gray arrows) and signature motifs ([A/S/G]NDV; [R/K]Px[P/A/R]x2[F/Y]) from R2R3-MYB anthocyanin-promoting proteins from Rosaceae were detected in all PsMYB10.1 sequences (Lin-Wang et al., 2010). Conserved motif [D/E]Lx2[R/K]x3Lx6Lx3R indicates a possible interaction with basic-helix-loop-helix (bHLH) proteins. Black arrows indicate key residues in PpMYB10.1 function (Zhou et al., 2018). **(B)** Polymorphisms found between a356-H1, a356-H3, a470-H2, a470-H4, and a243 upstream sequences: (a) G-box motif in the 44 bp indel; (b) polymorphisms in the R1-motif (Espley et al., 2009); (c) SNP S3_12879559 (Salazar et al., 2017); and (e) 8 bp insertion in exon 1 of a243, 4 bases before an in-frame STOP codon (shadowed). The start codon is marked as (d).

Allele a356 was present in haplotypes H1, H3, and H_i8 in the ABL collection (**Figure 2.3**); all accessions bearing skin-colored fruits but one (C74; H2/H6) had one or two copies of these haplotypes (**Supplementary Data 2.1**, **Supplementary Figure 2.6**), while those lacking anthocyanin pigments in the skin combined two copies of any of H2, H4, H5, H_i10, or H_i11. Consistently, most progenies containing either H1 or H3 had colored fruits (95.90 and 94.90% of cases, respectively), while those H2/H4 or H2/H2 fell mainly (95.12 and 95.24%, respectively) within the anthocyanin-less category, in particular with the mottled phenotype. In agreement with this, all ABL with the mottled phenotype had H2. In the progenies, combinations between H2 and H6 were either mottled (26.67%) or red (73.33%).

Apart from missing a356, only haplotypes H2, H4, and H5 amplified a470. The phase information, as well as phylogenetic analysis, are compatible with a356 and a470 being allelic of the same *MYB10.1* gene (from now on *PsMYB10.1a*). Therefore, *PsMYB10.1a* may have, at least, a356 and a470 as alleles, with a356 associated with anthocyanin color and a470 with its absence.

To validate the association, we analyzed a collection of 31 CV. Two new haplotypes were needed to explain the genotypes of 'Sunkiss' (H4/H_i12) and 'Gaia' (H5/H_i13), with H_i13 bearing a356 and conferring coloration (**Figure 2.3**). The marker was able to correctly predict the presence or absence of red color in the fruit skin in all cases: all colored cultivars carried at least one haplotype with the a356 allele while the rest had green or yellow fruits. In ABL and progenies, the H2 haplotype, in combination with other anthocyanin-less associated haplotypes, was observed in mottled fruits. This could not be validated in a more diverse collection since none of the CV showed this phenotype (**Supplementary Data 2.2**).

To investigate the C74 (H2/H6) outlier, we used the primer pair M101_f/r to identify possible null alleles in this ABL. The 1.5 kb sequence revealed a new *MYB10.1* allele identical to a470 except for a four-nucleotide deletion in the first intron (**Supplementary Data 2.4**). This allele (a467) was hidden by the monomorphic a466 in the electrophoresis profile. Using specific primers in all the germplasm, we found the a467 allele in 16 ABL and 6 CV, all with H6 or H2. While all genotypes with H6 had a467, only 13 of the accessions with H2 (27.08%) had this allele. Since C72 and C74 are H2/H6, it is uncertain whether a467 was in heterozygosis (carried in H6 only) or in homozygosis (carried in both H2 and H6). The segregation of a467 in P6 (with C11; H3/H6) confirmed that this allele co-segregates within H6, and therefore corresponds to an LG3 MYB10 gene (**Supplementary Figure 2.3**).

Expression of *PsMYB10.1a* in the Skin

The expression of the *PsMYB10.1a* alleles a356 and a470 in the fruit skin was evaluated in the black "BG" (H1/H2) and yellow "GJ" (H4/H$_i$9) varieties, at different stages of fruit development. A band with the expected size was obtained in the S2 and S3 stages of BG (**Figure 2.6**). Band sequencing revealed that this band corresponded exclusively to the a356 allele. No band was observed in GJ cDNA, indicating that a470 was not expressed in BG or in GJ. The alignment of a356-mRNA to different R2R3-MYB10s from databases detected a shared intron splicing mechanism (data not shown).

DISCUSSION

Major genes and QTLs for the traits related to anthocyanin coloration have been mapped along with the eight linkage groups of *Prunus* species (see summary in **Supplementary Table 2.6**). Several studies have demonstrated a major role of R2R3-MYB genes in the transcriptional control of the accumulation of anthocyanins in rosaceous crops. In particular, the expression of R2R3-MYB10 genes in fruit tissues has been correlated with the transcription of anthocyanin structural genes and pigment accumulation. The peach genome contains six R2R3-MYB10 genes, three of them in LG3, and three in LG6. Two of the LG6 genes (about 60 kb apart) co-localize with a Leaf color trait (Howad et al., 2005), while those in an LG3 (*PpMYB10.1*, *PpMYB10.2*, and *PpMYB10.3*) cluster in a genomic region of ca.70 kb co-localizing with an anther color gene (Ag) (Arús et al., 1994; Rahim et al., 2014). *PpMYB10.1* and *PpMYB10.3*, annotated as "Anthocyanin regulatory C1 protein" [AnnoMine gene description; PLAZA4.0 (Van Bel et al., 2018)], are expressed in fruit tissues at levels correlating with the accumulation of anthocyanins in the skin (Rahim et al., 2014). In Japanese plum, the sustained increase in the expression of *PsMYB10* during fruit ripening was in correlation with the accumulation of anthocyanins, suggesting the involvement of *PsMYB10* in the regulation of the transcriptional control during the biosynthesis of anthocyanins (González et al., 2016b). Considering that the modulation of MYB genes (including changes in either the promoter or their genic region) is enough to produce changes in the accumulation of anthocyanins (Ban et al., 2007) and that Salazar et al. (2017) identified the LG3 QTL for the skin color in Japanese plum, it is reasonable to consider that the genetic variation in *PsMYB10* is involved in the observed variation in fruit color. This motivated our search for polymorphisms in the Japanese plum LG3 R2R3-MYB genes to explain the observed variability in fruit color.

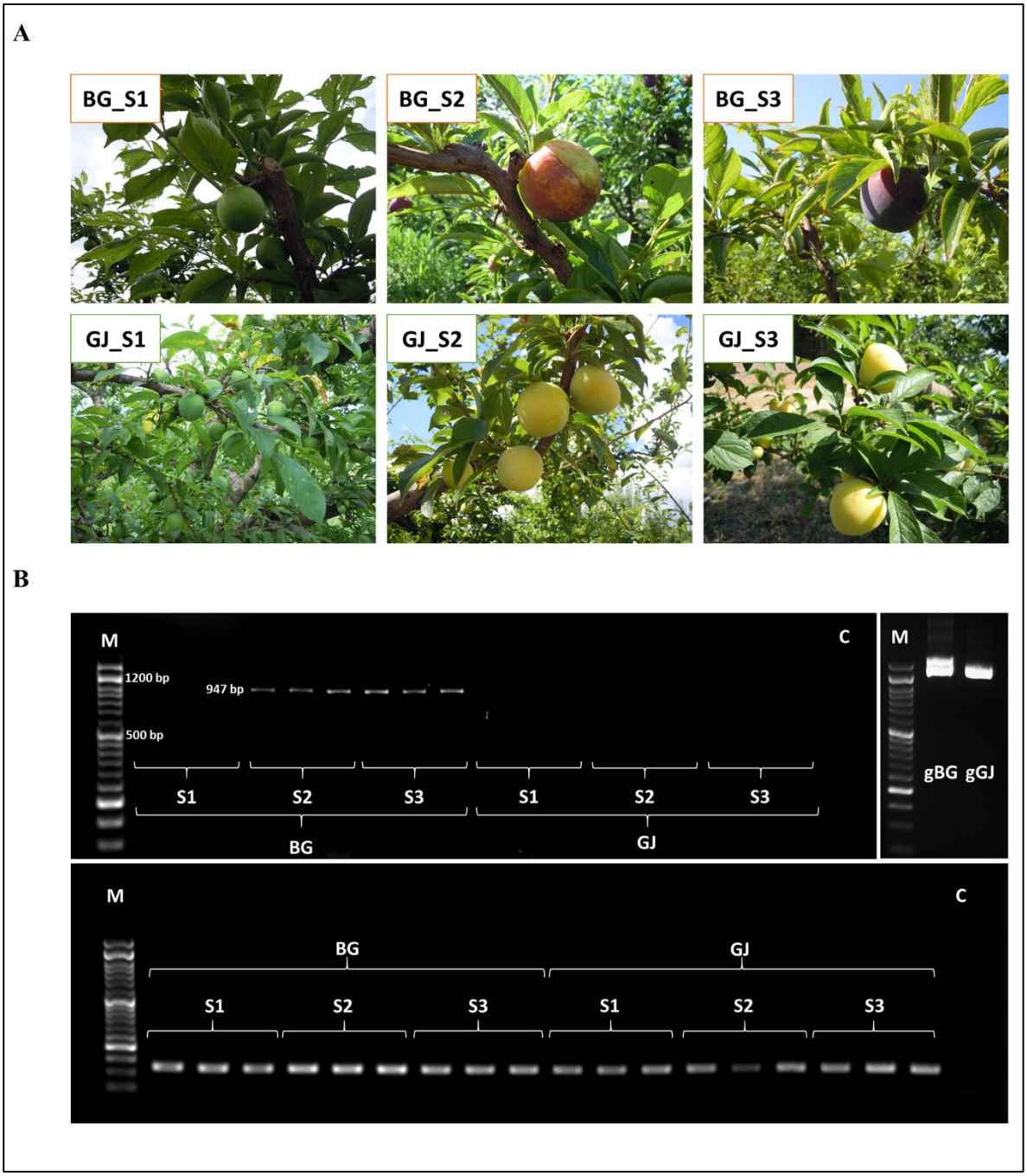


Figure 2.6. **(A)** RNA was extracted from the skin of 'Black Gold' (BG) and 'Golden Japan' (GJ) at three maturation stages (S1, S2, and S3). **(B)** Top row left: *PsMYB10.1* amplification from complementary DNA (cDNA) samples (BG_S2 and BG_S3). Band size 947 bp; top row right: *PsMYB10.1* amplification of genomic DNA (BG and gGJ). Band size 1.5 kb; bottom row: expression of the *MON* reference gene in all cDNA samples tested. Well M: DNA Ladder 50 bp ready-to-use (GeneON); well C: no template control.

While, only a few R2R3-MYB10 genes have been sequenced in Japanese plum, this gene family has been well characterized in other plant species, where systematic functional characterization identified highly conserved amino acid motifs and a characteristic R2R3-MYB DNA-binding domain (Kranz et al., 1998; Stracke et al., 2001; Zhang et al., 2020) with the homology values higher than 40% (Stracke et al., 2001). In this study, we took advantage of the high domain conservation and the high homology between a peach and Japanese plum (Mnejja et al., 2004) to obtain the partial sequence of 16 Japanese plum *MYB10* alleles. These alleles were amplified in a panel of Japanese plum accessions in numbers higher than the maximum expected for three loci in a diploid genome, confirmed through the progeny and haplotype segregation analysis, suggesting the gene duplication. The segregation of the bands in six biparental families allowed the definition of six haplotypes, each with three to five alleles. Combinations of these haplotypes were able to explain the genotype of most of the ABL and CV studied here, while the inference of additional haplotypes was needed to fully explain the genetic variability. Overall, this revealed high variability in this region, and high allelic variability (π = 0.204) when considering all alleles with a frequency >5% (12 in total), especially in the intronic regions.

Lin-Wang et al. (2010) isolated the R2R3-MYB10 homologs from cDNA of the major rosaceous crops. Here, we found some closely matched 12 alleles amplified at a frequency higher than 5%, grouping in three clusters in agreement with three tandemly duplicated PpMYB10 genes on LG3. Most of the alleles (7) clustered with the *PpMYB10.1* gene, contrasting with the phylogeny observed by Lin-Wang et al. (2010), where all but two of the isolated MYB *Prunus* alleles clustered with *PpMYB10.2*, as further confirmed for *PsMYB10* and *ParMYB10* genes by Rahim et al. (2014). Unlike Lin-Wang et al. (2010), by isolating the alleles from genomic DNA, we could identify a large number of *MYB10.1* alleles independent of their expression. On account of the high homology between the MYB10 genes, our strategy allowed us to assign some of the bands to gene alleles. In addition, the new grouping of previously isolated alleles will help to identify new candidate genes for fruit color as well as the further assignment of sequences to alleles or to novel genes in other *Prunus* species.

Haplotype reconstruction, homology, and phylogeny data suggested more than one *MYB10.1* gene in at least some *P. salicina* genotypes. Gene duplication is highly concordant with the number of alleles amplified in H1, H3, and H_i8 (with three alleles homologous to *MYB10.1*) and in H4, H_i9, H_i10, and H_i11 (with two alleles homologous to *MYB10.1*). Unlike *PsMYB10.1*, all haplotypes had a single copy of *PsMYB10.2* (with the monomorphic allele a466) and *PsMYB10.3* (either a443, a477, a492,

or a495). Similarly, sweet cherry (Shirasawa et al., 2017) and apricot (Jiang et al., 2019a) genomes have more than three MYB10 genes in the region. Although a fully assembled genome of *P. domestica* is not yet available, we detected a *MYB10.1* triplication in the scaffolds of its draft genome v1.0.a1 (https://www.rosaceae.org/species/prunus_domestica/genome_v1.0.a1). It is noticeable that not all haplotypes (observed or inferred here) carry the same number of *PsMYB10.1* alleles. This may be due to (i) a different number of LG3-*MYB10* copies in Japanese plum cultivars, (ii) a null allele due to mis-amplification or to same migration of different alleles of the same size in electrophoresis analysis, or (iii) a combination of (i) and (ii). The first hypothesis is highly possible considering the complex hybridization history of Japanese plum genotypes. Cultivated Japanese plums are complex interspecific hybrids between *P. salicina* and other species such as *Prunus simonii*, *P. cerasifera*, and American plums, with each cultivar having a different degree of these genomes in its genetic background, depending on the breeding history (Okie and Ramming, 1999; Boonprakob et al., 2001). In fact, individuals from a single species may have different genome sizes and genetic compositions, which fall within the concept of the pan-genome, i.e., all individuals within a species may have a core set of shared genes (pan-genome) plus a set of genes, similar to the insertion/deletion of genic copy number variants (CNVs) shared by only some individuals. Gene duplication, and in general gene CNV and the presence/absence of variants (PAV), is a frequently observed phenomenon in plants and is usually associated with both domestication and post-domestication diversification (Lye and Purugganan, 2019). In *Prunus*, the genes related to some quantitative traits are in hotspots (da Silva Linge et al., 2021), in certain cases located in clusters of duplicated genes (Bielenberg et al., 2004; Wells et al., 2015; Gu et al., 2016; da Silva Linge et al., 2021). Gene duplications can increase the gene product and result in altered patterns of gene expression (Innan and Kondrashov, 2010), which could explain the variability in the plum fruit color, including the mottled phenotype. A *de novo* sequence of *P. salicina* together with the sequence of a panel of cultivars will help to elucidate the impact of gene duplications in its phenotypic variation, especially in fruit color. Very recently, the *de novo* sequence of two *P. salicina* cultivars, 'Zhongli No.6' (https://www.rosaceae.org/Analysis/9019655) and 'Sanyueli' (Liu et al., 2020), has been released. These two sequences are still in their first version, with 318 Mb for 'Zhongli No.6' and 284 Mb for 'Sanyueli', which are ca. 53 and 19 Mb larger, respectively, than expected considering the size and homology with the peach genome (265 Mb). The alignments with both genomes of C20 and C46 Illumina sequences together with the BLAST analysis of the MYB alleles obtained here (data not shown) highlight the complexity of the region: the C20 and C46 sequences aligned in two MYB10

regions, 2 Mb apart, in the LG3 of the 'Zhongli No.6' genome, while gaps and misalignments were revealed when mapped against the 'Sanyueli' genome. The authors refer to these cultivars as Chinese plum and their interspecific hybridization history is not reported, so their kinship with the panel of varieties studied here (used in occidental breeding programs) is unknown.

The simplicity of the electrophoretic analysis that we used hides allelic complexity that can be revealed with the allele cloning and sequencing. We have proven this for the two alleles a356 and a466. In the first case, we observed polymorphisms in the upstream regulatory region of a356 in H1 and H3, although both alleles are equally associated with anthocyanin color. For a466, we isolated two alleles of almost the same size migrating together on electrophoresis, corresponding to *PsMYB10.1* (a467) and *PsMYB10.2* (a466). Likewise, as two duplicated genes may share alleles of equal size, resulting in a PCR band akin to a unique allele, we cannot discard the duplication of *PsMYB10.2* or *PsMYB10.3*. This emphasizes the utility of using haplotypes for the MAS since recombination in the region is unlikely.

The number of haplotypes is high when taking into account the expected reduced size of the region (73 kb in peach, 20 kb in almond, and 69 kb in sweet cherry genomes), which can be explained by the variability of the alleles and the complex hybridization history of Japanese plum. In *Prunus*, the size of this region varies independently of the number of MYB10 genes, the intergenic regions being shorter in sweet cherry and in almond (14.2 kb on average in a sweet cherry and 7.0 kb in an almond, compared to 32.8 kb in a peach). Such a high homology between the MYB genes, together with the unlikely recombination in the region, makes it difficult to assign the MYB alleles to their corresponding genes. Long-read DNA sequencing of the LG3-*MYB10* region in cultivars with different haplotypes should provide valuable data such as gene number and allelic assignment, together with gene order, the distance between them, and the size of the region.

Whole-genome sequence alignment of C20 and C46 against the peach reference genome found a low-sequence coverage in the intergenic *MYB10* sequences (<0.30), while it was considerably higher in the genic regions (>0.95). The alignment with the sweet cherry genome provided better coverage in the intergenic regions (>0.47). This highlights the importance of a high-quality reference sequence for highly complex regions, containing duplicated genes.

Primers were designed by sequence alignment to fully amplify the a356 and a470 alleles from different haplotypes. Their predicted amino acid sequence contained motifs conserved in other

anthocyanin-promoting R2R3-MYBs of the Rosaceae family as well as an intact bHLH domain (Lin-Wang et al., 2010). No polymorphisms were found in the protein sequences that could explain a possible loss of function, including the two key amino acids necessary for the PpMYB10.1 function identified by Zhou et al. (2018), which were unaltered in our sequences.

Our data suggest that a356 and a470 are allelic of *PsMYB10.1a*, with a356 being associated with the presence of anthocyanins in the fruit skin, and expressed in advanced stages of red fruits only, unlike a470. González et al. (2016a) sequenced the total RNA from the skin of 'Angeleno' (H1/H3; dark skin) and 'Lamoon' (unknown MYB10 genotype; yellow skin). Allele a356 was expressed in 'Angeleno' only, while no a470 reads were recovered (data not shown) confirming a possible loss of function of *MYB10.1a*-a470. This could be explained by the deletion of a G-box sequence in its upstream region. In Arabidopsis, the light-induced HY5 enhances the expression of the MYB ortholog *PAP1* (production of anthocyanin pigment 1) by binding to a G-box motif (CACGTG) in its promoter (Shin et al., 2013). This mechanism is conserved and has been described in an apple (An et al., 2017). Therefore, the deletion of this motif may prevent the binding of expression enhancers. Another cause of a470 loss of function could be the polymorphisms observed in the R1 motif, which has functional implications in *MYB10* expression. Espley et al. (2009) described multiple tandem copies of the motif being associated with red-fleshed phenotypes due to the ability of the same MYB10 protein to bind this sequence and enhance its own expression. The two polymorphisms found in this region, alone or in combination, may explain the lack of expression of a470 in the skin.

In this study, we have developed a molecular marker, which can predict the skin color in Japanese plum progenies and could be effectively used in breeding programs. The combination of the haplotypes obtained with the one-primer-pair molecular marker (which can be easily deduced from the allele segregation in the progeny) will identify, in a codominant manner, which seedlings will produce fruits with anthocyanin or non-anthocyanin coloration in the skin. Haplotype data have been shown to be highly informative for MAS (Aranzana et al., 2019), especially when the recombination within interesting regions is low as it occurs in the MYB10 region in *Prunus* LG3. The principal inconvenience of haplotype analysis is that at least two markers usually need to be run, increasing the cost of the analysis. Here, a unique primer can identify most of the haplotypes. Only H4/H5 and H4/H6 combinations amplify the same bands, and an additional primer pair to discern a467 from H6 is required. Given that the six haplotypes observed in this study can predict 91.36% of the phenotypes of the ABL and 83.87% of those in the panel of CV tested, we foresee a high

efficiency of this marker in other germplasms. For the remaining accessions (most containing low-frequency alleles), new haplotypes were inferred and need to be confirmed in families; however, color can still be predicted by the presence of a356 and a470 bands.

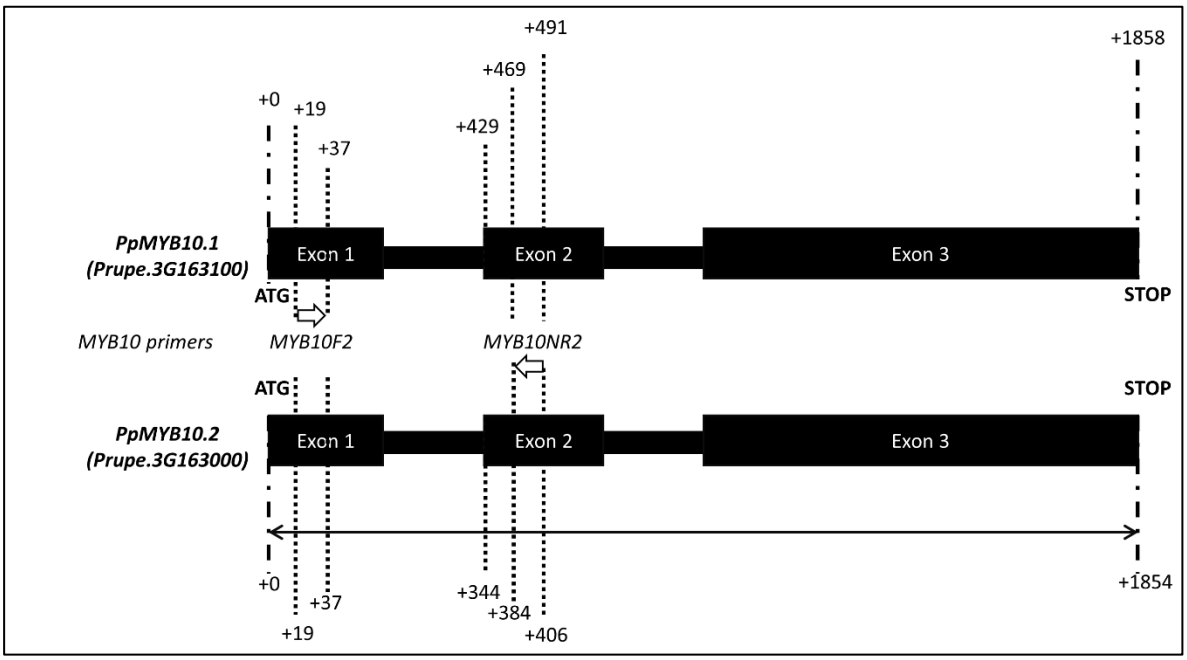
The color hue and tone have not been studied here. Salazar et al. (2017) found a SNP in 'Angeleno' (genotyped here as H1/H3) associated with the purple-skinned phenotypes in progenies, differentiating them from red phenotypes. We identified this SNP in the 5'UTR of a356; the nucleotide associated with higher intensity (G) was observed in a356-H1, while a356-H3 and other alleles not associated with skin color had the alternative nucleotide. In our panel of advanced lines, those with H1 are predominantly purple or black, which is consistent with the results of Salazar et al. (2017). There are many factors involved in coloration hues, such as the anthocyanin molecular structure, pH, copigmentation, temperature, and light (Khoo et al., 2017). Further studies considering anthocyanin levels as well as hue and tonalities are required to identify novel genes and markers for MAS.

CONCLUSIONS

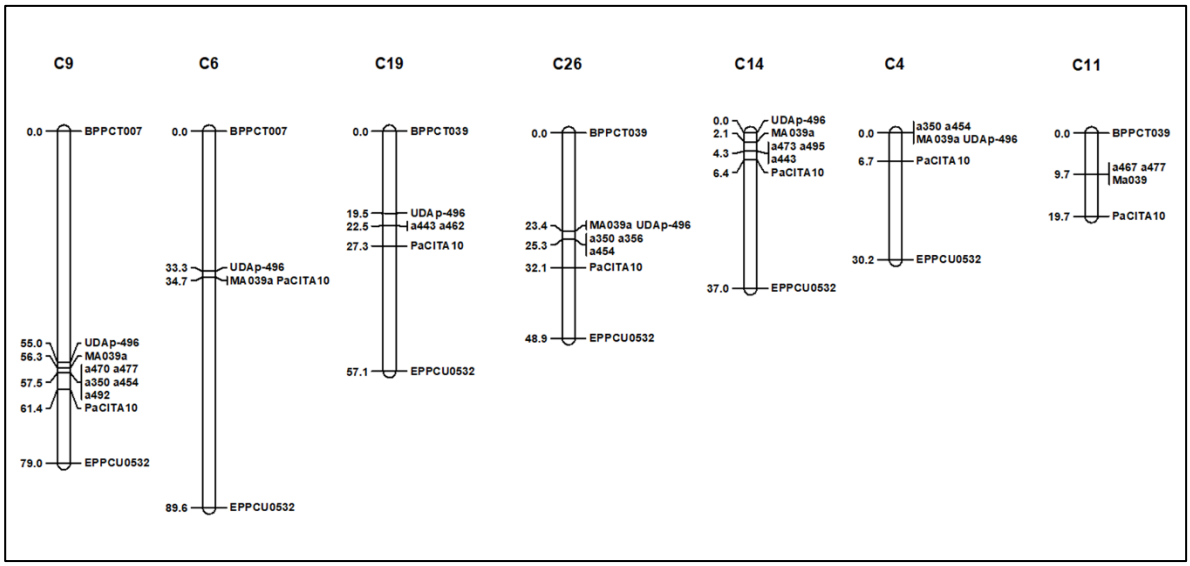
In this study, we have identified high levels of variability in the intronic and intergenic regions of the Japanese plum *MYB10* LG3 cluster, which contains at least three copies of the *MYB10.1* gene. In addition to the duplication of this gene, our data are compatible with higher levels of CNV. Methods including long-read sequences would be required to fully characterize the genetic variation in this *MYB10* cluster. Fruit color has an important impact on the choice of the consumer, therefore having good molecular markers for MAS is highly desired in breeding. Despite the complexity of the region, using allele cloning, phylogeny, progeny segregation, association test, and gene expression analysis, we succeed in identifying one allele (a356) associated with the skin color. Consequently, markers for this allele can be used with high efficiency for MAS. However, we could not find alleles associated with the flesh color. Here, we also found polymorphisms in the promoter of one of the alternative alleles (a470), which could explain the absence of this pigment in the skin. Further efforts characterizing this region may identify additional polymorphisms involved in fruit coloration.

SUPPLEMENTARY MATERIAL

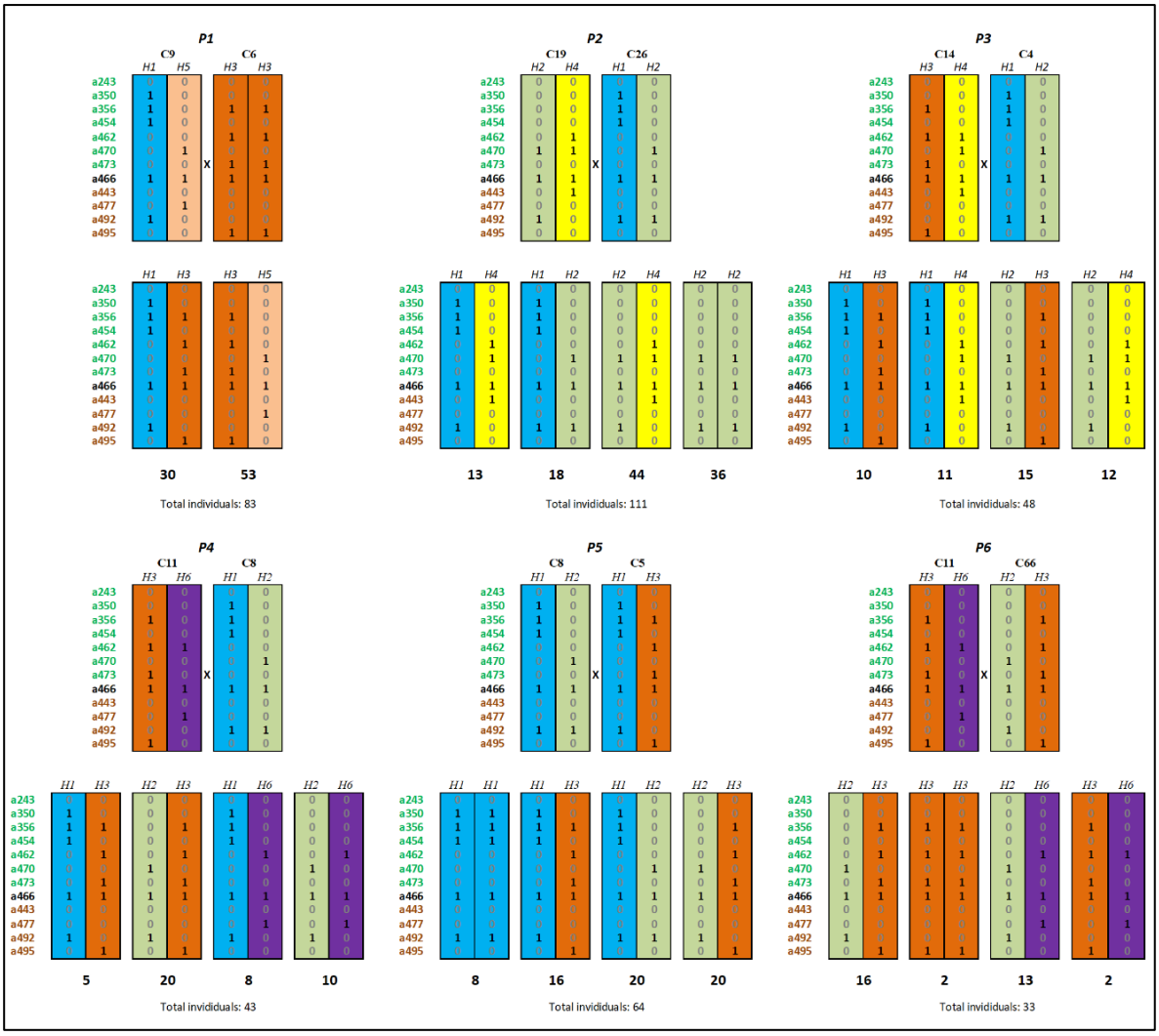
Supplementary Figure 2.1. **(A)** Fruits with mottled skin phenotype; **(B)** Detail of the skin speckles.



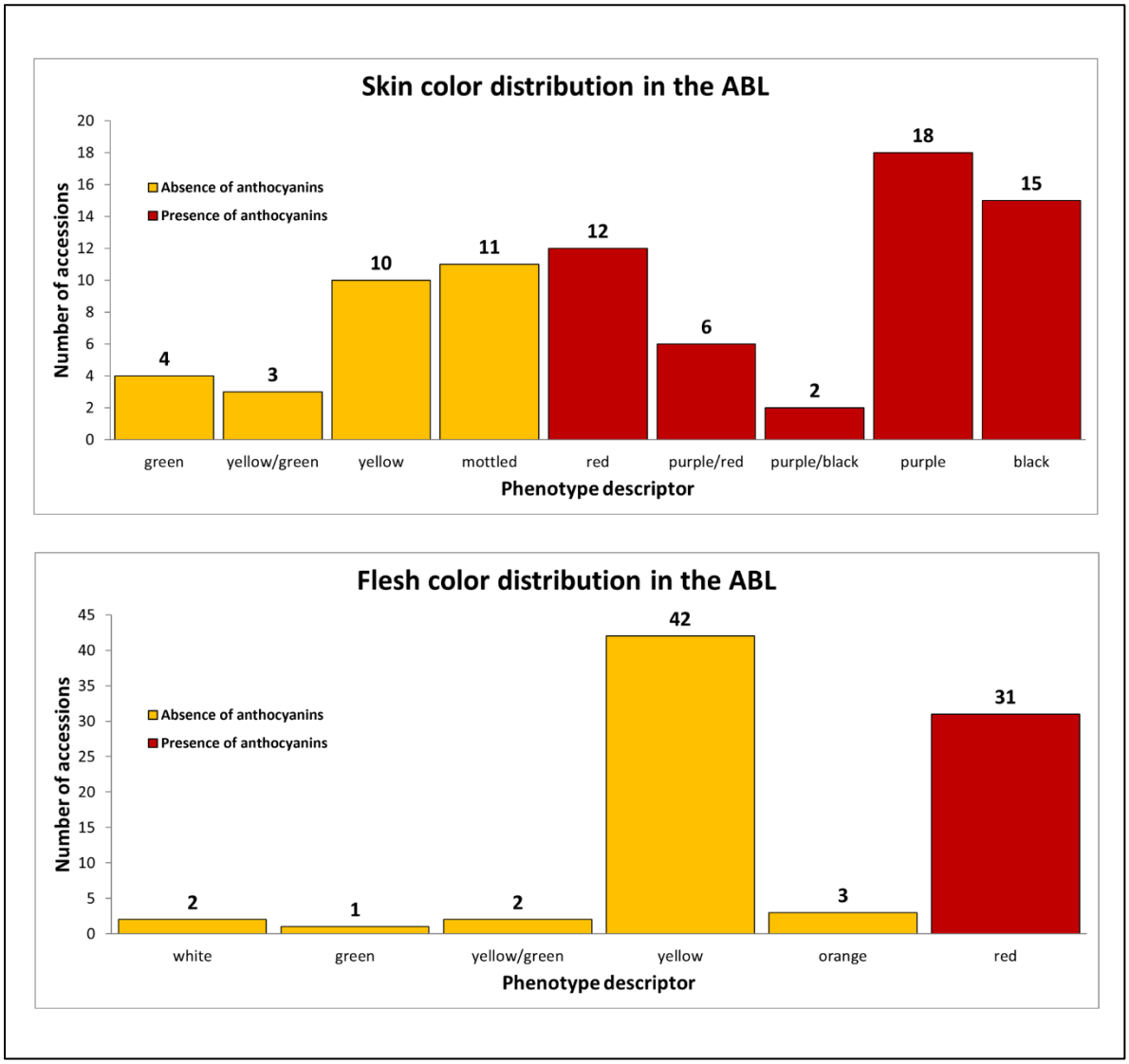
Supplementary Figure 2.2. Primer positions on the *PpMYB10.1* and *PpMYB10.2* genes.



Supplementary Figure 2.3. Genetic map of the Japanese plum *MYB10* alleles.

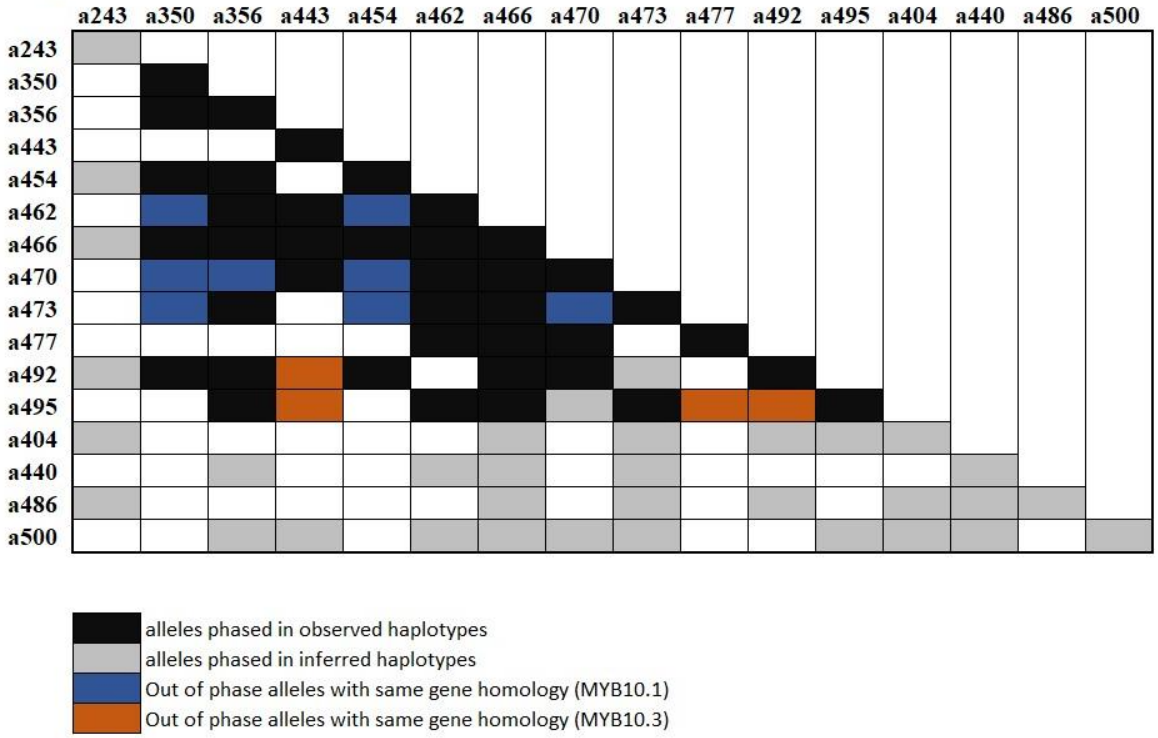


Supplementary Figure 2.4. Observed segregating Haplotypes in six F1 populations.



Supplementary Figure 2.5. Phenotype distribution in the ABL collection.

A)



B)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
H1	+												
H2	+	-											
H3	+	+	+										
H4	+	-	+	-									
H5	+	-	+	-	n.o								
H6	+	+	+	n.o	n.o	n.o							
H7	n.o	n.o	+	n.o	n.o	n.o	n.o						
H8	n.o	n.o	n.o	+	n.o	n.o	n.o	n.o					
H9	n.o	n.o	+	-	n.o	n.o	n.o	n.o	n.o				
H10	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o			
H11	n.o	n.o	+	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o		
H12	n.o	n.o	n.o	-	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o	
H13	n.o	n.o	n.o	n.o	+	n.o	n.o	n.o	n.o	n.o	n.o	n.o	n.o

+ : Anthocyanin presence
- : Anthocyanin absence
n.o: Combination not observed

Supplementary Figure 2.6. Allele (**A**) and Haplotype (**B**) combinations.

Supplementary Data 2.1 (1/2). Advance breeding lines collection (ABL): Parentals (C01 to C41).
The data for the P1 to P6 families can be found online: https://doi.org/10.3389/fpls.2021.655267

Germplasm	MYB10F2/MYB10NR2 Allele size (bp)												Phenotypic data		Haplotype combination		MYB10F2/MYB10NR4
	350	356	443	454	462	466	470	473	477	492	495	other	Flesh	Skin			a467
C01	1	1	0	1	0	1	1	0	1	1	0		yellow	red	H1	H5	0
C02	0	0	1	0	1	1	1	0	0	1	0		yellow	green	H2	H4	1
C03	1	1	0	1	0	1	0	0	0	1	0		yellow	black	H1	H1	0
C04	1	1	0	1	0	1	1	0	0	1	0		yellow	purple	H1	H2	0
C05	1	1	0	1	1	1	0	1	0	1	1		yellow	purple	H1	H3	0
C06	0	1	0	0	1	1	0	1	0	0	1		yellow	red	H3	H3	0
C07	1	1	0	1	0	1	1	0	0	1	0		red	purple	H1	H2	0
C08	1	1	0	1	0	1	1	0	0	1	0		red	black	H1	H2	0
C09	1	1	0	1	0	1	1	0	1	1	0		yellow	purple	H1	H5	0
C10	1	1	0	1	0	1	1	0	0	1	0		white	black	H1	H2	0
C11	0	1	0	0	1	1	0	1	1	0	1		yellow	red	H3	H6	1
C12	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C13	0	0	1	1	1	1	1	0	0	1	0	243	yellow	yellow	H4	*Hi9*	0
C14	0	1	1	0	1	1	1	1	0	0	1		yellow	purple	H3	H4	0
C15	0	0	1	0	1	1	1	0	1	0	0		yellow	yellow green	H4	H6	1
C16	0	0	0	0	0	1	1	0	1	1	0		red	mottled	H2	H5	0
C17	1	1	0	1	0	1	0	0	0	1	0		yellow	black	H1	H1	0
C18	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C19	0	0	1	0	1	1	1	0	0	1	0		red	green	H2	H4	0
C20	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C21	1	1	0	1	0	1	1	0	1	1	0		green	purple	H1	H5	0
C22	1	1	0	1	1	1	0	1	0	1	1		yellow	black	H1	H3	0
C23	1	1	0	1	0	1	1	0	0	1	0		red	black	H1	H2	0
C24	1	1	0	1	0	1	1	0	0	1	0		red	black	H1	H2	0
C25	1	1	0	1	0	1	1	0	0	1	0		red	purple red	H1	H2	0
C26	1	1	0	1	0	1	1	0	0	1	0		red	purple	H1	H2	0
C27	1	1	0	1	0	1	0	0	0	1	0		yellow	red	H1	H1	0
C28	0	1	0	0	1	1	0	1	1	0	1		yellow	black	H3	H6	1
C29	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C30	0	1	0	0	1	1	0	1	0	1	1	243, 404	yellow	black	H3	*Hi11*	0
C31	0	0	0	0	0	1	1	0	0	1	0		red	mottled	H2	H2	0
C32	0	0	0	0	0	1	1	0	1	1	0		red	mottled	H2	H5	0
C33	1	1	0	1	0	1	0	0	0	1	0		yellow	black	H1	H1	0
C34	1	1	0	1	1	1	0	1	0	1	1		yellow	red	H1	H3	0
C35	1	1	0	1	0	1	0	0	0	1	0		yellow	black	H1	H1	0
C36	1	1	0	1	0	1	1	0	0	1	0		red	purple	H1	H2	0
C37	0	0	1	0	1	1	1	0	0	1	0		white	yellow green	H2	H4	1
C38	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C40	1	1	1	1	1	1	1	0	0	1	0		yellow	purple	H1	H4	0
C41	0	0	1	0	1	1	1	0	0	1	0		red	mottled	H2	H4	0

Supplementary Data 2.1 (1/2). Advance breeding lines collection (ABL): Parentals (C42 to C86).
The data for the P1 to P6 families can be found online: https://doi.org/10.3389/fpls.2021.655267

Germplasm	MYB10F2/MYB10NR2 Allele size (bp)												Phenotypic data		Haplotype combination		MYB10F2/MYB10NR4
	350	356	443	454	462	466	470	473	477	492	495	other	Flesh	Skin			a467
C42	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C43	0	1	1	0	1	1	1	1	0	0	0	440, 500	red	purple black	H4	*Hi8*	0
C44	1	1	0	1	0	1	1	0	0	1	0		yellow	purple	H1	H2	n.d
C45	0	1	0	0	1	1	1	1	1	0	1		orange	purple	H3	H5	0
C46	1	1	0	1	0	1	1	0	0	1	0		red	purple	H1	H2	0
C47	1	1	0	1	1	1	0	0	1	1	0		yellow	red	H1	H6	1
C48	1	1	0	1	0	1	1	0	0	1	0		red	red	H1	H2	0
C49	1	1	0	1	0	1	1	0	0	1	0		yellow	black	H1	H2	1
C50	0	0	0	0	0	1	0	1	0	1	0	243, 404, 486	yellow	yellow	*Hi10*	*Hi11*	0
C51	0	1	0	1	1	1	0	1	0	1	1	243	yellow-red	purple	H3	*Hi9*	0
C52	0	1	1	0	1	1	1	1	0	0	1		yellow-green	purple	H3	H4	0
C53	0	1	0	0	1	1	1	1	0	1	1		orange	purple red	H2	H3	1
C54	0	0	0	0	0	1	1	0	0	1	0		red	mottled	H2	H2	1
C55	0	0	1	0	1	1	1	0	1	0	0		yellow	yellow green	H4	H6	1
C56	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C57	0	0	0	0	0	1	1	0	0	1	0		yellow	green	H2	H2	0
C58	0	1	1	0	1	1	1	1	0	0	1		yellow	red	H3	H4	0
C60	1	1	0	1	0	1	0	0	0	1	0		yellow	purple	H1	H1	0
C61	0	1	0	0	1	1	1	1	0	1	1		red	purple red	H2	H3	0
C62	0	1	0	0	1	1	1	1	0	1	1		red	red	H2	H3	0
C63	1	1	0	1	0	1	1	0	0	1	0		red	purple black	H1	H2	0
C64	1	1	0	1	0	1	1	0	0	1	0		red	purple red	H1	H2	0
C66	0	1	0	0	1	1	1	1	0	1	1		red	purple red	H2	H3	0
C67	0	0	1	0	1	1	1	0	0	1	0		red	mottled	H2	H4	0
C68	0	1	0	0	1	1	0	1	1	0	1		yellow	red	H3	H6	1
C69	1	1	0	1	0	1	1	0	0	1	0		red	purple red	H1	H2	0
C70	1	1	1	1	1	1	1	0	0	1	0		yellow-green	purple	H1	H4	0
C71	0	1	1	0	1	1	1	1	0	0	0	440, 500	yellow	purple	H4	*Hi8*	0
C72	0	0	0	0	1	1	1	0	1	1	0		red	mottled	H2	H6	1
C73	0	0	1	0	1	1	1	0	0	1	0		red	mottled	H2	H4	0
C74	0	0	0	0	1	1	1	0	1	1	0		red	red	H2	H6	1
C75	0	0	1	0	1	1	1	0	0	0	0		yellow	yellow	H4	H4	0
C76	0	0	1	0	1	1	1	0	0	1	0		yellow	green	H2	H4	1
C77	0	0	0	0	0	1	1	0	0	1	0		red	mottled	H2	H2	0
C78	0	1	0	0	1	1	0	1	1	0	1		yellow	red	H3	H6	1
C79	0	0	0	0	0	1	1	0	0	1	0		red	mottled	H2	H2	0
C80	1	1	0	1	0	1	1	0	0	1	0		red	black	H1	H2	0
C81	1	1	0	1	0	1	0	0	0	1	0		red	black	H1	H1	0
C82	0	0	0	0	0	1	1	0	0	1	0		red	mottled	H2	H2	0
C83	0	1	0	0	1	1	1	1	0	0	1	500	orange	purple	H3	*Hi7*	0
C86	1	1	0	1	0	1	1	0	1	1	0		purple	black	H1	H5	0

Supplementary Data 2.2. Commercial cultivars

Cultivar	MYB10F2/MYB10NR2 Allele size (bp)												Fruit skin color		Haplotype (H) combination		MYB10F2/MYB10NR4
	350	356	443	454	462	466	470	473	477	492	495	others	Expected	Described			a467
Angeleno	1	1	0	1	1	1	0	1	0	1	1		colored	black	H1	H3	0
Autumn Giant	0	1	0	0	1	1	0	1	0	0	1		colored	red	H3	H3	0
Black Amber	1	1	0	1	1	1	0	1	0	1	1		colored	black	H1	H3	0
Owen-T	1	1	0	1	0	1	0	0	0	1	0		colored	black/purple	H1	H1	0
Royal Diamond	1	1	0	1	1	1	0	1	0	1	1		colored	black	H1	H3	0
Fortune	0	1	0	0	1	1	0	1	1	0	1		colored	red	H3	H6	1
Black Diamond	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	0
Black Gold	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	0
Black Splendor	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	0
Black Star	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	1
Crimson Glo	1	1	0	1	0	1	1	0	0	1	0		colored	purple	H1	H2	0
Early Queen	1	1	0	1	0	1	1	0	0	1	0		colored	purple	H1	H2	1
Gaia	0	1	0	0	1	1	1	1	1	0	1	404	colored	red	H5	*Hi13*	0
John W	0	1	1	0	1	1	1	1	0	0	1		colored	purple	H3	H4	0
Laetitia	0	1	0	0	1	1	1	1	0	0	1	500	colored	red	H3	*Hi7*	0
Larry Ann	0	1	1	0	1	1	1	1	0	0	1		colored	purple	H3	H4	0
Pioneer	0	1	1	0	1	1	1	1	0	0	0	440, 500	colored	red	H4	*Hi8*	0
Primetime	1	1	0	1	0	1	1	0	0	1	0		colored	violet	H1	H2	0
Red Beauty	1	1	0	1	0	1	1	0	1	1	0		colored	red	H1	H5	0
Royal Garnet	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	1
Santa Rosa	0	1	0	0	1	1	1	1	0	1	1		colored	red	H2	H3	1
Sapphire	0	1	0	0	1	1	1	1	1	0	1		colored	red	H3	H5	0
Showtime	0	1	0	0	1	1	1	1	1	0	1		colored	red	H3	H5	0
Songría 15	1	1	0	1	0	1	1	0	0	1	0		colored	black	H1	H2	0
Golden Globe	0	0	1	0	1	1	1	0	0	0	0		not colored	yellow	H4	H4	0
Golden Japan	0	0	1	1	1	1	1	0	0	1	0	243	not colored	yellow	H4	*Hi9*	0
Kelsey	0	0	1	0	1	1	1	0	1	0	0		not colored	yellow	H4	H5	0
Sundew	0	0	1	0	1	1	1	0	0	0	0		not colored	yellow	H4	H4	0
SunGold	0	0	1	0	1	1	1	0	1	0	0		not colored	yellow	H4	H6	1
Sunkiss	0	0	1	0	1	1	1	0	0	0	0	500	not colored	yellow	H4	*Hi12*	0
TC Sun	0	0	1	0	1	1	1	0	1	0	0		not colored	yellow	H4	H5	0

Supplementary Data 2.3. Nucleotide diversity and percent identity matrix

	Nucleotide diversity, Pi values		
	Contig (bases 1 to 554)	EXONS	INTRON
MYB10 (12 cloned alleles)	0.204	0.071	0.244
MYB10.1 (7 alleles)	0.066	0.012	0.087
MYB10.3 (4 alleles)	0.007	0.000	0.010

Protein percent identity matrix created with Clustal																	
	PpMYB10.1	a356	a350	a454	a470	a473	EU153579.1	a243	a462	PpMYB10.2	EU153580.1	a466	PpMYB10.3	a443	a477	a492	a495
PpMYB10.1	100.00	98.15	98.15	98.15	98.15	98.15	98.15	96.00	88.46	94.44	88.89	90.74	81.48	85.19	85.19	85.19	85.19
a356	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
a350	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
a454	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
a470	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
a473	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
EU153579.1	98.15	100.00	100.00	100.00	100.00	100.00	100.00	98.00	90.38	94.44	90.74	92.59	83.33	87.04	87.04	87.04	87.04
a243	96.00	98.00	98.00	98.00	98.00	98.00	98.00	100.00	89.80	94.00	90.00	92.00	82.00	86.00	86.00	86.00	86.00
a462	88.46	90.38	90.38	90.38	90.38	90.38	90.38	89.80	100.00	86.54	86.54	88.46	78.85	80.77	80.77	80.77	80.77
PpMYB10.2	94.44	94.44	94.44	94.44	94.44	94.44	94.44	94.00	86.54	100.00	92.59	94.44	81.48	85.19	85.19	85.19	85.19
EU153580.1	88.89	90.74	90.74	90.74	90.74	90.74	90.74	90.00	86.54	92.59	100.00	98.15	81.48	85.19	85.19	85.19	85.19
a466	90.74	92.59	92.59	92.59	92.59	92.59	92.59	92.00	88.46	94.44	98.15	100.00	79.63	83.33	83.33	83.33	83.33
PpMYB10.3	81.48	83.33	83.33	83.33	83.33	83.33	83.33	82.00	78.85	81.48	81.48	79.63	100.00	92.59	92.59	92.59	92.59
a443	85.19	87.04	87.04	87.04	87.04	87.04	87.04	86.00	80.77	85.19	85.19	83.33	92.59	100.00	100.00	100.00	100.00
a477	85.19	87.04	87.04	87.04	87.04	87.04	87.04	86.00	80.77	85.19	85.19	83.33	92.59	100.00	100.00	100.00	100.00
a492	85.19	87.04	87.04	87.04	87.04	87.04	87.04	86.00	80.77	85.19	85.19	83.33	92.59	100.00	100.00	100.00	100.00
a495	85.19	87.04	87.04	87.04	87.04	87.04	87.04	86.00	80.77	85.19	85.19	83.33	92.59	100.00	100.00	100.00	100.00

EU153579.1 *PiMYB10 (P. domestica subsp. insititia)*
EU153580.1 *P. domestica*

Supplementary Data 2.4. *PsMYB10* sequences. These can be found online: https://doi.org/10.3389/fpls.2021.655267.

Supplementary Table 2.1. SSR markers used to construct LG3-*PsMYB10* genetic maps of P1, P2 and P3 families.

SSR marker	Families genotyped	Position in Peach Genome v2.0.a1	Reference
MA039a	P1, P2 and P3	Pp03: 17772071-17772260	Yamamoto *et al*., 2002
UDAp-496		Pp03: 16591013-16591165	Messina *et al*., 2004
PaCITA10		Pp03: 19499241-19499378	Lopes *et al*., 2002
EPPCU0532		Pp03: 26704358-26704530	Howad *et al*., 2005
BPPCT007	P1	Pp03: 3602939-3603083	Dirlewanger *et al*., 2002
BPPCT039	P2 and P3	Pp03: 6662508-6662632	Dirlewanger *et al*., 2002

Supplementary Table 2.2. Primer combinations used to amplify *PsMYB10* sequences.

Marker name	For/rev	Sequence	Ta (ºC)	Product range (bp)	Details
MYB10F2	For	GTGTGAGAAAAGGAGCTT	55	243-500	MYB10 marker
MYB10NR2	Rev	GATATTTGGCTTCAAATAGTTC			
M101_f	For	CTGTTATCAAGCCCTACATG	57		*MYB10.1*-a356 and *MYB10.1*-a470 full gene amplification
M101_r	Rev	GGTCTTTTGACAGCCG			
M102_f	For	CTGGCTGCAAGCATAC	57		*MYB10.2* full gene amplification
M102_r	Rev	GTGGGACAAACACTCTC			
M103_f	For	ATAGGAACTAGCAGGCAC	57		*MYB10.3* full gene amplification
M103_r	Rev	AGTTGCTAATAATTGCTACTAGG			
Pav3_9371	For	GAGTCGAGCTTGGGTC	54		Upstream amplification of *MYB10.1*-a356 (1R), a470 (2R) or a243 (5R)
MYB10_int1_1R	Rev	AGAGTTGGTACACAAGTGGGTATC	54		
MYB10_int1_2R	Rev	AGCACTAGTGACTATTGGGCC	54		
MYB10_int1_5R	Rev	CCTCGGCTCACGAGTAATATCACGAG	54		
M101_RT_F	For	TGGACACGAGACATTGCACG	60	950 bp	Genomic and cDNA *MYB10.1* amplification
M101_RT_R	Rev	CAATGGTCTTTTGACAGCCGC			
MYB10NR4	Rev	TTCCTGCACCTGTTCAAC	55		Used with MYB10F2 to discern a467

Supplementary Table 2.3. Sequence homology of cloned sequences with *Prunus* LG3-*PpMYB10* and NCBI sequences.

Fragment allele	Homologous gene from *Prunus persica* genome v2.0.a1	NCBI blastn		
		R2R3 MYB TF match	Identity (%)	Query coverage (%)
a243	*Prupe.3G163100* (*PpMYB10.1*)	*Prunus domestica subsp. insititia* R2R3 MYB transcription factor 10 gene, complete cds (EU153579.1)	97.84	93
a350			97.97	85
a356			96.82	84
a454			95.36	100
a462			94.51	100
a470			96.62	100
a473			99.16	100
a466	*Prupe.3G163000* (*PpMYB10.2*)	*Prunus domestica* R2R3 MYB transcription factor 10 gene, complete cds (EU153580.1)	98.93	100
a443	*Prupe.3G163300* (*PpMYB10.3*)	*Prunus avium* R2R3-MYB transcription factor (*MYBA1*) gene, complete cds (GU938680.1)	90.65	43
a477			90.65	22
a492			79.49	62
a495			79.57	62

Supplementary Table 2.4. Results of Illumina data aligned against three *Prunus* genomes.

		Genome	LG3	*MYB10* region	*MYB10* genic region	*MYB10* intergenic region
				P. persica		
depth	**C20**	22	23	24	40	17
	C46	74	73	63	135	36
coverage	**C20**	0.74	0.72	0.34	0.95	0.26
	C46	0.78	0.77	0.37	0.97	0.3
				P. dulcis		
depth	**C20**	22	22	28	30	27
	C46	78	78	85	101	66
coverage	**C20**	0.64	0.63	0.74	0.93	0.65
	C46	0.84	0.84	0.79	0.98	0.7
				P. avium		
depth	**C20**	23	24	13	11	14
	C46	73	77	30	21	33
coverage	**C20**	0.77	0.79	0.47	0.47	0.47
	C46	0.81	0.83	0.54	0.53	0.54

Supplementary Table 2.5. Results of $\chi^2_{(1df)}$ test in the panel of 81 accessions. The presence of allele a356 is statistically associated ($p=1.96\times10^{-18)}$ with the observation of anthocyanin-based color in the fruit skin.

Allele name	Allele frequency	Chi-square p-value	
		Skin color	Flesh color
a243	0.05	5.06×10^{-1}	1.06×10^{-1}
a350	0.42	2.64×10^{-8}	6.47×10^{-1}
a356	0.64	1.96×10^{-18}	3.65×10^{-1}
a443	0.31	2.21×10^{-6}	2.38×10^{-2}
a454	0.44	7.42×10^{-8}	9.19×10^{-1}
a462	0.54	7.55×10^{-2}	1.70×10^{-3}
a470	0.77	2.14×10^{-3}	7.16×10^{-4}
a473	0.27	5.22×10^{-4}	2.31×10^{-2}
a477	0.19	7.55×10^{-1}	2.23×10^{-1}
a492	0.73	2.08×10^{-1}	1.37×10^{-4}
a495	0.22	2.93×10^{-4}	2.12×10^{-2}

Supplementary Table 2.6. QTLs for fruit color in *Prunus*, online: https://doi.org/10.3389/fpls.2021.655267

Data Availability Statement

The data presented in the study are deposited in the EBI European Nucleotide Archive repository, accession number PRJEB43891.

Author Contributions

The experiments were conceived and designed by AF and MA. Experiments were conducted by AF. Data was analyzed by AF and MA. WH designed the MYB10F2/MYB10NR2 primers. Bioinformatics analysis was performed by AF, BG-G, FJ-R, and KA. Paper was written by AF and MA. All authors have critically revised the manuscript and approved the final document.

Funding

AF was a recipient of an FPI grant funded by the Spanish Ministry of Economy and Competitiveness (ref. BES-2016-079060F). FJ-R was a recipient of an FPI grant funded by the Spanish Ministry of Science and Innovation (ref PRE2019-087427). This research was supported by the Spanish Ministry of Science, Innovation and Universities (RTI2018-100795-B-I00). We acknowledge support from the CERCA Program (Generalitat de Catalunya), and the Severo Ochoa Program for Centres of Excellence in R&D 2016–2019 (SEV-2015-0533) and 2020–2023 (CEX2019-000902-S).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewers PM-G and JS declared past co-authorships with one of the authors BG-G and the absence of any ongoing collaboration with any of the authors to the handling editor.

Acknowledgments

We thank PLANASA company and PLANASA breeders, Mario Ortiz and Antonio García, for providing plant materials and phenotypic information.

REFERENCES

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014a). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell 26, 962–980. doi: 10.1105/tpc.113.122069

Albert, N. W., Davies, K. M., and Schwinn, K. E. (2014b). Gene regulation networks generate diverse pigmentation patterns in plants. Plant Signal. Behav. 9, 962–980. doi: 10.4161/psb.29526

Albert, N. W., Lewis, D. H., Zhang, H., Irving, L. J., Jameson, P. E., and Davies, K. M. (2009). Light-induced vegetative anthocyanin pigmentation in Petunia. J. Exp. Bot. 60, 2191–2202. doi: 10.1093/jxb/erp097

Alioto, T., Alexiou, K. G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., et al. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. Plant J. 101, 455–472. doi: 10.1111/tpj.14538

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

An, J. P., Liu, Y., Zhang, X., Bi, S., Wang, X., You, C., et al. (2020a). Dynamic regulation of different light intensity-modulated anthocyanin biosynthesis by BT2-TCP46-MYB1 in apple. J. Exp. Bot. 71, 3094–3109. doi: 10.1093/jxb/eraa056

An, J. P., Qu, F. J., Yao, J. F., Wang, X. N., You, C. X., Wang, X. F., et al. (2017). The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. Hortic. Res. 4, 1–9. doi: 10.1038/hortres.2017.23

An, J. P., Zhang, X. W., Bi, S. Q., You, C. X., Wang, X. F., and Hao, Y. J. (2020b). The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. Plant J. 101, 573–589. doi: 10.1111/tpj.14555

Aranzana, M. J., Decroocq, V., Dirlewanger, E., Eduardo, I., Gao, Z., Gasic, K., et al. (2019). *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects. Hortic. Res. 6:58. doi: 10.1038/s41438-019-0140-8

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, W597–W603. doi: 10.1093/nar/gks400

Arús, P., Messeguer, R., Viruel, M., Tobutt, K., Dirlewanger, E., Santi, F., et al. (1994). The European *Prunus* mapping project. Progress in the almond linkage map. Euphytica 77, 97–100. doi: 10.1007/978-94-011-0467-8_62

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. Plant Cell Physiol. 48, 958–970. doi: 10.1093/pcp/pcm066

Bielenberg, D. G., Wang, Y., Fan, S., Reighard, G. L., Scorza, R., and Abbott, A. G. (2004). A deletion affecting several gene candidates is present in the Evergrowing peach mutant. J. Hered. 95, 436–444. doi: 10.1093/jhered/esh057

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Boonprakob, U., Byrne, D. H., Graham, C. J., Okie, W., Beckman, T., and Smith, B. R. (2001). Genetic relationships among cultivated diploid plums and their progenitors as determined by RAPD markers. J. Am. Soc. Hortic. Sci. 126, 451–461. doi: 10.21273/JASHS.126.4.451

Bretó, M., Cantín, C., Iglesias, I., Arús, P., and Eduardo, I. (2017). Mapping a major gene for red skin color suppression (highlighter) in peach. Euphytica 213:14. doi: 10.1007/s10681-016-1812-1

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 17:66. doi: 10.1186/s13059-016-0924-1

Chagné, D., Carlisle, C. M., Blond, C., Volz, R. K., Whitworth, C. J., Oraguzie, N. C., et al. (2007). Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. BMC Genomics 8:212. doi: 10.1186/1471-2164-8-212

Chagné, D., Lin-Wang, K., Espley, R. V., Volz, R. K., How, N. M., Rouse, S., et al. (2013). An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiol. 161, 225–239. doi: 10.1104/pp.112.206771

da Silva Linge, C., Cai, L., Fu, W., Clark, J., Worthington, M., Rawandoozi, Z., et al. (2021). Multi-locus genome-wide association studies reveal fruit quality hotspots in peach genome. Front. Plant Sci. 12:644799. doi: 10.3389/fpls.2021.644799

Donoso, J. M., Picañol, R., Serra, O., Howad, W., Alegre, S., Arús, P., et al. (2016). Exploring almond genetic variability useful for peach improvement: mapping major genes and QTLs in two interspecific almond x peach populations. Mol. Breed. 36, 1–17. doi: 10.1007/s11032-016-0441-7

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

El-Sharkawy, I., Liang, D., and Xu, K. (2015). Transcriptome analysis of an apple (Malus x domestica) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. J. Exp. Bot. 66, 7359–7376. doi: 10.1093/jxb/erv433

Espley, R. V., Brendolise, C., Chagn,é, D., Kutty-Amma, S., Green, S., Volz, R., et al. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell 21, 168–183. doi: 10.1105/tpc.108.059329

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J. 49, 414–427. doi: 10.1111/j.1365-313X.2006.02964.x

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Feng, F., Li, M., Ma, F., and Cheng, L. (2013). Phenylpropanoid metabolites and expression of key genes involved in anthocyanin biosynthesis in the shaded peel of apple fruit in response to sun exposure. Plant Physiol. Biochem. 69, 54–61. doi: 10.1016/j.plaphy.2013.04.020

Feng, S., Wang, Y., Yang, S., Xu, Y., and Chen, X. (2010). Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. Planta 232, 245–255. doi: 10.1007/s00425-010-1170-5

Fischer, T. C., Gosch, C., Pfeiffer, J., Halbwirth, H., Halle, C., Stich, K., et al. (2007). Flavonoid genes of pear (Pyrus communis). Trees 21, 521–529. doi: 10.1007/s00468-007-0145-z

García-Gómez, B., Ruiz, D., Salazar, J., Rubio, M., Martínez-García, P., and Martínez-Gómez, P. (2020). Analysis of metabolites and gene expression changes relative to apricot (*Prunus armeniaca* L.) fruit quality during development and ripening. Front. Plant Sci. 11:1269. doi: 10.3389/fpls.2020.01269

García-Gómez, B., Salazar, J., Nicolás-Almansa, M., Razi, M., Rubio, M., Ruiz, D., et al. (2021). Molecular bases of fruit quality in *Prunus* species: an integrated genomic, transcriptomic, and metabolic review with a breeding perspective. Int. J. Mol. Sci. 22, 333. doi: 10.3390/ijms22010333

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant J. 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

González, M., Maldonado, J., Salazar, E., Silva, H., and Carrasco, B. (2016a). De novo transcriptome assembly of ‘Angeleno' and ‘Lamoon' Japanese plum cultivars (*Prunus salicina*). Genomics Data 9, 35–36. doi: 10.1016/j.gdata.2016.06.010

González, M., Salazar, E., Cabrera, S., Olea, P., and Carrasco, B. (2016b). Analysis of anthocyanin biosynthesis genes expression profiles in contrasting cultivars of Japanese plum (*Prunus salicina* L.) during fruit development. Gene Express. Patterns 21, 54–62. doi: 10.1016/j.gep.2016.06.005

González, M., Salazar, E., Castillo, J., Morales, P., Mura-Jornet, I., Maldonado, J., et al. (2016c). Genetic structure based on EST-SSR: a putative tool for fruit color selection in Japanese plum (*Prunus salicina* L.) breeding programs. Mol. Breed. 36:68. doi: 10.1007/s11032-016-0491-x

Gould, K. S. (2004). Nature's Swiss Army Knife: the diverse protective roles of anthocyanins in leaves. J. Biomed. Biotechnol. 2004, 314–320. doi: 10.1155/S1110724304406147

Gu, C., Wang, L., Wang, W., Zhou, H., Ma, B., Zheng, H., et al. (2016). Copy number variation of a gene cluster encoding endopolygalacturonase mediates flesh texture and stone adhesion in peach. J. Exp. Bot. 67, 1993–2005. doi: 10.1093/jxb/erw021

Guo, J., Cao, K., Deng, C., Li, Y., Zhu, G., Fang, W., et al. (2020). An integrated peach genome structural variation map uncovers genes associated with fruit traits. Genome Biol. 21, 1–19. doi: 10.1186/s13059-020-02169-y

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297–300. doi: 10.1093/nar/27.1.297

Honda, C., Kotoda, N., Wada, M., Kondo, S., Kobayashi, S., Soejima, J., et al. (2002). Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. Plant Physiol. Biochem. 40, 955–962. doi: 10.1016/S0981-9428(02)01454-7

Howad, W., Yamamoto, T., Dirlewanger, E., Testolin, R., Cosson, P., Cipriani, G., et al. (2005). Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics 171, 1305–1309. doi: 10.1534/genetics.105.043661

Illa, E., Sargent, D., Girona, E., Bushakra, J., Cestaro, A., Crowhurst, R., et al. (2011). Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. BMC Evol. Biol. 11:9. doi: 10.1186/1471-2148-11-9

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108. doi: 10.1038/nrg2689

Jiang, F., Zhang, J., Wang, S., Yang, L., Luo, Y., Gao, S., et al. (2019a). The apricot (*Prunus armeniaca* L.) genome elucidates Rosaceae evolution and beta-carotenoid synthesis. Hortic. Res. 6:128. doi: 10.1038/s41438-019-0215-6

Jiang, S. H., Sun, Q. G., Chen, M., Wang, N., Xu, H. F., Fang, H. C., et al. (2019b). Methylome and transcriptome analyses of apple fruit somatic mutations reveal the difference of red phenotype. BMC Genomics 20:117. doi: 10.1186/s12864-019-5499-2

Jin, W., Wang, H., Li, M., Wang, J., Yang, Y., Zhang, X., et al. (2016). The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). Plant Biotechnol. J. 14, 2120–2133. doi: 10.1111/pbi.12568

Jung, S., Lee, T., Cheng, C. H., Buble, K., Zheng, P., Yu, J., et al. (2019). 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Res. 47, D1137–D1145. doi: 10.1093/nar/gky1000

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food Nutr. Res. 61:1361779. doi: 10.1080/16546628.2017.1361779

Kim, H. Y., Saha, P., Farcuh, M., Li, B., Sadka, A., and Blumwald, E. (2015). RNA-seq analysis of spatiotemporal gene expression patterns during fruit development revealed reference genes for transcript normalization in plums. Plant Mol. Biol. Rep. 33, 1634–1649. doi: 10.1007/s11105-015-0860-3

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. 10, 236–242. doi: 10.1016/j.tplants.2005.03.002

Kong, J. M., Chia, L. S., Goh, N. K., Chia, T. F., and Brouillard, R. (2003). Analysis and biological activities of anthocyanins. Phytochemistry 64, 923–933. doi: 10.1016/S0031-9422(03)00438-2

Kosambi, D. D. (2016). "The estimation of map distances from recombination values," in DD Kosambi, ed R. Ramaswamy (New Delhi: Springer), 125–130. doi: 10.1007/978-81-322-3676-4_16

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., et al. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. Plant J. 16, 263–276. doi: 10.1046/j.1365-313x.1998.00278.x

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., and Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: an overview. Sci. World J. 2013:162750. doi: 10.1155/2013/162750

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. doi: 10.1093/molbev/msy096

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y., Xu, P., Chen, G., Wu, J., Liu, Z., and Lian, H. (2020). FvbHLH9 functions as a positive regulator of anthocyanin biosynthesis by forming a HY5-bHLH9 transcription complex in strawberry fruits. Plant Cell Physiol. 61, 826–837. doi: 10.1093/pcp/pcaa010

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T. K., et al. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol. 10:50. doi: 10.1186/1471-2229-10-50

Liu, C., Feng, C., Peng, W., Hao, J., Wang, J., Pan, J., et al. (2020). Chromosome-level draft genome of a diploid plum (Prunus salicina). Gigascience 9, giaa130. doi:10.1093/gigascience/giaa130.

Lü, P., Yu, S., Zhu, N., Chen, Y.-R., Zhou, B., Pan, Y., et al. (2018). Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. Nat. Plants 4, 784–791. doi: 10.1038/s41477-018-0249-z

Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. Trends Plant Sci. 24, 352–365. doi: 10.1016/j.tplants.2019.01.003

Mnejja, M., Garcia-Mas, J., Howad, W., Badenes, M. L., and Arús, P. (2004). Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and transferable to peach and almond. Mol. Ecol. Notes 4, 163–166. doi: 10.1111/j.1471-8286.2004.00603.x

Ogah, O., Watkins, C. S., Ubi, B. E., and Oraguzie, N. C. (2014). Phenolic compounds in Rosaceae fruit and nut crops. J. Agric. Food Chem. 62, 9369–9386. doi: 10.1021/jf501574q

Okie, W., and Ramming, D. (1999). Plum breeding worldwide. Horttechnology 9, 162–176. doi: 10.21273/HORTTECH.9.2.162

Panche, A., Diwan, A., and Chandra, S. (2016). Flavonoids: an overview. J. Nutr. Sci. 5:e47. doi: 10.1017/jns.2016.41

Paz-Ares, J., Ghosal, D., and Saedler, H. (1990). Molecular analysis of the C1-I allele from Zea mays: a dominant mutant of the regulatory C1 locus. EMBO J. 9, 315–321. doi: 10.1002/j.1460-2075.1990.tb08113.x

Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. Planta 240, 913–929. doi: 10.1007/s00425-014-2078-2

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302. doi: 10.1093/molbev/msx248

Salazar, J. A., Pacheco, I., Shinya, P., Zapata, P., Silva, C., Aradhya, M., et al. (2017). Genotyping by sequencing for SNP-based linkage analysis and identification of QTLs linked to fruit quality traits in Japanese plum (*Prunus salicina* Lindl.). Front. Plant Sci. 8:476. doi: 10.3389/fpls.2017.00476

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644. doi: 10.1086/502802

Shin, D. H., Choi, M., Kim, K., Bang, G., Cho, M., Choi, S. B., et al. (2013). HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in Arabidopsis. FEBS Lett. 587, 1543–1547. doi: 10.1016/j.febslet.2013.03.037

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., et al. (2017). The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res. 24, 499–508. doi: 10.1093/dnares/dsx020

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539. doi: 10.1038/msb.2011.75

Sneath, P.H. A., and Sokal, R. R., (1973). Numerical Taxonomy: The Principles and Practice of Numerical Classification. San Francisco, CA: WF Freeman and Co.

Socquet-Juglard, D., Christen, D., Devènes, G., Gessler, C., Duffy, B., and Patocchi, A. (2013). Mapping architectural, phenological, and fruit quality QTLs in apricot. Plant Mol. Biol. Rep. 31, 387–397. doi: 10.1007/s11105-012-0511-x

Sooriyapathirana, S. S., Khan, A., Sebolt, A. M., Wang, D., Bushakra, J. M., Lin-Wang, K., et al. (2010). QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). Tree Genet. Genomes 6, 821–832. doi: 10.1007/s11295-010-0294-x

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: join map. Plant J. 3, 739–744. doi: 10.1111/j.1365-313X.1993.00739.x

Stephens, M., and Donnelly, P. (2003). A Comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162–1169. doi: 10.1086/379378

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Curr. Opin. Plant Biol. 4, 447–456. doi: 10.1016/S1369-5266(00)00199-0

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol. 142, 1216–1232. doi: 10.1104/pp.106.088104

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Telias, A., Lin-Wang, K., Stevenson, D. E., Cooney, J. M., Hellens, R. P., Allan, A. C., et al. (2011). Apple skin patterning is associated with differential expression of MYB10. BMC Plant Biol. 11:93. doi: 10.1186/1471-2229-11-93

Tuan, P. A., Bai, S., Yaegaki, H., Tamura, T., Hihara, S., Moriguchi, T., et al. (2015). The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biol. 15:280. doi: 10.1186/s12870-015-0664-5

Ubi, B. E., Honda, C., Bessho, H., Kondo, S., Wada, M., Kobayashi, S., et al. (2006). Expression analysis of anthocyanin biosynthetic genes in apple skin: effect of UV-B and temperature. Plant Sci. 170, 571–578. doi: 10.1016/j.plantsci.2005.10.009

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., et al. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. 46, D1190–D1196. doi: 10.1093/nar/gkx1002

Velten, J., Cakir, C., and Cazzonelli, C. I. (2010). A spontaneous dominant-negative mutation within a 35S:: AtMYB90 transgene inhibits flower pigment production in tobacco. PLoS ONE 5:e9917. doi: 10.1371/journal.pone.0009917

Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarani, G., Vendramin, E., et al. (2017). The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. BMC Genomics 18:225. doi: 10.1186/s12864-017-3606-9

Vimolmangkang, S., Zheng, D., Han, Y., Khan, M. A., Soria-Guerra, R. E., and Korban, S. S. (2014). Transcriptome analysis of the exocarp of apple fruit identifies light-induced genes involved in red color pigmentation. Gene 534, 78–87. doi: 10.1016/j.gene.2013.10.007

Wang, H., Zhang, H., Yang, Y., Li, M., Zhang, Y., Liu, J., et al. (2020). The control of red colour by a family of MYB transcription factors in octoploid strawberry (Fragaria× ananassa) fruits. Plant Biotechnol. J. 18, 1169–1184. doi: 10.1111/pbi.13282

Wang, Y. C., Wang, N., Xu, H. F., Jiang, S. H., Fang, H. C., Su, M. Y., et al. (2018). Auxin regulates anthocyanin biosynthesis through the Aux/IAA-ARF signaling pathway in apple. Hortic. Res. 5, 1–11. doi: 10.1038/s41438-018-0068-4

Wang, Z., Meng, D., Wang, A., Li, T., Jiang, S., Cong, P., et al. (2013). The methylation of the PcMYB10 promoter is associated with green-skinned sport in Max Red Bartlett pear. Plant Physiol. 162, 885–896. doi: 10.1104/pp.113.214700

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wells, C. E., Vendramin, E., Jimenez Tarodo, S., Verde, I., and Bielenberg, D. G. (2015). A genome-wide analysis of MADS-box genes in peach [*Prunus persica* (L.) Batsch]. BMC Plant Biol. 15:41. doi: 10.1186/s12870-015-0436-2

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. Curr. Opin. Plant Biol. 5, 218–223. doi: 10.1016/S1369-5266(02)00256-X

Xu, Y., Feng, S., Jiao, Q., Liu, C., Zhang, W., Chen, W., et al. (2012). Comparison of MdMYB1 sequences and expression of anthocyanin biosynthetic and regulatory genes between Malus domestica Borkh. cultivar 'Ralls' and its blushed sport. Euphytica 185, 157–170. doi: 10.1007/s10681-011-0494-y

Yao, G., Ming, M., Allan, A. C., Gu, C., Li, L., Wu, X., et al. (2017). Map-based cloning of the pear gene MYB 114 identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. Plant J. 92, 437–451. doi: 10.1111/tpj.13666

Zhai, R., Wang, Z., Zhang, S., Meng, G., Song, L., Wang, Z., et al. (2016). Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (Pyrus bretschneideri Rehd.). J. Exp. Bot. 67, 1275–1284. doi: 10.1093/jxb/erv524

Zhang, J., Xu, H., Wang, N., Jiang, S., Fang, H., Zhang, Z., et al. (2018). The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. Plant Mol. Biol. 98, 205–218. doi: 10.1007/s11103-018-0770-5

Zhang, Z., Tian, C., Zhang, Y., Li, C., Li, X., Yu, Q., et al. (2020). Transcriptomic and metabolomic analysis provides insights into anthocyanin and procyanidin accumulation in pear. BMC Plant Biol. 20, 1–14. doi: 10.1186/s12870-020-02344-0

Zhou, H., Liao, L., Xu, S., Ren, F., Zhao, J., Ogutu, C., et al. (2018). Two amino acid changes in the R3 repeat cause functional divergence of two clustered MYB10 genes in peach. Plant Mol. Biol. 98, 169–183. doi: 10.1007/s11103-018-0773-2

Zhou, H., Lin-Wang, K., Wang, F., Espley, R. V., Ren, F., Zhao, J., et al. (2019). Activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation. N. Phytol. 221, 1919–1934. doi: 10.1111/nph.15486

Chapter 3

An efficient CRISPR-Cas9 enrichment sequencing strategy for characterizing complex and highly duplicated genomic regions. A case study in the *Prunus salicina* LG3-MYB10 genes cluster

Arnau Fiol[1], Federico Jurado-Ruiz[1], Elena López-Girona[2], Maria José Aranzana[1,3]*

[1]Centre for Research in Agricultural Genomics, CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain

[2]The New Zealand Institute for Plant and Food Research Limited (Plant & Food Research), Private Bag 11600, Palmerston North 4442, New Zealand

[3]Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain

This chapter has been submitted to Plant Methods and is currently under review

Preprint available at bioRxiv repository (January 2022), DOI: 10.1101/2022.01.24.477518

ABSTRACT

Background

Genome complexity is largely linked to diversification and crop innovation. Examples of regions with duplicated genes with relevant roles in agricultural traits are found in many crops. In both duplicated and non-duplicated genes, much of the variability in agronomic traits is caused by large as well as small and middle scale structural variants (SVs), which highlights the relevance of the identification and characterization of complex variability between genomes for plant breeding.

Results

Here we improve and demonstrate the use of CRISPR-Cas9 enrichment combined with long-read sequencing technology to resolve the *MYB10* region in the linkage group 3 (LG3) of Japanese plum (*Prunus salicina*), which has a length from 90 kb to 271 kb according to the *P. salicina* genomes available. We demonstrate the high complexity of this region, with homology levels between Japanese plum varieties comparable to those between *Prunus* species. We cleaved *MYB10* genes in five plum varieties using the Cas9 enzyme guided by a pool of crRNAs. The barcoded fragments were then pooled and sequenced in a single MinION Oxford Nanopore Technologies (ONT) run, yielding 194 Mb of sequence. The enrichment was confirmed by aligning the long reads to the plum reference genomes, with a mean read on-target value of 4.5% and a depth per sample of 11.9x. From the alignment, 3,261 SNPs and 287 SVs were called and phased. A *de novo* assembly was constructed for each variety, which also allowed detection, at the haplotype level, of the variability in this region.

Conclusions

CRISPR-Cas9 enrichment is a versatile and powerful tool for long-read targeted sequencing even on highly duplicated and/or polymorphic genomic regions, being especially useful when a reference genome is not available. Potential uses of this methodology as well as its limitations are further discussed.

Keywords: Cas9 enrichment, complex regions, long-read, targeted sequencing, variability, gene duplication, *MYB10*

BACKGROUND

The origins of agriculture and the development of new crops are tightly linked to whole genome duplications (WGD), and/or to chromosome rearrangements, which are among the major determinants of useful trait diversity [1-3] as well as of genome complexity [4]. In addition to WGD, small-scale duplications (SSD), usually produced by segmental duplications and gene expansion, also play an important role in crop innovation [5, 6]. Reduced selective pressure acting on duplicated genes may, in some cases, result in them acquiring novel functionalities that contribute to diversification and lead to important traits for agriculture [7-10].

Gene diversification results in high levels of diversity within gene families. For example, a substantial number of copies of celiac disease related α-gliadin genes are estimated in wheat (ranging from 25–35 to 100–150 copies per haploid genome), and are probably generated by duplication, deletion events and retrotransposon insertion [1, 12]. These genes show high variability between genomes [3], with at least half of the copies being inactive or pseudogenes [4].

In both duplicated and non-duplicated genes, a large fraction of the variability is caused by large as well as small and middle scale structural variants (SVs). For example, Chawla et al. [5] found that up to 10% of all genes in *Brassica napus* were affected by small- to mid-scale SV events. In tomato, Alonge et al. [6] found multiple SVs, in many cases mediated by transposable elements (TEs); many of the SVs responsible for changes in gene dosage and expression levels modified agronomic traits such as fruit flavor, size, and production. Recent studies of SVs between and within plant species have revealed extensive genome content variation [7], leading to the pan-genome concept, which is centered on the study of the entire gene repertoire of a species by sequencing multiple individuals [8].

The major effect of duplicated genes and SVs, especially in crop agronomic traits, highlights the importance of studying the variability between genomes in these genes for plant breeding.

Whole-genome sequencing (WGS) methods have contributed enormously to broadening our understanding of the genetic basis of useful traits. However, genome complexity still represents a challenge for genome assembly and annotation. In general, large divergent duplications with less than 97% homology can be resolved by WGS assembly, whereas reads of duplications with higher homology are frequently collapsed, producing assembly errors [9].

A large number of genomes have been obtained, and SNPs and small InDels discovered using short-read high throughput sequencing (SR-HTS), but long-read high throughput sequencing (LR-HTS) is required to resolve highly homologous duplicated regions and SVs, including repetitive regions and different forms of copy number variations, such as presence-absence variants [20]. Multiple reference genomes for crop species are currently being generated using LR-HTS methods [6, 21, 22]; however, although the cost of long-range sequencing methods is decreasing, the availability of large amounts of high quality DNA, whole genome data storage, alignment and computation costs may be drawbacks when the objective is to resolve and explore the variability in a large panel of varieties of a unique complex loci for an interesting agricultural trait.

Over the last few years, several skimming and targeted-enrichment sequencing methods have been released and successfully used to pinpoint small variants in single regions, especially in phylogenomics. However, a cost-effective method to scan the variability (including SV) in a complex or highly diverse region in a panel of genotypes (for example varieties or seedlings) using LR-HTS strategies is still lacking.

Recently, a methodology has been reported for long-read targeted sequencing using CRISPR-Cas9 technology to direct the sequencing adapters to the regions of interest [23, 24]. In this procedure, selected regions of high molecular weight dephosphorylated DNA are cut by the Cas9 enzyme directed by specific guide RNAs (gRNAs). The digested fragments are enriched by long-read sequencing thanks to the preferential ligation of sequencing adapters towards the 5’ phosphate groups generated at the cleaved ends. Gilpatrick et al. [23] used a pool of gRNAs to cut and sequence several loci associated with human breast cancer on the same run, resolving SNPs and SVs at the haplotype level. This methodology is simple to perform, as it does not require physical separation of the DNA for sequencing, cloning or amplification steps, thus allowing native DNA strands to be read, and it is not affected by amplification bias and allows visualization of allele-specific methylation patterns [23, 25]. CRISPR-Cas9 enrichment has recently been used on a plant genome (as a proof of concept) showing that, at the haplotype level, it can resolve the SV that causes red flesh color in apples [24]. Although this methodology was used to sequence a short (7.8 Kb) and low complex region in one apple variety, the successful results augur well for its use for fine-mapping in other situations.

The objective of this study was to evaluate the potential of CRISPR-Cas9 enrichment to identify the variability (in particular SV) in highly complex regions using a pool of DNA of highly diverse varieties. The region selected was the Japanese plum (*Prunus salicina*) *MYB10* region located on linkage group 3 (LG3), found recently to contain at least three *MYB10.1* (*MYB10.1a*, *MYB10.1b* and *MYB10.1c*) genes, one *MYB10.2* gene and one *MYB10.3* gene, while additional copies could not be discarded [26]. A marker system designed in this region was able to characterize the LG3-*PsMYB10* allelic organization in Japanese plum varieties and progenies into six haplotypes (H1 to H6), each with a different combination of the *MYB10.1–3* alleles [26]. For example, H1 and H3 haplotypes contained three *PsMYB10.1* alleles, one of them (*MYB10.1a*-a356) associated with the anthocyanin skin color; polymorphisms in the promoter and intron of this allele in H1 and H3 were observed, confirming the high variability of this region.

We designed seven gRNAs in conserved sites of the *PsMYB10* genes and pooled them, together with two guides flanking the region, to allow for multiple Cas9-mediated cuts in the DNA of five Japanese plum commercial varieties, all with a distinct and heterozygous genotype for the *MYB10* region, while sharing one haplotype by pairs. The fragments were labeled with a barcode system and pooled and sequenced in a single MinION Oxford Nanopore Technology (ONT) run. Reads alignment with reference genomes or *de novo* alignments successfully identified SNPs and SVs, most of them newly discovered and others observed in our previous work, so validating the methodology for sequencing and variant detection of regions containing complex duplications.

RESULTS

The Japanese plum LG3-*MYB10* region is highly complex

We used the Japanese plum LG3-MYB10 region as a model of high complexity. Japanese plum is a self-incompatible fruit tree with a complex history of interspecific crosses between diploid plums: the LG3-MYB10 region in *Prunus* genomes, particularly in Japanese plum, is highly diverse and contains a cluster of at least three paralogous *MYB10* gene copies.

To test the complexity of this region and validate its suitability for analysis, we identified and compared the assembly of the LG3-MYB10 region in the two Japanese plum genomes currently available (one for 'Sanyueli' [27] and one for 'Zhongli No. 6' [28] varieties). In the 'Sanyueli' genome (v1.0) the region spanned 135 kb in chromosome 4 (Chr4:12192580-12327184). Although the

genome annotation contained only one gene, homologous to *PsMYB10.2*, BLAST identified an additional *PsMYB10.2,* two *PsMYB10.1* and one *PsMYB10.3* gene copies in the region.

In the 'Zhongli No. 6' genome we identified two MYB10 regions 2 Mb apart in chromosome 3: one 271 kb long (from now on, Zhongli-1; LG03:30838572-31109359) and the other 90 kb long (Zhongli-2; LG03:28592935-28682529). The Zhongli-1 region had five *MYB10* genes annotated, corresponding to two copies of *MYB10.1*, two copies of *MYB10.2* and one copy of a *MYB10.3* gene. These genes were mis-annotated: the second exon was omitted in all but one of the *PsMYB10.1* copies, while in this copy the STOP codon in the last exon was not detected, causing the spurious automatic identification of five extra exons along the next 6 kb of sequence. By BLAST we identified three additional genes: one was homologous to *MYB10.1* and two homologous to *MYB10.2*. In the Zhongli-2 region, four *MYB10* genes were annotated: two *MYB10.1* copies, one *MYB10.2* and one *MYB10.3*. The second exon was also mis-annotated in most of the gene copies of this region. BLAST analysis did not identify additional *MYB10* genes.

To identify and visualize the differences in size and gene copy number and organization, we produced dot plots comparing the three LG3-MYB10 region assemblies (**Figure 3.1**). These reflected the high variability and poor homology along the region, with inverted assemblies, mis-assemblies and missing fragments. The Zhongli-1 *MYB10* region was assembled in an inverted orientation, when compared with 'Sanyueli' and Zhongli-2. Despite this, the Zhongli-1 (of 271 kb) and the Zhongli-2 (90 kb) alignments were collinear and with high homology (69% of sequence hits), with an evident gap of ~150 kb due to a missing fragment in Zhongli-2 explaining most of their difference in size (270 kb vs 90 kb). The percentage of hits between 'Sanyueli' and Zhongli-1 was lower (51%), with multiple misalignments along the region that explained the smaller size in 'Sanyueli'. Similarly, a discontinuous alignment and an inversion was also observed between Zhongli-2 and 'Sanyueli' (51% of sequence hits).

Similar comparisons were constructed for other *Prunus* species with more than one available reference genome (**Additional File 3.1**, **Additional File 3.2**). The MYB10 regions of the two sweet cherry (*P. avium*) genomes were highly collinear with 75% of sequence hits. In peach (*P. persica*), the region was collinear and highly similar in the 'Lovell' and 'Chinese Cling' genomes (85% of sequence hits), and three times larger than in the wild relative *P. mira,* with a mean of 55.5% of

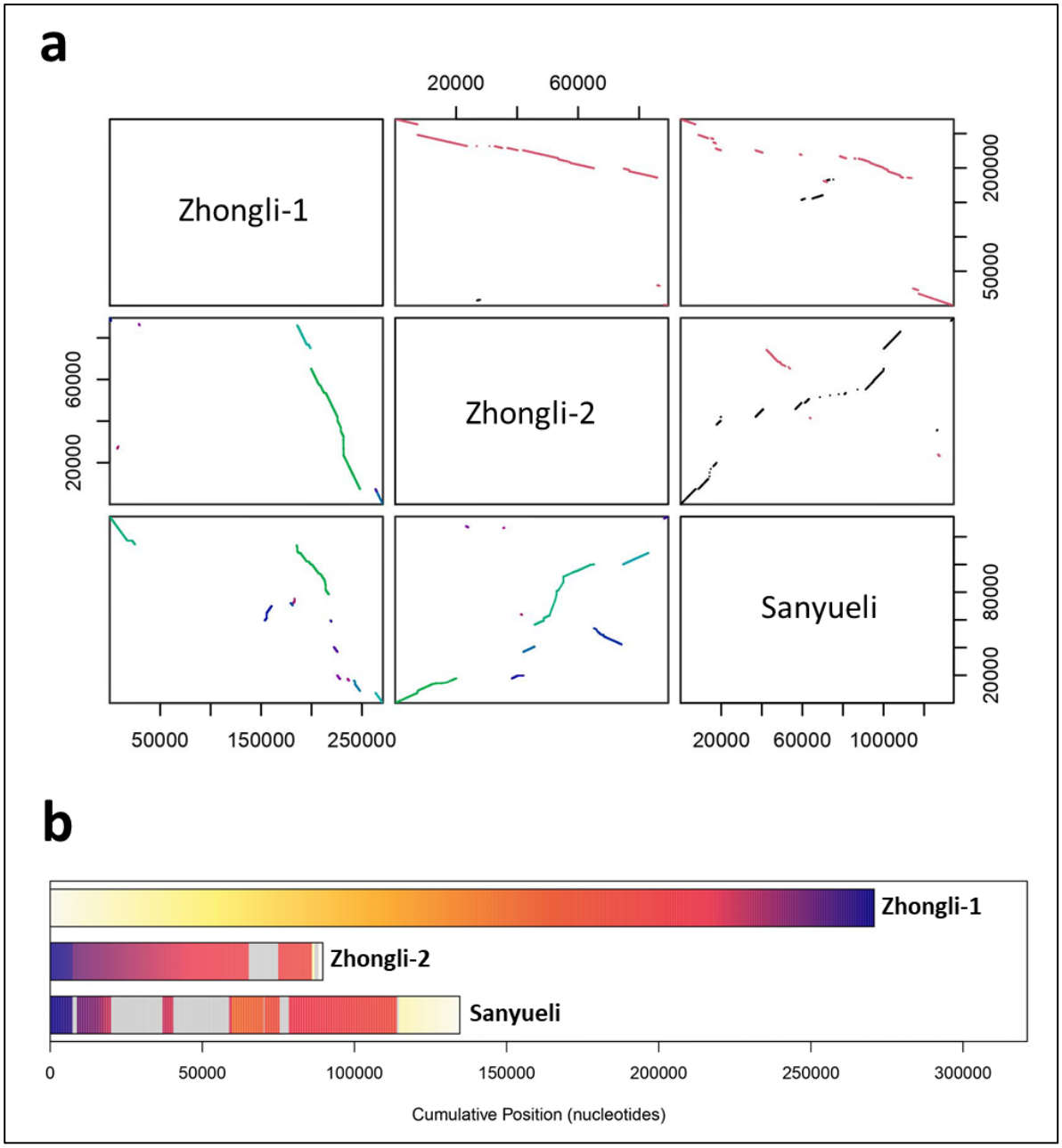


Figure 3.1. **a** Dot plot comparison of the three LG3-MYB10 assemblies in the two Japanese plum genomes published: 'Zhongli No. 6' (with two assemblies: Zhongli-1 and Zhongli-2) and 'Sanyueli'. Top right of the diagonal: homologous hits; inversions colored in red. Bottom left of the diagonal: homologous blocks colored by their score, from green to colder colors while the homology value decreases. Coordinates of each region are written in base pairs for both axes. **b** Distribution of the homologous sequences in the three compared genome regions. Grey areas correspond to sequences not found in the Zhongli-1 MYB10 region. Homologous regions are represented in the same yellow to purple colors.

sequence hits. In apricot, we compared five *P. armeniaca* accessions and three wild apricot species (two *P. sibirica* and one *P. mandshurica*). Homology between each *P. armeniaca* pairwise comparison ranged from 44% to 91% (mean of 67.7% of hits). When these were compared to the wild apricot genomes, the homology ranged from 58% to 93% (mean of 74.8% of hits) with *P. sibirica*, and from 47% to 59% (mean of 53.2% of hits) with the *P. mandshurica* genome. The mean percentage of sequence hits between the *Prunus* sections was 19.4% with values ranging from 8.6% to 37%.

The designed gRNAs successfully cut LG3-*MYB10* genes

For selective high-throughput sequencing of the desired LG3-*MYB10* region in Japanese plum varieties, we designed seven CRISPR RNAs (crRNAs) in genic regions (exons and introns) of the *PsMYB10.1*, *PsMYB10.2* and *PsMYB10.3* genes, including regions conserved between the genes to minimize the number of required crRNAs. The crRNAs were designed at the two DNA strands to ensure forward and reverse sequences and to direct the sequencing from the inside to the outside of the gene, covering the whole region with concatenated sequences (**Figure 3.2**). The *PsMYB10.1*, *PsMYB10.2* and *PsMYB10.3* sequences obtained in Fiol et al. (2021), the peach 'Lovell' reference genome, and other *Prunus* genomes were used to design the crRNAs with higher on-target capacity and lower off-target risk (see Methods). Therefore, crRNA-s1 was designed to cut within the exon 2 of *MYB10.1* and *MYB10.2* genes in the plus strand, while crRNA-s3 was directed to the same exon and strand in *MYB10.3* (**Additional File 3.3**). To avoid a putative mis-cut due to a possible SNP in the second exon of the *MYB10.2* cleavage site of crRNA-s1 (although only identified in *P. cerasifera*, *P. domestica* and *P. avium*) we designed crRNA-s2 for the alternative nucleotide. As reported by Fiol et al. (2021), variability in the introns was higher than in exons, so the crRNAs were designed taking into consideration the variants known. Introns of *PsMYB10.1*, *PsMYB10.2* and *PsMYB10.3* were targeted with crRNA-a1, crRNA-a3 and crRNA-a4, respectively, while crRNA-a2 was designed to include an SNP variant in one of the *PsMYB10.1* alleles present in the sample (*PsMYB10.1*-H1 in [26]). In addition, to cut the MYB10 genomic region at each flank, two crRNAs (crRNA-f1 and crRNA-f2) were designed in conserved exon sequences of the nearest flanking genes upstream (Prupe.3G162900 in peach, Pav_sc0000464.1_g320.1.mk in sweet cherry genomes) and downstream (Prupe.3G163400 in peach, Pav_sc0000464.1_g090.1.mk in sweet cherry genomes). (**Figure 3.2, Additional File 3.3**).

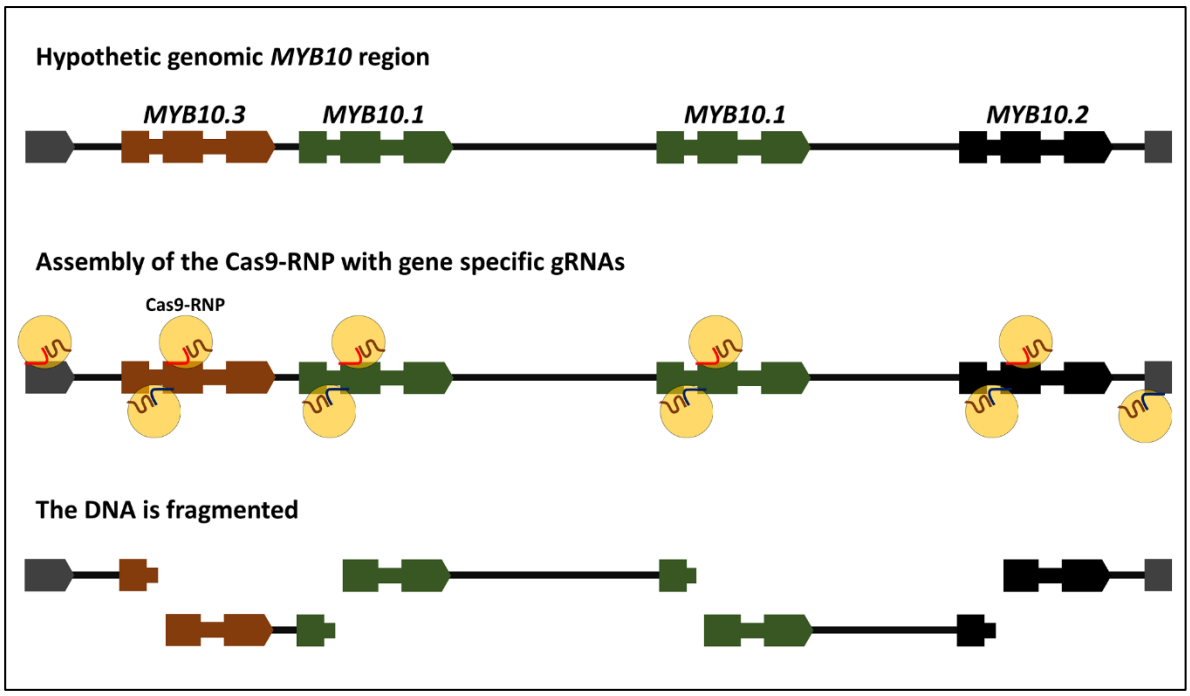


Figure 3.2. Schematic representation of the CRISPR RNA (crRNA) design to specifically cleave and sequence a hypothetic LG3-*MYB10* genomic region in Japanese plum. The guide RNAs (gRNAs) are formed by the interaction of trans-activating RNAs with crRNAs, then with the Cas9 enzyme are assembled into the Cas9- Ribonucleoprotein (Cas9-RNP) complex. The Cas9-RNPs allows the specific cleavage of the LG3-*MYB10* region, generating cuts at the two DNA strands for further sequencing in both directions from the inside to the outside of the genes. Gene number, order and size of the region, as well as the expected number of fragments were unknown at the time of the crRNA design.

We first validated the cleavage ability of the crRNAs in PCR amplified DNA. For that we amplified the *MYB10.1*, *MYB10.2* and *MYB10.3* complete genes in the plum variety 'Angeleno' and cleaved the products with a pool of the seven gRNAs targeting the *MYB10* genes only. As intron 2 in one of the *PsMYB10.1* alleles in 'Angeleno' is about 1.5 kb larger [26], the *MYB10.1* primers yielded fragments visualized in two bands in **Figure 3.3a** (the one containing the large intron is very faint in the gel, due to the preferential amplification of the shorter one). The fragments were cut and visualized in five bands. Three of them corresponded to the fragments 2 (526 bp), 3 (1078 bp), 4 (613 bp) and 5 (991 bp) in the *in-silico* design shown in **Figure 3.3b**. The two faint additional bands corresponded to fragments 3 and 5 (of about 2.5 kb) of the allele with the large intron 2, and to an undigested fraction of the product of the preferentially amplified copy. Similar results were observed for the *MYB10.2* amplicon, which was cut into four fragments. The band corresponding to the small gene fragment (fragment 6 in **Figure 3.3b**) spanning the exon and intron crRNA sites was not produced in either the *MYB10.1* or the *MYB10.2* cleavages. The pool was not able to cut the

MYB10.3 amplicon. A posterior sequencing of that amplicon revealed an unexpected 210-bp deletion affecting part of intron 1 and all of exon 2, and so lacking the two scission target sites.

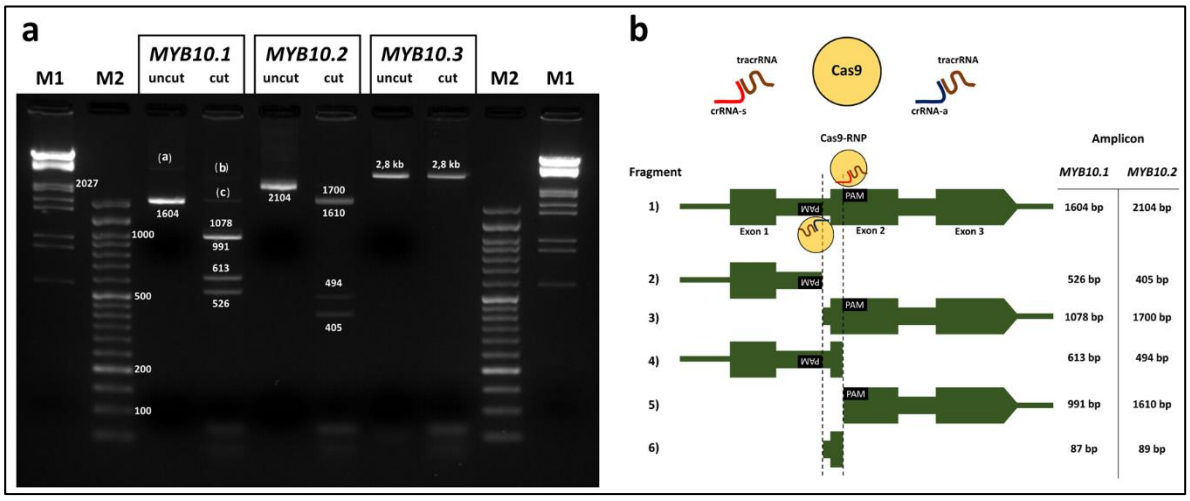


Figure 3.3. **a** MYB10 amplicon bands before and after cleavage with a pool of the guide RNAs (gRNAs) targeting the genes, with their size annotated in base pairs (bp). Bands (a, b) correspond to the faint bands from the allele with a large intron 1, (c) to the remaining undigested product. Wells M1: Lambda DNA HindIII-EcoRI digested molecular weight marker; M2: DNA Ladder 50 bp ready-to-use (GeneON). **b** Expected fragment sizes after *MYB10.1* and *MYB10.2* digestion with the Cas9-Ribonucleoprotein (RNP) complex. Guide-RNAs assembled with the trans-activating CRISPR RNA (tracrRNA) and crRNA-s (s1, s2 or s3) cut the exon 2 of the gene, gRNAs with crRNA-a (a1, a2, a3 or a4) cut by intron 1. Fragment 6 was not obtained after cleavage, fragment 3 and 5 bands overlapped in the gel.

Sequencing, alignment to the reference genomes and variant calling

The genomic DNA of five plum varieties (**Table 3.1**) was cleaved with the pool of gRNAs and barcoded. A pool of the output reactions was sequenced with the ONT MinION device, which produced 194 Mb of sequences in 33,794 reads of average mean length of 6,097 bp (**Table 3.1**). After demultiplexing the reads into varieties through their barcode, we obtained sequence yields ranging from 29.32 Mb to 45.64 Mb in reads of mean length from 4,110 to 7,778 bp; the average N50 was 14,510 bp.

The reads of each of the five varieties were aligned against the 'Sanyueli' and 'Zhongli No. 6' genomes. The coverage and depth of the alignments within the MYB10 region varied for each sample and reference genome used (**Table 3.2, Additional File 3.4**). Higher coverage was observed for the alignments against the shorter region of Zhongli-2 (90 kb long, 63.4% of coverage in average) while the coverage of the alignments against Zhongli-1 (54.2% of coverage, 271 kb long) was only slightly lower than that for 'Sanyueli' (57% of coverage, 135 kb long). The on-target alignment did

not correlate with the number of reads yielded, which was uneven for each variety. In particular, 'Golden Japan' was the variety with the lowest number of on-target aligned reads (168, 67 and 45 in 'Sanyueli', Zhongli-1 and Zhongli-2, respectively). The mean sequence depth was also unequal for each variety-reference genome combination. In general, for all varieties, a higher depth and higher number of on-target aligned reads was observed with 'Sanyueli' (**Table 3.2**). The variety with the lowest depth was 'Golden Japan' due to the poor alignment, while 'TC Sun' performed better.

Table 3.1. Sequencing statistics for each plum variety after read demultiplexing.

Variety	Barcode	Total reads	Yield (Mb)	Mean length (bp)	N50	Mean Q
'Angeleno'	NB01	4,995	38.85	7,777.9	19,049	11.0
'Black Gold'	NB02	11,106	45.64	4,109.7	10,288	10.9
'Fortune'	NB03	5,417	29.32	5,412.2	11,786	10.8
'Golden Japan'	NB04	5,063	34.98	6,908.0	16,740	11.0
'TC Sun'	NB05	7,213	45.27	6,275.8	14,686	10.9

In general, the depth along the region was unevenly distributed, with scattered regions of high depth (**Additional File 3.4**). This was caused by sequences that aligned partially and were clipped from one or both sides, indicating low homology with the genomes used as reference. Sequence clipping was not observed in 'TC Sun' aligned to Zhongli-1.

Variants were called for the alignment with 'Sanyueli', the one with the highest depth. A total of 3,261 SNPs in the *MYB10* region were found, 62.34% of them (2,033) being present in at least two varieties (**Figure 3.4a**), 98.5% with the same alternative allele (**Additional File 3.5**). All six observed MYB10 haplotypes (H1–H6) defined by Fiol et al. (2021) plus one inferred (H9) were present in the five varieties, three of them shared by pairs (H1 by 'Angeleno' and 'Black Gold'; H3 by 'Angeleno' and 'Fortune'; and H4 by 'TC Sun' and 'Golden Japan'). This was used to assign the phase of some of the polymorphisms: 'Angeleno' and 'Black Gold' had 259 private SNPs which may correspond to the shared H1; 'Angeleno' and 'Fortune' had 62 that could correspond to the shared H3. 'Fortune' and 'Black Gold' had 87 private SNPs which should be in H2 and H6. Unfortunately, due to the poor

Table 3.2. Alignment of each cultivar to the three Japanese plum *MYB10* reference regions.

Sample	MYB10 region								
	'Sanyueli' (135 kb)			Zhongli-1 (271 kb)			Zhongli-2 (90 kb)		
	On-target reads (%)	Cov. bases (%)	Mean depth	On-target reads (%)	Cov. bases (%)	Mean depth	On-target reads (%)	Cov. bases (%)	Mean depth
'Angeleno'	1,036 (5.63%)	73.3 kb (54.49%)	12.2x	489 (4.17%)	96.7 kb (35.7%)	3.8x	486 (4.15%)	48.7 kb (54.34%)	9.1x
'Black Gold'	1,300 (5.37%)	76.8 kb (57.07%)	19.1x	572 (2.58%)	123.0 kb (45.42%)	5.5x	611 (2.75%)	77.7 kb (86.31%)	14.1x
'Fortune'	1,692 (10.47%)	78.1 kb (58.05%)	23.2x	841 (8.60%)	165.1 kb (60.97%)	9.1x	616 (6.30%)	72.7 kb (81.18%)	18.5x
'Golden Japan'	168 (0.99%)	71.1 kb (52.85%)	2.0x	67 (0.63%)	133.8 kb (49.42%)	1.2x	45 (0.43%)	31.6 kb (35.31%)	1.0x
'TC Sun'	1,824 (7.04%)	84.5 kb (62.75%)	29.1x	677 (6.23%)	214.8 kb (79.33%)	19.3x	280 (2.58%)	53.6 kb (59.84%)	11.1x

On-target reads: reads that aligned to the region and the percentage from the total reads aligned to the genome; Cov. bases: the bases covered with at least one read and the percentage of the region size; Mean depth: the mean depth of coverage for the whole region.

quality of the 'Golden Japan' sequencing the phase of the SNPs in H4, H5 and H9 could not be inferred.

Two different software tools were used to call SVs and polymorphisms affecting more than 30 bps, identifying 287 variants: 72 in 'Angeleno', 63 in 'Black Gold', 68 in 'Fortune', 30 in 'Golden Japan' and 54 in 'TC Sun'. Most corresponded to sequence breakends (47.39%), followed by deletions (27.87%) and insertions (12.89%) (**Figure 3.4b**).

To check the accuracy of the SVs call we searched for the 44-bp insertion containing a G-box motif in the *PsMYB10.1a* promoter in the H1, H3 and H9 haplotypes in Fiol et al. (2021). The insertion was successfully annotated in 'Angeleno' (H1/H3), 'Black Gold' (H1/H2) and 'Golden Japan '(H4/H9), but not in 'Fortune' (H3/H6).

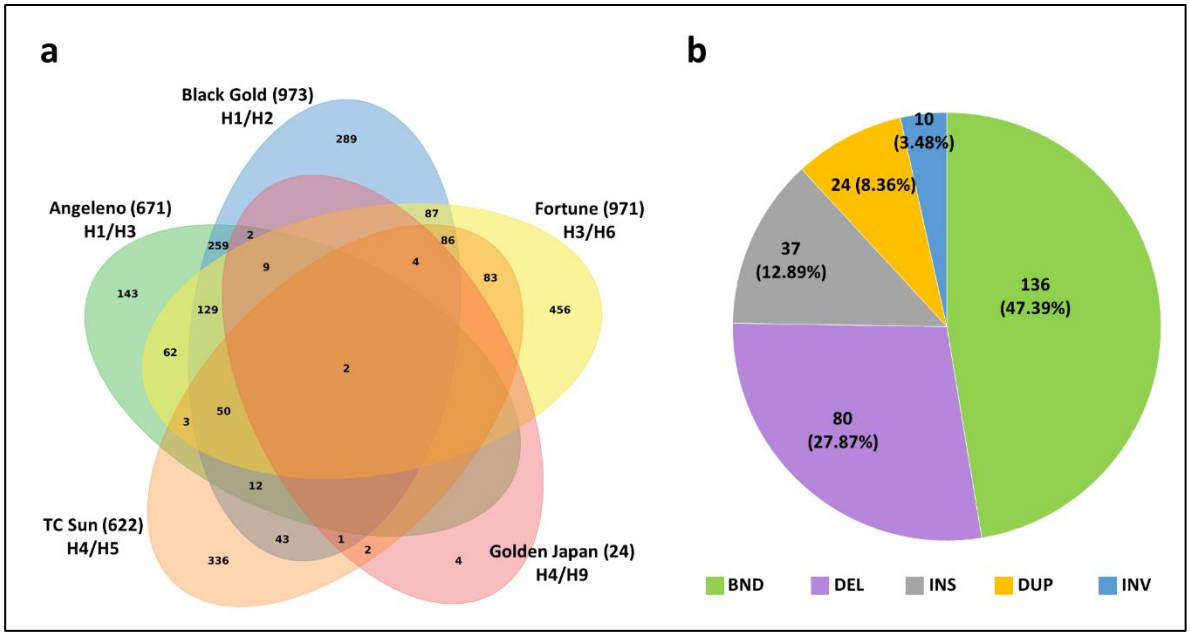


Figure 3.4. **a** Venn diagram showing the number of SNPs identified in the MYB10 region from each of the five Japanese plum samples sequenced and the overlap between them. The number next to the cultivar stands for their total SNP count. **b** Classification and percentage of all the structural variants identified: breakends (BND), deletions (DEL), duplications (DUP) and inversions (INV).

de novo assembly and homology visualization

Despite being able to identify new polymorphisms with this method, we carried out a *de novo* assembly to overcome the low coverage caused by sequence clipping in the regions with low homology. First we constructed an assembly for each variety, then we used the 'Zhongli No. 6' and 'Sanyueli' genomes to scaffold the contigs along the region. As expected, the degree of homology

and collinearity varied for each assembly (**Additional File 3.6**). In general, higher homology was observed in comparisons of the assemblies of the five varieties sequenced with the CRISPR-Cas9 selective cleavage strategy, particularly in those sharing a haplotype. In particular, the assemblies of 'Angeleno' and 'Black Gold' were always the most similar, with hits ranging from 41% to 53%, depending on the reference region used for assembly.

Identification of polymorphisms in the *PsMYB10.1a* promoter

Given the low homology between the samples and the 'Zhongli No. 6' and 'Sanyueli' genomes, we searched for polymorphisms among the *de novo* assemblies only. Using as an example a 250 bp contig upstream *PsMYB10.1*, previously sequenced with Sanger by Fiol et al. [26], we successfully identified the SNPs and InDels in heterozygosis, among them the 41 bp deletion in H3 and the 44 bp insertion in the H1, H3 and H9 haplotypes. The former (position 66 to 76 in the contig, **Figure 3.5**) was shared by 'Angeleno' (H1/H3) and 'Fortune' (H3/H6). The latter (position 200 to 243 in the alignment, **Figure 3.5**) was present in the varieties with H1, H3 or H9, in agreement with the results of Fiol et al. [26]. Most of the SNPs detected were consistent with those previously found, however we identified a few sequencing errors. This was the case of the SNPs in positions 52, 137, 175, and 244–245 in the alignment (**Figure 3.5**).

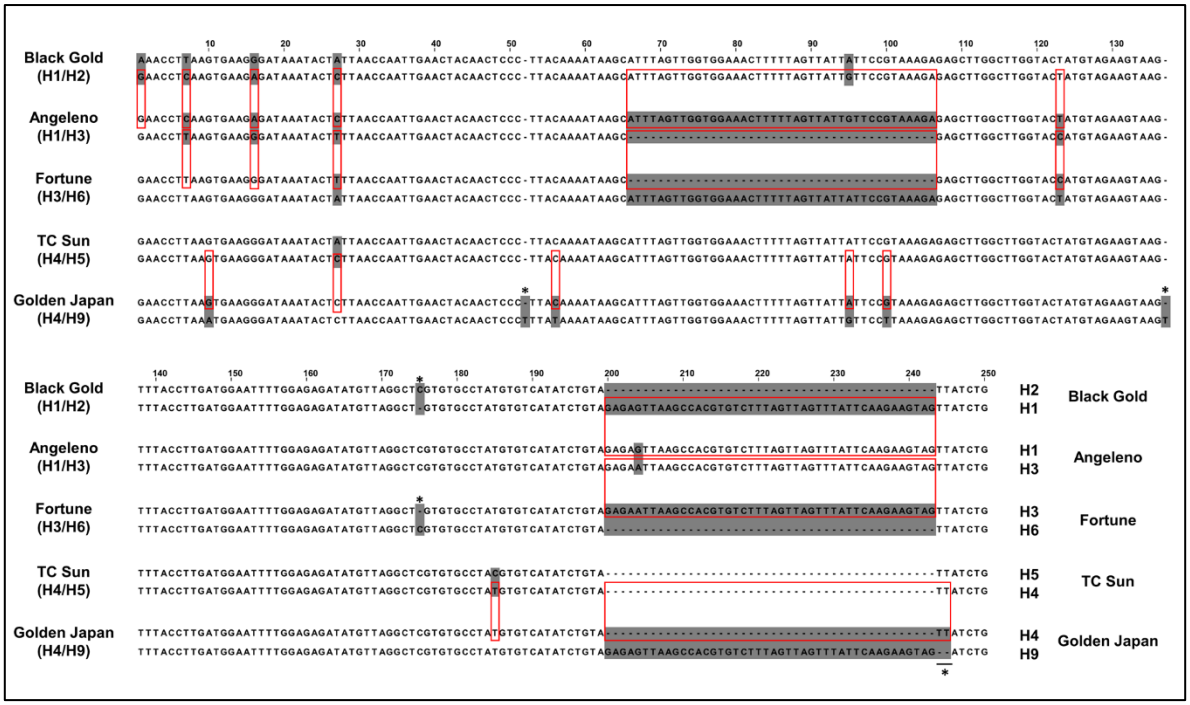


Figure 3.5. Sequence comparison of the five Japanese plum varieties sequenced. Heterozygous positions (shadowed bases) matched the position on the same haplotype in a different variety (red rectangles). Sequencing mistakes are marked with asterisks.

Despite these few sequencing errors, intrinsic to the ONT sequencing technology, the results validate this methodology for the resolution and identification of variability in highly complex regions.

DISCUSSION

In this work we confirm sequence enrichment CRISPR-Cas9 selective cleavage as a cheap and efficient high throughput strategy to identify variants and validate its use in highly complex regions in heterozygous genomes. Complex genomic regions in plant genomes are the basis of agriculture and crop diversity. They frequently include small and large variants, transposable elements (TE) and clusters of tandem or segmental duplicated multi-copy genes. Such genes are more likely to carry higher levels of variation [29] and be affected by presence/absence variation (PAV) than singleton genes [30], and often intervene in processes related to the evolution of agronomic traits such as secondary metabolisms, response to environmental stimuli, plant defense response and stress tolerance, flowering time, grain size and fruit traits [29, 31-34]. This makes characterizing such regions highly relevant for evolutionary studies as well as for plant breeding [35].

It is difficult, or even impossible, to completely resolve plant genome complexity by short read sequencing. Long-read approaches have improved the contiguity of genome assemblies, while special effort is currently being made to resolve particularly complicated regions of relatively small sizes in selected genotypes. In humans, several authors have used a strategy based on enrichment and sequencing using targeted cleavage of chromosomal DNA with Cas9 combined with Nanopore sequencing to resolve genomic SVs as well as mobile elements [23, 36-38]. In plants, this strategy has been validated to fine map short regions [24] and to characterize the transposon insertion landscape [39]. Here we extend this strategy for the characterization of much longer regions in a pool of genotypes in species without a reference genome or in regions with large variability in gene content.

As the region of interest (ROI), we selected the *Prunus* chromosome 3 *MYB10* region, which consists of a cluster of *MYB10* gene copies with a gene content variable among *Prunus* species. The genotypes studied were five commercial varieties of Japanese plum. Commercial varieties of this fruit crop derive from multiples crosses of *P. salicina* with other diploid plum species, which enhanced the fruit quality [40].

To reflect the complexity within the ROI we analyzed the synteny within and between *Prunus* sections, including two recent *P. salicina* genomes (one for the variety 'Sanyueli' and one for 'Zhongli No. 6') that were released when our study was already at a very advanced stage. Despite belonging to the same species, the level of homology in this region of these two Japanese plum varieties was low, comparable to that observed between cultivated and wild members from the same section of apricot and peach. By contrast, the homology within each of the *Prunus* sections studied (apricot, sweet cherries and peach) was higher or moderately higher than those obtained for Japanese plums. Such low level of homology in the ROI within *P. salicina* could be attributable to a possible interspecific origin of the varieties together with a putative mis-assembly of the region in the 'Zhongli No. 6' genome, in which two LG3-MYB10 regions were assembled 2 Mb apart instead of the single one expected for *Prunus*. The low identity between the *Prunus* ROI contrasts with the high synteny described at the whole genome level within the *Prunus* genus [41] and confirms the high complexity of this region in *Prunus* in general and in Japanese plums in particular, which validates its selection as a ROI for this study.

Previously in Fiol et al. [26], we observed six haplotypes from the segregation of *PsMYB10* alleles in six F1 Japanese plum progenies and advanced breeding lines, finding high allelic variability and the triplication of the *MYB10.1* gene in some haplotypes. Using the limited homology between *Prunus* genomes and by designing numerous PCR primer combinations, we were able to clone some of the alleles and phase the variants. This was a slow and costly strategy, and we were not able to obtain the full sequence of the haplotypes.

In contrast, here we complemented the Cas9 enrichment strategy with a multiplex approach to decrease the cost of sequencing six of the observed haplotypes (named H1 to H6). All the crRNAs were pooled for cleavage of the DNA in a single reaction, and the barcoded fragments of each variety were pooled, in equimolar proportions, for their sequencing on a MinION device. The five varieties selected carried the haplotypes in heterozygosis, three of them (H1, H3 and H4) shared by pairs. While this design was useful for optimizing the sequencing sample set and for phasing the variants, it was not determinant for the success of the methodology.

The crRNAs were designed within conserved regions of intron 1 and exon 2 of the *Prunus MYB10* genes. The *MYB* genes encode a family of transcription factors (TF) present in all eukaryotic organisms. The MYB10 proteins belong to one of the largest TF groups, the R2R3-MYB class, one of

the four *MYB* gene classes in seed plants [42]. As occurs with the majority of multi-copy gene families, multiple phylogeny and diversity studies in plant lineages have identified conserved motifs within the gene sequence. Such regions can be successfully used to direct crRNAs.

Design of the crRNAs

Here we targeted some of the conserved regions identified in Fiol et al. [26] in addition to two regions, one in each of the two closest genes up and downstream of the *MYB10* cluster, to border the ROI. This work was started several months before the two Japanese plum genomes were released. Although their availability could have helped in designing the crRNAs, our results indicate that this strategy can be successfully applied to those species lacking a reference genome.

To obtain the sequences of the genes and their upstream and downstream regions, we designed seven non-overlapping forward and reverse crRNAs able to direct the sequencing in both directions. Due to the small length of the *MYB10* sequences available and the high variability in the region, which limited the target region, the on-target activity of the crRNAs was far from optimal (51 to 68, on a scale of 1–100). Their cleavage activity was tested *in vitro* with *MYB10* amplicons of the variety 'Angeleno'. The results revealed that while the activity values would likely not perform well for *in vivo* gene editing, they were sufficient for *in vitro* cleavage with the system used here. In addition, *in vitro* digestion revealed that fragment 6 shown in **Figure 3.3b** was not obtained, indicating that Cas9 cannot simultaneously excise two targeted regions when these are only a few base pairs apart (about 100 bp in our design). This may be due to the Cas9 remaining bound in the cleaved DNA sequence on the 5'-side of the gRNA, allowing preferential ligation of the sequencing adapters to the available end and so providing sequencing directionality [23, 43, 44].

The pool of crRNAs did not cleave the *MYB10.3* amplicon tested. Further Sanger sequencing of the uncut amplicon showed that one of the gene copies missed both cleavage target sites. This allele was different from the two *MYB10.3* alleles previously identified in 'Angeleno', suggesting that the PCR primers had amplified an unknown duplicated *MYB10.3* copy. When we later visualized the genome alignments, reads were observed starting from the *MYB10.3* annotated position in all the samples, indicating that the Cas9 could indeed cleave this gene with the gRNA pool.

Sequencing and alignment to reference genomes

The crRNAs were pooled to cleave the DNA of each variety in a single reaction. This strategy made it possible to enclose the ROI and cleave the MYB10 genes along the whole region, regardless of their copy number and organization, which was unknown.

Pooling the barcoded cleaved DNAs lowered the sequencing costs, although it reduced the number of reads obtained per variety, which could have risked the depth of coverage. We obtained a total of 194 Mb of reads, 38.8Mb on average per variety, which represents 288 times the number of bases in the ROI of 'Sanyueli', 433 in that of Zhongli-2 and 143 times in Zhongli-1. The depth and coverage obtained after alignment were lower than expected from these figures. The lowest depth values were for the alignments with the 'Zhongli No. 6' genome, which was reasonable considering that sequences were aligned to the two assembled MYB10 regions, therefore reducing the depth at each of them. However, the depth in the alignments with 'Sanyueli' was only slightly better (from 12.2x to 29.1x, if we dismiss the extremely low values of 'Golden Japan'). The depth is highly determined by the percentage of coverage and of the on-target reads. We did not observe a correlation between the number of reads obtained and the coverage nor with the on-target ratios. Indeed, the sequencing statistics for 'Golden Japan' were similar to those obtained for the other varieties, while all the alignment values were extremely low, indicating a poor enrichment probably caused by an inefficient Cas9 cleavage reaction. The coverage varied per cultivar and reference genome used, although in general it was low (57% on average) due to the low homology between the commercial varieties and *P. salicina* genomes along the ROI, which caused terminal-clipped alignments. The terminal-clipped effect was minor only in the alignment of 'TC Sun' - Zhongli-1, which could indicate a shared haplotype between these two cultivars. The ratio of on-target alignments obtained here (8.6%) was higher than reported in other studies (4.61% for multiple cuts in Gilpatrick et al. [23]; 2.08 and 3.04% in López-Girona et al. [24]. This higher performance could be attributable to the higher number of cleavage sites designed, generating more Cas9-digested DNA fragments with compatible ends for their posterior ligation with the sequencing adapters. This conclusion was also reached by McDonald et al. [38], who obtained a mean on-target value of 44% in the CRISPR Cas9-enrichment of TEs.

Since the complexity of the ROI is out of our control, higher depths would be more easily obtained by increasing the number of cleavage sites. As the crRNAs are pooled in a single cleavage reaction,

the cost of the method is only increased by the cost of the crRNAs employed, which is a small proportion of the total cost. In addition, we consider that by reducing the number of pooled varieties, and therefore increasing the number of reads, we would not have obtained much higher coverage or depth. The complexity of the region, its size and the number of cutting sites need to be considered in deciding the number of pooled varieties.

Variant identification in the alignments to a reference genome

Our results indicate that the strategy described here, applied in a region determining agronomic traits in a panel of selected varieties, can be successfully used to rapidly extract useful information for marker-assisted selection. The generation of long reads offers a higher mapping certainty [45], which is extremely useful for regions with segmental duplications to pinpoint sequence variants with high confidence levels.

We identified SNPs and SVs along the ROI involved in the regulation of fruit skin color. For SNP calling we used Longshot software [46], which not only took into account the error tendency of the ONT long-reads to avoid false positives but has also been reported to outperform the SNP calling method using only short-read sequences in regions with duplications. A total of 129 SNPs were shared by the three cultivars with red-to-purple colored skin and absent in the two with yellow hues; since their phases are known, they could be selected, after validation, as alternatives to the haplotype (PCR) marker provided in Fiol et al. [26].

For SV calling we used Sniffles software [47] complemented with NanoVar [48] to overcome the low depth of 'Golden Japan'. The accuracy of the pipeline was tested by searching a previously Sanger-validated insertion of 44 bp on the *PsMYB10.1a* promoter in the haplotypes associated with red skin color (H1, H3) and in H9. Sniffles could positively detect the polymorphism in homozygosity on 'Angeleno' (H1/H3) and in heterozygosity on 'Black Gold' (H1/H2), while NanoVar could detect it even in low depth on 'Golden Japan' (H4/H9). The insertion was not called in 'Fortune' (H3/H6), although some reads were found in low frequency. This indicates that the tools complemented each other well in overcoming their limitations, and that SVs could still be called even in low depth**.** In addition, we were able to phase the variants, finding the 8 bp insertion in exon 1 of the *PsMYB10.1a* gene in the H9 of 'Golden Japan' in phase with the 44 bp insertion associated with red skin. This 8 bp insertion produces an early STOP codon, which explains the yellow skin of this cultivar despite carrying the 44 bp insertion (**Additional File 3.7**).

de novo assemblies

We have shown that reads can be aligned to a reference genome or assembled *de novo* in the case of its absence or of bad homology. Here, to recover the positions lost due to the lack of homology with the reference genomes, we used the pipeline for read correction and subsequent *de novo* assembly provided in López-Girona et al. [24]. One of the main limitations in our study was that reads produced from two very distant adjacent genes had little chance to overlap. This was the case in the Zhongli-1 region, where two adjacent genes 124 kb apart left 55 kb of the region uncovered. The other limitation for obtaining a whole-region assembly was that reads could not overlap the cleavage positions because these were produced in a single reaction. Therefore, despite reducing the cost, cleaving in a single reaction reduced the accuracy of the assembly, meaning that there should be a compromise between cost and coverage of the assembly when designing the experiment. Due to these limitations, the region was assembled in several contigs, which we put in order using a reference-guided scaffolding method.

The highest sequence conservation and collinearity was between the 'Angeleno', 'Black Gold' and 'Fortune' group, and between the 'Golden Japan' and 'TC Sun' group comparisons, which was in agreement with the SNP results and the shared haplotypes of the samples. Lower levels of sequence collinearity were observed between samples of the different groups, although generally in several discontinuous blocks of low homology. Given the high complexity of the region and the low conservation levels observed between some haplotypes, we conclude that it was suitable to design crRNAs aimed at the gene sequences, which might be more conserved than the intergenic regions [49, 50].

Variant identification in the de novo assemblies

The pipeline used to identify variants between *de novo* assemblies succeeded in the identification and assignment of polymorphisms into haplotypes rather than compiling them. The alignments could be separated into haplotypes in three samples and there was enough resolution to correctly identify SNPs and large InDels, improving the results obtained when using a reference genome (the 44 bp insertion was detected in 'Fortune'). Two samples lacked one of their haplotype sequences, in one case probably due to the high similarity between haplotypes ('TC Sun') and in the other due to the low depth of coverage ('Golden Japan'), therefore, where a reference genome is available,

the alignment to it and *de novo* assemblies should be combined to ensure all polymorphisms are identified.

Strengths, weaknesses, improvements and opportunities

We present here an economic and powerful strategy to extract sequences and polymorphisms from highly complex regions which may contain duplicated and presence/absence variants (PAVs). The main strengths of the strategy are i) that the design of the crRNAs does not necessarily rely on a reference genome, with information of conserved regions or partially cloned alleles being sufficient, ii) that it is economic as the DNA of each variety is cleaved in a single reaction and equimolar amounts of each variety are pooled for a unique ONT run, and iii) that polymorphisms and their phases can be extracted with high efficiency, even in comparison with a reference genome or by analyzing the *de novo* assemblies. The main weaknesses are i) that the search for polymorphisms might require the manual isolation and comparison of contigs (with genes that might have undergone extensive duplication events), ii) that ONT technology is prone to sequence errors (the sequencing errors identified here consist of one and two nucleotide InDels), some of them occurring in short homopolymer regions reported to be an issue in most basecallers [51], and iii) that by using a single Cas9 digestion, overlapping fragments to assemble the region in one contig are not obtained.

This methodology is sufficiently versatile to allow for future improvement. Digesting the DNA region separately with sub-pools of crRNAs would allow for fragment overlapping (this strategy would increase the cost of the experiment, so it is recommended for extremely complicated regions). The scaffolding could be improved by selecting genotypes with the ROI in homozygosis (although this may be difficult to obtain depending on the ROI or the plant materials).

This method offers new opportunities in the identification of genomic variability in agricultural crops, including methylation variants. Changes in methylation patterns may have drastic impacts on the phenotypes. Although we have not explored such application, the methodology used could have allowed it as in Giesselmann et al. [25] and in Gilpatrick et al. [23]. In fruit trees, methylation changes in the promoter of *MYB10* genes are associated with changes in fruit color. Also, as in other rosaceous fruit, the fruit color in some Japanese plum varieties is acquired after exposure to cold or light, indicating a possible epigenetic mechanism. Studying methylation changes in the *PsMYB10*

region using DNA of the appropriate tissues under certain given conditions may provide important information to understand these regulation mechanisms.

The method that we propose is computationally inexpensive. The most demanding software used in this process is the Flye assembler [52], the cost of which increases with the size of the assembly. In the benchmark tested by the authors, a genome of 4.6 Mb (such as *E. coli*) obtained only with long reads could be assembled using 2 Gb of RAM in about two hours using one CPU. For ROIs less than 1 Mb long it is reasonable to assume that the protocol reported here could be undertaken on most home computers.

CONCLUSIONS

The CRISPR-Cas9 enrichment strategy enabled long-read sequences from a highly polymorphic and duplicated region to be produced and polymorphisms related to fruit skin color to be deduced from five heterozygous Japanese plum varieties. The methodology was simplified through using a pool of crRNAs in duplicated conserved regions and the sequencing cost was optimized by multiplex sequencing while keeping the protocol computationally inexpensive. The tool could be further improved by taking into account methylation patterns and also adapted for the full assembly of complex regions, being also appropriate for those crops with a lack of a reference genome or with one unrepresentative for the genomic region of interest.

MATERIALS

Plant material and nucleic acid isolation

Young leaves from five commercial Japanese plum varieties (**Table 3.3**) were collected and kept at −80ºC. Prior to DNA extraction, nuclei were separated as described in Naim et al. [53] and detailed online [54], but the vegetal material was homogenized by grinding in liquid nitrogen using a pestle and mortar. DNA was extracted from the isolated nuclei using the Doyle CTAB method [55], introducing an RNAse treatment before the chloroform centrifugation step.

The quality of the extracted DNA was evaluated using Nanodrop and low voltage 0.8% agarose TAE gels, and quantified using a Qubit fluorometer.

Table 3.3. Commercial varieties selected for CRISPR-Cas9 targeted sequencing of their *MYB10* region, their fruit color and the MYB10 haplotype combination described in Fiol et al. [26].

Variety	Fruit color		MYB10 Haplotypes
	Skin	Flesh	
'Angeleno'	Black	Yellow	H1/H3
'Black Gold'	Black	Red	H1/H2
'Fortune'	Red	Yellow	H3/H6
'Golden Japan'	Yellow	Yellow	H4/H9
'TC Sun'	Yellow	Yellow	H4/H5

Design of guide RNAs targeting conserved sites of *MYB10* genes

The Alt-R CRISPR-Cas9 System (IDT®) was selected to design and assemble the CRISPR RNAs (crRNA) and the trans-activating crRNAs (tracrRNA) into functional guide RNAs (gRNAs).

The crRNA sequences were designed in the desired positions with the IDT custom gRNA design tool. Those with the highest on-target capacity were BLAST to the Peach genome v2.0.a1 [56] and the Sweet cherry genome v1.0 [57] to discard gRNAs with off-target risk regions. SNP variability identified in other genomes was also considered in the crRNA design. Those gRNAs directed at the most highly conserved LG3-*MYB10* sequences were finally selected.

Assembly of the Cas9 RNA-Ribonucleoprotein complex (RNP)

The designed crRNAs and the tracrRNA were resuspended to 100 µM in Nuclease-Free Duplex Buffer (IDT). To form the gRNA duplex, all the crRNAs were pooled in an equimolar mix and combined with the same volume of tracrRNA to a final concentration of 10 µM. The mixture was denaturalized for five minutes at 95ºC and allowed to cool at room temperature. To assemble the Cas9-RNP complex for a single cleavage reaction, 1 µL of the gRNA duplex and 0.08 µL of Alt-R® S.p. HiFi Cas9 Nuclease V3 (IDT) were mixed with 1X of CutSmart Buffer (NEB) and nuclease-free water to a total of 10 µL then incubated for 30 minutes at room temperature and kept on ice for immediate use.

Cleavage assay of the designed gRNAs with PCR amplicons

The assembled Cas9-RNPs activity was tested using PCR amplicons of *PsMYB10.1a*, *PsMYB10.2* and *PsMYB10.3* full gene sequences from 'Angeleno'. The primer sequences and their annealing temperature (Ta) are listed in **Additional File 3.8**. Each PCR reaction contained 1.5 mM $MgCl_2$ and 1X NH_4 buffers, 1U of BioTaq polymerase (Bioline), 0.2 mM dNTPs, 0.2 µM of each primer, 40 ng of DNA from 'Angeleno' and MilliQ water to a total of 25 µL. Thermocycler conditions were 94ºC for 1 minute, 35 cycles of 94ºC for 30 seconds, the Ta described for each primer pair for 20 seconds and 72ºC for 2 minutes, followed by a final extension step of 5 minutes at 72ºC. Reactions were purified using ExoSap-IT (Thermo Fisher) and then 3 µL of each amplicon was diluted with 32 µL of nuclease-free water before adding 10 µL of assembled Cas9-RNP. Cleavage reactions were incubated at 37ºC for 20 minutes and five minutes at 72ºC, then loaded in a 2% agarose TAE gel to visualize the cleaved bands compared to the uncut amplicons.

Comparison of the MYB10 region in several *Prunus* whole-genome assemblies

A total of fifteen genomes were added to the analysis: two Japanese plums, eight apricots, two sweet cherries, two peaches and a wild peach relative. For all the genomes, the MYB10 region was considered from the BLAST [58] regions of crRNA-f1 and crRNA-f2 sequences. All genomes and the regions considered in this study are listed in **Additional File 3.2**. The homology between the regions was calculated and visualized pairwise using the DECIPHER v2.0 [59] package for R version 4.0.3 [60]. The *MYB10* genes in the two Japanese plum genomes were identified by BLAST using the *PsMYB10* sequences from Fiol et al. [26].

CRISPR-Cas9 enrichment, library preparation and sequencing

The high-quality nuclei DNA from the five selected commercial plum varieties and the pooled crRNAs (100 µM) were sent to CNAG (Centre Nacional d'Anàlisi Genòmica, Barcelona) for CRISPR-Cas9 enrichment targeted sequencing using Oxford Nanopore Technology (ONT). For each sample, 1000 ng of DNA diluted in 27 µL of nuclease-free water was dephosphorylated by the addition of 3 µL of 10X CutSmart Buffer and 3 µL of Quick Calf Intestinal Phosphatase (NEB) and then incubated at 37ºC for 10 minutes, 80ºC for two minutes and 20ºC for two minutes. To each reaction, 10 µL of previously assembled Cas9-RNP complex was added together with 1 µL of dATP (10 mM) and 1 µL of Taq polymerase (NEB) and incubated at 37ºC for 20 minutes and at 72ºC for five minutes for A-tailing. Samples were tagged using barcodes NB01 to NB05 of the Native Barcoding Expansion 1-12 PCR-free kit (EXP-NBD104, ONT) and purified using Agencourt AMPure XP beads (Beckman Coulter).

The barcoded samples were quantified using a Qubit fluorometer and pooled equimolarly to reach 700 ng in 65 µL. The Cas9 Sequencing Kit (SQK-CS9109, ONT) was used to ligate the AMX adapters before purification of the fragments with AMPure XP Beads. The beads were washed twice with Long Fragment Buffer (LFB) and then resuspended in Elution Buffer (EB) for 30 minutes at room temperature. The sequencing library was prepared by mixing 12 µL of the DNA library, 37.5 µL of Sequencing Buffer (SQB) and 25.5 µL of Loading Beads (LB) (SQK-CS9109, ONT). The sample of pooled fragments was sequenced in a MinION (v9.4.1) flowcell on a GridION mK1 device operated by MinKNOW version 3.6.5 software.

Basecalling, quality filtering, adapter removal, alignment to reference and variant calling

Nanopore read sequences were basecalled and demultiplexed using Guppy (v3.2.10+aabd4ec) high accuracy model (ONT). MinIONQC v1.4.2 [61] was used to evaluate the read quality scores before and after trimming barcodes and adapters with Porechop v0.2.4 (https://github.com/rrwick/Porechop). Reads were aligned to *Prunus salicina* 'Sanyueli' v1.0 [27] and *P. salicina* 'Zhongli No. 6' v1.0 [28] whole genome assemblies by using Minimap2 [62], and were visualized in IGV [63] and represented in bwtool [64]. Alignment information was obtained with Samtools version 1.11 [65]. Variant calling was done with software specifically scripted for third-generation sequencing reads: Longshot v0.4.3 [46] was run for SNP calling, while both Sniffles v1.0.12 [47] and NanoVar v1.3.9 [48] were run, and their output files merged with VCFtools v0.1.16 [66], to obtain the list of SV calls. VFCtools was further used to compare the lists of SNP calls between the five 'Sanyueli' genome alignments, and the results represented in a Venn diagram using pyvenn (https://github.com/tctianchi/pyvenn).

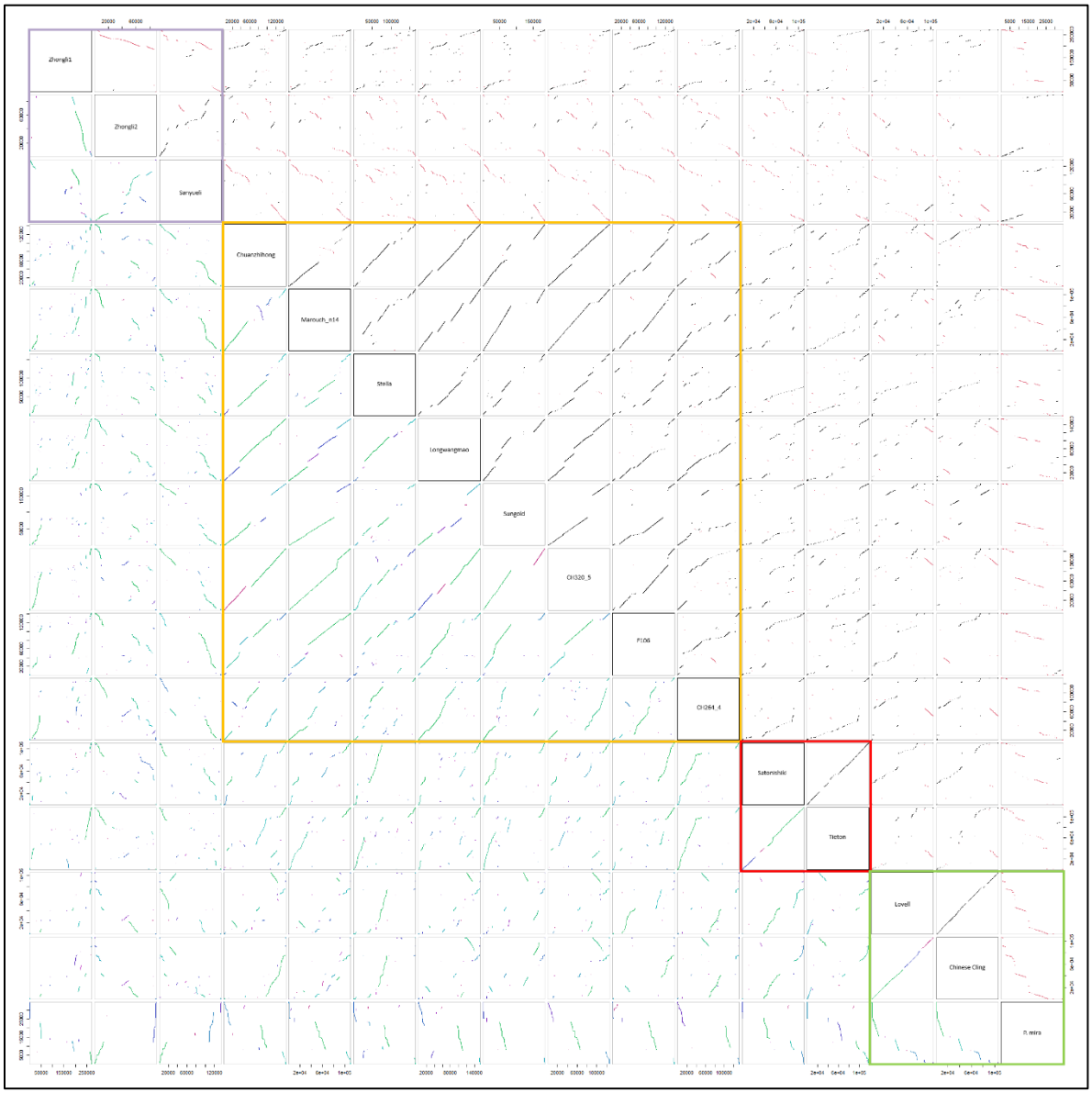
Reads correction, Flye assembly and reference-guided scaffolding

Canu v 2.1.1 [67] was used to correct the trimmed reads and perform an initial *de novo* assembly. Nanopolish v0.11.1 [68] was used to improve the assembly results. For each sequenced cultivar, their Canu corrected reads and polished contigs were used as input for Flye assembler v2.8.3 [52] enabling the option to keep haplotypes. The resulting contigs were genome-guided scaffolded with the RagTag software [69], using separately the three MYB10 regions from the two *P. salicina* available genomes as template reference sequences. DECIPHER v2.0 [59] was used to display the homology between the references and reference-guided sequences in a pairwise fashion and obtain their similarity values.

Identification of SNPs and InDels in the *PsMYB10.1a* promoter

A region of 250 bp in the *PsMYB10.1a* promoter containing several SNPs and large InDels was used for comparison. Sanger sequences of this region obtained for different haplotypes [26] were used in a BLAST analysis to recover the haplotypes in the *de novo* (Flye) contigs. In the case of the H4 sequence, which could not be recovered, reads were isolated from their alignment to the 'Zhongli No. 6' genome *PsMYB10.1a* promoter region and assembled manually. No Sanger sequences were available for H5 and H6, and their contigs on 'TC Sun' and 'Fortune' were isolated as alternative to H4 and H3 sequences. The selected contigs from each cultivar were aligned and compared using Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA).

ADDITIONAL FILES



Additional File 3.1. Dot plot comparing, pair-wise, the *MYB10* regions identified in 15 *Prunus* genomes, represented as in Figure 3.1a. The colored squares border the *Prunus* sections considered: purple for Japanese plums, orange for apricots, red for sweet cherries, and green for peaches and its wild relative. The image with full resolution can be found online: https://doi.org/10.1101/2022.01.24.477518

Additional File 3.2. Details of the *MYB10* region isolated from each of the fifteen *Prunus* genomes considered for pair-wise comparison. The results of the percentage of sequence hits and the number of shared blocks for every comparison is included.

Genome assemblies and the coordinates of the assembled MYB10 region considered for the comparative analysis

SECTIONS	Genome	MYB10 region coordinates	MYB10 region size (bp)	Reference
Japanese plums				
Zhongli-1	*Prunus salicina Zhongli No. 6 Genome v1.0*	LG03:30838572..31109359	270788	Huang et al., publication pending
Zhongli-2	*Prunus salicina Zhongli No. 6 Genome v1.0*	LG03:28592935..28682529	89595	
Sanyueli	*Prunus salicina Sanyueli Genome v1.0*	Chr4:12192580..12327184	134605	Liu et al., 2020
Apricots				
Chuanzhihong	*Prunus armeniaca Genome v1.0*	LG4:11257406..11401553	144148	Jiang et al., 2019
Marouch_n14	*Prunus armeniaca Marouch n14 Whole Genome v1.0*	chr3:14070728..14181189	110462	Groppi et al., 2021
Stella	*Prunus armeniaca Stella Whole Genome v1.0*	chr3:14773662..14939042	165381	Groppi et al., 2021
Longwangmao	*Prunus armeniaca Longwangmao Whole Genome v1.0*	Chr3:15339113..15493822	154710	Lin Wang et al., publication pending
Sungold	*Prunus armeniaca Sungold Whole Genome v1.0*	Chr3:15344002..15529870	185869	Lin Wang et al., publication pending
CH320_5	*Prunus sibirica CH320_5 Whole Genome v1.0*	chr3:23559518..23685710	126193	Groppi et al., 2021
F106	*Prunus sibirica F106 Whole Genome v1.0*	Chr3:14534446..14676083	141638	Lin Wang et al., publication pending
CH264_4	*Prunus mandshurica CH264_4 Whole Genome v1.0*	chr3:17172906..17304700	131795	Groppi et al., 2021
Sweet cherries				
Satonishiki	*Prunus avium Genome v1.0.a1*	chr3:12107728..12217510	109783	Shirasawa et al., 2017
Tieton	*Prunus avium Tieton Genome v2.0*	chr_3:23901605..24019020	117416	Wang et al., 2020
Peaches				
Lovell	*Prunus persica Genome v2.0.a1*	Pp03:18179394..18284092	104699	Verde et al., 2017
Ch_ Cling	*Prunus persica Chinese Cling Genome v1.0*	chr3:20361949..20468227	106279	Cao et al., 2021
P.mira	*Prunus mira Genome v1.0*	Pm03:10034003..10068377	34375	Cao et al., 2020 (preprint)

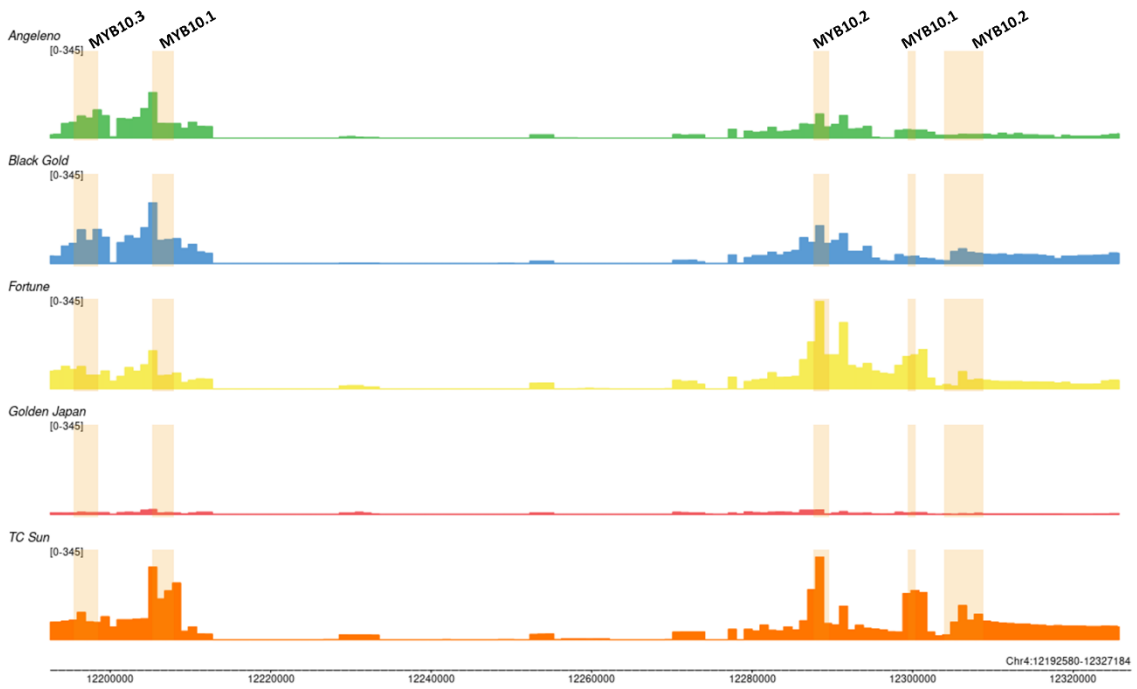
Sequence hit values and shared blocks between each identified MYB10 region on the selected Prunus genomes

MYB10 region	Zhongli-1	Zhongli-2	Sanyueli	Chuanzhihong	Marouch_n14	Stella	Longwangmao	Sungold	CH320_5	F106	CH264_4	Satonishiki	Tieton	Lovell	Ch_Cling	*P.mira*
Zhongli-1	1 seq	69% hits	51% hits	23% hits	29% hits	20% hits	23% hits	16% hits	17% hits	26% hits	23% hits	16% hits	9.1% hits	15% hits	13% hits	36% hits
Zhongli-2	7 blocks	1 seq	51% hits	25% hits	22% hits	23% hits	25% hits	23% hits	25% hits	26% hits	22% hits	14% hits	13% hits	15% hits	15% hits	24% hits
Sanyueli	11 blocks	10 blocks	1 seq	25% hits	23% hits	21% hits	28% hits	21% hits	21% hits	20% hits	22% hits	15% hits	13% hits	15% hits	20% hits	27% hits
Chuanzhihong	29 blocks	17 blocks	14 blocks	1 seq	61% hits	63% hits	87% hits	64% hits	88% hits	67% hits	59% hits	19% hits	16% hits	24% hits	13% hits	35% hits
Marouch_n14	24 blocks	10 blocks	11 blocks	6 blocks	1 seq	61% hits	87% hits	91% hits	93% hits	83% hits	52% hits	18% hits	11% hits	20% hits	16% hits	35% hits
Stella	34 blocks	21 blocks	23 blocks	6 blocks	11 blocks	1 seq	54% hits	44% hits	64% hits	59% hits	47% hits	9.3% hits	11% hits	10% hits	9.7% hits	37% hits
Longwangmao	27 blocks	13 blocks	17 blocks	3 blocks	5 blocks	6 blocks	1 seq	65% hits	87% hits	58% hits	58% hits	18% hits	18% hits	20% hits	20% hits	29% hits
Sungold	19 blocks	15 blocks	13 blocks	5 blocks	3 blocks	13 blocks	5 blocks	1 seq	87% hits	62% hits	50% hits	16% hits	11% hits	23% hits	23% hits	30% hits
CH320_5	23 blocks	13 blocks	12 blocks	2 blocks	1 block	7 blocks	3 blocks	2 blocks	1 seq	74% hits	37% hits	17% hits	14% hits	14% hits	13% hits	32% hits
F106	22 blocks	12 blocks	6 blocks	9 blocks	5 blocks	16 blocks	14 blocks	7 blocks	7 blocks	1 seq	42% hits	15% hits	15% hits	8.6% hits	12% hits	31% hits
CH264_4	18 blocks	11 blocks	9 blocks	10 blocks	8 blocks	19 blocks	11 blocks	15 blocks	14 blocks	16 blocks	1 seq	15% hits	9.2% hits	21% hits	20% hits	23% hits
Satonishiki	22 blocks	13 blocks	15 blocks	10 blocks	11 blocks	11 blocks	19 blocks	12 blocks	12 blocks	9 blocks	12 blocks	1 seq	75% hits	9.3% hits	10% hits	24% hits
Tieton	14 blocks	22 blocks	21 blocks	16 blocks	9 blocks	19 blocks	20 blocks	13 blocks	19 blocks	17 blocks	7 blocks	3 blocks	1 seq	10% hits	9.2% hits	21% hits
Lovell	16 blocks	9 blocks	12 blocks	17 blocks	15 blocks	13 blocks	18 blocks	15 blocks	16 blocks	11 blocks	13 blocks	8 blocks	13 blocks	1 seq	85% hits	56% hits
Ch_Cling	15 blocks	9 blocks	18 blocks	11 blocks	17 blocks	14 blocks	14 blocks	11 blocks	15 blocks	9 blocks	14 blocks	10 blocks	13 blocks	3 blocks	1 seq	55% hits
P.mira	8 blocks	3 blocks	4 blocks	7 blocks	7 blocks	11 blocks	6 blocks	5 blocks	4 blocks	6 blocks	4 blocks	7 blocks	8 blocks	7 blocks	7 blocks	1 seq

Additional File 3.3. Details of the crRNAs designed for Japanese plum LG3-*MYB10* region enrichment. The SNPs identified are shown underlined and crRNAs were designed including each variant.

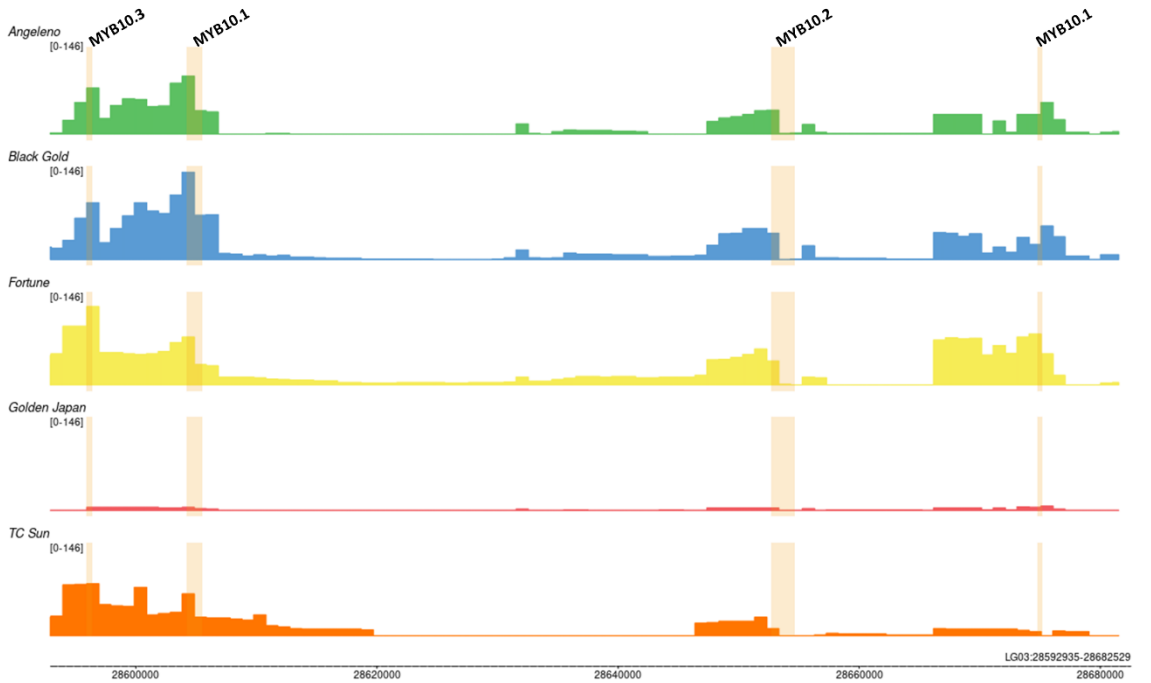
Name	Sequence 5'->3'	PAM	Target	Strand	On-target activity score
crRNA-s1	GGAAGAGCTGTAGACTAAGG	TGG	*MYB10.1* and *MYB10.2* (exon 2)	(+)	58
crRNA-s2	GGAGGAGCTGTAGACTAAGG	TGG	*MYB10.2* (exon 2)	(+)	50
crRNA-s3	GGAAGAGCTGCAGACTACGG	TGG	*MYB10.3* (exon 2)	(+)	57
crRNA-a1	ATAAGTCTCTTAGCACCCCT	CGG	*MYB10.1* (intron 1)	(-)	64
crRNA-a2	ATAAGTCTCTTAGTACCCCT	CGG	*MYB10.1* (intron 1)	(-)	67
crRNA-a3	AAACTTGTCATGAAATTATC	AGG	*MYB10.2* (intron 1)	(-)	51
crRNA-a4	ATTGTATAACATCTTTCTCG	AGG	*MYB10.3* (intron 1)	(-)	68
crRNA-f1	CGGGTGCAAGGCGTACCAAG	CGG	*Prupe.3G162900*	(-)	70
crRNA-f2	GACTACCCCTGCGAGTCCAG	AGG	*Prupe.3G163400*	(+)	61

Additional File 3.4 (1/3): Alignment in the Sanyueli MYB10 region



Additional File 3.4 (2/3): Alignment in the Zhongli-1 MYB10 region

Additional File 3.4 (3/3): Alignment in the Zhongli-2 MYB10 region



Additional File 3.4. Visualization of the depth of the sequences aligned to the 'Sanyueli', Zhongli-1 and Zhongli-2 regions. The pink bars highlight the coordinates with previously identified MYB10 gene sequences, where the Cas9 enzyme has two cutting points enabling sequencing in both directions.

Additional File 3.5. The total number of SNPs called from each sequenced sample and the count of the shared SNP positions between them.

Legend:

Barcode	Cultivar
NB01	Angeleno
NB02	Black Gold
NB03	Fortune
NB04	Golden Japan
NB05	TC Sun

One-on-one overlap SNP positions and alternate allele match

Sample	NB01	NB02	NB03	NB04	NB05
NB01	671				
NB02	463(460)	973			
NB03	255(251)	367(361)	971		
NB04	13(13)	18(18)	15(15)	24	
NB05	67(63)	198(192)	228(226)	9(9)	622

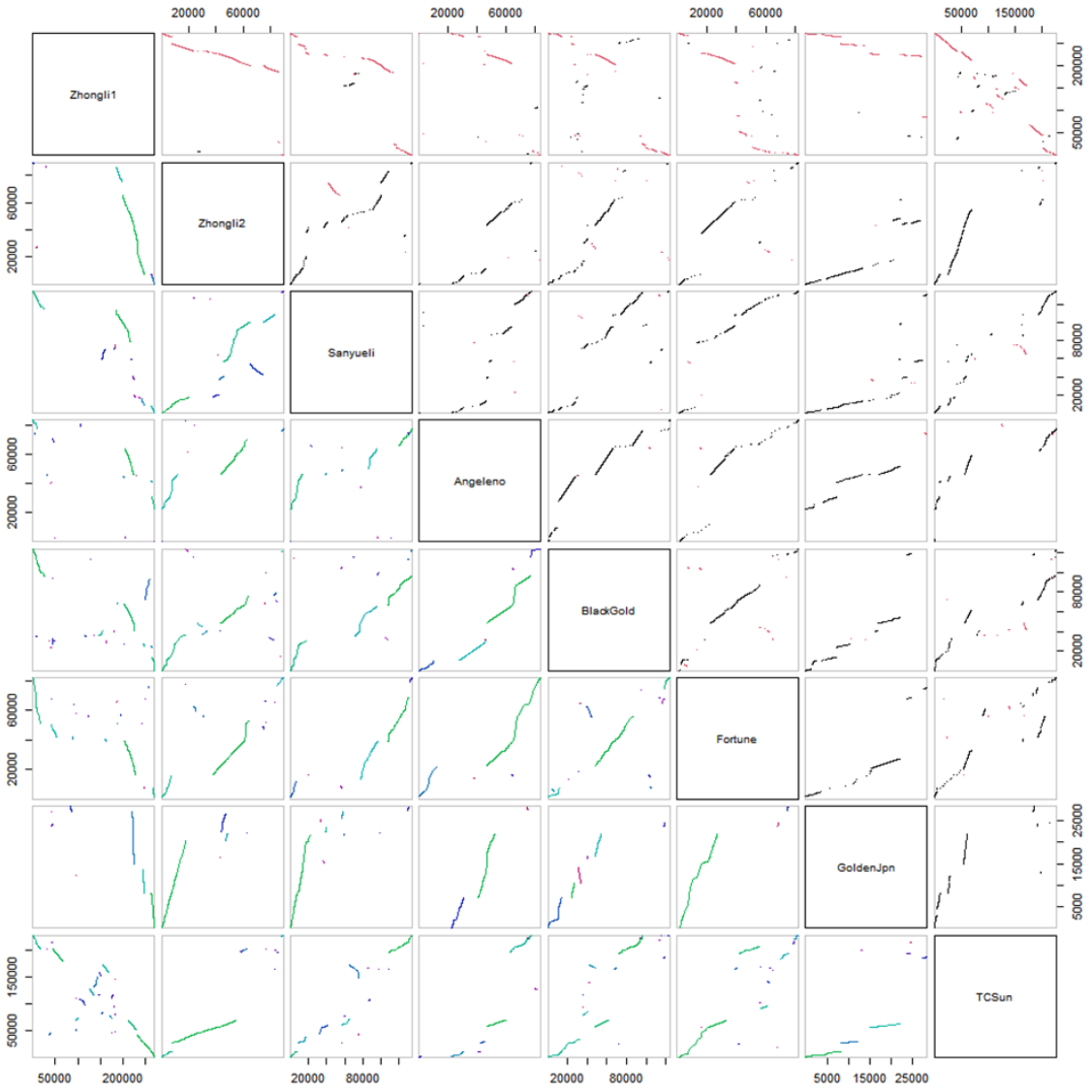
Number of Reference position matches (number of Alternative allele matches)

Multi-sample comparison of the shared SNP called positions

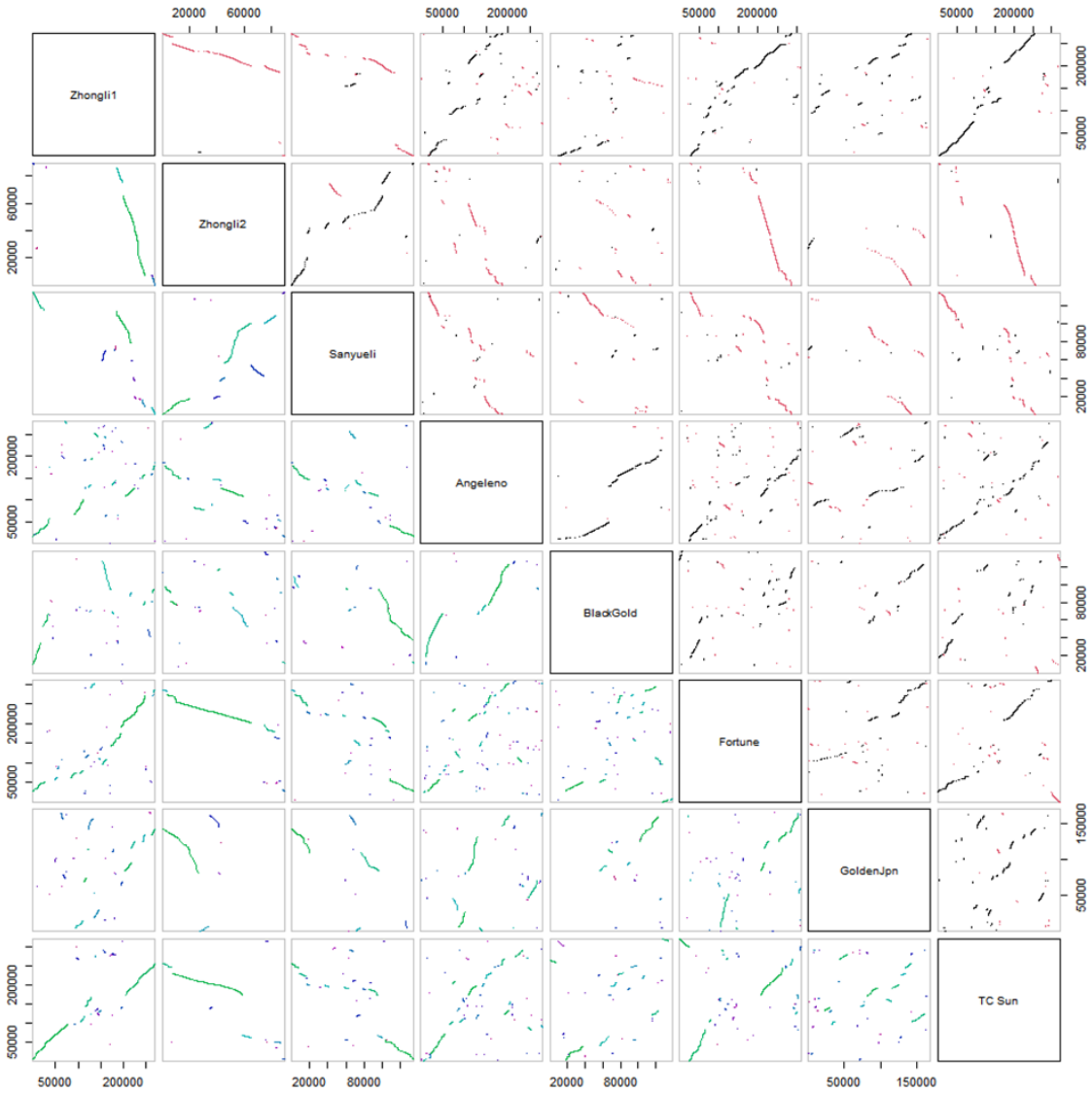
Shared SNP positions					
Shared SNP count	**Percentage from the total of each sample**				
	NB01	**NB02**	**NB03**	**NB04**	**NB05**
1		**0.10%**		**4.20%**	**0.20%**
2	**0.30%**	**0.20%**		**8.30%**	
2	**0.30%**	**0.20%**	**0.20%**	**8.30%**	**0.30%**
2				**8.30%**	**0.30%**
3	**0.40%**		**0.30%**		**0.50%**
4		**0.40%**	**0.40%**	**16.70%**	**0.60%**
4				**16.70%**	
9	**1.30%**	**0.90%**	**0.90%**	**37.50%**	
12	**1.80%**	**1.20%**			**1.90%**
43		**4.40%**			**6.90%**
50	**7.50%**	**5.10%**	**5.10%**		**8.00%**
62	**9.20%**		**6.40%**		
83			**8.50%**		**13.30%**
86		**8.80%**	**8.90%**		**13.80%**
87		**8.90%**	**9.00%**		
129	**19.20%**	**13.30%**	**13.30%**		
143	**21.30%**				
259	**38.60%**	**26.60%**			
289		**29.70%**			
336					**54.00%**
456			**47.00%**		

SNPs	Count	%
Total	3261	100
Shared	2033	62.34
Unshared	1228	37.66

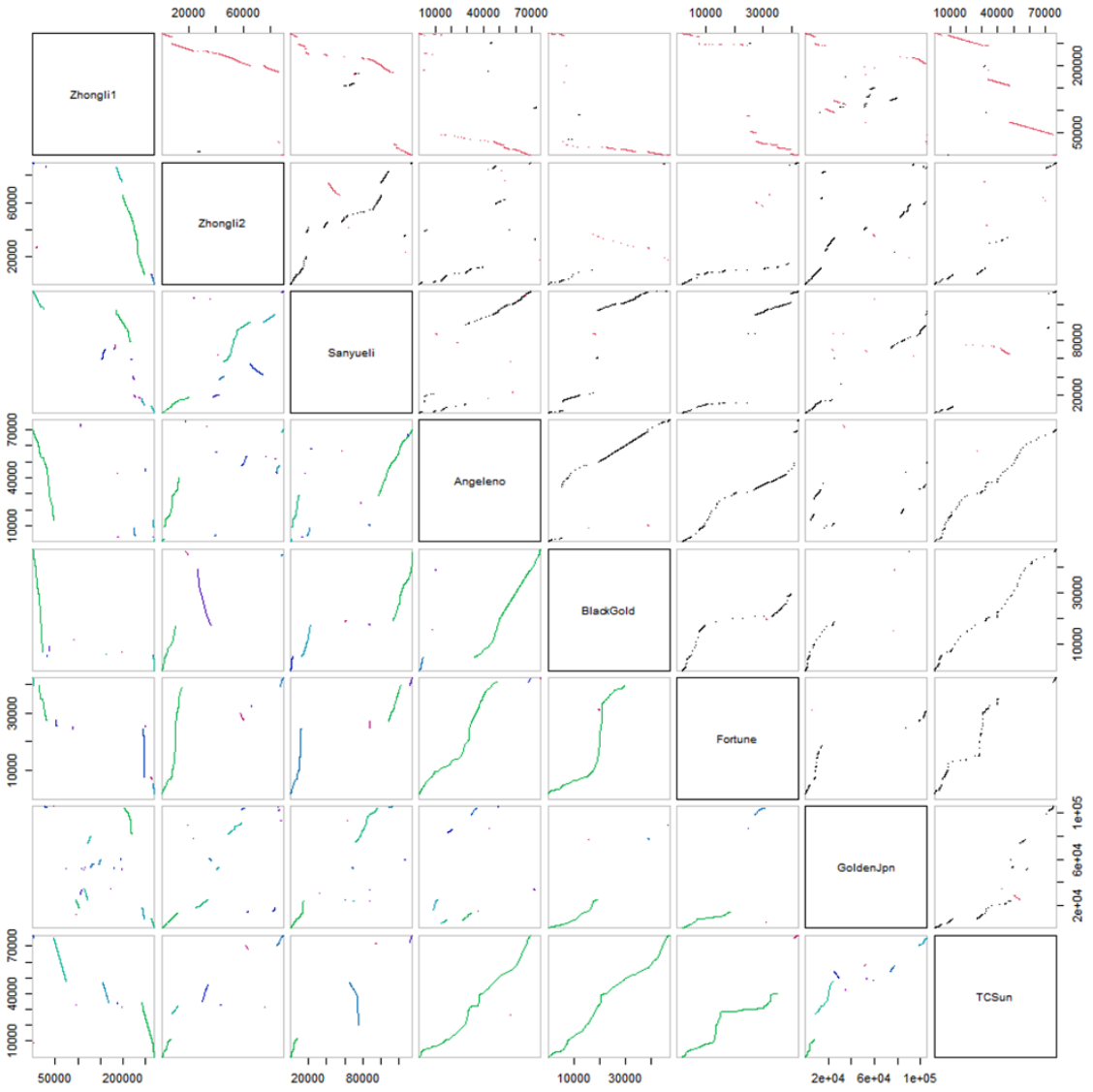
Additional File 3.6 (1/3): Contigs scaffolded with Sanyueli MYB10 region



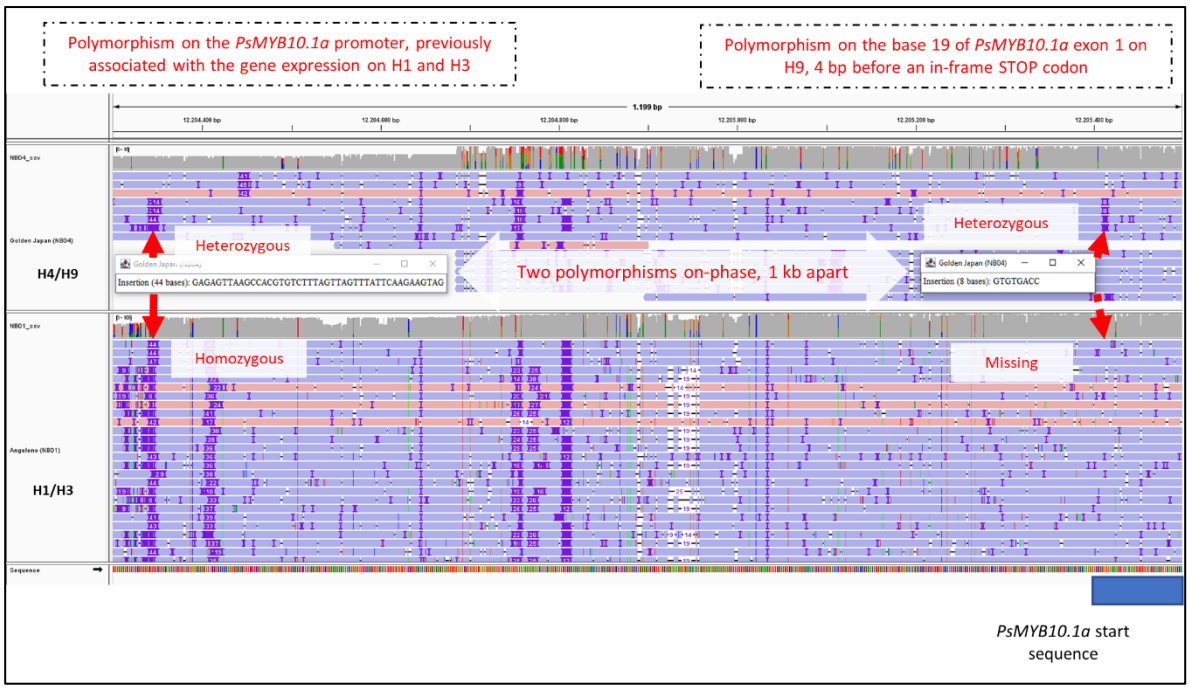
Additional File 3.6 (2/3): Contigs scaffolded with Zhongli-1 MYB10 region



Additional File 3.6 (3/3): Contigs scaffolded with Zhongli-2 MYB10 region



Additional File 3.6. Pair-wise visualization of the homologous hits (above diagonal) and homologous blocks (below) between each reference region and the *de novo* contigs scaffolded using the 'Sanyueli', Zhongli-1 or Zhongli-2 regions, represented in the same colors as in Figure 3.1a.



Additional File 3.7. Visualization of two phased variants 1 kb apart on H9. The 44 bp insertion is present in H1, H3 and H9 and was associated to the red skin color. The polymorphism is phased with an 8 bp insertion at the start of exon 1 of the *PsMYB10.1* gene, which explains its lack of function on H9. The reads from H4 do not show either of the two polymorphisms.

Additional File 3.8. Primer sequences and their annealing temperatures (Ta) used to PCR amplify the *MYB10* gene sequences.

Primer name	Sequence	Ta (ºC)	Description
M101_RT_F	TGGACACGAGACATTGCACG	57	*PsMYB10.1* amplification (Fiol et al., 2021)
M101_RT_R	CAATGGTCTTTTGACAGCCGC		
M102_f	CTGGCTGCAAGCATAC	57	*PsMYB10.2* amplification (Fiol et al., 2021)
M102_r	GTGGGACAAACACTCTC		
M103_f	ATAGGAACTAGCAGGCAC	57	*PsMYB10.3* amplification (Fiol et al., 2021)
M103_r	AGTTGCTAATAATTGCTACTAGG		

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw sequencing data presented in this study has been deposited at the EBI European Nucleotide Archive repository under the accession number PRJEB48338.

Competing interests

The authors declare that they have no competing interests.

Funding

AF is recipient of grant BES-2016-079060 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". FJR is recipient of grant PRE2019-087427 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future. This research was supported by project RTI2018-100795-B-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". We acknowledge support from the CERCA Programme ("Generalitat de Catalunya"), and the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533) and 2020-2023 (CEX2019-000902-S) both funded by MCIN/AEI /10.13039/501100011033.

Authors' contributions

The experiments were conceived and designed by AF and MJA. ELG provided support in the design of the experiment. Experiments were conducted by AF. Bioinformatics analysis were performed by AF and FJR. The paper was written by AF and MJA. All authors have critically revised the manuscript and approved the final document.

Acknowledgements

We thank Joan París, Ferran Contreras (ADV de Fruita del Baix Llobregat) and Jorge Naranjo (Tany Nature) for facilitating the plant material used in the experiment. We would like to thank David Chagne for thoughtful comments and helpful advice while reviewing our manuscript.

REFERENCES

1. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 2004;38:615-43.

2. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nat Rev Genet. 2009;10(10):725-32.

3. Gabur I, Chawla HS, Snowdon RJ, Parkin IA. Connecting genome structural variation with complex traits in crop plants. Theor Appl Genet. 2019;132(3):733-50.

4. Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then…and now: Stebbins revisited. Am J Bot. 2014;101(7):1057-78.

5. Giannuzzi G, D'Addabbo P, Gasparro M, Martinelli M, Carelli FN, Antonacci D, et al. Analysis of high-identity segmental duplications in the grapevine genome. BMC Genomics. 2011;12(1):1-14.

6. Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW. The roles of whole-genome and small-scale duplications in the functional specialization of Saccharomyces cerevisiae genes. Plos Genet. 2013;9(1):e1003176.

7. Zmienko A, Marszalek-Zenczak M, Wojciechowski P, Samelak-Czajka A, Luczak M, Kozlowski P, et al. AthCNV: A map of DNA copy number variations in the Arabidopsis genome. The Plant Cell. 2020;32(6):1797-819.

8. Vandepoele K, Simillion C, Van de Peer Y. Evidence that rice and other cereals are ancient aneuploids. The Plant Cell. 2003;15(9):2192-202.

9. Lye ZN, Purugganan MD. Copy number variation in domestication. Trends Plant Sci. 2019;24(4):352-65.

10. Xu C, Nadon BD, Kim KD, Jackson SA. Genetic and epigenetic divergence of duplicate genes in two legume species. Plant, cell & environment. 2018;41(9):2033-44.

11. Gu YQ, Crossman C, Kong X, Luo M, You FM, Coleman-Derr D, et al. Genomic organization of the complex α-gliadin gene loci in wheat. Theor Appl Genet. 2004;109(3):648-57.

12. Camerlengo F, Sestili F, Silvestri M, Colaprico G, Margiotta B, Ruggeri R, et al. Production and molecular characterization of bread wheat lines with reduced amount of α-type gliadins. BMC Plant Biol. 2017;17(1):1-11.

13. Ruiz M, Giraldo P, Royo C, Villegas D, Aranzana MJ, Carrillo JM. Diversity and Genetic Structure of a Collection of Spanish Durum Wheat Landraces. Crop Sci. 2012;52(5):2262-75.

14. Van Herpen TW, Goryunova SV, Van Der Schoot J, Mitreva M, Salentijn E, Vorst O, et al. Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics. 2006;7(1):1-13.

15. Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha S, Obermeier C, et al. Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. Plant Biotechnol J. 2021;19(2):240-50.

16. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell. 2020;182(1):145-61. e23.

17. Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. Genome Biol. 2021;22(1):1-19.

18. Tao Y, Zhao X, Mace E, Henry R, Jordan D. Exploring and exploiting pan-genomics for crop improvement. Mol Plant. 2019;12(2):156-69.

19. She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, et al. Shotgun sequence assembly and recent segmental duplications within the human genome. Cah Rev The. 2004;431(7011):927-30.

20. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21(1):1-16.

21. Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, et al. Pan-genome of wild and cultivated soybeans. Cell. 2020;182(1):162-76. e13.

22. Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nature Plants. 2020;6(1):34-45.

23. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat Biotechnol. 2020;38(4):433-8.

24. López-Girona E, Davy MW, Albert NW, Hilario E, Smart ME, Kirk C, et al. CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. Plant Methods. 2020;16(1):1-13.

25. Giesselmann P, Brändl B, Raimondeau E, Bowen R, Rohrandt C, Tandon R, et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. Nat Biotechnol. 2019;37(12):1478-81.

26. Fiol A, García-Gómez BE, Jurado-Ruiz F, Alexiou K, Howad W, Aranzana MJ. Characterization of Japanese Plum (*Prunus salicina*) *PsMYB10* Alleles Reveals Structural Variation and Polymorphisms Correlating With Fruit Skin Color. Front Plant Sci. 2021;12:1057.

27. Liu C, Feng C, Peng W, Hao J, Wang J, Pan J, et al. Chromosome-level draft genome of a diploid plum (*Prunus salicina*). GigaScience. 2020;9(12):giaa130.

28. Huang Z, Shen F, Chen Y, Cao K, Wang L. Chromosome-scale genome assembly and population genomics provide insights into the adaptation, domestication, and flavonoid metabolism of Chinese plum. Plant J. 2021.

29. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of gene duplication in plants. Plant Physiol. 2016;171(4):2294-316.

30. González VM, Aventín N, Centeno E, Puigdomènech P. High presence/absence gene variability in defense-related gene clusters of Cucumis melo. BMC Genomics. 2013;14(1):1-13.

31. Qiao X, Yin H, Li L, Wang R, Wu J, Wu J, et al. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (Pyrus bretschneideri). Front Plant Sci. 2018;9:161.

32. Gu C, Wang L, Wang W, Zhou H, Ma B, Zheng H, et al. Copy number variation of a gene cluster encoding endopolygalacturonase mediates flesh texture and stone adhesion in peach. J Exp Bot. 2016;67(6):1993-2005.

33. Barragan AC, Weigel D. Plant NLR diversity: the known unknowns of Pan-NLRomes. The Plant Cell. 2021;33(4):814-31.

34. Badet T, Croll D. The rise and fall of genes: origins and functions of plant pathogen pangenomes. Curr Opin Plant Biol. 2020;56:65-73.

35. Lallemand T, Leduc M, Landès C, Rizzon C, Lerat E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. Genes. 2020;11(9):1046.

36. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. Selective nanopore sequencing of human *BRCA1* by Cas9-assisted targeting of chromosome segments (CATCH). Nucleic Acids Res. 2018;46(14):e87-e.

37. Bruijnesteijn J, van der Wiel M, de Groot NG, Bontrop RE. Rapid characterization of complex genomic regions using Cas9 enrichment and Nanopore sequencing. bioRxiv. 2021.

38. McDonald TL, Zhou W, Castro CP, Mumm C, Switzenberg JA, Mills RE, et al. Cas9 targeted enrichment of mobile elements using nanopore sequencing. Nat Commun. 2021;12(1):1-13.

39. Kirov I, Merkulov P, Gvaramiya S, Komakhin R, Omarov M, Dudnikov M, et al. Illuminating the transposon insertion landscape in plants using Cas9-targeted Nanopore sequencing and a novel pipeline. bioRxiv. 2021.

40. Okie W, Ramming D. Plum breeding worldwide. Horttechnology. 1999;9(2):162-76.

41. Jung S, Jiwan D, Cho I, Lee T, Abbot AG, Sosinski B, et al. Synteny of *Prunus* and other model plant species. BMC Genomics. 2009;10:76.

42. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 2001;4(5):447-56.

43. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Cah Rev The. 2014;507(7490):62-7.

44. Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W, van Steensel B. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. Mol Cell. 2018;70(5):801-13. e6.

45. Li W, Freudenberg J. Mappability and read length. Frontiers in genetics. 2014;5:381.

46. Edge P, Bansal V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nat Commun. 2019;10(1):1-10.

47. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15(6):461-8.

48. Tham CY, Tirado-Magallanes R, Goh Y, Fullwood MJ, Koh BT, Wang W, et al. NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. Genome Biol. 2020;21(1):1-15.

49. Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. Genetics. 2000;154(1):459-73.

50. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. Genome Biol. 2002;3(2):1-9.

51. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019;20(1):1-10.

52. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6.

53. Naim F, Nakasugi K, Crowhurst RN, Hilario E, Zwart AB, Hellens RP, et al. Advanced engineering of lipid metabolism in Nicotiana benthamiana using a draft genome and the V2 viral silencing-suppressor protein. Plos One. 2012;7(12):e52717.

54. Hilario E. Plant nuclear genomic DNA preps 2018 [Available from: https://www.protocols.io/view/plant-nuclear-genomic-dna-preps-rncd5aw.

55. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. 1987.

56. Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. BMC Genomics. 2017;18(1):225.

57. Shirasawa K, Isuzugawa K, Ikenaga M, Saito Y, Yamamoto T, Hirakawa H, et al. The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding. DNA Res. 2017;24(5):499-508.

58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

59. Wright ES. Using DECIPHER v2. 0 to analyze big biological sequence data in R. R Journal. 2016;8(1).

60. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing.Vienna, Austria.: URL http://www.R-project.org/. 2020.

61. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics. 2019;35(3):523-5.

62. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.

63. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

64. Pohl A, Beato M. bwtool: a tool for bigWig files. Bioinformatics. 2014;30(11):1618-9.

65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

66. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8.

67. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Biotechfor. 2017;27(5):722-36.

68. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12(8):733-5.

69. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20(1):1-17.

Chapter 4

An LTR retrotransposon in the promoter of a *PsMYB10.2* gene associated with the regulation of fruit flesh color in Japanese plum

Arnau Fiol[1], Sergio García[1], Christian Dujak[1], Igor Pacheco[2], Rodrigo Infante[3], Maria José Aranzana[1,4]

[1]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193, Bellaterra, Barcelona, Spain

[2]Instituto de Nutrición y Tecnología de Alimentos (INTA), Universidad de Chile, El Líbano 5524, Santiago, Chile

[3]Departamento de Producción Agrícola, Facultad de Ciencias Agronómicas, Universidad de Chile, Santa Rosa 11315, Santiago, Chile

[4]IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Barcelona, Spain

This chapter has been submitted to Horticulture Research and is currently under review

Preprint available at bioRxiv repository (January 2022), DOI: 10.1101/2022.01.26.477575

ABSTRACT

Japanese plums exhibit wide diversity of fruit coloration. The red to black hues are caused by the accumulation of anthocyanins, while their absence results in yellow, orange or green fruits. In *Prunus*, MYB10 genes are determinants for anthocyanin accumulation. In peach, QTLs for red plant organ traits map in an LG3 region with three *MYB10* copies (*PpMYB10.1*, *PpMYB10.2* and *PpMYB10.3*). In Japanese plum the gene copy number in this region differs with respect to peach, with at least three copies of *PsMYB10.1*. Polymorphisms in one of these copies correlate with fruit skin color. The objective of this study was to determine a possible role of LG3-*PsMYB10* genes in the natural variability of the flesh color trait and to develop a molecular marker for marker-assisted selection (MAS). We explored LG3-*PsMYB10* variability, including the analysis of long-range sequences obtained in previous studies through CRISPR-Cas9 enrichment sequencing. We found that the *PsMYB10.2* gene was only expressed in red flesh fruits. Its role in promoting anthocyanin biosynthesis was validated by transient overexpression in Japanese plum fruits. The analysis of long-range sequences identified an LTR retrotransposon in the promoter of the expressed *PsMYB10.2* gene that explained the trait in 93.1% of the 145 individuals analyzed. We hypothesize that the LTR retrotransposon may promote the *PsMYB10.2* expression and activate the anthocyanin biosynthesis pathway. We provide a molecular marker for the red flesh trait which, together with that for skin color, will serve for the early selection of fruit color in breeding programs.

INTRODUCTION

Anthocyanins are plant pigments synthesized from the phenylpropanoid pathway and they confer red, purple or almost black coloration based on their structural conformation and accumulation levels[1]. These flavonoids protect plants against biotic and abiotic stress, and also enhance animal-mediated flower pollination and seed dispersion[2]. Some edible fruits are a good source of anthocyanins and their color is considered a fruit quality parameter, not only because of the visual attractiveness to the consumer, but also because anthocyanin ingestion has been correlated to several beneficial effects on human health[3]. The Rosaceae family includes species with edible fruits such as apples, pears, strawberries, peaches, cherries, apricots and plums, among others. Some of them have tremendous intraspecific fruit color variability, with anthocyanins conferring color to the whole fruit or single tissues, or not contributing to the final coloration at all.

The R2R3-MYB genes are among the largest transcription factor gene families in plants. They mostly regulate plant-specific processes, including anthocyanin biosynthesis by acting as transcriptional

activators as well as repressors[4,5]. In rosaceous species, *R2R3-MYB10* genes have been found to regulate anthocyanin levels in the fruit[6,7]. R2R3-MYB proteins regulate anthocyanin biosynthesis through interaction with a basic helix-loop-helix (bHLH) transcription factor and a WD-repeat protein, which results in the MBW complex that can bind to the promoters of several anthocyanin biosynthesis structural genes and enhance their expression[8-10]. Other R2R3-MYBs downregulate anthocyanin biosynthesis by diverging to other branches of the phenylpropanoid pathway or by repressing structural genes[11-13]. However, the *MYB10* genes remain the main factors leading to red-to-blue fruit coloration, and DNA polymorphisms with direct impact on its expression or protein function have been identified and correlated with fruit color variation in strawberry[14], apple[15,16], pear[17,18], peach[19,20], sweet cherry[21,22] and Japanese plum[23]. Transposons have also been reported to play a role in this variability[14,24]. Anthocyanin levels can also be modulated by ethylene production[25] and environmental factors such as light and temperature[26-28], adding an extra layer of complexity to their regulation.

Anthocyanins are synthetized in many plant tissues, constitutively or under certain conditions. In the *Prunus* genome, several QTLs for fruit color have been mapped in its eight linkage groups (LG), some of them in agreement with the position of a cluster of MYB10 genes in LG3[29-31]. In peach, this cluster contains three MYB10 genes (*MYB10.1*, *MYB10.2* and *MYB10.3*)[32], while an increased number of MYBs have been annotated in the syntenic region of the sweet cherry, apricot and Japanese plum genomes[33-35].

The multiple loci mapped in the *Prunus* genome for anthocyanin accumulation and the complexity of the phenotype highlight the complex regulation of synthesis of this pigment, in which several genes, differing in their spatial and temporal expression patterns, promote or repress the phenylpropanoid pathway. In peach, the recessive blood flesh (*bf*), mapped in LG4, promotes the accumulation of anthocyanins in the flesh after the pit hardening stage, i.e. long before fruit maturity[36], while the dominant blood flesh (*DBF*), mapped in LG5, controls the blood-flesh trait at the final stages of fruit maturation[37]. The flesh color around the stone (*Cs*) mapped in LG3[38] explains a localized pigment accumulation, unlike the highlighter (*h*) in the same LG that suppresses anthocyanin accumulation in the whole fruit[20,39]. Some of these traits, such as DBF, are regulated through the activation of MYB10 gene expression[40], while others are determined by polymorphisms in the MYB10 genes themselves[19,20,41].

Japanese plum (*Prunus salicina* L.) is one of the rosaceous crops with more fruit color variability. Fiol et al.[23] studied the allelic variability in the LG3-*PsMYB10* region. They found a triplication of the *PsMYB10.1* gene in some of the varieties and identified an allele in one of the gene copies highly associated with the skin anthocyanin color. In addition, they characterized the variability into haplotypes, providing an efficient molecular marker for marker-assisted selection (MAS) of skin color in breeding programs. However, they found no allele significantly associated with the red flesh color. To our knowledge, only two studies have found polymorphisms, mapped in LG6 and LG3, associated with the fruit flesh color in Japanese plum[42,43]. The number of individuals used in both studies was reduced, which limits the transfer of the markers to broader and more variable collections and motivates the search for markers more tightly linked to the trait.

The well-known role of *MYB10* genes in rosaceous fruit color variation, the weak association found between LG3 makers and flesh color in Japanese plum and the diverse number of alleles and gene copies identified by Fiol et al.[23] motivated a deeper exploration of the genetic variability of the LG3-*PsMYB10* region. For that, Fiol et al.[44] used CRISPR-Cas9 target enrichment sequencing in a pool of Japanese plum varieties, confirming the high level of variability in the MYB10 region between Japanese plum varieties compared to other *Prunus* genomes; and finding that the level of homology between Japanese plum varieties was comparable to that between *Prunus* species and their wild relatives. The CRISPR-Cas9 target sequencing approach yielded an excellent source of polymorphisms (SNPs and SVs) that can be used to find additional variants associated with fruit color, especially in the flesh.

The objective of this study was to explore the LG3-*PsMYB10* variability in the search for polymorphisms associated with the red flesh trait to further design a molecular marker for MAS in Japanese plum. By using the available MYB10 molecular marker provided in Fiol et al.[23], we found the association of one haplotype (H2) whose *PsMYB10.2* gene was expressed only in the mature flesh of red fruits. We validated the role of *PsMYB10.2* in the activation of anthocyanin biosynthesis by its transient expression in yellow fruits. After exploring the *de novo* sequences from the CRISPR-Cas9 experiment in Fiol et al.[44], we identified a copia-like LTR retrotransposon inserted in the promoter of the *PsMYB10.2* expressed genes only. The presence/absence of the LTR retrotransposon explained 93.1% of the red flesh color trait acquired during the last stage of ripening, becoming an efficient molecular marker for MAS in Japanese plum breeding programs.

RESULTS

The MYB10-H2 haplotype explains most of the flesh color variability

We studied the correlation between the LG3-MYB10 haplotypes described by Fiol et al.[23] and the fruit flesh color in a sample of 103 Japanese plum selections (**Figure 4.1**) (**Supplementary Data 4.1**). A χ^2 test found haplotype H2 highly correlated with red flesh (p-value=8.06×10^{-12}) (**Supplementary Table 4.1**). Forty-one produced fruits with red flesh, 33 of them (80.5%) with H2 in either homozygosis or heterozygosis. Haplotype H2 was present in seven non-red selections (11.29% in the non-red group), including one that was homozygous (C57, H2/H2). Five of them had an infrequent allele (a467) not observed in the H2 of red flesh selections (see Fiol et al.[23] for more details), indicating that there may be two different H2 haplotypes.

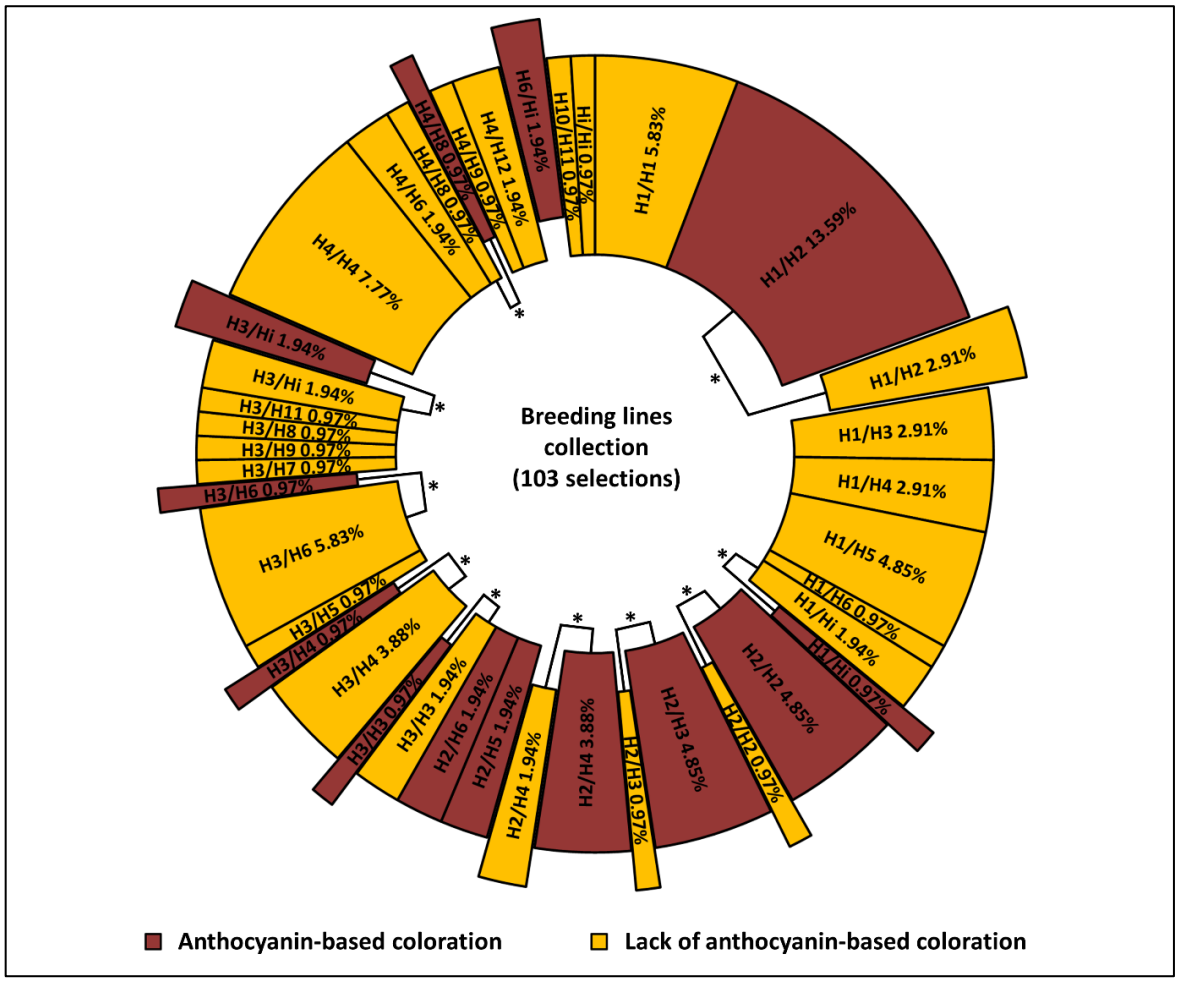


Figure 4.1. Graphic representation of haplotype combinations and flesh color in 103 Japanese plum selections. Portion size is proportional to the number of individuals. Ten of the haplotype combinations were found in both red and yellow fruits (*). Haplotype H2 was associated with the red flesh color in 84.47% of the tested genotypes; individuals with phenotype not explained by H2 (15.53%) are displaced out of the circle.

The *PsMYB10.2* gene is expressed in the red flesh

The haplotype H2 defined in Fiol et al.[23] was characterized by three *PsMYB10* alleles: a470 homologous to *PsMYB10.1*, a466 homologous to *PsMYB10.2*, and a492 homologous to *PsMYB10.3*. None of these alleles were exclusive to H2. We tested their expression in the flesh of mature fruits of five varieties with different flesh color and haplotype combinations. Only the *PsMYB10.2* allele (a466) was expressed in the flesh, and only in the red-fleshed fruits (**Figure 4.2**). The *PsMYB10.2* sequence amplified from mature fruit flesh cDNA had an approximate length of 800 bp, while the same primers amplified a fragment of around 2 kb in genomic DNA.

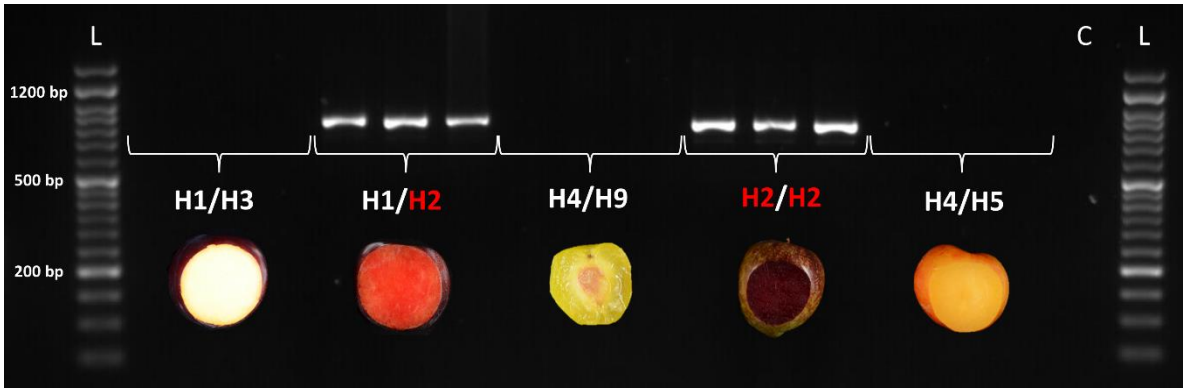


Figure 4.2. Expression of the *PsMYB10.2* gene in the mature flesh of five Japanese plum varieties (three fruits per variety). H1, H2, H3, H4, H5 and H9 are LG3-MYB10 haplotypes defined in Fiol et al.[23]. Well L: 50 bp ready-to-use DNA Ladder (GeneON); well C: no template control reaction.

The predicted PsMYB10.2 protein sequence has low variability

The RT-PCR bands, as well as the *PsMYB10.2* bands from genomic DNA of H1 to H4, were sequenced from C3, C31, C6 and C12, respectively. A complete match of the RT-PCR sequences with the exon sequences of the *PsMYB10.2-H2* gene was observed. The alignment of the predicted protein sequences of the *PsMYB10.2* alleles in the haplotypes H1 to H4 as well as that predicted from the NCBI *PsMYB10.2* GenBank sequence EU155161.1 revealed variation in some aminoacids (**Figure 4.3**). While the *PsMYB10.2*-H2 and H4 alleles coded for proteins with 243 amino acids; the *PsMYB10.2*-H1 and H3 alleles had three additional amino acids in their C-terminal sequence. All predicted proteins had homology higher than 95%, the proteins of H2 and H3 being the closest, with 99.18% homology, while H4 was slightly more distant to all the others, being 95.06% homologous with H1 and 95.88% with H2 and H3 (**Supplementary Data 4.2**). The amino acid sequences were also compared with the peach PpMYB10.2 translated protein (*Prupe.3G163000*) (**Figure 4.3**), where the functionality is significantly reduced due to the simultaneous occurrence of a lysine (K) and an

arginine (R) at positions 63 and 90, respectively[45]. In these positions, all the PsMYB10.2 proteins were monomorphic, with an arginine (R) in both positions. Motifs and other key amino acids reported by Stracke et al.[5], Lin-Wan et al.[7] and Yang et al.[46] were identified; none of the amionacid polymorphisms between the PsMYB10.2 proteins occurred at these positions.

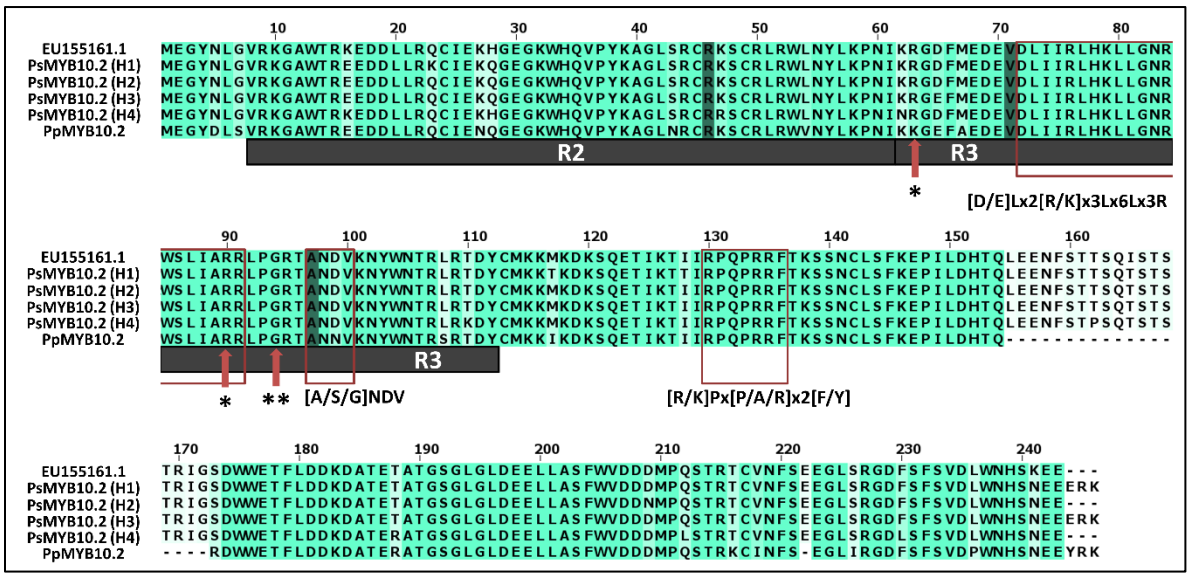


Figure 4.3. Alignment of the *in silico* proteins of the *PsMYB10.2* genes on haplotypes H1 to H4, the Japanese plum *PsMYB10.2* from the NCBI database (EU155161.1) and the peach *MYB10.2* (*PpMYB10.2*) from the Peach Genome v2[47]. The conserved R2 and R3 domains are shown. The motifs from orthologous MYBs (red rectangles) and the conserved residues from anthocyanin-promoting R2R3-MYBs from the Rosaceae family (shadowed residues) were monomorphic in the four haplotypes. Red arrows indicate key function residues reported by (*) Zhou et al.[45] and (**) Yang et al.[46].

The transient expression of *PsMYB10.2* promotes anthocyanin biosynthesis in the flesh of Japanese plum fruits

To determine whether the expression of *PsMYB10.2* promoted anthocyanin biosynthesis, its genomic sequence obtained from a red (H2/H2 genotype) and a yellow flesh (H4/H4) selections was transiently overexpressed in yellow flesh Japanese plum fruits. The fruits agroinfiltrated with either the *PsMYB10.2*-H2 or the *PsMYB10.2*-H4 alleles showed visible red stains on the yellow flesh in the region surrounding the injection (**Figure 4.4**), indicating that both coded proteins were functional. These red stains were not observed in other areas of the fruits, in fruits injected with the negative control vector (pBI121), and nor in noninjected fruits. The red stains in the flesh corresponded to a 20 to 103 fold increase of anthocyanin levels compared with the fruits injected with the control vector.

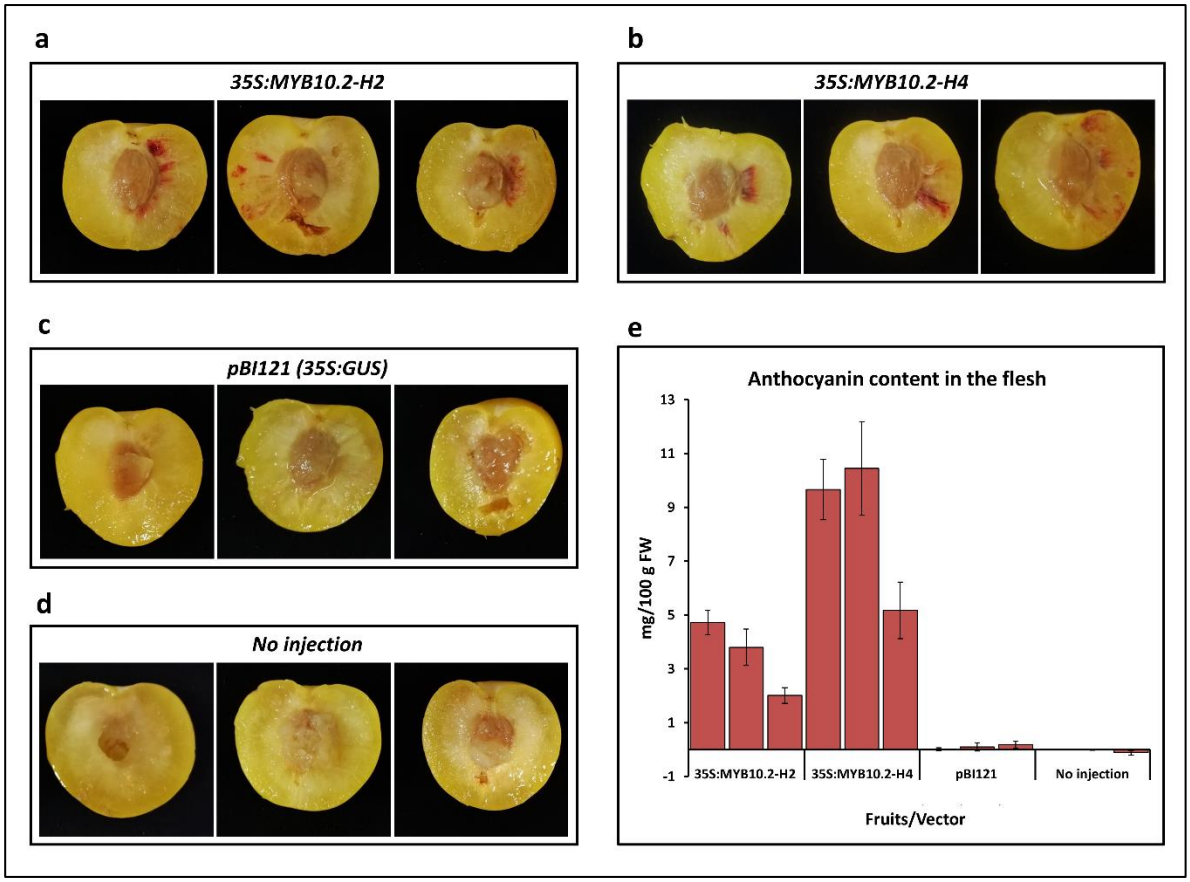


Figure 4.4. Transient overexpression of the *PsMYB10.2*-H2 and *PsMYB10.2*-H4 alleles in yellow flesh fruits and anthocyanin content. (**a**) fruits agroinfiltrated with the *PsMYB10.2*-H2 allele; (**b**) fruits agroinfiltrated with the *PsMYB10.2*-H4 allele. Negative controls: (**c**) fruits injected with the pBI121 vector and (**d**) noninjected fruits. (**e**) The anthocyanin content in the flesh of three fruits per treatment, error bars correspond to the standard deviation of each sample analyzed.

An LTR retrotransposon is inserted in the *PsMYB10.2-H2* promoter

To identify polymorphisms in the promoter that could be responsible for the lack of expression of the *PsMYB10.2* alleles in haplotypes other than H2, we did a BLAST search of the H1 and H2 allele sequences against the *de novo* assembly of the variety 'Black Gold' (H1/H2) obtained by Fiol et al. (submitted) through CRISPR-Cas9 enrichment sequencing. Only one contig for the *PsMYB10.2*-H2 allele was identified in 'Black Gold' (with 100% identity), with a sequence extending 3,142 bp upstream of the start codon. The BLAST search of this contig against all 'Black Gold' contigs identified an additional one 5,004 bp long. Both contigs overlapped 2,458 bp with 99.02% identity; the two sequences were merged to obtain a consensus sequence of 5,522 bp upstream of the *PsMYB10.2*-H2 gene start codon. For the *PsMYB10.2*-H1, we identified a contig with 99.7% identity in 'Black gold', which extended 9,339 bp upstream of the gene start codon. Alignment of the isolated H1 and

H2 contigs revealed a 2.8 kb insertion in H2, 633 bp upstream of the gene start codon (**Figure 4.5a**). The BLAST search of this 2.8 kb sequence against the LG3-*PsMYB10* region assemblies of the five varieties from Fiol et al. (submitted) ('Angeleno' H1/H3, 'Black Gold' H1/H2, 'Fortune' H3/H6, 'Golden Japan' H4/H9, and 'TC Sun' H4/H5) gave positive hits in 'Black Gold' only. The BLAST search of the same sequence against the 'Sanyueli' and 'Zhongli No.6' reference genomes[35,48] resulted in hits distributed along the eight chromosomes, none of them in the LG3-MYB10 region. Further search in the Rosaceae plants transposable elements database (RPTEdb) identified homology (E-value=0.0) with two *Prunus mume* Ty1-copia LTR retrotransposons: RLCopia_1_485_pmu and RLCopia_1_510_pmu. With the LTR_FINDER software, conserved sequences of long terminal repeat (LTR) retrotransposons were identified in the negative strand: the primer-binding site (PBS), the polypurine tract (PPT), the two LTR regions (762 bp and 772 bp long, with 97.4% of sequence identity), and the conserved 5' TG and 3' CA nucleotide pairs on them. The LTR retrotransposon insertion in the *PsMYB10.2* promoter was validated by long-range PCR. The amplification produced a band of 5 kb in a red flesh H2/H2 and of 2 kb in a yellow flesh H1/H1 selection (C31 and C27, respectively) (**Figure 4.5b**). The short band was also amplified in an H4/H4 selection (C12) and in C57, which was the only yellow flesh selection in the collection homozygous for the H2. Sequencing of the bands confirmed that the difference in size corresponded to the insertion of the LTR retrotransposon in H2 (**Figure 4.5c**). A characteristic of transposon insertions is the generation of a duplicated sequence at the genomic integration site, called Target Site Duplication (TSD). In the H2 sequence of the red flesh selection, a 5 bp TSD sequence was identified flanking the two LTR regions. This 5 bp sequence was found without duplication in the *PsMYB10.2* promoter of yellow flesh selections with H2 and H4, and was missing in the *PsMYB10.2*-H1 promoter.

The presence of the LTR retrotransposon correlates with red flesh color

The insertion of the LTR retrotransposon in the *PsMYB10.2* promoter was evaluated in the collection of 103 Japanese plum selections with a set of PCR primers, giving a 150 bp band when the LTR retrotransposon was present and a 300 bp band when it was absent (**Figure 4.5d**; **Supplementary Data 4.1**). Among the 41 selections with red flesh, the LTR retrotransposon was found in either homozygosis or heterozygosis in all the 32 genotyped with H2 plus one H4/H8 individual (C43). The eight red selections without the LTR retrotransposon acquired the red color at the very late stage of ripening. These included selection 98-99 and five seedlings of the same parent ('Sweet Pekeetah', with yellow flesh). In the yellow flesh group, only C4 (H1/H2) had the LTR retrotransposon insertion. RT-PCR analysis showed that the *PsMYB10.2* gene was expressed in the mature flesh of C4,

indicating that, in this selection, anthocyanin biosynthesis was repressed downstream in the pathway.

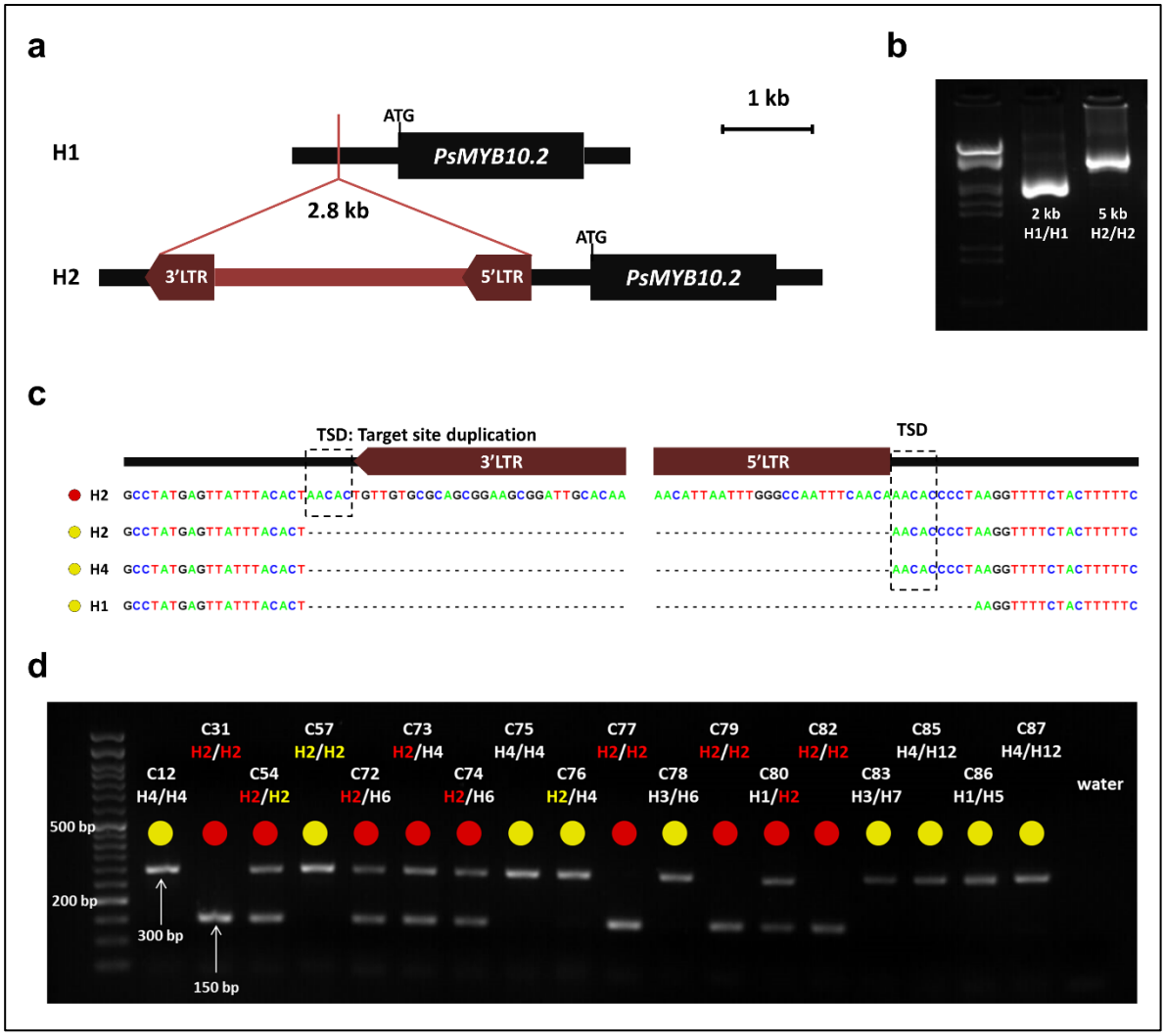


Figure 4.5. (**a**) Graphic representation of the 2.8 kb LTR retrotransposon insertion in the *PsMYB10.2* gene promoter in H2. (**b**) PCR amplification of the region in the *PsMYB10.2* promoter in C27 (H1/H1) and C31 (H2/H2) selections. The ladder in the first lane corresponds to Lambda DNA fully digested with HindIII and EcoRI enzymes. (**c**) Sequence of the retrotransposon insertion site in different haplotypes, the red and yellow circles correspond to red flesh and lack of red flesh, respectively. The TSD (target site duplication) produced by the mechanism of the retrotransposon insertion is indicated. (**d**) Agarose gel visualization of a representative set of Japanese plum selections, the 150 bp band corresponds to the presence of the LTR retrotransposon and the 300 bp band to its absence.

Our results indicate that there are two H2 haplotypes differing in the insertion of the LTR. While H2 selections lacking the LTR produce fruits without anthocyanins in their flesh, those with the LTR produce red-fleshed fruits. The LTR explained flesh color in 91.26% of the germplasm analyzed here.

In all cases, the LTR insertion was linked to the *PsMY10.2* gene expression, suggesting a role of this polymorphism in the promotion of the gene expression. The LTR was also inserted in C43 (H4/H8) suggesting that one of the two haplotypes has also two forms: one with the LTR retrotransposon inserted, producing red fruits, and one without.

The association of the LTR retrotransposon with red flesh color was further validated in a collection of 42 commercial varieties (eight with red flesh and 34 with yellow or orange flesh). The LTR was able to predict the presence or lack of anthocyanin flesh color in 97.6% of them. The prediction only failed for 'Black Beauty' (yellow flesh, H1/H2), which, as C4, had the LTR insertion in heterozygosis (**Supplementary Data 4.1**). Considering all the individuals of the study, the retrotransposon explained the flesh color of 93.1% of the 145 genotypes analyzed.

DISCUSSION

Activation of the *PsMYB10.2* gene promotes anthocyanin synthesis in the fruit flesh

In a previous study[23] we isolated *PsMYB10* alleles and found one highly associated with the red skin color and none with the flesh color. A PCR primer combination was then developed for marker-assisted selection (MAS). The MAS method relied on the organization of the alleles into haplotypes. Here we increased the sample size and looked for correlation between the haplotypes instead of the alleles, with the red/non-red flesh trait, identifying a strong correlation of the H2 haplotype with the red flesh phenotype (p-value=8.06×10^{-12}). This provides another example of the advantages of using haplotype data rather than only single polymorphisms[49-51]. However, the association did not explain the flesh color in 15.53% of the germplasm tested (consisting of advanced breeding lines and a few seedlings): these produced red flesh fruits without carrying the H2 haplotype or had the H2 haplotype but had yellow, orange or white flesh fruits. In Fiol et al.[23] we found that one allele, a467, was always present in H6 and in a few individuals with H2. This allele was in the H2 of most of the yellow flesh individuals. We hypothesized here that H2 could have two versions: one associated with the red flesh color, and one (hereby defined as H2*) non-associated.

The expression of the *PsMYB10.2* allele in H2 individuals with red flesh suggested that this gene had a functional role in the synthesis of anthocyanins in the flesh. This is in agreement with Fang et al.[52], who found that anthocyanin biosynthetic genes and the *PsMYB10.2* transcription factor were upregulated in a red flesh radiation-induced mutant of a yellow flesh Japanese plum variety. Here we show that the *MYB10.2* transcription factor is accountable for part of the natural variation of the

trait. Fang et al.[52] validated the gene function by transient overexpression in *Nicotiana*, co-infiltrating the *MYB10.2* and the *PsbHLH3* genes (required to activate anthocyanin synthesis through the R2R3-MYB-bHLH-WD-repeat complex, MBW). We performed the transient overexpression in yellow Japanese plum fruits, which did not require co-expression of the *bHLH3* gene with the *PsMYB10.2* to observe the pigment accumulation in the flesh. This indicates that, in the samples tested, the mechanism required for activation of the anthocyanin biosynthetic genes was functional and that the main determinant for anthocyanin accumulation in the flesh was the expression of the *PsMYB10.2* gene.

Functional amino acid modifications and key residues have been described in the *MYB10.2* protein of peach[5,7,45,46]. In Japanese plum, there were no relevant differences of expressed and non-expressed alleles of the *PsMYB10.2* gene at the translated amino acid level, suggesting that the differences in their functionality could be due to polymorphisms in the promoter or in the gene non-coding regions. Consistent with this hypothesis, the overexpression of *PsMYB10.2-H2* and *PsMYB10.2-H4* alleles in yellow flesh fruits initiated the synthesis of anthocyanins. Since the *PsMYB10.2*-H4 allele was not expressed in fruits and its protein was the most distant to that of the *PsMYB10.2*-H2 allele, the results hinted that all the *PsMYB10.2* alleles cloned may be *per se* functional and that the observed differences in flesh coloration were due to the activation of the *PsMYB10.2-H2* gene.

An LTR retrotransposon in the *PsMYB10.2* allele promoter region may activate the gene expression

The CRISPR-Cas9 enrichment strategy applied in the *PsMYB10* region[44] provided a reliable source of data to search for polymorphisms in the promoter of the *PsMYB10.2-H2* allele that could explain the gene expression. One of the advantages of long-read over short-read high-throughput sequencing technologies is the ability to detect structural variants[53], which served here to locate a 2.8 kb retrotransposon inserted in the *PsMYB10.2-H2* gene promoter. This large polymorphism involving a highly repeated sequence in the genome might have been difficult, or even impossible, to identify by other means.

Transposable elements can modulate gene expression by introducing transcription factor binding sites, transcription initiation sequences or *cis*-regulatory motifs that respond to environmental factors. Together with other genetic and epigenetic mechanisms they have been found to be

responsible for several cases of phenotypic variation in plants[54]. The retrotransposon identified here belongs to the long terminal repeat (LTR) family, which has already been reported to play a role in *MYB10* variability and the shift of anthocyanin fruit color in other rosaceous crops. The case reported in this study might be similar to that for red-skin apples, where an LTR retrotransposon in the *MdMYB10* promoter was associated and functionally validated with the increased gene expression[24]. In strawberry, an LTR retrotransposon disrupting the *MYB10* gene sequence was causative of the white-flesh phenotype, while in another allele a CACTA-like transposon inserted in the promoter was associated with the enhanced gene expression observed in the red fruits[14]. Outside the Rosaceae family, cases of transposons enhancing the expression of the orthologous R2R3-MYB gene have been reported in purple cauliflower, blood orange and pepper[55-57].

In the advanced breeding line collection studied here, we observed that the LTR retrotransposon was dominant for the red flesh color and improved the initial correlation of H2 (the LTR explained the phenotype of 91% of the lines versus the 84% explained by the H2 haplotype). The number of alleles shared between H2 and H2* and the TSD identified in H2 suggest that the latter originated after the insertion of the LTR retrotransposon in the original *PsMYB10.2*-H2* allele. Further functional studies are required to verify if the retrotransposon is the main determinant enhancing *PsMYB10.2*-H2 gene expression and to locate regulatory sequences within the transposable element or within the disrupted insertion site.

***PsMYB10* genes make a great contribution to the Japanese plum skin and flesh colors**

As for the advanced breeding lines, the fruit skin and flesh color in the panel of 42 commercial varieties were largely explained by the variability in the LG3-*MYB10* genes (only one outlier for the flesh color, none for the skin color markers), which confirmed the strong association of the region with the fruit color trait and the usefulness of the molecular markers designed. The MYB10 region contains several gene copies in tandem[44], a conformation that could have been generated by unequal crossing over by either homologous or non-homologous recombination[58].

After a gene duplication event, the subsequent fate of the copy can be i) conservation of the gene function (maintenance), ii) degeneration into a pseudogene or its removal (non-functionalization), iii) very rarely, the gain of a novel function (neo-functionalization) or, together with the original gene, iv) the accumulation of mutations which subdivide their functionality, complementing each other to achieve the function of the ancestral gene (sub-functionalization)[59,60]. This sub-

functionalization can be generated by changes in the gene regulatory sequences, altering their spatial and temporal expression[58]. This is likely the case in the *PsMYB10.1* and *PsMYB10.2* genes which greatly differ in their upstream regulatory sequences, causing one to be expressed in the skin and the other in the flesh, respectively, localizing anthocyanin synthesis in different fruit tissues. The haplotypes observed by Fiol et al.[23] (based on the segregation of the alleles of the *PsMYB10.1*, *PsMYB10.2* and *PsMYB10.3* gene copies) dominant for the red skin (H1, H3; both with the *PsMYB10.1*-356 allele) are recessive for yellow flesh, while the haplotype dominant for the red flesh color (H2; with an LTR in the promoter of the *PsMYB10.2* gene) is recessive for the yellow skin (**Figure 4.6**). As such, the presence of H2 with either H1 or H3 is required to confer anthocyanin coloration in both tissues. Okie (2008)[61] described that plums with yellow skin were transparent and if accompanied by red flesh they were mottled and bronze-like, matching the fruits in our study with background skin color (or mottled-type) and red flesh, which have H2 and any except H1 or H3.

Not all the fruit flesh color variability in Japanese plum is attributable to the identified MYB10 haplotypes

Fiol et al.[44] studied the synteny between the LG3-MYB10 region in the genome assemblies of *P. salicina* (two genomes), *P. avium* (two genomes), *P. persica* (two genomes), *P. mira*, *P. armeniaca* (five genomes), *P. sibirica* (two genomes) and *P. mandshurica,* demonstrating the high complexity of this region in Japanese plum, with homology levels between varieties comparable to those between *Prunus* species. This, together with the identification here of two H2 haplotypes (H2 and H2*) plus the insertion of the LTR in either H4 or H8 in C43 suggest that there may be additional polymorphisms in the LG3-*PsMYB10* region that could account for the variability in the flesh color not explained by the LTR.

In addition, similarly to what has been reported in peach, a single genomic region may not be responsible for all the red flesh color variability in Japanese plum. In the 103 selections, the only one non-anthocyanic with the LTR was C4 (H1/H2). When this line was crossed with another yellow selection (C14 [H3/H4]), some offspring (all with H2) had red flesh while, as with C4, some offspring with H2 were yellow[23]. We observed a similar segregation in the first year of fruiting of F1 progeny of C4 with C11 (yellow flesh, H3/H6) (data not shown), indicating that C4 can be a donor of the red flesh color trait through the LTR retrotransposon despite being yellow. Interestingly, the *PsMYB10.2* gene is expressed in the C4 flesh, which could be explained by the presence of a downstream

regulator of the anthocyanin biosynthesis pathway. The color prediction in the commercial cultivar collection failed only in 'Black Beauty', also H1/H2, which could be a similar case.

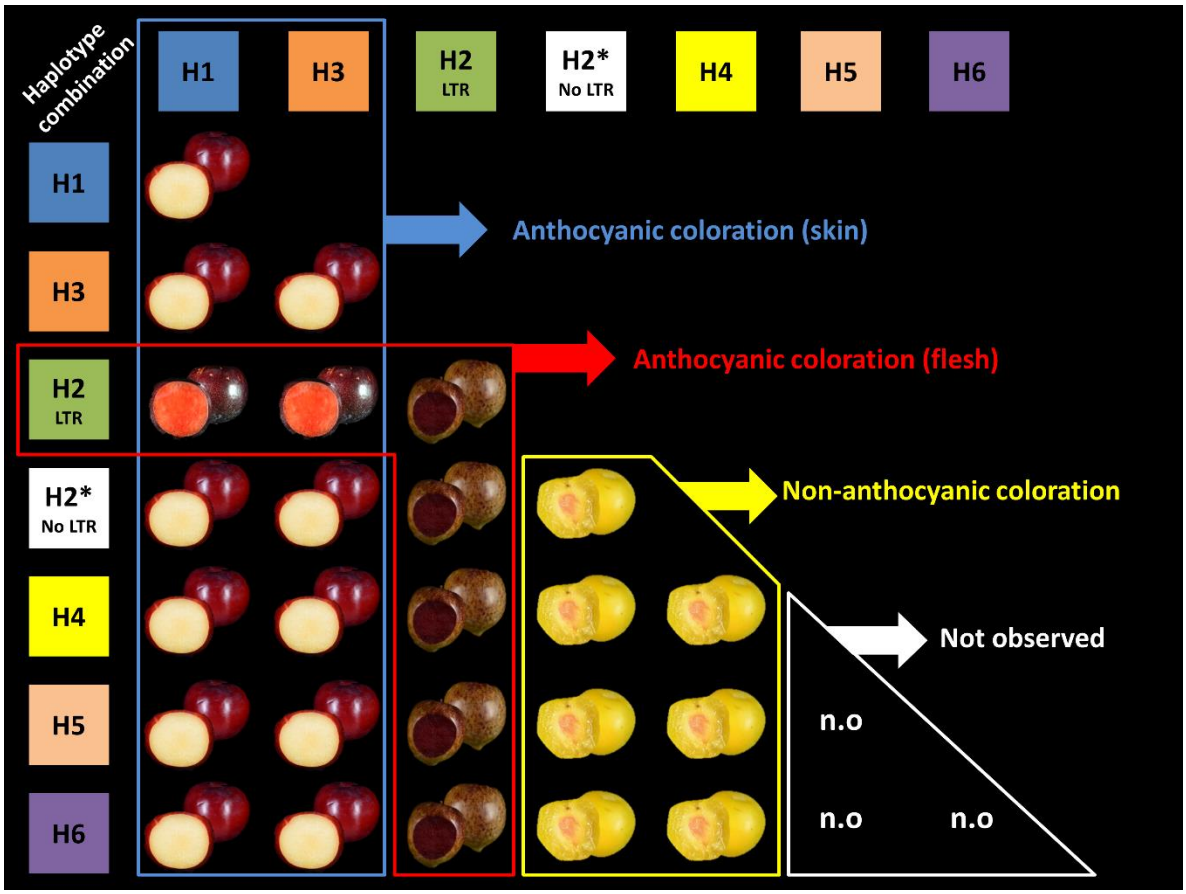


Figure 4.6. Japanese plum fruits showing the presence or absence of anthocyanin color in skin and flesh for every possible combination of the seven observed and most frequent MYB10 haplotypes found in the collections of Japanese plum selections and commercial varieties. The combinations H1/H2 and H2/H3 produce anthocyanins in the skin and flesh. We did not observe combinations with H5 and H6, but we expect they lack anthocyanins in both tissues.

A detailed observation of the trait showed that the phenotype in most of the red individuals without the LTR insertion was different to that observed in those with the insertion. While selections and varieties such as 'Black Gold' and 'Black Splendor' (H1/H2) accumulated the anthocyanin color progressively and homogeneously in the flesh during ripening (**Supplementary Figure 4.1a**), others without the insertion acquired color in the very late stages of fruit ripening and firstly around the skin. One example is selection 98-99 considered here as red flesh as in[42] although has been also described as yellow flesh[31,62,63] (**Supplementary Figure 4.1b**). Other five red-fleshed individuals without the LTR are offspring of 'Sweet Pekeetah', which does not have the retrotransposon and has yellow flesh: the late red flesh color might segregate through other genomic region/s. We noted

that, in some cases, the red pigmentation of the flesh was only localized near the red or black skin (**Supplementary Figure 4.1c**). While this could be due to the direct effect of another genomic region from that studied here, a possible contribution of the skin color to that of the flesh should not be discarded. Also not explained by the LTR retrotransposon insertion is the synthesis of anthocyanins induced by chilling stress (5°C). This phenotype has been reported in some yellow flesh commercial cultivars studied here: 'Friar' (H1/H5), 'Black Amber' (H1/H3) and 'Royal Diamond' (H1/H3)[64-66].

Additional QTL studies are required to identify more genomic regions involved in the regulation of flesh coloration. Due to appearing earlier in the maturation process, the homogenous red flesh color trait, if expressed, could have an epistatic effect over the presence of late coloration. As also noted by other authors, it will be interesting to study not only the presence of red flesh color but when[42] and where it appears, which would facilitate identifying other genomic regions controlling fruit coloration localized in specific fruit tissues and appearing in different ripening stages and/or environmental conditions.

CONCLUSION

In this study we found that most of the natural variability observed in the Japanese plum fruit flesh color is triggered by the expression of the *PsMYB10.2* gene. We have shown that when the *PsMYB10.2* gene is expressed in the flesh of mature fruits, the biosynthesis of anthocyanin is activated and the red pigment accumulates in the flesh. As has been demonstrated in other rosaceous species, the insertion of an LTR retrotransposon in the promoter of the gene may promote its expression. Apart from contributing to the knowledge of the genetic mechanisms behind the natural variation of Japanese plum fruit color, here we provide i) an improved protocol for the transient gene overexpression mediated by *Agrobacterium* in the flesh of Japanese plum fruits, that can be used for the validation of other candidate genes; ii) an efficient molecular marker for the flesh color which, together with the one previously developed by our group for skin color, can be used for marker-assisted breeding. However, as previously discussed the anthocyanin synthesis is complex, regulated by a network involving additional loci and environmental factors, and all the flesh color variability may not be determined by a single region. Therefore, further studies are needed to capture phenotypic variation caused by other loci.

MATERIALS AND METHODS

Plant material and nucleic acid isolation

A collection of 103 Japanese plum selections from two breeding programs was analyzed, consisting of 81 trees cultivated in Huelva (Spain) and 22 trees cultivated in Rinconada de Maipú (Chile) (**Supplementary Data 4.1**). Fruit color was visually phenotyped from several fruits at maturity for a minimum of two seasons. Flesh color was categorized as white, green, yellow, orange or red; skin color was classified as mottled, green, yellow, red, purple or black.

The fruits from five cultivars with different flesh color were obtained in triplicate at maturity from the local market. These were 'Angeleno' (pale yellow flesh), 'Golden Japan' (yellow flesh), 'Rose' (orange flesh), 'Black Gold' (red flesh) and a mottled-type plum (dark red flesh). The skin was removed and the flesh tissue was used for RNA and DNA extraction. The mature flesh of three fruits from the selection C4 was also used for RNA extraction. Transient gene expression assays were carried out using 30 yellow flesh fruits of 'Golden Japan' cultivar, harvested one week before full maturity stage.

To further validate the results, a DNA collection of 42 commercial varieties was analyzed (**Supplementary Data 4.1**). The DNA was extracted from the leaves in all cases except for 'Rose' and the mottled-type plum, where the DNA was extracted from the fruit. Fruit color descriptors for this collection were obtained from breeders' descriptions.

DNA was extracted following the CTAB procedure[67]. The RNA was extracted with the Maxwell RSC Plant RNA Kit (Promega) and was further DNase-treated with the TURBO DNA-free Kit (Invitrogen). The quality of the extracted nucleic acids was assessed using a Nanodrop ND-1000 Spectrophotometer and 0.8% agarose gels. All the studied material, including their fruit color phenotype and their genotyping results, can be found in **Supplementary Data 4.1**.

MYB10 genotyping and association with flesh color

All extracted DNAs were analyzed with the MYB10 markers (MYB10F2/MYB10NR2 and MYB10F2/MYB10NR4 primer pairs) to obtain their haplotype combination[23,68]. A χ^2 test was run on the collection of 103 Japanese plum selections to correlate the H1 to H6 haplotypes with the presence or absence of anthocyanin-based coloration of the fruit flesh. For this, it was assumed that

the red phenotype category indicated anthocyanin presence and the remaining categories (white, green, yellow and orange) its absence.

Expression of the *MYB10* genes in the fruit flesh

The RNA samples extracted from the fruit flesh were reverse transcribed using the PrimeScript RT-PCR Kit (TaKaRa) and the oligo(dt)20 primer. The expression of the MON reference gene[69] was used for quality control of the synthesized cDNAs. *PsMYB10.1*, *PsMYB10.2* and *PsMYB10.3* gene expression in the flesh of the red 'Black Gold' and the yellow 'Golden Japan' cultivars was evaluated by PCR. Primers and annealing temperatures used for each gene are described in **Supplementary Table 4.2**. The PCR reactions were carried out using 1.5 mM $MgCl_2$ and 1X NH_4 buffers for 1U of BioTaq polymerase (Bioline), 0.2 mM dNTP mix, 0.2 µM of each primer, 2 µl of cDNA and MilliQ water adjusted to a 10 µl reaction. PCR conditions were 1 minute at 94°C, 35 cycles of 94°C for 30 seconds, the annealing temperature of each primer pair for 20 seconds and 72°C for 2 minutes, followed by a final elongation step of 5 minutes at 72°C. Genomic DNA extracted from the fruit of each variety was added as amplification control. Amplification results were visualized in 2% agarose TAE gels after running 45 minutes at 120V. Amplification from the cDNAs only occurred with 'Black Gold' *PsMYB10.2*. The same reaction was repeated adding the cDNAs of 'Angeleno', 'Rose' and the mottled-type plums, and later repeated with the cDNA of C4. The *PsMYB10.2* gene product from 'Black Gold' was purified using PureIT ExoZAP Kit (Ampliqon) and sequenced by capillary electrophoresis with the same primers used to amplify the fragment.

***PsMYB10.2* sequencing on different haplotypes**

The *PsMYB10.2* alleles on haplotypes H1 to H4 were PCR-amplified from the genomic DNA of *PsMYB10* homozygous selections using the methodology described above. The selections were C3 (H1/H1), C31 (H2/H2), C6 (H3/H3) and C12 (H4/H4), being C31 the unique with red flesh color. The PCR purified products were ligated to the pCR™2.1-TOPO® vector (Invitrogen), following the manufacturer's instructions, and sequenced. The obtained sequences were aligned to the *Prunus salicina* MYB10 gene from the NCBI database (EU155161.1) and the *PsMYB10.2* cDNA sequence from 'Black Gold' using Sequencher 5.0 software (Gene Codes Corporation, Ann Arbor, MI, USA). The introns were removed, and the *in-silico* coded protein sequences obtained with the ExPASy Translate tool[70] were aligned to the peach *PpMYB10.2* (*Prupe.3G163000*) amino acid sequence using Clustal Omega[71] and visualized using Jalview v2[72].

Vector construction

The genomic 2 kb sequences from *PsMYB10.2*-H2 and *PsMYB10.2*-H4 were PCR-amplified from the C31 and C12 selections, respectively, using the Phusion Green High-Fidelity DNA Polymerase kit (Thermo Scientific) with primers that included restriction enzyme sites (**Supplementary Table 4.2**). The reactions were purified with the High Pure PCR product purification kit (Roche) and digested with SacI and XbaI enzymes in M buffer (TaKaRa). The gene sequence included a SacI restriction site in the second intron and, after 1 hour of XbaI digestion, the amplicon was partially digested with SacI by limiting the reaction to 2 minutes at 37°C. The pBI121 vector was fully digested with the same enzymes to remove the GUS gene, then the genic sequences were ligated next to the CaMV 35S promoter using T4 DNA Ligase (Thermo Scientific) to construct the 35S:MYB10.2-H2 and 35S:MYB10.2-H4 vectors. The ligation products were transformed into *Escherichia coli* JM109 competent cells and grown in LB media supplemented with 50 mg/l of kanamycin; screening was performed by direct colony PCR using M102_ex3_F and M13fwd primers (**Supplementary Table 4.2**). The vectors from positive colonies were extracted with the GenJet Plasmid Miniprep kit (Thermo Scientific) and sequenced before transformation into *Agrobacterium tumefaciens* EHA105 competent cells.

Transient gene expression in Japanese plum fruits

The *MYB10.2* function was validated by its transient overexpression in ripe fruits. The procedure was based on previous studies describing the genetic transformation in plum and the transient expression in fruits of closely related species[73-76], with some modifications to increase its efficiency. Specifically, *Agrobacterium tumefaciens* EHA105 cells transformed with the pBI121, the 35S:MYB10.2-H2 and the 35S:MYB10.2-H4 vectors were grown for 24 hours in liquid LB media with kanamycin and rifampicin (50 mg/l), then 1 ml of each culture was used to inoculate 15 ml of LB liquid media containing 20 µM of acetosyringone and 10 mM MES (2-(N-Morpholino)ethanesulfonic acid) at pH 5.6, and grown overnight. The final OD_{600} of the culture was quantified in a UV-2600 spectrophotometer (Shimadzu), then cells were pelleted by centrifugation and resuspended to a OD_{600} of 0.8 in injection buffer (10 mM magnesium chloride hexahydrate, 200 µM acetosyringone, 10 mM MES at pH 5.6) and kept in the dark at room temperature for 2-4 hours with occasional shaking. The 30 'Golden Japan' fruits used for the experiment were rinsed 2 minutes in a 30% bleach solution, washed with Tween-20 (0.05%) and completely dried. Ten fruits were injected with the *Agrobacterium* solution carrying the 35S:MYB10.2-H2 vector, ten more with the 35S:MYB10.2-H4, five with the pBI121, and the remaining five fruits were not injected. Using 1 ml hypodermic

syringes, a dose of 1 ml was evenly distributed on one side of each fruit by following the suture line, with a minimum of three injection points per fruit. Fruits were kept at room temperature in the dark and opened in halves after 13 days by cutting from the suture line. Anthocyanin was quantified using the differential pH spectrophotometry method as used in Fang et al.[77] in Japanese plum fruits, weighing approximately 0.1 g of flesh tissue extracted from the observable red flesh patches. The same approximate quantity was extracted from the injection sites in the case of the pBI121-treated fruits and randomly in the noninjected group of fruits, which served as negative controls.

Identification of polymorphisms in the *PsMYB10.2*-H2 promoter

To identify polymorphisms in the *PsMYB10.2* promoter, the CRISPR-Cas9 MYB10-enriched MinION sequencing data from Fiol et al.[44] was investigated. This data included sequences from the red flesh cultivar 'Black Gold' (H1/H2) and from the non-red flesh varieties 'Angeleno' (H1/H3), 'Fortune' (H3/H6), 'Golden Japan' (H4/H9) and 'TC Sun' (H4/H5). The blastn software[78] was run with the H1 and H2 *PsMYB10.2* input sequences in the 'Black Gold' assembly, selecting the contigs with the best score hits that extended upstream of the gene. The homologous *PsMYB10.2*-H2 contig ended prematurely and was extended by selecting the only contig that overlapped the upstream sequence after a blast analysis. By aligning and comparing the sequences in Sequencher 5.0, a large insertion on the H2 contig compared to the H1 was identified. This sequence was blast searched in the CRISPR-MinION sequences of the five varieties, and then on the *P. salicina* 'Sanyueli' v2.0[48] and *P. salicina* 'Zhongli No 6' v1.0[35] genomes using blastn software from GDR. The Blast search was repeated in the Rosaceae transposable elements database from the RPTEdb[79] and the sequence was further analyzed with LTR_FINDER software[80] to identify conserved structures from the LTR-type retrotransposon identified.

Validation of the insertion with long range PCR and marker design

To validate the large insertion, primers pM102F and pM102R (**Supplementary Table 4.2**) were designed and used in a long-range PCR reaction using the LongAmp® Taq DNA Polymerase (NEB) to amplify 40 ng of genomic DNA following the kit instructions. The red flesh selection used was C31 (H2/H2), the yellow flesh selections were C27 (H1/H1), C57 (H2/H2) and C12 (H4/H4). The PCR amplicons were run in a 2% agarose gel for 45 minutes at 120V. The PCR products were purified and sequenced by capillary electrophoresis as described above. The sequencing results were aligned with Sequencher 5.0 to validate the insertion site and used in Primer3 software[81] to design two primers flanking the 2.8 kb insertion on each side (LTR_F, LTR_R) and one within it (LTR_i)

(**Supplementary Table 4.2**). The PCR reaction for the LTR marker was carried out using 0.2 µM, 0.4 µM and 0.15 µM of LTR_F, LTR_i and LTR_R primers, respectively, mixed with 1.5 mM $MgCl_2$ and 1X NH_4 buffers, 1U of BioTaq polymerase, 0.2 mM dNTPs and MilliQ water to 10 µl. Thermocycler conditions were 94°C for 1 minute, 35 cycles of 94°C for 15 seconds, 56°C for 30 seconds and 72°C for 15 seconds, followed by one step at 72°C for 5 minutes. This reaction was used to genotype for the absence or presence of the retrotransposon in all the extracted DNA material, loading the 10 µl PCR results into TAE gels with 2% agarose which run for 40 minutes at 120V.

SUPPLEMENTARY MATERIAL

Supplementary Data 4.1 (1/3). Plant material (103 selections, 42 commercial varieties). Selections (C01 to C52).

Breeding lines collection (103 accessions)

Breedling lines	MYB10F2/MYB10NR2 Allele size (bp)												MYB10F2/MYB10NR4	Haplotype combination		Fruit color		(LTR_F/i/R)	
	350	356	443	454	462	466	470	473	477	492	495	other	a467			Skin	Flesh	150 bp	300 bp
C01	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	red	yellow		x
C02	0	0	1	0	1	1	1	0	0	1	0		1	H2	H4	green	yellow		x
C03	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	black	yellow		x
C04	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	yellow	x	x
C05	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	purple	yellow		x
C06	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	red	yellow		x
C07	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
C08	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
C09	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	purple	yellow		x
C10	1	1	0	1	0	1	1	0	0	1	0		1	H1	H2	black	white		x
C11	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red	yellow		x
C12	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C13	0	0	1	1	1	1	1	0	0	1	0	243	0	H4	*H9*	yellow	yellow		x
C14	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	purple	yellow		x
C15	0	0	1	0	1	1	1	0	1	0	0		1	H4	H6	yellow-green	yellow		x
C16	0	0	0	0	0	1	1	0	1	1	0		0	H2	H5	mottled	red	x	x
C17	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	black	yellow		x
C18	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C19	0	0	1	0	1	1	1	0	0	1	0		0	H2	H4	green	red	x	x
C20	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C21	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	purple	green		x
C22	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	black	yellow		x
C23	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
C24	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
C25	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	red-purple	red	x	x
C26	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
C27	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	red	yellow		x
C28	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	black	yellow		x
C29	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C30	0	1	0	0	1	1	0	1	0	1	1	243, 404	0	H3	*H11*	black	yellow		x
C31	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	mottled	red	x	
C32	0	0	0	0	0	1	1	0	1	1	0		0	H2	H5	mottled	red	x	x
C33	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	black	yellow		x
C34	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	red	yellow		x
C35	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	black	yellow		x
C36	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
C38	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C39	1	1	0	1	0	1	0	0	0	1	0		0	H1	H2	black	red	x	x
C40	1	1	1	1	1	1	1	0	0	1	0		0	H1	H4	purple	yellow		x
C41	0	0	1	0	1	1	1	0	0	1	0		0	H2	H4	mottled	red	x	x
C42	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
C43	0	1	1	0	1	1	1	1	0	0	0	440, 500	0	H4	*H8*	purple black	red	x	x
C45	0	1	0	0	1	1	1	1	1	0	1		0	H3	H5	purple	orange		x
C46	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
C47	1	1	0	1	1	1	0	0	1	1	0		1	H1	H6	red	yellow		x
C48	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	red	red	x	x
C49	1	1	0	1	0	1	1	0	0	1	0		1	H1	H2	black	yellow		x
C50	0	0	0	0	0	1	0	1	0	1	0	243, 404, 486	0	*H10*	*H11*	yellow	yellow		x
C51	0	1	0	1	1	1	0	1	0	1	1	243	0	H3	*H9*	purple	orange		x
C52	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	purple	yellow-green		x

Supplementary Data 4.1 (2/3). Plant material (103 selections, 42 commercial varieties). Selections (C53 to 98-99).

Breedling lines collection (103 accessions)

Breedling lines	MYB10F2/MYB10NR2 Allele size (bp)												MYB10F2/MYB10NR4	Haplotype combination		Fruit color		(LTR_F/i/R)		
	350	356	443	454	462	466	470	473	477	492	495	other	a467			Skin	Flesh	150 bp	300 bp	
C53	0	1	0	0	1	1	1	1	0	1	1		1	H2	H3	red-purple	orange		x	
C54	0	0	0	0	0	1	1	0	0	1	0		1	H2	H2	mottled	red	x	x	
C55	0	0	1	0	1	1	1	0	1	0	0		1	H4	H6	yellow-green	yellow		x	
C56	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x	
C57	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	green	yellow		x	
C58	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	red	yellow		x	
C60	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	purple	yellow		x	
C61	0	1	0	0	1	1	1	1	0	1	1		0	H2	H3	purple-red	red	x	x	
C62	0	1	0	0	1	1	1	1	0	1	1		0	H2	H3	red	red	x	x	
C63	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple black	red	x	x	
C64	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple red	red	x	x	
C66	0	1	0	0	1	1	1	1	0	1	1		0	H2	H3	purple red	red	x	x	
C67	0	0	1	0	1	1	1	0	0	1	0		0	H2	H4	mottled	red	x	x	
C68	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red	yellow		x	
C69	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple-red	red	x	x	
C70	1	1	1	1	1	1	1	0	0	1	0		0	H1	H4	purple	yellow-green		x	
C71	0	1	1	0	1	1	1	1	0	0	0	440, 500	0	H4	*H8*	purple	yellow		x	
C72	0	0	0	0	1	1	1	0	1	1	0		1	H2	H6	mottled	red	x	x	
C73	0	0	1	0	1	1	1	0	0	1	0		0	H2	H4	mottled	red	x	x	
C74	0	0	0	0	1	1	1	0	1	1	0		1	H2	H6	red	red	x	x	
C75	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x	
C76	0	0	1	0	1	1	1	0	0	1	0		1	H2	H4	green	yellow		x	
C77	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	mottled	red	x		
C78	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red	yellow		x	
C79	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	mottled	red	x		
C80	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x	
C82	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	mottled	red	x		
C83	0	1	0	0	1	1	1	1	0	0	1	500	0	H3	*H7*	purple	orange		x	
C85	0	0	1	0	1	1	1	0	0	0	0	500	0	H4	*H12*	yellow	yellow		x	
C86	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	black	yellow		x	
C87	0	0	1	0	1	1	1	0	0	0	0	500	0	H4	*H12*	yellow	yellow		x	
2B19	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	purple	red		x	
H9BPO75	0	0	0	0	1	1	0	1	1	0	0	243, 440	1	H6	Hi	purple	red		x	Sweet Pekeetah op.
10A	0	1	0	0	1	1	0	1	0	0	1	440	0	H3	Hi	red	yellow		x	
H10AP011	0	1	0	0	1	1	0	1	0	1	1	404, 480	0	H3	Hi	purple	red		x	Sweet Pekeetah op.
H10AP072	1	1	0	1	0	1	0	1	0	1	0	440, 500	0	H1	Hi	black	red		x	Sensation x Blue gusto
H10AP085	0	1	0	0	1	1	1	1	0	1	1		0	H2	H3	purple	red	x	x	Black Splendor x Sapphire
H11AP020	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	purple	red		x	Sweet Pekeetah op.
H11AP085	0	0	0	0	1	1	0	1	1	1	0	404, 480	1	H6	Hi	red	red		x	Sweet Pekeetah op.
H11BP093	0	1	0	0	1	1	0	1	0	0	1	440	0	H3	Hi	red	red		x	Sweet Pekeetah op.
H11AP108	0	1	0	0	1	1	0	1	0	1	1	404, 480	0	H3	Hi	red	yellow		x	Sweet Pekeetah op.
30A	1	1	1	1	1	1	1	0	0	1	0		0	H1	H4	black	yellow		x	
31A	0	1	1	0	1	1	1	1	0	0	0	440	0	Hi	Hi	red	yellow		x	
32-NN	0	1	0	0	1	1	0	1	0	0	1	440, 500	0	H3	Hi8	red	yellow		x	
33-23-85	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	purple	yellow		x	
37-118	1	1	0	1	1	1	1	0	1	1	0		0	H1	Hi	red	yellow		x	
43A54	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	red/yellow	yellow		x	
43A60	0	1	0	0	1	1	1	1	0	1	1		0	H2	H3	black	red	x	x	
549-60	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	red	yellow		x	
803-97	1	1	0	1	1	1	1	0	1	1	0		0	H1	Hi	red	yellow		x	
804	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red/green	green/yellow		x	
807	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red/green	yellow		x	
98-99	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	black	red		x	

Supplementary Data 4.1 (3/3). Plant material (103 selections, 42 commercial varieties). Commercial varieties.

Commercial varieties collection (42 cultivars)

Cultivar	MYB10F2/MYB10NR2 Allele size (bp)												MYB10F2/MYB10NR4	Haplotype (H) combination		Fruit color		(LTR_F/i/R)	
	350	356	443	454	462	466	470	473	477	492	495	others	a467			Skin	Flesh	150 bp	300 bp
Angeleno	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	black	pale yellow		x
Autumn Giant	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	red	yellow		x
Black Amber	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	black	amber		x
Black Beauty	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	yellow	x	x
Black Diamond	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
Black Gold	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
Black Splendor	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
Black Star	1	1	0	1	0	1	1	0	0	1	0		1	H1	H2	black	yellow		x
Canela	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	red	yellow		x
Crimson Glo	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
Early Queen	1	1	0	1	0	1	1	0	0	1	0		1	H1	H2	purple	yellow		x
Flavorosa	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
Fortune	0	1	0	0	1	1	0	1	1	0	1		1	H3	H6	red	yellow		x
Friar	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	black	yellow		x
Gaia	0	1	0	0	1	1	1	1	1	0	1	404	0	H5	*H13*	red	yellow		x
Golden Globe	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
Golden Japan	0	0	1	1	1	1	1	0	0	1	0	243	0	H4	*H9*	yellow	yellow		x
Golden Kiss	0	0	1	0	1	1	1	0	0	1	0		1	H2	H4	yellow	yellow		x
John W	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	purple	yellow		x
Kelsey	0	0	1	0	1	1	1	0	1	0	0		0	H4	H5	yellow	yellow		x
Laetitia	0	1	0	0	1	1	1	1	0	0	1	500	0	H3	*H7*	red	yellow		x
Larry Ann	0	1	1	0	1	1	1	1	0	0	1		0	H3	H4	purple	yellow		x
Mottled-type	0	0	0	0	0	1	1	0	0	1	0		0	H2	H2	mottled	dark red	x	
Owen-T	1	1	0	1	0	1	0	0	0	1	0		0	H1	H1	black/purple	yellow		x
Pioneer	0	1	1	0	1	1	1	1	0	0	0	440, 500	0	H4	*H8*	red	yellow		x
Primetime	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	purple	red	x	x
Red Beauty	1	1	0	1	0	1	1	0	1	1	0		0	H1	H5	red	yellow		x
Rose	0	0	1	0	1	1	1	0	1	0	0		0	H4	H5	yellow w/blush	orange		x
Royal Diamond	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	black	yellow		x
Royal Garnet	1	1	0	1	0	1	1	0	0	1	0		1	H1	H2	black	yellow		x
RR1	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	purple	yellow		x
Santa Rosa	0	1	0	0	1	1	1	1	0	1	1		1	H2	H3	red	amber		x
Sapphire	0	1	0	0	1	1	1	1	1	0	1		0	H3	H5	red	amber		x
Sensation	1	1	0	1	1	1	0	0	1	1	0		1	H1	H6	black	yellow		x
September King	0	1	0	0	1	1	0	1	0	0	1		0	H3	H3	purple	yellow		x
September Queen	1	1	0	1	1	1	0	1	0	1	1		0	H1	H3	purple	yellow		x
Showtime	0	1	0	0	1	1	1	1	1	0	1		0	H3	H5	red	amber		x
Songría 15	1	1	0	1	0	1	1	0	0	1	0		0	H1	H2	black	red	x	x
Sundew	0	0	1	0	1	1	1	0	0	0	0		0	H4	H4	yellow	yellow		x
SunGold	0	0	1	0	1	1	1	0	1	0	0		1	H4	H6	yellow	yellow		x
Sunkiss	0	0	1	0	1	1	1	0	0	0	0	500	0	H4	*H12*	yellow	yellow		x
TC Sun	0	0	1	0	1	1	1	0	1	0	0		0	H4	H5	yellow	yellow		x

Supplementary Data 4.2. Percent identity matrix (%) for the six MYB10.2 sequences compared.

	EU155161.1	PsMYB10.2-H1	PsMYB10.2-H2	PsMYB10.2-H3	PsMYB10.2-H4	PpMYB10.2
EU155161.1	100	97.12	97.94	97.94	96.3	91.07
PsMYB10.2-H1	97.12	100	98.35	98.37	95.06	91.19
PsMYB10.2-H2	97.94	98.35	100	99.18	95.88	91.96
PsMYB10.2-H3	97.94	98.37	99.18	100	95.88	92.51
PsMYB10.2-H4	96.3	95.06	95.88	95.88	100	89.73
PpMYB10.2	91.07	91.19	91.96	92.51	89.73	100

Supplementary Table 4.1. The results for the $\chi^2_{(1df)}$ test in the panel of 103 selections in the breeding collection considering the six most frequent haplotypes (H1 to H6) and the flesh color. The presence of H2 was statistically associated with the presence of red color.

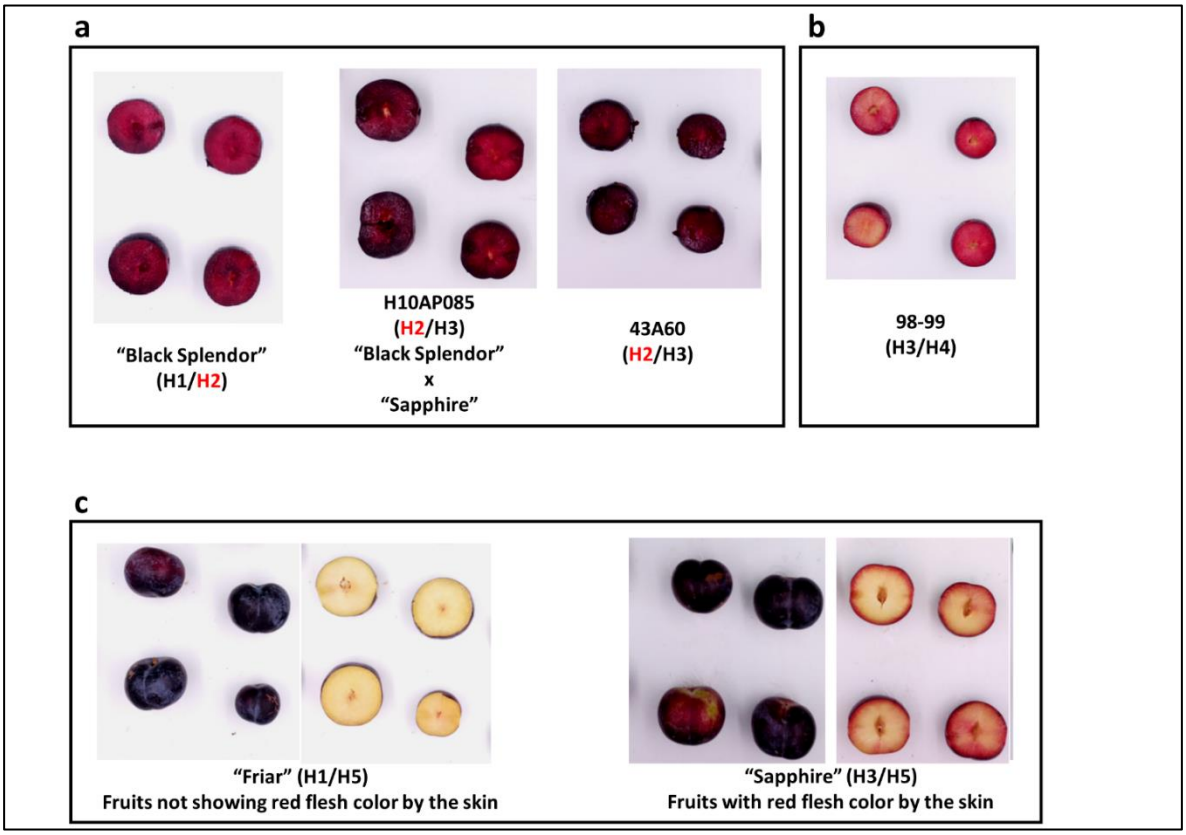
Haplotype	Frequency (%)	Chi-square p-value
H1	21.36	9.58×10^{-1}
H2	21.84	8.06×10^{-12}
H3	17.48	1.76×10^{-1}
H4	17.96	1.31×10^{-2}
H5	3.88	3.73×10^{-1}
H6	6.80	7.37×10^{-1}

Supplementary Table 4.2 (1/2). Primer sequences and their temperatures of annealing (Ta) used in the PCR reactions.

Primer name	Sequence	Ta (ºC)	Description
MYB10F2	GTGTGAGAAAAGGAGCTT	55	MYB10 marker (Fiol et al., 2021)
MYB10NR2	GATATTTGGCTTCAAATAGTTC		
MYB10NR4	TTCCTGCACCTGTTCAAC		
M101_RT_F	TGGACACGAGACATTGCACG	57	*PsMYB10.1* amplification (Fiol et al., 2021)
M101_RT_R	CAATGGTCTTTTGACAGCCGC		
M102_f	CTGGCTGCAAGCATAC	57	*PsMYB10.2* amplification (Fiol et al., 2021)
M102_r	GTGGGACAAACACTCTC		
M103_f	ATAGGAACTAGCAGGCAC	57	*PsMYB10.3* amplification (Fiol et al., 2021)
M103_r	AGTTGCTAATAATTGCTACTAGG		

Supplementary Table 4.2 (2/2). Primer sequences and their temperatures of annealing (Ta) used in the PCR reactions.

Primer name	Sequence	Ta (ºC)	Description
teM102_F2	ACGCATCTAGAACGCTGCACAAAAGAAAACA	62	Primers including XbaI and SacI restriction sites for *PsMYB10.2* amplification and insertion into pBI121 vector
teM102_R2	TCAGTGAGCTCGTCCCTTAACTTTTCAATGTGGG		
M102_ex3_F	ACAGGTTCTGGTCTTGGGTT	55	Screening of positive colonies carrying the *PsMYB10.2* insert
M13fwd	GTAAAACGACGGCCAGT		
pM102F	TGTTAGGCTGAAATGCAGGA	56	Amplification of the *PsMYB10.2* promoter using a long-range PCR
pM102R	GTATGCTTGCAGCCAG		
LTR_F	(*)TGAGTTAGGTTGCCTATGAGT	57	Molecular marker of the LTR retrotransposon. (*)can be fluorescently labeled for capillary electrophoresis
LTR_i	CGGCGCAGTATAAATCCTTG		
LTR_R	TTTTCGTTGATTGTTTGTCCAA		



Supplementary Figure 4.1. (**a**) Fruits of a commercial variety and two selections showing the red flesh color associated with the *PsMYB10.2*-H2 gene. Selection H10AP085 inherited the LTR retrotransposon and its color from 'Black Splendor'. (**b**) The fruits from selection 98-99, in which the red flesh appears very late in the maturation stage and is visually different from the former phenotype. (**c**) Two cultivars described with yellow flesh. Fruits from 'Friar' do not have red flesh around the skin compared to 'Sapphire' fruits.

ACKNOWLEDGMENTS AND FUNDING

AF is recipient of grant BES-2016-079060 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". CD was supported by "DON CARLOS ANTONIO LOPEZ" Abroad Postgraduate Scholarship Program, BECAL-Paraguay. This research was supported by project RTI2018-100795-B-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". We acknowledge support from the CERCA Programme ("Generalitat de Catalunya"), and the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533) and 2020-2023 (CEX2019-000902-S) both funded by MCIN/AEI /10.13039/501100011033.

We thank PLANASA company and their breeders Mario Ortiz and Antonio García for providing plant material and phenotype information. We also wish to thank Jorge Naranjo (Tany Nature), Alfonso Guevara (IMIDA) and David Ruiz (CEBAS-CSIC) for providing plant material from commercial varieties, and L. Maria Lois for providing the pBI121 vector.

AUTHOR CONTRIBUTIONS

The experiments were conceived and designed by AF and MJA; AF and SG conducted the gene transient expression experiments; AF and CD performed the gene expression analysis; IP and RI provided plant materials and contributed to phenotypic data collection; AF and MJ wrote the manuscript. All authors have critically revised the manuscript and approved the final document.

COMPETING INTERESTS

The authors declare no competing interests.

MATERIALS & DATA AVAILABILITY

Correspondence requesting materials or data should be addressed to M.J.A (mariajose.aranzana@irta.es)

REFERENCES

1. Saigo, T., Wang, T., Watanabe, M. & Tohge, T. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. *Curr. Opin. Plant Biol.* **55**, 93-99 (2020).

2. Winkel-Shirley, B. Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **5**, 218-223 (2002).

3. Khoo, H. E., Azlan, A., Tang, S. T. & Lim, S. M. Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food & nutrition research* **61**, 1361779, doi:10.1080/16546628.2017.1361779 (2017).

4. Albert, N. W. *et al.* A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *The Plant Cell* **26**, 962-980 (2014).

5. Stracke, R., Werber, M. & Weisshaar, B. The R2R3-MYB gene family in Arabidopsis thaliana. *Curr. Opin. Plant Biol.* **4**, 447-456 (2001).

6. Espley, R. V. *et al.* Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* **49**, 414-427, doi:10.1111/j.1365-313x.2006.02964.x (2007).

7. Lin-Wang, K. *et al.* An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol.* **10**, 50-50, doi:10.1186/1471-2229-10-50 (2010).

8. Koes, R., Verweij, W. & Quattrocchio, F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **10**, 236-242 (2005).

9. Heppel, S. C. *et al.* Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Mol. Biol.* **82**, 457-471 (2013).

10. Lai, Y., Li, H. & Yamagishi, M. A review of target gene specificity of flavonoid R2R3-MYB transcription factors and a discussion of factors contributing to the target gene selectivity. *Frontiers in biology* **8**, 577-598 (2013).

11. Ravaglia, D. *et al.* Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biol.* **13**, 68-68, doi:10.1186/1471-2229-13-68 (2013).

12. Zhou, H. *et al.* Activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation. *New Phytol.* **221**, 1919-1934 (2019).

13. Yan, H. *et al.* MYB-mediated regulation of anthocyanin biosynthesis. *Int. J. Mol. Sci.* **22**, 3103 (2021).

14. Castillejo, C. *et al.* Allelic variation of *MYB10* is the major force controlling natural variation in skin and flesh color in strawberry (*Fragaria spp*.) fruit. *The Plant Cell* **32**, 3723-3749 (2020).

15. Espley, R. V. *et al.* Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *The Plant Cell* **21**, 168-183 (2009).

16. Telias, A. *et al.* Apple skin patterning is associated with differential expression of *MYB10*. *BMC Plant Biol.* **11**, 93-93, doi:10.1186/1471-2229-11-93 (2011).

17. Feng, S., Wang, Y., Yang, S., Xu, Y. & Chen, X. Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor *PyMYB10*. *Planta* **232**, 245-255 (2010).

18. Pierantoni, L. *et al.* Mapping of an anthocyanin-regulating MYB transcription factor and its expression in red and green pear, *Pyrus communis*. *Plant Physiol. Bioch.* **48**, 1020-1026 (2010).

19. Tuan, P. A. *et al.* The crucial role of *PpMYB10.1* in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. *BMC Plant Biol.* **15**, 280, doi:10.1186/s12870-015-0664-5 (2015).

20. Bretó, M., Cantín, C., Iglesias, I., Arús, P. & Eduardo, I. Mapping a major gene for red skin color suppression (highlighter) in peach. *Euphytica* **213**, 14 (2017).

21. Starkevič, P. *et al.* Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (*Prunus avium* L.) *MYB10* and *bHLH* Genes. *Plos One* **10**, e0126991, doi:10.1371/journal.pone.0126991 (2015).

22. Jin, W. *et al.* The R2R3 MYB transcription factor *PavMYB10.1* involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). *Plant Biotechnol. J.* **14**, 2120-2133 (2016).

23. Fiol, A. *et al.* Characterization of Japanese Plum (*Prunus salicina*) *PsMYB10* alleles reveals structural variation and polymorphisms correlating with fruit skin color. *Front. Plant. Sci.* **12**, 1057 (2021).

24. Zhang, L. *et al.* A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* **10**, 1-13 (2019).

25. Cheng, Y., Liu, L., Yuan, C. & Guan, J. Molecular characterization of ethylene-regulated anthocyanin biosynthesis in plums during fruit ripening. *Plant Mol. Biol. Rep.* **34**, 777-785 (2016).

26. Takos, A. M. *et al.* Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* **142**, 1216-1232 (2006).

27. Niu, J. *et al.* Anthocyanin concentration depends on the counterbalance between its synthesis and degradation in plum fruit at high temperature. *Sci. Rep.* **7**, 1-16 (2017).

28. Fang, Z. *et al.* Postharvest temperature and light treatments induce anthocyanin accumulation in peel of 'Akihime'plum (*Prunus salicina* Lindl.) via transcription factor *PsMYB10. 1*. *Postharvest Biol. Tec.* **179**, 111592 (2021).

29. Sooriyapathirana, S. S. *et al.* QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). *Tree Genet. Genomes* **6**, 821-832, doi:10.1007/s11295-010-0294-x (2010).

30. Donoso, J. M. *et al.* Exploring almond genetic variability useful for peach improvement: mapping major genes and QTLs in two interspecific almond x peach populations. *Mol. Breeding* **36**, 1-17, doi:10.1007/s11032-016-0441-7 (2016).

31. Salazar, J. A. *et al.* Genotyping by sequencing for SNP-based linkage analysis and identification of QTLs linked to fruit quality traits in Japanese plum (*Prunus salicina* Lindl.). *Front. Plant. Sci.* **8**, doi:10.3389/fpls.2017.00476 (2017).

32. Rahim, M. A., Busatto, N. & Trainotti, L. Regulation of anthocyanin biosynthesis in peach fruits. *Planta* **240**, 913-929 (2014).

33. Shirasawa, K. *et al.* The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.* **24**, 499-508 (2017).

34. Jiang, F. *et al.* The apricot (*Prunus armeniaca* L.) genome elucidates Rosaceae evolution and beta-carotenoid synthesis. *Hortic. Res.* **6**, 128, doi:10.1038/s41438-019-0215-6 (2019).

35. Huang, Z., Shen, F., Chen, Y., Cao, K. & Wang, L. Chromosome-scale genome assembly and population genomics provide insights into the adaptation, domestication, and flavonoid metabolism of Chinese plum. *Plant J.* (2021).

36. Werner, D. J. & Okie, W. A history and description of the *Prunus persica*: plant introduction collection. *Hortscience* **33**, 787-793 (1998).

37. Shen, Z. *et al.* Characterization and genetic mapping of a new blood-flesh trait controlled by the single dominant locus DBF in peach. *Tree Genet. Genomes* **9**, 1435-1446 (2013).

38. Yamamoto, T., Yamaguchi, M. & Hayashi, T. An integrated genetic linkage map of peach by SSR, STS, AFLP and RAPD. *J Jpn Soc Hort Sci* **74**, 204-213 (2005).

39. Beckman, T., Alcazar, J. R., Sherman, W. & Werner, D. Evidence for qualitative suppression of red skin color in peach. *Hortscience* **40**, 523-524 (2005).

40. Zhou, H. *et al.* Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* **82**, 105-121, doi:10.1111/tpj.12792 (2015).

41. Guo, J. *et al.* An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol.* **21**, 1-19 (2020).

42. Salazar, J. A. *et al.* Development and applicability of GBS approach for genomic studies in Japanese plum (*Prunus salicina* Lindl.). *The Journal of Horticultural Science and Biotechnology* **94**, 284-294 (2019).

43. Acuña, C. V. *et al.* Characterization of genetic diversity in accessions of prunus salicina lindl: keeping fruit flesh color ideotype while adapting to water stressed environments. *Agronomy* **9**, 487 (2019).

44. Fiol, A., Jurado-Ruiz, F., Lopez-Girona, E. & Aranzana, M. J. An efficient CRISPR-Cas9 enrichment sequencing strategy for characterizing complex and highly duplicated genomic regions. A case study in the *Prunus salicina* LG3-MYB10 genes cluster *BioRxiv*, doi:https://doi.org/10.1101/2022.01.24.477518 (2022).

45. Zhou, H. *et al.* Two amino acid changes in the R3 repeat cause functional divergence of two clustered MYB10 genes in peach. *Plant Mol. Biol.* **98**, 169-183 (2018).

46. Yang, J. *et al.* A single amino acid substitution in the R2R3 conserved domain of the *BrPAP1a* transcription factor impairs anthocyanin production in turnip (Brassica rapa subsp. rapa). *Plant Physiol. Bioch.* **162**, 124-136 (2021).

47. Verde, I. *et al.* The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* **18**, 225, doi:10.1186/s12864-017-3606-9 (2017).

48. Liu, C. *et al.* Chromosome-level draft genome of a diploid plum (*Prunus salicina*). *GigaScience* **9**, giaa130 (2020).

49. N'Diaye, A. *et al.* Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *Plos One* **12**, e0170941 (2017).

50. Maldonado, C., Mora, F., Scapim, C. A. & Coan, M. Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: Hap LA4. *Plos One* **14**, e0212925 (2019).

51. Calle, A. & Wünsch, A. Multiple-population QTL mapping of maturity and fruit-quality traits reveals LG4 region as a breeding target in sweet cherry (*Prunus avium* L.). *Hortic. Res.* **7**, 1-13 (2020).

52. Fang, Z.-Z. *et al.* Activation of *PsMYB10. 2* transcription causes anthocyanin accumulation in flesh of the red-fleshed mutant of 'Sanyueli'(*Prunus salicina* Lindl.). *Front. Plant. Sci.* **12**, 1167 (2021).

53. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 1-16 (2020).

54. Wei, L. & Cao, X. The effect of transposable elements on phenotypic variation: insights from plants to humans. *Science China Life Sciences* **59**, 24-37 (2016).

55. Jung, S., Venkatesh, J., Kang, M.-Y., Kwon, J.-K. & Kang, B.-C. A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in *Capsicum annuum*. *Plant Sci.* **287**, 110181 (2019).

56. Butelli, E. *et al.* Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. *The Plant Cell* **24**, 1242-1255, doi:10.1105/tpc.111.095232 (2012).

57. Chiu, L.-W. *et al.* The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol.* **154**, 1470-1480 (2010).

58. Lallemand, T., Leduc, M., Landès, C., Rizzon, C. & Lerat, E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes* **11**, 1046 (2020).

59. Otto, S. P. & Yong, P. The evolution of gene duplicates. *Adv. Genet.* **46**, 451-483 (2002).

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

61. Okie, W. in *The Enciclopedia of Fruit and Nuts* (ed J. Janick & R.E. Paull) 694-705 (Cambridge University press, 2008).

62. Salazar, J. A. *et al.* Identification of loci controlling phenology, fruit quality and post-harvest quantitative parameters in Japanese plum (*Prunus salicina* Lindl.). *Postharvest Biol. Tec.* **169**, 111292 (2020).

63. Valderrama-Soto, D. *et al.* Detection of Quantitative Trait Loci Controlling the Content of Phenolic Compounds in an Asian Plum (*Prunus salicina* L.) F1 Population. *Front. Plant. Sci.*, 1331 (2021).

64. Crisosto, C. H., Garner, D., Crisosto, G. M. & Bowerman, E. Increasing 'Blackamber'plum (*Prunus salicina* Lindell) consumer acceptance. *Postharvest Biol. Tec.* **34**, 237-244 (2004).

65. Manganaris, G. A., Vicente, A. R., Crisosto, C. H. & Labavitch, J. M. Effect of delayed storage and continuous ethylene exposure on flesh reddening of 'Royal Diamond'plums. *J. Sci. Food Agr.* **88**, 2180-2185 (2008).

66. Wang, L., Sang, W., Xu, R. & Cao, J. Alteration of flesh color and enhancement of bioactive substances via the stimulation of anthocyanin biosynthesis in 'Friar' plum fruit by low temperature and the removal. *Food Chem.* **310**, 125862 (2020).

67. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. (1987).

68. Fiol, A., Howad, W., Surya, A. & Aranzana, M. Development of molecular markers for fruit skin color in Japanese plum (*Prunus salicina* Lindl.). *Acta Hortic.* **1307**, 221-226, doi:https://doi.org/10.17660/ActaHortic.2021.1307.34 (2021).

69. Kim, H.-Y. *et al.* RNA-seq analysis of spatiotemporal gene expression patterns during fruit development revealed reference genes for transcript normalization in plums. *Plant Mol. Biol. Rep.* **33**, 1634-1649 (2015).

70. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597-W603 (2012).

71. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

72. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

73. Canli, F. & Tian, L. Assessment of regeneration and transient expression factors for Agrobacterium-mediated transformation of *Prunus salicina* Lindl. *Eur. J. Hortic. Sci.* **74**, 66 (2009).

74. Xi, W., Feng, J., Liu, Y., Zhang, S. & Zhao, G. The R2R3-MYB transcription factor PaMYB10 is involved in anthocyanin biosynthesis in apricots and determines red blushed skin. *BMC Plant Biol.* **19**, 287 (2019).

75. Spolaore, S., Trainotti, L. & Casadoro, G. A simple protocol for transient gene expression in ripe fleshy fruit mediated by Agrobacterium. *J. Exp. Bot.* **52**, 845-850 (2001).

76. Zhao, Y. *et al.* Optimization and standardization of transient expression assays for gene functional analyses in strawberry fruits. *Hortic. Res.* **6**, 1-13 (2019).

77. Fang, Z.-Z., Zhou, D.-R., Ye, X.-F., Jiang, C.-C. & Pan, S.-L. Identification of candidate anthocyanin-related genes by transcriptomic analysis of 'Furongli'plum (*Prunus salicina* Lindl.) during fruit ripening using RNA-seq. *Front. Plant. Sci.* **7**, 1338 (2016).

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

79. Ma, K., Zhang, Q., Cheng, T. & Wang, J. Identification of transposons near predicted lncRNA and mRNA pools of Prunus mume using an integrative transposable element database constructed from Rosaceae plant genomes. *Mol. Genet. Genomics* **293**, 1301-1316 (2018).

80. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265-W268, doi:10.1093/nar/gkm286 (2007).

81. Kõressaar, T. *et al.* Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics* **34**, 1937-1938 (2018).

MAIN DISCUSSION

The *MYB10* genes are main determinants of fruit color in the Rosaceae family

Fruit color is an important breeding objective in rosaceous crops. The self-incompatibility and interspecific origin of Japanese plum (*Prunus salicina* Lindl.) has led to a huge phenotypic variability of this trait as well as of others of agronomic importance. Despite the economic relevance of Japanese plum in the market, up to date very few molecular markers have been developed and validated in germplasm and, therefore, the application of marker-assisted breeding in this crop is very limited. The discovery of the *MYB10* genes and their conserved role in the acquisition of fruit color (Lin-Wang et al., 2010) was the starting point for their study in members of the Rosaceae family. In many rosaceous and non-rosaceous species, the genetic variation and/or the differential expression of R2R3-MYB genes can have drastic effects on the fruit color. The source of this variation, leading to a non-functional protein or to differential expression levels, is highly diverse: SNPs, InDels, transposon insertions, tandem repetitions of an autoregulatory *cis*-element or differential methylation levels have been reported as causal variations (Espley et al., 2009; Telias et al., 2011; Castillejo et al., 2020). In some cases, these genes are activated under certain environmental or stress conditions, such as light or cold (Butelli et al., 2012; Fang et al., 2021b). The aim of this thesis was the study of the genetic variability regulating fruit color in Japanese plum to provide efficient molecular markers for their use in marker-assisted selection (MAS).

Up to date, the multiple studies of the *MYB10* genes in linkage group 3 (LG3) have successfully explained the fruit and organ color variation in closely related species and have allowed the development of markers for MAS in *Prunus* species, such as peach, sweet cherry and sour cherry (Stegmeir et al., 2015; Tuan et al., 2015; Sandefur et al., 2016; Bretó et al., 2017). Interestingly, variations in the *MYB10.1* gene sequence have been associated with different fruit color phenotypes in peach. Polymorphisms in this gene explained the red skin color (Tuan et al., 2015) and also the percentage of skin blushing (Bretó et al., 2017), although some samples escaped the association (3 out of 23 for the skin color, 6 out of 87 for the skin blush). The red flesh color around the stone was examined in 295 peach individuals (landraces and accessions) by Guo et al. (2020), who found a deletion in the promoter of the *MYB10.1* gene explaining part of the variability (the 86%, 61% and 77% of the individuals with the deletion in homozygosis, heterozygosis or without it, respectively). The occurrence of samples missing the genotype-phenotype association is probably a result from the high complexity of the fruit color trait, not exclusively regulated by the LG3-*MYB10* genes but by several QTLs scattered along the eight linkage groups (LGs) in *Prunus*. Yet, the *MYB10* genetic

variation largely explains the fruit color variation and is considered a main determinant not only in the *Prunus* genus but the whole Rosaceae family, which motivated the aim of this project.

In this thesis, by studying the LG3-*MYB10* gene variability we have successfully identified polymorphisms in the *PsMYB10.1* and *PsMYB10.2* genes highly correlated with the Japanese plum skin and flesh coloration, respectively; have validated the role of *PsMYB10.2* in promoting anthocyanin accumulation in the flesh and have developed efficient molecular markers for MAS. Moreover, we have improved a novel long-read targeted sequencing methodology, the CRISPR-Cas9 enrichment, providing a useful protocol for plant and animal studies requiring the targeted sequencing of gene duplicated regions in a pool of barcoded samples. Additionally, we established for the first time a protocol for the transient gene expression in Japanese plum fruits, which will be useful for genetic functional studies in the crop.

Genetic characterization of the LG3-*MYB10* region

Several genetic and genomic studies that will aid the breeding of Japanese plum have appeared during the development of this thesis. These include the first saturated linkage maps, from them the identification of QTLs for traits of agronomic interest (which included QTLs for fruit skin color), and the first assembly of a Japanese plum genome (Salazar et al., 2017; Carrasco et al., 2018; Liu et al., 2020). This thesis project started before these studies were available, and thus the possible implication of the *MYB10* genes in LG3 region of Japanese plum was yet unknown as also were their gene sequences. Consequently, we first used primers designed in conserved regions of the gene family by using the peach genome as reference. This gene conservation approach has been successful for the target fingerprinting of genes of the MYB, ERF, KNOX and MADS families in other species (Collard and Mackill, 2009; Poczai et al., 2011), which aimed at the amplification of a high number of polymorphic fragments. In this work, the strategy succeeded in the fingerprinting of LG3-*MYB10* genes in Japanese plum selections and varieties. The amplified bands were phased into haplotypes by studying their segregation in bi-parental families, and later cloned. By their segregation and sequence homology, some of the bands were found to be allelic and could be assigned to genes. The limitations of the approach used is the occurrence of off-target amplification and the overlap of fragments of similar size (Poczai et al., 2013), which we attempted to reduce by targeting exclusively the LG3-*MYB10* genes and through the use of capillary electrophoresis to

increase the resolution of the fragment separation. Our primers design did not produce off-target amplification as was demonstrated by the allele segregation and their homology with the LG3-*MYB10* genes. However, despite the precautions, we obtained several overlapping bands which were only revealed after cloning (allele a467 overlapping with a466) and after exploring the region assemblies of 'Sanyueli' and 'Zhongli' varieties (Chapter 3), which detected the duplication of the *MYB10.2* gene and revealed that the band 466 was made of DNA from two copies of the gene. Another limitation of the technique used here is the presence of null alleles (not amplified or above the capillary electrophoresis detection range) which may mask different haplotypes. This possibility was already hinted for H2 in Chapter 2 and was confirmed in Chapter 4. Our data indicates a high genetic variability of the region in Japanese plum, therefore its classification into haplotypes results useful for the identification of cultivars derived from a common ancestor. The characterization of the MYB10 region into haplotypes was already successful in sweet cherry by using several SNP markers included in an array (Sandefur et al., 2016). In our case of study, we could characterize the haplotypes with a single PCR reaction, increasing the method efficiency.

Polymorphisms in the introns were the cause for the length differences between all the alleles amplified. Markers designed in conserved sequences whose polymorphism rely in intron length differences are highly transferable between species (Poczai et al., 2013), being the closest example the molecular marker for the self-incompatibility used in Japanese plum, which is based on intron size differences of the *S-RNase* gene and is applied in several *Prunus* species (Guerra et al., 2012). Because LG3-*MYB10* genes are conserved in their function and sequence, the strategy and even the primers used here could likely be applied for fingerprinting the LG3-*MYB10* region in other close related species missing molecular markers for fruit color, such as apricot, European plum or myrobalan plum.

The allele segregation revealed the triplication of the *MYB10.1* gene in some haplotypes, confirmed later with the genomes of the Japanese plum varieties 'Zhongli' (Huang et al., 2021) and 'Sanyueli' (Liu et al., 2020) which also contain duplications of the *MYB10.2* gene. While the peach and almond genomes have only three LG3-*MYB10* genes in the cluster (Verde et al., 2017; Alioto et al., 2020), more than three gene copies were observed in the genomes of European plum, apricot and sweet cherry (Shirasawa et al., 2017; Jiang et al., 2019a; Callahan et al., 2021). The subgenus *Prunophora* (plums and apricots) and *Amygdalus* (peaches and almonds) are phylogenetically closer between them than to the *Cerasus* subgenus, which includes the sweet cherries (Shi et al., 2013). This would

indicate a loss of *MYB10* gene copies in the peach and almond genomes in respect to the ancestral species of the former subgenera or, alternately, the gain of gene copies in the other genomes. The comparative genetic analysis performed in Chapter 3 indicated that the region is highly variable between *Prunus* members and includes gene copy number variations (CNV), being the Japanese plum Zhongli-1 region the assembly with more copies identified. Interestingly, not only the size of the region and the *MYB10* gene number seemed to differ between species but also the position of each gene in the cluster. This hints that duplications events may have modulated the LG3-MYB10 region in *Prunus* resulting in CNV between species. In strawberry, segmental duplications have been reported to be the main event expanding the number of the R2R3-MYB family genes (Liu et al., 2021), which might explain the occurrence of gene duplications in the *Prunus* LG3-MYB10 region.

We could clone genes and upstream regulatory sequences thanks to the homology of Japanese plum with the *Prunus* genomes available when working in Chapter 2 (peach, sweet cherry and almond). Designing primers in non-conserved regions when a close reference genome is lacking results highly complicated, and the complexity increases tremendously when the region contains duplicated genes as the lack of PCR amplifications (due to low homology with the reference sequence used) or the amplification of extra bands (due to non-specific primers) occur often. For this reason, we explored the use of CRISPR-Cas9 enrichment methodology to fully characterize the genetic variability in the genic and non-genic regions of the *PsMYB10* cluster. Given the results obtained, we can state now that, in Chapter 2, it would have been a more time-saving strategy to perform first the CRISPR-Cas9 enrichment and then validate the identified polymorphisms with sequence cloning. However, since we had already cloned and detected several polymorphisms in the *PsMYB10* region, we could focus our study on the validation of the CRISPR-Cas9 strategy for the sequencing of complex and high duplicated regions in a pool of varieties, using the formerly known polymorphisms for its validation while identifying new ones.

As for the LG3-*MYB10* primer design, the strategy adopted to enrich the region relied in targeting conserved DNA sequences from the LG3-*MYB10* gene family, and resulted adequate to sequence the region despite its high variability and CNV. The gRNAs were designed prior to the availability of the two Japanese plum genomes (‘Zhongli’ and ‘Sanyueli’), therefore, we can confirm that the strategy can be successfully used for the enrichment and target sequencing of complex genome regions even in the absence of a reference genome. If the same experiment had to be done now, the assemblies of the two Japanese plum genomes published would be of great help, allowing for

the design of additional gRNAs inside regions that presented low read depth or lacked coverage. As discussed in Chapter 3, cleaving multiple sites with a single gRNA pool limits the assembly of the region in a single contig. To overcome it, more than one cleavage reaction with different pools of gRNAs should be performed separately to obtain overlapping fragments. Our approach will be useful for plant and animal studies requiring the sequencing and variant identification in duplicated regions with a single MinION run in a pool of samples, even when lacking a reference genome. In our case, we could identify SNPs and SVs with potential use for the MAS of this region associated with fruit color, including the LTR retrotransposon of 2.8 kb that was associated with flesh color in Chapter 4.

Identification of polymorphisms associated with fruit color and gene validation

The association analysis of the bands with the fruit color phenotype allowed the identification of one dominant allele (a356) associated with the red/purple/black skin color, proving that, as observed in other rosaceous species, the *MYB10* region variation was also determinant for the acquisition of anthocyanin fruit color in Japanese plum. As a result of the high genetic variability of the region, we detected that a356 was present in more than one haplotype (H1, H3, H_i8 and H_i13) and that it carried polymorphisms in the genic and upstream sequence when comparing the allele in H1 and H3. With the segregation and gene homology analysis, we could identify the recessive allele for the same locus (a470), also present in several haplotypes with some polymorphisms between them. A qualitative retrotranscription PCR (RT-PCR) was sufficient to visualize differences in the expression of the a356 and a470 alleles correlating with the red skin color in a switch-on/switch-off manner. However, the use of quantitative methods, such as the determination of anthocyanin and gene expression levels (the latter through a quantitative RT-PCR), would have been more precise. This has already been performed in another study, also correlating the *PsMYB10.1* expression with anthocyanin levels and red skin in Japanese plums (Fang et al., 2021b). Gene validation in fruit trees, even if possible, has the limitation of the time required to observe the phenotype, especially if it is related to fruit traits. Therefore, gene validation of *MYB10* genes has been usually done by transient gene expression in heterologous systems like *Nicotiana tabacum* or *N. benthamiana* (Starkevič et al., 2015; Zhou et al., 2015; Fang et al., 2021b). All the studies point out the *PsMYB10.1* gene as the responsible for the skin color in Japanese plum but, nonetheless, it should be validated in a homologous system. In this case, the protocol provided in this thesis should be useful to validate the gene and the polymorphisms associated with the differential expression of a356 and a470.

When searching for genotype-phenotype associations, haplotype data can provide results not reached by single marker analysis (N'Diaye et al., 2017), which was the case in Chapter 4 with the flesh color. The association of single alleles with flesh color failed in Chapter 2 because the *MYB10* marker was unable to capture the polymorphisms in the *PsMYB10.2* promoter, which were the cause of the trait variation. The insertion of the LTR retrotransposon may be recent and, therefore, the functional and non-functional copies of the gene (*PsMYB10.2*-H2) had no time to diverge and to accumulate changes detectable in the association analysis. In consequence, the band amplified for the *PsMYB10.2*-H2 allele (a466) was not exclusive of the haplotype. Both single allele and haplotype association methods should be used in other studies to identify significant association with the trait/s of analysis (Maldonado et al., 2019).

The large (2.8 kb), repetitive and terminal duplicated sequence found in our study would have been difficult to identify by other methods (such as the used in Chapter 2), validating the adequacy of the sequencing method used. Retrotransposons have already been reported to be able to increase the expression of genes, being the most similar cases from the studied here the occurring in red skin apples, which gain the color by a LTR retrotransposon activating the *MYB10* expression (Zhang et al., 2019); and red flesh peaches, in which a LTR retrotransposon might enhance the expression of the *BLOOD* (*BL*) gene, which is a *PpMYB10.1* activator (Hara-Kitagawa et al., 2020).

Contrarily to *PsMYB10.1* and fruit skin color, the correlation of *PsMYB10.2* expression with flesh color differed from the one reported for peach and sweet cherry, in which the expression of their *MYB10.2* homolog did not explain the fruit color (Rahim et al., 2014; Tuan et al., 2015; Jin et al., 2016). Therefore, due to the lack of evidences in other crops, we found the need to validate the *PsMYB10.2* gene function. Even though the genetic transformation is possible in plum (Urtubia et al., 2008) this is a long process and the transient overexpression in Japanese plum fruits was preferred. No protocol had been previously described for the transient gene expression in Japanese plum fruits and, consequently, we set up a new protocol based on studies done in other species. We adjusted several parameters to increase the gene expression efficiency: Maturity of the fruits, injection dose (concentration of *Agrobacterium tumefaciens* cells, injection volume), number of injections and their site, days of incubation and storage conditions (light/dark).

Soon after performing the transient expression experiments, the work from Fang et al. (2021a) was published, in which they also validated the *PsMYB10.2* function by transient overexpression in

Nicotiana leaves and by virus induced gene silencing (VIGS) in Japanese plum fruits. However, while their work focused in the gain of red flesh color of a radiation-induced mutant, our work was centered in the natural variation of the trait, relevant for breeding. They reported that the *PsMYB10.2* expression upregulated the *UFGT* and *GST* genes responsible for the synthesis and transport of anthocyanins to the vacuole, respectively, which altogether with our work provide strong evidence to propose the *PsMYB10.2* as candidate for the natural variation of red flesh color in Japanese plum. The anthocyanin-promoting activity of *PsMYB10.2* was found to be lower than *PpMYB10.1* (Fang et al., 2021a) and might be the reason explaining the increased anthocyanin levels observed in the skin over the flesh, which can reach seven fold as calculated for one cultivar (Fanning et al., 2014).

The transient expression protocol designed can be used for the validation of genes involved in fruit quality. It can prove useful to validate other genes regulating anthocyanin biosynthesis, or controlling other color traits such as the caused by carotenoids or chlorophylls. However, it is not limited to fruit color traits, as could be used to detect changes in other metabolic pathways if it is accompanied by an observable or quantifiable response. We were able to validate the function of the candidate gene using its entire sequence (exons and introns) from the PCR amplification of genomic DNA, therefore the method allows for testing expressed as well as not expressed alleles (such as the *PsMYB10.2*-H4). Interestingly, while setting up the protocol we also achieved the transient expression of the *PsMYB10.2* gene in apricot (data not shown), which also acquired red flesh color at the site of the injection. This would indicate that the protocol can be transferred to the fruits of other close *Prunus* species, and that the expression of the *MYB10* genes are also determinant for fruit color in apricot.

Design of molecular markers for fruit color useful for marker-assisted selection

In this work we have developed efficient molecular markers to select for the presence of anthocyanins in the fruit skin and flesh of Japanese plum fruits. For the molecular marker for skin color, with a single PCR reaction we are able to predict better the skin color phenotype than in González et al. (2016b), which required three. For the flesh color, this is the first time that a molecular marker has been validated in a large panel of Japanese plum varieties. Altogether, these two molecular markers greatly contribute to the marker-assisted breeding of the crop, predicting

the presence/absence of anthocyanins with an efficacy of 97.9% in the skin and 93.1% in the flesh in the 145 selections and varieties used in Chapter 4.

We reported here that the genetic variation in *PsMYB10.1* and *PsMYB10.2* genes is accountable for anthocyanin accumulation in the fruit skin and flesh tissues, respectively. This might be a case of subfunctionalization, in which the duplicated genes have subdivided the function of the ancestral gene by promoting anthocyanin synthesis in different tissues (Lallemand et al., 2020). The *PsMYB10.1* and *PsMYB10.2* gene copies responsible for the fruit color are 38 kb and 81 kb apart in 'Zhongli No. 6' v1 (Huang et al., 2021) and 'Sanyueli' v2 (Liu et al., 2020) genomes, respectively. If only the seven observed haplotypes are considered (H1 to H6, plus the H2*), the active copy of *PsMYB10.1* (with a356) is present in H1 and H3, and the active copy of *PsMYB10.2* (with the LTR retrotransposon insertion) in H2. Since the genes are physically close but their functionally active alleles are found in different phases (i.e. different haplotypes), they have few chances to be on-phase by recombination and co-segregate. The use of the molecular markers designed here will aid at selecting the most appropriate parentals when planning crosses and allow the early selection of their seedlings according to the objectives of each breeding program (**Figure D.1**).

Salazar et al. (2017) found a QTL in the LG3-*PsMYB10* region that explained the tendency from red to purple skin genotypes, which matched in our data with H1 haplotype, more present in black than red skin selections. However, this result is not conclusive due to many outliers and the MAS is, for now, only effective for the presence/absence of anthocyanin color. Similarly, the allelic variation in the *PpMYB10.1* gene also correlates with the intensity of coloration in the skin and flesh tissues (Tuan et al., 2015; Hara-Kitagawa et al., 2020), but also with some genotypes missing this association.

Given the high genetic variability found in the MYB10 region, we found plausible the existence of additional haplotypes and alleles also regulating anthocyanin synthesis. In agreement with this, we already found one red flesh accession with the LTR retrotransposon in a haplotype other than H2, as well as one allele (a243), possibly allelic of the *PsMYB10.1a* gene, that may also explain the lack of red skin color by a different polymorphism from the identified in a470. All this hints that additional polymorphisms might drive the fruit skin and flesh coloration regulated by the LG3-MYB10 region in Japanese plum, as also occurs in other rosaceous species (Jin et al., 2016; Castillejo et al., 2020).

With the resources provided here, researchers and breeders will be able to characterize their vegetal material and identify those haplotypes contributing to the segregation of fruit color in their germplasm and populations. However, similar to other LG3-MYB10 markers, these are not enough to explain all the observed variability, and future work will be required to identify other regions in the Japanese plum genome involved in the anthocyanin presence and color intensity, as well as the study of colors conferred by other pigments.

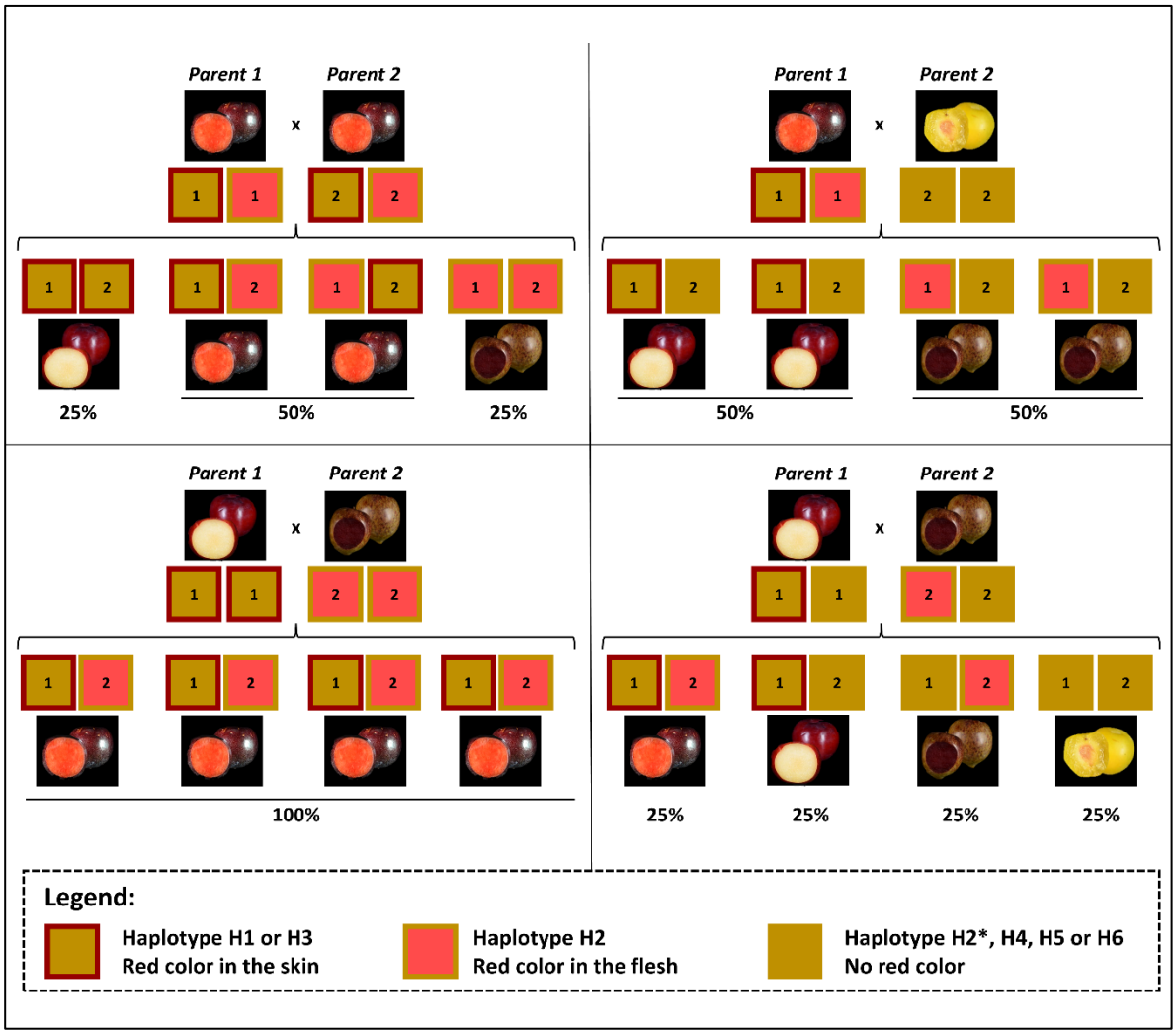


Figure D.1. The expected color segregation in four different crosses, taking only into account the seven observed MYB10 haplotypes (H1 to H6, with the H2*) and assuming a total lack of recombination in the region. Note that the parents of the two crosses at the bottom part of the figure share the same phenotypes while the trait segregation differs considerably, highlighting the utility of the molecular markers developed for breeding.

Future perspectives

The release of the Japanese plum genomes has provided a valuable resource to assist the breeding of the crop (Liu et al., 2020), because it is expected to facilitate studies of genome variation and the development of molecular markers for other traits of interest. The gene validation protocol designed here will be useful to validate other candidate genes or causal polymorphisms proposed from these studies. However, the protocol can still be improved by testing parameters that we did not consider: use of fruits still attached to the tree, the efficiency of different *Agrobacterium* strains or the incubation of the fruits in controlled conditions (temperature and humidity). We validated the *PsMYB10.2* gene function but not the *PsMYB10.1*, which should also be validated with the same protocol. The putative causal polymorphisms explaining the expression of these genes should also be validated by replacing the CaMV 35S promoter of the expression vector.

The MAS of the fruit color regulated by the *MYB10* genes can be further improved by using SNP data, and/or reducing the genotyping costs. In sweet cherry, even though the fruit color region was also characterized in haplotypes, a unique SSR marker is used for the routinely prediction of the trait in MAS (Sandefur et al., 2016), a strategy that we could adopt in the future by selecting, among the 129 SNPs exclusive for H1 and H3 and among the 289 SNPs exclusive for H2 identified in the CRISPR-Cas9 enrichment, potential candidates to move from the PCR markers designed here to SNP markers. We attempted to design a non-multilocus molecular marker for the skin color to give an alternative to those laboratories lacking capillary electrophoresis, but we did not succeed due to the multiple *PsMYB10.1* gene copies. The molecular marker for the flesh color can be visualized in either agarose gels or separated by capillary electrophoresis and visualized by adding a fluorochrome to the LTR_f primer. Since the PCR bands of the skin and flesh markers do not overlap, the two reactions can be loaded together in the ABI Genetic Analyzer or attempt to multiplex the PCR reactions to optimize the genotyping cost per sample (**Figure D.2**).

Consistent with previous studies of *MYB10* genes variation and QTL identification in peach, the *PsMYB10* genes explained a large part of the phenotypic variance but could not explain all the cases. QTL studies will be required to identify other genomic regions responsible for skin and flesh colors. We already detected two red flesh phenotypes differing in the spatial and temporal expression of the flesh color trait and one selection with an extra level of regulation (selection C4), which should

be further investigated. In Salazar et al. (2017) authors already identified a QTL for skin color in LG4. For future QTL studies it will be preferable to phenotype with a quantitative method, such as measuring the anthocyanin concentration or by colorimetric methods, which could aid at identifying more genomic regions involved in fruit coloration and be able to discriminate between red and black skin phenotypes. The coloration obtained by carotenoids and chlorophyll pigments should also be investigated to breed for the colors exposed when anthocyanins are absent, such as white (flesh), yellow, orange and green hues.

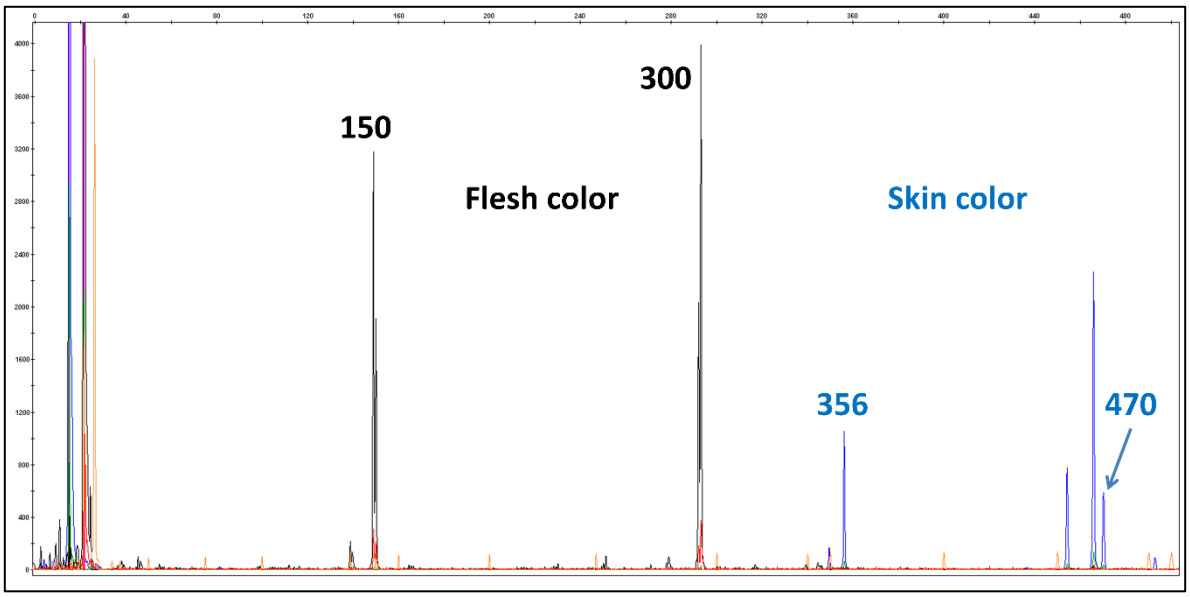


Figure D.2. Visualization of the pooled PCR reactions of the molecular markers for the skin and flesh color in the ABI Genetic Analyzer, for 'Black Gold' (H1/H2) cultivar (black skin, red flesh). The x-axis corresponds to the band length, the y-axis to the fluorescence intensity of each amplified band. The black peaks correspond to the marker for the flesh color, the blue for the skin color. The 150 bp and 356 bp bands correspond to the alleles associated with anthocyanin flesh and skin color, respectively, and the 300 bp and 470 bp bands to the recessive alleles for their loci.

CONCLUSIONS

1. A primer pair designed at conserved LG3-*MYB10* gene sequences in peach allowed to determine the allelic variability of MYB10 orthologs in Japanese plum. The allele segregation in 6 biparental families confirmed that they mapped in the LG3-*MYB10* region, and could be organized into haplotypes. The alleles were classified according to their MYB10 homology and served to identify a *MYB10.1* triplication in some Japanese plum individuals.

2. The band a356 (allele of *PsMYB10.1a*) amplified in several haplotypes (H1, H3, H_i8 and H_i13) explained in a dominant way the red skin color, while a470 was the alternative, recessive allele. Their expression in skin fruit tissue correlated with the appearance of red coloration.

3. The observed haplotypes were highly represented in the studied panel of breeding lines and commercial varieties (91.36% and 83.87% of the individuals, respectively) and easily identified with a PCR reaction, and therefore, useful for the early selection of the trait in marker-assisted selection breeding programs.

4. The whole-genome Illumina reads of two Japanese plum varieties aligned poorly to the peach, almond and sweet cherry genomes, but served to clone the *PsMYB10.1a* full gene sequences and their upstream regions. Putative causal polymorphisms regulating skin color were identified.

5. The allelic segregation and sequences indicated a high level of genetic variability in the MYB10 region in Japanese plum. This was later confirmed by target sequencing using CRISPR-Cas9 enrichment technology and by comparing the region assembly in 15 *Prunus* genomes. SNPs and SVs with potential use for the MAS of fruit color were extracted and phased into haplotypes H1 to H6 and H_i9, from both the genome alignment and the *de novo* sequences.

6. We validated the CRISPR-Cas9 enrichment sequencing strategy to allow, in a pool of 5 barcoded samples, the long-read sequencing and assembly of highly complex genomic regions. This can be applied in other studies requiring the sequencing of regions with high complexity, at a reduced cost per sample, even if lacking a reference genome.

7. The haplotype data was critical to identify association with the flesh color. The expression of *PsMYB10.2* gene in H2 correlated with the red flesh color, and an LTR retrotransposon in the gene promoter was discovered by exploring the CRISPR-Cas9 *de novo* sequences. Its identification improved the correlation with the flesh color trait and further validated the CRISPR-Cas9 enrichment method.

8. The *PsMYB10.2* gene function was validated by its transient overexpression in Japanese plum fruits, setting up a protocol that can be used to validate other candidate genes for fruit quality. The polymorphisms that putatively explain the *PsMYB10.1* and *PsMYB10.2* genes expression should be also validated by the same method.

9. The MYB10 region contributes greatly to Japanese plum fruit color and should be considered for the marker-assisted breeding of the crop. The two molecular markers developed here predict with an efficiency of 97.9% and 93.1% the absence/presence of anthocyanins in the fruit skin and flesh, respectively, in the 145 accessions and commercial varieties used in Chapter 4. The markers will allow breeders to characterize the haplotypes contributing to the segregation of fruit color in their parental lines, as well as to use them for marker-assisted selection.

10. Given the high variability in the studied region, additional polymorphisms contributing to fruit color should not be discarded. The molecular marker for the skin color and the CRISPR-Cas9 sequencing will be useful to identify new haplotypes and polymorphisms regulating fruit color.

11. The MYB10 region explained a large part of the natural variation in fruit color, but the fruit color is a complex trait also regulated by other genes. Thus, QTL analysis should be first performed to identify additional regions contributing to the absence/presence of anthocyanins, to their different accumulation levels, as well as to the coloration conferred by other pigments such as carotenoids and chlorophylls.

MAIN BIBLIOGRAPHY

Acuña, C. V., Rivas, J. G., Brambilla, S. M., Cerrillo, T., Frusso, E. A., García, M. N., et al. (2019). Characterization of Genetic Diversity in Accessions of Prunus salicina Lindl: Keeping fruit flesh color ideotype while adapting to water stressed environments. *Agronomy* 9. doi:10.3390/agronomy9090487.

Alioto, T., Alexiou, K. G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., et al. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J.* 101, 455–472. doi:10.1111/tpj.14538.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30. doi:10.1186/s13059-020-1935-5.

Aranzana, M. J., Decroocq, V., Dirlewanger, E., Eduardo, I., Gao, Z. S., Gasic, K., et al. (2019). Prunus genetics and applications after de novo genome sequencing: achievements and prospects. *Hortic. Res.* 6, 58. doi:10.1038/s41438-019-0140-8.

Arús, P. (2017). Molecular markers for plant genetics and breeding. *Contrib. to Sci.* 13. doi:10.2436/20.7010.01.256.

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol.* 48, 958–970. doi:10.1093/pcp/pcm066.

Beckmann, J. S., and Soller, M. (1986). Restriction fragment length polymorphisms in plant genetic improvement. *Oxford Surv. plant Mol. cell Biol.* 3, 196–250.

Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K., and Wang, J. (2019). Applications and analysis of targeted genomic sequencing in cancer studies. *Comput. Struct. Biotechnol. J.* 17, 1348–1359. doi:10.1016/j.csbj.2019.10.004.

Boonprakob, U., Byrne, D. H., Graham, C. J., Okie, W. R., Beckman, T., and Smith, B. R. (2001). Genetic Relationships among Cultivated Diploid Plums and Their Progenitors as Determined by RAPD Markers. *J. Am. Soc. Hortic. Sci.* 126, 451–461. doi:10.21273/JASHS.126.4.451.

Bretó, M. P., Cantín, C. M., Iglesias, I., Arús, P., and Eduardo, I. (2017). Mapping a major gene for red skin color suppression (highlighter) in peach. *Euphytica* 213, 14. doi:10.1007/s10681-016-1812-1.

Bueno, J. M., Sáez-Plaza, P., Ramos-Escudero, F., Jiménez, A. M., Fett, R., and Asuero, A. G. (2012). Analysis and Antioxidant Capacity of Anthocyanin Pigments. Part II: Chemical Structure, Color, and Intake of Anthocyanins. *Crit. Rev. Anal. Chem.* 42, 126–151. doi:10.1080/10408347.2011.632314.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., et al. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255. doi:10.1105/tpc.111.095232.

Byrne, D. H. (1990). Isozyme variability in four diploid stone fruits compared with other woody perennial plants. *J. Hered.* 81, 68–71. doi:10.1093/oxfordjournals.jhered.a110927.

Byrne, D. H. (2012). "Trends in fruit breeding," in *Fruit Breeding*, eds. M. L. Badenes and D. H. Byrne (Boston, MA: Springer US), 3–36. doi:10.1007/978-1-4419-0763-9_1.

Byrne, D., Raseira, M., Bassi, D., Piagnani, C., Gasic, K., Reighard, G., et al. (2012). "Peach," in *Fruit Breeding*, eds. M. L. Badenes and D. H. Byrne (New York (USA): Springer), 505-570.

Callahan, A. M., Zhebentyayeva, T. N., Humann, J. L., Saski, C. A., Galimba, K. D., Georgi, L. L., et al. (2021). Defining the 'HoneySweet' insertion event utilizing NextGen sequencing and a de novo genome assembly of plum (Prunus domestica). *Hortic. Res.* 8, 8. doi:10.1038/s41438-020-00438-2.

Carrasco, B., Díaz, C., Moya, M., Gebauer, M., and García-González, R. (2012). Genetic characterization of Japanese plum cultivars (Prunus salicina) using SSR and ISSR molecular markers. *Cienc. e Investig. Agrar.* 39, 533–543. doi:10.4067/s0718-16202012000300012.

Carrasco, B., González, M., Gebauer, M., García-González, R., Maldonado, J., and Silva, H. (2018). Construction of a highly saturated linkage map in Japanese plum (Prunus salicina L.) using GBS for SNP marker calling. *PLoS One* 13, e0208032. doi:10.1371/journal.pone.0208032.

Castillejo, C., Waurich, V., Wagner, H., Ramos, R., Oiza, N., Muñoz, P., et al. (2020). Allelic variation of MYB10 Is the major force controlling natural variation in skin and flesh color in strawberry (Fragaria spp.) fruit. *Plant Cell* 32, 3723–3749. doi:10.1105/tpc.20.00474.

Celli, G. B., Tan, C., and Selig, M. J. (2018). "Anthocyanidins and anthocyanins," in *Encyclopedia of Food Chemistry*, eds. L. Melton, F. Shahidi, and P. B. T.-E. of F. C. Varelis (Oxford: Academic Press), 218–223. doi:10.1016/B978-0-08-100596-5.21780-0.

Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., et al. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS One* 7, e31745. doi:10.1371/journal.pone.0031745.

Channuntapipat, C., Wirthensohn, M., Ramesh, S. A., Batlle, I., Arús, P., Sedgley, M., et al. (2003). Identification of incompatibility genotypes in almond (Prunus dulcis Mill.) using specific primers based on the introns of the S-alleles. *Plant Breed.* 122, 164–168. doi:10.1046/j.1439-0523.2003.00842.x.

Collard, B. C. Y., and Mackill, D. J. (2009). Conserved DNA-derived polymorphism (CDDP): A simple and novel method for generating DNA markers in plants. *Plant Mol. Biol. Report.* 27, 558–562. doi:10.1007/s11105-009-0118-z.

Cone, K. C., Burr, F. A., and Burr, B. (1986). Molecular analysis of the maize anthocyanin regulatory locus C1. *Proc. Natl. Acad. Sci. U. S. A.* 83, 9631–9635. doi:10.1073/pnas.83.24.9631.

Crisosto, C. H., Garner, D., Crisosto, G. M., and Bowerman, E. (2004). Increasing "Blackamber" plum (Prunus salicina Lindell) consumer acceptance. *Postharvest Biol. Technol.* 34, 237–244. doi:10.1016/j.postharvbio.2004.06.003.

D. Ruiz, P. Lambert, J.M. Audergon, L. Dondini, S. Tartarini, M. Adami, et al. (2010). Identification of Qtls for Fruit Quality Traits in Apricot. in *Acta Horticulturae* (International Society for Horticultural Science (ISHS), Leuven, Belgium), 587–592. doi:10.17660/actahortic.2010.862.93.

Dirlewanger, E., Graziano, E., Joobeur, T., Garriga-Calderé, F., Cosson, P., Howad, W., et al. (2004a). Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9891–9896. doi:10.1073/pnas.0307937101.

Dirlewanger, E., Kleinhentz, M., Laigret, F., Gómez-Aparisi, J., Rubio-Cabetas, M. J., Claverie, M., et al. (2004b). Breeding for a new generation of prunus rootstocks based on marker-assisted selection: A European initiative. in *Acta Horticulturae* (International Society for Horticultural Science (ISHS), Leuven, Belgium), 829–834. doi:10.17660/ActaHortic.2004.663.150.

Donoso, J. M., Eduardo, I., Picañol, R., Batlle, I., Howad, W., Aranzana, M. J., et al. (2015). High-density mapping

suggests cytoplasmic male sterility with two restorer genes in almond × peach progenies. *Hortic. Res.* 2, 15016. doi:10.1038/hortres.2015.16.

Donoso, J. M., Picañol, R., Serra, O., Howad, W., Alegre, S., Arús, P., et al. (2016). Exploring almond genetic variability useful for peach improvement: mapping major genes and QTLs in two interspecific almond × peach populations. *Mol. Breed.* 36, 1–17. doi:10.1007/s11032-016-0441-7.

Eduardo, I., Pacheco, I., Chietera, G., Bassi, D., Pozzi, C., Vecchietti, A., et al. (2011). QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet. Genomes* 7, 323–335. doi:10.1007/s11295-010-0334-6.

Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., et al. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* 21, 168–183. doi:10.1105/tpc.108.059329.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49, 414–427. doi:10.1111/j.1365-313X.2006.02964.x.

Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., et al. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J.* 76, 175–187. doi:10.1111/tpj.12283.

Fang, Z.-Z., Lin-Wang, K., Zhou, D.-R., Lin, Y.-J., Jiang, C.-C., Pan, S.-L., et al. (2021a). Activation of PsMYB10.2 Transcription Causes Anthocyanin Accumulation in Flesh of the Red-Fleshed Mutant of 'Sanyueli' (Prunus salicina Lindl.). *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.680469.

Fang, Z., Lin-Wang, K., Jiang, C., Zhou, D., Lin, Y., Pan, S., et al. (2021b). Postharvest temperature and light treatments induce anthocyanin accumulation in peel of 'Akihime' plum (Prunus salicina Lindl.) via transcription factor PsMYB10.1. *Postharvest Biol. Technol.* 179, 111592. doi:10.1016/j.postharvbio.2021.111592.

Fang, Z., Lin-Wang, K., Dai, H., Zhou, D., Jiang, C., Espley, R. V., et al. (2022). The genome of low-chill Chinese plum 'Sanyueli' (Prunus salicina Lindl.) provides insights into the regulation of the chilling requirement of flower buds. *Mol. Ecol. Resour.* n/a. doi:10.1111/1755-0998.13585.

Fang, Z. Z., Zhou, D. R., Ye, X. F., Jiang, C. C., and Pan, S. L. (2016). Identification of candidate anthocyanin-related genes by transcriptomic analysis of 'furongli' plum (Prunus salicina lindl.) during fruit ripening using RNA-seq. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01338.

Fanning, K. J., Topp, B., Russell, D., Stanley, R., and Netzel, M. (2014). Japanese plums (Prunus salicina Lindl.) and phytochemicals - breeding, horticultural practice, postharvest storage, processing and bioactivity. *J. Sci. Food Agric.* 94, 2137–2147. doi:10.1002/jsfa.6591.

FAOSTAT (2019). Food and agriculture Organization of the United Nations. Available at: https://www.fao.org/faostat [Accessed October 15, 2021].

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., et al. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. *Cell Rep.* 6, 32–43. doi:10.1016/j.celrep.2013.12.009.

Gil, M. I., Tomás-Barberán, F. A., Hess-Pierce, B., and Kader, A. A. (2002). Antioxidant capacities, phenolic compounds, carotenoids, and vitamin C contents of nectarine, peach, and plum cultivars from California.

J. Agric. Food Chem. 50, 4976–4982. doi:10.1021/jf020136b.

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., et al. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* 38, 433–438. doi:10.1038/s41587-020-0407-5.

Goldschmidt, E. E. (2013). The Evolution of Fruit Tree Productivity: A Review. *Econ. Bot.* 67, 51–62. doi:10.1007/s12231-012-9219-y.

Gómez, E., and Ledbetter, C. (1993). Transmission of Biochemical Flavor Constituents from Apricot and Plum to their Interspecific Hybrid. *Plant Breed.* 111, 236–241. doi:10.1111/j.1439-0523.1993.tb00634.x.

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* 53, 814–827. doi:10.1111/j.1365-313X.2007.03373.x.

González, M., Maldonado, J., Salazar, E., Silva, H., and Carrasco, B. (2016a). De novo transcriptome assembly of “Angeleno” and “Lamoon” Japanese plum cultivars (Prunus salicina). *Genomics Data* 9, 35–36. doi:10.1016/j.gdata.2016.06.010.

González, M., Salazar, E., Castillo, J., Morales, P., Mura-Jornet, I., Maldonado, J., et al. (2016b). Genetic structure based on EST–SSR: a putative tool for fruit color selection in Japanese plum (Prunus salicina L.) breeding programs. *Mol. Breed.* 36, 68. doi:10.1007/s11032-016-0491-x.

Gould, K. S. (2004). Nature’s Swiss army knife: The diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol.* 2004, 314–320. doi:10.1155/S1110724304406147.

Gu, C., Liao, L., Zhou, H., Wang, L., Deng, X., and Han, Y. (2015). Constitutive activation of an anthocyanin regulatory gene PcMYB10.6 is related to red coloration in purple-foliage plum. *PLoS One* 10, e0135159. doi:10.1371/journal.pone.0135159.

Gu, C., Wang, L., Wang, W., Zhou, H., Ma, B., Zheng, H., et al. (2016). Copy number variation of a gene cluster encoding endopolygalacturonase mediates flesh texture and stone adhesion in peach. *J. Exp. Bot.* 67, 1993–2005. doi:10.1093/jxb/erw021.

Guerra, M. E., López-Corrales, M., and Wünsch, A. (2012). Improved S-genotyping and new incompatibility groups in Japanese plum. *Euphytica* 186, 445–452. doi:10.1007/s10681-012-0636-x.

Guerra, M. E., and Rodrigo, J. (2015). Japanese plum pollination: A review. *Sci. Hortic. (Amsterdam).* 197, 674–686. doi:10.1016/j.scienta.2015.10.032.

Guerrero, B. I., Guerra, M. E., Herrera, S., Irisarri, P., Pina, A., and Rodrigo, J. (2021). Genetic diversity and population structure of japanese plum-type (Hybrids of p. salicina) accessions assessed by ssr markers. *Agronomy* 11. doi:10.3390/agronomy11091748.

Guo, J., Cao, K., Deng, C., Li, Y., Zhu, G., Fang, W., et al. (2020). An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol.* 21, 258. doi:10.1186/s13059-020-02169-y.

Hara-Kitagawa, M., Unoki, Y., Hihara, S., and Oda, K. (2020). Development of simple PCR-based DNA marker for the red-fleshed trait of a blood peach ‘Tenshin-suimitsuto.’ *Mol. Breed.* 40, 5. doi:10.1007/s11032-019-1068-2.

Hartmann, W., and Neumüller, M. (2009). "Plum Breeding," in *Breeding Plantation Tree Crops: Temperate Species* (New York, NY: Springer New York), 161–231. doi:10.1007/978-0-387-71203-1_6.

He, J., and Giusti, M. (2010). Anthocyanins: Natural colorants with health-promoting properties. *Annu. Rev. Food Sci. Technol.* 1, 163–187. doi:10.1146/annurev.food.080708.100754.

Huang, Z., Shen, F., Chen, Y., Cao, K., and Wang, L. (2021). Chromosome-scale genome assembly and population genomics provide insights into the adaptation, domestication, and flavonoid metabolism of Chinese plum. *Plant J.* 108, 1174–1192. doi:10.1111/tpj.15482.

Ingram, C. (1948). *Ornamental Cherries*. London (UK): Country life limited doi:10.2307/4119752.

Jiang, F., Zhang, J., Wang, S., Yang, L., Luo, Y., Gao, S., et al. (2019a). The apricot (Prunus armeniaca L.) genome elucidates Rosaceae evolution and beta-carotenoid synthesis. *Hortic. Res.* 6, 128. doi:10.1038/s41438-019-0215-6.

Jiang, S., Chen, M., He, N., Chen, X., Wang, N., Sun, Q., et al. (2019b). MdGSTF6, activated by MdMYB1, plays an essential role in anthocyanin accumulation in apple. *Hortic. Res.* 6, 40. doi:10.1038/s41438-019-0118-6.

Jin, W., Wang, H., Li, M., Wang, J., Yang, Y., Zhang, X., et al. (2016). The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (Prunus avium L.). *Plant Biotechnol. J.* 14, 2120–2133. doi:10.1111/pbi.12568.

Jo, Y., Lian, S., Cho, J. K., Choi, H., Chu, H., and Cho, W. K. (2015). De novo transcriptome assembly of two different Prunus salicina cultivars. *Genomics Data* 6, 262–263. doi:10.1016/j.gdata.2015.10.015.

Joobeur, T., Periam, N., De Vicente, M. C., King, G. J., and Arus, P. (2000). Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome* 43, 649–655. doi:10.1139/g00-040.

Joobeur, T., Viruel, M. A., De Vicente, M. C., Jáuregui, B., Ballester, J., Dettori, M. T., et al. (1998). Construction of a saturated linkage map for Prunus using an almond x peach F2 progeny. *Theor. Appl. Genet.* 97, 1034–1041. doi:10.1007/s001220050988.

Karp, D. (2015). Luther Burbank's Plums. *HortScience horts* 50, 189–194. doi:10.21273/HORTSCI.50.2.189.

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* 61. doi:10.1080/16546628.2017.1361779.

Khoo, H. E., Prasad, K. N., Kong, K. W., Jiang, Y., and Ismail, A. (2011). Carotenoids and their isomers: Color pigments in fruits and vegetables. *Molecules* 16, 1710–1738. doi:10.3390/molecules16021710.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: A colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242. doi:10.1016/j.tplants.2005.03.002.

Kong, J. M., Chia, L. S., Goh, N. K., Chia, T. F., and Brouillard, R. (2003). Analysis and biological activities of anthocyanins. *Phytochemistry* 64, 923–933. doi:10.1016/S0031-9422(03)00438-2.

Lallemand, T., Leduc, M., Landès, C., Rizzon, C., and Lerat, E. (2020). An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes (Basel).* 11, 1–40. doi:10.3390/genes11091046.

Lancaster, J. E., Grant, J. E., Lister, C. E., and Taylor, M. C. (1994). Skin color in apples - Influence of copigmentation and plastid pigments on shade and darkness of red color in five genotypes. *J. Am. Soc. Hortic. Sci.* 119, 63–69. doi:10.21273/jashs.119.1.63.

LaRue, J. H., and Norton, M. V (1989). "Japanese plum pollination," in *Peaches, Plums, and Nectarines. Growing and Handling for Fresh Market*, eds. J. H. LaRue and R. S. Johnson (Oakland, CA (USA): University of California, DANR), 49–55.

Lee, S., and Wen, J. (2001). A phylogenetic analysis of Prunus and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *Am. J. Bot.* 88, 150–160.

Li, X., Singh, J., Qin, M., Li, S., Zhang, X., Zhang, M., et al. (2019). Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (Pyrus). *Plant Biotechnol. J.* 17, 1582–1594. doi:10.1111/pbi.13085.

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T. K., et al. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol.* 10, 50. doi:10.1186/1471-2229-10-50.

Liu, C., Feng, C., Peng, W., Hao, J., Wang, J., Pan, J., et al. (2020). Chromosome-level draft genome of a diploid plum (Prunus salicina). *Gigascience* 9, giaa130. doi:10.1093/gigascience/giaa130.

Liu, H. J., and Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. *Plant J.* 97, 8–18. doi:10.1111/tpj.14139.

Liu, J., Wang, J., Wang, M., Zhao, J., Zheng, Y., Zhang, T., et al. (2021). Genome-Wide Analysis of the R2R3-MYB Gene Family in Fragaria × ananassa and Its Function Identification During Anthocyanins Biosynthesis in Pink-Flowered Strawberry. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.702160.

López-Girona, E., Davy, M. W., Albert, N. W., Hilario, E., Smart, M. E. M., Kirk, C., et al. (2020). CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods* 16, 121. doi:10.1186/s13007-020-00661-x.

Lozano, M., Vidal-Aragón, M. C., Hernández, M. T., Ayuso, M. C., Bernalte, M. J., García, J., et al. (2009). Physicochemical and nutritional properties and volatile constituents of six Japanese plum (Prunus salicina Lindl.) cultivars. *Eur. Food Res. Technol.* 228, 403–410. doi:10.1007/s00217-008-0946-3.

Lu, Z., Cao, H., Pan, L., Niu, L., Wei, B., Cui, G. C., et al. (2021). Two loss-of-function alleles of the glutathione S-transferase (GST) gene cause anthocyanin deficiency in flower and fruit skin of peach (Prunus persica). *Plant J.* 107, 1320–1331. doi:10.1111/tpj.15312.

Luo, Z., Xie, J., Xu, T., and Zhang, L. (2009). Delay ripening of "Qingnai" plum (Prunus salicina Lindl.) with 1-methylcyclopropene. *Plant Sci.* 177, 705–709. doi:10.1016/j.plantsci.2009.08.013.

Maldonado, C., Mora, F., Scapim, C. A., and Coan, M. (2019). Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: HAPla4. *PLoS One* 14, e0212925. doi:10.1371/journal.pone.0212925.

Marandel, G., Pascal, T., Candresse, T., and Decroocq, V. (2009). Quantitative resistance to plum pox virus in prunus davidiana P1908 linked to components of the eukaryotic translation initiation complex. *Plant Pathol.* 58, 425–435. doi:10.1111/j.1365-3059.2008.02012.x.

Marti, A. F. I., Saski, C. A., Manganaris, G. A., Gasic, K., and Crisosto, C. H. (2018). Genomic sequencing of

Japanese plum (Prunus salicina Lindl.) mutants provides a new model for rosaceae fruit ripening studies. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00021.

Massaglia, S., Borra, D., Peano, C., Sottile, F., and Merlino, V. M. (2019). Consumer preference heterogeneity evaluation in fruit and vegetable purchasing decisions using the best–worst approach. *Foods* 8. doi:10.3390/foods8070266.

Minas, I. S., Crisosto, G. M., Holcroft, D., Vasilakakis, M., and Crisosto, C. H. (2013). Postharvest handling of plums (Prunus salicina Lindl.) at 10°C to save energy and preserve fruit quality using an innovative application system of 1-MCP. *Postharvest Biol. Technol.* 76, 1–9. doi:10.1016/j.postharvbio.2012.08.013.

Minas, I. S., Forcada, C. F., Dangl, G. S., Gradziel, T. M., Dandekar, A. M., and Crisosto, C. H. (2015). Discovery of non-climacteric and suppressed climacteric bud sport mutations originating from a climacteric Japanese plum cultivar (Prunus salicina lindl.). *Front. Plant Sci.* 6, 1–16. doi:10.3389/fpls.2015.00316.

Mnejja, M., Garcia-Mas, J., Audergon, J. M., and Arús, P. (2010). Prunus microsatellite marker transferability across rosaceous crops. *Tree Genet. Genomes* 6, 689–700. doi:10.1007/s11295-010-0284-z.

N'Diaye, A., Haile, J. K., Cory, A. T., Clarke, F. R., Clarke, J. M., Knox, R. E., et al. (2017). Correction: Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map (PLoS ONE (2017) 12:1 (e0170941) DOI: 10.1371/journal.pone.0170941). *PLoS One* 12, e0170941. doi:10.1371/journal.pone.0187178.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* 32, 261–285. doi:10.1080/13102818.2017.1400401.

Netzel, M., Fanning, K., Netzel, G., Zabaras, D., Karagianis, G., Treloar, T., et al. (2012). Urinary excretion of antioxidants in healthy humans following queen garnet plum juice ingestion: A new plum variety rich in antioxidant compounds. *J. Food Biochem.* 36, 159–170. doi:10.1111/j.1745-4514.2010.00522.x.

Okie, W. R. (2008). "Prunus domestica (European plum) and Prunus salicina (Japanese plum)," in *The encyclopedia of fruit & nuts*, eds. J. Janick and R. E. Paull (Cambridge (UK): CABI), 694–705. doi:10.1079/9780851996387.0000.

Okie, W. R., and Hancock, J. F. (2008). "Plums," in *Temperate Fruit Crop Breeding: Germplasm to Genomics*, ed. J. F. Hancock (Dordrecht: Springer Netherlands), 337–358. doi:10.1007/978-1-4020-6907-9_11.

Okie, W. R., and Ramming, D. W. (1999). Plum Breeding Worldwide. *HortTechnology horttech* 9, 162–176. doi:10.21273/HORTTECH.9.2.162.

Okie, W. R., and Weinberger, J. H. (1996). "Plums," in *Fruit breeding, Volume 1: Tree and Tropical Fruits*, eds. J. Janick and J. N. Moore (New York (USA): John Wiley & Sons, Ltd), 559–608.

Payne, A., Holmes, N., Rakyan, V., and Loose, M. (2019). Bulkvis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198. doi:10.1093/bioinformatics/bty841.

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U. R., Stegmeir, T., et al. (2012). Development and Evaluation of a Genome-Wide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry. *PLoS One* 7, e48305. doi:10.1371/journal.pone.0048305.

Poczai, P., Varga, I., Bell, N. E., and Hyvönen, J. (2011). Genetic diversity assessment of bittersweet (Solanum

dulcamara, Solanaceae) germplasm using conserved DNA-derived polymorphism and intron-targeting markers. *Ann. Appl. Biol.* 159, 141–153. doi:https://doi.org/10.1111/j.1744-7348.2011.00482.x.

Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J. P. T., et al. (2013). Advances in plant gene-targeted and functional markers: A review. *Plant Methods* 9, 6. doi:10.1186/1746-4811-9-6.

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., et al. (2007). Phylogeny and classification of Rosaceae. *Plant Syst. Evol.* 266, 5–43. doi:10.1007/s00606-007-0539-9.

Quilot, B., Wu, B. H., Kervella, J., Génard, M., Foulongne, M., and Moreau, K. (2004). QTL analysis of quality traits in an advanced backcross between Prunus persica cultivars and the wild relative species P. davidiana. *Theor. Appl. Genet.* 109, 884–897. doi:10.1007/s00122-004-1703-z.

Rahim, M. A., Busatto, N., and Trainotti, L. (2014). Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 240, 913–929. doi:10.1007/s00425-014-2078-2.

Ramming, D. W., and Cociu, V. (1991). PLUMS (PRUNUS). in *Acta Horticulturae* (International Society for Horticultural Science (ISHS), Leuven, Belgium), 235–290. doi:10.17660/ActaHortic.1991.290.6.

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi:10.1016/j.tplants.2004.12.011.

Ravaglia, D., Espley, R. V., Henry-Kirk, R. A., Andreotti, C., Ziosi, V., Hellens, R. P., et al. (2013). Transcriptional regulation of flavonoid biosynthesis in nectarine (Prunus persica) by a set of R2R3 MYB transcription factors. *BMC Plant Biol.* 13, 68. doi:10.1186/1471-2229-13-68.

Rehder, A. (1940). *Manual of Cultivated Trees and Shrubs Hardy in North America Exclusive of the Subtropical and Warmer Temperate Regions*. 2nd ed. New York (USA): Macmillan.

Romero, C., Vilanova, S., Burgos, L., Martínez-Calvo, J., Vicente, M., Llácer, G., et al. (2004). Analysis of the S-locus structure in Prunus armeniaca L. Identification of S-haplotype specific S-RNase and F-box genes. *Plant Mol. Biol.* 56, 145–157. doi:10.1007/s11103-004-2651-3.

Rubio, J., Sánchez, E., Tricon, D., Montes, C., Eyquard, J. P., Chague, A., et al. (2019). Silencing of one copy of the translation initiation factor eIFiso4G in Japanese plum (Prunus salicina) impacts susceptibility to Plum pox virus (PPV) and small RNA production. *BMC Plant Biol.* 19, 440. doi:10.1186/s12870-019-2047-9.

Ruiz, D., Egea, J., Salazar, J. A., and Campoy, J. A. (2018). Chilling and heat requirements of Japanese plum cultivars for flowering. *Sci. Hortic. (Amsterdam).* 242, 164–169. doi:10.1016/j.scienta.2018.07.014.

Ruiz, D., García-Gómez, B. E., Egea, J., Molina, A., Martínez-Gómez, P., and Campoy, J. A. (2019). Phenotypical characterization and molecular fingerprinting of natural early-flowering mutants in apricot (Prunus armeniaca L.)and Japanese plum (P. salicina Lindl.). *Sci. Hortic. (Amsterdam).* 254, 187–192. doi:10.1016/j.scienta.2019.05.002.

Salazar, J. A., Pacheco, I., Shinya, P., Zapata, P., Silva, C., Aradhya, M., et al. (2017). Genotyping by sequencing for SNP-Based linkage analysis and identification of QTLs linked to fruit quality traits in japanese plum (Prunus salicina lindl.). *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00476.

Salazar, J. A., Pacheco, I., Silva, C., Zapata, P., Shinya, P., Ruiz, D., et al. (2019). Development and applicability of GBS approach for genomic studies in Japanese plum (Prunus salicina Lindl.). *J. Hortic. Sci. Biotechnol.* 94, 284–294. doi:10.1080/14620316.2018.1543559.

Salazar, J. A., Pacheco, I., Zapata, P., Shinya, P., Ruiz, D., Martínez-Gómez, P., et al. (2020). Identification of loci controlling phenology, fruit quality and post-harvest quantitative parameters in Japanese plum (Prunus salicina Lindl.). *Postharvest Biol. Technol.* 169, 111292. doi:10.1016/j.postharvbio.2020.111292.

Sandefur, P., Oraguzie, N., and Peace, C. (2016). A DNA test for routine prediction in breeding of sweet cherry fruit color, Pav-Rf-SSR. *Mol. Breed.* 36, 33. doi:10.1007/s11032-016-0458-y.

Schaid, D. J., Chen, W., and Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19, 491–504. doi:10.1038/s41576-018-0016-z.

Shi, S., Li, J., Sun, J., Yu, J., and Zhou, S. (2013). Phylogeny and Classification of Prunus sensu lato (Rosaceae). *J. Integr. Plant Biol.* 55, 1069–1079. doi:10.1111/jipb.12095.

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., et al. (2017). The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding. *DNA Res.* 24, 499–508. doi:10.1093/dnares/dsx020.

Sinecen, M., Temizkan, R., and Caner, C. (2015). Investigation of the morphological and color changes of damaged green plums during storage time using digital image processing techniques. *Gazi Univ. J. Sci.* 28, 133–139.

Sooriyapathirana, S. S., Khan, A., Sebolt, A. M., Wang, D., Bushakra, J. M., Lin-Wang, K., et al. (2010). QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (Prunus avium L.). *Tree Genet. Genomes* 6, 821–832. doi:10.1007/s11295-010-0294-x.

Spolaore, S., Trainotti, L., and Casadoro, G. (2001). A simple protocol for transient gene expression in ripe fleshy fruit mediated by Agrobacterium. *J. Exp. Bot.* 52, 845–850. doi:10.1093/jexbot/52.357.845.

Starkevič, P., Paukštyte, J., Kazanavičiute, V., Denkovskiene, E., Stanys, V., Bendokas, V., et al. (2015). Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (Prunus avium L.) MYB10 and bHLH genes. *PLoS One* 10, e0126991. doi:10.1371/journal.pone.0126991.

Stegmeir, T., Cai, L., Basundari, F. R. A., Sebolt, A. M., and Iezzoni, A. F. (2015). A DNA test for fruit flesh color in tetraploid sour cherry (Prunus cerasus L.). *Mol. Breed.* 35, 149. doi:10.1007/s11032-015-0337-y.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. *Curr. Opin. Plant Biol.* 4, 447–456. doi:10.1016/S1369-5266(00)00199-0.

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142, 1216–1232. doi:10.1104/pp.106.088104.

Telias, A., Lin-Wang, K., Stevenson, D. E., Cooney, J. M., Hellens, R. P., Allan, A. C., et al. (2011). Apple skin patterning is associated with differential expression of MYb10. *BMC Plant Biol.* 11, 93. doi:10.1186/1471-2229-11-93.

Topp, B. L., Russell, D. M., Neumüller, M., Dalbó, M. A., and Liu, W. (2012). "Plum," in *Fruit Breeding*, eds. M. L. Badenes and D. H. Byrne (Boston, MA: Springer US), 571–621. doi:10.1007/978-1-4419-0763-9_15.

Tuan, P. A., Bai, S., Yaegaki, H., Tamura, T., Hihara, S., Moriguchi, T., et al. (2015). The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. *BMC Plant Biol.* 15, 280. doi:10.1186/s12870-015-0664-5.

Tyson, J. R., O'Neil, N. J., Jain, M., Olsen, H. E., Hieter, P., and Snutch, T. P. (2018). MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. *Genome Res.* 28, 266–274. doi:10.1101/gr.221184.117.

Urtubia, C., Devia, J., Castro, Á., Zamora, P., Aguirre, C., Tapia, E., et al. (2008). Agrobacterium-mediated genetic transformation of Prunus salicina. *Plant Cell Rep.* 27, 1333–1340. doi:10.1007/s00299-008-0559-0.

Valderrama-Soto, D., Salazar, J., Sepúlveda-González, A., Silva-Andrade, C., Gardana, C., Morales, H., et al. (2021). Detection of Quantitative Trait Loci Controlling the Content of Phenolic Compounds in an Asian Plum (Prunus salicina L.) F1 Population. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.679059.

Vanderzande, S., Piaskowski, J. L., Luo, F., Edge Garza, D. A., Klipfel, J., Schaller, A., et al. (2018). Crossing the Finish Line: How to Develop Diagnostic DNA Tests as Breeding Tools after QTL Discovery. *J. Hortic.* 05, 1–6. doi:10.4172/2376-0354.1000228.

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., et al. (2012). Development and evaluation of a 9k snp array for peach by internationally coordinated snp detection and validation in breeding germplasm. *PLoS One* 7, e35668. doi:10.1371/journal.pone.0035668.

Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarani, G., Vendramin, E., et al. (2017). The Peach v2.0 release: High-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* 18, 225. doi:10.1186/s12864-017-3606-9.

Verde, I., Quarta, R., Cedrola, C., and Dettori, M. T. (2002). QTL analysis of agronomic traits in a BC1 peach population. in *Acta Horticulturae* (International Society for Horticultural Science (ISHS), Leuven, Belgium), 291–297. doi:10.17660/ActaHortic.2002.592.41.

Vieira, E. A., Nodari, R. O., Dantas, A. C. M., Ducroquet, J. P. H. J., Dalbó, M., and Borges, C. V. (2005). Genetic mapping of Japanese plum. *Crop. Breed. Appl. Biotechnol.* 5, 29–37. doi:10.12702/1984-7033.v05n01a04.

Wei, X., Shen, F., Zhang, Q., Liu, N., Zhang, Y., Xu, M., et al. (2021). Genetic diversity analysis of Chinese plum (Prunus salicina L.) based on whole-genome resequencing. *Tree Genet. Genomes* 17, 26. doi:10.1007/s11295-021-01506-x.

Weinberger, J. H., and Thompson, L. A. (1962). Inheritance of certain fruit and leaf characters in Japanese plums. *Proc Am Soc Hortic Sci* 81, 172–179.

Werner, D. J., Creller, M. A., and Chaparro, J. X. (1998). Inheritance of the blood-flesh trait in peach. *HortScience* 33, 1243–1246. doi:10.21273/hortsci.33.7.1243.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* 5, 218–223. doi:10.1016/S1369-5266(02)00256-X.

Wu, X., Gong, Q., Ni, X., Zhou, Y., and Gao, Z. (2017). UFGT: The key enzyme associated with the petals variegation in japanese apricot. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00108.

Wünsch, A., and Hormaza, J. I. (2004). S-allele identification by PCR analysis in sweet cherry cultivars. *Plant Breed.* 123, 327–331. doi:10.1111/j.1439-0523.2004.00984.x.

Xi, W., Feng, J., Liu, Y., Zhang, S., and Zhao, G. (2019). The R2R3-MYB transcription factor PaMYB10 is involved in anthocyanin biosynthesis in apricots and determines red blushed skin. *BMC Plant Biol.* 19, 287.

doi:10.1186/s12870-019-1898-4.

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., et al. (2017). Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Mol. Biol. Evol.* 34, 262–281. doi:10.1093/molbev/msw242.

Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48, 391–407. doi:10.2135/cropsci2007.04.0191.

Yoshida, K., Mori, M., and Kondo, T. (2009). Blue flower color development by anthocyanins: From chemical structure to cell physiology. *Nat. Prod. Rep.* 26, 884–915. doi:10.1039/b800165k.

Yoshida, M. (1987). "The origin of fruits. 2: Plums," in *Fruit Japan volume 42* (Japan), 49–53.

Yu, X. M., Rizwan, H. M., Li, P., Luo, S. X., Sherameti, I., Wu, W. F., et al. (2020). Comparative studies on the physiochemical properties, phenolic compounds and antioxidant activities in 13 Japanese plum cultivars grown in the subtropical region of China. *Appl. Ecol. Environ. Res.* 18, 3147–3159. doi:10.15666/aeer/1802_31473159.

Zhang, G., Sebolt, A. M., Sooriyapathirana, S. S., Wang, D., Bink, M. C., Olmstead, J. W., et al. (2010). Fruit size QTL analysis of an F1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genet. Genomes* 6, 25–36. doi:10.1007/s11295-009-0225-x.

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* 10, 1494. doi:10.1038/s41467-019-09518-x.

Zhang, Q. ping, Wei, X., Liu, N., Zhang, Y. ping, Xu, M., Zhang, Y. jun, et al. (2020). Construction of an SNP-based high-density genetic map for Japanese plum in a Chinese population using specific length fragment sequencing. *Tree Genet. Genomes* 16, 18. doi:10.1007/s11295-019-1385-y.

Zhao, Y., Dong, W., Zhu, Y., Allan, A. C., Lin-Wang, K., and Xu, C. (2020). PpGST1, an anthocyanin-related glutathione S-transferase gene, is essential for fruit coloration in peach. *Plant Biotechnol. J.* 18, 1284–1295. doi:10.1111/pbi.13291.

Zhebentyayeva, T., Shankar, V., Scorza, R., Callahan, A., Ravelonandro, M., Castro, S., et al. (2019). Genetic characterization of worldwide Prunus domestica (plum) germplasm using sequence-based genotyping. *Hortic. Res.* 6, 12. doi:10.1038/s41438-018-0090-6.

Zhou, H., Lin-Wang, K., Wang, H., Gu, C., Dare, A. P., Espley, R. V., et al. (2015). Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* 82, 105–121. doi:10.1111/tpj.12792.

Zhou, Y., Zhou, H., Lin-Wang, K., Vimolmangkang, S., Espley, R. V., Wang, L., et al. (2014). Transcriptome analysis and transient transformation suggest an ancient duplicated MYB transcription factor as a candidate gene for leaf red coloration in peach. *BMC Plant Biol.* 14, 388. doi:10.1186/s12870-014-0388-y.