Universitat Autònoma de Barcelona

Understanding and Exploiting Almond Genetic Diversity for Peach Breeding: Development of a Peach-Almond Introgression Line collection and Fine Mapping of Key Fruit-Related Genes

Naveen Kalluri

PhD Thesis

Understanding and Exploiting Almond Genetic Diversity for Peach Breeding: Development of a Peach-Almond Introgression Line collection and Fine Mapping of Key Fruit-Related Genes

Dissertation presented by Naveen Kalluri for the degree of Doctor in Genetics and Microbiology by Universitat Autónoma de Barcelona (UAB)

Faculty of Biosciences

This work was performed at Centre for Research in Agricultural Genomics (CRAG), Bellaterra.

Thesis directors	***Thesis Tutor***	***PhD candidate***
Dr. Pere Arús	*Francisco José Rodríguez-Trelles*	*Naveen Kalluri*
Dr. Iban Eduardo		

Barcelona, November 2021

Acknowledgements

This Thesis would not have been possible without the support and guidance of many people. I would like to express my gratitude to everyone that has been part of this journey.

Firstly, I would like to thank my supervisors. Dr. Pere Arús, it was such an honor to work with you. I am sure the words will fall short to describe about you. I felt like I learned something new every time I had a conversation with you. All the meetings that I had during my PhD were brainstorming and informative. I would forever be grateful for this opportunity. Dr. Iban Eduardo, a passionate researcher, on top of that even a great personality. I learned so many things from you in and out of research. Thanks for teaching me all the field work. I really appreciate your patience and the freedom you gave me to work and express myself. I definitely wouldn't have chosen better supervisors myself even if I had an opportunity to do so. Thank you both.

The Principal Investigators, Dr. Amparo Monfort, Dr. Maria José Aranzana and Dr. Werner Howad. Thank you for the help you provided whenever needed.

To Elena Del Blanco, the lab technician, who taught me all the in vitro work and treated me more like a family rather than a colleague. Also, the only person till date with whom I spoke and learnt a lot of Spanish. ¡Muchas gracias por todo!

The other lab technicians, Àngel Montejo, Fuensata Garcia, Esteve Collell and Alba Perez. Thank you for your help in the laboratory.

Dr. Konstantinos Alexiou, Thanks a lot for the constant bio-informatics support all along my PhD.

Joana Ribes, Thanks for all the genotyping work at the genomics platform.

My former PhD mate, Dr. Octávio Manuel Ribeiro Serra, Thank you for introducing me to the project and all the help provided at the beginning of my PhD.

My fellow PhDs, Arnau Fiol, Nuria Real, Neus Marimon, Laura Zingaretti, Lara Pereira, Pol Rey, Yu Zhang, Rong Zhang, Yang Zhang, Miguel Santo Domingo, Carlos Mayobre, Nathalia Zaracho, Lorena Areco, Christian Dujak, Federico Jurado, Felipe Pérez de los Cobos. Post docs, Dr. Jason Argyris, Dr. Mourad Mnejjia, Dr. Sílvia Fornalé, Dr. Valentino Ruggieri, Dr. Andrea Giordano, Dr. Liu Bin, Dr. Clementine Le Roux, Dr. Elli Koskela and Dr. Beatriz Ester Garcia. Thank you all for the great company.

The Bachelor and Master students, Bianca Estefani Vigil Santillan, Anna Maria Albert, Damian Dasilva Martinez, Lisa Pinochet, Cristina Peralta, Carolina Ribeiro and Enric Busquets. Thank you guys for all the help offered during the course of my PhD.

To the whole Plant Genomics Department, Thank you for such a wonderful work atmosphere.

CRAG greenhouse team, Thank you very much for taking care of the plants whenever needed.

A big thanks to the plant facility and greenhouse team at IRTA Caldes de Montbui, IRTA Gimenells and Mollerussa field stations for keeping a check on plants at all times.

Lastly and most importantly, I would like to thank my father, mother, brother and sister for their constant support and belief in me all the time. Also, I would like to dedicate this thesis to my parents.

This PhD thesis was carried out with the grant provided by BES-2016-077104 funded by MCIN/AEI/ 10.13039/501100011033 and by “ESF Investing in your future”.

Table of contents

Abbreviations

A – Additive

Alf – Almond fruit type

BC – Backcross population

BFT – Beginning of flowering time

BLF – Beginning of leaf fall

BS – Beginning of shooting

bp – Base pair

ºC – Celsius degrees

CC – Chlorophyll content

cM – Centimorgan

CRAG – Centre for Research in Agricultural Genomics

CSSL – Chromosome segment substitution line

CTAB – Cetyltrimethylammonium bromide

D – Dominant

DAF – Days after flowering

DBF2 – Dominant blood flesh

DNA – Deoxyribonucleic acid

dNTP – Deoxynucleoside triphosphate

E – 'Earlygold' map from the ('Texas' × 'Earlygold') × 'Earlygold'

E×E – 'Earlygold' F2 population

ELF – End of leaf fall

FDP – Fruit development period

FF – Fruit firmness

FP – Fruit production

FT – Flowering time

FW - Fruit weight

g – Grams

G – Linkage group

GA – Gene action

Hz – Hertz

IL – Introgression line

InDel – Insertion and deletion

IRTA – Institute of Agrifood Research and Technology

ISC – Intensity of red skin color

Jui – Juiciness

Kb – Kilo bases

LBL – Leaf blade length

LBW – Leaf blade width

LCS – Leaf color at senescence

l – Liter

LL – Leaf length

LOD – Logarithm of the odds

LP – Leaf perimeter

LS – Leaf surface

LSH – Leaf shape

LW – Leaf dry weight

M – Molarity

m - Meters

MAI – Marker assisted introgression

MAS – Marker assisted selection

Mb – Mega bases

MD – Maturity date

mg – Milli gram

min – Minutes

mm - Millimeter

mM – Millimolar

ml – Milliliter

ng – Nanogram

NIL – Near-isogenic line

NOR – Non ripening

O - Over dominant

OP – Open pollination

PAIL – Peach-almond introgression line

PCR – Polymerase chain reaction

PL – Petiole length

prIL – Pre-introgression line

QTL – Quantitative trait locus

QTLs – Quantitative trait loci

R^2 – Percentage of explained phenotypic variance

RIL – Recombinant inbred line

RNA – Ribonucleic acid

RPM – Revolutions per minute

SD – Standard deviation

s – Seconds

Sh - Showy

SNP – Single nucleotide polymorphism

SR – Slow ripening

SSC – Soluble solid content

SSR – Simple sequence repeat

T×E – 'Texas' × 'Earlygold' population

T1E - ('Texas' × 'Earlygold') × 'Earlygold' population

T_a – Annealing temperature

Taq – Thermus aquaticus

TA - Titratable acidity

VCF – Variant Call Format

µl – Microliter

µM – Micromolar

List of Tables

List of Figures

Summary

Peach (*Prunus persica*), one of the most important temperate fruit crops, has low levels of genetic variability. One of the ways to improve its diversity is by introgression of novel alleles from a closely related wild or cultivated *Prunus* species. For this study, almond (*Prunus dulcis*) was chosen, and the results obtained are based on an initial cross performed in the late 1970's as part of a rootstock breeding program using almond cultivar 'Texas' as female parent and peach cultivar 'Earlygold' as male parent. Later (2006), a large backcross one (BC1) generation to 'Earlygold' was produced, and a marker-based breeding strategy was developed to obtain plants with one or a few almond chromosomal fragments in the peach background only with two BC generations. In this thesis, based on the previous production of a set of BC2 lines with 2-3 almond introgressions, we have developed a complete introgression line (IL) collection, consisting of 67 lines that have a single almond fragment in the peach background; 39 ILs with the almond introgression in heterozygosis, covering 99% of the almond genome, and 28 with homozygous almond fragments, with 83% almond coverage. These collections were analyzed for some of the major genes that were expected to segregate and for some of the fruit-related QTLs that had been detected earlier with 'Texas' × 'Earlygold'-based progenies. Due to the partly heterozygous nature of our recurrent parent, 'Earlygold', which is expected to segregate in the IL collection, a QTL analysis was performed using its F2 progeny in a large set of 24 traits, where a total of 26 QTLs were identified. Only a major QTL for maturity date on chromosome 4, co-locating with other fruit-related characters and leaf color at senescence QTLs was considered as of potentially concern for IL analysis. The final part of the thesis involves the fine mapping of three major genes detected in the almond x peach populations: two that explain an important part of the differences between a peach and an almond fruit (*Alf*/*alf* that determines the thick and ripening mesocarp of peach and *Jui*/*jui* that determines the juciness of this mesocarp), and a third gene (*DBF2*/*dbf2*) that produces a red-fleshed fruit conferred by the almond allele. This study involved large-scale recombinant screening in various segregating populations and resulted in the location of *Alf*, *Jui* and *DBF2* genes in DNA fragments as short as 10-392 kb including 4 to 95 positional candidate genes. Most probable candidates were identified for *Alf* (*Prupe.4G187100* and *Prupe.4G188700*) and *DBF2* (*Prupe.1G519800*) genes from the analysis of variants and the prediction of their effects obtained from the study of 'Texas' and 'Earlygold' DNA resequence data and gene expression

analysis, providing a solid basis for their future cloning as responsible genes for the observed phenotypes.

Resumen

El melocotonero (*Prunus persica*), uno de los frutales más importantes, tiene un nivel de variabilidad genética bajo. Una de las formas de mejorar su diversidad es mediante la introducción de nuevos alelos de especies silvestres o cultivadas próximas del género *Prunus*. Para este estudio, se usó el almendro (*P. dulcis*) como donante de variabilidad, tomando como punto de partida un híbrido realizado a fines de la década de 1970 como parte de un programa de mejora de portainjertos con el almendro 'Texas' como parental femenino y el melocotonero 'Earlygold' como parental masculino. Más tarde (2006), se generó un numeroso retrocruzamiento (BC1) de este híbrido con 'Earlygold', y se desarrolló una estrategia de mejora basada en marcadores para obtener plantas con uno o unos pocos fragmentos cromosómicos de almendro en el fondo genético del melocotonero solo en dos generaciones de BC. En esta tesis, basándonos en la disponibilidad de un juego de líneas BC2 con 2-3 introgresiones de almendro, hemos desarrollado una colección completa de líneas de introgresión (ILs), formada por 67 ILs con un único fragmento de almendro en el fondo genético del melocotonero: 39 con la introgresión de almendro en heterocigosis, cubriendo el 99% del genoma del almendro, y 28 con fragmentos de almendro homocigotos, con 83% de cobertura. Estas colecciones se analizaron para algunos de los genes mayores que se espera que estén segregando en las ILs y para algunos de los QTLs de fruto previamente detectados en descendencias basadas en 'Texas' × 'Earlygold'. Debido a la naturaleza parcialmente heterocigótica del parental recurrente, 'Earlygold', que se espera que segregue en la colección de ILs, se realizó un análisis de 24 caracteres en su F2, identificando un total de 26 QTLs. Un QTL mayor para época de maduración situado en el cromosoma 4, que co-localizaba con otros QTLs de caracteres relacionados con la fruta y la hoja se consideró la única región potencialmente problemática para el análisis de las ILs. La parte final de la tesis consiste en el mapeo fino de tres genes mayores detectados en las poblaciones de almendro × melocotonero: dos que explican una parte importante de las diferencias entre el fruto del melocotonero y el almendro (*Alf*/*alf* que determina el mesocarpio grueso y que madura del melocotón y *Jui*/*jui* que determina la jugosidad de este mesocarpio), y un tercer gen, *DBF2*/*dbf2*, que produce un fruto de carne roja conferido por el alelo del almendro. Este estudio implicó la búsqueda a gran escala de recombinantes en varias poblaciones segregantes y resultó en la ubicación de los genes *Alf*, *Jui* y *DBF2* en fragmentos de ADN cortos, de 10-392 kb, que contenían entre 4 y 95 genes candidatos posicionales. Los candidatos más probables para *Alf* (*Prupe.4G187100* y

Prupe.4G188700) y *DBF2* (*Prupe.1G519800*) fueron identificados usando análisis de variantes y la predicción de sus efectos usando datos de resecuencia de 'Texas' y 'Earlygold' y de expresión génica, proporcionando una base sólida para su clonación futura como genes responsables de los fenotipos observados.

Resum

El presseguer (*Prunus persica*), un dels arbres fruiters més importants, té baixos nivells de variabilitat genètica. Una manera d'augmentar la seva diversitat és mitjançant la introducció d'al·lels nous d'altres espècies properes del gènere *Prunus*, silvestres o cultivades. Per a aquest estudi, es va triar l'ametller (*P. dulcis*), i els resultats obtinguts es basen en un encreuament inicial realitzat a finals de la dècada de 1970 com a part d'un programa de millora de portaempelts utilitzant l'ametller 'Texas' com a parental femení i el presseguer 'Earlygold' com a parental masculí. Més tard (2006), es va obtenir una nombrosa primera generació de retroencreuament (BC1) amb 'Earlygold', i es va desenvolupar una estratègia de millora basada en marcadors per tal d'obtenir plantes amb un o uns pocs fragments cromosòmics d'ametller en el fons de presseguer només en dues generacions de BC. En aquesta tesi, basada en la disponibilitat prèvia d'un conjunt de línies BC2 amb 2-3 introgressions d'ametller, hem desenvolupat una col·lecció completa de línies d'introgressió (IL), formada per 67 línies que tenen un sol fragment d'ametller en el fons de presseguer; 39 ILs amb la introgressió d'ametller en heterozigosi, que abasta el 99% del genoma de l'ametller, i 28 amb fragments d'ametller en homozigosi, amb un 83% de cobertura del genoma de l'ametller. Aquestes col·leccions van ser analitzades per a alguns dels gens majors que s'esperava que segreguessin i per a alguns dels QTLs relacionats amb caràcters del fruit detectats anteriorment en progènies de 'Texas' × 'Earlygold'. A causa de la naturalesa parcialment heterozigòtica del parental recurrent, 'Earlygold', que s'espera que se segregui en la col·lecció d'ILs, es va realitzar una anàlisi de QTLs en la seva F2 per 24 caràcters, on es van identificar un total de 26 QTLs. Només un QTL, el que determina a la data de maduresa en el cromosoma 4 i que co-localitza amb altres QTLs per caràcters del fruit i del color de les fulles, es va considerar com potencialment problemàtic per a l'anàlisi de les ILs. La part final de la tesi consisteix en el mapatge fi de tres gens majors detectats en les poblacions de presseguer × ametller: dos que expliquen una part important de les diferències entre el fruit d'un préssec i una ametlla (*Alf*/*alf* que determina el mesocarpi gruixut i que madura del presseguer i *Jui*/*jui* que fa que aquest mesocarpi sigui sucós), i un tercer gen, *DBF2*/*dbf2*, que produeix un fruit de carn vermella conferit per l'al·lel de l'ametller. Aquest estudi va implicar una cerca a gran escala de recombinants en vàries poblacions segregants i va permetre localitzar *Alf*, *Jui* i *DBF2* en fragments d'ADN curts, entre 10-392 kb, incloent de 4 a 95 gens candidats posicionals. Els candidats més probables per als gen *Alf* (*Prupe.4G187100* i *Prupe.4G188700*) i *DBF2* (*Prupe.1G519800*) es van trobar a partir de

l'anàlisi de variants i la predicció dels seus efectes obtinguts de l'estudi de les dades de reseqüència d'ADN i d'anàlisi d'expressió de 'Texas', 'Earlygold' i alguns dels seus descendents amb diferents fenotips, proporcionant una informació útil per a la seva futura clonació com a gens responsables dels fenotips observats.

1. General Introduction

1.1 Peach

1.1.1 Taxonomy and classification

Peach (*Prunus persica* (L.) Batsch) is an economically important fruit tree species that belongs to *Rosaceae* family, which comprises approximately 91 genera and 4,828 species (Christenhusz and Byng 2016). The *Rosaceae* family is classified into three sub families: *Rosoideae*, *Dryadoideae* and *Spiraeoideae* (Potter et al. 2007). Later, *Spiraeoideae* was named to *Amygdaloideae* based on the changes in International Code of Nomenclature for Algae, Fungi and Plants (Mc Neill et al. 2012). The *Prunus* genus is classified under the subfamily *Amygdaloideae*. It has over 200 species of flowering trees and shrubs, some of them being economically valuable stone fruits for their fleshy mesocarp and nut properties (Chin et al. 2014). The most widely accepted classification divides the *Prunus* genus into five subgenera: *Amygdalus* (peaches and almonds), *Prunus* (plums and apricot), *Cerasus* (cherries), *Padus* (bird cherries) and *Laurocerasus* (Laurel cherries) (Rehder 1940; Chin et al. 2014) (Figure 1.1). *P. persica* is sexually compatible with its wild realtives *P. davidiana* (Carr.) Franch, *P. mira* Koehne, *P. kansuensis* Rehd. and *P. ferganensis* (Kost. & Rjab) Kov. & Kost and with almond and its wild relatives (Yazbek and Oh 2013). The hybrids of almond and peach are fertile (Armstrong 1957) and employed as rootstocks for peach and almond. When peach is crossed with its closely related stone fruit crops (apricot, plum and cherry) sometimes results in successful hybrids which are largely sterile (Scorza and Okie 1991). The systematic classification is as follows:

Kingdom: *Plantae*

Subkingdom: *Tracheobionta*

Division: *Magnoliophyta*

Class: *Magnoliopsida*

Subclass: *Rosidae*

Order: *Rosales*

Family: *Rosaceae*

Subfamily: *Amygdaloideae*

Genus: *Prunus*

Subgenus: *Amygdalus*

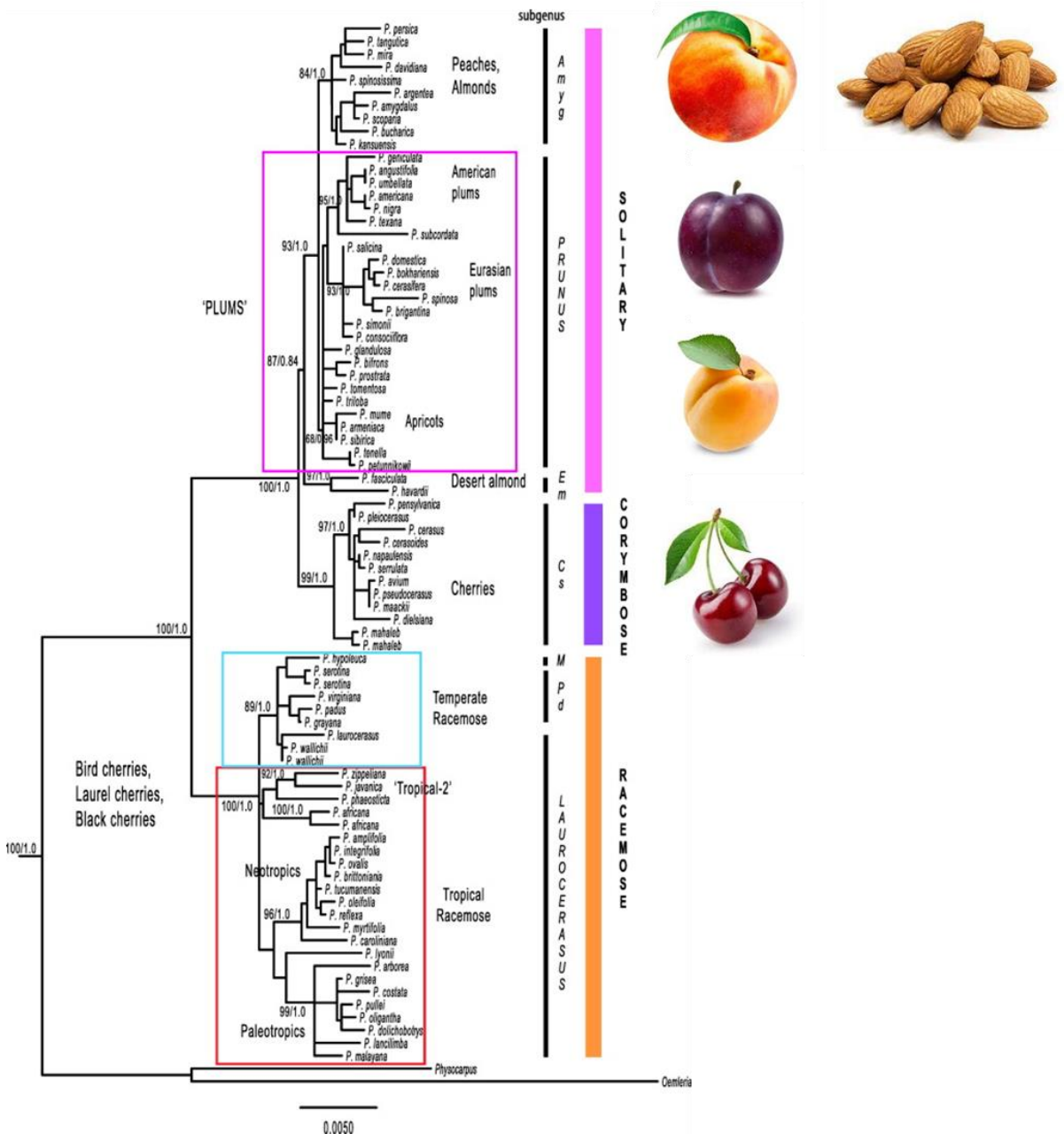


Figure 1.1: Maximum likelihood tree of the four-gene concatenated plastid DNA sequences for phylogeny of *Prunus.* Numbers above or below branches indicate maximum likelihood bootstrap support/Bayesian posterior probabilities. Abbreviations for subgenus: *Amyg*, *Amygdalus*; *Em*, *Emplectocladus*; *Cs, Cerasus*; *Lc*, *Laurocerasus*; *Pd*, *Padus*; *M*, *Maddenia* (Chin et al. 2014).

1.1.2 Geographic origin and distribution

Peach was originated in China and domesticated as early as 4000-5000 years ago (Faust and Timon 1995). Chinese have identified three groups of peaches based on where they were grown, and the types of peaches were different in each area (Li 1984; Wang 1985). A Southern group along the Yangtze River, a Northern group along the Yellow River and the third group

is located in the arid Northwest of China. Chinese still use the ancestral peach 'Wolda', also called as wild peach 'Yitao' or hairy peach 'Maotao' as rootstock for improved cultivars (Li 1984). Most of the current cultivars are largely native to China and hold a wider genetic diversity than the cultivated transect.

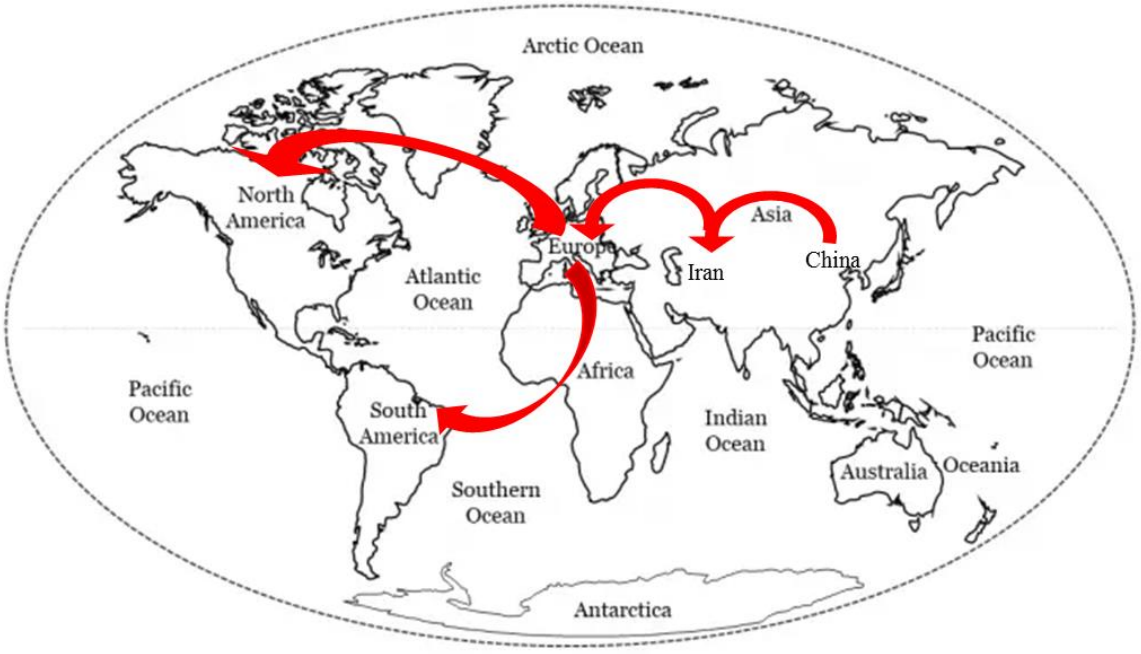


Figure 1.2: Peach origin and its geographic distribution across the world.

Peaches spread westwards from China to Persia (Iran) via the Silk Road, from where it was called as persica fruit (the scientific name *P. persica*). Peaches might have arrived in Japan at the same time as Persia, but the exact origin is yet unknown (Yamamoto et al. 2003). From Persia peaches entered to Greece in 2500 B.P. and to Rome 500 years later. It was dispersed to whole Europe by Romans and Greeks. Peaches were found in England by the 14th century (Bunyard 1938). Peaches entered to North and South America in the early 16th century through Spain and Portugal (Faust and Timon 1995) (Figure 1.2).

1.1.3 Peach fruit production

Peach is the third most important temperate tree fruit species next to apples and pears and the most widely grown species of the *Prunus* genus. Peaches are mainly grown in temperate zones, between latitudes 30° and 45° (Hancock et al. 2008) and need 100 to 1000 hours of chilling to break the bud dormancy. The annual worldwide peach production is estimated to be nearly 20

million metric tons. In 2019, peach was the 13th most produced fruit world-wide with a production of 25.74 million metric tons (Figure 1.3). The peach production has increased by more than 20% in the last 10 years, with 20.5 million metric tons in 2010 to 25.7 million metric tons in 2019 (Figure 1.4). China is the leading producer of world total peach production, which accounts for more than 50% of the total. In 2019, Asia had 70% of peach production, Europe accounted for 20% and rest of the world 10%. The countries with most production after China (15,841,928 tons) are Spain (1,545,610 tons) and Italy (1,224,940 tons) (Figure 1.4). Spain, with 1.5 million metric tons in an area of 77,000 hectares is the highest peach producer of Europe. The regions with most peaches produced were Catalonia (301.5 thousand metric tons), Aragon (254 thousand metric tons) and Murcia (235.4 thousand metric tons). There is an increase of 20% peach production in Spain in the last 10 years, from 1,286,456 tons in 2010 to 1,545,610 tons in 2019.

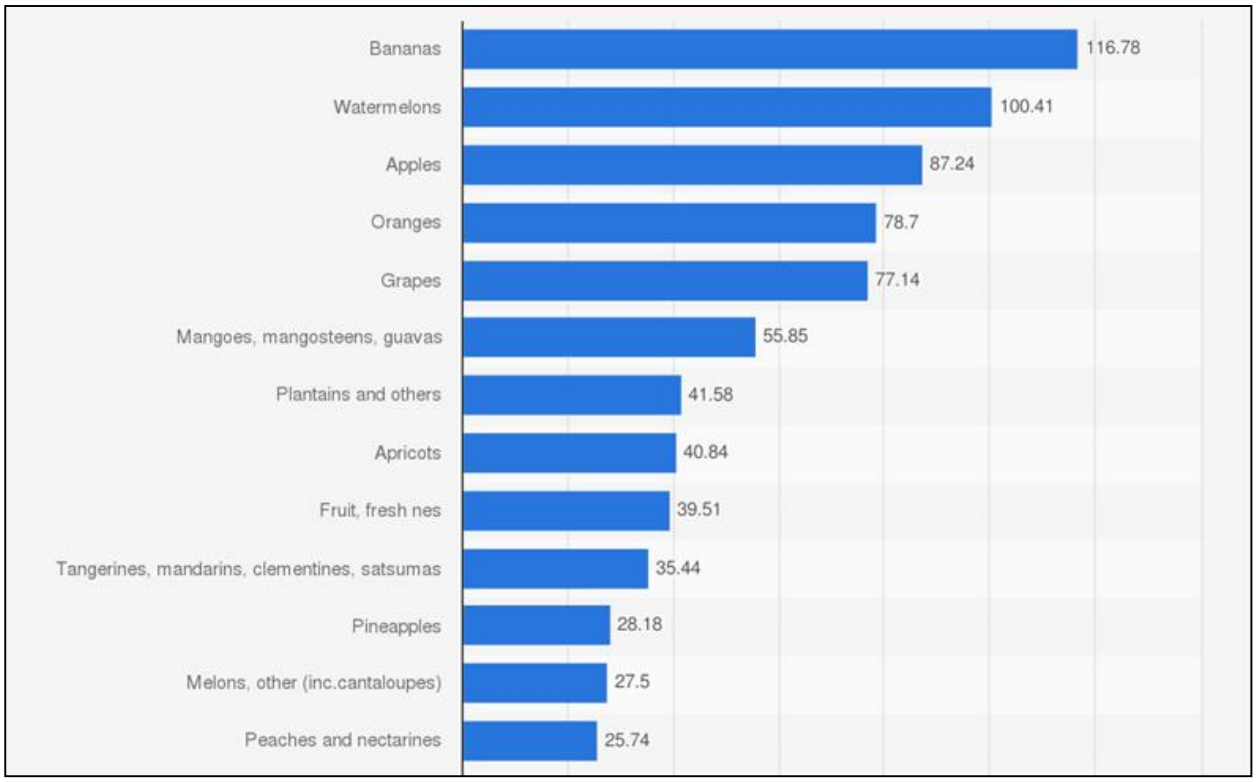


Figure 1.3: Topmost fruits production (in million metric tons) world-wide in 2019 (FAOSTAT 2019).

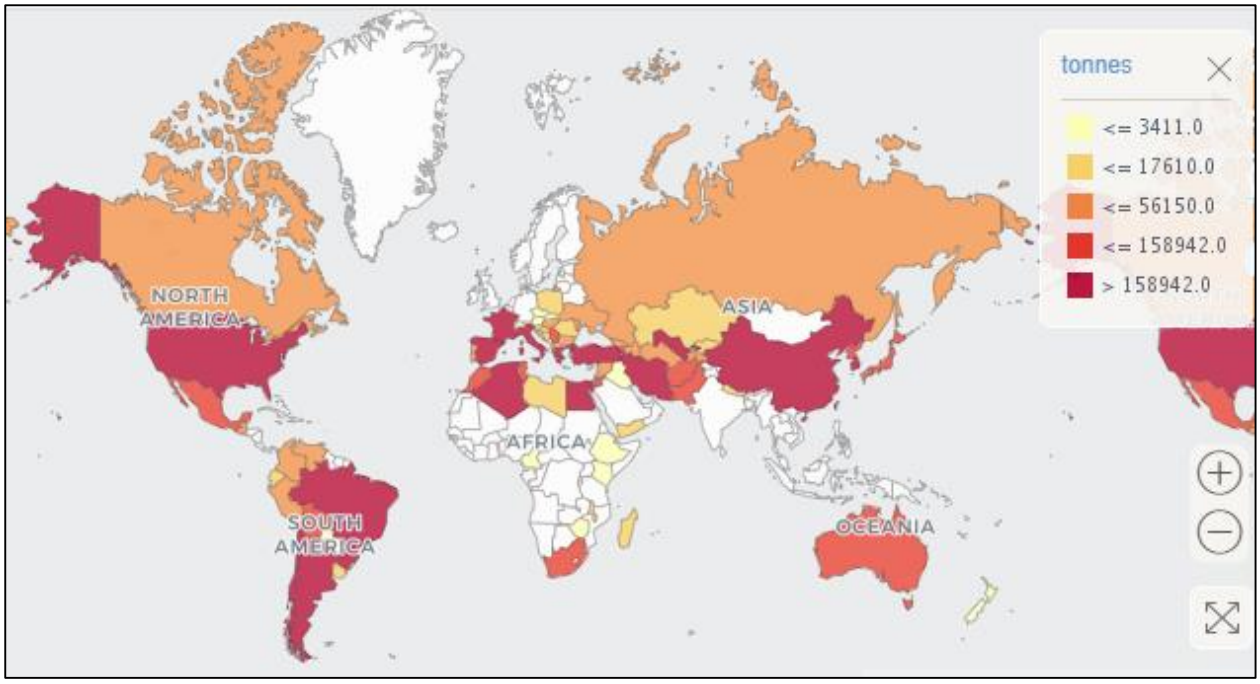


Figure 1.4: Peach fruit production world-wide in 2019 (FAOSTAT 2019).

1.1.4 Genetic diversity

The peach cultivars have great morphological variability, fruit shape round to flat; colors from yellow, white to red; flesh can be melting, non-melting or slow melting and free stone or cling stone type (Hancock et al. 2008). Other variations include differences in fruit size, flavor, flower type, double flower, anther color, tree architecture, disease and pest resistance and phenological characters such as bloom and maturity time (Dirlewanger et al. 2012). Many of these traits have a simple Mendelian inheritance and a much larger number of major genes have been found in peaches than in the rest of the *Prunus* crops (Aranzana et al. 2019). Possible explanations for these differences are that peaches have been genetically studied in more detail than any of the other *Prunus* due to their greater economic value, and that being peaches self-compatible, it is easier to recover rare alleles in homozygosis than in the other self-incompatible stone fruit.

In contrast, cultivated peach has a lower level of variability compared to the rest of the *Prunus* crops and the main reason for it is its selfing mating system (Arús et al. 2012). Selection during domestication or modern breeding has been more efficient at narrowing peach genetic variability due to inbreeding resulting from self-pollination, and this can be seen when examining the comparative variability of genetic markers of all kinds between peach and other related species (Byrne 1990; Mnejja et al. 2010; Velasco et al. 2016; Cao et al. 2014). Low

variability results in limited availability of genes of breeding interest which results in stagnation or low breeding progress for certain characters. One way to improve this situation would be to find novel sources of variability and incorporate them into peach. A possible source of novel variability comes from China, the center of origin for peach, which holds a more genetically diverse germplasm that can be used to breed cultivars with improved fruit quality, resistance to pests and diseases and climate change adaptation (Yoon et al. 2006; Li et al. 2013; Cao et al. 2014). Another source comes from the wild and cultivated peach relatives that contain an immense reservoir of new alleles, some of which potentially useful for peach breeding. Almond, one of the most variable *Prunus* (Byrne 1990; Mnejja et al. 2010), may prove a source of rusticity and biotic and abiotic resistance among other characters including fruit quality (Donoso et al. 2016).

1.15 Peach breeding

Peach modern breeding begun in the early 20th century. In 1850, 'Chinese Cling', a cultivar imported to North America from China, became one of the important cultivars in USA. The varieties derived from 'Chinese Cling' and other cultivars originally coming from Europe became the fresh market cultivars in North America (Byrne et al. 2012). The first breeding program in North America started in 1895, in Geneva, New York. After this, in Iowa (1905), Illinois (1907), California (1907), Ontario (1911), New Jersey (1914), Virginia (1914), Massachusetts (1918), and New Hampshire (1918). Other states also followed similar trend and began their own breeding programs (Okie et al. 2008). Private breeding programs started in California in the early 1930s to improve the local peaches and nectarines with melting flesh for the fresh market (Faust and Timon 1995). In Latin America, the breeding programs were initiated in Southern Brazil (1950) to develop non-melting and melting peaches and in Mexico (1980) to develop melting peaches (Byrne 2005), followed by Chile, Uruguay and Argentina. In Europe, the breeding program first began in Italy in 1920s, later in 1960s in France and subsequently in Spain, Romania, Serbia, Greece, Bulgaria, Ukraine and Poland. Most of the European cultivars developed were similar to the North American cultivars, as the cultivars used in the work were obtained from USA. In Asia, the breeding program started 60 years ago in Japan, followed by China (1970), Korea, India and Thailand. South Africa and Australia also started the breeding programs with the emphasis on fresh market peaches (Byrne et al. 2012). The 20th century has been a golden age for peach breeding, with more than 1000 new varieties being released (Sansavini et al. 2006).

In the recent times, public breeding programs have started to decline with an increase in the private breeding programs, which releases the majority of peaches and nectarines in USA, France and Spain (Byrne et al. 2012). The most noticeable advancements in the peach breeding have been, enhanced fruit size, higher yield, increase in flesh to stone ratio, and the expansion of the harvest period (Sansavini et al. 2006). However, breeders are still interested in other traits, specifically related to commercial and economic importance, such as adaptability to different environments, particularly reducing the chilling requirements to make possible growth in subtropical climate conditions, developing new varieties with a prolonged harvesting window and improving fruit quality related to flavor and aroma. One of the most important objectives of the breeding programs is enhancing the shelf life of peach fruits to be able to store them for longer periods of time and ship them to longer distances. Other traits of interest include novel fruit types, such as blood-flesh peaches, tree architecture (to adapt to more productive pruning strategies) and resistance to pests and diseases, such as sharka (Plum pox virus - PPV), powdery mildew (*Podosphaera pannosa*), brown rot (*Monilinia fruticola*), leaf curl (*Taphrina deformans*), *Xanthomonas* spp. and green aphid (*Myzus persicae*) (Byrne 2005; Sansavini et al. 2006). Even with limited variability, the current commercial peach gene pool is still enough to produce many new and improved cultivars yearly. However, most of the desired traits (particularly biotic and abiotic stress resistance, shelf life and quality) do not exist in the elite breeding pool, and crosses with exotic materials including closely related *Prunus* species are needed (Donoso et al.2016).

1.2 Genomic and molecular tools

1.2.1 The *Prunus* genome

Peach has been one of the model species for the *Rosaceae* family along with other important species like apple and strawberry (Shulaev et al. 2008). It is used for genomic studies due to its economic importance, diploid (2n = *2x = 16*) nature, self-compatibility and small genome size of 265 Mb. Peach also has a short juvenile period of 2-3 years compared to other fruit tree species and most woody perennial crops. The availability of molecular markers, linkage maps and the sequenced genome makes it the best genetically characterized species among the *Prunus* and other stone fruit trees. The first version of the peach genome (Peach v1.0) was obtained by the International Peach Genome Initiative (IPGI) (Verde et al. 2013) using a double haploid individual obtained from the peach cultivar ‘Lovell’. Sanger sequencing methodology

was used to assemble a genome sequence of 227 Mb, identifying eight pseudomolecules that represented the eight chromosomes of peach, covering 96% of its genome, and not revealing any recent whole-genome duplications. Four years later, a second version of the peach genome (Peach v2.0) was released (Verde et al. 2017) that improved the peach assembly. This reduced the gaps, corrected some errors in the assembly of the pseudomolecules, improved accuracy and increased the genome coverage from 96% to 99%. High-throughput linkage mapping and deep resequencing approaches were followed to improve the quality of the initial assembly. The available high quality peach genome v2.0 became a useful tool to study the genetic diversity and to understand the genetics of peach and other *Prunus* species.

Two whole-genome sequencies of almond have been recently produced, both using heterozygous genotypes, one of the French cultivar 'Lauranne', obtained basically with PacBio long DNA reads (Sánchez-Pérez et al. 2019), and the other with Illumina sequencing combined with Oxford Nanopore long reads in the cultivar 'Texas' (Alioto et al. 2020). Both covered a similar genome size, and one of them ('Texas') was compared with the peach reference genome sequence showing the expected high synteny. Whole-genome sequences of other *Prunus* species have been released, most in the last five years, and 26 of them are available at the Genome Database for Rosaceae (Jung et al. 2019) from peach and its relatives (7), almond (2), apricot (8), cherry (5), and plum (4), which are an invaluable resource for current and future genetics and evolutionary studies in this important genus.

1.2.2 Molecular markers

Molecular markers are based on DNA polymorphisms, base-pair substitutions or insertions/deletions, that occur in the genome of living organisms. They can be detected with various molecular techniques and usually inform about the characteristics of the genomic location that they occupy. They serve as a tool for detecting and quantifying genetic variation in individuals, populations and species, and are useful to associate phenotypic and genotypic variation. Some of the most used DNA-based markers are restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple-sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) (Table 1.1) (Nadeem et al. 2017). They are briefly described below:

Restriction fragment length polymorphisms (RFLPs) are hybridization-based markers, where DNA is digested with restriction enzymes and cut at specific recognition sites, generating a large number of fragments of different length that can be separated in gels through

electrophoresis, transferred to membranes, hybridized with labelled probes (radioactively or chemically) and the resulting identified positions (bands) viewed in photographic film and genetically interpreted (Botstein et al. 1980). RFLPs are usually codominant markers of excellent quality that can be transferable between individuals and species. The first high quality and saturated maps of plant species were obtained with RFLPs (Bernatzky and Tanksley 1986). However, the method is time consuming and expensive, and RFLPs have mostly been substituted by other PCR markers of similar quality when they became available.

Random amplified polymorphic DNAs (RAPDs): A PCR method based on differential amplification of genomic DNA, using short random oligonucleotide sequences (up to 10 bases long) as primers (Williams et al. 1990). Nucleotide changes at or near primer binding sites produce DNA polymorphisms. Genome knowledge is not required as universal primers are used in this method. Although this is a simple and affordable method, the main drawback is its low reproducibility. Also, RAPDs are usually dominant markers, so heterozygous individuals cannot be distinguished from homozygous (Bardakci 2001), which is a major limitation for most of their applications.

Amplified fragment length polymorphisms (AFLPs), combine elements of RFLP and RAPD technology (Vos et al. 1995). DNA is digested with two restriction enzymes and oligonucleotide adapters are ligated to the ends of the fragments obtained. Then primers are designed based on the adapters and the restriction site sequences to amplify a selected sample of the fragments. Amplified fragments are then separated by electrophoresis and identified. AFLPs are generally dominant markers, economically affordable, and more reproducible than RAPDs, that have successfully been used for certain applications such as bulked segregant analysis (Michelmore et al. 1991) although they have currently been substituted by other types of markers of better quality such as SSRs or SNPs.

SSRs or microsatellites: SSRs are tandem repeats of 1-6 nucleotide long DNA motifs that frequently present alleles with different numbers of repeats. Microsatellite existence was reported in various eukaryotes from yeasts to vertebrates. SSRs are present across the genome including coding and non-coding regions. The non-coding regions contain more SSRs than the coding regions. Over the years, SSRs have gained popularity in comparison to RFLPs and AFLPs due to their abundance, reproducibility, codominant inheritance and multi-allelic nature (Kalia et al. 2010) and because they can be obtained with simple and affordable methods. The assay requires very small amount of DNA (100 ng) that is submitted to PCR with primers

obtained from the flanking sequences of the SSR. Genotyping is performed after electrophoretic separation of the amplified fragments that can be done in agarose and acrylamide gels or with capillary electrophoresis.

Single-nucleotide polymorphisms (SNPs): The single nucleotide substitutions in the genome sequence are referred to as SNPs. All living organisms contain high numbers of SNPs, that are the most abundant source of DNA variability. The frequency of SNPs in plants ranges between 1 in every 100–300 bp (Xu 2010), distributed all over the genome and can be identified with an ample variety of molecular techniques. They are codominant markers, usually biallelic, and they are amenable to high-throughput methodologies that allows in certain cases to produce large numbers of data at very low costs. The high-throughput genotyping methods like GBS (Genotyping by sequencing), NGS (Next generation sequencing), chip-based NGS and allele specific PCR, makes SNPs as the most used markers for genetic analysis (Elshire et al. 2011; Shendure and Ji 2008).

Table 1.1: Comparison among the most widely used molecular markers for genotyping

	Molecular markers				
Characters	**RAPD**	**RFLP**	**AFLP**	**SSR**	**SNP**
Marker inheritance	Dominant	Codominant	Dominant	Codominant	Codominant
Method	Simple	Robust	Robust	Robust	Robust
DNA quality	Medium	High	High	Low	High
DNA quantity	Low	High	High	Low	Low
Transferability	Non-transferable	Transferable	Transferable	Transferable	Transferable
Cost	Inexpensive	Expensive	Inexpensive	Inexpensive	Inexpensive
Primers	Up to 10 random nucleotides	Low copy sequences	Sequence based	Sequence based	Allele specific
Polymorphism	Moderate	Moderate	Moderate	Very high	Moderate
Methodology	PCR	Southern blot	PCR	DNA-sequence	DNA-sequence
Reproducibility	Non-reproducible	Reproducible	Intermediate	Reproducible	Reproducible

1.2.3 Linkage maps

Molecular markers are widely used in genetic mapping and QTL analysis. The number of markers used for linkage map construction depends on the genome size of the species and its recombination rates. However, with recent advances in NGS, thousands of markers can be employed to obtain a high-resolution linkage map. Alternatively, for species with sequenced genomes, markers can be chosen to cover the whole genome at even physical or genetic distances. Generally, a segregating population with 50-250 individuals is used for construction of linkage map (Mohan et al. 1997). F2 and BC1 populations are the simplest mapping populations, as they can be obtained in short time and are easy to construct, but other populations such as RILs, segregating F1, DHLs etc. can be used for this purpose. These maps help in positioning the genes and QTLs, and provide information about the genetic distance between the markers. The two mainly used mapping functions are the Kosambi (recombination events influence the occurrence of adjacent recombination events) and the Haldane mapping function (no interference between cross over events is assumed) (Collard et al. 2005). Linkage maps provide marker-trait association data and help in choosing markers that could be used in marker assisted selection. To locate a QTL/gene of interest in genetic map it should contain adequate number of polymorphic markers evenly distributed in the genome.

An initial linkage map of *P. persica* was constructed by Chaparro et al. (1994) in an F2 population 'NC174RL' × 'Pillar' to 15 linkage groups, using one isozyme, two morphological markers and 83 RAPD markers. This was one of the first maps to be constructed with PCR technology, but as RAPDs are usually dominant, hardly transferable to other populations and with low reproducibility, their use in map construction was further avoided. A year later, the first linkage map was designed by Rajapakse et al. (1995) with RFLPs that are reproducible. This was constructed in a F2 progeny in eight linkage groups using seven morphological markers, 12 RAPDs and 46 RFLPs. However, these maps produced using peach progeny resulted in a high proportion of monomorphic markers due to its low level of genetic variability (Byrne 1990). To overcome this, the map construction was done using the F2 progenies of interspecific almond x peach crosses that were highly polymorphic. The map constructed with RFLPs in almond (cv. Texas) × peach (cv. Earlygold) F2 progeny (T×E), was used as a reference map for peach and other *Prunus* species (Joobeur et al. 1998). This map was further improved by addition of SSRs and sequence based markers (Aranzana et al. 2003; Dirlewanger et al. 2004). Currently, this map is widely used for locating major genes and QTLs across *Prunus* species.

1.3 Gene mapping and marker-assisted selection (MAS)

A large number of genes controlling important traits have been genetically described in peach (Aranzana et al. 2019). Most of the 31 major genes initially described in peach (Monet et al. 1996) have eventually been mapped in different populations of various *Prunus* species. Using the existing reference map (Joobeur et al.1998), Dirlewanger et al. (2004) integrated 28 major genes from different linkage maps into a single map based on the high synteny between the genomes of *Prunus* species. Presently, there are 53 major genes that are mapped across all the linkage groups in different *Prunus* crops (Table 1.2). These include traits related to disease resistance, flower, fruit or nut quality and vegetative characters. The majority of the genes mapped in Table 1.2 have commercial application in peach breeding. Those affecting fruits (white vs. yellow vs. red flesh, early vs. late maturity, melting vs. non-melting flesh, peach vs. nectarine, flat vs. round, clingstone vs. freestone, acid vs. non-acid and aborting fruit), flower (pollen sterility), tree architecture (normal vs. pillar) and disease resistance (green peach aphid, root-knot nematodes, sharka and powdery mildew).

Table 1.2: Fifty-three major genes mapped across all the linkage groups (G) in different species of *Prunus*

Trait name	Symbol	G	References
Sharka resistance	*Sharka*	G1	Hurtado et al. (2002); Lalli et al. (2008)
Green peach aphid resistance	*Rm2*	G1	Lambert and Pascal (2011)
Green peach aphid resistance	*Rm1*	G1	Pascal et al. (2017)
Flesh color (white/yellow)	*Y*	G1	Warburton et al. (1996); Bliss et al. (2002)
Anther color (yellow/anthocyanic)	*Ag2*	G1	Donoso et al. (1996)
Juiciness	*Jui*	G1	Donoso et al. (1996)
Evergrowing	*Evg*	G1	Wang et al. (2002)
Flesh color (normal/anthocyanic)	*DBF2*	G1	Donoso et al. (2016)
Flower color	*B*	G1	Jáuregui (1998)
Root-knot nematode resistance	*Mi*	G2	Jáuregui (1998); Lu et al. (1998); Yamamoto et al. (2001); Claverie et al. (2004); Gillen and Bliss (2005)
Root-knot nematode resistance	*Mj*	G2	Yamamoto et al. (2001, 2005)
Root-knot nematode resistance	*RMia*	G2	Duval et al. (2014)
Powdery mildew resistance	*Vr3*	G2	Donoso et al. (2016); Marimon et al. (2020)
Male fertility restorer	*Rf1*	G2	Donoso et al. (2015)
Shell hardness	*D*	G2	Arús et al. (1998)
Broomy (or pillar) growth habit	*Br*	G2	Scorza et al. (2002)
Double flower	*Dl*	G2	Chaparro et al. (1994); Meng et al. (2019)
Anther color (yellow/anthocyanic)	*Ag*	G3	Joobeur et al. (1998); Donoso et al. (2016)
Polycarpel	*Pcp*	G3	Bliss et al. (2002)
Flesh color around the stone	*Cs*	G3	Yamamoto et al. (2001)

Table 1.2 (Continued)

Trait name	Symbol	G	References
Flower color	*Fc*	G3	Yamamoto et al. (2001); Lu et al. (2021)
Skin color (Highlighter/Anthocyanic)	*H*	G3	Breto et al. (2017); Lu et al. (2021)
weeping tree	*pl*	G3	Pascal et al. (2017)
Temperature-sensitive semi-dwarf	*TssD*	G3	Lu et al. (2016)
Flesh color (normal/anthocyanic)	*Bf*	G4	Werner et al. (1998); Bliss et al. (2002)
Flower color	*Fc2*	G4	Donoso et al. (2016)
Late blooming	*Lb*	G4	Ballester et al. (2001)
Maturity date	*MD*	G4	Eduardo et al. (2011); Pirona et al. (2013); Donoso et al. (2016)
Slow ripening	*SR*	G4	Eduardo et al. (2015)
Almond fruit type	*Alf*	G4	Donoso et al. (2016)
Flesh adhesion (clingstone/freestone) / Flesh type (Melting/nonmelting)	*F-M*	G4	Dettori et al. (2001); Yamamoto et al. (2001); Verde et al. (2002); Dirlewanger et al. (2006)
Hybrid incompatibility	*Hls1*	G4	Tsuruta and Mukai (2015)
Non-acid fruit	*D*	G5	Dirlewanger et al. (1998, 1999); Etienne et al. (2002); Boudehri et al. (2009)
Flesh color (normal/anthocyanic)	*DBF*	G5	Shen et al. (2013)
Skin hairiness (nectarine/peach)	*G*	G5	Dirlewanger et al. (1998, 1999, 2006); Bliss et al. (2002); Lu et al. (2021)
Kernel taste (bitter/sweet)	*Sk*	G5	Bliss et al. (2002)
Leaf shape (narrow/wide)	*Nl*	G6	Yamamoto et al. (2001)
Plant height (normal/dwarf)	*Dw*	G6	Yamamoto et al. (2001)
Male sterility	*Ps*	G6	Dirlewanger et al. (1998, 2006); Eduardo et al. (2020)
Male fertility restorer	*Rf2*	G6	Donoso et al. (2015)
Fruit skin color	*Sc*	G6-G8	Yamamoto et al. (2001)
Leaf color (red/yellow)	*Gr*	G6-G8	Jauregui (1998); Yamamoto et al. (2001)
Powdery mildew resistance	*Vr2*	G6-G8	Pascal et al. (2017)
Botryosphaeria dothidea resistance	*Botd8*	G6-G8	Castillo et al. (2018)
Aborting fruit	*Af*	G6	Dirlewanger et al. (2006)
Fruit shape (flat/round)	*S*	G6	Dirlewanger et al. (1998, 2006)
Gametophytic self-incompatibility	*Si*	G6	Vilanova et al. (2003, 2005, 2006); Olmstead et al. (2008)
Double flower	*Di2*	G6	Pascal et al. (2017); Gattolin et al. (2018)
Root-knot nematode resistance	*Ma*	G7	Lecouls et al. (1999); Claverie et al. (2004)
Powdery mildew resistance	*Sf*	G7	Dirlewanger et al. (1996)
Leaf gland shape and presence	*E*	G7	Dettori et al. (2001)
Flower morphology	*Sh*	G8	Ogundiwin et al. (2009); Fan et al. (2010); Donoso et al. (2016)
Slow softening	*Sw*	G8	Ciacciulli et al. (2018)

MAS is a marker-based strategy to select for specific characters of simple inheritance (usually determined by one or a few genes/QTLs) based on the genotype of markers tightly linked to these genes/QTLs. Application of MAS to breeding programs saves time, as fruit related traits, or other traits expressed later in the development of the plant, can be selected at the seedling

stage using molecular markers, without the need of evaluation in the field for selection. This is particularly efficient in the case of species with long intergeneration periods such as fruit trees, because the savings of space, time and resources is larger than in herbaceous annual crops. For efficient MAS, the prerequisites are reliable markers, good quality DNA, genetic maps, and knowledge of marker-trait associations. Nowadays, it is widely used by plant breeders in most crops to select a desired trait or eliminate an unwanted trait. Molecular markers tightly linked or based on the causal polymorphism of the gene that determines the trait of interest are key for MAS (Byrne 2007). International collaborative research projects in both Europe (Laurens et al. 2012) and USA (Iezzoni et al. 2020), focusing on various rosaceous crops have represented an important advance in the understanding of the genetics and the development of markers adequate for breeding applications. Presently, MAS is being routinely used in peach for the selection of many traits, some of the most frequently selected are the acid vs. subacid fruit taste (*D*/*d*; Eduardo et al. 2014) and round vs. flat fruit (*Sh*/*sh*; Picañol et al. 2013), and to eliminate certain unwanted characters such as pollen sterility (*Ps*/*ps*; Eduardo et al. 2020) and dwarf trees (*Dw*/*dw*; Cantín et al. 2018).

1.4 Introgression lines (ILs)

Introgression is the transfer of DNA fragments between a donor species, usually an exotic line (wild relative or distant germplasm of the same species), and an elite cultivated line used as recurrent parent in a backcross breeding program. An introgression line (IL) is an individual nearly identical to the recurrent parent, except for a small DNA fragment introgressed from the donor parent (usually a single fragment), and are also defined as Near-Isogenic Lines (NILs) (Monforte and Tanksley 2000) or Chromosome Segment Substitution Lines (CSSLs) (Balakrishnan et al. 2019). IL collections are an advantageous alternative to other classical types of segregating populations used for genetic analysis, such as F2s, BC1s or RILs, and are especially indicated for the study of the genetics of traits between very distant materials, either between different species or between wild and cultivated germplasm (Eshed and Zamir 1995).

The advantages of ILs over other segregating populations are: (1) they are very effective for QTL analysis (Eshed and Zamir 1995), because of the high genetic similarity between each line and the recurrent parent, and the fact that ILs are immortal populations that can be replicated as much as needed, enabling an accurate estimation of the phenotype in a very

uniform genetic background, and allowing the detection of QTLs of very small effects that would require large populations of other types (Keurentjes et al. 2007); (2) the environment effects and the interaction of QTLs with the environment can be studied accurately; (3) the effects of each QTL can be well estimated including slight unwanted pleiotropic effects that may be difficult to detect in other population types with more heterogeneous genetic backgrounds; (4) the interactions between QTLs can be studied by creating additional lines with specific QTLs using two or more ILs that contain them (Gur and Zamir 2015), and (5) the introduction of a trait of interest into a commercial variety can be quick and straightforward, as there is a low percentage of foreign (donor) genome (Jeuken and Lindhout 2004).

The disadvantages of ILs are that development of a complete set may take a long time (many backcross generations), and is a laborious (large numbers of individuals need to be surveyed for obtaining a few with a single introgression) and expensive (need of many plants and markers per plant to cover the complete genome) process (Tuinstra et al. 1997). Another limitation is that for characters that need the combined action of more than one QTL to be expressed, such as disease resistances or color-related traits, the expected genotype may not be recovered in the IL collection, requiring a complementary analysis with other population types (F2, RILs) to identify the component QTLs, and later generating the appropriate combinations from crosses between single-introgression ILs. Despite these difficulties, introgression breeding with wild crop relatives using ILs has already had a significant impact and has a great potential in the development of modern cultivars, particularly new varieties with improved disease resistance and product quality (Zamir 2001).

An IL collection usually covers all the genome of the donor parent with as small introgressions as possible. The usual method for IL development is marker selection over successive backcross generations followed by at least one selfing generation to fix the introgressed fragment. The number of BC generations vary depending on the different strategies of the authors, the sizes of the populations at each stage and the number of chromosomes and recombination rates of each species, from three in tomato - *Solanum pennellii* (Eshed and Zamir 1995), four in tomato - *S. pimpinellifolium* (Giacomo et al. 2020), five in lettuce-*Lactuca saligna* (Jeuken and Lindhout 2004) and six in *Cucumis melo L.* (Eduardo et al. 2005). Less generations usually imply less time, but in general longer introgression fragments as there is less opportunity for recombination. Because of their many advantages as a resource for genetic analysis, ILs have been developed in many crop plant species that include tomato (Eshed and Zamir 1995; Giacomo et al. 2020), melon (Eduardo et al. 2005), strawberry (Urrutia et al.

2015), cabbage (Ramsay et al. 1996), lettuce (Jeuken and Lindhout 2004), rice (Xiao et al. 2016), wheat (Lu et al. 2020), maize (Szalma et al. 2007), barley (Von Koorf et al. 2004), *Arabidopsis thaliana* (Fletcher et al. 2013) and many others (Balakrishnan et al. 2019). To our knowledge, no IL collection has been developed in woody perennial species, so the peach-almond collection will be the first. This is a consequence of the long-term endeavor that such a project represents, in this case, three generations after the F1 hybrid almond × peach parent, but the expected benefits as a resource for *Prunus* geneticists and breeders are the same as in any other species.

One additional use of ILs is their use to facilitate the fine mapping of genes of interest to find diagnostic markers for MAS, and as a first step towards gene cloning. In the case of the 'Texas' × 'Earlygold' cross, the genetics of many characters has already been unveiled (Joobeur et al. 1998, Donoso et al. 2015; Donoso et al. 2016) and 11 major genes have been identified and mapped. One of them of especially high value, a strong resistance against powdery mildew (*Vr3*/*vr3*) has already been studied in detail (Marimon et al. 2010) and a region of 270 kb with 27 genes has been identified to contain *Vr3*, one of them, a Disease Resistance Protein RGA2 as a most probable candidate. In this thesis, additional steps have been undertaken to the fine mapping of three more genes of interest coming from almond, one the *DBF2*/*dbf2* that determines the peach red flesh character, with the allele producing the red flesh phenotype coming from almond, and two more genes that explain a good part of the difference between a peach and an almond fruit: the *Alf*/*alf* gene that determines the formation of a thick mesocarp as in peach, as opposed to the thin almond mesocarp, and the *Jui*/*jui* gene, that is involved in the generation of the juicy mesocarp of peach, compared to the non-juicy flesh of almond. Knowledge of the details of the genetics of these characters will be useful to understand the evolution of these two inter-compatible species, that have remarkable differences in other biological, populational and agricultural respects, and to design new strategies for gene interchange between peach and almond that are useful to generate new and improved varieties of both species.

2. Objectives

The main objective of this PhD thesis is to understand the genetic basis of the variability between peach and almond, and to enrich the peach genome with the introgression of novel genes of interest from almond. To achieve this, the following specific objectives are proposed:

1. Creation of a peach-almond introgression line (IL) collection, using 'Earlygold' peach as a recurrent parent and 'Texas' almond as a donor parent, as a tool for genetic analysis in the *Prunus* genus.

    1.1 Linkage map construction and QTL analysis for the traits segregating in the 'Earlygold' F2 progeny and comparison with those previously obtained with the BC1 population ('Texas' × 'Earlygold') × 'Earlygold'.

    1.2 To develop two IL collections of almond DNA fragments that cover the whole peach genome into the 'Earlygold' peach genetic background, one in heterozygosis and one in homozygosis, and analyze them with a set of traits of known inheritance.

2. Fine mapping of three major genes determining key differences between almond and peach fruit: almond fruit (*Alf*/*alf*), fruit juiciness (*Jui*/*jui)* and dominant blood flesh (*DBF2*/*dbf2*) found in almond × peach crosses, and identification of candidate genes.

3. Comparative QTL analysis in peach ‘Earlygold’ F2 and backcross progenies

This chapter corresponds to a published paper: Kalluri N, Eduardo I, Arús P (2021) Comparative QTL analysis in peach ‘Earlygold’ F2 and backcross progenies. Sci Hortic (https://doi.org/10.1016/j.scienta.2021.110726)

Abstract

Based on detailed maps, DNA sequences and phenotypic data, there is a great deal of information on the genetics and genomics of 'Earlygold', a historical peach cultivar from the US. The F2 between 'Texas' almond and 'Earlygold' peach (T×E) was used to construct the first saturated peach linkage map that later became the reference map for the *Prunus* genus. This population and the first backcross (Texas' × 'Earlygold') × 'Earlygold' (T1E) yielded information on QTLs for a large number of agronomic traits, and T1E is being used as the basis for constructing a set of introgression lines of 'Texas' fragments into the 'Earlygold' background, currently in progress. This paper describes the construction of a high-density SNP map for 'Earlygold' using an F2 population, and the QTL analysis of 24 traits. Results of maps and QTLs are compared with those from the 'Earlygold' parent of the T1E map, using the same set of markers and characters. Results show major differences between the two progenies in terms of numbers of markers mapped and the capability of detecting QTLs, with a large increase in the resolution of maps and QTLs when using the F2 progeny compared to the T1E pseudo-testcross. In addition, we provide data on leaf senescence color, studied for the first time in peach, with two consistent QTLs located in the same position as other color-related genes and QTLs.

Key words: *Prunus persica*, F2, backcross, QTL analysis, comparative mapping, leaf senescence color.

3.1 Introduction

Peach [*Prunus persica* (L.) Batsch], is an economically important stone fruit crop and one of the model species of the *Rosaceae* family (Shulaev et al. 2008). Like most cultivated members of the *Prunus* genus, such as almond, apricot, plum and cherry, peach is diploid (2n=*2x=16*) with a compact and sequenced genome of ~250 Mb (Verde et al. 2013). It has a self-compatible mating system and a short intergeneration period of 3-4 years, in contrast to other rosaceous fruit tree species that are usually self-incompatible and require a longer time for fruiting. Some of the major targets in the current peach breeding programs are difficult to meet, such as extended shelf life, better fruit quality and enhanced disease resistance, mainly due to the low levels of genetic variability of the elite peach materials (Micheletti et al. 2015).

'Earlygold', an old peach cultivar bred in the US, was crossed with 'Texas' almond in the peach rootstock breeding program of IRTA during the 80s and produced several hybrids. One of these, particularly fertile and prolific, was chosen to obtain an F2 progeny to construct the first saturated linkage map of *Prunus* (Joobeur et al. 1998), which later became the reference for the genus (Aranzana et al. 2003; Dirlewanger et al. 2004). The F2 population was useful for map construction, but only about a half of its progeny was fertile and fruited (Donoso et al. 2016), so a BC1 progeny with 'Earlygold' as recurrent parent was obtained, where most individuals produced fruit. In this cross, the 'Texas' cytoplasm resulted in male sterility in the peach nuclear genetic background unless at least one of two independent restorer factors from 'Texas' were present (Donoso et al. 2015). The BC1 progeny and further backcross and selfing generations have been used as a proof of concept for marker-assisted introgression (MAI), a fast strategy to obtain individuals with a single almond introgressed fragment in the BC2 (Serra et al. 2016). These materials are currently being used to develop an introgression line (IL) collection of almond fragments in the 'Earlygold' background.

In this paper we elaborate a high-density map of an 'Earlygold' F2 progeny, and examine its variability for a set of characters of agronomic interest, to understand their inheritance and to find useful marker-trait associations. We compared these results with those of a QTL analysis on the BC1 ('Texas' × 'Earlygold') × 'Earlygold' studied by Donoso et al. (2016), that allowed us to examine the QTLs of 'Earlygold' in two different genetic backgrounds. These data will also provide information useful to understand the importance of the 'Earlygold' allelic

variation in the characters that will be examined in the IL collection currently under construction.

3.2 Materials and Methods

3.2.1 Plant materials

The 'Earlygold' F2 population N = 75 (E×E) was used for the study. The trees are located at the IRTA experimental station of Torre Marimon (Caldes de Montbui, Spain), planted on their own roots in 2015. The spacing between the trees was 2.5 m and between the rows 4.0 m. The trees were thinned out during the second and third year of fruiting (2018 and 2019). Young leaves were collected from all the trees for DNA extraction (Doyle and Doyle 1990), to perform genotyping. The E×E population initially consisted of 81 plants, later two plants died and four did not produce genotypic data, resulting in 75 plants being used for mapping and phenotyping.

3.2.2 Phenotyping

The population was evaluated for 24 traits, 18 over three seasons in the years 2017 to 2019 and six in only two seasons: chlorophyll content of the leaves and leaf dry weight (only in 2018 and 2019), leaf color at senescence and early and late leaf fall (2019 and 2020), and beginning of shooting (2020 and 2021). These characters were also analyzed by Donoso et al. (2015, 2016), except for leaf color at senescence, fruit firmness and pH that we analyze here for the first time. The phenotyping methods were essentially identical to those used by these authors. The characters scored can be classified into four main categories and their measurement is described below:

Flower: The flower shape, showy (large petals) and non-showy (small petals) is determined by a major gene (*Sh*/*sh*) in peach, where showy flowers are homozygous (*shsh*) for the recessive allele (Bailey and French 1942).

Phenology: Flowering time (FT) was scored as the number of Julian days when 50% of the flowers were open. Beginning of shooting (BS) was the number of Julian days when 5% of the shoots start to appear. Maturity date (MD) was measured as the number of Julian days with 50% of the fruits mature, as determined by changes in the skin color and flesh firmness. Fruit development period (FDP) was scored as the difference in days between the flowering time

and maturity date. Beginning of leaf fall (BLF) was scored as the number of Julian days when 10% of leaves had dropped, and end of leaf fall (ELF) when 90% had dropped.

Fruit: Fruit weight (FW), in grams (g), was the average weight of six mature fruits per individual using a digital balance. Fruit production (FP) was estimated on a scale of 1 to 4 (1=no fruits, 2= <10 fruits, 3= 10-50 fruits and 4= >50 fruits). Intensity of red skin color (ISC) was visually determined by the % of red color at maturity (1=0-25%, 2=25-50%, 3=50-75% and 4=75-100%). Fruit firmness (FF) was evaluated with a hand penetrometer (Wagner, Model 53200), taking the average value of three fruits with the measurements from both sides for each fruit. Soluble solid content (SSC), expressed in Brix degrees, was measured from the juice of three fruits using a digital refractometer (Atago, Tokyo, Japan). Titratable acidity (TA) and pH were determined using a HI-84532-02 Titratable Acidity Mini Titrator and a pH meter (Hanna instruments, Rhode Island, USA) by diluting 5 ml of fruit juice with 45 ml of water and titrating with 0.5 M NaOH to a pH of 8.2. TA was calculated in g/l of malic acid.

Leaf: Chlorophyll content (CC) was estimated as the average from ten leaves per tree using SPAD 502 (Konica Minolta, Osaka, Japan). During the months of July-August, eleven leaves per tree were collected from the middle of the tree branches. The leaves were then scanned and their images stored as TIF files for further analysis. The leaf dimensions were measured using a Tomato Analyzer 3.0 (http://www.oardc.ohio-state.edu/vanderknaap) software. Leaf parameters analyzed (Figure 3.1) were leaf length (LL), petiole length (PL), leaf blade length (LBL), leaf blade width (LBW), leaf shape (LSH) as the ratio of LBL/LBW, leaf perimeter (LP) and leaf surface (LS). All the measurements were in cm, except for LSH that is a ratio, and LS that was measured in cm^2. Later, the leaves were placed in an incubator for 3 days at 60° C to determine the leaf dry weight (LW), in grams (g). Leaf color at senescence (LCS) was scored visually over two years, once a week, in September and October (Figure 3.2) using a scale of 1 (non-purple, including green yellow and red) and 2 (purple leaf).

The phenotypic data were analyzed statistically using JMP 14.0 software (SAS Institute, Cary, NC, USA). Correlations between different traits and years were calculated using the Spearman correlation coefficient.

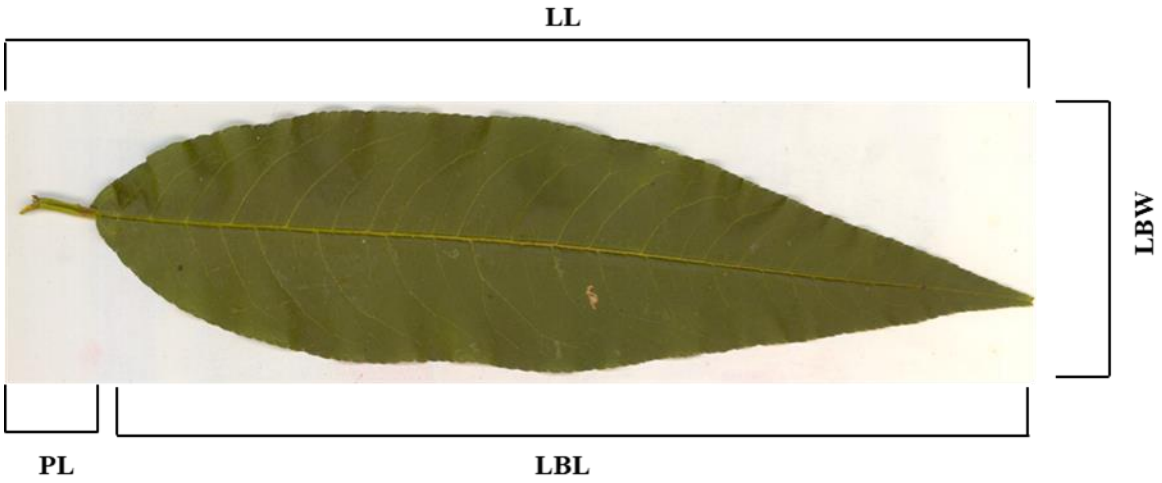


Figure 3.1: Schematic representation of leaf dimensions (PL petiole length, LBL leaf blade length, LL leaf length, LBW leaf blade width).

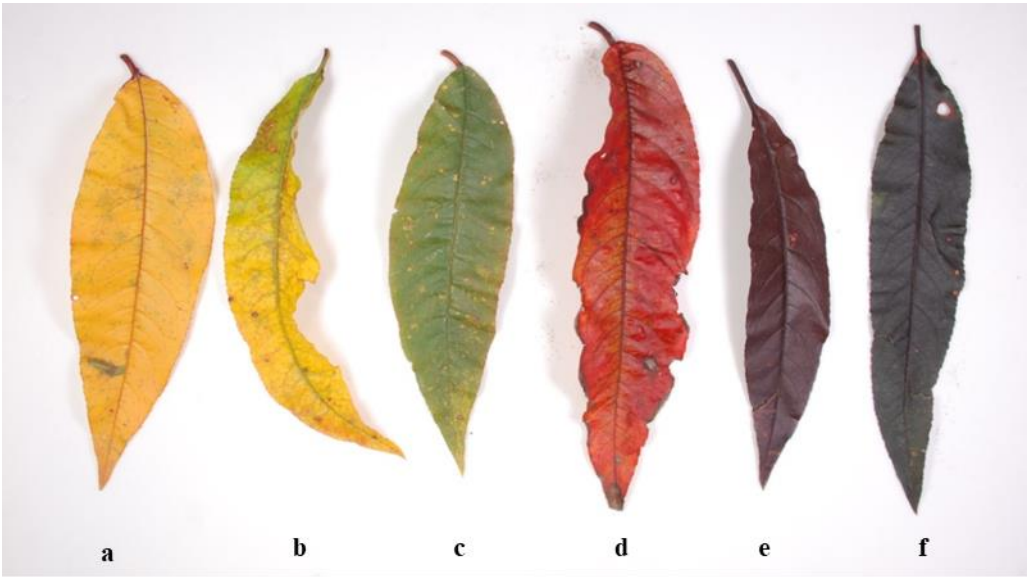


Figure 3.2: Leaf color at senescence a, b-yellow, c-green d-red and e,f-purple.

3.2.3 Construction of linkage map and QTL analysis

For linkage analysis, genotype data were obtained from the 9k International Peach SNP Consortium (IPSC) Illumina Infinium SNP array (Verde et al. 2012) in 75 plants of the E×E population. The linkage map was constructed using MapMaker/exp 3.0 (Lander et al. 1987) using the Kosambi distance function. We initially ordered the markers based on their physical position and established a set of bins (i.e. groups of markers with identical genotype for all the individuals), where each bin is separated from the adjacent bin by a single or a few

recombination events. Finally, a single SNP from each bin was selected for the dataset used to construct the linkage map. References for chromosome/linkage group numbers and orientation and physical positions were those of Dirlewanger et al. (2004) and Verde et al. (2017).

QTLs for all the traits were analyzed using the interval mapping method with the MapQTL 6.0 software (Van Ooijen et al. 2002). All the QTLs with a LOD $\geq$ 3.0 were considered as significant, as were those QTLs with a LOD $\geq$ 2.5 in one year and a LOD $\geq$ 3.0 in the rest. The QTLs were considered consistent if they were detected every year. The maps and positions of the QTLs were drawn using the map chart 2.3 software (Voorips 2002).

Gene action (GA) was established following the guidelines of Tanksley (1993), based on the values of additivity, a=(A-B)/2, and dominance d=H-[(A+B)/2], where A and B are the average values of the trait in the homozygous individuals for a given marker in the female and male parent, respectively, and H that of the heterozygotes. We considered gene action additive (A) if the quotient $|d/a| \leq 0.5$, dominant (D) if $0.5 < |d/a| \leq 1.5$, and overdominant (O) if $|d/a| > 1.5$.

3.3 Results

3.3.1 Linkage maps and comparisons

From the SNPs of the 9k chip, 1,640 were polymorphic in the E×E population and distributed in 269 bins. The E×E map covered a total genetic distance of 439.1 cM, detecting the expected eight linkage groups with a higher end of 76.2 cM for linkage group 2 (G2) to a lower end of 37.1 cM for G4. The overall physical length covered by this map was 186.7 Mb, 82.7% of the total physical distance (225.6 Mb) of the peach genome v2.0a1 (Verde et al. 2017). There were 24 gaps of >2Mb, the largest at the distal end of G4 (15.3 Mb), overall accounting for 116.4 Mb, equivalent to 52% of the total physical distance of the sequenced peach genome (Table S3.1).

The map of 'Earlygold' (the E map) constructed with the ('Texas' × 'Earlygold') × 'Earlygold' BC1 population (Donoso et al. 2015) using the same SNP chip plus 41 SSR markers, is similar in many respects to the one we present here (see comparisons in Table S3.1 and Figure S3.1), particularly with respect to the physical coverage (189.7 Mb), number of bins (214), number of gaps >2 Mb (23; the same as in the E×E map) and physical distance covered by these gaps (109.0 Mb). The genetic distance of the E map (521.2 cM) was 16% higher than that of the

E×E map, and some linkage groups were clearly longer than in the E×E map (Table S3.1), although not significantly longer (the paired t test of the differences of the genetic distances between the common markers in the extremes of each chromosome was 1.21). The most striking difference between the E and E×E maps was the number of SNP markers that could be mapped: 1,640 in E×E and only 1,050 in E, resulting in a marker density of 0.27 cM/marker in E×E compared to 0.47 in the E map. When examining the mapped markers in detail, we observed that their distribution was similar in both maps, and that most of the 1,050 SNPs of E (97%) were mapped in E×E, but of the SNPs mapped only in E×E (623) the majority (411; 66%) had a dominant pattern of inheritance, with only two genotypes from the three expected (3:1 ratio), instead of the usually codominant 1:2:1 SNP inheritance expected in an F2 progeny.

3.3.2 Trait phenotypic data and QTL analysis

All of the 24 traits scored were quantitative, except for flower shape (*Sh*), which was scored as qualitative and placed on G8 at the position 23.8 cM with the nearest marker (SNP_IGA_862006), with physical position 13,825,065 bp, cosegregating with the gene. Trait distributions and main features are described in Table S3.2 and Figure S3.2 Those that significantly departed from normal in more than one year were phenology-related (FT, BS, MD, FDP, FP, BLF and ELF), leaf color at senescence (LCS) and three fruit characters (ISC, FF and pH).

Correlation analysis was performed between all quantitative traits and the years in which they were measured (Table S3.3). The correlations between years for a particular trait were high and positive for FT, BS, MD, FDP, FW, SSC, CC, and LCS, and intermediate-to-low for the rest. For correlations between traits, there was a clear positive correlation between MD, FDP, FW and SSC, and between BLF and ELF. Most leaf traits (LP, LS, LBW, LL and LB) of the same year were strongly correlated, but the correlation decreased when comparing data from different years. A negative correlation was observed for MD and characters related to anthocyanin coloration of senescent leaves (LCS) and intensity of the red skin fruit color (ISC).

QTLs were detected for 19 traits, between one to four QTLs per trait, and none were identified for four traits (ELF, FP, LBW and LW), giving a total of 33 QTLs (Table S3.4 and Figure S3.3). Twelve QTLs of those detected (36%) were consistent in all years studied (Table 3.1, Table S3.4 and Figure 3.3). Most of the consistent QTLs detected explained the variability of phenology (FT, BS, MD and FDP) and fruit (FW, SSC, TA) traits, whereas of the QTLs for the leaf traits, only CC and LCS were consistent.

Considering only consistent QTLs (Table 3.1; Figure 3.1), two were found for flowering time, one on each of G7 and G8 with R^2 = 28.3-42.8% and R^2 = 16.5-20.9% respectively. Also, two QTLs for beginning of shooting, one on each of G7 and G4 with R^2 = 30.9-45.2% and R^2 = 19.0-24.9%, respectively. A QTL with a large effect for maturity date (R^2 = 80.6-82.0%) and fruit development period (R^2 = 80.1-85.0%) was found at the end of G4, and, at the same position, two other QTLs were identified for fruit weight and soluble solid content explaining 28.0-38.6% and 44.3-47.4% of the phenotypic variance, respectively. For titratable acidity, a QTL was detected at the end of G6, with R^2 = 17.3-34.9%. Three QTLs were identified for leaf-related traits, one for chlorophyll content on G3 (R^2 = 16.3-19.0%) and two for leaf color at senescence on G3 and G4 explaining 26.8-28.6% and 40.0-41.0% of phenotypic variance, respectively.

3.3.3 Comparison with the QTLs detected in the E map of the T1E population

Only six QTLs were detected on the map of 'Earlygold' obtained from the T1E population by Donoso et al. (2016), all of which we found using the E×E population. Two were consistent QTLs in E, one for end of flowering time, and the other for beginning of shooting, both located on G7 at the position of the consistent qFT7 for flowering time in the E×E map (Table 3.1). Two additional QTLs, for maturity date and fruit development period, both on G4 in the E map, correspond to the consistent qMD4 and qFDP4, respectively, identified here (Table 3.1). In the E map, these two latter QTLs were found only in one of the three years studied. Similarly, beginning of flowering time (BFT) that was mapped to the same position as qFT7, was detected in only two of the three years studied in E, and a QTL for beginning of shooting in G4 identified in only one of the three years in E probably corresponds to the consistent qBS4 in E×E. In all, for a set of 20 common characters, we found six QTLs in E and 26 in E×E (see Table 3.1 and Table S3.4).

Table 3.1: Summary of consistent QTLs detected in the E×E map ('Earlygold' F2 population) and E map ['Earlygold' map from the ('Texas' × 'Earlygold') × 'Earlygold' population]. Trait category, map type, QTL names, LOD score of the maximum peak, linkage group (G), map position of the maximum peak, percentage of explained phenotypic variance (R^2), additivity (a), dominance (d), d/a, gene action (GA).

Trait	Map	QTL name	LOD	G	Position (cM)	R^2	a	d	d/a	GA
Flowering time	E×E	qFT7	5.28-8.98	7	30.1-36.9	28.3-42.8	-2.07 to -2.75	-0.11 to -1.90	0.04 to 0.92	A-D
	E×E	qFT8	2.86-3.78	8	2.8-23.5	16.5-20.9	-0.12 to 1.91	-0.85 to 2.50	-0.45 to -18.66	A-O
End of flowering	E[1]	qFT7	2.50-4.65	7	36.8-40.8	8.9-14.3	3.07 to -4.24	-	-	-
Beginning of shooting	E×E	qBS7	5.70-9.27	7	34.9-35.9	30.9-45.2	-2.84 to -3.46	-0.62 to 0.15	-0.04 to 0.22	A
	E×E	qBS4	3.25-4.42	4	4.3	19.0-24.9	2.06 to 2.30	1.05 to 1.09	0.45 to 0.52	A-D
	E[1]	qBS7	4.24-8.13	7	41.4	12.1-23.5	3.72 to 7.03	-	-	-
Maturity date	E×E	qMD4	24.91-27.14	4	34.3-37.1	80.6-82.0	9.62 to 10.93	-1.23 to -5.18	-0.11 to -0.51	A-D
Fruit development period	E×E	qFDP4	23.84-30.12	4	37.1	80.1-85.0	9.27 to 11.32	-1.22 to -5.26	-0.10 to -0.49	A
Fruit weight	E×E	qFW4	4.85-7.73	4	33.6-37.1	28.0-38.6	12.18 to 18.32	-0.87 to -6.40	-0.35 to 0.21	A
Soluble solid content	E×E	qSSC4	8.77-9.51	4	36.4-37.1	44.3-47.4	1.23 to 1.62	-0.45 to 0.20	-0.27 to 0.14	A
Titratable acidity	E×E	qTA6	3.04-6.43	6	48.1-55.6	17.4-34.9	-0.67 to -0.90	0.09 to 0.24	-0.13 to -0.26	A
Chlorophyll content	E×E	qCC3	2.87-3.39	3	46.3-49.0	16.3-19.0	-0.47 to 0.12	1.15 to 3.28	-2.45 to 27.33	O
Leaf color at senescence	E×E	qLCS3	5.01-5.19	3	35.3	26.8-28.6	0.30 to 0.33	0.35 to 0.42	1.06 to 1.40	D
	E×E	qLCS4	7.88-8.48	4	36.4-37.1	40.0-41.0	-0.40 to -0.41	-0.006 to 0.01	-0.02 to 0.02	A

[1]Data from the E map are obtained from Donoso et al. (2016)

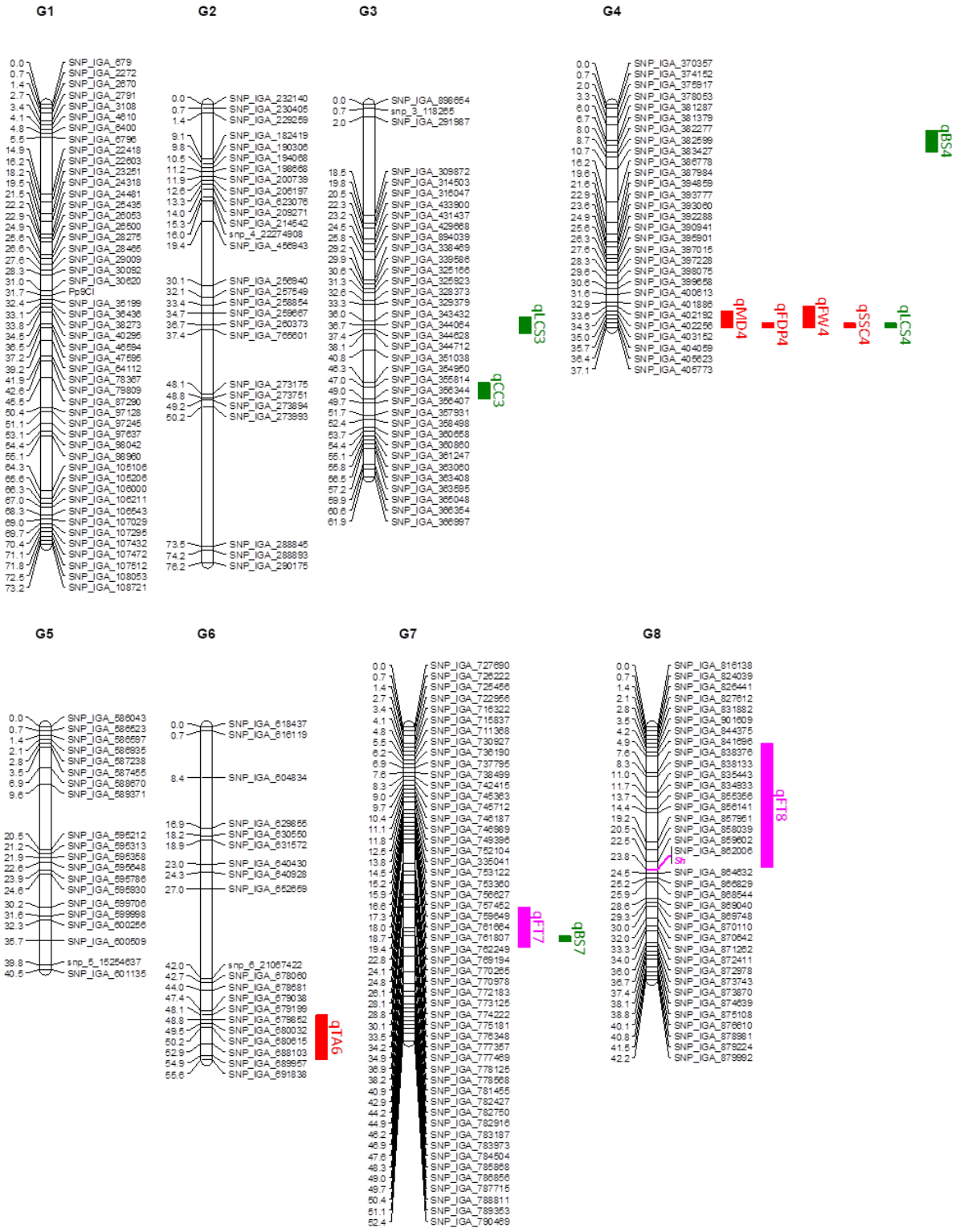


Figure 3.3: Map of the 'Earlygold' F2 progeny with the positions of a major gene (*Sh*) and 12 consistent QTLs over the years. Colors of the bars of QTLs are as follows: pink flower, red fruit and green leaf traits.

3.4 Discussion

Due to the self-compatibility system of peach, it is possible to obtain F2 progenies as well as F1 or backcross progenies for genetic analysis. F1 segregating progenies are frequently used for QTL analysis as they are the usual progeny type used in peach breeding. Since the parents used are partly heterozygous lines or cultivars, they segregate in the F1 progeny. These progenies, also termed pseudo-backcrosses (Grattapaglia and Sederoff 1994), are generally analyzed as two different backcross one (BC1) populations, each corresponding to one of the parents of the cross, and linked by markers that are heterozygous in both parents and then segregate 1:2:1 (or 1:1:1:1 when using multi-allelic markers). F2 populations are also used, especially to better understand the inheritance of specific characters as they recover the three possible genotypes, unlike BC1 populations that recover only two, making them suitable for the analysis of gene action (dominance, additivity, overdominance). F2 progenies are also more appropriate for constructing linkage maps, as they have more recombination events per meiosis (those of both parents), so more accurate maps can be constructed with the same number of progeny (Allard 1956).

Here we constructed a linkage map with the 9k peach IPGI SNP chip in the selfed progeny of 'Earlygold', which resulted in a high-density map with coverage of all eight peach chromosomes, with large fragments of DNA without segregating markers, approximately half of the 'Earlygold' genome, suggesting as in previous studies (Eduardo et al. 2013; Martínez-García et al. 2013; Donoso et al. 2015; Serra et al. 2017) that there are large regions of the peach genome identical by descent owing to the high level of coancestry of the cultivated gene pool. This map was similar in most respects (covered distance, chromosome length, gaps without markers) to the one obtained for 'Earlygold' in the interspecific backcross ('Texas' × 'Earlygold') × 'Earlygold' that was previously analyzed by Donoso et al. (2015). There was an important difference between these two maps, concerning the number of markers that could be integrated on the map: 1,640 SNPs in E×E, 56% more than the 1,050 in E. The fact that about two-thirds of the additional markers in E×E had dominant segregations suggests that a major cause for this difference is that these SNPs could have been discarded in the E map because they did not segregate (AA × Aa situations) or that segregations could have been strongly skewed compared to the two expected (1:1 and 1:2:1), resulting a loss of information compared to what can be obtained in an F2 population. While the quality of dominant markers is lower than codominant ones for mapping and QTL analysis, the results indicate the importance of

considering all possible segregation types before the initial marker filtering steps to maximize the available information for genetic analysis when using large sets of data.

The gene for flower type, *Sh*, was found as expected in G8 and at a similar map position as in Ogundiwin et al. (2009), who mapped this gene for the first time, and as in Donoso et al. (2016). Its physical position on chromosome 8 is also compatible with the results of Micheletti et al. (2015) and Cao et al. (2016), both using genome-wide association analysis.

A major QTL (qMD4) explaining >80% of the phenotypic variance for maturity date was found on G4. A QTL in this position has been detected by other authors in a very broad transect of peach materials as well as in other *Prunus* species (Dirlewanger et al. 2012; Hernández Mora et al. 2017; Serra et al. 2017; Aranzana et al. 2019; Rawandoozi et al. 2021; Chen et al. 2021). In certain cases, this trait has been integrated in the maps as a major gene *MD*/*md* (Eduardo et al. 2011; Pirona et al. 2013). A strong candidate gene for this character is an NAC transcription factor (*Prupe.4G186800.1*), orthologous to the *Nor* gene in tomato (Pirona et al. 2013). The same region includes a gene (*Sr*/*sr*) responsible for another character, slow ripening, that determines the presence of individuals producing fruit that do not ripen, remaining immature on the tree for a long time (Eduardo et al. 2015; Núñez-Lillo et al. 2015). A 26.6 kb deletion containing the *Prupe.4G186800.1* gene was associated with the *sr* allele (responsible for the slow ripening phenotype) suggesting that *MD* and *Sr* may be the same gene (Eduardo et al. 2015; Meneses et al. 2016). Another major gene mapped in this region is *Alf*/*alf* that determines the formation of the thick mesocarp characteristic of the peach fruit in almond × peach progenies (Donoso et al. 2016). Additionally, this region of chromosome 4 contains several QTLs involved in the inheritance of other fruit traits, including FDP, which can be considered as an alternative measurement of MD (Hernández Mora et al. 2017), and many other characters including SSC, fruit weight, acidity, pH and red skin color (Eduardo e al. 2011; Hernández Mora et al. 2017; Rawandoozi et al. 2021). It is unknown whether the concurrence of genes involved in the inheritance of so many characters in this region is due to the action of a highly polymorphic single gene with a broad set of pleiotropic effects, or to a cluster of genes that would constitute a hotspot for a diverse set of characters in peach and other *Prunus*. In this work, we identified consistent QTLs for four characters in this region, including FDP, SSC, FW and LCS: in all cases, late maturity was associated with an increase of these traits except for leaf color at senescence where late maturity and presence of anthocyanin color were negatively correlated. The QTLs of all traits at this region were essentially additive (Table 3.1) and with a generally large effect (R^2 = 28.0-85.0%). Overall, this region appears to have a major

impact on the phenotype and is usually the determinant of critical aspects of agronomic characters that are affected by phenology and fruit-related traits. It deserves detailed exploration at the genetic level to understand its nature and how different haplotypes may shape expression of the traits involved in this highly polymorphic region.

Two consistent QTLs found for flowering time (qFT7) and beginning of shooting (qBS7) mapped at the same location, suggesting that they could correspond to the same locus. These two characters differed in two QTLs, qBS4, located at a different position to the MD-related QTLs, and qFT8, indicating that these characters may be encoded by a partly overlapping set of genes. Their positions also coincide with those found in other peach and other *Prunus* species (Fan et al. 2006; Silva et al. 2005; Dirlewanger et al. 2012; Hernández Mora et al. 2017) where they have been studied along with a cohort of other QTLs that depend on the specific population studied.

We identified a consistent QTL for fruit juice titratable acidity (TA) on G6 (qTA6), at a different position of the major gene determining subacid vs. acid fruit (*D*/*d*) that maps on G5 (Dirlewanger et al. 2006) and is homozygous for the allele (*d*) that confers fruit acidity in 'Earlygold'. This QTL was already detected with strong evidence in the multiple progeny analysis performed by Hernández Mora et al. (2017), confirming its value as a factor that may be used to modulate fruit acidity in the absence of the subacid allele.

The leaf color at senescence was analyzed here for the first time in peach and we found two loci, both with strong effects (R^2 = 28.6-40.5%), one on G4 (qLCS4) and the other on G3 (qLCS3). A major gene for red vs. green leaf color (*Gr*/*gr*) has been described in peach and mapped to G6 (Lambert and Pascal 2011), which discards its involvement in the variability observed in the 'Earlygold' progenies. The position of qLCS3 corresponds to that of the gene that determines yellow vs. anthocyanin anther color (*Ag*/*ag*; Arús et al. 1994), the "highlighter" gene that controls peach fruit skin red blush (*H*/*h*; Bretó et al. 2017), and to various QTLs determining anthocyanin color in fruit skin and flesh of several stone fruit crops (Sooriyapathirana et al. 2010; Socquet-Juglard et al. 2013; Donoso et al. 2016; Calle et al. 2021; Fiol et al. 2021). The phenotypic variability of these genes in *Prunus* has been attributed to the variation of one or several tandemly-duplicated transcription factor genes (*PpMYB10.1, PpMYB10.2* and *PpMYB10.3*) involved in the anthocyanin metabolic pathway (Tuan et al. 2015; Fiol et al. 2021). While the QTL on G3 has been generally detected as the main determinant in the color-related traits studied in *Prunus*, it is often accompanied by a QTL on

G4, corresponding to the position of qLCS4 found here, such as for skin color in peach, almond × peach and Japanese plum crosses (Frett et al. 2014; Donoso et al. 2016; Salazar et al. 2017; Hernández Mora et al. 2017) and flesh color in cherry (Calle et al. 2021). The inheritance of non-anthocyanic vs. anthocyanic senescing leaves in E×E appears to have an oligogenic inheritance, based on the combination of genotypes of these two QTLs and their interaction. qLCS3 had a dominant gene action and qLCS4 was additive (see Table 3.1) with alleles B and A, respectively, being responsible for absence of anthocyanin color. Individuals having one of these alleles or both in homozygosis, selected using the SNPs closest to the LOD peaks of the QTLs (SNP_IGA_344086 for qLCS3 and SNP_IGA_405773 for qLCS4), were usually (90.2% of the cases) non-colored or anthocyanic, as predicted by the markers.

One remarkable finding of this research is that using an F1 population, i.e. with the 'Earlygold' QTLs studied using a backcross type segregation, we only found a subset (6) of the 26 QTLs that were detected using an F2 progeny. There are two reasons for this important difference. First, trait segregation between peach and almond, corresponding to QTLs heterozygous in the 'Texas' × 'Earlygold' hybrid parent in the F1 progeny, produced often by alleles with greater relative effects than those segregating within 'Earlygold', may have interfered with the identification of QTLs at the same or different genomic locations, resulting in a loss of efficiency in the detection of 'Earlygold' QTLs. And second, heterozygous QTLs in both 'Earlygold' and the 'Texas' × 'Earlygold' hybrid parent would segregate 3:1 or 1:2:1 in the progeny, but they were analyzed with markers that segregated 1:1, resulting in a reduction of power to identify QTLs; heterozygous individuals for the QTL cannot be used for genetic analysis in this case, halving the effective population size. While F1 progenies between partly heterozygous parents are often used for QTL analysis in clonally propagated species, a more efficient QTL analysis can be done with other population types (particularly F2 progenies). In addition, F2 populations allow for the analysis of QTL action (dominance, additivity and overdominance), which is not possible with backcross populations. One possible way to rescue as much information as possible from F1 segregating progenies is analyzing QTLs in two steps, the first one as a backcross, followed by a second analysis using only 1:2:1 segregating markers. The latter analysis would detect F2 segregating QTLs with more precision, as the information from both parents will be used, resulting in an increase of their LOD values in addition to the possible identification of new QTLs not significantly detectable when using a BC1 progeny for the analysis.

One of the reasons for the analysis of the segregation of ‘Earlygold’ is that this cultivar has been used as the recurrent parent for the introgression line collection peach/almond currently under construction. An inbred line is typically used for that in the IL collections available, but in the absence of peach inbred lines of the major commercial gene pool with sufficient vigor, we opted for ‘Earlygold’ which, as we have shown here and in previous studies (Donoso et al. 2015), has large homozygous regions adding up to approximately half of its genome. Knowing the segregating QTLs of this genotype is important as the IL collection will be segregating for them, and their variability may interfere with the analysis of the characters segregating between peach and almond, which are the relevant ones in this case. Our results indicate that there is essentially one region of concern, that of chromosome 4 that contains the QTLs related with maturity date and affects many other fruit traits. Having this region with a genetic composition that results in a similar expected phenotype for all plants of the IL collection is a clear conclusion from the data presented here.

3.5 Supplementary Material

Table S3.1: Comparison of the two 'Earlygold' maps, 'Earlygold' F2 population (E×E) with the ('Texas' × 'Earlygold') × 'Earlygold' BC1 population (E) using the peach 9k SNP chip. Linkage group (G), physical position in base pairs (bp) in version 2.0 (v2), in brackets is percentage of the physical distance covered by each group in proportion to the actual distance of the respective group, distance in centi-morgans (cM), number (Nb.) of SNPs, markers and bins, number of gaps each more than 2 mega base pairs (Mb), distance between extreme SNPs of E×E that are common with E (SNPs E×E vs. E), distance between extreme SNPs of E that are common with E×E (SNPs E vs. E×E), difference between common extreme SNPs between E×E and E (E×E-E).

G	Physical length (E×E) bp v2	Physical length (E) bp v2	cM (E×E)	cM (E)	Nb. SNPs E×E	Nb. markers E[1]	Nb. bins E×E	Nb. bins E	Nb. E×E gaps >2Mb (Mb)	Nb. E gaps >2Mb (Mb)	cM between extreme common SNPs (E×E vs. E)	cM between extreme common SNPs (E vs. ExE)	Difference (cM) between common fragments ExE-E
G1	35,993,817 (75.1%)	35,926,707 (75.1%)	73.2	103	273	189	49	34	7 (32.0)	7 (30.7)	73.2	102.7	-29.5
G2	30,329,738 (99.7%)	30,329,738 (99.7%)	76.2	68	176	99	27	13	4 (19.1)	4 (19.3)	76.2	65.2	11
G3	26,642,299 (97.4%)	26,704,254 (97.8%)	61.9	69.6	192	146	36	31	2 (7.3)	2 (4.8)	61.9	68.7	-6.8
G4	10,214,022 (39.5%)	12,558,056 (48.4%)	37.1	56.6	284	174	29	30	1 (15.3)	1 (13.3)	37.1	41.3	-4.2
G5	15,454,720 (83.6%)	15,454,720 (83.6%)	40.5	27.1	42	35	20	9	3 (14.7)	2 (12.5)	40.5	27.1	13.4
G6	27,839,827 (90.5%)	27,839,827 (90.5%)	55.6	84.1	153	116	20	23	5 (20.8)	5 (21.2)	55.6	84.1	-28.5
G7	20,519,494 (91.9%)	21,884,075 (97.8%)	52.4	66.4	274	194	52	38	-	-	52.4	61.4	-9
G8	19,675,823 (87.1%)	19,675,823 (87.1%)	42.2	46.7	246	138	36	36	2 (7.2)	2 (7.2)	42.2	43.1	-0.9
Total	186,669,740 (82.7%)	189,743,200 (84.0%)	439.1	521	1640	1091	269	214	24 (116.4)	23 (109.0)	439.1	493.6	-54.5

[1] Markers include 1050 SNPs and 41 of other types (see Donoso et al. 2015)

Table S3.2: E×E phenotypic data with trait type, trait name, trait acronym, year, number of individuals (N), the parental (‘Earlygold'), individuals of the population [mean, maximum, minimum and standard deviation (SD)], and probability of the Shapiro-Wilk test for normality of the distribution; *non-normally distributed traits.

	Trait name	Acronym	Year	N	Earlygold	Mean	Max	Min	SD	Shapiro-Wilk
Phenology	Flowering time	FT	2017	70	-	59.17	65	56	2.5	<0.001*
			2018	73	-	64.01	72	56	3.19	<0.001*
			2019	74	59	61.35	69	56	2.76	<0.001*
	Beginning of shooting	BS	2020	71	58	53.13	58	50	3.48	<0.001*
			2021	71	58	54.07	58	48	3.36	<0.001*
	Maturity date	MD	2017	71	157	154.21	173	145	9.28	<0.001*
			2018	68	159	156.78	180	145	9.45	<0.001*
			2019	73	168	160	179	149	8.46	<0.001*
	Fruit development period	FDP	2017	69	-	95.06	114	83	9.84	<0.001*
			2018	68	-	92.79	115	80	9.87	0.001*
			2019	73	111	98.67	116	86	7.95	<0.001*
	Fruit production	FP	2017	74	3	3	4	1	0.68	<0.001*
			2018	74	3	2.53	3	1	0.65	<0.001*
			2019	74	3	3.65	4	1	0.58	<0.001*
	Beginning of leaf fall	BLF	2019	74	309	285.66	333	247	20.83	<0.001*
			2020	71	-	276.42	293	245	11.97	<0.001*
	End of leaf fall	ELF	2019	74	333	324.01	339	261	15.7	<0.001*
			2020	70	-	293.41	307	251	9.71	<0.001*

Table S3.2 (Continued)

	Trait name	Acronym	Year	N	Earlygold	Mean	Max	Min	SD	Shapiro-Wilk
Fruit	Intensity of the red skin color	ISC	2017	71	3	2.45	4	1	0.77	<0.001*
			2018	72	2	2.26	4	1	0.98	<0.001*
			2019	73	2	2.75	4	1	1.04	<0.001*
	Fruit weight	FW	2017	71	79	82.23	133	20	24.09	0.516
			2018	68	91	75.62	122	43	17.97	0.0575
			2019	73	81	83	130	44	17.03	0.2228
	Fruit firmness	FF	2017	69	1	0.96	2.9	0.5	0.45	<0.001*
			2018	68	1.1	0.96	1.4	0.5	0.26	<0.001*
			2019	73	0.82	0.87	1.7	0.55	0.23	<0.001*
	Soluble solid content	SSC	2017	69	10.6	12.21	15.6	9.3	1.47	0.0309
			2018	68	11.4	11.77	16.5	8.5	1.86	0.1679
			2019	73	12.2	12.87	17.4	9.7	1.74	0.0335
	Titratable acidity	TA	2017	69	6.19	4.84	6.94	2.69	1.03	0.4382
			2018	66	6.49	5.19	7.39	3.21	0.97	0.2657
			2019	73	6.49	5.53	9.1	3.13	1.12	0.5512
	pH	pH	2017	69	3.5	3.62	4.4	2.9	0.23	<0.001*
			2018	68	3.1	3.14	3.6	2.8	0.23	0.002
			2019	73	3.4	3.53	4	3.2	0.16	0.0124

Table S3.2 (Continued)

	Trait name	Acronym	Year	N	Earlygold	Mean	Max	Min	SD	Shapiro-Wilk
Leaf	Leaf perimeter	LP	2017	46	3666	3920.31	4805.28	3332.91	335.7	0.0808
			2018	74	4336	4359.75	5199.06	3812.58	300.91	0.0283
			2019	74	-	4493.82	5322	3785.89	331.06	0.8983
	Leaf Surface	LS	2017	46	381119	460968.23	710193.91	289612	83541.86	0.0856
			2018	74	565475	593571.37	828082.27	453899.73	84672.37	0.0068
			2019	74	-	616850.28	827254.91	448740.45	84850.21	0.4363
	Leaf blade width	LBW	2017	46	430	458.64	625.82	361	49.74	0.0394
			2018	74	486	511.11	604	429.27	36.73	0.2019
			2019	74	-	532.09	624	454.91	35.73	0.9615
	Leaf length	LL	2017	46	1558	1624.73	1978	1373.2	134.95	0.0694
			2018	74	1900	1877.95	2235.55	1640.82	133.97	0.0044
			2019	74	-	1878.12	2202.64	1576.91	138.25	0.8471
	Leaf blade length	LBL	2017	46	1439	1514	1884.55	1287.1	131.33	0.0188
			2018	74	1768	1765.87	2098	1533	125.3	0.0045
			2019	74	-	1755.32	2043.91	1478.36	126.68	0.7386
	Leaf shape	LSH	2017	46	3.34	3.31	3.81	2.79	0.21	0.9841
			2018	74	3.63	3.46	3.83	3.11	0.14	0.6403
			2019	74	-	3.3	3.68	2.93	0.15	0.3968
	Petiole length	PL	2017	46	119	110.73	204.4	62.2	26.42	0.0053
			2018	74	132	112.08	144.36	58.91	14.79	0.0101
			2019	74	-	122.8	158.73	80.36	15.63	0.1372
	Leaf dry weight	LW	2018	74	-	3.75	5.11	2.77	0.55	0.0027
			2019	74	-	4.52	6.59	3.13	0.71	0.2642
	Chlorophyll content	CC	2018	74	-	36.15	39.42	32.13	1.6	0.7838
			2019	74	-	38.72	46.63	27.68	4.03	0.1454

Table S3.2 (Continued)

Trait name	Acronym	Year	N	Earlygold	Mean	Max	Min	SD	Shapiro-Wilk
Leaf color at senescence	LCS	2019	74	2	1.53	2	1	0.5	<0.001*
		2020	71	-	1.52	2	1	0.5	<0.001*

Table S3.3: Correlation coefficients between different traits and years estimated by pairwise method. The number next to the trait name corresponds to the year the data obtained. Colors: green indicates a positive correlation and red a negative correlation.

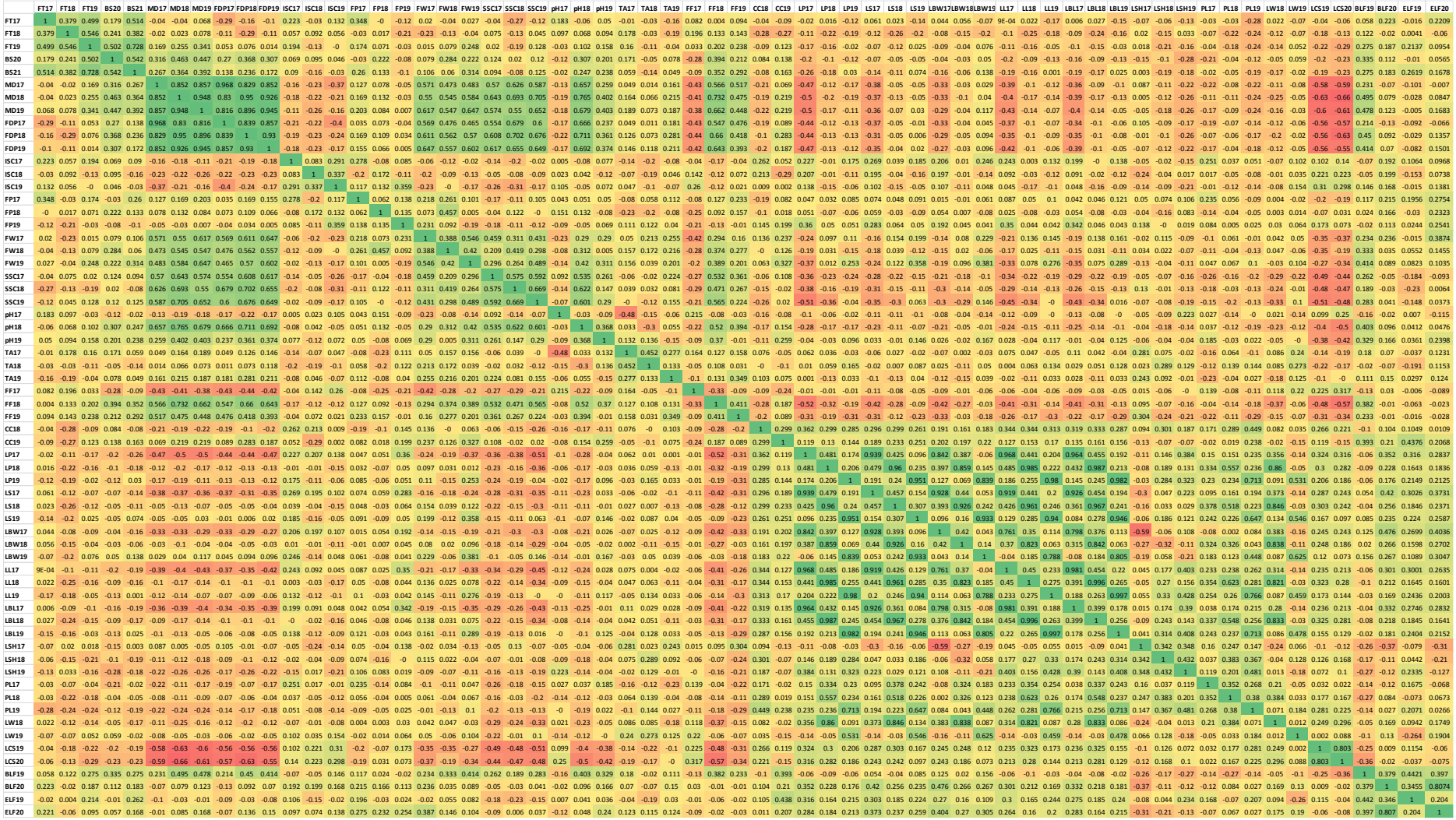


Table S3.4: QTLs identified in E×E map ('Earlygold' F2 population) and E map with the ('Texas' × 'Earlygold') × 'Earlygold' BC1 population. Trait category, acronym of the trait, map, year analyzed, linkage group (G), LOD score of the maximum peak, nearest marker, position of the maximum peak, physical position in base pairs, percentage of variance explained (R^2), average values of the trait, in female parent (A), heterozygotes (H), male parent (B) and other variables of additivity (a), dominance (d), (d/a), gene action (GA).

QTLs Phenology

Trait	Acronym	Map	Year	G	LOD	Nearest marker	Position (cM)	Position (bp v2)	R^2	A	H	B	a	d	d/a	GA
Flowering time*	FT	E×E	2017	7	7.12	SNP_IGA_775181	30.1	14002848	37.4	58.3	58.5	62.5	-2.07	-1.9	0.92	A
				8	3.26	SNP_IGA_862006	23.5	13825065	19.3	58.1	60.4	58.3	-0.12	2.24	-18.66	O
			2018	7	5.28	SNP_IGA_778125	36.9	15339643	28.3	61.5	64	66.7	-2.59	-0.11	0.04	A
				8	2.86	SNP_IGA_868544	26.9	15833068	16.5	63.4	65.4	62.5	0.46	2.5	5.43	O
			2019	4	4.9	SNP_IGA_381379	4.3	2337294	26.3	63	61.7	58.9	2.02	0.73	0.36	A
				7	8.98	SNP_IGA_778125	36.9	15339643	42.8	59	61.3	64.5	-2.75	-0.44	0.16	A
				8	3.78	SNP_IGA_831882	2.8	6560332	20.9	64	61.2	60.1	1.91	-0.85	-0.45	A
Beginning of flowering time	BFT	E	2011	7	2.5	SNP_IGA_769471	45.3	12223100	11	-	-	-	2.3	-	-	-
			2012	7	3.36	SNP_IGA_769471	28	12223100	8.7	-	-	-	2.84	-	-	-
End of flowering time	EFT	E	2011	7	3.62	SNP_IGA_776994	36.8	14942937	13.6	-	-	-	3.61	-	-	-
			2012	7	4.65	SNP_IGA_778138	40.8	15342839	14.3	-	-	-	3.07	-	-	-
			2013	7	2.5	SNP_IGA_776994	36.8	14942937	8.9	-	-	-	4.24	-	-	-

*traits common in E×E and E

A = additivity, D = dominance and O = overdominance

Table S3.4 (Continued)

QTLs Phenology

Trait	Acronym	Map	Year	G	LOD	Nearest marker	Position (cM)	Position (bp v2)	R2	A	H	B	a	d	d/a	GA
Beginning of shooting*	BS	E×E	2020	4	3.25	SNP_IGA_378053	4.3	1294398	19	54.6	53.7	50.5	2.06	1.09	0.52	D
			2021	4	4.42	SNP_IGA_375917	4.3	1117073	24.9	55.8	54.6	51.2	2.3	1.05	0.45	A
				7	9.27	SNP_IGA_778125	35.9	15339643	45.2	50.6	54.3	57.6	-3.46	0.15	-0.04	A
		E	2011	7	4.24	SNP_IGA_779386	41.4	15787610	12.1	-	-	-	3.73	-	-	-
			2012	7	7.51	SNP_IGA_779386	41.4	15787610	21.8	-	-	-	3.72	-	-	-
			2013	7	8.13	SNP_IGA_779386	41.4	15787610	23.5	-	-	-	7.03	-	-	-
Maturity date*	MD	E×E	2017	4	25.31	SNP_IGA_405773	37.1	9670782	80.6	166.5	151.2	146.2	10.17	-5.18	-0.51	D
			2018	4	24.91	SNP_IGA_405773	37.1	9670782	81.5	168.6	156.5	146.8	10.93	-1.23	-0.11	A
			2019	4	27.14	SNP_IGA_402256	34.3	8371579	82	170.5	158.8	151.3	9.62	-2.06	-0.21	A
		E	2012	4	4.92	BPPCT015	52.6	12558056	18.6	-	-	-	20.01	-	-	-
Fruit development period*	FDP	E×E	2017	4	24.52	SNP_IGA_405773	37.1	9670782	80.5	107.8	91.8	86.3	10.73	-5.26	-0.49	A
			2018	4	23.84	SNP_IGA_405773	37.1	9670782	80.1	105.1	92.5	82.4	11.32	-1.22	-0.1	A
			2019	4	30.12	SNP_IGA_405773	37.1	9670782	85	109	97.9	90.4	9.27	-1.76	-0.19	A
		E	2012	4	4.29	BPPCT015	52.6	12558056	16.9	-	-	-	18.98	-	-	-
Beginning of leaf fall*	BLF	E×E	2020	4	4.72	SNP_IGA_403152	35	8977975	25.5	291.5	293	269.6	10.97	12.45	1.13	D

Table S3.4 (Continued)

QTLs Fruit

Trait	Acronym	Map	Year	G	LOD	Nearest marker	Position (cM)	Position (bp v2)	R^2	A	H	B	a	d	d/a	GA
Fruit weight*	FW	E×E	2017	4	7.13	SNP_IGA_402192	33.6	8352011	37	103.3	78.6	66.6	18.32	-6.4	-0.35	A
			2018	4	4.85	SNP_IGA_405773	37.1	9670782	28	88.6	75.6	64.3	12.18	-0.87	-0.07	A
			2019	4	7.73	SNP_IGA_402192	33.6	8352011	38.6	95.6	85	68.8	13.37	2.82	0.21	A
Intensity of the red skin color*	ISC	E×E	2018	3	3.32	SNP_IGA_316047	21.5	7943434	19.1	2.3	2.5	1.3	0.48	0.65	1.35	D
			2019	3	5.07	SNP_IGA_898654	0	742389	27.4	3	3.1	1.9	0.58	0.61	1.05	D
				4	3.63	SNP_IGA_402192	32.6	8352011	20.5	2	3.1	2.8	-0.4	0.72	-1.8	O
				6	4.08	SNP_IGA_616119	3.7	1417402	22.7	1.8	3.2	2.4	-0.31	1.07	-3.45	O
Fruit firmness	FF	E×E	2018	3	3.07	SNP_IGA_325166	30.6	15430887	18.8	1	0.8	1	-0.01	-0.22	22	O
				4	10.1	SNP_IGA_405623	36.4	9640313	49.5	1.1	1	0.7	0.21	0.11	0.52	D
			2019	4	4.71	SNP_IGA_393777	22.9	5708791	25.7	1	0.8	0.7	0.14	-0.07	-0.5	A
Soluble solid content*	SSC	E×E	2017	4	8.77	SNP_IGA_405773	37.1	9670782	44.3	13.5	12	11	1.23	-0.28	-0.22	A
			2018	4	9.5	SNP_IGA_405773	37.1	9670782	47.4	13.6	11.5	10.3	1.62	-0.45	-0.27	A
			2019	4	9.51	SNP_IGA_405623	36.4	9640313	45.1	14.2	13	11.3	1.45	0.2	0.14	A
				6	3.14	SNP_IGA_688103	52.9	26118989	18	14.1	12.2	12.9	0.55	-1.25	-2.27	O
Titratable acidity*	TA	E×E	2017	4	3.28	SNP_IGA_381379	6.7	2337294	19.6	5.2	4.9	4	0.62	0.36	0.58	D
				6	6.43	SNP_IGA_691838	55.6	27308224	34.9	3.8	4.9	5.6	-0.9	0.24	-0.26	A
			2018	6	3.79	SNP_IGA_691838	55.6	27308224	23.2	4.5	5.2	5.8	-0.67	0.09	-0.13	A
			2019	6	3.04	SNP_IGA_679199	48.1	23496723	17.4	4.7	5.6	6.1	-0.67	0.14	-0.21	A
pH	PH	E×E	2018	4	7.68	SNP_IGA_405773	37.1	9670782	40.5	3.3	3.1	2.9	0.18	0.007	0.04	A
				6	3.51	SNP_IGA_688103	52.9	26118989	21.2	3.3	3	3.1	0.1	-0.15	-1.5	D
			2019	3	4.33	SNP_IGA_339586	29.9	17546326	23.9	3.5	3.4	3.6	-0.05	-0.14	2.8	O
				4	4.72	SNP_IGA_403152	35	8977975	25.8	3.5	3.5	3.4	0.07	0.11	1.57	O

Table S3.4 (Continued)

QTLs Leaf

Trait	Acronym	Map	Year	G	LOD	Nearest marker	Position (cM)	Position (bp v2)	R^2	A	H	B	a	d	d/a	GA
Chlorophyll content*	CC	E×E	2018	3	3.39	SNP_IGA_356344	49	22328362	19	35	36.6	35.9	-0.47	1.15	-2.45	O
			2019	3	2.87	SNP_IGA_354950	46.3	22093557	16.3	37	40.1	36.7	0.12	3.28	27.33	O
Petiole length*	PL	E×E	2018	7	3.46	SNP_IGA_782427	42.9	17205367	19.4	123.4	111.3	103.6	9.88	-2.21	-0.22	A
Leaf length*	LL	E×E	2017	4	4.13	SNP_IGA_404059	35.7	9220854	33.8	1534	1604.7	1746.3	-106.16	-35.38	0.33	A
			2018	7	3.62	SNP_IGA_782427	42.9	17205367	20.2	1992.2	1859.7	1819.7	86.22	-46.27	-0.53	D
Leaf blade length*	LBL	E×E	2017	4	3.97	SNP_IGA_404059	35.7	9220854	32.8	1439.6	1486.3	1635	-97.72	-51.01	0.52	D
		E×E	2018	7	3.34	SNP_IGA_782427	41.9	17205367	18.8	1867.6	1746.2	1714.4	76.57	-44.73	-0.58	D
Leaf shape*	LSH	E×E	2018	7	3.57	SNP_IGA_78247	41.9	17205367	19.9	3.7	3.2	3.1	0.3	-0.16	-0.53	A
Leaf perimeter*	LP	E×E	2017	4	5.19	SNP_IGA_404059	35.7	9220854	40.5	3664.3	3872.2	4247	-291.31	-83.42	0.29	A
		E×E	2018	7	3.16	SNP_IGA_782427	42.9	17205367	17.9	4601.8	4320.4	4238.5	181.65	-99.7	-0.54	D
Leaf surface*	LS	E×E	2017	4	3.79	SNP_IGA_404059	35.7	9220854	31.6	409103	447297	534716	-62806.5	-24612.8	0.39	A
Leaf color at senescence	LCS	E×E	2019	3	5.01	SNP_IGA_344628	35.3	19101956	26.8	1.6	1.6	0.9	0.33	0.35	1.06	D
				4	8.48	SNP_IGA_405773	37.1	9670782	41	1	1.5	1.9	-0.41	0.01	-0.02	A
			2020	3	5.19	SNP_IGA_343432	35.3	18681162	28.6	1.5	1.6	0.9	0.3	0.42	1.4	D
				4	7.88	SNP_IGA_405623	36.4	9640313	40	1.1	1.5	1.9	-0.4	-0.006	0.02	A

Figure S3.1: Mapping the corresponding bins between the E×E and E maps for each linkage group (G). The lines between the SNPs indicate the bin of one map identical to the bin of the corresponding map.

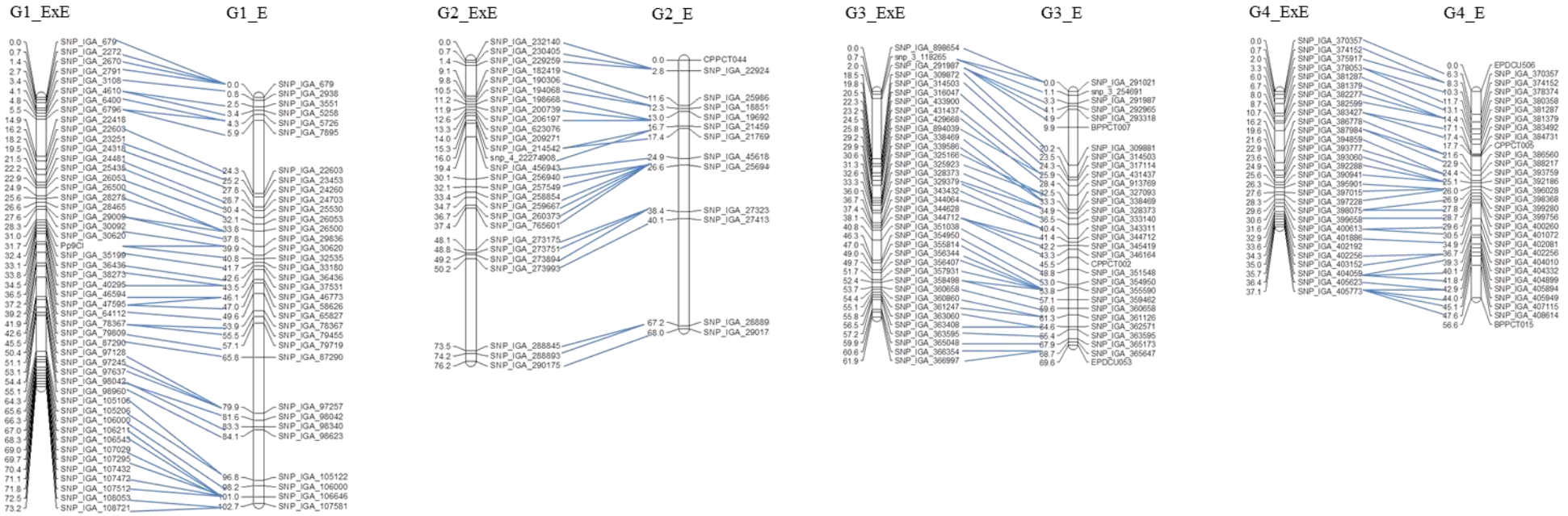
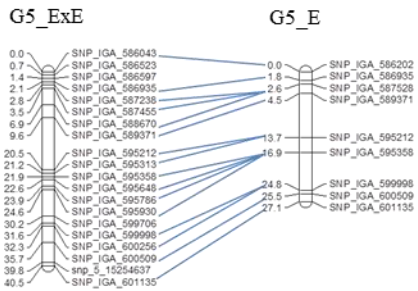
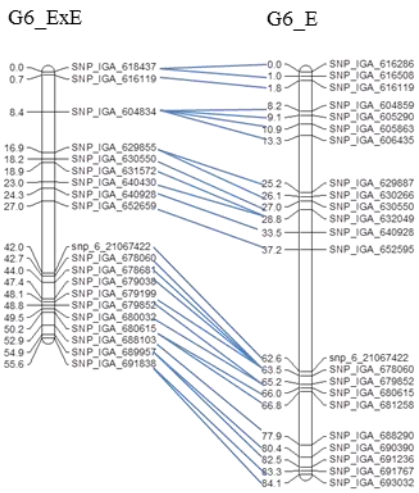
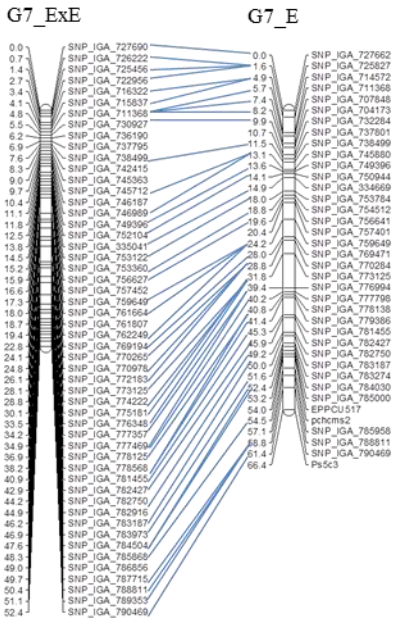


Figure S3.1 (Continued)







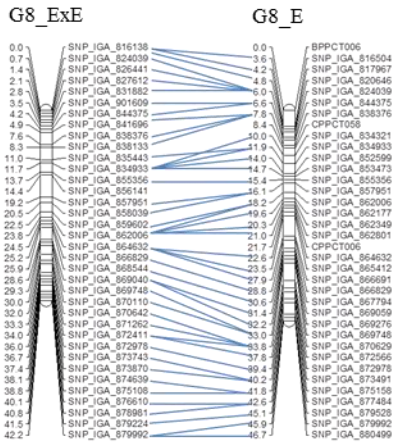


Figure S3.2: Histograms of E×E phenotypic data from 2017-2021. E indicates the value for 'Earlygold' and the names of the traits are indicated below each histogram.

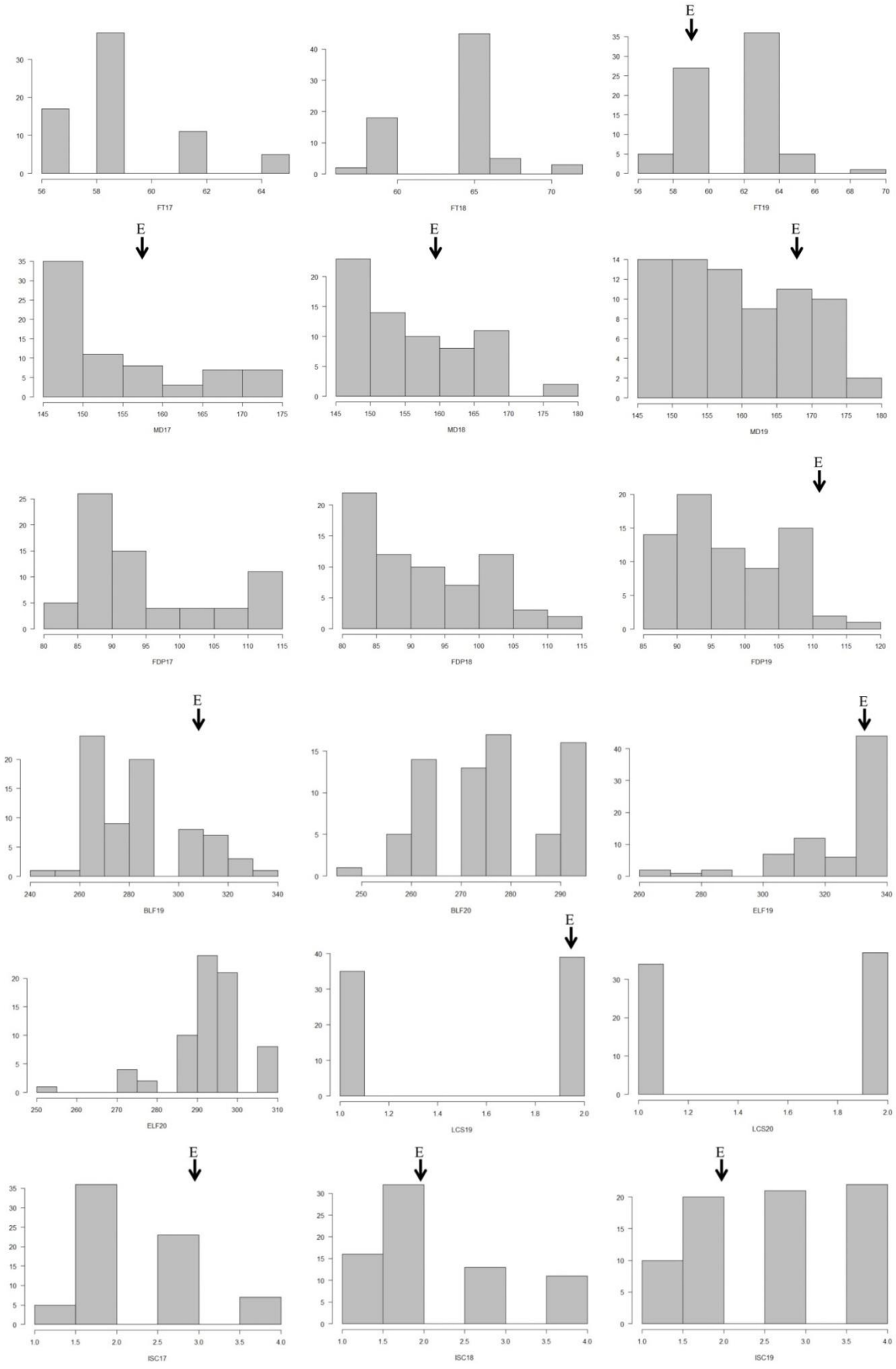


Figure S3.2 (Continued)

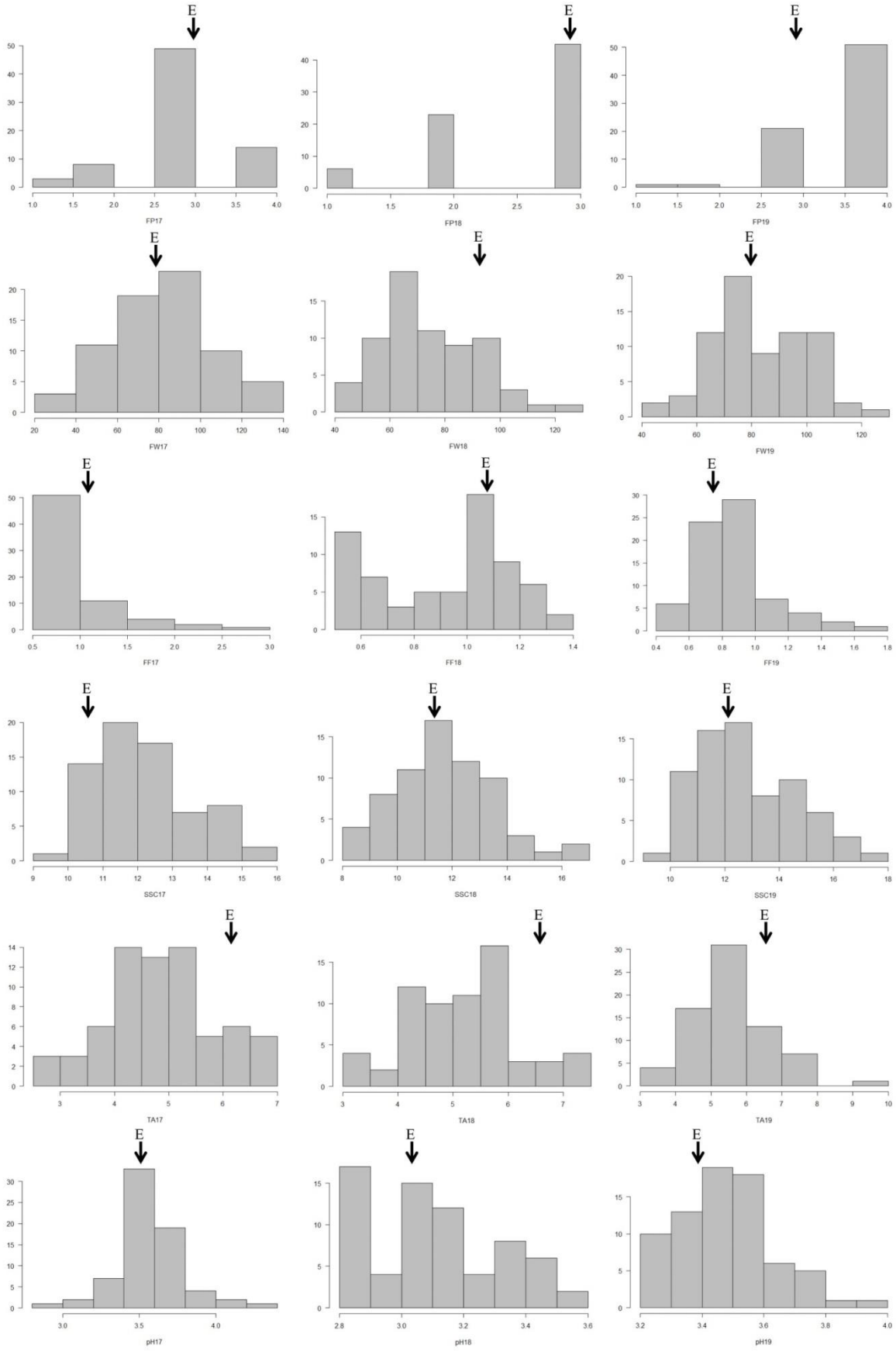


Figure S3.2 (Continued)

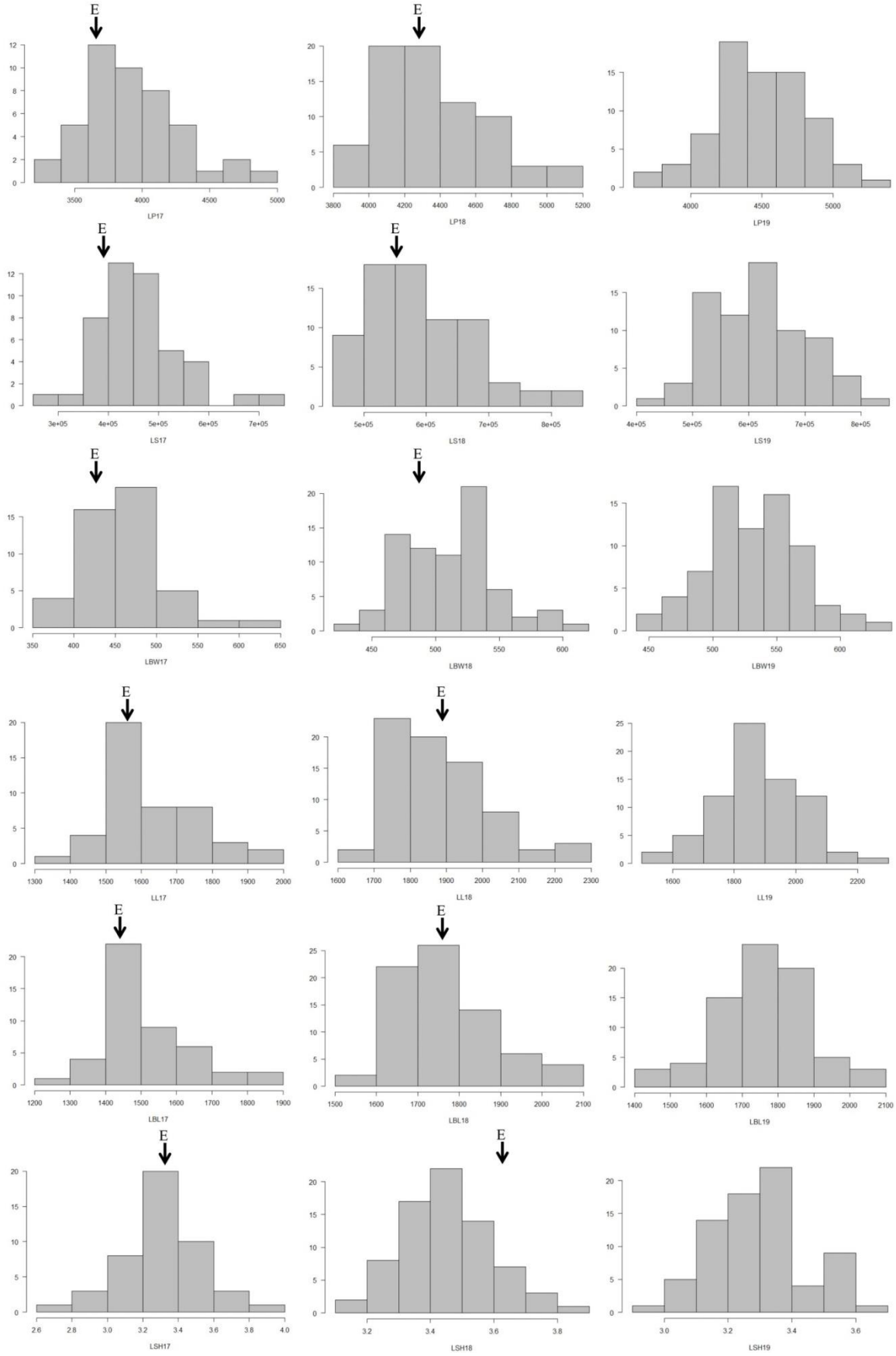


Figure S3.2 (Continued)

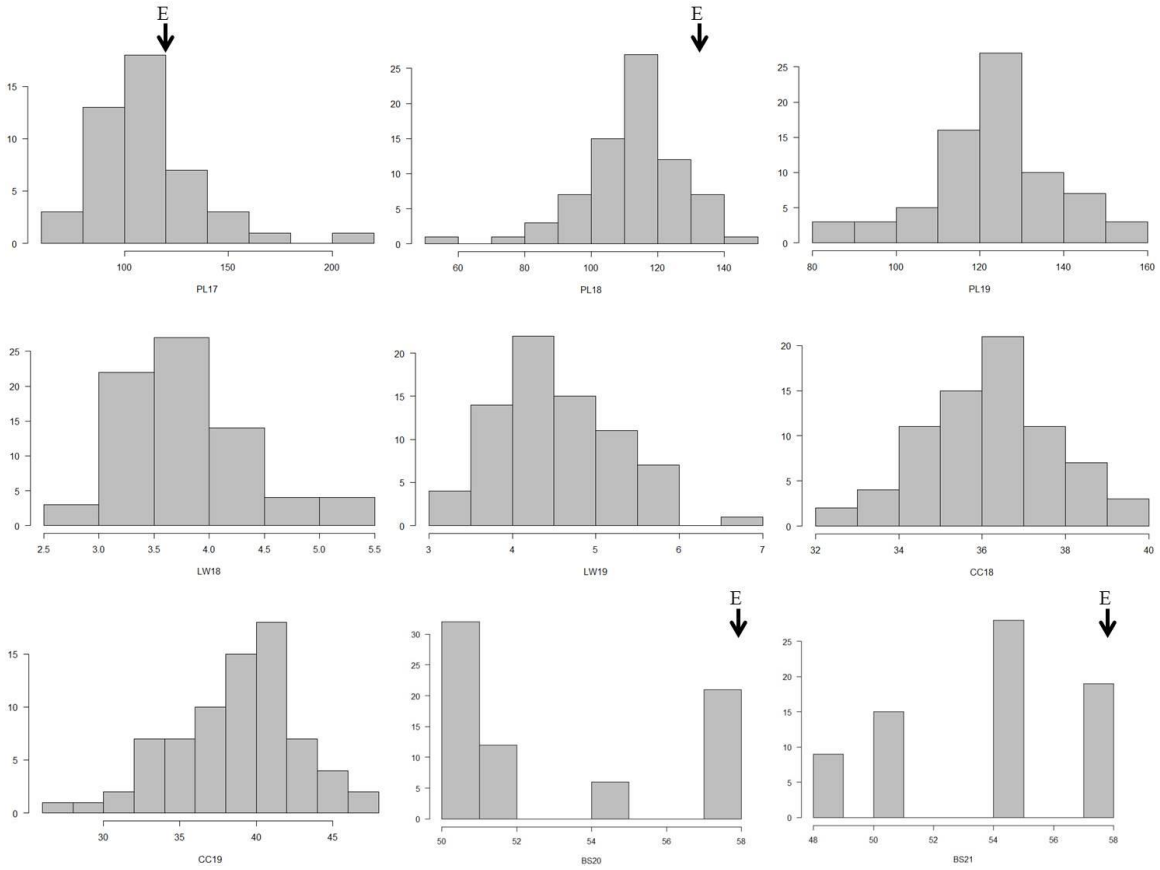


Figure S3.3: Graphical representation of all the QTLs identified from 2017-2021. Linkage group (G) name is indicated on top of each map. QTL names indicate trait name, year scored and percentage of explained variation. QTL bars indicate the LOD-1 interval.

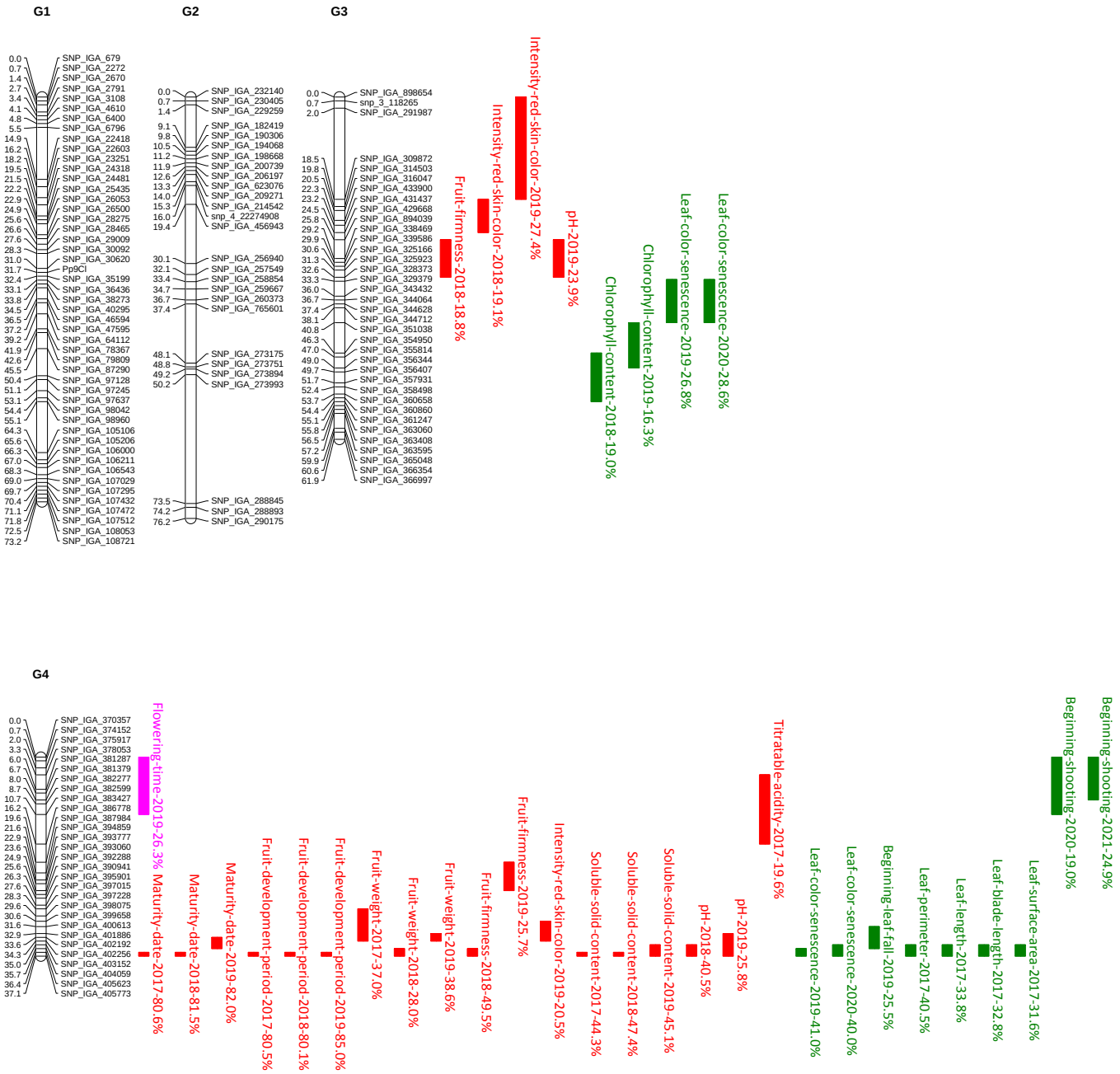
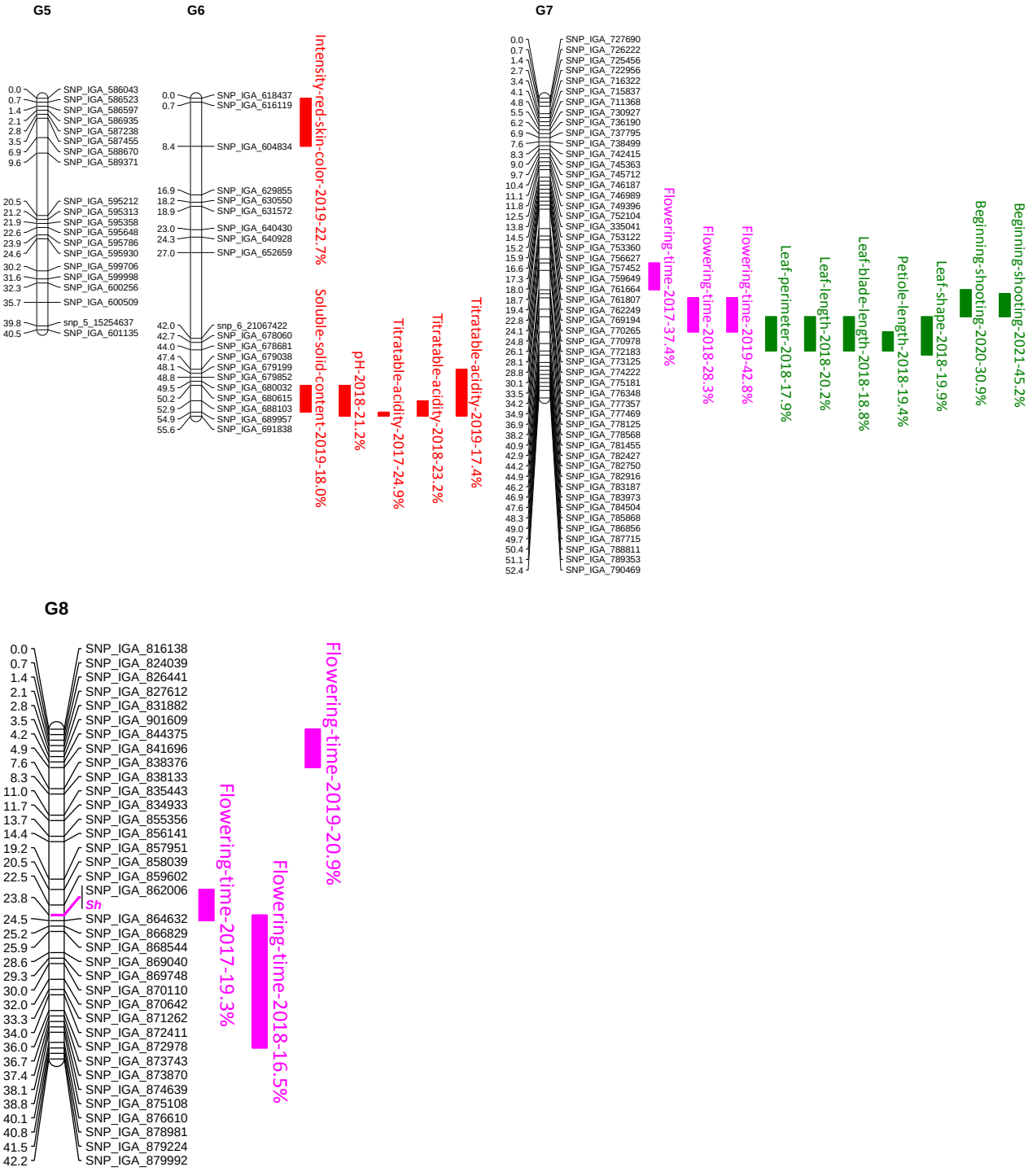


Figure S3.3 (Continued)



4. Construction of a collection of introgression lines of ‘Texas’ almond DNA fragments into the ‘Earlygold’ peach genetic background

Abstract

Peach [*Prunus persica* L. Batsch] is one of the major temperate fruit tree species, the commercial materials of which are known for having a low level of genetic variability. Almond [*P. dulcis* (Mill) DA Webb], a close relative of peach cultivated for its kernels, has a much higher level of diversity. The species are inter-compatible and often produce fertile hybrids, almond being a possible source of new genes for peach that could provide rusticity, adaptation to climate change, biotic stress resistance and even certain desirable fruit quality attributes. In this chapter we describe the development of a collection of peach-almond introgression lines (ILs), lines with a single fragment of almond (cv. Texas, syn Texas Prolific, syn. Mission) in the peach background (cv. Earlygold). Lines with few introgressions were selected with markers from two consecutive backcross progenies from a 'Texas' × 'Earlygold' F1 hybrid to obtain a small collection of trees with two or three introgressions of the second backcross progeny to 'Earlygold' (E2T). ILs were initially selected using a set of SSRs in all steps of this process, mainly in the selfed offspring of E2T, resulting in an initial set of 283 lines. A subset of these lines with full coverage of the almond genome were additionally genotyped with the 18k peach SNP chip, allowing for the final extraction of 67 lines, 39 with almond heterozygous introgressions covering 99% of the genome, and 28 with homozygous introgressions covering 83% of the genome. As a proof of concept, four major genes and four quantitative characters were examined in the selected ILs giving results generally consistent with previous information on the genetics of these characters. This collection is the first of its kind produced in a woody perennial species and promises to be a valuable tool for genetic analyses, including dissection of quantitative traits, positional cloning, epistasis and as prebreeding material to introgress almond genes of interest into the peach commercial gene pool.

4.1 Introduction

Introgression line (IL), near-isogenic line (NIL) or chromosome segment substitution line (CSSL) collections are sets of lines containing a single fragment of a donor genome in the background of a recurrent genome. These collections consist of tens to a few hundred lines, each with a different introgressed donor fragment, the sum of which cover most, or ideally all, the donor genome (Zamir 2001; Balakrishnan et al. 2017). They are usually developed by backcrossing to the recurrent parent in the initial generations, and selfing in the advanced backcross generations, using a set of markers with good genome coverage to identify the individuals of interest.

IL collections serve as valuable tools for genetics and breeding applications, as in the dissection of complex genetic traits (Eshed and Zamir 1995; Szalma et al. 2007; Szymański et al. 2020; Zhang 2021) and the fine mapping and positional cloning of genes or QTLs (Frary et al. 2003; Fridman et al. 2004; Xu et al. 2017). They are suitable for the study of the effects of specific QTLs in different environments and different genetic backgrounds (Balakrishnan et al. 2020), and the analysis of interallelic and epistatic interactions, as lines with different allelic dosages of one or more QTLs can be created by crossing selected ILs between them or with other lines (Gur and Zamir 2015). Given that ILs are usually constructed with a background of elite commercial lines, they provide an optimal resource to analyze the effects of QTLs of exotic materials in the breeding materials, which are masked in other populations used for genetic analysis by interactions with other loci of exotic origin. This makes IL libraries a direct source of improved materials for plant breeding as well as an invaluable tool for the evaluation and introgression of useful genes from wild or exotic materials in the cultivated gene pool, facilitating the use of the variability stored in exotic materials that can compensate for the loss of diversity resulting from domestication (Zamir 2001). The first IL collection was produced from a cross between tomato and the wild species *Solanum pennellii* (Eshed and Zamir 1992). Since then, IL collections have been extensively developed and used in many model and cultivated species, including most staple crops and horticultural herbaceous species (Balakrishnan et al. 2017). No examples exist for tree species, due to their long intergeneration periods that makes the process of IL generation extremely long.

Peach was used as the recurrent parent and almond (*P. dulcis*) as donor, sexually compatible species that originated from a common ancestor in central Asia about 5 Mya (Yu et al. 2018; Alioto

et al. 2020). Peach was domesticated in China ~5000 years ago (Faust and Timon 1995), while almond domestication is still unclear, although it probably occurred somewhere between the steppes of central Asia, and to the west of the Himalayas and the eastern Mediterranean shores (Ladizinski 1999). Peaches are cultivated for their fleshy mesocarp, as are other *Prunus* stone fruit species such as apricot, cherry, and plum, and almonds are cultivated for their seed: both are of major economic importance.

One of the key biological differences between peach and almond is that peach is self-compatible and has low levels of genetic variability, while almond is self-incompatible with a highly diverse genome (Byrne et al. 1990; Velasco et al. 2016). Almond appears to be a good source of new alleles that could provide useful variability conferring adaptation to climate change and disease and pest resistance in peach. Various studies have been undertaken to understand the inheritance of almond variability in the peach background using the offspring of a 'Texas' almond × 'Earlygold' peach F1 plant (MB1.37), selfed (the T×E population) and backcrossed to 'Earlygold' (the T1E population) (Donoso et al. 2015, 2016). A large backcross progeny was initially used to demonstrate the feasibility of a marker-based method (Marker-Assisted Introgression; MAI) to produce ILs from this interspecific hybrid in only two backcross generations, or one backcross plus one selfing generation (Serra et al. 2016). This was necessary to show that introgression from a distant source was possible within a reasonable timeframe, considering the long intergeneration period of tree species. Among these lines, those carrying a gene for resistance to peach powdery mildew from almond have already been incorporated in the IRTA peach breeding program (Marimon et al. 2021). Based on the T1E selections, Serra et al. (2016) extracted some ILs and developed a collection of individuals with two or three introgressions with the peach cytoplasm in a second backcross with T1E individuals (the E2T set). The objective here was to complete this work and develop and describe a nearly complete peach-almond IL collection, developed with the almond introgressed fragment in both heterozygosis and homozygosis, given that peaches can be easily reproduced by grafting. This collection is planned as a resource for fruit tree and *Prunus* researchers and breeders to better understand the genetics of valuable traits and to facilitate the use of almond variability to breed improved peach varieties.

4.2 Materials and Methods

4.2.1 Plant materials

In previous research, the F1 of the cross between almond 'Texas' used as female parent crossed with peach 'Earlygold' as pollen donor, named 'MB1.37', was backcrossed to 'Earlygold' as male parent, in the winter of 2006, 2007 and 2008 (Serra et al. 2016). A large offspring (N=1,095) was obtained, referred to as the T1E population, and a few (N=18) individuals, the pre-introgression line (prIL) set, carrying 2-4 almond introgressions were selected with simple-sequence repeat (SSR) markers covering the whole peach genome (Serra et al. 2016). Donoso et al. (2016) analyzed the inheritance of various agronomic traits with a group of N=190 T1E plants, most selected at random from the T1E initial population, but also including those of the prIL set. Due to the cytoplasmic male sterility conferred by the almond cytoplasm detected in the T1E population (Donoso et al. 2015), an additional backcross was performed using 'Earlygold' as the female parent (E2T; N=160), to ensure that the plants obtained were pollen-fertile. The T1E plants used as staminate parents to generate the E2T offspring were chosen to contain a low number of introgressions and to ensure full almond genome coverage, including mostly plants from the prIL set. A subset of 37 marker-selected plants of the E2T progeny to contain only two or three almond introgressions, the E2T set, was selected by Serra et al. (2016). In this work, the selfed (E2TS1) or backcrossed to 'Earlygold' (E3T) progeny of trees from the E2T set were used for this last step of IL extraction, where our objective was to obtain a collection of individuals with a single almond introgression, in both homozygosis and heterozygosis, and covering the full almond genome. The breeding scheme for IL extraction is given in Figure 4.1.

Plants from parents, hybrid and the T1E generation are kept in the experimental fields of IRTA at Cabrils and Gimenells (Catalonia, Spain) as described in Serra et al. (2016). The E2T set was grown at two different IRTA stations: in Caldes de Montbui grafted to 'Garnem' rootstock and in Mollerussa on their own roots. The selected ILs were planted at Caldes de Montbui on their own roots. For the E2T set and ILs the spacing between trees and rows is 2.5 × 4.0 m.

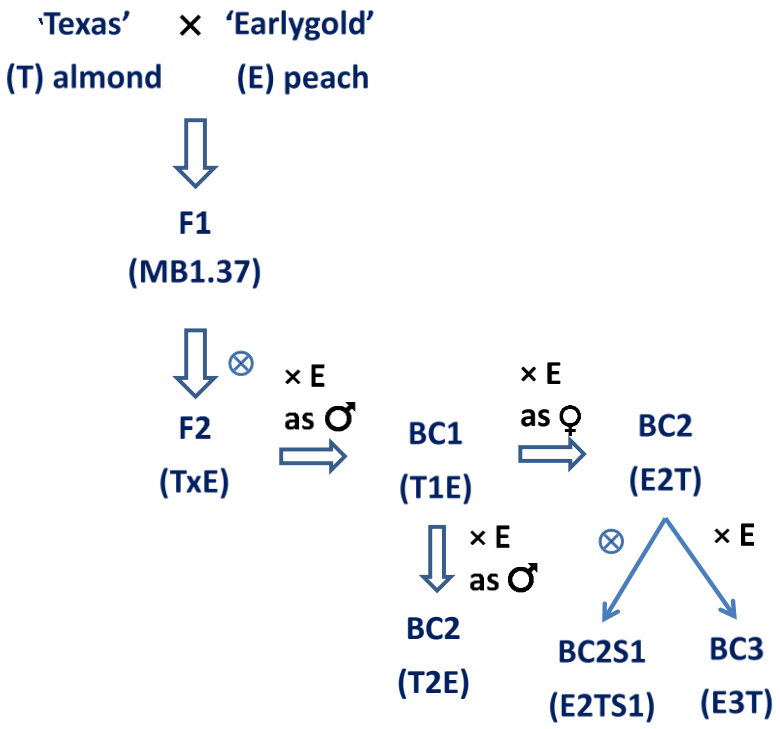


Figure 4.1: Breeding scheme for the extraction of introgression lines from the almond × peach cross.

4.2.2 *In vitro* embryo rescue

As the recurrent parent 'Earlygold' is an early maturing variety (May-June), most of the E2T individuals ripened early (June-July). The embryos of these individuals were not completely developed at fruit maturity time and *in vitro* embryo rescue was necessary to be able to germinate all seeds collected. The method used was based on previous work (Ramming 1990). The fruits were first immersed in a disinfectant solution of water and NaOCl (3.7%), thoroughly washed with water, then opened using a nutcracker to collect the seeds. The seeds were surface sterilized in a solution of 300 ml of NaOCl (3.7%), 700 ml H_2O and two drops of tween for 10 min, then washed with autoclaved water in a laminar flow hood. Seeds were opened and the embryos extracted and placed in test tubes with a solution of sucrose (30 g/l), Duchefa M0220 (2.46 g/l) and plant agar (8 g/l), with the addition of 10 ml of 100 µM benzyl aminopurine for seeds <5mm. These were kept in a cold chamber at 4ºC for 8-12 weeks or until the onset of radical growth, then transferred to a dark chamber (closed in carton box) at 23ºC for one week. After a week in the dark, the embryos were exposed to gradually increasing light conditions. Those plants with well-formed shoots and roots were moved to trays in the greenhouse, and finally to the field.

4.2.3 Genotyping and introgression line extraction

Genomic DNA was extracted from young leaves using the CTAB method (Doyle and Doyle 1990), in 96 well plates. Genotyping data from the E2T set were available for 113 SSRs with almost full genome coverage of the peach genome, spanning 212.2 Mb (359.3 cM), with the average interval between the markers being ~ 1.9 Mb (3.2 cM) (Serra et al. 2016; see Table S4.1 and Figure S4.1). Additionally, data from the same SSRs and the 9k peach SNP chip (Verde et al. 2012) had been obtained by Donoso et al. (2015) in the parents, the F1 and the T1E individuals used for mapping. To select the ILs in the progenies derived from plants of the E2T set we used the same 113 SSRs. The two markers at the extremes of the introgressions of each parent (Table S4.1) were genotyped in the corresponding E2TS1 and E3T progenies, and the individuals with a single introgression in homo or heterozygosis were selected. For plants having a recombination within one of the introgressed fragments, additional SSRs within this fragment were genotyped until its extremes were identified. Selected individuals were then genotyped with the additional SSRs with coverage of the complete genome.

SSRs were analyzed in the ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) by capillary electrophoresis following the PCR amplification method described in Aranzana et al. (2003). For the PCR for SSR amplification, the reaction mix (10 µl) contained 2 µl genomic DNA (200 ng), 1 µl of 10x NH_4 reaction buffer, 0.3 µl of 50 mM NH_4 $MgCl_2$, 0.2 µl of 10 mM dNTP, 0.2 µl of 10 µM forward primer labeled with fluorochrome (FAM, NED, VIC and PET), 0.2 µl of 10 µM reverse primer, 0.2 µl of 5 U lab Taq and 5.9 µl HPLC water. PCR was performed in a thermocycler (Applied Biosystems, USA) with the following conditions: a single cycle of initial denaturation at 94°C for 1 min, 35 cycles of denaturation at 94°C for 15 s, annealing step between 50 to 65°C for 15 s and an extension step at 72°C for 30 s. One cycle of final elongation was at 72°C for 5 min and at the final step, the PCR product was held at 4°C to complete. For capillary electrophoresis (Applied Biosystems ABI PRISM® 3130xl Genetic Analyzer), 2 µl of PCR product was mixed with 12 µl of formamide, 0.35 µl GeneScan500 LIZ (Applied Biosystems, USA) and denatured at 94°C for 3 min. Allele sizes were identified with the GeneMapper 5.0 software.

A selected set of 135 lines with a single introgression based on their SSR genotype, to cover the largest possible region of the genome, and prioritizing those lines covering whole chromosomes and lines that were already producing fruits, were genotyped with the 18k Illumina chip (Gasic et

al. 2019). For that, 1,000 ng genomic DNA was dried using a speed vacuum and eluted in 20 µl water to a final concentration of 50 ng/µl. DNA purity was checked by Nanodrop absorbance values, with ratios of 1.8 for 260/280 and 2.0 – 2.2 for 260/230 being considered pure. The samples were genotyped at Fondazione Edmund Mach facility, San Michelle all 'Adige, Italy.

The raw data from the genotyping platform was processed using Genome Studio Illumina 2.0. The two output files, final report and DNA report from the Genome Studio analysis were used as input files for genetic analysis using the ASSisT software (Di Guardo et al. 2015), an automatic SNP scoring tool, along with a generated pedigree file and map file. The population was set to germplasm, allowed missing data range to 1.0 and all the other parameters (call rate tolerance, p-value segregation distortion, unexpected genotype threshold per individual and frequency rare allele) set to zero. This way, 41.3% of the SNPs were approved, which was the highest compared to the default and all other settings we tried. We separated the SNPs in two sets, one that was heterozygous in the MB1.37 hybrid and homozygous in 'Earlygold', which was used for identification of the introgressed almond fragments, and the other with the remaining SNPs, i.e., heterozygous in 'Earlygold' irrespective of their genotype in the hybrid, which served to identify the genotype of the 'Earlygold background. Given that peach varieties are known to have large regions identical by descent, chromosomal fragments ≥2.5 Mb long (approximately 1% of the peach genome) with >3 in total or >2 consecutive SNPs in heterozygosis, were considered identical by descent and consequently with two identical haplotypes for 'Earlygold'. SNP data from IL genotyping have been graphically represented with GGT2 (Van Berloo 2008), using the physical location of the markers in the Lovell peach genome reference sequence v2.0 (Verde et al. 2017) as their positional reference.

To identify the ILs obtained individually, we used an internal code to select an initial collection with SSRs (see Tables S4.2 and S4.3 and Figures S4.2 and S4.3). For the final IL collection using the data of the 18 k SNP chip, we coined a new terminology: first a group of five letters and a number (the first four letters "PAIL" to indicate Peach-Almond Introgression Line, the fifth letter is E or O indicating heterozygous or homozygous ILs, respectively) and a digit 1-8 corresponding to the chromosome number where the introgression is located. The second part of the IL name is a dash followed by four digits corresponding to the position in Mb of the two extremes of the introgressed fragments. For example, line PAILO5-1219 is a homozygous peach-almond

introgression line located on chromosome 5 and in the region spanning from the 12th to the 19th Mb. A small letter at the end of the IL indicates lines with the same fragment, e.g., PAILE2-0130a and PAILE2-0130b.

4.2.4 Phenotyping

The IL collection was evaluated for traits known to have different phenotypes based on the previous information of 'Texas' × 'Earlygold' progenies (Donoso et al. 2016). These include several major genes and a few major quantitative trait loci (QTLs). The major genes were: juiciness (*Jui/jui*) as the presence and absence of juice in the fruits at maturity; blood flesh color (*DBF2/dbf2*) as red or yellow flesh color at maturity; maturity date (*MD/md*) as early or late maturing based on the number of Julian days when more than half of the fruits reach ripening stage, estimated on parameters such as fruit firmness and visual observation of fruit skin color change; and powdery mildew resistance (*Vr3/vr3*), scored as resistant or susceptible based on the absence or presence of fungal infection on the leaves. A quantitative leaf character, petiole length (PL), was measured in mm with a ruler in three to six average size leaves of each IL. We measured three additional quantitative fruit traits, taking data from three samples (average of 3-4 fruits per sample) per tree: fruit weight (FW), titratable acidity (TA) and soluble solids content (SSC), measured as described in Donoso et al. (2016). Statistical analysis of the data for quantitative traits was performed using the DescTools package of R. The overall data were studied with a one-way analysis of variance, and Dunnet's test used to compare the mean of each IL with that of the 'Earlygold' control.

4.3 Results

4.3.1 Development of a peach-almond IL collection

A total of 8,467 fruits from different generations of the 'Texas' × 'Earlygold' cross were obtained from 2011 to 2020 and used to construct this IL collection (Table 4.1, Figure 4.1). Results to 2015 have been reported earlier (Serra et al. 2016). Since that date, 4,916 fruits from the offspring of the E2T set were used for *in vitro* embryo rescue, giving rise to 1,276 seedlings (Table 4.1), as

described here. Table S4.1 and Figure S4.1 show the genotypes for 113 SSRs in the plants of the E2T set used as parents for the development of the ILs. Using these SSRs we extracted 146 ILs (97 in heterozygosis and 49 in homozygosis) that, along with the 137 lines (109 in heterozygosis and 28 in homozygosis) previously selected (Serra et al. 2016), resulted in a full set of 283 ILs covering the complete genome of almond 'Texas' in the 'Earlygold' background in heterozygosis (206 lines; Figure S4.2) and 94% of the almond genome in homozygosis (77 lines; Figure S4.3). Of the 206 heterozygous ILs, 65 had unique introgressed fragments and in 141 these fragments were in common with other lines (Figure S4.2), whereas for the 77 homozygous ILs, 29 were unique and in the remaining 48 the introgressed fragments were in common with other lines (Figure S4.3). The number of lines covering each linkage group in the heterozygous ILs were G1 (20), G2 (27), G3 (21), G4 (19), G5 (24), G6 (60), G7 (16) and G8 (19). For the 77 homozygous ILs, almond fragments were located on G1 (11), G2 (12), G3 (3), G4 (8), G5 (8), G6 (25), G7 (5) and G8 (5). The missing regions in homozygosis were the ends of G3 and G8 (Table 4.2). One of the lines (37P18-44) had two introgressed fragments, one in heterozygosis (G1) and another in homozygosis (G2). For all chromosomes in the heterozygous ILs and for chromosome 2, 4 and 6 in homozygous ILs, at least one of the ILs spanned the whole chromosome distance covered by the markers used.

4.3.2 Selection of an IL collection with 18k SNP genotyping

We selected 135 ILs from the total collection to be genotyped for the 18k SNP chip, including 81 heterozygous ILs (100% almond genome coverage) and 54 homozygous ILs (94% coverage). A total of 6,624 SNPs were identified with the appropriate segregations. Markers with 'Earlygold' and the MB1.37 hybrid in homozygosis (417) and with missing data in either of them (85) were discarded. The remaining 6,122 SNPs nearly covered the entire peach genome (98.4%), with an average density of one SNP every 37kb. Approximately half (3,080) of these SNPs were homozygous for 'Earlygold' and heterozygous for MB1.37 and were used to determine the positions of the almond introgressions. The remaining 3,042 SNPs, which were heterozygous for 'Earlygold and with any genotype for MB1.37, were used to establish the composition of the 'Earlygold' background.

From the 135 lines analyzed with the 18k SNP chip, four did not produce reliable results, three did not detect any introgressed fragments and of the remaining 128 lines, 39 were selected for the IL

collection in heterozygosis, 38 with a single DNA introgressed fragment from almond and one with two fragments, a major one on G1 and the other on G8. The genome coverage was 99%: including all but one small almond DNA fragment at the proximal end of chromosome 2 (1.35 Mb) that could not be detected only with the SSRs used (Figure 4.2, Table 4.2 and Table S4.2).

Table 4.1: Introgression lines (ILs) selected with SSRs from 2011-20 and their origin.

Year	fruits	seedlings	selected ILs	IL HET	IL HOM	T2E	E2T	E2TS1	E3T
2011-15	3551	836	137	109	28	8	28	101	-
2016	370	146	21	13	8	-	-	21	-
2017	1511	320	26	19	7	-	-	26	-
2018	782	324	25	14	11	-	-	25	-
2019	1853	418	62	42	20	-	-	53	9
2020	400	68	12	9	3	-	-	12	-
Total	8467	2112	283	206	77	8	28	238	9

Twenty-eight additional lines with introgressed fragments in homozygosis were also selected, covering 83% of the genome with a large gap on G4 (Figure 4.3, Table 4.2 and Table S4.3) and smaller ones on all groups except G5. While most of these plants had a single major fragment in homozygosis, two had two fragments, one in homozygosis and the other in heterozygosis. Nine of the selected lines had part of their introgression in homozygosis and part in heterozygosis, indicating their origin from gametes with recombinations in the almond fragment. These lines would require an additional round of selection in their selfed offspring to obtain completely single-fragment homozygous ILs. The remaining lines genotyped with the SNP chip (61) were discarded as they contained additional fragments longer than 2.5 Mb.

The SNP analysis provided a much more detailed picture of the genome and allowed identification of eight small fragments, <1% of the genome (<2.5 Mb), that were not detected by the SSRs: one in G1 (0.97 Mb), G2 (0.82 Mb), G3 (0.52 Mb), G5 (1.40 Mb), and G7 (2.17 Mb), and three in G8 (0.91, 1.29 and 1.60 Mb). Their positions are indicated in Figures 4.2 and 4.3. They were not considered when selecting or discarding ILs, although one of the homozygous ILs contains only the largest fragment, the one on G7, in homozygosis. Introgressed fragments were generally large (see Table 4.2), as expected considering the low number of generations used to obtain them, ranging from 4.0-36.9 Mb (average 17.1 Mb) in the heterozygous ILs and from 2.2-36.3 Mb

Table 4.2: Heterozygous and homozygous IL collections selected with the 18k SNP chip. Numbers of lines, fragments identified and genome coverage.

	ILs HET								ILs HOM						
Linkage group	**Mb total**	**ILs**	**Fragments ***	**Smallest IL fragment (Mb)**	**Largest IL fragment (Mb)**	**Average IL fragment (Mb)**	**Mb covered**	**Coverage (%)**	**ILs**	**Fragments***	**Smallest IL fragment (Mb)**	**Largest IL fragment (Mb)**	**Average IL fragment (Mb)**	**Mb covered**	**Genome coverage (%)**
G1	47.8	4	5	14.0	36.9	24.4	47.8	100	3	3	17.0	36.3	24.0	36.3	80
G2	30.4	7	5	4.5	29.1	16.8	29.1	96	5	4	19.2	29.1	23.4	29.1	96
G3	27.4	5	5	6.0	26.4	16.1	27.4	100	2	2	6.0	26.4	16.2	26.4	96
G4	25.8	4	4	4.0	24.3	18.3	25.8	100	2	1	11.1	11.1	11.1	11.1	43
G5	18.5	3	3	6.0	14.0	8.7	18.5	100	3	4	6.0	14.5	11.5	18.5	100
G6	30.8	9	4	8.5	30.8	21.4	30.8	100	8	8	5.1	26.5	14.6	26.5	86
G7	22.4	4	4	11.0	22.4	18.7	22.4	100	2	2	2.2	19.6	10.9	21.8	97
G8	22.6	3	5	6.4	19.5	12.4	22.6	100	3	4	8.5	15.2	11.8	18.4	81
Total	225.7	39	35			17.1	224.4	99	28	28			15.4	188.1	83

*Number of different non-overlapping fragments in which the chromosome is divided by the ILs of the collection

(average 15.4 Mb) in the homozygous ILs. Several chromosomes were completely or almost completely (>95%) covered with a single introgression: these were G2, G3, G6, and G7, the two former in both homozygosis and heterozygosis.

Segregation in the 'Earlygold' background was also studied with the SNP chip. Results (Figures 4.4 and 4.5) show that there were large regions without segregating markers, covering approximately half of the genome (50.5%) as observed previously (Donoso et al. 2015; Kalluri et al. 2021). The rest of the genome of each IL segregated in large blocks for the two alleles of 'Earlygold'. Knowledge of the 'Earlygold' genotype for each specific IL is important to incorporate the possible effects of a segregating background in the genetic analysis of trait variability.

4.3.3 Major genes and QTLs analyzed in the population

The data were obtained from single trees of different ages, some too young to produce fruit, some grown on their own roots and others grafted. For these reasons, and as a preliminary attempt to understand the potential of ILs for genetic analysis, here we studied traits that were determined by major genes and could be analyzed as qualitative, as well as certain quantitative characters that produced consistent QTLs in the T×E and T1E populations (Donoso et al. 2016). Three of the major genes expected to segregate in the ILs, juiciness (*Jui*), blood flesh (*DBF2*) and resistance to powdery mildew (*Vr3*), had the expected phenotypes in all ILs that could be studied, 28, 28 and 63, respectively (Table S4.4). For the two linked genes *Jui* and *DBF2* on chromosome 1, all ILs had the juicy, yellow flesh phenotype except for PAILE1-2348 and PAILE1-3448 that were both non-juicy and red-fleshed. For *Vr3*, only ILs with almond introgressions in the region of chromosome 2 that contains the gene (Marimon et al. 2020) were resistant (PAILE2-0119, PAILE2-0125, PAILE2-0130a, PAILE2-0130b, PAILE2-1625, PAILO2-0123, PAILO2-0125a, PAILO2-0125b, PAILO2-0130*, and PAILE2-0625).

Maturity date (*MD*), a gene located in the central region of chromosome 4 (Pirona et al. 2013), was scored as qualitative in the spring/summer of 2021. In this case, the ILs without the fragment from almond that included *MD* and 'Earlygold' (28) had an average number of 167.7 (± 12.4) Julian days to maturity, whilst two ILs (PAILE4-0226b and PAILE4-0526) that contained the *MD* fragment matured more than one month later (the 214th day), as expected considering the delaying

Figure 4.2: Graphical genotype of the 39 heterozygous ILs of almond ('Texas') in the peach ('Earlygold') background based on the 18k SNP chip. G1 to G8 are the eight linkage groups of *Prunus*.

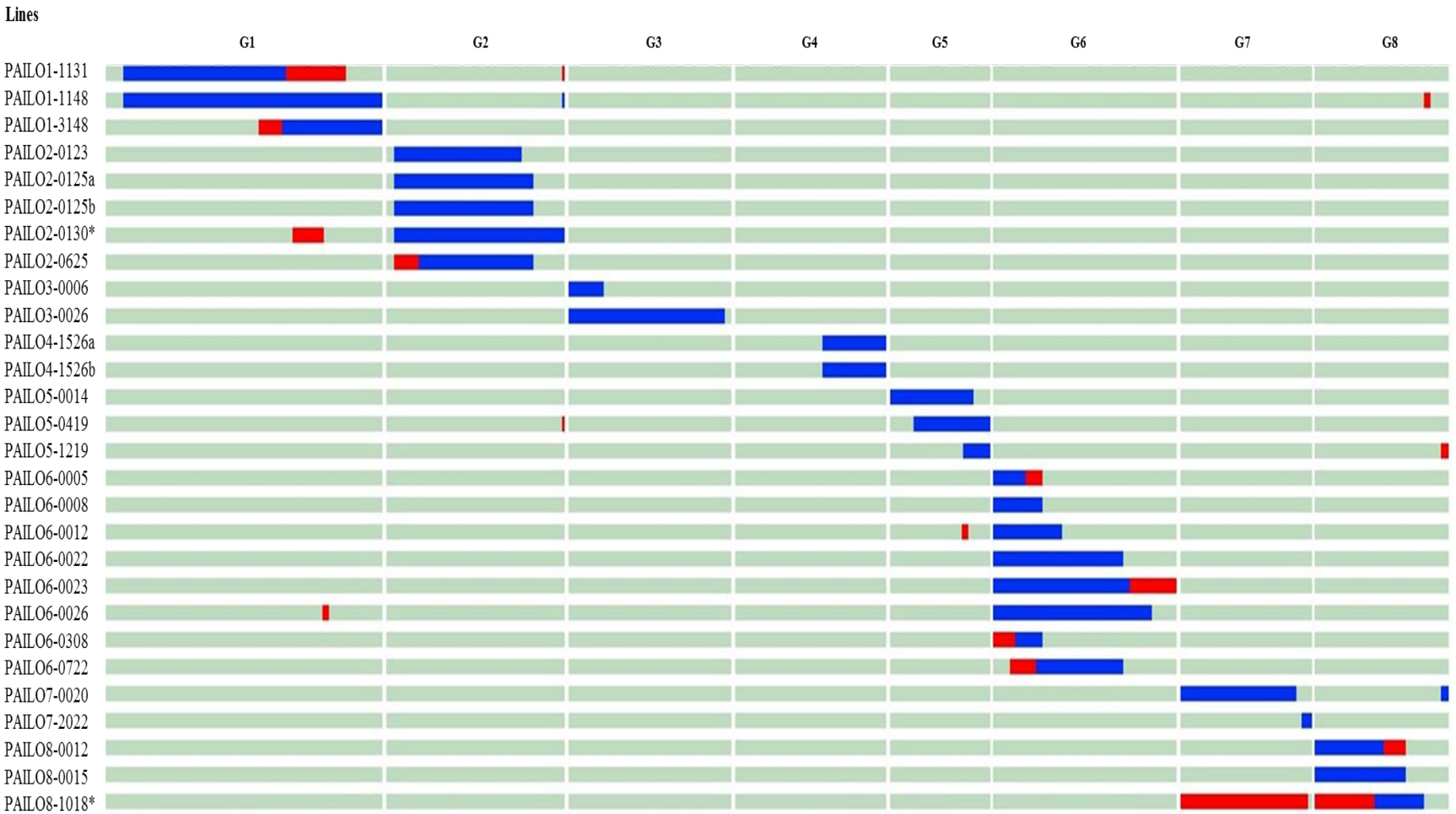


Figure 4.3: Graphical genotype of the 28 homozygous ILs of almond ('Texas') in the peach ('Earlygold') background based on the 18k SNP chip. Red fragments are heterozygous almond introgressions. G1 to G8 are the eight linkage groups of *Prunus*.

effect of the almond allele. However, PAILE4-0226a which contained the same almond fragment as PAILE4-0226b, matured on the 179th day, i.e., within the range of 'Earlygold' and the early maturing ILs. This can be explained by the variability of the 'Earlygold' background, where PAILE4-0226b and PAILE4-0526 share a late maturity allele (*A*) of 'Earlygold' (Kalluri et al. 2021; see also Figure 4.4), whereas PAILE4-0226a had the alternative (*B*) allele at this locus, that confers earliness, and in heterozygosis with the almond allele produces a phenotype of intermediate maturity date.

Quantitative measurements were taken for three fruit traits, weight (FW), soluble solid contents (SSC) and titratable acidity (TA) (Table S4.4). A highly significant increase in fruit weight compared to 'Earlygold' was observed in the homozygous ILs that covered that region (PAILO6-0005 and PAILO6-0008) with average values of 144 and 148 g/fruit compared to 83g of 'Earlygold' (73-78% weight increase). Another IL in this region, PAILO6-0308, had FW values similar to 'Earlygold' indicating that the QTL was located at the extreme of the chromosome (0-4 Mb). When looking at heterozygous ILs in this region (PAILE6-0008b, PAILE6-0009, PAILE6-0831, PAILE6-0031b and PAILE6-0031c), we observed an inconsistent pattern with some lines having values similar to 'Earlygold'(PAILE6-0031c), and others significantly higher (PAILE6-0008b and PAILE6-0831) or lower (PAILE6-0009 and PAILE60031b) values. Additionally, several lines had a highly significant ($P<0.001$) weight increase in ILs with introgressions at the central region of G2 and G8, and the proximal end of G3. For SSC, only one highly significant QTL ($P<0.001$) was observed in the proximal end of G6 (PAILO6-0308) with the almond allele producing an increase of the value of this trait, and for TA, a QTL ($P<0.05$) occurred in the same region (PAILE6-0008b) resulting in a decrease of acidity in the heterozygous individual, although it was not confirmed in the homozygous line containing the same fragment (Table S4.4).

Petiole length was a variable trait in the crosses between almond and peach studied by Donoso et al. (2016), where typically almond had a long petiole, peach a short one and the progeny segregated between these two extremes. In the IL collection all plants had a short petiole like in peach except for PAILE8-0516 and PAILO8-1018 that had a highly significantly higher length of intermediate size, similar to that of the interspecific hybrid. Four more lines, corresponding to three different chromosomal positions, had a significantly shorter petiole than 'Earlygold' (PAILE2-2630, PAILO1-1148, PAILO1-3148 and PAILO6-0722).

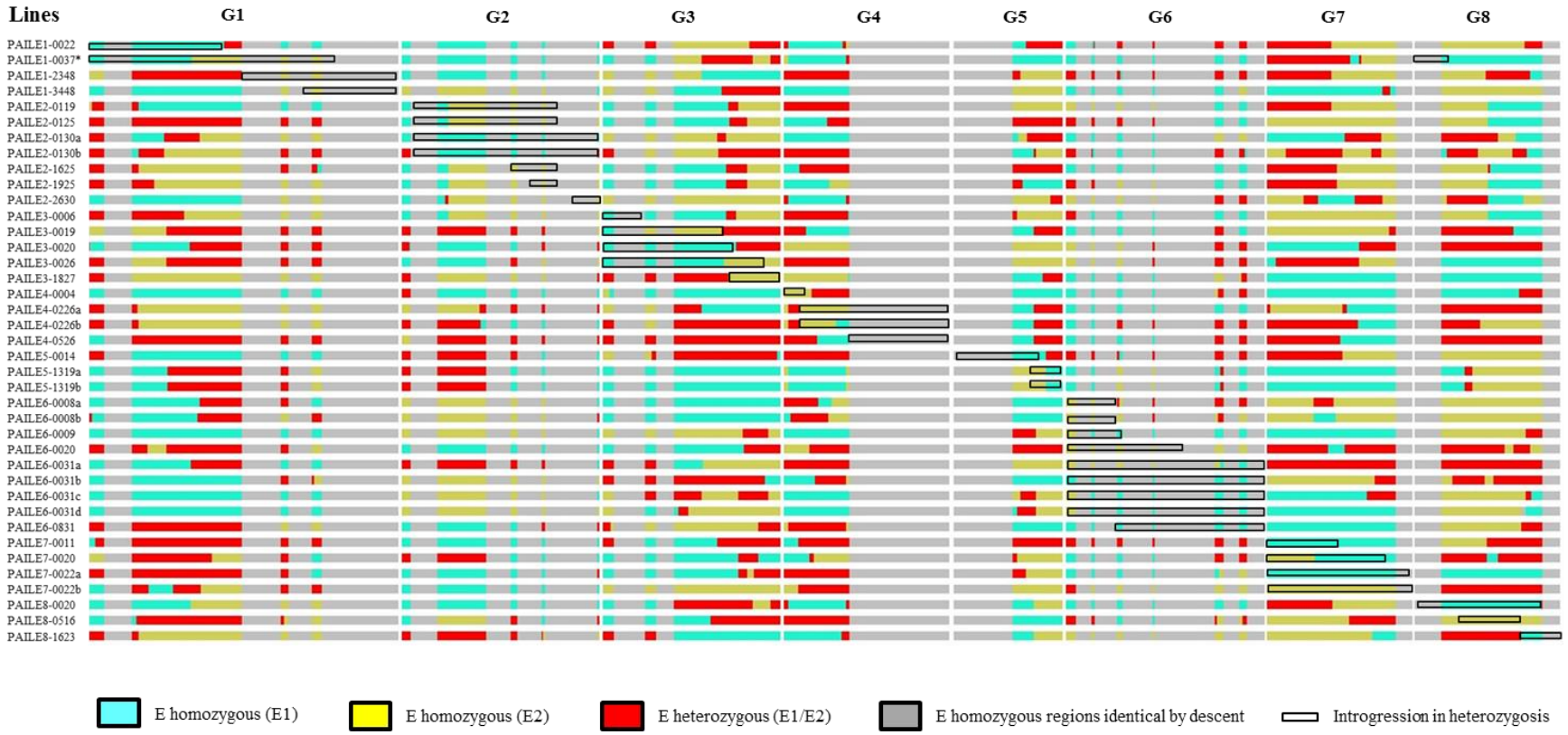


Figure 4.4: Graphical genotype of the 'Earlygold' background in the selected heterozygous ILs.

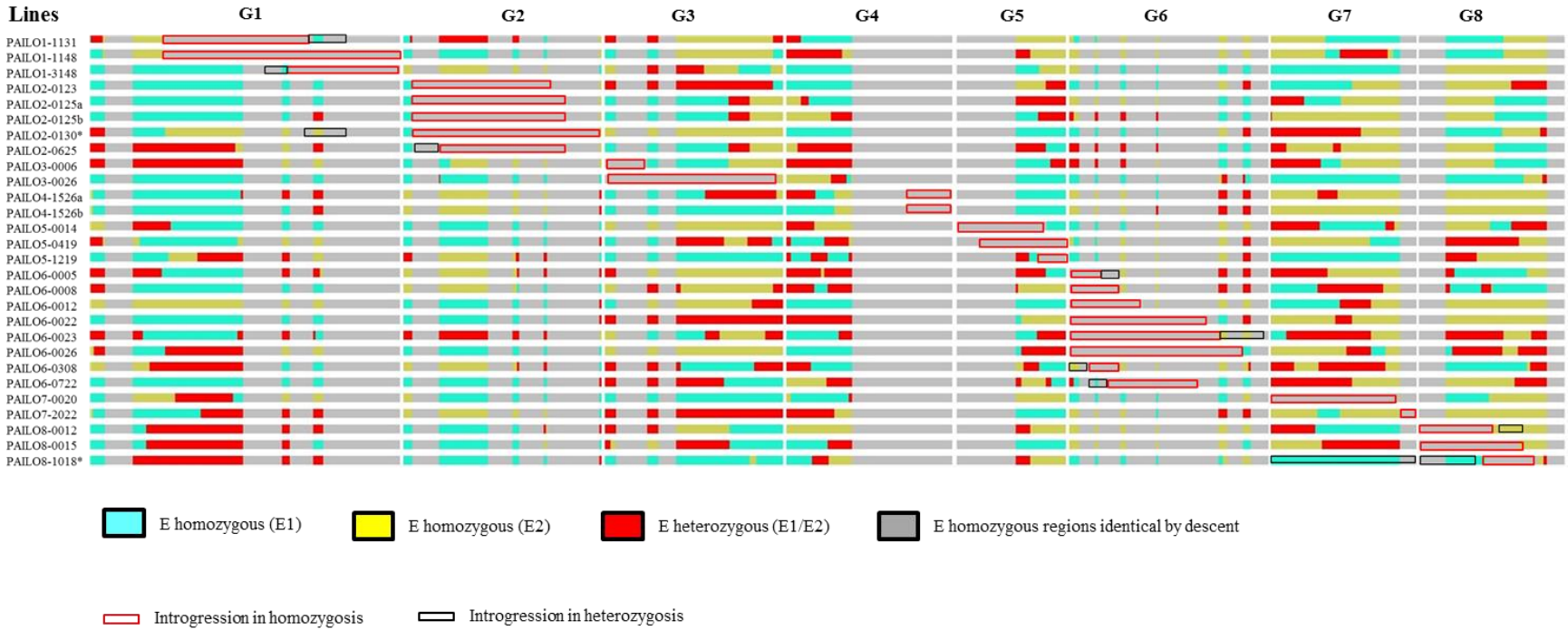


Figure 4.5: Graphical genotype of the 'Earlygold' background in the selected homozygous ILs.

4.4 Discussion

Plant collections adequate for efficient genetic analysis are a key resource for the progress in the genetics and genomics of crop species. Introgression line sets are optimal for the dissection of quantitative traits and to facilitate genetic analysis and gene cloning, particularly in the background of lines of agricultural interest, so avoiding problems related with donor-donor epistatic effects and making possible the detection of subtle pleiotropic effects, usually difficult to identify in other population types such as F2, BC1 or RILs (Tanksley and Nelson 1996). However, the construction of such populations is especially difficult in woody perennials, as it represents an enormous investment in time and resources, a consequence of the large size of the individuals and the long intergeneration time of these species. In this paper we present a peach-almond introgression collection that we believe is the first to have been produced in a woody perennial species. One advantage of certain tree crops is that they can be clonally reproduced, which has given us the possibility of constructing two collections, one with 39 ILs that contain introgressed almond fragments in heterozygosity and covers 99% of the almond genome, and another with 28 homozygous ILs with 83% of genome coverage. This continues research begun in 2006 in our lab with the initial objective of finding fast approaches to introgress useful genes from the highly variable *Prunus* wild or cultivated species, almond in this case, to enrich the much narrower variability of cultivated peach. This initial research led to an incomplete first set of introgression lines (Serra et al. 2016) with partial genome coverage (64% in heterozygosis and 14% in homozygosis). Here, we selected the ILs from the selfed progeny of a set of individuals with a low number of introgressions (2-3). This has considerably simplified our work, as peach has low seed production compared to other tree species, particularly when artificial crossing is required and because, given the early maturity time of most of the materials we used, in vitro embryo rescue was necessary, making the process more time-consuming, unpredictable and expensive.

Using a set of SSRs we selected 283 ILs, 135 of which were genotyped in depth with the 18k SNP chip. SNP results were particularly informative as they increased the marker density more than 50-fold (from an average of one SSR every 2,015 kb to one SNP every 37 kb) allowing us to identify and discard plants having more than one introgression. The final collection of 67 individuals with a single introgression (with a few exceptions with two introgressions) could therefore be selected

with great certainty, and with a precise position of the boundaries of each almond fragment. The accuracy of this analysis also made possible the identification of several smaller introgressions (<1% of the genome) that were undetected by the SSR markers, and of some missing fragments at the end of certain linkage groups. Our initial expectations were that these fragments would occur only very sporadically because of the relatively low level of recombination of peach chromosomes and specifically of this interspecific cross (1.2-1.4 crossovers per chromosome on average; Donoso et al. 2015), but the fact that we recovered them suggests that a high-density genotyping step before elaborating a final set of ILs is an advisable option.

One aspect to consider in the analysis of this IL collection is that a commercial cultivar ('Earlygold') was used as a recurrent parent. 'Earlygold' is not an inbred line, which has the drawback that segregation for chromosomal fragments heterozygous in this variety may interfere with the interpretation of genes/QTLs from the introgressed almond fragments in the ILs. At the beginning of this project we considered, but discarded, the option of switching to a more homozygous line. The first main reason was that the availability of homozygous lines is scarce in peach, and most of them are either weak individuals, probably as a consequence of inbreeding, or genotypes, such as the Spanish non-melting flesh varieties, that are genetically distant from the major commercial peach gene pool (Aranzana et al. 2003). Secondly, including a new recurrent parent would have delayed the construction of this IL collection by two generations, equivalent to 7-9 years. A third reason is that we expected that the variability detected by the almond genome would often be of a sufficiently different nature and genome position compared to that of peach, as shown in previous results based on the analysis of 'Texas' × 'Earlygold' progenies (Donoso et al. 2015, 2016), making interferences between almond/peach vs. peach/peach easily detectable and interpretable. Additionally, 'Earlygold' is heterozygous in less than half of its genome (Donoso et al. 2016; Kalluri et al. 2021, this work), because of the existence of large DNA fragments identical by descent, as in many other modern peach varieties (Micheletti et al. 2015), due to the recent history of co-ancestry of modern peach breeding programs (Scorza et al. 1985), so we were only expecting a relatively narrow window of segregation to occur in its offspring. The inheritance analysis of a large set of characters of the F2 of 'Earlygold' is available (Kalluri et al. 2021), and relevant for this work as it identified a limited number of QTLs. Only one of these, a major QTL located on G4 in the region of *MD* that, in addition to maturity date, partially affects other fruit characters such as fruit weight, soluble solid contents and leaf color at senescence. The genotypes

for the ‘Earlygold’ background in the IL collection are given in Figures 4.4 and 4.5 and may be useful to other users to identify the genotypes of the recurrent parent and consider or discard possible effects of the background on the expression of their character of interest.

Using the plants of the IL collections, we examined the segregation of certain major genes previously described in ‘Texas’ × ‘Earlygold’ progenies, and for four of them (*Jui*, *DBF2*, *Vr3* and *MD*) we observed the expected phenotypes. Other genes that segregate in the IL collection, such as two independent male fertility restorer genes (*Rf1*/*rf1* and *Rf2*/*rf2*) (Donoso et al. 2015) were not expected to produce any phenotypes as all E2T individuals had the peach cytoplasm. The almond fruit (*Alf*) gene that determines the formation of the thick mesocarp (Donoso et al. 2016) typical of peach, would produce only almond fruit types in the individuals homozygous for the recessive almond (*alf*) allele, and the ILs homozygous at the corresponding region of G4 were not available in homozygosity. Other lines having this fragment in homozygosity, even with additional introgressed fragments, were too young to produce fruit. The case of one IL (PAILE4-0226a) with an early maturity date phenotype while it was expected to mature late as it had the almond introgression, can be explained by the interaction of the almond allele with an allele of ‘Earlygold’ that confers early maturity (Donoso et al. 2016; Kalluri et al. 2021), as opposed to other ILs that had the late allele of ‘Earlygold’. This would be a clear case of segregation in the ‘Earlygold’ background that would result in an unexpected phenotype in an IL. Considering the importance of this region for phenological and fruit related characters, it would be interesting to fix one of the alleles of ‘Earlygold’ at this locus in all lines to have a uniform phenotype for maturity date which may avoid interference with other characters, particularly fruit characters, that may be associated with this region.

Three quantitative characters were studied in fruits of some ILs that were already fruiting. For one character, fruit weight, a major QTL was found by Donoso et al. (2016) at the proximal end of G6, where the allele of almond produced an increase in fruit weight with a high percentage of explained phenotypic variance (R^2=19-21%). A major QTL for fruit weight was detected at the same chromosomal region by combining linkage mapping and association studies in a collection of peach genotypes (Cirilli et al. 2021). We also detected a major increase in fruit weight in this region that was very strong in the homozygous ILs but gave inconsistent results in heterozygous ILs, suggesting a major dominance component of the QTL and possibly that QTLs in neighboring

regions of the same chromosome may be involved in the expression of this trait. This is an especially interesting QTL, as fruit size is a character of great commercial value in peach and the use of the almond allele, alone or in combination with other peach alleles may have an immediate application in breeding. Results for SSC and TA were less conclusive, with the identification of a QTL for both traits also in the proximal end of G6. QTLs for SSC and TA at the beginning and end of chromosome 6 have been identified by Hernández Mora et al. (2017) using a multi-progeny approach that included peach × peach and peach × other species (almond one of them). Finally, petiole length was significantly longer in the lines PAILE8-0516 and PAILO8-1018 than in 'Earlygold' (43-70% longer) and the rest of ILs, coinciding with the position of a QTL with large effects (R^2=15-39%) previously identified (Donoso et al. 2016) for this character. Other ILs with petiole lengths shorter than 'Earlygold' were also found, suggesting the presence of additional QTLs not identified in biparental crosses, as well as the existence alleles that may determine shorter petioles in the species (almond) that has the long petiole phenotype.

The results on the trait analysis reported here were obtained on single plants grown on their own roots and from seedlings obtained in various years (see Tables S4.2 and S4.3), some of which were still in their juvenile period. While they can be accurate for traits with clear alternative phenotypes that can be scored qualitatively, adequate conditions for the analysis of metric traits are required. The next step in this research is the production of grafted replicates from the complete IL collection, to be planted in two sites. Once these plants reach the production stage, we will start phenotyping them for the traits that we consider of interest. We also plan to continue the production of new ILs to complete the collection, especially that of homozygous ILs where there are still some gaps. We are also expanding this collection to include ILs with wild crop relatives of peach, particularly *P. davidiana* and *P. mira*. One direction of our current research is to fine map, and eventually clone, the main major genes detected so far in peach × almond crosses. We have already started with the *Vr3* gene that confers resistance to powdery mildew (Marimon et al. 2020), and are advancing with others such as (*DBF2, Jui* and *Alf*). We believe that this is a valuable resource for the understanding of the genetics of fruit trees and offer any plants in this collection, or the full collection, to members of the scientific community that would like to use them for research purposes.

4.5 Supplementary Material

Table S4.1: Genotypes of the E2T set of lines with two and three almond ('Texas') introgressions in heterozygosity on the peach ('Earlygold') background for 113 SSRs. The lines with three introgressions are marked with asterisks (*), unknown data (-). Linkage groups are from G1 to G8. Genotypes are H= heterozygous peach-almond; A=both alleles from peach.

SSR	cM	E2T-022-11	E2T-022-06*	E2T-031-23	E2T-092-28	E2T-022-23*	E2T-031-06*	E2T-031-25	E2T-031-15*	E2T-092-02	E2T-092-25	E2T-092-20*	E2T-092-23	E2T-022-08	E2T-031-14*	E2T-092-13	E2T-079-19	E2T-079-17	E2T-051-20*	E2T-079-07*	E2T-051-12*	E2T-079-03	E2T-079-10	E2T-031-26	E2T-051-07	E2T-502-10	E2T-031-19*	E2T-031-13*	E2T-502-17	E2T-051-11	E2T-962-05	E2T-051-04	E2T-502-18	E2T-079-15	E2T-241-11	E2T-724-03*	E2T-502-27*	E2T-092-11*
G1																																					
EPDCU3122	0.8	H	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP96-018	1.4	H	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EPPCU5331	12.2	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
PACITA005	12.2	H	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
PceGA59	17.0	H	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP96-005	19.1	H	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EPPCU1090	19.1	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPPCT003	21.2	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPPCT026	32.1	H	H	H	H	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H
EPDCU3489	35.3	H	H	H	H	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	H
BPPCT016	43.6	A	H	H	H	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
CPPCT019	46.7	A	H	H	H	H	H	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
CPPCT042	47.7	A	H	H	H	H	H	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
EPDCU2862	47.7	A	H	H	H	H	H	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
CPPCT053	48.7	A	H	H	H	H	H	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
CPPCT029	49.2	A	H	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
BPPCT028	51.2	A	H	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	H
G2																																					
calSG2	0.0	A	A	A	A	A	H	H	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPPCT044	0.0	A	A	A	A	A	H	H	H	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
AMPA93	1.4	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-
MA024a	5.4	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP98-025	6.5	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPDCT044	7.5	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BPPCT004	10.1	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EPDCU4017	12.2	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BPPCT001	12.7	A	A	A	A	A	H	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPSCT044	17.5	A	A	A	A	A	A	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M1a	19.6	A	A	A	A	A	A	A	H	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP96-013	21.6	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPDCT004	25.2	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP98-411	26.3	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
pchgms1	28.6	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BPPCT030	32.2	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPPCT043	32.2	A	A	A	A	A	A	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPSCT021	38.1	A	A	A	A	A	A	A	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
PceGA34	45.1	A	A	A	A	A	A	A	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPSCT034	48.8	A	A	A	A	A	A	A	H	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDA-023	49.9	A	H	A	A	A	A	A	H	A	A	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
G3																																					
EPPCU5990	0.0	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP97-403	6.5	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BPPCT007	6.5	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BPPCT039	15.5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	-	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EPDCU3083	16.6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CPPCT002	28.2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
UDP96-008	30.6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
EPPCU0532	42.6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
G4																																					
BPPCT010	0.0	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
EPDCU5060	1.1	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
pchgms2	7.9	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
CPPCT011	14.6	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
CPPCT005	14.6	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
CPDCT045	24.7	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
UDP96-003	29.5	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	H	A	A	H	H	A	A	A	A	A	A	A	A	A	A
M12a	31.6	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A
EPPCU2000	37.9	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A
BPPCT015	37.9	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A
UDA-021	41.5	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A
CPPCT046	41.5	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A
UDA-027	45.2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	H	A	A	A	A	A	A	A	H	A	A
Ps12	45.8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	H	H	A	A	A	A	A	A	A	H	A	A

Table S4.1 (Continued)

G5																																	
CPPCT040	0.0	A	A	H	A	A	H	H	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A	A
BPPCT026	4.0	A	A	H	A	A	H	H	A	A	A	A	A	-	H	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A	A
UDP97-401	5.4	A	A	H	A	A	H	H	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A	A
BPPCT017	15.3	A	A	H	A	A	H	H	-	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A	A
CPSCT006	15.3	A	A	H	A	A	H	H	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A	A
BPPCT037	17.9	A	A	H	A	A	H	H	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	A	A	H	H	H	A	A	A	A	A	A	A
pchgms4	19.5	A	A	A	A	A	H	H	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	H	A	H	H	H	A	A	A	A	A	A	A
CPPCT013	21.2	A	A	A	A	A	H	H	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	H	A	H	H	H	A	A	A	A	A	A	A
EPDCU5183	26.9	A	A	A	A	A	H	H	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	H	H	A	H	H	H	A	A	A	A	A	A	A
EPDCU4658	28.4	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	A	A	H	H	A	A	A	A	A	A	A
BPPCT038	29.4	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	A	A	A	H	A	A	A	A	A	A	A
CPSCT022	36.9	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	A	A	A	-	A	A	A	A	A	A	A
BPPCT014	38.1	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A
G6																																	
CPP21413	0.0	H	A	A	H	A	A	A	A	A	H	H	H	A	A	H	H	A	H	H	H	H	H	A	A	H	A	A	A	H	H	H	A	H	A	A	A
CPP21245	1.1	H	A	A	H	H	A	A	A	A	H	H	H	A	A	H	H	A	H	H	H	H	H	A	A	H	A	A	A	H	H	H	H	H	A	A	A
Ps7a2	1.1	H	A	A	H	H	A	A	A	A	H	H	H	A	A	H	H	A	H	H	H	H	H	A	A	H	A	A	A	H	H	H	H	H	H	A	A
CPP20836	1.1	H	A	A	H	H	A	A	A	A	H	H	H	A	A	H	H	A	H	H	H	H	H	A	A	H	A	A	A	H	H	H	H	H	H	A	A
UDP96-001	10.6	H	A	A	H	H	A	A	A	A	H	H	A	A	A	H	A	H	H	H	H	H	A	A	H	A	H	A	H	H	H	H	H	H	H	H	H
BPPCT008	16.0	H	A	A	H	H	A	A	A	A	H	H	A	A	A	A	H	A	H	H	H	H	H	A	A	H	A	H	A	A	H	H	H	H	H	A	H
CPSCT012	17.0	A	A	A	H	H	A	A	A	A	H	H	A	A	A	A	H	A	H	H	H	H	H	A	A	H	A	H	A	A	H	H	H	H	H	A	A
pchcms5	20.7	A	A	A	H	H	A	A	A	A	H	H	A	A	A	A	A	A	-	A	-	A	A	A	A	H	A	H	A	A	H	H	H	A	A	A	A
BPPCT025	24.1	A	A	A	H	H	A	A	A	A	H	-	A	A	A	A	A	A	-	A	-	A	A	A	A	H	A	H	A	A	A	H	H	A	A	A	A
CPPCT047	27.6	A	A	A	H	H	A	A	A	A	H	-	A	A	A	A	A	A	-	A	-	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
UDP98-412	34.4	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
MA040a	34.4	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
AMPA130	34.9	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
MA14a	34.9	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
EPPCU4092	37.0	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	H	A	A	A	A
CPPCT030	43.0	A	A	A	H	H	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A
CPPCT021	43.0	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A
G7																																	
CPSCT004	5.0	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H
CPPCT039	6.7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H
pchgms6	11.4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H
UDP98-408	12.9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
CPPCT057	18.0	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
CPPCT033	24.1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
MA20a	25.8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
PMS02	27.9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
EPPCU5176	30.6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
pchcms2	31.1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	H	A	A	A	H	A	A	A	A	H	H	H	A	A	A	H
CPPCT017	39.1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	H	A	A	A	A	A	H	H	A	A	A	H
EPDCU3392	41.9	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	H	A	A	A	A	A	H	H	H	A	A	H
Ps5c3	45.3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	H	A	A	A	A	A	H	H	A	A	A	H
G8																																	
CPSCT018	0.0	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	H	H	A	H
BPPCT006	7.8	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H
CPDCT034	7.8	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H
CPPCT058	7.8	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H
CPPCT006	13.7	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H
M6a	21.9	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	H
UDP98-409	34.8	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A
EPDCU3454	38.5	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A
EPDCU3117	43.4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A

Table S4.2: Characteristics of the selected set of introgression lines (ILs) of almond fragments in the peach background in heterozygosis using the 18k SNP chip.

Line name	Initial code	Origin [1]	cs	Beginning of introgression (bp)	Marker in the proximal extreme	End of introgression (bp)	Marker in the distal extreme	Size (bp) of introgression	Age [2]	Fruit production [3]	Comments
PAILE1-0022	E2T-022-13	E2T‡	1	367,055	Peach_AO_0000794	22,241,226	Peach_AO_0070852	21,874,171	6	Yes	-
PAILE1-0037*	51P19-05	E2TS1	1	367,055	Peach_AO_0000794	36,948,923	Peach_AO_0109269	36,581,868	1	No	More than one fragment
PAILE1-2348	11P15-04	E2T‡	1	23,494,321	Peach_AO_0074055	47,719,653	Peach_AO_0138791	24,225,332	5	Yes	-
PAILE1-3448	21P15-39	E2TS1	1	33,759,798	Peach_AO_0101543	47,719,653	Peach_AO_0138791	13,959,855	5	Yes	-
PAILE2-0119	19P15-95	E2TS1	2	1,349,947	SNP_IGA_230270	19,091,424	Peach_AO_0277409	17,741,477	5	Yes	-
PAILE2-0125	19P15-58	E2TS1	2	1,349,947	SNP_IGA_230270	24,822,389	Peach_AO_0295724	23,472,442	5	Yes	-
PAILE2-0130a	56P19-18	E2TS1	2	1,349,947	SNP_IGA_230270	30,092,550	Peach_AO_0308231	28,742,603	1	No	-
PAILE2-0130b	56P19-23	E2TS1	2	1,349,947	SNP_IGA_230270	30,092,550	Peach_AO_0308231	28,742,603	1	No	-
PAILE2-1625	19P15-65	E2TS1	2	16,477,856	Peach_AO_0266461	24,822,389	Peach_AO_0295724	8,344,533	5	Yes	-
PAILE2-1925	19P15-04	E2TS1	2	19,128,706	Peach_AO_0277526	24,822,389	Peach_AO_0295724	5,693,683	5	Yes	-
PAILE2-2630	55P19-25	E2TS1	2	25,893,886	Peach_AO_0299018	30,092,550	Peach_AO_0308231	4,198,664	1	No	Small fragment
PAILE3-0006	19P15-77	E2TS1	3	199,920	Peach_AO_0309082	6,088,325	Peach_AO_0329942	5,888,405	5	Yes	-
PAILE3-0019	13P15-06	E2T‡	3	199,920	Peach_AO_0309082	19,025,208	Peach_AO_0380211	18,825,288	5	Yes	-
PAILE3-0020	01P19-10	E3T	3	199,920	Peach_AO_0309082	20,039,043	Peach_AO_0383420	19,839,123	1	No	-
PAILE3-0026	01P19-02	E3T	3	199,920	Peach_AO_0309082	26,417,128	Peach_AO_0403611	26,217,208	1	No	-
PAILE3-1827	04P20-01	E2TS1	3	18,221,847	SNP_3_12878608	27,311,852	Peach_AO_0405265	9,090,005	< 1	No	-
PAILE4-0004	02P20-01	E2TS1	4	109,846	Peach_AO_0405407	3,954,996	Peach_AO_0422387	3,845,150	< 1	No	Small fragment
PAILE4-0226a	T2E-193-06	T2E‡	4	1,559,815	Peach_AO_0415788	25,439,465	Peach_AO_0525706	23,879,650	6	Yes	-
PAILE4-0226b	T2E-193-29	T2E‡	4	1,559,815	Peach_AO_0415788	25,439,465	Peach_AO_0525706	23,879,650	6	Yes	-

[1] Origin: generation where the line has been selected; ‡plants that were grafted on 'Garnem' rootstock [2] Age: years from transplanting till 2021; [3] Trees producing fruit in 2021.

Table S4.2 (Continued)

Line name	Initial code	Origin [1]	cs	Beginning of introgression (bp)	Marker in the proximal extreme	End of introgression (bp)	Marker in the distal extreme	Size (bp) of introgression	Age [2]	Fruit production [3]	Comments
PAILE4-0526	T2E-193-19	T2E‡	4	4,996,274	Peach_AO_0425793	25,439,465	Peach_AO_0525706	20,443,191	6	Yes	-
PAILE5-0014	11P15-13	E2T‡	5	277,967	Peach_AO_0526579	13,961,772	SNP_IGA_598118	13,683,805	5	Yes	-
PAILE5-1319a	61P19-09	E2TS1	5	12,552,668	Pp05_12552668	17,034,882	Peach_AO_0590939	4,482,214	1	No	-
PAILE5-1319b	67P19-09	E2TS1	5	12,552,668	Pp05_12552668	17,034,882	Peach_AO_0590939	4,482,214	1	No	-
PAILE6-0008a	30P18-31	E2TS1	6	90,829	Peach_AO_0592816	8,478,666	SNP_IGA_631014	8,387,837	2	No	-
PAILE6-0008b	28P15-21	E2TS1	6	90,829	Peach_AO_0592816	8,478,666	SNP_IGA_631014	8,387,837	2	No	-
PAILE6-0009	E2T-003-04	E2T‡	6	90,829	Peach_AO_0592816	9,281,763	Peach_AO_0617940	9,190,934	6	Yes	-
PAILE6-0020	30P18-18	E2TS1	6	90,829	Peach_AO_0592816	19,790,349	Peach_AO_0659816	19,699,520	2	No	-
PAILE6-0031a	28P15-30	E2TS1	6	90,829	Peach_AO_0592816	30,560,829	Peach_AO_0692360	30,470,000	5	Yes	-
PAILE6-0031b	T2E-304-06	T2E‡	6	90,829	Peach_AO_0592816	30,560,829	Peach_AO_0692360	30,470,000	6	Yes	-
PAILE6-0031c	04P17-01	E2TS1	6	90,829	Peach_AO_0592816	30,560,829	Peach_AO_0692360	30,470,000	3	No	-
PAILE6-0031d	30P18-10	E2TS1	6	90,829	Peach_AO_0592816	30,560,829	Peach_AO_0692360	30,470,000	2	No	-
PAILE6-0831	09P15-04	E2T‡	6	7,932,882	SNP_IGA_629062	30,560,829	Peach_AO_0692360	22,627,947	5	Yes	-
PAILE7-0011	03P19-03	E3T	7	209,533	Peach_AO_0693189	11,012,670	SNP_IGA_762094	10,803,137	1	No	-
PAILE7-0020	03P19-06	E3T	7	209,533	Peach_AO_0693189	19,626,488	Peach_AO_0772754	19,416,955	1	No	-
PAILE7-0022a	72P19-04	E2TS1	7	209,533	Peach_AO_0693189	21,814,037	Peach_AO_0777606	21,604,504	1	No	-
PAILE7-0022b	01P19-05	E3T	7	209,533	Peach_AO_0693189	22,231,288	SNP_IGA_792877	22,021,755	1	No	-
PAILE8-0020	31P18-01	E2TS1	8	47,906	Peach_AO_0778829	19,513,212	Peach_AO_0867586	19,465,306	2	No	-
PAILE8-0516	53P17-03	E2TS1	8	5,159,866	Peach_AO_0811535	16,340,573	SNP_IGA_870207	11,180,707	3	Yes	-
PAILE8-1623	24P15-27	E2TS1	8	16,188,664	Peach_AO_0855225	22,462,250	Peach_AO_0875402	6,273,586	5	Yes	-

Table S4.3: Characteristics of the selected set of introgression lines (ILs) of almond fragments in the peach background in homozygosis using the 18k SNP chip. All ILs have been selected in the E2TS1 generation.

Line name	Initial code	cs	Beginning of introgression (bp)	Marker in the proximal extreme	End of introgression (bp)	Marker in the distal extreme	Size (bp) of introgression	Age [1]	Fruit production [2]	Comments
PAILO1-1131	75P19-03	1	11,452,860	Peach_AO_0032086	31,554,767	Peach_AO_0096835	20,101,907	1	No	-
PAILO1-1148	57P18-44	1	11,452,860	Peach_AO_0032086	47,719,653	Peach_AO_0138791	36,266,793	2	No	-
PAILO1-3148	30P18-03	1	30,835,956	Peach_AO_0095680	47,719,653	Peach_AO_0138791	16,883,697	2	No	-
PAILO2-0123	06P17-15	2	1,349,947	SNP_IGA_230270	22,838,257	Peach_AO_0290341	21,488,310	3	No	-
PAILO2-0125a	19P15-50	2	1,349,947	SNP_IGA_230270	24,822,389	Peach_AO_0295724	23,472,442	5	No	-
PAILO2-0125b	19P15-120	2	1,349,947	SNP_IGA_230270	24,822,389	Peach_AO_0295724	23,472,442	5	No	-
PAILO2-0130*	37P18-44	2	1,349,947	SNP_IGA_230270	30,092,550	Peach_AO_0308231	28,742,603	2	No	More than one fragment
PAILO2-0625	19P15-116	2	5,591,372	SNP_IGA_182333	24,822,389	Peach_AO_0295724	19,231,017	5	No	-
PAILO3-0006	19P15-74	3	199,920	Peach_AO_0309082	6,088,325	Peach_AO_0329942	5,888,405	5	Yes	-
PAILO3-0026	02P20-14	3	199,920	Peach_AO_0309082	26,417,128	Peach_AO_0403611	26,217,208	< 1	No	-
PAILO4-1526a	28P15-34	4	14,698,356	Peach_AO_0459684	25,439,465	Peach_AO_0525706	10,741,109	5	Yes	-
PAILO4-1526b	28P15-46	4	14,698,356	Peach_AO_0459684	25,439,465	Peach_AO_0525706	10,741,109	5	Yes	-
PAILO5-0014	16P17-02	5	277,967	Peach_AO_0526579	14,038,734	Peach_AO_0584194	13,760,767	3	No	-
PAILO5-0419	09P17-01	5	3,993,107	Peach_AO_0548234	17,034,882	Peach_AO_0590939	13,041,775	3	No	-
PAILO5-1219	67P19-03	5	12,477,386	SNP_IGA_594745	17,034,882	Peach_AO_0590939	4,557,496	1	No	-
PAILO6-0005	20P15-66	6	90,829	Peach_AO_0592816	5,435,066	SNP_IGA_621556	5,344,237	5	Yes	Small fragment

[1] Age: years from transplanting till 2021; [2] Trees producing fruit in 2021

Table S4.3 (Continued)

Line name	Initial code	cs	Beginning of introgression (bp)	Marker in the proximal extreme	End of introgression (bp)	Marker in the distal extreme	Size (bp) of introgression	Age[1]	Fruit production[2]	Comments
PAILO6-0008	20P15-26	6	90,829	Peach_AO_0592816	8,478,666	SNP_IGA_631014	8,387,837	5	Yes	-
PAILO6-0012	28P18-13	6	90,829	Peach_AO_0592816	11,579,097	Peach_AO_0626234	11,488,268	2	No	-
PAILO6-0022	06P17-19	6	90,829	Peach_AO_0592816	21,789,663	Peach_AO_0665354	21,698,834	3	No	-
PAILO6-0023	50P19-24	6	90,829	Peach_AO_0592816	22,842,917	SNP_IGA_677625	22,752,088	1	No	-
PAILO6-0026	37P18-13	6	90,829	Peach_AO_0592816	26,493,790	SNP_IGA_688827	26,402,961	2	No	-
PAILO6-0308	20P15-45	6	3,445,617	Peach_AO_0601606	8,478,666	SNP_IGA_631014	5,033,049	5	Yes	Small fragment
PAILO6-0722	34P18-05	6	7,073,013	Peach_AO_0611109	21,789,663	Peach_AO_0665354	14,716,650	2	No	-
PAILO7-0020	20P17-06	7	209,533	Peach_AO_0693189	19,601,766	Peach_AO_0772632	19,392,233	3	No	-
PAILO7-2022	28P15-65	7	20,209,201	Peach_AO_0774300	22,231,288	SNP_IGA_792877	2,022,087	5	Yes	Small fragment
PAILO8-0012	73P19-03	8	47,906	Peach_AO_0778829	11,802,085	Peach_AO_0839772	11,754,179	1	No	-
PAILO8-0015	22P17-04	8	47,906	Peach_AO_0778829	15,286,511	Peach_AO_0852224	15,238,605	3	Yes	-
PAILO8-1018*	72P19-22	8	9,882,943	Peach_AO_0831487	18,415,523	Peach_AO_0863859	8,532,580	1	No	More than one fragment

Table S4.4: Phenotypic data measured in the IL collections in homozygosis and heterozygosis. For Juiciness (*Jui*): J = Juicy; NJ= non-juicy; for flesh color (*DBF2*): Y= yellow flesh; R=Red flesh; for resistance to powdery mildew (*Vr3*): S= susceptible; R= resistant; for maturity date (MD): Julian days until fruit maturity. Quantitative traits are the mean of three samples in all cases except for those with a [1] that are based on 6 samples. Dunett's test for the comparison with 'Earlygold': * P≤0.05; ** P≤0.01; *** P≤0.001.

	Qualitative				**Quantitative**			
	Jui	***DBF2***	***Vr3***	***MD***	**FW**	**SSC**	**TA**	**Petiole length**
Earlygold	J	Y	S	158	82.67	10.23	3.91	0.98[1]
PAILE1-0022	J	Y	S	155	-	8.43	2.89	1.13
PAILE1-0037*	-	-	S	-	-	-	-	0.57
PAILE1-2348	NJ	R	S	165	106.00**	10.10	4.15	1.10
PAILE1-3448	NJ	R	S	194	104.33**	10.70	4.50	1.03
PAILE2-0119	J	Y	R	179	114.00***	12.03	3.13	1.13
PAILE2-0125	J	Y	R	179	115.67***	12.40	4.80	1.03
PAILE2-0130a	-	-	R	-	-	-	-	0.70
PAILE2-0130b	-	-	R	-	-	-	-	0.63
PAILE2-1625	J	Y	R	179	119.33***	12.03	3.11	1.07
PAILE2-1925	J	Y	S	173	111.33***	11.93	3.92	0.97
PAILE2-2630	-	-	S	-	-	-	-	0.50**
PAILE3-0006	J	Y	S	179	111.67***	9.00	2.84	0.90
PAILE3-0019	J	Y	S	173	123.00***	9.13	5.19	0.83
PAILE3-0020	-	-	S	-	-	-	-	0.77
PAILE3-0026	-	-	S	-	-	-	-	0.67
PAILE3-1827	-	-	-	-	-	-	-	-
PAILE4-0004	-	-	-	-	-	-	-	-
PAILE4-0226a	J	Y	S	179	84.67	11.83	4.75	1.03
PAILE4-0226b	J	Y	S	214	71.00	-	-	0.87
PAILE4-0526	J	Y	S	214	102.67*	-	-	1.10
PAILE5-0014	J	Y	S	165	106.67**	9.90	3.16	1.07
PAILE5-1319a	-	-	S	-	-	-	-	0.83
PAILE5-1319b	-	-	S	-	-	-	-	0.63
PAILE6-0008a	-	-	S	-	-	-	-	0.73
PAILE6-0008b	J	Y	S	161	101.00*	10.23	2.54*	1.13
PAILE6-0009	J	Y	S	158	57.00***	6.77**	5.35	0.77
PAILE6-0020	-	-	S	-	-	-	-	1.33
PAILE6-0031a	-	-	S	-	-	-	-	1.00
PAILE6-0031b	J	Y	S	145	60.67**	7.10*	2.71	1.07
PAILE6-0031c	J	Y	S	161	93.67	9.97	3.43	-
PAILE6-0031d	-	-	S	-	-	-	-	0.97
PAILE6-0831	J	Y	S	165	106.00**	13.83**	4.33	1.00

Table S4.4 (Continued)

	Qualitative				Quantitative			
	Jui	***DBF2***	***Vr3***	***MD***	**FW**	**SSC**	**TA**	**Petiole length**
PAILE7-0011	-	-	S	-	-	-	-	0.83
PAILE7-0020	-	-	S	-	-	-	-	0.70
PAILE7-0022a	-	-	S	-	-	-	-	1.07
PAILE7-0022b	-	-	S	-	-	-	-	0.77
PAILE8-0020	-	-	-	-	-	-	-	-
PAILE8-0516	J	Y	S	161	97.33	13.30*	3.66	1.67[1]***
PAILE8-1623	J	Y	S	161	109.33***	11.43	3.38	1.20
PAILO1-1131	-	-	S	-	-	-	-	-
PAILO1-1148	-	-	S	-	-	-	-	0.53*
PAILO1-3148	-	-	S	-	-	-	-	0.60*
PAILO2-0123	-	-	R	-	-	-	-	0.80
PAILO2-0125a	-	-	R	-	-	-	-	1.13
PAILO2-0125b	-	-	R	-	-	-	-	0.90
PAILO2-0130*	-	-	R	-	-	-	-	-
PAILO2-0625	-	-	R	-	-	-	-	0.97
PAILO3-0006	J	Y	S	168	102.33*	11.23	3.78	0.97
PAILO3-0026	-	-	-	-	-	-	-	-
PAILO4-1526a	J	Y	S	161	80.00	9.60	2.84	0.70
PAILO4-1526b	J	Y	S	165	83.33	10.50	2.86	1.13
PAILO5-0014	-	-	S	-	-	-	-	0.87
PAILO5-0419	-	-	S	-	-	-	-	0.97
PAILO5-1219	-	-	S	-	-	-	-	-
PAILO6-0005	J	Y	S	187	144.33***	11.63	3.55	0.93
PAILO6-0008	J	Y	S	168	148.33***	13.20*	4.63	1.07
PAILO6-0012	-	-	S	-	-	-	-	0.80
PAILO6-0022	-	-	S	-	-	-	-	1.10
PAILO6-0023	-	-	S	-	-	-	-	0.70
PAILO6-0026	-	-	S	-	-	-	-	0.67
PAILO6-0308	J	Y	S	194	84.67	16.87***	2.41*	0.73
PAILO6-0722	-	-	S	-	-	-	-	0.67[1]*
PAILO7-0020	-	-	S	-	-	-	-	1.00
PAILO7-2022	J	Y	S	155	76.33	8.30	2.74	1.33
PAILO8-0012	J	Y	S	152	-	-	-	0.97
PAILO8-0015	-	-	S	-	-	-	-	1.20
PAILO8-1018*	-	-	S	-	-	-	-	1.40[1]**

Figure S4.1: Graphical genotype of the E2T set with two and three almond introgressions in heterozygosity based on 113 SSR markers. Lines marked with asterisk (*) have three introgressions and the rest two introgressions each. Gray color corresponds to the peach 'Earlygold' background, red to the 'Texas' almond introgressions and light gray for missing data.

Figure S4.2: Graphical representation of the 206 heterozygous ILs of almond fragments (red) in the peach background (gray) based on SSR markers. G1 to G8 are the eight linkage groups of *Prunus*.

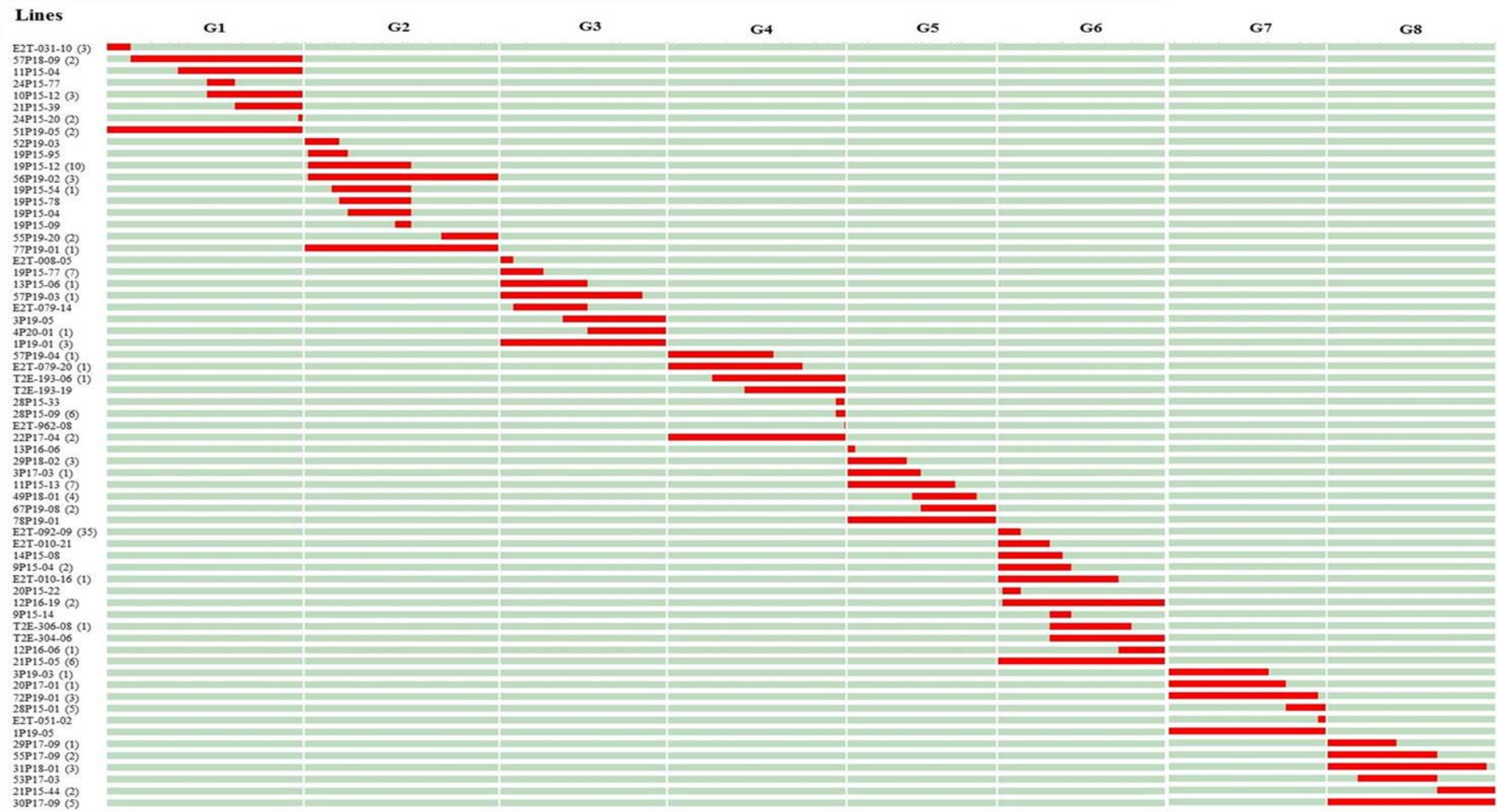
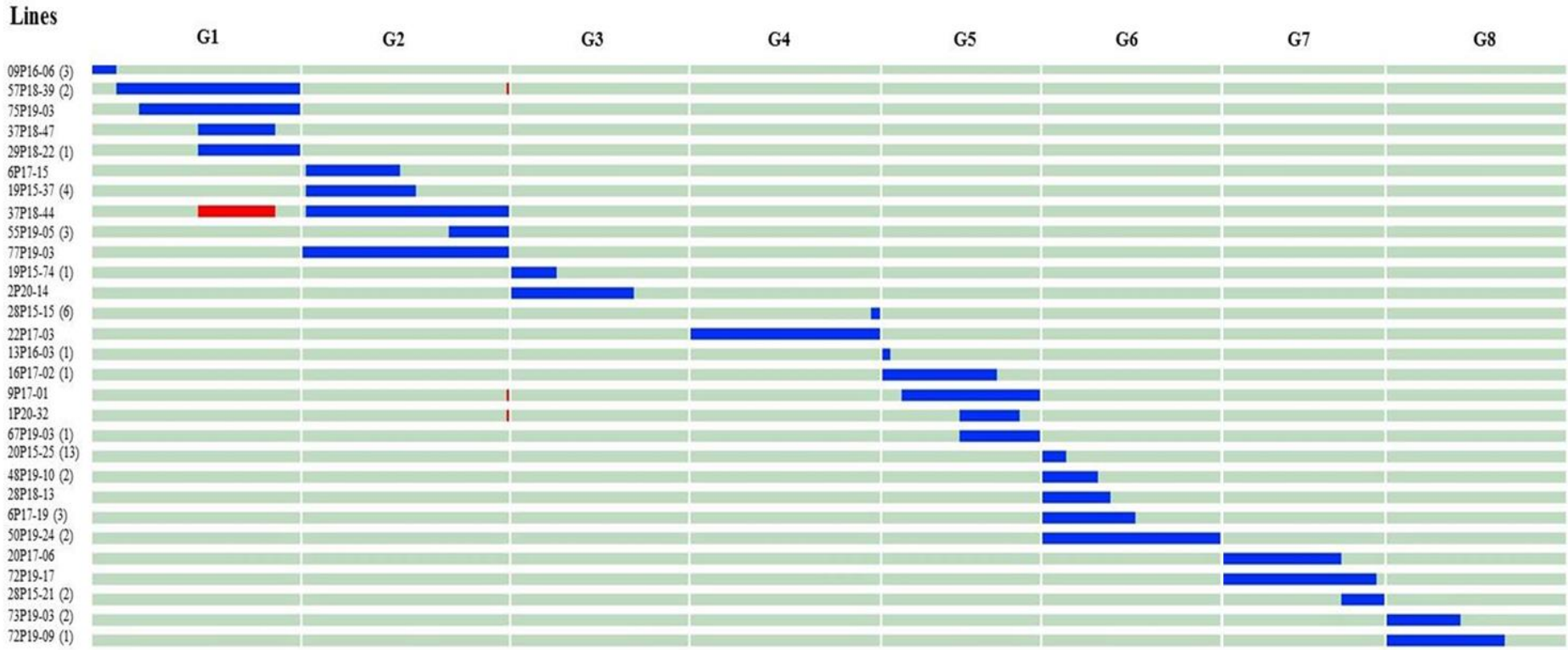


Figure S4.3: Graphical representation of the 77 homozygous ILs of almond fragments (blue) in the peach background (gray) based on SSR markers. Red fragments are heterozygous almond introgressions. G1 to G8 are the eight linkage groups of *Prunus*.



5. Fine mapping of three major genes identified in almond × peach interspecific crosses

5.1 Introduction

To introgress new and useful genetic variability from almond to peach, two almond × peach interspecific populations, an F2 (T×E) and a BC1 (T1E), were obtained from a cross between the almond cultivar 'Texas' and the peach cultivar 'Earlygold'. In these two populations, nine major genes were identified and mapped (Donoso et al. 2016) including two anther color genes (*Ag/ag* and *Ag2/ag2*), flower color (*Fc2/fc2*), flower type (*Sh/sh*; showy/non-sowy), powdery mildew resistance (*Vr3/vr3*), maturity date (*MD/md*), almond fruit type (*Alf/alf*; almond vs. peach), juiciness (*Jui/jui*) and blood flesh (*DBF2/dbf2*). In this chapter we describe the fine mapping of three of those genes (*Alf*, *Jui* and *DBF2*). Determining their position and identifying the candidate gene/s would help to understand the physiological mechanisms of these traits and would facilitate their selection in segregating populations.

The first two genes, *Alf* and *Jui*, determine the main differences between peach and almond fruits. Peach fruits have thicker juicy mesocarps with ripening capacity, while almonds have thin mesocarps that do not ripe. At early stages of the fruit development, almond and peach look very similar, but later on peach develops a thick mesocarp and reaches the ripening stage (Tonutti et al. 1991; Rodriguez et al. 2019). In almond, when the seed is completely developed, the mesocarp is dry and leathery (Martínez-Gómez et al. 2016). Thus, peaches are grown for their juicy fleshy fruit, while almonds for their edible seed. The *Alf* almond allele is recessive and causes the almond fruit phenotype. *Alf* was initially mapped in the T×E population in linkage group 4 (G4; Pp04: 10922662-12523245) in a genomic region of 6.3 cM and 1.6 Mb with 274 annotated genes in the peach reference genome v2.0 (Verde et al. 2017). The *Jui* peach allele is recessive and produces the juicy phenotype. *Jui* was mapped in the T1E population in G1 (Pp01: 35198093-35398680) in a genomic region of 8.3 cM and 200 kb with 38 annotated genes.

The third gene, *DBF2*, is responsible for the red flesh color of the fruit mesocarp. Although there are plenty of white and yellow flesh peaches available commercially, there are only a handful of red flesh peach commercial cultivars, such as 'Nectavigne', 'Diablotina' or the Indian blood peaches. Red flesh peaches contain high levels of anthocyanins, that are antioxidants that possess health benefits (Konczak and Zhang 2004). *DBF2* almond allele is dominant and responsible for the red flesh color. *DBF2* was mapped in the T1E population at the end of G1 (Pp01: 41709139-

42754924) in a genomic region of 2.5 cM and 1.04 Mb with 277 annotated genes. In addition to *DBF2*, two other peach genes producing fruit red flesh have been reported, *Bf*, mapped at the top of G4 (Gillen and Bliss 2005), and *DBF*, located at the top of G5 (Shen et al. 2013).

Sometimes, traits of agronomical importance for breeding such as fruit quality and disease resistance (Dirlewanger et al. 2004), or the ones presented in the previous paragraphs, are mapped as major genes covering a large genomic region with a high number of genes on it. To identify the genes responsible of these traits it is necessary to do a fine mapping of the region to increase mapping resolution and reduce the list of positional candidate genes as much as possible. This process involves the screening of many individuals with molecular markers flanking the target gene to select individuals that present a recombination between them. These recombinant individuals are used to saturate the genomic region of interest with more markers and to phenotype them to increase the mapping resolution (Jaganathan et al. 2020). The first fine mapping in fruit species was reported in tomato, where a QTL affecting fruit yield was fine mapped using an introgression line (IL) population (Eshed and Zamir 1995). In tree species, fine mapping is often more difficult, because obtaining and managing large number of individuals is a long and more expensive process, as they require a large amount of space and maintenance. In the case of peach, seed germination is another limitation, as the germination percentage is often quite low, especially for the early maturing individuals. In spite of these limitations, several fine mapping studies have been reported for different traits in different tree species including black peel color in pomegranate (Trainin et al. 2021), Alternaria brown spot resistance in citrus (Cuenca et al. 2016), black spot susceptibility in pear (Terakami et al. 2016), fire blight resistance in apple (Emeriewen et al. 2021) and powdery mildew resistance (Marimon et al. 2020), pollen sterility (Eduardo et al. 2020) and plant height (Lu et al. 2016) in peach.

The objective of this work was (i) to fine-map the three major genes *Alf*, *DBF2* and *Jui*, by identifying new recombinant individuals in the genomic regions where the genes were located and saturating it with molecular markers, (ii) to understand when these genes were expressed during fruit development using fruit images and (iii) to analyze the positional candidate genes by predicting the effects of the polymorphisms between the parental lines and performing an expression analysis during the fruit development.

5.2 Materials and methods

5.2.1 Plant materials

Ten individuals were selected based on their fruit characteristics to study the major genes *Alf, Jui* and *DBF2* during the fruit development. These were, the peach cultivar 'Earlygold', the almond cultivar 'Texas', the F1 hybrid MB1.37, three F2 individuals T×E15, T×E32 and T×E47 and four BC1 individuals T1E226, T1E427, T1E492 and T1E505. All these individuals were grafted on 'Garnem' rootstocks (Felipe 2009) and planted at a field station in Gimenells (Lleida, Spain). Standard agricultural practices were followed. Fruits from the ten individuals selected for the fruit development study were observed each two weeks since their pollination until they were ripened. They were measured (Figure 5.1), pictures were obtained, and fruits were harvested for the microscopic and gene expression analysis.

Two sets of individuals were used to obtain recombinant individuals. The first one was composed of 1,069 MB1.37 open pollinated individuals obtained from 2015 and 2017. The second set was composed of 2,770 open pollinated individuals derived from 54 T1E individuals that were obtained between 2018 and 2019. The 54 parental T1E individuals were heterozygous in the genomic regions where at least one of our target genes (*Alf, Jui* and *DBF2*) were located. All the T1E individuals selected were late maturing (to avoid embryo rescue) and fertile (able to produce fruits). The MB1.37 (F1) individual was located at the IRTA Experimental station of Cabrils (Spain), and the T1E individuals were situated at the IRTA Experimental station of Gimenells (Lleida, Spain). All of them were grafted on to 'Garnem' rootstocks. The spacing followed within and between the rows was 1.5 × 3.5 m.

5.2.2 Fruit measurements

Two or three fruits were collected each two weeks during fruit development, pictures were taken, and measurements were done using a ruler. These measurements included fruit diameter (red line - horizontal section of the fruit distance from one end to the other), mesocarp diameter (green line - distance between the stone and the pericarp) and stone diameter (black line - the diameter of the stone). All the measurements were always done in the section with maximum diameter (red line) (Figure 5.1).

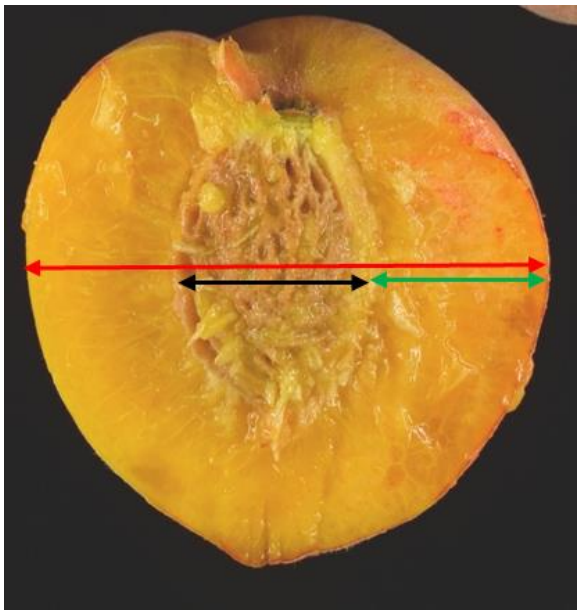


Figure 5.1: Transversal section of a peach fruit. Red arrow – fruit diameter, green arrow – mesocarp diameter and black arrow – stone diameter.

5.2.3 Seed germination

To identify recombinant individuals, open pollinated fruits from MB1.37 and selected T1E individuals were harvested from the field at fruit maturity (estimated from visual color change and manual evaluation of firmness) and stored in a cold chamber at 4ºC. The fruits were opened using a nutcracker, seeds were collected and cleaned in a solution of water and 30% bleach for 10 min in a falcon tube. After, they were washed with distilled water and rinsed in a fungicide solution (Merpan 1 g/l). Seeds were dried by letting them on a filter paper at room temperature overnight. Stratification was done in plastic containers with perlite and fungicide solution at 4ºC for 8-12 weeks. Every week humidity levels were monitored and, if necessary, they were sprayed with water to maintain the moisture. After the stratification period or at the onset of radical growth, seeds were transferred to trays with substrate and vermiculite (2:1) and finally moved to the greenhouse for germination.

5.2.4 DNA extraction

Once the seeds germinated, seedlings were labelled and a piece of leave from each plant was collected in a 96 well plate. For DNA extraction we used two methods. Initially we used, the CTAB method (Doyle and Doyle 1990) and later we adopted the alkaline lysis method (Lu et al. 2020),

as it was faster and cheaper. For the CTAB method, 340 µl of the CTAB solution were added to the well of the 96 well plates where leave tissue was collected. Then, plant tissue was crushed using tungsten balls and the TissueLyser (Qiagen, Germany) at 30 Hz (hertz) for 2 min. The solution was incubated at 65ºC for 40 min. After that, 340 µl of chloroform were added, mixed well and centrifuged at 3,000 rpm for 15 min to separate the supernatant. The supernatants were collected in a new 96 well plate and were precipitated with an equivalent amount of isopropanol and centrifuged for 30 min at 3,000 rpm to obtain a white DNA pellet. Two washing steps were performed with ethanol to remove the impurities. Finally, the pellets were dried and eluted in 100 µl Milli-Q H_2O. The alkaline lysis method is a two-step process. In the first one, 67 µl of 0.3 M NaOH solution and a tungsten ball were added to each sample into 96 well plates and grinded with the TissueLyser (Qiagen, Germany) at 30 Hz for 2 min. In the second step, the mixture was incubated at 96ºC for 1 min and 200 µl of 0.75 M Tris-HCl (pH 7.5-7.8) was added to it. The plate was shaken thoroughly and centrifuged at 3,000 rpm for 1 min to separate the supernatant. The DNA concentration and DNA quality were measured using the Nanodrop (Thermoscientific, USA).

5.2.5 Screening of recombinant individuals

Recombinant screening was performed by selecting two flanking markers for each gene (*Alf*, *Jui* and *DBF2*), one upstream and one downstream, according to Donoso et al. (2016). To further saturate the genomic region containing the genes of interest, different molecular markers including SSRs (simple sequence repeats), InDels (insertion/deletion) and SNPs (single nucleotide polymorphism) were designed using the resequencing data from 'Texas', 'Earlygold' and MB1.37 already available in the lab (Serra 2017). Raw Illumina data for 'Texas', 'Earlygold' and MB1.37 are available at the European Nucleotide Archive under the accession numbers ERS4540423, ERS3508161 and ERS4540424 respectively.

Integrative genomics viewer (IGV) software (Robinson et al. 2011) was used to visualize the resequencing data and identify polymorphisms. The peach genome sequence v2.0 (Verde et al. 2017) was used as reference. For SSRs, the SSR annotation from the IGA webpage was used (http://services.appliedgenomics.org/projects/drupomics/). For InDels, the insertions/deletions were detected at the region of polymorphism and markers were designed around them using the fasta sequence with Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/). SNPs were designed at the point

where there was a single nucleotide change between almond and peach sequences. Primer picker software (www.kbioscience.co.uk) was used to design the primers. A 200 bp fasta sequence was taken with at least 50 bp of the sequence present on either side of the SNP. All primers are presented in Tables 5.1, 5.2 and 5.3.

5.2.6 Genotyping with SSRs and InDels

For SSR and InDel genotyping, PCR was performed in reactions of 10 µl containing 1 µl of lab (10x) buffer, 0.3 µl of (50 mM) $MgCl_2$, 0.2 µl of (10 mM) dNTP, 0.2 µl each of forward and reverse primers (10 µM), 0.2 µl lab taq (5U/µl), 5.90 µl of Milli-Q H_2O and 2 µl of DNA (100 ng/µl). The forward primer of SSRs was labelled with one of the four fluorescent dyes (FAM, NED, VIC and PET) at its 5' end. Two generic tag primers, *tagF* (labelled with fluorescent dye at 5' end) and *tagR* were designed by Hayden et al. (2008) to reduce the cost of fluorescent primer labelling, and thus the forward and reverse primers (SSRs-tag) were designed with a tail identical to *tagF* and *tagR* respectively, to use *tag* primers in the assay of SSRs. The PCR reaction volume for the SSRs-tag was the same as mentioned above, with a few modifications: *tagF* and *tagR* primers each of 0.2 µl (10 µM), 0.04 µl each of forward and reverse primers (10 µM) and Milli-Q H_2O of 5.82 µl. For a faster and efficient SSR genotyping, a multiplex PCR can be performed, where several markers can be amplified in the same reaction (Hayden et al. 2008).

PCR amplifications were performed in a thermocycler (Applied Biosystems, USA) in a five-step process. For SSRs and InDels, initialization at 94ºC for 1 min, 35 cycles of denaturation (94ºC for 15 s, annealing temperature (T_a) specific for each primer for 15 s, elongation at 72ºC for 30 s) and a final elongation at 72ºC for 5 min. The PCR product was held at a temperature of 4ºC. The PCR amplification for SSR-tag includes an initial denaturation at 94ºC for 1 min, 60 cycles of denaturation: 20 cycles at 94ºC for 15 s, 63ºC for 30 s and 72ºC for 1 min, followed by 40 cycles at 94ºC for 15 s, 54ºC for 30 s and 72ºC for 1 min.

Table 5.1: Markers used in the fine mapping of *Alf*. Marker type includes SSRs (standard labelling), SSRs-tag (tag labelling) and InDels. The position is from *P. persica* v.2.0 reference genome. The markers designed in this study include, SSR15636, SS15662 and InDels. M12a and EPPCU2000 available from Donoso et al. (2016).

Marker type	Marker name	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'	Position (bp)
SSR	M12a	AGGTGCCTCATCTTCTTCTCT	GTGTGGTGAGGGGTGAGAGC	Pp04: 9,219,594
SSR	EPPCU2000	CACTTTGTTTCTCTCCTTGCTTC	TCCTATAGCCTTGCCTCGAC	Pp04: 12,478,769
SSR-tag	SSR15636	TCAGACCGCATTTCAAAGAA	ACAACATGTGGTGTGGTGGT	Pp04: 11,069,299
SSR-tag	SSR15662	CGAAATGCACTC	TTGGTCTCCCA	Pp04: 11,228,933
InDel	InDel11067	AGATTGGGAGTTGGTATGGCC	GCTGCTGTTTGCTTGGATCC	Pp04:11,067,095
InDel	InDel11120	TGGACCATAAGGAAGCCAAG	TTGAACCTAACCATGCCGTA	Pp04:11,119,601
InDel	InDel11147	AGATGGAGTCCAGAGCCAGT	AGGCGTCTTCTGGCTTAGTG	Pp04:11,146,913
InDel	InDel11163	CCAAAAGGCGGAGAAAGGTA	GGAAAATTAGAAAATCACTAGGAGGA	Pp04:11,163,343
InDel	InDel11172	CATTCAGCCTCATCAACCTGT	CAGAAGATTCTGCAGCCACA	Pp04:11,172,926
InDel	InDel11175	GCGATGAAGGTCCCACAGT	TTGTTGCACATCTCAAGTCAAG	Pp04:11,175,774
InDel	InDel11191	CCGATGCCGTATATGTGTGT	CATTGCATAGTGGGTTTGAGTT	Pp04:11,191,991
InDel	InDel11196	ACACGTTGCTTTCGCTTTCT	TGAAAAGACACCAGCACAGC	Pp04:11,196,845
InDel	InDel11206	TGGAGGCTTACATGGCTTTC	CCAATGGTAACAAAAACTCTTGG	Pp04:11,206,105
InDel	InDel11210	CATCAGCAGGTAAGGTGCCA	ACTCAAGGAGGCTTCAGGGA	Pp04:11,210,233
InDel	InDel11254	TAGGGCCAGCTGTCATTTTT	TGATGTGGCATGTTGCCTAT	Pp04:11,253,858
InDel	InDel11303	GCCTGCAATGAAGAAAGGTC	TTGGGTAAGTGCCTTTGGAC	Pp04:11,303,453

Table 5.2: Markers used in the fine mapping of *DBF2*. Marker type includes SSRs (standard labelling), SSRs-tag (tag labelling), InDels and SNPs (two forward primers and a common reverse primer). The position is from *P. persica* v.2.0 reference genome. The new markers developed in this study were, SSR6105, SSR6125, InDels and SNPs. CPPCT029 and BPPCT028 described in Donoso et al. (2016).

Marker	Marker name	Forward primers sequence 5'-3'		Reverse primer sequence 5'-3'	Position (bp)
SSR	CPPCT029	CCAAATTCCAAATCTCCTAACA		TGATCAACTTTGAGATTTGTTGAA	Pp01: 41,168,265
SSR	BPPCT028	TCAAGTTAGCTGAGGATCGC		GAGCTTGCCTATGAGAAGACC	Pp01: 44,130,041
SSR-tag	SSR6105	AATGCAGAGACAGGGAGGAC		GCTGTTGCTGTTGGTATTGC	Pp01: 42,638,623
SSR-tag	SSR6125	GTCAAGTGTGCTCCACATGC		CAGTGGGTCGGTGCTAATTT	Pp01: 42,723,336
InDel	InDel42115	CGGAGAGTTTTCGATGACA		AACATAAGAGAGTTCACGGG	Pp01:42,115,821
InDel	InDel42311	AGTGGTGTTCTCTATCTCCA		AACCAAACATGAATCGCAG	Pp01:42,311,741
InDel	InDel42471	TGGGCTCTCTGCTCTGTTCT		TGGTGTACGCAATTATCGTG	Pp01:42,471,315
InDel	InDel42525	TCGGATTTCTGTTAACGGTT		GAAAGGAAGAAGAGGTGACC	Pp01:42,525,134
InDel	InDel42584	GAAGGCAGGGATTCCTTTTC		TGTTATGTGTCCCTGCTCCA	Pp01:42,584,009
InDel	InDel42666	TTTGTGAATTTTTATTTCCCTCTC		CATTTTTACAAGGCATTCTTCA	Pp01:42,666,650
InDel	InDel42718	GGCCAATAACGAAATTTACCA		ATGGTAACTGCCAACCCATT	Pp01:42,718,200
SNP	SNP42675	CCAAATGACCCGATTCTGGATTTC	CCCAAATGACCCGATTCTGGATTTA	GATGGCAGAAGCTCTTTTAGCTCGTT	Pp01: 42,675,790
SNP	SNP42677	CCCATAAAATTACCACTATTTGGCCAA	CCATAAAATTACCACTATTTGGCCAG	AAAAGGAGGGAAGGGCTTTGAAATGAAAA	Pp01: 42,677,206
SNP	SNP42687	AATTCATACTTATGGATATTATGCTAATATTA	AATTCATACTTATGGATATTATGCTAATATTG	AAGATAAAAGTTAAAGTTAAAAAAGCAACC	Pp01: 42,687,251
SNP	SNP42696	ATTTGTAACTAATAATATAATATTCTCTAACTA	TGTAACTAATAATATAATATTCTCTAACTG	GTGAACAAAAATTAATTACAAACATGTTAT	Pp01: 42,686,025

Table 5.3: Markers used in the fine mapping of *Jui*. Marker type includes SSRs (standard labelling), SSRs-tag (tag labelling), InDels and SNPs (two forward primers and a common reverse primer). The position is from *P. persica* v.2.0 reference genome. The markers in the below list, EPDCU3489, BPPCT016 and SNP109223 already mentioned in Donoso et al. (2016). The newly designed markers include, SSR4996, the rest of SNPs and all InDels.

Marker	Marker name	Forward primers sequence 5'-3'		Reverse primer sequence 5'-3'	Position (bp)
SSR	EPPCU3489	AAATCAGCTCCCATCACTCC		AGCTGAGTGGAACCAGAGGA	Pp01: 34,001,221
SSR	BPPCT016	GATTGAGAGATTGGGCTGC		GAGGATTCTCATGATTTGTGC	Pp01: 37,047,997
SSR-tag	SSR4996	TGCTTTGGTAGCAAAAACCA		TTCCTGACGATGCAGATGAG	Pp01: 35,251,878
InDel	InDel34892	AGGCTTCCTCCCAGAAAGAG		TGTGTGTGCTTGCATCTTGA	Pp01:34,892,317
InDel	InDel34907	ATAGCGTGCCAGGTGTTTTC		CCGTACCCTTCTTTCCTGTG	Pp01:34,907,350
InDel	InDel34933	ATGCCGTTAGGTTTGTAGGC		CGTGTACGTCACCACTGTCC	Pp01:34,933,506
InDel	InDel34972	GGATGCAGAGTCACCTGGAC		ACTTTGGGGTCCCATAAACC	Pp01:34,972,191
InDel	InDel35000	TGCTCCAACACCAGAGAAAA		GATTTAAGGGTGGGTGACCA	Pp01:35,000,890
InDel	InDel35036	TTGCTTGTGTTCTTCGGTTG		AGGATCCTTCCAGCATCAGA	Pp01:35,036,070
InDel	InDel35085	TTGGTTGAAGACCACTTTTTGTT		TTTGATAAAGTTCCTTTGCCTCT	Pp01:35,085,392
InDel	InDel35147	TTCAAGGTGGCAATCTGTAAAA		AATTACATCGCCAAAAATACCG	Pp01:35,147,490
InDel	InDel35173	TACTTTTTGGTCCCATGAGTTGT		TGTTTATTTGGGGTTTAGTCCAG	Pp01:35,173,747
InDel	InDel35245	TGGAGGAAAAGTCGGATTTG		GAATTAATGAAGGTTAATGGCACA	Pp01:35,245,432
InDel	InDel35271	CAGCATTTAATTTCCTGTTCCTG		CATCAAGAACGCAATCATCTACA	Pp01:35,271,415
InDel	InDel35336	CCGTTCCCTGGATTTACTGA		CCCCATCAATGAAAGCAAAT	Pp01:35,336,891
InDel	InDel35351	GGATGGAGATTCCCAAAACTATG		GCAAATTTTGGCATTGAAAACT	Pp01:35,351,260
SNP	SNP35183	GTGATATTACAGACACAATTAAAGAGACA	GTGATATTACAGACACAATTAAAGAGACG	CGGTCTGTCTGTCTGTCTTTTGATTAATT	Pp01:35,183,714
SNP	SNP35194	GCAGCCTTAAATAAAACTTACAACATATG	AATGCAGCCTTAAATAAAACTTACAACATATT	GAGGGATATGGAGAGAGCAAGAGAT	Pp01:35,194,967
SNP	SNP35200	CAAATTGACAAATGCGTAAAGAAGATGAAA	AAATTGACAAATGCGTAAAGAAGATGAAG	AACTTCCAATCATCCTGCAGTTTCCAATT	Pp01:35,200,523
SNP	SNP35204	GCCACGTGCACTCAGCTGC	GCCACGTGCACTCAGCTGG	CGGTGCAGCCGGGGGTGTT	Pp01:35,204,671
SNP	SNP35211	GAAAAGGATAATAACAGTCCAAATGCTC	ATGAAAAGGATAATAACAGTCCAAATGCTT	CCAGTTTTCTTATAGAAGAAGCATTTCAAT	Pp01:35,211,229
SNP	SNP35224	GGATATCTAGATAGAACTAAAGGATACAG	AGGATATCTAGATAGAACTAAAGGATACAT	CTACAAATCTGGTAGTGTTTTTAGGACAAT	Pp01:35,224,589
SNP	SNP35244	CCAATGCAAGGCATGTTCTGCC	CCCAATGCAAGGCATGTTCTGCT	AGGCGGAGAGAAAAGGGTTGAAGAT	Pp01:35,244,343
SNP	SNP35247	AGTAGTCGTGACTAGGAACAGAAAC	GAGTAGTCGTGACTAGGAACAGAAAT	GGACTTGGAGTTTGGAAGGAAGGTA	Pp01:35,247,471
SNP	SNP35252	TGCTTTCAGGACAGGGCCAC	GTTGCTTTCAGGACAGGGCCAA	CCTCAAGGGGGAGCCTCTAATTATT	Pp01:35,252,325

Table 5.3 (Continued)

Marker	Marker name	Forward primers sequence 5'-3'		Reverse primer sequence 5'-3'	Position (bp)
SNP	SNP35264	CAAGAGGCCAAGAATATGAAATTTTACG	CCAAGAGGCCAAGAATATGAAATTTTACA	GATTGAACCCCCTTTGATATTTTTTTAAA	Pp01:35,264,164
SNP	SNP35224	GGATATCTAGATAGAACTAAAGGATACAG	AGGATATCTAGATAGAACTAAAGGATACAT	CTACAAATCTGGTAGTGTTTTTAGGACAAT	Pp01:35,224,589
SNP	SNP35244	CCAATGCAAGGCATGTTCTGCC	CCCAATGCAAGGCATGTTCTGCT	AGGCGGAGAGAAAAGGGTTGAAGAT	Pp01:35,244,343
SNP	SNP35247	AGTAGTCGTGACTAGGAACAGAAAC	GAGTAGTCGTGACTAGGAACAGAAAT	GGACTTGGAGTTTGGAAGGAAGGTA	Pp01:35,247,471
SNP	SNP35252	TGCTTTCAGGACAGGGCCAC	GTTGCTTTCAGGACAGGGCCAA	CCTCAAGGGGGAGCCTCTAATTATT	Pp01:35,252,325
SNP	SNP35264	CAAGAGGCCAAGAATATGAAATTTTACG	CCAAGAGGCCAAGAATATGAAATTTTACA	GATTGAACCCCCTTTGATATTTTTTTAAA	Pp01:35,264,164
SNP	SNP35269	TCATTTATTATTTTGACTTGTGATTTATTGGC	CATTTATTATTTTGACTTGTGATTTATTGGG	CTTCCTGAAGAGCTTCACAAAATAAATGTA	Pp01:35,269,192
SNP	SNP35545	GTTCAGACATCATTTCCTCTCTAACT	GTTCAGACATCATTTCCTCTCTAACC	AATGTCCTGTAAGCAGGTTATCATTCAATT	Pp01:35,545,389
SNP	SNP35644	GGTCTTCTTTGATTAACCCTAAAATGGA	GTCTTCTTTGATTAACCCTAAAATGGG	GTGCCCGCACCTAATTTCCAGGTT	Pp01:35,644,865

For SSRs, 2 µl of PCR product along with 12 µl of Hi-Di formamide (Applied Biosystems, USA) and 0.35 µl of GeneScan 500 LIZ (Applied Biosystems, CA, USA) were used as a reaction mix. Three PCR products each of 2 µl can be mixed for analysis if they contain different fluorochromes and PCR fragments of different size. Even four PCR products can be combined, but in that case the volume of Hi-Di formamide should be 22 µl. This mix was denatured at 94ºC for 3 min and capillary-electrophoresed with ABI Prism 3130 Genetic analyzer (Applied Biosystems, USA). The data was analyzed using the Genemapper 5.0 software (Applied Biosystems, USA) to determine SSR allele sizes. For InDels, 10 µl PCR product was mixed with bromophenol blue dye (10x) prior to loading into agarose gel. DNA was separated with electrophoresis and bands visualized in agarose gels (2%). Separation of DNA fragments were observed based on their size and the size was estimated in reference to a DNA ladder.

5.2.7 KASPar SNP genotyping

KASPar genotyping was performed using a PCR platform, LightCycler 480 (Applied Biosystems, USA) with the program conditions: 1 cycle of hot-start 94ºC for 15 min at a ramp rate 4.4ºC/s, 10 cycles of touch-down 94ºC for 20 s at 4.4ºC/s and 61ºC for 1 min at 2.2ºC/s, 26 cycles of PCR 94ºC for 20 s at 2.2ºC/s and 55ºC for 1 min at 2.2ºC/s, 1 cycle of read plate 37ºC 1 min at 2.2ºC/s and 37ºC for 1 s at 4.4ºC/s. PCR mix for each reaction contained 4 µl of master mix of KASPar (2x), 0.11 µl of KASPar assay, 1.89 µl of Milli-Q H_2O and 2 µl (100 ng/µl) DNA. KASPar assays were prepared by mixing three primers, two forward and one common reverse A1, A2 and C1 (Table 5.2 and 5.3) respectively in the concentrations 12 µM, 12 µM and 30 µM. The results were analyzed using Fluidigm SNP genotyping analysis software (Smith and Maughan 2015).

5.2.8 Phenotyping

The almond fruit (ALF) trait was scored as almond or peach type depending on the development of a fleshy mesocarp and a change in external and internal fruit color at maturity. Fruits that remain green and firm were classified as almond type and fruits with a change in skin color from visual observation, thick mesocarp development and loss of firmness were classified as peach type. Juiciness (JUI) was scored as a fruit capacity to produce juice or not at maturity and blood flesh

(DBF2) was scored as the presence or absence of fruit red flesh color at maturity (Donoso et al. 2016).

5.2.9 SNP effect (SnpEff)

SnpEff software (Cingolani et al. 2012) was used to annotate and predict SNP and InDel effects in the *Alf* region (Pp04: 11119601-11303453) and in the *DBF2* region (Pp01: 42677206-42687251). The input file for SnpEff analysis is in Variant Call Format (VCF). Annotated genomic locations include exon, intron, gene upstream, gene downstream, intergenic region, splice site region. The genetic variant effects by impact were classified as high (STOP codon), moderate (amino acid change), low (codon change with same amino acid) and modifier (no evidence of impact).

5.2.10 RNA extraction

To study the expression of major genes, *Alf* and *DBF2* during the fruit development, mesocarp tissue from the parents ('Texas', 'Earlygold'), F1 hybrid (MB1.37), F2 (T×E15 and T×E32) and BC1 (T1E492 and T1E505) individuals was used for RNA extraction. RNA was extracted from fruits of the above individuals at different developmental stages, from initial fruit formation to until fruit maturity. The sampling was done every two weeks. Fruits were collected from the field, the skin was peeled, flesh was extracted, cut into small pieces, and immediately frozen in liquid nitrogen and stored at -80ºC. Prior to the extraction, the fruit material was ground into fine powder using mortar and pestle. RNA extraction was done using the Spectrum Plant Total RNA kit (Sigma Aldrich, Germany): To 100 mg of tissue powder, 500 µl lysis solution were added for tissue lysis, incubated at 56ºC for 5 min, centrifuged and the lysate pipetted through a filtration column. To the clear lysate, 500 µl binding solution was added. This mix was then passed through a binding column to bind RNA to the column. The column was then washed with 500 µl each of wash solution 1 and wash solution 2, dried and RNA was eluted into 50 µl elution solution. To remove any traces of genomic DNA, a DNAse treatment was done using the turbo DNA-free kit (Invitrogen, Lithuania): To 50 µl RNA, 5 µl of turbo DNAse buffer (10x) and 2 µl turbo DNAse enzyme were added and incubated at 37ºC for 30 min. To this mix, 5 µl DNAse inactivation reagent was added, kept at RT for 5 min and centrifuged to collect the supernatant. The extracted RNA quality and quantity were measured using Nanodrop absorbance (260/230) values and bands detected in a 1% agarose gel.

5.2.11 Expression analysis

Three biological replicates were assayed for each time point per individual. The samples were sent to Novogene (UK) to perform RNA-seq analysis. RNA-seq results obtained were analyzed by using two programs, Kallisto v.0.46.1 (Bray et al. 2016) and Sleuth v.0.30.0 (Pimentel et al. 2017). Kallisto is an RNA-seq quantification program used for sequence data alignment. It works by pseudoaligning reads to a reference genome (in our case is almond and peach) and generates transcripts that are compatible with each read and also avoids individual base alignments. Kallisto has high performance (speed) and accuracy among the existing RNA-seq quantification tools. Sleuth program is used for the differential analysis of gene expression data, which relies mainly on the biological differences in transcript or gene expression. It has a high control for the false discovery rate (FDR), which is key for identifying differentially expressed genes, especially in experiments with few replicates. Sleuth is executed through an interactive shiny app, which utilizes Kallisto quantifications and bootstraps for a quick and precise analysis of data from RNA-seq experiments. It has a higher sensitivity compared to other methods that are used for differential analysis of gene expression data.

5.3 Results

5.3.1 Study of peach and almond fruit development

To better understand when our traits of interest were expressed and what could be the physiological mechanism producing the differences between almond and peach fruits, we observed and measured the fruits of different individuals during fruit development (Figure 5.2). The selected individuals were ‘Texas’, ‘Earlygold’, MB1.37, three F2 individuals (T×E15, T×E32 and T×E47) and four BC1 individuals (T1E492, T1E226, T1E505 and T1E427). They were selected based on *Alf*, *Jui* and *DBF2* phenotypes, including early and late maturity individuals. Two individuals produced almond type fruits (‘Texas’ and T×E15), three individuals produced non-juicy red flesh late ripening fruits (MB1.37, T1E226 and T1E492), three individuals produced juicy and early ripening fruits (‘Earlygold’, T×E32 and T1E505), and two individuals produced juicy and late ripening fruits (T×E47 and T1E427).

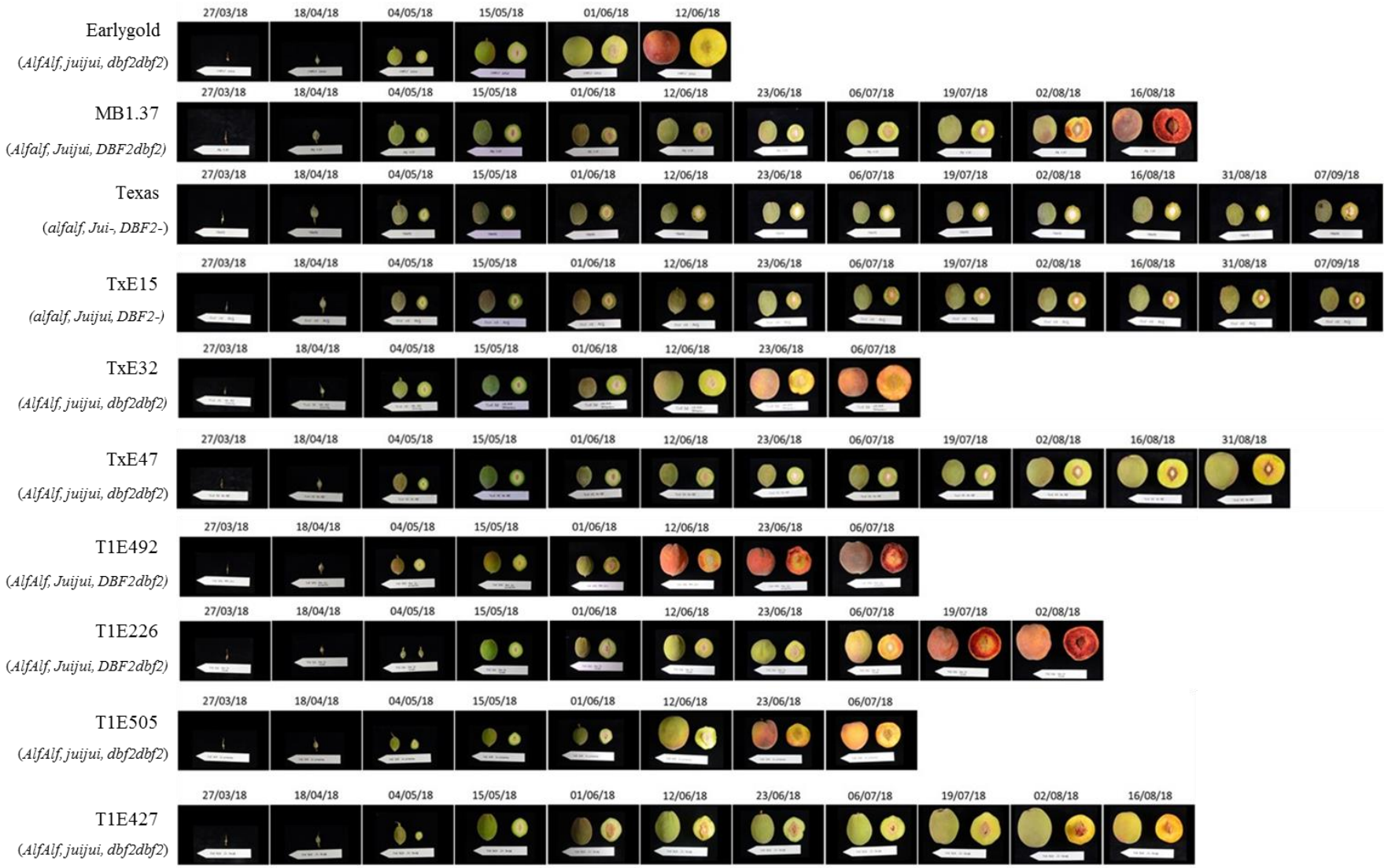


Figure 5.2: Fruit development pattern with an interval of every two weeks from 30 DAF (days after flowering) until the fruit maturity, in the parents ('Earlygold', 'Texas'), F1 hybrid (MB1.37), F2 (T×E15 and T×E32) and BC1 (T1E492, T1E226, T1E505 and T1E427) individuals. The allelic configuration of *Alf*, *Jui* and *DBF2* for each individual are indicated in the parenthesis.

5.3.1.1 Fruit growth and ripening and flesh color during fruit development

We could observe four different patterns of ripening behavior during fruit ripening. The first one was for the almond type fruit individuals (‘Texas’ and T×E15) whose fruit mesocarp was very thin and never became ripen. The other three patterns were for the peach like fruit individuals, that were early ripening (‘Earlygold’), intermediate ripening (T×E32, T1E492 and T1E505) and late ripening (MB1.37, T×E47, T1E226 and T1E427).

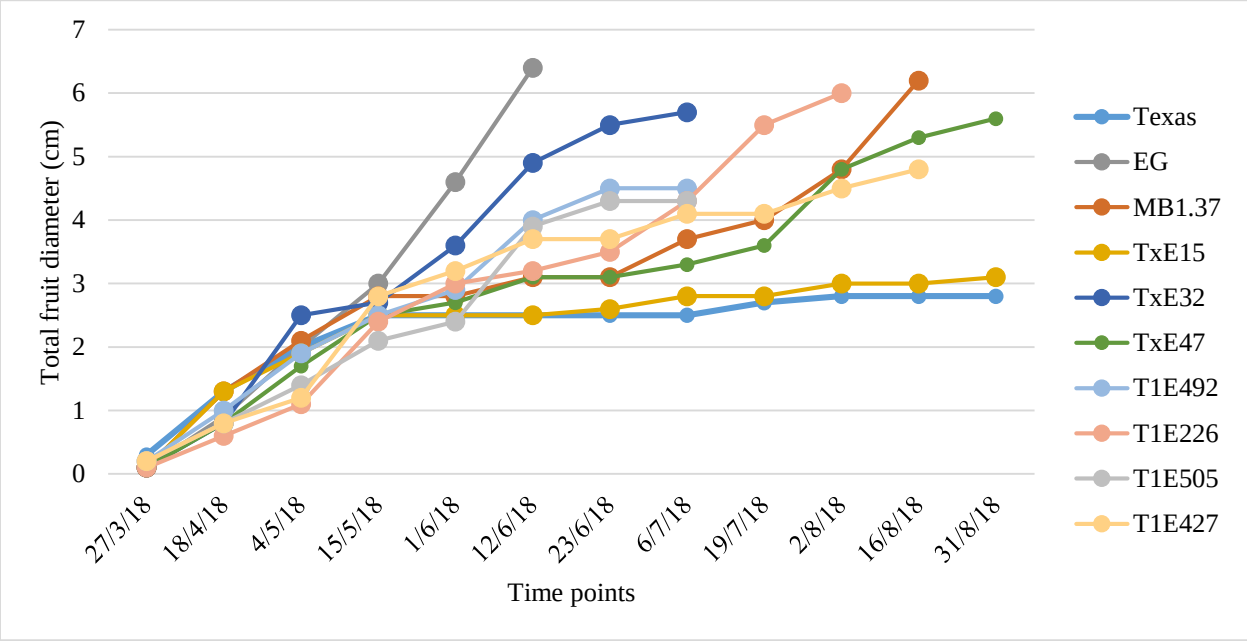


Figure 5.3: Fruit diameter measurements of almond × peach individuals for every two weeks from 30 DAF until the fruit maturity.

Fruit diameter, mesocarp thickness and stone diameter were measured during fruit ripening. Fruit diameter and mesocarp thickness presented a very similar evolution (Figures 5.3 and 5.4). Until the 15th of May (approximately eight weeks after pollination), all the individuals presented the same growth pattern. From this point on we could observe three growing patterns. Almond type fruits (‘Texas’ and T×E15) practically did not increase their diameter until the fruits were harvested and their mesocarp did not get ripen. Early ripening peach type individuals continued with an exponential fruit growth until fruit maturity. Finally, medium and late ripening peach type individuals presented two exponential growth phases separated by a low or no growth phase. In the case of stone diameter, all the individuals, including peach and almond types, presented the

same growth pattern (Figure 5.5). The size increased from 4th of May to 12th of June. At this point lignification started and the stone dimeter did not increase until the harvest date.

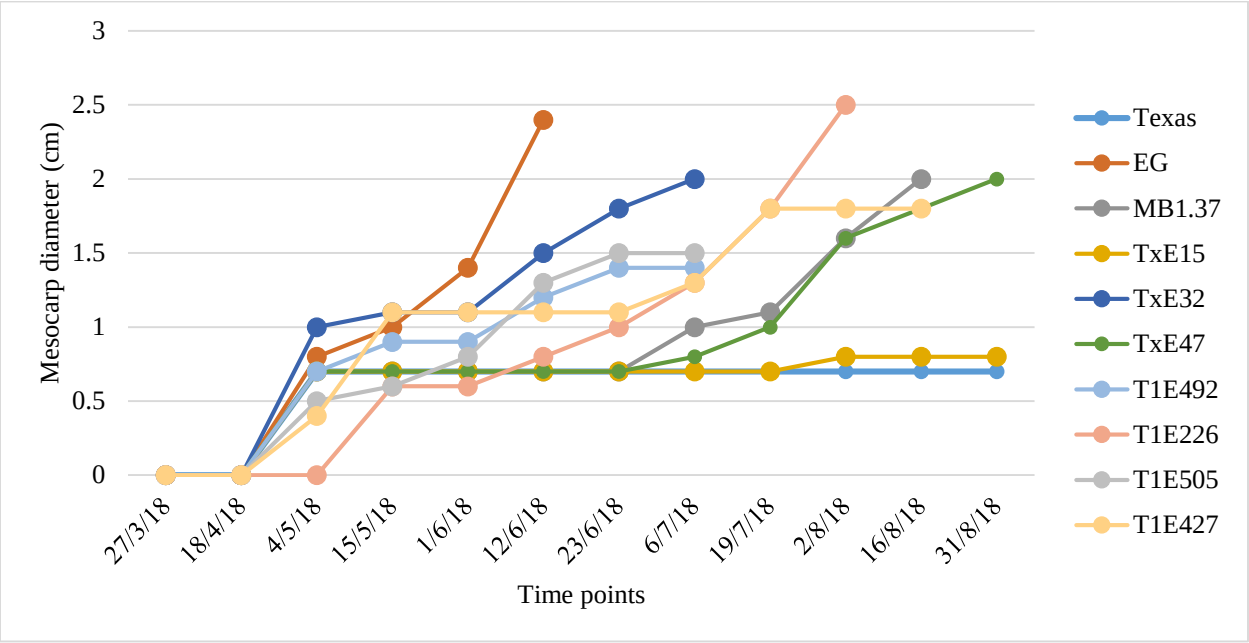


Figure 5.4: Mesocarp diameter measurements of almond × peach individuals for every two weeks from 30 DAF untill the fruit maturity.

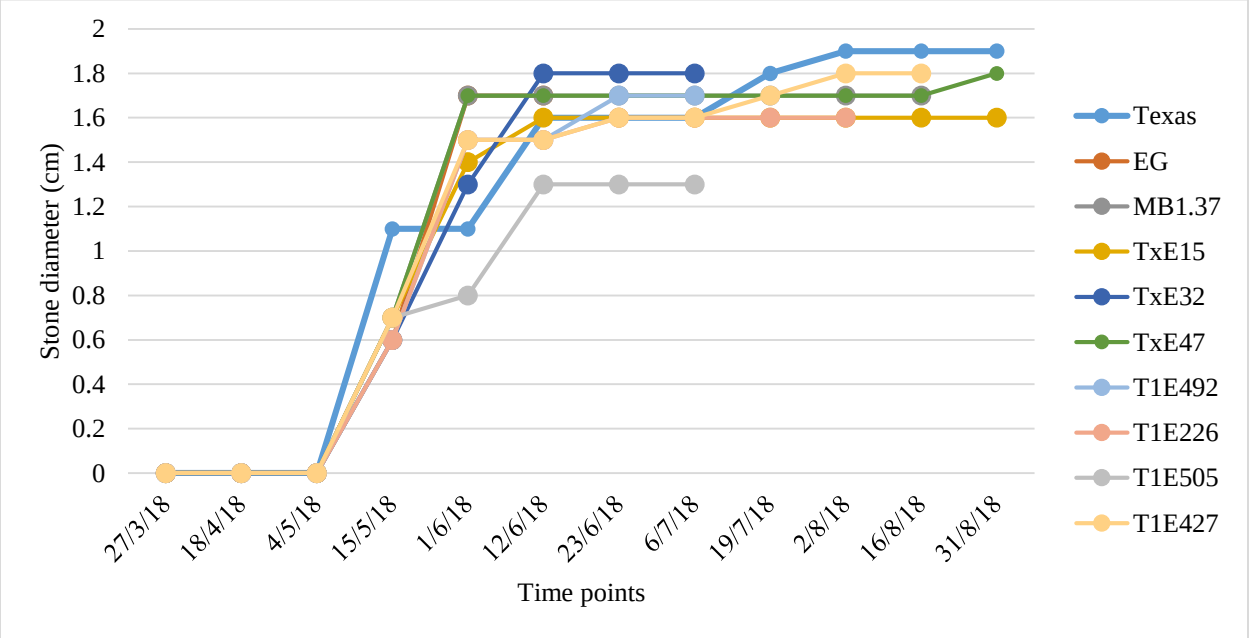


Figure 5.5: Stone diameter measurements of almond × peach individuals for every two weeks from 30 DAF untill the fruit maturity.

Focusing on the mesocarp color we could observe that the red flesh color of the individuals presenting the *DBF2* almond allele (MB1.37, T1E226 and T1E492) started to appear approximately two weeks before fruit ripening. We could also observe that the fruits from the T1E492 individual presented yellow color around the stone at the fruit ripening stage, indicating that this population is also segregating for this trait.

The juicy character started to be appreciated in juicy fruits ('Earlygold', T×E32, T×E47, T1E427 and T1E505) very close to the fruit ripening stage, when the mesocarp started to lose its firmness.

Based on these observations we selected the key three time points to collect samples for the expression analysis study. The first one was the 4th of May, where all individuals presented a similar size and stage of development. The second one represented an intermediate stage, and the date was different for each individual (Table 5.4). Finally, the last one was the ripening stage that corresponded to the harvest date of each individual.

Table 5.4: Time points (S1, S2 and S3) where fruit samples were collected for each individual for expression analysis.

				Early stage	**Intermediate stage**	**Ripening stage**
Genotype	Fruit type	Juiciness	Flesh color	S1	S2	S3
Texas	Almond	-	-	04-May	06-Jul	07-Sep
T×E15	Almond	-	-	04-May	06-Jul	07-Sep
Earlygold	Peach	Juicy	Yellow	04-May	01-Jun	12-Jun
T×E32	Peach	Juicy	Yellow	04-May	12-Jun	06-Jul
T1E505	Peach	Juicy	Yellow	04-May	12-Jun	06-Jul
MB1.37	Peach	Non-juicy	Red	04-May	06-Jul	16-Aug
T1E492	Peach	Non-juicy	Red	04-May	12-Jun	06-Jul

5.3.2 Fine mapping of *Alf*

The *Alf* gene was initially located between the markers SNP_IGA_410955 and EPPCU2000 in an interval of 1.6 Mb from the physical position Pp04: 10922662-12523245 (Donoso et al. 2016). To increase the map resolution, we developed a set of molecular markers, including 12 InDels and two SSRs, based on the resequencing data from the parental lines. These markers were genotyped in the two TxE individuals determining the position of the gene in the T×E population (T×E59 and T×E85), for which the phenotype was already available. This data allowed us to reduce the region where *Alf* was located to a region of 183 kb (Pp04: 11119601-11303453), between the markers InDel11120 and InDel11303 (Table 5.6). To increase the mapping resolution even further, we looked for new recombinant individuals in this region. For that, we screened, between 2015 and 2019, 3,828 plants derived from open pollinations of MB1.37 (1,069) or selected T1E individuals (2,759). We used the two SSRs flanking the *Alf* gene, M12a and EPPCU2000. We initially used SSRs because they can be efficiently genotyped using the ABI sequencer. As *Alf* is recessive for the almond allele, the only useful recombinant individuals selected were those that were homozygous for the almond allele for one of the markers and heterozygous or homozygous for the peach allele for the other marker. From the screenings performed with these two SSRs, 403 individuals were identified as recombinants (Table 5.5) and 910 individuals were discarded because of the missing data at least for one marker. The recombinant individuals were then genotyped with InDel11120 and InDel11303, reducing the number of recombinants to 13. Finally, these 13 individuals were genotyped with all the set of markers developed in the region to confirm the genotypes (Table 5.6). These 13 individuals have not produced any fruit yet, therefore the position of *Alf* could not be further reduced.

Table 5.5: Individuals genotyped for fine mapping of the *Alf* gene.

Year	Population	Genotype code	Female parent	Individuals genotyped	Recombinants
2015	F2	T×E	MB1.37	185	28
2017	F2	51P17	MB1.37	884	84
2018	BC1S1	71P18	T1E-020	2	-
2018	BC1S1	72P18	T1E-021	33	3
2018	BC1S1	74P18	T1E-024	12	-
2018	BC1S1	79P18	T1E-034	32	2
2018	BC1S1	80P18	T1E-035	18	2
2018	BC1S1	84P18	T1E-040	25	1
2018	BC1S1	85P18	T1E-043	2	1
2018	BC1S1	93P18	T1E-064	259	20
2018	BC1S1	99P18	T1E-079	22	3
2018	BC1S1	103P18	T1E-096	5	1
2018	BC1S1	106P18	T1E-101	1	-
2018	BC1S1	107P18	T1E-104	19	3
2018	BC1S1	108P18	T1E-108	2	1
2018	BC1S1	111P18	T1E-123	12	1
2018	BC1S1	118P18	T1E-197	4	-
2018	BC1S1	120P18	T1E-241	2	-
2018	BC1S1	123P18	T1E-219	5	-
2018	BC1S1	129P18	T1E-239	1	-
2018	BC1S1	134P18	T1E-333	5	-
2018	BC1S1	135P18	T1E-335	67	2
2018	BC1S1	150P18	T1E-463	21	3
2018	BC1S1	151P18	T1E-467	12	-
2018	BC1S1	156P18	T1E-500	24	3
2019	BC1S1	96P19	T1E-021	39	1
2019	BC1S1	102P19	T1E-032	28	4
2019	BC1S1	103P19	T1E-034	241	39
2019	BC1S1	104P19	T1E-035	123	24
2019	BC1S1	108P19	T1E-040	61	2
2019	BC1S1	109P19	T1E-043	15	2
2019	BC1S1	117P19	T1E-064	234	20
2019	BC1S1	118P19	T1E-065	14	1

Table 5.5 (Continued)

Year	Population	Genotype code	Female parent	Individuals genotyped	Recombinants
2019	BC1S1	130P19	T1E-101	82	11
2019	BC1S1	131P19	T1E-104	180	23
2019	BC1S1	133P19	T1E-116	32	2
2019	BC1S1	135P19	T1E-123	98	13
2019	BC1S1	142P19	T1E-197	76	5
2019	BC1S1	145P19	T1E-201	127	13
2019	BC1S1	147P19	T1E-219	21	4
2019	BC1S1	148P19	T1E-220	113	22
2019	BC1S1	149P19	T1E-226	77	4
2019	BC1S1	162P19	T1E-344	45	1
2019	BC1S1	165P19	T1E-389	226	19
2019	BC1S1	167P19	T1E-410	63	8
2019	BC1S1	170P19	T1E-424	50	4
2019	BC1S1	175P19	T1E-463	89	10
2019	BC1S1	176P19	T1E-467	43	8
2019	BC1S1	181P19	T1E-500	97	5
Total				3828	403

Table 5.6: Genotypes and phenotypes of the 15 recombinant individuals. The two red lines denote the recombination breakpoint. The two flanking markers, InDel11120 and InDel11303 define the *Alf* genomic region of 183 kb. A (almond fruit) allele from the parent 'Texas' and H (peach fruit) allele from the parent 'Earlygold'/hybrid (MB1.37), missing data (-).

Marker name	**Group**	**Position in bp (v2.0)**	T×E59	T×E85	103P19 -22	117P19 -125	142P19 -75	165P19 -84	103P19 -148	109P19 -04	104P19 -113	130P19 -54	181P19 -13	145P19 -88	103P19 -38	148P19 -07	117P19 -105
M12a	Pp04	9,219,594	A	A	A	A	A	A	A	A	A	A	H	H	A	A	H
SSR15636	Pp04	11,069,299	A	A	A	A	A	A	A	A	A	A	H	H	A	A	-
InDel11067	Pp04	11,067,095	A	A	A	A	A	A	A	A	A	A	H	H	A	A	-
IndDel11120	Pp04	11,119,601	A	A	A	A	A	A	A	A	A	A	H	H	A	A	-
InDel11147	Pp04	11,146,913	A	H	H	H	H	H	H	A	A	A	H	H	A	A	-
InDel11163	Pp04	11,163,343	A	H	H	H	H	H	H	H	A	A	H	H	A	A	-
InDel11172	Pp04	11,172,926	A	H	H	H	H	H	H	H	A	A	H	H	A	A	-
InDel11175	Pp04	11,175,774	A	H	H	H	H	H	H	H	A	A	H	H	A	A	-
Alf	**Pp04**	-	**A (Almond)**	**H (Peach)**	-	-	-	-	-	-	-	-	-	-	-	-	-
InDel11191	Pp04	11,191,991	A	H	H	H	H	H	H	H	H	A	H	H	A	A	-
InDel11196	Pp04	11,196,845	A	H	H	H	H	H	H	H	H	A	H	H	A	A	-
InDel11206	Pp04	11,206,105	A	H	H	H	H	H	H	H	H	H	H	H	A	A	-
InDel11210	Pp04	11,210,233	A	H	H	H	H	H	H	H	H	H	A	H	A	A	-
InDel11254	Pp04	11,253,858	A	H	H	H	H	H	H	H	H	H	A	A	A	A	-
Indel11303	Pp04	11,303,453	H	H	H	H	H	H	H	H	H	H	A	A	H	H	-
SSR15662	Pp04	11,228,933	H	H	H	H	H	H	H	H	H	H	A	A	H	H	A
EPPCU2000	Pp04	12,478,769	H	H	H	H	H	H	H	H	H	H	A	A	H	H	A

5.3.2.1 Identification of candidate genes

The 183 kb genomic region containing the *Alf* gene has 31 annotated genes (Table 5.7) according to the *P. persica* Genome Annotation v2.1 retrieved from the Genome Database for Rosaceae (https://www.rosaceae.org/). Nineteen out of the 31 genes did not have any functional annotation and 12 did. The functions of the annotated genes were obtained from their UniProt gene code (https://www.uniprot.org/). Three were annotated as transcription factors and included, *Prupe.4G187100,* annotated as a NAC transcription factor necessary for normal seed development and morphology, *Prupe.4G188700*, annotated as protein ULTRAPETALA 2 that negatively regulates cell accumulation in shoot and floral meristems and, *Prupe.4G190000,* annotated as B3-domain containing transcription repressor VAL2 involved in seed maturation. Other three had a role in transport, including *Prupe.4G187500*, annotated as trafficking protein particle complex subunit 10 that transports vesicles from endoplasmic reticulum to Golgi complex, *Prupe.4G189700*, annotated as ADP-ribosylation factor GTPase-activating protein AGD5 that transport vacuolar cargo and promote plant growth and, *Prupe.4G189900*, annotated as cytochrome c-type biogenesis ccda-like chloroplastic protein 1 that transports NADH from stroma to thylakoid lumen. Two were involved in cell growth as *Prupe.4G188000,* annotated as receptor-like protein kinase THESEUS 1 required for cell elongation during vegetative growth and, *Prupe.4G188900,* annotated as Transmembrane protein 184 C that promotes cell growth by tumor suppression. *Prupe.4G187300* was annotated as peroxisomal membrane protein 2 that contributes to permeability of peroxisomal membrane. *Prupe.4G187400* was annotated as peptidyl-prolyl cis-trans isomerase CYP37 that accelerates protein folding. *Prupe.4G187800* was annotated as PAP-specific phosphatase HAL2-like that regulates sulfur flux. Finally, *Prupe.4G188600*, annotated as Lysine-specific histone demethylase 1 homolog that promotes flowering by FLOWERING LOCUS C (FLC) repression.

Table 5.7: List of candidate genes identified in the *Alf* genomic region.

Gene ID	Physical position v2.0	Description	Pathway
Prupe.4G187000	Pp04: 11134679-11135942	n/a	Unknown
Prupe.4G187100	Pp04: 11138518-11140641	NAC transcription factor 25 (*Arabidopsis thaliana*)	Transcription factor
Prupe.4G187200	Pp04: 11145857-11146604	n/a	Unknown
Prupe.4G187300	Pp04: 11147111-11151935	Peroxisomal membrane protein 2 (*Bos taurus*)	Peroxisomal membrane pore-forming activity
Prupe.4G187400	Pp04: 11153240-11151935	Peptidyl-prolyl cis-trans isomerase CYP37 (*Arabidopsis thaliana*)	Protein folding
Prupe.4G187500	Pp04: 11161053-11175010	Trafficking protein particle complex subunit 10 (*Dictyostelium discoideum*)	Transport
Prupe.4G187600	Pp04: 11161413-11161773	n/a	Unknown
Prupe.4G187700	Pp04: 11178009-11180316	Methyl transferase	Unknown
Prupe.4G187800	Pp04: 11184964-11190022	PAP-specific phosphatase HAL2-like (*Arabidopsis thaliana*)	Sulfur metabolism
Prupe.4G187900	Pp04: 11192916-11193704	n/a	Unknown
Prupe.4G188000	Pp04: 11206310-11210427	Receptor-like protein kinase THESEUS 1 (*Arabidopsis thaliana*)	Cell growth
Prupe.4G188100	Pp04: 11211569-11213268	Late embryogenesis abundant (LEA) protein related	Unknown
Prupe.4G188200	Pp04: 11215309-11217498	n/a	Unknown
Prupe.4G188300	Pp04: 11218458-11219376	n/a	Unknown
Prupe.4G188400	Pp04: 11220112-11220971	n/a	Unknown
Prupe.4G188500	Pp04: 11220929-11222217	Late embryogenesis abundant (LEA) protein related	Unknown
Prupe.4G188600	Pp04: 11223366-11227007	Lysine-specific histone demethylase 1 homolog 2 (*Arabidopsis thaliana*)	Flowering
Prupe.4G188700	Pp04: 11227089-11228967	Protein ULTRAPETALA 2 (*Arabidopsis thaliana*)	Transcription factor
Prupe.4G188800	Pp04: 11232893-11238310	Uncharacterized protein YnbB (*Bacillus subtilis*)	Unknown

Table 5.7 (Continued)

Gene ID	Physical position v2.0	Description	Pathway
Prupe.4G188900	Pp04: 11238694-11241828	Transmembrane protein 184C (*Pongo abelii*)	Cell growth
Prupe.4G189000	Pp04: 11242310-11245089	n/a	Unknown
Prupe.4G189100	Pp04: 11244108-11247472	n/a	Unknown
Prupe.4G189200	Pp04: 11247548-11249911	n/a	Unknown
Prupe.4G189300	Pp04: 11250220-11252990	Monodehydroascorbate reductase, seedling isozyme (*Cucumis sativus*)	Unknown
Prupe.4G189400	Pp04: 11254886-11256301	n/a	Unknown
Prupe.4G189500	Pp04: 11271534-11272507	n/a	Unknown
Prupe.4G189600	Pp04: 11274431-11275614	n/a	Unknown
Prupe.4G189700	Pp04: 11278755-11285726	Probable ADP-ribosylation factor GTPase-activating protein AGD5 (*Arabidopsis thaliana*)	Transport
Prupe.4G189800	Pp04: 11289619-11292044	n/a	Unknown
Prupe.4G189900	Pp04: 11294659-11299414	Cytochrome c-type biogenesis ccda-like chloroplastic protein 1 (*Oryza sativa subsp. Japonica*)	Transport
Prupe.4G190000	Pp04: 11300877-11316907	B3 domain-containing transcription repressor VAL2 (*Arabidopsis thaliana*)	Transcription factor

5.3.2.2 Variant calling and SnpEff

A SnpEff analysis was performed to predict the impact of the polymorphisms present between peach and almond in the *Alf* region (Pp04: 11119601-11303453). 2,160 variants detected between 'Texas' and 'Earlygold' resequences, all of them being SNPs. These variants produced 6,207 effects on the sequences (Table 5.8). 5,953 (96%) were non-coding variants, 143 (2.3%) were low-impact variants, 102 (1.7%) were moderate impact variants, and nine (0.15%) were high-impact variants. The predicted nine high-impact variants causing protein truncation included four candidate genes. A protein trafficking gene (*Prupe.4G187500*) with three high-impact variants, late embryogenesis abundant (LEA) protein (*Prupe.4G188500*), with four high-impact variants, and two other genes with unknown annotation (*Prupe.4G187600* and *Prupe.4G189600*), with one

high-impact variant each (Table 5.9). Other 23 candidate genes were predicted to have variants with moderate impact.

Table 5.8: Effect of variants detected and their impact predicted by SnpEff in *Alf* region

Effect	Number	Percentage (%)	Impact
Frameshift_variant	6	0.10%	High
splice_donor_variant	1	0.016%	High
stop_gained	2	0.032%	High
missense_variant	102	1.645%	Moderate
synonymous_variant	109	1.758%	Low
splice_region_variant	21	0.339%	Low
5_prime_UTR_premature_start_codon_gain_variant	13	0.21%	Low
5_prime_UTR_variant	102	1.65%	Modifier
3_prime_UTR_variant	173	2.79%	Modifier
upstream_gene_variant	1,983	31.98%	Modifier
intron_variant	608	9.81%	Modifier
intergenic_region	1,082	17.45%	Modifier
downstream_gene_variant	2,005	32.33%	Modifier
Total	6,207		

Table 5.9: High impact variants detected for the candidate genes of *Alf* region and their nucleotide changes

Variant position	Gene ID	Effect	Nº variants	Peach genome	Almond genome
Pp04: 11,161,725	Prupe.4G187600	frameshift variant	1	GTTT	GTT
Pp04: 11,171,150	Prupe.4G187500	stop gained	1	C	A
Pp04: 11,171,549	Prupe.4G187500	frameshift variant	1	CATATATAT	CATATATATAT
Pp04: 11,171,554	Prupe.4G187500	frameshift variant	1	AT	ATTT
Pp04: 11,220,938	Prupe.4G188500	stop gained	1	G	T
Pp04: 11,220,964	Prupe.4G188500	frameshift variant	1	AGG	AGGG
Pp04: 11,221,333	Prupe.4G188500	frameshift variant	1	CTTT	CTTTT
Pp04: 11,222,069	Prupe.4G188500	splice donor variant	1	C	G
Pp04: 11,275,588	Prupe.4G189600	frameshift variant	1	CCGCG	CCGCGCG

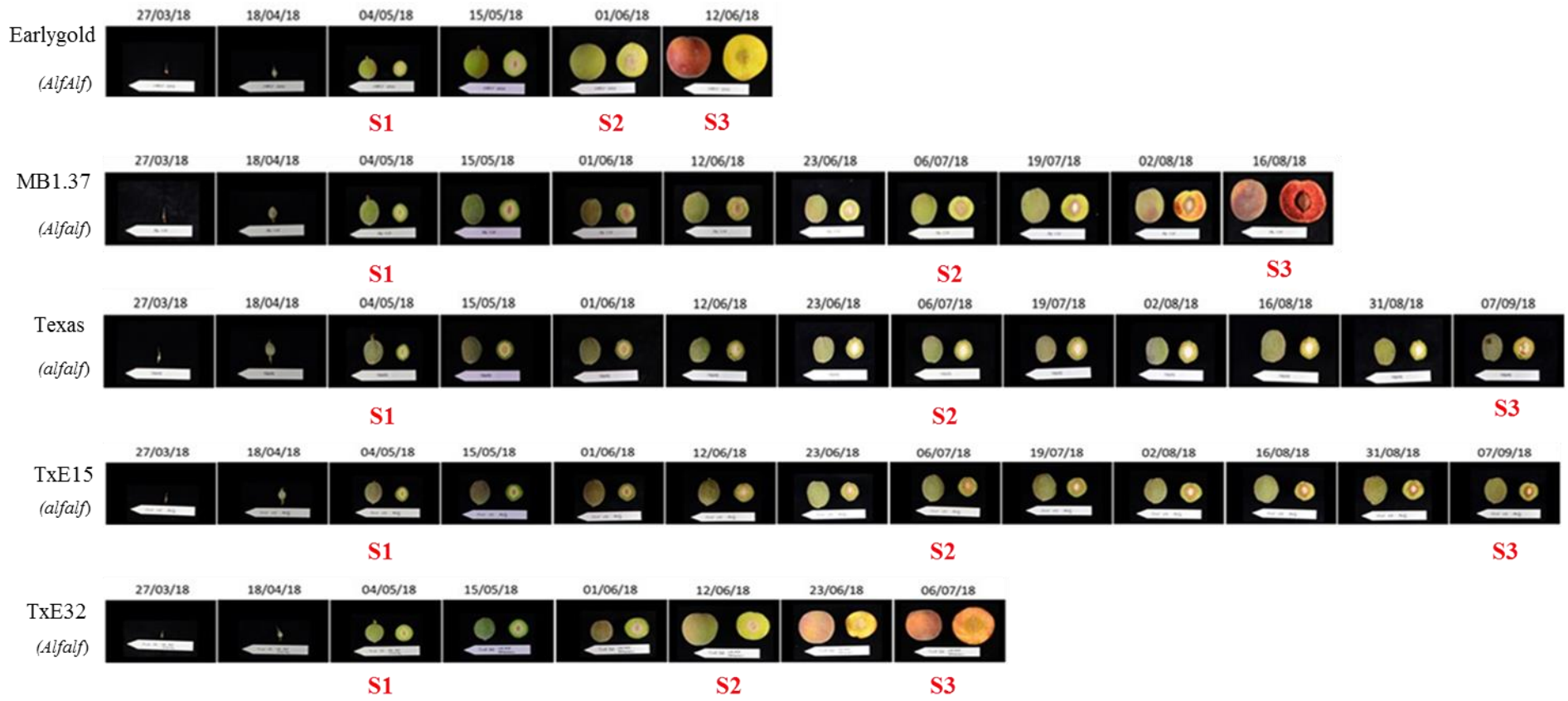


Figure 5.6: Fruit developmental stages of 'Earlygold', 'Texas', MB1.37 and F2 individuals (T×E15 and T×E32). The allelic configuration of *Alf* for each individual is indicated in the parenthesis. The images were taken from 30 DAF for every two weeks until fruit maturity. S1, S2 and S3 were the time points selected for the expression analysis.

5.3.2.3 *Alf* Gene expression analysis

To determine the expression profiles of the 31 positional candidate genes (Table 5.7), we performed an RNA-seq analysis using fruit mesocarp tissue at different developmental stages selected based on our previous observations (see section 5.3.1.1). The three developmental stages included the initial stage (S1), where peach and almond fruits are still very similar in size, an intermediate stage of development (S2) depending on the maturity date of each individual, and the ripening stage (S3). Five individuals were selected for the study including three with peach type fruits ('Earlygold', MB1.37 and T×E32) and two with almond type fruits ('Texas' and T×E15). The sample time points analyzed are presented in Figure 5.6.

Table 5.10: Differentially expressed candidate genes of *Alf* at all three stages of fruit development (S1, S2 and S3).

Stage	Candidate gene	P-value	Q-value
S1	Prupe.4G189700	$4.58e^{-162}$	$1.06e^{-158}$
	Prupe.4G189400	$4.10e^{-110}$	$5.3e^{-107}$
	Prupe.4G189100	$6.37e^{-41}$	$2.33e^{-38}$
	Prupe.4G187300	$1.20e^{-30}$	$3.42e^{-28}$
	Prupe.4G187100	$7.71e^{-30}$	$2.06e^{-27}$
S2	Prupe.4G189700	$1.02e^{-145}$	$1.02e^{-141}$
	Prupe.4G187700	$3.03e^{-33}$	$1.36e^{-30}$
	Prupe.4G187100	$2.01e^{-32}$	$8.76e^{-30}$
	Prupe.4G189100	$3.12e^{-27}$	$8.14e^{-25}$
	Prupe.4G189400	$1.23e^{-23}$	$32.36e^{-21}$
S3	Prupe.4G189700	$2.92e^{-186}$	$4.44e^{-182}$
	Prupe.4G189100	$9.46e^{-39}$	$4.04e^{-36}$
	Prupe.4G188000	$3.10e^{-29}$	$8.49e^{-27}$

Although we had whole genome RNAseq data, we focused our attention on the 31 positional candidate genes. Their expression was compared during the three stages of fruit development between individuals with peach and almond fruit types. From the threshold values set at each stage, we identified five genes (*Prupe.4G189700*, *Prupe.4G189400*, *Prupe.4G189100*, *Prupe.4G187300* and *Prupe.4G187100*) that were differentially expressed in S1, five genes in S2 (*Prupe.4G189700*, *Prupe.4G187700*, *Prupe.4G187100*, *Prupe.4G189100* and *Prupe.4G189400*) and three genes in S3 (*Prupe.4G189700*, *Prupe.4G189100* and *Prupe.4G188000*) (Table 5.10, Figure S5.1). Two

genes, *Prupe.4G189100* and *Prupe.4G189700*, were differentially expressed between peaches and almonds in all three stages of development.

5.3.3 Fine mapping of *DBF2*

The *DBF2* gene was initially mapped between two SNPs, upstream SNP_IGA_121740 and downstream SNP_IGA_123023 in a genomic region of 1.05 Mb (Pp01: 41709139-42754294) (Donoso et al. 2016). To fine map this region, we designed new molecular markers from the parents resequence data, that included seven InDels, four SNPs and two SSRs (Table 5.2). This set of markers was genotyped in the two T1E individuals (T1E62 and T1E464) that defined the position of the gene, and the phenotype data was already available. This information helped us to narrow down the *DBF2* region to 978 kb (Pp01: 41709139-42687251), delimited by the markers SNP_IGA_121740 and SNP42687. To further reduce this region, we searched for additional recombinant individuals. For that, we screened a segregating BC2 population (51) developed during the NIL collection development. In addition, we also genotyped 3,835 plants between 2015 and 2019, obtained from open pollinations of MB1.37 (1,069) and of some T1E individuals (2,766) that were heterozygous for the target region. The screening was initially done with the two SSRs flanking the *DBF2* gene, CPPCT029 and BPPCT028. SSRs were chosen for initial screening, because genotyping can be performed in a high throughput manner using the ABI Genetic Analyzer. As *DBF2* almond allele is dominant, the only useful recombinant individuals were those where one marker was homozygous for the peach allele and the other marker was heterozygous for peach and almond. The screenings performed using the two SSRs resulted in 250 recombinants and 696 individuals with missing data in at least one of the markers (Table 5.11). The selected 250 recombinants genotyped with SNP_IGA_121740 and SNP42687 markers, resulted in 30 recombinant individuals. Two of those individuals have already produced fruits and phenotypes were available. One of them, 21P15-27, had the closest recombination to the *DBF2*. This data reduced our genomic region from 978 kb to 10 kb (Pp01: 42677206-42687251) (Table 5.12). The recombinant individual 103P19-211 also had a recombination in the same region but phenotype information is not available yet.

Table 5.11: Individuals genotyped for fine mapping of the *DBF2* gene.

Year	Population	Genotype code	Female parent	Individuals genotyped	Recombinants
2015	BC2	21P15	E2T-092-11	51	14
2015	F2	T×E	MB1.37	185	16
2017	F2	51P17	MB1.37	884	54
2018	BC1S1	72P18	T1E-21	33	9
2018	BC1S1	74P18	T1E-24	12	1
2018	BC1S1	79P18	T1E-34	32	3
2018	BC1S1	80P18	T1E-35	18	1
2018	BC1S1	84P18	T1E-40	25	1
2018	BC1S1	85P18	T1E-43	2	-
2018	BC1S1	93P18	T1E-64	259	17
2018	BC1S1	99P18	T1E-79	22	1
2018	BC1S1	103P18	T1E-96	5	1
2018	BC1S1	107P18	T1E-104	19	3
2018	BC1S1	108P18	T1E-108	2	-
2018	BC1S1	111P18	T1E-123	12	-
2018	BC1S1	118P18	T1E-197	4	2
2018	BC1S1	120P18	T1E241	2	1
2018	BC1S1	123P18	T1E-219	5	2
2018	BC1S1	134P18	T1E-333	5	1
2018	BC1S1	135P18	T1E-335	67	2
2018	BC1S1	150P18	T1E-463	21	7
2018	BC1S1	151P18	T1E-467	12	1
2018	BC1S1	155P18	T1E-488	2	-
2018	BC1S1	156P18	T1E-500	24	2
2019	BC1S1	96P19	T1E-21	39	4
2019	BC1S1	97P19	T1E-22	9	-
2019	BC1S1	102P19	T1E-32	28	4
2019	BC1S1	103P19	T1E-34	241	14
2019	BC1S1	104P19	T1E-35	123	10
2019	BC1S1	108P19	T1E-40	61	4
2019	BC1S1	109P19	T1E-43	15	2
2019	BC1S1	117P19	T1E-64	234	8
2019	BC1S1	118P19	T1E-65	14	1
2019	BC1S1	130P19	T1E-101	82	7

Table 5.7 (Continued)

Year	Population	Genotype code	Female parent	Individuals genotyped	Recombinants
2019	BC1S1	131P19	T1E-104	180	6
2019	BC1S1	133P19	T1E-116	32	-
2019	BC1S1	135P19	T1E-123	98	8
2019	BC1S1	142P19	T1E-197	76	1
2019	BC1S1	145P19	T1E-201	127	6
2019	BC1S1	147P19	T1E-219	21	1
2019	BC1S1	148P19	T1E-220	113	7
2019	BC1S1	149P19	T1E-226	77	4
2019	BC1S1	162P19	T1E-344	45	5
2019	BC1S1	165P19	T1E-389	226	8
2019	BC1S1	167P19	T1E-410	63	4
2019	BC1S1	170P19	T1E-424	50	-
2019	BC1S1	175P19	T1E-463	89	3
2019	BC1S1	176P19	T1E-467	43	2
2019	BC1S1	181P19	T1E-500	97	2
Total				3886	250

Table 5.12: Genotypes and phenotypes of the three *DBF2* recombinant individuals. The two red lines represent the recombination break point. *DBF2* genomic region of 10 kb is defined by the markers SNP42677 and SNP42687. Phenotype data from T1E62 and 21P15-27 was used as a molecular marker to indicate the position of the recombinants. B (yellow color) allele from the the parent 'Earlygold' and H (red color) allele from MB1.37, unknown data (-).

Marker name	Group	Position in bp (v2.0)	T1E62	21P15-27	103P19-211
CPPCT029	Pp01	41,168,265	H	B	B
InDel42115	Pp01	42,115,821	H	B	B
InDel42311	Pp01	42,311,741	H	B	B
InDel42471	Pp01	42,471,315	H	B	B
InDel42525	Pp01	42,525,134	H	B	B
InDel42584	Pp01	42,584,009	H	B	B
SSR6105	Pp01	42,638,623	H	B	B
InDel42666	Pp01	42,666,650	H	B	B
SNP42675	Pp01	42,675,790	H	B	B
SNP42677	Pp01	42,677,206	H	B	B
DBF2	**Pp01**	-	**H (Red color)**	**H (Red color)**	-
SNP42687	Pp01	42,687,251	B	H	H
SNP42696	Pp01	42,696,025	B	H	H
InDel42718	Pp01	42,718,200	B	H	H
SSR6125	Pp01	42,723,336	B	H	H
BPPCT028	Pp01	44,130,041	B	H	H

5.3.3.1 Identification of candidate genes

Table 5.13: List of candidate genes identified in the *DBF2* genomic region

Gene ID	Position v2.0	Description	Pathway
Prupe.1G519800	Pp01: 42674777-42677556	UDP-glycosyltransferase 85A2 (*Arabidopsis thaliana*)	Anthocyanin biosynthesis
Prupe.1G519900	Pp01: 42678375-42680309	UDP-glycosyltransferase 85A2 (*Arabidopsis thaliana*)	Anthocyanin biosynthesis
Prupe.1G520000	Pp01: 42680643-42684196	UDP-glycosyltransferase 85A2 (*Arabidopsis thaliana*)	Anthocyanin biosynthesis
Prupe.1G520100	Pp01: 42684510-42686498	UDP-glycosyltransferase 85A2 (*Arabidopsis thaliana*)	Anthocyanin biosynthesis

The 10 kb *DBF2* region of the peach genome contains four annotated candidate genes (Table 5.13) according to the *P. persica* Genome Annotation v2.1 (https://www.rosaceae.org/). The four genes were annotated as UDP-glycosyltransferase (UGT) 85A2. We performed a BLAST analysis to check if we could obtain more information about our four positional candidate genes. For the four candidate genes the highest homology was found for a 7-deoxyloganetin glucosyltransferase gene.

5.3.3.2 Variant calling and SnpEff

To predict the DNA variant that could be responsible of the blood flesh phenotype, a SnpEff analysis was performed using the resequence data from 'Texas' and 'Earlygold' for the *DBF2* region (Pp01: 42677206-42687251). We identified 183 variants including 161 SNPs, 10 insertions and 12 deletions. These variants predicted 649 effects including 624 non-coding variants, 12 (1.85%) were moderate impact variants and 13 (2%) were low-impact variants (Table 5.14). No high impact variants were detected for the candidate genes.

Table 5.14: Effect of variants and their impact predicted by SnpEff in *DBF2* region.

Effect	Count	Percentage	Impact
missense_variant	12	1.85%	Moderate
synonymous_variant	8	1.23%	Low
splice_region_variant	2	0.31%	Low
5_prime_UTR_premature_start_codon_gain_variant	3	0.46%	Low
5_prime_UTR_variant	17	2.62%	Modifier
3_prime_UTR_variant	37	5.69%	Modifier
upstream_gene_variant	271	41.85%	Modifier
intron_variant	50	7.69%	Modifier
intergenic_region	59	9.08%	Modifier
downstream_gene_variant	190	29.23%	Modifier
Total	649		

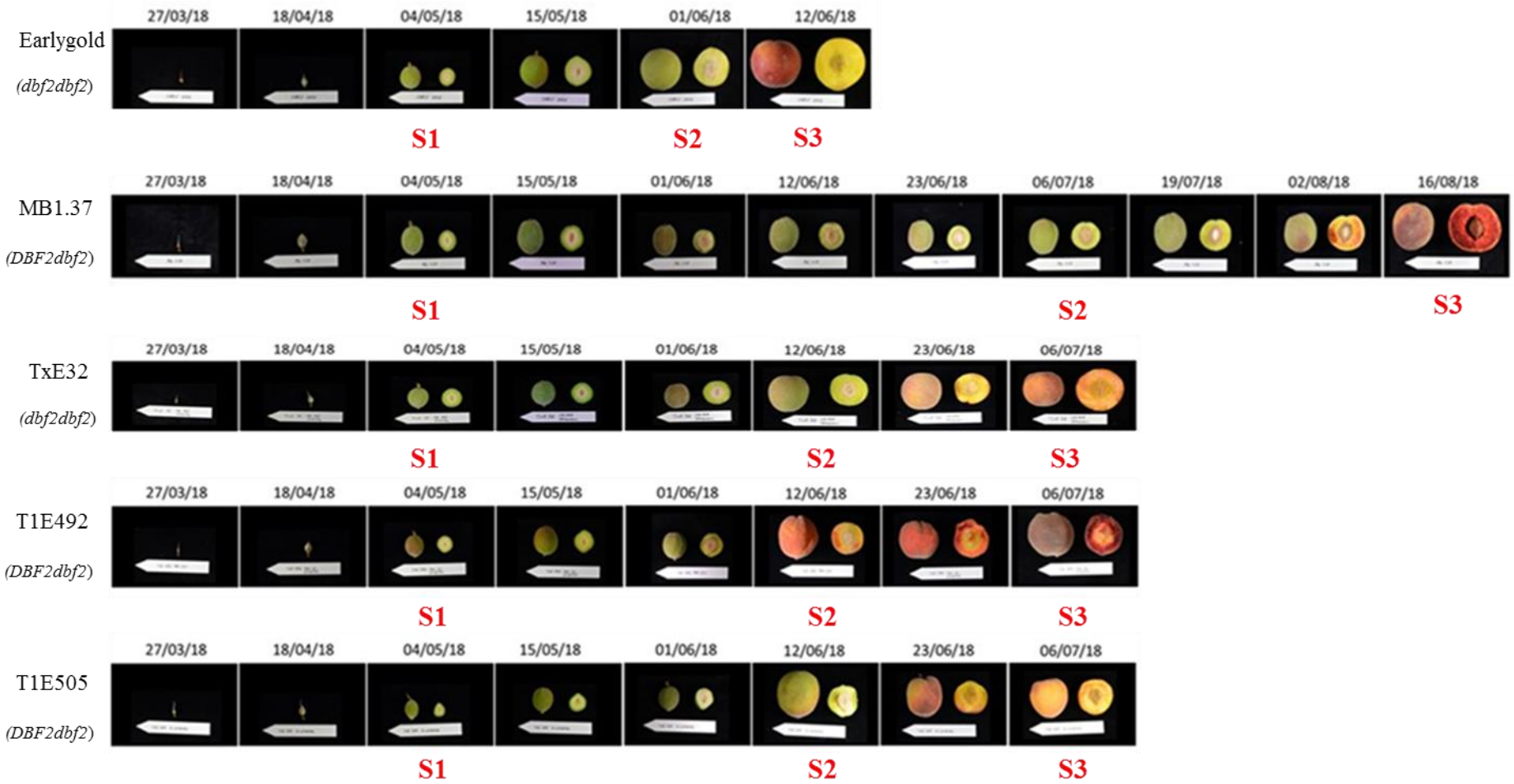


Figure 5.7: Fruit developmental stages of, parent ('Earlygold'), F1 hybrid (MB1.37), F2 individual (T×E32) and BC1 individuals (T1E492 and T1E505). The allelic configuration of *DBF2* for each individual is indicated in the parenthesis. The images were taken from 30 DAF for every two weeks until fruit maturity. S1, S2 and S3 represent the time points selected for expression analysis.

5.3.3.3 *DBF2* Gene expression analysis

RNA-seq analysis was performed to identify the expression levels of the four positional *DBF2* candidate genes (Table 5.13). For this, we used fruit mesocarp tissue from different fruit developmental stages, starting from 60 DAF for every two weeks till the fruit maturity. Five individuals were selected for this study including three yellow flesh individuals ('Earlygold', T×E32 and T1E505) and two red flesh individuals (MB1.37 and T1E492). Three developmental stages were selected for analysis, the early stage of fruit development (S1) where all the individuals had similar flesh color, an intermediate stage (S2) individuals at different maturity with slight changes in mesocarp color, the maturity stage (S3) with a complete yellow or red flesh color (Figure 5.7). Only *Prupe.1G519800* was differentially expressed between red and yellow flesh individuals in S1 stage (Table 5.15, Figure S5.2).

Table 5.15: *DBF2* candidate genes differential expression at all the three stages, S1, S2 and S3. *Prupe*.1G519800 differentially expressed at S1 stage (Q-value < 0.005).

Stage	Candidate gene	P-value	Q-value
S1	Prupe.1G519800	7.32e-06	0.0006
	Prupe.1G520100	-	-
	Prupe.1G519900	0.85	0.95
	Prupe.1G520000	0.88	0.96
S2	Prupe.1G519800	0.65	0.95
	Prupe.1G519900	0.019	0.32
	Prupe.1G520000	0.50	0.92
	Prupe.1G520100	-	-
S3	Prupe.1G519800	0.004	0.052
	Prupe.1G519900	0.033	0.181
	Prupe.1G520000	-	-
	Prupe.1G520100	-	-

5.3.4 Fine mapping of *Jui*

According to Donoso et al. (2016), the *Jui* gene was mapped between two SNPs, SNP_IGA_107095 and SNP_IGA_107417, in a genomic region of 200 kb with a physical position between Pp01: 35198093-35398680 defined by the two recombinant individuals T1E8 (upstream) and T1E724 (downstream). To saturate this region, we designed 27 new markers that included one SSR, 13 InDels and 13 SNPs (Table 5.3), and genotyped the initial recombinant individuals T1E8 and T1E724. This narrowed down the target genomic region to 66 kb (Pp01: 35198093-35264164). To further reduce the size of this region we looked for new recombinants. For that, we genotyped, between 2017 and 2019, 3,616 plants obtained from open pollinations of MB1.37 (884) and of *Jui* heterozygous T1E individuals (2,722). We used the two SSRs flanking the *Jui* gene, EPDCU3489 and BPPCT016 and identified 333 new recombinants, 2,757 non-recombinants and 526 individuals with missing data in at least one of the markers. The 333 recombinants were genotyped with the two new markers surrounding the gene, from which four recombinants were detected. During the last year of this PhD, we realized that the recombinant T1E724 phenotype was scored incorrectly (juicy instead of non-juicy). That relocated the initial position of *Jui* to a new region defined by the individuals T1E8 and T1E694 in a 980 kb region delimited by the markers SNP_IGA_107095 and SNP_IGA_109223 with a physical position Pp01: 35198093-36177446. The genotype data of the T1E724 individual obtained with the new markers developed indicated that it had a recombination closer to *Jui* than the T1E8 individual (Table 5.16). This narrowed down the *Jui* genomic region from 980 kb to 392 kb (Pp01: 35252325-35644865). Almost all the individuals selected from the screenings performed between 2017 and 2019 had to be discarded, because the selection was performed in the region proposed by Donoso et al. (2016), and most of this region was located outside of where *Jui* is currently situated. There was one individual (104P19-93) with a recombination in this interval but still has not produced any fruit (Table 5.16). The *Jui* final genomic region (392 kb) contains 95 annotated genes according to the *P. persica* genome Annotation v2.1 (https://www.rosaceae.org/). As we consider that this is still a very long list it is not presented here, neither the SnpEff nor the expression analysis were performed.

Table 5.16: Genotypes and phenotypes of the four *Jui* recombinant individuals. Recombinants T1E694 and T1E724 define the *Jui* genomic region of 392 kb, delimited by two markers SNP35252 and SNP35644. The two red lines indicate the recombination breaking point. Phenotype data was used as a molecular marker to indicate the position of the recombinants. B (juicy) allele from the parent 'Earlygold' and H (non-juicy) allele from the F1 hybrid MB1.37, unknown data (-).

Marker name	Group	Position in bp (v2.0)	T1E8	T1E694	T1E724	104P19-93
EPPCU3489	Pp01	34,001,221	B	B	B	B
InDel34892	Pp01	34,892,317	B	B	B	B
InDel34907	Pp01	34,907,350	B	B	B	B
InDel34933	Pp01	34,933,506	B	B	B	B
InDel34972	Pp01	34,972,191	B	B	B	B
InDel35000	Pp01	35,000,890	B	B	B	B
InDel35036	Pp01	35,036,070	B	B	B	B
InDel35085	Pp01	35,085,392	B	B	B	B
InDel35147	Pp01	35,147,490	B	B	B	B
InDel35173	Pp01	35,173,747	B	B	B	B
SNP35183	Pp01	35,183,714	B	B	B	B
SNP35194	Pp01	35,194,967	B	B	B	B
SNP_IGA_107095	Pp01	35,198,093	B	B	B	B
SNP35200	Pp01	35,200,523	H	B	B	B
SNP35204	Pp01	35,204,671	H	B	B	B
SNP35211	Pp01	35,211,229	H	B	B	B
SNP35224	Pp01	35,224,589	H	B	B	B
SNP35244	Pp01	35,244,343	H	B	B	B
InDel35245	Pp01	35,245,432	H	B	B	B
SNP35247	Pp01	35,247,471	H	B	B	B
SSR4996	Pp01	35,251,878	H	B	B	B
SNP35252	Pp01	35,252,325	H	B	B	B
SNP35264	Pp01	35,264,164	H	B	H	H
SNP35269	Pp01	35,269,192	H	B	H	H
InDel35271	Pp01	35,271,415	H	B	H	H
Jui	**Pp01**	**-**	**H (Non-juicy)**	**B (Juicy)**	**H (Non-juicy)**	-
InDel35336	Pp01	35,336,891	H	B	H	H
InDel35351	Pp01	35,351,260	H	B	H	H
SNP_IGA_107417	Pp01	35,398,680	H	B	H	H
SNP35545	Pp01	35,545,389	H	B	H	H
SNP35644	Pp01	35,644,865	H	H	H	H
SNP_IGA_109223	Pp01	36,177,446	H	H	H	H
BPPCT016	Pp01	37,047,997	H	H	H	H

5.4 Discussion

5.4.1 Differences between peach and almond fruit development

Peaches and almonds are closely related species with very similar and syntenic genomes (Dirlewanger et al. 2004) that are also inter-compatible producing fertile offspring. They are consumed by their fruit, that are drupes in both cases, although the edible part is the mesocarp in peaches and the seed in almonds. The fact that the genetics behind part of their main fruit differences is controlled by monogenic traits, such as the *Alf* and *Jui* genes studied in this chapter, confirms their high level of resemblance. From other aspects they are clearly different and molecular analysis has shown that they are different species that evolved about 5 Mya from a common ancestor in Central Asia (Delplancke et al. 2016; Yu et al. 2018; Alioto et al. 2020).

In this study, we tried to fine-map and identify the molecular basis of the genes responsible for the fruit type (*Alf*) and fruit juiciness (*Jui*) traits previously analyzed in almond × peach segregating progenies (Donoso et al. 2016). In addition, we also included a gene coming from almond and responsible of providing red flesh color to the fruits (*DBF2*), that could have an impact in peach breeding. For that, we started observing the main differences between peach and almond fruit development including individuals that differed in those characteristics: peach and almond fruit types, juicy and non-juicy, and yellow and red flesh types. We could observe that almond and peach fruits followed a different fruit growth pattern that affected the mesocarp but not the endocarp, that presented a very similar pattern of development in peaches and almonds. Although peach and almond type fruits follow a similar pattern of growth during the first eight weeks, from this point until the harvest date, their development patterns diverge. While almonds stop their growth and do not ripe, peaches increase their size (faster or slower depending on their early, medium or late ripening time) and suffer the typical changes associated to the fruit ripening process, including changes in flesh and skin color, increase in sugar content, and changes in texture, among others. Previous studies reported that peach, like other stone fruits, follows a double sigmoid growth pattern with four distinct stages of fruit development including two alternative growth phases (Lombardo et al. 2011). Here we could observe that this was the case for medium or late ripening individuals but not for the early ones, that basically presented a continuous exponential growth from pollination until fruit ripening. It also was possible to identify when the differences between the traits studied started to be evident. For *Alf* we observed that the divergence between almond and peach types could be seen from the eighth week on. For *DBF2* and *Jui*, the differences between juicy and non-juicy fruits and between yellow and red fleshed ones started to be evident only in the last two weeks of development, immediately before fruit ripening and could only be appreciated in peach like fruits.

5.4.2 Fine mapping strategy

The fine mapping strategy that we used was efficient, but it could be further improved. One of the first aspects that could be improved and that we already modified during this PhD is the origin of the individuals to screen. In the first year we used selfed individuals from the F1 interspecific hybrid. Although it looks an adequate strategy because all the individuals will be segregating for all the target genes, we realized that most individuals were weak or infertile and the probability of obtaining a fruiting plant that could be phenotyped from the selected recombinant individuals was very low. To avoid these problems, we used individuals obtained from the open pollinations of T1E trees that were heterozygous for one or more of the target traits. When possible we chose late ripening T1E individuals, as they usually have a better germination rate than early or intermediate ripening plants.

Another important aspect in a fine mapping project is the type of markers used for the selection. In this work we started using SSRs because they can be partially automatized and multiplexed using different fluorophores and because we have all the equipment in the lab and we do not need to resort to external services. This last factor, although it is not important in normal conditions, it was under movement restrictions due to the Covid19 pandemic. Nowadays individual SNP markers can be genotyped at very competitive prices and in 384 well plates, therefore SNP markers could be a very good alternative to SSRs. The other type of markers used were InDels, which are cheap and useful for genotyping alternative, but only for small sample sizes.

5.4.3 *Alf* fine mapping and candidate gene analysis

Before this work, *Alf* was mapped in a 1.6 Mb genomic region in G4 where there are 274 annotated genes in the peach reference genome v2.0 (Donoso et al. 2016). The recombinant screening performed for 3,828 individuals between 2015-19 resulted in the identification of 403 recombinants. The two initial recombinants, T×E59 and T×E85, were genotyped with new molecular markers developed and we narrowed down the region to 183 kb (Pp04: 11119601-11303453) containing 31 annotated genes. This region could be further reduced when the other 13 identified recombinants will produce fruits and phenotyping data will be available. To find which among the 31 positional candidates were responsible for *Alf*, a gene expression analysis and prediction of the effects of the genetic variants using the SnpEff software were performed. Among the variants identified in the genomic region, only those with high impact on the protein were initially considered. There were four genes predicted with high impact variants and seven genes that were differentially expressed in some of the developmental stages between almond and peach type fruits. None of the candidate

genes was included in both lists. We did not find in the literature any relationship between the genes with high impact variants and fruit size or ripening. One of our differentially expressed candidate genes, *Prupe.4G187100*, was annotated as a NAC transcription factor 25. *Prupe.4G187100* was more expressed in peach-like fruits than in almond-like fruits during the three stages of development, although the difference was not significant in S3. NAC transcription factors are one of the largest plant TF families involved in plant development and stress responses (Olsen et al. 2005; Jensen et al. 2010). A phylogenetic analysis of NAC transcription factors of different species clustered *Prupe.4G187100* gene with the non-ripening (NOR) tomato gene (Pirona et al. 2013). The cause of the NOR mutation is a deletion of two adenines (Giovannoni JJ 2004) that produced a truncated NOR protein of 186 amino acids which disrupts the transcriptional activation region but preserves the complete DNA-binding region. This fact suggested that the *NAC-NOR* gene was a loss of mutant and that the gene was involved in the first steps of the regulation of tomato fruit ripening (Osorio et al. 2011). Other recent works that use edited versions of the *NAC-NOR* gene suggest that the *NAC-NOR* gene mutation is a gain of function because the *NAC-NOR* truncated version is limiting the binding of other transcription factors to the promoters of several genes involved in fruit ripening (Guo et al. 2021). The *NAC-NOR* gene in normal tomato has a very low expression in the first steps of fruit development and then has peak of expression in the breaker stage, when pink or red color starts to be visible. In the *nor* mutant, the expression remains very low even at the ripening stage (Wang et al. 2020). This different pattern of expression is like what we could observe between almonds and peaches, where almond presented a very low expression of *Prupe.4G187100* during all the stages analyzed. Although this could explain the differences between peaches and almonds in relation to their ripening capabilities, we did not find any explanation for the differences observed in mesocarp size. In peach it exists a natural mutant where this gene is deleted and that is called slow ripening (SR; Nuñez-Lillo et al. 2015). In this mutant, ripening is almost completely inhibited but mesocarp size is not very affected. This indicates that although *Prupe.4G187100* could be partially explaining the differences between peach and almond fruits, another factor is needed.

We also found an interesting gene, *Prupe.4G188700*, annotated as protein ULTRAPETALA 2 (ULT2). Although this gene has no major impact variants and is not differentially expressed in the developmental stages studied, it has been reported that ULTRAPETALA genes play a role in the control of the initial development of reproductive organs in *Arabidopsis thaliana* (Monfared et al. 2013). Therefore, if this gene was the causal factor, the differences between peach and almond fruit development could be established at earlier stages of development than the ones studied here.

The major gene *Alf* is crucial to understand the physiological differences between almond and peach fruit development. It could also explain the different evolution paths of almonds and peaches, and how almonds ended up with a dry leathery mesocarp and peaches developed a fleshy and edible mesocarp. Until we will have a shorter list of candidate genes or a strong candidate to start the functional validation of the causal gene, we thought how these genes could be used from a peach breeding point of view.

5.4.4 *DBF2* fine mapping and candidate gene analysis

Before the fine mapping, the dominant blood flesh gene *DBF2* was located in a 1.05 Mb genomic region on G1, with 277 annotated genes in the peach reference genome v2.0 (Donoso et al. 2016). The fine mapping process done in this work narrowed down this genomic region to 10 kb between positions Pp01: 42677206-42687251. This region contains four candidate genes annotated as a UDP-glycosyltransferase 85A2 (UGT85A2). To find evidence that supports some of these genes as responsible for red flesh color, we performed expression analysis and variant prediction in the coding region of candidate genes with the SnpEff software. From the variants analysis, there were no high impact variants detected for all the four candidate genes, while all of them had at least one variant with moderate impact. One of the four genes, the *Prupe.1G519800,* was differentially expressed at the S1 initial stage. For that reason, it was considered our strongest candidate gene.

UDP-glycosyl transferases (UGTs) are a large gene family involved in the glycosylation processes that play an important role in regulating secondary metabolite availability. Many of them have been linked to different anthocyanin-related traits in many species including peach (Cheng et al. 2014), Arabidopsis (Yonekura-Sakakibara and Hannada 2011), strawberry (Song et al. 2016), kiwifruit (Montefiori et al. 2011) or sweet potato (Wang et al. 2018). In peach, 168 UGTs have been identified and classified in 16 groups based on their sequence similarity (Wu et al. 2017). Our candidate genes were placed in group G, containing a total of 34 UGTs. To try to have a deeper understanding on the function of our candidate genes we performed a BLAST analysis with our stronger candidate *Prupe.1G519800*. The highest e value corresponded to a 7-deoxyloganetin glucosyltransferase gene. This gene was already reported as a candidate gene responsible for anthocyanin biosynthesis in apricot (García-Gómez et al. 2020).

Unlike in *Alf,* for *DBF2* we have a very short list of candidate genes, with *Prupe.1G519800* being the strongest candidate. For that reason, we could start its functional validation. Although peach transformation is a very inefficient process (Ricci et al. 2020), several studies used different protocols based on *Agrobacterium tumefaciens* for transient expression in fleshy fruits (Wu et al. 2019) or in leaves (Zhou et al. 2014; Tuan et al. 2015). In particular, Wu et al. (2019) overexpressed

a PpUGT85A2 responsible of linalool glycosylation introducing it in *A. tumefaciens* and infiltrating peach fruit cubes kept in artificial medium.

Although peach transformation could be an option to incorporate this trait in peach breeding programs, its efficiency is very low, and GMOs are not allowed in many countries. An alternative way to introgress *DBF2* gene into peach breeding programs would be through marker assisted introgression (MAI) (Serra et al. 2016). In the collection of ILs presented in chapter 4 we have two introgression lines, PAILE1-2348 and PAILE1-3448, with a single introgression from almond and containing the *DBF2* almond allele producing red flesh. These lines can be considered as prebreeding material that can be used in peach breeding programs to introgress the trait. In addition, the molecular markers described in this chapter could be used to select the plants with red flesh, increasing the efficiency of the process.

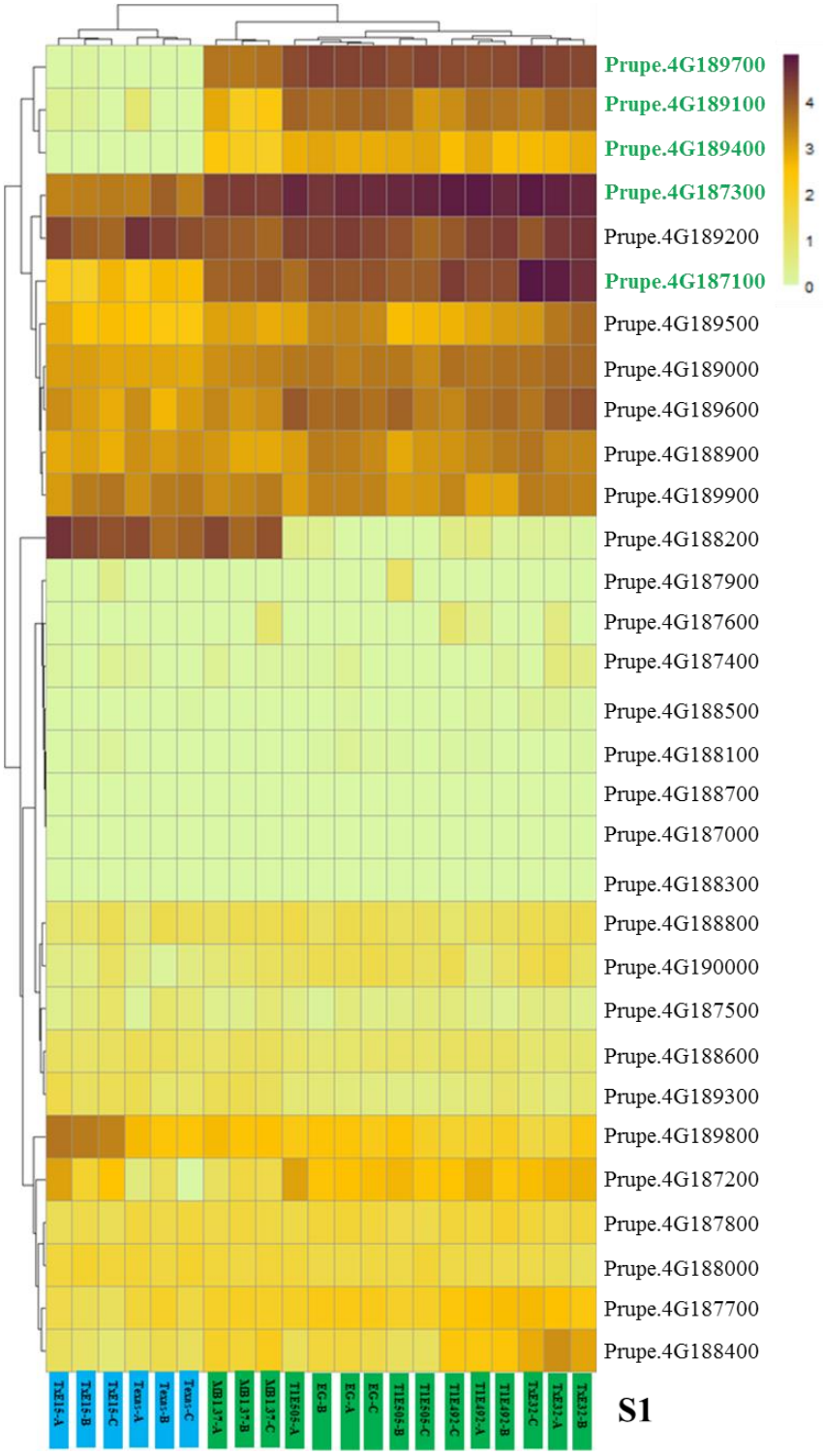
5.4.5 *Jui* genomic region saturation with new markers

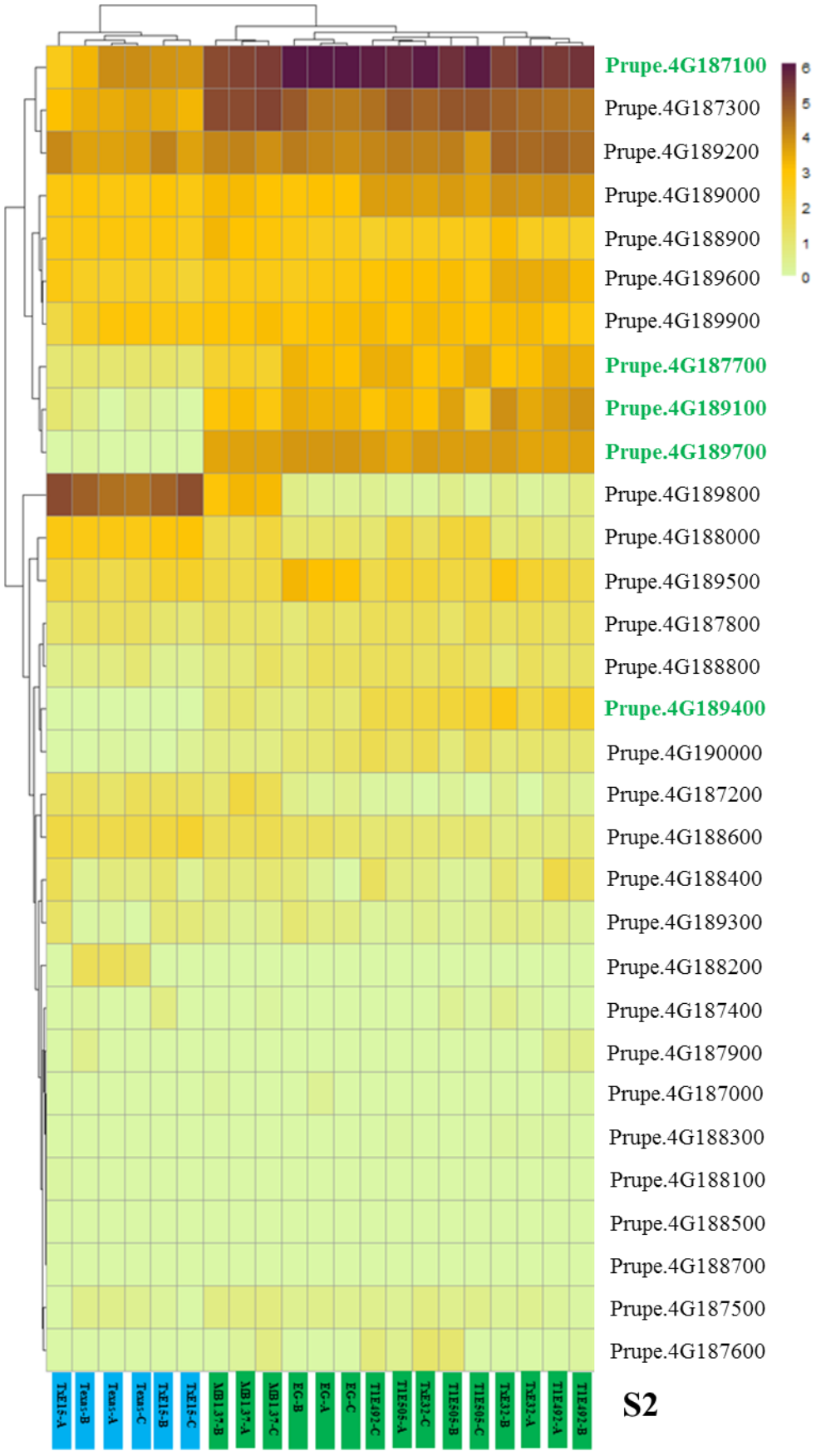
The *Jui* gene was initially located in a 200 kb (Pp01: 35198093-35398680) genomic region in G1, where 38 genes were annotated in the peach reference genome v2.0. In the last year of this PhD, we realized that the phenotype of one of the initial recombinant individuals from the T1E population determining *Jui* position, was scored incorrectly. The alternative recombinant individual increased the size of the genomic region where *Jui* was located from 200 kb to 980 kb (Pp01: 35198093-36177446). Saturating this region with molecular markers in the new recombinant individuals allowed us to increase the resolution of the mapping to a region of 392 kb (Pp01: 35252325-35644865), where 95 genes were annotated. Given that this is a very high number, we postponed the analysis of variants and gene expression until this list is substantially reduced. This region could be further narrowed down when the 104P19-93 recombinant will produce fruits and we could score the phenotype.

To avoid further errors in *Jui* phenotyping it would be convenient to develop a more reliable method of phenotyping. Several methods have been already described for assessing the juice content related to the mealiness symptoms observed during cold storage. One of them consisted of collecting pieces of mesocarp without the skin, homogenate them, and then centrifuge and weigh the supernatant and use it as a measure of the apparent juice content (Lill and van der Mespel 1988). Other approaches that have been tested include compressing a piece of flesh on top of a piece of absorbent paper and determining the area, a sensory panel test or measuring the weight difference between fresh fruits and after drying them in the oven (Baltazar et al. 2020).

5.5 Supplementary Material

Figure S5.1: *Alf* candidate genes differential expression at three fruit developmental stages, S1, S2 and S3. Almond type individuals indicated in blue color and peach type in green color.




Prupe.4G187100
Prupe.4G187300
Prupe.4G189200
Prupe.4G189000
Prupe.4G188900
Prupe.4G189600
Prupe.4G189900
Prupe.4G187700
Prupe.4G189100
Prupe.4G189700
Prupe.4G189800
Prupe.4G188000
Prupe.4G189500
Prupe.4G187800
Prupe.4G188800
Prupe.4G189400
Prupe.4G190000
Prupe.4G187200
Prupe.4G188600
Prupe.4G188400
Prupe.4G189300
Prupe.4G188200
Prupe.4G187400
Prupe.4G187900
Prupe.4G187000
Prupe.4G188300
Prupe.4G188100
Prupe.4G188500
Prupe.4G188700
Prupe.4G187500
Prupe.4G187600
6
5
4
3
2
1
0
TxE15-A
Texa-B
Texa-A
Texa-C
TxE15-B
TxE15-C
MB137-B
MB137-A
MB137-C
EG-B
EG-A
EG-C
TIE492-C
TIE505-A
TxE31-C
TIE505-B
TIE505-C
TxE31-B
TxE31-A
TIE492-A
TIE492-B

S2

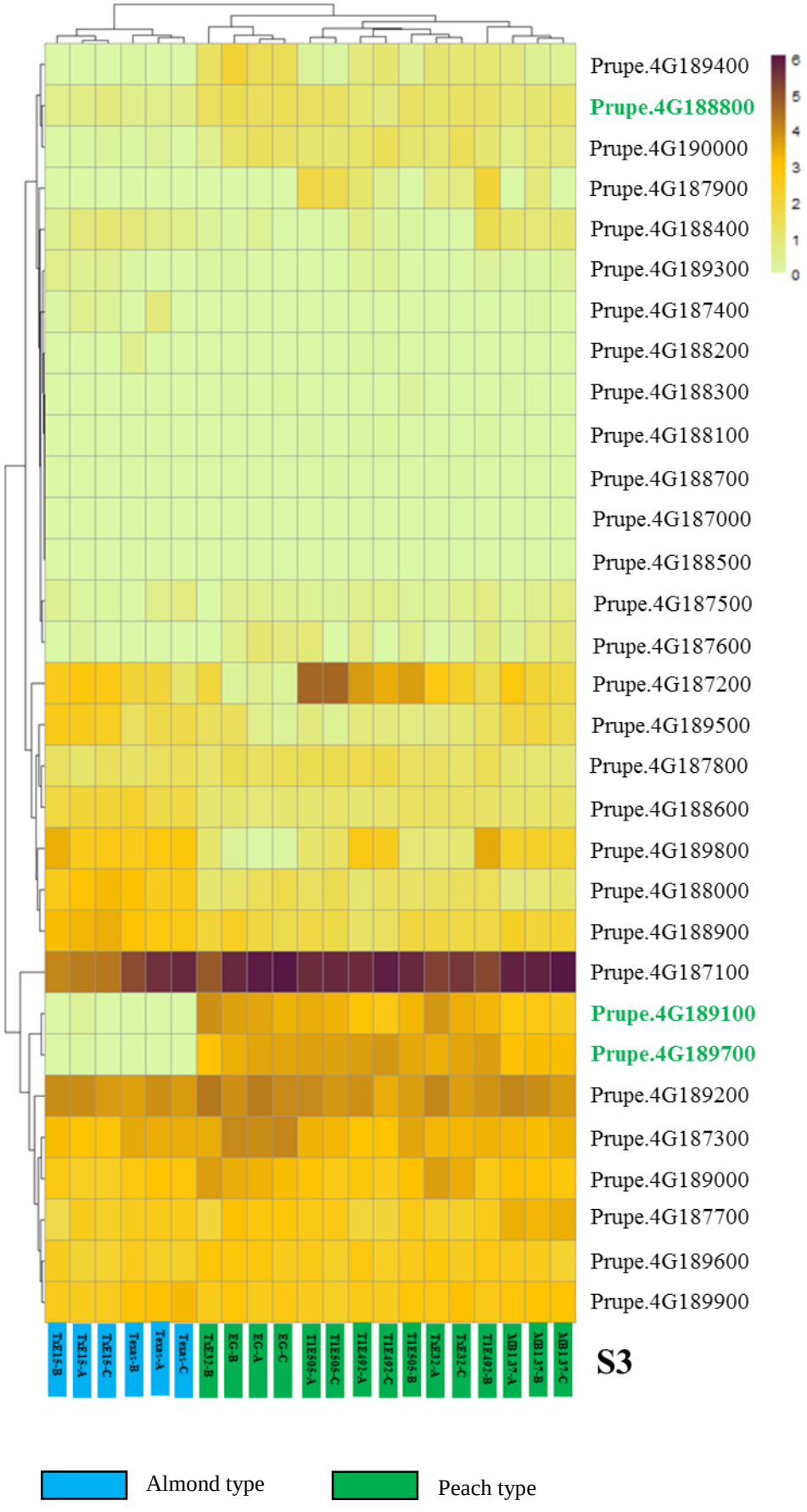
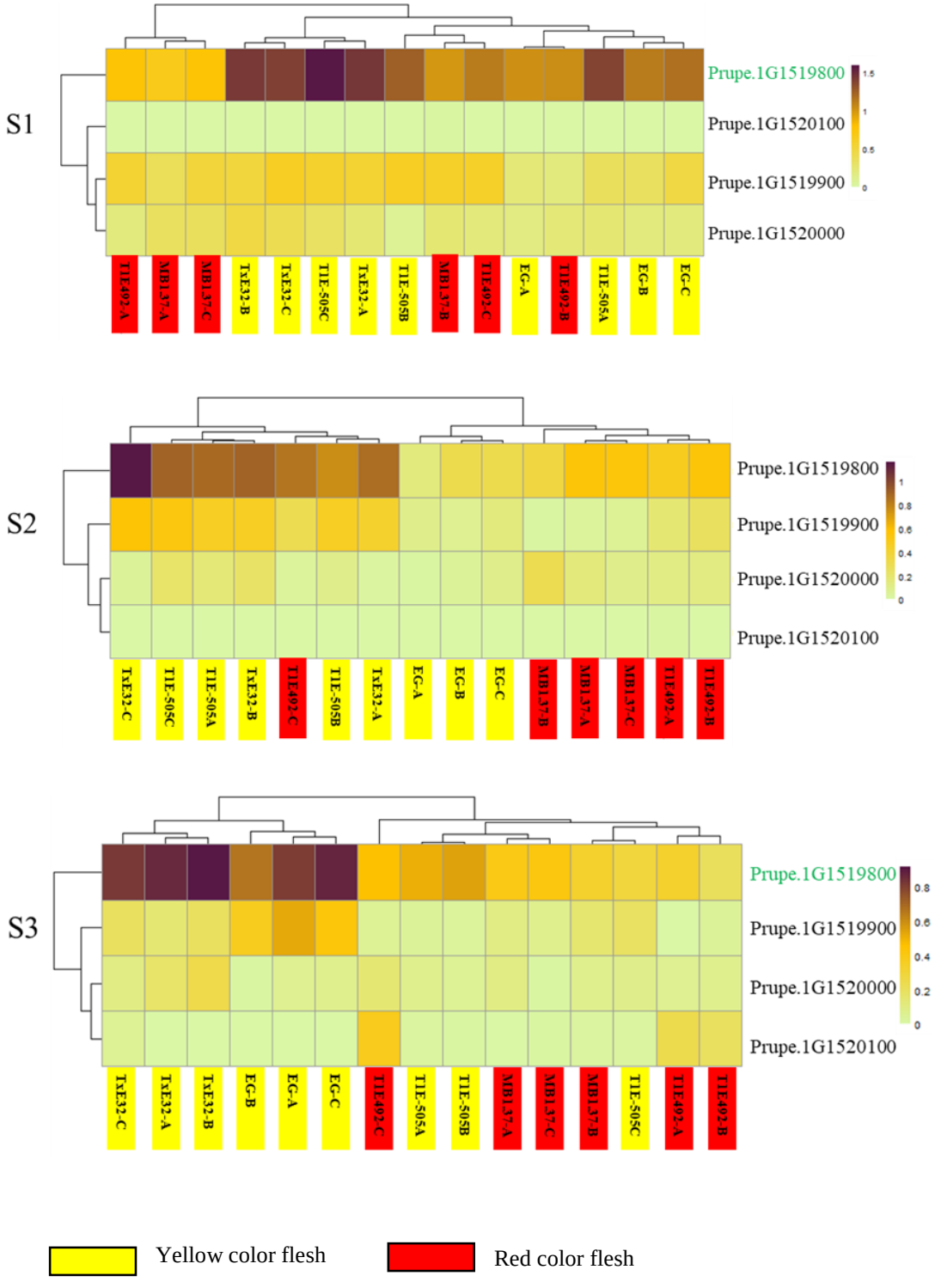

Prupe.4G189400
Prupe.4G188800
Prupe.4G190000
Prupe.4G187900
Prupe.4G188400
Prupe.4G189300
Prupe.4G187400
Prupe.4G188200
Prupe.4G188300
Prupe.4G188100
Prupe.4G188700
Prupe.4G187000
Prupe.4G188500
Prupe.4G187500
Prupe.4G187600
Prupe.4G187200
Prupe.4G189500
Prupe.4G187800
Prupe.4G188600
Prupe.4G189800
Prupe.4G188000
Prupe.4G188900
Prupe.4G187100
Prupe.4G189100
Prupe.4G189700
Prupe.4G189200
Prupe.4G187300
Prupe.4G189000
Prupe.4G187700
Prupe.4G189600
Prupe.4G189900
6
5
4
3
2
1
0
TxE15-B
TxE15-A
TxE15-C
Texs-B
Texs-A
Texs-C
TxE32-B
EG-B
EG-A
EG-C
TlE505-A
TlE505-C
TlE492-A
TlE492-C
TlE505-B
TxE32-A
TxE32-C
TlE492-B
MB13*-A
MB13*-B
MB13*-C
S3
Almond type
Peach type

Figure S5.2: *DBF2* candidate genes differential expression at different fruit developmental stages, S1, S2 and S3. Individuals with yellow color flesh were represented in yellow color and red color flesh in red color.



6. General Discussion

Peach and almond are two cultivated species of the *Prunus* genus with high global economic importance, being Spain the second worldwide producer of both. They share a small and diploid genome (approx. 250 Mb), with eight chromosome pairs that originated from a common ancestor in Central Asia approximately 5 million years ago (Yu et al. 2018; Alioto et al. 2020). Their sexual compatibility made Charles Darwin believe that they were the same species and considered that "peaches are almonds modified in a marvelous way" (Darwin 1868). However, peaches and almonds have many other important differences, some of which obvious, such as the distinct morphology of their product (the mesocarp or the seed contained in the fruit), and that of the leaves and other organs of the plant. Others are not so evident, such as a different mating system, selfing in peaches and cross-pollinating in almonds, which determines a very different level of genetic variability in the two species, where peaches are much less variable than almonds. This thesis intends to contribute to understand the genetics of some traits that segregate in peach and interspecific almond × peach populations, and to develop genetic resources for the analysis of the variability of almond in the peach genomic background that facilitate the introgression of interesting almond genes into the more genetically impoverished peach genome.

Several studies were previously done using F2 and BC1 progenies of almond × peach crosses using 'Texas' as the almond donor parent and 'Earlygold' as the recurrent parent, allowing to understand the genetics of certain aspects of the genetics of each species and their interspecific interaction. First, it was possible to identify a cytoplasmic male-sterility system, where individuals with the almond cytoplasm in an essentially peach background did not produce pollen unless they had an almond allele in at least one of the two independent restorer genes (Donoso et al. 2015). This is another indication of the great genetic distance between peach and almond. The consequences of that were the need to perform crosses with 'Earlygold' as female parent to recover its cytoplasm and to produce fertile individuals, which implied a supplementary generation to conclude our work, and the obtention of early maturing plants that required the use of embryo rescue for germination, increasing the time and resources needed and probably decreasing the number of individuals obtained in each generation.

The second important observation was that the genetics of many of the measured traits was controlled by major genes and major QTLs when analyzing these progenies: 11 major genes that could be analyzed qualitatively and seven of the 33 QTLs that explained at least 20% of the phenotypic variability (Donoso et al. 2016). Most of these genes were also located in places different from other loci mapped for the same character using intraspecific peach crosses (Aranzana et al. 2019). This showed that almond is providing a great deal of novel variability with alleles that produce large phenotypic effects.

To extract more information, particularly relevant information applicable to peach breeding, we needed to analyze the variability with a much finer approach to unravel QTLs that produce more subtle phenotypes and to estimate their effects in the peach background. For this reason, we started the construction of the peach-almond IL collection that we present here. One of the first decisions that had to be taken was which genotype to use as recurrent parent, considering that peach commercial varieties are partly heterozygous. We opted for 'Earlygold' essentially due to the lack of other obviously better alternatives (vigorous homozygous peaches that belong to the major commercial gene pool), to save time (changing the recurrent parent would have delayed these results for a few additional generations), and because we knew from earlier work that 'Earlygold' has approximately half of its genome identical by descent, a circumstance that has been confirmed twice in this thesis (in the F2 of 'Earlygold' and in the IL collection). The accurate analysis of the genetics of quantitative traits with the IL collection required a prior study of the segregation of 'Earlygold' that was done using its F2 progeny and that yielded a set of 12 consistent QTLs. The positions and effects of these QTLs should be taken into consideration when analyzing these characters in the IL collection, as they can potentially interfere with the phenotypes caused by the alleles coming from the almond introgressed fragments (Kalluri et al. 2021; Chapter 3).

The comparison between the results of QTL analysis in the F2 population and the 'Earlygold' component (E) of the T1E progeny resulted in a much larger number of QTLs in the F2 population compared to E using in both cases the same SNP chip and with a similar number of plants. The reasons are that the E map and QTL analysis was based on backcross type segregations (1:1), whereas in E×E marker segregations were 1:2:1, allowing the estimation of the gene action and a finer estimation of the effects of each QTL. In addition, the number of expected recombinations was double in the case of E×E than in E, resulting in a more accurate estimation of the QTL map position. This could have been partly solved in E considering only the 1:2:1 segregating markers, although in this case the markers used would be limited to those that were heterozygous for both E and the T×E hybrid, strongly reducing the power of the analysis. We found also many more markers segregating in E×E than in T1E, but most new markers segregated as dominant (3:1) suggesting that they were discarded in T1E because one of their alleles could not be interpreted by the SNP platform used, behaving as a null and determining a non-segregating pattern. These additional markers contributed, although individually to a lesser extent than codominant markers, to have an increased resolution in the E×E progeny.

The QTL with most relevant effects in the segregation of E×E was that of chromosome 4 coinciding with the position of the *MD* locus. Apart from the high magnitude of its effects, this QTL is in a region where other important QTLs are located affecting different characters. One of them, FDP,

can be considered a different measurement of the maturity date, and given the results of the QTL analysis a slightly more accurate one. Other traits with QTLs at this region may be physiologically related to maturity date, with later-maturing fruit being heavier (FW) and producing more SSC, while other characters are apparently unrelated, such as leaf color at senescence. In addition, one of the major genes identified, the almond fruit gene (*Alf*), is also located at this region. With the current information, it is unclear whether these phenotypes are caused by pleiotropy of a unique gene or by different genes located at the same region. The existence of heterozygous ILs for this region may easily generate a set of subILs with recombinations at this region, providing an opportunity to unveil the genetic nature of these traits, by separating the effects of different linked genes, provided that the cause of the observed variation is oligogenic. This has already started with the search for recombinants to fine map *Alf* (Chapter 5), where the candidate proposed for *MD* is a NAC transcription factor (*Prupe.4G186800*) (Eduardo et al. 2015). We found that this gene is very close, but out of the range of the genomic region of chromosome 4 that contains *Alf*, suggesting that at least in this case the causal genes of *MD* and *Alf* would be different.

From the enormous gene pool of almond variability (Velasco et al. 2015), the part that we are analyzing in the ILs is very small: a single gamete of 'Texas'. Some important characters that almond can transfer, such as disease resistances, are known to exist only in certain almond individuals, and not in all the species, as it occurs in sharka (Cirilli et al. 2016) or in gummosis (Mancero-Castillo et al. 2018) resistance. While the current IL collection may give important information on major differences between peach and almond, specific ILs will be needed for characters present in other almond or other exotic genotypes. One approach to understand and exploit exotic variability was proposed by Serra et al. (2016) where a few (15-25) pre-introgression lines (prILs; lines with 2-4 introgressions), may be obtained using molecular markers from large BC1 or BC2 progenies of different crosses of peach with almond or other related wild species. These prILs may be initially used to find the position of major genes/QTLs, and later ILs may be obtained from prILs with genes of interest in one backcross or selfing generation. This last step can also be done with elite breeding lines as recurrent parents that may result in materials with commercial value, immediately or after one more crossing step.

One interesting characteristic of the set of ILs obtained is that they can be analyzed in homozygous and heterozygous state, meaning that gene action can be studied in detail. This is an advantage compared to the usual IL collections made by homozygous ILs that require an additional step of production of hybrids between a homozygous IL and the recurrent parent to obtain the heterozygous line. Heterozygous lines also provide an opportunity to obtain sublines that have shorter introgressed fragments to generate data for fine mapping of specific regions or to separate possible

loci affecting the same character that are located on the same chromosome. The sizes of the introgressed fragments in our IL collection are large, due to the short number of generations used for its extraction. Certain lines of the heterozygous collection cover >90% of a chromosome and this occurs in seven of the eight *Prunus* chromosomes (all but chromosome 1). This would facilitate the initial identification of QTLs of interest by phenotyping only a few lines containing entire (or almost entire) chromosomes, and later generate line collections of the "Stepped Aligned Inbred Recombinant Strains; STAIRS" type (Koumproglou et al. 2002), facilitating the fine mapping of genes of interest.

When examining the progenies of distant crosses, it is usual to find genes that have unexpected effects, in the sense that the exotic donor parent allele increases the value of a character that it has to a very limited extent but is largely present in the cultivated types (Tanksley and McCouch 1991; Tanksley and Nelson, 1996; Zamir 2001). This is one of the main reasons why wild crop relatives are a valuable source of variability for crop species. In our peach × almond crosses we have found several examples of it in certain major genes or QTLs. One of them is the blood fruit flesh gene (*DBF2*), where the allele coming from almond, confers the red flesh phenotype, while the almond fruit does not seem to express this trait. Another one of high interest is the fruit weight major QTL located at the beginning of chromosome 6 that was found mainly in 'Texas' × 'Earlygold' derived materials (Donoso et al. 2016; Hernández Mora et al. 2017) and in the ILs (Chapter 4) and where the almond allele produces fruit with increased weight. A major QTL at the same position has recently been found examining a collection of non-flat peaches (Cirilli et al. 2021), suggesting that the causal gene is variable in both peach and almond. Finally, in an example of lesser economic interest, the length of the petiole is higher in almonds than in peaches, but one of the ILs with an almond fragment at the end of chromosome 2, produced significantly shorter petioles. The detailed analysis of the peach-almond IL collection is likely to produce more examples of this kind, and others that may identify genes for expected advantageous effects of the almond parent, particularly, rusticity, yield stability, drought and heath tolerance, and resistance to other pests and diseases.

The collection of ILs in homozygosis and heterozygosis covered 83 and 99%, respectively, of the almond gamete genome that was transferred to the MB1.37 hybrid from 'Texas'. While it is difficult to establish at this point slight differences in vigor or other plant aspects that may indicate inbreeding effects due to the current heterogeneity of the IL collection (age, rootstock, etc.), all plants of the collection have a normal appearance. In the case of homozygous plants, we have examples of individuals homozygous for the 16% of the genome that was not included in the final IL set, that have a normal growth. In all, this indicates that the original gamete of almond does not carry lethal or deleterious alleles that affect seriously the viability or plant growth in its homo- or heterozygous

progeny. Considering that the obligate cross-pollinating behavior of almond due to its functional gametophytic self-incompatibility system has allowed it to maintain a large level of variability, it would have not been surprising that almond would carry alleles conferring deleterious phenotypes in homozygosis, that we have not found in our IL collection.

Two genes of almond have immediate potential to be incorporated into peach and have a commercial value. One is the gene of powdery mildew resistance (*Vr3*), where ILs with small introgressed fragments have already been obtained and integrated into IRTA's breeding program (Marimon et al. 2021). The other is the gene for blood flesh *DBF2*, where some of the ILs obtained here could be used as starting materials for the generation of cultivar with red flesh. However, these ILs have relatively long fragments that include also the de dominant allele of *Jui* coming from almond that results in non-juicy fruit and that would require the generation of recombinant individuals between these two loci with the DBF2-/*juijui* combination. As previously mentioned, a third locus of interest for peach is the fruit size increase QTL on chromosome 6, although this QTL needs further confirmation with a more uniform set of replicated ILs that will be available in the coming years.

Although our original goal was to look for almond genes that could be useful for peach breeding, knowledge of the genetics of interspecific populations has made possible to identify initially unsuspected applications in the opposite direction. Two of the genes that were fine mapped, *Alf* and *Jui*, account for a large part of the differences between the fruits of almond and peach. The in depth understanding of their molecular nature will help us to understand some key elements involved on how a fruit can evolve from dry, like almond and the ancestral species that generated peaches and almonds (Yu et al. 2018), to a fleshy and edible fruit, like peach. One interesting consequence of the simple inheritance observed for this character is that it would be easy to introduce the alleles of peach that confer the peach-like fruit into the almond gene pool. Assuming that the peach alleles act in a similar way as in peach when integrated in an almond background, one would expect to obtain almond trees that produce peach-like fruit. This means also to have all the variability carried by the almond, including the one that confers traits of rusticity and adaptation to drier and hotter conditions expected to be brought by climate change. These activities have started already in IRTA and CRAG, where the MB1.37 hybrid has been crossed with almond cultivars and their progeny selected with markers for the dominant *Alf* and the recessive *jui* alleles from peach. One more generation is needed to obtain *juijui* individuals that would have juicy in addition to thick mesocarps and to start to evaluate the potential of this approach to generate materials of agricultural value.

7. Conclusions

1. A high-density SNP map for the ‘Earlygold’ F2 progeny was constructed with 1,640 SNPs, 56% more than the 1,050 obtained for ‘Earlygold’ with the same set of markers in the BC1 (‘Texas’ × ‘Earlygold’) × ‘Earlygold’. Both maps detected the same putatively identical-by-descent genome regions accounting for approximately half of the ‘Earlygold’ genome.

2. QTL analysis has proven to be more efficient with the ‘Earlygold’ F2 population, with 26 QTLs detected, than the BC1 population, where only six QTLs were identified for the same characters analyzed and with similar population sizes. This shows the greater power of F2 compared to BC1 (or F1 segregating) progenies for inheritance analysis.

3. Two consistent QTLs were reported for the leaf senescence color trait in the ‘Earlygold’ F2 population, at the same positions as other color-related genes and QTLs of other plant organs. This trait was correlated with maturity date, where early-maturing individuals had anthocyanic leaves while those maturing later had non-anthocyanic leaves.

4. Several fruit related QTLs, maturity date (MD), fruit development period (FDP), soluble solid content (SSC) and titratable acidity (TA) were identified at the same position in G4, corresponding with the effects of a gene hotspot in this region or with MD having possible pleiotropic effects with all the other traits.

5. Two introgression line (IL) collections of almond DNA fragments in the peach background were developed with a total of 67 lines: 39 with heterozygous introgressions (99% almond genome coverage) and 28 with homozygous introgressions (83%). This is the first collection of this kind reported in a woody perennial, and it has required a 15-year period to be developed from the F1.

6. The 18 k SNP chip used to refine an initial selection of ILs made with SSRs improved marker density by more than 50-fold: from one SSR every 2.0 Mb to one SNP every 37 kb. This made possible to obtain a robust final set of ILs and to identify several smaller introgressions, most <1% of the peach genome, that were not detected using only SSRs.

7. The SNP analysis also led to obtain the genotypes for the ‘Earlygold’ background in the IL collections. This information, along with that obtained with the QTL analysis of the ‘Earlygold’ F2, is vital to identify possible background effects.

8. The phenotypes of traits determined by four major genes, maturity date (*MD*), juiciness (*Jui*), blood flesh color (*DBF2*), and powdery mildew resistance (*Vr3*), and four quantitative characters, fruit weight (FW), soluble solid content (SSC), titratable acidity (TA) and petiole length (PL) were examined in the ILs. Results on their genetics were generally consistent with the previous information available from other almond × peach populations.

9. Some of the introgression lines carrying valuable genes from almond, such as those responsible for the peach powdery mildew resistance and the blood flesh color identified in this research, have already been incorporated into IRTA's peach breeding program.

10. Fine mapping narrowed down the genomic region of two major genes identified in the T×E and T1E populations. Almond fruit type (*Alf*) from 1.6 Mb to 183 kb, and dominant blood flesh (*DBF2*) from 1.05 Mb to 10 kb. These new regions contain 31 (*Alf*), and four (*DBF2*) positional candidate genes. For the juiciness gene (*Jui*) the results presented were useful to correct the initial estimation of 200 kb for the target region of this gene to 392 kb containing 95 candidate genes.

11. For *Alf*, DNA sequence variation and expression studies were inconclusive. Preliminarily, *Prupe.4G187100* (NAC transcription factor) and *Prupe.4G188700* (ULTRAPETALA 2), that were reported to have a role in fruit development and ripening in other species are our probable candidates. This could also be confirmed from our new recombinants once they start to produce fruit, further reducing the current genomic region.

12. A UDP-glycosyltransferase gene (*Prupe.1G519800*) was identified as a possible candidate for *DBF2* from expression analysis studies. Functional validation of this candidate by cloning it into *Agrobacterium tumefaciens* to study transient gene expression in *Nicotiana benthamiana* leaves or in *P. persica* fruits, may provide additional evidence of its causal nature for the red flesh trait.

8. Bibliography

Alioto T, Alexiou KG, Bardil A, Barteri F, Castanera R, Cruz F, Dhingra A, Duval H, Marti AFI, Frias L, Galan B, Garcia JL, Howad W, Gomez-Garrido J, Gut M, Julca I, Morata J, Puigdomenech P, Ribeca P, Cabetas MJR, Vlasova A, Wirthensohn M, Garcia-Mas J, Gabaldon T, Casacuberta JM, Arus P (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

Aranzana MJ, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Ryder CD, Testolin R, Abbott A, King GJ, Iezzoni AF, Arús P (2003) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. Theor Appl Genet 106:819-25

Aranzana MJ, Decroocq V, Dirlewanger E, Eduardo I, Gao ZS, Gasic K, Iezzoni A, Jung S, Peace C, Prieto H, Tao R, Verde I, Abbott AG, Arús P (2019) Prunus genetics and applications after de novo genome sequencing: achievements and prospects. Hortic Res 6:25

Arús P, Messeguer R, Viruel M, Tobutt K, Dirlewanger E, Santi F, Quarta R, Ritter E (1994) The European *Prunus* mapping project Progress in the almond linkage map. Euphytica 77:97-100

Arús P, Ballester J, Jáuregui B, Joobeur T, Truco M, Carmen de Vicente M (1998) The European *Prunus* mapping project: update on marker development in almond. Acta Hort 484:331–338

Arús P, Verde I, Sosinski B, Zhebentyayeva T, Abbott AG (2012) The peach genome. Tree Genet Genomes 8:531-547

Bailey JS, French AP (1942) The inheritance of blossom type and blossom size in peach. Proc Am Soc HortScience 40:248-250

Balakrishnan D, Surapaneni M, Mesapogu S, Neelamraju S (2019) Development and use of chromosome segment substitution lines as a genetic resource for crop improvement. Theor Appl Genet 132:1-25

Balakrishnan D, Surapaneni M, Yadavalli VR, Addanki KR, Mesapogu S, Beerelli K, Neelamraju S (2020) Detecting CSSLs and yield QTLs with additive, epistatic and QTL×environment interaction effects from *Oryza sativa* × *O. nivara* IRGC81832 cross. Sci Rep 10:7766

Ballester J, Socias i Company R, Arús P, de Vicente MC (2001) Genetic mapping of a major gene delaying blooming time in almond. Plant Breed 120:268–270

Baltazar P, Correa EC, Diezma B (2020) Instrumental procedures for the evaluation of juiciness in peach and nectarine cultivars for fresh consumption. Agronomy 10:152

Bardakci F (2001) Random Amplified Polymorphic DNA (RAPD) Markers. Turk J Biol 25:185-196

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887-898

Bliss FA, Arulsekar S, Foolad MR, Becerra V, Gillen AM, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK (2002) An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. Genome 45:520–529

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-31

Boudehri K, Bendahmane A, Cardinet G, Troadec C, Moing A, Dirlewanger E (2009) Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. BMC Plant Biol 9:59

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525-7

Breto MP, Cantin CM, Iglesias I, Arús P, Eduardo I (2017) Mapping major gene for red skin color suppression (highlighter) in peach. Euphytica 213:14

Bunyard EA (1938) The history and cultivation of the peach and nectarine. J Royal Hort Soc 63:114-121

Byrne DH (1990) Isozyme variability in four diploid stone fruits compared with other woody perennial plants. J Hered 81:68–71

Byrne DH (2005) Trends in stone fruit cultivar development. J Am Soc Hortic Sci 15:494–500

Byrne DH (2007) Molecular marker use in perennial plant breeding. Acta Hort 751:163–168

Byrne D, Raseira M, Bassi D, Piagnani M, Gasic K, Reighard G, Moreno M and Pérez S (2012) Peach. In: Badenes ML, Byrne DH, (Eds) Fruit breeding (handbook of plant breeding). Springer, New York, pp 513–569

Calle A, Serradilla MJ, Wünsch A (2021) QTL mapping of phenolic compounds and fruit colour in sweet cherry using a 6+9K SNP array genetic map. Sci Hortic 280:109900

Cantín CM, Arús P, Eduardo I (2018) Identification of a new allele of the *Dw* gene causing brachytic dwarfing in peach. BMC research notes 11:386

Cao K, Zheng ZJ, Wang LR, Liu X, Zhu GR, Fang WC, Cheng SF, Zeng P, Chen CW, Wang XW, Xie M, Zhong X, Wang XL, Zhao P, Bian C, Zhu YL, Zhang JH, Ma GS, Chen CX, Li YJ, Hao FG, Li Y, Huang GD, Li YX, Li HY, Guo J, Xu X, Wang J (2014) Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. Genome Biol 15:15

Cao K, Zhou Z, Wang Q, Guo J, Zhao P, Zhu G, Fang W, Chen C, Wang X, Wang X, Tian Z, Wang L (2016) Genome-wide association study of 12 agronomic traits in peach. Nat Commun 7:13246

Castillo DM, Beckman TG, Harmon PF, Chaparro JX (2018) Amajor locus for resistance to *Botryosphaeria dothidea* in *Prunus*. Tree Genet Genomes 14:26

Chaparro JX, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theor Appl Genet 87:805–815

Chen C, Guo J, Cao K, Zhu G, Fang W, Wang X, Li Y, Wu J, Xu Q, Wang L (2021) Identification of candidate genes associated with slow-melting flesh trait in peach using bulked segregant analysis and RNA-seq. Sci Hortic 286:110208

Cheng J, Wei G, Zhou H, Gu C, Vimolmangkang S, Liao L, Han Y (2014) Unraveling the mechanism underlying the glycosylation and methylation of anthocyanins in peach. Plant Physiol 166:1044-1058

Chin SW, Shaw J, Haberle R, Wen J, Potter D (2014) Diversification of almonds, peaches, plums and cherries – Molecular systematics and biogeographic history of *Prunus* (Rosaceae). Mol Phylogenet Evol 76:34–48

Christenhusz MJM, Byng JW (2016) The number of known plant species in the world and its annual increase. Phytotaxa 261:201-217

Ciacciulli A, Cirilli M, Chiozzotto R, Attanasio G, Linge CDS, Pacheco I, Rossini L, Bassi D (2018) Linkage and association mapping for the slow softening (SwS) trait in peach (*P. persica* L. Batsch) fruit. Tree Genet Genomes 14:93

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80-92

Cirilli M, Geuna F, Babini AR, Bozhkova V, Catalano L, Cavagna B, Dallot S, Decroocq V, Dondini L, Foschi S, Ilardi V, Liverani A, Mezzetti B, Minafra A, Pancaldi M, Pandolfini T, Pascal T, Savino VN, Scorza R, Verde I, Bassi D (2016) Fighting sharka in peach: Current limitations and future perspectives. Front Plant Sci 7:1290

Cirilli M, Baccichet I, Chiozzotto R, Silvestri C, Rossini L, Bassi D (2021) Genetic and phenotypic analyses reveal major quantitative loci associated to fruit size and shape traits in a non-flat peach collection (*Prunus persica* L. Batsch). Hortic Res 8:232

Claverie M, Bosselut N, Lecouls AC, Voisin R, Lafargue B, Poizat C, Kleinhentz M, Laigret F, Dirlewanger E, Esmenjaud D (2004) Location of independent root-knot nematode resistance genes in plum and peach. Theor Appl Genet 108:765–773

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196

Cuenca J, Aleza P, Garcia-Lor A, Ollitrault P, Navarro L (2016) Fine mapping for identification of *Citrus Alternaria* brown pot candidate resistance genes and development of new SNP markers for marker-assisted selection. Front Plant Sci 7:1948

Darwin C (1868) The variation of animals and plants under domestication. John Murray, New York

Delplancke M, Yazbek M, Arrigo N, Espindola, A, Joly H, Alvarez N (2016) Combining conservative and variable markers to infer the evolutionary history of *Prunus* subgen. *Amygdalus* s.l. under domestication. Genet Resour Crop Evol 63:221–234

Dettori MT, Quarta R, Verde I (2001) A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. Genome 44:783–790

Di Guardo M, Micheletti D, Bianco L, Putten HJK, Longhi S, Costa F, Aranzana MJ, Velasco R, Arús P, Troggio M, Weg EWV (2015) ASSIsT: an automatic SNP scoring tool for in- and outbreeding species. Bioinformatics 31:3873-3874

Dirlewanger E, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. Theor Appl Genet 97:888–895

Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Guye A, Plomion C, Monet R (1999) Mapping QTL controlling fruit quality in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 98:18–31

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howard W, Arús P (2004) Comparative mapping and marker assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci USA 101:9891-9896

Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet Genomes 3:1–13

Dirlewanger E, Quero-García J, Le Dantec L, Lambert P, Ruiz D, Dondini L, Illa E, Quilot-Turion B, Audergon J-M, Tartarini S, Letourmy P, Arús P (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity 109:280-292

Donoso JM, Picañol R, Eduardo I, BatlIe I, Howad W, Aranzana MJ, Arús P (2015) High-density mapping suggests a cytoplasmic male sterility system with two restorer factors in almond x peach progenies. Hortic Res 2:15016

Donoso JM, Picañol R, Serra O, Howad W, Alegre S, Arús P, Eduardo I (2016) Exploring almond genetic variability useful for peach improvement: mapping major genes and QTLs in two inter-specific almond × peach populations. Mol Breeding 36:16

Doyle JJ, Doyle JL (1990) A rapid total DNA preparation procedure for fresh plant tissue. Focus 12:13–15

Duval H, Hoerter M, Polidori J, Confolent C, Masse M, Moretti A, Ghelder CV, Esmenjaud D (2014) High-resolution mapping of the *RMia* gene for resistance to root-knot nematodes in peach. Tree Genet Genomes 10:297–306

Eduardo I, Arús P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI 161375. Theor Appl Genet 112:139–148

Eduardo I, Pacheco I, Chietera G, Bassi D, Pozzi C, Vecchietti A, Rossini L (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

Eduardo I, Chietera G, Pirona R, Pacheco I, Troggio M, Banchi E, Bassi D, Rossini L, Vecchietti A, Pozzi C (2013) Genetic dissection of aroma volatile compounds from the essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. Tree Genet Genomes 9:189–204

Eduardo I, López-Girona E, BatlIe I, Reig G, Iglesias I, Howad W, Arús P, Aranzana MJ (2014) Development of diagnostic markers for selection of the subacid trait in peach. Tree Genet Genomes 10:1695-1709

Eduardo I, Picañol R, Rojas E, Batlle I, Howad W, Aranzana MJ, Arús P (2015) Mapping of a major gene for the slow ripening character in peach: co-location with the maturity date gene and development of a candidate gene-based diagnostic marker for its selection. Euphytica 205:627-636

Eduardo I, Tomas CD, Alexiou K, Giovannini D, Pietrella M, Carpenedo S, Raseira MCB, Battle I, Cantin CM, Aranzana MJ, Arús P (2020) Fine mapping of the peach pollen sterility gene (*Ps*/*ps*) and detection of markers for marker-assisted selection. Mol Breeding 40:57

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Emeriewen OF, Richter K, Flachowsky H, Malnoy M, Peil A (2021) Genetic Analysis and Fine mapping of the fire blight resistance locus of *Malus* × *arnoldiana* on linkage group 12 reveal first candidate genes. Front Plant Sci 12:667133

Eshed Y, Abu-Abied M, Saranga Y, Zamir D (1992) *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. Theor Appl Genet 83:1027-1034

Eshed Y, Zamir D (1995). An introgression-line population of *Lycopersicon pennelli* in the cultivated tomato enables the identification and fine mapping of yield-associated QTLs. Genetics 141:1147–1162

Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Dumas LS, Cosson P, Pronier V, Monet R, Dirlewanger E (2002) Candidate genes and QTL for sugar and organic acid content in peach (*Prunus persica* (L.) Batsch). Theor Appl Genet 105:145–159

Fan S, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG (2010) Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). New Phytol 185:917–930

Faust M, Timon B (1995) Origin and dissemination of peach. Hortic Rev 17:331–379

Felipe AJ (2009) 'Felinem, 'Garnem' and 'Monegro' almond x peach hybrid rootstocks. HortScience 44:196–197

Fiol A, García-Gómez BE, Jurado-Ruiz F, Alexiou K, Aranzana MJ (2021) Characterization of Japanese plum (*Prunus salicina*) PsMYB10 gene alleles reveals structural variation and polymorphisms correlating with fruit skin color. Front Plant Sci 12:655267

Fletcher RS, Mullen JL, Yoder S, Bauerle WL, Reuning G, Sen S, Meyer E, Juenger TE, Mckay JK (2013) Development of a next generation NIL library in *Arabidopsis thaliana* for dissecting complex traits. BMC Genom 14:655–668

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit Size. Science 289:85-88

Frett TJ, Reighard GL, Okie WR, Gasic K (2014) Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. Tree Genet Genomes 10:367-381

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786-1789

García-Gómez BE, Ruiz D, Salazar JA, Rubio M, Martínez-García PJ, 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

Gasic K, Linge CDS, Bianco L, Troggio M, Rossini L, Bassi D, Aranzana MJ, Arús P, Verde I, Peace C, Iezzoni A (2019) Development and evaluation of a 9K SNP addition to the peach IPSC 9K SNP array v1. HortScience 54:S188

Gattolin S, Cirilli M, Pacheco I, Ciacciulli A, Linge CDS, Mauroux JB, Lambert P, Cammarata E, Bassi D, Pascal T, Rossini L (2018) Deletion of the miR172 target site in a TOE-type gene is a strong candidate variant for dominant double-flower trait in Rosaceae. Plant J 96:358-371

Giacomo MD, Luciani MD, Cambiaso V, Zorzoli R, Rodriguez GR, Costa JHPD (2020) Tomato near isogenic lines to unravel the genetic diversity of *S. pimpinellifolim* LA0722 for fruit quality and shelf life breeding. Euphytica 216:126

Gillen AM, Bliss FA (2005) Identification and mapping of markers linked to the *Mi* gene for root-knot nematode resistance in peach. J Am Soc Hortic Sci 130:24-33

Giovannoni JJ (2004) Genetic regulation of fruit development and ripening. Plant Cell 16:S170-S180

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137

Guo ZH, Zhang YJ, Yao JL, Xie ZH, Zhang YY, Zhang SL, Gu C (2021) The NAM/ATAF1/2/CUC2 transcription factor PpNAC.A59 enhances PpERF.A16 expression to promote ethylene biosynthesis during peach fruit ripening. Hortic Res 8:209

Gur A, Zamir D (2015) Mendelizing all components of a pyramid of three yield QTL in tomato. Front Plant Sci 6:1096

Hancock JF, Scorza R, Lobos GA (2008) Peaches. In: Hancock JF (Ed.) Temperate fruit crop breeding, Springer, Netherlands, pp 265-298

Hayden MJ, Nguyen TM, Waterman A, Chalmers KJ (2008) Multiplex-Ready PCR: A new method for multiplexed SSR and SNP genotyping. BMC Genom 9:80

Hurtado MA, Romero C, Vilanova S, Abbott AG, Llácer G, Badenes ML (2002) Genetic linkage maps of two apricot cultivars (*Prunus armeniaca* L.), and mapping of PPV (sharka) resistance. Theor Appl Genet 105:182–191

Hernández Mora JR, Micheletti D, Bink M, van de Weg E, Cantín C, Nazzicari N, Caprera A, Detori MT, Micali S, Banchi E, Campoy JA, Dirlewanger E, Lambert P, Pascal T, Troggio M, Bassi D, Rossini L, Verde I, Quilot-Turion B, Laurens F, Arús P, Aranzana MJ (2017) Integrated QTL detection for key breeding traits in multiple peach progenies. BMC Genom 18:404

Iezzoni AF, McFerson J, Luby J, Gasic K, Whitaker V, Bassil N, Yue CY, Gallardo K, McCracken V, Coe M, Hardner C, Zurn JD, Hokanson S, van de Weg E, Jung S, Main D, Linge CD, Vanderzande S, Davis TM, Mahoney LL, Finn C, Peace C (2020) RosBREED: bridging the chasm between discovery and application to enable DNA-informed breeding in rosaceous crops. Hortic Res 7:177

Jaganathan D, Bohra A, Thudi M, Varshney RK (2020) Fine mapping and gene cloning in the post-NGS era: advances and prospects. Theor Appl Genet 133:1791-1810

Jáuregui B (1998) Identification of molecular markers linked to agronomic characters in an interspecific almond × peach progeny. University of Barcelona, Spain

Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C, Skriver K (2010) The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. Biochem J 426:183-96

Jeuken MJW, Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theor Appl Genet 109:394–401

Joobeur T, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map in *Prunus* using an almond x peach F2 progeny. Theor Appl Genet 97:1034-1041

Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, Frank M, Ru SS, Hough H, Evans K, Peace C, Olmstead M, DeVetter LW, McFerson J, Coe M, Wegrzyn JL, Staton ME, Abbott AG, Main D (2019) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Res 47:D1137-D1145

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2010) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309-334

Kalluri N, Eduardo I, Arús P (2021) Comparative QTL analysis in peach 'Earlygold' F2 and backcross progenies. Sci Hortic 293:110726

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175:891–905

Konczak I, Zhang W (2004) Anthocyanins-more than nature's colours. J Biomed Biotechnol 5:239-240

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ (2002) STAIRS: a new genetic resource for functional genomic studies of Arabidopsis. Plant J 31:355-364

Ladizinsky G (1999) On the origin of almond. Genetic Resour Crop Evol 46:143-147

Lalli DA, Abbott AG, Zhebentyayeva TN, Badenes ML, Damsteegt V, Polák J, Krška B, Salava J (2008) A genetic linkage map for an apricot (*Prunus armeniaca* L.) BC1 population mapping Plum pox virus resistance. Tree Genet Genomes 4:481–493

Lambert P, Pascal T (2011) Mapping *Rm2* gene conferring resistance to the green peach aphid (*Myzus persicae Sulzer*) in the peach cultivar "Rubira". Tree Genet Genomes 7:1057-1068

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Laurens F, Aranzana MJ, Arús P, Bassi D, Bonay J, Corelli L, Durel CE, Mes J, Pascal T, Patocchi A, Peli A, Quilot B, Salvi S, Tartarini S, Troggio M, Vecchietti A, Velasco R, Weg EVD (2012) Review of fruit genetics and breeding programmes and a new European initiative to increase fruit breeding efficiency. Acta Hort 929

Lecouls AC, Cabetas MJR, Minot JC, Voisin R, Bonnet A, Salesses G, Dirlewanger E, Esmenjaud D (1999) RAPD and SCAR markers linked to the *Ma1* root-knot nematode resistance gene in Myrobalan plum (*Prunus cerasifera* Her.) Theor Appl Genet 99:328-335

Lill RE, Van der Mespel GJ (1998) A method for measuring the juice content of mealy nectarines. Sci Hortic 36:267-271

Li Z (1984) Peach germplasm and breeding in China. HortScience 19: 348-351

Li XW, Meng XQ, Jia HJ, Yu ML, Ma RJ, Wang LR, Cao K, Shen ZJ, Niu L, Tian JB, Chen MJ, Xie M, Arús P, Gao ZS, Aranzana MJ (2013) Peach genetic resources: diversity, population structure and linkage disequilibrium. BMC Genet 14:84

Lombardo VA, Osorio S, Borsani J, Lauxmann MA, Bustamante CA, Budde CO, Andreo CS, Lara MV, Fernie AR, Drincovich MF (2011) Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. Plant physiol 157:1696-1710

Lu J, Hou J, Ouyang Y, Luo H, Zhao J, Mao C, Han M, Wang L, Xiao J, Yang Y, Li X (2020) A direct PCR-based SNP marker-assisted selection system (D-MAS) for different crops. Mol Breeding 40:9

Lu L, Liu H, Wu Y, Yan G (2020) Development and characterization of near-isogenic lines revealing candidate genes for a major 7AL QTL responsible for heat tolerance in wheat. Front Plant Sci 11:1316

Lu Z, Niu L, Chagne D, Cui G, Pan L, Foster T, Zhang R, Zeng W, Wang Z (2016) Fine mapping of the *temperature-sensitive semi-dwarf* (*Tssd*) locus regulating the internode length in peach (*Prunus persica*). Mol Breeding 36:20

Lu Z, Cao H, Pan L, Niu L, Wei B, Cui G, Wang L, Yao JL, Zeng W, Wang Z (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

Lu ZX, Reighard GL, Nyczepir AP, Beckman TG, Ramming DW (1998) Inheritance of resistance to root-knot nematodes in peach rootstocks. Acta Hort 465:111–116

Mancero-Castillo D, Beckman TG, Harmon PF, Chaparro JX (2018) A major locus for resistance to *Botryosphaeria dothidea* in *Prunus*. Tree Genet Genomes 14:26

Marimon N, Luque J, Arús P, Eduardo I (2020) Fine mapping and identification of candidate genes for the peach powdery mildew resistance gene *Vr3*. Hortic Res 7:175

Martínez-Gómez P, Sánchez-Pérez R, Dicenta F (2006) Fruit development in almond. Acta Hort 726

Martínez-García PJ, Parfitt DE, Ogundiwin EA, Fass J, Chan HM, Ahmad R, Lurie S, Dandekar A, Gradziel TM, Crisosto CH (2013) High density SNP mapping and QTL analysis for fruit quality characteristics in peach (*Prunus persica* L.). Tree Genet Genomes 9:19–36

McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Prud'homme van Reine WF, Smith GF, Wiersema JH, Turland NJ (2012) International code of nomenclature for algae, fungi and plants (Melbourne code) adopted by the 18th International Botanical Congress Melbourne, Australia, July 2011

Meneses C, Ulloa-Zepeda L, Cifuentes-Esquivel A, Infante R, Cantin C, Batlle I, Arús P, Eduardo I (2016) A codominant diagnostic marker for the slow ripening trait in peach. Mol Breed 36:77

Meng G, Zhu G, Fang W, Chen C, Wang X, Wang L, Cao K (2019) Identification of loci for single/double flower trait by combining genome-wide association analysis and bulked segregant analysis in peach (*Prunus persica*). Plant Breed 138:360-367

Micheletti D, Dettori MT, Micali S, Aramini V, Pacheco I, da Silva C, Foschi S, Banchi E, Barreneche T, Quilot-Turion B, Lambert P, Pascal T, Iglesias I, Carbó J, Wang LR, Ma RJ, Li XW, Gao ZS, Nazzicari N, Troggio M, Bassi B, Rossini L, Verde I, Laurens F, Arús P, Aranzana MJ (2015) Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. PLoS One 10:e0136803

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:1212-1217

Mnejja M, Garcia-Mas J, Audergon J-M, Arús P (2010) *Prunus* microsatellite marker transferability across rosaceous crops. Tree Genet Genomes 6:689-700

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breeding 3:87-103

Monet R, Guye A, Roy M, Dachary N (1996) Peach mendelian genetics: a short review and new results. Agronomie 16:321–329

Monfared MM, Carles CC, Rossignol P, Pires HR, Fletcher JC (2013) The ULT1 and ULT2 trxG genes play overlapping roles in Arabidopsis development and gene regulation. Mol Plant 6:1564-79

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome 43:803-813

Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC (2011) Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). Plant J 65:106-118

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yildiz M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G, Baloch FS (2017) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Agriculture and Environmental Biotechnology, 261-285

Núñez-Lillo G, Cifuentes-Esquivel A, Troggio M, Micheletti D, Infante R, Campos-Vargas R, Orellana A, Blanco-Herrera F, Meneses C (2015) Identification of candidate genes associated with mealiness and maturity date in peach [*Prunus persica* (L.) Batsch] using QTL analysis and deep sequencing. Tree Genet Genomes 11:86

Ogundiwin EA, Peace CP, Gradziel TM, Parfitt DE, Bliss FA, Cristosto CH (2009) A fruit quality gene map of *Prunus*. BMC Genom 10:587–600

Okie WR, Bacon T, Bassi D (2008) Fresh market cultivar development. In: D.R. Layne and D. Bassi (Eds.), The peach: botany, production and uses. CAB International 139–174

Olmstead JW, Sebolt AM, Cabrera A, Sooriyapathirana SH, Iriarte G, Wang D, Chen CY, Knaap van der E, Iezzoni AF (2008) Construction of an intra-specific sweet cherry (*Prunus persica* L.) genetic linkage map and synteny analysis with the *Prunus* reference map. Tree Genet Genomes 4897-910

Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci 10:79-87

Osorio S, Alba R, Damasceno CMB, Lopez-Casado G, Lohse M, Zanor MI, Tohge T, Usadel B, Roce JKC, Fei Z, Giovannoni JJ, Fernie AR (2011) Systems biology of tomato fruit development: Combined transcript, protein, and metabolite analysis of tomato transcription Factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. Plant Physiol 157:405-425

Pascal T, Aberlenc R, Confolent C, Hoerter M, Lecerf E, Tuéro C, Lambert P (2017) Mapping of new resistance (*Vr2*, *Rm1*) and ornamental (*Di2*, *pl*) Mendelian trait loci in peach. Euphytica 213:132

Picañol R, Eduardo I, Aranzana MJ, Howad W, Batlle I, Iglesias I, Alonso JM, Arús P (2013) Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. Euphytica 190:279-288

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L (2017) Differential analysis of RNA-seq incorporating quantification uncertainity. Nat Methods 14:687-690

Pirona R, Eduardo I, Pacheco I, Linge CDS, Miculan M, Verde I, Tartarini S, Dondini L, Pea G, Bassi D, Rossini L (2013) Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. BMC Plant Biol 13:166

Potter D, Eriksson T, Evens RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS (2007) Phylogeny and classification of Rosaceae. Plant Syst Evol 266:5-43

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theor Appl Genet 90:503-510

Ramming DW (1990) The use of embryo culture in fruit breeding. HortScience 25:393-398

Ramsay LD, Jennings DE, Bohuon EJR, Arthur AE, Lydiate DJ, Kearsey MJ, Marshal DF (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome 39:558–567

Rawandoozi ZJ, Hartmann TP, Carpenedo S, Gasic K, da Silva Linge C, Cai LC, Van de Weg E, Byrne DH (2021) Mapping and characterization QTLs for phenological traits in seven pedigree-connected peach families. BMC Genom 22: 187

Rehder A (1940) A manual of cultivated trees and shrubs hardy in North America exclusive of the subtropical and warmer temperate regions, 2nd ed. Macmillan, New York, USA

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP (2011) Integrative Genomics Viewer. Nat Biotechnol 29:24-26

Rodriguez CE, Bustamante CA, Budde CO, Müller GL, Drincovich MF, Lara MV (2019) Peach fruit development: A comparative proteomic study between endocarp and mesocarp at very early stages underpins the main differential biochemical processes between these tissues. Front Plant Sci 10:715

Ricci A, Sabbadini S, Prieto H, Padilla IMG, Dardick C, Li Z, Scorza R, Limera C, Mezzetti B, Perez-Jimenez M, Burgos L, Petri C (2020) Genetic transformation in peach (*Prunus persica* L.): challenges and ways forward. Plants 9:971

Salazar JA, Pacheco I, Shinya P, Zapata P, Silva C, Aradhya M, Velasco D, Ruiz D, Martinez-Gomez P, 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 8:476

Sánchez-Pérez R, Pavan S, Mazzeo R, Moldovan C, Cigliano RA, Cueto JD, Ricciardi F, Lotti C, Ricciardi L, Dicenta F, López-Marqués RL, Møller BL (2019) Mutation of a bHLH transcription factor allowed almond domestication. Science 364:1095-1098

Sansavini S, Gamberini A, Bassi D (2006) Peach breeding, genetics and new cultivar trends. Acta Hort 713:23-48

Scorza R, Mehlenbacher SA, Lightner GW (1985) Inbreeding and coancestry of freestone peach cultivars of the Eastern-United-States and implications for peach germplasm improvement. J Am Soc Hortic Sci 110:547-552

Scorza R, Okie WR (1991) Peaches (*Prunus*). In: Moore JN and Ballington RB (Eds) Genetic Resources of Template Fruit and Nut Crops. Acta Hort 290:177–234

Scorza R, Melnicenco L, Dang P, Abbott AG (2002) Testing a microsatellite marker for selection of columnar growth habit in peach (*Prunus persica* (L.) Batsch). Acta Hort 592:285–289

Serra O, Donoso JM, Picañol R, Batlle I, Howad W, Eduardo I, Arús P (2016) Marker-assisted introgression (MAI) of almond genes into the peach background: a fast method to mine and integrate novel variation from exotic sources in long intergeneration species. Tree Genet Genomes 12:96

Serra O (2017) Towards increasing genetic variability and improving fruit quality in peach using genomic and bioinformatic tools (Doctoral dissertation, Universitat Autònoma de Barcelona, Spain). www.tdx.cat/handle/10803/460882

Serra O, Giné-Bordonaba J, Eduardo I, Bonany J, Echeverría G, Larrigaudière G, Arús P (2017) Genetic analysis of the slow melting flesh character in peach. Tree Genet Genomes 13:77

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 10:1135-45

Shen Z, Confolent C, Lambert P, Poëssel J-L, Quilot-Turion B, Yu M, Ma R, Pascal T (2013) 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

Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arús P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE (2008) Multiple models for Rosaceae genomics. Plant Physiol 147:985-1003

Silva C, Garcia-Mas J, Sánchez AM, Arús P, Oliveira MM (2005) Looking into flowering time in almond (*Prunus dulcis* (Mill) D. A. Webb): the candidate gene approach. Theor Appl Genet 110:959-968

Smith SM, Maughan PJ (2015) SNP genotyping using KASPar assays. Methods Mol Biol 1245:243-56

Socquet-Juglard D, Christen D, Devènes G, Gessler C, Duffy B, Patocchi A (2013). Mapping architectural, phenological, and fruit quality QTLs in apricot. Plant Mole Biol Rep 31:387-397

Song C, Zhao S, Hong X, Liu J, Schulenburg K, Schwab W (2016) A UDP-glucosyltransferase functions in both acylphloroglucinol glucoside and anthocyanin biosynthesis in strawberry (*Fragaria* x *ananassa*). Plant J 85:730-742

Sooriyapathirana SS, Khan A, Sebolt AM, Wang D, Bushakra JM, Lin-Wang K, Allan AC, Gardiner SE, Chagné D, Iezzoni AF (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

Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW, Holland JB (2007) QTL mapping with near-isogenic lines in maize. Theor Appl Genet 114:1211–1228

Szymański J, Bocobza A, Panda S, Sonawane P, Cárdenas PD, Lashbrooke J, Kamble A, Shahaf N, Meir S, Bovy A, Beekwilder J, Tikunov Y, Fuente IR, Zamir D, Rogachev I and Aharoni A (2020) Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. Nat Genet 52:1111-1121

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205-233

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191-203

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066

Terakami S, Moriya S, Adachi Y, Kunihisa M, Nishitani C, Saito T, Abe K, Yamamoto T (2016) Fine mapping of the gene for susceptibility to black spot disease in Japanese pear (*Pyrus pyrifolia* Nakai). Breed Sci 66:271-280

Tonutti P, Casson P, Ramina A (1991) Ethylene biosynthesis during peach fruit development. J Am Soc Hortic Sci 116:274-279

Trainin T, Harel-Beja R, Bar-Ya'akov I, Ben-Simhon Z, Yahalomi R, Borochov-Neori H, Ophir R, Sherman A, Doron-Faigenboim A, Holland D (2021) Fine Mapping of the "black" Peel color in

pomegranate (*Punica granatum* L.) strongly suggests that a mutation in the anthocyanidin reductase (ANR) gene is responsible for the trait. Front Plant Sci 12:642019

Tuan PA, Bai S, Yaegaki H, Tamura T, Hihara S, Moriguchi T, 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

Tuinstra MR, Ejeta G, Goldsbrough PB (1997) Heterogenous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005-1011

Urrutia M, Bonet J, Arús P, Monfort A (2015) A near-isogenic line (NIL) collection in diploid strawberry and its use in the genetic analysis of morphologic, phenotypic and nutritional characters. Theor Appl Genet 128:1261-1275

Van Berloo R (2008) GGT 2.0: Versatile software for visualization and analysis of genetic data. J Hered 99:232-236

Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) MapQTL® 4.0. Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen

Velasco D, Hough J, Aradhya M, Ross-Ibarra J (2016) Evolutionary genomics of peach and almond domestication. G3: Genes Genomes Genetics 6:3985-3993

Verde I, Quarta R, Cerdrola C, Dettori MT (2002) QTL analysis of agronomic traits in a BC1 peach population. Acta Hort 592:291– 297

Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troggio M, Sosinski B, Aranzana MJ, Arús P, Iezzoni A, Morgante M, Peace C (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

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel LA, Decroocq V, Sosinski B, Prochnik S, Mitros T, Policriti A, Cipriani G, Dondini L, Ficklin S, Goodstein DM, Xuan P, Del Fabbro C, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B, Bielenberg D, Pirona R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arús P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45:487-494

Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, Paris R, Aramini V, Gazza L, Rossini L, Bassi D, Troggio M, Shu S, Grimwood J, Tartarini S, Dettori MT, Schmutz J (2017) The peach v2.0 release: high re-solution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. BMC Genom 18:225

Vilanova S, Romero C, Abbott AG, Llacer G, Badenes ML (2003) An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits. Theor Appl Genet 107:239-247

Vilanova S, Romero C, Llacer G, Badenes ML (2005) Identification of self-(in)compatibility alleles in apricot by PCR and sequence analysis. J Am Soc Hortic Sci 130:893-898

Vilanova S, Badenes ML, Burgos L, Calvo JM, Llacer G, Romero C (2006) Self-compatibility of two apricot selections is associated with two pollen-part mutations of different nature. Plant Physiol 142:629-641

Von Korff M, Wang H, Leon J, Pillen K (2004) Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum*) as donor. Theor Appl Genet 109:1736–1745

Voorrips RE (2002) Mapchart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP – a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407-4414

Wang H, Wang C, Fan W, Yang J, Appelhagen I, Wu Y, Zhang P (2018) A novel glycosyltransferase catalyses the transfer of glucose to glucosylated anthocyanins in purple sweet potato. J Exp Bot 69:5444-5459

Wang R, Lammers M, Tikunov Y, Bovy AG, Angenet GC, Maagd R (2020) The *rin*, *nor* and *Cnr* spontaneous mutations inhibit tomato fruit ripening in additive and epistatic manners. Plant Sci 294:110436

Wang Y (1985) Peach growing and germplasm in China. Acta Hort 173:51-55

Wang Y, Georgi LL, Reighard GL, Scorza R, Abbott AG (2002) J Hered 93:352–358

Warburton ML, Becerra-Velasquez VL, Goffreda JC, Bliss FA (1996) Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. Theor Appl Genet 93:920–925

Werner DJ, Creller MA, Chaparro JX (1998) Inheritance of the blood-flesh trait in peach. HortScience 33:1243– 1246

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Res 18:6531-6535

Wu B, Gao L, Gao J, Xu Y, Liu H, Cao X, Zhang B, Chen K (2017) Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). Front Plant Sci 8:389

Wu B, Cao X, Liu H, Zhu C, Klee H, Zhang B, Chen K (2019) UDP-glucosyltransferase PpUGT85A2 controls volatile glycosylation in peach. J Exp Bot 70:925-936

Xiao C, Hu J, Ao YT, Cheng MX, Gao GJ, Zhang QL, He GC, He YQ (2016) Development and evaluation of near-sogenic lines for brown planthopper resistance in rice cv. 9311. Sci Rep 6:38159

Xu Q, Xu X, Shi Y, Qi X, Chen X (2017) Elucidation of the molecular responses of a cucumber segment substitution line carrying Pm5.1 and its recurrent parent triggered by powdery mildew by comparative transcriptome profiling. BMC Genom 18:21

Xu Y (2010) Molecular plant breeding. Wallingford: CABI, UK, pp 734

Yamamoto T, Shimada T, Imai T, Yaegaki H, Haji T, Matsuta N, Yamaguchi M, Hayashi T (2001) Characterization of morphological traits based on a genetic linkage map in peach. Breed Sci 51:271–278

Yamamoto T, Mochida K, Hayashi T (2003) Shanhai Suimitsuto, one of the origins of Japanese peach cultivars. J Jpn Soc HortScience 72:116–121

Yamamoto T, Yamaguchi M, Hayashi T (2005) An integrated genetic linkage map of peach by SSR, STS, AFLP and RAPD. J Jpn Soc HortScience 74:204–213

Yazbek M, Oh SH (2013) Peaches and almonds: phylogeny of *Prunus* subg. *Amygdalus* (Rosaceae) based on DNA sequences and morphology. Plant Syst Evol 299:1403-1418

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J 66:182-193

Yoon J, Liu D, Song W, Liu W, Zhang A, Li S (2006) Genetic diversity and ecogeographical phylogenetic relationships among peach and nectarine cultivars based on simple sequence repeat (SSR) markers. J Am Soc Hortic Sci 131:513-521

Yu Y, Fu J, Xu Y, Zhang J, Ren F, Zhao H, Tian S, Guo W, Tu X, Zhao J, Jiang D, Zhao J, Wu W, Wang G, Ma R, Jiang Q, Wei J and Xie H (2018) Genome re-sequencing reveals the evolutionary history of peach fruit edibility. Nat Commun 9:5404

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zhang G (2021) Target chromosome-segment substitution: A way to breeding by design in rice. The Crop Journal 9:658-668

Zhou Y, Zhou H, Lin-Wang K, Vimolmangkang S, Espley RV, Wang L, Allan AC, Han Y (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