U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Marker assisted selection, fine mapping and identification of candidate genes for three major traits of *Prunus persica* L. (Batsh)

Carolina Ribeiro Campos

Internship Report

MSc in Agricultural Engineering

Department of Geosciences, Environment and Spatial Planning's

2020

Supervisor

Susana Maria Pinto de Carvalho, Assistant Professor, Faculdade de Ciências da Universidade do Porto

Co-supervisors

Iban Eduardo Muñoz, Researcher, IRTA/Crag

Naveen Kalluri, Ph.D Student, Crag

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Acknowledgments

At the end of another stage, I would like to thank all the people that supported me since the beginning of my master's degree and through the development of this dissertation.

First, I want to address my words of acknowledgment to my supervisor, Prof. Susana Carvalho and to Prof. Jaime Motos Ramos for sharing their love for the agriculture in every class, for the inspirations as teachers and for all the guidance and support. Without them the opportunity to realize this work never would have arisen.

To my co-supervisors, Iban Eduardo Muñoz and Naveen Kalluri for the opportunity to work with them, for receiving me so well in their team, for the support, dedication, time spend explaining again when a didn't understand something, availability for help anytime and of course for all the knowledge shared. To all the team members Nathalia, Neus and Filipe for the spectacular work environment, and to Pere Arús for the opportunity to share knowledge with such an important researcher. I also want to thank to IRTA, Crag and all the people from the laboratory, for all the sympathy, and availability for help.

To the Faculdade de Ciências da Universidade do Porto and the Erasmus + team, for the scholarship and accompaniment during my stay in Barcelona. To all the teachers of the Master in Agricultural Engineering, for sharing knowledge and tireless monitoring and to all my classmates for the friendship and team spirit over these years.

Lastly, to all my family, for supporting all my dreams and decisions and for making it possible to carry out this work even away from them in such difficult times. To my boyfriend Miguel, for all the love, support and companionship all the times and trough this journey.

Thank you all.

Abstract

Prunus persica L. (Batsh), is an economically important temperate fruit crop and one of the model species of the Rosaceae family. Peach breeding programs include different objectives as prolonged shelf life, fruit quality and disease resistance. One of the main limitations is the genetic basis of the commercial cultivars being very narrow. Currently the main sources of new alleles for breeding programs are the more variable ancient oriental cultivars, the landraces, wild species and the close cross compatible relatives, such us almond [Prunus dulcis (Miller) D. A. Weeb]. From the interspecific cross between 'Texas' almond x 'Earlygold' peach, eight major genes have been identified and mapped. Two of them, almond fruit type (Alf) and juiciness (Jui), define the main differences between almond and peach fruits. Other interesting gene was blood flesh (DBF2). In the first chapter of this report, we fine mapped them by saturating target regions with more markers developed from the sequences of the parental lines, and by identifying new recombinant individuals. The aim was to reduce the target genomic regions and the list of candidate genes for each gene to facilitate functional validation. We were able to reduce the genomic region of *Jui* and *DBF2* to sixty-nine and one point four Kb respectively, and the number of candidate genes to eleven and one. For *Alf* gene we identified twelve new possible recombinants but we have to wait to phenotype them to reduce the target region.

In the second chapter we compared three DNA extraction methods in thirteen different species: the alkaline lysis method used in the first chapter for the DNA extraction from the T1E population; the standard CTAB method used worldwide; and the fast and inexpensive DNA extraction method (SILEX) applicable to a high range of plant species and tissues, based on the CTAB method with a DNA silica matrix recovery. In this approach we concluded that the CTAB extraction method continues to be the one that provides the best results for the majority of the applications.

Key-words: almond fruit type, blood flesh, juiciness, molecular markers, screening.

Resumo

O pessegueiro *Prunus persica* L. (Batsh), é uma árvore de fruto de clima temperado economicamente importante e uma das espécies modelo da família das Rosáceas. Os programas de melhoramento do pessegueiro incluem diferentes objetivos como prolongar a vida útil, qualidade e resistência a doenças dos frutos. Uma das principais limitações dos programas de melhoramento do pessegueiro é a base genética das cultivares comerciais ser muito estreita. Atualmente, as principais fontes de novos alelos para programas de melhoramento são as cultivares orientais mais antigos, as *landraces*, as espécies selvagens e os parentes compatíveis, como a amendoeira [Prunus dulcis (Miller) D. A. Weeb]. Do cruzamento interespecífico entre 'Texas' amêndoeira x 'Earlygold' pessegueiro, onze genes principais foram identificados e mapeados. Dois deles, fruto tipo amêndoa (*Alf*) e suculência (*Jui*), definem as principais diferenças entre os frutos da amendoeira e pessegueiro. Outro gene interessante identificado foi a polpa do fruto vermelho cor de sangue (*DBF2*). No primeiro capítulo deste relatório, mapeou-se com precisão os três genes, saturando as regiões-alvo com marcadores moleculares desenvolvidos a partir das sequências das linhagens parentais e identificando novos indivíduos recombinantes. O objetivo foi reduzir as regiões genómicas alvo e a lista de genes candidatos para cada gene para facilitar a sua validação funcional. Reduzimos a região genómica de *Jui* e *DBF2* para sessenta e nove e um ponto quatro Kb, respetivamente, e o número de genes candidatos para onze e um. Para o gene *Alf* identificámos doze novos possíveis recombinantes, mas temos que esperar para conhecer o seu fenótipo e assim reduzir a região alvo.

No segundo capítulo, comparamos três métodos de extração de DNA em treze espécies diferentes: o método da lise alcalina usado no primeiro capítulo para a extração de DNA da população T1E; o método padrão CTAB usado mundialmente e o método de extração de DNA rápido e barato (SILEX) aplicável a uma grande variedade de espécies de plantas e tecidos, baseado no método CTAB e através da recuperação de uma matriz de sílica de DNA. Nesta abordagem concluímos que dos métodos testados o método de extração CTAB continua a ser o que fornece os melhores resultados para a maioria das aplicações.

Palavras-chave: Fruto tipo amêndoa, marcadores moleculares, polpa cor de sangue, seleção suculência.

Index

List of tables

List of figures

Abbreviations

ABI – Applied Biosystems

AFLP – Amplified fragment length polymorphism

Alf – Almond Fruit type

ALMELO – Almond × peach population (under development)

BC – Backcross population

bp – Base pair

cm – Centimeters

cM – Centimorgan

CRAG – Center for research in agricultural genomics

CTAB – Cetrimonium bromide

DBF2 – Blood flesh

DNA – Deoxyribonucleic acid

dsDNA – Doble strain DNA

EST – Expressed sequence tag

FAO – Food and Agricultural Organization of the United Nations

g – Grams

GBS – Genotype by sequencing

HAL – Histidine ammonia-lyase

InDel – Insertion or Deletion

IRTA – Institut de reserca I tecnologia agroalimentaries

Jui – Juiciness

Kb – Kilo base

Kg – Kilograms

MAI – Marker assisted introgression

MAS – Marker assisted selection

µl – Microlite

ml – Milliliters

ng – Nanograms

NGS – New generation sequencing

°C – Celsius degrees

PAP – 3'(2') - phosphoadenosine 5'-phosphate

PACE – PCR Allele Competitive Extension

PCR – Polymerase chaine reaction

Pp – Peach chromosome

QTL – Quantitative trait *locus*

QTLs – Quantitative trait loci

RAPD – Random amplified polymorphic DNA

RFLP – Restriction fragment length polymorphism

RNA – Ribonucleic acid

SILEX – Silica matrix Extraction

SNP – Single nucleotide polymorphism

SSR – Simple sequence repeat or microsatellite

ssDNA – Single strain DNA

STR – Short tandem repeats

T×E – ‘Texas’ × ‘Earlygold’ population

T1E – (‘Texas’ × ‘Earlygold’) × ‘Earlygold’ population

THESEUS – Receptor-like protein kinase

UDP – Uridine 5'-diphospho-glucuronosyltransferase

UTR – Untranslated region

UFGT – UDP-glucose flavonoid 3-O-glucosyltransferase

1. INTRODUCTION

Peach, *Prunus persica* L. (Batsh), is an economically important temperate tree fruit and one of the model species for the Rosaceae family [1]. Its genome has been sequenced [2] and recently a new version was released (Peach v2.0.a1 (v2.1)), and important information on the genetics of major genes and Quantitative Trait Loci (QTLs) was obtained in the last two decades [3,4]. Very early on, geneticists, realized that the genetic basis of the commercial cultivars was very narrow [5], later confirmed with molecular markers [6,7]. The main sources of new alleles for breeding programs were the more variable ancient oriental cultivars, the landraces and the close cross compatible relatives of peach, such us almond [*Prunus dulcis* (Miller) D.A. Webb] and several wild species (*Prunus davidiana* Carr., *Prunus mira* Koehne and *Prunus kansuensis* Rehder) [6,8]. Interspecific hybrids between peach and these species have been used for rootstock development over the years [9]. Although they have been used in some genetic analysis for fruit quality [10] or tree architecture [11], there is no evidence of a single example of a peach cultivar currently under commercial cultivation which contains beneficial genes from other species.

Some years ago, an ‘almond × peach’ (TxE) backcross 1 progeny was developed at IRTA to obtain a high-density linkage map for Prunus that became its reference map [12]. This map provided the scientific community with a common terminology and orientation for linkage groups and a large set of transferable markers that were used as anchors for constructing other maps [3] and to align the physical map assembled for the construction of the whole genome sequence [2].

Another interspecific population between almond and peach was developed and analyzed with the objectives of understanding the genetic variability of 42 phenotypic traits for flower, phenology, fruit quality, leaf and disease resistance, exploring possible alleles from almond that could be introgressed in peach commercial cultivars. In total, eight major genes and thirty-two QTLs were detected, with consistent behavior over the years. These new alleles identified from almond were responsible for important traits such as, blood flesh or powdery mildew resistance and were considered useful for the introduction of new variability into the peach gene pool [13]. In 2017 an approach was created for the introgression of almond alleles in peach, and propose a basic scheme, termed marker-assisted introgression (MAI), that allows a first survey of the variability that could be introduced from almond to peach, identifying the map positions of major genes dominant for the almond or additive, and plants with a desired almond chromosome fragment in the peach background to be obtained only two generations after the interspecific hybrid [14].

Some of the major genes identified and mapped by Donoso et al. (2016), were Almond Fruit type (*Alf/alf*), Juiciness (*Jui/jui*) and Dominant Blood Flesh (*DBF2/dbf2*) (Figure 1). The first two traits explain fundamental aspects that define the fruit of peach with respect to that of almons: *Alf* and *Jui*, for its thick and juicy mesocarp, respectively. *DBF2* is involved in the synthesis of red anthocyanin pigments in cells of the flesh.

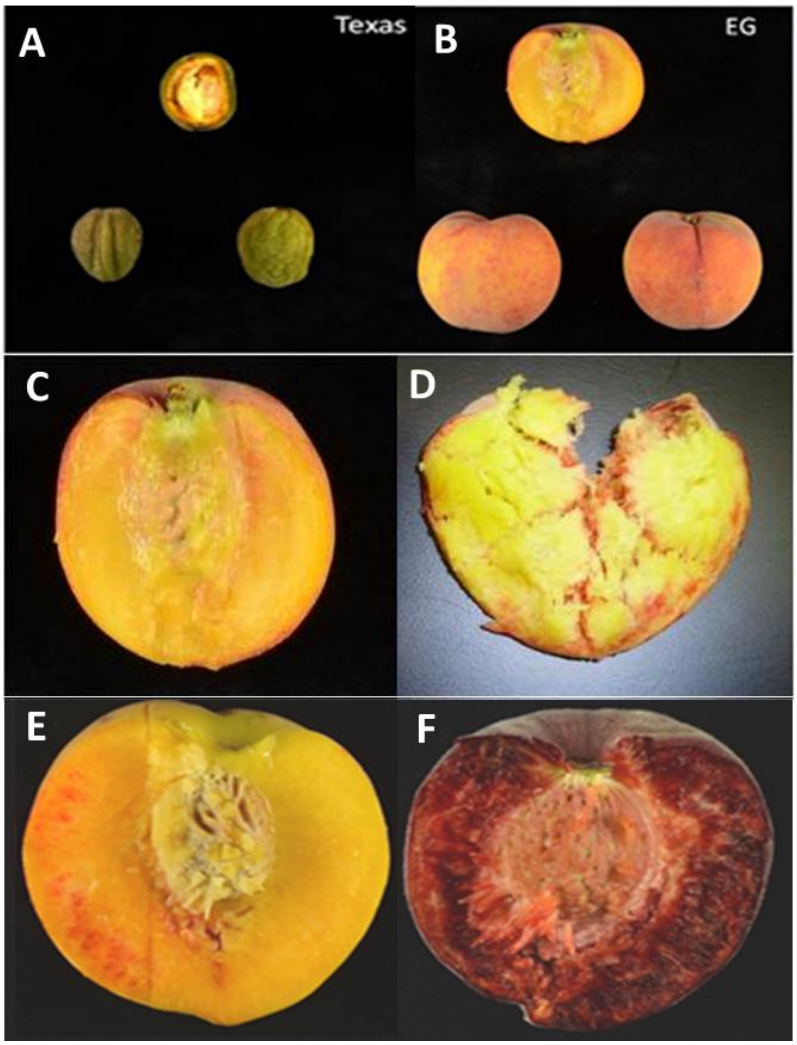

Figure 1. Photos of the phenotype produced by the three major genes under study. 1A: 'Texas' (alf/alf); 1B: 'Earlygold' or (Alf/alf) or (Alf/Alf); 1C: Juiciness (jui/jui); 1D: Non juicy (Jui/jui) or (Jui/Jui); 1E: Without blood flesh (dbf2/dbf2); 1F: Blood flesh (DBF2/dbf2) or (DBF2/DBF2).

In 2019 twenty genetic markers including SSRs, Indels and SNPs were used for a marker assisted selection (MAS) of segments of almond's genome into a peach genetic background, specifically in the regions where these major genes are mapped and a fine mapping was done. In total thirteen new recombinants and 21 putatively recombinants for *Jui* plants were found after fine mapping these 20 genetic markers. Their phenotype allowed to narrow down the regions of *Alf* and *DBF2* also.

Before this work was started, *Alf* was inserted in a region of 186 Kb with 29 candidate genes, *Jui* had a region of 132 Kb and 37 candidate genes, and *DBF2* was mapped in a region of 377 Kb, and had an annotation of 13 candidate genes.

In the present work, the same genetic markers as in the previous work were used for a MAS of *Alf*, *Jui* and *DBF2*. Fine mapping was done in these regions to detect any recombination event, near to where these genes lay, that could produce a distinguishable phenotype when the trees mature and help this way to trace down the position of these genes by crossing the phenotypic and genotypic data of each individual.

A second experiment was carried out during the development of this study, with the objective of improving the methodology used for the DNA extraction of peach and other crop species under research in the laboratory were the work was caried out.

The extraction of good quality DNA with a high yield is a limiting factor in plants' genetic analysis. DNA quality from each line should be good to allow a proper genetic analysis, especially for approaches using NGS techniques. High quality of DNA is characterized by predominantly high molecular weight fragments with an A260/280 ratio between 1.8 and 2.0 and the lack of contaminating substances, such as polysaccharides and phenols [15].

Pure and rapid DNA extraction is a pre-requisite for most advanced techniques such as genetic mapping, fingerprinting, marker-assisted selection, and for evaluating authenticity of exported varieties. The extraction of high-quality DNA from plant tissue is time consuming, arduous, and costly due to multiple steps and the high cost of liquid nitrogen. In addition, the problems associated with the available commercial kits are their high cost and low yield of DNA [16,17]. Several methods to isolate DNA from plant tissues are available; however, these methods produce either small amounts or DNA of inconsistent quality. DNA quality also depends on the application, for example the quality needed for SSR genotyping is not the same than for GBS (genotyping by sequencing), therefore we need different methods according to the final application.

Thirty years have passed since the description of CTAB (cetyltrimethylammonium bromide) protocol from Doyle and Doyle (1990). At that time the protocol had the purpose of getting high molecular weight DNA suitable for digestion with restriction endonucleases. The main problem was finding a procedure that worked with not only a group of plants but for many different species. Over the years the CTAB method for fresh tissues was been used successfully in many plant species, tissues and laboratories worldwide [18–21], and have been adapted to plates of 96-wells and to atomatized robots to increase the number of DNA extractions that can be done in a short period of time. Nowadays most of the DNA extraction methods are modified versions of CTAB extraction with some crop-to-crop limitations and

differ in time and cost. The main cause of the differences in the CTAB protocol is the composition of cell walls and intracellular components such as nucleus mitochondria and cellulose. CTAB is a cationic surfactant added in the DNA extraction buffer, which dissociates and selectively precipitates DNA from histone proteins [22]. Several modifications of this protocol have been implemented in order to minimize contamination by other compounds of specific tissues of species [23–25]. These modifications, apart from being species or tissue-specific and frequently not removing completely interfering compounds, are time-consuming due to many handling steps, and thus are not suitable for high-throughput applications [26,27].

One example of a DNA extraction protocol developed from the CTAB is the SILEX method, which is faster and inexpensive, and applicable to a wide range of plant species and tissues developed in the present year[28]. This approach emerged because commercial kits based on silica matrices avoid many of these issues by optimizing the conditions in which only DNA can bind to the silica surface. Therefore, contaminants such as polysaccharides, polyphenols and proteins can be easily removed [29]. They also tend to be faster than the standard CTAB protocol, being the preferred option for sequencing studies in which many samples must be evaluated [30,31].

Another method tested in this section, is the alkaline lysis [32]. This methodology was tested in rice, wheat, and maize and rapeseed and involves only four simple steps for the DNA extraction. This method was also used in the first chapter of this report and was very efficient, fast and low-cost.

This way, the main objectives of the current study were:

- Identification of new recombinants for the fine mapping of important traits (Alf, Jui, DBF2) identified in interspecific crosses between peach and almond, and track down their specific position;
- Reduce the number of candidate genes for Alf, Jui and DBF2.
- Compare the concentration and quality of the DNA extracted through three different DNA extraction, for some of the main plant species under study in the different groups of CRAG laboratory.

2. LITERATURE REVIEW

2.1. Peach

2.1.1. Taxonomy and phylogeny of the genus Prunus

The Rosaceae family is an important plant family with approximately 90 genera and 3000 species which includes a large number of economically important crops and ornamental species [33][14]. Among them, in addition to ornamental species such as rose and hawthorn, are listed some of the most important fruit species like apple, pear, almond, apricot, plum, cherry, strawberry, blackberry and raspberry. The most recent classification for the family [34] divides it into three sub families: *Rosoideae*, *Dryadoideae* and *Spiraeoideae,* although the name *Spiraeoideae* was corrected in 2011 to *Amygdaloideae* based on changes in the International Code of Nomenclature for Algae, Fungi and Plants [35], belonging precisely to this family the genus *Prunus.*

This genus comprises 200 species of trees and shrubs, some of them economically relevant fruit and nut crops. The infrageneric classification of *Prunus* by Rehder (1940) is the most accepted nowadays and consists in five subgenera: *Amygdalus* (peaches and almonds), *Cerasus* (cherries), *Prunus* (plums), *Laurocerasus* (evergreen laurel cherries) and *Padus* (bird cherries) [36] (Figure 2).

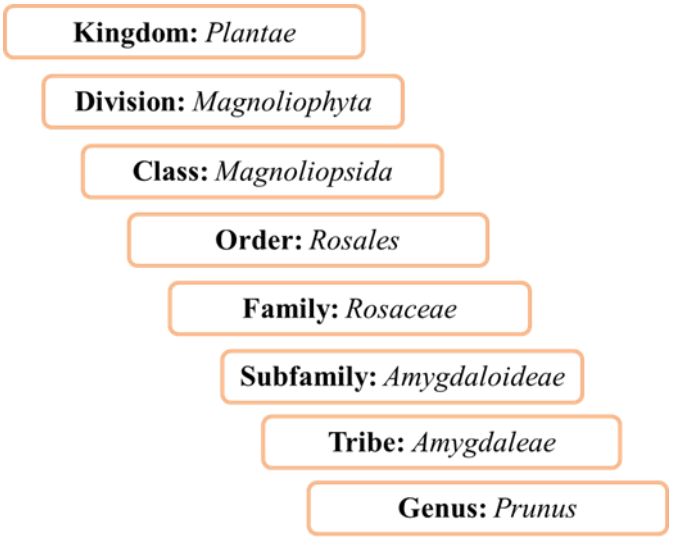

Figure 2. Systematic classification of Prunus genus.

Peach [*Prunus persica* (L.) Batsch] is a genetically well characterized model for research about Prunus species and other Rosaceae fruit trees. With a small size diploid genome (2n =

2x =16; 230 Mbp), and relatively short generation time (two to four years), peach has become a model species for fruit genetic studies [33]. It shares the *Amygdalus* subgenus with almond (*P. dulcis* (Mill.) D. A. Webb) and also with several wild relatives like *P. davidiana* (Carr.) Franch, *P. mira* Koehene, *P. ferganensis* (Kost and Rjab) Kov. & Kost and *P. kansuensis* Rehd [9,36].

All the previous peach relatives originate from China, as well as some regions of Nepal, India and other peripheral countries. The genetic proximity of peach and the aforementioned species makes it possible to produce fertile hybrids between them. Other *Prunus* distantly related may also be hybridized with peach, but produce generally sterile hybrids [37] (Figure 3).

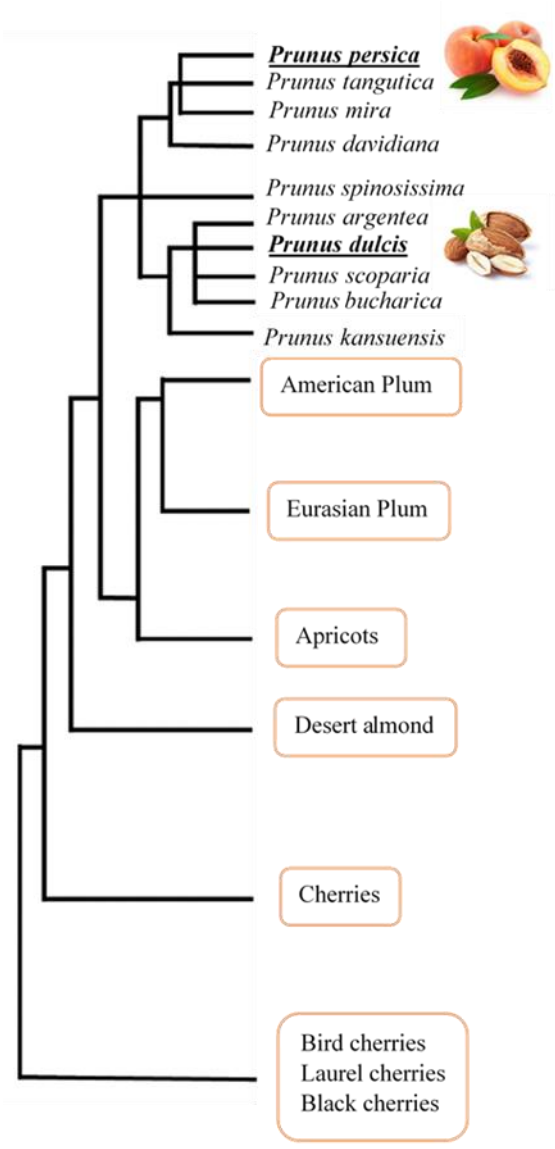

Figure 3. The phylogeny of Prunus based on the plastid DNA sequences of four genes. Adapted from: Chin et al. (2014).

The majority of the interspecific crosses between almond and peach were performed in the past mainly for the development of rootstocks for peach. Interspecific crosses between almond

and peach were historically done also for rootstock development, for use primarily in calcareous soils, since these hybrids often tolerate iron chlorosis. In the last years a growing interest has emerged in the use of these related species for peach breeding and almond has become an interesting choice for introgressing new genes into peach, mainly due to the high variability of the species [13].

2.1.2. Origin and dissemination

Domesticated peach in China was dispersed westward to Europe via the ancient Silk Road through Persia (present day Iran) in the final centuries B.C, and from Europe to the Americas during the 16th century [38]. Today, peach is widely cultivated in temperate and subtropical zones throughout the world [39] (Figure 4).

Figure 4. Geographic distributions of the Amygdalus accessions and the dispersal route of domesticated peach (shown with white dotted lines). Source: Yu, Y. et al. (2018).

The direct wild ancestor of peach remains unknown, and is likely extinct, so there is controversy about the origin and evolutionary history of this fruit. It has been proposed that peach was originally domesticated in northwest China around 4000-5000 years ago [38] while, fossil evidence indicates that peach cultivation and domestication could date back to at least seven thousand and five hundred years ago in the Yangtze River valley of southern China [40]. Surprisingly, peach endocarp fossils from 2.6 million-year-ago (Mya) found recently in Kunming, southwest China, are indistinguishable from endocarps of the modern peach cultivars, and studying of these fossils suggested that peaches may have acquired their modern-like edible fruits long prior to domestication, perhaps mediated by frugivorous primates [41]. It is therefore possible that there was a very long period of pre-selection that

occurred in natural environments that may have enabled the development of edible fruit (fleshy and palatable mesocarp) long before outward dispersal for cultivation and domestication [39].

In consequence, Chinese germplasm and local varieties (i.e. varieties not obtained in breeding programs) may constitute the main source of diversity for modern occidental breeding programs [42].

2.1.3. Economic importance

Peach is one of the most important fruits worldwide. The production of peaches, nectarines and flat fruits has doubled in the last two decades, mainly due to more efficient agricultural practices and the development of new varieties and rootstocks well adapted to specific climate conditions. The overall world production of peach reached 24 million tons in 2018 (FAOSTAT, 2020). China is the top producer, with more than half of world's total production in 2018 (≈15M tonnes), followed by Italy (≈1.1M tonnes), Greece and Spain (≈0.9M tonnes) (Figure 5).

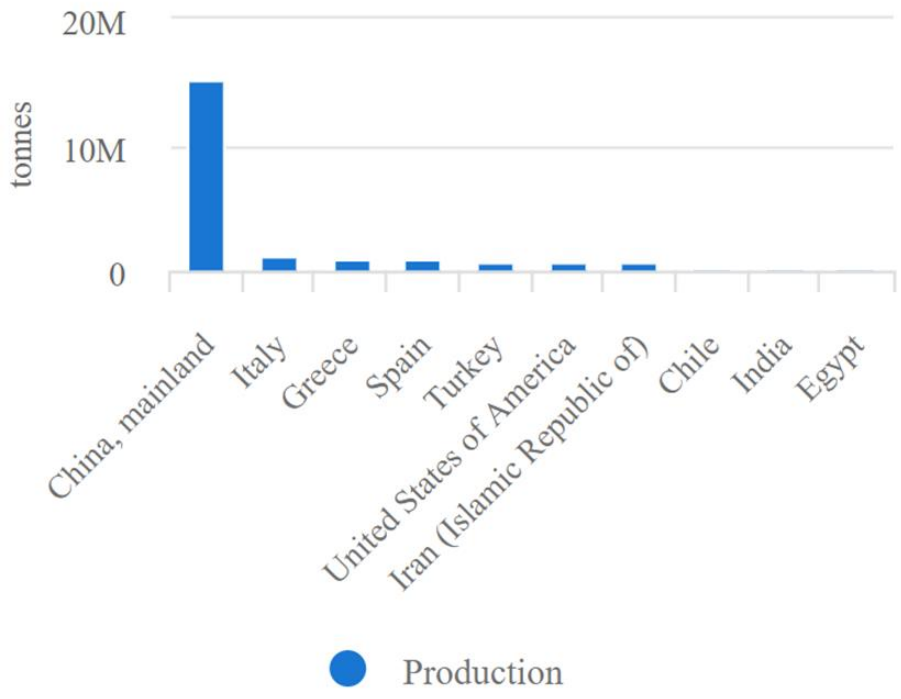

Figure 5. The 10 most important producers of peach and nectarines in the year 2018. Source: FAOSTAT (2020).

In Spain the peach tree, boasting 81,300 ha of cultivated land, is the most important sweet fruit species and at the same time the most exported. With an average annual production for the period of 2015-2017 of 1,431,000 Spain is the third larger producer in Europe and the leading exporter worldwide [43].

In Portugal, the peach and nectarine orchards are mostly located in the inland center region. In the years of 2015/2016 peach production reached 46,000 tones and is expected to remain stable in the years to come [44].

2.1.4. Genetic characterization and molecular breeding

Genetically, peach has long been one of the best characterized species in the Rosaceae family [3], and its considered, along with *Malus* × domestica Borkh and *Fragaria vesca* L. a model specie for the development of genetic studies, due to several advantageous characteristics [33,45,46]. Compared with the other two model species (Fragaria *vesca* and Malus × domestica Borkh), and the genome of *Prunus mume* [47]. Further genetic advantages of peach are the absence of a self-incompability system, which permits the creation of F2 populations, and the short intergeneration period of two to four years that is much lower than the majority of fruit tree species (five to ten years). These characteristics facilitate and accelerate the generation of segregating populations for genetic mapping studies.

The current scientific community has several tools for the genetic and genomic study of peach. The advantageous characteristics of the species together with the genetic studies carried out mainly in the USA and Europe during the 20th century and the efforts of different countries and research groups in the USA, Spain, Italy, Chile and France, culminated in the sequencing and release of the species genome [2]. The integration of all available genetic and genomic information triggered the beginning of the post-genomic era, which is characterized by the availability of a complete genome and new DNA and RNA sequencing technologies. Thus, we are facing a revolution in the use of new high-performance analysis techniques, which may indicate a shift in the scientific paradigm where the identification of genomic regions of interest and their subsequent application in breeding are concerned [48].

2.2. Genetic Markers

2.2.1. History, improvements and classification

Genetic markers are biological characteristics established by the genetic variants between an individual, organism or species and, if they are located in genes or are closely linked to them, can be used as ‘signs’, ‘flags’, ‘probes’ or ‘tags’ of such a gene. The first genetic markers used were the morphological (‘classical’ or ‘visible’) markers which themselves are phenotypic characters or variants. These were the ones used in the early plant breeding. The first biochemical markers used were the isozymes, i.e. the genetic variants of a specific enzyme.

The utility of such markers was limited due to their small numbers of potential marker loci, low levels of polymorphism between closely related individuals and their not always consistent expression [49,50]. Thus, future prospects go through the control of biotechnology as a fundamental condition to obtain a greater probability of success in crop improvement [51]. Within biotechnology, the study and use of DNA markers for plant breeding provide an encouraging picture [52]. It should not be forgotten that many of the characteristics that have been pointed out and that concern the agricultural sector, such as resistance to pests or productivity, are genetically determined. The regions of the genome in which the genes associated with a particular quantitative trait are located are called QTLs, quantitative trait loci [53].

The use of DNA markers associated with important agronomic factors is widespread in the improvement of various types of crops and are also used globally to optimize efficiency in the production of other types of food, such as vegetables and pastures [54–56]. To this end, new approaches due to the increasing availability of data provided by the sequencing of complete genomes and transcriptomes are fundamental results. In fact, the complete genome of many species with agronomic interest such as rice [57] or tomato [58] already exists. These new technologies offer a large amount of genomic sequences in a short period of time and do so at a low price [59]. Thus, genetic improvement is expected to benefit from this new circumstance and optimize both the efficiency and accuracy of the whole process [54].

As mentined before molecular markers have been used in recent years in the agronomic sector as powerful tools for the analysis of genetic variation as they offer an efficient way of linking phenotypic and genotypic variation [60,61]. However, not all markers are equally valid. The characteristics that a good marker has to fulfil will depend, to a large extent, on the size and composition of a plant population and the number of genes segregating in a population. However, in any case, all molecular markers analysis techniques must meet the following criteria: (1) reliability, since molecular markers should be very close to an investigated locus. Once the results are improved using several markers if they are flanking at a loci or intragenic; (2) highly polymorphic, to discriminate between different genotypes, and to be evenly distributed in the genome; (3) a simple, cheap, and fast technique; and (4) requiring very little genetic starting material to carry out the analysis [54].

DNA markers can be classified based on: (1) the methodology for their detection, which can be southern or hybridization-based, polymerase chain reaction (PCR)-based and DNA sequence based; (2) their dominant or codominant polymorphism and (3) their location in relation to a gene, in which they can be classified into random molecular markers (anonymous or neutral markers), gene targeted markers and functional markers. Random markers are

distributed all across the genome while gene targeted markers are found within genes not necessarily involved in phenotypic variation, e.g. un-translated regions (UTRs) of EST sequences. Functional markers are located in the polymorphism casually associated with a phenotypic trait variation, so they are totally linked to the allelic forms in the locus and the functional motifs. Random and gene targeted markers can be used to tag functional variations if QTL studies establish an association between marker and trait, however and unlike functional markers, the association can be broken through recombination [50].

The choice of one DNA marker over another will depend on the research's goal. A comparison between the most widely-used DNA markers is shown in table 1.

Table 1. Comparison of the five most widely used DNA markers in plants. A. Mutation at enzyme restriction or PCR priming site, B. Insertion or deletion between enzyme restriction or PCR priming sites, C. Change of tandem repeat units between enzyme restriction. Source: López Girona (2014).

Characteristics	DNA markers				
	RFLPs	RAPDs	AFLPs	SSR	SNPs
Methodology	Southern blot	PCR	PCR	DNA sequence	DNA sequence
Molecular basis	A,B	B	A,B	C	D
Genomic Coverage	Low copy coding region	Whole genome	Whole genome	Whole genome	Whole genome
Inheritance	Codominant	Dominant	Dominant	Codominant	Codominant
Polymorphism	Medium	Medium	High	Very High	Medium
No. Loci / marker	1 to 3	1 to 20	10 to 100	1 to 2	1
No. Alleles / locus	Multiallelic	2	2	Multiallelic	2
DNA quantity (ug) / reaction	5 to 10	0,02	0,5-1	0,05	0,05
DNA quality	High	Moderate	High	Moderate	High
Reproducibility	Very High	Moderate	High	Very High	Very High
Cost development / analysis	High/High	Low/Low	Low/Low	Low/Moderate	High/Low
Automation	Low	Medium	High	High	High
Suitable utility in deversity, genetics and breeding	Genetics	Diversity	Diversity and Genetics	All purposes	All purposes
Type of probs/primers	Low copy DNA or cDNA clones	Usually 10 bp random nucleotides	Specific sequence	Specific sequence	Allele-specific PCR primers
Effective multiplex ratio	Low	Medium	High	High	Medium to high

In peach, RFLPs, PAPDs and AFLPs markers have been used for genetic diversity studies [62–64], for synteny studies [27,65,66], for cultivar identification [67–69] and for construction of linkage maps [45,48,70–73].

The evolution of main keywords in articles related to molecular markers from 2000 to 2016 proves that nowadays the SSR and SNPs markers are the most widely used and were the ones used in the present research work (Figure 6).

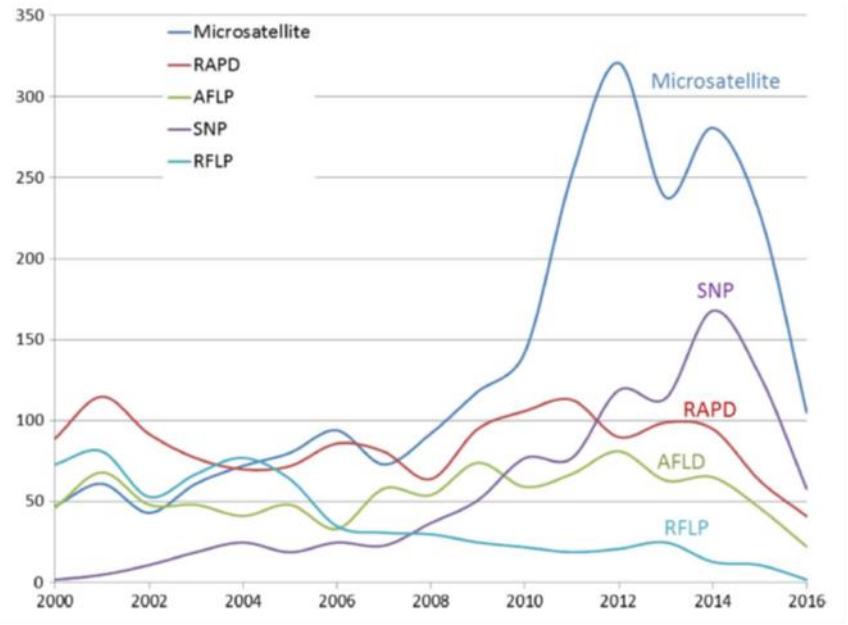

Figure 6. Evolution of main keywords in articles related to molecular markers from 2000 to 2016. Source: Garrido-Cardenas et al. (2018).

2.2.2. Microsatellites or SSRs, simple sequence repeats

Microsatellites, also known as short tandem repeats (STRs) or as simple sequence repeats (SSRs), are repeats of up to 100 times of simple sequences of 1–8 base pairs [74]. These elements are present in both coding and non-coding regions of all eukaryotic and prokaryotic genomes studied to date, even being present in chloroplast and mitochondrial DNA [75,76].

Among individual genotypes, the number of repeat units may vary since the tandem arrays of SSR motifs change. Accordingly, with additional repeated units, the genotypic variety also increases. Likewise, motif length also affects the number of repeats as shorter motifs contain a higher number of repeats than larger ones (e.g., tetranucleotide). Notwithstanding, in smaller motifs, there is a greater feasibility of genotyping errors due to slipped-strand mispairing (stuttering) during the polymerase chain reaction (PCR), while longer and perfect SSR loci display more prominent allelic fluctuation [77,78].

Microsatellites have been utilized liberally over previous years since they are profoundly informative with a high mutation rate per locus per generation (10^{-7} to 10^{-3}) [77], locus have specificity, high intraspecific polymorphism, high reproducibility and ease of scoring, are multiallelic, and frequent transpacific presence across related taxa. Additionally, the codominance nature of SSRs allows for the direct measurement of heterozygosity and only requires small amounts of DNA for data collection, another characteristic of SSRs (1 ng of DNA per reaction) [79–81]. Notably, they have been widely applied for different purposes, such

as (1) genetic diversity and (2) discovering QTLs [82]; (3) linkage map construction between gene and marker [83]; (4) MAS for desired traits[84]; (5) forensics and parentage analysis (SSRs with core repeats three to five nucleotides long are preferred) [85]; (6) cultivar DNA fingerprinting [86], (7) genome-wide association study (GWAS) [87]; (8) gene flow estimation and crossing over rates[88]; (9) haplotype determination [89]; (10) harnessing heterosis [90]; (11) germplasm characterization; and (12) genetic diagnostics, characterization of transformants, and the study of genome organization [84,91–94]. However, the high cost for SSR development, the presence of more null alleles, and the occurrence of homoplasy are some of the weak points of microsatellites [95].

The primers used in the PCR reactions for the analysis of microsatellites may be labelled with either a fluorophore, with a radioactive element or lacking labelling altogether. Depending on whether one option or another is used, the detection systems will be different and can be used from laser detection systems with automatic reading to simple agarose gels. The main advantages of this type of markers are both their large quantity and variety [96] as well as their co-dominant inheritance, which provides, in contrast to dominant markers, the complete genetic information. That is why they are probably the most widely used molecular markers in labs across the world [54]. They have also been used in different plants such as *Arabidopsis thaliana*, maize (*Zea mays*), soybean (*Glycine max*), rice (*Oryza sativa*) and wheat (*Triticum aestivum*) [97–99].

2.2.3. SNPs, Single-nucleotide polymorphisms

A single-nucleotide polymorphism is said to exist when a single-nucleotide change (A, T, C, or G) is observed by comparing the DNA of different members of a species. These changes in a single position are used as an effective genetic marker in practically all the studied species, both animal [100] and vegetal [101], due to its great abundance, and its importance has become remarkable in the genetic analyzes in recent years. Due to their characteristics, they are extremely useful in a multitude of analysis, being able to evaluate a large number of loci and discriminating efficiently between homozygous and heterozygous alleles. In addition, SNPs are homogeneously distributed throughout the genome, they have low mutation rates, and they show high heritability, making them ideal markers. Depending on the type of mutation that occurs, the SNPs can be classified into: (1) transversions, with changes in nucleotides C/G, A/T, C/A, and T/G; (2) transitions, appearing C/T or G/A changes; and (3) indels, produced by insertions or deletions of a single nucleotide. In plants, thanks to the recent development of different molecular techniques such as massive sequencing [102], it has been

possible to design high-performance routine SNP analysis that allows for the study of thousands of positions at a time.

They are considered the ultimate form of molecular marker because a nucleotide base is the smallest unit of inheritance and they are the most abundant genetic marker in all organisms. The 90% of human genetics variation is due to SNPs, with one SNP every 100-300 base pairs [103]. SNP variability in peach is lower; with an estimated average of 1 SNP every 598 base pairs [104].

The selection of SNPs enables the selection of desired lines in large-scale populations. The marker can be used to modulate the cultivation program for the determination of the relevant feature and improvement of the crop more economically using new-generation technologies than using traditional methods [105]. Today, plant breeding is dependent on SNPs and similar differences for fast and cost-effective analysis of germplasm and feature mapping [106].

Because the desired trait is under genetic control, phenotypic experiments can be attempted faster, and the breeder not only does the early trait selection but also can transmit the desired allele to a large number of populations [107]. These sequences enable screening of more than 1 million SNPs for each species.

These polymorphisms can be used as simple genetic markers that can be identified around almost any gene. The usage of SNPs in detecting relationships between allelic forms of a gene and phenotypes, especially common diseases with multifactor genetics, high-resolution genetic map construction, linkage disequilibrium based association mapping, genetic diagnostics, genetic diversity analysis, cultivar identification, and phylogenetic analysis, creates great potential for characterization of genetic resources [108].

SNPs can also be used to discover new genes and their functions by affecting gene expression and transcriptional and translational promoter activities. Therefore, they may be responsible for phenotypic variations between individuals in improving agronomic features. It is also important to know the location of SNP in the genome, because if SNP is present in the coding region, it can greatly affect the activity and thermostability level of an enzyme or a similar product [109].

Some authors consider the InDels as SNPs [110], and its name indicates insertions or deletions (insertion-deletion) of nucleotide fragments of different sizes at the same site in the genome sequence between the same or closely related species and is a gap in sequence derived from alignment of the homologous sequence [111,112]. InDels are widely distributed across the genome and occur in high density and large numbers in a genome. The InDel polymorphic molecular marker is a PCR-amplified marker that is based on specific primers designed from

both sides of the site of sequence of an insertion or deletion. It is essentially a length polymorphic marker still, and electrophoresis can be used for genotyping [113,114]. InDel molecular markers have the advantage of high accuracy and good stability, which help to avoid confusion in subsequent analysis due to marker specificity and complexity, as is often seen in other length polymorphic markers. Furthermore, mixed or highly degraded DNA samples can be successfully amplified with InDel markers, and effectively typed. Because of its abundance, convenient typing platform and other advantages, InDel molecular markers have been applied to genetic analyses of animal and plant populations, molecular assisted crops and farmed animal breeding, human forensic genetics, medical diagnostics and other research areas. The development of the InDel molecular marker located on functional genes, combined with chromosome walking and fine gene mapping, has enabled the application of these molecular markers in the screening of genes related to important economic traits, which is conducive to the further development and utilization of these valuable genes [115].

2.2.4. Applications

2.2.4.1. Marker Assisted Selection (MAS)

Marker-assisted selection is the selection for a trait on the basis of genotype using associated markers rather than the phenotype of the trait. Molecular markers are usually incorporated in a variety of breeding applications, such as cultivar identification, assessing the genetic diversity and purity of a genus, among others [116]. It makes use of DNA-based markers that are tightly linked to the gene of interest to either assist or replace phenotypic evaluation. Identification of plants that carry specific genes or QTLs is based on their genotype and this is done by determining the allele of the molecular (DNA-based) marker [117]. The use of molecular marker information to identify and select specific genotypes also gives a clear understanding of the molecular marker. Before now, breeders used to rely on information about how plants or animals performed to assess whether or not their genes were of high enough quality, a practice that is time consuming, laborious and less effective [118–120]. A direct handle on the genes controlling the traits of interest could lead to much faster progress. [121,122].

For tree species like peach, whose intergeneration period can last up to four years, this selection step represents a severe delay on the creation of new varieties for both public and private breeding programs. Other limitations of traditional selection are the pyramiding of several alleles (i.e. for durable resistance to specific diseases) or the identification of genetically superior lines when the phenotypic variance is influenced by the environment [123].

Marker Assisted Selection (MAS) is especially important for traits that manifest late on a plant's life cycle, such as fruit traits [9]. In peach, the application of MAS is possible for a few Mendelian traits, mainly for low-acidity, skin pubescence, flesh color, stone adhesion, flesh texture and fruit shape [104] and also for pollen sterility and aborting fruit [12,124], slow ripening [125], resistance to green peach aphid [126] and skin color [127]. In addition to the monogenic traits already mapped for peach, several complex traits were also detected for the species in several segregating populations during last couple decades [3]. Despite the discovery of markers associated with these QTLs, further work is still necessary before integrating these into breeding programs [9]. Moreover, solutions should be found to other important shortcomings such as the lack of trained staff, or the need of simplified and cheap strategies to screen large progenies [128]. MAS also has its disadvantages, as said before this methodology uses molecular markers known to be associated with trait of interest or phenotypes to select plants with desirable allele effecting target trait. It is efficient only for those traits that are controlled by fewer numbers of quantitative trait loci (QTLs) having the major effect on trait expression, whereas for complex quantitative traits which are governed by large number of minor QTLs, the method is even inferior to conventional phenotypic selection [129]. To overcome this disadvantage, research communities were looking for solutions over decades how to deal with the complex traits and come out in the form of GS (Genomic selection). GS estimates the genetic worth of the individual based on large set of marker information distributed across the whole genome, and is not based on few markers as in MAS. The GS develops the prediction model based on the genotypic and phenotypic data of training population (TP), which is used to derive genomic estimated breeding values (GEBVs) for all the individuals of breeding population (BP) from their genomic profile [130]. GS is relevant because it has a grater accuracy in predictive terms for agronomic traits than MAS.

2.2.4.2. Marker Assisted Introgression (MAI)

Introgression simply means the transfer of a genomic fragment containing genes of interest from one species to another through hybridization and repeated backcrossing [131]. Introgression is defined as the process where a target gene or QTL from a plant in population 'A' is inserted to another plant in population 'B' by crossing the two of them and then repeatedly backcrossing to 'B' which is known as the recipient and/or recurrent parent. In this case, DNA markers are useful in controlling the presence of the target gene or QTL. This is also useful in accelerating the recovery of the background genome to the recipient type. Introgression using molecular markers is very effective in incorporating genes or QTLs from landraces, because

the time required to produce an improved variety and the issue of linkage drag are reduced [132,133].

In the year of 2016 this breeding strategy was proposed in order, to obtain lines of perennial species with a single introgressed fragment from a compatible species two generations after the interspecific hybrid. This strategy allowed the enrichment of the genome with genes from a wild or exotic relative in a short timeframe and with an intermediate step that allows for an initial exploration of genes/QTLs that the donor species can provide to the target crop. This new method involved three phases: (1) creating a large backcross one (BC1) population to select, with molecular markers, a reduced number of individuals (called prIL set) with a low number of introgressions; (2) phenotyping the prIL set for the traits of interest and inferring the inheritance and map position of segregating major genes/QTLs based on the known genotypes of the prILs; and (3) advancing selected lines carrying the traits of interest to a subsequent generation of backcross or selfing to obtain individuals with a single introgression in the background of the elite commercial germplasm. The proof of concept of this strategy was implemented by using peach as the recurrent species and almond as the donor [134], and the scheme of the MAI is showed in figure 7.

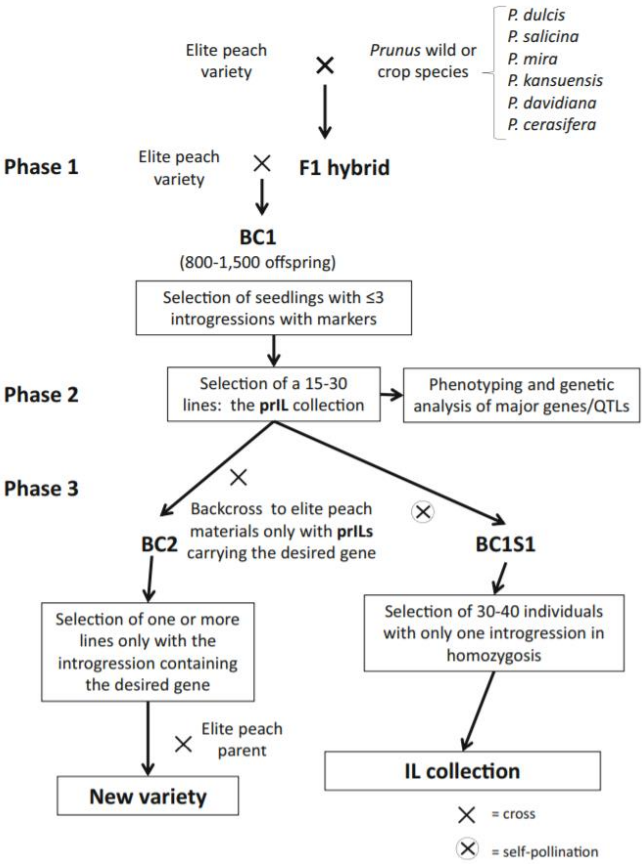

Figure 7. Scheme of the marker assisted introgression (MAI) strategy for the integration of exotic chromosome fragments into the elite peach genome. Source: Serra O. et al (2016).

2.2.4.3. Fine mapping

A genetic fine map of a specific locus will usually have as its goal the identification and location of markers that flank the targeted region of genome with replicable signals of association, with the aim of localizing the causal variants. In most cases, markers with positive association and causal polymorphisms are less than one centimorgan (cM) apart. In some cases, comparisons of map positions can be accurately made with results of other species, examining syntenic regions for defined causal polymorphisms and for similar traits.

Fine-mapping is a necessary step to identify causal polymorphisms and it involves the identification of recombinants with recombination's very close to the casual gene of the target trait, and saturating the region with more molecular markers. Nowadays, next generation sequencing techniques allow to resequence the genomic region of interest to identify all variants.

2.2.4.4. Predictive Breeding

Genomic prediction is an advanced form of marker-assisted selection in which genetic markers covering the whole genome are used so that all genes and QTLs associated with traits of interest are in linkage disequilibrium with the markers. Genomic prediction includes two steps: (1) train the prediction model using lines with both genotypic and phenotypic data as training sets and (2) predict breeding values using genotypic information of the unphenotyped lines, the testing sets, with the trained model and estimated parameters from step 1. Genomic prediction is becoming a widely adopted tool for breeding. It makes use of historical genotypic and phenotypic information to predict genotyped but not phenotyped lines for selection decisions. The advantage of genomic prediction is that the prediction accuracy will be improved with high quality phenotypic and genotypic data accumulated across years in the training sets. Accuracies of genomic prediction are affected by the quality of genotypic and phenotypic data and by an appropriate statistical model [135].

Genomic selection (GS) or prediction is referred as the simultaneous use of genome-wide markers to predict an individual's genotypic or breeding value for both observed and unobserved individuals. Multi-trait GS models can make use of information on correlated traits to improve prediction accuracy[136]. Early work on GS in plants was mainly focused on unobserved individuals, in the marker assisted recurrent selection (MARS) context. GS can be beneficial for observed individuals as well if entry-mean heritability is low[137].

To date, the development of high-throughput phenotyping systems for plants has largely focused on measuring traits of individual plants in greenhouses or growth chambers. However, many phenotypic responses of interest for crop improvement, especially those related to yield potential and abiotic stress tolerance, involve suites of traits that are best measured as expressed among communities of plants that grow in agronomically relevant edaphic and climatic conditions. Furthermore, field-based systems are more readily incorporated into applied plant breeding programs. Thus, there is growing interest in adapting agricultural machinery and electronic sensors for field-based high-throughput phenotyping [138,139]. Potential applications are mainly envisaged for genetic research (e.g. in detection of quantitative trait loci, QTL) and crop improvement but also include monitoring of the crop response to soil and management variability (i.e. precision agriculture).

Several decades-long investigations in molecular genetics and biochemistry of bacteria and viruses have allowed researchers to develop new methods of manipulating DNA through creation of various vector systems and tools for their delivery into the cell. All of these developments allow successful creation of not only transgenic microorganisms but also genetically modified higher organisms including various plant and crop species. Creation of novel tools for breeding and biotechnology, an application area of genetic engineering, has received significant focus resulting in accelerated development of useful tools. However, conventional genetic engineering strategy has several issues and limitations, one of which is the complexity associated with the manipulation of large genomes of higher plants [140]. Currently, several tools that help to solve the problems of precise genome editing of plants are at scientists' disposal, such as TALENs (transcription activator-like effector nucleases), CRISPR/Cas (clustered regularly interspaced short palindromic repeats) and genome editing with engineered nuclease (GEEN) [141].

3. MATERIALS AND METHODS

1st Experiment

3.1. Plant Material

From 2013 to date, several progenies of different generations derived from 'Texas' and 'Earlygold' crosses (Figure 8) were screened to search for recombinants to fine map *Alf*, *Jui* and *DBF2* genes.

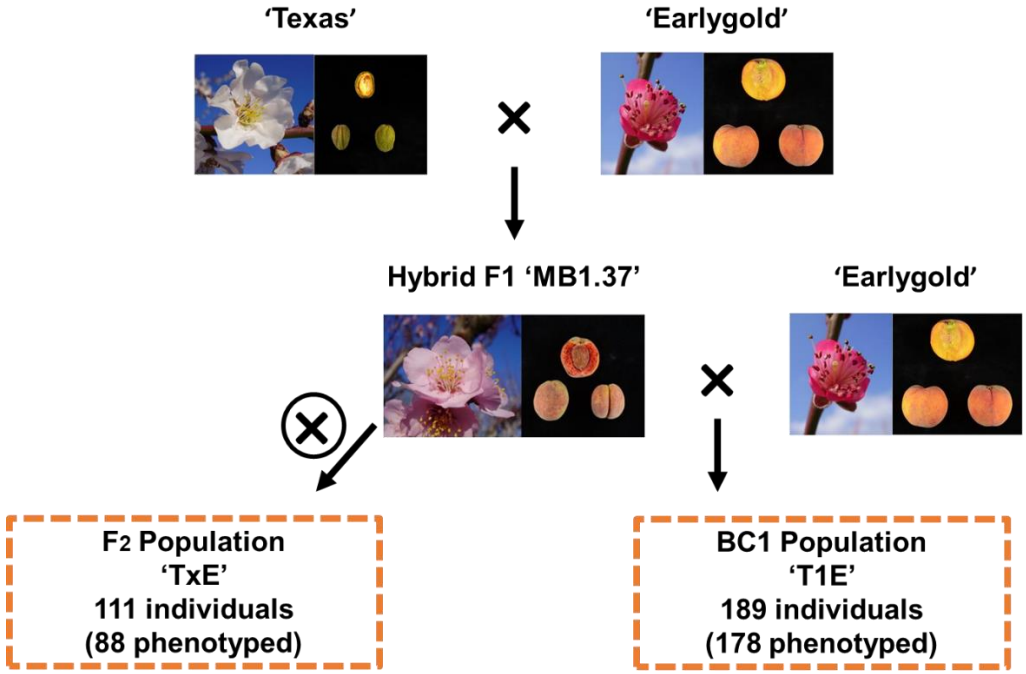

Figure 8. Schematic representation of the first progenies derived from 'Texas' × 'Earlygold' crosses [13], that originate the subsequent generation.

From the T1E population, around 25 individuals were selected based on their heterozygosity concerning the three regions of the study (the *Jui* and *DBF2* region in the group 1 and the *Alf* region in the group 4); thus, they possessed almond and peach alleles in at least two of the three central genes analyzed in this work and their maturity date (MD) is late.

The entire process of extracting the seeds was carried out in the laboratory in the most aseptic manner as possible. First, were opened the fruits with special scissors to extract the seed. After drying the seeds for 2 to 3 days on filter paper at room temperature they were stored at 4ºC. After extracting the seeds from all families, the stratification process was started by putting them for 12 weeks at 4ºC in trays, filled to half of its capacity with perlite (Europerl A-13) previously saturated with water. The trays were observed so as to discard contaminated

seeds and change the perlite if necessary. After stratification, the roots began to elongate and the seeds were sown in forest plastic trays with substrate (2: 1 Floragard TKS-1 seed, with macro and micronutrients) and perlite. The trays were properly labeled and placed on tables in a greenhouse. During the plant's growth (2 to 2.5 months), moisture content, temperature and humidity were controlled. When the plants had four true leaves, fertirrigation (HNO_3 + NH_4NO_3 + KPO_4H_2 and MG $(NO_3)_2$ + Hortrilon + Fertrilon) through drippers was applied. Phytopathological treatments were always been carried out preventively.

3.2. DNA extraction

Young leaves from the plants were collected and placed in DNA extraction plates of 96 wells. The DNA extraction method used starts by adding 67 µl of 0.3MNaOH solution and two 4 mm diameter stainless steel balls into each sample well. Followed by grinding the samples at 50 Hz for 60 s using TissueLyser and pacing the deep-well plate into a water bath at 96°C for 1 min. After this 200µl of 0.75MTris-HCl (pH 7.5–7.8) were added and the samples were centrifuged at 3000g for 1 min to separate the supernatants, which were transferred to a new 200-µl 96-well PCR plate. These supernatants were diluted 5x with ultrapure water in a final 96-well plate [32]. In total 24 plates were extracted following the scheme represented (Figure 9).

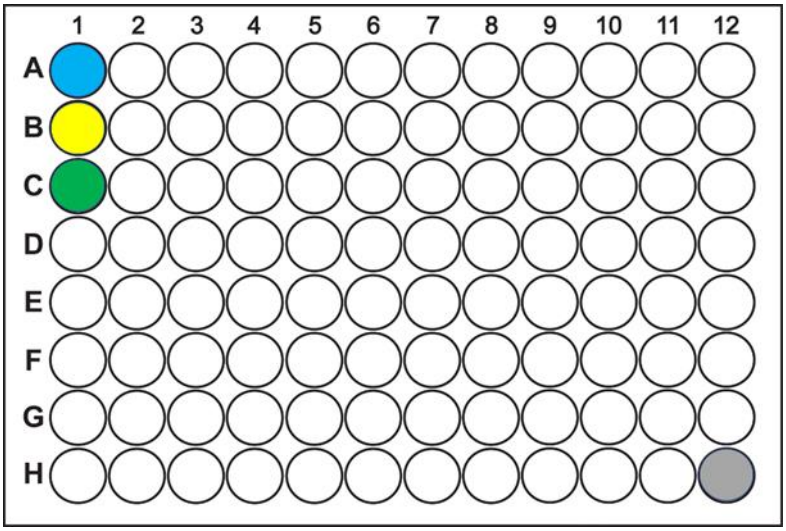

Figure 9. Representation of the 96 well plate scheme used for PCR. The circles represent the wells with DNA samples. In each well containing DNA from the seedlings. Blue, yellow and green wells contained DNA from 'Texas', 'Earlygold' and the hybrid MB 1.37 respectively. In grey a sample of water as the negative control.

DNA was quantified by applying the Beer-Lambert equation relating absorbance, and extinction co-efficient to DNA concentration. Absorbance readings were measured at 260 nm and the extinction coefficients used were 50 for dsDNA and 33 for ssDNA. Absorbance

readings were taken on a NanoDrop spectrophotometer (Wilmington, USA), and the DNA dilutions were kept in a freezer at -20ºC until genetic analysis was done.

3.3. Genotyping

3.3.1. First round of selection

To reduce the time and labor required to carry out the first screening and since the study comprised the analysis of a large number of plants, for *Alf*, *Jui* and *DBF2* we selected 2 SSR markers surrounding the genes. In some cases, if a marker could not be genotyped for technical reasons, a third SSR marker was used (Table 2). Each set of SSRs was genotyped in the families where these genes were heterozygous. The genotyping of the two microsatellites closest to each trait was made both upstream and downstream. When genotyping did not work or primer stock was finished, the second closest SSR was used for genotyping; two SSRs were used in these cases. Each marker was identified as homozygous for almond or peach, or heterozygous when both alleles were present.

The recombinant individuals between two markers surrounding one of the target traits were selected for a second round of selection, while the remaining non-recombinant individuals, were discarded.

Table 2. Microsatellites used for genotyping in the first round of selection. The group to where they belong is represented and the positions of the markers.

Marker	Group	Position
M12a	4	9.208.608
Alf		**7cM**
SSR15662	4	11.228.933
EPPCU2000	4	12.467.623
CPPCT026	1	32.766.347
EPDCU3489	1	33.027.380
Jui		**9cM**
BPPCT016	1	36.075.308
CPPCT029	1	40.195.426
DBF2		**8cM**
SSR6125	1	42.723.336
BPPCT028	1	45.685.797

3.3.2. Second round of selection: Fine mapping

The SSR markers used in the second round of selection were inside the region defined by the markers used in the first round. The genetic markers used are represented on table 3 and include 2 SSRs for *Alf* and 3 for *DBF2*, and 13, 10 and 7 InDels for *Alf*, *Jui* and *DBF2* respectively. In this last round of selection, for the first time in this research project 12 SNP's were also designed 11 for *Jui* and 1 for *DBF2*. These markers allowed us to narrow down the genomic region where these 3 genes are located.

Table 3. Indels and SSR's used for genotyping in the second round of selection.

Region	Markers		
Upstream		Indel34972	SSR5939
		Indel35000	Indel41927
		Indel35036	Indel42115
		Indel35085	Indel42311
		Indel35147	Indel42525
	SSR15636	Indel35173	Indel42584
	SSR15639	SNP35183	SSR6105
	Indel11067	SNP35194	Indel42666
	Indel11120	SNP35200	SNP42675
Genotype	***Alf* 183 Kb**	***Jui* 97 Kb**	***DBF2* 10 Kb**
Downstream	Indel11147	SNP35204	Indel42718
	Indel11163	SNP35211	SSR6125
	Indel11172	SNP35224	
	Indel11175	SNP35244	
	Indel11191	Indel35245	
	Indel11196	SNP35247	
	Indel11206	SNP35252	
	Indel11210	SNP35264	
	Indel11254	SNP35269	
	Indel11303	Indel35271	
	Indel11330	Indel35336	
		Indel35351	

3.4. PCR

The polymerase chain reaction (PCR) is a method which allows the amplification of a specific region of DNA. Two short oligonucleotides, or primers, were designed to match DNA sequence at either end of the region of interest. In the presence of the thermostable enzyme Taq polymerase I, the portion of DNA lying between these two primers is amplified. Two PCR protocols were followed depending if the SSR primers were labelled with different fluorochromes (SSR15662 and SSR6125) or with a tag sequence (M12a, EPPCU2000, CPPCT026, EPDCU3489, BPPCT016, CPPCT029 and BPPCT028) for the first round of selection.

The PCR was performed in a 96-well microtiter plate or in 0.2 ml tubes in a Thermocycler (2700 SimplyAmp, Thermo Fisher®). A Master Mix sufficient for the number of planned reactions was prepared, allowing for a 1x reaction mix once the DNA template was added. The volume and contents of the reactions are represented below in table 4 and 5 for the SSRs labelled with different fluorochromes and for SSRs with a tag sequence, respectively.

Table 4. PCR reaction for SSRs labelled with different fluorochromes (FAM, VIC, NED or PET).

Reagent	Quantity
H2O (HPLC)	5.90 µl
lab (10x)	1.00 µl
MgCl2 (50 mM)	0.30 µl
dNTP (10 mM)	0.20 µl
Primer F (10 µM)	0.20 µl
Primer R (10 µM)	0.20 µl
Taq (5 U/µl)	0.20 µl
1x cocktail	8.00 µl
DNA (20 ng/µl)	2.00 µl
Total	**10.00 µl**

Table 5. PCR reaction for SSRs with a tag sequence.

Reagent	Quantity
H2O (HPLC)	5.82 µl
lab (10x)	1.00 µl
MgCl2 (50 mM)	0.30 µl
dNTP (10 mM)	0.20 µl
Primer tagF (10 µM)	0.20 µl
Primer tagR (10 µM)	0.20 µl
Primer F (10 µM)	0.04 µl
Primer R (10 µM)	0.04 µl
Taq (5 U/µl)	0.20 µl
1x cocktail	8.00 µl
DNA (20 ng/µl)	2.00 µl
Total	**10.00 µl**

PCR amplifications were performed with the specific cycling profile for SSR primers labelled with different fluorochromes and for the ones with a tag sequence (Table 6).

Table 6. PCR cycling profile for the SSRs labelled with different fluorochromes and SSRs with a tag sequence.

PCR Program		
Primers	**Labelled with fluorochromes**	**With a tag sequence**
Denaturation	94°C 1 min	94°C 1 min
Annealing	[94°C 15s; Ta (57-65°C depending on the primer used) 15s; 72°C 30s] 35 cicles	[94°C 15s; 63°C 30s; 72°C 60s] 20 cicles [94°C 15s; 54°C 30s; 72°C 60s] 40 cicles
Extension	72°C 5 min	72°C 5 min

For the second round of selection, we used markers such as SSR's and InDels. The adopted PCR profile for the SSR's was the same used the primers with a tag sequence; for the InDels the same profile utilized in the primers labelled with different fluorochromes was also used, however the latter one's extension time was adapted according to the size of each InDel.
The reaction products were visualized by capillary electrophoresis (SSRs) and agarose gel electrophoresis (InDels).

3.5. Capillary electrophoresis

This electrophoresis procedure was used in the first and second round of selection for all the SSR markers after the PCR. First, 2µl of the resulting PCR amplification products were placed

in a specific plate used for this procedure and mixed with 12 µl of Hi-Di Formamide, 0.35 µl of ROX o LIZ. A denaturation step was performed for 3 min at 94 °C in the Thermocycler and then the plate was analyzed by capillary electrophoresis in an ABI 3130xl Genetic Analyzer (Thermo Fisher®) capillary sequencer. Each PCR product was loaded into a capillary containing a polyacrylamide matrix in which the electrophoresis was performed, the emitted florescence was captured and the molecular mass of the PCR product was determined. An electropherogram was then generated which shows fluorescence peaks corresponding to each amplified allele. The resulting electropherogram corresponding to the outputs were then analyzed with GeneMapper® Software Version 4.0 Microsatellite Analysis (Applied Biosystems) and with GeneMarker® Software Version 2.6.3 (SoftGenetics LLC®), flexible fragment analysis software's packages that provides quality DNA sizing and alleles as well as providing and allowing for the analysis of the peaks.

3.6. Agarose Gel Electrophoresis

Electrophoresis through agarose gels is used to separate DNA fragments by size. The technique is simple, rapid to perform, and capable of resolving fragments of DNA that cannot be separated adequately by other procedures. Electrophoresis uses an electrical field to move the negatively charged DNA through an agarose gel matrix toward a positive electrode. Shorter DNA fragments migrate through the gel more quickly than longer ones. This way we can determine the approximate length of a DNA fragment by running it on an agarose gel alongside a DNA ladder (a collection of DNA fragments of known lengths). The location of bands of DNA within the gel can be determined directly by staining with low concentrations of fluorescent intercalating dyes, such as ethidium bromide [142].

In the present work agarose electrophoresis was used to analyze all the InDels used for the second round of selection.

An agarose gel was prepared (2.5%) by melting 7.5 g of agarose in 300 ml of TBE. Ethidium bromide was added to a concentration of 20µl/100µl and 2µl of loading buffer was added to the totality of the resulting PCR product. Electrophoresis was performed in 1x TBE with a voltage ranging between 100-200 volts for the time required to obtain satisfactory separation. The DNA was visualized under UV light using *QuantumCapture* software.

3.7. SNP's design

SNP's polymorphisms in the resequencing data were detected using Integrative Genomics Viewer software version 2.5 (https://software.broadinstitute.org/software/igv/). The

comparison between the resequencing information from Earlygold and Texas allows the identification of all the polymorphisms between peach and almond in the regions where the genes of interest are located. The last-version of almond (*Texas)* and peach (*Earlygold)* reference genomes used were downloaded from the Genome Database of Rosaceae (https://www.rosaceae.org).

For the SNP design this program displays the parental genomes and compares it with *Prunus persica* Genome v2.0.a1 (peach) as a reference sequence. The program shows the similarities of the readings in gray and the differences with respect to the reference genome in bright colors (Figure 10).

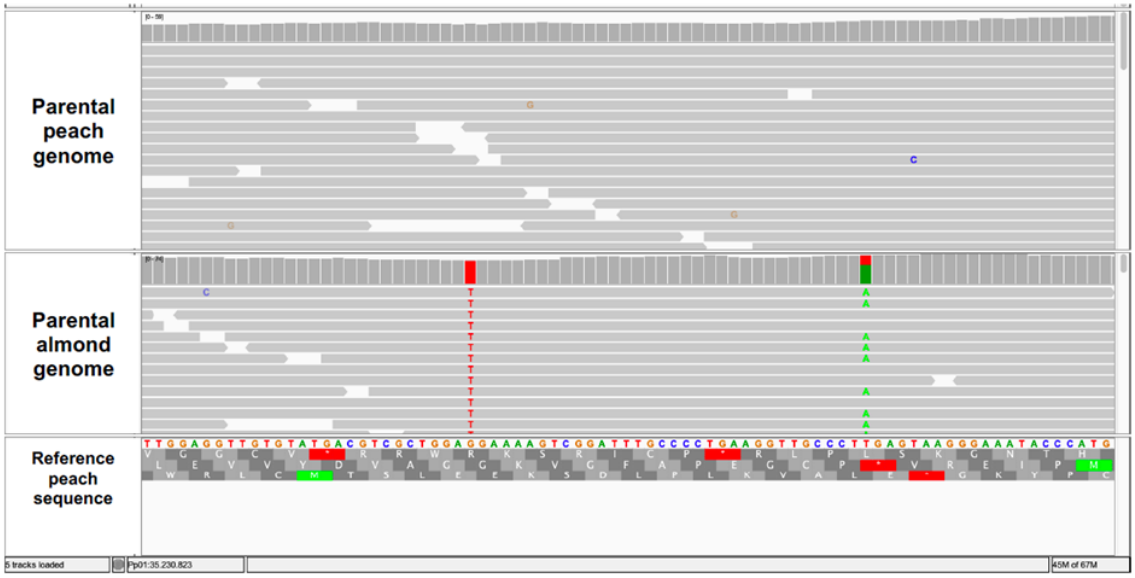

Figure 10. Screenshot of the Integrative Genome Viewer software showing the genomes of 'Earlygold' (above), 'Texas' (in the middle) and P. persica reference sequence (at the bottom). It displays two SNP's in the position 35230823 bp (T instead of G, in red letters) and 35230856 bp (A instead of T, green letters) of Linkage Group 1 in almond's genomic sequence.

The primers for SNP detection were designed using the Primer Picker Lite tool from KASPar SNP Genotyping System (KBiosciences, Herts, UK). It was necessary to have the flanking sequence of the SNP (≈ 200 bp in total) to design them. With this program two allele specific primers (A1 and A2) were obtained, that differed in the nucleotide of the SNP and in a specific tail for each fluorescence (VIC/FAM) and a common primer (C1).

3.8. PACE (PCR Allele Competitive Extension)

PACE (PCR Allele Competitive Extension) genotyping is a fluorescent, competitive allele-specific PCR genotyping technology. It is ideal for biallelic discrimination of single nucleotide polymorphisms (SNPs) and insertions and deletions (Indels) at specific loci. For the Protocol PACE2.0 to be performed we pursued a mix of the tree primers (PACE 2.0 assay) designed

for a specific SNP, a solution was prepared (12 µl of A1, 12 µl of A2, 30 µl of C1 and 46 µl of MilliQ water, for a total of 100 µl) and 2 µl of the DNA samples were placed in an optical plate. The PCR mix (5 µl of Master Mix of PACE 2.0 Genotyping Master Mix, containing Taq polymerase, universal fluorescent reporting cassette, dNTPs, buffer, performance enhancers, MgCl2 at 4.4 mM (2.2 mM at 1x concentration) and the passive reference dye, 5-carboxy-Xrhodamine, succinimidyl ester (ROX)) and 0.138 of PACE 2.0 assay was prepared and added to each sample present in the plate. The SNP genotyping was then performed by qPCR through a LightCycler 480 device (Roche Diagnostics, Spain) using the PACE Genotyping program (Table 7). When the program was finished, we analyzed the results using the program Endpoint Genotyping.

Table 7. PCR program performed in the LightCycler 480, for the PACE Genotyping.

Name	Cicle	Temperature	Time	Ramp rate	Quantification
Hot-start	1	94 °C	0:15:00	4,4	None
Touch-down	10	94 °C	0:00:20	4,4	None
		65 °C	0:01:00	2,2	None
PCR	35	94 °C	0:00:20	2,2	None
		57 °C	0:01:00	2,2	None
Read plate	1	37 °C	0:01:00	2,2	None
		37 °C	0:00:01	4,4	Single

3.9. Selection of candidate genes

To select the candidate genes for each trait we used the annotation file from *Prunus persica* Genome v2.0.a1 (peach), available in the laboratory data bases. (https://www.rosaceae.org/species/prunus_persica/genome_v2.0.a1).

2nd Experiment

3.10. Plant Material

To test and compare the three extraction protocols proposed (Alkaline lysis method, CTAB and SILEX), young and old leaf tissues was harvested from 13 species: plum, apricot, pear and blackberry. For strawberry we collected leaves from *Fragaria vesca* (diploid) and *Fragaria* × annanasa (octoploid), and for asparagus we test leaf and spear tissue. We collected young leaf tissue from hybrids of peach and almond, peach, melon, almond, cannabis, grape and

apple. The material was collected and placed in DNA extraction plates of 96 wells directly from plants in the field or greenhouse, using the scheme showed on table 8.

Table 8. Scheme used in the 96 well-plate for the DNA extraction using the three protocols. The nomenclature young and old is used for each species to identify the type of tissue collected. Leaf and spear are the tissues collected and used for the DNA extraction of asparagus.

Plate scheme for DNA extractions											
Peach x Almond	**Peach**	**Plum**	**Melon**	**Almond**	**Apricot**	**Cannabis**	**Strawberry**	**Grape**	**Apple**	**Asparragus/ Pear**	**Blackberry**
Rec	**Pc**	**P**	**M**	**Al**	**A**	**C**	**S**	**G**	**Ap**	**As/Pr**	**B**
young	young	young	young	young	young	young	vesca	young	young	As Leaf	young
young	young	young	young	young	young	young	vesca	young	young	As Spear	young
young	young	young	young	young	young	young	vesca	young	young	As Spear	young
young	young	young	young	young	young	young	vesca	young	young	As Spear	young
young	young	old	young	young	old	young	octaploid	young	young	Pr young	old
young	young	old	young	young	old	young	octaploid	young	young	Pr young	old
young	young	old	young	young	old	young	octaploid	young	young	Pr old	old
young	young	old	young	young	old	young	octaploid	young	young	Pr old	old

3.11. Alkaline lysis extraction method

The first DNA extraction method started by adding 67 µl of 0.3MNaOH solution and a 4 mm diameter stainless ball into each sample well of the 96 well-plate. Followed by grinding at 50 Hz for 60s each side using TissueLyser and placing the deep-well plate into a water bath at 96 °C for 1 min. After this 200 µl of 0.75MTris-HCl (pH 7.5–7.8) were added and the samples were centrifuged at 3000g for 1 min to separate the supernatants, which were transferred to a new 200-µl 96-well PCR plate (Figure 11). These supernatants were diluted 5x with ultrapure water. A DNA spectrophotometer *NanoDrop Technologies* (Wilmington, USA) was used to quantify the final concentration of DNA in ng/µL and its quality.

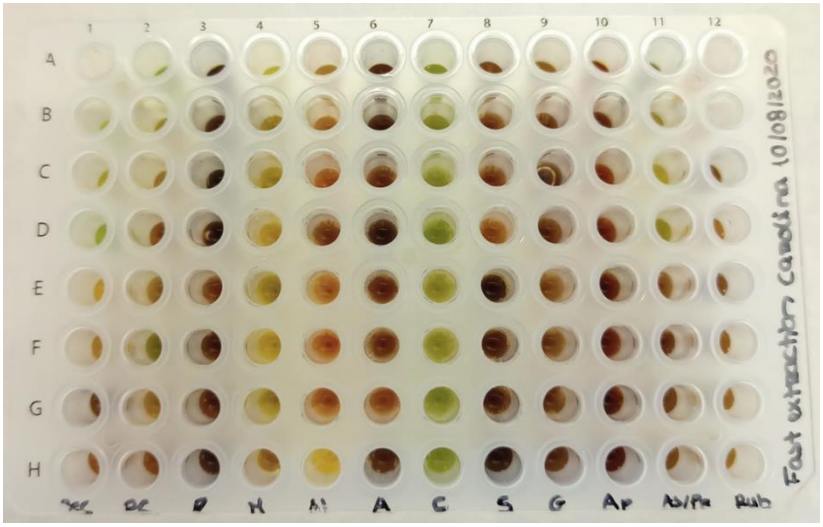

Figure 11. Picture of the alkaline lysis 96-well PCR plate before the dilution 5x with ultrapure water.

3.12. CTAB extraction method

The CTAB protocol for DNA extraction used was adapted to the conditions of the laboratory in which this study was carried out and begins with the addition of 4 mm diameter stainless ball into each sample well of the 96 well-plate and 340 µl of Doyle buffer (mixed with mercaptoethanol; 20 µl for 20 ml of Doyle). After grinding the material at 30 Hz for 2 min each side using TissueLyser, the plate was incubated at 65ºC for 40 minutes at least. 340 µl of chloroform were added and mix by rigorous shaking. Then a centrifugation at 3000 rpm for 15 min was done and 100 µl of the aqueous phase was transferred in to a PCR plate, followed by the addition of 100 µl of isopropanol. A new centrifugation step is performed at 3000 rpm for 30 minutes, and the supernatant was discarded. 100 µl of 70% ethanol were added and the centrifugation is repeated for 10 min. In the end the supernatant was eliminated and the plate was left l on the bench for 24 h. In the next day the DNA was resuspended in 180 µl of HPLC water for dilution 10x. After applying a spin in the centrifuge to the plate the DNA was quantified using the previous spectrophotometer *NanoDrop Technologies* (Wilmington, USA) for the same propose.

3.13. SILEX extraction method

In the present year a new protocol based on the standard CTAB method with a DNA silica matrix recovery was developed [28]. This was adapted to the laboratory conditions (Annex 1) and tested through the collection of leaf material (≈ 15-25 mg) and placing it in tubes of 1.1 ml. Tungsten bead, 437 µl of extraction buffer (2% (w/v) CTAB, 2% (w/v) PVP-40, 20 mM EDTA, 100 mM Tris HCl (pH 8.0) and 1.40 M NaCl) and 6 µl of β-marcaptoethanol were added to the 96 well-plate. After the samples were grinded using the TissueLyser (30 HZ, 1:30 sec. each side) and 2 µl of Rnase (10 mg/ml) were added before an incubation step was performed in a thermalblock for 30 min at 65 ºC. At next the samples were placed on ice for 5 min. and 306 µl of protein precipitation buffer (24 parts of chloroform and 1 part of isoamyl alcohol) were added and gently vortexed. A centrifugation at 3 000 rpm for 15 min. at room temperature, to recover around 300 µl of the supernatant phase to a new 2 ml tube is done and 178 µl of binding buffer (2.5 M NaCl and 20% PEG 8000) and 267 µl of absolute ethanol are added and gently invert the tube by hand until complete mixing. Addition of 8 µl of the silica matrix buffer (5 g of silicon dioxide, 50 ml and 10 µl of HCl 36% per ml of silica matrix solution obtained) and mix gently during 5 min by hand, spin down the silica for 5 min and the supernatant must be discarded by decantation. In the end of this protocol 700 µl of washing buffer (fresh prepared ethanol 70%) were added and

shake gently by hand until a uniform dispersion of the silica was obtained; the silica was spin down for 5 min, and the supernatant gently discarded by decantation and let dry at room temperature overnight. In the next day 100 µl of elution buffer (10 mM Tris HCl (pH 8.0) and 1 mM EDTA (pH 8.0)) were added and a gentle shake was given by hand util the pellet was resuspended and the plate was incubated 5 min at 65 ºC. The plate is centrifugated at 3 000 rpm for 20 min at room temperature and 90 µl of the supernatant were transferred to a new tube. As in the previous protocols the DNA is quantified using Nanodrop.

3.14. NanoDrop Spectrophotometer

Nucleic acid samples can be readily checked for concentration and quality using the NanoDrop Spectrophotometer. Based on the Beer-Lambert Law (A=Ɛcl, where *A=absorbance, Ɛ=extinction coefficient, c=concentration and l=path length*), that draws a direct correlation between absorbance and concentration. While nucleic acids absorb at many wavelengths, they have a peak absorbance of UV light at 260nm. Thus, the amount of light absorbed in this region can be used to determine the concentration of RNA or DNA in solution by applying the Beer-Lambert law. However, the Beer-Lambert equation is only linear for absorbances between 0.1 and 1.0. This translates to concentrations between 10.0 ng/uL and 3700 ng/uL when using the Nanodrop. Values above 20 ng/µl are considered reasonable DNA concentration, of course these values vary accordingly to the objective after the extraction.

3.15. Agarose gel Electrophoresis

The agarose gels were prepared (1%) by melting 3 g of agarose in 300 ml of TBE. Ethidium bromide was added to a concentration of 20µl/100µl and 2µl of loading buffer was added to the totality of the resulting PCR product. Electrophoresis was performed in 1x TBE with a voltage ranging between 100-200 volts for the time required to obtain satisfactory separation. The DNA was visualized under UV light using *QuantumCapture* software.

4. RESULTS

1st Experiment

The individuals of the T1E population were selected if they were heterozygous for one of the target genes and if they showed a medium or late ripening to avoid resorting to the embryo rescue (Table 9).

Table 9. List of the T1E possible recombinants. In the first column is showed the progenies number of different generations derived from 'Texas' and 'Earlygold' selected to produce seedlings. MD corresponds to the maturity date of the fruits in each plant individual. Is also showed the number of fruits, seeds and seedlings obtained from each individual and the percentage of germination.

T1E	MD	Nº Fruits	Nº Seeds	Nº seedlings	% Germination
21	29-Aug	50	51	39	76.47
22	1-Aug	41	42	9	21.43
32	1-Aug	153	145	28	19.31
34	29-Aug	235	265	241	90.94
35	1-Aug	213	208	123	59.13
40	29-Aug	75	77	61	79.22
43	1-Aug	50	57	15	26.32
64	29-Aug	185	254	234	92.13
65	1-Aug	33	22	14	63.64
101	1-Aug	109	108	82	75.93
104	14-Aug	164	189	180	95.24
116	1-Aug	89	81	32	39.51
123	29-Aug	130	131	98	74.81
197	1-Aug	166	186	76	40.86
201	1-Aug	204	196	127	64.80
219	1-Aug	86	65	21	32.31
220	1-Aug	118	150	113	75.33
226	1-Aug	142	139	77	55.40
344	14-Aug	65	63	45	71.43
389	29-Aug	226	235	226	96.17
410	1-Aug	174	217	63	29.03
424	14-Aug	73	73	50	68.49
463	29-Aug	118	110	89	80.91
467	1-Aug	111	120	43	35.83
500	14-Aug	194	196	97	49.49
TOTAL:		**3204**	**3380**	**2183**	**Average: 61**

In 2019 open pollinated fruits from these 25 individuals selected were harvested and the seeds were extracted and germinated. In total we obtained, 3204 fruits. An average of 128.16 fruits were obtained per family, with, only four that produced less than 50 fruits (included) and nine that produced more than 150 fruits. The number of seedlings obtained by family was between nine and 241 individuals with an average of 87.32. Only seven families produced less than 50 seedlings, and four more than 150.

4.1. First round of selection

In the first selection stage, the 2183 individuals resulting from BC1 selfing were genotyped with the 9 SSRs mentioned above to select the individuals that presented a recombination between the markers surrounding the different target genes. The genotype of each marker was annotated, based on the emitted florescence of the PCR product. The electropherogram generated by GeneMapper and GeneMarker softwares shows the fluorescence peaks corresponding to each amplified allele, as represented on figures 13, 14 and 15 for the M12a SSR marker.

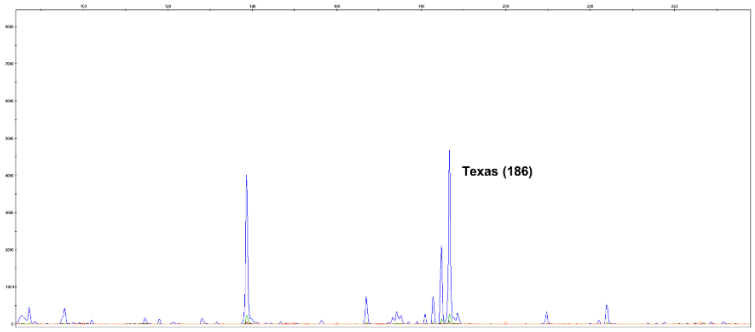

Figure 12. Screenshot of the GeneMapper software showing the electropherogram generated for the Alf region, screened with the marker M12a, for 'Texas'.

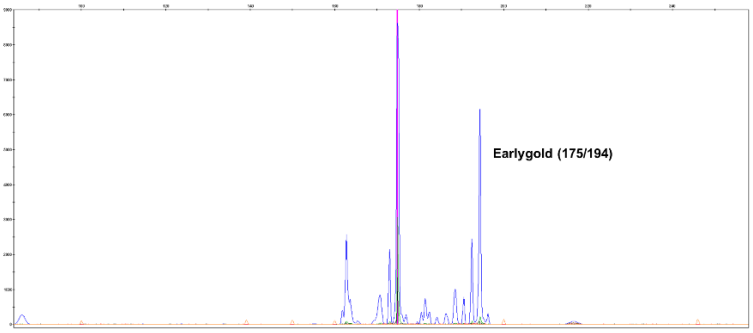

Figure 13. Screenshot of the GeneMapper software showing the electropherogram generated for the Alf region, screened with the marker M12a, for 'Earlygold'.

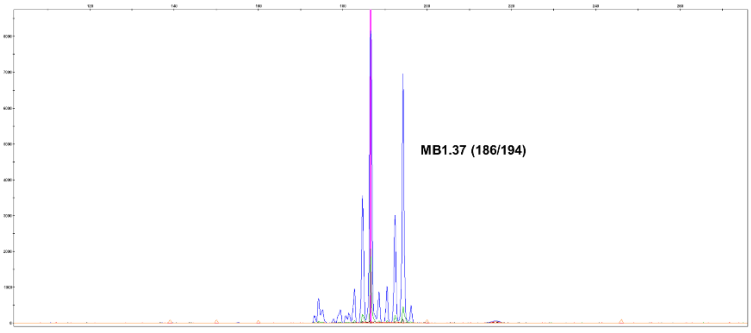

Figure 14. Screenshot of the GeneMapper software showing the electropherogram generated for the Alf region, screened with the marker M12a, for the hybrid MB 1.37.

Marker M12a presented the genotypes 186, 175/194 and 186/194 for T, EG and MB1.37, that were the almond parental, the peach and F1 hybrid respectively. We can observe that the hybrid MB1.37 presented one allele coming from 'Texas' (186) and another one coming from 'Earlygold' (194). The genotyped individuals were scored as homozygous for 186 or 194 or as heterozygous. Based on the electropherograms generated for each marker we identified the individuals that were recombinants in the genomic region containing each gene of interest (Table 10).

Table 10. Representation of the fluorescence emitted by each marker chosen for the screening of Alf, Jui and DBF2 and the respective annotation to identify recombinants.

Marker	Group	Position	Texas	Earlygold	MB1.37		Example of Recombinats	
M12a	4	9.208.608	186	175/194	175/186	186/194	186	186/194
Alf		**7cM**						
SSR15662	4	11.228.933	227/247	255	255		255	247
EPPCU2000	4	12.467.623	122	120	120/122			
CPPCT026	1	32.766.347	150	176	150/176			
EPDCU3489	1	33.027.380	155	170	155/170		170	155/170
Jui		**9cM**						
BPPCT016	1	36.075.308	86	96	86/96		86/96	96
CPPCT029	1	40.195.426	178/190	190	178/190		178/190	190
DBF2		**8cM**						
SSR6125	1	42.723.336	260	280	260/280		280	260/280
BPPCT028	1	45.685.797	165	163	163/165			

In the almond × peach crossing population *Alf* and *Jui* are recessive genes in the peach genome, whereas *DBF2* is dominant and the responsible allele comes from the almond. For *Alf*, the individuals were selected as recombinants if for one marker were homozygous for the almond allele and for the other were heterozygous or homozygous for peach.

For *DBF2* the almond allele is dominant, therefore the individuals were selected as recombinants if one marker was homozygous for the peach allele and for the other was heterozygous showing the genotype of almond or hybrid in the upper part or heterozygous.

For *Jui*, the peach allele is the recessive and we selected the recombinant if one of the markers was homozygous for peach and the other was heterozygous for peach and almond alleles.

Table 11 summarizes the efficiency of the markers used in detecting recombinants and the number of recombinant individuals determined for each genome region.

Table 11. Summary of the efficiency of each marker during the screening, and number of recombinants selected for each genome region, that will proceed to the second round of selection.

Marker	Group	Position	Nº Individuals (Analyzed)	Nº Individuals (Failed Analyzis)	Nº Recombinants Selected	Nº Individuals Discarded
M12a	4	9.208.608	1831	330	239	1944
Alf		**7cM**				
SSR15662	4	11.228.933	1427	758		
EPPCU2000	4	12.467.623	164	204		
CPPCT026	1	32.766.347	370	182	199	1984
EPDCU3489	1	33.027.380	1621	587		
Jui		**9cM**				
BPPCT016	1	36.075.308	2009	163		
CPPCT029	1	40.195.426	1810	373	112	2071
DBF2		**8cM**				
SSR6125	1	42.723.336	1722	461		
BPPCT028	1	45.685.797	55	37		

For *Alf* the marker M12a has the one that allow us to analyze the higher number of individuals and for *Jui* and *DBF2* was the BPPCT016 and CPPCT029 respectively. The marker that showed the higher number of analysis failed was the SSR15662 from *Alf*.

In the first round of selection a total of 2183 individuals were genotyped with three to nine SSRs depending on the number of target genes that were heterozygous in the parental line. In total we identified 550 recombination events with 239 possible recombinants for *Alf*, 199 for *Jui*, and 112 for *DBF2*. In total, 1633 plants were dismissed.

4.2. Second round of selection

The second round of selection was done with the 550 recombinant individuals described in the previous section. To further narrow down the positions of these recombination events we used 4 SSRs, 30 Indels and 12 SNPs depending on the target gene of each plant. The SSRs were visualized as described before using GeneMapper and GeneMarker software's through the electropherograms, the InDels were visualized and analyzed in agarose gels (Figure 16).

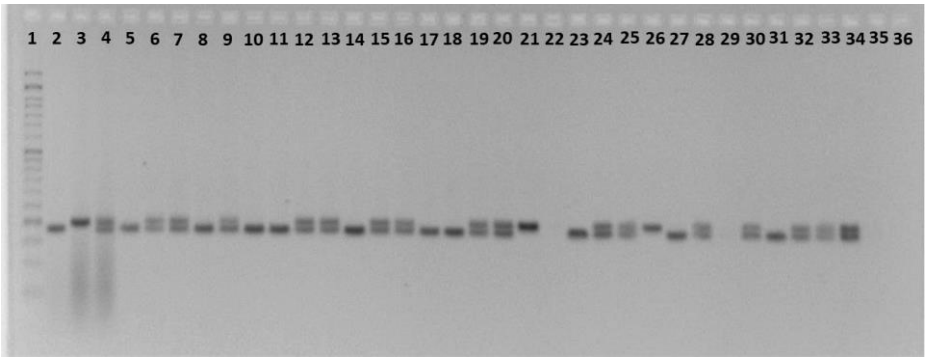

Figure 15. Portion of an agarose gel that shows the genotype of 31 samples (four missing 22,29,35 and 36) for the Indel11175, downstream of Alf gene. The molecular weight marker is on the first column (1) on the left and the rest of the columns show both peach (3, 21 and 26) and almond (2, 5, 8, 10, 11, 14, 17, 18, 23, 27 and 31) homozygous states and heterozygous individuals (4, 6, 7, 9, 12, 13, 15, 16, 19, 20, 24, 25, 28, 30, 32, 33 and 34).

The SNP genotyping was performed by PACE genotyping through a LightCycler 480 device (Roche Diagnostics, Spain) using the PACE Genotyping program, as described before (section 3.8). And the results were analyzed using the program Endpoint Genotyping (Figure 17).

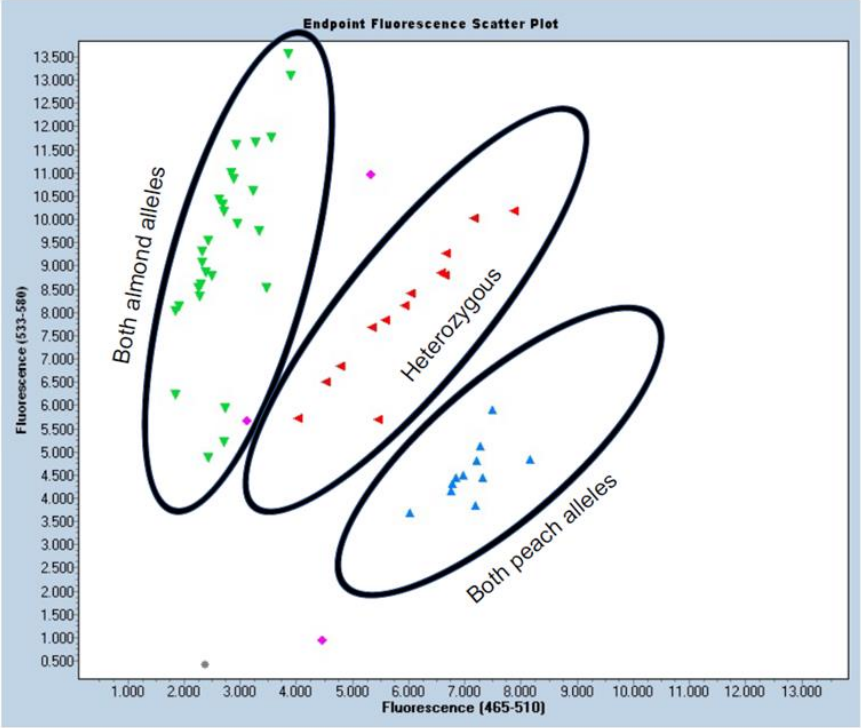

Figure 16. Screenshot of the Endpoint Genotyping program layout result. Three theoretical groups can be distinguished. The bottom group made up of blue dots, represents those genotypes that were homozygous for peach allele. The red group represents heterozygous genotypes, and the green almond homozygous individuals. The dots in pink were classified as unknown and the grey one is the negative control classified as 'non parental allele present'.

In total we identified 20 new recombinant individuals, 13 for *Alf*, 4 for *Jui* and 3 for *DBF2*, that were planted in the fields of IRTA Torre Marimon until they will be phenotype the next years. More details about the markers genotyped are present in the next section separately for the three target genes.

4.2.1. Fine mapping

Alf, *Jui* and *DBF2* traits were identified and mapped by Donoso et al. (2016). *Alf* segregated on G4 between the markers M12a and EPPCU2000, corresponding to the physical position 32.8. *Jui* was mapped to the distal position of G1 between the markers CPPCT026 and BPPCT016, in the physical position 41.5 and *DBF2* to the end of G1 between the markers CPPCT029 and BPPCT028, in the physical position 53.0. (Table 12 and Figure 18).

Table 12. Description of the genes and their segregation in the interspecific almond × peach progenies T × E and T1E. Adapted from: Donoso et al. (2016).

Trait	Symbol	Texas[a]	Earlygold[a]	F1[a]	Map[b]	Expected segregation	Observed segregation	χ^{2c}	Map position (LG:cM)	Closest marker[d]	Physical position[e]
Almond fruit type	*Alf/alf*	Almond (*alf/alf*)	Peach (*Alf/Alf*)	Peach (*Alf/alf*)	T × E	3:1	30 *Alf*-:14 *alf/alf*	0.13	4:32.8	SNP_IGA_412380 (1.1)	4:10922662 12523245
Juiciness	*Jui/jui*	Non-juicy (*Jui*-)	Juicy (*jui/jui*)	Non-juicy (*Jui/jui*)	T1E	1:1	65 *jui/jui*:64 *Jui/jui*	0.08	1:41.5	SNP_IGA_107417 (0.3)	1:35198093 35398680
Blood flesh	*DBF2/ dbf2*	– (*DBF2*-)	Yellow (*dbf2/dbf2*)	Red (*DBF2/dbf2*)	T1E	1:1	73 *DBF2/dbf2*:57 *dbf2/dbf2*	1.97	1:53.0	snp_1_41535285 (0.1)	1:41709139 42754924

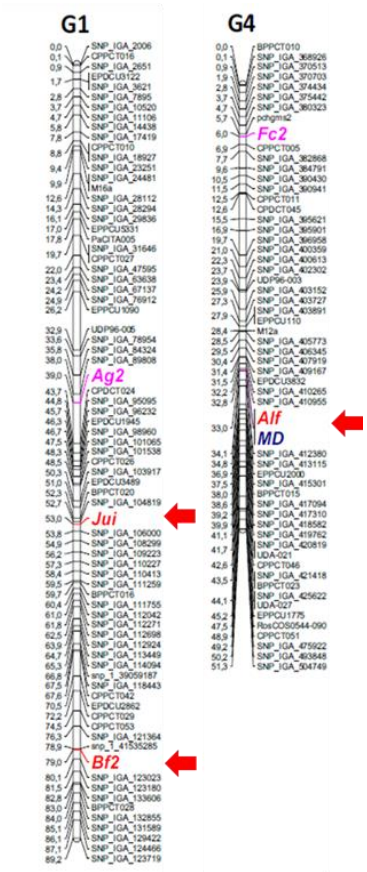

Figure 17. Map with the physical positions of the genes mapped by Donoso et al. (2016). In red is marker de name of the genes under study responsible for fruit traits. Source: Donoso et al. (2016).

On the years before this master project more recombinant individuals were already selected and some of them were phenotyped. These individuals were TxE059, TxE085, TxE004, TxE202, T1E724, 25P15-32, T1E8, T1E183, T1E694, T1E62, 21P15-27, T1E197, T1E343 and T1E464. To narrow down the region where the target genes are located, molecular markers (Indels and SSRs) were developed in the past. In the present work we also applied SNPs with the same objective as the other markers previously tested. The genotyping and phenotype results for the recombinant individuals for *Alf*, *Jui* and *DBF2* are presented in tables 13, 14 and 15 respectively.

Table 13. Genotypes of individuals with a possible recombinant breakpoint near Alf. A, B and H represent the alleles responsible for 'Texas', 'Earlygold' and hybrid 'MB1.37' genotypes respectively. C represents the genotype when 'Earlygold' and hybrid are not distinguished by the marker and the numbers correspond to the size of the pics for the electropherograms resulting from the analysis of the SSRs. The red lines surround the region were the breakpoint is believed to be.

Distance	Marker	Position	Controls Genotype					Individuals Genotype													Controls Genotype	
			T	EG	MB1.37	TxE059	TxE085	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	TxE004	TxE202
	M12a	9208608	186	175/194	186/194	186	186	186	186	186	186	186	186	186	186	175/194	175/186	186	186		186	186
	SSR15636	11058333	202/205	210	210	205	205	205	205	205		205		205	205		210	205	205		210	210
	SSR15639	11066811	204/210	226	204/226	204								204				204			226	226
	Indel11067	11067236	A	B	H	A	A	A	A	A		A	A	A	A			A	A		H	H
	Indel11120	11119850	A	B	H	A	A	A	A	A	A	A	A	A	A	H	H	A	A		H	H
	Phenotype					A	C														C	C
	Indel11147	11147190	A	B	H	A	H	H	H	H	H	H	A	A	A	H	H	A	A		H	H
	Indel11163	11153064	A	B	H				H	H	H	H	H	A	A	H	H	A	A			H
	Indel11172	11162651	A	B	H	A	H	H	H	H	H	H	H	A	A	H	H	A	A			H
183 Kb	Indel11175	11165424	A	B	H	A	H	H	H	H	H	H	H	A	A	H	H	A	A		H	
	Indel11191	11181771	A	B	H	A	H	H	H	H	B	B	H	H	A		H	A	A			H
	Indel11196	11186338	A	B	H	A	H	H	H	H	H	H	H	H	A	H	H	A	A		H	
	Indel11206	11195658	A	B	H	A	H	H	H	H	H	H	H	H	H	H	H	A	A			H
	Indel11210	11210804	A	B	H	A	H	H	H	H					H	A	H				H	H
	Indel11254	11254057	A	B	H	A	H	H	H	H	H	H	H	H	H		A	A	A		H	H
	Indel11303	11303223	A	B	H	H	H	H	H	H	H	H	H	H	H	A	A	H	H		H	H
	Indel11330	11330342	A	C	C	C	C		C					C							C	C
	EPPCU2000	12467623	122	120	120/122	120/122	120/122					120/122			120/122						120/122	120/122

Table 14. Genotypes of individuals with a possible recombinant breakpoint near Jui. A, B and H represent the alleles responsible for 'Texas', 'Earlygold' and hybrid 'MB1.37' genotypes respectively. The numbers correspond to the size of the pics for the electropherograms resulting from the analysis of the SSRs. The red lines surround the region were the breakpoint is believed to be.

Distance	Marker	Position	Controls Genotype					Individuals Genotype				Controls Genotype		
			T	EG	MB1.37	T1E724	25P15-32	102P19-21	165P19-215	131P19-142	104P19-93	T1E8	T1E183	T1E694
	CPPCT026	31,792,329	150	176	150/176	176	150/176	176	176	150/176	176	176	176	176
	EPPCU3489	33,027,380	155	170	155/170	170	155/170	170	170	155/170	170	170	170	170
	Indel34972	33,998,283	A	B	H	B	H	B	B			B	H	
	Indel35000	34,026,756	A	B	H	B	H	B	B	H	B	B	H	
	Indel35036	34,062,108	A	B	H	B	H	B	B	H	B	H	H	B
	Indel35085	34,111,389	A	B	H	B	H	B	B	H	B	H	H	B
	Indel35147	34,173,520	A	B	H	B	H				B	H	H	B
	Indel35173	34,199,641	A	B	H	B	H	B	B	H	B	H	H	B
	SNP35183	34,209,841	A	B	H	B	H	B	B	H	B	H	H	
	SNP35194	34,221,098	A	B	H	B	H	B	B	H	B	H	H	B
	SNP35200	34,226,652	A	B	H	B	B	B	B	H	B	H	H	
	Phenotype					B	B					H	H	B
	SNP35204	34,230,699	A	B	H	B	B	B	H	B	B	H	H	B
	SNP35211	34,237,359	A	B	H	B	B	B	H	B	B	H	H	B
69Kb	SNP35224	34,250,624	A	B	H	B	B	B	H	B	B	H	H	B
	SNP35244	34,256,057	A	B	H	B	B	H	H	B	B	H	H	B
	Indel35245	34,256,978	A	B	H	B	B	H	H	B		H	H	B
	SNP35247	34,273,606	A	B	H	B	B	H	H	B	B	H	H	
	SNP35252	34,278,504	A	B	H	B	B	H	H	B	B	H	H	
	SNP35264	34,290,349		B	H	H	B	H	H	B	H	H	H	B
	SNP35269	34,295,319		B	H	H	B	H	H	B	H	H	H	B
	Indel35271	34,297,238	A	B	H	H	B	H	H	B	H	H	H	B
	Indel35336	34,363,000	A	B	H	H	B	H	H	B	H	H	H	B
	Indel35351	34,377,195	A	B	H	H	B	H	H	B	H	H	H	B
	BPPCT016	36,075,308	86	96	86/96	86/96	96	86/96	86/96	96	86/96	86/96	86/96	86/96

Table 15. Genotypes of individuals with a possible recombinant breakpoint near DBF2. A, B and H represent the alleles responsible for 'Texas', 'Earlygold' and hybrid 'MB1.37' genotypes respectively. The numbers correspond to the size of the pics for the electropherograms resulting from the analysis of the SSRs. 0The red lines surround the region were the breakpoint is believed to be.

Distance	Marker	Position	Controls Genotype					Individuals Genotype			Controls Genotype		
			T	EG	MB1.37	T1E62	21P15-27	149P19-18	104P19-96	103P19-211	T1E197	T1E343	T1E464
	CPPCT029	40195426	178	190	178/190	178/190	190	178/190	190	190	178/190	190	190
	SSR_5939	40502945					B						
	Indel41927	40953671				H					H	B	H
	Indel42115	41142482				H	B				H	B	H
	Indel42311	41338395	A	B	H	H	B	H	B	B	H	B	H
	Indel42525	41551835				H	B				H	B	H
	Indel42584	41610491	A	B	H	H	B	H	B	B	H	B	H
	SSR_6105	41665544	270	300	270/300	270/300	300	270/300	300		270/300	300	270/300
	Indel42666	41693407	A	B	H	H	B	H	B		H	B	H
	SNP42675	41702598	A	B	H	H	B	H	B	B	H	B	H
1.4 Kb	Phenotype					H	H				H	B	H
	reseq rec point 062	41703992				H							
	reseq rec point 062	41703993				B							
	Indel42718	41744798	A	B	H	B	H	B	H	H	H	B	H
	SSR_6125	41750304	260	280	260/280	280	260/280	280	260/280	260/280	260/280	280	260/280

The markers surrounding the target genes in the genetic map presented by Donoso et al. (2016), covered a genomic region of 7, 9 and 8 cM for *Alf*, *Jui* and *DBF2* respectively. In the past years the saturation with molecular markers allowed the reduction of the region were *Alf* is inserted to 183 Kb, for *Jui* to 132 Kb and for *DBF2* to 377 Kb. Through the fine mapping in the present work we were able to narrow down the region of *Jui* and *DBF2* to 69 Kb and 1.4Kb respectively. For *Alf* we have to wait to phenotype the new recombinants found to narrow down the region.

The region where *Alf* is localized contain 29 annotated genes (Table 16). With the reduction of the region where *Jui* and *DBF2* are located we were also able to decrease the number of candidate genes responsible for these two traits. Before this year's fine mapping we had a total of 37 candidate genes for *Jui* and 13 candidate genes for *DBF2*. The final number of candidate genes for these two traits after this approach is 11 candidate genes for *Jui* and 1 candidate gene for *DBF2* (Table 17 and 18).

Table 16. Alf candidate genes after this year's fine mapping.

Gene	Position			Predicted Function
	Location	Start	Stop	
Prupe.4G187100.1	Pp04	11138518	11140641	NAC transcription factor 25 (*Arabidopsis thaliana*)
Prupe.4G187200.1	Pp04	11145857	11146604	n/a
Prupe.4G187300.1	Pp04	11147111	11151935	Peroxisomal membrane protein 2 (*Bos taurus*)
Prupe.4G187400.1	Pp04	11153240	11158321	Peptidyl-prolyl cis-trans isomerase CYP37, chloroplastic (*Arabidopsis thaliana*)
Prupe.4G187500.1	Pp04	11161053	11175010	Trafficking protein particle complex subunit 10 (*Dictyostelium discoideum*)
Prupe.4G187700.1	Pp04	11178009	11180316	n/a
Prupe.4G187800.1	Pp04	11184964	11190022	PAP-specific phosphatase HAL2-like (*Arabidopsis thaliana*)
Prupe.4G187900.1	Pp04	11192916	11193704	n/a
Prupe.4G188000.1	Pp04	11206310	11210427	Receptor-like protein kinase THESEUS 1 (*Arabidopsis thaliana*)
Prupe.4G188100.1	Pp04	11211569	11213268	n/a
Prupe.4G188200.1	Pp04	11215309	11217498	n/a
Prupe.4G188300.1	Pp04	11218458	11219376	n/a
Prupe.4G188400.1	Pp04	11220112	11220971	n/a
Prupe.4G188500.1	Pp04	11220929	11222217	n/a
Prupe.4G188600.1	Pp04	11223366	11227007	Lysine-specific histone demethylase 1 homolog 2 (*Arabidopsis thaliana*)
Prupe.4G188700.1	Pp04	11227089	11228967	Protein ULTRAPETALA 2 (*Arabidopsis thaliana*)
Prupe.4G188800.1	Pp04	11232893	11238310	Uncharacterized protein YnbB (*Bacillus subtilis* (strain 168))
Prupe.4G188900.1	Pp04	11238694	11241828	Transmembrane protein 184C (*Pongo abelii*)
Prupe.4G189000.1	Pp04	11242310	11245089	n/a
Prupe.4G189100.1	Pp04	11244108	11247472	n/a
Prupe.4G189200.1	Pp04	11247548	11249911	n/a
Prupe.4G189300.1	Pp04	11250220	11252990	Monodehydroascorbate reductase, seedling isozyme (*Cucumis sativus*)
Prupe.4G189400.1	Pp04	11254886	11256301	n/a
Prupe.4G189500.1	Pp04	11271534	11272507	n/a
Prupe.4G189600.1	Pp04	11274431	11275614	n/a
Prupe.4G189700.1	Pp04	11278755	11285726	Probable ADP-ribosylation factor GTPase-activating protein AGD5 (*Arabidopsis thaliana*)
Prupe.4G189800.1	Pp04	11289619	11292044	n/a
Prupe.4G189900.1	Pp04	11294659	11299414	Cytochrome c-type biogenesis ccda-like chloroplastic protein 1 (*Oryza sativa* subsp. *japonica*)
Prupe.4G190000.1	Pp04	11300877	11316907	B3 domain-containing transcription repressor VAL2 (*Arabidopsis thaliana*)

Table 17. Jui candidate genes after this year's fine mapping.

Gene	Position			Predicted Function
	Location	Start	Stop	
Prupe.1G397400.1	Pp01	35195006	35196774	Peroxidase 18 (*Arabidopsis thaliana*)
Prupe.1G397500.1	Pp01	35196836	35199864	Putative pentatricopeptide repeat-containing protein At3g25970 (*Arabidopsis thaliana*)
Prupe.1G397600.1	Pp01	35202922	35205915	Protein ASPARTIC PROTEASE IN GUARD CELL 2 (*Arabidopsis thaliana*)
Prupe.1G397700.1	Pp01	35206904	35208871	Protein ASPARTIC PROTEASE IN GUARD CELL 2 (*Arabidopsis thaliana*)
Prupe.1G397800.1	Pp01	35217508	35218879	n/a
Prupe.1G397900.1	Pp01	35229142	35231096	Protein ASPARTIC PROTEASE IN GUARD CELL 1 (*Arabidopsis thaliana*)
Prupe.1G398000.1	Pp01	35233176	35237006	Protein ASPARTIC PROTEASE IN GUARD CELL 1 (*Arabidopsis thaliana*)
Prupe.1G398100.1	Pp01	35243497	35245491	Protein ASPARTIC PROTEASE IN GUARD CELL 2 (*Arabidopsis thaliana*)
Prupe.1G398200.1	Pp01	35247719	35251222	Protein ASPARTIC PROTEASE IN GUARD CELL 1 (*Arabidopsis thaliana*)
Prupe.1G398300.1	Pp01	35253708	35254845	Protein ENHANCED DISEASE RESISTANCE 2-like (*Arabidopsis thaliana*)
Prupe.1G398500.1	Pp01	35259778	35267247	Uncharacterized protein sll0005 (*Synechocystis sp.*) (strain PCC 6803 / Kazusa)

Table 18. DBF2 candidate genes after this year's the fine mapping.

Gene	Position			Predicted Function
	Location	Start	Stop	
Prupe.1G519800.1	Pp01	42674777	42677556	UDP-glycosyltransferase 85A2 (*Arabidopsis thaliana*)

2nd Experiment

To compare the three methods of extraction and verify the purity and quality of the DNA we used Nanodrop spectrophotometer and agarose gel analysis. The average of the results for each group of samples and extraction method obtained in the Nanodrop are presented in Table 19.

Table 19. Representation of the average results obtained in the Nanodrop spectrophotometer for the DNA extractions using alkaline lysis, CTAB and SILEX extraction methods. In red are marked the lower values of DNA concentration (ng/µl)(under or closer to 20 ng/µl) and for the A260/280 and A260/230 ratios (under 1.5).

Species	Rec	Pc	P young	P old	M	Al	A young	A old	C	S vesca	S octo	G	Ap	As Leaf	As Spear	Pr young	Pr old	B young	B old	Average
Alkaline lysis extraction																				
ng/ul	422.0	296.4	181.8	494.6	330.3	362.1	524.3	276.1	200.6	302.5	254.8	406.0	28.0	308.1	696.4	440.5	322.1	221.2	295.0	334.9
260/280	1.3	1.2	1.1	1.1	1.4	1.3	1.2	1.3	1.6	1.2	1.1	1.4	-2.2	1.5	1.7	1.1	1.2	1.1	1.2	1.1
260/230	0.8	0.7	-0.2	1.2	0.9	1.1	1.1	0.8	0.7	0.8	-0.7	0.8	-0.1	0.8	0.7	1.3	0.8	0.8	0.9	0.7
CTAB extraction																				
ng/ul	43.9	102.8	202.9	68.1	393.6	252.7	123.6	66.9	390.6	174.5	228.5	380.7	50.0	65.0	1059.0	317.5	95.5	135.9	83.5	222.9
260/280	1.9	1.9	1.7	2.2	2.0	1.9	1.8	2.3	2.1	1.8	2.0	2.0	1.9	1.5	2.1	1.9	2.0	1.8	1.9	1.9
260/230	1.4	1.7	1.2	1.0	2.0	1.5	1.4	0.7	2.3	1.3	1.8	1.9	1.5	0.6	2.2	2.0	2.0	1.4	1.8	1.6
SILEX extraction																				
ng/ul	61.6	61.3	232.2	89.9	339.0	144.3	95.7	67.6	122.5	86.9	339.1	281.5	53.5	67.1	692.9	101.4	42.8	87.7	55.4	159.1
260/280	2.3	2.2	1.9	2.3	2.1	2.1	2.0	2.4	2.2	1.9	1.9	2.0	2.4	2.1	2.1	2.2	2.3	2.1	2.1	2.1
260/230	0.2	0.2	0.5	0.2	0.8	0.4	0.3	0.2	0.4	0.3	0.3	0.7	0.2	0.2	1.2	0.3	0.2	0.3	0.2	0.4

Observing the table, we can see that for the alkaline lysis method the values of the DNA concentration are all above 20 ng/µl. The only value that was closer to this number was the one from apple, the higher value of concentration obtained was 696.4 ng/µl for the asparagus spear. For the CTAB and SILEX the DNA concentration values are all above 20 ng/µl, for these two protocols the higher DNA extraction value was also observed for the asparagus spear the lowest were 43.9 from the recombinants and 42.8 from pear, respectively, for each protocol.

For the A260/280 and A260/230 ratios we can clearly see that the values obtained in the alkaline lysis protocol were the worst in the generality of the results. The lower value obtained in this protocol for A260/280 ratio was -2.2 for apple and for A260/230 ratio -0.8 from the strawberry octoploid. The only considerable values (under 1.5) were observed in cannabis, asparagus leaves and spear. In the CTAB and SILEX extraction method all the values for the A260/280 ratio are equal or superior to 1.5, with the higher value observed in the apricot old leaves (2.3) and the lowest (1.5) in the asparagus leaves for the CTAB method, and the highest value (2.4) in the apricot old leaves and apple and the lower (1.9) for peach young leaves and both strawberry samples in the SILEX method.

In the A260/230 ratio the values were all under 1.5 with the lower number (0.2) for the majority of the samples and the highest (1.2) for asparagus spear. The values for the CTAB were all above 1.5 or very closer to it with the exception of the value obtained for the asparagus leaves (0.6).

To assure the Nanodrop results, three agarose gels were performed with the DNA samples that resulted from each protocol of extraction (Figure 18 and 19), the samples were loaded in the same order as in the 96-well plate (notice that for the table 19 we used average results of the samples, in the gel we have a band representing every sample extracted) and the λ/Hind III Eco RI was used to confirm the DNA concentration. For the alkaline lysis method, the bands were not visible in the resulting gel, probably due to the high levels of impurity's in the samples provided, as we can see also by the values given by the Nanodrop for the A260/280 and A260/230 ratios. For the CTAB some operational mistakes were made but through the resulting gel we can confirm the concentration of the DNA extracted (Figure 18).

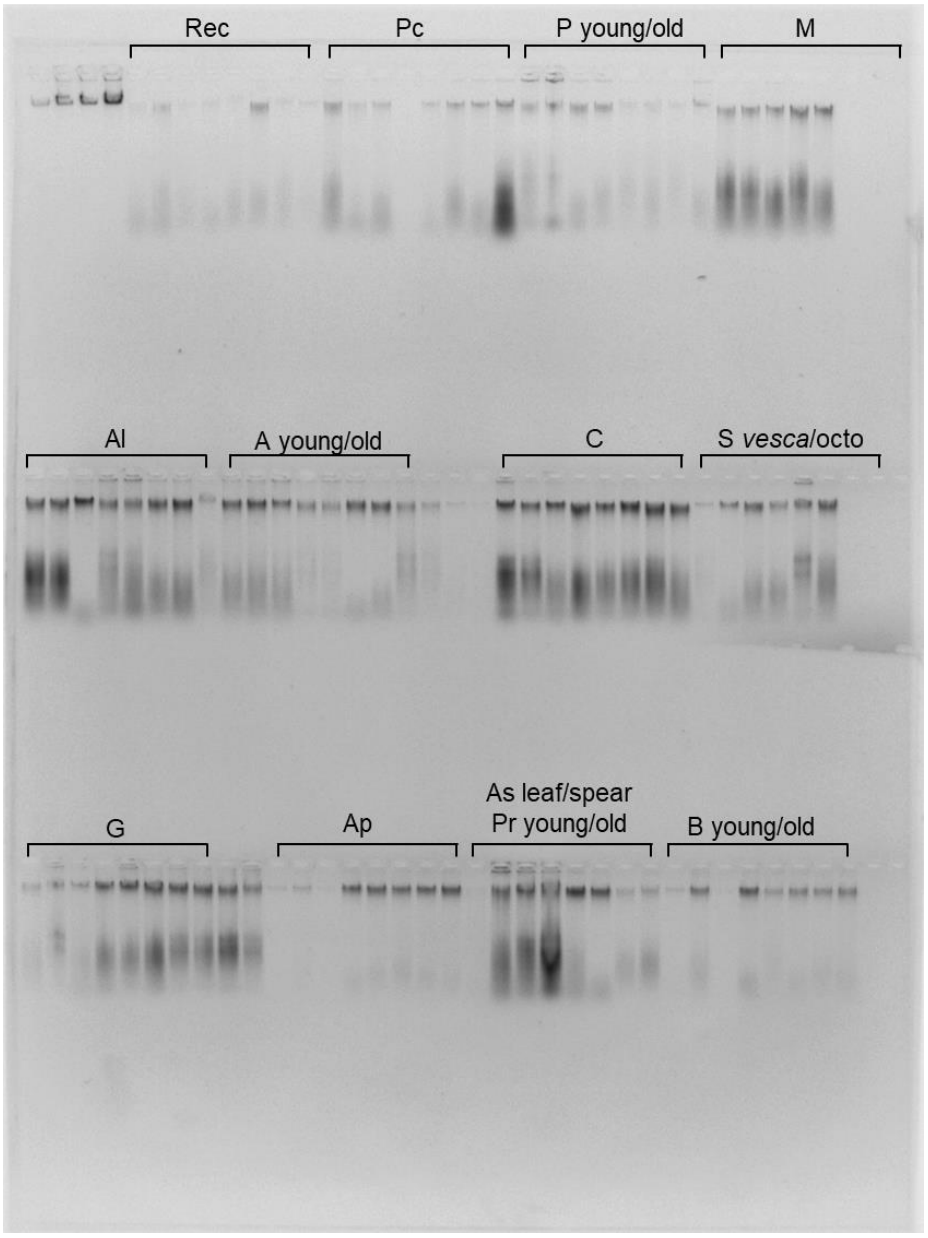

Figure 18. Agarose gel that shows the genotype of the samples used to test the three DNA extraction protocols, as result of the CTAB extraction method. In the first 4 wells is the DNA ladder ʎ/Hind III Eco R I in the concentrations of 1, 2, 4 and 6 µl.

Through the DNA ladder ʎ we can know the DNA concentration of each sample on the agarose gel, 1 µl of the ladder correspond to a concentration of 50 ng, 2 µl to 100 ng, 4 µl to 200 ng and 6 µl to 300 ng. This way we can see that the general results observed in the gel are accordingly to the ones obtained with the Nanodrop, the lower concentration value was obtained for the recombinants (43.9 ng/ µl) and we can see by the intensity of the bands in the gel corresponding to the recombinants that they match with the intensity of the DNA ladder at 1 µl, which means that the samples have less than 50 ng/µl. The samples with the higher DNA concentration in the Nanodrop for the CTAB protocol are the ones from the asparagus spear (1059 ng/µl) as we can confirm in the second, third and fourth sample walls present in the gel

in the asparagus section, in the Nanodrop results the concentration for this sample was higher than the rest and in the gel we also see a very intense band, although is not possible to compare it with the DNA ladder once we only have a representation from the ladder until 300 ng.

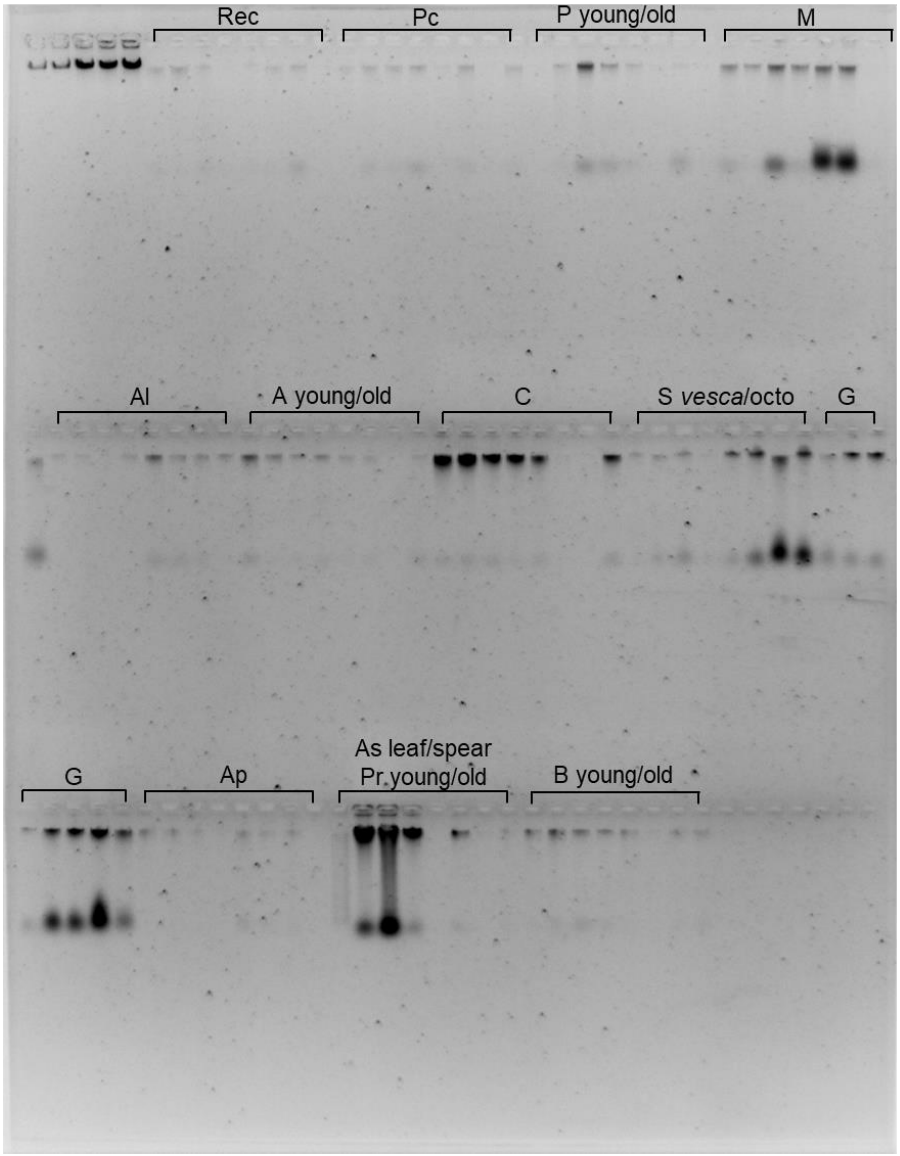

Figure 19. Agarose gel that shows the genotype of the samples used to test the three DNA extraction protocols, as result of the SILEX extraction method. In the first 5 wells is the DNA ladder λ/Hind III Eco R I in the concentrations of 1, 2, 4, 6 and 8 µl.

Comparing the results for the SILEX (Figure 19) extraction method not all the samples seam so clear as the ones extracted with the CTAB, but again we see the lower concentration for the pear samples and that is also showed in the gel with very light band corresponding to the 1 µl of the DNA ladder, the higher concentrations are again for the asparagus spear samples represented again in the gel by the band with the higher intensity corresponding to 500 ng of the DNA ladder.

5. DISCUSSION

1st Experiment

5.1. Number of recombinant individuals obtained

We estimated the expected number of recombinant individuals according to the genetic distance between the two markers that were used in the first round of screenings (Table).

Table 20. Results for the expected and observed number of recombinant individuals in the first screening.

First round of selection			
Gene	**Distance (cM)**	**Expected nº**	**Observed nº**
Alf	7	153	239
Jui	9	196	199
DBF2	8	175	112

The expected number was calculated using the distance between the two markers used in the first round of selection for each trait. One centimorgan means that there is one recombination each 100 individuals. Therefore, when the genetic distance is 7 cM we expect 7 recombinant individuals each 100 individuals analyzed. In the present study we should obtain a lower number of recombinants because we only selected a subgroup of the total number of recombinations that occurred because some of them were not informative for the fine mapping depending on the genetic behavior of each gene (dominant or recessive). Hear we obtained 2183 individuals that were analyzed for the three traits. For *Alf* and *DBF2* we observed a number of recombinants considerably different, that was higher for *Alf* and lower for *DBF2*.

This can be explained because in same plates that were not working with the closer marker were performed with the next marker known available in the laboratory, and as the distance between the markers increases the number of possible recombinants to for *Alf*. This can also be explained because the resolution of this approach is not so high as is for the fine mapping or because the markers, PCR or GeneMapper and GeneMarker softwares didn't work properly. For *Jui* the expected and observed numbers are very close.

5.2. Efficiency of the different molecular markers

Focusing on the molecular markers used, we know that SSRs, SNPs and InDels are considered to be very valuable, and this kind of markers appear in large number in model crops as *Prunus persica*, covering a big part of the genome. SSRs are repeats of up to 100 times of simple sequences of 1–8 base pairs [74], while with SNPs and its associates InDels

are a variation in the DNA sequence that affects only one base in the genome sequence between individuals of a specie or between pairs of chromosomes of an individual [102].

From the user point of view, SSRs methodology takes time, which makes its efficiency lower when the objective is to analyze a large number of samples. The main reason why we used them is because we can genotype them in our laboratory, which make them cheaper for us.

They were used mainly in the first round of selection when we had to analyze 2183 seedlings corresponding to 24 plates of 96-wells, for two or three makers, when needed, of the three traits under study, which makes at least a total of 12 PCR plates for each one of the 24 initial plates, although the PCR plates have often been repeated to correct operators or machinery errors. Beyond this the preparation of the master mix and the PCR on the thermocycler for a SSR primer with a tag takes more than 3 hours, which is an inefficient time for a day of work, without counting the time dispend in the ABI Genetic Analyzer and analyzing the samples in the GeneMarker and GeneMapper software's. The markers fulfilled their goal since they allowed the selection of only 550 possible recombinants reducing the initial sample number almost 75%.

To produce the InDels and SNPs we need to send them to a proper service that construct them, and all this process is expensive and time consuming. InDels methodology was not an option for the first round of selection for the reason referred previously and because InDels need agarose gels that is very time consuming for such a big number of samples. The preparation of the master mix and the PCR were also needed to amplify the DNA quantity although the protocol used for the InDels was the same one as for the markers with a fluorochrome so less cycles were needed to increase the same amount of DNA. The next step was the agarose gel electrophoresis which was also very time consuming because after loading the samples the gel always as to run at least 45 minutes before it can be analyzed in the *QuantumCapture* software.

In the end the most efficient protocol was the one performed for the SNPs, despite the fact that we cannot produce them. The PACE genotyping master mix was always prepared in the laboratory, because its common, we only mixed it with the DNA samples and let it to run in the LightCycler 480 device for 2 hours. In the end it takes 10 minutes to analyze the results in the Endpoint Genotyping. These two last markers described allowed us the saturation with many different markers in different positions and reduction of the number of possible recombinants to 20 in the total of the three traits.

5.3. *Alf* and *Jui* fine mapping

As referred previously almond and peach can be crossed and produce fertile offspring. But they are distinct species in many aspects: they have been selected under domestication in different environments and for different characteristics, peach is a fleshy fruit while from almond we select the seed (core) to eat. Almond is self-incompatible, highly variable, adapted to hot and dry conditions while peach is self-compatible, with a narrow gene pool and requiring optimal growing conditions for production (temperate climate). It has been shown that almond could be a good source of genes for peach breeding, particularly those related with disease resistance and abiotic stress, which are scarce in the peach commercial germplasm [143]. *Alf* and *Jui* are two major genes that determinate crucial differences between peach and almond fruits and introgressing these two genes could be a good way to engineer plants that have the most of almond variability and produce an edible fleshy fruit [13,134]. For *Alf* and *Jui* the direct application in breeding programs is not so clear as for *DBF2*. But there is a project called 'ALMELO' running in the same group where this research is being developed. These genes also seam extremely interesting to further understand the domestication process and the differences between peach and almond.

For *Alf* there are 37 candidate genes. One of them is annotated as NAC gene. NAC transcription factors are involved in fruit growth, development, senescence, ripening, stress response, and secondary cell wall and vascular development [144]. In an approach conduced with *Fragaria* × *ananassa* fruits, six strawberry NAC proteins were found to play different important regulatory roles in the process of development and ripening of the fruit, providing the basis for further functional studies [145]. PAP-specific phosphatase HAL2-like and Receptor-like protein kinase THESEUS 1 were also annotated candidate genes found trough fine mapping of a major locus controlling maturity date in peach [73], so these three genes seem to be the best candidate genes for this trait.

Exploring the list of candidate genes annotated for *Jui* it was not possible to find any bibliography refereed to them as responsible for fruit development or texture, so we have to wait for the phenotyping of the recombinants found this year to narrow down the target region or do a candidate gene validation through the approaches explained below.

5.4. *DBF2* fine mapping

Anthocyanins are flavonoid pigments that are responsible for the red, blue, and purple pigmentation of fruits, flowers, foliage, seeds, and roots [146]. The accumulation of anthocyanin pigments in fruit and vegetables is also an important determinant of ripeness and quality.

Anthocyanin-rich fruits and vegetables are brighter and more attractive to consumers, making them more marketable. Recently, anthocyanins have attracted even more interest from the public and the research community due to their potential to positively impact human health based on their antioxidant properties. It has also been suggested that anthocyanins can reduce the incidence of certain cancers, coronary heart diseases, oxidative stress, and other age-related diseases [147–151]. Their association with human health, particularly their capacity for pigmentation, has led to an increased demand among consumers for anthocyanin-containing products and to plant breeders doubling their efforts to identify natural foods, especially fruits that are rich in anthocyanins. For these reasons, the investigation of anthocyanin biosynthesis in fruit has become a popular area of research, with the aim of developing novel fruit cultivars with a higher anthocyanin content [152].

Some peach cultivars exhibit an intensely pigmented fruit mesocarp with a red-violet color, referred to here as blood-flesh peach, that represented an attractive starting point for the development of novel fruit varieties with a high anthocyanin content. Three different genes have been identified for red flesh in peach. Werner et al. (1998) demonstrated that the blood-flesh phenotype of 'Harrow Blood' cultivar was controlled by a single recessive locus, designated *bf* (blood-flesh) and Gillen and Bliss (2005) mapped the *bf* locus to the top of G4. Shen et al. (2013) results are in agreement with a model based on a single dominant gene controlling the blood-flesh trait and the trait was been referred after that as *DBF* (Dominant Blood-Flesh) and located on G5. Our gene of interest, *DBF2*, was mapped in the G1 [13]. The first two genes described have been mapped but there are no candidate genes available. For *DBF2*, we were able to fine mapped it to a region with only one candidate gene annotated as UDP-glycosyltransferase 85A2. As far as we know this was the first time that fine mapping in trees was performed to this level of resolution. It takes a lot of years and work, but is possible to get to this level of specificity with fine mapping in trees.

Glycosyltransferases are found in all living organisms, catalyzing the transfer of a glycosyl moiety from an activated donor to an acceptor molecule, forming a glycosidic bond. These glycosyl transfer reactions have been highlighted as the most important biotransformation on earth, since in quantitative terms they account for the assembly and degradation of the bulk of biomass [153]. A unique signature motif has been identified in the amino acid sequence of many of these glycosyltransferases, leading to their classification into a single UDP-glycosyltransferase [154]. Anthocyanins are major secondary metabolites that are responsible for color variation in plants and that exhibit health-promoting properties [155], several recent studies have demonstrated the involvement of UDP-glycosyltransferases in the coloration of different species, Peng et al. (2020) found 8 UDP- glycosyltransferases (UDP-

glycosyltransferase 88B1, UDP-glycosyltransferase 73C3, UDP-glycosyltransferase 90A1, UDP-glycosyltransferase 76F1, UDP-glycosyltransferase 88B1, UDP-glycosyltransferase 89A2, UDP-glycosyltransferase 73E1, UDP-glucose iridoid glucosyltransferase) expressed genes encoding biosynthetic enzymes potentially involved in the Kiwifruit flavonoid and anthocyanin pathway. Ni et al. (2018) also reported that the activity of UDP-glucose flavonoid-3-O-glycosyltransferase enzyme was significantly higher in all red-skinned cultivars, of Japanese apricot, suggesting that it is the potential vital regulatory gene for biosynthesis of anthocyanin, in this work the UDP-glycosyltransferase 85A2 was reported as a potential candidate gene for this characteristic. In *Vitis vinifera* the same candidate gene (UDP-glycosyltransferase 85A2), was reported as one of the responsible for the red leaf coloration [156]. All of these works strengthen the idea that the gene that we found is responsible for the Dominant Blood Flesh trait in peach and in the future, it can be used to develop new commercial varieties with red flesh.

5.5. Candidate genes validation

Functional validation in peach is difficult because peach transformation is very inefficient. For this reason, we should reduce the number of candidate genes as much as possible. An alternative to peach transformation is to analyze the effect of the polymorphism available in the candidate genes between the two parentals, using the SNPeffect software. The genes with no polymorphisms could be discarded as candidates. Another approach is expression analysis to see if candidate genes are expressed in the appropriate tissues and developmental stages according to the phenotypes produced.

5.5.1. Compare the sequence of almond and peach and look for polymorphism's trough SNPeffect

The information produced by new recombinant individuals helps to trace down a gene by shortening the region known to contain it. Nevertheless, this process reaches a point where the expected number of recombination is low due to the small size of this genomic region.

To overpass this situation, the *SnpEff* software can be used to identify all the different polymorphisms between parental almond's and peach's using genomes by comparing two genomics regions. The output of this program predicts the impact of the nucleotide difference observed, classifying its effect in an increasing scale of modifier, low, moderate and high impact.

SNPeffect analysis was done previously in the group were this work was developed for *Jui* and *DBF2*. For *Jui* three candidate genes with high impact variants were annotated as aspartic-type endopeptidase activity enzymes. Specifically, as xylanase inhibitor. These enzymes degrade the linear polysaccharide xylan into xylose, part of the hemicellulose, a major component of plant cell walls. Through this previous approach, we concluded that peach's alleles are longer than the ones from almond because of the loss of a stop codon (gained 23 bp later) and the appearance of an early starting codon (448 bp earlier than in almond). This can lead to an incorrect fold of the final protein and therefore result in the loss of function of the xylanase inhibitor protein. This could mean that even when expressed and receiving the proper activation signaling, the inhibitor does not work and consequently xylanases are free to degrade the hemicelluloses present in the cell wall. If this is true, while almond can regulate and stop softening of the mesocarp when needed, peach cannot and continues to degrade the hemicellulose present in its cell walls.

SNPeffect also allowed in the past to predict and explore the presence of anthocyanins in the flesh and the role of *DBF2* in the synthesis pathway of red-pigments in a critical step that transforms anthocyanidins in anthocyanins. Allowing the discovery of several intervenient in the process, as the enzyme *UDP-glucose: flavonoid-3-O glucosyltransferase* (UFGT), *Leu-rich-repeat receptor-like kinase* and *MYB transcription factors*.

5.5.2. Expression analysis during the development

Expression analysis is used for determining gene functions and interactions among genes in the development and life cycles of organisms. The majority of experiments aiming to identify gene function or structure and the dynamics of gene regulatory networks have been conducted on a few model species, such as *Arabidopsis thaliana* or *Drosophila melanogaster*. Then, the obtained data are applied to other organisms. However, even for model species, only 20%–30% of genes have been the subjects of genetic experiments [157]; thereby, the utilization of additional data is essential for, at the very least, the indirect characterization of gene function. Detailed profiles of gene expression can be used for this purpose [158]. This approach is being applied in the group where this study was caried out through the extraction of RNA from peach fruits in different developmental stages.

5.5.3. Genetic transformation

Almost 30 years have passed since the first published report on the regeneration of transformed peach plants [159]. Nevertheless, the general applicability of genetic transformation to this species has not yet been established. In the absence of an efficient peach

transformation system, progress in determining gene function will remain slow. As an alternative, a highly efficient transformation method in European plum (*P. domestica L.)* has shown to be a useful tool for functional genomics studies in *Prunus* spp [160]. However, peach genetic engineering is not only significant for gene function studies. The lack of efficient peach genetic transformation protocols precludes the application in peach of new biotechnological tools such as RNA interference (RNAi), trans-grafting, cisgenesis/intragenesis, or genome editing in peach breeding programs, as are currently being applied in other fruit tree species [161], and can be applied in futures approaches to validate candidate genes.

2nd Experiment

In terms of practicality comparing the three extraction methods, the alkaline extraction method seems to be very fast and low-cost which is a big advantage if we are dealing with a big volume of samples. Even though the DNA concentration is very high, its quality was not good, once we got the worst results for the A260/280 and A260/230 ratios in this protocol which means that the samples have big amounts of contaminants (proteins and organic compounds). For genotyping this protocol is very suitable, if we want to use it in fast screenings with PCR-based markers as SSRs or InDels (used in the materials and methods 3.2.), but might not work in species that were not exemplified in this work such as rice or wheat.

The CTAB method is one of the most extensively used worldwide for the DNA extraction in plants. Again, in this report, it seems to be the one that produced better results. In the generality of the samples, there was an equilibrium between the concentration of the DNA extracted and the amount of proteins and organic acids that can cause contamination in the samples. Taking into account the performance, the CTAB protocol takes a more time to be executed, and in general the groups that work in the laboratory have a large amount of samples to work with, which makes this protocol less efficient. However, DNA purity is much higher than in the alkaline lysis method, a characteristic that is sometimes necessary for procedures that will be performed after DNA extraction, such as DNA sequencing or genotyping.

The authors of the SILEX extraction method claim that this new approach is faster and inexpensive, and that works for different NGS applications, combining the advantages of commercial kits (high-quality DNA, fast and broad range of species spectrum) with those of a CTAB-based method (high yield and inexpensive) being suitable for routinely DNA screening. Although we did not obtain good results, we did only few extractions and perhaps the

methodology could be further improved, since the CTAB protocol was not providing quality DNA in some species under research in the laboratory.

From the point of view of the operator the initial steps of the SILEX compared with the CTAB are simpler. We prepared a big amount of all the reagents and buffers to be used in this extraction method in the beginning, and they can be saved for several months without deterioration. Other advantages are the fact that no toxic salts such as guanidinium thiocyanate or sodium iodide at high concentrations are used and the total cost of all the reagents and consumables is only 0.12 € per sample and for multiple simultaneous manual extractions, each sample requires less than 2 min per person. In this respect, in the SILEX method, the silica matrix used for each extraction cost less than 0.001 € and the washing buffer is only water and ethanol, a common non-toxic reagent in most molecular biology laboratories [28].

The reason why this protocol is faster is the absence of the precipitation step required in the CTAB. And it is indeed faster, and for us is practicality is in fact bigger than the one from the CTAB, but it also can be improved. For example, the original SILEX protocol is performed in Eppendorf's, which is suitable in the case of a few samples. To be applicable in our laboratory it should suit the extraction in a 96-well plate, because we always work with a big number of samples, and this is the reason why we adapted it. This way, the adaptations that we made could be one of the main reasons why the results obtained are not so good as the ones from the original protocol. We can see by the results that the DNA concentration value is between the optimal values and it is not contaminated with proteins, but the contamination with organic compounds was very high, which means that for example this DNA is not enough pure to be used in sequencing for example.

6. CONCLUSIONS

In this work we fine mapped the *Alf*, *Jui* and *DBF2* genes already described by Donoso et al. (2016), that provided a first survey on the almond variability available for peach breeding. The search on these three major genes has been done for several years and some remarkable conclusions were achieved in this report.

- The use of SSRs was extremely useful and cost efficient in first round of selection for the analysis of a large number of samples. The InDels and SNPs have also showed to be very effective markers due to their abundance in genomes. SNPs were the most efficient markers in terms of hand labor.
- A total of 12, 4 and 3 new possible recombinants for *Alf*, *Jui* and *DBF2* respectively was found after the fine mapping with SSRs, Indels and SNPs. In the future their phenotype will allow to narrow down the regions of *Alf* and *Jui* major genes.
- After the phenotyping and marker saturation of recombinants already available in the laboratory from previous years, the target regions containing *Jui* and *DBF2* decreased form 132 and 377 kb to 69 and 1.4 kb respectively;
- The number of candidate genes for *Jui* droped thanks to the study of the SNPs present in almond and peach, and the number of candidate genes was reduced from 37 to 11.
- For *DBF2* the number of candidate genes also droped, from 13 to 1. And as far as we know we are facing the first case of fine mapping in trees that get so far in the resolution level.

The second experiment was a side project that emerged during the development of my work at CRAG as an opportunity to contribute with some valuable information that could streamline the work of all the research groups working in the same laboratory as me. All the species tested in this section, with the exception of blueberry, are species under study in the laboratory, and while testing and comparing the three protocols I wrote an adapted SILEX protocol to a 96-well plate (Annex 1) to be used by all the researchers when needed.

The main conclusions that we reached are based on the fact that the CTAB extraction method remains the best to be applied in the generality of the species that we tested, but we also tested the SILEX in the first place because for some species, as plum and apricot, the quality of the DNA extracted with the CTAB is not the best for the needed applications. Although we didn't obtain good results for the SILEX, we did only few extractions and perhaps the methodology could be further improved or better adapted to the species where the CTAB doesn't work so well.

7. List of References

1. Shulaev, V., Korban, S. S., Sosinski, B., Abbott, A. G., Aldwinckle, H. S., Dandekar, A. M., Lewers, K., Brown, S. K., Folta, K. M., Iezzoni, A., Main, D., Aru, P., Davis, T. M., Gardiner, S. E., Potter, D., & Veilleux, R. E. (2008). *Multiple Models for Rosaceae Genomics [OA]*. *147*(July), 985–1003. https://doi.org/10.1104/pp.107.115618

2. Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M. T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J., Vendramin, E., Meisel, L. A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., … Rokhsar, D. S. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, *45*(5), 487–494. https://doi.org/10.1038/ng.2586

3. Arús, P., Verde, I., Sosinski, B., Zhebentyayeva, T., & Abbott, A. G. (2012). The peach genome. *Tree Genetics and Genomes*, *8*(3), 531–547. https://doi.org/10.1007/s11295-012-0493-8

4. Aranzana, M. J., Decroocq, V., Dirlewanger, E., Eduardo, I., Gao, Z. S., Gasic, K., Iezzoni, A., Jung, S., Peace, C., Prieto, H., Tao, R., Verde, I., Abbott, A. G., & Arús, P. (2019). Prunus genetics and applications after de novo genome sequencing: achievements and prospects. *Horticulture Research*, *6*(1). https://doi.org/10.1038/s41438-019-0140-8

5. Scorza, R., Mehlenbacher, S. A., & Lightner, G. W. J. J. of the A. S. for H. S. (1985). *Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germplasm improvement*.

6. Li, X., Meng, X., Jia, H., Yu, M., Ma, R., Wang, L., Cao, K., Shen, Z., Niu, L., Tian, J., Chen, M., Xie, M., Arus, P., Gao, Z., & Aranzana, M. J. (2013). *Peach genetic resources : diversity , population structure and linkage disequilibrium*.

7. Micheletti, D., Dettori, M. T., Micali, S., Aramini, V., Pacheco, I., Da Silva Linge, C., Foschi, S., Banchi, E., Barreneche, T., Quilot-Turion, B., Lambert, P., Pascal, T., Iglesias, I., Carbó, J., Wang, L. R., Ma, R. J., Li, X. W., Gao, Z. S., Nazzicari, N., … Joséaranzana, M. (2015). Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLoS ONE*, *10*(9), 1–19. https://doi.org/10.1371/journal.pone.0136803

8. Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., Cheng, S., Zeng, P., Chen, C., Wang, X., Xie, M., Zhong, X., Wang, X., Zhao, P., Bian, C., Zhu, Y., Zhang, J., Ma, G.,

Chen, C., … Wang, J. (2014). Comparative population genomics reveals the domestication history of the peach, Prunus persica, and human influences on perennial fruit crops. *Genome Biology*, *15*(7), 1–15. https://doi.org/10.1186/s13059-014-0415-1

9. Byrne, D. H., Raseira, M. B., Bassi, D., Piagnani, M. C., Gasic, K., Reighard, G. L., Moreno, M. A., & Pérez, S. (2012). Peach. In *Fruit breeding* (pp. 505–569). Springer.

10. Quilot, B., Wu, B. H., Kervella, J., Génard, M., Foulongne, M., & Moreau, K. (2004). QTL analysis of quality traits in an advanced backcross between Prunus persica cultivars and the wild relative species P. davidiana. *Theoretical and Applied Genetics*, *109*(4), 884–897. https://doi.org/10.1007/s00122-004-1703-z

11. Carrillo-Mendoza, O., Chaparro, J. X., & Williamson, J. (2013). Branching and blind node incidence in interspecific backcross families of peach. *HortScience*, *48*(9), 1119–1124. https://doi.org/10.21273/hortsci.48.9.1119

12. Dirlewanger, E., Graziano, E., Joobeur, T., Garriga-Calderé, F., Cosson, P., Howad, W., & Arús, P. (2004). *Comparative mapping and marker-assisted selection in Rosaceae fruit crops*. *101*(26), 9891–9896. https://doi.org/10.1073/pnas.0307937101

13. Donoso, J. M., 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 interspecific almond × peach populations. *Molecular Breeding*, *36*(2), 1–17. https://doi.org/10.1007/s11032-016-0441-7

14. Ribeiro Serra, O. M. (2017). Towards increasing genetic variability and improving fruit quality in peach using genomic and bioinformatic tools. *TDX (Tesis Doctorals En Xarxa)*. http://www.tdx.cat/handle/10803/460882

15. Kasem, S., Rice, N., & Henry, R. J. J. P. genotyping I. S. N. P. technology. (2008). *14 DNA Extraction from Plant Tissue*. 219.

16. Kang, H. W., Cho, Y. G., Yoon, U. H., & Eun, M. Y. (1998). A Rapid DNA Extraction Method for RFLP and PCR Analysis from a Single Dry Seed. *Plant Molecular Biology Reporter*, *16*(1), 90–90. https://doi.org/10.1023/A:1007418606098

17. Von Post, R., Von Post, L., Dayteg, C., Nilsson, M., Forster, B. P., & Tuvesson, S. (2003). A high-throughput DNA extraction method for barley seed. *Euphytica*, *130*(2), 255–260. https://doi.org/10.1023/A:1022863006134

18. Abdel-Latif, A., & Osman, G. (2017). Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. *Plant Methods*, *13*(1), 1–9.

https://doi.org/10.1186/s13007-016-0152-4

19. Cota-Sánchez, J. H., Remarchuk, K., & Ubayasena, K. (2006). Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant Molecular Biology Reporter*, *24*(2), 161–167. https://doi.org/10.1007/BF02914055

20. Ibrahim, R. I. H. (2011). Article A modified CTAB protocol for DNA extraction from young flower petals of some medicinal plant species. *Geneconserve 10(40): 165-182*, *10*(40), 165–182.

21. Letebrhan, K., Marie, K., & Hewan, D. (2020). Comparison and optimization for DNA extraction of okra (Abelmoschus esculentus L. Moench). *African Journal of Biotechnology*, *19*(6), 353–361. https://doi.org/10.5897/ajb2019.16991

22. Irfan, M., Ting, Z. T., Yang, W., Chunyu, Z., Qing, M., Lijun, Z., & Feng, L. (2014). *Modification of CTAB protocol for maize genomic DNA extraction*. *May*.

23. Vaillancourt, B., & Buell, C. R. (2019). *High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing*. 1–13. https://doi.org/10.1101/783159

24. Healey, A., Furtado, A., Cooper, T., & Henry, R. J. (2014). Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods*, *10*(1), 1–8. https://doi.org/10.1186/1746-4811-10-21

25. Martínez-González, C. R., Ramírez-Mendoza, R., Jiménez-Ramírez, J., Gallegos-Vázquez, C., & Luna-Vega, I. (2017). Improved method for genomic DNA extraction for Opuntia Mill. (Cactaceae). *Plant Methods*, *13*(1), 1–10. https://doi.org/10.1186/s13007-017-0234-y

26. Barbier, F. F., Chabikwa, T. G., Ahsan, M. U., Cook, S. E., Powell, R., Tanurdzic, M., & Beveridge, C. A. (2019). A phenol/chloroform-free method to extract nucleic acids from recalcitrant, woody tropical species for gene expression and sequencing. *Plant Methods*, *15*(1), 9–12. https://doi.org/10.1186/s13007-019-0447-3

27. Illa, E., Eduardo, I., Audergon, J. M., Barale, F., Dirlewanger, E., Li, X., Moing, A., Lambert, P., Le Dantec, L., Gao, Z., Poëssel, J. L., Pozzi, C., Rossini, L., Vecchietti, A., Arús, P., & Howad, W. (2011). Saturating the Prunus (stone fruits) genome with candidate genes for fruit quality. *Molecular Breeding*, *28*(4), 667–682. https://doi.org/10.1007/s11032-010-9518-x

28. Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Fortea, E., Ferrante, P., Schmidt, M., Díez, M. J., Usadel, B., Giuliano, G., & Prohens, J. (2020). SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods*, *16*(1). https://doi.org/10.1186/s13007-020-00652-y

29. Kovačević, N. (2016). Magnetic Beads Based Nucleic Acid Purification for Molecular Biology Applications. In M. Micic (Ed.), *Sample Preparation Techniques for Soil, Plant, and Animal Samples* (pp. 53–67). Springer New York. https://doi.org/10.1007/978-1-4939-3185-9_5

30. Rana, M., Aycan, M., Takamatsu, T., & Kaneko, K. (2019). *Optimized Nuclear Pellet Method for Extracting Next-Generation Sequencing Quality Genomic DNA from Fresh Leaf Tissue*. 1–11.

31. Martin, S. L., Parent, J., Laforest, M., Page, E., Kreiner, J. M., & James, T. (2019). *Population Genomic Approaches for Weed Science*.

32. Lu, J., Hou, J., Ouyang, Y., & Luo, H. (2020). *A direct PCR – based SNP marker – assisted selection system (D-MAS) for different crops*. *2005*.

33. Abbott, A., Georgi, L., Yvergniaux, D., Inigo, M., Sosinski, B., Wang, Y., Blenda, A., & Reighard, G. J. A. H. (2002). *Peach: the model genome for Rosaceae*. *1*, 145–156.

34. Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., Kerr, M., Robertson, K. R., Arsenault, M., Dickinson, T. A., & Campbell, C. S. (2007). Phylogeny and classification of Rosaceae. In *Plant Systematics and Evolution* (Vol. 266, Issues 1–2). https://doi.org/10.1007/s00606-007-0539-9

35. McNeill, J., Barrie, F. R., Buck, W. R., Demoulin, V., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Marhold, K., & Prado, J. (2012). International Code of Nomenclature for algae, fungi and plants. *Regnum Vegetabile*, *154*.

36. Chin, S.-W., Shaw, J., Haberle, R., Wen, J., & Potter, D. (2014). Diversification of almonds, peaches, plums and cherries–molecular systematics and biogeographic history of Prunus (Rosaceae). *Molecular Phylogenetics and Evolution*, *76*, 34–48.

37. Scorza, R., Okie, W. R. J. G. R. of T. F., & 290, N. C. (1991). *Peaches (Prunus)*. 177–234.

38. Faust, M., & Timon, B. J. H. R. (1995). *Origin and dissemination of peach*. *17*, 331–379.

39. 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., & Xie, H. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nature Communications*, *2018*, 1–13. https://doi.org/10.1038/s41467-018-07744-3

40. Zheng, Y., Crawford, G. W., & Chen, X. J. P. one. (2014). *Archaeological evidence for peach (Prunus persica) cultivation and domestication in China*. *9*(9).

41. Su, T., Wilf, P., Huang, Y., Zhang, S., & Zhou, Z. J. S. reports. (2015). *Peaches preceded humans: fossil evidence from SW China*. *5*, 16794.

42. Xie, R., Li, X., Chai, M., Song, L., Jia, H., Wu, D., Chen, M., Chen, K., Aranzana, M. J., & Gao, Z. (2010). Evaluation of the genetic diversity of Asian peach accessions using a selected set of SSR markers. *Scientia Horticulturae*, *125*(4), 622–629.

43. Iglesias, I. (2018). *Tendencias e Innovación en Agronomía*.

44. U.S. Department of Agriculture. (2016). Stone Fruit Annual EU-28. *Global Agricultural Information Network*, *SP1618*.

45. Verde, I., Lauria, M., Dettori, M. T., Vendramin, E., Balconi, C., Micali, S., Wang, Y., Marrazzo, M. T., Cipriani, G., Hartings, H., Testolin, R., Abbott, A. G., Motto, M., & Quarta, R. (2005). Microsatellite and AFLP markers in the Prunus persica [L. (Batsch)]xP. ferganensis BC1linkage map: Saturation and coverage improvement. *Theoretical and Applied Genetics*, *111*(6), 1013–1021. https://doi.org/10.1007/s00122-005-0006-3

46. Bielenberg, D., Gasic, K., & Chaparro, J. X. (2009). An introduction to Peach (Prunus persica). In *Genetics and genomics of Rosaceae* (pp. 223–234). Springer.

47. Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Wang, J., Yang, W., Tao, Y., Yuan, Z., Fan, G., Xing, Z., Han, C., Pan, H., Zhong, X., Shi, W., Liang, X., Du, D., Sun, F., Xu, Z., … Wang, J. (2012). The genome of Prunus mume. *Nature Communications*, *3*(May), 1–8. https://doi.org/10.1038/ncomms2290

48. Salazar, J. A., Rasouli, M., Moghaddam, R. F., Zamani, Z., Imani, A., & Martínez-Gómez, P. (2014). Low-Cost Strategies for Development of Molecular Markers Linked to Agronomic Traits in Prunus. *Agricultural Sciences*, *05*(05), 430–439. https://doi.org/10.4236/as.2014.55044

49. Paterson, A. H. (1996). *Genome mapping in plants*. (RG Landes., p. 330).

50. López Girona, E. (2014). *Genetic architecture of agronomic traits in peach [Prunus persica (L.) Batsch]: subacid, flat shape and nectarine*. Universitat Autònoma de

Barcelona.

51. Lucht, J. M. (2015). Public acceptance of plant biotechnology and GM crops. *Viruses*, *7*(8), 4254–4281. https://doi.org/10.3390/v7082819

52. Lateef, D. D. (2015). DNA Marker Technologies in Plants and Applications for Crop Improvements. *Journal of Biosciences and Medicines*, *03*(05), 7–18. https://doi.org/10.4236/jbm.2015.35002

53. Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1–2), 169–196. https://doi.org/10.1007/s10681-005-1681-5

54. Garrido-Cardenas, J. A., Mesa-Valle, C., & Manzano-Agugliaro, F. (2018). Trends in plant research using molecular markers. *Planta*, *247*(3), 543–557. https://doi.org/10.1007/s00425-017-2829-y

55. Xiong, J.-S., Ding, J., & Li, Y. J. H. research. (2015). *Genome-editing technologies and their potential application in horticultural crop breeding*. *2*(1), 1–10.

56. Eathington, S. R., Crosbie, T. M., Edwards, M. D., Reiter, R. S., & Bull, J. K. J. C. S. (2007). *Molecular markers in a commercial breeding program*. *47*, S-154-S-163.

57. Sasaki, T., & Burr, B. J. C. opinion in plant biology. (2000). *International Rice Genome Sequencing Project: the effort to completely sequence the rice genome*. *3*(2), 138–142.

58. Consortium, T. G. (2012). *The tomato genome sequence provides insights into fleshy fruit evolution*. *485*(7400), 635.

59. Garrido-Cardenas, J. A., Garcia-Maroto, F., Alvarez-Bermejo, J. A., & Manzano-Agugliaro, F. J. S. (2017). *DNA sequencing sensors: an overview*. *17*(3), 588.

60. Varshney, R. K., Graner, A., & Sorrells, M. E. J. T. in B. (2005). *Genic microsatellite markers in plants: features and applications*. *23*(1), 48–55.

61. Grover, A., & Sharma, P. C. J. C. reviews in biotechnology. (2016). *Development and use of molecular markers: past and present*. *36*(2), 290–302.

62. Nagaty, M. A., El-Assal, S. E., & Rifaat, M. M. J. A. J. O. A. S. (2011). *Characterization of the genetic diversity of peach cultivars in Taif by RAPD-PCR*. *8*(7), 708–715.

63. Bouhadida, M., & Mart, J. P. (2007). *Chloroplast DNA Diversity in Prunus and Its Implication on Genetic Relationships*. *132*(5), 670–679.

64. Warburton, M. L., Bliss, F. A., & Davis, C. (1996). *Genetic Diversity in Peach (Prunus persica L . Batch) Revealed by Randomly Amplified Polymorphic DNA (RAPD) Markers and Compared to Inbreeding Coefficients*. *121*(6), 1012–1019.

65. Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M. J., Poizat, C., Zanetto, A., Arús, P., & Laigret, F. (2002). Development of microsatellite markers in peach [Prunus persica (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (Prunus avium L.). *Theoretical and Applied Genetics*, *105*(1), 127–138. https://doi.org/10.1007/s00122-002-0867-7

66. Vilanova, S., Sargent, D. J., Arús, P., & Monfort, A. (2008). Synteny conservation between two distantly-related Rosaceae genomes: Prunus (the stone fruits) and Fragaria (the strawberry). *BMC Plant Biology*, *8*, 1–12. https://doi.org/10.1186/1471-2229-8-67

67. Lu, Z.-X., Sosinski, B., Reighard, G. L., Baird, W. V., & Abbott, A. G. (1998). Construction of a genetic linkage map and identification of AFLP markers for resistance to root-knot nematodes in peach rootstocks. *Genome*, *41*(2), 199–207. https://doi.org/10.1139/gen-41-2-199

68. Aranzana, M. J., Pineda, A., Cosson, P., Dirlewanger, E., Ascasibar, J., Cipriani, G., Ryder, C. D., Testolin, R., Abbott, A., King, G. J., Iezzoni, A. F., & Arús, P. (2003). A set of simple-sequence repeat (SSR) markers covering the Prunus genome. *Theoretical and Applied Genetics*, *106*(5), 819–825. https://doi.org/10.1007/s00122-002-1094-y

69. Rojas, G., Méndez, M. A., Muñoz, C., Lemus, G., & Hinrichsen, P. (2008). Identification of a minimal microsatellite marker panel for the fingerprinting of peach and nectarine cultivars. *Electronic Journal of Biotechnology*, *11*(5), 1–12. https://doi.org/10.2225/vol11-issue5-fulltext-1

70. 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 Biology*, *9*, 1–14. https://doi.org/10.1186/1471-2229-9-59

71. Dhanapal, A. P., Martínez-García, P. J., Gradziel, T. M., & Crisosto, C. H. (2012). First genetic linkage map of chilling injury susceptibility in peach (Prunus persica (L.) Batsch) fruit with SSR and SNP markers. *Journal of Plant Science and Molecular Breeding*, *1*(1), 3. https://doi.org/10.7243/2050-2389-1-3

72. Martínez-García, P. J., Fresnedo-Ramírez, J., Parfitt, D. E., Gradziel, T. M., & Crisosto, C. H. (2013). Effect prediction of identified SNPs linked to fruit quality and chilling injury

in peach [Prunus persica (L.) Batsch]. *Plant Molecular Biology*, *81*(1–2), 161–174. https://doi.org/10.1007/s11103-012-9989-8

73. Pirona, R., Eduardo, I., Pacheco, I., Da Silva Linge, C., 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 Biology*, *13*(1). https://doi.org/10.1186/1471-2229-13-166

74. Hamada, H., & Kakunaga, T. J. N. (1982). *Potential Z-DNA forming sequences are highly dispersed in the human genome*. *298*(5872), 396–398.

75. Provan, J., Powell, W., & Hollingsworth, P. M. (2001). Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution*, *16*(3), 142–147. https://doi.org/10.1016/S0169-5347(00)02097-8

76. Chung, S. M., Staub, J. E., & Chen, J. F. J. T. A. G. (2003). *The development and evaluation of consensus chloroplast SSRs for chloroplast genetic analysis*. *107*, 757–767.

77. Buschiazzo, E., & Gemmell, N. J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays*, *28*(10), 1040–1050. https://doi.org/10.1002/bies.20470

78. Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F., & Makova, K. D. (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research*, *18*(1), 30–38. https://doi.org/10.1101/gr.7113408

79. Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, *29*(2), 294–307. https://doi.org/10.1590/S1415-47572006000200018

80. Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

81. Fan, L., Zhang, M. Y., Liu, Q. Z., Li, L. T., Song, Y., Wang, L. F., Zhang, S. L., & Wu, J. (2013). Transferability of Newly Developed Pear SSR Markers to Other Rosaceae Species. *Plant Molecular Biology Reporter*, *31*(6), 1271–1282. https://doi.org/10.1007/s11105-013-0586-z

82. Jonah, P. M., Bello, L. L., Lucky, O., Midau, A., & Moruppa, S. M. (2011). *Review : The Importance of Molecular Markers in*. *11*(5).

83. Garcia, A. A. F., Kido, E. A., Meza, A. N., Souza, H. M. B., Pinto, L. R., Pastina, M. M., Leite, C. S., Da Silva, J. A. G., Ulian, E. C., Figueira, A., & Souza, A. P. (2006). Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics*, *112*(2), 298–314. https://doi.org/10.1007/s00122-005-0129-6

84. Zhao, D. wei, Yang, J. bo, Yang, S. xiong, Kato, K., & Luo, J. ping. (2014). Genetic diversity and domestication origin of tea plant Camellia taliensis (Theaceae) as revealed by microsatellite markers. *BMC Plant Biology*, *14*(1), 1–12. https://doi.org/10.1186/1471-2229-14-14

85. Cipriani, G., Marrazzo, M. T., Di Gaspero, G., Pfeiffer, A., Morgante, M., & Testolin, R. (2008). A set of microsatellite markers with long core repeat optimized for grape (Vitis spp.) genotyping. *BMC Plant Biology*, *8*, 1–13. https://doi.org/10.1186/1471-2229-8-127

86. Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: An overview of the recent progress in plants. *Euphytica*, *177*(3), 309–334. https://doi.org/10.1007/s10681-010-0286-9

87. Zargar, S. M., Raatz, B., Sonah, H., Muslimanazir, Bhat, J. A., Dar, Z. A., Agrawal, G. K., & Rakwal, R. (2015). Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology*, *18*(5), 293–308. https://doi.org/10.1007/s12892-015-0037-5

88. Edh, K., Widén, B., & Ceplitis, A. (2007). Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic Brassica cretica (Brassicaceae). *Molecular Ecology*, *16*(23), 4972–4983. https://doi.org/10.1111/j.1365-294X.2007.03585.x

89. Harrigan, R. J., Mazza, M. E., & Sorenson, M. D. (2008). Computation vs. cloning: Evaluation of two methods for haplotype determination. *Molecular Ecology Resources*, *8*(6), 1239–1248. https://doi.org/10.1111/j.1755-0998.2008.02241.x

90. Ramya, A. R., Ahamed, M. L., Satyavathi, C. T., Rathore, A., Katiyar, P., Bhasker Raj, A. G., Kumar, S., Gupta, R., Mahendrakar, M. D., Yadav, R. S., & Srivastava, R. K. (2018). Towards defining heterotic gene pools in pearl millet [pennisetum glaucum (L.) R. Br.]. *Frontiers in Plant Science*, *8*(March), 1–11. https://doi.org/10.3389/fpls.2017.01934

91. Gao, H., Jiang, K., Geng, Y., & Chen, X. Y. (2012). Development of microsatellite

primers of the largest seagrass, Enhalus acoroides (Hydrocharitaceae). *American Journal of Botany*, *99*(3), 99–101. https://doi.org/10.3732/ajb.1100412

92. Mohan Jain, S., & Brar, D. S. (2009). Molecular techniques in crop improvement: 2nd edition. In *Molecular Techniques in Crop Improvement: 2nd Edition: Vol. V* (Issue December 2016). https://doi.org/10.1007/978-90-481-2967-6

93. Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. de F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

94. Maris, L., & Ritter, O. (2014). Application of Microsatellite Molecular Markers in Studies of Genetic Diversity and Conservation of Plant Species of Cerrado. *Journal of Plant Sciences*, *1*(1), 1. https://doi.org/10.11648/j.jps.20130101.11

95. Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, *32*(2), 261–285. https://doi.org/10.1080/13102818.2017.1400401

96. Adal, A. M., Demissie, Z. A., & Mahmoud, S. S. (2015). Identification, validation and cross-species transferability of novel Lavandula EST-SSRs. *Planta*, *241*(4), 987–1004. https://doi.org/10.1007/s00425-014-2226-8

97. Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., & Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, *156*(2), 847–854. https://doi.org/10.1007/s11295-010-0287-9

98. Morgante, M., Hanafey, M., & Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, *30*(2), 194–200. https://doi.org/10.1038/ng822

99. Zhao, X., Tan, Z., Feng, H., Yang, R., Li, M., Jiang, J., Shen, G., & Yu, R. (2011). Microsatellites in different Potyvirus genomes: Survey and analysis. *Gene*, *488*(1–2), 52–56. https://doi.org/10.1016/j.gene.2011.08.016

100. Kim, J. J., Han, B. G., Lee, H. I., Yoo, H. W., & Lee, J. K. (2010). Development of SNP-based human identification system. *International Journal of Legal Medicine*, *124*(2), 125–131. https://doi.org/10.1007/s00414-009-0389-9

101. Ganal, M. W., Polley, A., Graner, E. M., Plieske, J., Wieseke, R., Luerssen, H., & Durstewitz, G. (2012). Large SNP arrays for genotyping in crop plants. *Journal of Biosciences*, *37*(5), 821–828. https://doi.org/10.1007/s12038-012-9225-3

102. Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. https://doi.org/10.1038/nrg3012

103. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., … Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science*, *280*(5366), 1077–1082. https://doi.org/10.1126/science.280.5366.1077

104. Aranzana, M. J., Illa, E., Howad, W., & Arús, P. (2012). A first insight into peach [Prunus persica (L.) Batsch] SNP variability. *Tree Genetics and Genomes*, *8*(6), 1359–1369. https://doi.org/10.1007/s11295-012-0523-6

105. Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D., Cuguen, J., Bergelson, J., & Roux, F. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. *Molecular Ecology*, *22*(16), 4222–4240. https://doi.org/10.1111/mec.12396

106. Morgil, H., Gercek, Y. C., & Tulum, I. (2020). Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. In *The Recent Topics in Genetic Polymorphisms*. IntechOpen.

107. Brennan, A. C., Méndez-Vigo, B., Haddioui, A., Martínez-Zapater, J. M., Picó, F. X., & Alonso-Blanco, C. (2014). The genetic structure of Arabidopsis thaliana in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biology*, *14*(1). https://doi.org/10.1186/1471-2229-14-17

108. Freudenthal, J. A., Ankenbrand, M. J., Grimm, D. G., & Korte, A. (2019). *The GWAS-Flow Software*. 1–7.

109. Somerville, C., & Koornneef, M. (2002). A fortunate choice: The history of Arabidopsis as a model plant. *Nature Reviews Genetics*, *3*(11), 883–889. https://doi.org/10.1038/nrg927

110. Lopes, W. S. (2018). *Tendências da literatura científica sobre uso de marcadores*

moleculares em estudos de diversidade genética de peixes. 52.

111. Weber, J. L., David, D., Heil, J., Fan, Y., Zhao, C., & Marth, G. (2002). Human diallelic insertion/deletion polymorphisms. *American Journal of Human Genetics*, *71*(4), 854–862. https://doi.org/10.1086/342727

112. Tao, S., Fan, Y., Wang, W., Ma, G., Liang, L., & Shi, Q. (2008). Patterns of Insertion and Deletion in Mammalian Genomes. *Current Genomics*, *8*(6), 370–378. https://doi.org/10.2174/138920207783406479

113. Hyten, D. L., Cannon, S. B., Song, Q., Weeks, N., Fickus, E. W., Shoemaker, R. C., Specht, J. E., Farmer, A. D., May, G. D., & Cregan, P. B. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics*, *11*(1). https://doi.org/10.1186/1471-2164-11-38

114. Jander, G., Norris, S. R., Rounsley, S. D., Bush, D. F., Levin, I. M., Last, R. L., Llc, C. G., & Street, S. (2002). *Arabidopsis Map-Based Cloning in the Post-Genome Era*. https://doi.org/10.1104/pp.003533.responsible

115. Yang, J., He, J., Wang, D., Shi, E., Yang, W., Geng, Q., & Wang, Z. (2016). Progress in research and application of InDel markers. *Biodiversity Science*, *24*(2), 237–243. https://doi.org/10.17520/biods.2015205

116. Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

117. Ramalingam, J., Savitha, P., Alagarasan, G., Saraswathi, R., & Chandrababu, R. (2017). Functional marker assisted improvement of stable cytoplasmic male sterile lines of rice for bacterial blight resistance. *Frontiers in Plant Science*, *8*(June), 1–9. https://doi.org/10.3389/fpls.2017.01131

118. Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J. P. T., & Hyvönen, J. (2013). Advances in plant gene-targeted and functional markers: A review. *Plant Methods*, *9*(1). https://doi.org/10.1186/1746-4811-9-6

119. Okporie, E. O., Chukwu, S. C., & Onyishi, G. C. (2013). Phenotypic recurrent selection for increase yield and chemical constituents of Maize (Zea mays L.). *World Applied Sciences Journal*, *21*(7), 994–999. https://doi.org/10.5829/idosi.wasj.2013.21.7.2863

120. Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *In the Light of Evolution*, *1*, 205–223. https://doi.org/10.17226/11790

121. Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, N. K., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*, *14*(11), 22499–22528. https://doi.org/10.3390/ijms141122499

122. Kumar, S., & Rao, M. (2014). *Conventional and Molecular Breeding for Bacterial Leaf Blight and Blast Resistance in Rice*. *3*(3), 1–3.

123. Ru, S., Main, D., Evans, K., & Peace, C. (2015). Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. *Tree Genetics and Genomes*, *11*(1). https://doi.org/10.1007/s11295-015-0834-5

124. Eduardo, I., de Tomás, C., Alexiou, K., Giovannini, D., Pietrella, M., Carpenedo, S., Bassols Raseira, M., Batlle, I., Cantín, C., Aranzana, M., & Arús, P. (2020). Fine mapping of the peach pollen sterility gene (Ps/ps) and detection of markers for marker-assisted selection. *Molecular Breeding*, *40*(6). https://doi.org/10.1007/s11032-020-01139-3

125. Meneses, C., Ulloa-Zepeda, L., Cifuentes-Esquivel, A., Infante, R., Cantin, C. M., Batlle, I., Arús, P., & Eduardo, I. (2016). A codominant diagnostic marker for the slow ripening trait in peach. *Molecular Breeding*, *36*(6). https://doi.org/10.1007/s11032-016-0506-7

126. Lambert, P., Campoy, J. A., Pacheco, I., Mauroux, J. B., Da Silva Linge, C., Micheletti, D., Bassi, D., Rossini, L., Dirlewanger, E., Pascal, T., Troggio, M., Aranzana, M. J., Patocchi, A., & Arús, P. (2016). Identifying SNP markers tightly associated with six major genes in peach [Prunus persica (L.) Batsch] using a high-density SNP array with an objective of marker-assisted selection (MAS). *Tree Genetics and Genomes*, *12*(6). https://doi.org/10.1007/s11295-016-1080-1

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

128. Byrne, D. H. (2005). Molecular marker use in perennial plant breeding. *IV International Symposium on Rose Research and Cultivation 751*, 163–167.

129. Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H., & Reif, J. C. (2014). Bridging the

gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity*, *112*(6), 638–645. https://doi.org/10.1038/hdy.2014.1

130. Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829.

131. Chukwu, S. C., Rafii, M. Y., Ramlee, S. I., Ismail, S. I., Oladosu, Y., Okporie, E., Onyishi, G., Utobo, E., Ekwu, L., Swaray, S., & Jalloh, M. (2019). Marker-assisted selection and gene pyramiding for resistance to bacterial leaf blight disease of rice (Oryza sativa L.). *Biotechnology and Biotechnological Equipment*, *33*(1), 440–455. https://doi.org/10.1080/13102818.2019.1584054

132. Dwivedi, S., Tripathi, R. D., Srivastava, S., Mishra, S., Shukla, M. K., Tiwari, K. K., Singh, R., & Rai, U. N. (2007). Growth performance and biochemical responses of three rice (Oryza sativa L.) cultivars grown in fly-ash amended soil. *Chemosphere*, *67*(1), 140–151. https://doi.org/10.1016/j.chemosphere.2006.09.012

133. Hospital, F. (2009). Challenges for effective marker-assisted selection in plants. *Genetica*, *136*(2), 303–310. https://doi.org/10.1007/s10709-008-9307-1

134. Serra, O., Donoso, J. M., 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 Genetics and Genomes*, *12*(5). https://doi.org/10.1007/s11295-016-1056-1

135. Bernal-Vasquez, A. M., Möhring, J., Schmidt, M., Schönleben, M., Schön, C. C., & Piepho, H. P. (2014). The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics*, *15*(1), 1–17. https://doi.org/10.1186/1471-2164-15-646

136. Jia, Y., & Jannink, J. L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, *192*(4), 1513–1522. https://doi.org/10.1534/genetics.112.144246

137. Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., & Jannink, J. L. (2014). Optimal design of preliminary yield trials with genome-wide markers. *Crop Science*, *54*(1), 48–59. https://doi.org/10.2135/cropsci2013.03.0154

138. Montes, J. M., Melchinger, A. E., & Reif, J. C. (2007). Novel throughput phenotyping platforms in plant genetic studies. *Trends in Plant Science*, *12*(10), 433–436. https://doi.org/https://doi.org/10.1016/j.tplants.2007.08.006

139. White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., Feldmann, K. A., French, A. N., Heun, J. T., Hunsaker, D. J., Jenks, M. A., Kimball, B. A., Roth, R. L., Strand, R. J., Thorp, K. R., Wall, G. W., & Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, *133*, 101–112. https://doi.org/10.1016/j.fcr.2012.04.003

140. Amitai, G., Sorek, R., McDermott, D. F., Atkins, M. B., Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., Horvath, P., Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., Yang, S.-H., Nishimasu, H., Cong, L., … Charpentier, E. (2016). TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Science*, *6*(1), 914 LP – 929.

141. Kamburova, V. S., Nikitina, E. V, Shermatov, S. E., Buriev, Z. T., Kumpatla, S. P., Emani, C., & Abdurakhmonov, I. Y. (2017). *Genome Editing in Plants : An Overview of Tools and Applications*. *2017*.

142. Green, M. R., & Sambrook, J. (2019). Analysis of DNA by agarose gel electrophoresis. *Cold Spring Harbor Protocols*, *2019*(1), 6–15. https://doi.org/10.1101/pdb.top100388

143. Marimon, N., Luque, J., Arús, P., & Eduardo, I. (2020). Fine mapping and identification of candidate genes for the peach powdery mildew resistance gene Vr3. *Horticulture Research*, *7*(1). https://doi.org/10.1038/s41438-020-00396-9

144. Moyano, E., Martínez-Rivas, F. J., Blanco-Portales, R., Molina-Hidalgo, F. J., Ric-Varas, P., Matas-Arroyo, A. J., Caballero, J. L., Muñoz-Blanco, J., & Rodríguez-Franco, A. %J P. one. (2018). *Genome-wide analysis of the NAC transcription factor family and their expression during the development and ripening of the Fragaria× ananassa fruits*. *13*(5), e0196953.

145. Martı, J., Blanco-portales, R., Moyano, E., Caballero, L., Molina-hidalgo, J., Ric-varas, P., Matas-arroyo, A. J., Mu, J., & Rodrı, A. (2018). *Genome-wide analysis of the NAC transcription factor family and their expression during the development and ripening of the Fragaria × ananassa fruits*. 1–23.

146. Tanaka, Y., Sasaki, N., & Ohmiya, A. J. T. P. J. (2008). *Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids*. *54*(4), 733–749.

147. Butelli, E., Titta, L., Giorgio, M., Mock, H., Matros, A., Peterek, S., Schijlen, E. G. W. M., Hall, R. D., Bovy, A. G., Luo, J., & Martin, C. (2008). *Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors*. *26*(11), 1301–1308. https://doi.org/10.1038/nbt.1506

148. Williams, C. M., Abd, M., Mohsen, E., Vauzour, D., Rendeiro, C., Butler, L. T., Ellis, J. A., Whiteman, M., & Spencer, J. P. E. (2008). *Free Radical Biology & Medicine Blueberry-induced changes in spatial working memory correlate with changes in hippocampal CREB phosphorylation and brain-derived neurotrophic factor (BDNF) levels*. *45*, 295–305. https://doi.org/10.1016/j.freeradbiomed.2008.04.008

149. Dragsted, L. O., Krath, B., Ravn-haren, G., Vogel, U. B., Vinggaard, A. M., Bo, P., Loft, S., Rasmussen, S. E., Sandstrom, B., & Pedersen, A. (2006). *Biological effects of fruit and vegetables*. *July 2005*, 61–67. https://doi.org/10.1079/PNS2005480

150. Shin, W., Park, S., & Kim, E. (2006). *Protective effect of anthocyanins in middle cerebral artery occlusion and reperfusion model of cerebral ischemia in rats*. *79*, 130–137. https://doi.org/10.1016/j.lfs.2005.12.033

151. Ross, J. A., & Kasum, C. M. (2002). *D IETARY F LAVONOIDS : Bioavailability , Metabolic Effects , and Safety*. 19–34. https://doi.org/10.1146/annurev.nutr.22.111401.144957

152. Xie, R., Zheng, L., He, S., Zheng, Y., Yi, S., & Deng, L. (2011). *Anthocyanin biosynthesis in fruit tree crops : Genes and their regulation*. *10*(86), 19890–19897. https://doi.org/10.5897/AJBX11.028

153. Campbell, J. A., Davies, G. J., Bulone, V., & Henrissat, B. (1998). Correction: A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities (Biochemical Journal (1997) 326 (929-939)). *Biochemical Journal*, *329*(3), 719. https://doi.org/10.1042/bj3290719

154. Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., Gigleux, S. F., Green, M., Hum, D. W., Iyanagi, T. %J P., & Genomics. (1997). *The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence*. *7*(4), 255–269.

155. 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. *Journal of Experimental Botany*, *69*(22), 5445–5459. https://doi.org/10.1093/jxb/ery305

156. Yang, Y., Zhu, J., Chen, M., Zhao, H., Liu, P., Ding, H., Jia, Z., & Du, H. (2020). Phenotypic and transcriptome analysis on red leaf and green leaf Eucommia ulmoides Oliver. *Acta Physiologiae Plantarum*, *42*(4), 1–11. https://doi.org/10.1007/s11738-020-03045-9

157. Klepikova, A. V, Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., & Penin, A. A. (2016). *A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling*. 1058–1070. https://doi.org/10.1111/tpj.13312

158. Assis, R., & Bachtrog, D. (2013). *Neofunctionalization of young duplicate genes in Drosophila*. *110*(43). https://doi.org/10.1073/pnas.1313759110

159. Smigocki, A. C., & Hammerschlag, F. A. J. J. A. S. H. S. (1991). *Regeneration of plants from peach embryo cells infected with a shooty mutant strain of Agrobacterium*. *116*(6), 1092–1097.

160. Petri, C., & Webb, Æ. K. (2008). *High transformation efficiency in plum (Prunus domestica L .): a new tool for functional genomics studies in Prunus spp .* 581–591. https://doi.org/10.1007/s11032-008-9200-8

161. Limera, C., Sabbadini, S., Sweet, J. B., Mezzetti, B., & Kühn-institut, J. (2017). *New Biotechnological Tools for the Genetic Improvement of Major Woody Fruit Species*. *8*(August), 1–16. https://doi.org/10.3389/fpls.2017.01418

8. ANNEXES

ANNEX 1

DNA Extraction SILEX (SILica matrix Extraction) protocol

1. Collect leaf material (≈ 15-25 mg) and place it in to the tubes;
2. Add tungsten beads prior to DNA extraction;
3. Handle the tubes vertically to keep the plant material at the bottom of the tube. Add 437 µl of extraction buffer;
4. Add 6 µl of β-marcaptoethanol and gently mix the tube until complete homogenization;
5. Grind the samples using the tissue lyser (30 HZ, 1:30 secs each side);
6. Add 2 µl of Rnase (10 mg/ml) and incubate in a termoblock for 30 min at 65 ^{o}C [OPTIONAL];
7. Put the samples on ice for 5 min. Add 306 µl of protein precepitation buffer and gentely vortex;
8. Centrifugate at 3 000 rpm for 15 min at room temperature;
9. Recover around 297 µl of the supernatant phase and transfer it to a new 2 ml tube;
10. Add 178 µl of binding buffer and gentely invert the tube by hand until complete mixing;
11. Add 267 µl of absolute ethanol and gentely invert the tube again for a few seconds until complete mixing;
12. Add 8 µl of silica matrix buffer and mix gentely during 5 min by hand;
13. Spin down the silica for 5 min and discard the supernatant by decantation;
14. Add 700 µl of washing buffer (Fresh prepered ethanol 70%) and shake gentely by hand until a uniform dispertion of the silica is obtained;
15. Spin down the silica for 5 min, gentely discard the supernatant by decantation and let dry at room temperature over night (make sure the ethanol is completly evaporated);

16. Add 100 µl of elution buffer, shake gentely by hand util the pellet is resuspended and incubate 5 min at 65 ºC;
17. Centrifugate at 3 000 rpm for 20 min at room temperature and transfer 90 µl of supernatant to a new tube;
18. Quantify the DNA using Nanodrop.

Reagent preparation

- ***Extraction buffer***

2% (w/v) CTAB, 2% (w/v) PVP-40, 20 mM EDTA, 100 mM Tris HCl (pH 8.0) and 1.40 M NaCl) (critical step: Avoid sample thaw before adding the extraction buffer. The buffer my be stored for several months at room temperature.

- ***Protein precipitation buffer***

24 parts of chloroform and 1 part of isoamyl alcohol. It may be stored for several months at room temperature.

- ***Binding buffer***

2.5 M NaCl and 20% PEG 8000. It may be stored for several months at room temperature.

- ***Silica matrix buffer***

Mix 5 g of silicon dioxide with 50 ml of MiliQ water and let stand for 24 h, discard the supernatant and resuspend the pellet in 50 ml of MiliQ water and let stand for another 5 h. Discard the supernatant and resuspend the pellet in 1:1 (v/v) MiliQ water. Add 10 µl of HCl 36% per ml of silica matrix solution obtained. It may be stored for several months at room temperature.

- ***Washing buffer***

Fresh prepered ethanol 70%. It may be stored for several months at room temperature.

- ***Elution buffer***

10 mM Tris HCl (pH 8.0) and 1 mM EDTA (pH 8.0). It may be stored for several months at room temperature.


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO