UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di scienze agrarie
Department of Crop Production
PhD School in Plant Biology and Crop Production
Disciplinary sector: Genetics (AGR07)

FRUIT WEIGHT IN PEACH: ASSESSING THE GENETIC POTENTIAL THROUGH PHENOTYPIC AND GENOMIC TOOLS

PhD program coordinator: Prof. Piero Bianco
Supervisor: Prof. Daniele Bassi
Co-supervisor: Laura Rossini, PhD

PhD candidate: Cássia da Silva Linge

July 10th, 2012

Contents

Chapter 1

Summary

The trait fruit weight is of great agronomic importance for the commercial production of peach. In view of conducting a study of association mapping, the genetic diversity of peach accessions from the germplasm bank of 'MAS.PES' was evaluated using amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR). The results indicated that SSR markers were more informative and showed a high level of homozygosity in the accessions under evaluation. In studies meant for breeding of this species, the identification of quantitative trait loci (QTLs) explaining the phenotypic differences of this trait is important, because they can be used as a genomic tool in marker-assisted selection (MAS). In the present study, QTLs were detected through two approaches: 1. Linkage analysis, starting from a F2 population of 123 individuals from 'NJ Weeping' (small fruit) × 'Bounty' (large fruit), in which the software *JoinMap* was used to establish the relationship of linkage between the segregating markers, while the software *MapQTL* was employed to associate the data with the obtained phenotypic linkage map. 2. Association mapping, carried out on 70 peach accessions from the 'MAS.PES' germplasm collection, in which the software *Structure* was used to analyse population structure: associations between markers and phenotypic traits were identified with the software *TASSEL*. As a result of linkage analysis, 877 single-nucleotide polymorphism (SNP) markers were grouped into 8 linkage groups and 34 QTLs related to fruit weight and size

were identified. The results of the association analysis, performed using the General Linear Model (GLM), suggested the existence of 39 markers associated with fruit weight and size.

Chapter 2

Introduction

2.1 Botanical aspects

The family Rosaceae is of high economic importance in temperate regions. It consists of more than 100 genera such as *Malus*, *Pyrus*, *Fragaria* and *Prunus*, which represent a significant number of species of commercial value (Dirlewanger et al., 2002; Shulaev et al., 2008).

The genus *Prunus* has more than 200 species of trees and shrubs that are sources of nuts, oil, wood, ornamental plants and fruits such as apricot (*Prunus armeniaca* L.), cherry (*Prunus avium* L.), plum (*Prunus domestica* L.) and peach (*Prunus persica* L.), which cater to different tastes and demands on the consumer market (Lee and Wen, 2001; Shulaev et al., 2008; Boudehri et al., 2009; Gilani, 2010).

Peach is a diploid species native to China, predominantly autogamous with less than 5% of cross-fertilization (Scorza et al., 1985). It is a medium-sized tree, up to 8 m in height. It displays pivoting roots, with subsequent formation of side branches, which reach 50–60 cm deep, depending on soil type. The lanceolate leaves with serrate-crenate margins usually have green colour, but may present a purple colour. The flowers are hermaphrodite, usually with five petals ranging from white to red, being most commonly

pink. The fruits are drupes, mostly round or elongated, and have a delicate aroma and a velvety epidermis covered by trichomes which varies in intensity or, in the case of nectarines, are absent. The fruit weight ranges from less than 50 g in wild forms to 80–110 g in early maturing genotypes and more than 680 g in late maturing varieties (Li, 1984); the commercial standard requires 180–230 g, depending on the purpose of use.

2.2 Use and economic importance of peach

According to data from FAOSTAT (2010), world peach production in 2010 was more than 20.2 million tons; 49.13% was produced in Asia, 28.97% in Europe, 16.21% in America, 4.91% in Africa, and 0.77% in Oceania. Italy is the second largest producer with an annual production of 1.59 million tons, second only to China with 10.72 million tons.

The species is a good choice for cultivation, both for commercial value of the fruits that can be consumed most commonly raw or canned, or as jam, jelly and juice, and for its medicinal, nutritional and organoleptic properties.

In nutritional terms, peach fruits are rich in water, magnesium, potassium, manganese, copper, iodine and iron. They are also sources of fibre, pectin, carbohydrates, and vitamins A, C and B complex (Júnior, 2007; Tavarini et al., 2011). In accordance with USDA (2011), a fruit with a weight of 147 g provides approximately 57 calories, 0.37 mg of lipids (fat), 14 g carbohydrates, 2.2 g of fibres, 279 mg potassium, 1.7 mg vitamin E and 9.7 mg vitamin C. These values show that the consumption of peach is an excellent choice for a healthy diet based on fruits and vegetables.

In medical terms, the supply of antioxidants, compounds that are effective against free radicals that damage DNA, proteins and lipids, is beneficial and effective against heart disease, cancer and aging (Campbell et al., 2011). Rossato (2009) conducted research to study the antioxidant activity and phenolic compounds present in the skin, pulp and peach jam, indicating that

the fruits can be used as sources of antioxidants in a diet and for industrial applications. Likewise, Segantini et al. (2012) detected significant concentrations of vitamin C in the flesh and positive correlations with the antioxidant activity. According to Li et al. (2011), the flowers are also a source of natural antioxidants for the food and cosmetic industries. As the leaves exhibit cholinomimetic action, they may be used in treatment for constipation (Gilani et al., 2000).

Mishra and Dubey (1990) showed that oil extracted from the leaves presents antifungal properties, as such oil extracted in the concentration of 0.1% inhibited mycelial growth of *Aspergillus flavus*.

2.3 Peach genome and development of the 9K SNP array v1

The peach has a small diploid genome (about 290 Mb) compared with other species of fruit trees. Since it also has a relatively short juvenile phase (two to three years) and has many Mendelian traits, peach is considered a model species for the family Rosaceae (Baird et al., 1994; Bassi and Monet, 2008; Boudehri et al., 2009; Brandi et al., 2011).

Peach genome sequencing began in 2002, but only in 2007, the Joint Genome Institute (JGI) announced a list of plants that would be the target of genome sequencing at JGI. The sequencing and publication of the results took place in 2010 (`http://www.rosacea.org`). The Project *DRUPOMICS*, supported by the Italian Ministry of Agriculture, joined the American group. This consortium was later expanded to the International Peach Genome Initiative (IPGI), which includes researchers from USA, Italy, Spain, France and Chile (Arús et al., 2012).

The availability of the peach genome along with the ease of data acquisition by using new generation sequencing has allowed for efficiently identifying

a large number of markers such as SNPs. Thus, the International Peach SNP Consortium (IPSC), that includes institutions from the US, Italy, Spain, developed the 9K SNP array v1, which consists of a platform using SNP genotyping (Verde et al., 2012). A total of 56 peach accessions divided into 12 pools were re-sequenced using the Illumina and Roche/454 sequencing technologies. The validation of these SNPs with the Illumina GoldenGateH assay were performed on a subset of the predicted SNPs. Filtration was performed until reaching a number of 8144 SNPs distributed on eight chromosomes. The stages of development of the 9K SNP array v1 are schematically described in Fig. 2.1.

The development of the 9K SNP array v1 benefits breeding programs, because it allows whole genome fingerprinting, genome wide associations and population-based analyses. In this study, the 9K SNP array v1 was used for linkage analysis (chapter 5) and association mapping (chapter 6).

2.4 Peach breeding

Peach breeding programmes have traditionally targeted the nectarine trait that is controlled by a monogenic recessive gene (Rivers, 1906; Blake, 1932; Blake and Connors, 1936), red skin overcolour (a quantitative trait), low or sub-acid as a dominant and monogenic character (Monet, 1979) and many other features. However, current breeding goals focus on the development of new cultivars with attributes that meet the requirements of the consumer market. In Europe, the consumption of this species has decreased over the last years, possibly because of the low quality of the fruits on the market (Vecchietti et al., 2009). Thus, the main objectives of peach breeding programs relate to fruit quality (Folta and Gardiner, 2009), i.e. a set of characteristics such as good-looking fruits with excellent flavour, colour, shape, texture, firmness and fruit weight (Genard et al., 2007; Infante et al., 2008; Junior et al., 2011).

The use of molecular markers in breeding programs through marker-

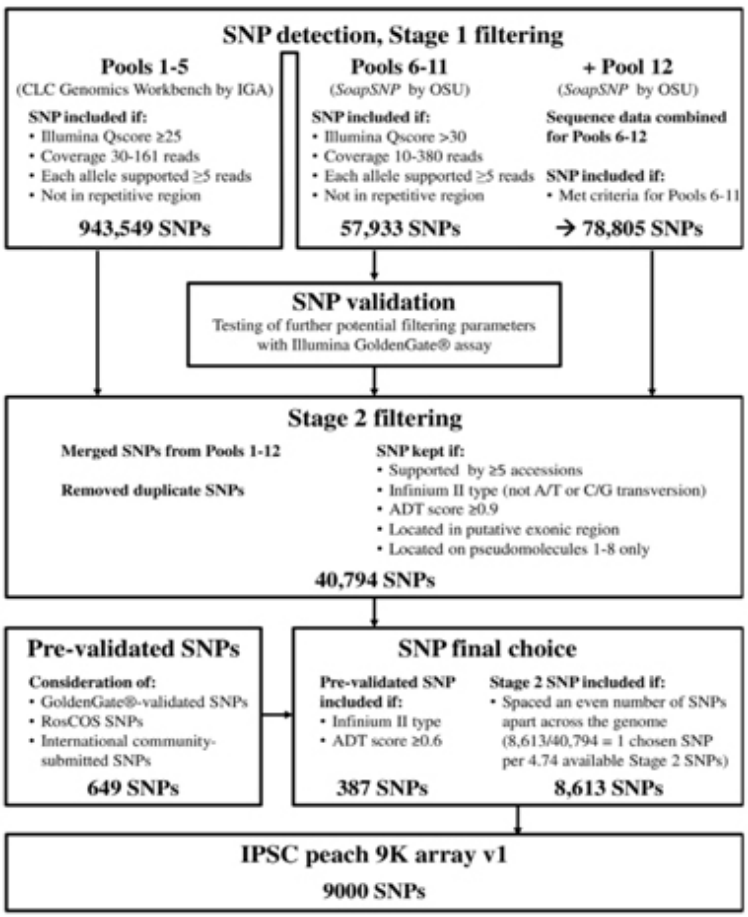

Figure 2.1: Workflow of the International Peach SNP Consortium (IPSC) peach 9 K SNP array v1 development, taken from Verde *et al.*, 2012.

assisted selection (MAS) allows optimising the selection process, especially as the traits are strongly influenced by environment. In comparison to traditional breeding methods, MAS increases the selection efficiency and reduces in the long term the time required for the release of new cultivars (Rodrigues et al., 2010). Although it requires prior knowledge on the genome regions where the loci are found, it detects very tightly linked markers and traits of interest (Abbott et al., 2009).

Han et al. (2012) applied MAS in peach and identified one sequence-related amplified polymorphism (SRAP) marker linked to the Me07Em02 colour around the stone (Cs locus). Gillen and Bliss (2006) identified two RFLPs and one RAPD marker in linkage to a locus for the Mi-root nematode resistance. In the study by Dirlewanger et al. (2006), an AFLP was found that co-segregated with the G gene ("nectarine" recessive character) and two markers (one AFLP and one RFLP) for the S gene ("saucer shaped" dominant character).

2.5 Fruit weight

Fruit weight is an important parameter on the market, since it influences the purchase of the product (Infante et al., 2008). Consumers are naturally driven to aesthetically homogeneous fruits of high weight and diameter, free of physiological and pathological alterations (Bertoglio, 2010). In peach, fruit growth is represented by a double sigmoid curve with four distinct phases: a first phase characterized by cell division, a second phase with reduced growth and lignification of the pit (endocarp), a third phase with cellular expansion and a fourth phase which corresponds to the maturation (Chalmers and van den Ende, 1975; El-Sharkawy et al., 2007; Lombardo et al., 2011). Phase I takes about 2–6 weeks after flowering. The fruit growth occurs rapidly and is due to cell division (Liu et al., 2006; Marini, 2002). The duration of phase II depends on the cultivar, ranging from days for early cultivars to weeks for late cultivars, and brings along an increase in fruit size (El-Sharkawy et al., 2007; Dardick et al., 2010; Lombardo et al., 2011; Marini, 2002). In Phase

III, the fruit grows intensively due to cell expansion in the last weeks (or months) before harvest, accumulating water, organic acids and minerals in the cell vacuole (Coombe, 1976; Marini, 2002). Phase IV is divided in two steps: firstly, the fruit reaches its final size; and secondly, it ripens in an ethylene-dependent manner (Trainotti et al., 2003; Lombardo et al., 2011). Studies conducted with peach, apple and sweet cherry to evaluate the fruit growth in the four phases showed that the final size of the ripe fruits is determined by the number and size of mesocarp cells and intercellular spaces (Yamaguchi et al., 2002; Harada et al., 2005; Olmstead et al., 2007). This trait is strongly affected by environmental factors such as plant health, water, weather conditions such as temperature fluctuations and cultural practices such as thinning or load of fruits per tree, soil moisture and nutrition. Lopez et al. (2007) studied the influence of temperature on fruit growth in peach and found that high temperatures in spring promote a reduction in fruit size as the tree does not provide sufficient resources to support high growth rates under these circumstances. Tibola et al. (2007) evaluated the degree of compliance of peach orchards in integrated production regime and noted broken branches and small fruits in those orchards where the intensity of thinning was inadequate. Likewise, studies indicate that the relative position of the fruit to the assimilating sources influences growth; fruits positioned in crown areas, more exposed to direct sun light, tend to have increased growth compared to those in the inner position (Li et al., 2005; Guedes et al., 2008). Referring to water availability, Mercier et al. (2009) reported that a water deficit during the cell expansion phase leads to a significant decrease in fruit weight and an increase in soluble solids.

The final fruit size is also associated with the flow of water and photoassimilates transported through the xylem and phloem. The distribution among plant organs results in changes in size, total production and individual fruit weight (Peil and López-Galvéz, 2002; Morandi et al., 2007, 2008). In peach, sorbitol and sucrose are the primary photosynthetic products of translocated carbon. Knowledge of their concentrations in cell sources provides information to better understand possible modes of phloem loading

and the mechanisms for regulation of growth and final fruit size (Nadwodnik and Lohaus, 2008). Several authors have examined how the availability of carbohydrates influences growth and final fruit size. Morandi et al. (2008) studied how the metabolism of carbohydrates in the peach fruit is affected by source-sink changes and, specifically, how the availability of sorbitol and sucrose regulates the enzymatic activity and fruit growth. The results indicated that the supply of carbon can be changed by environmental stress or cultural practices, and that the availability of carbon promotes differences in fruit size and sugar content.

2.6 Genetic and molecular dissection of fruit weight

Fruit weight is a quantitative trait resulting from the combined action of several genes and environmental factors. Several studies have been carried out in view of obtaining peach fruits with better physical and chemical qualities (Morandi et al., 2008; González-Agüero et al., 2008; Morandi et al., 2007). However, there are few studies aiming at the knowledge of genes and chromosomal regions linked to traits such as sugar content, acidity, firmness, shape and fruit weight (Quilot et al., 2005; Fernie et al., 2006; Causse et al., 2007; Bertin et al., 2009; Eduardo et al., 2011).

Quilot et al. (2005) analysed the genetic control of fruit quality of peach with an ecophysiological model, using a QTL mapping approach; QTLs were detected for all genotypic parameters, and a large number of the same QTLs was found over two years. Referring to QTLs specifically associated with fruit weight or fruit size, Yuan et al. (2008) observed the presence of four QTLs and pointed out their importance for plant breeding in cucumber. In sweet cherry, Zhang et al. (2010) detected the presence of three QTLs associated with fruit size and one QTL for the number of mesocarp cells. In papaya, Blas et al. (2012) identified 14 QTLs associated with fruit weight and form, with phenotypic effects ranging from 5 to 23%. In apple, Devoghalaere et al.

(2012) found six QTLs for fruit weight using the 'Royal Gala' × 'Braeburn' (RG×BB) and 'Starkrimson' × 'Granny Smith' (STK×GS) genetic maps.

In tomato, fruit weight and size have already been extensively studied. In total, 28 loci have been associated with these traits (Grandillo et al., 1999; Paran and van der Knaap, 2007; Zhang et al., 2012). Frary et al. (2000) identified the QTL *fw2.2* that explains 30% of the variation in fruit size in this species. This locus codes for a negative repressor of cell division with activity during the cellular division of the fruit (Frary et al., 2000; Cong et al., 2002; Tanksley, 2004). Other two loci, *fasciated* (chromosome 11) and *locule-number* (chromosome 2) have been identified as provoking changes in fruit size via changes in the number of carpels in the flower (Tanksley, 2004). *Fasciated YABBY-like* encodes the transcription factor controlling organ polarity, and the mutation that increases locule number and fruit weight is caused by a 294 kb inversion knocking out the gene (Cong et al., 2008; Huang and van der Knaap, 2011; Zhang et al., 2012). On chromosome 11, a new fruit weight locus called *fw11.3* was mapped near *fasciated*. In this region, 22 candidate genes for this trait were found (Huang and van der Knaap, 2011). *SUN* and *Ovate* loci control fruit elongation. *SUN* encodes a protein that is a positive regulator of growth resulting in elongated fruit (Xiao et al., 2008; Rodriguez et al., 2011). *Ovate* with a single mutation, leading to a premature stop codon, causes the transition of tomato fruit from round- to pear-shaped (Liu et al., 2002).

In view of the results obtained in tomato and other species, it is important to detect QTLs that control this trait in peach.

2.7 Genetic approaches for dissection of QTLs in plants

Linkage analysis and association mapping are two methods for studying the mechanisms that control the genetic architecture of complex traits (Lander

and Schork, 1994; Risch and Merikangas, 1996; Mackay, 2001; Doerge, 2002; Darvasi and Shifman, 2005; Yu et al., 2008). Linkage maps are used to identify chromosomal regions of agronomic interest with a relatively low coverage of markers, while association mapping provides a high resolution with a high coverage of markers, either via prior information on candidate genes or by a genome-wide scan (Thornsberry et al., 2001; Hirschhorn and Daly, 2005; Yu et al., 2008).

Traditional QTL analysis based on linkage maps requires the establishment of a biparental population with crossing of contrasting genotypes. Phenotypic measurements are collected for the quantitative trait of interest and analysed in conjunction with genotyping information from molecular marker data covering the genome. Initially, a linkage map is constructed through the following steps (Toledo et al., 2008):

I. distributing the markers in groups, called linkage groups;

II. sorting them within the group; and

III. estimating the distance between adjacent or flanking markers within each group.

Finally, statistical methods allow to map the QTLs and to estimate their positions and their effects through associations between the markers and the evaluated phenotypic traits (Jangarelli et al., 2010).

The main methods of linkage mapping are: analysis of single markers (Stuber et al., 1987), interval mapping (Lander and Botstein, 1989), composite interval mapping (Zeng, 1994), its extension by mapping multiple characters (Jiang and Zeng, 1995) and mapping of multiple intervals (Kao et al., 1999).

Analysis of single markers does not require complete linkage maps and allows to detect possible associations between the genotypes of the markers and quantitative traits. The statistical methods used to detect associations

are t-tests, analysis of variance (ANOVA) and linear regression. However, linear regression is the most commonly used, because the coefficient of determination describes, based on the marker, the phenotypic variation (Collard et al., 2005). A disadvantage of this method is the possibility of confusing the position and the effect of the QTL in addition to underestimating the QTLs, since their estimators are biased by the recombination fraction between marker and QTL (Collard et al., 2005; Bento, 2006).

The interval mapping approach is based on the segregation information of pairs of adjacent markers, using the maximum likelihood method to estimate the frequency of recombinants and the magnitude of the effect of the QTL in the interval between two linked markers (Lynch and Walsh, 1998; Carneiro and Vieira, 2002). This method is statistically more powerful compared to single marker analysis (Lander and Botstein, 1989; Liu, 1998; Collard et al., 2005). The only drawback of this method is that other QTLs outside of the interval are ignored.

For example, the construction of a map using m marks leads to likelihood map for each interval (M1–M2, M2–M3, etc.). Thus, the information of marker pairs at considered intervals is taken into account, built after a map with all the intervals (Toledo, 2006). Thus, the statistics are calculated for each interval m–1, thereby limiting the efficiency of this method, since information on markers outside the mapped range, that may contain association with the trait of interest, is not considered (Lander and Botstein, 1989; Zeng, 1993; Toledo, 2006; Bento, 2006). To overcome this limitation, Zeng (1993) proposed to use a multiple regression model, thereby creating the method of composite interval mapping.

Thus, information from markers adjacent to the considered interval and markers present in other linkage groups is considered as cofactors (Toledo, 2006). This method is currently one of the most widely employed methods.

The mapping of multiple characters is an expansion of the interval map-

ping, using statistical models in a multivariate approach. The methodology includes the effects through which a character is influenced by more than one genetic locus and the pleiotropic effects, i.e., different traits may share more than one common locus (Li et al., 2006; Sabadin, 2008). The method also allows evaluating multiple environments estimating the effect of the interaction of QTLs with the environment and studying the nature of the genetic correlation between traits (Bento, 2006).

Multiple interval mapping incorporates epistasis into the model and considers multiple intervals simultaneously. In the presence of epistasis, this method is more advantageous than the others as it provides greater efficiency and accuracy in the identification of QTLs, contributes to the understanding of epistasis and is more efficient in marker-assisted selection because of the identification of epistatic effects (Kao et al., 1999; Carneiro and Vieira, 2002).

Although association mapping was initially used in studies of human diseases, its field of application has expanded in recent years to plants including crops and fruit species (Remington et al., 2001; Agrama et al., 2007; Cockram et al., 2008; Murray et al., 2009). Association methods detect valid genetic associations for the entire population and not only for a specific cross, showing higher allelic variation, higher accuracy and resolution of QTLs (Breseghello and Sorrells, 2006; Sorkheh et al., 2008). The construction of an association map involves six steps:

I. Selection of individuals;

II. Measurement of phenotypic characteristics;

III. Genotyping of individuals;

IV. Determining the population structure;

V. Quantifying the linkage disequilibrium using molecular marker data; and

VI. Identifying markers associated with the trait of interest.

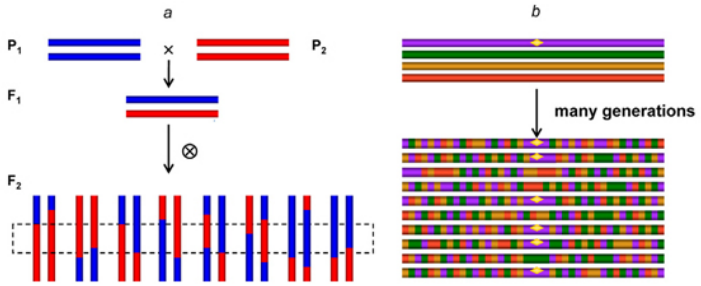

Figure 2.2: Comparison of recombination events considered in linkage analysis (a) with designed mapping populations from controlled cross and association mapping (b) within diverse collections, taken from Zhu et al. (2008).

The analysis of population structure aims to prevent the occurrence of correlations between unrelated loci (Zhu et al., 2008; Abdurakhmonov and Abdukarimov, 2008).

In association mapping, the selection of individuals is conducted in natural populations, germplasm banks, breeding lines, landraces, thus exploiting all the recombination events that occurred over the evolutionary history of the species, contrary to what occurs in the linkage analysis (Fig. 2.2) in which, besides being necessary to develop a population obtained by a controlled experiment, the recombination events are occurring during the development of the population to be mapped (Myles et al., 2009; Wang et al., 2012). Association mapping emerged as an interesting alternative to traditional QTL mapping, since it allows mapping QTLs in natural collections, thus exploiting a broader diversity and presenting a potential for improving the traits and safety of germplasm (Yu et al., 2008; Myles et al., 2009).

Phenotypic analysis involves a large number of individuals. Data collection, the adequate number of replications and experimental design are still an object of study in association mapping (Zhu et al., 2008). Sometimes, evaluations may be needed in unbalanced designs of experiments and in various environments (Ersoz et al., 2009). It is also possible to use the historical phenotypic data from cultivars, without the need to develop new studies (Sorkheh et al., 2008).

Various types of genetic markers can be used to investigate association (Ersoz et al., 2009), including dominant markers like AFLP (D'hoop et al., 2008). The simple sequence repeats (SSR) markers are useful in these studies, as reported in in rice (Agrama et al., 2007), sorghum (Casa et al., 2008) and cotton (Abdurakhmonov et al., 2009). The SNP markers are potentially more useful in comparison with other genetic markers (Ersoz et al., 2009).

The main step in association mapping is to identify the existence of substructure in the population and to detect possible bias due to factors such as adaptation or domestication (Thornsberry et al., 2001; Wright and Gaut, 2005; Ersoz et al., 2009). Population structure is the presence of subpopulations in the sample in which individuals are more closely related to each other than the average pair of individuals taken at random in the population (Breseghello and Sorrells, 2006). When the trait of interest (e.g., disease resistance) is more prevalent in one subpopulation than others, the trait will be associated with any marker allele that is in high frequency in that subpopulation (Ewens and Spielman, 1995; Lander and Schork, 1994; Pritchard and Rosenberg, 1999). Therefore, estimates of the population structure are a prerequisite for the successful implementation of the association mapping approach in admixed populations (Simko and Hu, 2008). Among the main methods to estimate the level of population structure, principal component analysis (PCA) and cluster analysis, based on Bayesian statistics, are prominent, e.g. as provided by the software *Structure* (Pritchard et al., 2000a).

The quantification of linkage disequilibrium in statistical terms is measured by deviations between the average measurements of the frequencies observed for certain combinations of alleles and the expected frequencies, considering that segregation of these alleles is independent. Thus, in situations with dependence of alleles, the values of these deviations are different from zero, creating a disequilibrium (Lopes, 2011).

The analysis of association between markers and traits of interest is per-

formed by statistical software packages such as SAS (SAS-Institute, 1999), R (Ihaka and Gentleman, 1996) and *TASSEL* (Bradbury et al., 2007).

In summary, linkage mapping based on bi-parental populations is an efficient tool for identifying chromosomal regions conditioning the variation for traits of interest. However, these chromosomal regions may influence the trait only in the specific cross considered (Brachi et al., 2010). Association mapping does not have this drawback. However, it can generate false positives due to population structure (Wang et al., 2005; Shriner et al., 2007; Brachi et al., 2010; Wang et al., 2012). Researchers tend to use both methods to detect QTLs since they are complementary and each method contributes to a better understanding of the inheritance of complex traits.

Chapter 3

Objectives

The aim of my thesis was to identify genomic markers linked to fruit weight in peach, allowing the use of these markers in breeding programmes through marker-assisted selection (MAS).

For this purpose, the present study was divided in three stages, as summarized in the three chapters 4, 5 and 6 (Fig. 3.1).

In Fig. 3.1, all steps in red boxes correspond to activities that I conducted in the laboratory of plant genomics in Parco Tecnologico Padano, located in Lodi, Italy, or on the experimental field, located in Bologna county (Italy) where the collection of germplasm accessions and also the F2 population ('NJ Weeping' × 'Bounty'), evaluated in chapter 5, reside. The gray box represents the activities conducted in other laboratories to which DNA samples were sent to be genotyped using the Peach 9K SNP array v1.

In chapter 4, the objective was to understand the genetic diversity of accessions from the germplasm collection of a breeding program called 'MAS.PES'. The results were used to select accessions to build a panel of genetic diversity to be studied in association mapping. To analyse diversity, AFLP and SSRs were used. For AFLP markers, a total of four primers combinations for a total of 153 accessions of peaches was chosen. Fifteen SSR markers distributed

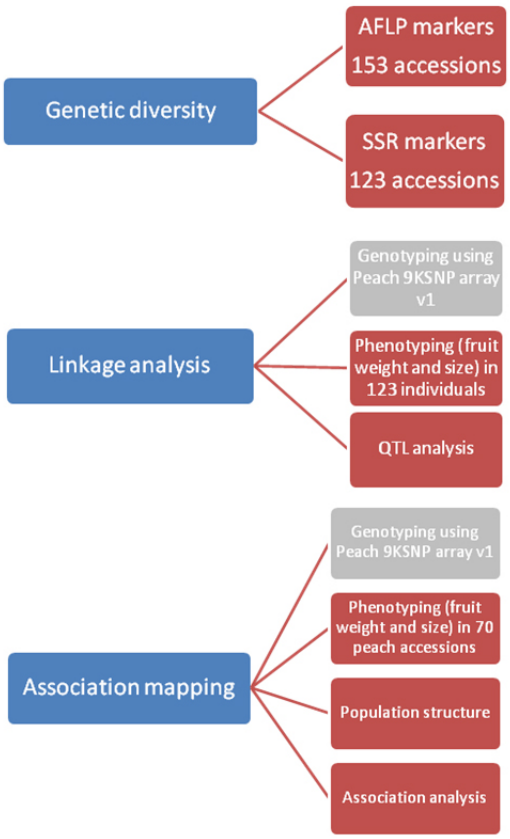

Figure 3.1: Stages of the project to study fruit weight in peach: assessing the genetic potential through phenotypic and genomic tools; the gray box corresponds to the analysis provided by other laboratories through the Illumina platform.

on eight chromosomes of peach were used to detect genetic diversity in 123 accessions.

The second objective (chapter 5) was to build a linkage map in an F2 breeding population by SNP markers from the peach 9K SNP array v1. Phenotyping was conducted in a total population of 123 F2 individuals from the cross 'NJ Weeping' × 'Bounty', segregating for fruit weight and size. QTL mapping was performed using the generated map and obtained phenotypic data.

The third objective (chapter 6) was to perform a preliminary association mapping study and identify markers associated with fruit weight and size. 70 peach accessions showing variation for the trait of interest were genotyped with the 9K SNP array v1. The accessions were phenotypically evaluated according to fruit weight and size. To avoid false positive associations, the population structure was estimated using a Bayesian analysis. The association between the markers and the phenotype data was obtained using the general linear model (GLM) in which the analysis of the effects of the population structure were introduced.

The present work provides new information on the genetic architecture of the trait fruit weight in peach, representing a possible milestone for breeding programmes.

Chapter 4

Analysis of genetic diversity in peach by AFLP and SSR markers

4.1 Introduction

Although peach (*Prunus persica* L.) has already undergone many generations of breeding, there is a strong commercial interest to provide new cultivars that satisfy changing consumer interests on the market (Infante et al., 2008). Conventional breeding methods are expensive and time-consuming because peach requires three years from seed to bear fruit allowing the first phenotypic evaluations (Varshney and Tuberosa, 2009). Furthermore, phenotypic variation is dependent on both genetic and environmental factors, implying that breeding based on phenotypic selection may not afford the maximum genetic gain (Hesse, 1975).

Accessing the genetic diversity of a species directly at the DNA level through the use of molecular markers is a powerful tool to reveal the genotypic differences between individuals. Markers are not influenced by environmental factors (Rajapakse, 2003; Cheong, 2012) and are useful in the early steps of selection within large progenies, because those individuals with

unwanted characters can be removed immediately after seed sprouting and allow to focus only on individuals with the desired trait (Collard and Mackill, 2008; Cheong, 2012).

A variety of markers is available, which differ according to the technology used to detect variation. They are divided into hybridization markers and amplification markers (Lins, 2008). The main hybridization-based markers are RFLPs (Restriction Fragment Length Polymorphisms) and minisatellites or VNTRs (Variable Number of Tandem Repeats) that were used in various plant species (Keim et al., 1992; Lamy et al., 1994; Lebrun et al., 1998). The development of technologies based on DNA amplification (Polymerase Chain Reaction; PCR) led to the advent of molecular markers such as RAPDs (Random Amplified Polymorphic DNA) (Williams et al., 1990) and AFLPs (Vos et al., 1995), ISSRs (Inter Simple Sequence Repeats) and SSRs (Simple Sequence Repeats) (Sørensen et al., 2009). All these types of markers have been used in genetic diversity studies in peach (Lima et al., 2003; Hu et al., 2005; Aranzana et al., 2010).

AFLP is a powerful DNA fingerprinting technology based on selective amplification of a subset of genomic restriction fragments using PCR (Vos et al., 1995). The technique involves cleavage of the total DNA by restriction enzymes resulting in a large number of fragments. Short DNA sequences (adapters) are coupled to the ends of restriction fragments that bind to specific primers during PCR. The amplified fragments are separated on polyacrylamide gels or through capillary electrophoresis and visualized with autoradiography or fluorescence methods (Oliveira et al., 2005; Pacheco Cruz, 2010). AFLPs are dominant markers that provide an effective, rapid and economical tool for detecting a large number of polymorphic genetic markers without prior molecular information about the corresponding loci (Teneva, 2009).

SSR markers represent DNA regions of short tandem repeat sequences of one to four nucleotides, which are amplified by PCR using primers with 20

to 30 bases, complementary to sequences flanking the microsatellite (Bento, 2006). These markers require an initial development step as the sequences flanking the repeat need to be known in order to design primers. Despite their higher costs compared to AFLPs, SSRs have the advantage of showing co-dominance, high polymorphism and high reproducibility, which are important characteristics for an analysis based on molecular markers (Sansavini, 1998; Sosinski et al., 2000).

SSR and AFLP markers provide high reproducibility and high variability which may increase the reliability and resolution of phylogenies. Since both types of markers allow genome-wide sampling, the probability that the data sets and phylogenies represent the evolutionary affinities within the species rather than that of the individual traits is very high (Koopman et al., 2008).

The aim of the present study was to score the genetic diversity and structure of a population of peach accessions from a germplasm bank of the 'MAS.PES' breeding program by AFLP and SSR markers with a view to support breeding programmes.

4.2 Materials and methods

The 'MAS.PES' peach germplasm bank includes accessions from Europe, Asia and America. Genomic DNA of the accessions was extracted from young leaves by the Power Plant DNA isolation kit (MO BIO Laboratories, Carlsbad, CA, USA), according to the manufacturer's instructions.

4.2.1 AFLP analysis

Four AFLP primer combinations (E32M50, E32M60, E36M60 and E36M59) were used on 153 peach accessions (Table 4.1). Digestion, ligation and pre-amplification were carried out according to the protocol by Vos et al. (1995). Customized forward primers, which were 5-labelled with 6-FAM (Applied Biosystems, Foster City, USA), were used instead of radioactive labelled

primers for the selective fluorescent amplification process. Selected amplification products were diluted tenfold; 10 μl of deionised formamide and 0.15 μl of GeneScan 1200-LIZ internal size standard (Applied Biosystems, Foster City, USA) were added to 2 μl of each dilution. AFLP fingerprinting was carried out with a fluorescent protocol using the ABI3730 DNA Analyzer for the automatic electrophoretic separation and the GeneMapper 4.0 software (Applied Biosystems) for scoring polymorphic peaks. Polymorphic peaks were converted to a binary data matrix and analysed in the *Darwin* software (Perrier and Jacquemond-Collet, 2006). For the calculation of genetic similarities, the similarity coefficient of Jaccard (Sneath and Sokal, 1973) was used. For the construction of the phylogenetic tree, the Neighbour Joining (NJ) distance method was employed and statistical support for phylogenetic groupings was estimated with the bootstrap method (1000 replications). The similarity matrices were subjected to Principal Coordinates Analysis (PCoA) with the purpose of visualizing the accessions through graphical dispersion.

4.2.2 SSR analysis

Fifteen primer pairs (Table 4.2) were used on 123 peach accessions (Table 4.4). The primers were selected on the basis of an earlier study conducted by Pacheco Cruz (2010) based on the PIC (Polymorphic Information Content) values and coverage in the genome. SSR reactions were performed using the multiplex-ready PCR protocol (Hayden et al., 2008). The particularity of this protocol consists of the inclusion of short generic primers tagF (fluorescently labelled with FAM, VIC, NED or PET fluorescent dyes) and tagR (unlabelled) whose sequences were 5'–ACGACGTTGTAAAA–3' and 5'–CATTAAGTTCCCATTA–3', respectively.

PCR was carried out in a final volume of 8 μl with the following conditions: 1–20 ng of genomic DNA, one PCR reaction buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-Cl pH 8.8 and 0.1% Tween-20), 3 mM $MgCl_2$, 0.2 mM of each dNTP, 100 nM of each tag primer, 40 nM of SSR-specific primers (Sigma Life Sciences, Italy) in either singleplex or multiplex reactions, 0.01 U of Eu-

Table 4.1: Peach accessions analysed by AFLP markers

194RXIII43	Da Tian Tao	Morsiani 60
391 C12 XXXIV 86	Dialona	Morsiani 90
A15	Dixired	Moscatello
A219	DOFI 88338026	Neve
A318	Dolores	NJ307
Afra T15-9	Doris	NJC112
Agate	Durbin	NJC113
Albatros	Early Rich	Orion
Alexa	Early Silver	P2152N
Aliblanca	Early Top	Paola Cavicchi Precoce
Alice	Elbertita	Poppa di Venere
Alice Col	Elegant Lady	PZ 1
Alipersie	Fei Cheng Bai Li	Rasciadente Gialla
Alired	Flavorcrest	RedHaven
Alirosada	Forli 1	Regina Bianca
Alitop	Forlivese	Regina d'ottobre
Alix	Francoise	Rising Star
Alma	Gemini	Romagna Bright
Amanda	GG30	Romagna Queen
Amarillo de Agosto I	Gladis	Romagna Red
Amarillo de Agosto II	Glohaven	Romagna Top
Angelo Marzocchela	Goldcrest	Romamer 2
Antonina	Grezzano	Rosa Del West
Antony	Harrow Blood	Rossa di Trenti
Aurora	Honey Kist	Royal Lee
Azurina	Honey Royal	Royal Pride
Azuritte	IF7910001	Royal Summer
Beicme bin	IFF331	Rubia
Bella Contadina	IFF691	Ruby Rich
Bella di Borgo d'Ale	IP1	Sweet Melody × W Glory
Bella di Cesena	Iskra	S5898128
Bella di Cesena Precoce	J. H. Hale	San Giorgio
Bella di Piangipane	Jing Yu	San Varano 3
Big Top	June Princess	Sel 97517
Blazing Star	Kamarat	Sentry
Blushing Star	Kaweah	Pesca Settembrina
BO05015080	Kevina	Settembrina Polpa Rossa
Borgia	Kurakata Wase	Snow King
Botto	KV930278	Spring belle
Buco Incavato C	KV930386	(Stark) RedGold
Buco Incavato I	KV930455	Summer Sweet
Buco Incavato II	Limonet	Sweet Fire
Calred 27-48	Maillara	Sweet Silver
Capucci 18	Maria Dorata	Tardibelle
Cesarini	Maria Regina	Torquoise
Cinzia	Marli	Valley Sweet
Compact RedHaven	Maruja	Vespignani 2
Corindon	Maura	Vista Rich
Cormonese Migliorata	Maycrest	Zee Diamond
Cristal Rose	Mayglo	Zee Glo
Cristina	Merril Gem Free	Zephir

Table 4.2: Primer sequences used in microsatellite analysis

Primer	Primer F sequence (5–3)	Primer R sequence (5-3)
UDP98-409	GCTGATGGGTTTTATGGTTTTC	CGGACTCTTATCCTCTATCAACA
UDP98-412	AGGGAAAGTTTCTGCTGCAC	GCTGAAGACGACGATGATGA
BPPCT025	TCCTGCGTAGAAGAAGGTAGC	CGACATAAAGTCCAAATGGC
BPPCT007	TCATTGCTCGTCATCAGC	CAGATTTCTGAAGTTAGCGGTA
BPPCT015	ATGGAAGGGAAGAGAAATCG	GTCATCTCAGTCAACTTTTCCG
BPPCT001	AATTCCCAAAGGATGTGTATGAG	CAGGTGAATGAGCCAAAGC
UDP96-005	GTAACGCTCGCTACCACAAA	CCTGCATATCACCACCCAG
BPPCT017	TTAAGAGTTTGTGATGGGAACC	AAGCATAATTTAGCATAACCAAGC
EPPCU5176	ATGACCACACAGAATCACCC	GATCCTCAGCCCGAGTCAAT
CPPCT006	AATTAACTCCAACAGCTCCA	ATGGTTGCTTAATTCAATGG
BPPCT038	TATATTGTTGGCTTCTTGCATG	TGAAAGTGAAACAATGGAAGC
UDP98-022	CTAGTTGTGCACACTCACGC	GTCGCAGGAACAGTAAGCCT
UDP96-008	TTGTACACACCCTCAGCCTG	TGCTGAGGTTCAGGTGAGTG
CPDCT045	TGTGGATCAAGAAAGAGAACCA	AGGTGTGCTTGCACATGTTT
CPSCT039	GCCGCAACTCGTAAGGAATA	TCCACCGTTGATTACCCTTC

roTaq DNA polymerase (EuroClone, Italy). Thermal cycling comprised an initial denaturation step of 2 min at 95 °C; 20 pre-amplification cycles of 30 s at 92 °C, 30 s at 60 °C, 30 s at 72 °C; 40 amplification/labelling cycles of 15 s at 92 °C, 30 s at 54 °C, 30 s at 72 °C; and a final extension of 5 min at 72 °C followed by 25 min at 25 °C. After amplification, PCR products were diluted with 10 μl of distilled water. Subsequently, 5 μl of dilution labelled with a different dye was pooled in a 1:1:1:1 proportion (multi-pooling), ethanol-precipitated and re-suspended in 20 μl with distilled water. Capillary electrophoresis was performed with a mix including 2 μl purified PCR multi-pool, 10 μl formamide and 0.15 μl GeneScan500 LIZ-250 size standard using an ABI Prism 3730 DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions; allele size was determined using the GeneMarker demo version 1.70 (SoftGenetics).

Phylogenetic analysis was performed by the *Darwin* software, using the method of neighbour-joining distance and estimating statistical support for phylogenetic groupings with the bootstrap method (1000 replications).

Determining the total number of alleles, the observed and expected heterozygosity (Ho and He), the fixation index or F-statistic of Wright using the *G*enAlEx 6.3 software (Peakall and Smouse, 2006) allowed estimating

the genetic diversity.

4.2.3 Population structure analysis

Population structure analysis was conducted with the software *Structure* (Pritchard *et al.*, 2000), based on Bayesian statistics using 15 SSR markers (Table 4.3). The "admixture model" of ancestry and correlated allele frequencies was adopted to analyse the data set. No preliminary information on the number of subpopulations was considered. The proportion of the ancestry of each individual was tested considering a number K from 1 to 10, with 5 iterations for each value of K. The settings for burning and MCMC (Markov Chain Monte Carlo) were 20,000 and 200,000, respectively. To determine the number K, the model established by (Evanno et al., 2005) was adopted using the *S*tructure Harvest (Earl and vonHoldt, 2012).

4.3 Results

4.3.1 AFLP analysis

The use of four primer combinations allowed the acquisition of 171 molecular markers to differentiate the peach accessions. The number of polymorphic fragments for the markers E32M50, E32M60, E36M59 and E36M60 were 41, 47, 45 and 38, respectively. The dendrogram shown in Fig. 4.1 demonstrates the formation of two groups, with the presence of subgroups in group 2. The cophenetic correlation coefficient (measure of reliability in the clusters of accessions with respect to the genetic similarity matrix) was 0.97.

The analysis of graphic dispersion via Principal Coordinates Analysis (Fig. 4.2) showed results similar to those observed in the dendrogram. It is possible to observe that the first group is separated from the second group and other subgroups by the axis 1. The first axis explained about 17%, while the second explained 9% of the observed variability.

Table 4.3: Peach accessions analysed by 15 microsatellite markers

Accessions		
189CXIIKLI62	Elegant Lady	Red Elegant
193 QXXVI 131	Fei Cheng Bai Li	Redhaven
193QXXVII 111	Forli 1	Rich Lady
193QXXVII 111	Francesca	Rising Star
194 RXXVI 12	GloHaven	Romagna 3000
391 C12 XXXIV 86	Grenat	Romagna Bright
A15 – P1	Harrow Blood	Romagna giant
A219 – P1	Hardired	Romagna Gold
A9	Helena Cling	Romagna Red
Alirosada	Honey Blaze	Romagna Star
Aliblanca	Honey Glo	Romagna Top
Alicecol	Honey Kist	Rosa Dardi
Alipersié	Honey Royal	Royal Estate
Alired	IFF 691	Royal Glory
Alitop	IFF 331	Royal Jim
Alma	Ionia	Royal Lee
Ambra	Iskra	Royal Majestic
Andross	June Princess	Royal Time
Angelo Marzocchella	Kakamas	Rubia
Antony	Kaweah	Ruby Rich
Autumn Grand	Kevina	S5898:128
Azurite	KV930386	Salkaja
Big Top	KV930455	Sentry
Blazing Star	Laura	Siberian C
Blushing Star	Magique	Soleada
Bolero	Maillar	Springbelle
Bonia	Maura	S 6699
Bordó	Maycrest	(Stark) RedGold
Botto	Mayfire	Summer Rich
Capucci 18	Merril Gem Free	T16
Chimarrita	Nadia	Tardigold
Contender	Nectagrand 1	Tardiva Spadoni
Corindon	NJ 307	Tasty Free
Cristina	NJ Weeping(PI 91459)	Torquoise
Dixired	NJC113	Venus
Dolores	Oro A	Vista Rich
Dourado	P5/645	Zao Xia Lu
Early Rich	Paola Cavicchi Precoce	Zee Diamond
Early Top	Pieri 81	Zee Glo
Early Zee	Rasciadente Bianca	Zee Lady
Elbertita	Rasciadente Gialla	Zephir

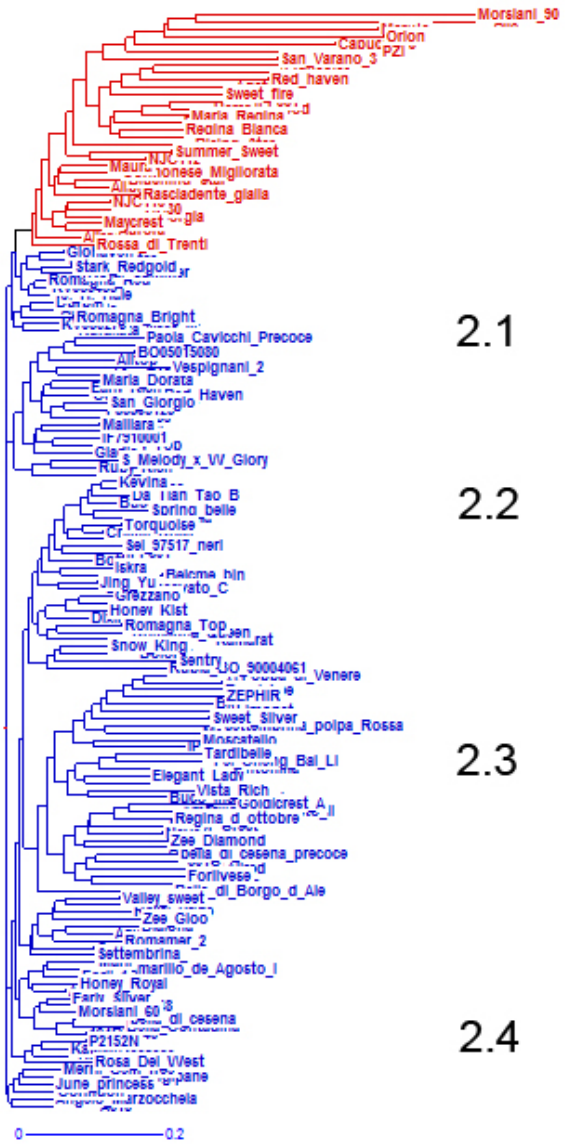

Figure 4.1: Neighbour joining tree of 153 peach accessions from the analysis of 171 AFLPs using four primers combinations; the tree was constructed by means of Jaccard's genetic distance.

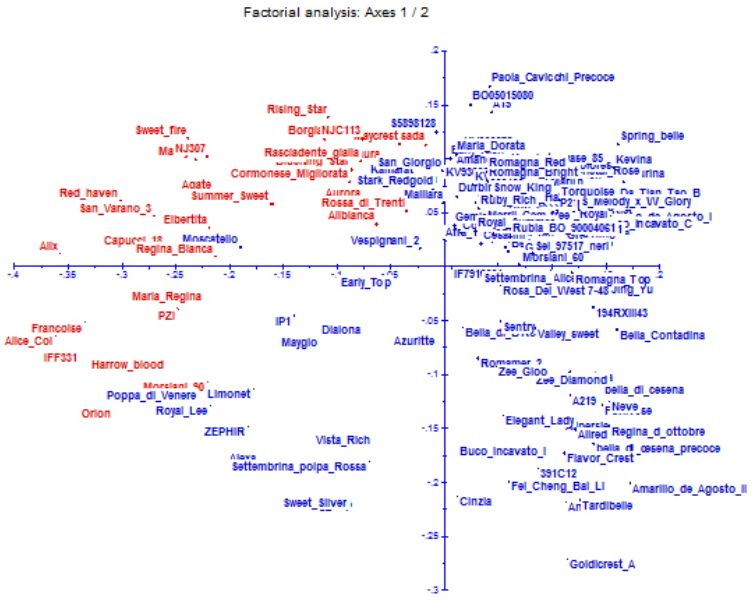

Figure 4.2: Principal coordinates analysis of 153 peach accessions obtained by four AFLP primers combinations

4.3.2 Microsatellite analysis

The 15 studied SSRs amplified 115 alleles, with an average of 7.7 alleles per locus. Markers BPCCT025, BPPCT001, and BPPCT017 were the most polymorphic with an amplification of 11, 11, and 10 alleles, respectively (Table 4.4). Values of expected heterozygosity ranged from 0.441 (DP98-409) to 0.732 (BPPCT001) with an average of 0.6. The marker UDP98-412 showed the highest observed heterozygosity (0.634), while marker UDP96-008 had the lowest heterozygosity (0.244). For all loci, the fixation indexes were positive with an average of 0.3.

The phylogenetic tree (Fig. 4.3) demonstrated the formation of three groups. A cophenetic correlation of 0.93 was obtained, slightly lower compared to the value observed in the phylogenetic tree obtained by AFLP markers. However, it is worth highlighting that only 15 microsatellite markers were used in the present study.

Regarding the population, the method by Evanno et al. (2005) revealed

Table 4.4: Variability of 123 peach accessions analysed by 15 microsatellite markers. Locus name, number of alleles per locus (A), effective number of alleles (Ae), observed heterozygosity (Ho), expected heterozygosity (He), and Wright's fixation index (F) are shown.

Locus	A	Ae	Ho	He	F
BPPCT017	10	2.512	0.385	0.602	0.360
UDP96-008	7	1.967	0.244	0.492	0.504
EPPCU5176	8	3.203	0.585	0.688	0.149
CPPCT045	5	2.521	0.545	0.603	0.097
BPPCT038	8	2.384	0.361	0.580	0.379
UDP98-022	5	2.722	0.472	0.633	0.255
CPPCT006	5	2.778	0.438	0.640	0.316
CPSCT039	7	2.458	0.434	0.593	0.268
BPCCT025	11	2.291	0.390	0.564	0.307
UDP98-412	8	3.245	0.634	0.692	0.083
UDP98-409	5	1.790	0.358	0.441	0.190
BPPCT001	11	3.736	0.469	0.732	0.360
BPPCT015	8	2.141	0.388	0.533	0.272
UDP96-005	8	2.104	0.434	0.525	0.172
BPPCT007	9	2.846	0.528	0.649	0185
Average	7.7	2.6	0.4	0.6	0.3

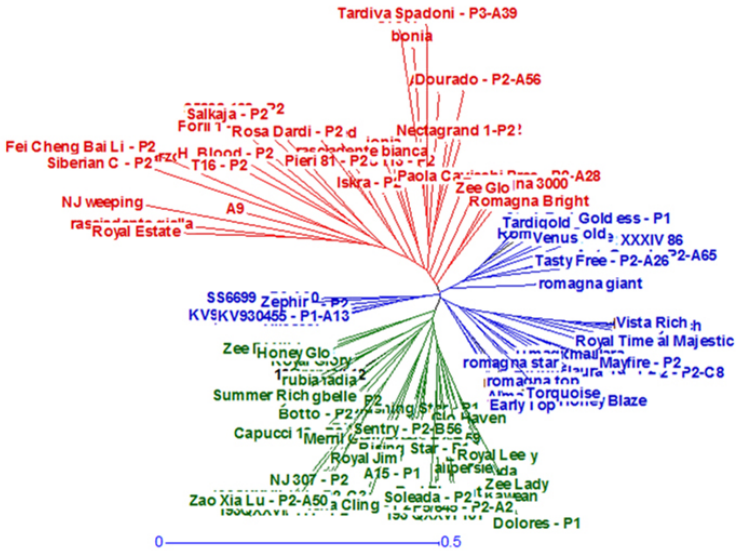

Figure 4.3: Phylogenetic tree of 123 peach accessions by 15 SSR markers; the tree was built by neighbour joining and simple matching coefficient.

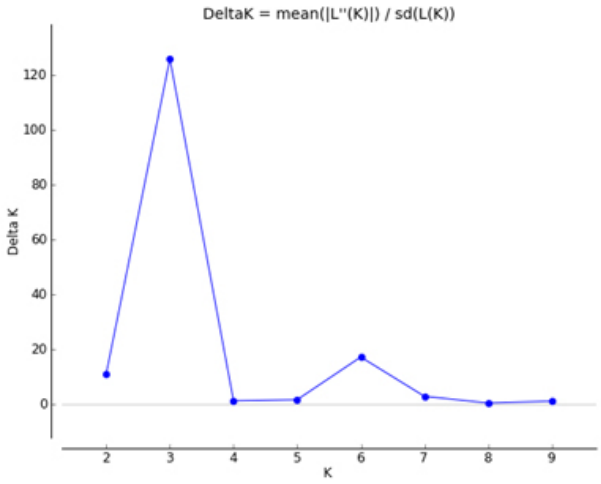

Figure 4.4: Number of clusters (K) in 123 accessions obtained by the method of Evanno et al. (2005)

three clusters (Fig. 4.4). This number is in agreement with that obtained from the phylogenetic tree built with the software *Darwin*. However, some accessions such as 'Paola Cavicchi Precoce', 'Romagna 3000', 'Zee Glo', 'Romagna Bright' and 'Cristina' were identified as belonging to cluster 1 by the software *Darwin*, although they were classified as belonging to cluster 3 by the software *Structure*.

In Table 4.5, it is possible to observe the values of ancestry expressed for the 123 accessions. The majority of accessions showed, in their cluster, high values of ancestry (Fig. 4.5).

The accessions 'Hardired', 'IFF 691', 'Nectagrand 1', and 'NJ113' in cluster 1, 'Helena Cling', 'Iskra', 'Paola Cavicchi Precoce', 'Royal Glory', 'Royal Jim', 'Yao Xia Lu' and 'Zee Diamond' in cluster 2, 'KV930386', 'Romagna 3000' and 'S 6699' in cluster 3 were the accessions showing ancestry values below 0.65.

Considering the number of clusters obtained by the software *Structure*, the number of alleles per locus (A), effective number of alleles (A), observed

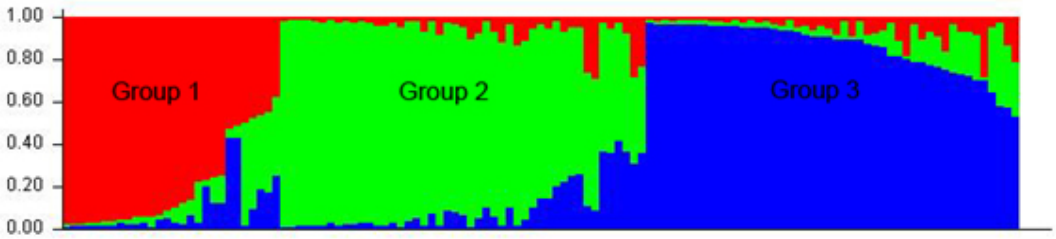

Figure 4.5: Ancestry representation of 123 accessions from the germplasm bank of 'MAS.PES'; values obtained with the software *Structure*.

heterozygosity (Ho), expected (He) heterozygosity, and Wright's fixation index (F) were analysed for each population (Table 4.6).

Marker BPPCT017 amplified ten alleles in cluster 1. The highest values of observed and expected heterozygosity were identified in markers BPCCT025 and BPPCT001, respectively. For all markers, the fixation rates were positive.

In cluster 2, some negative fixation indices were observed. Markers that amplified the highest number of alleles were EPPCU5176, BPPCT038, BPCCT025, BPPCT001 and UDP96-005.

As observed in cluster 2, six was the highest number of alleles amplified by marker in cluster 3. This value was observed in markers BPCCT025, BPPCT001 and BPPCT015. The values of both observed and expected heterozygosity ranged from 0.125 to 0.771 and 0.156 and 0.724, respectively.

Cluster 1 showed a higher average of amplified alleles per locus (6.40), expected heterozygosity (0.62) and Wright's fixation index (0.50), while cluster 2 had the highest average values of observed heterozygosity (0.49).

Table 4.5: Mean ancestry values for the three genetic clusters by the software *Structure*.

Accessions	K1	K2	K3	Cl.	Accessions	K1	K2	K3	Cl.
A219	0.75	0.12	0.13	1	Redhaven	0.05	0.69	0.26	2
A9	0.94	0.05	0.01	1	Rising Star	0.05	0.85	0.10	2
Andross	0.76	0.12	0.12	1	Rosa Dardi	0.07	0.92	0.02	2
A. Marzocchella	0.94	0.03	0.03	1	Royal Glory	0.02	0.56	0.42	2
Bonia	0.90	0.07	0.03	1	Royal Jim	0.07	0.57	0.37	2
Chimarrita	0.76	0.21	0.03	1	Royal Lee	0.02	0.93	0.05	2
Dourado	0.91	0.03	0.06	1	Rubia	0.01	0.96	0.03	2
Fei Cheng Bai Li	0.97	0.01	0.02	1	Sentry	0.02	0.89	0.08	2
Forli 1	0.94	0.02	0.03	1	Soleada	0.02	0.95	0.02	2
H. Blood	0.88	0.09	0.03	1	Springbelle	0.05	0.70	0.25	2
Hardired	0.46	0.44	0.09	1	Summer Rich	0.03	0.89	0.08	2
IF 691	0.53	0.04	0.43	1	Zao Xia Lu	0.48	0.50	0.02	2
Ionia	0.86	0.07	0.06	1	Zee Diamond	0.02	0.61	0.37	2
Nectagrand 1	0.46	0.35	0.19	1	Zee Lady	0.01	0.97	0.01	2
NJ weeping(PI 91459)	0.98	0.02	0.01	1	189CXIIKLI62	0.02	0.02	0.96	3
NJC113	0.46	0.38	0.17	1	Aliblanca	0.03	0.23	0.74	3
Oro A	0.98	0.01	0.02	1	Alice Col	0.08	0.14	0.77	3
Rasciadente Bianca	0.77	0.03	0.20	1	Alitop	0.04	0.02	0.94	3
Rasciadente Gialla	0.95	0.02	0.03	1	Alma	0.02	0.02	0.96	3
Royal Estate	0.97	0.01	0.02	1	Ambra	0.08	0.02	0.90	3
S5898:128	0.96	0.01	0.03	1	Antony	0.06	0.08	0.87	3
Salkaja	0.93	0.02	0.05	1	Aut. Grand	0.01	0.02	0.97	3
Siberian C	0.96	0.02	0.02	1	Azurite	0.04	0.05	0.91	3
T16	0.96	0.02	0.02	1	Big Top	0.01	0.03	0.96	3
Tardiva Spadoni	0.97	0.01	0.02	1	Bordó	0.04	0.30	0.65	3
391 C12 XXXIV 86	0.50	0.05	0.44	1	Corindon	0.08	0.21	0.71	3
193QXXVII 111	0.05	0.94	0.01	2	Cristina	0.08	0.04	0.88	3
193QXXVII 111	0.02	0.96	0.02	2	Early Rich	0.01	0.02	0.96	3
194 RXXVI 12	0.02	0.94	0.04	2	Early Top	0.03	0.18	0.79	3
193 QXXVI 131	0.01	0.97	0.02	2	Early Zee	0.06	0.02	0.91	3
A15 – P1	0.04	0.94	0.02	2	Francesca	0.27	0.01	0.72	3
Alirosada	0.03	0.94	0.03	2	Honey Blaze	0.10	0.11	0.79	3
Alipersié	0.02	0.97	0.02	2	Honey Glo	0.02	0.08	0.90	3
Alired	0.05	0.58	0.37	2	Honey Kist	0.04	0.04	0.92	3
Blazing Star	0.01	0.97	0.02	2	Honey Royal	0.07	0.05	0.87	3
Blushing Star	0.03	0.82	0.15	2	June Princess	0.02	0.02	0.96	3
Bolero	0.02	0.88	0.10	2	KV930386	0.13	0.29	0.58	3
Botto	0.10	0.85	0.05	2	KV930455	0.07	0.20	0.73	3
Capucci 18	0.27	0.65	0.09	2	Laura	0.17	0.02	0.81	3
Contender	0.07	0.88	0.05	2	Magique	0.01	0.01	0.97	3
Dixired	0.09	0.90	0.02	2	Maillar	0.02	0.01	0.97	3
Dolores	0.02	0.97	0.01	2	Mayfire	0.08	0.02	0.90	3
Elbertita	0.02	0.95	0.03	2	Rich Lady	0.01	0.02	0.96	3
Elegant Lady	0.01	0.96	0.03	2	Rom. 3000	0.02	0.39	0.59	3
GloHaven	0.02	0.95	0.03	2	Rom. Bright	0.01	0.05	0.94	3
Grenat	0.03	0.86	0.10	2	Rom. Giant	0.05	0.04	0.91	3
Helena Cling	0.36	0.38	0.26	2	Rom. Gold	0.01	0.02	0.97	3
IFF 331	0.10	0.89	0.01	2	Rom. Red	0.02	0.03	0.95	3
Iskra	0.23	0.41	0.36	2	Rom. Star	0.02	0.09	0.89	3
Kakamas	0.11	0.87	0.02	2	Romagna Top	0.02	0.16	0.82	3
Kaweah	0.04	0.95	0.02	2	Royal Majestic	0.01	0.01	0.97	3
Kevina	0.24	0.65	0.11	2	Royal Time	0.06	0.20	0.74	3
Maura	0.01	0.79	0.20	2	Ruby Rich	0.01	0.02	0.97	3
Maycrest	0.07	0.71	0.22	2	S 6699	0.20	0.26	0.54	3
Merril Gem Free	0.04	0.89	0.07	2	(Stark) RedGold	0.02	0.02	0.96	3
Nadia	0.05	0.80	0.15	2	Tardigold	0.01	0.02	0.97	3
NJ 307	0.03	0.96	0.02	2	Tasty Free	0.01	0.02	0.97	3
P5/645	0.13	0.86	0.02	2	Torquoise	0.10	0.07	0.83	3
P. Cavicchi Precoce	0.27	0.42	0.31	2	Venus	0.05	0.02	0.94	3
Pieri 81	0.07	0.87	0.06	2	Vista Rich	0.03	0.03	0.95	3
Red Elegant	0.02	0.91	0.08	2	Zee Glo	0.15	0.09	0.76	3
					Zephir	0.07	0.16	0.77	3

Table 4.6: Variability of three peach clusters identified by population structure analysis with 15 microsatellite markers; locus name, number of alleles per locus (A), effective number of alleles (Ae), observed heterozygosity (Ho), expected heterozygosity (He) and Wright's fixation index (F) are shown.

Cluster	Locus	N	A	Ae	Ho	He	F
Cluster 1	BPPCT017	26	10	3.896	0.385	0.743	0.483
	UDP96-008	26	4	1.957	0.038	0.489	0.921
	EPPCU5176	26	7	2.748	0.385	0.636	0.395
	CPPCT045	26	4	2.039	0.308	0.510	0.396
	BPPCT038	25	8	2.178	0.360	0.541	0.334
	UDP98-022	26	5	4.375	0.462	0.771	0.402
	CPPCT006	25	4	2.822	0.360	0.646	0.442
	CPSCT039	26	6	1.945	0.304	0.486	0.374
	BPCCT025	26	9	4.552	0.500	0.780	0.359
	UDP98-412	26	6	4.048	0.462	0.753	0.387
	UDP98-409	26	5	1.714	0.192	0.416	0.538
	BPPCT001	24	8	4.702	0.375	0.787	0.524
	BPPCT015	26	6	1.781	0.000	0.439	1.000
	UDP96-005	26	6	3.549	0.423	0.718	0.411
	BPPCT007	26	8	2.585	0.269	0.613	0.561
	Average		6.40	2.99	0.32	0.62	0.50
Cluster 2	BPPCT017	48	4	1.777	0.542	0.437	-0.239
	UDP96-008	49	5	2.028	0.286	0.507	0.436
	EPPCU5176	49	6	3.018	0.592	0.669	0.115
	CPPCT045	49	4	2.564	0.694	0.610	-0.138
	BPPCT038	49	6	3.011	0.592	0.668	0.114
	UDP98-022	49	4	2.030	0.469	0.507	0.075
	CPPCT006	48	4	2.801	0.583	0.643	0.093
	CPSCT039	46	3	2.203	0.568	0.546	-0.039
	BPCCT025	49	6	1.547	0.204	0.354	0.423
	UDP98-412	49	5	2.199	0.592	0.545	-0.086
	UDP98-409	49	4	1.348	0.204	0.258	0.210
	BPPCT001	47	6	2.861	0.404	0.651	0.379
	BPPCT015	46	3	2.205	0.457	0.547	0.165
	UDP96-005	49	6	2.347	0.592	0.574	-0.031
	BPPCT007	49	5	2.043	0.571	0.511	-0.119
	Average		4.73	2.27	0.49	0.54	0.09
Cluster 3	BPPCT017	48	2	1.306	0.229	0.234	0.021
	UDP96-008	48	3	1.868	0.313	0.465	0.327
	EPPCU5176	48	4	2.657	0.688	0.624	-0.102
	CPPCT045	48	4	1.989	0.521	0.497	-0.048
	BPPCT038	48	3	1.185	0.125	0.156	0.200
	UDP98-022	48	5	2.004	0.479	0.501	0.044
	CPPCT006	48	5	1.990	0.333	0.498	0.330
	CPSCT039	46	5	1.889	0.391	0.471	0.169
	BPCCT025	48	6	2.238	0.521	0.553	0.058
	UDP98-412	48	5	3.625	0.771	0.724	-0.064
	UDP98-409	48	5	2.379	0.604	0.580	-0.042
	BPPCT001	47	6	2.901	0.595	0.655	0.092
	BPPCT015	47	6	2.066	0.511	0.516	0.011
	UDP96-005	47	4	1.330	0.277	0.248	-0.114
	BPPCT007	48	5	2.472	0.625	0.595	-0.050
	Average		4.53	2.13	0.47	0.49	0.06

4.4 Discussion

The genetic diversity within peach accessions from the 'MAS.PES' germplasm collection (Italy) was analysed in order to obtain a base for future research on association mapping.

For this purpose, AFLP and microsatellite markers were used in this study. Both systems for polymorphism detection demonstrated their efficiency in the analysis of genetic diversity of peach accessions.

In the case of AFLP markers, 171 polymorphic fragments were obtained and it was possible to verify the formation of two groups in the dendrogram, while some subdivisions were observed in the second group. However, it was not possible to separate the groups based on their genetic or geographic origin, nor on any phenotypic trait. This result was similar to that observed by Aranzana et al. (2003) who evaluated 210 peach cultivars by nine AFLP primer combinations. However, in this study, the presence of yellow fleshed and melting peaches such as 'Tardibelle', 'Elegant Lady', 'Vista Rich' and 'FlavorCrest' was observed in subgroup 2.3.

Another factor that hindered the separation of groups is that many accessions were rather similar as per their genetic distance. This occurs because most of the studied accessions have high level of inbreeding, despite their distinct geographical origin.

Considering the limitations associated with the dominant nature of AFLP markers, SSR markers were chosen to complete the analysis of genetic diversity. The analysis of diversity by SSR markers was performed using a smaller number of accessions. This occurred because some accessions were still in non fruiting stage, complicating the phenotypic analyses performed in subsequent stages (chapter 6).

Markers BPCCT025, BPPCT001, and BPPCT017 showed the highest

polymorphism. Aranzana et al. (2010) obtained similar results using the same markers in a study of genetic diversity on 224 peach accessions, detecting amplification of 10, 9 and 9 alleles, respectively. The average of 7.7 allele per locus observed in the present study was greater than that observed by Li et al. (2008); Aranzana et al. (2010); Bouhadida et al. (2011); Sitther et al. (2012). However, this value depends on the number and variability of the accessions used in each study (Garcia et al., 2012).

In agreement with Bouhadida et al. (2011), marker BPPCT001 was the most informative, with a total of 3.74 effective alleles (Ae). The observed heterozygosity was lower than expected, suggesting a high level of homozygosis compared to the expected value according to the equilibrium of Hardy-Weinberg. Observed values below the expected heterozygosity were also reported by Aranzana et al. (2002); Chalak et al. (2006); Aranzana et al. (2010); Paula et al. (2012) which suggest the presence of inbreeding that occurs naturally in autogamous species such as peach. The number of clusters obtained by analysis of population structure was in accordance with the analysis performed by the software *Darwin*, although it was observed that some accessions were shown in a different group compared to *Structure*. However, this could be due to the different approaches adopted in these methods to classify accessions within each group. For example, the accession 'Zee Glo', was shown in group 1 of the *Darwin* dendrogram, together with 'NJ Weeping', 'Fei Cheng Bai Li' among others. However, the software *Structure* placed 'Zee Glo' (76% of co-ancestry) in group 3, together with 'Ruby Rich', 'Vista Rich' and 'Tardigold', among others. However, approximately 15% of the ancestry of this accession also belongs to group 1.

The accessions 'Stark Redgold' and 'Venus' were classified in the same group in both approaches. 'Stark Redgold' is one of the parents of 'Venus'. A similar situation was observed in accessions 'Ambra' and 'Mayfire', where 'Mayfire' is a parent of 'Ambra'. The accessions 'Contender' and 'Redhaven', 'Elegant Lady' and 'Merrill Gem Free', 'Blazing Star' and 'Blushing Star', genetically related, were also classified in the same group by both methods,

demonstrating their efficiency.

The number of alleles per locus (A), effective number of alleles (Ae) observed heterozygosity (Ho) and expected heterozygosity (He), and Wright's fixation index (F) were analysed considering the three clusters obtained by the software *Structure*. Studying the genetic diversity of grapevines, Andres et al. (2012) also evaluated these parameters in the clusters identified by *Structure*.

In some loci of clusters 2 and 3, negative fixation indices were observed, indicating high heterozygosity for these loci. The most informative markers were BPPCT001 in cluster 1, EPPCU5176 in cluster 2 and UDP98-412 in cluster 3. Cluster 1 showed the greatest mean value of amplified alleles, while exhibiting a greater homozygosity. Cluster 2 showed the highest observed heterozygosity. This value was rather similar to that observed in cluster 3. The mean fixation indices were positive in all clusters, revealing high level of homozygotes relative to the expected value in the Hardy-Weinberg equilibrium.

As previously mentioned, it was possible to assess the genetic diversity through both markers. Although the four primer combinations allowed a higher amplification of the fragments (171), the SSR markers have proved to be more informative, showing a better distribution of the accessions in groups, including the use of different approaches. SSR markers are also advantageous in view of the future goal for the selection of accessions (panel of diversity) to be assessed in studies of association mapping.

Although AFLP markers allow the construction of association mapping, studies have shown that the use of such markers requires adapted statistical methods to determine the genetic structure of the population (Zhu et al., 2008). Moreover, markers such as SSRs and SNPs are more powerful and informative in determining the population structure and kinship matrix (Zhu et al., 2008). For the next stages of analysis, a higher number of microsatellite

markers should be used, namely in the process of detecting markers associated with traits of interest.

Chapter 5

Linkage analysis and QTL mapping in a F2 population

5.1 Introduction

Genetic maps are important instruments for plant breeding. Linkage maps provide a basis for genetic dissection of quantitative traits by mapping Quantitative Trait Loci (QTLs), besides allowing the identification of individual genes and the construction of genome-wide physical maps. This is exploited in Marker-Assisted Selection (MAS) in diverse crops (Ogundiwin et al., 2009).

Peach linkage maps are available in the literature, including a total of more than 2000 markers (Shulaev et al., 2008). Using nine different F2 families, Chaparro et al. (1994) constructed a linkage map using isozymes and Random Amplification of Polymorphic DNA (RAPD) markers. Rajapakse et al. (1995) built another map composed of Restriction Fragment Length Polymorphism (RFLP) markers, RAPD markers and morphological markers and based on 71 F2 individuals from the cross 'New Jersey Pillar' × 'KV 77119'. In the study by Lu et al. (1998) with 169 AFLP markers in 55 individuals of an F2 breeding population derived from the crossing of 'Lovell' × 'Nemared', two genes involved in resistance to nematodes were mapped in linkage group 1. The map developed by Dirlewanger et al. (1998) com-

bined morphological traits, isozymes, RFLPs, RAPDs, inter-microsatellite amplification (IMA), and AFLPs in an F2 breeding population composed of 63 individuals derived from the crossing of the non-acid peach 'Ferjalou' × 'Jalousia' and the acid round nectarine 'Fantasia'. The same population was further characterised, increasing the number of individuals to 207 and adding 82 new microsatellite markers (Dirlewanger et al., 2006). A genetic linkage map of 211 markers was constructed for a peach progeny population, Pop-DG, derived from the peach cultivar 'Dr. Davis' and the fresh market cultivar 'Georgia Belle' (Ogundiwin et al., 2009). Using an F2 population developed from 'Contender' × 'Fla.92-2C', contrasting in the trait chilling requirement, Fan et al. (2010) constructed a map using AFLP and SSR markers.

Maps from interspecific progenies have also been published, such as the Prunus reference map derived from a cross of almond ('Texas') with peach ('Earlygold'), which presents a higher saturation, with a total of 562 markers and an average density of 0.9 cM between markers (Joobeur et al., 1998; Aranzana et al., 2003; Dirlewanger et al., 2004; Ogundiwin et al., 2009). Three populations (F1, F2 and BC2) derived from a cross between the clone P1908 of *P. davidiana* and the peach cultivar 'Summergrand' were used in the construction of the map proposed by Foulongne et al. (2003). Verde et al. (2005) increased the number of SSR and AFLP markers in order to enrich the map obtained by crossing a ('*P. persica*' × '*P. ferganensis*') × '*P. persica* BC1' progeny developed in previous studies by Quarta et al. (1998, 2000); Dettori et al. (2001).

Once the maps are constructed, they can be used for the genetic dissection of quantitative traits of agronomic importance. In this approach, the positions of QTLs in the genome are determined and their genetic effects estimated, such as the additive, dominance and other effects in the adopted model (Toledo et al., 2008).

QTLs for different traits have been mapped in peach. Yamamoto et al. (2001) mapped sixteen QTLs related to flowering time, ripening time, fruit

dropping time and fruit weight. Fan et al. (2010) identified 20 QTLs associated with the traits chilling requirement, heat requirement and bloom date in an F2 population developed from the cross 'Contender' × 'Fla.92-2C', contrasting in the trait chilling requirement. Maturity date, external fruit skin overcolour, juice total soluble solids, titrable acidity and pH were studied by Eduardo et al. (2011), in an F1 population derived from 'Bolero' × 'OroA' and an F2 population of 'Contender' × 'Ambra'; the authors detected up to two QTLs per trait in each population, with some of the traits being located in the same region. Pacheco Cruz (2010) identified markers associated with resistance to brown rot in an F1 population from 'Contender' × 'Elegant Lady'. Han *et al.* (2012) identified a marker (Me07Em02) linked to the locus Cs (flesh colour around the stone) in F1 progenies derived from 'Chongyanghong' × 'Yanhong'. Ogundiwin et al. (2009) determined the genomic locations of 133 fruit quality candidate genes with an intraspecific peach population, Pop-DG, and the Prunus reference map, T × E.

Referring to QTLs associated with fruit weight, a two-year study by Dirlewanger et al. (1999) identified a reproducible QTL in chromosome 6, by assessing an F2 population from the cross 'Ferjalou Jalousia' × 'Fantasia'. Consistent results were obtained in the same population by Etienne et al. (2002). Evaluating F2 seedlings from cross 'Akame' × 'Juseito', Yamamoto et al. (2001) detected three QTLs in linkage group 3 and one QTL in linkage group 6. Cantín et al. (2010) identified QTLs in linkage group 4 in a population from 'Venus' × 'Big Top'. Consistent with these previous studies, Eduardo et al. (2011) found stable QTLs on linkage groups 4 and 6 in the F2 population from 'Contender' × 'Ambra' over two years. One problem with integrating and comparing data from these studies was linked to the use of different molecular markers that make comparisons of genetic positions difficult. However, the recent publication of the peach genome sequence (Arús et al., 2012) provides a new basis for these comparative approaches allowing alignment of different genetics maps to the reference sequence.

Most of the above-mentioned maps and the mapped QTLs were obtained

using markers of the type RFLP, AFLP and SSR. These maps are mostly low-density maps, and QTLs placed on the maps are unable to provide precise and complete information about the numbers and the locations of the genes or QTLs controlling the traits (Martínez-García et al., 2012). However, the availability of the peach genome along with the ease of data acquisition by using next generation sequencing has allowed for efficiently identifying a large number of sequence-based markers such as SNPs. SNP markers have several advantages for genetic mapping over other molecular markers. SNPs have fewer detection/evaluation errors than SSRs and can result in greater precision in QTL mapping (Ball et al., 2010; Yu et al., 2011; Martínez-García et al., 2012).

Although some studies have already been conducted to detect QTLs associated with fruit weight, the identification of genes controlling fruit weight has been one of the goals of breeding programs in peach in recent years, since consumers are naturally driven to buy an aesthetically homogeneous fruit of large size, free of physiological and pathological variations (Bertoglio, 2010).

In this context, the aim of the present study is to construct a linkage map using a SNP genotyping array for peach (9K SNP array v1) and to identify QTLs associated with fruit weight, using an F2 population from a cross of two accessions highly contrasting for this trait.

5.2 Material and methods

5.2.1 Plant material

An F2 population of 123 individuals was derived from self-pollination of an F1 seedling derived from a cross 'NJ Weeping' × 'Bounty'.

Individuals of the F2 population are located in the experimental field in Tebano Faluza in Emilia Romagna region, Italy. The population segregates for maturity date, fruit weight and size, flesh texture and colour, peduncle

length, flower type and colour and tree habit. In the present study, only the data concerning fruit weight and size were evaluated.

'NJ Weeping' is an ornamental peach with small white fruits and late ripening. 'Bounty' is a medium ripening peach variety with large yellow fruits (more than 160 g).

5.2.2 Total DNA extraction

DNA was extracted from young leaves, using the "DNeasy Plant Mini Kit" (Qiagen) according to the manufacturer's instructions. For DNA quantification, Quant-iT Picogreen (Invitrogen) was used and DNA concentration of each sample was estimated based on a standard concentration curve. The final concentrations of all DNA samples were adjusted to 50 ng/μl.

5.2.3 Genotyping

The 123 F2 plants were genotyped by means of the 9K SNP array v1 (Illumina Peach Infinium Chip), developed by the International Peach SNP Consortium (IPSC), with a total of 8144 SNPs distributed over eight peach chromosomes (Verde et al., 2012). For linkage analyses, SNPs were initially selected discarding markers that showed more than 20% missing data. Then, the polymorphism of the SNPs in relation to the 123 individuals was evaluated. Finally, the SNPs were selected based on the GenTrain scores. GenTrain scores correspond to the reliability of SNP detection based on the distribution of genotype classes and range from 0 to 1. According to Illumina, the recommended value for maintaining a SNP is a GenTrain score of 0.25 (Fan et al., 2003). However, Esteras et al. (2012) used the GenTrain score of 0.6 for their SNP-based genetic map of *Cucurbita pepo*. Thus, the present study considered only SNP markers with GenTrain scores greater than 0.6.

5.2.4 Analysis of phenotypic characteristics

In order to obtain fruits showing the maximum phenotypic expression, thinning of fruits in all 123 individuals of the F2 population was performed in spring 2011. At the time of collection (summer 2011), ten fruits per plant with the greatest weight were assessed by standard scales. Fruit height, width and depth were measured by a calliper. Fruit height (FH) is the distance between the apex and the stem cavity of the fruit, fruit width (FWD) represents the distance between the two sides of the fruit, and fruit depth (FD) is the distance between the suture line and the opposite side to it.

Genetic parameters for the F2 generation were estimated according to Allard (1960) in the software *GENES* (Cruz, 2007): phenotypic (σ^2_P), environmental (σ^2_E) and genetic variance (σ^2_G), where phenotypic variance $\sigma^2_P = \sigma^2_{F2}$; environment variance $\sigma^2_E = \frac{1}{4}[\sigma^2_{P1} + 2\sigma^2_{F1} + \sigma^2_{P2}]$; with σ^2_{P1} being the variance of the first parent, σ^2_{P2} the variance of the second parent, σ^2_{F1} the variance of F1 and σ^2_{F2} the variance of F2.

Broad-sense heritability (h^2) for the traits FW, FH, FWD and FD was estimated by the model proposed by Allard (1960): $h^2 = \sigma^2_G/\sigma^2_P$.

Considering the importance that the data follow a normal distribution, a normality analysis using the Shapiro–Wilk test was performed in the program *PAST* (Hammer et al., 2001), which was also used to construct histograms of frequency distribution.

5.2.5 Construction of the genetic map

The genetic map was build in the software *JoinMap* v. 3.0 (van Ooijen, 2006), considering the markers segregating in the F2 progeny. The recombination threshold value was set to 0.40 and the Kosambi mapping function was used to convert recombination frequencies into map distances. The linkage groups were constructed with the option "Create groups using the groupings trees". Markers classified by *JoinMap* as suspect linkage were removed from the map

calculation after review in order to avoid bias in the analysis.

5.2.6 Identification of QTLs

The software *MapQTL* v. 4.0 (Van Ooijen, 2009) was used for detecting and mapping QTLs in the 'NJ Weeping' × 'Bounty' F2 population. First, the data were subjected to the random permutation test (with 10,000 replicates) in order to calculate the critical LOD (logarithm of odds) of each QTL. Then, Interval Mapping (IM) and Multiple QTL Mapping (MQM) were employed. The detection of QTLs was performed via IM with a 95% significance ($p>0.05$) to identify QTLs with significant main effects. Subsequently, the module MQM was used to detect possible QTLs masked by QTLs identified by IM. Using the option "automatic cofactor selection" in the MGM strategy, markers that flank the QTLs identified by IM were detected as cofactors.

5.3 Results

5.3.1 Phenotypic traits

The results concerning the phenotypic data and heritability are presented in tables 5.1 and 5.2. There are considerably contrasting values between the parents as evidenced by the parental mean values (Table 5.1). 'NJ Weeping' revealed low values of weight and fruit dimensions, while 'Bounty' was heavier and larger (Fig. 5.1). The divergence between the parents was important for the study, because it allowed a greater segregation in the F2 breeding population for the evaluated traits.

The evaluated traits showed heritability coefficients above 80%, indicating a high contribution of genetic variance in relation to phenotypic variance (Table 5.2). The highest heritability coefficient was found for fruit weight (88.62%), while fruit height showed the lowest coefficient (84.51%).

Table 5.1: Average, maximum and minimum values of fruit weight and size for 'NJ Weeping' (P1) and 'Bounty' (P2) parents and the F2 population.

Statistical Descriptor	Weight (g)	Height (mm)	Width (mm)	Depth (mm)
Mean Value P1	42.47	40.10	38.40	39.60
Maximum Value P1	51.91	42.03	42.30	43.60
Minimum Value P1	34.45	37.10	35.47	36.10
Standart desviation P1	5.47	2.10	2.01	1.96
Mean Value P2	160.02	60.06	60.01	63.30
Maximum Value P2	188.39	67.15	65.67	67.50
Minimum Value P2	139.27	55.20	58.67	60.80
Standart desviation P2	2.15	2.78	1.15	2.25
Mean Value F2	88.59	52.22	52.14	53.14
Maximum Value F2	199.55	69.20	74.20	70.60
Minimum Value F2	32.96	38.80	38.47	38.20
Standar desviation F2	29.45	6.19	6.34	6.16

Figure 5.1: Graphical presentation of the mean values for 'NJ Weeping' (P1) 'Bounty' (P2), and their F2 progeny; data from Table 5.1

Table 5.2: Estimates of phenotypic (σ_P^2), environmental (σ_E^2), genotypic variance (σ_G^2) and broad-sense heritability (h^2) of fruit weight and size in the F2 population ('NJ Weeping' × 'Bounty').

	Trait			
	Weight (g)	Height (mm)	Width (mm)	Depth (mm)
Phenotypic Variance (σ_P^2)	860.93	38.23	39.94	37.73
Environmental Variance (σ_E^2)	97.91	5.91	5.36	4.51
Genetic Variance (σ_G^2)	763.02	32.31	34.57	33.21
Broad-sense Heritability (h^2)	88.62%	84.51%	86.57%	88.02%

Referring to the normality test of Shapiro–Wilk (Table 5.3), the Ho hypothesis was accepted for all traits indicating that all data have a normal distribution (Fig. 5.2).

Table 5.3: Shapiro–Wilk test values for the quantitative traits

Traits	Shapiro–Wilk test
Weight	0.0918
Height	0.4218
Width	0.3096
Depth	0.9762

Shapiro–Wilk $p_{normal} < 0.05$ normal distribution can be rejected.

All correlations were significant at the level of 1% probability (Table 5.4). The highest correlation was found between FD and FWD with a value of 0.979. The traits related to fruit size were correlated with FW (0.841, 0.950 and 0.940, respectively) demonstrating that heavier fruits have higher values of height, width and depth, as expected.

5.3.2 Linkage mapping

After SNP filtering, a total of 2,390 markers that segregated in the F2 population were obtained. Finally, 877 SNP markers were used to construct a genetic map. The reduction is due to the elimination of some markers, which, due to high recombination frequencies, showed a tendency to broaden the genetic distance and were then classified as suspect linkage by *JoinMap*; other markers were eliminated because they were found in the same loci. The 877

Table 5.4: Correlation values for fruit weight and size.

	Fruit Weight	Fruit Height	Fruit Width	Fruit Depth
Fruit Weight	1			
Fruit Height	0.841**	1		
Fruit Width	0.950**	0.891**	1	
Fruit Depth	0.940**	0.892**	0.979**	1

**: Significant at 1% probability by t-test

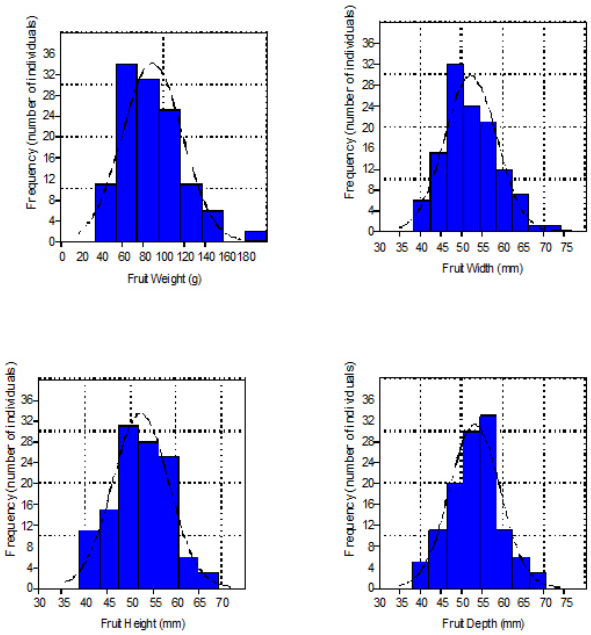

Figure 5.2: Frequency distribution of fruit weight, fruit height, fruit width and fruit depth in the F2 population ('NJ Weeping' × 'Bounty')

markers were distributed in eight linkage groups, which correspond to the eight chromosomes of the peach genome, with a total coverage of 541.06 cM and an average distance between markers of 0.61 cM. Thus, the map features a high density of markers (Fig. 5.3)

In almost all linkage groups, the total number of markers was above 100, with the exception of linkage groups 2 and 5, which showed a total number of 51 and 91 markers, respectively. The highest numbers were found in linkage groups 1 and 4, with 180 and 121 markers, respectively.

The size of the linkage groups ranged from 108.69 cM in linkage group 1 to 56.03 cM in linkage group 2. The other linkage groups 3 to 8 spanned 61.34 cM, 66.55 cM, 61.80 cM, 60.71 cM, 64.40 cM and 61.57 cM, respectively.

The genetic distance between pairs of adjacent markers varied from 0.004 to 7.446 cM for SNP_IGA_79719 and SNP_IGA_79809 in linkage group 1 and

SNP_2_22274363 and SNP_IGA_286418 in linkage group 2, in concordance with the study by Martínez-García et al. (2012) who constructed an SNP map and found the largest genetic distance between pairs of adjacent markers in the linkage group 2. Large areas without marker coverage were not observed.

According to the χ^2 test, the following markers showed segregation distortion: SNP_IGA_65694, SNP_IGA_66456, SNP_IGA_67620. Markers SNP_IGA_69306, SNP_IGA_79719, SNP_IGA_79809, SNP_IGA_80547, SNP_IGA_82861, SNP_IGA_83053, SNP_IGA_84580 were significant at the level $p<0.05$, while SNP_IGA_78367 and SNP_IGA_79455 at $p<0.01$. However, these markers were not removed from the analysis, given that the method for the construction of linkage groups adopted by the *JoinMap* software is based on the independence of the LOD score which is not affected by the distortion of segregation (van Ooijen, 2006; Pacheco Cruz, 2010; Alheit et al., 2011).

5.3.3 Analysis of QTLs

The values of critical LOD (threshold) obtained by permutation test were 3.5, 3.6, 3.5 and 3.6 for FW, FH, FWD, FD, respectively. The numbers of cofactors for each trait were 8, 6, 10, and 10, respectively.

QTLs associated with these traits were identified in almost all linkage group, with the exception of linkage group 8 (Table 5.5).

A total of eight QTLs were mapped for fruit weight, accounting for 54% of the phenotypic variation observed. For this trait, QTLs were mapped in linkage groups 1, 2, 3, 4, 6 and 7. The QTL associated with marker SNP_IGA_669440, located in linkage group 6, showed the highest LOD score with a value 10.98 (10.9% of the phenotypic variation observed). When considering this marker, the fruits of individuals with genotype AA, weighted on average 78.00 g, individuals with a heterozygous genotype (AB) 97.80 g

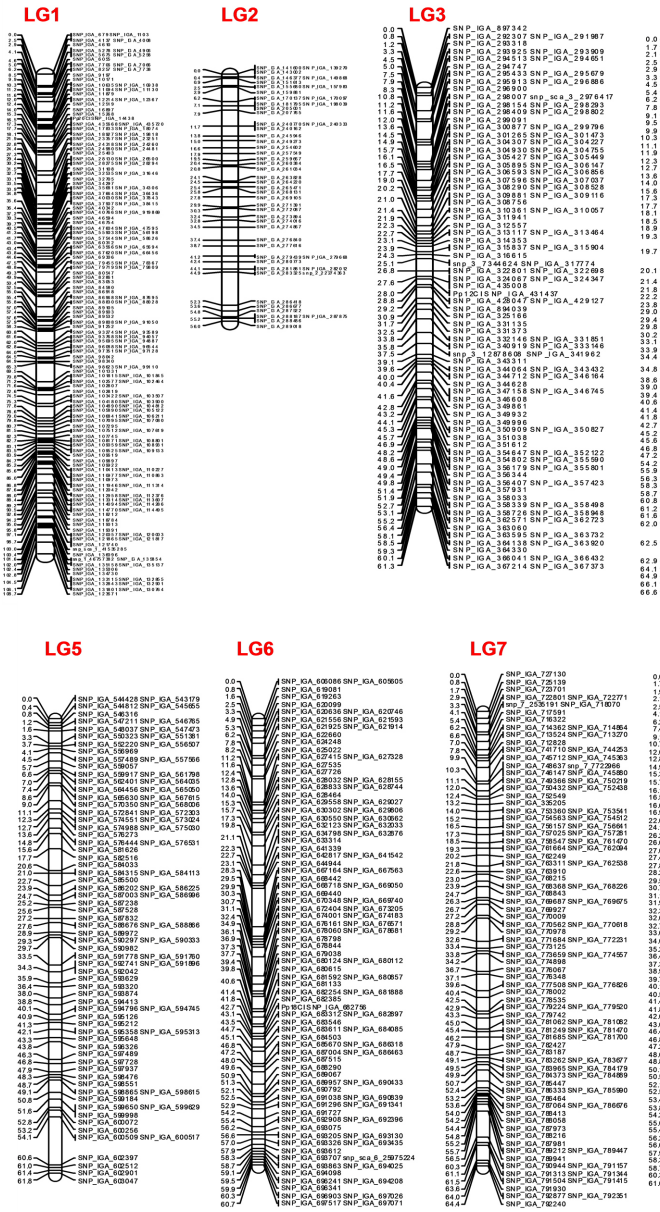

Figure 5.3: Genetic linkage mapping of eight chromosomes in the F2 population 'NJ Weeping' × 'Bounty', using 877 SNP markers; positions and distances between markers were calculated by means of the Kosambi function.

and those with the genotype BB 109.41 g (Table 5.6).

For fruit height, QTLs were identified in linkage groups 1, 2, 3, 4 and 7. Of the six QTLs mapped, two were identified in linkage group 1, which together explained 18.8% of the observed phenotypic variance. However, the QTL associated with marker SNP_IGA_776826 in linkage group 7 explained the highest percentage of the phenotypic variation (20.9%). The observed LOD score values varied from 15.23 to 3.51, while the coefficient of determination of all QTLs was 62.10% overall.

For fruit width, the ten mapped QTLs explained 55.7% of the observed phenotypic variance. The markers SNP_IGA_678798, SNP_IGA_776826 and SNP_IGA_32535 in linkage groups 6, 7 and 1 were associated with QTLs with larger effects and explained 12.1%, 7.5% and 7.0% of the observed phenotypic variance with LOD scores of 14.35, 9.85 and 9.31, respectively. Considering the marker with the greatest effect on fruit width, individuals with genotype AA showed an estimated average of 47.70 mm, individuals with genotype BB 54.92 mm and heterozygous individuals (AB) 50.80 mm (Table 5.6).

Ten QTLs were detected for fruit depth. LOD scores ranged from 4.31 to 14.43, the latter was attributed to marker SNP_IGA_678798, located in linkage group 6, (explaining 12.3% of the phenotypic variance). The QTLs observed for this trait were the same as observed for fruit width, differing in the LOD scores and the percentage of phenotypic variance explained for both additive and dominance effects.

5.4 Discussion

Although linkage maps were previously constructed with the aim to map QTLs associated with fruit weight and size in peach, the development of the map of the F2 population from 'NJ Weeping' × 'Bounty' provides an important tool for genetic analysis, considering that parents are highly contrasting for this trait.

Table 5.5: Linkage group locations, positions, markers and variations explained (%) by the QTLs which control fruit weight and size in the 'NJ Weeping' × 'Bounty' F2 population.

QTLs	LG	Position	Nearest markers	LOD	% Expl.
Fruit Weight	6	30.305	SNP_IGA_669440	10.98	10.9
	7	40.44	SNP_IGA_778002	9.63	9.2
	4	41.84	SNP_IGA_410794	8.59	8.1
	1	100.38	SNP_IGA_136096	6.22	5.6
	2	24.126	SNP_IGA_263828	5.93	5.3
	7	6.602	SNP_IGA_713270	5.3	4.7
	1	32.709	SNP_IGA_32535	4.77	4.2
	3	48.992	SNP_IGA_356179	4.6	4
Fruit Height	7	39.621	SNP_IGA_776826	15.23	20.9
	1	100.38	SNP_IGA_136096	11.09	14
	2	13.751	SNP_IGA_245946	10.59	13.2
	4	64.092	SNP_IGA_513616	4.92	5.4
	1	31.474	SNP_IGA_31108	4.42	4.8
	3	48.583	SNP_IGA_355590	3.51	3.8
Fruit Width	6	36.911	SNP_IGA_678798	14.35	12.1
	7	39.621	SNP_IGA_776826	9.85	7.5
	1	32.709	SNP_IGA_32535	9.31	7
	1	100.38	SNP_IGA_136096	8	5.9
	4	64.092	SNP_IGA_513616	7.28	5.3
	2	14.986	SNP_IGA_249273	7	5
	4	33.126	SNP_IGA_404899	6.81	4.9
	5	12.333	SNP_IGA_574551	4.1	2.8
	3	48.583	SNP_IGA_354802	3.95	2.7
	7	6.602	SNP_IGA_713524	3.77	2.5
Fruit Depth	6	36.911	SNP_IGA_678798	14.43	12.3
	1	32.709	SNP_IGA_32535	9.96	7.7
	7	39.621	SNP_IGA_776826	9.27	7.1
	2	14.986	SNP_IGA_249273	7.99	6
	1	100.38	SNP_IGA_136096	7.4	5.5
	4	64.092	SNP_IGA_513616	7.11	5.2
	4	33.126	SNP_IGA_404899	4.72	3.3
	5	12.333	SNP_IGA_574551	4.46	3.1
	3	48.583	SNP_IGA_354802	4.31	3
	7	6.602	SNP_IGA_713524	4.33	3

Table 5.6: Linkage group locations, positions, markers, explained variations (%), estimated mean associated with individuals "AA", "AB" and "BB" and additive and dominant effects for QTLs which control fruit weight, analysed in 2011 in the 'NJ Weeping' × 'Bounty' F2 population.

QTLs	LG	Nearest markers	% Expl.	mu_AA	mu_AB	mu_BB	Additive	Dominance
FW	6	SNP_IGA_669440	10.9	78.70	97.84	109.41	-15.36	3.78
	7	SNP_IGA_778002	9.2	77.68	98.37	110.43	-16.37	4.31
	4	SNP_IGA_410794	8.1	66.00	94.07	122.11	-28.06	0.01
	1	SNP_IGA_136096	5.6	83.38	94.94	104.73	-10.67	0.88
	2	SNP_IGA_263828	5.3	83.44	97.12	104.67	-10.61	3.07
	7	SNP_IGA_713270	4.7	87.55	80.51	100.57	-6.51	-13.55
	1	SNP_IGA_32535	4.2	86.16	87.25	101.95	-7.89	-6.81
	3	SNP_IGA_356179	4	84.52	91.90	103.59	-9.53	-2.16
FH	7	SNP_IGA_776826	20.9	46.78	50.88	55.32	-4.27	-0.17
	1	SNP_IGA_136096	14	47.72	52.24	54.37	-3.33	1.19
	2	SNP_IGA_245946	13.2	47.72	50.02	54.38	-3.33	-1.03
	4	SNP_IGA_513616	5.4	48.55	52.75	53.54	-2.49	1.70
	1	SNP_IGA_31108	4.8	49.05	50.09	53.05	-2.00	-0.96
	3	SNP_IGA_355590	3.8	49.12	50.60	52.98	-1.93	-0.45
FWD	6	SNP_IGA_678798	12.1	47.71	50.81	54.92	-3.61	-0.51
	7	SNP_IGA_776826	7.5	48.05	52.45	54.59	-3.27	1.14
	1	SNP_IGA_32535	7	48.76	50.09	53.87	-2.56	-1.22
	1	SNP_IGA_136096	5.9	49.17	52.73	53.46	-2.14	1.41
	4	SNP_IGA_513616	5.3	48.46	52.10	54.18	-2.86	0.78
	2	SNP_IGA_249273	5	49.12	50.71	53.52	-2.20	-0.61
	4	SNP_IGA_404899	4.9	48.00	51.58	54.63	-3.31	0.26
	5	SNP_IGA_574551	2.8	49.97	53.14	52.66	-1.34	1.82
	3	SNP_IGA_354802	2.7	49.55	51.08	53.08	-1.77	-0.24
	7	SNP_IGA_713524	2.5	49.78	49.69	52.85	-1.53	-1.63
FD	6	SNP_IGA_678798	12.3	48.46	51.78	55.60	-3.57	-0.25
	1	SNP_IGA_32535	7.7	49.33	51.10	54.73	-2.70	-0.94
	7	SNP_IGA_776826	7.1	48.94	52.87	55.12	-3.09	0.84
	2	SNP_IGA_249273	6	49.74	51.20	54.32	-2.29	-0.83
	1	SNP_IGA_136096	5.5	50.20	53.80	53.86	-1.83	1.77
	4	SNP_IGA_513616	5.2	49.29	53.13	54.77	-2.74	1.10
	4	SNP_IGA_404899	3.3	49.38	52.23	54.68	-2.65	0.20
	5	SNP_IGA_574551	3.1	50.49	53.56	53.57	-1.54	1.53
	3	SNP_IGA_354802	3	50.20	51.93	53.86	-1.83	-0.10
	7	SNP_IGA_713524	3	50.32	50.45	53.74	-1.71	-1.58

The results from the phenotypic analysis of the F2 population revealed a normal distribution for all examined traits. The high heritability coefficients (above 80%) demonstrate the great influence of the genetic variance on the observed phenotypic variance. When the trait has a low heritability, the accuracy and the detection power of QTLs may be reduced (Wu, 1999). On the other hand, when the heritability coefficients are high, such as those observed in this study, the QTLs explained more phenotypic variation (Li et al., 2010).

The use of the peach 9K SNP array V1 (Verde et al., 2012), developed by the International Peach SNP Consortium (IPSC), provided a large number of high quality SNPs for construction of a dense genetic map. It was however necessary to carry out a filtering step to discard SNPs with low GenTrain or GenCall scores. Myles et al. (2011); Mahanil et al. (2012) recommended the elimination of SNPs with GenTrain scores <0.3 or GenCall scores <0.2. These authors also suggest discarding individuals or SNPs with 20 % or more missing data. Considering the abundance of SNPs deriving from genotyping with the 9K SNP chip, in the present study, SNPs with GenTrain scores <0.6 were discarded in order to select high quality SNPs only. The 9K SNP array v1 (Verde et al., 2012) allowed the construction of a highly saturated map, with average distances between markers of 0.61 cM, providing an ideal basis for mapping of QTLs. (Martínez-García et al., 2012) obtained similar results, constructing a consensus SNPs map in peach, derived from 'Dr. Davis' × 'Georgia Belle' and 'Dr. Davis' × 'F8,1-42' with 588 SNPs covering 454.80 cM and an average density of 0.81 cM/marker. Myles et al. (2010) evaluated the efficiency of this technology using the Vitis9KSNP platform developed for grapevine. According to their results, Vitis9KSNP offers sufficient resolution to distinguish among *V. vinifera* cultivars, between *V. vinifera* and wild *Vitis* species, and even among diverse wild *Vitis* species.

Another advantage of the peach 9K SNP array v1 is the short time for genotyping compared to other methods. While it was possible to obtain data from 8,144 SNP markers in few months, it would take years to identify SSR

or AFLP markers that were polymorphic and allowed the construction of a map with the same density of markers as observed in this study.

All traits presented high and significant correlation coefficients. This is an expected result, since fruits with higher weight usually have a larger size as well. Correlating the data on fruit weight and size, Zhang et al. (2010) noted that these traits are highly correlated in sweet cherry as well.

The high correlation coefficients were confirmed in the QTL analysis, where it was possible to observe coincident QTL between correlated traits. This was the case for traits fruit width and fruit depth (showing a highly significant positive correlation 0.979). QTLs associated with FWD are associated also with FD. In peach, overlapping QTLs associated with FDW and FD were also reported in the study by Cantín et al. (2010) in linkage group 4. Also, Santos et al. (2010) identified QTL clusters for size and weight in linkage groups F4 and M11 in cashew apple.

QTLs associated with fruit weight were identified in linkage groups 1, 2, 3, 4, 6 and 7, whereby linkage group 6 explained most of the observed phenotypic variance. Dirlewanger et al. (1999) have mapped a QTL on chromosome 6 after two years of assessment. Similarly, Yamamoto *et al.* (2001) identified QTLs associated with this trait in linkage group 6. Eduardo et al. (2011) found stable QTLs over two years of evaluation in the same linkage group.

The QTL identified on linkage group 2, which explains about 5.3% of the observed phenotypic variance, corresponds to the same region where QTLs for fruit weight have been mapped in sweet cherry (Zhang et al., 2010; Cabrera, 2011).

Mapping of QTLs in linkage group 3 is also consistent with results from Yamamoto et al. (2001). In the present study, a QTL in the same linkage group was identified, between markers SNP_IGA_355590 (48.58 cM) and

SNP_IGA_356179 (48.92 cM). According to Illa et al. (2010), putative candidate genes involved in fruit growth and maturity exist in this region in peach.

Referring to the traits involved in fruit size (height, width and depth), a QTL identified in linkage group 4, in the interval between the markers SNP_IGA_402569 (30.21 cM) and SNP_IGA_404899 (33.13 cM) for the trait fruit width, corresponds to a similar QTL identified for the same traits in the study conducted by Quilot et al. (2004).

In the case of marker SNP_IGA_713270, the increasing allele (allele associated with increased values of the traits) was recessive, in contrast with marker SNP_IGA_574551 for which the increasing allele was dominant.

In terms of the possibility to genetically dissect and manipulate FW in peach, the obtained results are encouraging. In order to confirm the stability of the QTLs, phenotypic evaluations will be repeated and the results will be compared to those from previous studies. This Meta-QTL analysis will be helpful in identifying QTLs/markers that are stable across different crosses and may be used in MAS programmes.

Chapter 6

Preliminary study of association mapping in traits associated with peach fruits

6.1 Introduction

Since the cost of sequencing and genotyping technologies is decreasing, association mapping or linkage disequilibrium mapping is emerging as a powerful tool for identifying QTLs (Zhao et al., 2007).

Association mapping allows exploring the natural diversity based on existing cultivars without the need to develop new mapping populations. Furthermore, it allows studying recombination events of the ancestors considering the main alleles present in the population under evaluation in the identification of significant associations between marker and phenotype (Zhu et al., 2008; Simko and Hu, 2008; Pasam et al., 2012). Other advantages are the high probability of obtaining high-resolution maps and the possibility of using previous phenotypic data (Abdurakhmonov and Abdukarimov, 2008). However, the biggest drawback is related to possible false positive associations that occur due to the genetic structure of the populations in some germplasm sets (Stich et al., 2008; Ersoz et al., 2009).

Population structure, i.e. the presence of subgroups, occurs naturally. When a particular trait of interest is more prevalent in one subpopulation than in others, the trait will be associated with any marker allele that is in high frequency in this subpopulation (Ewens and Spielman, 1995; Lander and Schork, 1994; Pritchard and Rosenberg, 1999; Simko and Hu, 2008). For this reason, it is crucial to understand the genetic structure of the population in order to avoid false positive associations between markers and genes affecting the trait. To avoid this problem, population structure can be assessed with marker information from genetic markers and the results are incorporated in the analysis of association as a correction factor (Pritchard et al., 2000a; Sakiroglu, 2009; Myles et al., 2009; Lopes, 2011).

Several statistical approaches have been developed to consider the population structure and levels of relatedness between individuals, thereby avoiding false positive associations (Zhu et al., 2008; Souza, 2011). Among these methods, cluster analysis based on Bayesian statistics estimated using the software *Structure* (Pritchard et al., 2000a), in which these results are incorporated in further statistical analysis through the matrix Q, is widely used (Pritchard and Rosenberg, 1999; Pritchard et al., 2000a; Falush et al., 2003; Zhu et al., 2008). Mixed Linear Model (MLM) considers the estimates of the population structure and multiple levels of relatedness and incorporates these factors in the analysis through the matrix of population effects (Q) and the kinship matrix (K) (Yu et al., 2006; Sun et al., 2010; Souza, 2011).

Association mapping has been used as an alternative approach to QTL analysis on biparental populations with the objective of detecting associations between markers and phenotypic traits of agronomic importance. This approach has shown good results in various crops. In soybean, Li et al. (2011) identified a total of 21 SSR markers associated with high oil content, high protein content, drought tolerance, soybean cyst nematode resistance and soybean mosaic virus resistance. In rice, a total of 25 marker-trait associations were identified, seventeen being in regions where QTLs associated with the given trait had previously been reported (Agrama et al., 2007).

In barley, Pasam et al. (2012) identified 107 QTLs associated with heading date, plant height, thousand grain weight, starch content and crude protein content. Reif et al. (2011) mapped QTLs for the traits sedimentation volume and test weight in wheat. In alfalfa, one SSR and one SNP associated with biomass yield were reported in the study conducted by Sakiroglu (2009).

In peach, Aranzana et al. (2010) evaluated the genetic structure of the population and linkage disequilibrium in 224 cultivars. The results showed high linkage disequilibrium suggesting that association mapping can be used for dissection of traits of agronomic importance in this species. Evaluating 104 peach landrace accessions from six Chinese geographical regions, Cao et al. (2012) identified significant associations for flesh colour around the stone, red pigment in the flesh, flesh texture, flesh adhesion, flesh firmness, fruit weight, chilling requirement, flowering time, ripening time, and fruit development period. Markers associated with all traits were identified, many of which were located in regions where QTLs had previously been identified by Yamamoto et al. (2001); Wang et al. (2010); Peace et al. (2005); Illa et al. (2010); Abbott et al. (1997). However, the limited number of markers considered in the study by Cao et al. (2012) (53 SSRs) strongly limits the possibility of dissecting the traits of interest.

To overcome these limitations, in the present work, we chose a high-throughput genotyping platform to increase marker density and ensure better coverage of the genome for QTL detection. In order to detect associations between SNP markers and traits related to fruit weight and size, a preliminary study of association mapping was conducted, using the general linear model (GLM), on a set of 70 peach accessions from the 'MAS.PES' programme, selected for their genetic diversity and contrasting FW values. The 9K SNP array v1 (Verde et al., 2012) was used to assess the population structure and to conduct association mapping in order to detect associations between SNP markers and fruit weight and size.

Table 6.1: Peach accessions selected for the diversity panel in the association analysis.

Accessions		
194RXXIII43	Forli 1	Romagna Red
391 C12 XXXIV 86	Harrow Blood	Rosa Dardi
A15	Honey Blaze	Royal Estate
A219	IF 691	Royal Jim
Aliblanca	IFF 331	S5898:128
Alicecol	Ionia	S6699
Alipersié	Iskra	Siberian C
Alma	June Princess	Stark Red Gold
Angelo Marzocchella	Kaweah	Tardigold
Antony	Kevina	Vista Rich
Autumn Grand	KV930455	Zao Xia Lu
Azurite	Maycrest	Zee Diamond
Big Top	Mayfire	Zee Lady
Blushing Star	Merril Gem Free	Soleada
Bolero	Nadia	Rasciadente Bianca
Bordó	Nectagrand	Rasciadente Gialla
Capucci 18	NJ 307	Bounty
Chimarrita	NJC113	Pieri 81
Dolores	NJ Weeping	Hardired
Early Top	Tardiva Spadoni	Helena Cling
Early Zee	Rich Lady	OroA
Elbertita	Rising Star	Contender
Elegant Lady	Romagna 3000	
Fei Cheng Bai Li	Romagna Bright	

6.2 Materials and methods

6.2.1 Plant material

The diversity panel consisted of 70 accessions (Table 6.1) obtained from the germplasm collection of the 'MAS.PES' program, located in Imola (BO). The accessions were selected based on traits of interest and also based on the study of genetic diversity conducted in chapter 4, prioritizing contrasting individuals relating to fruit weight and size and also selecting individuals from different groups. The accessions were planted with a spacing of 1 m within and 4 m between rows with a total of three trees for each accession.

6.2.2 Total DNA extraction, genotyping and analysis of phenotypic traits

In order for fruits to express their phenotypic potential, fruit thinning was carried out for the 70 accessions. DNA extraction, genotyping and phenotypic analysis of the traits fruit weight, height, width and depth were performed according to the methodology described in chapter 5. Data normality was evaluated with the Shapiro–Wilk test in the software PAST (Hammer et al., 2001).

6.2.3 Analysis of diversity and population structure

A "neighbour-joining" phylogenetic analysis was conducted using 400 SNPs in the software *Darwin*, estimating statistical support for phylogenetic groupings with the "bootstrap" method (1000 replications).

The analysis of population structure was conducted with the software *Structure* (Pritchard et al., 2000a), based on Bayesian statistics using the same 400 SNP markers as in the phyologenetic analysis. The "admixture model" was adopted as ancestry model and correlated allele frequencies were used to analyse the data set. No preliminary information on the number of cluster was considered. The proportion of ancestry of each accession was tested considering a K number of 1 to 10, with 5 iterations for each K value. The settings for the length of the burn-in period and MCMC (Markov Chain Monte Carlo) were 20,000 and 200,000, respectively. To determine the K number, the model Delta K, as established by Evanno et al. (2005) was adopted by means of Structure harvest (Earl and vonHoldt, 2012).

6.2.4 Linkage disequilibrium

The squared allele-frequency correlation r^2, was calculated for all possible combinations of alleles to estimate the extent of linkage disequilibrium (LD) in the peach accessions, using the software *TASSEL* 2.01 (Bradbury et al., 2007). The weighted average of r^2 values was obtained by further weighting

the corresponding allele frequencies.

The significance of pairwise LD (p-value) among all possible pairs was also evaluated by *TASSEL* with the rapid permutation test.

The 95th percentile of this approximate normal distribution was assumed as the threshold of the r^2 value to declare the presence of LD among molecular markers (Breseghello and Sorrells, 2006).

6.2.5 Analysis of association

The associations between markers and phenotypic data were calculated with the software *TASSEL* applying the general linear model (GLM). The matrix Q of the effects of the population was incorporated into this model. This matrix was incorporated into the analysis of association as a correction factor in order to avoid false positive associations, as recommended by Pritchard et al. (2000b); Yu et al. (2006); Myles et al. (2009); Lopes (2011); Souza (2011); Sakiroglu et al. (2012).

To determine the positive associations, the methodology based on the study of Pasam et al. (2012), who adopted an arbitrary threshold P-value of 0.03, was used for all traits. In the present study, a threshold P-value of 0.001 was chosen.

6.3 Results

6.3.1 Phenotypic analysis

According to the results obtained by the Shapiro–Wilk normality tests, all traits follow a normal distribution (Table 6.2). Histograms with the frequency distributions of the traits fruit weight, height, width and depth are shown in Fig. 6.1.

Table 6.2: Shapiro–Wilk test statistic values for the quantitative traits

Fruit Trait	Shapiro–Wilk (p_{normal})
Weight	0.4025
Height	0.4718
Width	0.3071
Depth	0.1199

Shapiro–Wilk $p_{normal} < 0.05$ normal distribution can be rejected.

The average values of the traits fruit weight, height width and depth are shown in Table 6.3. The accession 'NJ Weeping' exhibited the smallest fruit weight (42.4 g), while the accession with the highest weight was 'Chimarrita' (288.6 g). The accession 'Oro A' had the lowest fruit height with an average of 37.9 mm. 'Nadia' had the highest fruit height (75.73 mm) and also the highest fruit width. Concerning fruit depth, the highest value was observed in 'Bolero' and the smallest fruit width and depth were found in 'NJ Weeping'.

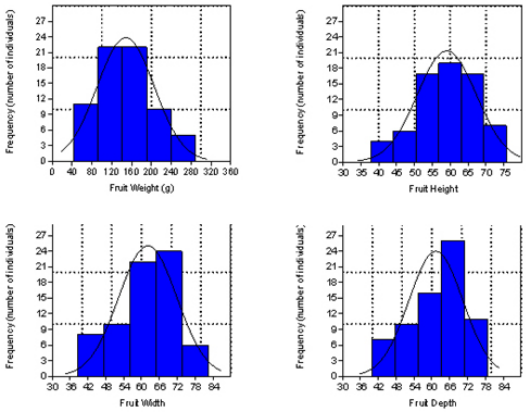

Figure 6.1: Frequency distribution of fruit weight (FW), height (FH), width (FWD) and depth (FD) in 70 accessions of peach (year 2011).

The correlation coefficients were highly significant and positive for all traits (Table 6.4). The highest correlation coefficient was observed between traits fruit width and fruit depth (0.974). These high correlation coefficients

Table 6.3: Fruit weight (FW), height (FH), width (FWD) and depth (FD) of the 70 peach accessions evaluated in year 2011.

Accessions	FW	Std	FH	Std	FWD	Std	FD	Std
194RXXIII43	184.29	9.53	70.22	2.21	73.72	2.55	67.58	2.22
391 C12 XXXIV 86	168.11	10.94	59.28	2.26	65.81	1.14	67.25	1.85
A15	84.58	4.93	47.00	4.97	48.44	5.43	46.33	3.46
A219	92.25	20.10	52.31	4.41	52.03	5.06	52.61	4.86
Aliblanca	202.78	18.89	69.42	2.71	73.29	2.63	71.78	2.14
Alicecol	166.66	9.68	63.69	2.63	66.22	3.65	64.39	2.17
Alipersié	113.72	11.88	56.83	2.01	58.53	2.01	58.03	2.39
Alma	177.74	27.43	64.03	4.50	67.31	3.99	66.56	3.62
Ang, Marzocchella	126.13	10.98	53.67	2.22	60.19	2.05	56.64	2.06
Antony	174.29	10.66	65.36	2.05	66.53	1.57	66.94	2.78
Autumn Grand	78.82	11.86	50.31	2.72	51.35	2.32	51.96	3.59
Azurite	173.15	12.91	63.92	2.34	69.25	2.09	69.58	2.17
Big Top	177.24	9.86	68.58	2.11	68.75	2.05	65.39	1.88
Blushing Star	101.38	5.39	54.61	1.33	56.56	1.64	58.86	1.19
Bolero	241.69	25.44	74.84	2.93	77.78	3.28	78.80	3.11
Bordó	138.85	5.39	56.56	2.66	61.83	1.96	58.69	1.61
Capucci 18	231.35	28.68	64.89	3.82	72.14	3.68	72.11	4.66
Chimarrita	288.62	25.56	67.81	2.30	79.11	4.76	76.58	4.61
Dolores	106.94	11.50	50.83	2.60	55.03	1.48	54.67	2.60
Early Top	142.96	25.66	55.94	1.37	59.83	2.98	56.58	3.06
Early Zee	175.10	13.75	63.08	2.68	63.86	1.87	64.67	1.26
Elbertita	168.67	30.19	72.63	3.59	63.99	4.84	67.68	3.97
Elegant Lady	223.61	20.04	68.53	2.11	71.92	2.90	70.94	2.23
Fei Cheng Bai Li	276.94	34.61	71.36	3.62	77.94	4.45	72.33	3.25
Forli 1	99.58	5.95	50.92	2.09	56.00	1.87	54.56	2.21
Harrow Blood	50.55	5.42	40.32	1.59	42.22	1.91	41.61	1.72
Honey Blaze	154.44	16.72	61.33	3.30	64.14	3.22	63.67	2.99
IF 691	195.47	17.85	65.83	2.90	71.64	3.09	71.36	3.68
IFF 331	124.11	12.42	57.89	2.42	59.06	2.36	58.75	2.50
Ionia	151.83	12.14	61.00	3.09	64.11	3.90	61.78	3.15
Iskra	107.92	5.97	56.14	1.13	57.64	2.17	56.58	0.84
June Princess	110.65	16.31	58.56	3.84	55.33	3.93	55.78	3.56
Kaweah	137.01	21.63	58.02	2.27	64.44	3.93	66.60	3.17
Kevina	128.15	10.06	56.36	2.26	64.82	1.85	61.44	2.08
KV930455	111.73	10.51	54.17	2.75	57.50	2.50	57.58	2.17
Maycrest	169.44	12.04	59.11	2.51	70.64	4.79	63.69	2.12
Mayfire	109.16	13.93	53.00	1.51	55.56	2.89	55.00	3.60
Merril Gem Free	125.19	9.31	59.58	1.72	60.78	1.90	59.28	2.28
Nadia	269.18	20.14	75.73	4.26	82.39	3.27	77.91	2.91
Nectagrand	158.76	11.35	58.00	1.94	65.75	2.21	63.25	2.67
NJ 307	185.10	15.19	63.44	1.96	67.36	1.93	65.47	2.32
NJC113	96.33	9.62	53.78	3.21	52.17	3.05	53.42	2.59
NJ Weeping	42.93	5.97	40.19	2.02	38.56	2.01	39.75	1.96
Tardiva Spadoni	65.69	7.46	47.31	1.68	44.75	1.94	47.78	2.29
Rich Lady	161.01	10.52	57.50	2.00	64.33	2.31	64.72	1.94
Rising Star	104.36	8.06	56.72	1.05	56.92	2.65	58.47	1.42
Romagna 3000	158.38	13.11	62.82	2.07	65.00	2.36	66.69	2.92
Romagna Bright	221.44	16.94	68.14	2.71	72.28	2.59	72.75	3.17
Romagna Red	160.01	8.41	64.28	2.09	65.36	2.61	65.64	1.83
Rosa Dardi	199.02	29.30	61.03	6.35	71.28	4.92	68.11	2.62
Royal Estate	242.70	22.37	70.84	2.38	79.09	2.52	73.28	3.22
Royal Jim	142.14	12.68	60.55	2.61	67.74	2.26	63.55	2.45
S5898:128	61.05	6.48	47.94	2.57	44.61	2.41	45.47	2.26
S6699	93.77	14.70	53.47	2.73	50.50	2.46	50.75	2.31
Siberian C	75.39	7.50	47.64	2.66	45.75	1.81	48.22	1.89
Stark Red Gold	178.74	17.44	63.97	2.40	65.31	2.75	64.42	2.22
Tardigold	101.32	6.65	55.75	1.68	56.71	1.65	56.41	1.24
Vista Rich	188.62	19.73	67.81	1.94	71.25	3.63	71.25	3.03
Zao Xia Lu	89.86	9.67	55.11	2.11	53.36	3.43	54.58	3.53
Zee Diamond	133.62	12.43	62.36	2.72	61.22	2.90	59.72	2.11
Zee Lady	170.83	18.58	61.36	3.04	63.61	3.20	66.31	2.29
Soleada	219.74	15.34	65.06	2.40	71.72	1.90	72.00	1.93
Rasc, Bianca	58.26	5.81	45.31	1.12	46.86	1.64	46.15	2.17
Rasc, Gialla	52.64	5.08	42.89	1.22	45.97	2.17	46.52	2.53
Bounty	179.57	2.15	60.06	2.78	60.01	1.15	63.30	2.25
Pieri 81	221.24	13.34	66.06	2.16	70.83	2.85	69.81	2.17
Hardired	131.19	6.49	55.06	2.76	58.75	1.33	57.56	0.84
Helena Cling	190.52	14.49	61.97	2.16	67.19	1.66	66.61	2.59
OroA	49.97	5.18	37.86	2.86	43.89	3.16	40.83	3.49
Contender	221.26	14.18	65.47	2.50	71.33	1.95	70.78	3.35

Table 6.4: Correlation values for fruit weight and size.

	Weight	Height	Width	Depth
Weight	1			
Height	0.909**	1		
Width	0.961**	0.932**	1	
Depth	0.950**	0.949**	0.974**	1

**: Significant at 1% probability in t-test

were also observed in chapter 5 in which the same traits in the F2 population from crossing 'NJ Weeping' × 'Bounty' were evaluated.

6.3.2 Genotyping

The Peach 9K Infinium SNP Chip containing a total of 8144 SNPs was used to genotype the 70 accessions. After SNP filtering, which excluded markers with 10% missing data and those with Genetrain values smaller than 0.6, a total of 5102 markers segregated for these accessions. Of these, 400 SNPs (50 markers per chromosome) were used to construct the phylogenetic tree and for studies of population structure. In total, 4702 markers were used in the analysis of association between markers and phenotypic data.

6.3.3 Analysis of genetic diversity and population structure

A preliminary analysis of genetic diversity using 400 SNP markers was conducted based on genetic dissimilarity in the *Darwin* software (Fig. 6.2). The formation of two groups with high bootstrap values of up to 100% and 94% cophenetic correlation was observed. The cophenetic correlation (a statistical method used to increase the reliability of results obtained in dendrograms, (Kopp et al., 2007) was 94% revealing a high fit between the graphical representation of the genetic distance and the matrix of calculated genetic distance.

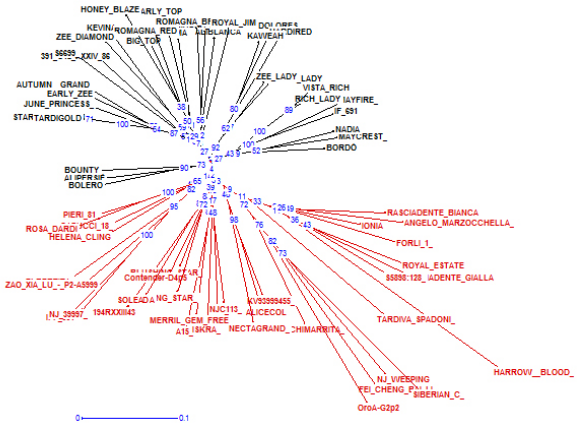

Figure 6.2: Phylogenetic tree of 70 peach accessions obtained by 400 SNP markers; the tree was constructed using neighbour joining and simple matching coefficient.

In order to detect the effects of the population structure and to correct these effects in association analysis, the Bayesian method in the software *Structure* was used. The determination of the exact number of clusters or subpopulations (K) with the method DeltaK as developed by Evanno et al. (2005) showed two clusters (Fig. 6.3).

The distribution of accessions in the two clusters as well as the shared ancestry of each individual can be seen in Fig. 6.4.

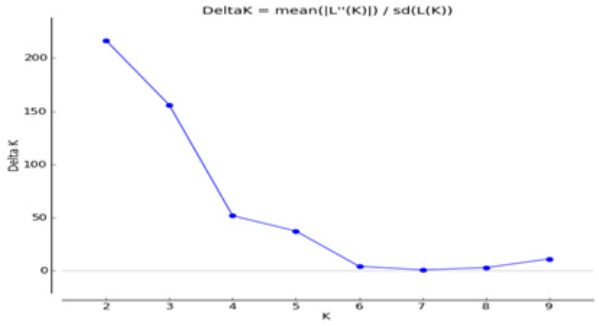

Figure 6.3: Number of clusters (K) in 70 peach accessions obtained by the method of Evanno et al. (2005)

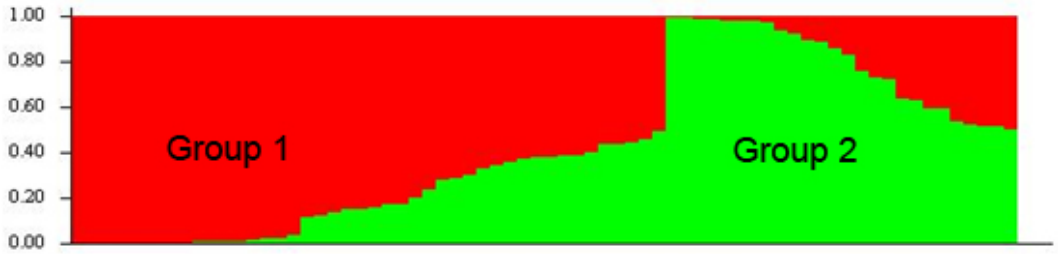

Figure 6.4: Graph of the ancestry of 70 peach accessions from the germplasm collection; values were obtained by the software *Structure*.

A large part of the accessions exhibit a fraction of its ancestry in the other cluster and most of the ancestry was observed in the cluster in which it had been classified. The values of ancestry membership for each accession are shown in Table 6.5.

Cluster 1 consisted of 44 accessions, which mostly showed ancestry values above 0.65 (Table 3). Some accessions such as '391C12XXXIV86' and 'A15' showed high admixture values between the clusters 1 and 2, 0.5364 and 0.4636 and 0.5012 and 0.4988, respectively. In such a situation, some authors suggest a minimum value of ancestry to classify an individual in a specific cluster, and when a specific accession does not reach this minimum value, it is classified as hybrid (Lopes, 2011). The same was observed in cluster 2 with the accessions 'Alice Col', 'Harrow Blood', 'Nectagrand' and 'Capucci 18'.

6.3.4 Linkage disequilibrium

The analysis with the software *TASSEL*, using 4,702 SNP markers, lead to a total of 233,826 SNP pairs considering all accessions. For all chromosomes, the maximum, minimum and medium values of r^2, D' and p values were calculated (Table 6.6). The chromosomes 2, 5 and 7 showed r^2 values higher than 0.6. The largest maximum value of r^2 (0.81) was observed on chromosome 7, while the lowest observed maximum values of r^2 (0.49) were found on chromosomes 6 and 8. Although chromosome 7 showed the highest r^2 value, the highest average value was observed in chromosome 5 (0.05). The lowest p

Table 6.5: Mean ancestry values for the two genetic clusters (software *Structure*).

Accessions	K1	K2	Cl.	Accessions	K1	K2	Cl.
194RXXIII43	0.5598	0.4402	1	ROMAGNA BRIGHT	0.996	0.004	1
391 C12 XXXIV 86	0.5364	0.4636	1	ROMAGNA RED	0.7976	0.2024	1
A15	0.5012	0.4988	1	ROYAL JIM	0.6986	0.3014	1
ALIBLANCA	0.979	0.021	1	SOLEADA	0.634	0.366	1
ALIPERSIÉ	0.8808	0.1192	1	STARK RED GOLD	0.996	0.004	1
ALMA	0.7594	0.2406	1	TARDIGOLD	0.9958	0.0042	1
ANTONY	0.996	0.004	1	VISTA RICH	0.9924	0.0076	1
AUTUMN GRAND	0.9732	0.0268	1	ZEE DIAMOND	0.9726	0.0274	1
AZURITE	0.9936	0.0064	1	ZEE LADY	0.9896	0.0104	1
BIG TOP	0.85	0.15	1	S6699	0.476	0.524	2
BLUSHING STAR	0.6534	0.3466	1	A219	0.1648	0.8352	2
BOLERO	0.8214	0.1786	1	ALICECOL	0.4972	0.5028	2
BORDÓ	0.8432	0.1568	1	ANGELO MARZOCCHELLA	0.008	0.992	2
BOUNTY	0.7142	0.2858	1	CAPUCCI 18	0.4618	0.5382	2
Contender-D4p5	0.5926	0.4074	1	CHIMARRITA	0.239	0.761	2
DOLORES	0.997	0.003	1	FEI CHENG BAI LI	0.0766	0.9234	2
EARLY TOP	0.8266	0.1734	1	FORLI 1	0.0154	0.9846	2
EARLY ZEE	0.994	0.006	1	HARROW BLOOD	0.4838	0.5162	2
ELBERTITA	0.5502	0.4498	1	IFF 331	0.4046	0.5954	2
ELEGANT LADY	0.9584	0.0416	1	IONIA	0.0148	0.9852	2
HARDIRED	0.991	0.009	1	ISKRA	0.1084	0.8916	2
HELENA CLING	0.6116	0.3884	1	NECTAGRAND	0.4842	0.5158	2
HONEY BLAZE	0.8354	0.1646	1	NJ 307	0.365	0.635	2
IF 691	0.7104	0.2896	1	NJ WEEPING	0.0616	0.9384	2
JUNE PRINCESS	0.996	0.004	1	NJC113	0.2726	0.7274	2
KAWEAH	0.989	0.011	1	OroA	0.004	0.996	2
KEVINA	0.6698	0.3302	1	PIERI 81	0.3604	0.6396	2
KV930455	0.5632	0.4368	1	RASCIADENTE BIANCA	0.0062	0.9938	2
MAYCREST	0.626	0.374	1	RASCIADENTE GIALLA	0.0234	0.9766	2
MAYFIRE	0.8642	0.1358	1	ROSA DARDI	0.269	0.731	2
MERRIL G, FREE	0.6094	0.3906	1	ROYAL ESTATE	0.0092	0.9908	2
NADIA	0.6166	0.3834	1	S5898:128	0.1054	0.8946	2
RICH LADY	0.993	0.007	1	SIBERIAN C	0.0206	0.9794	2
RISING STAR	0.615	0.385	1	TARDIVA SPADONI	0.1364	0.8636	2
ROMAGNA 3000	0.872	0.128	1	ZAO XIA LU	0.3998	0.6002	2

value and thus the most significant r^2 values were observed on chromosome 5.

In Fig. 6.5, the graphs of r^2 values are shown as a function of distance. The linkage disequilibrium clearly decays as a function of distance in all chromosomes.

The square root-transformed distribution of r^2 values of SNPs mapping on different chromosomes allowed setting an appropriate threshold at a value of 0.20 beyond which LD values were considered significant. The value of 0.20 calculated for this LD threshold excluded most of the r^2 values of SNPs mapping to the same chromosomal region. Significant marker pairs were observed in all chromosomes.

6.3.5 Association analysis

The analysis of association between traits related to fruit weight and size (height, width and depth) was conducted with the general linear model (GLM).

The results of the preliminary study of association with GLM are shown in Table 6.7 and Fig. 6.5. QTLs were identified on chromosomes 1, 2, 4, 5 and 7 for the trait fruit weight. The strongest association was observed on chromosome 7, where marker SNP_IGA_779594 who was responsible for 25% of the phenotypic variance observed among the accessions. Two markers were found on chromosome 5 together accounting for about 41% of the observed phenotypic variance (SNP_IGA_553456 and SNP_IGA_602901). It is noteworthy that association studies by GLM and MLM were performed on one marker at a time and thus can result in a sum of r^2 above 100% (Souza, 2011).

Referring to the trait fruit height, four positive and significant associations were found for the SNP markers SNP_IGA_374610, SNP_IGA_388388, SNP_IGA_443952, SNP_IGA_388527 located on chromosome 4, accounting for about 19%, 15%, 17% and 17% of the observed phenotypic variance.

Table 6.6: Maximum, medium and minimum of linkage disequilibrium in the eight peach chromosomes.

Chromosome	Descriptive Statistic	r^2	D'	pDiseq
1	MIN	0	0	4.46E-06
1	MAX	0.54066	1	1
1	MEAN	0.012287	0.171641	0.47417
2	MIN	0	0	7.75E-10
2	MAX	0.66502	1	1
2	MEAN	0.022814	0.247464	0.436261
3	MIN	0	0	1.24E-09
3	MAX	0.58387	1	1
3	MEAN	0.014646	0.180951	0.372282
4	MIN	0	0	4.22E-10
4	MAX	0.57891	1	1
4	MEAN	0.017746	0.191869	0.525986
5	MIN	0	0	3.73E-14
5	MAX	0.77258	1	1
5	MEAN	0.058297	0.287878	0.178373
6	MIN	0	0	1.41E-06
6	MAX	0.49265	1	1
6	MEAN	0.016689	0.203871	0.505985
7	MIN	0	0	2.77E-11
7	MAX	0.81823	1	1
7	MEAN	0.017762	0.182567	0.401312
8	MIN	0	0	5.73E-07
8	MAX	0.49265	1	1
8	MEAN	0.014024	0.17801	0.632478

r^2 = square of the correlation coefficient between the two loci
D'= standardized D (Lewotin, 1964)

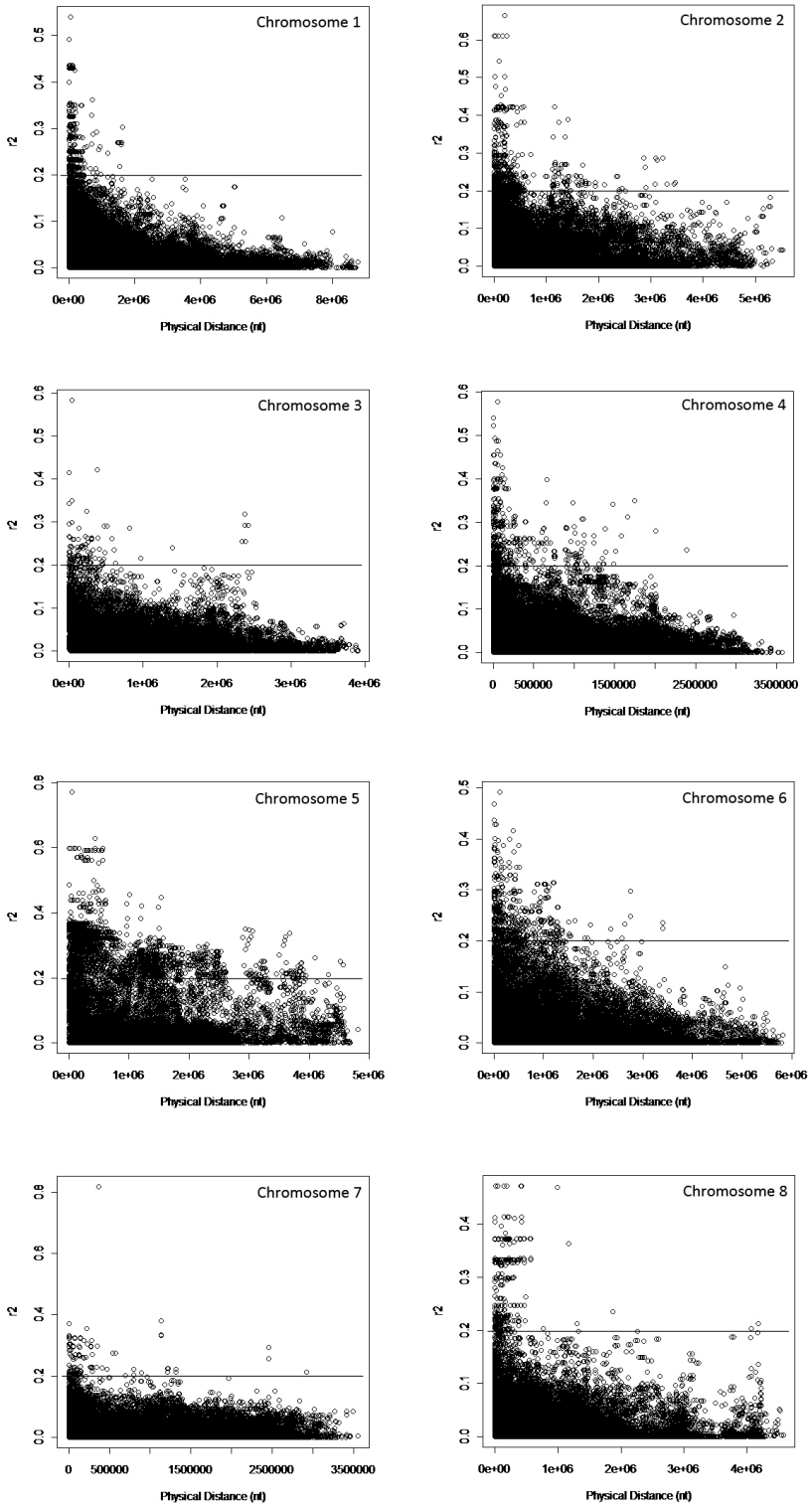

Figure 6.5: Linkage disequilibrium decay as a function of distance: the r^2 threshold line is at 0.20.

Two associations on chromosome 7 were observed for the SNP markers SNP_IGA_767644 and SNP_IGA_779594, the latter being also associated with the trait fruit weight.

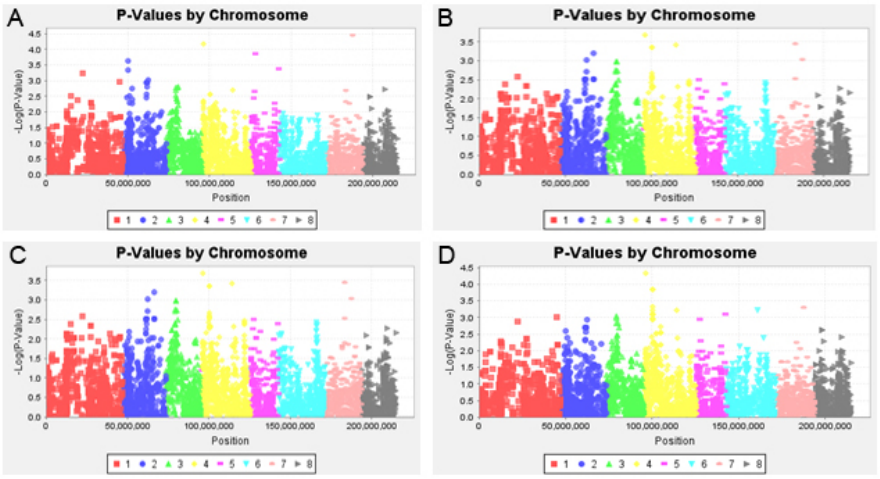

Figure 6.6: Genome-wide associations for traits fruit weight (A), height (B), width (C) and depth (D). The vertical axis shows -log10(P) values of the P-value of the marker-trait association. The peaks above minimum threshold of 3 (P-value = 0.001) can be considered as significantly associated.

Fruit width and depth showed some markers in common, consistent with results presented in chapter 5 in the study on the F2 population 'NJ Weeping' × 'Bounty'. Markers SNP_IGA_388457 and SNP_IGA_388802 explained 18% of the observed phenotypic variance for this trait. Markers SNP_IGA_669440, SNP_IGA_388125, SNP_IGA_388168, SNP_IGA_388234, SNP_IGA_388258, SNP_IGA_310361 and SNP_IGA_131496 were associated only with the trait fruit width and each one of them explained 17% of the observed phenotypic variance, respectively.

The markers associated with fruit depth not associated with fruit width were SNP_IGA_386970 and SNP_IGA_263828, located on chromosomes 4 and 2, respectively.

Table 6.7: Genome-wide associations for traits fruit weight (FW), height (FH), width (FWD) and depth (FD) by the GLM model.

Trait	Marker	Locus	Locus_pos	marker_F	marker_p	r^2
FW	SNP_IGA_779594	7	16243710	12.02624	3.52E-05	0.25
	SNP_IGA_374610	4	994204	11.20867	6.57E-05	0.24
	SNP_IGA_553456	5	2477325	10.18742	1.39E-04	0.22
	SNP_IGA_174498	2	3279542	15.10599	2.38E-04	0.17
	SNP_IGA_602901	5	16707379	8.754179	4.24E-04	0.19
	SNP_IGA_174475	2	3279135	8.695859	4.62E-04	0.20
	SNP_IGA_78954	1	22331148	8.37797	5.72E-04	0.19
	SNP_IGA_263828	2	15656993	7.748311	9.49E-04	0.17
FH	SNP_IGA_374610	4	994204	9.666013	2.11E-04	0.19
	SNP_IGA_767644	7	12014754	9.006549	3.53E-04	0.18
	SNP_IGA_443952	4	18818056	14.03736	3.83E-04	0.15
	SNP_IGA_388388	4	4720743	8.689349	4.47E-04	0.17
	SNP_IGA_388527	4	4741427	8.689349	4.47E-04	0.17
	SNP_IGA_277359	2	19444908	8.2559	6.31E-04	0.17
	SNP_IGA_779594	7	16243710	7.779602	9.25E-04	0.16
	SNP_IGA_263828	2	15656993	7.734538	9.60E-04	0.16
FWD	SNP_IGA_374610	4	994204	11.65101	4.74E-05	0.23
	SNP_IGA_388388	4	4720743	10.16596	1.42E-04	0.21
	SNP_IGA_388527	4	4741427	10.16596	1.42E-04	0.21
	SNP_IGA_388457	4	4738206	8.639093	4.65E-04	0.18
	SNP_IGA_779594	7	16243710	8.558379	4.96E-04	0.18
	SNP_IGA_669440	6	18315130	8.382238	5.92E-04	0.17
	SNP_IGA_443952	4	18818056	12.98574	6.09E-04	0.15
	SNP_IGA_388125	4	4691366	8.282894	6.17E-04	0.18
	SNP_IGA_388802	4	4769975	8.016641	7.64E-04	0.17
	SNP_IGA_602901	5	16707379	7.985735	7.84E-04	0.17
	SNP_IGA_388168	4	4697983	7.958937	8.01E-04	0.17
	SNP_IGA_388234	4	4701823	7.958937	8.01E-04	0.17
	SNP_IGA_388258	4	4704021	7.958937	8.01E-04	0.17
	SNP_IGA_310361	3	5894946	7.816663	9.08E-04	0.17
	SNP_IGA_131496	1	44780943	7.722088	9.80E-04	0.17
FD	SNP_IGA_374610	4	994204	14.16818	7.82E-06	0.25
	SNP_IGA_779594	7	16243710	8.629403	4.69E-04	0.17
	SNP_IGA_388388	4	4720743	8.425062	5.51E-04	0.17
	SNP_IGA_388527	4	4741427	8.425062	5.51E-04	0.17
	SNP_IGA_386970	4	4410111	8.375689	5.73E-04	0.17
	SNP_IGA_263828	2	15656993	8.235752	6.41E-04	0.17
	SNP_IGA_443952	4	18818056	12.22161	8.57E-04	0.13
	SNP_IGA_602901	5	16707379	7.824982	8.92E-04	0.16
	SNP_IGA_404899	4	9432878	7.357696	0.001304	0.15
	SNP_IGA_404914	4	9433562	7.357696	0.001304	0.15
	SNP_IGA_405055	4	9467499	7.357696	0.001304	0.15

6.4 Discussion

In the present study, the 9K SNP genotyping array v1 (Verde *et al.*, 2012) developed by the International Peach SNP Consortium (IPSC) was used, which allowed the identification of 5,102 polymorphic markers in the 70 accessions studied, proving the efficiency of this technology and the variability of the selected accessions.

High and significant correlation values were observed for all traits, similar to those observed in chapter 5, where the same traits in the F2 population of 'NJ Weeping' and 'Bounty' had been studied and high correlation coefficients had been observed.

In the *Structure* software, it was crucial to determine the exact number of clusters (K) in the analysis in order to obtain the actual population structure. Sometimes, the distribution of log Pr(X/K) increases continuously with increasing K values, as it occurred in the analysis by Cao et al. (2012) and in the present study. Evanno et al. (2005) developed a method based on DeltaK, where the actual number of clusters (K) occurs at the maximum value (Fig. 4.4). In this study, the number K or the number of observed clusters is 2 (Fig. 4.5). In the analysis, it was considered that each accession can have a part of its genome derived from another cluster, which is different from where it belongs. Therefore, the option admixture model was used. The presence of shared ancestry between clusters was confirmed in many accessions (Table 6.5) such as 'A15' which showed a value of ancestry fragmenting almost halfway between the two clusters. Among the formed clusters, most of the accessions forming the second cluster are characterized by fruits with a low average fruit weight.

Regarding the comparison between the phylogenetic tree and information about the population structure, both methods identified the formation of two groups. However, as also noted in chapter 4, some accessions were classified into different groups. This is because the analysis conducted in *Darwin* is

based on genetic distance, while the analysis conducted in *Structure* is based on allele frequency. However, in general, both methods were efficient to determine the population structure.

The analysis of association between markers and fruit weight resulted in eight markers associated with this trait. The marker SNP_IGA_263828 (linkage group 2) was also associated with this trait in the study conducted in the F2 population of 'NJ Weeping' and 'Bounty' in chapter 5. Notably, this marker is found in collinear position to a major QTL for fruit weight in sweet cherry (Zhang et al., 2010; Cabrera, 2011).

The marker SNP_IGA_78954, located on chromosome 1, is in a neighbouring region to gene ppa022891m, predicted to encode a ring finger protein. Interestingly, some ring finger proteins were recently proposed to play a role in apple fruit development (Li et al., 2011).

The marker SNP_IGA_553456 is in a region near transcript ppa017022m related to mitogen-activated protein kinases (MAPKs), a subfamily of the protein kinases associated with diverse biological processes, including cellular division (Krysan et al., 2002).

Similarly, the markers SNP_IGA_388388 and SNP_IGA_388527, significantly associated with FWD, are adjacent to genes ppa016617m and ppa020949m. This region has 14 genes associated with wall-associated receptor kinases (WAKs). According to Wagner and Kohorn (2001), WAKs may have significance in the control of cell expansion. The involvement of WAKs with cellular expansion is also reported by Kohorn and Kohorn (2012). The amount of accessions and the fact that the phenotypic analysis was performed in only a year of assessment are to be considered as limiting factors to confirm the associations found between markers and fruit weight and size. Small population sizes may lead to bias when estimating the population structure and the degree of relatedness between individuals in comparison to populations with a higher number of individuals (Wang et al., 2012). These potential

drawbacks in the estimation of population structure can lead to false positive associations. Thus, it is important to emphasize that the present study is of preliminary character. More accessions and phenotypic evaluations will be conducted in order to detect new QTLs. Further studies will be carried out, considering the mixed linear model as well.

Chapter 7

Final remarks

Fruit weight is a trait of high economic importance and has thus been studied in several breeding programs. The results of the present study provide important information for a better understanding of the mechanisms that control the genetic bases of this trait.

In chapter 4, the objective was to assess the genetic diversity using AFLP and SSR markers. Based on this information and other criteria such as contrasting fruit weight values, a panel of accessions to be used in association mapping for frui weight was chosen. In the present study, SSR markers were more informative, although a higher number of markers is needed for other stages of the study, in case of association mapping. Overall, a high level of homozygosity in the accessions under evaluation was observed, similar to results by Aranzana et al. (2010) who pointed out that this is due high level of co-ancestry.

In chapter 5, genotyping of the biparental cross 'NJ Weeping' × 'Bounty' with the 9K SNP array v1 (Verde et al., 2012) allowed the construction of a high density genetic map comprising 877 markers with an average marker density of 0.65 cM. 34 QTLs related to fruit weight and size were identified, some of which have been already reported in the literature. New phenotypic evaluations will be conducted over the next two years in order to confirm

these QTLs.

In future assessments, the markers contained in these regions may be used as reference to validate these QTLs in other populations that segregate for fruit weight and size, and may eventually be used in MAS as a tool for breeding for fruit weight.

In chapter 6, the preliminary study of association mapping to detect markers associated with fruit weight and size gave promising results. 39 markers were found to be associated with fruit weight, height, width and depth. Some of these chromosomal regions correspond to QTLs mapped in chapter 5 in the F2 population. Others were found in regions near transcripts involved in biological processes related to fruit development such as cell division and cell expansion. However, it should be noted that this is a preliminary study. Further analysis will be conducted using MLM (mixed linear model) in which the effect of the Kinship matrix (matrix K) will be introduced in addition to the effect of the population matrix (matrix Q). Phenotypic evaluations will also be repeated, and the number of accessions used in the association mapping will be expanded.

Literature

A.G. Abbott, S. Rajapakse, B. Sosinski, Z.X. Lu, K. Sossey-Alaoui, M. Gannavarapu, G. Reighard, R.E. Ballard, W.V. Baird, R. Scorza, and A. Callahan. Construction of saturated linkage maps of peach crosses segregating for characters controlling fruit quality, tree architecture and pest resistance. *Acta Horticulturae*, 465:41–49, 1997.

A.G. Abbott, B. Sosinski, and A. Orellana. *Plant Genetics and Genomics: Crops and Models*, volume 6, chapter Functional Genomics in Peach. Springer, New York, 2009.

I.Y. Abdurakhmonov and A. Abdukarimov. Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int. J. Plant. Genomics*, 2008:574927, 2008.

I.Y. Abdurakhmonov, S. Saha, J.N. Jenkins, Z.T. Buriev, S.E. Shermatov, B.E. Scheffler, A.E. Pepper, J.Z. Yu, R.J. Kohel, and A. Abdukarimov. Linkage disequilibrium based association mapping of fiber quality traits in G. hirsutum L. variety germplasm. *Genetica*, 136(3):401–417, Jul 2009.

H.A. Agrama, G.C. Eizenga, and W. Yan. Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, 19:341–356, 2007.

K.V. Alheit, J.C. Reif, H.P. Maurer, V. Hahn, E.A. Weissmann, T. Miedaner, and T. Wurschum. Detection of segregation distortion loci in triticale (x

Triticosecale Wittmack) based on a high-density DArT marker consensus genetic linkage map. *BMC Genomics*, 12:380, 2011.

R.W. Allard. *Principles of plant breeding*. John Wiley, New York, 1960.

M.T. Andres, A. Benito, G. Perez-Rivera, R. Ocete, M.A. Lopez, L. Gaforio, G. Munoz, F. Cabello, J.M. Martinez Zapater, and R. Arroyo-Garcia. Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. *Mol. Ecol.*, 21(4):800–816, Feb 2012.

M.J. Aranzana, E.L.K. Abbassi, W. Howad, and P. Arús. Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genet.*, 11:69, 2010.

M.J. Aranzana, J. Garcia-Mas, Carbó, and P. Arús. Development and variability analysis of microsatellite markers in peach. *Plant Breeding*, 121: 87–92, 2002.

M.J. Aranzana, A. Pineda, P. Cosson, E. Dirlewanger, J. Ascasibar, G. Cipriani, C.D. Ryder, R. Testolin, A. Abbott, G.J. King, A.F. Iezzoni, and P. Arús. A set of simple-sequence repeat (SSR) markers covering the Prunus genome. *Theor. Appl. Genet.*, 106(5):819–825, Mar 2003.

P. Arús, I. Verde, B. Sosinski, T. Zhebentyayeva, and A. Abbott. The peach genome. *Tree Genetics & Genomes*, 8:531–547, 2012.

W.V. Baird, A.S. Estager, and J.K. Wells. Estimating nuclear-DNA content in peach and related diploid species using laser flow-cytometry and DNA hybridization. *J. Am. Soc. Hort. Sci.*, 119(6):1312–1316, 1994.

A.D. Ball, J. Stapley, D.A. Dawson, T.R. Birkhead, T. Burke, and J. Slate. A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (Taeniopygia guttata). *BMC Genomics*, 11 (218):1–15, Apr 2010.

D. Bassi and R. Monet. *The peach*, volume 1, chapter Botany and taxonomy, pages 1–36. CAB International, Wallingford, 2008.

D.A.V. Bento. *Mapeamento de QTLs para produção de grãos e seus componentes em uma população de milho tropical*. PhD thesis, Universidade de São Paulo, 2006.

N. Bertin, M. Causse, B. Brunel, D. Tricon, and M. Gnard. Identification of growth processes involved in QTLs for tomato fruit size and composition. *Journal of Experimental Botany*, 60:237–248, 2009.

M. Bertoglio. *Studio sulla dimensione del frutto nel pesco (Prunus persica L. Batsch)*. PhD thesis, Università degli studi di Milano, 2010.

M.A. Blake. The J.H. Hale peach as a parent in peach crosses. *Proceedings of the American Society for Horticultural Science*, 29:131–136, 1932.

M.A. Blake and C.H. Connors. Early results of peach breeding in New Jersey. *New Jersey Agricultural Experiment Station Bulletin*, 599, 1936.

A. Blas, Q. Yu, O. Veatch, R. Paull, P. Moore, and R. Ming. Genetic mapping of quantitative trait loci controlling fruit size and shape in papaya. *Molecular Breeding*, 29:457–466, 2012.

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

M. Bouhadida, M. Moreno, M. Gonzalo, J. Alonso, and Y. Gogorcena. Genetic variability of introduced and local spanish peach cultivars determined by ssr markers. *Tree Genetics & Genomes*, 7:257–270, 2011.

B. Brachi, N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, J. Bergelson, J. Cuguen, and F. Roux. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. *PLoS Genet.*, 6(5):e1000940, May 2010.

P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23:2633–2635, 2007.

F. Brandi, E. Bar, F. Mourgues, G. Horvath, E. Turcsi, G. Giuliano, A. Liverani, S. Tartarini, E. Lewinsohn, and C. Rosati. Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.*, 11:24, 2011.

F. Breseghello and M.E. Sorrells. Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.*, 46(1323-1330), 2006.

A. Cabrera. *Genetic Analysis and Fruit weight QTL fine mapping in Sweet cherry (Prunus avium L.).* PhD thesis, Ohio State University, 2011.

O.E. Campbell, I.A. Merwin, and O.I. Padilla-Zakour. Nutritional quality of New York peaches and apricots. *New York Fruit Quarterly*, 19(4):12–16, 2011.

C.M. Cantín, C.H. Crisostob, E.A. Ogundiwinb, T. Gradzielb, J. Torrentsc, M.A. Morenoa, and Y. Gogorcenaa. Chilling injury susceptibility in an intra-specific peach [Prunus persica (L.) Batsch] progeny. *Postharvest Biology and Technology*, 58:79–87, 2010.

K. Cao, L. Wang, G. Zhu, W. Fang, C. Chen, and J. Luo. Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (Prunus persica) landraces in China. *Tree Genetics & Genomes*, pages 1–16, 2012.

M.S. Carneiro and M.L.C. Vieira. Mapas genéticos em plantas. *Bragantia*, 61(2):89–100, 2002.

A.M. Casa, G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. Community resources and strategies for association mapping in sorghum. *Crop Science*, 48:30–40, 2008.

M. Causse, J. Chaïb, L. Lecomte, M. Buret, and F Hospital. Both additivity and epistasis control the genetic variations for fruit quality traits in tomato. *Theoretical and Applied Genetics*, 115:429–442, 2007.

L. Chalak, A. Chehade, A. Elbitar, P. Cosson, A. Zanetto, E. Dirlewanger, and F. Laigret. Morphological and molecular characterization of peach accessions (Prunus persica L.) cultivated in Lebanon. *Lebanese Science Journal*, 7(2):23–31, 2006.

D.J. Chalmers and B. van den Ende. Productivity of peach trees: factors affecting dry-weight distribution during tree growth. *Ann. Bot.*, 39:423–432, 1975.

J.X. Chaparro, D.J. Werner, D. O'Malley, and R.R. Sederoff. Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theor. Appl. Genet.*, 87:805–815, 1994.

E.J. Cheong. Biotechnological approaches for improvement and conservation of Prunus species. *Plant Biotechnol. Rep.*, 6:17–28, 2012.

J. Cockram, J. White, F.J. Leigh, V.J. Lea, E. Chiapparino, D.A. Laurie, I.J. Mackay, Powell W., , and O'Sullivan D.M. Association mapping of partitioning loci in barley. *BMC genetics*, 9:16, 2008.

B.C. Collard and D.J. Mackill. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 363(1491):557–572, Feb 2008.

B.C.Y. Collard, M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142:169–196, 2005.

B. Cong, L. S. Barrero, and S. D. Tanksley. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.*, 40(6):800–804, Jun 2008.

B. Cong, J. Liu, and S.D. Tanksley. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl. Acad. Sci. USA*, 99:13606–13611, 2002.

B.G. Coombe. The development of fleshy fruits. *Annual Review of Plant Physiology*, 27:507–528, 1976.

C.D. Cruz. Genes versão 2007. Aplicativo computacional em genética e estatística, 2007.

C.D. Dardick, A.M. Callahan, R. Chiozzotto, R.J. Schaffer, M.C. Piagnani, and R. Scorza. Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to Arabidopsis dehiscence. *BMC Biol.*, 8:13, 2010.

A. Darvasi and S. Shifman. The beauty of admixture. *Nat. Genet.*, 37: 118–119, 2005.

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

F. Devoghalaere, T. Doucen, B. Guitton, J. Keeling, W. Payne, T.J. Ling, J.J. Ross, I.C. Hallett, K. Gunaseelan, G.A. Dayatilake, R. Diak, K.C. Breen, D.S. Tustin, E. Costes, D. Chagne, R.J. Schaffer, and K.M. David. A genomics approach to understanding the role of auxin in apple (Malus x domestica) fruit size control. *BMC Plant Biol.*, 12:7, 2012.

B.B. D'hoop, M.J. Paulo, R.A. Mank, H.J. van Eck, and F.A. van Eeuwijk. Association mapping of quality traits in potato (Solanum tuberosum L. *Euphytica*, 161:47–60, 2008.

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

E. Dirlewanger, E. Graziano, T. Joobeur, F. Garriga-Caldere, P. Cosson, W. Howad, and P. Arús. Comparative mapping and marker-assisted se-

lection in Rosaceae fruit crops. *Proc. Natl. Acad. Sci. U.S.A.*, 101(26): 9891–9896, Jun 2004.

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

E. Dirlewanger, Crosson P., M. Tavaud, M.J. Aranzana, C. Poizat, A. Zanetto, P. Arús, and F. Laigret. 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.). *Theor. Appl. Genet.*, 105:127–138, 2002.

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

R.W. Doerge. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.*, 3:43–52, 2002.

D.A. Earl and B.M. vonHoldt. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2):359–361, 2012.

I. Eduardo, I. Pacheco, G. Chietera, D. Bassi, C. Pozzi, A. Vecchietti, and L. Rossini. QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genetics & Genomes*, 7:323–335, 2011.

I. El-Sharkawy, W.S. Kim, A. El-Kereamy, S. Jayasankar, A.M. Svircev, and D.C. Brown. Isolation and characterization of four ethylene signal transduction elements in plums (Prunus salicina L.). *J. Exp. Bot.*, 58(13): 3631–3643, 2007.

E.S. Ersoz, J. Yu, and E.S. Buckler. Applications of linkage disequilibrium and association mapping in maize. In Alan L. Kriz and Brian A.

Larkins, editors, *Molecular Genetic Approaches to Maize Improvement*, volume 63 of *Biotechnology in Agriculture and Forestry*, pages 173–195. Springer Berlin Heidelberg, 2009.

C. Esteras, P. Gomez, A.J. Monforte, J. Blanca, N. Vicente-Dolera, C. Roig, F. Nuez, and B. Pico. High-throughput SNP genotyping in Cucurbita pepo for map construction and quantitative trait loci mapping. *BMC Genomics*, 13:80, 2012.

C. Etienne, C. Rothan, A. Moing, C. Plomion, C. Bodenes, L. Svanella-Dumas, P. Cosson, V. Pronier, R. Monet, and E. Dirlewanger. Candidate genes and QTLs for sugar and organic acid content in peach [Prunus persica (L.) Batsch]. *Theor. Appl. Genet.*, 105(1):145–159, Jul 2002.

G. Evanno, S. Regnaut, and J. Goudet. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8):2611–2620, 2005.

W.J. Ewens and R.S. Spielman. The transmission/disequilibrium test: history, subdivision, and admixture. *Amer. J. Hum. Genet.*, 57:455–464, 1995.

D. Falush, M. Stephens, and J.K. Pritchard. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164:1567–1587, 2003.

J.B. Fan, A. Oliphant, R. Shen, B.G. Kermani, F. Garcia, K.L. Gunderson, M. Hansen, F. Steemers, S.L. Butler, P. Deloukas, L. Galver, S. Hunt, C. McBride, M. Bibikova, T. Rubano, J. Chen, E. Wickham, D. Doucet, W. Chang, D. Campbell, B. Zhang, S. Kruglyak, D. Bentley, J. Haas, P. Rigault, L. Zhou, J. Stuelpnagel, and M.S. Chee. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, 68: 69–78, 2003.

S. Fan, D.G. Bielenberg, T.N. Zhebentyayeva, G.L. Reighard, W.R. Okie, D. Holland, and A.G. Abbott. Mapping quantitative trait loci associ-

ated with chilling requirement, heat requirement and bloom date in peach (Prunus persica). *New Phytol.*, 185(4):917–930, Mar 2010.

FAOSTAT. Statistical databases, food and agriculture organization of the United Nations, 2010. URL `http://faostat.fao.org`.

A.R. Fernie, Y. Tadmor, and D. Zamir. Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology*, 9:196–202, 2006.

K.M. Folta and S.E. Gardiner. *Plant Genetics and Genomics: Crops and Models*, volume 6, chapter Genetic and genomics of Rosaceae. Springer, New York, 2009.

M. Foulongne, T. Pascal, P. Arús, and J. Kervella. The potential of Prunus davidiana for introgression into peach [Prunus persica (L.) Batsch] assessed by comparative mapping. *Theor. Appl. Genet.*, 107(2):227–238, Jul 2003.

A. Frary, TC. Nesbitt, S. Grandillo, E. van der Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289:85–88, 2000.

M.O.S. Garcia, C.T. Karia, R.M. Resende, L. Chiari, M.L. Vieira, M.I. Zucchi, and A.P. Souza. Identification of Stylosanthes guianensis varieties using molecular genetic analysis. *AoB Plants*, 2012:pls001, 2012.

M. Genard, N. Bertin, C. Borel, P. Bussieres, H. Gautier, R. Habib, M. Lechaudel, A. Lecomte, F. Lescourret, P. Lobit, and B. Quilot. Towards a virtual fruit focusing on quality: modelling features and potential uses. *J. Exp. Bot.*, 58(5):917–928, 2007.

A.H. Gilani, N. Aziz, S.M. Ali, and M. Saeed. Pharmacological basis for the use of peach leaves in constipation. *J. Ethnopharmacol.*, 73(1–2):87–93, Nov 2000.

S.A. Gilani. A molecular phylogeny of selected species of genus Prunus L. (Rosaceae) from transcribed spacer ITS spacer DNA. *African Journal of Biotechnology*, 9:4867–4872, 2010.

A.M. Gillen and F.A. Bliss. Identification and mapping of markers linked to the Mi gene for root-knot nematode resistance in peach. *J. Americ. Soc. Hort. Sci.*, 130(1):24–33, 2006.

M. González-Agüero, L. Pavez, F. Ibáñez, I. Pacheco, R. Campos-Vargas, L.A. Meisel, A. Orellana, J. Retamales, H. Silva, M. González, and V. Cambiazo. Identification of woolliness response genes in peach fruit after post-harvest treatments. *Journal of Experimental Botany*, 59(8): 1973–1986, 2008.

S. Grandillo, H.M. Ku, and S.D. Tanksley. Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor. Appl. Genet.*, 99:978–987, 1999.

P.D.A. Guedes, O.S. Almeida, O.L. Lemos, and T.N.H. Reboucas. Relação fonte-dreno na formação de frutos: uma revisão bibliográfica. *Diálogos e Ciências*, 13:1–13, 2008.

O. Hammer, D.A.T. Harper, and R.D. Ryan. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1):1–9, 2001.

J.C. Han, G.J. Liu, R.F. Chang, and X.Z. Zhang. The sequence-related amplified polymorphism (SRAP) markers linked to the color around the stone (Cs) locus of peach fruit. *African Journal of Biotechnology*, 11(42): 9911–9914, May 2012.

T. Harada, W. Kurahashi, M. Yanai, Y. Wakasa, and T. Satoh. Involvement of cell proliferation and cell enlargement in increasing the fruit size of malus species. *Sci Hort (Amsterdam)*, 105:447–456, 2005.

M.J. Hayden, T.M. Nguyen, A. Waterman, and K.J. Chalmers. Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. *BMC Genomics*, 9:80, 2008.

C.O. Hesse. *Advances in Fruit Breeding, Temperate fruits*, chapter Peaches, pages 285–335. Purdue University Press, West Lafayette, USA, 1975.

J.N. Hirschhorn and M.J. Daly. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, 6:95–108, 2005.

D. Hu, Z. Zhang, D. Zhang, Q. Zhang, and J. Li. Genetic relationship of ornamental peach determined using AFLP markers. *HortScience*, 40(6): 1782–1786, 2005.

Z. Huang and E. van der Knaap. Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. *Theor. Appl. Genet.*, 123(3): 465–474, Aug 2011.

R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. *Journal of Computational & Graphical Statistics*, 5:299–314, 1996.

E. Illa, I. Eduardo, J.M. Audergon, F. Barale, E. Dirlewanger, X.W. Li, A. Moing, P. Lambert, L.L. Dantec, Z.S. Gao, J.L. Poëssel, C. Pozzi, L. Rossini, A. Vecchietti, P. Arús, and W Howad. Saturating the Prunus (stone fruits) genome with candidate genes for fruit quality. *Mol. Breed.*, 28(4):667–682, 2010.

R. Infante, P. Martínez-Gómez, and S. Predieri. Quality oriented fruit breeding: Peach [Prunus persica (L.) Batsch]. *Journal of Food, Agriculture & Environment*, 6(2):342–356, 2008.

M. Jangarelli, R.F. Euclydes, and P.R. Cecon. Desempenho fenotípico ao utilizar diferentes densidades de marcadores moleculares no mapeamento genômico. *Biosci. J. Uberlândia*, 26(4):626–631, 2010.

C. Jiang and Z. B. Zeng. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, 140(3):1111–1127, Jul 1995.

T. Joobeur, M.A. Viruel, M.C. Vicente, B. Jáuregui, J. Ballester, M.T. Dettori, I. Verde, M.J. Truco, R. Messeguer, I. Batlle, R. Quarta, E. Dirlewanger, and P. Arús. Construction of a saturated linkage map for Prunus using an almond x peach F2 progeny. *Theor. Appl. Genet.*, 97: 1034–1041, 1998.

A.W. Junior, C.H. Bruckner, C.M. Cantín, M.A.M. Sánchez, and C.D. Cruz. Divergência genética entre progênies de pessegueiro em Zaragoza, Espanha. *Rev. Bras. Frutic. Jaboticabal*, 33(1):303–310, Mar 2011.

D.H. Júnior. *Análise de ácido clorogênico em amostras de pêssego – Prunus persica L. Batsch: otimização e validação de método*. PhD thesis, Universidade federal do Rio Grande do Sul, Porto Alegre, 2007.

C.H. Kao, Z.B. Zeng, and R.D. Teasdale. Multiple interval mapping for uantitative trait loci. *Genetics Bethesda*, 152(3):1203–1216, 1999.

P. Keim, W. Beavis, J. Schupp, and R. Freestone. Evaluation of soybean RFLP marker diversity in adapted germ plasm. *Theor. Appl. Genet.*, 85: 202–212, 1992.

B.D. Kohorn and S.L. Kohorn. The cell wall-associated kinases, WAKs, as pectin receptors. *Front. Plant Sci.*, 3:88, May 2012.

W.J. Koopman, V. Wissemann, K. De Cock, J. Van Huylenbroeck, J. De Riek, G.J. Sabatino, D. Visser, B. Vosman, C.M. Ritz, B. Maes, G. Werlemark, H. Nybom, T. Debener, M. Linde, and M.J. Smulders. AFLP markers as a tool to reconstruct complex relationships: A case study in Rosa (Rosaceae). *Am. J. Bot.*, 95(3):353–366, Mar 2008.

M.M. Kopp, V.Q. Souza, J.L.M. Coimbra, V.K. Luz, N. Marini, and A.C. Oliveira. Melhoria da correlação cofenética pela exclusão de unidades experimentais na construção de dendrogramas. *Revista da FZVA Uruguaiana*, 14(2):46–53, 2007.

P.J. Krysan, P.J. Jester, and M.R. Sussman. An Arabidopsis mitogen activated protein kinase kinase kinase gene family encodes essential positive regulators of cytokinesis. *The Plant Cell*, 14:1109–1120, 2002.

F. Lamy, S. Andre, T. Robert, and A. Sarr. Use of isozymes and RFLP markers to assess genetic diversity in introgression procedures between adapted and non adapted maize populations (Zea mays L.). *Hereditas*, 120:165–173, 1994.

E. Lander and D. Botstein. Mapping Mendelian fators underlying quantitative trait using RFLP linkage maps. *Genetics Bethesda*, 121:185–199, 1989.

E.S. Lander and N.J. Schork. Genetic dissection of complex traits. *Science*, 265:2037–2048, 1994.

P. Lebrun, Y.P. Ncho, M. Seguin, L. Grivet, and L. Baudouin. Genetic diversity in coconut (Cocos nucifera L.) revealed by restriction fragment length polymorphism (RFLP) markers. *Euphytica*, 101(1):103–108, 1998.

S. Lee and J. Wen. A phylogenetic analysis of Prunus and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *Am. J. Bot.*, 88(1):150–160, Jan 2001.

R.C. Lewotin. The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics*, 49:49–67, 1964.

R. Li, S.W. Tsaih, K. Shockley, I.M. Stylianou, J. Wergedal, B. Paigen, and G.A. Churchill. Structural model analysis of multiple quantitative traits. *PLoS Genet.*, 2(7):e114, Jul 2006.

T.H. Li, Y.X. Li, Z.C. Li, H.L. Zhang, Y.W. Qi, and T. Wang. Simple sequence repeat analysis of genetic diversity in primary core collection of peach (Prunus persica). *J. Integr. Plant Biol.*, 50(1):102–110, Jan 2008.

W.D. Li, S.H. Li, S.H. Yang, J.M. Yang, X.B. Zheng, X.D. Li, and H.M. Yao. Photosynthesis in response to sinksource manipulation during different phenological stages of fruit development in peach trees: regulation by stomatal aperture and leaf temperature. *J. Hortic. Sci. Biotechnol.*, 80: 481–487, 2005.

Y. Li, B. Wu, Y. Yu, G. Yang, C. Wu, and C. Zheng. Genome-wide analysis of the RING finger gene family in apple. *Mol. Genet. Genomics*, 286(1): 81–94, Jul 2011.

Z. Li, T. Peng, Q. Xie, S. Han, and J. Tian. Mapping of QTL for tiller number at different stages of growth in wheat using double haploid and immortalized F2 populations. *J. Genet.*, 89(4):409–415, Dec 2010.

M.R. Lima, E. Augustin, E. Choer, and M.C.B. Raseira. Caracterização de cultivares de pessegueiro e de nectarineira por marcadores moleculares. *Pesq. agropec. bras. Brasília*, 38(3):349–355, Mar 2003.

A.B. Lins. *Distancias genéticas estimadas com marcadores moleculares e associação com performance de híbridos de Theobroma cacao L.* PhD thesis, Universidade estadual de Santa Cruz, Ilhéus, Brazil, Jul 2008.

B. Liu. *Statistical Genomics: Linkage, Mapping and QTL Analysi.* CRC Press, Boca Raton, 1998.

J. Liu, K. Nada, X. Pang, C. Honda, H. Kitashiba, and T. Moriguchi. Role of polyamines in peach fruit development and storage. *Tree Physiol.*, 26 (6):791–798, Jun 2006.

J. Liu, J. Van Eck, B. Cong, and S.D. Tanksley. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):13302–13306, Oct 2002.

V.A. Lombardo, S. Osorio, J. Borsani, M.A. Lauxmann, C.A. Bustamante, C.O. Budde, C.S. Andreo, M.V. Lara, A.R. Fernie, and M.F. Drincovich. Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. *Plant Physiol.*, 157(4):1696–1710, Dec 2011.

F.C.C. Lopes. *Mapeamento genético de cana-de-açúcar (Saccharum spp.) por associação empregando marcadores SSR e AFLP.* PhD thesis, Universidade de São Paulo, 2011.

G. Lopez, R.S. Johnson, and T.M. DeJong. High spring temperatures decrease peach fruit size. *California Agriculture*, 61:31–34, 2007.

Z.X. Lu, B. Sosinski, G.L. Reighard, W.V. Baird, and A.G. Abbott. 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, 1998.

M. Lynch and B. Walsh. *Genetics and analysis of quantitative traits.* Sunderland Sinauer Associates, 1998.

T.F. Mackay. The genetic architecture of quantitative traits. *Annu. Rev. Genet.*, 35:303–339, 2001.

S. Mahanil, D. Ramming, M. Cadle-Davidson, C. Owens, A. Garris, S. Myles, and L. Cadle-Davidson. Development of marker sets useful in the early selection of Ren4 powdery mildew resistance and seedlessness for table and raisin grape breeding. *Theor. Appl. Genet.*, 124(1):23–33, Jan 2012.

R. Marini. How to grow big peaches, 2002. URL `http://njaes.rutgers.edu/peach/orchard/bigpeaches.pdf`.

P. Martínez-García, D. Parfitt, E. Ogundiwin, J. Fass, H. Chan, R. Ahmad, S. Lurie, A. Dandekar, T. Gradziel, and C. Crisosto. High density SNP mapping and QTL analysis for fruit quality characteristics in peach (Prunus persica L.). *Tree Genetics & Genomes*, pages 1–18, 2012.

V. Mercier, C. Bussi, F. Lescourret, and M. Génard. Effects of different irrigation regimes applied during the final stage of rapid growth on an early maturing peach cultivar. *Irrigation Science*, 27:297–306, 2009.

A.K. Mishra and N.K. Dubey. Fungitoxic properties of prunus persica oil. *Hindustan Antibiotic Bulletin*, 32:91–93, 1990.

R. Monet. Transmission génétique du caracter fruit doux chez le pêcher. incidence sur la selection pour la qualité. *Proceedings of Eucarpia, Fruit Section, Tree Fruit Breeding*, pages 273–279, 1979.

B. Morandi, L.C. Grappadelli, M. Rieger, and R. Lo Bianco. Carbohydrate availability affects growth and metabolism in peach fruit. *Physiologia Plantarum*, 133:229–241, 2008.

B. Morandi, M. Rieger, and L.C. Grappadelli. Vascular flows and transpiration affect peach (Prunus persica Batsch) fruit daily growth. *J. Exp. Bot.*, 58:3941–3947, 2007.

S.C. Murray, W.L.M. Rooney, T.S.E.H. Mitchell, and S. Kresovich. Sweet sorghum genetic diversity and association mapping for brix and height. *The Plant Genome*, 2:48–62, 2009.

S. Myles, A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, A. Reynolds, J.M. Chia, D. Ware, C.D. Bustamante, and E.S. Buckler. Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. USA*, 108:3530–3535, 2011.

S. Myles, J. M. Chia, B. Hurwitz, C. Simon, G. Y. Zhong, E. Buckler, and D. Ware. Rapid genomic characterization of the genus Vitis. *PLoS ONE*, 5(1):e8219, 2010.

S. Myles, J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*, 21(8):2194–2202, Aug 2009.

J. Nadwodnik and G. Lohaus. Subcellular concentrations of sugar alcohols and sugars in relation to phloem translocation in Plantago major, Plantago maritima, Prunus persica, and Apium graveolens. *Planta*, 227:1079–1089, 2008.

E.A. Ogundiwin, C.P. Peace, T.M. Gradziel, D.E. Parfitt, F.A. Bliss, and C.H. Crisosto. A fruit quality gene map of Prunus. *BMC Genomics*, 10: 587, 2009.

P.R.D. Oliveira, D.C. Scotton, D.S. Nishimura, and A. Figueira. Análise da diversidade genética por AFLP e identificação de marcadores associados à resistência a doenças em videira. *Rev. Bras. Frutic. Jaboticabal*, 27(3): 454–457, Dec 2005.

J.W. Olmstead, A.F. Iezzoni, and M.D. Whiting. Genotypic differences in sweet cherry fruit size are primarily a function of cell number. *J. Amer. Soc. Hort. Sci.*, 132:697–703, 2007.

I.A. Pacheco Cruz. *Brown rot resistance in peach: a genomics approach*. PhD thesis, Università degli studi di Milano, Nov 2010.

I. Paran and E. van der Knaap. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.*, 58: 3841–3852, 2007.

R.K. Pasam, R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.*, 12:16, 2012.

L.A. Paula, V.J. Bianchi, and J.C. Fachinello. Caracterização molecular e variabilidade genética entre porta-enxertos de pessegueiro com base em marcadores codominantes. *Pesq. agropec. bras. Brasília*, 47(2):193–199, Feb 2012.

C.P. Peace, C.H. Crisosto, and T.M. Gradziel. Endopolygalacturonase: a candidate gene for freestone and melting flesh in peach. *Mol. Breed.*, 16: 21–31, 2005.

R. Peakall and P.E. Smouse. GenAlEx 6: genetic analysis in Excel. *Molecular Ecology Notes*, 6:288–295, 2006.

R.M.N. Peil and J. López-Galvéz. Effect of fruit removal on growth and biomass partitioning in cucumber. *Acta Horticulturae*, 588:69–74, 2002.

X. Perrier and J.P. Jacquemond-Collet. Darwin software, 2006. URL `http://darwin.cirad.fr`.

J.K. Pritchard and N.A. Rosenberg. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.*, 65: 220–228, 1999.

J.K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000a.

J.K. Pritchard, M. Stephens, N.A. Rosenberg, and P. Donnelly. Association mapping in structured populations. *Amer. J. Hum. Genet.*, 67:170–181, 2000b.

R. Quarta, M.T. Dettori, A. Sartori, and I. Verde. Genetic linkage map and QTL analysis in peach. *Acta Horticulturae*, 521:233–241, 2000.

R. Quarta, M.T. Dettori, I. Verde, A. Gentile, and Z. Broda. Genetic analysis of agronomic traits and genetics linkage mapping in a BC1 peach population using RFLPs and RAPDs. *Acta Horticulturae*, 465:51–59, 1998.

B. Quilot, J. Kervella, M. Genard, and F. Lescourret. Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. *J. Exp. Bot.*, 56(422):3083–3092, Dec 2005.

B. Quilot, B.H. Wu, J. Kervella, M. Genard, M. Foulongne, and K. Moreau. QTL analysis of quality traits in an advanced backcross between Prunus persica cultivars and the wild relative species P. davidiana. *Theor. Appl. Genet.*, 109(4):884–897, Aug 2004.

S. Rajapakse. Progress in application of molecular markers to genetic improvement of horticultural crops. *Acta. Hortic.*, 625:29–36, 2003.

S. Rajapakse, L.E. Belthoff, G. He, A.E. Estager, R. Scorza, I. Verde, R.E. Ballard, W.V. Baird, A. Callahan, R. Monet, and A.G. Abbott. Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. *Theor. Appl. Genet.*, 90:503–510, 1995.

J.C. Reif, M. Gowda, H.P. Maurer, C.F. Longin, V. Korzun, E. Ebmeyer, R. Bothe, C. Pietsch, and T. Wurschum. Association mapping for quality traits in soft winter wheat. *Theor. Appl. Genet.*, 122(5):961–970, Mar 2011.

D.L. Remington, J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA*, 98:11479–11484, 2001.

N. Risch and K. Merikangas. The future of genetic studies of complex human diseases. *Science*, 273:1516–1517, 1996.

S. Rivers. The cross-breeding of peaches and nectarines. Report on third international conference on genetics, 1906.

J.I.S. Rodrigues, F.D. Miranda, A. Ferreira, L.L. Borges, M.F.S. Ferreira, P.I.V. Good-God, N.D. Piovesan, E.G. Barros, C.D. Cruz, and M.A. Moreira. Mapeamento de QTL para conteúdos de proteína e óleo em soja. *Pesq. agropec. bras. Brasília*, 45(5):472–480, May 2010.

G.R. Rodriguez, S. Munos, C. Anderson, S.C. Sim, A. Michel, M. Causse, B.B. Gardener, D. Francis, and E. van der Knaap. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.*, 156(1):275–285, May 2011.

S.B. Rossato. *Potencial antioxidante e compostos fenólicos de pêssegos (Prunus persica L. Batsch).* PhD thesis, Universidade do Rio Grande do Sul, 2009.

P.K. Sabadin. *Mapeamento de QTLs e base genética da correlação entre caracteres em uma população de milho tropical.* PhD thesis, Universidade de São Paulo, 2008.

M. Sakiroglu. *Population structure, genetic diversity, phylogenetic analyses, and association mapping of biofuel traits in wild diploid alfalfa (Medicago sativa L.) accessions.* PhD thesis, University of Georgia, Athens, Georgia, 2009.

M. Sakiroglu, S. Sherman-Broyles, A. Story, K.J. Moore, J.J. Doyle, and E.C. Brummer. Patterns of linkage disequilibrium and association mapping in diploid alfalfa (M. sativa L.). *Theor Appl Genet*, Apr 2012.

S. Sansavini. Biotecnologie frutticole: le nuove frontiere delle ricerche per il miglioramento genetico e la propagazione delle piante da frutto. *Frutticoltura Bologna*, 5:75–81, 1998.

F.H.C. Santos, J.J.V. Cavalcanti, and F.P. Silva. Detection of quantitative trait loci for physical traits of cashew apple. *Crop Breeding and Applied Biotechnology*, 10:101–109, 2010.

SAS-Institute, editor. *SAS language reference*. SAS Institute, 8th edition, 1999.

R. Scorza, S.A. Mehlenbacher, and G.W. Lightner. Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germoplasm improvement. *American Society for Horticultural Science Journal*, 110:547–552, 1985.

D.M. Segantini, S. Leonel, G.P.P. Lima, S.M. Costa, and A.R.P. Ramos. Caracterização da polpa de pêssegos produzidos em São Manuel-SP. *Ciência Rural*, 42(1):52–57, Jan 2012.

D. Shriner, L.K. Vaughan, M.A. Padilla, and H.K. Tiwari. Problems with genome-wide association studies. *Science*, 316, 2007.

V. Shulaev, S.S. Korban, B. Sosinski, A.G. Abbott, H.S. Aldwinckle, K.M. Folta, A. Iezzoni, D. Main, P. Arús, A.M. Dandekar, K. Lewers, S.K. Brown, T.M. Davis, S.E. Gardiner, D. Potter, and R.E. Veilleux. Multiple models for Rosaceae genomics. *Plant Physiol.*, 147(3):985–1003, Jul 2008.

I. Simko and J. Hu. Population structure in cultivated lettuce and its impact on association mapping. *J. Amer. Soc. Hort. Sci.*, 133(1):61–68, 2008.

V. Sitther, D. Zhang, S.A. Dhekney, D.L. Harris, A. Yadav, and W.R. Okie. Cultivar identification, pedigree verification, and diversity analysis among peach cultivars based on simple sequence repeat markers. *J. Amerc. Soc. Hort. Sci.*, 137:114–121, Mar 2012.

P.H. Sneath and R.R. Sokal. *Numerical taxonomy: The principles and practice of numerical classification.* W.H. Freeman, San Francisco, 1973.

R.A.P. Sørensen, J. Stuurman, J.R. van der Voort, and J. Peleman. *Genomics Approaches and Platforms*, chapter Genomics-Assisted Crop Improvement. Springer Netherlands, 2009.

K. Sorkheh, L.V. Malysheva-Otto, M.G. Wirthensohn, S. Tarkesh-Esfahani, and P. Martínez-Gómez. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genetics and Molecular Biology*, 31(4):805–814, 2008.

B. Sosinski, M. Gannavarapu, and L.D. Hager. Characterization of microsatellite markers in peach [Prunus persica (L.) Batsch]. *Theoretial and Applied Genetics New York*, 101:421–428, 2000.

F.F. Souza. *Diversidade génetica, estretura populacional e mapeamento associativo em Coffea canephora*. PhD thesis, Universidade Federal de Viçosa, Minas Gerais, Brazil, 2011.

B. Stich, J. Mohring, H.P. Piepho, M. Heckenberger, E.S. Buckler, and A.E. Melchinger. Comparison of mixed-model approaches for association mapping. *Genetics*, 178(3):1745–1754, Mar 2008.

C.W. Stuber, M.D. Edwards, and J.F. Wendell. Molecular marker facilitated investigations quantitative trait loci in maize. ii. factors influencing yield and it component traits. *Crop Science Madison*, 27:639–648, 1987.

G. Sun, C. Zhu, M.H. Kramer, S.S. Yang, W. Song, H.P. Piepho, and J. Yu. Variation explained in mixed-model association mapping. *Heredity (Edinb.)*, 105(4):333–340, Oct 2010.

S. D. Tanksley. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell*, 16 Suppl:S181–189, 2004.

S. Tavarini, M.I. Gil, F.A. Tomas-Barberan, B. Buendia, D. Remorini, R. Massai, E. Degl'Innocenti, and L. Guidi. Effects of water stress and rootstocks on fruit phenolic composition and physical/chemical quality in Suncrest peach. *Annals of Applied Biology*, 158(2):226–233, 2011.

A. Teneva. Molecular markers in animal genome analysis. *Biotechnology in Animal Husbandry*, 25(5-6):1267–1284, 2009.

J.M. Thornsberry, M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.*, 28:286–289, 2001.

C.S. Tibola, J.C. Fachinello, C.V. Rombaldi, L. Picolotto, and L. Krüge. Análise de confomidade na adoção das normas de produção integrada de pêssego. *Ciência Rural Santa Maria*, 37(4):1149–1152, 2007.

E.R. Toledo, R.A. Leandro, C.L.S. Junior, and A.P. Souza. Mapeamento de QTLs: uma abordagem Bayesiana. *Rev. Bras. Biom. São Paulo*, 26(2): 107–114, 2008.

R.R. Toledo. *Mapeamento de QTLs utilizando as abordagens Clássica e Bayesiana*. PhD thesis, Universidade de São Paulo, 2006.

L. Trainotti, D. Zanin, and G. Casadoro. A cell wall-oriented genomic approach reveals a new and unexpected complexity of the softening in peaches. *J. Exp. Bot.*, 54:1821–1832, 2003.

USDA. U.S. Department of Agriculture, National Nutrient Database for Standard Reference, 2011. URL `http://www.ars.usda.gov/ba/bhnrc/ndl`.

J.W. van Ooijen. JoinMap 4 software for the calculation of genetic linkage maps in experimental populations, 2006.

J.W. Van Ooijen. MapQTL software for the mapping of quantitative trait loci in experimental populations of diploid species, 2009.

R.K. Varshney and R. Tuberosa. *Genomics Approaches and Platforms*, chapter Genomics-assisted crop improvement: an overview. Springer Netherlands, 2009.

A. Vecchietti, B. Lazzari, C. Ortugno, F. Bianchi, R. Malinverni, A. Caprera, I. Mignani, and C. Pozzi. Comparative analysis of expressed sequence tags

from tissues in ripening stages of peach (Prunus persica L. Batsch). *Tree Genetics & Genomes*, 5:377–391, 2009.

I. Verde, N. Bassil, S. Scalabrin, B. Gilmore, C.T. Lawley, K. Gasic, D. Micheletti, U.R. Rosyara, F. Cattonaro, E. Vendramin, D. Main, V. Aramini, A.L. Blas, T.C. Mockler, D.W. Bryant, L. Wilhelm, M. Troggio, B. Sosinski, M.J. Aranzana, P. Arús, A. Iezzoni, M. Morgante, and C. Peace. Correction: Development and Evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm. *PLoS One*, 7(6), 2012.

I. Verde, M. Lauria, M.T. Dettori, E. Vendramin, C. Balconi, S. Micali, Y. Wang, M.T. Marrazzo, G. Cipriani, H. Hartings, R. Testolin, A.G. Abbott, M. Motto, and R. Quarta. Microsatellite and AFLP markers in the Prunus persica [L. (Batsch)]xP. ferganensis BC(1)linkage map: saturation and coverage improvement. *Theor. Appl. Genet.*, 111(6):1013–1021, Oct 2005.

P. Vos, R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. AFLP: a new technique for dna fingerprinting. *Nucleic Acides Res.*, 23(21):4407–4414, 1995.

T.A. Wagner and B.D. Kohorn. Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *The Plant Cell*, 13:303–318, 2001.

H. Wang, K.P. Smith, E. Combs, T. Blake, R.D. Horsley, and G.J. Muehlbauer. Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor. Appl. Genet.*, 124(1):111–124, Jan 2012.

K.L. Wang, K. Bolitho, K. Grafton, A. Kortstee, S. Karunairetnam, T.K. McGhie, R.V. Espley, R.P. Hellens, and A.C. Allan. An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant. Biol.*, 10:50, 2010.

W.Y.S. Wang, B.J. Barratt, D.G. Clayton, and J.A. Todd. Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.*, 6:109–118, 2005.

J.G.K. Williams, A.R. Kubelik, J.A. Livak, K.J. nad Rafalski, and S.V. Tingey. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Res.*, 18:6531–6535, 1990.

S.I. Wright and B.S. Gaut. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.*, 22:506–519, 2005.

R.L. Wu. Mapping quantitative trait loci by genotyping haploid tissues. *Genetics*, 152(4):1741–1752, Aug 1999.

H. Xiao, N. Jiang, E.K. Schaffner, E.J. Stockinger, and E. van der Knaap. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science*, 319:1527–1530, 2008.

M. Yamaguchi, T. Haji, M. Miyake, and H. Yaegaki. Studies on the varietal differences and yearly deviation of mesocarp cell numbers and lengths and fruit weight among commercial peach [Prunus persica (L.) Batsch] cultivars and selections, wild types, and their hybrids. *Journal of the Japanese Society for Horticultural Science*, 71(4):459–466, 2002.

T. Yamamoto, T. Shimada, T. Imai, H. Yaegaki, T. Haji, N. Matsuta, M. Yamaguchi, and T. Hayashi. Characterization of morphological traits based on a genetic linkage map in peach. *Breeding science*, 51(4):271–278, 2001.

H. Yu, W. Xie, J. Wang, Y. Xing, C. Xu, X. Li, J. Xiao, and Q. Zhang. Gains in QTL detection using an ultra-high density snp map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE*, 6(3): e17595, 2011.

J. Yu, J.B. Holland, M.D. McMullen, and E.S. Buckler. Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178 (1):539–551, Jan 2008.

J. Yu, G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, 38(2):203–208, Feb 2006.

X.J. Yuan, X.Z. Li, J.S. Pan, G. Wang, S. Jiang, X.H. Li, S.L. Deng, H.L. He, M.X. Si, L. Lai, A.Z. Wu, L.H. Zhu, and R. Cai. Genetic linkage map construction and location of QTLs for fruit-related traits in cucumber. *Plant Breeding*, 127(2):180–188, 2008.

Z.B. Zeng. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA*, 90:10972–10976, 1993.

Z.B. Zeng. Precision mapping of quantitative trait loci. *Genetics*, 136:1457–1468, 1994.

G. Zhang, A. Sebolt, S. Sooriyapathirana, D. Wang, M. Bink, J. Olmstead, and A. Iezzoni. Fruit size QTL analysis of an F1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genetics & Genomes*, 6:25–36, 2010.

N. Zhang, M.T. Brewer, and E. van der Knaap. Fine mapping of fw3.2 controlling fruit weight in tomato. *Theor. Appl. Genet.*, 125(2):273–284, Jul 2012.

K. Zhao, M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram, and M. Nordborg. An Arabidopsis example of association mapping in structured samples. *PLoS Genet.*, 3(1): e4, Jan 2007.

C. Zhu, M. Gore, E.S. Buckler, and J. Yu. Status and prospects of association mapping in plants. *The Plant Genome*, 1:5–20, Mar 2008.

Glossary

AFLP	Amplified Fragment Length Polymorphism
DNA	Desoxyribonucleic Acid
FD	Fruit Depth
FH	Fruit Height
FW	Fruit Weight
FWD	Fruit Width
GLM	General Linear Model
IPSC	International Peach SNP Consortium
IM	Interval Mapping
IMA	Inter-Microsatellite Amplification
JGI	Joint Genome Institute
LD	Linkage Disequilibrium
LOD	Logarithm of Odds
MAPK	Mitogen-Activated Protein Kinase
MAS	Marker-Assisted Selection
MLM	Mixed Linear Model
MQM	Multiple QTL Mapping
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
QTL	Quantitative Trait Locus
SNP	Single-Nucleotide Polymorphism
SRAP	Sequence-Related Amplified Polymorphism
SSR	Simple Sequence Repeat
WAK	Wall-Associated Receptor Kinase

Acknowledgments

I thank God for the gift of life; my advisor Daniele Bassi for his trust, patience, understanding — I will always be grateful for the opportunity to develop a work so beautiful and rewarding; my co-advisor Laura Rossini for her encouragement, conversations, ideas and the enthusiasm with which she talks about science; the Università degli Studi di Milano and the Italian government for granting the scholarship; Stefano Biffani, Agostino Fricano, Igor Pacheco, Raul Pirona and Giorgiana Chietera for their support in clarifying some doubts and help in difficult times; my dear colleagues at the Plant Genomic Lab of the Parco Tecnologico Padano from whom I learned a lot; Stefano Foschi, Martina Lama, Emanuele Quattrini, Alice, Danilo, and Luciano for their help with collecting and analysing the phenotypes; my colleagues from the PhD school in Plant Biology and Production; all professors and staff at the Federal University Recôncavo of Bahia, where my academic journey began; my mother Eliana for being the best mom in the world; my father Edmilson who is far away, yet still present in my dreams; my sisters Carla and Evelyn for their encouragement and the strong feeling that we nurture what unites us; my grandmother Nina for understanding my absence and always giving me support, even while in tears when I had to leave; my wonderful family, my aunts, uncles and cousins, Meire, Joy and Euliane; my husband Jens for his patience, understanding, love, affection, companionship, and encouragement; my mother-in-law Ursula for treating me like a daughter; my best friend Tanilda Magalhaes for the 23 years of our friendship and for always believing in my choices in life; all my friends whom I left in Brazil, but have been so close at the same time, Alex, Diego, Olivia, Vania, Adriana, Dreide, Juliana and Soraya, who taught me that true

feelings of friendship are recognized and not conquered; the Brazilian community in Italy, my friends Gustavo, Rachel, Mirela, Humberto, Glaucia and Gislaine; last, but not least, my peach trees for making the study feasible. To all those who contributed directly or indirectly to achieving my dream, my sincere thanks.