UNIVERSITÀ DEGLI STUDI DI MILANO

Department of Agricultural and Environmental Sciences

School of Agriculture, Environment and Bioenergy

PhD Thesis

GENOMIC APPROACHES AND PHENOTYPIC ANALYSES FOR IMPROVING THE SELECTION OF FRUIT QUALITY TRAITS IN PEACH [*Prunus persica* L. (Batsch.)] AND APRICOT (*Prunus armeniaca* L.)

Tutor:
Dr. Anna Spinardi

Co-Tutor:
Dr. Marco Cirilli

PhD candidate:
Irina Baccichet

PhD Coordinator: Prof. Daniele Bassi

XXXIII PhD cycle

SUMMARY

LIST OF ABBREVIATIONS AND SYMBOLS

BLUE: Best Linear Unbiased Estimator

BLUP: Best Linear Unbiased Predictor

BV: Breeding Value

CIT: Citrate content trait

DM%: Dry matter in percentage

EBV: Estimated Breeding Value

FW: Fresh weight measured as grams (g)

GABA: Gamma-aminobutyric acid

G-BLUP: Genomic Best Linear Unbiased Predictor

GEBV: Genomic Estimated Breeding Value

GLS: Generalized Least Squares

GP: Genomic Prediction

GRM: Genomic Relationship Matrix (additive)

GRM_{194}: Genomic Relationship Matrix (additive) using the peach reference population

GRM_{796}: Genomic Relationship Matrix (additive) using only the peach seedlings of five crosses

GRM_{990}: Genomic Relationship Matrix (additive) using the total peach high-quality genotypes dataset

GS: Genomic selection

GWAS: Genome-wide association studies

h^2: Narrow-sense hereditability

HD: Harvest day (expressed as Julian Days)

HPLC: High Performance Liquid Chromatography

HTG: High-throughput genotyping

HTTP: High-throughput phenotyping

I_{AD}: Index of Absorbance Difference (represents the chlorophyll content)

LA: Low acid

LD: Linkage disequilibrium

LG: Linkage group

LMM: Linear Mixed Model

JD: Julian days

MAL: Malate content

MAP: Markers-assisted pyramiding

MARS: Markers-recurrent-assisted selection

MAS: Markers-assisted selection

MC: malate-citrate content ratio

MME: Mixed Model Equation

OA: Organic acid

PA: Prediction Ability

PACC: Prediction Accuracy

PBV: Predicted Breeding Value

PCA: Principal Component Analysis

PCoA: Principal Coordinate Analysis

PEP: Phosphoenolpyruvate enzyme

PEPC: Phosphoenolpyruvate carboxylase enzyme

PEV: Predicted Error Variance

QTL: Quantitative trait locus

ρ: Spearmans' correlation coefficient

REML: Restricted Maximum Likelihood

RR-BLUP: Ridge Regression Best Linear Unbiased Predictor

SSC: solids-soluble content

SsGBLUP: Single Step Genomic Best Linear Unbiased Predictor

SNP: Single nucleotide polymorphisms

TA: Titratable acidity (expressed as g/L of malic acid)

TBV: True Breeding Value

TCA cycle: Trycarboxylic acid cycle

$\hat{u}$: BLUP of the random effect inside the LMM

UHPLC-HRMS: Ultra-High-Performance Liquid Chromatography–High-Resolution Mass Spectrometry

LIST OF FIGURES

CHAPTER 5

LIST OF TABLES

CHAPTER 5

ABSTRACT

Peach [*Prunus persica* L. (Batsch.)] and apricot (*Prunus armeniaca* L.) are two *Prunus* species developing delicious fruits and they are mostly grown in temperate areas of the world. Both species have been cultivated since ancient times, being greatly appreciated for their unique taste and benefits to human health. Peaches and apricots can be consumed either fresh or processed through canning or drying, depending on the preference of a specific region and the use-purpose. So far, many efforts have been done in enhancing disease and pest resistance, in increasing the environmental adaptability to sub-tropical areas, in extending the harvest season or in improving fruits appearance. However, consumers frequently complain about the lack of taste in peach and apricot fruits sold on markets, encouraging the ongoing breeding projects to include organoleptic properties among their objectives. Among all the fruit quality-related traits, acidity plays a pivotal role affecting both consumers' and market acceptance. To match with market trends, an extensive characterization of peach and apricot fruits attributes under the highlighted necessity of renewing the varieties cultivated worldwide is required. In this thesis, a peach and apricot collection of 201 and 164 accessions, respectively, was screened for many important fruit organoleptic attributes, with a specific focus on acidity and organic acids content. Fruits acidity was titrated and ten organic acids (cis-aconitate, citrate, fumarate, galacturonate, malate, oxalate, quinate, shikimate, succinate and tartrate) separation was accomplished by HPLC technique coupled to UHPLC-HRMS validation. Analyses were performed on peach pulp and on apricot pulp and peel. The final aim was in-depth dissecting the peach and apricot panel for these organoleptic parameters to exploit the existing variability within their germplasm. Malate and citrate were the most abundant organic acids in both species, with pattern more genotype than year dependent. Results suggested that seasonality effects on fruits acidity and almost all the considered-organic acids were very low. Among the other organic acids, quinate and succinate reached large concentrations in peach and apricot, respectively, while tartrate was interestingly present more in apricot peel than pulp. The availability of a reference genome in peach has allowed to further characterize peach fruit quality traits. Over the past, peach fruits acidity has been partially elucidated at genetics level, identifying one major locus and making it a breeding target. Although the great advances, the selection accuracy and the long-time required for releasing new varieties on markets still hamper peach breeding progress. To in-depth unravel acidity trait and to speed up the selection of newly developed individuals in peach, a total of 1,190 accessions were genotyped for performing genomics analyses. Two different studies were carried out in this thesis: genomic selection (GS) and genome wide association studies (GWAS). Results confirmed that GS seems feasible in peach not only for acidity but also for organic acids content, in particular for patterns of malate and citrate. GWAS confirmed the presence of one major locus acting as dominant in peach fruits but revealed other significant associations on chromosomes 1, 7 and 8.

In summary, this thesis includes the first efforts in an in-depth and at multilevel dissection of acidity in peach and apricot, applying different approaches on a large panel of individuals. This thesis attempts to provide a complete overview with results that may be useful for future researches, studies and successful breeding programmes.

1. GENERAL INTRODUCTION

1.1 TAXONOMY, ORIGIN AND DOMESTICATION OF PEACH AND APRICOT

Peach [*Prunus Persica* (L.) Batsch] and apricot (*Prunus armeniaca* L.) are two species belonging to *Prunus* genus included in *Rosaceae* family (**Fig. 1.1** and **Fig. 1.2**, respectively). *Prunus* genus encompasses other species such as almond [*Prunus dulcis* (L.) Batsch], sweet cherry (*Prunus avium* L.), sour cherry (*Prunus cerasus* L.) and plum (*Prunus domestica* L. and *Prunus salicina* Lindl.). Peach and apricot develop stone fruits characterized by an epicarp (peel), a fleshy mesocarp (pulp) and a lignified endocarp (stone) protecting the unique seed. In peach, epicarp can be fuzzy or smooth (nectarines) conferred by a monogenic trait (*G locus* dominant for the presence of trichomes) mapped on chromosome 5 (Dirlewanger *et al.*, 2006). Both are climacteric. Peach and apricot species can be grown for ornamental purposes or collection of fruits, used for fresh consumption or industrial processing.

FIGURE 1.1. **Peach: from the left flowers, fruit-set stage and developed-fruit are depicted.** The pictures were taken in the peach germplasm collection maintained at the 'Mario Neri' farm of 'Centro Ricerche Produzioni Vegetali' (*CRPV*) located in Imola (Bologna, Italy).

FIGURE 1.2. **Apricot: from the left flowers, fruit-set stage and developed-fruit are depicted.** The pictures were taken in the apricot germplasm collection maintained at the 'Mario Neri' farm of 'Centro Ricerche Produzioni Vegetali' (*CRPV*) located in Imola (Bologna, Italy).

1.1.1 TAXONOMY OF PEACH

Peach is a diploid species (2n = 2x = 16) with a genome size of about 256 Mb. Self-compatibility, intense selection carried out by man over the century and vegetative propagation have reduced the peach genetic variability leading to a low heterozygosity.

Despite hundreds of different peach cultivars, the total wild relatives are reduced to five: *Prunus mira* (K.), *Prunus davidiana* (Carr.) Franch., *Prunus davidiana* var. *potatinii* (Rehd.), *Prunus kansuensis* (Rehd.) and *Prunus ferganensis* (Kost and Rjab.) Kov. and Kost. (Cao *et al.*, 2014). These five species do not set high-quality fruits but are often used as rootstock in order to increase drought tolerance in scions (for example *Prunus davidiana*) or to introgress pathogens resistance genes in commercial varieties (Bassi and Monet, 2008).

1.1.2 ORIGIN AND DOMESTICATION OF PEACH

Peach origin remains not fully unravelled. The ancient Persian area (largely corresponding to the actual Iran) was the first centre of origin proposed, justifying the botanical name given by Carl von Linnè in 1758 (Bassi and Monet, 2008). Recently, a newly proposed theory suggested that a selection process underpinned millions of years before the domestication performed by man was likely carried out by frugivore-species (Yu *et al.*, 2018). Domestication events dealt with fruit-quality attributes such as larger size, attractive skin colour and pleasant flavour. Peach and almond divergence from the common ancestor occurred about 4.99 million of years ago followed by the differentiation of *P. mira* (about 4.13 millions of years ago) and *P. davidiana* (3.47 millions of years ago). *P. kansuensis* speciation occurred about 2.47 millions of years ago from the same *P. persica* population (Yu *et al.*, 2018). The same study located the peach centre of origin in the southern-western of China and linked the divergence processes to several uplift events in the Tibetan area, also justified by the discovery of endocarp fossils (dated 8500-7500 years ago) in the Yangzi river valley (Zheng *et al.*, 2014; Su *et al.*, 2015; Yu *et al.*, 2018).

Human domestication process began around 4000-5000 years ago in the North-West of China while the peach cultivation was introduced in Europe at the ending centuries B. C. following the historic Silk Road (Faust and Timon, 1995). Then, European settlers introduced the peach cultivation in the Americas during the 16th century (Abbott *et al.*, 2002). Many landraces -most of all characterized by non-melting texture- originated in North America due to the seed propagation done by native populations (Bassi and Monet, 2008).

1.1.3 TAXONOMY OF APRICOT

Apricot (*Prunus armeniaca* L.) belong to the *Prunus* subgenus of *Prunophofora* and is diploid (2n = 2x = 16). Apricot has an estimated genome size of 220.36-220.56 Mb (Jiang *et al.*, 2019). Compared to peach, apricot genome is smaller but largely heterozygous (Jiang *et al.*, 2019) probably due to self-fertilization absence. Apricot can be inter-crossed with other *Prunus* species (i.e. plum) producing natural hybrids and increasing the existing variability (Zhebentyayeva *et al.*, 2012).

Apricot taxonomy is affected by different classification systems complicating the standardization. Six apricot species are widely recognized:

- *Prunus armeniaca* (L.) the main species for most of the apricot cultivars currently cultivated;
- *Prunus brigantina* (Vill.) commonly known as 'Briancon' apricot or 'Alpine plum';
- *Prunus holosericeae* (Batal.) known as 'Tibetan apricot';
- *Prunus mandshurica* (Maxim.) known as 'Manchurian apricot';
- *Prunus sibirica* (L.) known as 'Siberian apricot';
- *Prunus mume* (Sieb.) commonly called 'Japanese apricot' and originally cultivated only for ornamental purposes (Zhebentyayeva *et al.*, 2012).

1.1.4 ORIGIN AND DOMESTICATION OF APRICOT

Apricot origin and domestication have not been fully elucidated over the past centuries, despite of many efforts done. Vavilov identified three apricot origin centres based on the high germplasm diversity observed. The first centre was in mountain regions of central and western China. The second centre was placed in central Asia, corresponding to the Afghanistan, North-West of India, Pakistan, Kashmir, Tadjikistan, Uzbekistan, the Chinese province of Xeinjing and the western Thian-Shan mountainous chain. The last centre was identified in the areas of Caucasus and Asia Minor.

Apricot introduction in Europe probably followed several different routes (Faust *et al.*, 1998) and happened about 3000 - 4000 years ago. Then, the spread from Europe to the Americas occurred via the Atlantic Ocean while from China via the Pacific Ocean (Faust *et al.*, 1998).

1.2 BREEDING PROGRAMS IN PEACH AND APRICOT

1.2.1 OVERVIEW OF BREEDING STRATEGIES ADOPTED IN FRUIT-TREE CROPS

Over the past decades, traditional selection conducted by breeders mainly relied on phenotypic observations. The first reason concerns the lack of knowledge about the considered species' genome while the second regards the technologies unsuitableness to dissect the genetic architecture of target traits. Unlike genetic traits inherited in a discrete manner (Mendelian inheritance), complex or quantitative traits are hardly to characterize, complicating the selection. Quantitative trait loci (QTLs) define continuous phenotypic variation in target populations, being the genetic variance explained by interactions between genetic and non-genetic effects. Discrimination between genetic and non-genetic effects is difficult to obtained and time-consuming, affecting the genetic gain in the process of seedlings selection. Understanding the genetic mechanisms of target traits under polygenes' interactions could make the newly improved variety performances more stable over the years. Several strategies have been adopted to improve the selection process. Progeny testing found hampers in fruit-tree crops because parents' selection based on the evaluation of progeny performance. Progeny testing-based selection is time-consuming and requires a lot of

space and high cost to maintain all the seedlings, especially because marketable fruit quality attributes can be evaluated only after the plant reaches the physiological adult stage. Another selection-based strategy relies on the use of molecular markers associated with the character of interest. Molecular markers availability has increased over the last decades, becoming more affordable. Nowadays, single nucleotide polymorphisms (SNPs) are the most popular markers used for differentiating individuals. They are very abundant along the genome and can be detected by many multiplex genotyping technologies (Rafalski, 2002). SNPs are becoming more feasible, affordable and rapid. Many arrays have been developed for fruit tree crops including peach [*Prunus persica* (L.) Batsch] (Verde *et al.*, 2012), apple [*Malus* x *domestica* Borkh.] (Chagnè *et al.*, 2012; Bianco *et al.*, 2016), sweet cherry [*Prunus avium* L.] and sour cherry [*Prunus cerasus* L.] (Peace *et al.*, 2012), grapevine [*Vitis vinifera* L.] (Troggio *et al.*, 2007) and kiwifruit [*Actinidia chinensis* Planch] (Fraser *et al.*, 2009). This next-generation sequencing (NGS) technology is fast automatable projecting plant researchers to new and exciting study perspectives (i.e. genome-wide association studies, high-density genetic maps development and genomes evolution studies) (Iwata *et al.*, 2013) paving the way to the high-throughput genotyping (HTG) era. Genome-wide association studies (GWAS) use mapping populations carrying the target trait to explore the genome and find molecular markers associated to a chromosomal region at a significant statistical level. Obtaining molecular markers uniquely associated to target traits allows breeders to perform markers-assisted selection (MAS), markers-recurrent-assisted selection (MARS) and markers-assisted pyramiding (MAP). MAS and MARS strategies detect a reduced number of markers in linkage disequilibrium (LD) with the character of interest. Molecular markers are used for screening the individuals with the aim of finding at least one candidate parent for future crosses and thus introgressing the trait. MAS appeared more promising than progeny testing, accelerating the phenotypic selection and avoiding the need of keeping all the plants until the adult stage (Iwata *et al.*,2013). The cost of MAS in a breeding program depends on the trait and the capability of discriminating and evaluating the interesting locus effect (Jannink *et al.*, 2010). MAS finds main limitations in focusing on a reduced markers subset discarding the ones under the significant threshold, in hardly detecting the same QTLs across different environments (Crossa *et al.*, 2017) and in analysing large segregating populations. These aspects often decrease the possibility to detect all the significant marker-traits associations, becoming more evident in non-model crops as perennial fruit tree species (Jannink *et al.*, 2010; Resende *et al.*, 2012; Desta and Ortiz, 2014).

In this scenario, a newly developed approach - as genomic selection (GS; Meuwissen *et al.*, 2001; Heffner *et al.*, 2009)- appear less deficient than MAS and more powerful than other breeding strategies previously adopted in fruit-tree crops breeding.

1.2.2 STATE-OF-THE ART OF BREEDING PROGRAMMES IN PEACH

Peach breeding programmes of the past years mainly focused on fighting pathogen diseases as powdery mildew [*Sphaerotheca pannosa* (Wallr.: Fr.) Lev.] (Foulongne *et al.* 2003) and brown rot (*Monilinia fructigena* 'Honey' in Whetzel. and *Monilinia laxa* Aderh. and Ruhl.) (Adaskaveg *et al.*, 2008). However, fruit taste remains a bottleneck of the improved peach variety and one of the main consumers' complaints (Bassi and Selli, 1990; Cirilli *et al.*, 2016). Many efforts have been done in understanding the genetic architecture of fruit quality-related traits in order to improve the breeding strategy and moved the selection of novel peach accessions toward an improved fruit quality. Slow ripening (Ramming, 1991), flat shape (Dirlewanger *et al.*, 1998; Dirlewanger *et al.*, 1999; Bliss *et al.*, 2002; Guo *et al.*, 2002), flesh colour (Falchi *et al.*, 2013), freestone and melting texture (Peace *et al.*, 2005) have been already elucidated. Peach fruits trait as and low-acid content (Yoshida, 1970; Monet, 1979; Dirlewanger *et al.*, 1998) have still to be fully characterized, becoming a fertile soil for further investigations.

1.2.3 STATE-OF-THE ART OF BREEDING PROGRAMMES IN APRICOT

In the last decades, apricot breeding programmes mainly focused on seven aspects:

1) Pest and pathogens resistance with an attention to Sharka virus (*Plum Pox virus*, *PPV*), one of the greatest diseases in apricots production (Dondini *et al.*, 2011; Decroocq *et al.*, 2014; De Mori *et al.*, 2019).
2) Larger adaptability to different environments, mainly achieved through the improvement of varieties with low and high chilling requirements (Bassi and Audergon, 2006);
3) Minor attitude to fruit cracking, mostly due to rains when the fruit is close to be ripen;
4) Tree structure;
5) Low kernel seed bitterness (Cervellati *et al.*, 2012);
6) Self-fertility (Vilanova *et al.*, 2005);
7) Blooming time (Salazar *et al.*, 2013).

1.2.4 PHENOTYPING IN PEACH AND APRICOT

In recent years, progress in the application of genomic technologies has improved plant breeding efficiency enlarging the gap between genotype and phenotype. HTG has become more affordable and cost-effective while phenotyping still remains a bottleneck. Nowadays, guaranteeing the final success of breeding activities implies massive phenotyping. Many efforts have been made, mostly in annual crops, trying to solve this bottleneck. Fruit tree crops, such as peach and apricot, are not receiving the increasing interest by phenomics approaches. Deep phenotyping is costly and time consuming, therefore hardly affordable by public and private sectors. The main challenges of phenomics are to increase the quality of the collected data, improve non-destructive high-throughput phenotyping (HTP) tools and platforms, combine

simultaneous and automated analyses for exploring plants on large scale and implement new software for images acquisition and data processing. Unfortunately, at the moment, HTP-based studies have been carried out mainly on model organisms (Granier *et al.*, 2006; Jansen *et al.*, 2009; Arvidsson *et al.*, 2011; Granier and Vile, 2014) and annual crops such as wheat (*Triticum* spp.) (Boyle *et al.*, 2015; Holman *et al.*, 2016), maize (*Zea mays* L.) (Grift *et al.*, 2011; Montes *et al.*, 2011; Brichet *et al.*, 2017) and rice (*Oryza sativa*) (Yang *et al.*, 2013; Tanger *et al.*, 2017). HTP data can improve the accuracy of fruit tree crops selection for target traits providing high-quality data that are pivotal in developing prediction model in GS (Cobb *et al.*, 2013). The combination of HTP and approach as GS should improving fruit tree crops breeding by shortening breeding cycle length and the interval between generations, all leading to a higher genetic gain in fields (Meuwissen *et al.*, 2001; Heffner *et al.*, 2009).

1.3 PEACH AND APRICOT PRODUCTION IN ITALY

Peach and apricot cultivation has a long tradition in Italy, reaching production of 1,198,648 tonnes and 236,137 tonnes in 2018, respectively (ISTAT; www.dati.istat.it). Following Campania region (with 306,070 tonnes produced), Emilia-Romagna is the second largest peach producer in Italy, reaching 271,780 tonnes in 2018 (ISTAT). Emilia-Romagna highlands are important for apricot production with 62,711 tonnes in 2018 (ISTAT). *CRPV* (*Centro Ricerche Produzioni Vegetali*, www.crpv.it) is one of the leading cooperatives company of Emilia-Romagna region, mainly located in Cesena in the Po Valley (Northern Italy) (**Fig. 1.3**). Several activities are carried out at *CRPV*, including peach and apricot breeding.

FIGURE 1.3. **Peach and apricot fields at 'CRPV' farm.** The photos were taken in the peach (A) and apricot (B) germplasm collection maintained at the 'Mario Neri' farm of *CRPV* (*Centro Ricerche Produzioni Vegetali*) located in Imola (Bologna, Italy).

1.4 OBJECTIVES OF THE PhD THESIS

This PhD thesis is framed into apricot and peach breeding programs, in particular the Italian *MAS.PES* project in Northern Italy. Main *MAS.PES* objective is the introduction of apricot and peach accessions featuring with improved-organoleptic characteristics and pathogens resistance.

This thesis includes two main parts.

I. The first sections (i.e. *Chapter 2* and *Chapter 3*) concern the phenotypic characterization of fruit quality-related attributes along two years (seasons of 2017 and 2018) in a large peach and apricot collection grown at the *Centro Ricerche Produzioni Vegetali* (CRPV, *Crop Production Research Centre*) located near Imola (North-Eastern Italy).

II. The second part (i.e. *Chapter 4* and *Chapter 5*) includes the genetic dissection of acidity trait and ten organic acids profiles, all pivotal in determining the overall fruit quality and increasing consumers' satisfaction. Many accessions, selections and seedlings of peach (a total amount of 1,190 individuals) were collected and genotyped using the 18K SNP chip. Genetic architecture was investigated in peach through two strategies:

    1- Genomic selection (GS) or genome-wide selection approaches (*Chapter 4*);

    2- Genome wide association studies (GWAS) (*Chapter 5*).

2. PHENOTYPIC CHARACTERIZATION OF FRUIT QUALITY-RELATED ATTRIBUTES IN A LARGE COLLECTION OF PEACH ACCESSIONS

2.1 INTRODUCTION

Peach [*Prunus Persica* L. (Batsch)] is a perennial crop setting climacteric fleshy fruits commonly found on our table and important for their nutritional intake to human diet. Peach is widely cultivated in temperate regions since ancient times and is one of the most valuable crops in the world, with an achieved production of 24,453,425 tonnes in 2018 (FAOSTAT). China is the largest peach producer in the world with 15,217,797 tonnes. In Europe, Italy is the leading producer with 1,090,678 tonnes followed by Spain (with an achieved production of 303,809 tonnes) and France (with a total of 184,064 tonnes) in 2018 (FAOSTAT). However, the consumers' consumption has decreased in the last decade. Consumers are mostly disappointed for the low quality of fruits commercialized and generally picked before the maturity stage (Etienne *et al.*, 2002; Cirilli *et al.*, 2016). In the last years, many efforts have been done in releasing new or improved cultivars with larger fruit size, pleasing texture and attractive skin colour. However, consumers seem not fully satisfied and continuously criticize the lack of the characteristic "peach aroma" (Bassi and Selli, 1990; Cirilli *et al.*, 2016). In general, fruit quality is determined by several parameters as texture, skin colour, juiciness, volatile compounds and taste. Taste is strictly linked to water-soluble compounds (largely represented by mono- and di-saccharides and by organic acids content and profile), aroma to volatile molecules while bitterness or astringency depends on phenolic substances (Colaric *et al.*, 2005; Predieri *et al.*, 2006). Palatability and sweetness perception mostly depend on the sugars-acids balance at maturity stage, rather than on sugars alone (Colaric *et al.*, 2005; Batista-Silva *et al.*, 2018). Sugars (SSC) and organic acids (OAs) are different metabolites acting in carbon metabolism and in biosynthetic pathways for producing secondary compounds as amino acids (Batista-Silva *et al.*, 2018). Unlike soluble sugars - mainly translocated to "sink" organs through the phloematic system-, OAs are synthetized directly in the fruit (Etienne *et al.*, 2013), reaching higher concentrations in the first fruit growth stages (Bae *et al.*, 2014; Batista-Silva *et al.*, 2018). Then, the accumulated OAs become the preferred metabolic substrates during fruit respiration (Etienne *et al.*, 2002). A major classification of peach accessions is between low acid (LA) and acid, based on the fruit juice pH above or below 3.8 - 4.0, respectively (Yoshida, 1970; Dirlewanger *et al.* 1998). Low-acid peaches have SSC-OAs ratio almost up to 4 times higher than the acid ones (Monet, 1979; Moing and Svanella, 1998). LA peach varieties mainly characterize the Chinese market compared to acid or normal acids peaches widely spread on the

European one (Reimer, 1906). Low acidity in peach has been partially unravelled at genetic level with the discovery of *Locus D* (where *D* means "Doeux" in French) at the beginning of chromosome 5 (Monet, 1979; Dirlewanger *et al.*, 1998; Micheletti *et al.*, 2005). *Locus D* has been described as the major dominant locus controlling low acidity in peach and explaining the largest part of the observed variability for this fruit attribute. In addition to the overall peach fruits acidity, a great variability has been reported for OA content. OA patterns vary quantitatively and qualitatively among species and varieties. Peaches accumulate mostly malate and citrate in pulp cells vacuole, both weak OAs (Colaric *et al.*, 2005; Etienne *et al.*, 2013). Other OAs as shikimate, fumarate, succinate and ascorbate were reported in peach fruits, but all observed in trace amounts (Moing and Svanella, 1998; Etienne *et al.*, 2013). Several metabolic pathways seem to be involved in OAs synthesis, accumulation and storage. The first malate and citrate metabolic pathway occurs in the cytosol and regards the OAs synthesis from phosphoenolpyruvate (PEP) carboxylation. OAs accumulation into vacuoles follows a "acid-trap" mechanism based on malate-citrate dissociation property (Martinoia *et al.*, 2007). Malic and citric acid are dissociated in the cytosol (pH neutral) and protonated inside the vacuoles, generating an electrochemical potential gradient that drives the continuous OAs transport across membranes. However, different phosphoenolpyruvate carboxylase enzyme (PEPC) activity seems not related to malic acid concentration observed between low-acid and acid peaches (Moing *et al.*, 2000). The second pathway concerns the decarboxylation of OAs (malate and oxaloacetate) with the following activation of gluconeogenesis pathway during the fruit ripening (Sweetman *et al.*, 2009). This metabolic way reduces OAs content in mesocarp cells. The third proposed way consists the conversion of tricarboxylates into dicarboxylates through several possible cycles (i.e. tricarboxylic acid cycle in mitochondria, glyoxylate cycle for synthetizing flavonoids and isoprenoids or for converting into succinate glutamate and γ-aminobutyric acid) (Etienne *et al.*, 2013). This multiplicity of metabolic pathways suggested -and probably involved in OAs metabolism and catabolism- complicates the achievement of an in-depth characterization of these aspects, all pivotal in increasing the overall peach fruits quality.

In this chapter, a large peach germplasm collection was recorded for several fruit quality-related attributes, giving a particular attention to fruit acidity and OAs qualitative and quantitative profiles. This phenotypic dissection for titratable acidity (TA) and OAs content represents a first effort in characterizing 201 peach germplasm accessions in order to front the continuous attempts to cope with market demand. The achieved results may be used as a scaffold for further studies and for planning future breeding programmes. Surely future breeding programmes should include improved pests and diseases resistance and larger adaptability in different environments, but all these aspects should have a continuous glance at fruit organoleptic attributes, in order to guarantee the economic success for growers, breeders and retailers (Colaric *et al.*, 2005; Delgado *et al.*, 2013; Batista-Silva *et al.*, 2018).

2.2 MATERIALS AND METHODS

2.2.1 PLANT MATERIAL AND EXPERIMENTAL DESIGN

The considered peach germplasm collection is maintained at the *Centro Ricerche Produzioni Vegetali* (CRPV, *Crop Production Research Centre*, www.crpv.it) located near Imola (North-East of Italy). Peach trees are grown on *GF-677* rootstock. A total of 138 and 183 peach accessions were recorded for several fruit quality characters in season 2017 and 2018, respectively. Fruit weight, harvest day (expressed as Julian days), chlorophyll content, firmness, titratable acidity, dry matter (expressed in percentages), ten organic acids (i.e. oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) and SSC were considered. A total of 30 fruits in season 2017 and 20 fruits in season 2018 were randomly picked from different branches of two tree replicates for each genotype. Peach fruits harvesting lasted from the end of May to the beginning of September. Ripeness degree was visually evaluated and furtherly confirmed by chlorophyll content (I_{AD}) using the DA-meter portable device (Ziosi *et al.*, 2008). Peaches were harvested at the full maturity stage ("ready-to eat").

2.2.2 NON-DESTRUCTIVE ANALYSES OF PEACH FRUIT-QUALITY PARAMETERS

Several preliminary fruit attributes were recorded on the harvest day. Then, the following evaluations regarded titratable acidity (TA), OAs profiles and dry matter in percentage (DM%). Harvest day (HD) was recorded when fruits were picked and was expressed as Julian days (*JD*). Fresh weight (FW) was measured (g) for each fruit. Chlorophyll content (*Index of Absorbance Difference*, I_{AD}) was calculated as the average value read for each face of the intact fruit using a DA-meter portable spectrometer (*Sintéleia* S.r.l., Bologna, Italy). The instrument reads at two different wavelengths (680 nm and 720 nm) the absorbance peak (maximum at 680nm) of the bound between chlorophyll-A and chloroplast thylakoids. During peaches ripening, chlorophyll content decreases due to catabolic events (Gross, 1987) leading to lower I_{AD}.

2.2.3 DESTRUCTIVE ANALYSES OF PEACH FRUIT PARAMETERS

Firmness is measured as the fruit pulp resistance to the pressure executed by a constant rate digital penetrometer (Andilog Centor AC TEXT08). The penetrometer punctured one side of the fruit, after having removed a round area of the peel from the middle of the peach by a slicer. The analysis performance used a digital penetrometer equipped with a metal plunger of 8 mm of diameter, flat on the tip and set at 5 mms- 1 speed. Firmness was expressed in Newton (N). The total amount of sugars was estimated measuring the solid-soluble content (SSC% or °Brix) through a digital refractometer (Atago, Milan, Italy). SSC values were measured squeezing two peach pieces cut from the stalk to the bottom and from the exocarp to the ligneous endocarp in both fruit cheeks. DM% was calculated as the ratio between five equally-sized fruit pieces weighted before and after drying in an oven set at 60°C for 72 hours.

2.2.4 TITRATABLE ACIDITY ANALYSIS

TA analysis was performed only on fruit pulp because consumers frequently peel the peach before consuming it, especially when the fruit is fuzzy. After the total removal of the peach fruit peel, three biological juice replicates (50 mL for each replicate) were prepared from the remaining samples. The juices were made using a blender and stored in a refrigerator (-18°C). Once unfrosted, peach juices were centrifuged at 5000 rpm for 20 minutes at 4°C to precipitate the heavier particles and, then, 5 mL were diluted to 50mL with ultrapure water (18.2 MΩ.cm at 25°C). Peach juices acidity was measured using an automatic titration instrument (CRISON, Crison Instrument, Spain). By successive 0.1 N NaOH (Merck, KGaA, Germany) addition, each sample reached the equivalence point (pH 8.3). Peach acidity was expressed as g/L of malic acid following the formula:

$$TA\left({}^{g}/_{L}\, of\ malic\ acid\right) = (mL - b) * C_t * F_t * \frac{M}{V})$$

where *mL* is the volume of the titrant added, *b* is zero because the dilution was performed using water, C_t is the concentration of NaOH (0.1 N), F_t is a conversion factor equal to 1, *M* is the molecular weight (*MW*) of malic acid (67.05 g/mol) and *V* is the sample juice volume (5 mL).

2.2.5 DETECTION OF TEN ORGANIC ACIDS

Ten OAs detection (i.e. oxalate, cis-aconitate, citrate, galacturonate, tartrate, malate, quinate, succinate, shikimate and fumarate) was performed using the high-pressure liquid chromatography (HPLC) technique coupled to UHPLC-MS validation. HPLC analyses were carried out at a UV wavelength of 210 nm and each run lasted 30 minutes. The linearity was obtained by the injections of pure standards solutions (Fluka-Sigma-Aldrich; St. Louis, MO, USA) sequentially diluted from the stock solution. All standards were diluted in ultrapure water and filtered through a nylon membrane (0.45 μm). Retention time (t_r; **Table 2.1**) was calculated injecting the standard alone and then mixed with the others at different relative concentrations. A further validation for the OAs presence/absence was carried out injecting internal standards into some samples. A total of 100 μL of 0.5 % (w/v) Ethylenediaminetetraacetic acid (EDTA) were added to each mL of OAs standard and peach sample, to chelate calcium ions to avoid the interference with the column resin. From the stock juices, 2 mL were transferred into an Eppendorf® tube and centrifuged at 14,000 rpm for 15 minutes to remove any residual cell debris and particulate matter. The supernatant was filtered using a luer-lock nylon membrane syringe filters (0.45 μm, Chromacol, Welwyn Garden, UK) and 5 μL were injected into a Perkin Elmer series LC200 pump (Perkin Elmer, Norwalk, CT) equipped with a Jasko 975 UV/VIS detector (JASCO 28600, Mary's Court, Easton, MD). The analysis was performed using an Aminex HPX-87H cation exchange column (Bio-Rad Laboratories, Inc.) 300 mm x 7.8 mm i. d., 9 μm particle size, protected with a 40 mm x 4.6 mm Micro-Guard Cation H Cartridge. The analyses were performed at 65° C with (flow rate of 0.6 mL/min) using 4 mM H_2SO_4 as eluent under isocratic conditions. Chromatograms were acquired

and processed via Perkin Elmer's *Total Chrom Workstation 6.2*. The areas were converted in concentrations (ng/μL) using the calibration curves previously calculated.

TABLE 2.1. Organic acids considered in the analysis with the corresponding retention time.

Organic acid	**Retention time (minutes)**
Oxalic acid	6.33
Cis-aconitic acid	6.87
Citric acid	7.62
Tartaric acid	8.12
Galacturonic acid	8.18
Malic acid	9.11
Quinic acid	9.59
Succinic acid	11.17
Shikimic acid	11.51
Fumaric acid	13.07

Furthermore, the OAs patterns of some samples were validated qualitatively and quantitatively through the ultra-high-performance liquid chromatography–high-resolution mass spectrometry (UHPLC-HRMS) method. The tests were performed using an Acquity UHPLC separation module (Waters, Milford, MA, USA) coupled with a model Exactive Orbitrap MS through a HESI-II probe for electrospray ionization (Thermo Scientific, USA) set in negative ion mode. The OAs separation was carried out using -3.0 kV of spray voltage. The capillary and tube lens voltage were respectively of -27 V and -80 V while for the skimmer was -16 V. The sheath gas flow rate was 55 (arbitrary units) and the auxiliary gas flow rate was 15 (arbitrary units). The temperature for the heater and the capillary was set at 120°C and 320°C, respectively. A 1.8 μm HSS T3 column (150 x 2.1 mm, Waters) was used for separation at a flow rate of 0.45 mL/min. The eluents were 0.05% HCOOH in MilliQ-treated water (solvent A) and CH_3CN (solvent B). Five μL of the sample were separated by the UHPLC using the following elution gradient: 0 % B for 5 min, 0 - 80% B in 1 min, 80 % B for 3 min and then return to initial conditions in 1 min. The column and samples were kept at 40° C and 15° C, respectively. The UHPLC eluate was investigated in full scan MS in the range (m/z)- 50 - 1000 u. The resolution, AGC target, maximum ion injection time and mass tolerance were 50 K, 1E6, 100 ms and 2 ppm, respectively. The ion with m/z 91.0038 u, corresponding to the formic acid dimer [2 M - H]-, has been used as lock mass. The MS data were processed using *Xcalibur* software (Thermo Scientific, USA).

2.2.6 STATISTICAL DATA ANALYSIS

All statistical analyses performed on peach quality-related data used packages implemented in *R-Studio* software (version 3.6.3). Statistical analyses were carried out at first for each year and, then, averaging

fruit-quality related observations across years. TA values distribution was tested for normality using the Shapiro-Wilks method. Minimum and maximum values were reported for each OA observed in the analysed peach samples. Correlation matrix was built using Spearman's method because almost all the fruit quality attributes did not follow a normal distribution. Correlation matrix was calculated using *corrplot* package (version 0.84) in *R Studio*. Principal Component Analysis (PCA) was performed for representing in a bi-dimensional space the contribution of each fruit parameter to the total variability observed in peach germplasm collection. PCA analysis was carried out using *RStat* and *factoextra* package (version 1.0.7) in *R Studio*.

2.3 RESULTS

2.3.1 PEACH TITRATABLE ACIDITY

A total of 117 and 180 accessions were recorded for TA in the seasons 2017 and 2018, respectively. The total amount of unique accessions was 201 while 96 accessions were recorded for both years. Similar TA mean values were observed between the two years of evaluations on peach fruit quality. In harvest season 2017, TA values ranged from 1.27 g/L of malic acid in 'Ornella' to 11.16 g/L of malic acid in 'Blushing Star', with a mean value of 5.32 g/L of malic acid. In the harvest season 2018, the minimum TA value was 1.15 g/L of malic acid in 'Ornella' and the maximum was 12.52 g/L of malic acid in 'Romagna Bright', with the mean of 6.18. TA distribution and frequency for 201 unique peach genotypes were calculated (**Fig. 2.1** and **Fig. 2.2**). Biological replicates in each year and two years-replicated measurements for the 96 accessions were averaged. In the global peach germplasm collection considered, TA values ranged from a minimum of 1.18 g/L of malic acid in 'Ornella' to a maximum of 12.31 g/L of malic acid in 'Early Top'. Most of the peach accessions had 4-8 g/L of malic acid. TA data were tested for normality using Shapiro-Wilk normality test. TA data distribution seemed not to follow the normality (*p-value* of 0.012), probably justifying the bi-modal density curve reported in TA histogram.

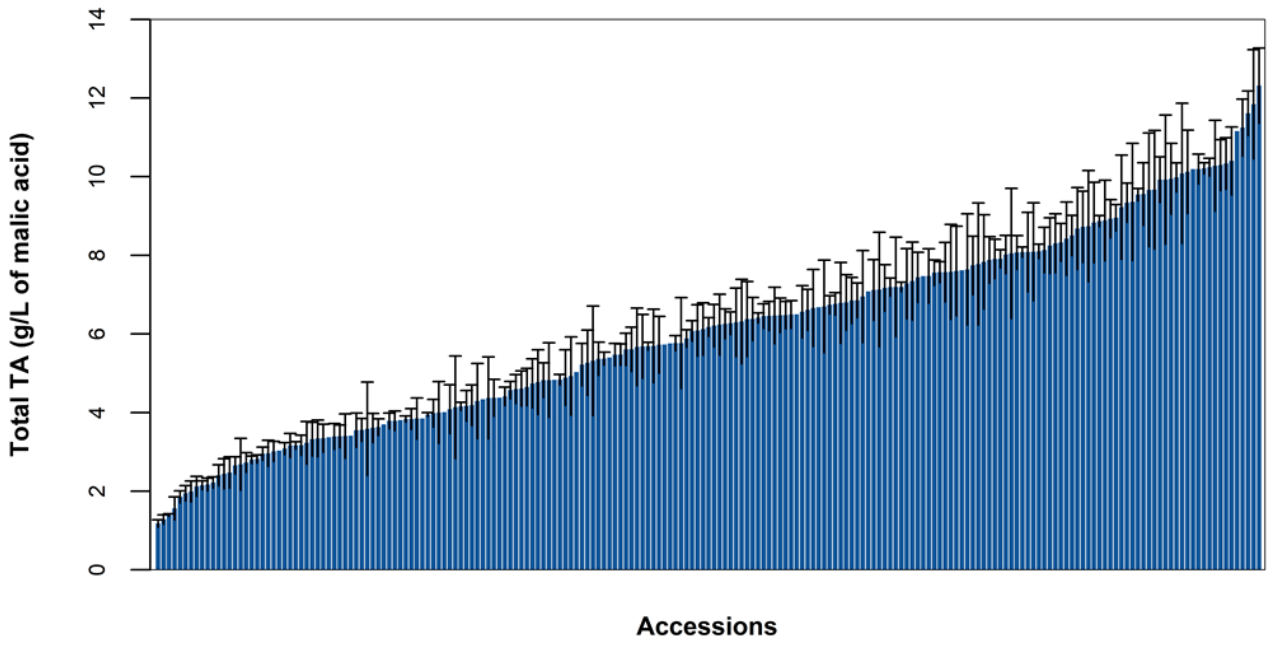

FIGURE 2.1. **TA values (expressed as g/L of malic acid) in peach germplasm collection (201 unique accessions).** Biological replicates and two years-replicated measurements were averaged. Vertical bars represent the standard deviation calculated in each peach accession. In the peach germplasm collection considered, TA values ranged from a minimum of 1.18 g/L of malic acid in 'Ornella' to a maximum of 12.31 g/L of malic acid in 'Early Top'.

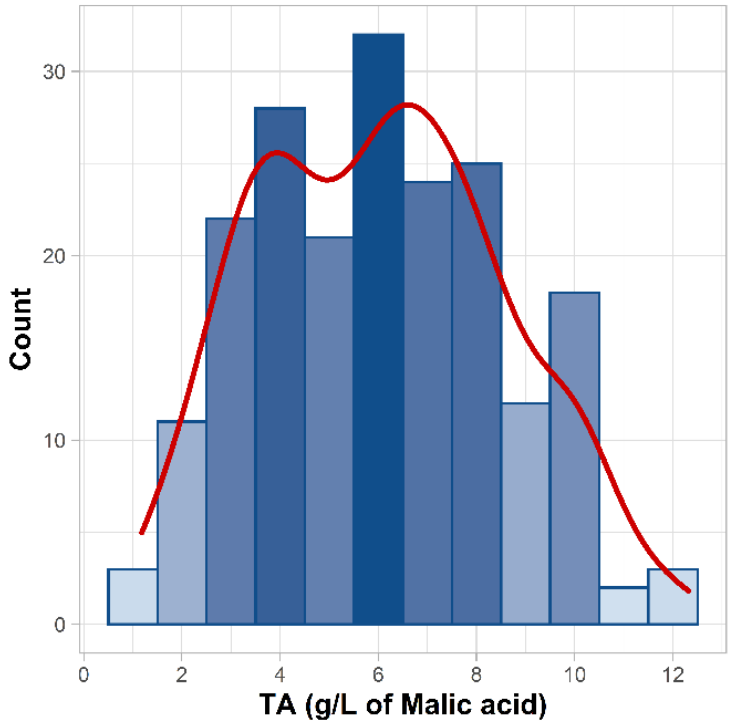

FIGURE 2.2. **TA values (expressed as g/L of malic acid) frequency in the studied peach germplasm collection (201 unique accessions).** Biological replicates and two years-replicated measurements were averaged. Most of the peach accessions were in the range between 4 and 8 g/L of malic acid. TA data distribution seemed to not follow the normality (*p-value* of 0.012) basing on the result of Shapiro-Wilk's test.

2.3.2 ORGANIC ACID PROFILES IN PEACH FRUITS

Each OA content (ng/μL) was calculated from the peach chromatograms obtained (**Fig. 2.3**), with minimum, mean and maximum values (**Table 2.2**). OAs qualitative and quantitative profiles were variegated among peach accessions, although malate and citrate were the most abundant. Malate was the predominant OA, reaching a higher concentration than citrate with the exception in 'BO05030133', 'BO09001134', 'BO05021034' and 'BO0904003' where the citrate-malate ratio observed was 1.18, 1.10, 1.06 and 1.04, respectively. Malate minimum content was 608.69 ng/μL in 'Xia Hui' while the maximum concentration was observed in 'Diamond Ray' with 8977.43 ng/μL. The lowest citrate amount was found in 'Bolivia' (49.27 ng/μL) while the largest was in 'BO05021034' (5861.29 ng/μL). Fumaric acid was greatly present in the peach collection, although not large concentrations were reached. In fact, fumarate content ranged between 1.66 ng/μL (in 'BO99024032') and 26.78 ng/μL (in 'IFF143'). Among OAs generated through trycarboxylic acid (TCA) cycle, succinate reached a large concentration of 1296.75 ng/μL in 'Fei Cheng Bai Li' while oxalate and cis-aconitate were most abundant in 'Carota' with 55.70 ng/μL and 'Michelini' with 71.19 ng/μL, respectively. Galacturonate, tartrate and shikimate were mostly present in traces, reaching maximum amount of 1552.43 ng/μL in 'Turquoise', 112.61 ng/μL in 'Bolivia' and 24.59 ng/μL in 'BO06006070', respectively. Among OAs not produced through TCA cycle, quinic acid was largely present with a maximum content of 3543.90 ng/μL in 'Bolina'. Among the three most frequently observed OAs in the peach collection (i.e. malate, citrate and fumarate), only malate seemed to follow a normal distribution (*p-value* of 0.12).

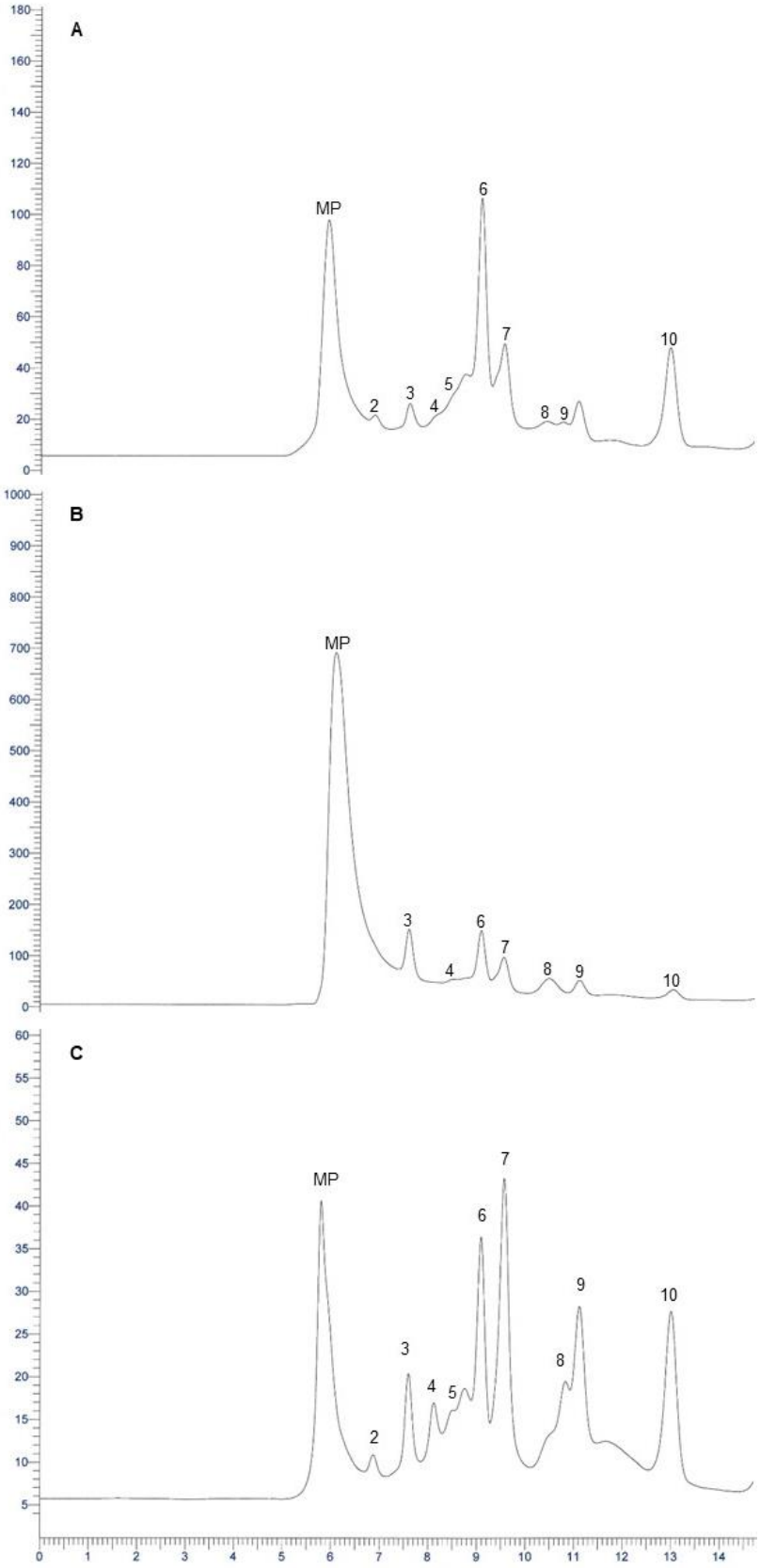

FIGURE 2.3. **Chromatograms of 'Ornella' (A), 'Bolina' (B) and 'Xia Hui' (C) reporting their OA profiles.** Each peak represents an OA identified by the comparison with the standards retention time (t_r) and confirmed by spiking standard stock solution into some samples. Based on t_r, the peaks reported are: *MP*, mobile phase; *1*, oxalic acid (not present in the three accessions); *2*, cis-aconitic acid; *3*, citric acid; *4*, tartaric acid; *5*, galacturonic acid; *6*, malic acid; *7*, quinic acid; *8*, succinic acid; *9*, shikimic acid and *10*, fumaric acid.

TABLE 2.2. OAs content (ng/μL) detected through the HPLC analysis in the 201 unique peach accessions.

Organic acid	Minimum value (ng/μL)	Maximum value (ng/μL)	Mean (ng/μL)
Oxalic acid	0	55.70	0.29
Cis-aconitic acid	0	71.19	1.33
Citric acid	49.27	5861.29	1521.21
Tartaric acid	0	112.611	6.38
Galacturonic acid	0	1552.425	49.73
Malic acid	608.69	8977.43	4173.48
Quinic acid	0	3543.90	783.17
Succinic acid	0	1296.75	183.79
Shikimic acid	0	24.59	7.85
Fumaric acid	1.66	26.78	7.58

2.3.3 CORRELATION TEST AMONG FRUIT-QUALITY ATTRIBUTES

Correlations (Spearman's method) were tested among the fruit quality attributes and OAs pattern in the peach collection (201 accessions), at first between each year (**Table 2.3** and **Table 2.4**) and then across years (**Fig. 2.4**). FW, I_{AD}, firmness, TA, SSC, DM% and HD were significantly correlated between years. In particular, TA and HD had the highest positive correlations (correlation coefficient of 0.86 and 0.94, respectively) suggesting a greater genotype-dependence than a seasonality influence. Almost all OAs were correlated across years. The presence in trace amounts of some OAs as succinate and tartrate could explain the absence of strong correlation coefficients across years (correlation of 0.05 and 0.17, respectively). Correlations across years were tested averaging replicated measurements for each peach genotype. TA strongly positive correlated with malate (ρ = 0.82), citrate (ρ = 0.80) and less with cis-aconitate (ρ = 0.29) and succinate (ρ = 0.31). On the other hand, peach acidity seemed negatively affected by fumarate content (ρ =-0.66). Malate concentration increased with larger content of citrate (ρ = 0.54) cis-aconitate (ρ = 0.40) and succinate (ρ = 0.43) but decreased with the presence of fumarate (ρ = -0.54). Fumarate was positively related to shikimate (ρ = 0.35), that was negatively affected by larger amount of oxalate (ρ = -0.22) and malate (ρ = -0.13). However, oxalic acid occurred in trace amount in the peaches analysed. Galacturonate concentration seemed likely independent from the presence or absence of the other OAs being negatively related only with HD (ρ = -0.62) and DM% (ρ = -0.33). Tartrate content was larger when high quinate concentration were found (ρ = 0.38). SSC showed strong correlations with DM% (ρ =0.83) and HD (ρ =0.86) likely because sugars represent the largest amount of the soluble solids at the end of the fruit ripening.

TABLE 2.3. Correlation test among fruit quality attributes recorded in each harvest season (2017 and 2018) for the peach collection. Correlations were significant '*' at the 5% level, '**' at the 1% level, '***' at the 0.1% level. *TA*, titratable acidity; *FW*, fresh weight (g); *HD*, harvest day (as Julian days); I_{AD}, chlorophyll content index; *SSC*, soluble-solids content.

	TA (2018)	**FW** (2018)	**HD** (2018)	**IAD** (2018)	**DM%** (2018)	**Firmness** (2018)	**SSC** (2018)
TA (2017)	0.86***						
FW (2017)		0.57***					
HD (2017)			0.94***				
I_{AD} (2017)				0.32**			
DM% (2017)					0.62***		
Firmness (2017)						0.57***	
SSC (2017)							0.77***

TABLE 2.4. Correlation test among OAs content recorded in each harvest season (2017 and 2018) for the peach collection. Correlations were significant '*' at the 5% level, '**' at the 1% level, '***' at the 0.1% level. Correlations were not significant ' '.

	Malate (2018)	**Citrate** (2018)	**Fumarate** (2018)	**Cis-aconitate** (2018)	**Tartrate** (2018)	**Galacturonate** (2018)	**Quinate** (2018)	**Succinate** (2018)	**Shikimate** (2018)
Malate (2017)	0.58***								
Citrate (2017)		0.64***							
Fumarate (2017)			0.59***						
Cis-aconitate (2017)				0.23*					
Tartrate (2017)					0.17				
Galacturonate (2017)						0.51***			
Quinate (2017)							0.53***		
Succinate (2017)								0.05	
Shikimate (2017)									0.53***

As expected, TA and SSC were negatively correlated (ρ =-0.22). SSC weakly decreased with higher amounts of galacturonate (ρ =-0.40), tartrate (ρ =-0.29), citrate (ρ=-0.28), oxalate (ρ =-0.25), succinate (ρ =-0.22), malate (ρ =-0.17), cis-aconitate (ρ =-0.17) and quinate (ρ =-0.14).

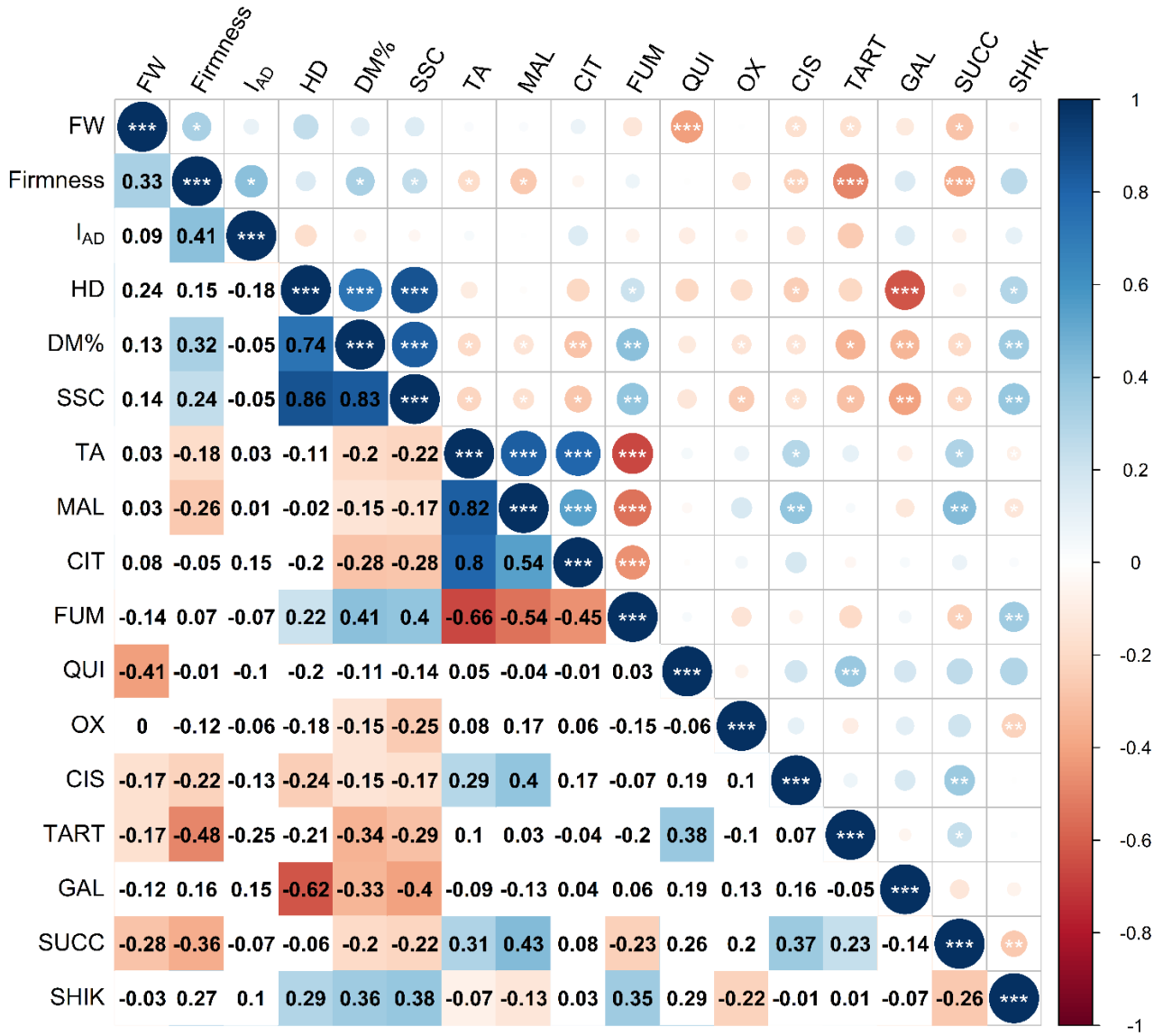

FIGURE 2.4. **Correlation test among fruit quality attributes in the peach collection.** Biological replicates and replicated measurements were averaged across years (2017 and 2018). A total of 201 peach accessions were considered in investigating correlations among fruit quality attributes. Parameters were significantly correlated '*' at the 5% level, '**' at the 1% level, '***' at the 0.1% level. Correlations were not significant ' '.*FW*, fresh weight (g); I_{AD}, chlorophyll content index; *HD*, Harvest day; *DM%*, Dry matter in percentage, *SSC*, soluble-solids content and *TA*, titratable acidity; *MAL*, malate content; *CIT*, citrate content; *FUM*, fumarate content; *QUI*, quinate content; *OX*, oxalate content; *CIS*, cis-aconitate content; *TART*, tartrate content; *GAL*, galacturonate content; *SUCC*, succinate content; *SHIK*, shikimate content.

2.3.4 PCA ANALYSIS

PCA analysis (**Fig. 2.5**) was carried out among fruit quality attributes in the peach collection. The first two principal components (i.e. *Dim1*, *Dim2*) captured 35.8% of the total variability observed in peach accessions (96 peach genotypes). *Dim1* was scarcely affected by firmness and I_{AD}, while it was more dependent on TA, citrate, malate and fumarate content. TA, malic acid and citric acid were grouped together on the bottom of PCA plot confirming the higher correlations found. *Dim2* had stronger associations with DM%, SSC and HD while was less affected by FW, quinate, cis-aconitate, tartrate, galacturonate, shikimate and succinate. TA, malate and citrate characterized more 'Romagna Bright' and 'BO99014002' while DM%, SSC and HD were more represented in 'Honey Royal', 'Romagna Gold', 'Zephir' and 'Oriane'. 'BO05041018' and 'Turquoise' were more prominent in fumarate and quinate content, respectively.

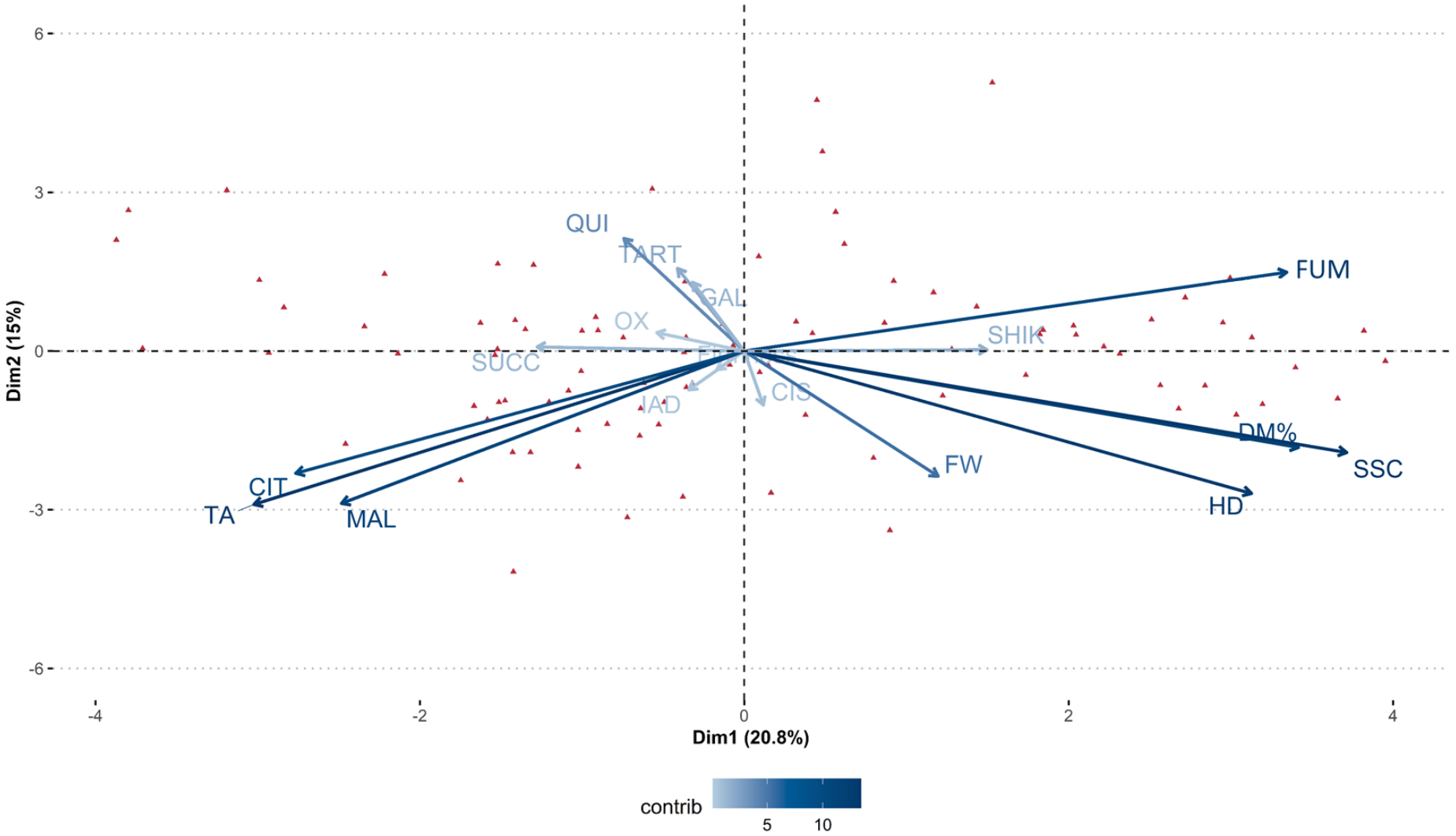

FIGURE 2.5. **PCA plot showing the relation among fruit quality traits.** Red points are peach accessions recorded for both years (96 unique genotypes). Colours become lighter when the variable contribution is lower. *DM%*, Dry matter in percentage; *FW*, fresh weight (g); *HD*, harvest day (expressed as Julian days); *IAD*, chlorophyll content index; *SSC*, soluble solids content (°Brix); *TA*, titratable acidity (g/L of malic acid); *CIS*, cis-aconitate content; *CIT*, citrate content; *FUM*, fumarate content; *GAL*, galacturonate content; *MAL*, malate content; *OX*, oxalate content; *QUI*, quinate content; *SHIK*, shikimate content; *SUCC*, succinate content; *TART*, tartrate content.

2.4 DISCUSSION

A large peach collection was in-depth characterized through several approaches to acquire a complete knowledge about several fruit quality parameters related to overall peach taste and consumer's satisfaction. A large variability was observed among different peach accessions, useful for planning future breeding programmes. Sugars seems to be the driving factor in the consumer's sweetness perception (Colaric *et al.*, 2005; Delgado *et al*., 2013). However, sweetness is also influenced by acids content (Colaric *et al.*, 2005; Delgado *et al.*, 2013). Therefore, acidity becomes an important indicator of peaches quality (Esti *et al.*, 2002; Harker *et al.*, 2002; Cantìn *et al.*, 2010; Etienne *et al.*, 2013). Peach acidity seemed to follow a bimodal distribution (**Fig. 2.2**), feature of quantitative trait loci (QTL) distribution found in fruit tree crops. Acidity content inversely correlated with DM%, probably depending on the dilution effect related to mesocarp cells enlargement during peach growth (Ruffner, 1982; Famiani *et al.*, 2005; Walker *et al.*, 2011; Famiani *et al.*, 2016). TA in the peach collection seemed more genotype- than year-dependent, being stable between years (seasons 2017 and 2018). OAs composition in peach affect the final taste perception in mouth. OAs concentration decreases along with the fruit ripening process, being used as metabolic substrates (Giovannoni *et al.*, 2004; Wu *et al.*, 2003; Etienne *et al.*, 2013; Batista-Silva *et al.*, 2018). According with previous studies, malate and citrate were the most abundant OAs observed in the peach collection (Moing and Svanella, 1998; Etienne *et al.*, 2002; Bureau *et al.*, 2013) and their ratio was measured being associated with the taste perception (Colaric *et al.*, 2005). Only four selections (i.e. 'BO05030133', 'BO09001134', 'BO05021034' and 'BO0904003') had larger content of citrate than malate. Malate and citrate were the major contributors to TA in peach collection, based on correlations observed (ρ of about 0.80) and PCA results. In disagreement with previous works (Génard *et al.*, 1994; Wu *et al.*, 2003), malate was positively correlated to citrate content ($\rho = 0.54$). Fumarate was the third more present acid in the peach collection, although without reaching high concentrations (range of 1.66- 26.78 ng/μL). Among the OAs produced by TCA cycle, succinate was abundant (with a maximum of 1296.75 in 'Fei Cheng Bai Li') but not detected in all peaches. Oxalate and cis-aconitate were detected in traces in almost all peach accessions, reaching high amounts only in 'Carota' (with 55.70 ng/μL) and 'Michelini' (with 71.19 ng/μL), respectively. Among OAs not produced by TCA cycle, quinate was abundant although not present in all accessions. Tartaric acid generally represents other species as grapes, but low amounts were observed also in the peach collection, with a maximum content in 'Bolivia' (112.61 ng/μL). Galacturonate and shikimate are precursors of pectin and aromatic amino acids (Maeda and Dudareva, 2012), respectively, and both were found in traces in the peach collection.

2.5 CONCLUSIONS

Collecting data on fruit-quality-related attributes is time-consuming and labour intense but becomes necessary in breeding programmes. The narrow diversity of recently released peach varieties on the market

has hampered the effort in increasing consumers' satisfaction. This study has revealed a large diversity for fruit acidity in the peach collection, representing a promising pool of phenotypic records to support next quality-oriented breeding programmes. Malate and citrate were the major contributors to the overall peach acidity, with profiles more dependent on genotype. Further sensory tests are not excluded in order to detect possible relationships among OAs patterns and consumer's degree of liking. A large peach dissection is required for updating the fresh fruit market and guarantying the economic success, both for growers and retailers.

2.6 PUBLICATION

Scientia Horticulturae 278 (2021) 109865

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Characterization of fruit quality traits for organic acids content and profile in a large peach germplasm collection

Irina Baccichet [a], Remo Chiozzotto [a], Daniele Bassi [a], Claudio Gardana [b], Marco Cirilli [a,*], Anna Spinardi [a,*]

[a] Department of Agricultural and Environmental Sciences (DISAA), University of Milan, Milan, Italy
[b] Department of Food, Environmental and Nutritional Science (DeFENS), University of Milan, Milan, Italy

ARTICLE INFO

Keywords:
Acidity
Consumers' acceptance
Fruit quality
Peach germplasm
Prunus persica
Organic acids

ABSTRACT

Organoleptic properties play a pivotal role in determining peach fruit quality, affecting both consumers' and market acceptance. In spite of a narrow genetic diversity, peach [*Prunus persica* L. (Batsch.)] has a remarkable range of fruit taste, mainly driven by the relative amount of sugars and acids. Characterization and exploitation of such variability is a major objective of fruit-quality orientated breeding programmes. In this study, a peach collection of 201 accessions was dissected for important fruit quality traits, with a particular focus on acidity and organic acids content. Fruit acidity was titrated and ten organic acids (oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) were detected through HPLC and UHPLC-MS. Malate and citrate are the most abundant accounting for the 62 % and the 22.6 % of total organic acids, respectively, and suggesting their largest contribution to the overall peaches acidity. Results reveal also a genotype-dependent contribution of specific organic acids and a low seasonality-effect on peach acidity and organic acid patterns. Collectively, this work provides an overview of the phenotypic variation associated with organic-acid related traits useful for supporting the ongoing breeding works.

Baccichet I., Chiozzotto R., Bassi D., Gardana C., Cirilli M. and Spinardi A. (2021). Characterization of fruit quality traits for organic acids content and profile in a large peach germplasm collection. *Scientia Horticolturae* 278.

3. PHENOTYPIC CHARACTERIZATION OF FRUIT-QUALITY-RELATED ATTRIBUTES IN A LARGE COLLECTION OF APRICOT ACCESSIONS

3.1 INTRODUCTION

Despite the molecular markers' availability and powerful phenotyping tools for developing new strategies in selecting novel or improved genotypes, apricot (*Prunus armeniaca* L.) market relies on a narrow range of cultivars (Bassi and Selli, 1990; Bassi and Audergon, 2006). Fruit organoleptic characterization remains the bottleneck in apricot breeding programmes, one of the most versatile and delicious temperate fruit tree crops grown since ancient times (Faust *et al.*, 1998; Gurrieri *et al.*, 2001; Ruiz *et al.*, 2010). Apricot is mainly cultivated in the Mediterranean area that accounts for more than 50 % of the global production (FAOSTAT). The larger producers in 2018 were Turkey (750,000 tonnes), Algeria (242,243 tonnes), Italy (229,020 tonnes) and Spain (176,289 tonnes). Apricots can be consumed fresh or processed through drying, canning and juicing, making necessary a deep characterization of apricot genotypes for satisfying different market demands (Bassi and Selli, 1990; Crivelli, 1995; Bassi and Audergon, 2006). Consumers commend apricot fruits aspect with their round to little oblong shape and velvety skin but complain about the lack of pleasing taste and aroma at the retail customer's end (Gurrieri *et al.*, 2001; Ledbetter *et al.*, 2006). Consumers are willing to pay more if high fruit quality is guaranteed and the product expectations are satisfied (Lange *et al.*, 2000; Harker *et al.*, 2002; Ledbetter *et al.*, 2006). Apricot fruit quality minimum standards vary greatly among country being subjected to consumer's preferences and targeted markets (Harker *et al.*, 2002). Flavour, skin colour, firmness, fruit size with absence of flesh adhesion to the pit, fruit cracking and gummy texture are only few pomological determinants of overall apricots quality (Souty *et al.*, 1990; Génard *et al.*, 1994; Bassi and Audergon, 2006; Ruiz and Egea, 2008). Moreover, fruit quality is directly influenced by tree growth and load, genotype-by-environment interaction, seasonality, pruning, grafting and other agronomical practices usually adopted in orchards (Laurens *et al.*, 2000; Miller and Scorza, 2010). Early fruits harvesting, long distance shipment and post-harvest handling affect the market acceptability, the consumer's behaviour in repurchasing, competitiveness on the market and the final commercial value (Bassi and Selli, 1990; Audergon *et al.*, 1991*a*; Bassi *et al.*, 1996; Bassi and Audergon, 2006; Sadar *et al.*, 2016; Xi *et al.*, 2016). Beside to exalting apricot aroma, mostly related to β-ionone and γ-decalactone volatile compounds (Gurrieri *et al.*, 2001; Kader, 2008; Xi *et al.*, 2016; Zhang *et al.*, 2019), balanced sugars (or soluble-solid content, SSC) and total organic acids (OAs) content at harvest time is pursued (Bassi and Selli, 1990; Bassi *et al.*, 1996; Borsani

et al., 2009; Bae *et al.*, 2012). Sugars and OAs are related and equally contribute to the consumers' quality perception, becoming an index of ripeness and the driving factors of purchase (Palmer and List, 1973; Souty *et al.*, 1990; Audergon *et al.*, 1993; Bartolozzi *et al.*, 1997; Fan *et al.*, 2017; Zhang *et al.*, 2019). SSC-acidity ratio was described as very variegated in the cultivars considered ranging from 1.9 to 3.5 (Bassi and Selli, 1990). Many low-acid apricot genotypes are grown in Turkey, leader country in drying apricot, where the top rated cultivar 'Hacihaliloglu' has a total SSC/acidity content of 20-25 % and 0.5 %, respectively (Akca and Askin, 1995; Esitken and Guleriuz, 1995; Ayanoglu and Kaska, 1995; Bassi and Audergon, 2006). In apricot, acidity may be a fertile soil for further investigations. Unlike peach, no major locus controlling fruit acidity has been detected yet, although the phenotypic distribution in progenies showed a continuous trend, peculiarity of other fruit-related quantitative trait loci (QTLs) (Bassi and Negri, 1991; Bassi *et al.*, 1996; Bartolozzi *et al.*, 1997; Bassi *et al.*, 2010; Ruiz *et al.*, 2010). Acidity mainly depends on the balance between biosynthesis, decomposition and accumulation of OAs in fruit mesocarp cells (Xi *et al.*, 2016; Zhang *et al.*, 2019). OAs are transported into vacuoles and stored in larger concentrations than in the cytosol through specific proton pumps activity as V-ATPase (Ma *et al.*, 2017; Zhang *et al.*, 2019). Malate, citrate and secondly quinate have been detected as the most abundant OAs in apricot flesh and peel accounting for more than 95% of total OAs at fruit maturation stage (Bassi *et al.*, 1996; Gurrieri *et al.*, 2001; Hasib *et al.*, 2002; Xi *et al.*, 2016; Fan *et al.*, 2017), similar to peach fruit (Moing *et al.*, 1998). Knowledge about the genetic mechanisms underlying OAs synthesis, accumulation and storage in apricot is still scarce. However, it has been reported that apricot has three copies of citrate synthase -compared to two found in other *Prunus* species- highly expressed during the ripening process (Jiang *et al.*, 2019). An aluminium-activated malate transporter (*ALMT9*) is related to low acidity content when down-regulated in apricot (Zhang *et al.*, 2019). Titratable acidity (TA) and OAs accumulation content are larger at the early stage of fruit growth and decrease at the fruit full-ripen stage, since they are used as respiratory substrates (Etienne *et al.*, 2002; Bae *et al.*, 2012; Xi *et al.*, 2016). Differences in OAs content and profiles were observed between pulp and peel during apricot fruit ripening (Xi *et al.*, 2016). OAs pattern likely vary among apricot varieties influencing consumers' sensorial perception and seems to be dependent on genotype and ripening stage (Guichard and Souty, 1988; Audergon *et al.*, 1990; Audergon *et al.*, 1991*b*; Bassi and Selli, 1990; Bassi *et al.*, 1996; Hasib *et al.*, 2002; Sadar et al, 2016). OAs content was observed to be more diverse than in peach. Malate/citrate content ratio is extremely diversified ranging from 0.2 to 8.8 (Gurrieri *et al.*, 2001) and apricot genotypes accumulate selectively malic or citric in mesocarp cells (Bassi and Selli, 1990). Galacturonate, quinate and tartrate were abundantly detected in apricot fruit flesh followed by minor quantities of succinate, shikimate, oxalate and fumarate (Bartolozzi *et al.*, 1997; Gurrieri *et al.*, 2001; Hasib *et al.*, 2002; Bae *et al.*,2012).

This work dissected a large collection of apricot accessions and selections for fruit acidity separately in pulp and peel because consumers eat the whole fruit without peeling it. Therefore, both apricots pulp and

peel tissues can contribute individually and differently to the final acidity perception. Apricot genotypes revealed similar TA records and almost similar OAs profiles between pulp and peel, albeit some exceptions of acidity higher in pulp than in peel -and *viceversa*- occurred. The rich diversity found in apricot collection can pave the way for developing genotypes suitable to specific processing uses, to face the bland taste frequently reported by consumers and for improving the breeding efficiency.

3.2 MATERIALS AND METHODS

3.2.1 PLANT MATERIAL AND EXPERIMENTAL DESIGN

The collection of apricot accessions and selections is maintained at the *Centro Ricerche Produzioni Vegetali* (CRPV, *Crop Production Research Centre*, www.crpv.it) located near Imola (North-East of Italy). Apricot trees are grown on *Mirabolan 29C* rootstock. Ten uniform apricot fruits were randomly picked at full maturity stage ("ready-to-eat") from 94 and 128 genotypes in the seasons 2017 and 2018, respectively. A total of 164 unique genotypes were included in the apricot germplasm collection. Fruits of each accession and selection were recorded for fruit fresh weight (FW), harvest day (HD), chlorophyll content, firmness, fruit dry matter (DM) and soluble-solids content (SSC). Fruit pulp and peel of each genotype were separately characterized for titratable acidity (TA) and 10 organic acids (oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate).

3.2.2 NON-DESTRUCTIVE ANALYSES OF APRICOT FRUIT-QUALITY PARAMETERS

Several non-destructive analyses were carried out before collecting data on titratable acidity (TA) and OAs profiles of fruit pulp and peel separately. HD corresponded to the fruit picking date and was expressed as Julian days (*JD*). Individual fruit FW was determined in grams (g) using a precision scale. The chlorophyll content (*Index of Absorbance Difference*, I_{AD}) was calculated as the average value read for each fruit cheek using a DA-meter portable spectrometer (*Sintéleia* S.r.l., Bologna, Italy).

3.2.3 DESTRUCTIVE ANALYSES ON APRICOT FRUIT-QUALITY PARAMETERS

Firmness was measured by a constant rate digital penetrometer (Andilog Centor AC TEXT08) test after having removed a 1.5 cm round area of the peel from the middle of both fruit faces by a slicer. The penetrometer was equipped with a flat metal plunger (8 mm) for 1 cm puncture and motorized by a basic test stand (BATDRIVE) set at 5 mm/s of speed. Firmness was expressed in Newton (N). DM content percentage (DM%) was obtained cutting five equal-sized pulp pieces randomly selected for each genotype. DM% data were calculated as the ratio of sample weight (g) before and after oven-drying at 60°C for 72 hours. Sugars were determined as SSC -being described as highly correlated (Dirlewanger *et al.*, 1999; Gurrieri *et al.*, 2001)- from the three biological replicates of pulp juices prepared for each apricot genotype to subsequently determine TA and OAs profiles. SSC readings were performed using a digital refractometer

(Atago, Milan, Italy), after pulp juices centrifugation at 5000 rpm for 20 minutes at 4°C for precipitating the heavier particles. SSC values were expressed as °Brix.

3.2.4 APRICOT FRUIT TITRATABLE ACIDITY ANALYSIS

TA determination was carried out preparing three biological replicates of juice of fruit pulp and peel separately. After the evaluation of FW, chlorophyll content, firmness and DM%, all the fruit peel was carefully removed from the samples. Four grams of fruit peel for each replicate were diluted 1:10 in bi-distilled water and mixed by a blender. Fruit pulp replicates (50 mL each) were made using a blender after the complete removal of the peel. After centrifugation at 5000 rpm for 20 minutes at 4°C, 5 mL of pulp and peel juices were collected and diluted to 50 mL with ultrapure water (18.2 MΩ/cm at 25 °C). TA analysis was performed using an automatic titration instrument (CRISON, Crison Instrument, Spain). Acidity was determined by successive addition of 0.1 N NaOH (Merck, KGaA, Germany) and was expressed as g/L of malic acid following the formula:

$$TA\left({}^{g}/_{L}\, of\ malic\ acid\right) = (mL - b) * C_t *\ F_t * \frac{M}{V})$$

where *mL* is the volume of the titrant added, *b* is zero because the dilution was performed using ultrapure water, C_t is the concentration of NaOH (0.1 N), F_t is a conversion factor equal to 1, *M* is the molecular weight (*MW*) of malic acid (67.05 g/mol) and *V* is the sample juice volume (5 mL).

3.2.5 DETERMINATION OF ORGANIC ACIDS QUALITATIVE AND QUANTITATIVE PROFILES

Detection of ten OAs (i.e. oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) was carried out through high-pressure liquid chromatography (HPLC) technique. OAs determination was performed reading juice samples at UV wavelength of 210 nm and each run lasted 30 minutes. Stock solutions of OAs standards (Fluka-Sigma-Aldrich; St. Louis, MO, USA) were prepared by dissolving in ultrapure water (18.2 MΩ/cm at 25 °C). Sequential dilutions of each standard stock solution were injected into the column to build calibration curves. The retention time of each OA (t_r, **Table 3.1**) was determined by injecting the standard alone and, then, the mixed solutions at different concentrations and compositions. Another validation for the characterization of the target OA was done by spiking internal standards into some juice samples. To avoid interference between calcium ions and column resin, 100 μL of 0.5 % (w/v) Ethylenediaminetetraacetic acid (EDTA) were added to standard solutions and juices. Samples and standard solutions were passed through a 0.45 μm nylon membrane filters (CHROMACOL, LTD, UK) before the HPLC analysis. After the centrifugation of peel and pulp samples, 2 mL of clarified supernatant were collected into an Eppendorf® tube and furtherly centrifuged at 14,000 rpm for 15 minutes at 4 °C. Determination was performed by injecting 5 μL of samples into a Perkin Elmer LC200 series HPLC system equipped with a Jasko 975 UV/VIS detector (JASCO 28600, Mary's Court, Easton, MD) and an

Aminex HPX-87 Ion Exclusion column (300 x 7.8 mm; Bio-Rad Laboratiories, Inc.). The operating conditions were set at 65 °C (column temperature) with a flow rate of 0.6 mL/min and using 4 mM H_2SO_4 as elution solvent under isocratic conditions. Data processing was carried out by *Chrom Workstation 6.2* software. OAs peaks were identified by comparing the relative retention times. Manual integration of each OAs peak avoided the over-estimation of the areas in the chromatograms. The areas were quantified and converted into concentrations (ng/μL) using the calibration curves previously estimated. Furthermore, the OAs profiles of some samples were validated qualitatively and quantitatively through the ultra-high-performance liquid chromatography–high-resolution mass spectrometry (UHPLC-HRMS) method. The tests were performed using an Acquity UHPLC separation module (Waters, Milford, MA, USA) coupled with a model Exactive Orbitrap MS through a HESI-II probe for electrospray ionization (Thermo Scientific, USA) set in negative ion mode. The OAs separation was carried out using -3.0 kV of spray voltage. The capillary and tube lens voltage were respectively of -27 V and -80 V while for the skimmer was -16 V. The sheath gas flow-rate was 55 (arbitrary units) and the auxiliary gas flow-rate was 15 (arbitrary units). The temperature for the heater and the capillary was set 120°C and 320°C, respectively. A 1.8 μm HSS T3 column (150x2.1 mm, Waters) was used for separation at a flow-rate of 0.45 mL/min. The eluents were 0.05 % HCOOH in MilliQ-treated water (solvent A) and CH_3CN (solvent B). Five μL of the sample were separated by the UHPLC using the following elution gradient: 0% B for 5 min, 0-80% B in 1 min, 80 % B for 3 min and then return to initial conditions in 1 min. The column and samples were kept at 40 and 15 °C, respectively. The UHPLC eluate was investigated in full scan MS in the range $(m/z)^-$ 50-1000 u. The resolution, AGC target, maximum ion injection time and mass tolerance were 50 K, 1E6, 100 ms and 2 ppm, respectively. The ion with *m/z* 91.0038 u, corresponding to the formic acid dimer $[2M-H]^-$, has been used as lock mass. The MS data were processed using *Xcalibur* software (Thermo Scientific).

TABLE 3.1. Organic acids considered in the analysis with the corresponding retention time.

Organic acid	**Retention time (minutes)**
Oxalic acid	6.33
Cis-aconitic acid	6.87
Citric acid	7.62
Tartaric acid	8.12
Galacturonic acid	8.18
Malic acid	9.11
Quinic acid	9.59
Succinic acid	11.17
Shikimic acid	11.51
Fumaric acid	13.07

3.2.6 STATISTICAL DATA ANALYSES

All statistical analyses were performed using *RStudio* (version *1.3.1056*) in *R* environment (version *3.6.3*). Data of fruit-quality attributes were reported as the means ± standard error of replicates and elaborations were performed firstly for each year and then across years. TA distributions of apricot pulp and peel were tested by Shapiro-Wilks normality test. Distributions with minimum, maximum and mean values were estimated for each OA extracted in pulp and peel samples. Spearman's correlation coefficients were computed among fruit-quality parameters using the *corrplot* package (version 0.84). Principal components analysis (PCA) was performed on apricot data for discriminating the most relevant fruit attributes among the others. Singular values decomposition (SVD) of each principal component, followed by scaling and centring, was carried out for explaining the variance found in the apricot collection dataset. PCA analysis was performed using *RStat* and *factoextra* packages (version 1.0.7) in *RStudio*.

3.3 RESULTS

3.3.1 TITRATABLE ACIDITY

TA of fruit pulp and peel was determined separately in all tested apricots. Compared to peach dataset (i.e. *Chapter 2*), apricot has a larger range of variation for fruit TA. In season 2017, TA of pulp ranged from a minimum of 5.85 g/L of malic acid in 'BO06634241' to a maximum of 21.36 g/L of malic acid in 'BO03614108'. TA of peel varied from a minimum of 2.16 g/L of malic acid in 'BO03614029' to a maximum of 26.33 g/L of malic acid in 'BO05634124'. Mean values for TA of pulp and peel were 13.13 g/L and 15.58 g/L of malic acid, respectively. In season 2018, TA of pulp ranged from a minimum of 3.69 g/L of malic acid in 'BO06603111' to a maximum of 27.29 g/L of malic acid in 'Pricia'. TA of peel covered the range of 3.79 - 30.23 g/L of malic acid observed in 'BO06603111' and 'BO03614029', respectively. Mean values for fruit pulp and peel were 12.61 g/L and 12.78 g/L of malic acid, respectively. TA records of 58 genotypes replicated in the two seasons were averaged. In 164 unique apricot genotypes dataset, TA of pulp ranged from a minimum of 3.69 g/L of malic acid in 'BO06603111' to a maximum of 23.65 g/L of malic acid in 'BO06628081' where the mean was 12.64 g/L of malic acid. TA of peel varied from 3.79 g/L of malic acid in 'BO06603111' to 24.87 g/L of malic acid in 'Zebra'. Frequencies and distributions of TA values recorded in fruit pulp and peel for 164 unique genotypes are reported in **Fig. 3.1** and **Fig. 3.2**. Normality distribution of TA was tested through Shapiro-Wilks test with a *p-value* threshold of 0.05. Unlike in peel (*p-value* = 0.03), TA values of pulp follow the normal distribution (*p-value* = 0.11). Ratio of TA values between fruit peel and pulp was calculated for each accession to quantify the contribution of peel TA to the overall fruit acidity. The largest number of individuals had a similar acidity content in pulp and peel, with values ranging around 1 – 1.25 (**Fig. 3.3**). The minimum (0.50) and maximum (2.09) values were recorded in 'BO92618086' and 'Harval', respectively.

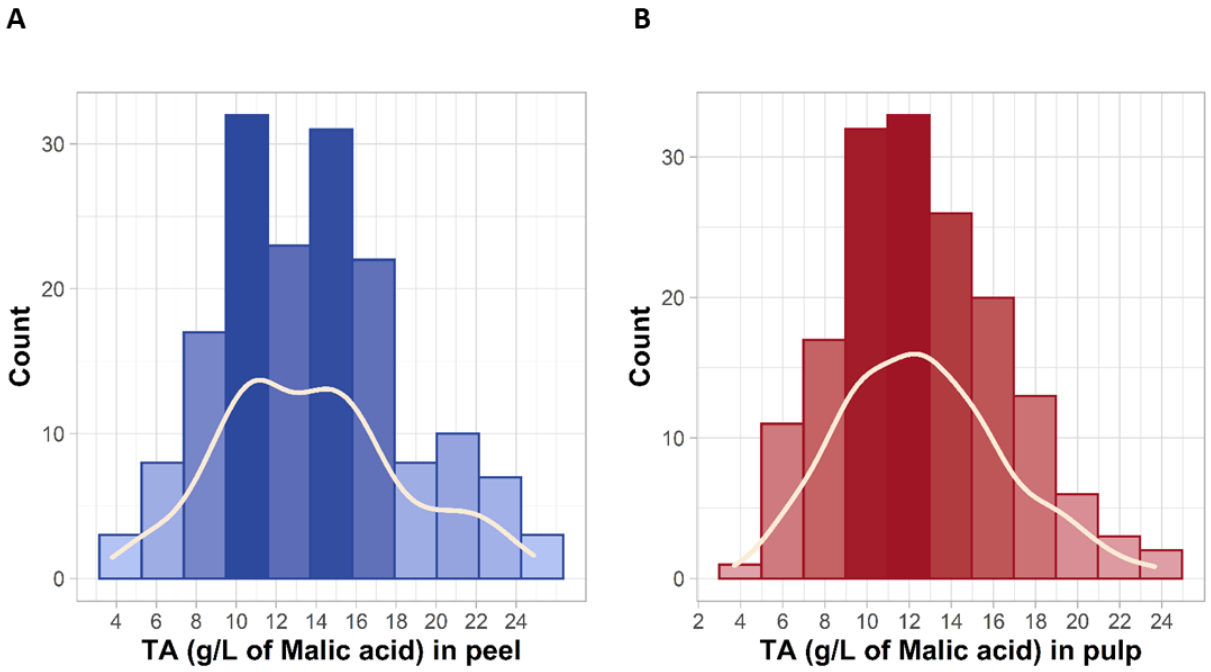

FIGURE 3.1. **Frequencies of TA records (expressed as g/L of malic acid) in apricot peel (A) and pulp (B).** TA records of 164 unique apricot genotypes were averaged when collected in both years. Most of the apricot accessions ranged between 10-15 g/L of malic acid in peel (**A**) and 9-13 g/L of malic acid in pulp (**B**). The white curve confirms the Shapiro-Wilk's test output. TA data of peel deviated from the normal distribution (*p-value* = 0.03) while TA records of pulp were normally distributed (*p-value* = 0.11).

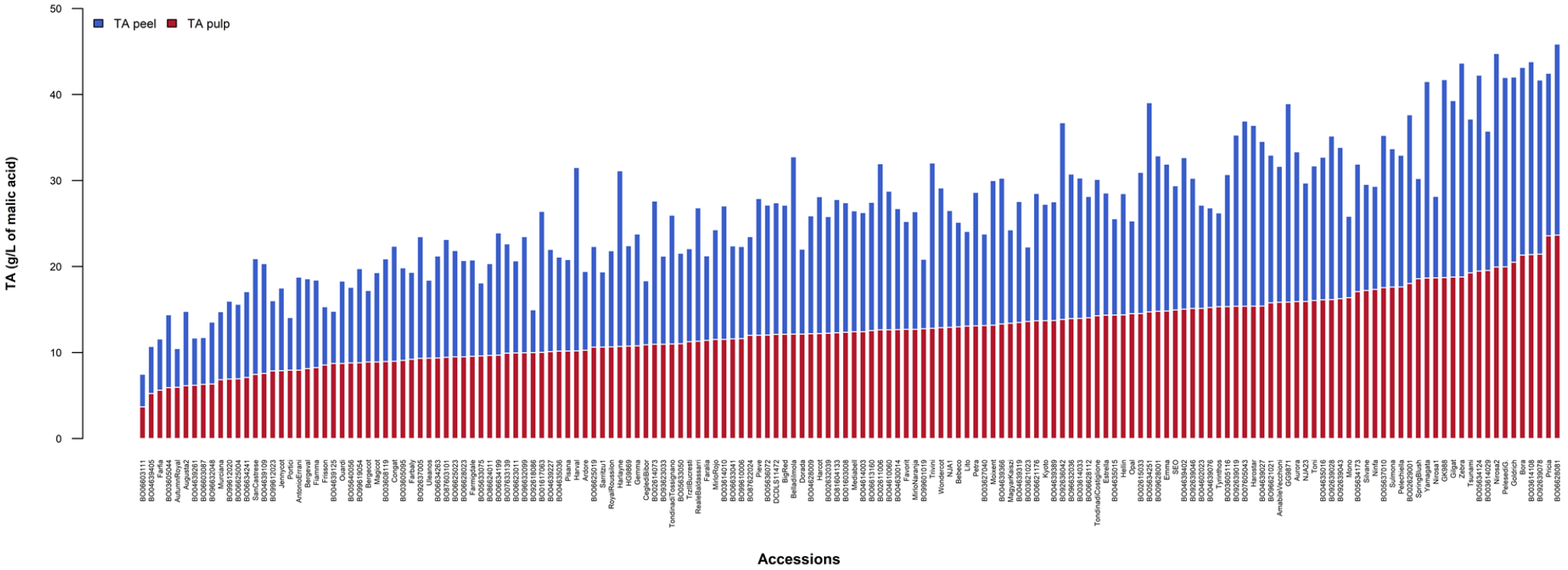

FIGURE 3.2. **Distribution of TA values (expressed in g/L of malic acid) in fruit pulp (in red) and peel (in blue).** Replicated measurements of same genotypes were averaged across years. TA of pulp (red) ranged from a minimum of 3.69 g/L of malic acid in 'BO06603111' to a maximum of 23.65 g/L of malic acid in 'BO06628081' and the mean was 12.64 g/L of malic acid. TA of peel varied from 3.79 g/L of malic acid in 'BO06603111' to 24.87 g/L of malic acid in 'Zebra' and mean value of 12.74 g/L of malic acid.

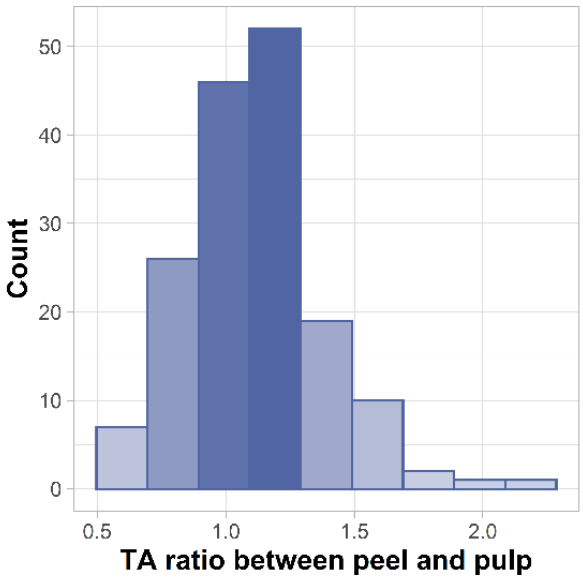

FIGURE 3.3. **Histogram of TA ratio between fruit peel and pulp.** The contribution of fruit peel to the overall acidity was investigated through the ratio of TA in peel and pulp. The largest number of individuals had similar TA acidity in pulp and peel. The minimum (0.50) and maximum (2.09) values were recorded in 'BO92618086' and 'Harval', respectively.

3.3.2 ORGANIC ACIDS PROFILES IN APRICOT PULP AND PEEL

Ten organic acids (oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) were determined qualitatively and quantitatively in apricot pulp and peel, separately (**Table 3.2** and **Table 3.3**). Data collected over the two seasons showed no within-year variability of OAs relative composition among the biological replicates of pulp and of peel. A great variability was observed among apricot accessions, but OAs patterns were similar between pulp and peel of each genotype (**Fig. 3.4**). Malic acid and citric acid were the two most abundant OAs, ranging in apricot pulp from a minimum of 1.78 mg/mL (in 'Gilgat') and 0.51 mg/mL (in 'BO04602023') to a maximum of 24.49 mg/mL (in 'Bora') and 17.09 mg/mL (in 'BO06628081'), respectively. In apricot peel, malic acid and citric acid varied from concentrations of 0.87 mg/mL (in 'Mono') and 0.16 mg/mL (in 'Royal Roussilon') to 29.12 mg/mL (in 'Bora') and 29.19 mg/mL (in 'BO04639027'), respectively. Frequencies and distributions of citrate content found in the apricot collection are reported in **Fig. 3.5** and **Fig. 3.7 (A)**, respectively. Malate concentrations in tested peel and pulp with the corresponding frequencies and distributions are shown in **Fig. 3.6** and **Fig. 3.7 (B)**. Mean value of citrate and malate content in apricots pulp averaged on 6.70 mg/mL while in peel mean was of 6.53 mg/mL and 5.62 mg/mL, respectively.

TABLE 3.2. Quantitative profiles (ng/μL) of ten OAs detected in apricot pulp collection through the HPLC analysis. Concentrations of replicated apricot genotypes were averaged across years. Minimum, maximum and mean values were estimated for each OA found in fruit pulp.

Organic acid	Minimum value (ng/μL)	Maximum value (ng/μL)	Mean (ng/μL)
Oxalic acid	0	447.23	28.56
Cis-aconitic acid	0	198.81	3.32
Citric acid	507.88	17094.26	6713.37
Tartaric acid	0	72.46	7.74
Galacturonic acid	0	269.98	38.27
Malic acid	1683.35	24493.37	6736.65
Quinic acid	0	2045.50	189.64
Succinic acid	0	2560	448.29
Shikimic acid	0	413.06	7.47
Fumaric acid	2.93	57.61	20.69

TABLE 3.3. Quantitative profiles (ng/μL) of ten OAs detected in apricot peel collection through the HPLC analysis. Concentrations of replicated apricot genotypes were averaged across years. Minimum, maximum and mean values were estimated for each OA found in fruit peel.

Organic acid	Minimum value (ng/μL)	Maximum value (ng/μL)	Mean (ng/μL)
Oxalic acid	0	123.90	21.31
Cis-aconitic acid	0	16.22	2.08
Citric acid	157.47	29185.05	6529.13
Tartaric acid	0	2088.64	57.15
Galacturonic acid	0	1247.22	188.79
Malic acid	865.51	29115.72	5617.44
Quinic acid	0	1588.37	139
Succinic acid	0	3279.58	167.50
Shikimic acid	0	240.34	8.62
Fumaric acid	1.02	29.71	8.99

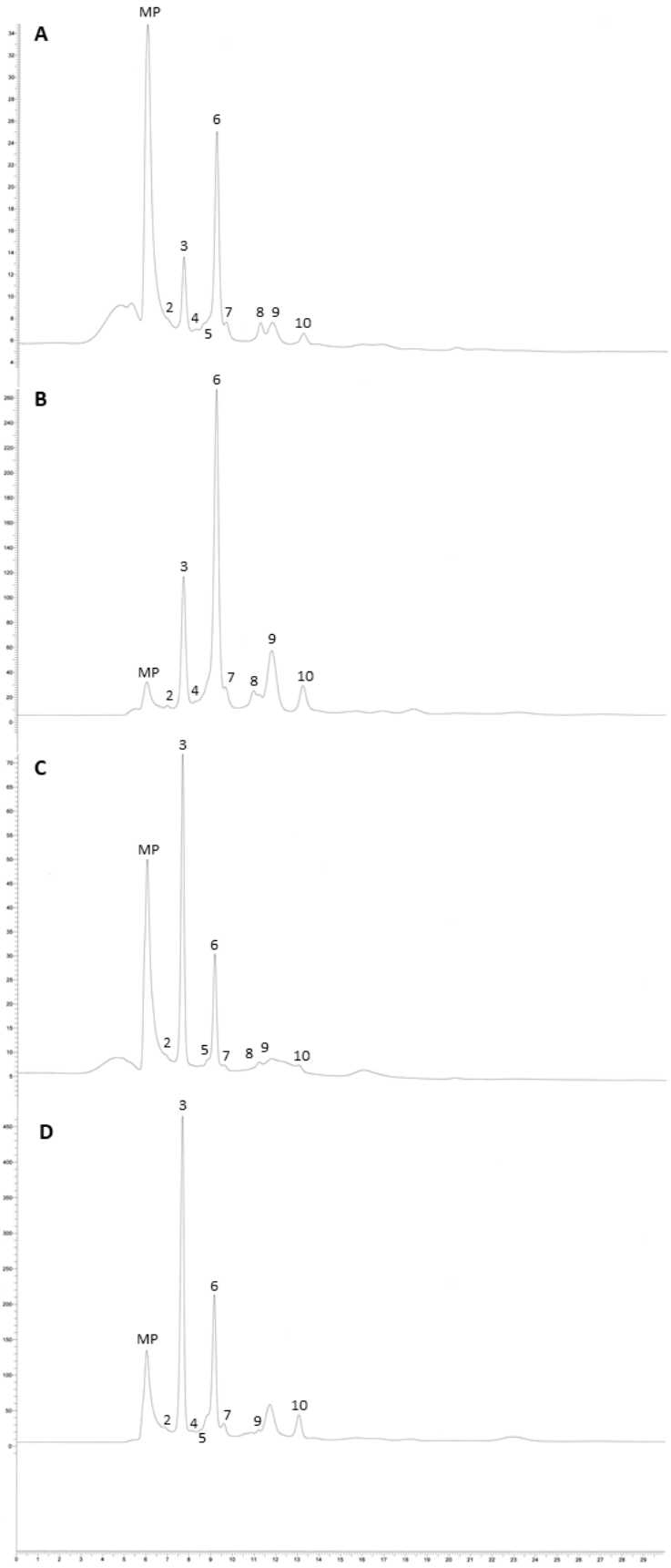

FIGURE 3.4. Chromatograms of 'Lito' peel (A) and pulp (B) and of 'Zebra' peel (C) and pulp (D). The peaks reported are: *MP,* mobile phase; *1,* oxalic acid (not present); *2,* cis-aconitic acid; *3,* citric acid; *4,* tartaric acid; *5,* galacturonic acid; *6,* malic acid; *7,* quinic acid; *8,* succinic acid; *9,* shikimic acid and *10,* fumaric acid.

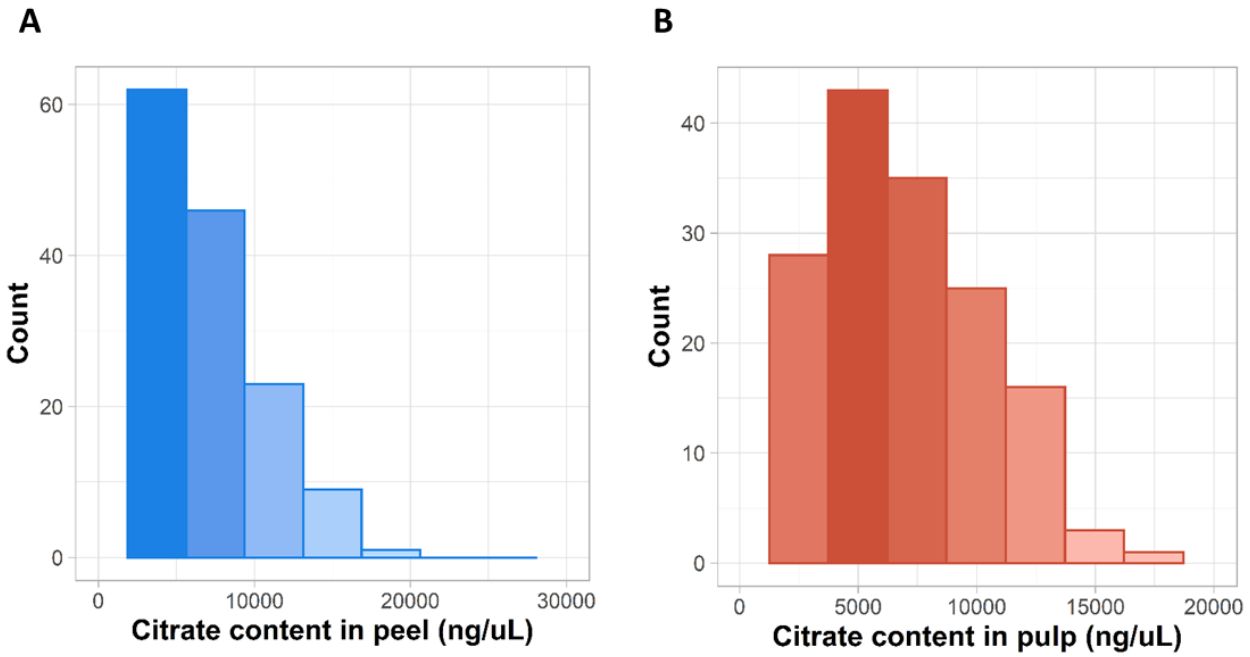

FIGURE 3.5. **Frequencies of citrate content (ng/µL) in fruit peel (A) and pulp (B) in the apricot collection**. Citrate was one of the major OAs detected in the 164 apricot genotypes considered. The largest part of apricot accessions had citrate content in the range of about 2500-5000 ng/µL and 5000-7500 ng/µL in peel and pulp, respectively.

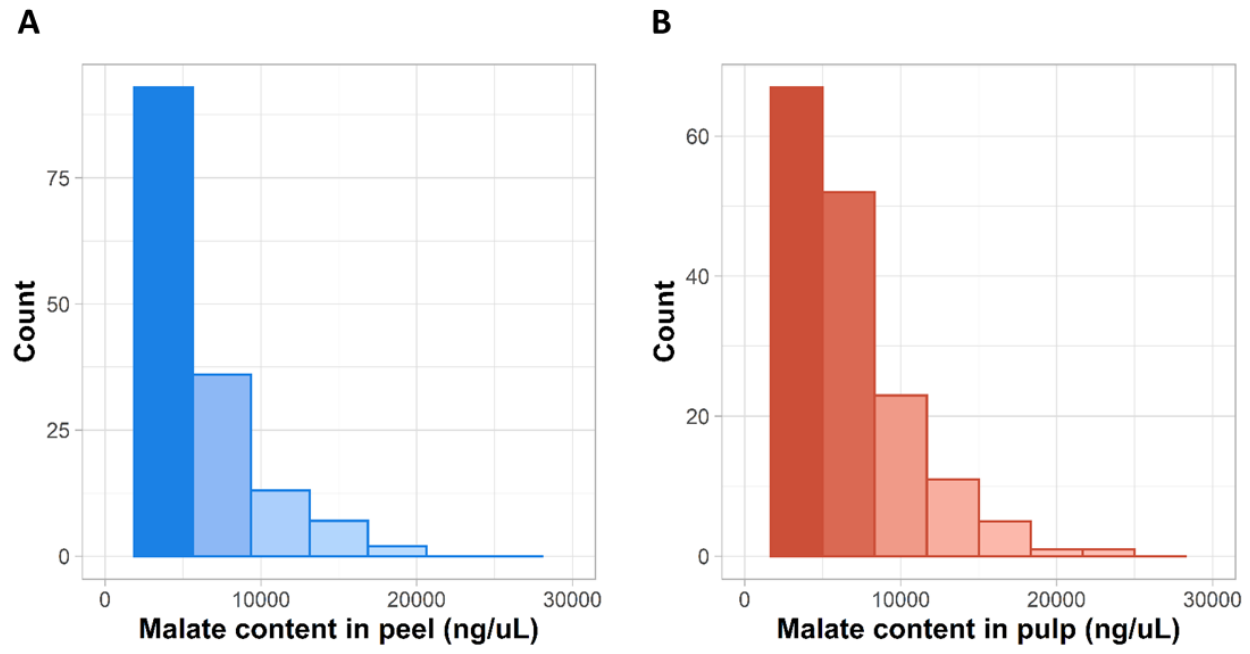

FIGURE 3.6. **Frequencies of malate content (ng/µL) in fruit peel (A) and pulp (B) in the apricot collection**. Malate was one of the two major OAs detected in the 164 apricot genotypes considered. The largest part of apricot accessions had malate content in the range of about 5000 ng/µL and 2500-7500 ng/µL in peel and pulp, respectively.

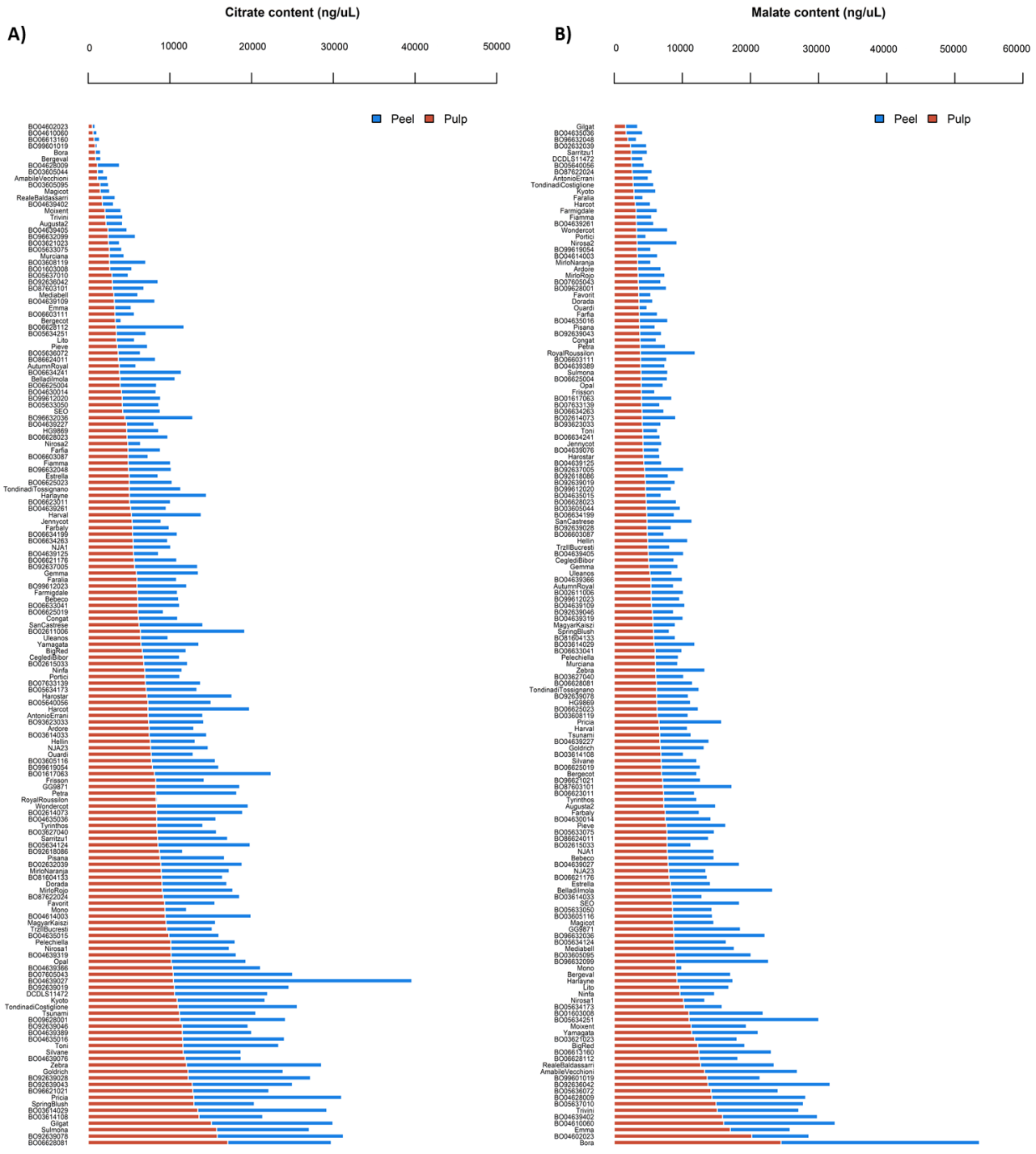

FIGURE 3.7. **Distributions of citrate and malate content (ng/µL) in fruit pulp (in red) and peel (in blue) of apricot collection.** Citrate and malate were the major OAs in both pulp and peel. The measurements of replicated genotypes were averaged across years. The range of citrate (A) was from 507.88 ng/µL in 'BO04602023' and 157.47 ng/µL in 'Royal Roussilon' to 17094.26 ng/µL in 'BO06628081' and 29185.05 ng/µL in BO04639027 in pulp and peel, respectively. Malate content of apricots pulp and peel respectively ranged from 1683.35 ng/µL in 'Gilgat' and 865.51 ng/µL in 'Mono' to 24493.37 ng/µL and 29115.72 ng/µL, both found in 'Bora'. Mean values for citrate and malate content were of about 6700 ng/µL in apricot pulp and in peel of 6529.13 - 5617.44 ng/µL, respectively.

Malate/citrate ratio was calculated for both pulp and peel in apricot collection. Almost all the apricot genotypes ranged between 0.1-1 in apricot pulp and peel (**Fig. 3.8**), with similar and highly correlated (ρ = 0.88***) malate/citrate ratio in pulp and peel (**Fig. 3.9**). Therefore, malate/citrate ratio obtained for each accession suggested that OAs seemed to be more related to genotype rather than affected by seasonality. Exceptions in malate/citrate ratio distribution were 'Royal Roussillon', 'Nirosa 2', 'Bergecot', 'BO03605095' and 'BO04628009'.

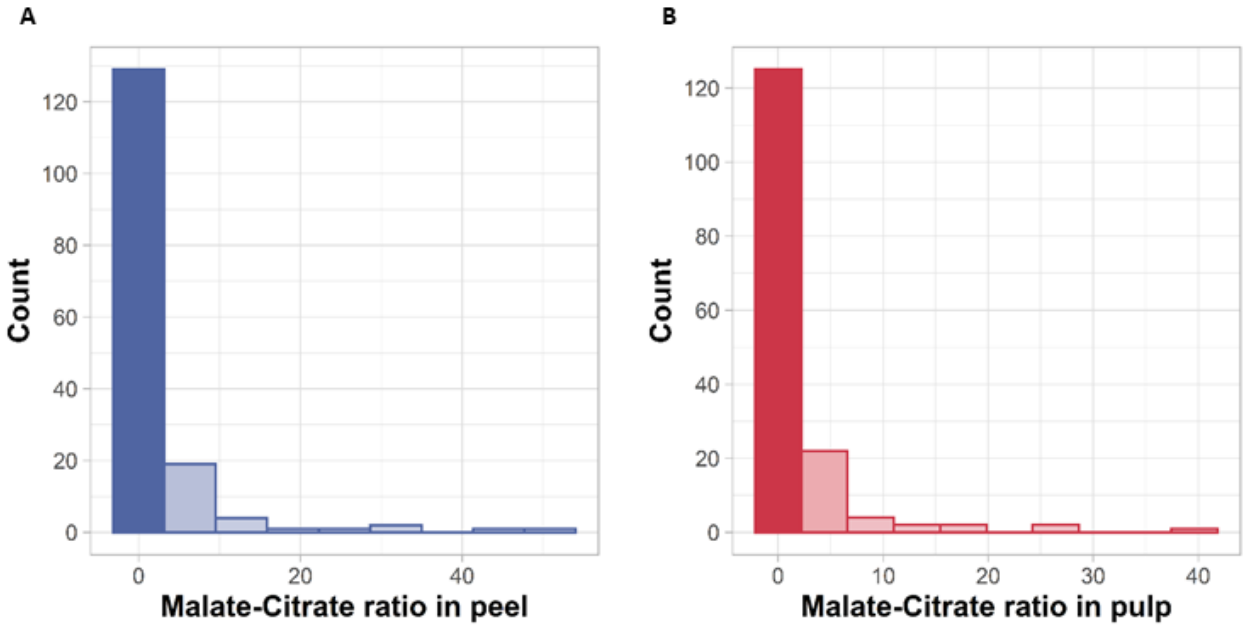

FIGURE 3.8. **Frequencies of malate-citrate content ratio in fruit peel (A) and pulp (B) in apricot collection**. Malate and citrate were the two major OAs detected in the 164 apricot genotypes considered. The measurements of replicated genotypes were averaged across years. The largest part of apricot accessions had balanced malate-citrate content between fruit peel and pulp, mostly ranging between 0.1-1. Interestingly, malate-citrate ratio was similar between fruit pulp and peel of same accession.

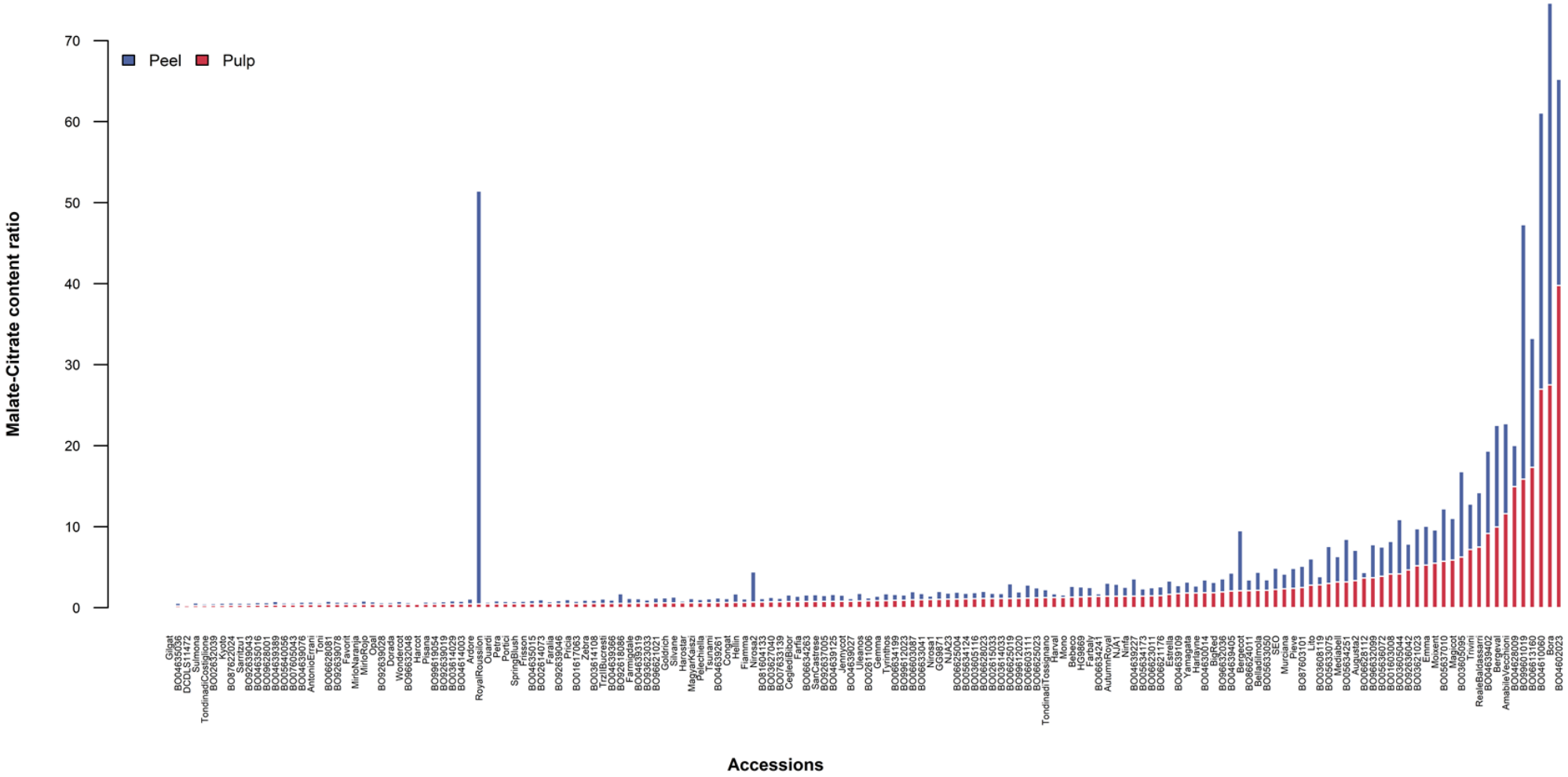

FIGURE 3.9. **Distributions of malate-citrate content ratio in fruit pulp (in red) and peel (in blue) of apricot collection.** The measurements of replicated genotypes were averaged across years. Malate-citrate ratio was similar and highly correlated (r= 0.88***) between fruit pulp and peel in almost all apricot genotypes, suggesting that OAs patterns were more genotype than year-dependent. Exceptions in ratio distribution are 'Royal Roussillon', 'Nirosa 2', 'Bergecot', 'BO03605095' and 'BO04628009'.

Among ten OAs considered, malate, citrate and fumarate were detected in all apricot accessions, both in fruit pulp and peel. Fumarate content in pulp and peel was lowest in 'Nirosa 2' (2.93 ng/μL) and in 'Estrella' (1.02 ng/μL), respectively. The highest fumarate concentration was found in 'Autumn Royal' reaching 57.61 ng/μL and 29.71 ng/μL in pulp and peel, respectively. Among OAs related to the tricarboxylic acid (TCA) cycle, succinic acid was the third most abundant OA detected in almost all the apricot genotypes. Except two selections and two cultivars ('BO03605044', 'BO04639261', 'Pelechiella' and 'San Castrese') where succinate was not found in fruit pulp, this OA ranged from 1.10 ng/μL (in 'Fiamma') to 2560 ng/μL (in 'Yamagata'). No succinate concentrations were detected in apricot peels of many accessions, although it reached a relevant maximum content of 3.28 mg/mL in 'BO04638027'. Oxalate and cis-aconitate mostly occurred at low concentrations in fruit pulp with the highest content of 0.45 mg/mL in 'Congat' and 0.20 mg/mL in 'Harleyne', respectively. The maximum amount of oxalate and cis-aconitate in fruit peel was 0.12 mg/mL in 'GG9871' and 16.22 ng/μL in 'Trivini', respectively. Among OAs not produced through TCA cycle, quinic acid was abundantly detected in apricot pulp and peel reaching concentrations of 2.05 mg/mL in 'BO92618086' and 1.6 mg/mL in 'BO04635036'. Tartaric acid was more abundant in apricot peel than in pulp reaching a maximum of 2.09 mg/mL in 'Royal Roussilon'. Shikimate was largely abundant in fruit pulp while in fruit peel, except three accessions ('Gemma', 'Magicot' and 'Spring Blush'), was detected in almost all tested apricot genotypes with a maximum content of 0.41 mg/mL and 0.24 mg/mL, respectively, both found in the same selection ('BO92618086'). In the end, galacturonate was not always detected in apricot genotypes although highest content of 0.27 mg/mL ('BO02615033') and 1.25 mg/mL ('Bergecot') were found in pulp and peel, respectively.

3.3.3 CORRELATION AMONG FRUIT-QUALITY ATTRIBUTES IN APRICOT PULP AND PEEL

Correlations among TA, OAs and other fruit quality attributes (i.e. HD, SSC, FW, DM% and I_{AD}) in apricot pulp and peel were calculated firstly between years (**Table 3.4**, **Table 3.5** and **Table 3.6**). Almost all the OAs and fruit parameters considered showed a relatively high stability across the two years, suggesting that OAs patterns were more genotype-dependent. In particular, among the ten OAs, contents of citrate, malate, galacturonate and shikimate were highly correlated across years (ρ of 0.78, 0.69, 0.64 and 0.60 respectively) in apricot peel. Among fruit quality traits, TA of apricot pulp and harvest day (HD) were significantly the less year-dependent with ρ of 0.83 and 0.81, respectively. Citrate, malate and shikimate were significantly stable between 2017 and 2018, with corresponding ρ of 0.88, 0.79 and 0.61. Malate/citrate ratio was also constant between the two years with ρ of about 0.90 in both fruit pulp and peel. According with the high stability of the largest part of fruit-quality attributes, data of each apricot accession were averaged across years and correlations were tested (**Fig. 3.10**).

TABLE 3.4. Spearman's test correlation matrix among some fruit quality attributes recorded across the two harvesting season 2017 and 2018. The correlation coefficients are significant when *p-value* < 0.05 (* significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level). Correlations were not significant at ' '.*TA* is titratable acidity measured separately for pulp and peel (g/L of malic acid), *SSC*, soluble solids content (°Brix); *FW*, fresh weight (g); *HD*, harvest day (Julian days); *IAD*, chlorophyll absorbance index and *DM%*, dry matter percentage.

	TA (pulp)$_{2018}$	**TA (peel)$_{2018}$**	**TA (peel:pulp)$_{2018}$**	**SSC$_{2018}$**	**FW$_{2018}$**	**HD$_{2018}$**	**IAD$_{2018}$**	**Firmness$_{2018}$**	**DM%$_{2018}$**
TA (pulp)$_{2017}$	0.83***								
TA (peel)$_{2017}$		0.55***							
TA (peel:pulp)$_{2017}$			0.43*						
SSC$_{2017}$				0.48***					
FW$_{2017}$					0.44**				
HD$_{2017}$						0.81***			
IAD$_{2017}$							0.16		
Firmness$_{2017}$								0.44	
DM%$_{2017}$									0.65***

TABLE 3.5. Spearman's test correlation matrix among apricot peel OAs recorded across the two harvesting season 2017 and 2018. The correlation coefficients are significant when *p-value* < 0.05 (* significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level). *MC* is the ratio between malate and citrate content detected in each genotype.

	Oxalate$_{2018}$	**Cis-aconitate$_{2018}$**	**Citrate$_{2018}$**	**Tartrate$_{2018}$**	**Galacturonate$_{2018}$**	**Malate$_{2018}$**	**Quinate$_{2018}$**	**Succinate$_{2018}$**	**Shikimate$_{2018}$**	**Fumarate$_{2018}$**	**MC$_{2018}$**
Oxalate$_{2017}$	-0.07										
Cis-aconitate$_{2018}$		-0.07									
Citrate $_{2018}$			0.78***								
Tartrate$_{2018}$				0							
Galacturonate$_{2018}$					0.64***						
Malate$_{2018}$						0.69***					
Quinate$_{2018}$							0.26				
Succinate$_{2018}$								0.25			
Shikimate$_{2018}$									0.60***		
Fumarate$_{2018}$										0.18	
MC$_{2018}$											0.89***

TABLE 3.6. Spearman's test correlation matrix among apricot pulp OAs recorded across the two harvesting season 2017 and 2018. The correlation coefficients are significant when *p-value* < 0.05 (* significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level). *MC* is the ratio between malate and citrate content obtained for each genotype.

	Oxalate$_{2018}$	**Cis-aconitate$_{2018}$**	**Citrate$_{2018}$**	**Tartrate$_{2018}$**	**Galacturonate$_{2018}$**	**Malate$_{2018}$**	**Quinate$_{2018}$**	**Succinate$_{2018}$**	**Shikimate$_{2018}$**	**Fumarate$_{2018}$**	**MC$_{2018}$**
Oxalate$_{2017}$	0										
Cis-aconitate$_{2018}$		0.26									
Citrate $_{2018}$			0.88***								
Tartrate$_{2018}$				0.41***							
Galacturonate$_{2018}$					0.36*						
Malate$_{2018}$						0.79***					
Quinate$_{2018}$							0.47***				
Succinate$_{2018}$								0.25**			
Shikimate$_{2018}$									0.61***		
Fumarate$_{2018}$										0.54***	
MC$_{2018}$											0.91***

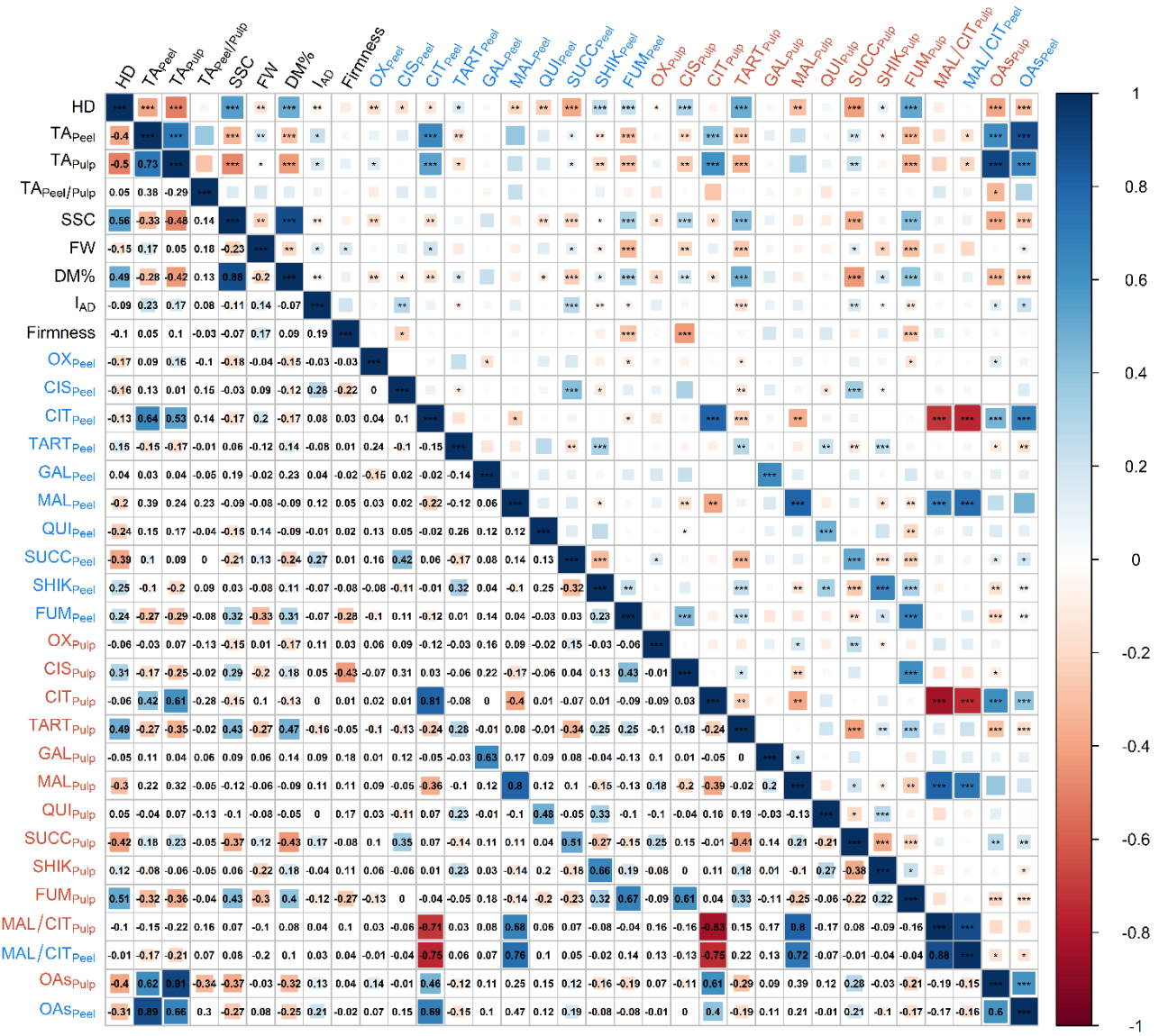

FIGURE 3.10. Spearman's rank correlation among TA, OAs content and other fruit quality attributes in apricot collection. Spearman's correlation test among fruit quality traits was carried out on 164 unique apricot genotypes. Samples were collected in two harvesting seasons (2017 and 2018) followed by separate analyses on pulp (labels are coloured in red) and peel (labels in blue). Correlations are significant at 5% "*", 1% "**" and 0.1% "***" level. Correlation coefficients are coloured on the left square when significant. *HD*, harvest day (expressed as Julian days); *TA*, fruit titratable acidity of pulp and peel, separately; $TA_{Peel/Pulp}$, fruit TA ratio between pulp and peel of each apricot accession; *SSC*, soluble-solids content (°Brix); *FW*, fresh weight (g); *DM%*, dry matter in percentages; I_{AD}, chlorophyll absorbance index; *OX*, oxalate; *CIS*, cis-aconitate, *CIT*, citrate, *TART*, tartrate, *GAL*, galacturonate, *MAL*, malate, *QUI*, quinate, *SUCC*, succinate, *SHIK*, shikimate, *FUM*, fumarate; *MAL/CIT*, ratio between malate and citrate content (ng/µL) detected in fruit pulp and peel of each apricot accession; *OAs*, total sum of OAs content (ng/µL) found in samples of fruit pulp and peel.

Significant correlations were found among many fruit-quality attributes in apricot collection. TA was strongly correlated between apricot pulp and peel (ρ= 0.73) and affected by total OAs content, with coefficients of 0.89 and 0.91 in apricot fruit peel and pulp, respectively. Content of almost all OAs (except of oxalate and cis-aconitate) in peel was related to the one in pulp with correlation coefficients ranging from about 0.3 to 0.8. Although both citrate and malate were significantly more present that other OAs, citrate seemed to largely contribute to the overall TA in peel (ρ = 0.64) and pulp (ρ = 0.61).

In particular, in fruit pulp, malate content was negatively affected by citrate concentrations (ρ = -0.39). Total OAs concentration in apricot peel and pulp strongly affected the TA recorded for each genotype (ρ = 0.89 for peel and ρ = 0.91 for pulp), giving a possible alternative way to support the acidity values recorded in apricot collection. Among other OAs in apricot peel, succinate content was positively related to cis-aconitate concentration (ρ = 0.42). In apricot pulp interesting correlations were found between shikimate and succinate (ρ = -0.38) and between fumarate and cis-aconitate (ρ = 0.61). SSC and DM% were both negatively related to TA and positively correlated to HD (correlation coefficient of about 0.5).

3.3.4 PCA ANALYSIS

Principal component analysis (PCA) was carried out on TA, OA contents in pulp and peel and the other fruit-related quality. The first two components (i.e. *Dim1* and *Dim2*) accounted for 30.8 % of the total fruit-quality variation in apricot collection (**Fig. 3.8**). Before the analysis, TA ratio between apricot peel and pulp was removed from the dataset having a not significant contribution to the variability explanation in apricot dataset. Moreover, total sum of OAs content in fruit pulp and peel was removed because redundant of TA values recorded for each genotype (correlation coefficients were of 0.89 for peel and 0.91 for pulp). In particular, *Dim1* was positively and strongly associated with TA of pulp and peel but negatively associated with HD, SSC, DM% and fumarate concentrations. *Dim2* was positively and strongly affected by citrate content but negatively related to malate concentrations and malate/citrate ratio, both in pulp and peel. PCA analysis provided a simplified classification of fruit-quality records, particularly TA and OAs content in both fruit pulp and peel, collected for 164 apricot genotypes. Furthermore, by adding apricot accessions names, PCA analysis provided a classification of apricot accessions showing that 'Yamagata', 'BO05634124', 'BO04639027', 'Tsunami' and 'Pricia' mostly contributed to TA of pulp and peel while 'BO04639405', 'Autumn Royal', 'BO04639261', 'BO92618086' and 'Augusta 2' greatly represented HD, SSC, fumarate content and DM%. Malate content and malate-citrate ratio were affected by 'BO99601019', BO04628009', 'Amabile Vecchioni' and 'Bergeval' while 'Gilgat' contributed to citrate concentrations.

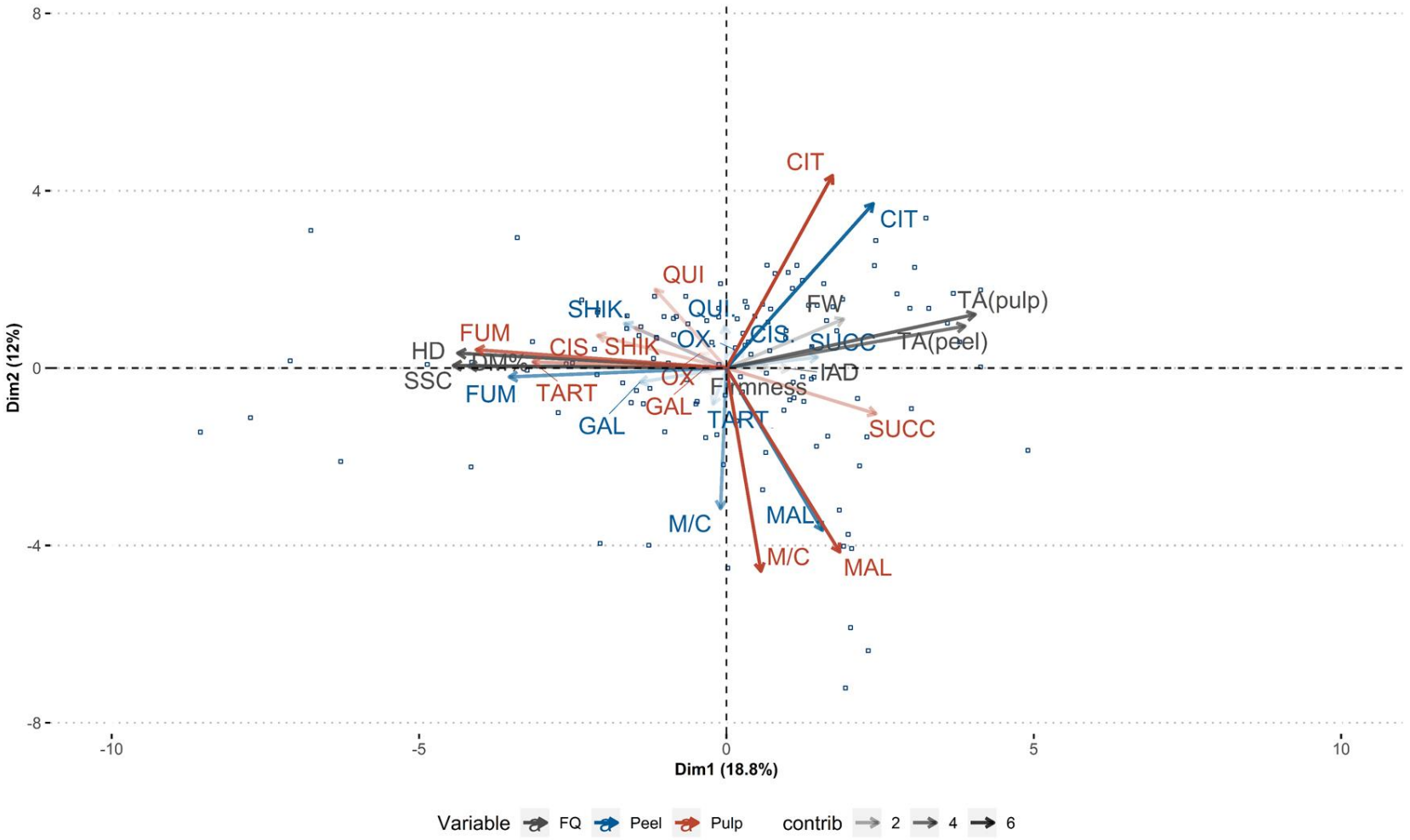

FIGURE 3.11. PCA plot on fruit quality attributes in apricot collection. PCA analysis was carried out on TA, ten OAs and other six fruit-quality related traits. Acidity and OAs profiles were analysed in fruit pulp and peel, separately. The first two components (i.e. *Dim1* and *Dim2*) explained 30.8 % of the total variability in apricot collection. *HD*, harvest day (expressed as Julian days); *TA*, fruit titratable acidity of pulp and peel, separately; *SSC*, soluble-solids content (°Brix); *FW*, fresh weight (g); *DM%*, dry matter in percentages; *IAD*, chlorophyll absorbance index; *CIS*, cis-aconitate content; *CIT*, citrate content; *FUM*, fumarate content; *GAL*, galacturonate content; *MAL*, malate content; *OX*, oxalate content; *QUI*, quinate content; *SHIK*, shikimate content; *SUCC*, succinate content; *TART*, tartrate content.; *M/C*, ratio between malate and citrate content (ng/μL) detected in fruit pulp and peel of each apricot accession.

3.4 DISCUSSION

A large dissection of apricot genotypes was carried out, with a particular focus on TA and OAs profiles in fruit pulp and peel, separately. Consumers' satisfaction is mainly driven by sugars/acids ratio, although TA and OAs pattern heavily influence taste and sweetness perception (Bartolozzi *et al.*, 1997; Colaric *et al.*, 2005; Ruiz and Egea, 2008; Etienne *et al.*, 2013). The close relation between sweetness and consumers' acceptability has been previously investigated in peach (Delgado *et al.*, 2013; Echeverría *et al.*, 2015) and in apricot (Fan *et al.*, 2017) but without considering a large germplasm collection. High but balanced sugars/acids content is desired by consumers. The constant complains about the poor apricot quality and the demand of high-quality apricots have highlighted the need of deep phenotyping of apricot genotypes for acids content. A large variability among apricot accessions was found in TA and OAs profiles, more dependent on genotypic than on seasonal effects (Bassi *et al.*, 1996). However, to avoid effects related to agronomical practices daily performed in orchards and validate fruit-quality related data across years, plants were uniformly maintained. TA and OAs content were evaluated separately in pulp and peel to capture the contribute of each fruit component to the final acidity. TA greatly varied in apricot collection, with range of 3.69-23.65 g/L of malic acid and 3.79-24.87 g/L of malic acid in pulp and peel, respectively. No significant differences were found between fruit pulp and peel of each accession, with a strong correlation of 0.73. Similarly, OAs pattern showed no relevant variations between pulp and peel of same genotype confirming previous work results (Xi *et al.*, 2016). OAs profiles remained almost stable within each apricot accession, also when little variations in absolute OAs concentrations were detected. OAs content in ripe apricots results from the balance among OAs synthesis, catabolism, transport and vacuole storage (Ruffner *et al.*, 1984). At plant level, OAs ensure redox equilibrium generating ionic gradients across membranes and supply substrates for other related metabolic pathways. Sourness perception is not only related to total acids concentration but also to the qualitative composition of OAs, due to different sensorial impact of each acid on taste during apricot consumption. Total OAs content correlated to overall TA in pulp and peel (with correlation coefficients of 0.89 and 0.91, respectively) suggesting a possible two-ways of expressing apricot acidity. Similar to peach (Moing et., 1998), malate and citrate were the most abundant reaching maximum concentration of 17.09 mg/mL and 24.49 mg/mL in pulp, respectively, and of 29.19 mg/mL and 29.12 mg/mL in peel, respectively. Malate/citrate ratio was strongly related between fruit pulp and peel but the relationships with TA was very weak and not significant. Similar malate-citrate ratio was found between pulp and peel of each accession validating the possible genetic basis underlying OAs pattern in apricot. Although not the prominent OA in pulp, malate was described as sourer but less strong than citrate (Colaric *et al.*, 2005; Xi *et al.*, 2016). The decreased malate and citrate content during *Citrus* fruit post-harvest were described as adverse factors to the overall fruit quality (Sun *et al.*, 2013).

Among OAs of TCA cycle, succinate was the third most abundant OA both in apricot pulp and peel, reaching relevant concentration in 'Yamagata' pulp (2.56 mg/mL) and 'BO04635036' peel (1.59 mg/mL). Succinic acid content increased at before the fruit full-maturity stage (ρ = 0.27 in peel and ρ = 0.17 in pulp) and was strongly related to cis-aconitate levels (ρ of about 0.4 in pulp and peel). Succinate is found in other species as broccoli and sugar beets but is also frequently used as acidulant in food processing. Furthermore, a previous work identified succinic acid as an enhancer of umami taste in certain food (Kaneko *et al.*, 2006). Fumaric acid naturally occurs in other fruit crops as papaya, pear and plum. Fumarate was always detected in apricot germplasm collection, although low concentrations were found in pulp (range of 2.93 – 57.61 ng/μL) and peel (range of 1.02 – 29.71 ng/μL). Fumarate was strongly correlated to cis-aconitate in pulp (ρ = 0.61) and negatively contributed to TA (ρ = - 0.36 in pulp and ρ = -0.27 in peel). Compared to other OAs, fumarate is slightly soluble in water and this could explain the positive correlation found with DM% (ρ = 0.40). Higher content of cis-aconitate in apricot pulp seemed to result in fruit softening, being negatively correlated to fruit firmness (ρ = - 0.43). Cis-aconitate seemed to reach higher amounts in peel when fruit was not mature (correlation coefficient of 0.28 with I_{AD}). Albeit in traces in almost all accessions, oxalate reached a maximum concentration of 0.47 mg/mL in 'Congat' pulp. In general, oxalate is considered a secondary OA in fruit (Moing *et al.*, 2000). Relevant oxalate concentrations were found in unripe kiwifruits, probably useful for modulating soluble calcium concentration in the plant (Rassam *et al.*, 2007). Among OAs metabolized through TCA cycle, oxalate and cis-aconitate were more year-dependent, both in fruit pulp and peel.

Quinate, shikimate, galacturonate and tartrate are OAs not produced by TCA cycle. Quinate content was more stable in fruit pulp across years and was reported at low concentration also in other species as peach (Moing *et al.*, 1998; Etienne *et al.*, 2013; Baccichet *et al.*, in press). Shikimate is generated from Pentose phosphate pathway or from phosphoenolpyruvate (PEP) and it is a precursor of aminoacids (i.e. Tyrosine) and secondary metabolites in plants such as flavonoids and phenylpropanoids. Galacturonic acid is the most important constituent of pectins, polysaccharides building the cell walls in plants. Fruit ripening is characterized by pectins depolymerisation and generally leads to pulp softening and decreased fruit firmness (Femenia *et al.*, 1998; Brummel, 2006; Goulao and Oliveira, 2008; Kovács *et al.*, 2008). On the contrary, galacturonate was not related to firmness in the considered apricot accessions. Tartrate characterizes OAs profile of other species as grapevine and with oxalate and L-threonate it seems to be a catabolic product of ascorbate (Vitamin C) in plants (DeBolt *et al.*, 2007). According with previously results (Xi *et al.*, 2016), higher tartrate contents were observed in peels than in pulps reaching maximum of 2.09 mg/mL and 72.46 ng/μL, respectively.

Different correlations were found among OAs and HD, both in pulp and peel. Early-ripening apricot accessions were recorded with higher amounts of cis-aconitate, tartrate and fumarate in pulp and with higher shikimate and fumarate levels in peel. On the other hand, late-ripening apricots had higher contents of quinate and

succinate in peel and of malate and succinate in pulp. As expected, SSC and TA were negatively related to each other in pulp (ρ= -0.48).

3.5 CONCLUSION

In this chapter a large collection of apricot germplasm was explored for several fruit-quality related attributes. Although the narrow diversity of the apricots sold on markets, apricot germplasm was extremely diversified, making apricot breeding an interesting challenge for selecting new varieties and for the overall fruit quality improvement. Acidity range of variation was wider than in peach (i.e. *Chapter 2*), where citrate and malate were the most abundant OAs in all the tested individuals. Moreover, the developed-approach of screening apricot pulp and peel separately has provided important information on organoleptic properties of each accession, revealing a genotype-dependence of OA patterns and amounts between fruit pulp and peel. However, more research will be useful to better understand all metabolic pathways involved in the accumulation of OAs in apricots. Further studies about the impact of OAs on taste may be carried out in the future.

4. GENOMIC PREDICTIONS-BASED SELECTION FOR FRUIT ACIDITY AND CONTENT OF MALATE AND CITRATE IN A LARGE PEACH COLLECTION

4. 1. PEACH FRUIT ACIDITY, GENOMICS AND BREEDING: A GENERAL PREFACE

Peach genome was sequenced and assembled from the double haploid '*Lovell*' by the 'International Peach Genome Initiative' (*IPGI*). The small genome dimension combined with high quality of assembled reads lead peach genome to become a reference for other *Rosaceae* species (Verde *et al.*, 2012; Verde *et al.*, 2017). The 9K SNP array V1 was released in 2012 ((Verde *et al.*, 2012) followed by the improved version releasing in 2017 (Verde *et al.*, 2017). Peach was a breeding target for specific traits, as disease resistance or fruit quality-related attributes. Traditional breeding - driving strategy of the last decades - finds some hampers in the long juvenile period of peach trees, in human vision-based selection, in screening large seedlings populations for target traits and in a reduced genetic variation of peach accessions grown. Moreover, genotype-environment interactions (*GxE*) and management strategies in orchards can alter individuals' phenotypic expression - especially when the target trait behaves as quantitative - decreasing individual's phenotypic performance stability in following years. New breeding strategies are required for facing all those aspects and satisfying consumers' demand of superior peaches quality. Low-acid trait (*Locus D* where *D* means "sweet" in French) has been characterized at genetics level, revealing a dominance inheritance and becoming a relatively easy trait to breed (Byrne, 2002). However, *Locus D* captures only a part of acidity variability in peach accessions grown worldwide. Especially in a period characterized by lack of funding for agriculture, genomic selection (GS) can play a key role in peach breeding, basing on genomic predictions (GP) of the individuals' phenotypic values before reaching the adult stage and identifying candidate parents carrying characters of interest.

In this chapter, GP-based selection was implemented on a large peach collection of accessions genotyped and phenotyped for fruit TA and OAs content (i.e. *Chapter 2*) and tested on only-genotyped seedlings from five crosses in order to predict individuals' phenotype.

4. 2 GENOMIC SELECTION BACKGROUND

4.2.1 GENOMIC SELECTION: WHERE IT ALL BEGAN

GP-based selection was developed and implemented in dairy cattle in 2001 (Meuwissen *et al.*, 2001) revolutionizing the bovine breeding. In fact, it was possible to predict the genetic merits of bulls used for artificial insemination to give birth to cattle able to produce higher milk amount. GS increased the gain for

dairy industry anticipating the bulls screening to the age of 12 months instead of 7 years (Goddard and Hayes, 2009; Voss-Fels *et al.*, 2018).

The main principle of GP is to have a reference or training population of individuals both genotyped and recorded for the character of interest and use the information available for developing a statistical model. All the single nucleotide polymorphisms (SNPs) effects are estimated simultaneously from reference population to obtain a prediction equation where everyone's genetic merit is calculated by best liner unbiased predictors (BLUPs) (Henderson, 1963; Henderson, 1975; Henderson, 1984). After the training, GP model is used in a population of individuals only genotyped for predicting their genomic estimated breeding values (*GEBVs*, the genetic potential inherited from parents to the progeny) and so their phenotypic performance. Meuwissen *et al.* (2001) demonstrated that prediction accuracy is higher for GS than for *MAS*. GS has become the preferred strategy in dairy cattle industry replacing progeny test (Hickey *et al.*, 2017). GS seems very promising not only for dairy cattle but also for plants.

4.2.2 MAIN ADVANTAGES OF GS

GP-based strategies help breeders to identify candidate parents for controlled- crosses estimating their genetic merits (Meuwissen *et al.*, 2001). GP-based strategy allows breeders to plan more accurately controlled-crosses because candidate parents are chosen based on genotypic information rather than phenotypic records collected over the previous years (Gezan *et al.*, 2017). Furthermore, GS seems to cover the *MAS* failures bringing several advantages to fruit tree-crops breeding programmes, as seen in apple (Chagnè *et al.*, 2007; Bus *et al.*, 2009; Kumar *et al.*, 2012*a*; Kumar *et al.*, 2012*b*). Unlike in MAS, GP fits all the genome wide distributed markers simultaneously without establishing a threshold and discarding the ones less statistically significant (Meuwissen *et al.*, 2001). Similar to MAS, GP-based selection assumes that the target trait is in linkage disequilibrium (LD) with at least one SNP in order to capture with higher probability minor loci, contributing less to the phenotypic target trait but essentials in estimating the individuals breeding value (Kumar *et al.*, 2012*b*; Voss-Fels *et al.*, 2018). Then, selection based on individuals' GEBV is accurate including the estimation of each marker effect (Meuwissen *et al.*, 2001; Goddard and Hayes, 2009; Heffner *et al.*, 2009; Heffner *et al.*, 2010). GS is time-saving compared to progeny test and MAS approaches shortening breeding programmes length.

4.2.3 STATE-OF-THE-ART OF GS STUDIES IN FRUIT TREE CROPS

GS has found lesser applications in plants than in animals and the largest part have been carried out on annual, inbreed and model crops as alfalfa (Annicchiarico *et al.*, 2015), maize (Lorenzana and Bernardo, 2009; Zhao *et al.*, 2012; Windhausen *et al.*, 2012; Riedelsheimer *et al.*, 2012), rice (Xu, 2013; Spindel *et al.*, 2015), wheat (Heffner *et al.*, 2011; Poland *et al.*, 2012), sorghum (Hunt *et al.*, 2018; Fernandes *et al.*, 2018), barley (Zhong *et al.*, 2009; Lorenz *et al.*, 2012; Heslot *et al.*, 2013) and sugar beet (Biscarini *et al.*, 2014).

In perennial crops there are some applications on strawberry (Gezan *et al.*, 2017), raspberry (Stephens *et al.*, 2012), apple (Kumar *et al.*, 2012a; Kumar *et al.*, 2013; Kumar *et al.*, 2015; Muranty *et al.*, 2015; Hardner *et al.*, 2016*c*), grapevine (Fodor *et al.*, 2014), loblolly pine (Zapata-Velenzuela *et al.*, 2013), eucalyptus (Resende *et al.*, 2012*b*; Hardner *et al.*, 2016*a*), Japanese pear (Iwata *et al.*, 2013), sweet cherry Hardner *et al.*, 2016*b*) and peach (Biscarini *et al.*, 2017). GS on several fruit quality-related traits (i.e. fruit firmness, titratable acidity, appearance, red-flesh coverage and absence of physiological defects) in apple evidenced the great advantages of GP-based approach in simultaneously fitting all the molecular markers and capturing a larger genetic variance proportion (Kumar *et al.*, 2013). Also, prediction models were adopted in European peach varieties to predict fruit TA, SSC and fresh FW revealing that GS seems feasible in this species (Biscarini *et al.*, 2017).

4.2.4 LINEAR MIXED MODELS IN GS STUDIES

GP-based approaches deal with the "large *p* small *n*" problem where the amount predictors (i.e. molecular markers; *p*) is larger than the number of observations (*n*) (Jannink *et al.*, 2010). Especially with the high-throughput genotyping (HTG) – not always in conjunction with a high-quality phenotyping-phenotyping the problem is very common, leading to consider predictors as fixed, overfitting of molecular marker effects and thus obtaining a reduced-predictive ability (Lorenz *et al.*, 2012). In developing genomic prediction equation, linear mixed models (LMMs) are very useful and flexible allowing missing data, correlation among traits, and heterogeneous variances in the dataset (Welham *et al.*, 2004). The general LMM equation is:

$$y = Xb + Zu + e$$

where *y* is a vector (*nx1* where *n* corresponds to the records number) of the collected observations, *b* is an unknown vector of fixed effects (*px1*, where *p* represents the number of levels of the fixed effects), *X* is the known design (or incidence) matrix (*nxp*) showing the relationships between records and fixed effects, *u* is the unknown vector of random effects (*qx1* where *q* represents the number of levels of the random effects), *Z* is the known design (or incidence matrix; *nxq*) associating the observations in *y* to random effects and *e* is the unknown vector (*nx1*) of residual errors. In LMMs, vectors of random and residual effects are uncorrelated and follow a normal distribution:

$$u \sim N_q(0, G)$$

$$e \sim N_n(0, R)$$

Variance is:

$$Var\begin{pmatrix} u \\ e \end{pmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G(\sigma_g) & 0 \\ 0 & R(\sigma_r) \end{bmatrix}\right)$$

where *G* and *R* are the variance-covariance matrix functions of σ_g and σ_r parameters, respectively. *G* and *R* are square and symmetric, positive definite and non-singular. The vector of the observations *y* has an assumed multivariate normal distribution with mean equal to *Xb* and variance:

$$var(y) = V = ZG(\sigma_g)Z^T + R\ (\sigma_r)$$

When there is only a single random effect, *Z* matrix is an identity matrix where all diagonal entries are equal to 1. If there are additional random terms, *u* can be partitioned as $u = [u_1^T, u_2^T, u_3^T, \ldots u_s^T]$ and *G* matrix results from the direct sum (⊕) of each structure:

$$G = \oplus_{i=1}^{s} G_i = \begin{bmatrix} G_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & G_s \end{bmatrix}$$

Variance of random effect *u* with multiple terms becomes:

$$Var(u) = A\ \sigma_u^2$$

$$\sigma(u_i, u_s) = A\ \sigma_s^2$$

where *A* is the numerator relationship matrix (or kinship matrix) among plant individuals. This matrix is symmetric and can based on pedigree information or on molecular markers. Off-diagonal elements of the matrix are calculated on the genome fraction shared by two individuals. Residuals are usually defined as uncorrelated. In the simplest case, with only one residual effect, *R* matrix is the product between residual variance and identity matrix for the residual term. Covariance of *u* and *e* is zero and variance of *y* becomes:

$$Var(y) = A\ \sigma_u^2 + I\ \sigma_e^2$$

In multilevel LMM, residual error is partitioned*, R* matrix is specified by the direct sum of structures and variance of *y* becomes:

$$var(y) = V = \sigma_u^2 ZAZ^T + I\ \sigma_e^2$$

According with Henderson (1950, 1059, 1963, 1975), when *b, A* and *R* matrices are known, the best linear predictor of *k'b + m'u* is:

$$k'b + m'GZ'V^{-1}(y - Xb)$$

The parameters *b, A* and *R* need to be estimated for σ_g and σ_r and variance estimations ($\widehat{b}$ and $\widehat{u}$, respectively for fixed and random effects) have to be carried out before any statistical inference. Variance components can be estimated with different approaches, such as Restricted Maximum Likelihood (*REML*; Patterson and Thompson, 1971) and Bayesian statistics. Assuming that *V* is known, Henderson (1950, 1959,

1963, 1975) proposed Mixed Model Equations (MME) for estimation of fixed effects and prediction of random effects. Henderson's equations give both the predictors for random effects and the generalized least squares (*GLS*) equations for computing the fixed components. Henderson's equation main advantage consists in not requiring the *V* matrix inversion. Unlike *V* – that can be a large matrix-, inversion of G^{-1} and R^{-1} is less costly. Fixed effects can be estimated by *GLS* equations:

$$\widehat{b} = X'V^{-1}y\,(X'V^{-1}X)^{-1}$$

where $\widehat{b}$ becomes the 'Empirical Best Linear Unbiased Estimator' (E-BLUE) of *b*. An estimator is defined as 'best' when it has the lowest variance and the highest correlation between predicted breeding values (PBVs) and true breeding values (TBVs). In random effects, $\widehat{u}$ is the BLUP of *u* where the prediction of u is unique (Henderson, 1975). When BLUPs refer to breeding values (BVs), they are called PBVs or estimated breeding values (*EBVs*). If EBVs are based on the genotypes, *GEBVs* represent the genetic merit.

4.2.5 PREDICTION ERROR VARIANCE AND RELIABILITY

Prediction error variance (PEV) is useful to evaluate the quality of predictions. It corresponds to the error variances of the difference between *u* and $\widehat{u}$. PEV *c*alculation is very useful in detecting model precision (Heslot *et al.*, 2015; Hardner *et al.*, 2017). Hence:

$$PEV(\widehat{u}) = Var(u - \widehat{u})$$

It derives that when *PEV* is small, individuals' *PBVs* are very similar to *TBVs* (Lynch and Walsch, 2018). Reliability (ρ) of the random effect *u* - or coefficient of determination ratio between *Var (*$\widehat{u}$*)* and *Var (u)* - represents the portion of *u* accounted by *PBVs* (Lynch and Walsh, 2018):

$$\rho = 1 - \frac{PEV(\widehat{u})}{\sigma^2(u)}$$

4.2.6 GENOMIC RELATIONSHIP MATRIX AND GBLUP

Genetic matrix contains the genetic information necessary for computing each individuals' BV. It can be based on pedigree or on molecular markers. Pedigree-based matrix captures the portion of genes shared by relatives and the fraction of alleles identical by descendent (*IBD*) (Myles *et al.*, 2009; Kumar *et al.*, 2013). Pedigree-based matrices seem to be less precise, with assignments of missing connections among relatives and wrong attributed-relationships. These lead to a reduced hereditability estimation (VanRaden, 2007; VanRaden, 2008; Hayes *et al.* 2009). Molecular markers-based genomic relationship matrix (GRM) contains the fraction of alleles shared by different individuals or the alleles fraction in a QTL shared by two genotypes (VanRaden, 2007; VanRaden, 2008). The method for estimating markers effect is also called SNP-BLUP (Moser *et al.*, 2009). Basing on the assumption that alleles have an additive effect, GRM computes variances and covariance among genotypes over all molecular markers in an approach termed genomic BLUP (GBLUP)

(Hayes *et al.* 2009; Daetwyler *et al.* 2010; Heslot *et al.* 2012; Lin *et al.* 2014; Kumar *et al.*, 2015). Molecular markers-based matrix is more informative because includes the Mendelian sampling not considered in pedigree-based matrices (Daetwyler *et al.*, 2007; Hill and Weir, 2011; Heslot *et al.*, 2015). The genetic material distributed from parents to offspring can be obtained by parents' pedigree, but the deviation of progeny performances from the population mean range depends on the chromosomes number and on map length (Hill and Weir, 2011). GP with a molecular markers-based matrix predicts *BVs* that include both parents' effects (i.e. the *BVs* mean between the parents used in controlled-cross) and seedlings deviation from them (Heslot *et al.*, 2015; Crossa *et al.*, 2017). GRM built following the first VanRaden's method becomes:

$$GRM = \frac{ZZ'}{2\sum p_i\,(1-p_i)}$$

where p_i is the frequency of the second allele for the *i* locus. The denominator scales the GRM making it similar to the kinship matrix *A*. *GEBVs* are effectively computed across loci:

$$GEBV = \sum_{i=1}^{i} Z\hat{u}_i$$

where $X\hat{b}$ are the solutions for the mean of *y* and *R* is the variance-covariance matrix of residual effects. The second method was introduced for human studies and accounts for different markers weight on the target trait (Leutenegger *et al.*, 2003; Amin *et al.* 2007). *GRM* is obtained as follows:

$$GRM = \frac{ZDZ'}{n}$$

where *n* is the number of SNPs, *D* is the diagonal matrix of weights calculated as:

$$\frac{1}{2\,p_i\,(1-p_i)}$$

The third method differs from the others in computing GEBVs through the GRM inversion (Garrick, 2007), as follows:

$$\hat{u} = \left[R^{-1}GRM^{-1}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right]\text{^}{-1}\,R^{-1}\left(y - X\hat{b}\right)$$

4.2.7 PREDICTION ABILITY AND PREDICTION ACCURACY

Prediction ability (*PA*) is the Pearson's correlation between the vector of GEBVs and TBVs (Gezan *et al.*, 2017). Because TBVs can be only estimated, phenotypic values adjusted for fixed effects estimation replace *TBVs* (Lorenz *et al.*, 2011; Lin et l., 2014; Gezan *et al.*, 2017). Prediction accuracy (*PACC*) is the ratio

between the correlation of GEBVs -TBVs and the square root of narrow sense hereditability (h^2) calculated for the specific validation dataset. *PACC* indicates the GEBVs prediction accuracy to selection response (Falconer and Mackay, 1996; Lorenz *et al.*, 2011) and it is greatly useful in predicting the performance of future seedlings generation without phenotyping it. *PACC* ranges from 0 to 1 and more the value is close to 1, higher is the predictability of the developed model. Also, *PACC* is useful in comparing GP models and choosing the best one in data fitting (Falconer and Mackay, 1996; Lorenz *et al.*, 2011; Daetwyler *et al.*, 2013). In addition, high *PA* and *PACC* are desirable for selecting candidate parents used in planning new crosses.

Many aspects affect *PA* and *PACC* (Goddard, 2009; Wurschum *et al.*, 2013; Desta and Ortiz. 2014):

- Reduced size of the reference population decreases the accuracy in the predictions. GP models are developed from genotypic and phenotypic information in the analysed reference population. Larger population sizes increase the opportunity to reliably estimate the alleles effect, also at low frequencies, being the inbreeding rate less frequent and genetic drift absent (Falconer and Mackay, 1996; Lin *et al.*, 2014; MacLeod *et al.*, 2016).
- Quality of the phenotypic records collected. Phenotyping can limit the achievement of reference population adequate size, especially in fruit tree crops where collecting large high-quality phenotypic data is still difficult to achieve, in contrast to model plant species.
- Relatedness among individuals belonging to the reference and testing population. Based on the *PACC* equation for GBLUP-based prediction models, the relatedness is calculated as:

$$r = \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}}$$

where N_p is the total amount of individuals in reference population, h^2 is the narrow-sense hereditability of target trait and M_e is the effective number of independent chromosome segments (Goddard, 2009; Daetwyler *et al.*, 2010). In population genetics, M_e can be estimated as two times the product between the effective population size and the genome length expressed in Morgans (Hayes *et al.*, 2009). Therefore, individuals in the *GRM* should be closer genetically, leading to longer haplotypes shared between reference and validation populations and lower quantity of independent chromosome segments (Goddard, 2009; Daetwyler *et al.*, 2010; Lin *et al.*, 2014).

- Density of the molecular markers. When markers amount is larger, more probable is capturing genetic variants widely spread along the genome and dissecting the genetic architecture of target traits. Large-density genotyping with SNPs has increase the possibility of discriminating between rare and causal variants for the target trait, providing more significant associations between traits and genomic variance and thus more accurate GEBVs (Daetwyler *et al.*, 2012).

- *LD.* In absence of equilibrium there is a non-random association between alleles at two loci. If molecular markers and target QTL are in *LD*, they are associated. A high correlation between them implies that marker effects continue across different population and in the future generations.
- Genetic architecture of target traits. *PACC* is influenced by the number of loci controlling the target trait because it is assumed that marker effects are normally distributed in GP models. If kurtosis occurs, *PACC* decreases (Meuwissen *et al.*, 2001; Habier *et al.*, 2007).
- Hereditability of the character of interest. Higher is h^2, more accurate is GEBVs prediction because dominance, epistasis and non-genetic effects less affect the target trait expression level (Kumar *et al.*, 2015). Seedlings performances is more stable across the years, with higher correlation between additive genetic variance and phenotypic variance.

4.2.8 STATISTICAL APPROACHES ADOPTED FOR DEVELOPING GP MODELS

SNPs effects are generally the random component in LMM and are all summed. Several GP models try to overcome the "large *p* small *n*" problem and to estimate more precise BVs (Meuwissen *et al.*, 2001; Heffner *et al.*, 2010). Several approaches can be used in developing GP models, that differ in the underlying assumptions of random effects. When the phenotypic records amount is lesser than the number of molecular markers, GLS estimates are poorly predictive because markers are treated as fixed effects and data suffer for multicollinearity. This differs from regression models now commonly adopted in GS studies (Meuwissen *et al.*, 2001; Hayes *et al.* 2009; Daetwyler *et al.* 2010; Heslot *et al.* 2012; Desta and Ortiz, 2014; Lin *et al.* 2014). Regression models are divided into parametric and non-parametric (Desta and Ortiz, 2014).

Ridge Regression Best Linear Unbiased Predictor (RR-BLUP)- based approach assumes that all marker effects follow a normal distribution and have a common variance (Meuwissen *et al.*, 2001; Habier *et al.*, 2007), leading to an equally shrunken-predictor effects toward zero using λ parameter (Meuwissen *et al.*, 2001; Jannink *et al.*, 2010; Heffner *et al.*, 2011). G-BLUP based approach belongs to RR-BLUP group (Habier *et al.*, 2007; VanRaden, 2008).

Bayesian statistics –based methods are another class of parametric methods and assume that each marker effects follow a normal distribution but without fixed variance. Bayesian approaches require algorithms as Markov-Chain-Monte-Carlo (*MCMC*) for computing the parameters (Lorenz *et al.*, 2011). *BayesA* assumes that each marker has its own variance and shrink toward zero each markers effect at different degree (Xu, 2003; Lorenz *et al.*, 2011). *BayesB* adopts that every marker has its own variance with a Student's *t*-distribution but a part of them has no effect (Meuwissen *et al.*, 2001). *BayesB* can be useful when the variance is not presence at every *locus*. In *BayesC*, molecular markers have a single variance equal for all the SNPs (Habier *et al.*, 2011). As for *BayesB* some *SNPs* have a probability π of not having effect while the remaining have a probability *1-* π of non-zero effect. Parametric models can capture only additive genetic effects while

non-parametric models can evaluate also non-additive genetic effects. Non-parametric approaches are less adopted being not so efficient in computing the genetic gains in future generations of crosses, especially under the not full inheritance of non-addittive effects by the progeny (Lin *et al.*, 2014).

Several extensions of each approach exist, and some can fit better the data than others, according with the association between molecular markers and target traits (Daetwyler *et al.*, 2010; Clark *et al.*, 2011; Zhao *et al.*, 2013). For example, weighted single step GBLUP (wssGBLUP) is an extension of the classic GBLUP method and is more flexible because - similarly to Bayesian statistics- builds GRM attributing different weight on each molecular marker variance (Teissier *et al.*, 2018; Oget *et al.*, 2019). In ssGBLUP-based strategy - where each SNP capture the same phenotypic variance amount-, GRM can be built following VanRaden's method.

In wssGBLUP, GRM *(GRM_w)* developed by Wang *et al.* (2012) becomes:

$$GRM_w = \frac{ZDZ'}{2\sum p_i\,(1-p_i)}$$

where *D* is added and corresponds to the diagonal matix where each diagonal elements is the SNP weight. WssGBLUP can attribute more weight on SNPs having greater effect on target traits, reflecting better the genetic architecture of the character of interest (Teissier *et al.*, 2018; Oget *et al.*, 2019).

4.3 MATERIALS AND METHODS: PEACH

4.3.1 PEACH PHENOTYPIC DATASET

The peach panel included 201 individuals between accessions and selections, all recorded for titratable acidity (TA) and ten OAs content (i.e. oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) in two harvest seasons (2017 and 2018). Statistical analysis on TA and OAs content were carried out in *Chapter 2*.

4.3.2 PLANT MATERIAL COLLECTION, DNA EXTRACTION AND QUANTIFICATION

Leaves of 1,190 peach individuals were collected from young shoots during the spring and stored at -80°C until DNA extraction. The individuals' pool included several accessions (a total of 217 unique genotypes), selections (equal to 94 unique genotypes) and seedlings from 4 biparental controlled crosses and one self-cross (population name '11014') (**Table 4.1**). All the plants are maintained at 'Centro Ricerche Produzioni Vegetali' (*CRPV*, *Crop Production Research Centre*, www.crpv.it) located near Imola (North-East of Italy). The leaves were grounded using liquid nitrogen for improving tissue lysis. DNA extraction was performed following the *CTAB* method of Doyle and Doyle (1990). Quality and quantity of DNA extracted was checked by electrophoresis on agarose gel and by dsDNA HS assay kit for Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Peach accessions genotyping was performed through the peach 18K SNP chip and using the peach genome version 2.0 as reference (Verde *et al.*, 2017).

TABLE 4.1. Name of the crosses with the corresponding number of seedlings genotyped for each population.

Population name	Number of peach individuals genotyped
CxEL	247
DxP	191
11014	93
11001	159
11004	189

4.3.3 ALLELE FREQUENCY CALCULATION, DATA FILTERING, IMPUTATION OF MISSING VALUES AND MINOR ALLELE FREQUENCY (MAF) CALCULATION IN THE PEACH GENOTYPES DATASET

4.3.3.1 Allele frequency calculation

Nucleobases (AA, CC, GG, TT) were counted for each SNP (16,028) in the complete peach genotypes dataset (1,190 unique individuals) under the assumption of Hardy-Weinberg equilibrium. Based on the calculated frequencies, the entire genotypic dataset was converted into the bi-allelic codification format where GG and AA were formatted as AA, CC and TT were formatted as BB and heterozygous combinations as AB.

4.3.3.2 Genotypic data filtering

Missing data rate was calculated for each SNP and it was reported along the eight peach chromosomes using *ggplot2* package in *RStudio* software (**Fig. 4.1**).

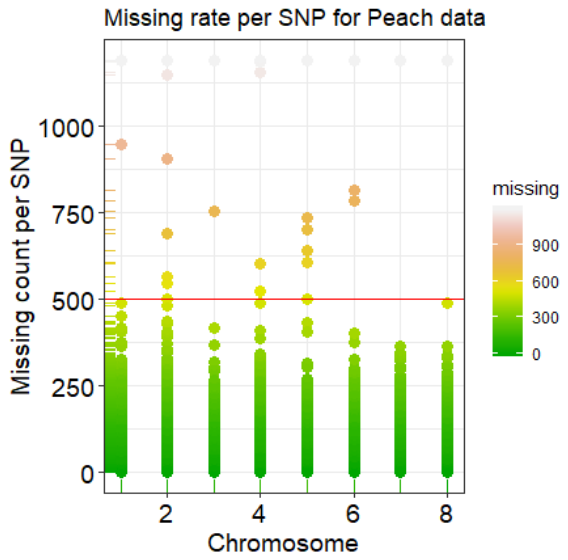

FIGURE 4.1. Missing rate per SNP using the complete peach dataset (1,190 individuals and 16,028 SNPs).

Both SNPs and genotypes with a missing value rate higher than 10% were removed. The reduced peach genotypes dataset contained 9,881 SNPs and 990 individuals. Based on the highest amount of missing data, four SNPs with duplicated positions were removed (**Table 4.2**).

The high-quality genotypic dataset included 9,877 SNPs and 990 individuals genotyped, for a total of 279 among accessions and selections and 711 seedlings of five populations.

TABLE 4.2. SNPs with duplicate positions in the peach genotypes dataset.

SNP	Chromosome	Position (bp)	Missing values
Pp33Cl	1	28,184,376	34
Pp7Cl	1	28,184,376	10
snp_3_130507	3	992,345	6
snp_scaffold_3_130507	3	992,345	5
Peach_AO_0654153	6	17,880,635	87
SNP_IGA_662881	6	17,880,635	78
Pp08_12463247	8	12,463,247	63
SNP_IGA_857951	8	12,463,247	71

4.3.3.3 imputation of missing SNPs using F-IMPUTE software

Missing values in the high-quality peach genotypes dataset were imputed using the *F-Impute* software version 3.0 (Sargolazei *et al.*, 2014). The program required bi-allelic format and missing values coded as 5. Imputation process assumed that individuals more related had longer haplotypes compare to accessions with higher genetic distance. This means that all the samples were related to each other but at different relationship degrees. Chromosome position was sorted within the map position (bp).

4.3.3.4 Minor allele frequency (MAF) calculation

Minor allele frequency (MAF) was calculated for each SNP in the imputed-genotypes file assuming Hardy-Weinberg equilibrium. SNPs with MAF lower than 0.01 were discarded in order to discriminate between common and rare alleles that, despite being at very low frequency, could have larger effects on target traits. The final high-quality peach genotypes dataset contained 7,865 SNPs and 990 individuals.

4.3.4 LINKAGE DISEQUILIBRIUM

Linkage disequilibrium (LD) was calculated on reference population as Pearson's squared correlation coefficient (r^2) between each pair of SNPs (Hill and Robertson, 1968). The distance between pairs of markers was calculated using *PLINK* software (Purcell *et al.*, 2007). Although *PLINK* was first developed for human applications, it can be use also for plants and livestock. It is an open-source toolset useful for several analyses on the whole-genome associations. *Openair* package (version 2.7.0) in *RStudio* software was used for binning the distance between each markers pair every 1000 kb and for each fixed interval the mean r^2 was calculated. Plot of *LD* decay was generated using *ggplot2* package (version 3.2.1) in *RStudio*.

4.3.5 PRELIMINARY ANALYSES ON GRM OF PEACH ACCESSIONS RECORDED FOR FRUIT TITRATABLE ACIDITY AND CONTENT OF MALATE AND CITRATE

4.3.5.1 GRM of reference population, inverse of GRM, plot and quality check of the matrix

Individuals both phenotyped and genotyped (i.e. reference population) were pulled out from the complete high-quality peach genotypes dataset (7,865 SNPs and 990 accessions). Training population contained 194 unique genotypes recorded for peach TA and content of malate (MAL) and citrate (CIT). After dataset transposition, peach genotypes format was converted from *AA*, *AB*, *BB* to -1, 0, 1 respectively, following the guidelines of *sommer* package (version 4.0.9) in *RStudio*. The additive *GRM* (i.e. GRM_{194}) was estimated following the first method reported by VanRaden (VanRaden, 2008) and using *sommer* package. Then, the inverted additive GRM of peach reference population was plotted as a heat map using *ggplots* package (version 3.0.0.3) in *RStudio*.

4.3.5.2 PCoA analysis

Principal Coordinate Analysis (PCoA) was carried out for representing the reference population in two-dimensional space according with the Euclidean distance calculated among them. GRM was converted into a correlation matrix (the diagonal values were equal to 1). Correlation matrix was centred and scaled to obtain a distance matrix where diagonal values were equal to 0 and off-diagonal values were all positives. Multi-dimensional scaling (*MDS*) was performed to decompose the distance matrix into eigen-values. Therefore, the variation amount accounted for each *MDS* axis was calculated and represented in a plot using *ggplot2* package (version 3.2.1) in *RStudio*. To better capture the population structure, peach individuals were attributed to different clusters based on their degree of similarities and dissimilarities using *factoextra* package (version 1.0.6) in *RStudio*. Peach genotypes within the same cluster were more similar to those in other clusters. The adopted cluster algorithm was *k-mean* (Hartigan and Wong, 1979), that required *a priori* definition of clusters number (*K*) in order to partition the analysed reference population. The clustering quality was suggested using 'silhouette' approach, that measured the similarity between each peach individual and its membership cluster compared to other clusters (Rousseeuw, 1987).

4.3.5.3 GRM of peach accessions not recorded for titratable acidity, malate and citrate content

Peach individuals not belonging to the training population (in total 796 genotypes) were selected from the high-quality peach genotypes dataset (7,865 SNPs and 990 accessions). Genotypes were converted into the numeric format. Additive GRM (i.e. GRM_{796}) was estimated following VanRaden' approach (VanRaden, 2008). GRM was inverted and visualized in a heat map using *ggplots* package (version 3.0.1.2).

4.3.5.4 GRM of the complete high-quality peach genotypes dataset

The total amount of individuals genotyped (990 unique genotypes) was used for building the GRM (i.e. GRM_{990}). After dataset transposition, genotypes format was converted from AA, AB, BB to -1, 0, 1 respectively. Additive *GRM* was built following the first method proposed by VanRaden (2008) using *sommer*

package. *GRM* was inverted and its structure was plotted as an heat map using *ggplots* package (version 3.0.0.3) in *RStudio*.

4.3.6 GP MODEL FOR FRUIT TA IN PEACH USING ONLY THE REFERENCE POPULATION GRM

4.2.6.1 Hypotheses of different GP models for titratable acidity

According with the experimental design and general LMM equation, the most complete GP model developed for TA was:

$$TA = \mu_{TA} + Y_{TA} + G_{TA} + G_{TA}:Y_{TA} + P_{TA} + P_{TA}:Y_{TA} + r$$

where μ_{TA} was the overall mean for TA trait (expressed in g/L of malic acid), *Y* was the year of observation (harvest season 2017 or 2018), *G* was the genotypic effect across years, *G:Y* was the effect of each year on *G*, *P* was the non-genetic permanent environment effect on each accession across years, *P:Y* was the effect of each year on *P,* and *r* included the direct sum of residuals. Fixed effects included $\mu_{TA}\ and\ Y_{TA}$ while random effects were $G_{TA} + G_{TA}:Y_{TA} + P_{TA} + P_{TA}:Y_{TA}$. Six different models (**Table 4.3**), nested into the first one, were developed and fitted the data using *ASReml* software (version 4; Gilmour *et al.*, 2009) in order to test the effect of each parameter inside the LMM. *ASReml* is a not open-source package (*VSN International*, www.vsni.co.uk.) implemented for the *R* environment based on the *REML* approach. *ASReml* estimated the variance parameters to solve the LMMs, to be used used for estimating fixed effects and predicting random effects.

TABLE 4.3. Six different LMMs tested for GP of TA trait.

Model number	**Linear mixed model equation**
1	$TA = \mu_{TA} + Y_{TA} + G_{TA} + G_{TA}: Y_{TA} + P_{TA} + P_{TA}: Y_{TA} + r$
2	$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA} + P_{TA}: Y_{TA} + r$
3	$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA} + r$
4	$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA}: Y_{TA} + r$
5	$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA}: Y_{TA}$
6	$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA} + P_{TA}: Y_{TA}$

Model testing was carried out using GRM_{194}. For each model, the significance of fixed and random effects was inferred through Wald's test and likelihood ratio test (Wilks, 1938), respectively. Random effects inference assumed that fixed effects were identical under *REML,* enabling the comparison of changes in fitting among the six models.

4.3.6.2 Estimation of narrow-sense hereditability and non-genetic effects

Variances estimation lead to h^2 calculation as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_P^2 + (\sigma_{P:Y}^2/2) + (\sigma_r^2/6)}$$

where the total variance of *P:Y* was divided by the years number of analysis while residuals components were divided by the product between the years number of observation and the three biological replicated- juices for each peach genotype. Then, variance components of *P* across years and the influence of each year on *P* were estimated fitting the data with the following LMM (genetic effects were excluded):

$$TA = \mu_{TA} + Y_{TA} + P_{TA} + P_{TA} : Y_{TA}$$

where μ_{TA} is the overall mean for TA trait (expressed in g/L of malic acid), *Y* is the year of observation and P_{TA}*:* Y_{TA} is the effect of each year on *P*. Fixed effects include $\mu_{TA} + Y_{TA}$ while random components are $P_{TA} + P_{TA} : Y_{TA}$. TA values predicted over the samples were averaged between the two harvest seasons to estimate *PA* and *PACC*.

4.3.6.3 Cross-validation of the reference population

Cross-validation in GS approach is a pivotal step for training the GP model (Kumar *et al.*, 2012*a*; Biscarini *et al.*, 2017; Gezan *et al.*, 2017; Piaskowsky *et al.*, 2018; Hardner *et al.*, 2019).

Cross-validation can help to discriminate the best model basing on *PA* and *PACC* calculation and determine the model capability in fitting data. Cross-validation was performed randomly partitioning the reference population (194 unique genotypes) into 5-groups ("folds") of equal size (four groups contained 39 and one 38 peach accessions). One-fold was removed across all years and the remaining four folds were used as reference population in GP model fitting, estimating variance components and predicting individuals' BV. Cross-validation loop was done discarding in turn one of the five folds. *PA* and *PACC* were calculated for each loop.

4.3.6.4 Population structure: hierarchical cluster analysis and dendrogram

A total of 279 among peach accessions and selections were taken from the high-quality genotypes dataset to explore the individuals' structure and parents' distribution inside the analysed dataset. The main objective was allocating the eight parents of the four biparental crosses and the one self-crossed inside the population structure. The final purpose was investigating if the parents were genetically distant in order to associate the large observed variability observed for fruit TA, MAL and CIT with a genetic basis. Except of one parental lost, the remaining eights ('391C12XXXIV86', 'Contender', 'Elegant Lady', 'Dulcebo', 'Pulchra', 'BO96028059', 'BO95009009', 'BO06013049') were genotyped. The remaining accessions were divided in two groups: one included peach accessions both genotyped and recorded for TA, MAL and CIT and the other one contained 85 individuals only genotyped. Hierarchical cluster analysis was carried in *RStudio*. The

dissimilarity structure was obtained with a complete method and the parents' position inside the dendrogram was visualized using *dendextend* package (version 1.3.4) in *RStudio*.

4.3.7 GP MODEL FOR MALATE CONTENT IN PEACH USING ONLY THE GRM OF REFERENCE POPULATION

4.3.7.1 Hypotheses on different GP models

Based on the experimental design, six LMMs were tested for malate content (MAL) (**Table 4.4**) using *AsREML* package. First model was the most complete while the others were nested into the first one. Fixed and the random effects were estimated from the model fitting and each nested model was compared in a pairwise manner through the likelihood ratio test (Wilks, 1938).

TABLE 4.4. Six different *LMMs* tested for MAL content.

Model number	Linear mixed model equation
1	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + G_{MAL}: Y_{MAL} + P_{MAL} + P_{MAL}: Y_{MAL} + r$
2	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL} + P_{MAL}: Y_{MAL} + r$
3	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL} + r$
4	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL}: Y_{MAL} + r$
5	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL}: Y_{MAL}$
6	$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL} + P_{MAL}: Y_{MAL}$

4.3.7.2 Estimation of non-genetic effects and narrow-sense hereditability, cross-validation, PA and PACC calculation

Estimation of non-genetic effects, h^2, *PA* and *PACC* for MAL was performed following the same steps done for TA. Cross-validation was performed dividing the accessions into 5-fold cross equally-sized, as for fruit TA trait. Cross-validation loop run for five times and each loop released the estimated PA and PACC values.

4.3.8 GP MODEL FOR CITRATE CONTENT IN PEACH USING ONLY THE REFERENCE POPULATION GRM

4.3.8.1 Hypotheses on different GP models

Based on the experimental design, six LMMs were tested for citrate content (CIT) (**Table 4.5**) using *AsREML* package in *RStudio*, similar to fruit TA and MAL analysis. First tested model was the most complete while the others were nested into the first. Fixed and random effects were estimated from the model fitting and each nested model was compared in a pairwise manner through the likelihood ratio test (Wilks, 1938).

TABLE 4.5. Six different *LMMs* tested for CIT trait.

Model number	Linear mixed model equation
1	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + G_{CIT}: Y_{CIT} + P_{CIT} + P_{CIT}: Y_{CIT} + r$
2	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + P_{CIT} + P_{CIT}: Y_{CIT} + r$
3	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + P_{CIT} + r$
4	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + P_{CIT}: Y_{CIT} + r$
5	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + P_{CIT}: Y_{CIT}$
6	$CIT = \mu_{CIT} + Y_{CIT} + G_{CIT} + P_{CIT} + P_{CIT}: Y_{CIT}$

4.3.8.2 Estimation of non-genetic effects and narrow-sense hereditability, cross-validation, PA and PACC calculation

Estimation of non-genetic effects, h^2, *PA* and *PACC* for citrate content was performed following the same steps done for TA and MAL traits in peach. Cross-validation was performed dividing the peach accessions into 5-folds cross equally-sized. Cross-validation loop run for five times and each loop released the estimated PA and PACC values.

4.3.9 GS ANALYSIS FOR MULTI-TRAITS: FRUIT TITRATABLE ACIDITY AND MALIC ACID CONTENT

After fitting the model for each trait separately, *GP* approach was extended simultaneously to MAL and TA traits. Fitting multiple traits improves accuracy of predictions, especially if target traits are correlated, and can be more helpful for breeders to select the best candidates for two traits simultaneously. Based on the high correlation observed between TA and MAL in the peach germplasm collection (ρ of 0.82 in *Chapter 2*), multi-traits analysis was performed. The chosen GP model for each trait was:

$$TA = \mu_{TA} + Y_{TA} + G_{TA} + P_{TA} + P_{TA}:Y + r$$

$$MAL = \mu_{MAL} + Y_{MAL} + G_{MAL} + P_{MAL} + P_{MAL}:Y + r$$

In multi-traits analysis, the genetic effect (*G*) became a variance-covariance matrix:

$$G_{TA,MAL} = \begin{bmatrix} \sigma^2_{G\,TA} & \sigma_{G\,TA,MAL} \\ \sigma_{G\,MAL,TA} & \sigma^2_{G\,MAL} \end{bmatrix}$$

The non-genetic effect (*P*) was calculated as:

$$P_{TA,MAL} = \begin{bmatrix} \sigma^2_{P\,TA} & \sigma_{P\,TA,MAL} \\ \sigma_{P\,MAL,TA} & \sigma^2_{P\,MAL} \end{bmatrix}$$

Residuals were:

$$r_{TA,MA} = \begin{bmatrix} \sigma^2_{r\,TA} & \sigma_{r\,TA,MA} \\ \sigma_{r\,MA,TA} & \sigma^2_{r\,MA} \end{bmatrix}$$

In multi-traits analysis, phenotypic records constituted the fixed effects while random effect was represented by a correlation matrix with heterogeneous variances. Residuals were obtained as the direct products of the combination between year and the unique number index given to each phenotypic record and the specific trait. The direct product specified a different variance structure for each random term.

4.3.10 ALLELE WEIGHTS CALCULATION FOR FRUIT TITRATABLE ACIDITY IN THE PEACH COLLECTION

Alleles weight calculation was performed in order to detect possible locus/loci with a larger effect on TA, despite of *Locus D* on the 5th chromosome. Computation of TA values for each peach individual was performed by back solving the LMM equation. The estimated effect of each *SNP* ($\hat{a}$) on TA trait was estimated by converting GEBVs calculated in GBLUP-based approach (Wang *et al.*, 2012; Pryce *et al.*, 2015). Equation of GEBVs conversion into SNP coefficients was

$$\hat{a} = H'(HH')^{-1}\ \hat{u}$$

where *H* is the matrix containing the number of individuals in the reference population (194 unique genotypes) for the number of *SNPs* (expressed in the 0, 1, 2 format) and $\hat{u}$ is the vector of GEBVs only genotyped (i.e. testing population) under ssGBLUP procedure. GEBVs estimation (vector $\hat{u}$) for testing population was:

$$\hat{u} = H_{ref}\ \hat{a}$$

where H_{ref} is the GRM_{194} of the 194 individuals belonging to the reference population. SNP coefficients were computed fitting the model with *ASReml* software. The weight of each allele was derived as follows:

$$\hat{a} = \frac{H'(HH')^{-1}\ \hat{u}}{2N}$$

where *N* is the number of SNPs (7865). Each allele effect was represented in a Manhattan plot using *qqman* package (version 0.1.4) in *RStudio*.

4.4 RESULTS

4.4.1 LD ANALYSIS

LD decreased of 50% around 200 kb and, then, declined slowly until 1000 kb **(Fig. 4.2)**.

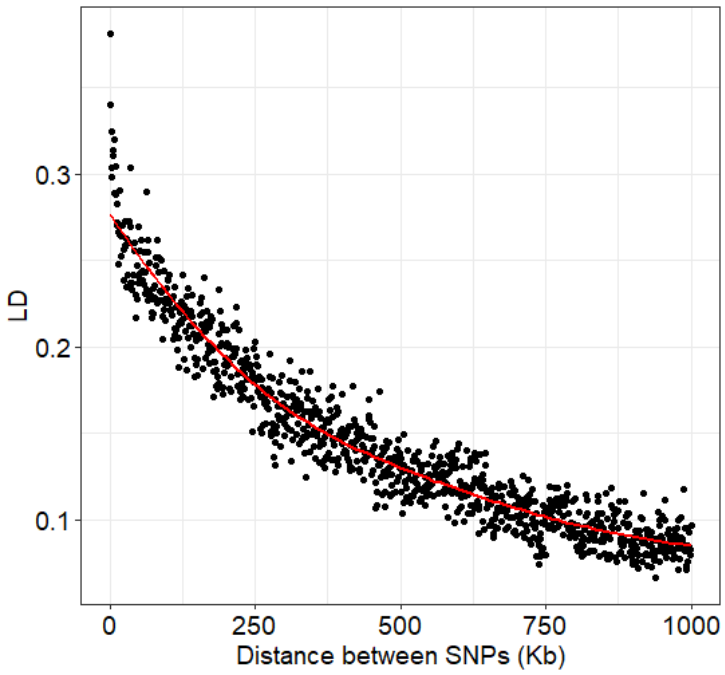

Fig. 4.2 Decay of *LD* in a large high-quality peach genotypes collection. LD declined of the 50% around 200 kb and continued decreasing slowly until 1000 kb.

4.4.2 POPULATION STRUCTURE AND DENDROGRAM OF PEACH ACCESSIONS AND SELECTIONS

Population structure was analysed using the three different GRMs built using 194 (i.e GRM_{194}), 796 (i.e. GRM_{796}) and 990 peach individuals (i.e. GRM_{990}) (**Fig. 4.3 A**, **Fig. 4.3 B** and **Fig. 4.4**, respectively). Reference population structure was almost absent, including peach accessions largely heterogeneous in their descent. When seedling of five crosses were included in GRM_{990}, each population clustered showing a structured-genotypic dataset (**Fig. 4.4**). This evidence was also proved by the parents' position inside the dendrogram (**Fig. 4.5**).

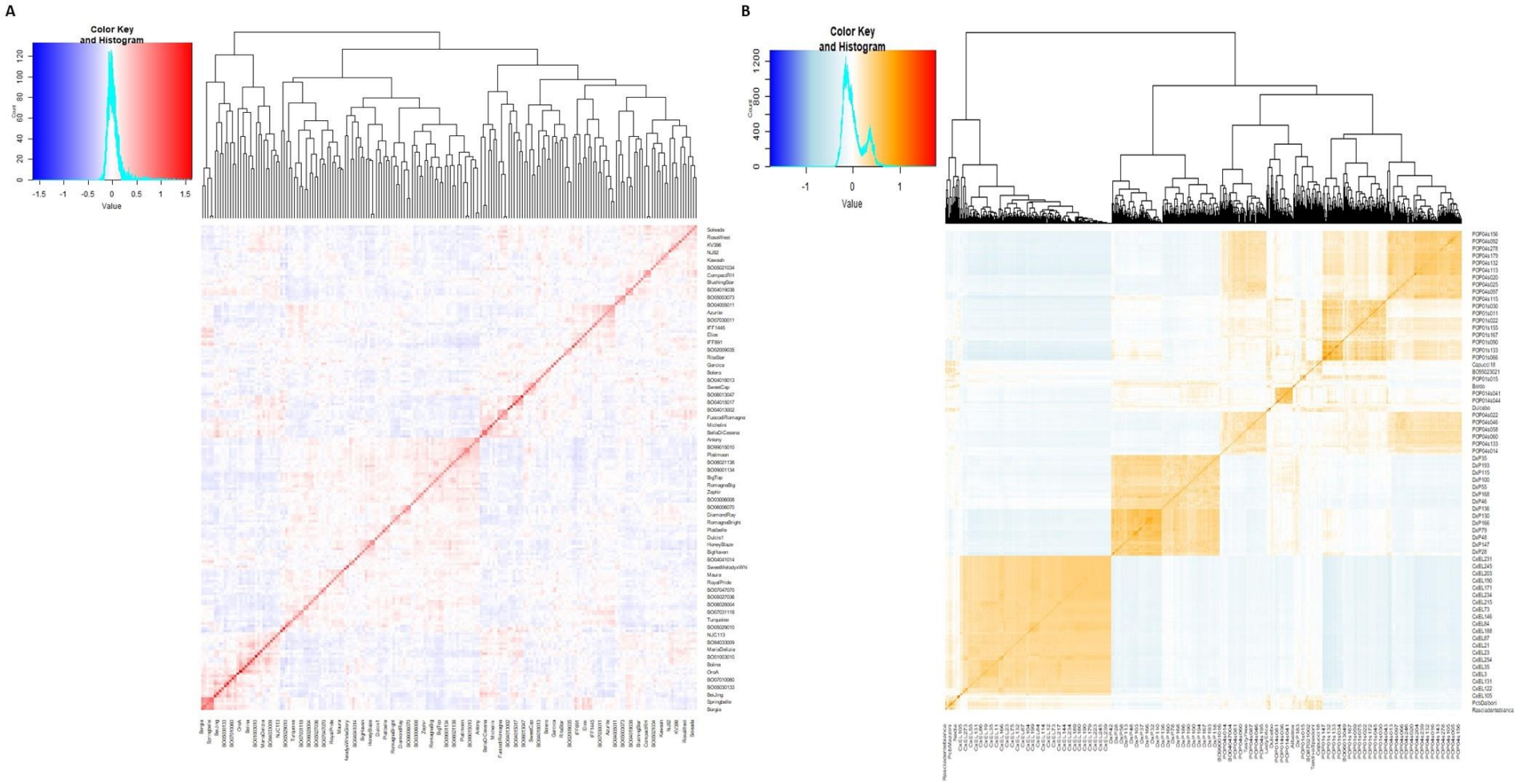

FIGURE 4.3. **Heat maps of the *GRM* made by 194 individuals (*GRM_{194}*) belonging to the reference population (A) and 796 individuals only genotyped (B; *GRM_{796}*).** Heat maps showed the pairwise similarities among the accessions. GRMs were computed following the first method proposed by VanRaden (2008). Relatedness degree (from -1.5 to 1.5) is shown as a colours range. The highest positive values are close to the red while the lowest values are near to the blue. Clustering dendograms are reported on the top of each heat map.

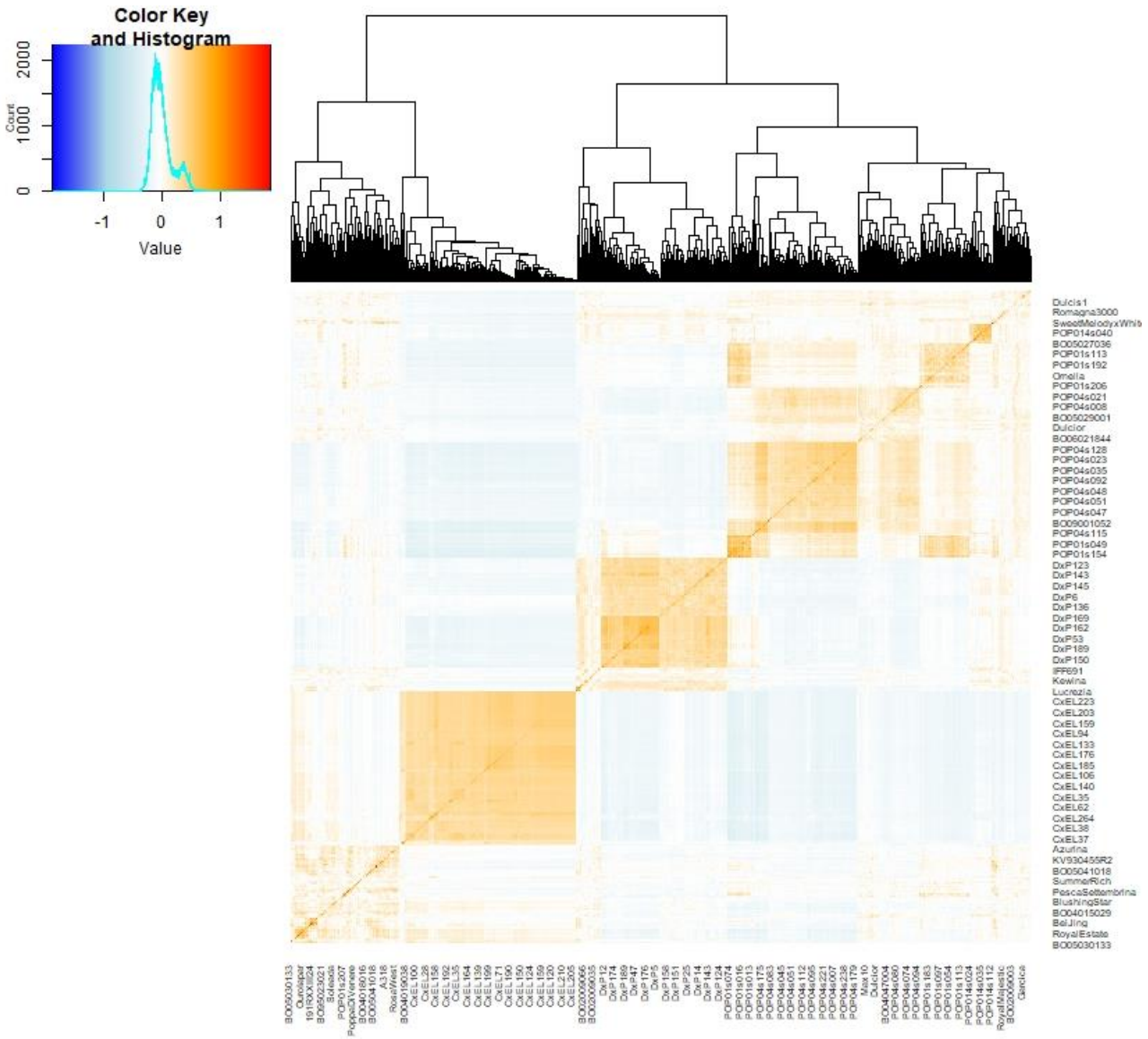

FIGURE 4.4. **Heat map of the GRM made by all the individuals (990 in total) having high-quality genotypes.** The heat map shows the intensity of the values among the peach individuals as a range of colours. The highest positive values are close to the red while the lowest values are near to the blue. Five crosses were easily distinguished within the peach collection.

4.4.3 PCO ANALYSIS

The first two PCo axes (i.e. *PCoA 1* and *PCoA 2*) accounted for the 39.88 % of total genetic variation observed in the training population. The best result was obtained by dividing the totality of peach accessions into three clusters. Peach individuals within the same cluster showed a higher similarity degree (**Fig. 4.6 B**).

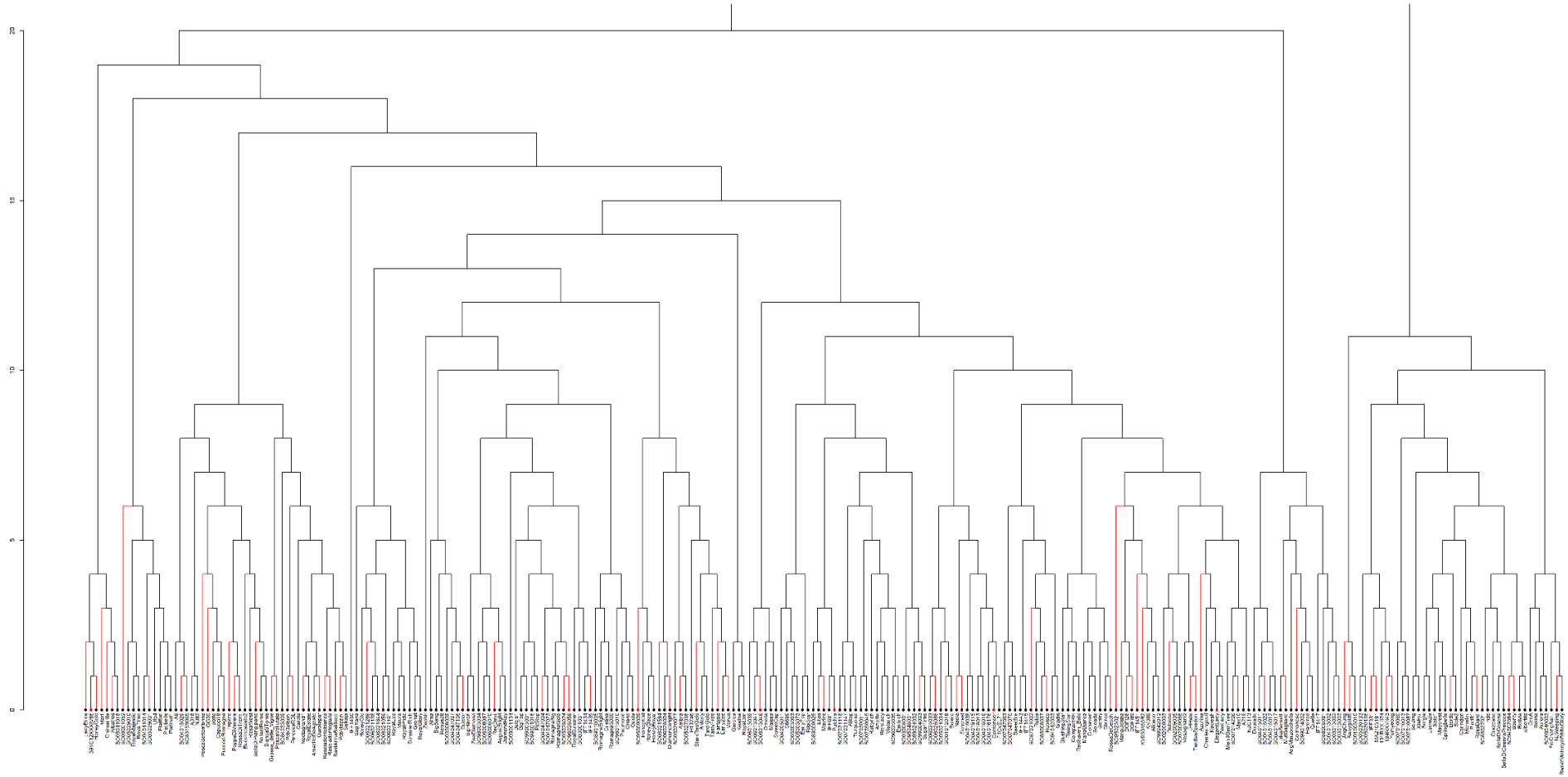

FIGURE 4.5. **Dendrogram of the 279 accessions/selections in the peach population.** The total of 279 accessions/selections are displayed in the dendrogram cut at height 20 for a clearer visualization of the structure. The branch colour is black when the individual is phenotyped for *TA* trait and is red when there are no records for TA. Then, the red nodes indicate the parents (‘*391C12XXXIV86*’, ‘*Contender*’, ‘*Elegant Lady*’, ‘*Dulcebo*’, ‘*Pulchra*’, ‘*BO96028059*’, ‘*BO95009009*’, ‘*BO06013049*’) of the unrecorded crosses population.

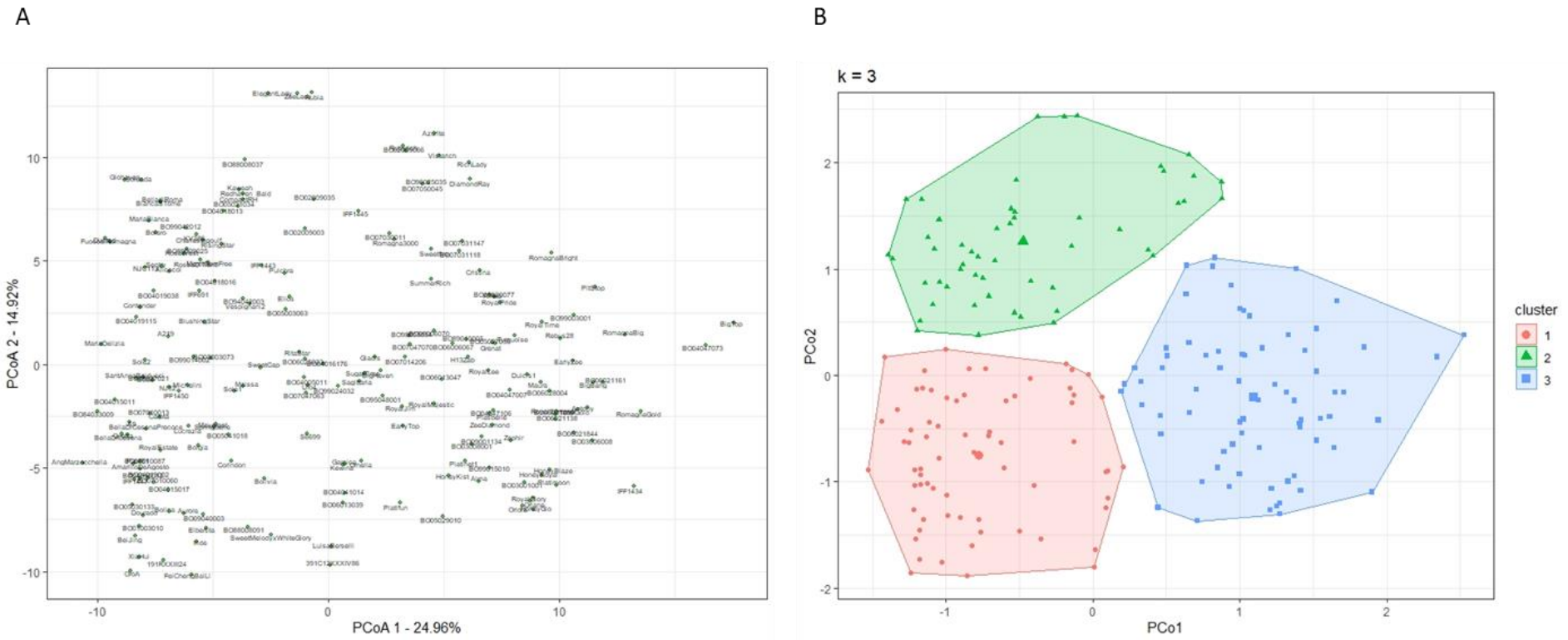

FIGURE 4.6. **PCo (A) and cluster (B) analysis on the reference population (194 high-quality peach genotypes).** The first two axes explained the 39.88% of total genetic variance among peach accessions in the reference population (**A**). The clustering, based on the (dis)similarity degree among peach accessions, split the reference population in three groups. Peach individuals belonging to the same group were more similar than those located in different clusters.

4.4.4 GP MODELS ON FRUIT TITRATABLE ACIDITY TRAIT

Based on the likelihood ratio test performed for every nested model (**Table 4.6**), the second model (nested into the first and without *G:Y* component) significantly fitted better the TA data than the first one (*p-value* < 0.01) and the others. This GP model was adopted for predicting TA values of testing population (796 peach genotypes). The estimated h^2 was 0.93, suggesting that TA trait was strongly genotype-dependent and lesser affected by non-genetic effects. At first, PACC – estimated only cross-validating the reference population- was on average of 0.7. A strong positive correlation (equal to 0.81) existed between MAF values calculated on reference population and the entire high-quality genotypes dataset (990 genotypes), separately. Also acidity values predicted from model fitting using GRM_{194} and GRM_{990}, separately, were highly correlated (r= 0.99 with a *p-value* lower than 0.001).

TABLE 4.6. Summary of the variances calculated for each random component in the six GP models for TA trait with the corresponding standard errors (i.e. *Std. Error*) and maximum Log-likelihood. *G*, genotypic effect across years; *G:Y*, effect of each year on *G*; *P*, non-genetic permanent environment effect on each accession across years; *P:Y*, the effect of each year on *P* and *r*, direct sum of residuals.

Model number	Random components in GP models								Maximum Log-likelihood value
	G	P	G:Y	P:Y	r(2017): units	r(2018): units	r(2017)	r(2018)	
1	6.38	0.14	0.39	0.64	0.22	0.40	1	1	-424.13
Std. Error	1.24	0.38	0.29	0.21	0.07	0.03			
2	6.7	0.02		0.91	0.23	0.40	1	1	-425.91
Std. Error	1.27	0.39		0.17	0.08	0.03			
3	6.7	0.81			1.82	0.41	1	1	-448.27
Std. Error	1.32	0.40			0.27	0.03			
4	6.75			0.91	0.23	0.40	1	1	-425.91
Std. Error	0.95			0.16	0.08	0.03			
5	6.81			0.86			1	1	-427.23
Std. Error	0.96			0.16					
6	6.73	0.03		0.85			1	1	-427.23
Std. Error	1.27	0.39		0.17					

4.4.5 GP MODELS ON PEACH FRUIT MALIC ACID CONTENT TRAIT

Variance components for each random factor and maximum Log-likelihood were estimated in the nested models developed for MAL predictions, suggesting that the second model increased the PACC (*p-value* < 0.01) (**Table 4.7**). The estimated h^2 was 0.79. PACC, estimated performing cross-validation using the reference population, was equal to 1. Hence, cross-validation and model fitting was extended to all high-

quality peach genotypes dataset (990 individuals). A perfect positive correlation (equal to 1) existed between malate content values predicted with the GRM_{194} and GRM_{990}, separately.

TABLE 4.7. Summary of the variances calculated for each random component in the six GP models for MAL trait with the corresponding standard errors (i.e. *Std. Error*) and maximum Log-likelihood. *G*, genotypic effect across years; *G:Y*, effect of each year on *G*; *P*, non-genetic permanent environment effect on each accession across years; *P:Y*, the effect of each year on *P* and *r*, direct sum of residuals.

Model number	Random components in GP models								Maximum Log-likelihood value
	G	P	G:Y	P:Y	r(2017): units	r(2018): units	r(2017)	r(2018)	
1	2.01	1.81e-07	1.97e-07	9.43e-01	2.79e-01	2.05e-01	1	1	-258.76
Std. Error	0.39			0.15	0.12	0.02			
2	2.01	1.57e-07		9.43e-01	2.79e-01	2.05e-01	1	1	-258.76
Std. Error	0.51			0.16	0.12	0.02			
3	2.93	0.38			2.12	0.22	1	1	-281.25
Std. Error	0.59	0.19			0.30	0.02			
4	2.01			0.94	0.28	0.21	1	1	-258.76
Std. Error	0.39			0.15	0.12	0.02			
5	1.99			0.97					-259.03
Std. Error	0.39			0.14					
6	1.99	1.50e-06		9.73e-01					-259.03
Std. Error	0.39			0.143					

4.4.6 GP MODELS ON PEACH FRUIT CITRIC ACID CONTENT TRAIT

Citric acid was the second most abundant OA in the analysed peach fruits (i.e. *Chapter 2*). Variance components for each random factor and maximum Log-likelihood were estimated (**Table 4.8**) followed by the likelihood ratio test, that evidenced as the second GP model achieved the highest model fitting accuracy. The estimated h^2 was 0.74, lower than for malate content and TA. Moreover, *PACC* estimated was lower than in MAL trait analysis but similar to TA trait, averaging on 0.7. Comparison of CIT content predictions obtained by fitting the model with the GRM_{194} and GRM_{990}, separately, indicated almost a perfect correlation (equal to 0.99).

TABLE 4.8. Summary of the variances calculated for each random component in the six GP models for CIT trait with the corresponding standard errors (i.e. *Std. Error*) and maximum Log-likelihood. *G*, genotypic effect across years; *G:Y*, effect of each year on *G*; *P*, non-genetic permanent environment effect on each accession across years; *P:Y*, the effect of each year on *P* and *r*, direct sum of residuals.

Model number	Random components in the model								Maximum Log-likelihood value
	G	P	G:Y	P:Y	r(2017): units	r(2018): units	r(2017)	r(2018)	
1	0.59	0.07	0.10	0.29	0.05	0.13	1	1	-61.00
Std. Error	0.18	0.09	0.08	0.07	0.02	0.01			
2	0.66	0.01		0.39	0.05	0.13	1	1	-63.23
Std. Error	0.18	0.09		0.06	0.02	0.01			
3	0.77	0.44			0.60	0.13	1	1	-100.81
Std. Error	0.22	0.11			0.09	0.01			
4	0.67			0.39	0.05	0.13	1	1	-63.24
Std. Error	0.14			0.06	0.02	0.01			
5	0.69			0.38					-66.45
Std. Error	0.15			0.06					
6	0.67	0.02		0.38					-66.43
Std. Error	0.18	0.09		0.07					

4.4.7 MULTI-TRAITS ANALYSIS

The second nested model developed fitted TA and MAL traits simultaneously using the GRM_{990}. Maximum Log-likelihood value was equal to -529.41 and random effects were estimated by fitting the data (**Table 4.9**).

TABLE 4.9. Estimation of maximum Log-likelihood and random effects (with the corresponding standard error) in the multi-trait analysis using GRM_{990}. *TA*, titratable acidity; *MAL*, malic acid content trait; *G:TA*, genotypic effect for TA trait across years; *G:Y*, effect of each year on correlation between MAL and TA; *P:TA*, non-genetic permanent environment effect on each accession across years for TA; *P:TA*, non-genetic permanent environment effect on each accession across years for MAL; *P:Y*, the effect of each year on *P* on for correlation between MAL and TA and *r*, direct sum of residuals.

Random components in GP model	Value	Std. Error
G:TA	5.99	0.90
G:MAL	1.90	0.34
G:Y:cor(MAL,TA)	0.98	0.02
P:TA	1.08	0.19
P:MAL	0.93	0.13
P:Y:cor(MAL, TA)	0.61	0.07
r(Y:Sample:Trait)	1	
r (Y:Sample:cor(MAL, TA)	0.59	0.04
r(Y:Sample:TA)	0.39	0.03
r(Y:Sample:MAL)	0.26	0.03

4.4.8 ALLELE WEIGHTS ESTIMATION FOR TITRATABLE ACIDITY TRAIT

Allelic effects displayed in a Manhattan plot confirmed the *Locus D* on the chromosome 5 as the major QTL controlling peach fruits acidity (**Fig. 4.7**). Furthermore, two other minor loci were observed, respectively on chromosomes 2 and 7.

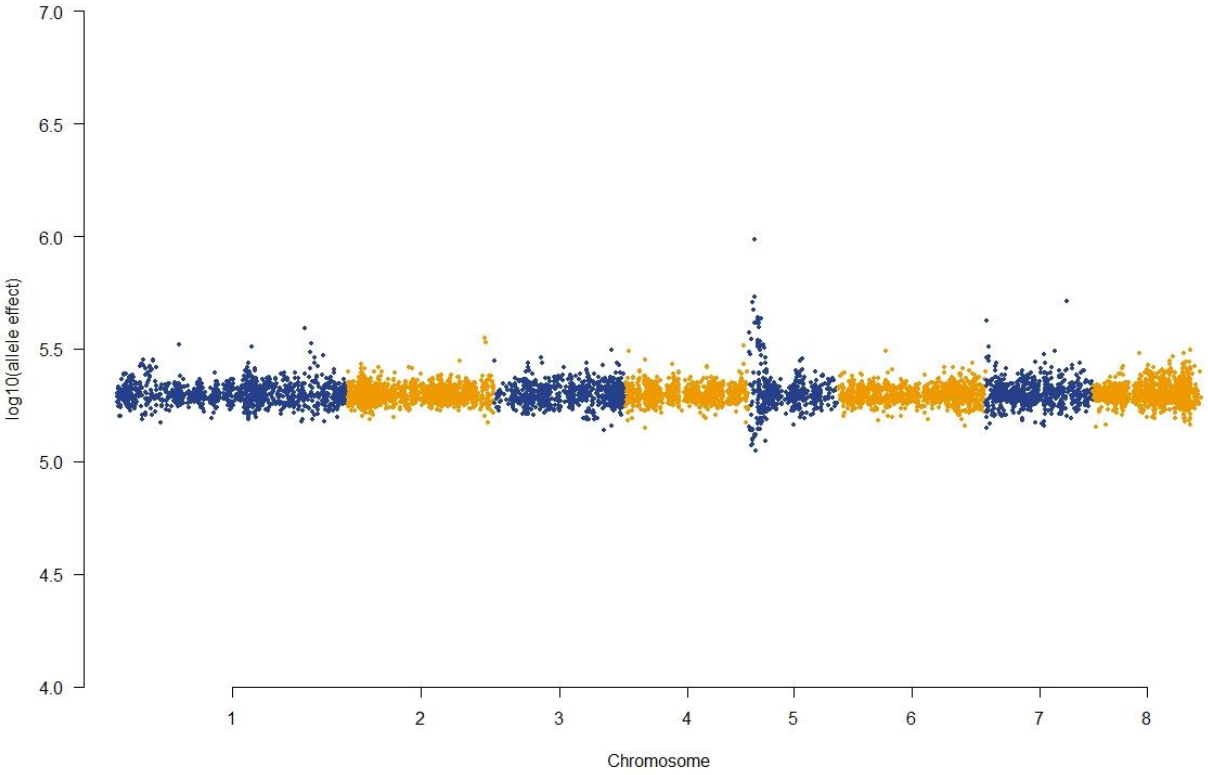

FIGURE 4.7. **Manhattan plot of the estimated allelic weights on acidity trait in the analysed peach collection.** The estimated weight (expressed in log_{10}) of each allele is represented along the peach genome. With the exception of *Locus D* on chromosome 5, other two genomic regions seemed to have an effect on peach fruits TA. These two minor loci were observed on chromosomes 1 and 7.

4.5 DISCUSSION

This chapter reports a GP-based strategy in selecting peach accessions for fruit acidity and content of two OAs, malate and citrate. Malate and citrate were the most abundant OAs described in the analysed peach collection (i.e. Chapter 2), making interesting a possible GP models applicability on peach. Low acidity major locus in peach fruits (*Locus D*) has been already discovered and mapped on chromosome 5, making this trait a target for peach breeding programmes through MAS approach (Dirlewanger *et al.*, 1998; Micheletti *et al.*, 2005). Unfortunately, this QTL partially explain the acidity range of variation observed in peach varieties grown worldwide. In exploring large peach populations, the main bottleneck is collecting phenotypic records for fruit quality related traits. In fact, phenotyping in peach, coupled with the relative

long growing period (around 3-4 years) and the need of repeated data measurements over years (Topp *et al.*, 2008; Biscarini *et al.*, 2017), has limited the breeding programmes advances. In this scenario, GP-based approach could speed up the selection of improved fruit quality seedlings in peach. The feasibility of GS was already partially investigated in peach (Biscarini *et al.*, 2017) but without considering OAs, strongly related to the overall fruit titratable acidity (i.e. *Chapter 2*).

In this thesis, GP models for fruit acidity and content of malate and citrate were developed and tested on a large peach collection (194 accessions characterized for TA and OAs pattern and a total of 1,190 accessions genotyped). GP models considered a permanent environmental effect, where the environmental effect was related only to the seasonality influence rather than a complete different multi-sites effect. GP-strategy based on ssGBLUP method, where GBLUPs were used for GPs and estimation of narrow sense hereditability (h^2), prediction ability and prediction accuracy.

To increase the GP accuracy, population structure should be accounted (Zhong *et al.*, 2009; Fodor *et al.*, 2014; Zhang *et al.*, 2019). *LD* extent analysis revealed a relatively rapid decay in tested peach collection, probably related to the peach collection breeding history. LD was already widely investigated in peach, also by comparing *LD* decay between Oriental and Occidental germplasm (Li *et al.*, 2013; Micheletti *et al.*, 2005). Genetic relationships among individuals of reference population was explored through PCoA, indicating three main clusters. The distribution inside the dendrogram of the eight genotyped parents of the five crosses confirmed the peach reference population heterogeneity.

Additive-genetic relationship matrix was built and used as random effect in the second mixed-model to fit fruit TA, MAL and CIT data. The analyses performed only considered additive genetic effect, mostly because the separation between additive and non-additive genetic effects required several peach genotype replications in different environments. In this study, peach individuals were all grown on the same rootstock and in the same orchard, making possible to consider only permanent environmental effect.

Prediction accuracy of GP model for TA averaged on 0.7, in contrast with the very high narrow-sense hereditability estimated (0.93). A possible reason could be related to the major and dominant effect of *Locus D*, underlying the peach fruit low acidity. TA values distribution of reference population seemed to not follow a discrete distribution (i.e. *Chapter 2*), peculiarity of many other fruit attributes under QTLs control as peach pulp texture (Peace *et al.*, 2005). The achieved result suggested that other possible small effect genes could influence the final observed fruit acidity in peach. Allele weights estimation further evidenced the presence of other two minor loci likely acting on peach fruit acidity. The first was identified on chromosome 1 while the second on chromosome 7. Further GWAS should be carried out for confirming the proposed oligo-genic architecture of acidity trait in peach. This could lead to achieve higher GPs accuracy. Then, the work performed represented the first effort in trying to include fruit MAL and CIT in peach GS. The estimated h^2 (on average 0.75) and the provided GP accuracy evidenced the great gain for peach selection in considering

these traits. MAL analysis showed a very high prediction accuracy (equal to 1), proving this trait as an easy breeding target for improving the peach fruit quality. MAL showed an accuracy higher than TA and CIT traits, indicating an achieved superior genetic gain per unit of time in GP-based screening.

Based on the great results obtained for fruit MAL trait, multi-trait analysis was performed with TA trait. Multi-analyses traits is generally considered more accurate in predictions, especially when traits are correlated (Heslot *et al.*, 2015). Moreover, this approach seems very cost-effective when GP accuracy is high, obtaining genomic predictions for two traits simultaneously (Heslot *et al.*, 2015).

To improve the genomic predictions for TA, a possible approach concerns the wssGBLUP, that assumes different SNPs weight on target trait. This approach seems to be more flexible, similar to Bayesian statistics (Teissier *et al.*, 2018; Oget *et al.*, 2019). With wssGBLUP, GRM is built in a different way from VanRaden's method (VanRaden, 2008), including each SNP weight into the matrix diagonal (Teissier *et al.*, 2018; Oget *et al.*, 2019). Further molecular studies seem necessary for further elucidate the TA genetic architecture in in peach.

GP-based selection seems advantageous in peach breeding, for many aspects. The possibility of shortening up the generation intervals and estimating the seedling performance earlier than MAS are only few of the aspects contributing to reduce phenotyping costs. GPs strategy has been mainly used to select candidate parents for future controlled crosses but GS with more accurate GP may be useful to predict TA, MAL and CIT content in future seedlings generation, also with a guaranteed repeatability of the GP models developed.

4.6 CONCLUSIONS

The analyses done in this chapter represents the first efforts in applying GP models not only on peach fruit acidity but also on OAs content. The obtained results seemed promising for more deep studies for fully characterizing fruit acidity in peach at the genetics' level and supporting peach breeding programmes. The possible repeatability may become very useful in shortening the time, bringing benefits to the global peach fruit production. Surely, the GP modelling performed in this thesis chapter represents only the starting point for further analyses in order to extend GS to other seedlings population and to support the candidate parents' identification in planning new crosses. Improvements of prediction model will be done for making more effective this selection strategy in peach.

5. GENOMIC WIDE ASSOCIATION MAPPING IN A LARGE PEACH COLLECTION

5.1 INTRODUCTION

Peach titratable acidity (TA) and organic acids (OAs) content varies during ripening (Batista-Silva *et al.*, 2018), although was stable at maturity stage across years in the analysed-peach accessions (i.e. 201 unique genotypes; *Chapter 2*). A great variability in OAs qualitative and quantitative was observed among peach varieties, more dependent on genotypic than annual effects (*Chapter 2*). Over years, peach acidity has been investigated leading to a differentiation between low-acid (LA) and normal acid varieties. LA peaches have fruit juice pH above 3.8 – 4.0 (Yoshida, 1970; Dirlewanger *et al.*, 1998; Boudheri *et al.*, 2009) and reduced malate and citrate content probably related to a decreased phosphoenolpyruvate carboxylase activity (*PEPC*) (Moing *et al.*, 1998; Moing *et al.*, 2000). LA peaches group (also defined as "honey peaches") is mainly a Far-East countries commodity while European and American areas consume and appreciate more normal-acid type (Reimer, 1906). Genetic architecture of LA trait or locus *D* (*D* is for "Doux" meaning "sweet" in French) has been partially elucidated, being mapped on the chromosome 5 (Dirlewanger *et al.*, 1998; Micheletti *et al.*, 2005). Locus *D* is the major locus controlling acidity in peach genotypes (Dirlewanger *et al.*, 1998; Micheletti *et al.*, 2005) and plays the primary role acting as dominant and capturing the largest part of phenotypic variation observed in peach fruits. The characterization of Locus *D* has suggested that targeted-selection toward LA or normal acid peach should be promising, making breeding advances relatively rapid and more cost-effective. However, *Locus D* does not fully explain the variegated OAs profiles and TA values found in the large peach collection studied (*Chapter 2* and *Chapter 4*). Based on previous results (*Chapter 2* and *Chapter 4*), TA in peach seemed better described by a mixed genetic architecture with one major locus and few minor loci involved rather than by a polygenic effect. LA trait has been reported in other fruit tree crops as apple and *Citrus*. In apple, acidity has been proposed as a QTL with two major loci on chromosomes 8 and 16 (Liebhard *et al.*, 2003; Zhang *et al.*, 2012; Kumar *et al.*, 2012*a*, Ma *et al.*, 2016). Other minor loci for apple acidity trait were reported on chromosomes 1 and 6 (Xu *et al.*, 2012). In LA *Citrus* accessions, investigations on acidity trait identified a down-regulated expression of *PH1* and *PH5* at vacuoles level (Faraco *et al.*, 2014; Strazzer *et al.*, 2019). Another gene *Noemi* has been detected as the major determinant of *Citrus* fruit acidity and LA accessions shows large deletions or insertions by retrotransposons in this gene (Butelli *et al.*, 2019). In the past decades, several possible genes were proposed to have a role in controlling malate and citrate content. In apple, a dominant *Ma* allele was identified on chromosome 16 and associated with the presence of large malate content. This allele showed an incomplete dominance over the LA allele *ma*

(Maliepaard *et al.*, 1998, King *et al.* 2000; King *et al.*, 2001; Liebhard *et al.* 2003; Xu *et al.* 2012; Khan *et al.* 2013). Another major locus for MAL trait was detected on chromosome 8, flanking the acidity one (Kenis *et al.*, 2008; Zhang *et al.*, 2012; Kumar *et al.*, 2012*a*). In LA *Citrus* accessions, the reduced content of citrate seems dependent on *acitric* gene (Fang *et al.*, 1997). However, both *Ma* and *acitric* seemed to have no homologous genes in peach (Boudheri *et al.*, 2009). Thus, identifying possible genetic determinants of peach acidity-related loci seems attractive, especially under the highlighted necessity of including fruit taste among the objectives in breeding programmes. The main objective of genome-wide association studies (GWAS) is identifying genomic source of variations highly associated to target trait in a collection of accessions, enabling the markers-assisted selection (MAS) deployment in successive seedling generations.

In this chapter, GWA analysis was performed on TA and OA profiles in the peach collection (201 unique accessions recorded in two harvest seasons). The main objectives were to elucidate the genetic architecture of TA and OAs pattern in peach and advance genomics-assisted breeding of peach quality-related attributes. The results obtained could be useful also for refining the genomic prediction models previously adopted (i.e. *Chapter 4*).

5.2 MATERIALS AND METHODS

5.2.1 PEACH PHENOTYPIC DATASET

The peach panel included 201 individuals between accessions and selections, all recorded for titratable acidity (TA) and ten OAs content (i.e. oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) in two years (2017 and 2018). Based on the strong correlations observed across years for the fruit quality attributes (i.e. *Chapter 2*), replicated measurements for each genotype were averaged. Statistical analysis on TA were carried out in *Chapter 2*. Hereby, data of the OAs tested (i.e. oxalate, cis-aconitate, citrate, tartrate, galacturonate, malate, quinate, succinate, shikimate and fumarate) were binned to calculate their frequencies and identify possible associations with variations along the genome. Statistical analysis on OA content was performed using *ggplot2* package in *RStudio*.

5.2.2 GENOTYPING OF PEACH ACCESSIONS

Leaves from 200 peach accession were collected during spring and stored at – 80° C until DNA extraction. DNA isolation followed the method described by Doyle and Doyle (1987). Quality and quantity of DNA extracted was checked by electrophoresis on agarose gel and by dsDNA HS assay kit for Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Genotyping of 200 individuals was performed through the peach 18K SNP chip and using the peach genome version 2.0 as reference (Verde *et al.*, 2017). Genotypic data were filtered removing SNPs with a missing rate higher than 10% and minor allele frequency (MAF) lower than 5%. A total of 11,559 SNPs passed the quality filters and were included in the following analyses.

5.2.3 POPULATION STRUCTURE AND PRINCIPAL COMPONENT ANALYSIS

Population structure was inferred using *ADMIXTURE* software (v 1.3.0), a maximum likelihood- based clustering method (Alexander *et al.*, 2009). The analysis was performed by running the software with *K* values *a priori* established from 2 to 6, which corresponded to the fixed number of source population. Then, the results were compared using cross-validation error rates (*CV-error*) per each *K* value tested in order to avoid a poor resolution of the population structure. The adopted number of cluster *K* was chosen based on the ΔK method, which estimated the logarithmic probability rate of change between successive *K* values (Evanno *et al.*, 2005). Then, the Q-matrix was derived for the *K* value chosen. The kinship matrix was calculated in *TASSEL* as a genetic distance matrix among individuals and visualized through a heat map. Principal Component analysis (PCA) was performed using *factoextra* package in *RStudio.* PCA combined the population stratification with the tested traits.

5.2.4 GENOME WIDE ASSOCIATIONS ANALYSIS

Association analysis was performed on a total of 200 accessions using different GWA algorithms in *GAPIT* (Lipka *et al.*, 2012): mixed linear model (MLM), fixed and random model circulating probability unification (FarmCPU) and linkage disequilibrium iteratively nested keyway (Blink). Population structure was taken into account including the Q-matrix with *K =2* as a covariate. The inclusion of population stratification adjusted the SNP-trait association by reducing false positive call and increasing the statistical power (Price *et al.*, 2010). Compared to FarmCPU and Blink, MLM included also the kinship matrix (K-matrix). To reduce confounding effects, FarmCPU modelled the genetic association without the K-matrix and through maximum likelihood method. Blink method based on the Bayesian Information Content (*BIC*) to further reduce spurious associations between SNPs and target trait. Univariate GWAS for tested traits were visualized in Manhattan plots. In each Manhattan plot, the peach genome was plotted against the *x*-axis and the degree of correlation between each SNP and the tested trait was plotted against the *y*-axis. When strong associations occurred between genomic locus/loci and traits, larger peak heights were observed in the plots. Then, quantile-quantile (Q-Q) plots of *P*-values were investigated in order to assess the tested-algorithm performances. The comparison among algorithms was based on the relationship between the observed (on *y-axis*) and the expected (on *x-axis*) negative logarithms base 10 of the *P*-values, assuming that *P*-values were uniformly distributed. Each algorithm was visualized as dotted lines in the 95% confidence interval for the Q-Q plot under the null hypothesis of no associations between SNPs and target trait. The determination of SNP significance based on a conservative threshold inferred by Bonferroni's correction for a type one error rate of 0.05 (grey horizontal line in Manhattan plots) and related to the effective number of independent tests. SNPs were determined to be significantly associated with the target trait when were fitted above the Bonferroni's adjusted-threshold.

5.3 RESULTS

5.3.1 PHENOTYPIC DATA

Among all the OAs considered in the peach panel , malate and citrate were the most abundant accounting for the 62% and the 22.6% of the total content, respectively, and with concentrations averaging on 3000-6000 ng/μL and 1000 ng/μL (**Fig. 5.1**), respectively. Moreover, both these two OAs were the greatest contributors to the overall peaches TA (i.e. *Chapter 2*). Among all the other OAs, the widest range of content variation was observed for succinate (from 0 to 1400 ng/μL), although often undetectable in many accessions in many. A large content of quinate (between 500 and 1000 ng/μL **Fig. 5.1**) and galacturonate (on average of 50 ng/μL) was also found, followed by fumarate (on average of 5 ng/μL), shikimate (on average of 5 ng/μL), cis-aconitate (between 1 and 1.5 ng/μL), tartrate (between 0 and 10 ng/μL) and oxalate (absent in almost all peaches collected (**Fig. 5.1**).

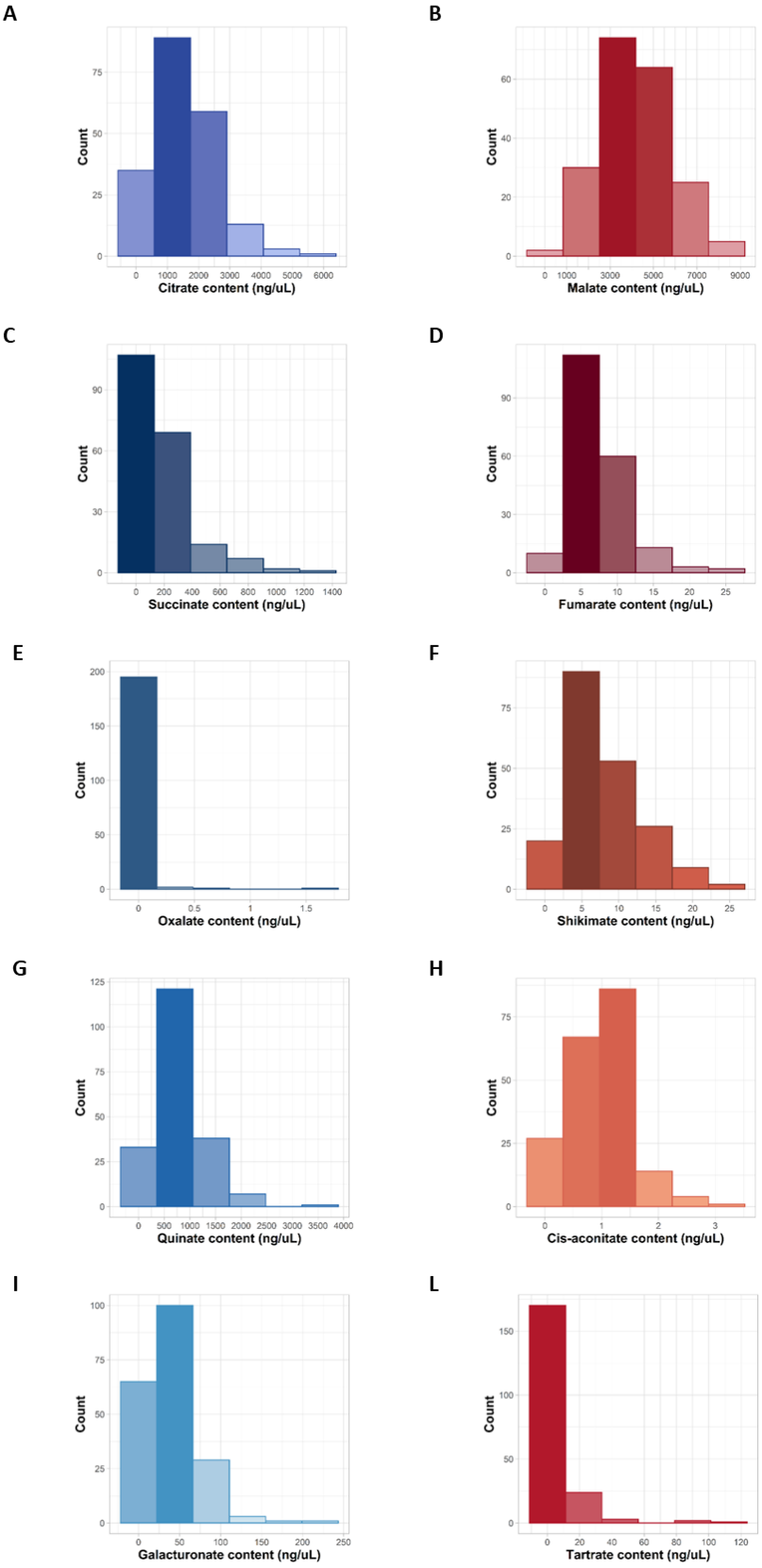

FIGURE 5.1. **OAs profiles (ng/µL) in the peach collection (200 unique accessions).** Replicated measurements for each accession were averaged across years. *A*, citrate; *B*, malate; *C*, succinate; *D*, fumarate; *E*, oxalate; *F*, shikimate; *G*, quinate; *H*, Cis-aconitate; *I*, galacturonate and *L*, tartrate. Carota', 'Michelini' and 'Turquoise' were removed from the histograms of oxalate, cis-aconitate and galacturonate to avoid outliers in the reported range.

5.3.2 POPULATION STRUCTURE AND KINSHIP ANALYSIS

According to the minimization of CV-error for a *K* value equal to 2, structure analysis divided the peach collection panel into two main clusters with the remaining classified as admixed (**Fig. 5.2**). Peach accessions with probability score greater or equal to 0.75 were assigned to one of the two sub-populations while the others attributed to admixture group. The observed population stratification appears related to the history of the analysed peach collection: sub-population 1 (*G1*) mostly consisting of Occidental traditional and not breeding-derived accessions (including some selections deriving from them) and sub-population 2 (*G2*) comprising most of the Occidental breeding-derived accessions and selections (either peach or nectarines). Most of the peach accessions belongs to the admixed group, which also includes some individuals with Oriental ancestry.

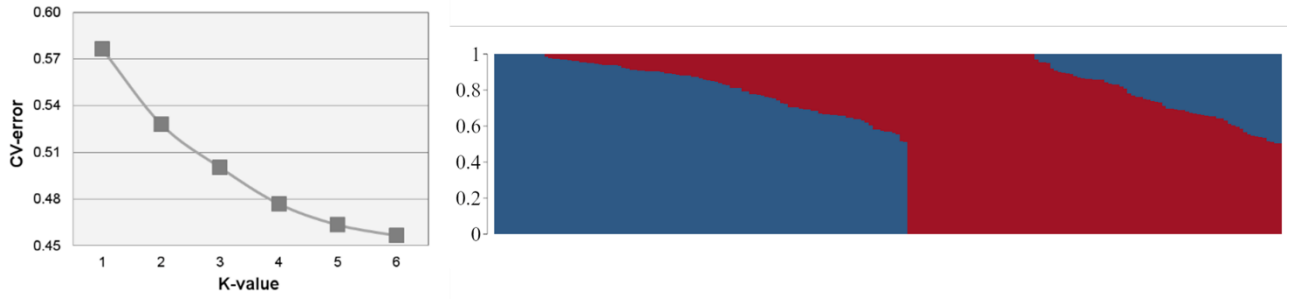

FIGURE 5.2. Structure analysis with K =2 on individual ancestry in the peach collection. The assignment probability of belonging to one of the two groups is reported on *y*-axis. Each individual is represented as a vertical bar coloured in red or blue basing on the membership sub-population. The first sub-population encompassed most of the Occidental genotypes not breeding-derived while the second sub-population included accessions derived from Occidental breeding. Many accessions were assigned as admixed, including many selections and others with Oriental origin.

5.3.2 PRINCIPAL COMPONENT ANALYSIS COMBINED WITH POPULATION STRATIFICATION

PC analysis revealed that the variance explained by each eigenvalue dropped after the third PC (i.e. *Dim3*) (**Fig. 5.3**). The first two PCs (i.e. *Dim1* and *Dim2*) accounted for the 43 % of the total variability for TA and OA content (**Fig. 5.4**). PCA grouped together TA, malate and citate, which had a cumulative positive contribution of 82.8 % to *Dim1*. Shikimate, quinate, tartrate and oxalate accounted for the 84.9 % of total variability in *Dim2*, while specific OAs contributions to *Dim3* was more variable. PCA confirmed the absence of stratification for the target traits, clustering together the different peach sub-populations (i.e. *G1*, *G2* and *admixed*). Low and normal acid accessions were not divided into different sub-populations, as the first group range from 'Bei Jing' (fruit TA of 1.39 g/L of malic acid) to 'Romagna Bright' (11.84 g/L of malic acid) while 'Ornella' (1.18 g/L of malic acid) and 'Early Top' (12.32 g/L of malic acid) belonged to the admixed cluster. Some accessions analysed were positioned outside their membership groups, among them 'Bolinha' characterized by very high fruit quinate content.

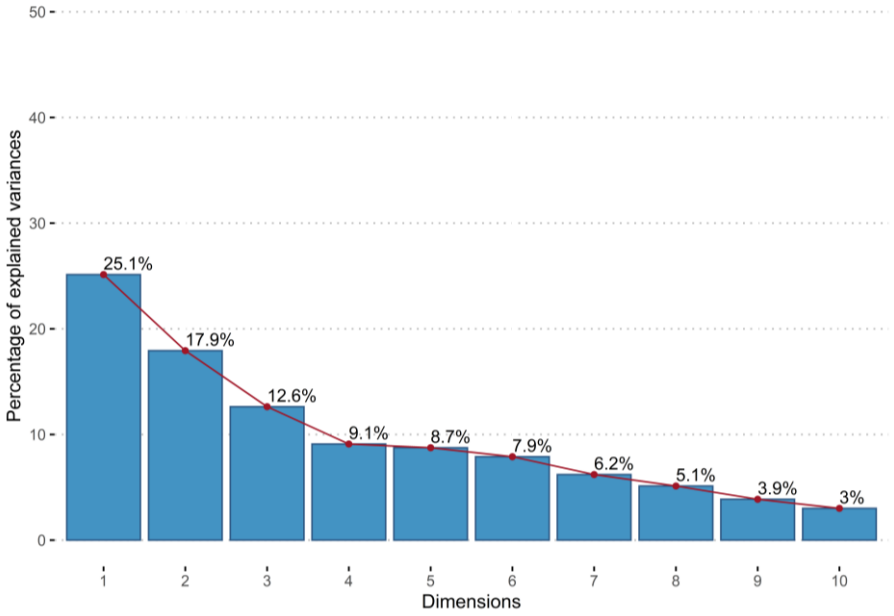

FIGURE 5.3. **Screen plot of the variables' contribution to the first ten PCs.** The first three PCs accounted for the 55.6 % of total variability observed for TA and OA content in the analysed peach panel (200 unique accessions).

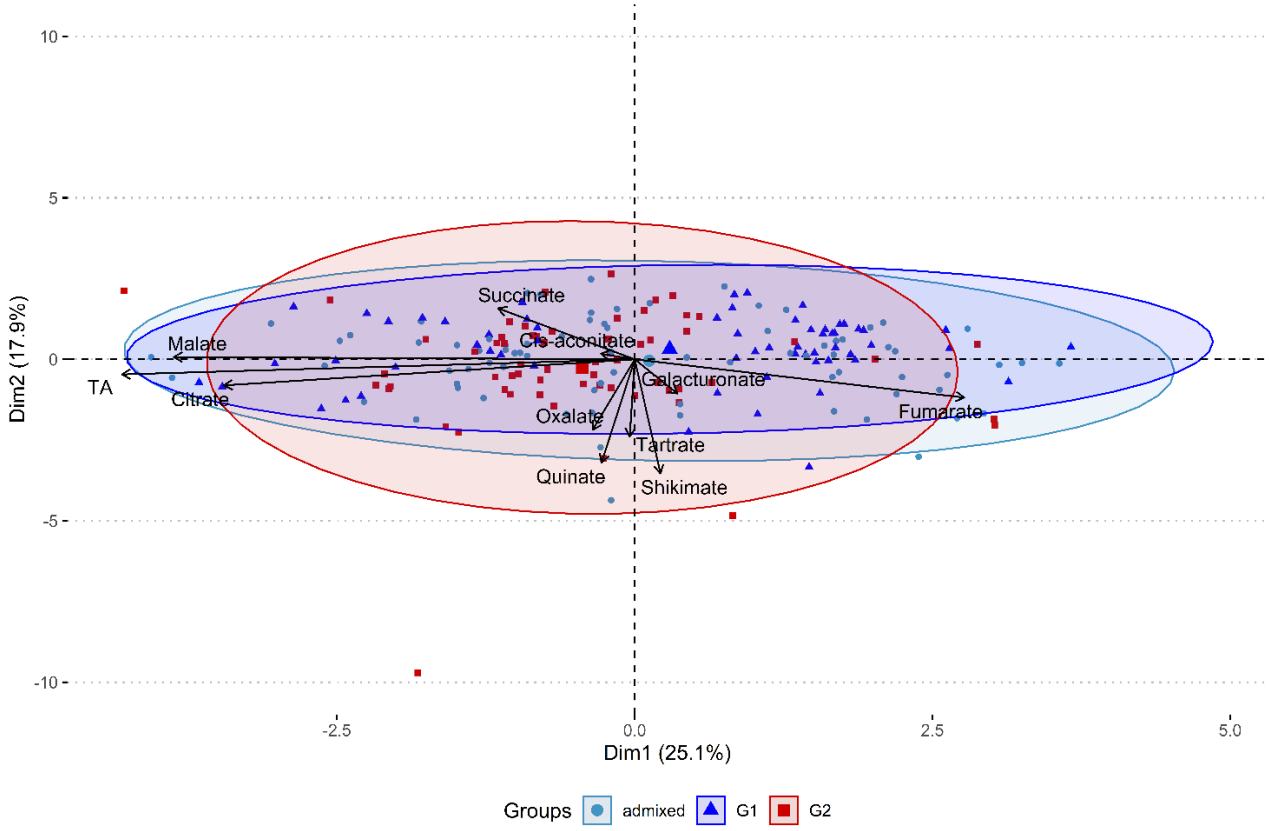

FIGURE 5.4. **Population structure of 200 peach accessions combined with the target traits by PCA.** The first two PCs (i.e. *Dim1* and *Dim2*) accounted for the 43 % of the total variability existing for TA and OA content TA, malate, citrate and fumarate explained the 96.8 % of the totality variables contribution to the first dimension (i.e. Dim1). Shikimate, quinate, tartrate and oxalate accounted for 84.9 % of the total variability in *Dim2*. PCA confirmed the absence of population stratification reporting a not clear separation of the three different membership sub-populations (i.e. *G1*, *G2* and *admixed*).

5.3.3 GWA ANALYSIS FOR FRUIT TITRATABLE ACIDITY IN THE PEACH PANEL

Association analysis for TA trait confirmed the presence of *Locus D* on chromosome 5 as the major QTL controlling peach fruits acidity in all the tested algorithms (**Fig. 5.5; Fig.5.6** and **Table 5.1**). As expected (i.e. *Chapter 2*), fruit TA was low influenced by seasonality having a stronger genotype-dependence and leading to similar associations across different years (**Fig.5.5**). These results proved the feasibility of the peach dataset used for GWAS and the possibility of improving selection accuracy for TA trait in peach. All the models showed a good fitting for *p*-values, as supported by the Q-Q plot distributions of the negative log10 *P*-values observed versus estimated. Among all the SNPs co-localized in *D Locus*, the strongest association (*p-value* of 1.09E-31) for TA was observed for 'Peach_AO_0526196' positioned at 104.5 Kb. Beside of *D Locus*, association analysis found other SNPs on chromosomes 1, 7 and 8 (**Fig.5.6** and **Table 5.1**). In particular, MLM model found that TA trait was strongly associated to a single SNP on chromosome 1 with a *p-value* of 5.38E-10 ('SNP_IGA_37540') and on chromosome 8 with a *p-value* of 2.09E-9 ('Peach_AO_0804157'). Also FarmCPU detected a single strong association on chromosome 8, but this marker ('Peach_AO_0792146') was located at 1.66 Mb from 'Peach_AO_0804157'. On chromosome 7, MLM model identified two associated SNPs at about 12 Mb (*p-value* of 2.34E-07) and at about 17.5 Mb (*p-value* of 1.83E-05), which were also detected by Blink and FarmCPU models, respectively. In addition, Blink and FarmCPU models identified another associated marker ('Peach_AO_0400285 ') at about 25.3 Mb on chromosome 3, but with a poor signal (*p-value* of 5.67E-05) and slightly under the adjusted-threshold. The further examination of Q-Q plot (**Fig.5.6**) indicated that MLM modelled the expected *versus* observed *p-values* along the diagonal line, curving less at the end.

TABLE 5.1. **SNPs associated with fruit TA in the peach panel (200 accessions).** *Chr*, chromosome position; *p-value* reported under the three different GWA models tested; *MAF*, Minor Allele Frequencies; *M*, MLM model; *F*, FarmCPU model; *B*, Blink model.

SNP	Chr	Position (bp)	p-value	MAF	Model
Titratable acidity					
SNP_IGA_37540	1	12,986,177	5.38E-10	0.15	M
SNP_IGA_347845	3	19,790,364	9.27E-06	0.08	F
Peach_AO_0400285	3	25,319,177	5.67E-05	0.30	FB
Peach_AO_0526196	5	104,490	1.09E-31	0.19	FBM
Peach_AO_0746017	7	11,985,355	2.34E-07	0.16	MB
Peach_AO_0765343	7	17,531,971	1.83E-05	0.46	FM
Peach_AO_0792146	8	2,639,406	2.69E-07	0.13	F
Peach_AO_0804157	8	4,293,876	2.09E-09	0.20	M

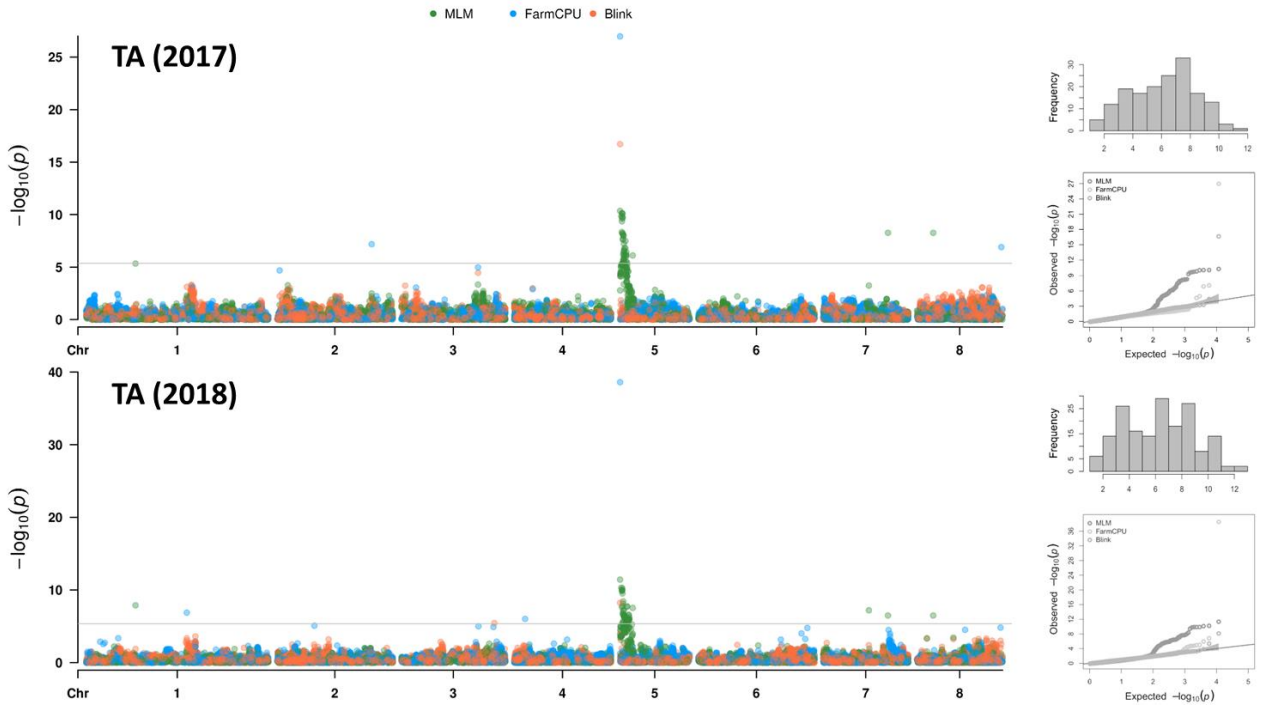

FIGURE 5.5. **Manhattan plot of GWAS for TA carried out in each harvest seasons (2017 and 2018).** Genomic association analysis was performed for peach fruits TA measured in 2017 (116 accessions) and 2018 (178 accessions), separately. Each tested algorithm has a different colour. Grey horizontal lines represent the Bonferroni's adjusted-threshold for a type one error rate of 0.05 based on the effective number of independent test. Chromosomes are reported on *x*-axis. The *y*-axis reports the negative logarithm base 10 (- log10) of the *P*-values estimated. GWAS confirmed the *D Locus* as the major QTL controlling peach fruits TA in both years, proving the validity of the peach dataset used for GWAS

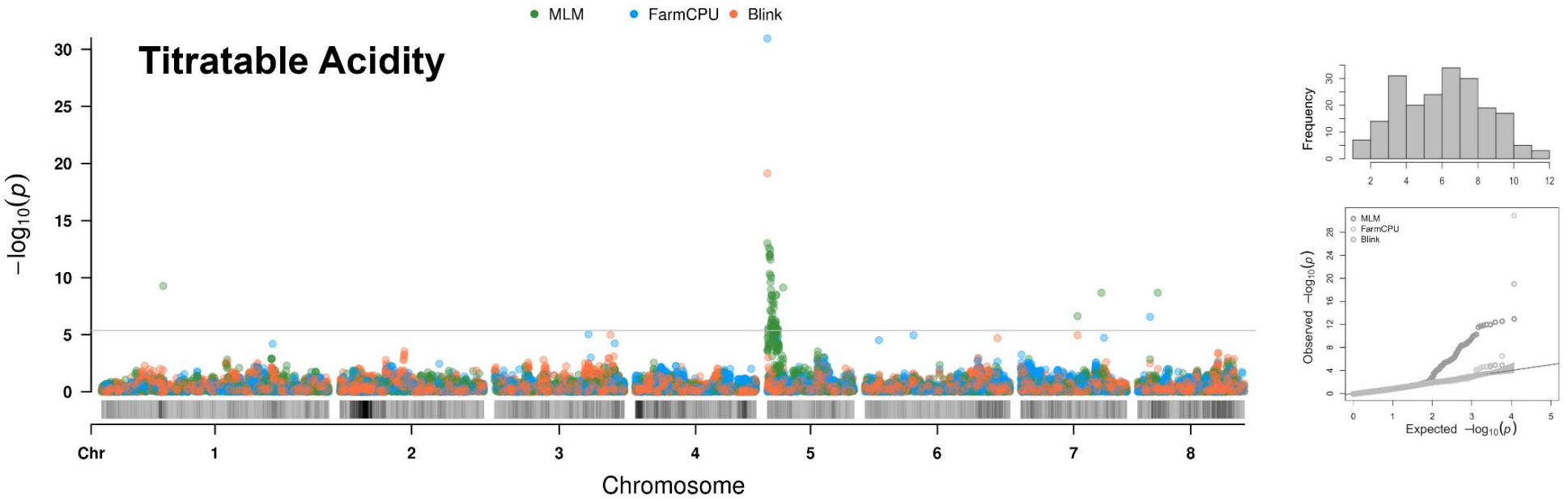

FIGURE 5.6. **Manhattan and Q-Q plots for TA in the peach panel (200 accessions).** Each tested algorithm for SNP-TA associations is coloured differently. The grey horizontal line corresponds to the Bonferroni's adjusted-threshold for a type one error rate of 0.05 based on the effective number of independent test. The *y*-axis reports the negative logarithm base 10 of *P*-values estimated. Among all the three models compared in Q-Q plot, MLM (adjusted for population structure and kinship) seemed to curve less at the end of the diagonal line.

5.3.4. GWA ANALYSIS FOR MALATE AND CITRATE CONTENT IN THE PEACH PANEL

GWA analysis for MAL and CIT in peach fruits revealed SNP-trait associations on chromosomes 1, 3, 5, 6, 7 and 8, but some of them were slighter (**Fig. 5.7** and **Table 5.2**). All models identified 'Peach_AO_0526196' as the strongest SNP associated with MAL (*p-value* of 1.04E-28) at the beginning of chromosome 5, having an effect not only on TA but also on this OA content (**Table 5.1** and **Table 5.2**). This SNP had a MAF of 0.19 for both traits. 'Peach_AO_0527873' showed the strongest association with the citrate content trait at about 585.6 Kb on chromosome 5, with a *p-value* of 3.73E-08 and MAF of 0.21. In addition, FarmCPU and Blink models detected other two SNPs ('Peach_AO_0527873' and 'SNP_IGA_548512') associated with citrate content on chromosome 5, spanning a region of about 917.8 Kb and having a *p-value* of 3.73E-08 and 5.86E-22, respectively. MLM model revealed the same positive association on chromosomes 1 ('SNP_IGA_37540') and 8 ('Peach_AO_0804157') for both MAL and CIT, although the association was poorer for citrate trait. Interestingly, these SNPs were identified also in GWAS for TA trait (**Table 5.1** and **Table 5.2**). Two significant SNPs were detected on chromosome 3 for CIT ('SNP_IGA_358781' with a *p-value* of 1.69E-06) and malate ('Peach_AO_0397899' with a *p-value* of 3.24E-08), respectively. Both molecular markers were comprised between two SNPs ('SNP_IGA_347845' and 'Peach_AO_0400285') associated with TA and spanning a region of about 5.53 Mb on chromosome 3 (**Table 5.1** and **Table 5.2**). Association analysis for MAL and CIT identified the same significant marker ('Peach_AO_0763761') at about 16.98 Mb on chromosome 7, flanking 'Peach_AO_0765343' observed in GWAS for TA. Other markers on chromosome 6 seemed to contribute to CIT and MAL in peaches with 'SNP_IGA_635504' (located at about 10.2 Mb with a *p-value* of 8.73E-11) and 'Peach_AO_0648451' (positioned at about 15.7 Mb with a *p-value* of 1.80E-07), respectively. Based on the Q-Q plots comparison, MLM and Blink seemed to achieve a greater capability of capturing associations between SNPs and these target traits (**Fig. 5.7**). Compared to MLM and Blink models, FarmCPU showed more SNPs close to the Bonferroni's adjusted- threshold in Manhattan plots.

TABLE 5.2. SNPs associated with fruit TA and content of malate and citrate in the peach panel (200 accessions). *Chr*, chromosome position; *p-value* reported under the three different GWA models tested; *MAF*, Minor Allele Frequencies; *M*, MLM model; *F*, FarmCPU model; *B*, Blink model.

SNP	Chr	Position (bp)	p-value	MAF	Model
Malate					
SNP_IGA_37540	1	12,986,177	7.95E-08	0.15	M
Peach_AO_0397899	3	24,509,565	3.24E-08	0.24	F
Peach_AO_0526196	5	104,490	1.04E-28	0.19	FMB
Peach_AO_0648451	6	15,761,835	1.80E-07	0.23	F
Peach_AO_0763761	7	16,981,730	1.91E-08	0.20	M
Peach_AO_0804157	8	4,293,876	1.91E-08	0.20	M
Citrate					
SNP_IGA_37540	1	12,986,177	3.54E-06	0.15	M
SNP_IGA_358781	3	23,028,795	1.69E-06	0.25	B
Peach_AO_0527873	5	585,598	3.73E-08	0.21	M
SNP_IGA_548512	5	1,503,387	5.86E-22	0.21	FB
SNP_IGA_635504	6	10,276,430	8.73E-11	0.48	BF
Peach_AO_0763761	7	16,981,730	1.57E-06	0.20	M
Peach_AO_0804157	8	4,293,876	1.57E-06	0.20	M

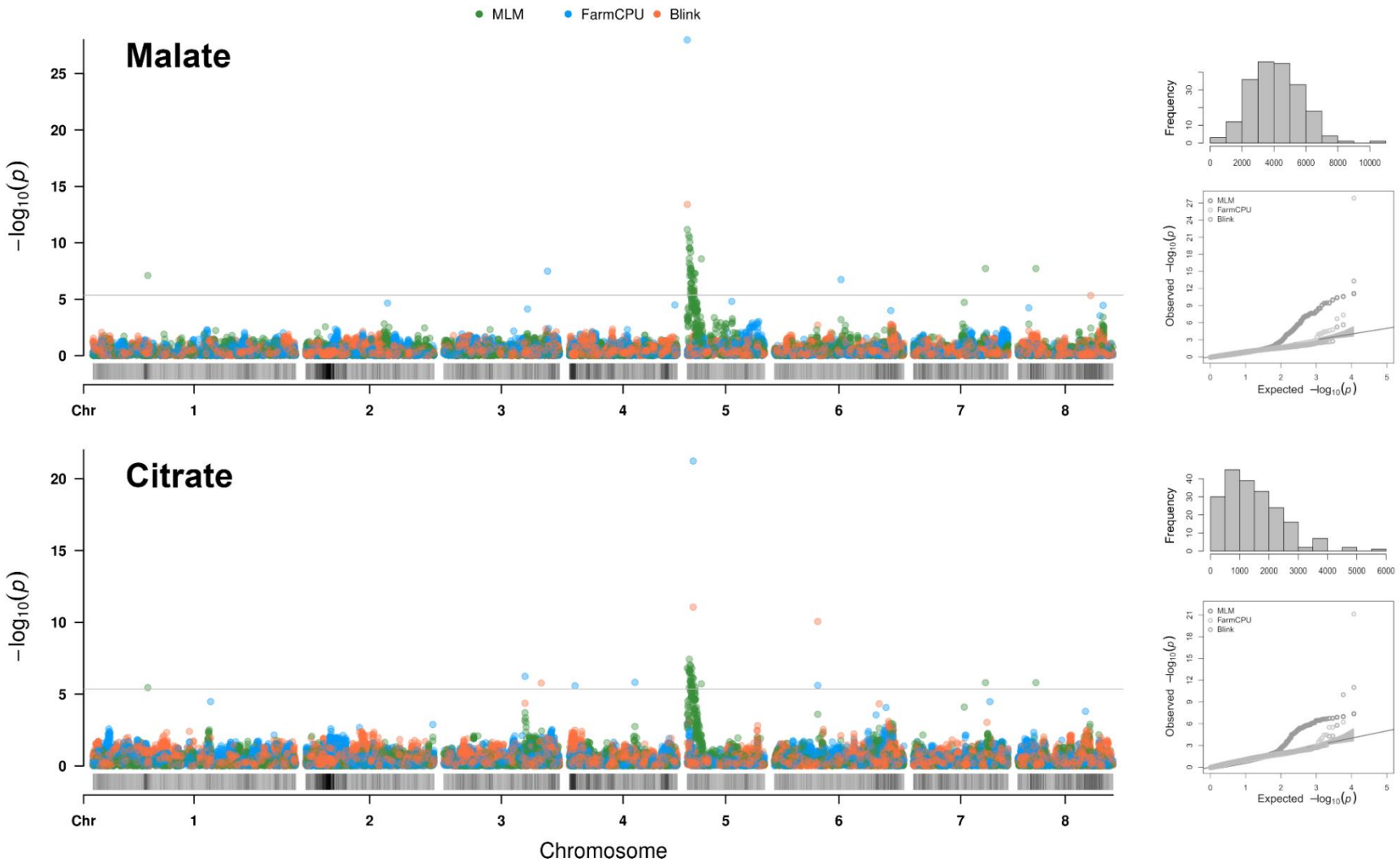

FIGURE 5.7. **Manhattan and Q-Q plots for malate and citrate content in the peach panel (200 accessions).** Each tested algorithm for SNP-traits association is coloured differently. The grey horizontal line corresponds to the Bonferroni's adjusted-threshold for a type one error rate of 0.05 based on the effective number of independent test. The *y*-axis reports the negative logarithm base 10 of *P*-values estimated. Among all the three models compared in Q-Q plot, MLM seemed to curve less at the end of the diagonal line.

5.3.5. GWA ANALYSIS FOR THE OTHER ANALYSED-ORGANIC ACIDS CONTENT IN THE PEACH PANEL

Among the OAs produced through the tricarboxylic acid (TCA) cycle, GWAS for fumarate content reported the strongest signal (**Table 5.3**). This association consisted of a single SNP ('Peach_AO_0526196') on chromosome 5, the same of TA and MAL traits. With the exception of fumarate among all the analysed-OAs of TCA cycle, only oxalate content seemed strongly (*p-value* of 4.47E-08) related to a marker ('SNP_IGA_560930') on chromosome 5 (**Table 5.3**). In addition, GWA for fumarate content revealed other two markers ('Peach_AO_0763761' and 'Peach_AO_0804157') on chromosomes 7 and 8, respectively, with a significant effect also on fruit TA, MAL and CIT traits. Only GWA analysis for cis-aconitate and succinate content showed a minor locus ('Peach_AO_0109161' and Peach_AO_0137377', respectively) positioned at about 36.8 Mb and at about 47.4 Mb on chromosome 1, respectively (**Table 5.3**). Three significant SNPs ('SNP_IGA_305394', 'Peach_AO_0347872' and 'Peach_AO_0381457') were located on chromosome 3 for content of fumarate, oxalate and succinate, respectively. Among them, only the marker associated with succinate content trait flanked 'SNP_IGA_347845' found in GWAS for TA, spanning a region of about 317.8 Kb (**Table 5.1** and **Table 5.3**). A further evidence of the cross-link among TCA cycle OAs metabolic pathways was in the presence of an association signal for succinate content ('SNP_IGA_611891') on chromosomes 6, at about 7.5 Mb from the one of citrate content (**Table 5.2** and **Table 5.3**).

TABLE 5.3. SNPs associated with content of TCA cycle OAs (i.e. cis-aconitate, fumarate, oxalate and succinate) in the peach panel. *Chr*, chromosome position; *p-value* reported under the three different GWA models tested; *MAF*, Minor Allele Frequencies; *M*, MLM model; *F*, FarmCPU model; *B*, Blink model.

SNP	Chr	Position (bp)	*p*-value	MAF	Model
Cis-Aconitate					
Peach_AO_0109161	1	36,875,199	4.52E-04	0.17	B, F, M
Fumarate					
SNP_IGA_305394	3	5,183,078	3.55E-07	0.32	F
Peach_AO_0426069	4	5,073,655	2.28E-06	0.37	F
Peach_AO_0526196	5	104,490	1.37E-11	0.19	B, F, M
Peach_AO_0763761	7	16,981,730	7.98E-06	0.20	M
Peach_AO_0804157	8	4,293,876	7.98E-06	0.20	M
Oxalate					
Peach_AO_0347872	3	11,002,319	4.37E-07	0.05	B, F, M
SNP_IGA_560930	5	4,274,186	4.47E-08	0.12	B, F, M
Succinate					
Peach_AO_0137377	1	47,431,446	7.54E-05	0.12	B
Peach_AO_0381457	3	19,472,512	5.83E-07	0.48	B
SNP_IGA_611891	6	2,803,790	3.71E-05	0.28	B

Among the OAs not produced by TCA cycle (i.e. galacturonate, quinate, shikimate and tartrate), 'SNP_IGA_88772'showed the strongest association with shikimate content trait at 26 Mb on chromosome 1, with a *p-value* of 4.87E-12 and MAF of 0.06 (**Fig. 5.8** and **Table 5.4**). This marker localized between 'SNP_IGA_37540' (seen in GWAS for fruit TA and content of malate and citrate) and 'Peach_AO_0090886' detected in the GWAS for galacturonate content (**Table 5.1** and **Table 5.4**). Association analysis reported a locus ('Peach_AO_0021918') related to quinate content at the beginning of the same chromosome, also suggesting that it could have and additive effect on 'SNP_IGA_37540' associated with TA, MAL and CIT traits in a region of almost 6 Mb (**Fig. 5.8**, **Table 5.1**, **Table 5.2** and **Table 5.4**). Among the non TCA cycle OAs, GWA analysis for tartrate and quinate identified two markers ('SNP_IGA_364517' and 'SNP_IGA_783342') on chromosomes 3 and 7, respectively, which were distant about 121 Kb and 275 Kb from the SNPs associated with TA (**Fig. 5.8**, **Table 5.1** and **Table 5.4**). Similarly, association analysis found a marker ('SNP_IGA_810565') associated with galacturonate content trait on chromosome 8, with *a p-value* of 1.29E-04 and located at about 837 Kb from the one for TA trait (**Table 5.1** and **Table 5.4**). Other significant markers associated with shikimate content were distributed over chromosomes 4 and 6, while a single signal for tartrate content was identified at the end of chromosome 8. Among the non TCA cycle OAs, the only association found on

chromosome 5 was with quinate content ('Peach_AO_0573863'), although it was the slightest for this OA (*p-value* of 1.38E-06) (**Fig. 5.8** and **Table 5.4**). Compared to citrate and oxalate, 'Peach_AO_0573863' showed the greatest distance (of about 10.4 Mb) from the SNP associated with TA, malate and fumarate content and co-localized with *Locus D* on chromosome 5.

TABLE 5.4. SNPs associated with content of non TCA cycle OAs (i.e. galacturonate, quinate, shikimate and tartrate) in the peach panel. *Chr*, chromosome position; *p-value* reported under the three different GWA models tested; *MAF*, Minor Allele Frequencies; *M*, MLM model; *F*, FarmCPU model; *B*, Blink model.

SNP	Chr	Position (bp)	*p*-value	MAF	Model
Galacturonate					
Peach_AO_0090886	1	29,263,033	2.42E-04	0.14	BF
SNP_IGA_810565	8	3,476,627	1.29E-04	0.09	BF
Quinate					
Peach_AO_0021918	1	7,027,644	3.43E-07	0.14	B
Peach_AO_0573863	5	10,516,519	1.38E-06	0.21	B
SNP_IGA_783342	7	17,807,009	3.59E-06	0.18	F
Shikimate					
SNP_IGA_88772	1	27,062,432	4.87E-12	0.06	BFM
Peach_AO_0411086	4	996,927	9.28E-08	0.12	F
SNP_IGA_693707	6	27,789,777	6.73E-06	0.05	F
Tartrate					
SNP_IGA_364517	3	25,440,837	6.14E-07	0.09	BFM
Peach_AO_0858471	8	17,090,260	8.46E-07	0.14	BFM

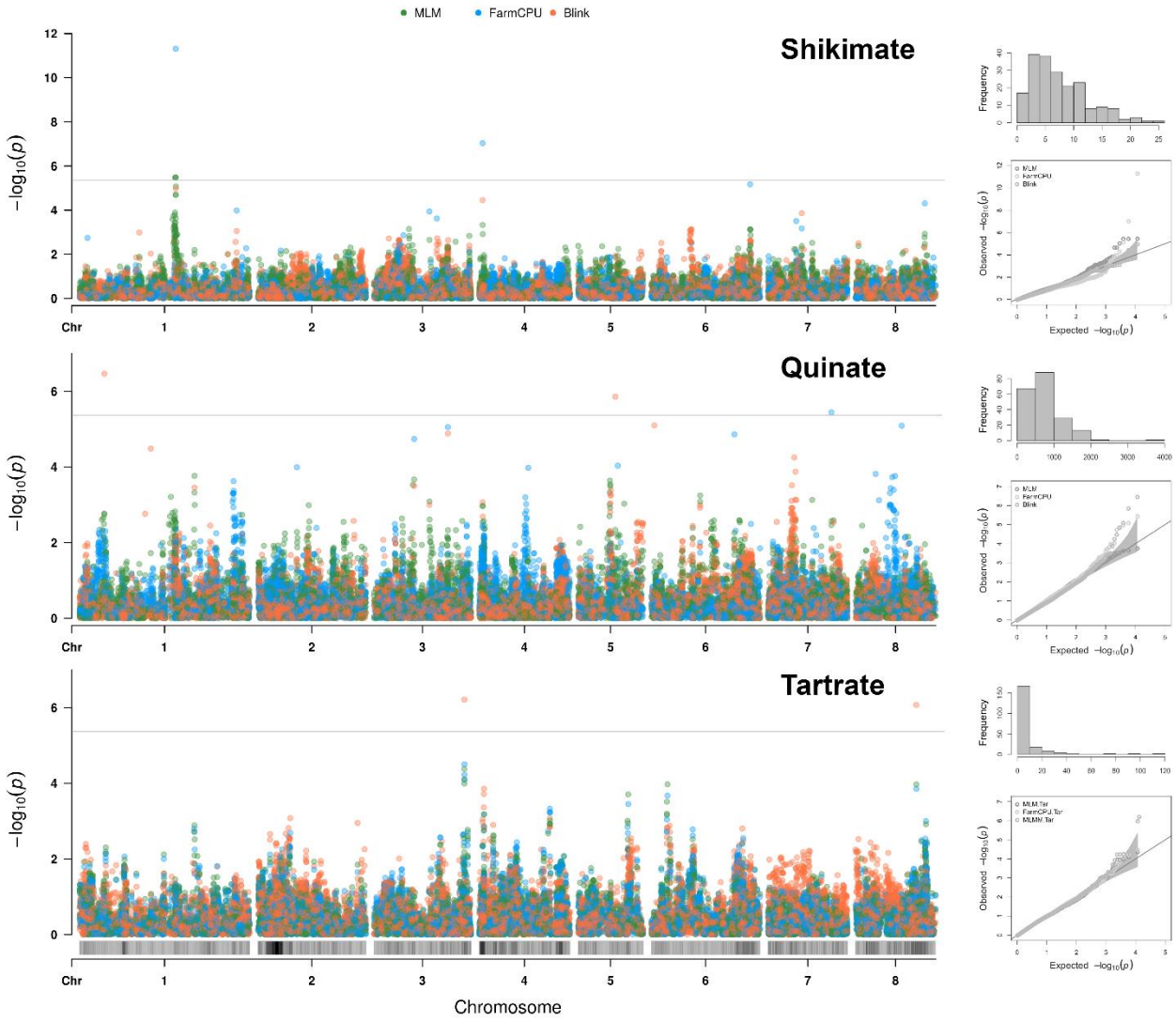

FIGURE 5.8. Manhattan and Q-Q plots for shikimate, quinate and tartrate content in the peach panel (200 accessions). Each algorithm for SNP-traits association is coloured differently. The grey horizontal line corresponds to the Bonferroni's adjusted-threshold for a type one error rate of 0.05 based on the effective number of independent test. The *y*-axis reports the negative logarithm base 10 of *P*-values estimated.

5.4 DISCUSSION

Fruit titratable acidity (TA) and organic acid (OA) profiles are two of the most important attributes contributing to the overall sweetness perception and to consumers' satisfaction (Colaric *et al.*, 2005). Only the major locus controlling peaches acidity has been discovered and mapped so far (Dirlewanger *et al.*, 1998), but without capturing all the variability described among peach varieties. Peach acidity seemed to be a complex trait, with one major locus and probably other minor loci underlying the trait. In this chapter, the study provided a first insight into an in-depth dissection of other genomic loci responsible for acidity variability and of ten OAs contributing to the final peaches acidity. The availability of a complete and large fruits-quality related dataset has enabled to further explore this trait through a GWA analysis-approach to find other possible genomic determinants of peaches acidity. As previously described in *Chapter 2*, malate

and citrate were the two most abundant OAs in the peach panel studied, followed by quinate and succinate. Most of the other OAs were observed at very low concentrations in the analysed-peaches, with a reduced range of variations and present at low frequency (for example, oxalate, cis-aconitate and tartrate). Population stratification was analysed being a factor affecting the accuracy of GWA analysis (Price *et al.*, 2010). The total 200 peaches genotypes used in all the analyses comprise diverse genetic materials, mostly breeding derived Estimation accuracy was achieved by a different model basing on the considered target trait. For example, MLM seemed to better capture SNP-trait association in GWAS for TA while Blink achieved greater accuracy in SNP-tartrate association. Interestingly, GWA analysis for TA confirmed the presence of other two minor loci (especially the ones on chromosomes 1 and 7, respectively) observed in Manhattan plot for the allele weights estimation (i.e *Chapter 4*). However, these SNPs linked to the TA in peach fruits still need to be validated before applying them in marker assisted selection (MAS). Among all the ten OAs considered, malate and fumarate were strictly linked to TA on chromosome 5, identifying the same SNP mapped near *Locus D*. Also, CIT trait had a marker flanking the SNP for TA, malate and fumarate content on chromosome 5, covering a region of about 481 Kb, all positioned in the reported position for the Locus D (Dirlewanger *et al.*, 1998). GWA analysis for oxalate and quinate identified a significant marker positioned more distant on the same chromosome, but with a significant effect on these traits. In all the OAs panel, Locus D showed dominance on the other minor loci detected in association analysis. Results evidenced that OAs had additive effects on the identified TA loci, especially on chromosomes 1, 3, 5, 7 and 8. The presence of OAs in peach flesh depends on the balance between OA synthesis, transport and storage into vacuoles (Etienne *et al.*, 2002), that could reflect the multiplicity of minor QTLs detected. Among all the OAs in the peach panel, malate and citrate were not only the most abundant in content but also they had the largest amount of QTLs – on a total of 6 chromosomes interested- observed in GWAS. Among the other TCA cycle OAs, fumarate was the second largest contributors to the fruit TA in the analysed-peach accessions, as also suggested by PCA. Among the non TCA cycle OAs, quinate and shikimate added a strong effect to fruit TA on chromosome 1, with the other QTLs located along different chromosomes and with a unique distribution pattern. Analysis on galacturonate and cis-aconitate had a very poor resolution, although their content was correlated across the two harvest seasons (i.e. *Chapter 2*). A possible reason could be the largest amount of peach accessions without the presence of these OAs compared to malate, citrate and fumarate, that were always observed in the analysed panel.

5.5 CONCLUSIONS

The analyses performed in this chapter represent the first effort in investigating more in-depth the peach acidity trait. The clarification of peach fruit acidity architecture is desired for breeding purposes in order to improve the selection accuracy. The achieved results confirmed the already proposed oligogenic architecture of acidity trait in peach. Also, the results seemed to be promising for further study on peach genetic diversity

to fully characterize TA and OA qualitative and quantitative profiles. The more precise identification of possible minor loci involved in the overall peaches acidity could be exploit in future breeding programmes, reducing costs and increasing the selection efficiency. However, before applying the identified markers in MAS, a SNP validation has to be carried out in order to confirm the association with the target trait and remove any false positives. In addition, an association studies for fruit acidity and organic acids content could be performed on larger seedling population in order to confirm the proposed loci. These aspects will consider in future studies.

6. GENERAL DISCUSSION AND CONCLUSIONS: NEXT CHALLENGES AND FUTURE PERSPECTIVES

So far, peach and apricot fruit-related breeding strategies mainly improved fruit size, firmness and skin colour with less attention to the organoleptic quality; as a consequence, a "poor taste" is frequently reported by consumers, as widely discussed in the previous chapters and in previous works (Fideghelli *et al.*, 1998; Moreno, 2005; Cantín *et al.*, 2009; Stanley *et al.*, 2014; Cirilli *et al.*, 2016). Under the need of including an improved taste among the objectives of stone fruits breeding programmes, this PhD thesis mainly focused on acidity and organic acids content, both assessed as great contributors of sweetness perception and consumers' satisfaction (Esti *et al.*, 2002; Harker *et al.*, 2002; Byrne *et al.*, 2005; Colaric *et al.*, 2005; Delgado *et al.*, 2013).

Among the main constraints in pursuing markets' demand, the long-time for developing and releasing novel peach and apricot varieties represents a bottleneck for breeding. A deep and high-quality phenotyping is still one of the most critical causes of this delay, although it allows to expand the plant material available for breeders.

The in depth phenotyping performed on peach and apricot germplasm collections dissected a large number of accessions and collected fruit quality related information useful for next breeding programmes. Fruit acidity showed a diverse range of variation in the peach panel (i.e. *Chapter 2*), where the maximum value was 12.31 g/L of malic acid (in 'Early Top'). However, this fruit attribute was more variegated in apricot accessions (i.e. *Chapter 3*), reaching a concentration of 26.33 g/L of malic acid in 'BO05634124' peel. This larger variability could be at least in part related to recent apricot breeding history, with the introduction of PPV-resistant and self-incompatibility apricot materials with high fruit acidity (i.e. *Chapter 1*).

Unlike peach (Yoshida, 1970; Dirlewanger *et al.*, 1998; Boudheri *et al.*, 2009), a scarce knowledge about fruit acidity in apricot exists. Apricot cultivars are not clearly classified into acids and low-acids based on their pH while major loci controlling acidity trait have been not mapped yet. However, some Turkish accessions (such as 'Hacıhaliloğlu' and 'Kabaaşı') or some novel cultivars show a reduced acidity content (Tricon *et al.*, 2010; Karabulut *et al.*, 2018), suggesting a future trend toward apricot breeding for low-acidity traits.

Phenotypic results obtained in both germplasm collections seems to reflect findings of similar works, although they generally relied on a smaller panel of assessed accessions (Yoshida, 1970; Audergon *et al.*, 1991*a*; Audergon *et al.*, 1991*b*; Bassi *et al.*, 1996; Gurrieri *et al.*, 2001; Chen *et al.*, 2006; Ruiz and Egea, 2008; Bureau *et al.*, 2009; Fan *et al.*, 2017). Also, fruit acidity was described as more genotype than year-dependent in both species (i.e. *Chapter 2* and *Chapter 3*), with a very slight variation across years, with a negligible

seasonal effect. In peach, orchard management scarcely influences fruit titratable acidity (Cummings and Reeves, 1971; Souty *et al.*, 1999), furtherly validating the hypotheses of a prevailing genetic control.

Malate and citrate were the most abundant organic acids found in the whole peach and apricot collections, in agreement with several other previous studies (Byrne *et al.*, 1991; Bassi *et al.*, 1996; Moing and Svanella, 1998; Gurrieri *et al.*, 2001; Etienne *et al.*, 2002; Colaric *et al.*, 2005; Schmitzer *et al.*, 2011; Bureau *et al.*, 2013; Fan *et al.*, 2017). Citrate and malate content appeared strongly and positively correlated with titratable acidity in peach while only citrate seemed to play a role in increasing apricot acidity. Both *Prunus* species showed that fumarate was negatively correlated to fruit acidity. Apart from a few, most of peach accessions accumulated larger content of malate than citrate: this can be considered a peculiar trend of this species. Compared to peach, apricot accessions seemed to predominantly accumulate malate or citrate, an evidence supported by other studies albeit, in relatively small panels (Souty *et al.*, 1991; Gurrieri *et al.*, 2001; Karabulut *et al.*, 2018). This finding suggested that organic acids metabolism, accumulation and catabolism seemed more under genetic control rather than being influence by seasonal effects, with a highly stable relative composition of the most abundant organic acids over years. Also, this genotype-dependent accumulation of malate or citrate seemed consistent between pulp and peel, strengthening that malate/citrate ratio is a useful chemical index for classifying apricot cultivars (Souty *et al.*, 1976; Gurrieri *et al.*, 2001). A different contribution of each organic acid presence to the final taste perception is not excluded in the future, especially for identifying a possible relationship with an increased degree of liking and acceptance at market's level.

Harvest day (equal to the full fruit maturity stage) influenced soluble-solids content and dry matter in percentage while no effects were recorded on titratable acidity in the peach panel. On the other hand, in apricot, harvest day seemed negatively related to titratable acidity both in pulp and peel.

All the phenotypic data collected along the PhD thesis will be useful for supporting ongoing genomics analyses, especially in apricot that seems a fertile soil for in-depth investigations. Among the fruit attributes evaluated in this PhD, titratable acidity and organic acids contents might be further explored in GWAS with a high-quality genotypic dataset. Apricot genome was recently released using *PacBio* sub-reads aligned on *Illumina* fragments, with an estimated heterozygosity rate of 0.900-0.902 % (Jiang *et al.*, 2019). The lack of an apricot reference genome has probably hampered the advances in apricot studies in the last decades. A future genotyping of the apricot accessions assessed in this PhD seems highly feasible in next years and could provide additional, and possibly more in depth information on genetic determinisms of acidity-related traits. In turn, this will be crucial to identify candidate gene(s) and/or variant(s) for either marker assisted selection or to improve the knowledge on molecular and biochemical mechanisms underlying organic acids accumulation in fruit tissues.

With the availability of a high-quality peach reference genome, fruit acidity was explored at genomics level, providing preliminary results useful for other applications. To speed up the selection of candidate parents and to estimate the seedlings' phenotypic performance, genomic prediction-based approach was tested on a large peach collection (i.e. *Chapter 4*). The performed genomic selection study focused on peach fruit acidity and content of malate and citrate. Genomic predictions were promising for the considered peach accessions and population, although further studies will be undertaken to improve the prediction accuracy and to guarantee the repeatability of the prediction model developed. The feasibility "on field" of the prediction model can allow to extend this selection approach to larger seedlings populations, leading to a real cost-effectiveness of breeding programmes. To obtain a long-term genomic selection strategy, a further clarification genetic architecture of peach fruit acidity trait became pivotal. Population structure affects both genomic predictions and GWAS (Zhong *et al.*, 2009; Price *et al.*, 2010) and this study widely explored it in the peach collection. This study (i.e. *Chapter 4* and *Chapter 5*) reported a non-absolute population structure, in particular for the 200 peach accessions recorded for fruit acidity and organic acids content. GWA analysis suggested an inter-crossed effect of organic acids to the overall observed peach fruit acidity, paving the way for other studies (i.e. *Chapter 5*). Results have suggested the presence of both dominance and additive effects underlying peach fruit acidity (i.e. *Chapter 5*). Dominance was confirmed for the known Locus *D* associated with low-acidity in peach while each organic acid seemed to add effects on acidity (i.e. *Chapter 5*). Further investigations on all these aspects seem to be attractive and necessary to fully characterize these fruit-related traits. GWAS provided statistical evidences that few genomic regions were associated with the target traits but without testing if the variants were functional and thus responsible for the phenotypic variation. In this PhD thesis, GWA analyses were performed but no real associations were found between genetic loci and phenotypic variants paving the way for deeper studies in the future. To find plausible causal variants and to quantify their effects on the final phenotypic performance, a fine-mapping should be undertaken as well as a DNA re-sequencing with a larger SNP density. Compared to GWAS, fine mapping might allow to discriminate between functional variants and tag-SNPs, which are generally chosen because in LD with flanking SNPs and often only correlated with functional variants. In this study, GWAS provided a list of SNPs useful for identifying the genomic regions and restricting the interval of interest. However, the identification of candidate parents and segregating populations – in addition to the assessed peach germplasm collection- are required for a high-resolution fine-mapping. Unfortunately, compared to annual crops, this aspect is greatly hampered in peach - as in apricot- by long intergeneration time and space required for maintaining large individual numbers in mapping crosses (Aranzana *et al.*, 2019), slowing down the advances. Moreover, GWAS results combined with a fine-mapping could provide information for quantifying by gene-expression the strength of SNPs effect on target traits. Another strategy for discriminating between statistical associations in GWAS and real causal variants might be positional cloning, especially with the presence of few major loci

likely explaining the largest part of the phenotypic effect. In this way, it could be allowed to force the consideration of these QTLs as discrete traits and to perform a genetic linkage analysis (calculation of the logarithm of the odds score) which is based on the genetic recombination frequencies generated by meiotic cross-overs. Also, RT-PCR might be used to validate the candidate genes involved in peach fruit titratable acidity and organic acid content and to quantify their level of expression between low-acid and normal acid accessions.

This PhD thesis has provided first efforts about the presence of a valuable phenotypic variability for future breeding activities aimed at the improvement of overall fruit quality in peach and apricot. Certainly, results allowed to clarify many aspects before proceeding with a more precise characterization of titratable acidity and organic acids contents in peach and apricot, as previously affirmed. A possible sensory evaluation of organic acid composition in both *Prunus* species might be carried out to understand if a different impact exists on taste. Moreover, more research on the metabolic and catabolic mechanisms responsible for accumulation, storage and degradation of organic acids might be considered, although it requires a different experimental design with fruit collection at diverse fruit maturity stages. Among the genomics' topics covered in the thesis, genomic prediction should deserve further attention in next years as well as GWA analysis.

Certainly, the overarching goal of the phenotypic and genetic dissection of this PhD thesis is to support ongoing peach and apricot breeding programmes, bringing benefits both for agriculture and retailers.

REFERENCES

Abbott A. G., Georgi L., Inigo M., Sosinski B., Yvergniaux D., Wang Y., Blenda A. and Reighard G. (2002). Peach: the model genome for *Rosaceae*. *Acta Horticolturae* 575, 145–155.

Adaskaveg J. E., Schnabel G. and Förster H. (2008). Diseases of peach caused by fungi and fungal-like organisms: Biology, epidemiology and management. In *The Peach Botany, Production and Uses* edited by D. Layne and D. Bassi, CAB International, Wallingford (UK), pp. 352-40.

Alexander D. H., Novembre J. and Lange K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Resource* 19, 1655–1664.

Ayanoglu H. and Kaska N. (1995). Apricot selection studies in the Mediterranean region of Turkey. Tenth International Symposium on Apricot Culture. *Acta Horticulturae* 384, 177-182.

Akca Y. and Askin A. (1995). Clonal selection in the apricot cultivar 'Hacihaliloglu'. Tenth International Symposium on Apricot Culture. *Acta Horticulturae* 384, 169-172.

Amin N., Van Duijn C. M. and Aulchenko Y. S. (2007). A genomic background-based method for association analysis in related individuals. *PLoS ONE* 2 (12).

Annicchiarico P., Nazzicari N., Li X., Wei Y., Pecetti L. and Brummer E. C. (2015). Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics* 16 (1020).

Aranzana M. J., Decroocq V., Dirlewanger E., Eduardo I., Gao S. Z., Gasic K., Iezzoni A., Jung S., Peace C., Prieto H., Tao R., Verde I., Abbott A. G. and Arús P. (2019). *Prunus* genetics and applications after de novo genome resequencing: achievements and prospects. *Horticulture Research* 6 (58).

Arvidsson S., Pèrez-Rodriguez P. and Roeber M. B. (2011). A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modelling for robust quantification of genotype effects. *New Phytologist* 191 (3), 895-907.

Audergon J. M., Souty M. and Breuils L. (1990). Amélioration génétique pour l'obtention d'abricots de qualité. *IX Colloque sur les researches fruitiéres* (4-5-6 Décember, Avignon, France)

Audergon J. M., Duffilol J. M., Souty L., Breuils L. and Reich M. (1991*a*). Biochemical and physicochemical characterization of 400 apricot varieties. Consequences I apricot selection and improvement process. *Acta Horticulturae* 293, 111-119.

Audergon J. M., Reich M. and Souty M. (1991*b*). Abricot: variations des critéres de qualité. *Arboriculture fruitiére* 436, 35-45.

Audergon J. M., M. Souty, L. Breuils, M. Reich, and J. M. Duffilol (1993): Recherches sur les determinismes genetiques associes aux parameters de qualite des fruits chez l'abricot. Agriculture. Programme de recherche Agrimed. *Deuxiemes rencontres sur l'abricotier.* Avignon (France), 89-97.

Bae H., Yun S. K., Jun J. H., Yoon I. K., Nam E. Y. and Kwon J. H. (2014). Assessment of organic acid and sugar composition in apricot, plumcot, plum, and peach during fruit development. *Journal of Applied Botany and Food Quality* 87, 24-29.

Bartolozzi F., Bertazza G., Bassi D. and Cristoferi G. (1997). Simultaneous determination of soluble sugars and organic acids as their trimethylsilyl derivatives in apricot fruits by gas-liquid chromatography. *Journal of Chromatography* 758, 99-107.

Bassi D. and Selli R. (1990). Evaluation of fruit quality in peach and apricot. *Advances in Horticultural Science* 2, 107-112.

Bassi D. and Negri P. (1991). Ripening date and fruit traits in apricot progenies. *Acta Horticolturae* 293, 133-140.

Bassi D., Bartolozzi F. and Muzzi E. (1996). Patterns of hereditability of carboxylic acids and soluble sugars in fruit apricot (*Prunus armeniaca* L.). *Plant breeding* 115, 67-70.

Bassi D. and Audergon J. M. (2006). Apricot Breeding: update and perspectives. *Acta Horticulturae* 701, 279-294.

Bassi D. and Monet R. (2008). Botany and Taxonomy. Edited in *The Peach Botany, Production and Uses* by D. Layne and D. Bassi, CAB International, Wallingford (UK).

Bassi D., Rizzo M. and Foschi S. (2010). Breeding apricot in Northern Italy. *Acta Horticulturae* 862, 151-158.

Batista-Silva W., Nascimento V. L., Medeiros D. B., Nunes-Nesi A., Ribeiro D. M., Zsogon A. and Araùjo W. L. (2018). Modification in organic acid profiles during fruit development and ripening: correlation or causation? *Frontiers in Plant Science* 9 (1689).

Beruter J. (1993). Characterization of the permeability of excised apple tissue for sorbitol. *Journal of Experimental Botany* 44, 519-528.

Bianco L., Cestaro A., Linsmith G., Muranty H., Denancé C., Théron A., Poncet C., Micheletti D., Kerschbamer E., Di Pierro E. A., Larger S., Pindo M., Van de Weg E., Davassi A., Laurens F., Velasco R., Durel C. E. and Troggio M. (2016). Development and validation of the Axiom® Apple480K SNP genotyping array. *Plant Journal* 86, 62-74.

Biscarini F., Stevanato P., Broccanello C., Stella A. and Saccomani M. (2014). Genome-enabled predictions for binomial traits in sugar beet populations. *BMC Genetics* (15).

Biscarini F., Nazzicari N., Bink M., Arus P., Aranzana M/ J., Verde I., Micali S., Pascal T., Quilot-Turion B., Lambert P., Da Silva Linge C., PAcheco I., Bassi D., Stella A. and Rossini L. (2017). Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. *BMC Genomics* 18 (432).

Bliss F. A., Arulsekar S., Foolad M. R., Becerra V., Gillen A. M., Warburton M. L., Dandekar A. M., Kocsisne G. M. and Mydin K. K. (2002). An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. *Genome* 45, 520-529.

Borsani J., Budde C. O., Porrini L., Lauxmann M. A., Lombardo V. A., Murray R., Andreo C. S., Drincovich M. F. and Lara M. V. (2009). Carbon metabolism of peach fruit after harvest: changes in enzymes involved in organic acid and sugar level modifications. *Journal of Experimental Botany* 60 (6), 1823-1837.

Boudheri K., Bendahmane A., Cardinet G. G., Troadec C., Moing A. and Dirlewanger E. (2009). Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. *BMC Plant Biology* 9 (1).

Boyle R., Corke F. and Howarth C. (2015). Image-based estimation of oat panicle development using local texture patterns. *Functional Plant Biology* 42, 433–43.

Brichet N., Fournier C., Turc O., Strauss O., Artzet S., Pradal C., Welcker C., Tardieu F. and Cabrera-Bosquet L. (2017). A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform. *Plant Methods* 13 (1).

Brummell D. A., Dal Cin V., Crisosto C. H. and Labavitch J. M. (2004). Cell wall metabolism during maturation, ripening and senescence of peach fruit. *Journal of Experimental Botany* 55, 2029–2039.

Bureau S., Quilot-Turion B., Signoret V., Renaud C., Maucort M., Bancel D. and Renard M. C. G. (2013). Determination of the composition in sugars and organic acids in peach using mid infrared spectroscopy: comparison results according to datasets and different reference methods. *Analytical Chemistry* 85 (23).

Bus V. G. M., Esmenjaud D., Buck E., Laurens F. (2009). Application of genetic markers in rosaceous crops. In: *Genetics and genomics of the Rosaceae* edited by Folta K. M., Gardiner S. E. Springer (New York), 563-600.

Butelli E., Licciardello C., Ramadgu C., Durand-Hulak M., Celant A., Reforgiato Reupero G., Froechlicher Y. and Martin C. (2019). *Noemi* controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern citrus varieties. *Current Biology* 29 (1), 158-164.

Byrne D. H., Nikolic A. N. and Burns E. E. (1991). Variability in sugars, acids, firmness and colour characteristics of 12 peach genotypes. Journal of the American Society for Horticulture Science 116, 1004-1006.

Byrne D. H. (2002). Peach breeding trends: a worldwide perspective. *Acta Horticulturae* 592, 49-59.

Cantín C. M., Gogorcena Y. And Moreno M. A. (2009). Analysis of phenotypic variation of sugar profile in different peach and nectarine [*Prunus persica* (L.) Batsch.] breeding progenies. *Journal of the Science of Food and Agriculture* 89, 1909-1917.

Cantín C. M., Gogorcena Y. and Moreno M. A. (2010). Phenotypic diversity and relationships of fruit quality traits in peach and nectarine [*Prunus persica* (L.) Batsch.] breeding progenies. *Euphytica* 171, 211-226.

Cao K., Zheng Z., Wang L., Liu X., Zhu G., Fang W., Cheng S., Zeng P., Chen C., Wang X., Xie M., Zhong X., Wang X., Zhao P., Bian C., Zhu nY., Zhang J., Ma G., Chen C., Li Y., Hao F., Li Y., Huang G., Li Y., Guo J., Xu X. and Wang J. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biology* 15 (415).

Cervellati C., Paetz C., Dondini L., Tartarini S., Bassi D., Schneider B. and Masia A. (2012). A qNMR approach for bitterness phenotyping and QTL identification in an F1 apricot progeny. *Journal of Biotechnology* 159 (4).

Chagné D., Carlisle C., Blond C., Volz R. K., Whitworth C. J., Oraguzie N. C., Crowhurst R. N., Allan A. C., Espley R. V., Hellens R. P., Gardiner S. E. (2007). Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. *BMC Genomics* 8 (212).

Chagnè D., Crowhurst R. N., Troggio M., Davey M. W., Gilmore B., Lawley C., Vanderzande S., Hellens R. P., Kumar S., Cestaro A., Velasco R., Main D., Rees J. D., Iezzoni A., Mockler T., Wilhelm L., Van de Weg E., Gardiner S. E., Bassil N. and Peace C. (2012). Genome-Wide SNP Detection, validation, and development of an 8K SNP Array for Apple. *PLoS ONE* 7.

Chen J. Y., Zhang H. and Matsunaga R. (2006). Rapid determination of the main organic acid composition of raw Japanese apricot fruit juices using Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry* 54 (26), 9652 – 9657.

Cirilli M., Bassi D. and Ciacciulli A. (2016). Sugars in peach fruit: a breeding perspective. *Horticulture Research* 3 (15067).

Clark R. T., MacCurdy B., Jung J. K., Shaff J. E., McCouch S. R., Aneshansley D. J. and Kochian L. V. (2011). Three-Dimensional root phenotyping with a novel imaging and software platform. *Plant Physiology* 156 (2), 455-465.

Cobb J.N., DeClerck G., Greenberg A., Clark R. and McCouch S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* 126, 867–887.

Colaric M., Veberic R., Stampar F. and Hudina M. (2005). Evaluation of peach and nectarine fruit quality and correlations between sensory and chemical attributes. *Journal of the Science of Food and Agriculture* 85, 2611-2616.

Crivelli G. (1995). Researches on raw material and pre-treatments suitable for high-quality dried apricots. *Atti I. V. T. P. A*., vol XVIII.

Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., de los Campos G., Burgueño J., González-Camacho J. M., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rutkoski J. And Varshney R. K. (2017). Genomic selection in plant breeding: methods, models and perspectives. *Trends in Plant Science* 22(11), 961-975.

Cummings G. A. and Reeves J. (1971). Factors influencing chemical characteristics of peaches. *Journal of the American Society for Horticultural Science* 96, 320–322.

Daetwyler H. D., Villanueva B., Bijma P. and Woolliams J. A. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics* 124 (6), 369–376.

Daetwyler H. D., Pong-Wong R., Villanueva B. and Woolliams J. A. (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031.

Daetwyler H. D., Swan A. A., van der Werf J. H. J. and Hayes B. J. (2012) Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. *Genetics Selection Evolution* 44 (33).

Daetwyler H., Hayden M., Bansal U., Bariana H. and Hayes B. (2013). Genomic selection for disease and morphological traits in diverse wheat landraces. In *Plant and Animal Genomes XXI.* 11–16 January, San Diego (CA).

DeBolt S., Melino V. and Ford C. M. (2007). Ascorbate as a biosynthetic precursor in plants. *Annals of Botany* 99, 3-8.

Decroocq S., Chague A., Lambert P., Roch G., Audergon J. M., Geuna F., Chiozzotto R., Bassi, D., Dondini L., Tartarini S., Salava J., Krška B., Palmisano F., Karayiannis I. and Decroocq V. (2014). Selecting with markers linked to the *PPVres* major QTL is not sufficient to predict resistance to *Plum Pox Virus* (*PPV*) in apricot. *Tree Genetics and Genomes* 10 (5).

Delgado C., Crisosto G. M., Heymann H. and Crisosto C. H. (2013). Determining the primary drivers of liking to predict consumer's acceptance of fresh nectarines and peaches. *Journal of Food Science* 78, 605-614.

De Mori G., Falchi R., Testolin R., Bassi D., Savazzini F., Dondini L., Tartarini S., Palmisano F., Minafra A., Spadotto A., Scalabrin S. and Geuna F. (2019). Resistance to *Sharka* in apricot: comparison of phase-reconstructed resistant and susceptible haplotypes of 'Lito' chromosome 1 and analysis of candidate genes. *Frontiers in Plant Science* 10.

Desta A. Z. and Ortiz R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* 19 (9), 592-601.

Dirlewanger E., Pronier V., Parvery C., Rothan C., Guy A. and Monet R. (1998). Genetic linkage map of peach. *Theoretical and Applied Genetics* 97, 888-895.

Dirlewanger E., Moing A., Rothan C., Svanella L., Pronier V., Guye A., Plomion C. and Monet R. (1999). Mapping QTLs controlling fruit quality in peach [*Prunus Persica* (L.) Batsch]. *Theoretical and Applied Genetics* 98, 18-31.

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

Dondini L., Lain O., Vendramin V., Rizzo M., Vivoli D., Adami M., Guidarelli M., Gaiotti F., Palmisano F., Bazzoni A., Boscia D., Geuna F., Tartarini S., Negri P., Castellano M., Savino V, Bassi D. and Testolin R (2011) Identification of QTL for resistance to *Plum Pox virus* strains *M* and *D* in 'Lito' and 'Harcot' apricot cultivars. *Molecular Breeding* 27, 289–299.

Echeverría G., Cantín C. M., Ortiz A., López M. L., Lara I. and Graell J. (2015). The impact of maturity, storage temperature and storage duration on sensory quality and consumer satisfaction of 'Big Top®' nectarines. *Scientia Horticolturae* 190, 179-186.

Esitken A. and Guleriuz M. (1995). Investigation of harvest criteria and physical and chemical changes during fruit development of some apricot cultivars in Erzincan. Tenth International Symposium on Apricot Culture. *Acta Horticolturae* 384, 613-618.

Esti M., Cinquanta L., Sinesio F., Moneta E. and Di Matteo M. (2002). Physicochemical and sensory fruit characteristics of two sweet cherry cultivars after cool storage. *Food Chemistry* 76, 339-405.

Etienne C., Rothan C., Moing A., Plomion C., Bodénès C., Svanella-Dumas L., Cosson P., Pronier V., Monet R. and Dirlewanger E. (2002). Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch.]. *Theoretical and Applied Genetics* 105, 145-159.

Etienne C., Génard M., Lobit P., Mbeguié-A-Mbéguié D. and Bugaud C. (2013). What controls fruit acidity? A review of malate and citrate accumulation in fruit cells. *Journal of Experimental Botany* 64 (6), 1451-1469.

Evanno G., Regnaut S. and Goudet J. (2005). Detecting the number of clusters of individuals using the software *STRUCTURE*: A simulation study. *Molecular Ecology* 14(8):2611-20

Falchi R., Vendramin E., Zanon L., Scalabrin S., Cipriani G., Verde I., Vizzotto G. and Morgante M. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *The Plant journal: for cell and molecular biology*, 76(2), 175-187.

Falconer D. S. and Mackay T. F. C. (1996) 'Introduction to quantitative genetics. Fouth edition (Longmans, Harlow, UK)

Famiani F., Cultrera N., Battistelli A., Casulli V., Proietti P., Standardi A., Chen Z. H., Leegood R. C. and Walker R. P. (2005). Phoshoenolpyruvate carboxykinase and its potential role in the catabolism of organic acids in the flesh of soft fruit during ripening. *Journal of Experimental Botany* 56, 2959-2969.

Famiani F., Farinelli D., Moscatello S., Battistelli A., Leegood R. C. and Walker R. P. (2016). The contribution of stored malate and citrate to the substrate requirements of metabolism of ripening peach (*Prunus persica* L. Batsch) flesh is negligible. Implications for the occurrence of phosphoenolpyruvate carboxynase and gluconeogenesis. *Plant physiology and biochemistry* 101, 33-42.

Fan X., Zhao H., Wang X., Cao J. and Jiang W. (2017). Sugar and organic acid composition of apricot and their contribution to sensory quality and consumer satisfaction. *Scientia Horticulturae* 225, 553-560.

Fang D. Q., Federici C. T. and Roose M. L. (1997) Development of molecular markers linked to a gene controlling fruit acidity in citrus. *Genome* 40, 841–849.

Faraco M., Spelt C., Bliek M., Verweij W., Hoshino A., Espen L., Prinsi B., Jaarsma R., Tarhan E. and de Boer A. H. (2014) Hyper acidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower colour. *Cell Reports* 6, 32– 43.

Faust M. and Timon B. (1995). Origin and dissemination of peach. *Horticultural Reviews* 17, 331–379.

Faust M., Suranyi D. and Nyujto F. (1998). Origin and dissemination of apricot. *Horticulture Research* 22, 225–260.

Femenia F., Sánchez E. S., Simal S. and Rosselló C. (1998). Developmental and ripening-related effects on the cell wall of apricot (*Prunus armeniaca* L.) fruit. *Journal of the Science of Food and Agriculture* 77, 487–493.

Fernandes S. B., Dias K. O. G., Ferreira D. F. and Brown P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics* 131 (3), 747-755.

Fideghelli C., Della Strada G., Grassi F. and Morico G. (1998). The peach industry in the world: present situation and trends. *Acta Horticulturae* 465, 29-40.

Fodor A., Segura V., Denis M., Neuenschwander S., Fournier-Level A., Chatelet P., Homa F. A. A., Lacombe T., This P. and Cunff L. L. (2014). Genome-wide prediction methods in highly diverse and heterozygous species: Proof-of-concept through simulation in grapevine. *PLoS ONE* 9.

Foulongne M., Pascal T., Pfeiffer F. and Kervella J. (2003). QTL for powdery mildew resistance in peach × *Prunus davidiana* crosses: consistency across generations and environments. *Molecular Breeding* 12, 33-50.

Fraser L. G., Tsang G. K., Datson P. M., De Silva H. N., Harvey C. F., Gill G. P., Crowhurst R. N. and McNeilage M. A. (2009). A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. *BMC Genomics* 10, 10-102.

Garrick D. J. 2007. Equivalent mixed model equations for genomic selection. *Journal of Dairy Science* 90 (Suppl. 1):376. (Abstr.)

Génard M., Souty M., Holmes S., Reich M. and Breuils L. (1994) Correlation among quality parameters of peach fruit. *Journal Science of Food and Agriculture* 66, 241-245.

Gezan S. A., Osorio L. F., Verma S. and Whitaker V. M (2017). An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research* 4 (16070).

Gilmour A. R., Gogel B., Cullis B. and Thompson R. (2009). *ASREML* User Guide Release 3.0. Hemel Hempstead, VSN International Ltd, Hemel Hempstead (UK).

Giovannoni J. J. (2004). Genetic regulation of fruit development and ripening. *The Plant Cell* 16, 170-180.

Goddard M. E. and Hayes B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* 10, 381–391.

Goulao L. F. and Oliveira C. M. (2008). Cell wall modifications during fruit ripening: When a fruit is not the fruit. *Trends in Food Science and Technology* 19, 4–25.

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M , Dauzat M., Hamard P., Thioux J. J., Rolland G., Bouchier-Combaud S., Lebaudy A., Muller B., Simonneau T. and Tardieu F. (2006) *PHENOPSIS*, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* 169, 623–635.

Granier C. and Vile D. (2014). Phenotyping and beyond: modelling the relationships between traits. *Current Opinion in Plant Biology* 18, 96-102.

Grift T. E., Novais J. and Bohn M. (2011). High-throughput phenotyping technology for maize roots. *Biosystems Engineering* 1 (110), 40-48.

Gross J. (1987). Pigments of fruit. Series: food science and technology. Oxford, UK: Academic.

Guichard E. and Souty M. (1988). Comparison of the relative quantities of aroma compounds found in fresh apricot (*Prunus armeniaca* L.) from six different varieties. *Z Lebensem Unters Forsch* 186, 301-307.

Guo, J., Jang, Q., Zhang, K., Zhao, J. and Yang, Y. (2002) Screening for molecular marker linked to saucer gene of peach fruit shape. *Acta Horticulturae* 592, 267–271.

Gurrieri F., Audergon J. M., Albagnac G. and Reich M. (2001). Soluble sugars and carboxylic acids in ripe apricot fruit as parameters for distinguishing different cultivars. *Euphytica* 117, 183-189.

Habier D., Fernando R. L. and Dekkers J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397.

Habier D., Fernando R. L., Kizilkaya K. and Garrick D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12 (186).

Hardner C. M., Healey A. L., Downes G., Herberling M. and Gore P. L. (2016*a*). Improving prediction accuracy and selection of open-pollinated seed-lots in *Eucalyptus dunnii* (Maiden) using a multivariate mixed model approach. *Annals of Forest Science* 73, 1035–1046.

Hardner C. M., Hayes B. J., Kumar S., Vanderzande S., Lichun C., Piaskowski J., Quero-Garcia J., Campoy J. A., Barreneche T., Giovannini D., Liverani A., Charlot G., Villamil-Castro M., Oraguzie N. and Peace C.

(2016*b*). Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Horticulture Research* 6.

Hardner C., Evans K., Brien C., Bliss F. and Peace C. (2016c). Genetic architecture of fruit quality traits in apple following short-term storage and implications for genetic breeding and selection. *Tree Genetics and Genomes* 12 (20).

Harker F., Marsh K., Young H., Murray S., Gunson F. and Walker S. (2002). Sensory interpretation of instrumental measurements: sweet and acid taste of apple fruit. *Postharvest Biology and Technology* 24, 241-250.

Hartigan J. A. and Wong M. A. (1979). A K-means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1), 100 -108.

Hasib A., Jaouad A., Mahrouz M. and Khouili M. (2009). HPLC determination of organic acids in moroccan apricot. *Ciência e Tecnologia de Alimentos* 3 (4), 207-211.

Hayes B. J., Bowman P. J., Chamberlain A. J. and Goddard M. E. (2009). Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92 (2), 433-443.

Heffner E. L., Sorrells M. E. and Jannink J. L. (2009). Genomic selection for crop improvement. *Crop Science* 49, 1-12.

Heffner E. L., Lorenz A. J., Jannink J. L. and Sorrells M. E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science* 50, 1681–1690.

Heffner E. L., Jannink J. L. and Sorrells M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4(1).

Henderson C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics* 21, 309-310.

Henderson C. R., Kempthorne O., Searle S. R. and von Krosigk C. N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics* 13, 192-218.

Henderson C. R. (1963). Selection index and expected genetic advance. In: *Statistical Genetics and Plant Breeding* edited by Hanson, W. D. and Robinson H. F., pp. 141-163. National Academy of Sciences (National Research Council), Washington DC.

Henderson C. R. (1975). Best linear unbiased estimation and predictor under a selection model. *Biometrics* 31, 423-447.

Henderson C. R. (1984). Applications of Linear Models in Animal Breeding. University of Guelph, Ontario, Canada.

Herbers K. and Sonnewald U. (1998). Molecular determinants of sink strength. *Current Opinion of Plant Biology* 1, 207-216.

Heslot N., Yang H. P., Sorrells M. E. and Jannink J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Science* 52, 146–160.

Heslot N., Jannink J. L. and Sorrells M. E. (2013). Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science* 53, 921–933.

Heslot N., Jannink J. C. and Sorrells M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science* 55.

Hickey J. M., Chiurugwi T., Mackay I. and Powell W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics* 49 (9), 1297 1303.

Hill W. G. and Robertson A. (1968). Linkage Disequilibrium in Finite Populations. *Theoretical and Applied Genetics*, 38, 226-231.

Hill WG, Weir B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research* 93, 47–64.

Holman F. H., Riche A. B., Michalski A., Castle M., Wooster M. J. and Hawkesford M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing* 8 (12).

Hunt C. H., van Eeuwijk F. A., Mace E. S., Hayes B. J., Jordan D. R. (2018). Development of genomic prediction in Sorghum. *Crop Science* 58 (2).

Iwata H., Hayashi T., Terakami S., Takada N., Sawamura Y. and Yamamoto T. (2013). Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breeding Science* 63, 125–140.

Jannink J. L., Lorenz A. J. and Iwata H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics* 9 (2), 166-177.

Jansen M., Gilmer F., Biskup B., Nagel K. A., Rascher U., Fischbach A., Briem S., Dreissen G., Tittmann S., Braun S., De Jaeger I., Metzlaff M., Schurr U., Scharr H. and Walter A. (2009). Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via *GROWSCREEN FLUORO* allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants. *Functional Plant Biology* 36 (11) 902-914.

Jiang F., Zhang J., Wang S., Yang L., Luo Y., Gao S. Zhang M., Wu S., Hu S., Sun H. and Wang Y. (2019). The apricot (*Prunus armeniaca* L.) genome elucidates *Rosaceae* evolution and β-carotenoid synthesis. *Horticulture Research* 6 (1).

Kader A. A. (2008). Flavour quality of fruits and vegetables. *Journal of the Science of Food and Agriculture* 88(11), 1863-1868.

Kaneko S., Kumazawa K., Masuda H., Henze A. and Hofmann T. (2006). Molecular and sensory studies on the umami taste of Japanese green tea. *Journal of Agricultural and Food Chemistry* 54(7), 2688-2694.

Karabulut I., Bilenler T., Sislioglu K., Gokbulut I., Seyhan F., Ozdemir I. S. and Ozturk B. (2018). Effect of fruit canopy positions on the properties of apricot (*Prunus armeniaca* L.) varieties. *Journal of Food Biochemistry* 42.

Kenis K., Keulemans J.and Davey M (2008) Identification and stability of QTLs for fruit quality traits in apple. *Tree Genetics and Genomes* 4, 647–661.

Khan S. A., Beekwilder J., Schaart J. G., Mumm R., Soriano J. M., Jacobsen E. and Schouten H. J. (2013). Differences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. *Tree Genetics and Genomes* 9, 475–487.

King G. J., Maliepaard C., Lynn J. R., Alston F. H., Durel C. E., Evans K. M., Griffon B., Laurens F., Manganaris A. G., Schrevens E., Tartarini S. and Verhaegh J. (2000). Quantitative genetic analysis and

comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* 100, 1074–1084.

King G. J., Lynn J. R., Dover C. J., Evans K. M. and Seymour G. B. (2001). Resolution of quantitative trait loci for mechanical measures accounting for genetic variation in fruit texture of apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* 102, 1227–1235.

Kovács E., Merész P., Kristóf Z. and Németh-Szerdahelyi E. (2008). Ripening and microstructure of apricot (*Prunus armeniaca* L.). *Acta Alimentaria* 37(1), 23–39.

Kumar S., Chagné D., Bink M. C. A. M., Volz R. K., Whitworth C., Carlisle C. (2012a). Genomic Selection for fruit quality traits in apple (*Malus* x *domestica* Borkh.). *PLoS ONE* 7 (5).

Kumar S., Chagné D., Bink M. C. A. M., Volz R. K., Bus V. G.M. Chagné D. (2012b). Towards genomic selection in apple (Malus × domestica Borkh.) breeding programmes: Prospects, challenges and strategies *Tree Genetics and Genomes* 8 (1), 1-14.

Kumar S., Garrick D. J., Bink M. C. A. M., Whitworth C., Changé D. and Volz R. K. (2013). Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* 14 (1).

Kumar S., Molloy C., Muñoz P., Daetwyler H., Chagné D. and Volz R. K. (2015). Genome-enabled estimates of additive and non-additive genetic variances and prediction of apple phenotypes across environments. *G3: Genes, Genomes, Genetics* 5 (12), 2711–2718.

Lange C., Issanchou S. and Combris P. (2000). Expected versus experienced quality: trade-off with price. *Food Quality and Preference* 11, 289-297.

Laurens F., Audergon J., Claverie J., Duval H., Germain E., Kervella J., Le Lezec M., Lauri P. E. and Lespinasse J. M. (2000). Integration of architectural types in French programmes of ligneous fruit species genetic improvement. *Fruits* 54, 441-449.

Ledbetter C., Peterson S. and Jenner J. (2006). Modification of sugar profiles in California adapted apricots (*Prunus armeniaca* L.) through breeding with Central Asian germplasm. *Euphytica* 148, 251-259

Leutenegger A. L., Prum B., Gènin E., Verny C., Lemainque A., Clerget-Darpoux F. and Thompson E. A. (2008). Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics* 73, 516-523.

Li X-w, Meng X-q, Jia H-j, Yu M-l, Ma R-j, Wang L-r, Cao K., Shen Z-j., Niu L., Tian J-b., Chen M-j., Xie M., Arus P., Gao Z-s. and Aranzana M. J. (2013). Peach genetic resources: diversity, population structure and linkage disequilibrium. *BMC Genetics* 14 (84).

Liebhard R., Kellerhals M., Pfammatter W., Jertmini M. and Gessler C. (2003) Mapping quantitative physiological traits in apple (*Malus* × *domestica* Borkh.). *Plant Molecular Biology* 52, 511– 526.

Lin Z., Hayes B. J., Daetwyler H. D. (2014). Genomic selection in crops, trees and forages. A review. *Crop Pasture Science* 65(11).

Lipka A. E., Tian F., Wang Q., Peiffer J., Li M., Bradbury P. J., Gore M. A., Buckler E. S. and Zhang Z. (2012). *GAPIT*: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399.

Lorenz A. J., Chao S., Asoro F. G., Heffner E. L., Hayashi T., Iwata H., Smith K. P., Sorrells M. E. and Jannink J. L. (2011). Genomic selection in plant breeding. Knowledge and prospects. *Agronomy and Plant Genetics* 110, 77–123.

Lorenz A. J., Smith K. P. and Jannink J. L. (2012) Potential and optimization of genomic selection for *Fusarium* head blight resistance in six- row barley. *Crop Science* 52(4).

Lorenzana R. E. and Bernardo R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* 120, 151–161.

Lynch M. and Walsch B. (2018). *Evolution and Selection of Quantitative Traits*. Edited by Oxford University Press (UK).

Ma B. Q., Zhao S., Wu B. Wang D., Peng Q., Owiti A., Fang T., Liao L., Ogutu C., Korban S. S., Li S. and Han Y. (2016). Construction of a high-density linkage map and its application in the identification of QTLs for soluble sugar and organic acid components in apple. *Tree Genetics and Genomes*, 12 (1).

Ma Q. J., Sun M. H., Lu J., Liu Y. J., Hu D. G. and Hao Y. J. (2017). Transcription factor AREB2 is involved in soluble sugar accumulation by activating sugar transporter and amylase genes. *Plant Physiology* 174, 2348–2362.

MacLeod I. M., Bowman P. J., Vander Jagt C. J., Haile-Mariam M., Kemper K. E., Chamberlain A. J., Schrooten C., Hayes B. J. and Goddard M. E. (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* 17 (144).

Maeda H. and Dudareva N. (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants, Annual Review of Plant Biology 63.

Maliepaard C., Alston F. H., van Arkel G., Brown L. M., Chevreau E., Dunemann F., Evans K. M., Gardiner S., Guilford P., van Heusden A. W., Janse J., Laurens F., Lynn J. R., Manganaris A. G., den Nijs A. P. M., Periam N., Rikkerink E., Roche P., Ryder C., Sansavini S., Schmidt H., Tartarini S., Verhaegh J. J., Vrielink-van Ginkel M. and King G. J. (1998). Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* 97, 60–73.

Martinoia E., Maeshima M. and Neuhaus N. E. (2007). Vacuolar transporters and their essential role in plant metabolism. *Journal of Experimental Botany* 58, 83-102.

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

Micheletti D., Dettori M. T., Micali S., Aramini V., Pacheco I., Da Silva Linge C., Foschi S., Banchi E., Barreneche T., Quilot-Turion B., Lambert P., Pascal T., Iglesias I., Carbò J., Wang L., Ma R., Li X., Gao Z., Nazzicari N., Troggio M., Bassi D., Rossini L., Verde I., Laurens F., Arùs P. and Aranzana J. (2005). Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLOS ONE* 10 (9).

Miller S. and Scorza R. (2010). Response of two novel peach tree growth habits to in-row tree spacing, training system and pruning: effect on growth and pruning. *Journal of The American Pomological Society* 64, 199-217.

Moing A. and Svanella L. (1998). Compositional changes during the fruit development of two peach cultivars differing in juice acidity. *Journal of the American Society for Horticultural Science* 123, 770-775.

Moing A., Rothan C., Svanella L., Just D., Diakou P., Raymond P., Gaudillère J.P. and Monet R. (2000). Role of phosphoenolpyruvate carboxylase in organic acid accumulation during peach fruit development. *Physiologia Plantarum* 108, 1-10.

Monet R. (1979). Transmission génétique du caractère 'fruit doux' chez le pêcher. Incidence sur la sélection pour la qualité. In*: Proceeedings of Eucarpia Fruit Section Symposium*. Tree Fruit Breeding. INRA, Angers, France, 273–276.

Montes J. M., Technow F., Dhillon B. S., Mauch F. and Melchinger A. E. (2011). High-throughput non-destructive biomass determination during early plant development in maize under field conditions. *Field Crops Research* 121 (2), 268-273.

Moreno M. A. (2005). Seleccion de patrones y variedades demelocotonero. *Vida Rural* 206, 28–31.

Moser G., Tier B., Crump R. E., Khatkar M. S. and. Raadsma H. W. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genetics Selection Evolution 41(56).

Muranty H., Troggio M, Sadok I. B., Al Rifaı M., Auwerkerken A., Banchi E., Velasco R., Stevanato P., van de Weg E., Di Guardo M., Kumar S., Laurens F. and Bink M. C. A. M. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research* 2 (15060).

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

Oget C., Teissier M., Astruc J. M., Tosser-Klopp G. and Rupp R. (2019). Alternative methods improve the accuracy of genomic prediction using information from a causal point mutation in a dairy sheep model. *BMC Genomics* 20 (719).

Palmer J. K. and List D. M. (1973). Determination of organic acids in foods by Liquid Chromatography. Journal of Agricolture and Food Chemistry 21(5), 903-906.

Patterson H. D. and Thompson R. (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545–555.

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

Peace C., Bassil N., Main D., Ficklin S., Rosyara U.R., Stegmeir T., Sebolt A., Gilmore B., Lawley C., Mockler T.C., Bryant D.W., Wilhelm L. and Iezzoni A. (2012). Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLOS ONE* 7(12).

Piaskowski J., Hardner C., Cai L., Zhao Y., Iezzoni A. and Peace C. (2018). Genomic hereditability in sweet cherry reveal non-addittive genetic variance is relevant for industry-prioritized traits. *BMC Genetics* 19 (23).

Pinosio S., Marroni F., Zuccolo A., Vitulo N., MAriette S., Sonnante G., Aravanopoulos F. A., Ganopoulos I., Palasciano M., Vidotto M., Magris G., Iezzoni A., Vendramin G. G. and Morgante M. (2020). A draft genome of sweet cherry (Prunus avium L.) revelas genome-wide and local effects of domestication. *The Plant Journal* 103 (4).

Poland J., Endelman J., Dawson J., Rutkoski J., Wu S., Manes Y., Dreisi- gacker S., Crossa J., Sánchez-Villeda H., Sorrells M., Jannink J. L. (2012). Genomic selection in wheat breeding using genotyping- by-sequencing. *Plant Genome* 5(3):103.

Predieri S., Ragazzini P. and Rondelli R. (2006). Sensory evaluation and peach fruit quality. *Acta Horticolturae* 713, 429-434.

Price A. L., Zaitlen N. A., Reich D. and Patterson N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11, 459–463.

Pryce J. E., Gonzalez-Recio O., Nieuwhof G., Wales W. J., Coffey M. P., Hayes B. J. and Goddard M. E. (2015). Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. *Journal of Dairy Science* 98, 7340–7350.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A. R., Bender D., Maller J., Sklar P., de Bakker P.i.W., Daly M. J. and Sham P.C. (2007). *PLINK*: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

Rafalski A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in Plant Biology* 5 (2), 94-100.

Ramming D. W. (1991). Genetic control of a slow-ripening fruit trait in nectarine. *Canadian Journal of Plant Science* 71, 601–603.

Rassam M., Bulley S. M. and Laing, W. A. (2007). Oxalate and ascorbate in Actinidia fruit and leaves. *Acta Horticolturae* 753, 479-485.

Reimer F. S. (1906). The honey peach group. *Florida Agricultural Experiment Station Bulletin* 73.

Resende Jr M. F. R., Munoz P., Acosta J. J., Peter G. F., Davis J. M., Grattapaglia D., Resende M. D. V. and Kirst M. (2012*a*). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193, 617–624.

Resende M. D., Resende M. F., Sansaloni C. P., Petroli C. D., Missiaggia A. A., Aguiar A. M., Abad J. M., Takahashi E. K., Rosado A. M., Faria D. A., Pappas Jr G. J., Kilian A. and Grattapaglia D. (2012*b*). Genomic selection for growth and wood quality in eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194(1), 116–28.

Riedelsheimer C, Czedik-Eysenberg A., Grieder C., Lisec J., Technow F., Sulpice R., Altmann T., Stitt M., Willmitzer L. and Melchinger A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* 44, 217–220.

Rousseeuw P. J. (1987). Silhouttes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 53-65.

Ruffner H. P. (1982). Metabolism of tartaric and malic acid in *Vitis*: a review-Part B. *Vitis* 21, 346-358.

Ruffner H. P., Possner D., Brem S. and Rast D. M. (1984). The physiological role of malic enzyme in grape ripening. *Planta* 160, 444—448.

Ruiz D. and Egea J. (2008). Phenotypic diversity and relationships of fruit quality traits in apricot (*Prunus armeniaca* L.) germplasm. *Euphytica* 163, 143-158.

Ruiz D., Lambert P., Audergon J. M., Dondini L., Tartarini S., Adami M., Gennari F., Cervellati C., De Franceschi P., Sansavini S., Bureau S., Gouble B., Reich M., Renard C. M. G. C., Bassi D. and Testolin R. (2010). Identification of QTLs for fruit quality traits in apricot. *Acta Horticulturae* 862, 587–592.

Sadar N., Krajnc U. A., Tojnco S., Tijskens L. M. M., Schouten R. E. and Unuk T. (2016). Development and distribution of quality related compounds in apples during growth. *Scientia Horticolturae* 213, 222-231.

Salazar J. A., Ruiz D., Egea J. and Martinez-Gomez P. (2013). Inheritance of fruit quality traits in apricot (*Prunus armeniaca* L.) and analysis of linked quantitative trait loci (QTLs) using simple sequence repeat (*SSR*) markers. *Plant Molecular Biology Reporter* 31, 1506-1517.

Sánchez-Pérez R., Pavan S., Mazzeo R., Moldovan C., Cigliano R. A., Del Cueto J., Ricciardi F., Lotti C., Ricciardi L., Dicenta F., López-Marqués R. L. and Moller B. L. (2019). Mutation of bHLH transcription factor allowed almond domestication. *Science* 364 (6445).

Sargolzaei M., Chesnais J. P. and Schenkel F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15.

Schmitzer V., Slatnar A., Mikulic-Petkovsek M., Veberic R., Krska B. and Stampar F. (2011). Comparative study pf primary and secondary metabolites in apricot (*Prunus armeniaca* L.) cultivars. *Journal of the Science of Food and Agriculture* 91(5).

Shirasawa K., Isuzugawa K., Ikenaga M., Saito Y., Yamamoto T., Hirakawa H. and Isobe S. (2017). The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Research* 12, 60–68.

Souty M., Audergon J. M. and Chambroy Y. (1990). Abricot: les critéres de qualité. *L'Arboriculture fruitiére* 430, 16-20.

Souty M., Audergon J. M. and Duprat F. (1991). Physical and biochemical criteria for apricot varieties characterisation. *Acta Horticolturae* 293, 95-110.

Souty M., Génard M., Reich M. and Albagnac G. (1999). Influence de la fourniture en assimilats sur la maturation et la qualite de la peache (Prunus persica L., cv. Suncrest). *Canadian Journal of Plant Science* 79, 259-268.

Spindel J., Begum H., Akdemir D., Virk P., Collard B., Redoña E., Atlin G., Jannink J. L. and McCouch S. R. (2015) Genomic selection and association mapping in rice (*Oryza sativa*). Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics* 11 (2).

Stanley J., Marshall, R., Tustin, S. and Woolf A. (2014). Pre-harvest factors affect apricot fruit quality. *Acta Horticulturae* 1058, 269-276.

Stephens M. J., Alspach P. A., Beatson R. A., Winefield C., Zealand N. and Buck E. (2012). Genetic parameters and breeding for yield in red raspberry. *Journal of the American Society for Horticultural Science*137 (4), 229-235.

Strazzer P., Spelt C. E., Li S., Bliek M., Federici C. T., Roose M. L., Koes R. and Quattrocchio F. M. (2019). Hyperacidification of *Citrus* fruits by a vacuolar proton-pumping *P-ATPase* complex. *Nature communication* 10 (744).

Su T., Wilf P., Huang Y., Zhang S. and Zhou Z. (2015). Peaches preceded humans: fossil evidence from SW China. *Scientific Reports* 5.

Sun Y., Qiao L., Shen Y., Jiang P., Chen J. and Ye X. (2013). Phytochemical profile and antioxidant activity of physiological drop of *Citrus* fruits. *Journal of Food Science* 78 (1), 37-42.

Sweetman C., Deluc L. G., Cramer G. R., Ford C. M. and Soole K. L. (2009). Regulation of malate metabolism in grape berry and other developing fruits. *Phytochemistry* 70, 1329-1344.

Tanger P., Klassen S., Mojica J. P., Lovell J. T., Moyers B. T., Baraoidan M., Naredo M. E. B., McNally K. L., Poland J., Bush D. R., Leung H., Leach J. E. and McKay J. K. (2017). Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. *Scientific Reports* 7 (42839).

Teissier M., Larroque H. and Graniè C. R. (2018). Weighted single-step genomic *BLUP* improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genetics Selection Evolution* 50 (31).

Topp B. L., Sherman W. B. and Raseira M. C. B. (2008). Low-chill cultivar development. In: *The Peach: Botany, Production and Uses* (pp. 106-138) edited by D. Layne and D. Bassi, CAB International, Wallingford (UK).

Tricon D., Bourguiba H., Ruiz D. and Blanc A. (2010). Evolution of apricot fruit quality attributes in the new released cultivars. *Acta Horticulturae* 814 (814), 571-576.

Troggio M., Malacarne G., Coppola G., Segala C., Cartwright D. A., Pindo M., Stefanini M., Mank R., Moroldo M., Morgante M., Grando M. S. and Velasco R. (2007). A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring 'Pinot Noir' Bacterial Artificial Chromosome contigs. *Genetics* 176, 2637-2650.

VanRaden, P. M. (2007). Genomic measures of relationship and inbreeding. *Interbull Bullettin* 37, 33–36.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414-4423.

Verde I., Bassil N., Scalabrin S., Gilmore B., Lawley C. T., Gasic K., Micheletti D., Rosyara U. R., Cattonaro F., Vendramin E., Main D., Aramini V., Blas A. L., Mockler T. C., Bryant D. W., Wilhelm L., Troggio M., Sosinski B., Aranzana M. J., Aru´P., Iezzoni A., Morgante M. and Peace C. (2012). Development and evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP detection and validation in breeding germplasm. *PLOS ONE* 7.

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

Vilanova S, Romero C, Llácer G, Badenes ML, Burgos L (2005) Identification of self-(in)compatibility alleles in apricot by PCR and sequence analysis. *Journal of the American Society of Horticultural Science* 130, 893-89.

Vizzotto G., Casatta E., Bomben C., Bregoli A. M., Sabatini E. and Costa G. (2002). Peach ripening as affected by *AVG*. *Acta Horticulturae* 592, 561–566.

Voss-Fels K. P., Cooper M. and Hayes B. J. (2018). Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics* 132, 669–686.

Walker R. P., Battistelli A., Moscatello S., Chen Z. H., Leegood R. C. and Famiani F. (2011). Phosphoenolpyruvate carboxykinase in cherry (*Prunus avium* L.) fruit during development. *Journal of Experimental Botany* 62, 5357-5365.

Wang H., Misztal I., Aguilar I., Legarra A. and Muir W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetic Resources* 94 (2), 73–83.

Welham S. J., Cullis B. R., Gogel B., Gilmour A. and Thompson, R. (2004). Prediction in linear mixed models. *Australian and New Zealand Journal of Statistics* 46 (3),325-347.

Wilks S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* 9, 60–62.

Windhausen V. S., Atlin G. N., Hickey J. M., Crossa J., Jannink J. L., Sorrells M. E., Raman B., Cairns J. E., Tarekegne A., Semagn K., Beyene Y., Grudloyma P., Technow F., Riedelsheimer C. and Melchinger A. E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 – Genes, Genomes, Genetics* 2, 1427–1436.

Wu B. H., Quilot B., Kervella J., Génard M. and Li S. H. (2003). Analysis of genotypic variation of sugar and acid contents in peaches and nectarines through the Principal Component Analysis. *Euphytica* 132, 375-384.

Würschum T., Reif J. C., Kraft T., Geert J. and Yusheng Z. (2013). Genomic selection in sugar beet breeding populations. *BMC Genetics* 14.

Zapata-Valenzuela J., Whetten R. W., Neale D., McKeand S. and Isik F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3 – Genes, Genomes, Genetics* 3.

Zhang, Q., Ma, B.Q., Li, H. et al. (2012) Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. *BMC Genomics*, 13, 537.

Zhang Q., Chen W., Sun L., Zhao F., Huang B., Yang W., Tao Y., Wang J., Yuan Z., Fan G., Xing Z., Han C., Pan H., Zhong X., Shi W., Liang X., Du D., Sun F., Xu Z., Hao R., Lv T., Lv Y., Zheng Z., Sun M., Luo L., Cai M., Gao Y., Wang J., Yin Y., Xu X., Cheng T. and Wang J. (2012). The genome of *Prunus mume*. *Nature Communications* 1318 (3).

Zhang Q., Feng C., Li W., Qu Z., Zeng M. and Xi W. (2019). Transcriptional regulatory networks controlling taste and aroma quality of apricot (*Prunus armeniaca* L.) during ripening. *BMC Genomics* 20(45).

Zhao Y. S., Gowda M., Liu W. X., Würschum T., Maurer H. P., Longin F. H., Ranc N. and Reif J. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics* 124, 769–776.

Zhebentyayeva T., Ledbetter C., Burgos L. and Llàcer G. (2012). Apricot. In: *Fruit Breeding* edited by Badenes M. L. and Byrne D. H., printed by Springer.

Zheng Y., Crawford G. and Chen X. (2014). Archaeological evidence for peach (*Prunus persica*) cultivation and domestication in China. *PLOS ONE* 9.

Zhong S., Dekkers J. C. M., Fernando R. L. and Jannink J. L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines. A Barley case study. *Genetics* 182 (1), 355–364.

Ziosi V., Noferini M., Fiori G., Tadiello A., Trainotti L., Casadoro, G. and Costa, G. (2008). A new index based on vis-spectroscopy to characterize the progression of ripening in peach fruit. Postharvest biology and technology 49 (3), 319-329.

Yang W., Duan L., Chen G., Xiong L. and Liu Q. (2013). Plant phenomics and high-throughput phenotyping: Accelerating rice functional genomics using multidisciplinary technologies. *Current Opinion in Plant Biology* 16 (2), 180-187.

Yoshida M. (1970). Genetical studies on the fruit quality of peach varieties. I. Acidity. *Bulletin of the Horticultural Research Station* 9, 1-15.

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

Xi W. P., Zheng H. W., Zhang Q. Y. and Li W. H. (2016). Profiling taste and aroma compound metabolism during apricot fruit development and ripening. *International Journal of Molecular Sciences* 17.

Xu K., Wang A. and Brown S. (2012). Genetic characterization of the *Ma* locus with pH and titratable acidity in apple. *Molecular Breeding* 30, 899– 912.

Xu S. Z. (2013) Genetic mapping and genomic selection using recombination breakpoint data. *Genetics* 195, 1103–1115.

ACKNOWLEDGMENTS

For me, my PhD experience consisted not only of researches, happiness and struggles but also of people. People who contributed to achieve results, giving me suggestions and advices but most of all who made me enjoy my PhD. I thank those persons who crossed my PhD path.

Dr. Anna Spinardi: You always gave me personal support, insightful teachings and guidance during my PhD. Thank you infinitely.

Prof. Bassi: I was lucky to work on a such large collection of peach and apricot. Thank you for every suggestion, and feedback on my work but most of all, thank you for having encouraged me to learn more by going to Australia.

Prof. Bruce Topp and Dt. Craig Hardner: It was great to have met you and learned from you spending six months in of the most amazing place worldwide. Believe me when I say that all your teachings will remain firstly in my heart. I greatly thank you.

Dr. Marco Cirilli: "Whoever PhD student finds a Marco, finds a treasure". Forgive me if sometimes I was "n'accollo", but you helped me to stay "'na crema". You were "'na spada" in all this PhD. I will never-endingly thank you.

Dr. Remo Chiozzotto: It is so hard to find such an extremely skilled-person, patient and taking care of everything as you are. I know that I was messy, but I hope that one day- not too far please...- you will forget my sticky laboratory bench and all the peach juice spread around the laboratory. Never-endingly thank you.

Unimi colleagues: Thank you all for every smile, nice thoughts and coffee together. My colleagues Dr. Leila Caramanico and Davide Bianchi are the first of this long list.

Unimi people: From Dr. Gabriella De Lorenzis to Dr. Debora Tura passing through each student, I am grateful for every nice thoughts and advice you had for me.

UQ people (St. Lucia Campus): Camille Branchereau (INRA, France), Ulduz Vafadarshamasbi, Daniel Garza, Michael Bird, Charlotte Rambla and Dr. Mulusew Fikre. It means a lot for me to have met you. I hope to see you again, in the meantime good luck with your PhD! Thank you for having joined me in Ozzy's adventures.

UQ people (Nambour): Jasmine Nunn and Katie Nieuwenhuis. You are such amazing persons, good luck for all your researches and I hope to meet you again. Thank you for the great times we spent together.

Maroochy Research Facility team: Dougal M. Russell and Dr Mobashwer Alam. Thank you for all the teachings. It was a great opportunity to work with you.

CRPV team: I am greatly grateful to Martina Lama and Stefano Foschi, for the technical support and assistance in the fields.