Investigating the fruit texture genetic control in apple and its interplay with the production of volatile compounds using multi-family based analysis and genome wide association mapping

Mario Di Guardo

Thesis committee

Promotor

Prof. Dr. R.G.F. Visser
Professor of Plant Breeding
Wageningen University & Research

Co-promotors

Dr. W.E. van de Weg
Senior Scientist, Plant Breeding
Wageningen University & Research

Dr. F. Costa
Senior Scientist, Edmund Mach Foundation
Istituto Agrario di San Michele all'Adige, San Michele all'Adige,Trento, Italy

Other members

Prof. Dr. L.F.M. Marcelis, Wageningen University & Research
Prof. Dr. R.D. Hall, Wageningen University & Research
Prof. Dr. J.J.B. Keurentjes, University of Amsterdam & Wageningen University & Research
Dr J.J. Mes, Wageningen University & Research

Investigating the fruit texture genetic control in apple and its interplay with the production of volatile compounds using multi-family based analysis and genome wide association mapping

Mario Di Guardo

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr. A.P.J. Mol,
in the presence of the Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 21 September 2017
at 11 a.m. in the Aula

Mario Di Guardo

Investigating the fruit texture genetic control in apple and its interplay with the production of volatile compounds using multi-family based analysis and genome wide association mapping, 177 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)

With references, with summary in English.

ISBN: 978-94-6343-205-4

DOI: https://doi.org/10.18174/414274

Table of Contents

Chapter 1

General Introduction

1) Rosaceae

The Rosaceae family is composed of approximately 3.000 species distributed all over the word, with a dominant presence in temperate climates. This family is characterized by a myriad of types, ranging from herbaceous plants to shrubs and trees. Many members of this family can be easily recognized by their edible fruits, particularly appreciated for flavors, colors and nutritional properties (eg. apple, pear, peach, apricot, cherry, plum, strawberry, raspberry). Other species, instead, are prized for their ornamental value (roses), nuts production (almonds) or even for timber production (black cherry; Table 1). The total world production of the edible rosaceous fruits in 2014 is estimated to 84.6 million tons. Taking also into account the annual production of nuts and roses (cut flowers and plants), the rosaceous compartment could be worth at least €55 billion/year at the farm gate, with a consumer value of about €180 billion/year (www.faostat.fao.org; Folta and Gardiner 2009).

The family of Rosaceae includes several species characterized by a considerable genetic diversity (Table 1; Shulaev et al., 2008). The karyotype (x) commonly ranges from seven to nine chromosomes, with exceptions. In fact, while the subfamily Rosoideae (including strawberry, raspberry and rose) is characterized by x = 7 (Stauct et al., 1989; Nybom et al., 2003), Amygdaloideae by x = 8 (apricot, sweet cherry, almond and peach; Foolad et al. 1995; Dirlewanger et al. 1998; Vilanova et al. 2003; Peace et al. 2012), and Spiraeoideae by x = 9 (Evans and Campbell, 2002), the genus Malus and Pyrus (Jung et al., 2012) are distinguished by x = 17. To date, different models have been proposed to explain this unique number of chromosomes within Rosaceae. The first and well accepted hypothesises involved a 'wide-hybridization' based on an allopolyploidization event between Spireoid (x = 9) and Amygdaloid (x = 8) ancestors (Chevreau et al., 1985). Successive phylogenetic studies based on molecular markers highlighted an alternative two-step hypothesis. Apple and pear would have been originated by an autopolyploidization (or hybridization) event between two taxa with x = 9, followed by diploidization and a subsequent aneuploidic reduction (x = 17). The research carried out by Velasco et al. (2010) suggested the genus Gillenia (or similar taxa) as the most probable ancestor of domesticated apple. This hypothesis agrees with the geographical finding that both genus Gillenia and the first fossil evidence of specimens belonging to the tribe Pyreae (that includes both apple and pear) were found in North America. This theory was further supported by the sequencing of the 'Golden Delicious' genome (Velasco et al., 2010), which indicated that the apple genome has derived from a relatively recent duplication which occurred between 48 and 50 million years ago. The molecular phylogeny of several apple genes also indicated most probably a monophyletic origin of Pyreae and Gillenia (Velasco et al., 2010; Jung et al., 2012).

Table 1. Overview of some of the most important economical species belonging to Rosaceae family. For each of those the taxonomy, the main uses and some information on the genome (ploidy level, chromosome number and genome size) are reported.

Subfamily	Genus	Species	Common Name	Uses	Ploidy	Chromosome Number	Genome Size	Released Genome sequence
Amygyloideae	Prunus	armeniaca	Apricot	Fresh and processed fruit	2n	x = 8	240 Mb	
		avium	Sweet cherry	Fresh and processed fruit	2n	x = 8	388 Mb	
		cerasus	Tart (sour) cherry	Fresh and processed fruit	4n	x = 8	599 Mb	
		dulcis	Almond	Fresh and processed fruit	2n	x = 8	240 Mb	
		persica	Peach, nectarine	Fresh and processed fruit	2n	x = 8	290 Mb	Verde et al. 2013
		serotina	Black cherry	Timber species	4n	x = 8	490 Mb	
Maloideae	Malus	x domestica	Apple	Fresh and processed fruit	2n, 3n	x = 17	750 Mb	Velasco et al. 2010
	Pyrus	communis	European pear	Fresh and processed fruit	2n	x = 17	577 Mb	Chagne et al. 2014
Rosoideae	Fragaria	x ananassa	Strawberry	Fresh and processed fruit	8n	x = 7	240 Mb	VanBuren et al. 2016
	Fragaria	vesca	Woodland strawberry	Fresh and processed fruit	2n	X = 7	240 Mb	Shulaev et al. 2011
	Rubus	occidentalis	Black raspbarry	Fresh and processed fruit	2n	X = 7	240 Mb	
	Rosa	spp.	Rose	Cut flowers, landscape ornamental, medicinal	2n, 3n, 4n	x = 7	600 Mb	

To this end, the availability of a reference genome for apple provided a great contribution to the determination of its phylogeny. The ‘Golden Delicious’ genome was the first one to be published within the family of Rosaceae (Velasco et al., 2010), followed by strawberry (Shulaev et al., 2011), peach (Verde et al., 2013), pear (Chagné et al., 2014) and black raspberry (VanBuren et al., 2016)

2) Apple

Apple (Malus x domestica Borkh.) is the most widely cultivated species of the genus Malus. The center of origin has been established in Central Asia, where the wild ancestor of apple, Malus sieversii, (Harris et al., 2002) was found. The denomination Malus x domestica has been generally accepted as the appropriate scientific name for domesticated apple (Korban et al., 1984).

2.1) Economic Importance

Apples are cultivated in all temperate and subtropical world zones, even though a prominent part of the production is centered in the cool temperate regions of the world, spanning from 35° to 50° latitude (Kellerhals, 2009). Apple is one of the most appreciated fruit crops world-wide, giving its nutritive properties and excellent storability, which can guarantee its availability for a year-round period. With more than 50 million tons produced every year (www.faostat.fao.org), apple is the most important rosaceous species for annual world fruit production and it is ranked second in order of economic importance within the fruit crops after banana and immediately before grape and orange (www.faostat.fao.org). Although apple is mainly cultivated for fresh consumption, a considerable proportion of fruit is also produced for industrial processing (snack, juice, cider and brandy).

2.2) The apple genome

The physical length of the apple genome is estimated to be 750 Mb (Velasco et al., 2010), organized into 2n = 34 chromosomes. Although some varieties are known to present a different level of ploidy (Tab. 1), most of the apple cultivars are diploid (2n=2x), with a total genetic length estimated around 1230 cM (Di Pierro et al., 2016).

2.3) Apple breeding

The first controlled apple cross has been attributed to Thomas A. Knight (1806). Since then, breeding underwent an important improvement both in terms of strategies and techniques. The driving forces in apple breeding include the selection of fruits with improved appearance

(size, color; Kenis et al., 2008; Di Guardo et al., 2013; Costa, 2015), flesh textural characteristics (Costa et al., 2012; Longhi et al., 2012; Amyotte et al., 2017) and resistance against pathogens (Seglias & Gessler 1997; Hemmat et al., 2002; Gygax et al., 2004). More recently, increasing efforts have also been made to enhance nutritional properties (Guan et al., 2015) and environmental sustainability (Kumar et al., 2014; Allard et al., 2016).

The main constraints in apple breeding are however represented by the long juvenile period (it may easily take five years from seed to a fruit bearing tree), the size of the plant (demanding large experimental plots) and the genetic complexity due to the self-incompatibility proper of apple. This implies that offspring grown from seed does not resemble the mother apple tree (Cornille et al., 2014). At the same time the creation of near isogenic lines (NIR), backcross (BC) or recombinant isogenic lines (RIL) populations, mating schemes useful to fix the desired traits on self-compatible crops (Monforte and Tanksley, 2000), is unfeasible. The self-incompatibility of apples, combined with the above-mentioned limitations related to woody crops, makes the direct selection of interesting phenotypes a time and space-demanding process. Milestones in apple breeding were the introduction of vegetative propagation by grafting and the use of dwarfed apple trees as rootstocks (Cornille et al., 2014). A recent impulse to apple breeding has been represented by the employment of genetic information to assist and guide the selection process. Details on this issue will be further discussed in the coming paragraphs, while the establishment, use and implications of genetic informations in apple breeding are the red thread of the thesis.

3) Molecular markers

The essence of breeding is the selection of the best performing individuals with superior and improved traits. Selections can be made when differences among individuals become evident and measurable. The use of markers to explain diversity can be dated back to the XIX century: Gregor Mendel employed phenotype-based markers (plant height, pod shape and colour, seed shape and colour etc.) on his pioneering studies on pea. The same type of marker was further used to explain genetic linkage (Morgan, 1909), the most significant exception to Gregor Mendel's low of independent assortment (Agarwal et al., 2008). Phenotypic markers were also employed in apple: acidity of leave juices for instance was used as an estimator for fruit acidity (Visser and Verhaegh, 1978). Although these types of markers gave a great impulse to the first genetic discoveries, they showed important limitations in terms of reproducibility (phenotypic markers can be strongly affected by the environment) and availability (they can be used only when the phenotype is expressed and became visible). From this point, an extraordinary step forward was represented by the discovery of DNA and the unravelling of its role in biological processes. These findings, together with the advances in the techniques to

isolate and amplify nucleic acids, enabled the development of a new type of markers (named molecular markers), based exclusively on the intrinsic characteristics of an individual's genome. Each DNA region (locus) characterized by multiple forms (polymorphic: polys = multiple, morphe = forms) within a population and inherited in a Mendelian fashion can be considered as a molecular marker, which allows the grouping of a set of individuals into as many subgroups as the number of forms (alleles) identified at that locus. This finding has been widely used for molecular taxonomy studies, identification of individuals (fingerprinting), germplasm characterization, estimation of genome size, forensic analysis, generation of linkage maps and marker-trait association studies. The last two, in particular, enabled the shift from the traditional breeding into the well-known molecular plant breeding.

The use of molecular markers, rather than phenotypic markers, relies on the fact that: (i) molecular markers are not influenced by pleiotropic or epistatic effects or environmental factors; (ii) they are potentially uniformly distributed over the genome (introns and exons) allowing the identification of differences among individuals despite their phenotype; (iii) in most of the cases they are co-dominant allowing, therefore, the detection of both homozygous and heterozygous genetic configurations and (iv) molecular markers are not influenced by the physiology of the trait in analysis (Khan et al., 2015).

Although each type of molecular marker is characterized by a unique combination of advantages and constraints, the choice of the appropriate marker techniques mainly depends on their final application. Factors that must be taken into account in choosing a marker system are the required quantity and quality of DNA, the availability of a (preferably) public collection of DNA markers for the species under investigation, the costs and time needed for genotyping and the level of throughput/automation (Williams et al., 1993). To date many types of molecular marker techniques are available: the first widely used molecular marker was the Restriction Fragment Length Polymorphism (RFLP) (Botstein et al., 1980). This technique has serious limitation in throughput since it requires high quantity of sample DNA and the entire process can take several weeks to complete.

During the last 50 years, molecular markers experienced a great evolution in both properties and potential applications. From the first type of hybridization-based marker (RFLP) the discovery of the PCR mechanism has greatly contributed to define a novel generation of markers (e.g. RAPD, CAPS, SCAR, SSR and AFLP) easier to apply and efficient (Williams et al., 1990; Vos et al., 1995; Selkoe & Toonen 2006; Jaillon et al., 2007; Sato et al., 2012). Despite the technology at the base of their detection, molecular markers can be classified on the base of their mode on inheritance and allelic detection. Markers can be in fact distinguished in codominant (normally with a mendelian inheritance and able to differentiate heterozygous from homozygous) and dominant (which can be inherited following a non-mendelian fashion and not

efficient to distinguish the hetero from the homozygous). Ultimately, the advent of sequencing and high-throughput genotyping systems has revolutionized the accessibility and the application of molecular markers for a myriad of different applications (Boutet et al., 2016).

Nowadays, the most widely used molecular marker technique is based on the analysis of single nucleotide polymorphisms (SNPs) naturally occurring in each organism, which represent the most abundant source of variation within each genome. This polymorphism originates from mutation and DNA repairing events leading to variation of one nucleotide between homologous chromosomes. SNPs are present both in coding and non-coding DNA regions, with a higher proportion in non-coding regions, since they do not undergo selective pressure. Within a coding region a SNP is either non-synonymous, resulting in an amino acid sequence change, or synonymous (Sunyaev et al., 2000). In the latter case, the nucleotide change does not alter the amino acid sequence, although a synonymous SNP can interfere with the translation process through the modification of mRNA splicing (Richard and Beckmann, 1995). SNP density, moreover, varies significantly across species: maize has a dense SNP coverage (one SNP every 60-120 bp) (Ching et al., 2002) while SNP frequency decreases to one every 1.000 bp in humans (Shastry et al., 2002). In apple a SNP is found approximately every 288 bp (Chagné et al., 2012). During the past 20 years, automation and miniaturization in SNP-based marker technologies enabled to increase the number of markers analysed while reducing cost and time requirements (Deschamps et al., 2012). The emergence of fluorescent labelling and massively parallel array systems made the simultaneous scoring of up to hundreds of thousands of markers a reality. Several SNP array genotyping platforms are commercially available to date, including Illumina Infinium® platform (Illumina Inc.) and the Axiom® platform (Affymetrix). Specific SNP arrays have been already developed for apple: from the initial 8K Infinium® array (Chagné et al., 2012), to the 20K Infinium® array (Bianco et al., 2014) and the 480K Axiom® array (Bianco et al., 2016).

4) Molecular Breeding

The apple panorama harbours a great diversity in terms of physiology, production and disease resistance. This phenotypic variability is mainly due to the underlying genetic complexity derived from multiple loci interacting together. A genomic region containing one or more genes affecting a quantitative trait (for the most agronomically important) is called Quantitative Trait Locus (QTL). The conceptual basis for the genetic dissection of multi-genic traits is relatively straightforward. QTL mapping relies on finding an association between a genetic marker and the phenotype of interest. QTL mapping requires the availability of either a genetic linkage map, in which the molecular markers are grouped and ordered according to their recombination frequencies within the studied population, or a physical map in which the order of the markers

is based on their physical distances. Once all individuals in the mapping population are genotyped and phenotyped for the trait of interest, QTL mapping can take place. Mapping of a QTL has become a common first step towards the understanding of the genetic basis of complex genetic traits. Many different techniques have been proposed so far, which can be mainly grouped into what has been described as linkage or association mapping approaches (Fig. 1), both sharing the same guiding principle: the *'recombination's ability to break up the genome into fragments that can be correlated with phenotypic variation'* (Myles et al., 2009).

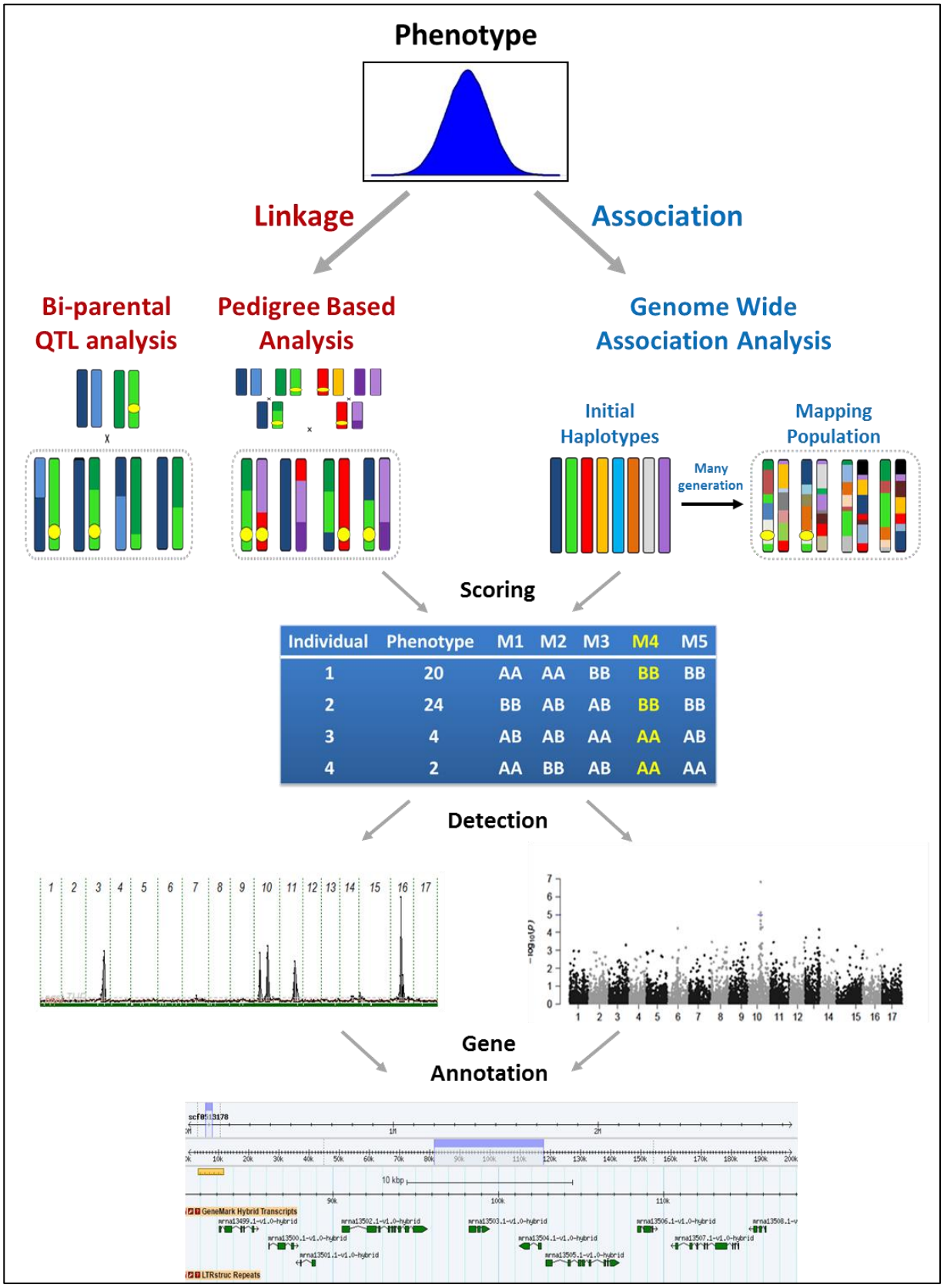


Figure 1. From phentype to causal gene identification: workflow of different marker – trait association techniques (bi-parental QTL analysis, Pedigree Based Analysis and Genome Wide Association Analysis) highlighting differencies and similarities.

4.1) Bi-parental QTL analysis

One of the first approaches, and maybe the most widely used technique to dissect the genomic regions involved in a trait of interest, is the bi-parental QTL analysis. This technique makes use of a full-sib (FS) family that arose from crossing two parents, often characterized by a divergent phenotype for the trait under investigation. This approach is based on the analysis of the phenotypic variation and the genotypic segregation of the offspring to detect which of the parental marker alleles are linked to the phenotype of interest (Fig. 1). Key factors for this analysis are the quality of the phenotypic data, the size of the FS family, the marker density and the heritability of the trait. All aspects have a direct influence on the quality and reliability of the analysis, since the number of individuals reflects the number of meiosis (and therefore the number of examined recombinations) and marker density influence the size of the confidence interval of the QTL (Mauricio et al., 2001). The ongoing bi-parental QTL analyses of complex traits in outcrossing plants contributed to the understanding of quantitative trait genetics through the discovery of many QTLs. However, only few of these QTL have been adopted by breeders for marker-assisted breeding (MAB) due to various reasons (van de Weg et al., 2004; Bink et al., 2014) including:

- Most QTL discoveries have been based on germplasm with a narrow genetic basis (often just a single family). As a consequence, only a small proportion of the total number of relevant QTLs has been detected, which may explain only a limited fraction of the total genetic variance present within a breeding program (many QTL will not segregate in the selected single progeny).
- Many useful functional alleles within a QTL are missed as these are not present or do not segregate in specific single mapping families.
- For most QTLs, little is known about their mode of action and robustness across different genetic backgrounds, i.e., the estimated magnitude of the QTL may be different for families derived from other parents.

To overcome these limitations novel approaches have been recently adopted, such as Pedigree Based Analysis (on FS families) and Genome Wide Association Analysis (on germplasm collections).

4.2) Pedigree Based Analysis

The interest in the use of multiple genetically related FS families in dissecting quantitative trait variation has grown rapidly during the last decades (Blanc et al., 2006; Yu et al., 2008; Würschum et al., 2012). In the presence of pedigree structures, the explicit modelling of familiar

relatedness in QTL and association mapping approaches may significantly improve the detection power (Bink and Van Arendonk, 1999; Yu et al., 2006). Until recently, the experimental set-up of such QTL studies was often restricted to pre-defined fixed designs such as factorial or bi-allelic state to allow standard statistical analyses. To better explore available FS families, more flexible statistical procedures are required to utilize complex pedigree relationships (Fig.1). The Pedigree Based Analysis (PBA) approach supports the use of multiple segregating FS-families into a single analysis, making use of pedigree information, molecular marker data, and the Identity By Descent (IBD) concept, thus allowing generalized conclusions across wide germplasms (Van de Weg et al. 2004). The Bayesian approach uses probability models that through the analysis of quantities that we observe and measure (phenotype, molecular markers) infers unknown quantities we are interested in (QTLs) (Bink et al. 2008). This novel approach steps forward in the genetic analysis of a trait since it supports the discovery of QTLs as well as their characterization by estimating location, size, mode of action (additive, dominance), ancestral origin, and linked marker haplotype(s) (Bink et al., 2014). Pedigree records allow a deeper exploration of QTLs variation than single QTL analysis, since the number of alleles analysed is generally a function of the number of founders: more founders are taken into account, more probable it will be that a valuable QTL allele is included in the germplasm in analysis (Bink et al., 2008). On the other hand, performing QTL analysis in such pedigreed populations is challenging since it requires a high level of data integration, pedigree records have to be curated and molecular markers checked for consistent calling across the entire germplasm. Also, pedigrees must be trimmed if ancestors are not available (anymore) for genotyping or if they cannot be imputed. Next, adequate phasing might require enrichment of the germplasm with direct descendants of the major founding cultivars. Finally, the QTL models used by the software may not encompass the actual mode of QTL performance, such as the presence of intra and inter locus interactions and the presence of QTL that contain multiple functional alleles with different effect.

4.3) Association Mapping

While in classical QTL mapping, carried out in outcrossing species, only the recombination events within the two parents used to create the mapping population are exploited (Breseghello & Sorrells 2006; Myles et al., 2009; Rafalski et al., 2010), the association studies are based on the analysis of individuals with unobserved ancestry. A statistical significant association between genotypes at one or more marker loci and the phenotype is usually considered to be evidence of close physical linkage between the marker locus and a QTL for the trait of interest (Pritchard et al., 2000). This relationship strictly depends on the linkage disequilibrium (LD), which is a parameter describing the non-random association

between two genomic regions within the same chromosome. In a population of fixed size undergoing random mating, repeated recombination events would interrupt the order of the ancestral genetic segments of contiguous chromosomes, reaching the point that all alleles in the population are independent (linkage equilibrium). The rate of LD decay depends on several factors including population size, number of founding chromosomes (representing the actual allelic variability of the population under examination), the number of generations that occurred between the founders and the examined germplasm and the number of markers used for the chromosome haplotyping. The association mapping can be exploited at two levels, candidate gene based association and genome wide association (Rafalski et al., 2010). The first approach requires a good understanding of both the physiology and genetics of the trait under investigation, since the analysis is aimed to test whether a correlation between one or more (known) functional markers and the trait of interest exists or not. Since the traits of agronomical interest are mostly polygenic, this approach does not allow a genome-wide investigation of the genomic regions influencing the trait under consideration (Palaisa et al., 2003; Pozniak et al., 2007). On the other hand, genome wide association studies (GWAS) do not need a deep prior knowledge of the genetics of the trait since ideally all chromosomes should have adequate marker coverage and the marker-trait association analysis is evaluated genome-wide. The principle behind GWAS is that one (or more) of the molecular markers being considered is either causal for the trait (functional marker) or in LD with the causal locus.

The design of a GWAS experiment must consider several factors such as the choice of the composition of the population and the marker density. An inaccurate set up of these two aspects will lead to type one and/or type two errors. Ideal germplasm in GWAS should not present population structure (a structured population is a set of individuals characterized by distinct subgroups related by kinship). Population structure, if not adequately corrected and taken into account, can give rise to spurious associations (Yu et al., 2008). Different methods and software have been proposed to calculate population structure (Pritchard et al., 2000, Rafalski et al., 2010). One of the most widely used methods involves a Principal Component Analysis (PCA) based on the genotypic data available (Patterson et al., 2006), while software named STRUCTURE (Pritchard et al., 2000) calculates the most probable number of subpopulations (k) on the dataset in analysis. Basic concept in designing GWAS experiment is the knowledge about the LD decay level present in the species under investigation. With a rapid LD decay (short LD extent) a higher number of makers is required to target the QTL since the chromosome will be broken into several haplotype blocks. On the contrary, with a slow LD decay (large LD extent) a minor number of markers will be instead needed to identify the region harbouring the QTL. As a consequence, with a rapid LD decay there will be also a higher chance to tackle the causal gene controlling the phenotype of interest.

A limitation of GWAS is the detection of functional variants that have low frequencies in the germplasm to be analysed. Low frequency alleles have little influence on the population as a whole and their signal is therefore difficult to detect (even if these low alleles have an enormous effect on the phenotype). In such case GWAS can be coupled with linkage mapping, by creating a controlled cross and increasing the frequency of the rare alleles to have in the end a better understanding of its specific effect on the phenotype.

5) Marker Assisted Breeding (MAB)

Plant breeding is based on the phenotypic selection of superior genotypes via the use of ad-hoc designed segregating progenies (Francia et al., 2005). Traditionally, the selection process relies on a phenotypic evaluation of the promising seedlings assuming a direct correlation between favourable phenotype and favourable genotype. From a theoretical point of view this assumption is correct, but in real life a fairly good relation between phenotype and genotype can be postulated only for traits being not strongly influenced by environmental changes. The phenotype is therefore the result of the interaction between these two variables as following:

$$P = G + E + (G \times E)$$

where P is the phenotype of interest, G the genotype and E is the environment.

To overcome the limitations and constraints proper of the traditional selection based only on the phenotype, breeders started to implement molecular markers associated to the traits needing improvement as a selection tool, enabling the so-called Marker Assisted Breeding (MAB). While in traditional breeding the selection of the best performing individuals passes through the selection of the best phenotype, in MAB, polymorphic DNA regions (molecular markers) are linked to differences in phenotype providing a more precise and efficient selection procedure. With the availability of the first molecular markers and the consequent generation of the first genetic maps, MAB has become affordable both for monogenic traits governed by major genes as well as for polygenic traits controlled by several genes (Fig. 2). A successful application of MAB requires as a pre-requisite the availability of a tight association between markers and the genes affecting the phenotype. These markers can be then used to (i) identify, among segregating progenies, the most suitable individuals based on their allelic composition and (ii) trace favourable alleles across generation (following the IBD approach) to accumulate favourable alleles in progenies.

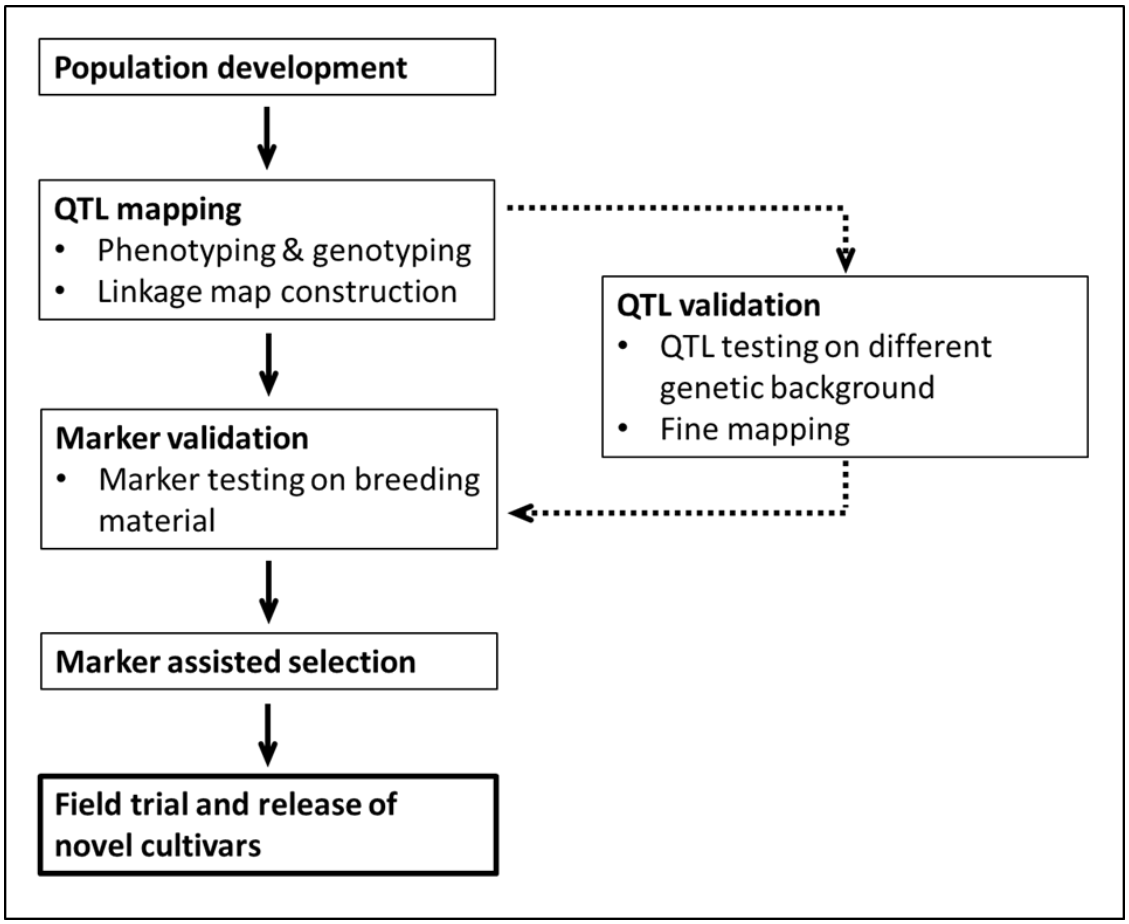


Figure 2. Workflow of the experimental design for marker assisted breeding.

Nowadays, most of the molecular markers used in MAB in apple breeding programmes are associated to monogenic traits or to genes with relevant impact on the phenotype. Most of these genes are related to disease resistance (R genes), such as apple scab (Vf gene; Chevalier et al., 1991; Tartarini et al., 1999) or powdery mildew (Pl1, Pl2, Plw and Pld genes; Dunemann et al., 2007) or quality traits such as the columnar growth (Co; Baldi et al., 2013) or Ma gene, the latter involved in the regulation of the acidity content in fruit (Maliepaard et al., 1998). Nevertheless, many studies have been carried out on quantitative traits as well (Calenge et al., 2005; Kumar et al., 2012; Amyotte et al., 2017) to define genomic regions linked to important traits for breeding. The advance in computation power and the possibility of generating large-scale marker arrays enabled the determination of the genetic basis of potentially any agronomical important trait. The complete chromosome haplotyping will allow breeders to design superior genotypes ‘in silico’ by combining the most favourable alleles at all the loci of interest towards a Breeding By Design approach (Peleman and Van Der Voort, 2003).

6) Breeding for fruit quality

From the second half of the XX century the consumers became more demanding for an improved quality of the fruits, especially considering their impact on the human health. Fruit quality is, essentially, the resultant of the interplay between colour, flavour, texture, size, and nutritional properties (Bourne et al., 2002; Sansavini et al., 2004; Costa et al., 2011). Albeit all these characteristics contribute to what is perceived as fruit quality, breeding efforts were historically aimed to improve productivity traits such as yield, shelf life and disease resistance. Different studies have been carried out to identify genomic regions involved in fruit quality. Chagné et al., (2013) identified a transcriptional factor, MYB1, located on chromosome 9 directly related to fruit skin colour regulation. Other members of MYB transcriptional factors control also fruit flesh and foliage anthocyanin pigmentation (Maliepaard et al., 1998; Chagné et al., 2012) while other studies were focused on deciphering the genetic control of productivity traits such as biennial bearing (Guitton et al., 2012). In apple great attention was also devoted to the understanding of the genetic and physiological mechanisms regulating fruit texture, since textural properties are directly related to storability and consumer appreciation (King et al., 2000; Costa et al., 2010; Bink et al., 2014). On the other side, another -perhaps less investigated aspect of apple quality- is flavour. The concept of quality in fruit is not static, but evolves through time. In past decades, aroma components were neglected in place of other fruit quality traits, ending with the release of novel apple cultivars with enhanced textural or resistance properties but with poor flavour components. Nevertheless, apple aroma is well appreciated by consumers (Baldwin et al., 2002), and represents an important aspect influencing consumer's choice. For this reason, a major goal in apple breeding would be the combination of good textural performance with a favourable aromatic bouquet. The interplay between these two traits will be further elaborated and discussed in the chapter five of this thesis, while chapters three and four are specifically dedicated to fruit texture.

6.1) Texture

Fruit ripening per se is a fundamental process making fruits edible and desirable to enable the release of seeds by animals in the outer space, but it is also the natural programmed process leading to senescence, characterized by a severe firmness loss. Fruit flesh generally tends to soften with the ongoing of the fruit ripening due to a decreased cell-to-cell adhesion, finally resulting in a progressive cell separation, which generally ends with a poor juice release and dry mouthfeel in overripe and mealy fruit. Beside this, fruit ripening can be also accompanied by a reduction in the cell turgor due to an increased concentration of solutes in the cell space as well as to wall loosening (Brummell and Harpster 2001). In the modern

horticulture, fruit texture plays two fundamental roles for apple. First, fruit texture is the most appreciated feature by consumers, and since it is directly perceived by the human sense it has the ability to drive the final choice. In addition, fruit texture, related to the properties of the cell wall structure, influence directly the storability of the fruit, thus its commercial success and postharvest management.

Fruit texture is nowadays considered as a multi-trait feature, being composed by mechanical and acoustic components. The first one is represented by sub-phenotypes related to the physical resistance exerted by the cell wall upon external pressure, such as firmness, mealiness, sponginess and hardness. The acoustic signature, instead, is in essence based on the air pressure released by the cell wall breaking, and comprehend crispness and crunchiness (Vincent et al., 1998). The importance of fruit texture relies also in the fact that the texture component is often highly correlated with the juiciness perception (Bourne et al., 2002; Varela et al., 2007). Among the acoustic components, crispness, in particular, is the feature most appreciated by consumers, due to the general thought that a crispy fruit is also characterized by a general healthy state (Hampson et al., 2000). Crispness is based on the rupture of turgid cell wall, after which a sound pressure wave generates a vibration between molecules around their equilibrium, producing the typical sound perceived as crispness (Duizer et al., 2001; Demattè et al., 2014). In the opposite case, with a low cell wall turgidity or for an excessive pectin polysaccharide solubilisation of the middle lamella, the cells tend to separate instead of breaking down. This eventually results in a rubbery dry and mealy texture (Reeve 1953; Reeve 1970; Niklas 1992; Andani et al., 2001). Differences in the regulation of the cell wall/middle lamella polysaccharide biosynthesis among cultivars is the final cause of the great variation in fruit texture within the apple panorama.

6.1.1) Physiology of fruit texture and cell-wall metabolism

Fruit texture largely depends on the characteristics of the cortex cell wall, in particular on (i) the chemical binding of the polysaccharide structure, (ii) the internal turgor pressure, and (iii) the intercellular adhesion between cells. The dissolution of the middle-lamella, a structure rich in pectin devoted to control the cell-to-cell adhesion, is one of the first changes observed in a ripe fruit by electron microscopy, followed by the disruption of the ordered structure of the primary cell wall, indicating that significant degradation has occurred. Throughout fruit development, maturation and ripening, changes in fruit texture are governed by the synergistic and coordinated action of several enzymes progressively acting during the cell wall disassembly to dismantle the cell wall polysaccharide architecture, as well as altering the linkage between polymers. The timing and intensity of these processes can vary among species and cultivars, leading to distinct types of texture behaviour. For instance, dissolution of the middle-lamella can

initiate at early stages during ripening in species such as tomato, or in late ripening phases in crispy types of fruits, such as apple (Seymour et al., 2002; Toivonen and Brummell, 2008). The cell wall structure can be interpreted as a network of rigid inextensible cellulose microfibrils held together by an interpenetrating and coextensive network of glycan, pectin and structural glycoproteins (Cosgrove 2000; Brummell 2006). In dicots, the cellulose microfibrills are coated and crossed-linked by a glycan matrix, of which the xyloglucan is generally the most abundant (Brummell and Harpster, 2001). The xyloglucan matrix firmly binds the cellulose microfibrils between them. The spaces in the cellulose matrix glycan network are filled by highly hydrated pectins which can also form a network, held together by ester bounds between pectin molecules and ionic calcium cross links. The cellulose/glycan matrix can be embedded and bound to the pectin network by covalent links established between xyloglucan molecules and pectin polysaccharides.

In several fruits, the most active enzymes responsible for the pectin modification are polygalacturonase (PG) and pectin methylesterase (PME), while those acting on the primary cell wall are xyloglucan endotransglycosylase (XET) and expansins (Exp). Among them. PG is a known major enzyme involved in pectin disassembly by biochemically catalysing the hydrolytic cleavage of α (1–4) galacturonan (Brummell and Harpster, 2001; Costa et al., 2010; Farrokhi et al., 2013) and it is normally encoded by multigene families (Sitrit and Bennett, 1998). The PG action in apple (as well as in other climacteric fruits) is physiologically related with the presence of the plant hormone ethylene, required to trigger and coordinate several changes of the fruit ripening (Liu et al., 1985), and influencing the cascade activation of many ripening related genes (Rose et al., 1998; Cosgrove 2000; Cosgrove 2001). The direct physiological relationship between the ethylene production and the rate of fruit softening in climacteric fruit was also established in apple, where increased expression of the final enzymes in the ethylene pathway (Md-ACS1 and Md-ACO1) paralleled mRNA accumulation of two genes related to the cell wall disassembly (Md-Exp and Md-PG1) (Wakasa et al., 2006; Zhu and Barritt, 2008). As proof of this relationship it has also been demonstrated that the application of 1-MCP (ethylene competitor) efficiently inhibits the transcription of Md-PG1 (Wakasa et al., 2006; Costa et al., 2010).

6.1.3) New Phenotyping Approaches

To date, the overall fruit texture complexity in apple has been largely assessed using a penetrometer, a simple equipment measuring the maximum force needed to enter a probe into the fruit flesh. Although this device has been extensively used for fruit firmness characterization, it has also been recently documented that it might be not sufficient for a reliable and precise investigation of several fruit texture sub-phenotypes (Harker et al., 2003; Costa et al., 2011;

Longhi et al., 2012). In order to efficiently perform a more comprehensive characterization of fruit texture, a new equipment combining a classical penetrometer and an acoustic device was employed here to profile the complete fruit texture behaviour. Since the phenotype resolution is the key factor for a precise QTL characterization (Schauer et al., 2006), this novel approach can be applied to identify new sets of QTLs related to important texture components, such as fruit crispness.

6.1.4) Genetic dissection through QTL mapping

Over the past 15 years several QTL mapping studies have been successful in identifying many genomic regions involved in fruit firmness, crispness and juiciness (Maliepaard et al., 1998; King et al., 2000; Liebhard et al., 2003; Costa et al., 2005; Costa et al., 2008; Kenis et al., 2008; Costa et al., 2010; Bink et al., 2014; Cappellin et al., 2014; Chagné et al., 2014; Costa 2014). However, most of the discovered QTLs have not yet been sufficiently characterized and validated. Most of them still have large confidence intervals due to small germplasm size or low marker density. Consequently, also their performance has not yet been clarified, and QTL genotypes of founders and major parents remained mostly unresolved. The identification of a high number of QTLs is not surprising considering the complexity of the involved process such as the high number of genes controlling cell wall metabolism, which is estimated to represent almost 10% of the entire gene set of the genome in the plant species sequenced (McCann and Rose, 2010). Consequently, a high number of genes are available for which a mutation may have considerable effects on texture. QTL analysis have identified different candidate genes involved in fruit texture changes, including Md-ACO1, Md-Exp7, Md-ERS1 and Md-PG1, which are located on four different genomic regions on linkage groups LG1, LG 3 and LG10, respectively (Costa et al., 2005; Costa et al., 2008; Tatsuki et al., 2009; Costa et al., 2010; Harb et al., 2012; Longhi et al., 2012), and involved in different important physiological pathways, such as ethylene production (Md-ACO1), ethylene perception (Md-ERS1) and the degradation of polysaccharides (Md-Exp7 and Md-PG1).

6.2) Flavour and aroma

Flavour is one of the most important and distinctive features in apple, resulting from the interaction between taste (mainly determined by the interplay between sweetness, sourness, acidity and astringency) and aroma (due to the production of odour-active volatile compounds). Whereas taste is determined by relatively few components, aroma results from the simultaneous presence of a vast number of organic volatile compounds (VOCs) whose composition is species, and often, variety-specific. In apple more than 370 VOCs have been measured (Dixon

and Hewett, 2000), although not all of them affect apple aroma directly (Espino-Díaz et al., 2016). The VOCs that greatly contribute to apple aroma belong to the classes of esters, alcohols and aldehydes (Dimick et al., 1983). Their mode of action and relative presence can vary considerably through the entire economic life of apple: the aromatic profile of immature apple is mainly due to the presence of aldehydes (Paillard 1990), while in ripe apple the relative presence of aldehydes decreases in favour of alcohols and especially esters (Kakiuchi et al. 1986). Albeit the characteristic apple aroma is given by mainly alcohols and esters, their relative presence - as well as the overall VOCs composition - varies among cultivars (Kakiuchi et al., 1986; Pooter et al., 1986; Holland et al., 2005) and within a cultivar, since VOCs production can be influenced by several biotic and abiotic factors (Espino-Díaz et al., 2016). Despite the high number of VOCs produced, only a small proportion generates the specific 'flavour fingerprint' characteristic of each fruit (Farneti, 2014). Aldehydes, which are a major component of the aroma of unripe apples, derive from the catabolism of either fatty acids or of the aliphatic branched-chain amino acids such as leucine, isoleucine and valine (Rowan et al., 1996). More than 25 aldehydes were reported in apple (Dimick et al., 1983), with hexanal, trans-2-hexanal and butanal as the ones with the highest concentration (Espino-Díaz et al., 2016). The reduction of aldehydes by the enzyme alcohol-dehydrogenase (ADH) leads to the formation of alcohols (Bartley and Hindley 1980). In particular, linear alcohols derived from the catabolism of fatty acids, while branched amino acids are the precursor of branched alcohols. The most abundant alcohol components affecting ripe apple aroma are: 2-methyl-1-butanol, 1-butanol, 1-hexanol, 1-propanol and 2-methyl-1-propanol (Espino-Díaz et al., 2016). In apple, at maturity stage, alcohols range from 6 to 16% of the total VOCs composition. Esters, the most abundant VOC in ripe apple, are derived from alcohols through the action of the alcohol acyltransferase (AAT) enzyme that transfers an acyl group from acyl-CoA to the hydroxyl group of the alcohol. In ripe apple esters relative frequency can vary from 80% (measured in Golden Delicious) (López et al., 1998a) to 98% (in Stark Delicious (López et al., 1998b), confirming their role in the aromatic bouquet of apples.

Breeding for fruit with improved aroma characteristics represents an intriguing challenge for breeders and researchers. Breeding for such trait requires a deep understanding of the VOCs biochemistry and their genetic regulation together with sophisticated equipment for phenotyping. Moreover, selection for other agronomical or quality traits such as yield, fruit size and fruit firmness might have negatively affected fruit flavour since this characteristic was not usually taken into consideration during the selection process. Towards the improvement of the fruit quality in apple, novel ad-hoc breeding programmes should be designed and supported by the use of markers to combine, through a DNA-guided selection, storability features ensuring a long-term shelf-life and favourable flavour. Several studies have been carried out to shed light

on the genetic regulation of aromatic compounds. The role of AAT in flavour formation in particular has already been described in several species such as strawberry (Aharoni et al., 2000), banana (Beekwilder et al. 2004) and melon (El-Sharkawy et al., 2005). In apple the role of this gene in esters production was also confirmed by several QTL-based studies (Dunemann et al., 2012; Cappellin et al., 2014;Both et al., 2017).

One of the aspects that hampered the set-up of specific breeding programmes designed to improve the aromatic profile of novel cultivars is the lack of devices for fast, not expansive and reliable phenotyping (Cappellin et al., 2014). One of the most used instruments for VOCs phenotyping is a SPME-GC-MS (Solid Phase Micro Extraction – Gas Chromatography – Mass Spectrometry). This device represents a valuable analytical instrument although the phenotyping process remains laborious and time-demanding making this device not suitable for high-throughput screening. To overcome this limitation a PTR-MS (Proton-Transfer Reaction - Mass Spectrometry) has been employed for phenotyping (Lindinger et al., 2005). This instrument is characterized by a high throughput assessment due to the direct injection of VOCs, although its analytical determination of each distinct compound is not such accurate as a standard GC-MS. (Zini et al., 2005). Recently a new generation of PTR-MS was made available (Jordan et al., 2009) implementing a time-of-flight mass analyser (ToF), PTR – ToF – MS. This device combined the time-effectiveness of a PTR-MS with a much higher analytical accuracy (~ 6000 times higher with respect to the PTR-MS). PTR-ToF_MS was already successfully employed in QTL discovery studies in apple (Soukoulis et al., 2013; Cappellin et al., 2014; Farneti et al., 2015a; Farneti, et al., 2015b).

7) Scope and thesis outline

This thesis provides novel insights on the genomic regions and candidate genes underlying traits related to fruit quality that are of economic importance in modern apple breeding.

In Chapter 2 ASSIsT, a novel software for efficient calling and filtering of SNPs from Illumina Infinium arrays is presented. This software is developed to filter SNPs according to their performances and can be employed in the analysis of different population types such as full-sib families (backcross, F_1, F_2 and cross-population) and unrelated individuals. Three-allelic SNPs (caused for example by null alleles) are detected and their genetic call is re-edited taking into account a multi-allelic genetic model.

In Chapter 3 the dynamics of fruit firmness QTL is assessed over post-harvest storage using a pedigree based analysis (PBA) QTL discovery approach with 24 full-sib families using penetrometer data. Phenotypic data came from four different storage periods from harvest to 4 months of cold storage during three consecutive years.

In Chapter 4 a different phenotyping approach was used to dissect and distinguish fruit texture into mechanical and acoustic parameters. Phenotypic data were assessed using a TAXT-AED texture analyser that measures both the mechanical and acoustic component of fruit texture on 6 full-sib families (and their ancestors) and on a germplasm collection. The genomic regions involved in both fruit texture components were detected through two statistical approaches and experimental designs. The first is represented by a PBA analysis, while the second is a genome-wide association study (GWAS).

In Chapter 5 the interplay between fruit texture and aroma profiles was investigated through a GWAS approach on a germplasm collection. The entire apple ‘volatilome’ was measured using a PTR -ToF -MS device coupled with an artificial chewing device, while fruit texture was measured using TAXT-AED texture analyser.

Finally, in Chapter 6 the results from the experimental chapters are integrated and related to the current knowledge of fruit firmness and aroma regulation. Furthermore, use and implications of MAB in apple breeding are discussed and future perspectives are delineated.

References

Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports **27**, 617–631.

Aharoni A, Keizer LCP, Bouwmeester HJ, et al. 2000. Identification of the SAAT Gene Involved in Strawberry Flavor Biogenesis by Use of DNA Microarrays. The Plant Cell **12**, 647–661.

Allard A, Bink MCAM, Martinez S, Kelner JJ, Legave JM, Di Guardo M, Di Pierro EA, Laurens F, van de Weg EW, Costes E. 2016. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. Journal of Experimental Botany, erw130.

Amyotte B, Bowen AJ, Banks T, Rajcan I, Somers DJ. 2017. Mapping the sensory perception of apple using descriptive sensory evaluation in a genome wide association study. PlosOne **12**, e0171710.

Andani Z, Jaeger SR, Wakeling I, Macfie HJH. 2001. Mealiness in apples : towards a multilingual consumer vocabulary. Sensory and Nutritive Qualities of Food **66**, 872–879.

Baldi P, Wolters PJ, Komjanc M, Viola R, Velasco R, Salvi S. 2013. Genetic and physical characterisation of the locus controlling columnar habit in apple (Malus x domestica Borkh.). Molecular Breeding **31**, 429–440.

Baldwin EA 2002. Fruit flavor, volatile metabolism and consumer perceptions. CRC Press: Boca Raton, FL.

Bartley I, Hindley SJ. 1980. Alcohol Dehydrogenase of Apple. Journal of Experimental Botany **31**, 449–459.

Beekwilder J. 2004. Functional Characterization of enzymes forming volatile esters from strawberry and banana. Plant Physiology **135**, 1865–1878.

Bianco L, Cestaro A, Sargent DJ, et al. 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh). PLoS ONE **9**, e110377.

Bianco L, Cestaro A, Linsmith G, et al. 2016. Development and validation of the Axiom® Apple480K SNP genotyping array. The Plant Journal,.

Bink MCAM, Van Arendonk JAM. 1999. Detection of quantitative trait loci in outbred populations with incomplete marker data. Genetics **151**, 409–420.

Bink MCAM, Boer MP, Ter Braak CJF, Jansen J, Voorrips RE, Van De Weg WE. 2008. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica **161**, 85–96.

Bink MCAM, Jansen J, Madduri M, et al. 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theoretical and applied genetics. **127**, 1073–90.

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theoretical and Applied Genetics **113**, 206–224.

Both V, Thewes FR, Brackmann A, de Oliveira Anese R, de Freitas Ferreira D, Wagner R. 2017. Effects of dynamic controlled atmosphere by respiratory quotient on some quality parameters and volatile profile of 'Royal Gala' apple after long-term storage. Food Chemistry **215**, 483–492.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics **32**, 314–31.

Bourne M. 2002. Food texture and viscosity: concept and measurement. Academic press.

Boutet G, Alves Carvalho S, Falque M, Peterlongo P, Lhuillier E, Bouchez O, Lavaud C, Pilet-Nayel M-L, Rivière N, Baranger A. 2016. SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. BMC Genomics **17**, 121.

Breseghello F, Sorrells ME. 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics **172**, 1165–1177.

Brummell DA and Harpster M. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. In: Carpita, N.C. and Campbell, M. and Tierney M, ed. Plant Cell Walls. Springer Netherlands, 311–340.

Brummell DA 2006. Cell wall disassembly in ripening fruit. Functional Plant Biology **33**, 103–119.

Calenge F, Drouet D, Denancé C, Van De Weg WE, Brisset MN, Paulin JP, Durel CE. 2005. Identification of a major QTL together with several minor additive or epistatic QTLs for resistance to fire blight in apple in two related progenies. Theoretical and Applied Genetics **111**, 128–135.

Cappellin L, Farneti B, Di Guardo M, Busatto N, Khomenko I, Romano A, Velasco R, Costa G, Biasioli F, Costa F. 2014. QTL Analysis coupled with PTR-ToF-MS and candidate gene-based association mapping validate the role of Md-AAT1 as a major gene in the control of flavor in apple fruit. Plant Molecular Biology Reporter **33**, 239–252.

Chagné D, Crowhurst RN, Troggio M, et al. 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS ONE **7**, e31745.

Chagné D, Lin-Wang K, Espley R V., et al. 2013. An ancient duplication of apple MYB Transcription Factors is responsible for novel red fruit-flesh phenotypes. Plant Physiology **161**, 225–239.

Chagné D, Crowhurst RN, Pindo M, et al. 2014. The draft genome sequence of European pear (Pyrus communis L. 'Bartlett'). PLoS ONE **9**, 1–12.

Chevalier M, Lespinasse Y, Renaudin S. 1991. A microscopic study of the different classes of symptoms coded by the Vf gene in apple for resistance to scab (Venturia inaequalis). Plant Pathology **40**, 249–256.

Chevreau E, Lespinasse Y, Gallet M. 1985. Inheritance of pollen enzymes and polyploid origin of apple (Malus x domestica Borkh.). Theoretical and Applied Genetics **71**, 268–277.

Ching A, Jung M, Dolan M, et al. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics **3**, 19.

Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. 2014. The domestication and evolutionary ecology of apples. Trends in Genetics **30**, 57–65.

Cosgrove DJ. 2000. Loosening of plant cell walls by expansins. Nature **407**, 321–326.

Cosgrove DJ. 2001. Wall Structure and Wall Loosening. A Look Backwards and Forwards. Plant physiology **125**, 131-134

Costa F, Stella S, van De Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S. 2005. Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus x domestica Borkh.). Euphytica **141**, 181–190.

Costa F, van De Weg WE, Stella S, Dondini L, Pratesi D, Musacchi S, Sansavini S. 2008. Map position and functional allelic diversity of Md-Exp7, a new putative expansin gene associated with fruit softening in apple (Malus x domestica Borkh.) and pear (Pyrus communis). Tree Genetics and Genomes

4, 575–586.

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, van de Weg WE. 2010. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus x domestica Borkh.). Journal of Experimental botany **61**, 3029–39.

Costa F, Cappellin L, Longhi S, et al. 2011. Assessment of apple (Malus×domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biology and Technology **61**, 21–28.

Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F. 2012. Texture dynamics during postharvest cold storage ripening in apple (Malus x domestica Borkh.). Postharvest Biology and Technology **69**, 54–63.

Costa F. 2014. MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. Tree Genetics and Genomes **11**, 1–11.

Costa F. 2016. Mechanical investigation to assess the peel contribution in apple fruit. Postharvest Biology and Technology **111**, 41–47.

Demattè ML, Pojer N, Endrizzi I, Corollaro ML, Betta E, Aprea E, Charles M, Biasioli F, Zampini M, Gasperi F. 2014. Effects of the sound of the bite on apple perceived crispness and hardness. Food Quality and Preference **38**, 58–64.

Deschamps S, Llaca V, May GD. 2012. Genotyping-by-Sequencing in Plants. Biology **1**, 460–483.

Di Guardo M, Tadiello A, Farneti B, Lorenz G, Masuero D, Vrhovsek U, Costa G, Velasco R, Costa F. 2013. A Multidisciplinary Approach Providing New Insight into Fruit Flesh Browning Physiology in Apple (Malus x domestica Borkh.). PLoS ONE **8**, 1–15.

Di Pierro EA, Gianfranceschi L, Di Guardo M, et al. 2016. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Horticulture Research **3**, 16057.

Dimick PS, Hoskin JC, Acree TE. 1983. Review of apple flavor — State of the art*. Critical Reviews in Food Science and Nutrition **18**, 387–409.

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

Dixon J, Hewett EW. 2000. Factors affecting apple aroma/flavour volatile concentration: A Review. New Zealand Journal of Crop and Horticultural Science **28**, 155–173.

Duizer L. 2001. A review of acoustic research for studying the sensory perception of crisp, crunchy and crackly textures. Trends in Food Science and Technology **12**, 17–24.

Dunemann F, Peil A, Urbanietz A, Garcia-Libreros T. 2007. Mapping of the apple powdery mildew resistance gene Pl1 and its genetic association with an NBS-LRR candidate resistance gene. Plant Breeding **126**, 476–481.

Dunemann F, Ulrich D, Malysheva-Otto L, Weber WE, Longhi S, Velasco R, Costa F. 2012. Functional allelic diversity of the apple alcohol acyl-transferase gene MdAAT1 associated with fruit ester volatile contents in apple cultivars. Molecular Breeding **29**, 609–625.

El-Sharkawy I, Manríquez D, Flores FB, Regad F, Bouzayen M, Latché A, Pech J-C. 2005. Functional characterization of a melon Alcohol Acyl-transferase gene family involved in the biosynthesis of Ester volatiles. Identification of the crucial role of a Threonine residue for enzyme activity. Plant Molecular Biology **59**, 345–362.

Espino-Díaz M, Sepúlveda DR, González-Aguilar G, Olivas GI. 2016. Biochemistry of apple aroma: A review. Food Technology and Biotechnology **54**, 375-394.

Evans RC, Campbell CS. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. American Journal of Botany **89**, 1478–1484.

Farneti B. 2014. Tomato quality: from the field to the consumer Interactions between genotype, cultivation and postharvest conditions. PhD thesis, Wageningen University, Wageningen, NL ISBN 978-94-6257-080-1

Farneti B, Busatto N, Khomenko I, Cappellin L, Gutierrez S, Spinelli F, Velasco R, Biasioli F, Costa G, Costa F. 2015a. Untargeted metabolomics investigation of volatile compounds involved in the development of apple superficial scald by PTR-ToF-MS. Metabolomics **11**, 341–349.

Farneti B, Khomenko I, Cappellin L, Ting V, Romano A, Biasioli F, Costa G, Costa F. 2015b. Comprehensive VOC profiling of an apple germplasm collection by PTR-ToF-MS. Metabolomics **11**, 838–850.

Farrokhi J, Darvishzadeh R, Hatami Maleki H, Naseri L. 2013. Evaluation of Iranian native apple (Malus X domestica Borkh.) germplasm using biochemical and morphological characteristics. Agriculturae Conspectus Scientificus **78**.

Folta KM, Gardiner SE. 2009. Genetics and genomics of Rosaceae (Springer, Ed.).

Foolad MR, Arulsekar S, Becerra V, Bliss FA. 1995. A genetic map of Prunus based on an interspecific cross between peach and almond. Theoretical and Applied Genetics **91**, 262–269.

Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G. 2005. Marker assisted selection in crop plants. Plant Cell, Tissue and Organ Culture **82**, 317–342.

Guan Y, Peace C, Rudell D, Verma S, Evans K. 2015. QTLs detected for individual sugars and soluble solids content in apple. Molecular Breeding **35**, 135.

Guitton B, Kelner JJ, Velasco R, Gardiner SE, Chagné D, Costes E. 2012. Genetic control of biennial bearing in apple. Journal of Experimental Botany **63**, 131–149.

Gygax M, Gianfranceschi L, Liebhard R, Kellerhals M, Gessler C, Patocchi A. 2004. Molecular markers linked to the apple scab resistance gene Vbj derived from Malus baccata jackii. Theoretical and Applied Genetics **109**, 1702–1709.

Hampson CR, Quamme HA, Hall JW, MacDonald RA, King MC, Cliff MA. 2000. Sensory evaluation as a selection tool in apple breeding. Euphytica **111**, 79–90.

Harb J, Gapper NE, Giovannoni JJ, Watkins CB. 2012. Molecular analysis of softening and ethylene synthesis and signaling pathways in a non-softening apple cultivar, 'Honeycrisp' and a rapidly softening cultivar, 'McIntosh'. Postharvest Biology and Technology **64**, 94–103.

Harker FR, Gunson FA, Jaeger SR. 2003. The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. Postharvest Biology and Technology **28**, 333–347.

Harris SA, Robinson JP, Juniper BE. 2002. Genetic clues to the origin of the apple. Trends in Genetics **18**, 426–430.

Hemmat M, Brown SK, Weeden NF. 2002. Tagging and Mapping Scab Resistance Genes from R12740-7A Apple. Journal of the American Society for Horticultural Science **127**, 365–370.

Holland D, Larkov O, Bar-Ya'akov I, Bar E, Zax A, Brandeis E, Ravid U, Lewinsohn E. 2005. Developmental and varietal differences in volatile ester formation and Acetyl-CoA: alcohol acetyl transferase activities in apple (Malus x domestica Borkh.) Fruit. Journal of Agricultural and Food

Chemistry **53**, 7198–7203.

Jaillon O, Aury J-M, Noel B, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463–7.

Jordan A, Haidacher S, Hanel G, Hartungen E, Märk L, Seehauser H, Schottkowsky R, Sulzer P, Märk TD. 2009. A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). International Journal of Mass Spectrometry **286**, 122–128.

Jung S, Cestaro A, Troggio M, et al. 2012. Whole genome comparisons of Fragaria, Prunus and Malus reveal different modes of evolution between Rosaceous subfamilies. BMC Genomics **13**, 129.

Kakiuchi N, Moriguochi S, Fukida H, Ichimura N, Kato Y, Banba Y. 1986. Composition of Volatile Compounds of Apple Fruits in Relation to Cultivars. Journal of the Japanese Society for Horticultural Science **55**, 280–289.

Kellerhals M. 2009. Introduction to Apple (Malus × domestica Borkh.). In: Folta KM, Gardiner SE, eds. Genetics and Genomics of Rosaceae. New York, NY: Springer New York, 73–84.

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

Khan F. 2015. Molecular Markers : An Excellent Tool for Genetic Analysis. Journal of Molecular Biomarkers & Diagnosis **6**.

King GJ, Maliepaard C, Lynn JR, et al. 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.

Korban SS, Skirvin RM, others. 1984. Nomenclature of the cultivated apple. HortScience **19**, 177–180.

Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C. 2012. Genomic selection for fruit quality traits in apple (Malus??domestica Borkh.). PLoS ONE **7**, 1–10.

Kumar S, Volz RK, Chagné D, Gardiner S. 2014. Breeding for apple (Malus × domestica Borkh.) fruit quality traits in the genomics era. Genomics of plant genetic resources. Springer, 387–416.

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

Lindinger C, Pollien P, Ali S, Yeretzian C, Blank I, Märk T. 2005. Unambiguous Identification of Volatile Organic Compounds by Proton-Transfer Reaction Mass Spectrometry Coupled with GC/MS. Analytical Chemistry **77**, 4117–4124.

Liu Y, Hoffman NE, Yang SF. 1985. Promotion by Ethylene of the Capability to Convert 1-Aminocyclopropane-1-carboxylic Acid to Ethylene in Preclimacteric Tomato and Cantaloupe Fruits. Plant Physiology **77**, 407–411.

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus x domestica Borkh.). Journal of Experimental Botany **63**, 1107–1121.

López ML, Lavilla T, Recasens I, Riba M, Vendrell M. 1998a. Influence of Different Oxygen and Carbon Dioxide Concentrations during Storage on Production of Volatile Compounds by Starking Delicious Apples. Journal of Agricultural and Food Chemistry **46**, 634–643.

López ML, Lavilla MT, Riba M, Venderell M. 1998b. Comparison of volatile compounds in two seasons in apple: 'Golden Delicious' and 'Granny Smith'. Journal of Food Quality **21**, 155–166.

Maliepaard C, Alston FH, Van Arkel G, et al. 1998. Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers. Theoretical and Applied Genetics **97**, 60–73.

Mauricio R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nature reviews genetics **2**, 370–81.

McCann M, Rose J. 2010. Blueprints for building plant cell walls. Plant physiology **153**, 365.

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

Morgan TH. 1909. What are 'Factors' in Mendelian explanations? Journal of Heredity **os**-**5**, 365–368.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. the Plant Cell **21**, 2194–2202.

Niklas KJ. 1992. Plant biomechanics: an engineering approach to plant form and function. University of Chicago press.

Nybom H, Esselink GD, Werlemark G, Vosman B. 2003. Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog-roses, Rosa L. Sect. Caninae DC. Heredity **92**, 139–150.

Paillard NMM. 1990. Flavour of apples, pears and quinces. Developments in food science.

Palaisa K a, Morgante M, Williams M, Rafalski A. 2003. Applications of linkage desequilibrium and association mapping in crop plants. The Plant Cell **15**, 1795–1806.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genetics **2**, 2074–2093.

Peace C, Bassil N, Main D, et al. 2012. Development and Evaluation of a Genome-Wide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry. PLoS ONE **7**.

Peleman JD, Van Der Voort JR. 2003. Breeding by design. Trends in Plant Science **8**, 330–334.

Pooter HL De, Acker MR Van, Schamp NM. 1986. Aldehyde metabolism and the aroma quality of stored 'Golden Delicious' apples. Phytochemistry **26**, 89–92.

Pozniak CJ, Knox RE, Clarke FR, Clarke JM. 2007. Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. Theoretical and Applied Genetics **114**, 525–537.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**, 945–959.

Rafalski JA. 2010. Association genetics in crop improvement. Current Opinion in Plant Biology **13**, 174–180.

Reeve RM. 1953. Histological Investigation of Texture in Apples. Journal of Food Science **18**, 604–617.

Reeve RM. 1970. Relationships of histological structure to texture of fresh and processed fruits and vegetables. Journal of Texture Studies **1**, 247–284.

Richard I, Beckmann JS. 1995. How neutral are synonymous codon mutations? Nature Genetics **10**, 259.

Rose J, Hadfield K, Labavitch J, Bennett A. 1998. Temporal sequence of cell wall disassembly in

rapidly ripening melon fruit. Plant physiology **117**, 345–61.

Rosyara UR, Bink MC a M, van de Weg E, Zhang G, Wang D, Sebolt A, Dirlewanger E, Quero-Garcia J, Schuster M, Iezzoni AF. 2013. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. Molecular Breeding **32**, 875–887.

Rowan DD, Lane HP, Allen JM, Fielder S, Hunt MB. 1996. Biosynthesis of 2-Methylbutyl, 2-Methyl-2-butenyl, and 2-Methylbutanoate esters in 'Red Delicious' and 'Granny Smith' apples using deuterium-labeled substrates. Journal of Agricultural and Food Chemistry **44**, 3276–3285.

Sansavini S, Donati F, Costa F, Tartarini S. 2004. Advances in apple breeding for enhanced fruit quality and resistance to biotic stresses: new varieties for the european market. Journal of Fruit and Ornamental Plant Research **12**, 13–52.

Sato S, Tabata S, Hirakawa H, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**, 635–641.

Schauer N, Semel Y, Roessner U, et al. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology **24**, 447–454.

Seglias NP, Gessler C. 1997. Genetics of apple powdery mildew resistance from Malus zumi (P/2). IOBC WPRS Bulletin **20**, 195–208.

Selkoe KA, Toonen RJ. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters **9**, 615–629.

Seymour GB, Manning K, Eriksson EM, Popovich AH, King GJ. 2002. Genetic identification and genomic organization of factors affecting fruit texture. Journal of Experimental botany **53**, 2065–2071.

Shastry BS. 2002. SNP alleles in human disease and evolution. Journal of human genetics **47**, 561–566.

Shulaev V, Korban SS, Sosinski B, et al. 2008. Multiple Models for Rosaceae Genomics. Plant Physiology **147**, 985–1003.

Shulaev V, Sargent DJ, Crowhurst RN, et al. 2011. The genome of woodland strawberry (Fragaria vesca). Nature Genetics **43**, 109–116.

Sitrit Y, Bennett AB. 1998. Regulation of Tomato Fruit Polygalacturonase mRNA Accumulation by Ethylene: A Re-Examination. Plant Physiology **116**, 1145–1150.

Soukoulis C, Cappellin L, Aprea E, Costa F, Viola R, Märk TD, Gasperi F, Biasioli F. 2013. PTR-ToF-MS, A Novel, Rapid, High Sensitivity and Non-Invasive Tool to Monitor Volatile Compound Release During Fruit Post-Harvest Storage: The Case Study of Apple Ripening. Food and Bioprocess Technology **6**, 2831–2843.

Stauct G. 1989. The species of Fragaria, their taxonomy and geographical distribution. Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, 23–34.

Sunyaev S, Ramensky V, Bork P. 2000. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends in Genetics **16**, 15–17.

Tartarini S, Gianfranceschi L, Sansavini S, Gessler C. 1999. Development of reliable PCR markers for the selection of the Vf gene conferring scab resistance in apple. Plant Breeding **118**, 183–186.

Tatsuki M, Hayama H, Nakamura Y. 2009. Apple ethylene receptor protein concentrations are affected by ethylene, and differ in cultivars that have different storage life. Planta **230**, 407–417.

Toivonen PMA, Brummell DA. 2008. Biochemical bases of appearance and texture changes in fresh-cut fruit and vegetables. Postharvest Biology and Technology **48**, 1–14.

van de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM. 2004. Pedigree genotyping: a new pedigree-based approach of QTL identification and allele minining. Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, 45–50.

VanBuren R, Bryant D, Bushakra JM, et al. 2016. The genome of black raspberry (Rubus occidentalis). Plant Journal **87**, 535–547.

Varela P, Salvador A, Hernando I, Pérez-Munuera I, Quiles A, Ángeles Lluch MA, Fiszman SM. 2007. Eating quality of 'Flor de Invierno' pears: Chemical and structural aspects. International Journal of Food Science and Technology **42**, 1052–1058.

Velasco R, Zharkikh a, Affourtit J. 2010. The genome of the domesticated apple (Malus x domestica Borkh.). Nature genetics **42**, 833–839.

Verde I, Abbott AG, Scalabrin S, et al. 2013. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nature genetics **45**, 487–94.

Vilanova S, Romero C, Abbott AG, Ll??cer G, Badenes ML. 2003. An apricot (Prunus armeniaca L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits. Theoretical and Applied Genetics **107**, 239–247.

Vincent JFV. 1998. The quantification of crispness. Journal of the Science of Food and Agriculture **78**, 162–168.

Visser T, Verhaegh JJ. 1978. Inheritance and selection of some fruit characters of apple. II. The relation between leaf and fruit pH as a basis for preselection. Euphytica **27**, 761–765.

Vos P, Hogers R, M B, Reijans M, T L. 1995. A new technique for DNA fingerprinting. Nucleic Acids Research **44**, 388–396.

Wakasa Y, Kudo H, Ishikawa R, Akada S, Senda M, Niizeki M, Harada T. 2006. Low expression of an endopolygalacturonase gene in apple fruit with long-term storage potential. Postharvest Biology and Technology **39**, 193–198.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey S V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research **18**, 6531–6535.

Williams JGK, Hanafey MK, Rafalski JA, Tingey S V. 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods in enzymology **218**, 704–740.

Würschum T. 2012. Mapping QTL for agronomic traits in breeding populations. Theoretical and Applied Genetics **125**, 201–210.

Yu J, Pressoir G, Briggs WH, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics **38**, 203–8.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics **178**, 539–551.

Zhu Y, Barritt BH. 2008. Md-ACS1 and Md-ACO1 genotyping of apple (Malus x domestica Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genetics and Genomes **4**, 555–562.

Zini E, Biasioli F, Gasperi F, Mott D, Aprea E, Märk TD, Patocchi A, Gessler C, Komjanc M. 2005. QTL mapping of volatile compounds in ripe apples detected by proton transfer reaction-mass spectrometry. Euphytica **145**, 269–279.

Chapter 2

ASSIsT: an Automatic SNP ScorIng Tool for in- and outbreeding species

Mario Di Guardo[1,2,3$], Diego Micheletti[2,4$,] Luca Bianco[2], Herma J.J. Koehorst-van Putten[1], Sara Longhi[1], Fabrizio Costa[2], Maria J. Aranzana[4], Riccardo Velasco[2], Pere Arús[4], Michela Troggio[2] and Eric W. van de Weg[1]

[1]Plant Breeding, Wageningen University & Research, PO Box 386, 6700 AJ Wageningen, The Netherlands.

[2]Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy

[3] Graduate School Experimental Plant Sciences, Wageningen University, PO Box 386, 6700 AJ Wageningen, The Netherlands

[4]IRTA. Centre de Recerca en Agrigenómica CSIC-IRTA-UAB, Campus UAB-UB – Edifici CRAG, Beellaterra (Cerdanyola del Vallés), 08193 Barcelona, Spain.

[$] The authors have contributed equally to this work

Published in Bioinformatics (2015)
31 (23): 3873-3874. doi: 10.1093/bioinformatics/btv446

Abstract

ASSIsT (Automatic SNP ScorIng Tool) is a user-friendly customized pipeline for efficient calling and filtering of SNPs from Illumina Infinium arrays, specifically devised for custom genotyping arrays. Illumina has developed an integrated software for SNP data visualization and inspection called GenomeStudio® (GS). ASSIsT builds on GS derived data and identifies those markers that follow a bi-allelic genetic model and show reliable genotype calls. Moreover, ASSIsT re-edits SNP calls with null alleles or additional SNPs in the probe annealing site. ASSIsT can be employed in the analysis of different population types such as full-sib families and mating schemes used in the plant kingdom (backcross, F1, F2), and unrelated individuals. The final result can be directly exported in the format required by the most common software for genetic mapping and marker-trait association analysis. ASSIsT is developed in Python and runs in Windows and Linux.

Availability: The software, example data sets and tutorials are freely available at http://compbiotoolbox.fmach.it/assist/

Introduction

Advances in whole genome genotyping (WGG) technologies enabled the investigation of several hundred thousand SNP markers simultaneously on a genome-wide scale. To date, Illumina (GoldenGate® and Infinium®) and Affimetrix (Axiom®) are the most widely used array-based genotyping platforms worldwide. Illumina has developed GenomeStudio®, a proprietary software with a graphical user interface (GUI) for SNP data visualization and filtering that enables the selection of high-quality markers showing robust performance across the examined germplasm. However, the actual filtering of such SNPs requires a deep understanding of the performance of SNP markers, genetic segregation patterns, and familiarity with the many tools and parameters in GenomeStudio® (GS). ASSIsT accounts for this by offering a user friendly, automated pipeline that builds on the results of Illumina's GenCall algorithm (Kermani. 2006) as incorporated in GS.

In addition to filtering, ASSIsT also re-edits GS-calls in order to better explore the available information for SNPs showing null alleles or additional SNP clusters due to additional polymorphisms at the probe annealing site. This re-editing enhances correct SNP calling and reduces unnecessary removal of potentially valuable markers

Methods

The analysis and selection of SNPs performed by ASSisT is based on the calls produced by Illumina's GenCall algorithm (Kermani 2006). A two tiers approach that employs a bi-allelic genetic model, and then a tri-allelic model is used to classify SNPs on the basis of their real performance on examined germplasm. The tri- allelic model is used to describe more complex segregation patterns due to null-alleles or alleles with variable signal intensity due to additional SNP, as the bi-allelic genetic model used by GS cannot account for such polymorphisms (Troggio et al. 2013; Gardner et al. 2014; Pikunova et al. 2014; Bassil and Davis et al. 2015). In this case, ASSIsT may re-edit GS-calls by applying de novo filters using the original light intensity data and the segregation patterns in the germplasm.

Results

ASSIsT supports the analyses of different population types, such as full-sib families (e.g. human, livestock, cross pollinating plants), mating schemes common in plants (backcross, F1, F2), and individuals with unknown genetic relationships. ASSIsT's Graphical User Interface (GUI) allows easy parameter setting and provides a visual output of the SNP clustering analysis. The results produced by ASSIsT can be directly exported to the input format of the most widely used software for genetic and marker-trait association analysis (FlexQTL™, GAPIT, JoinMap®,

PLINK, Structure and Tassel). This straightforward integration will improve marker performance in association and QTL mapping studies. ASSIsT is developed in Python (www.python.org). Its source code is released under the GNU General Public Licence (GNU-GPLv3) to allow its integration into bioinformatic pipelines.

ASSIsT requires three input files: a pedigree file in which the parents of each sample are reported and two standard report files from GS ('Final Report' and 'DNA Report'). The two GS reports are standard output of commercial service companies; therefore, ASSIsT does not necessarily require access to GS. A map file with the genetic or physical position of the markers may also be included. This information is mandatory for exporting results in Structure or PLINK formats.

ASSIsT allows pre-selection of the stringency of the filtering procedure by customizing the following parameters: (i) Proportion of missing data, (ii) Call Rate threshold, (iii) Segregation distortion (X^2 P-value), (iv) Frequency of not allowed genotypes (structured germplasm) and (v) Minor Allele Frequency.

The first step of the filtering analysis is a quality check of the individuals; samples with a high proportion of unexpected marker genotypes due to outcrossing, different ploidy levels, and DNA admixture, among other causes, are considered deviating germplasm and further excluded from the analysis. Samples with poor DNA quality (call rate significantly lower than the average of the dataset) will not be considered in the analysis either. All discarded samples are listed in the 'summary' output file.

Only "robust" markers, (i.e., those showing a clear cluster separation and few No Calls) are allowed through the initial filtering. These markers can show two (one homozygous and one heterozygous) or three clusters (two homozygous and one heterozygous). For some markers, the AB cluster might result in two distinct sub-clusters, due to additional SNPs at the probe site, which may lead to differential hybridization efficiency and to distinct classes of signal intensity within a marker allele. The variation in signal intensity, generally ignored by GS, is considered by ASSisT instead. For instance, a cross between two heterozygous parents generates three genotype clusters at a single locus (e.g., CT×CT produces ¼CC + ½CT + ¼TT). When one allele (let us say T) shows two distinct intensity classes, it may be interpreted as CT×Ct, which gives ¼CC + ¼CT + ¼Ct + ¼Tt. The discernment between the two heterozygous classes (CT and Ct) makes this marker fully informative in inheritance studies, where as 'classical' heterozygotes are not informative in the generation of genetic linkage maps as it is not possible to determine the parental origin of the alleles. Additional SNPs in the probe, as well as INDELS (Pikunova et al. 2014), may also give rise to null alleles, due to the lack of signal in one of the DNA templates, which results in additional clusters. GS cannot currently account for this

scenario; thus, informative markers are lost. Conversely, ASSIsT succeeds in the analysis of the majority of such markers (A0×A0, A0×00 and A0×B0), allowing a more efficient marker calling.

All the above-mentioned SNP classes are suitable for the generation of genetic linkage maps or for marker-trait association studies. Discarded markers are grouped according to their performance considering absence or severe distortion in segregation, presence of not allowed genotypes in segregating families, and number of No Calls.

ASSIsT has been used to analyze SNP markers of several bi-parental full-sib families and germplasm of apple (Bianco et al. 2014), peach, melon and grape. For each family, approximately 99% of the “approved” (those that passed the filtering procedure) SNPs showed to have high quality data as they integrated smoothly in the generation of high-quality genetic linkage maps. The remaining 1% presented several types of issues, largely related to the presence of paralog loci where the AB cluster was too close or even merged to one of the two homozygous clusters.

ASSIsT thus proved to be an effective tool for genotyping studies as it allows to easily filter informative and well performing SNP and to recover potentially useful SNPs from indels or regions of high sequence divergence, feeding them directly to the most common downstream analysis tools through its easy interface.

Acknowledgements

Funding: This work has been co-funded by the EU seventh Framework Programme by the FruitBreedomics project N°. 265582: Integrated Approach for increasing breeding efficiency in fruit tree crops (www.FruitBreedomics.com). The views expressed in this work are the sole responsibility of the authors and do not necessary reflect the views of the European Commission.

References

Bassil, NV, Davis TM, Zhang H, et al. 2015. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. BMC genomics, **16**, 155.

Bianco L, Cestaro A, Sargent DJ et al. 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (Malus × domestica Borkh.). PLoS One, **9**, e110377.

Gardner KM, Brown P, Cooke TF, Cann S, Costa F, Bustamante, C, Velasco R, Troggio M, Myles, S. 2014. Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3: Genes| Genomes| Genetics, **4**, 1681-1687.

Kermani BG. 2006 Artificial intelligence and global normalization methods for genotyping. US Patent 20060224529.

Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Visser RGF, van de Weg, E. 2014. 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree genetics & genomes, **10**, 261-271.

Troggio M, Šurbanovski N, Bianco L et al. 2013. Evaluation of SNP data from the Malus infinium array identifies challenges for genetic analysis of complex genomes of polyploid origin. PLoS One, **8**, e67407.11

Appendix

ASSIsT Reference manual

Di Guardo M, Micheletti D, Bianco L, Koehorst-van Putten HJJ,
Longhi S, Costa F, Aranzana MJ, Velasco R,
Arús P, Troggio M, van de Weg WE

Version 1.02
April 28, 2017

Contents

ASSIsT is a tool for efficient filtering of Illumina Infinium/BeadExpress based SNP markers. This software can analyse different types of experimental populations: Cross-pollinated (CP – F_1), Back Cross (BC), F_2 and collections of unrelated individuals (Germplasm). It is possible to export the filtered data in several formats according to the most widely used software for marker-trait association analysis.

1 Getting started

ASSIsT is written in Python; therefore, it can run virtually in any platform with python installed.

1.1 Availability

Source code and Windows executables (built using pyinstaller) are available for download at:

- http://compbiotoolbox.fmach.it/assist
- http://bioinformatics.tecnoparco.org/fruitbreedomics/assist-tool

When using ASSIsT, please cite: Di Guardo and Micheletti et al. 2015, referenced as: Di Guardo M, Micheletti D, Bianco L, Koehorst-van Putten HJJ, Longhi S, Costa F, Aranzana MJ, Velasco R, Arús P, Troggio M, van de Weg EW (2015) ASSIsT: An Automatic SNP ScorIng Tool for in- and out-breeding species. Bioinformatics, DOI: 10.1093/bioinformatics/btv446

1.2 What's new in version 1.02

- Fixed the shift of two individuals in the gtypes.csv output file with Germplasm population type.
- Added the ability to deal with crosses derived from self-pollination.

1.3 Installing ASSIsT

The Windows executables is distributed as a zip archive. It not necessary to install ASSIsT, just extract the ASSIsT_Windows.xx.zip archive (xx is the version number).

The source code is a collection of Python scripts. They can be executed from any operating system with Python 2.7 installed. The following additional Python modules are needed to run the software:

- PyQt4 (v.4.8 or higher)

- NumPy (v.1.8 or higher)
- matplotlib (v.1.3 or higher)
- SciPy (v.0.14 or higher)

1.4 Running ASSIsT

In Windows, double click on ASSIsT.exe to start ASSIsT. To run ASSIsT from the source code, execute ASSIsT.py from a command line shell.

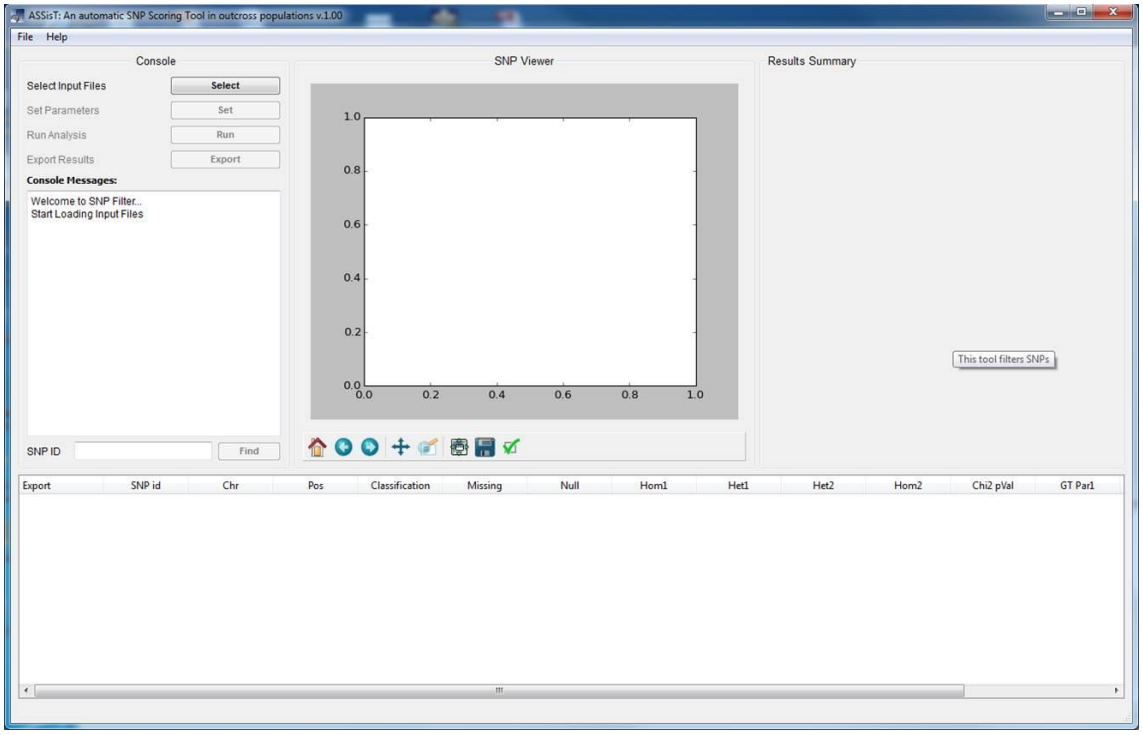


Figure 1: ASSIsT layout.

The analysis starts by loading the four input files (GenomeStudio® Final Report, Genome Studio® DNA Report, pedigree, and map file) using the Select button. Example data files are available at the previously mentioned web-pages. Example data are provided for Cross Pollinators (CP) and F_2-germplasm (F_2 for inbreeding crops). Please note that the map file is not mandatory. To load the files, click on and select the appropriate input file or enter the file name (with the full path) in the text box.

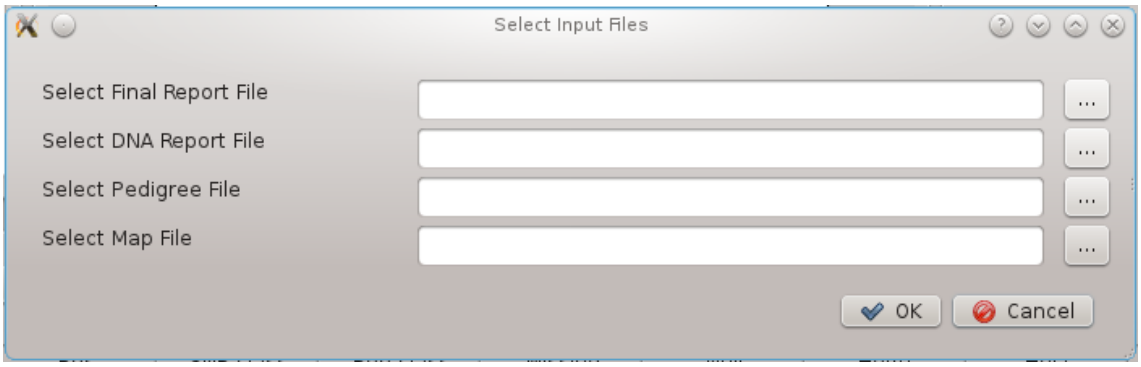


Figure 2: Menu opened by clicking on the Select button

After the Input files are correctly imported, it is necessary to set up the filtering parameters by clicking on the Set button. The first parameter to set is the population type ('CP (F_1)', BC, F_2 or Germplasm) and then the related statistical and germplasm parameters (see below for details).

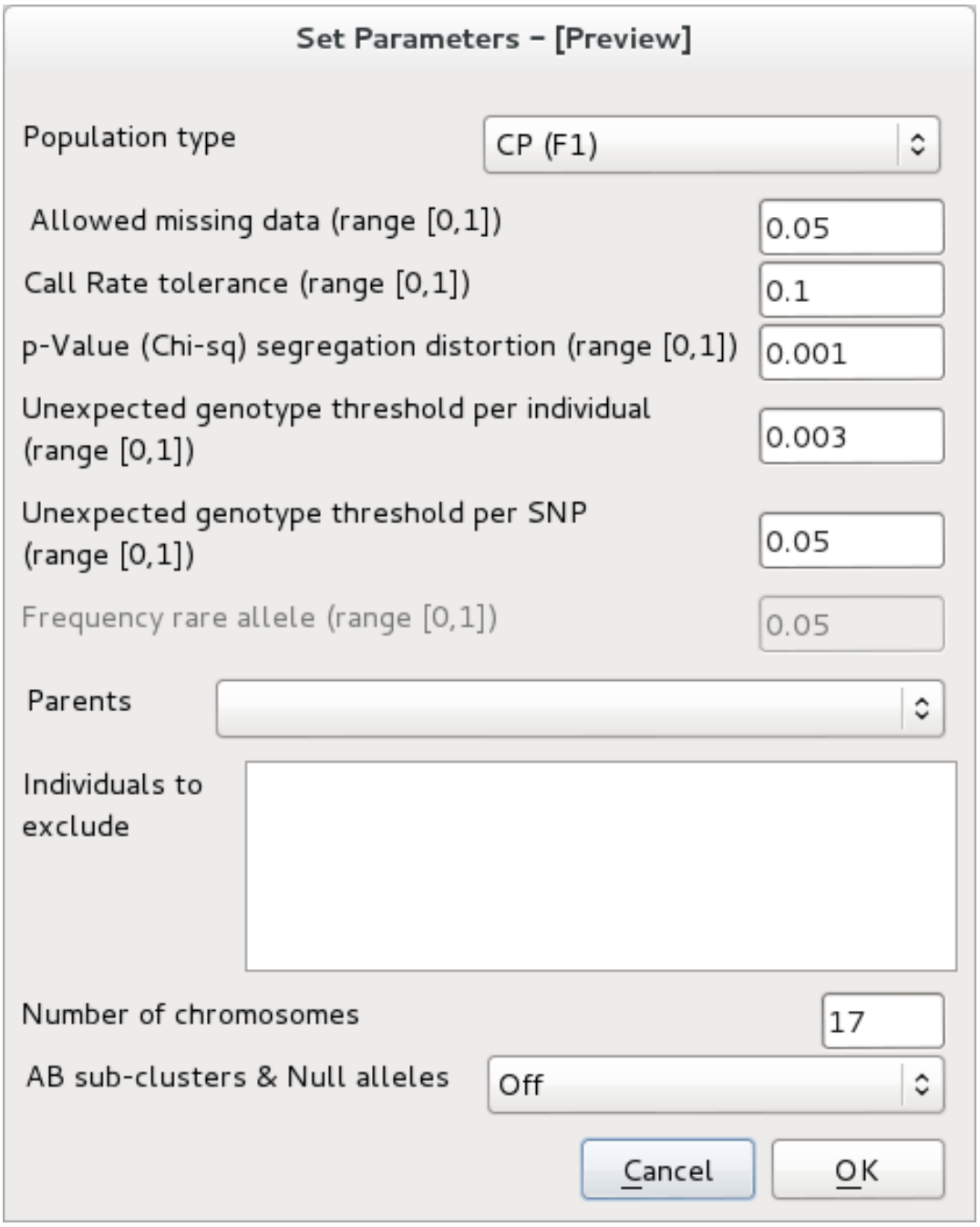


Figure 3: Dialog to select analysis parameter.

When this step is completed, the Run button becomes available, and the analysis can be performed by clicking on it.

At the end of the analysis, it is possible to choose the files to export by clicking Export. Some outputs provide more detailed information on the performance of the filtering analysis itself, e.g., “Summary”, “Custom gtypes”, “Custom SNP information table” and “Custom Mendel error report”.

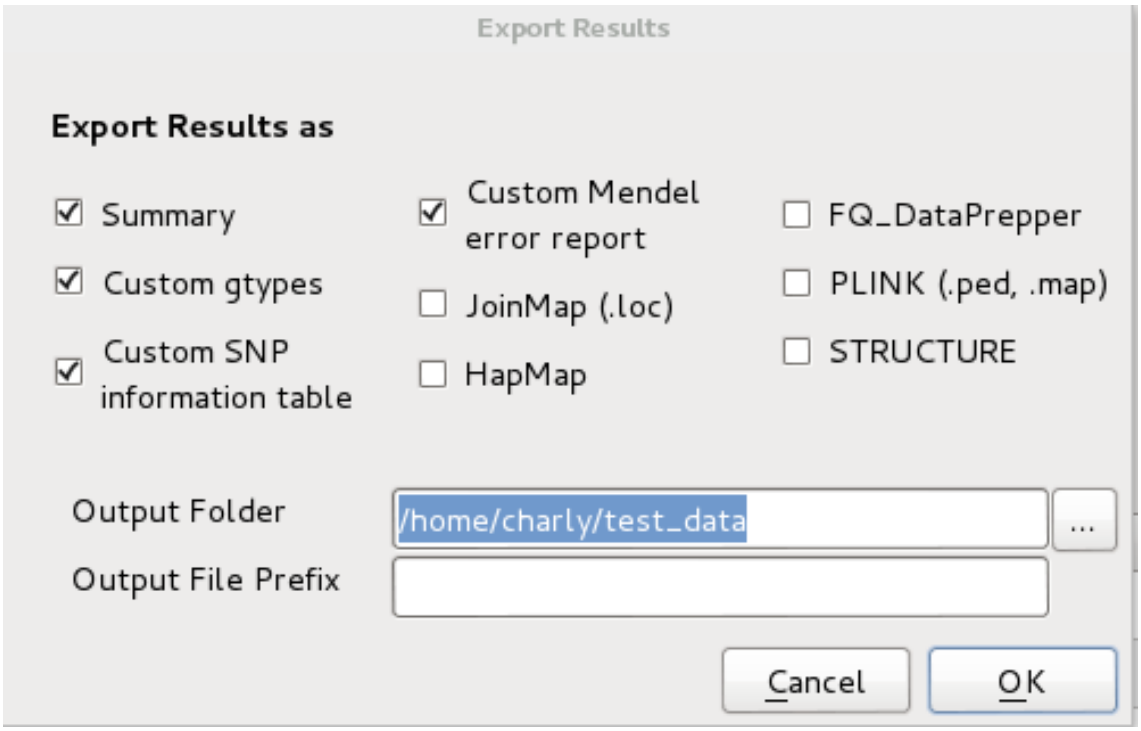


Figure 4: Dialog to select the files to Export.

Additionally, it is possible to export the results in additional formats (JoinMap®, PLINK, HapMap, FlexQTL™DataPrepper and STRUCTURE) that can be used as inputs to third-party programmes.

Note: Through the export section, a customized prefix can be added to the names of the output files.

2 Input files

Sample and marker names have to be consistent in all 4 input files.

Genome Studio Final Report: Using the Report Wizard (Open GenomeStudio® → Analysis →Reports →Report wizard), select Final Report, and press Next. On the following page, use the “redo with the best 10th Percentile GC Score” option, and press Next. If some samples have been excluded from the GenomeStudio® project you need to remove or zero-out your sample in the report. Make your choice according to how you want to account the excluded data and press Next (if no samples have been excluded from GenomeStudio® project this page is not displayed). The format of the Final Report must be set to “Standard” (the other choices are “Matrix” or “3rd Party”). The Final Report should include the following columns (in the specified order):

1. SNP Name
2. Sample ID
3. Allele1 - Top (or Allele1-AB or Allele1-design, depending on the desired type of output)
4. Allele2 - Top (or Allele2-AB or Allele2-design, has to be in line with the choice for allele1)
5. GC Score
6. GT Score
7. Theta
8. R

Select Group by SNP (and not by "sample"). In General Option, select "Tab" as the field delimiter. Press Next, and select the folder in which the file has to be stored, and enter its name. Press Finish.

DNA Report: Using the Genome Studio Report Wizard (see the above section on the GenomeStudio® Final Report) select DNA Report. Press Next once or twice (twice if there are excluded samples in the GenomeStudio® project). Use the redo with the best 10th Percentile GC Score option and press Next once or twice. Finally select SampleID.

Pedigree: The pedigree file is composed of 3 columns: The first contains the list of individuals to analyse, while in the second and third columns, the female and male parents are reported. Be aware that this order (sample, mother, father) is important for some of the output files and that the only compulsory column is the first. In case of populations derived from a self pollination write both in second and third column the name of the selfed parent. Tabulation ("tab") must be used as the field delimiter. The file must also include the following header row:

//SampleID Mother Father

Sample names should not include white spaces (blanks) or special characters (non-ASCII symbols, http://it.wikipedia.org/wiki/ASCII).

Map file (Optional): The map file specifies the physical or genetic coordinates of the Genotyped SNPs. This file must include three columns: the SNPid, the chromosome, and the position. The position can be expressed in base pairs (bp), Megabase (Mb), or centiMorgan (cM). The following file header line is necessary:

//SNPid Chromosome Position

3 Customizable parameters

Population type: Type of population analysed. The possible choices are 'CP (F_1)', 'BC', 'F_2' and 'Germplasm'. The Back-cross (BC) population is analysed as Cross-pollinated (CP) population, as the segregation types are the same (ABxAA or ABxAB and occasionally ABxAC=EFxEG). This tool does not make any assumption on the Grand-parental origin of the alleles. If the populatin type is "CP (F_1)" or "BC" and the two parents name are identical ASSIsT consider the population as derived from a self-pollination and the data of the parent are duplicated to simulate two independent individuals.

Allowed missing data: Frequency of allowed missing data (No Call) by SNP and by individuals. The default value is 0.05. Range [0,1].

Call Rate tolerance: Maximum tolerance for the distance between an individual call rate and the analysed population mean. This parameter is used to exclude individuals (rather than SNP markers) for which too many SNPs could not be called. The default value is 0.1. Range [0,1]

p-Value (Chi-sq) segregation distortion: p-value of the Chi-squared test to test the allelic segregation. This check is based on the Hardy-Weinberg equilibrium test for unstructured germplasm or the expected and observed segregation ratio's when bi-parental populations are analysed. The lower the threshold, the more distortion is allowed. The default value is 0.001. Range [0,1].

Unexpected genotype threshold per individual: Proportion of allowed unexpected genotypes for each individual (Trio's Mendel Errors). This parameter is applied at a very final stage of the filtering process, after having accounted for "AB- sub-clusters and Null alleles" (see below) and is used to exclude individuals that have high probability to be not true to type or that have erroneous pedigree records. The default value is 0.003. Range [0,1].

Unexpected genotype threshold per SNP: Proportion of allowed unexpected genotypes for each SNP (Trio's Mendel Errors). Unexpeced calls will be made missing as long as their proportion does not exceed this threshold. When the threshold is exceeded, this SNP will be excluded. Note that this option is only available when Population Type "CP (F_1)", "BC" or "F_2" is selected. The default value is 0.05. Range [0,1].

Frequency rare allele: Maximum frequency to define an allele as rare. Note that this option is only available when Population Type "Germplasm" is selected. The default value is 0.05. Range [0,1].

Parents: Parents ('CP (F1)', BC) or grandparents (F2) of the analysed experimental population. This parameter is used to specify which segregating family will be analysed. Note that this option is not available when germplasm is selected as Population type.

Individuals to exclude: It is possible to select the individuals to remove prior to the filtering analysis.

Number of chromosomes: Chromosome number of the species in the analysis. Since the tool was developed for apple, the default value is 17.

AB sub-clusters & Null alleles: Find and score markers with the AB cluster split into two sub-clusters and markers that show a null-allele. Note that this option is only available when Population Type 'CP (F_1)' or 'BC' is selected.

4 Output files format

summary.txt: Gives an overview of the assay performances both by markers (number of markers for each class) and by individuals (number of samples analysed and list of individuals that did not pass the quality check: outcross or individuals with poor DNA quality). Moreover, it presents the data and parameter settings that were used for the analyses.

gtypes.csv: Custom file reporting the genotypes of all the successfully genotyped SNPs for each individual in the pedigree file. Each row represents a SNP. The individuals of the analysed population are sorted lexicographically, and the identified outcrosses are reported at the end of each row. The file contains the following information: SNP name (SNP id), Chromosome (Chr) position (Pos), Classification of SNP performance (Classification), Number of No Calls (Missing), number of individuals for each genotype (HomozygousNull, Hom1, Het1, Het2, Hom2), Chi-Squared p-value and the genotypes for each individual analysed. Note that Hom stands for homozygous (AA or BB), Het for heterozygous (AB, Ab, aB, AO, BO) and HomozygousNull for the contemporary presence of a null allele in both chromosomes (OO). O is used to indicate a null allele while a lowercase a or b indicates the presence of an additional SNP at the A or B probe site, respectively.

mendel_error.tsv: For each unexpected genotype, the individual involved is specified together with the genotypes of the two parents and the marker name (only for "CP (F1)", BC or F2, and for the SNP that passed filtering).

snp_table.csv: Reports the segregation and classification information for each SNP (excluded and included). Each line reports the information for a single SNP in the following order: SNP id, genetic position (Chr and Pos), whether the marker has been exported

(Exported), Classification of SNP performance (Classification), number of missing values (Missing), number of individuals for each allelic class (HomNull, Hom1, Het1, Het2, Hom2), Chi-Squared pvalue. The genotype of the parents (GT Par1, GT Par2) is reported only when an experimental population is analysed, while the minor allele frequency (MAF) information is provided only when a Germplasm set is analysed. Het2 represents the second heterozygous state and is present only for the SNPs with AB cluster showing a significant split in two sub-clusters or for the A0 x BO cross.

joinmap.loc: Input file for JoinMap®. This file is created only for the CP population. More details on the file format are available on the JoinMap® user manual beginning on page 46. This file contains only the "approved" markers while the "discarded" markers are left out.

FlexQTL DataPrepper: This output is helpful when preparing an input file for FlexQTL™.

PLINK: Creates the ped and map file that can be used as an input for PLINK (Purcell et al. 2007); this file includes all the SNPs that pass the quality filtering. Details on the file format can be found at pngu.mgh.harvard.edu/ purcell/plink/data.shtml.

HapMap: creates a file including all the SNPs that pass the quality filtering. File format specifications can be found at www.broadinstitute.org.

Structure: Input file for Structure. The file includes two header lines with the marker position and the relative distance between them. Each individual is stored in a single line. The missing data are coded as '-9', while the nucleotides are stored as digits (1=A, 2=C, 3=G, 4=T).

5 SNP classification

A pre-screening of the fully genotyped germplasm is performed to identify poorly performing SNPs and individuals with low DNA quality. In this first phase, all the SNPs showing more than 75% of NoCall in GenomeStudio® are classified as Failed and excluded from further analysis. Additionally, the accessions showing a CallRate lower than the average CallRate minus the "Call Rate tolerance" are also excluded. The remaining SNPs are further classified based on their performances on the accessions from the pedigree file.

Robust: All the successfully genotyped SNPs in which the segregation follows Mendelian rules and the number of NoCall is lower than the maximum allowed in the dataset. In "CP (F_1)" and "BC" populations, the SNPs can be segregated into two or three clusters depending on the parent genotype (AAxAB and ABxAB). In "F2" and "Germplasm", populations, the SNPs show three clusters with a not significant p-value for the Chi-squared test.

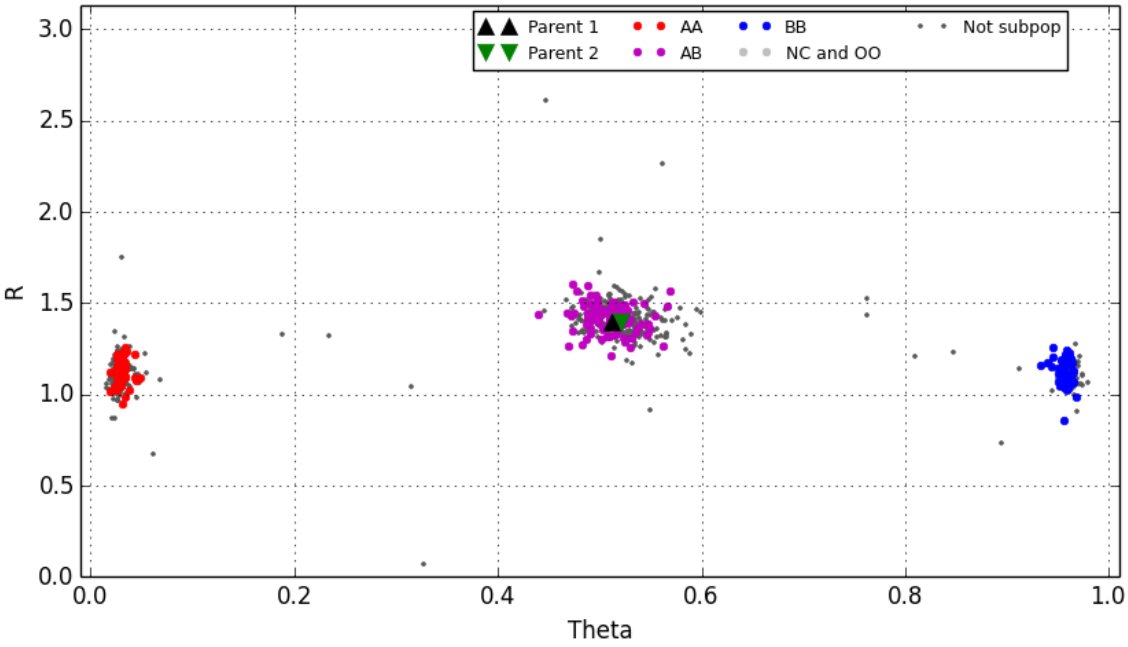


Figure 5: Plot of a Robust SNP.

NullAllele-Failed: This class may appear when 'AB sub-clusters & Null alleles' is set to 'Off'. NullAllele-Failed are the SNPs in which the frequency of the HomozygousNull genotypes (No Call with an intensity of the luminous signal, R, lower than the threshold for null-alleles) is higher than the 'Unexpected genotype threshold per SNP' in 'CP (F_1)', BC or F_2 or higher than the 'Frequency rare allele' in 'Germplasm'. When 'AB sub-clusters & Null alleles' is set to 'On' the NullAllele-Failed are the SNPs in which the frequency of HomozygousNull exceeds the 'Unexpected genotype threshold per SNP', or for null-allele including segregation patterns that ASSIsT does not account for (see last page of the manual), or when the segregation is too skewed (Chi-squared p-value lower than the maximum allowed distortion) to fall in Null_2_Clusters or Null_4_Clusters.

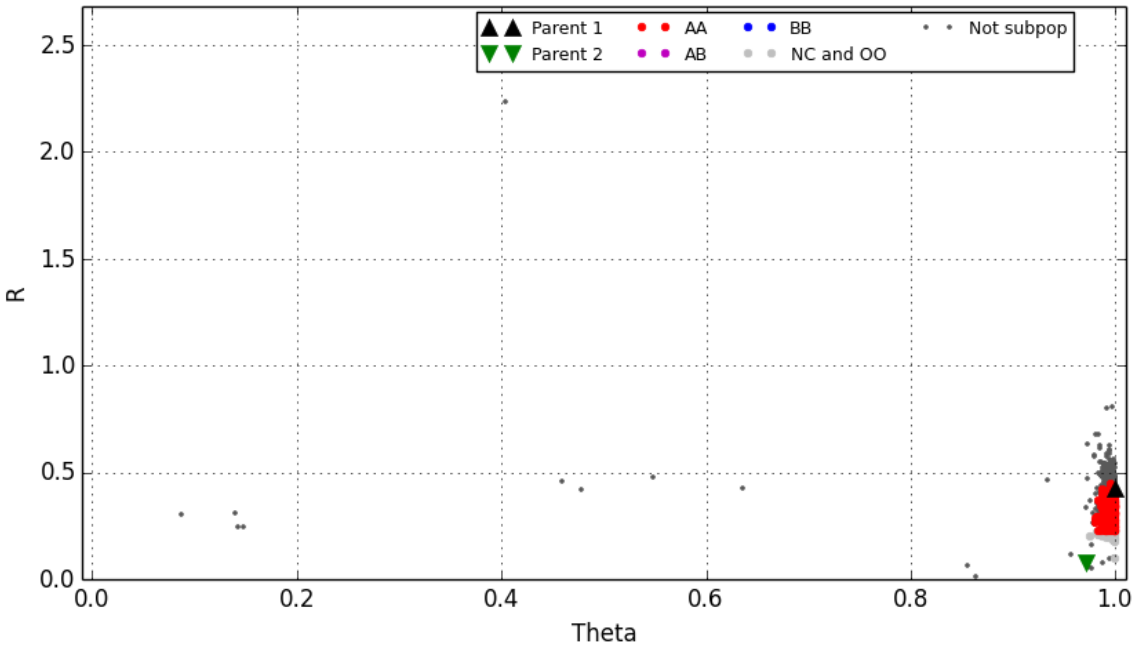


Figure 6: Plot of a NullAllele-Failed SNP.

Null_2_Clusters: SNPs fall in this category if the frequency of homozygous Null genotypes is higher than the 'Unexpected genotype threshold per SNP' and if the frequency of one of the homozygous as well as heterozygous classes are lower than the 'Unexpected genotype threshold per SNP'. The presence of the null allele is coded with 'O'. According to the genotypes of the two parents it is possible to distinguish two different classes of markers: If parents are AO x OO, half of the offspring will be AO and half will be OO. If both parents are AO, one quarter of the offspring will be AA, half will be AO and one quarter will be OO. AA and AO clusters are often partially or totally merged so ASSIsT will score the marker according to the presence/absence of the A allele in the offspring. This re-calling analysis results in two genotype configurations: A- (the second allele is not specified as it is not possible to determine whether the genotype is AA or AO). This class can be present only in 'CP (F_1)' and 'BC' populations when 'AB sub-clusters & Null alleles' is set to 'On'. Note that ASSIST does not account for crosses of type AB x OO or AB x AO.

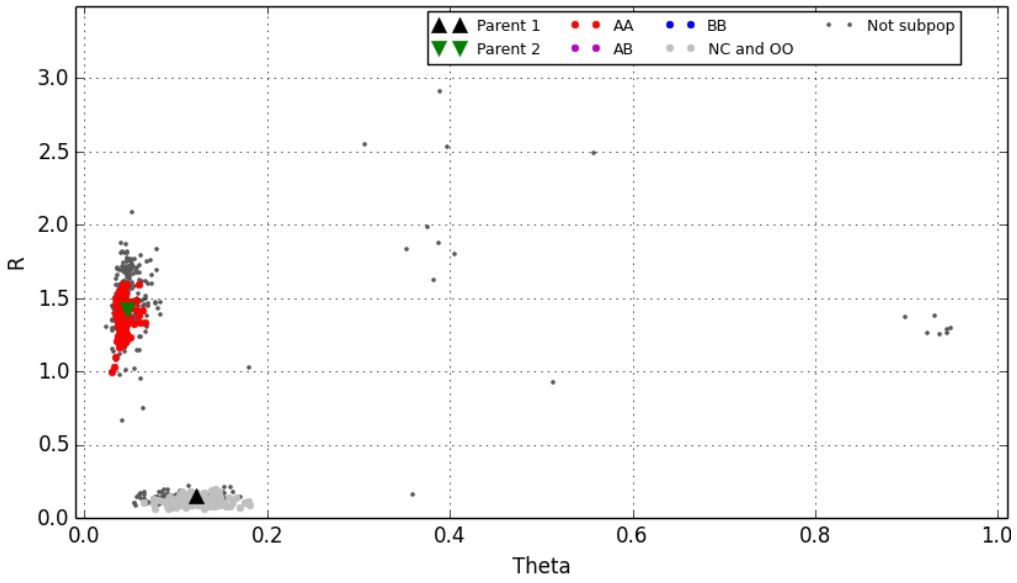


Figure 7: Plot of a Null-one-parent 2 Clusters SNP.

Null_4_Cluster: SNPs fall in this category when the frequencies of AA, AB, BB and HomNull (Both chromosomes with null allele) are higher than the 'Unexpected genotype threshold per SNP' and one of the parents is initially called AA and the other BB. Based on the observed segregation pattern of the family (1/4 AA: 1/4 AB: 1/4 BB: 1/4 OO), these parents are recoded as AO and BO, and their progeny is recoded as AO, AB, BO and OO, respectively. This class can be present only in a 'CP (F_1)' and 'BC' population when 'AB sub-clusters & Null alleles' is set to 'On'.

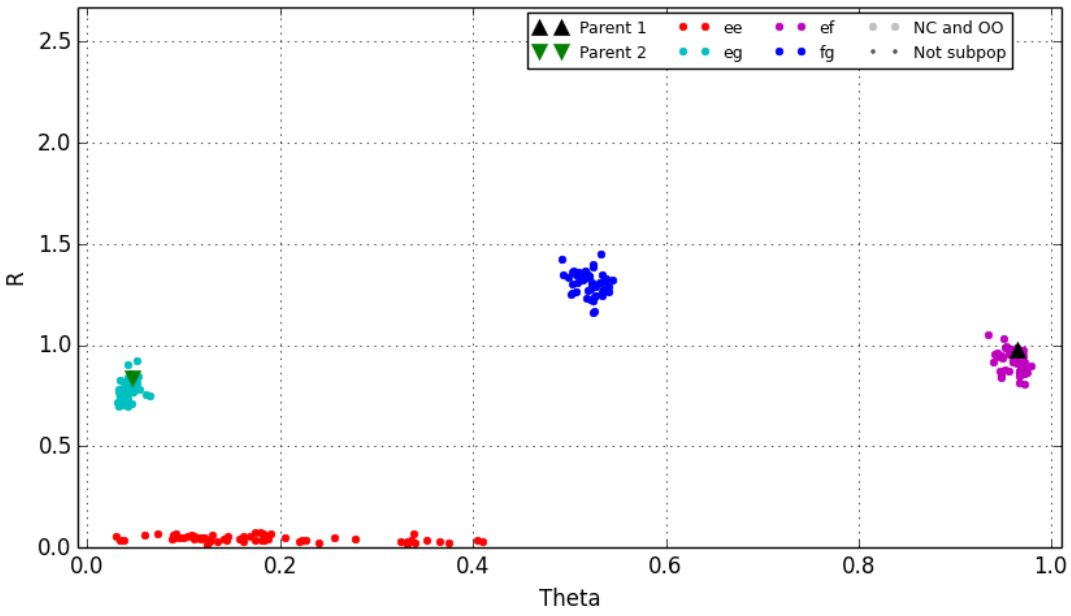


Figure 8: Plot of a Null-two-parents 4 Cluster SNP.

AB_2_sub-clusters: The separation of the AB cluster into two distinct sub-clusters is tested when the 'AB sub-cluster & NullAllele' option is activated. The presence of 2 sub-clusters within the AB genotypes is assessed by looking at the presence of one gap in the derivatives of the distances between the Theta of a contiguous data point. To exclude spurious separation at the lower or higher bound of the AB cluster, please be aware that the derivate is computed after a 10% trim of the extreme values of Theta. The separation is accepted when less than three consecutive values are over 2 * 95th percentile of the derivative distribution. This class can be present only in the 'CP (F_1)' and 'BC' populations.

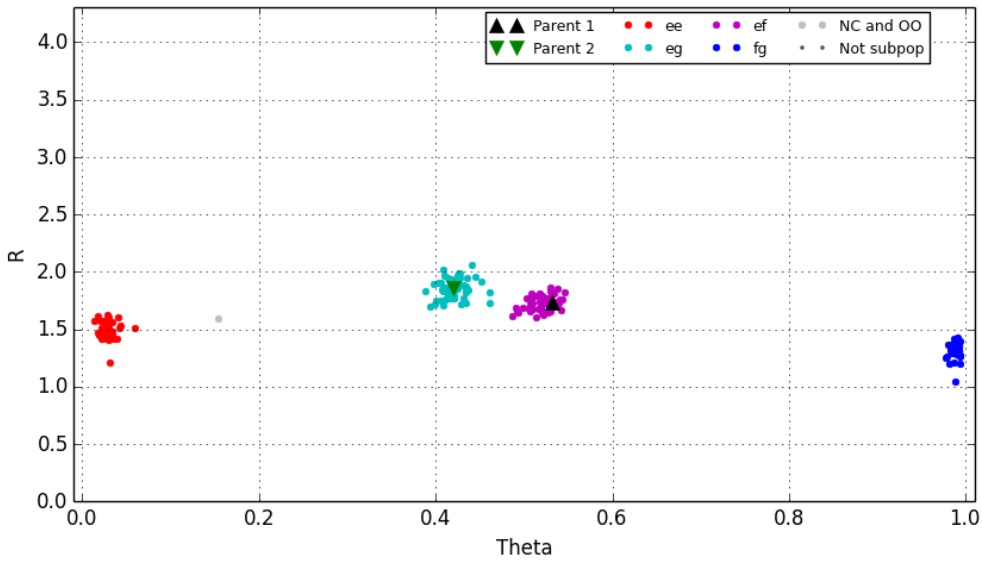


Figure 9: Plot of a AB 2 sub-clusters SNP

OneHomozygRare_HWE: The SNPs are classified as 'OneHomozygRare_HWE' when the frequency of one homozygote cluster is lower than the threshold for the 'Frequency rare allele' but the proportions of the three genotype classes respect the Hardy-Weinberg equilibrium. This class can be present only in the 'Germplasm' population. In this case the 'Frequency rare allele' is actually used as genotype frequency and not as allelic frequency to warn the users about the presence of clusters comprising few individuals. This situation in some cases can hide the presence of an unspecific annealing that causes a shift of part of the homozygote cluster at higher or lower Theta values in correspondence to the heterozygote cluster.

OneHomozygRare_NotHWE: The SNPs are classified as 'OneHomozygRare_ NotHWE' when the frequency of one homozygote cluster is lower than the threshold for the 'Frequency rare allele' and the proportions of the three genotype classes does not follow the Hardy-Weinberg equilibrium. This class can be present only in the 'Germplasm' population. In this case the 'Frequency rare allele' is actually used as genotype frequency and not as allelic

frequency to warn the users about the presence of clusters comprising few individuals. This situation in some cases can hide the presence of an unspecific annealing that causes a shift of part of the homozygote cluster at higher or lower Theta values in correspondence to the heterozygote cluster.

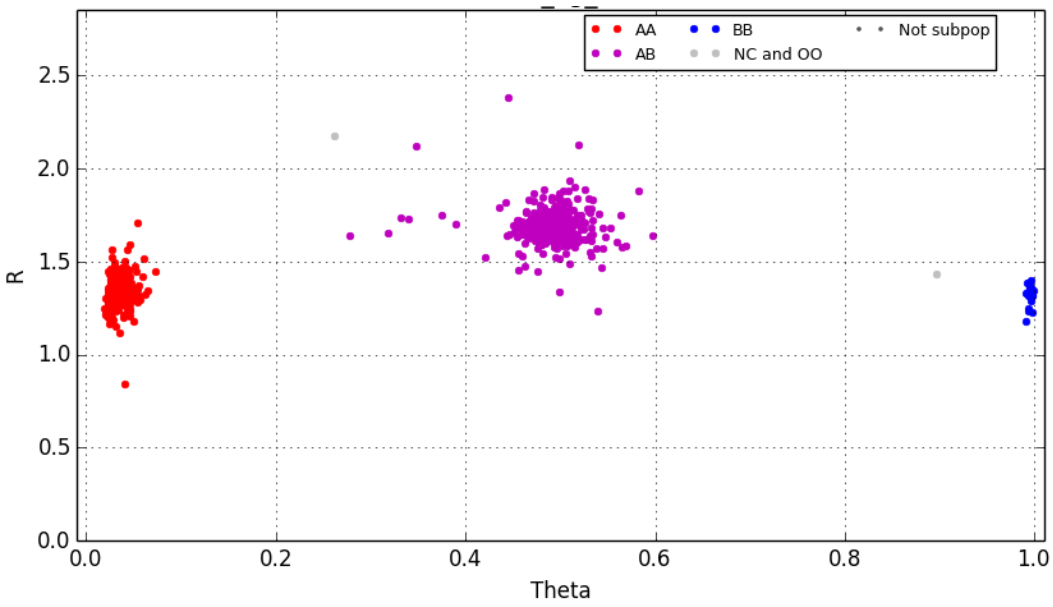


Figure 10: Plot of a OneHomozygRare_(Not)HWE SNP

Monomorphic: The SNPs are classified as False-SNP when a single Genotype class is present and its frequency is higher than the rare allele frequency threshold.

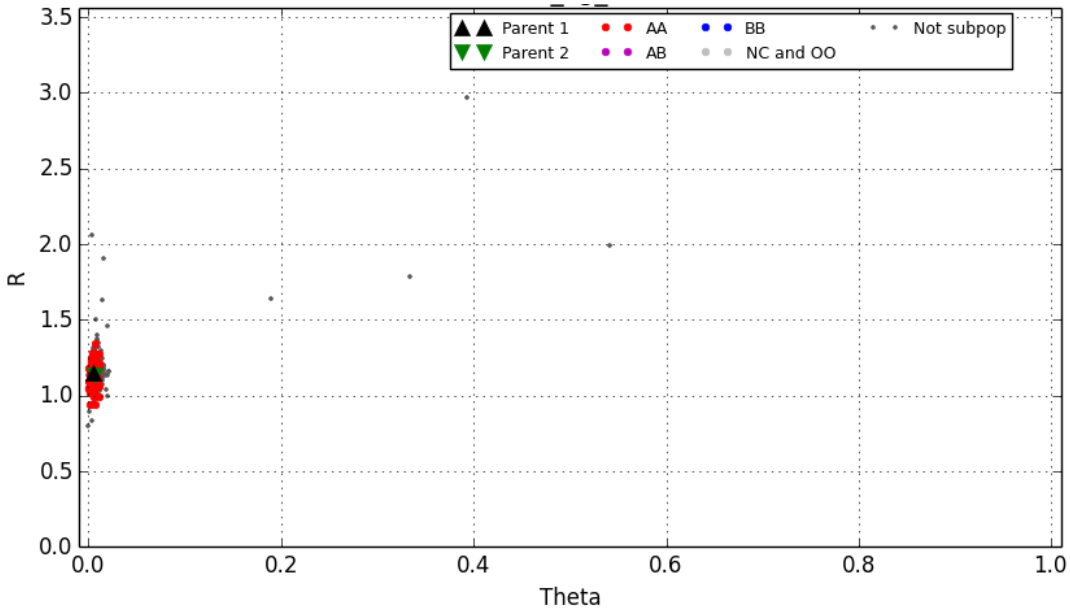


Figure 11: Plot of a Monomorphic SNP.

DistortedAndUnexpSegreg: The segregation in the full-sib families shows a severe skewedness (Chi-squared p-value higher than the value set in the parameters), or one of the genotype classes is missing in a “Germplasm” population, or a genotype class occurs which is not supported by the parental genotypes. This could be due to for instance a AB x AO marker.

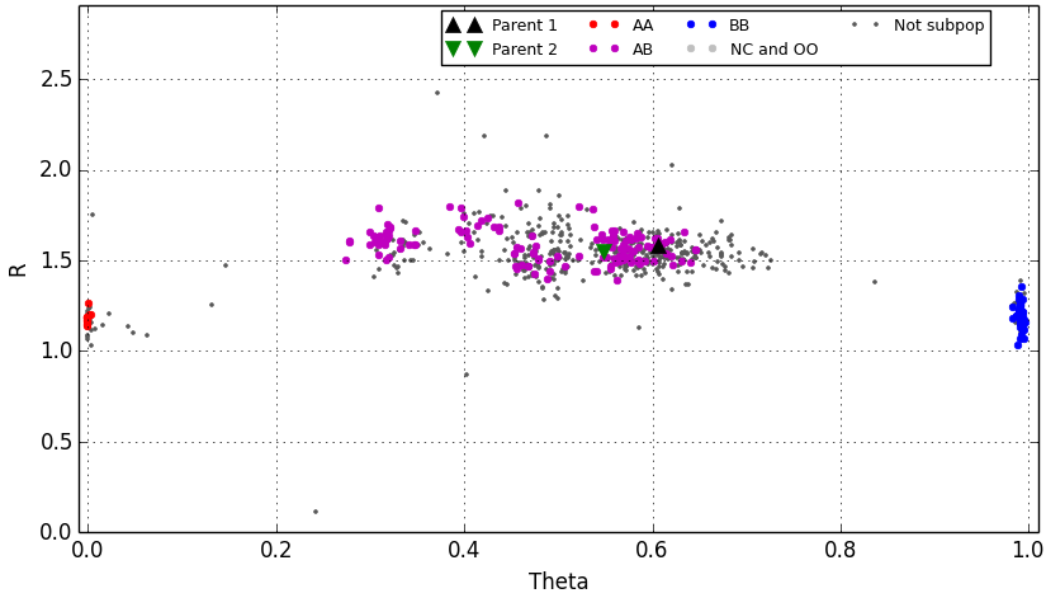


Figure 12: Plot of a Distorted SNP.

OneClassMissing: In an “F_2” population, SNPs fall into this class when one of the three genotypes has a frequency lower than the rare allele frequency threshold.

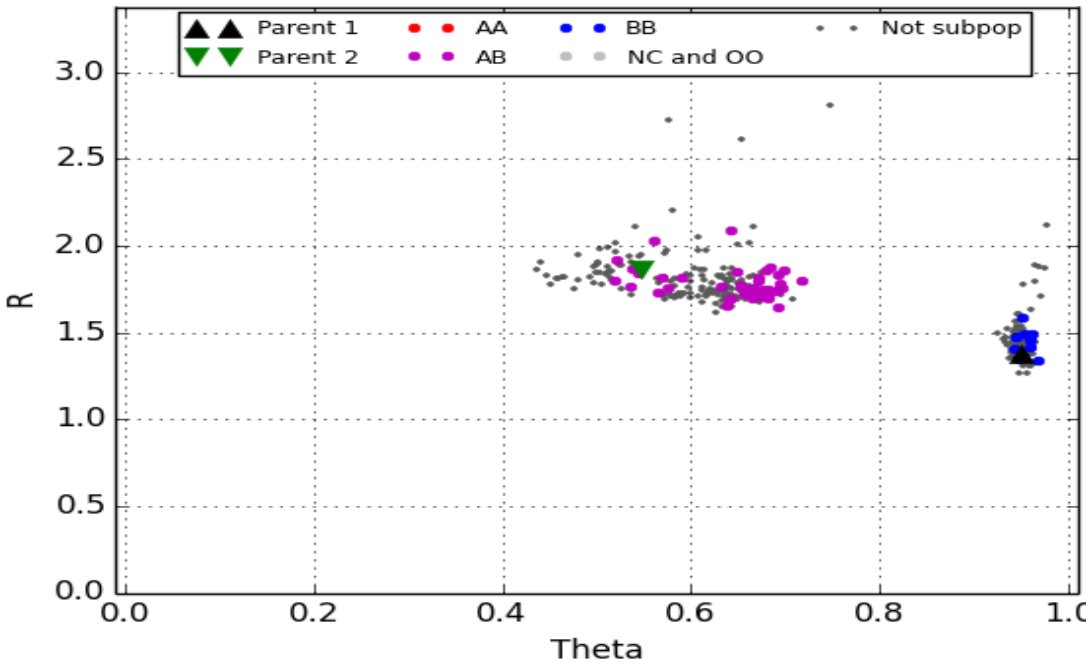


Figure 13: Plot of a oneClassMissing SNP.

ShiftedHomo: In a “Germplasm” population, SNPs are classified as ShiftedHomo when one of the two homozygous classes are absent. This is normally due to an unspecific annealing that causes a shift of the cluster at higher or lower Theta values, depending on the allele that is the concern of the paralogy.

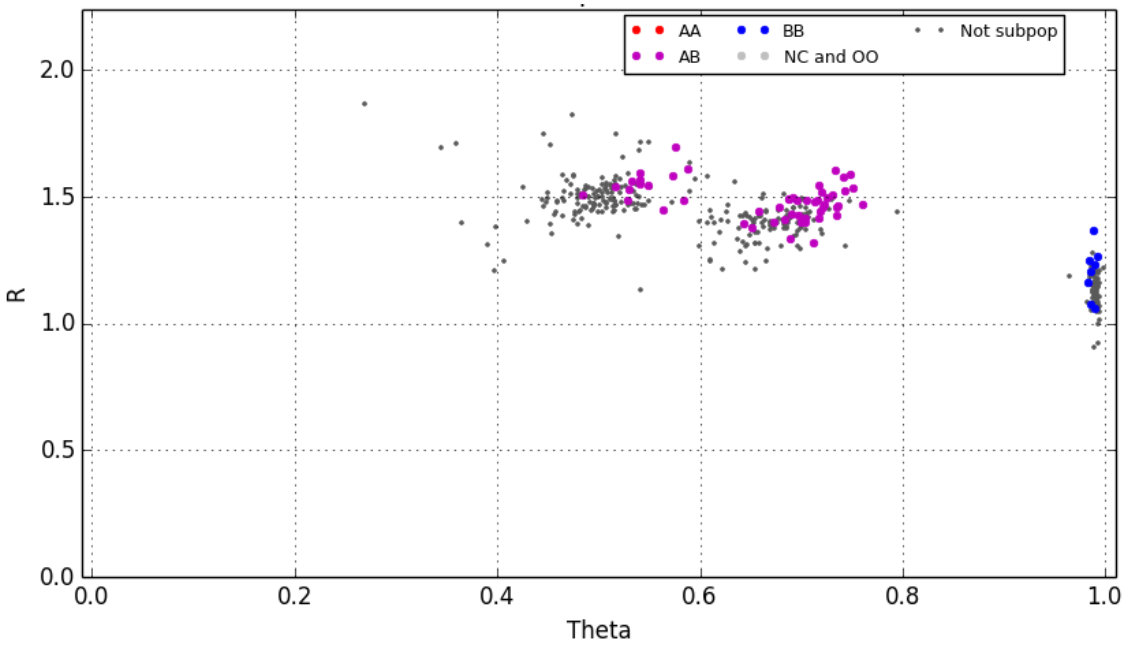


Figure 14: Plot of a shiftedHomo SNP.

Failed: All the SNPs that show a high rate of no-call, that have a mean signal intensity <0.4 or that do not fall in any other class are classified as Failed

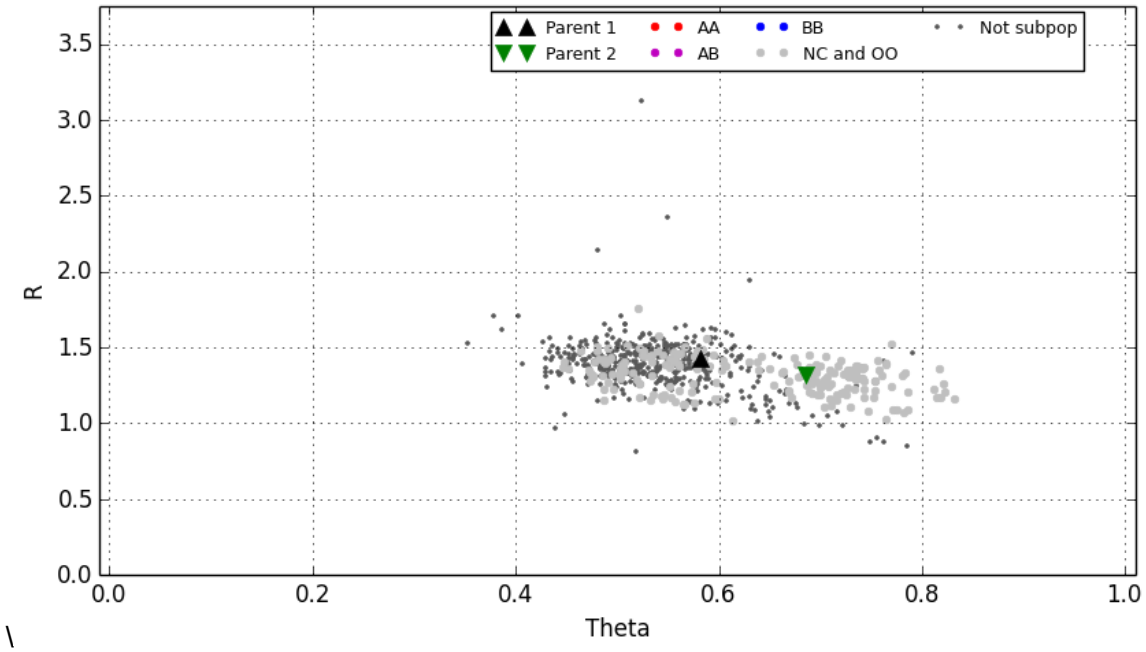
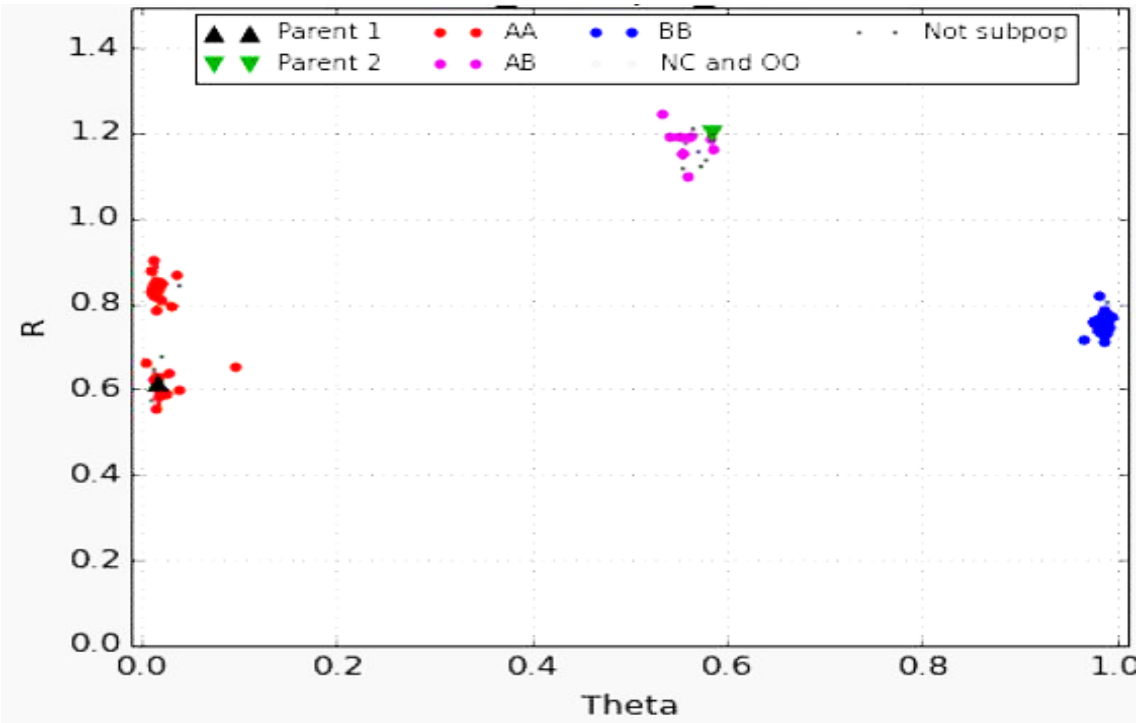


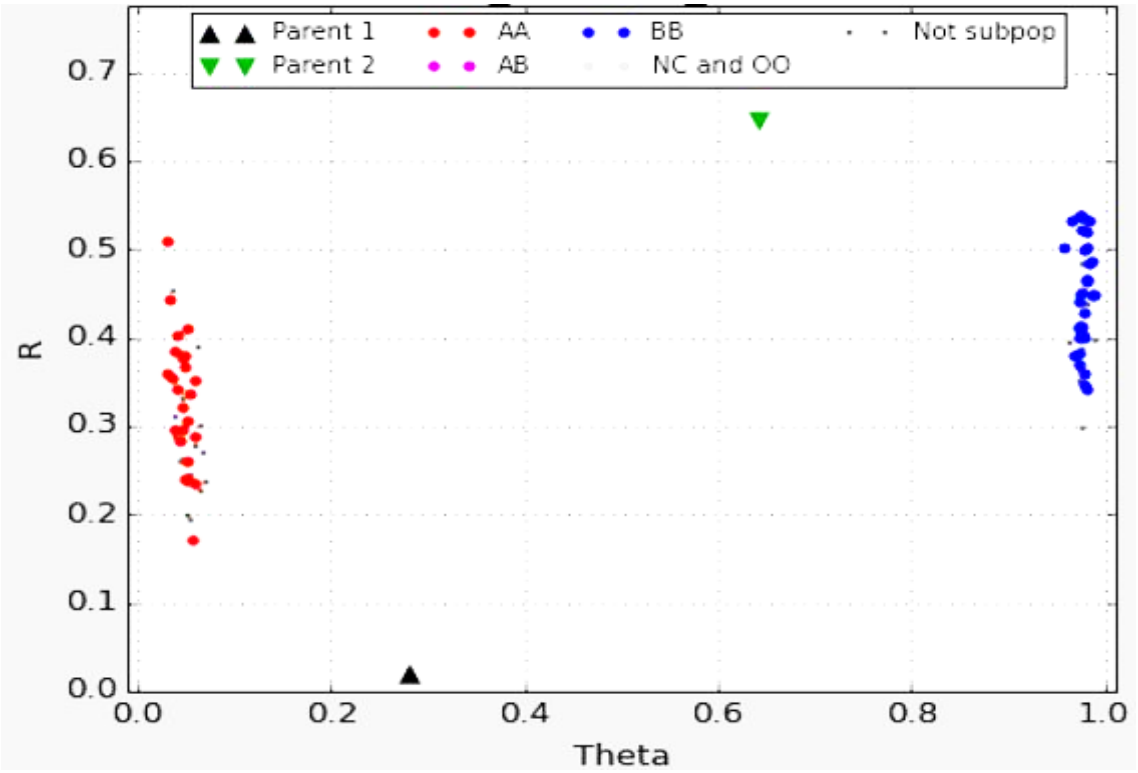
Figure 15: Plot of a Failed SNP

6 Prospects for further development

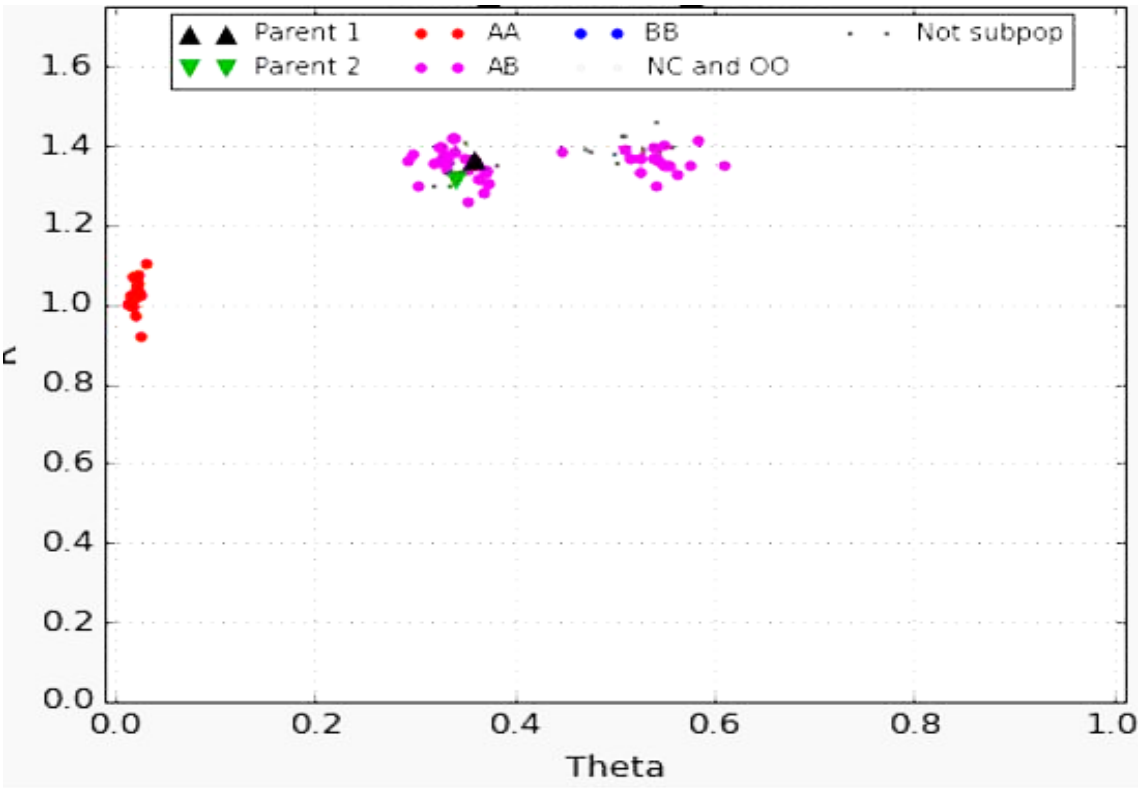
To our knowledge ASSIsT is the first software that identifies and calls null-alleles from SNP markers. The parental SNP-genotype combinations considered are AO x AO, AO x OO and AO x BO. The combinations AB x AO (case (a)) and AB x OO (case (b)) do show equally good prospects based on our results on SNP that were filtered and called using Excel based procedures developed in-house. These were not incorporated into ASSIsT due to time constraints. Currently, neither GenomeStudio® nor ASSIsT supports automated calling of SNP for which one of the clusters for homozygous individuals (AA or BB) is in between x=0.4 and x=0.6, which is true for part of the paralogous SNP. Part of these SNP markers do show three well separated clusters (case (c)), and thus have good prospects for calling through alternative procedures. Another useful extension could be the further classification of excluded markers. Currently, markers with non-allowed genotypes and markers with segregation distortion are both assigned to the class "Distorted and unexpected segregation".



(a) AO x AB →1/4 AO + 1/4 AA + 1/4 AB + 1/4 B0



(b) OO x AB →1/2 AO + 1/2 BO



(c) AB x AB →1/4 AA + 1/2 AB + 1/4 BB

Chapter 3

Unravelling genome-wide QTL patterns for fruit firmness in apple over different storage periods using the multi-family Pedigree Based Analysis (PBA) approach

Mario Di Guardo[1,2], Marco CAM Bink[3], Johannes Jansen[3], Richard GF Visser[4], Fabrizio Costa[1], Eric van de Weg[4].

[1]Fondazione Edmund Mach, via Mach 1, 38010 San Michele all'Adige, Trento (Italy)

[2]Graduate School Experimental Plant Sciences, Wageningen University, PO Box 386, 6700 AJ Wageningen, The Netherlands

[3]Biometris, Wageningen University and Research, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA Wageningen, The Netherlands

[4]Plant Breeding, Wageningen University and Research, Droevendaalsesteeg 1, PO Box 386, 6700 AJ Wageningen, The Netherlands

Abstract

The quality of fruit is represented by a series of physiological and biochemical modifications occurring throughout the entire process of fruit development, until the onset of fruit ripening. Among several phenomena, the cell wall degradation is certainly one of the most evident aspects, finally leading to fruit softening. Although a loss of firmness is necessary to enable the fruit to become more palatable and attractive, an excessive softening also generates severe fruit decay, especially during postharvest storage. To dissect the genetic control of fruit firmness in apple during postharvest storage, the multi-family QTL discovery approach named Pedigree Based Analysis (PBA) was used, employing 24 bi-parental families connected by a common pedigree structure. The association between fruit firmness, assessed over four postharvest treatments, from harvest to four months of storage, and 7112 SNP markers, further assembled into 1113 haploblocks, resulted in the identification of ten QTLs distributed on eight linkage groups. Three QTLs were common to all storage periods (FF-LG10$_b$, FF-LG14 and FF-LG15), five were common for two to three periods (FF-LG1$_a$, FF-LG1$_b$, FF-LG3, FF-LG6 and FF-LG10$_a$) while two QTLs were highly specific to harvest (FF-LG11) or four months of storage (FF-LG16). Candidate genes were identified for some of these QTLs, including MdPG1 for the common and strong FF-LG10$_b$ QTL.

The comparison of the QTL pattern unravelled a QTL dynamic over storage, shedding light on the specific genetic control ongoing during storage and shelf life. The IBD (Identity by Descent) analysis allowed tracing the QTL-allele flow over the pedigree, defining specific breeding signatures for each individual highlighting in particular the effect of HB-10-32, a haplotype related to the MdPG1 gene and coincident with the major QTL for fruit texture. The different impact on the phenotype of two types of markers, SNP and haplotype is shown and their properties in breeding programs towards a more informative and accurate assisted selection of the most favourable fruit quality features are discussed.

Introduction

Change in fruit firmness is the physiological result of important modifications occurring in the primary cell wall and middle lamella (Hadfield and Bennett, 1998; Rose and Bennett, 1999; Brummell, 2006). These structures, made up of cellulosic microfibrilles embedded into a hemicellulosic matrix (mainly xyloglucan), are continuously disassembled by reversible and irreversible degradation processes operated by a myriad of cell wall modifying proteins, encoded by multi-gene families (Rose et al. 1998; Cosgrove 2000; Brummell and Harpster 2001; Powell et al. 2003; Nishiyama et al. 2007; Bennett and Labavitch 2008). Loss of firmness is one of the most important and evident processes taking place during fruit maturation in both climacteric and non-climacteric fruits (Giovannoni, 2001).

In apple, firmness plays a fundamental role in the definition of fruit quality, for two basic reasons. Firstly, firm fruits are more prone to a better and favourable postharvest performance including storability. This aspect is of economic relevance, since storability determines the economic life-time of harvested fruit. On the other hand, a favourable firmness and crispness is nowadays preferred by consumers, whereby soft and mealy apples are generally associated with lower quality (Harker et al., 2003; Costa et al., 2011, 2012). To this end, in the last decades, fruit firmness became one of the most investigated topics in apple, as well as in many other fruit species. To unravel its complex genetic base, several QTL discovery studies have been carried out. QTL mapping is widely accepted as a valuable strategy to unravel the genetic architecture of a trait. Both QTL analysis based on bi-parental populations and association studies on functional markers in cultivar collections enabled the tracing of important candidate genes in the control of fruit firmness through ethylene production and cell-wall degradation, such as Md-ACO1, MdACS1, Md-Exp7 and MdPG1 (Sunako et al. 1999; Harada et al. 2000; Oraguzie et al. 2004; Costa et al. 2005, 2010; Longhi et al. 2012). Diagnostic markers suitable for marker assisted breeding (MAB) have been identified (Harada et al. 2000; Costa et al. 2005, 2010) or ad-hoc developed, such as MdPG1 (Longhi et al. 2013b; Nybom et al. 2013), and used for further examination of their predictive value in candidate-gene based association studies (Zhu and Barrit 2008; Longhi et al. 2013a; Nybom et al. 2008, 2013). Their joint predictive power was 15% of the observed variation in firmness at harvest and 18% in softening rate on a suite of 127 cultivars of various age and origin. Cultivars with the same allelic composition for these four genes could still show large differences in softening (Nybom et al. 2013). This indicated that additional, not yet identified, genes of major effect may exist. Indeed, other firmness QTLs have been reported in single or multi-family QTL discovery and genome wide association studies (GWAS) for which no candidate genes have been proposed yet (Maliepaard et al. 1998; King et al. 2000, Liebhard et al. 2003; Kenis et al. 2008; Bink et al. 2014), and of which some showed up in multiple independent studies. Although these works shed light on the genomic regions

controlling fruit firmness, they provided a static and somehow simplified picture of fruit firmness since just one time point was considered, usually harvest (Maliepaard et al. 1998; King et al. 2000; Liebhard et al. 2003; Kenis et al. 2008) or two months of cold storage (Longhi et al. 2013a; Bink et al. 2014).

Moreover, most of the above-mentioned QTL discovery studies were based on single bi-parental families, which can only discover those loci and alleles that segregate in the two examined parental cultivars, usually a subset of the genetic variability explored in breeding. Moreover, the size of these study populations is usually limited to up to a few hundred individuals, thus putting constraints on the accuracy of QTL positions. These limitations explain the recent interest in the use of multiple families in QTL and GWAS in apple (Kumar et al. 2012, 2013; Bink et al. 2014; Allard et al. 2016), enlarging the number of genes, alleles and genetic backgrounds that may be simultaneously analysed. Recently, a new strategy known as Pedigree Based Analysis (PBA), has been presented as valid approach for the simultaneous analyses of multiple pedigreed families (van de Weg et al., 2004; Bink and van Eeuwijk, 2009). As this approach explores known genetic relationships, it allows the use of families from ongoing breeding programs. Therefore, it is possible to employ already existing germplasm, like in LD studies on unstructured germplasm, which is of particular interest for crops with a long juvenile phase and a long economic life such as apple. Dedicated software has been developed, FlexQTL™ (Bink et al. 2012, 2014, www.flexqtl.nl), which makes use of Bayesian statistics and Identity By Descent (IBD) concept to explore known genetic relationships by linking genomic regions of mapping families and successive generations at the molecular marker level. The PBA approach was developed and implemented within the frameworks of two EU-Projects: HiDRAS (Gianfranceschi and Soglio, 2004) and FruitBreedomics (www.FruitBreedomics.com) and the USDA SCRI project RosBREED (www.rosbreed.org, Iezzoni et al. 2010). QTLs have already been detected with this approach in several fruit species, such as cherry (Rosyara et al., 2013), apple (Bink et al., 2014; Guan et al., 2015; Allard et al., 2016), peach (Fresnedo-Ramírez et al., 2015, 2016), and strawberry (Roach et al., 2016).

Finally, besides the represented genetic diversity and germplasm size, the power of QTL discovery and characterization may be affected by marker density. In previous single family studies, genetic diversity and sample size were usually the limiting factors, whereas in the first PBA study on fruit firmness in apple, marker density put a serious constraint on the detection power, as the ~1300cM genome of over 1347 individuals was represented by just 87 SSR markers (Bink et al., 2014).

In this study, we report QTL results from a PBA analyses on four different post-harvest storage periods using an experimental scheme of 24 full-sib families and a genome-wide high-

density SNP set from a 20K Infinium® SNP array (Bianco et al. 2014). QTLs that are either stable along the different storage periods or trait-specific were detected, shedding light onto the genomic regions controlling changes in fruit firmness.

Materials and Methods

Plant material

In this study, a total of 1216 phenotyped and genotyped individuals, represented by 1049 progenies and 167 progenitors of 24 full sib (FS) families from the HiDRAS project (Gianfranceschi and Soglio 2004) were used. The progenitors included 28 direct parents, 33 founders (of which 2 were also direct parents), and 41 individuals from intermediate generations. To maximize phasing of SNP markers and to allow extrapolation of obtained QTL-genotypes for the direct parents to a wider breeding germplasm, cultivars and four full-sib families (for which no firmness data were available) were added that had at least one of the above progenitors as parent. This included up to five additional first and second generations cultivars and selections from founders. Triploid individuals were a priori excluded from the analysis.

The 24 FS families were generated by crossing 28 cultivars and selections following the mating scheme outlined in Supplementary Table 1. Specific cultivars have been employed in more than one cross, such as ‘Pinova’ and ‘Gala’ which served as parent in five and four different crosses respectively. FS families were retrieved from five different breeding programs ongoing in four EU countries (INRA-France, JKI-Germany, RCL-Italy, RIPF-Poland, SGGW-Poland). The FS families and wider germplasm are mutually connected through a pedigree structure (Suppl. Fig 1).

Phenotypic Data

In this paper, historic phenotypic data from the HiDRAS project (Gianfranceschi and Soglio 2004) were used, the generation of which is described in Kouassi et al. (2009) and Bink et al. (2014). In short, fruit firmness was assessed with a standard digital penetrometer, which measured the maximum force needed to penetrate the cylindrical 1 cm probe into 7 mm of a peeled fruit portion. Per each individual, ten homogeneous apples were assessed on two opposite sites for a total of twenty measurements. Fruit phenotyping was repeated for three consecutive years, from 2003 to 2005, and for each experimental year fruit firmness was assessed at four different storage periods: at harvest (M_1), after 2 months of cold storage (M_2), after 2 weeks of shelf-life following two months of cold storage (M_3), and after 4 months of cold storage (M_4). A reference set of thirty cultivars was planted at each of five locations and used to

harmonize phenotypic data. Each observation was modelled as a linear function of a grand mean, year, location and genotype using GenStat software (GenStat Committee, 2004). The number of seedlings phenotyped, across the entire dataset, progressively decreased from M_1 (1049 at harvest) to M_4 (893 after four months of cold storage; Table 1) due to limitations in the amount of fruit that could be harvested from a single seedling and due to decay during storage. In case of insufficient fruit availability, the initial stages were favoured.

Table 1. Descriptive statistics and QTL details. For each phenotyping date (Trait) the number of individuals (Records), the phenotypic mean (μ_P), variance (σ^2_P) and heritability (h^2) is reported. For each genomic region with posterior QTL intensity exceeding the posterior probability threshold the linkage group (LG), the genetic region of the QTL (QTL region (cM)), the length of the QTL interval (Length), the bin with the highest 2lnBF within the QTL (Mode), the chromosome-wise 2lnBF (Evidence 2lnBF), the marker more strongly associated to the QTL (Haplotype), the QTL probability (Prob) and the QTL additive effect size (AEt1) are also specified.

Trait	Records	μ_P	σ^2_P	h^2	LG	QTL region (cM)	Length	Mode	Evidence [2lnBF]	Haploblock	Prob	AEt1
M_1	1076	8.74	2.4	0.40	1	44 - 50	6	48	30.2	HB-01-29	0.98	1.03
					3	47 - 62	15	57	29.7	HB-03-35b	1.02	0.93
					10	42 - 82	40	49	28.9	HB-10-32	1.18	0.46
					11	0-16	16	8	5.6	HB-11-7b	0.79	0.59
					14	38 - 69	31	60	4.7	HB-14-33b	0.79	0.34
					15	1 - 35	34	2	3.5	HB-15-1	1.12	0.50
M_2	1050	7.65	2.9	0.46	1	28 - 41	13	33	30.2	HB-01-22	0.91	0.74
					3	46 - 58	12	54	13.7	HB-03-33	1.01	0.91
					6	1 -32	31	12	6.3	HB-06-8	0.65	0.55
					10	12 - 31	19	24	7.5	HB-10-17	0.87	0.51
						43 - 52	9	49		HB-10-32	0.87	0.68
					14	38 - 65	27	52	2.7	HB-14-29	0.51	0.32
					15	16 - 34	18	22	28.5	HB-15-8c	0.92	0.59
M_3	956	6.33	2.3	0.40	1	21 - 57	36	33	6.5	HB-01-22	0.94	0.48
					3	49 - 76	27	60	4.6	HB-03-35c	0.78	0.66
					10	19 - 35	16	24	6.9	HB-10-17	0.70	0.42
						35 - 51	16	49		HB-10-32	1.15	0.75
					14	34 - 62	28	58	5.8	HB-14-30	0.93	0.51
					15	1 - 34	33	27	14.5	HB-15-8c	1.12	0.49
M_4	893	6.33	2.5	0.49	1	38 - 62	24	49	28.9	HB-01-29	1.10	0.68
					6	1- 28	27	2	8.8	HB-06-1	1.06	1.15
					10	14 - 41	27	23	3.9	HB-10-16b	0.52	0.39
						47 - 51	4	49		HB-10-32	0.99	0.90
					14	39 - 61	27	53	7.2	HB-14-30	0.95	0.38
					15	1 - 34	33	21		HB-15-8c	1.09	0.46
					16	15 - 43	28	35	7.9	HB-16-14	0.98	0.49

Generation of SNP data

Marker data were generated at the Fondazione Edmund Mach within the EU project FruitBreedomics, employing a 20K Infinium® array (Bianco et al., 2014) and following the protocol described by Chagné et al. (2012) and Antanaviciute et al. (2012). Initial marker calls were obtained through Illumina's genotyping software GenomeStudio®. For each FS family, SNP markers were further filtered and genotype calls were edited through a pre-runner of the ASSIsT software (Di Guardo et al. 2015; Chapter 2 of this thesis).

Genetic linkage map

At the onset of this work, neither genetic linkage maps nor a reference genome sequence were available by which the 20K Infinium® SNP markers could be adequately ordered. A consensus genetic linkage map was constructed based on the bi-parental linkage maps of 21 single full-sib families, which included 18 of our current QTL mapping families complemented with the families 'Jonathan' x 'Prima', DLO-12 (D1980-15-25 x D19773-01-41) and 'Telamon' x 'Breaburn'. The bi-parental maps had been created using JoinMap® 4.1 (Van Ooijen 2011) with the multipoint maximum likelihood mapping algorithm approach for cross pollinators and the Haldane mapping function. The individual families contained from 5570 (I_CC) to 8454 (Jonathan x Prima) SNP markers with an average of 6950 and spanning 1124 ('Gala' x 'Pinova') to 1551 cM (12_K) with an average of 1.300 cM. During the performance of the current study, these same bi-parental families were used for the creation of a multi-family integrated genetic linkage map, more details on the bi-parental families and their maps can be found in Di Pierro et al. (2016).

For the construction of the consensus map, the integrated bi-parental 'Jonathan' x 'Prima' map was used as framework on which markers mapped on the other families were added manually family by family (by decreasing family size) and focal point by focal point according to their relative distances among common SNP markers. Focal points do hereby refer to the design of the 8K and 20K Infinium® SNP arrays (Chagné et al. 2012; Bianco et al. 2014), where SNP markers located in clusters within a region of at most 100 Kbp or 10 Kbp respectively, defined as Focal Point (FP). Markers within a focal point were ordered according to the 'Golden Delicious' genome sequence (Velasco et al. 2010) v2, in some cases this resulted in double recombinations between nearby markers that disappeared with marker re-ordering.

Filtering of SNP markers

Since an integrated QTL analyses across multiple families requires consistent genotyping among families and progenitors, mapped SNP markers were filtered for stable genotyping performance by making use of FlexQTL's reports on inconsistent marker data and double recombination events. Markers with null-alleles, targeting multiple loci and/or showing clustering issues due to duplicated genes/allo-polyploidy or additional polymorphisms at the probe site were finally removed. This resulted in a set of 7112 SNP markers showing a consistent performance across the germplasm employed here. Along this process, pedigree records were simultaneously validated and, where feasible, corrected or completed. Furthermore, marker order of the consensus map was further improved with FlexQTLs' graphical genotyping plots (Young and Tanksley, 1989) using new double recombination events in the wider germplasm of pedigreed cultivars.

SNP Haploblocks

To both reduce computation time and increase marker informativeness, bi-allelic SNP markers were converted in multi-allelic haploblock (HB) markers. SNP markers were phased with FlexQTL™, while allowing distorted segregation and treating singletons as missing values through FlexQTL™ parameters (MSegDelta=1 and DeleteDR=0 respectively). Next, HBs were first sized with VisualFlexQTL such that no recombination occurred among their set of SNPs in the genotyped progenitors. Next, multi-cM HBs were split when the genetic distance between its first and a next SNP would become larger than 1 cM, to decrease the number of mapping progenies that would have recombination within a HB. HBs with many SNP (>28) were split in equal parts to reduce the chance for a HB to harbour SNP calling errors. Finally, a gene specific SNP, Rvi6_42M10SP6_R193, from Malus floribunda 821 for the Rvi6 (Vf) gene for scab resistance, was entered as a single SNP to allow positioning of QTLs relative to this specific gene because the Rvi6-region seems to be a hot-spot for QTL from various traits (unpublished data). Finally, haplotypes were assigned to haploblocks using PediHaplotyper (Voorrips et al. 2016). In the end, 1114 HBs were generated, carrying from 1 to 28 SNPs comprising a total of 7096 SNPs. (Suppl. Table 2), and further used in the QTL-analyses. Genetic positions of our HBs were adjusted to a pre-runner of the final integrated map of Di Pierro et al. (2016) of 1281cM, which was shorter than the reference Jonathan x Prima map (1339 cM) and only slightly longer than the final integrated map (1267 cM) (Di Pierro et al. 2016). HB-characteristics (SNP composition, genetic positions) are given in Supplementary Table 2. A visual presentation of the consensus HB map is given in Suppl. Fig 2.

QTL analysis

Genome-wide QTL detection was performed using a Bayesian approach (Bink et al., 2007, 2012) implemented in FlexQTL™ software (Bink et al., 2014). The relation between the observed phenotype (**y**) and unknown parameters was expressed in the following linear model:

$$y = \mu + Wa + e \quad (1)$$

where (**μ)** is an intercept representing the phenotypic mean, (**W**) is a design matrix of a vector of regressions (**a**) on the QTL covariates and (**e)** is the residual error of the model**.** A major feature of the Bayesian approach is the implicit search among models with different numbers of QTL that best explain the phenotypic trait variation. The number of QTL in the model is therefore a random variable itself, implying that the number of columns of the design matrix **W** is not fixed. The entries of **W** correspond to the QTL genotypes, which are continuously updated conditional on the QTL positions and individuals' phenotypes. Each QTL is assumed as bi-allelic, therefore the QTL genotypes can be identified in three different conformations: 'QQ', 'Qq', 'qq'. In this study, the 'Q' allele is associated to a high phenotypic value, and thus, in case of firmness, to the favourable allele. The analyses were carried out with an additive genetic model, therefore the QTL elements in W are equal to 1 for 'QQ' genotypes, -1 for 'qq' and 0 for 'Qq' (no dominance effect). The modelling operated by FlexQTL is based on the independent assignment of 'Q' and 'q' alleles to founders, therefore the QTL genotypes of the analysed individuals is a priori unknown, implying that the frequency of the 'Q' allele among the founders (f_a) is also a random variable with a uniform prior distribution between 0 and 1. The Bayesian modelling, moreover, assigned normal prior distribution to the vector **a** and **e** (1) with mean 0 and variance σ^2_a and σ^2_e being respectively the per QTL explained variance and the residual variance. The phenotypic mean (**μ)**, the QTL position (**λ**) and the QTL allele frequency (**f_a**) were attributed to a uniform distribution, as no prior assumptions can be taken about the phenotypic value, the position of the QTL and the frequencies of the positive 'Q' allele. The number of the QTL is explained by a Poisson distribution. In this survey, different mean values of the Poisson distribution (N_{QTL}) were tested to assess the robustness of the Bayesian model.

The joint posterior distribution cannot be calculated analytically due to the large number of possible genotype combinations occurring in such a complex pedigree-structured experimental scheme. This limitation is, however, overtaken by the use of Markov chain Monte Carlo (MCMC) simulation, describing the joint posterior distribution as:

$$f(\mu, a, \sigma^2_a, \sigma^2_e, f_a, \lambda \mid y) \quad (2)$$

Inference on the number of QTL

Per chromosome, a pairwise comparison of models with increasing number of QTLs (1 vs. 0, 2 vs. 1 etc.) and based on a Bayes Factor (BF) was carried out to infer the number of QTL. After taking the natural logarithm of the Bayes Factor (2lnBF), which holds a similar scale to frequentist's LOD scores, test statistic values greater than 2, 5 and 10 indicate a positive, strong or decisive evidence of a model vs. the other, respectively. For the inference on the number of QTLs, we considered genomic regions for which the $2\ln BF_{1\text{-}0}$ was higher than 5 (strong to decisive evidence) for at least one storage period, and genomic regions showing at least positive evidence for at least two storage periods (Table 1). This same approach was used for inference of two QTLs on a single linkage group, using $2\ln BF_{2\text{-}1}$ values.

QTL representing markers

In general, the QTL mode was used to select the QTL representing markers. In cases where no clearly dominating peak was present within the QTL interval, we used the median as point of reference.

IBD and visualizations

FlexQTL™ was also used to estimate IBD probabilities at dedicated map positions conditional on marker data and pedigree records. Pedigree information was visually depicted using Pedimap software (Voorrips, 2012), while the other illustrations were created with R scripts (R Development Core Team) as integrated in VisualFlexQTL, the Graphical User Interface of FlexQTL™ .

Candidate gene annotation

Candidate gene annotation was carried out on the 1.0 version of the apple reference genome (https://www.rosaceae.org/tools/ncbi_blast). The probe sequences of the SNPs composing the closest haploblock to the mode of the QTL were blasted on the apple reference genome to retrieve their physical position. SNP showing a divergent physical position than the other SNP composing the haploblock were discarded. The physical position of the haploblock (resulting from averaging the physical position of the underlying SNPs) was compared to its genetic position in a colinearity analysis to detect possible mismatches between genetic and physical mapping.

The genomic region interrogated spanned 400 Kb up and downstream the SNP of interest, in agreement with the mean LD decay in apple (Di Guardo et al. 2017, chapter 4 of this thesis).

Results and Discussion

Fruit firmness trait distribution

The fruit firmness across all full-sib families showed a continuous, slightly skewed distribution for each of the four storage periods (M_1, M_2, M_3 and M_4) (Fig. 1), though the underlying individual families could show a more skewed distribution (Suppl. Fig. 3). The highest values of fruit firmness were registered at harvest (M_1), with an overall mean of 8.3 $Kgcm^{-2}$, which decreased to 6.3 $Kgcm^{-2}$ after four months of cold storage (Table 1, Fig. 1). The distributions of M_3 and M_4 coincided; apparently, the overall effect of two weeks of shelf-life is similar to that of two additional months of cold-storage.

The largest fruit firmness loss in absolute terms was detected between M_1 and M_2, with a decrease of 1.1 units. The decrease between M_2 and M_3, as well as between M_2 and M_4 was slightly less: 0.9 units. In relative terms, the decrease rates were the same (12%). The squared phenotypic correlations (r^2) among the four storage periods varied between 0.67 for M_1- M_3 to 0.86 for M_2-M_4. The r^2 between fruit firmness at harvest and after two and four months of cold storage was 0.81 and 0.73 respectively suggesting firmness at harvest as a good indicator for firmness after storage: fruits showing good firmness at harvest have a high probability to be also relatively firm after storage.

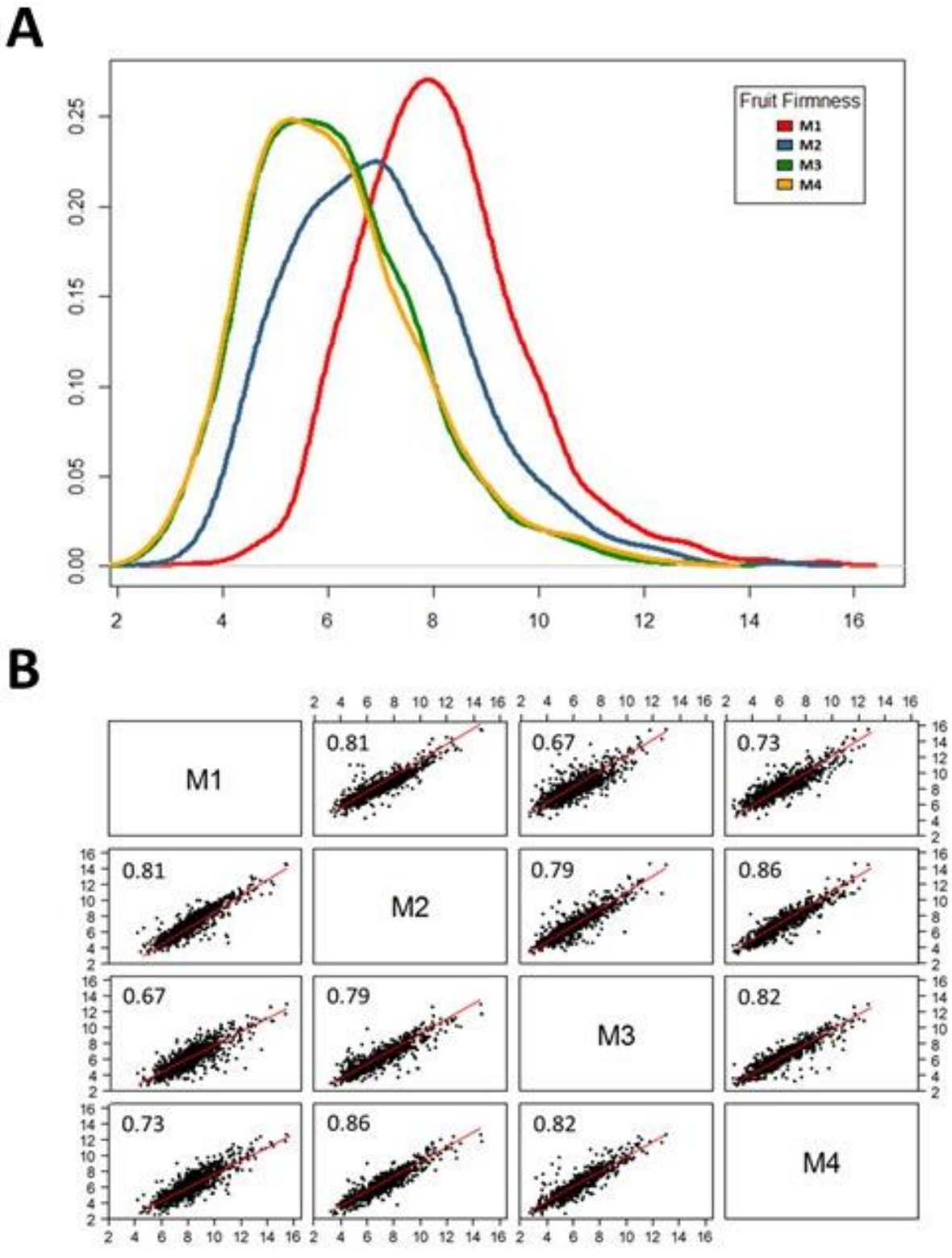


Figure 1. A; Frequency distribution of fruit firmness assessed at the four scoring dates, coded as: M_1 (red line) harvest, M_2 (blue line) after two months of cold storage, M_3 (green line) after two weeks of room temperature shelf-life following two months of cold storage, M_4 (orange line) after four months of cold storage. **B** Correlation between the four storage periods, for each of those the r^2 is reported.

Convergence of the QTL analysis

The MCMC simulations performed by FlexQTL™ showed convergence for all the analysis performed, as demonstrated by the effective chain size (ECS) values, always higher than 200 (data not shown), as well as by the trace plots for the error variance, the genetic variance and the heritability estimate (Suppl. Fig. 4). Convergence was demonstrated also by the trace plots for the posterior QTL positions (Fig. 2A and Suppl. Fig. 5A), which showed a stable pattern over the MCMC simulation.

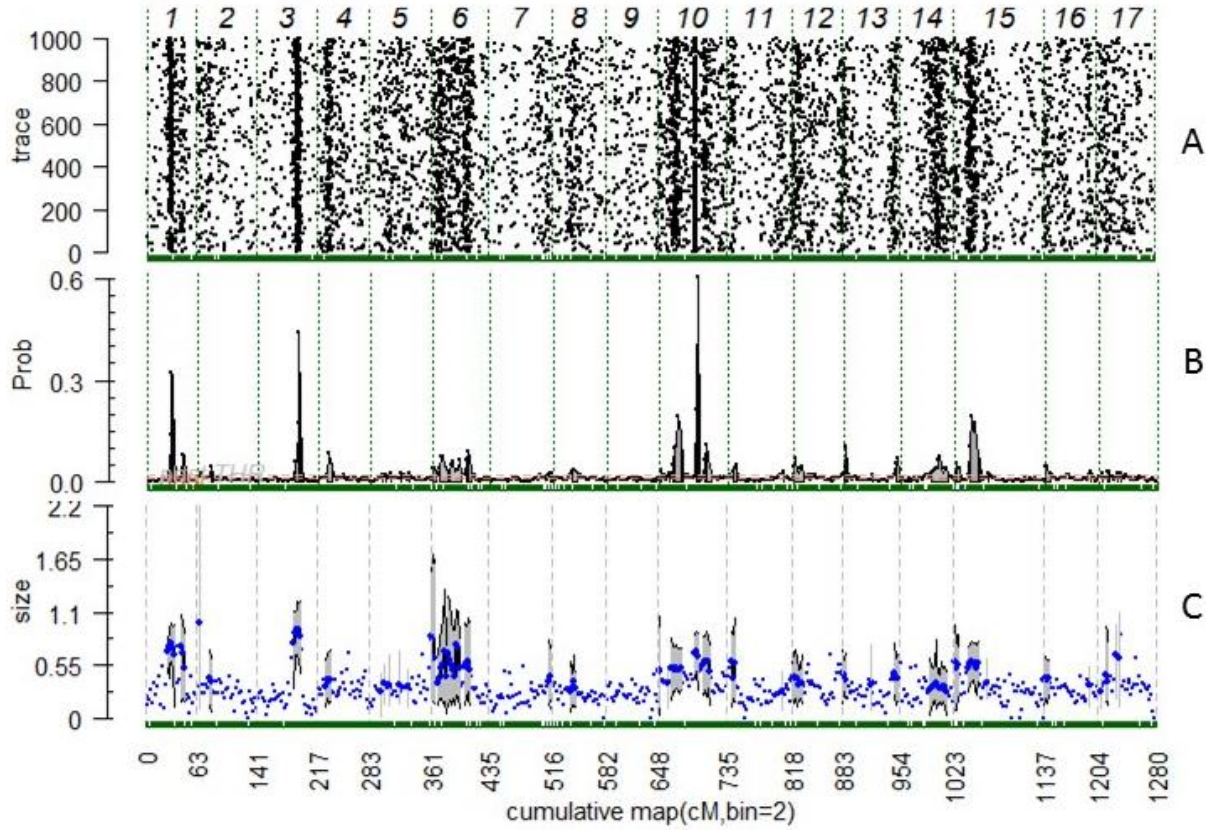


Figure 2. Trace plots of the sampled simulations (**A**) and the QTL posterior positions (**B**) and the posterior mean (blue dots) and 90 % credible region (gray surfaces) estimates (per 2 cM bins) of additive QTL effects (**C**) along the genome at M2. Chromosomes are delimited by grey dashed vertical lines.

QTL discovery

The genome-wide analysis resulted in ten genomic regions that passed our criteria to be considered as QTL for fruit firmness (Table 1, M&M). These ten detected QTLs were located on eight chromosomes and their position is visualized by the QTL intensity profiles (Fig. 2B and Suppl. Fig. 5B, Table 1). Their estimated QTL effect sizes are shown on Fig. 2C, Suppl. Fig. 5C and Table 1. Seven signals for the putative presence of QTLs did not pass our inclusion criteria (having a 2lnBF < 5 and showing up just in one storage period) and will be ignored. Two were

specific for M1 (on LG15 and 17), two for M3 (on LG5 and 17) and three for M4 (on LG 9, 13 and 15).

QTLs common to all four storage periods

Three of the detected QTLs were common to all four storage periods, hypothesizing these genomic regions as major regulators for fruit firmness. They are located on the LG 10, 14 and 15 and named FF-LG10b, FF-LG14 and FF-LG15a respectively. These QTLs and their profiles are further detailed below. The presence of common major QTLs agreed with the high phenotypic correlations among the four storage periods implying that some common genetic regions were involved in changes in fruit firmness and were not influenced by different storage periods.

FF-LG10b has its mode (the genomic region showing the highest probability of harbouring a QTL) at 49 cM. Its evidence was always decisive and it showed increasing size along storage, from 0.46 at M_1 to 0.90 at M_4 (Table 1). The closest haploblock to the mode was HB-10-32 (48.6 cM). This haploblock includes SNPs designed within the sequence of the Polygalacturonase MdPG1 gene, known to encode a cell modifying protein involved in the middle lamella pectin degradation and consequently in the control of firmness loss in apple (Costa et al., 2010) as well as in other climacteric fruit species, such as peach (Lester et al., 1996), kiwifruit (Atkinson and Gardner, 1993) and tomato (Sitrit and Bennett, 1998). In peach, in particular, two genes encoding endopolygalacturonase were detected on LG 4 (Gu et al. 2005), an orthologue of the apple LG 10 (Jung et al. 2012).

FF-LG14 has its mode between 52 and 60 cM. Its evidence varied around the inclusion threshold (BF>5) and its estimated size was similar among M_1, M_2, and M_4 (0.33, 0.32 and 0.37 respectively), and was somewhat higher for M_3 (0.51). The higher effect at M_3, suggests that the role of this QTL increases during shelf-life at room-temperature. The LG14 QTL-region includes a gene coding for a pectinesterase (MDP0000202127) and expansin (MDP0000151618, MDP0000193025) enzyme, and the gene MdRIN (MDP0000326390), of which a homolog in tomato is involved in fruit ripening (Barry et al. 2000).

FF-LG15a has its mode between 21-27 cM, whereby M_1 had a wide QTL interval (1-62 cM) reflecting uncertainty on the location of the QTL. Its evidence was always decisive but still decreased from M_2 on. Its estimated size varied only slightly among the storage periods (0.46-0.56) showing its maximum at M_2 (Table 1).

QTLs specific to one or more storage periods

Seven of the detected QTLs were specific to one or more storage periods, and occurred on six chromosomes.

FF-LG1a and FFQTL1b are both located on LG1 and have their mode at 33cM and 47-48cM respectively (Table 1). They showed up alternately: FF-LG1b was detected at M_1 and M_4 and FF-LG1a at M_2 and M_3. FF-LG1a's first showing up at M_2 may indicate that this QTL becomes active after harvest. Its presence at M_2 and M_3 and absence at M_4 may indicate temporary activity during specific physiological stages during which it seems to dominate and/or mask FF-LG1b. Their estimated sizes decreased in time as follows: FF_{QTL}1b from 1.03 at M_1 to 0.68 at M_4 and FF_{QTL}1a from 0.74 at M_2 to 0.48 at M_3 (Table 1). However, the partial confounding of these two QTL may have inflated these estimates.

The parental cultivars showed a particular pattern for their QTL genotypes. M_1 and M_4 showed highly similar QTL genotypes for FF_{QTL}1b while M_2 and M_3 showed highly similar genotypes for FF-LG1a. However, the two QTLs showed some commonalities and differences of interest. Some of the M_1-heterozygous parents were homozygous 'qq' at M_2 and M_3, and all carried a relatively short chromosome fragment from M. floribunda 821, whereas all the parents for which QTL genotype remained the same, had inherited a long fragment of M. floribunda 821 (Fig. 3). This phenomenon might indicate the presence of two true QTLs that i) derived from a single source, ii) that showed stage specific expression, although the presence of QTL modelling issues could be an alternative explanation. Also, previous reports indicate difficulties in the unequivocal assignment of QTLs for this region as King et al, (2000) reported a QTL for firmness (assessed shortly after harvest) that was located on a similar position as FF-LG1b. Whereas the putative presence of two QTLs was indicated by Maliepaard et al. (2001), while elaborating on the data of King et al (2000).

We found FF-LG1b to co-localize with the pectinesterase gene MDP0000320566 which locates close to the SNP markers SNP_FB_0440719 and ss475876885 of haploblock marker HB-01-29, the most associated marker to this QTL.

The presence of FF-LG1a at both M_2 and M_3 indicates its potential value for breeding of cultivars of good firmness and shelf-life for intermediate storage periods. Statements on the use of FF-LG1b need further clarification on its irregular discovery.

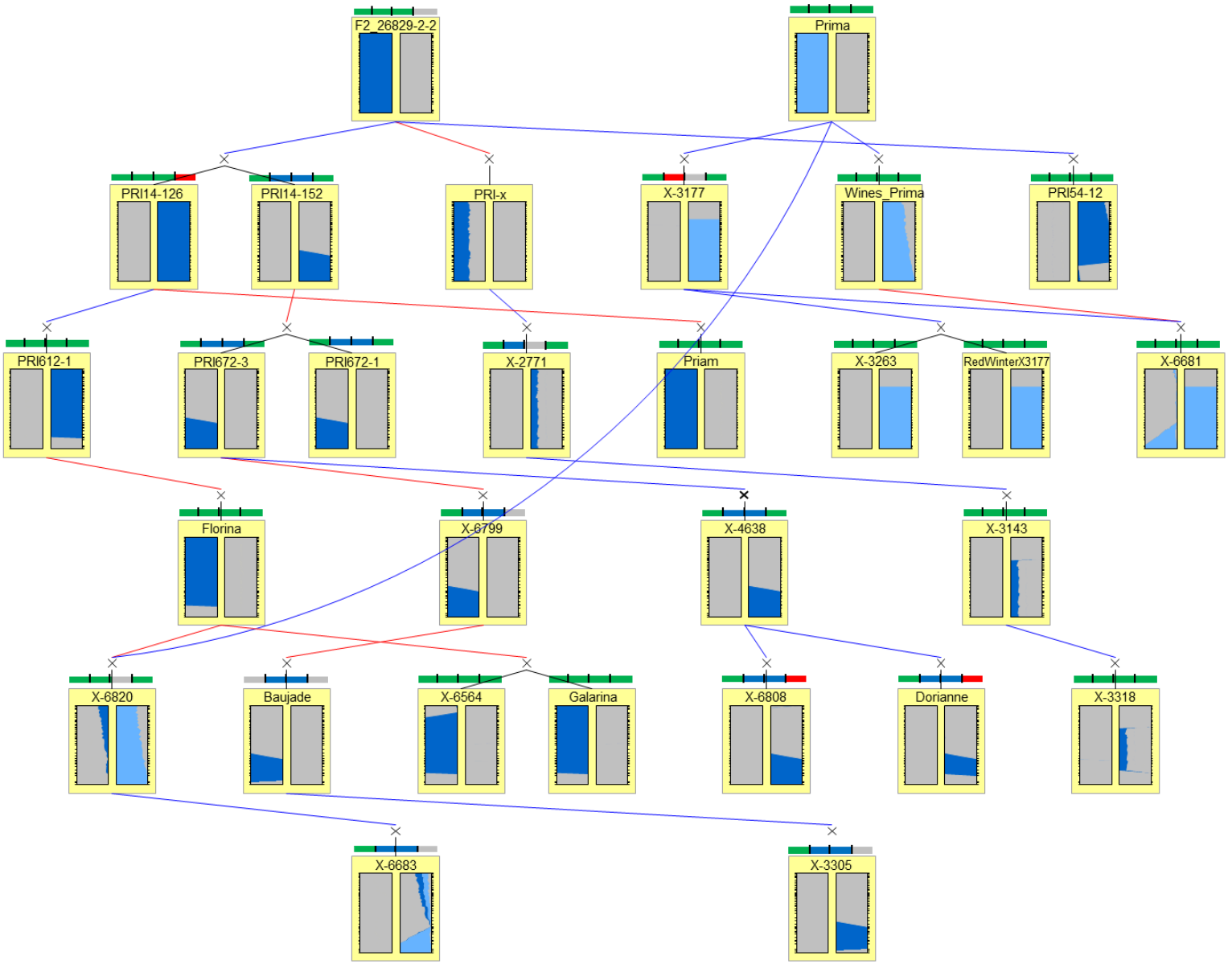

F2_26829-2-2
Prima
PRI14-126
PRI14-152
PRI-x
X-3177
Wines_Prima
PRI54-12
PRI612-1
PRI672-3
PRI672-1
X-2771
Priam
X-3263
RedWinterX3177
X-6681
Florina
X-6799
X-4638
X-3143
X-6820
Baujade
X-6564
Galarina
X-6808
Dorianne
X-3318
X-6683
X-3305

Figure 3. Allelic flow of the chromosome harm carrying Rvi6 (Vf) resistance allele of Malus floribunda 821 (here named F2_26829-2-2) (in blue), and its homologue (light blue). The horizontal bars above each box reported the QTL genotype (blue – 'qq', green – 'Qq' and red – 'QQ') at FF-LG1a (for M_2 and M_3) and FF-LG1b (for M_1 and M_4).

FF-LG3: This QTL has its mode from 54 to 60 cM and it was detected in the first three storage periods (M_1 - M_3), with evidence progressively decreased from decisive (29.7) to positive (4.6). The QTL-effect showed the same trend, passing from 0.93 to 0.66 (Table 1). FF-LG3 thus showing a decreasing effect size over postharvest ripening. The mode of this QTL coincides with an expansin gene (MDP0000306058, MDP0000128827) which is located between the SNP markers SNP_FB_0546048 and SNP_FB_0546475 of HB-03-35c.

FF-LG6: This QTL has its mode at 2-12 cM, and thus has an uncertain position. It had strong evidence at M_2 and M_4 and its estimated size increased from 0.55 to 1.15, suggesting its involvement during cold storage but not when room temperature is restored.

FF-LG10a: This QTL was detected at all storage periods but M_1. It had its mode at 23-24 cM, and showed decreasing evidence and decreasing size over time (from 0.51 to 0.39). The two QTLs on LG10 thus showed opposing trends due to which the estimated effect of FF-LG10a relative to that of FF-LG10b decreased from 74% at M1 to 43% at M4. These two QTLs were also observed by Costa et al. (2010), whereby FF-LG10b coincides with their central QTL, as both harbour MdPG1. Our and their study revealed also similar QTL-dynamics: a decrease in effect of FF-LG10a and an increase of FF-LG10b from harvest to two months after storage. The same opposing patterns were observed in both studies for apples that had been kept at room temperature either after storage (this study) or directly after harvest (Costa et al. 2010).

FF-LG-11: This QTL showed up in M_1 only. It had strong evidence, and an estimated effect of 0.59,

FF-LG16: This QTL showed up in M_4 only. It had strong evidence, and an estimated effect of 0.49.

Estimation of QTL genotypes

The QTL regions that together describe the regulation of fruit firmness (Fig.2 and Supp. Fig. 5) have been identified through the simultaneous analysis of 24 FS families. Not all the QTLs identified segregated in each family. Information on the segregating QTL will provide insight on the number of families and individuals for which a QTL was postulated, and will also allow to estimate the QTL genotypes of the parents (and related germplasm). This information will support breeders in the design of new genetic studies and in guiding the selection of breeding parents and seedlings.

The assignment of QTL genotypes starts from the heterozygous 'Qq' loci as these are a direct function of the postulated segregating QTL (green bars in Fig. 4). Thus, homozygous 'QQ' and 'qq' genotypes were assigned, by combining inheritance information on the already assigned 'Q' and 'q´ along pedigrees (via linked molecular markers), and the phenotypic value of the respective individual. Overview of the QTL genotypes estimated for all the parental cultivars, provides both QTL-wise and family-wise information, showing for each QTL how many parents and families contributed to its discovery (columns in Fig. 4), and, for each parent and family, the number of heterozygous and segregating QTL (rows in Fig.4). For each QTL and storage period, multiple parents showed to be heterozygous leading to various numbers of segregating families per QTL, ranging from a maximum of 15 (FF-LG1b at M_1) to a minimum of 3 (FF-LG3 at M_2 and M_3).

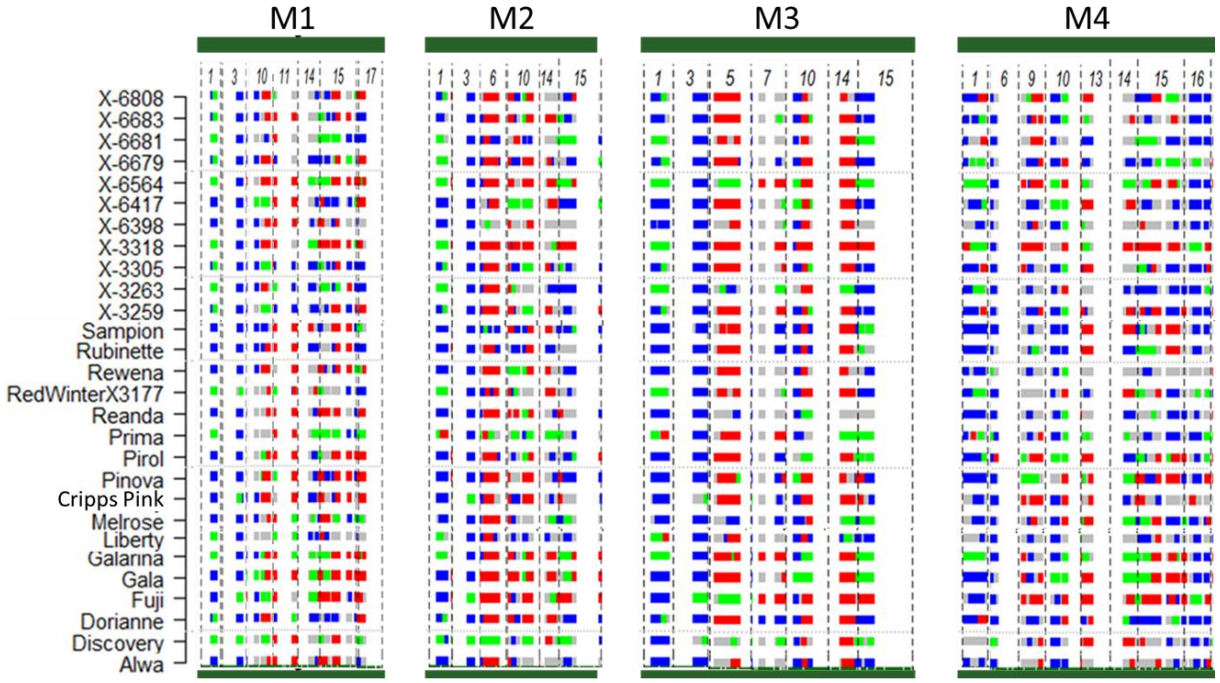


Figure 4. Posterior estimates of QTL genotype probabilities for the 28 parents of the full-sib families along the genome. Estimates are displayed for those genomic regions with at least positive evidence for QTL presence. Red colour indicates evidence for 'QQ' QTL-genotype, blue for 'qq' and green for 'Qq' QTL genotypes. Grey indicates ignorable evidence for any QTL genotype. Chromosomes are delimited by grey dashed lines.

Since the most important changes in sensory perception occur during the first 2 months of cold storage (Kouassi et al., 2009), this storage period was used as reference in our analyses.

Fig. 5 presents the QTL model by which FlexQTL™ explained the observed phenotypic data at M_2 in terms of parental QTL genotypes and QTL sizes. The phenotypic mean of the full-sib families ranged from 5.6 (for 12 P) to 9.2 $Kgcm^{-2}$ (12 F), with the number of Q-alleles ranging from 4 (AlSa) to 18 (12 F). The phenotypic variation within a family correlated with the total

number of heterozygous QTL genotypes across the two parents, spanning from 0 (AlSa and AlPi) to 8 (DiPr) (Fig. 5 A, B). The size of the segregating QTL, ranged from 0.32 (FF-LG14) to 0.91 (FF-LG3) (Table 1). Also, the mean value of fruit firmness of the families correlated with the total number of *'Q'*-alleles and thereby with the QTL genotypes of the parents (Fig 5 C): Indeed, the family with the least overall firmness (12 P) had one of the lowest number of '*Q'*-alleles, whereas the family with the highest overall firmness (12 F) had the highest proportion of '*QQ'* genotypes (Fig. 5 B, C).

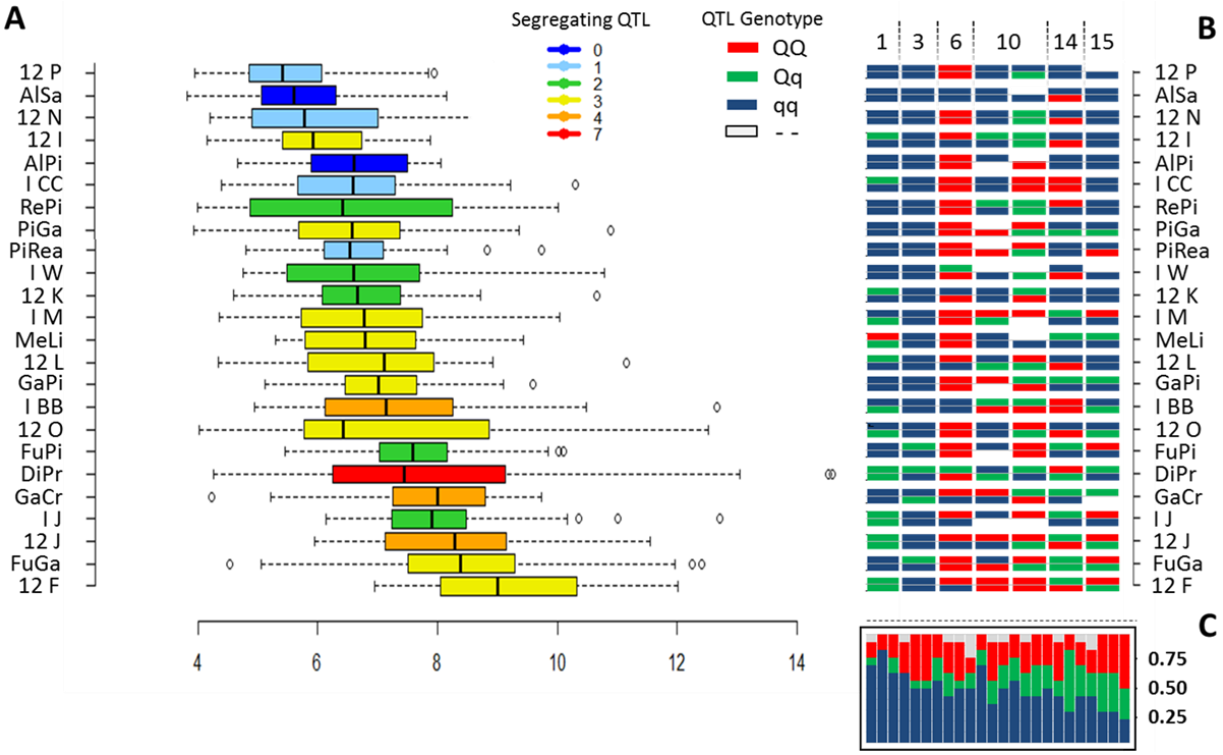


Figure 5. Trait distribution and number of QTL segregating within the 24 FS families analysed. In panel A the boxplots of the fruit firmness assessed at M_2 for each FS family is shown. In panel B, the QTL genotypes for both parents (mother upperline and father bottom line) are presented. Red color indicates 'QQ' genotypes, 'qq' genotypes are plotted in blue, while heterozygous 'Qq' individuals are colored in green. White color is instead used to indicate ignorable evidence for QTL genotype. In panel C, the barplot of the combined frequency of the QTL estimated genotypes for the two parents of each FS family is highlighted. Families (bars in the graph) are ordered in ascending order according to their mean firmness value

Knowledge on QTL-genotypes provides guidance for the choice of parents in genetic research and in breeding: 'Prima' x 'X-3318' would give the widest genetic variation since together the two parents had the largest range in QTL genotypes across all QTLs, whereas 'Dorianne' x 'X-6679' would allow to examine the performance of FF-LG1a under a stable genetic background, as these individuals are assumed to be homozygous for all other QTLs.

The estimated QTL-genotypes may also give an indication on the relevance of a QTL in breeding. In our analyses, FlexQTL™ estimated FF-LG6 to be homozygous 'QQ' in all parents

but two. This may suggest that this locus has been under high selection pressure for the *'Q'* allele. In contrast, FF-LG15a hardly had homozygous 'QQ' parents, indicating good prospects for further increases in firmness, if desired. However, before drawing firm conclusions, the robustness of the genetic model should be examined, as alternative models of similar probability may be feasible for the genetic description and interpretation of the phenotypic data. Our current results on LG 6 and LG15 are a good example for this. The parents that were proposed here to be homozygous 'QQ' for these two QTLs, were genotyped as *'qq'* in Bink et al. 2014, and vice versa for our current *'qq´* genotypes. Apparently, FF-LG6 and FF-LG15 are somehow mutual dependent in our dataset for the assignment of their homozygous genotypes.

Evaluation of QTL performance by single QTL segregating families

Among the twenty-four families employed in this investigation, four harboured only one segregating QTL at M_2, implying that, for these families, the genetic contribution to the phenotypic variation could be attributed to the effect of just one segregating locus, which makes them very suitable to examine the mode of action of a single QTL on a more stable genetic background.

Family 12 N: Both parents of this family were heterozygous for FF-LG10b. The mean firmness of the 'Qq' genotype of the family was intermediate to that of the two homozygous classes (Fig. 6 and Suppl. Fig. 6), thus following an additive mode of action. The estimated QTL size was 1.25 $Kgcm^{-2}$ per dose, which is almost double of the germplasm wide estimation of 0.68 $Kgcm^{-2}$ (Table 1). This large size may be the result of the overall low firmness of this family, under which condition any additional QTL may have a large effect, whereas with increasing overall firmness, additional Q-alleles may show decreasing additional value.

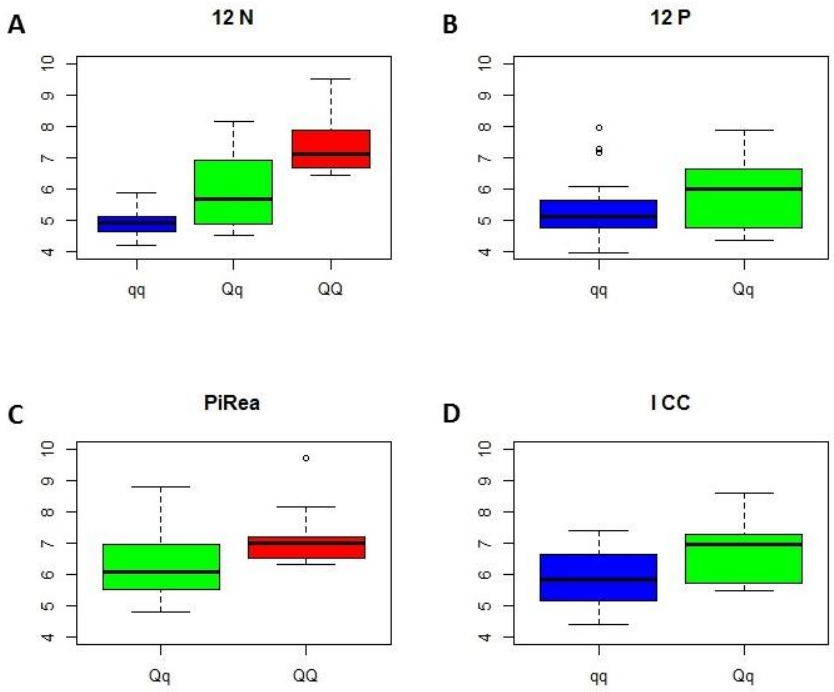


Figure 6. Boxplots for the firmness distribution at M_2 for the four families having a single segregating QTL. FF-LG10b segregates in families 12 N, 12 P and PiRea while FF-LG1a on family I CC. For each family, the QTL genotypes are reported on the x axes while the range of variability of fruit firmness is reported on the y axes.

Family 12 P: One parent was heterozygous 'Qq', and the other homozygous 'qq'. The heterozygous parent ('X-3305') was in common with family 12 N. The 'Qq' and 'qq' progenies had an average firmness of 5.9 and 5.3 $Kgcm^{-2}$, which resulted in an estimated QTL size of 0.6 $Kgcm^{-2}$, which value is slightly lower than the germplasm-wide estimate.

Family PiRea was a 'QQ'x'Qq' cross. The 'QQ' and 'Qq' progenies had an average fruit firmness of 7.1 and 6.0 $Kgcm^{-2}$ respectively, which resulted in an estimated QTL size of 1.1 $Kgcm^{-2}$, similar to that of the 12 N family. The overall firmness of this family is somewhat higher than that of the previous two, which is consistent with the germplasm wide QTL model proposed by FlexQTL™, whereby this family has 10 additional non-segregating Q-alleles whereas the above families 12 N and 12 P only had respectively six and four such alleles.

Summarizing, FF-LG10b seems to have a crucial role on the fruit firmness regulation for all these three families, as demonstrated by the recurrent presence. This QTL showed to be stable across storage (M_2-M4) with regard to statistical evidence, position and interval size (Table 1).

Family I_CC segregated solely for FF-LG1b, and is characterized as a 'Qq'x'qq' cross. The 'qq' and 'Qq' progenies had an average firmness of 5.9 and 6.7 $Kgcm^{-2}$, which resulted in

an estimated QTL size of 0.8 $Kgcm^{-2}$ in agreement with the germplasm wise estimation of the FF-LG1a (0.74 $Kgcm^{-2}$, Table 1, Fig 2C).

Use of haploblock and single SNP markers in breeding

Our genome wide QTL analyses resulted in the detection of six QTLs that had at least strong evidence to be associated to fruit firmness after two months of cold storage (M_2) (Table 1). The length of their QTL intervals varied considerably: from 9 cM for FF-LG10b to 31 cM for FF-LG6 (Table 1). The QTL analysis was carried out using HB haploblocks rather than single SNPs. HB markers combine the favourable features of both SSR and SNP markers: multi-allelism, automated genotyping and high genomic coverage combined to improved computation performance. The use of HB markers coupled with the estimation of the QTL genotype eased the tracing of either the sources of 'Q' and 'q' alleles and the historic recombination events (Voorrips et al. 2016). However, for the performance of genotyping assays, HB markers need to be translated back to their most informative SNP markers. In the following lines, we elaborate on these topics.

Marker HB_10-32 coincided with the mode for the FF_LG10b. The set of parents had five and three haplotypes that were consistently linked with the QTL-allele for high and low firmness respectively, while one haplotype, Hap-21, showed a mixed association, being linked to high firmness with all parents except 'Liberty' (Tables 2 and 3). Each of these parental haplotypes could be traced back to one to seven founders (Tables 2 and 3). Among them, previously unknown, distant genetic relationships were revealed. E.g., three of these founders ('Esopus Spitzenburg', 'Winesap' and 'Red Winter'') arose before 1800 (Suppl. Table 3) and could not be distinguished by age based on historic records. Phased haplotype data for their two LG10 homologs are presented in Fig. 7. Focusing on the region around HB_10-32, it appeared that 'Esopus Spitzenburger' shared extended haplotypes of 38 cM with 'Winesap' and of 17 and 43 cM for the two homologs of 'Red Winter' and suggesting the presence of distant genetic relationships. These common extended haplotypes always represented two continuous regions of 'Esopus Sptzenburg', one from each of the two homologs as would occurred in case of recombination, and the site of transition was identical indicating a single origin. These observations suggest that 'Esopus Spitzenbug' served as a distant progenitor of both Winesap' and 'Red Winter' even through both its maternal and paternal line. Likewise, the extended haplotypes of the younger founder 'Dr Oldenburg' might descent from 'Esopus Spizenburg' and 'Winesap' but not from 'Red Winter'. Extended haplotypes of founders might thus be helpful to clarify ancestral lineages and to identify the most senior sources of a functional QTL allele. They also give insight in the robustness of the obtained QTL estimates. The Identification of the Hap-3–Q association for multiple genetically related individuals that were treated as independent

founders in the QTL discovery analyses, but that might be IBD based on the size of common extended haplotypes, can be perceived as an indication on the robustness of the QTL-genotype estimates.

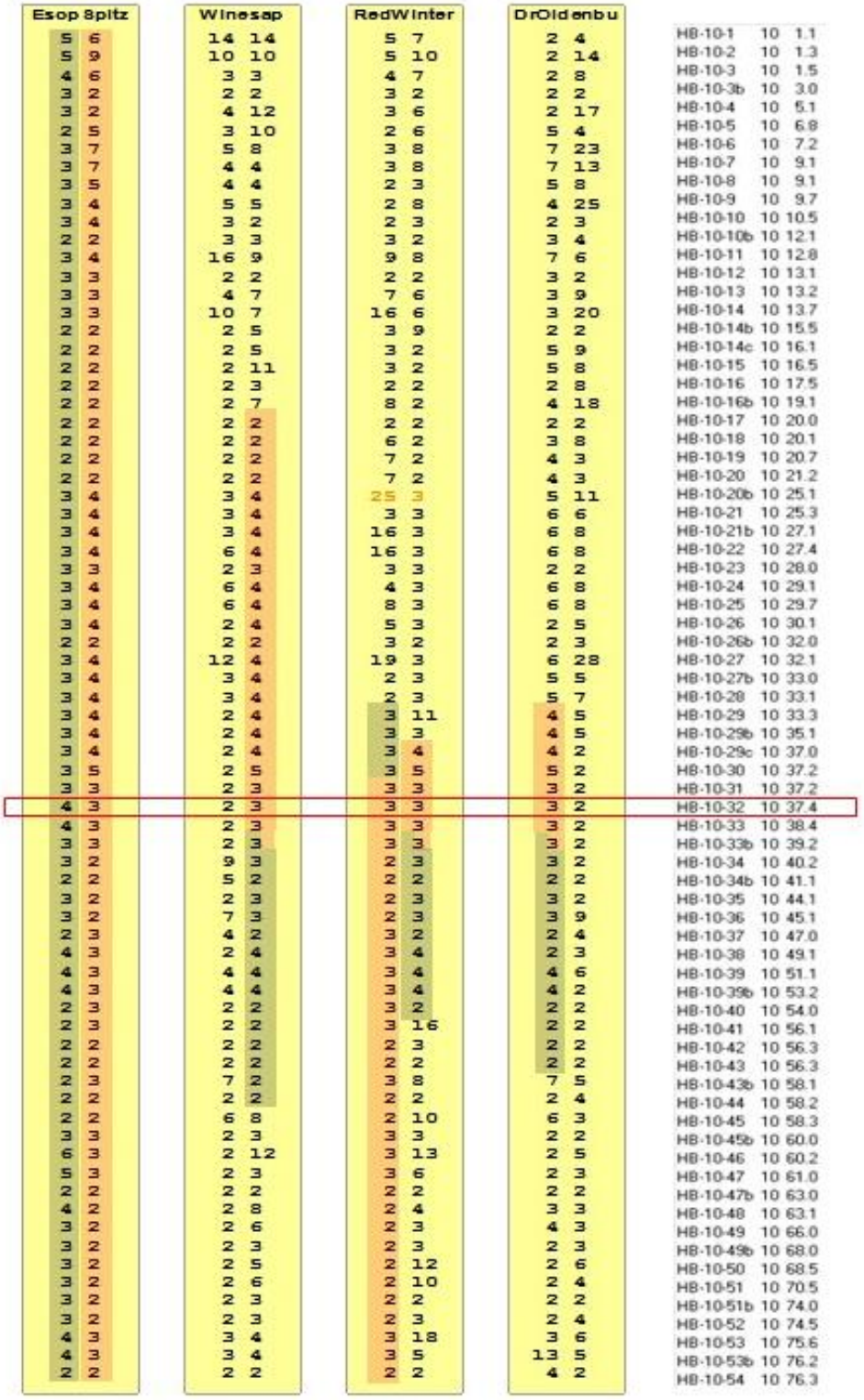


Figure 7. Haplotye composition of four founders for their two homologs of LG10. Orange and green colored boxes mark extended haplotypes common to 'Esopus Spitzenburg' that carry the Hap-3 and Hap-4 of HB_10-32 for high and low firmness respectively.

Three haplotypes were associated with FF-*LG10b*'s q-allele for low firmness. The most prevalent in our study germplasm was Hap-2, which was traced back to seven founders of which some major ones such as 'Golden Delicious', 'Winesap' and 'Cox' (Table 3). Next, one haplotype (Hap-21) seemed to have an ambiguous association as it was associated with a Q-allele for high firmness for three out of the four parents with this haplotype ('Reanda', 'Rewena', and 'Sawa'), whereas in 'Liberty' it was associated with the q-allele for low firmness (Table 2). This inconsistent QTL genotyping for a haplotype that came from the same source by IBD was possible only because the pedigrees that led to 'Borowitsky' were not known at the time that the QTL analysis was performed (Table 3), due to which their genetic relationships were not accounted for. As 'Liberty' was represented by only one full-sib family, and as the QTL genotype of the other parent of this family ('Melrose') was inconclusive, the QTL genotype estimate of 'Liberty' might be incorrect.

Table 2. FF_LG10b-QTL genotypes of the 28 parents of the full-sib families, the marker alleles associated to each of the QTL-alleles for HB_10-31, -32 and -33, and the successive progenitors where these alleles had been inherited from.

Parent	QTL-allele ♀ ♂	Haplotypes HB_10-31 to 33	Successive ancestors (founders in Bold)
Alwa	?	2--2--2	Macoun, **Jersey Black**
	?	3--3--3	**F_Alwa**
Discovery	Q	3--16--18	**Beauty of Bath**
	q	2--2--2	Worcester Pearmain
Dorianne	Q	3--4--2	X-6823, **Golden Delicious**
	Q	3--4--2	X-4638, (PRI672-3), **Golden Delicious**
Fuji	Q	3--4--4	**Ralls Janet**
	Q	3--3--3	Delicious, **Winesap**
Gala	q	2--2--2	Kids Orange Red, Cox
	Q	3--4--2	**Golden Delicious**
Galarina	Q	3--3--3	Florina, PRI612-1, Delicious, **Winesap**
	q	2--2--2	Gala, Golden Delicious
Liberty	q	2--2--2	Macoun, **Jersey Black**
	q	3--21--9	PRI54-12, **Wealthy**, **(Borowitsky)**
Melrose	?	3--3--3	Jonathan**, Esopus Spitzenburg**
	?	3--3--3	Delicious, **Winesap**
Cripps Pink	Q	3--4--2	**Golden Delicious**
	Q	2--27--9	**F_Lady Williams**
Pinova	Q	3--3--3	Clivia, **Dr Oldenburg**
	Q	3--4--2	**Golden Delicious**

Pirol	Q	3--4--2	**Golden Delicious**
	q	2--2--2	Alkmene, Cox
Prima	q	2--2--5	PRI14-510, F2-26829-2-2
	q	2--5--10	NJ123249, ?, Melba, McIntosh
Reanda	q	2--2--2	Clivia, Cox
	Q	3--21--9	**BX44.14, (Borowitsky)**
Rewena	q	2--2--2	Dr.Oldenburg
	Q	3--21--9	**BX44.14**, (**Borowitsky**)
Rubinette	q	2--2--2	Golden Delicious
	q	2--11--2	Cox
Sampion	q	2--2--2	Golden Delicious
	q	2--2--2	Lord Lambourne, James Grieve, F_James Grieve
Sawa	Q	3--21--9	Fantazja, Linda, **Wealthy**, **Borowitsky**
	q	2--5--10	Primula, Melba, McIntosh
X-3259	q	3--5--4	Chantecler, Clochard
	Q	3--16--18	Coop-17, **F_III_#2** (= Winesap open pollinated)
X-3263	?	3--3--3	**Red Winter**
	?	2--2--5	X-3177, Prima, PRI14-510, **F2-26829-2-2**
X-3305	q	3--5--4	Chantecler, Clochard
	Q	3--13--12	Baujade, **Granny Smith**
X-3318	Q	3--3--3	Fuji, Delicious, **Winesap**
	Q	3--3--3	X-3143, **Winesap**
X-6398	?	2--2--5	(X-6820, Prima, PRI14-510, F2-26829-2-2)
	?	3--4--2	(X-4598, F_X-4598, Golden Delicious)
X-6417	Q	3--4--2	**Golden Delicious**
	q	3--5--4	TN_R10A08, Clochard
X-6564	Q	3--3--3	Florina, PRI612-1, Delicious, **Winesap**
	Q	3--4--2	Gala, **Golden Delicious**
X-6679	Q	3--4--2	X-6823, **Golden Delicious**
	Q	3--3--3	Coop-17, **Winesap**
X-6681	q	2--2--2	Winesap x Prima, Winesap
	q	2--2--5	X-3177, Prima, F2-26829-2-2
X-6683	Q	3--3--3	X-6820, Florina, PRI612-1,Delicious, **Winesap**
	q	2--2--2	Golden Delicious
X-6808	Q	3--4--4	X-4638, **Clochard**
	q	2--2--2	Golden Delicious

The five haplotypes with a consistent association with the Q-allele seemed to represent two sub-groups based on their underlying SNP-genotypes, with ‘Esopus Spitzenburg’ and ‘Borowitsky’ as representative members (Table 3). These two sub-groups could be distinguished from each other as well as from the haplotypes for the q-allele by means of two SNP markers, one of which could be selected from a set of five SNPs (FEM_cg_9, _11, _17, _18 and FB_SNP_0832811) each of those was fully predictive for the presence/absence of the ‘Borowitsky’ like group of haplotypes, and a second SNP (ss475879805) which was fully predictive for the presence/absence of the ‘Esopus Spitzenburg’ like group (Table 3). These SNP might represent different mutations that may or may not differ in effect. To examine whether they might be associated to functionally different Q-alleles, full-sib families were selected that segregated for both SNPs as such families provide the most balanced representation of the other QTLs among the different FF-LG10b QTL genotypes, thus approaching an unbiased estimation of allele effects. Only QTL FF-LG10a could not be accounted for in this approach because of its genetic linkage with FF-LG10b. The two sub-groups did not show a significant difference in the firmness frequency distribution of their ‘Qq*’* heterozygous progenies (Suppl. Fig. 7) according to a Kolmogorov-Smirnov two sample (P=0.37). Due to the uniqueness and predictive power of each of the two SNP markers, no phasing is needed in their use in Marker Assisted Breeding, which raises great promises on applicability.

Four of the five alternative SNPs for the ‘Esopus Spitzenburg’ ’source came from an in depth examination of sequence polymorphism for the Md-PG1 gene among nine cultivars (Baumgartner et al., 2016).

Table 3. SNP sequences for the fully genotyped haplotypes of HB-10-32 that carry the B allele for firmness of marker FB_SNP_0832819. All these haplotypes were associated with a QTL allele for high firmness except haplotype 11. Haplotype 11 be differentiated from the others by means of two SNP markers (presented in bold). These markers were selected because being nearest to FB_SNP_0832819 while (1) differentiating the haplotype that was most similar to haplotype 11 (haplotype 27), and (2) separating all remaining haplotypes from haplotype 11. SNP-calls identical to that of haplotype 11 are marked by dark-grey filled boxes

QTL-allele	Haplotype	SNP markers and their genetic position (cM)														Founders in which the haplotype occurs
		FB SNP 0029600	**ss475879805***	FB_SNP_0029606	FEM_cg_8	**FEM_cg_9 ***	FEM_cg_10	**FEM_cg_11 ***	FEM_cg_14	**FEM_cg_17 ***	**FEM_cg_18 ***	FEM_cg_19	FB_SNP_**0832811 ***	FB_SNP_0832819	FB_SNP_0832820	FB_SNP_0030341	
		37.23	37.24	37.25	37.30	37.31	37.32	37.33	37.35	37.36	37.37	37.38	37.42	37.43	37.44	38.00	
Q/q	21	B	A	A	B	B	B	B	A	B	B	B	A	B	B	A	Borowitsky
Q	**4**	B	A	B	A	B	A	B	A	B	B	B	A	B	B	B	Golden Delicious, Borowitsky
	13	A	A	A	B	B	B	B	A	B	B	B	A	B	B	A	Granny Smith
	27	A	A	B	B	B	B	B	A	B	B	B	A	B	B	A	F_Lady williams
	3	A	B	A	B	A	B	A	B	A	A	B	B	B	A	B	Dr.Oldenburg, Esopus Spitzenburg, F_Alwa, Red Winter, Winesap
	16	A	B	A	B	A	B	A	A	A	A	B	B	B	A	B	Beauty of Bath
q	2	A	A	A	B	A	B	A	A	A	A	A	B	A	A	B	Cox, Dr. Oldenburg, F_James Grieve, F2_26829-2-2, Golden Delicious, Winesap, Worcester Pearmain
	5	A	A	A	B	A	B	A	A	A	A	B	B	A	A	B	Clochard, McIntosh
	11	A	A	A	B	A	B	A	A	A	A	B	B	B	A	B	Cox

Conclusion

Fruit firmness is one of the most important components of fruit texture, a multi-factorial trait greatly influencing overall fruit quality. In this work, the Pedigree Based Analysis approach was employed to decipher the genetic control underlying its inheritance. Besides the unravelling of genomic loci involved in the determination of this trait, it also revealed the dynamics of these loci across different postharvest storage periods. While some QTLs (FF-LG10b, FF-LG14 and FF-LG15) remained stable throughout the entire storage periods other QTLs such as FF-LG3 decreased in size from M_1 to M_3 and disappeared at M_4. Other QTLs such as the two located at LG 1 were located in different position according to the storage period. The overall pattern of QTL intervals identified across multiple families together with their estimated genotypes, provided relevant information on the most valuable candidate markers to be used towards the selection of progenies of high fruit quality (Table 1, Suppl. Table 2). PBA bases the QTL detection on the simultaneous analysis of many families. For each genomic region, information on which parental line of the FS families carries a favourable genetic configuration are given (Fig.4 and Fig. 5). The combined use of HB-markers, QTL genotypes and accurate pedigree records allowed the tracing of the sources of favourable alleles and to hypothesize the presence of common sources among founders. The analysis of the QTL that coincided with MdPG1 lead to the differentiation of two distinct groups of founders, represented by 'Esopus Spitzenburg' and 'Borowitsky', that probably represent two independent mutations that each led to a Q-allele of similar effect. Predictive single SNP markers were identified for each of these two sources. The gained knowledge on the location and performance of QTLs for texture and the QTL genotypes of a series of potential breeding parents will be helpful to breeders to set-up ad-hoc breeding schemes directed to the development of new cultivars with outstanding texture and storage characteristics.

Supplementary data

Supplementary data can be found at:

https://drive.google.com/drive/folders/0B1u0mN44Aw3pZ05IZG9XQW9yY2M?usp=sharing

Acknowledgments

Partners of the HiDRAS and Isafruit project and Wannes Keulemans are acknowledged for the use of phenotypic data. François Laurens, Walter Guerra, Andreas Peil, Stefano Tartarini for sharing leave tissue of the discovery germplasm, Andrea Patocchi for sharing DNA of Polish

HiDRAS germplasm and Yolanda Noordijk for DNA isolation of the non-Polish germplasm. The authors want to thank also Michela Troggio and Elisa Banchi for running the iSCANs and Herma Koehorst - van Putten, Sara Longhi, Hélène Muranty, Giulia Pagliarani, Thomas Leitschka, Larisa Garkava-Gustavsson and Johannes Kruisselbrink for their contribution in the generation of the consensus genetic linkage map. This work was co-funded by the EU 'Quality of Life and Management of Living Resource and seventh Framework Programme by respectively the HiDRAS project No. QLK5-2002-01492 'Disease Resistant Apples for a Sustainable Agriculture" and the FruitBreedomics Project No. 265582: Integrated Approach for increasing breeding efficiency in fruit tree crops (www.FruitBreedomics.com). The views expressed in this work are the sole responsibility of the authors and do not necessary reflect the views of the European Commission. The authors are also grateful to the PhD School of Wageningen University and the GMPF program of FEM to have financed the PhD fellowship of Mario Di Guardo

References

Allard A, Bink MCAM, Martinez S, Kelner J-J, Legave J-M, di Guardo M, Di Pierro EA, Laurens F, van de Weg EW, Costes E. 2016. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. Journal of Experimental Botany, erw130.

Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troggio M, Sargent DJ. 2012. Development of a dense SNP-based linkage map of an apple rootstock progeny using the Malus Infinium whole genome genotyping array. BMC Genomics **13**, 203.

Atkinson RG, Gardner RC. 1993. A polygalacturonase gene from kiwifruit (Actinidia-deliciosa). Plant Physiology **103**, 669–670.

Barry CS, Llop-Tous MI, Grierson D. 2000. The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. Plant physiology **123**, 979–986.

Baumgartner IO, Kellerhals M, Costa F, Dondini L, Pagliarani G, Gregori R, Tartarini S, Leumann L, Laurens F, Patocchi A. 2016. Development of SNP-based assays for disease resistance and fruit quality traits in apple (Malus × domestica Borkh.) and validation in breeding pilot studies. Tree Genetics and Genomes **12**.

Bennett a B, Labavitch JM. 2008. Ethylene and ripening-regulated expression and function of fruit cell wall modifying proteins. Plant Science **175**, 130–136.

Bianco L, Cestaro A, Sargent DJ, et al. 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh). PLoS ONE **9**, e110377.

Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE. 2007. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica **161**, 85–96.

Bink MCAM, van Eeuwijk F. 2009. A Bayesian QTL linkage analysis of the common dataset from the 12th QTLMAS workshop. BMC proceedings **3 Suppl 1**, S4.

Bink MCAM, Jansen J, Madduri M, et al. 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theoretical and applied genetics **127**, 1073–90.

Bink MC a M, Totir LR, ter Braak CJF, Winkler CR, Boer MP, Smith OS. 2012. QTL linkage analysis of connected populations using ancestral marker and pedigree information. Theoretical and Applied Genetics **124**, 1097–1113.

Brummell DA2006. Cell wall disassembly in ripening fruit. Functional Plant Biology **33**, 103–119.

Brummell DA, Harpster M. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. In: Carpita, N.C. and Campbell, M. and Tierney M, ed. Plant Cell Walls. Springer Netherlands, 311–340.

Chagné D, Crowhurst RN, Troggio M, et al. 2012. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. PLoS ONE **7**, e31745.

Cosgrove DJ. 2000. Loosening of plant cell walls by expansins. Nature **407**, 321–326.

Costa F, Stella S, Van De Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S. 2005. Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus x domestica Borkh). Euphytica **141**, 181–190.

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, Van de Weg WE. 2010. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus x domestica Borkh.). Journal of experimental botany **61**, 3029–39.

Costa F, Cappellin L, Longhi S, et al. 2011. Assessment of apple (Malus × domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biology and Technology **61**, 21–28.

Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F. 2012. Texture dynamics during postharvest cold storage ripening in apple (Malus x domestica Borkh.). Postharvest Biology and Technology **69**, 54–63.

Di Guardo M, Micheletti D, Bianco L, et al. 2015. ASSIsT: An Automatic SNP ScorIng Tool for in- and outbreeding species. Bioinformatics, btv446.

Di Guardo M, Bink MCMC, Guerra W, et al. 2017. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. Journal of Experimental Botany, **68**, 1451-1466.

Fresnedo-Ramírez J, Bink MCAM, van de Weg E, Famula TR, Crisosto CH, Frett TJ, Gasic K, Peace CP, Gradziel TM. 2015. QTL mapping of pomological traits in peach and related species breeding germplasm. Molecular Breeding **35**, 166.

Fresnedo-Ramírez J, Frett TJ, Sandefur PJ, et al. 2016. QTL mapping and breeding value estimation through pedigree-based analysis of fruit size and weight in four diverse peach breeding programs. Tree Genetics & Genomes **12**, 25.

Gianfranceschi L, Soglio V. 2004. The European project HiDRAS: innovative multidisciplinary approaches to breeding high quality disease resistant apples. Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, 327–330.

Giovannoni J. 2001. Molecular biology of fruit maturation and ripening. Annual Review of Plant Physiology and Plant Molecular Biology **52**, 725–749.

Guan Y, Peace C, Rudell D, Verma S, Evans K. 2015. QTLs detected for individual sugars and soluble solids content in apple. Molecular Breeding **35**, 135.

Hadfield KA, Bennett AB. 1998. Polygalacturonases: many genes in search of a function. Plant Physiology **117**, 337–343.

Harada T, Sunako T, Wakasa Y, Soejima J, Satoh T, Niizeki M. 2000. An allele of the 1-aminocyclopropane-1-carboxylate synthase gene (Md-ACS1) accounts for the low level of ethylene production in climacteric fruits of some apple cultivars. Theoretical and Applied Genetics **101**, 742–746.

Harker FR, Gunson FA, Jaeger SR. 2003. The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. Postharvest Biology and Technology **28**, 333–347.

Iezzoni A, Weebadde C, Luby J, Chengyan Y, van de Weg E, Fazio G, Main D, Peace CP, Bassil N V, McFerson J. 2010. ROSBREED: enabling marker-assisted breeding in Rosaceae. Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, 389–394.

King GJ, Maliepaard C, Lynn JR, et al. 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.

Kouassi AB, Durel CE, Costa F, et al. 2009. Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genetics and Genomes **5**, 659–672.

Lester D, Sherman WB, Atwell BJ. 1996. Endopolygalacturonase and the Melting Flesh (M) Locus in Peach. Journal of the American Society for Horticultural Science **121**, 231–235.

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus x domestica Borkh.). Journal of Experimental Botany **63**, 1107–1121.

Longhi S, Cappellin L, Guerra W, Costa F. 2013a. Validation of a functional molecular marker suitable for marker-assisted breeding for fruit texture in apple (Malus × domestica Borkh.). Molecular Breeding **32**, 841–852.

Longhi S, Hamblin MT, Trainotti L, Peace CP, Velasco R, Costa F. 2013b. A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (Malus x domestica Borkh). BMC plant biology **13**, 37.

Morgan J, Richards A. 2002. The new book of Apples: the definitive guide to apples, including over 2000 varieties. Ebury Press.

Nishiyama K, Guis M, Rose JKC, et al. 2007. Ethylene regulation of fruit softening and cell wall disassembly in Charentais melon. Journal of Experimental Botany **58**, 1281–1290.

Oraguzie NC, Iwanami H, Soejima J, Harada T, Hall a. 2004. Inheritance of the Md-ACS1 gene and its relationship to fruit softening in apple (Malus x domestica Borkh.). Theoretical and Applied Genetics **108**, 1526–1533.

Powell ALT, Kalamaki MS, Kurien P a, Gurrieri S, Bennett AB. 2003. Simultaneous transgenic suppression of LePG and LeExp1 influences fruit texture and juice viscosity in a fresh market tomato variety. Journal of agricultural and food chemistry **51**, 7450–5.

Roach JA, Verma S, Peres NA, Jamieson AR, van de Weg WE, Bink MCAM, Bassil N V., Lee S, Whitaker VM. 2016. FaRXf1: a locus conferring resistance to angular leaf spot caused by Xanthomonas fragariae in octoploid strawberry. Theoretical and Applied Genetics.

Rose JKC, Bennett AB. 1999. Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: Parallels between cell expansion and fruit ripening. Trends in Plant Science **4**, 176–183.

Rose J, Hadfield K, Labavitch J, Bennett A. 1998. Temporal sequence of cell wall disassembly in rapidly ripening melon fruit. Plant physiology **117**, 345–61.

Rosyara UR, Bink MC a M, van de Weg E, Zhang G, Wang D, Sebolt A, Dirlewanger E, Quero-Garcia J, Schuster M, Iezzoni AF. 2013. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. Molecular Breeding **32**, 875–887.

Sitrit Y, Bennett AB. 1998. Regulation of Tomato Fruit Polygalacturonase mRNA Accumulation by Ethylene: A Re-Examination. Plant Physiology **116**, 1145–1150.

Voorrips RE, Bink MCAM, van de Weg WE. 2012. Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. The Journal of heredity **103**, 903–7.

van de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM. 2004. Pedigree genotyping: A new pedigree-based approach of QTL identification and allele mining. Acta Horticulturae **663**, 45–50.

van Ooijen JW. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Research **93**, 343–349.

Young ND, Tanksley SD. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theoretical and Applied Genetics **77**, 95–101.

Zhu Y, Barritt BH. 2008. Md-ACS1 and Md-ACO1 genotyping of apple (Malus x domestica Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genetics and Genomes **4**, 555–562.

Chapter 4

Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association

Mario Di Guardo[1,2,] Marco CAM Bink[3,4,] Walter Guerra[5], Thomas Letschka[5], Lidia Lozano[5], Nicola Busatto[1], Lara Poles[6], Alice Tadiello[1], Luca Bianco[1], Richard GF Visser[7], Eric van de Weg[7] and Fabrizio Costa[1]

[1]Fondazione Edmund Mach, via Mach 1, 38010 San Michele all'Adige, Trento (Italy)

[2]Graduate School Experimental Plant Sciences, Wageningen University, PO Box 386, 6700 AJ Wageningen, The Netherlands

[3]Biometris, Wageningen University and Research Centre,

[4]Hendrix Genetics P.O. box 114, 5830 AC, Boxmeer, The Netherlands (current address)

[5]Laimburg Research Centre for Agriculture and Forestry, via Laimburg 6, 39040 Ora (BZ) Italy

[6]Innovation Fruit Consortium (CIF), via Mach 1, 38010 San Michele all'Adige, Trento (Italy)

[7] Plant Breeding, Wageningen University and Research, Droevendaalsesteeg 1, PO Box 386, 6700 AJ Wageningen, The Netherlands

Published in Journal of Experimental Botany (2017) 68 (7): 1451-1466. doi: 10.1093/jxb/erx017

Abstract

Fruit texture is a complex feature composed by mechanical and acoustic properties relying on the modifications occurring in the cell wall throughout fruit development and ripening. Apple is characterized by a large variation in fruit texture behavior that directly impacts both consumer's appreciation and postharvest performance. To comprehensively decipher the genetic control of fruit texture, two complementing QTL mapping approaches were employed. The first was represented by a PBA carried out on six full-sib pedigreed families, while the second was a GWAS performed on a collection of 233 apple accessions. Both plant materials were genotyped with a 20K SNP array and phenotyped with a sophisticated high-resolution texture analyzer. The overall QTL results indicated the fundamental role of chromosome 10 in controlling the mechanical properties, while chromosomes 2 and 14 were more associated to the acoustic response. The latter QTL, moreover, showed a consistent relationship between the QTL estimated genotypes and the acoustic performance assessed among seedlings. The in silico annotation of these intervals revealed interesting candidate genes potentially involved in the fruit texture regulation, as suggested by the gene expression profile. The joint integration of these approaches sheds light on the specific control of fruit texture, enabling important genetic information to assist the selection of valuable fruit quality apple varieties.

Introduction

Fruit ripening is an orchestra of physiological changes occurring to render fruits more attractive and palatable (Giovannoni, 2001) . This important quality feature depends on the dismantling of the primary cell wall polysaccharide complex by a series of cell wall modifying proteins (Brummell and Harpster, 2001; Brummell, 2006). This synergic enzymatic action leads to different types of fruit texture behaviors in apple, from soft and mealy to firm and crisp, suggesting that rather than a single trait, fruit texture can be therefore considered as a multiple feature, distinguished in specific mechanical and acoustic properties (Szczesniak, 2002; Zdunek et al., 2010; Costa et al., 2011). In the last decades, the control of fruit texture has represented a major goal towards the improvement of shelf-life performance (Matas et al., 2009). This aspect is of crucial importance, especially in case of year-round fruit marketability and over-seas shipping. The use of transgenic lines (Kramer and Redenbaugh, 1994) and whole transcriptome platforms has in fact identified several genes deputed to the cell wall metabolism, such as expansin, pectin acetylesterase, xyloglucan endotransglycosylase, pectin methylesterase, pectate lyase and polygalacturonase (Rose et al., 2002 and 2004; Marin-Rodriguez et al., 2003; Eriksson, 2004; Brummell, 2006; Vicente et al., 2007; Janssen et al., 2008; Moore et al., 2008; Soglio et al., 2009; Costa et al., 2010a; Jimenez-Bermudez et al., 2012; Segonne et al., 2014). In support of these studies, also genome-wide mapping of QTLs controlling fruit firmness is compelling, with the final purpose to identify the most valuable molecular markers suitable for marker assisted breeding programs. An exhaustive knowledge of the fruit texture genetic make-up is in fact essential to guide the selection of the most valuable ideotypes in breeding by design approaches (Peleman and Van Der Voort, 2003). In this regard it is worth underlining that the majority of QTLs associated to fruit texture characteristics have been focused mainly on one measurement, fruit firmness, and usually restricted to one or few bi-parental mapping populations (King et al., 2000; Harada et al., 2000; Liebhard et al., 2003; Oraguzie et al., 2004; Kenis et al., 2008; Costa et al., 2010b; Chagné et al., 2014; Ben Sadok et al., 2015). To overcome this constraint, an important effort was represented by the simultaneous analysis of multiple populations connected in a common pedigree scheme, namely pedigree based analysis (PBA) (Bink et al., 2008). This method has been already successfully employed to target QTLs in apple (Bink et al., 2014; Allard et al., 2016) as well as in cherry (Rosyara et al., 2013) and peach (Fresnedo-Ramírez et al., 2015, 2016). In addition to this, also association mapping study has been largely implemented as a complementary strategy to classical QTL mapping (Rafalski, 2010). Although this approach was initially employed in annual crops (Thornsberry et al., 2001; Weber et al., 2008; Stracke et al., 2009) and forest trees (Neale and Savolainen, 2004; González-Martínez et al., 2007, 2008; Eckert et al., 2009; Neale and Kremer, 2011), it has also been recently exploited in fruit tree crops, such

as grapevine (Cardoso et al., 2012), peach (Micheletti et al., 2015) and apple (Kumar et al., 2013, 2015). Especially in the latter species, a major QTL for fruit firmness was observed on chromosome 10, which coincided with MdPG1, a gene known to encode for a polygalacturonase playing a pivotal role in the depolymearization of pectins (Sitrit and Bennett, 1998; Brummell and Harpster, 2001). These investigations, however, are characterized by a low phenotyping resolution, to date recognized as the major operational bottleneck limiting the power of genetic analysis (Cobb et al., 2013).

To this end, the dissection of the fruit texture complexity was carried out with a texture analyzer, an instrument that has already demonstrated its reliability in dissecting the apple fruit texture into mechanical and acoustic signatures (Costa et al., 2012). To step forward in the deciphering of the genetic control of fruit texture in apple, a double approach was employed. Initially six full-sib pedigreed families were investigated through a PBA approach to detect QTLs associated to mechanical and acoustic fruit texture features. These regions were further complemented and validated by genome wide association study (GWAS) performed on a large apple germplasm collection to exploit a much larger quote of both genetic and phenotypic variation.

Materials and Methods

Plant Materials

In this study, two groups of plant materials were employed. The first was represented by a pedigree composed by thirteen parental lines and six full-sib populations (for a total of 416 individuals). The scheme generated with PediMap (Voorips et al., 2012; Supplementary Fig. S1 and Supplementary Tab. S1) shows the maternal and paternal descendants from founders to progenies. The second group was represented by a collection of 387 apple accessions (Malus × domestica species; Supplementary Tab. S2), retrieved from the general apple variety repository available at the Fondazione Edmund Mach (FEM). The germplasm collection and two full-sib progenies (namely ‘FjDe’: ‘Fuji’x’Delearly’ and ‘FjPL’: ‘Fuji’ x ‘Cripps Pink’) were planted at the experimental orchard of FEM, while the other four full-sib families (‘GDFj’: ‘Golden Delicious’ x ‘Fuji’, ‘GaPL’: ‘Gala’ x ‘Cripps Pink’, ‘GaPi’: ‘Gala’ x ‘Pinova’ and ‘FjPi’: ‘Fuji’ x ‘Pinova’) were chosen from the breeding program ongoing at the Laimburg Research Centre for Agriculture and Forestry. Within each group of plant materials (pedigreed full-sib families and germplasm), trees were in a full fruit bearing stage at the time of the analysis. Moreover, while individuals from full-sib families were characterized by a single and original tree, the apple accessions included in the germplasm collection were represented by triplicates.

Fruit harvesting and high-resolution texture phenotyping

Fruit, from both the pedigreed full-sib families and cultivar collection, were harvested at a commercial maturity stage according to typical fruit characteristics, such as starch degradation (selected at 7 on the base of 1 to 10 scale) and skin color. After harvest, fruit were stored for two months at 2°C with 95% of RH. Prior phenotyping, apples were maintained at 20 °C overnight.

The high-resolution phenotyping of fruit texture was carried out with a computer controlled Texture Analyzer TAXT-AED (StableMicroSystem, Godalming, UK), according to the protocol described by Costa et al. (2011 and 2012). Since the texture analyzer was equipped with an AED (Acoustic Envelop Device), for each sample (flesh disc) a simultaneous profiling of the mechanical displacement and acoustic response was acquired. The combined profile was further processed with an ad hoc macro for the digital definition of twelve parameters, related to both mechanical (eight) and acoustic (four) texture properties (specified in Fig. 1). The phenotype was assessed for two consecutive experimental years and represented by BLUP (Best Linear Unbiased Prediction, “lme4” R package), which adjusted the mean by reducing the error variance.

SNP genotyping

Genomic DNA was isolated from young leaves with the Qiagen DNeasy Plant Kit (Qiagen). DNA quantity and quality was measured with a Nanodrop ND-8000 (ThermoScientific, USA). The genotyping was carried out with the 20K Infinium array (Illumina; Bianco et al., 2014; www.fruitbreedomics.com). The SNP data, initially processed with the GenomeStudio Data Analysis software, were finally elaborated with ASSIsT (Di Guardo et al., 2015), a dedicated stand-alone pipeline to filter and re-edit SNP call with distorted segregation pattern due to the presence of null-allele (Pikunova et al., 2013).

Linkage mapping and SNP marker consensus order

The array of SNPs segregating within the six bi-parental pedigreed populations was used to create a bi-parental integrated map for each full-sib family, using the software JoinMap 4.1 (Van Ooijen, 2006). Markers were initially clustered in linkage groups with a minimum LOD value of 3.0 and further ordered with a recombination frequency of 0.45 and Haldane mapping function. The six genetic maps were finally merged into a consensus map using BioMarcator software v4.2 (Sosnowski et al., 2012), through the implementation of the ConsMap module. The new reference and harmonized marker order (Supplementary Tab. S3) was exploited as map input file for both PBA and GWAS analysis.

QTL mapping by Pedigree Based Analysis

The identification and mapping of QTLs at genome-wide scale was carried out with FlexQTL™ (Bink and van Eeuwijk, 2009; Bink et al., 2014; www.wur.nl/en/show/flexqtl.htm). The Bayesian approach on which the software is based compares different models, considered as a random variable. The linear model is expressed as follow:

$$y = \mu + Wa + e$$

where **y** is the observed phenotype, **μ** is the phenotypic mean, **W** is a matrix of a vector of regression on the QTL covariates (**a**) and **e** is the residual error of the model. The model operated by FlexQTL™ is based on independent assignment of QTL alleles to founders, therefore the genotype of the several individuals is a priori unknown. Since the joint posterior distribution for the number of QTLs cannot be analytically computed due to the high number of genotype combinations, a MCMC simulation is used. The Effective Chain Size (ECS, Gianola and Sorensen, 2002), used to assess the sensitivity of the posterior inference, was considered statistically significant with values higher than 100 for phenotypic mean (μ), QTL explained variance (σ^2_a), QTL residual variance (σ^2_e) and number of expected QTLs on the a priori distribution (N_{QTL}), respectively. In the analysis carried out here, 500.000 iterations were performed and a thinning of 500 was applied to reduce computation storage. For each run the convergence quality was represented by a trace plot for visual inspection (Supplementary Fig. S2). For each chromosome, the number of QTL(s) was inferred comparing models with increased number of QTLs. In FlexQTL™, the most likely number of QTLs was inferred based on Bayes Factors (BF), which represents the ratio of the marginal likelihood under one model compared to the marginal likelihood under a second model. A 2lnBF ≥ 2, 5 or 10 indicates a positive, strong or decisive evidence of a model, respectively. Moreover, each QTL is considered as bi-allelic, with three possible genotypic conformations: ‘QQ’, ‘Qq’ and ‘qq’. The analysis was carried out with an additive genetic model, assigning to ‘QQ’ and ‘qq’ a value of 1 and -1, respectively, while ‘Qq’ was equal to 0. The QTL genotype of each individual included into the pedigree is a priori unknown and the alleles are assigned to founders tracing their transmission to offspring. To reduce computation time and increase marker informativeness, the initial set of 10.695 markers was finally converted into 1.045 haploblocks with PediHaplotyper (Voorips et al., 2016)

Linkage Disequilibrium Analysis

The SNP markers exploited on the germplasm collection were initially employed to estimate the linkage disequilibrium (LD) decay. From the 20K SNPs, 7.378 markers were actually used for LD analysis, excluding rare alleles with a MAF < 0.05 and those showing an

incongruent physical/genetic position. The pair-wise r^2 between SNP markers was calculated with PLINK (Purcell et al., 2007). For both chromosome-wise and genome-wide scale, the LD decay was depicted by plotting the pair-wise r^2 value against the corresponding physical distance on the genome (bp). The estimation of the LD decay distance was defined by crossing the r^2 baseline (based on the 95th percentile of the marker distribution, according to Breseghello and Sorells, 2006) and the locally weighted polynomial regression based-fitting curve (LOESS) fitted to the plot ("stats" R package). For each chromosome, the linkage disequilibrium level was also depicted by partitioning the chromosomal regions into segments of strong LD with Haploview (Barrett et al., 2005).

Population Structure and Genome-wide association mapping

The level of genetic stratification was assessed with STRUCTURE v2.3.1 (Pritchard et al., 2000). To this end, 17 SSRs (Supplementary Tab. S4) were amplified according to the protocol reported in Di Guardo et al. (2013). The SSR genetic data were further used to compute the posterior probability (Pr(X|K)), given a specific number of group K (ranging from K=2 to K=8). The computation was carried out performing five independent runs of 1.000.000 burn-in generations and considering the admixture model. The most probable number of populations was identified with STRUCTURE HARVESTER (Earl and vonHoldt, 2012) and the final population structure matrix (Q) was further implemented as covariate in GWAS analysis.

The marker-trait association analysis was carried out with TASSEL v3 software (Bradbury et al., 2007; http://www.maizegenetics.net). The significance of the association was tested implementing both the general linear model (GLM) and the mixed linear model (MLM). The GLM (Pritchard et al., 2000) was computed correcting for population structure. The MLM model (Zhang et al. 2010), instead, included both fixed and random effects, allowing the incorporation of genetic relationship as follow:

$$y=X\beta+Zu+e$$

where **y** represents the phenotype (vector of observation), **β** is an unknown vector containing fixed effects (marker and Q), **u** is an unknown vector of random additive genetic effects, **X** and **Z** are the known designed matrices and **e** is the vector of random residuals. The MLM considered both Q (population structure) and K (kinship) matrixes as covariate for population and parental relationship correction (false positive). Significant associations were selected according to the P-value ≤ 0.05, after FDR correction for multiple comparison according to the procedure of Benjamini and Hochberg (1995) using "stats" R package. For each trait considered in the association, the choice of the model was suggested by the visual inspection of the Q-Q plot, obtained with the "qqplot" R package.

Gene expression analysis by RT-qPCR

To assess whether a change in gene expression corresponds to a different QTL estimated genotype, the transcription profiling of MdACO1 and MdPG1 was assessed at harvest and after two months of cold storage. To achieve this goal, the three parental lines ('Delearly', 'Fuji' and 'Cripps Pink') together with four seedlings for 'FjDe' (45, 125, 10 and 14) and 'FjPL' (23, 25, 35 and 68) were selected. Fruit mesocarp was cut, frozen in liquid nitrogen, grinded into a fine powder and stored at -80°C until processing. RNA extraction, quantification and RT-qPCR were carried out according to the methods described in Busatto et al. (2015). The final Ct is represented by the average of three independent normalized expression values for each sample and an actin gene (MdACT) was employed as housekeeping (Di Guardo et al., 2013). For each gene, a couple of discriminant and specific primer was designed (Supplementary Tab. S5), using Primer3 (http://primer3.ut.ee) and Primique (http://cgi-www.daimi.au.dk/cgi-chili/primique/ front.py).

Results and Discussion

High-resolution phenotyping of fruit texture behavior in apple

The apple fruit texture was assessed with a novel and sophisticated texture analyzer (Costa et al., 2011 and 2012). The overall phenotypic fruit texture variability was initially represented by a PCA plot (Fig. 1).

For both groups of plant materials, the fruit texture parameters were similarly oriented, with a consistent incidence of the two principal components chosen to define the PC-hyperspace. Comparing the two plots, PC1 explained 71,6% of the total phenotypic variance in the pedigreed full-sib families (Fig. 1 A and B) and 79.6% in the germplasm collection (Fig. 1 C and D), while PC2 accounted for 19.9% and 12.7%, respectively. Individuals were distributed following the loadings' projection (Fig. 1 B and D) represented by the twelve texture parameters, which clearly discriminated the two signatures. In both scenarios, the variables were in fact distinctively oriented towards two PCA quadrants. The mechanical parameters (highlighted by the numerical code from 1 to 8) were for the most plotted in the PC1 positive/PC2 negative quadrant, while the acoustic parameters (9-12) were projected on the quadrant defined by positive values for both PCs, besides the number of force peaks (8). Although this index is considered a mechanical parameter it is more correlated with the group acoustic indices, justified by the mechanism behind the generation of the acoustic response and pressure progression (Vincent, 1998).

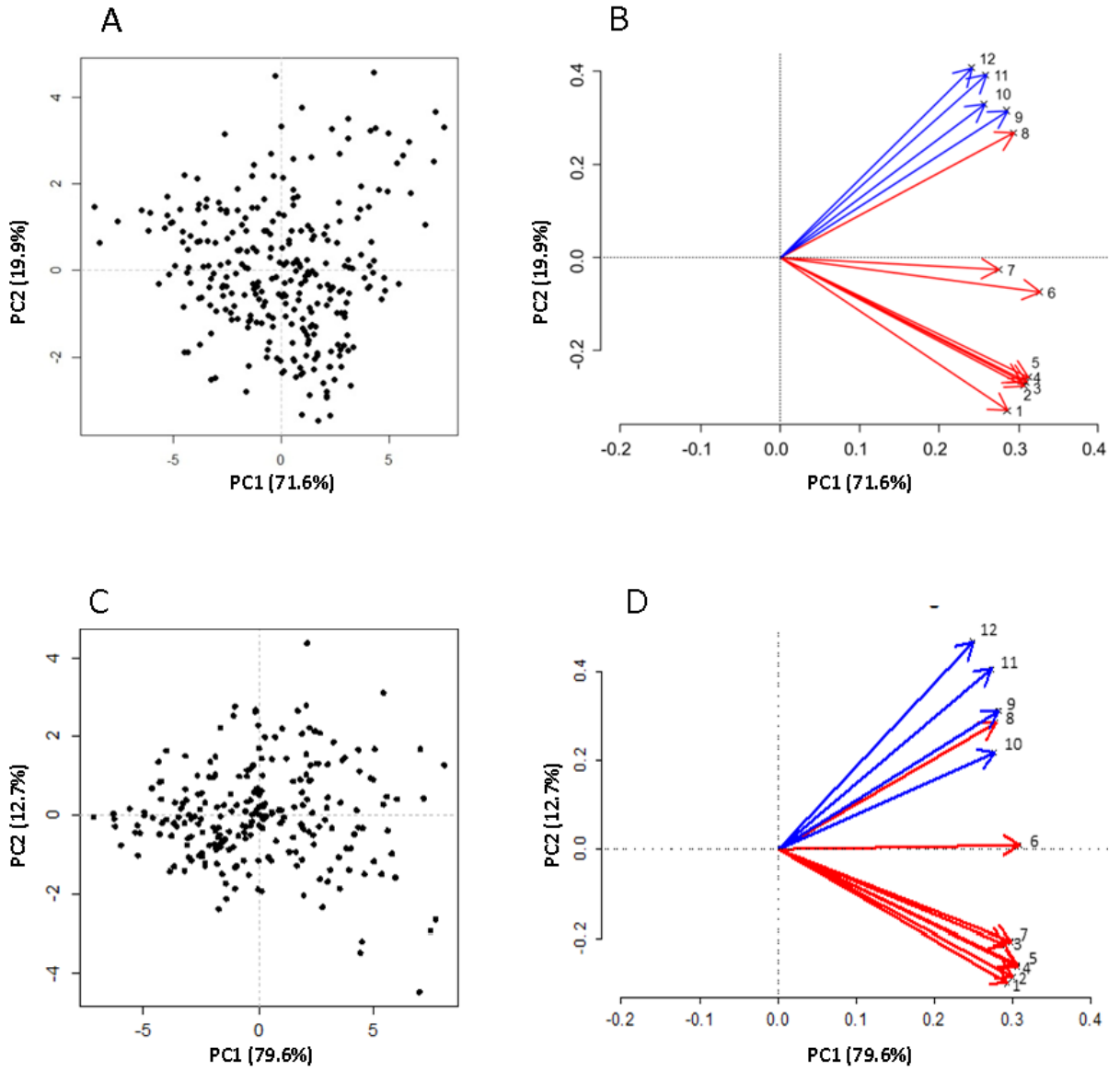


Figure 1. 2D-PCA plots depicting the fruit texture variability assessed in the six pedigreed-families (panels A and B) and germplasm collection (panels C and D). For both plant materials, the sample distribution over the hyperspace defined by the first two PCs (A and C) and the loading projections (B and D) are illustrated. In the loading panels, red colored arrows indicates the mechanical parameters, while acoustic parameters are visualized with blue arrows. Each parameter is coded with a number as follow, 1: initial force, 2: maximum force, 3: final force, 4: mean force, 5: area, 6: force linear distance, 7: Young's module, 8: number of force peaks, 9: maximum acoustic pressure, 10: mean acoustic pressure, 11: acoustic linear distance, 12: number of acoustic peaks.

Fruit texture QTL discovery through Pedigree Based Analysis (PBA)

Each fruit texture parameter obtained from the phenotypic dissection of fruit texture was finally exploited in marker-trait association studies. In the attempt to map the QTLs related to this feature, the Bayesian approach was initially employed. QTLs were identified and mapped on thirteen chromosomes, on which the posterior QTL intensity exceeded the posterior

probability threshold (2lnBF>2; Table 1). The overall genome-wide QTL overview (Supplementary Fig. S3) distinguished specific probability profiles for the two groups of texture related parameters, acoustic and mechanical. For simplicity, the QTL differences are highlighted in Fig. 2, comparing the profiles of the maximum force (mechanical) with the number of acoustic peaks (acoustic).

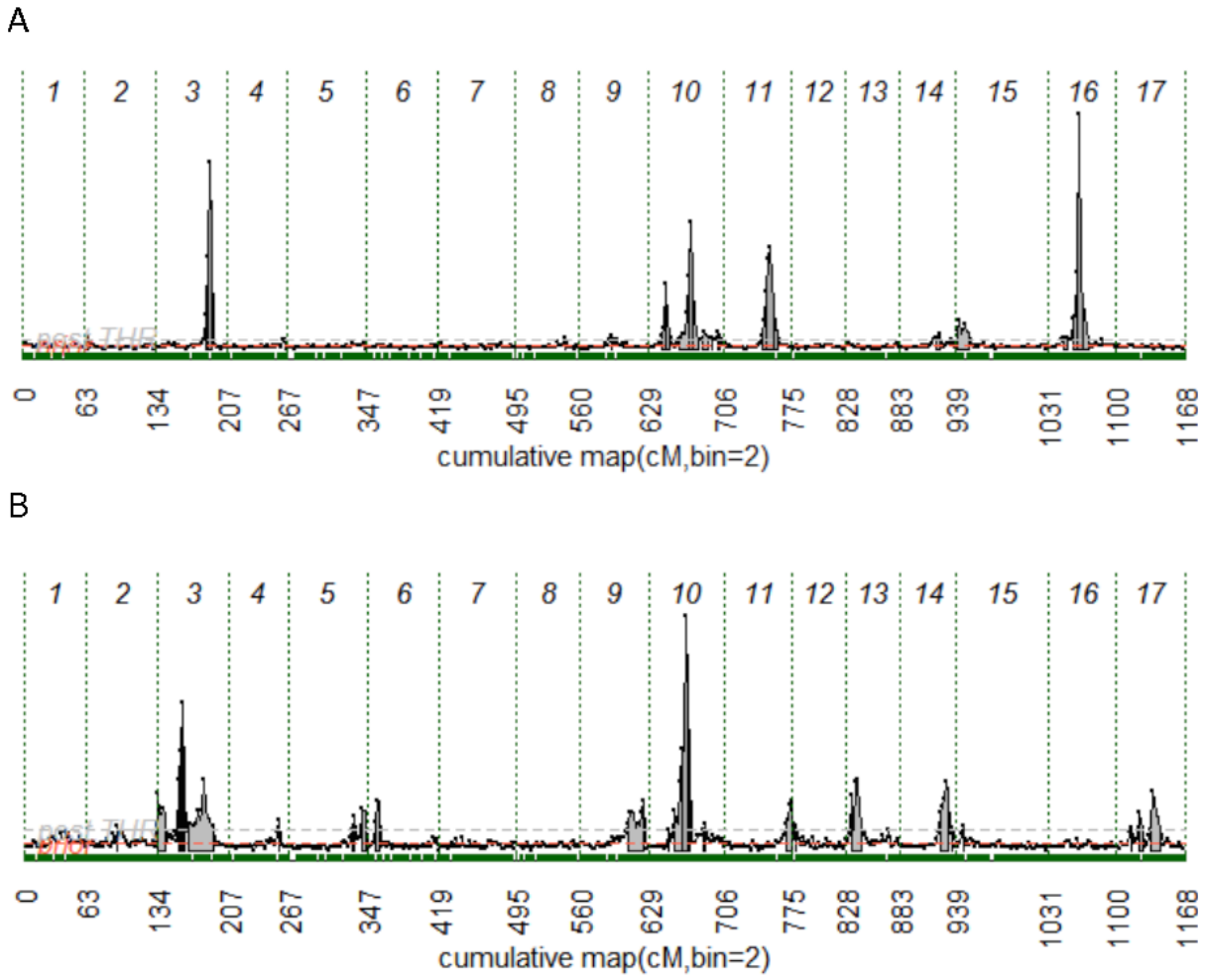


Figure 2. QTL probability pattern for the maximum force (panel A) and number of acoustic peaks (panel B). The grey area below each QTL profile indicates the 90% credible region associated to each estimated mean. For both panels, chromosomes are delimited by dashed vertical lines and numbered from 1 to 17. Below each figure the cumulative distance of the consensus genetic map is reported.

The chromosome 3 and 10 turned out to be in common between the two texture features, showing QTLs for both sub-traits (Fig. 2; Supplementary Fig. S3), with distinction. For the maximum force (Fig. 2A) the QTL mapped on chromosome 3 is located on a single genomic region (2lnBF1/0 = 13.6), with a mode at 55 cM and an allelic effect (AEt1) of 1.54 (Table 1). This position was also similar to the rest of the mechanical parameters, spanning from 55 to 57 cM (Supplementary Fig. S3 and Table 1).

The only difference was observed for the number of force peaks, which showed the QTL at 10 cM. This observation, however, additionally confirms the association of this parameter to the group of acoustic parameters. In the case of the acoustic peaks (Fig. 2B), two QTLs were instead suggested (2lnBF2/1 = 2.7), with a mode at 3 and 28 cM and an allelic effect of 10.55 and 9.39, respectively. Among the acoustic sub-traits, the QTL on chromosome 3 was also identified for the acoustic linear distance, but at 28 cM. On chromosome 10, a single QTL associated to the number of acoustic peaks (Fig. 2B) was shown (2lnBF1/0 = 10.6) and located at 40 cM with an allelic effect of 18.25. This position was also similar to the rest of the acoustic parameters (at 40 and 42 cM), beside the mean acoustic pressure that showed the QTL peak at 35 cM, as its was also observed for the number of force peaks (Table 1). For the maximum force, two QTLs were instead observed on this chromosome (2lnBF2/1 = 4.2). The first was located at 20 cM, with an allelic effect of 0.91, while the second was mapped at 45 cM with an AEt1 of 1.92. These two regions were also consistent across the mechanical parameters (spanning between 19 and 20 cM for the first QTL and 44-46 cM for the second) with two exceptions. The force linear distance showed in fact only one QTL (with a low probability and effect) at 49 cM, while for the Young's module (or elasticity module) no QTL was observed (Table 1 and Supplementary Fig. S3). Beside these, other QTLs were identified with a more specific pattern. The two major genomic intervals showing QTLs associated with mechanical parameters were mapped on chromosomes 11 and 16 (Fig. 2 and Supplementary Fig. S3). The QTLs positioned on chromosome 11, across the several mechanical parameters, were located between 41 and 49 cM (Table 1).

Table 1. For each trait assessed, the variance (Var), the mean (Mean) and the linkage group (LG) on which the QTL is identified are reported. For each QTL the interval, the length, the mode (in cM), the 2ln Bayes factor (BF) for the presence of 1 or 2 QTLs (1/0 and 2/1, respectively), the probability (Prob), the additive effect size (AEt1), the additive variance (AVt1) and the weighted Additive variance (wAVt1) are also shown. Each parameter refers to a haploblock, specified in the last column.

Trait	Var	Mean	LG	Interval	Length	Mode	2lnBF (1/0)	2lnBF (2/1)	Prob	AEt1	AVt1	wAVt1	Haplotype
Number of Acoustic Peaks	678.4	1.23	3	1-19	18	3	-	2.7	0.31	10.56	55.69	17.39	fp03_03
			3	21-60	39	28	6.2		0.98	9.39	42.86	42.26	fp03_28
			9	37-69	32	61	2.9		0.52	10.27	51.38	27.08	fp09_61
			10	21-64	43	40	10.6		1.10	18.25	142.51	157.11	fp10_40
			13	7-29	22	13	2.6		0.33	10.36	52.18	17.45	fp13_13
			14	39-56	17	48	2.8		0.46	16.48	134.27	62.55	fp14_47
Acoustic Linear Distance	1857766.77	79.35	1	23-63	40	41	3.9		0.75	618.22	185782.32	140368.87	fp01_36
			3	1-61	60	28	5.2		1.14	484.32	111028.38	127177.97	fp03_28
			10	27-45	18	40	29.4		1.02	1254.33	641967.08	658178.37	fp10_40
Mean Acoustic Pressure	2.12	0.1	1	23-63	40	36	10.7		1.30	0.99	0.45	0.59	fp01_36
			10	22-44	22	35	29.5		1.03	1.43	0.91	0.94	fp10_35
Maximum Acoustic Pressure	8.37	0.22	1	21-58	37	33	5.9		0.79	1.86	1.53	1.22	fp01_33
			6	1-34	33	10	2.9		0.62	1.18	0.68	0.42	fp06_10
			10	31-45	14	42	29.6		1.01	2.93	3.46	3.52	fp10_42
			11	51-70	19	67	2.7		0.51	2.06	2.02	1.03	fp11_67
Number of Force Peaks	0.08	0.02	1	17-63	46	35	7.5		1.07	0.16	0.01	0.01	fp01_35
			3	1-20	19	10	4.6		0.70	0.12	0.01	0.01	fp03_10
			5	62-74	12	71	2.5		0.42	0.17	0.01	0.01	fp05_71
			8	21-65	44	51	3.9		0.70	0.08	0.003	0.002	fp08_51
			9	36-68	32	57	2.2		0.42	0.09	0.004	0.002	fp09_57
			10	25-45	20	35	29.5		1.02	0.27	0.03	0.03	fp10_35

Trait	Var	Mean	LG	Interval	Length	Mode	2lnBF (1/0)	2lnBF (2/1)	Prob	AEt1	AVt1	wAVt1	Haplotype
Young's Module	0.05	0.02	1	11-43	32	26	4.5		0.62	0.06	0.002	0.001	fp01_26
			3	43-61	18	57	15.7		0.98	0.10	0.005	0.01	fp03_57
			4	1-32	31	8	3.4		0.32	0.09	0.004	0.001	fp04_07
			8	43-60	17	51	13.3		1.00	0.10	0.005	0.01	fp08_51
			11	1-32	31	14	7.9		1.07	0.12	0.01	0.01	fp11_15
Force Linear Distance	26.85	0.4	3	45-59	14	56	29.2		0.99	2.64	2.91	2.89	fp03_56
			8	32-54	22	50	6.2		0.88	1.73	1.45	1.29	fp08_50
			10	31-77	46	49	4.1		0.85	1.99	1.87	1.59	fp10_45
			11	34-59	25	41	12.6		1.00	2.69	3.61	3.64	fp11_41
			16	1-44	43	32	29.1	2.7	1.29	2.67	3.09	3.99	fp16_32
			16	60-69	9	65	-		0.08	1.44	1.02	0.09	fp16_65
Area	52001.19	17.2	3	47-58	11	56	29.6		0.99	114.80	5948.64	5924.70	fp03_56
			10	10-30	20	19	-	4.3	0.42	67.95	2289.38	962.74	fp10_19
			10	32-77	45	45	10.1		1.27	125.24	7219	9209.01	fp10_45
			11	37-57	20	49	16		1.01	135.67	9043.08	9170.45	fp11_49
			16	13-45	32	32	29.1	2.7	1.28	167.25	11316.16	14481.05	fp16_32
			16	51-68	17	62	-		0.09	55.53	1540.73	151.90	fp16_62
Mean Force	7.73	0.2	3	50-58	8	56	29.6		0.97	1.39	0.87	0.85	fp03_56
			10	12-26	14	19	-	4.2	0.39	0.81	0.33	0.13	fp10_19
			10	28-77	49	45	10.9		1.27	1.51	1.07	1.37	fp10_45
			11	34-60	26	47	13.7		1.04	1.66	1.37	1.42	fp11_47
			16	12-43	31	32	28.9	3.1	1.34	1.92	1.57	2.10	fp16_32
			16	56-69	13	69	-		0.07	0.67	0.23	0.02	fp16_68

Trait	Var	Mean	LG	Interval	Length	Mode	2lnBF (1/0)	2lnBF (2/1)	Prob	AEt1	AVt1	wAVt1	Haplotype
Final Force	9.77	0.21	3	38-60	22	55	15.7	5.5	1.02	1.28	0.78	0.80	fp03_56
			10	11-26	15	19	-		0.62	1.02	0.51	0.32	fp10_19
			10	28-53	25	44	11.1		1.02	2.29	2.28	2.33	fp10_43
			11	35-58	23	45	29.8		1.03	2.01	1.89	1.96	fp11_45
			16	13-59	46	32	28.4	4.2	1.53	2.27	2.25	3.44	fp16_32
			16	61-69	8	64	-		0.05	0.79	0.32	0.02	fp16_64
Maximum Force	10.95	0.22	3	41-63	22	55	13.6	4.2	1.02	1.54	1.09	1.12	fp03_56
			10	10-29	19	20	-		0.38	0.91	0.42	0.16	fp10_20
			10	31-77	46	45	10.5		1.20	1.92	1.68	2.02	fp10_45
			11	36-66	30	47	15.9		1.05	2.01	1.96	2.07	fp11_47
			16	10-44	34	32	28.8		1.38	2.43	2.46	3.40	fp16_32
Initial Force	7.1	0.17	3	48-59	11	56	29.5	3.3	0.97	1.33	0.80	0.78	fp03_56
			10	14-26	12	20	-		0.15	0.75	0.28	0.04	fp10_20
			10	30-77	47	46	7.4		1.33	1.17	0.68	0.91	fp10_46
			11	38-62	24	49	12.3		1.02	1.55	1.19	1.23	fp11_49
			15	1-19	18	4	2.9		0.54	1.07	0.56	0.31	fp15_04
			16	10-42	32	31	29.4		1.17	1.90	1.46	1.71	fp16_31

On this chromosome, it is also interesting to note the QTL positioned at 14 cM (2lnBF 7.9 and AEt1 of 0.11) and related to the Young's module. The different and original positioning of this QTL within the class of mechanical parameters can be due to the fact that the Young's module depends on the elasticity of the sample (ratio between stress and strain) rather than fruit firmness. The particular behavior observed for the Young's module, besides its projection over the PCA plot (Fig. 1B and D), is moreover validated by the QTL profile detected on chromosome 16. As aforementioned reported for chromosome 11, also this QTL is specifically associated to the mechanical parameters (Supplementary Fig. S3), with a mode located around 32 cM, exception made for the Young's module, where this QTL was not detected (similarly to chromosome 10).

In the second sub-set of QTLs (related to the acoustic parameters), several genomic regions located on chromosome 1 and associated to mean and maximum acoustic pressure, number of force peaks as well as the Young's module were identified. About force peaks, other QTLs were moreover mapped on chromosome 5, 8 and 9. This latter chromosome, together with 13 and 14, was also associated to the number of acoustic peaks (Fig. 2B and Supplementary Fig. S3 and Table 1). The simultaneous presence of QTLs detected on chromosome 9 for both number of force and acoustic peaks strengthens the relationship between these two parameters. Especially for the number of acoustic peaks, chromosome 9 and 14 resulted to be the most important, showing an allelic effect of 10.27 and 16.48, respectively (Table 1). In particular, the QTL on chromosome 14 is characterized by an estimated genotype (Fig. 3A) consistent with the acoustic performance (assessed as number of acoustic peaks) of the six parental cultivars (Fig. 3B and C). Among the group, 'Pinova' and 'Fuji' were distinguished by the highest acoustic response, as depicted in the 2D-PCA plot (Fig. 3B) and loading projection (Fig. 3C).

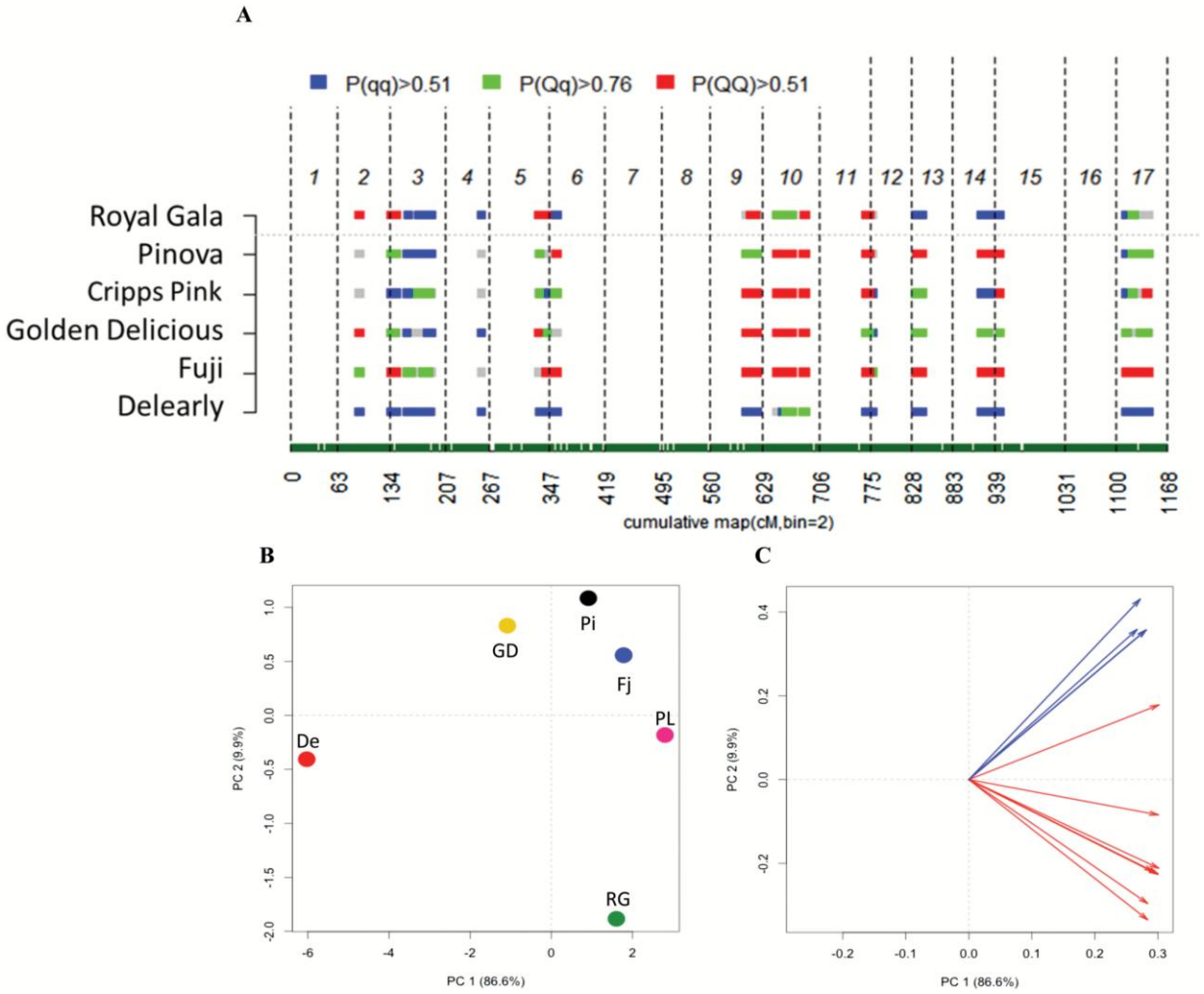


Figure 3. Posterior estimate of the QTL genotype probabilities computed for the number of acoustic peaks for each of the six parental cultivars (panel A). Each chromosome is indicated with numbers on the top, while on the bottom the cumulative genetic distance is reported. Each row represents a cultivar, named on the left side. Blue, green and red colored bars indicate 'QQ', 'Qq' and 'qq' QTL estimated genotype, respectively. In panel B is reported the 2D-PCA plot showing the distribution of the six parental cultivars on the base of their textural performance. Each cultivar is coded as follow, De: 'Delearly', RG: 'Gala', GD: 'Golden Delicious', PL: 'Cripps Pink', Pi: 'Pinova' and Fj: 'Fuji'. In panel C, instead, the variable projection is shown. Mechanical and acoustic parameters are depicted with red and blue colored arrows, respectively, according to Fig. 1.

The superior crispness performance of these two apple cultivars is also confirmed by the homozygous state of the positive estimated QTL allele ('QQ'). On the contrary, cultivars with a mealy texture, such as 'Delearly' and 'Royal Gala', are plotted on the other extreme of the 2D-PCA plot, showing a 'qq' genotype for this QTL. The effect of the estimated allele for the QTL on chromosome 14 was further investigated on the six progenies (Fig. 4). FlexQTL™ estimated a 'QQ' genotype for 'Fuji' and 'Pinova', a heterozygous 'Qq' genotype only for 'Golden Delicious' and a 'qq' genotype for the other three varieties ('Delearly', 'Royal Gala' and 'Cripps Pink'). Taking into account that the 'Q' allele is deputed to increase the phenotypic performance, it is

worth noting that only the seedlings of ‘FjPi’ and half of the progeny of ‘GDFj’ (distinguished by a ‘QQ’ genotype) were distinguished by the highest acoustic response (Fig. 4), underlying the role of this QTL in the control of the acoustic properties in apple.

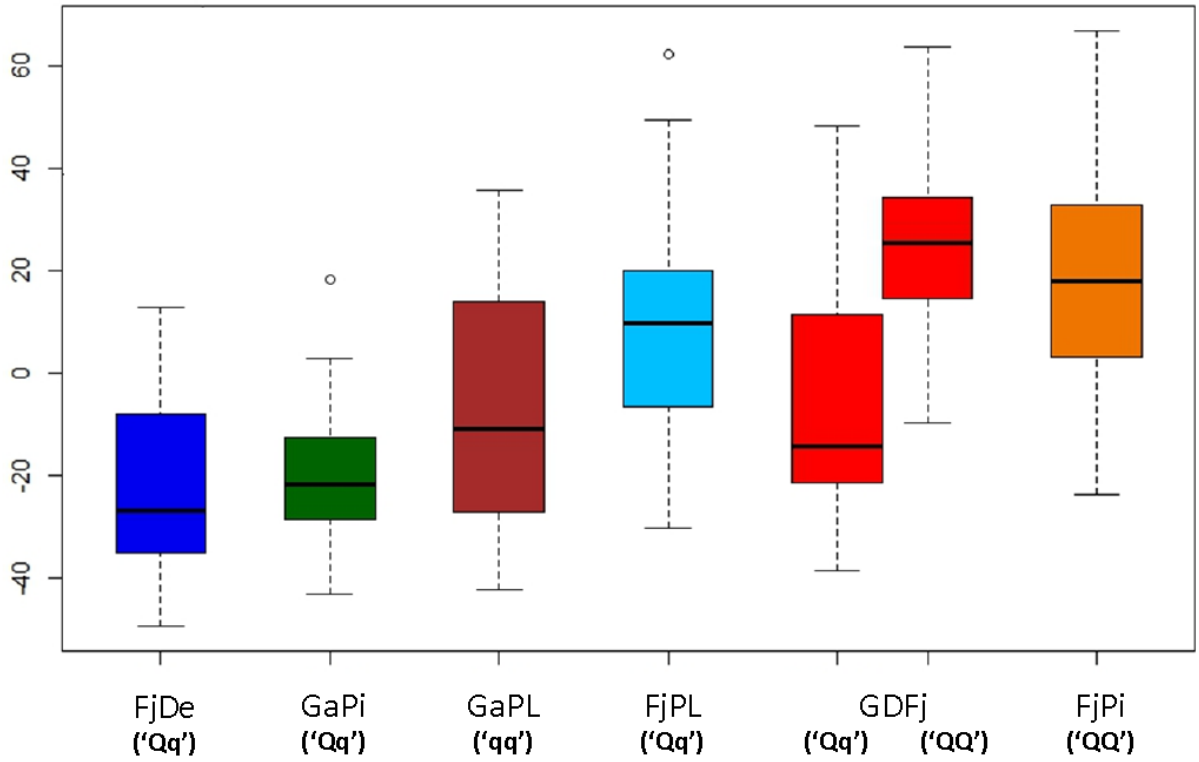


Figure 4. Box plot indicating the phenotype performance of the number of acoustic peaks for the six full-sib families. For each population the estimated genotype for the QTL mapped on chromosome 14 is also indicated. Each box comprehends the upper and lower quantile, with the median being represented by a horizontal solid line. Outliers are pointed by dots.

Analysis of the Linkage Disequilibrium in domesticated apple

The level of linkage disequilibrium (LD) was determined to verify the genetic associations between loci and to scan the LD decay over each chromosome. To this end, from the total set of 10.695 selected SNP markers, 7.378 were effectively employed in the computation, after subsequent filtering steps. From the initial dataset, besides SNPs with MAF<0.05, markers showing an inconsistent position between the physical location on the genome and the consensus genetic map were also excluded (Supplementary Fig. S4). In the germplasm collection investigated here (and represented by 387 apple accessions), the LD decay was estimated to extend for an average up to r^2=0.19 at the genome-wide level (Fig. 5A) corresponding to approximately 400 kb, and spanning from a a maximum of r2=0.28 for chromosome 16 (Fig. 5B) to a minimum of r2=0.13 for chromosome 17 (Fig. 5C).

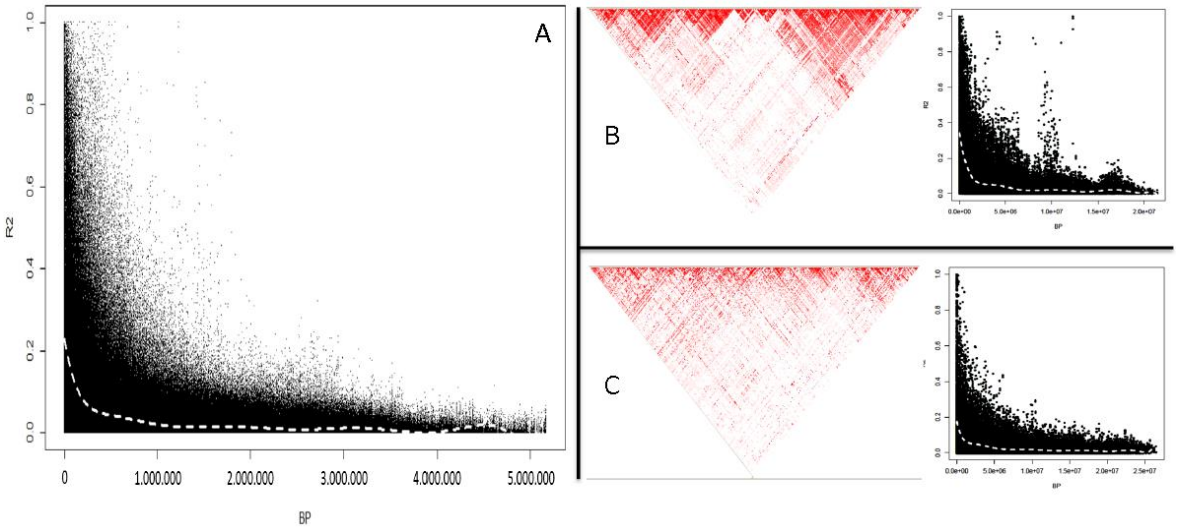


Figure 5. Linkage disequilibrium (LD) decay plot showing the r^2 value between all possible pair-wise marker combinations against their physical distance over the genome. The LD plot is shown at genome-wide level (panel A) as well as for chromosome 16 (panel B) and 17 (panel C). For these two, the pattern of LD was also depicted with Haploview heatmaps. For the LD decay plot, the white dashed line indicates the LOESS fitting curve.

In parallel, the presence of distinct LD blocks over the genome was illustrated with a LD heat-map (Supplementary Fig. S5), highlighting specific genetic fixation for each apple chromosome. Amongst all, chromosome 16 is characterized by the highest LD value, showing a LD block of 2675 Kb (Fig. 5B). These results indicate that the LD extent in apple is shorter than in peach (Micheletti et al., 2015) but larger than other species, e.g. grapevine (Myles et al., 2009).

Fruit texture genetic dissection by GWAS

The QTLs identified with the PBA approach were further complemented by GWAS. From the entire germplasm collection assessed to estimate the LD decay, 233 accessions were used for both population structure and marker-trait association. Individuals were assigned to three sub-populations (K=3; Supplementary Fig. S6) following the plateau criterion (Falush et al., 2007), the non-parametric Wilcoxon test (Rosenberg et al., 2002) and the ΔK method proposed by Evanno et al. (2005). Beyond this point, the mean log-likelihood values tend to a platform together with an increased standard deviation, which became clearly evident from K=5.

From the 20K SNPs present in the array, 10.558 were finally exploited in the GWAS computation, performed with the MLM model implemented in TASSEL v3.0. As first attempt to dissect the genetic control of apple fruit texture, the two principal components (PC1 and PC2) were initially implemented as phenotypic traits. In this case, the PC1 was considered to capture the overall texture variability, being the entire group of parameters oriented towards its

projection (Fig. 1C and D). PC2, instead, was employed to discriminate the mechanical from the acoustic subset of variables. The MLM module identified for PC1 a major QTL on chromosome 10 (Supplementary Fig. S7A) with ten markers exceeding the FDR corrected P-value threshold. The phenotypic variability explained by the markers spanned from 9 to 13% and resulted positioned on the consensus map between 42 and 47 cM (Table 2).

Table 2. SNPs exceeding the statistical threshold in the association analysis computed with PC1 (depicted in Supplementary Fig. S7A). For each SNP, the chromosome on which the marker is mapped (chr), the genetic position (cM), the P-value, the name (SNP) and the percentage of phenotypic variability explained (r^2) are indicated.

chr	cM	P-value	SNP	r^2
5	4.841	5.19E-06	FB_0597458_L5_PA	0.125730941
10	42.241	2.44E-06	FB_0832819_L10_41_1	0.134019865
10	42.241	4.31E-05	FB_0028781_L10_PA	0.117406088
10	42.241	4.40E-05	FEM_cg_19	0.10254861
10	42.843	2.60E-05	FEM_cg_9	0.108222257
10	42.843	2.60E-05	FB_0832811_L10_41_1	0.108222257
10	42.843	2.60E-05	FEM_cg_11	0.108222257
10	42.843	2.60E-05	FEM_cg_18	0.108222257
10	42.843	2.60E-05	FEM_cg_17	0.108222257
10	47.79775	1.68E-05	FB_0032582_L10_PA	0.094505374

Within this marker set it is worth to underline five SNPs, coded as FEM_cg_9, 11, 17, 18 and 19, which are custom SNPs specifically designed on polymorphisms discovered re-sequencing the full length of MdPG1, a gene playing a key role in the fruit softening process in apple (Wakasa et al., 2006; Costa et al., 2010b; Longhi et al., 2012). Especially FEM_cg_19, further named Md-PG1SNP (Baumgartner et al., 2016), is a SNP highly correlated with the microsatellite marker Md-PG1SSR10kd, previously associated with the fruit texture behavior in apple (Longhi et al., 2013). On the contrary, when PC2 was implemented into the association analysis not a single SNP was identified as statistically associated (Supplementary Fig. S7B). It is moreover worth noting that when single texture sub-traits were used as phenotype, the association result was consistent with the profile obtained for PC1, as shown in Supplementary Fig. S7 (C and D) for the maximum force and number of acoustic peaks, respectively. Although the two groups of variables are oriented towards two different PCA quadrants they are however commonly projected along the PC1 orientation (Fig. 1D). Within the panel of apple accessions employed here, PC1 accounts for 79.6% of the total phenotypic variance, influencing thus the genetic association of each single texture parameter. PC2, instead, is orthogonally oriented with respect to PC1 and more related to the difference between the two groups of variables, therefore more effective in the dissection of the genetic control of the two texture signatures.

To decipher more specifically the genetic regulation of fruit texture properties, a second round of association was performed. Since it is already well known that crisp apples are more appreciated by consumers, from the initial set of accessions used in GWAS, genotypes distinguished by the unfavorable homozygous allelic configuration for MdPG1 and MdACO1 were removed. The effect of these two genes on the fruit texture in apple largely depends on the interaction of the physiological processes they control. MdPG1 is involved in the dismantling process of the cell wall/middle lamella structure (Brummell and Harpster, 2001; Brummell, 2006) and its effect in apple seems to be more relevant than in other climacteric species. In tomato, in fact, the role of this gene alone does not impact significantly the fruit texture physiology (Sheehy et al., 1988; Smith et al., 1988; Giovannoni et al., 1989). The other gene, MdACO1, regulates the last step of the ethylene biosynthesis (Bleecker and Kende, 2000). Although in climacteric fruit the amount of this hormone is known to control several processes (Rose et al., 1998; Costa et al., 2005; Wakasa et al., 2006), it has been also proposed the coexistence of ethylene –dependent and –independent regulation of fruit texture, as demonstrated in melon (Nishiyama et al., 2007) as well as in apple (Tadiello et al., 2016). This dual mechanism can explain why QTLs for fruit firmness have been collocated with the MdPG1 but not MdACO1. To investigate the consistency between the QTL genotypes estimated by FlexQTL™ and the expression of a gene included in the corresponding genomic interval, the transcript accumulation of MdPG1, together with MdACO1, was assessed. The transcript profiling was carried out in two groups of seedlings chosen between ‘FjDe’ and ‘FjPL’ populations. Among the four individuals selected in the first population, two (FjDe_10 and FjDe_14) were characterized by a ‘QQ’ genotype estimated for the QTL on chromosome 10 (and coincident with the genetic position of MdPG1), while the other two (FjDe_45 and FjDe_125) were distinguished by a ‘Qq’ genotype. These two QTL genotypes are moreover consistent with the allelotype configuration of Md-PG1SNP. The ‘Q’ and ‘q’ alleles are in fact linked to the ‘C’ and ‘T’ allele of this marker, in agreement with the genotype of the parental lines. ‘Fuji’ (‘QQ’ estimated genotype) is indeed distinguished by a ‘CC’ allelic state for Md-PG1SNP, while ‘Delearly’ has an ‘TC’ allelotype. The ‘T’ allele (related to ‘q’ QTL allele) segregates therefore within the ‘FjDe’ population, contributing to a loss of fruit firmness. This association is further validated by the Pearson correlation value (R^2: -0.8) between the MdPG1 expression and the fruit firmness assessment depicted in Fig. 6.

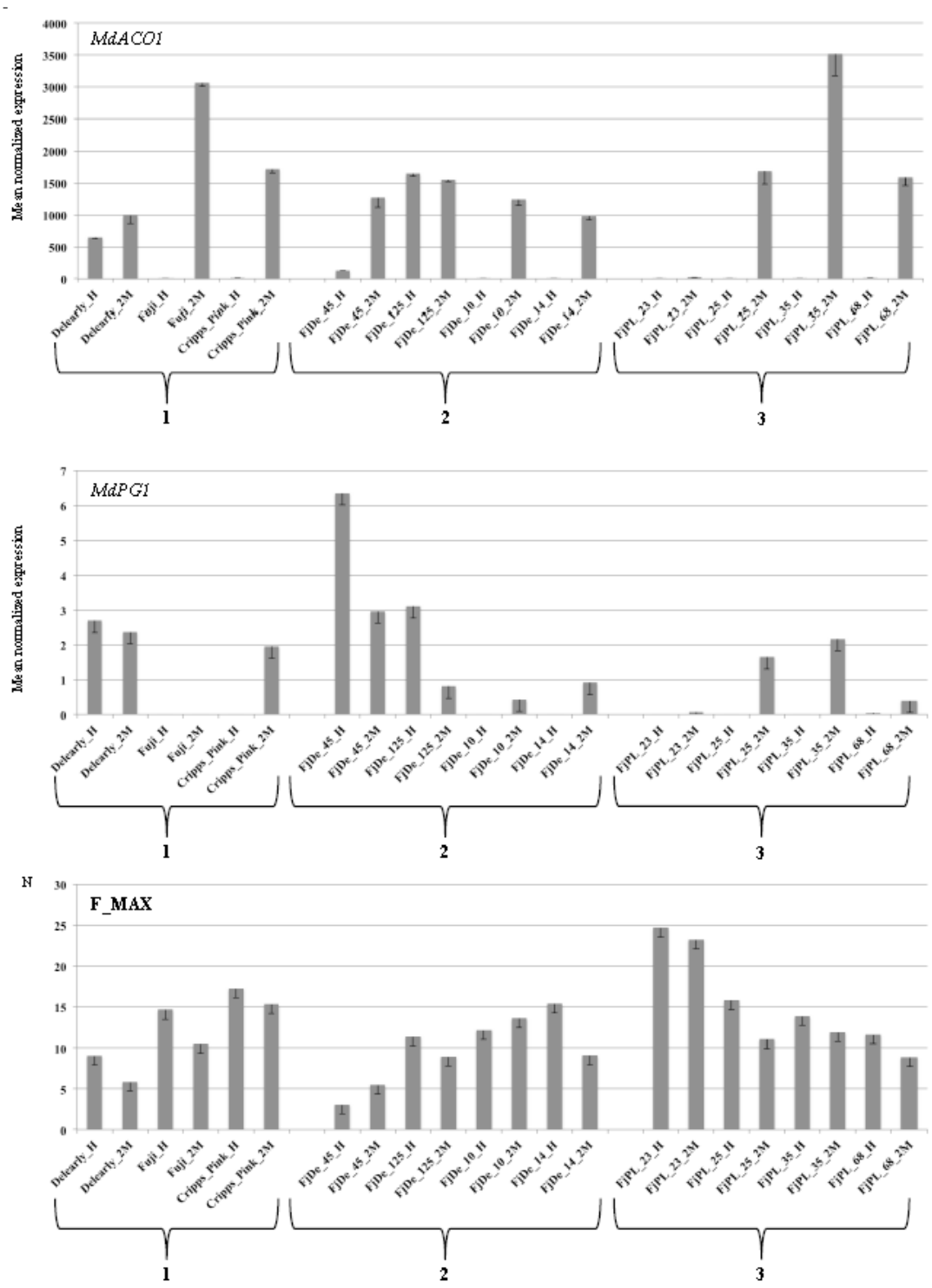


Figure 6. MdACO1 and MdPG1 expression profile and fruit firmness assessment. The transcript accumulation together with the phenotype measurement have been performed in three groups of plant materials indicated by numerical codes as 1: parental lines, 2: four individuals of the 'FjDe' population and 3: four individuals of the 'FjPL' population. Each genotype is also distinguished by two samples, H: harvest, PH: postharvest. On the first two panels (referring to the MdACO1 and MdPG1 expression) the y-axes indicate the mean normalized gene expression, while for the last one (maximum force) Newton (N) is instead reported.

In both groups of parental lines (group 1 in Fig. 6) and 'FjDe' individuals (group 2) it is clear that high loss of firmness corresponds to high MdPG1 expression. The genotypes

distinguished by a ‘q’ allele (FjDe_45, FjDe_125 as well as the parental variety ‘Delearly’) are in fact characterized by a considerable MdPG1 expression already at harvest. This observation was additionally confirmed by the analysis carried out on the second population, ‘FjPL’. Since the two parents (‘Fuji’ and ‘Cripps Pink’) do not segregate for this QTL, all the four seedlings (group 3) are characterized by a ‘QQ’ estimated genotype (MdPG1SNP_CC). Due to the role exerted by these two genes, the breeding activities oriented towards the selection of firm and crisp apple do not longer consider cultivars with these unfavorable genotypes as parental lines. Although this second panel of accessions is composed by only 64 individuals, it captures the real phenotypic variability nowadays used by breeders. This second GWAS analysis was carried out with a general linear model (GLM), selected on the base of the Q-Q plot inspection. The result of this re-shaped phenotypic variance, obtained by fixing the effect of the two loci, was evident in the association analysis depicted in Fig. 7 and computed for the maximum force (Fig. 7A) as well as the number of acoustic peaks (Fig. 7B).

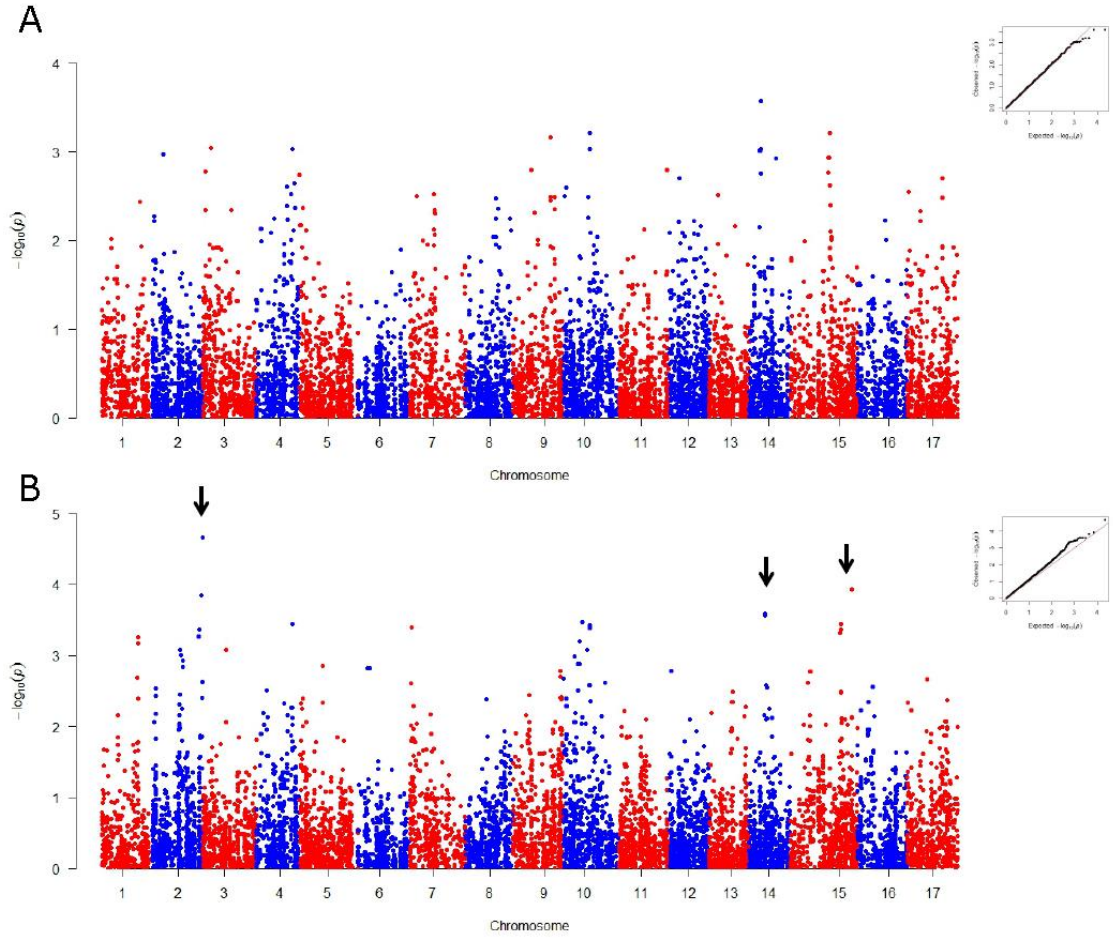


Figure 7. Manhattan plot illustrating the association between SNP markers and two texture sub-traits, maximum force (A) and number of acoustic peaks (B), computed in the selected panel of apple accessions. X and y axes report the number of chromosomes and the $-\log_{10}$(P-value), respectively. For both panels also the Q-Q plot is reported.

In both associations, no major QTL on chromosome 10 was detected, especially in the case of the mechanical parameter. On the contrary, when the number of acoustic peaks was considered, other regions were identified and located on chromosomes 2, 14 and 15 (Fig. 7B and Supplementary Tab. S6). These results provide evidences about the distinct genetic control for the two texture properties in apple, and suggest the role of chromosome 14 in the determination of acoustic properties, as underlined by the PBA results.

QTl anchoring and in silico gene annotation

To further investigate the role of the QTL mapped on chromosome 14 in the regulation of the acoustic component of fruit texture, an in silico searching and annotation of the candidate genes included in the genomic interval, together with their transcription profiling, was carried out. Genes were searched and annotated within an interval of 400kb (LD block) up and down-stream from the QTL peak determined by both PBA and GWAS approach (Table 3). Among them it is important to highlight important categories. The first is represented by Proline-Rich Proteins (PRPs). This class of cell-wall modifying proteins (CWMP) presents proline and hydroxyproline peptides and seems to be involved in the cell wall metabolism of several species, such as cotton, carrot and Arabidopsis (John and Keller, 1995; Fowler et al., 1999; Holk et al., 2002). It is also interesting to note that PRP-related genes are expressed in immature watermelon fruit (Guo et al., 2011), therefore not related with the late ripening dismantling process of the cell wall. Another important CWMP class is represented by expansin, a type of protein involved in the architectural re-modeling of the cell wall operating a disruption of the non-covalent bounds between hemicellulose matrix and cellulose microfibrils (Cosgrove, 2000), exposing the cell wall structural polymers to the action of other cell wall modifying proteins. Within the QTL computed with GWAS, a xylanase (xylo-glucangalactosyltransferase) was also found. Heteroxylan are a divergent group of polymers contributing to the cell wall structure, although in dicots are less abundant than xyloglucan (Johnston et al. 2013). Xylanase are thus involved in the re-assessment of the cellulose-non-cellulosic framework and their role has been already reported in fruit. This gene was in fact expressed during the fruit softening process in papaya, according to an ethylene dependent pattern (Manenoi and Paull, 2007). In the end, it is also worth mentioning the identification of a Fucosyltransferase, an enzyme involved in the xyloglucan metabolism, as documented in Arabidopsis (Roche et al., 2016). To further assess the mode of action of these genes in the regulation of fruit texture, the whole transcriptomic survey presented by Tadiello et al. (2016) was re-examined. In this particular case, the expression profile of two expansins (MDP0000423907 and MDP0000193025) and a fucosyltransferase (MDP0000230681) was selected from the analysis carried out with the WGAA array. The expression analysis of these elements was retrieved form three specific

samples of the reference apple cultivar ‘Golden Delicious’, at harvest and after one week of postharvest shelf-life ripening at both normal and 1-MCP treated condition (Supplementary Fig. S8). As described in Tadiello et al. (2016), the control postharvest sample is characterized by an important loss of acoustic performance with regards to harvest, while the application of the ethylene competitor (1-MCP) effectively limited the cell wall degradation. Interestingly, the genes here identified and associated to the regulation of the acoustic component of texture are not stimulated by ethylene and do not participate to the major cell wall degrading events, since their expression do not change from harvest to postharvest. One expansin, in particular (MDP 0000423907), is induced when the acoustic performance is promoted (1-MCP treatment), meaning that its role rather than on the dismantling process leading to softening is more involved in the maintenance of the cell wall architectural structure.

Table 3. Gene annotation within the QTL interval mapped on chromosome 14 and identified through both GWAS (upper part) and PBA (lower part) approaches. For each gene is reported the NCBI and GDR gene ID, gene function and conting coordinates. In addition, the closest SNP with its relative genomci information is also provided.

Gene ID	Gene Function	MDC Name	MDC Start	MDC End	Flanking SNP	MDC Name	MDC Start	MDC End
gi|658000264|ref|XM_008394357.1 | MDP0000126350	Proline-rich protein	MDC011541.208	17995930	8011488	FB_0240837_L14_PA	MDC018357.590	17637582	17670933
gi|658000266|ref|XM_008394359.1 | MDP0000155446	Proline-rich protein	MDC013704.625	18033491	8049518	FB_0240837_L14_PA	MDC018357.590	17637582	17670933
gi|658000211|ref|XM_008394331.1 | MDP0000423907	Pistil-specific extensin-like protein	MDC010624.545	16734875	6765614	RB_20380789_L14_41_1	MDC021233.163	17000850	17040029
gi|658000230|ref|XM_008394341.1 | MDP0000775334	Xylo-glucangalactosyltransferase	MDC007808.273	17603144	7620726	RB_21013914_L14_PA	MDC018357.590	17637582	17670933
gi|658063270|ref|XM_008369329.1 | MDP0000151618	Expansin-A1	MDC015296.157	25922508	5924244	FB_0254699_L14_PA	MDC012426.235	26135397	26155901
gi|658001413|ref|XM_008394948.1 | MDP0000193025	Expansin-A1	MDC009270.341	25922966	5925030	FB_0254699_L14_PA	MDC012426.235	26135397	26155901
gi|658001426|ref|XM_008394954.1 | MDP0000230681	Fucosyltransferase	MDC014060.237	26024637	6027504	FB_0254699_L14_PA	MDC012426.235	26135397	26155901

Conclusion

Fruit texture in apple is composed by multiple sub-traits, most of which (especially the acoustic ones) are poorly investigated, in particular for genetic purpose. So far, this limitation led to the identification of markers suitable to assist the selection of fruit firmness only, although, for apple, the most preferred feature by consumers is crispness. The coupling of PBA with GWAS enabled the genetic deciphering of the fruit texture control, identifying important QTLs associated to both texture features. The comparison of the results obtained by the two genetic approaches highlighted an inventory of genomic intervals specifically associated to mechanical and acoustic parameters, respectively, hypothesizing as these sub-traits are effectively controlled by different genetic mechanisms. In a near future, in the new high-quality breeding materials, the alleles of the markers currently in use will be quickly fixed as a result of recurrent rounds of marker informed parental selection (MAPS marker assisted parent selection) and the subsequently assisted selection of seedlings (MASS – marker assisted seedling selection). The information presented here can therefore be taken into consideration to design novel markers useful to identify novel apple accessions distinguished by superior fruit crispness.

Supplementary data

Supplementary data can be found at:

https://drive.google.com/drive/folders/0B1u0mN44Aw3pZ05IZG9XQW9yY2M?usp=sharing

Acknowledgment

This work was co-funded by the EU seventh Framework Programme by the FruitBreedomics Project No. 265582: Integrated Approach for increasing breeding efficiency in fruit tree crops (www.FruitBreedomics.com). The views expressed in this work are the sole responsibility of the authors and do not necessary reflect the views of the European Commission. The authors are also grateful to Marco Fontanari for participating in the phenotyping activity, Marco Moretto for bioinformatics assistance and Elisa Banchi for SNP genotyping. This work was also supported by the EPS PhD School of Wageningen University and the GMPF program of FEM.

References

Allard A, Bink MCAM, Martinez S, Kelner J-J, Legave J-M, di Guardo M, Di Pierro EA, Laurens F, van de Weg EW, Costes E. 2016. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. Journal of Experimental Botany, doi:10.1093/jxb/erw130.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics **21**, 263–265.

Baumgartner IO, Kellerhals M, Costa F, Dondini L, Pagliarani G, Gregori R, Tartarini S, Leumann L, Laurens F, Patocchi A. 2016. Development of SNP-based assays for disease resistance and fruit quality traits in apple (Malus × domestica Borkh.) and validation in breeding pilot studies. Tree Genetics & Genomes **12**, 35.

Ben Sadok I, Tiecher A, Galvez-Lopez D, Lahaye M, Lasserre-Zuber P, Bruneau M, Hanteville S, Robic R, Cournol R, Laurens F. 2015. Apple fruit texture QTLs: year and cold storage effects on sensory and instrumental traits. Tree Genetics & Genomes **11**:119

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289–300.

Bianco L, Cestaro A, Sargent DJ, et al. 2014. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh.) (P Jaiswal, Ed.). PLoS ONE **9**, e110377.

Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE. 2008. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica **161**, 85–96.

Bink MC a M, van Eeuwijk F a. 2009. A Bayesian QTL linkage analysis of the common dataset from the 12th QTLMAS workshop. BMC proceedings **3 Suppl 1**, S4.

Bink MCAM, Jansen J, Madduri M, et al. 2014. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **127**, 1073–90.

Bleecker AB, Kende H. 2000. Ethylene: A Gaseous Signal Molecule in Plants. Annual Review of Cell and Developmental Biology **16**, 1–18.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics **23**, 2633–2635.

Breseghello F, Sorells ME. 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics **172,** 1165–1177.

Brummell D a. 2006. Cell wall disassembly in ripening fruit. Functional Plant Biology **33**, 103–119.

Brummell D, Harpster M. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. In: Carpita, N.C. and Campbell, M. and Tierney M, ed. Plant Cell Walls. Springer Netherlands, 311–340.

Busatto N, Farneti B, Tadiello A, Velasco R, Costa G, Costa F. 2015. Candidate gene expression profiling reveals a time specific activation among different harvesting dates in 'Golden Delicious' and 'Fuji' apple cultivars. Euphytica **208,** 401-413.

Cardoso S, Lau W, Eiras Dias J, Fevereiro P, Maniatis N. 2012. A candidate-gene association study for berry colour and anthocyanin content in vitis vinifera L. PLoS ONE **7.**

Chagne´ D, Dayatilake D, Diack R, Oliver M, Ireland H, Watson A, Gardiner SE, Johnston JW, Schaffer RJ, Tustin S. 2014. Genetic and environmental control of fruit maturation, dry matter and firmness in apple (Malus × domestica Borkh.). Horticulture Research **1,** 14046; doi:10.1038/hortres.2014.46

Cobb JN, DeClerck G, Greenberg A, Clark R, 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.

Cosgrove DJ. 2000. Loosening of plant cell walls by expansins. Nature **407**, 321–326.

Costa F, Stella S, Van De Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S. 2005. Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus × domestica Borkh.). Euphytica **141**, 181–190.

Costa F, Alba R, Schouten H, Soglio V, Gianfranceschi L, Serra S, Musacchi S, Sansavini S, Costa G, Fei ZJ, Giovannoni J. 2010a. Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. BMC Plant Biol. **10**: 229.

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, Van de Weg WE. 2010b. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus × domestica Borkh.). Journal of experimental botany **61**, 3029–39.

Costa F, Cappellin L, Longhi S, et al. 2011. Assessment of apple (Malus × domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biology and Technology **61**, 21–28.

Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F. 2012. Texture dynamics during postharvest cold storage ripening in apple (Malus × domestica Borkh.). Postharvest Biology and Technology **69**, 54–63.

Di Guardo M, Tadiello A, Farneti B, Lorenz G, Masuero D, Vrhovsek U, Costa G, Velasco R, Costa F. 2013. A multidisciplinary approach providing new insight into fruit flesh browning physiology in apple (Malus × domestica Borkh.) PlosONE **8**,10

Di Guardo M, Micheletti D, Bianco L, et al. 2015. ASSIsT: An Automatic SNP ScorIng Tool for in- and outbreeding species. Bioinformatics, btv446.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4**, 359–361.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky K V., St. Clair JB, Neale DB. 2009. Association genetics of coastal Douglas fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. Cold-hardiness related traits. Genetics **182**, 1289–1302.

Eriksson EM. 2004. Effect of the Colorless non-ripening Mutation on Cell Wall Biochemistry and Gene Expression during Tomato Fruit Development and Ripening. Plant Physiology **136**, 4184–4197.

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

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. Molecular Ecology Notes **7**, 574–578.

Fowler TJ, Bermnhardt C, Tierney ML. 1999. Characterization and expresion of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. Plant

Physiology, **121**, 1081-1092.

Fresnedo-Ramírez J, Bink MC a. M, van de Weg E, Famula TR, Crisosto CH, Frett TJ, Gasic K, Peace CP, Gradziel TM. 2015. QTL mapping of pomological traits in peach and related species breeding germplasm. Molecular Breeding **35**, 166.

Fresnedo-Ramírez J, Frett TJ, Sandefur PJ, et al. 2016. QTL mapping and breeding value estimation through pedigree-based analysis of fruit size and weight in four diverse peach breeding programs. Tree Genetics & Genomes **12**, 25.

Gianola D, Sorensen D. 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Statistics for Biology and Health. sl: Springer.

Giovannoni JJ, DellaPenna D, Bennett AB, Fischer RL. 1989. Expression of a chimeric polygalacturonase gene in transgenic rin (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. Plant Cell **1**, 53–63

Giovannoni J. 2001. Molecular biology and fruit maturation and ripening. Annual Review of Plant Physiology and Plant Molecular Biology **52**, 725–749.

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. 2007. Association genetics in Pinus taeda L. I. wood property traits. Genetics **175**, 399–409.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB. 2008. Association genetics in Pinus taeda L. II. Carbon isotope discrimination. Heredity **101**, 19–26.

Guo S, Liu J, Zheng Y, Huang M, Zhang H, Gong G, et al. 2011. Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. BMC genomics, **12**, 454.

Harada T, Sunako T, Wakasa Y, Soejima J, Satoh T, Niizeki M. 2000. An allele of the 1-aminocyclopropane-1-carboxylate synthase gene (Md-ACS1) accounts for the low level of ethylene production in climacteric fruits of some apple cultivars. Theoretical and Applied Genetics **101**, 742–746.

Holk A, Klumpp L, Scherer GF. 2002. A cell wall protein down-regulated by auxin suppresses cell expansion in Daucus carota (l). Planta Mol. Biol. **50**, 295-305.

Janssen BJ, Thodey K, Schaffer RJ, et al. 2008. Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. BMC plant biology **8**, 16.

Jimenez-Bermudez S, Redondo-Nevado J, Mun˜ oz-Blanco J, Caballero JL, Lo´ pez-Aranda JM, Valpuesta V, Pliego-Alfaro F, Quesada MA, Mercado JA. 2002. Manipulation of strawberry fruit softening by antisense expression of a pectate lyase gene. Plant Physiology **128**, 751–759.

John ME, Keller G. 1995. Characterization of mRNA for a proline-rich protein of cotton fier. Plant Physiology, **108**, 669-676.

Johnston SL, Prakash R, Chen NJ, Kumagai MH, Turano MH, et al. 2013. An enzyme activity capable of endotransglycosylation of heteroxylan polysaccharides is present in plant primary cell walls. Planta, **237**, 173-187.

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

King GJ, Maliepaard C, Lynn JR, et al. 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.

Kramer MG, Redenbaugh K. 1994. Commercialization of a tomato with an antisense polygalacturonase

gene: The FLAVR SAVR tomato story. Euphytica **79**, 293–297.

Kumar S, Garrick DJ, Bink MC, Whitworth C, Chagné D, Volz RK. 2013. Novel genomic approaches unravel genetic architecture of complex traits in apple. BMC Genomics **14**, 393.

Kumar S, Rowan D, Hunt M, Chagné D, Whitworth C, Souleyre E. 2015. Genome-wide scans reveal genetic architecture of apple flavour volatiles. Molecular Breeding **35**.

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

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus × domestica Borkh.). Journal of Experimental Botany **63**, 1107–1121.

Longhi S, Hamblin MT, Trainotti L, Peace CP, Velasco R, Costa F. 2013. A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (Malus × domestica Borkh.). BMC plant biology **13**, 37.

Manenoi A, Paull RE. 2007. Papaya fruit softening, endoxylanase gene expression, protein and activity. Physiologia Plantarum **131**, 470-480.

Marı n-Rodriguez MC, Smith DL, Manning K, Orchard L, Seymour GB. 2003. Pectate lyase gene expression and enzyme activity in ripening banana fruit. Plant Molecular Biology **51**, 851–857.

Matas AJ, Gapper NE, Chung MY, Giovannoni JJ, Rose JK. 2009. Biology and genetic engineering of fruit maturation for enhanced quality and shelf-life. Current Opinion in Biotechnology **20**, 197–203.

Micheletti D, Dettori MT, Micali S, et al. 2015. Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. PLoS ONE **10**, 1–19.

Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A. 2008. Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiologia Plantarum **134**, 237–245.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. the Plant Cell Online **21**, 2194–2202.

Neale DB, Savolainen O. 2004. Association genetics of complex traits in plants. Trends in Plant Science **9**, 1360–1385.

Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. Nature reviews. Genetics **12**, 111–122.

Nishiyama, K., Guis, M., Rose, J.K.C. et al. 2007. Ethylene regulation of fruit softening and cell wall disassembly in Charentais melon. J. Exp. Bot. **58**, 1281–1290.

Oraguzie NC, Iwanami H, Soejima J, Harada T, Hall A. 2004. Inheritance of the Md-ACS1 gene and its relationship to fruit softening in apple (Malus × domestica Borkh.). Theoretical and Applied Genetics **108**, 1526–1533.

Peleman JD, Van Der Voort JR. 2003. Breeding by design. Trends in Plant Science **8**, 330–334.

Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Van de Weg E. 2013. 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree Genet Genomes. **10**:261-271. DOI 10.1007/s11295-013-0679-8

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus

genotype data. Genetics **155**, 945–959.

Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics **81**, 559–575.

Rafalski JA. 2010. Association genetics in crop improvement. Current Opinion in Plant Biology **13**, 174–180.

Rocha J, Cicéron F, Lerouxel O, Breton C, de Sanctis D. 2016. The galactoside 2- α-L-fucosyltransferase FUT1 from Arabidopsis thaliana: crystallization and experimental MAD phasing. Acta Cryst **F72**, 564-568.

Rose J, Hadfield K, Labavitch J, Bennett A. 1998. Temporal sequence of cell wall disassembly in rapidly ripening melon fruit. Plant physiology **117**, 345–61.

Rose JKC, Braam J, Fry SC, Nishitani K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant and Cell Physiology **43**, 1421–1435.

Rose JKC, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS. 2004. Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant Journal **39**, 715–733.

Rosenberg N a, Rosenberg N a, Pritchard JK, et al. 2002. Genetic structure of human populations. Science (New York, N.Y.) **298**, 2381–2386.

Rosyara UR, Bink MC a M, van de Weg E, Zhang G, Wang D, Sebolt A, Dirlewanger E, Quero-Garcia J, Schuster M, Iezzoni AF. 2013. Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. Molecular Breeding **32**, 875–887.

Segonne SM, Bruneau M, Celton J-M, Le Gall S, Francin-Allami M, Juchaux M, Laurens F, Orsel M, Renou J-P. 2014. Multiscale investigation of mealiness in apple: an atypical role for a pectin methylesterase during fruit maturation. BMC plant biology **14**, 375.

Sitrit Y, Bennett AB. 1998. Regulation of Tomato Fruit Polygalacturonase mRNA Accumulation by Ethylene: A Re-Examination. Plant Physiology **116**, 1145–1150.

Soglio V, Costa F, Molthoff JW, Weemen-Hendriks WMJ, Schouten HJ, Gianfranceschi L. 2009. Transcription analysis of apple fruit development using cDNA microarrays. Tree Genetics & Genomes **5**, 685–698.

Sheehy R, Kramer M, Hiatt W. 1988. Reduction of polygalacturonase activity in tomato fruit by antisense RNA. Proc. Natl. Acad. Sci. **85**, 8805–9

Sosnowski O, Charcosset A, Joets J. 2012. Biomercator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. Bioinformatics **28**, 2082–2083.

Smith C, Watson C, Ray J, Bird C, Morris P, et al. 1988. Antisense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. Nature **334**, 724–26

Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, Piepho HP. 2009. Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theoretical and Applied Genetics **118**, 259–273.

Szczesniak AS. 2002. Texture is a sensory property. Food Quality and Preference **13**, 215–225.

Tadiello A, Longhi S, Moretto M, Ferrarin A, et al., 2016. Interference with ethylene perception at receptor level sheds light on auxin and transcriptional circuits associated with the climacteric ripening of apple fruit (Malus × domestica Borkh.The Plant Journal, doi: 10.1111/tpj.13306

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet **28**, 286–289.

Van Ooijen JW. 2006. JoinMap®4, Software for the calculation of genetic linkage mapsin experimental populations. Kyazma B.V., Wageningen, Netherlands.

Velasco R, Zharkikh A, Affourtit J, et al. 2010. The genome of the domesticated apple (Malus × domestica Borkh.). Nature Genetics **42**, 833–839.

Vicente AR, Powell A, Greve LC, Labavitch JM. 2007. Cell wall disassembly events in boysenberry (Rubus idaeus L. x Rubus ursinus Cham. & Schldl.) fruit development. Functional Plant Biology **34**, 614–623.

Vincent JF V. 1998. The quantification of crispness. Journal of the Science of Food and Agriculture **78**, 162–168.

Voorrips RE, Bink MCAM, Kruisselbrink JW, Koehorst- van Putten HJJ, van de Weg WE. 2016. PediHaplotyper: Software for consistent assignment of marker haplotypes in pedigrees. Mol Breeding. In Press

Voorrips RE, Bink MCAM, van de Weg WE. 2012. Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. J. Hered. **103(6),** 903-7.

Wakasa Y, Kudo H, Ishikawa R, Akada S, Senda M, Niizeki M, Harada T. 2006. Low expression of an endopolygalacturonase gene in apple fruit with long-term storage potential. Postharvest Biology and Technology **39**, 193–198.

Weber AL, Briggs WH, Rucker J, Baltazar BM, Sanchez-Gonzalez JDJ, Feng P, Buckler ES, Doebley J. 2008. The genetic architecture of complex traits in teosinte (Zea mays ssp. parviglumis): New evidence from association mapping. Genetics **180**, 1221–1232.

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DA, Ordovas JM, Buckler ES. 2010. Mixed linear model approach adapted for genomewide association studies. Nat Genet **42,** 355-60.

Zdunek A, Cybulska J, Konopacka D, Rutkowski K. 2010. New contact acoustic emission detector for texture evaluation of apples. Journal of Food Engineering **99**, 83–91.

Chapter 5

Genome-wide association study unravels the genetic control of the apple volatilome and its interplay with fruit texture

Brian Farneti[1], Mario Di Guardo[1,2], Iuliia Khomenko[1,3], Luca Cappellin[1], Franco Biasioli[1], Riccardo Velasco[1], Fabrizio Costa[1]

[1]Research and Innovation Centre, Fondazione Edmund Mach, via Mach 1, 38010 San Michele all'Adige, Trento (Italy)

[2]Graduate School Experimental Plant Sciences, Wageningen University, PO Box 386, 6700 AJ Wageningen, The Netherlands

[3]Institute for Ion Physics and Applied Physics, University of Innsbruck, Technikerstr. 25/3, 6020 Innsbruck, Austria

Published in Journal of Experimental Botany (2017) 68 (7): 1467-1478. doi: 10.1093/jxb/erx018

Abstract

Fruit quality represents a fundamental aspect guiding consumers' preferences. Amongst apple quality traits, volatile organic compounds and texture features play a major role. PTR-ToF-MS, coupled with an artificial chewing device, was used to profile the entire apple volatilome of 162 apple accessions, while the fruit texture was dissected with a TAXT-AED texture analyzer. The array of volatile compounds was grouped in seven major groups and implemented into a genome-wide association analysis carried out with 9142 SNPs. Marker-trait associations were identified on seven chromosomes co-locating with important candidate genes for aroma, such as MdAAT1 and MdIGS. The integration of volatilome and fruit texture data conducted with a Multiple Factor Analysis unraveled a contrasting behavior, underlying an opposite regulation of the two fruit quality aspects. The association analysis implementing the first two principal components identified two QTLs located on chromosome 10 and 2, respectively. In the end, the distinction of the apple accessions on the base of the allelic configuration of two functional markers, MdPG1 and MdACO1, shed lights on the type of interplay existing between fruit texture and the production of volatile organic compounds.

Introduction

Fruits are functional components of the human diet, daily supplying important elements such as sugars, organic acids, vitamins and fibers. These elements change during the fruit developmental processes, to render the fruit more attractive and palatable, especially at the onset of ripening. Overall, the array of these compounds defines the quality of a fruit, though as a degree of excellence (Klee, 2010). Fruit quality can be therefore defined by the achievement of four principal quality factors: appearance, flavor, texture and nutritional properties (Costa et al., 2011). Among these, appearance, texture and flavor directly impact the postharvest performance and consumers' appreciation (Harker et al., 2008; Cliff et al., 2016), therefore the fruit marketability.

Physiologically, fruit texture depends on the dismantling process occurring on the polysaccharide cell wall architecture coordinated by cell wall modifying proteins. Fruit texture is composed by a series of sub-trait components, distinguished in mechanical and acoustic properties (Varela et al., 2006; Costa et al., 2011 and 2012). Beside the preference of consumers for crisp and juicy apple, firmer fruit are also more prone to long-time storage, facilitating shipping and ensuring a time-wise fruit market availability. In addition, also fruit flavor is another important quality factor, which results from the combination of a large array of primary and secondary metabolites. While taste is primarily related to non-volatile metabolites (mainly sugars, organic acids, free amino acids and salts), aroma is rather represented by the interaction of a wide blend of volatile organic compounds (VOCs) with human receptors. The apple aroma depends on the interaction of more than 370 VOCs (Dimick and Hoskin, 1983; Fuhrmann and Grosch, 2002; Ulrich and Dunemann, 2012; Farneti et al., 2015a) synthesized by the fruit during ripening and enahanced upon cellular disruption by biting and mastication (Contreras and Beaudry, 2013; Farneti et al., 2015b). Among them only a minor set of chemical compounds, for the most esters, alcohols and aldehydes, can be distinctly perceived (Holland et al., 2005; Ulrich and Dunemann, 2012).

Although these characteristics are essential factors for the fruit quality excellence, breeding efforts have historically been mainly oriented to improve fruit appearance and storability. Selection for yield, fruit size, color, and shelf life properties might have had unintended negative consequences on other fruit quality traits, for instance aroma, as already suggested for strawberry, peach and tomato (Goff and Klee, 2006; Klee, 2010; Rambla et al., 2014). The selection of firm apple accessions, thus distinguished by a higher storability, has been also facilitated by the identification and subsequent validation of QTLs and functional markers, associated to both ethylene and texture (Costa et al., 2005, 2010; Zhu and Barritt, 2008; Longhi et al., 2012 and 2013; Baumgartner et al., 2016). Moreover, this quality drop-off

has been enhanced by the fact that breeding for aroma occurred practically by chance (not assisted), since aroma is not yet considered as a discriminating trait in the early selection phase. This situation is also strengthened by the complex and time-consuming phenotyping protocols ordinarily used, which makes the analytical screening of large plant material unfeasible. This limit reduced to only few the scientific reports on QTL mapping related to apple aroma (Dunemann et al., 2009 and 2012; Rowan et al., 2009; Kumar et al., 2015; Yauk et al., 2015). In most of the cases, VOCs have been monitored and quantified with SPME-GC-MS equipment. Although it represents a valuable and accurate technique, it is laborious and time consuming. In this sense, Proton Transfer Reaction- Time of Fligh- Mass Spectrometry (PTR-ToF-MS) might represent a valid alternative to profile VOCs more time-efficiently (Lindinger et al., 1998; Jordan et al., 2009) also in apple (Zini et al., 2005; Costa et al., 2013; Cappellin et al., 2014). Beside the headspace concentration of VOCs, the interaction between aromatic compounds and human receptors should be also considered (Farneti et al., 2015b). Differences in VOC releasing behaviours, due to the textural and physicochemical properties of the food matrix, may influence the perception of aroma during food consumption (Farneti et al., 2013). VOCs are in fact released from the food matrix and then transported to mouth and nose receptors (Buettner et al., 2008). The modification of the food matrix and the long incubation time normally required by static based methodologies can drastically alter the in vitro VOC profile (Dewulf et al., 2002; Biasioli et al., 2011; Farneti et al., 2013). The employment of a strategy suitable to monitor the VOC emission released during chewing is therefore preferable. To this end, Farneti et al. (2013 and 2015a) developed an analytical system based on an artificial chewing device coupled to a PTR-MS in order to detect the VOC kinetics during food matrix processing.

In this survey, an apple collection was employed and assessed for both texture and aroma. To date these two fruit quality traits have been merely assessed separately, thus a comprehensive and exhaustive investigation on their relationship is lacking. The analysis of these datasets using multivariate statistical approaches (Multiple Factor Analysis-MFA and Functional Principal Component Analysis-FPCA) together with a high dense SNP genotyping platform, enabled a genome-wide association study addressed to step forward into the understanding of the genetic relationship between these two important fruit quality traits.

Materials and methods

Plant materials

In this investigation, a collection of 162 apple accessions was chosen within the germplasm repository available at the Fondazione Edmund Mach (Trento, Italy). Each genotype, planted in triplicate, was in adult and fruit bearing phase at the time of the analysis.

Trees were maintained with standard agronomic practices for fruit thinning, pruning and pest/disease control. Apples were harvested at commercial ripening stage determined following the changes of skin and seed color as well as the chlorophyll content degradation assessed non-destructively with a Da-Meter (TR, Forli, Italy, Ziosi et al., 2008). For each apple accession, a minimum of twenty homogeneous fruit were collected and stored for two months in a cold cellar (2–4 °C with ~ 95 % RH). After postharvest storage, fruit were removed and maintained at room temperature (~20 °C) overnight before texture and VOC analysis.

Fruit texture phenotyping

A subset of five apples/genotype was used for the fruit texture sub-traits characterization. The measurements were carried out with a Texture Analyzer (Stable MicroSystem, Godalming, UK). Protocol was detailed in Costa et al. (2011 and 2012). Briefly, for each genotype 20 measurements (four technical per five biological replicates) were carried out. The instrument was equipped with a 4 mm flat head probe and an AED (Acoustic Envelop Device) for the simultaneous acquisition of the mechanical and acoustic profiles, further processed with an ad hoc macro for the digital definition of twelve parameters. Out of these, eight were related to the mechanical signature of texture (Yield Force, Maximum Force, Final Force, Mean Force, Area, Force Linear Distance, Young's Module, Number of Force Peak) and four to the acoustic response (Maxim Acoustic Pressure, Mean Acoustic Pressure, Acoustic Linear Distance, Number of Acoustic Peaks; fully described in Costa et al., 2011, Supplementary Tab S1).

Dynamic VOC fingerprinting in apple

To profile the emission of VOCs during artificial chewing, another batch of five apples/genotype was assessed according to the method described by Farneti et al. (2015b). The chewing device was composed by a cylindrical glass cuvette (800 ml) sealed with a cap and equipped with a manual notched plunger (Fig. 1A). All device elements were made of polytetra-fluoroethylene. The fruit sample was represented by an apple flesh cylinder (1.7 cm diameter and 5 cm height) isolated from each fruit. Before crushing, the headspace VOC concentration of the apple flesh cylinder was measured for 60 s. The artificial chewing was performed by pressing the notched plunger 5 times within 10 s and VOC analysis continued for 60 s following crushing (Fig. 1B, C). This setting was optimized in preliminary trials in order to assure a variability lower than 5% for analysis repeated on the same fruit. The headspace was drawn from the chewing device to the PTR-ToF-MS at 2.4 L/h. VOCs were then assessed by direct injection of the headspace mixture into a commercial PTR-ToF-MS 8000 apparatus

(Ionicon Analytik GmbH, Innsbruck, Austria), set with the condition described in Farneti et al. (2015b).

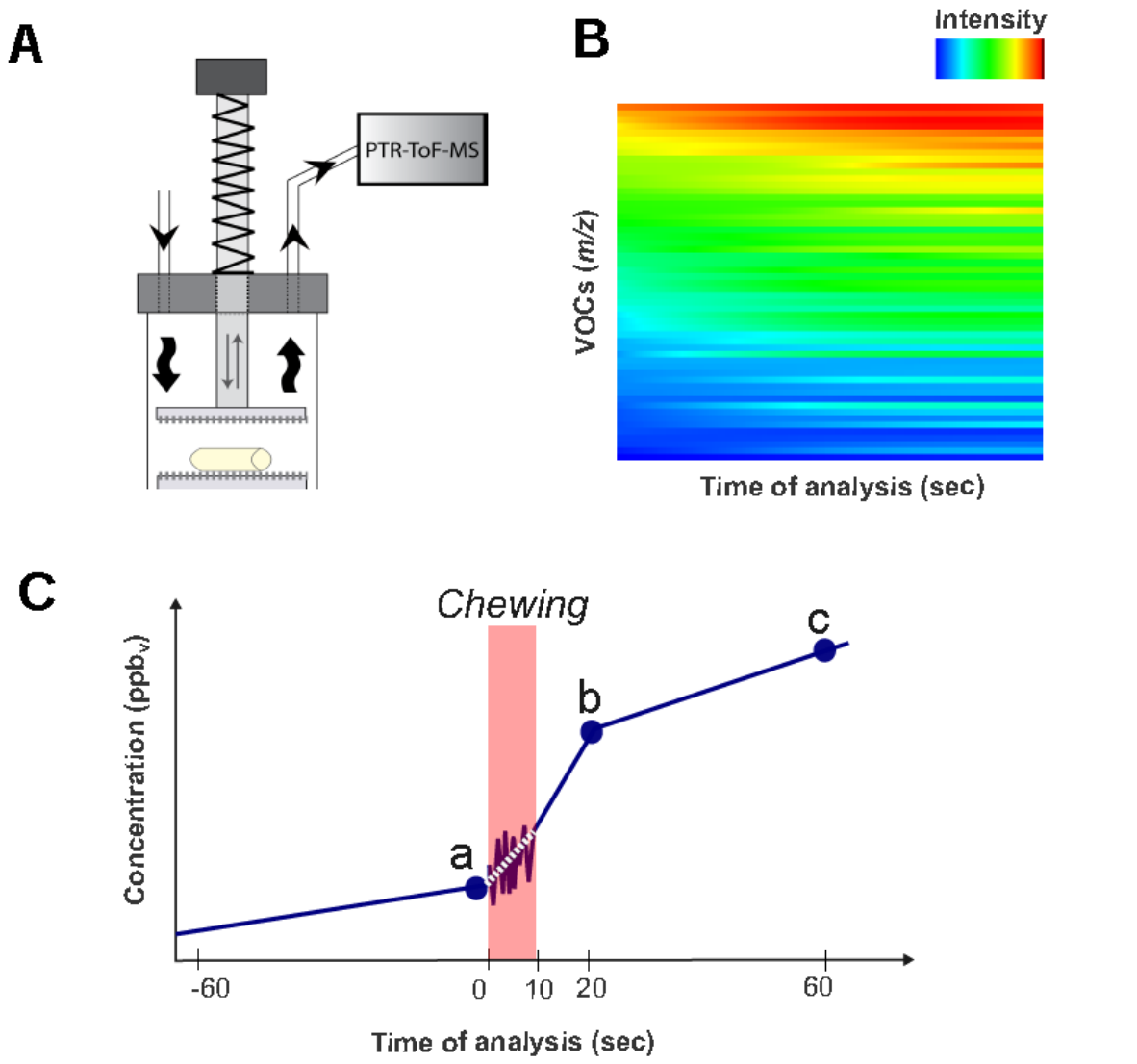


Figure 1. Schematic representation of the dynamic VOC fingerprinting of apple fruit assessed by PTR-ToF-MS coupled with an artificial chewing device (panel A) composed by a cylindrical glass cuvette (800 mL) sealed with a cap and a notched plunger. In panel B is reported the 3D heatmap of the on-line VOC dynamic fingerprinting carried out in 'Golden Delicious', selected as reference cultivar. For graphically purpose the VOCs are reported with a decreasing ordered based on their initial level (before chewing). The headspace VOC concentration was measured for 60 s before and after crushing, for a total of 120 s. In order to simplify the analysis of the entire dynamic VOC profiling only three specific time points of the entire dynamic were primarily compared: (a) before the artificial chewing (0 second), (b) 20 seconds and (c) 60 seconds after the fruit processing.

All apple cultivars were measured in five independent replicates for each measuring data point. The analysis of PTR-ToF-MS spectral data, compound annotation, spectra correction through Poisson statistics, internal calibration, noise reduction, baseline removal and compound quantification proceeded according to Cappellin et al. (2011a, b and 2012).

SNP genotyping

Young leaves collected from each apple accession were used for nucleic acid extraction. gDNA was isolated with Qiagen DNeasy Plant Kit. DNA quantity and quality was measured with a Nanodrop ND-8000 (ThermoScientific, USA). The 162 accessions were genotyped with the 20K Infinium SNP Array (Illumina), ad hoc designed for apple (Bianco et al., 2014). SNP data were filtered with ASSIsT (Di Guardo et al., 2015) obtaining a final set of 11.277 polymorphic markers. However, the identification and selection of the final set of markers to be used in GWAS cannot rely exclusively on SNP qualitative parameters (eg: relative amount of missing calls, call rate ratio and segregation distortion), since marker physical position may result inaccurate, due to the high sequence homology between homoeologue chromosomes and the high heterozygosity of the apple genome (Velasco et al., 2010). To this end, from the total number of SNP markers included in the 20K Illumina Infinium Array, 9.142 were in the end employed for GWAS analysis. This subset resulted therefore reliable within the 162 apple accessions (based on the filtering processed by ASSIsT) and positioned on the consensus genetic map described in a parallel study (Di Guardo et al., 2017, Chapter 4).

Marker-Trait association by GWAS

Filtered SNP data and phenotypic assessment (represented by both texture and VOC analysis) were jointly analyzed in a marker-trait association approach. For this purpose, the software TASSEL v3.0 was employed and the genome-wide association study (GWAS) was computed implementing two models: GLM and MLM. The GLM model (Pritchard et al., 2000) was performed taking into consideration population structure (Q matrix) to correct for genetic stratification. The membership of each individual into sub-population was implemented into the model as covariates and it was represented by principal components (PCA). The second model adopted here to find marker-trait association was the mixed linear model (MLM, Yu et al., 2006), which also considered the kinship matrix (population relatedness) to correct for false association. This model is expressed by the Hendersen matrix as follows:

$$Y=X\beta+Zu+e$$

where **Y** is the vector of observation, **β** is a vector containing fixed effects (including genetic markers and population structure), **u** is a vector of random additive genetic effects for multiple background QTL, **X** and **Z** are the known design matrices and **e** is the unobserved vector of random residuals. Significant associations were selected on the threshold of P-value ≤ 0.05, corrected for multiple comparisons according to the FDR (false discovery rate) procedure reported by Benjiamini and Hochberg (1995), calculated with the “stats” package of R (R core

developing team). The model used per each trait was selected on the base of the visual inspection of the Q-Q plot (“qqplot” R package).

Statistical analysis

The array of protonized VOC masses was reduced by applying noise and correlation coefficient thresholds. The first removed peaks with mean intensity less than 25 ppbv and not significantly different from blank samples (Farneti et al., 2015a). The latter excluded peaks having correlation higher than 99%, which correspond for the most to isotopes of monoisotopic masses. During the progression of the notched plunger, the VOC profile/signal was, as expected, not stable, thus, 10 seconds of artificial mastication were removed and substituted by cubic spline interpolation (Fig. 1C). Further analyses were therefore carried with smoothed curves. Each VOC dynamic was characterized by mean intensity at three specific time points: before chewing (a: 55 sec ± 5 sec), immediately after chewing (b: 80 sec ± 5 sec) and at the end of the measurement (c: 120 sec ± 5 sec). To represent the general changes of VOC profiles, Principal Component Analysis (PCA) was performed on the log-transformed data on these three data points. A Multiple Factor Analysis (MFA) was used to further compare fruit texture with VOC profiles before fruit crushing, in a way similar to classical static headspace analysis. Moreover, a multivariate Functional Principal Component Analysis (FPCA, Ramsay and Silverman, 2005) was implemented to perform the analysis on whole VOC pattern released during the artificial chewing. To this end, 50 linear combinations of parabolic b-spline basis objects (the highest order equal 3) were constructed for the creation of functional data object from VOC dataset.

Visualization of significant VOC correlations ($P < 0.01$; $R > 0.50$) was conducted by the generation of a Correlation Analysis Network with Cytoscape (version 3.2.1; Cline et al., 2007). The ClusterONE plugin (Nepusz et al., 2012) was used to identify putative metabolite clusters by finding regions of significant local density. R 3.2.0 internal statistical functions and the external packages “ChemometricsWithR”, “FactoMineR”, “Funclustering”, “fda”, “ggplot2” were used for the multivariate statistical methods employed in this work. Regarding the texture analysis, each combined mechanical-acoustic profile was analyzed with the Exponent v4.0 software (Stable MicroSystem, Godalming, UK).

Results and discussion

High resolution VOC phenotyping

In this study, the interplay between fruit texture and aroma was investigated by a comprehensive high-resolution phenotyping assessment (Supplementary Tab. S2). Fruit were analyzed after a period of cold storage (2 months), taking also into account that both volatilome and texture undergo important changes during the postharvest phase (Fellman et al., 2003; Newcomb et al., 2006; Schaffer et al., 2007). The analysis of aroma was performed on apple cut flesh portion, since the release of aroma is distinct and dominating in processed fruit (for instance during mastication) rather than intact (Farneti et al., 2015b). It is in fact known that VOCs, indeed, can be distinguished in two categories (Yahia, 1994): those produced by whole fruit and those synthesized during chewing. Fruit cutting highly stimulates the changes in both concentration and composition of VOCs in the headspace, due to an increased exposure of the food matrix to air (de Roos, 2003; Arvisenet et al., 2008). This operation triggers several chemical reactions as a consequence of cell disruption, such as lipid oxidation and the consequently synthesis of aldehydes. In order to evaluate the VOCs affecting aroma perception, the methodology proposed by Farneti et al. (2015b), based on the real-time analysis of volatiles emitted during in vitro mastication, was used (Fig. 1A). The PTR-ToF-MS setting employed in this investigation enabled a full scan of the entire volatilome in 1 second, allowing the real time monitoring for most VOCs emitted by the fruit during chewing. This detailed characterization permitted the development of a dynamic VOC fingerprinting before and after the fruit in vitro mastication (Fig. 1B and Supplementary Fig. S1). The entire VOC profile, assessed for the 162 apple accessions by PTR-ToF-MS, was reduced from 590 to 33 masses, applying noise and correlation coefficient thresholds. The resulting array of VOCs (Supplementary Tab. S3) was comparable to the dataset described by Farneti et al. (2015a). The blend of VOCs detected in apple fruit for the most includes alcohols (i.e. m/z 33.033, 47.049, or 85.101), aldehydes (i.e. m/z 83.086, 99.081, or 101.097), phenylpropenes (i.e. m/z 134.072 and 149.097) and esters (i.e. m/z 61.027, 89.059, or 117.091). Among these classes, esters, of both straight and branched types, are to date recognized as the most relevant VOCs in apple aroma (Holland et al., 2005; Ulrich and Dunemann, 2012). Similarly, to the results presented by Farneti et al. (2013 and 2015b), VOCs were differently released from the food matrix according to their chemical nature and, more likely, to the textural properties of the apple flesh (Supplementary Fig. S1). Overall, esters (i.e. m/z 61.028, 89.059, and 117.091) and alcohols (i.e. m/z 43.054, 57.069, and 71.086) were rapidly released after fruit structure crushing, while for other molecules, such as acetaldehyde (m/z 45.033) and acetone (m/z 59.049), the emission was less influenced by the sample disruption. In addition to this, several other compounds, mainly C6-aldehydes, such

as hexanal (m/z 83.086 and 101.097) and hexenals (m/z 81.070 and 99.081), revealed a third trend characterized by a constant and linear production, delayed of about 20 s from the initial tissue disruption (Supplementary Fig. S1).

As first attempt to simplify the analysis of the entire VOC profiling during the in vitro mastication, we initially compared only three specific time points of the entire VOC dynamic: i) before the artificial chewing (0 second), ii) 20 seconds and iii) 60 seconds after the fruit processing (Fig. 1C). The effect of the artificial mastication on apple VOC profiling is depicted in the PCA plot defined by the implementation of the first two principal components, explaining together 65.5% of the total apple volatilome variability (Fig. 2A and Supplementary Fig. S2).

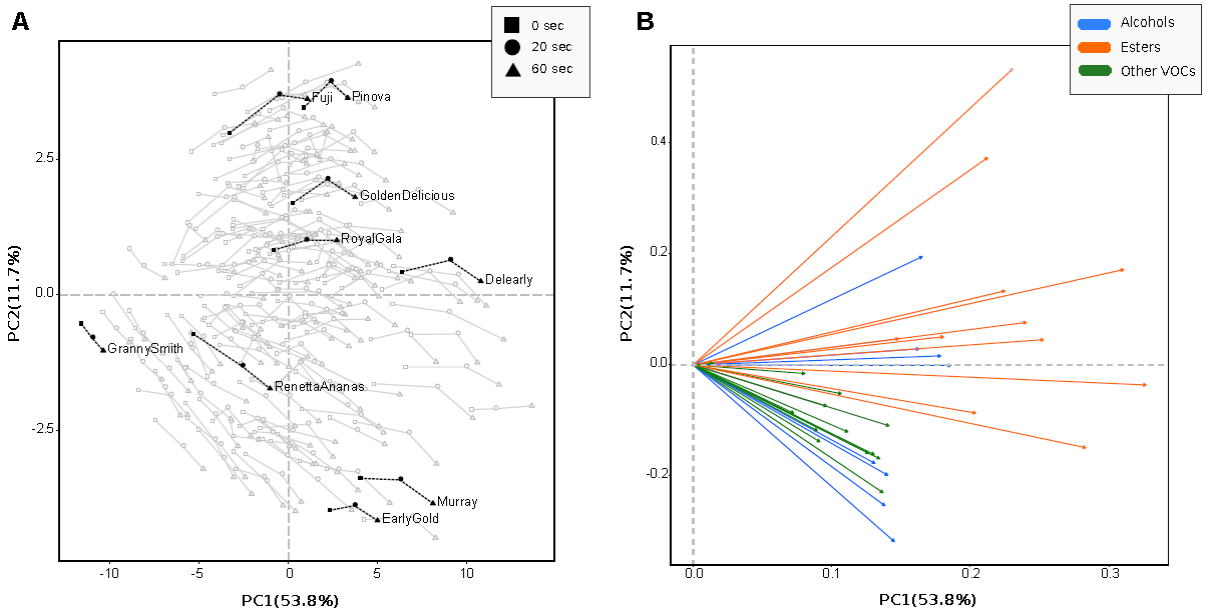


Figure 2. Principle Component Analysis (PCA) plot (A) and loading projection (B) of the VOC distribution assessed by PTR-ToF-MS during the artificial chewing. Plot “A” depicts the VOC profile distribution of the apple cultivars over the PCA score plot defined by the first two principle components. Within the germplasm collection, nine cultivars were arbitrarily highlighted ('Delearly', 'Early Gold', 'Fuji', 'Golden Delicious', 'Granny Smith', 'Murray', 'Pinova', 'Renetta Ananas', and 'Royal Gala'). Different symbols (square, circle, and triangle) indicate the time of assessment during the artificial chewing (0, 20, and 60 seconds after chewing). Each data point is the average of five biological replicates. Plot “B” shows the projection of the 33 significant VOC mass peaks reported using different colors according to the chemical family. The mass peaks identity is reported in the Supplementary Fig. S2.

According to the loading plot (Fig. 2B), the first principle component (PC1, 53.8%) mainly correlates with the quantitative concentration of VOC, while the second (PC2, 11.7%) was more related to the qualitative distinction of VOCs (chemical composition). Positive PC2 values indicate, for instance, a higher concentration of esters (i.e. m/z 61.027, 43.017, or 71.085) and amyl alcohols (i.e. m/z 71.085), while negative values are linked to a greater abundance of methanol (m/z 33.033), acetaldehyde (m/z 45.033), ethanol (m/z 47.049), and C6-aldehydes (m/z 81.07 or 83.086). Beyond the differentiation of apple cultivars based on their aromatic

profiles, samples are also differentially distributed on the PCA-hyperspace according to the VOC assessment carried out at three specific time points after chewing (0, 20, 60 seconds). Although the VOC magnitude was enhanced by mastication, a genetic and physic (apple flesh structure) regulation was suggested.

Volatilome QTL mapping

To depict the general behaviour existing among the VOCs assessed within the germplasm collection, the apple volatilome was analysed through a correlation network (Fig. 3A).

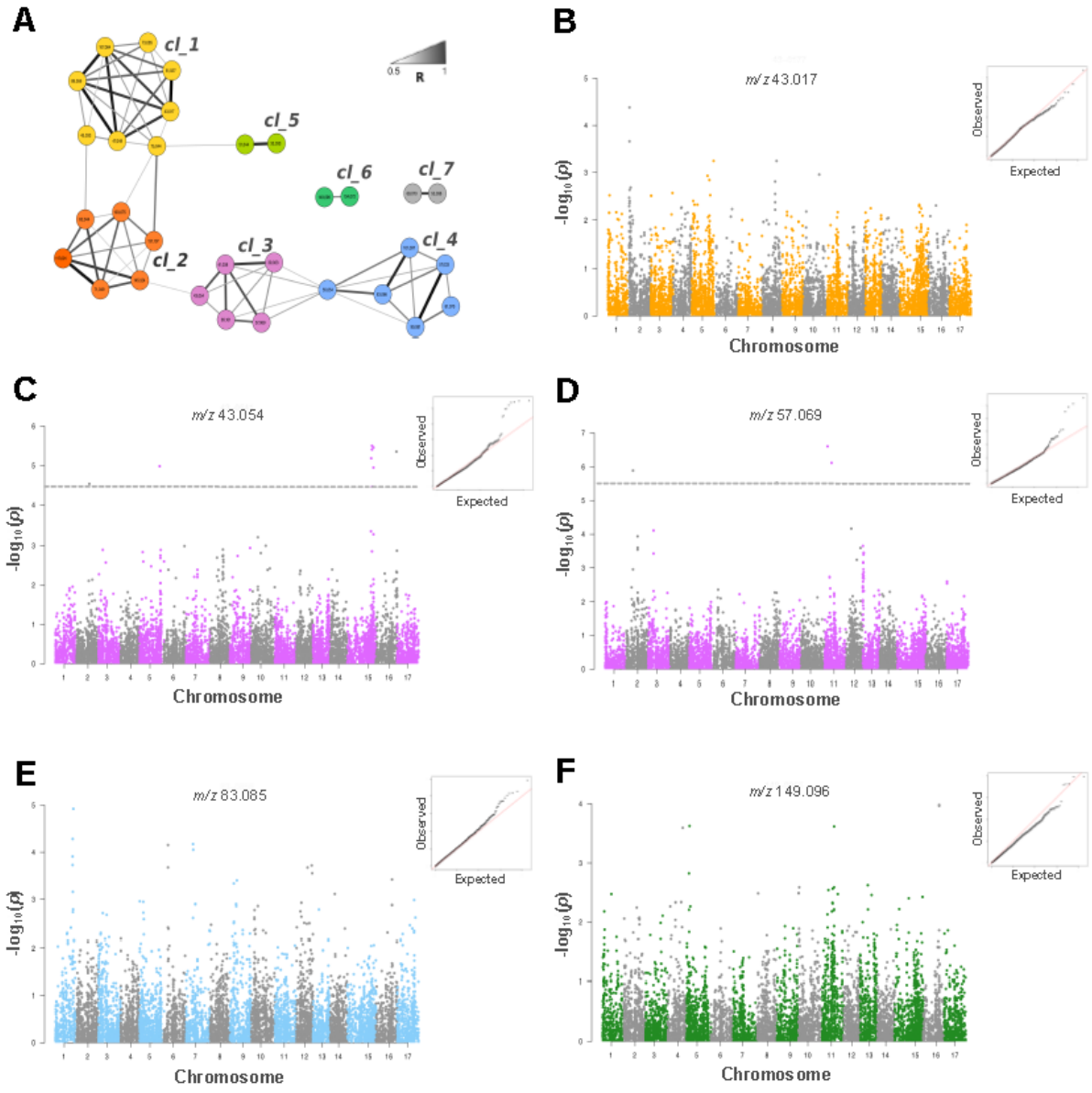


Figure 3. Correlation Analysis Networks (CAN) of VOCs (A) and genome-wide association results (from B to F). The CAN (A) is obtained by determining the significant Pearson correlations ($p < 0.01$) among the 33 VOC mass peaks assessed by PTR-ToF-MS at 60 seconds after the artificial mastication. The color-coding of the edges (gradient from light gray to black) denotes the level of correlation (R from 0.5 to 1).

Positive and negative correlations are shown by solid and dashed lines, respectively. Significant VOC clusters, identified by Cytoscape ClusterONE plugin, were highlighted with different colors. Each cluster is also defined by a numerical code according to the VOC category as follow: 1, esters; 2, ethanol and acetaldehyde; 3, alcohols; 4, aldehyde; 5, methanol; 6, phenylpropenes and 7, unknown compounds. For GWAS analysis, five Manhattan plots (from B to F) showing the SNP association (based on P-value – log10 transformed) with selected representative masses (reported on the top of each plot and depicted with the same color of the CAN grouping) are illustrated. For each Manhattan plot, the Q-Q plot and the corrected P-value threshold (FDR ≤ 0.05), when possible, are also indicated.

The network, created from a significant Pearson correlation matrix (P ≤ 0.01, threshold 0.05) among the set of 33 masses identified seven main groups of VOCs (Fig. 3A; Supplementary Tab. S3). The first two groups (1 and 2) mainly include ester compounds, ethanol and acetaldehyde. The high positive correlation between esters and the two anaerobic metabolites (acetaldehyde and ethanol) is generally observed in several fruit species, since the latter compounds are involved in the synthesis of several aroma volatiles during fruit ripening (El Hadi et al., 2013). Acetaldehyde is generally accumulated during ripening, also under aerobic conditions (Fidler, 1968), and it is formed from pyruvate by action of pyruvate decarboxylase (PDC) enzyme. The two immediate products formed from acetaldehyde are ethanol, produced by alcohol dehydrogenase (ADH), and acetyl coenzymeA (CoA), obtained by the activity of the enzyme aldehyde dehydrogenase (ALDH; Cossins, 1978). While acetyl CoA is the precursor of acetate esters, acyl-CoA is involved in the formation of longer esters (Gilliver and Nursten, 1976). The amount of ethyl ester (such as ethylacetate and ethyl butanoate) is correlated with the content of ethanol (Pesis, 2005). The third and fourth group of masses were mainly composed by alcohols and C6-aldehydes, respectively (Supplementary Tab. S3). The remaining three clusters (5, 6 and 7) were instead composed by methanol (m/z 33.033 and its water-cluster m/z 51.044) and phenylpropenes (m/z 134,072 and 149.096), together with unknown fragments. Methanol, similarly to ethanol and acetaldehyde, is positively correlated (R= 0.63) with an ester included in group 1, finding agreement with the involvement of methanol in the methyl esters synthetic pathway (i.e. methyl acetate, m/z 75.044).

To identify the most significant genomic regions involved in the genetic control of the apple volatilome, a genome-wide association study was performed with MLM model. The genetic dissection was carried out selecting the most reliable VOCs (mass) within each network analysis group, based on Q-Q plot. The choice of the representative mass/cluster is also justified by the fact that the PTR-ToF-MS device accurately detects the nominal mass of a molecule, important to identify an array of compounds with similar structure and thus with a similar quality impact. The association between the set of SNP and m/z 43.017, a common ester fragment selected to represent group 1, identified a major QTL on chromosome 2 (Fig. 3B and Supplementary Tab. S4). Although the most significant SNP (RB_1979331_L2_PA) does not cross the adjusted threshold, this genomic region coincided with MdAAT1. This gene belongs

to the alcohol acyl-transferase family and it is deputed to catalyse the transacylation from acyl-CoA to ester (esterification). In several fruits AAT is essential to control flavour biogenesis during the fruit ripening phase. It has been in fact documented (Aharoni et al., 2000; Beekwilder et al., 2004; El-Sharkawy et al., 2005 and Dunemann et al., 2012) that esters are the most important compounds within the aromatic bouquet of fruit. In apple, moreover, esters are, amongst others, the dominating compound, contributing for the 80% of the entire aromatic blend. For this species, the expression profile of MdAAT1 resulted to be consistent with the production of esters and the accumulation of ethylene as well (Schaffer et al., 2007). This gene (MDP0000637737) was further retrieved on the apple genome assembly (GDR database, Jung et al., 2014) within an interval of 400 kb, established as LD extent present in domesticated apples (Di Guardo et al.,2017, Chapter 4).

For the alcohol group (depicted in group 3, Fig. 3A) two masses were selected and further implemented in marker trait-association study (m/z 43.054 and m/z 57.069). These two compounds were selected considering that they did not show a high correlation value (R= 0.35), hypothesizing the identification of two distinct groups. For m/z 43.054, statistically significant SNPs were found in chromosomes 2, 5, 15 and 16 (Fig. 3C and Supplementary Tab. S4). Interesting candidate genes were identified in the homoeolougus pair of chromosomes 2 and 15 (within the LD interval of 400 kb). In the QTL region on chromosome 2 an alcohol dehydrogenase-1 like (ADH) gene was annotated (MDP0000523942). The action of ADH is devoted to reduce aldehydes (previously reduced by AcylCoA) to alcohols (that will be further converted to esters by AAT; De Filippi et al., 2005). In other climacteric species such as tomato, this gene is expressed during ripening (Chen and Chase, 1993) and functional validation demonstrated that fruit with an enhanced ADH activity were distinguished by a higher concentration of alcohols and a more typical flavour of ripe fruit (Speirs et al., 1998). On the same chromosome region, also MdACS3 (MDP0000247533) was identified. This element, which belongs to the 1-aminocyclopropane-1-carboxylate family, is involved in the early phase of the apple ethylene production (Wang et al., 2009). Since its expression precedes MdACS1 (mapped on chromosome 15; Costa et al., 2005), this element might play an important role in the transition phase from the ethylene system 1 to system 2, supporting thereby the direct role of this hormone in controlling the VOC production. On chromosome 15, instead, another alcohol acethyl transferase gene was identified (MDP0000528775), supporting the interplay between alcohol and ester in apple. Chromosomes 2 and 15 were already associated to the accumulation rate of alcohol compounds in apple by Kumar and colleagues (2015), however no genes were identified in these regions. For the second alcohol compound implemented in the GWAS analysis (m/z 57.069), a QTL was located on chromosome 11 (Fig. 3D and Supplementary Tab. S4), on which a short-chain dehydrogenase reductase3b-like gene (SDR) was annotated

(MDP0000313884). This element encodes for one domain NAD(P)(H)-dependent enzyme characterized by a wide range of substrates, including alcohols and aromatic compounds (Kallberg et al., 2002; Persson et al., 1995). This gene, also known as alcohol dehydrogenase, is involved in the regulation of alcohol/aldehyde ratio (Moummou et al., 2012). In particular, SDR genes contribute to the aroma biosynthesis in tomato, converting phenyl acetaldehyde to the corresponding alcohol (Tieman et al., 2012). SNP markers associated to these compounds (Supplementary Tab. S4) showed a P-value exceeding the statistical threshold corrected for multiple comparisons (FDR ≤ 0.05).

From the aldehyde compounds group (group 4), QTLs were identified and located on chromosome 1 and 7, respectively (coincident with FB_0442970_L1_PA, FB_0697476_L7_PA; Fig. 3E and Supplementary Tab. S4). Although no relevant gene was identified for chromosome 1, an additional alcohol dehydrogenase (ADH) gene (MDP0000077529) was annotated on chromosome 7. This association, together with the aforementioned reported genes, strengths the role of this region in the regulation between alcohol and aldehyde VOC categories.

The last VOC implemented in the GWAS analysis is m/z 149.096 (Fig. 3F and Supplementary Tab. S4), corresponding to phenylpropenes. The ordering of the SNP markers based on their P-value allowed the detection of three genomic regions located on chromosome 5 (RB_14354679_L5_PA), 11 (FB_0086581_L11_PA) and 16 (FB_0362423_L16_PA), respectively. While for the QTLs on chromosomes 5 and 16 any significant gene involved in the phenylpropenes synthesis was not annotated, an isoeugenol synthase-1 like gene (MDP0000141131) was targeted on chromosome 11. This gene is involved in the biosynthetic pathway of phenylpropanoid (PhP-Vs), a VOC category with multiple roles, from attractors to pollinators and defence, to important contributors of the typical “spicy/smoky” aroma of fruits (Karapinar, 1990; Koeduka et al., 2006; Pasay et al., 2010). Isoeugenol synthase (IGS), as eugenol synthase (EGS), is a NADPH-dependent enzyme converting coniferyl acetate (synthesized from phenylanine along the phenypropanoid pathway) into isoegenol and eugenol, respectively (Koeduka et al., 2009; Aragüez et al., 2013). Both enzymes (IGS and EGS) can utilize the analogous substrate coumaryl acetate to produce t-anethol and chavicol, respectively. The methylation of the para-hydroxy groups on the benzene ring by O-methyltransferase (OMTs) catalyses the final formation of t-anethol and estragole. The latter compound was shown to be highly accumulated in ripe fruit of ‘Royal Gala’ apple (Yauk et al., 2015). In our investigation, this same apple cultivar was included in a group (together with ‘Delearly’, ‘Delblush’, ‘Golden Delicious’, ‘Prima’, ‘Delicious’ and ‘Pinova’, for example) characterized by the highest accumulation of phenylpropenes during artificial chewing. The emission of PhP-Vs may mostly depend from the deconjugation of their glycosylate precursor, operated by glycosidase upon fruit disruption, rather than its de novo biosynthesis (Rambla et

al., 2014). This mechanism, at the base of the identification of NSGT1 (NON-SMOKY GLYCOSYLTRANSFERASE1), a gene involved in the conversion of non-cleavable triglycosides form of phenylpropanoids (preventing deglycosylation and release; Tikunov et al., 2013), supports the methodology adopted here of determining VOCs during fruit artificial disruption rather than by intact fruit. The targeting of this IGS gene and associated SNP markers might open more opportunities to select against or in favour of this class of compounds, important for the aromatic blend in many fruit species.

Dynamic VOC profiling and GWAS analysis shed light on the interplay between volatilome production and fruit texture properties

To step forward into the inter-relationship between texture properties and the aromatic pattern of apple, both mechanic and acoustic signature of fruit texture were assessed. The overall texture variability, examined by Principle Component Analysis (PCA, Supplementary Fig. S3), revealed a distinct classification of apple cultivars based on these parameters. All 162 cultivars are uniformly spread over the PCA score plot defined by the first two PCs, expressing together 97.9 % of the total variability. The first principle component (PC1, 82.6%) corresponds to the overall apple textural performance, while the second component (PC2, 15.3%) mainly refers to the subtler classification based on the prominence of mechanical (i.e. flesh firmness) or acoustic (i.e. crispness) parameters. Although this characterization has been performed in previous investigations (Costa et al., 2011; Longhi et al., 2012), the results presented here were obtained with the largest apple collection employed to date for fruit texture investigation. The texture dataset was further integrated with the volatilome data (Supplementary Tab. S2), represented by the array of VOCs fingerprinted at 60 s after the in vitro mastication and assessed using a Multiple Factor Analysis (MFA, Fig. 4 and Supplementary Fig. S4).

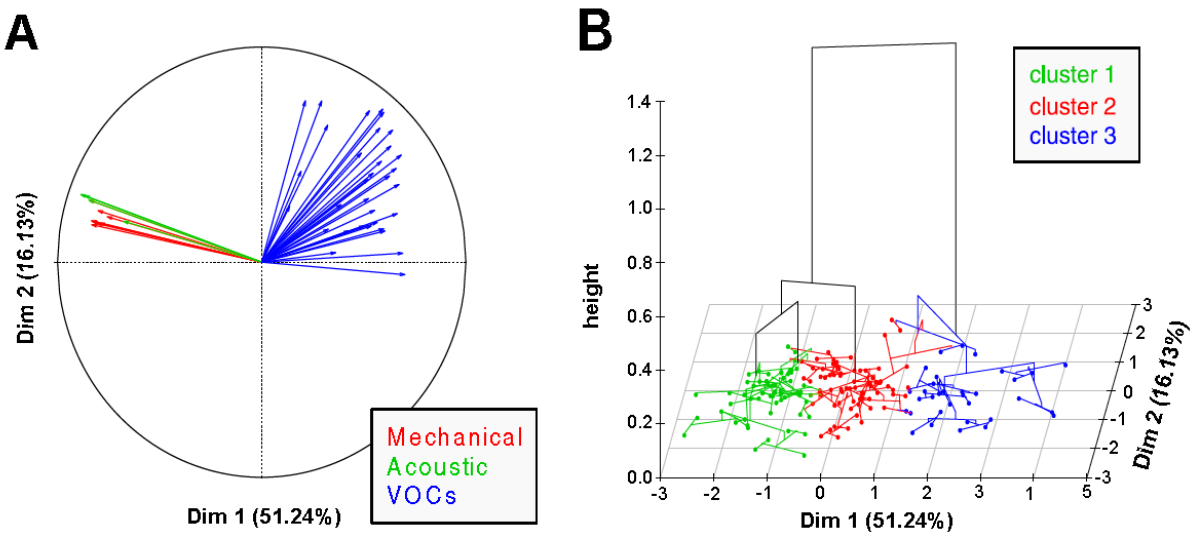


Figure 4. Multiple Factor Analysis (MFA) of texture properties (mechanical and acoustic) and VOCs. Plot "A" shows the MFA projection of the 33 VOC masses assessed by PTR-ToF-MS at 60 seconds after the artificial mastication (grey color) together with eight mechanical (orange color) and four acoustic (black color) parameters. Plot "B" depicts the hierarchical clustering of the 162 apple cultivars and their distribution over the MFA score plot defined by the first two dimensions. Apple cultivars are also grouped into three significant clusters highlighted by three colors (green, red, and blue). The detailed hierarchical clustering representation is also reported in the Supplementary Fig. S4.

In this computation, the samples were distributed according to the first two dimensions, accounting together to 67.4% of the total phenotypic variance. The loadings' projection (depicted in Fig. 4A) shows that VOC and texture variables were oppositely oriented, suggesting a contrasting physiological behaviour. Most of apple cultivars distinguished by a high aromatic volatile production are therefore characterized by a low texture performance and vice-versa. This distinction, which is for the most plotted according to the 1st dimension (explaining 51.2% of the entire variability), is moreover magnified in the hierarchical clustering represented over the MFA 2D-plot (Fig. 4B). In this plot, cultivars distributed according to their phenotypic values are also grouped into three significant clusters, defined by the specific weight of the most predominant phenotype (Supplementary Tab. S5, Supplementary Fig. S4). The first two clusters (cluster 1 and cluster 2) are distinguished by a low VOC production. Moreover, while in cluster 1 accessions with high acoustic performance are included, in cluster 2 apple accessions with high values for mechanical parameters are grouped. The last group, namely cluster 3, is instead represented by varieties with low texture performance but high VOC production. Although VOC and texture regulations belong to two distinct physiological pathways, they seem to be controlled by two mechanisms negatively correlated between each other. To shed light on this regulation, the VOC profiles were further assessed applying a functional principal component analysis (FPCA, Fig. 5 and Supplementary Fig. S5), which considered the entire profile of each VOC mass released during the artificial chewing process. FPCA is an exploratory multivariate technique that allows the analysis of functional data, essentially curves and trajectories. In this context, the VOCs released during the in vitro mastication are considered as "functional", since they are single entities rather than merely sequences of individual observations (Ramsay and Silverman, 2005). The distribution of the apple accessions on the FPCA plot defined by the first two PCs (PC1: 50.1%; PC2: 12.3%), showed a consistent grouping of the varieties into three clusters (Fig. 5A) according to the previous MFA hierarchical plot (Fig. 4B).

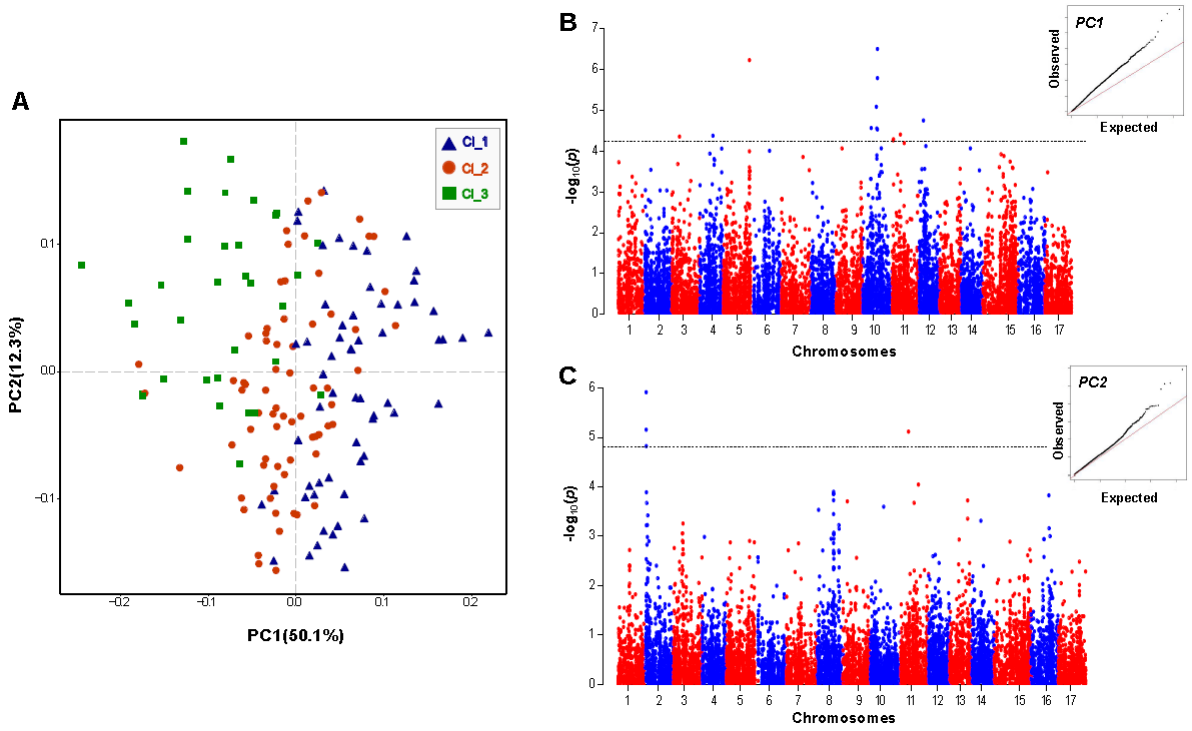


Figure 5. Functional Principle Component Analysis (FPCA) of the apple volatilome assessed with a PTR-ToF-MS before and after processing. In panel A the FPCA plot illustrating the distribution of the 162 apple cultivars employed here is shown. The cultivar identity is detailed in Supplementary Fig. S5. In the plot the three clusters based on the MFA hierarchical cluster (Fig. 4) are also depicted with green triangles (Cl_1), red circles (Cl_2) and blue squares (Cl_3), respectively. In panel B and C is instead reported the genome-wide association results for PC1 (B) and PC2 (C), respectively. For each Manhattan plot, the Q-Q plot and the FDR correction threshold are also reported.

To genetically dissect the VOC control in apple, the FPCA components were further employed as phenotypic trait into GWAS computation. The implementation of PC1 in the analysis, enabled the detection of a major QTL on chromosome 10, with a cluster of SNPs exceeding the FDR adjusted threshold (Fig. 5B, Supplementary Tab. S4). According to the Manhattan plot, this QTL coincided with MdPG1, a polygalacturonase gene encoding for an enzyme involved in the depolymerization of pectin and playing a major role in the control of the fruit firmness in apple (Brummell and Harpster, 2001; Brummell, 2006). The fact that this locus, known to be associated to the fruit firmness variation in apple (King et al., 2000; Maliepaard et al., 2001; Costa et al., 2010; Longhi et al., 2012 and 2013; Kumar et al., 2013) was associated to the quantitative variation of VOC production in apple reinforces the hypothesized interplay between texture and aroma. When the PC2 was instead used in the GWAS analysis, another QTL was identified and located on chromosome 2 (Fig. 5C, Supplementary Tab. S4). Among the SNPs exceeding the adjusted threshold a marker associated to MdAAT1 was found (FB_0451368_L2_PA). This gene, as aforementioned reported, is a major candidate in the formation of ester. The association between MdAAT1 and PC2 is instead more related to the type of aroma (rather than its overall production), which in apple is for the most related to the

ratio between ester/alcohol (Ulrich and Dunemann, 2012; Farneti et al., 2015a; Newcomb et al., 2006). This result unraveled that the quantitative and qualitative VOC production in apple are under a different genetic control, confirmed by the different genetic association obtained implementing PC1 or PC2. In particular, the association between the principal component related to the general amount of volatile and MdPG1 support the role of the fruit texture structure in the release of aroma in apple.

The selection in favor of fruit firmness negatively impacts the production of VOCs in apple

Within the several aspects of apple fruit quality, fruit texture and flavor are dominating features for their effect on the postharvest performance and consumer preference (Baietto and Wilson, 2015). Despite the relevance of both traits, the breeding for fruit quality is fundamentally based on fruit firmness for two reasons. Fruit texture (especially firmness) is easy to measure and validated functional markers are already available for an anticipated-assisted selection of the most performing individuals, such as MdACS1, MdACO1 and MdPG1 (Costa et al., 2005; Zhu and Barritt, 2008; Baumgartner et al., 2016). Also for fruit aroma a molecular marker based on alcohol acyltransferase gene has been recently developed (MdAAT1; Dunemann et al., 2012). However, since the association between the type of aroma selected by this marker and the consumer preference is lacking, this tool is not yet used in breeding assisted selection. To this end, for a better understanding of the relationship between fruit texture and aroma, the distribution of the apple accessions over the FPCA and arranged into three clusters (Fig. 5A and Supplementary Tab. S5) was reconsidered on the base of the allelotype configuration of two functional markers, MdPG1 and MdACO1 (Fig. 6A).

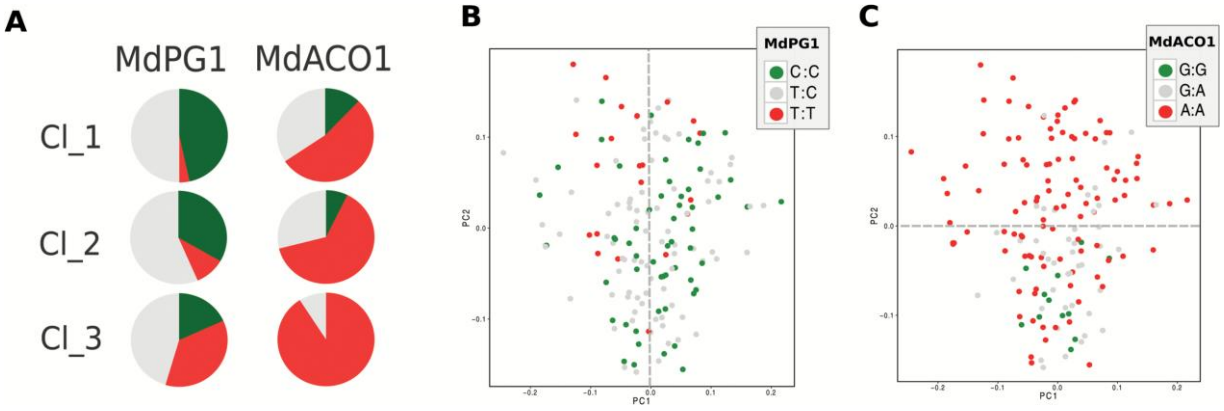


Figure 6. MdPG1 and MdACO1 allelotype configuration of the apple varieties distributed on the FPCA plot. For each panel, green and red color are for the favorable and un-favorable homozygous allelic state, while the heterozygous are reported in grey. The pie charts depicted in panel A show the proportion of allele for the two functional markers in the three clusters defined by the MFA analysis. Panel B and C illustrate the cultivar distribution of the FPCA plot, colored according to the allelotype of MdPG1 (B) and MdACO1 (C), respectively. In the legend of panel B and C, the allelism of the two SNPs associated to both genes is also indicated.

Amongst others, these two markers were specifically selected for sharing their genetic position on chromosome 10, co-locating with QTLs associated to fruit firmness and softening (Costa et al., 2010; Longhi et al., 2012; Kumar et al., 2013). The distribution of the apple cultivars based on MdPG1 allelism depicted a clear distinction around the PC1 (Fig. 6B). While most of the cultivars characterized by a homozygous unfavorable allelic configuration (TT), promoting fruit softening due to an intense pectin enzymatic depolymerization, are plotted on the negative PC1 area, the favorable homozygous genotypes (CC) are for the most located on the positive PC1 quadrants. This distribution confirmed, moreover, the distinction of the three clusters identified through both MFA and FPCA (Fig. 4 and Fig. 5). According to this new elaboration, the cluster 3 (represented by varieties with a low texture properties and high VOC production) was distinguished by a high proportion of unfavorable TT alleles for MdPG1, with regards to clusters 1 and 2. The allelic distribution for MdPG1 is moreover consistent with MdACO1. The cluster 3 is therefore characterized by the dominant presence of AA allelotype for this gene, associated to a high ethylene production, while in cluster 1 and 2 this ratio decreases in favor of the GG allelotype, which is associated to a low ethylene production (Fig. 6A). The varieties distinguished by the two MdACO1 allelotype are furthermore distributed around the PC2 (Fig. 6C), with the cultivars included in cluster 3 mostly plotted in the PC2 positive area. The breeding towards firm fruit would therefore select apple fruit distinguished also by a low aroma production. This relationship finds an explanation related with the physiological role of the two genes employed here. MdPG1 is involved in the control of fruit texture, which depends on the degradation of the cell wall polysaccharide structure. Firm apples are therefore characterized by a more intact and solid cell wall, which can prevent the release of VOCs due to a more tight cell compartmentation. On the other side, MdACO1, is the last gene governing the final biochemical formation of ethylene, a plant hormone triggering and coordinating several ripening processes (Giovannoni, 2001; Bennett and Labavitch, 2008). The amount of ethylene, moreover, has been already correlated with the fruit softening in apple (Costa et al., 2005 and 2010; Wakasa et al., 2006) as well as the rate of VOC production (Schaffer et al., 2007). To this end, the actual breeding in favor of postharvest would most likely also decrease the aromatic blend in apple fruit.

The exploitation of the genetic variability existing within the apple germplasm can allow a valuable combination of alleles for the selection of high quality ideotype. This goal, however, can be achieved only with a more informed and precise identification of the most performing cultivars to be used as superior parental lines.

Supplementary data

Supplementary data can be found at:

https://drive.google.com/drive/folders/0B1u0mN44Aw3pZ05IZG9XQW9yY2M?usp=sharing

Acknowledgments

The authors want to thank Elisa Banchi for the SNP genotyping and Marco Fontanari for his valuable support in texture phenotyping.

References

Aharoni A, Keizer LCP, Bouwmeester HJ, et al. 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell **12**, 647–661.

Aragüez I, Osorio S, Hoffmann T, Rambla JL, Medina-Escobar N, Granell A, Botella MÁ, Schwab W, Valpuesta V. 2013. Eugenol production in achenes and receptacles of strawberry fruits is catalyzed by synthases exhibiting distinct kinetics. Plant Physiology **163**, 946–958.

Arvisenet G, Billy L, Poinot P, Vigneau E, Bertrand D, Prost C. 2008. Effect of apple particle state on the release of volatile compounds in a new artificial mouth device. Journal of agricultural and food chemistry **56**, 3245–53.

Baietto M, Wilson AD. 2015. Electronic-Nose applications for fruit identification, ripeness and quality grading. Sensors, 899-931.

Baumgartner IO, Kellerhals M, Costa F, Dondini L, Pagliarani G, Gregori R, Tartarini S, Leumann L, Laurens F, Patocchi A. 2016. Development of SNP-based assays for disease resistance and fruit quality traits in apple (Malus × domestica Borkh.) and validation in breeding pilot studies. Tree Genetics and Genomes **12,** 12-35

Beekwilder J, Alvarez-Huerta M, Neef E, Verstappen FWW, Bouwmeester HJ, Aharoni A. 2004. Functional Characterization of Enzymes Forming Volatile Esters from Strawberry and Banana. Plant Physiology **135**, 1865–1878.

Bennett AB, Labavitch JM. 2008. Ethylene and ripening-regulated expression and function of fruit cell wall modifying proteins. Plant Science **175**, 130–136.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289–300.

Bianco L, Cestaro A, Sargent DJ, et al. 2014. Development and validation of a 20K Single Nucleotide Polymorphism (SNP) whole genome genotyping array for apple (Malus × domestica Borkh.). PLoS ONE **9**.

Biasioli F, Yeretzian C, Märk TD, Dewulf J, Van Langenhove H. 2011. Direct-injection mass spectrometry adds the time dimension to (B)VOC analysis. Trends in Analytical Chemistry **30**, 1003–1017.

Brummell DA. 2006. Cell wall disassembly in ripening fruit. Functional Plant Biology **33**, 103–119.

Brummell DA, Harpster MH. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. Plant Molecular Biology **47**, 311–340.

Buettner A, Otto S, Beer A, Mestres M, Schieberle P, Hummel T. 2008. Dynamics of retronasal aroma perception during consumption: Cross-linking on-line breath analysis with medico-analytical tools to elucidate a complex process. Food Chemistry **108**, 1234–1246.

Cappellin L, Biasioli F, Granitto PM, Schuhfried E, Soukoulis C, Costa F, Märk TD, Gasperi F. 2011a. On data analysis in PTR-TOF-MS: From raw spectra to data mining. Sensors and Actuators B: Chemical **155**, 183–190.

Cappellin L, Biasioli F, Schuhfried E, Soukoulis C, Mark TD, Gasperi F. 2011b. Extending the dynamic range of proton transfer reaction time-of-flight mass spectrometers by a novel dead time correction. Rapid Communications in Mass Spectrometry **25**, 179–183.

Cappellin L, Soukoulis C, Aprea E, Granitto P, Dallabetta N, Costa F, Viola R, Märk TD, Gasperi F, Biasioli F. 2012. PTR-ToF-MS and data mining methods: A new tool for fruit metabolomics. Metabolomics **8**, 761–770.

Cappellin L, Farneti B, Di Guardo M, Busatto N, Khomenko I, Romano A, Velasco R, Costa G, Biasioli F, Costa F. 2015. QTL Analysis Coupled with PTR-ToF-MS and Candidate Gene-Based Association Mapping Validate the Role of Md-AAT1 as a Major Gene in the Control of Flavor in Apple Fruit. Plant Molecular Biology Reporter **33**, 239–252.

Chen AR., Chase JT. 1993. Alcohol dehydrogenase 2 and pyruvate decarboxylase induction in ripening and hypoxic tomato fruit. Plant Physiology and Biochemistry **31**, 875–885.

Cliff MA, Stanich K, Lu R, Hampson CR. 2016. Use of descriptive analysis and preference mapping for early-stage assessment of new and established apples. Journal of the Science of Food and Agriculture **96**, 2170–2183.

Cline MS, Smoot M, Cerami E, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. Nature protocols **2**, 2366–2382.

Contreras C, Beaudry R. 2013. Lipoxygenase-associated apple volatiles and their relationship with aroma perception during ripening. Postharvest Biology and Technology **82**, 28–38.

Cossins EA. 1978. Ethanol metabolism in plants. In: Hook, D.D. Crawford, R.M.M (Eds.), Plant Life in Anaerobic Environments. Science Publishers, Ann Arbor, MI, pp. 169-202

Costa F, Stella S, Van De Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S. 2005. Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus × domestica Borkh.). Euphytica **141**, 181–190.

Costa F, Alba R, Schouten H, et al. 2010. Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. BMC plant biology **10**, 229.

Costa F, Cappellin L, Longhi S, et al. 2011. Assessment of apple (Malus × domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biology and Technology **61**, 21–28.

Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F. 2012. Texture dynamics during postharvest cold storage ripening in apple (Malus × domestica Borkh.). Postharvest Biology and Technology **69**, 54–63.

Costa F, Cappellin L, Zini E, Patocchi A, Kellerhals M, Komjanc M, Gessler C, Biasioli F. 2013. QTL validation and stability for volatile organic compounds (VOCs) in apple. Plant Science **211**, 1–7.

De Filippi BG, Kader AA, Dandekar AM. 2005. Apple aroma: alcohol acyltransferase, a rate limiting step for ester biosynthesis, is regulated by ethylene. Plant Science **168**, 1199-1210.

Dewulf J, Van Langenhove H, Wittmann G. 2002. Analysis of volatile organic compounds using gas chromatography. Trends in Analytical Chemistry **21**, 637–646.

Di Guardo M, Micheletti D, Bianco L, et al. 2015. ASSIsT: An automatic SNP scoring tool for in- and outbreeding species. Bioinformatics **31**, 3873–3874.

Di Guardo M, Bink CAM M, Guerra W, et al., 2016. Deciphering the genetic control of fruit texture in apple by multiple family based analysis and genome-wide association. Journal of Experimental Botany (submitted)

Dimick PS, Hoskin JC. 1983. Review of apple flavor - state of the art. Critical reviews in food science and nutrition **18**, 387–409.

Dunemann F, Ulrich D, Boudichevskaia a., Grafe C, Weber WE. 2009. QTL mapping of aroma compounds analysed by headspace solid-phase microextraction gas chromatography in the apple progeny 'Discovery' × 'Prima'. Molecular Breeding **23**, 501–521.

Dunemann F, Ulrich D, Malysheva-Otto L, Weber WE, Longhi S, Velasco R, Costa F. 2012. Functional allelic diversity of the apple alcohol acyl-transferase gene MdAAT1 associated with fruit ester volatile contents in apple cultivars. Molecular breeding **29**, 609–625.

El Hadi MAM, Zhang FJ, Wu FF, Zhou CH, Tao J. 2013. Advances in fruit aroma volatile research. Molecules **18**, 8200–8229.

El-Sharkawy I, Manríquez D, Flores FB, Regad F, Bouzayen M, Latché A, Pech JC. 2005. Functional characterization of a melon alcohol acyl-transferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. Plant Molecular Biology **59**, 345–362.

Farneti B, Alarcón AA, Cristescu SM, Costa G, Harren FJM, Holthuysen NTE, Woltering EJ. 2013. Aroma volatile release kinetics of tomato genotypes measured by PTR-MS following artificial chewing. Food Research International **54**, 1579–1588.

Farneti B, Khomenko I, Cappellin L, Ting V, Romano A, Biasioli F, Costa G, Costa F. 2015a. Comprehensive VOC profiling of an apple germplasm collection by PTR-ToF-MS. Metabolomics **11**, 838–850.

Farneti B, Khomenko I, Cappellin L, Ting V, Costa G, Biasioli F, Costa F. 2015b. Dynamic volatile organic compound fingerprinting of apple fruit during processing. LWT - Food Science and Technology **63**, 21–28.

Fellman JK, Rudell DR, Mattinson DS, Mattheis JP. 2003. Relationship of harvest maturity to flavor regeneration after CA storage of 'Delicious ' apples. Postharvest Biology and Technology **27**, 39–51.

Fidler JC. 1968. The Metabolism of Acetaldehyde by Plant Tissues. Journal of Experimental Botany **19**, 41–51.

Fuhrmann E, Grosch W. 2002. Character impact odorants of the apple cultivars Elstar and Cox Orange. Die Nahrung **46**, 187.

Gilliver PJ, Nursten HE. 1976. The source of the acyl moiety in the biosynthesis of volatile banana esters. Journal of the Science of Food and Agriculture **27**, 152–158.

Giovannoni J. 2001. Maturation and ripening. Annual Review of Plant Biology **52**, 725–749.

Goff SA, Klee HJ. 2006. Plant volatile compounds: Sensory cues for health and nutritional value? Science **311**, 815–819.

Harker FR, Kupferman EM, Marin AB, Gunson FA, Triggs CM. 2008. Eating quality standards for apples based on consumer preferences. Postharvest Biology and Technology **50**, 70–78.

Holland D, Larkov O, Bar-Ya'akov I, Bar E, Zax A, Brandeis E, Ravid U, Lewinsohn E. 2005. Developmental and varietal differences in volatile ester formation and acetyl-CoA: Alcohol acetyl transferase activities in apple (Malus domestica Borkh.) fruit. Journal of Agricultural and Food Chemistry **53**, 7198–7203.

Jordan A, Haidacher S, Hanel G, Hartungen E, Märk L, Seehauser H, Schottkowsky R, Sulzer P, Märk TD. 2009. A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). International Journal of Mass Spectrometry **286**, 122–128.

Jung S, Ficklin SP, Lee T, et al. 2014. The Genome Database for Rosaceae (GDR): Year 10 update. Nucleic Acids Research **42**.

Kallberg Y, Oppermann U, Jörnvall H, Persson B. 2002. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignments in completed genomes. European Journal of Biochemistry **269**, 4409–4417.

Karapinar M. 1990. Inhibitory effects of anethole and eugenol on the growth and toxin production of Aspergillus parasiticus. International Journal of Food Microbiology **10**, 193–199.

King GJ, Maliepaard C, Lynn JR, et al. 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.

Klee HJ. 2010. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. New Phytologist, 44–56.

Koeduka T, Fridman E, Gang DR, et al. 2006. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. Proceedings of the National Academy of Sciences of the United States of America **103**, 10128–10133.

Koeduka T, Orlova I, Baiga TJ, Noel JP, Dudareva N, Pichersky E. 2009. The lack of floral synthesis and emission of isoeugenol in Petunia axillaris subsp. parodii is due to a mutation in the isoeugenol synthase gene. Plant Journal **58**, 961–969.

Kumar S, Garrick DJ, Bink MCAM, Whitworth C, Chagné D, Volz RK. 2013. Novel genomic approaches unravel genetic architecture of complex traits in apple. BMC Genomics **14**.

Kumar S, Rowan D, Hunt M, Chagné D, Whitworth C, Souleyre E. 2015. Genome-wide scans reveal genetic architecture of apple flavour volatiles. Molecular breeding **35**.

Lindinger W, Hansel A, Jordan A. 1998. Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level. Chemical Society Reviews **27**, 347–354.

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus x domestica Borkh.). Journal of Experimental Botany **63**, 1107–21.

Longhi S, Cappellin L, Guerra W, Costa F. 2013. Validation of a functional molecular marker suitable for marker-assisted breeding for fruit texture in apple (Malus × domestica Borkh.). Molecular Breeding **32**, 841–852.

Maliepaard C, Sillanpää MJ, van Ooijen JW, Jansen RC, Arjas E. 2001. Bayesian versus frequentist analysis of multiple quantitative trait loci with an application to an outbred apple cross. Theoretical and Applied Genetics **103**, 1243–1253.

Moummou H, Tonfack LB, Chervin C, Benichou M, Youmbi E, Ginies C, Latché A, Pech JC, van der Rest B. 2012. Functional characterization of SlscADH1, a fruit-ripening-associated short-chain alcohol dehydrogenase of tomato. Journal of Plant Physiology **169**, 1435–1444.

Nepusz T, Yu H, Paccanaro A. 2012. Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods **9**, 471–472.

Newcomb RD, Crowhurst RN, Gleave AP, et al. 2006. Analyses of expressed sequence tags from apple. Plant Physiology **141**, 147–166.

Pasay C, Mounsey K, Stevenson G, Davis R, Arlian L, Morgan M, Vyszenski-Moher D, Andrews K, Mccarthy J. 2010. Acaricidal activity of eugenol based compounds against scabies mites. PLoS ONE **5**.

Persson B, Krook M, Jornvall H. 1995. Short-chain dehydrogenases/reductases. Advances in Experimental Medicine and Biology **372**, 383–395.

Pesis E. 2005. The role of the anaerobic metabolites, acetaldehyde and ethanol, in fruit ripening, enhancement of fruit quality and fruit deterioration. Postharvest Biology and Technology **37**, 1–19.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**, 945–959.

Rambla JL, Tikunov YM, Monforte AJ, Bovy AG, Granell A. 2014. The expanded tomato fruit volatile landscape. Journal of Experimental Botany **65**, 4613–4623.

Ramsay J, Silverman BW. 2005. Functional data analysis. Series in statistics: Springer.

de Roos KB. 2003. Effect of texture and microstructure on flavour retention and release. International Dairy Journal **13**, 593–605.

Rowan DD, Hunt MB, Alspach P a, Whitworth CJ, Oraguzie NC. 2009. Heritability and genetic and phenotypic correlations of apple (Malus domestica Borkh.) fruit volatiles in a genetically diverse breeding population. Journal of agricultural and food chemistry **57**, 7944–52.

Schaffer RJ, Friel EN, Souleyre EJF, et al. 2007. A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway. Plant physiology **144**, 1899–912.

Speirs J, Lee E, Holt K, Yong-Duk K, Scott NS, Loveys B, Schuch W. 1998. Genetic manipulation of alcohol dehydrogenase levels in ripening tomato fruit affects the balance of some flavor aldehydes and alcohols. Plant Physiology **117**, 1047-1058.

Tieman D, Bliss P, McIntyre LM, et al. 2012. The chemical interactions underlying tomato flavor preferences. Current Biology **22**, 1035–1039.

Tikunov YM, Molthoff J, de Vos RCH, et al. 2013. Non-smoky glycosiltransferase1 prevents the release of smoky aroma from tomato fruit. Plant Cell **25**, 3067–3078.

Ulrich D, Dunemann F. 2012. Towards the development of molecular markers for apple volatiles. Flavour and Fragrance Journal **27**, 286–289.

Varela P, Chen J, Fiszman S, Povey MJW. 2006. Crispness assessment of roasted almonds by an integrated approach to texture description: Texture, acoustics, sensory and structure. Journal of Chemometrics **20**, 311–320.

Velasco R, Zharkikh A, Affourtit J, et al. 2010. The genome of the domesticated apple (Malus × domestica Borkh.). Nature genetics **42**, 833–9.

Wakasa Y, Kudo H, Ishikawa R, Akada S, Senda M, Niizeki M, Harada T. 2006. Low expression of an endopolygalacturonase gene in apple fruit with long-term storage potential. Postharvest Biology and Technology **39**, 193–198.

Wang A, Yamakake J, Kudo H, Wakasa Y, Hatsuyama Y, Igarashi M, Kasai A, Li T, Harada T. 2009. Null mutation of the MdACS3 gene, coding for a ripening-specific 1-aminocyclopropane-1-carboxylate synthase, leads to long shelf life in apple fruit. Plant Physiology **151**, 391–399.

Yahia EM. 1994. Apple flavor. Horticultural Reviews **16**, 197-233

Yauk YK, Chagné D, Tomes S, Matich AJ, Wang MY, Chen X, Maddumage R, Hunt MB, Rowan DD, Atkinson RG. 2015. The O-methyltransferase gene MdoOMT1 is required for biosynthesis of methylated phenylpropenes in ripe apple fruit. Plant Journal **82**, 937–950.

Yu J, Pressoir G, Briggs WH, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics **38**, 203–208.

Zhu Y, Barritt BH. 2008. Md-ACS1 and Md-ACO1 genotyping of apple (Malus x domestica Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genetics and Genomes **4**, 555–562.

Zini E, Biasioli F, Gasperi F, Mott D, Aprea E, Märk TD, Patocchi A, Gessler C, Komjanc M. 2005. QTL mapping of volatile compounds in ripe apples detected by proton transfer reaction-mass spectrometry. Euphytica **145**, 269–279.

Ziosi V, Noferini M, Fiori G, Tadiello A, Trainotti L, Casadoro G, Costa G. 2008. A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit. Postharvest Biology and Technology **49**, 319–329.

Chapter 6

General Discussion

The results presented in this thesis provide novel insights on the genomic regulation of fruit texture and the production of Volatile Organic Compounds (VOCs). The entire experimental procedure matches with the statement reported by Zamir (2001) citing that: *'Plant breeding is* the art and science of the genetic improvement of crops to produce new varieties that have *increased productivity and quality'.* Besides the novel insights on the genomic regulation of fruit texture and the production of VOCs, in this thesis the importance of the definition of a robust marker set, the employment of a high-resolution phenotyping method and the design of an accurate experimental scheme is discussed.

The accuracy of genetic data is a fundamental requirement to enable reliable marker-trait association. To this end, a novel software tool, named ASSIsT, designed to provide a better and more precise filtering of SNPs has been presented and discussed in Chapter 2. Particular care was given to the markers showing additional alleles, thus presenting different segregation patterns compared to the usual bi-allelic SNPs. The existence of such additional alleles can be explained as (i) insertion or deletion spanning a polymorphic annealing site giving raise to null-alleles, and (ii) secondary polymorphism at the probe-site close to the primary targeted polymorphism giving raise to null-alleles or marker alleles that differ in signal intensity (Carlson et al., 2006; Pikunova et al., 2014). The presence of such additional alleles often results in the discarding of potential valuable markers, although of biological value (Illumina technical note). While the inclusion of markers with additional alleles occurred so far through a manual editing process (Montanari et al., 2013), with ASSIsT this operation became completely automatized. ASSIsT allows the identification of markers showing robust performances on the germplasm in analysis, the user can define either the stringency of the filtering performance and decide the inclusion/exclusion of the three-allelic markers. The software, moreover, is also provided with an easy-to-use graphic user interface, to enable its use to a wide range of researchers.

In marker-trait association analysis, the resolution and accuracy of phenotyping is recognized as a crucial factor. The choice of the phenotyping approach to use is fundamental and it depends on the cost/time effectiveness of the methods that need to be applied. Till now, the phenotyping of fruit texture was mainly based on mechanical measurements, while the assessment of the acoustic response requires more sophisticated equipment (Duizer, 2001). For this reason, the acoustic response of fruit texture was for the most sensorialy evaluated. This methodology, however, presents serious limitation due to the phenotypic scale used (often hedonic), thus related to pleasure or acceptability rather than on the intensity of attributes such as hardness, toughness, crispness and mealiness (Abbott et al., 1984). To overcome this limitation, in chapter 4 a fully equipped TAXT-AED texture analyser was employed (StableMicroSystem, Godalming, UK) to implement into the QTL analysis novel quality aspects. Following the same philosophy, also for the aroma of apple a novel and sophisticated instrument

was used, represented by a PTR-ToF-MS coupled with an artificial chewing device that simulates, in vitro, the mastication (and subsequent VOCs release) of an apple fruit portion. In such case, the sensory evaluation is even more difficult than for texture due to the high number of VOCs influencing apple aroma.

In this work for the analysis of marker-trait associations two methods were employed, namely pedigree based analysis (PBA) and genome wide association studies (GWAS). The pros and cons of the two different genetic approaches have already been outlined in the general introduction. A common positive feature of both methods is that the allelic variability investigated can virtually encompass the whole allelic variability present in nature. More than the type of association analysis, emphasis should be given to the choice of the individuals to include in the analysis. This decision should reflect the scientific question under investigation. If the purpose is the identification of novel markers suitable for marker assisted breeding (MAB), ideally the individuals should represent at least the actual allelic variability of the parental cultivars available for breeding purposes.

Fruit quality

Although varying with context, quality of fresh fruits includes several properties such as colour, texture, flavour and health promoting compounds (Abbott, 1999; Bourne, 2002; Costa et al., 2011; Farneti, 2014). Different genetic approaches have been carried out to date to get insight on the regulation of these different aspects of fruit quality (Maliepaard et al., 1998; Hampson et al., 2000; Kumar et al., 2012; Farneti et al., 2015). A better comprehension of the genomic regions regulating a trait of interest would provide useful information to researchers to gain knowledge about the physiology of a specific trait. The aim of this thesis was an in-depth study of two important quality aspects such as texture and aroma in apple, defining the genomic regions involved in the control of these features and the existing inter-traits interaction (Fig. 1).

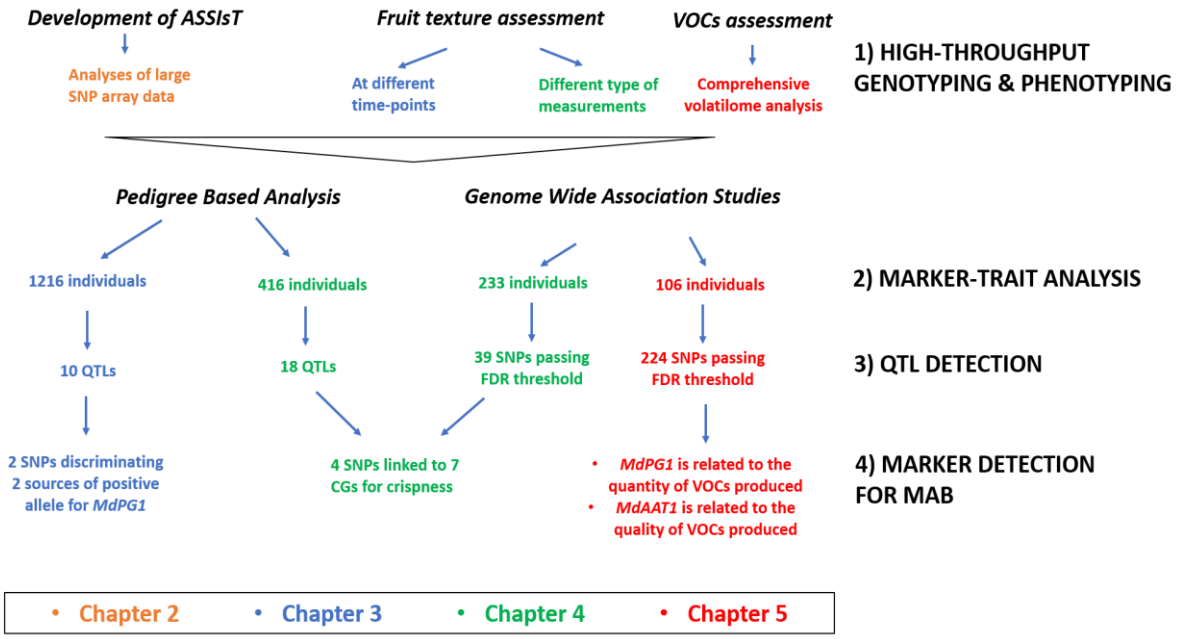


Figure 1. Flowchart of the upstream and downstream processes related to marker-trait association analysis as described in this PhD thesis. The major achievement in terms of software or molecular markers for MAB are outlined.

Breeding for fruit texture

Fruit texture represents one of the most investigated quality traits in apple because of its direct influence on both storability and consumer appreciation. Fruit texture is perceived by different senses: tactile and acoustic among others. The tactile - mechanical - component of texture is mainly perceived by the mouth and will be further defined as firmness, while the acoustic component is perceived by the hearing receptors and will be addressed as crispness. Since the two components involve different senses, they are measured and analysed using different equipment. Penetrometer is the most widely used instrument to measure firmness while for crispness different approaches are available (see Vincent, 1998 and Duizer, 2001 for more detailled reviews). In the presented thesis, a TAXT-AED texture analyser equipped with two distinct devices (StableMicroSystem, Godalming, UK) was employed to assess with a high resolution both firmness and crispness.

To address both the mechanical and acoustic component of texture, two different experimental plans were designed. While chapter 3 focused on the dissection of the genomic regions involved in the change of fruit firmness during storage, Chapter 4 was instead devoted to an in-depth analysis of the two main components influencing fruit texture: fruit firmness and fruit crispness. Although both chapters were focused on the textural properties of apple, the two experiments were designed to answer distinct, although complementary, scientific and practical

questions. Chapter 3 presents a detailed analysis of the different genomic regions associated with the firmness of fruits after four months of cold storage. This provided useful information about the storage performance of the largest germplasm collection assessed to date representing different breeding programmes from different countries from all over Europe. Chapter 4 was instead designed on the assumption that traditional phenotyping methods, based on the use of a penetrometer, do not provide enough resolution to capture the entire phenotypic complexity of fruit texture. In fact, although fruit firmness is an important and appreciated component of fruit texture, fruit crispness accounts for 90% of the variation in texture appreciation (Hampson et al., 2000), which is a key trait influencing consumers' preferences.

Fruit firmness changes during storage. Ripening comprises a complex series of physiological changes eventually leading to fruits with more attractive and palatable aspects (Brummell and Harpster, 2001; Costa et al., 2010). However, an excessive loss of firmness is responsible for undesirable post-harvest deterioration, limiting storability, transportability, shelf life and overall consumer appreciation. Insight into the genomic regions related to the control of fruit firmness during storage could represent a valuable tool for setting-up novel breeding programmes aimed to develop new apple varieties with superior storability properties. Better storage performances will have direct positive repercussions on the reduction of fruit loss during post-harvest management. The mapping of QTLs associated to fruit firmness assessed at four different time points (from harvest to four months of cold storage) shed light on the genomic regions possibly regulating this phenomenon. Three QTLs (located on chromosome 10, 14 and 15) remained stable across the whole storage period, while seven other QTLs were detected in specific moments only (Table 1, Chapter 3). Within the QTL intervals on chromosome 10 and 14 (named FF-LG10b and FF-LG14, respectively) several candidate genes were identified. It is worth noting that the haploblock with the highest significance for FF-LG10b harbours the polygalacturonase gene MdPG1, known to encode a protein involved in the middle lamella pectin degradation and consequently in the control of firmness loss in apple (Costa et al., 2010). The work presented by Costa and colleagues identified FF-LG10b through the analysis of one specific segregating family. MAB tool development benefits from a multi-family experimental plan since different sources of positive alleles can be identified and used in breeding programmes, thereby taking into account a much larger genetic diversity consequently limiting genetic erosion. The loss of genetic variability represent a serious problem in crops like apple: approximately 7000 varieties were grown in the United Stated at the beginning of the 20th century (Noiton and Alspach, 1996) while, one century later, the number of cultivated apples dropped to less than 400 (Gepts, 2006) although intensive apple production is dominated by few cultivars (Way, 1991). The discovery of one or more QTLs influencing the trait under investigation is a necessary, but not sufficient, condition for marker assisted breeding.

Successive steps are indeed needed to develop and validate low cost, high throughput genetic tests that can be used for routine MAB applications (Peace et al., 2017). For this reason, the QTL discovery described in chapter 3 was followed by a detailed analysis on the predictive value of the SNP markers most tightly linked to the FF-LG10b QTL. This resulted in a hypothesis on the presence of two independent mutations in the MdPG1 gene that each led to a QTL allele for increased firmness. Their most senior representative founders were the USA cv 'Esopus Sptzenburg' and the Ukrainian cv 'Borowitsky' (also known as 'Duchess of Oldenburg'). In our study, both sources for a positive allele had a similar effect on firmness. For each, a single highly predictive SNP could be identified.

Fruit texture components: most of the published QTLs associated to fruit texture were detected focusing mainly on firmness. The improvement of fruit firmness had a great impact on the reduction of the fruit losses during storage and transportation, but a relatively lower impact on what concerns consumer appreciation (Hampson et al., 2000). The reason behind the high number of studies focusing on fruit firmness only can be led back to the phenotyping approach employed to date. The most widely used instrument for fruit texture phenotyping is a penetrometer, which measures the force needed to penetrate the probe of the instrument into a portion of fruit flesh (Harker et al., 1996). For a long time, the penetrometer represented the instrument of choice/preference for such measurements thanks to its cost-effectiveness, portability and practice of use. On the other hand, it suffers from serious limitations in terms of phenotyping resolution, since what it measures is essentially the hardness of the flesh (Costa et al., 2011). Technological developments have extended our ability to describe complex biological systems such as fruit texture. For this reason, in chapter 4 and 5, the fruit firmness was assessed with more sophisticated equipment. It is already well known that one of the most important genes influencing apple textural properties is MdPG1 (Costa et al., 2010; Longhi et al., 2012, 2013; Chapter 3); therefore, it is not surprising that most of the elite cultivars currently used as parental lines in novel breeding schemes are homozygous for the positive allele of this gene ('Fuji', 'Pinova' and 'Cripps Pink' for instance). For this reason, emphasis was given to the definition of a subset of elite cultivars that can represent the actual allelic variability used in modern breeding. Fruit texture is regulated by many different genes, even though MdPG1 is one of the most important ones. Other genes can also play important, although minor, roles with respect to textural changes, whose effect can be masked by the effect of MdPG1. Molecular markers in strong linkage disequilibrium (LD) with MdPG1 (MdPG1$_{SSR}$,Md-PG1-SNP) (Longhi et al., 2013a; Baumgartner et al., 2016) represent a valuable tool for novel cultivar selection. However, in modern breeding programs using these selections as parental cultivars the favourable alleles of these markers will become fixed (homozygous allelic configuration), requiring the identification of new markers associated to a different quote of phenotypic

variance. The analyses described in Chapter 4 were conducted on a subset of elite cultivars using a GWAS approach as well as on 6 full-sib employing a PBA approach. The choice of the germplasm was conducted specifically to explore the portion of phenotypic variability not influenced by MdPG1. Textural changes within cultivars were analysed using twelve different parameters (distinctive of both the mechanical or the acoustic components of texture), and interesting QTL intervals located on chromosome 2, 14 and 15 were identified, representing potential regions for the design of markers for the selection of cultivars with superior crispness performance.

Breeding for apple aroma

Flavour is the result of the interaction between taste and aroma. The latter represents one of the distinctive characteristics making apples recognizable and particularly appreciated by consumers (Baldwin, 2002). Despite this, breeding for aroma did not register the same interest such as other traits related for instance to production or resistance against biotic or abiotic stresses. The reason can be assigned to the fact that breeding has historically favoured traits related to yield and appearance. At the same time, it needs to be taken into consideration that breeding for aroma is hampered by its complex genetic control and the need for sophisticated equipment for phenotyping. Also breeding for fruit yield, size, colour or resistance might have had an unintended negative repercussion on aroma, in line to what has been already described for other crops such as strawberry, peach and tomato (Goff and Klee, 2006; Klee, 2010; Rambla et al., 2014). This tendency was confirmed by the multiple factor analysis presented in Chapter 5 (Fig.4). The projection of the loadings related to VOCs and textural components have opposite directions, resulting in an inverse correlation between positive textural properties and high VOCs release. Furthermore, apples with good aroma (and poor texture) were mostly ancient cultivars confirming the somewhat negative repercussions that modern breeding programmes had on aroma properties. GWAS analysis using the whole VOC pattern produced during chewing highlighted the different genetic regulations influencing the quantity and the quality of the VOCs released. In an attempt to distinguish several apple cultivars on the basis of their aromatic fingerprinting through a Principal Component Analysis, two PCs were identified. The first (PC1) was specifically associated to the quantitative production of the entire aromatic bouquet, while the second (PC2), explaining a lower quote of phenotypic variance, was more related to a qualitative distinction of the type of aroma, especially defined by the ester/alcohol ratio. The implementation of these two components into a GWAS investigation unravelled the interplay existing between aroma and the structural properties of the cell wall. When the PC1 was in fact implemented, a significant association was identified on chromosome 10, in LD with MdPG1, validating the theory that a different structure of the fruit

cells might influence the aromatic release, possibly due to a loss of cellular compartmentation. The PC2, on the other hand, enabled the detection of another QTL, mapped in the proximity of MdAAT1, a gene involved in the biosynthesis of esters, which account for the 80% of the entire aromatic bouquet in apple fruit (Newcomb et al., 2006; Ulrich and Dunemann, 2012). Since the aroma of ripe apple is essentially the result of the ratio between alcohols and esters, the role of MdAAT1 seems to be more related to the quality determination of the aromatic profile rather than the quantity of the VOCs produced. The identification of these two QTLs might represent the opportunity to guide a more accurate selection of novel ideotypes combining valuable and appreciated fruit quality features.

Conclusion and further perspectives

For many economic sectors, globalisation resulted in a proliferation of industries that leads to what some economists have described as a state of hypercompetition (D'Aveni, 1998; Axelson and Axelson, 2000). To survive in such a competitive market, apple industries are creating novel varieties with superior fruit quality characteristics. This process is often accompanied by the differentiation of new products in brands held by private clubs (Pink Lady®, the commercial name of the variety 'Cripps Pink' is an example) to make the new product captivating and clearly recognizable by consumers (Harker et al., 2003) and to keep prices attractive to producers by controlling production volumes. The demand of new cultivars is also stimulated by two other factors. The consumers continuously seek for new flavour and taste, generating thus novel requests of accessions with desirable traits. The second aspect is represented by the problems associated to climate change. The global variation we are experiencing in terms of global warming lead to important modifications in the environment, making the cultivation of dedicated species less profitable because of lower quality and/or quantity than before. To face this problem, without changing the horticultural settings, novel and more resilient accessions, characterized by higher performance, are needed.

Breeding, especially in fruit crops, is a time-consuming process. The result of traditional breeding is, despite improvements in the breeding process such as MAB, still difficult to predict. This is mainly due to the heterozygosity of most of the genomes of these species as well as the complexity of most agronomical important traits of which our genetic knowledge is still limited. The use of molecular markers to help and guide selection cannot be simply reconducted to finding a statistical association between a marker and a trait. Great attention should be devoted to the upstream analysis related to the quality control of both phenotypic and genotypic data and, after the detection of a candidate molecular marker, to its validation on different genetic backgrounds and environments. Once a molecular marker proved its efficiency in different environments and genetic backgrounds, it can become a valuable tool to assist breeding in the

selection of novel valuable accessions. In classical breeding for apple, the seedlings originated from a controlled cross are planted in open field and their phenotypes evaluated over a period of 8-10 years. The identification of a molecular marker linked to the phenotype of interest can allows an early selection of the seedlings (screened at a 3-4 leaves stage), resulting in a much more efficient space management, since only the individuals with favourable genotype for the trait of interest will be considered. Other horticultural strategies can be potentially coupled to marker selection to fasten the breeding fruit bearing processes, such as in vitro flowering and induced flowering under proper greenhouse conditions. The ultimate potential use of marker-trait associations can potentially lead to a complete chromosome haplotyping for the species of interest enabling the so-called Breeding By Design, in which novel cultivars can be created by combining positive alleles for all agronomical traits (Fig. 2).

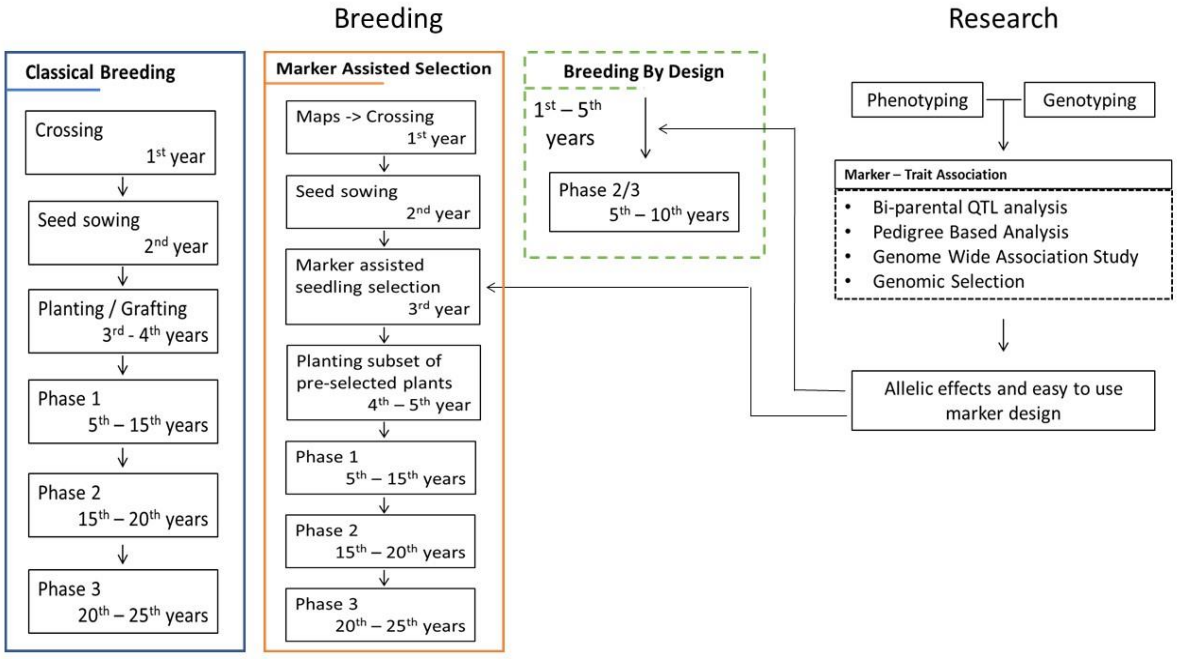


Figure 2. Proposed workflow and interaction between research and different types of breeding selection strategies

As a future perspective, raw phenotypic data should be collected in public databases making them accessible to researchers worldwide. Such a public accessible repository will greatly help researchers to learn from what has been done in the past when making decisions about future breeding strategies (Zamir, 2013). This thesis could represent a valuable tool for breeders presenting several molecular markers in linkage with agronomical traits of interest. However, these finding have still to be confirmed on larger germplasm sets and time frames. On a longer term, the lack of an accurate marker validation can weaken the impact of the finding reported in scientific papers and represent a break point between the breeding sector and

academia. This last point should be taken into serious consideration since breeding makes use of what is discovered by applied science and applied science (such as part of this thesis) has its raison d'être on the application in breeding.

References

Abbott JA, Watada AE, Massie DR. 1984. Sensory and Instrument Measurement of Apple Texture. Journal of the American Society for Horticultural Science **109**, 221–228.

Abbott JA. 1999. Quality measurement of fruits and vegetables. Postharvest Biology and Technology **15**, 207–225.

Axelson L, Axelson J. 2000. Hypercompetition on horticultural marjets. Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, 485–492.

Baldwin EA 2002. Fruit flavor, volatile metabolism and consumer perceptions. CRC Press: Boca Raton, FL.

Baumgartner IO, Kellerhals M, Costa F, Dondini L, Pagliarani G, Gregori R, Tartarini S, Leumann L, Laurens F, Patocchi A. 2016. Development of SNP-based assays for disease resistance and fruit quality traits in apple (Malus × domestica Borkh.) and validation in breeding pilot studies. Tree Genetics and Genomes **12**.

Bourne M. 2002. Food texture and viscosity: concept and measurement. Academic press.

Brummell DA, Harpster M. 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. In: Carpita, N.C. and Campbell, M. and Tierney M, ed. Plant Cell Walls. Springer Netherlands, 311–340.

Carlson CS, Smith JD, Stanaway IB, Rieder MJ, Nickerson DA. 2006. Direct detection of null alleles in SNP genotyping data. Human molecular genetics **15**, 1931–1937.

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, Van de Weg WE. 2010. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus x domestica Borkh.). Journal of Experimental Botany **61**, 3029–3039.

Costa F, Cappellin L, Longhi S, et al. 2011. Assessment of apple (Malus × domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biology and Technology **61**, 21–28.

D'Aveni RA. 1998. Waking up to the new era of hypercompetition. The Washington Quarterly **21**, 183–195.

Duizer L. 2001. A review of acoustic research for studying the sensory perception of crisp, crunchy and crackly textures. Trends in Food Science and Technology **12**, 17–24.

Farneti B. 2014. Tomato quality: from the field to the consumer Interactions between genotype, cultivation and postharvest conditions. PhD thesis, Wageningen University, Wageningen, NL ISBN 978-94-6257-080-1

Farneti B, Busatto N, Khomenko I, Cappellin L, Gutierrez S, Spinelli F, Velasco R, Biasioli F, Costa G, Costa F. 2015. Untargeted metabolomics investigation of volatile compounds involved in the development of apple superficial scald by PTR-ToF--MS. Metabolomics **11**, 341–349.

Gepts P. 2006. Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. Crop Science **46**, 2278–2292.

Goff SA, Klee HJ. 2006. Plant Volatile Compounds: Sensory Cues for Health and Nutritional Value? Science **311**, 815–819.

Hampson CR, Quamme HA, Hall JW, MacDonald RA, King MC, Cliff MA. 2000. Sensory evaluation

as a selection tool in apple breeding. Euphytica **111**, 79–90.

Harker FR, Maindonald JH, Jackson PJ. 1996. Penetrometer measurement of apple and kiwifruit firmness: operator and instrument differences. Journal of the American Society for Horticultural Science **121**, 927–936.

Harker FR, Gunson FA, Jaeger SR. 2003. The case for fruit quality: An interpretive review of consumer attitudes, and preferences for apples. Postharvest Biology and Technology **28**, 333–347.

Illumina Technical Note (https://support.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf)

Klee HJ. 2010. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. New Phytologist **187**, 44–56.

Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C. 2012. Genomic selection for fruit quality traits in apple (Malus x domestica Borkh.). PLoSONE **7**, 1–10.

Longhi S, Moretto M, Viola R, Velasco R, Costa F. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus x domestica Borkh.). Journal of Experimental Botany **63**, 1107–1121.

Longhi S, Cappellin L, Guerra W, Costa F. 2013a. Validation of a functional molecular marker suitable for marker-assisted breeding for fruit texture in apple (Malus × domestica Borkh.). Molecular Breeding **32**, 841–852.

Longhi S, Hamblin MT, Trainotti L, Peace CP, Velasco R, Costa F. 2013b. A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (Malus x domestica Borkh). BMC plant biology **13**, 37.

Maliepaard C, Alston FH, Van Arkel G, et al. 1998. Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers. Theoretical and Applied Genetics **97**, 60–73.

Montanari S, Saeed M, Knäbel M, et al. 2013. Identification of Pyrus single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and Interspecific Pyrus Hybrids. PLoS ONE **8**, 1–11.

Newcomb RD, Crowhurst RN, Gleave AP, et al. 2006. Analyses of expressed sequence tags from apple. Plant physiology **141**, 147–166.

Noiton DAM, Alspach PA. 1996. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. Journal of the American Society for Horticultural Science **121**, 773–782.

Peace CP, Hayes K, Miller J, Trexler C, Bliss F, Carlisle C. 2017. DNA-informed breeding of rosaceous crops: promises, progress and prospects. Horticulture Research **4**, 17006.

Pikunova A, Madduri M, Sedov E, Noordijk Y, Peil A, Troggio M, Bus VGM, Visser RGF, van de Weg E. 2014. 'Schmidt's Antonovka' is identical to 'Common Antonovka', an apple cultivar widely used in Russia in breeding for biotic and abiotic stresses. Tree genetics & genomes **10**, 261–271.

Rambla JL, Tikunov YM, Monforte AJ, Bovy AG, Granell A. 2014. The expanded tomato fruit volatile landscape. Journal of Experimental Botany **65**, 4613–4623.

Ulrich D, Dunemann F. 2012. Towards the development of molecular markers for apple volatiles. Flavour and fragrance journal **27**, 286–289.

Vincent JFV. 1998. The quantification of crispness. Journal of the Science of Food and Agriculture **78**, 162–168.

Way, RD, Aldwinckle, HS, Lamb RC, et al. 1991. Apples (Malus). Genetic Resources of temperate fruit and nut crops 290, 3-46.Zamir D. 2001. Improving plant breeding with exotic genetic libraries. Nature reviews. Genetics **2**, 983–989.

Zamir D. 2013. Where Have All the Crop Phenotypes Gone? PLoS Biology **11**, 1–4.

Summary

Although varying with context, quality of fresh fruits includes several properties such as color, texture, flavor and health promoting compounds. This thesis focused on two important quality aspects, namely texture and aroma in apple, and defining the genomic regions involved in the control of these two features. The genetic control of texture and VOCs production have been investigated using two marker-trait association analysis approaches: Pedigree Based Analysis (PBA) and Genome Wide Association Study (GWAS).

Chapter 2 presents ASSIsT (**A**utomatic **S**NP **S**cor**I**ng **T**ool), a software dedicated for the efficient calling and filtering of SNPs from Illumina Infinium® arrays. The availability of high-throughput SNP genotyping platforms makes the genome-wide scanning of large sample sets affordable and exhaustive. However, the genomic data produced requires ad-hoc software for reliable SNP scoring. ASSIsT builds on GenomeStudio® derived data and identifies markers showing reliable genotype calls (bi-allelic segregation pattern). In addition, ASSIsT identifies and re-edits SNP calls of markers showing additional alleles (null alleles or additional SNPs in the probe annealing site). ASSIsT can be employed in the analysis of several population types such as full-sib families, backcross, F_1, F_2 and unrelated individuals. The analysis settings can be customized making the filtering process more or less stringent and output data can be exported according to the format requested by the most widely used software for marker-trait association analysis.

Chapter 3 aimed to dissect the genetic control of fruit firmness in apple during storage through PBA and employing 24 bi-parental families (1216 individuals) connected by a common pedigree structure. Fruit firmness, was assessed over four postharvest storage periods (harvest; two months of cold storage; two months of cold storage and two weeks of shelf life; four months of storage). Ten QTLs were identified encompassing eight linkage groups, which unravelled a QTL dynamics over storage shedding light on the specific genetic control at each time-point. Three QTLs were common to all storage periods (FF-LG10$_b$, FF-LG14 and FF-LG15), five were common for two to three periods (FF-LG1$_a$, FF-LG1$_b$, FF-LG3, FF-LG6 and FF-LG10$_a$) while two QTLs were specific to harvest (FF-LG11) or four months of storage (FF-LG16). The IBD (Identity by Descent) analysis allowed to trace the founder sources of QTL-alleles. Emphasis was hereby given to QTL FF-LG10$_b$, a major QTL for fruit texture that

coincided with the polygalacturonase gene MdPG1. Analysis of the SNP haplotypes of these founders hypothesized the presence of two independent mutations, although with similar effect, on the MdPG1 gene.

Chapter 4: aimed to comprehensively decipher the genetic control of fruit texture. Two complementing QTL mapping approaches were employed together with a novel and high sophisticated phenotyping device for fruit texture. The PBA was carried out on six full-sib pedigreed families (416 individuals), while the GWAS was performed on a collection of 233 apple accessions. The texture analyser employed (TAXT-AED texture analyser) allowed the measurement of both the mechanical properties (firmness) and the acoustic properties (crispness) of fruit texture. The QTL results indicated chromosome 10 being associated in changes of the mechanical properties of fruit texture, while chromosomes 2 and 14 were more associated to the acoustic response. The in silico annotation of these intervals revealed the presence of candidate genes potentially involved in fruit texture regulation as suggested by the gene expression profiles.

In **Chapter 5** the interplay between texture and volatile organic compounds (VOCs) was investigated in 162 apple accessions. Texture was assessed using the TAXT-AED texture analyser of chapter 4, while VOCs production was analysed with a PTR-ToF-MS, coupled with an artificial chewing device. The array of volatile compounds was implemented into a GWAS identifying seven chromosomes harbouring important candidate genes for aroma, such as MdAAT1 and MdIGS. Next, volatilome and fruit texture data were integrated employing a Multiple Factor Analysis (MFA), revealing a negative correlation between these two features. The association analysis performed between these components and marker genotypes revealed two major QTLs. The PC1, associated to the quantitative production of the volatilome, enabled the identification of the first QTL on chromosome 10, coincident with the genetic position of MdPG1. The PC2, related to a qualitative distinction of the type of aroma, allowed instead the identification of another QTL on chromosome 2, collocated with MdAAT1, a gene involved in the biosynthetic pathway of esters.

In **Chapter 6** the novel insights outlined in chapters 3 to 5 on the genomic regions involved in changes in fruit texture and VOCs production were discussed and integrated in a marker assisted breeding (MAB) framework. MAB indeed is a process that required more than a decade to be completed and attention should be given to all the steps eventually leading to the commercialization of a novel variety (from phenotyping and genotyping to field trials). Novel

markers regulating textural performance and/or VOCs production were presented that could be used in new MAB programmes to enhance fruit textural properties or aroma of the apple of the future.

Acknowledgement

The thesis you just red (or you will read if, as I do, you prefer to go straight to the acknowledgement at first) is the result of five years of exciting and challenging work. Looking back to this period I wish to express my gratitude to all people that contributed to this achievement with their understanding, suggestions and support.

I would like to express my gratitude to my supervisor and promoter, Prof. Dr. Richard Visser, not only for giving me the possibility to conduct this study but also for his help and ability in putting my work in a wider context.

Many thanks also to you, Eric, one of the first persons I met in the Netherlands during my master thesis, thanks to you I learned about this PhD project.

I would like to thank Riccardo Velasco for his help in writing the project proposal and, more important, for his support.

A particular thanks to Fabrizio, who represented (and represents) much more than a supervisor. You are a reference both from a personal and a scientific point of view.

A big thanks to the colleagues at FEM: Daniele, Lucia, Annarita, Iuliia, Alice, Matteo, Sudarshan, Carmen and Samia for making FEM an amazing place to work and have fun. I would like to thank also Michela, Luca, Diego, Dan and Kristof, the doors of your offices were always open to discuss about work or any other topics. A special thanks to Brian for the nice discussions and Nicola, my favourite photographer.

A special thanks to the amazing people I met at the animal science group at the Università Cattolica del Sacro Cuore: professor Ajmone Marsan, Licia, Riccardo, Mario, Elia, Roberta, Elisa, Marcello, Marco, Solange and Chiara for letting me feel at home in a city and a job that were not familiar to me.

Ad Alessandra, Stefano, Gaetano, Cetti e Marco per la loro amicizia e vicinanza durante questi anni di dottorato.

Ad Annalisa ed Enrico che mi hanno accolto come se ci conoscessimo da sempre. A Lori, Mimmo, Antonietta ed Enrico, parenti di affetto.

I am especially thankful to my friends, from the ones with whom I shared memories from the kindergarten to the ones I met just few months ago, I hope that the saying that the value of a person is measured on the quality of his friends is true. The first page of this thesis is dedicated to you.

Ad Andrea e Caterina, perché i nostri cuori battono all'unisono.

A mia mamma e mio papà, che percorrono la vita tenendosi per mano e che ai diplomi, preferiscono incorniciare biglietti di treno.

A Lara, mio presente e mio futuro che mi ha insegnato a declinare la vita in prima persona plurale e a mio nonno per aver creato la comunità degli affetti.

Curriculum vitae

Mario Di Guardo was born on December 25th, 1986, in Catania, Italy. In 2009, he graduated cum laude at the University of Agronomical study of Catania with the thesis entitled 'Using of plants for soil remediation from heavy metal and organic pollutants'. He attended the master in Plant Biotechnology at the University of Florence. In 2011, he won an Erasmus studio scholarship and spent six months at the Plant Breeding department of the Plant Research International (Wageningen UR). He graduated cum laude in 2012 with the thesis entitled: 'Microsatellite scoring in tetraploid rose and determination of allele dosage'. In August 2012, he started a PhD project entitled: 'Fine mapping and genetic dissection of important apple fruit quality trait by pedigree based analysis'; the results of this work are presented in this thesis. From November 2016, he is employed as Postdoc at the Institute of Zootechnic at the Università Cattolica del Sacro Cuore, Piacenza (Italy) to work on GWAS analysis and gene annotation of qualitative traits linked to milk production on dairy cattle.

Publications

Paper published in refereed journals

Di Guardo M, Bink MCAM, Guerra W, Letchcka T,Lonzano L, Busatto N, Poles L, Tadiello A, Visser R, van de Weg E, Costa. Multiple-Family Pedigree Based Analysis and Genome Wide Association Study decipher the genetic control of mechanical and acoustic textural components in apple fruit. JExBot 2017, doi: 10.1093/jxb/erx017

Farneti B, **Di Guardo M**, Khomenko I. Costa F. Genomewide association study coupled to high performing PTRToF-MS unraveled the genetic interplay between fruit texture and volatilome in apple fruit. JExBot 2017, doi: 10.1093/jxb/erx018

Di Pierro EA, Gianfranceschi L, **Di Guardo M**, Herma JJ Koehorst-van Putten, Johannes W Kruisselbrink, Sara Longhi, Michela Troggio, Luca Bianco, Hélène Muranty, Giulia Pagliarani, Stefano Tartarini, Thomas Letschka, Lidia Lozano Luis, Larisa Garkava-Gustavsson, Diego Micheletti, Marco CAM Bink, Roeland E Voorrips, Ebrahimi Aziz, Riccardo Velasco, François Laurens & W Eric van de Weg. A high-density, multi-parental SNP genetic

map on apple validates a new mapping approach for outcrossing species. Horticulture Research 2017, 3, 16057.

Allard A, Bink MCAM, Martinez S, Kelner JJ, Legave JM, **Di Guardo M**, Di Pierro EA, Laurens F, van de Weg EW, Costes E: Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multi-parental population. JExBot 2016, doi: 10.1093/jkb/erw130

Muranty H, Troggio M, Sadok I Ben, Rifaï M Al, Auwerkerken A, Banchi E, Velasco R, Stevanato P, van de Weg WE, **Di Guardo M**, Kumar S, Laurens F, Bink MCAM: Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res 2015, 2:15060.

Di Guardo M, Micheletti D, Bianco L, Koehorst- van Putten HJJ, Longhi S, Costa F, Aranzana MJ, Velasco R, Arús P, Troggio M, van de Weg EW: ASSIsT: An Automatic SNP ScorIng Tool for in- and outbreeding species. Bioinformatics 2015: btv446.

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, **Di Guardo M**, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troggio M: Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus × domestica Borkh). PLoS One 2014, 9:e110377.

Cappellin L, Farneti B, **Di Guardo M**, Busatto N, Khomenko I, Romano A, Velasco R, Costa G, Biasioli F, Costa F: QTL Analysis Coupled with PTR-ToF-MS and Candidate Gene-Based Association Mapping Validate the Role of Md- AAT1 as a Major Gene in the Control of Flavor in Apple Fruit. Plant Mol Biol Report 2014, 33:239–252

Di Guardo M, Tadiello A, Farneti B, Lorenz G, Masuero D, Vrhovsek U, Costa G, Velasco R, Costa F: A Multidisciplinary Approach Providing New Insight into Fruit Flesh Browning Physiology in Apple (Malus x domestica Borkh.). PLoS One 2013, 8:1–15.

Vukosavljev M**, Di Guardo M**, van de Weg WE, Arens P, Smulders MJM. Quantification of Allele Dosage in tetraploid Roses. 2013, ScienceMed

Paper submitted

van de Weg E, **Di Guardo M**, Jänsch M, Socquet-Juglard D, Costa F, Broggini G, Kellerhals M, Troggio M, Laurens F, Durel CE, Patocchi A: Epistatic fire blight resistance QTL alleles in the apple varieties ‘Enterprise’ and X-6398 discovered and characterized through pedigree-informed analysis (submitted).

The research described in this thesis was financially supported by Edmund Mach Foundation (GMPF program), Plant Breeding Wageningen University and Research and by the EU seventh Framework Programme by the FruitBreedomics Project No. 265582: Integrated Approach for increasing breeding efficiency in fruit tree crops.

Cover design was done by agilecolor design studio/atelier (www.agilecolor.com) and the thesis layout by the author

Printed by GVO, Ede, The Netherlands