ALMA MATER STUDIORUM-UNIVERSITÀ DEGLI STUDI DI BOLOGNA

DOTTORATO DI RICERCA IN COLTURE ARBOREE ED AGROSISTEMI
FORESTALI ORNAMENTALI E PAESAGGISTICI
CICLO XXV
SETTORE SCIENTIFICO-DISCIPLINARE: AGR/03
SETTORE CONCORSUALE DI AFFERENZA: 07/B2

Molecular strategies for genetic diversity analysis and development of markers linked to resistance traits in apple

Presented by

WEI LIANG

Tutor:
Dott. Stefano Tartarini

Coordinator:
Prof. Luca Corelli Grappadelli

Co-tutor:
Dott. Luca Dondini

April 2013

Table of Contents

Chapter 1

General Introduction

1.1. Origin of the Domesticated Apple

The domesticated apple (*Malus × domestica* Borkh.) is belonging to the Rosids superorder, which include one-third of all flowering plants (Hummer and Janick 2009), and in particular to the Rosaceae family which include different forms with its closely related fruit (*Pyrus* and *Cydonia*) and ornamental (*Amelanchier, Aronia, Chaenomeles, Cotoneaster, Crateagus, Pyracantha* and *Sorbus*) genera (Challice 1974). There is no agreement among taxonomists as to how many species the *Malus* genus comprises, as the different species are widely compatible and readily interbreed (Korban 1986). Most of the *Malus* species are diploids (2n = 34), but a few are triploid (e.g., *M. hupehensis* and *M. coronaria*), or tetraploid (e.g., *M. sargentii*), while some species show variable levels of ploidy (Way et al., 1989). The characteristic inter-fertility among *Malus* species has been deployed in apple breeding for the introgression of pest and disease resistance genes. For this reason as well as the assumed inter-specific origin of the eating apple in general (Korban 1986; Korban and Chen 1992; Robinson et al*.,* 2001), it seems appropriate to identify the domesticated apple as *M.* x *domestica* Borkh. A survey of the molecular differences at 23 genes across the genus *Malus* clearly showed that the *M*. × *domestica* cultivars appear more closely related to accessions of the wild species *M. sieversii* and less closely related to accessions of *M. sylvestris*, *M. baccata*, *M. micromalus* and *M. prunifolia.* Furthermore, these data support the formation of the *M. × domestica* gene pool directly from *M. sieversii* (Velasco et al., 2010). Whereas the haploid (x) chromosome numbers of most Rosaceae are 7, 8 or 9, Pyreae have a distinctive x = 17. Pyreae have long been considered an example of allopolyploidization between species related to extant Spiraeoideae (x = 9) and Amygdaleoideae (x = 8) (Velasco et al., 2010). The pairwise comparison of 17 apple chromosomes highlighted strong collinearity between large segments of some chromosomes pairs (3-11, 5-10, 9-17, and 13-16) , and between shorter segments of other

chromosomes (1 -7, 2-7, 2-15, 4-12, 12-14, 6-14, and 8-15; Velasco et al. 2010; see figure 1.1).

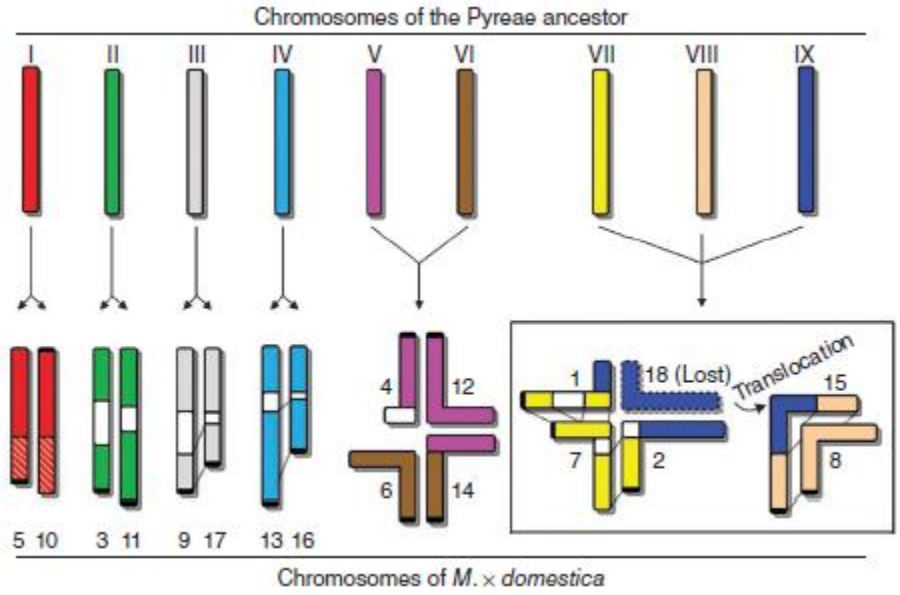


Figure 1.1. A model explaining the evolution from a 9-chromosome ancestor to the 17-chromosome karyotype of extant Pyreae, including the genus *Malus*. A Genome wide duplication followed by a parsimony model of chromosome rearrangements is postulated. Shared colors indicate homology between extant chromosomes. White fragments of chromosomes indicate lack of a duplicated counterpart. The white-hatched portions of chromosomes 5 and 10 indicate partial homology (modified from Velasco et al., 2010).

Therefore, a recent genome wide duplication has shaped the genome of the domesticated apple (Velasco et al., 2010), as already hypothesized by Evans and Campbell (2002) that first proposed a within-lineage polyploidization event to explain the chromosome number of apple.

1.2. Breeding Strategy

The science of breeding started with the first controlled cross-pollination carried out by Thomas Knight early in the nineteenth century (Brown, 1975). The traditional method of

apple breeding by selecting the best phenotypes from seedlings grown from open-pollinated seeds was replaced by deliberate hybridization about 200 years ago. However, initially little progress was made in improving apple cultivars through controlled crossing, which has been attributed to poor selection of parents (Janick et al*.,* 1996). Apple is self-incompatible and highly heterozygous, which results in very diverse progeny with only a few of them being a major improvement of the parents. As most characters are under polygenic control, low efficiency in genetic improvement of breeding lines together with a long juvenile period make breeding in this crop a slow and expensive process.

1.3. DNA markers and genomics

In conventional plant breeding, genetic variation is usually identified by visual selection. However, with the development of molecular biology, it can now be identified at the molecular level based on changes in the DNA. These variations may have direct effects on the phenotype or more often they are simply genetically linked to a trait. Molecular markers are biological features that are determined by allelic forms and can be used as tags to keep track of a chromosome region or a gene. In modern genetics, polymorphism is the relative difference at any genetic locus across a genome. Desirable genetic markers should meet the following criteria: (i) high level of genetic polymorphism; (ii) codominance (so that heterozygous can be distinguished from both homozygous genotypes); (iii) clear distinct allele features (so that different alleles can be identified easily); (iv) even distribution on the entire genome; (v) neutral selection (without pleiotropic effects); (vi) easy detection (so that the whole process can be automated); (vii) low cost of marker development and genotyping; and (viii) high reproducibility (so that the data can be accumulated and shared between laboratories) (Xu, 2010).

Molecular changes can be identified by many techniques to highlight the DNA variation among individuals. Once the DNA has been extracted from plants, variation in samples can be identified using a polymerase chain reaction (PCR) and/or hybridization process followed by, agarose gel electrophoresis, polyacrylamide gel electrophoresis (PAGE) or capillary electrophoresis (CE) to identify distinct molecules based on their sizes. Genetic markers are used to tag and track genetic variation in DNA samples. During the 1980s and 1990s, various types of molecular markers such as restriction fragment length polymorphism (RFLP) (Botstein et al., 1980), randomly amplified polymorphic DNA (RAPD) (Botstein et al., 1980; Welsh and Mc Clelland, 1990), amplified fragment length polymorphsim (AFLP) (Vos et al., 1995), simple sequence repeat (SSR; Tautz, 1989) and single nucleotide polymorphism (SNP; Wang et al., 1998) were developed. Figure 1.2 shows the molecular mechanism of several major DNA markers and the genetic polymorphisms that can be generated by restriction site or PCR priming site mutation, insertion, deletion or by changing the number of repeat units between two restriction or PCR priming sites and nucleotide mutation resulting in a single nucleotide polymorphism (Xu, 2010). Only a selection of markers adopted in this thesis will be discussed in this section.

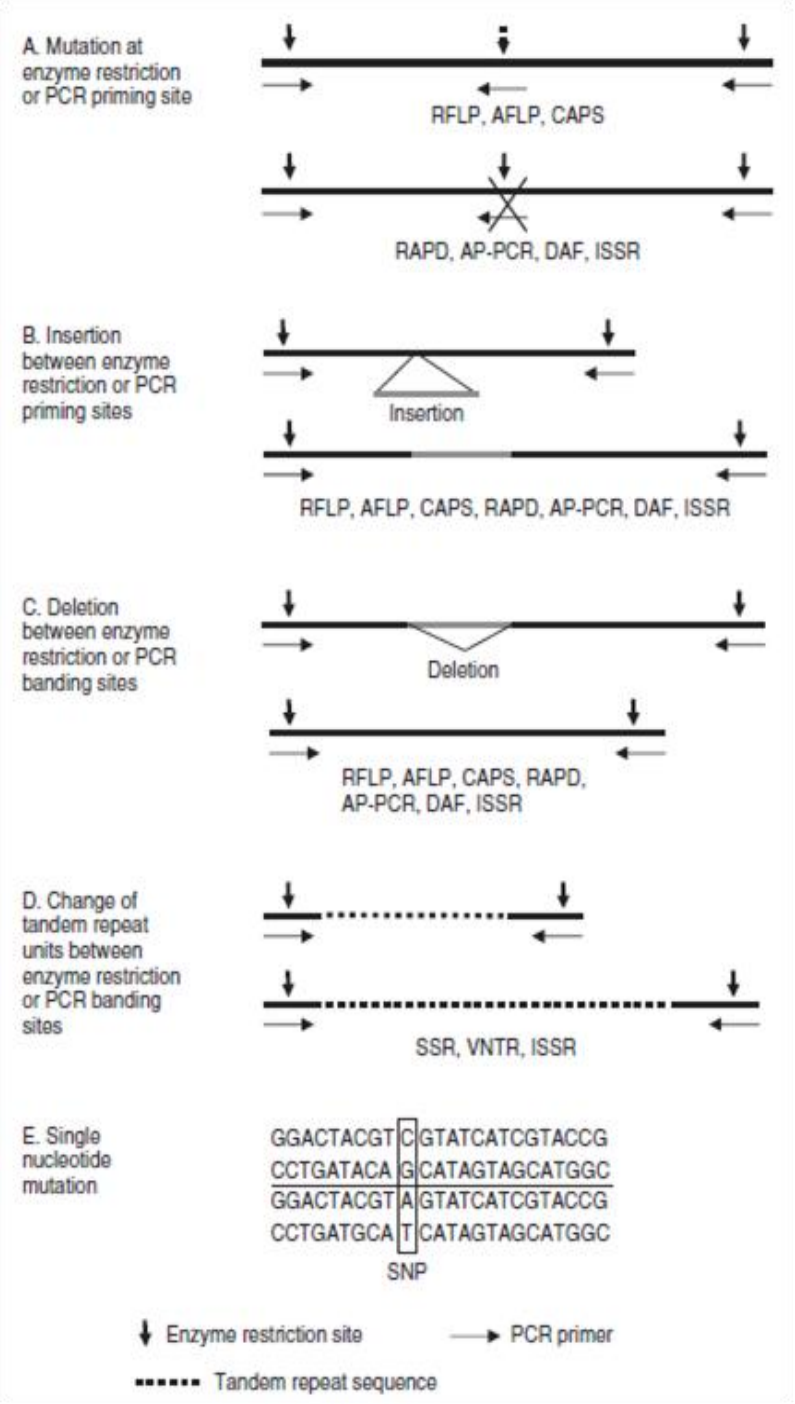


Figure 1.2. Molecular basis of major DNA markers. Parts A–E show different ways in which DNA markers (listed below each diagram) can be generated. The cross in part A indicates that mutation has eliminated the priming site. VNTR, variable number of tandem repeat; CAPS, a DNA marker generated by specific primer PCR combined with RFLP; ISSR, inter simple sequence repeat (Xu, 2010)

1.3.1. RAPD

Williams et al. (1990) and Welsh and McClelland (1990) independently described the utilization of a single, random-sequence oligonucleotide primer in a low stringency PCR (35–45°C) for the simultaneous amplification of several discrete DNA fragments referred to as random amplified polymorphic DNA (RAPD) and arbitrary primed PCR(AP-PCR), respectively. The principle of RAPD consists of a PCR on the DNA of the individual under study using a short primer, usually ten nucleotides, of arbitrary sequence. The short primer which binds to many different loci therefore is used to amplify random sequences from a complex DNA template including a limited number of mismatches. This means that the amplified fragments generated by PCR depend on the length and size of both the primer and the target genome. Ten-base oligomers of varying GC content (ranging from 40 to 100%) are usually used. If two hybridization sites are similar and close to one another (at max 3000 bp) and in opposite directions, the PCR amplification will take place. The amplified products (of up to 3.0 kb) are usually separated on agarose gels and visualized using ethidium bromide staining. The use of a single 10-mer oligonucleotide promotes the generation of several discrete DNA products and these are considered to originate from different genetic loci. Polymorphisms result from mutations or rearrangements either at or between the primer binding sites and are visible in conventional agarose gel electrophoresis as the presence or absence of a particular RAPD band. RAPDs predominantly provide dominant markers but homologous allele combinations can sometimes be identified with the help of detailed pedigree information. RAPDs have several advantages and for this reason they were widely used (Karp and Edwards, 1997): (i) neither DNA probes nor sequence information is required for the design of specific primers; (ii) the procedure does not involve blotting or hybridization steps thus making the technique quick, simple and efficient; (iii) RAPDs require relatively small amounts of DNA (about 10 ng per reaction) and the procedure can be automated; they are also capable of

detecting higher levels of polymorphism than RFLPs; (iv) development of markers is not required and the technology can be applied to virtually any organism with minimal initial development; (v) the primers can be universal and one set of primers can be used for any species. In addition, RAPD products of interest can be cloned, sequenced and converted into other types of PCR-based markers such as sequence tagged sites (STS), etc. Reproducibility affects the way in which RAPD bands can be standardized for comparison across laboratories, samples and trials and whether RAPD marker information can be accumulated or shared. Due to frequently observed problems with reproducibility of overall RAPD profiles and specific bands, this marker class is now treated with reserve in favour of more reliable markers.

1.3.2. SSR

Microsatellites, also known as SSRs, short tandem repeats (STRs) or sequence-tagged microsatellite sites (STMS), are tandemly repeated units of short nucleotide motifs that are 1–6 bp long. Di-, tri- and tetra-nucleotide repeats such as $(CA)_n$, $(AAT)_n$ and $(GATA)_n$ are widely distributed throughout the genomes of plants and animals (Tautz and Renz, 1984). One of the most important attributes of microsatellite loci is their high level of allelic variation, making them valuable as genetic markers. The unique sequences bordering the SSR motifs provide templates for specific primers to amplify the SSR alleles via PCR. Referred to as simple sequence length polymorphisms (SSLPs), they pertain to the number of repeat units that constitute the microsatellite sequence. The predominant mutation mechanism in microsatellite tracts is 'slipped strand mispairing' (Levinson and Gutman, 1987). When slipped-strand mispairing occurs within a microsatellite array during DNA synthesis, it can result in the gain or loss of one or more repeat units depending on whether the newly synthesized DNA chain or the template chain loops out. The relative propensity for either chain to loop out seems to depend in part on the sequences making

up the array and in part on whether the event occurs on the leading (continuous DNA synthesis) or lagging (discontinuous DNA synthesis) strand (Freudenreich et al., 1997). Microsatellites may be obtained by screening sequences in databases or by screening libraries of clones. Once a plant species has been completely sequenced, the entire set of available SSRs in the genome can be easily accessed through online databases. The simple method of SSR genotyping is to separate PCR products by denaturing PAGE and silver staining PCR amplicons. This assay can usually distinguish alleles which differ by 2–4 bp or more. Semi-automated SSR genotyping can be carried out by assaying fluorescently labelled PCR products for length variants on an automated DNA sequencer (e.g. Applied Biosystems 3730/3730xl DNA analyzer). One drawback of fluorescent SSR genotyping is the cost of end-labelling primers with the necessary fluorophores. SSR markers are characterized by their hypervariability, reproducibility, codominance, locus specificity and random dispersion throughout the genome. In addition, SSRs are reported to be more variable than RFLPs or RAPDs. The advantages of SSRs are that they can be readily analyzed by PCR and are easily detected on polyacrylamide gels. SSLPs with large size differences can be also detected on agarose gels. SSR markers can be multiplexed, either functionally by pooling independent PCR products or by true multiplex-PCR. Their genotyping throughput is high and can be automated. In addition, start-up costs are low for manual assay methods (once the markers have been developed) and SSR assays require only very small DNA samples (~100 ng per individual). The disadvantages of SSRs are the labor-intensive development process particularly when this involves screening genomic DNA libraries enriched for one or more repeat motifs (although SSR-enriched libraries can be commercially purchased) and the high start-up costs for automated methods. Next generation sequencing approaches reduce the effects of these disadvantages because it made easily available several most of the SSRs in a specific genome.

1.3.3. SNP

A single nucleotide polymorphism or SNP (pronounced *snip*) is an individual nucleotide base difference between two DNA sequences from the two homologous chromosomes. SNPs can be categorized according to nucleotide substitution as either transitions (C/T or G/A) or transversions (C/G, A/T, C/A or T/G). For example, sequenced DNA fragments from two different individuals, AAGC**C**TA to AAGC**T**TA, contain a single nucleotide difference. In this case there are two alleles: C and T. C/T transitions constitute 67% of the SNPs observed in plants (Edwards et al., 2007). As a nucleotide base is the smallest unit of inheritance, SNPs provide the ultimate form of molecular marker. SNPs may fall within coding sequences of genes, non-coding regions of genes or in the intergenic regions between genes at different frequencies in different chromosome regions. SNPs within a coding sequence will not necessarily change the amino acid sequence of the protein that is produced due to redundancy in the genetic code. A SNP in which both forms lead to the same polypeptide sequence is termed synonymous, while if a different polypeptide sequence is produced they are non-synonymous. SNPs that are not in protein coding regions may still have consequences for gene splicing, transcription factor binding or the sequence of non-coding RNA. Hayden et al. (2009) described a new and simple PCR-based SNP genotyping method. The name is Temperature-switch PCR (TSP). TSP is a biphasic three-primer PCR system with a universal primer design that permits amplification of the target locus in the first phase of thermal cycling before switching to the detection of the alleles. Each TSP marker comprised a set of three primers: a pair of locus-specific primers flanking the SNP and designed to amplify the unigene sequence, and an allele-specific primer designed to assay the SNP harbored within that sequence (Figure 1.3).

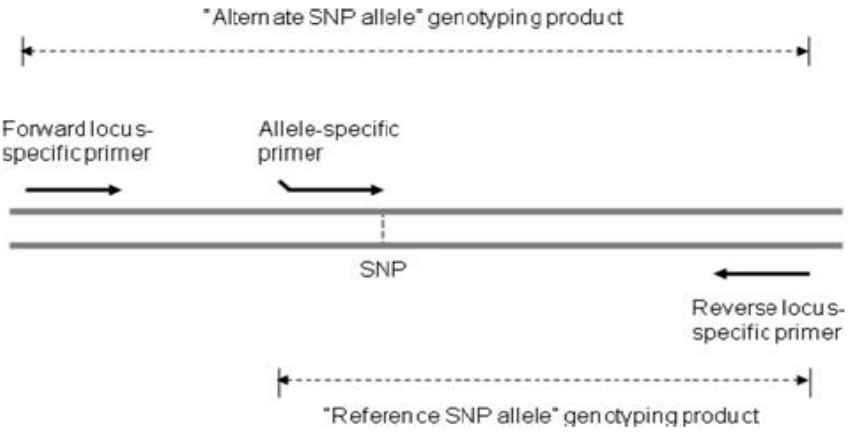


Figure 1.3. Diagram from Hayden et al. (2009) illustrating the position and orientation of PCR primers for three primers, allele-specific TSP markers relative to matching template DNA

The locus-specific primers were designed to work at 60–65℃ and to amplify a PCR fragment greater than 400 bp in length. The locus-specific primers were positioned such that the forward primer designed to the same template DNA strand as the allele-specific primer, was located at least 100 nucleotides from the SNP. The allele specific primer was designed with a melting temperature of 43–48℃ and to have a 3' nucleotide complementary to one of the SNP alleles present at the locus. A short (2 or 3 bp) arbitrary nucleotide sequence, non-complementary to the target DNA, was added to the 5' end of the allele-specific primer. This 5' -tail sequence was designed to increase the melting temperature of the allele-specific primer to about 53℃ once the non-complementary sequence was incorporated into PCR product. The amplification products can be easily separated in agarose gels and visualized after ethidium bromide staining.

1.4. Genetic linkage mapping

In order to utilize the genetic information provided by molecular markers more efficiently,

it is important to know the locations and relative positions of molecular markers on chromosomes. The construction of genetic linkage maps using molecular markers is based on the same principles as those used in the preparation of classical genetic maps: selection of molecular markers and genotyping system; selection of parental lines from the germplasm collection that are highly polymorphic at marker loci; development of a population or its derived lines with an increased number of molecular markers in the segregated population; genotyping each individual/line using molecular markers; and constructing linkage maps from the marker data and conducting linkage analysis using computer program. Commonly used software programs include Mapmaker/EXP (Lander et al., 1987; Lincoln et al., 1993) and MapManager QTX (Manly et al., 2001), which are freely available from the internet. JoinMap is another commonly-used program for constructing linkage maps that was developed also for outbred species like apple (Stam, 1993). The recombination frequency between two linked genetic markers can be defined in units of genetic distance known as centiMorgans (cM) or map units. If two markers are found to be separated in one of 100 progeny, those two markers are 1 cM apart. However, 1 cM does not always correspond to the same length of physical distance in base pairs. Areas in the genome where recombination is frequent are known as recombination hot spots; there is relatively little DNA per cM in these hot spots and it can be as low as 200 kb/cM. In other areas recombination may be suppressed and 1 cM will represent more DNA and in some regions the physical to genetic distance can be up to 1500 kb/cM (Xu, 2010).

1.4.1. Brief History of Genetic Mapping in Apple

The earliest genetic maps of apple were developed in the USA and took advantage of the ready availability of RAPD markers during the nineties. They also included a small number of isoenzyme markers (Hemmat et al., 1994; Conner et al., 1997). These maps were specific to the genetic background of the mapping parents because of the poor

transferability of RAPD markers. For that reason, an international initiative based in Europe developed a genetic map with a number of codominant transportable markers. These were mostly RFLPs plus a few microsatellite markers (Maliepaard et al., 1998). The most complete map to date is constructed with 129 microsatellites, as well as larger numbers of dominant AFLPs and RAPDs to assist in filling in gaps (Liebhard et al., 2003). Within the framework of the HiDRAS European project a new set of 148 SSR markers were identified from genomic libraries, publicly available EST sequences, and SSR markers of other species closely related to *Malus* and mapped on the previously reported apple reference linkage map Fiesta x Discovery (Silfverberg-Dilworth et al., 2006). Such robust PCR-based saturated reference maps are essential for whole genome scanning and for understanding complex traits controlled by several QTLs. A dense genetic map from the cross 'M.9' ('Malling 9') and 'R.5' ('Robusta 5') apple rootstocks was also constructed (Celton et al., 2009). In the last years, new high throughput methods for SNP genotyping and mapping have been established (Khan et al., 2011 and Chagné et al., 2012).

1.5. Mapping quantitative trait locus (QTLs)

Quantitative variation in phenotype (quantitative trait) can be explained by the combined action of many discrete genetic factors or polygenes, each having a rather small effect on the overall phenotype and being influenced by the environment. The contribution of each quantitative locus at a phenotypic level is expressed as an increase or decrease in trait value and it is not possible to distinguish the effect of various loci acting in this manner from one another based on phenotypic variation alone. A chromosomal region linked to or associated with a marker which affects a quantitative trait was defined as a quantitative trait locus (QTL) (Geldermann, 1975).

In simple terms, QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers. Markers are used to partition the mapping population into different genotypic groups based on the presence or absence of a particular marker locus and to determine whether significant differences exist between groups with respect to the trait being measured (Tanksley, 1993; Young, 1996). A significant difference between phenotypic means of the groups (either 2 or 3), depending on the marker system and type of population, indicates that the marker locus being used to partition the mapping population is linked to a QTL controlling the trait. The closer a marker is from a QTL, the lower the chance of recombination occurring between marker and QTL. Therefore, the QTL and marker will be usually inherited together in the progeny, and the mean of the group with the tightly-linked marker will be significantly different ($P < 0.05$) to the mean of the group without the marker (Figure 1.4).

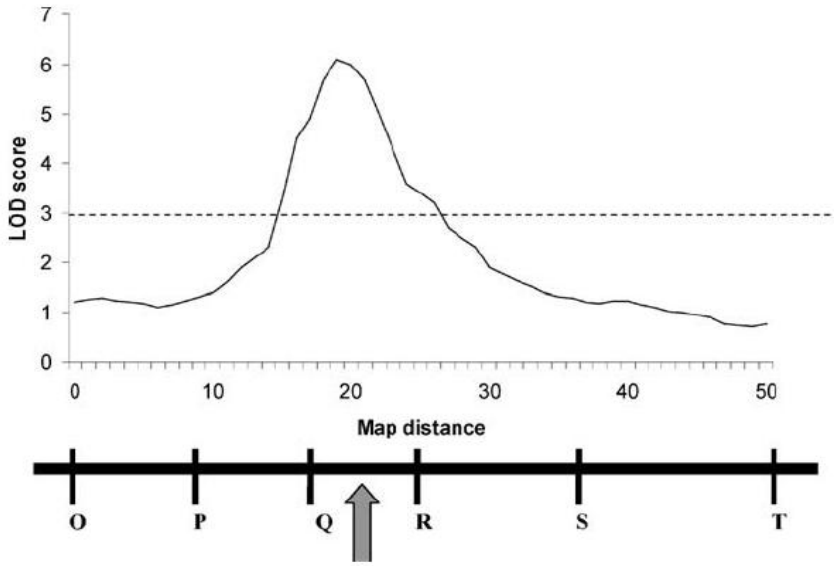


Figure 1.4. Hypothetical output showing a LOD profile for a chromosome. The dotted line represents the significance threshold determined by permutation tests. The output indicates that the most likely position for the QTL is near marker Q (indicated by an arrow). The best flanking markers for this QTL would be Q and R.

When a marker is loosely-linked or unlinked to a QTL, there is independent segregation of the marker and QTL. In this situation, there will be no significant difference between

means of the genotype groups based on the presence or absence of the loosely inked marker (Figure 1.4). Unlinked markers located far apart or on different chromosomes to the QTL are randomly inherited with the QTL; therefore, no significant differences between means of the genotype groups will be detected.

Three widely-used methods for detecting QTLs are single-marker analysis, simple interval mapping (SIM) and composite interval mapping (Liu, 1998; Tanksley, 1993). The SIM makes use of linkage maps and analyses intervals between adjacent pairs of linked markers along chromosomes simultaneously, instead of analyzing single markers (Lander and Botstein, 1989). The use of linked markers for analysis compensates for recombination between the markers and the QTL, and is considered statistically more powerful compared to single-point analysis (Lander and Botstein, 1989; Liu, 1998). Many researchers have used MapMaker/QTL (Lincoln et al., 1993b) and QGene (Nelson, 1997), to conduct SIM.

Interval mapping methods produce a profile of the likely sites for a QTL between adjacent linked markers. In other words, QTLs are located with respect to a linkage map. The results of the test statistic for SIM and CIM are typically presented using a logarithmic of odds (LOD) score or likelihood ratio statistic (LRS). There is a direct one-to-one transformation between LOD scores and LRS scores (the conversion can be calculated by: LRS = 4.6 × LOD) (Liu, 1998). These LOD or LRS profiles are used to identify the most likely position for a QTL in relation to the linkage map, which is the position where the highest LOD value is obtained. A typical output from interval mapping is a graph with markers comprising linkage groups on the *x* axis and the test statistic on the *y* axis. The peak or maximum must also exceed a specified significance level in order for the QTL to be declared as 'real' (i.e. statistically significant). The determination of significance thresholds is most commonly performed using permutation tests (Churchill and Doerge, 1994). Briefly, the phenotypic values of the population are

'shuffled' whilst the marker genotypic values are held constant (i.e. all marker-trait associations are broken) and QTL analysis is performed to assess the level of false positive marker-trait associations (Churchill and Doerge, 1994; Hackett, 2002; Haley and Andersson, 1997). This process is then repeated (e.g. 500 or 1000times) and significance levels can then be determined based on the level of false positive marker-trait associations. Before permutation tests were widely accepted as an appropriate method to determine significance thresholds, a LOD score of between 2.0 to 3.0 (most commonly 3.0) was usually chosen as the significance threshold.

1.6. Marker-assisted selection (MAS)

'Marker-assisted selection' (also 'marker-assisted breeding' or 'marker-aided selection') may greatly increase the efficiency and effectiveness in plant breeding compared to conventional breeding methods. Once markers that are tightly linked to genes or QTLs of interest have been identified, prior to field evaluation of large numbers of plants, breeders may use specific DNA marker alleles as a diagnostic tool to identify plants carrying the genes or QTLs (Michelmore, 1995; Ribaut et al., 1997; Young, 1996). The advantages of MAS include: (i) time saving from the substitution of complex field trials (that need to be conducted at particular times of year or at specific locations, or are technically complicated) with molecular tests; (ii) elimination of unreliable phenotypic evaluation associated with field trials due to environmental effects; (iii) selection of genotypes at seedling stage; (iv) gene 'pyramiding' or combining multiple genes simultaneously; (v) avoid the transfer of undesirable or deleterious genes ('linkage drag'; this is of particular relevance when the introgression of genes from wild species is involved); (vi)selecting for traits with low heritability; (vii) testing for specific traits where phenotypic evaluation is not feasible (e.g.

quarantine restrictions may prevent exotic pathogens to be used for screening) (Xu, 2010). The preliminary aim of QTL mapping is to produce a comprehensive 'framework' that covers all chromosomes evenly in order to identify markers flanking those QTLs that control traits of interest. There are several more steps required, because even the closest markers flanking a QTL may not be tightly linked to a gene of interest (Michelmore, 1995). This means that recombination can occur between a marker and QTL, thus reducing the reliability and usefulness of the marker. By using larger population sizes and a greater number of markers, more tightly-linked markers can be identified; this process is termed 'high-resolution mapping' (also 'fine mapping'). Therefore, high-resolution mapping of QTLs may be used to develop reliable markers for marker assisted selection (at least <5 cM but ideally <1 cM away from the gene) and also to discriminate between a single gene or several linked genes (Michelmore, 1995; Mohan et al., 1997).

1.7. The apple genome and apple bioinformatics resources

Genome resources for apple have been developed over the past 10 years, culminating in the sequencing of the 'Golden Delicious' genome (Velasco et al., 2010). In details, the total contig length (603.9 Mb) covers about 81.3% of the apple genome that was estimated of about 742.3 Mb. Most of the apple genome (500.7 Mb) consists of repetitive elements and a total of 57,386 genes were predicted for the apple genome that is the highest number of genes reported among plants so far (Velasco et al., 2010). The availability of the apple whole-genome sequence is causing a rapid acceleration of apple genetics and genomic research by providing new tools to identify genes and other functional elements, to enable study of the evolution of plant genome structure, as well as the more efficient development of improved apple varieties. There are three public points of access to apple

genome data: IASMA/FEM, GDR Db, NCBI and Phytozome. In addition, apple genome data are integrated into the PLAZA 2.0 comparative genomics platform. The apple genome can be explored using the open source Generic Genome Browser (GBrowse) genome viewer (http://gmod.org/wiki/Gbrowse) at the IASMA/FEM website (http://genomics.research.iasma.it/gb2/gbrowse/apple/), and at the Genome Database for Rosaceae (http://www.rosaceae.org). GBrowse enables the viewing of apple genome data using a wide range of starting points, from chromosome/scaffold coordinates, to the gene or contig name, as well as gene annotation terms (Figure 1.5). Once a user has identified a region of interest, data can be downloaded for 'in house' use.

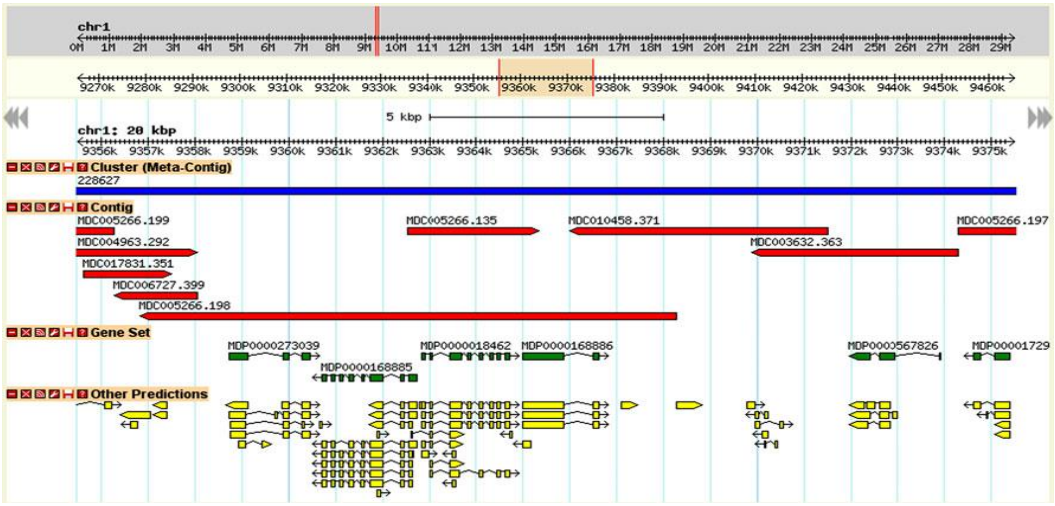


Figure 1.5. Representative view of apple data in GBrowse. The apple scaffold is indicated as Cluster (Meta-Contig) and is depicted in blue; genome contigs belonging to this scaffold are shown as red arrows. Scaffolds and contig coordinates refer to the chromosome (top ruler).Gene predictions are divided into two groups: Gene predictions are indicated as Gene Set (green) and Other Predictions (yellow); the distinction comes from the consensus approach applied in the apple gene prediction phase (Velasco et al., 2010).

The HiDRAS website (http://www.hidras.unimi.it/) was realized in the framework of an European research project 'High-quality Disease Resistance in Apples for Sustainable Agriculture' (HiDRAS). This website contains a SSR markers data repository, including all the available information of primer sequences, the PCR annealing temperature, allele size

and range of reference cultivars, and the map position. A total of 351 already developed and mapped SSR markers info are available for downloading. This database is searchable and easy to use.

1.8. References

Andersson-Eklund, L., L. Marklund, K. Lundström, C. Haley, K. Andersson, I. Hansson, M. Moller and L. Andersson (1998). "Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross." Journal of Animal Science 76(3): 694-700.

Botstein, D., R. L. White, M. Skolnick and R. W. Davis (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." American journal of human genetics 32(3): 314.

Brown AG (1975) Apple: In: Janick J, Moore JN (eds) Advances in Fruit Breeding. Purdue University Press, West Lafayette, IN, USA, PP 3-37

Bus, V., W. Van de Weg, C. Durel, C. Gessler, F. Calenge, L. Parisi, E. Rikkerink, S. Gardiner, A. Patocchi and M. Meulenbroek (2004 a). "Delineation of a scab resistance gene cluster on linkage group 2 of apple." Acta Hortic 663: 57-62.

Bus, V. G., F. N. Laurens, V. De Weg, W. Eric, R. L. Rusholme, E. H. Rikkerink, S. E. Gardiner, H. Bassett, L. P. Kodde and K. M. Plummer (2005 b). "The Vh8 locus of a new gene-for-gene interaction between Venturia inaequalis and the wild apple Malus sieversii is closely linked to the Vh2 locus in Malus pumila R12740-7A." New Phytologist 166(3): 1035-1049.

Calenge, F., A. Faure, M. Goerre, C. Gebhardt, W. Van de Weg, L. Parisi and C.-E. Durel (2004). "Quantitative trait loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of Venturia inaequalis." Phytopathology 94(4): 370-379.

Challice, J. (1974). "Rosaceae chemotaxonomy and the origins of the Pomoideae." Botanical Journal of the Linnean Society 69(4): 239-259.

Churchill, G. A. and R. W. Doerge (1994). "Empirical threshold values for quantitative trait mapping." Genetics 138(3): 963-971.

Conner, P. J., S. K. Brown and N. F. Weeden (1997). "Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars." Journal of the American Society for Horticultural Science 122(3): 350-359.

Crute, I. R. and D. Pink (1996). "Genetics and utilization of pathogen resistance in plants." The Plant Cell 8(10): 1747.

N. E. N. Erdin, S. T. S. Tartarini, G. B. G. Broggini, F. G. F. Gennari, S. S. S. Sansavini, C. G. C. Gessler and A. P. A. Patocchi (2006). "Mapping of the apple scab-resistance gene Vb." Genome 49(10): 1238-1245.

Durel, C., F. Calenge, L. Parisi, W. Van de Weg, L. Kodde, R. Liebhard, C. Gessler, M. Thiermann, F. Dunemann and F. Gennari (2004). "An overview of the position and robustness of scab resistance QTLs and major genes by aligning genetic maps of five apple progenies." Acta Hort (ISHS) 663: 135-140.

Durel, C., L. Parisi and F. Laurens (1999). Does the Vf gene maintain a residual resistance to apple scab

despite its breakdown by Venturia inaequalis race 6 strains? Eucarpia symposium on Fruit Breeding and Genetics 538.

Durel, C., L. Parisi, F. Laurens, W. Van de Weg, R. Liebhard and M. Jourjon (2003). "Genetic dissection of partial resistance to race 6 of Venturia inaequalis in apple." Genome 46(2): 224-234.

Edwards, D., J. W. Forster, D. Chagné and J. Batley (2007). "What Are SNPs?" Association mapping in plants: 41-52.

Evans, R. C. and C. S. Campbell (2002). "The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes." American Journal of Botany 89(9): 1478-1484.

Geldermann, H. (1975). "Investigations on inheritance of quantitative characters in animals by gene markers I. Methods." TAG Theoretical and Applied Genetics 46(7): 319-330.

Gygax, M., L. Gianfranceschi, R. Liebhard, M. Kellerhals, C. Gessler and A. Patocchi (2004). "Molecular markers linked to the apple scab resistance gene Vbj derived from Malus baccata jackii." TAG Theoretical and Applied Genetics 109(8): 1702-1709.

Hackett, C. A. (2002). "Statistical methods for QTL mapping in cereals." Plant molecular biology 48(5): 585-599.

Hayden, M. J., T. Tabone and D. E. Mather (2009). "Development and assessment of simple PCR markers for SNP genotyping in barley." TAG Theoretical and Applied Genetics 119(5): 939-951.

Hemmat, M., N. Weedon, A. Manganaris and D. Lawson (1994). "Molecular marker linkage map for apple." Journal of heredity 85(1): 4-11.

Janick, J. and J. N. Moore (1975). Advances in fruit breeding, Purdue University Press.

Jourjon, M., C. Durel, B. Goffinet and F. Laurens (2000). An example of application for MCQTL software: fine characterisation of a QTL for apple scab resistance. Proceedings of the Plant and Animal Genome Conference VIII, San Diego, Calif. Available at http://www. intl-pag. org/pag/8/abstracts/pag8663. html.

Karp, A., K. J. Edwards, M. Bruford, S. Funk, B. Vosman, M. Morgante, O. Seberg, A. Kremer, P. Boursot and P. Arctander (1997). "Molecular technologies for biodiversity evaluation: opportunities and challenges." Nature biotechnology 15(7): 625-628.

Korban, S. (1986). "Interspecific hybridization in Malus." HortScience 21.

Korban, S., H. Chen, F. Hammerschlag and R. Litz (1992). "Apple." Biotechnology of perennial fruit crops.: 203-227.

Lander, E. S. and D. Botstein (1989). "Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps." Genetics 121(1): 185-199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, L. A. Newberg and L. Newburg (1987). "MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations." Genomics 1(2): 174.

Levinson, G. and G. A. Gutman (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." Molecular biology and evolution 4(3): 203-221.

Liebhard, R., M. Kellerhals, W. Pfammatter, M. Jertmini and C. Gessler (2003 a). "Mapping quantitative

physiological traits in apple (Malus× domestica Borkh.)." Plant molecular biology 52(3): 511-526.

Liebhard, R., B. Koller, L. Gianfranceschi and C. Gessler (2003 b). "Creating a saturated reference map for the apple (Malus× domestica Borkh.) genome." TAG Theoretical and Applied Genetics 106(8): 1497-1508.

Liebhard, R., B. Koller, A. Patocchi, M. Kellerhals, W. Pfammatter, M. Jermini and C. Gessler (2003 c). "Mapping quantitative field resistance against apple scab in a'Fiesta'×'Discovery'progeny." Phytopathology 93(4): 493-501.

Lincoln, S., M. Daly and E. Lander (1993). "Mapping genes controlling quantitative traits using MAPMAKER/QTL version 1.1: a tutorial and reference manual." Whitehead Inst., Cambridge, MA.

Liu, B. H. (1997). Statistical genomics: linkage, mapping, and QTL analysis, CRC.

Maliepaard, C., F. Alston, G. Van Arkel, L. Brown, E. Chevreau, F. Dunemann, K. Evans, S. Gardiner, P. Guilford and A. Van Heusden (1998). "Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers." TAG Theoretical and Applied Genetics 97(1): 60-73.

Manly, K. F., J. Cudmore, Robert H and J. M. Meer (2001). "Map Manager QTX, cross-platform software for genetic mapping." Mammalian Genome 12(12): 930-932.

Michelmore, R. (1995). "Molecular approaches to manipulation of disease resistance genes." Annual review of phytopathology 33(1): 393-427.

Michelmore, R. W. and B. C. Meyers (1998). "Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process." Genome Research 8(11): 1113-1130.

Nelson, J. C. (1997). "QGENE: software for marker-based genomic analysis and breeding." Molecular breeding 3(3): 239-245.

Parisi, L., Y. Lespinasse, J. Guillaumes and J. Krüger (1993). "A new race of Venturia inaequalis virulent to apples with resistance due to the Vf gene." Phytopathology 83(5): 533-537.

Parlevliet, J. and J. Zadoks (1977). "The integrated concept of disease resistance: A new view including horizontal and vertical resistance in plants." Euphytica 26(1): 5-21.

Patocchi, A., B. Bigler, B. Koller, M. Kellerhals and C. Gessler (2004). "Vr 2: a new apple scab resistance gene." TAG Theoretical and Applied Genetics 109(5): 1087-1092.

Patocchi, A., M. Walser, S. Tartarini, G. A. Broggini, F. Gennari, S. Sansavini and C. Gessler (2005). "Identification by genome scanning approach (GSA) of a microsatellite tightly associated with the apple scab resistance gene Vm." Genome 48(4): 630-636.

Primmer, C., A. Møller and H. Ellegren (1996). "A wide-range survey of cross-species microsatellite amplification in birds." Molecular Ecology 5(3): 365-378.

Ribaut, J.-M., C. Jiang, D. Gonzalez-de-Leon, G. Edmeades and D. Hoisington (1997). "Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies." TAG Theoretical and Applied Genetics 94(6): 887-896.

Roberts, A. and I. Crute (1994). "Apple scab resistance from Malus floribunda 821 (Vf) is rendered ineffective by isolates of Venturia inaequalis from Malus floribunda." Norwegian Journal of Agricultural Sciences.

Robinson, J., S. Harris and B. Juniper (2001). "Taxonomy of the genus Malus Mill.(Rosaceae) with emphasis

on the cultivated apple, Malus domestica Borkh." Plant Systematics and Evolution 226(1): 35-58.

Sansavini, S., S. Tartarini and F. Gennari (1999). Scab-tolerance segregation from two Italian apple germplasm cultivars to increase durability of resistance. Eucarpia symposium on Fruit Breeding and Genetics 538.

Sansavini, S., B. Vinatzer, M. Barbieri and S. Tartarini (2000). Advances in Marker-Assisted Selection for Scab Resistance in Apple and Cloning of the Vf Gene. International Symposium on Apple Breeding for Scab Resistance 595.

Silfverberg-Dilworth, E., C. Matasci, W. Van de Weg, M. P. Van Kaauwen, M. Walser, L. Kodde, V. Soglio, L. Gianfranceschi, C. Durel and F. Costa (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree Genetics & Genomes 2(4): 202-224.

Tanksley, S. D. (1993). "Mapping polygenes." Annual review of genetics 27(1): 205-233.

Tartarini, S., F. Gennari, D. Pratesi, C. Palazzetti, S. Sansavini, L. Parisi, A. Fouillet, V. Fouillet and C. Durel 2004). "Characterisation and genetic mapping of a major scab resistance gene from the old Italian apple cultivar 'Durello di Forlì'." Acta Hort.(ISHS) 663: 129-133.

Tautz, D. and M. Renz (1984). "Simple sequences are ubiquitous repetitive components of eukaryotic genomes." Nucleic Acids Research 12(10): 4127-4138.

Van Ooijen, J. (2004). "MapQTL® 5." Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen.

Velasco, R., A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S. K. Bhatnagar, M. Troggio and D. Pruss (2010). "The genome of the domesticated apple (*Malus x domestica* Borkh.)." Nature genetics 42(10): 833-839.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. v. d. Lee, M. Hornes, A. Friters, J. Pot, J. Paleman and M. Kuiper (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic acids research 23(21): 4407-4414.

Way RD, Aldwinckle HS, Lamb RC, Rejman A, Sansavini S, Shen T, Watkins R, Westwood MN, Yoshida Y (1989) Apples (*Malus*). Acta Hort 290:1-62

Welsh, J. and M. McClelland (1990). "Fingerprinting genomes using PCR with arbitrary primers." Nucleic acids research 18(24): 7213-7218.

Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey (1990). "DNA polymorphisms amplified by arbitrary primers are useful as genetic markers." Nucleic acids research 18(22): 6531-6535.

Young, N. (1996). "QTL mapping and quantitative disease resistance in plants." Annual review of phytopathology 34(1): 479-501.

Chapter 2

Characterization of genetic diversity in an Italian apple germplasm collection

2.1. Introduction

The domesticated apple (*Malus × domestica* Borkh.) is one of the most widely cultivated fruit tree and the fourth most economically important following citrus, grape and banana (Hummer and Janick 2009). There are more than 10,000 documented cultivars of apples, resulting in a range of desired characteristics, but only a few major cultivars now dominate the world fruit production (Janick et al. 1996), such as McIntosh (1800s), Jonathan (1820s), Cox's Orange Pippin (1830s), Granny Smith (1860s), Delicious (1870s), Golden Delicious (1890s) and Braeburn (1940s). New cultivars including hybrids increasingly derived from adapted, genetically related and elite modern cultivars with a minimum of disturbance to the genotypic structure have resulted in a narrow genetic base. This was the consequence of the high frequencies involvement of popular commercial cultivars as parentages in the contemporary apple breeding efforts around world. Moreover, the selection and release of mutants of popular cultivars have also accelerated the trend toward a narrow genetic base in commercial apple cultivars (Brooks and Olmo, 1994). As the consequence, the more genetically variable but less productive primitive ancestors have been almost excluded from most breeding programs. The dominance of a few elite cultivars and substitution and consequent loss of a primitive cultivar makes the risk of genetic vulnerability in apple increased. To prevent such losses, and in further step to make the genetic vulnerable condition under control, the primitive ancestors should be adequately conserved for possible future use. Apart from genetic vulnerability mainly resulting from genetic uniformity, greater genetic diversity in apple breeding is highly desirable to develop new cultivars for most traits, including yield, fruit characteristics, plant size and architecture, and resistance to biotic and abiotic stresses (Xu, 2010).

Traditionally, germplasm is defined as a morphologically distinct biological object. Different plant species or cultivars from the same species can be distinguished from each other morphologically based on size, colour and shape. With the development of molecular biology, the concept of germplasm has been generalized and broadened. Germplasm defined as the genetic materials that represent an organism. In other words, germplasm is a comprehensive genetic resource of reserved sum total of genes, gene combinations or genotypes embodied as cultivars available for the genetic improvement of plants (Xu, 2010).

Just having thousands of germplasm accessions available is not helpful if they are not appropriately utilized (Xu, 2010). The evaluation of germplasm resources is a prerequisite for their utilization in apple improvement (Xu, 2010). Vast genetic resources are available for apple plant, but, to date, few of them have been well characterized, either phenotypically or genotypically. Morphological evaluation of germplasm is based on the assessment of phenotypic variation of each accession for a clearly defined characteristic that is recognizable in the whole plant. Unfortunately, exotic accession which is perceived to be a poor bet for the improvement of most traits based on phenotypic examination, may contain some superior genes (alleles) for the improvement of some traits, but they lie buried amid the thousands of accessions maintained in genebanks (de Vicente and Tanksley, 1993; Xiao et al., 1998). Obviously, visual evaluation is not efficient enough to identify all of these features.

Marker-assisted germplasm evaluation (MAGE) aims to complement phenotypic evaluation by helping to define the genetic architecture of germplasm resources and by identifying and managing germplasm that contains alleles associated with traits of economic importance (Xu, Y. et al., 2003). Molecular markers may allow for characterization based on genes, genotypes and genomes which provides more precise information than classical phenotypic or passport

data. Many features can be revealed by molecular markers, such as unique alleles, allele frequency and heterozygosity, mirror the genetic structure of germplasm resources at the molecular level (Lu et al., 2009). On a more fundamental level, molecular marker information can lead to the identification of useful genes contained in collections and aid in the transfer of these genes into well-adapted cultivars.

Microsatellites are a group of repetitive DNA sequences that represent a significant portion of higher eukaryotic genomes (Powell et al., 1996). Such tandem arrays of SSR motifs mutate at a high rate of 10^{-7} to 10^{-3} mutations per locus per generation (Buschiazzo and Gemmell, 2006). Therefore, the number of repeat units may be variable among individual genotypes, making SSRs highly polymorphic and useful for genetic analysis. SSR markers offer several advantages over other molecular markers as the first choice of markers for molecular evaluation of germplasm. Firstly, microsatellite loci are very abundant and relatively evenly distributed across genomes (Buschiazzo and Gemmell, 2006; Kelkar et al., 2008). Secondly, the most important feature of SSR markers is that they can detect multiple alleles per locus (10 alleles in a population). This hypervariability makes SSR markers more informative per locus than single nucleotide polymorphism (SNP) markers, which are generally biallelic (Kelkar et al., 2008). Besides being multiallelic and codominant at a single locus, genetic variability can be easily detected by PCR using primers designed around the SSR repeats and analyzing fragments to detect length-polymorphisms (Csencsics et al., 2010; Dutta et al., 2011). Last but not least, microsatellite PCR amplification protocols are standard, require as little as ~1 ng of DNA per reaction, and are amenable for high-throughput analysis, including fluorescent automated genotyping and multiplexing (Csencsics et al., 2010; Lepais and Bacles, 2011).

Fluorescence-based SSR detection and allele sizing on an automated DNA fragment analyzer

is one of the fastest and most accurate methods for SSR genotyping (Ziegle et al., 1992; Mansfield et al., 1994). This procedure is based on the separation of fluorescently labeled SSR amplicons by capillary, and requires that one of the PCR primers used for SSR amplification is labeled with a fluorescent dye (Ziegle et al., 1992; Oetting et al., 1995). An advantage of fluorescence-based SSR genotyping is that several SSRs can be simultaneously separated in a single capillary providing that the SSR fragments have non-overlapping sizes. In instances where SSR allele sizes are overlapping, coseparation can be achieved by labeling the SSR products with fluorescent dyes that have different emission wavelengths. Two approaches are commonly used to multiplex SSR genotyping. Post-PCR multiplexing (also called multi-pooling) involves the pooling of individual PCR assays for two or more SSRs prior to electrophoresis (Heller, 2001). PCR multiplexing involves simultaneously amplifying two or more SSRs in a single PCR (Henegariu et al., 1997), and is typically used to repetitively genotype a small number of markers. The combined use of these approaches can be used to achieve highly paralleled, multiplexed SSR genotyping (Tang et al., 2003; Ponce et al., 1999). Hayden et al. (2008) presented a new method for multiplex PCR that simplifies assay development and provides several technological advantages that facilitate fluorescence-based SSR genotyping and the multiplexed preparation of DNA templates for SNP detection. This method, termed multiplex-ready PCR, combines the advantages of the M13 tailed primer method and multiplex PCR in a single step, closed tube assay. It does not rely on expensive reagents, is suitable for the amplification of any published sequence and requires only the synthesis of primers in the multiplex-ready format for its deployment.

Recently, approximately 300 SSR markers have been developed and genetically mapped in apple (Guilford et al., 1997; Liebhard et al., 2002; Silfverberg-Dilworth et al., 2006; Han and Korban 2008; Celton et al., 2009; He et al., 2011). These SSR markers have been successfully used to assess genetic diversity and relationships among different apple germplasm, three

literatures reported genetic identity and variations of local apple cultivars in Spain (Pereira-Lorenzo et al., 2007; Pereira-Lorenzo et al., 2008; Urrestarazu et al., 2012), in Czech apple cultivars (Patzak et al., 2012), and indigenous Swedish apple cultivars (Garkava-Gustavsson et al., 2008). Moreover, the genetic diversity of wide apple species like *Malus sieversii* (Richards et al., 2009) and *Malus sylvestris* (L.) Mill. was also analyzed (Reim et al., 2012).

A large number of SSRs make it possible the characterization of specific genomic region in a wide panel of apple accession and align it with the apple genome sequence (Velasco et al., 2010). One of the most important regions for apple breeding is in chromosome 9 and control red skin color (Chagné et al., 2007). Red skin in apple reflect an anthocyanin accumulation and it is correlated to high expression of flavonoid genes during fruit development (Honda et al., 2002). Several recent studies have focused on the underlying mechanisms regulating the expression of these genes in fruit skin. Studies on diverse plant species revealed that anthocyanin expression is controlled, at least in part at the transcriptional level, usually by an R2R3-MYB and/or a basic helix-loop-helix (bHLH) transcription factor (Lin-Wang et al., 2010). In grape the red coloration of the berries depends on a coordinated increase in expression of a number of genes in the anthocyanin biosynthetic pathway during ripening (Boss et al., 1996), that is controlled by a single genetic locus (Kobayashi et al., 2004) containing four MYB genes; at least two of them are mutated in white grapes. In apple (*Malus x domestica*) three different groups have identified the R2R-MYB transcription factors responsible for anthocyanin accumulation (Takos et al., 2006; Ban et al., 2007; Espley et al., 2007). Recently a Myb10 gene has been characterized in several species belonging to Rosaceae family (Lin-Wang et al., 2010; Pierantoni et al., 2010) (Figure 2.1). MdMyb10 is known to be located in apple linkage group 9 (Chagnè et al., 2007).

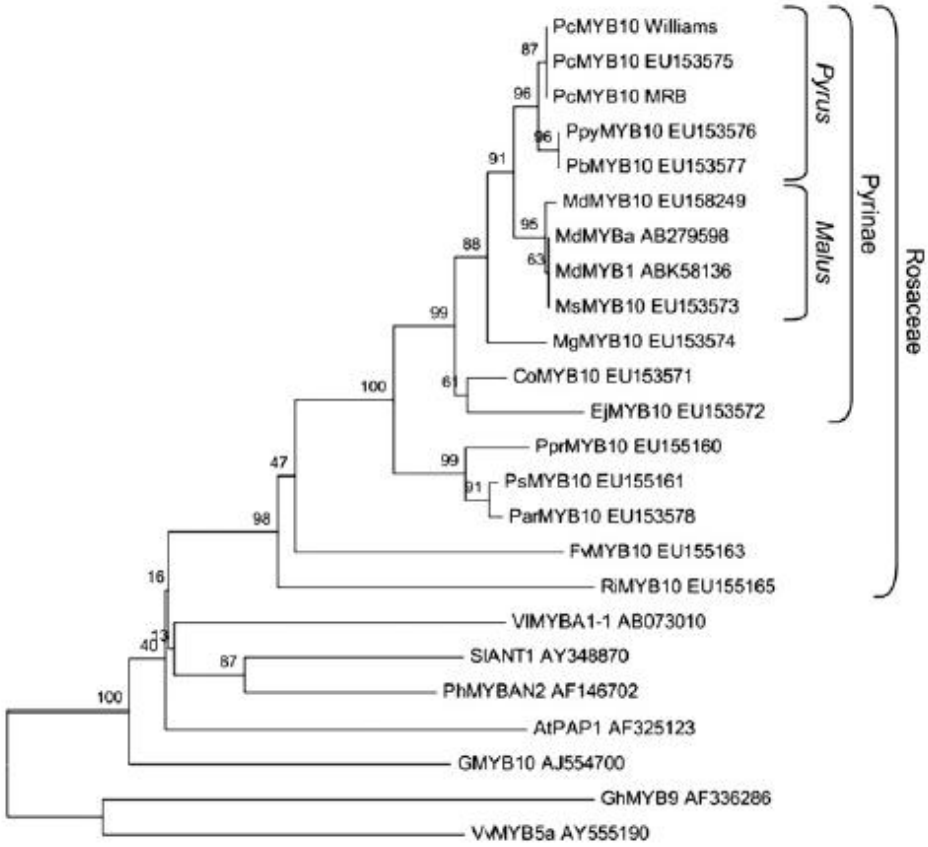


Figure 2.1. Plant MYB factors controlling anthocyanin biosynthesis. Phylogenetic tree constructed by neighbor-joining methods using full-length deduced amino acid sequences of plant R2R3-type MYB transcription factors, including the ones isolated from pear in the present study (PcMYB10 from 'Williams' and 'Max Red Bartlett'). Previously reported MYB transcription factor sequences were retrieved from the EMBL or GenBank databases: *Pyrus communis* (PcMYB10; EU153575), *Pyrus pyrifolia* (PpMYB10; EU153576), *Pyrus bretschneideri* (PbMYB10; EU153577), *Malus x domestica* (MdMYB1; ABK58136), *Malus X domestica* (MdMYB10; EU158249), *Malus x domestica* (MdMYBa; AB279598), *Malus sylvestris* (MsMYB10; EU153573), *Mespilus germanica* (MgMYB10; EU153574), *Cydonia oblonga* (CoMYB10; EU153571), *Eriobotya japonicus* (EjMYB10; EU153572), *Prunus persica* (PprMYB10; EU155160), *Prunus salicina* (PsMYB10; EU155161), *Prunus armeniaca* (ParMYB10; EU153578), *Fragaria vesca* (FvMYB10; EU155163), *Rubus idaeus* (RiMYB10; EU155165), *Gerbera hybrida* (GMYB10; AJ554700), *Vitis labruscana* (VlMYBA1-1; AB073010), *Solanum lycopersicum* (SlANT1: AAQ55181), *Petunia hybrida* (PhMYBAN2; AF146702), *Arabidopsis thaliana* (AtPAP1; AF325123), *Gossypium hirsutum* (GhMYB9; AF336286), and *Vitis vinifera* (VvMYB5a; AY555190)] (Pierantoni et al 2010).

2.2. Aims of study

Apple germplasm resources are known as a repository for the genetic improvement of the cultivated species and its evaluation is an essential prerequisite for their utilization in apple breeding. The characterization of germplasm collections is therefore one of the primary goals for the development of innovative breeding programs because the availability of genetic diversity is highly desirable to develop new cultivars and to widen the genetic base of the species. This is particularly for traits that have a limited variability in the commercial cultivars, including yield, fruit characteristics, plant size and architecture, resistance to biotic and abiotic stresses.

To date, DCA-BO holds a wide apple germplasm collection of old Italian accessions rarely used in breeding programs before. This collection was phenotyped for many traits in the last 25 years and collected data are available for association mapping studies. The traditional phenotypic characterization can be today complemented by the Marker-assisted germplasm evaluation (MAGE) which aims to define the genetic architecture of germplasm resources and to identify and manage germplasm that contains alleles associated with traits of economic importance. Microsatellite markers have been favored over other markers to establish unique genetic identities or fingerprints and to assess genetic diversity due to their high polymorphism level, reproducibility, and relative ease of analysis.

The primary goal of this study is to investigate and evaluate genetic diversity and relationship within this large collection of apple germplasm through the use of SSR markers. The molecular characterization of the collection will allow at first:

- to identify synonyms and homonyms genotypes that are not easy to assess by using

the standard phenotypic descriptors

- to identify triploid accessions thanks to the power of single locus SSR markers and discard these genotypes because they are useless in breeding programs
- to evaluate the relationships among genotypes and identify the unique and more diverse genotypes that have a priority for the establishment of a preservation program
- to select the plants for the constitution of an Italian 'apple core collection' that could be further genotyped for genome wide association studies

Finally, the suitability of the collection for association studies will be tested for a key trait as fruit over-colour. To this extent, four SSRs spanning the LG9 (where the red fruit skin color gene is located) will be used to estimate allele frequencies and to evaluate their association with the phenotypic trait.

2.3. Materials and methods

2.3.1. Plant material and DNA extraction

A collection of 424 apple accessions with few repetitions was investigated, including 386 accessions from Italian germplasm, 28 from other countries and 10 as reference cultivars ('Prima', 'Fiesta', 'Gala', 'Golden Delicious', 'Red Chief', 'Fuji', 'Granny Smith', 'Discovery', 'Florina' and 'Law Red Rome Beauty') (Table 2.1). For each accession, the genomic DNA was extracted from a 30 mg fine powder ground from freeze-dried young leaves following the standard CTAB protocol (Doyle and Doyle, 1990). The gDNA was quantified on the Nanodrop^{TM} ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and diluted to a working solution of 10ng/ul.

Table 2.1. All the 424 apple accessions studied in the germplasm

Code	Accession name	Code	Accession name	Code	Accession name	Code	Accession name
90	ABBONDANZA	A89	CALVILLA BIANCA INVERNALE	D93	DI BONACARDO	D16	GRENOBLE BIS
91	ABBONDANZA ROSSA	B89	CALVILLA ESTIVA	C91	DI LUGLIO	98	GRILL
E91	AGOSTINA	L47	CALVILLA GIALLA	DIS	DISCOVERY	I33	GRIS CANAVIOT
70	AGRE	A92	CALVILLA ROSSA	99	DOLCE	L26	GRIS COMPOSTA
51	ALMA MELA	A84	CALVILLA S. SALVATORE (MI)	L95	DOLCINO	I20	GRIS ED BERTI
C02	AMASYA	B83	CALVILLA S. SALVATORE (TN)	I71	DOMINICI	I56	GRIS ROUS
88	ANNURCA	E25	CAMARON	D35	DRAP DORE	29	GROSSA VINO ROSSO
57	ANNURCA 1	C51	CAMPANINO (MI)	H27	DREI	114	GROSSO VINO ROSSO
59	ANNURCA 2	C53	CAMPANINO (TN)	G10	DUNIZZA GIALLA	1	IMPERATORE MATUZ RUNCO
89	ANNURCA ROSSA DEL SUD	F62	CANNAMELE (PA)	H12	DURELLA	E06	KING DAVID
I59	AOSTA	E61	CANNAMELE (TN)	28	DURELLO	D56	LAVINA
A14	APIONE	G58	CANONICO	87	DURELLO DI FORLI	LRRB	LAW RED ROME BEAUTY
75	APPIA (RT)	111	CAPO D'ASINO 1	E57	DURELLO DI FORLI	E11	LAZZERUOLA
D90	APPICCADORZA	103	CAPO D'ASINO 2	34	EDELROTER	F92	LEDERER
B11	APPIO (TN)	L21	CARBONE	E08	EGREMONT RUSSET	62	LIMONCELLA
B14	APPIO ROSSEGGIANTE	L79	CARDINALE	EL	ELSTAR	F74	LIMONCELLA
A11	APPIO SS	H29	CARLA	73	EPPIA	83	LIMONCELLA URIDDU
61	APPIONA	100	CARLA ROSSA	L89	FACCIA ROSSA	H62	LISCIO CUMIANA
F31	ARKANSAS BLACK	120	CARLE	B71	FERRO CESIO	H20	LOCALE DI TRENTO
E31	ARKANSAS BLACK (RA)	115	CARLO	B65	FERRU	G21	LOCALE GROSSETO
92	ASTRAKAN BIANCO	110	CARPENEDOLA DORATA	B46	FIESTA	G85	LOSA
F35	ASTRAKAN ROSSO	C80	CASCIANA (RODELLA)	FIE	FIESTA	I22	LOSA D' GIAVENO
D86	BACCAIANA	13	CATALINA	E46	FIOR ACACIA PC	G65	LUSIN
C88	BACCALARISCA	G51	CATTIVA	E43	FIOR DI CASSIA PC1	M9	M9
G77	BAUSSAN	C70	CAVECH	FLO	FLORINA	I47	MADAMA
I17	BELA D' BARGE	C75	CAVICCHIO	D01	FOIASSA	H50	MADONNA
A77	BELFIORE GIALLO (MI)	119	CAVICCHIO DI LEVIZZANO	H88	FRAGONI	G92	MAGNANA (TN)
B77	BELFIORE GIALLO (FO)	H24	CERES	A56	FRANCESCA (MI)	I44	MAGNANA D'MANAVELA
A80	BELFIORE TRENTO (TN)	L40	CIODU	B57	FRANCESCA Ai8	E66	MAIOLINO (CT)
B80	BELFIORE TRENTO (PD)	G41	CIPOLLA	B60	FRANCESCA (TN)	F65	MAIOLINO (PA)
E86	BELLA DEL GIARDINO	H47	CIRINO	39	FUJI	MFA	MALUS FLORIBUNDA
50	BELLA DI BOSCO	F96	CLIVIA	102	GAETANA	G95	MARCON (TN)
E29	BELLA DI BOSKOOP	L29	CLOT	18	GALA	H75	MATAN
93	BELLA SCARLATTA	96	COCOINE	21	GALUBRIA	H95	MAZZON (BO)
E49	BEN DAVIS	D44	COMMERCIO (FO)	G68	GALUBRIA (TO)	D71	MELA BASTONE
A95	BENONI	C44	COMMERCIO (MI)	B96	GAMBAFINA	G23	MELA CERINA
I26	BIANC BRUSC	L43	CONTESSA	I68	GAMBE FINE PIATTE	H83	MELA CONO
L07	BIANCHI	14	CONTESSA COLOMBERA	L24	GAMBEFINE LUNGHE	108	MELA D'AUTUNNO
G37	BIANCUCCIA	I07	CONVENTINA GUBBIO	L14	GAROLA	D79	MELA DEL SANGUE
38	BINOTTA	G73	COSSA	85	GELATA	112	MELA DI GENOVA
G29	BISMARK	H78	COUMBA	E69	GELATO (CT)	B93	MELA DONNA
COX	BLEDISLOE COX	95	CV BASSI CUNEO	F68	GELATO (PA)	20	MELA DURELLA
6	BOLOGNOLA	B07	DAMA	E72	GELATO COLA	A52	MELA FORESTIERA
G63	BOMMINO	A69	DE FERRU	H39	GHIACCIOLA	41	MELA FUJONA
56	BOURAS	C86	DE OZZU	B74	GIALLA V.COLLEM.	71	MELA GIALLA I
BR2	Braeburn	E52	DECIO	G71	GIAN D'ANDRE'	72	MELA GIALLA II
BR1	Braeburn (Clone STAR)	12	DECIO(FE)	L17	GIAON	107	MELA GIALLA SENZA RUGGINE
9	BUSIARD	DEL	DELICIOUS	GD	GOLDEN DELICIOUS	A74	MELA GIAPPONE
D83	CADDINA	H57	DELIEGGIO	36	GOLDNOBEL	F47	MELA GOLD SIM VIL COLLEM
C78	CALERA	42	DELLA LIRA	GS	GRANNY SMITH	H14	MELA MAGGIAIOLA
D74	CALIMANO	G32	DELLA SERRA	H80	GRASSI VINO COTTO	A35	MELA MEZZA ROSA
A86	CALVILLA (MI)	DELO	DELORINA	C13	GRENOBLE (PC)	D26	MELA RENETTA GIALLA
B87	CALVILLA (TO)	86	DEMOCRAT	I38	GRENOBLE (TO)	A26	MELA ROSA (PD)

In code, numbers mean the DNA sample number, letters mean the field No. of each accession

(To be continued)

(Continued)

Code	Accession name	Code	Accession name	Code	Accession name	Code	Accession name
B25	MELA ROSA (TN)	E38	POMELLA BRUSCA	D28	RENETTA GRIGIA TORRIANA	30	SAN BARIL
81	MELA ROSATA	E40	POMELLA GENOVESE	D24	RENETTA LOCALE	C63	SAN GIOVANNI (FI)
B54	MELA ROSSA	F38	POMELLA VERDE BRISCA	15	RENETTA ROSETTA	C58	SAN GIOVANNI (MO)
122	MELA ROSSA ALESSANDRO	22	POMI ROSINI	123	RENETTA ROSSA	D59	SAN GIOVANNI (PT)
69	MELA ROZZA	31	POMO ROSSO LILLO	52	RENETTA RUGGINE	H35	SAN PAOLO
G04	MELA SASSA	104	PONTELLA	C32	RENETTA WALDER	64	SANT' AGOSTINO
H06	MELA SASSA D'INVERNO	I62	PORTA	40	RICCARDA	D66	SANT ANNA
G16	MELA TARDIVA 1	PRI	PRIMA	19	RODELLA	I49	SAPIS
F86	MELA TELLINA	I77	PRON	L03	ROS DEL POVER	118	SAPORITO
76	MELA TOSTA	F18	PUM BARGNOL	I86	ROS ED LINOT GIACHE'	C95	SCHLUNER
16	MELA VIOLETTA	L56	PUM BIANC	I96	ROS MAJOR	H03	SCODELLINO
E89	MELA ZAMBONI	E16	PUM BSEUNT	L31	ROS TOMASIN	G80	SCONOSCIUTA BENECH
63	MELE AGRE	L51	PUM D'ACACIA	B18	ROSA (FI)	E59	SEL CAMPIGNA 5
35	MELE D'INVERNO	L59	PUM DARCUCU	84	ROSA D'AOSTA	F59	SEL CAMPIGNA 6
11	MELE GALLIANE	A65	PUM FER	B20	ROSA DI CALDARO	D54	SEL MARZABOTTO 1
I05	MELE UBRIACHE	C65	PUM GIUAN	49	ROSA DI SAN LORENZO	A37	SEL. IDICE No.1
39	MELO DI METZECHI	C69	PUM PERSEGH	105	ROSA DI SAN LORENZO	B38	SEL. IDICE NO.2
45	MELO DODI	B46	PUM ROS	B23	ROSA D'OLIVETO	A41	SEL. IDICE No.3
121	MELO DURONE	B29	PUM ROSA	37	ROSA GENTILE	B41	SEL. IDICE No.4
A71	MELO FERRO (PC)	A29	PUM ROSA BRUSC	A20	ROSA MN (TN)	A44	SEL. IDICE No.5
B68	MELO FERRO (PD)	A50	PUM ROSON	24	ROSA ROMANA	55	SEL. MONTEROMANO
D14	MELO GRENOBLE (BO)	L61	PUM RUGIN	A23	ROSA ROMANA	113	SERPENTE
C16	MELO GRENOBLE (TN)	A48	PUM RUS	43	ROSA ROMANA GENTILE	3	SICILIA GRANDE
L68	MELO RANDAZZO	B62	PUM RUSNEIN	A32	ROSA SS	116	SICILIA PICCOLA
C11	MELO VERDONE	E20	PUM SALAM	B35	ROSATA ROSSULINA	I73	SOLIE
A17	MERAVIGLIA RIGOTTI	F23	PUM SIGULA	C47	ROSMARINA BIANCA (MI)	8	SOLIE D'CUMIANA
F03	MERTON WORCESTER	F13	PUM TOSC	D47	ROSMARINA BIANCA (TN)	B04	SONADORE
C83	MIALI	68	PUMA OLIO	B45	ROSSA	H87	SONALIO
C56	MODENESE	79	PUMA ROSA BIANCA	B32	ROSSA DI LAVENO	117	SPENNACCHINO
80	MORELLA	44	PUMA TENERELLA	7	ROSSA DI VALENZA	54	SPEZZANO ALBANESE
F05	MORELLO	67	PUMA TENERELLA	10	ROSSA MANFENERA	5	SPITZLEDERER
G57	MOSCARELLO	F41	PUMELLA BIS	B49	ROSSA X	27	TINELLA
G44	MUSA	82	PUMELLA DOLCE	L05	ROSSI	94	TONNORELLA
46	MUSO DI BUE	E74	PUMMONCELLO	H68	ROSSI JAHIER	F89	TRAVAGLINO
E80	MUSONA	G01	PUPPINO FERRARESE	48	ROSSO DELLA VALLE	A05	TREMPA ORRUBIA
2	MUSONA MUSABO	D41	RAMBOUR FRANK	47	ROSSO DI AGOSTO	G47	TURCO
G35	NESTA	C41	RAMBOUR FRANK (MI)	77	ROTELLA	C04	VERDONE (FE)
A08	OXIU	C25	RANETTA	L77	ROTOLARI	C09	VERDONE (PC)
H32	PANAIA	REA	REALKA	33	ROUS D'BORSETTA	L71	VERDONE INVERNALE
97	PAOLUCCIA (LT)	RED	RED CHIEF	32	ROUS GIAIET	58	VERGINELLA
60	PAOLUCCIA (VT)	F08	RED GOLD	H91	RUBIS	66	VIGNONE
65	PARADISA	H42	REGINA	C37	RUGGINE RENETTA	E82	VIOLETTA
A60	PARADISA	REI	REINDERS	D38	RUGGINE ROGGIA	G52	VIRCHIATA
109	PARMENA DORATA	D21	REINETTE FRANCHE	E96	RUNSE (TN)	H53	VIRTICCHIARO
106	PARMENA ROSSA	I82	RENE INVERSO	I88	RUS CAVALLOTTA	G26	WAGENER
G83	PASAROT	78	RENETTA ANANAS	I80	RUS D' MUSLOT	I10	WINTER BANANA
I35	PIATLIN	23	RENETTA BIANCA	L39	RUS D' TOURINN	I15	WINTER LEMON
L93	PINA	C24	RENETTA CANADA	A63	RUSIN BLEU	E01	WINTER WINESAP
PIN	PINOVA	4	RENETTA CHAMPAGNE	I28	RUSIN D' BIPU	74	ZANIBONI
25	PLATTLEDERER	D32	RENETTA CHAMPAGNE	L64	RUSTAIEN	B03	ZAZZARI
I41	POM D'LA MADLENA	C21	RENETTA GRENOBLE	17	S. GIUSEPPE	D96	ZEUKA
26	POM SANDRI	53	RENETTA GRIGIA (TN)	F25	SAMBOUN	E78	ZITELLA
18	POMA	C29	RENETTA GRIGIA (TO)	E14	SAMOGIA	101	ZUCCHERINA

In code, numbers mean the DNA sample number, letters mean the field No. of each accession

2.3.2. SSR primer selection and amplification

A set of 16 single-locus SSRs was chosen from HiDRAS (High-quality Disease Resistant Apples for a Sustainable Agriculture) website (http://users.unimi.it/hidras/), mainly on the basis of their distribution across the apple genome (table 2.2), except for 4 SSRs from linkage group 9 for the association analysis on apple fruit skin color (CN444542, CH01H02, Hi05e07 and CH01f03b).

These 16 PCR primers for SSRs were synthesized with generic non complementary nucleotide sequences at their 5'-end (Hayden et al., 2008). Specifically, the forward and reverse primer for each marker was synthesized with the nucleotide tail sequence 5' ACGACGTTGTAAAA 3' and 5' CATTAAGTTCCCATTA 3', respectively. Two generic tag primers, tagF and tagR, with the sequences 5' ACGACGTTGTAAAA 3' and 5' CATTAAGTTCCCATTA 3', respectively, were also synthesized. The tagF primer was labeled at its 5'-end with one of the following fluorescent dyes: VIC, FAM, NED, and PET (Applied Biosystems, Warrington, UK). The amplification of SSR markers was performed using PCR protocol as described by (Hayden et al., 2008). Multiplex-ready PCR achieves consistent amplification yields by amplifying SSRs in two stages in a single-step, closed-tube reaction (Figure 2.2).

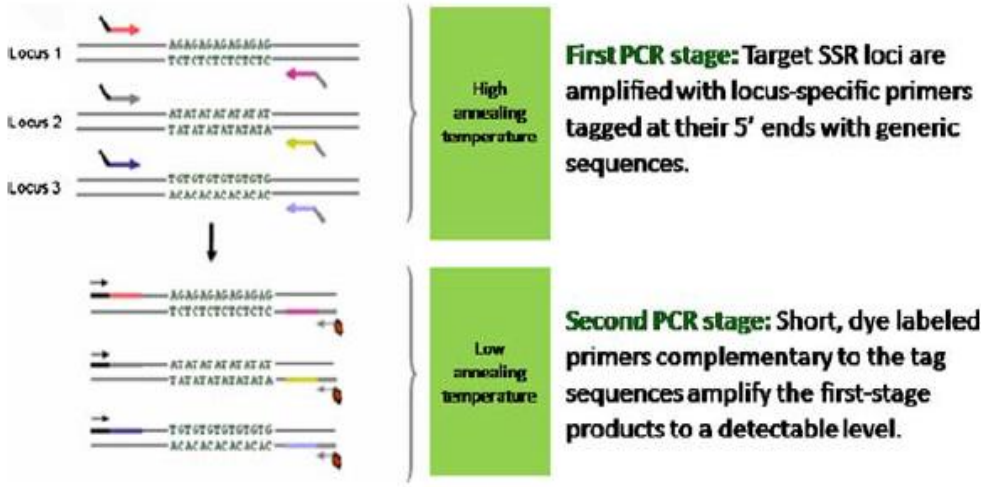


Figure 2.2. Diagrammatic representation of multiplex-ready PCR (Hayden et al., 2008)

In the first stage, low concentrations of locus-specific primers tagged at their 5' ends with generic non-complementary nucleotide sequences amplify the target SSR loci from genomic DNA. The locus-specific primers become fully incorporated into PCR product, which have the

generic nucleotide sequences at their 3' and 5' ends. These nucleotide sequences serve as primer binding sites for the second stage of PCR, and help to reduce amplification bias between SSR loci by normalizing primer hybridization kinetics (Vos et al., 1995). This provides for the uniform amplification of SSRs during the second PCR stage, and results in a relatively uniform yield of SSR product for each marker within a multiplexed reaction. In the second stage of amplification, universal primers corresponding to the nucleotide tag sequences amplify the first stage PCR products to a detectable level.

PCR was performed in a 10 ul reaction mixture containing 1x reaction buffer (Applied Biosystems, Foster City, CA, USA), 3 mM $MgCl_2$ (Applied Biosystems, Foster City, CA, USA), 200 uM dNTPs, 76 nM each of dye-labeled tagF and unlabeled tagR primer, 10 ng genomic DNA, 0.5 U DNA polymerase (Applied Biosystems, Foster City, CA, USA), and 5 nM each locus-specific primer. A modified thermal cycling condition compared to the one reported by (Hayden et al., 2008) included: an initial denaturation step of 10 min at 95 ℃ to heat activate the DNA polymerase, followed by 20 cycles with the profile: 30 s at 92 ℃, 1 min 30 s at 60 ℃, and 1 min at 72 ℃ for 20 cycles, followed by 40 cycles of 15 s at 92 ℃, 30 s at 54 ℃, and 60 s at 72 ℃, with a final extension step of 10 min at 72℃ as the ending.

2.3.3. SSR fragments analysis

PCR products of sixteen SSRs were assigned to four multi-pooling groups (MPG) in order to facilitate a highly paralleled SSR genotyping on ABI 3730 DNA analyzer available at Bologna Sant'Orsola hospital. In practice, four different SSR amplification products labeled with four fluorescent dyes were combined to one single MPG based on the range of their allele length. If the range of 4 SSR alleles size can be well separated without overlapping on the marker panel, these 4 SSRs PCR products will be strongly recommended as one multi-pooling group (Table 2.2).

Table 2.2. The combination of 4 mutli-pooling groups

MPG No.	SSR	dye-labelled	MPG No.	SSR	dye-labelled
MPG 1	CH02c09	VIC	MPG 3	CH01h01	VIC
	CH01a09	FAM		ch01f03b	FAM
	CHVf1	NED		CH02c06	NED
	CH03g07	PET		CH01h10	PET
MPG 2	CH01f02	VIC	MPG 4	CH05C06	VIC
	GD12	FAM		Hi05e07	FAM
	CH04c07	NED		CH01h02	NED
	CH02d08	PET		CN444542	PET

For ABI 3730 analysis, a detailed protocol used to prepare SSR products for electrophoresis was as follows: PCR products labeled with different fluorescent dyes were pooled in a total volume of 25 ul at a ratio of 1 : 1 : 1 : 2 for VIC : FAM : NED : PET, according to differences in the relative fluorescence of each fluorophore; 3 ul of PCR products mixture was added 7 ul of deionized formamide containing 0.2 ul of GeneScan500 LIZ size standard (Applied Biosystems). This multi-pooling mixture containing 4 different SSR amplification products was separated on a capillary electrophoresis ABI 3730 DNA sequencer. Raw fragment size data were analyzed using Peak Scanner Software v1.0 (Applied Biosystems), and all automated results were manually reviewed.

The entire germplasm collection was subjected to the screening and scoring with all the 16 SSRs of choice. The resulting missing data for a few accessions were recovered with a missing data collection procedure in order to fill the germplasm dataset as much as possible.

2.3.4. Statistical Analysis

The number of alleles per locus (A), the observed and expected heterozygosity (Hobs and Hexp), the effective number of alleles (Ae), the number of rare alleles per locus (number of alleles with frequency <0.01) and the number of genotype specific alleles (those present in only one accession) and F-statistics were calculated with the SPAGeDi ver. 1.4 software (Hardy and Vekemans, 2002). This program supports analyses of data sets containing individuals with different ploidy levels. To determine the genetic uniqueness of each accession, and to quantify redundancy, the multilocus DNA profile of all the accessions was compared pairwise using the program Cervus ver. 3.0 (Kalinowski et al., 2007) under the

identity analysis with the setting of minimum number of matching loci is 15 and allow 0 mismatch. Polymorphic information content (PIC) (Botstein et al., 1980; Hearne et al., 1992) was also calculated with same program. The collected data were organized in a square matrix in which the code '0' was used for allele absence and the code '1' for presence of allele (code for missing data was 9). The genetic distance between cultivars was then analyzed with DICE coefficient (2a/2a + b + c; where a = co-presence in two genotypes; b = band present in genotype 1 and c = band present in genotype 2; DICE 1945) implemented by the SimQual procedure of NTSYSpc 2.0 program (Rohlf, 1994). The dendrogram was constructed using the unweighted pair group method average (UPGMA) clustering and drawn with NTSYSpc 2.0 program (Rohlf, 1994). The Cophenetic Correlation Coefficient, which measures the correlation between the distances observed (calculated, in this case, with the Dice coefficient) and those resulting in the constructed tree, has been calculated as a measure of goodness of fit of the tree. The Mantel test (Matix Comparison Plot in NTSYSpc 2.0, using normalized mantel stat to compute the significance of the given correlation) was carried out to investigate matrix correlation.

2.3.5. Linkage analysis for red skin overcolor

A number of 4 SSRs (CN444542, CH01H02, Hi05e07 and CH01f03b) has been used to characterise the genomic region controlling red skin overcolor in linkage group 9. SSRs were genotyped in the whole set of 424 accessions of the apple germplasm but only the 366 ones, for which phenotypic data about skin color were available, have been used for linkage analysis. Accessions were grouped in 4 classes according to their red overcolor coverage (no red skin, low, medium and high overcolor). Allele frequencies within each class of overcolor for all the SSRs have been calculated as well as all their respective chi-square values in order to identify the alleles linked to the trait. A chi-square threshold of 11.3 (0.01 significance with 3 degree of freedom) was used to estimate linkage between allele and trait. BLAST analysis of the SSR sequences (http://www.hidras.unimi.it/), and of the transcription factor MdMYB10 (EU518249) on the apple whole genome sequence (http://www.rosaceae.org/) was used as validation of the linkage analysis.

2.4. Results and discussion

2.4.1. Polymorphism detected by SSRs

Except for the CH02c06, the remaining 15 SSRs obtained clear and easily scored amplified peaks. When scoring the unambiguous peaks, the accuracy of scoring was checked by comparing SSR profile of cultivars with the profiles of their similar cultivars if available. CH01h02 amplified two loci in our gemplasm collections with one locus of a narrow allele range from 230 bp to 233 bp and one locus of a long allele range from 268 bp to 285 bp. Due to the less informativeness of the locus of short allele range with only 3 alleles diversity, CH01h02 was considered as a single locus SSR and only the locus with a wider allele range was scored. For the rest of SSRs, single locus amplification was obtained. Thus, fifteen amplified polymorphic loci were analyzed statistically, yielding a total of 266 polymorphic alleles in the set of 424 accessions.

Number of alleles per polymorphic locus ranged from 9 (CH03g07) to 31 (Hi05e07) with a mean 17.7 (Table 2.3). The average effective number of alleles (Ae) was 5.72, ranging from 3.13 (CHvf1) and 9.74 (CH01a09) (Table 2.5). In general, allele frequencies were distributed unevenly within the investigated locus, each locus had 1 to 4 more common alleles (Table 2.4). A total of 117 rare alleles were identified from all the SSR loci studied varying from 3 (CH02c09, CH01f03b and CH05c06) to 19 (Hi05e07). Out of 117 rare alleles, 38 were genotype-specific alleles considered as present only once in one specific accession, which seems to indicate a substantial level of diversity not previously used for breeding. Except CH03g07, CH02d08, CH01h01, CH01h02 and CH05c06, the remaining SSRs had genotype-specific alleles ranging from 1 (CH02c09 and CH01f03b) to 10 (Hi05e07).

Table 2.3. Chracteristics of the 15 SSRs studied

SSR	LG	Primer[a] (5'-3')		Allele	Total No.	Genotyoe-specific	Rare
		Fwd	Rvs	range (bp)	of alleles	allele No.	allele
CH01a09	14	GATGTGGTTCCAGAAGCTAC	CACATGCATGAAAAGCATAT	212-243	24	4	10
CH02c09	15	TTATGTACCAACTTTGCTAACCTC	AGAAGCAGCAGAGGAGGATG	261-287	12	1	3
CH03g07	3	AATAAGCATTCAAAGCAATCCG	TTTTTCCAAATCGAGTTTCGTT	150-169	9	0	4
CHVf1	1	ATCACCACCAGCAGCAAAG	CATACAAATCAAAGCACAACCC	157-204	18	4	10
GD12	3	TTGAGGTGTTTCTCCCATTGGA	CTAACGAAGCCGCCATTTCTTT	169-220	17	4	6
CH01f02	12	ACCACATTAGAGCAGTTGAGG	CTGGTTTGTTTTCCTCCAGC	190-252	25	4	11
CH02d08	11	TCCAAAATGGCGTACCTCTC	GCAGACACTCACTCACTATCTCTC	237-288	17	0	6
CH04c07	14	GGCCTTCCATGTCTCAGAAG	CCTCATGCCCTCCACTAACA	124-179	23	3	11
ch01f03b	9	GAGAAGCAAATGCAAAACCC	CTCCCCGGCTCCTATTCTAC	169-214	13	1	3
CH01h01	17	GAAAGACTTGCAGTGGGAGC	GGAGTGGGTTTGAGAAGGTT	135-167	14	0	5
CH01h10	8	TGCAAAGATAGGTAGATATATGCCA	AGGAGGGATTGTTTGTGCAC	118-154	17	3	12
CH01h02	9	AGAGCTTCGAGCTTCGTTTG	ATCTTTTGGTGCTCCCACAC	260-280	12	0	5
CH05C06	16	ATTGGAACTCTCCGTATTGTGC	ATCAACAGTAGTGGTAGCCGGT	130-162	15	0	3
Hi05e07	9	CCCAAGTCCCTATCCCTCTC	GTTTATGGTGATGGTGTGAACGTG	213-266	31	10	19
CN444542	9	ATAAGCCAGGCCACCAAATC	GTTTGCAGTGGATTGATGTTCC	152-197	19	4	9
Total					266	38	117

[a] Silfverberg-Dilworth et al. (2006), Liebhard et al. (2002) and Hokanson et al. (1998)

Table 2.4. Allele frequency of 15 SSRs in germplasm with 424 accessions

SSR name	Allele size (bp) / Frequency																		
CH01a09	**212**	214	**216**	***218***	220	221	**222**	223	224	225	227	228	**230**	231	232	**233**	**234**	**235**	**236**
	0.0012	0.0825	0.0035	0.0008	0.0098	0.0747	0.0012	0.0177	0.0656	0.0609	0.0177	0.1144	0.0043	0.1050	0.1946	0.0024	0.0043	0.0024	0.0043
	237	238	***240***	242	243														
	0.0208	0.0963	0.0008	0.0118	0.1030														
CH02c09	261	**267**	269	271	273	**275**	277	279	281	283	285	***287***							
	0.1156	0.0094	0.0248	0.0570	0.3235	0.0024	0.1211	0.0291	0.0350	0.1761	0.1053	0.0008							
CH03g07	150	152	154	**156**	158	160	**162**	**167**	169										
	0.2671	0.0843	0.0788	0.0075	0.1623	0.3798	0.0035	0.0035	0.0130										
CHvf1	157	163	**165**	**167**	***168***	170	172	**175**	**182**	**186**	**188**	190	192	**194**	196	**200**	202	**204**	
	0.0149	0.0102	0.0020	0.0043	0.0012	0.4855	0.0601	0.0035	0.0012	0.0047	0.0075	0.0134	0.0983	0.0012	0.0228	0.0031	0.2649	0.0012	
GD12	***169***	171	179	180	182	183	184	185	**186**	189	191	**193**	**197**	204	212	**214**	**220**		
	0.0008	0.0189	0.1376	0.4501	0.1262	0.0126	0.0189	0.1050	0.0012	0.0373	0.0224	0.0047	0.0012	0.0177	0.0350	0.0012	0.0094		
CH01f02	190	**192**	200	202	203	205	209	**211**	212	213	**214**	**215**	217	**219**	221	223	225	**227**	**230**
	0.0236	0.0012	0.2488	0.0106	0.0684	0.0259	0.1380	0.0047	0.0951	0.1053	0.0071	0.0024	0.0480	0.0055	0.0110	0.0161	0.0236	0.0012	0.0024
	234	236	**238**	**246**	**248**	**252**													
	0.0503	0.1026	0.0047	0.0012	0.0012	0.0012													
CH02d08	**237**	239	241	243	245	**247**	249	254	256	**258**	260	278	280	282	**284**	286	**288**		
	0.0035	0.0138	0.1309	0.2319	0.0248	0.0035	0.0649	0.0432	0.0873	0.0043	0.1281	0.0291	0.0173	0.0118	0.0063	0.1946	0.0047		
CH04c07	124	126	**130**	**132**	134	136	138	140	142	**144**	**146**	149	151	***155***	**160**	**161**	163	165	**167**
	0.0837	0.1533	0.0012	0.0071	0.0417	0.2209	0.0444	0.0786	0.1435	0.0059	0.0059	0.0232	0.0440	0.0012	0.0024	0.0012	0.0122	0.0444	0.0047
	171	**173**	**175**	***179***															
	0.0731	0.0047	0.0012	0.0016															
CH01f03b	169	175	**185**	187	189	191	193	200	**204**	206	208	212	**214**						
	0.1600	0.0130	0.0024	0.0272	0.2100	0.0493	0.0122	0.3058	0.0012	0.0355	0.1592	0.0162	0.0083						
CH01h01	**135**	**140**	142	144	146	149	151	153	155	**157**	***159***	161	163	**167**					
	0.0094	0.0024	0.0200	0.0849	0.1089	0.1757	0.2091	0.0833	0.0145	0.0059	0.0047	0.0318	0.2469	0.0024					
CH01h10	**118**	**122**	124	**128**	131	**133**	***134***	135	136	138	**141**	143	**145**	**148**	**150**	**152**	**154**		
	0.0024	0.0012	0.1950	0.0012	0.4488	0.0024	0.0008	0.1221	0.0634	0.0102	0.0055	0.1269	0.0032	0.0032	0.0071	0.0035	0.0032		
CH01h02	**260**	262	264	**266**	268	**270**	272	274	276	**277**	278	280							
	0.0047	0.0118	0.4431	0.0095	0.0703	0.0024	0.1232	0.0983	0.1722	0.0087	0.0454	0.0103							
CH05c06	**130**	132	134	136	138	140	142	**144**	147	149	151	153	155	157	**162**				
	0.0036	0.2828	0.0182	0.1106	0.0470	0.0158	0.0269	0.0059	0.1264	0.0288	0.2563	0.0130	0.0355	0.0209	0.0083				
Hi05e07	213	***215***	**217**	***218***	**219**	220	221	222	**223**	**224**	226	228	**230**	232	234	**236**	***237***	238	240
	0.0265	0.0008	0.0012	0.0008	0.0083	0.1884	0.0103	0.0503	0.0095	0.0048	0.0523	0.0483	0.0012	0.0297	0.0127	0.0012	0.0008	0.0103	0.0447
	241	**243**	244	***247***	**251**	**253**	***255***	257	**259**	**261**	***265***	**266**							
	0.1979	0.0020	0.0127	0.0008	0.0020	0.0012	0.0008	0.2664	0.0063	0.0048	0.0008	0.0024							
CN444542	152	154	156	158	***160***	162	164	**166**	168	**170**	**171**	173	**175**	179	**181**	183	**187**	***189***	**197**
	0.3003	0.0342	0.0354	0.0118	0.0012	0.0114	0.2614	0.0035	0.0637	0.0035	0.0012	0.1863	0.0012	0.0106	0.0094	0.0590	0.0012	0.0024	0.0024

The rare alleles are indicated in bold, the genotype-specific alleles are indicated with underline and genotype-specific alleles only present in the triploid genotypes are indicated in italic

A total of 119 accessions were identified with three alleles at least in one locus, from which 87 accessions with three alleles in more than three loci (83 accessions in at least four loci and four accessions in three loci). In the remaining 32 accessions, 29 presented only one locus with three alleles and three in two loci. After careful examination in 32 accessions, a third allele normally with biggest size compared to the first and second allele was present and this could be due to possible contamination or stutter peaks, as confirmed by the already known diploid cultivars (e.g. Delicious, M9 and Pum Bianc). In this case, 32 accessions were categorized as diploid by removing the biggest allele, while the rest 87 accessions were identified as true triploids in our study. Interestingly, except CH03g07, CH01f02, CH02d08, CH01f03b, CH01h02 and CH05c06, the triploid specific alleles were detected in the other nine SSRs (Table 2.4). The fact that most alleles of this kind were in the middle of normal diploid alleles with 1~3 bp more or less ruled out the possibility of stutter peaks. A better explanation for these triploid specific alleles was due to the mutation of simple repeated sequence numbers.

Reliable genetic markers are essential for efficient differentiation of cultivars and to establish the genetic relationships among them. All SSR loci analyzed in this study displayed a high degree of polymorphism with 9 – 31 alleles per locus. The overall allelic diversity showed by the set of 15 SSR used revealed a high genetic variation in the apple germplasm evaluated. When compared to other wide scale studies of apple genetic diversity, the average number of alleles per locus (17.73) was similar (18.50) to that reported by van Treuren et al. (2010) in their analysis of 695 local genotypes from several Dutch collections, and higher than the one obtained (12.30) by Pereira-Lorenzo et al. (2007) and Gharghani et al. (2009) who studied both pools of ≈100 local and commercial genotypes. Those levels or polymorphism reveal a high genetic diversity in the studied apple germplasm, in consistence to the level of polymorphism reported by van Treuren et al. (2010) and by Pereira-Lorenzo et al. (2007).

2.4.2. Genetic diversity

HExp summarizes the fundamental genetic variation of a population or species for a single parameter (Berg and Hamrick, 1997) and, for this reason, it is a commonly used parameter

that allows comparison with the literature. The estimate of HExp in Italian germplasm of this study (0.806) (Table 2.5) was in agreement with the level of polymorphism reported for local apple cultivars in Northwestern Spain (0.80, Pereira-Lorenzo et al., 2007) and it was slightly higher than that obtained for a reduced set of Spanish apple cultivars from different geographical origins (0.73, Pereira-Lorenzo et al. 2008) or than those reported for Swedish, and Bosnian material (Garkava-Gustavsson et al., 2008; Gasic et al., 2010; Guarino et al., 2006).

The inbreeding coefficient F_{IS} ranged from 0.155 (CH01h01) to −0.104 (CHvf1), in the overall set, with a quite low average value of -0.022 ($P<0.001$) for all loci, suggesting no loss of heterozygosity among the accessions analyzed (Table 2.5). The F_{IS} values were closer to those reported by Pereira-Lorenzo et al. (2007; −0.088) and Gasi et al. (2010; 0.023) in *Malus x domestica* germplasm than to those reported for wild *Malus* (Coart et al., 2003; Gharghani et al., 2009; Richards et al., 2009).

Table 2.5 Measures of genetic diversity of 424 accessions based on 15 SSR loci

Locus	A	Ae	AR	HObs	HExp	PIC	F_{IS}	P value
CH01a09	24	9.740	23.880	0.861	0.897	0.890	0.006	0.684
CH02c09	12	5.560	11.990	0.843	0.820	0.791	-0.053	0.006
CH03g07	9	3.930	9.000	0.652	0.745	0.709	0.036	0.148
CHvf1	18	3.130	17.980	0.757	0.681	0.633	-0.104	0.000
GD12	17	3.970	16.890	0.742	0.748	0.717	-0.036	0.082
CH01f02	25	8.160	24.890	0.855	0.877	0.865	-0.007	0.653
CH02d08	17	7.140	17.000	0.840	0.860	0.842	-0.026	0.127
CH04c07	23	8.410	22.940	0.917	0.881	0.870	-0.062	0.000
CH01f03b	13	5.190	12.970	0.810	0.807	0.543	-0.066	0.001
CH01h01	14	6.160	14.000	0.656	0.838	0.806	0.155	0.000
CH01h10	17	3.650	16.990	0.792	0.726	0.680	-0.126	0.000
CH01h02	12	3.880	12.000	0.693	0.743	0.715	0.031	0.187
CH05c06	15	5.580	15.000	0.857	0.821	0.801	-0.070	0.000
Hi05e07	31	6.380	30.880	0.814	0.843	0.827	-0.021	0.237
CN444542	19	4.930	18.960	0.787	0.797	0.585	-0.008	0.715
Mean	17.73	5.720	17.690	0.790	0.806	0.779	-0.022	0.000

Number of alleles per locus (A), number of effective number of alleles (Ae), allelic richness (expected number of alleles among 878 gene copies)(AR), observed (HObs) and expected (HExp) heterozygosity, polymorphic information content (PIC), and F_{IS} value were included

PIC value ranged from 0.543 (CH01f03b) to 0.890 (CH01a09) with a mean value of 0.779, suggesting the 15 SSRs selected for this Italian germplasm study contained a high polymorphic information content. Moreover, the AR also provided evidence proving the genetic diversity of our germplasm.

2.4.3. Cultivar identification

Comparison of SSR profiles revealed 44 groups of diploid accessions that had the same SSR profile. The size of those groups varied from two to seven accessions, and involved 114 as a whole (Table 2.6). Four groups (group 2, 9, 16 and 44) containing two accessions with identical name revealed that the repetition of accession selection occurred. Besides, two identical accessions with the same name were also found in the group 5 and 9. So a total of six accessions, namely Renetta Champagne, Puma Tenerella, Limoncella, Abbondanza Rossa, Rosa di San Lorenzo and Fiesta were resampled in this germplasm. Some duplicates were expected, as the accessions received very similar or identical denominations only with little change of collection place, for example, the group 13 of two Paoluccia one from Viterbo and one from Latina were actually the same genotype. More examples of this kind replicates included group 29, group 30 (with two Belfiore Giallo), group 31 and group 37. However, most groups of duplicates belonged to synonyms category, meaning accessions contained in these groups with different names but the same genotype profiles (Table 2.6). For example, the Rosa Mantovana and Rosata Russolina from group 26 both produced bicolour (green and red) and flat-globose apples (Figure 2.3).

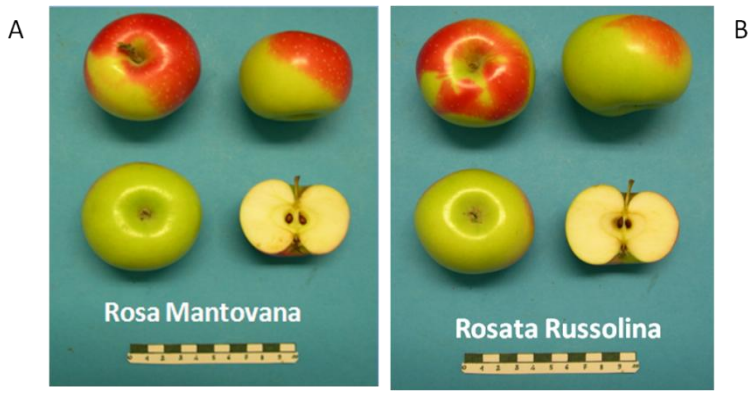


Figure 2.3. Picture of two synonym accessions from germplasm. A.Rosa Mantovana and B. Rosata Russolina

Table 2.6 Marker duplicate groups of diploid accessions with the same allelic profile at 15 SSR loci

Marker duplicate group number	Diploid accessions with the same allelic profile at 15 SSR loci				
1	IMPERATORE MATUZ RUNCO (1)	SAPORITO (118)	MORELLA (80)	CV BASSI CUNEO (95)	MELA FORESTIERA (A52)
	MELA SASSA D'INVERNO (H06)	LAW RED ROME BEAUTY (LRRB)			
2	RENETTA CHAMPAGNE (4)	RENETTA CHAMPAGNE (D32)			
3	DECIO(FE) (12)	MODENESE (C56)	DECIO (E52)		
4	CATALINA (13)	ROSSA DI VALENZA (7)			
5	RENETTA ROSETTA (15)	SICILIA GRANDE (3)	DELLA LIRA (42)	LIMONCELLA (62)	LIMONCELLA URIDDU (83)
	PUMMONCELLO (E74)	LIMONCELLA (F74)	MUSA (G44)		
6	MELA VIOLETTA (16)	VIOLETTA (E82)			
7	DURELLO (28)	MORELLO (F05)	PINA (L93)		
8	ROUS GIAIET (32)	ROSSI JAHIER (H68)			
9	PUMA TENERELLA (44)	PUMA TENERELLA (67)			
10	RENETTA GRIGIA (TN) (53)	RUGGINE RENETTA (C73)	RENETTA GRIGIA TORRIANA (D28)		
11	SPEZZANO ALBANESE (54)	VERGINELLA (58)	GELATA (85)		
12	ANNURCA 1 (57)	ANNURCA 2 (59)	ANNURCA ROSSA DEL SUD (89)	SEL. IDICE No.5 (A44)	VIRCHIATA (G52)
13	PAOLUCCIA (VT) (60)	PAOLUCCIA (LT) (97)			
14	PARADISA (65)	MELA ZAMBONI (E89)	SAN PAOLO (H35)		
15	MELA GIALLA I (71)	MELA GIALLA II (72)			
16	ABBONDANZA ROSSA (90)	ABBONDANZA ROSSA (91)			
17	CAPO D'ASINO 2 (103)	CAPO D'ASINO 1 (111)			
18	PONTELLA (104)	APPIA (RT) (75)			
19	ROSA DI SAN LORENZO (105)	PLATTLEDERER (25)	ROSA DI SAN LORENZO (49)		
20	PARMENA ROSSA (106)	PARMENA DORATA (109)	GRILL (98)		
21	MELA GIALLA SENZA RUGGINE (107)	ASTRAKAN BIANCO (92)			
22	MELA DI GENOVA (112)	MELA TOSTA (76)			

In the bracket, numbers mean the DNA sample number, letters mean the field No. of each accession

(Continued)

Marker duplicate group number	Diploid accessions with the same allelic profile at 15 SSR loci				
23	TREMPA ORRUBIA (A05)	CALIMANO (D74)			
24	OXIU (A08)	DE OZZU (C86)			
25	MERAVIGLIA RIGOTTI (A17)	RUBIS (H91)			
26	ROSA MANTOVANA (A20)	ROSATA ROSSULINA (B35)			
27	ROSA SS (A32)	DAMA (B07)			
28	PUM RUS (A48)	PUM ROSA (B29)			
29	FRANCESCA (MI)(A56)	FRANCESCA (TN) (B60)			
30	BELFIORE GIALLO (MI) (A77)	BELFIORE TRENTO (TN) (A80)	BELFIORE GIALLO (FO) (B77)	PUM BSEUNT (E16)	
31	CALVILLA S. SALVATORE (MI) (A84)	CALVILLA S. SALVATORE (TN) (B83)			
32	ZAZZARI (B03)	WINTER BANANA (I10)			
33	GAMBAFINA (B96)	BEN DAVIS (E49)			
34	Braeburn (Clone STAR) (BR1)	Braeburn (BR2)			
35	GRENOBLE (PC) (C13)	MELO GRENOBLE (TN) (C16)			
36	RAMBOUR FRANK (MI) (C41)	COMMERCIO (FO) (D44)			
37	SAN GIOVANNI (FI) (C63)	SAN GIOVANNI (PT) (D59)	ASTRAKAN ROSSO (F35)	DELLA SERRA (G32)	POM D'LA MADLENA (I41)
38	FIOR DI CASSIA PC1 (E43)	VERDONE INVERNALE (L71)			
39	CANNAMELE (TN) (E61)	MAIOLINO (CT) (E66)			
40	MUSONA (E80)	RED CHIEF (RED)			
41	GRIS CANAVIOT (I33)	GRIS ROUS (I56)			
42	MADAMA (I47)	CLOT (L29)			
43	PUM BIANC (L56)	M9 (M9)			
44	FIESTA (B46)	FIESTA (FIE)			

In the bracket, numbers mean the DNA sample number, letters mean the field No. of each accession

Table 2.7 Marker duplicate groups of triploid accessions with the same allelic profile at 15 SSR loci

Marker duplicate group number	Triploid accessions with the same allelic profile at 15 SSR loci				
1	MELO VERDONE (C11)	VERDONE (PC) (C09)			
2	MELO FERRO (PC) (A71)	PUM ROSON (A50)			
3	MELA DURELLA (20)	PUM ROSA BRUSC (A29)	ROSA GENTILE (37)	ROSA ROMANA (24)	ROSA ROMANA GENTILE (43)
4	RENETTA CANADA (C24)	DE FERRU (A69)			
5	AGOSTINA (E91)	RANETTA (C25)	SEL. IDICE No.1 (A37)		
6	CARLO (115)	CONTESSA COLOMBERA (14)	SANT ANNA (D66)		
7	CADDINA (D83)	RENETTA LOCALE (D24)			
8	CASCIANA (RODELLA) (C80)	PIATLIN (I35)			

In the bracket, numbers mean the DNA sample number, letters mean the field No. of each accession

Six accessions from the group 1 namely Imperatore Matuzalem Runco, Saporito, Morella, Cv Bassi Cuneo, Mela Forestiera and Mela Sassa d'Inverno were identical to the reference cultivar Law red Rome Beauty and in group 40 Musona was identical to Red Chief, which may indicate the introgression of foreign cultivars into native populations. Among the analyzed genotypes there were also some clones of known cultivars: Abbondanza and Abbondanza Rossa and, Annurca and Annurca Rossa del Sud. As expected, no differences were observed. For the triploids, eight groups of marker duplicates containing eighteen accessions (Table 2.7) were identified. No identical accession repeated in the triploid marker duplicates groups. Eight groups were all groups of synonyms. Therefore, 114 accessions identified as duplicates of 44 groups in the 337 diploid collection reduced the diploid genotype number to 276. The number of unique triploid genotypes reduced from 87 to 77 excluding the replicated genotypes. Sums of 353 distinct genotypes were identified through the identity analysis using pairwise comparison of SSR profile inside the whole germplasm.

2.4.4. Cluster analysis

The result of the 15 SSRs data analyzed using NTSYS software showed a value of R= 0.65 indicated that the dendrogram faithfully represented the molecular data for low genetic distances (within the same cluster), while there is a risk of failure for long distances (between distinct clusters) (Figure 2.4 A-H). This value is due to the use of a relatively limited number of markers, which does not allow a precise reconstruction of the genetic relationships in a population so wide. Almost all the marker duplicate groups identified before can be detected on the dendrogram in the cluster groups with the R value of 1, suggesting the correctness of genotypes replicates identified by the Cervus software. A further analysis is needed to confirm the differences between two very similar genotypes because in some cases the presence of a different and specific allele was found (i.e. Appiccadorza and Baccalarisca, Zitella and Cattiva) but in other cases the polymorphism seems to be due to the presence/absence of one or few alleles that may be due to a less efficient PCR amplification in one genotype.

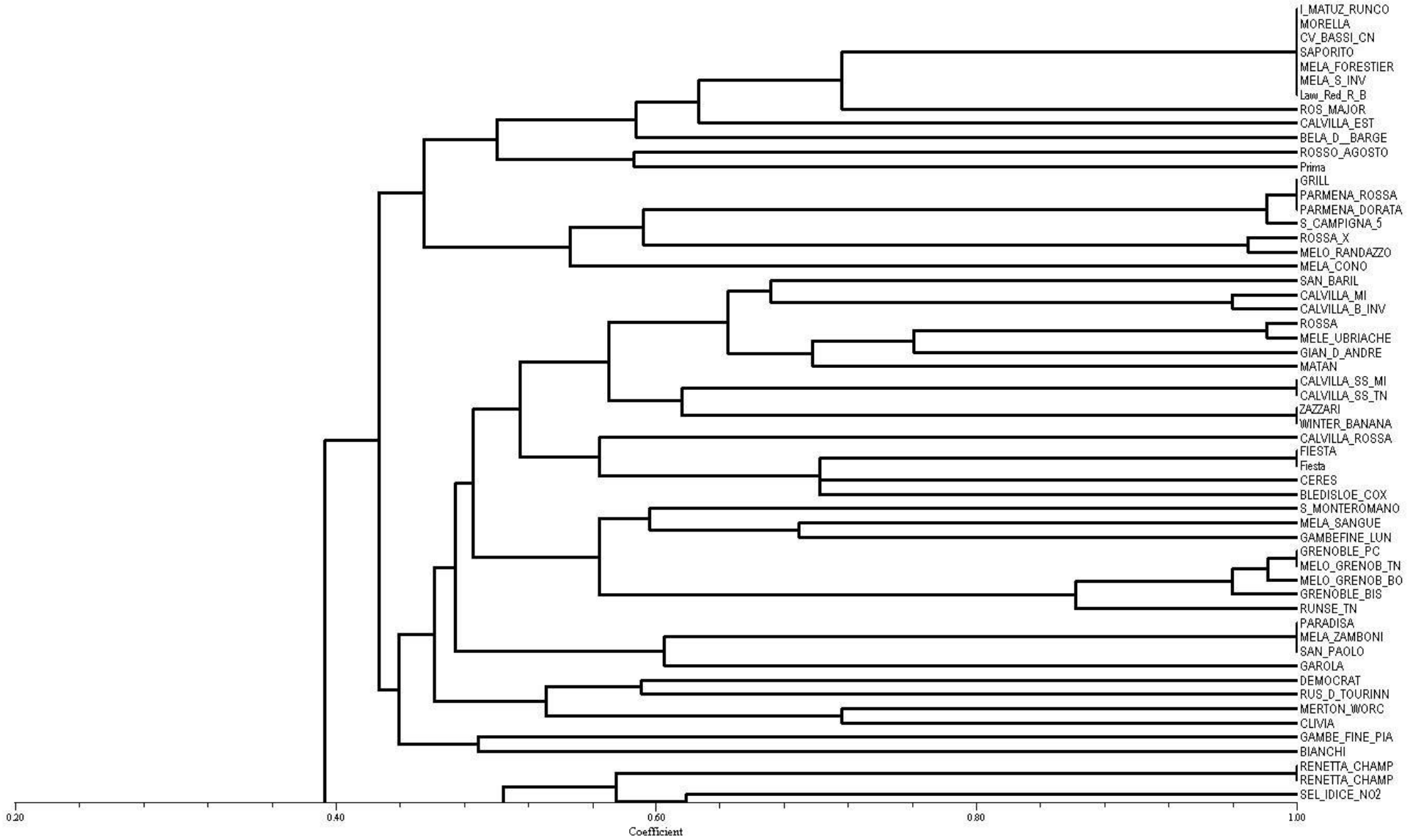


Figure 2.4 A. Dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

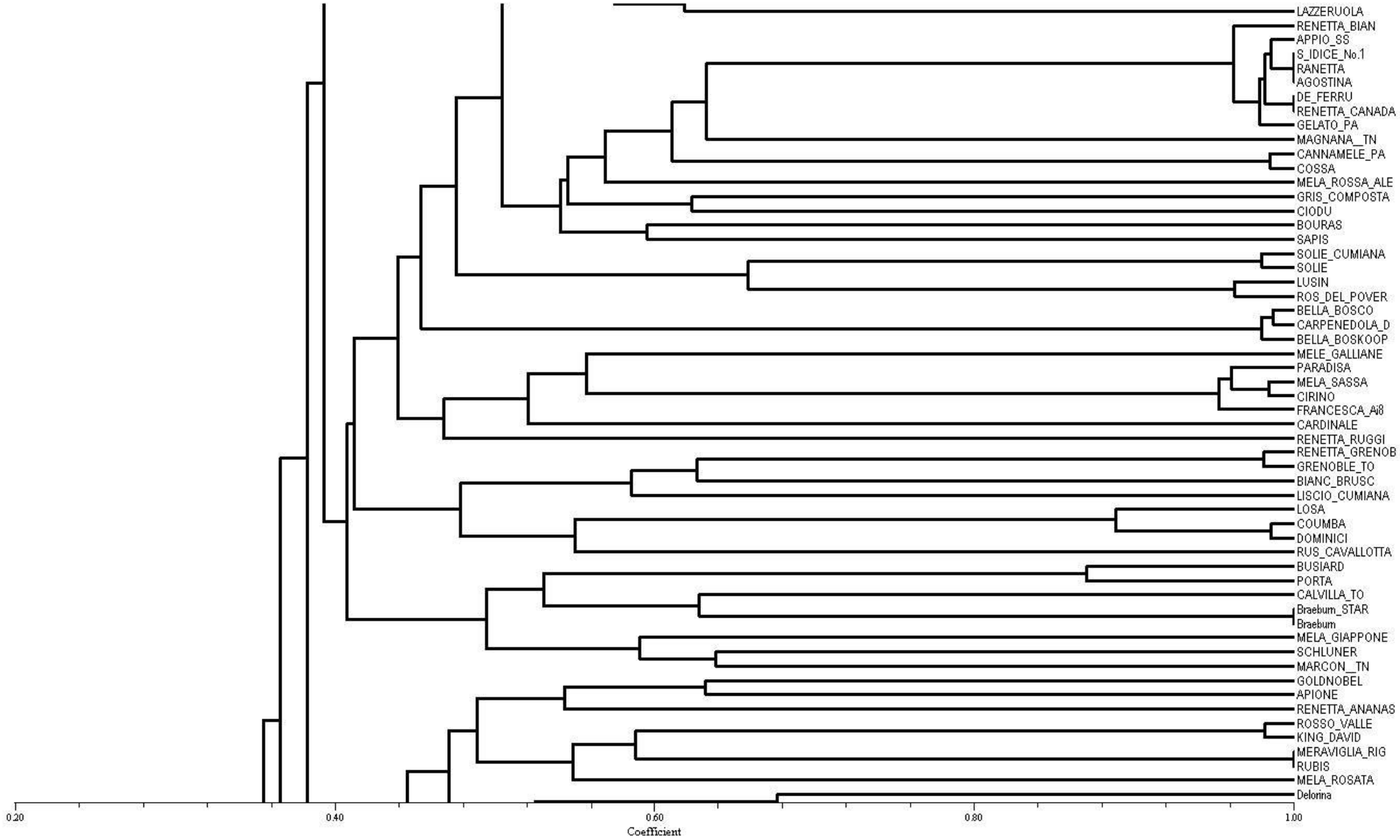


Figure 2.4 B. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

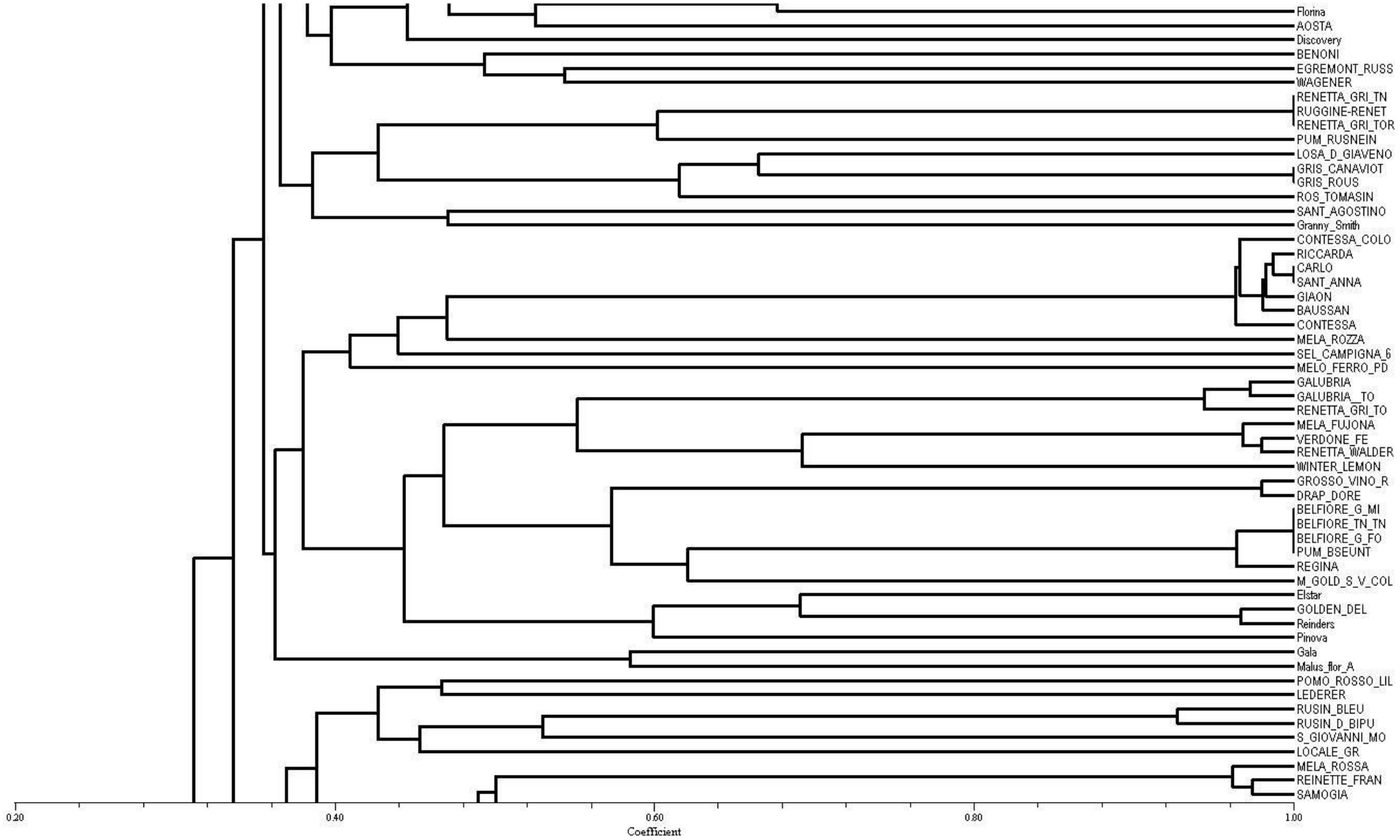


Figure 2.4 C. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

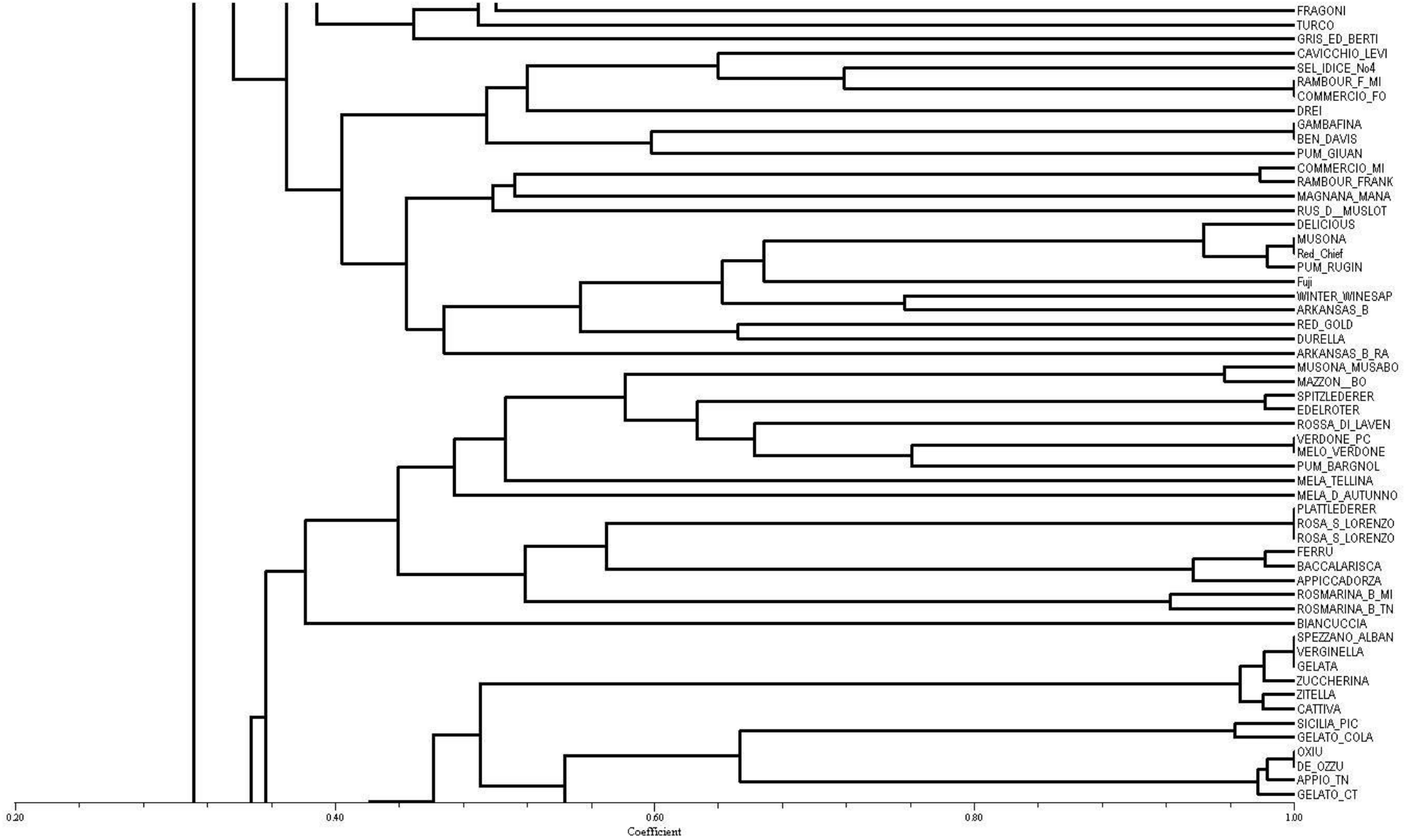


Figure 2.4 D. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

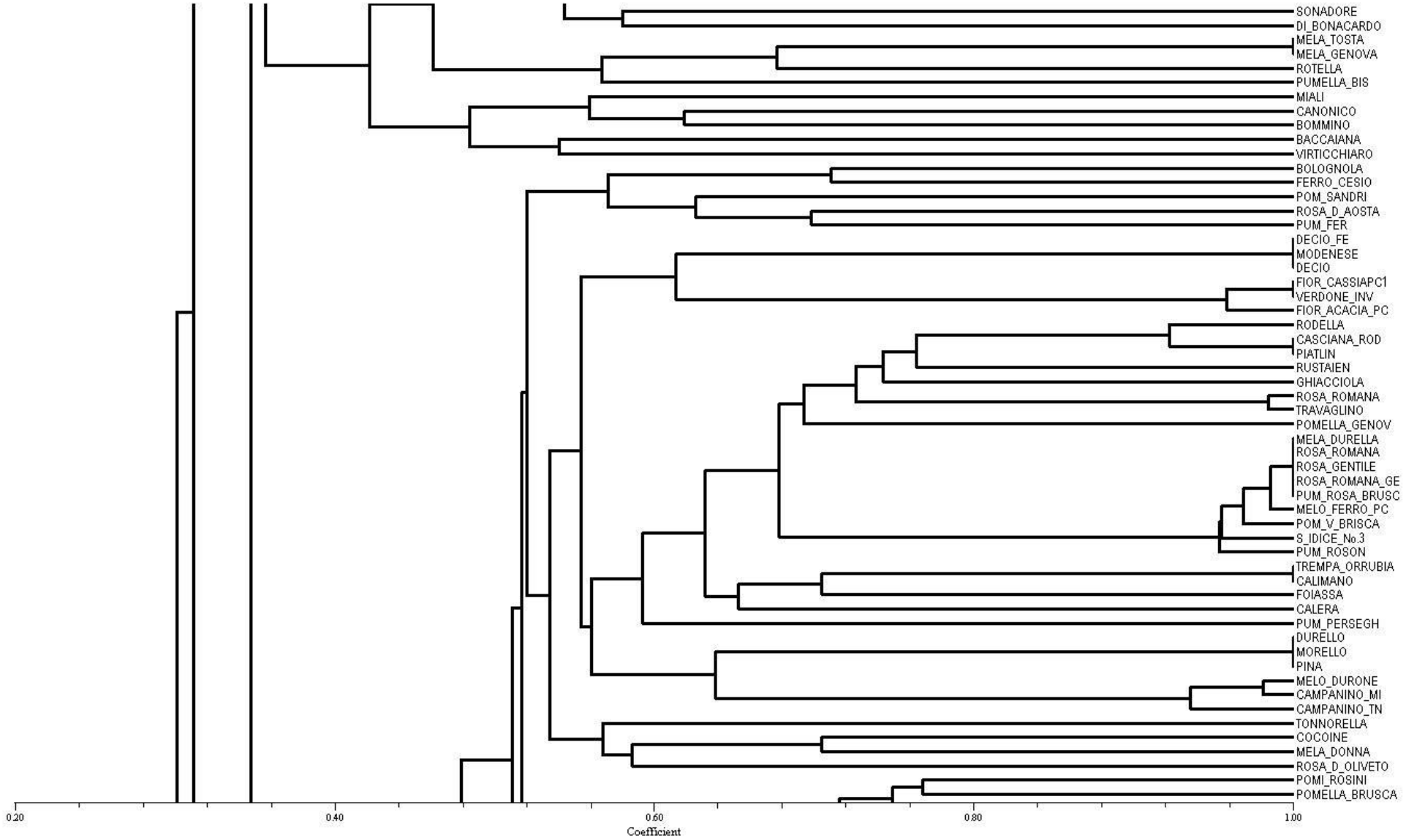


Figure 2.4 E. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

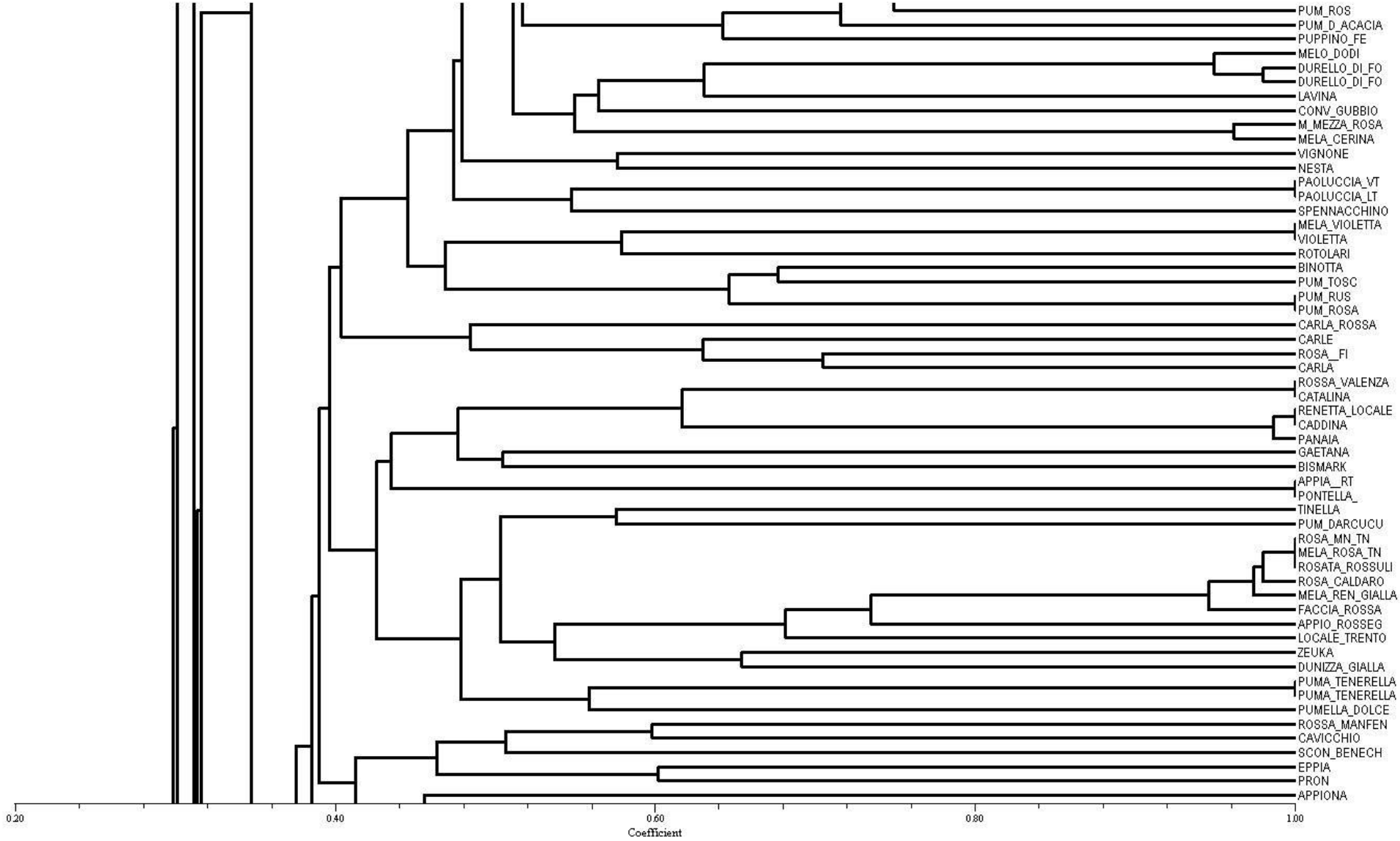


Figure 2.4 F. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

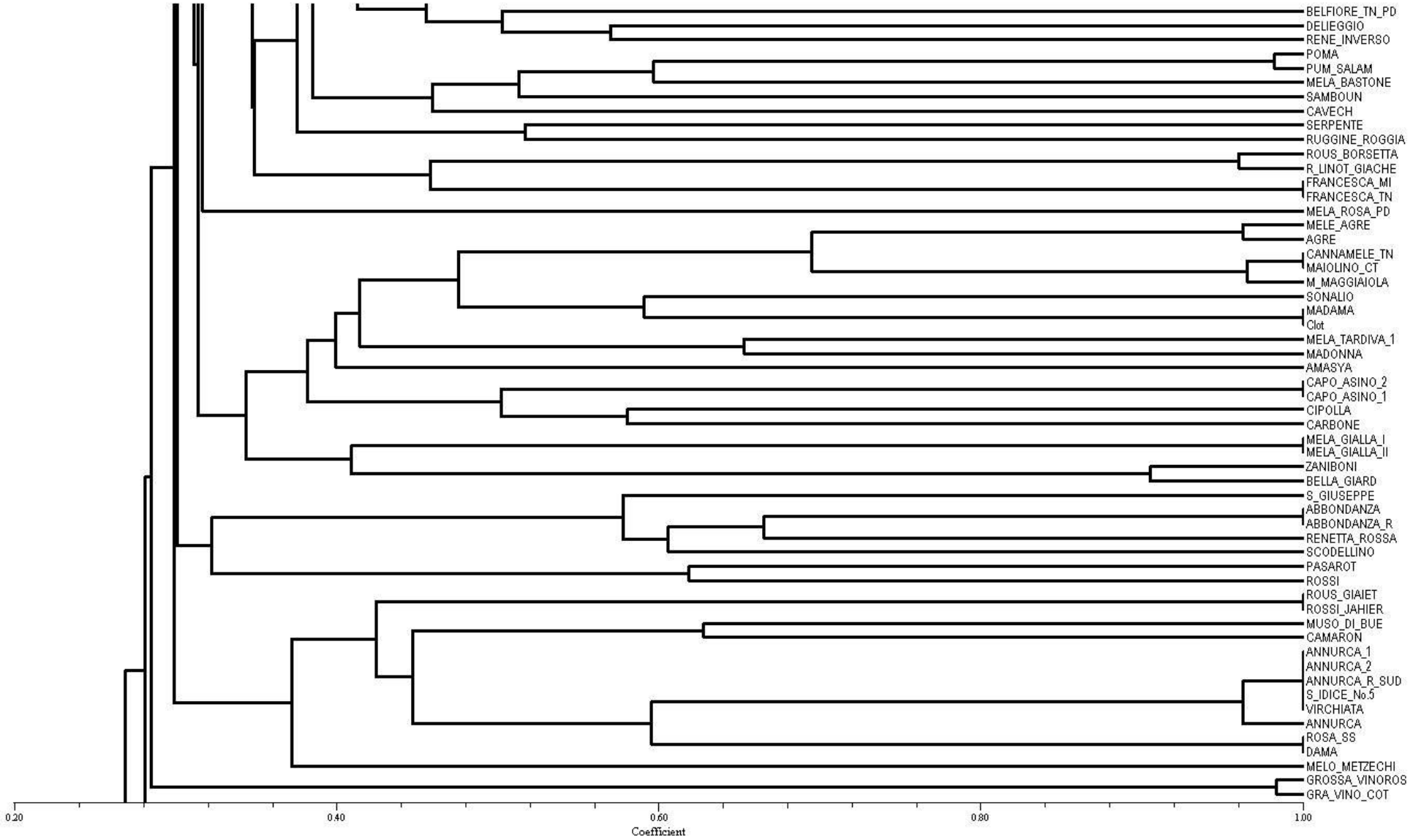


Figure 2.4 G. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

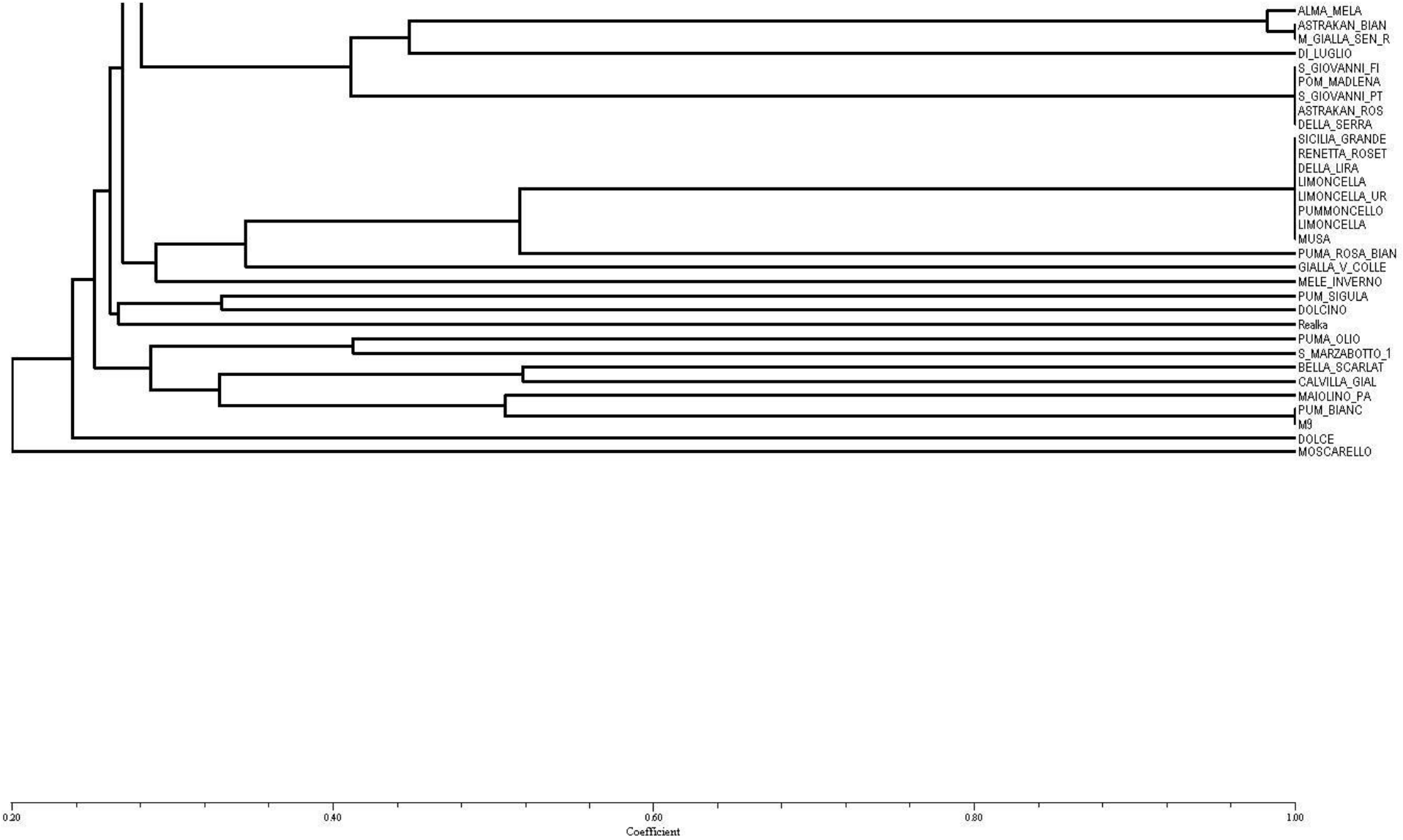


Figure 2.4 H. Continued dendogram of 424 apple accessions from Italian germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 266 alleles found in 15 SSR loci

Three big cluster groups spreading over the dendrogram comprised of only triploid accessions (Figure 2.5) with a value of R more than 0.92 were detected.

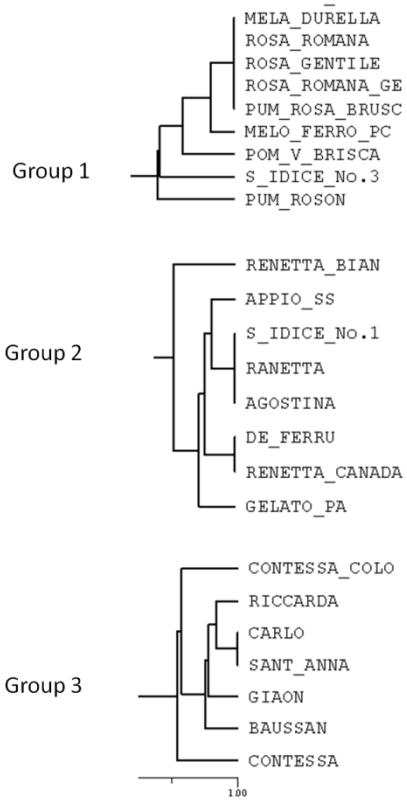


Figure 2.5. Three groups of triploids clustered on the dendrogram

Moreover, some accessions of the similar name but not clustered in the same groups classified as homonymous were also identified. For example, Verdone (PC) and Verdone (FE) were two accessions with identical name but collected from different place, Piacenza and Ferrara, respectively. The clustering results (Figure 2.6 A) showed that Verdone (PC) identical to Melo Verdone was actually triploid while Verdone (FE) was diploid and similar to the accession Renetta Walder, suggesting that Verdone (PC) and Verdone (FE) were two different apple genotypes. Moreover, the phenotype profiles (Figure 2.6 B and C) showed that the apple skin color of Verdone (FE) was totally green while the Verdone (PC) was with a little bit of red flush, providing another strong evidence to prove that they are in fact two different genotypes in spite of the identical name. Therefore, Verdone (PC) and Verdone (FE)

were a good example of homonyms in our germplasm.

Figure 2.6. A. Verdone (FE) and Verdone (PC) were clusted in different group on the dendrogram; B and C. Picture of two homonym accessions from germplasm: Verdone (PC) (B) and Verdone (FE) (C).

Other examples are with pairs of accessions that have been collected in different places in Italy: Calvilla (MI) and Calvilla (TO); Cannamele (PA) and Cannamele (TN); Melo Ferro (PD) and Melo Ferro (PC); San Giovanni (FI)/San Giovanni (PT) and San Giovanni (MO); Rosa (SS) and Rosa (FI); Maiolino (PA) and Maiolino (CT); Appio (TN) and Appio (SS); Gelato (CT) and Gelato (PA); Grenoble (TO) and Grenoble (PC).

Based on the level of variation observed in the study, the estimated probability of observing identical multilocus SSR profiles by chance was 2.89×10^{-30}, indicating a high resolution potential of the marker set we used. Consequently, those accessions with identical SSR profiles may be considered duplicates in terms of collection management. The potential redundancy found in other apple or pear collections from Spain and Netherland (Bassil et al., 2008; Pereira-Lorenzo et al., 2008; van Treuren et al., 2010; Ferreira dos Santos et al., 2011) was reported around 30% ~ 50%, whereas a rather low redundancy rate around 17% was identified in our apple germplasm. Therefore, a relative higher diversity value contained in our apple germplam collections compared to others germplasms. High levels of duplication within a collection must be avoided due to the high maintenance cost of field collections (Hokanson et al., 1998), but a certain level of duplication provides a safety backup system. The SSR profiles obtained in this study, along with morpho-agronomic evaluation will be used to critically re-examine the composition of the collections and the possibility of removal of certain accessions duplicated within collections. Moreover, apparent misidentifications observed in our study, which were six accessions identical to Law Red Rome Beauty and one accession identical to Red Chief, revealed that foreign cultivars was penetrating Italian's

native apple populations and the resultant varieties were given local names that make them indistinguishable without the use of molecular marker analyses.

2.4.5 Characterization by SSRs of the region controlling red skin color in apple germplasm

The region on LG9 which control red skin color in apple has been analyzed on the whole accession panel by 4 SSRs (CN444542, CH01H02, Hi05e07 and CH01f03b). These SSRs allowed to identify a total of 75 alleles and their relative frequency was analyzed according to the distribution of the genotypes in classes of skin overcolor (high, medium, low and no overcolor). For this fact only the 366 (68, 97, 58, 143 respectively in the above described classes) varieties for which data about overcolor were available have been used for linkage analysis.

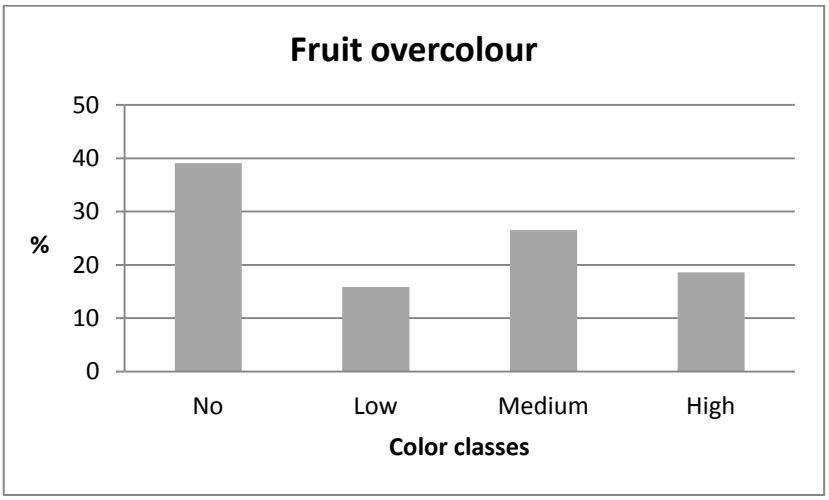


Figure 2.7. Distribution of fruit overcolour in the analyzed germplasm collection

In figure 2.7 is shown the distribution in classes of overcolour of the analyzed germplasm collection. About 40% of the plants have no coloration at all while 46% have medium to high level of red skin coverage. The low overcolour class include mainly accessions with a very low red color that are difficult to classify for two possible reasons. Some red accessions

could have a very low coloration due the unfavorable environmental conditions that hamper the fruit color development. In fact, these conditions are typical of the Pianura Padana valley where these plants are grown. On the other hand some other accessions present high levels of red blushing but only on the more exposed fruits within a tree. Therefore, other genetic factors could be involved in the formation of fruit red blush but the main gene controlling fruit red color is missing.

The SSR marker which evidenced the strongest linkage with red skin color was CN444542. In fact 66 accessions over 68 (97.1% frequency) that are grouped in the class of high overcolor and 82 on 97 (84.5% frequency) in the medium overcolor one are carrying the allele of 164bp size. On the contrary only 11 on 143 of the not red varieties had this allele (Table 2.8).

Table 2.8. Frequencies and chi-square analysis of the CN555452 SSR alleles. In red are highlighted chi-square values indicating a linkage with red overcolor while in yellow the significant chi-square values for the absence of red color.

CN444542 allele size and distribution of the observed frequencies																		
	152	154	156	158	160	162	164	166	168	170	171	173	175	179	181	183	189	197
high overcolor	38,2	7,4	4,4	1,5	0,0	1,5	97,1	1,5	8,8	0,0	0,0	17,6	0,0	0,0	0,0	7,4	0,0	0,0
medium overcolor	47,4	3,1	4,1	2,1	1,0	3,1	84,5	1,0	7,2	1,0	0,0	19,6	0,0	0,0	0,0	13,4	3,1	0,0
low overcolor	77,6	10,3	1,7	1,7	0,0	1,7	32,8	0,0	8,6	1,7	0,0	32,8	0,0	3,4	0,0	1,7	0,0	1,7
no color	53,8	7,0	9,8	4,2	0,0	3,5	7,7	0,7	18,9	0,7	0,7	59,4	0,7	3,5	3,5	17,5	0,0	0,7
expected freq.	53,0	6,6	6,0	2,7	0,3	2,7	48,6	0,8	12,3	0,8	0,3	36,9	0,3	1,9	1,4	12,0	0,8	0,5
chi-square																		
high overcolor	4,1	0,1	0,4	0,6	0,3	0,6	48,2	0,5	1,0	0,8	0,3	10,0	0,3	1,9	1,4	1,8	0,8	0,5
medium overcolor	0,6	1,8	0,6	0,2	2,1	0,0	26,5	0,1	2,1	0,1	0,3	8,1	0,3	1,9	1,4	0,2	6,3	0,5
low overcolor	11,4	2,2	3,1	0,4	0,3	0,4	5,2	0,8	1,1	1,0	0,3	0,5	0,3	1,2	1,4	8,8	0,8	2,5
no color	0,0	0,0	2,4	0,8	0,3	0,2	34,5	0,0	3,5	0,0	0,7	13,8	0,7	1,3	3,3	2,5	0,8	0,0

I

The chi-square test (3 degree of freedom and threshold of 11.345 for 0.01 significance) confirmed that the 164bp allele is in a very strong linkage with this trait. This analysis also evidenced that an allele of 173bp is more common in non red varieties (13.8 of chi-square) while 45 of the 58 accessions with low overcolor (77,6% and 11.4 of chi-square) presented an allele of 152bp. Many of the varieties that are carrying this 152bp allele in this class are lacking of the 164bp allele (38 on 45).

It is interesting to note that most of red varieties that are lacking the 164 bp allele are clustering together in small clusters. This is the case of the 'Rosa' , 'Campanino' and 'Paoluccia' groups (see figure 2.8).

Within the group of 11 accessions with no fruit coloration carrying the 164bp allele there are 4 accessions with almost completely russeted fruits that could mask the possible presence of some red color (Lederer, Mela Rozza, Gris Canaviot and Bella di Boskoop). Other fully russeted accessions do not have the 164 allele and they are located in various different positions in the whole cluster (i.e. Gris d'la Composta, Egremont Russet, Renetta Grigia di Torriana, Ruggine-Roggia, and Pum Rusnein).

A

Cluster	Accession name	Color	CN444542 allele	
			1	2
1	MELO DURONE	M	152	-
	CAMPANINO (MI)	M	152	-
	CAMPANINO (TN)	M	152	-
2	ROSA MANTOVANA (TN)	L	152	-
	MELA ROSA (TN)	M	152	-
	ROSATA ROSSULINA	M	152	-
	ROSA DI CALDARO	M	152	-
	MELA RENETTA GIALLA	M	152	-
	FACCIA ROSSA	M	152	-
	APPIO ROSSEGGIANTE	H	152	156
3	PAOLUCCIA (VT)	M	152	170
	PAOLUCCIA (LT)	L	152	170
	SPENNACCHINO	M	152	-

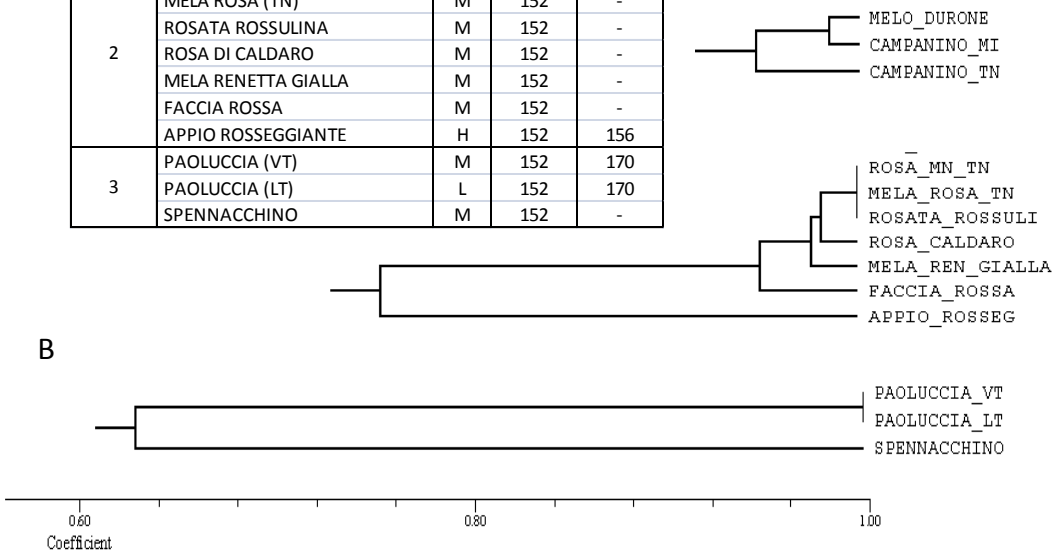


Figure 2.8. CN444542 allelic composition of the red fruit accessions without the 164 allele (A) and their overall degree of relatedness after cluster analysis (B)

The strong linkage of CN444542 was confirmed by BLAST analysis on the apple whole genome sequence: in fact CN444542 sequence (acc. number CN444542.1) was located in scaffold 9 position 29456305 at 15589bp of distance from a MdMYB10 transcription factor (a sequence from the cultivar Royal Gala, acc. number EU518249, was used for BLAST) which was located in position 29471894. A large literature described MdMYB10 as the gene that play the key role in the control of the anthocyanin biosynthesis in apple (Ban et al 2007, Espley et al 2007; Takos et al 2006; Chagnè et al 2007). The same role for this gene was also

described in other species belonging to the Rosaceae family as pear, peach and apricot (Lin-Wang et al 2010; Pierantoni et al 2010).

Data about allele frequency distribution for the other analyzed SSRs did not evidenced linkage with the red skin overcolor. They are far from CN444542 about 10cM (CH01H02), 26cM (Hi05e07 and 45 cM (CH01F03; data from http://www.hidras.unimi.it/). These distances are very high and not compatible with Linkage Disequilibrium that in apple decays very fast. This element should be taken into account in planning association mapping approaches in this species. In fact even an 8k SNP array for apple was demonstrated to have not resolution high enough for association mapping studies using unrelated germplasm (Micheletti et al 2011; Chagné et al 2012).

2.5. Conclusions

A total of 353 unique apple genotypes including 276 diploids and 77 triploids were identified through the germplasm genetic diversity analysis with a set of 15 polymorphic SSR markers containing high PIC values. A relatively low accession redundancy rate of 17% suggested a high diversity level of our germplasm collections compared to the other holders of apple and pear germplasm in Europe. Besides, the number of triploids (22%) is higher than expected, and this can be explained with the habit to select large fruit size genotypes, typical of polyploids. Additionally, the identification of 7 accessions which are identical to reference cultivars emphasized the importance of verifying germplasm repository collections with powerful tools such as molecular markers, to avoid synonymous repetitions, remove duplications and distribute correctly identified cultivars to nurseries for propagation. However, this wide panel of selected unique genotypes could be defined as an Italina core collection that will be used to improve the knowledge of genetic variability. By exploring the phenotypic and allelic diversity available in our apple germplasm it will be possible to identify the genomic regions contributing to the genetic control of major horticultural traits through genome-wide association genetics. An EU-funded project Fruit Breedomics (2011-2015) is currently running and the core collection identified in this thesis will be deeply genotyped with an Affimetrixs high-density SNP chip that is under development at Fondazione E. Mach, San Michele all'Adige (TN).

2.6. References

Ban Y., Honda C., Hatsuyama Y., Igarashi M., Bessho H., Moriguchi T., “Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin”, Plant Cell Physiol. 48(9) (2007) 58–970.

Berg, E. E. and J. Hamrick (1997). "Quantification of genetic diversity at allozyme loci." Canadian Journal of Forest Research 27(3): 415-424.

Boss P.K., Davies C., Robinson S.P., “Analysis of the Expression of Anthocyanin Pathway Genes in Developing *Vitis vinifera* L. cv Shiraz Grape Berries and the Implications for Pathway Regulation”, Plant Physiol. 111(1996):1059-1066

Botstein, D., R. L. White, M. Skolnick and R. W. Davis (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." American journal of human genetics 32(3): 314.

Buschiazzo, E. and N. J. Gemmell (2006). "The rise, fall and renaissance of microsatellites in eukaryotic genomes." Bioessays 28(10): 1040-1050.

Celton, J.-M., D. Tustin, D. Chagne and S. Gardiner (2009). "Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from Malus ESTs and Pyrus genomic sequences." Tree Genetics & Genomes 5(1): 93-107.

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

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, et al. (2012) “Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple”. PLoS ONE 7(2): e31745.

Coart, E., X. Vekemans, M. J. Smulders, I. Wagner, J. Van Huylenbroeck, E. Van Bockstaele and I. Roldán-Ruiz (2003). "Genetic variation in the endangered wild apple (Malus sylvestris (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers." Molecular Ecology 12(4): 845-857.

Csencsics, D., S. Brodbeck and R. Holderegger (2010). "Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (Typha minima) using next-generation sequencing technology." Journal of Heredity 101(6): 789-793.

Cummins, J. N. (1994). "Register of new fruit and nut varieties Brooks and Olmo: list 36." HortScience 29(9): 942-969.

Dutta, S., G. Kumawat, B. P. Singh, D. K. Gupta, S. Singh, V. Dogra, K. Gaikwad, T. R. Sharma, R. S. Raje and T. K. Bandhopadhya (2011). "Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]." BMC plant biology 11(1): 17.

Espley R.V., Hellens R.P., Putterill J., Stevenson D.E., Kutty- Amma S., Allan A.C., “Red colouration in apple fruit is due to the activity of the MYB transcription factor”, MdMYB10, Plant J. 49 (2007) 414-427.

Ferreira dos Santos, A. R., A. M. Ramos-Cabrer, M. B. Díaz-Hernández and S. Pereira-Lorenzo (2011). "Genetic variability and diversification process in local pear cultivars from northwestern Spain using microsatellites." Tree Genetics & Genomes 7(5): 1041-1056.

Garkava-Gustavsson, L., K. Antonius, D. McDonagh, T. Gallagher and H. Nybom (2010). Do We Preserve Unique Apple Germplasm? XXVIII International Horticultural Congress on Science and Horticulture for People (IHC2010): III International Symposium on 918.

Garkava-Gustavsson, L., A. Kolodinska Brantestam, J. Sehic and H. Nybom (2008). "Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis." Hereditas 145(3): 99-112.

Gasi, F., S. Simon, N. Pojskic, M. Kurtovic and I. Pejic (2010). "Genetic assessment of apple germplasm in Bosnia and Herzegovina using microsatellite and morphologic markers." Scientia Horticulturae 126(2): 164-171.

Gharghani, A., Z. Zamani, A. Talaie, N. C. Oraguzie, R. Fatahi, H. Hajnajari, C. Wiedow and S. E. Gardiner (2009). "Genetic identity and relationships of Iranian apple (Malus× domestica Borkh.) cultivars and landraces, wild Malus species and representative old apple cultivars based on simple sequence repeat (SSR) marker analysis." Genetic resources and crop evolution 56(6): 829-842.

Guarino, C., S. Santoro, L. De Simone, O. Lain, G. Cipriani and R. Testolin (2006). "Genetic diversity in a collection of ancient cultivars of apple (Malus × domestica Borkh.) as revealed by SSR-based

fingerprinting." Journal of horticultural science & biotechnology 81(1): 39-44.
Guilford, P., S. Prakash, J. Zhu, E. Rikkerink, S. Gardiner, H. Bassett and R. Forster (1997). "Microsatellites in Malus X domestica (apple): abundance, polymorphism and cultivar identification." TAG Theoretical and Applied Genetics 94(2): 249-254.
Han, Y. and S. S. Korban (2008). "An overview of the apple genome through BAC end sequence analysis." Plant molecular biology 67(6): 581-588.
Hardy, O. J. and X. Vekemans (2002). "SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels." Molecular Ecology Notes 2(4): 618-620.
Hayden, M. J., T. Nguyen, A. Waterman, G. McMichael and K. J. Chalmers (2008). "Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat." Molecular breeding 21(3): 271-281.
He, Q., X. W. Li, G. L. Liang, K. Ji, Q. G. Guo, W. M. Yuan, G. Z. Zhou, K. S. Chen, W. E. van de Weg and Z. S. Gao (2011). "Genetic diversity and identity of Chinese loquat cultivars/accessions (Eriobotrya japonica) using apple SSR markers." Plant Molecular Biology Reporter 29(1): 197-208.
Hearne, C. M., S. Ghosh and J. A. Todd (1992). "Microsatellites for linkage analysis of genetic traits." Trends in Genetics 8(8): 288-294.
Henegariu, O., N. Heerema, S. Dlouhy, G. Vance and P. Vogt (1997). "Multiplex PCR: critical parameters and step-by-step protocol." Biotechniques 23(3): 504-511.
Hokanson, S., A. Szewc-McFadden, W. Lamboy and J. McFerson (1998). "Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a Malus× domestica Borkh. core subset collection." TAG Theoretical and Applied Genetics 97(5): 671-683.
Honda C., Kotoda N., Wada M., KondoS., Kobayashi S., SoejimaJ., Zhang Z., Tsuda T., Moriguchi T., “Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin”, Plant Physio. Biochem. 40 (2002) 955–962.
Hummer, K. E. and J. Janick (2009). "Rosaceae: taxonomy, economic importance, genomics." Genetics and Genomics of Rosaceae: 1-17.
Janick, J. and J. N. Moore (1996). Fruit breeding. Volume I. Tree and tropical fruits, John Wiley and Sons.
Kalinowski, S. T., M. L. Taper and T. C. Marshall (2007). "Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment." Molecular ecology 16(5): 1099-1106.
Kelkar, Y. D., S. Tyekucheva, F. Chiaromonte and K. D. Makova (2008). "The genome-wide determinants of human and chimpanzee microsatellite evolution." Genome research 18(1): 30-38.
Kobayashi S., Goto-Yamamoto N., Hirochika H., “Retrotransposon-induced mutations in grape skin color”, Science 304 (2004) 982–982.
Kumar, S., K. Tamura and M. Nei (1994). "MEGA: molecular evolutionary genetics analysis software for Kobayashi S., Goto-Yamamoto N., Hirochika H., “Retrotransposon-induced mutations in grape skin color”, Science 304 (2004) 982–982. microcomputers." Computer applications in the biosciences: CABIOS 10(2): 189-191.
Liebhard, R., L. Gianfranceschi, B. Koller, C. Ryder, R. Tarchini, E. Van de Weg and C. Gessler (2002). "Development and characterisation of 140 new microsatellites in apple (Malus x domestica Borkh.)." Molecular Breeding 10(4): 217-241.
Lin-Wang K., Bolitho K., Grafton K., Kortsee A., Karunairetnam S., McGhie T.K., Espley R.V., Hellens R.P., Allan A.C., “An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae”. BMC Plant Biol. 2010, 10:50.
Lu, Y., J. Yan, C. T. Guimaraes, S. Taba, Z. Hao, S. Gao, S. Chen, J. Li, S. Zhang and B. S. Vivek (2009). "Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms." TAG Theoretical and Applied Genetics 120(1): 93-115.
Mansfield, J. C., H. Holden, J. Tarlow, F. Di Giovine, T. McDowell, A. Wilson, C. Holdsworth and G. Duff (1994). "Novel genetic association between ulcerative colitis and the anti-inflammatory cytokine interleukin-1 receptor antagonist." Gastroenterology 106(3): 637.
Micheletti D, Troggio M, Zharkikh A, Costa F, Malnoy M, et al. (2011) “Genetic diversity of the genus Malus and implications for linkage mapping with SNPs”,Tree Genetics & Genomes 7: 857–868.
Oetting, W., H. Lee, D. Flanders, G. Wiesner, T. Sellers and R. King (1995). "Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers." Genomics 30(3): 450-458.

Patzak, J., F. Paprštein, A. Henychová and J. Sedlák (2012). "Genetic diversity of Czech apple cultivars inferred from microsatellite markers analysis." Horticultural Sci.(Prague) 39.

Pereira-Lorenzo, S., A. Ramos-Cabrer, A. Gonzalez-Diaz and M. Diaz-Hernandez (2008). "Genetic assessment of local apple cultivars from La Palma, Spain, using simple sequence repeats (SSRs)." Scientia Horticulturae 117(2): 160-166.

Pereira-Lorenzo, S., A. M. Ramos-Cabrer and M. B. Díaz-Hernández (2007). "Evaluation of genetic identity and variation of local apple cultivars (Malus× domestica Borkh.) from Spain using microsatellite markers." Genetic Resources and Crop Evolution 54(2): 405-420.

Ponce, M., P. Robles and J. Micol (1999). "High-throughput genetic mapping in Arabidopsis thaliana." Molecular and General Genetics MGG 261(2): 408-415.

Powell, W., G. C. Machray and J. Provan (1996). "Polymorphism revealed by simple sequence repeats." Trends in plant science 1(7): 215-222.

Reim, S., A. Proft, S. Heinz and M. Höfer (2012). "Diversity of the European indigenous wild apple Malus sylvestris (L.) Mill. in the East Ore Mountains (Osterzgebirge), Germany: I. Morphological characterization." Genetic Resources and Crop Evolution 59(6): 1101-1114.

Richards, C. M., G. M. Volk, A. A. Reilley, A. D. Henk, D. R. Lockwood, P. A. Reeves and P. L. Forsline (2009). "Genetic diversity and population structure in Malus sieversii, a wild progenitor species of domesticated apple." Tree Genetics & Genomes 5(2): 339-347.

Rohlf, F. J. N.-P. (1994). "Version 2.10." New York: Applied Biostatistics Inc.

Silfverberg-Dilworth, E., C. Matasci, W. Van de Weg, M. P. Van Kaauwen, M. Walser, L. Kodde, V. Soglio, L. Gianfranceschi, C. Durel and F. Costa (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree Genetics & Genomes 2(4): 202-224.

Tang, S., V. K. Kishore and S. J. Knapp (2003). "PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower." TAG Theoretical and Applied Genetics 107(1): 6-19.

Tanksley, S. (1993). "QTL analysis of transgressive segregation in an interspecific tomato cross." Genetics 134(2): 585-596.

Takos A.M., Jaffe F., Jacob S.R., Bogs J., Robinson S.P., Walker A.R., "Light induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples", Plant Physiol. 142 (2006) 1216–1232.

Urrestarazu, J., C. Miranda, L. G. Santesteban and J. B. Royo (2012). "Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers." Tree Genetics & Genomes: 1-18.

Van Treuren, R., H. Kemp, G. Ernsting, B. Jongejans, H. Houtman and L. Visser (2010). "Microsatellite genotyping of apple (Malus× domestica Borkh.) genetic resources in the Netherlands: application in collection management and variety identification." Genetic resources and crop evolution 57(6): 853-865.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. v. d. Lee, M. Hornes, A. Friters, J. Pot, J. Paleman and M. Kuiper (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic acids research 23(21): 4407-4414.

Xiao, J., J. Li, S. Grandillo, S. N. Ahn, L. Yuan, S. D. Tanksley and S. R. McCouch (1998). "Identification of trait-improving quantitative trait loci alleles from a wild rice relative, Oryza rufipogon." Genetics 150(2): 899-909.

Xu, Y. (2010). Molecular plant breeding, Cabi.

Xu, Y., T. Ishii, S. McCouch, T. Mew, D. Brar, S. Peng, D. Dawe and B. Hardy (2003). "Marker-assisted evaluation of germplasm resources for plant breeding." Mew, TW, Brar, DS, Peng, S., Dawe. D. & Hardy, B.(Eds.): Rice Science: Innovations and Impact for Livelihood.-Manila (IRRI): 213-229.

Ziegle, J. S., Y. Su, K. P. Corcoran, L. Nie, P. Eric Mayrand, L. B. Hoff, L. J. McBride, M. N. Kronick and S. R. Diehl (1992). "Application of automated DNA sizing technology for genotyping microsatellite loci." Genomics 14(4): 1026-1031.

Chapter 3

Fine mapping of R genes

3.1. Introduction

Apple scab, caused by the fungal pathogen *Venturia inaequalis* (Cooke) G.Wint., is a widespread and one of the most devastating diseases for apples (*Malus x domestica* Borkh) particularly in temperate zones with humid growing seasons. Both leaves and fruit can be affected by this fungal pathogen. If uncontrolled, the fungus may defoliate trees and blemish fruits to a point where they are unmarketable with marked production losses. Fungicides represent around 30% of the total pesticides used in Europe (Gullino and Kuijpers, 1994) and their use is important in intensive production system, such as fruit orchards. In some years and in some parts of Europe, the number of fungicide treatment against apple diseases exceeds 30, with around 20 to control scab. Such a large amount of chemical treatment raises numerous ecological problems and consumer health concerns, in addition to the economic cost (Lespinasse et al., 2002).

The use of resistant cultivars, which can be grown with much less pesticide treatment, could reduce the cost to the growers and may also contribute to a cleaner environment and to a reduction of fungicide residuals on apples for consumers. Therefore, the selection for the resistant cultivars is now a central theme in most apple breeding programs. Two different resistance gene sources are known, namely qualitative and quantitative resistance. For qualitative resistances, cultivars are highly resistant and the resistance is transmitted to the progeny in simple a mendelian fashion by a single or limited number of loci conferring the resistance. Whereas, the quantitative resistance is assumed to be polygenic, the result of the combined action of many factors each providing a modest contribution to the overall resistance.

In *Malus* species, a total of 16 major scab resistance genes and various quantitative trait loci (QTLs) conferring resistance to a broad spectrum of *V. inaequalis* strains, have been discovered and mapped (Calenge et al., 2004; Schmidt and Van de Weg, 2005; Gessler et al., 2006; Gardiner et al., 2006; Bus et al., 2009; Patocchi et al., 2009; Soriano et al., 2009). In Table 3.1 are reported the main differential hosts and the main resistance genes identified so far in apple (Patocchi et al., 2009). Throughout all the manuscript we will use the new

nomenclature of the apple scab resistance genes proposed by Bus et al. (2009), and in some cases the historical name in brackets.

Table 3.1. New set of differential hosts and major scab resistance genes in apple (Patocchi et al. 2009)

Host	Accession	Origin of the plant material	R-gene	
			Historical name	New name[g]
h0	'Gala'	ACW	None known	
h1	'Golden Delicious'	ACW	V*g*	*Rvi1*
h2	TSR34T15	ACW	V*h2*	*Rvi2*
h3	Q71[a]	Plant and Food Research	V*h3.1*	*Rvi3*
h4	TSR33T239	INRA	V*h4*	*Rvi4*
h5	9-AR2T196	INRA	V*m*	*Rvi5*
h6	'Priscilla'	INRA	V*f*	*Rvi6*
h7	F1 *M. floribunda* 821[b]	ACW	V*fh*	*Rvi7*
h8	B45[c]	Plant and Food Research	V*h8*	*Rvi8*
h9	J34[d]	Plant and Food Research	V*dg*	*Rvi9*
h10	A 723-6[e]	ACW	V*a*	*Rvi10*
h11	*M. baccata jackii*	ACW	V*bj*	*Rvi11*
h12	Hansen's baccata #2	ACW	V*b*	*Rvi12*
h13	'Durello di Forlı˙'	DCA	V*d*	*Rvi13*
h14	'Du lmener Rosenapfel'	INRA		*Rvi14*[f]
h15	GMAL 2473	ACW	V*r2*	*Rvi15*

[a] F1 of 'Geneva' x 'Braeburn'

[b] Currently not available, temporarily replaced by *M. floribunda* 821

[c] F1 of 'Pacific Beauty' x *M. sieversii* GMAL4302-X8

[d] F1 of 'Gala' x 'Dolgo'

[e] F1 of 'Worcester Pearmain' X PI172623

[f] First apple scab resistance gene named according the nomenclature proposed by Bus et al. (2009)

[g] For the historical apple scab resistance gene *Vr* no new name is currently available (see Bus et al. 2009)

3.1.1. Apple Scab Resistance Genes Mapping

Apple scab resistance genes have received the most attention by genetic mapping groups, because of the significance of the economic impact of apple scab on production. It is also relatively easy to identify markers for major resistances to apple scab in comparison with other resistances. This is because of the relative reliability of phenotypes obtained from glasshouse screening of very young seedlings from mapping populations for response to infection by *Venturia inaequalis*. A number of major apple scab resistance genes have now been assigned to linkage groups: *Rvi6* (*Vf*; Maliepaard et al., 1998; Vinatzer et al., 2001) to LG 1, *Rvi5* (*Vm*) to LG 17 (Patocchi et al., 2005), *Rvi15* (*Vr2)*, *Rvi2* (*Vh2), Rvi4 (Vh4), Vt57, Rvi11 (Vbj)* and *Rvi8* (*Vh8)* to LG 2 (Bus et al.,2004; Gygax et al., 2004; Patocchi et al., 2004; Bus et al., 2005a,b), *Rvi13* (*Vd)* to LG10 (Tartarini et al., 2004), *Rvi1* (*Vg)* and *Rvi12* (*Vb)* to LG

12 (Durel et al., 1999; Erdin et al., 2006).

To date, the most successful breeding programs for apple scab resistant cultivars have been based on the introgression into susceptible cultivars of the major resistance gene *Rvi6* (*Vf)* from *Malus floribunda* 821 (Janick et al., 1996). The *Rvi6* gene conferring resistance to five out of seven known races of *V. inaequalis* has held up quite well in the orchards for over 80 years, until 1993 when two new races of the pathogen (races 6 in Germany and 7 in England) became able to produce scab lesions on some apple cultivars carrying *Rvi6* (Parisi et al., 1993; Roberts and Crute, 1994; Benaouf and Parisi, 2000). Unlike an annual species where if the resistance is overcome, a new variety can be planted and be profitable the following year, it can take at least 10 years to recoup the cost of planting the orchard. This is because an apple orchard remains in the ground for a production period of around 15 years and it takes several years for the trees to come into commercial production. It is therefore imperative that the resistance of new commercial varieties is durable. The combination of functionally different resistances genes in the same genotype, also known as “pyramiding” was proposed as a promising strategy to obtain a durable scab resistance over time (Gessler and Blaise, 1994; Laurens and Lespinasse, 1994; Fischer, 1995; Evans, 1997).

Most of the disease resistance genes characterized to date in apple are single dominant genes. Such genes commonly confer resistance to the pathogen in a gene-for-gene manner and are therefore in theory, easily overcome by the pathogen’s ability to mutate to virulence (Crute and Pink, 1996). It is likely that durable resistance to apple pathogens will be established through the pyramiding of different resistance genes with different resistance specificities into a single cultivar.

3.1.2. Apple Scab Resistance QTLs Mapping

Pathogen resistance conferred by QTLs would be a valuable addition to breeding portfolios of major resistance genes, as incorporating QTLs into a single cultivar is likely to be more effective than the combining of major genes alone (Parlevliet and Zadoks, 1977). QTLs have been identified for apple scab resistance using the reference genetic maps constructed in the following populations; Prima × Fiesta (Durel et al., 2003), Fiesta × Discovery (Liebhard et al., 2003c) and Discovery × TN10-8 (Calenge et al., 2004).

Durel et al. (2003) used two monoconidial strains of race 6 to identify QTLs controlling resistance in both Fiesta and Prima. Detailed QTL analysis using both MCQTL (Jourjon et al., 2000) and MapQTL (VanOoijen, 2004) software identified four genomic regions that were significantly involved in partial resistance, characterized by a reduction in sporulation. One of these regions was located close to the original *Rvi6* gene and it is possible that the observed partial resistance was due to a closely linked gene, or a result of a residual effect of the overcome *Rvi6* gene (Durel et al., 2003). The remaining three additional regions identified on LG 15, 11 and 17 were novel locations for association with scab resistance.

Liebhard et al. (2003c) carried out extensive assessment of field resistance to apple scab over a three year period involving three different geographical sites. Using MapQTL, eight QTLs were identified that contributed to apple scab resistance; six for leaf scab and two for fruit scab. Interestingly, although Discovery demonstrated a greater degree of resistance, most of the identified QTLs were attributed to Fiesta, the more susceptible parent, indicating a high degree of homozygosity at the resistance gene loci in Discovery that prevented their detection in the progeny because of the lack of segregation. The high levels of resistance observed in individuals during the study confirmed that Discovery was a strong resistant parent for breeding (Liebhard et al., 2003c). The strongest scab resistance QTL from Prima × Fiesta mapped to LG 17 (Liebhard et al., 2003c), coinciding with a scab resistance QTL that Durel et al.(2003) identified, and similarly LG 11 was identified in both studies as possessing a region of interest. One of the QTLs detected by Liebhard et al. (2003c) that accounted for 4% of the phenotypic variability was located on LG 12 in a position comparable to *Rvi1* (*Vg;* Van de Weg, unpublished data).

Calenge et al. (2004) used a panel of eight monoconidial isolates to inoculate replicated progeny from a Discovery × TN10-8 cross, resulting in the identification of numerous QTLs across seven linkage groups (with MapQTL), depending upon the isolate used. Combining QTLs with overlapping confidence intervals and close likelihood peaks revealed three major QTLs on LG 1, LG 2 and LG 17. The region identified on LG 1 corresponds to the region around *Rvi6* that Durel et al. (2003) identified as contributing between 16.0% and 17.8% of phenotypic variation, and the QTL identified on LG 17 (Calenge et al., 2004) is also in

agreement with a QTL mapped in both Fiesta and Discovery that explained 23% of the observed phenotypic variability. Calenge et al. (2004) also detected additional QTLs on LG 5, 13 and 15 to only one or two isolates and a QTL on LG 2 that appeared to control more broad-spectrum resistance to apple scab. This QTL spans a region around the major scab resistances *Rvi11*, *Rvi2 and Rvi8* (Calenge et al., 2004; Durel et al., 2004). The identification of isolate-specific QTL indicates that some partial resistance QTL could be involved in a pathogen-mediated recognition response, similar to major genes.

In this context, marker-assisted selection (MAS) could be a useful tool in order to accelerate apple breeding programs, for example, selecting the parents for the future crosses. However, in the MAS process, knowledge of the genetic distance between a marker and gene of interest is essential for its reliable application. Tightly-linked markers located on both sides of target genes are needed for accurate selection, reducing the risk of selecting unfavorable individuals resulting from recombination between the markers and the target gene.

3.1.3. The *Rvi13 (Vd)* scab resistance gene

The old Italian apple cultivar 'Durello di Forli' has been described as both conferring high field tolerance to apple scab (3b type reaction) and a stellate necrotic reaction in glasshouse grown seedlings exposed to the EU-D-42 race 6 reference strain of *V. inaequalis.* The QTL analysis of scab resistance in a progeny derived from the cross 'Durello di Forli' x 'Fiesta' made it possible to identify three main QTLs on LG10, 11 and 17. Although the scab resistance from 'Durello di Forli' is considered as quantitative (Sansavini et al., 2000), only a strong QTL on LG10 derived from this Italian cultivar was identified. The other QTLs were derived from the 'Fiesta' parent. The inoculation of the same progeny with a monoconidial strain of race 6 (EU-D-42) made it possible to identify a previously unknown major gene for scab resistance. This gene was mapped on LG10 and has been called *Rvi13* (*Vd*; Tartarini et al., 2004).

The markers OPAF07/880 (RAPD) and G63Tru91a (CAPS) that flank *Rvi13* are spanning about 10 cM at the top of LG10 (Tartarini et al., 2004). Four additional NBS markers identified by NBS profiling method were also found at the top of LG10 in the F1 progeny of 'Discovery' X

'TN10-8' (D X T) (Calenge et al., 2005). Moreover, two apple resistance genes homologues–derived markers (ARGH8 and ARGH25) which were highly considered as candidate genes for apple scab resistance were also developed and mapped on the LG10 of two different maps of Fiesta and Discovery, respectively (Baldi et al., 2004).

The availability of the genome sequence of the apple cultivar 'Golden Delicious' reported by Velasco et al. (2010) (http://www.rosaceae.org/), made it possible to easily design and develop new molecular markers in specific regions using an *in silico* approach, because the nucleotide sequences associated with their locations in the apple genome are anytime ready to be downloaded for 'in house' use. Besides, the predicted genes in each contigs were also able to be downloaded from the GDR website in the form of the whole gene sequences or coding sequences. Some literatures reported construction of fine genetic maps in apple by using this *in silico* approach. Until now, three literatures (Moriya et al., 2012, Baldi et al., 2012 and Bai et al., 2012) about *Co* locus, a gene controlling columnar growth habit and one literature (Xu et al., 2012) about *Ma* locus, a major gene controlling malic acid content in apple fruit, were all based on the GD genome sequence to develop and map SSR markers.

The availability of tightly linked markers to these resistances is an important prerequisite for MAS and pyramiding resistance genes to reinforce resistance. Therefore, a high resolution mapping of the apple resistance gene *Rvi13* is of great importance to narrow down the region conferring the *Rvi13* resistance on the LG10.

3.1.4. The *Rvi5 (Vm)* scab resistance gene

The inheritance of apple scab resistance from *M. micromalus* 245–38 and *M. atrosanguinea* 804 is complex (Shay and Hough, 1952; Dayton and Williams, 1970). Dayton and Williams (1970) identified a "pit-type" gene, which masked a "3-type" gene allelic with *Rvi6* in these two sources of resistance. The "pit-type" gene is known as *Rvi5* and induces a hypersensitive reaction upon inoculation with the races 1 to 4 of *V. inaequalis.* The *Rvi5* gene has been overcome by race 5 of *V. inaequalis*, which was discovered in England (William and Brown, 1968) but no further resistance breakdown have been reported after the first identification. For this reason, *Rvi5* has been used quite rarely in apple breeding. The Canadian breeding

program introduced two cultivars ('Murray' and 'Rouville') after two generations from the first cross with the wild species. Therefore, in these varieties the *Rvi5* resistance was included in a *Malus* x *domestica* background with an acceptable fruit quality that makes this gene interesting for pyramidization with the *Rvi6* gene and other apple scab resistance genes. Patocchi et al. (2005) reported the discovery of a genomic region carrying the *Rvi5* gene together with identification of a SSR marker, Hi07h02, tightly-linked to this gene by GSA (Genome Scanning Approach). Starting from this knowledge in 2004 an *Rvi5* gene cloning project was created as a joint collaboration between the Department of Fruit and Woody Plant Science of the Bologna University and the Fondazione E. Mach, S. Michele all'Adige, Trento.

A first contig of two BAC clones was identified by Cova, (2008) through the screening of the Murray BAC library with the Hi07h02 marker and a step of chromosome walking in the direction of the *Rvi5* gene (Figure 3.1). The second BAC was recently extracted at Fondazione E. Mach by the screening with the B-SSR (Komjanc et al., personal communicaiton) (Figure 3.1). A preliminary sequence of this clone was assembled in 6 contigs. To date, no overlapping between the available BAC sequence from the *Rvi5* locus was identified. The *Rvi5* region in the apple genome of Golden Delicious was not well covered making very difficulty the development of new markers directly from the genome sequence (Cova, 2008).

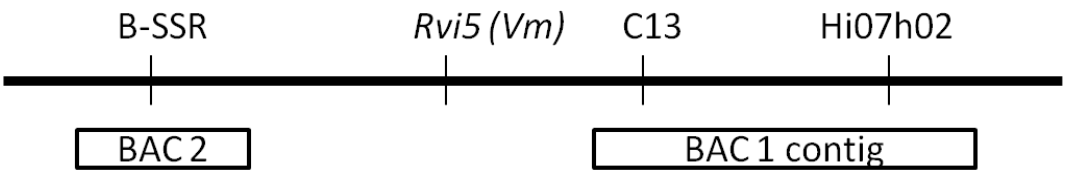


Figure 3.1. The figuration of two BAC clones surrounding of the *Rvi5 (Vm)* gene; BAC1 contig sequence from Cova, 2008 and BAC2 sequence is indicated by Komjanc, 2012 (pers. comm.)

3.1.5. Plant resistance genes

Induction of plant defense signaling involves the recognition of specific pathogen effectors

by the products of specialized host genes called R-genes (Belkhadir et al., 2004). A significant effort by several laboratories in the past 5–10 years has resulted in the identification of many R genes from model and crop species (Bent, 1996; Ellis et al., 2000 and Jones et al., 2001). Functional R genes isolated so far encode resistance to bacterial, viral, fungal, and even nematode and insect pathogens with very different lifestyles, outside or inside the plant cell. Despite this wide range of pathogen taxa and their presumed pathogenicity effector molecules, plant resistance genes can be broadly divided into eight groups based on their amino acid motif organization and their membrane spanning domains (Gugurani et al., 2012; Figure 3.2).

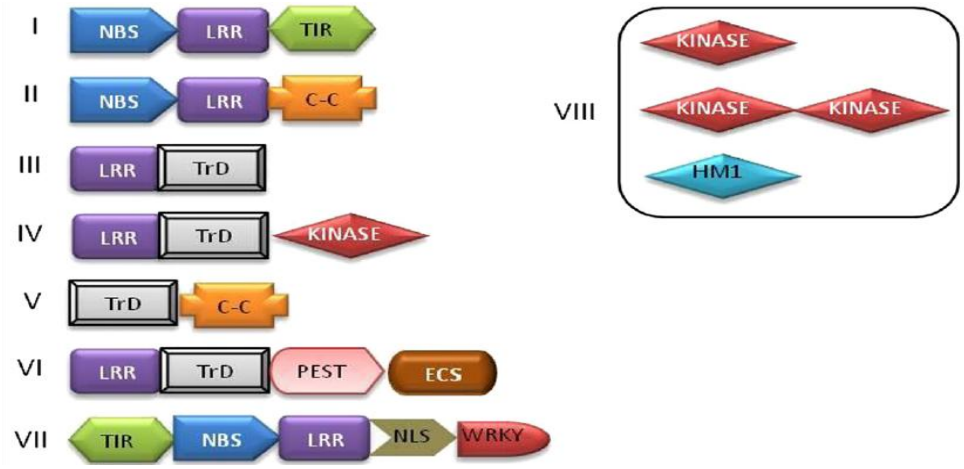


Figure 3.2. Major classes of plant resistance (R) genes based on the arrangement of the functional domains. LRR - Leucine rich repeats; NBS - Nucleotide-binding site; TIR Toll/Interleukin -1- receptors; C – C - Coiled coil; TrD - Transmembrane domain; PEST - Protein degradation domain (proline-glycine-serine-threonine); ECS - Endocytosis cell signaling domain; NLS - Nuclear localization signal; WRKY - Amino acid domain; HM1 - *Helminthosporium carbonum* toxin reductase enzyme (Gugurani et al., 2012).

The LRRs (Leucine rich repeats) represents the components having an important role for recognition specificity and these domains are present in the majority of R proteins (Jones, 2001). First major class of R-genes include the genes encoding for cytoplasm proteins with a nucleotide-binding site (NBS), a C terminal leucine rich repeat (LRR) and a putative coiled coil domain (CC) at the N- terminus (Gururani et al., 2012). The second class of resistance genes consists of cytoplasmic proteins which possess LRR and NBS motifs and an N-terminal domain with homology to the mammalian toll-interleukin-1-receptor (TIR) domain. Third major class of resistance genes family devoid of NBS motif consists of extra cytoplasmic leucine rich repeats (eLRR), attached to a transmembrane domain (TrD) (Gururani et al.,

2012). eLRRs are known to play an important role for certain defense proteins such as, polygalacturonase inhibiting proteins (PGIPs) (Jones et al., 1997) even though they are not directly involved in pathogen recognition and activation of defense genes (Jones, 2001 and Stotz et al., 2000). The *HcrVf2* gene conferring scab resistance in apple is belonging to the eLRR class (Belfanti et al., 2004). The fourth class of resistance genes consists of an extracellular LRR domain, a transmembrane domain (TrD) and an intracellular serine-threonine kinase (KIN) domain (Song et al., 1995). The fifth class of resistance genes contain the putative extracellular LRRs, along with a PEST (Pro-Glu-Ser-Thr) domain for protein degradation, and short proteins motifs (ECS) that might target the protein for receptor mediated endocytosis (Thomma et al., 2011). The *Arabidopsis* RPW8 protein is an example of the sixth major class of resistance genes which contains a membrane protein domain (TrD), fused to a putative coiled coil domain (CC) (Wang et al., 2009), whereas, the seventh major class of resistance genes includes the *Arabidopsis RRS1*-R gene conferring resistance to the bacterial phytopathogen *Ralstonia solanacearum*, and it is a new member of the TIR-NBS-LRR R protein class. *RRS1*-R has a C-terminal extension with a putative nuclear localization signal (NLS) and a WRKY domain (Deslandes, et al., 2002 and 2003). The WRKY domain is a 60 amino acid region that is defined by the conserved amino acid sequence WRKYGQK at its N terminal end, together with a novel zinc-finger-like motif. The last major class of resistance genes includes the enzymatic R-genes which contain neither LRR nor NBS groups. For example the maize *Hm1* gene which provides protection against southern corn leaf blight caused by the fungal pathogen *Cochliobolus carbonum* (Johal and Briggs, 1999). Unlike other resistance genes, Hm1 encodes the enzyme HC toxin reductase, which detoxifies a specific cyclic tetra-peptide toxin produced by the fungus (HC toxin) that is essential for pathogenicity. Therefore, cereal resistance genes like *Hm1* can be seen to encode a range of different proteins that in some cases have obviously very different functions.

3.2. Aims of Study

Apple scab caused by *Venturia inaequalis* can markedly affect yields and apple fruit quality therefore resistance sources have received the most attention in the world by the apple breeding group. Many scab resistance genes have been identified so far but most of the available resistant cultivars are derived from a single resistance source i.e *Malus floribunda* 821 carrying the *Rvi6* gene. The discovery of new races of *Venturia* able to overcome the *Rvi6* resistance induced a marked change in the breeding strategy. The introduction of new scab resistance genes in breeding programs and particularly the pyramidization of different resistances in a single genotype would have a positive impact on resistance durability.

The selection of plants with multiple resistances against the same pathogen is very difficult and time consuming with traditional genetic methods but marker-assisted selection (MAS) could be a useful tool to accelerate apple breeding programs. However, in the MAS process, tightly-linked markers located on both sides of target genes are needed for accurate selection, reducing the risk of selecting unfavorable individuals resulting from recombination between the markers and the target gene.

The Apple breeding program of the Bologna University started more than 30 years ago and a lot of work have been done to introgress and characterize the genetics of different scab resistance genes. In the last years special attention has been paid to pyramidize different scab resistances and to develop tools to be used for marker assisted selection.

The main aims of this study are:

- to characterize two different sources of scab resistance used in the breeding program of the Bologna University: *Rvi13* (*Vd*) and *Rvi5* (*Vm*).
- to develop and map various SSRs and SNPs markers on the top of LG10 in order to better define the QTL peak region which is responsible for *Rvi13* resistance; the marker development will make use of the availability of the recently published apple genome sequence of 'Golden Delicious'.
- to develop and map various SNPs markers on the bottom of LG17 in order to make a

fine-mapping of the region responsible for monogenic *Rvi5* resistance; the marker development will make use of the sequence of two Murray BAC clones that have been identified in a joint "*Rvi5* gene cloning" project between Bologna University and FEM, San Michele all'Adige, Trento.

-

The availability of suitable closely linked markers will open the way for the development of high-throughput MAS to be used in the early selection in future breeding programs. Finally, the availability of the apple genome sequence and the fine-mapping around resistance genes is essential to identify possible candidate genes that could be further characterized by genetic transformation.

3.3. Materials and methods

3.3.1. Fine mapping of scab resistance gene *Rvi13 (Vd)*

3.3.1.1. Plant material and DNA extraction

An F1 population of 381 seedlings derived from the cross between 'Durello di Forli' and 'Fiesta' was analyzed. Genomic DNA extraction was performed following the standard CTAB protocol (Doyle and Doyle, 1990). Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) was used to quantify the gDNA and the working solution was prepared at 50ng/ul. The inoculation test and scab infection severity evaluation was performed by Tartarini, et al. (2004), scab scale data was available to use in this study.

3.3.1.2 Initial re-mapping of the *Rvi13* region

In order to verify the trueness-to-type after DNA extraction, two markers RAPD OPAF07/880 and SSR CH02b07 were repeated. For the RAPD, PCR amplification was performed using the Bio-Rad PTC-100® thermal cycler with the oligo OPAF07/880 (5'- GGAAAGCGTC-3') on the 'Durello di Forli' and 'Fiesta' (D X F) progeny plants along with the parents. The following PCR condition was used to amplify: 0.5 uM primer, 2 mM $MgCl_2$, 125 uM dNTPs, 0.6 Units DNA Polymerase (Fisher Molecular Biology, Hampton, NH, USA), 1 X reaction buffer, and 50 ng of genomic DNA in a final volume of 20 ul. The thermal cycler condition used for the RAPD amplification was same as described by Tartarini (1996). The PCR products were separated on a 2% agarose in TAE gel electrophoresis, stained with ethidium bromide, and photographed.

The CH02b07 SSR marker was analyzed together with Hi02d04, another SSR at the top of the LG10 (Silfverberg-Dilworth et al., 2006). SSR data were retrieved from the HiDRAS (High-quality Disease Resistant Apples for a Sustainable Agriculture) website (http://users.unimi.it/hidras/). Both primer pairs were purchased from Biofab research (Pomezia) and the SSR was screened on the entire mapping population. The SSR reaction was carried out in a volume of 17.5 ul containing 50 ng of genomic DNA, 0.1 uM of each

primer, 2 mM $MgCl_2$, 100 uM dNTPs, 0.5 Units DNA Polymerase (Fisher Molecular Biology, Hampton, NH, USA) and 1 X reaction buffer. PCR amplifications were performed on the Applied Biosystems 2720 thermal cycler under the following conditions: an initial denaturation step of 5 min at 95 ℃ was followed by 32 cycles of annealing (60 ℃ for 45s), extension (72 ℃ for 1min) and denaturation (94 ℃ for 30s) with a final extension of 10 min at 72 ℃. Polyacrylamide gel electrophoresis (PAGE) visualized by silver staining was used for separating SSR amplification products.

The genetic linkage mapping was performed with the software JoinMap 3.0 (Van Ooijen and Voorrips, 2001) using the Kosambi mapping function. The LOD value chosen for grouping the markers was equal to 7. The final visual representation of the map was generated with MapChart (Voorrips, 2001).

3.3.1.3. Mapping of ARGH markers

The two Sequence Characterized Amplified Region (SCAR) markers ARGH8 and the ARGH25, mapping on the top of LG 10 (Baldi et al., 2004) were searched for Single Nucleotide Polymorphisms (SNPs) allowing their mapping. With the aim to transform SCAR marker in a Cleaved Amplified Polymorphic Sequence (CAPS) marker, PCR amplification of both markers was performed on Durello di Forli and Fiesta using 50 ng of genomic DNA as described in Baldi et al. (2004). The resulting amplicons obtained were digested with *TaqI, Tru1I* and *RsaI* following manufacturer's instructions (Fisher Molecular Biology, Hampton, NH, USA) and subsequently run on a polyacrylamide gel in order to identify possible polymorphism between the parents. Once the polymorphism was confirmed, all the progeny seedlings were analyzed. Meanwhile the sequencing of the PCR products from the ARGH8 and ARGH25 markers was also conducted so as to know exactly nucleotide sequences of the two NBS-LRR resistance gene homologues.

3.3.1.4. Development and mapping of new SSR markers from GD genome sequence and QTL mapping

A sequence-based approach was used to develop markers in the *Rvi13* region by exploiting the draft sequence of apple genome (Velasco et al., 2010). The strategy was first to establish

the connections between already mapped markers and their corresponding DNA sequences in the apple genome so that the region could be determined physically, and then to explore the DNA sequences in the region for developing SSR markers closer to the resistance gene. In practice, nucleotide sequences of the SSR markers Hi02d04 and CH02b07 were obtained from the website HiDRAS (http://www.hidras.unimi.it/) and BLAST searched against the apple genome sequence at GDR (Genome Database for Rosaseae) (http://www.rosaceae.org/species/apple). Both contigs were confirmed to be of chromosome 10 origin. Long contigs not overlapping with other contigs at the beginning of chromosome 10 were used as first choice for SSR markers development. DNA sequences of various representative contigs on the chromosome 10 were downloaded and analyzed for the presence of SSRs. Tandem Repeats Finder software (http://tandem.bu.edu/trf/trf.html) was employed to search for the repeated motif under the default parameter settings. Primer pairs bracketing the SSR sequences were designed with the program Primer3 publicly available at http://frodo.wi.mit.edu/primer3/. The ideal annealing temperature (Tm) of the primers was set at 60 ℃. A total of 24 SSR primer pairs were designed and analyzed. SSR amplification and amplicons separation were performed as described above for CH02b07 and Hi0204.

The validation of the newly developed SSR markers was conducted as follows: the new marker was firstly tested on 20 randomly selected progeny plants and the two parents, as long as the marker showed strong consistence with the dataset constructed by the confirmed *Rvi13*-linked markers (in detail, OPAF07/880, Hi02d04 and CH02b07), the entire mapping population was applied to screen with this marker. The linkage mapping analysis was performed in the same way as described above.

Interval mapping analysis using MapQTL 4.0 (Van Ooijen et al., 2002) was performed for QTL identification. Two-year pooled data of apple scab resistance scale was used. A LOD threshold of 7 was used for declaring the putative QTL as significant. The percentage of the variation explained by the significant QTL was estimated by MapQTL.

3.3.1.5. Transformation of the OPAF07/880 RAPD marker into a easy-to-use marker

Development and mapping of SCAR marker OPAF07. The corresponding polymorphic OPAF07/880 RAPD band derived from Durello di Forli was excised from the gel and purified according to the GenElute™ Gel Extraction Kit protocol (Sigma-Aldrich). The purified fragment was ligated into the pGEM®-T Easy Vector (Promega) and heat-shock transformed into chemically competent *E.coli* cell. Positive colonies recognized by the white color were picked up with sterilized tooth sticks from plates, suspended in 30ul ddH_2O and 1ul was used as PCR template. The size of the insert was verified by colony-PCR with the pUC/M13 forward and reverse primers and by comparison with the GeneRuler 100 bp DNA Ladder (Fisher Molecular Biology, Hampton, NH, USA). Reaction conditions for the colony-PCR (20ul) were as follows: 1 X reaction buffer, 5 mM $MgCl_2$, 200 uM dNTPs, 0.2 uM of each primer, 1 Units DNA polymerase (Fisher Molecular Biology, Hampton, NH, USA) and bacterial DNA. The thermal cycling program contained 95 ℃ for 5 min followed by 35 cycles of 1 min at 95 ℃, 1 min 30s at 52 ℃, and 1min 30s at 72 ℃, ending with 15min extension at 72 ℃. Amplification products were separated on 1% agarose gel and visualized on an Image Station 440CF (Kodak, Rochester, NY, USA) by ethidium bromide staining. Plasmid DNA extraction from two positive colonies was conducted according to the protocol of plasmid purification kit (MACHEREY-NAGEL, Germany). Specific primer pairs were designed from the Durello di Forli OPAF07/800 RAPD sequence using the Primer3 v0.4.0 (http://frodo.wi.mit.edu/primer3/). The original OPAF07 primer sequence was extended from 10 bp to 18 bp for forward primer (5'- **GGAAAGCGTC**CAGTGGTG-3') and transformed to 20 bp for reverse primer (5'-CGTCGATGCCCCATATAACA-3'). PCR amplification with SCAR marker was performed as described by Liebhard et al. (2002) for SSR in 20 ul.

PCR-based screening of the BAC library. A BAC library from the cultivar Florina (Vinatzer et al., 1998) was already available at the Department of Fruit Tree and Woody Plant Sciences (University of Bologna). This library consists of 36,864 BAC clones with an average insert size of 120 kb, representing approximately 5 x apple haploid genome equivalents. PCR-based screening on the entire the Florina BAC library with specific primer pairs of AF07 SCAR

marker was carried out under a bi-dimensional pooling approach. Positive BAC clones were picked from the library, singularized and tested by colony-PCR with the same primers used for the screening. The amplicons were visualized on an Image Station 440CF (Kodak, Rochester, NY, USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining. Positive clones harboring the AF07 fragments were carefully picked out from the corresponding wells from Florina BAC library and cultivated on the LB solid medium with ampicilin ready for the plasmid extraction. The conventional alkaline extraction procedure (Birnboim and Doly, 1979) was used to extract and purify plasmids.

Plasmid sequencing. In attempt to obtain more upstream and downstream nucleotide sequence information of the AF07 fragment on the top of the LG10, one primer pair designed at the two extremes region inside of AF07 fragment with the opposite direction was used to sequencing the positive plasmids harboring the AF07 fragment. Along with the T7 and SP6 BAC ends primer pairs, the purified plasmid DNA products were subjected to sequencing by Bio-Fab Research srl (Pomezia, Italy).

3.3.1.6. Candidate gene-based SNP discovery

As long as the QTL region was well defined, the physical position of loci responsible for the apple scab resistance on the top of LG 10 was traced back to the apple genome assembly contig sequence through the two SSR markers flanking the QTL. Candidate genes were then selected in this region from apple genome sequence on the basis of their position. The nucleotide sequence of candidate genes was downloaded from Gbrowse at GDR website with the aim to identify specific SNPs for Durello di Forli via the sequence comparison of the candidate gene amplicons with Fiesta. Specific PCR primer pairs were designed containing candidate genes using Primer3 v0.4.0 (http://frodo.wi.mit.edu/primer3/). The design criteria were: a final product size ranging from 800 to 1200 bp; a primer size ranging from 18 to 27 bp (optimum: 20 bp); a primer melting temperature ranging from 57°C to 63°C (optimum: 60°C); and a percentage of G and C bases ranging from 40% to 55%. The self-complementarity and the 3' self-complementarity were set to a maximum of 4 and 1, respectively, in order to prevent the formation of primer dimers during PCR. The designed primers were submitted to a BLASTN analysis against the apple genome assembly to guard against annealing to multiple locations.

PCR amplifications for the candidate genes were performed in a 20-ul volume containing 50 ng of genomic DNA, 0.1 uM gene-specific primers, 1.5 mM $MgCl_2$, 200 uM dNTPs, 1 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1 x reaction buffer. The reaction included an initial 10 min denaturation at 95 ℃, followed by 30 cycles of 45 s at the optimized annealing temperature, 2 min at 72 ℃, and 30 s at 95℃, with a final extension of 7 min at 72℃. Amplicons were run on a 2% TAE agarose gel to check the gene amplification quality. The optimization of PCR condition was needed once the target candidate gene amplification was not enough specific. Prior to the sequencing of the candidate gene fragments which was conducted by Bio-Fab Research srl (Pomezia, Italy) with both forward and reverse specific primer pairs, the PCR products of candidate genes were purified using Wizard® SV Gel and PCR Clean-Up System Kit (Promega) following the instruction provided. The final sequence assembly was manually performed with the SeqMan software (Lasergene v8.0). SNPs detection was realized directly by the Chromas version 2.22 in the form of double peaks in two different colors.

3.3.1.7. SNP genotyping by Temperature Switch PCR (TSP) approach

TSP primer design. TSP assay primers were designed based on the strategy described by Hayden et al. (2009) and Tabone et al. (2009). For each SNP marker, locus specific (LS) primers amplifying the region surrounding the SNP of interest as well as a allele-specific (AS) primer fully complementary to the sequence of both parents, but matching the nucleotide specific of Durello di Forli at the SNP locus, were designed using Primer 3 Software. LS primers, in our case identical to the candidate gene primers, have an optimum melting temperature (Tm) of 62℃ (range of 60–65℃) and to amplify a PCR product greater than 400 bp. The AS primer was designed to have a core region with an optimum Tm of 45 ℃ (range of 43–48 ℃) and a non-complementary 2-3 bp nucleotide sequence added to the 5' end region that increased the overall optimum primer Tm to 53℃ (range of 52–55 ℃). The forward or reverse AS primer was positioned in at least 60 bp distance from the corresponding forward/reverse LS primer to ensure a clear distinction between the larger LS and the smaller AS PCR product.

Optimization of the TSP genotyping protocol. The TSP method described by Hayden et al. 2009 and Tabone et al. 2009 served as basis for the development of the TSP assay protocol for SNP genotyping. A number of technical details were modified, including standard PCR reagents, the 50 ng diluted DNA instead of 20ng desiccated DNA, increase or decrease of TSP component concentration and so on. TSP assays was performed using 1.5 U AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA, USA), 100 ng/ul bovine serum albumin Fraction V (Thermo Fisher Scientific), 1 x reaction buffer, 1.5 mM $MgCl_2$, 200 uM dNTPs, 0.1 uM each locus-specific primer, 0.5 uM allele-specific primer and 50 ng genomic DNA in a total reaction volume of 10 ul.

Thermal cycling was performed on Applied Biosystems 2720 thermal cycler consist of an initial denaturation step (95℃ for 10 min); 20 cycles of 95 ℃ for 30 s, 60 ℃ for 30 s, 72 ℃ for 1min 30 s in order to enrich the LS product at a relatively high annealing temperature; 5 cycles of 95 ℃ for 10 s and 45 ℃ for 30 s to enable a possible incorporation of the alleles-specific primer into the enriched LS PCR product at a low annealing temperature; 10 cycles of 95 ℃ for 10 s, 53 ℃ for 30 s and 72 ℃ for 5 s to facilitate a competitive amplification of LS and AS PCR products. The TSP thermal cycling program described above was deployed merely as the first preliminary trail for each SNP to gain more information for the further improvement of TSP protocol. In other words, for each individual SNP, it was necessary to include one essential optimization step for TSP themal cycling condition, which could guarantee an accurate SNP genotyping without mistakes. Total 10 ul of the PCR products were analyzed on 2% agarose gels and stained with ethidium bromide. Once the best TSP amplification system was set up, screening on the entire recombinants seedlings of this SNP marker was performed. Whether the tested seedling plants carry reference or SNP allele at the analyzed loci could then be determined by the presence of either the smaller or the larger PCR product, respectively.

3.3.2. Fine mapping of scab resistance gene *Rvi5 (Vm)*

3.3.2.1. Plant material and genomic DNA extraction

Seventeen F1 recombinant plants from two large mapping populations of more than 1200 seedlings ('Murray' X 'Golden Delicious' and 'Murray' X 'Galaxy') were identified in a previous study (Cova, 2008). Genomic DNA extraction from freeze-dried young leaves of recombinant seedlings and three parents was performed following the standard CTAB protocol (Doyle and Doyle, 1990). Nanodrop^{TM} ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) was used to quantify the gDNA. All gDNA were diluted to 50ng/ul for working solution.

3.3.2.2. Re-mapping of the available SSR markers

Four SSR markers were already available, Hi07h02 reported by Patocchi et al. (2005), and the remaining three SSR markers G-SSR, B-SSR and C13 were all unpublished (Cova, 2008). The primer pairs of all 4 SSR markers were purchased and tested on 16 recombinant seedlings in order to confirm the available dataset and the orientation of these 4 SSRs in the mapping population. The SSR amplification was performed as follows: 50 ng of genomic DNA, 0.1 uM of each primer, 2 mM $MgCl_2$, 100 uM dNTPs, 0.5 Units DNA Polymerase (Fisher Molecular Biology, Hampton, NH, USA) and 1 X reaction buffer in a total of 17.5 ul volume. PCR amplifications were performed on the Applied Biosystems 2720 thermal cycler under the following conditions: an initial denaturation step of 5 min at 95 ℃, followed by 32 cycles of annealing (60 ℃ for 45s), extension (72 ℃ for 1min) and denaturation (94 ℃ for 30s) with a final extension of 10 min at 72 ℃. Polyacrylamide gel electrophoresis (PAGE) visualized by silver staining was used for scoring SSR amplicons.

3.3.2.3. Candidate gene-based SNP discovery

The available BAC clone sequences were subjected to the open reading frame (ORF) prediction performed using the FGENESH 2.6 software (Salamov and Solovyev, 2000). Deduced protein sequences of resulting ORFs were then compared by BLASTX to all protein sequences deposited at NCBI database. Candidate genes were then selected mainly based on their relativity to the pathogen disease resistance pathway. With the aim to identify

specific SNPs for Murray via the sequence comparison of the candidate gene amplicons with Golden Delicious and Galaxy, 5 specific PCR primer pairs were designed on the gene sequence using Primer3 v0.4.0 (http://frodo.wi.mit.edu/primer3/). The design criteria, PCR reaction and conditions, amplicon purification for candidate genes are already described in the *Rvi13* section.

3.3.2.4. SNP genotyping by Temperature Switch PCR (TSP) approach

TSP assay was already described in the *Rvi13* section. For each SNP marker, locus specific (LS) primers amplifying the region surrounding the SNP of interest as well as a allele-specific (AS) primer fully complementary to the sequence of Golden Delicious and Galaxi, but matched at the 3' end the nucleotide from Murray, were designed using Primer 3 Software.

3.4. Results and discussion

3.4.1. Fine mapping of scab resistance gene *Rvi13 (Vd)*

3.4.1.1. Initial re-mapping of available markers flanking the *Rvi13*

The results of the screening of the RAPD marker OPAF07/880 and two SSR markers Ch02b07 and Hi02d04 with the entire 381 progeny seedlings descended from Durello di Forli X Fiesta, confirmed that all the molecular markers were strongly associated with *Rvi13*. Four main bands with different sizes were randomly amplified with the RAPD primer OPAF07, but the 880 bp band resulted tightly linked to the *Rvi13* gene (Figure 3.3 A).

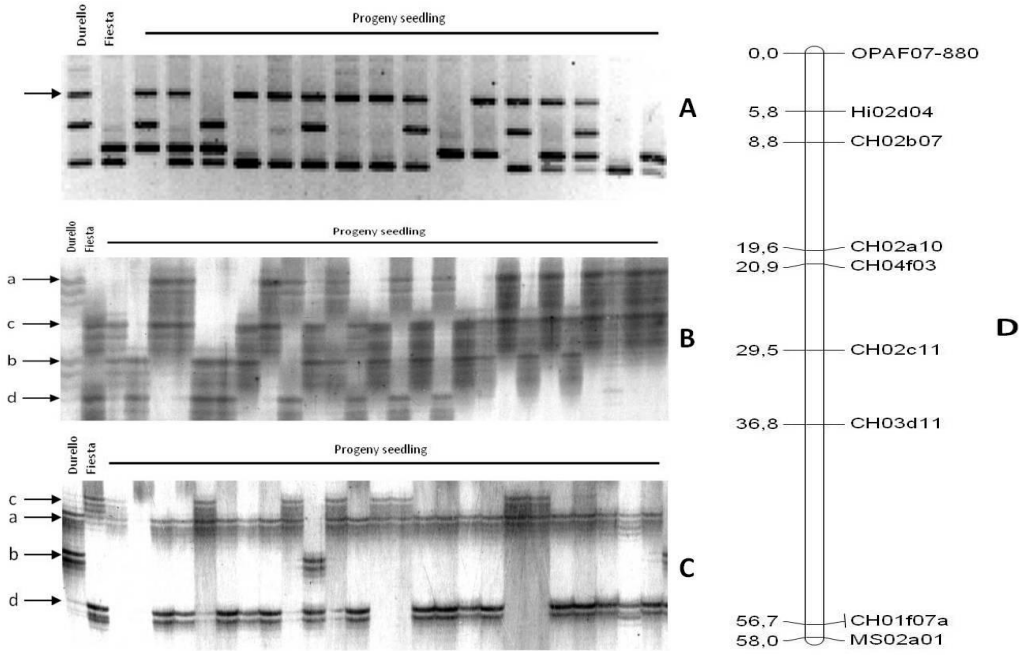


Figure 3.3. A. Agarose gel profile of OPAF07/880 RAPD marker; the band associated with the resistance gene is indicated with the arrow, the dominant marker is segregating; B and C. Polyacrylamide gel profile of CH02b07 and Hi02d04 SSR markers; both markers showed four alleles segregating in a "ab X cd" pattern; D. Genetic linkage map based on data from 381 progeny seedlings of the cross Durello di Forli x Fiesta using the polymorphic markers available.

The two SSR markers' polymorphic amplicons between Durello di Forli and Fiesta were

segregating in the F1 mapping population (Figure 3.3 B and C). Linkage analysis of the new and the already available SSR data demonstrated that the OPAF07/880 was mapped at the very top of the LG10, while Hi02d04 and CH02b07 were mapped at 5.8 cM and 8.8 cM relative to the OPAF07-880 respectively (Figure 3.3 D).

3.4.1.2. Mapping of ARGH markers

No differences were identified between the Durello di Forli and Fiesta in both amplicons of ARGH8 and ARGH25 by direct amplification using the primers from Baldi et al. (2004). So the ARGH8 and ARGH25 sequences were blasted on the GDR website in order to search for the corresponding contigs of the apple genome. The BLASTN results revealed that ARGH 25 sequence was matching with a contig of the chromosome 10. The best hit for ARGH 8 was with a contig from chromosome 5, which was known as homeologous to LG10 (Table 3.2).

Table 3.2. BLANSTN results of the ARGH8 and ARGH25

ARGHs Name		Primer (5'-3')*	Results from BLASTN
ARGH8	Fwd	CTAAAATAAACCGTACTCTAATTG	MDC002167.405
	Rvs	GTCTTTCTTAGGTTGTTCTTTC	chr5:16179878..16190195
ARGH25	Fwd	CAAACATCATCGTAATTTTGACG	MDC007061.242
	Rvs	CATACTCTTCATGAGGATAATTC	chr10:26905408..26910174

***Baldi et al., (2004)**

The SNP (C/T=Y) at position 128 bp of ARGH25 sequence from Durello di Forli fell exactly inside *TaqI* restriction site. Therefore, one allele (TCGA) was digested by the enzyme while the second allele (TTGA) was not cut (Figure 3.4). Digestion of the ARGH25 amplicon with *TaqI* resulted in a polymorphic CAPS marker in which two fragments of size 127 bp and 277 bp were obtained in Durello di Forli and no evidence of *TaqI* restriction sites in Fiesta. The mapping of ARGH25 polymorphism confirmed that this sequence was derived from LG10 (Figure 3.5). Since no recombinant was detected between CH02b07 and ARGH25 marker, the ARGH25 was cosegregating with CH02b07 SSR marker.

ARGH25 primer Fwd

CAAACATCATCGTAATTTTGACGGTGCAAGTTTCCTTGAAAACGTGAGGGA
AAAGAAACTGGTTGATTTGCAAGAACAATTTCTTTCTGATATCTTGAAACC
GGCCTACATAGGGGTAAGTTGTA**TYGA**TGAAGGGACCAAGGAGATAACAAA

TaqI

GGTTAGGCAACGGAAGGGTGCTTGTCGTAATTGATGATTTAGACAGCGTTG
AACAGATAGATGCTTTGGCTATAAAACCTGACTCCTTTGGTCCAGGGAGCA
GAATCATTGTAACTTCAAGAAACAAACATTTGCTGAAGATACTTAATGTGG
ATAAGATATGTCCTCTGCCATCAATGAGTGAAAGAGAAGCTCTTGAGCTAC
TTAGTTTGCATGCTTTTAGAAAGAATTATCCTCATGAAGAGTATG

ARGH25 primer Rvs

Figure 3.4. ARGH25 sequence from Durello di Forli, the arrow indicates one SNP inside the *TaqI* restriction enzymes site; the ARGH25 primer pair is indicated with boxes.

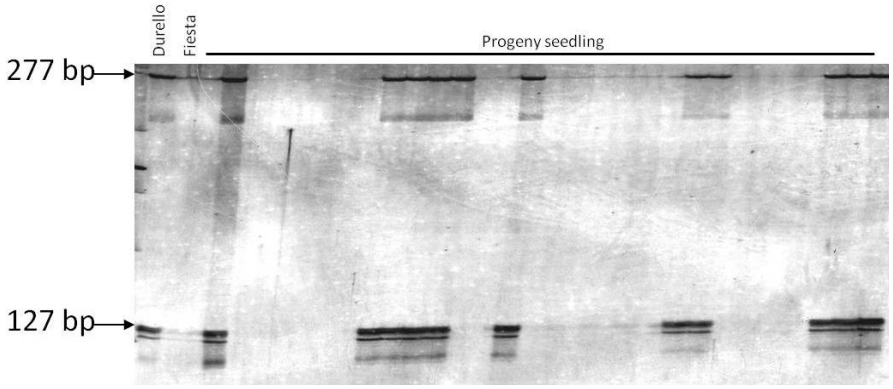


Figure 3.5. Polyacrylamide gel profile of the ARGH25 CAPS markers. The band at about 400 bp is the undigested amplification products from Fiesta while the two fragments of 127 and 277 bp derived from the *TaqI* digestion of the ARGH25 amplicon of Durello di Forli.

3.4.1.3. New SSR markers development from the GD genome

Three rounds of SSR markers design and evaluation were conducted in order to achieve a relatively high resolution map covering the top of the LG10, and one step further towards the fine mapping of *Rvi13* scab resistance locus (Table 3.3).

In the first round of SSR markers development, eight primer pairs were analyzed (Table 3.3). Except for DCA10SSR4, which was monomorphic and DCA10SSR6, which was multilocus, the remaining six SSRs all exhibited polymorphism between Durello di Forli and Fiesta. However, after screening the polymorphic SSRs on the entire progeny, the segregating bands of DCA10SSR1 and DCA10SSR3 were consistent with Hi02d04 and CH02b07 (Figure 3.6), the other four SSRs resulted not mapping on the LG10.

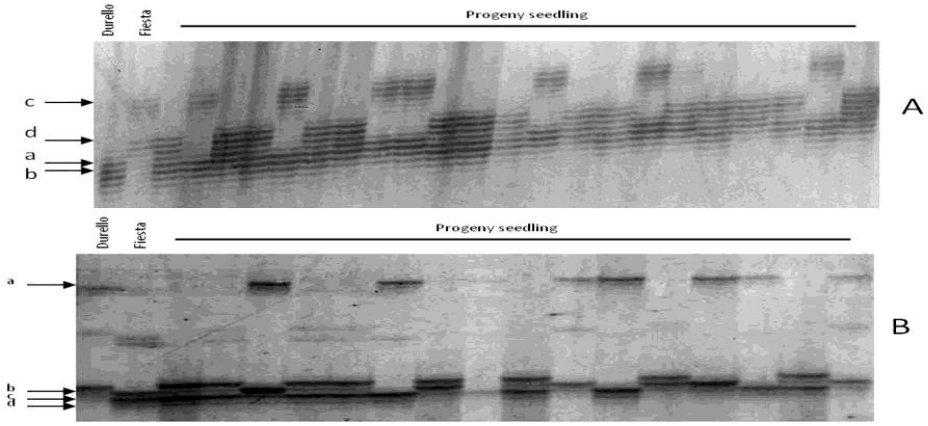


Figure 3.6. A and B. Polyacrylamide gel profile of CH02b07 and Hi02d04 SSR markers; both markers shows four alleles segregating in a "ab X cd" pattern.

In the second round of SSR markers development, other eight SSR markers, named from DCA10SSR14 to DCA10SSR21, were analyzed. However, the PCR amplification and polyacrylamide gel revealed that only DCA10SSR17 and DCA10SSR21 were polymorphic and really mapped on the LG10 (Figure 3.7).

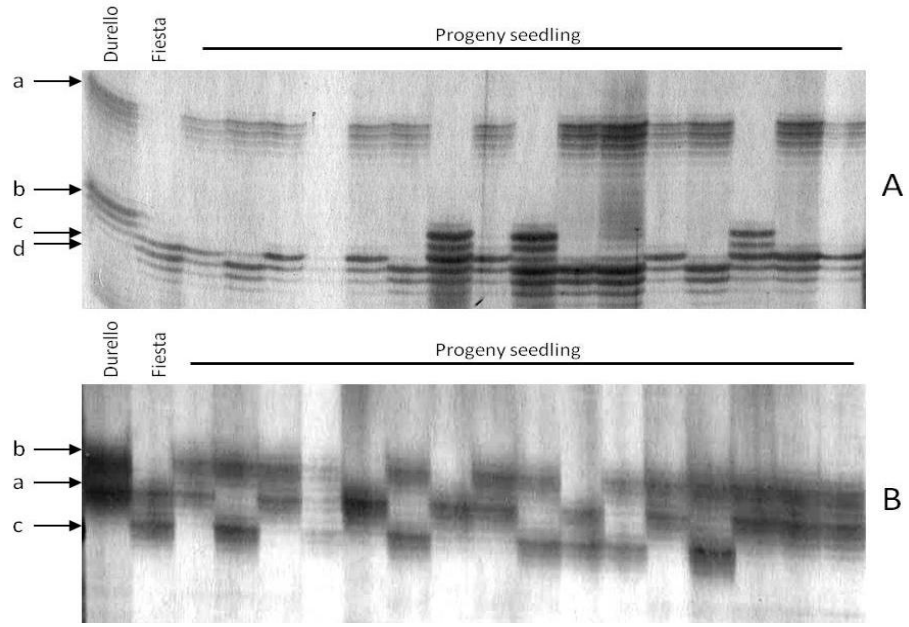


Figure 3.7. A. Polyacrylamide gel profile of DCA10SSR17 marker, four alleles segregating in a "ab X cd" pattern; B. Polyacrylamide gel profile of DCA10SSR21 marker, three allele segregating in "ab x ac" patter.

DCA10SSR14 marker was multilocus but no locus was mapped on the LG10. For DCA10SSR15 and 20, these two SSR markers did not produce any segregating bands on the polyacrylamide gel, suggesting they were monomorphic. Although the DCA10SSR16, 18 and 19 were polymorphic SSR markers and designed on the contigs downloaded from the apple genome assembly website, unexpectedly they were not mapped on the LG10 because of the inconsistent segregating bands with the LG10 SSR markers.

Table 3.3. Detailed information of SSR markers analyzed in this study

Round No.	SSR name		Primers sequence (5'-3')	Apple source contigs	Apple genome coordinates	Polymorphsim	Multilous	Segregation on LG 10
	CH02b07	Fwd	CCAGACAAGTCATCACAACACTC	MDC008148.490	ch10:5600392..5606934	Yes	No	Yes
		Rvs	ATGTCGATGTCGCTCTGTTG					
	Hi02d04	Fwd	TGCTGAGTTGGCTAGAAGAGC	MDC015312.249	ch10:2710638..2721470	Yes	No	Yes
		Rvs	GTTTAAGTTCGCCAACATCGTCTC					
Round 1	DCA10SSR1	Fwd	GCGCATTATTCATAACATTACTAGC	MDC015291.258	ch10:240028..292677	Yes	No	Yes
		Rvs	GGCTTCTTGTTGCTGAGAGG					
	DCA10SSR3	Fwd	AACCCGGCCCAATAATAAAG	MDC013845.347	ch10:890245..905668	Yes	No	Yes
		Rvs	CGGAAACCCTGTATGTGCTT					
	DCA10SSR4	Fwd	CCTGAGGGAATGGATCATGT	MDC013762.165	chr10:1534651..1555470	No	Yes	No
		Rvs	AAATTGCTCTCACGCGAACT					
	DCA10SSR6	Fwd	GGAATTTCACTTCCAGGGATT	MDC009854.509	chr10:3111004..3129006	Yes	Yes	No
		Rvs	GGAATTGCATTGCATTGGTT					
	DCA10SSR7	Fwd	TGTTTACCGAGGGAGAGGTG	MDC016990.256	chr10:3215437..3235755	Yes	No	No
		Rvs	TGCCGATGAATCGTTTACAA					
	DCA10SSR9	Fwd	TTGAAATCCCAAGGAAGCAC	MDC011274.217	chr10:4283926..4294698	Yes	No	No
		Rvs	CTCGAAAGCAAAGCAGAAGC					
	DCA10SSR12	Fwd	ATCTTAGGGATTGGGCAACA	MDC006904.636	chr10:4522907..4546658	Yes	No	No
		Rvs	AGAAAGAAAAGTCGAAGGGAGA					
	DCA10SSR13	Fwd	CCCAGACCCAATATCCTTCA	MDC019496.75	chr10:4907193..4920833	Yes	No	No
		Rvs	CACGTGGACTTTCGGATCTT					
Round 2	DCA10SSR14	Fwd	TCCTGAATGAGTCCTGAGTCC	MDC000285.563	chr10:1..10768	No	Yes	No
		Rvs	AAAGGAAGAGGGAGGAGCAG					
	DCA10SSR15	Fwd	CGAGTAGTGTTGTCGATGTGG	MDC022095.180	chr10:121840..133356	No	No	No
		Rvs	TTAGGACGTCATGCGTATTGTT					
	DCA10SSR16	Fwd	TTCCCATTAATTATTCCATCAGTG	MDC022909.139	chr10:166457..175337	Yes	No	No
		Rvs	CCTCTTCCACAACCACCATC					
	DCA10SSR17	Fwd	GGATGCTCGGCTGTACAAGT	MDC029126.38	ch10:1076249..1096702	Yes	No	Yes
		Rvs	GGTGCATACACCAACTTACCAA					

(Continued)

Round No.	SSR name		Primers sequence (5'-3')	Apple source contigs	Apple genome coordinates	Polymorphsim	Multilous	Segregation on LG 10
Round 2	DCA10SSR18	Fwd	TTCGTTTAACTCCAAGCGAAA	MDC018507.307	chr10:1213913..1241440	Yes	No	No
		Rvs	TCTCCCCTCTCGAGTTTCAA					
	DCA10SSR19	Fwd	AAACTTCCCACTCCTCCCATA	MDC011437.437	chr10:1426941..1504935	Yes	No	No
		Rvs	TCGTTGAAAGAAAATTGCACA					
	DCA10SSR20	Fwd	CGAATGTCACAGCATCCATC	MDC018546.292	chr10:1581124..1613226	No	No	No
		Rvs	AGAGCCTCTCCTGGGGATT					
	DCA10SSR21	Fwd	TCTCCACCTCCCTATTAACACAA	MDC042829.5	ch10:2008671..2027387	Yes	No	Yes
		Rvs	CTCCATTTACACACAGGGAGAA					
Round 3	DCA10SSR2	Fwd	TTGCTTTGGAATGCTTGAGA	MDC005799.536	ch10:718885..729454	Yes	No	Yes
		Rvs	CTTCTGGTCCCATCCTTTGA					
	DCA10SSR22	Fwd	AATGGTTTTGTGGAGCCAAC	MDC022095.200	chr10:111770..128249	Yes	No	Yes
		Rvs	TTGGCTCTGAAGGAAATGCT					
	DCA10SSR26	Fwd	GTTAGCATAGCCGGTTGGTC	MDC013126.201	chr10:32650..42448	Yes	No	Yes
		Rvs	CTCGACGACACATTGTACGC					
	DCA10SSR27	Fwd	GTCGAGCGAAAACACACAGA	MDC013126.201	chr10:32650..42448	Yes	Yes	Yes
		Rvs	GGGTTGGGGTGTTACATTCA					
	DCA10SSR28	Fwd	GCACGGTGTAGAGTACCTTGG	MDC022357.121	chr15:25874579..25909257	Yes	No	No
		Rvs	CAGGTCGCAAGACACGATTA					
	DCA10SSR29	Fwd	CATAGCTATTGCGCGTTTCA	MDC006516.55	ch10:1936488..1949522	Yes	No	Yes
		Rvs	CGTCGATTCTGTTGCTTTCA					
	DCA10SSR30	Fwd	TGCATAAAACCCCGATATTTG	MDC008416.202	ch10:1156516..1187334	Yes	No	Yes
		Rvs	ACAATTTTGACGTCCGCAAG					
	DCA10SSR31	Fwd	TTCGTTCTTGCCAACAGAAT	MDC013187.318	chr6:11837941..11866766	Yes	No	No
		Rvs	GCCTTTCTTCACCACCTCAG					

The difficulties encountered in developing new SSR markers at the top of LG10, suggested to investigate the synteny between the apple and peach genome. Since, apple (sub-family Maloideae) and peach (sub-family Prunoideae) are genera of the Rosaceae family. Then, peach is a model plant for the family Rosaceae due to its small genome size of ~ 230 Mb and the draft genome generated from the '"Lovell" was released by the international Peach Genome Initiative at the Genome Database for Rosaceae (http://www.rosaceae.org/peach/genome). Compared with apple, peach has a relative small diploid genome and the current version of peach genome (Verde et al., in press) has more complete coverage than the apple genome (Velasco et al., 2010). A BLASTN search against the peach genome was performed using the six confirmed SSR markers from apple LG10 (Hi02d04, CH02b07, DCA10SSR1, DCA10SSR3, DCA10SSR17 and DCA10SSR 21) and this analysis revealed that except for the DCA10SSR17, the sequences of the contigs for other five SSR markers development were homologous to the peach genome scaffold 8. Therefore, to develop new SSR markers, apple contigs from LG10 supposed to exploit SSRs for primer design, were firstly blasted against the peach genome to confirm the high alignment score on the scaffold 8, and then the ordinary primer pair design procedure was performed. Actually, nine primer pairs were designed in the third round SSR markers development, named from DCA10SSR22 to DCA10SSR31. In fact, DCA10SSR28 and DCA10SSR31 were directly designed based on the beginning sequences of peach scaffold 8, in detail, DCA10SSR28 was at 650,000-750,000 bp from the top of the scaffold 8 and DCA10SSR31 was at 1810,000-1840,000 bp. Moreover, the DCA10SSR2 which was designed in the first round but without performing the analysis was also analyzed in this round because of a putative scab resistance related gene found on the contig containing DCA10SSR2 with high similarity to the peach scaffold 8. With the exception of the two markers designed on the peach genome, the other six SSR markers identified in this round were all proved to be useful LG10 SSR markers with polymorphism between Durello di Forli and Fiesta parents by the PCR amplification and electrophoresis on the polyacrylamide gel (Figure 3.8). Because DCA10SSR26 and 27 were designed on the same contig MDC013126.201 and segregation type of DCA10SSR26 ("ab x cd") was more informative than the "a0 X 00" of DCA10SSR27, the DCA10SSR26 was preferred for genotyping the whole progeny. So other five SSR markers were identified in round three with help of the peach synteny analysis.

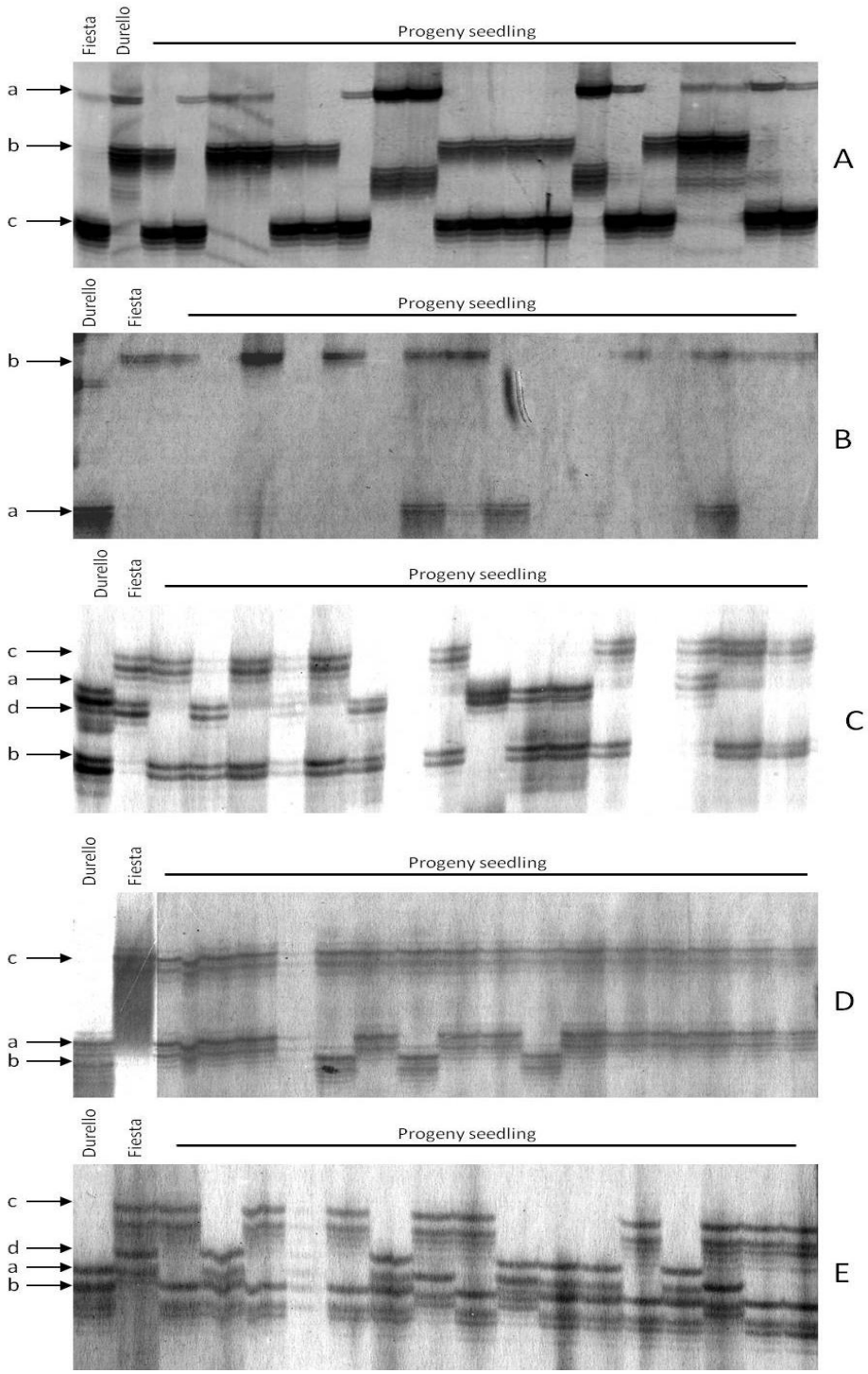


Figure 3.8. A. Polyacrylamide gel profile of DCA10SSR2 marker; three alleles segregating in a "ab X ac" pattern. B. Polyacrylamide gel profile of DCA10SSR22 marker; two alleles segregating in a "ab X aa" pattern; C and E. Polyacrylamide gel profile of DCA10SSR26 and DCA10SSR30 markers, four alleles segregating in a "ab X cd" pattern; D. Polyacrylamide gel profile of DCA10SSR29 marker; three alleles segregating in a "ab x cc" pattern.

Summarizing, a rather low efficiency of SSR markers development based on the released "Golden Delicious" apple genome sequence has been found in our study. In the first and second round of SSR markers development, a total of 16 primer pairs were designed on the 16 contigs downloaded from chromosome 10 on the GDR website. Excluding the five (31.25%) monomorphic and multilocus SSRs, among the remaining eleven polymorphic SSR markers, seven (43.75%) SSRs were not mapping on LG10. In the end, 4 SSR markers were successfully mapped on the top of LG10 with a development rate of only 25%. Some difficulties in developing markers directly on the apple genome in specific regions has been recently reported by researchers that also worked on chromosome 10 in order to make the fine-mapping of the Columnar (*Co*) gene (Bai et al., 2012; Moriya et al., 2012). The SSR development in specific regions of the apple chromosome 10 sequence with the additional examination step on peach genome (scaffold 8) using the BLASTN tool on the GDR website, gave positive results. Among the eight SSRs developed in the third round, seven were polymorphic and six were successfully mapped on the LG10. The 75% of success in SSR markers development observed in the third round was relatively high when compared to 25% in the first and second rounds. The use of the apple/peach synteny was recently reported also by Bai et al. (2012) that confirmed the good similarity between apple chromosome 10 and peach chromosome 8 in the *Co* gene region. However, the attempt to use directly the peach genome scaffold 8 sequence to exploit markers on specific regions of apple LG10 not covered by the apple genome has been justified ineffective in our study, even if only two SSRs were tested with this methodology (DCA10SSR28 and DCA10SSR31). Additional mapping investigations performed on the 5 polymorphic SSR markers not mapping on LG10, showed that the DCA10SSR9, DCA10SSR12 and DCA10SSR16 are on LG4 and DCA10SSR18 and DCA10SSR19 on LG14. This result was slightly unexpected because both chromosomes 4 and 14 are not homeologous to chromosome 10, as reported by Velasco et al., 2010 (see figure 1.1). Therefore, these mapping results could be used to correct a possible assembly error in the apple genome.

3.4.1.4. Genetic linkage mapping with newly developed markers

Using the segregation data of all the eleven SSR markers (nine newly developed SSR markers and two from published SSR markers Hi02d04 and CH02b07), a genetic linkage map has been calculated (Figure 3.9). All the eleven SSR markers were linked in a single linkage group at the

top of apple chromosome 10 with a total length 8.7 cM.

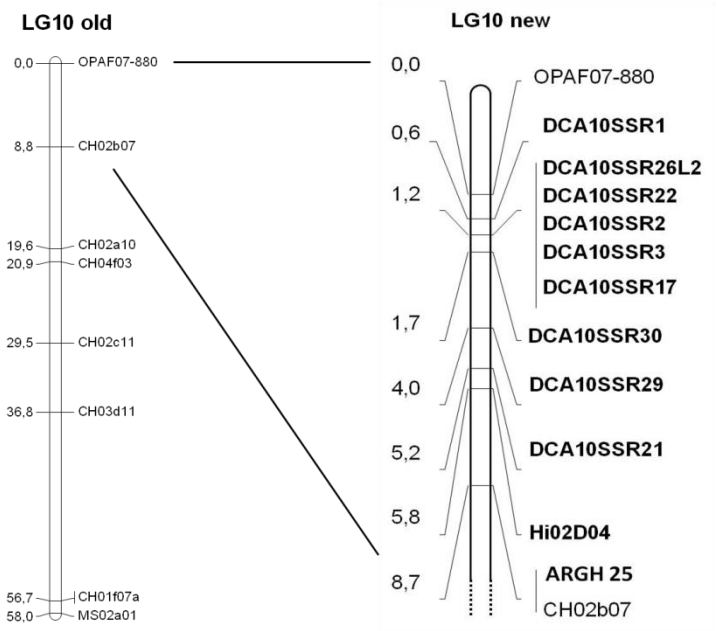


Figure 3.9 Comparison of the genetic linkage mapping of LG10 covered with different number of molecular markers. A. Picture of linkage map on LG10 using only polymorphic markers reported by Tartarini, ete al. (2004); B. A broaden picture of top part of LG 10 from saturated with new SSR markers identified in this study based on 381 progenies from the cross "Durello di Forli" X "Fiesta".

A total of 32 recombinants were found between RAPD marker OPAF07/800 and SSR marker CH02b07 by analyzing 381 plants. Among the 32 recombinants, no recombinant seedling was found between DCA10SSR2, DCA10SSR3, DCA10SSR22, DCA10SSR26 and DCA10SSR17, therefore the order of these cosegregating SSR markers couldn't be determined. Eleven recombinants were identified between CH02b07 and Hi02d04, three between DCA10SSR21 and Hi02d04, two between DCA10SSR29 and DCA10SSR21, eight between DCA10SSR30 and DCA10SSR29, two between the group of cosegregating SSRs and DCA10SSR30, and five between DCA10SSR1 and OPAF07/880. One key informative recombinant was also identified between DCA10SSR1 and the group of cosegregating SSR markers that made the order of marker DCA10SSR1 closer to the RAPD marker. A discrepancy was found between an SSR position in the apple genome and its position in the linkage map because the informative recombinant identified between DCA10SSR1 and DCA10SSR26 suggested that the contig MDC015291.258 carrying DCA10SSR1 was actually much closer to the top of LG10 than the MDC0131226.201 carrying the DCA10SSR26 while from Golden Delicious genome the two

contigs are in the opposite order. Table 3.4 demonstrated the comparison position in the apple and peach genome of the 11 SSR markers studied in this work. From DCA10SSR3 to the CH02b07, the consistence of corresponding position order for each SSR marker between the GD apple genome and peach genome was perfect, which was also in accordance with our genetic linkage mapping result. But for the DCA10SSR1, the corresponding position in peach genome is supporting our finding that the DCA10SSR1 is located at the top of the chromosome 10 in respect to its position in the apple genome.

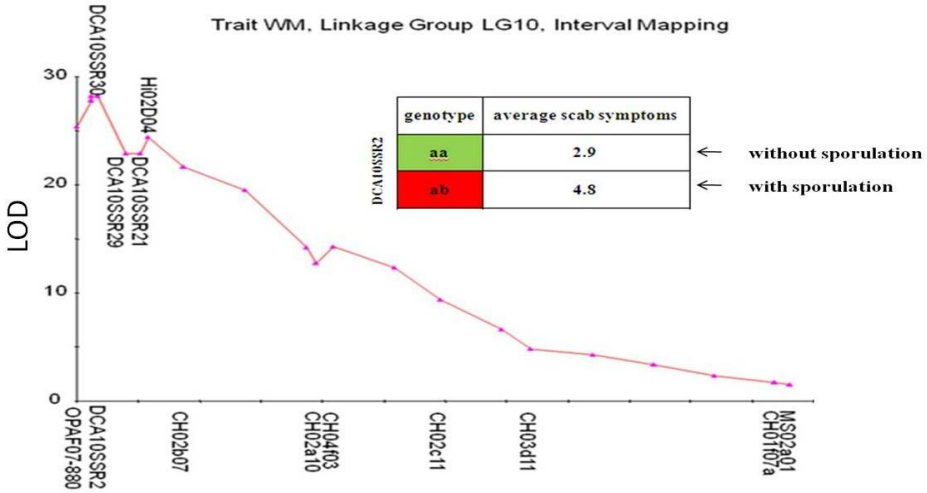
Table 3.4. The comparison of 11 SSR markers in the apple and peach genome

Apple genome cordinates	Mapping order in Du x Fu population		Peach genome cordinates
ch10:240028..292677		DCA10SSR1	scaffold 8: ~ 37300
chr10:32650..42448	Cosegregating	DCA10SSR26	
chr10:111770..128249		DCA10SSR22	scaffold 8: ~ 63000
ch10:718885..729454		DCA10SSR2	
ch10:890245..905668		DCA10SSR3	scaffold 8: ~ 1100000
ch10:1076249..1096702		DCA10SSR17	scaffold 8: ~ 1505000
ch10:1156516..1187334		DCA10SSR30	scaffold 8: ~ 1576000
ch10:1936488..1949522		DCA10SSR29	scaffold 8: ~ 2520000
ch10:2008671..2027387		DCA10SSR21	scaffold 8: ~ 2614000
ch10:2710638..2721470		Hi02d04	scaffold 8: ~ 3398000
ch10:5600392..5606934		CH02b07	scaffold 8: ~ 5817000

3.4.1.5. QTL analysis and relationship between susceptibility scale and genotype

In order to facilitate the investigation of the relationship between genotype and the susceptibility scale, progeny seedlings that inherited the SSR alleles from one of the Durello chromosomes were recorded as genotype “aa” and from the other chromosome were recorded as “ab”. Genotyping data of entire mapping population were transformed into “aa” and “ab” with all the 11 SSR identified on the top of LG10. Through the interval mapping algorithm method, the QTL analysis was performed on a sub-population of 171 progeny seedlings in which no missing data were present. Main peak presumably responsible for the *Rvi13* scab resistance was detected. The highest peak was located in the SSR cosegregating group with a very high LOD score about 28. Another peak was also identified beside the main peak in correspondence of the SSR Hi02d04 with LOD score about 24 (Figure 3.10). Closer

examination of the DCA10SSR2 data (as the representative SSR marker of the cosegregating SSR group) on the entire mapping population and scab resistance data, revealed that the average scab symptom of the “aa” genotypes was 2.9 while the “ab” genotypes possessed an average scab symptom of 4.8. Therefore, the “aa” was the resistance genotype but “ab” was the susceptible genotype (Figure 3.10 and 3.11). The resulted “aa” resistance genotype from QTL analysis was used to trace back alleles in coupling with the *Rvi13* resistance for all the 11 SSR markers (Table 3.5).



MW	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	3.0	4.0	4.0	4.0	4.0	5.0	5.0	5.0	5.0	5.0	6.0	6.0	6.0	6.0	6.0	6.0
Plant	187	189	194	201	202	113	120	128	130	131	139	105	142	137	151	154	155	110	116	117	119	121	253	258	259	265	270	271
DCA10SSR2	aa	aa	aa	aa	aa	aa	aa	aa	aa	aa	aa	aa	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab	ab

Figure 3.10. QTL of scab resistance identified on the LG10 of scab tolerant cultivar “Durello di Forli” and relationship between susceptibility scale and genotype from DCA10SSR2 in the progeny seedlings from cross of “Durello di Forli” and “Fiesta”

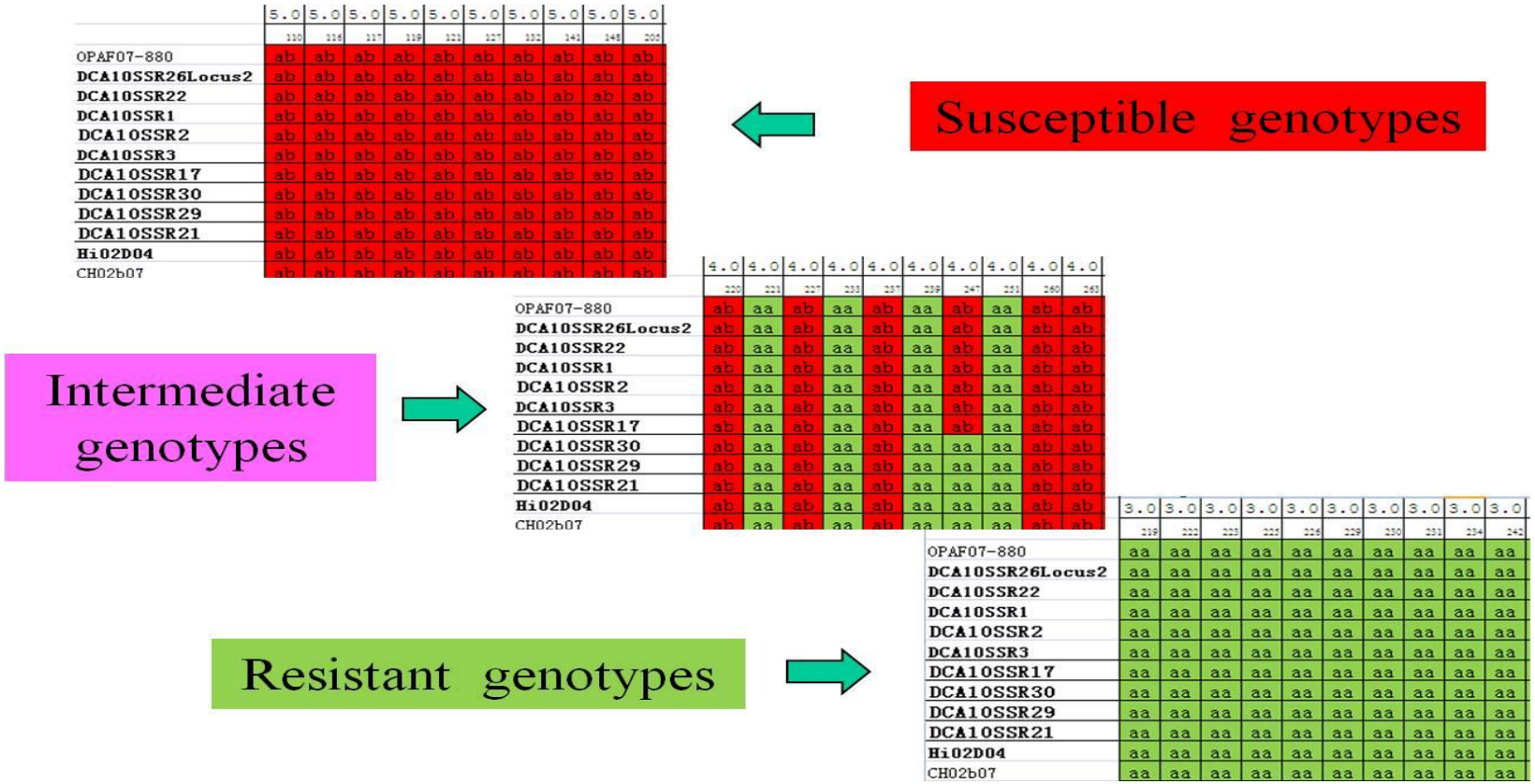


Figure 3.11. Association of the susceptible scale with two different genotypes ("aa" and "ab") transformed from all molecular markers evaluated in this study in the progeny seedlings of "Durello di Forli" and "Fiesta"

Table 3.5. SSR markers mapping on the top of LG10 based on the cross of "Durello di Forli" and "Fiesta"

SSR name	Repeat sequence	Segregation type in Du X Fi population	Durello alleles in coupling with resistance
DCA10SSR26	CT	ab x cd	b
DCA10SSR27	GA	a0 x 00	a
DCA10SSR22	CATA-AT	ab x aa	b
DCA10SSR1	TA-GA	ab x cd	a
DCA10SSR2	CT	ab x cc	a
DCA10SSR3	CT	ab x cd	a
DCA10SSR17	GA	ab x cd	a
DCA10SSR30	GA	ab x cd	b
DCA10SSR29	TA-GA	ab x cc	a
DCA10SSR21	TA-GT	ab x ac	b
Hi02d04	CT	ab x cd	b
CH02b07	CT	ab x cd	a

3.4.1.6. Analysis of RAPD OPAF07/880

The RAPD OPAF07/880 segregating band from Durello di Forli was successfully cloned and sequenced. The development of SCAR marker on the OPAF07/880 sequence obtained from the cloned band was unable to exhibit stable amplification even after various attempts of optimization. The Durello di Forli parent exhibited the band but also Fiesta sometimes amplified a band of the same size. With the attempt to know detailed sequence information so as to find SNPs in the RAPD sequence, the AF07 amplicon from Durello di Forli was used for direct sequencing, but no SNPs were identified. The obtained AFO7 nucleotide sequence was also used as a query to search the homologous contigs on the apple genome through BLASTN server of the GDR website. The results showed that the AF07 shared high sequence identity of the contig MDC010342.280 (5762 bp), but this small contig was in the unanchored category in the "Golden Delicious" genome. The chromosome 0 was a presumed chromosome served as a collection of unanchored repeated sequences without unambiguous position information that are very abundant in the apple genome and which assembly is very difficult (Velasco et al., 2010). In this case, acquisition of the sequence flanking the AF07 through BLASTN tool on the GDR website was not feasible. Thanks to the available BAC library from the cultivar Florina, the AFO7 SCAR marker primer pairs served as probes

to perform PCR-based screening of BAC library. This analysis revealed that two BAC clones amplified the AF07 marker: 72E11 and 89O11. The plasmids extracted from these two positive clones were sequenced with T7 and SP6 primers along with two new primers designed inside the AF07 sequence with an opposite direction in order to get sequence information on the AF07 flanking regions (forward primer 5'- GGAAAGCGTCCAGTGGTG-3' and reverse primer 5'-CGTCGATGCCCCATATAACA-3'). The BLASTN results were summarized in Table 3.6.

Table 3.6. BLASTN results of sequences from positive clones with the T7 and SP6 primers and specific primers developed on the RAPD OPAF07/880 sequence

Positive clones name	Primer for sequencing	BLASTN results on GDR	Size (bp)
72E11	AF07 sequencing primer Fwd	MDC010342.280 unanchored:76688032..76693793	5762
	AF07 sequencing primer Rvs	MDC011683.619 chr10:66513..71408	4896
		MDC010342.280 unanchored:76688032..76693793	5762
	SP6	MDC000285.365 chr10:675550..677908	2359
		MDC011720.199 chr10:758392..764419	6208
		MDC000285.562 chr10:758606..776965	18360
89011	SP6	MDC020040.455 chr10:61517..68160	6644
		MDC000158.245 chr10:61517..64119	2603
		MDC011068.496 chr10:63556..69415	5860
		MDC011068.528 chr10:83574..93539	9966
	T7	MDC009544.134 chr10:10432078..10441093	9016
	AF07 sequencing primer Fwd	sequencing quality was not good	
	AF07 sequencing primer Rvs	sequencing quality was not good	

Since the majority of BLASTN results of 72E11 and 89O11 plasmids exhibiting a high similarity to various contigs from top of the LG10 enabled us to propose a hypothesis about the contigs surrounding AF07 based on the apple genome even if the distance between the T7 and SP6 ends were not compatible with the expected size for a single BAC clone (Figure 3.12). These sequence information are going to be used to develop new reproducible markers in the AF07 region.

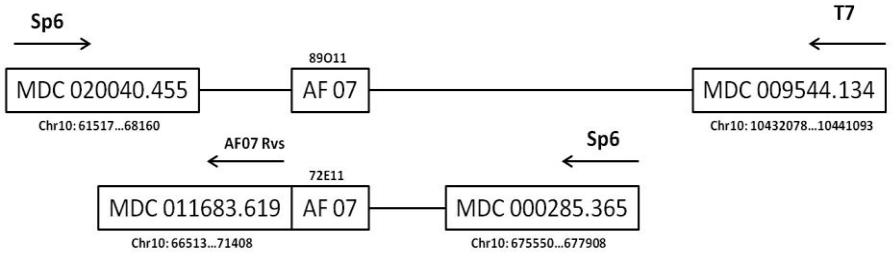


Figure 3.12. A hypothesis on contigs surrounding AF07 marker based on the sequence of positive plasmids carrying the RAPD fragment of Florina BAC library

Taking advantage of the Florina BAC library, we obtained four sequence fragments with high similarity to the contigs from LG10 on GD apple genome. Considering a relatively high rate of genetic diversity present in apple, including a supposedly high level of presence/absence gene variation between different haplotypes (Velasco et al., 2010), the next steps has to be still tested through new markers development and mapping in our segregating progeny.

3.4.1.7. Candidate gene-based SNP discovery

Since the result of QTL analysis suggested that the region responsible for the *Rvi13* scab resistance with the highest LOD was inside a group of cosegregating SSR markers, some candidate genes were selected for detection of SNPs. Take into account of the rather low efficiency of LG10 SSR markers development, the SSR cosegregating contigs mapped on LG10 were mainly used for SNP development. The consensus gene set under these contigs was searched mainly for putative genes involved in the pathogen resistance pathway. Three rounds of candidate gene primer design were conducted in order to identify SNPs.

In the first round, 8 candidate genes (CGs) were identified and analyzed. Except for ARGH8 and ARGH25, two ARGH genes found in published literature, another two candidate genes MDP0000637744 from the contig MDC005799.536 (the same as DCA10SSR2) and MDP0000182552 from an overlapping contig with MDC005799.536 were also found homologous to "*Malus floribunda* clone ABGA003014CT putative NBS-LRR disease resistance protein gene" through the BLASTN search on the NCBI website using their coding sequence as the query (Table 3.7). Moreover, another three genes of interest, a WHY transcription factor (MdWhy) from the contig MDC004168.448, a glucanase (MdGluca) from contig MDC015291.258 (the same as DCA10SSR1) and a phosphatase 2C (MdPhosph) from MDC013845.347 (same as DCA10SSR3) were also selected for SNPs discovery (Table 3.7). In addition to the candidate gene found on LG10 from apple genome, one candidate gene on an hypothetical protein (MdHyp) was identified around 100 kb on the homologous peach scaffold 8 region based on the high level of synteny with apple LG10 which was also confirmed during SSR markers development (Table 3.7).

Table 3.7. Detailed list of three rounds candidate genes development for SNP detection

Round No.	CG name*		Consensus gene name on the GD genome	Contig		CG Primers	Size (bp)	PCR amplifiaction	Amplicons purification	SNP No. in Durello	TSP pimer design
Round 1	MdGluca		MDP0000345322	MDC015291.258	Fwd	CGCACATGTGAGAAATCCAG	893	Yes	PCR product	>1	Yes
					Rvs	GACATTGTCTCCAGCGTCAT					
	MdPhosph		MDP0000296528	MDC013845.347	Fwd	GACAAGCTCCTCAAAGCACTG	742	Yes 3 different amplicons	Gel cutting	No	No
					Rvs	CAGAGGAGCTCAAGGGATAA					
	MdHyp1			MDC016545.264 chr11:22110221..22114434	Fwd	GTGTCCTGGACGGCTATGAT	699	Yes	PCR product	No	No
					Rvs	TTCCATGGCTTCCTCAAATC					
	MdWhy			MDC004168.448	Fwd	AGAGGTTGGCGCTCTGATAA	719	Yes	PCR product	2	Yes
					Rvs	TCTGGTCTTTAGCAGGTTGTTG					
	ARGH	ARGH8		MDC002167.405 chr5:16179878..16190195	Fwd	CTAAAATAAACCGTACTCTAATTG	380	Yes 2 different amplicons	Gel cutting	>4	No
					Rvs	GTCTTTCTTAGGTTGTTCTTTC					
		ARGH25	MDP0000136651	MDC007061.242	Fwd	CAAACATCATCGTAATTTTGACG	420	Yes 2 different amplicons	Gel cutting	3	No
					Rvs	CATACTCTTCATGAGGATAATTC					
		MDP744	MDP0000637744	MDC005799.536	Fwd	CCATCTTCCATTGTCCTGAAA	705	Yes	PCR product	9	Yes
					Rvs	CTTCTGGTCCCATCCTTTGA					
		MDP552	MDP0000182552	MDC005799.573	Fwd	CCATCTTCCATTGCCCTTCT	709	Yes	PCR product	No	No
					Rvs	CTTCTGGTCCCATCCTTTGA					
Round 2	MDP130		MDP0000319130	MDC008416.202	Fwd	CGGGGAAGAACCTCAACA	963	Yes	PCR product	4	Yes
					Rvs	TCAGAATGCCCAGATGAACA					
	MDP053		MDP0000263053	MDC012944.362	Fwd	GAATTCATTGTTGCGCTGAA	976	Yes	PCR product	No	No
					Rvs	GTCAATAGAGACCGAATTGTAGCA					
Round 3	MDP616		MDP0000196616	MDC013126.201	Fwd	AGCATCCGCACCTCCTCT	1458	Durello no amplification			No
					Rvs	ACCGGCTATGCTAACAACTCC					
	MDP337		MDP0000231337	MDC022095.200	Fwd	GTATCCACAGCCCTCATCCA	1126	Yes	PCR product	1	Yes
					Rvs	TGAATGGGAATGATCCAACA					
	MDP526		MDP0000296526	MDC013845.347	Fwd	CTTGTGCTGCTTGTCTCGAA	1100	Yes	PCR product	>3	Yes
					Rvs	AAGGTGACAGCACCGAGAAG					

*MdGluca = glucanase; MdPhosp = phosphatase 2C; MdHyp = hypothetical protein; MdWhy = WHY transcription factor; MDP744, MDP552, ARGH8 and ARGH25 = LRR/NBS gene; MDP130 = transcriptor co-suppressor; MDP053 = hypothetical protein; MDP616= No significant homology; MDP337 = Programmed cell death (PCD) inhibitor; MDP526 = DNAJ heat shock protein.

Except for MdPhosph, the specific amplicons of other five candidate genes were obtained with the expected size. So, the sequencing of these five gene fragments was performed directly on PCR products. In the case of MdPhosph, three amplicons of different sizes were amplified. The 742 bp band was cut from the agarose gel and used to conduct PCR product purification. All candidate genes in the first round gained sequencing of high quality. However, only three candidate genes MdGluca, MdWhy and MDP0000637744 contained SNPs in the Durello di Forli parent.

At least 9 SNPs were identified in candidate gene MDP0000637744, while no SNP was detected in the candidate gene MDP0000182552. The alignment of two sequences revealed that the two sequences were identical apart from the 9 SNPs found in the MDP0000637744, implying that MDP552 was allelic to the gene MDP0000637744 of Durello di Forlì. Therefore, it seems that in this region only one gene is present with two alleles while the heterozygosity of the GD genome may have artificially produced duplications in the assembled sequence, as happened in the construction of a physical map of the heterozygous grapevine Cabernet Sauvignon (Moroldo et al. 2008). But the position of MDP0000182552 and MDP0000637744 exhibited on the GDR website were clearly not overlapping (Figure 3.13).

Figure 3.13. Picture from the GDR website of position of MDP0000637744 and MDP0000182552, through resequencing the specific amplicon of MDP0000637744 and MDP0000182552 revealing the MDP0000182552 actually was one allele of gene MDP0000637744, this finding proved a error made by assembly GD apple sequences.

In the second round, 2 candidate genes MDP0000319130 from the contig MDC008416.202 (as DCA10SSR30) and MDP0000263053 from contig MDC012944.362 were analyzed, but only candidate gene MDP0000319130 showed polymorphic SNPs.

In the third round, 3 candidate genes MDP000019616 from contig MDC013126.201 (as DCA10SSR26), MDP0000231337 from contig MDC022095.200 (as DCA10SSR22) and MDP0000296526 from contig MDC013845.347 (as DCA10SSR3) were analyzed. MDP0000231337

and MDP0000296526 gave expected amplicons in both parents Durello di Forli and Fiesta. But for candidate gene MDP000019616, no amplification product was obtained from Durello di Forli. Therefore this gene was discarded. More than one SNP were detected in both candidate genes MDP0000231337 and MDP0000296526 after closer examination of the sequencing results.

3.4.1.8. SNP genotyping and mapping

For TSP primer design on suitable SNPs, it was generally followed the criteria described in the materials and methods. In our study, another criterion of great importance should be point out. TSP genotyping method was extremely dependent on the changes in the temperature, one little modification of annealing temperature from the TSP primer could make the reliability of the TSP genotyping methods unjustified, Considering the general criteria Hayden et al., (2009) proposed and new criterion found in our study together with the normal criteria for the primer design, six TSP primers were designed in each candidate gene (MdWhy, MdGluca, MDP0000637744, MDP0000319130, MDP0000231337 and MDP0000296526) to genotype the alleles harboring the SNPs and further mapping these SNP markers (Table 3.8).

Table 3.8. Detailed list of 5 new SNP markers

SNP name	SNP alleles in Durello	SNP alleles in Fiesta	SNP allele assayed	Locus-specific primers	Tm (°C)	Reference SNP allele size (bp)	TSP primers[a,b]	Tm (°C)	Alternate SNP allele size (bp)
MdGluca-TSP	C/A=M	A/A	C		62	893	**CG**TGTTATATATTGTAAATCTTTTG	54	278
MDP0000231337-TSP	T/A=W	A/A	T		60	1126	**GC**CGAAATGACTAAACCAA	56	338
MDP0000637744-TSP	A/G=R	G/G	A	Same as CGs primers	60	705	**TC**CATGGTTGTACTAAACTGA	53	314
MDP0000296526-TSP	G/T=K	T/T	G		60	1100	**GG**GTATTATCCACATTATTGAC	55	307
MDP0000319130-TSP	T/C=Y	C/C	T		62	963	**CC**GAGCTATTTGACAATCATA	54	275

[a] A SNP allele complementary to the allele-specific primer produces the 'reference SNP allele' (smallest) PCR product, whereas those without complementarity produce the 'alternate SNP allele' (largest) product.

[b] The arbitrary 5' -nucleotide sequence indicated in underlined, bold font is non-complementary to the target DNA

Underlined alleles were found to be in coupling with the scab resistance

Table 3.9. Optimized TSP program for each SNP marker

TSP program name		TSP programs										SNP name
TSP-62	Temperature	95 °C	94 °C	62 °C	72 °C	94 °C	47 °C	94 °C	53 ~ 55 °C	72 °C	16 °C	MdGluca-TSP MDP0000319130-TSP
	Time	10 min	30 s	30 s	1min 30 s	10 s	30 s	30 s	30 s	5 s	∞	
	Cycles			20		5			10			
TSP 2	Temperature	95 °C	94 °C	60 °C	72 °C	94 °C	42 °C	94 °C	53 °C	72 °C	16 °C	MDP0000637744-TSP
	Time	10 min	30 s	30 s	1min 30 s	10 s	30 s	30 s	30 s	5 s	∞	
	Cycles			20		5			10			
TSP 4	Temperature	95 °C	94 °C	60 °C	72 °C	94 °C	46 °C	94 °C	55 °C	72 °C	16 °C	MDP0000296526-TSP
	Time	10 min	30 s	30 s	2 min	10 s	30 s	30 s	30 s	5 s	∞	
	Cycles			20		5			10			
TSP 5	Temperature	95 °C	94 °C	60 °C	72 °C	94 °C	48 °C	94 °C	56 °C	72 °C	16 °C	MDP0000231337-TSP
	Time	10 min	30 s	30 s	2 min	10 s	30 s	30 s	30 s	5 s	∞	
	Cycles			25		5			10			

Since TSP primers designed for MdGluca and MDP00000319130 possessed an identical annealing temperature for locus-specific primers (62 ℃) and TSP primers (47 ℃ and 54 ℃ with the 5'tag), the same TSP program named TSP-62 was set up (Table 3.9). After preliminary trail on a small number of progeny to confirm the effectiveness of the TSP program used for genotyping SNPs, 32 recombinant seedlings were subjected to screening with MdGluca and MDP00000319130 TSP primers. The agarose gel result (Figure 3.14) of MdGluca TSP genotyping demonstrated that 278 bp allele-specific band harboring the C nucleotide originated from Durelllo di Forli was segregating in the progeny, and the analysis of recombinant seedlings suggested that the MdGluca TSP marker was successfully developed. For candidate gene MDP00000319130, the allele-specifc band of the size 275 bp carrying the A nucleotide derived from Durello di Forli (Figure 3.15) was also obtained unambiguous segregation in the recombinant seedlings, corroborated the SNP found in the MDP00000319130 candidate gene was correct.

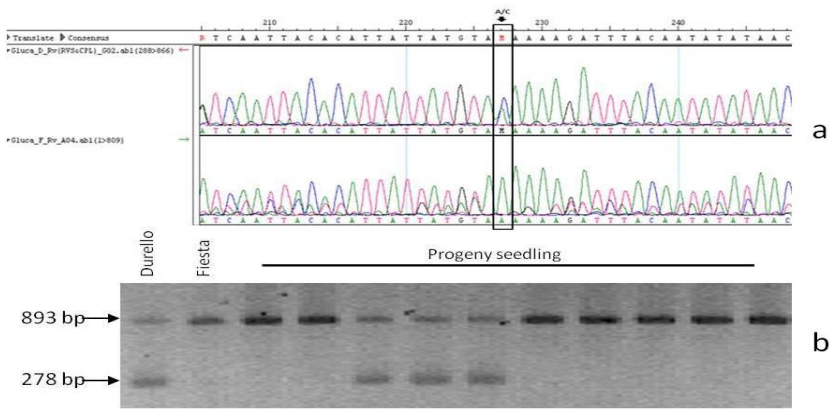


Figure 3.14. a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of the MdGluca amplicon. The A/C polymorphism in Durello is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 278 bp band was the polymorphic band carrying the C nucleotide.

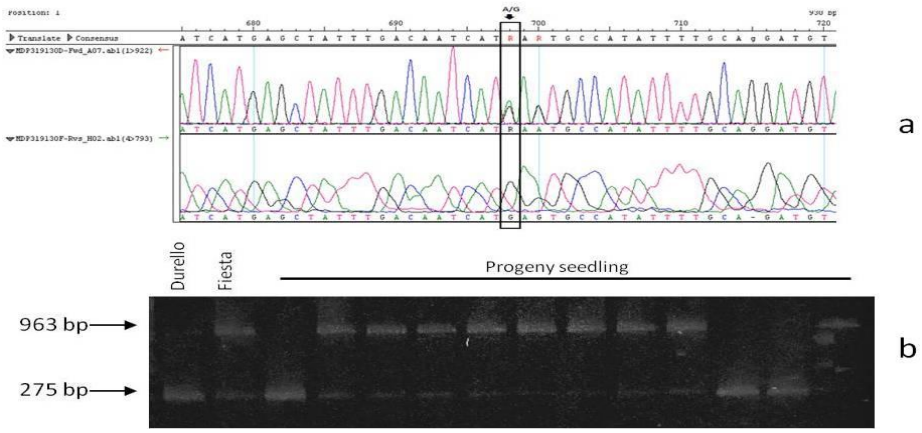


Figure 3.15. a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of the MDP0000319130 amplicon. The A/G polymorphism in Durello is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 275 bp band was the polymorphic band carrying the A nucleotide.

In the candidate gene MdWhy, a T/G single nucleotide polymorphism was found in the Durello di Forli parent. After performing the initial TSP trial on some seedlings, the segregation pattern of the allele-specific band with 595 bp fragment carrying the A nucleotide demonstrated that the segregation data were not fitting the results of other LG10 markers genotyping (Figure 3.16). So, MdWhy gene was not mapping on LG10 and it was not futher analyzed. Moreover, the unsuccessful mapping of SNP found in the candidate gene MdWhy on LG10 implicated that the MDP0000168614 as the source gene is not from LG10.

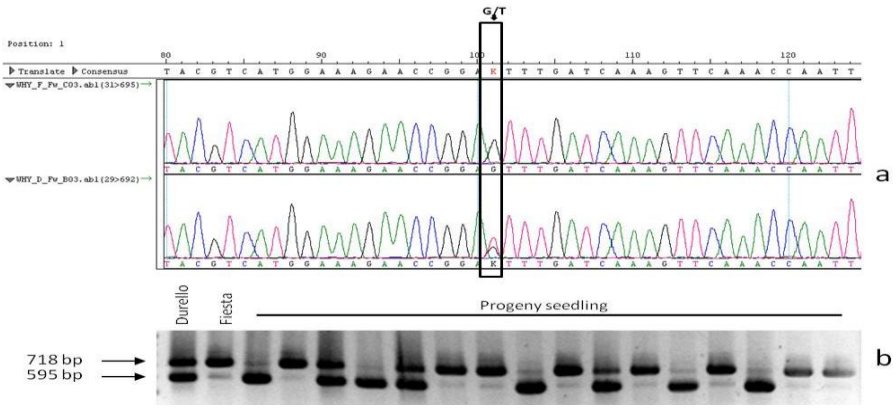


Figure 3.16. a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of MdWhy amplicon. The T/G polymorphism in Durello di Forli is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 595 bp band was the polymorphic band carrying the A nucleotide.

The TSP amplification protocol of the MDP0000637744 was modified to take into account of the lower annealing temperature for both locus specific primer and the TSP primer. A new TSP program named TSP2 was built up for MDP0000637744 , TSP2 program contained a 60℃ of Tm for locus specific primers and 42℃ for TSP primers without 5’ tag (Table 3.9). The 314 bp specific–allele band containing the A nucleotide possessing a consistent segregation with LG10 markers, suggesting MDP0000637744 TSP primer and PCR program was successful for genotyping the SNP (Figure 3.17).

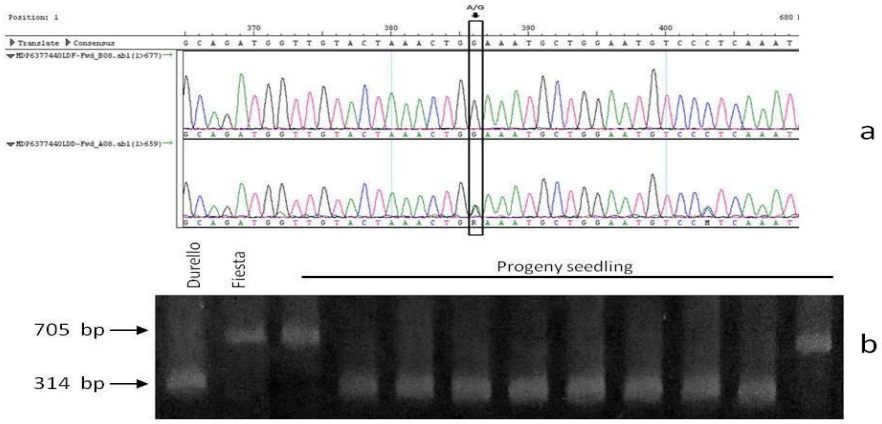


Figure 3.17. a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of MDP0000637744 amplicon. The A/G polymorphism in Durello di Forli is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 314 bp band was the polymorphic band carrying the A nucleotide.

For candidate gene MDP0000296526, TSP4 PCR program was set up to consider the TSP primer annealing temperatures at 46℃ and 55℃ with the 5’ tag (Table 3.9). The agarose gel (Figure 3.18) showed that the 307 bp band was polymorphic between two parents Durello di Forli and Fiesta. The result of scoring the 307 bp band in the progeny corroborated that the T/G SNP identified in the candidate gene MDP0000296526 was from LG10.

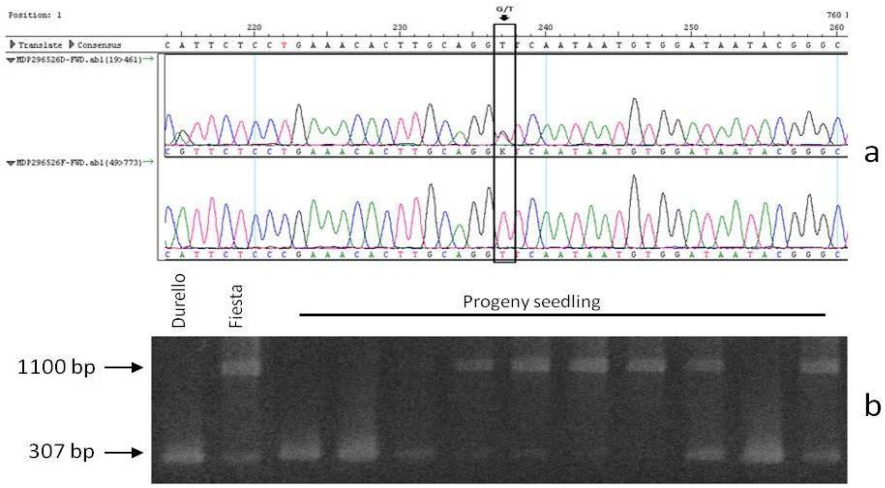


Figure 3.18 a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of MDP0000296526 amplicon. The T/G polymorphism in Durello di Forli is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 307 bp band was the polymorphic band carrying the G nucleotide.

TSP5 PCR program (Table 3.9) was modified for TSP primer with annealing temperature at 48℃ and 56℃ with 5' tag for genotyping the progeny seedlings for the T/A kind of SNP inside the candidate gene MDP0000231337 from Durello di Forli. Because of the consistent genotyping score of the 338 bp allele-specific band carrying the T nucleotide with the genotyping results of the LG10 markers on progeny seedlings (Figure 3.19), justifying the MDP0000231337-TSP primer designed for SNP genotyping was a reliable SNP marker.

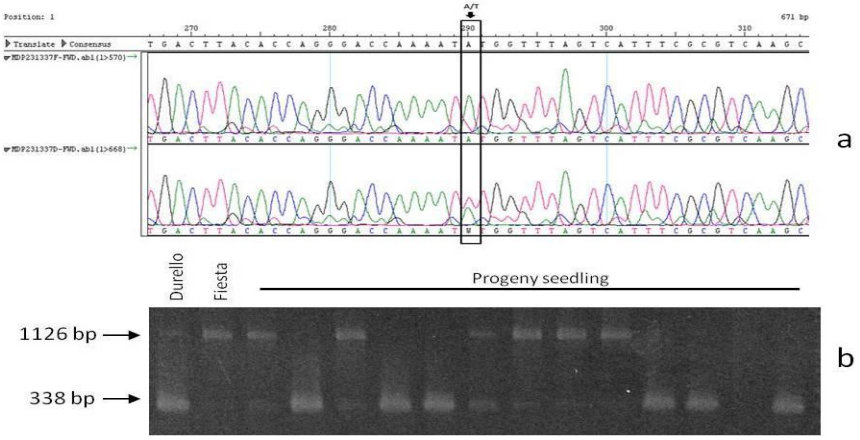


Figure 3.19. a. Portion of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing of MDP0000231337 amplicon. The T/A polymorphism in Durello di Forli is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the recombinants and 2 parents by the TSP method, the 338 bp band was the polymorphic band carrying the T nucleotide.

Closer examination of the genotyping data obtained from all the recombinant seedlings with the five newly developed SNP markers, showed that only the SNP identified in the candidate gene MDP0000296526 was in coupling with the *Rvi13* scab resistance. In particular, the allele-specific band containing the G nucleotide descended directly from Durello di Forli. The remaining four specific Durello di Forli SNPs found in MdGluca, MDP0000637744, MDP00000319130 and MDP0000231337, were actually in repulsion with the scab resistance. Because the polymorphic allele-specific bands from these four candidate genes was scored as scab susceptible genotype.

3.4.2. Fine mapping of scab resistance gene *Rvi5 (Vm)*

3.4.2.1. Mapping of SSR markers for *Rvi5*

Screening of the four available markers with the recombinant seedlings found in the large Murray-derived progenies showed that: except for the marker B-SSR, the remaining three markers (C13-SSR, G-SSR and Hi07h02) yielded the expected SSR segregation (Figure 3.20 a-c).

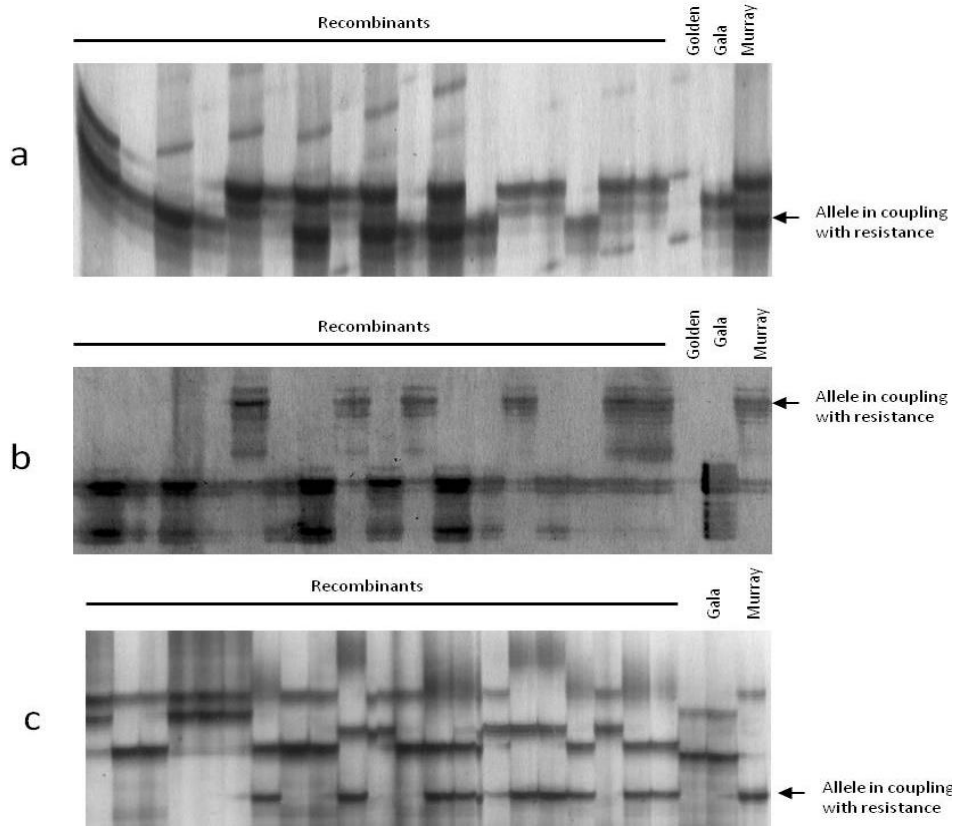


Figure 3.20. Polyacrylamide gel profile of 17 recombinants segregation pattern of the three SSR markers flanking the *Rvi5* region. a G-SSR. b C13-SSR. c Hi07h02. Bands in coupling with the resistance are indicated with the arrow.

The difference of the segregation patterns between the three SSR markers on the recombinant enabled us to map C13 and Hi07h02 on the same side of *Rvi5* locus and the G-SSR on the opposite side (Table 3.10). Moreover, four recombinant seedlings (Q22, N47, B39 and O30) found between C13 and Hi07h02 revealed that the marker C13 SSR was closer to the *Rvi5* locus. When the B-SSR was first tested, unclear results were obtained after

amplification and electrophoresis, leading to a possible loss of information. So the B-SSR marker was discarded in our study. A little discrepancy was also found between our data and the previous dataset. The genotyping of G-SSR on the recombinant seedling M42-pato was scored as susceptible in our analysis, whereas the old data showed it was resistance type. From our data, it seems the G-SSR marker was much closer to the *Rvi5* locus than the old scoring.

Table 3.10. Position of 3 SSR markers localized through the segregation pattern differences in the recombinant seedlings

Recombinants	OLD DATA	NEW DATA	*Rvi5*	NEW DATA	NEW DATA
	G-SSR	G-SSR		C13-SSR	Hi07h02
H05	R	R	R	R	R
Z33	R	R	R	R	R
A05	S	S	S	S	S
A19	S	S	S	S	S
C16	S	S	S	S	S
G14	S	S	S	S	S
G15	S	S	S	S	S
Q22	S	S	S	S	R
N47	S	S	S	S	R
O32	S	S	S	R	R
S01	uu	R	R	R	R
B39	R	R	R	R	S
O30	R	R	R	R	S
O13	R	R	R	S	S
R20	R	R	S	S	S
M42_pato	R	S	S	S	S

Notes: uu = missing data; R in blue box means allele in coupling with the *Rvi5* resistant allele; S in green box means allele in repulsion with the *Rvi5* resistant allele.

3.4.2.2. Candidate gene-based SNP discovery and mapping

Three rounds of candidate gene primer design were conducted in order to search for SNPs in the selected gene sequences. The Murray BAC clone sequence identified by the C13-SSR and Hi07h02 markers was focused on to detect SNP in the first round, while the BAC on the G-SSR side was analyzed in the second and third rounds. In total, fifteen primer pairs were designed on the sequences surrounding the *Rvi5* locus (Table 3.11).

Table 3.11. Detailed list of three rounds candidate genes development for SNP detection

Round No.	CG name		CG Primers	Size (bp)	PCR amplifiaction		As SCAR marker	Amplicons purification for sequencing	SNP No. in Murray	TSP pimer design
Round 1	Shaggy-like-Ser	Fwd	CACATCAAAAGCCGGAAAAA	740	2 different amplicons of 740 bp and ~900 bp		No	Gel cutting	3	Yes
		Rvs	GCGCGATAGAGGATTAGACG							
	Phosphatase2C	Fwd	GCCAAAAATGTTCACGAGGT	777	Yes		No	PCR products	4	Yes
		Rvs	ATGCCAAGATTTTGGACTGC							
	Mlo like	Fwd	AAAGGTGTGGCATCGGTAAG	843	Yes		No	PCR products	6	Yes
		Rvs	CTGAAAACCGCCATTAGAGC							
	WRKY	Fwd	AATCTCCTGGTTTTGATACATTCA	850	Yes		No	PCR products	4	Yes
		Rvs	CAATATGAATGGGGGAGGAG							
Round 2	Contig1-1	Fwd	CTTCAACCACTTGCGATTCA	984	Murray	1 specific amplicon of 984	Yes	No		No
		Rvs	TTTAACGAGGTTCCTCAACAGTC		GD/Galaxy	no amplicon				
	Contig1-2	Fwd	GGCCGAAAACTGGGTAAGAT	1400	Murray	1 specific amplicon of 1400	Yes	No		No
		Rvs	GACGCCTTACAAACCAAAGG		GD/Galaxy	no amplicon				
	C3-1	Fwd	TCCAACTTTTAAGTAGAATACCAAACG	1022	Yes		No	PCR products	11	Yes
		Rvs	GGAGAAGAAATTACTGGCGATG							
	C5-1	Fwd	GTTTGCGAGACAAACACCTG	1037	Murray	1 specific amplicon of 1037 bp	Yes	No		No
		Rvs	CTGCAGTGGATGAGATCGAG		GD/Galaxy	1 specific amplicon of ~ 700 bp				
	C7-1	Fwd	TTGTAAGTGATCAGTTGTGGCAGT	1150	Yes		No	PCR products	2	Yes
		Rvs	TCCTTCATATAGCCAGTTTCTTCA							
	C7-2	Fwd	ACATAGCTGCTGGAAAATCG	1171	Murray	2 amplicons of ~ 900 bp and 1171 bp	No	No		No
		Rvs	GCTTCGTCACCACTGGTTTT		GD/Galaxy	1 specific amplicon of ~ 900 bp				
	C8	Fwd	CTCGCCTTAAAATTGCATCA	1279	Yes		No	PCR products	No	No
		Rvs	GTTTGCTCTGCGAAGTACGA							

(To be continued)

(Continued table)

Round No.	CG name		CG Primers	Size (bp)		PCR amplifiaction	As SCAR marker	Amplicons purification for sequencing	SNP No. in Murray	TSP pimer design
Round 3	C1-1	Fwd	ATTGGTGATAAGGGTGGCAAG	860		Yes	No	PCR products	No	No
		Rvs	CACCTGAGGTCGCCTACAA							
	C1-2	Fwd	TGGTTGTCCAAAACCCGTA	850	Murray GD/Galaxy	1 specific amplicon of 850 bp no amplicon	Yes	No		No
		Rvs	AAGCATTCATCACATGGTAGTCA							
	C3-2	Fwd	ATGGATGCTTTCCGAGATTC	883		Yes	No	PCR products	No	No
		Rvs	GATGGTGCTCTTGCCAATTC							
	C3-3	Fwd	TCTTCTTCCTGATGCTCCAGA	900		Yes	No	PCR products	>2	Yes
		Rvs	TGCTCCATGCTTACAAGGAA							
	C5-2	Fwd	TGGTAACTGTTAATTGACTACCATGTG	832		Yes	No	PCR products	No	No
		Rvs	AATTCCCCACCCCACAAT							

In the first round, four putative genes involved in the resistance were identified, namely Shaggy-like kinase, Phosphatase2C, Mlo-like and WRKY transcription factor. In *Arabidopsis*, Jonak et al. (2000) showed that *WIG*, a *SHAGGY*-like homolog, responded to wounding, suggesting SHAGGY-like kinases are also involved in the plant response to stress. Members of the complex family of WRKY transcription factors have been implicated in the regulation of transcriptional reprogramming associated with plant immune responses (Eulgem et al., 2007). A small clade (subgroup IIa) of WRKY genes, comprising *AtWRKY18, AtWRKY40,* and *AtWRKY60*, play important and partly redundant functions in regulating plant disease resistance (Eulgem et al., 2007). Elliott et al. (2002) reported a *Mlo*-like gene in wheat and demonstrated that it is associated with powdery mildew resistance. Protein phosphatase 2Cs (PP2Cs) have been demonstrated to play critical roles in regulation of plant growth/development, abscisic acid signaling pathway and adaptation to environmental stresses (Hu et al., 2009). The same authors reported the cloning and molecular characterization of a novel rice protein phosphatase 2C gene, *OsBIPP2C2*, and they identified that the *OsBIPP2C2a* over expressing transgenic tobacco plants showed constitutive expression of defense-related genes. These four candidate genes, each obtained amplicons of expected size successfully. Excluding the Shaggy-like, other three candidate genes (Phosphatase2C, Mlo-like and WRKY) were specifically amplified. So the purification of these three genes was performed directly on the PCR products. In the case of Shaggy-like, two kinds of amplicons were produced with different size, one was around 900 bp and the other was 740 bp expected. For sequencing, the 740 bp band cut from the agarose gel and purified. All candidate genes in the first round gained sequencing of high quality. Several SNPs were identified in each candidate gene (Table 3.11).

The preliminary assembly of the BAC2 sequence resulted in six contigs which sequences were used for SNP detection in the second and third round. The preliminary BLASTN analysis of these contigs on GDR website revealed that most of the contigs did not produce significant alignment with the chromosome 17 but chromosome 7, 3, 13 and 15; only contig C3 and contig C5 showed expected high homology with chromosome 17 sequences of Golden Delicious (Table 3.12).

Table 3.12. Results of BLASTN of 6 Murray BAC clone contigs on the apple genome

Contig name	Length	First 3 sequences producing significant alignments from the result of BLASTN	Score bits	E-value
Contig 1	42302 bp	MDC015069.308 chr7:21533090..21541539	8510	0.0
C1	14205 bp	MDC019637.224 chr3:27865217..27885408	1.15E+04	0.0
C3	1092 bp	MDC010450.933 chr17:24088848..24102477	4105	0.0
C5	18962 bp	MDC010450.952 chr17:24056412..24071813	5981	0.0
C7	46643 bp	MDC001743.356 chr12:18189625..18196951	8324	0.0
C8	6443 bp	MDC016308.423 chr13:25757322..25762075	6181	0.0

Avoiding unspecific primer pairs developed from apple LGs rather than LG17, CGs primer design was changed to search for the sequences with high similarity of LG17 from the six BAC clone contigs. So the chosen sequences were firstly blasted against the apple genome to confirm their location in LG17, and then the primer pairs were designed on flanking sequences. All sequences repeated in the apple genome were discarded for marker design. A total of twelve primer pairs were developed on distinct sequences in the second and third round, but only three SNPs were merely detected in contig C3 and C7 contig, namely C3-1, C3-3 and C7-1 (Table 3.11). However, four unexpected SCAR markers, namely contig1-1, contig1-2, C1-2 and C5-1 were found polymorphic in the process of development specific primer pair on the LG17 (Figure 3.21). These four SCARs were also used for fine-mapping of the *Rvi5* region.

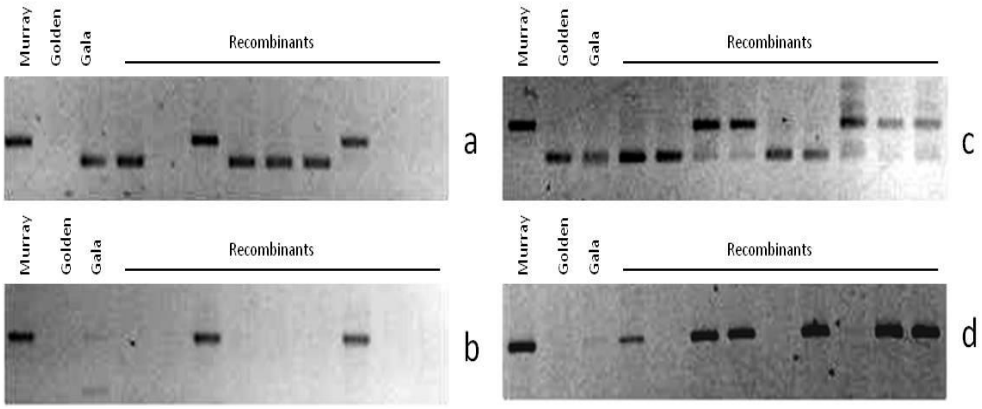


Figure 3.21. Agarose gel profile of segregation pattern of the four new SCAR markers flanking the *Rvi5* region in the 17 recombinants. a Contig 1-1; b Contig 1-2; c C5-1; and d C1-2. The specific Murray bands in the four SCAR markers were in coupling with resistance

Seven TSP primers were designed for SNP mapping and five of them resulted polymorphic and were tested on recombinant plants (Table 3.13). The analysis of the remaining TSP markers from

the C3 contig, named 3-1 and C3-3, was not yet completed.

Since TSP primers designed for Phosphtase 2c, Mlo-like and WRKY possessed a similar annealing temperature, no matter with or without the non-complementary 2-3 bp nucleotide sequence added to the 5' end, the same TSP program named TSP2 was performed (Table 3.14). After preliminary trial on a small number of progeny to confirm the effectiveness of the TSP program used for genotyping SNPs, 17 recombinant seedlings were subjected to the screening with these three TSP primers. A new TSP program was set up for Shaggy, named TSP1 (Table 3.14).

The agarose gel genotyping results (Figure 3.22, 3.23, 3,24 and 3.25) demonstrated that: in Phosphtase 2c, 244 bp allele-specific band harboring the G nucleotide originated from Murray was in coupling with the resistance, while in Mlo-like and WRKY, the allele-specifc bands in coupling with Murray resistance were 243 bp and 635 bp respectively. The Shaggy-like TSP genotyping showed that the 165 bp bands carrying the T nucleotide is in coupling with resistance. The scoring dataset of 17 recombinants with these three SNP markers showed that Phosphatase 2c, Mlo-like and Shaggy-like were co-segregating due to no recombinants found between them, whereas four informative recombinants (Q2, N47, B39 and O30) were scored for WRKY which resulted co-segregating with the C13-SSR. Therefore, WRKY is closer to the *Rvi5* locus in respect to the other three SNPs (Table 3.15).

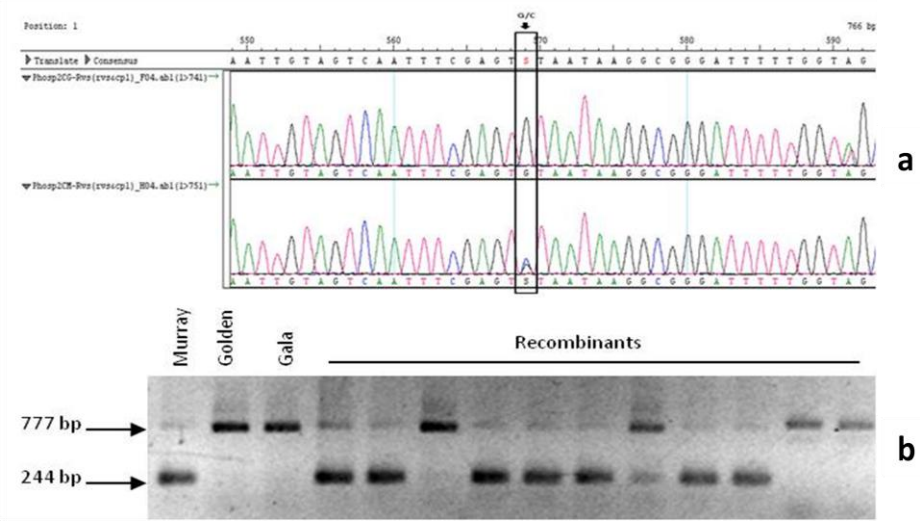


Figure 3.22. a. Portion of the two sequence chromatograms of Murray and Galaxy obtained with direct sequencing of Phosphatase2C amplicon. The G/C polymorphism in Murray is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the 17 recombinants and 3 parents by the TSP method, the 244 bp band is the polymorphic band carrying the C nucleotide.

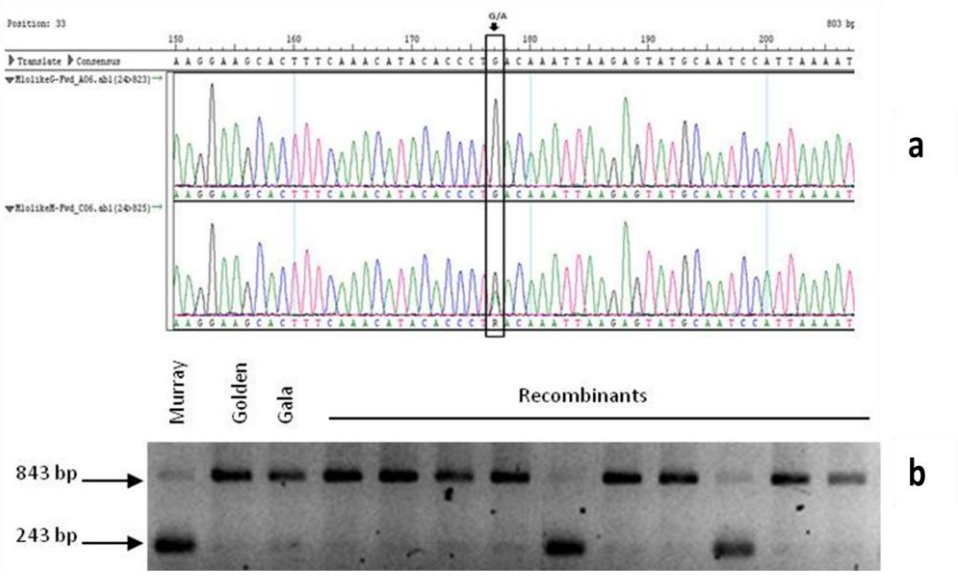


Figure 3.23. a. Portion of the two sequence chromatograms of Murray and Galaxy obtained with direct sequencing of Mlolike amplicon. The G/A polymorphism in Murray is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the 17 recombinants and 3 parents by the TSP method, the 243 bp band is the polymorphic band carrying the A nucleotide.

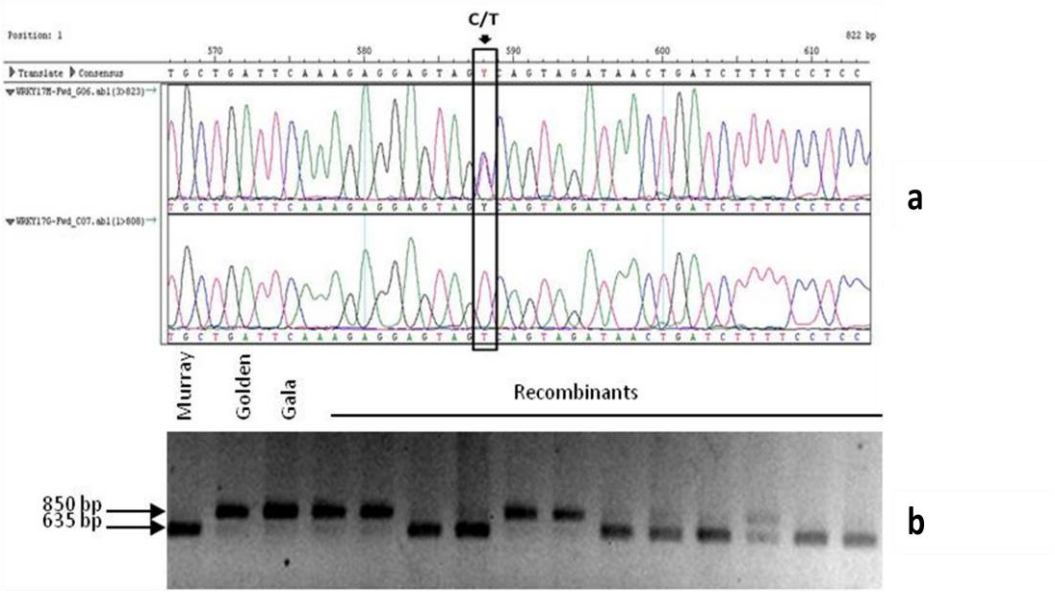


Figure 3.24. a. Portion of the two sequence chromatograms of Murray and Galaxy obtained with direct sequencing of WRKY amplicon. The C/T polymorphism in Murray is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the 17 recombinants and 3 parents by the TSP method, the 635 bp band is the polymorphic band carrying the C nucleotide.

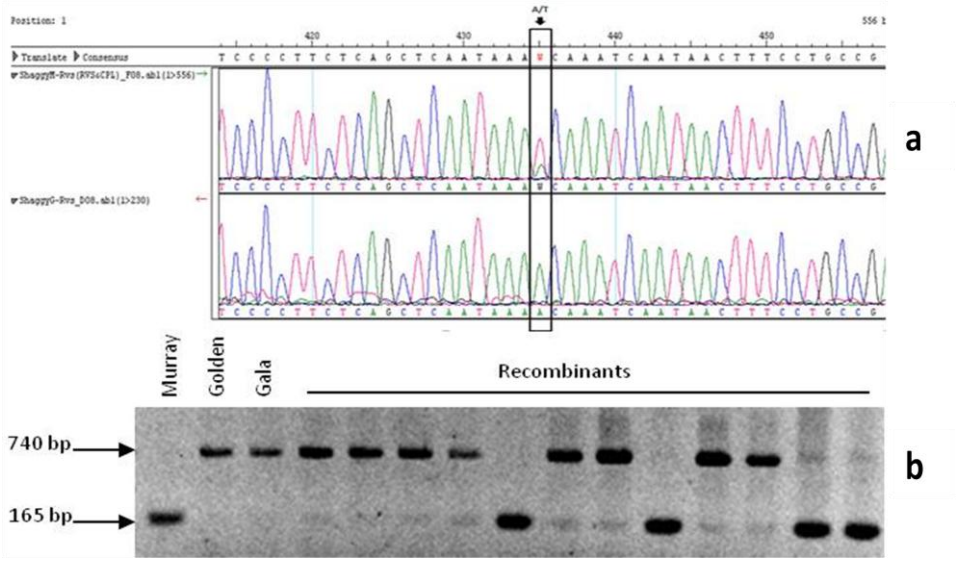


Figure 3.25. a. Portion of the two sequence chromatograms of Murray and Galaxy obtained with direct sequencing of Shaggy-like-Ser amplicon. The A/T polymorphism in Murray is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the 17 recombinants and 3 parents by the TSP method, the 165 bp band is the polymorphic band carrying the T nucleotide.

For the C7-1 SNP genotyping, initial TSP program was set up with 5 s 72 ℃ in the third round cycle to perform the amplification of allele-specific band. But loss of 951 bp bands in the agarose gel suggested that the specific allele containing the SNP was not successfully amplified. Considering the relative bigger size of the specific SNP allele (951 bp) in C7-1 SNP genotyping compared to other four SNP alleles (normally less than 500 bp), 5 s more 72 ℃ extension time was add to the third round in the TSP program (Table 3.14). The modified TSP program (TSP 6) was performed successfully to genotyping the C7-1 SNP (Figure 3.26). C7-1 was mapped together with the two SCAR marker Contig1-1 and Contig 1-2 by scoring on the recombinants (Table 3.15).

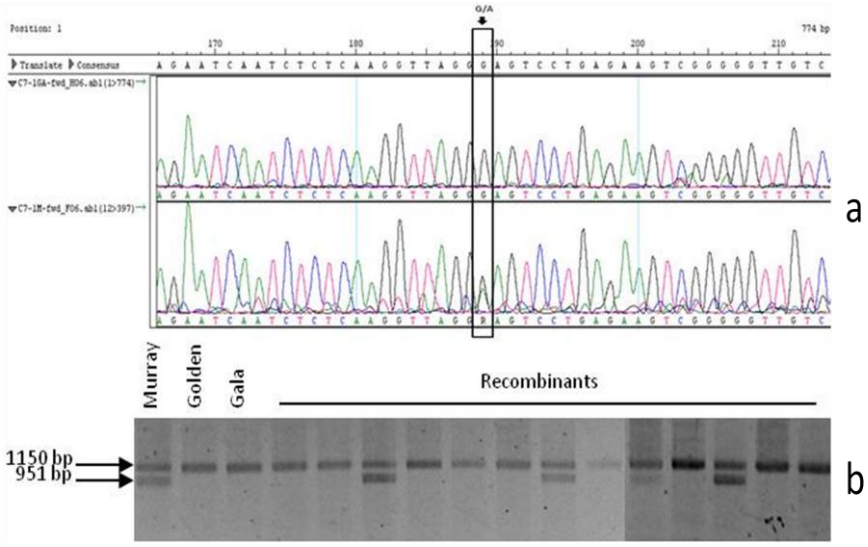


Figure 3,26. a. Portion of the two sequence chromatograms of Murray and Galaxy obtained with direct sequencing of C7-1 amplicon. The G/A polymorphism in Murray is indicated with a box. b. Agarose gel profile of the SNP genotyping results on the 17 recombinants and 3 parents by the TSP method, the 951 bp band is the polymorphic band carrying the A nucleotide.

Finally, the results of the fine mapping of SCAR markers (C1-2 SCAR, C5-1 SCAR, Contig1-1 SCAR and Contig 1-2 SCAR) and C7-1 SNP marker developed from the BAC2 clone are co-segregating and they closer to *Rvi5* than G-SSR (Table 3.15).

Table 3.13. Detailed list of 5 new SNP markers

SNP name	SNP alleles in Murray	SNP alleles in Golden/Gala	SNP allele assayed	Locus-specific primers	Tm (°C)	Reference SNP allele size (bp)	TSP primers[a, b]	Tm (°C)	Alternate SNP allele size (bp)
Shaggy-TSP	T/A=W	A/A	T		62	740	**GG**CTTCTCAGCTCAATAAAT	54	165
Phosphatase2C-TSP	C/G=S	G/G	C		60	777	**CC**GTAGTCAATTTCGAGTC	53	244
Mlolike-TSP	A/G=R	G/G	A	Same as Candidate genes primers	60	843	**CC**GCATACTCTTAATTTGTT	53	243
WRKY-TSP	C/T=Y	T/T	C		60	850	**CG**AAGATCAGTTATCTACTGG	53	635
C7-1-TSP	A/G=R	G/G	A		60	1150	**GG**AATCTCTCAAGGTTAGGA	53	951

[a] **A SNP allele complementary to the allele-specific primer produces the 'reference SNP allele' (smallest) PCR product, whereas those without complementarity produce the 'alternate SNP allele' (largest) product.**

[b] **The arbitrary 5' -nucleotide sequence indicated in underlined, bold font is non-complementary to the target DNA**

Underlined alleles were found to be in coupling with the scab resistance

Table 3.14. Optimized TSP programs for each SNP marker

TSP program name		TSP programs										SNP name
TSP 1	Temperature	95 °C	94 °C	62 °C	72 °C	94 °C	45 °C	94 °C	54 °C	72 °C	16 °C	Shaggy-TSP
	Time	10 min	30 s	30 s	1min 30 s	10 s	30 s	30 s	30 s	5 s	∞	
	Cycles			20		5			10			
TSP 2	Temperature	95 °C	94 °C	60 °C	72 °C	94 °C	42 °C	94 °C	53 °C	72 °C	16 °C	Phosphatase2C-TSP
	Time	10 min	30 s	30 s	1min 30 s	10 s	30 s	30 s	30 s	5 s	∞	Mlolike-TSP
	Cycles			20		5			10			WRKY-TSP
TSP 6	Temperature	95 °C	94 °C	60 °C	72 °C	94 °C	48 °C	94 °C	53 °C	72 °C	16 °C	
	Time	10 min	30 s	30 s	1min 30 s	10 s	30 s	30 s	30 s	10 s	∞	C7-1-TSP
	Cycles			20		5			10			

Table 3.15. Mapping of 4 SCAR markers and 5 SNP markers developed in this study through the segregation pattern observed in the *Rvi5*-region recombinant seedlings

Recombinants	C1-2	G	C5-1	Contig1-1	Contig1-2	C7-1	*Rvi5*	C13	WRKY	Mlo-like	Phosph2C	Shaggy
	SCAR	SSR	SCAR	SCAR	SCAR	SNP		SSR	SNP	SNP	SNP	SNP
H05	R	R	R	R	R	R	R	R	R	R	R	R
Z33	R	R	R	R	R	R	R	R	R	R	R	R
A05	S	S	S	S	S	S	S	S	S	S	S	S
A19	S	S	S	S	S	S	S	S	S	S	S	S
C16	S	S	S	S	S	S	S	S	S	S	S	S
G14	S	S	S	S	S	S	S	S	S	S	S	S
G15	S	S	S	S	S	S	S	S	S	S	S	S
Q22	S	S	S	S	S	S	S	S	S	R	R	R
N47	S	S	S	S	S	S	S	S	S	R	R	R
O32	S	S	S	S	S	S	S	R	R	R	R	R
S01	R	R	R	R	R	R	R	R	R	R	R	R
B39	R	R	R	R	R	R	R	R	R	S	S	S
O30	R	R	R	R	R	R	R	R	R	S	S	S
O13	R	R	R	R	R	R	R	S	S	S	S	S
R20	R	R	R	R	R	R	S	S	S	S	S	S
M42_pato	R	S	S	S	S		S	S	S	S	S	S

In order to verify the specificity of markers for the Rvi5 region, eleven reference cultivars (Golden Delicious, Galaxy, Delicious, Braeburn, Cox, F2-26829-2-2, Granny Smith, Jonathan, McIntosh and Durello di Forli and Fiesta) were also genotyped by the newly developed four candidate genes based SNP markers (Phosphatase 2c, Mlo-like, WRKY and Shaggy-like) (Figure 3.27). Interestingly, none of the reference cultivars was carrying the Murray alleles, suggesting that these new markers are highly specific to Murray.

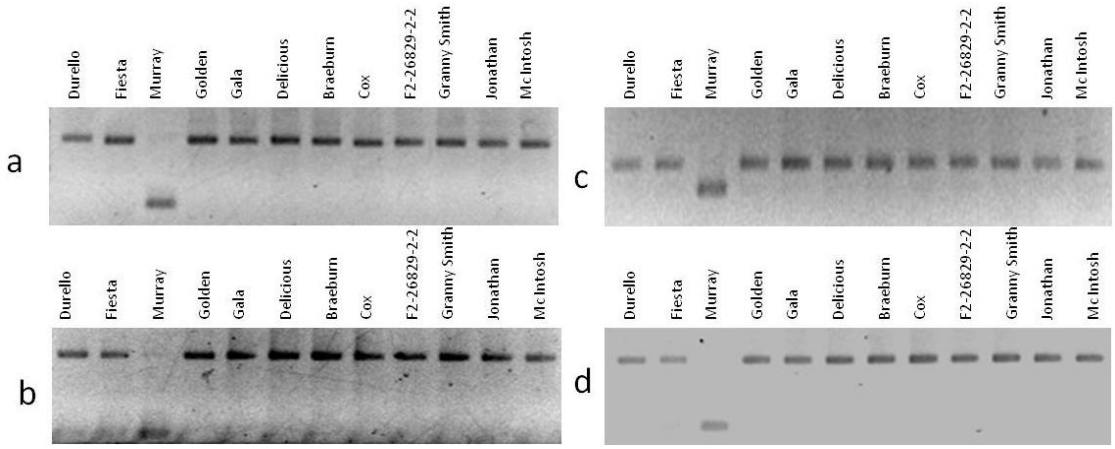


Figure 3.27. Agarose gel profile of genotyping results of 11 reference cultivars with the 4 new candidate gene based SNP markers. a Phosphtase 2c, reference cultivars were CC, only Murray was GC; b. Mlolike and c. WRKY, reference cultivars were TT, only Murray was TC; d. Shaggy, reference cultivars were AA, only Murray was TA.

3.5 Conclusions

3.5.1. Fine mapping of the *Rvi13*

Based on the apple genome sequencing in combination with the synteny analysis with the peach genome, nine new polymorphic SSR markers were identified and mapped on the top of the LG10 in the Durello di Forli x Fiesta progeny comprising 381 seedlings. This region is responsible for *Rvi13* apple scab resistance as evinced by the major QTL previously identified (Tartarini et al., 2004). The fine-mapping improved the definition of the QTL peak in which a group of SSR markers are co-segregating (DCA10SSR2, DCA10SSR3, DCA10SSR17, DCA10SSR22 and DCA10SSR26). A second smaller peak was also identified close to the Hi02d04 SSR. Interestingly, for each QTL a putative gene-specific marker was mapping exactly under the peak, providing a strong evidence for the implication of these genes in conferring *Rvi13* scab resistance. In particular, the MDP0000637744 which is codifying for a protein with an LRR/NBS domains is probably the best candidate gene. In fact, the nucleotide-binding site (NBS)/leucine-rich repeat (LRR) class of receptor-like gene is by far the largest class of known resistance genes in plants. Then, ARGH25 is also a resistance gene analog as already described by Baldi et al., (2004). In addition, availability of the apple genome made it possible to identify directly the sequences of some candidate genes that could be involved in the resistance response. Among these genes, there are a glucanase, a PCD-inhibitor, a transcription co-repressor, and a gene coding for a heat shock protein. The SSR and SNP markers identified in this study under the major QTL were reliable and are ready to use for MAS. The optimized TSP protocols provide an opportunity for cheap SNP genotyping simply by agarose gel electrophoresis, making the SNP genotyping easy to deploy. Once the 5 candidate gene based-SNPs are applied to the SNP array, high throughput protocols can be adopted in order to speed up the marker assisted selection towards *Rvi13*

scab resistance in apple. Finally, the fine mapping of markers tightly-linked to *Rvi13* scab resistance, evidenced discrepancies with the published GD apple genome sequence that can be used for a more precise assembly in this specific genomic region.

3.5.2. Fine mapping of the *Rvi5*

In this work, five SNPs were identified through re-sequencing of four putative resistance related genes and 1 anonymous-PCR products on the basis of the available Murray BAC clone sequences partially covering the *Rvi5* region. The five SNPs were mapped very close to the *Rvi5* locus together with additional four SCAR markers. The analysis on seventeen recombinants previously identified in more than 1200 seedlings derived from Murray, the *Rvi5* donor of resistance, made it possible to make a very precise fine-mapping of the *Rvi5* region now highly saturated with molecular markers. The *Rvi5* gene is now bracketed by two SNP markers (C7-1 SNP and WRKY) and only three recombinant plants can be used for the gene identification. The presence of recombinants between the markers designed on candidate genes and the *Rvi5* position suggests that these genes are not the main responsible of this resistance.

These markers are directly suitable for MAS because they were found to be highly specific for the Murray chromosome carrying the *Rvi5* gene. For that they can be promptly used for MAS both by using standard gel-based protocols or high throughput approaches. The availability of several tightly-linked markers is a pre-requisite for the selection of resistant plants with a reduced linkage drag.

3.6. References

Andersson-Eklund, L., L. Marklund, K. Lundström, C. Haley, K. Andersson, I. Hansson, M. Moller and L. Andersson (1998). "Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross." Journal of Animal Science 76(3): 694-700.

Belfanti, E., Silfverberg-Dilworth, E., Tartarini, S., Patocchi, A., Barbieri, M.; Zhu, J., Vinatzer, B., Gianfranceschi, L., Gessler, C., Sansavini, S. 2004. The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. PNAS 101 (3): 886-890

Botstein, D., R. L. White, M. Skolnick and R. W. Davis (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." American journal of human genetics 32(3): 314.

Brown AG (1975) Apple: In: Janick J, Moore JN (eds) Advances in Fruit Breeding. Purdue University Press, West Lafayette, IN, USA, PP 3-37

Bus, V., W. Van de Weg, C. Durel, C. Gessler, F. Calenge, L. Parisi, E. Rikkerink, S. Gardiner, A. Patocchi and M. Meulenbroek (2004 a). "Delineation of a scab resistance gene cluster on linkage group 2 of apple." Acta Hortic 663: 57-62.

Bus, V. G., F. N. Laurens, V. De Weg, W. Eric, R. L. Rusholme, E. H. Rikkerink, S. E. Gardiner, H. Bassett, L. P. Kodde and K. M. Plummer (2005 b). "The Vh8 locus of a new gene-for-gene interaction between Venturia inaequalis and the wild apple Malus sieversii is closely linked to the Vh2 locus in Malus pumila R12740-7A." New Phytologist 166(3): 1035-1049.

Calenge, F., A. Faure, M. Goerre, C. Gebhardt, W. Van de Weg, L. Parisi and C.-E. Durel (2004). "Quantitative trait loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of Venturia inaequalis." Phytopathology 94(4): 370-379.

Challice, J. (1974). "Rosaceae chemotaxonomy and the origins of the Pomoideae." Botanical Journal of the Linnean Society 69(4): 239-259.

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, et al. (2012) Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. PLoS ONE 7(2): e31745

Churchill, G. A. and R. W. Doerge (1994). "Empirical threshold values for quantitative trait mapping." Genetics 138(3): 963-971.

Conner, P. J., S. K. Brown and N. F. Weeden (1997). "Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars." Journal of the American Society for Horticultural Science 122(3): 350-359.

Crute, I. R. and D. Pink (1996). "Genetics and utilization of pathogen resistance in plants." The Plant Cell 8(10): 1747.

Deslandes, L., J. Olivier, F. Theulières, J. Hirsch, D. X. Feng, P. Bittner-Eddy, J. Beynon and Y. Marco (2002). "Resistance to Ralstonia solanacearum in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes." Proceedings of the National Academy of Sciences **99**(4): 2404-2409.

Deslandes, L., J. Olivier, N. Peeters, D. X. Feng, M. Khounlotham, C. Boucher, I. Somssich, S. Genin and Y. Marco (2003). "Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus." Proceedings of the National Academy of Sciences **100**(13): 8024-8029.

N. E. N. Erdin, S. T. S. Tartarini, G. B. G. Broggini, F. G. F. Gennari, S. S. S. Sansavini, C. G. C. Gessler and A. P. A. Patocchi (2006). "Mapping of the apple scab-resistance gene Vb." Genome 49(10): 1238-1245.

Durel, C., F. Calenge, L. Parisi, W. Van de Weg, L. Kodde, R. Liebhard, C. Gessler, M. Thiermann, F. Dunemann and F. Gennari (2004). "An overview of the position and robustness of scab resistance QTLs and major genes by aligning genetic maps of five apple progenies." Acta Hort (ISHS) 663: 135-140.

Durel, C., L. Parisi and F. Laurens (1999). Does the Vf gene maintain a residual resistance to apple scab despite its breakdown by Venturia inaequalis race 6 strains? Eucarpia symposium on Fruit Breeding and Genetics 538.

Durel, C., L. Parisi, F. Laurens, W. Van de Weg, R. Liebhard and M. Jourjon (2003). "Genetic dissection of partial resistance to race 6 of Venturia inaequalis in apple." Genome 46(2): 224-234.

Edwards, D., J. W. Forster, D. Chagné and J. Batley (2007). "What Are SNPs?" Association mapping in plants: 41-52.

Evans, R. C. and C. S. Campbell (2002). "The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes." American Journal of Botany 89(9): 1478-1484.

Geldermann, H. (1975). "Investigations on inheritance of quantitative characters in animals by gene markers I. Methods." TAG Theoretical and Applied Genetics 46(7): 319-330.

Gygax, M., L. Gianfranceschi, R. Liebhard, M. Kellerhals, C. Gessler and A. Patocchi (2004). "Molecular markers linked to the apple scab resistance gene Vbj derived from Malus baccata jackii." TAG Theoretical and Applied Genetics 109(8): 1702-1709.

Hackett, C. A. (2002). "Statistical methods for QTL mapping in cereals." Plant molecular biology 48(5): 585-599.

Hayden, M. J., T. Tabone and D. E. Mather (2009). "Development and assessment of simple PCR markers for SNP genotyping in barley." TAG Theoretical and Applied Genetics 119(5): 939-951.

Hemmat, M., N. Weedon, A. Manganaris and D. Lawson (1994). "Molecular marker linkage map for apple." Journal of heredity 85(1): 4-11.

Janick, J. and J. N. Moore (1975). Advances in fruit breeding, Purdue University Press.

Jourjon, M., C. Durel, B. Goffinet and F. Laurens (2000). An example of application for MCQTL software: fine characterisation of a QTL for apple scab resistance. Proceedings of the Plant and Animal Genome Conference VIII, San Diego, Calif. Available at http://www. intl-pag. org/pag/8/abstracts/pag8663. html.

Karp, A., K. J. Edwards, M. Bruford, S. Funk, B. Vosman, M. Morgante, O. Seberg, A. Kremer, P. Boursot and P. Arctander (1997). "Molecular technologies for biodiversity evaluation: opportunities and challenges." Nature biotechnology 15(7): 625-628.

Khan M.A., Han Y., Zhao Y.F., Korban S.S. 2011. A high-throughput apple SNP genotyping platform using the GoldenGate™ assay. Gene494:196-201

Korban, S. (1986). "Interspecific hybridization in Malus." HortScience 21.

Korban, S., H. Chen, F. Hammerschlag and R. Litz (1992). "Apple." Biotechnology of perennial fruit crops.: 203-227.

Lander, E. S. and D. Botstein (1989). "Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps." Genetics 121(1): 185-199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, L. A. Newberg and L. Newburg (1987). "MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations." Genomics 1(2): 174.

Levinson, G. and G. A. Gutman (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." Molecular biology and evolution 4(3): 203-221.

Liebhard, R., M. Kellerhals, W. Pfammatter, M. Jertmini and C. Gessler (2003 a). "Mapping quantitative physiological traits in apple (Malus× domestica Borkh.)." Plant molecular biology 52(3): 511-526.

Liebhard, R., B. Koller, L. Gianfranceschi and C. Gessler (2003 b). "Creating a saturated reference map for the apple (Malus× domestica Borkh.) genome." TAG Theoretical and Applied Genetics 106(8): 1497-1508.

Liebhard, R., B. Koller, A. Patocchi, M. Kellerhals, W. Pfammatter, M. Jermini and C. Gessler (2003 c). "Mapping quantitative field resistance against apple scab in a'Fiesta'×'Discovery'progeny." Phytopathology 93(4): 493-501.

Lincoln, S., M. Daly and E. Lander (1993). "Mapping genes controlling quantitative traits using MAPMAKER/QTL version 1.1: a tutorial and reference manual." Whitehead Inst., Cambridge, MA.

Liu, B. H. (1997). Statistical genomics: linkage, mapping, and QTL analysis, CRC.

Maliepaard, C., F. Alston, G. Van Arkel, L. Brown, E. Chevreau, F. Dunemann, K. Evans, S. Gardiner, P. Guilford and A. Van Heusden (1998). "Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers." TAG Theoretical and Applied Genetics 97(1): 60-73.

Manly, K. F., J. Cudmore, Robert H and J. M. Meer (2001). "Map Manager QTX, cross-platform software for genetic mapping." Mammalian Genome 12(12): 930-932.

Michelmore, R. (1995). "Molecular approaches to manipulation of disease resistance genes." Annual review of phytopathology 33(1): 393-427.

Michelmore, R. W. and B. C. Meyers (1998). "Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process." Genome Research 8(11): 1113-1130.

Nelson, J. C. (1997). "QGENE: software for marker-based genomic analysis and breeding." Molecular breeding 3(3): 239-245.

Parisi, L., Y. Lespinasse, J. Guillaumes and J. Krüger (1993). "A new race of Venturia inaequalis virulent to apples with resistance due to the Vf gene." Phytopathology 83(5): 533-537.

Parlevliet, J. and J. Zadoks (1977). "The integrated concept of disease resistance: A new view including horizontal and vertical resistance in plants." Euphytica 26(1): 5-21.

Patocchi, A., B. Bigler, B. Koller, M. Kellerhals and C. Gessler (2004). "Vr 2: a new apple scab resistance gene." TAG Theoretical and Applied Genetics 109(5): 1087-1092.

Patocchi, A., M. Walser, S. Tartarini, G. A. Broggini, F. Gennari, S. Sansavini and C. Gessler (2005). "Identification by genome scanning approach (GSA) of a microsatellite tightly associated with the apple scab resistance gene Vm." Genome 48(4): 630-636.

Primmer, C., A. Møller and H. Ellegren (1996). "A wide-range survey of cross-species microsatellite amplification in birds." Molecular Ecology 5(3): 365-378.

Ribaut, J.-M., C. Jiang, D. Gonzalez-de-Leon, G. Edmeades and D. Hoisington (1997). "Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies." TAG Theoretical and Applied Genetics 94(6): 887-896.

Roberts, A. and I. Crute (1994). "Apple scab resistance from Malus floribunda 821 (Vf) is rendered ineffective by isolates of Venturia inaequalis from Malus floribunda." Norwegian Journal of Agricultural Sciences.

Robinson, J., S. Harris and B. Juniper (2001). "Taxonomy of the genus Malus Mill.(Rosaceae) with emphasis on the cultivated apple, Malus domestica Borkh." Plant Systematics and Evolution 226(1): 35-58.

Sansavini, S., S. Tartarini and F. Gennari (1999). Scab-tolerance segregation from two Italian apple germplasm cultivars to increase durability of resistance. Eucarpia symposium on Fruit Breeding and Genetics 538.

Sansavini, S., B. Vinatzer, M. Barbieri and S. Tartarini (2000). Advances in Marker-Assisted Selection for Scab Resistance in Apple and Cloning of the Vf Gene. International Symposium on Apple Breeding for Scab Resistance 595.

Silfverberg-Dilworth, E., C. Matasci, W. Van de Weg, M. P. Van Kaauwen, M. Walser, L. Kodde, V. Soglio, L. Gianfranceschi, C. Durel and F. Costa (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree Genetics & Genomes 2(4): 202-224.

Tanksley, S. D. (1993). "Mapping polygenes." Annual review of genetics 27(1): 205-233.

Tartarini, S., F. Gennari, D. Pratesi, C. Palazzetti, S. Sansavini, L. Parisi, A. Fouillet, V. Fouillet and C. Durel 2004). "Characterisation and genetic mapping of a major scab resistance gene from the old Italian apple cultivar 'Durello di Forlì'." Acta Hort.(ISHS) 663: 129-133.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17, 6463-6471.

Tautz, D. and M. Renz (1984). "Simple sequences are ubiquitous repetitive components of eukaryotic genomes." Nucleic Acids Research 12(10): 4127-4138.

Van Ooijen, J. (2004). "MapQTL® 5." Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen.

Velasco, R., A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S. K. Bhatnagar, M. Troggio and D. Pruss (2010). "The genome of the domesticated apple (Malus [times] domestica Borkh.)." Nature genetics 42(10): 833-839.

Vinatzer, B. A., Patocchi, A., Gianfranceschi, L., Tartarini, S., Zhang, H-B., Gessler, C., and Sansavini, S. 2001. Apple (Malus sp.) contains receptor-like genes homologous to the Cf resistance gene family of tomato with a cluster of such genes co-segregating with Vf apple scab resistance. Mol. Plant-Microbe Interact. 14: 508–515.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. v. d. Lee, M. Hornes, A. Friters, J. Pot, J. Paleman and M. Kuiper (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic acids research 23(21): 4407-4414.

Way RD, Aldwinckle HS, Lamb RC, Rejman A, Sansavini S, Shen T, Watkins R, Westwood MN, Yoshida Y (1989) Apples (*Malus*). Acta Hort 290:1-62

Wang, D.G., Fan, J.B., Siao, C.J., et al., Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome, Science, 1998, vol. 280, pp. 1077–1082.

Welsh, J. and M. McClelland (1990). "Fingerprinting genomes using PCR with arbitrary primers." Nucleic acids research 18(24): 7213-7218.

Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey (1990). "DNA polymorphisms amplified by arbitrary primers are useful as genetic markers." Nucleic acids research 18(22): 6531-6535.

Young, N. (1996). "QTL mapping and quantitative disease resistance in plants." Annual review of phytopathology 34(1): 479-501.