GENOMIC AND TRANSCRIPTIONAL ANALYSIS
OF ALLERGEN GENES IN APPLE
(Malus x domestica)

presentata da

Giulia Pagliarani

Co-relatrori:
Dr. ROBERTA PARIS
Dr. STEFANO TARTARINI

Relatore:
Chiar.mo Prof. SILVIERO SANSAVINI

Coordinatore:
Prof. LUCA CORELLI GRAPPADELLI

Ciclo XXII

Anno Accademico 2009-2010
Ad Alfonso e alle sue fragole
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>General introduction</td>
<td>1</td>
</tr>
<tr>
<td>Aim of the thesis</td>
<td>15</td>
</tr>
<tr>
<td><strong>Chapter 1 -</strong> Genomic organization of the <em>Mal d 1</em> gene cluster on LG 16</td>
<td>17</td>
</tr>
<tr>
<td><strong>Chapter 2 -</strong> Development of a PCR-based tool for expression analysis of 20 individual <em>Mal d 1</em> genes</td>
<td>49</td>
</tr>
<tr>
<td><strong>Chapter 3 -</strong> Specific <em>Mal d 1</em> genes expression analysis on apple fruit tissues</td>
<td>73</td>
</tr>
<tr>
<td><strong>Chapter 4 -</strong> Specific <em>Mal d 1</em> genes (PR10) expression analysis on apple leaves</td>
<td>93</td>
</tr>
<tr>
<td><strong>Chapter 5 -</strong> Identification and mapping of new <em>Mal d 2</em> and <em>Mal d 4</em> genes</td>
<td>113</td>
</tr>
<tr>
<td>General conclusions</td>
<td>131</td>
</tr>
<tr>
<td>Summary</td>
<td>137</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>139</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>141</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>219</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>221</td>
</tr>
</tbody>
</table>
General introduction

**Apple (Malus x domestica)**

Apple belongs to the genus Malus within the Rosaceae family. This family includes several important genera that are the most important deciduous fruit crops. It include pear (Pyrus communis), peach (Prunus persica), cherry (Prunus avium), plum (Prunus domestica), apricot (Prunus armeniaca), almond (Prunus dulcis), as well as other valuable ornamental plants including rose (Rosa spp.), medlar (Mespilus germanica), and hawthorn (Crataegus monogyna). Among these various genera, Malus serves as the most commercially valuable in temperate climate regions of the world.

The allopolyploid origin for all the Maloideae species was proposed (Sax, 1933). In particular, it has been suggested that they derived from the genome duplication of the hybrid between the Spiraeoideae (x=9) and Prunoideae (x=8) subfamilies. The origin of the cultivated apple is likely the results of this interspecific hybridization, so the binomial Malus x domestica Borkh. has generally been accepted to underline its hybrid origin. Up to now, there is no conclusive answer to the origin of the domesticated apple but the presence of multiple loci for the same molecular markers in genetic maps revealed the presence of several homeologous linkage groups (LGs) such as linkage group pairs 2-7, 4-12, 5-10 and 13-16 (Maliepaard et al., 1998; Liebhard et al., 2002; 2003a; 2003b; Gao et al., 2005a, b, c). These duplicated markers directly support the hybrid origin of the cultivated apple. Moreover, it is known that this genome duplication process suffered by apple and the other Maloideae species, may have activated retrotransposons (Madlung et al. 2005) for which an important role also in the apple genome evolution have been proposed (Sun et al., 2008). In fact, the expansion of particular loci and clusters has been partly attributed to the insertion of retrotransposons (Sassa et al., 2007). The majority of apple cultivars is diploid (2n = 34), self-incompatible and with an highly heterozygous genome.

Apple is also an important model species for functional genomics research among woody perennial angiosperms due to its relative small genome size, 1.54 pg DNA/2C nucleus or 750 Mb per haploid genome, which is similar to that of the sorghum (Sorghum bicolor) genome and about the same size as the tomato (Solanum lycopersicum) genome (Gasic et al., 2009). Several efforts are under way to develop a substantial resource of molecular markers for genotyping and marker-assisted selections (MAS) and the apple molecular maps currently have more than 1200 markers distributed on the 17 LGs (Maliepaard et al., 1998; Liebhard et al., 2002, 2003a, 2003b; Xu and Korban, 2002; Silfverberg-Dilworth et al., 2006; Naik et al., 2006; Chagné et al., 2008; Celton et al., 2009). Moreover, identifying molecular markers linked to major genes/quantitative trait loci (QTL) contributing to desirable economic traits has become an important goal in apple genetics studies. Other genetic tools such as bacterial artificial chromosome (BAC) libraries and expressed sequence tags (ESTs) have been recently developed. At least three bacterial artificial chromosome (BAC) libraries have been constructed from different genotypes.
General introduction

(Vinatzer et al., 1998; Xu et al., 2001). These BAC libraries have been successfully used for cloning genes of interest (Vinatzer et al., 2001; Xu and Korban, 2002; Han et al., 2007). In a recent analysis of high-quality apple ESTs obtained from different tissues, under different conditions, and from different genotypes, the total number of apple unigenes obtained was ~33,000 (Gasic et al., 2009), comparable to that previously estimated by Neucomb et al. (2006). A more accurate estimate of the total number of genes in the apple genome can be made by comparing the size of the EST-derived unigenes set and the percentage of predicted genes in genomic DNA (e.g., through BAC sequences) that are represented by a unigene match. Computational comparison of apple unigenes against the databases have allowed the assignment of putative functional roles for about 80% of the transcripts. The remaining 20% of sequences, having no matches to any sequences in public databases, may represent apple specific genes or genes most likely associated with tree formation (Gasic et al., 2009). The most abundant protein families represented within the apple unique sequences resulted the protein kinases followed by leucine-rich repeat family and Tyr protein kinases. Within transcription factors, the MYB transcription factor family was the most common in apple sequences. Interestingly, also two apple allergens families resulted in the fifty most common apple protein families: the Bet v 1-like (PR-10) and the Thaumatin-like (PR-5) proteins (Gasic et al., 2009). Similarly to tomato (van der Hoeven et al., 2002), the majority of apple unigenes (80%) with no matches in the Arabidopsis genome have unknown functions and they have no matches in other genome databases. Hence, Gasic et al., (2009) suggested that these may represent fast-evolving genes that have aquired new functions in apple and related taxa. The majority of these novel genes are confined to apple and to other species of the Rosaceae family. Among them, there is the Mald 1 gene family.

Recently, a first draft of the physical map of the genome has been reported by Han et al. (2007) and efforts are underway to anchor this physical map to the genetic map. In addiction, the whole apple genome sequence will be soon published (Shulaev et al., 2008; R. Velasco, Istituto Agrario di San Michele All’Adige, Italy, personal communication).

Food allergy

Food allergy is an hypersensitive reaction to normally harmless substances (allergens, mostly proteins) and involves humoral immune responses, mediated by immunoglobuline (Ig) E synthesized by B-lymphocytes (Bohle, 2004). During sensitization, the immune system produces specific IgE antibodies against food allergens. The IgE antibodies bind to high affinity receptors an the surface of mast cells and basophiles. Upon re-exposure to the ingested food, cross-linking of the IgE antibodies with the allergens triggers the release of mediators (i.e. histamine) causing the acute phase of the allergic reactions. Two classes of food allergy are reported. The class-I food allergy, in which the immune reaction takes place in the gastrointestinal tract, is especially found in children. The allergens involved in this class are very stable and resistant to heat and digestion.
processes. A class-II food allergy is initiated by proteins that come into contact with the immune system through inhalation; i.e. the inhalation of pollen from several tree species and grasses. Food allergy symptoms are usually mild local reactions in the oral cavity, the so-called oral allergy syndrome (OAS). Especially, fresh fruits and vegetables can cause such allergy problems. Other symptoms are also visible in organs like skin (urticaria, atopic eczema), gastro-intestinal tract (cramps, diarrhoea, vomiting) nose and lungs (rhinitis and asthma) and cardiovascular system (anaphylactic shock) (Fernandez-Rivas and Miles, 2004). IgE-mediated food allergy is a disease affecting all age groups and because the only treatment is still avoidance, it can affect the quality of life in a profoundly negative way (Mills et al., 2007). Figures coming from some clinical studies estimate that food allergy affects almost 4% of the adult population in the United States (Sampson et al., 2005), and about 2–4% of European adults (Bruijnzeel-Koomen et al., 1995).

Allergens are named according to the source material, with the first three letters of the genus and the first letter of the species name (King et al., 1995). For instance, “Mal d” represents apple (*Malus x domestica*) allergens. Different allergens are designed by Arabic numbers according to the time of their identifications. Additional numbers and letters can be added to distinguish isoallergens and variants. The plain text refer to proteins, italics refer to genes. For example, Mal d 1 refers to first apple allergen and *Mal d 1* refers to this gene class.

**Allergy cross-reactivity**

Allergic cross-reactivity can occur in a patient reacting to similar allergens from different origins through the same IgE antibody type (Jenkins et al., 2005) since most food allergens belong to a limited number of protein families and show similarities in primary and/or tertiary structure across different sources (Breiteneder and Mills, 2006). In fact, IgE antibodies are not necessarily specific to a unique allergen. They will bind the allergen that initiated the IgE production by the immune system (primary sensitization) with the highest affinity, but they can also bind structurally similar allergens. The major birch pollen allergen, Bet v 1, is the most relevant sensitizing protein causing this type of food allergy (Wensing et al., 2002) but minor allergens such as Bet v 2, Bet v 5 and Bet v 6 have also been shown to be involved (Karamloo et al., 2001). Bet v 1 belongs to the PR-10 family (Breiteneder and Radauer, 2004). Other members of this protein family are present in various foods, such as fruits of *Rosaceae* (e.g. Mal d 1 in apple, Pru a 1 in cherry, Pyr c 1 in pear), vegetables of *Apiaceae* (e.g. Api g 1 in celery, Dau c 1 in carrot), hazelnut (Cor a 1), soybean (Gly m 4), mungbean (Vig r 1) and peanut (Ara h 8). Thus, Bet v 1-specific IgE antibodies can bind to these dietary proteins which may cause immediate hypersensitivity reactions upon consumption of the respective foods. Hence, Bet v 1-specific IgE antibodies react preferentially with allergens in *Rosaceae* fruits sharing between 56% and 59% of amino acid
similarity with Bet v 1 and less frequently with homologues in vegetables of the Apiaceae family sharing 37–41% (De Amici et al., 2002).

**Fruit allergy**

Fruits of the *Rosaceae* family, especially apple and peach, are reported as the plant food most frequently involved in allergic reactions (Fernandez-Rivas et al., 2006). The major allergens from *Rosaceae* belong to few protein families as Pathogenesis Related proteins of class 10 (PR-10), Lipid Transfer Proteins (LTP), Thaumatine-like proteins (TLP) or profilins. The official list of allergens in *Rosaceae* fruits according to the Allergen Nomenclature Sub-Committee (http://www.allergen.org) are shown in Table 1.

<table>
<thead>
<tr>
<th>Fruit</th>
<th>Protein Family</th>
<th>PR-10</th>
<th>TLP (PR-5)</th>
<th>LTP (PR-14)</th>
<th>Profilins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple (<em>Malus x domestica</em>)</td>
<td>Mal d 1</td>
<td>Mal d 2</td>
<td>Mal d 3</td>
<td>Mal d 4</td>
<td></td>
</tr>
<tr>
<td>Peach (<em>Prunus persica</em>)</td>
<td>Pru p 1</td>
<td>Pru p 3</td>
<td>Pru p 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweet cherry (<em>Prunus avium</em>)</td>
<td>Pru av 1</td>
<td>Pru av 2</td>
<td>Pru av 3</td>
<td>Pru av 4</td>
<td></td>
</tr>
<tr>
<td>Apricot (<em>Prunus armeniaca</em>)</td>
<td>Pru ar 1</td>
<td>Pru ar 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pear (<em>Pyrus communis</em>)</td>
<td>Pyr c 1</td>
<td>Pyr c 3</td>
<td>Pyr c 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European plum (<em>Prunus domestica</em>)</td>
<td>Pru d 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strawberry (<em>Fragaria ananassa</em>)</td>
<td>Fra a 1</td>
<td>Fra a 3</td>
<td>Fra a 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red raspberry (<em>Rubus idaeus</em>)</td>
<td>Rub i 1</td>
<td>Rub i 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the clinical relevance and prevalence, major and minor allergens are classified for different geographical areas (Andersen et al., 2009). In most cases, the fruit allergy originates after sensitization with a pollen source and a further cross-reaction with allergens from fruits (Asero et al., 2007). They usually involve labile allergens, such as Bet v 1-like allergens, and therefore sensitization is only possible through the respiratory passages (class-II food allergy). This situation is predominant in Central- and Northern Europe. Although fruit allergies of class II are predominant, few fruit allergens capable of causing sensitization independently of pollen allergens (class-I food allergy) has been identified (Salcedo et al., 2007). Allergens involved in this case are resistant to proteolysis, such as LTP, and sensitization is thought to take place in the gastrointestinal tract (Nicoletti et al., 2007). This situation is reported to be predominant in Mediterranean areas. The *Rosaceae* fruit allergy patterns in Italy is a special case, as both the Mediterranean and the Northern and Central European allergy patterns can be found in this country. In fact, the frequency of LTP sensitization was higher in the Southern parts of the country, where pollen-associated fruit allergy was almost absent. Conversely, pollen-associated fruit allergies dominated in the Northern parts of the country (Asero et al., 2009). Allergies of both classes can involve one or more fruits depending on the degree of cross-reactions to other fruits. In addition, a patient can be co-sensitized to more allergens, further increasing
General introduction

the potential number of allergy-causing fruits. Presence of sensitization to one or more allergens in a patient does not necessarily predict clinical relevance, as sensitizations can also be latent (Bousquet et al., 2006). Furthermore, the reported dominant allergy-causing fruits vary between countries. In Northern and Central Europe, allergy to apple is most widespread (Eriksson et al., 2004; Zuberbier et al., 2004), while strawberry and peach most often cause allergy in Southern Europe (Cuesta-Herranz et al., 2000; Chen et al., 2008).

Apple allergy

Apple consumption is highly recommended for a healthy diet because of its efficacy in reducing the risk of stroke, heart disease and lung cancer (Knekt et al., 1996; 2000; Le Marchand et al., 2000) but, unfortunately, it is also one of the most important allergenic fruits. For many reasons apple can nowadays be considered a model food for the study of fruit allergy although it does not rank among the most dangerous allergenic foods like peanut or tree nuts. In fact, apple allergenicity shows different clinical relevance across Europe, and this would allow the availability of enough patients to study. Moreover, apples are important components of a healthy diet, and therefore their avoidance can have a significant negative impact for health. Finally, apple is an important crop for European agriculture and a broad variety of cultivars with a variable degree of allergenicity is grown in various European countries.

Apple allergy is mainly due to the presence of four classes of allergens: Mal d 1, Mal d 2, Mal d 3 and Mal d 4 (http://www.allergen.org). Below these allergen classes are described more in detail.

**Mal d 1** is a multigene family containing 18 different loci mainly clustered on the homeologous linkage groups (LG) 13 and 16; four different sub-families (I - IV) have been described in relation to sequence similarity and presence/length of intron. The comparisons of **Mal d 1** coding sequences has revealed different levels of identity: 71 - 83% between sub-families; 86 - 98% within a sub-family; and 98 - 100% between alleles of a single gene (Gao et al., 2005a). **Mal d 1** genes code for a 17 - 18 kDa cytoplasmatic protein of 158 - 159 amino acids (aa), classified as a PR-10 (van Loo et al., 2006). High expression levels of Mal d 1 have been found in ripe apple fruit and in mature leaves (Pühringer et al., 2003). A clear distribution between apple skin and flesh have been reported for Mal d 1 transcripts with higher levels in skin (Pagliarani et al., 2009). The cross-reaction between Mal d 1 and Bet v 1 , due to their high sequence and structural homology, is well reported (Ebner et al., 1991; Vanek-Krebitz et al., 1995). In particular, this allergen is the main responsible for apple allergy in North and Central Europe where many birch pollen sensitised patients (50-70%) suffer from oral allergy symptoms after eating fresh apples (Son et al., 1999). Only a limited number of Mal d 1 proteins and mRNAs have been traced back in apple fruit so far (Helsper et al., 2002; Pühringer et al., 2003; Beuning et al., 2004, Botton et al., 2008) suggesting that not all Mal d 1 isoallergens are likely to be involved in allergic reactions. Furthermore, it has been demonstrated that different **Mal d 1** isoallergens and variants, as
well as mutants of specific isoallergens, have different binding affinities to IgE (Ma et al., 2006) indicating that the Mal d 1 protein composition, as well as the presence and expression of specific alleles, may have significant and different effects on allergenicity.

While Mal d 2 is the second apple allergen to be identified (Krebitz et al., 2003), its relevance to allergenicity is still unclear. It is an apoplastic, 31 kDa protein of 246 aa encoded by 1,119 - 1,121 nucleotides (nt) organised in two exons (61 and 680 nt) and one intron (378 - 380 nt). Gao et al. (2005b) assumed that the intron size was locus specific and two loci named Mal d 2.01A and Mal d 2.01B, respectively, were both found to be located on LG 9. Other different Mal d 2 ESTs have been retrieved in the database suggesting the presence of more Mal d 2 gene in the apple genome. The N-terminus of the mature protein is about 50% identical to the superfamily of TLPs, also known as PR-5 proteins, with antifungal activity (Krebitz et al., 2003). TLPs contain 16 conserved cysteine residues, forming eight disulphide bonds essential for the overall folding of the protein and possibly for their anti-fungal and allergenic potential (van Loon et al., 2006). This stabilized structure contributes to the protein's resistance to low pH conditions, heat-induced denaturation and proteolysis suggesting for this allergen the ability to sensitize the allergic patient in the gastrointestinal tract (Breiteneder, 2004). Up to now Mal d 2 is known as one of the major protein constituents of mature apple (Oh et al., 2000) with a predominant gene expression in fruit flesh than in skin (Pagliarani et al., 2009).

Mal d 3 is small, apoplastic protein of 9 kDa, belonging to the PR-14 protein family (non-specific LTP) and characterised by high resistance to pepsin hydrolysis and thermal denaturation (van Loon et al., 2006). This allergen is the main responsible of allergic reactions in the Mediterranean areas, where the apple allergy is less frequent but it can provoke more severe symptoms (Diaz-Perales et al., 2002). Gao et al. (2005c) described two Mal d 3 genomic sequences called Mal d 3.01 and Mal d 3.02 that were mapped on LG 12 and 4, respectively. Both genes contain a single exon of 348 nt, of which the first 72 nt code for a putative signal peptide (Kader, 1996). The two Mal d 3 sequences share 89% identity in their coding sequences, although the similarity was very low in the upstream non coding region. LTPs are mainly found in aerial plant organs like leaves, seeds, flowers, and fruits, with expression being low or even nil in roots, and accumulate preferentially in exposed surfaces such as fruit skin (Borges et al., 2006). It is known that the Mal d 3.01 gene is expressed in apple fruit (Diaz-Perales et al., 2002; Pagliarani et al., 2009).

Mal d 4 is a small (12 - 15 kDa) cytosolic protein belonging to the profilin protein family. Profilins are found in all eukaryotic cells and their allergenic potency has been frequently reported (Asero et al., 2003). Profilins were first described as minor allergens in birch pollen (Bet v 2) and it is known that sensitization to profilin is mediated by pollen (Valenta et al., 1992). There are no reports of IgE antibodies for profilins independent from pollen sensitization. Again, the clinical presentation appears to be different depending on the geographical background of the patients (Ballmer-Weber et al., 2002; Wensing et al.,
Three distinct profilin sequences have been reported in apple, with 75 - 80% identity in both their coding and amino acidic sequences: *Mal d 4.01, Mal d 4.02, and Mal d 4.03* (Gao et al., 2005b). All apple profilins have a coding sequence of 396 nt and two introns of different sizes in conserved positions. *Mal d 4.01* was mapped on LG 9, *Mal d 4.02* on LG 2, and *Mal d 4.03* on LG 8 (Gao et al., 2005b).

Apple allergenicity is influenced both by apple genotype and by many external factors like growing practices, storage conditions and fruit processing (Bolhaar et al., 2005; Botton et al., 2008). At present, the only therapy for allergy is the avoidance of apples and related fruits, even if this deprives the sufferer’s diet of an important source of vitamins, minerals, and fibers. The identification of hypoallergenic apple genotypes and the use of cultural practices that might reduce the amount of allergens in the fresh fruit represent nowadays important goals in the apple growing and breeding.

**PR proteins and plant defense**

Plants possess both preformed and inducible mechanisms to respond to pathogen invasion. Extant morphological barriers, secondary metabolites and antimicrobial proteins must be avoided or overcome by pathogens to be able to invade a plant. Once contact has been established, elicitors produced and released by the pathogen induce further defenses, comprising the reinforcement of cell walls, the production of phytoalexins, and the synthesis of defense-related proteins (Vlot et al., 2008). The term “defense-related” refers to the fact that these proteins are induced in association with defense responses but does not by imply itself a functional role in defense. However, because some of these proteins have at least potential antimicrobial activities, a role in resistance to pathogens appears plausible. In the past few years, plant gene expression studies have been collected showing that in both compatible and incompatible plant-pathogen interactions, hundreds of genes are up- and down-regulated. In many cases, differences between susceptibility and resistance are associated with differences in the timing and magnitude of these changes rather than with the expression of different sets of genes (Tao et al., 2003). Most of these defense-related proteins correspond to pathogenesis-related proteins (PRs) or the products of so-called systemic acquired resistance (SAR) genes, which were identified several years ago as being associated with resistance reactions of plants to various pathogens (van Loon et al., 1998). The term PR proteins encompasses very different plant proteins, such as chitinases, glucanases, endoproteinases and peroxidases, as well as small proteins such as defensins, thionins and lipid transfer proteins (LTPs). Some of the PR-proteins not are only induced *de novo* upon pathogen attack but also under abiotic stresses such as wounding or other physical or chemical stress. Surprisingly, in some case they are constitutively expressed in some organs or during certain developmental stages (van Loon et al., 2006). Defense-related proteins commonly occur as families of closely related homologues proteins. At gene level, sequences are annotated on the basis of homology to an arbitrary member of the
family without knowledge of whether and where the gene(s) are expressed. This redundancy of sequences hampers the genomic and functional study of these gene families because of the difficulty to readily distinguish between related members. Indeed, it is still unclear why some genes are prone to duplication and they are subsequently retained during evolution while others not. Also the properties or functions in common within large gene families are still unidentified. Inside the gene family, a fine gene expression control and a diversification of function have been proposed (Friedman and Baker, 2007).

Interestingly, a considerable percentage of identified plant allergens can be grouped in PR protein families (Hoffmann-Sommergruber, 2002). In particular, plant-derived allergens have been identified with sequence similarity to PR-protein families 2, 3, 4, 5, 8, 10 and 14. These proteins share some characteristics that are relevant for plant-derived allergens: they are usually rather small proteins (5±70 kDa), stable at low pH and resistant to proteolysis. Therefore new insights into plant cell metabolism and defense-related strategies that plants have developed may contribute to clarify the biological function of the allergenic proteins. A crucial challenge for the future will be understanding the function of each member of the allergen protein families and which effects the suppression of one or more allergenic proteins might have on the defense system of the plant in order to reduce the allergenicity. The double face of the medal, Mal d 1 as allergens and as PR-10 proteins, have to be considered and coordinated to drive the apple breeding processes.

As regards to apple, all the major allergens except one (Mal d 4) are classified as PR proteins. As mentioned above, the major allergen Mal d 1 belong to the PR-10 protein family. PR-10 proteins are cytosolic proteins isolated from various species such as asparagus, parsley, bean, pea and potato (Hoffman-Sommergruber, 2002) and more recently from peach (Zubini et al., 2009). Several PR-10 are up-regulated upon pathogen infection, have a direct and selective antifungal activity, or accumulate in overwintering organs of tree species, suggesting a key role in selective defence mechanisms against microbes and fungi and in protecting plants from abiotic stresses (Flores et al., 2002; Chadha and Das, 2006). A number of PR-10 proteins are also constitutively expressed at different plant growth development stages and/or in different tissues and organs, suggesting a role in development regulation, beside plant stress response.

Despite the ubiquitous occurrence in the plant kingdom, the molecular mechanisms through which PR-10 regulates these important plant processes are not understood yet. Several PR-10 members were reported to act as RNases. This activity can be crucial during plant defence for controlling the burst of transcription that occurs upon stress sensing or for the apoptotic processes activated in pathogen-infected cells in order to limit the pathogen invasion. Additional roles of PR-10 proteins could be related to their ability to bind hydrophobic ligands, such as fatty acids, flavonoids and steroids (Neudecker et al., 2001; Mogensen et al., 2002; Koistinen et al., 2005) or plant hormones such as cytokinins (CKs) (Pasternak et al., 2006; Fernandes et al., 2008) recently emerging as important components of the plant defence strategy repertoire (Chung et al., 2008). Indeed, their concentration is
highly affected by pathogen invasion and, in turn, their high concentration induces apoptosis (Carimi et al., 2003) and expression of other PR proteins, such as PR-1 (Memelink et al., 1987).

As regards apple, the induction of apple PR-10 proteins by pathogen and abiotic factors has been reported by Pühringer et al. (2000) and by Paris et al. (2009). In particular, the upregulation of the Mal d 1 gene has been reported in apple after inoculation with Venturia inaequalis, that is the causal agent of apple scab (Paris et al., 2009). This let to suppose an involvement of this gene in apple scab defense response even if a role cannot be supposed yet. Moreover, no information on the involvement and role of different Mal d 1 genes in apple response to biotic and abiotic stress have been reported yet.
References


Liebhard R, Koller B, Gianfranceschi L, Gessler C (2003c) Creating a saturated reference map for the apple (Malus 3 domestica Borkh.)


13

General introduction


Zubini P, Zambelli B, Musiani F, Ciurli S, Bertolini P, Baraldi E. 2009. The RNA hydrolysis and the cytokinin binding activities of PR-10 proteins are differently performed by two isoforms of the Pru p 1 peach major allergen and are possibly functionally related. Plant Physiol. 150(3):1235-47.
Aims of the thesis

Considering that:

- Cultivated apple (*Malus x domestica*) is diffuse worldwide;
- Apple is one of the fruits most often causing food allergies.
- High level of variability in the allergic reaction have been reported among apple genotypes and among patients.
- As other *Rosaceae* fruits, apple allergens primarily belong to four protein families: Mal d 1 (PR-10), Mal d 2 (Thaumatine-like proteins, TLPs or PR-5), Mal d 3 (Lipid Transfer Proteins or LTPs), and Mal d 4 (profilins).
- Mal d 1, Mal d 2, Mal d 3, and Mal d 4 are encoded by gene families and this redundancy of sequences hampers their genomic and functional studies.
- Mal d 1 is a particularly relevant apple allergen since its cross-reactivity with the main birch pollen allergen, Bet v 1. Also cross-reaction among apple profilin and pollens has been reported.
- Not all Mal d 1 isoallergens are likely to be involved in allergic reactions.
- Despite the ubiquitous occurrence in the plant kingdom, the molecular mechanisms through which PR-10 proteins act in the plant-defense processes are not understood, yet.

The main goal of this thesis was to increase the knowledge on apple allergen gene families by combining genomic and transcriptional approaches. More in detail, five main aims have been pursued as explained in the five chapters of the thesis.

Since the first step to understand apple allergy mechanisms is the identification of apple allergen genes, the aim of the first chapter was to highlight on the genomic organization of the complex *Mal d 1* gene family in the apple genome. In particular, the knowledge about the gene family composition and the physical positioning of homologous members within the family was pursued in order to create a solid base for the further genomic and functional analysis. This step will be crucial also for breeding strategies for hypo-allergenic cultivars.

Due to the high sequence homology among individual members, up to now, the study of allergens gene families was hampered by the difficulties to readily distinguish between related members. The aim of the second chapter was to develop and validate a tool for the specific gene expression analysis of known *Mal d 1* genes. This tool was validated by checking the expression of the different Mal d 1 genes in leaves and fruits but it will be useful also to determine the levels of *Mal d 1* transcripts in other different tissues and conditions.

Starting from the knowledge acquired in Chapter 2, the aim of the third chapter was to carefully evaluate the whole Mal d 1 gene expression profile in apple fruits of different
genotypes with a different allergenicity. This in order to speculate on possible correlations between gene expression levels and degree of apple allergenecity.

Up to now, the biological function of apple PR-10 proteins have been not analyzed in detail despite several different functions and hypothetical involvements in plant responses to a broad range of stresses have been proposed. The aim of Chapter 4 was to exploit the knowledge acquired in Chapter 2, by investigating the gene expression profile of \textit{Mal d 1} genes in young apple leaves upon challenge with a biotic stress (\textit{V. inaequalis}) and to use these results to speculate on the specific involvement of each member of the family in the plant defense mechanisms.

Finally, the aim of Chapter 5 was to make a step forward in the genomic characterization of other apple allergen families. In particular, since some indications on the presence in the apple genome of other LTP and profilin genes are available, this work was focused on the discovery and positioning in the genetic map new \textit{Mal d 2} and \textit{Mal d 4} genes.
Chapter 1

Genomic organization
of the *Mal d 1* gene cluster on LG 16
Introduction

Apple (Malus x domestica, Borkh) is one of the most important fruit species worldwide. The genus Malus belongs to the Rosaceae family that includes other important genera such as pear (Pyrus), stone fruits (Prunus spp.) and strawberry (Fragaria). Nowadays, apple is considered also an important model species for functional genomics research among woody perennial angiosperms due to its relatively small genome size, 750 Mb per haploid genome (Gasic et al., 2009) and availability of its genomic sequence (R. Velasco, Istituto Agrario di San Michele All’Adige, Italy, personal communication). Malus x domestica is a highly heterozygous diploid species (2n = 34) and an allopolyploid origin have been postulated (Chevreau et al. 1985). Due to this ancient duplication, homeologous chromosomes with large colinear regions have been identified in the apple genome, such as LG5 and LG10, LG9 and LG17 or LG 13 and LG 16 (Maliepaard et al., 1998; Liebhard et al., 2002; Chen et al., 2008).

Genetic studies on apple are evolving fast and many genetic tools have been recently developed. The apple molecular map currently has more than 1200 markers on the 17 LGs (Maliepaard et al., 1998; Liebhard et al., 2002, 2003a, 2003b, 2003c; Xu and Korban, 2002a; Naik et al., 2006; Chagné et al., 2008; Celton et al., 2009). Furthermore, at least three apple Bacterial Artificial Chromosome (BAC) libraries have been constructed from different genotypes (Vinatzer et al., 1998; Xu et al., 2001, 2002) that have been successfully used for cloning genes of interest (Vinatzer et al., 2001; Xu and Korban, 2002b; Han et al., 2007). Finally, many collections of Expressed Sequence Tags (ESTs) have been developed obtaining a total number of apple unigenes of ~ 33,000 (Gasic et al., 2009). Computational analysis of apple unigenes has yielded to assign putative functional roles to about 80% of apple ESTs. The remaining 20% of sequences with unknown functions or not matching with any other sequence in public databases may represent apple specific genes or genes most likely associated with tree formation. Gasic et al. (2009) suggested that these sequences probably represent fast-evolving genes that have acquired new functions in apple and related taxa. Interestingly, Mal d 1 genes were included into this category.

Genes of the Mal d 1 family encode for Pathogenesis Related proteins of class 10 (PR-10) for which a key role in selective defense mechanisms against biotic and abiotic stress was proposed (van Loon et al., 2006). The molecular mechanisms by which Mal d 1 proteins regulate these important plant responses are not yet understood. Several PR-10 protein members were reported to hydrolyze RNA (Liu et al., 2006; Yan et al., 2008), others were found to be able to bind hydrophobic ligands or plant hormones (Zubini et al., 2009) that are signal elements induced in the regulation of plant growth, development and plant defense response (Chung et al., 2008). Beside its functional classification as PR-10 protein, Mal d 1 is considered the major apple allergen (Fernandez-Rivas et al., 2006). Allergic reactions caused by Mal d 1 belong to class-II allergies (Breiteneder, 2004), which mainly
affect northern and central European populations and are often associated with birch pollen allergy due to cross-reactivity between Mal d 1 and the major birch pollen allergen, Bet v 1. The two amino acid sequences are 64.5% identical (Vanek-Krebitz et al., 1995).

Genomic and linkage mapping studies revealed that Mal d 1 is a complex gene family consisting of at least 18 members which at the protein level are indicated as isoallergens. Except for Mal d 1.05 on LG 6 and the unmapped Mal d 1.03G, all the Mal d 1 genes were located in two clusters on the two homoeologous LG 13 and 16 (Gao et al., 2005). Two additional Mal d 1-related genes, Mal d 1m and Mal d 1n were previously described (Beuning et al., 2004) and their complete coding sequences are available in the public databases but they have not yet been mapped.

Mal d 1 genes were classified by Gao et al. (2005) into four subfamilies based on DNA sequence similarity, which sub-division showed to co-incide with the length of the intron. Subfamilies I–III contain members with a single intron which sizes are specific for each subfamily. Subfamily IV included only intronless gene members. Subfamily I includes reference sequences that were formerly classified as Mal 1.01 and Mal d 1.02 with an intron length of 168 and 171 nt, respectively. Subfamily II includes Mal d 1.04 and Mal d 1.05 sequences with an intron length of 111 and 119 nt, respectively. Subfamily III contains sequences classified as Mal d 1.06A-C with intron lengths ranging from 128 to 153 nt. Subfamily IV includes intronless sequences like Mal d 1.03A-G, Mal d 1.07, Mal d 1.08 and Mal d 1.09. Comparison of Mal d 1 coding sequences revealed different levels of identity: 71–83% among the four subfamilies, 86–98.1% among genes within a subfamily, and 98.3–100% among alleles of a single gene (Gao et al., 2005). Genes of subfamilies I and IV strengthen the hypothesis of the amphidiploid origin of the apple genome being present on both the homoeologous LG 13 and LG 16. Moreover, the intronless genes of subfamily IV are located on the same region as the intron-containing genes of subfamily I on both LGs. To date, genes of subfamilies II and III were mapped only on LG 16, suggesting that the Mal d 1 clusters on these two LGs evolved differently (Gao et al., 2005).

The role of Mal d 1 in resistance response and allergenicity is not known yet, both for the family as a whole as for its individual members. Differences in functionality among members can be expected as only for some PR-10 isoforms of peach a RNA hydrolysis activity have been reported (Zubini et al., 2009). Also a variability in the cytokinin binding specificity for PR-10s of birch was found by Markovic-Housley et al. (2003). Moreover, PR-10 isoforms of apple differ for the tissues in which they are expressed and for the degree in which their expression is affected by culture and storage conditions (Puehringer et al., 2003; Beuning et al., 2004, Botton et al., 2008). Indeed only a limited number of Mal d 1 proteins and mRNAs have been traced back in apple fruit so far (Helsper et al., 2002; Puehringer et al., 2003; Beuning et al., 2004, Botton et al., 2008) suggesting that not all Mal d 1 isoallergens are likely to be involved in allergic reactions. Furthermore, it has been demonstrated that different Mal d 1 isoallergens and variants, as well as mutants of specific isoallergens, have different binding affinities to IgE (Ma et al., 2006) indicating that the
relative Mal d 1 protein composition, as well as the presence of specific alleles, may have significant and different effects on allergenicity. Gao et al. (2008) corroborated the hypothesis of a different involvement in allergic reaction for different Mal d 1 genes, suggesting a strong association of Mal d 1.04 and Mal d 1.06A with apple allergenicity. It is becoming evident that assessing the total amount of Mal d 1 proteins is not sufficient to understand apple allergy and, indeed, the identification of specific Mal d 1 genes implicated in allergic reactions from the many existing allergen genes should get high priority in allergy research.

This work highlights on the genomic organization of the complex Mal d 1 gene family on LG16. To reach this goal, an apple BAC library of the cultivar Florina was screened through a PCR-based protocol and two BAC clones containing several members of the Mal d 1 gene cluster on LG 16 were fully sequenced. The physical map was anchored to the genetic map providing new information as number of isoallergens in the cluster, their orientation and physical distances.
Materials and Methods

Apple BAC library

A BAC library from the cultivar Florina was used (Vinatzer et al., 1998). The library is made by 36,864 BAC clones stored in 96 plates (384-wells). Its clones have an average insert size of 120 kb and represents approximately 5 x apple haploid-genome equivalents. A bi-dimensional pooling mechanism was performed as described by Cova (2007). The horizontal pool consisted of 96 samples (plate pool) each containing all the BAC clones from a single 384-wells plate. The vertical pool consisted in 4 X 96 samples prepared by bulking the clones of a specific position (i.e. A1, A2, ecc...) from all the original 384-wells plates (96 clones/well). Plasmids from the BAC clone pools were extracted by the alkaline extraction procedure (Birnboim and Doly, 1979).

PCR-based screening of the BAC library

The PCR-based screening of the whole library was carried out with 4 primer pairs specifically designed for one of each of the four Mal d 1 sub-families. A general Mal d 1 primer pair was also designed on consensus regions. These primers (listed in Table 1) where designed with the Primer3 software (http://frodo.wi.mit.edu/primer3/). All the BAC clones identified after the screening steps were picked up from the library, singularized and tested by colony-PCR with the same primers used for the screening.

All the PCR amplifications were performed in a 17.5 µl volume containing 200 ng of DNA from BAC library pools, 0.1 µM gene-specific primers (Table 1), 1.5 mM MgCl₂, 100 µM dNTPs, 0.5 Unit DNA Polymerase (Fisher Molecular Biology, Hampton, NH, USA) and 1X reaction buffer. The reaction included an initial 3 min denaturation at 94°C, followed by 35 PCR cycles (45 s at each optimised annealing temperature, 2 min at 72°C, and 30 s at 94°C), with a final extension of 10 min at 72°C. The amplicons were visualised on an Image Station 440 CF (Kodak, Rochester, N.Y., USA) after electrophoresis in 1.5% (w/v) agarose gels and ethidium bromide staining.

Analyses of positive BAC clones

Plasmid DNA from each positive BAC clone was purified as described by Untergasser (2006). Isoallergen-specific primer pairs (listed in Table 2) were designed adding in some case deliberate SNPs to increase specificity, as reported by Gao et al. (2005). Just one primer pair were designed for amplify all the Mal d 1.03 genes. They were used to identify by PCR the Mal d 1 genes physically located on each BAC clone. The specificity of some primers was enhanced by deliberate mismatch at 3’-end of the primer sequences. All PCR amplifications were performed in a 20 µl volume containing 50 ng of plasmid DNA, 0.1 µM gene-specific primers, 1.5 mM MgCl₂, 100 µM dNTPs, 0.5 Unit AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1X
reaction buffer. The reaction included an initial 10 min denaturation at 95°C, followed by 30 PCR cycles (45 s at the optimised annealing temperature, 2 min at 72°C, and 30 s at 95°C), with a final extension of 7 min at 72°C.

Moreover, plasmid DNA (approximately 20 µg) from positive BAC clones, was digested with 2 U EcoRI overnight at 37°C. Digested DNA fragments were loaded onto 1% Ultra Pure agarose gel (Lonza, Basel, Switzerland), and electrophoresed at 35 V overnight. Images of EcoRI-digested DNA fragments of positive BAC clones were used to identify overlapping BAC clones.

**Shotgun sequencing**

Two *Mal d 1* BAC clones were subjected to shotgun sequencing at Macrogen Inc. (Korea). For each BAC clone, a six-fold sequence coverage was assembled by Greenomics (The Netherlands). Gaps between contigs were filled by direct sequencing performed with primer pairs (listed in Table 4) specifically designed with the software PrimerSelect (Lasergene® v8.0) on all the contig-ends. Single run sequencing was performed by Bio-Fab Research srl (Pomezia, Italy). The final assembly was manually performed with the SeqMan software (Lasergene® v8.0).

**Sequences analysis and annotation**

BAC sequences were annotated using the gene prediction program GENESCAN (Burge and Karlin, 1997). Predicted genes were searched for similarity to known proteins by BLASTP (Altschul et al., 1990) with E-value cut off E<15 against the National Center for Biotechnology Information (NCBI) non-redundant protein database. The predicted genes were also searched for similarity against the NCBI nucleotidic sequences database by BLASTN (Altschul et al., 1990) with E-value cut off E<25. Predicted amino acid sequences of *Mal d 1* isoallergens were aligned using ClustalW (Jeanmoung et al., 1998). BALSTX (Altschul et al., 1990) was used to further verify the correspondances between predicted ORFs and proteins in the databases.

The promoter region sequences (~1300bp) of each *Mal d 1* gene found on the two BAC clones were analyzed using the program PLANTPAN to find transcription factor binding sites.

**BAC clones anchoring on a genetic map**

The two BAC sequences were analyzed with the software Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html) and 10 SSR primers (listed in Table 9) were designed from regions flanking repetitive stretches, using Primer3 software (http://frodo.wi.mit.edu/primer3/). PCR amplification and gel electrophoresis were performed as described previously (Gianfranceschi et al., 1998). These SSR markers were applied on a Durello di Forlì × Fiesta population (population size n = 174) and added to the available molecular marker linkage map using JoinMap 3.0® (van Ooijen and Voorrips 2001) using the
Kosambi mapping function. The LOD value chosen for grouping LG16 markers was equal to 7. The final visual representation of the map was generated with MapChart (Voorrips 2001).
Results and discussion

BAC library screening

The screening of the Florina BAC library for Mal d 1 sequences, conducted with the four subfamily representing primers (Table 1), resulted in 20 positive BAC clones named MC-1 to MC-20.

Table 1 Primer pairs used for the PCR-based BAC screening.

<table>
<thead>
<tr>
<th>Primer names</th>
<th>Sequence (5’-3’)</th>
<th>Ta (°C)</th>
<th>Product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subfam I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Md1-1For</td>
<td>TGCAAACTATTACGGCTAGGG</td>
<td>60</td>
<td>176</td>
</tr>
<tr>
<td>Md1-1Rev</td>
<td>CTCAACAGGCTAGGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subfam II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Md1-2For</td>
<td>CGGAGGCTAGGGAGGT</td>
<td>60</td>
<td>315</td>
</tr>
<tr>
<td>Md1-2Rev</td>
<td>CATCTCATCAGGGAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subfam III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Md1-3For</td>
<td>GCTCCAAAGGCTCAAAAAC</td>
<td>57</td>
<td>232</td>
</tr>
<tr>
<td>Md1-3Rev</td>
<td>TCTACCTCTGCTCAAATGAA</td>
<td>60</td>
<td>198</td>
</tr>
<tr>
<td>Subfam IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Md1-4For</td>
<td>TCACTTTGGGAAAGGTAGCAC</td>
<td>60</td>
<td>375; 471; 489; 528</td>
</tr>
<tr>
<td>Md1-4Rev</td>
<td>TGGTGTGTTAG1GCTGGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Md1-5For</td>
<td>GTCAATATCGCTTGTCTGCTGAG</td>
<td>57</td>
<td>375; 471; 489; 528</td>
</tr>
<tr>
<td>Md1-5Rev</td>
<td>TCTTTGGCAGCCTTGACATGCTTCTC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The generic primer pair gave different product lengths due to the different intron lengths of the isoallergens belonging to different subfamilies.

Further analysis of these BAC clones was performed both by their amplification with 11 gene specific primers for Mal d 1 genes mapped by Gao et al. (2005) (listed in Table 2) and by enzymatic digestions (Figure 1A). This analysis made it possible to identify the (partly) overlapping clones and their putative location based on the known map position of some apple allergen genes (Gao et al., 2005) as summarized in Table 3. In detail, five clones gave amplifications with primers specific for Mal d 1 genes of LG16 (MC-1, -12, -14, -16, -20) and therefore they should derive from chromosome 16. Analogously, 11 clones gave amplifications for Mal d 1 genes of LG13 (Table 3) and were thus supposed to be from chromosome 13. Finally, four clones gave amplifications with primers for Mal d 1.05 (MC-2, -3, -4, -5), the only isoallergen mapped on LG 6 and therefore they should derive from chromosome 6. Primers specific for Mal d 1.04 didn’t amplify on any BAC clone. It is not possible to exclude the presence of further BAC clones containing Mal d 1 genes because of failure of the screening can not be excluded nor lack of some genome regions in the BAC library. Two clones (MC-2 and -10) gave amplicons with primers for genes that are thought to be located on different linkage groups: MC-2 for Mal d 1.03 of LG13 and Mal d 1.05 of LG6 and MC-10 for Mal d 1.01 and 1.03 of LG13 and Mal d 1.06A of LG16. This may be due to the presence of still unidentified isoallergen genes.
Table 2. Primer pairs used for the analysis of BAC clones. Underlined and italicized nucleotides in the primers sequences are deliberate introduced SNP to increase specificity (Gao et al., 2005).

<table>
<thead>
<tr>
<th>Isoallergens</th>
<th>Primer names</th>
<th>Sequence (5’-3’)</th>
<th>Ta (°C)</th>
<th>Product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mal d 1.01</strong></td>
<td>Mal d 1.01general F</td>
<td>TGACTCGATTGACGAAGCAAG</td>
<td>54</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.01general R</td>
<td>TTCAATGGTTTCTGGGTGAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.02</strong></td>
<td>Mal d 1.029 F</td>
<td>GCCCTGGAACCATCAAGAGT</td>
<td>51</td>
<td>419</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.02general R</td>
<td>GTGCTCTCTCTGTCTGATCATCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.03A-P</strong></td>
<td>Mal d 1.03general F</td>
<td>CTGACAACCTCTCCCCAACAGA</td>
<td>54</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.03general R</td>
<td>GTGTTGTTAGTGCTGCTGATTGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.04P</strong></td>
<td>1.0404F</td>
<td>TAATTCATTGCCAGCGCCGC</td>
<td>54</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>1.04 generalR</td>
<td>TCAGAGATGCGATCAATTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.05A</strong></td>
<td>Mal d 1.0502 SNP96 F</td>
<td>TGTGATGTTGAACTGCTACCGT</td>
<td>52</td>
<td>358</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.05general R</td>
<td>TTGTTGGTCTAAGCGATCTTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.06A</strong></td>
<td>1.06A03 SNP37F</td>
<td>CGAAACCGAATACGCGCTAAA</td>
<td>54</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>1.06A generalR</td>
<td>GCATCCCCCTCAAAGTCAACCGATACGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.06B</strong></td>
<td>1.06B01 SNP229F</td>
<td>ATTTTCTCAATTAATCGTCTTTTTC</td>
<td>54</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>1.06B generalR</td>
<td>CTTTGCGATGATCAATCTCCTACACCTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.06C</strong></td>
<td>1.06C generalF</td>
<td>TAACATTTGCCTCTACCATCTCCTG</td>
<td>54</td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>1.06C03 SNP417R</td>
<td>CCAACTTAATCTCATAAGAGATCTCCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.07</strong></td>
<td>Mal d 1.07general F</td>
<td>CAACCTTTGTTGACCAATACGTCTC</td>
<td>51</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.07general R</td>
<td>CTTGGTGTGTTGATGGTGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.08</strong></td>
<td>Mal d 1.08general F</td>
<td>CGACTGCTTGGATGGTGATG</td>
<td>51</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.08general R</td>
<td>CTAGCTTGGATGGTGATCAGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mal d 1.09</strong></td>
<td>Mal d 1.09general F</td>
<td>GAGCTGGAAACATTAGAAGATAG</td>
<td>54</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.09general R</td>
<td>GCCCTTGTGATGCGAACCTGT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*General primer pair for all Mal d 1.03 isoallergens

*Isoallergens for which were used primer pairs developed at Plant Research International (The Netherlands)

Based on the EcoRI restriction patterns of the BAC clones (examples are reported in Figure 1A), 6 groups were distinguished, as is indicated in the last column of Table 3: two groups for LG 16 (group I and II), two groups for LG 13 (III and IV) and two groups for LG 6 (V and VI). The putative map position of some BAC clones is schematically presented in Figure 1B. The restriction analysis made it possible to add information for the above mentioned putative inconsistencies in amplification patterns of MC-10. MC-10 showed to have overlapping digestion fragments with MC-8/13 and MC17 (MC-8 and MC-17 restriction pattern are reported in Figure 1A). These last three clones were supposed to derive from LG13 making more consistent the localization on LG13 also for MC-10 as well. The amplification obtained with Mal d 1.06A primers on MC-10 is probably due to a new Mal d 1 genes similar to Mal d 1.06A on LG13. Despite the clear amplification of MC-15 with Mal d 1.03 primers, it remained ungrouped because its digestion pattern doesn’t not clearly fit to any of the other BAC clones (data not shown). Probably this clone contains a fragment of DNA at the extreme of the Mal d 1 cluster.
Table 3 Results of BAC clone amplifications with some gene specific primer pairs. X indicates the positive amplification. In black are indicated the two clones chosen for sequencing. The final column (Group) refers to overlapping clones depending on digestion patterns analysis (see fig 1A).

<table>
<thead>
<tr>
<th>BAC clone</th>
<th>LG 13</th>
<th>LG 16</th>
<th>LG 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC-1</td>
<td>1.01</td>
<td>1.03A-F</td>
<td>1.02</td>
</tr>
<tr>
<td>MC-12</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MC-14</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>MC-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-15</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-6</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-7</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-9</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-11</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-18</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-19</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-8</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-13</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-10</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-17</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-2</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-3</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-4</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MC-5</td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: A) Profile of 11 EcoRI-digested BAC clones, derived from group I to IV. Each group contains overlapping clones. B) Genetic map position of apple Mal d 1 isoallergens (Gao et al, 2005) and BAC clones related to each LG. BAC clones are indicated with boxes. Box sizes are not in scale with BAC length.

BAC clones sequencing

For sequencing, the cluster on LG 16 was preferred to the cluster on the homoeologous LG 13 because containing the highest number and largest diversity in Mal d 1 isoallergen genes, among which Mal d 1.04 and Mal d 1.06A, proposed to be related to allergenicity (Gao et al., 2008). Hence, of the 20 BAC clones containing Mal d 1 genes, two not overlapping but representative BAC clones from the sub-families I and IV on LG 16 (MC-12 and MC-20) were chosen for sequencing. MC-12 was chosen within group I because it seem to contain more Mal d 1 isoallergen genes than MC-1 and MC-14 (Table 3). Within
group II, MC-20 was chosen instead of MC-16 for its length, which was estimated to be ~130 kb based on NotI digestion results, whereas C-20 has only around ~70 kb (Figure 2).

![Figure 2: A) 10kb marker. B) Mid range marker. C) NotI digestion of BAC clones of group 2. In detail, in the first line the 10 kb marker, in the second lane the low range marker, in the third lane the mid range marker, in the fourth line the MC-16 and in the fifth lane MC-20.](image)

The first assembly output gave four contigs for each clone. A further sequencing step performed with primers designed on both ends of each contig (listed in Table 4) allowed the assembly of a single full-length sequence for both clones. This approach is summarized in Figure 3. In particular, for MC-12 the sequence resulted of 125’046 nt (Figure 3A), for MC-20 it resulted of 132’896 nt (Figure 3B). The full-length sequences of the two BAC clones are reported in Appendix 1A and 1B, respectively.

| Table 4. Primers used for direct sequencing step on BAC clones MC-12 and MC-20. |
|-------------------------------|------------------|----------------|
| BAC clone | Contig | Primer name | Primer sequence (5'-3') |
| MC-12 | C1 | 1-MC12-C1For | AGCAGAAATGCTTCGTTCTGCGT TT |
| | C2 | 2-MC12-C2Rev | AGGGGAGTTATGCC AAAAT |
| | 3-MC12-C2For | GCTAGTCAGGTCGGGATTTC |
| | C3 | 4-MC12-C3Rev | TCCTGGGAATGGAAACACCTT |
| | 5-MC12-C3For | AATGCCAAGCTTTCAAGAT |
| | C4 | 6-MC12-C4Rev | GATTGGGATATGGGTAT |
| MC-20 | C1 | 7-MC20-C1For | TAATGTGGGGATGGGTAT |
| | C2 | 8-MC20-C2Rev | TCTTGGAGTCCCACCTGTT |
| | 9-MC20-C2For | CAGAACATGCTTTCAAGAT |
| | C3 | 10-MC20-C3Rev | CCAATTTCCGACGTTAGGAA |
| | 11-MC20-C3For | CTTGGAGCTTTCAAGAT |
| | C4 | 12-MC20-C4For | CATCGTGTTCTACTGCTCA |
Figure 3: Outline of steps occurred to close the gaps between contigs of clone MC-12 (A) and MC-20 (B). Box sizes are not in scale with BAC length.

**Sequences annotation**

The gene predictor software (GENESCAN) identified 52 open reading frames (ORFs): named from ORF1 to ORF32 in the clone MC-12 and from ORF33 to ORF52 in MC-20. The putative functions of these ORFs were recorded based on their BLASTP scores and in Table 5 and 6 are reported the descriptions of the proteins with the highest similarities with the 52 ORFs. Of the total 52 ORFs, 16 ORFs showed homology to known genes; 10 were homologous to retrotransposons, 18 showed no significant homology to sequences in the databases and 8 were similar to hypothetical proteins from genome sequencing of *Vitis vinifera*. As regards ORFs homologues to known genes, the majority (13/16) were similar to *Mal d 1* sequences. In particular, MC-12 carries nine *Mal d 1*-like ORFs, and MC-20 four. Their sequences were further analyzed with BLASTN. Results and additional information (exon/intron length) are synthesized in Tables 7 and 8 for MC-12 and MC-20, respectively. Eight out of thirteen *Mal d 1*-like genes had previously been mapped on LG16. The remaining five sequences were not previously known to be part of LG16. Of these five, two were identical to EST-sequences previously identified as *Mal d 1*-like genes named as *Mal d 1m* and *Mal d 1n* (Beuning et al., 2004), one is similar to *Mal d 1.03G*, and one was previously identified as the pseudogene *Mal d 1ps2* (AY827730) (Gao et al., 2005). We can now assign these four genes to LG16. The remaining sequence is completely new and is coded by ORF 15. Below, results on these five genes will be further described.
Table 5. Predicted ORFs on MC-12 analyzed with BALSTP software. In bold are indicated ORFs similar to *Mal d 1* genes.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Direction¹</th>
<th>Start position²</th>
<th>Lenght (nt)³</th>
<th>Lenght (aa)</th>
<th>Protein description as reported by the software</th>
<th>Accession number</th>
<th>E-value</th>
<th>Source</th>
<th>Conserved domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>↓</td>
<td>519</td>
<td>1351</td>
<td>446</td>
<td>Reverse transcriptase (RNA-dependent DNA polymerase) domain containing protein</td>
<td>AAT85304</td>
<td>4,00E-92</td>
<td>Oryza sativa Japonica Group</td>
<td>RVE superfamily</td>
</tr>
<tr>
<td>ORF2</td>
<td>↓</td>
<td>2837</td>
<td>1105</td>
<td>222</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF3</td>
<td>↓</td>
<td>5457</td>
<td>6506</td>
<td>2162</td>
<td>Retrotransposon protein</td>
<td>ABM55240</td>
<td>0.00</td>
<td>Beta vulgaris</td>
<td></td>
</tr>
<tr>
<td>ORF4</td>
<td>↑</td>
<td>19700</td>
<td>602</td>
<td>159</td>
<td>Mal d 1.06A02</td>
<td>AAX18298</td>
<td>1,00E-84</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF5</td>
<td>↓</td>
<td>22252</td>
<td>701</td>
<td>67</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF6</td>
<td>↑</td>
<td>24478</td>
<td>1022</td>
<td>212</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF7</td>
<td>↓</td>
<td>29652</td>
<td>231</td>
<td>76</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF8</td>
<td>↓</td>
<td>30449</td>
<td>961</td>
<td>161</td>
<td>Mal d 1-like</td>
<td>AAS00052</td>
<td>1,00E-86</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF9</td>
<td>↑</td>
<td>33317</td>
<td>4098</td>
<td>600</td>
<td>Retrotransposon protein, putative, Ty1-copia subclass</td>
<td>ABF96902</td>
<td>4,00E-48</td>
<td>Oryza sativa japonica group</td>
<td>RVE superfamily; RVT_2 superfamily</td>
</tr>
<tr>
<td>ORF10</td>
<td>↑</td>
<td>39968</td>
<td>2791</td>
<td>89</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF11</td>
<td>↓</td>
<td>45546</td>
<td>651</td>
<td>159</td>
<td>Mal d 1 allergen</td>
<td>AAD26548</td>
<td>2,00E-85</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF12</td>
<td>↑</td>
<td>47853</td>
<td>3177</td>
<td>68</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF13</td>
<td>↑</td>
<td>51932</td>
<td>663</td>
<td>123</td>
<td>Hypothetical protein</td>
<td>CAN78839</td>
<td>2,00E-29</td>
<td>Vitis vinifera</td>
<td></td>
</tr>
<tr>
<td>ORF14</td>
<td>↑</td>
<td>53816</td>
<td>507</td>
<td>168</td>
<td>Copia-type polyprotein</td>
<td>AAG51247</td>
<td>1,00E-39</td>
<td>Arabidopsis thaliana</td>
<td></td>
</tr>
<tr>
<td>ORF15</td>
<td>↓</td>
<td>59735</td>
<td>861</td>
<td>146</td>
<td>Mal d 1-like</td>
<td>AAS00052</td>
<td>2,00E-38</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF16</td>
<td>↑</td>
<td>63931</td>
<td>1861</td>
<td>132</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF17</td>
<td>↑</td>
<td>66452</td>
<td>428</td>
<td>75</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF18</td>
<td>↓</td>
<td>68017</td>
<td>702</td>
<td>163</td>
<td>Mal d 1-like</td>
<td>AAS00053</td>
<td>6,00E-92</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF19</td>
<td>↑</td>
<td>70566</td>
<td>415</td>
<td>104</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF20</td>
<td>↑</td>
<td>71477</td>
<td>1540</td>
<td>96</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF21</td>
<td>↓</td>
<td>73770</td>
<td>633</td>
<td>159</td>
<td>Mal d 1.06B02</td>
<td>AAX20971</td>
<td>2,00E-84</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF22</td>
<td>↓</td>
<td>76305</td>
<td>812</td>
<td>119</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF23</td>
<td>↓</td>
<td>78970</td>
<td>283</td>
<td>97</td>
<td>Mal d 1.06C04 *</td>
<td>AAX20989</td>
<td>2,00E-42</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF24</td>
<td>↑</td>
<td>79408</td>
<td>4443</td>
<td>1390</td>
<td>Hypothetical protein</td>
<td>XP_002280852</td>
<td>0.00</td>
<td>Vitis vinifera</td>
<td>RVE superfamily; RVT_2 superfamily</td>
</tr>
<tr>
<td>ORF25</td>
<td>↓</td>
<td>87785</td>
<td>2574</td>
<td>338</td>
<td>Retrotransposon protein, putative</td>
<td>ABA95230</td>
<td>1,00E-51</td>
<td>Oryza sativa japonica group</td>
<td>RVT_2 superfamily</td>
</tr>
<tr>
<td>ORF26</td>
<td>↓</td>
<td>92218</td>
<td>608</td>
<td>159</td>
<td>Mal d 1.06C</td>
<td>AAX18306</td>
<td>5,00E-85</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF27</td>
<td>↓</td>
<td>95324</td>
<td>470</td>
<td>120</td>
<td>Mal d 1.06C*</td>
<td>AAX18306</td>
<td>3,00E-47</td>
<td>Malus domestica</td>
<td>Polyketide_cyc 2 superfamily</td>
</tr>
<tr>
<td>ORF28</td>
<td>↓</td>
<td>96401</td>
<td>397</td>
<td>101</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>Start</td>
<td>Stop</td>
<td>Length</td>
<td>Description</td>
<td>Accession</td>
<td>E-value</td>
<td>Organism</td>
<td>Domain</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>ORF29</td>
<td>↓</td>
<td>97358</td>
<td>9486</td>
<td>Transposon protein, putative</td>
<td>ABA94905</td>
<td>2.00E-17</td>
<td><em>Oryza sativa japonica</em> group</td>
<td>Transposase domain</td>
<td></td>
</tr>
<tr>
<td>ORF30</td>
<td>↓</td>
<td>107669</td>
<td>2433</td>
<td>Putative non-LTR retroelement reverse transcriptase</td>
<td>AAD21778</td>
<td>6.00E-35</td>
<td>Arabidopsis thaliana</td>
<td>RT_like superfamily, NTP binding site</td>
<td></td>
</tr>
<tr>
<td>ORF31</td>
<td>↓</td>
<td>110916</td>
<td>5126</td>
<td>Hypothetical protein</td>
<td>CAN79553</td>
<td>4.00E-71</td>
<td>Vitis vinifera</td>
<td>HMA, Heavy-metal-associated domain</td>
<td></td>
</tr>
<tr>
<td>ORF32</td>
<td>↑</td>
<td>117249</td>
<td>8154</td>
<td>Hypothetical protein</td>
<td>XP_002273305</td>
<td>6.00E-76</td>
<td>Vitis vinifera</td>
<td>Glycosyl hydrolases family 28</td>
<td></td>
</tr>
</tbody>
</table>

¹The arrows are directed downward when the gene is directed from T7-end to Sp6-end and vice versa.
²Start position counting from T7-end of the clone.
³Length of the gDNA nucleotide sequences.
*Not complete sequences.
Table 6 Predicted ORFs on MC-12 analyzed with BALSTP software. In bold are indicated ORFs similar to Mal d 1 genes.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Direction¹</th>
<th>Start position²</th>
<th>Lenght (nt)³</th>
<th>Lenght (aa)</th>
<th>Protein description</th>
<th>Accession number</th>
<th>E-value</th>
<th>Source</th>
<th>Conserved domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF33</td>
<td>↓</td>
<td>6</td>
<td>1977</td>
<td>622</td>
<td>RNA-directed DNA polymerase</td>
<td>ABD28426</td>
<td>4.00E-179</td>
<td>Medicago truncatula</td>
<td></td>
</tr>
<tr>
<td>ORF34</td>
<td>↑</td>
<td>2241</td>
<td>3000</td>
<td>404</td>
<td>No significant similarities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF35</td>
<td>↓</td>
<td>5408</td>
<td>480</td>
<td>163</td>
<td>Putative reverse transcriptase</td>
<td>AA40500</td>
<td>7.00E-36</td>
<td>Solanum demissum</td>
<td></td>
</tr>
<tr>
<td>ORF36</td>
<td>↓</td>
<td>9630</td>
<td>5803</td>
<td>680</td>
<td>Hypothetical protein</td>
<td>CAN74865</td>
<td>8.00E-39</td>
<td>Vitis vinifera</td>
<td>PMD, Plant mobile domain</td>
</tr>
<tr>
<td>ORF37</td>
<td>↑</td>
<td>16117</td>
<td>1110</td>
<td>142</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF38</td>
<td>↑</td>
<td>17717</td>
<td>5550</td>
<td>835</td>
<td>Hypothetical protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF39</td>
<td>↓</td>
<td>24713</td>
<td>950</td>
<td>74</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF40</td>
<td>↑</td>
<td>27462</td>
<td>3347</td>
<td>422</td>
<td>Hypothetical protein</td>
<td>XP_002267217</td>
<td>2.00E-61</td>
<td>Vitis vinifera</td>
<td>RVT_2 superfamily</td>
</tr>
<tr>
<td>ORF41</td>
<td>↓</td>
<td>34860</td>
<td>6920</td>
<td>1526</td>
<td>Retrotrasposon protein</td>
<td>ABM55240</td>
<td>0.00</td>
<td>Beta vulgaris</td>
<td>Retrotransposon gag protein; RVP_2, Retroviral aspartyl protease; RT_LTR; RnaseH1</td>
</tr>
<tr>
<td>ORF42</td>
<td>↑</td>
<td>43949</td>
<td>480</td>
<td>159</td>
<td>Mal d 1.03G</td>
<td>AAX18324</td>
<td>7.00E-85</td>
<td>Malus domestica</td>
<td>Pathogenesis-related protein Bet v I family</td>
</tr>
<tr>
<td>ORF43</td>
<td>↓</td>
<td>61000</td>
<td>480</td>
<td>159</td>
<td>Mal d 1.07</td>
<td>AAX18307</td>
<td>5.00E-86</td>
<td>Malus domestica</td>
<td>Pathogenesis-related protein Bet v I family</td>
</tr>
<tr>
<td>ORF44</td>
<td>↓</td>
<td>63412</td>
<td>480</td>
<td>159</td>
<td>Mal d 1.0093</td>
<td>AAX20996</td>
<td>5.00E-87</td>
<td>Malus domestica</td>
<td>Pathogenesis-related protein Bet v I family</td>
</tr>
<tr>
<td>ORF45</td>
<td>↑</td>
<td>66259</td>
<td>6528</td>
<td>349</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF46</td>
<td>↓</td>
<td>74350</td>
<td>480</td>
<td>159</td>
<td>Mal d 1.08</td>
<td>AAX18309</td>
<td>6.00E-87</td>
<td>Malus domestica</td>
<td>Pathogenesis-related protein Bet v I family</td>
</tr>
<tr>
<td>ORF47</td>
<td>↓</td>
<td>76606</td>
<td>9082</td>
<td>460</td>
<td>Hypothetical protein</td>
<td>XP_002271465</td>
<td>5.00E-63</td>
<td>Vitis vinifera</td>
<td>DUF789, Protein of unknown function (DUF789)</td>
</tr>
<tr>
<td>ORF48</td>
<td>↑</td>
<td>87704</td>
<td>213</td>
<td>106</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF49</td>
<td>↑</td>
<td>100005</td>
<td>3648</td>
<td>248</td>
<td>HEAT repeat-containing protein</td>
<td>NP_197125</td>
<td>4.00E-36</td>
<td>Arabidopsis thaliana</td>
<td></td>
</tr>
<tr>
<td>ORF50</td>
<td>↓</td>
<td>114365</td>
<td>3129</td>
<td>674</td>
<td>Multi-domain protein (a sensor histidine kinase and a response regulator)</td>
<td>XP_002519174</td>
<td>0.00</td>
<td>Ricinus communis</td>
<td>REC Response_reg, Response regulator receiver domain; myb-like DNA-binding domain, SHAQKYF class</td>
</tr>
<tr>
<td>ORF51</td>
<td>↑</td>
<td>118236</td>
<td>675</td>
<td>224</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF52</td>
<td>↑</td>
<td>119188</td>
<td>7853</td>
<td>1050</td>
<td>Protein COBRA precursor, putative</td>
<td>XP_002519175</td>
<td>0.00</td>
<td>Ricinus communis</td>
<td>COBRA superfamily</td>
</tr>
</tbody>
</table>

¹The arrows are directed downward when the gene is directed from T7-end to Sp6-end and vice versa
²Start position counting from T7-end of the clone
³Length of the gDNA nucleotide sequences
\textbf{Mal d 1m – Mal d 1.10.} The \textit{Mal d 1m} sequence (AY428588), completely corresponding here to ORF8, was previously derived from an EST. We now report for the first time its full-length genomic sequence. Its intron length is 475 nt, which is longer than any of the other known \textit{Mal d 1} introns. This deviating intron size makes that it doesn’t fit clearly with any of the four sub-families described so far. The protein sequence has 161 amino acid and the most similar \textit{Mal d 1} amino acid sequence present in the database is \textit{Mal d 1} 1.04 (92% similarity). Following official allergen nomenclature (King et al., 1995) we now propose to denote this gene as \textit{Mal d 1.10}.

\textbf{Mal d 1n – Mal d 1.11.} Similarly, to \textit{Mal d 1.10} also the \textit{Mal d 1n} sequence, completely corresponding here to ORF15, was previously derived from an EST (AY428589). Its full-length genomic sequence is reported here for the first time. Its intron length (208 nt) does doesn’t fit with any known sub-family. The protein sequence includes 163 amino acid, which is three amino acids more than the protein sequences respect most of the other \textit{Mal d 1} iso-allergens (Figure 4). Its protein sequence was most similar to \textit{Mal d 1} 1.06A (67%). We here denote this gene as \textit{Mal d 1.11}.

\textbf{Mal d 1.03G.} ORF42 shows an high homology with \textit{Mal d 1.03G01} (AY822733). It has 5 non-synonomous SNPs, corresponding to two substitutions in position 110 (asparagine \rightarrow serine) and 132 (histidine \rightarrow glutamine), were detected. Sequence similarities exceeding 95% are thought to refer to different variants of the same iso-allergen or, at the genetic level, different alleles of the same locus. Hence, the ORF42 sequence of Florina corresponds to a new allele (allele 02) of for \textit{Mal d 1.03G}. This is the first event in which a \textit{Mal d 1.03} like gene is found on LG16 as all other known \textit{Mal d 1.03} like genes are present on LG13.

\textbf{Mal d 1ps2.} \textit{Mal d 1ps2} (AY827730) was previously found by Gao et al. (2005). Whereas Gao et al. (2005) could not map this sequence, we could assign it to LG16 as ORF23 has a highly similar sequence, which is truncated after 283 nt instead of 471. The most similar protein sequence is Mal d 1.06C (AAX20989) and this is the reason of its annotation in Table5, but ORF23 is truncated. Since 19 SNPs were found between this BAC sequence and \textit{Mal d 1ps2.02}, the Florina sequence can thus be classified as a new allele (variant 03)

\textbf{ORF15 - Mal d 1.12} The coding sequence of ORF15 has high similarity with several Bet v 1-like sequences. By BLASTX software, the highest similarity have been found with the allergen Pru du 1.03 of \textit{Prunus dulcis x Prunus persica} (85% of identities). Mal d 1.04 was the most similar \textit{Malus} sequence (81%). Its predicted protein contains 146 AAs, hence it is significantly shorter than any other \textit{Mal d 1} protein. It can be considered as a new \textit{Mal d 1} family member which we propose to name \textit{Mal d 1.12}.

As we discussed the five genes that for the first time could be assigned to LG16, we now further present results on the previously mapped \textit{Mal d 1} genes.

In MC-12 BAC clone, 4 ORFs correspond to already known and mapped allergen genes (ORF4, ORF11, ORF21 and ORF27). ORF4 resulted similar to the allele 02 of \textit{Mal d 1.06A}
(AY827697) but with 1 synonymous SNP in the coding region. In a previous work by Gao et al. (2008), variant 02 of Mal d 1.06A was related to low allergenicity as measured by Skin Prik Tests.

Figure 4: Genomic organization of Mal d 1 genes on LG16. A) Physical map of two BACs. The new isoallergens are indicated in boxes. The isoallergens mapped for the first time is underlined. Transposable elements are represented as gray boxes. B) Genetical map of LG 16 in Durello di Forlì. SSRs developed on the sequences of the two BACs are indicated in bold. In red are indicated retrotransposon elements.

Also a relevant allele dosage effect was proposed because in the cultivar Santana and Priscilla, considered the cultivars with lower allergenic potential, an homozygous genotype for the allele 02 was found, in intermediated allergenic cultivars the allele 02 was found at a
single dosage whereas all high allergenic cultivars lacked this allele. Since Florina has at least one allele 02 for the gene Mal d 1.06A, it is candidate to be an intermediate or low allergenic cultivar but further analysis on the complete allelic composition and immunological data of Florina are needed. ORF11 showed just one synonymous SNP in the coding region in respect to allele 01 of the isoallergen Mal d 1.02 (AY827654). ORF21 resulted identical to the isoallergen Mal d 1.06B02 (AY827712). ORF26 was highly homologous to the allele 02 of the isoallergen Mal d 1.06C (AY827725) except for a synonymous SNP in the coding region. ORF27 was classified as a pseudogene due to the presence of stop codons in the sequence and for the high homology with the known pseudogene (AY827730). In Table 5 it is annotated as Mal d 1.06C (AAX18306) since this is the most similar protein sequence but ORF27 showed a not complete sequence.

On the BAC clone MC-20 only intronless Mal d 1-like sequences with the conserved full-length of 480 nt were found. In particular, ORF43 showed high homology with the isoallergen Mal d 1.0701 (AY822717) but with 5 non-synonymous SNPs causing an amino acid substitution in position 71 (lysine → arginine). This BAC sequence of Florina can thus be classified as a new Mal d 1.07 variant (variant 03). ORF44 resulted identical to the isoallergen Mal d 1.0903 (AY822721) and, finally, ORF46 was identical to the isoallergen Mal d 1.0801 (AY822719).

Looking at the allelic composition of the Mal d 1 genes on MC-12 and at the haplotypes of Jonathan (Gao et al., 2008) it is possible to assume that this haplotype comes from Florina's mother Jonathan. As regards MC-20, no hypothesis can be done since this clone contains only intron-less Mal d 1 genes and no information are available on the alleles of Jonathan for this genes.

Genomic organization of Mal d 1 genes on LG16

The genomic organization of the Mal d 1 gene cluster of isoallergens on LG16 was investigated by the availability of the two BAC clones sequences. First of all, it should be noted that all the four intronless isoallergens are on the clone MC-20 and located in a region of just around 30 kb. The seven isoallergens on MC-12 are spread in a region of about 75 kb. The average distance among isoallergens is ranging from 10 to 15 kb. Two exceptions are the distance between Mal d 1.11 and Mal d 1.06B (~5 kb) and between Mal d 1.07 and Mal d 1.09 that are the two closest isoallergens in this cluster (~2 kb). Furthermore, in both clones all the isoallergens are in the same direction inside the clone (headed in direction Sp6-end) except Mal d 1.06A on MC-12 and Mal d 1.03G on MC-20 that are in the opposite direction (Figure 4). These results might be useful to better understand the further gene expression data since it is reported that the intergenic regions and the gene orientation can influence gene expression levels. For instance, in Arabidopsis it was demonstrated that a small intergenic region is able to drive tissue-specific expression of two adjacent genes (Bondino and Valle, 2009).
Table 7 Predicted Mal d 1-like ORFs on MC-12 analyzed with BALSTN software.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Mal d 1 gene/allele</th>
<th>Direction</th>
<th>Full-length (nt)</th>
<th>CDS (nt)</th>
<th>Intron/Exon</th>
<th>Accession number</th>
<th>Description</th>
<th>SNPs in CDS</th>
<th>SNPs in intron</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF4</td>
<td>Mal d 1.06A.02</td>
<td>↑</td>
<td>602</td>
<td>480</td>
<td></td>
<td>AY827697</td>
<td>Malus x domestica clone 220903B10 Mal d 1.06A02 (Mal d 1.06A) gene</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF8</td>
<td>Mal d 1.10ª</td>
<td>↓</td>
<td>961</td>
<td>486</td>
<td></td>
<td>AY428588</td>
<td>Malus x domestica clone 1m Mal d 1-like mRNA</td>
<td>0</td>
<td>not available</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF11</td>
<td>Mal d 1.02.01</td>
<td>↓</td>
<td>651</td>
<td>480</td>
<td></td>
<td>AY827654</td>
<td>Malus x domestica clone 250803B10 Mal d 1.0201 (Mal d 1.02) gene</td>
<td>1</td>
<td>0</td>
<td>1.00E-156</td>
</tr>
<tr>
<td>ORF15</td>
<td>Mal d 1.12ª</td>
<td>↓</td>
<td>861</td>
<td>486</td>
<td></td>
<td>AY822733</td>
<td>Malus x domestica clone 231103G3 Mal d 1.03G01 (Mal d 1.03G) gene</td>
<td>-</td>
<td>-</td>
<td>3.00E-25</td>
</tr>
<tr>
<td>ORF18</td>
<td>Mal d 1.11ª</td>
<td>↓</td>
<td>702</td>
<td>495</td>
<td></td>
<td>AY428589</td>
<td>Malus x domestica clone 1n Mal d 1-like mRNA</td>
<td>0</td>
<td>not available</td>
<td>4.00E-160</td>
</tr>
<tr>
<td>ORF21</td>
<td>Mal d 1.06B.02</td>
<td>↓</td>
<td>633</td>
<td>480</td>
<td></td>
<td>AY827712</td>
<td>Malus x domestica clone 220903F10 Mal d 1.06B02 (Mal d 1.06B) gene</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF23</td>
<td>Mal d 1ps2.03</td>
<td>↓</td>
<td>283</td>
<td>-</td>
<td></td>
<td>AY827730</td>
<td>Malus x domestica clone 220903A7 Mal d 1ps2 pseudogene</td>
<td>19</td>
<td>-</td>
<td>2.00E-113</td>
</tr>
<tr>
<td>ORF26</td>
<td>Mal d 1.06C.02</td>
<td>↓</td>
<td>608</td>
<td>480</td>
<td></td>
<td>AY827725</td>
<td>Malus x domestica clone 231103F11 Mal d 1.06C02 (Mal d 1.06C) gene</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF27</td>
<td>Mal d 1ps2.04</td>
<td>↓</td>
<td>470</td>
<td>-</td>
<td></td>
<td>AY827730</td>
<td>Malus x domestica clone 220903A7 Mal d 1ps2 pseudogene</td>
<td>5</td>
<td>-</td>
<td>0.00</td>
</tr>
</tbody>
</table>

ªAllergen gene names proposed following the official allergen nomenclature

Table 8 Predicted ORFs on MC-20 analyzed with BALSTN software.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Mal d 1 gene/allele</th>
<th>Direction</th>
<th>Full-length (nt)</th>
<th>Intron/Exon</th>
<th>Accession number</th>
<th>Description</th>
<th>SNPs in CDS</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF42</td>
<td>Mal d 1.03G02</td>
<td>↑</td>
<td>480</td>
<td>Intronless</td>
<td>AY822733</td>
<td>Malus x domestica clone 231103G3 Mal d 1.03G01 (Mal d 1.03G) gene</td>
<td>5</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF43</td>
<td>Mal d 1.0703</td>
<td>↓</td>
<td>480</td>
<td>Intronless</td>
<td>AY822717</td>
<td>Malus x domestica clone 231103B3 Mal d 1.0701 (Mal d 1.07) gene</td>
<td>5</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF44</td>
<td>Mal d 1.0903</td>
<td>↓</td>
<td>480</td>
<td>Intronless</td>
<td>AY822721</td>
<td>Malus x domestica clone 231103C1 Mal d 1.0903 (Mal d 1.09) gene</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>ORF46</td>
<td>Mal d 1.0801</td>
<td>↓</td>
<td>480</td>
<td>Intronless</td>
<td>AY822719</td>
<td>Malus x domestica clone 231103F3 Mal d 1.0801 (Mal d 1.08) gene</td>
<td>0</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Comparison of genetic and physical map of the *Mal d 1* cluster on LG16

Six of the ten new SSR markers (reported in Table 9) developed on BAC clone sequences, have been mapped on LG16 of the molecular marker linkage map of Durello di Forlì × Fiesta (Figure 4A and B). The remaining 4 SSRs were not polymorphic in that cross. The location of SSRs allowed to anchor the physical map to the genetic map and confirmed the location of the two BAC clones to be on LG16. Since two recombinant plants were found between ssr744aMC-12 and ssr744bMC-12, the BAC orientation of MC-12 within the LG could be determined. The physical distance of about 30 kb between these two SSRs corresponded to a genetical distance of ±1.2 cM (Figure 4) resulting in a 25 kb/cM region. For SSRs developed on MC-20 no recombinant individuals were found hence the orientation of this clone is not yet defined.

Table 9 Primer pairs for SSR markers developed on MC-12 and MC-20.

<table>
<thead>
<tr>
<th>BAC clone</th>
<th>Primer name</th>
<th>Primer sequence (5'-3')</th>
<th>Position (nt)¹</th>
<th>Ta (°C)</th>
<th>Segregation in DuXiF1 population²</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC-12</td>
<td>SSR384For</td>
<td>CCAATATCCCTCTCCCACAAC</td>
<td>10.083-10.248</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR384Rev</td>
<td>AGATGGGAGGCTGTGTAAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744AFor</td>
<td>TCAACATCCAAGCCACACAA</td>
<td>49.406-49.612</td>
<td>57</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744ARev</td>
<td>GGTGTTTCTGTAGCCTGCTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744BFor</td>
<td>CCAATATCCCTCTCCCACAAC</td>
<td>78.221-78.418</td>
<td>57</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744BRev</td>
<td>CATAGCTTCTCCACGCAGCT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744CFor</td>
<td>CCGTACAGAGCAGGAGGATTG</td>
<td>102.343-102.529</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR744CRev</td>
<td>TCCAATGGTGACAAATTTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR316For</td>
<td>CCGTAACTCTTTTTGGGATA</td>
<td>115.950-116.110</td>
<td>60</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-12</td>
<td>SSR316Rev</td>
<td>CATTACACAGGAGGAGGAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR1aFor</td>
<td>TTTGCAAAGGATGGATTGACT</td>
<td>5.841-6.048</td>
<td>60</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR1aRev</td>
<td>GCAATGGGAGGCTGTGTAAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR1For</td>
<td>AATCGGAGGTGAAATCGGTG</td>
<td>14.888-15.154</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR1Rev</td>
<td>CAGTTCTCTGGGAGGAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR3For</td>
<td>AAGGAGGAGGCTGTGTAAGG</td>
<td>40.986-41.169</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR3Rev</td>
<td>TCGGGTCTCTGGGAGGAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR2For</td>
<td>AATCGGAGGTGAAATCGGTG</td>
<td>114.473-114.647</td>
<td>60</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR2Rev</td>
<td>TGGTTCTCTGGGAGGAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR4For</td>
<td>GTGCGACTGAGTTTCTGTG</td>
<td>127.180-127.365</td>
<td>60</td>
<td>Yes</td>
</tr>
<tr>
<td>MC-20</td>
<td>SSR4Rev</td>
<td>TCATTCAACAGGAGGAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Positions count from the T7-End of Florina BAC clones sequences
²All the SSR resulted polymorphic in Durello di Forli

Combining the whole physical map obtained with the BAC sequences and the genetic map obtained with SSR markers of the *Mal d 1* cluster on LG16, the genetic distance of around 2 cM corresponds to a physical region of at least 260 kb (Figure 4). Comparing the order of *Mal d 1* genes in the previous map of LG16 (Gao et al., 2005) and the physical order of the isoallergens reported here, there is an overall agreement but also three discrepancies were found (Figure 1 for the genetic map and 4 for the physical map). Firstly, on the genetic map, *Mal d 1.09* was located above the SSR marker CH05a04 and at 1.5 cM from *Mal d 1.07* and *Mal d 1.08* but, in the physical map it is located between *Mal d 1.07* and *Mal d 1.08.*
Moreover, in the BAC sequences the marker CH05a04 was not found. Secondly, Mal d 1.02 was mapped 0.4 cM from Mal d 1.06A, Mal d 1.06B and Mal d 1.06C but but on the physical map Mal d 1.02 is located between Mal d 1.06A and Mal d 1.06B. Thirdly, on the genetic map Mal d 1.04 co-localized with Mal d 1.02, Mal d 1.07 and Mal d 1.08 whereas on the physical map Mal d 1.04 could not be traced despite the presence of Mal d 1.02, -1.07 and -1.08. One possible explanation for this absence could be the fact that the two clones do not overlap each other and so Mal d 1.04 can be located in the unsequenced region between MC-12 and MC-20. These results confirm that the fine positioning in a molecular marker linkage map of highly homologous members of a gene family is very cumbersome.

Analysis of upstream regions of Mal d 1 genes

Many findings suggest that a complex network of regulating elements contributes to modulate PR-10 gene expression (Liu and Ekramoddoullah, 2006). Walther et al. (2005), by analyzing Arabidopsis expression datasets, hypothesized that genes differentially expressed in response to several different stimuli, as is known for PR protein genes, should contain a greater number of distinct transcription factors (TFs) binding sites (or cis-elements) in their upstream regions than genes that respond to relatively few stimuli. The analysis of upstream regions of each Mal d 1 isoallergens of the BAC sequences seems to confirm this hypothesis because several putative motifs have been found in all promoter fragments. In particular, some cis-elements related to responses to biotic and abiotic stress were retrieved in all the promoter regions, such as i) GATA-boxes, which are a short (GATA) repeat known to be conserved among several light-responsive promoters (Reyes et al., 2004), ii) TGAC-containing W-boxes, which are specifically recognized by salicylic acid (SA)-induced WRKY DNA binding proteins and are known to be responsible for responses to fungal elicitors (Yu et al., 2001), iii) GT-1 elements (GGTAAA) which are consensus binding sites in many light-regulated genes (Terzaghi and Cashmore, 1995). Additionally, other motifs were found in all the promoter regions but one like iv) the characteristic motif of the Sequences Over-Represented in Light-Induced Promoter (SORLIP, Hudson and Quail, 2003), which was only absent with Mal d 1.11, v) the ERD-element (ACGT), a cis-element involved in dehydration stress response and dark-induced senescence (Simpson et al. 2003) which lacks only with Mal d 1.10, vi) a low temperature responsive element (LTRE, ACCGACA, Maestrini et al., 2009) lacked only with Mal d 1.02. Finally, An ARF (auxin response factor) element, that is a binding site for the promoters of primary/early auxin response genes of Arabidopsis thaliana (GAGACA, Guilfoyle and Hagen, 2007), was found only in the promoter of Mal d 1.10. Although it remains a challenge to distinguish potential cis-elements that serve as genuine TF binding sites from genomic background noise, the combining differential gene expression patterns with the analysis of cis-elements in the Mal d 1 promoters will undoubtedly contribute to elucidating the regulating mechanisms of Mal d 1 genes. This part will be addressed more in detail in Chapter 3 and 4.
Deduced amino acid sequences

The alignment of Mal d 1 deduced amino acid sequences (one sequence for isoallergen) is reported in Figure 5. The most conserved part in the sequences of PR-10 proteins is a glycine-rich loop (P-loop motif GXGGXGXXK) (Spangfort et al. 1997) that is frequently found in nucleotide-binding proteins (Radauer et al., 2008). In fact, the threedimensional structure of Bet v 1, investigated by X-ray crystallography and magnetic resonance spectroscopy (Gajhede et al., 1996), revealed that the P-loop is involved in the formation of a large cavity in the protein. The alignment shows that this region is quite conserved as only just few substitutions appeared in this domain among Mal d 1 proteins: a lysine replaced by a glutamine in Mal d 1.08 and by a methionine in Mal d 1.09. Also the third glycine is replaced with a glutamic acid in Mal d 1.11 and with a arginine in Mal d 1.12. Moreover, the alignment showed some gaps among the sequences, for instance the gap of three amino acids between Mal d 1.11 and the majority of the other isoallergens. The division into subfamilies based on nucleotide sequences was also reflected at the proteomic level, as it is clear looking at the phylogram tree (Figure 6). Interestingly, Mal d 1.10 entered in the same group of subfamily II but Mal d 1.11 and Mal d 1.12 formed a new group which we define here as subfamily V.

Regarding the characteristics that confers the ability to induce allergic responses in susceptible individuals, is still missing complete understanding. A high sequence similarity among proteins increases the chance of shared epitopes, whereas a single amino acid change may drastically influence the extent of allergenicity. Wagner et al. (2008) showed that few amino acid changes between three Bet v 1 isoforms were located on the surface of the proteins causing a difference in the IgE induction. Bet v1.0401 and Bet v 1.1001 do not induce IgE synthesis while Bet v 1.0101 can induce IgE that only partly cross-reacts with the 2 other isoforms. Bet v 1.01 and Bet v 1.04 were included in the Mal d 1 alignment (Figure 4). For instance, in position 8 the isoforms Bet v 1.01 presented a “T” like Mal d 1.04, -1.05, -1.06A/B/C and -1.10, while only the hypoallergenic variant Bet v 1.04 presented an “I”. In position 57 the Bet v 1.01 has a “S” and the hypoallergenic variant Bet v 1.04 had a N. Also the apple isoforms Mal d 1.03A/B/C/D/E/F/G, Mal d 1.07 and Mal d 1.11 presented a N in that position. Moreover, in a previous crystallographic study of Bet v 1-antibody complex (Ghosh and Bhattacharya, 2007) it was found that the relevant antigenic surface for these class of allergens consist of 16 residues, a continous stretch (E45-T52) along with a few others (R70, D72, H76 and K97). In particular, it is known that E45 residue is located centrally in the binding pocket of Bet v 1a and fits well into the groove on the antibody sufrace. As it is reported in Figure 5, the residue in position 45 is conserved both in apple and birch among all the amino acid sequences reported, but Mal d 1.11.
Figure 5: Alignment of predicted amino acid sequences of Mal d 1 isoforms. For the isoforms retrieved from the database the accession number is reported. The isoforms derived from the BAR library sequences are indicated with –bar. Also two Bet v 1 isoforms were included: Bet v 1.01 and Bet v 1.04 that is considered a hypoallergenic variant. Hydrophobic residues crucial for the binding activity are indicated big black boxes. Important amino acid substitutions among Mal d 1 sequences are indicated with red circles; substitutions between Bet v 1 sequences are indicated with small black boxes and imporant sequences and amino acid putatively important for the IgE recognition are indicated with arrows.
In this isoallergen the $E_{45} \rightarrow S_{45}$ substitution is reported. In the work of Spangfort et al. (2003) they performed the same artificial mutation ($E_{45} \rightarrow S_{45}$) on Bet v 1 and no effect on either local or global conformation of the protein was reported but a substantially reduced capacity to bind human IgE was found. These results combined with the substitution in position 57 reported above, let suppose that Mal d 1.11 may be an hypoallergenic isoforms. Proteomic experiments are needed to confirm this hypothesis. Moreover, in a recent work by Zaborsky et al. (2010) crystallographic studies and dynamic light scattering revealed that Bet v 1.04 demonstrated a high tendency to form aggregates due to a serine to cysteine exchange at residue 113. These aggregation of Bet v 1.04 triggers the establishment of a protective IgE titer and supports the use of Bet v 1.04 as a promising candidate for specific immunotherapy of birch pollen allergy. According to the similarity among Mal d 1 and Bet v 1 at the nucleotidic and amino acidic levels, it is likely also an oligomerization at least for some Mal d 1 proteins with the consequent influence in their ability to provoke the allergic reaction but any data is available for apple.

As regard the biological function, in the Bet v 1 structure some hydrophobic residues were identified as crucial for this binding activity such as $V_{67}$, $I_{85}$, and $I_{98}$ (Ghosh and Bhattacharya, 2007). These amino acid show to be conserved in both allergenic and non-allergenic members of the Bet v 1 family suggesting a similar biological function for them. On the contrary, some substitutions were found in these positions for Mal d 1 isoforms: Mal d 1.11 for the position 67 ($V_{67} \rightarrow A_{67}$), Mal d 1.12 for the position 85 ($I_{85} \rightarrow V_{85}$) and Mal d 1.02 and Mal d 1.10 for the position 98 ($I_{98} \rightarrow V_{98}$).

![Phylogram tree](image)

Figure 6: Phylogram tree derived from the alignment reported in Figure 5. The gray arrows indicated the new isoforms for LG16.

Other genes in the cluster

Three ORFs similar to known genes but not belonging the Mal d 1 gene family were found on MC-20. ORF49 is highly homologous to an HEAT repeat-containing protein. The HEAT domain is a tandemly repeated element occurring in a number of cytoplasmic
proteins, including elongation factor 3 (EF3), the 65 Kd alpha regulatory subunit of protein phosphatase 2A (PP2A) and the yeast PI3-kinase TOR1 (Ma and Bork, 1995). Arrays of HEAT repeats consists of 3 to 36 units forming a rod-like helical structure and appear to function as protein-protein interaction surfaces. It has been noted that many HEAT repeat-containing proteins are involved in intracellular transport processes. This could be correlated with the role of binding and transporter of plant hormones proposed for Mal d 1 proteins (Fernandes et al., 2008) and their close position in the genome might be due to the need of a coordinated regulation of these genes.

**ORF50** is homologous to a multi-domain protein consisting of a sensor histidine kinase and a response regulator, acting in response to environmental changes. In this two-component system the signal is transduced from histidine kinase to response regulator through phosphoryl transfer, which is a quite well know cell signaling mechanism (Yamada and Shiro, 2008). Mal d 1 proteins, as hormones transporter (i.e. cytokinins, brassinosteroids) are considered also important in signal transduction during the response to biotic and abiotic stresses (Zubini et al., 2009) and so their role could be related to the flanking multi-domain protein.

The last protein (**ORF52**) resulted similar to a COBRA protein that has been identified previously as a potential regulator of cellulose biogenesis. Cellulose microfibrils are the primary anisotropic material in the cell wall and thus are likely to be the main determinant of the orientation of cell expansion. Hence, this protein is important for the cell since the orientation of cell expansion is a process at the heart of plant morphogenesis (Wasteneys and Fujita, 2006). It is known that Mal d 1 can bind hydrophobic ligands like as the components of cellulose (Koistinen et al., 2005) so it is not possible to exclude an action of the COBRA protein on the regulation and/or the function of Mal d 1 proteins.

Furthermore, in the ORFs identified as hypothetical proteins (See Table 5 and 6) some conserved domain were found with BLASP. On MC-12, **ORF31** is highly homologous to an hypothetical protein of *Vitis vinifera* with an Heavy-Metal-Associated (HMA) domain that is a conserved domain of approximately 30 amino acid residues found in a number of proteins that transport or detoxify heavy metals, as the CPx-type heavy metal ATPases or copper chaperones (Zhou et al. 2009). Many metal-responsive transcriptional regulators have been described and the proteins containing this domain are quickly induced after different kinds of stress as cold and drought or during leaf senescence (Barth et al., 2004). **ORF32** is highly similar to an hypothetical protein of *Vitis vinifera*. It contains two conserved domains: a domain of for? the glycosyl hydrolase of family 28 and a WD-40 domain. WD-40 domain is a conserved domain found in a number of eukaryotic proteins that cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly (Stacey et al., 1999). Considering that the HEAT repeat-containing protein is important in the protein-protein interaction and that histidine kinase, a response regulator, HMA domain and WD-40 domain are directly related to the signal transduction pathway, the presence of these other genes close to the Mal d 1 cluster
let suppose an important role for all this genomic region during the response of the plant to external and internal stimuli. Further functional studies are needed to confirm this hypothesis.

Finally, a large accumulation of retrotransposons and retrotransposon-like elements was observed in this region (Eickbush et al., 2008). In particular they were 10 of 26 ORFs with a certain similarity with known proteins (for MC-12 the ORF1, -3, -9, -14, -24, -29, -30 and for MC-20 the ORF33, -35, -41), like as reverse transcriptase (RNA-dependent DNA polymerase), putative NTP binding site, RNase H, RNA/DNA hybrid binding site, integrase core domain, CHRromatian Organization MOdifier domain (CHROMO domain) and retrotransposon gag proteins (Tables 5 and 6). Other three ORFs showed similarity to hypothetical proteins of *Vitis vinifera* but with conserved domain related to retrotransposons (for MC-12 the ORF24 with a conserved integrase core domain and a reverse transcriptase domain; for MC-20 the ORF36 with a conserved Plant Mobile domain and the ORF40 with another reverse transcriptase domain). Retrotransposons, the most common class of transposable elements, represent a major fraction of the repetitive DNA of most eukaryotes and it is known that retrotransposon rearrangements play an important role in the plasticity of eukaryotic genomes. In fact, the presence of many retrotransposon elements could also be an indication of the evolution of *Mal d 1* gene family. It should be noted that apple belongs to the *Maloideae* family and, as all the *Maloideae* members, it is considered to be of polyploid origin. Therefore, it is possible to assume that the expansion of the family occurred during this polyploidization of the apple genome because it is known that this process can activate retrotransposons with consequent genes duplication (Madlung et al., 2005). The similar location of *Mal d 1* clusters on of two homeologous LG (13 e 16) enforces this hypothesis. Moreover, other gene clusters have been detected in apple as *O*-Methyltransferase genes (Han et al., 2007) or F-Box genes (Sassa et al., 2007), and the expansion of these loci has also been partly attributed to the activity of retrotransposons.

For what concern the presence of many homologous genes in the genome, it is thought that the pressure to conserve protein sequences and structures is associated with an increase in copy number of certain genes during evolution. This could be the case of *Mal d 1* genes because the duplication rates of this genes may be partly explained by pathogen-mediated-selection, considering that the most rapidly evolving families were associated with pathogen defense (Wagner et al., 2008). Also recent analysis of the complete eukaryotic Arabidopsis genome sequence provided evidences that the gene decay rates of homologous genes is biological function-dependant. In particular, Maere et al. (2005) found that the gene copies encoding regulatory proteins (transcription factors, proteins with kinase activity or binding activity) and signal transduction proteins, tend to retain when generated by polyploidization. This finding supports both the hypothesis regarding the formation of *Mal d 1* family during the apple genome polyploidation and the hypothesis that many copies of *Mal d 1* genes were maintained during evolution due to their role in signal transduction.
Finally, it has been reported that the retrotransposon-rich region of a chromosome may reflect its centromeric location (Ma et al., 2007) so the retrotransposons abundance found here let suppose a centromeric position in the chromosome for Mal d 1 cluster. On the contrary, the finding of large stretches of DNA related to a single cM for this region make not realistic the hypothesis its centromeric location. Other specific analysis are needed to resolve this question.
Conclusions

In this work, the study of the *Mal d 1* gene family has been effectively addressed by the retrieval and sequencing of two BAC clones containing the cluster of *Mal d 1* allergen genes on the LG16 of the apple genome. This approach, based on the analysis of large contiguos blocks of DNA sequences, revealed to be suitable to study genes with a cluster organization and many findings regarding the number, gene orientation, physical distances and full-length sequences were obtained. Also the anchoring of the physical and the genetic map of the region has been successfully achieved. New *Mal d 1* nucleotidic sequences have been found on LG16 suggesting that other isoforms may be present in the apple genome and that the complexity of the genetic base of resistance and allergenicity will increase. It is likely that, through further study on the homeologous LG13, other new *Mal d 1* genes will find out. Therefore, the knowledge about the gene family composition and the physical positioning of homologous members within the family will be crucial for further association studies, as QTL mapping studies. New lights has been thrown also on the *Mal d 1* gene cluster organization and evolution. In fact, in this work emerged the hypothesis that the duplication of *Mal d 1* members could have occurred during the polyploidation of the apple genome, as the presence of many retrotransposons elements in the cluster. Why the retention of many homeologous copies during evolution occurred is not clear yet. The evolution of new functions or the distribution of existing ones among isollergens may be a possible explanation, together with a selection-driven preservation of all the copies. For *Mal d 1* genes, in particular, a pathogen-mediated selection can be involved in the evolution of this gene family. Understanding the functions of each isoallergen and their peculiar mode of action (expression profile, activity, regulation) in the plant and in response to different biotic and abiotic stimuli will be the challenge of further studies.
References

Chapter 1


Zubini P, Zambelli B, Musiani F, Ciurli S, Bertolini P, Baraldi E. 2009. The RNA hydrolysis and the cytokinin binding activities of PR-10 proteins are differently performed by two isoforms of the Pru p 1 peach major allergen and are possibly functionally related. Plant Physiol. 150:1235-47.
Chapter 2

Development of a PCR-based tool for expression analysis of 20 individual *Mal d 1* genes
Introduction

Results of the whole-genome and EST sequencing projects clearly pointed out that in plants genomes many genes exist in multiple copies as members of multi-gene families. A variety of parameters could be used to define gene families but those most commonly accepted are the degree of primary sequence identity and shared functional motifs or domains (Wu et al., 2003). This extensive gene redundancy seems to have arisen from genome polyploidization and it has been particularly significant in the evolutionary history of flowering plants (Vision et al., 2000). In the Arabidopsis genome, for example, an high percentage of annotated genes (65%) is belonging to gene families (Wortman et al., 2003).

An important characteristic of plant food allergens is their being part of gene families (Shewry et al., 2002). Bet v 1 is the major birch pollen allergen family encoding for pathogenesis-related proteins 10 (van Loon et al., 2006) and allergens belonging to the Bet v 1-like gene family are the most abundant and widely spread in plants. This family, together with the profilin family, accounts for more than 65% of all food allergens (Hoffmann-Sommergruber and Mills, 2009). In the apple genome, the Bet v 1-like allergens are grouped in the Mal d 1 family (Breiteneder and Radauer, 2004).

The genomic organization of Mal d 1 gene family was firstly examined by Gao et al. (2005) through the identification and mapping of genomic sequences. A total of 18 Mal d 1 genes (from Mal d 1.01 to Mal d 1.09) were identified, sixteen of which are organized in a duplicated cluster located on the two homeologous linkage groups (LGs) 13 and 16. In Chapter 1, the genomic organization of the Mal d 1 gene cluster on LG 16 has been further investigated by obtaining most of the sequence of this genomic region and its anchoring to a genetic map. New findings regarding number, gene orientation, physical distances and full-length sequences were obtained. Most notably, the three previously known but unmapped isoallergens Mal d 1m, Mal d 1n and Mal d 1.03G (Beuning et al., 2004; Gao et al., 2005), were mapped on LG16. Moreover, an isoallergen never described before was found for this cluster. According to the official allergen nomenclature (King et al., 1995), the isoallergen genes Mal d 1m, Mal d 1n and the new isoallergen have been named Mal d 1.10, Mal d 1.11 and Mal d 1.12, respectively (Chapter 1). Twenty Mal d 1 loci able to code for complete proteins were identified up to now and they share an high sequence homology, both at nucleotidic (from 54 to 98%) and proteomic (from 45 to 100%) level. An high similarity between Mal d 1 and Bet v 1 sequences was already reported (between 56% and 68% of amino acid similarity, Bohle, 2007) and this similarity is considered as the base of the IgE-mediated cross-reactivity between birch and apple: after primary sensitization to the pollen allergen Bet v 1, the majority of birch allergic patients tend to develop allergic symptoms also after the ingestion of the Mal d 1 proteins present in apple (Yagami, 2002). Today, up to 70% of birch-pollen-sensitized patients suffer from an oral allergy syndrome after eating apples (Fernandez-Rivas, 2003).
Differences in allergenicity among apple cultivars were reported by many authors (Vieths et al., 1994; Son et al., 1999; Bolhaar et al., 2006; Kootstra et al., 2007). This finding proved the importance of genetic factors in the determination of apple allergenicity and raised questions on its genetic base. Nowadays, there is the hypothesis that the qualitative characteristics of the different proteins belonging to the Mal d 1 family can contribute, together with the other Mal d apple allergens, in the determination of the allergy degree, as can be argued from the differences in binding capacity of birch pollen-specific IgE to two Mal d 1 proteins (Bolhaar et al, 2005) or from the association of particular protein variants (Mal d 1.04 and Mal d 1.06A) to apple allergenicity (Gao et al., 2008). Most notably, it has to take in consideration that the differences in total amount of just Mal d 1 could not explain differences in allergenicity among apple cultivars (Asero et al. 2006). In fact, the question regarding the the impact of quantity and quality of Mal d 1 isoallergens has not yet been solved. Some proteomic and transcriptomic expression studies on Mal d 1 are available in literature. Some authors used anti-Mal d 1 or anti-Bet v 1 antibodies to measure the Mal d 1 total content in different apple cultivars and tissues (Marzban et al., 2005; Zuidmeer et al., 2006; Herndl et al., 2007) but the quantitation of Mal d 1 proteins in fruits was limited to the detection of the total amount, without distinguish among the isoforms. Moreover, proteomic studies of plant allergens are faced by a number of obstacles which are caused by the nature of plant material, as a low protein content accompanied by a plethora of proteases and interfering compounds (Oberhuber et al., 2008). Quantitative Real Time PCR (qPCR) was also used to investigate the expression of some isoallergens (Mal d 1.01, 1.02, 1.03 and 1.04 ) in fruit of different cultivars, in different growing conditions or stage of development (Puehringer et al., 2003; Botton et al., 2008; Pagliarani et al., 2009) but the specificity of the primers was not always sufficient. Moreover, it has to be considered also that the qPCR approach provides a description of the mRNA levels of regarding the precise moment when the sample are collected but does not take into account the possible protein accumulation in the cell. All these studies have the common result that only a limited number of different Mal d 1 genes proteins and mRNAs were traced back in apple fruit (Helsper et al., 2002; Puehringer et al., 2003; Beuning et al., 2004; Gao et al., 2008). This latter issue highlight the importance of the researches on the expression of specific Mal d 1 genes to investigate the behavior of each isoallergen in different cultivars, fruit tissues or stress conditions.

Although the Mal d 1 genomic resources available in literature and acquired in Chapter 1 can be extremely useful in taking an informative picture of the genomic organization of this gene family, the ‘functional’ characterization of each gene is still a challenge. Due to the high sequence homology among individual members, a specific technology must be employed to allow gene specific functional studies. Currently qPCR is the most precise and quick method for measuring gene expression (Larionov et al., 2005), especially to distinguish among highly similar genes belonging to gene families. Therefore, to unravel the complexity of Mal d 1 expression, a set of twenty highly specific primer pairs
for qPCR was developed, each able to recognize and amplify the alleles of only one gene. Single nucleotide polymorphisms (SNPs) were exploited to distinguish among isoallergen genes, since they constitute the most common DNA sequence variations found in genomes of most organism, including the apple (Newcomb et al., 2006). Accurate validation and optimization were performed to ensure and document specificity. Moreover, it is known that the qPCR method generates a large amount of raw numerical data and processing may notably influence final results. In particular, the data processing can be based either on standard curves or on the comparative Ct method. At the moment, the comparative Ct method, that is based mainly on the assessment of the PCR efficiency, is preferred in relative PCR. It permit the analysis of changes in gene expression in a given sample relative to another reference sample (such as an untreated control sample). On the contrary, the standard curve is often used for absolute PCR. It is based on creation of a standard curve and then it is possible to compare unknowns to the standard curve and extrapolate a value. In this work also an accurate analysis to choose the best method for processing the qPCR data was performed. A preliminary gene expression analysis has also been performed for apple fruits and leaves to investigate the tissue-specific expression of each Mal d 1 isoallergen. The development of this Mal d 1 expression tool is the base for the following studies of specific gene expression profiling under a range of plant tissues and conditions, as will be reported in Chapter 3 and 4.
Materials and Methods

Mal d 1 sequences
For this study, Mal d 1 sequences were retrieved from the literature (Gao et al., 2005; 2008), from the key-word research in the GeneBank database and from Chapter 1. All the different sequences are listed in Appendix 2.

Alignment and primer design
Fine isoallergen coding sequences alignments was performed with the software Lasergene® v8.0-MegAlign (DNASTAR, Inc., Madison, WI USA) looking for SNPs among Mal d 1 sequences. For sequence alignment, all the available allelic variants indicated in Appendix 2 were included. Firstly, regions conserved among the alleles of a gene/locus but variable among genes/loci were identified that were. Secondly, specific primer pairs were designed in SNPs-containing regions with the software Primer3 version 0.4.0 (http://frodo.wi.mit.edu/primer3/). To ensure high specificity and efficiency during qPCR amplification, generally a stringent set of criteria was used for primers design (Udvardi et al., 2008) but sometimes the choice was constrained by the SNPs position. These main criteria included the positioning of the primers in 3' untranslated region (3'UTR), the primer lengths of 18-24 nucleotides, a guanine-cytosine content of 20-80% and PCR amplicon lengths of 80-200 base pairs. Each primer pair was also tested with the software PrimerSelect® v8.0-MegAlign for the formation of primer homo and heterodimers.

qPCR conditions and specificity validation of primer
Designed primers were first tested on gDNA of Florina for their ability to produce an amplicon. All the PCR amplifications were performed in a 17.5 µl volume containing 50 ng of DNA, 100 nM gene-specific primers, 1.5 mM MgCl2, 100 µM dNTPs, 0.5 Unit AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1X reaction buffer. The reaction included an initial 10 min denaturation at 95°C, followed by 33 PCR cycles (45 s at 60°C, 2 min at 72°C, and 30 s at 95°C), with a final extension of 7 min at 72°C. The amplicons were visualised on an Image Station 440 CF (Kodak, Rochester, N.Y., USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining.

Next, their specificity was validated through four different steps: 1) using the BLAST N against the available databases 2) checking their ability to give amplification only on the corresponding BAC plasmid DNA, 3) checking their ability to produce a single melting curve peak during gene expression study, and 4) sequencing of Florina gDNA amplicons and checking the correspondance of the retrieved sequences and the targeted gene, and 5) amplifying gDNA of other genotypes at the optimized conditions to check their ability to produce an amplicons also on other tempaltes. Below, these validation steps are further elaborated.
1) **In silico analysis of Mal d 1 primer pairs**

In order to further validate the specificity of the primers against a wider range of alleles, an *in silico* analysis was performed using the software BLAST N (Altschul et al., 1997). By the insertion of our primer sequences as input, the perfect correspondence of the output sequences only with the targeted gene was evaluated.

2) **Specificity on BAC clones**

Firstly, basic qPCR conditions reactions were performed in triplicate in a final volume of 10 µl, containing 5 µl of Power SYBR® Green Master Mix 1X, 100 nM of each primer, PCR-grade water and 1 ng of plasmid DNA or a pool of 1:9 cDNA from apple fruits and apple leaves. Reactions were incubated at 50 °C for 2 min and at 95° for 5 min to activate the AmpliTaq Gold® DNA Polymerase, followed by 40 cycles at 95°C for 15 sec and 60°C for 1 min. The plasmid DNA used for this first specificity test derived from the BAC clones known from Chapter 1. Each primer pair was tested on two different BAC plasmids: one containing the specific gene targeted by the primer pair (positive control) and one containing other *Mal d 1* genes (negative control). The presence or absence of amplification was evaluated according to the detectability of the raw dye fluorescence by the qPCR machine. Where amplification was detected also in negative control, the conditions were further optimized in order to increasing gene specificity by adjusting primer concentration (from 100 to 70 nM) and annealing temperature (from 60 to 63°C).

3) **Production of a single melting curve peak during the gene expression study**

As described above, plasmid DNA and cDNA from Florina fruits and leaves were amplified by qPCR. To ensure the absence of aspecific PCR products and primer dimers, an heat dissociation protocol (from 60°C to 95°C) was also performed and a dissociation curve for each samples was generated. The StepOne Software version 2.1 (Applied Biosystem) was used to analyse the fluorescence data. The production of a single melting curve peak for a sample means that the amplicon is specific and not a mix of different amplicons.

4) **Amplicon direct sequencing**

The primers specificity for each *Mal d 1* isoallergen was further examined by the direct sequencing of PCR products starting from gDNA of Florina. All the PCR amplifications were performed in a 17.5 µl volume containing 50 ng of gDNA, optimized concentration of gene-specific primers, 1.5 mM MgCl₂, 100 µM dNTPs, 0.5 Unit AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1X reaction buffer. The reaction included an initial 10 min denaturation at 95°C, followed by 33 PCR cycles (45 s at the optimized annealing temperature, 2 min at 72°C, and 30 s at 95°C), with a final extension of 7 min at 72°C. The amplicons were visualised on an Image Station 440 CF.
(Kodak, Rochester, N.Y., USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining. Sequencing was performed by Bio-Fab Research srl (Pomezia, Italy). Sequences were analyzed both with Chromas Lite 2.01, BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Lasergene® v8.0-MegAlign (DNASTAR, Inc., Madison, WI USA) softwares to check the correspondence of the retrieved sequences and the targeted gene.

5) Test of Mal d 1 specific primers on other genotypes

In order to verify the ability of the primer pair to give amplification also on other genotypes, a qualitative test was performed. An end point PCR was performed with each primer pair on gDNA of Gala, Fiesta, Jonathan, Jonagold, Durello di Forlì using the optimized conditions. All the PCR amplifications were performed in a 17.5 µl volume containing 50 ng of DNA, 100/70 nM gene-specific primers (depending on the primer), 1.5 mM MgCl2, 100 µM dNTPs, 0.5 Unit AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1X reaction buffer. The reaction included an initial 10 min denaturation at 95°C, followed by 33 PCR cycles (45 s at 61/63°C depending on the primer, 2 min at 72°C, and 30 s at 95°C), with a final extension of 7 min at 72°C. Water-samples were added as negative controls. The amplicons were visualised on an Image Station 440 CF (Kodak, Rochester, N.Y., USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining.

Gene expression study

- Plant material, RNA extraction and cDNA synthesis

Apple fruits were collected at the Cadriano Experimental Station, Bologna University, Italy, from Florina trees at commercial harvest time, Apple skin and flesh were separately frozen in liquid nitrogen and stored at −80°C until RNA extraction. Fruits RNA extractions were carried out according to Pagliarani et al., 2009 starting from 6-8 g of frozen tissue. Florina plants, grafted on M9 were grown in greenhouse and inoculated with a suspension of V. inaequalis conidia (at least 2×10^5 conidia/ml). Young expanded leaves were collected at 0, 24, 48, 72, and 96 h post inoculation, stored separately at −80°C until RNA extraction. Total RNA was extracted from 1 g of leaves, according to Paris et al.,(2009) and quantified using a Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific,Wilmington, DE, USA). 40 µg of DNA-free RNA were treated with 10 Units DNasel (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK) at 37°C for 20 min. The cultivar Florina was chosen for this work because of the availability of the sequences and physical map of the Mal d 1 genes cluster on LG16 (Chapter 1). First-strand cDNA was synthesized according to Paris et al. (2009), starting from 1 µg DNA-free RNA. The cDNA was diluted 1:9 and its quality was verified by the amplification with actin specific primers.
- qPCR analysis

The amplifications were performed in 96-well plates with a StepOne™ Real-Time PCR instrument (Applied Biosystems, Foster City, CA). A pool of cDNA derived from leaves collected at different times after the inoculation with *V. inaequalis* and a pool of cDNA derived from fruit skin and flesh were tested. The reactions were performed in triplicate in a final volume of 10 µl, containing 5 µl of Power SYBR® Green Master Mix 1X, 70-100 nM of each primer, PCR-grade water and 1 ng of DNA or a pool of 1:9 cDNA from apple skin and flesh. Reactions were incubated at 50 °C for 2 min and at 95° for 5 min to activate the AmpliTaq Gold® DNA Polymerase, followed by 40 cycles at 95°C for 15 sec and 60/63°C for 1 min. The gene expression was evaluated as presence or absence according to the detectability of the raw dye fluorescence. The Ct values were used for a preliminary analysis of gene expression levels. To ensure the absence of aspecific PCR products and primer dimers, an heat dissociation protocol (from 60°C to 95°C) was always performed and a dissociation curve for each samples was generated. The StepOne Software version 2.1 (Applied Biosystem) was used to analyse the fluorescence data.

- Evaluation of primer pair relative efficiency

Amplicons obtained with each specific primer pair and with primers for actin have been used for the preparation of standard curves which consisted in a ten-fold dilution series of the amplicons over six dilution points. These standard curve samples were used as reference for qPCR amplifications with all the specific primer pairs as described above. The optimal threshold was chosen automatically by the StepOne Software version 2.1 (Applied Biosystem) and was used to calculate the threshold cycles (Ct) value for each standard curve point. Ct values in each dilution were measured in duplicate and were plotted against the logarithm of their initial template concentration. Each standard curve was generated by a correlation coefficient (R²) of the plotted points. From the slope of each standard curve, PCR amplification efficiency (E) was calculated according to the equation: 

\[ E = 10^{(-1/slope)} - 1 \] (Lee et al., 2006). To test the relative efficiency of each primer pair with *actin*, the ΔCt (C_T of the gene - C_T of actin) was calculated for the standard curve points and plotted vs the logarithm of their initial template concentration. The slopes of these charts were evaluated: if the absolute slope value results < 0.1, the efficiency of the two systems is approximately equal and the comparative Ct method can be used for the analysis of the data, otherwise quantisation has to be performed using the standard curves method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

- Survey of DataBases for Mal d 1-related ESTs

In order to increase our knowledge on *Mal d 1* expression as to substantiate results of the previous expression study, an *in silico* analysis was performed on EST sequences as present in public Databases. Using the software BLAST N (Altschul et al., 1997), ESTs for
Mal d 1-related sequences were searched by the insertion of key-words or known Mal d 1 sequences as input. Since the high homology among Mal d 1-like sequences, an accurate analysis of the output sequences was performed by Lasergene® v8.0-MegAlign (DNASTAR, Inc., Madison, WI USA) software to distinguish among alleles and different genes. Sequences with 5% mismatch (gDNA) were considered as alleles. Only the Mal d 1.03 and -1.06 genes deviate from this general rule. Regard them, sequences with less then 5% dissimilarity can still be of different genes. This fact is reflected by the denotation of these genes as they are distinguished from each other not by numbers anymore (requiring the 5% threshold) but letters after a common number (Gao et al., 2005).
Results and discussions

Mal d 1 isoallergens alignment, primer design

A total of 63 different Mal d 1-related sequences, (Appendix 2) derived both from literature, databases and BAC clone sequencing (see Chapter 1) were aligned. Due to the high homology, their fine alignment was needed to highlight the SNPs for each isoallergen. Figure 1A reports the alignment of a specific fragment for part of the examined sequences and all the present SNPs.

![Alignment of Mal d 1 isoallergens](image)

Figure 1: A) Portion of a simplified Mal d 1 sequences alignment. Only one sequence for each isoallergen and BAC sequences (28 on 53 sequences) were included in this alignment. The reverse primers specific for Mal d 1.01 and Mal d 1.02 are showed in red boxes. The SNPs at the 3' end of each primer, indicated with red arrows, are important for the primer specificity. B) Portion of a complete alignment of all the allelic variants known for Mal d 1.01 reported by Gao et al., 2008 and the specific reverse primer for this gene.

The SNPs-containing regions were used to design a set of 20 primer pairs, each one specific for a different Mal d 1 gene (Table 1). These primers are suitable for qPCR gene expression analysis with the SYBR-Green chemistry. In fact, they all amplify short amplicons (from 79 to 200 bp) to ensure the efficiency of the Taq polymerase processivity.

In Figure 1A are reported as example the reverse primer sequences of Mal d 1.01 and Mal d 1.02.
1.02 isoallergens. Firstly, all the primers were tested with an \textit{in silico} analysis performed using the software BALST N. For all a perfect match have been found only with sequences corresponding to the targeted gene (all the allele variants). In Figure 1B is showed as primers specific for a gene cannot distinguish among allele variants. In particular the example of \textit{Mal d 1.01} reverse primers is reported. Secondly, the primer pairs were tested on gDNA of Florina and for all amplification was obtained (data not shown) validating their ability to produce an amplicon.

**qPCR conditions and specificity validation of primers**

The chemistry used in this study is not gene-specific since SYBR GREEN is a dye able to bind each double-stranded DNA fragment. Indeed, many efforts were spent to avoid the generation of aspecific amplicons by optimising PCR conditions. It was necessary to individually optimize the conditions for each qPCR reaction and to validate the specificity.

**Specificity on BAC clones**

In detail, the annealing temperature and the primers concentration for the qPCR amplifications were adjusted for each primer pair in order to obtain an amplification signal only in the positive controls represented by the BAC clone containing the \textit{Mal d 1} isoallergen under study. The scenario for the optimized conditions is reported in Table 1. Some examples of the optimization steps are reported in Figure 2. For some isoallergens (i.e. \textit{Mal d 1.02}) it was sufficient to modify the annealing temperature (Figure 2A); for others (i.e. \textit{Mal d 1.03F}) also the primers concentration was changed to reach a specific amplification (Figure 2B). Sometimes, unspecific amplifications resulted in the water controls suggesting the need to slightly change the amplification conditions to avoid the primer dimers formation (i.e. for \textit{Mal d 1.11} in Figure 2C). As appears from Table 1, to guarantee the specificity of the amplification it was necessary, in some cases, to apply very stringent conditions (i.e. Ta=63 or 70 nM primers), despite the risk of reducing the efficiency of amplification and so the final expression value.
Table 1. Isoallergen-specific primer pairs designed for qPCR analysis. It continues on the next page

<table>
<thead>
<tr>
<th>Isoallergens</th>
<th>Primer name</th>
<th>Primer sequence (5'-3')</th>
<th>Start* (bp)</th>
<th>Lenght (bp)</th>
<th>Ta (°C)</th>
<th>Primer conc. (nM)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 1.01</td>
<td>qMd1.01/02F</td>
<td>GATTGAGAGGAGATGCTTTGACA</td>
<td>258</td>
<td>103</td>
<td>63</td>
<td>100</td>
<td>76.83</td>
</tr>
<tr>
<td></td>
<td>qMd1.01R</td>
<td>GTAATGACTGATGCTTTGATAG</td>
<td>258</td>
<td>111</td>
<td>62</td>
<td>100</td>
<td>79.06</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>qMd1.01/02F</td>
<td>GATTGAGAGGAGATGCTTTGACA</td>
<td>6</td>
<td>167</td>
<td>63</td>
<td>100</td>
<td>81.45</td>
</tr>
<tr>
<td></td>
<td>qMd1.02R</td>
<td>TTGTGTTGTAGTGGCTTGAAT</td>
<td>21</td>
<td>96</td>
<td>70</td>
<td>70</td>
<td>78.55</td>
</tr>
<tr>
<td>Mal d 1.03A</td>
<td>qMd1.03AF</td>
<td>ATCTGAGTTCACCTCCTGATT</td>
<td>70</td>
<td>258</td>
<td>111</td>
<td>62</td>
<td>77.44</td>
</tr>
<tr>
<td></td>
<td>qMd1.03AR</td>
<td>ACTGCTTGTTGTGAATTCTTT</td>
<td>276</td>
<td>79</td>
<td>63</td>
<td>100</td>
<td>79.5</td>
</tr>
<tr>
<td>Mal d 1.03B</td>
<td>qMd1.03BF</td>
<td>TGTTTTCACATACGAGTTGGAAT</td>
<td>15</td>
<td>156</td>
<td>62</td>
<td>70</td>
<td>79.5</td>
</tr>
<tr>
<td></td>
<td>qMd1.03BR</td>
<td>ATACGATCCGAGTTGCAATCTC</td>
<td>15</td>
<td>169</td>
<td>70</td>
<td>70</td>
<td>79.72</td>
</tr>
<tr>
<td>Mal d 1.03C</td>
<td>qMd1.03CF</td>
<td>ATCTGAGTTCACCTCCTGATT</td>
<td>70</td>
<td>258</td>
<td>111</td>
<td>62</td>
<td>77.44</td>
</tr>
<tr>
<td></td>
<td>qMd1.03CR</td>
<td>ACTGCTTGTTGTGAATTCTTT</td>
<td>276</td>
<td>79</td>
<td>63</td>
<td>100</td>
<td>79.3</td>
</tr>
<tr>
<td>Mal d 1.03D</td>
<td>qMd1.03DF</td>
<td>CACAGAATTGAGCGGTTGGAAT</td>
<td>208</td>
<td>119</td>
<td>70</td>
<td>70</td>
<td>77.29</td>
</tr>
<tr>
<td></td>
<td>qMd1.03FR</td>
<td>TCCAACCTTACATGCTTTTCTAA</td>
<td>337</td>
<td>122</td>
<td>62</td>
<td>70</td>
<td>78.32</td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>qMd1.04F</td>
<td>GGGTATGTTAAGCAAGAGCTCA</td>
<td>196</td>
<td>193</td>
<td>61</td>
<td>100</td>
<td>76.97</td>
</tr>
<tr>
<td></td>
<td>qMd1.04R</td>
<td>TGGTCTCTAATAGTGGCTTCTAGC</td>
<td>343</td>
<td>124</td>
<td>63</td>
<td>70</td>
<td>80.32</td>
</tr>
<tr>
<td>Mal d 1.05</td>
<td>qMd1.05F</td>
<td>ATCTGAGTTCACCTCCTGATT</td>
<td>70</td>
<td>258</td>
<td>111</td>
<td>62</td>
<td>77.44</td>
</tr>
<tr>
<td></td>
<td>qMd1.05R</td>
<td>ACTGCTTGTTGTGAATTCTTT</td>
<td>276</td>
<td>79</td>
<td>63</td>
<td>100</td>
<td>79.5</td>
</tr>
<tr>
<td>Mal d 1.06A</td>
<td>qMd1.06AF</td>
<td>CACAGAATTGAGCGGTTGGAAT</td>
<td>208</td>
<td>119</td>
<td>70</td>
<td>70</td>
<td>77.29</td>
</tr>
<tr>
<td></td>
<td>qMd1.06AR</td>
<td>TCCAACCTTACATGCTTTTCTAA</td>
<td>337</td>
<td>122</td>
<td>62</td>
<td>70</td>
<td>78.32</td>
</tr>
<tr>
<td>Mal d 1.06B</td>
<td>qMd1.06BF</td>
<td>CACAGAATTGAGCGGTTGGAAT</td>
<td>208</td>
<td>119</td>
<td>70</td>
<td>70</td>
<td>77.29</td>
</tr>
<tr>
<td></td>
<td>qMd1.06BR</td>
<td>TCCAACCTTACATGCTTTTCTAA</td>
<td>337</td>
<td>122</td>
<td>62</td>
<td>70</td>
<td>78.32</td>
</tr>
<tr>
<td>Mal d 1.06C</td>
<td>qMd1.06CF</td>
<td>GTCCTAAAGAACCTCAGAATTCTA</td>
<td>103</td>
<td>116</td>
<td>63</td>
<td>70</td>
<td>76.74</td>
</tr>
<tr>
<td></td>
<td>qMd1.06CR</td>
<td>GTCCTAAAGAACCTCAGAATTCTA</td>
<td>103</td>
<td>116</td>
<td>63</td>
<td>70</td>
<td>76.74</td>
</tr>
<tr>
<td>Mal d 1.07</td>
<td>qMd1.07F</td>
<td>CACCTTTGTTACAGTACAGTGC</td>
<td>234</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>77.91</td>
</tr>
<tr>
<td></td>
<td>qMd1.07R</td>
<td>TAGTGGCTGTAGCTCCTTTGATAC</td>
<td>234</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>77.91</td>
</tr>
<tr>
<td>Mal d 1.08</td>
<td>qMd1.08F</td>
<td>TAGTGGCTGTAGCTCCTTTGATAC</td>
<td>234</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>77.91</td>
</tr>
<tr>
<td></td>
<td>qMd1.08R</td>
<td>TAGTGGCTGTAGCTCCTTTGATAC</td>
<td>234</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>77.91</td>
</tr>
<tr>
<td>Mal d 1.09</td>
<td>qMd1.09F</td>
<td>TTTTTTCAATAGGGCCGAGCTCA</td>
<td>8</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>81.72</td>
</tr>
<tr>
<td></td>
<td>qMd1.09R</td>
<td>GTAATCAGAAGCCGCTTCCAACA</td>
<td>8</td>
<td>126</td>
<td>61</td>
<td>100</td>
<td>81.72</td>
</tr>
<tr>
<td>Mal d 1.10</td>
<td>qMd1.10F</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
<tr>
<td></td>
<td>qMd1.10R</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
<tr>
<td>Mal d 1.11</td>
<td>qMd1.11F</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
<tr>
<td></td>
<td>qMd1.11R</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
<tr>
<td>Mal d 1.12</td>
<td>qMd1.12F</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
<tr>
<td></td>
<td>qMd1.12R</td>
<td>CAGGCTTCTGTACAGTACAGTGC</td>
<td>60</td>
<td>157</td>
<td>61</td>
<td>100</td>
<td>79.81</td>
</tr>
</tbody>
</table>

*Refers to the start position of the forward primer in the coding sequence
Chapter 2

Figure 2: qPCR amplification plots of two *Mal d 1* isoallergens on plasmid DNA and fruit cDNA. A) Amplification of *Mal d 1.02* on MC-12 plasmid DNA as positive control and MC-20 plasmid DNA as negative control. B) Amplification of *Mal d 1.03F* on MC-08 plasmid DNA as positive control and MC-12 plasmid DNA as negative control. C) Amplification of *Mal d 1.11* on MC-12 plasmid DNA as positive control and MC-20 plasmid DNA as negative control. For all the three isoallergens, in the lower chart are reported specific amplification and specific conditions.

Melting curve analysis

During this set up of qPCR experiments, also the analysis of the amplicon melting curves obtained with each primer pair provided a further validation regarding the primers. In fact, at the optimized conditions, single peaks in the heat dissociation curves were obtained indicating also the absence of primer dimers. The melting temperatures (Tm) varied between 76.38 and 81.72 °C (Table 1). Examples for *actin*, *Mal d 1.04*, *Mal d 1.07* and *Mal d 1.10* are reported in Figure 3.

Finally, the specificity of all the primers under optimized conditions was further confirmed by melting curve analysis on pooled cDNA of fruits and leaves of the
cultivar Florina. For Mal d 1.03B and Mal d 1.05 it was not possible to perform this validation since no amplification was observed on the pooled cDNA sample.

Amplicons sequencing

One more corroboration for the primers specificity was obtained by direct sequencing of the amplicons derived from Florina DNA at the optimized conditions. In fact, the annotation of amplicon sequences, reported in Table 2, revealed always the expected single target sequence.

Table 2. Annotation of Mal d 1 amplicon sequences by BALSTN and MegAlign softwares. It continues in the next page.

<table>
<thead>
<tr>
<th>Mal d 1 isoallergens</th>
<th>Primer for sequencing</th>
<th>Sequence annotation</th>
<th>ID*</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 1.01</td>
<td>qMd1.01R</td>
<td>Mal d 1.0105/1.0109</td>
<td>AY827645/AY827644</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>qMd1.02R</td>
<td>Mal d 1.0201/1.0209</td>
<td>AY827660/AY827659</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.03A</td>
<td>qMd1.03AF</td>
<td>Mal d 1.03A01</td>
<td>AY822722</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.03B</td>
<td>qMd1.03BF</td>
<td>Mal d 1.03B01/1.03B02</td>
<td>AY822724/AY822723</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.03C</td>
<td>qMd1.03CF</td>
<td>Mal d 1.03C01/1.03C03</td>
<td>AY822726/AY789266</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.03D</td>
<td>qMd1.03DR</td>
<td>Mal d 1.03D01</td>
<td>AY822727</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.03E</td>
<td>qMd1.03ER</td>
<td>Mal d 1.03E02</td>
<td>AY789269</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.03F</td>
<td>qMd1.03FR</td>
<td>Mal d 1.03F01</td>
<td>AY789271</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.03G</td>
<td>qMd1.03GF</td>
<td>Mal d 1.03G01</td>
<td>AY822733</td>
<td>3</td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>qMd1.04F</td>
<td>Mal d 1.0404/0405</td>
<td>AY827665/AY827664</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.05</td>
<td>qMd1.05F</td>
<td>Mal d 1.0501/1.0503</td>
<td>AY827688/AY827682</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.06A</td>
<td>qMd1.06AF</td>
<td>Mald1.06A02</td>
<td>AY827698</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.06B</td>
<td>qMd1.06BF</td>
<td>Mald1.06B02/B03/B04/B05.02</td>
<td>AY827718/AY827716/AY827714/AY827710</td>
<td>2</td>
</tr>
<tr>
<td>Mal d 1.06C</td>
<td>qMd1.06CF</td>
<td>Mal d 1.06C06</td>
<td>AY827728</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.07</td>
<td>qMd1.07F</td>
<td>1.0702</td>
<td>AY789258</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.08</td>
<td>qMd1.08F</td>
<td>1.0801.01</td>
<td>AY822718</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.09</td>
<td>qMd1.09F</td>
<td>Mal d 1.0901/1.0902/1.0903</td>
<td>AY822721/AY822720/AY789262</td>
<td>-</td>
</tr>
<tr>
<td>Mal d 1.10</td>
<td>qMd1.10F</td>
<td>Mal d 1-like clone 1m</td>
<td>AY428588</td>
<td>2</td>
</tr>
<tr>
<td>Mal d 1.11</td>
<td>qMd1.11F</td>
<td>Mal d 1-like clone 1n</td>
<td>AY428589</td>
<td>1</td>
</tr>
<tr>
<td>Mal d 1.12</td>
<td>qMd1.12F</td>
<td>Identical the Mal d 1.12 BAC sequence</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

* AY827645 and AY827664 derived from cv Red Delicious; AY827644 and AY827688 derived from cv Fuji; AY827660, AY827659 AY827665, AY827682 and AY827718 derived from cv Discovery; AY927222, AY922722, AY922723, AY922726, AY922727, AY922733, AY922718, AY922721 and AY922720 derived from the cv Golden Delicious; AY789266, AY789271, AY789258 and AY789262 derived from cv Prima; AY82629 derived from cv Fiesta; AY827698, AY827714 and AY827728 derived from cv Ingrid Marie; AY827716 derived from the cv Cox; AY827710 derived from cv Jonathan; AY428586 and AY428589 derived from cv Royal Gala.

Some SNPs were also detected revealing the presence of new Mal d 1 alleles in the genome of the cultivar Florina. An example of the alignment of new amplicon sequences of subfamily II with databases sequences is reported in Figure 4.
Chapter 2

Figure 4: Alignment among Mal d 1.04 alleles (AY789242, AY827665, AY827666), Mal d 1.05 alleles (AY789245, AY789247) and Mal d 1.04 and Mal d 1.05 sequences (named Flo Md1.04F and Flo Md 1.05F) derived from the amplicons. SNPs between amplicons and isoallergens sequences are indicated with arrows.

Amplification of different genotypes with Mal d 1 specific primer pairs

The ability of Mal d 1 specific primer pairs to give amplification also on other genotypes respect Florina was confirmed performing end-point PCRs at the optimized conditions on gDNA of Gala, Fiesta, Durello di Forlì, Jonathan and Jonagold. For all the primers a clear single band was obtained confirming the locus-specificity of these primers and not an allele-specificity. The only exception have been found for Mal d 1.12 primers because they were not able to amplify gDNA of the cultivar Fiesta. This fits with the fact that Mal d 1.12 was the only gene for which just the sequence of Florina was available (see Chapter 1). Since the lack of sequences information, it is likely that Mal d 1.12 primers are specific for some alleles and not able to amplify others. In Figure 5 are reported examples of end-point PCRs.

Figure 5: End-point amplifications of Mal d 1.02, -1.06B, -1.07, -1.12 and actin genes on the genotypes: Florina (FL), Durello di Forlì (DU), Gala (GL), Fiesta (FS), Jonathan (JO) and Jonagold (JG). The negative controls are indicated with CN.
PCR efficiency and qPCR data analysis

Relative PCR efficiency for each target gene was evaluated and for the reference gene (actin) and a slight variation was found among them. This result was expected because of differences in stringency of PCR conditions and considering that the primer regions were not chosen by an appropriate software but depending on the presence of SNPs. In order to minimize the influence of PCR conditions on gene expression values, in each plate the actin amplification should be performed under the same conditions as the relative Mal d 1 primer pair.

qPCR data can be analysed mainly with two method: the standard curve method allows to obtain an absolute quantitation of unknown samples. First it has to be created a standard curve and then it is possible to compare unknowns to the standard curve and extrapolate a value. On the contrary, the comparative Ct method permit a relative quantitation by the analysis of changes in gene expression in a given sample relative to another reference sample (such as an untreated control sample). It is well known that data processing can seriously affect interpretation of qPCR results (Larionov et al., 2004) indeed the choise of the method is of a crucial importance. In this work the best method for the qPCR data analysis was established considering the differences in the relative PCR efficiency of primers. In detail, the standard curve method was preferred at the comparative Ct method. In fact, most of the slopes of the relative efficiency curves (Table 3) didn’t fit in the range (between –0,1 and 0,1) required for the use of the comparative Ct method for the analysis of qPCR data. Only for Mal d 1.02 and Mal d 1.03F a slope in that range was found. In Figure 6 are reported, as example, the two relative efficiency curves of the isoallergens belonging to subfamily I, Mal d 1.01 and Mal d 1.02.

![Relative Efficiency Curves](image)

**Figure 6:** Plots of relative efficiency (log input amount versus ΔCt) for Mal d 1.01 (left plot) and Mal d 1.02 (right plot). The actin was the reference gene.

The standard curve method requires the addition of the standard samples in all the plates and for all the tested genes, incuding reference gene and so it is considered more expensive and time consuming compared to other methods. Despite this deficiency, when reliability of results prevails over costs and labor load, the standard curve approach may have advantages. In fact, the standard curve method simplifies calculations and avoids practical and theoretical problems currently associated with PCR efficiency assessment. In
this way the comparisons among results obtained with different primer pairs and in different qPCR plates will be more reliable.

Table 3. Slopes of log input vs ΔCt (Ct Mal d 1 gene- Ct actin) for the 18 expressed isoforms qPCR.

<table>
<thead>
<tr>
<th>Mal d 1 isoallergen</th>
<th>Slope</th>
<th>Mal d 1 isoallergen</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 1.01</td>
<td>-0.791</td>
<td>Mal d 1.06A</td>
<td>-0.8705</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>0.0519</td>
<td>Mal d 1.06B</td>
<td>-0.8616</td>
</tr>
<tr>
<td>Mal d 1.03A</td>
<td>-0.1378</td>
<td>Mal d 1.06C</td>
<td>-0.6125</td>
</tr>
<tr>
<td>Mal d 1.03C</td>
<td>-0.6816</td>
<td>Mal d 1.07</td>
<td>-1.0369</td>
</tr>
<tr>
<td>Mal d 1.03D</td>
<td>-1.0113</td>
<td>Mal d 1.08</td>
<td>-0.2212</td>
</tr>
<tr>
<td>Mal d 1.03E</td>
<td>-1.3793</td>
<td>Mal d 1.09</td>
<td>-1.691</td>
</tr>
<tr>
<td>Mal d 1.03F</td>
<td>0.0799</td>
<td>Mal d 1.10</td>
<td>-0.7929</td>
</tr>
<tr>
<td>Mal d 1.03G</td>
<td>-0.4563</td>
<td>Mal d 1.11</td>
<td>0.166</td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>-0.4486</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mal d 1 expression as subtracted from EST-Databases

The Mal d 1 transcript composition in fruits and leaves of the cultivar Florina was investigated by qPCR, the results are reported in Table 4.

Table 4. Results of the screening conducted by qPCR of Mal d 1 isoallergens expression in fruit and leaf of the cultivar Florina (on the left) and results of the EST database analysis (on the right). The gene expression is indicated with an X. In gray are highlighted the genes for which no expression was found neither with the in vitro nor in silico analysis.

<table>
<thead>
<tr>
<th>Mal d 1 gene</th>
<th>Expression in cultivar Florina</th>
<th>EST database analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fruits</td>
<td>Leaves*</td>
</tr>
<tr>
<td>Mal d 1.01</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.02</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.03A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.03B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.03C</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.03D</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.03E</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.03F</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.03G</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.04</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.06A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.06B</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.06C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.07</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.08</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 1.10</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mal d 1.11</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mal d 1.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pool of leaves collected after 0, 24, and 48h after the inoculation with the fungus V. inaequalis.
Half of the Mal d 1 genes (10/20) have been found expressed in fruits while 17/20 were expressed in leaves treated with the fungus V. inaequalis. A possible explanation for this predominance of gene expression in leaves compared with ones expressed in fruits can be found in the biological function of these genes. In fact, since it is known that PR-10 proteins, as all the other PR proteins, are induced under a range of biotic and abiotic stresses (van Loon et al., 2006), the pooling of leaves challenged with the fungus may have caused the induction of some Mal d 1 genes that otherwise wouldn’t have been expressed or expressed at very low levels. Only two isoforms (Mal d 1.03B and Mal d 1.05) resulted not expressed neither in fruit nor in leaves. Since their nucleotidic sequences did not show any stop codon, they should encode for complete proteins. Therefore, these two genes may have a context-specific expression (other tissues or particular conditions). On the contrary, for 8 Mal d 1 genes detectable expression was retrieved in both leaves and fruits. The results of this experimental work were compared to that of an in silico approach adopted to find out the Mal d 1-related ESTs in the databases (Table 5). This approach resulted in 10/20 Mal d 1 genes found in ESTs from fruits; 5 genes have been found in ESTs derived from leaves and 6 genes in ESTs from roots. In Table 5 are indicated also few information on other tissues as flower or bud. Moreover, three isoallergens (Mal d 1.03A, -1.06C and -1.12) did not found any matches in the ESTs databases even if they were classified as expressed by qPCR. The absence of a sequence in the EST database do not necessarily imply that this Mal d 1 isoallergen is not expressed at all because of the limited availability and representatives of apple cDNA libraries, particularly for genes with a low expression or expressed only after the induction with specific stimuli, as for tissues examined with very few cultivars.

Peculiar isoallergens expression patterns were detected in different tissues with both approaches, suggesting a spatial distribution and different regulatory mechanisms for Mal d 1 transcripts. Concerning the Mal d 1 genes in fruit, only ten on twenty isoallergens showed a detectable level of expression (Table 4). In general, these data were also confirmed by the in silico analysis (Table 5). For instance, results regarding the subfamily III are in good agreement since only Mal d 1.06A and Mal d 1.06B resulted expressed in fruit. However, looking at the subfamily IV, the ESTs for the isoallergens Mal d 1.03C, Mal d 1.03E, Mal d 1.03F, and Mal d 1.07 were retrieved from apple fruit cDNA libraries but no expression was revealed by qPCR for Mal d 1.03C and Mal d 1.03E. This discrepancy might be due to the different genotype or other external factors probably involved in the modulation of the Mal d 1 genes expression. Moreover, for Mal d 1.01, Mal d 1.02, Mal d 1.06A and Mal d 1.11, the expression in fruits was already reported in literature (Puehringer et al., 2003; Beuning et al.; 2004, Botton et al, 2008, 2009; Gao et al., 2008; Pagliarani et al., 2009). Regarding Mal d 1 expression in leaves, 17 isoforms out of 20 showed to be expressed in our qPCR experiments (Table 4), where in the GenBan EST of only seven Mal d 1 genes could be traced (Tables 4 and 5). A possible explanation for this difference can
be found in the biological function of these genes. In fact, since it is known that PR-10 proteins, as all the other PR proteins, are induced under a range of biotic and abiotic stresses (van Loon et al., 2006), the pooling of leaves challenged or not with the fungus may have caused the induction of some Mal d 1 genes that otherwise wouldn’t have been expressed or expressed at very low levels. This consideration can be helpful also to explain the predominance of Mal d 1 genes resulted expressed in leaves by qPCR compared with ones expressed in fruits. Moreover, for Mal d 1.04 and Mal d 1.11, in the databases were found ESTs derived from young leaves challenged with fungal pathogen V. inaequalis, confirming their induction under this biotic stress. This suggest also a functional specificity for the corresponding Mal d 1 proteins.

The spatial distribution of Mal d 1 transcripts in other tissues was shown by the *in silico* analysis. For instance, for Mal d 1.03D and Mal d 1.08, only ESTs from roots were retrieved. On the contrary, many isoallergens were contemporary present in different tissues, namely in flowers (Mal d 1.01, Mal d 1.02, Mal d 1.10, Mal d 1.11, Mal d 1.03E, Mal d 1.03F and Mal d 1.03G) or buds (Mal d 1.02, Mal d 1.06A, Mal d 1.11 and Mal d 1.07). Also in cotton, that have a polyploid genome as apple, a variable expression levels and silencing depending homolog gene and organ was found. In this case, it was suggested that some silencing events can be epigenetically induced during the polyploidization (Adams et al., 2003). This might be valid also for Mal d 1 family in apple. Relatively little is known regarding the functional consequences and evolutionary importance of expression modification after the genome duplication, although a subfunctionalization is is conceivable. Despite this, only more detailed analysis may shed light on the question if in the Mal d 1 family there is a functional redundancy or a diversification of biological activities.

An example of accurate gene expression quantification of Mal d 1 gene family will be presented in Chapter 3 and 4 but some preliminary consideration regarding the levels of expression can be made. In fruit, a general low level of transcripts amount was detected, except for Mal d 1.01 and Mal d 1.02, that were already reported as the most expressed Mal d 1 genes (Puehringer et al., 2003; Botton et al., 2008). Among the other isoallergens, Mal d 1.03A, Mal d 1.03F, Mal d 1.06A and Mal d 1.12 resulted with a Ct value close to 35 indicating a very low expression. Considering the comparable Ct obtained for actin in leaves and fruits, the higher levels of expression of Mal d 1 genes in leaves than in fruit appear clear but, also this time, it has to take in consideration that the leaves were challenged with V. inaequalis. Also in leaves Mal d 1.01 and Mal d 102 were the prevailing isoallergens but a considerable expression was detected also for Mal d 1.03C, Mal d 1.03F, Mal d 1.06A and Mal d 1.11. Mal d 1.08 and Mal d 1.09 were the lower expressed gene. Further analysis on specific samples of fruit (skin and flesh) and of leaves (challenged and not challenged) are needed for a deeper comprehension of the Mal d 1 gene family expression profile.
Table 5. Analysis of *Mal d 1* ESTs in the database. Isoforms for which not ESTs were found are highlighted in gray.

<table>
<thead>
<tr>
<th>Isoallergen</th>
<th>LG</th>
<th>Tissue</th>
<th>EST</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subfamily I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mal d 1.01</em> (AY789238)</td>
<td>13</td>
<td>fruit</td>
<td>CN995671, CO067772, AT000383, AT000488, AT000238, CO418705, CO417908</td>
</tr>
<tr>
<td></td>
<td></td>
<td>flower</td>
<td>CN499628, CO067265</td>
</tr>
<tr>
<td><em>Mal d 1.02</em> (AY789240)</td>
<td>16</td>
<td>fruit</td>
<td>AT000305, CO722244, CO417338, CO751748, CO755442</td>
</tr>
<tr>
<td></td>
<td></td>
<td>flower</td>
<td>CN498980, CN498915</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bud</td>
<td>CO885939, CO723162, CO886486</td>
</tr>
<tr>
<td><em>Mal d 1.03</em> (AY789261)</td>
<td>13</td>
<td>leaves challenged with apple scab</td>
<td>DR997552, DR999339</td>
</tr>
<tr>
<td><em>Mal d 1.04</em> (AY789242)</td>
<td>16</td>
<td>fruit</td>
<td>AT000346, CK900679, CV092112, CV128440, DY439734, CV464743</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bud</td>
<td>CO885448</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leaf</td>
<td>CV524623</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stem</td>
<td>CV997946</td>
</tr>
<tr>
<td><em>Mal d 1.05</em> (AY789245)</td>
<td>6</td>
<td>fruit</td>
<td>AT000364, CK900679, CV092112, CV128440, DY439734, CV464743</td>
</tr>
<tr>
<td><em>Subfamily II</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mal d 1.06A</em> (AY789248)</td>
<td>16</td>
<td>fruit</td>
<td>AT000364, CK900679, CV092112, CV128440, DY439734, CV464743</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bud</td>
<td>CO885448</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leaf</td>
<td>CV524623</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stem</td>
<td>CV997946</td>
</tr>
<tr>
<td><em>Mal d 1.06B</em> (AY789252)</td>
<td>16</td>
<td>fruit</td>
<td>AT000364, CK900679, CV092112, CV128440, DY439734, CV464743</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leaf</td>
<td>EB121962, EG974779</td>
</tr>
<tr>
<td></td>
<td></td>
<td>root</td>
<td>EB151385, EB151964, EB152631</td>
</tr>
<tr>
<td><em>Mal d 1.08C</em> (AY789255)</td>
<td>16</td>
<td>leaf</td>
<td>CN873661, EB123144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>flower</td>
<td>CN996594, CN996563</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fruit</td>
<td>CO541245, CO541264, CO755638, EB125297, EB152534, EB125582, DY255601</td>
</tr>
<tr>
<td>No subfamily</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mal d 1.09</em> (AY428598)</td>
<td>16</td>
<td>fruit</td>
<td>CN999026, CN994391, CO576341, CO415469, CO756114, CO752639, CV128071, CV466349</td>
</tr>
<tr>
<td></td>
<td></td>
<td>flower</td>
<td>CO885269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bud</td>
<td>CO900039</td>
</tr>
<tr>
<td><em>Mal d 1.10</em> (BAC seq)</td>
<td>16</td>
<td>cell culture</td>
<td>EB152817</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Isoallergen</th>
<th>LG</th>
<th>Tissue</th>
<th>EST</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mal d 1.03A</em> (AY789263)</td>
<td>13</td>
<td>fruit</td>
<td>CK900634, CO418155</td>
</tr>
<tr>
<td><em>Mal d 1.03B</em> (AY789265)</td>
<td>13</td>
<td>leaf</td>
<td>EB974766, EB121653, CO693908</td>
</tr>
<tr>
<td><em>Mal d 1.03C</em> (AY789266)</td>
<td>13</td>
<td>cell culture</td>
<td>EB153482</td>
</tr>
<tr>
<td><em>Mal d 1.03D</em> (AY789267)</td>
<td>13</td>
<td>root</td>
<td>DT042899</td>
</tr>
<tr>
<td><em>Mal d 1.03E</em> (AY789270)</td>
<td>13</td>
<td>flower</td>
<td>EE663662, EB176958, EB777475</td>
</tr>
<tr>
<td><em>Mal d 1.03F</em> (AY789271)</td>
<td>13</td>
<td>leaf</td>
<td>CO418484, CO899602, CO021280</td>
</tr>
<tr>
<td><em>Mal d 1.03G</em> (AY789274)</td>
<td>13</td>
<td>root</td>
<td>EB121057, EB149863</td>
</tr>
<tr>
<td><em>Mal d 1.07</em> (AY789257)</td>
<td>13</td>
<td>root</td>
<td>DT003343</td>
</tr>
<tr>
<td><em>Mal d 1.08</em> (AY989259)</td>
<td>13</td>
<td>cell culture</td>
<td>EB152576, EB155679, EB156359</td>
</tr>
<tr>
<td><em>Mal d 1.09</em> (AY989262)</td>
<td>16</td>
<td>root</td>
<td>EB151988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leaf</td>
<td>EG974799</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fruit</td>
<td>CN994295, CO755618, CO755631, CO755661, DY255953</td>
</tr>
<tr>
<td><em>Mal d 1.11</em> (AY428589)</td>
<td>16</td>
<td>root</td>
<td>EB152649, EB151974, EB151528</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leaf</td>
<td>EB151950, EB151857, EB152699, EB152168, EB152262, EB152309, EB152699, EB152833</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cell culture</td>
<td>EB155990, EB156054</td>
</tr>
</tbody>
</table>
Conclusions

In this work the availability of a comprehensive inventory of Mal d 1 isoallergens (Gao et al., 2008; Arens & Van de Weg unpublished) and the more recent genomic information of Chapter 1 were used for the development of a highly specific qPCR tool for gene expression analysis: a set of gene specific primers for 20 Mal d 1 genes, standard reference curves, and protocols for qQPC using SYBR Green chemistry and including methods for data analysis. As a first application, this tool was applied on leave and fruit tissues of the cultivar Florina in order to identify the Mal d 1 allergen genes that are expressed in these studies. Specific primers were developed and validated for 20 Mal d 1 genes and applied. Despite their high sequence similarity, a differential expression was showed according to the apple tissue: many isoallergens resulted expressed in both tissues while some revealed a tissue-specificity. Moreover, for some genes the presence of transcripts was reported here for the first time.

The isoallergens expressed in fruits are all candidates for the determination of apple allergenicity, while those only expressed in leaves are probably more interesting for their putative involvement in stress responses. Two isoallergens were expressed nor in fruits nor in leaves and they are probably specific of other plant tissues or activated in response to other types of stimuli. Whether the co-expression of so many highly similar transcripts in a particular tissue implies functional redundancy or a functional distribution among the gene family is still not clear. Thanks to the tool developed in this work in depth studies on expression levels of all the currently known Mal d 1 genes became feasible and, through the analysis of these genes in different external conditions, development stages and tissues, new knowledge on their function will be gained. Finally, research on specific Mal d 1 genes could be relevant to gain insights towards the relative importance of their quality and quantity in the apple allergy.
References


Son DY, Scheurer S, Hoffmann A, Haustein D, Vieths S. 1999. Pollen-related food allergy: Cloning and immunological analysis of isoforms and mutants of Mal d 1, the major apple allergen, and Bet v 1, the major birch pollen allergen. Eur J Nutr. 38:201–215


Chapter 3

Specific *Mal d 1* genes expression analysis on apple fruit tissues
Introduction

Regular fruit intake promotes good health and has been demonstrated to be preventive for development of several chronic diseases (Vainio and Weiderpass, 2006; Hamer and Chida, 2007). However, like other Rosaceae species, apples can cause severe allergic reactions. The prevalent allergy in Northern and Central Europe is named class 2 fruit allergy or pollen-food syndrome and results from cross-reactivity between the major apple allergen, Mal d 1, and Bet v 1, the major birch pollen allergen. This allergy is exclusively linked to mild and local allergic symptoms together referred to as the oral allergy syndrome (van Ree, 1997). Although several Rosaceae fruits can be involved, apple allergy is mostly associated with birch pollen allergy (Anderson et al., 2009). Based on sequence similarity, both Bet v 1 and Mal d 1 belong to the pathogenesis-related proteins of class 10 (PR 10) (van Loon et al., 2006). Allergens of the PR-10 family are intracellular proteins of 158 - 159 amino acids, with a molecular mass around 17 - 18 kD (Breiteneder and Radauer, 2004) with a still not clear biological function. PR-10 proteins are unstable to pepsin digestion and it has been demonstrated that IgE reactivity to PR-10 proteins is absent following fruits heat-treatments of fruits (Bohle et al., 2006). A high degree of structural homology has been demonstrated among Mal d 1 from apple, Bet v 1 from birch (Vanek-Krebitz et al., 1995) and PR-10 proteins from different Rosaceae fruits (Gaier et al., 2008).

In accordance with this, the cross-reactivity between Bet v 1, Mal d 1, Pyr c 1, and Pru av 1 have been demonstrated in inhibition experiments with serum pools from fruit-allergic patients (Ebner et al., 1995; Fernandez-Rivas and Cuevas, 1999).

A number of Mal d 1 isoforms is present in apple fruits (Son et al., 1999; Helsper te al., 2002, Puehringer et al., 2003). The diversity among isoforms is also reflected at the genomic level. In fact, Mal d 1 appears as a complex multigene family containing at least 20 different loci mainly clustered on the homeologous linkage groups (LG) 13 and 16 (Gao et al., 2005; see Chapter 1). To date, all the isoallergens have been classified in four different sub-families (I - IV) in relation to the presence and length of intron, except for three new Mal d 1 genes described in Chapter 1. The comparison of Mal d 1 coding sequences has revealed different levels of identity: 71 - 83% between sub-families; 86 - 98% within a sub-family; and 98 - 100% between alleles of a single gene (Gao et al., 2005). Modelling of the Mal d 1 structure on the basis of the known Bet v 1 structure revealed a very similar folding (Markovic-Housley et al., 2003). On the other hand, the native IgE-binding conformation of Mal d 1 cannot be easily maintained during the protein extraction because of the interference of different oxidation proteins like as polyphenol oxidases and peroxidases (Vieths et al., 1994). Mal d 1 is reported to be a labile protein and it is easily affected by these endogenous enzymes. For these reasons, techniques avoiding the protein extraction might be more reliable for studying the Mal d 1 family. For instance, although hard to apply for large screening studies, prick-to-prick tests for the direct evaluation of fruits allergenicity can
be a good choice among the in vivo tests. As regards the molecular analysis on fruits, the study of the gene expression at the transcriptomic level can be considered as an informative approach.

From patients’ experience it is known that the severity of apple allergic reactions is not only related to the specific sensitivity of the individual, but is also dependant on many fruits external factors, like growing and storage conditions (Botton et al., 2008) and internal factors, like genotype (Bolhaar et al., 2005). Considering the genotype, cultivars Golden Delicious, Fiesta and Gala were classified as highly allergenic cultivars while Santana is low allergenic (Gao et al., 2008; Bolhaar et al., 2005). The different degree of allergenicity among cultivars raised a crucial question on its origin. Allergenicity may depend on the total amount of Mal d 1 proteins as suggested by Son et al. (1999) but little evidence support this hypothesis. To date, the linear response between total Mal d 1 protein content and allergenicity estimates is lacking. On the other hand, also qualitative Mal d 1 proteins characteristics could be involved, as can be argued from the differences in specific-IgE binding capacity of two Mal d 1 protein variants (Ma et al., 2006; Koostra et al., 2007). Gao et al. (2005) reported also the association of the allergenicity degree and some Mal d 1 genes allelic variants. Which is the principal factor determining different levels of allergenicity is still a question under discussion because this is a complex mechanism in which many variable are involved.

Starting by the knowledge at the genetic and transcriptional level regarding Mal d 1 gene family (described in Chapter 1 and 2), the specific gene expression of each isoallergens was investigated in different genotypes and in apple fruit tissues. In particular, a qPCR approach was used here and an association between the gene transcript amounts and cultivars allergenicity was proposed.
Materials and methods

Plant material
Apple fruits from the genotypes: Jonagold, Jonathan, Durello di Forlì, Florina, Fiesta and Gala were collected at the Cadriano Experimental Station, Bologna University (Italy). All samples were collected at commercial harvest time. Apple skin and flesh were immediately separated and frozen in liquid nitrogen and stored at –80°C until RNA extraction. For both skin and flesh the RNA extractions were carried out starting from 6-8 g of frozen tissue deriving from a pool of three different fruits, according to Pagliarani et al., 2009. For Jonathan, Florina, and Fiesta the RNA was extracted from two different pools to obtain a biological replicate.

qPCR analysis
The qPCR analysis were conducted on three technical replicates, as indicated in Materials and Methods of Chapter 2. Only the ten genes resulted expressed in fruits in Chapter 2 were deeply tested here: Mal d 1.01, Mal d 1.02, Mal d 1.03A, Mal d 1.03F, Mal d 1.06A, Mal d 1.06B, Mal d 1.07, Mal d 1.08, Mal d 1.11 and Mal d 1.12. The sequence of Mal d 1 isoallergens specific primer pairs and optimized conditions of amplification are listed in Table 1 of Chapter 2. The ability of these primer pairs to amplify at the optimised condition in the 6 genotypes used in this study was checked in Chapter 2. A preliminary comparison among three different reference genes (actin, GADPH gene and ubiquitine) was performed in order to choose the gene with the most stable expression in fruit skin and flesh. Primers used for these reference genes were: MdActF/MdActR reported in Paris et al. (2009); GAPDH-For (5’-ATTGCGAGTGGACGCGT-3’)/GAPDH-Rev (5’-GGAGGAGTGACTGTCAGCAGA-3’); UBC-For (5’-CGAATTTTTGCTCGAGCGT-3’)/UBC-Rev (5’-AATGATAAGCAGCTACGAGCC-3’). The qPCR raw data were analyzed with the standard curve method, as explained in Chapter 2, and with actin as reference gene. The final results represent the transcript amount levels of Mal d 1 genes normalized with the transcript amount levels of actin, so they can be defined also as relative expression levels. They are expressed as Arbitrary Unit (A.U.).

Sequences analysis
Both the nucleotidic and amino acid sequences were aligned with the softwares ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and MegAlign (Lasergene® v8.0) using standard parameters. The promoter regions were with analyzed using the program PLANTPAN to find transcription factor binding sites.

Statistical analysis

In order to gain information on the variables most effective in discriminating and grouping the 6 genotypes, the Principal Component Analysis (PCA) was performed with the STATISTICA Software using the Mal d 1 genes normalized expression values of the tertiles for all 6 genotypes. Common component coefficients, eigenvalues, relative and cumulative proportions of the total variance explained by single Mal d 1 genes transcript amounts were calculated. The first two components having maximum variance were then selected for the ordination analysis. Eigenvectors from the matrix of correlation among variables were extracted and used to create bidimensional plots.
Results and discussion

Comparison of *Mal d 1* genes expression levels in Florina apple fruits

The qPCR analysis performed in this work started with the test of reference genes: three putative reference genes, *GAPDH*, *ubiquitin* and *actin*, have been amplified on all the samples. Since *actin* resulted the most stable among different tissues (data not shown), as already reported by Paris et al. (2009) and Pagliarani et al. (2009), it was chosen as reference gene for our expression analysis.

The expression levels of the ten *Mal d 1* isoallergens found expressed in apple fruits (see Chapter 2) were evaluated firstly in skin and flesh of the cultivar Florina and a huge variation have been found among them (Figure 1). In particular, respect to *actin*, that showed a cycle treshold (Ct) around 23-24, *Mal d 1* genes revealed lower expression levels (Ct from 26 to 34) except *Mal d 1.01* (Ct of 21) and *Mal d 1.02* (Ct of 15). This explains why in some case the normalization led to very low relative expression values. According to the level of expression, these genes were divided in three main groups: the highly expressed ones (*Mal d 1.01* and *Mal d 1.02*), the isoallergens with an intermediate expression (*Mal d 1.03F, Mal d 1.06B, Mal d 1.07, Mal d 1.08* and *Mal d 1.11*) and the low expressed (*Mal d 1.03A, Mal d 1.06A and Mal d 1.12*). More in detail, as it is indicated by the charts scales (Figure 1A), the relative gene expression of the first group resulted 100-fold and 10,000-fold higher than the relative gene expression of the second and third group, respectively.

Figure 1: Quantification, by qPCR, of the transcript levels of the isoallergens *Mal d 1.01, Mal d 1.02, Mal d 1.03A, Mal d 1.03F, Mal d 1.06A, Mal d 1.06B, Mal d 1.07, Mal d 1.08, Mal d 1.11, Mal d 1.01* and *Mal d 1.12* in apple fruit skin and flesh at harvest of the cv. Florina (Panel A) and Gala (Panel B). Relative levels of allergen expression in cvs Florina (Panel A) and Gala (Panel B). The levels of expression were calculated using the standard curve method and transcript accumulation is reported as relative expression level, normalized in respect to *actin*, and expressed in arbitrary units (A.U.). The mean values reported in the charts resulted from three technical replicates ± SEM. Black bars represented the relative expression in the apple skin and grey bars the relative expression in flesh.
The predominance of the two isoallergens, Mal d 1.01 and Mal d 1.02 was observed already reported in literature (Puehringer et al., 2003; Botton et al., 2008). For the other genes, the expression level was reported here for the first time.

Thanks to the availability of the Mal d 1 gene cluster on LG 16 sequence (Chapter 1) the comparison of the upstream regions (1.300 bp) of three representative genes (Mal d 1.02, Mal d 1.06A and Mal d 1.07) was performed. Unfortunately, among the big amount of motifs recognized by the software (data not shown), the discrimination of the cis-elements really involved in the determination of the gene expression is hard. Despite this, three cis-elements (Table 1) were found only in the highly expressed gene (Mal d 1.02) and not in the low expressed ones (Mal d 1.06A and Mal d 1.07) suggesting for them an hypothetical involvement even if, up to date, these cis-elements are reported to act in other cellular mechanisms. In fact, the first, UP2ATMSD, is a cis-element known as regulator of the gene expression during initiation of axillary bud outgrowth in Arabidopsis (Tatematsu et al., 2005).

The second, P1BS, is a sequence found in the upstream regions of phosphate starvation responsive genes from several plant species (Rubio et al., 2001) Finally, PIATGAPB was found in the Arabidopsis GAPB gene promoter and resulted involved in the reduction of light-activated gene transcription (Chan et al., 2001). Only further functional studies on these regions will clarify their real importance in the differential expression of the homologous members of Mal d 1 gene family.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Site*</th>
<th>Strand</th>
<th>Motif</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>UP2ATMSD</td>
<td>900</td>
<td>+</td>
<td>AAACCCTA</td>
<td>Arabidopsis</td>
</tr>
<tr>
<td>P1BS</td>
<td>1209</td>
<td>+</td>
<td>GTATATAC</td>
<td>Arabidopsis/tomato/Medicago/barley</td>
</tr>
<tr>
<td>P1BS</td>
<td>1209</td>
<td>-</td>
<td>GTATATAC</td>
<td>Arabidopsis/tomato/Medicago/barley</td>
</tr>
<tr>
<td>PIATGAPB</td>
<td>89</td>
<td>+</td>
<td>GTGATCAC</td>
<td>Arabidopsis</td>
</tr>
<tr>
<td>PIATGAPB</td>
<td>89</td>
<td>-</td>
<td>GTGATCAC</td>
<td>Arabidopsis</td>
</tr>
</tbody>
</table>

* Distance from ATG-codon

The comparison of the promoter regions of Mal d 1.02, Mal d 1.06A and Mal d 1.07, revealed also that the TATA-box is the only conserved portion albeit its position slightly differ: -118 for Mal d 1.02, -112 for Mal d 1.07, and -115 for Mal d 1.06A. In Figure 2 the 200 bp upstream regions of the three genes were aligned and the similarity among the sequences was about 35%.
As described by Gao et al. (2005), the genes of Mal d 1 family can be divided into subfamilies depending on the sequence similarity and corresponding to the presence and length of introns. If the promoter regions of other isoallergens are included in the alignment (Mal d 1.08 and Mal d 1.06B; Figure 3), it clearly appeared that the promoter regions similarity is higher between genes belonging the same subfamily. In particular, if only the -150 bp regions are considered, the similarity between Mal d 1.06A and Mal d 1.06B (subfamily III) is 86% and between the Mal d 1.07 and Mal d 1.08 (subfamily IV) is 79%. On the contrary, if different subfamilies are compared, the similarity is around 45-50% (Table 2). These results indicated that also the -150 bp upstream region of a Mal d 1 genes seems a typical feature of genes belonging to a subfamily.

Table 2. Score of similarity among -150 bp upstream regions of Mal d 1.02 (subfamily I), Mal d 1.06A and Mal d 1.06B (subfamily III), Mal d 1.07 and Mal d 1.08 (subfamily IV). The lines regarding the scores within a subfamily are highlighted in grey.

<table>
<thead>
<tr>
<th>SeqA Name</th>
<th>Len(nt)</th>
<th>SeqB Name</th>
<th>Len(nt)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mald102</td>
<td>150</td>
<td>2 Mald106A</td>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>1 Mald102</td>
<td>150</td>
<td>3 Mald106B</td>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>1 Mald102</td>
<td>150</td>
<td>4 Mald107</td>
<td>150</td>
<td>46</td>
</tr>
<tr>
<td>1 Mald102</td>
<td>150</td>
<td>5 Mald108</td>
<td>150</td>
<td>43</td>
</tr>
<tr>
<td>2 Mald106A</td>
<td>150</td>
<td>3 Mald106B</td>
<td>150</td>
<td>86</td>
</tr>
<tr>
<td>2 Mald106A</td>
<td>150</td>
<td>4 Mald107</td>
<td>150</td>
<td>54</td>
</tr>
<tr>
<td>2 Mald106A</td>
<td>150</td>
<td>5 Mald108</td>
<td>150</td>
<td>58</td>
</tr>
<tr>
<td>3 Mald106B</td>
<td>150</td>
<td>4 Mald107</td>
<td>150</td>
<td>54</td>
</tr>
<tr>
<td>3 Mald106B</td>
<td>150</td>
<td>5 Mald108</td>
<td>150</td>
<td>52</td>
</tr>
<tr>
<td>4 Mald107</td>
<td>150</td>
<td>5 Mald108</td>
<td>150</td>
<td>79</td>
</tr>
</tbody>
</table>

Interestingly, the Mal d 1 genes have been found to share also an higher homology in the intron sequences among genes belonging to the same subfamily, as reported in Figure 4. In particular, the 79% similarity was found between intron sequence of Mal d 1.01 and Mal d 1.02.
Figure 3: Alignment of the promoter regions (-250 bp from the start-codon) of Mal d 1.02 (subfamily I), Mal d 1.06A and Mal d 1.06B (subfamily III), Mal d 1.07 and Mal d 1.08 (subfamily IV). Start codons and TATA-boxes are indicated with boxes. The nucleotides in common for the three sequences are indicated with stars; the nucleotides in common between Mal d 1.06A and Mal d 1.06B are highlighted in violet; the nucleotides in common between Mal d 1.07 and Mal d 1.08 are highlighted in grey.
Figure 4: Alignment of nucleotidic sequences of Mal d 1.01 and Mal d 1.02, belonging to subfamily I, and Mal d 1.06A and Mal d 1.06B, belonging to subfamily III. Start codons, stop codons and the extreme nucleotides of exons were indicated with boxes. In the intron sequences, the identical nucleotides were hightlighted with the same colors for the members of the same subfamily.

After these findings it is difficult to conclude if the -150 bp upstream region might act in the transcriptional control of Mal d 1 genes belonging the same subfamily. In fact, Mal d 1.01 and Mal d 1.02 share a comparable level of expression in Florina fruit but, on the contrary, Mal d 1.06A and Mal d 1.06B expression levels differed consistently. Indeed it is reasonable to assume a putative involvement of other elements in the determination of gene expression levels, for example the sequence structure. In fact, the relationship between sequence structure (gene size, especially the size of non-coding regions, intron sequence or gene orientation) and the expression level for plant genes was already reported (Yang, 2009; Bondino and Valle, 2009, Sanzol et al., 2009).

As well as among genes, a clear differential expression has been found between apple skin and flesh. An higher abundance of transcripts has been generally found in apple skin than in flesh for most of Mal d 1 genes (Figure 1A), confirming the data previously published by Pagliarani et al. (2009) for the Mal d 1.02. However, an opposite behaviour was found for the two new isoallergens, Mal d 1.11 and Mal d 1.12, that resulted most
expressed in flesh than in skin. Beuning et al. (2004) have already reported the presence of Mal d 1 in ESTs (corresponding to Mal d 1.11) in fruits. By contrast, for Mal d 1.12 no data regarding its expression were available, probably due to its very low expression level.

Fernandez-Rivas and Cuevas (1999), after both in vivo and in vitro tests, reported that skin of Rosaceae fruits (apple, pear and peach) induce an higher allergic response than flesh. In this case, the higher amount of the majority of Mal d 1 genes in skin fits with the immunological results and might be associated to the higher allergenicity. In the same study it was also reported that 40% of apple allergic patients tolerate the ingestion of the flesh of these fruits. This means that also the low amount of Mal d 1 allergens in apple flesh may be sufficient to provoke allergy in the 60% of sufferers. A significant variability of allergic responses among patients, genotypes and apples were already described by Asero et al. (2006) and also Ricci et al. (2010, submitted), in a study performed with Prick-to-Prick tests, reported for some cultivars (i.e. in Jonathan and Jonagold) even a stronger reaction for flesh than skin. As a consequence, it is conceivable that also small variations in the amount of Mal d 1 isoforms in similar apple samples may be sufficient to increase or reduce the reactivity in allergic patients. Hence, it is not possible to exclude a role in the allergic reaction also for the low expressed Mal d 1 isoallergens. In this case, Mal d 1.11 and Mal d 1.12 are the best candidates to be relevant for the allergenicity to apple flesh.

Expression levels of Mal d 1 genes in the fruits of different genotypes

In order to unravel the importance of internal factors, as the genotype, in the induction of allergy, the Mal d 1 gene expression profiles were also analysed in fruits of cultivar Gala. As shown in Figure 1B, also in this case the qPCR analysis allowed to divide the isoallergens in the three groups. A general correspondance between Florina and Gala for what concern this division of isoallergens and a comparable magnitude of expression between genotypes was found, except the higher expression of Mal d 1.06A in Gala and other slight differences. To investigate further this putative variability in Mal d 1 gene expression among genotypes, the investigation was extended to four more genotypes (Jonagold, Jonathan, Durello di Forlì and Fiesta). Most notably, the ability of Mal d 1 specific primers to give amplification on these genotypes was described in Chapter 2. In particular, Fiesta was excluded from the gene expression analysis for Mal d 1.12 because the primers for this gene are not able to amplify this genotype.

A different degree of allergenicity among cultivars is reported in literature (Bolhaar et al., 2005; Gao et al., 2008) indicating that the genotype can directly influence the allergic reactions in atopic individuals. The data available in literature for these genotypes regarding allergenicity or Mal d 1 protein amounts are summarized in Table 3.
Table 3. Summary of data available in literature for the genotypes allergenicity or protein amount.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Technique</th>
<th>Gala</th>
<th>Fiesta</th>
<th>Florina</th>
<th>Durrello di Forlì</th>
<th>Jonathan</th>
<th>Jonagold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ricci et al., 2010</td>
<td>Prick-to-Prick Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skin: HIGH</td>
<td>Skin: HIGH</td>
<td>Skin: INTERMEDIED</td>
<td>Skin: LOW</td>
<td>Skin: LOW</td>
<td>Skin: LOW</td>
</tr>
<tr>
<td>Son et al., 1999</td>
<td>Total Mal d 1 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>determination</td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>LOW</td>
</tr>
<tr>
<td>Bolhaar et al., 2004</td>
<td>Prick-to-Prick Test</td>
<td>HIGH</td>
<td>HIGH</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Marzban et al., 2005</td>
<td>Total Mal d 1 protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LOW</td>
</tr>
<tr>
<td></td>
<td>determination</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>LOW</td>
</tr>
</tbody>
</table>

The complex expression profile of Mal d 1 genes in different genotypes is reported in Figure 5. These results are not easily explainable but it is clear that, beside the variability regarding the allergenicity degree, also at the transcriptional level there is great genotype-depending variability.

Exploiting the data reported in Table 3, it was possible to speculate on the contribution of each Mal d 1 gene in the allergenicity of the cultivars. More in detail, Mal d 1.06A and Mal d 1.07 were the two isoallergens for which the expression was better positively correlated with the allergenic information. In fact, they resulted mainly expressed in Gala and Fiesta that are well characterized as high allergenic genotypes. Looking to the other genes, Mal d 1.03F and Mal d 1.12 resulted negatively correlated to allergenicity due to their absence in Gala and Fiesta. For the other isoallergens (Mal d 1.01, Mal d 1.02, Mal d 1.03A, Mal d 1.06B, Mal d 1.08 and Mal d 1.11) a correlation to allergenicity is not clear.

By the Principal Component Analysis (PCA) it was possible to quantify this hypothetical association. In was possible to conduce this analysis only with the data regarding fruit skin due to the almost absent expression of many Mal d 1 genes in pulp. The analysis of the three replicates normalized values of Mal d 1 gene expression in skin, the 6 genotypes resulted clustered in three groups (Figure 6A): Gala and Fiesta were closely clustered and discriminated from Jonathan, Jonagold and Durrello di Forlì. Florina was located apart resulting to be differentiated from the first two groups.

The first two factors with eigenvalues >1 were able to explain more than 80% of the total quantitative variation found among cultivars. In particular, the first factor, which explains the 53.7% of the total variation, divides Fiesta, Gala and Florina from the others. This this factor appeared mainly affected by Mal d 1.01 (91.6%) and Mal d 1.02 (92.9%) expression, as it is shown in Figure 6B. Since Gala, Fiesta and Florina are the three cultivars with the higher general transcripts amounts, this factor can be explained looking at the clear predominance of expression of Mal d 1.01 and Mal d 1.02 respect the others. More interestingly, the second factor, which explains the 28.5% of the total variation and which is what clearly distinguishes between the high allergenic cultivars (Gala and Fiesta) and the others, appeared mainly correlated with Mal d 1.06A (85.3%) and Mal d 1.07 (90.4%) expression,
closely followed by *Mal d 1.03A* expression (68.4%), as is reported in Figure 6B. Indeed, these results corroborate the hypothesis of a particular correlations among these isoallergens and allergenicity.

**Figure 5**: Quantification, by qPCR, of the transcript levels of the isoallergens *Mal d 1.01, Mal d 1.02, Mal d 1.03A, Mal d 1.03F, Mal d 1.06A, Mal d 1.06B, Mal d 1.07, Mal d 1.08, Mal d 1.11 and Mal d 1.12* in apple fruit skin and flesh at harvest. The levels of expression were calculated on the cv Jonagold, Jonathan, Durello di Forlì, Florina, Fiesta and Gala using the standard curve method and transcript accumulation is reported as mean normalized expression, in arbitrary units (A.U.). The values in the charts resulted from three technical replicates ± SE. Black bars represented the relative expression in the apple skin and grey bars the relative expression in flesh.
Figure 6: Results of the PCA. A) Grouping of the 6 genotypes, using Euclidean coefficients based on the Mal d 1 genes expression data. In the charts the points: Three groups of genotypes were identified according to the first two factors: group I containing the replicates 13, 14 and 15 of the cultivar Fiesta and the replicates 16, 17 and 18 of the cultivar Gala; group II containing the replicates 4, 5 and 6 of the genotypes Durello di Forlì; 7, 8 and 9 of the cultivar Jonagold and 10, 11 and 12 of the cultivar Jonathan; group III containing the replicates 1, 2, and 3 of the cultivar Florina. The groups are indicated with circles. B) Disposition of the variables (Mal d 1 genes) in the two factors plot. The genes more involved in the determination of the factor 2 are indicated with a box.

The reliability of the expression data for the isoallergens was tested by the study of their transcript levels on a biological replicate of fruit samples of Jonathan, Florina and Fiesta. The pattern of expression was always confirmed and, most notably, the higher Mal d 1.06A gene expression in Fiesta than the others genotypes. In Figure 7 are reported, as example, the results on the biological replicate for Mal d 1.01, -1.02, -1.06A and -1.06B.

Figure 7: Quantification, by qPCR, of the transcript levels of the isoallergens Mal d 1.01, -1.02, -1.06A and -1.06B in a biological replicate of apple fruit skin and flesh at harvest. The levels of expression were calculated on the cv Jonagold, Jonathan, Durello di Forlì, Florina, Fiesta and Gala using the standard curve method and transcript accumulation is reported as mean normalized expression, in arbitrary units (A.U.). The values in the charts resulted from three technical replicates ± SE. Black bars represented the relative expression in the apple skin and grey bars the relative expression in flesh.
By combining the gene expression results obtained here with the allelic-diversity study conducted by Gao et al. (2008) on some intron-containing Mal d 1 genes, some statement can be made. In fact, Gao et al. proposed an association between allergenicity and the allelic composition of two isoallergens located in the LG16, Mal d 1.04 and Mal d 1.06A. In particular, they found association between low allergenicity and the pseudo-alleles ps1 and ps2 of Mal d 1.04 and the allele -02 for Mal d 1.06A. The absence of Mal d 1.04 transcripts in fruits (see Chapter 2) suggested that this allergen can not be relevant for allergenicity.

On the contrary, after the transcriptomic study performed at gene (locus) level in this work, the correlation of the isoallergens Mal d 1.06A with allergenicity became more consistent but the question regarding the different level of allergenicity among cultivars remain unresolved. In fact, it is not yet clear if the degree of allergenicity between genotypes with a different haplotype for the Mal d 1.06A gene, is due to an involvement of the three amino acid substitutions or to a different transcript amounts, as appeared in our study. Mal d 1.06A protein variants differ for three amino acid substitutions in distinct positions (Figure 8).

The changes in the protein sequences might affect the binding affinity among IgE-epitopes. Albeit this, the quantitative hypothesis seems more realistic since in some case the Mal d 1.06A expression level appeared more related to allergenicity than allelic composition. In fact, Gala and Jonathan showed the same Mal d 1.06A alleles: -01 and -02 (Gao et al., 2008) but a difference in gene expression and allergenicity degree have been reported: Gala howed higher allergenicity and higher expression levels. Why different genotypes showed a different levels of Mal d 1 transcripts is not clear. It is hard to assume that the
few SNPs in the coding regions may affect the level of expression indeed, also in this case, a putative role of some gene feature, as the intron sequences, is more likely. Other functional analysis are need to further investigate this aspect.

Finally, also by Gao et al. (2008) no association was observed for the protein variants coded by the *Mal d 1.01* (on linkage group 13), -1.02, -1.06B, -1.06C genes (all on linkage group 16), nor by the *Mal d 1.05* gene (on linkage group 6) and it is in agreement with the results retrieved by the gene expression profiling (Figure 5). All these findings highlight the importance of a multidisciplinary approach including genomic, transcriptomic, proteomic and medical studies, in the study of a complex mechanism such apple allergy.

All these considerations lead to the conclusion that both the qualitative and quantitative aspects may play a role in the determination of allergenicity. In fact, to distinguish relevant members for the allergenicity among the complex gene family, the qualitative characteristics of *Mal d 1* isoallergens seem to prevail over those quantitative, considering the prevalence of *Mal d 1.06A* over *Mal d 1.02* although the great difference in their gene expression (*Mal d 1.02* >> *Mal d 1.06A*). On the contrary, considering the pattern of expression of a single isoallergen putatively involved in allergenicity, small variation in its amount in different samples seem to influence the allergic response indicating the importance of quantitative aspects.

Since it is possible to assume that there is no naturally occurring cultivars without *Mal d 1* allergens and that the genotype is an important factor in the determination of allergenicity, the hypo-allergenicity has to be searched in other genotypes, for instance among wild genotypes or germoplasm. The knowledge acquired in this work can be considered a good base since the gene expression of the two isoallergens hypothetically best related to allergenicity, *Mal d 1.06A* and *Mal d 1.07*, might be used to test new materials as marker for the degree of allergenicity. For this aim the segregating population of Durello di Forlì X Fiesta might represent a good resource due to the opposite degree of allergenicity of the two parents (Table 1) and the clearly distinguishable levels of expression of *Mal d 1.06A* and *Mal d 1.07* between the high allergic cultivar (Fiesta) and the low allergic genotypes (Durello di Forlì). Further proteomic and clinical studies will be helpful to corroborate these hypothesis. Moreover, considering the putative involvement of isoallergens quantitative aspects in the determination of allergenicity, a special attention should be given to the apple growth and storage conditions. In fact, in previous study a significant effects on the *Mal d 1* content of fruits have been reported for enviromental factors as orchard elevations or shading and postharvest storage conditions as modified atmospheric conditions (Sancho et al., 2006; Botton et al., 2009).
Conclusions

In this work the specific expression levels of Mal d 1 isoallergen genes was studied for the first time in apple fruits. The analysis was performed on six different genotypes and was focused only on the 10 Mal d 1 members that were already reported to be expressed in fruits. The results showed a significant variability, mainly depending on Mal d 1 gene but also on tissue and genotype and so a complex gene expression profile was obtained. The results suggested a certain influence of both the qualitative (which Mal d 1 gene) and quantitative aspects (level of important Mal d 1 gene) of Mal d 1 genes transcription profile in the apple allergy mechanism. In fact, by combining these expression data with the knowledge about allergenicity of the different genotypes available in literature, it was possible to speculate on the correlation of each gene with the degree of the allergic reaction in different cultivars. Mal d 1.06A and Mal d 1.07 resulted the two genes with the transcript amount positively correlated to the degree of allergenicity and so a particular importance was proposed for them albeit they are not the most expressed Mal d 1 genes in apple. This hypothesis needs further validation by in vivo and in vitro tests, both at medical and biological level because only a multidisciplinary approach might solve a complex phenomenon such as apple allergy. In particular further proteomic analysis will be helpful to confirm the levels of expression also at the protein level.
References


Chapter 4

Specific *Mal d 1* genes (PR10) expression analysis on apple leaves
Introduction

Pathogenesis-related (PR) proteins are plant-specific proteins currently classified into 17 functional families that are rapidly and strongly induced upon pathogen infection or in related situations (van Loon et al., 2006). The term pathogenesis-related refers to the fact that these proteins are induced in association with resistance responses but does not by itself imply a functional role in defence (van Loon et al., 2006). In the past few years, the changes in transcriptomes during the plant response to biotic stress have been deeply analysed and, among the hundreds of modulated genes, PR proteins inductions have been detected in both compatible and incompatible plant-pathogen interactions (Liu and Ekramoddoullah, 2006).

PR-10 is a large PR protein family that comprises ubiquitous intracellular proteins with still unclear biological function. This family contains more than one hundred members isolated from several plant species and tissues. PR-10 proteins are typically small (17-18 kDa proteins of 151 to 163 residues) cytoplasmic proteins, with an acidic isoelectric point and with a conserved three-dimensional structure, consisting of seven-stranded $\beta$-sheet wrapped around a C-terminal $\alpha$-helix and two additional small $\alpha$-helices (van Loon et al., 2006; Markovic-Housley et al., 2003). The primary amino acid sequence presents a conserved P-loop motif (GxGGxGxxK) localized in the region between $\beta2$ and $\beta3$ that is hypothetically involved in the RNase activity of some PR-10 proteins as a binding site. Moreover, an hydrophobic pocket has been identified in the crystal structure of PR-10 from Betula verrucosa and Lupinus albus (Markivic-Housley et al., 2003).

So far, there are many reports of cloning, expression and characterization of multiple members of the PR-10 family in plant genomes. Most PR-10 sequences have been found clustered in plant chromosomes (Kleine-Tebbe et al., 2002; Hoffmann-Sommergruber and Radauer, 2004; Gao et al., 2005). In apple, the Mal d 1 multigene family is containing at least 20 different genes, mainly clustered on the two homeologous linkage groups (LG 13 and 16). Comparison of Mal d 1 sequences revealed high levels of identity among members, ranking from 73% to 99% of positives amino acid matches (Gao et al., 2005, see Chapter 1) and the homogeneity is suggested to be maintained by concerted evolution. The cluster organization of this gene family may have facilitated this evolution (Gao et al., 2005, see Chapter 1).

The inducible expression of PR-10 genes has been widely investigated in a number of plant species and in response to different biotic stresses, as viruses, bacteria and fungal pathogens (Liu and Ekramoddoullah, 2006), suggesting a key role in selective defence mechanisms (Flores et al., 2002; Chadha and Das, 2006). In apple, a GUS activity regulated by the promoter of a PR-10 gene resulted induced by viral (TMV) and fungal (Botrytis cinerea) attacks (Pühringer et al., 2000). Moreover, Beuning et al. (2004) and Paris et al. (2009) found an induction of Mal d 1 transcripts in leaves of apple challenged with the
fungus *V. inaequalis*. Upon the infection with various types of pathogens, the association between the accumulation of PRs and the production of SAR genes is often taken to represent a causal relationship (Ryals et al., 1996). After the arrival of the mobile signal, the infected tissues start to produce salicylic acid which induce PR proteins locally (Verberne et al., 2003). Several studies demonstrated the induction of PR-10 under SA application, as for example in pear and apple, suggesting for these proteins a putative role both in local and in systemic reactions (Ziadi et al., 2001; Faize et al., 2004). The induction of PR-10 genes was also reported by abscisic acid and methyl jasmonate in lily anthers (Wang et al., 1999). These data suggest that defence-related compounds are involved in the signal transduction pathway leading to PR-10 activation.

Many abiotic stresses have also been demonstrated to induce PR proteins productions. For instance, wounding in birch, potato or white pine (Liu and Ekramoddoullah, 2006), salinity in roots of rice (Moons et al., 1997) and drought stress in maritime pine and hot pepper (Dubos and Plomion, 2001; Park et al., 2004). As an important environmental factor, cold hardiness affects PR-10 expression in white pine, peach and mulberry inducing the accumulation of these proteins during cold acclimation (Ekramoddoullah et al., 1995; Wisniewski et al., 2004 and Ukaji et al., 2004).

Finally, some PR-10 proteins display constitutive expression patterns unrelated to stress responses but developmentally regulated in different plants tissues and organs. For instance, it is well know that PR-10 expression is ripening-related in apple fruits (Atkinson et al., 1996; Goulao and Oliveira, 2006; Pagliarani et al., 2009). Beside the expression in fruits, PR-10 in apple are known to be expressed in young and mature leaves, flowers and roots (Puehringer et al., 2003).

Despite the ubiquitous occurrence in the plant kingdom, the molecular mechanisms through which PR-10 regulates all these important plant processes are not understood yet. Several *in vitro* microbial inhibition experiments led to suppose that PR-10 members could act as RNases (Park et al., 2004; Zubini et al., 2009). This activity can be crucial during plant defence for controlling the burst of transcription that occurs upon stress sensing or for the apoptotic processes activated in pathogen-infected cells to limit the pathogen invasion. This activity could also be directly involved in the degradation of pathogenic viral RNA. Although the *in vitro* tests, results of PR-10 genes over-expression in transgenic plants are not consistent because the resistance was not always enhanced (Wang et al., 1999; Truesdell et al., 1997). Recent data indicate that not all PR-10 proteins possess the RNase catalytic property, as reported in peach by Zubini et al. (2009). Therefore it has been proposed that RNase activity of PR-10 proteins might be incidental and of no crucial biological importance (Biesiadka et al., 2002). Additional roles of PR-10 proteins could be related to their ability to bind hydrophobic ligands, such as fatty acids, flavonoids (Neudecker et al., 2001; Mogensen et al., 2002; Koistinen et al., 2005) or plant hormones such as cytokinins that are recently emerging as important components of the plant defence strategy repertoire (Pasternak et al., 2005; Fernandes et al., 2008; Chung et al., 2008). Also
a role in steroid hormone-mediated disease resistance or developmental regulation is proposed since the ability of PR-10s to bind brassinosteroids (Markovic-Housley et al., 2003). The hydrophobic pocket in the 3D structure of PR-10 proteins is considered responsible for these non-covalent interaction with apolar ligands, and perhaps with other substrates for enzymatic activity. Moreover, Zubini et al. (2009) reported that there are specific binding affinities among different PR-10 isoforms for plant functional ligands or enzymatic substrates indicating that little changes in the structure of highly homologous isoforms can lead to a subfunctionalization within the family. Finally, the apple PR-10 proteins have been reported to bind also another protein, named MdAP (Puehringer et al., 2003), suggesting its potential role in signalling. Mogensen et al. (2002) proposed that the interaction of PR-10s with other ligands may be exploited in signalling pathway or it can be responsible for protein storage. Moreover, an high PR-10 concentration is also reported to induce apoptosis (Carimi et al., 2003) and the expression of other PR proteins, such as PR-1 (Memelink et al., 1987). Therefore, it is not reasonable to assign a unique function of PR-10 proteins throughout the plant kingdom. This is also supported by the existence of a huge number of different isoforms, with high sequence similarity but with different mode of expression, constitutive or induced by many different factors.

Since the presence of many members of this family in a single species, up to now the functional study of PR-10 proteins have been hampered by the difficulty to evaluate the contribution of each specific member of PR-10 family in plant defence mechanisms. This is also the case for the PR-10 family of apple. As they are important pollen- and food-derived allergens (Breiteneder and Ebner, 2000; Hoffmann-Sommergruber, 2000; Breiteneder and Radauer, 2004), many recent studies are focused on their implications in allergenicity. Less is known about the possible involvement of apple PR-10 proteins in plant defence responses to biotic and abiotic stresses (Ziadi et al., 2002; Beuning et al., 2004; Paris et al., 2009). In this study, through the use of the set of specific Mal d 1 primer pairs for gene expression analysis developed in Chapter 2, the modulation of 17 different Mal d 1 genes was investigated for the first time in apple leaves and upon challenge with the fungus Venturia inaequalis that is the causal agent of apple scab, one of the more severe apple disease in climate regions. The results showed different mode of expression for all the expressed Mal d 1 genes both in the intensity and in the timing of the responses among genes, suggesting a diversification of biological role within this protein family.
Materials and methods

Plant material and inoculation with *Venturia inaequalis*

Young expanded leaves from the scab-resistant cv. Florina were inoculated with a suspension of *V. inaequalis* conidia in a dark chamber growth with 100% of humidity, as described in Paris et al. (2009). The plants were pre-acclimated in the chamber for 24 hours. Water was sprayed on other Florina plants (mock inoculation) and young expanded leaves collected and used as control. A second inoculation experiment was performed in order to obtain biological replicates. The expression of 5 selected *Mal d 1* genes (*Mal d 1.02*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.07* and *Mal d 1.10*) was also tested on young expanded leaves from the scab-susceptible cv. Gala and from two independent scab-resistant transgenic Gala lines transformed with the *HcrVI2* gene, called Ga2-2 and Ga2-21 (Belfanti et al., 2004). All the leaves were collected at 0, 24, 48, 72, and 96 h post inoculation, stored separately at −80°C until RNA extraction.

Total RNA extraction

Total RNA was extracted from 1 g of leaves, according to Paris et al. (2009) and quantified using a Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). 40 µg of DNA-free RNA were treated with 10 Units DNaseI (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK) at 37°C for 20 min. First-strand cDNA was synthesized according to Paris et al. (2009), starting from 1 µg DNA-free RNA. The cDNA was diluted 1:9 and its quality was verified by the amplification with *actin* specific primers (Paris et al., 2009).

qPCR analysis

The qPCR analysis were conducted on three technical replicates, as indicated in Materials and Methods of Chapter 2. Two biological replicates were tested for each Florina sample. *Mal d 1* isoallergens specific primer pairs and conditions of amplification are listed in Table 1 of Chapter 2. The raw data were analyzed with the standard curve method and the *actin* reference gene as indicated in Paris et al., 2009. The final results are expressed as Arbitrary Unit (A.U.).

Sequences analysis

Both the nucleotidic and amino acid sequences were aligned with the softwares ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and MegAlign (Lasergene® v8.0) using standard parameters. The promoter regions were with analyzed using the program PLANTPAN to find transcription factor binding sites.
Results and discussion

The preliminary screening regarding the Mal d 1 genes expression on young leaves challenged with the fungus V. inaequalis revealed that all the genes are expressed except the Mal d 1.03B, Mal d 1.05 and Mal d 1.12 (Chapter 2). Here, for the first time, the transcripts levels of the expressed genes were studied in detail at different times after the inoculation and in different genotypes.

Basal expression of Mal d 1 genes in young leaves of Florina

The comparison of Mal d 1 genes in young, not challenged leaves of Florina have revealed a huge variation of the basal expression among different isoforms (Figure 1). In particular Mal d 1.11 resulted the highest expressed gene followed by Mal d 1.02 Mal d 1.01, Mal d 1.03C/D/E/F/G, Mal d 1.07, Mal d 1.09 and Mal d 1.10 showed an intermediate expression, with Mal d 1.01 10 times more expressed then the other genes. Finally, Mal d 1.04, Mal d 1.06A, Mal d 1.06C and Mal d 1.08 resulted the lowest expressed.

Differences in the levels of expression among Mal d 1 genes were also retrieved in apple fruits (Chapter 3) with only slight differences. In fact, in fruits Mal d 1.02 was the most expressed gene, followed by Mal d 1.01. These results suggest that the tissues specificity for Mal d 1 genes, reported in Chapter 2, is not only qualitative (specific genes in specific tissues) but also quantitative (different levels of the genes in different tissues).

The comparison of the 5' regulating region of the Mal d 1 genes expressed in leaves, when available (Chapter 1), revealed that the TATA-box is the only long stretch that shows high similarity among all the sequences, except for Mal d 1.09, Mal d 1.11 and Mal d 1.12 (Figure 2A). The similarity of this region increases significantly if only -150 bp of the promoter regions of the same subfamily are considered. In fact, the homology reaches the 85% within the subfamily III and the 80% within the subfamily IV as it is shown also by the
cladogram in Figure 2B. Since differences have been found in the expression results for genes of the same subfamily, probably the responsive elements have to be searched far from the start codon, in the not conserved regions.

![Figure 2: Alignment (panel A) and cladogram (panel B) of 150 bp upstreaming regions of Mal d 1.02, Mal d 1.03G, Mal d 1.06A, Mal d 1.06B, Mal d 1.06C, Mal d 1.07, Mal d 1.08, Mal d 1.09, Mal d 1.10, Mal d 1.11 and Mal d 1.12. The cardinal numbers indicate the subfamilies.](image)

**Expression of Mal d 1 genes in young leaves of Florina challenged with V. inaequalis**

Although the Malus-V. inaequalis interaction is probably the most well characterized woody tree plant-pathogen interaction from the molecular standpoint, the mechanism of the incompatible interaction is still under study. PR-10 proteins have been shown to be transcriptionally activated upon microbial attack or after treatment with fungal elicitors (Puhringer et al., 2000; Elvira et al., 2008). In order to gain insight into the behaviour of the gene family coding for PR-10 proteins during the incompatible Florina-V. inaequalis interaction, the qPCR technology and the tool developed in Chapter 2 were used to test the specific expression of the Mal d 1 genes in young leaves of Florina at different
times after the in vivo inoculation with the fungus and the mock inoculation. The cv. Florina was chosen because it is a scab-resistant genotype, with a resistance level classified as 2/3a in the Chevalier’s scale for the appearance of chlorotic lesions in young leaves 21 days after the attack of the fungus. The resistance is mainly due to the resistance gene HcrVf, cloned from this cultivar (Vinatzer et al., 1998) and successfully used for the production of transgenic apple plants (Belfanti et al., 2004). The important responses of apple to V. inaequalis have been reported in early stages, within 24-48h when a cascade of reactions set up that end with the block of the growth of the fungus, registered after 72h (Komjanc et al., 1999). Among the genes induced during the Malus-V. inaequalis interaction, genes coding for the PR-10 proteins were already reported (Beuning et al., 2004; Gau et al., 2004; Paris et al., 2009) but here for the first time the analysis was extended to 17 different Mal d 1 genes. Moreover, the differences in direction, timing and magnitude of the modulation among each member of the family were detected and discussed in order to gain information on the putatively different biological functions of these highly similar genes.

In this work, the gene expression profiling was performed on samples collected at 0, 24, 48, 72 and 96h upon challenge. By calculating the ratio of expression at different times after the inoculation and the expression at T0 (Table 1) a significative modulation (log2 ratio ≥1 or ≤ 0.5 as respectively up- and down-regulated) have been reported for all the Mal d 1 homologous genes albeit with clear differences among genes.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Timing</th>
<th>1.01</th>
<th>1.02</th>
<th>1.03A</th>
<th>1.03C</th>
<th>1.03D</th>
<th>1.03E</th>
<th>1.03F</th>
<th>1.04</th>
<th>1.06A</th>
<th>1.06C</th>
<th>1.07</th>
<th>1.08</th>
<th>1.09</th>
<th>1.10</th>
<th>1.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. inaequalis</td>
<td>T24/T0</td>
<td>3.4</td>
<td>2.2</td>
<td>3.9</td>
<td>4.3</td>
<td>3.3</td>
<td>3.7</td>
<td>4.9</td>
<td>4.1</td>
<td>3.1</td>
<td>3.1</td>
<td>3.6</td>
<td>3.1</td>
<td>3.1</td>
<td>4.6</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>T48/T0</td>
<td>3.4</td>
<td>2.2</td>
<td>3.9</td>
<td>4.3</td>
<td>3.3</td>
<td>3.7</td>
<td>4.9</td>
<td>4.1</td>
<td>3.1</td>
<td>3.1</td>
<td>3.6</td>
<td>3.1</td>
<td>3.1</td>
<td>4.6</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>T72/T0</td>
<td>3.6</td>
<td>2.7</td>
<td>3.6</td>
<td>4.0</td>
<td>3.5</td>
<td>3.6</td>
<td>4.5</td>
<td>3.3</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>4.1</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>T96/T0</td>
<td>3.6</td>
<td>2.7</td>
<td>3.6</td>
<td>4.0</td>
<td>3.5</td>
<td>3.6</td>
<td>4.5</td>
<td>3.3</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
<td>4.1</td>
<td>3.9</td>
</tr>
<tr>
<td>water</td>
<td>T24/T0</td>
<td>4.0</td>
<td>2.0</td>
<td>3.6</td>
<td>2.4</td>
<td>3.6</td>
<td>2.2</td>
<td>3.5</td>
<td>3.5</td>
<td>3.1</td>
<td>3.1</td>
<td>2.2</td>
<td>3.1</td>
<td>3.1</td>
<td>2.2</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>T48/T0</td>
<td>3.0</td>
<td>1.6</td>
<td>1.5</td>
<td>1.7</td>
<td>3.7</td>
<td>4.2</td>
<td>2.1</td>
<td>2.0</td>
<td>3.3</td>
<td>0.5</td>
<td>2.7</td>
<td>3.1</td>
<td>2.7</td>
<td>0.7</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>T72/T0</td>
<td>1.1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.3</td>
<td>3.7</td>
<td>4.2</td>
<td>2.1</td>
<td>2.0</td>
<td>3.3</td>
<td>0.5</td>
<td>2.7</td>
<td>3.1</td>
<td>2.7</td>
<td>0.7</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>T96/T0</td>
<td>0.1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.8</td>
<td>1.9</td>
<td>4.2</td>
<td>2.1</td>
<td>2.0</td>
<td>3.3</td>
<td>0.5</td>
<td>2.7</td>
<td>3.1</td>
<td>2.7</td>
<td>0.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>

More in detail, an up-regulation for all the genes was reported (Figure 3 and 5) except for Mal d 1.10 that was strongly down-regulated by the attack of the fungus (Figure 4). Also by the suppression subtractive hybridization performed by Paris et al. (2009) to identify apple genes that are differentially expressed after V. inaequalis inoculation, two different ESTs for Mal d 1 genes, one by ‘forward’ and the other by ‘reverse’ subtraction, were collected. Moreover, elements for both the activation and repression of expression have been found in the PR-10 promoter of Pinus monticola (Liu and Ekramoeddoullah, 2004).
Figure 3: Quantification, by qPCR, of the transcript levels of the genes Mal d 1.01, Mal d 1.02, Mal d 1.03C, Mal d 1.03F, Mal d 1.03G, Mal d 1.04, Mal d 1.06A, Mal d 1.07, Mal d 1.08 and Mal d 1.11 in Florina young leaves at different times (0, 24, 48, 72 and 96 hours) after the inoculation. The black lines represent the expression in leaves treated with V. inaequalis, the dashed gray lines represent the expression in leaves treated with water. The levels of expression were calculated using the standard curve method and transcripts accumulations are reported as normalized expression, in arbitrary units (A.U.). The values in the charts resulted from three technical replicates ± SE.
Interestingly, among the complexity of cis-elements that have been found in the upstream regions of the Mal d 1 genes (Chapter 1), the 5’ regulating region of Mal d 1.10 was the only one without the cis-element CATGTG. This motif was reported as essential for the up-regulation of the gene erd1 in Arabidopsis by dehydration stress and dark-induced senescence (Simpson et al., 2003). This difference in the upstream regions might be responsible for the opposite type of modulation (up- or down-regulation) observed for Mal d 1.10 in respect to all the other tested genes. The biological meaning of this divergent behaviour found in highly similar genes is still unclear. Several evidence regarding the multifunctional feature of PR-10 family have been found in literature (Zubini et al., 2009; Markovic-Housley et al., 2003). Two hypothesis can be made at this regard. The first hypothesis is based on the assumption of a not essential role for Mal d 1.10 during the Malus-V. inaequalis interaction. In this case a strategy of the cell to save energy can be an hypothesis for the decrease of Mal d 1.10 expression after the challenging, since the large energetic cost for the protein synthesis process during the plant response to an external attack. The second hypothesis regards the putative role of PR-10 proteins in the modulation of endogenous hormones, such as cytokinins, and the ligand specificity of the different PR-10 proteins (Koistinen et al., 2005) during the stress defence. Probably a fine regulation and balance of plant hormones is required to guide the signalling pathway cascade of the defence response to external attack and this might be an explanation for the need of the cell to increase the expression of some members of the PR-10 family and to decrease the levels of others. An antagonistic induction mechanism for the PR proteins was proposed also by Colditz et al. (2007). Using the RNAi approach, they carried out the knock down of the Medicago Truncatula PR10-1 gene and an antagonistic induction of other PR proteins, which are normally repressed due to PR-10 expression, was found. In this case, the silencing of a PR-10 gene and the following increasing of the PR-5b resulted in an increased tolerance of Medicago to the fungus A. euteiches after in vitro infection. Although in their study genes belonging two different class of PR proteins, PR-10 and PR-5, were involved, a similar mechanism inside the complex and heterogeneous PR-10 family can not be excluded.

The magnitude of the gene expression modulation varied significantly among the Mal d 1 genes. Considering the maximum induction for each gene, seven resulted had a log2 ratio > 4 (Mal d 1.01, Mal d 1.03C/F/G, Mal d 1.04, Mal d 1.06A, Mal d 1.07); six had a log2 ratio between 2 and 4 (Mal d 1.02, Mal d 1.03A/D/E, Mal d 1.06B, Mal d 1.08 and Mal d 1.09); and two had a log2 ratio between 1 and 2 (Mal d 1.06C and Mal d 1.11). The down-regulation of Mal d 1.10 resulted with a log2 ratio of -4.8. In particular, the highest modulation have been detected for Mal d 1.06A that presented a maximum log2 ratio of 11 and Mal d 1.11 resulted the least induced with a log2 ratio of 2 (Figure 6).
Figure 4: Absolute quantification, and Relative Quantitation by qPCR, of the transcript levels of the Mal d 1.10 gene in Florina young leaves at different times (0, 24, 48, 72 and 96 hours) after the inoculation. In the upper chart is reported the levels of expression calculated using the standard curve method and transcripts accumulations are reported as normalized expression, in arbitrary units (A.U.). The values in this charts resulted from three technical replicates ± SE. The black lines represent the expression in leaves treated with *V. inaequalis*, the dashed gray lines represent the expression in leaves treated with water. In the lower chart is reported the relative quantitation of Mal d 1.10 described as log₂ of the ratio between the expression at different times after the inoculation and the expression at T0. The black bars represent the induction in leaves treated with *V. inaequalis*, the gray bars represent the induction in mock-inoculated leaves.

Figure 5: Quantification, by qPCR, of the transcript levels of the genes Mal d 1.03D, Mal d 1.03E, Mal d 1.06B, Mal d 1.06C, Mal d 1.03A and Mal d 1.09 in Florina young leaves at different times (0, 24, 48, 72 and 96 hours) after the inoculation. The black lines represent the expression in leaves treated with *V. inaequalis*, the dashed gray lines represent the expression in leaves treated with water. The levels of expression were calculated using the standard curve method and transcripts accumulations are reported as normalized expression, in arbitrary units (A.U.). The values in the charts resulted from three technical replicates ± SE.
Mal d 1.11 was reported as the most expressed at T0 indeed, its low up-regulation may be due to a sort of threshold reached by the Mal d 1.11 transcripts or by a constitutive feature of this genes probably not directly involved in the biotic stress response. On the contrary, considering the very low transcripts level in leaves at T0 of Mal d 1.06A, it’s strong up-regulation after the challenge with the fungus let suppose a crucial role for this gene in the defence mechanism.

**Figure 6:** Up-regulation of gene expression of Mal d 1.06A (black bars) and Mal d 1.11 (striped bars) at different times (24, 48, 72 and 96h) after challenge with V. inaequalis described as log2 of the ratio between the expression at different times after the inoculation and the expression at T0. The threshold for a significative induction of expression is indicated on the chart.

For what concern the timing of modulation, almost all the genes started to be significantly modulated respect the T0 from the first time of collection (24h) till the last (96h) but different patterns have been detected for the maximum of expression. An early single peak of modulation (at 24h) was detected for Mal d 1.07, Mal d 1.08, Mal d 1.03D, Mal d 1.03E, Mal d 1.06B and Mal d 1.06C (Figure 3, 4 and 5). For Mal d 1.07 and Mal d 1.08 after the 24h occurred a slow decrease of expression still the 96h (first column of Figure 3); for Mal d 1.03D, Mal d 1.03E, Mal d 1.06B and Mal d 1.06C a strong reduction of expression was detected at 48h but a slow increase was reported at 72 and 96h (first column of Figure 5). For Mal d 1.04 and Mal d 1.06A a huge single peak of expression was detected only at 48h. A slow increase and a single peak at 72h was found for Mal d 1.01, Mal d 1.02, Mal d 1.03F and Mal d 1.11. The same timing but with an opposite results was detected for expression of Mal d 1.10 that reached the peak of down-regulation at 72h after inoculation. Finally, two peaks of expression (24 and 72h) were detected for Mal d 1.03G, Mal d 1.03C, Mal d 1.03A and Mal d 1.09. Since the important plant transcriptional changes that result in the defence response during the incompatible interaction between Florina leaves and V. inaequalis attack are supposed to occure in the early times (24 and 48h), an important role might be argued for Mal d 1 genes mainly modulated in this period (Mal d 1.07, Mal d 1.08, Mal d 1.03D, Mal d 1.03E, Mal d 1.06B, Mal d 1.06C, Mal d 1.04 and Mal d 1.06A).

This complex expression profiles obtained after the attack of the fungus seems to fit well with the putative involvement of PR-10 members in the signalling pathway activated in the plant cell in response to stress. In birch the interaction of PR-10 proteins with several important ligands, such as phenolic compounds, cytokinines or brassinosteroids, have been demonstrated (Koistinen et al., 2005). It is noteworthy that cytokinines and brassinosteroids are involved in the regulation of the processes of growth and development but are also considered important signalling molecules. It is conceivable that the PR-10 steroid carrier
system proposed for PR-10 of birch (Markovic-Housley et al., 2003; Koistinen et al., 2005) can be true also in the apple defence. In fact, this putative role in the transport and regulation of endogenous hormones concentration and the ligand specificity among PR-10 proteins might allow the availability of high levels of signal molecules readily delivered to their receptors for a quick response to external stimuli. Indeed, a complex network modulating the PR-10 gene expression is proposed and it can also act through a feed-back mechanism. This may be a possible reason for the double peak in the patterns of \textit{Mal d 1.03G}, \textit{Mal d 1.03C}, \textit{Mal d 1.03A} and \textit{Mal d 1.09}. Up to now, there are no elements to interpret the different timing of induction of \textit{Mal d 1} genes but a diversification of functions for genes induced in the same time is likely because otherwise this redundancy it would be a loss of energy for the cell. Further functional analysis on the specific genes are needed to solve this question.

Regarding the comparison of \textit{Mal d 1} transcripts accumulation in leaves challenged with the fungus and in mock inoculated leaves, a modulation was reported for almost all the genes also in the control samples but to a lower extent. For the \textit{Mal d 1} genes reported in Figure 3 and 4 a significative difference was revealed between challenged and not leaves, with a stronger up-regulation in treated samples respect to the controls. These results suggest an involvement of these genes in the response to the specific biotic stress applied in this work. On the contrary, in Figure 5 are grouped the \textit{Mal d 1} genes for which a not clear distinction between treated and controls were reported. In particular, for \textit{Mal d 1.03D}, \textit{Mal d 1.03E}, \textit{Mal d 1.06B} and \textit{Mal d 1.06C} an opposite pattern of induction was reported: the fungus stimulated an higher expression of these genes after 24h from the attack but, although a strong decrease of expression was reported in leaves treated with the fungus at 48h, the maximum of expression in leaves challenged with water was detected at 48h. For \textit{Mal d 1.03A} and \textit{Mal d 1.09}, the expression in leaves challenged with \textit{V. inaequalis} was not significantly different from the finding in the leaves treated with water. Indeed, for these last genes other components respect the attack of the fungus have been occurred in the determination of the gene expression induction, as the wounding during the collection of samples or some experimental conditions. These results are in accordance with the data obtained by the work of Liu et al. (2005) in which plants of Arabidopsis were transformed with a PR-10 of \textit{Pinus Monticola} and a strong induction of this protein was reported after the infection with the pathogen \textit{P. Syringae} as well as after the challenge with water. Similar results were reported for other PR proteins in plants of \textit{Capsicum chinense} challenged with pepper mild mottle virus or with water, where beside in the infected plants, PR-1 and PR-3 mRNAs were detected also in control plants although to a lesser extent (Elvira et al., 2008). It is well known that PR-10 proteins are able to respond to a large range of abiotic stresses (Ukaji et al., 2004; Park et al., 2004; Liu and Ekramoddoullah, 2006) and, in fact, Liu and Ekramoddoullah (2004) demonstrated also the induction of the PR-10 protein around the wounding sites in the control leaves. Also the different patterns of expression in control
leaves among the Mal d 1 genes contribute to stress the hypothesis of different biological function within this gene family and the hypothesis of an indirect role for some of these genes in the defence mechanism of the plant to the fungus.

The results regarding the modulation of the Mal d 1 gene expression upon the V. inaequalis biotic stress were confirmed on a biological replicate of Florina albeit with a lower intensity. In Figure 7 are reported as example the comparisons of the modulation of Mal d 1.04 and Mal d 1.06A in the two replicates. Since it is known that PR-10 proteins are sensible to a variety of external stimuli, these results confirm how the experimental conditions (the plant conditions, the inoculum, the darkness, and humidity) can influence their gene expression and the modulation.

**Figure 7:** Comparison of the modulation of gene expression of Mal d 1.04 (left chart) and Mal d 1.06A (right chart) on two biological replicates of young leaves of Florina at different times (24, 48, 72 and 96h) after the challenging with V. inaequalis and with water described as log2 of the ratio between the expression at different times after the inoculation and the expression at T0.

**Mal d 1 gene expression on leaves of other genotypes**

Because of their peculiar mode of expression, five Mal d 1 genes (Mal d 1.02, Mal d 1.04, Mal d 1.06A, Mal d 1.07 and Mal d 1.10) were chosen for further gene expression analyses in other genotypes: the scab-susceptible genotype Gala and two independent scab-resistant transgenic Gala lines transformed with the HcrVf2 gene, called Ga2-2 and Ga2-21 (Belfanti et al., 2004). Among Ga2-2, Ga2-21 and Florina different level of resistance were reported. Ga2-21 is classified a fully resistant genotype with the value of 1 in the Chevalier’s scale. On the contrary, Ga2-2 and Florina presented a 2/3a value in the Chevalier’s scale due to the presence of chlorotic lesions on the leaves after the attack of the fungus.

The basal, not induced expression level of the tested Mal d 1 genes in the cvs. Florina and Gala and in the GM Gala lines Ga2-2 and Ga2-21 is reported in Figure 8. It is noteworthy the higher expression in the transgenic lines, in particular in Ga2-21, respect to both cultivars for all the Mal d 1 genes, except for Mal d 1.10 that showed an opposite trend, with the major expression in Gala. For Mal d 1.10, a sort of inverse relation between the level of the Mal d 1.10 transcripts and the level of resistance of the plant to the fungus have been reported. Most notably, the comparison of the expression among the genetically modified lines and the wild type genotypes let suppose that the insertion of the transgene for
the resistance *HcrVf2* under a strong promoter in the Gala genotype provoked the increase of the basal level of protection of the cell.

The analysis of gene expression at different hours post inoculation revealed a strong modulation of PR-10 genes also in these genotype (Figure 9 and Table 2). Of particular interest was the up- or down-regulation reported in Gala even if with different intensity and timing. This is in agreement with the claim that PRs are induced also in the plant-pathogen compatible interactions (van Loon et al., 2006; Elvira et al., 2008). This observation make more consistent the hypothesis of a role in the signalling pathway for PR-10 proteins despite the first hypothesis of a putative direct activity against pathogens. In particular, Gala showed always a more linear trend of expression (Figure 9) compared to Florina. A possible explanation for this profile of expression might be that in this susceptible genotype the modulation of PR-10 genes is due expecially to the age of the plant or other abiotic stresses (i.e. darkness or temperature) and not to the pathogen. As regard the expression in the transgenic lines after the inoculation with the fungus, the Ga2-21 line, that is characterized by a high phenotypic level of resistance, showed the maximum of up-regulation always after 72h from the inoculation. On the contrary, in Ga2-2 line, that is characterized with a phenotypic level of resistance similar to Florina, showed the maximum of up-regulation always after 48h from the inoculation. Considering also the difference in the phenotypic level of resistance, a slightly different mechanism of response to the fungus between these two lines might be the reason for the shift of the modulation peak.

Table 2: Levels of induction of 5 *Mal d 1* genes after the inoculation with the *V. inaequalis* expressed as log2 of the ratio between the gene expression at different times after the inoculation (24, 48, 72 and 96 hours) and the gene expression at T0. The highest value of induction of each gene is highlighted in gray (up-regulation) or black (down-regulation).

<table>
<thead>
<tr>
<th></th>
<th>Mal d 1.02</th>
<th>Mal d 1.04</th>
<th>Mal d 1.06A</th>
<th>Mal d 1.07</th>
<th>Mal d 1.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>T24/T0</td>
<td>0.3</td>
<td>0.5</td>
<td>1.9</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>T48/T0</td>
<td>0.5</td>
<td>1.3</td>
<td>2.7</td>
<td>1.2</td>
<td>3.5</td>
</tr>
<tr>
<td>T72/T0</td>
<td>1.8</td>
<td>1.3</td>
<td>3.2</td>
<td>2.7</td>
<td>3.7</td>
</tr>
<tr>
<td>T96/T0</td>
<td>0.7</td>
<td>0.3</td>
<td>3.5</td>
<td>1.7</td>
<td>2.7</td>
</tr>
</tbody>
</table>
Beside the timing, also the magnitude of expression varied among genotypes. In fact, Mal d 1.04 and Mal d 1.07 were clearly higher expressed in Ga2-21 and Ga2-2 respect the wild type genotypes. For Mal d 1.02, a quite high expression was reported for all the genotypes but Ga2-21 reached the maximum value. Mal d 1.06A showed a huge peak of expression at T48h in Ga2-2, as it was reported in Florina. Comparable extent of Mal d 1.06A expression have been detected in Ga2-21 and in Gala wild type. Finally, despite the range of basal expression of Mal d 1.10, a great decrease of expression was detected for this gene in all the genotypes and they all reached an expression close to zero after 96h from the inoculation. The characteristic expression patterns of PR-10 genes in each genotype let suppose that a fine balance of PR-10 proteins, and probably other PR proteins, is needed to obtain each resistant phenotype.

Figure 9: Quantification, by qPCR, of the transcript levels of the genes Mal d 1.02, Mal d 1.04, Mal d 1.06a, Mal d 1.07 and Mal d 1.10 in young leaves of Florina (Flo), Gala wild type (Gwt), Gala line 2-21 (G21) and Gala line 2-2 (G2) at different times (0, 24, 48, 72 and 96 hours) after the inoculation with V. inaequalis. The levels of expression were calculated using the standard curve method and transcripts accumulations are reported as normalized expression, in arbitrary units (A.U.). The values in the charts resulted from three technical replicates ± SE.
Conclusions

Although a gene does not have to be up- or down-regulated to play a key role in a biological process, screening for differentially expressed genes is one of the most straightforward approaches to reveal the molecular basis of a biological system. In this work it was possible, for the first time, to monitor separately the expression of all the PR-10 genes in young leaves of Florina after challenging with the fungus V. inaequalis and many indications regarding the multifunctional feature of the PR-10 family were retrieved. In particular, a clear modulation for all the tested genes have been reported but with a peculiar expression profile for each gene for what concern the direction, the timing and the magnitude of modulation. Moreover, throughout the comparison among fungus-treated and controls samples, the PR-10 genes more involved in the specific Malus-V. inaequalis interaction have been identified. Taking in consideration the double nature of the proteins encoded by Mal d 1 genes, as PR proteins and as fruits allergens, the knowledge acquired in this work will be helpful in the future also to minimize the impact of the PR-10 proteins in the apple allergenicity without compromising the mechanism of response of the plants to stress conditions.

In contrast to the earlier findings concerning the putative RNase activity, the modulation of PR-10 genes during the Gala-V. inaequalis compatible interaction as well as in the Florina-V. inaequalis incompatible interaction, contribute to validate the hypothesis of an indirect role for at least some of this proteins in the induced resistance. In particular, after this work, a putative involvement in the fine and complex network of the plant signal transduction have been suggested for PR-10 proteins. Considering the diversity in the pattern of Mal d 1 gene expression among different resistant genotypes, also a crucial importance for the coordinated expression and balancing of the different transcripts amounts inside the family members have been proposed. Interestingly, a different balance seems to create a different resistant phenotype. The simultaneous up- and down-regulation of different Mal d 1 genes after the challenging with the fungus seems to support this hypothesis. The modulation of PR-10 transcripts in leaves treated with water confirm their ability to respond also to abiotic stress.

Further gene expression analysis upon other pathogens, mechanic injury or environmental conditions will be helpful for the comprehension of the specific biological function of each specific Mal d 1 genes. Moreover, studies at the proteomic level will shed light on PR-10 proteins stability or ligand binding specificity. The subfunctionalization proposed in this work inside the PR-10 gene family might be an explanation for the redundance of highly similar genes maintained during the evolution and also for all the efforts spent by the cell for coding many so similar proteins.
References


Zubini P, Zambelli B, Musiani F, Ciurlì S, Bertolini P, Baraldi E. 2009. The RNA hydrolysis and the cytokinin binding activities of PR-10 proteins are differently performed by two isoforms of the Pru p 1 peach major allergen and are possibly functionally related. Plant Physiol. 150:1235-47.
Chapter 5

Identification and mapping of new \textit{Mal d 2} and \textit{Mal d 4} genes
Introduction

In general, consumption of fruit is strongly recommended to improve health but this can cause allergic reactions in certain individuals. Fruits from the Maloideae, like apple and pear, and Prunoideae such as peach, sweet cherry, plum, apricot and almond (Marzban et al., 2005) have been reported to cause allergic reactions in an increasing proportion of European citizens. General fruit avoidance has negative effects on the health of allergic patients and also affects their quality of life. Therefore, the low allergenic property of fruit is worth considering in new fruits breeding programs. As genotypes of the same species differ in allergenicity (Bolhaar et al., 2005), the selection and breeding of new, low-allergenic cultivars is becoming feasible by a multidisciplinary approach, where doctors and plant geneticists collaborate to obtain low allergenic fruits (Hoffmann-Sommergruber, 2005; Gao et al., 2008). To date four apple allergens have been identified: Mal d 1, a pathogenesis-related protein 10 (PR-10 protein, birch allergen Bet v 1 homologues), Mal d 2, a thaumatin-like proteins (TLP, PR-5 proteins), Mal d 3, a non-specific lipid transfer proteins (nsLTPs, PR-14 proteins) and Mal d 4, a profilin (Breiteneder and Ebner, 2000). This work focused on the two allergens Mal d 2 and Mal d 4.

Mal d 2 was the first TLP described as a plant food allergen (Hsieh et al., 1995) and up to now it is known as one of the major protein constituents of mature apple (Oh et al., 2000) with a predominant gene expression in flesh fruit than in skin (Pagliarani et al., 2009). TLPs belong to the PR-5 family (van Loon et al., 2006). Several researchers provide evidence that TLPs play a role in plant defense against pathogens (Krebitz et al., 2003; Venisse et al., 2002). The majority of TLPs feature a molecular mass of about 22 kDa, showing a characteristic pattern of 16 cysteine residues, forming 8 disulfide bonds that are responsible for protein stability and compactness (Breiteneder, 2004). This stabilized structure contributes to the protein’s resistance to low pH conditions, heat-induced denaturation and proteolysis (Breiteneder, 2004). Numerous studies (Krebitz et al., 2003; Fernandez-Rivas et al., 2006) have been performed on the frequency of sensitization to Mal d 2. Albeit recent immunological data revealed that patients from Spain and Italy demonstrated higher IgE reactivity to recombinant Mal d 2 proteins than those from the Netherlands and Austria (Fernandez-Rivas et al., 2006), only limited information on the biochemical and immunological properties of the purified natural protein are available.

Up to now, two copies of the Mal d 2 gene (Mal d 2.01A and Mal d 2.01B) have been identified in the apple genome, which differed in the length of a single intron (378 or 380 nt) and in only one amino acid in the signal peptide. Both Mal d 2.01A and Mal d 2.01B were located in the same position on linkage group 9 (Gao et al, 2005). However, beside Mal d 2.01A and Mal d 2.01B, also two other ESTs were identified (Gao et al., 2005) corresponding to two putative new Mal d 2 genes in the apple genome since their levels of
sequences similarity. The names of *Mal 2.02* and *Mal d 2.03* for the new isoallergens have been proposed by Gao et al. (2005).

*Mal d 4* encode for a profilin. Profilins are protein families classified as minor allergens involved in pollen allergy and they play a role in pollen-associated food allergy (van Ree et al., 1995). They are small (12-15 kDa) cytosolic proteins found in the eukaryotics cell. Profilin cDNAs from numerous plant species have been cloned and their deduced amino acid sequences are typically 70% to 80% similar. Their protein features are strikingly conserved in respect to their length (most are 131 to 134 amino acid), domains and structure and the basic biological functions of profilins have been attributed to cell elongation, cell shape maintenance and flowering (Ramachandran et al. 2000), seedling development (McKinney et al. 2001) and pollen tube growth (McKenna et al. 2004). In pollen, their abundance increases 10-fold to 100-fold during maturation (Radauer and Hoffmann-Sommergruber 2004). Apart from pollen, many fruits contain profilins and their allergenic potency has been frequently reported (Scheurer et al. 2001; Asero et al. 2003; Westphal et al. 2004). As regards to profilin-related allergy, the cross-reaction between *Bet v 2* (birch pollen profilin) and *Mal d 4* have been reported (Vieths et al., 2002). Four *Mal d 4* genes (*Mal d 4.01A* and *B*, *Mal d 4.02A* and *Mal d 4.03A*) have been characterized in the apple genome and their complete gDNA sequences varied among genes in length from 862 to 2017 nt (Gao et al., 2005). *Mal d 4.01* appeared to be duplicated in two copies and located on linkage group 9. *Mal d 4.02A* and *Mal d 4.03A* were single copy genes located on linkage group 2 and 8, respectively (Gao et al., 2005).

After the first deep genomic characterization and linkage mapping study of two classes of apple allergen, *Mal d 2* (TLP) and *Mal d 4* (profilin), performed by Gao et al. (2005), the idea of the existence in the apple genome of other isoallergens belonging to these two gene families has been developed. This study was a step forward in the characterization of the TLP and profilin gene families in apple by the identification and mapping of two *Mal d 2* and one *Mal d 4* further loci.
Materials and methods

Databases research, sequences analysis and design of isoallergens specific primers

The *in silico* analysis of *Mal d 2* and *Mal d 4* sequences was performed using the software BLAST N (Altschul et al., 1997). ESTs for *Mal d 2* and *Mal d 4*-related sequences were searched both by key-words method in GenBank/EMBL database. A cut off of $E^{-15}$ was used. The alignments of the sequences were obtained with the MegAlign program (DNASTAR, Madison, WI, USA) and the specific primer pairs were designed with the PrimerSelect program (DNASTAR, Madison, WI, USA) in the variable regions. The primers are listed in Table 2.

Full-length sequenceing of *Mal d 4.04*

The specific primer pair for *Mal d 4.04* was used to screen a BAC library available at PRI (Wageningen, The Netherlands) with a PCR-based method. The genetic source is 1980-015-025 derived from selection of the breeding program of PRI. The BAC library has a 12 x coverage. The average BAC clone size is 80 kb. One positive clone was used to obtain the full-length sequence of *Mal d 4.04* through a primer walking approach. The primers used are listed in Table 2.

Mapping genes on molecular linkage groups

Two molecular markers linkage maps derived from Prima x Fiesta (PM x FS, $n=141$) and Durello di Forlì x Fiesta (DU X FS $n=174$) were used to map *Mal d 2* and *Mal d 4* genes. In particular, amplicons for *Mal d 2.02*, *Mal d 2.03* and *Mal d 4.04* from the genomic DNA of Prima, Fiesta and Durello di Forlì, were obtained by direct sequencing. All the PCR amplifications were performed in a 17.5 µl volume containing 50 ng of DNA, 100 nM gene-specific primers, 1.5 mM MgCl$_2$, 100 µM dNTPs, 0.5 Unit AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1X reaction buffer. The reaction included an initial 10 min denaturation at 95°C, followed by 33 PCR cycles (45 s at the corresponding annealing temperature, 2 min at 72°C, and 30 s at 95°C), with a final extension of 7 min at 72°C. The amplicons were visualised on an Image Station 440 CF (Kodak, Rochester, N.Y., USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining. Sequencing was performed by Greenomics (Plant Research International, The Netherlands). Sequences and single nucleotide polymorphism (SNPs) were analyzed both with Chromas Lite 2.01 and Seqman program (DNASTAR, Madison, WI, USA). For *Mal d 2.02* and *Mal d 4.04* genes, the amplicons of seedlings were sequenced and the SNPs retrieved were used to map the genes. *Mal d 2.03* gene was mapped using the Temperature Switch PCR (TSP) method, as reported by Hayden et al., 2009. Grouping and mapping of sequence specific data were performed with 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 map of LGs involved was generated with MapChart (Voorrips 2001).
Results and discussion

Mal d 2 - Thaumatin-like gene family

The results of the EST database analysis conducted for Thaumatin-like genes are reported in Table 1 and, interestingly, ESTs for the new genes Mal d 2.02 and Mal d 2.03 were retrieved from fruit tissue indicating a putative involvement in the allergic reaction. The alignment of the nucleotidic sequences of Mal d 2.01A (AY792359), Mal d 2.01B (AY92602) and the ESTs corresponding to Mal d 2.02 and Mal d 2.03 is reported in Figure 1. Considering the longest EST for Mal d 2.02 (CV082311), an homology of 90% was found with Mal d 2.01A, 91% with Mal d 2.01B and 74% with Mal d 2.03. For what concerns the longest EST for Mal d 2.03 (CO901275), a lower similarity was found with both Mal d 2.01A and Mal d 2.01B (76%). The alignment of the partial deduced amino acid sequences and the phylogenetic tree are reported in Figure 2. The partial deduced amino acid sequences of Mal d 2.02 and Mal d 2.03 showed an homology of 88% and 75% with the Mal d 2.01 sequences. The SNPs identified with the nucleotidic alignments were used to design Mal d 2.02 and Mal d 2.03 specific primer pairs as showed in Figure 3A and 3B, respectively. These primers are listed in Table 2.

Table 1. Mal d 2 and Mal d 4 genes sequences in GenBank/EMBL. Genes studied in this work are highlighted in yellow. All the sequence lengths are indicated in bp.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reference genomic seq.</th>
<th>EST</th>
<th>Tissue</th>
<th>Total length</th>
<th>Coding Ex1</th>
<th>Int1</th>
<th>Ex2</th>
<th>Pop¹</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mal d 2.01A</strong></td>
<td>AY792359</td>
<td>CO754260, CN996077, CN4900732, CO952474, AF4949393, AF4949394, CV10401, CN994747</td>
<td>Fruit, bud, leaf</td>
<td>1.119</td>
<td>741</td>
<td>61</td>
<td>378</td>
<td>680</td>
<td>9</td>
</tr>
<tr>
<td><strong>Mal d 2.01B</strong></td>
<td>AY792602</td>
<td>Not yet</td>
<td>Not known</td>
<td>1.121</td>
<td>741</td>
<td>61</td>
<td>380</td>
<td>680</td>
<td>9</td>
</tr>
<tr>
<td><strong>Mal d 2.02</strong></td>
<td>Not yet</td>
<td>CO804477, CV892311, CO723595, CN445021</td>
<td>Fruit, bud</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>PMxF/S</td>
</tr>
<tr>
<td><strong>Mal d 2.03</strong></td>
<td>Not yet</td>
<td>CO866730, CO901275, CN490042, CV984040, CO866711, CN444138, CO866347, CN491711</td>
<td>Fruit, bud, flower</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DUxF/S</td>
</tr>
</tbody>
</table>

¹Populations used for mapping. PM, Prima; FS, Fiesta; DU, Durello di Forlì
Table 2. Sequences and information about primers for Mal d 4.04, Mal d 2.02 and Mal d 2.03.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer name</th>
<th>sequence (5’-3’)</th>
<th>Ta (° C)</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 2.02</td>
<td>Mal d 2.02Fmap</td>
<td>CCAGAAGTAGATGTTCCACAGATT</td>
<td>58</td>
<td>Mapping by direct sequencing</td>
</tr>
<tr>
<td></td>
<td>Mal d 2.02Rmap</td>
<td>TCAAGGGCTTTCCACATGTAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 2.03</td>
<td>Mal d 2.03Fmap</td>
<td>ACGATACCTTCCACACAGGT</td>
<td>60</td>
<td>Mapping by TSP</td>
</tr>
<tr>
<td></td>
<td>Mal d 2.03Rmap</td>
<td>AGCTCAGCTGGGCAAACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp76M2.03F</td>
<td>CCGGTGCGTGGTAGAGCAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mal d 2.04</td>
<td>Md d 4 group 1F1</td>
<td>ATTTGACGCGCCAGGACAG</td>
<td>60</td>
<td>BAC library screening, first step of primer walking, mapping by direct sequencing</td>
</tr>
<tr>
<td></td>
<td>Md d 4 group 1R1</td>
<td>GGGTCTCAAATCTCTTTCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Md d 4 group 1 step 2</td>
<td>CGAGACACTGCCATCAAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Md d 4 group 1 step 2</td>
<td>GAGGAGGATGACTGTACACAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Md d 4 group 1 step 3</td>
<td>GTCGACCGCATGCTGGTGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Md d 4 group 1 step 3</td>
<td>GCAACCTCTAGTATTTGTAATCTT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Portion of the nucleotidic sequences alignment of Mal d 2.01A (AY792599), Mal d 2.01B (AY92602) and the ESTs corresponding to Mal d 2.02 (CN445021, CO723595, CO904477 and CV082311) and Mal d 2.03 (CN495042, CO901275 and CV084040).
Figure 2: A) Alignment of the amino acid predicted sequences of Mal d 2.01A, Mal d 2.01B, Mal d 2.02 and Mal d 2.03 from the amino acid sequences AY92599, AY92602, CO904477 and CO901275, respectively. B) Phylogenetic tree of the TLP sequences with the percentage of homology.

Figure 3: Representation of the sequence variability in the Mald 2.02 (A) and Mal d 2.03 (B) primer regions. The primer regions are indicated with boxes and the identical nucleotides among sequences are highlighted in yellow.
Through the direct sequencing of the amplicons obtained with these specific primer pairs from the parents of the two mapping populations, it was possible to find polymorphisms to map these new *Mal d 2* loci. More in detail, for *Mal d 2.02*, the SNP found in position 112 (G/T) among the sequences of the cultivars Prima and Fiesta, the parents of a European reference mapping population (Maliepard et al., 1998), was used to follow the pattern of segregations in 100 seedlings and this made it possible to map this gene on the bottom of LG 17 of Fiesta (Figure 4). On the contrary, for *Mal d 2.03* no polymorphism was retrieved among the sequences of the cultivars Prima and Fiesta. Therefore, the Durello di Forlì x Fiesta mapping population was investigated and a SNP polymorphism among the two parents in position 76 was found. Through the TSP method, *Mal d 2.03* locus have been mapped on the bottom of LG 4 of Durello di Forlì (Figure 5). Beside the fact that most *Mal d 1* genes were mapped on the two homeologous LG 13 and 16 and two *Mal d 3* genes on the two homeologous LG 4 and 12, in this work also for the *Mal d 2* gene family a similar consideration can be made. In fact, the finding of a *Mal d 2* gene on LG 17 can be considered a corroboration of the duplicated nature of the apple genome (Maliepaard et al., 1998) since *Mal d 2.01A/B* locus was already reported on LG 9 (Gao et al., 2005) that is known to be homeologous to LG 17. With regard to the *Mal d 2.03* gene position on LG 4, no corresponding genes have been reported up to now in the homeologous LG 12.

Therefore, additional *Mal d 2* loci are expected in this position. The hypothesis regarding the complex multigene family nature for *Mal d 2* is supported by several findings. For instance, the presence of similar TLP genes in the apple genome have been already reported by Oh et al. (2000). Moreover, other *Mal d 2* loci might be expected considering the synteny among *Malus* and *Prunus* genomes. In fact, Chen et al. (2008) reported the presence of TLP of

Figure 4: A) Portions of the two sequence chromatograms of Prima and Fiesta obtained with direct sequencing and three sequence chromatograms of seedlings as examples. The T/C polymorphism is indicated within a box. B) LG 17 of Fiesta. The *Mal d 2.02* sequence position is highlighted in yellow.
Prunus in a region of G1 homeologous to apple LG8 and 19; on G3 that is homeologous to apple LG9; on G7 that is homeologous to LG2 and, finally, on G8 that is homeologous to apple LG5. Considering these findings, other Mal d 2 loci might be located on the top of LG8 and on the top of LG5 of the apple genetic map.

Figure 5: A) Portions of the two sequence chromatograms of Durello di Forlì and Fiesta obtained with direct sequencing. The G/T polymorphism is indicated with a box and the allele specific primer used for the TSP method is indicated with a light gray box and a light gray arrow. B) Screening of the two parents, Durello di Forlì (DU) and Fiesta (FS) and 16 seedlings with the TSP method. The presence of the two bands in Durello di Forlì, due to the polymorphism in the sequence, was used to map the gene. C) LG 4 of Durello di Forlì. The Mal d 2.03 sequence position is highlighted in yellow.

Moreover, Mal d 2 genes encode for proteins also classified as PR-5 proteins and their involvement in the defense response of the plant to pathogens have been reported. For instance, different PR-5 ESTs sequences were found out in a study on the Malus-Erwinia amylovora interaction (Vennisse et al., 2002).

The retention by the evolution of many homologues for the genes involved in the plant resistance mechanism is already reported (Liu et al., 2010) and, in this case, it can further contribute to corroborate the hypothesis of the presence of other Mal d 2 genes in the apple genome.
Mal d 4 – profilin gene family

The research for apple profilins in the public databases (Table 3) revealed the presence of 3 ESTs belonging to this family but not matching perfectly with any Mal d 4 genes known up to date. In particular, through the alignment of these ESTs and the coding sequences of Mal d 4.01, -4.02 and -4.03, the 95% of homology have been found among the ESTs, the 77% between the longest EST (CO066117) and Mal d 4.01, the 74% between the ESTs and Mal d 4.02 and the 76% between the ESTs and Mal d 4.03 (Figure 6). Therefore, the presence of these similar ESTs lead to suggest the presence of a new Mal d 4 isoallergen in the apple genome and according to the official allergen nomenclature (King et al., 1995), the name of Mal d 4.04 was proposed. Interestingly, the ESTs corresponding to the new profilin gene derived from flower tissue. This finding could indicate a minor involvement of this gene in the apple allergen but further specific gene expression analysis are needed to investigate the presence of Mal d 4.04 in apple tissues.

Table 3. Mal d 4 genes sequences in GenBank/EMBL. Genes studied in this work are highlighted in yellow. All the sequence lengths are indicated in bp.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reference genomic seq</th>
<th>EST</th>
<th>Tissue</th>
<th>Total length</th>
<th>Coding Ex1</th>
<th>Int1</th>
<th>Ex2</th>
<th>Int2</th>
<th>Ex3</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal d 4.01</td>
<td>AY792606</td>
<td>CN488488, Co068373, CN992251, CO755896</td>
<td>Fruit, bud, flower</td>
<td>2.017</td>
<td>396</td>
<td>123</td>
<td>343</td>
<td>138</td>
<td>1.278</td>
<td>135</td>
</tr>
<tr>
<td>Mal d 4.02</td>
<td>AY792610</td>
<td>CV084991, CN492781, Co066892, CO756641, CO756641, CN491268, Co0899800</td>
<td>Fruit, bud, flower</td>
<td>862/863</td>
<td>396</td>
<td>123</td>
<td>248</td>
<td>138</td>
<td>218/219</td>
<td>135</td>
</tr>
<tr>
<td>Mal d 4.03</td>
<td>AY792614</td>
<td>CO755279, Co067336, CO067114, Co051739, CV086316, Co415968, CN494198, Co067937</td>
<td>Fruit, bud, flower</td>
<td>1.030</td>
<td>396</td>
<td>123</td>
<td>386/388</td>
<td>138</td>
<td>248</td>
<td>135</td>
</tr>
<tr>
<td>Mal d 4.04 from the BAC clone</td>
<td>CO066117, CN492105, CN996826</td>
<td>Flower</td>
<td>1.120</td>
<td>402</td>
<td>129</td>
<td>531</td>
<td>138</td>
<td>187</td>
<td>135</td>
<td>2</td>
</tr>
</tbody>
</table>
The SNPs found by *Mal d* 4 nucleotide sequences alignments were used to design a *Mal d* 4.04 specific primer pair, as it is reported in Figure 7 and Table 1.

![Figure 6: Alignment of apple profilin nucleotidic sequences: AF129426/AY792605 for *Mal d* 4.01; AF129427/AY792610 for *Mal d* 4.02; AF129428/AY792615 for *Mal d* 4.03 and the ESTs CO066117, CN492105 and CN996826.](image)

The SNPs found by *Mal d* 4 nucleotide sequences alignments were used to design a *Mal d* 4.04 specific primer pair, as it is reported in Figure 7 and Table 1.

![Figure 7: Representation of the sequence variability in the *Mal d* 4.04 primer regions. The primer regions are indicated with boxes and the identical nucleotides among sequences are highlighted in yellow.](image)

This specific primer pair was used for the PCR-based screening of an apple BAC library. From a positive BAC clone obtained from the screening, the full-length DNA sequence of *Mal d* 4.04 was gained through 4 steps of primer walking. The primers used are listed in
Table 1. The 2.111 bp sequence that was obtained (Figure 8) was analysed with a gene predictor software and the results are reported in Table 2 and Figure 8. In particular, a coding sequence of 402 nt was predicted, interrupted by two introns of 531 and 187 bp, respectively.

TGTCCCCTCCCCCATTTTCTCTTCTCTCCCTGTGTTATATTTGGTCTTTTCTCAAACCACCCCCAAATGAGGCACAAACACAA
TATATATAATCCACCAACACAAACACAAATCCTACCTTCTCTCCACAGAAAAAATTTCTCAGTATCTTGATACAAAACACAAAGAAC
AGCATAACACAGGGAGACGCTTGCACACTACACATCAGAGTATCGAAAG

← Md4newREV-STEP3
GATGTCGCAATTGACGCCAGGGACACGACACTGAGGCTGTCGGAAGGGACAGCTCTTCCTTCAGAAAAATTCACAGCTATCTGTACACCAACAAACAAGAACACA

← Md4newREV-STEP2
AGCATAACACAGGGAGACGCTTGCACACTACACATCAGAGTATCGAAAG

← Md4newREV-STEP1
CTGACTTTTGACACTTGCGATTGAGGCTCAATCTCTTTTTTCTGGACTTTGCTTTG

Figure 8: Full-length DNA sequence of Mal d 4.04 from a BAC clone plus 222 bp of the upstream region and 769 bp of the 3' - UTR region. The coding sequence is highlighted in yellow; the start- and stop-codon are in bold and indicated with boxes. Primers used for the screening of the BAC library and for the first step of sequencing are highlighted in black; primers used for the further primer walking steps are indicated with boxes.

The predicted Mal d 4.04 amino acid sequence showed, among the 4 Malus profilin sequences known to date, the highest similarity to other profilin sequences. In particular, the highest homology was found with a profilin of Ricinus communis (XP_002514198). Also...
profilins from *Betula pendula* (ABG48509) and *Corylus avellana* (ABG81302) reported an high similarity, as it is reported in Figure 9. Among the Malus sequences retrieved from the BLAST analysis, the *Mal d 4.01* (AF129426) revealed an homology of 80%, followed by *Mal d 4.03* (AF129428) with 76% of homology and, finally, *Mal d 4.02* (AF129427) with 72% of homology. Most notably, in the *Mal d 4.04* sequence the terminal portion of the protein is the most variable region, compared to the other profilins showed in Fig. 9A. This portion corresponds to the conserved G-actin-binding region (Radauer and Hoffmann-Sommergruber, 2004). This characteristic feature of the protein might affect its biological function and introduce the hypothesis of a function variability inside the gene family, as it is reported for the *Mal d 1* gene family (see Chapter 4).

The new *Mal d 4* isoallergen was mapped in the Prima x Fiesta segregating population. By direct sequencing with *Mal d 4.04* specific primers, a G/T polymorphism between the two parents in position 768 from the start codon (Figure 10A) was found. The segregation of this polymorphism in 100 seedlings allowed the mapping of the *Mal d 4.04* on LG 2 of Fiesta (Figure 10B), very close to *Mal d 4.02* position. Interestingly, the cluster organization for *Mal d 4* genes in apple is reported here for the first time.
Figure 10: A) Portions of the two sequence chromatograms of Prima and Fiesta obtained with direct sequencing. The G/T polymorphism is indicated with a box. B) LG 2 of Fiesta. The Mal d 4.04 sequence position is highlighted in yellow.
Conclusions

The elucidation of the genetic basis of the apple allergy is hampered by the several biological and medical factors involved in this complex reactions. First of all, the presence of four classes of apple allergens and the fact that patients differ in their reactions to these classes of allergens have to be considered. Secondly, the existence of different members (isoallergens) for all the allergen classes may complicate the clarification of the role of individual genes in the allergic reactions. In this work the mapping of two new loci for the TLP family (*Mal d 2.02* and *Mal d 2.03*) and one new locus for the profilin family (*Mal d 4.04*) was performed. In particular, the existence of other loci (LG4 and LG17) for the TLP in apple respect the one reported up to now (LG9) was demonstrated here for the first time. Moreover, despite an overall conservation also among profilins of *Malus* and other species, the predicted amino acid sequence of the new member of the profilin family revealed an high variability in the G-actin-binding region suggesting the hypothesis of a different biological function for this protein. Although further functional analysis, such as proteomic and transcriptomic analysis are needed to well understand their biological characteristic and their involvement in the apple allergy, the identification and location of the all the member of the *Mal d 2* and *Mal d 4* gene families can be considered a fundamental step. In fact, considering that for the conventional breeding it is essential to investigate the allelic diversity of the apple allergen genes, this task might be accomplished only once a genomic basal knowledge of these families will be acquired.
References


General conclusions

Apple (*Malus x domestica*) is a model for *Rosaceae*, because it offers many genetic and genomic resources, like collections of expressed sequence tags (ESTs), bacterial artificial chromosome (BAC) libraries, physical and genetic maps, molecular markers and efficient genetic transformation protocols. Apple can be considered a model species also to study fruit allergy because of the amount of information already available. Then, even if a regular apple consumption is encouraged to enhance human health and prevent various diseases, apple is one of the fruits more frequently reported to provoke allergic reactions, although it does not rank among the most dangerous allergenic foods. Third, apple can be also considered a model system for studying plant-pathogen interaction because, the first resistance gene in a fruit tree species was found in apple and the *Malus-V. inaequalis* system that is nowadays the most studied plant/pathogen interaction in fruit trees.

The main goal of this thesis was to go deep in the knowledge of apple allergen gene families focusing mainly on the larger apple allergen gene family *Mal d 1* and, at a lesser extent, on *Mal d 2* and *Mal d 4* gene families. Starting from previous knowledge, new advances in the characterization of allergen gene families have been reached using different approaches. First of all, a genomic approach was used for the genetic characterization of the allergen gene families of *Mal d 1* (Chapter 1), *Mal d 2* and *Mal d 4* (Chapter 5). A methodological approach was then used in Chapter 2 to set up a tool to discriminate and quantify the gene expression of each allergen gene with the use of a set of highly specific primer pairs and the quantitative Real-time PCR technique. Finally, this transcriptional approach was used to unravel the gene expression profile of each member of *Mal d 1* gene family in apple fruits and leaves (Chapter 3) and in response to one of the most severe apple pathogens, *Venturia inaequalis* (Chapter 4). General conclusions are reported below, divided for each chapter.

- **Chapter 1**
  The study of large contiguous blocks of DNA sequences was a suitable approach to characterize the genomic organization of *Mal d 1* genes on apple. Here, for the first time, the complete nucleotidic sequence of the whole *Mal d 1* cluster on LG16 of the cultivar Florina was reported. This knowledge allowed to acquire many new findings on the number and orientation of allergen genes, their physical distances, their regulatory sequences and the presence of other genes or pseudogenes in that genomic region. By this approach, three new members were discovered within the *Mal d 1* gene family, co-localizing with the Mal d 1 genes known to be in the cluster on LG16 and this result clearly point out that other isoforms may be present in the apple genome. These findings is also suggesting that the complexity of the genetic base of resistance and allergenicity will increase in the next future with the increasing of current knowledge. Also the anchoring of the physical and genetic map of this region has been successfully
achieved suggesting a relation among physical (nt) and genetic distance (cM). New lights has been thrown also on the \textit{Mal d 1} gene cluster evolution. In fact, in this work the hypothesis of the duplication of \textit{Mal d 1} members could have occurred during the polyploidation of the apple genome is further confirmed, as emerge from the presence of many retrotransposons elements within the cluster. The availability of a comprehensive inventory of all the genes belonging to a allergen family, especially in the case of extended gene families like \textit{Mal d 1}, is a basal prerequisite for further studies. In addition, genomics data are useful to predict biochemical and physiochemical characteristics of the protein regarding its molecular weight, secondary and tertiary structure, thermal stability and resistances to proteolysis.

- **Chapter 2**
  The accurate quantification of the expressions of different isoallergens is a difficult task. To solve this problem, an highly specific tool for gene expression analysis was developed and preliminary validated. A set of gene specific primers for the 20 \textit{Mal d 1} genes was developed. As a first application, this tool was applied on leaves and fruit tissues of the cultivar Florina in order to identify the \textit{Mal d 1} allergen genes that are expressed in different tissues. This specific approach regarding the expression of each member of the gene family was performed here for the first time and also the presence of transcripts for some genes have not been demonstrated before. The differential expression retrieved in this study revealed a tissue-specificity for some \textit{Mal d 1} genes. The possibility to distinguish the level of expression of each \textit{Mal d 1} gene expressed in fruits (10/20) may be an useful tool to find out \textit{Mal d 1} genes candidates for the determination of apple allergy. On the contrary, the genes expressed in leaves (17/20) are probably more interesting for their putative involvement in stress responses. The availability of this tool make it possible to evaluate the \textit{Mal d 1} genes behavior on different samples and conditions creating the base to unravel their biological functions.

- **Chapter 3**
  In this work the specific expression levels of 10 \textit{Mal d 1} isoallergen genes, found to be expressed in fruits, was studied for the first time in skin and flesh of apples of different genotypes. A complex \textit{Mal d 1} gene expression profile was obtained, with high variability according to the tissue and genotype. Although we don’t know the exact mechanism that undergo allergencity, an higher or lower level of gene expression in fruits led to identify some \textit{Mal d 1} genes putatively involved in the determination of allergenicity among cultivars. At this regard, the gene expression of \textit{Mal d 1.06A} and \textit{Mal d 1.07} resulted positively associated with the degree of allergenicity and, indeed, considered putatively relevant for allergenicity albeit they were not the most expressed \textit{Mal d 1} genes in apple. These results suggested a certain influence of both the qualitative (which gene) and quantitative aspects (how much a gene is expressed) of \textit{Mal d 1} genes transcription profiles in the determination of apple allergy. Which factors are the main responsible for the differential expression of a particular \textit{Mal d 1} gene
among cultivars is still under discussion. Since the high coding sequence homology among Mal d 1 alleles, the non-coding regions, as upstream or intron regions, might be the most important in determining the differences in their expression.

- **Chapter 4**

Screening for differentially expressed genes is one of the most straightforward approaches to reveal the molecular basis of a biological system. Here for the first, it was possible to monitor separately the expression of all the PR-10 genes in young leaves of Florina after challenging with the fungus V. inaequalis. A clear modulation for all the tested genes have been reported but with a peculiar expression profile for each gene for what concern the direction, the timing and the magnitude of modulation. These differences seem to confirm the hypothesis of a subfunctionalization inside the family as proposed for birch PR-10 proteins, despite their high sequence and structural similarity. Moreover, the modulation of PR-10 genes both in compatible (Gala-V. inaequalis) and incompatible (Florina-V. inaequalis) interactions contributed to validate the hypothesis of an indirect role for at least some of these proteins in the induced defense responses. After this work, a putative involvement in the fine and complex network of the plant signal transduction may also be suggested for PR-10 proteins. Considering the diversity in the pattern of Mal d 1 gene expression among different resistant Vf-genotypes (transgenic and wild type), the balancing of the different transcripts levels inside the gene family seems to contribute to the different resistance phenotypes. Finally, a certain modulation of PR-10 transcripts retrieved also in leaves treated with water confirm their ability to respond also to abiotic stresses. Taking into consideration the double nature of the proteins encoded by Mal d 1 genes, as PR proteins and as fruits allergens, the knowledge acquired in this work about the PR-10 genes putatively more involved in the specific Malus-V. inaequalis interaction will be helpful, in the future, to drive the apple breeding for hypo-allergenic genotypes without compromising the mechanism of stress response of the plants.

- **Chapter 5**

The presence of four classes of apple allergens and the fact that patients differ in their reactions have to be considered in the study of the complex mechanism of apple allergy. In this work two other classes of apple allergens were taken into account: Mal d 2 (TLP) and Mal d 4 (profilin). In particular, the existence of other loci (Mal d 2.03 on LG4 and Mal d 2.02 on LG17) for the TLP in apple respect the one reported up to now (Mal d 2.01A/Mal d 2.01A on LG9) was demonstrated for the first time. Moreover, one new locus for profilins (Mal d 4.04) was mapped on LG2, close to the Mal d 4.02 locus, suggesting a cluster organization also for this gene family. The predicted amino acid sequence of the new member of the profilin family revealed an high variability in the G-actin-binding region suggesting a different biological function of this protein in respect to the other profilins. Despite this, no information is available up to now regarding the involvement of these genes in the allergic reaction but the identification and location of
all the members of the *Mal d 2* and *Mal d 4* gene families can be considered a crucial step for many further studies, as for instance the QTL association study.

**Future directions**

Apple allergy is a multifactorial-determined disease. Indeed, should be considered globally from molecular biology, cell biology, agronomical and medical sciences, as well as social and environmental sciences.

Considering the importance to unravel the whole composition of allergen gene families, the genomic research of new members of these families becomes a huge task and have to be continued. As regards *Mal d 1* family, an approach suitable for this aim is the fine investigation of the cluster on LG13 as performed for LG16. Moreover, thanks to the synteny between apple and pear and also between apple and *Prunus*, other *Mal d* genes could be found by using the information of allergen positions in one species as references for the others. Finally, when the whole apple genome sequence will be available, the search of *Mal d* homologues will be faster and most effective.

An important step forward to study apple allergens will be the development of proteomic approaches. Unfortunately, to obtain recombinants proteins for all the *Mal d 1* genes for their specific quantification will be hard due to the high sequence similarity. A possible approach to solve this problem may be the development of monoclonal recombinant variants for all known *Mal d 1* allergens. Also the modelling of deduced variants and the 3-D structures will be helpful to elucidate the IgE- *Mal d 1* interactions.

Since the importance attributed to the amount of particular *Mal d 1* transcripts on apple allergenicity, further specific studies on the effects of cultural and post-harvest conditions on *Mal d 1* gene expression will be helpful.

The identification of high and low allergenic cultivars represents an important step for understanding the apple allergy mechanism. Indeed, tests for allergenicity of cultivars have to be improved and extended to a broad range of genotypes and patients. The allelic diversity analysis on the high and low allergenic cultivars for all the allergens will provide the genetic background to understand the basis of the low allergenicity. This step from genes to alleles will allow the develop of molecular markers that might be used to effectively address apple breeding for hypo-allergenicity but *in vivo* and *in vitro* tests will be still essential when selected plants will be at the fruiting stage.

Still unresolved are the questions regarding functions and evolution of allergen gene families. For instance, why the retention of many homeologous copies during evolution occurred is not clear yet. The evolution of new functions or the distribution of existing ones among isollergens may be possible explanations, together with a selection-driven preservation of all the copies. For *Mal d 1* genes, in particular, a pathogen-mediated selection can be involved in the evolution of this gene family. Further gene expression analysis upon other pathogens, mechanic injury or enviromental conditions will be helpful for the comprehension of the specific biological function of each specific *Mal d 1* genes.
Understanding the functions of each isoallergen and their peculiar mode of action (expression profile, activity, regulation) in the plant and in response to different biotic and abiotic stimuli will be the challenge of further studies.

The knockout strategy for the introduction of hypo-allergenicity has been applied in rice and soybean. The production of an apple plant with a significant reduction of the overall expression of Mal d genes from an existing economically successful cultivar using the silencing approach seems an attractive time-saving alternative than crossing strategies but, in this case, the biological function of Mal d proteins, as allergens and as PR proteins or profilin, have to be carefully considered in order to not compromise the life of the cell.
General conclusions
Summary

Apple consumption is highly recommended for a healthy diet and is the most important fruit produced in temperate climate regions. Unfortunately, it is also one of the fruits that most often provokes allergy in atopic patients and the only treatment available up to date for these apple allergic patients is the avoidance. Apple allergy is due to the presence of four major classes of allergens: Mal d 1 (PR-10/Bet v 1-like proteins), Mal d 2 (Thaumatine-like proteins), Mal d 3 (Lipid transfer protein) and Mal d 4 (profilin). In this work new advances in the characterization of apple allergen gene families have been reached using a multidisciplinary approach.

First of all, a genomic approach was used for the characterization of the allergen gene families of Mal d 1 (task of Chapter 1), Mal d 2 and Mal d 4 (task of Chapter 5). In particular, in Chapter 1 the study of two large contiguous blocks of DNA sequences containing the Mal d 1 gene cluster on LG16 allowed to acquire many new findings on number and orientation of genes in the cluster, their physical distances, their regulatory sequences and the presence of other genes or pseudogenes in this genomic region. Three new members were discovered co-localizing with the other Mal d 1 genes of LG16 suggesting that the complexity of the genetic base of allergenicity will increase with new advances. Many retrotransposon elements were also retrieved in this cluster. Due to the development of molecular markers on the two sequences, the anchoring of the physical and the genetic map of the region has been successfully achieved. Moreover, in Chapter 5 the existence of other loci for the Thaumatine-like protein family in apple (Mal d 2.03 on LG4 and Mal d 2.02 on LG17) respect the one reported up to now was demonstrated for the first time. Also one new locus for profilins (Mal d 4.04) was mapped on LG2, close to the Mal d 4.02 locus, suggesting a cluster organization for this gene family, as is well reported for Mal d 1 family.

Secondly, a methodological approach was used to set up an highly specific tool to discriminate and quantify the expression of each Mal d 1 allergen gene (task of Chapter 2). In particular, a set of 20 Mal d 1 gene specific primer pairs for the quantitative Real time PCR technique was validated and optimized. As a first application, this tool was used on leaves and fruit tissues of the cultivar Florina in order to identify the Mal d 1 allergen genes that are expressed in different tissues. The differential expression retrieved in this study revealed a tissue-specificity for some Mal d 1 genes: 10/20 Mal d 1 genes were expressed in fruits and, indeed, probably more involved in the allergic reactions; while 17/20 Mal d 1 genes were expressed in leaves challenged with the fungus Venturia inaequalis and therefore probably interesting in the study of the plant defense mechanism. In Chapter 3 the specific expression levels of the 10 Mal d 1 isoallergen genes, found to be expressed in fruits, were studied for the first time in skin and flesh of apples of different genotypes. A complex gene expression profile was obtained due to the high gene-, tissue- and genotype-
variability. Despite this, *Mal d 1.06A* and *Mal d 1.07* expression patterns resulted particularly associated with the degree of allergenicity of the different cultivars. They were not the most expressed *Mal d 1* genes in apple but here it was hypotized a relevant importance in the determination of allergenicity for both qualitative and quantitative aspects of the *Mal d 1* gene expression levels. In Chapter 4 a clear modulation for all the 17 PR-10 genes tested in young leaves of Florina after challenging with the fungus *V. inaequalis* have been reported but with a peculiar expression profile for each gene. Interestingly, all the Mal d 1 genes resulted up-regulated except *Mal d 1.10* that was down-regulated after the challenging with the fungus. The differences in direction, timing and magnitude of induction seem to confirm the hypothesis of a subfunctionalization inside the gene family despite an high sequence and structure similarity. Moreover, a modulation of PR-10 genes was showed both in compatible (Gala-*V. inaequalis*) and incompatible (Florina-*V. inaequalis*) interactions contribute to validate the hypothesis of an indirect role for at least some of these proteins in the induced defense responses. Finally, a certain modulation of PR-10 transcripts retrieved also in leaves treated with water confirm their ability to respond also to abiotic stress.

To conclude, the genomic approach used here allowed to create a comprehensive inventory of all the genes of allergen families, especially in the case of extended gene families like *Mal d 1*. This knowledge can be considered a basal prerequisite for many further studies. On the other hand, the specific transcriptional approach make it possible to evaluate the *Mal d 1* genes behavior on different samples and conditions and therefore, to speculate on their involvement on apple allergenicity process. Considering the double nature of Mal d 1 proteins, as apple allergens and as PR-10 proteins, the gene expression analysis upon the attack of the fungus created the base for unravel the Mal d 1 biological functions. In particular, the knowledge acquired in this work about the PR-10 genes putatively more involved in the specific Malus-*V. inaequalis* interaction will be helpful, in the future, to drive the apple breeding for hypo-allergenicity genotype without compromise the mechanism of response of the plants to stress conditions.

For the future, the survey of the differences in allergenicity among cultivars has to be be thorough including other genotypes and allergic patients in the tests. After this, the allelic diversity analysis with the high and low allergenic cultivars on all the allergen genes, in particular on the ones with transcription levels correlated to allergencity, will provide the genetic background of the low ones. This step from genes to alleles will allow the develop of molecular markers for them that might be used to effectively addressed the apple breeding for hypo-allergenicity. Another important step forward for the study of apple allergens will be the use of a specific proteomic approach since apple allergy is a multifactor-determined disease and only an interdisciplinary and integrated approach can be effective for its prevention and treatment.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>Bet v</td>
<td>Betula verrucosa</td>
</tr>
<tr>
<td>Ct</td>
<td>Threshold cycles</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tag</td>
</tr>
<tr>
<td>Fra a</td>
<td>Fragaria ananassa</td>
</tr>
<tr>
<td>IgE</td>
<td>Immunoglobulin E</td>
</tr>
<tr>
<td>LG</td>
<td>Linkage group</td>
</tr>
<tr>
<td>LTP</td>
<td>Lipid transfer protein</td>
</tr>
<tr>
<td>Mal d</td>
<td>Malus domestica</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker assisted selection</td>
</tr>
<tr>
<td>OAS</td>
<td>Oral allergy syndrome</td>
</tr>
<tr>
<td>PR</td>
<td>Pathogenesis Related</td>
</tr>
<tr>
<td>Pru ar</td>
<td>Prunus armeniaca</td>
</tr>
<tr>
<td>Pru av</td>
<td>Prunus avium</td>
</tr>
<tr>
<td>Pru d</td>
<td>Prunus domestica</td>
</tr>
<tr>
<td>Pru p</td>
<td>Prunus persica</td>
</tr>
<tr>
<td>Pyr c</td>
<td>Pyrus communis</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative PCR</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci</td>
</tr>
<tr>
<td>R-gene</td>
<td>Resistance gene</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>Rub i</td>
<td>Rubus ideaus</td>
</tr>
<tr>
<td>SAR</td>
<td>Systemic acquired resistance</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SPT</td>
<td>Skin prick test</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat</td>
</tr>
<tr>
<td>TLP</td>
<td>Thaumetine-like protein</td>
</tr>
<tr>
<td>TMV</td>
<td>Tobacco mosaic virus</td>
</tr>
</tbody>
</table>
Appendix

Appendix 1A

MC-12 clone sequence (from T7-End to Sp6-End)

Text underlined in yellow: *Mal d 1* genes (the coding sequences are indicated in bold)
Text underlined in gray: other genes
Text not underlined but in bold: retrotransposon elements
Text inside boxes: SSR primers

Retrotransposon protein, putative (ORF1) ↓
aaagccagatttgccttcacacaacgagttgtagctcagcccaagcatcct
ccccgccgcatgtaaatctgtagatgaattcatgctctccccctc
tgtcaaaccttcgctgtagctcaggtcatcttttgtaaatctct
ctcaagctcaaacctgcatgtaaatctgattagttaaaagtcct
cagcggcatgtaaatctgattagttaaaagtccttacacaacgag
ttttctctctccaacttctacttactctgtgctttagctttgctgtaat
ccgaaattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Appendix

20001

20501

21001

21501

22001

22501

23001

Mal d 1.06A

MGVLYTEIASVIPPARLYNALVLD

DNLIPKAPQVKTEIREDGGGVTI

KKVSFGSEGYSEYVVKHVEGIDKNEFDNF

DYSVSEIGIDKIEKISYEIKLVANDQGSIKKSINTSHYHTKGDVEIKEEHVK

AGKDKAGLFLKLYIVNLVPAYDN.
tctcactatattatatattatcctccatatttttaactgcaatttgctccaattatgacaccatctgagtagttgttagtatgctttttgtgcagcgtctgatttttgctgatgctagttctgattttttgctggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
26501
gattttggtttatattggttttatcaatctcctgttttaaattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
cactttcttacagcaacacacactctcacactgagttcttggaagggctc
tgcaacactgtgccccagaaagaggaagagcagtccagactgttggatatc
tatatatttttatttggctaaataatattacgtgcaaatattcttattgg
tgtgtgctttttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Appendix

Mal d 1.02

MGVTFENETYSEIPPRLKFALDVA
DNLPKIAAPQAIKHAELDQGGPGT
KRITFEGSQYVVKHDVSIGIN
SAYXILLEGCCDLEITIEKVFSTKVL
SSCSSIIKSHYHTKDVEIKEEHVK
AGKEKALQLKIESLYKHPADAYN.
acttctgtaaagcaaggttcaaaagctcacttaaaaagggagaattctta
ttaaataagaaaaaacaactactatggttattaatataacactctgaact
taaagaaaaagcagcacgccttttttacagtcttcataattctgtttgaa
tagacttgaatgcgagtttaccttttgaaatactaagtttctcttgaact
tataaaaagacaactacatggttattaaatagctttactcttgaactagaagaaagagcaactgcccacgattttacaggtcctataattgtggtaa
gagactatgtcaaaaatagagaaaatcaagtatagggattattggactcct
acatgactaacaaccctaaaaactagaaaccctagtagtaacggctagaa
ctcttaacaacttcaacatcctcccttaaactgaataccacagatattca
gtctactgaaacttgcaaacacccagtaaacctcgcatcttcaagggac
atcaaccttcaatggttttgttaagacatatgcaatttgttcttgagttg
aacactgcaccagctccaaaacccactttgaccaaatctcgaagaaaat
gaaaccttgcatcaatatgcttgctacgaccatgcatcacaggattcttt
gaaagtttgatcgtcgaaacattatcacaatataccaaggttggactaca
cctgctcctgattaagactcttcaagatccttcttaaccatacaacttggc
atatactcgatgctgctgccatgaattccacttcaatggtggataaagta
acaactggttgttttttttaggaccaagaaacacgacactaatacttatt
aaccacgtaaaacctaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
cagactggttgttttttttaggaccaagaaacagcacctaaactcattat
aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
cactatccgtatatctcttggtgtaaaagctcaatagtccttttaacgtaccttagagccttttttgtgtg
Hypothetical protein
(ORF13) ↑

Appendix
Appendix

Copia-type polyprotein (ORF14)↑
159

Appendix

IKAGREKALGMYRVVETYLLQNPDAY

IKAGREKALGMYRVVETYLLQNPDAY

IKAGREKALGMYRVVETYLLQNPDAY

IKAGREKALGMYRVVETYLLQNPDAY
Appendix

74501

Mal d 1.06B

MGVLTYETEASIIPPARLYNALVLDA
DNLIPKIAPQAVKTVEILEGDGGVGTI
KKVSFGSEGYNYVHKVHDKDNDV
VYISILSEAIKIEISKIEYKTWE
SGGSHIKNXHYTGTKDFEIKEHVK
AGKERAGLFLKLENYLVNPDAYN
84501
atatcagagaatatgcaacaatattaacctattagcataagtaagtaaa
gaatattaaataaatatttacctgtacatagtatctgcttttgcgtttgatgc
gaaaccatgtagcaacctatggtctcaatggtttcgaattaatgtcggttga
cacattggaaagaacaaaaaatataactacacattacaaaccttacaaaa
tatcataatcgtacagtgaactcacaactctctcataacacataagcacactagc
ttacaacagcactctagatatctttctctatcattcatataaaaaatcatgtta
tagacaccaaaaaacctcattagtcgcaagttttccgacagaacaaagaaat
aatagttgatattgctgaaagaagctcgcgcaatacgcaattaacagtactgag
actactgtaaaaaggtcatattgtaaggacatagtgggcagaagggaagc
gaagggatttttttttttttttttgattgcgaacacattgacacttatgtttcatgagtttcaatattttcatcaaacatcaataaatgttttcgaaattgc
caccattggaaagaacaaaataatactaaccaattacaaatttcatcaaa
tatcaatcgtatcgtagaactcacaacttcatcaaacataagcacatgac
ttacacgagcctatagatcattttcctctaattcatataaaatcatttga
agtacaccaaaccttacatgtgcgaagttttccagacaacaaagaaaaat
aaatgatgtaatcctgaaaagactcgcgcaatacgcaattaacagtgagt
actactgtaaaaaggtcatattgtaaggacatagtgggcagaagggaagc
gaagggatttttttttttttttttgattgcgaacacattgacacttatgtttcatgagtttcaatattttcatcaaacatcaataaatgttttcgaaattgc
85001
tttgtgatgttacatataattattatgtgtatgtagtacatatttttattttgttattttgatcttttatttttttatttttttttttttttttttttttttatttttttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Appendix

Retrotransposon protein, putative (ORF25)↓

88001

88501

90001

90501

91001

167
acccaccccaagcgaccgaattggcctggtacaccagcaaagacacccgc
tcagagaaagtttaaggctcgtgaggccacagaaccgacccagatct
tctctgctgctttttctcagcaagttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tcagagaaagtttaaggctcgtgaggccacagaaccgacccagatct
tctctgctgctttttctcagcaagttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
gcccccaatattctgtgtgtttggtcctcagcagtctattttggaggattc
cccccaatcaatcgtgacctttcaactacaatcagaacgatcttccggtt
tcagagaaacgctcgtcagagtttttttttagttatcattggaacttt
tgctcctctttgtgtcatccctcatctctccaattctctcaatagata
gtatttacccccaaatcaattgtatgtggaaaccctaatcccaaaattg
Mal d 1.06C↓
MGVLTYETEYAVIPPARLYNALVDA
DNLIPIAPAVKVEILEGGDVYTI
KKVSIPEGSEYVVHVKVVEGIDKDNF
VSYLSLEIDSAIKQNIYSEIKLVA
SNGSIKNIHVHTKGDVEIKEENVK
AGKERAHGLKLFIEHNLVANPARDY
Mal d 1ps pseudogene

Transposon protein, putative (ORF29)
acatatatatgtaacacaattaagttgctcgaacctatcaaatattaaacgttt
tgtctaatattagttgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca
tcgaagacagtcatcgaatctactattaggtgaatcaactgaaagcatcgaattcccccgcgtaca

101501

AcTGCAaatTTGACCTGCAATTGCCTTACTGCAATTTAATCTTTT

102001

MC-12SSRC744CF

103001

MC-12SSRC744CRev
Putative non-LTR retroelement reverse transcriptase (ORF30)↓
Hypothetical protein with Heavy-Metal-Associated (HMA) domain
(ORF31)
tgtggaaacttaatattcctcctaaagttaagatgtttgcgtggttgttc
attcatgatagacttcaaactaggaaaagattaagttatattgtaattaa
tagtgataagatgtgtcttttttgtaattccatagacgaggaccaatgc
atttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Appendix

11501

115501

116001

116501

117001

117501

120001

Hypothetical protein with Glycosyl hydrolases domain and WD40 domain (ORF 32) ↑
aagtagaaacaaaagatgtgagggtatgacatattgcatatgtgaatcatctct
tagtaaacagacataagggagattgtccittgtgaggctttagaatcctcc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttctgcacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
cgcagaggtttcctgcaacagacatcatcaccagaataaaatcatcactccc
Appendix

Appendix 1B

MC-20 clone sequence (from T7-End to Sp6-End)

Text underlined in yellow: Mal d 1 genes (the coding sequences are indicated in bold)
Text underlined in gray: other genes
Text not underlined but in bold: retrotransposon elements
Text inside boxes: SSR primers

1  attcctgtctgtgattcgcgaatgcaggaaggttggtggttggagtgtgctcttgaggtagtattaggttggtggtgacgtttttggtgtgtggtggtgtgtggtgaacgtcaggtgcaggtgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
Putative reverse transcriptase
Appendix

9001

Hypothetical protein with PMD, Plant mobile domain
(ORF36) ↓

9501

10001

10501

11001

11501

12001

Hypothetical protein with PMD, Plant mobile domain
(ORF36) ↓

11501
gaccttattattgttcgttgattaaggatgcatacttagtttgcatgcatgaatttgatgctagaatataagggagtttcatctaattgttatgaactta
tattcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
tctttagtttgtgattttgctattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
tctttagtttgtgattttgctattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
ttttcgtaagtagtggaggttgctagtcacaatcgtgttaagtaaattcc
ttgtcaatgcttatgattttcacaaagcttaatgatctttgattgtatct
cctattatgtaggttagctcattgattattttgctgaattttgagtc
agctttttgctctttcctctgttgaatttttgctgtattttgagtc
12501
13001
Appendix

16001

16501

17001

17501

18001

18501

19001

183

Hypothetical protein (ORF38)
Hypothetical protein with a domain for Reverse transcriptase (RNA-dependent DNA polymerase) (ORF40) ↑
Retrotransposon protein with domain for Retrotransposon gag protein; RVP_2, Retroviral aspartyl protease; RT_LTR; RnaseH (ORF41)↓

Retrotransposon protein with domain for Retrotransposon gag protein; RVP_2, Retroviral aspartyl protease; RT_LTR; RnaseH (ORF41)↓
Appendix

36501
tggcccaagctttttaacagaatgcaattctgcctctcctccccctctccttaaatg
acaccccttggagaccaatgtacaactttatggaattttttttagtgaat
aaggttttggagaggggggttttgagaatttttatttttttggtttgagc
acaccaatcacttttttttttttttttggggtgttaatctagtattttctggat
acgagcagaatgtaatgtaatgtaatgtaatgtaatgtaagttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
tattgaatttacattaagaatagagaaaataatatttaataaacaactgatt
gagtcatcattatgtgtgcaaaaaaaccttttttttagataattttg
tttgaaaaaaacaatttaacaactaatatacattatctagtgcac
gtccaanaagaccccttttttttaacccgcattacaacccccttttaggtgtt
tgaacgctgttcataaatggagaataaatattatatgacacactgtgaa
tacatcattatgtgcctattgtgaggactaatttttttaaaaatgtacaa
aaagaattataattaattaaaacacgctgtgaatatgcacaaaaatatcccc
tacattttttgatatttttttggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
catttttgtccaaaaatatattttggaacgtgtgtgacccccaanaaggtt
tggtcttataatttttaaatggtactcgaagattgcatc
aatccccaaatggttggaaatagttttcccttggtcactctgtgcagccact
catcgaattgtgtgcaatctgactttttttttttgtcagctattatta
agcgcggactttttttgtcagctttttttgggctcaaatatccctccttttaattttttgggttgcctttgaaattttttgttttggggt
agcccgttttaacgtctggtacacaaatgaacatccagaaaaacgcata
gtcatatagtacaccaacaggaaaaatctctacgctccactaagga
tcggaaatagttgtgagaacatacgtttttaatgattgaatgtacttcaat
gtacggtaccttcttacctgcagctatcagctaagctcgtacagtgtgctagtgcatgtaaaggtgatatttaaatggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
53501
aatctaatataagatactcaacctctataatacttgattatgttgctata
tagcttaaatctcgctttgagattgatgttggatagaagttttgggaaaatattgatttt
ttgcttaaaattttgagaagatttggggacttggaagttttgggaaaatattgatttt
ttgcttaaaattttgagaagatttggggacttggaagttttgggaaaatattgatttt
atccgcaacatatttataaatggatgagatgtttttttttttttaaatttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Mal d 1.08↓

MGVTYESETTSVIPPARLFNATALDGDELI
AKLAPQAVKSEILEEGGAVTVQKII
EGSTNGYVKKIDVIDKDNFVYKYSMEGDA
ISETIEKISYETTILAVSSGISIKRTCHY
HTKGDVEINEHKLASKESSHLLKLVENYL
LEHQDAYN
Hypothetical protein with a conserved domain with unknown function: DUF789 (ORF47)
acttatggagatgatgtgtgtgactcaacacccttcgtatataattct
cactatcatatgactcgtttaacaccctgtatcatatcatgtgatgtat
ttcactgatttacaaatgccccaaatgttaaaccctcttggtccacaccctt
ggcatatgtatatatatattataatatagatagcatttcctttttccacacctttta
taaatatgtaatggaaccttttctctctctctctcttttggcccaaatatatgggt
gcagaaaaagacgaacetcaaatataaccaacagccttctgtgatgaattaagt
aactacctgtgaaaaaccataacactgtgttatacatcaaatctatgatc
cgcgattttcaactggtcagcagctcataaaaagcctttttttttttttttgta
tatcataactgtgacacctataaatccccatgtcgtgtgtttgacgttgaatgct
atcccagagatgcagatcatacagcttctcttcagcttggttataagacgcttt
gattttgaatggtggtttgacggtggtggtgtctttttgacggtggtggtgtggtcctt
ttatagagcgttggttttatatatctctactctgggaggttgattgtggaa
cctatatgtcactatgcttttaaatgctagacacatcaatataactccccatg
tgtgtggtgctccctttttggaggtgatgactctgtctggtgtggtgtgctatctc
tggtcactcagaactgtcctctgctcactctgtgactgtgctatctc
tggctcatcagttctgctcactctgtgactgtgctatctc
tttgccattgtgctcactctgtgactgtgctatctc
ttcgccattgtgctcactctgtgactgtgctatctc
ttcgccattgtgctcactctgtgactgtgctatctc
ttcgccattgtgctcactctgtgactgtgctatctc
ttcgccattgtgctcactctgtgactgtgctatctc
HEAT repeat-containing protein (ORF49) ↑
atacagttcacaatatataattaacgtcttttcactcttcatgtagtttttcacat
ccaaaaagcaaaaaagcaattaagatgaatgctgatctggttttttgaccat
tatatatatcttcacacacacacagcagaacactgtcgagctactatcata
tcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Appendix

MC-20SSR96C2For

MC-20SSR96C2Rev

Multi-domain protein
(a sensor histidine kinase and a response regulator)

(ORF50) ↓

212
Protein COBRA precursor, putative (ORF52) ↑
agccagacctgattgctttctcttaaattcgaagaggcatcgttc
tccgacactcagacgactttttcaagtgctctgacagagttaaaacatgtgaagctggcagctcccactaccgtgctttgag
tgagaaagatggaaggctaggctagaagtcgaagctcccactaccgtgcctttcttgctagtgcagcgctatgtgcagc
tctcatctcctaaacactcatatgtgttctcaggcgagcttcagctccaccttgcctttcttcaggttaaaccagagataac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
tcaggttaaaaactggaagctggcagctcccactaccgtgctttcaggttaaaccagagataacccttctcttcaggttttac
## Appendix 2

### Mal d 1 sequences added in the alignment of Chapter 2

<table>
<thead>
<tr>
<th>N.</th>
<th>Gene</th>
<th>Allele</th>
<th>Source*</th>
<th>GeneBank accession num.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Mal d 1.01</em></td>
<td>Mal d 1.0105.01</td>
<td>PM</td>
<td>AY789236</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Mal d 1.0105.02</td>
<td>FS</td>
<td>AY789238</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Mal d 1.0105.03</td>
<td>PM</td>
<td>AY789237</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Mal d 1.0109</td>
<td>GD</td>
<td>AY026910</td>
</tr>
<tr>
<td>5</td>
<td><em>Mal d 1.02</em></td>
<td>Mal d 1.0201.01</td>
<td>PM</td>
<td>AY789240</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Mal d 1.0201.09</td>
<td>PM</td>
<td>AY789239</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Mal d 1.0209.01</td>
<td>DS</td>
<td>AY827660</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Mal d 1.0201.02</td>
<td>FL</td>
<td>Not yet</td>
</tr>
<tr>
<td>9</td>
<td><em>Mal d 1.03A</em></td>
<td>Mal d 1.03A01.01</td>
<td>FS</td>
<td>AY789263</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Mal d 1.03A01.01</td>
<td>GD</td>
<td>AY822722</td>
</tr>
<tr>
<td>11</td>
<td><em>Mal d 1.03B</em></td>
<td>Mal d 1.03B01</td>
<td>PM</td>
<td>AY789264</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Mal d 1.03B02</td>
<td>FS</td>
<td>AY789265</td>
</tr>
<tr>
<td>13</td>
<td><em>Mal d 1.03C</em></td>
<td>Mal d 1.03C01</td>
<td>PM</td>
<td>AY789266</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>Mal d 1.03C02</td>
<td>GD</td>
<td>AY822725</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>Mal d 1.03C03</td>
<td>GD</td>
<td>AY822726</td>
</tr>
<tr>
<td>16</td>
<td><em>Mal d 1.03D</em></td>
<td>Mal d 1.03D01</td>
<td>PM</td>
<td>AY789267</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>Mal d 1.03D02</td>
<td>FS</td>
<td>AY789268</td>
</tr>
<tr>
<td>18</td>
<td><em>Mal d 1.03E</em></td>
<td>Mal d 1.03E01</td>
<td>PM</td>
<td>AY789270</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>Mal d 1.03E02</td>
<td>GD</td>
<td>AY822729</td>
</tr>
<tr>
<td>20</td>
<td><em>Mal d 1.03F</em></td>
<td>Mal d 1.03F01.01</td>
<td>PM</td>
<td>AY789271</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Mal d 1.03F01.02</td>
<td>PM</td>
<td>AY789272</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>Mal d 1.03F02.01</td>
<td>FS</td>
<td>AY789273</td>
</tr>
<tr>
<td>23</td>
<td><em>Mal d 1.03G</em></td>
<td>Mal d 1.03G01</td>
<td>GD</td>
<td>AY789273</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>Mal d 1.03G01</td>
<td>FS</td>
<td>AY789274</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>Mal d 1.03G02</td>
<td>FL</td>
<td>Not yet</td>
</tr>
<tr>
<td>26</td>
<td><em>Mal d 1.04</em></td>
<td>Mal d 1.0404.01</td>
<td>PM</td>
<td>AY789242</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>Mal d 1.0404.02</td>
<td>PM</td>
<td>AY789243</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>Mal d 1.0405</td>
<td>DS</td>
<td>AY827665</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>Mal d 1.0406</td>
<td>IM</td>
<td>AY827666</td>
</tr>
<tr>
<td>30</td>
<td><em>Mal d 1.05</em></td>
<td>Mal d 1.0501</td>
<td>PM</td>
<td>AY789245</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>Mal d 1.0502</td>
<td>FS</td>
<td>AY789247</td>
</tr>
<tr>
<td>32</td>
<td><em>Mal d 1.06A</em></td>
<td>Mal d 1.06A01.01</td>
<td>PM</td>
<td>AY789250</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>Mal d 1.06A01.02</td>
<td>GS</td>
<td>AM283491</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>Mal d 1.06A02.01</td>
<td>PM</td>
<td>AY789248</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>Mal d 1.06A03</td>
<td>RD</td>
<td>AM283501</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>Mal d 1.06A02.02</td>
<td>PS</td>
<td>AY827692</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>Mal d 1.06A02.03</td>
<td>IM</td>
<td>AY827698</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>Mal d 1.06A02.04</td>
<td>DS</td>
<td>AY827700</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>Mal d 1.06A02.05</td>
<td>FL</td>
<td>Not yet</td>
</tr>
<tr>
<td>40</td>
<td><em>Mal d 1.06B</em></td>
<td>Mal d 1.06B01</td>
<td>FS</td>
<td>AY789251</td>
</tr>
<tr>
<td>41</td>
<td></td>
<td>Mal d 1.06B02</td>
<td>PM</td>
<td>AY789252</td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>Mal d 1.06B03</td>
<td>PM</td>
<td>AY789253</td>
</tr>
<tr>
<td>43</td>
<td></td>
<td>Mal d 1.06B02</td>
<td>FL</td>
<td>Not yet</td>
</tr>
<tr>
<td>44</td>
<td><em>Mal d 1.06C</em></td>
<td>Mal d 1.06C01</td>
<td>FS</td>
<td>AY789254</td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>Mal d 1.06C02</td>
<td>PM</td>
<td>AY789255</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>Mal d 1.06C03</td>
<td>PM</td>
<td>AY789256</td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>Mal d 1.06C04.01</td>
<td>DS</td>
<td>AY827726</td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>Mal d 1.06C05</td>
<td>RD</td>
<td>AY827727</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>Mal d 1.06C06</td>
<td>IM</td>
<td>AY827728</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>Mal d 1.06C02</td>
<td>FL</td>
<td>Not yet</td>
</tr>
<tr>
<td>51</td>
<td><em>Mal d 1.07</em></td>
<td>Mal d 1.0701</td>
<td>FS</td>
<td>AY789257</td>
</tr>
<tr>
<td>52</td>
<td></td>
<td>Mal d 1.0702</td>
<td>PM</td>
<td>AY789258</td>
</tr>
<tr>
<td></td>
<td>Mal d 1.0703</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Mal d 1.08</td>
<td><strong>PM</strong></td>
<td>AY789259</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Mal d 1.0801.01</td>
<td>PM</td>
<td>AY789261</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Mal d 1.0801.02</td>
<td>PM</td>
<td>AY789261</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>Mal d 1.0801.01</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>Mal d 1.0901</td>
<td><strong>PM</strong></td>
<td>AY789262</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Mal d 1.0902</td>
<td>GD</td>
<td>AY822720</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>Mal d 1.0903</td>
<td>GD</td>
<td>AY822721</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Mal d 1.0903</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Mal d 1.1001</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Mal d 1.1101</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>Mal d 1.1201</td>
<td>FL</td>
<td>Not yet</td>
<td></td>
</tr>
</tbody>
</table>

*PM=Prima, FS=Fiesta; GD=Golden Delicious; RD=Red Delicious; IM=Ingrid Marie; PS=Priscilla; DS=Discovery; GS=Granny Smith; RG=Royal Gala.
All the sequences derived from genomic DNA except the sequence of the cultivar Florina that derived from the plasmidic DNA of the BAC clones
Acknowledgements

This thesis was carried out part in the ISAFRUIT Project funded by the European Commission under the Thematic Priority 5 - Food Quality and Safety of the 6th Framework Programme of RTD (Contract No. FP6-FOOD-CT-2006-016279), and in part in the project funded by the University of Bologna and coordinated by Professor Donatella Serafini-Fracassini, under the thematic: Cross allergens in pollen and fruits: modulation by climate changes of their allergenic potential (2006). The views and opinions expressed in this thesis are purely those of the writers and may not, in any circumstances, be regarded as stating an official position of the European Commission.

It is my great pleasure to express my gratitude to those have given me support, help and friendship throughout these years. First of all, I would like to thank my promoter Prof. Silviero Sansavini, and my co-promoters Dr. Roberta Paris and Dr. Stefano Tartarini for their scientific supervision during all my work and the writing of the final dissertation. On many occasions I would be lost without their help to find the right track. In particular, I would like to express my gratitude to Roberta also for her constant help in the lab work. Secondly, many thanks to Dr. Eric van de Weg and Dr. Paul Arens for their supervision and support during my period at the Plant Research International (Wageningen, The Netherlands). I hope to have the opportunity to spend other periods at the PRI and that the collaboration between our two groups may continue. A special thanks goes to Eric for his accurate reading of this manuscript, for the critical comments on the experimental chapters and for the great contribution to making them logical and concise.

I would like to express my gratitude also to Professor Lars Ove Dragsted (University of Copenhagen) for agreeing to do the external reviewer of the thesis despite the short time available to read it.

Many thanks go to my colleagues of DCA and PRI for all the help in experimental techniques or just for their encouragement to hold on in the difficult moments of my PhD.

I also thank colleagues involved in the project founded by Bologna University, Dr. Giampaolo Ricci, Professor Stefano Del Duca, Dr. Gianluca Tasco, for collaboration and useful exchanges of ideas in these years.

Thanks to Dr. Henk Schouten for allowing me to use the apple BAC library available at the PRI for the work described in Chapter 5.
Heartfelt thanks go to Taia (my portuguese sister), Matteo and Stathis for their contribute to making me feel at home during my stay in Wageningen.

I am grateful to all my big family and friends that always supported and encouraged me during these three years. A special mention is for Francesco who is not here yet but already is in my heart.

Finally, I'm sure that stay by my side in these last years must not been easy. For this reason and for a thousand of other reasons, the last and biggest thanks is for Andrea.

I dedicated this thesis to my grandfather Alfonso who strongly supported me in all my challenges. I was far away from him and my family for my PhD research during the last period of his life. Probably I’ll never completely forgive my self for this absence but I am confident that now he is watching me and is proud of my result. Grazie nonno.

Giulia