

ALMA MATER STUDIORUM-UNIVERSITÀ DEGLI STUDI DI BOLOGNA

FACOLTÀ DI AGRARIA

DOTTORATO DI RICERCA IN COLTURE ARBOREE ED AGROSISTEMI
FORESTALI ORNAMENTALI E PAESAGGISTICI

SETTORE SCIENTIFICO-DISCIPLINARE
DI AFFERENZA: AGR/03

**GENOMIC AND TRANSCRIPTIONAL ANALYSIS
OF ALLERGEN GENES IN APPLE**

(Malus x domestica)

presentata da

Giulia Pagliarani

Co-relatori:

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

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

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 includes 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 result 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 (Maliapaard 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 has 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 a 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 (Maliapaard 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

(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 acquired 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 addition, 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 immunoglobulin (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 on 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), Thaumatin-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 1. List of allergens from *Rosaceae* fruits and family classification

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

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

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., 200) 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 Loon 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; 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 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.,

2002). 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 itself imply 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 are not only induced *de novo* upon pathogen attack but also under abiotic stresses such as wounding or other physical or chemical stress. Surprisingly, in some cases 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 homologous 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 led 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

- Andersen MB, Hall S, Dragsted LO. 2009. Identification of European Allergy Patterns to the Allergen Families PR-10, LTP, and Profilin from Rosaceae Fruits. *Clin Rev Allergy Immunol* Oct 23. [Epub ahead of print.]
- Asero R, Antonicelli L, Arena A, Bommarito L, Caruso B, Crivellaro M et al. 2009. EpidemAAITO: features of food allergy in Italian adults attending allergy clinics: a multi-centre study. *Clin Exp Allergy* 39(4):547–555.
- Asero R, Ballmer-Weber BK, Beyer K, Conti A, Dubakiene R, Fernandez-Rivas M et al. 2007. IgE-Mediated food allergy diagnosis: current status and new perspectives. *Mol Nutr Food Res* 51(1):135–147.
- Asero, R.; Mistrello, G.; Roncarolo, D.; Amato, S.; Zononi, D.; Barocci, F.; Caldironi, G. 2003. Detection of clinical markers of sensitization to profilin in patients allergic to plant-derived foods. *J. Allergy Clin. Immunol.* 112, 427–432.
- Ballmer-Weber, B. K. 2002. Lipid transfer protein as a potential panallergen? *Allergy* 57, 873–875.
- Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA. 2004. Characterisation of Mal d 1-related genes in Malus. *Plant Mol Biol* 55:369-388.
- Bohle B. 2004. T lymphocytes and food allergy. *Mol Nutri Food Res* 48:424-433.
- Bohaar ST, Zuidmeer L, Ma Y, Ferreira F, Buijnzeel-Koomen CA, Hoffmann-Sommergruber K, van Ree R, Knulst AC. 2005. A mutant of the major apple allergen, Mal d 1, demonstrating hypo-allergenicity in the target organ by double-blind placebo-controlled food challenge. *Clin Exp Allergy*. 35(12):1638-44.
- Borges, J. P.; Jauneau, A.; Brule´, C.; Culerrier, R.; Barre, A.; Didier, A.; Rouge´, P. 2006. The lipid transfer proteins (LTP) essentially concentrate in the skin of Rosaceae fruits as cell surface exposed allergens. *Plant Physiol. Biochem.* 44, 535–542.
- Botton A., Lezzer P., Dorigoni A., Barcaccia G., Ruperti B., Ramina A. (2008) - Genetic and environmental factors affecting allergen-related gene expression in apple fruit (*Malus domestica* L. Borkh). *J Agric Food Chem.*, 56(15), 6707-16.
- Bousquet J, Anto JM, Bachert C, Bousquet PJ, Colombo P, Cramer R et al. 2006. Factors responsible for differences between asymptomatic subjects and patients presenting an IgE sensitization to allergens. A GA(2)LEN project. *Allergy* 61:671–680.
- Breiteneder H, Mills C. 2006. Structural bioinformatic approaches to understand cross-reactivity. *Mol Nutr Food Res*. 50:628-32. Review.
- Breiteneder H, Radauer C. 2004. A classification of plant food allergens. *J Allergy Clin Immunol* 113(5):821–830.
- Breiteneder H. 2004. Thaumatin-like proteins – a new family of pollen and fruit allergens. *Allergy*; 59: 479–481.
- Buijnzeel-Koomen C, Ortolani C, Aas K, Bindslev-Jensen C, Björkstén B, Moneret-Vautrin D, Wüthrich B. 1995. Adverse reactions to food. European Academy of Allergology and Clinical Immunology Subcommittee. *Allergy*. 50:623-35.
- Carimi F, Zottini M, Formentin E, Terzi M, Lo Schiavo F. 2003. Cytokinins: new apoptotic inducers in plants. *Planta*. 216:413-21.
- Celton JM, Chagné D, Tustin SD, Terakami S, Nishitani C, Yamamoto T, Gardiner SE. 2009. Update on comparative genome mapping between Malus and Pyrus. *BMC Res Notes*. 14;2:182.
- Chadha P, Das RH. 2006. A pathogenesis related protein, AhPR10 from peanut: an insight of its mode of antifungal activity. *Planta*. 225(1):213-22.
- Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EH, Gardiner SE, Korban SS. 2008. Development of a set of SNP markers present in expressed genes of the apple. *Genomics*. 92:353-8.
- Chung KM, Igari K, Uchida N, Tasaka M. 2008. New perspectives on plant defense responses through modulation of developmental pathways. *Mol Cells*. 26:107-12.
- Cuesta-Herranz J, Lazaro M, Figueredo E, Igea JM, Umpierrez A, De-Las-Heras M. 2000. Allergy to plant-derived fresh foods in a birch- and ragweed-free area. *Clin Exp Allergy* 30:1411–1416.
- De Amici M, Moratti R, Quaglini S, Vignini MA, Mosca M. 2002. IgE to Bet v 1/Bet v 2 in UAS. *Allergy*. 57:181-2.

- Diaz-Perales A, Garcia-Casado G, Sanchez-Monge R, Garcia-Selles FJ, Barber D, Salcedo G. 2002. cDNA cloning and heterologous expression of the major allergens from peach and apple belonging to the lipid-transfer protein family. *Clin Exp Allergy* 32:87–92.
- Ebner C, Birkner T, Valenta R, Rumpold H, Breitenbach M, Scheiner O, Kraft D. 1991. Common epitopes of birch pollen and apples--studies by western and northern blot. *J Allergy Clin Immunol.* 88:588-94.
- Eriksson NE, Moller C, Werner S, Magnusson J, Bengtsson U, Zolubas M. 2004. Self-reported food hypersensitivity in Sweden, Denmark, Estonia, Lithuania, and Russia. *J Investig Allergol Clin Immunol* 14:70–79.
- Fernandes H, Pasternak O, Bujacz G, Bujacz A, Sikorski MM, Jaskolski M. 2008. *Lupinus luteus* pathogenesis-related protein as a reservoir for cytokinin. *J Mol Biol.* 378:1040-51.
- Fernandez-Rivas M, Bolhaar S, Gonzalez- Mancebo E, Asero R, van Leeuwen A, Bohle B, Ma Y, Ebner C, Rigby N, Sancho AI, Miles S, Zuidmeer L, Knulst A, Breiteneder H, Mills C, Hoffmann-Sommergruber K, van Ree R. 2006. Apple allergy across Europe: how allergen sensitization profiles determine the clinical expression of allergies to plant foods. *J Allergy Clin Immunol* ; 118: 481–488.
- Fernandez-Rivas M, Miles S. 2004. Food allergies: clinical and psychosocial perspectives. In: Mills ENC, Shewry PR (eds) *Plant Food Allergens*, Blackwell, pp1-23.
- Flores T, Alape-Giron A, Flores-Diaz M, Flores HE. 2002. Ocatin, a novel tuber storage protein from the andean tuber crop oca with antibacterial and antifungal activities. *Plant Physiol*; 128: 1291–302.
- Friedman AR, Baker BJ. 2007. The evolution of resistance genes in multi-protein plant resistance systems. *Curr Opin Genet Dev.* 17:493-9.
- Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ. 2005a. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet.* 111:171-83.
- Gao ZS, Weg WE, Schaart JG, Arkel G, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJ. 2005b. Genomic characterization and linkage mapping of the apple allergen genes Mal d 2 (thaumatin-like protein) and Mal d 4 (profilin). *Theor Appl Genet*; 111: 1087–1097.
- Gao, Z. S.; van de Weg, W. E.; Schaart, J. G.; van der Meer, I. M.; Kodde, L.; Laimer, M.; Breiteneder, H.; Hoffmann-Sommergruber, K.; Gilissen, L. J. W. J. 2005c. Linkage map positions and allelic diversity of two Mal d 3 (non-specific lipid transfer protein) genes in the cultivated apple (*Malus domestica*). *Theor. Appl. Genet.* 110, 479–491.
- Gasic K., Gonzalez D.O., Thimmapuram J., Liu L., Malnoy M., Gong G., Han Y., Vodkin L.O., Aldwinckle H.S., Carroll N.J., Orvis K.S., Goldsbrough P., S., Pape D., Fulton L., Martin J., Theising B., Wisniewski M.E., Fazio G., Feltus F.A., Korban S.S. 2009. Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. *The plant genome*, 2: 23-38.
- genome. *Theor Appl Genet* 106: 1497–1508.
- Han Y, Gasic K, Marron B, Beever JE, Korban SS. 2007. A BAC-based physical map of the apple genome. *Genomics* 89: 630–637.
- Helsper JPPFG, Gilissen LJWJ, van Ree R, America Antoine HP, Cordewener Jan HG, Bosch D. 2002. Quadrupole time-of-flight mass spectrometry: A method to study the actual expression of allergen isoforms identified by PCR cloning. *J Allergy Clin Immunol* 110:131-138.
- Hoffmann-Sommergruber. 2002. Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions* 30, 930-935.
- Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills EN. 2005. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *J Allergy Clin Immunol.* 115(1):163-70.
- Kader, J. C. Lipid-transfer proteins in plants. *Ann. ReV. Plant Physiol. Plant Mol. Biol.* 1996, 47, 627–654.
- Karamloo F, Scheurer S, Wangorsch A, May S, Hausteine D, Vieths S. 2001. Pyr c 1, the major allergen from pear (*Pyrus communis*), is a new member of the Bet v 1 allergen family. *Journal of Chromatography B- Analytical Technologies in the Biomedical and Life Sciences* 756(1–2):281–293.

- King TP, Hoffman D, Lowenstein H, Marsh David G, Platts Mills Thomas AE, Thomas W. 1995. Allergen nomenclature. *J Allergy Clin Immunol* 96:5–14.
- Knekt, P.; Isotupa, S.; Rissanen, H.; Heliovaara, M.; Jarvinen, R.; Hakkinen, S.; Aromaa, A.; Reunanen, A. 2000. Quercetin intake and the incidence of cerebrovascular disease. *Eur. J. Clin. Nutr.* 54, 415–417.
- Knekt, P.; Jarvinen, R.; Reunanen, A.; Maatela, J. 1996. Flavonoid intake and coronary mortality in Finland: A cohort study. *Br. Med. J.* 312, 478–481.
- Koistinen KM, Soininen P, Venäläinen TA, Häyrinen J, Laatikainen R, Peräkylä M, Tervahauta AI, Kärenlampi SO. 2005. Birch PR-10c interacts with several biologically important ligands. *Phytochemistry*. 66(21):2524-33.
- Krebitz, M.; Wagner, B.; Ferreira, F.; Peterbauer, C.; Campillo, N.; Witty, M.; Kolarich, D.; Steinkellner, H.; Scheiner, O.; Breiteneder, H. 2003. Plant-based heterologous expression of Mal d 2, a thaumatin-like protein and allergen of apple (*Malus domestica*) and its characterization as an antifungal protein. *J. Mol. Biol.* 329, 721–730.
- Le Marchand, L.; Murphy, S. P.; Hankin, J. H.; Wilkens, L. R.; Kolonel, L. N. 2000. Intake of flavonoids and lung cancer. *J. Natl. Cancer Inst.* 92, 154–160.
- Liebhart R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 40 new microsatellites in apple (*Malus domestica* Borkh.). *Mol Breed* 10: 217–241.
- Liebhart R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003a) Mapping quantitative physiological traits in apple (*Malus x domestica* Borkh.). *Plant Mol Biol* 52: 511–526.
- Liebhart R, Koller B, Patocchi A, Kellerhals M, Pfammatter W, Jermini M, Gessler C (2003b) Mapping quantitative field resistance against apple scab in a 'Fiesta' x 'Discovery' progeny. *Phytopathology* 93: 493–501.
- Liebhart R, Koller B, Gianfranceschi L, Gessler C (2003c) Creating a saturated reference map for the apple (*Malus domestica* Borkh.)
- Ma Y, Gadermaier G, Bohle B, Bolhaar S, Knulst A, Markovic-Housley Z, Breiteneder H, Briza P, Hoffmann-Sommergruber K, Ferreira F. 2006. Mutational analysis of amino acid positions crucial for IgE-binding epitopes of the major apple (*Malus domestica*) allergen, Mal d 1. *Int Arch Allergy* 139:53-62.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41:221-30.
- Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, et al. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor Appl Genet* 97: 60–73.
- Memelink J, Hoge JH, Schilperoort RA. 1987. Cytokinin stress changes the developmental regulation of several defence-related genes in tobacco. *EMBO J.* 6:3579-3583.
- Mills EN, Mackie AR, Burney P, Beyer K, Frewer L, Madsen C, Botjes E, Crevel RW, van Ree R. 2007. The prevalence, cost and basis of food allergy across Europe. *Allergy*. 62:717-22. Review.
- Mogensen JE, Wimmer R, Larsen JN, Spangfort MD, Otzen DE. 2002. The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. *J Biol Chem*; 277:23684–92.
- Naik S, Hampson C, Gasic K, Bakkeren G, Korban SS. 2006. Development and linkage mapping of E-STS and RGA markers for functional gene homologues in apple. *Genome* 49: 959–968.
- Neudecker P, Schweimer K, Nerkamp J, Scheurer S, Vieths S, Sticht H, et al. 2001. Allergic cross-reactivity made visible: solution structure of the major cherry allergen Pru av 1. *J Biol Chem*; 276:22756–63.
- Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, et al. 2006. Analyses of expressed sequence tags from apple. *Plant Physiol* 141: 147–166.
- Nicoletti C, Temblay JN, Ivory K. 2007. The immune system and allergic responses to food. In: Mills C, Wichers H, Hoffmann-Sommergruber K (eds) *Managing allergens in food*. Woodhead, Cambridge, pp 29–46.
- Oh, D. H.; Song, K. J.; Shin, Y. U.; Chung, W. I. Isolation of a cDNA encoding a 31-kDa, pathogenesis-related 5/thaumatin-like (PR5/TL) protein abundantly expressed in apple fruit (*Malus domestica* cv. Fuji). *Biosci., Biotechnol., Biochem.* 2000, 64, 355–362.
- Pagliarini G., Paris R., Tartarini S., Sansavini S. Cloning and expression of the major allergen genes in apple fruit. *JHSB ISAFRUIT Special Issue* 176–181.

- Paris, R., Cova, V., Pagliarini, G., Tartarini, S., Komjanc, M. and Sansavini, S. (2009). Expression profiling in HcrVf2-transformed apple plants in response to *Venturia inaequalis*. *Tree Genetics & Genomes*, 5, 81–91.
- Pasternak O, Bujacz GD, Fujimoto Y, Hashimoto Y, Jelen F, Otlewski J, Sikorski MM, Jaskolski M. 2006. Crystal structure of *Vigna radiata* cytokinin-specific binding protein in complex with zeatin. *Plant Cell*. 18:2622–34.
- Puehringer H, Zinoecker I, Marzban G, Katinger H, Laimer M. 2003. MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene* 321:173-183.
- Pühringer H, Moll D, Hoffmann-Sommergruber K, Watillon B, Katinger H, Laimer da Câmara Machado M. 2000. The promoter of an apple Ypr10 gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. *Plant Sci*;152:35–50.
- Salcedo G, Sanchez-Monge R, Barber D, az-Perales A. 2007. Plant non-specific lipid transfer proteins: an interface between plant defence and human allergy. *Biochimica et Biophysica Acta-Molecular and Cell Biology of Lipids* 1771(6):781–791.
- Sampson HA. 2005. Food allergy--accurately identifying clinical reactivity. *Allergy*. 60 Suppl 79:19-24. Review.
- Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T. 2007. S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. *Genetics* 175:1869-81.
- Sax K. 1933. The origin of *Pomoideae*. *J Am Soc Sci* 30: 147-150.
- Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arús P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE. 2008. Multiple models for Rosaceae genomics. *Plant Physiol*. 147(3):985-1003.
- Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, Van Kaauwen MPW, Walser M, Kodde LP, Soglio V, Gianfranceschi L, Durel CE, Costa F, et al (2006) Microsatellite markers spanning the apple (*Malus domestica* Borkh.) genome. *Tree Genet Genomes* 2: 202–224.
- 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(4):201-15.
- Sun HY, Dai HY, Zhao GL, Ma Y, Ou CQ, Li H, Li LG, Zhang ZH. 2008. Genome-wide characterization of long terminal repeat -retrotransposons in apple reveals the differences in heterogeneity and copy number between Ty1-copia and Ty3-gypsy retrotransposons. *J Integr Plant Biol*. 50(9):1130-9.
- Tao Y, Xie Z, Chen W, Glazebrook J, Chang HS, Han B, Zhu T, Zou G, Katagiri F. 2003. Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*. 15:317-30.
- Valenta R, Duchene M, Ebner C, Valent P, Sillaber C, Deviller P, Ferreira F, Tejkl M, Edelmann H, Kraft D, et al. 1992. Profilins constitute a novel family of functional plant pan-allergens. *J Exp Med*. 175(2):377-85.
- Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S. 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell*.14(7):1441-56.
- van Loon L.C., Rep M., Pieterse C.M. 2006. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- van Loon LC, Bakker PA, Pieterse CM. 1998. Systemic resistance induced by rhizosphere bacteria. *Annu Rev Phytopathol*. 36:453-83.
- Vanek-Krebitz M, Hoffmann-Sommergruber K, Laimer da Camara Machado M, Susani M, Ebner C, Kraft D, Scheiner O, Breiteneder H. 1995. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem Biophys Res Commun*. 14;214:538-51.
- Vinatzer B.A., Zhang H.B., Sansavini S. 1998. Construction and characterization of a bacterial artificial chromosome library of apple. *Theor Appl Genet.*, 97:1183-1190.

- Vinatzer, B.A., A. Patocchi, L. Gianfranceschi, S. Tartarini, H.B. Zhang, C. Gessler, S. Sansavini. 2001. Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with Vf apple scab resistance. *Mol. Plant Microbe Interact.* 14:508–515.
- Vlot AC, Klessig DF, Park SW. 2008. Systemic acquired resistance: the elusive signal(s). *Curr Opin Plant Biol.* 11:436-42.
- Wensing M, Akkerdaas JH, van Leeuwen WA, Stapel SO, Bruijnzeel-Koomen CA, Aalberse RC, Bast BJ, Knulst AC, van Ree R. 2002. IgE to Bet v 1 and profilin: cross-reactivity patterns and clinical relevance. *J Allergy Clin Immunol.* 110:435-42.
- Wensing, M.; Akkerdaas, J. H.; van Leeuwen, W. A.; Stapel, S. O.; Bruijnzeel-Koomen, C. A.; Aalberse, R. C.; Bast, B. J.; Knulst, A. C.; van Ree, R. 2002. IgE to Bet v 1 and profilin: Cross-reactivity patterns and clinical relevance. *J. Allergy Clin. Immunol.* 110, 435–442.
- Xu M, Song J, Cheng Z, Jiang J, Korban SS (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene Vf. *Genome* 44: 1104–1113.
- Xu ML, Korban SS. 2002. AFLP-derived SCARs facilitate construction of a 1.1 Mb sequence-ready map of a region that spans the Vf locus in the apple genome. *Plant Mol Biol* 50: 803–818.
- Zuberbier T, Edenharter G, Worm M, Ehlers I, Reimann S, Hantke T et al. 2004. Prevalence of adverse reactions to food in Germany—a population study. *Allergy* 59:338–345.
- 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 (Thaumatin-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 allergens 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 allergenicity.

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 *Mal d 1* genes in young apple leaves upon challenge with a biotic stress (*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 *Mal d 2* and *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) were 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 μ l volume containing 200 ng of DNA from BAC library pools, 0.1 μ M gene-specific primers (Table 1), 1.5 mM MgCl₂, 100 μ M 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 μ l volume containing 50 ng of plasmid DNA, 0.1 μ M gene-specific primers, 1.5 mM MgCl₂, 100 μ M 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 $E < E^{-15}$ against the National Center for Biotechnology Information (NCBI) non-redundant protein database. The predicted genes were also searched for similarity against the NCBI nucleotide sequences database by BLASTN (Altschul et al., 1990) with E-value cut off $E < E^{-25}$. 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.

	Primer names	Sequence (5'-3')	Ta (°C)	Product (bp)
Subfam I	Md1-1For	TGTCAAACTATTACAGCGCTAGTGG	60	176
	Md1-1Rev	CTTCAATCAAAGTGTAGGCGTATGA		
Subfam II	Md1-2For	CGAAGCGGATGGAGGTGT	60	315
	Md1-2Rev	GATCCTCCATCAGGAGATGC		
Subfam III	Md1-3For	GCTCCACAAGCAGTCAAACT	57	232
	Md1-3Rev	TTCACTCCCTGTCAAATGAAAA		
Subfam IV	Md1-4For	TCAACTTTGGTGAAGGTAGCACA	60	198
	Md1-4Rev	TGGTGTGGTAGTGGCTGGTA		
Generic*	Md1-5For	GTACAATGCCTTTGTTCTTGATG	57	375; 471; 489; 528
	Md1-5Rev	TCTTGCCAGCCTTGACATGCTCTTC		

*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).

Isoallergens	Primer names	Sequence (5'-3')	Ta (C°)	Product (bp)
<i>Mal d 1.01</i>	Mal d 1.01general F	TGACTCGATTGACGAAGCAAG	54	166
	Mal d 1.0105 R	TTCAATGTTTCCCTTGGTG <u>AGA</u>		
<i>Mal d 1.02</i>	Mal d 1.0209 F	GCCCTGGAACCATCAAGAAT	51	419
	Mal d 1.02general R	GTGCTCTTCCTTGATCTCA <u>CCAT</u>		
<i>Mal d 1.03A-F*</i>	Mal d 1.03general F	CTGACAACCTCATCCCCAAGA	54	380
	Mal d 1.03general R	GTGTGGTAGTGGCTGGTG <u>GT</u>		
<i>Mal d 1.04^a</i>	1.0404F	TAATTCATTTGCGAGCGGC	54	159
	1.04 generalR	TCAGGAGATGCGATCAATTTG		
<i>Mal d 1.05^a</i>	Mal d 1.0502 SNP96 F	TTGATGGTGATAACCTCATCC <u>IG</u>	52	358
	Mal d 1.05 general R	TTTGGTCTCATAAGCGATCTTCTC		
<i>Mal d 1.06A^a</i>	1.06A03 SNP37F	CGAAACCGAATACGCCTAAA	54	387
	1.06A generalR	GCATCCCCTTCAATCAAGCTATAGC		
<i>Mal d 1.06B^a</i>	1.06B01 SNP229F	ATTTTCTCCATTAACCTGGTAATCTTT	54	183
	1.06B generalR	CTTTGTCAATTCCTCAACCTT		
<i>Mal d 1.06C^a</i>	1.06C generalF	TCAACTATTTTCTCCATTAACCTTGCT	54	249
	1.06C03 SNP417R	CCAACCTAATCTCATAAGAGATCTCCTG		
<i>Mal d 1.07</i>	Mal d 1.07general F	CAACTTTGTGTACCAATACAGT <u>CTC</u>	51	138
	Mal d 1.07general R	C <u>ITTGGT</u> ATGGTAGTGG <u>G</u> TGA		
<i>Mal d 1.08</i>	Mal d 1.08general F	CGACTGCTCTTGATGGTGATG	51	343
	Mal d 1.08general R	CTAGCCTTGAGATGCTCTTCA <u>GTA</u>		
<i>Mal d 1.09</i>	Mal d 1.09general F	GAGCTGGAACCATTATGAAGAT <u>IAG</u>	54	200
	Mal d 1.09general R	GCGCTTGATGATCGAAC <u>IGT</u>		

*General primer pair for all *Mal d 1.03* isoallergens

^aIsoallergens 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 patten 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).

<i>Mal d 1</i> isoforms on apple genome												
BAC clone	LG 13		LG 16							LG 6	Group	
	1.01	1.03A-F	1.02	1.06A	1.06B	1.06C	1.07	1.08	1.09	1.05		
MC-1			X	X								I
MC-12			X	X	X	X						
MC-14					X	X						
MC-16							X	X	X			II
MC-20							X	X	X			
MC-15		X										?
MC-6		X										III
MC-7		X										
MC-9		X										
MC-11		X										
MC-18		X										
MC-19		X										
MC-8		X										IV
MC-13		X										
MC-10	X	X		X								
MC-17	X											
MC-2		X								X		V
MC-3										X		
MC-4										X		VI
MC-5										X		

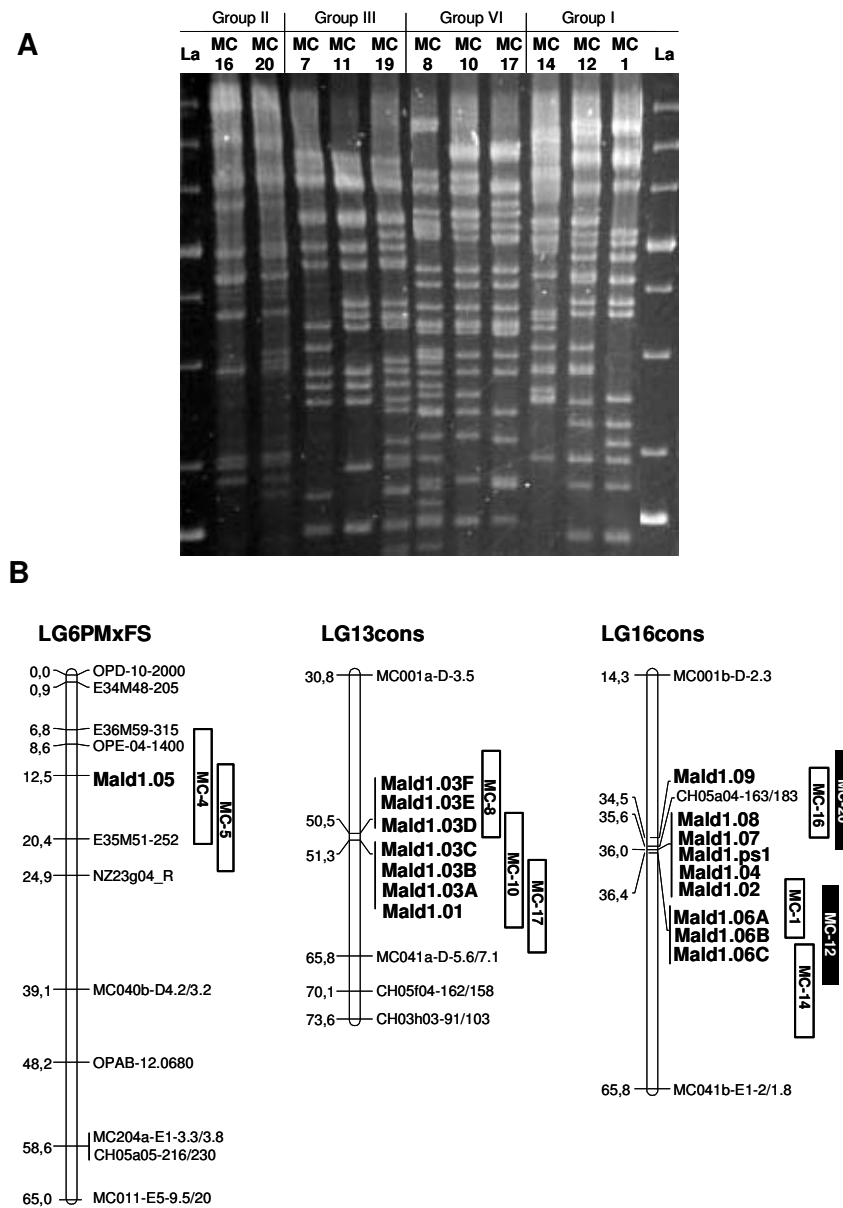


Figure 1: A) Profile of 11 EcoRI-digested BAC clones, derived from group I to IV. Each group contains overlapping clones. B) Genetic map position of apple *Mal d 1* isoallergens (Gao et al, 2005) and BAC clones related to each LG. BAC clones are indicated with boxes. Box sizes are not in scale with BAC length.

BAC clones sequencing

For sequencing, the cluster on LG 16 was preferred to the cluster on the homeologous 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* digeston results, whereas C-20 has only around ~70 kb (Figure 2).

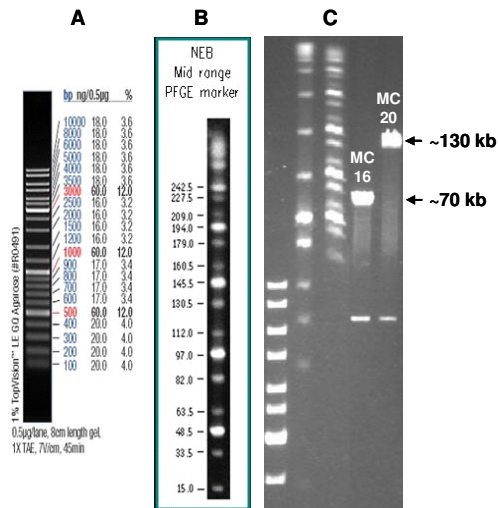


Figure 2: A) 10kb marker. B) Mid range marker. C) *NotI* digestion of BAC clones of group 2. In detail, in the first line the 10 kb marker, in the second lane the low range marker, in the third lane the mid range marker, in the fourth line the MC-16 and in the fifth lane MC-20.

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	AGCAGAAATGCTTCGTCGTT
	C2	2-MC12-C2Rev	AGGGGAGGTTATGCC AAAAT
		3-MC12-C2For	GCTAGTTCAGGTCGGATTTC
	C3	4-MC12-C3Rev	TCCTGGAATGAAACACCTT
		5-MC12-C3For	AATGCCAAGGCTTTCAAGAT
	C4	6-MC12-C4Rev	GATTGGGATATGACCGTTGG
MC-20	C1	7-MC20-C1For	TAATGTGGGCGATTGGGTAT
	C2	8-MC20-C2Rev	TCCTTTGAGTTTTCCACCTGTT
		9-MC20-C2For	CAGAACATGTTTGC GTTGAA
	C3	10-MC20-C3Rev	CCGAAATCCACAATCAAA
		11-MC20-C3For	CTGTGAGGGCTTTGGATAGG
	C4	12-MC20-C4For	CATCGGTTTGCTACTGCTCA

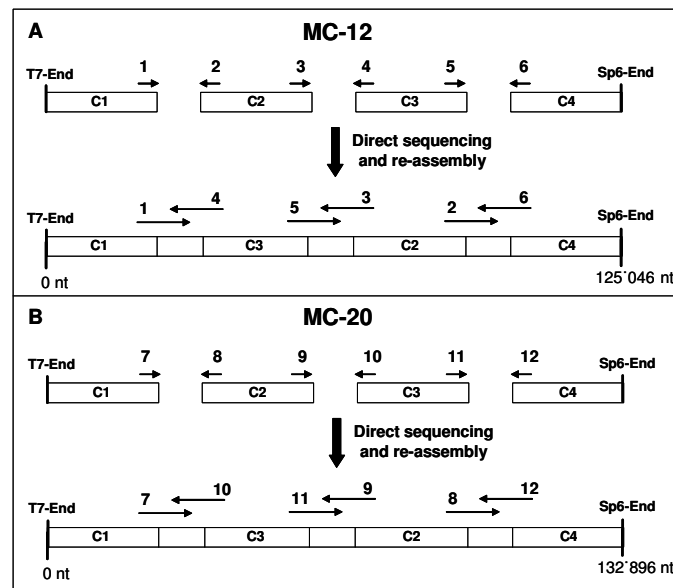


Figure 3: Outline of steps occurred to close the gaps between contigs of clone MC-12 (A) and MC-20 (B). Box sizes are not in scale with BAC length.

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.

ORF	Direction and position on BAC clone				BLASTP results				
	Direction ¹	Start position ²	Lenght (nt) ³	Lenght (aa)	Protein description as reported by the software	Accession number	E-value	Source	Conserved domains
ORF1	↓	519	1351	446	Reverse transcriptase (RNA-dependent DNA polymerase) domain containing protein	AAT85304	4,00E-92	<i>Oryza sativa</i> Japonica Group	RVE superfamily
ORF2	↓	2837	1105	222	No significant similarity				
ORF3	↓	5457	6506	2162	Retrotransposon protein	ABM55240	0.00	<i>Beta vulgaris</i>	Reverse transcriptase (RNA-dependent DNA polymerase); putative NTP binding site; RNase H; RNA/DNA hybrid binding site; integrase core domain; CHRromatic Organization MODifier domain (CHROMO domain)
ORF4	↑	19700	602	159	Mal d 1.06A02	AAX18298	1,00E-84	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF5	↓	22252	701	67	No significant similarity				
ORF6	↑	24478	1022	212	No significant similarity				
ORF7	↓	29652	231	76	No significant similarity				
ORF8	↓	30449	961	161	Mal d 1-like	AAS00052	1,00E-86	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF9	↑	33317	4098	600	Retrotransposon protein, putative, Ty1-copia subclass	ABF96902	4,00E-48	<i>Oryza sativa</i> japonica group	RVE superfamily; RVT_2 superfamily
ORF10	↑	39968	2791	89	No significant similarity				
ORF11	↓	45546	651	159	Mal d 1 allergen	AAD26548	2,00E-85	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF12	↑	47853	3177	68	No significant similarity				
ORF13	↑	51932	663	123	Hypothetical protein	CAN78839	2,00E-29	<i>Vitis vinifera</i>	
ORF14	↑	53816	507	168	Copia-type polyprotein	AAG51247	1,00E-39	<i>Arabidopsis thaliana</i>	
ORF15	↓	59735	861	146	Mal d 1-like	AAS00052	2,00E-38	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF16	↑	63931	1881	132	No significant similarity				
ORF17	↑	66452	428	75	No significant similarity				
ORF18	↓	68017	702	163	Mal d 1-like	AAS00053	6,00E-92	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF19	↑	70566	415	104	No significant similarity				
ORF20	↑	71477	1540	96	No significant similarity				
ORF21	↓	73770	633	159	Mal d 1.06B02	AAX20971	2,00E-84	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF22	↑	76305	812	119	No significant similarity				
ORF23	↓	78970	283	97	Mal d 1.06C04 *	AAX20989	2,00E-42	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF24	↑	79408	4443	1390	Hypothetical protein	XP_002280852	0.00	<i>Vitis vinifera</i>	RVE superfamily; RVT_2 superfamily
ORF25	↓	87785	2574	338	Retrotransposon protein, putative	ABA95230	1,00E-51	<i>Oryza sativa</i> japonica group	RVT_2 superfamily
ORF26	↓	92218	608	159	Mal d 1.06C	AAX18306	5,00E-85	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF27	↓	95324	470	120	Mal d 1.06C*	AAX18306	3,00E-47	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily
ORF28	↓	96401	397	101	No significant similarity				

ORF29	↓	97358	9486	576	Transposon protein, putative	ABA94905	2,00E-17	<i>Oryza sativa japonica</i> group	Transposase domain
ORF30	↓	107669	2433	261	Putative non-LTR retroelement reverse transcriptase	AAD21778	6,00E-35	<i>Arabidopsis thaliana</i>	RT_like superfamily, NTP binding site
ORF31	↓	110916	5126	443	Hypothetical protein	CAN79553	4,00E-71	<i>Vitis vinifera</i>	HMA, Heavy-metal-associated domain
ORF32	↑	117249	8154	868	Hypothetical protein	XP_002273305	6,00E-76	<i>Vitis vinifera</i>	Glycosyl hydrolases family 28
					Hypothetical protein	XP_002271327	0.00	<i>Vitis vinifera</i>	WD40 domain; NUC189 domain

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 nucleotidic 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.

ORF	Direction and position on BAC clone				BLASTP results				
	Direction ¹	Start position ²	Lenght (nt) ³	Lenght (aa)	Protein description	Accession number	E-value	Source	Conserved domains
ORF33	↓	6	1977	622	RNA-directed DNA polymerase	ABD28426	4,00E-179	<i>Medicago truncatula</i>	
ORF34	↑	2241	3000	404	No significant similarities				
ORF35	↓	5408	480	163	Putative reverse transcriptase	AAT40500	7,00E-36	<i>Solanum demissum</i>	
ORF36	↓	9630	5803	680	Hypothetical protein	CAN74865	8,00E-39	<i>Vitis vinifera</i>	PMD, Plant mobile domain
ORF37	↑	16117	1110	142	No significant similarity				
ORF38	↑	17717	5550	835	Hypothetical protein		6,00E-38	<i>Vitis vinifera</i>	
ORF39	↓	24713	950	74	No significant similarity				
ORF40	↑	27462	3347	422	Hypothetical protein	XP_002267217	2,00E-61	<i>Vitis vinifera</i>	RVT_2 superfamily
ORF41	↓	34860	6920	1526	Retrotransposon protein	ABM55240	0.00	<i>Beta vulgaris</i>	Retrotransposon gag protein; RVP_2, Retroviral aspartyl protease; RT_LTR; RnaseH;
ORF42	↑	43949	480	159	Mal d 1.03G	AAX18324	7,00E-85	<i>Malus domestica</i>	Pathogenesis-related protein Bet v I family
ORF43	↓	61000	480	159	Mal d 1.07	AAX18307	5,00E-86	<i>Malus domestica</i>	Pathogenesis-related protein Bet v I family
ORF44	↓	63412	480	159	Mal d 1.0903	AAX20996	5,00E-87	<i>Malus domestica</i>	Pathogenesis-related protein Bet v I family
ORF45	↑	66259	6528	349	No significant similarity				
ORF46	↓	74350	480	159	Mal d 1.08	AAX18309	6,00E-87	<i>Malus domestica</i>	Pathogenesis-related protein Bet v I family
ORF47	↓	76606	9082	460	Hypothetical protein	XP_002271465	5,00E-63	<i>Vitis vinifera</i>	DUF789, Protein of unknown function (DUF789)
ORF48	↑	87704	213	106	No significant similarity				
ORF49	↑	100005	3648	248	HEAT repeat-containing protein	NP_197125	4,00E-36	<i>Arabidopsis thaliana</i>	
ORF50	↓	114365	3129	674	Multi-domain protein (a sensor histidine kinase and a response regulator)	XP_002519174	0.00	<i>Ricinus communis</i>	REC Response_reg, Response regulator receiver domain; myb-like DNA-binding domain, SHAQKYF class
ORF51	↑	118236	675	224	No significant similarity				
ORF52	↑	119188	7853	1050	Protein COBRA precursor, putative	XP_002519175	0.00	<i>Ricinus communis</i>	COBRA superfamily

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

³Lenght of the gDNA nucleotidic sequences

Mal d 1m – Mal d 1.10. The *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 *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 *Mal d 1* amino acid sequence present in the database is *Mal d 1.04* (92% similarity). Following official allergen nomenclature (King et al., 1995) we now propose to denote this gene as *Mal d 1.10*.

Mal d 1n – Mal d 1.11. Similarly, to *Mal d 1.10* also the *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 not 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 *Mal d 1* iso-allergens (Figure 4). Its protein sequence was most similar to *Mal d 1.06A* (67%). We here denote this gene as *Mal d 1.11*.

Mal d 1.03G. ORF42 shows a high homology with *Mal d 1.03G01* (AY822733). It has 5 non-synonymous SNPs, corresponding to two substitutions in position 110 (asparagine → serine) and 132 (histidine → 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 *Mal d 1.03G*. This is the first event in which a *Mal d 1.03* like gene is found on LG16 as all other known *Mal d 1.03* like genes are present on LG13.

Mal d 1ps2. *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 *Mal d 1ps2.02*, the Florina sequence can thus be classified as a new allele (variant 03)

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 *Prunus dulcis x Prunus persica* (85% of identities). *Mal d 1.04* was the most similar *Malus* sequence (81%). Its predicted protein contains 146 AAs, hence it is significantly shorter than any other *Mal d 1* protein. It can be considered as a new *Mal d 1* family member which we propose to name *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 *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 *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.

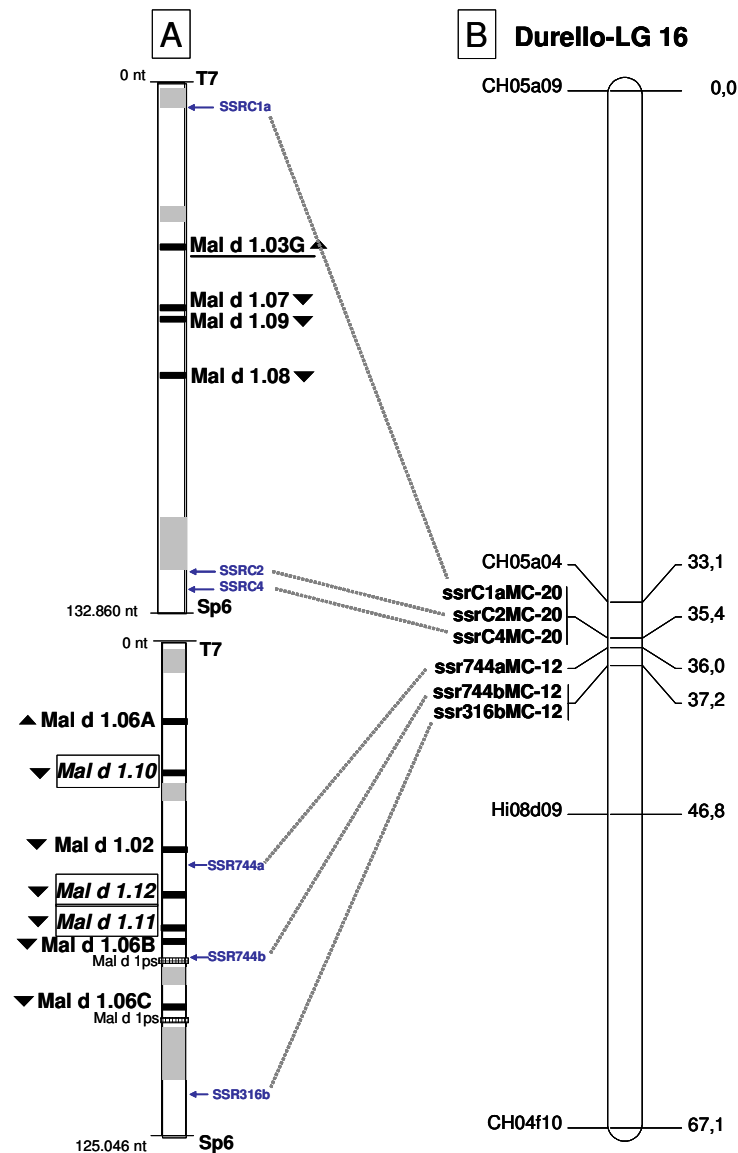


Figure 4: Genomic organization of Mal d 1 genes on LG16. A) Physical map of two BACs. The new isoallergens are indicated in boxes. The isoallergens mapped for the first time is underlined. Transposable elements are represented as gray boxes. B) Genetical map of LG 16 in Durello di Forlì. SSRs developed 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 O2 for the gene *Mal d 1.06A*, it is candidate to be an intermedied 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 O1 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 O2 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 O3). **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.

Mal d 1 genes features on BAC clone						BLASTN results				
ORF	Mal d 1 gene/allele	Direction	Full-length (nt)	CDS (nt)	Intron/Exon	Accession number	Description	SNPs in CDS	SNPs in intron	E-value
ORF4	<i>Mal d 1.06A.02</i>	↑	602	480	Exon I:184; Intron:122; Exon II: 296	AY827697	Malus x domestica clone 220903B10 Mal d 1.06A02 (Mal d 1.06A) gene	1	0	0.00
ORF8	<i>Mal d 1.10^a</i>	↓	961	486	Exon I:184; Intron:475; Exon II: 302	AY428588	Malus x domestica clone 1m Mal d 1-like mRNA	0	not available	1,00E-156
ORF11	<i>Mal d 1.02.01</i>	↓	651	480	Exon I:184; Intron:171; Exon II: 296	AY827654	Malus x domestica clone 250803B10 Mal d 1.0201 (Mal d 1.02) gene	1	0	0.00
ORF15	<i>Mal d 1.12^a</i>	↓	861	486	Exon I:184; Intron:375; Exon II: 302	AY822733	Malus x domestica clone 231103G3 Mal d 1.03G01 (Mal d 1.03G) gene	-	-	3,00E-25
ORF18	<i>Mal d 1.11^a</i>	↓	702	495	Exon I:184; Intron:208; Exon II: 310	AY428589	Malus x domestica clone 1n Mal d 1-like mRNA	0	not available	4,00E-160
ORF21	<i>Mal d 1.06B.02</i>	↓	633	480	Exon I:184; Intron:153; Exon II:296	AY827712	Malus x domestica clone 220903F10 Mal d 1.06B02 (Mal d 1.06B) gene	0	0	0.00
ORF23	<i>Mal d 1ps2.03</i>	↓	283	-	with stop codons and truncated	AY827730	Malus x domestica clone 220903A7 Mal d 1ps2 pseudogene	19	-	2,00E-113
ORF26	<i>Mal d 1.06C.02</i>	↓	608	480	Exon I:184; Intron:128; Exon II:296	AY827725	Malus x domestica clone 231103F11 Mal d 1.06C02 (Mal d 1.06C) gene	1	0	0.00
ORF27	<i>Mal d 1ps2.04</i>	↓	470	-	with stop codons	AY827730	Malus x domestica clone 220903A7 Mal d 1ps2 pseudogene	5	-	0.00

^aAllergen gene names proposed following the official allergen nomenclature

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

Mal d 1 genes features on BAC clone					BLASTN results			
ORF	Mal d 1 gene/allele	Direction	Full-length (nt)	Intron/Exon	Accession number	Description	SNPs in CDS	E-value
ORF42	<i>Mal d 1.03G02</i>	↑	480	Intronless	AY822733	Malus x domestica clone 231103G3 Mal d 1.03G01 (Mal d 1.03G) gene	5	0.00
ORF43	<i>Mal d 1.0703</i>	↓	480	Intronless	AY822717	Malus x domestica clone 231103B3 Mal d 1.0701 (Mal d 1.07) gene	5	0.00
ORF44	<i>Mal d 1.0903</i>	↓	480	Intronless	AY822721	Malus x domestica clone 231103C1 Mal d 1.0903 (Mal d 1.09) gene	0	0.00
ORF46	<i>Mal d 1.0801</i>	↓	480	Intronless	AY822719	Malus x domestica clone 231103F3 Mal d 1.0801 (Mal d 1.08) gene	0	0.00

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 $\pm 1,2$ 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.

BAC clone	Primer name	Primer sequence (5'-3')	Position (nt) ¹	Ta (°C)	Segregation in DuXFi opulation ²
MC-12	MC-12SSR384For	CCTAAATCCTCCCTCCCCAAAAC	10.083-10.248	60	No
	MC-12SSR384Rev	AGATCGGACGTTGGTGTAGG			
	MC-12SSRC744AFor	TCAACATCCAAATCCCACAA	49.406-49.612	57	Yes
	MC-12SSRC744ARev	GGTGTCTTTGAGCCTCCTG			
	MC-12SSRC744BFor	CCAAATCCACCCCTAGTTTG	78.221-78.418	57	Yes
	MC-12SSRC744BRev	CATACGTTCTCCCACCGACT			
	MC-12SSRC744CFor	CCTGACACAACCCGATAACC	102.343-102.529	60	No
	MC-12SSRC744CRev	TCCAATGGTCACCAATTTTT			
	MC-12SSR316For	CGGTGATTCTTTGGGGATA	115.950-116.110	60	Yes
	MC-12SSR316Rev	CATTACAACGAGGAGCAGCA			
MC-20	MC-20SSRC1aFor	TTTGCAAAGGATGGATTGACT	5.841-6.048	60	Yes
	MC-20SSRC1aRev	GCAATGGCGTTCTAGGATTC			
	MC-20SSRC1For	AATCGGAGTTGAATCGGTTG	14.888-15.154	60	No
	MC-20SSRC1Rev	CaGTTGTGAGCTTCGAAGAATG			
	MC-20SSRC3For	AGGGTGAGGATTGGATGTTG	40.986-41.169	60	No
	MC-20SSRC3Rev	TCGTTCTCGGGATAGGTGTC			
	MC-20SSRC2For	AATCCATGTTGGGAGACAGG	114.473-114.671	60	Yes
	MC-20SSRC2Rev	TGTGGTCGACGATGATCCTA			
MC-20SSRC4For	GTGCGACTGTTTCTGTGTGC	127.180-127.365	60	Yes	
MC-20SSRC4Rev	TCATTCAAAGACGGCAACAA				

¹Positions count from the T7-End of Florina BAC clones sequences

²All the SSR resulted polymorphic in Durello di Forlì

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 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 in 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 continuous stretch (E₄₂-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 surface. 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.

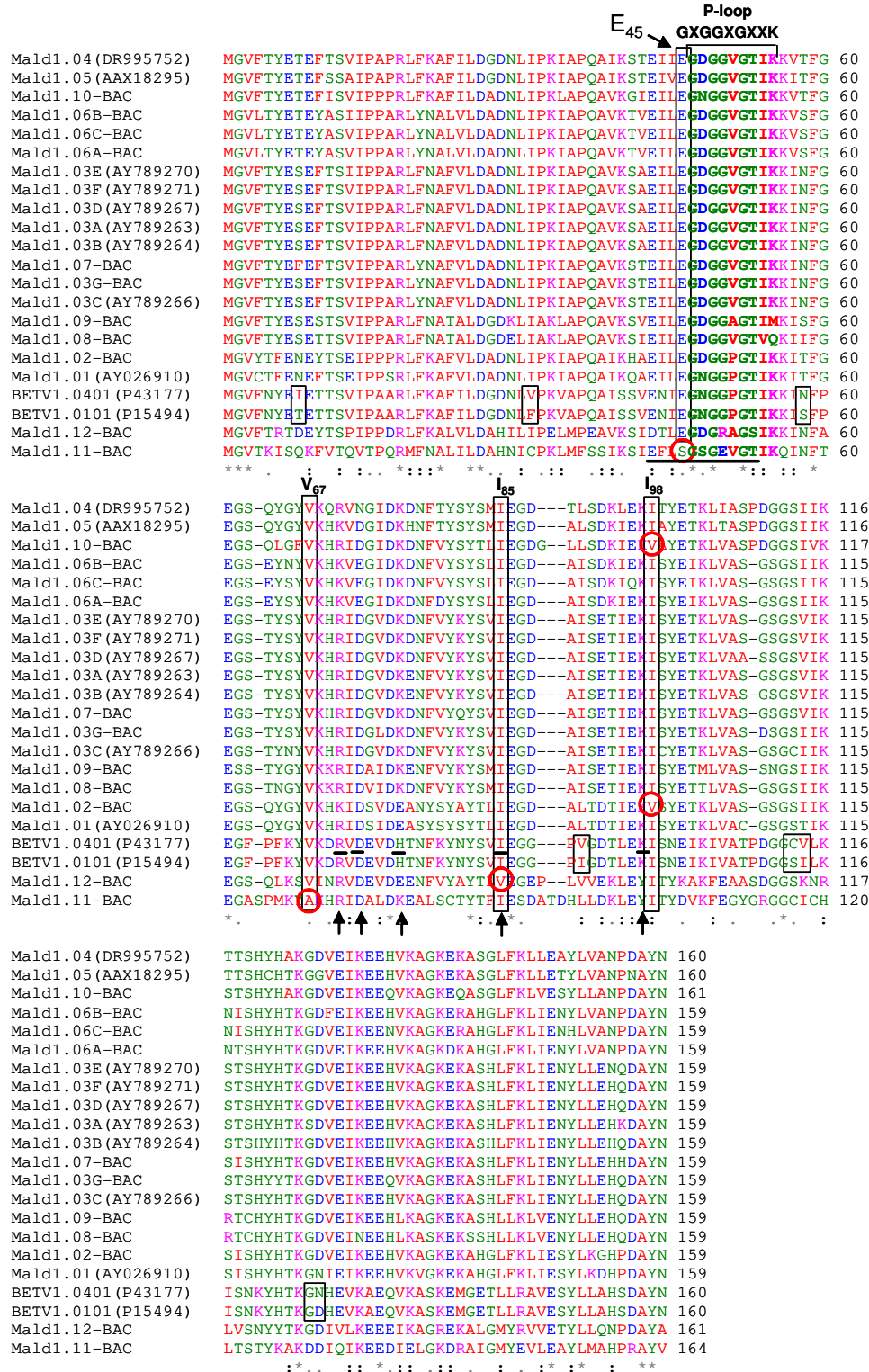


Figure 5: Alignment of predicted amino acid sequences of Mal d 1 isoforms. For the isoforms retrieved from the database the accession number is reported. The isoforms derived from the BAC library sequences are indicated with -BAC. Also two Bet v 1 isoforms were included: Bet v 1.01 and Bet v 1.04 that is considered an hypoallergenic variant. Hydrophobic residues crucial for the binding activity are indicated by 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 important sequences and amino acid putatively important for the IgE recognition are indicated with arrows.

In this isoallergen the E₄₅→S₄₅ substitution is reported. In the work of Spangfort et al. (2003) they performed the same artificial mutation (E₄₅→S₄₅) 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 date 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 V₆₇, I₈₅, and I₉₈ (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 (V₆₇→A₆₇), Mal d 1.12 for the position 85 (I₈₅→V₈₅) and Mal d 1.02 and Mal d 1.10 for the position 98 (I₉₈→V₉₈).

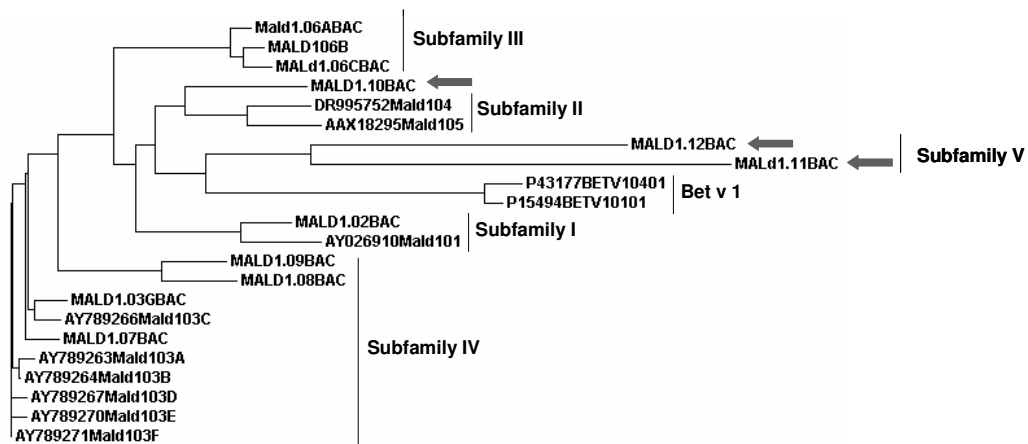


Figure 6: Phylogram tree derived from the alignment reported in Figure 5. The gray arrows indicated the new isoforms for LG16.

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, Chromatin 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 these 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 polyploidization 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 contiguous 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 isoallergens 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

- Andrade MA, Bork P. 1995. HEAT repeats in the Huntington's disease protein. *Nat Genet.* 11:115-6.
- Barth O, Zschiesche W, Siersleben S, Humbeck K. 2004. Isolation of a novel barley cDNA encoding a nuclear protein involved in stress response and leaf senescence. *Physiol Plant.* 121:282-293.
- Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA. 2004. Characterisation of Mal d 1-related genes in *Malus*. *Plant Mol Biol* 55:369-388.
- Birnboim, H.C. and J. Doly. 1979. A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. *Nucleic Acids Research* 7:1513-1523.
- Botton A, lezzer p, Dorigoni A, Barcaccia G, Ruperti B, Ramina A. 2008. genetic and environmental factors affecting allergen-related gene expression in apple fruit (*Malus domestica* L. Borkh) *Agr Food Chem* 56: 6707-6716.
- Breiteneder, H. 2004. Thaumatin-like proteins - a new family of pollen and fruit allergens. *Allergy.* 59, 479-481.
- Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78-94.
- Celton, J.-M., D.S. Tustin, D. Chagné, S.E. Gardiner. 2009. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. *Tree Genet. Genomes* 5:93–107.
- Chagné, D., K. Gasic, R. Crowhurst, Y. Han, H.C. Bassett, D.R. Bowatte, T.J. Lawrence, E. Rikkerink, S. Gardiner, S.S. Korban. 2008. Development of a set of SNP markers in expressed genes of the apple. *Genomics* 92:353–358.
- Chen L., Zhang S., Illa E., Song L., Wu S., Howad W., Arús P., van de Weg E., Chen K., Gao Z. 2008. Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genomics*, 17, 9:543.
- Chevreau E, Lespinasse M, Gallet M. 1985. Inheritance of pollen enzymes and polyploid origin of apple (*Malus x domestica* Borkh.). *Theor. Appl. Genet.* 71: 268-277.
- Chung KM, Igari K, Uchida N, Tasaka M. 2008 New perspectives on plants defense responses through modulation of developmental pathways. *Mol Cells* 26: 107–112
- Cova V. 2008. Isolamento di marcatori microsatelliti strettamente associati al gene di resistenza a ticchiolatura Vm in melo (PhD Thesis, Università di Bologna 2008, 114 pp).
- Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134:221-34.
- Fernandes H, Pasternak O, Bujacz G, Bujacz A, Sikorski MM, Jaskolski M. 2008. *Lupinus luteus* pathogenesis-related protein as a reservoir for cytokinin. *J Mol Biol* 378: 1040–1051.
- Fernandez-Rivas, M., Bolhaar, S, Gonzalez-Mancebo, E., Asero, R., van Leeuwen, A., Bohle, B., Ma, Y., Ebner, C., Rigby, N., Sancho, A. I., Miles, S., Zuidmeer, L., Knulst, A., Breiteneder, H., Mills, C., Hoffmann-Sommergruber, K. van Ree, R. 2006. Apple allergy across Europe: how allergen sensitization profiles determine the clinical expression of allergies to plant foods. *Journal of Allergy and Clinical Immunology.* 118, 481-488.
- Gao Z, van de Weg EW, Matos CI, Arens P, Bolhaar ST, Knulst AC, Li Y, Hoffmann-Sommergruber K, Gilissen LJ. 2008. Assessment of allelic diversity in intron-containing *Mal d 1* genes and their association to apple allergenicity. *BMC Plant Biol.* 13:8:116.
- Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ. 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet.* 111:171-83.
- Gasic K., Gonzalez D.O., Thimmapuram J., Liu L., Malnoy M., Gong G., Han Y., Vodkin L.O., Aldwinckle H.S., Carroll N.J., Orvis K.S., Goldsbrough P., S., Pape D., Fulton L., Martin J., Theising B., Wisniewski

- M.E., Fazio G., Feltus F.A., Korban S.S. 2009. Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. *The plant genome*, 2, (1): 23-38.
- Ghosh D, Gupta-Bhattacharya S. 2007. Structural insight into protein T1, the non-allergenic member of the Bet v 1 allergen family-An in silico analysis. *Mol Immunol*. 45:456-62.
- Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C. 1998. Simple sequence repeats for the genetic analysis of apple. *Theor Appl Genet* 96:1069–1076.
- Han, Y., K. Gasic, B. Marron, J.E. Beever, S.S. Korban. 2007. A BAC based physical map of the apple genome. *Genomics* 89:630–637.
- Helsper JPDFG, Gilissen LJWJ, van Ree R, America Antoine HP, Cordewener Jan HG, Bosch D. 2002. Quadrupole time-of-flight mass spectrometry: A method to study the actual expression of allergen isoforms identified by PCR cloning. *J Allergy Clin Immunol* 110:131-138.
- Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, Gibson T. J. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23: 403–405.
- King TP, Hoffman D, Lowenstein H, Marsh David G, Platts Mills Thomas AE, Thomas W. 1995. Allergen nomenclature. *J Allergy Clin Immunol* 96:5–14.
- Koistinen KM, Soininen P, Vena"la"inen TA, Ha"yrinen J, Laatikainen R, Pera"ky"la" M, Tervahauta AI, Ka"renlampi SO. 2005. Birch PR-10c interacts with several biologically important ligands. *Phytochemistry* 66:2524–2533
- Liebhart R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C 2002 Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Mol Breed* 10: 217–241.
- Liebhart R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C 2003a Mapping quantitative physiological traits in apple (*Malus x domestica* Borkh.). *Plant Mol Biol* 52: 511–526.
- Liebhart R, Koller B, Patocchi A, Kellerhals M, Pfammatter W, Jermini M, Gessler C 2003b Mapping quantitative field resistance against apple scab in a 'Fiesta' x 'Discovery' progeny. *Phytopathology* 93: 493–501.
- Liebhart R, Koller B, Gianfranceschi L, Gessler C. 2003c. Creating a saturated reference map for the apple (*Malus x domestica* Borkh.) genome. *Theor Appl Genet* 106: 1497–1508.
- Liu JJ, Ekramodoullah AK. 2006. The family 10 of plant pathogenesisrelated proteins: their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol Mol Plant Pathol* 68: 3–13
- Ma J, Wing RA, Bennetzen JL, Jackson SA. 2007. Evolutionary history and positional shift of a rice centromere. *Genetics*.177:1217-20.
- Ma Y, Gadermaier G, Bohle B, Bolhaar S, Knulst A, Markovic-Housley Z, Breiteneder H, Briza P, Hoffmann-Sommergruber K, Ferreira F. 2006. Mutational analysis of amino acid positions crucial for IgE binding epitopes of the major apple (*Malus domestica*) allergen, Mal d 1. *Int Arch Allergy* 139:53-62.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic Arabidopsis polyploids. *Plant J.* 41::221-30.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A.* 102:5454-9.
- Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, et al. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor Appl Genet* 97: 60–73.
- Marković-Housley Z, Degano M, Lamba D, von Roepenack-Lahaye E, Clemens S, Susani M, Ferreira F, Scheiner O, Breiteneder H. 2003. Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *J Mol Biol.* 325:123-33.
- Naik S, Hampson C, Gasic K, Bakkeren G, Korban SS. 2006 Development and linkage mapping of E-STS and RGA markers for functional gene homologues in apple. *Genome* 49: 959–968.
- Ooijen J van, Voorrips R. 2001. Joinmap® Version 3.0: Powerful genetic linkage mapping made easy. Plant Research International, Wageningen, The Netherlands.

- Puehringer H, Zinoecker I, Marzban G, Katinger H, Laimer M. 2003. MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene* 321:173-183.
- Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, Ushijima K, Kusaba M, Hirano H, Koba T. 2007. S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. *Genetics* 175:1869-81.
- Schöll I, Kalkura N, Shedziankova Y, Bergmann A, Verdino P, Knittelfelder R, Kopp T, Hantusch B, Betzel C, Dierks K, Scheiner O, Boltz-Nitulescu G, Keller W, Jensen-Jarolim E. 2005. Dimerization of the major birch pollen allergen Bet v 1 is important for its in vivo IgE-cross-linking potential in mice. *J Immunol.* 175:6645-50.
- Spangfort MD, Mirza O, Ipsen H, Van Neerven RJ, Gajhede M, Larsen JN. 2003. Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. *J Immunol.* 171:3084-90.
- Stacey MG, Hicks SN, von Arnim AG. 1999. Discrete domains mediate the light-responsive nuclear and cytoplasmic localization of Arabidopsis COP1. *Plant Cell.* 11:349-64.
- Untergasser A. 2006. Maxiprep – Alkaline Lysis. Untergasser's Lab. Summer 2006 (http://www.untergasser.de/lab/protocols/maxiprep_alkaline_lysis_v1_0.htm).
- van Loon L.C., Rep M., Pieterse C.M. 2006. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- Vanek-Krebitz M, Hoffmann-Sommergruber K, Laimer da Camara Machado M, Susani M, Ebner C, Kraft D, Scheiner O, Breiteneder H. 1995. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem Biophys Res Commun.* 214:538-51.
- Vinatzer B.A., Zhang H.B., Sansavini S. 1998. Construction and characterization of a bacterial artificial chromosome library of apple. *Theor Appl Genet.*, 97:1183-1190.
- Vinatzer, B.A., A. Patocchi, L. Gianfranceschi, S. Tartarini, H.B. Zhang, C. Gessler, S. Sansavini. 2001. Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with Vf apple scab resistance. *Mol. Plant Microbe Interact.* 14:508–515.
- Wagner S, Radauer C, Bublin M, Hoffmann-Sommergruber K, Kopp T, Greisenegger EK, Vogel L, Vieths S, Scheiner O, Breiteneder H. 2008. Naturally occurring hypoallergenic Bet v 1 isoforms fail to induce IgE responses in individuals with birch pollen allergy. *J Allergy Clin Immunol.* 121:246-52.
- Wasteneys GO, Fujita M. 2006. Establishing and maintaining axial growth: wall mechanical properties and the cytoskeleton. *J Plant Res.* 119:5-10.
- Xu ML, Korban SS. 2002a. AFLP-derived SCARs facilitate construction of a 1.1 Mb sequence-ready map of a region that spans the Vf locus in the apple genome. *Plant Mol Biol* 50: 803–818.
- Xu, M.L., and S.S. Korban. 2002b. A cluster of four receptor-like genes reside in the Vf locus that confers resistance to apple scab disease. *Genetics* 162:1995–2006.
- Xu, M.L., J. Song, Z. Cheng, J. Jiang, S.S. Korban. 2001. A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 for positional cloning of the apple scab resistance gene Vf. *Genome* 44:1104–1113.
- Xu, M.L., S.S. Korban, J. Song, J. Jiang. 2002. Constructing a bacterial artificial chromosome (BAC) library of the apple cultivar GoldRush. *Acta Hortic.* 595:103–111.
- Yamada S, Shiro Y. 2008. Structural basis of the signal transduction in the two-component system. *Adv Exp Med Biol.* 631:22-39.
- Yan Q, Qi X, Jiang Z, Yang S, Lujia H. 2008. Characterization of a pathogenesis-related class 10 protein (PR-10) from *Astragalus mongholicus* with ribonuclease activity. *Plant Physiol Biochem* 46: 93–99
- Zhou S, Sauvé R, Thannhauser TW. 2009. Proteome changes induced by aluminium stress in tomato roots. *J Exp Bot.* 60:1849-57.
- 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 organisms, 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 permits 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 templates. 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 correspondance 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 μ l, containing 5 μ l 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 60C 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 fuorescence 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 μ l volume containing 50 ng of gDNA, optimized concentration of gene-specific primers, 1.5 mM MgCl₂, 100 μ M 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 correspondance 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 pairs 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 DNaseI (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 phisical 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, a 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 (R_2) of the plotted points. From the slope of each standard curve, PCR amplification efficiency (E) was calculated according to the equation: $E = 10^{(-1/\text{slope})} - 1$ (Lee et al., 2006). To test the relative efficiency of each primer pair with *actin*, the ΔC_t (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

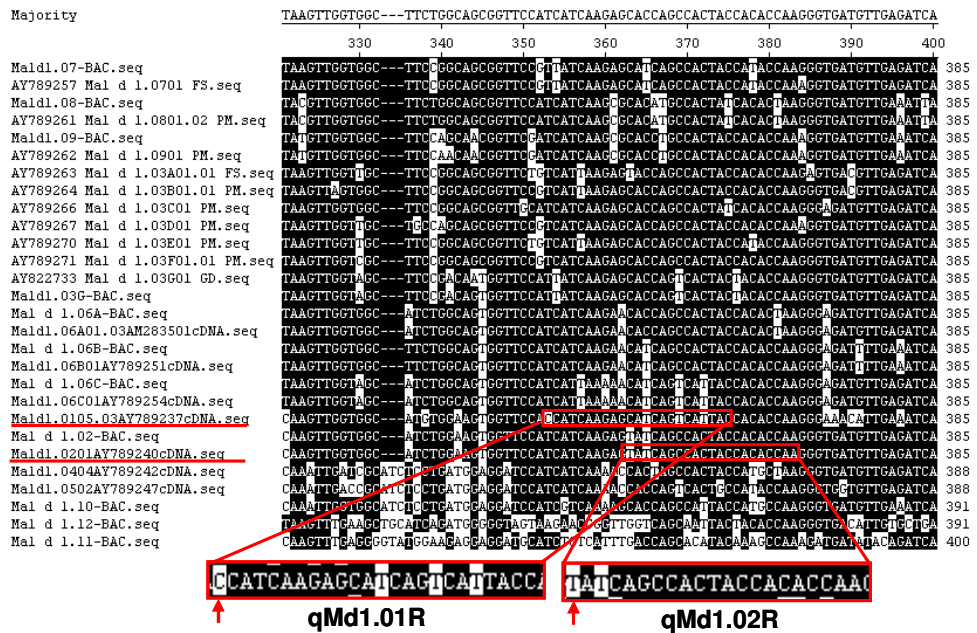
Mal d 1-related sequences were searched by the insertion of key-words or known *Mal d 1* sequences as input. Since the high homology among *Mal d 1*-like sequences, an accurate analysis of the output sequences was performed by Lasergene® v8.0-MegAlign (DNASTAR, Inc., Madison, WI USA) software to distinguish 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 than 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.

A



B

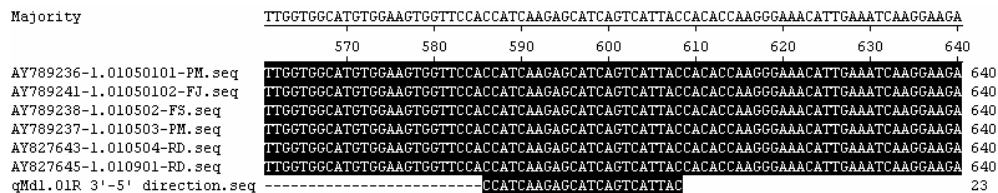


Figure 1: A) Portion of a simplified *Mal d 1* sequences alignment. Only one sequence for each isoallergen and BAC sequences (28 on 53 sequences) were included in this alignment. The reverse primers specific for *Mal d 1.01* and *Mal d 1.02* are showed in red boxes. The SNPs at the 3' end of each primer, 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 isoallergens. Firstly, all the primers were tested with an *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 *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 *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. *Mal d 1.02*) it was sufficient to modify the annealing temperature (Figure 2A); for others (i.e. *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 *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. $T_a=63$ 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

Isoallergens	Primer name	Primer sequence (5'-3')	Start*	Length (bp)	T _a (C°)	Primer conc. (nM)	T _m (C°)
<i>Mal d 1.01</i>	qMd1.01/02F	GATTGAAGGAGATGCTTTGACA	258	103	63	100	76,83
	qMd1.01R	GTAATGACTGATGCTCTTGATGG				100	
<i>Mal d 1.02</i>	qMd1.01/02F	GATTGAAGGAGATGCTTTGACA	258	111	62	100	79,06
	qMd1.02R	TTGGTGTGGTAGTGGCTGATA				100	
<i>Mal d 1.03A</i>	qMd1.03AF	ATCTGAGTTCACCTCCGTCATT	21	96	63	70	78,55
	qMd1.03AR	ACTGCTTGTGGTGAATCTTT				70	
<i>Mal d 1.03B</i>	qMd1.03BF	TGTTTTACATACGAATCCGAA	6	167	63	100	81,45
	qMd1.03BR	TGATCTTCTTAATGGTTCCTACGC				100	
<i>Mal d 1.03C</i>	qMd1.03CF	CTCGAAACAATTGAGAAAATCTG	276	79	63	100	77,44
	qMd1.03CR	GCTGGTGTCTTGATGATGC				100	
<i>Mal d 1.03D</i>	qMd1.03F	ATACGAATCCGAGTTCACCTCT	15	156	62	70	79,5
	qMd1.03DR	ATCTTCTTAATGGTCCAACCTCT				70	
<i>Mal d 1.03E</i>	qMd1.03F	ATACGAATCCGAGTTCACCTCT	15	169	62	70	79,72
	qMd1.03ER	TTCACCGAAGTTGATCTTCTTAATA				70	
<i>Mal d 1.03F</i>	qMd1.03FF	CACAGAATTGACGGGGTG	208	119	63	70	77,29
	qMd1.03FR	CCGGAAGCGACCAACTTA				70	
<i>Mal d 1.03G</i>	qMd1.03GF	ATTATCAAGAGCACCAGTCACTACT	337	122	62	70	78,32
	qMd1.03GR	TCCAAGAGGTAGTCTCAATCAA				70	
<i>Mal d 1.04</i>	qMd1.04F	GGGTATGTTAAGCAAAGGGTCA	196	193	61	100	76,97
	qMd1.04R	TGATCTCAACATCACCTTAGC				100	
<i>Mal d 1.05</i>	qMd1.05F	ATCAAACCACTAGTCACTGCCAT	343	124	63	70	80,32
	qMd1.05R	GGTTGGCCACAAGGTAGGTT				70	
<i>Mal d 1.06A</i>	qMd1.06AF	CTATAGCTATAGCTTGATTGAAGGG	243	167	61	100	76,83
	qMd1.06AR	TTCCAACCTTAACATGTTCTTCT				100	
<i>Mal d 1.06B</i>	qMd1.06BF	AAACCGAATACGCATCCATT	20	106	61	100	79,09
	qMd1.06BR	ACAGTTTTGACTGCTTGTGGAG				100	
<i>Mal d 1.06C</i>	qMald1.06CF	GCTCCACAAGCAGTCAAACT	103	116	63	70	76,74
	qMald1.06CR	TCAACCTTGTGCTTCACATAACTA				70	
<i>Mal d 1.07</i>	qMd1.07For	CAACTTTGTGTACCAGTACAGTGTC	234	126	61	100	77,91
	qMd1.07Rev	TAGTGGCTGATGCTCTTGATAAC				100	
<i>Mal d 1.08</i>	qMd1.08F	TCTTCGGTGAAGGTAGCACAA	173	200	61	100	78,72
	qMd1.08R	ACCCTTAGTGTGGTAGTGGCAT				100	
<i>Mal d 1.09</i>	qMd1.09For	TTTTACATACGAATCCGAGTC	8	126	61	100	81,72
	qMd1.09Rev	GGATCTCAACGCTCTTCACA				100	
<i>Mal d 1.10</i>	qMald1.10F	CAAGGCTTTCATCCACGAC	60	157	61	100	79,81
	qMald1.10R	GATTCTGTGCTTTACAAACCCT				100	
<i>Mal d 1.11</i>	qMald1.11F	GGAGGATGCATCTGTCAATTTG	343	130	62	100	76,38
	qMald1.11R	CCATGAGATAGGCTTCCAAAACCT				100	
<i>Mal d 1.12</i>	qMd1.12F	CAAGGGTGACATTGTGCTG	372	110	61	100	78,02
	qMd1.12R	CATAGGCATCAGGATTTTTGG				100	

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

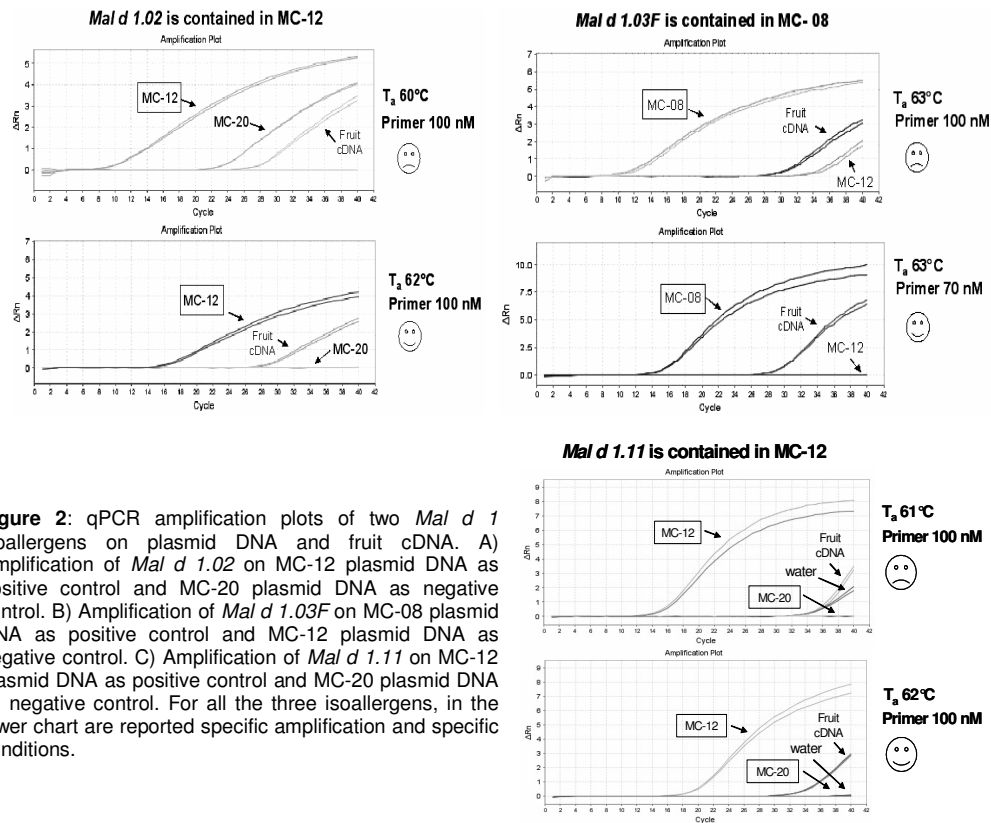


Figure 2: qPCR amplification plots of two *Mal d 1* isoallergens on plasmid DNA and fruit cDNA. A) Amplification of *Mal d 1.02* on MC-12 plasmid DNA as positive control and MC-20 plasmid DNA as negative control. B) Amplification of *Mal d 1.03F* on MC-08 plasmid DNA as positive control and MC-12 plasmid 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 (T_m) 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.

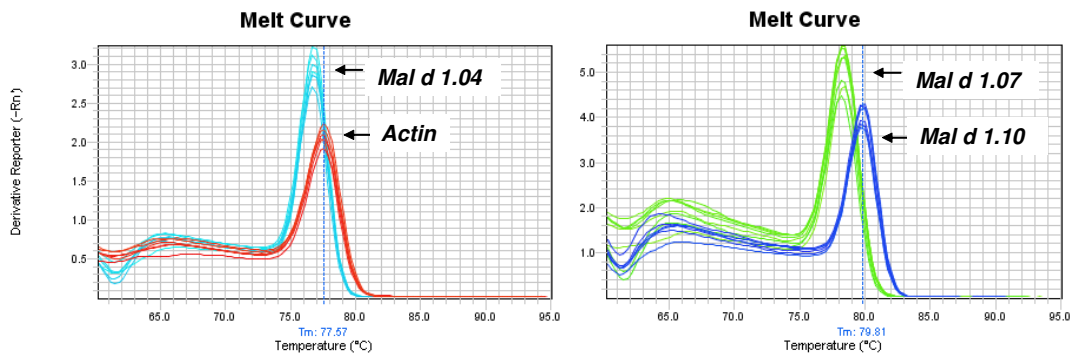


Figure 3: Examples of melting curves obtained in two different qPCR plates at the optimized conditions. In both plots, each primer pair showed a characteristic and single melting curve peak. *Actin* represents the reference gene.

Finally, the specificity of the 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.

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

* 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; AY822722, AY822724, AY822723, AY822726, AY822727, AY822733, AY822718, AY822721 and AY822720 derived from the cv Golden Delicious; AY789266, AY789271, AY789258 and AY789262 derived from cv Prima; AY789269 derived from cv Fiesta; AY827698, AY827714 and AY827728 derived from cv Ingrid Marie; AY827716 derived from the cv Cox; AY827710 derived from cv Jonathan; AY428588 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.

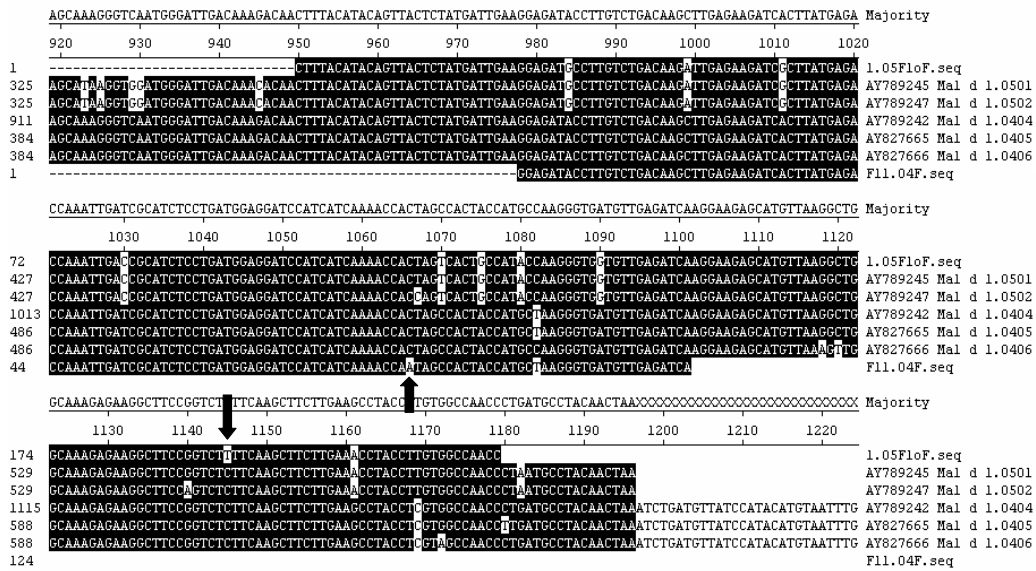


Figure 4: Alignment among *Mal d 1.04* alleles (AY789242, AY827665, AY827666), *Mal d 1.05* alleles (AY789245, AY789247) and *Mal d 1.04* and *Mal d 1.05* sequences (named Flo Md1.04F and Flo Md 1.05F) derived from the amplicons. SNPs between amplicons and isoallergens sequences are indicated with arrows.

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 exaples of end-point PCRs.

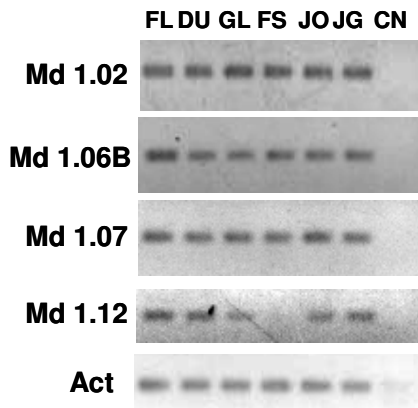


Figure 5: End-point amplifications of *Mal d 1.02*, *-1.06B*, *-1.07*, *-1.12* and *actin* genes on the genotypes: Florina (FL), Durello di Forlì (DU), Gala (GL), Fiesta (FS), Jonathan (JO) and Jonagold (JG). The negative controls are indicated with CN.

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 methods: 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 permits 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 choice 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*.

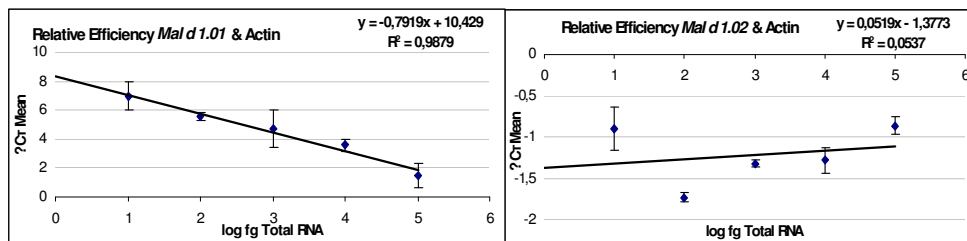


Figure 6: Plots of relative efficiency (log input amount versus ΔC_T) for *Mal d 1.01* (left plot) and *Mal d 1.02* (right plot). The *actin* was the reference genes.

The standard curve method requires the addition of the standard samples in all the plates and for all the tested genes, including 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.

<i>Mal d 1</i> isoallergen	Slope	<i>Mal d 1</i> isoallergen	Slope
<i>Mal d 1.01</i>	-0,791	<i>Mal d 1.06A</i>	-0,8705
<i>Mal d 1.02</i>	0,0519	<i>Mal d 1.06B</i>	-0,8616
<i>Mal d 1.03A</i>	-0,1378	<i>Mal d 1.06C</i>	-0,6125
<i>Mal d 1.03C</i>	-0,6816	<i>Mal d 1.07</i>	-1,0369
<i>Mal d 1.03D</i>	-1,0113	<i>Mal d 1.08</i>	-0,2212
<i>Mal d 1.03E</i>	-1,3793	<i>Mal d 1.09</i>	-1,691
<i>Mal d 1.03F</i>	0,0799	<i>Mal d 1.10</i>	-0,7929
<i>Mal d 1.03G</i>	-0,4563	<i>Mal d 1.11</i>	0,166
<i>Mal d 1.04</i>	-0,4486	<i>Mal d 1.12</i>	-0,984

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 .

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

*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 GenBank 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 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 1.02* 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.

	Isoallergen	LG	Tissue	EST		Isoallergen	LG	Tissue	EST
Subfamily I	<i>Mal d 1.01</i> (AY789238)	13	fruit	CN995671, CO067772, AT000383, AT000088, AT000238, CO418705, CO417908		<i>Mal d 1.03A</i> (AY789263)	13		
			flower	CN493628, CO067265		<i>Mal d 1.03B</i> (AY789265)	13		
	<i>Mal d 1.02</i> (AY789240)	16	fruit	AT000305, CO722244, CO417338, CO751748, CO755402		<i>Mal d 1.03C</i> (AY789266)	13	fruit	CK900634, CO418155
			flower	CN490990, CN492915				leaf	EG974766, EB121653, CO960908
bud			CO865939, CO723162, CO866496	cell culture				EB153482	
Subfamily II	<i>Mal d 1.04</i> (AY789242)	16	leaves challenged with apple scab	DR995752, DR993339		<i>Mal d 1.03D</i> (AY789267)	13	root	DT042809
	<i>Mal d 1.05</i> (AY789245)	6					<i>Mal d 1.03E</i> (AY789270)	13	flower
Subfamily III	<i>Mal d 1.06A</i> (AY789248)	16	fruit	AT000364, CK900679, CV092112, CV128420, DY439734, CV466473		<i>Mal d 1.03F</i> (AY789271)			13
			bud	CO865485			flower	CN497094	
			leaf	CV524623			fruit	CO418484, CO899602, CO902128	
			stem	CV997946			leaf	EB112057, EB149683	
	<i>Mal d 1.06B</i> (AY789252)	16	fruit	CO866217, EB115143, EG631380, EB115278, EB125175		<i>Mal d 1.03G</i> (AY789274)	13	root	EB151349, EB151491, EB151562, EB151739, EB151851, EB151780, EB152174, EB152146, EB152817
			leaf	EB121962, EG974779				flower	CO068711
			root	EB151385, EB151964, EB152831				root	DT003343
<i>Mal d 1.06C</i> (AY789255)	16				<i>Mal d 1.07</i> (AY789257)	16	cell culture	EB155976, EB155679, EB156359	
No subfamily	<i>Mal d 1.10</i> (AY428588)	16	leaf	CN873661, EB123144				<i>Mal d 1.08</i> (AY789259)	16
			flower	CN996594, CN995063	bud	CV084268			
			fruit	CO541245, CO541264, CO754636, EB125297, EB125314, EB125582, DY256001	root	EB151988			
			bud	CO866812, CO899926	leaf	EG974799			
	<i>Mal d 1.11</i> (AY428589)	16	fruit	CN996026, CN994391, CO576341, CO415469, CO756114, CO752639, CV128571, CV466349		<i>Mal d 1.09</i> (AY789262)	16	root	EB152408, EB151974, EB155128
	flower	CO052696	root	EB151590, EB151857, EB152698, EB152168, EB152262, EB152309, EB152690, EB152833					
<i>Mal d 1.12</i> (BAC seq)	16					cell culture	EB155990, EB156054		

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

- Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci U S A*. 100:4649-54.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389-402. Review.
- Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA. 2004. Characterisation of Mal d 1-related genes in *Malus*. *Plant Mol Biol* 55:369-388.
- Bolhaar ST, van de Weg WE, van Ree R, Gonzalez-Mancebo E, Zuidmeer L, Buijnzeel-Koomen CA, Fernandez-Rivas M, Jansen J, Hoffmann-Sommergruber K, Knulst AC, Gilissen LJ. 2006. In vivo assessment with prick-to-prick testing and double-blind, placebo-controlled food challenge of allergenicity of apple cultivars. *J Allergy Clin Immunol*. 117:189.
- Bolhaar ST, Zuidmeer L, Ma Y, Ferreira F, Buijnzeel-Koomen CA, Hoffmann-Sommergruber K, van Ree R, Knulst AC. 2005. A mutant of the major apple allergen, Mal d 1, demonstrating hypo-allergenicity in the target organ by double-blind placebo-controlled food challenge. *Clin Exp Allergy*. 35:1638-44.
- Botton A, Lezzer P, Dorigoni A, Ruperti B, Ramina A. 2009. Environmental factors affecting the expression of apple (*Malus x domestica* L. Borkh) allergen-encoding genes. *JHSB ISAFRUIT Special Issue* 182–187.
- Botton A., Lezzer P., Dorigoni A., Barcaccia G., Ruperti B., Ramina A. 2008. Genetic and environmental factors affecting allergen-related gene expression in apple fruit (*Malus domestica* L. Borkh). *J Agric Food Chem.*, 56, 6707-16.
- Breiteneder H, Radauer C. 2004. A classification of plant food allergens. *J Allergy Clin Immunol*. 113:821-30.
- Fernández Rivas M. 2003. Cross-reactivity between fruit and vegetables. *Allergol Immunopathol (Madr)* 31:141-6. Review.
- Gao Z, van de Weg EW, Matos CI, Arens P, Bolhaar ST, Knulst AC, Li Y, Hoffmann-Sommergruber K, Gilissen LJ. 2008. Assessment of allelic diversity in intron-containing Mal d 1 genes and their association to apple allergenicity. *BMC Plant Biol*. 13:8:116.
- Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ. 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet*. 111:171-83.
- Helsper JPF, Gilissen LJW, van Ree R, America Antoine HP, Cordewener Jan HG, Bosch D. 2002. Quadrupole time-of-flight mass spectrometry: A method to study the actual expression of allergen isoforms identified by PCR cloning. *J Allergy Clin Immunol* 110:131-138.
- Herndl A, Marzban G, Kolarich D, Hahn R, Boscia D, Hemmer W, Maghuly F, Stoyanova E, Katinger H, Laimer M. 2007. Mapping of *Malus domestica* allergens by 2-D electrophoresis and IgE-reactivity. *Electrophoresis*. 28:437-48.
- Hoffmann-Sommergruber K, Mills EN. 2009. Food allergen protein families and their structural characteristics and application in component-resolved diagnosis: new data from the EuroPrevall project. *Anal Bioanal Chem*. 395:25-35. Review.
- King TP, Hoffman D, Lowenstein H, Marsh David G, Platts Mills Thomas AE, Thomas W. 1995. Allergen nomenclature. *J Allergy Clin Immunol* 96:5–14.
- Kootstra HS, Vlieg-Boerstra BJ, Dubois AEJ. 2007. Assessment of the reduced properties of the Santana apple. *Ann Allergy Asthma Immunol*. 99:522-5.
- Larionov, A., Krause, A. and Miller, W. 2005. A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*, 6:62.
- Marzban G, Mansfeld A, Hemmer W, Stoyanova E, Katinger H, da Câmara Machado ML. 2005. Fruit cross-reactive allergens: a theme of uprising interest for consumers' health. *Biofactors*. 23:235-41. Review.

- Oberhuber C, Ma Y, Marsh J, Rigby N, Smole U, Radauer C, Alessandri S, Briza P, Zuidmeer L, Maderegger B, Himly M, Sancho AI, van Ree R, Knulst A, Ebner C, Shewry P, Mills EN, Wellner K, Breiteneder H, Hoffmann-Sommergruber K, Bublin M. 2008. Purification and characterisation of relevant natural and recombinant apple allergens. *Mol Nutr Food Res*. 52 Suppl 2:S208-19.
- Pagliarani G., Paris R., Tartarini S., Sansavini S. Cloning and expression of the major allergen genes in apple fruit. *JHSB ISAFRUIT Special Issue* 176–181.
- Paris, R., Cova, V., Pagliarani, G., Tartarini, S., Komjanc, M. and Sansavini, S. 2009. Expression profiling in HcrVf2-transformed apple plants in response to *Venturia inaequalis*. *Tree Genetics & Genomes*, 5, 81–91.
- Puehringer H, Zinoecker I, Marzban G, Katinger H, Laimer M. 2003. MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene* 321:173-183.
- Radauer C, Lackner P, Breiteneder H. 2008. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *MMC Evol Biol*. 15;8:286.
- Rutledge RG, Côté C. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res.*31:e93.
- Shewry PR, Beaudoin F, Jenkins J, Griffiths-Jones S, Mills EN. 2002. Plant protein families and their relationships to food allergy. *Biochem Soc Trans*. 30(Pt 6):906-10. Review.
- 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 Nutri*. 38:201–215
- Tichopad A, Dilger M, Schwarz G, Pfaffl MW. 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res*. 31:66-88.
- Udvardi MK, Czechowski T, Scheible WR. 2008. Eleven golden rules of quantitative RT-PCR. *Plant Cell*. 20:1736-7.
- van Loon L.C., Rep M., Pieterse C.M. 2006. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- Vieths S, Jankiewicz A, Schoning B, Aulepp H. 1994. Apple allergy: the IgE-binding potency of apple strains is related to the occurrence of the 18-kDa allergen. *Allergy* 49:262–271.
- Vision TJ, Brown DG, Tanksley SD. 2000. Science. The origins of genomic duplications in *Arabidopsis*.v290(5499):2114-7.
- Wortman JR, Haas BJ, Hannick LI, Smith RK Jr, Maiti R, Ronning CM, Chan AP, Yu C, Ayele M, Whitelaw CA, White OR, Town CD. 2003. Annotation of the *Arabidopsis* genome. *Plant Physiol*. 132:461-8.
- Wu, C H., Huang, H., Yeh, L S., Barker, W C. 2003. Protein family classification and functional annotation. *Computational Biology and Chemistry* 27, 37-47.
- Yagami T. 2002. Allergies to cross-reactive plant proteins. Latex-fruit syndrome is comparable with pollen-food allergy syndrome. *Int Arch Allergy Immunol*. 128:271-9. Review.
- Zuidmeer L, van Leeuwen WA, Kleine Budde I, Breiteneder H, Ma Y, Mills C, Sancho AI, Meulenbroek EJ, van de Weg E, Gilissen L, Ferreira F, Hoffmann-Sommergruber K, van Ree R. 2006. Allergenicity assessment of apple cultivars: hurdles in quantifying labile fruit allergens. *Int Arch Allergy Immunol*. 141:230-40.

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 et 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'-ATTGGCAGTGTGCGACGTT-3')/ GAPDH-Rev (5'-GGAGGAGTC AATGGTGGAGGA-3'); UBC-For(5'-CGAATTTGTCCGAAGGCGT-3')/UBC-Rev(5'-AATGAT TGTCACAGCAGCCA-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 threshold (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.

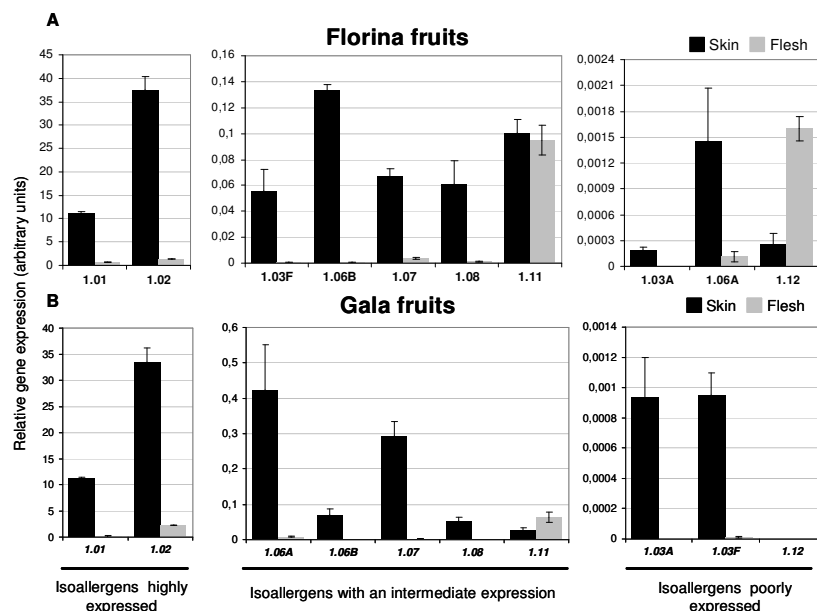


Figure 1: Quantification, by qPCR, of the transcript levels of the isoallergens *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.03A*, *Mal d 1.03F*, *Mal d 1.06A*, *Mal d 1.06B*, *Mal d 1.07*, *Mal d 1.08*, *Mal d 1.11* and *Mal d 1.12* in apple fruit skin and flesh at harvest of the cv. Florina (Panel A) and Gala (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 \pm 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 1. *Cis*-elements found in the *Mal d 1.02* upstream region and not in *Mal d 1.06A* and *Mal d 1.07* upstream regions with the software PLACE.

Factor	Site*	Strand	Motif	Species
UP2ATMSD	900	+	AAACCCTA	Arabidopsis
P1BS	1209	+	GTATATAC	Arabidopsis/tomato/Medicago/barley
P1BS	1209	-	GTATATAC	Arabidopsis/tomato/Medicago/barley
PIATGAPB	89	+	GTGATCAC	Arabidopsis
PIATGAPB	89	-	GTGATCAC	Arabidopsis

* 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%.



Figure 2: Alignment of the promoter regions (-200 bp from the start-codon) of *Mal d 1.02*, *Mal d 1.06A* and *Mal d 1.07*. 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 1.02* and *Mal d 1.07* are highlighted in black; the nucleotides in common between *Mal d 1.02* and *Mal d 1.06A* are highlighted in yellow and the nucleotides in common between *Mal d 1.06A* and *Mal d 1.07* are highlighted in grey.

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 to 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.

	SeqA Name	Len(nt)	SeqB Name	Len(nt)	Score
1	Mald102	150	2 Mald106A	150	50
1	Mald102	150	3 Mald106B	150	50
1	Mald102	150	4 Mald107	150	46
1	Mald102	150	5 Mald108	150	43
2	Mald106A	150	3 Mald106B	150	86
2	Mald106A	150	4 Mald107	150	54
2	Mald106A	150	5 Mald108	150	58
3	Mald106B	150	4 Mald107	150	54
3	Mald106B	150	5 Mald108	150	52
4	Mald107	150	5 Mald108	150	79

Interestingly, the *Mal d 1* genes have been found to share also a 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*.

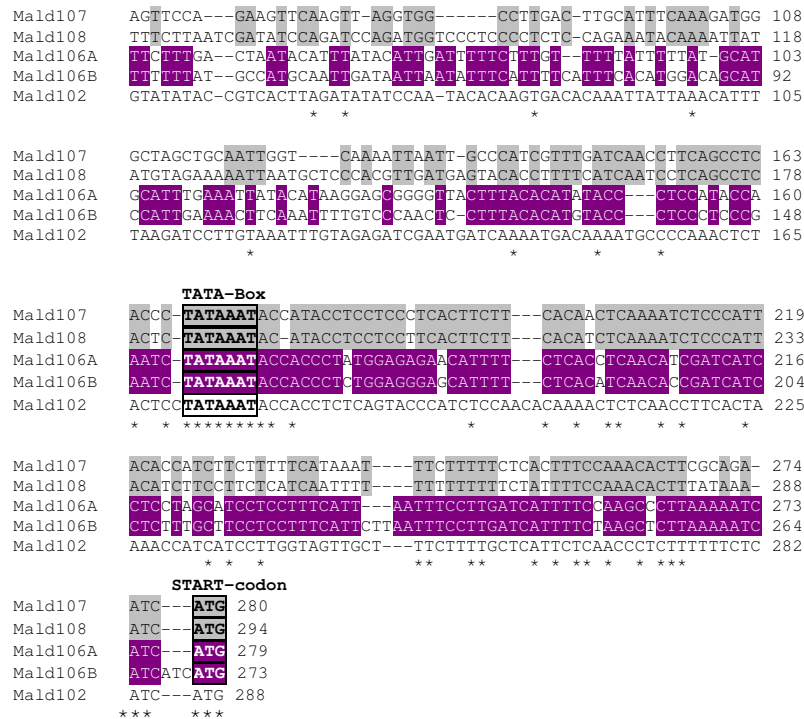


Figure 3: Alignment of the promoter regions (-250 bp from the start-codon) of *Mal d 1.02* (subfamily I), *Mal d 1.06A* and *Mal d 1.06B* (subfamily III), *Mal d 1.07* and *Mal d 1.08* (subfamily IV). Start codons and TATA-boxes are indicated with boxes. The nucleotides in common for the three sequences 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.

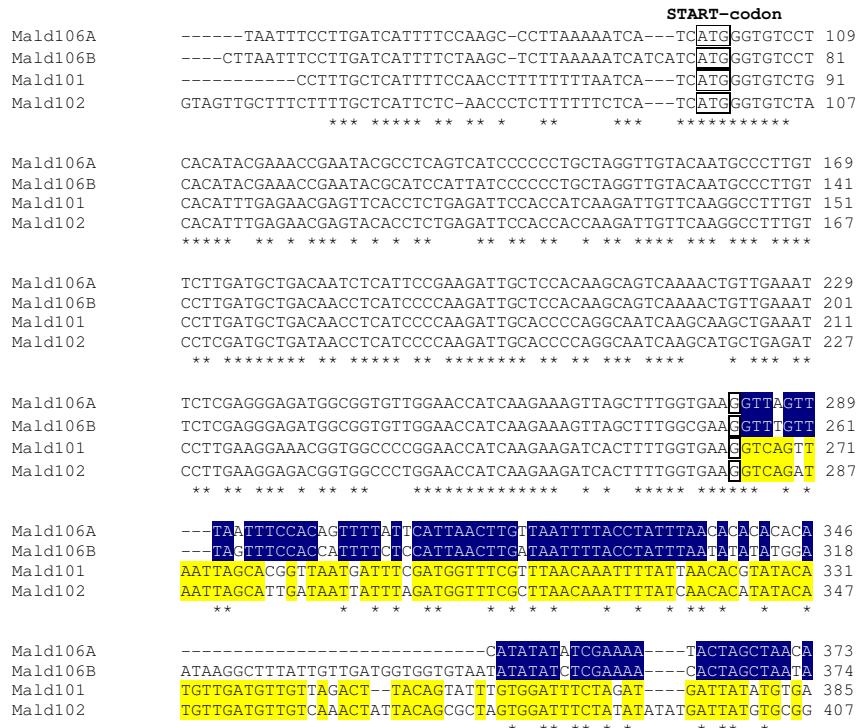




Figure 4: Alignment of nucleotidic sequences of *Mal d 1.01* and *Mal d 1.02*, belonging to subfamily I, and *Mal d 1.06A* and *Mal d 1.06B*, belonging to subfamily III. Start codons, stop codons and the extreme nucleotides of exons were indicated with boxes. In the intron sequences, the identical nucleotides were highlighted 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 to 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 1n 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.

Reference	Technique	Apple genotypes					
		Gala	Fiesta	Florina	Durello di Forlì	Jonathan	Jonagold
Ricci et al., 2010 (submitted)	Prick-to-Prick Test	Skin: HIGH	Skin: HIGH	Skin: INTERMEDIATED	Skin: LOW	Skin: LOW	Skin: LOW
		Flesh: INTERMEDIATED	Flesh: HIGH	Flesh: INTERMEDIATED	Flesh: LOW	Flesh: HIGH	Flesh: HIGH
Bolhaar et al., 2004	Prick-to-Prick Test	HIGH	HIGH	/	/	/	/
Marzban et al., 2005	Total Mal d 1 protein determination	/	/	/	/	/	LOW
Son et al., 1999	Total Mal d 1 protein determination	/	/	/	/	/	LOW

The complex expression profile of *Mal d 1* genes in different genotypes is reported in Figure 5. These results are not easily to 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 Durello 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.

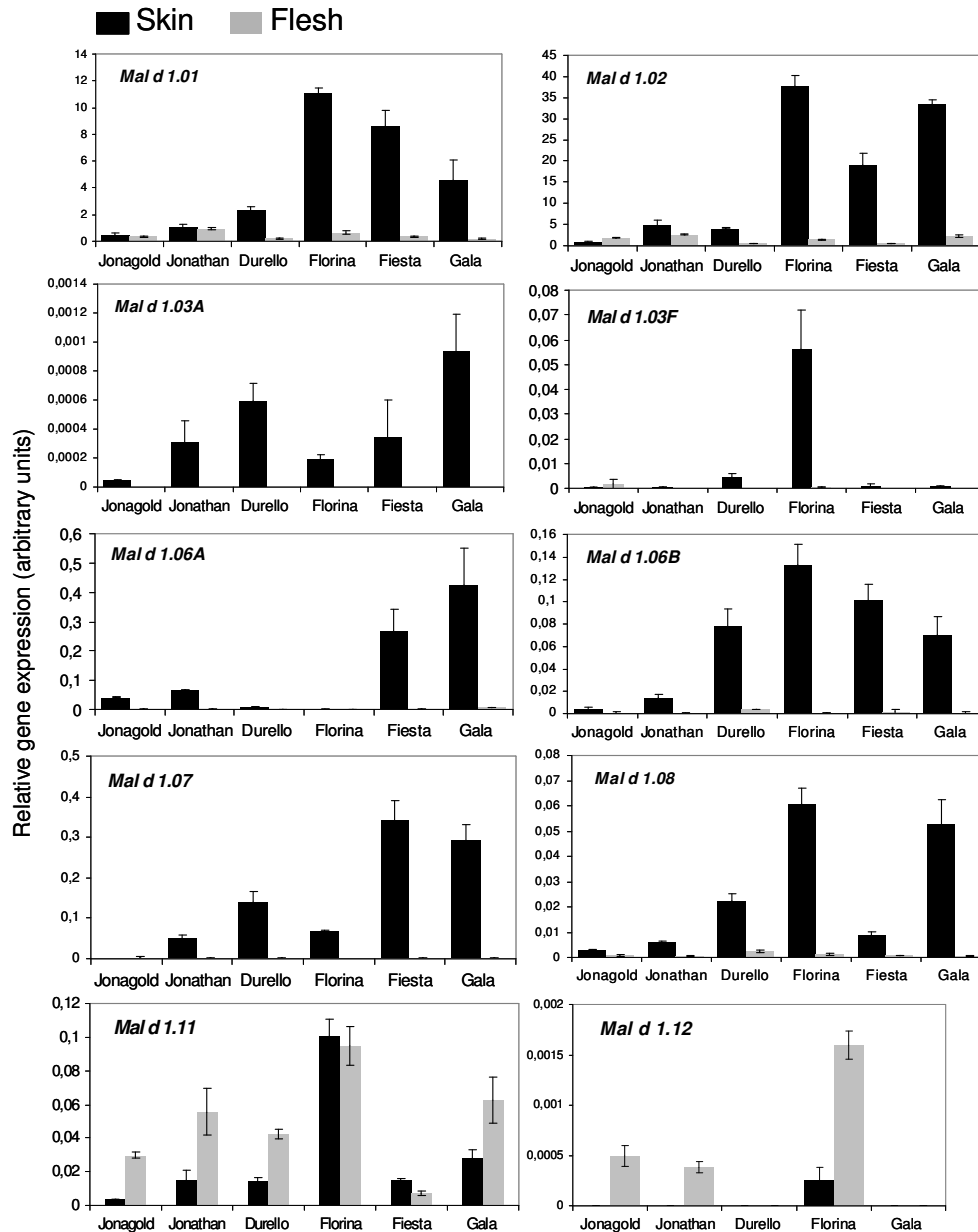


Figure 5: Quantification, by qPCR, of the transcript levels of the isoallergens *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.03A*, *Mal d 1.03F*, *Mal d 1.06A*, *Mal d 1.06B*, *Mal d 1.07*, *Mal d 1.08*, *Mal d 1.11* and *Mal d 1.12* in apple fruit skin and flesh at harvest. The levels of expression were calculated on the cv Jonagold, Jonathan, Durello di Forli, 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 \pm SE. Black bars represented the relative expression in the apple skin and grey bars the relative expression in flesh.

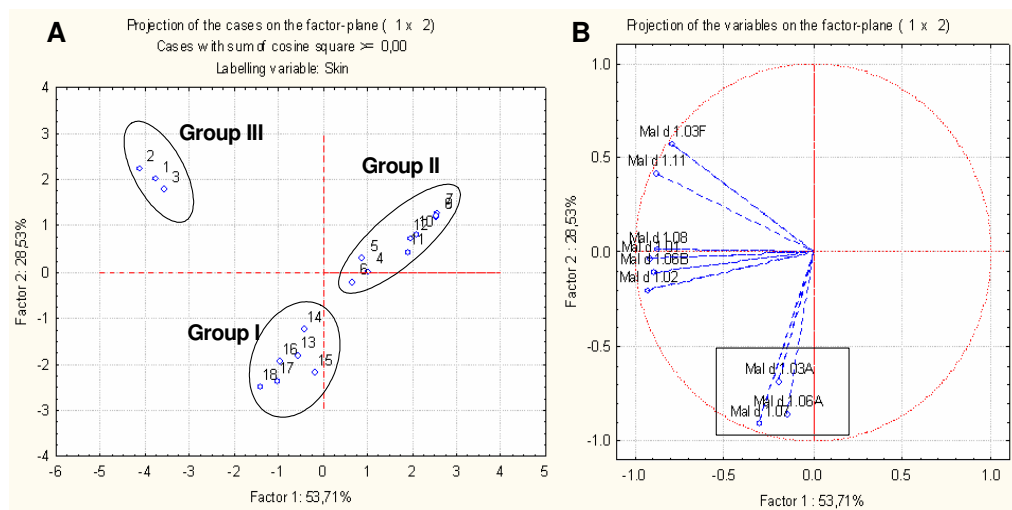


Figure 6: Results of the PCA. A) Grouping of the 6 genotypes, using Euclidean coefficients based on the *Mal d 1* genes expression data. In the charts the points: Three groups of genotypes were identified according to the first two factors: group I containing the replicates 13, 14 and 15 of the cultivar 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*.

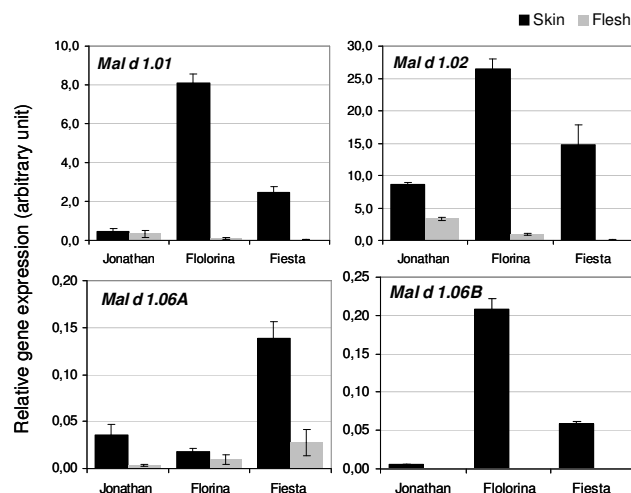


Figure 7: Quantification, by qPCR, of the transcript levels of the isoallergens *Mal d 1.01*, *-1.02*, *-1.06A* and *-1.06B* in a biological replicate of apple fruit skin and flesh at harvest. The levels of expression were calculated on the cv Jonagold, Jonathan, Durello di Forlì, Florina, Fiesta and 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 \pm 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.

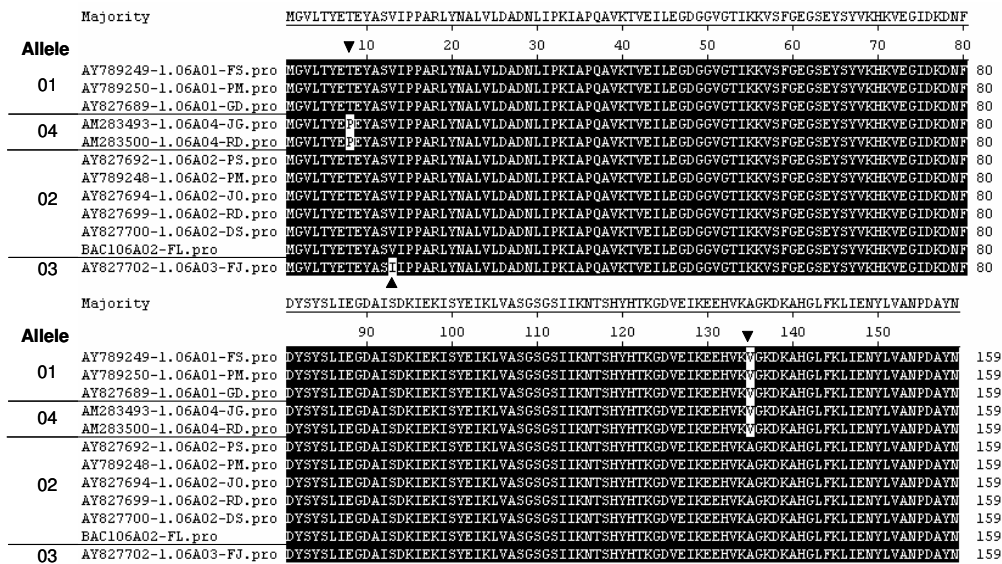


Figure 8: Alignment of the predicted amino acid sequences of *Mal d 1.06A* allele variants of different cultivars: Fiesta (FS), Prima (PM), Golden Delicious (GD), James Grieve (JG), Red Delicious (RD), Priscilla (PS), Jonathan (JO), Discovery (DS), Florina (FL) and Fuji (FJ). The amino acid substitutions among allele variants are indicated with arrows.

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 showed 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 allergenic 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 environmental 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

- Andersen MB, Hall S, Dragsted LO. 2009. Identification of European Allergy Patterns to the Allergen Families PR-10, LTP, and Profilin from Rosaceae Fruits. *Clin Rev Allergy Immunol*. [Epub ahead of print].
- Asero R, Marzban G, Martinelli A, Zaccarini M, Machado ML. 2006. Search for low-allergenic apple cultivars for birch-pollen-allergic patients: is there a correlation between in vitro assays and patient response? *Eur Ann Allergy Clin Immunol*. 38:94-8.
- Asero R, Mistrello G, Roncarolo D, Casarini M. 2000. Detection of allergens in plantain (*Plantago lanceolata*) pollen. *Allergy*. 55:1059-62.
- Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA. 2004. Characterisation of Mal d 1-related genes in *Malus*. *Plant Mol Biol* 55:369-388.
- Bohle B, Zwoelfer B, Heratizadeh A, Jahn-Schmid B, Antonia YD, Alter M et al. 2006. Cooking birch pollen-related food: divergent consequences for IgE- and T cell-mediated reactivity in vitro and in vivo. *J Allergy Clin Immunol* 118:242–249.
- Bolhaar ST, Zuidmeer L, Ma Y, Ferreira F, Bruijnzeel-Koomen CA, Hoffmann-Sommergruber K, van Ree R, Knulst AC. 2005. A mutant of the major apple allergen, Mal d 1, demonstrating hypo-allergenicity in the target organ by double-blind placebo-controlled food challenge. *Clin Exp Allergy*. 35:1638-44.
- Bolhaar STHP, van de Weg WE, Van Ree R, Gonzalez-Mancebo E, Zuidmeer L, Bruijnzeel-Koomen CAFM, Fernandez-Rivas M, Jansen J, Hoffmann-Sommergruber K, Knulst AC, Gilissen LJWJ. 2004.: In vivo assessment with prick-to-prick testing and double-blind, placebo-controlled food challenge of allergenicity of apple cultivars. *J Allergy Clin Immunol*. 114: 1080-1086.
- Bondino HG and Valle EM. 2009. A small intergenic region drives exclusive tissue-specific expression of the adjacent genes in *Arabidopsis thaliana*. *BMC Molecular biology* 10:95.
- Botton A., Lezzer P., Dorigoni A., Barcaccia G., Ruperti B., Ramina A. 2008. Genetic and environmental factors affecting allergen-related gene expression in apple fruit (*Malus domestica* L. Borkh.). *J Agric Food Chem.*, 56, 6707-16.
- Botton A., Lezzer P., Dorigoni A., Ruperti B., Ramina A. 2009. Environmental factors affecting the expression of apple (*Malus X domestica*) allergen-encoding genes. *JHSB ISAFRUIT Special Issue* 182-187.
- Breiteneder H, Radauer C. 2004. A classification of plant food allergens. *J Allergy Clin Immunol*. 113:821-30.
- Chan CS, Guo L, Shih MC. 2001. Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of *Arabidopsis thaliana*. *Plant Mol Biol*. 46:131-41.
- Ebner C, Hirschwehr R, Bauer L, Breiteneder H, Valenta R, Ebner H et al. 1995. Identification of allergens in fruits and vegetables—Ige cross-reactivities with the important birch pollen allergens Bet-V-1 and Bet-V-2 (birch profilin). *J Allergy Clin Immunol* 95:962–969.
- Fernandez-Rivas M, Cuevas M. 1999. Peels of Rosaceae fruits have a higher allergenicity than pulps. *Clin Exp Allergy* 29 :1239–1247.
- Fernández-Rivas M, van Ree R, Cuevas M. 1997. Allergy to Rosaceae fruits without related pollinosis. *J Allergy Clin Immunol*. 100:728-33.
- Gaier S, Marsh J, Oberhuber C, Rigby NM, Lovegrove A, Alessandri S et al 2008. Purification and structural stability of the peach allergens Pru p 1 and Pru p 3. *Mol Nutr Food Res* 52: S220–S229.
- Gao Z, van de Weg EW, Matos CI, Arens P, Bolhaar ST, Knulst AC, Li Y, Hoffmann-Sommergruber K, Gilissen LJ. 2008. Assessment of allelic diversity in intron-containing Mal d 1 genes and their association to apple allergenicity. *BMC Plant Biol*. 13;8:116.
- Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ. 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet*. 111:171-83.
- Hamer M, Chida Y 2007. Intake of fruit, vegetables, and antioxidants and risk of type 2 diabetes: systematic review and meta-analysis. *J Hypertens* 25:2361–2369.

- Helsper JPPG, Gilissen LJWJ, van Ree R, America Antoine HP, Cordewener Jan HG, Bosch D. 2002. Quadrupole time-of-flight mass spectrometry: A method to study the actual expression of allergen isoforms identified by PCR cloning. *J Allergy Clin Immunol* 110:131-138.
- Ma Y, Gadermaier G, Bohle B, Bolhaar S, Knulst A, Markovic-Housley Z, Breiteneder H, Briza P, Hoffmann-Sommergruber K, Ferreira F. 2006. Mutational analysis of amino acid positions crucial for IgE-binding epitopes of the major apple (*Malus domestica*) allergen, Mal d 1. *Int Arch Allergy Immunol.* 139:53-62.
- Marković-Housley Z, Miglierini G, Soldatova L, Rizkallah PJ, Müller U, Schirmer T. 2000. Crystal structure of hyaluronidase, a major allergen of bee venom. 8:1025-35.
- Marzban G, 1, Puehringer H, Dey R, Brynda S, Ma Y, Martinelli A, Zaccarini M, van der Weg E, Housley Z, Kolarich D, Altmann F and Laimer M. 2005. Localisation and distribution of the major allergens in apple fruits. *Plant Science* 169, I2: 387-394.
- Pagliarani G., Paris R., Tartarini S., Sansavini S. Cloning and expression of the major allergen genes in apple fruit. *JHSB ISAFRUIT Special Issue* 176–181.
- Paris, R., Cova, V., Pagliarani, G., Tartarini, S., Komjanc, M. and Sansavini, S. 2009. Expression profiling in HcrVf2-transformed apple plants in response to *Venturia inaequalis*. *Tree Genetics & Genomes*, 5, 81–91.
- Puehringer H, Zinoecker I, Marzban G, Katinger H, Laimer M. 2003. MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene* 321:173-183.
- Ricci G, Dondi A, Belotti T, Baldi E, Tartarini S, Paris R, Pagliarani G, Serafini-Fracassini D, Casadio R, Giannetti A, Masi M. 2010. Submitted to the *International Journal of Food Science & Technology*.
- Rubio V, Linhares F, Solano R, Martín AC, Iglesias J, Leyva A, Paz-Ares J. 2001. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes Dev.* 15:2122-33.
- Sancho AI, Gillabert M, Tapp H, Shewry PR, Skeggs PK, Mills EN. 2006. Effect of environmental stress during grain filling on the soluble proteome of wheat (*Triticum aestivum*) dough liquor. *J Agric Food Chem.* 56:5386-93. Epub 2008 Jun 14.
- Sanzol J. 2009. Pistil-function breakdown in a new S-allele of European pear, S21*, confers self-compatibility. *Plant Cell Rep.* 28:457-67.
- 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(4):201-15.
- Tatematsu K, Ward S, Leyser O, Kamiya Y, Nambara E. 2005. Identification of cis-elements that regulate gene expression during initiation of axillary bud outgrowth in *Arabidopsis*. *Plant Physiol.* 138(2):757-66. Epub 2005 May 20.
- Vainio H, Weiderpass E. 2006. Fruit and vegetables in cancer prevention. *Nutrition and Cancer-An International Journal* 54:111–142.
- van Loon L.C., Rep M., Pieterse C.M. 2006. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- Vanek-Krebitz M, Hoffmann-Sommergruber K, Laimer da Camara Machado M, Susani M, Ebner C, Kraft D, Scheiner O, Breiteneder H. 1995. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem Biophys Res Commun.* 214:538-51.
- Vieths S, Hoffmann A, Holzhauser T, Müller U, Reindl J, Haustein D. 1998. Factors influencing the quality of food extracts for in vitro and in vivo diagnosis. *Allergy.* 53:65-71.
- Vieths S, Jankiewicz A, Schöning B, Aulepp H. 1994. Apple allergy: the IgE-binding potency of apple strains is related to the occurrence of the 18-kDa allergen. *Allergy.* 49:262-71.
- Yang H. 2009. In plants, expression breadth and expression level distinctly and non-linearly correlate with gene structure. *Biol Direct.* ;4:45; discussion 45.

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 β -sheet wrapped around a C-terminal α -helix and two additional small α -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 β 2 and β 3 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* (Markovic-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 acclimatation (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 known that PR-10 expression is ripening-related in apple fruits (Atkinson et al., 1996; Goulao and Oliveira, 2006; Pagliarani et al., 2009). Besides 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 *HcrVf2* 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 than 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.

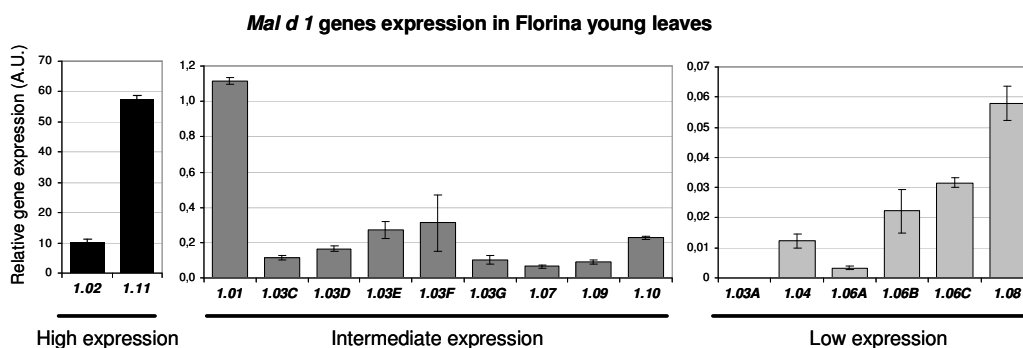


Figure 1: Quantification, by qPCR, of the transcript levels of the genes *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.03A*, *Mal d 1.03C*, *Mal d 1.03D*, *Mal d 1.03E*, *Mal d 1.03F*, *Mal d 1.03G*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.06B*, *Mal d 1.06C*, *Mal d 1.07*, *Mal d 1.08*, *Mal d 1.09*, *Mal d 1.10* and *Mal d 1.11* in Florina, not challenged young leaves. 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 \pm SE. Black bars represented the relative expression in the highly expressed genes, the dark grey bars the relative expression of genes with an intermediate expression and the light grey bars the low expressed genes.

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 tissue 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.

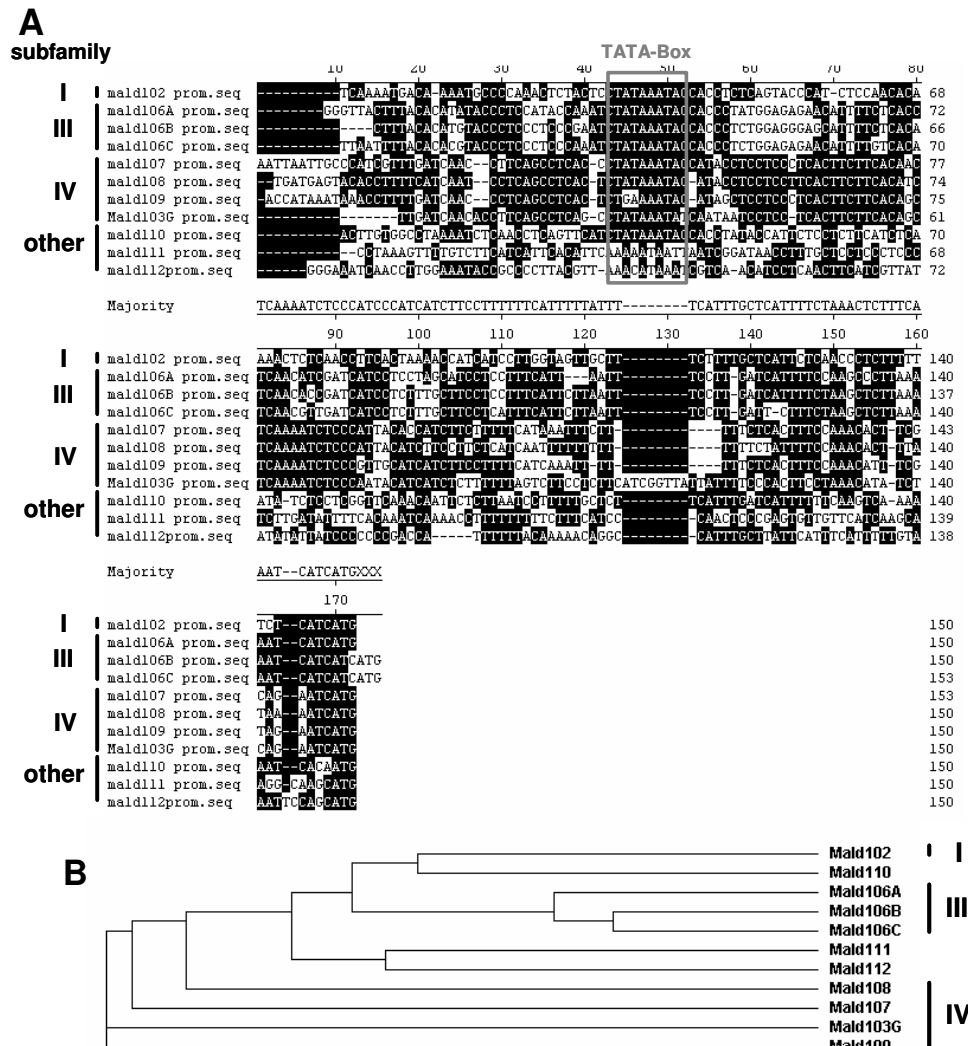


Figure 2: Alignment (panel A) and cladogram (panel B) of 150 bp upstreaming regions of *Mal d 1.02*, *Mal d 1.03G*, *Mal d 1.06A*, *Mal d 1.06B*, *Mal d 1.06C*, *Mal d 1.07*, *Mal d 1.08*, *Mal d 1.09*, *Mal d 1.10*, *Mal d 1.11* and *Mal d 1.12*. The cardinal numbers indicate the subfamilies.

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 *HcrVf2*, 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 (\log_2 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 1: Levels of induction of *Mal d 1* genes after the inoculation with the *V. inaequalis* and with water (controls) expressed as \log_2 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). The highest and the lowest values among all the genes are indicated with boxes.

Treatment	Timing	<i>Mal d 1</i> genes																
		1.01	1.02	1.03A	1.03C	1.03D	1.03E	1.03F	1.03G	1.04	1.06A	1.06B	1.06C	1.07	1.08	1.09	1.10	1.11
<i>V. inaequalis</i>	T24/T0	3,4	2,2	3,9	4,3	3,9	3,7	4,9	4,1	3,1	3,1	3,6	2,1	4,9	2,4	3,1	-1,8	1,3
	T48/T0	3,4	2,3	2,0	2,4	3,1	1,6	5,4	3,6	6,2	11,0	2,2	0,2	4,3	1,6	2,1	-3,2	1,3
	T72/T0	4,6	2,7	3,6	4,0	3,5	3,0	5,8	4,7	3,3	3,9	2,2	0,8	4,4	1,1	3,5	-4,8	2,0
	T96/T0	4,0	2,0	3,6	2,4	3,6	3,2	3,9	3,5	1,5	3,1	2,2	1,6	4,1	1,1	3,3	-2,4	1,2
water	T24/T0	1,7	0,5	2,3	1,3	2,8	3,6	1,2	1,3	0,3	0,4	1,7	1,6	1,7	0,2	1,7	-0,6	-0,1
	T48/T0	3,0	1,6	1,5	1,7	3,7	4,6	2,1	2,0	3,3	0,5	2,7	2,5	2,8	0,7	1,8	-0,3	-0,3
	T72/T0	1,1	1,3	-1,1	0,4	3,0	3,1	2,6	1,9	3,3	0,3	0,0	0,0	1,9	0,5	2,8	-1,2	0,2
	T96/T0	4,5	2,8	0,8	1,9	4,0	4,2	4,8	4,0	5,3	2,8	1,6	0,6	2,7	0,3	1,4	-3,7	1,2

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 Ekramoddoullah, 2004).

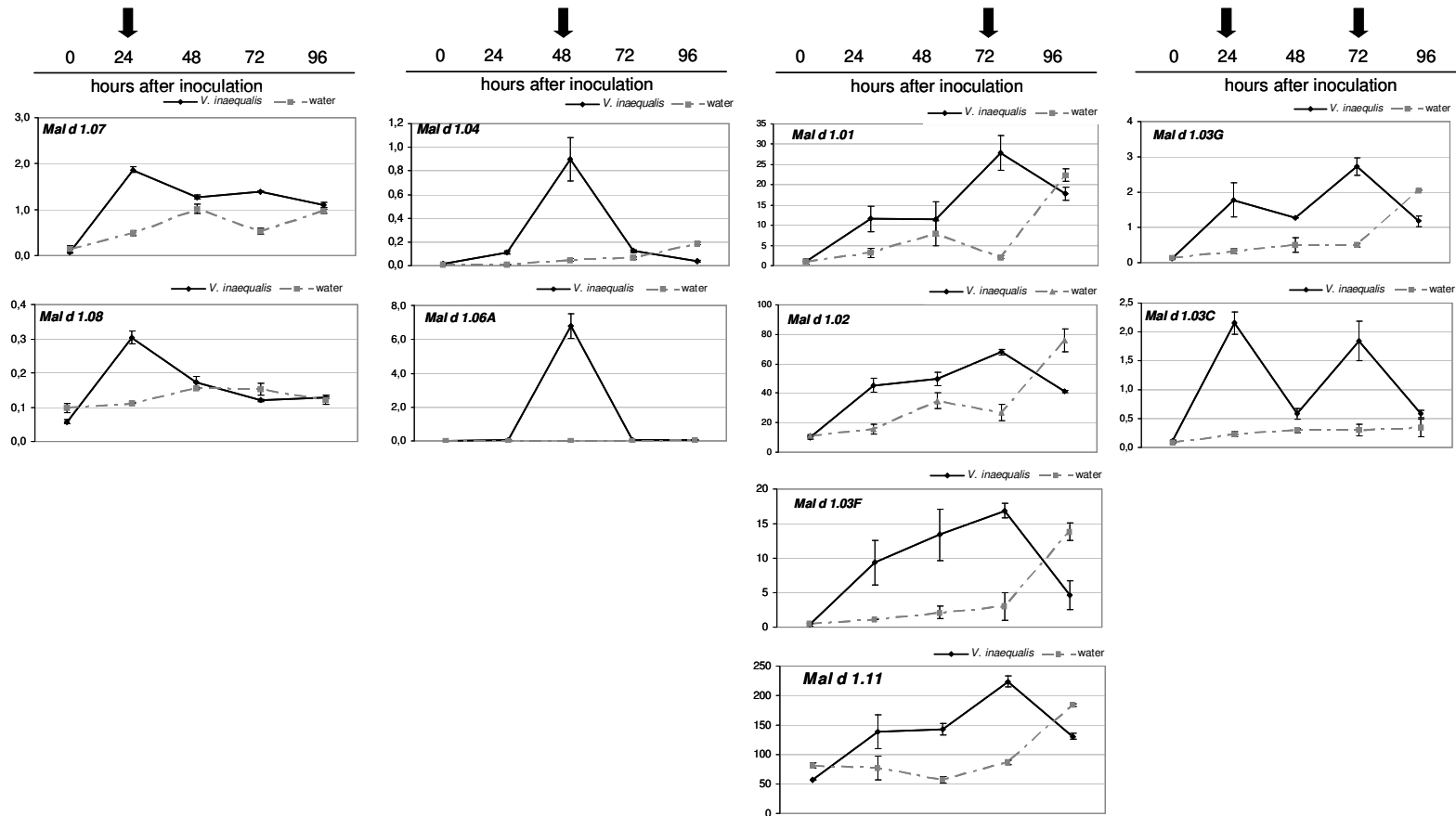


Figure 3: Quantification, by qPCR, of the transcript levels of the genes *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.03C*, *Mal d 1.03F*, *Mal d 1.03G*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.07*, *Mal d 1.08* and *Mal d 1.11* in Florida young leaves at different times (0, 24, 48, 72 and 96 hours) after the 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 \pm 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 knockdown 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 \log_2 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 \log_2 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 \log_2 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 \log_2 ratio of -4.8. In particular, the highest modulation have been detected for *Mal d 1.06A* that presented a maximum \log_2 ratio of 11 and *Mal d 1.11* resulted the least induced with a \log_2 ratio of 2 (Figure 6).

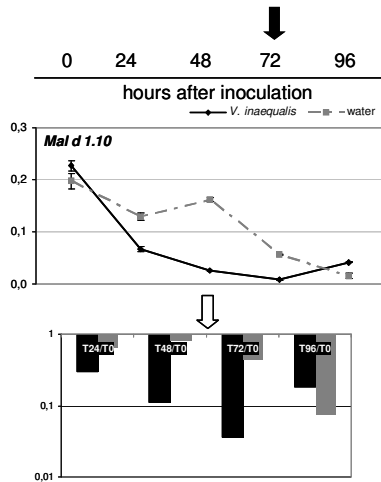


Figure 4: Absolute quantification, and Relative Quantitation by qPCR, of the transcript levels of the Mal d 1.10 gene in Florina young leaves at different times (0, 24, 48, 72 and 96 hours) after the inoculation. In the upper chart is reported the levels of expression calculated using the standard curve method and transcripts accumulations are reported as normalized expression, in arbitrary units (A.U.). The values in this charts resulted from three technical replicates \pm 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 \log_2 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 .

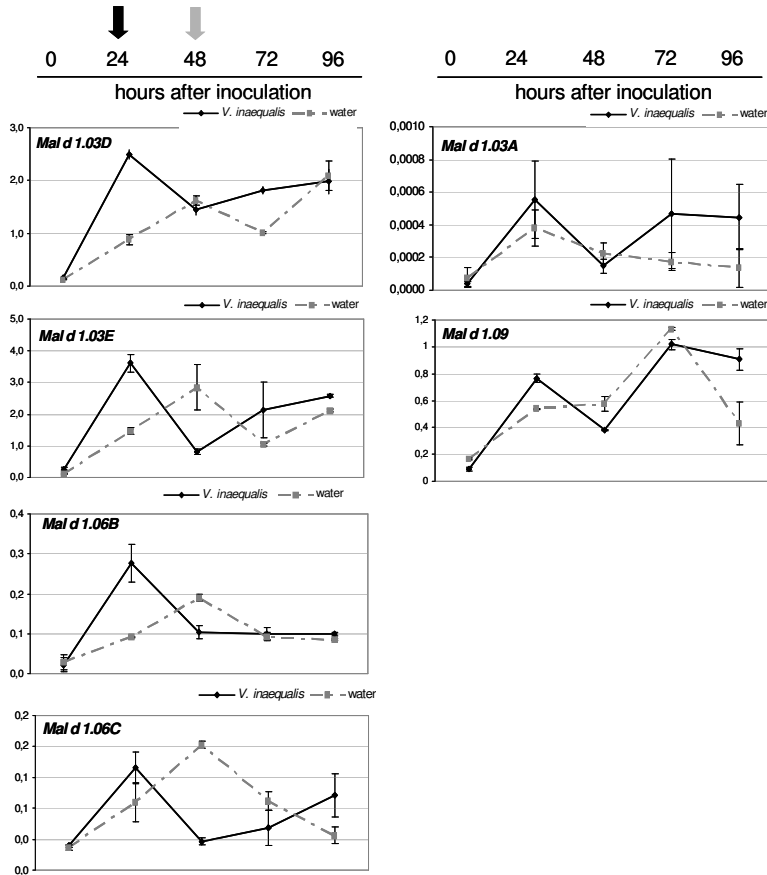
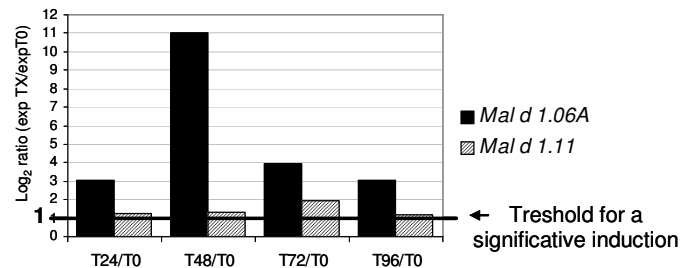


Figure 5: Quantification, by qPCR, of the transcript levels of the genes *Mal d 1.03D*, *Mal d 1.03E*, *Mal d 1.06B*, *Mal d 1.06C*, *Mal d 1.03A* and *Mal d 1.09* in Florina young leaves at different times (0, 24, 48, 72 and 96 hours) after the inoculation. The black lines represent the expression in 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 \pm 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 \log_2 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 occur 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 *Mal d 1.03G*, *Mal d 1.03C*, *Mal d 1.03A* and *Mal d 1.09*. Up to now, there are no elements to interpret the different timing of induction of *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 *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 *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 *Mal d 1* genes for which a not clear distinction between treated and controls were reported. In particular, for *Mal d 1.03D*, *Mal d 1.03E*, *Mal d 1.06B* and *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 *Mal d 1.03A* and *Mal d 1.09*, the expression in leaves challenged with *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 *Pinus Monticola* and a strong induction of this protein was reported after the infection with the pathogen *P. Syringae* as well as after the challenge with water. Similar results were reported for other PR proteins in plants of *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 sperimental conditions (the plant conditions, the inoculum, the darkness, and humidity) can influence their gene expression and the modulation.

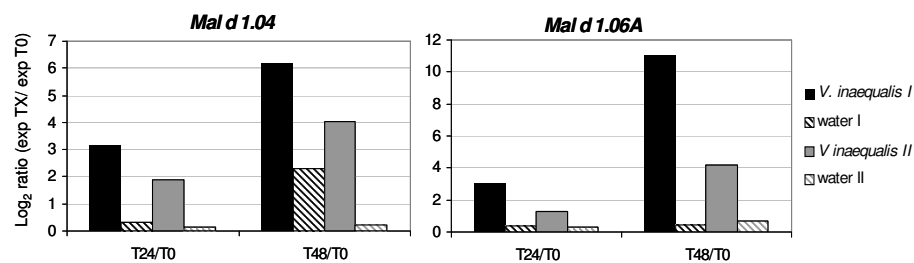


Figure 7: Comparison of the modulation of gene expression of *Mal d 1.04* (left chart) and *Mal d 1.06A* (right chart) on two biological replicates of young leaves of Florina at different times (24, 48, 72 and 96h) after the challenging with *V. inaequalis* and with water described as log₂ of the ratio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

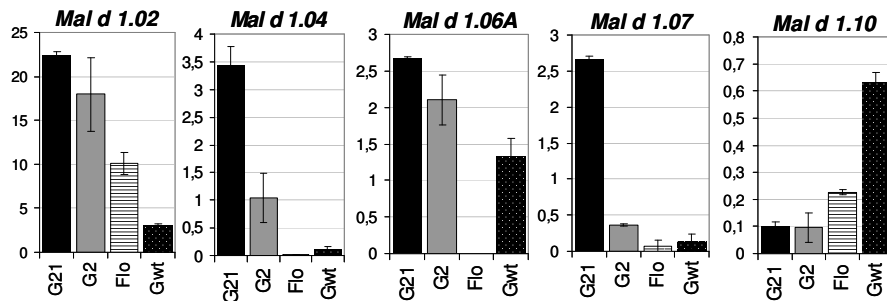
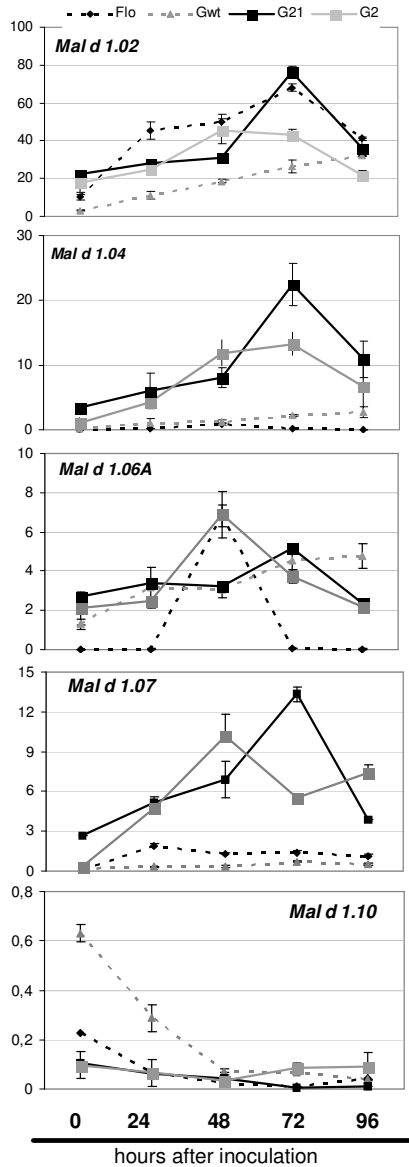


Figure 8: Quantification, by qPCR, of the basal transcript levels of the genes *Mal d 1.02*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.07* and *Mal d 1.10* in young leaves of different genotype: Florina (white bars with vertical lines), Gala (black bars with points), G21 (black bars) and G2 (gray bars). 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 \pm SE.

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 especially 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 \log_2 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).

	<i>Mal d 1.02</i>			<i>Mal d 1.04</i>			<i>Mal d 1.06A</i>			<i>Mal d 1.07</i>			<i>Mal d 1.10</i>		
	G2-21	G2-2	Gwt	G2-21	G2-2	Gwt	G2-21	G2-2	Gwt	G2-21	G2-2	Gwt	G2-21	G2-2	Gwt
T24/T0	0,3	0,5	1,9	0,8	2,0	3,3	0,4	0,2	1,3	0,9	3,7	1,2	-0,8	-0,6	-1,1
T48/T0	0,5	1,3	2,7	1,2	3,5	3,6	0,3	1,7	1,2	1,4	4,8	1,4	-1,3	-1,6	-3,1
T72/T0	1,8	1,3	3,2	2,7	3,7	4,4	1,0	0,8	1,8	2,3	3,9	2,4	-4,0	-0,2	-3,3
T96/T0	0,7	0,3	3,5	1,7	2,7	4,7	-0,2	0,1	1,9	0,6	4,3	2,0	-3,3	-0,1	-4,1



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 gens *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 \pm 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 co-ordinated 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 redundancy 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

- Atkinson RG, Perry J, Matsui T, Ross GS, Macrea EA. 1996. A stress-, pathogenesis-, and allergen-related cDNA in apple fruit is also ripening-related. *N Z J Crop Hort Sci*; 24:103–7.
- Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S. 2004. The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proc Natl Acad Sci U S A*. 101:886-90.
- Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA. 2004. Characterisation of Mal d 1-related genes in *Malus*. *Plant Mol Biol* 55:369-388.
- Biesiadka J, Bujacz G, Sikorski MM, Jaskolski M. 2002. Crystal structures of two homologous pathogenesis-related proteins from yellow lupine. *J Mol Biol*;319:1223–34.
- Bohle B, Zwoelfer B, Heratizadeh A, Jahn-Schmid B, Antonia YD, Alter M et al. 2006. Cooking birch pollen-related food: divergent consequences for IgE- and T cell-mediated reactivity in vitro and in vivo. *J Allergy Clin Immunol* 118:242–249.
- Breiteneder H, Ebner C. 2000. Molecular and biochemical classification of plant-derived food allergens. *J Allergy Clin Immunol*. 106:27-36. Review.
- Carimi F, Zottini M, Formentin E, Terzi M, Lo Schiavo F. 2003. Cytokinins: new apoptotic inducers in plants. *Planta*. 216:413-21.
- Chadha P, Das RH. 2006. A pathogenesis related protein, AhPR10 from peanut: an insight of its mode of antifungal activity. *Planta*. 225:213-22.
- Chung KM, Igari K, Uchida N, Tasaka M. 2008. New perspectives on plants defense responses through modulation of developmental pathways. *Mol Cells* 26: 107–112
- Colditz F, Niehaus K, Krajinski F. 2007. Silencing of PR-10-like proteins in *Medicago truncatula* results in an antagonistic induction of other PR proteins and in an increased tolerance upon infection with the oomycete *Aphanomyces euteiches*. *Planta*;226:57-71. Epub 2007 Jan 20.
- Dubos C, Plomion C. 2001. Drought differentially affects expression of a PR-10 protein in needles of the maritime pine (*Pinus pinaster* Ait.) seedling. *J Exp Bot*; 52:1143–4.
- Ekramoddoullah AKM, Taylor D, Hawkins BJ. 1995. Characterization of a fall protein of sugar pine and detection of its homologue associated with frost hardiness of western white pine needles. *Can J Forest Res*; 25:1137–47.
- Elvira MI, Galdeano MM, Gilardi P, García-Luque I, Serra MT. 2008. Proteomic analysis of pathogenesis-related proteins (PRs) induced by compatible and incompatible interactions of pepper mild mottle virus (PMMoV) in *Capsicum chinense* L3 plants. *J Exp Bot* 59:1253-65.
- Faize M, Faize L, Koike N, Ishizaka M, Ishii H. 2004. Acibenzolar-s-methyl-induced resistance to Japanese pear scab is associated with potentiation of multiple defense responses. *Phytopathology*. 94:604-12.
- Fernandes H, Pasternak O, Bujacz G, Bujacz A, Sikorski MM, Jaskolski M. 2008. *Lupinus luteus* pathogenesis-related protein as a reservoir for cytokinin. *J Mol Biol* 378: 1040–1051
- Flores T, Alape-Giron A, Flores-Diaz M, Flores HE. 2002. Ocatin, a novel tuber storage protein from the andean tuber crop oca with antibacterial and antifungal activities. *Plant Physiol*; 128: 1291–302.
- Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ. 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet*. 111:171-83.
- Gau AE, Koutb M, Piotrowski M, Kloppstech K. 2004. Accumulation of pathogenesis-related proteins in the apoplast of a susceptible cultivar of apple (*Malus domestica* cv. Elstar) after infection by *Venturia inaequalis* and constitutive expression of PR genes in the resistant cultivar Remo. *Eur J Plant Pathol* 110:703–711

- Goulaoa L F and Oliveira C M. 2006. Molecular identification of novel differentially expressed mRNAs up-regulated during ripening of apples. *Plant Science*. 172 : 306-318.
- Hoffmann-Sommergruber K and Radauer C. Bet v 1-homologous allergens. In: Mills ENC, Shewry PR, editors. *Plant food allergens*. Oxford: Blackwell; 2004. p. 125–40.
- Hoffmann-Sommergruber K. 2000. Plant allergens and pathogenesis-related proteins. What do they have in common? *Int Arch Allergy Immunol*;122:155-66. Review.
- Kleine-Tebbe J, Vogel L, Crowell DN, Haustein UF, Vieths S. 2002. Severe oral allergy syndrome and anaphylactic reactions caused by a Bet v 1-related PR-10 protein in soybean, SAM22. *J Allergy Clin Immunol*. 110:797–804.
- Koistinen KM, Soininen P, Venäläinen TA, Häyrinen J, Laatikainen R, Peräkylä M, Tervahauta AI, Kärenlampi SO. 2005. Birch PR-10c interacts with several biologically important ligands. *Phytochemistry*. 66:2524-33.
- Komjanc M, Festi S, Rizzotti L, Cattivelli L, Cervone F, De Lorenzo G. 1999. A leucine-rich repeat receptor-like protein kinase (LRPKm1) gene is induced in *Malus x domestica* by *Venturia inaequalis* infection and salicylic acid treatment. *Plant Mol Biol*;40:945-57.
- Liu J J and Ekramoddoullah Abul K.M. 2006. The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiological and Molecular Plant Pathology* 68, 3-13.
- Liu JJ, Ekramoddoullah AK, Piggott N, Zamani A. 2005. Molecular cloning of a pathogen/wound-inducible PR10 promoter from *Pinus monticola* and characterization in transgenic *Arabidopsis* plants. *Planta*;221:159–69.
- Liu JJ, Ekramoddoullah AKM. 2004. Characterization, expression and evolution of two novel subfamilies of *Pinus monticola* (Dougl. ex D. Don) cDNAs encoding pathogenesis-related (PR)-10 proteins. *Tree Physiol*;24:1377–85.
- Marković-Housley Z, Degano M, Lamba D, von Roepenack-Lahaye E, Clemens S, Susani M, Ferreira F, Scheiner O, Breiteneder H. 2003. Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *J Mol Biol*. 325:123-33.
- Mogensen JE, Wimmer R, Larsen JN, Spangfort MD, Otzen DE. 2002. The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. *J Biol Chem*; 277:23684–92.
- Moons A, Prinsen E, Bauw G, Van Montagu M. 1997. Antagonistic effects of abscisic acid and jasmonates on salt stress-inducible transcripts in rice roots. *Plant Cell*; 9:2243–59.
- Neudecker P, Schweimer K, Nerkamp J, Scheurer S, Vieths S, Sticht H, et al. 2001. Allergic cross-reactivity made visible: solution structure of the major cherry allergen Pru av 1. *J Biol Chem*; 276:22756–63.
- Pagliarani G., Paris R., Tartarini S., Sansavini S. Cloning and expression of the major allergen genes in apple fruit. *JHSB ISAFRUIT Special Issue* 176–181.
- Paris, R., Cova, V., Pagliarani, G., Tartarini, S., Komjanc, M. and Sansavini, S. 2009. Expression profiling in HcrVf2-transformed apple plants in response to *Venturia inaequalis*. *Tree Genetics & Genomes*, 5, 81–91.
- Park CJ, Kim KJ, Shin R, Park JM, Shin YC, Paek KH. 2004. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant J*; 37:186–98.
- Pasternak O, Biesiadka J, Dolot R, Handschuh L, Bujacz G, Sikorski MM, et al. 2005. Structure of a yellow lupin pathogenesis-related PR-10 protein belonging to novel subclass. *Acta Crystallogr*;D16:99–107.
- Puehringer H, Zinoecker I, Marzban G, Katinger H, Laimer M. 2003. MdAP, a novel protein in apple, is associated with the major allergen Mal d 1. *Gene* 321:173-183.
- Pühringer H, Moll D, Hoffmann-Sommergruber K, Watillon B, Katinger H, Laimer da Câmara Machado M. 2000. The promoter of an apple Ypr10 gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. *Plant Sci*;152:35–50.
- Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner HY, Hunt MD. 1996. Systemic Acquired Resistance. *Plant Cell*. 8:1809-1819.

- Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. 2003. Two different novel cis-acting elements of *erd1*, a *clpA* homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. *Plant J* 33:259-70.
- Truesdell GM, Dickman MB. 1997. Isolation of pathogen/stress-inducible cDNAs from alfalfa by mRNA differential display. *Plant Mol Biol*;33:737-43.
- Ukaji N, Kuwabara C, Takezawa D, Arakawa K, Fujikawa S. 2004. Accumulation of pathogenesis-related (PR) 10/Bet v 1 protein homologues in mulberry (*Morus bombycis* Koidz.) tree during winter. *Plant Cell Environ*;27:1112-21.
- van Loon L.C., Rep M., Pieterse C.M. 2006. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- Verberne MC, Hoekstra J, Bol JF, Linthorst HJ. 2003. Signaling of systemic acquired resistance in tobacco depends on ethylene perception. *Plant J*. 35:27-32.
- Wang Y, Nowak G, Culley D, Hadwiger LA, Fristensky B. 1999. Constitutive expression of pea defense gene DRR206 confers resistance to blackleg (*Leptosphaeria maculans*) disease in transgenic canola (*Brassica napus*). *Mol Plant-Microbe Interact*;12:410-8.
- Wisniewski M, Bassett C, Arora R. 2004. Distribution and partial characterization of seasonally expressed proteins in different aged shoots and roots of 'Loring' peach (*Prunus persica*). *Tree Physiol* ;24:339-45.
- Ziadi S, Poupard P, Brisset M N, Paulin J P, Simoneau P. 2001. Characterization in apple leaves of two subclasses of PR-10 transcripts inducible by acibenzolar-S-methyl, a functional analogue of salicylic acid. *Physiological and Molecular Plant Pathology*. 59: 33-43.
- 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 5

Identification and mapping of new *Mal d 2* and *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 d 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 eukaryotic 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 values $<E^{-15}$ 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 sequencing 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 Forli 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 Forli, were obtained by direct sequencing. All the PCR amplifications were performed in a 17.5 μ l volume containing 50 ng of DNA, 100 nM gene-specific primers, 1.5 mM $MgCl_2$, 100 μ M 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 tissues indicating a putative involvement in the allergic reaction. The alignment of the nucleotide sequences of *Mal d 2.01A* (AY792599), *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 nucleotide 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.

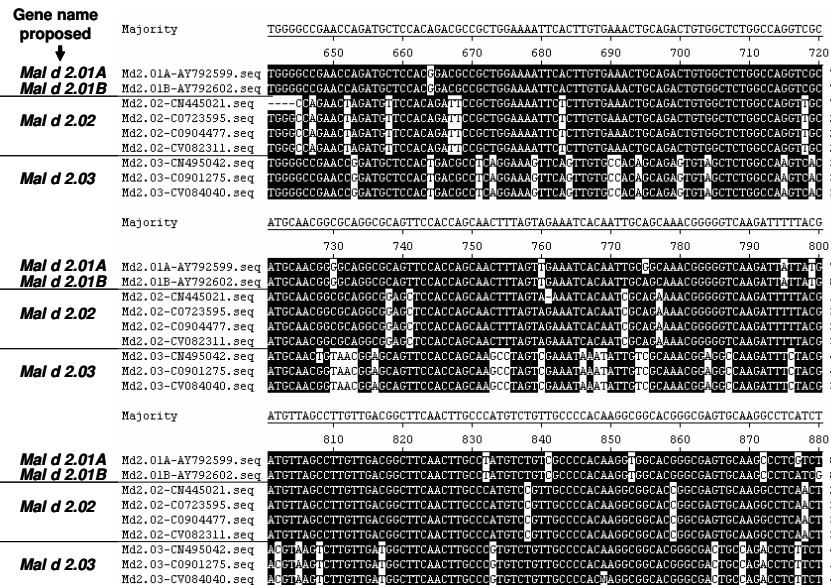
Table 1. *Mal d 2* and *Mal d 4* genes sequences in GenBank/EMBL. Genes studied in this work are highlighted in yellow. All the sequence lengths are indicated in bp.

Gene	Reference genomic seq.	EST	Tissue	Total length	Coding	Ex1	Int1	Ex2	Pop ^a	LG
<i>Mal d 2.01A</i>	AY792599	CO754260, CN996077, CN490732, CO052447, AF494393, AF494394, CV150401, CN994747	Fruit, bud , leaf	1.119	741	61	378	680		9
<i>Mal d 2.01B</i>	AY792602	Not yet	Not known	1.121	741	61	380	680		9
<i>Mal d 2.02</i>	Not yet	CO904477, CV082311, CO723595, CN445021,	Fruit, bud	-	-	-	-	-	PMxFS	17
<i>Mal d 2.03</i>	Not yet	CO866730, CO901275, CN495042, CV084040, CO866711, CN444138, CO866347, CN491711	Fruit, bud , flower	-	-	-	-	-	DUxFS	4

^aPopulations 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*.

Gene	Primer name	sequence (5'-3')	Ta (°C)	Use
<i>Mal d 2.02</i>	Mald2.02Fmap	CCAGAACTAGATGTTCCACAGATT	58	Mapping by direct sequencing
	Mald2.02Rmap	TCAGCCGCTTTCACTTGTAAC		
<i>Mal d 2.03</i>	Mald2.03Fmap	ACGATCACCTTCCACCAACAGT	60	Mapping by TSP
	Mald2.03Rmap	AGCTCAGCTGGGCAAACA		
	Asp76Md2.03F	CCGGTTCGAGTTAGCAT		
<i>Mal d 4.04</i>	Md4group1F1	ATTGACGGCCAGGGACAG	60	BAC library screening, first step of primer walking, mapping by direct sequencing
	Md4group1R1	GGTTCCTCAAAATCCTTGTTG		
	Md4gr1REVstep2	CCAGACACTGCCATCAAGA		
	Md4gr1FORstep2	GAGGAGATGACTGGTATCAACAAG		
	Md4gr1REVstep3	GGTCGTCCAGTAGGTCTG		
<i>Mal d 4.04</i>	Md4gr1FORstep3	GGCAAGCTCTAGTATTTGGAATCT	Primer walking	
	Md4gr1REVstep4	CTGCAACAACCTCACCAACA		

**Figure 1:** Portion of the nucleotide 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).

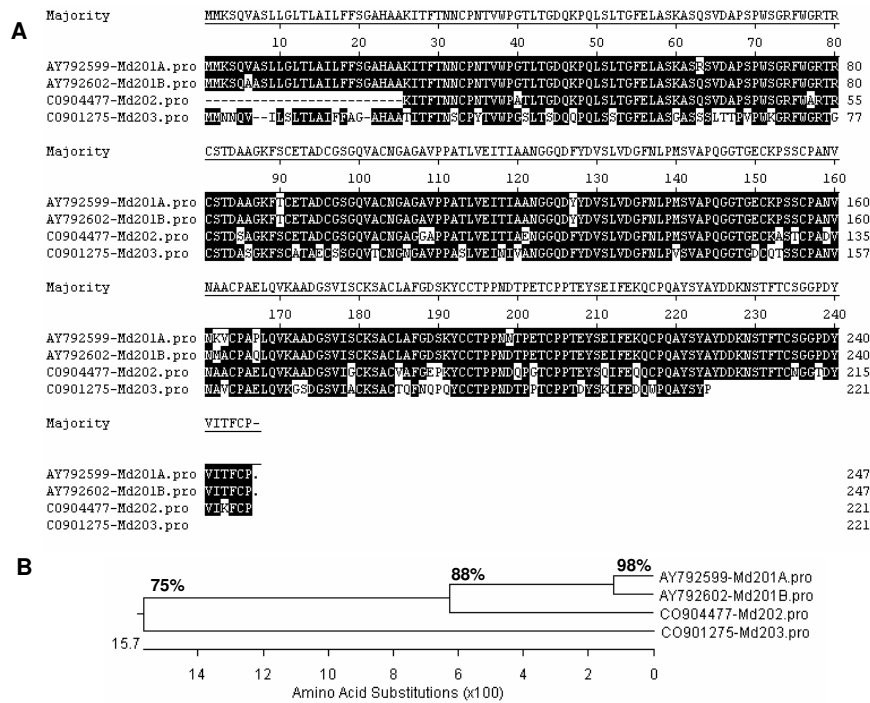


Figure 2: A) Alignment of the amino acid predicted sequences of Mal d 2.01A, Mal d 2.01B, Mal d 2.02 and Mal d 2.03 from the amino acid sequences AY92599, AY92602, CO904477 and CO901275, respectively. B) Phylogenetic tree of the TLP sequences with the percentage of homology.

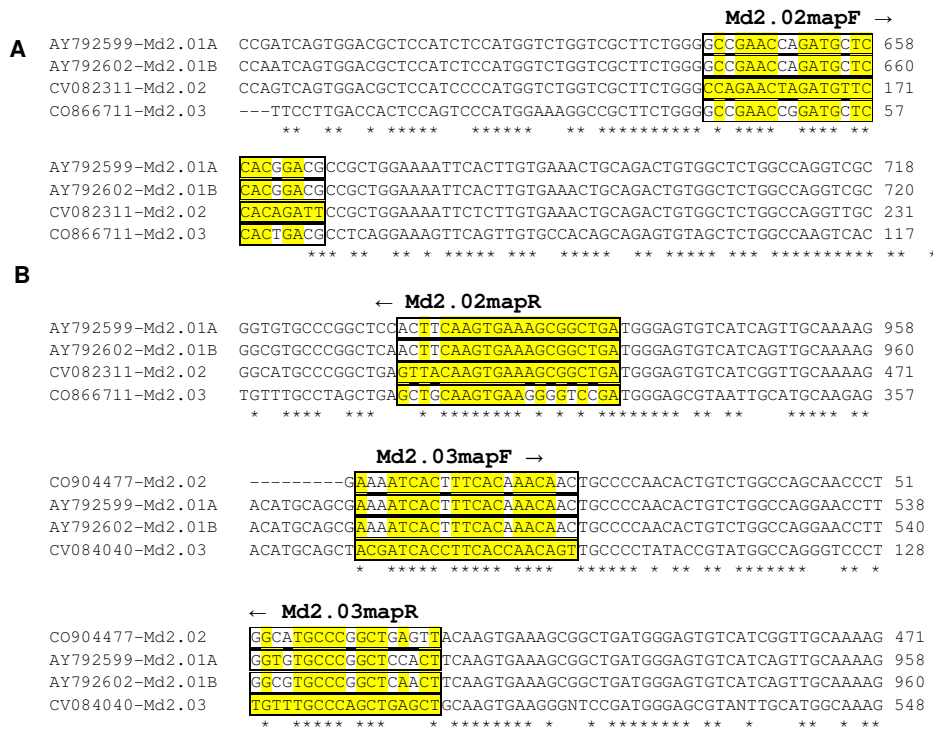


Figure 3 : Representation of the sequence variability in the *Mal d 2.02* (A) and *Mal d 2.03* (B) primer regions. The primer regions are indicated with boxes and the identical nucleotides among sequences are highlighted in yellow.

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 (Maliepaard 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 LG4 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 LG4, no corresponding genes have been reported up to now in the homeologous LG 12.

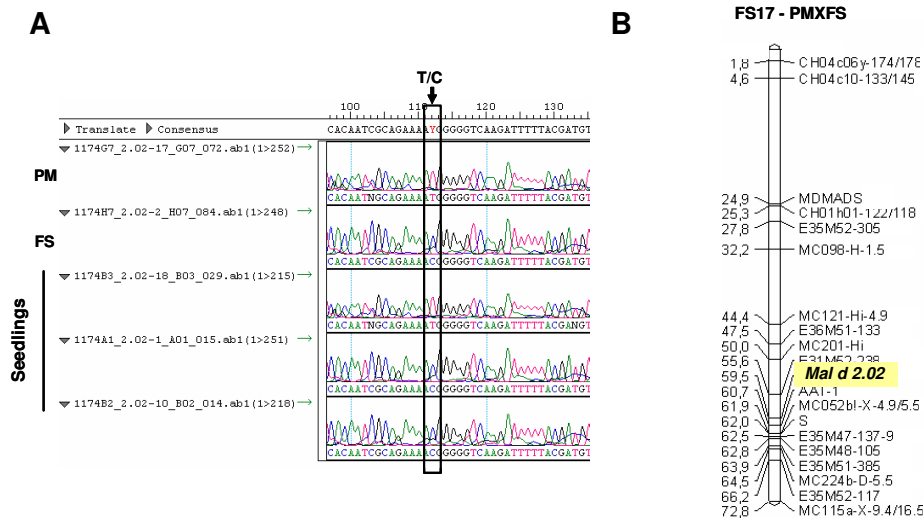


Figure 4: A) Portions of the two sequence chromatograms of Prima and Fiesta obtained with direct sequencing and three sequence chromatograms of seedlings as examples. The T/C polymorphism is indicated within a box. B) LG 17 of Fiesta. The *Mal d 2.02* sequence position is highlighted in yellow.

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

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.

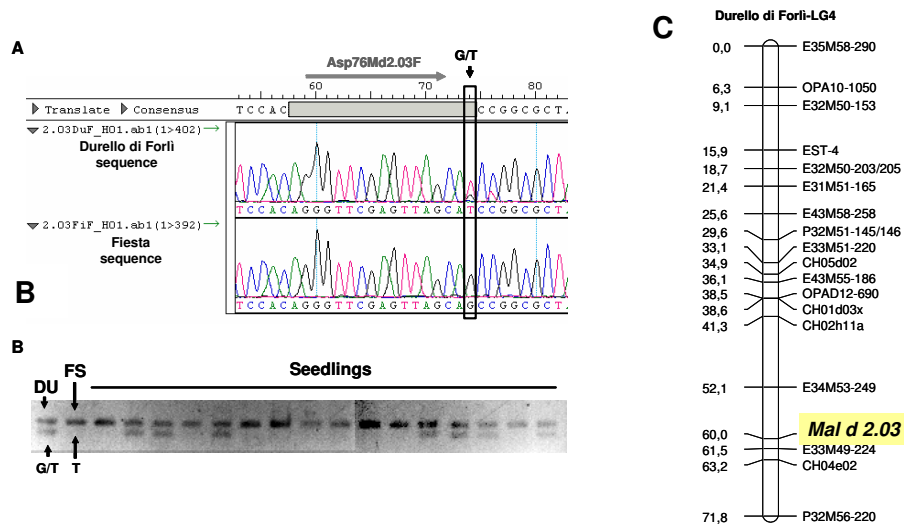


Figure 5: A) Portions of the two sequence chromatograms of Durello di Forli and Fiesta obtained with direct sequencing. The G/T polymorphism is indicated with a box and the allele specific primer used for the TSP method is indicated with a light gray box and a light gray arrow. B) Screening of the two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

Gene	Reference genomic seq	EST	Tissue	Total lenght	Coding	Ex1	Int1	Ex2	Int2	Ex3	LG
<i>Mal d 4.01</i>	AY792606	CN488488, Co068373, CN993251, CO755896	Fruit, bud, flower	2.017	396	123	343	138	1.278	135	9
<i>Mal d 4.02</i>	AY792610	CV084691, CN492781, CO066892, CO756641, CO756641, CN491268, CO899800	Fuit, bud, flower	862/863	396	123	248	138	218/219	135	2
<i>Mal d 4.03</i>	AY792614	CO755279, CO867336, CO867114, CO051739, CV086318, CO415968, CN494198, CO067937	Fruit, bud, flower	1.030 /1.032/ 1.033	396	123	386/388 /389	138	248	135	8
<i>Mal d 4.04</i>	from the BAC clone	CO066117, CN492105, CN996826	Flower	1.120	402	129	531	138	187	135	2

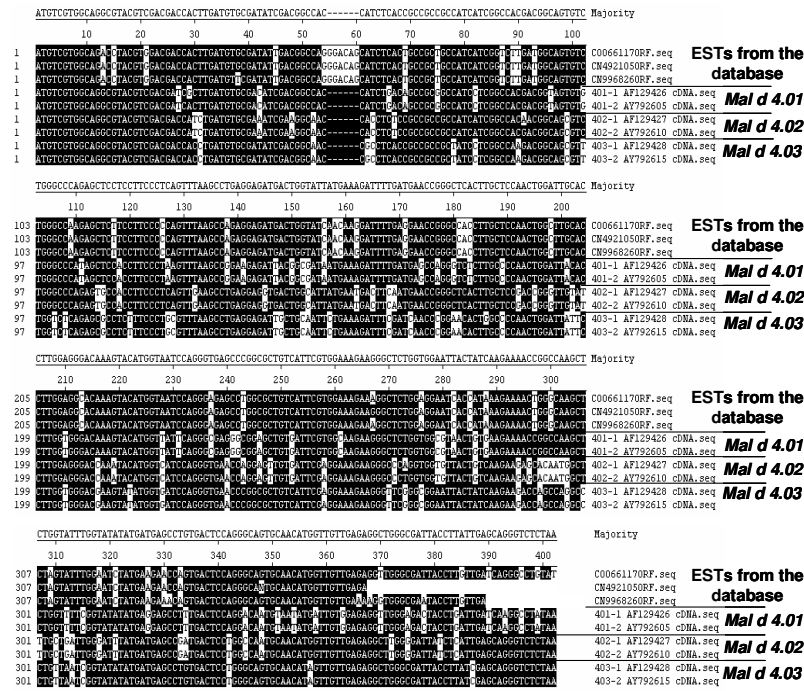


Figure 6: Alignment of apple profilin nucleotide sequences: AF129426/AY792605 for *Mal d 4.01*; AF129427 /AY792610 for *Mal d 4.02*; AF129428/AY792615 for *Mal d 4.03* and the ESTs CO066117, CN492105 and CN996826.

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.

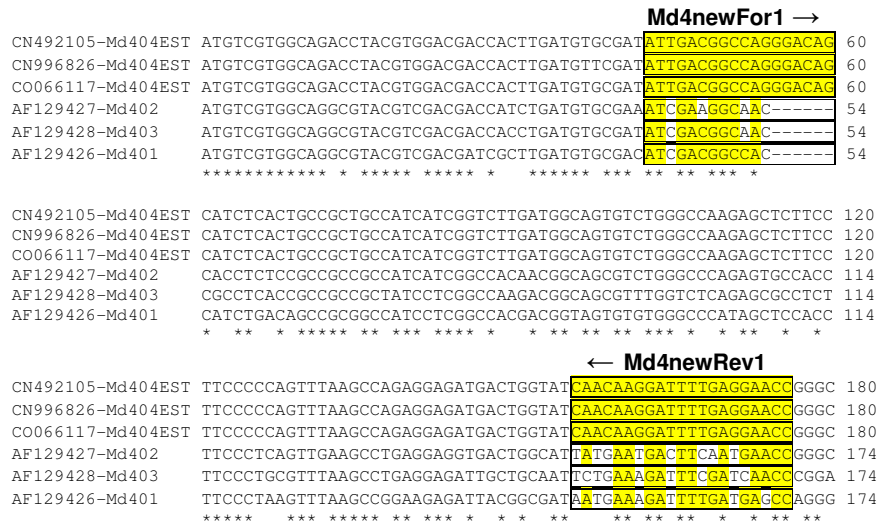


Figure 7: Representation of the sequence variability in the *Mal d 4.04* primer regions. The primer regions are indicated with boxes and the identical nucleotides among sequences are highlighted in yellow.

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.

```

TGTCCTCCCTCCCTCCACATTTCTCTCTCTCCTCTGTTTATAAATTTGCTTTCTCAAACCAACCCCAATGGAGTCCAAACCCAA
TATTAATAACTCCACCAACAAACAATCACCTCTTCTCAGAAAAATTCTCACAGCTATCTGTACACCAACAACAAGAACACA

                                  ← Md4newREV-STEP3
AGCATAACACAAGGCAGACTCTTGCCTACAACATCAGAGGTAATCGAAAAATGTCGTGGCAGACCTACGTGGACGACACT
T                                  M S W Q T Y V D H L

Md4newFor1 →                                  ← Md4newREV-STEP2
GATGTGCGATATTGACGGCCAGGGACAGCATCTCACTGCCGCTGCCATCATCGGTCTTGATGGCAGTGTCTGGGCCAAGAGC
M C D I D G Q G Q H L T A A A I I G L D G S V W A K S
TCTTCCTTCCCCAGGTCCAACCCTACCTNTTACCTTTCACTATTGCGTGCCTTCCCTTGTGCTTTCTTTTGGACTTGTGTTG
S S F P Q

CTAGATAATATTTGGTTATATCGTCACACTGTGAATAGAGACCATCTCGAAGCTTGCATAGTCATGATCAACAATCCATGTTTAC
AATCTGACAACATAGTATATTATCATCGTTTTTGTGTATGTTGTTGGTTC AACATAGTTGCATTACATGCTTTACTTTCATCTG
ATCCATACATTCAAAATATCATAAGATATTGATTTCTTATTGCCACATAATATATGATTATATTGCTTAATGTGAAGCGGACACA
CATTCAAGCATATTTTGTGATGATCAATAGTATGTGCTCACAAATTTGACAACGTAATATACTTATATGGCCGCTCCTATTGTATGC
TGGTTGGTGTCAATGTAGCTGATTGCACATGCAGCATTGTGAACCTGTGGTTGATCTTTTCATACAGAAATGCATGTGCTTG

                                  Md4newFOR-STEP2 →                                  ← Md4newRev1
TGGATTAATTAGGGTTCATTATAATGCAGTTTAAGCCAGAGGAGATGACTGGTATCAACAAGGATTTTGAGGAACGGGCCAC
L F K P E E M T G I N K D F E E P G
CTTGCTCCAACTGGCTTGCACCTTGGAGGCACAAAGTACATGGTAAATCCAGGGAGAGCCTGGCGCTGTCATTTCGTGGAAGAA
H L A P T G L H L G G T K Y M V I Q G E P G A V I R G K
GGTACGAGCCTCCCATATTGTCGATAATGTATCAATAACGTATCAATACATGTTGTAACCTATAATGATGTCAACACATCATCGAT
TTCATATCAACAATTTGTTCTTTTGTCCGTTCCATAATCGGACTACACATACTTATTGCACAACCTTTAAACTAAGATCTCCTTTGT

                                  Md4newFOR-STEP3 →
AATTGGTTTTTTAGGGCTCTGGAGGAATCACCATAAAGAAAACTGGCAAGCTCTAGTATTTGGAATCTATGAAGAACCAGTG
K G S G G I T I K K T G Q A L V F G I Y E E P
ACTCCAGGGCAGTGCAACATGGTTGTTGAGAGTTGGGCGATTACCTTGTGATCAGGGCCCTTAGGCTAAATAATATTCTAC
C V T P G Q C N M V V E S W A I T L L I R A
ATGTTATGATCTCCTAGCATTGATTTTGGATCTTATTTTCTTTTGTGTTTGGGGTTTTGGTCCCAGTTTGCATCTCCTTCATC

Md4newFOR-STEP4 →
CAAAACCTAAAACTGGTGGTGAAGTTGTTGCAGACAAACCATGTGACAAGGACTAAAAGCAGTTTGATGTTTGAATTGTTT
GAATATTGCATAAACTCTATATTTCATGATATATTTTGTATCCTTTTATCCCTCCGTTTCCCTTAATCTCTTGTGGTTTTTCT
TAAAGAATCTTTATCTGGCAAAAGCTTGCCAGCACATATTGAATCGCTGGTTGTAATTTATGATTTGGTTTTTATTTTCTCA
CAAAAGTCATGAGATGTGCTTGACATGTTTGGTACAATAATGAATCAAATACTAGTTGCATTGCATATAAAGAAGGACGTGTCG
ACGAGATGAAAGAACTTATTACAACATAAATATTATTTAATAATATCTTAAATTAATAATATGCGTCAACCATGTTAATNTTT
GATATAACAATACGTCATAGGCTTAATATATTGCAGATGTGATATCCCGATTGCATAAAACAATTCTTAGTCATTGATGCACTTAC
AAGTCACATTGCTAGTTGGTAACCAATGGTCTTCAAAGAATAAATCCTCCATTTTGGCGTTTGATTAATGGTTTTGCTGGCAAG
GGAATATTTATCATCATTTTTNATTTAATGGTGGGAAAATCCACCATCATGGATTGCCATG

```

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 shwed 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).

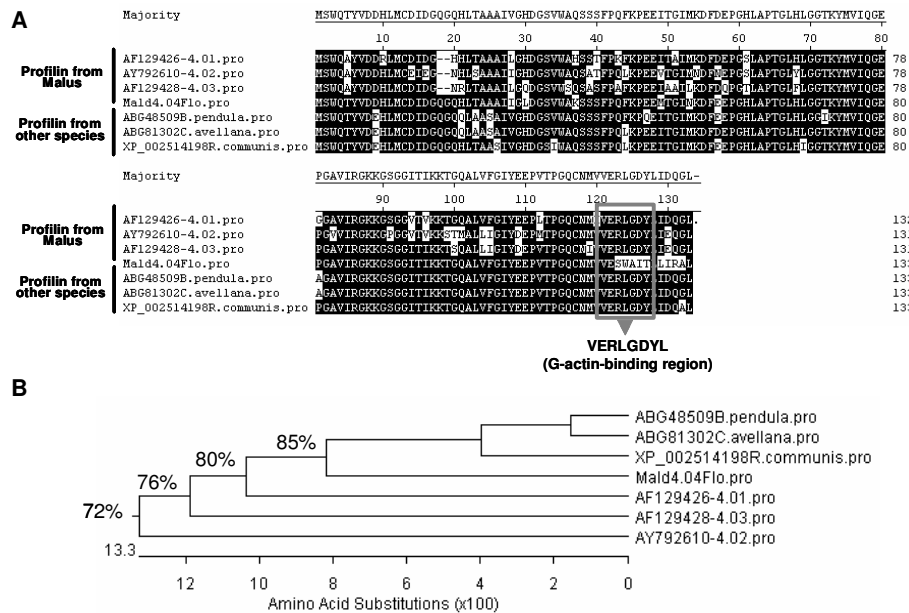


Figure 9: A) Alignment of the amino acid predicted sequences of *Mal d 4.04* from the BAC library, *Mal d 4.01* (AF129426), *Mal d 4.02* (AF129427), *Mal d 4.03* (AF129428), profilin from *R. communis* (XP_002514198), profilin from *C. avellana* (ABG81302) and profilin from *B. pendula* (ABG48509). B) Phylogenetic tree of the profilin sequences with the percentage of homology.

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.

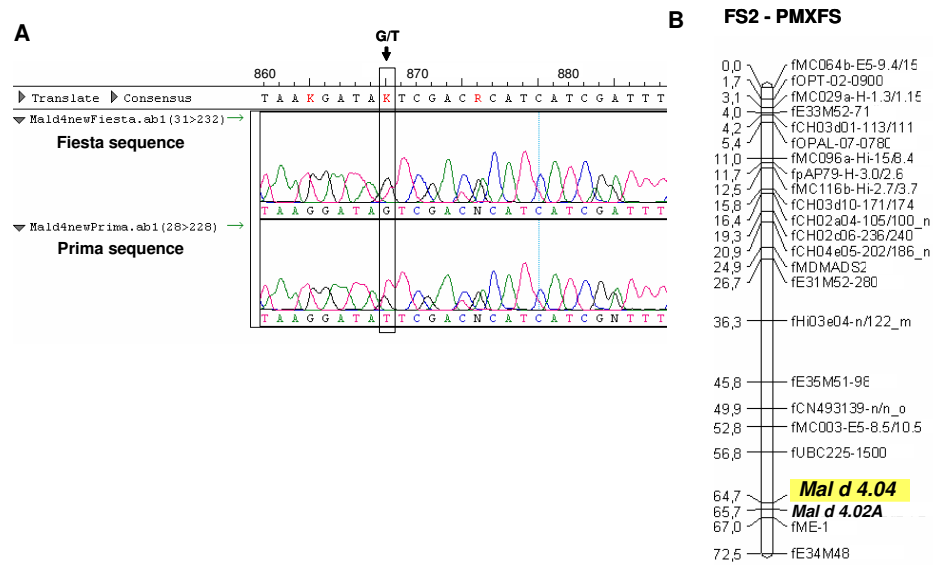


Figure 10: A) Portions of the two sequence chromatograms of Prima and Fiesta obtained with direct sequencing. The G/T polymorphism is indicated with a box. B) LG 2 of Fiesta. The Mal d 4.04 sequence position is highlighted in yellow.

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 *nel* 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 a 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 all the members 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

- Bolhaar ST, Zuidmeer L, Ma Y, Ferreira F, Bruijnzeel-Koomen CA, Hoffmann-Sommergruber K, van Ree R, Knulst AC. 2005. A mutant of the major apple allergen, Mal d 1, demonstrating hypo-allergenicity in the target organ by double-blind placebo-controlled food challenge. *Clin Exp Allergy*. 35:1638-44.
- Breiteneder H, Ebner C. 2000. Molecular and biochemical classification of plant-derived food allergens. *J Allergy Clin Immunol*. 106:27-36. Review.
- Breiteneder H. 2004. Thaumatin-like proteins – a new family of pollen and fruit allergens. *Allergy*; 59: 479–481.
- Fernandez-Rivas M, Bolhaar S, Gonzalez- Mancebo E, Asero R, van Leeuwen A, Bohle B, Ma Y, Ebner C, Rigby N, Sancho AI, Miles S, Zuidmeer L, Knulst A, Breiteneder H, Mills C, Hoffmann-Sommergruber K, van Ree R. 2006. Apple allergy across Europe: how allergen sensitization profiles determine the clinical expression of allergies to plant foods. *J Allergy Clin Immunol* ; 118: 481–488.
- Gao Z, van de Weg EW, Matos CI, Arens P, Bolhaar ST, Knulst AC, Li Y, Hoffmann-Sommergruber K, Gilissen LJ. 2008. Assessment of allelic diversity in intron-containing Mal d 1 genes and their association to apple allergenicity. *BMC Plant Biol*. 8:116.
- Gao ZS, Weg WE, Schaart JG, Arkel G, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJ. 2005. Genomic characterization and linkage mapping of the apple allergen genes Mal d 2 (thaumatin-like protein) and Mal d 4 (profilin). *Theor Appl Genet*; 111: 1087–1097.
- Hsieh, L. S.; Moos, M.; Lin, Y. 1995. Characterization of apple 18 and 31 kDa allergens by microsequencing and evaluation of their content during storage and ripening. *J. Allergy Clin. Immunol*. 96, 960–970.
- Krebitz, M.; Wagner, B.; Ferreira, F.; Peterbauer, C.; Campillo, N.; Witty, M.; Kolarich, D.; Steinkellner, H.; Scheiner, O.; Breiteneder, H. 2003. Plant-based heterologous expression of Mal d 2, a thaumatin-like protein and allergen of apple (*Malus domestica*) and its characterization as an antifungal protein. *J. Mol. Biol*. 329, 721–730.
- Marzban G, 1, Puehringer H, Dey R, Brynda S, Ma Y, Martinelli A, Zaccarini M, van der Weg E, Housley Z, Kolarich D, Altmann F and Laimer M. 2005. Localisation and distribution of the major allergens in apple fruits. *Plant Science* 169, I2: 387-394.
- McKenna, S. T.; Vidali, L.; Hepler, P. K. Profilin inhibits pollen tube growth through Actin-binding, but not poly-L-proline-binding. *Planta* 2004, 218, 906–915.
- McKinney, E. C.; Kandasamy, M. K.; Meagher, R. B. Small changes in the regulation of one Arabidopsis profilin isovariant, PRF1, alter seedling development. *Plant Cell* 2001, 13, 1179–1191.
- Oh, D. H.; Song, K. J.; Shin, Y. U.; Chung, W. I. Isolation of a cDNA encoding a 31-kDa, pathogenesis-related 5/thaumatin-like (PR5/TL) protein abundantly expressed in apple fruit (*Malus domestica* cv. Fuji). *Biosci., Biotechnol., Biochem*. 2000, 64, 355– 362.
- Pagliarani G., Paris R., Tartarini S., Sansavini S. Cloning and expression of the major allergen genes in apple fruit. *JHSB ISAFRUIT Special Issue* 176–181.
- Radauer C, Hoffmann-Sommergruber K. 2004. Profilins. In: Mills ENC, Shewry PR (eds) *Plant food allergens*. Blackwell, Oxford, pp 105–124.
- Ramachandran, S.; Christensen, H. E.; Ishimaru, Y.; Dong, C. H.; Chao-Ming, W.; Cleary, A. L.; Chua, N. H. Profilin plays a role in cell elongation, cell shape maintenance, and flowering in Arabidopsis. *Plant Physiol*. 2000, 124, 1637–1647.
- Scheurer, S.; Wangorsch, A.; Nerkamp, J.; Skov, P. S.; Ballmer-Weber, B. K.; Wu"thrich, B.; Hausteiner, D.; Vieths, S. 2001. Crossreactivity within the profilin panallergen family investigated by comparison of recombinant profilins from pear (*Pyr c 4*), cherry (*Pru av 4*), and celery (*Api g 4*) with birch pollen profilin Bet v 2. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci*. 756, 315–325.
- van Loon L.C., Rep M., Pieterse C.M. 2006 - Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol.*, 44, 135-62.
- van Ree, R.; Fernandez-Rivas, M.; Cuevas, M.; van Wijngaarden, M.; Aalberse, R. C. 1995. Pollen-related allergy to peach and apple: An important role for profilin. *J. Allergy Clin. Immunol*. 95, 726–734.

- Asero, R.; Mistrello, G.; Roncarolo, D.; Amato, S.; Zononi, D.; Barocci, F.; Caldironi, G. 2003. Detection of clinical markers of sensitization to profilin in patients allergic to plant-derived foods. *J. Allergy Clin. Immunol.* 112, 427–432.
- Westphal, S.; Kempf, W.; Foetisch, K.; Retzek, M.; Vieths, S.; Scheurer, S. 2004. Tomato profilin Lyc e 1: IgE cross-reactivity and allergenic potency. *Allergy* 59: 526–532.
- Venisse JP, Malnoy M, Faize M, Paulin JP, Brisset MN. 2002. Modulation of defense responses of *Malus* spp. during compatible and incompatible interactions with *Erwinia amylovora*. *Mol Plant Microbe Interact* 15:1204–1212.
- Hoffmann-Sommergruber K; SAFE consortium. 2005. The SAFE project: 'plant food allergies: field to table strategies for reducing their incidence in Europe' an EC-funded study. *Allergy*. 60:436-42. Review.
- Vieths S, Scheurer S, Ballmer-Weber B. 2002. Current understanding of cross-reactivity of food allergens and pollen. *Ann N Y Acad Sci.* 964:47-68. Review.
- Chen L, Zhang S, Illa E, Song L, Wu S, Howad W, Arús P, van de Weg E, Chen K, Gao Z. 2008. Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genomics.* 17:9:543.
- Liu JJ, Sturrock R, Ekramoddoullah AK. 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Rep.* [Epub ahead of print].
- King TP, Hoffman D, Lowenstein H, Marsh David G, Platts Mills Thomas AE, Thomas W. 1995. Allergen nomenclature. *J Allergy Clin Immunol* 96:5–14.

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 *Mal d 1* gene cluster evolution. In fact, in this work the hypothesis of the duplication of *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 *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 *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 *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 *Mal d 1* genes. The possibility to distinguish the level of expression of each *Mal d 1* gene expressed in fruits (10/20) may be an useful tool to find out *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 *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 *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 *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 allergenicity, an higher or lower level of gene expression in fruits led to identify some *Mal d 1* genes putatively involved in the determination of allergenicity among cultivars. At this regard, the gene expression of *Mal d 1.06A* and *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 *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 *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 *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 recombinant 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.

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 fruit 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 (Thaumatin-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 Thaumatin-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 hypothesized 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 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 allergenicity, 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.

Abbreviations

BAC	Bacterial artificial chromosome
Bet v	<i>Betula verrucosa</i>
Ct	Threshold cycles
EST	Expressed sequence tag
Fra a	<i>Fragaria ananassa</i>
IgE	Immunoglobuline E
LG	Linkage group
LTP	Lipid transfer protein
Mal d	<i>Malus domestica</i>
MAS	Marker assisted selection
OAS	Oral allergy syndrome
PR	Pathogenesis Related
Pru ar	<i>Prunus armeniaca</i>
Pru av	<i>Prunus avium</i>
Pru d	<i>Prunus domestica</i>
Pru p	<i>Prunus persica</i>
Pyr c	<i>Pyrus communis</i>
qPCR	Quantitative PCR
QTL	Quantitative trait loci
R-gene	Resistance gene
RNAi	RNA interference
Rub i	<i>Rubus ideaus</i>
SAR	Systemic acquired resistance.
SNP	Single nucleotide polymorphism
SPT	Skin prick test
SSR	Simple sequence repeat
TLP	Thaumetidine-like protein
TMV	Tobacco mosaic virus

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

1 aagtgcagacatattaccatggcgggggtattacttggcagctacgcaatt
attgagtttctgggattctcagatatacggccaatcatatcccaggggat
ccaacttagcggctaacgagatggcacagctagcctcaggagtgccaata
catgagagaaaatattgggatagatgtcgagatccaaagaagaaccttcc
ttccatcttagaaaggggattcagctagatgtaattggtgctaaaggccg
agatagaagattggaggtcacctatcatcatcatttgaaggatccttct
tcgcctacaagcaagaatgatagacagcaagcaaccaagatgtcctatg
ggcaaagaacttgctaagaaaaactccatagggttactggtgaaatgct
tggggcaggaagaatccatgagggtaattggccaagtacatgaatgagta
tgtggagcacatcaggctggaacgaagatgagatggttgcttaggaggt
501 tggttattctggcccaacatggaaaaagattgcaagctctatgcccag
gttgtgaagaatgtcaaaggcatggacctctccagcatgtgcccttggt
cctttaaatccagtggtcaagccttggcccttcagaggatggcgatgga
cttcatcgggcaaatatccggcttctagtaaagggcatacctcataa
ttgtagcaacggattacttccaagtgggtggaagcatcagcagtaaaa
tcataaacttcagccacagtcaagaatttatcgagaccaagattttgca
catatttggggtgccgaaactatagtacggatcgtggaccatctttta
tttcaaaagaagttgaggagtttgaagcaagtaaaaaaagatgatc
cagtcagtcgactaccctcaatcaatgaccaggcagaagccagcaa
caagatcctgggtcaacattatcaagaggatgggtattgatagtcagaaa
1001 aatggcatgaaaaactggggaacacttgtgggcatacagaacttctaag
agagcaggaacagggacaactccttatgctttaacttctgggcaagatgc
aatgctcccctggaaatcaatgtaagttctgtaagaattcaaaatcaat
ttggattgcatagtgaaagatatacgaagccatgtgtcaaggaattgaa
gacttggatgccgccgaattgaagcttgaaccagattcaagaaggaaa
gagagtcgttggcccgacttatacaaaaagggtgaagatgaaatcttca
aagaaggggatttagtatggaagacagtcctccctcaggagctcagctc
aggggcttggaaaaatggagccgacatgggaaggctccttctattatag
tcgcttttggatagagggggataacttggcggaccttgagggaaaaca
ggcataaaactcccattaatgttaatttctaaagaaatattgtcctacg
1501 ttgtgggatctcatagattgttatattgaagaggatgcagggtaaagtaa
gcttcgggatataatgtcagtgatatttactaatcattcaagaagat
ggaaagacaagagttgctgagataatgtatatttcttgcacaaattga
tgaaatttcaaaaagattcaaagctgacgtattacaatcaattctcctgc
tggatgcaatggtatcttcagctgatctccgatgtctcatggatagaa
cccaatcaggtgatcgggttgtgcccgaacatgagcctggtagaggatc
ctcagacaggaccatcacataggtgatggtcaagcccgacttatacca
caataagcaggaagacaagccttcgagaggaaaccgctgaccaagggtg
ttggagatagcgggggtgttcgaccaagggtgttagagatagcagtccttt
gatgaaagcccttttctcagaaaaggagtttaggaaaccccactca
2001 aagcctaaggaacccatttctcagtttctgaacgctcgaggaagatggaa
gagaatgctgaagaggatttgatggcttgacagaagaaattccaggccag
aatttatagggcccaagggtagttcaaggctcgtgagaagcaaggg
gcacggggttgccctcatctgtctagaaaacacgaagtccaaaagatt
gagatacacgcatgcggatattcaatcaatgatggcgccctgggattccaa
ccttaacataattgaaggggactgttggatcaaaacttgaaacttt
gggctcaagaaaggagctagcccaaaaggaggcccgaaagggaagataac
cgaagcccagtcaaagcccagaaggcagaaatggcctttctgacttgat
gcagctcatgaagaccacttgettacctaaccaaaaggccaagtggatag
gctgaccagctacacagcccataaaaagtaactttatgatggcagtaacgt
2501 aaaaaaactgatgagtcacaccttcaaacccgattcgggcaagattgc
tgctacaggagaccaagccgagatgtctatataagaagaagagaacca
ggaataaggacactcaatcaaacaaacaaatgcaagcacaacactctgccc

Retrotransposon protein,
putative (ORF1) ↓

aagccagatttgccttcaaacgaagttgtagtcagcccaagcattcat
 cccttgccgggataaaccttcaactcaactcttgaatagctctgctacct
 tgttcaaccttgctgtatcgattcatttttgtaaatctctctccctc
 cctgaagcttccatcccttgataaaccacaaagatatgaacctgcaaga
 cgaccaagcttgcccgaaggttaaaccttgcccagcttctttgtttt
 tgtcttttcattcattagcctagcaatatattgtatatttccagttatat
 tcaagtagttatttctagcaaatgactattaaaaggctttctcaatact
 3001 tggatccgatttgaatatatcttcagtggtcaaaccagatccaagtccta
 agccctcgggcatgtaaatctgattagttaaaagtccttagcctcaaggc
 ataaaaagaaccttatgaggacttaacccatccacgacaatccttgata
 aacggaagtcogaagttacttggggcgaagtaatcgacctacctcta
 gtcttattgctattacattatgattatcatagaacgcttaagcatgggaa
 ggatcctgataggaagcctcaaggcctacacctaagccctacgaaggca
 cctatttggattcattccgttctgcggtctaggacgtttgagtataagaa
 cgaactctaatgggaagcctcaaggcctacacctaaggccccacgaaggc
 acctatttgggttcattctacctctcgaactcggataatcagaagataaa
 tagttataaagactcatgcattcacgagaacaaatgatgtgttcgagc
 3501 actggtttagttattgataggaagcctcaaggcctacacctaaggcccca
 caaaggcactattataactaacacggtttctcggggatatacatataca
 accaaaactcgacatccatttcatatctgccataactaaagatggcagtg
 cagcctgagcacttaaaagttaaccttgattgtgagcctcaaggcctac
 atctaagacccccacaaaggcaacctctaaagttaaactctgtccttctct
 tcagcattgcccgaagccttgcccgaagcagcacttgcccgaacaaagccc
 gatagtgaaagcagaagcccgaagcagcctcaaccagcaccacacctccc
 ggggagctgtgtgctcgtcggccaggaacatccccgacggaattcctg
 acgcaaacactttctttagtccttttatttttaattatatttctttg
 ctttcttaaatattttgtgtgctcagtttcaatagttagtggtttttt
 4001 ccatttaataagcttgctcgcattaaaattaatgggatataatagttaat
 atgtacttgggttttttaattatttttatttttattttatattagtgtaacg
 atgagtagattgaatttgtatagatgctcttcatagatgagttttcttt
 caaacctttatataatgaataaactttatgatattagatttattgatgttc
 cagtaattggatgggactgtttaaactatcagttttgtatgcttcatt
 tttattagataaattagattttcttttgatccatgtgcttatgtctatac
 agataaagagagcaattagtgaggttcaagcattaggagcaatatttcac
 tttcaaggaacttcatttggaggtgtctaaagaaaagtgaagtatttcaa
 gtgacaatttcttgtttttaaaggcataatttggtaattcatgtgtagtt
 4501 taaatcacgggtggcgtgttaactttgcacctcttgtaatacaatttgg
 tattcttagtataaattattgggtgggtgtgttagctttgtacatctgtg
 atacgttttgactatagattattcatgaaatattgtattatagtaaatg
 tgcaattttgtgtttctttatataattttaaattaatttttatttttt
 agaaaacattttataacaggtttatgtgtccacgagcactacaaaataa
 aagctcatctaccatgataattttatcacgggctttatgtgtccgttata
 aaaactacaaatttatgacgggcataagaagcagctgataaataacaatt
 ttataccacagacaaatttagtctgttgtaaaatagcagtgaaatgatgat
 ggcacaaagtgtacattgtaatttactaagccaccatagataatgctca
 tgggtcaatactgtctgtatccatagcgcctttttgttatctacctcgga
 aaaatgtgtggtattgtacatgcttaattgaaaattttggaggaggctgg
 5001 acacccaaattttgcaaaatttcaaacctatgtaattatctgtattattta
 agaaaataaaaactgaaacaataatcgtttagaaccagataaacaccaca
 ccacatccatcgtgcaaaactagaccagtaaaagaaatcgaagtagtttt
 ctaaacgtatgctcatgaagtgaattaaattctgaaaccactatccac
 ttctcccacgacagtttctccacaaaggaaaagctatgttgtttattcg
 atcctctcacgtactttttttatttgcaatttttagttttatctggtttatt
 ttattttatcgattagattcctocattactttgggtattaaattctgcaatt
 gttttgttgttttgggttcaactgtatttctgactctccattatgtttatt
 aggggttttgataaaagctagatttcttgtttcatcgactttggggtagaa
 ttgtacaatggctggccagtcgaagagtgaaatccacctacgatttcaactg
 5501 gcaaggttggtttttgacaaaaattatttttcttttccatctactc
 tacaatttgaagttttggttaaattttttgaacattaacaatcgaacgc
 tctcatattttgttttttaggtttactgggtgtatcagcaaatcttcca
 tctatttactcctcctttgttaaataatctctataatttaggtttgatt
 tttatgcttcaggattcgtagtttacaatatattcttttaaatttctagg
 ttaagattggctgataatgctttctccgattgagttggctggcgcagag
 accaaattggagcggctcctccttggagggaatctgctagaattctttt
 ccgcaaatttctttcttttagtccttttatttttttaattattatctt
 tgctttcttaaatattttgtgtgctcagtttcaatagtttagtggtttt
 ttcccatttaataagcttgctcgcattaaaattaatgggatataatagtta
 6001 atatgtacttgggttttttaattatttttttttttatttagtgtaaat

Retrotransposon
 containing
 Reverse transcriptase (RNA-
 dependent DNA polymerase);
 putative NTP binding site;
 RNase H; RNA/DNA hybrid
 binding site; integrase
 core domain; CHRromatician
 Organization Modifier
 domain (CHROMO domain)
 (ORF3) ↓

cgatgagttgattgaatttgtacatgatgctcttcatagatgagtttct
ttcaaactttatatatgaataaactttatgatattagattttattgatgt
tccagtaattggatgggactgtttaaactatcagttttgttatgcttca
ttttattagataaattagattttctttgatccatgtgcttatgtctat
6501 acagataaagagcaattagtgggttcaagcattaggagcaatttc
actctcaaaggacttcatttggagggtgtctaaagaaaaggaagtattc
aagtgacaatttccttgtttttaaaggcatatttggtaattcatgtgtag
tttaaatacagggtggcgttgttaaccttgcacctctgtaatacaattt
ggattcttagtataaatcattgggtgtttagctttgtacatcttg
tgatcagtttggactatatgattattcatgaaatattgtattatagtaaa
tgtgcaattttgttctttatatatttttaattaatttttatttatt
ttgaaaaacattttataacaggcattgatcatggcatttagtaatccac
aacggctgtagcgcattgtttaaactctaatccatcaggggactgtcc
gtgattatagtaattaaacacttcttgcctggtaaaagttatga
ctttttatcacgtctagatacttaacgggtatattgtccatgggtaattg
caatttaccacgggcattcatcgtgaaagatgtgggttcttccagtg
gagccatgtgctcaataaaacaaagaatataatacatcttctgctcc
taataagttagtaggatttgaatgccaatgctttcaagatggaaaag
7001 aaagtgtcactactcctgttctgtctacagattttagggtat
tgattaactatgttcttctcagtcctattcattgttattcaacc
ctacaatgactgccatcttacagaattgataagccattgttctcaat
ataaaccctcttaataatcagttcaaaattcaaataaataaaaaatc
ccaacaatttgggcttgaaggccaatccagaactaaccgaaatcccca
accccaaaaattaggaattcttaacttctggttctgatttctggttttaga
atcattctccgaaaatttattgattgggatttgggtcgaattcctaccg
aaccggcctagtgaaccctaatattttatctactataatattggacaat
accattaaaagctgttctatataaacactcttcaagtcatgact
ctttgaattaaaaatttcccttttataaatttttagattgtcaatgtt
7501 gattcccttttatcaatttagatttcttgttatttttggatgtat
ttgttgatttggcccctaaacttgtatagagggtcaatttctctataa
acttttaatttagcagattacctccatgaatttttataattagccaatt
cccacatgaacttttaacttttagccgattaaaccctaaaatattatagag
taaattacacaaaactatctcaactataggtcaaaccacaatctcatacc
ttatggtttaaaaattgggagtttaacgaaaaactacatttttatacta
aaaagtcaaacatggattattcattttatcctttattttgtccttca
ttaaactcaagtttcaagccatttcttagttttctttaaataat
acaatgtcatacctcatcttacgaatttattgcagtgtagacctcgtc
aaattttctatcaattttctattaaatgttgacgtgacttaacaataa
8001 cccactccctaccgactaaataaaaaattaaataataaaatta
agaattaaaagcgtgggatccacccccacaaaccatcattcttcccca
ttacaccgacgacatcccatctccaggcaccacactcaccacctgctt
ccccagtacgcccgcctcctcctcctcaaacccccagacagcagaagc
caccacgtcgtcatgctcatcctcctcggcttcttcaagggctgctgct
gcccgtctctcttactcttctcctcctcctagcactcctcgtcctcgccgtc
gtgctcgtcatcatacctccgctcagtcctcctcctcctcctcagc
agtacgtgctcttcaaccgtaaccccaacaaggtcgggatcaagta
cgacgagtcagggttactgtcatgtaccttgagatttcatgggcaaggc
ctccatcccggcttctaccaggacgcccagactatcgccaagtctcg
ccaccatcgccatcaatcgagtcattttctgacgctgctaatttagtct
8501 ggcagcttccctcaatgtgttctgcatcaaaaccacttcccaattt
ttctttgaattctagttcggctactaagaaaatcacagaaccactatct
tcggtgggggagagttcttggagcttgggttcttccagaagagaaagggt
tgaagtggaggtgaaagaaggggaggagcaacggaggagggggaggt
cgaaggtgtggcggaggaggttgaggtggaggcaggtggtgaggttgggt
accggagatgggatgtctcgatgtaccggggaagaaggatgggttgtg
gggtaggtcccacgtttttaaataaaaaattaaatgatttaattttt
aatattttaaattgatttaattttttaaatttttaatttaatttaatt
atgtaccctttctagcatttctcctagaattttacatgtcaaatggtga
gtgaaaaaataacagaaaagataactagcatttctcttatctccaaca
9001 agtcacgcaccatttacttgatcaacgcataatcacatagagactactgc
agggttagtctaataagaccttttgaacaaggaagtaaatccatgccct
accatcaagtatgagttctgtcttataggaccaattggaagtgttttct
aacatgtatgcaaagtgtgggaatttgaagcagaatgcttatgaagta
gaatcccattgttcataagcatttctcaatttctcagggaaaagaagatg
aagatggtgtagttttttaaatttaatttatttttttattttttct
ataaagtaataaataatttctattatttgaagcattagttatgagggt
aagttgtcaatttattagcaaatgttgactgaaggggctatgcacaagt
gaattaaaataaagagggtgttaattgtaattttcaaacgtgaggagat

9501 tttgtgaaaacttgtaaacttctagaggtgttattgttaattaacccaac
 cattgcaacatatttatcacgtctgcttttccgaaatttcctgtgtgctt
 ctctctttgtcactttcaacatgaattttcaatataattgaaaaaccaca
 tcgatattagtttagttatggtataccaagtatatcgtcgtccaaaatt
 aaactttttatcttttcttctcaatcatttgaccaaacttcttgttctga
 acaagttagagctcaagttacataggagaacacgtctgcacatttgtaa
 ttttaagggggcctttattcaacgaattttgaggtgaaatttatgaaattg
 gaagggaaaaataataactccaaggtaaatcatcaaataatttacattt
 aactagtatatatgttttatttattgagattaaaatcaaatacaacaa
 ttgtcacagcccgtcccagagttactaaataccgaggacgtgaaattact
 aaaacgccttaaacgggattaaggtgcgtaaaatttggtatttgttttac
 10001 actaagatcctaaactagggttagaattatattagttatgtgtaatttaa
 tttgtgtttggacttaggtggggtcctaccccctaaatccctccccaaa
accggtgggatgtctctctcatctccctcacactctctctctcttct
 ttcatctcagtatcactctctctcttactctcgaactctctcaagctc
 tcggcaaacaccaagaacaatcacaatcctacaccaacgtccgatctaa
 gatcaccacgacactccttaggacctcacgagcacgggaataccagttt
 aggttaagttctatttcgagaaccctagttttaaaccaccccgtgtaagg
 cactgttcacgaacttaattttgggtgtgttttaggttaacaaagatca
 ccagctcttaggaagttccaaggaggtcggagtgctcgtttgaaaca
 aaatggacgtcgggatcgtcgggttcgagtttgccgaaattgaagaatt
 10501 ttgaaaggtttgatctgtgtgttttaggccttaaaccctccaacgtg
 atagtacatgtgaaatgcttcattttggtataaaatcggaagaaattggt
 tgaaaaacgaaggagaatagtgatttgaaatttccagaaaccggcga
 acagccgcccggcaccggcactcgccggagaagacagggaatctccgt
 caagtttgacggaatattccgacgccgttagttaacttaacggaaaccg
 ttagtttttaacggaatattccgggaatattcctaaccgcttactgta
 gccgtcagtgctggcacgtggctgcgctgggcccgcgctaggtccg
 tgccacgtcagggcgtggggcgcgtgaggactcaaaaattattttaa
 aaataggggatgatcctgaggttgttaggtcactgtggtatattcata
 tacccaatttgagcatcgtatgaagaattaattacctagtttggttagg
 11001 tgcggttaattgtgcgttaaatgattgttttaagttatttcacttctaggt
 ggaacatttaacaggacgagcacatccaggggctcaagggggtacga
 cccggcgacataaccagtgagtggtgctttctatatatatataacc
 tataactcgatttcccagaaatcaaatttaaatgaaaaatattgaa
 atgaaatgaaatgatgtgattgccatgcatagaatgttatggatatta
 tgaactgtcgtatgatgcatatgatgatggtgctgtggacgcacagg
 taagtttaattactgttatgttaggatgatgatattgagctcata
 tcctgaccatggtttagtgcttatagttaccgcacgcacgcctcgc
 cttggatccaagtagctggtcgtacagtcacgcggagtggtacga
 cgggcccagtcgtagagtgtagtgagattatgactggtgggtgacctag
 11501 gttattgtatcacagatgattgatgagagaagcactagagcgaatattacc
 atgagtcgttcagactacattaggtggtccgacttatgtgcagaaggcc
 ggacaggtcacgcggagtgactccggcagagagtgagattgatagatgtt
 gagctctaggttcaatcgttcagggctattagaggccctccgattgatta
 tttctttacctgattatattatgtaaatgcattcatactacactgttga
 aattggcatggtacattctttattgaaatctgttataagattgatgattga
 gacagttgagatataatgctataactattttctgggaaagtatacag
 gttttccaaaaggggtataaatgtggatttatgaaatgttttgaaaa
 gctttattttcgcccactcacgttttctgttttcccccctccaggttct
 agttgatagttgaggcgttggcctacgagactgcttcggcgttctga
 12001 cagactaaataaatgtaggattcacccgagggtgttgaattagttatg
 ttccctacttgactgcacctagatgcttatgctctgtttatgtgtgttag
 tacactcttatgcacatagaatgctcggttgtaaaataactgcaattagt
 gttttcgttgactcgtattttattttaaatcgtttccgcttgccgtttgg
 ttacgtcacactcacgtgacggccagcacgtcctagtcttcgggtaggg
 tgtgtcagtttggatcagagcataggttgcaagtcctgataactaatga
 gttcttctattgattttgttatgtttctgtcagaattatgcccctcgt
 agagctcctgcgcgtgctgctgagcctaatttccccgatataactcagtt
 aggggcaatggctcaagcttttcaggctaatatccgtcctcctcaga
 gaacgcccgtagagacgatgtataatctgaaactgaaacttttgaggga
 12501 aatgaaggttatgaaggggcagaaaagtggttagatcgaattgagcagac
 ctttcaagtgatgcaaagtcaagggaaacctgccagctaatagatgggtgg
 agaccaccacctggttttgggcccgtgagccagctgctggtggataaat
 cagtcgaggcacatggcacctgaaagggcagccgaatgggaggtatttaa
 ggaaaaatttatgaagagatttgttccctccggaatataatagatcgcaaga
 aacaggaattccaccagctgaagcagaggaatattgtcgcacatgagtac

MC-12SSR384For

MC-12SSR384Rev

tacaggaagtttactgatttatcccgttatgattctgatttagcgggtaa
tcaagcagaatgcttcgctcgtttcaagctaggatctaagaagaagtaca
gaacgtttgccaatgcacttccctgtgccgattatcatgagatattttgag
13001 attctggtcaggatggaagactctgataatcttccggacagtgmggatga
agaggataaagggtaatggtcagaagaaaaatgagaaaagtaaggtgtt
ccattccaggacctcgtcagacgcagaatttcaagaaaagtggaatgagt
tcgagttcttccagtggtggatttagtgccacaggtccgaggagaggagg
tggtaggtttggtaatggacctagattttctggtcagagaggctttggt
atactggtagttcgggtcctccgttatgtcgccgttgaatttccgacat
catggggaatgtaggagaagcagtggtgcatgctttacatgtggtcagac
aggacatagagctatgtatgtccccagaatcagcagagggccccagcagc
ctggttatgcaacatcagcaccgactcaacagaactttaattcaggcagt
tatggccaaggtggtcgtgggtggtgcttatcactatcagggatgatgctgc
13501 tcttatgctccgggacagatcactatcccaggatccttattttcaga
gtggatattctcaggatcagggagggtatacttcatatccgtctatgcca
gctagcggatctcagtggtatcaagggggccagccccacagagcggagt
tgctgctagtagtacagggcgttttagccgctgcccagggcaggtcaag
gcgctactcatcagggacgaggttaaccagagtgccagaggtcgtggagga
cgacagtcagctcagggcaggttaaccacatcgtacaagatgctca
gaaccatccagacttgattatgggtacgttgaatgttcttgggtcattttg
ctaaagtcttgattgatgtgggtgctacacactctgtgattttctataca
tttgctcaataacgcaacctcatccttccctctagggtttgatttaga
gtttgctatgcttagaggggataaatgttatggtgatagttttactctg
14001 ggtgctcagtagtggtagataatgtcattatgcctgctaactttatcca
ttagacatcgtggattttgatgtgattttaggggcggttggttgcat
taatcgcgccatatagattggttacgggaaatcagttacttttcttctc
ctggactacctgaggttacttttgggtgaaagaagtggggtgaggcat
gggtgtatttccgccataagggcgaagaaattggtatccaaggggtgtca
gggatacttggcacatgtggttttgaatgatggtgatccggtagtggtg
aggaagtcggagtagtcagacactatcctgatgtatttccagatgattta
cctgggtgctccagatagagatgtggaattttcgattgatttgcttcc
aggtacgaacctatctcttggctccttataagaatggcaccagcggaa
taagagagttgaaaattcaattgcaagagtttaattgataaagggttct
14501 cagcctagtcttccacttggggagctccagtggtgtttgtaagaaagaa
agatggaaactttgagattgtgcattgattacagggcaattgaatcgggtga
cgataaaaaaccgttatcccttgcctcgtatagatgatttggttgatcag
ctcaaaggtgctgtgtgttttctaagatcgatttgaggctgggtatta
tcaattaaagattaaagatgaggatggtcataaaaacagctttcaggactc
gttatggcattatgagtttttgggtgatgccatttggattaacgaacgct
cctgcagctttcatgagattgatgaatgaagtattccaggaatatcttga
caagtttgttattgttttattgatgacatcctggatatttctaagtcga
aatcagaccatattcgacatcttaacttgggtgtaaggaattgagggaa
caccgattatagccaaatttagtaaatgtgaattttggctggatcaagt
15001 agcatttttgggacatgtggtatcagctcagggaaattcaagtggatcctc
aaaagatagcagcagtggaatttgggaacaacctcgacggctcactgag
gtgaggagtttcttgggttggcaggttattatagacggtttgttcaaga
cttttctatgattgccttgccattaacgaagtttaaccaggaaggatgta
agtttgagtgaggatgaaagtgtgagcaaagtttccaacagttgaagtat
tgccttactcatgcacctgtgttagtacttctctgatgataatggtaactt
cgagatctatagtgatgcttcttgaatgggttgggttgtgtttgatgc
agcatagttagagtgatgcttctagcagtggaagactcatgaa
agaaactatccgactcagatcttgagttggcagctattgtgttgcctt
gaagatttggagacactatctctatggcgagaaatgtaagatctttaccg
15501 atcataaaagccttcagtaaccttttactcagcatgatcttaactctcgt
cagagaagatgggttggaaattggttaagtattatgattgcagattgagta
ccatccgggtcgtgcaaatgtggttagctgatgctctgagtaggaagcctc
aagggcgacttaatgctttgtataaccagtcgtgttctcttctggcagag
ttgagatctactggagtagaattagaatgggaagagcaagtgaaagctt
tcttgccaaatctcaagtcagccaattttaattgatcgggtgcttgag
ctcaatcgttggatgaagaaattcaagaattgatcaatttaagaaatgaa
gggaagaaagaaagatctcaagatccgaggatcagatggtatgcttatgca
agagaacaggatgatgtgcctaataatgaggaactgaagaaagaaatct
tggatgaagcacattgttccgcttatgctatgacccagggaggaactaaa
16001 atgtatcataccattcgaccattctactattggccgggtatgaagagaga
gattcggagtagtgagtaggtgatttctgctgccagcaggttaaggctg
aaagaaagaaagcgttttgggagattgcaaccacttcccgctcccagtg
aaatgggagaatataacgatggattttgtgtataagttgccacgtacaca
aaatgggtttgatggcatttgggtggtttagatcgactaccaaatcag

cacattttattccagtaagggagaagtactctttaaataagttggctcag
ttattcatalatcgaaggttgtaaagatcatgggtgccagtgaaatattat
ttctgatcgagatccaagatttacttctaagttttggatagcttttcaag
aagctttgggtactagattgctttacagtacagcttatcatccacagaca
gatggtcagtacagagaaactattcagacactcgaggatagtttgagatc
16501 ctctgtgatgcaatttgggtgattcttggcatgatcgttggatttgatgg
agtttgcctataataatagttttcattcgagtattggaatgtctccattt
gaagcactttatggtaaagcttgccgtacgccattatgttggtcagaggt
tggtgaaagaataacttgagggccagagatttgggatgagactactcaga
atattcaggaattaataactaacctgaaagtggcccaggatagacaaaag
agctcagcggatcgacataccacggacagaatgtataatgtgggcgacta
tgtttttctgaaattatcgcttggagaggtgtgggtcgcttggaaaaga
aaggaaagctaagtcccaggtacattggaccttatgagatcactgagagg
attggtgaaagtgtcttacaggttgagctacctccagagttgtctaaggt
acataatgtgttccatgtctcgatgctcggcattatgtgtcagatcctt
17001 cacatgtgattcctcctcaaccattggaattaatccggacttgacgtac
gatgaggaaccagtgactatactggattggaaagataagaccttgaggaa
taagaccgtgagcttgggtgaaagatgttgagaaaccattcagctgaag
aagctacttgggagacagaagatcggatgagagatagtatccgaggtta
ttctatgacttttgatgttctgggttagtgattggaatttcggggacgaaa
ttctataaggaggggagatgtcacagcccgtcccagagttactaaatac
cgaggacgtgaaattactaaaacgcccttaaacgggattaaggtgcgtaa
atltggatattgttttacactaagatcctaaactagggtagaattataat
tagtatgtgttaatttaatttgggttggacttaggggtggggtcctacc
17501 ctaaatcctccccaaaacccgtgggagtgctctctctcatctccctc
acactctctctctcttcttctcatttcagtatcactctctctcttactc
togaactctctcaagctctcggacaaaacccaagaacaatcacatccta
caccaacgtccgatctaagatcaccacgacactccttaggacctcacgag
cacgggaataccagtttaggtaagttctatctcgagaaccctagtttta
aaacaccccgtgtaatggcactgttcacgaacttaattttgggtgtgttt
taggttaaaacaagatcaccacgagctcttaggaagttccaaggaggctcg
gagtgccctcggttgaacaaaatggacgtcgggatcgtcgggtcagattt
ggccgaaattgaagaattttgaaaggtttgatcttgtgtttttaggcct
taaaaccctccaacgtgatagtacatgtgaaatgctcattttgggtata
18001 aatgcgaaagaatttgggttgaaaaacgaaggagaatagtggaatttga
atlttccagaaaaccggcgaacagccgccggcgaccggcgactcgccggag
aagacagggaaatcttccgtcaagtttgacggaatattccgacgccgttag
ttaactttaacggaaccgttagtttttaacggaatattccgggaatatt
cctaacgccgttcaactgtagccgtcagtggtgacctggcagctggctgcg
tgggcccgcgctaggtccgtgccacgtcagggcgcgtgggggcccgtgagg
actccaaaaattatlttaaaaataggggatgatcctgaggttgtgtagg
tcaactgtgttatattcatataaccaatttgagcatcgtatgaagaattaa
ttacctagtttgggttaggtgcttaattgtgctttaaatagattgtttta
agttatttcaactctaggtggaacatttaacgaggacgagcacatccagg
ggcgtcaagggggttacgacccggcgacataaccagttagtgggattgct
18501 ttctatataatataatataactcagatttccccagaaatcaaat
aatgaaaaatataatgaaatgaaatgaaatagatgtgattgccatgca
tagaatgttatggatattatgaactgtcgtatgatgcatatagatgat
gggtgcttgagcgcacaggttaagtttaataactgttatgttgaggatg
atgatataatgagctcatatcctgcaccatggtttagtgcttatagatt
caccgcatcgcacgctcgcttggatccaagtatagctggtcgtacagt
ccacgcgagtggttacgacgagggccagtcgtagagtgtagtgagattat
gactggtgggtgaccttaggttattgtatacagatgattgatgagagaag
cactagagcgaatattaccatgagtcgttcagactacattagggtgggtcc
gacttatgtgcagaaggccggacaggtcacgaggagtgactccggcagag
19001 agtgagattgatagatgttgagctctaggttcaatcgttcagggctatta
gaggccctccgattgatatttcttttacctgattatattatgttaatgc
attcactacactgttgaattggcatggtacattcttattgaaatcgt
ttataagattgatgattgagacagttgagatataatgctatataactatt
tttctgggaaagtatacaggttttccaaaagggttataaatgtggatt
tatgaaatgttttggaaaagctttatlttgcgccactcagttttctgtt
tttgcgccctccaggttctagttgatagttgagggcgttgggtggcctacga
gactgcttcggcgttctgacagactaaataaatgtaggattcacccgagg
gtgttgaataatagttatgttctacttgactgcacctagatgcttatgc
tctgtttatgtgtgtttagtacactctatgcacatagaatgctcggttg
19501 taataactgcaattagtggttttctgttactcgtattttattatataat
cgtttccgcttgcgtttgggttacgtcacactcacgtgacggccagcacgt
cctagtcctcgggttagggtgtgtcaacaatcctttttaatcatatccaa

agtgcgaagtcaagaattgctttggtacgataataaattgaaccacacag
 ctgagtaaacagagaaagcaaacacaagattggtgccagatggatggtc
 ttatggttaggcatcaggattggccacaaggtagttctcaataagcttga
 acagaccatgggccttgtctttgccagcctaacatgttcttctttgatc
 tcaacatctcccttagtggtagtggctgggttcttgatgatggaacc
 actgcagatgccaccaacttaatttcataagagatcttctcaatttgt
 cagaaatggcatccccttcaatcaagctatagctatagtcaaagttatct
 20001 ttgtcaattccctcaacctgtgcttcacatagctgtattcactccctgt
 caaatgaaatgaaatataaattaagatggttagctagtattttcgatata
 tatgtgtgtgtgtgtaataaggtaaaattaacaagttaatgaataaaac
 tgtggaaattaaactaaccttcaccaaaagctaaacttcttgatggtcca
 acaccgcatctccctcgagaatttcaacagttttgactgcttggtagc
 aatctcggatgagattgtcagcatcaagaacaagggcattgtacaacc
 tagcaggggggatgactgagggcgtattcggtttcgtatgtgaggacacc
 atgatgattttaagggcttgaaaatgatcaaggaaattaatgaaagga
 ggatgctaggaggatgatcgatggtgaggtgagaaaatgttctctccata
 20501 ggggtggtatttatagatttggatggagggatatagtgtaagtaacccc
 gctccttatgtataattcaaatgcatgataaaaaataaaaaaaaagaaa
 aatcaatgtataaatgtattagtcaaagaaagccaccaatagaaactaa
 tacaatttattatgaagatattatgaaggtgtaacatttcttcga
 cataaaagaaaagacaatcttacaatgtcgggtgggctgttgatgatgag
 atgcttacaaccataatttggataatggcacaggtggcgaggatgacac
 aggagaaatcatgatcagtagctatagataaggtataacatacaaga
 acttggctctcgaaggctaattaggtgtaacttttttttaatttttaa
 tgcatacaaaaaaaaaaaaaaaaaacttgaacattattctctcacttt
 gttgataccataagaaaagaaagtagtattccttgattgaataaaga
 ttattgttcaactaacacagtagtaactgaggttatttaagtaaatttt
 21001 atcacactatcatggtgatgttacgctataaatgttttggtagaggata
 gatgtaattagtgagcgtcatgaaattcttcattattgtattcactttt
 ttgagctccttacttttacatttttaagatggttataatgatgttaaaat
 caacttttaggctccttaactgtaacaatgatgattttacgttgtgct
 tgtcattttacactatcgaagtataatattatcatgaacgtgcttgcat
 tacaanaatcctgctcaagggatcctgtttgattttctggctcgcctac
 tgatcatagctatgaatacccttttttttttaaggtgctgtagacta
 ttgagtgtaacaatttttgtattttatgaaaaaataatcatccatgtaa
 aatgtgttagctatgtaaaatataagaaaacttggatgaacactcatta
 tatacttgaacaaaactttttgtatgttctagatattttatactataca
 21501 tatcatctatttgaaatttatatttccaatataattatataatttttag
 tataagtaaatatttattgtatgtttataataaaataaataaaactaa
 aaacagacctcctagacgctatgctagtctgaagctcttactcgtcggc
 tgatttgacgttttagaacgttggatataaggatattatatttttca
 atcatataaagtagaaaatattctatatttttctcattatccggatct
 tttatgtcagaccaagaaaaaacatggttggatgatatgtatggtgaacg
 tggacggatttttaggacaatcggcttttggacgtatcaagacaaat
 tggtagcaatttgcagatcagattgggttagatggtagttgttgaacg
 tgggctgattttgaggacaatcagctgttggacgtatacaaacaaat
 22001 tgttactaattgcatgatcagattgattaaacggataaagcttgaanaat
 taattcctactatttctttggcccgaatacaaaaatgatccacgtcttt
 gcatatataaacgtacataatcatttctatgtcaataaaaaaacgtctct
 attgaaaagctgtttcttttttcaaatattattattttaaggggagga
 ggacagtttgaactattagaatagttttgggaaagggagtttcgaatct
 gggacgtgtgggtggaanaatagggctagaaaaaatcccgaatatacca
 aatgtacccaaaaaactcccgatcctgagttgaaaaaatctcgaaccaa
 aatcctgaaagtttcagttccacaaccccaaccgatcctgaaccattcgg
 tacgggaatcggctcaccatttttaagagtttggtaatcccaaaccaaag
 cgaaaccttaatatataatattcttttgttatgataattaattatgt
 gtgcttttgttttaggtaagcggcttaagactgtcttccagcttcatgc
 22501 cactacaacaaacaagctatctaaaacctataggggtgtttgttactct
 cattaagtgtagtggactgacaggactaaggggttagtccagctccatgt
 ttgttccaacatagactaggtttaagagactaagctctcactcgtccg
 acaaaaacgttgcaggggtcttagcaagacccccgaaaaggagcgg
 actgctagctcctgcttcgccccatcttctgctcgggtcacccctccccg
 ctttgccccatctctgcttcgcttacccttccccgctttgatgctctc
 ccttccaaaaatcctaattgagcttcgcccgtatttctctcccagaaaa
 gttagaaatgtttatcagagagataaaaaagagagaacaatggagctcca
 tgactttgctgacgagaggagcatcaaaaatcgggtacaatatttcggtt
 23001 tagttgggcaaacatggaactcagattgtcgtcaggcatttatgtaatt

Mal d 1.06A ↑

MGVLTYETEYASVIPPARYNALVLDA
 DNLIPKIPQAVKTVIILEGDDGGVTI
 KKVSVFEGESEYSYVKHKVEGIDKDNF
 DYSYSLIEGDAISDKIEKISYEIKLVA
 SGSGSIKNTSHYHTKGDVEIKEEHVK
 AGKDKAHGLFKLIENYLVANPDAYN.

tctcactatataaatttaaatcccatttttttaactgcaatttgctccaat
atatgaccatctgagtagttggttagtatcttggtgttggctctttgttg
ttaaattttgttttatgggttttatccggtgatgcttgttttatgaatttt
gtttttgtgtgcaaaaaacatgtataatagattggtgtatttggtgcag
gctatcggatttgagggcgatagcgaatttttgtatatattttttatta
ttagaaaattatgaatgcaacaatgctagaaattgtaggatttgaggct
tggatttgtgagttatctttgaaatctcaaattgtaggattttttgtct
aatgatctctgcaatcaagggttagggtctagaggctgggtttgttg
23501 gggttagcttccaatgatcttccataatcgatttttgcagttttttaa
gagttaaaggtgatgcattactcaggcatgagttgttcttatgtatgct
tgttatgatgaggttcagccaaagtcataagaaagtcgtaggaatgaa
aaaggagactgagaagatcagtggtgatgataaattgggccaagatcaaaa
gttaggtcattgaaaagaagaagggaagaatataagtaccacaaaggaa
ggaccagaatcctatatggaaaaatattgatacaagaaggacgcattga
tgaactaatgggtaacaccattataaaaaagagagataatgatgtaa
tcaaatatgattctctttccatctttggcatcccttatgtgcaataagat
tgctcatgttaagcaatgtgagacttgttcttcttcttctgtgtgaggt
tatcctaattatggtttccattgatttgttatggttggttttgtctattag
24001 tttttgttgggttttctctggatgagatttaattgatataatgcattga
ctagctatggtgcaactctgcataaaatcatttcttgatttgcagtttct
catggctgctgaagtctatttgaataggatagttgtatgtgtattgctga
gagcgaatcccagcacctgaattagtttgcagtatttgttgataccattt
gatgatcatagtgatgactggatttgggtgggcaaaactttcacagaaa
gagttttagtttaattttgtgctgaaatttgcagggtgaaaccactgtt
gaagaattgcccaggattgatggaacataatgggttggatggtttcaact
atataaagtgtcttctctggatttattgtgctattaatttatataatagataaaa
atcgcagaaaagcatgctcaagggtgagcagtagcttgaataaagagag
gtagtcaaccaagtgctgaagccatattacaattacataccggtttatt
24501 ctccagtcaccagtcgcataaacaatggtaagggttcttgcatacaaaaa
gaagtaaaagaagcagtgacttgattcatcaagaattctgcttctctgtt
aatcagttccaattgagccataagtagtcatcgggttctctagttcttt
tggcgggtctctccaaaagacattgatgttccaagttctctcagaagt
cttctttagtagacagcaattttctaagtttatttgggagtgcttcttca
gaagacatcagttgatcgggtattgctgcttgagaagacattgactttcc
aaattcttttgcagtgcttcttcagaagacaccaaatctcctagttctt
tgggtagttcttcagaagacatcgggtgatagagtaatgctgcttgagaa
tacatcgactttccaaatctttttgcagatgtcttctgttttttact
aaactggcaattttacaatgtctacttgcagcaggtggttctactgtg
gacatgctcaagcttggtagactttgcttgggtgctagaattgttgttg
25001 gttataagctgaaacgacgcgtccgaagttgggcttccaaacttctcaag
aagatcagagatgattgatgcaactggatcgatttataatgatacatt
taaagtgttgttctccatagaagaatgattctccaacaatagttgtttc
tctagagggtgcaatttttatagctttccgagattatagtaaacactgct
tctgacaattctgacatgatcaacataatattgttaccctcgtgttttat
tatgttaatgataaatttacaagctaaccatcaacagtgctctttttc
tgtactttggaattgttgcactctgctgatttagcctcgttgcagatca
atgtgatttaagtcggatgaattctgaatatacagtcggaagggtattt
ttttgcttgatattgtctgcatgaatacttgttcatactctccaagctta
25501 ggttcttctccttttctgtttttccatcgaaatttcaatatgttggcg
catattcgcttgaatttgtttcgtcattgttagattcttattttcgtcat
tgtcaatcagaccaggctcatttttctcccggtgttttacggaagttgt
atataaattggatagtatgaaaaacagttgcatttacatcttctaagct
actatcattatacatttatattatcttctctacggatgctccatagga
ttagaaaaagatagctacttggtagcttggacgatgggtgcttctagttt
tagacatactccttgggaattcaaacctgctcgttcttatgactgag
ttcatgtgatagcactcagctcctttgccaacattggcatcccacagt
agaatacccctacaaatagctgatttctttaatatactcgattctttgtt
tcgttgctctactagatgcatgatttagggttctggtgctctactagtcaa
gtttggaacaatgctacttttgaataatgtatgttcttctggtggcaag
26001 aaaatttgggtttttgtatttttattgagctcactctcatccaatattat
taattttataatatacttataaaaaaacaattaaaatagaacaat
aatgtagtcctagctgagcaaacaccaaatgatgatagtagtctgtttc
attatcttcttcttggttcaacattgcaaccaaacgcttcaactaagtag
tacaacttaactagctcaatccagtcagtttagtctcaagctaatc
caactgaaacagctgatacaacaacgtaccatatactttcagcat
cattttcagctctgcatccacatgcatgttctcaataaaacaagaa
taaattatacatcttctgctcctaataagtaggatgtgaatgccaaggctt
tcaagatggaaggaaagtgatcaatactcctgttgttatcagctctaca

gatttttagggttattgattaataactctgttgtttatcaatctcctgtaa
26501 ttggttaggggtgtgttatccacacactccatcttacttctcacacacccc
ttgataatztatgtccgctgatcttcttcaattcatctgatccgacggcc
gaaaattaaaaaggtgtgtgagaagtaaaatgggggtgtgtggatcaca
ccccaatgtttattcaccocctacagtgactgccatctttacagaattg
ataagcctattgtttgtctatataaaccocacttgggctaactaatatc
aatccaaaattcaaataaattaaaaaatcccaacaatttaggcttgaagg
cccaatcccaactatcccgaaatcctcccaacaaattaggaatccga
aatttcgtttcggtttcggattttagaatcttgtcccaagaattatgat
ttgggattcagttcgaaatcccgacctgaactagccctagtggaaaccaa
27001 cactctatccacaattacccocctgaacttttataagtagtcaagttccc
catgaacttttaacttttagccgattacccocctaaaatattataaatagcc
aatttacccocctagcgttagattttaaaattttcatccatccaattttc
cattctttttgcccataacattgtcatgtgttctccacgtgatcaaaat
ggataaaaatgaaattttttaaatccagcgtcaggagaaaaattagcta
tttataaaagttcaggggtaaatcggctaaaattaaagttcaatggggaa
attggctaatataaaaagttcaaagggtaatggctaaaatgaaggttaa
gggagaaattgacacctctctacaagttcatgaggcaaatgacaaaata
tgcttatttttaactgtgatctttttgtgtgcttcagattatgaataaa
tatctttattcaaaaatacaaatttatatttaacgggttaaaaaatagag
tatagaaagtcacaaaataccataaaaaatggctttttatattagcaccoc
27501 acaaaatgttgaatgcaactcacattaaaattagtattaaatgacttat
ttaccctactatgcaatgacaattttgactttattaggttaaaaaataa
ttaaaaaattatatacacccoccaatcacttttaagttattttatgt
tcttttaactatacaaacocctttcaaccocctcattgttgaaaaaacctc
accaatgaaactacaacatgttgattgagaaaatttatttttgatgaat
gaaacacgaaatcgatgttttgaaacttcatatgaggttagacttcata
tttttttatgttttagtgattttgatgacttcaataattattgaatcct
gcttttgatgcattctcaagcttatatgttcgtattcatcttttaggga
tctcatggattaaatgaagaattaaacacocctaaaattactaccaaatt
28001 aaaaactaaagacatgggttttaaggtgaaattgaaaaaacccacata
tgctttaaactcaagcgacatctttgtcaaaatccagtagacatttt
ttgtccaaatgaaggtgtatattaagaatgtgggggtgtgagggattatg
aggaaatagtgggggtatattaagataattttttaattgaaaaagcaaatg
tggagtgatattaagtagtgagggtttattcaataattggtggaggttaa
tataataagcctaaaaatttgttgctaacatagatcgataagttttata
agttgagcagtcattttcatgctcaataaatgtttgttaagttcgttgtgt
agtacgagaagttagagatttcgggtcactacagagtggtggcaataaaaag
gcaccgcaccgcagcagcaaggcgcaaaaacaaagtctttaaacaaga
aaattcaaaggtaatgacaatatccaatgttgactgctagttgccattgc
28501 tgaaccaaaccacaaaagtcagaaaaaggagaaacaaaataaaaacocca
gctacataggtttcctgcattaccaccaaagtttagtttaagaagaaac
aaaaatgtgagagtaattttttagatcgtaaaaatatacactatttctaaa
atctctgcctaacattggctaaatcttgtctacgcgctaggcgggtgac
accgctcctgactaattcataggcatttgaaaattaagaatgacacatag
actgcctaggcgcocctcctagcttactcagaccgcctaggcatcaactt
agatcgggactctcacttggaacaaaaatagataacttgcatattgcatt
attttctaataattataatttatatgtcatcatattttgcaatttatata
ccaaatgtaattatgtatttttaagtataaacaaaaactattttataag
gtatataataaaatctacttaaatccgcttaaatctcgocctggccgctaa
29001 gtgcccagaccocctaccattgcccactagtgccatcatcttcagaact
ttgaaaaataactatgagtagctgttgtgggttaatatgctgaaacaatt
cgtgaagttcttactaaggctgggttcgggtatgggataccgaactgaaacc
accatcctgatttccaaactgaaatatttcggaacgggatttcaattccc
gaaattgaaocgaatcttaagattcccgaacttcaggatgcccaaat
tttttcgaaatatacaggacggatcgggattgatacttcatgcccacaaatcc
aattcatacgaagccaaattgcaccataactctgcaattcaaaaacaaaa
taaaaattctgtaaaaagctcaaatgatcaaacaaacagatcccaaac
aaaattcaaatcaataaagaatcagatttacatccattattattatta
ttattattattattatttgagaaatgactacaaccattaaaaaaaacca
cataattaatcatgaagacctgaaaggatgggtgttgctccgatgcgggtg
29501 atggcaagtgggcgaagctgttagcgaggaacggaggatttggggtgagc
agagatggcgcagatgaaagaagagctattttgggggtttgggaatcgaa
aagatcacagagtaggaagaagaagagaggaagaagggaacgggtggca
ggacatgaagagacaactgttatggcagtagtcgtagccactgacctcag
aaatgcagaggcagaggagagaacagaggcgaagaagtgagaactcaaga
gagatgagaggggtgaagattgaggcttcgggtgagaaatgtgagagagtg
agagtgtaaacccataacttagattggacgggttagatttatatctcaac

caatccaacggtccaaatagtttataatttctagtttaaatataacat
 attatataatataatattgttgggtcgggtaccggaacttaaggt
 30001 ttatcgaaccaaataccataccgaaagtttcggactatttcgggtcggg
 ttaccgaagtttcgggattttcgggttgggttttttcggggcgagttt
 gagaaatttctgatcgggtcgggtttttcagctttttttccagcccta
 gttcttacctctaaattagaaacgaaatttcttcattatttcgactttc
 caaaaatcaggttttaactttcatgaccgacaccaaataaaaacagaaa
 ttcaaagttgaaagttggaagaccacctagaaacccggctccttcatagt
 tccttaatttgtccaactcctcgtcgacaccgagcaaaataatttgggtgg
 aaattaaatttttggagataatgtccgataataactaaatgtagtaa
 caattgcataaactgggtgatgatctacaccgacagttcacagcttcaca
 acacttgggtcctaaaatctcaacctcagttcatctataaataaccacta
 30501 taccattctcctcttcatctcaatctcctcgggtcaaaacaattctctt
 aatccttttctccttcatcttcatcttcaagtcataaaatacaca
tgggagttttcaacctatgaaacggagttcatctcgggtcattccccacct
agattgttcaaggctttcatcctcgacgccgataacctatcccgaagct
tgctccacaagcagtgaaagggcattgaaatccttgaaggaaacggagggg
ttggaaccattaagaaagttaccttcggcgaaggttggatgggtgtccct
 atttctttatgctcaagtgctcctccttcttctccttcaaaaatcagc
 aaaacctcatgaaactcatagccttgggttggattttctccatgaaacttaa
 attaggttataagccatctgatcttgaagagaaatgacttcatatgcgc
 cttttatttatgtatgtccattgatttttttttaatttattcaacacaa
 31001 tggacaaaaataaataaaggtgcataatgacgagtaaaaatcacttccct
 cttttaaagtgccaattatcccttgcctcacatcgatttaggggtctta
 aataacggaattttcttttggggggggggaggggggaggggattgtttc
 ttgatttctccacttttttagtatctgatgaataatgatggtaaaaa
 atttgtggcaatgacaagcattatttaaaactaatttgaatattttagt
 ttttctag**gcagccaattagggttggtaaaagcacagaatcgatgggattg**
acaaggacaattttgtatcacagttacactttgattgaaggagatggttt
ttgtccgacaaaattgagaaggtagcttatgagaccaaattgggtggcatc
tctgatggaggtccatcgtcaaaagcaccagccattaccatgccaagg
gtgatgttgaatcaaggaagagcaagttagggtggtaaaagcaggct
tccggtctcctcaagcttggtaaaagctacctcctggccaacctgatgc
 31501 **ctacaactaa**atctgaagttattcatatttataatagcttgagctccca
 gcttaacagctgaaggtgtgaaggggtcgtttaaatatttcttgaccagt
 tgttcatggtagtagattgaatatctatatgttgtaattgtattctat
 aataattcgatcggattacccttaataaaaaataaataataataactc
 gagactcgaatttaattttacttataggtggataaaaatgttgaagtgc
 aaattttctagtgtgttttctttaaagttaaagaagaatagttaaaa
 gttgtcactcctcgaatttgtgtcaataatcatggttatttttaggaaaa
 tttagtttctggctccttagtttctaataatttattgactaagacctatcag
 tttttaaattttcatcaaagtccttagcattaatataatgaattaca
 32001 tgaagttaatgtttagcctattttatttgacaaaggtgatggggctgg
 ccgcaaggagagagagagagtagagagagatgatttgtaaaattgtagggg
 aatataatgttatccctcctacatagtgcccttatttatagtagtaaaa
 ggagagaatataatttctcctcctcaaggaatacaagttgtaataaggaag
 ggataactagaatacaaatctaactaggtttacacaatacacacttaaac
 taggaaaaattacaacactcccccttaagtggtgtaataactcaaggtaga
 ttagcatcatgcagaagttgatgaagttgacttgtcgacactgattcta
 aggaataacacttattctaataaaggtaggaacttgcataagtaagtga
 tctcactaaaaaaaaaccctaaaggctatgacaaaaaccgatgagggaca
 aagtcctatagctaaagaaaaatgtgtgagaaatgcaaagtcaaaagaaa
 cgtctacgagacgtcatcagggatagaccagcccaaggtgggtgcctcg
 32501 ttaaacctagtttagatagcaaaaaccagtgggaaaaatgctcataatc
 gtaaggaataagagtacattaagatcaagcaaatattcttcaggatactc
 cctctgagtttaacacaattccaagagaaatagcaatgttacaactcag
 acagtttacgcataactaatccatgaacaagcttctaaaacgtcatcttt
 gataatgattttagtgaagaggtcggctagattgtattgtgaacgaatttg
 cgtgactacaatcttctgatgctcatgttgttgatgtgaaaagaagaact
 ttggttcaatgtgttttgggtgtgtctccttgatataagagttgttcgat
 gcatgctacgttgtattcaagatggccgttgggacatcaacgacaggat
 aaagatctcaagagcttcgaatatggcctactactgctctcaaccaaag
 cattcccagttgtttcatgtaaggcgataatttcagcttgggttagacaa
 33001 cgtggcaactaaggtctgtttagttgacctcaagagattgcagtgccctc
 caacggtaaaagacataaccatttgaaaacgcaccttgtgaggatcagat
 aagatcatgctgttggcataaccaacaaggcgagaatcgactcgagaagc
 agaggggtcggcatcactctaggattcatagggatagaacaaacccaaat

Mal d 1.10 ↓
(Mal d 1 m)

MGVFTYETEFISVIPPRFLKAFILDA
 DNLIPKLAPQAVKGIIELEGNGGVGTI
 KKVTFGEQSGLGFVKHRIDGIDKDNF
 VYSYTLIEGDGLLSDKIEKVAYETKLV
 ASPDGGSIKSTSHYHAKGDVEIKEEQ
 VKAGKEQASGLFKLVESYLLANPDAY
 N

ctgtagtaccttaaggtaacagaagatgtctttcacgccaatctagtgt
 ctgcatgtaggtagcattattgtatgttgcacaaaagattaacagtgaggag
 atgtcagggtcctaaggtat**taagctaagtaacaataaagcgccatagcact**
taataaggtacttttaggctccaaaatctcttcatcatcctcatttggac
ggaagagatctcgTTTTGCATCTAGCGATCAAATGACCACAATAGTACTC
 33501 **gaatgctttgctttatcctcattaagcggcgcaacaccttctgggtgta**
gttcgattgatgtactaggattccatccgaacaatgctctatctcgagac
cgagacagtatcgagcttacctagatcttcatctcaaatctgacttc
aagtgcgtggcagtttccgacacctcaggagtttgatgacattcat
gtcatcgacataattgcaacaatcgcaaatctggaatgtgacttcttaa
gtaacacacaagggcatagttcgTTGTTCATATACCTGACTAGTCAA
tacttactcagacggttatcacattctccggattgttcaaaccgta
gagtgaacacttcagccgtatcgagagcgtgttctggggttagaaacat
ttgaactagtcaatgtaagtcttggggaacttcatgtaaatccata
 34001 **tcaagatccctatatagatagcggttactacgtccatcagctacatact**
cagttttcaaaaactccaaaactgataaggtagcaaaaagtaatccat
ccataatgggtgaataagtttCGTCATAGTCAATCATAAGGCGTTGTGAG
aagccttGTGCAACAAGACGAGCTTTGTAATGCACAATTTGTTCTTCTC
attacgTTTTCCGAACGAAAACCTACTGTAGCCAAAGAGCTTCA
catgtggaagtaggagccacaggttcaaacaccttatgtttgcaagcgaatcg
agttcgacttggatttctgtttccagtttaaccaatcagttttatgtcg
acattcatcaatgaaatgcagttcaatgttatcgctcaacatgattgcaa
atgccaatgcctcgatcaatcatatcatttctacgccacacgtcatct
aaagtagcataatagactgaaatcccacgattctcggaatgaggattgt
ctcttcaaggacacttccataaatttagaatttCGTCATGAATGGATAGA
 34501 **atgagtaagcgtatagtcagattcacggtaggctcttcgggaccttgtgct**
gtggatttctcttccgggggtgtgaatcatttaaccaaggggtctgcca
tgcttttGTGTAGGGCCAAATGATTGGCTAGCCGGTAATGTACGTGGATC
actaagattgggtgtcccaggcttccaggaaggaagtcctcgtaacattt
gtatccctgcttgcaggcgtatgtgcagctggaatgtgatcttctc
acgcgcgtcaaatgggtgaaaacatttggtagctctgagctatacaatg
gagatcaatatacgtgcaacttcagtttcagactgagtggtgcaggat
ctaaatgagacaaaatgggggtcatccacaataattcacgtcattcttca
agaacgtaaacattcttatttgcctcaatgacgggaaaactatctcata
gaagtgcagatccatgaaataagcggtaaacagatcgcatgtcaaaggtt
 35001 **cttagtagcgaataatcgaatgagaatcatatctgacatagattctcatc**
cttcactaaggccccattttttacataagggcgaggagatcggcatata
tacagcacaacaaaaaacatgcagatagataagtcgagttcgtatccgg
taaccaattgaaatggcacaataatagttgggttgcaataggcctcaggcg
gaccaacttgttgcgtgccatattgcatagcccaagcagcgatcaaaa
gtttggtagctatgaccaacgatcgagcaatcatttgaagcgttaatg
aaagcatctgccaggtcatactaggtgtgaacatggggtacaaaatgttc
aacttcaacccccactgacatgcaatagtcataaaagtttttagatgtga
attctccagcatttccaatcgaatataattgatcagataatcagggtgg
taagcctcgagcttgataaacctgagctaacagtttggagaatgtagcgtt
 35501 **ccttgtggacaacaagcacacgtgtgaccaacgtgtggaagcatcaacca**
aaaccataaagatctaaatgggtccgcatggagagtgaaatcaatccataa
atgtccccctgaatctgttatagaaaaatgggaggattcaaacaaatctt
gtcataaagaaggcttggtaataagcttccattgaacatgtttgacatg
cgattccttggatcgaacctaaacttcaggctagtggatgcccggtgat
gatttaaggattcagcgcattgctattcatctaggatgtcccaaacgatt
atgccaaaagtgaatttctgctgctgggtcccagtagtagggccagccgat
agtggaattctatagggcgtatggctcgtatgtagttagttagccactcgagata
cgctccatctttctagaatatacttttggccatattcgtatgaagttat
 36001 **gcatataaagtttaactccgTTTTCTACATGGGTTTCAACGTGGTAATTGT**
tatctctaatgtccttgaaacttaacaacgttcttccggaacgtggagaa
taaagtgtctcaacaatgggtcaagattataccattggacaacattacacg
tgcttctcgtatccttctatcagggtggatagcctgaaagggttgtca
aaggtgtattcttaggtatgaagttagtgaataagatgcgttcaactcaaa
acgatgtgtcattacactatctgccagacaactaactccccattagtca
tacctagaatgaaaaataaattgaaatcggtcacatgcataaaaattcta
aaaacaaatcaattcaaaaataatttcatatattcaaggatggtaatta
atgaaaacttaatatccaaaaacaaatcaccaaaaaataatccaaacata
aaggaaaaactgttcaaaaatcaacccccaaaaacaaatgaagggttcaac
 36501 **cccattattcaatttcaactaaactaaaatgtttatgtctaacattataa**
tcttccatattgggtggatcctcttgaaagtcataaacctccatcattgt
agtcttgggtcgtccactgtacgaagtttgattcaaaccttcttacaac
gagaattatattcatctccaaccttcttgggagctcggcaaacacatgac

Retrotransposon protein,
 putative, Ty1-copia
 subclass
 (ORF9) ↑

caatggctctttgaaccacagcgatagcacatgtctgcatccatgtttca
agtgtttacccttattcttgaagttcggggccttagaaacgaggttga
gagcttctggtctttgtttctcccttggatgggccttgactatgttgac
cttgggtggatggcttctagccactcttaccatgcctcttttggcgttt
gtacgtgatttgtgctataatgtgttcaagcacaacagcatccaata
gggcaagcttgatgattatcatcaatagctagttctgcttttcagcgag
aagtaaaatagagatcaaatcgagaacctagtgaacttctaagctctat
37001 attgttgttgcaggacaatattagacgtagagaaggtcaaaaaggtctt
tttaggagatcctctttagtcaaagtttgttacaanaacttgaaaagtga
tcagattcgacaaacttcagaattatattcattcacatacttaagtctt
ggaagcgcaaatgttgcagatgtcttctcaggcaagaataagtcatt
tttgatgatcaaacgatcagccaaagcgaccatagtgcacgtgaatcc
ttctcagcaagatactcaatttgcaagacgtcatggatagtcttccgat
gaagatcatagcaatggctttttcagctttgccaacagggttgcctct
cttcttcaatggcaagcgaagttctttgcagtgaggtggagcttcaca
tcattggacccacttgggtagttctttccaagatgggaaagtttaagttt
gttcaaatcgacatgttctctacacaaaatagatggacaagatgtggtt
37501 agtataatggagaaaaaaatcaatccattcacataggagtagaatataca
ggttataatagacatgtattggtttaaatttgcagtaaaaaattcgggtt
ttcatgggtgatgttttaaatgaaaacttcagggttttaataaggcat
gttttaagaacttcaagtttcaaaaataattatgaacttcatgttcata
tattccaacaggcaaacacaaatataatcaaaaaatagcaaatagaac
ttcaggtctacgattgtaattgaattcattaattgaacttcggggccaaat
ttaaagtgtgagtgaaaaactataaaaccaaatttgttataaaatata
aacaatagaactcagggcccaaaaatataaggccaaagccacgggtggc
tgagcaaattttgggttagggacaaaacaggccgtaacaggcaaacccaa
aaaaaaatatacaatgcagccgttgcacgcaggccgaaaaatttgggttt
38001 tttttttcacgggctaatacaagccttaataatgaatagtgaggctgcaggc
ccaatatagtaggccggatcatgtgtgacgcaatcgggggaacacata
gcatcaggagagttgagatccgggtgcggcacacaatgtcgggacaccata
acccttaggtcagtggttgggttgagccatgtattgggtgcacatggg
agccattctaaaaaaaattctctttttttttttctcggggccgagaa
aggtcagcccaagatttggggcttagggttcaaccaagaacccccat
ccggagtgggtgcagtcactggggacaaaggccagtgccaccactgcag
acggccatattgtgggaaattttctcgggtgcaaggtcagaggttcgaca
acaagaaagttatgggggatgaaatgagtaacctcgacatcgaggggaaaa
aaaacctaaagaaattcaaatggaaatttcaaccaaattcgatgaaattc
38501 ttccgggaattcaatgttatagaatcaaaattatgataccatgtcaaat
tggtcacaaaattgcacgtcaaagccgtggtctcgttgcaagcgaagtag
attttctgggttggacgccatgagagtcaggtttcaaatgagctttgat
gctccatgcttgcgggtcatgggttcgaagaacctatgtgggtgattt
tgattttttttctctttttttttcaattcaattcatatataattcaatt
tcatgcataatgcaattcaatattttaatgtatgtacataatattgatgca
attcgtagcaaatcaaatcacataattcaacaattatgaatataatg
catatacaatttatgtagaatctaaaattcaaaaaacgtaagttgggt
catgcattatggtgaatattcatgctattagggtgcacaaaatcgaga
taaaaaccgttgttttgaagaacctaatcagtgatgataggatgaac
39001 tgggttgatgcagaaaaactctcaacgttagattcctaaacgtagccgt
agaagcgtgctgataacgtattgtaacctattttatcaacaaaggtgag
ggggccggcagcaaggagaagagagagtgagagggatgtgttgtagaat
tgtagaagactgtgtatgttatccttctacatagtgctttatttatag
tagtaaaaggagagaagatattccttcatcctcaaggaatacaagttgta
ataggaaatgataactaaaaccaaattcaatataggatttacacaatcgc
acttaactaggaaagttacaatagttacataattttatattaaaaaa
attagtaattgcttttagagtttaatactcacacccttaataaactccta
attaattttcaattcaagcatttttcaaatataaaaagtaagggttttt
tcatcacaaaatggtccctaaaattggcctaactcatcatgttggctcttc
39501 aatttgaaaatcaatcaatgtcatttctgaaattgactaccgcaaatcaa
cttggctcttaccgttttaacogttaaattttctgttaatagggcataca
agtaaaagccccgtgtccaattaaatagtgccatgtggaaaaatctataaa
atcataaaaagacgaatttaaaaaccataaaaaaatcccaaaactaaa
aaagaaaaaaagaaaaagaaaaagaagatgaagcaggcagcatcatcgtc
ttcctaagatgccataaccagccaccacttatgccagatcctaagcagc
aaccactccttccaaatccaaccaccagccaaaacccttgattctttg
actctattcgttgatctgttggctcgaccagccaccatctccctctgc
atatacttccatcgaccaggcagtggttgcgtgtgctgatgcgaca
cgtcttctatccatcgatctactataaacacogcttgcaatttcgttctc
40001 tttctgcaacaccacgtgctcctcttgaaccatattcttgctttctc

cactttcttacagcaacacacactctcacactgagttcttggagggtc
tgcaactgttgccagaacgaggagcaagggtcagactggatTTTTT
tataattattttatttggctaaataataatcatggcactaaatcattgg
ttgtgtggccacatgtgtgctatgttagcaaaactaacagaaaacttgat
ggaattaaacggcaaggccaagtTgattTgtgataggcaatctcagaga
ttacattgattgattttcaatttcatggaccatcttgatggTgaaagtca
atTTtagggatcattTgtgataaaaaacccaaaaataaattTggaataagt
ttgtaaccattaaattttTgtaaaaaattgatatgatccattctTaaat
40501 ataccaattTgtTaaaattTgtgcccTTTTTaaatgtaccattTTTTTa
tataaattTcattcaatctttctctaatccattTTTTTaaacttataTg
ggtaacttctatttcgTtgattTgagaatgatctatgtcaagTtTaaTa
gaaaaattgaaataatgtTattaagcctaaatgattTgttagatgctatt
TgactTcagTtTgtTTTTctTTTTctTactaatcctTctatagTtt
gattTtacaggcctctTgaattTgaattTgtgattggaaattggagTAcgC
acatggagTataggtTaaaaccactactattTTTTctTactTTTTcctT
aaatgtctTttgagTctgagattgagTctctTtagattTgTaaattTccga
caactatgcataccaagaaagccaatgTTTTctTtaggTataaagaac
41001 caaatacatTaatatataTatataTatataTatataTctTcaacgaatgc
ttTgctgtTTTTactTTTTaattgtTtagTtcaatggctgtTggatact
TaaaaatattattcgTtcaaatgtctctTggatgcataattaaatcgat
ataaattctgtctTTTTtagatggataattaaatcgatagattatgtTa
ccgTttgtTTTTatcagTtTactTattTgaatcaaatgccaattctaatat
TctattTgTcaagccctgaaTtaaaacatagTTTTTcacattaaattgA
gTTTTccatgattTgaaactctctTTTTgtTcaaattTgtTaccagTt
TgtgtTTTTTTTTcatctagcaggtggatattTaaattTgataagTcg
TgcgTaaaacaccataatcataaaaaggccctccgcatgcgctgcaggg
aggctagTtaattctTatggcgatgaggtgggacatgtgaaaaatacagT
TTTTctTattTccaaatataTatgaagaaaaaaatcatggaaacat
41501 actcttagatagTttcgattcagaggcgtacatgcactTactctTatca
TTTTcactagaattactgtagaattgcatatactTcaaacgaataaaTg
TataaactatagccactTtaactctTattTattTgcgggTcatggaaa
tagaaacgtgcagTggTggagaaaaggccttagagatgcgTTTTctTct
accaatgcTTTTctgtagTaaTatacaagTaatagagagaaaaaaat
TgctctTcaagTgtataaggtcaaataggaaagTattatctgtattacat
agaaatcctactataatactacctatgtTaaatattagTtacaacactTc
TTTTgagTgtgTaaatattcaagTtagTtagTgcatacaatctTtagata
ggaaTcgaagaagTtgatgaagTcgtctTgataggcattTattTatag
gcaacgcgagTctcaaatagggaactgcataaggagTaaatct
42001 cacaacacTgaaaatagTaaaatctgataggcaaaaatccatagTgTg
aggaaaaaagTgagTaaatgTaaatagTcaaaatTaaactatctagaatg
cgagTaaaatgcacacaaaaggTatgaccaacctataattggTgtctctTc
aaaacctcgttagacaacaaaaactcaatggaacaaatggTcccaatcgt
TggaaaaagaatacacTtaggtcaaacaaatataTtacaggatattcctc
TtgagTtgcaaaaactcccTtgagaactacaagcgtTgcagatcaaca
gTttacgcatagTaaTtctTaaacaagcTctTaaaaagTtagactTtgac
aatgaggtcggcaagacaattTgactTtaattTtTcgatgctctTgctaa
TaatgggagaagaaaaactTcagcgcataatgTttgctattTgtctctTt
gatgtatcattTctTgatccggTcaatctaaagcgacattTgtctTcataat
42501 cattTgcaggacatcaacgacgaaacgaaaacatcaatgatagaacgaa
TctggatgtgctctgaatagTcaataaagTctcagTcaaaaactc
atgtgtagcctagTgaaggcgagaattTtagcatattcaagaagTcgc
aactaatgtctgTttggTgactTacaagatattTcgatgattTgcaacta
TgaagacataaccatctgaaaatgtgtctggTgtgtgTcagataaataa
cctacgTtaacataTccaacaaagcaagcataattccgaagaccaaagg
gtgatccatagTactTttTgTgataacaaaagatataTtaacatcattgc
agTgtTgacataattgacgcgtTactataTctTgctaaaagattatcaacg
aaagatgtgTttcaaaaatgtctcattTgaaactgcccggattatacacactg
cattagaaatgaagattTaaTgagattTgaaacagaatggcaaaaactTat
TggctgcaaacctgcaataattTgtTcaatgagctTataTctatattat
43001 TaccgcagaagagctTatggTacgaataaaaactattTtagTcaactatgat
tagTtgatagagcaacaaaaaaatTattTaatTggTggTgaaatc
aaataatgaaacgtTgtacaggctataactTgcagaatccattgaaTtctg
gataatataTgggTggTgaaatagTtTaaacaaaaacaaaaataaTattg
atTattTgTaaaggaataaaaagatgtTcgataactTtaattgattTcaa
cagTgctaggaaactTatattTaatTgattccaatagctgctaggaaac
TattataTaaTgagattagctagTactTtggTagggtTTTTcactgtgg
gtagTattattcaagcattTctTTTTactcaaatattTttattTaatTTTT
TTTTattgattTTTTTcaattcattTaaTtaacggTaaaaaatccaaat

43501 gaatacgtaaaatataaaaaatagtgtgtgtgataaacactattctttattt
 ttacagaaagacataaaaataaaatagtcagcataaaatgttatcttgttct
 atgatgaccttccctccagctttcggcagaattattgcgcaaacctttttta
 ctgggtttttttataaaagtaagttaattggccacattcatgagatact
 acacgctcaagctgtaagttgaagcatgttgttgcgtccgfatgtggtta
 tttggtaatgtaaacctgctccattaatccggatgaggatctctccagatt
 ctctctacagggatcataaggatcttcacattctaactcgttcatcgtata
 ttatgcgatcaattttctgcaaatattatctttaaattttaaataat
 aaaattcaataatttatgaccgcacaatgtataatgaatacctaataatg
 tgaggattcttaagattcccacaatttggatccagatgggatccaaaccc
 44001 cattaatcctaccatttagttcatgtatgaatgaagcattgaattttgg
 tttcttaccattactttatgagagattatocacacacttcaaattact
 tctctcacatttttttaatttttttaataattttatacacaccactaacact
 tgatagaaaaatattaaaaaataaaagggtgtgaaattgggtgtgtgtaa
 atataatctccctattttattaatctcaaatctcataaattagcctataa
 ctgttagcaacaagcctaaacgcccattcaaggaaaactacaatataat
 accactaattctaaaaaatcttagcagcgtctgccactaagagatgttga
 acagaagcagatcacttgatcgaaacaaattaaatttagtgatcacgata
 aactaaggttaactagttatcagccacaaaatttctcatgataaata
 aatcagacggagttacaaatcaaaacttggattatttgtcacaatgt
 agatatccatccccagataatcaggtaccataagttattcagcgtttg
 44501 taacgaattttatftaattccaagtaatttaataaggagttactccagtg
 caagatatagaatftaatttagagaatftaacgaaaagttccggttct
 gttcattftaacgaaaaatcacattfttatattaaaaagcaattctgggt
 actattcactftaccctftatfttgcctftattgttaaaactcaaagttt
 tcaagttactftttatftaattfttctfttaattftataactaagaagttgtg
 agaaagcaactagaaatftaatgagagaaattgataggcaacggttgactgc
 cagttgccattgttgaaacaaaaagaactctcaagttctcaacaccccat
 cctcatcctftcccaaatftaaacaagtgaaaacaaaatttaccagatga
 taattaaccaagccaaactaaaaagttftaacattftataattattgggtc
 aattttgtcataataatgtgtgctctfttatcctfttgggtgacctfttccac
 45001 ctgaaactatgattftaaaaagctccaacatgftattfttcaaaaagcat
 cgatcgtctcttacattcctacctftttagaatgttaccatctaaatcaag
 taagatctgcagactgtcactagaatctgaaaagcagccttgaaagttga
 aaccctatgtgcatgcacatccagttaccggggatgatcatcaaagggaag
 aattgaaaccatacatctagaaagcctftatfttcttgggtacaaaatca
 coggcctcaaaaactcagttggacctftttatftaatctattcaaatgagct
 ttatttgccaaaacatggattfttaattaaagatagttactatctacgcac
 tcaattftaattctcacacattcctcataaattfttgttgcgagtcgaat
 taattgaaagaagaccaatgactacaatgaacaatgatgtgtgagaggtaa
 aatgaggttatataccgtcacttagatataccaatacacaagtgacac
 45501 aaattatftaaacattfttaagatcctftgtaattftgtagagatcgaatgat
 caaaatgacaaaatgccccaaactctactcctataaataaccacctctcag
 taccactctcaacacaaaactctcaacctcactaaaaccatcatcctt
 ggtagttgcttcttfttggctcattctcaacctcfttttctcatcattggtg
 gtgtctcacattftgagaacaggtacacctctgagattccaccaccaaga
 ttgttcaaggcctfttgcctcgatgctgataacctcatcccaagattgc
 accccaggcaatcaagcatgctgagatccttgaaggagacgggtggcctg
 gaaccatcaagaagatcactfttgggtgaaggtcagataaattagcattgat
 aattatfttagatggttctcgttaacaaattfttatcaacacatatacatgt
 46001 tgatgttgtcaaacatttacagcgttagtggttctatataatgatta
 tgtgctggtttatattcatagttcaacaataaataaattcatttgcctcaca
 ggcagccaatacggctcgtgaagcacaagatcgactcggttgacgaagc
 aaactactcatacgcctacactfttattgaaggagatgctftgacagaca
 ccattgagaaggtctcttacgagaccaagttgggtggcatctggaagtggt
 tccatcatcaagagatcagccactaccacaccaagggtgatgttgagat
 caaggaagagcagctcaaggctggcaaaagagaaggctcatggttgggtca
 agcttattgagagctacctcaagggccaccccgacgcatacaactcaatt
 aatatggttctatggtatfttgggtgtcacccctcagtcgaaggttgtgtg
 ggctfttcttgtatgctfttctagtcagccaaaagtagccatggttftt
 gagcactftggctfttgaaaaaaagttatggttgtgatcatcttctggt
 46501 tgccctftatgtatfttgggttgaaaaaaagttgcagactgtaagtgatt
 atctataaatataatgaaatcagttaatgtctcacttgcaccattgatc
 atftfttcttfttattatfttattfttgaactaactcttgaaaacctaa
 tagagattatattcacacacctcaaatftacttctctcacacctfttftaat
 tftftaatatfttctatacaccactaacacttgatagaaaaatattaaaa
 aattaaaaagggtgtaggagaagtaatttggagtgctgaaatataatctct

Mal d 1.02 ↓

MGVYTFENEYTSIIPPRLFKAFVLDA
 DNLIPKIAPOAIKHAEILEGGGPGTI
 KKITFGEGSQYGYVKHKIDSVD EANY
 SYAYTLIEGDALDITIEKVSYETKLV
 SSGSIIKSISHYHTKGDVEIKEEHVK
 AGKEKAHGLFKLIESYLKHPDAYN.

caaacctaaaatgttaattatacatcttttagaggcaaatctattgaca
 cgggaaatttttgagtttcaccgattgattataaacaataatattgtttac
 acatttttaaataaccatattgcatccttttcatatcgtaacaagtgt
 47001 gatgggacaaattctctccttttctaatcccttaagtcttcaaacatt
 atctaacaacacagattttgtcacaactctgcagcataaccagaacca
 caaatattccaagtcatcaaatgaagtggcttccccctgccccagag
 ccaaaaacaaaccataacatccgctcaagagttggacatcgataaggtac
 tatgaatgtagattattagtagtagtcttgacaaatttgaggttaata
 ctacttttgggttgaaaaacgacaacctaaattaggactcatgctttt
 aaataactaaaagcgaaaaagaaaaaaataaaccaatctttattaaaa
 tgaatgttaattttgaaaaacacttcaattgcttcatttaaaaagcatc
 aattatgtgattcttaaaagaaatcacgtcaagtgtttttaaaccaaaa
 ctattttctcaaaagtgttttcaattattataaaaaacacatttccaaacg
 47501 agataattaaagcctactgctgctgatatcccaaaagaaaagtaactgt
 tccaacgaaaatgctaagtaataaagactgctgctcagtaattatcata
 aactttcgtgataaactataagtagttaagatctactgagtaatttata
 ttacaccttttttttatttacacatttttattattttgatcattgata
 ttttaagtaattcctgaatccaataattaaaaattaaaaataatattgct
 aactaaaaatgaatgtatgaataaatgtcttagaactagacaaacatg
 aaccggctagtttagaaatacaatacttcaaatagataatacaaccacaa
 atagaaaattcaaacagataaggagacggctccactagttgggtgacccat
 gacctagtgcatgtgcatccatgatccatccacaaatgtgaccagataca
 aattccaacaaatctgagaaagatattgcagaaagagaaaaaccattatt
 48001 taaaattttgaaccacatgtcattcagattatagttagttggtattcttc
 ttttaagaagttaggtttctcctctcttattttcctcttattcactctt
 cttttatttgaaacgggtcacgggttaggtcacgtcaatattttatattaatt
 tttttataaaaagaagaaagacaaaatagagaatgtgagagtgagaggag
 gggatggaagataaggagagagaatcctagtccttatcacttgtaagt
 aaaatgttttaggttgcgaatcttgacagataacaaattcaataccaaat
 gattgttcattatgtaggttagccgaattccctcttcccttaattgtacat
 tatcgttgactaaaaaaattttatataaattggaacgactttcttggc
 cggtagatcgtgtgacttgagcccgtagctcagtcctcgtcaaccaagga
 tttttaggggttggcataattaaagtatgatgaaccagatacgaagaaa
 48501 taatgacgaaaaattaaaattgatgcagcttaattatccaacaaaaatgg
 tgataaagaattaaagatggtagtttgcttcccagacaaggtcttctggtt
 ttaaggaatacatccaattaaaaagcatatataaaaagtcgttgggtgct
 agacttttctttggagtagacttgggtttattgtcatttgaaggttgcgc
 cagaggaccacaaaattcaggattgaaagagtaaggcttaagacgtatac
 gacaaaataactcactcagcaacttactacaatgtacatgagattgaag
 ggttaataataactacgttttaagtgaaggaataataccgtcgtgatgt
 tgacatgacttgcttataacccttactccactatttatgtaacaaaag
 tttcttacgtgatagtagggttagagcaactctacttgaagccatccctc
 tgagcaatttactatttattcatcatgtgaacagtaactgtcctaataga
 49001 acagtaactgttttttgcatcttcacccatgcacttaaatagccctagca
 atagtaataaaaattattatattttttttacaaaataataactaaataa
 atttatttgaatttccgataagatattttaaactcgttctcgttgcgcca
 cgtgtcatttaccacaaacgactattattagtcataagatttcgacaaaga
 ctcgaactttcaaaacatgtaaaatcaaagtaacaaactttattaactgg
 aaattcacacccaaaagcacacagaaaaatccaagcaagatccataaggc
 agcatgaattacatacacaatgaactgggtagaactcaacttatcatcaa
 aatcatttctgttttatttgtaccatcatataaaaaccattttctttaa
 cagcttcaacatccaaatcccacaaaattttttaaacataaagcaacaaaa
 tacacaatttaacgcatattccaaatcccaatcacaaaatgcatgagaga
 49501 gagagagagagagagagagagagagagagagagagagaagaactgactt
 gttgtgacaaaacccaactctcagaacctagtgaagtagtcaaggaggc
 tcaagaacacccagaacggagatgagatttttccagtcagtgattcaaa
 gtgaaaaaaaaaatggatatttttaattggcagaagatagatttgagggt
 gaaatattggaaaccagatcaaaaagtaaaagagatttaaaagagattggg
 tattggcatgtgagactggcaagctgctgcgaacctaaagaaattggttc
 aactaaaacgggttcgactggaccgacgcccaggcatggcgtcagcctgac
 gtcacaccctcgggctgaagggttagcaatccctcacgggctggccct
 cgggctcccacctcactcgggctgcccaacgctgaagccgaaccaccg
 gccctcgggctctttcctctcgcctatcccttgagcaaggtccaaccgtg
 50001 gacttgctcttaaggggtgattttcagacgatcagaattcattccaaaac
 ttgtggatctaattgaaagctttgaattattcgaataacaacaacttatt
 atgtgttaaattatccaaactaattatttgttacatgaccacttgtagat
 gcgcatgtcaatattggagggttggttagaaccaagaaatttgcaaga

MC-12SSR744aFor

MC-12SSR744aRev

acttctgtaaagcaaggttcaaaagctcacttaaaaaggagaattctta
 ttaaatgaaaaacaactactatgggttatataatagctttactcttgaact
 agaagaaagagcaactgcccacgattttacaggtcctataattgtggtaa
 agactatgtcaaaaaatagagaaaatcaagatagggattattggactcct
 acatgactaacaccctaaaaactagaaacctagtagtaacggctagaa
 50501 atcttaacaacttcaacatcctcccttaaactgaataccacagatattca
 gtctactgaaacttgcaaacacccagtaaacctcgcaccttcagaaaagc
 atcaaccttcaatgggtttgttaagacatagcaatttgttcttgagttg
 aacactgcaaccagctccaaaaccacttttgaccaaatctcgaagaaaat
 gaaaccttgcacatcaatagcttgctacgaccatgcatcacaggattcttt
 gaagtttgatcgtcgaaacattatcacaaatataccaaggttggactaca
 ctgctcctgattaaagactctcaagatccttcttaaccataacaacttggc
 atatactcgatgctgctgccatgaattccacttcaatgggtggataaagta
 acaactgggtgttttttttaggaccaagaacagcacctaaactcattat
 51001 aaacacgtaacctgaggtgctctttctatcatcttgaacaccagtatagt
 cactatccgtatataccaataagctcatcaattcctccctcttgtagaaa
 atccccaggtcaatagtccttttaacgtaccttagagcctttttgtgtg
 ctgcaagtggagactcagttggacggtccatctatctactgatcaactaa
 atcgggacgtgtggctattatatacaataaggtttcccaccatttgtttgt
 aaacagtgctattagcttcgactccttgtgagtttaaaaccagggaccat
 aggattaagaactgagttacactaatccacggttaaatctctctaagacct
 cttgagcactacttctgctgactgataaagatcccatcaaaaactttgtatc
 acttcaataccaaggaaatatactcatttttccaaggccagtcagtcgaa
 ctctaccatcatagacttcttaaattgtttgaacattgtttcatcattgc
 cagtaaaaatgaaatcatcaacataaagacacactatataaaatttacct
 51501 cattctgctgtcttaaatgaacaaggtatgctcataaggacacttgttgaa
 tccctccttaatgaaataggactctatacgaactatatacaggctcggggag
 cttgttttaagtccatataaagcctttttaagtcgatataccttagactca
 taccattttgttcatatccaggtggctgatcgataaacacatctttatt
 gatctccctatgcataaatgcagatttgacatcaagctgatagatcacct
 aattcttttgagcagcaagggaatgacaattcgaatgggtgctaggtga
 gccactggtgcgaataacctctgcataatcgatctcatatcttgacagta
 tcccttagcaactagtcgagctttgtacttgtttatttccacattctcat
 tgagttttttcttaaacacccactttaatccaattgttttccctcatggg
 ggtaactccatcaattcccatgtactgttctt**ctatatggcttgtatctc**
 52001 **ttgatccataacttgtctccattttccaaacttcacatcttcttcaaagg**
tcgtgggatcttgattagtaaataacaccaagtgcagccatgctagcattc
tcactcaagatagaccttgacctgttcataatctctcatccaaattgg
tggctctcgagttcgtcctatacagaggtgagatctcatcattttgaatgt
ctccttcaaccaatgagccactttgatcagactcatattcagttctctccc
acaacttcatcaatttaagacttgaggtcaatttccttcttcaatagct
aatttcttcatcaatatccactcaagatctgccaatatagcctcttgt
gactgtcatcccaactccattggtgatcatcttcaaacacaacatctcgg
tttataactattttatgtaaaacaggatcaaatagtcgataagccttaga
ttcctcactttaccctaagcaaatatgcacttataactcttggcatcgagct
 52501 **tcacccttttgttgcaggtacataaacaatgagcaatgcaacccaaaatt**
ctgaaatgatccaacaatagtttgtgtccattccaagcctcctcaggtgt
cttgttcttcacagcaaaagttagggcatcgattcaacacatgcat_{cgctcc}
 gattcactactgcatgccagaaagttttgggaatttactttgctgataac
 atgtttcgaccatattcattattgtgctgattcgtcatttctacgacact
 gttttgttggagtgtaagctgcctatgaatccattctcatgacagaaa
 tttgtaaaactcatgcgaggtgaactctccttctcgatctgttctcaaaact
 tcttatgaatgccccaaattctttttcaactcgtgctttgaaggatttga
 atgtgccaaatgcttttgacttttccattaaagaaatagaccatgttttt
 53001 atgctgaaatcatcaataaaggtgatgagataccttttgttgttgcga
 ggtaaggttaattggctgcatattatcagcatgcatcaactgaagaatct
 tggaggctctccataagttttctttgggaaatagatctcgttgatgtctt
 cctaccaatcaatcttcacacacccttgagaggctgtgaagtgaggcaa
 tctttccaccatttgcctttgttgaagcaactttaaagccattccggctga
 gatgcccatagcgacagtgccaaagtcttgccttgatctgtgggtcaacgag
 gagaacaattttgttccctttgggtggacaacaagccagcattacaacat
 tctattatgattcatcttggctccatgattaaacctctctcagagtgaa
 agatcttgcattttccatggtgaatgagattgtgaagccctttatcttgc
 aactgactaatgcttaacaaattatttttcaaatttggcacgtagaaaac
 acatgtgatcacatgcacaatgacattcacctccattcgcacatttccct
 53501 tgccctgcacgtcaaaactcgaatcatttccggtttcaactgattatctga
 aatgctatcaagctcacaaaacatgtcctgcttaccacatatatgattg
 ctgtagccagaataaaaaaactatagtgtccattgttgcctccttgaa

Hypothetical protein
 (ORF13) ↑

ctccacaagtgttatcaacaacatctcttctcttctccacatagtttg
 ttttctcttgcttaggtttcttaggacactcccactgaaagtgcctagt
 tcatggcagttatagcattctaggggtgacttatcaaaccctagcctacc
 ccgtccttttctctgctcgataagttcctcgatctcgtcctctttact
 agactccatgggtgat**ttagagtgcttgttcatccaacacatgctgattt**
atcctttgctcatgaccaaagactactatgaagttcatctatggacag
ggtgtccagatcattagattccttaattgaacatacaacataatcatatt
 54001 **tcaaagtcgtggacctcaaaatctcaatgatccacatcatccatc**
ttctcccgtgagttctcatattgttggttatggaaagtgtactccaaa
gtagtattgattgtttctccattcttcatgtgcagcacctcgaactctt
tgcaagagtcgtgaattgagcagccttacctgtgcaatcccttgatat
ttttgtttcaaggaatcccatatatccttcttcagaatggtctccaatat
tgaccgatcaatggcttggacaaatagttcttgccttcaaatccttca
gcttcagttcatctagagtccttcttctgacttactggaatctgatccc
cctactacagcagcgcgcatccccatctccaccaagttccaatacttttggg
 acgaagaaaattctccattagcatgctctagtgttgtagtgtccatcaa
 accttggaaatggctggttgcaagaaactgttctcaaataccatgtctgt
 54501 **ttcttctactctagcacctgggttgcaagctttttaaatacaagccccctgac**
agggtcttgataccaaatgttagaaccaagaaatttgcaagaaactctg
taaggcaaggtttgaaaactcacataaacaagggagaattcttattaat
gaaaaacaactactatggctattaatagccttacacttgaactagaaga
aagagcaactgccacaattttacaggtactataatcgtggtaaagacta
cgtcaaaaacagaaaatgcaagcttttccctcaagataaaggattattga
actcctactgctgacgacccctaaaaatgagaaccctagtgtaacgg
ctagaatcctaacaacttcaacaagttggtctagacagctctagactcat
gagtcagtgaggcagagaaaaataaagaaagaaacgatcagagaaacgat
agaaacgaaacataaagtcactgtatgtatattgagcattgacaactca
 55001 **accttcccttaattacctcttcagtcactaagattgagatcatatttg**
acagctaacccccctgtatttttaaatcgtcaaaactatttattttataa
ataaaatctaattgggacttagaagaagcaacttggaaaaatattatgcg
tgttgatttgagcaatctgatctgatctacataaaatcttctatatata
tttgcgtaaagaataagtttcttaattgtgatcaaatgtctaagcctgtt
ttttaatgtaataaagaaatcgttcaaattgggtgtaagcctaaaactagg
attaaggcttttgttttcttttcttttcttttcttttcttttcttttct
aatgggtgaaatgtatagagtcattgcatgatgggggtgggttagaactga
tgagtagattttacattacatattttaccccaatcttaatttttttgggt
tattattttggaagaattttgatacttttaaatgtattttcaatatatga
 55501 **gttccgacttcttctgggaaaaacgtgaccaaattgatgaattttggag**
taattctagttggaggacgttcgtgagtcacttagcttgatcgtgtcaaa
atatcataattttccaccaatcgattattttctgacaatagaataaaaga
gcagtgccgaatgctaaagtgacattttgggcttgattacgagctttgga
gcccaagatgaactattttggagttggcccttctggagatgtgttcaga
atgttccgatcttataaaaaagctatcatgggcccggatttgggaagagat
ttgagggtaaaaactgactgaacagttgccttgcagattctcctagctta
attagaattttattctctataaatattttatttttatttgggttccctagt
tgataaaaaggactttattaggattatgttatagctgatgagaggggat
aaggctttggctggccttactcttttgacgacttcttttaggcgactttg
 56001 **gagatagaaaattggctaaagttcttgaagattttcagacttttacttta**
aggtatttttaattctttttttccataataatatttttaattatgaata
tgcgtaactaatcagttttgttaggacgaagccgtgagccttaacatgaa
tatgtaatttttatttaattgcttgtgattgattgcatgcatgctttgaa
ttatttactgctgttttaactatctaattgtcataatgtttgatcacca
atagctggagctaggaatttttactgagtagatcttatgcttgatcacc
aatacggagagctaggaatttttaccgggtgggctagcttaagattta
tatctttataacaaaaaacatgatattgtattttcacagtatcagcta
gctaaagaaaattttgacaaggtgagccatgaatttttggttattaacaa
atgtagagaaaaaagaatgtaagaaaaagagatagtttgaactgtat
 56501 **tatatcattttatataatcaaaccaagagaatttggattagtactaaat**
atatggtaagctacattcgaaaatcaataaaaaatttatatgtcaatttta
tttttactaaaagctacctgggctatagccaaggtagccttaagtg
gctccgcaagtgaccaatgatctttgaaaaagtaatttgatataattt
tgggtcaaaaggttctctgaaattgacgctagcttcatgtggttagtaatt
gtaatttctacttaggatgaacatcacgttttaagggttctattgttttct
aaaggttttcataaaactaatgagtgtttatgttcatatttgatcca
aagctccagacgcttacatgttagatatacgttatatgttggagattcc
aagtaggacatatattaagaaaatctaaccttcaaaaatagcatgggtaa
 57001 **ttcataagtgataggaagaactacgtatgattgttaaagtgatagcagaa**

Copia-type polyprotein
 (ORF14) ↑

cccatttgtctctttaaattgattttctttaaattgttttatagtcatt
 tatttaattatTTTTTattTTTaaattcgtTTTTaatattTTTaaatttcaaa
 atcaacattttccaaactttgTTTTaaataattaattaagagttggtttt
 aaatacaaaattattcattcaattcctgtggaaaacgatcttacttcaact
 actataatcagcaaaccttacactcttacaagtattttaagtgttttat
 cactcattttgcaaattgtaaaaatcataatcaagaatcttattgttaact
 aatctatcacctaactctgacataattttataaataaattagtatgaggagc
 gatcgaatcgaaattgtttcacttcaataaaaaactaaattgaaataagct
 aatatacaaaaaagcatgcaagctataaggaaaaaaaagggttaatta
 57501 caatttaccctctcaagtttgagggtcaattacgatccatatacaacatatt
 taaaacatttttaattttatataatttaactatttcatttcaatttcatt
 caaccattacttaactgttaaatcaaccgttaaaaatagtgtggtgcaa
 atatatggcccacttttgatgactgtcaaaaataaattgaatatt
 atttaaaaaataaacctaaattaaaataaagggtgacccccaccagaaattt
 gttcctgtgtcatcgttcatccaacactgccaccgcatcgtcgattgg
 agccttgaccctgccctcctcatcctcagattccacggcctcgccacga
 cggctcggcgatgtgacctatggcgttatcagctgcatagcgtttgac
 ggcggtcgtcgagattagtgggaaccattcttccagatccacgccc
 cagctcgtcgccatcactcgattcctcccaccccgctctatctgtaa
 58001 cccacccatcctcgttagaaatcaaacagctaaaatcacttttctcca
 actaacctaaaccctcgctatggccctgaaaactcgtcagcaaaacc
 agatgcaaatctgcttcccaaatcccactcttaaaatcctcacttcccta
 atccccaaacatggtgaaaaccaggatgaacccaacgcccctcacttccc
 aaatctcacagccacatacaaccaccctctcatccactttatccatcaa
 caatccctcatccttctcgtcggatttgaatcggacctcaagccaagt
 gtggatgtggtccaggcaaccagctggaggacagcgtgaagaagctcaag
 agccccaaaacttcacctggtaggggaatttggagccgcatccgcacc
 ttccgcacccctccgctctcaggtgggtcgagcgggtccgggttgtcaa
 gcacgaccgagagaagaggatttttgggggtggggtatgggagaaagggg
 58501 tggctcgggtgagatttttgggggtgggtgatgggtggggaagacgaatg
 agatttttgatttgcgcgctcgttatttaacgatttacttaatgaattt
 tttaaatataaatttaaaaataaaatgataagtaaatataagaaattga
 aatatttttaaaaataaaaaatatacaaaagctagacgttaaaaagcgt
 tccagacatttttgtcatcatgtgctcatagtgcaatcgtccatgac
 acggaccctacactcttactgatcatgtccctcctgttatttgattt
 ttggatcattactaacaaggcgcaatctctcgaaaaatacaattaattt
 taacaaggctaatttttttaatttgaattcactgggataatgtaggaga
 ctacataatttatactatattgatgtaccatttatgtgaaagtgtatata
 aattaattgatgtggcagctacaaatggtagcattaatatggtagatgtg
 59001 tctcctaacgttgaatagcgatccctcatttagtctcattgtgtctgtca
 aaaaaagattgtatgtatgaacttatggatccttaaccaaatgaaga
 tgtgttatttagtcagttgccaattatttctgtgttttttttggaaac
 tgaaaacttgttagtaattaatTTTtagttttaaaaaataaaaaatca
 aaattgaaagttacttttatgagattttaaattttcgttggcttttaaat
 atcatttagtagtatggtctagcagattctctcttcttgtgaagtgaga
 agtcttagactcgaatctccttgatagtgagttcgaaaccaatttatctc
 accactaagtgtaaatataatagtggttttcttttacaatgatagca
 ctagtgaattaaagttgttaaaaggagaaaaactagtgagattaaccca
 aactatgatgcataaacatgattgttttcatcatagcggtaaacatag
 59501 tatgtgtaaatatctgattttgaaaaaaaacaaaaatttggcttgc
 aatattcaaagtagtaactaaacaattttcgttttaaaaactaaaataa
 ttatcaaataaaccctgtaaatatctgatttttaaaaaaaaacaaaaa
 atttggcttgcaatattcaaaagcagtaactaaacaattttcgttttaaa
 aactaaaataaattagcaaatagaccctgaaaaaaatattcattaaaaaa
 aaaaagccctttattaagaagcagctgaaacgcctcatgcattgtgtca
 tttcaataaaaaataaataaaaaagattctttcttttcttctcctgtag
 cggattgggaaatcaaccttggaataccgccccttactgtaaacataaa
 tctgtaacatcctcaactcactcgttatataatattcccccccgccat
 tttttacaaaaacaggccatttgcttattcatttcatttttgaatttcc
 60001 agc**atggggtgtcttcaactcgtacagatgagataacttcccaatccctcc**
ggacaggttgttcaaggccttggctccttgatgctcacatcctgatcccag
agctcatgccagaggctgttaagagcattgacaccctcgaaggtgacgga
agagcaggaagcatcaagaagatcaacttcgctgaaggcattgtaaaataat
 caactggctaccatacaaacactgaggtgaaatattctcgtactagactat
 aactcaaaagattctataaccactatttgagcagatactcacttttttt
 tttcataaaagagaagattaggcagatttagctcttttaaataggatgat
 gcaaagcatccactctcaactagctaaagtgaaagattgttttgggtccc
 ctttgcgctccctctaataaataatacaacttatttgaatagttgtatca

Mal d 1.12 ↓

MGVFTRTDEYTSPIPPDRLFKALVLDA
 HILPELMPEAVKSIDTLEGDGRAGSI
 KKINFAEGSQLKSVINRVDEVDENF
 VYAYTLVEGEPLVVEKLEYITYKAKFE
 AASDGGSKNRLVSNYYTKGDIVLKEE

60501 gattgatagtttggccggcaacataatatatgcccggagaacacacccc IKAGREKALGMRYVETYLQNPDAY
 acatgttgatcattttataatgtaaaaagtgttgaatacatgtaataat A
 tttttcatgtagtagccagctcaaactctgtgataaacccgggtgacgaa
 gtcgacgaagagaacttgtatatgcttacacttgggtgaaggagaacc
 attagtagtgggaaacttgaatacatcacttataaggctaaagtgaag
 ctgcatcagatggggtagtaagaaccggttggcagcaattactacac
 aagggtgacattgtgctgaaagaggaagaataaaaagctggcagggaaa
 ggccttggggatgtacagagttgtgaaacctatctcctccaaaatcctg
 atgcctatgcttaaatttttaggggtggggtagggggcggtgggggtttt
 gccttcactagcttgtggggattgataaatagaaatagcttgacaatgc
 ctgtgactacggccattggaacatacttctccatgtattgacttcaata
 61001 aaaagaatttctccatgattttgacatactcattttcatctcctacaaac
 tttccgtaatttccggtcatttgatcctcctctttcaacaatttaatggt
 cagaaataaataaaaatgtacagagaataagatgggtgagggaataat
 tgccccttgctcttatttatttataaagcctataagggttagaat
 gggctcagctgaaaggacataaactatggggccacaatctcccaaaaa
 ccaatccttgactgatgggctcagggatggccatttgaccatgggtact
 tgtatccatggattctttattctaataggatcataaataaaatggatta
 gatgggttgggtatttgggtgagatcaagtcttctgcatcgaaaataga
 tatgacctctataactctaataagtttgacacataataaatttattac
 aataaaatagaaaagtgaataaataaatacatgtcagtcacttatta
 61501 aaatcacaagaaccacacctattttgagtgcataaaaatttagtctctt
 tttgttgggtctattatacccacttataagaaaagatcataaacact
 cctctcatggttccgctccatgtctccttcccattgccttcattgtctc
 tatgtctctccttctcggctccttcccgagaccggcaaaccttcc
 attgcccgcactcctacatccacctcctcctttccattctgtctcttct
 cttttgtttttgtttttgtttttgttttttttttttttttttttttt
 tttgtttttttgtttttgtgtcttctccttagcatttcagatccg
 caaggaaggaatgaatgtcaaccaaattcttcttcttcttcttcttct
 ttcactataattgacgttaaaaaagtagtgaatcatgttcaatataatt
 gacgttaaacgtgtaaaacaacaagagttataaaaggtataactttgaa
 62001 aagtcacataagcatttctgataatagaataaaatacaacaaaaagtgat
 agaggccatcaaatcataaagatcagagatgctggtgaaagctctcatg
 ctaaaatgatgaagaaaagaggaaggttttgggttttcttggttctgatc
 ctctgctgcctaaaggctggaagcatatggtgagggaccattgataca
 agtttcttatttgattttttatatttgttctttgatttagcctcacatg
 catggcacatgactcattttaacgaaaattctaacggagttgatgaaaa
 gacaatagtgtcgttttggatgagtgaggatcaatggtaaatgaattt
 ttggtgagggaccattgctctaatttgattaaagctgagagacaatttc
 tccaattcgattaaaattgagagaccattgatacaatttaccaaaaaaat
 atcaatattgactgtaagctatgattataggtaaatatttttttttta
 62501 gtcaaggtcaatattgttatagtgagtaaaacaaagagctaaatcatata
 tgatttgatcatataattcatgtaagagtggtactaaaccatccattgt
 tttatttgtccaccaattatattcccacactataaaaaggtaataaa
 taacttgctattttctcaccaccattattccaattataaccttttaa
 tatcttaagttctttaaattttatatttctgtaatatgataataaat
 aaatataatatacgtatatacatgtaaatatatttatctcaagaaaa
 aaaccaaccatgggacatatttagcagtaattcaaacacaacaagaat
 ccatcaaggcagacatggaaggtgtgacgaccgctcctacttttcactg
 taattttcttttgggtgtaaatgacggctatacccttattggacaa
 63001 aaatggaatctaacgtggatgtggttttagctttcagtttcattttcct
 gggttcgtactaaaatagtagtactcatcgctatgaatgcgtaagcgcgtg
 agagtcgaatcagatctgtaacgaaaaagttacgctcgttaagatgt
 tagtcacaggtaattttatacacgcgcgtagcaatttactgtgtcaaac
 attgtcgaatgacataattgtacataaaaatttgtacaaatccatctaa
 tccttaaaagggtccggtcccttcttccacccatcaaccatcaaac
 tttgcttttcttcttctcctctccaccaatcaaatgttttccctcaa
 ccttgcaccaatcagaaatcaaaaggttctcctcctcacttaaaacttt
 catttcaactctctctcgggtgaacaaccatagagacaccaccaccac
 cacacaaacgtcgtccgacaacttcttcaactctgtcacgccacatta
 ccatcaccactactagcaggttgcctccgatgagatagcgagtttctca
 63501 gctaagaactaggacttacctgccaatgaccattaccactgctcactt
 tcttctccagcgaacttaattttctgatcaagaaacgatttgaagaaa
 gaaactcccagatttttggaggcgaaccgtggccatcaggccactttca
 aaccgtttccttgatcaacacaaccacgactgagcttcaaataggata
 gccctgttcaactacatttctagctttatattcatttttctactgaaa
 atgggttaagaatcgaactagaattagccttcaaagttgggaagaattc
 ctgatttctggaatttggatccgtggtcatttggccactttcaggcca

aattcctgggttaacacagaccacatttgaactttttgtgaatatggatcg
attcataaacattcctaattttaatttggctattatagaagtgttttgg
64001 tgagaatcaagctcgttaggggctcgggaagtggccaaaaaactgcag
aaaatcgtggagaatgcgagtatacagctattcacaagcgcgatgaacgcg
cgtgaccaccacatgccatcggctcgtgcaacgcggtgtaggttgtgca
gcttcttgtgacagcagctgacagcgcgtatggcttccggccgacatgtg
ggtgacacgtgtgtgtctgtgacgtcgagtaaatcgattttatataat
aaccctaggtttgagcaatctacggaggttttatttaagtttccatgat
gttttattaaactaagttaataagtaatttcacataaagggaaacata
ctcgaggatttggcagggttaagcaaggctaggaggctcgtttcgactgc
gtatctgtaagtgggcagtttattatatacctatacatatgaaagtcc
64501 ctaaatgtggagttaacttcaactttacgcttatattgccaagtattcgt
attgtgataatgataatgctgttatatgggtgggatattactg
tcatgactttcatacatatctgtacatgctcatcttgcctgactgatgt
agtactcaccctagggtcaggccagtccttcatgtgatgttcacagcc
ttggatccaagttagggtgccagtccttgtgggcatattgcattaggcgt
ccgatttgtatgtgacgtgcttctgcaccagtcctcagctgatcgtagta
ctagaacgtaaatgattacaccagttctgttctgtgcaaaaattattgt
tcaaaactatgtgtcagcttagatggatgagcacttagctatacgtgat
atcacatgactatattactacaacatttgatataatcatgacttggcatat
ttctgggtataatgctgttattattgtttatatttgtaccatacttga
ccaatcccaaaactactgagcaccggtcaacgttatacctgcaaggacct
agaagagtttccctccaaccaggaggccaatcatagtagcagacgtgtca
65001 acatcaaaagccaatcatagcacaacacgtgtcaacatcagaagccaatc
acaacacaacatgtgtcaatgtcagaacaagctagaactcttctataa
aaggagatcattctcccacaatatttccaaatgtcatttgtactaaatca
ttcacttgtactcactaaatgagagcttgaacctatgacttgtgtaaac
ccttcacaattaatgagaacttctcactccatggacgtagccaatctag
gtgaaccacgtacatcttgtgtttgcttccctatctctatccatttaca
atctctatccatttatacacttatacacttagtgaccggagcaatctag
cgaaggtcacaacttaacacttctgttgtaccaaaatcctcactgatt
ttgtgcatcaacagtttatacacttatacagtagtattttaaagaaactat
acttgtttacgggtgagaggttagtataataaaaataaagggtttttaca
65501 aaacggtgttttactgaccactcaacttgttcttccgcccctccaggt
tagtagttgtgcttgcgtgtaaacgaagactttggcaagatcttagtaga
tgggtaccttagcggataaacctcaccctaattactgtactgttcttat
actctaccattacatgtgaaatgggtgcaatcccgtcaccgcacactc
ttttgttttaggcacttttaggtttaaatttattcactttttcacatcac
ttacccttatgggtacgtcacctcacggtgatggccaacatgcttcgacc
ttcccaggtcatgggtgtgtcagaaggtccacagttacctattttctaggt
caagctcgaagaaggatggagatggcaaaagatgtactagtacataata
tcatcataaaaaaccaagtcaagtgccaatataagattgaatatcaaac
tacggacaagaaatagtagtatgtcaatgcactgctgcaacttgggtacat
66001 gtgttagtgagaattacttgttgccttcttctttttattggattga
gggttagtctcgaagatatttaagagagaaaaaggaactgaaatgagc
agattgaccttgtttagtcacaaaatcattgaggacttgggtacaacag
aaaatgtaagtgtgtgaccttactagattgctctggtcactagtgtga
ataagtagtaaatggatagggacacggaagcaaatacaagatgtacgtg
atcaccagattggctacgtccacggagtagaggattctcattaatgtg
tgaaaggtttacacaagtatataggttcaagctctcctttagttagtact
agtgaatgatttagtacaatgacattaggaaatattgtgagagaatgat
ctctatttataagaagagagtttctagtttcattttgacattgacacgtgt
cgtgttatgattggcttttgatgttgacacgtgtcgcgctatgattggct
66501 tctgatgtcgacacgtgtcgcgctgtgatggccttctagtggaggaa
actcttctgggtccttgacgggtataacgttaaccgggtgctcaatagttc
gggattgggtcaagtaggtacaaataacctcaaaagcaaaaactctcgcc
actcgctgcaaaaacaatcaaacctccaaaacccaacaacaccactacc
cccaagtttcagccgtgccaacacataagccctaccccatgcccacttt
cgcagcaggtcatagccttataaccaccacatgcatggctccaccat
atattccccttgcgtgtacacaatctccattccgcccagaattagattaag
gctgtcatccattgtgagagaaagggatttgaaaaggaggagctggca
tcgacgccaccgttagggaggtgtgcgggtgtggattgggtgggcttga
cagtaggggtgggttttttaattttttttttcttttttttctgcccattt
67001 caactaatgatgatgtggcaaatgttagtatttattgtttaagtcctta
aatgtcattttataagaggagatatttgcattgacatttatatttttaa
gtggatagagtataagacagtggggtgggcttaacatcttcatgtaatg
agggttcgaatgttttttgggttaataaacaagatttaaatatattcacttt
cgatctttacagtaatgaaagggataaaaagatttctgatcattat

67501 cagtggaagaatacagttaatttccttccacgtatatttgtttgagctat
 tttgatcttataggcttgacgtgcaatgattgatgtcatttaagatta
 ttaaggcgaggaatattatggttttgacactaatcaatcttagtactaa
 tttttgggtgcgctccctattttaaatagtttaggctatcttatttttagc
 gactcaatctgaaacaatagtcacatcagttgcccataacattcaagctt
 cctccttgtgagaaaaggaatagaattttattcaaatattgcctatttta
 taaatggagataaatggtaataatagattggttgcacacaataggtgt
 tacaaaatttgaagaggtgatgggtgacttaacatcagtagggaaaca
 taattggtcggtagtggtggtcggtagcacaattggttccaaagcttgaa
 aaactctacgagcaccaaaacttgaacattgcatggaatataatgtagtc
 ttgacaaagttagatttgattattgcttaataataagatgagctagatctt
 ctttcagagcatgaaaataaaagatatcatgtaactttgaaaattccatg
 tcaactagaagcctgacatttcaagcaatcttccctcctaaatgtaaaacat
 accatgtacctttcagagcatgaaaatttcatgtgactccaagtcatcca
 gttttgcatttttaaagaggggaccatcggattccctccccttgtcgaaa
 68001 agtccacattaggttacatttgggtccaactatacatcgcatccaccat
 tacatattgcatcgggaaaaaataatattcataccaacaacttccaagc
 aaagcagcagcagcccaccacctaagttttgtcttcatcattcacattc
 aaaaataatataatcggataacctttgtctcctccctccttgatatttt
 cacaatcaaaaccttttttttctttcatcccaactcccgagtgttgtt
 catcaagcaaggcaagc**atgggtgtaccaagatcagccagaagtttgtg**
actcaggtgacgccacagaggatgttcaacgccttgatcttggatgccca
caacatctgccccagctcatgttctcatcaattaaaagtattgaattcc
ttagcgggtcaggagaagttggaacctcaaacagatcaacttcactgaa
 68501 **ggtaaatgttaggcaaatataaagtcaaacctctttttcttggaaaaatt**
ggataataaatcaaatctcaattggtgaaaccgatcaaaataaaccaaat
attaatgatatagacaataatattgaaaaacaaaatcattttggtaattt
tccattttctttctctgtttttacatgaaattcaatatctaatttgttg
gggatgtgcagctagtcctatgaaatagctaagcacaggattgatgctt
tggaacaaggaggcactcagctgcacctacacctcatcgaaagcgatgca
acggatcacttgttggaacaagcttgaatacataacatatgatgtcaagtt
tgaggggtatggaagaggaggatgcatctgtcatttgaccagcacataca
aagccaaagatgatatacagatcaaagaagaagacattgagcttggaag
gatagagctattgggatgtatgaagttttggaagcctatctcatggcgca
ccctcgcgctcactgtgaaaatttttaagaaaactttcagtagtcgat
 69001 gagtttttgggaattctgctgctgcaatataatctaacaatattggtcatg
 ttgttttccctttttccgtaagtgcttttcttctgttctttcctaaact
 ttattctgtgtgacaagattgtgattaagaaatatttagtttcggttct
 aatttacttaacaggggatctgttgtgtagaagtagtttatccattttt
 tattatcagatagttcctccaaatatactcagcggcatccatcttaagtaa
 agtggaaaggttaaatctcagatgataaacagattggaattctagccggaa
 attccttaattggcctctgtcaactgtttttatttttagtaaatcacttctt
 ttgatttactggagtggaaatgattttaaactcccgcaaaataagaaacta
 gttgcagaaaacccctaaattgacagctgatttaagaattacgaaagaga
 tgaattgagtaaaaagtggtatcagaaccgatatgagacgaggcctagaa
 69501 atactttgtaaatgtcataaaaaaggacctataacaatcttgagcgcga
 tcaaacacactaagccccaaactcaagtttacttccgagattagcaggca
 gtccatgaccactttacactctctattacactctctataaatgtgttga
 cactacaggtctatctgtgattgatcaaatcttgacatcgatcccaat
 gataattacctcaggtgtaacatttttctgattgtgtgtgtgtgtatt
 ttctctactacaattctgccttgtccaaaccccaaacatgttttctacca
 aacaccaattgatcatccaattaatcaaccacaacaagggaaagaacta
 acacagcccagtgatattcttaactttgttctaacatttaaaaagttac
 aaaagcagagagaacttggacctataagggatcccaaggatgctacag
 70001 tgtgacaaaattttgtaagcggccatccatattgttttaaggacaataaatc
 aatagtgacgggaatctaggaacgaataaggatccctatcaatataaaag
 caaaaggaaaaaaaatttttaagcccaacatatattttatctgttgaa
 tcggtgaagtttgaaacccaagaacaaaagtcttttttaattgacaat
 ggaaatgtttaagtagctttaaataatgcaaaagggcacagaagaatacaaat
 aattgatcattgataaattcaaggatcatttttattgatttttaaatcaca
 acaatcaaatgaggtgctataataatctcattaccattttggctaaaaa
 aatccttttttattaaggaaaactaatgaaatgacttgaaaactttgag
 ttttaacgataaggataaaaataaagggtaaagtgaatggtaccatgattg
 attttttagtgttaaatgtggtttttgttaaagtgaatagatcggaaag
 tttttgttaagttctcatattattaacagatggttttagctgggtcgat
 70501 ttgggatccgtaccaaaatgggatcccaattccgaaatagacattttt
 cggcatgagattctgaaaacaaaaccgaaccgatatttcggtaactctg

Mal d 1.11 ↓

(Mal d 1 n)

MGVTKISQKFVTVQVTPQRMFNALILDA
 HNICPKLMFSSIKSIEFLSGSGEVGTI
 KQINFTEGASPMKYAKHRIDALDKEA
 LSCTYTFIESDATDHLDDKLEYITYDV
 KFEgyGRGGCICHLTSTYKAKDDIQIK
 EEDIELGKDRAIGMYEVLEAYLMAHP
 RAYV

aatgttgggattccccaaaaatttcgggatttcgggtacaattcgggatt
tgtcgaccttctcctcaaatcatcgacgtggctaccgtctcccggaggcc
aaggctcagaactagaacgagatcaggaaccaccattgacctcaacgagc
tcacgggagtcacgatattcaatgatggcggcttctaggttctaacaacga
tcattcctcacagatccaaggagaaatggagctctcgagttctcgagttc
tctctcgtaccatcattgaatatcgtaccatgttttttcttagtctgac
ataaaaaagatacctgataaagagaaaaatataataactatctacat
71001 ctaaaaaatattcctataactaaagttttaaagtgtttagttggacggcg
gacaaaggcctaacgcatagacgtttatgcagatttttttattttatta
actttattgtataacatatagcactataacatatataattatactataat
acataaatttcacatgtaaaatatagaagtagggtgggcacggggccga
gctcaattttgaagggatcgaccggatccggattttaaagaaaacggggc
gagccaagacgggtccatcaaattttaatacagtaaccagacctaaccgg
atccgttcaaaatgggccaattcggaccggatccacgggtccttttttt
tttcgggtctcttcttcttcttttagatctcccgcagagctcgttgggatg
gcgtgatctactaccacatgccaccaactatgtgacgagatcaacgtcgt
71501 ctgacgacggcaacaaagtcaaggcatcgacgacaccatgggacctggaa
aagcttcgoggcctgggacctggaaaaggggttggctcttcgatgaag
taatcaacagcgtcaacgacaccgactgctcactcctctcccactcaa
ctcgtcctcgtctagtttggaaactcgtcctggaggctgtctagctctt
atctgcaatctgaatcatcgtaaatcctaattgctgcaagattccagag
agcttcaaatgttcggggacgatttcgctcctgaggctatcaacatggag
caaacctcgaggcgaaggcgtcgtctcctggatcgacgagcagttttgt
gggtgctcgggagcactgggtgtggaggaggagatggcctctttgacat
cgtccttaattggagctccttgatgggaggctgtcaagcaaggtagag
aagttgagggtgagcctagggaaaatgtggcgagggttctgtcaaggg
gccacattcggctcccgtttgaagagtcctctatgaaggctagatcctt
72001 tggagtaggatgtaaagctcctgaaagcagaggaggcggagaggagaa
gatgggtggtagctcctgatgatgtcttcgaaaatcaattgggggtgg
aaatcggttcgggccaaggcccgcttcttccaggatttagaccgctcac
gccccctattaacctaagaccgctgccccgccccatgttgaattttaa
aattttgaccgctcctaaaccgcttagaccgccccgatctgtgggtcc
aggatccgtttgcccaccctaatatgaagtaactataacatattaaaaac
ctggaaactttggttacaagcacataatgattgttcatccaagttttttt
atattttacatagctaacacatttttagatggatgaatcctttttctga
tatgtaaaaaatttcacaatcaatagctacgataccttttttaaaaaata
72501 taaaaaataatagggtattgatagctataattaaacgatatttttgaat
gtaatacagctattataactttgatagtttaatatgatgtccacaacgtaa
tatcatcattatacagcttaaggagccaaaagttgaatttaacgtca
tgtttgtaactgttggatgttagaccctaatccaatagctaataggctcact
acagttaataaggattggtttggaaatgaagtagggttagggtttagtta
atatgctataacttggactccaagtaaaagttaatattccagttaatttgg
gagaataaggactcctgatctatcctatttgaacagatatttggagatta
atatctttacaataaagaaccttactgtcatcattataggaagatacat
tagagttaaagttagtataaataactaagcaaaagtcctgcacatccatc
gaaacaaactccaaagaattcccttaggttcgagttgtatagtggtgggtg
agtgtgctgaaagccattgctaaagttctagggtgctcgaagttgttg
73001 tgtgctcgtgatttgctacactgctaagaatgacttctcaaggctcagtgct
ttgtgggtgttttaactacatgaacttgctagggttgaagtaaaaccag
taaaacataaaaagtaaggaacctaaacagcccaccgacattgtaagggtg
tctttatgccgaataataactcagcatcttcttctctcttctgtattat
taaccgaataaataacttggctgcttaaagatgagtagaaaactcttaatta
ctcaaaatttatcaaaattgtttgttctatttatagaacttcgtgttt
ataattaaaaccggttcatattataaatcacactataaagattgtgtatgc
aagaaatcaataaaaattaaggctcgttagtcatcaaacgttttaaaac
aatgaacggtatttggtaaaagcattacaacccgtctatgtatttgcac
aatagattaaactaaatgaccttagtttaatttttttttaagagataa
73501 tttttatagtgatcattcatagatgaatagttctgatcataagcaca
ccctgtgattaaaacactaaaagttttgagtagaaatttctacttatctt
taagtagccaagttgttctattaaccctcaattcttgcacatttccat
tggatttctgaattgattgggtgcttagacaatagaaacgaaaatatta
ttttgttttggaaaattaaaatattaataagttggcccaccagtagaaa
taagttgatttttctttttctaatcaagttctagaattttattgggtggc
tttctttgactaatatattatgcattgattttttttttatgccatgca
attgataatataatattcattttcatttcacatggacagcatccattgaa
aacttcaaaatttgcctcaactcctttacacatgtaccctcctcccga
tctataaataccaccctcggaggagcattttctcacatcaacaccgat
74001 catcctcttgcctcctccttctcattcttaatttcttgcacattttcta

agctcttaaaaatcatcatc[atg]ggtgtcctcacatcgaacccaatac
 gcattccattatccccctgctaggtgtacaatgcccttgccttgatgc
 tgacaacctcatcccgaagattgctccacaagcagtaaaactggtgaa
 ttctcgagggagatggcggtgttgaaccatcaagaaagttagctttggc
 gaaggtttgtttagttccaccattttccatttaactgataattttac
 ctatttaatatatatggaataaggctttattggtgaggtggtgtaatat
 atatctcgaaaacactagctaataatcttaattttctgtttcattttccat
 ttgacagggagtgaatacaactatgtgaagcacaaggtgaggggaattga
 caaagacaactttgtgtacagctatagtttgattgaaggagatgccattt
 ctgacaaaattgagaagatctctatgagattaagttggtggcttctggc
 agtggttccatcatcaagaacatcagccactaccacccaagggagattt
 tgaattcaaggaagagcatgttaaggctggcaagaaagagccatggtc
 tgttcaagctcattgagaactaccttggccaatcctgatgcctacaac
 taaaacctttcatatggcaacaatcttatctcttgcctttctctgtttatt
 caattgtgtggtttaatttattatcgtgccaaagcaattctttgacttgc
 actttggatgatgattaaggagtgatgttgtatttgatctttaatctcaat
 aataaaatgtatactagttatcgtaaattatctcatgatatttaactt
 ggagtattattttcccttttaattgtgtgaatccacatccgatctgtt
 acggtgtaaaactcaatgacacatagaaaggaaaagaaagttattccact
 aaggttagaccaatctaactcaaaagttgactggttgggttgttgggatgg
 aagattctaataggcgaaaagatgaaagactttaacattacgtttggacga
 agaaaattaaaattactaaggaatttaaaatgatagaaattaaaatgacg
 agattttattttctaaaatttgtaaatttcttgggttggccaacctaaaa
 gaacaatgaaattgaaacggaatttgttatttttaactcccaatcgta
 gaaattgagaagtgacacatatttatatggaatttgaactgggatttag
 agatcccaattccaaatttttttgcacgcggaattttaaaattctatg
 tttatgaattcaacaagagaattggtgcctgtcaatttataaattctaa
 cgtttaccaaaaattccatgtttatttccctcatccaaacatagcgtaaat
 ttccctgaaaaaaaaaaaaagacttgcctgcatagcgttctgggtgacca
 ttatatccatcttcatatcccacaaattaccatttcagtgggctattct
 tccagctcatgaacaaaatccttgcctagtgaggcaatttgcctatgttga
 atttcagaaaaggctgcacccagtgcaaacggctcccgtttacgcaggg
 tttgggagaggtgaatgtcggetagccttaccocccatttatggagaggct
 gctcccaagtctcgaaccogagacctaccgctcatgggcaaggaacttg
 ccacgcaccaagtgaacctctatgttgaatttcagaaagaaagaaaa
 atttccaaaagcttattcaaatggccatggcttattgaaaggggaagaata
 gcgccacaccgccaccctagtgagtagtgagggtgtttagtgatccatcca
 aatttatagaatagaatatcttcttaaggaaaatcctccctctccaacc
 aaaaacctcaaaactctgcaaaattttcaaccgtcggctccttgttttggc
 ttaggaccggcggtggcctctcctgcctcttctagtgctccttccctcat
 agcctttccaatcctcgcctccttaacccttccccgtccgctcctcccac
 gagcctcactttctttactatcattccgctatccctcccccttctcaa
 ttttgcctcggtaacttgggtgctgcccagcaaccttctccccactttt
 cgtgggcaaccaattttccaaactttgttgcaatcctcagaccgattctc
 gagatctattgocgtacagggattcctccctgtctggggctagatctgc
 ggttttgggttgggtgtcatttgggtgattgggtgggtcggaccctcctc
 tatgggttgccttctcgattttgtggctcttgggtgattgtgttgggtg
 gcaactgccttgogataaaagtgtcaatggcgagggtaaatgcgatgggtgc
 atgctaattggcgaggttatgggttgggtggcattttgacgacgatggc
 agttctgactggatccccctcccgcccccaccttccgggtgctcgtcccg
 gctacagctagcgtcccacgcctcctgccaattctgctcgtgcatctgca
 acccatgccttctgccaattctgcacctgcttctccactatggacacgtc
 ggttctttagcctaggtcctttggggggttttttttttttggctgatttg
 ctccggcaagtgtgcccgtacggcattttttagcctcattgggttgtga
 tggcttttggaggggttaggtgcttggatgtcacgctctcatctatgtg
 gcatatctacctcactcttgttctagtggtgtagtttgggggtggccg
 tttaggggttttaataatttgccttttgcctaatgatcactagagcata
 tgtaccgtctggcacctttgtcattgggtgggtgttgggtagggggtg
 gattttttgcataaattgcctgccaaccgaattgccaactgtgtcatat
 cgattttggccaatttttggtaaccaatcggtaatggtacggattgtga
 ccataccgaaatttcatggtacggatattggtaattggatattaccacg
 gtattaccgtaccataccgaaatacaatataatttataatttataaatt
 atataatagataattatataacacatatttataatattccaataatttga
 aaataaaacctccttatgttagcattgttgggtgactttgatctcct
 gaaattgtggaattcaaacacaaaagaattcctaactctttcaciaaatag
 ccaaaatattttgtaacccccacttctactgttgctagtgaaaatgcat
 ttagecctagggaggagcttgttgacccttttagggtatccttgactcct

Mal d 1.06B ↓

MGVLTYETEFYASIPPARLYNALVLDA
 DNLIPKIPQAVKTVILEGGVGTI
 KKVSVFEGESEYNYVKHKVEGIDKDNF
 VYSYSLIEGDAISDKIEKISYEIKLVA
 SSGSIIKNISHYHTKGDFEIKEEHVK
 AGKERAHGLFKLIENYLVANPDAYN

77501 aaataatgaaggcactagtgcgattagtggttgcttagggcagatga
 agtaaaacttctacaaggaatcaacagaagatagcttttattttacaagg
 aaatggaagaagagaaaataagaaagatagctttaatttttttagtaaa
 tttgttattaaatgggtttcaattttaatccttcttgctgctaattaata
 tgttttcttggttgtaatgtaggcttgccaaaactcaatcttctaca
 gttcaatgcctcctccaaaagcttcgacatctcaacttgaataattgat
 caatgaattctccattttgaagtttacttccaattttcatctggtttgaa
 actttaatttctatttggattgtaacttctatttggattgtaagcttaa
 ttttcatgatgaatcttgatgcacatttgaattattatagtggtgaatt
 gtaaatgatgaataatagatgaatttaactataatgtatgagataagg
 tcaaaaaaaaaaagtttgccttgaattggattaaaaaaaaacacaacaga
 78001 ttttgcctccaaaacaaaatggcaattaccattgtcataccatgccaacg
 aacttttggcaataccgaacttcggtataccgaaagttagtacggta
 ggtaatagattttaccaaaaccgaaagttaggtacggtaattggtaacagg
 gttttggtagcgtaacgta[ccaaatccaccctagttt]tggtaatgaa MC-12SSR744bFor
 attttctcgttctgtccaaaaaaaaatataataataataataataat
 aataaaatataatataatataatataatataatataatataatataat
 atataatgaattgtagagagagaaaaaggatgggagcagttgacaa[ag] MC-12SSR744bRev
 [tcgggtgggagaacgtat]gtcgtccaagttagagctagggtcggctcctga
 ggcttcttcccttggttgataacttggcatcaatttttcttctgtttt
 tgtccaactgaataactcagcatttggacctcagaattttttctt
 78501 gcacatgtattgggaacttccatcagattttccgacataattgggtgtgc
 tattcaatattcatttcaaaaatgacatgtgcattccacataattacttg
 gcattcacaataattagttgaatttttctaacgaattatattctaccgg
 tggcctaatttcttggacaaatatacattgttgcaggaatttgaatatt
 atatgtttccctttcacatggacaacaaatattgaaaatttgaatatt
 gtctgaactccttttggttatagggcccactttccacacttacccttggc
 cctaactcagctctcctaatgatcacatttttccacaacataataatctt
 cctcttggcttccctcatttcttcttaatttttggatcatttttcta
 agctcttagaaatcatcacc[atgtgtgtcttcacatacgaacccaatac
 gtcttctcatccccctgctaggttgcacaatgcccttgttcttgatgc
 79001 tgacaactctcattccaaaatgtctccacaagcaatcaaaactactgaa
 ttcttgaggagatggaggtgttggaccattaaagaagattggcttgg
 gaaggaagtgaatacagctatgtgaagcacaaggtggatgggattgacaa
 agataacttctcacaattacagtttaaatcgaaggagatgctatttcta
 agataaattgaaatagaatttgaagaataatccctctattacttcttgg
 aaatataacatatacaatcctatacaatccttattgaaataaggaaacaa
 ataattaactttataaacaacccctcctaataaaggactcagcaccac
 actcccccttaagttgggtgcatagatgtcacacatgcccaacttgacaag
 tgagtcttcaaaacttgcactacacacagcatgcgtgagaacgtctgcaa
 gttgctcctccgatttcacaaaaggattgacacaacttctctcaagt
 79501 ttctcttaataaaaatgccgatcaacctccacatgttttgttctatcatg
 ctgtaccggattatcggctatatctctcgtgattgattatcacaatata
 atctcatgatctccttggcttgaaccacaagccctcaagaagttccga
 agccaaagaatttcacaaatgccgtggccatcctctataactcggcctc
 ggctgacgatctcgataccaccttctgttcttacttcgccatgtaacca
 aattctcgcctacaaaagtaaaatcccgaagttagaccgtcagatcactc
 acatttctgccaatcagcatctatgaaccctcaacattcaaatgtcc
 atgcttctgatataaaactccttttccaggtgcagacttcaaatatgcta
 agatacgcataacagcagccatgatctacacttggatgaatgcataaat
 tgacttactacactcactgcatatgcaatctggctcgtgtgtgagccaa
 80001 gtaaatagctcctacaagtctctgatctcctttatcaattgatt
 cctgattcggatccaaacatagatgatgttttcaacaataggagatct
 actggcttcatcccagcataaccggttcttttaacaaatccaaaacata
 cttacgttgtgacaagaaaatcctttcaagatcgagcaacttcaactc
 caagaaaacttcaaatctccagactcttatttcaactcagccgca
 aggttactctgcaacttgaatctcatcaaaatcatctccagtaataat
 catgtcatctacatagatgattaaagctgtcacttttctcgtcttctgt
 taacaacaatgtatgatcagaatgactctgataatccaactcctttc
 ataacttgagtaaacctatcaaacatgcaacgaggtgattgcttaagtcc
 ataaagagacttacgcaaccgacataactccggtatttctaccaactgt
 80501 acccaggaggaaaatccatataacttcttcccaagtcccattggaga
 aaagcattttcacatcaactgttttaatggccagtccatattagcagc
 taagatatacagaacacgtaccgtattcatcttagcaaccgggtgagaa
 gtttctgataatctacaccataagtttgggtgtatcctttagctaccaac
 ctctgttctgactctattgatccatcagcttggatattgacagtaaa
 caccatcgacatccaactgttttctccttctggtaatgacgtcacct

MC-12SSR744bFor

MC-12SSR744bRev

Mal d lps2
pseudogeneHypothetical protein
(ORF24) ↑

cccaggtattattcttctgtaatgccaacatttcttccattgcattt
gcccactttggatcctttagagcctcctccacacgggttggtagctgaat
81001 agcctccatattgtccacaaaatcttgcgctctgatgcaagaccattgc
atgagacatagttggctattggatacttcactttaccttccggagaaaac
ctgtcaggtggtacaccccggattgtgtctcgggtggcaatataatgtatt
tacatcattacttgcacagttacataattatccacaatacttacctcag
gtatatccagagaagattcattgggcagcacagtatagctaaagagagag
gggggtacggtagcagtagcagaatcgataaggaacgactgtgggtgttc
tgatggcacggcaatctgttgccgtggggttgcgaaggtactgcagcag
tttctgcaatagtagtatgttccctgagtatcatgaagctcgacagcaggt
actacgggaagaagtgccgagtagtaccatagtagtatttctctctaggtccaa
ccatccaagatcaccacaaaatgttctcccctggtgatcggaagatgaag
81501 atacgggagcataaaaacaattctgactcagaaaacgtcacatccatggta
acatacatatggcgagtagtgcgggatgataaacactgataccctttttgggtg
agaagaaaaaccaacaaagacacaccggagagcacagggatccaacttag
tacgatgaaacttgtgaatatgcacataaccacacacccaaaataccggg
ggggtgagagtaaaagttagacaccatcggtacatgtaagtggagacttg
aagaggagttagaaaatcaatcaccggagagggcatacagatttattacat
aaacagcataagtaacagcagctggccaaaaaaccttaggaacagaggcg
cccagaaggagagcagcagctgttctaggatattggcgattttccggtc
agccaccatatttctgctgggagtagtgggaacaggtgttccatggagaa
tgccttgatcacaaagaaactttgacaactcacgattaatatactacca
82001 ccattatcagaacgggaaactttaatcacagatgaatattgagtagtgaat
catctgagaaaacgtagcggaaaaccatactcacatcactttttttttta
agacataaaccaagtcatacgggtgcaatcatcaacaaacgtaacaaac
caccgaattccagaagaaggaacaataggagagggggcccaaacatcaga
gtgcacaagttcaaacggtaaagctcttttattcactactaggaggaacg
aagcagtaggcttttagcaagaatacatacttcacaatgtaagtctgat
tcatatttccactaaatagactaggtacaaaatgccgtaaatatccaaa
agatgcatgtcctaaccgacgtagcaacaaccacacttcttgtaaatatc
tagattgtgacgcgcaaacagcagtagctctgccagtagcagcgtcgtcc
acatagtagcaacccctctctcttagtgccacgccaattatctcctgggt
82501 ctgaatatcctgaagtagacaaaaggacggatcattagaacaacacaat
ccaattgttcagtaacctgtggtatggataacagatgatgggataaagaa
ggaaacaagcaagcaatgatgtaagggaaagtgcaggagttagatgcacagt
gccggcaccacacacaggtgccagagtagcattggtagttgatataat
tccggtagatgtagtcatatattggaatataatttttatcatatgtcatg
tgatccttgccccagagtagcagaatccaacctgtatggtgctgagttc
ggaggctaaagagaacacgtcccacggtagcctgtgtaggaagcagtagc
ccctagtagcagtagcaataagttgtgacgaggtagcactagaacaaggcacg
acctcctgggctgttgggtgtgcccagcaagggaggcacagcctccact
ggaaaccactgctccacgctctttggcagcgaattttttctttaattccg
83001 gaaaccaatcaggcactccatgcttagtaaaagcagtagcctcagtagta
cgagttgtccacagaaagtacacttcaaatcatctttattttctcgggt
ggaagcagtagaatttaaagattgattagaagaattatagagcagagaag
caccagttggccgagaaatcatagccagggagggcaccctccttgattg
ttactaccacccatcatagtaggcatgtcgttagcttccagccgaacaac
agaaaacacctcctcaacagatggtagtaggctgagtagcgtaaaatcac
tagcactttatcaaagacatcatccaaacccacaaaaagcatacaca
cggtagaattgaatttcttcccgaagggtttgaaatccacagcattc
83501 catcctgatgggtctccttgatctaactcctgcaaacggacttgagat
ccgcataatacacaccaaacttgccgaccttcttgacgaagtcggaagat
ttaacctcaattcataaatttgggaaatccgaacctcataataggt
ccgagccacctcttccaaacttcttagcagtagcaggtgaataaaaa
atcccatgatagcaggtccatggaattgatcaaccaccttttacgatt
gcattcctgtctccatgtttcatagtcagcaatctcctcagcaggtc
cttgatactccagtcacaaaaccttctacctcttccagcagtagtgc
tctctaaaaccttgaccaaagcggataatttgatccattcaatttcag
gtgttgggtacagcagatacatcatgctgaagaacaacaggattcaggt
acgagcactgcttccaatttcaaaacaccactaccttccgagatgcca
tattacctcgggaacaaaatcagagaaaaatgcacgtaggaacagtagt
cccaccacactgcaatacacacggcactactacccccactcacgcgaact
gaggtctgcgggaagaaggcagaaaccaatccggcgggaggaaaggaaaa
84001 acaggggtacaggaaaaaaaagggttacaggaggaactagggttacggc
taggctcgtgataccaaatagaatttgacaagaataatcgtctattact
ttctttgaaatatacatatacaatcctatacaatccttattgaataag
gaaacaaataaattaactttataaacaacacctcctaataaaggactcac
gaccaacaaattgtaattattttttaattcttttttgattaaaaataatt

atatacagaatatgacaacttatactattagcatagaattgaaaaagtaa
agaattattaaataatatttactgtacatagttaatcgcggttttgattgc
gaacacattgacacttattgttcatgagtttcaaattgttttcgaaattgc
caccattggaagaacaaaataactaaccaattacaaatttcatcaaa
84501 tatcaatcgtatcgtagaactcacaacttcatcaaacataagcacatgac
ttacacgagcctatagatcattttcctctaattcatataaaaatcatttga
agtacaccaaaccttacctgtggaagttttccagacaacaaagaaaaat
aaatgatgtaaatcctgaaaagactcgcgcaatcgcgaattaacagtgagt
actactgtaaaaaggtcatattgtaaggacatagtgggcagaagggaaagc
aaggattttttttttttttgttgttgaacaaagcaaggatgttttga
ctgaataattagttttacaattcagtaagtgagaatcgaagtaaaaaagt
ctcatatcgtgatgatcttctactcatcccacctcattcggtttcttag
gtttgggaagctcgatcaacccttgctctaataaagatgtgaaaaatct
tcaacatcctcattttctaaaggaatattctttgatttgtgttctttata
85001 aaaccttgcaagggttacaaaaagttgagttactttttatattattaatt
tttgtgatgtttactatatttatgtgatgagtagtacaattggtttgagt
gtttctatttaaaacacgactatttttaagggatgtgctatccacacacc
ccattttacttcccacacacctcattttacttctcacacacctttgata
atttctgtcggttgatcttctcaattcattccgatctgacggccgaaaa
ttaaaaaggtgtgtgagaaataaaatggggtgtgtggatcacaccct
atttttaatagagttagaaatgaccaaatagcacaaatttaagatggcg
gagagataggtgggcccatttttttttttttttttttttttttttttt
85501 atgataagggtaacgaagggagagaaaaatccaaacttctgtatatta
atattttagtttatgtgcaatggaatttcacatgtgaaagggatgagtg
gaaagaaaacgatccttgccggatctatttttttttttttttttttttt
tctacaattatgttcgtttatcgtacattatacggtttagaaatgatttaa
aatttaaaattcaaatataaatcgcagagaggctactacaacatatttat
cacatctacttgtttgatatttccattagctagaaatgtgtacttctaac
tctatcacttccaaaagaaattttcaatgtaattgaaaccaaggttctaa
aagacaataggcactaaacgagcggcgggctgacacctattgcctaggcg
aactaggcggatttaagtaaatctatttatatttcatataaaataagtaatt
86001 gtttatacttaaaatataataatttcatcataaactacaaaatagaatg
acatatataataaagatattggacataatgaaaataggggaacaagca
tataatgtgtgttcatttaagtatccaacaagtttttacaatttatgaa
aaataaataaaatgcaaaaagaaagttatctattttctgtctaagtaagt
cgcaacctagtcggtgcctaagcgggctctgggcaggtaggcggatgtctc
agcaggtctaagcggcctttcttaattttcaaacgcctagtcattataaa
agcgggtgcccagcggccaagcgtccgcctagcgcctaggccaagcctag
gcccactaggcagggatttttagaatagtgcaatgaagacgagtggttg
aaggaagttaatgttccacctcaaatgcggaataaggttgggatatgcaa
aagtccctgagttttcaatgtggggcaaaattctgcatctttacatgca
ccctcacatgtgaaaatactcaaataaaacacgtagacaacacatatcaa
ataaacatgtgtggtcattatacttttacacataagccaaatagtttga
86501 atttataaatgataaatacaattcacacatgaatctagcatcaaacatac
catattcagatcaatgaggattaagaaatgaagaaacaaataaagtaatt
aggttttaaaagaaaacatgcaaacgcaagacagcaaacacaaaactg
aaaagcaatgcgtagaattaatgtgagaattgacctactcgacctgctaa
tcgaaattttgtcatgaattaaaattatttttttgaaatgggtattataaa
atttgcaagttgattgtatgatcactattggatgaattgaaggtggaaaac
ctaaaccacctcacaatgtatggaagtggttgttactttggcaacatg
atttccaagttttgaacaaacgaatattatgttgggtgtaaacattaatc
gatcaaaaatagcaaacattatcctctaccatcgtcatgaaaacggttttt
tgtttttgtacttatatagtacttaaggttgtcttatgtgttccgatgc
ttattcatgatgtactttcaagaatttatctagtcgaacttgtgtttta
87001 cattcttaacggttgtagttattttttcataatcgtaaaactttcctgtat
aatcaacattttcaaaagaaaaagaagtttgtatattaacttttgtttcc
ttttctaataatgatgtgctagtgacccaatttgaatctcctacaagtaa
atgaaacacaaaacaaaatacaaggttaaatcaaaatttggacgacttaaa
tatttttcttaactaggtatcattggctgcatgattctctctcctttaa
atttcttctccttctagtttgttcttatttgaacggttacaattaaatca
cgtcaacatcaacatttatgttaacttctttataaaaaataaaaaagatg
aagataaaatgttgatttctttataaaaaatacaaaataaaagatgaagata
aaagtaagaaagcatggcagaaaaggggagatgatttggaaaggaagagta
tcaaactctgttagattcagacacaatttattttctgcatttttgggtct
87501 tcatgttcgaaatttaataaaaaatttgaacaaaaataaataaaatttag
atcaaatgatttgactttgattcaaaattagataatttgaaggtcaaac
cacaggaaacagataaaacgggaaggtaaatagcagagttaaaaaggaaa
agattatacaaaaattacaatggtaagcagtaaatgaaggaaggaactg

ttagttaatcaagttaataacctcgccaaatgatagacaagaggatttatt
 gcaggagtgtgtctttgtgctctggagaatctggaagtctcggaatgaaa
 tggatattaatggagtccttgtaaccctatagagatggttcatgtggtt
 aggaacaaaatttggagtttogaagtgtcaaggtgctaaaggtgctaa
 aggtgctaaaggtgctagagtggatgcagacatcaaatttggttgtgpc
 atcgatgattgggacatctatcattcagttaaatgaagcatttattttat
 ttttcgctacgcaattgtagtgttcagagttcaaagtgaacatgtgt
 88001 catggctaaagagtcatagtgccctcctttcccgttaagtgattttaagcta
 ctttaccatttaatttgatacattctgatgtgtggggctcggcaaaagtt
 acttctaattggattccggttggtttgttacttttattgatgattgtactcg
 agttcttgggtgttcttaataagaataaagagtgatgttccggtgcttc
 ttcaagaattttgtacaatgggtgtctactcagttttagaccaaaagttat
 ggtcttcgggtctgataatggaggagaatagtgaatcacactttggcat
 gttttttcgtgattatagttatttaccggagcactactctgtttaca
 ctccaaaaaatgggtgttccgaacggaagaatcgtcaaataatggaggt
 tgctcgctccttgatgttggataagtgttcttaatacatttggggatc
 atattgttttggctgctgtgatttgatcaatggagttccaagtagggtt
 88501 cttgatttttagacactgcttgatgtgctacaaaaacatgttctccttgt
 ttctatatcaaagcttcttatcatcagaggagtaacttgatgcataatgt
 tctccgatgtgtctttattgggtatgctaacaaataaaaaatgctataagt
 gttatcatcctccgactcaaaaaacttatattaccatggatgtcaccttc
 cacgggaaatctcatatttgtgaaaccctcttccgactctccatttca
 gggggagaggatgagtgagtgcatattcgaatagatggatagatgatg
 tgttacagacagagttgggaacataacctattatgttgcgtgatactgat
 cagtcggctacagatagtgatcggctcacctgtcatccccgaactttct
 acttacgacagatcagatagccagtgagttgtctgttagtatgtctga
 cgaattaccttctgatgatcggttgcctgctgctggatgtttaacaagt
 89001 tgcttgatgatggtttgcagtgatgatcttctaacagtttgatacaa
 gatgatggcatacatgaggtaaattctgacgatttttctacttatcagtt
 gcctccacgagctaatcgtggaaaacctaaagtacagttatgagcctgata
 tggctccaaagcctaaatcctatcaacaattatgtgttactcatcgt
 ttgtccaaacctatgcatctacatagtcagctatctagtgtatcaat
 tccaacaaaattgcaagatactttgtctgacccttagtgggttaatgcca
 tgaaagttgagatggaagcttagaaaaagaattctacttgggatttgggt
 cctttaccaaatgggaagaatattggtgagtagatgggtattcgctat
 taagcagaaagcagatggttctattgaccgggtataaagccagattagtgt
 ctaaaagttataactcaaacctatgggggtggactatcaggaaacttttgc
 89501 catgttgcacaaacttaataactgtgctgttcttctgtccttaacagcaa
 ccaggattgtcctttgttgaatttgatgttaagaatgctttccttcatg
 gtgatcttaaggaagaagtttatatggatcttccacatggatcagaaca
 tctcttggaaaatgtgttgcaggttacgaaaggctttgaatggttt
 gaaacaatctcctagatagtggttttggagggttacaatttcaatgaaga
 agtttgggtatataccagagtcatttagatcatactttggttctaagcga
 caaaatggtaagctaacctgcatgtattttagatgatgacatgatag
 gactggtgatgactcagaaggagatacaaacacctcaaaagtaacctagcta
 ctaaaattgagataaaaagaattgggtgaattgaagtattttcttggaaatc
 aaggttgacagatccaagcatggatatttttttctcaacgaaagatg
 90001 ttcttgatttggtagctaaaacatgtatgttagattgcaaacctgttgac
 actctgatttagcagaatcatcggctgggcttatttctagatcaagttcc
 tactcataaggaacgatgtcaaaggcttgtgaggagattaatttatttgt
 ctcacactcgcctgacattgcttatgtagttagtggttaagttagttt
 atgcactcaatagtgaaagctcatatggatgccgtaaccctgattttgag
 gtacttgaatgtcgaagggtatcacagatgcagattgggcaggttctact
 gatcgggtgatctacatctggatactttacgtttgtgggtggtaatttagt
 tacttggagaagtaagaaataaaaagtgggtggctaggtcaagtgacagaag
 ctaagtttctgtagtatgtctcatgggtgatgtgagttgttgggtgaaa
 90501 aaattgttgaagatccttaggtttaaaccogagtggtgctatgaaacttca
 ttgtgataacaaggctactattgagattgctcataatccagtgtaacatg
 atcgaacaaaatagtggagattgggtcaacatttcatcaaggaaaagttg
 gatgctggaattattatgtttccatttgtgtgatctgaagatcaacttgc
 tgatgttcttaactaagttgtgtctagtgtgttttccaactcgcttga
 caagttgggcatgctgacatctttgcaccaacttgagagtgagtggtgc
 gagtataatattagttaaagtgtaaatagtagtttattgtctcatattgg
 tgagtacatagttaataccaattataactcctatatatactccacttga
 agattaatgaatataataaatttctaataattgtgatataatattttg
 gtatttaacatgttttagtttgataggaccttcttccaccatcaccacg
 91001 acctgtccccgccccgaccctaatccgctctgagcggctgtaggaggaaa

Retrotransposon protein,
putative (ORF25) ↓

accaccceaagcgaccgaattggcctggtacaccagcaaacacccgc
 cttagagaaagttaagggcatctgtgagggccacaaagaaccgccacgatc
 tctcgtgcagtattgggctacgaaaaacctagcagacatcacctccttct
 cggctcgtctctttcatcgaagttttagttcatctcctgcaacctttt
 tggctcctcgtttgcatcccatcatcatctccaatctctctcatagata
 gctattatccccaaatcaattgatagttgaaaccctaaaatcccaattg
 gcccaaatctcgttgggattctcagacgagttctatttgaggaatca
 tccagagttcaaacatggtacgatccattattttggttactggaatttg
 91501 ggattggttgtgcttcaactccatggcgaagatgtagagagtttgtaggg
 acgagaaaagctacgggagatttcttagatgaagtgtgattgaaggggagg
 gtgggtgatggcgggtgggattctcaggaagaccaagggatctgagga
 tttttggttatttttttttaacttttctttaataatatttttaatat
 tttaaattttttattaagtgttagctacgtgacataaatgtggatgaaa
 aatataacgaatatattatattgaaataaaaataatactttatgtaagtga
 aatgttctaagatattgtgaaggaattgaaatcgacctcaaaccttaaggg
 ggtgtattcaattgggattttaaggattttaattcttttaataatgaaatca
 tggggattcaatcaggattttaagtgttctctgaaattctaggtgtat
 tcaattagaatttttaaaatagtttattaaatccttagaaatccaggtgt
 92001 attcaattagattttaaagaagttataacattccaggtgtattcaatt
 agaaattgattttaagaatttgataaagtttaaggaattagaggggaattg
 gagagatttcgtagtgtattttaagatccacaaatctcacctcttcca
 tgagatttcgaggggaattgaaatcaaaattttatatggaatctctacaaat
 caattaaactccataaaaaatccatggatttataaatccattaaaatctct
 cacattctcaattgaatacaccccccttaagatggtaaggtgaaatttacc
 cataaaatattaatagatgcagtagaaaatagttgattttttttttctta
 atcaagttctagaattttattgggtggcttctttgactaatatattttatg
 cattgattttttttattttatgccatgcaattgaaaattagttttcagt
 ttctttcacatggacagcatccattgaaaacttcaaattttgtcccaac
 tcttttagatatagggggttaattttacacacgtaccctccctcccaaat
 92501 ctataaataaccaccctctggagagaaacattttgtcacatcaacgttgatc
 atcctcatttgcttctcctcatttcttcaatttcttctgattctttctaa
 ctcttaaaaatcatcatcatggtgtcctcacatagaaactgaatacgc
 ctccgtcatccccagctaggtgtacaatgctctgttcttgatgctg
 acaacctcatccccagattgctccacaagcagtcaaaactggtgaaatt
 ctcgagggagatggcgggtgttggaaacctcaagaaagttagctttggcga
 aggtttgtttagtttcaactattttctccattaaactgctaattttacct
 atttaatatatacaccccacatatatctcgacaataactagctagtatct
 taatttctatttctttttctattttgacagggagtgaaatagttatgtg
 aagcacaaaggttgagggaattgataaagataaactttgtgacagctacag
 93001 tttgattgaaggagatgccatttctgacaaaattcagaagatctcttatg
 agattaagtggtagcatctggcagtggtccatcattaaaaacatcag
 cattaccacaccaagggagatgttgagatcaaggaagaaaatgtcaaggc
 tggcaaaagaaagggccatggtttgtcaagcttattgagaaccaccttg
 tggccaatcctgatgctacaactaaaccatcgctctggcaacaatctt
 gtgccttgcttctctatttattcagctgtgtggttcaatttactatcat
 accaaaacaagctctttgacttgcaacttggatagattaaagaataatgt
 tgtatttgatttttaactgaataataaaacatgtactagttatatgtaa
 attatttgatgacacatttaacttggagtaaatatttttcccttgcaatat
 catgaatttcacattcaatctgctatgggtaaaactcaattacacataga
 93501 aaggaaaaagaaagttatacgactaaaattagaccaattaaactcaaaagt
 caactgtttgggtgttgggatgcaaggttcttatgacaaaagaataaaa
 gaccaaatccttgggaaaaaataaaataaaacccctcaaattaacaa
 atgagcaaatgtggtctcttatatgtatgagtggtgaaacttgacttcaa
 ctgttccagaacaagaagtttgggtcaaatgactaagaatagaaaagataa
 aaagtctaagtttgataaacatgactaaacttaaatatattatgtcgagaa
 attttgaatttttctcttcccttctttaccattattataagttcaacaaa
 ttaaaaaataagcaagtggttctttgacacagtggtgaaagaagttgaa
 ccttgcaaaacgacgtgggctcaaacattttttatgactaatctaacaaa
 ctctatcatttgataaaaaaaacaaatataataaaaaacctaatttaa
 94001 aatggcccgccacctatctctctcgcaccttttaatttgtgctatt
 gtcattttacactaataaaaaataaattaatgggtcacacggctttaccaa
 aataataggacaaaaaatctataactgctcgaaaattgatcattaataga
 tcaagggatattttgtgataaaacatgtaataaaaatcaataaaaattaa
 aaaaagaaatcaacttccaaagaaaagaaaacgtgcaactagagtcacaaa
 ctgaagaagttttgtgaaatgagagataaatattcaaacctatgatataaa
 aaatccacttctataaacgtggattcaaatcccttcacatgtttcgttc
 ctgaaataactagttatggaagctcatagcacgatcgattgtgacgagaaa

Mal d 1.06C ↓

MGLTYETEYASVIPPARLYNALVLDA
 DNLIPKIAPQAVKTVILEGDDGVGTI
 KKVSYFGESEYSYVHKHKEGIDKDNF
 VYSYSLIEGDAISDKIQKISYEIKLVA
 SSGSISIKNISHYHTKGDVEIKEENVK
 AGKERAHGLFKLIENHLVANPDAYN

ctgatggggtgtttgtgaatgtaggggaagagaattctcagaatttggtt
 94501 ttgtgtggcggaagtgtggcctttgagagcaagccaaaactagcaggtga
 tttgatggaatcgggagaaggaacttcccaagttgaagttatTTAATAAAT
 ctaaacagtataagttgtaattatataatgtataattatttgaattcagt
 ttgtattgtgatcactccaaatTTTTTTTTTTTTTgaaaatagagtacga
 tattcattaataatagaggaatcgtacaaacaaagaaaatagtgcatat
 aaagggctagataaaaagtctgtcgaagatttcaaccagctcaagaaa
 aaaaaatctcaccatttcatgggtgtctaatgtatcgaaaatattagcat
 ttttctctctTTTTTggcogagactgaataactgagcatttggacct
 cacattTTTTTcattatcagacttccggacacaattggtgtgctactca
 atattcatttgaaaatgacatatgcaatccacatattacttggcattcag
 95001 aataactgttgaattttataaataaattatattctactgggtggtgaa
 tttctttgactaataaattggtgtacggaatttgaagattatagtttcc
 cttttcacatggacaacaaatattgaaaaattggaaatttgtctgaactc
 cttttgttagaggggccactttccacacataacccttagctctaactctgt
 aaataccgtctctgggtgctcacatttgtcacaccaatgtcaatcttccct
 ctttgtaattgtctccctcatttctgttcttaattTTTTTggatcatttt
 tctaagctcttagacatcatcacc**atgggtgtcttcacatacgaaccga**
atacgcctctgtcatccccctgctaggtgtgacaatgcccttgttcttga
tgctgacaatctcattccaaagattgctccacaagcaatcaaaactgccg
 aaattcttgaggagatggaggtgttggaaacctcaagaagattagcttt
 95501 ggtgaaggaggtgaatacagctttgtgaaagcacaaggtggatgggataga
 caaggataactttgtgtataactacattttaaattgaaggagatgccattt
 ctgagacgattaagaagatctcttagagattaagttggtggcatctggc
 aatggttccatcataaagaacatcagccactaccacaccaaattgagatgt
 tgagattaaggaagagcatgttaaagctggcaaagaagtttgttcaaac
ttattgagaactcccttgggtgccaatcttgggtgcatacaactaaaatcgt
 ctatgatattgcatgatgtttatcgtcgggtgtggatttcaatttttgctt
 tttattatcatataccacttccactactggctttgacataactttgga
 gttgtgatcaataacctatgaatggttatttgaatttcaatattcaatgc
 atttagtaaaaaacagcttgcgcgacgaacaatattcgtcgcgcgaca
 96001 cgtacttgagcgcgaaaaataaacttcatcgcgcaaaaatagagcgcgca
 ctagtttctgcgcgcaagaacacttgcgcgcatgaataatctcaaaaaat
 tgcacgacgaataatctctggcatcacataaagtgggcaagaaaatgtag
 ggtggcgcgaaaaacttgcgcgacgaaatggagaatttgcacgacctata
 attcgtcgcacaagttcctgttagttttgaccatttactgcaggtgaatc
 cgaggaattaaactgcatttaatacaggtgttgcgcgacaaaaggattc
 atcgcgcaaggtcacacccacctataaataattcgttctcgttctccta
 attcttcaatttctattatccttattcttcaaaattctgttctgtgatg
 gatgacggataagaacaaggtgagactgtacatgatgcaaacccgcatt
 atttaaggatcattttgagtttgcgcgcatgogatcgtttagccatgggg
 96501 aataattgtcccactgttgagtggtgttcttgggatgatgacttgggaa
 tgtgaagaaggatgtgatggatgaagtaatagtgagatgtaattcttga
 tggattaaataaagcttgtgaaattttagttatagtgtgaaattaat
 tatttgcatttacgtgtaacagtgtgaatataactcttgatgatgtagga
 acgaacagttgatgaagttgatggataatgcgttggatggagggttacaac
 cgatggcgttatgaagcttgcggaatggaccaggatcatccaaatagta
 gttttaaatttatgtttgtatgacaatatttaaatttaagggtgttttt
 tatgttacgtattcgaacttaattatgtttcaattatgtatttggactta
 attaggtttcaattcatgttggttttttattgaaacctaatgtaagcac
 cttatcctcggtttatatgtgtttacaatcttcaatgaatcgtactta
 97001 tgctgaaaaataattgtatggattaagtaattaattatataagactaaat
 attaaggcattaattatttatagaataatatttcaggtattaattattt
 atagaataaataaataattattttattcgtcacgcaatattttgagcgaca
 atagcttcatcgcgtaaggaatctttgtgcaacgaacacatgtatttgc
 gcgcaagcaggtcagcatgcgttcataaccttccaattaaatgcacatt
 gatgatgcattttgcgcgcatgaatattttgtcacgcaaaaggtgtcctag
 tgcataataaacacaatttcagagccctagttttaaataaaatggttt
 aaaaaatgatggccgattctagagaaggtcccggcaaaaaaaactt**gta**
gtgcaaatagagaggtatcgacgaaagcaagtggtgtacttgaaggtaa
aaatttggattaggcgtactccgaatttaagcatgacattgggaatttgg
 97501 **tgccggagggtattgttctctcgagttttagtcttggaaaagggttctgag**
gagctaaagaagtcctgctgggagagttatcggtaggtttggtaataa
tgaataattattttgttaaaccgtaataatttttaattactaatttgt
tgttacttgcatttaattgagggtcattgggatgttgatgaaaccgatgag
aagcaacggatttatgtggatggacttttcaagcagagtcttctgcagtg
gaagttttagtgttgcgaagtggaggattaaagttgattaggggtgtttt
tttttttttaaggattttgtggactttataatttctatttgggatagt

Mal d lps
pseudogene

Transposon protein,
putative (ORF29) ↓

aaattagttttgttaattaatttaatgtttaattagtttgagtttctga
tttaaaataaaaataaaattttacaatatatacaaaataaaaaataaaataga
98001 aaaaaaattaaaaaaaatgcatgtgacgacgatacattcattcgtcgt
gcaatgtgttgaaaagaataaattgtaaaagtaaggacataaaaataaata
aatggtaaaccttgcgacgacgaataacttatttttccattgacgcaaatgacg
tttgcgacccatcttgcgacgcaatggtgttgcgacgacgaatacctat
tttgcgacgcaatcgttgcgacgcaaatggtgttgcgacgacgaatacctat
aaccatgtgcagattgtgcctctgcacgcatctcttcaattttgctcaat
ccgtcatccgtcacctgtcattttcattttctctcaatttcatcatctaa
tactccctccattttcttcttttggatcgagctatctcgggtgatct
ggaagcgtttttgtcattacggtaaaagtattaatgtaaaattcatact
aacgccatttaattctgttctctataggatattgtgtgtgattagcatt
98501 ggtggagaacgattgagaaggtattttcttgcgttttatttttaacata
taaaactaattaaattatggtgaacttagtttgtattgtttattttgtgt
tgtgaaattatatttatatggtttaaatttgcagtgacgtacaaa
tatgtgtgtgatgtaggagtttaattgaaattaactcgtacttgttt
ttatttttagtaaaacttagtttctggtcgaggattagttggatttcgt
gcatggatgcaaaagcatgactgcagccaattaattctgaattgtccc
attatttgaacagtttgaatgcacagttaggatagctagtattggtt
gaaggattaagtctcattgtggagatttcggtatcatctctgtacgttc
ttcgggtgtacacctgaagaattgtattgagatgctacaaaattcatc
cacgtcgaatctaattctgatgtaaccgtaaaattacgtgacatggaaaa
99001 aatgccaaccaagagttattcccgggtgacgagactttccgttctcat
agccatttgaactaatgcatggaaaaataaagtatcgtatgtcgaacca
gtgttctgtttacttttgggggtttcaagagaatgcttccgaaggaca
attgtttgcaaaagaccatagacacacacaaaagggtttaaattggtctt
ggattgggttatgaaaaaattcatacttgcaaaaacaattgtattttatt
ctacaaagagaataaagcattggataaatgccctgtatgcaatgagtcga
ggttcaaaatgatctcagaatagaatgactaagatcccacaaaaagtt
atgcgttatctgcccctgaaacctaggttgcagcattgtatattgtgac
gcatactgccacagacatgagatggcataaggaaaaacgggtaacaacg
atgtgatcggcatcctgcagatgggagggatggaaagagttcgatcga
atgttccccgagtttgcgtgctgattctcgtaacgttcgattgggactgc
99501 cactgaaagatttaattccgtttgggggttttaaccaacaccacagtt
ggccgattttcgtatttccgtataattgccgcatggaatgtatggaa
aaataacatgatgatgactctattgataactgaaaatcctggtaagt
aatcgataatacttacggccgttagtggatgagctaaaagatttatgga
cacacgtgtgtgcacatcagataaattgactagcatgatgttcaacttg
agggtcagtgatgtggactgtgaacaatttccccgcatatgcaatggt
ttccgagtgagcactatgggttatatggcatgccctgtatgcaagaaag
acataacatctagttggcagcgaacaaaagtttgttaccttagtcatcga
agatggttgccttgggaccacgagtggtgagagaatgataaagagtttga
100001 cggggagaaaagagcatcgtctcagacccagagaatggtccgggtgctcaga
ttttgaaacagcttaaccttttgattttgctcctttagggaacaatgtc
agtaggacaagaccttctacacatgaattggacgcataaagccatattt
ttgagttctcgtatttagtccaaactaaaattaagacacaacctcgtggt
atgcatgttgagaaaaatgtatttgacacatagggtcagcacaattctaga
tattaagggaagacaaaagacacaatacaagctcgtcttgatttggaa
gaaatgggaataaaaacgggggttatggatgaaatggaacaatgataaagcc
ataaggatcttgcattttttcaataaaaaccgaatgataagaaagaatt
tttaaagcttttctcgtctgtaaaagttcccgatgggtatgcttcaata
tcgcacgttgcgtgaatgttgaaggggttaattagctggcttaagaggt
catgactatcatgtggttctgcaacacttacttctgtgggtattcaaca
100501 tctcgccgatgtagtgaaccaatcatgtgtgttccagttttttttgc
aattgacgtcaagaacgttgcgaaaaaacggacgttaatacagttgcacca
tgacattgtgcaagttctatgcaagtttgaaaagatatttctccactct
tttcacaagtatgatacacgtgatggttcaacttaccaaatgaggcattgc
ttgttgaccagttcaacttcaatggatgtatccaataaaaaggatata
ttcctcgtcttttttattaataatataatataattacgattagtaata
tgtgtatcatatcgttttcttttggcaggcttcttgagacttgaagaaa
agtgtacgaaacaaagcaaagccgaatgatccattataaaagcttgagt
ggcatataacgcgcttacttttggtaattgatttaaggatggtgaga
caactttcaatcgtcctcctcgcaataatgacgaaggtgtgagaaaggag
101001 aaactttcagtttttgcacaaattgctcgaccattcggagatcctatttg
tggcagtcattttccaagaaggacatggaggtagcattgggtcgtac
tcaacaattgtgacgagacactgtcatacctagacgagcatgaaaatag
atgaagtaggcacatcctttatattgtatgccaagaaatattgtgaatt
gtttccgacgtggtttgtcgaatatgtaaaattgaaggttttattg

acatatatattgtaccaattaagttgctcgaactatcaaatTTAACGTT
 TGTCTAATATTAGGTGAATCAACTGAAAGCATCGAATCCCCCGCGTACA
 CTGAAGAGTTATATAAECTGGCGTCAAAACGAATTCGCGTTGAATTGTAC
 101501 CGGGTGGCATGTCAATGGCGTCAAGTCTTAGCAGGTGCACGAGATGA
 CAAGTTGTGACGCAAAACAACGATATTCATGTCCCAGGGGAGCGAAA
 GTACAGACATTGAATCTATGGTAACTAACAGTGTGCGTGAATGCTT
 TACCAAGACTGGTGCCAAGTGATCCTATTTAAGTGTGATGGTTAATAAC
 AAACCCAAATAGGAATGGAAGTGTGAAAAGAGATCATGGATTACTATTAG
 TGAACATTAGCAAACTGGTACGATGACGACCCTTACATCTTAGCAACC
 ATGGAAAAATAAATTATGTATCTAGATGACCTTAAAGCCGGAAGGGTGTG
 GAAAGTTGTCAGAAGATGGATCATAGGAACGTGTATGTATACCAGAAC
 AAAACCGTACTGACAACGACATCGACAATGTTGCTGACCAACGTGGAATC
 TTCGATGGAAAGTGGTGCAGAACACTCCGAGATACTAATCTTATCTAAG
 AACCATTCAGATAGAAGGAGTATCTTCAATTGAAATACCGATTCACTCA
 ATCATAATTGACCTCGTAATATTGCACGATACGATGGTACAGTGTCTTAC
 102001 AAACGACAATGGTGACGTACCTATCGATGATGAGGAATGAGACACTGAAA
 ATGATGATAGCAACGAAAGTGAAGTTATTATAGTTCGGATGAAGATTA
 GGAATATATTGTAACCTTTTTTTTGTAAAACACATGAAATGGTTAATG
 CATTATATAGAGTGTGAAAAAATGGTATAAATGATCCATTTTATAT
 GTGAATTGCTTGTAAATAAATTTGGTATGTTGGTATTTTATAATAAATT
 GTAACAAAAATAAAAAAATTA AAAACCATTTTAAAACATAATTGCTG
 ACGCATAACAATTTTTGGTGCAGCAAGCTAGGCTGGCAATTCCTGACAC
 AACC CGATAACCGACACGACAAGCAAAATATAATTTTATAATTTTATA
 ACCCTAAAAGAAATTACTATAATAATTATATATTAATTTAACAATTTT
 ATACCCTAAAACCTATAAAAATTATATATATATATATATATATATAA
 102501 TTTAATTTAAAAATTGGTGACCATTGGATTGACTTAAATACATACCA
 TTCAACTCTTAAGTGTATACGTACAATAAATAAATTTAAATCTACTA
 AACTAATGGATACCACAAAATCTTATGTTATACTTAATGAAAGTATAAAT
 AAATCTAAGTGTAGTGAATGTATCTTTTGTGGGATACGCATTGTAC
 GAAACTAATTTCAACGATTC AACCATCAAACCTGTTTGTATATGGTTCGA
 GATCGCATAAGCCGAAAATCACAGAAAACAACATTCAAAATCAAGTAA
 CAGGCAAAAAAATTCGACGGTTATAACA AAAAATCATGATTTAACAGT
 TATTTAAATTTACGCGGGAATATTGCGTGACGAATACATATATTTCACT
 GCACAAAGTATATTTGCGCGACGAAATATATTTATTCGTGCGACAAGTGT
 CTGAAATTTCAATTAAGTGTCTTTCTTTGGGAAATTTAAAAAATCTG
 103001 CGCGACGATGTAGCCTGCGAGACAAGACTCTTCGCGCAAGACTC
 TATATTTAACCCCCCTCTTCCCCTCACTTCCCTCTTCTTCTTTTCT
 CCTGTGTGATACCCTCTCTCTCATTCTGTATCTCCATCTTTCTCTCG
 GCGTGCAGCTCTCGTCTCCAAGAGTCACTTCTCTTAGTTTTTATTT
 TCAATTTCTTTTACAATTTGGGATAGAATTAGGATTTTTTTTTT
 GTTTTCTTTCTTTTCAAATTTGGGATAGAATTAGGCGAGTCAAGTTA
 GATTTAGGTTGCTAAATTTGGGGTTACATTTAGGTTCTTAAATTAGATT
 TTTTAGGTTCTTAATTTGGGTTAGCTTTTGTATCTTAAATTAGGGGAACG
 ATTTTGGGCAATGAGTTAGATTAGGTTTTCTTAAATTTGGGGTTAGATT
 TAGGTTCTTAAATTTAGGTTCTTAAATTTGGTGGCGATTGTCATACTGC
 103501 AATCAAGTTTATGTC AAGTGTCAAGAGGTTATCATAGTGAGGGATA
 ACAATAGTTTCTATAGTTTGAAGATGGAAGTGCACATGCGAAGATTAG
 TTCTGTAACCTCCACATGCTTCCCTTGTGTTTTGTTTTCTCTATGATTG
 CAGTTGCTCTCTTCAATATGCAGACATGAACACATCTAGAACATG
 AGTTCAAGCAGTTGGCATTTTCTATTTCAAATTTTGTGTTCTGGAGA
 AAAGGTCAGGAATTTGGATTTTTTTTTTCACTCAAATTTGTGTTCTCCT
 GGAGAATGATATTGTGTTGTAAGTTTTATATTCTAGATGGGATTTATT
 TGCTCTCTGAAGAAATTTGTTATCTTTTTTGGATACTGGGTGAGGTTT
 TTTGTCAACGAATACTTCTTCTGCTAATAATGTCATTCTTAAACAAAAT
 ACTTTCTATGATAAGATTGTCATTCTTGATTTTTTAAATTTTCAGGAGA
 104001 AAAATTCATAGCTTTTGGACGTACACTTGGTCACTGCACTTCTTAT
 GGTAATTTAATACTATTTCCATAATTTGTTACTTCCATTGCTTCT
 AGTATTGTTGTTGACTCAAGTCTAGTGGTCTGCTCGAAAGTGTCCC
 CATAGATTTGATATTGTTGTGGTTAACTGCTTTTTTAAAGGTTCACTTG
 ATGAGAACTTACATATCAGAATTATATGCTTTGGTACTCTCTTTCACT
 GGTTTTCTGATTGATTCACTGACCAATATTGTGCTTCAATTTTTTAAA
 TAGAAGGTTATCATCTGCTTCTAGGAGATGCAAAAAGGTTATCAGT
 TATCTTCTGCTTCCCTTGTTTAATTGTGTTATAGCACCTGTTATGTGT
 ATTTGGCGTATATATTGACTTTGTAATTTCTAAGATGCTAATGCTTCT
 GTCTTCTAGCTATGATGCTTGTGTTCCATGGTATTGTTGAATATGTGCA
 104501 TACTTTTTGTTTTCCATCCATTAACCTGGTGCATGTTCTTGGTTTTCTAC
 CTACCCTCAAGGACCAAAATGCTAGCTTGTGACTTATTTGTCAATTGAA

MC-12SSRC744CFOR

MC-12SSRC744CRev

105001 ttcctgacatacacatagcacgtacatactagatcttgcctccaactgc
 tcctttatgctagtttctcttttaagtataaattgtgggtctatatcc
 tatatgagttgattagtttgtgtatagtacatcagttttaaaatgttt
 ctgccaacttgaattgggtttagaattagggtttttaaaatcgttta
 aatthaagtgaattatagcatgtgacggtaagttaaatcttagtgt
 acgaaacaaattgattccaatgttagatgtcaaacctcattaaaacc
 cgaagggcgatgacttctacacctagtcgacgactacatctactactgc
 catcactgctccggcaagatggaccataggccaattaatctgggtgacc
 cagttggtcctccagtcacacaggcgacggcatcgtcgacttcttggtg
 gcgttgctgttagcactcgacgtattcaccggcgccccgcactacgaa
 ctgagatgttgccattgggatccacaattgatgcctcgggtacacggcca
 agtatacaatcctaattcactcccttattcccatgttaactgaatgtt
 ttttttaggtgtaattgttaacttattgtcatgtaaggattaaaaagttg
 tcatcgttccaaacctttattaaactattgtgatcattgtgttgatgc
 gtatattgtttatgttattttcaagatttgggaaatgttttaattatag
 gggaggttatgccaaaattagtaggattwgtkattagtttgggttaa
 tgattatcttcatctcttgcacgaagaaaaacaccagggaccttg
 105501 tgcgcaattgaagcggccaaggtcaccgggtgaccaacgggtcatatcc
 caatcgaatcagatgagtgacatcgggcagcatcaacggcggagtagcat
 agtgcattagccatgacattgggcatgtcgtgcagacctttgccttag
 cattgccttttaagtctttcatgtaaaatattatattatatttatt
 actgtcttttattactaatactttcaaaatattgttacattgcagacaa
 actacaatggaggacatggacgaggacatggtcacgtacctcaatcag
 ctcttccgaacactacaagcagtggaagagtgaccctgcaccaatgtt
 tccagcaatttgatgatccgcaggtcactctcgatgagggtgtcgaag
 gagttgaaggaccgacaagatagttgggttgacttgggtcattttca
 106001 ggagccaagcttatgctgcgtttttattacgacttctttgttttactt
 ttttttaatttttactaatgtttattatattttttattatttttct
 taaagtattacaactaatacgtttattttttttataacaaaagaaggcg
 aaggcgaaggcgaacaagatcaatcaggagaataagactcttctccacca
 ttcgggttcgagcccttctcttataggatggaggcgcgacggaaggtat
 atattacatctttcatttccaatttaaaatgccgctaataaaattttt
 tttcttactaacattttccttattttttcgatttcagaagggttcaaaa
 tttccagagattgacgtctttgctaacattatggtcggcctaggaatga
 gttgaccgagtcctttcatgtaagtaattcttatgcatcaacatttat
 actaattattcaaatgttttctattaataatccttttttttcacagg
 106501 cgactatggtggagaaaagtcagtcggtgcttcaagagtcgacctcctag
 ctctctcggacacgcctattgagtagtggtatccccctgagaatgcagg
 gttcacatcatgacagagacgttgaccagacttccggtcggagaccaa
 ggacttattgtcgggggatggaaaatgcccaagcggcgagagagtagagcc
 tcgcatcctagcagtcfaatggccaggttacggcttgacgcaagaagt
 tgtagcctaaggagtcagctagcatctacaagtcfaatgtcgatgc
 ttgtacaaacctcagttcctccggaatacatctcccatgttatctaca
 tcatcgccctcacgccttccacactgagcaagcacagcaatcaggctc
 gtgcactctgaccccgctcccaaccagcagcaagatcctttaa^{acgaca}
 ttocctatagatttttccgcttttttttcatagttttgttctgctttctt
 107001 ttaaaactttatatttgtacatacatttttttatattataaataaaaat
 tgttttttatatattaatttgcctttttttgtcattagtagaaatctt
 ttttttcttttttcttttttttgataaaaattacattgcgcaaca
 ctgaccagtcgtcgcgcattaccaagtgcctctaactttttcgtcatgc
 aaacttttcgagacaaactagtcacaaagtcacttgcaccttttatat
 cgtaaagcgaacaccttctgtogtataactttagcccgacgagtaataaa
 gcgtcgcacaattttttttgcacgaggttgaattgttcgtcgcgc
 aatgtctttttcacgacaaaataaccagctttttgggtcaaatgcaatcc
 taagaaaactagggctttttccattaatctaattggcaatcataagaa
 aactagggctctctcttatttacaatttaaaactgggtttttgggcacgc
 107501 aacgatttgggtgccaaggaagaattcgggttggtctgtgattcaattaaa
 tgggtcctcacaagcgtggttgccgggttctagagcttaaatgctcaa
 tccttgcaaaatataatgggtttgtagcttatctgatgttcaataaatt
 ggggaacatgagatagaagcctatgtgagagttcatgtgctatgcccgt
 atctggggaacaagtaaagatgcagggcgtgaagtagtgaagatctttg
 agagcgtgacgaaggttgatgcaagaacctgacattgtgaaaataaga
 ttcaatttcattggatgctcaaggtaatggtgggattctcgtacatga
 ccagggggtaaagctctggaattgtagctcctctaggtttataaattcta
 tctctaactttttatcttttagacaaaatgaggatagatggatagttatc
 tctagtttaattttattgactctatgtttgatgatgatgagtcctttat
 108001 catgccatgtcacaataatcttgtttgtgattataaaagtgttaattttg
 atttctttgatgttacgaatgatcttgcctttatacctagttggtgatag

Putative non-LTR
 retroelement reverse
 transcriptase
 (ORF30) ↓

tcaaatgtgaaaagtgaggaggctgaatctaaggagaacactagaaat
 ggtttatgcttttcttaaagtagaggtggctctatgagtgaggccca
 aaataggaatgatgagtccaaagaaaagagatgtgaagtgatgctgt
 aaaaaggcctagtgggtgaccctcatgccactataggatactgag
 tccaaaaaagagaatccaataggaattgcaattaaggaagagttagt
 atcttgaagaattagattgattaagtgtcttgcagtggaactgtagg
 gta ttaggagaaatgaatttcttagtaattttaattttatatgtagt
 ttagct
 aaattggacatTTTTGCTTGATGGAGATTAAGGTGGATGTTcataaggt
 108501 ccctaggaacttttattccaagttttttgacaatattttttatttgatc
 caattggccgactctggaggcctctgcttatgttggaaatgtagtaaggt
 tttgttagtataatttcttctcctccgagatataatccattggtctataaa
 ggacttagtactaatttagagcaattagttactttcatttatgcctttc
 cccaaaaacatatgcagtgtagcctgtggaggaggctcctcaagttagat
 ccaattcataatccttgggtgcttagttggagattttaacaatattgttga
 gctttcagagaaggttggggtaagcaaccataccacggacatgaaca
 attttatcaatttcttaatcgtggcaaccttttgcctttaaatgtgtct
 ggagttccttttacttggactaattgtcatacagatcatcgggtggttt
 caagagattggataggggtggtaaccctttatggttaaatatggtcc
 109001 ctaattatcatttgcatacctatcctatttttggttctgatcatagttct
 attttattatctctaaattctagtgggagaagaataactacttctcgtt
 taaatttgaggctatgtggcttaatcatcctcagtttaaacggttttgc
 aaaatttggcattgtgaccctagatataatcctaaagataaattcagag
 tgtgtgctggtaagtttaatttctagtttgaatgaacaaggacact
 tttggcaattaaaggagaaaaagaaagaattggtatataaaattgatga
 gcttcaaaaacaaattatgtttaacaactctggtgatttgatgatggaaa
 aatctctgtgagatgaattgaatcatgtgtaatttgtgaagaaattatg
 tgagcccagagagctaaaactaattggctgcaacttggggataagaatac
 taaattcttcaactgtggctactattcgttaaaaaaaagaaatgaat
 109501 taaaaagtcaaagatgagaatggtttttgggtggtctgttgggtcagggcc
 ttgaacaggttttgtgacagattttaaaggcacttttctgtgacagt
 cctccctggatgcttacctaagtaatttagcggatatacctcaaccttg
 tgttctaatcatcataatgaaagttgtgtctctgttacagatgagg
 aaatttgggatgctgtgagtagtattggagctctaaagccctgggtctt
 gatggtttaagtgtggcctttttcatggttgttgggaccaaattaaggg
 ctctgttctcgttgggtcaaagatttttccaaaatggttccagcttaa
 gactcataaatcactaacatcactttagttcctaaaaatataatccg
 gaaaattgtagcaactataagcctattagtttgtgtaattgttggttataa
 aatttccaaaagatcatcattaagagactaaaacctctattagatcttt
 gcatctctcaaaatcaaggtgttttctcgggtatagccaatcaggat
 110001 aatattcttattgctcacgaaatggttctgctgatttcaataggatgaaagg
 aagaactggagctatgggtgtcaagctggatctggaaaaggcatatgatt
 agttaaacgggattacattcggatgttcttcttagatttgggtttcat
 aatcatttgggtgaattagttatggaatgtattacttctccttctccta
 tgtgcttataaatgggacacctcattgttatttttatccgaagtaggaa
 ttgcactaaggacccttgtccccttatattttcatcatttgtatggaa
 ccgttaagttagacagcttaataccctaacaataaccctaaagtctcata
 tgggtatttgcctcacctagaggatttaaaatctccaatctagtgtttg
 ctaatgactgtctcctgtttgctaaggccacgcccagggtgaggaat
 atcttcaaatccttaatatgttttctaaaacggcagggccaacaaataaa
 110501 ctttcataagtcgtccttgtatttttcaagaataccacttctccaacta
 gaattgatattgtaataattttacataatcaacataaaaactactattgtt
 aagtacctaggaattcataatattgtttctagaaaagaccaagttaatgg
 taatgagttacttaagagggttaagcaaaatagctggttggaaagcaag
 cactcttcaagagcggggaggcttacttacttaactctaactaacgg
 gtatgcctaatcatgttatgtcttgaaggtcctgacaagattgtc
 aataagcttactaaagagtgtagggatttttctgggttaaggacaaga
 aaatctatccagtggttgaattcggctctatgctcctaagaaaatggga
 ggttggagtagaagacttgatcattttaaataatgcttgcctcgtctaa
 gttgggttgaaggtaatgactaacgaggataacttgggtgaaagttag
 111001 taaaaagtaagtaccatagaaaggaaaattttctagataataaaatcatt
 cttttacttggaaaggtatcttgaaggtagggatgtattcttaagggt
 tatgagatgtagagtggttaaggttaaggacattttgttatggtcccaca
 attgggttttcttaccattgttccatctcttaccctaatattgacaac
 caaaacattaatgggacttgaaggttaatgaatttattgataattattg
 ttgggataggagaaatcttccgcaagtggttagtaatgaattgtgagga
 aaatttgcataattacaattcccttgcaggattccaaagatgtactagtt
 tgggggtccaaatacaaatggtaagttctctgttaaatcagctacttggat
 ccataaaggagcgttgactcaatctacaagcataagtttaattaaga

Hypothetical protein with
 Heavy-Metal-Associated
 (HMA) domain
 (ORF31) ↓

111501 tgtggaaacttaatatctcctcctaaagttaagatggttgcgtgggtgttc
atccatgatagacttcaactaggaaaagattaagttatattgtaattaa
tagtgataagatgtgtcctttttgtaattccatagacgaggaccaaagtc
atctatcttttaattgtggtgtcactaaattgatttgggaccttgttca
atlaagtttgggcaatttcaactaactaactcgcactcagcagtagactg
gctaaattccttaagcatggtgataaggaagctcctagtgtttgagct
tgtgtttgctcctttgttggcaatttgaatgaccgtaataataggtg
ttcagaaatgtttttcccacccagcaagatctaagttattggctatttc
tatgggtgagcgatatttaagccaatgtaaggtttaaagaagaata
ctaagtacttccacaattgttaaatggcatccgctagataagagtttt
gtaagttaaattttgatggttctgtggctggagacaaggttgcgtcgag
112001 ttttgccttaaggaatcataatggacagattattagaggtggagcttaa
atctggatggtgccactatttctaaagtggaggctaaggctctcagagag
ggcctctttttgctcagaggaatggctataaaaagattatgattgcaggt
gactctaagttggctatacagtcagtggtggaccgggtgagacaccatg
gaatatggttcccaatattgaagatataagtgacgggtgaccaatttg
attgcatgtactagaagcatattatagagaggctaactttgtggcagat
gtcttcgctcgcactggccttcttattactgattgtcatatttggaaatca
ttgtattccttcatttgccttctcactctcacttgacacttg
gtaggtgctagatggttggagctggctactgtatataatttaattgggaa
aaggatcctcgcggatcatttttctggagattccgggaatcccatgatc
112501 gtgaccgtacgatcacgatgaacgggtcacgatatttaatttttaaatat
aaataatacctaacgaaaactgaccacagatatacaatgaacgggtcacg
atcatgggatcccagaaaaggatcggcgaggatcttttctatatac
acttccactccaatattgtgagacttgagagtgaattatataccgagga
aatgatgaattttgcatgcccttataaataattgagctactctccatatt
gctaattggttttatagtggaatctcaatttcttcatggtattacagca
ggttgtcccatgggtgaagtctaacgccacacgtgctcacgtcacccaat
ttgtgttgcacttgctatgcttaaaaatttggccacatgtaagagggcg
tgttgagagtgaatcttacctcagggaaagaacaggccttacctgtgctt
ataagtaattgggttacttctcatattgccaattgattttatggtagaac
113001 acaacttctttatagtatcagagcaggttgttccatttgtgaagccaa
cggccacacgtactctacatgacccatttatgttgtccatgtattaggc
ttgaaaatttgcataatgaggggtgtattgatagtgaaatccttcatcgga
gaaagttatagtggaaccccaattttcttcatggcatcagagtaggttg
tcccacatgtgaagttcaacagccacacgtactccatgtcacccattgg
tatggttcacgtatttagacttgaaaatcgtcacacagacgtattgagag
ttaatccttcatattggagaaaagaggaactttgcatgtgcttacaagtaat
tcgactacttcccataattaccaattatattatgatgaaatatacaacttc
ttcacagtacacaataaataatgcactctttttctatacggatcgaaact
caaaagcatcaagcaatttctgttagataaacaggagcatcaagtaatttat
113501 atggcagaaatccgtacgtcttgaggaaagtcgcttggatgactagaaa
acttgtaaatgcaaaactgttccaatttcttccagatacataaattatt
agcagagtgaaaactatgtgttcttgggctgtttttcacaatgaccaa
gtcaagcaacttcaaaactggtaggtcttgaccgttttgggtgcatatc
gctaagattcttatcaattctaaaaaaacaaattcatgagttataata
tagttaaataaccgttttatgatgaagcaaatagtgtaataaccatttc
acaggttataacatataaattgactggttataactaagtaccacataata
tctttccaaaatgggatcaaaaatggttagctagaatagaagatggtta
agctcattcagactaggatggcgttcgattaaagtgcattagaaaaatta
atgtatgtcaccaaaagacagctgcatgataatcaattcacgtgctata
114001 atactggttgattctttttacaattatagaggccattcagacattaataac
tccgataagaacataaaaacttattttatcctactaaaacacagttatgta
agtagtataactaatctcccacaaatgggtccaaatattttatctaga
atataatataagatggttaagctcatccagaaccaggctagcacactcaa
gaagcagcattagggaaagaaaaaagagactcaaaattagtttattaaa
gttttggagtacacactttaaactcttctgtgaaagaatctgcaactcc
aatctccatcaactttacgactaaaatgaaagaaggtcatgcttccca
accctaatttttgtgatttgcattgtgaaaacacccatctttgcttgt
gcaaaaaggattccacaatattagtggaactccagttaaaacgtcgaaag
ctaagctcaatgaaaatgtaacaaaatcatctgatcatctccttttat
114501 ttctttatccaaggacggtggatattcctaataatctcgaagacagacgt
cctcaatcgcacacagccatcaattccatcacacagaaagcaaatctt
tttggcacagcattccatcgtccatataaaaaccgaccttcttccatc
cttttaacacatttgacattgatttttgcaggttaagttaatcaattcga
aaaggagattccttgcattgcccacatcttttgcctttctattattcctt
gocatttttatttttatttttgcattgtgacaaaataaaatgccaccta
ttaatccattgcccctcctctctccctccctccctccacctcttatt

115001 g c c t c t c a t g g a g g t a t g g c c t a a a t t a a t g g c a a t c a c a t g c t t c c a c
 a t t a c a t g t a t c a g t c c a t g c t t a c a t t a a t t t t t c c g t g t c t c g a a t g c
 a g a a c c t c g g g t g a c g g a g a t t c a t g t c c g g a t g g a c t g t a a c g g a t g t
 g t g c a a g a t c a a g a a g c g t a c a t g g c a t c a a t g g t a a g a a t t a a g
 a a t c c a t g t t t t t t t t t g t t t t t t g g t t t t t g g t a t g t t t c t t a c a c t a a
 g a a c c c a g c t t t t a t t a g c t t g t a g t t t t t a t g t t t t g a g t g a g g t g
 t t t t t t t t t c t t c t g c a g g t a t a t a t g a t c t t t a c a t t g a c t t c c c t c a
 a c a a a g t t g a c a a a a t t g g g t g g g c g g a t c c g g a g a a g g t a g t a a a g g
 c g a t t a a g a a g a c g a g g a a g a t t g c c a c c a t a t g t t c c c a c a c a g a a c a a
 c a a a c g g a a c c t g c t c c a c c a c c g c c a g a g c a a g c g c c g g a g a a t g g t t c
 t c c a g c t c t g a g g c a g a a a c c c c c t c c t t c a g a a g c a c c a c c a c c g g
 c t g a a g c a g c t c c g c a g c t g a a g t g g c c c c g c c a g c a g a g c c a a c a a g a
 g a a t c c c c g c g c t g a a a a t c c c a c a c c c g a g c c a a c a g t a c c a c c t g t
 115501 g t c t g c t g a g a c c a a c c c c g g c c a a c a a a t a c a c c a c c a g a g a c c a a g a g
 a c g t t g g a g a a a c t c a t a c c a t a t a c c a c c a c c a c c t g a t t a t g g c t a t
 a g a t a t g g c t a c t c c c a a g g c t a c a a t g g c t a c t g g a c a g a t a c c a t a a
 c a g c c a a g g a c c t c c a c a g g a g c c t g c c a t a t c c c c g c c c g a t g g g a c
 c t c c t c c a a t g a t c c a c g c c c a a t g g a c c c t c c t c c a a t g a g c c a c g c c
 c c g g t c t c c c c c g a t g a g c c a c g t c c c a a t g g g t c c t c c c c c g a t g a g
 c c a c t c c c a g t g g t t c c t c c c c g a t g a g c c a c g c c c c g a t g t g t c c c a
 c c c a c c t c c g c t a t g c a t g t g a c a c a g c t a c a a c a t a c a g a c c a
 t c a c c t t a t a t c a c a g a a t t g a a t a t a t c c a g c c a c c c c a c a a c c c t c
 a c a t t t c a g c a g g a t g a a c c a t t a c a a c g a g c a t t t c a g c a g g g c g a a c c
 116001 a t t a c a a c g a g g a g c a g c a c g a c g t c a a c g g c a a c g g a a a c a t c a c a t c a
 a t g t t t a g c a c g a a a a t c c g a a t g c g t g t a c c g t a a t g t a g a g a g a g a g
 a g a g a g a g a g a g a a c g g t a c a a a a g g a g g g g t a c t g t t t a t c c c c a a a
a g a a t c a c g g c a a g a a c g t t a a t g g a g a a t a t g t t a g c a g g a a a a t a t a a
 t t t a a t g g a a t t a t a t t t t t c t t a a t c g t g t g a t a t t a t g g g a a g a a a g a
 a t g a g a t t c c a a g c t t g t a a g c g a a g a a a g a a t t g a t c c t a a a t c c t c t c
 c a a c a g g g t a c t a c a t a c t g c c g g c a a c c t g g a g t t t c c c a a a t c g t g t
 t c t g c c g a a a g t a a t c t g a a a a c c a t t g t t c c c a g t t t c t c t c c t c a c a
 g a a t g c a a c a c c a a g t t c c a a g a t c a a a c t t c t g c t c t g g a g a g g a g c
 t t c c t g g c t t a a t a t t c a t c g g a g a a a t c t t a c t c g g a t t t t c t a a t t t
 a t a g a t a t a c g a a c a a t t t t c a a t t a c g a g c a a a g a c c o t t t c t t a t a a t
 t a a g t a g t g a t a t t t t g g a c a a a a t g c g a a t g a a c c g a a c g c o t a a t a t a
 c a a t t a c c a g c a a g a t a a a t a g t a c g c g g t a a c a c a a t t t a t a c a a a a t c
 g c t t g t g c t t t t g c a a g t c a c a g a a t a a t t g a t a t c t t g g a g a a t t c t a a
 a t t t c g t c a c c a t g g t a a t g g g t a g a g g g g t a a c c c c a g a g g a a a t g a a a
 a c t g t g g g t g t t g a a c t t g g a c a g g g c c a g a a a c c g a c a t t a a a a g t t g
 c t g a t t a a c g a a c g g g a t g t t g a a c a g a c t g c a c c g c a t c g a g c c a t t
 c t g t c c t a a a t c c a c a a g g t g c a c a g g a a a t g c a a c t t a t g a g c a g a a a a
 c t g c a a a a t a a a c c a t a a g c c a t g a a a t c c a a t a a c a c a a c t g t g t t c c a
 117001 t g t c g t c a t a t a t t c a a t a t c a a c t a a a a t g t t c c a a a t c c t t g a g a a a a
 a c t a a g c a t g g a c c a g c t a g a c c g t a a t t a g a c a c c c a a t t c a a a a g a a t
 a c a t c t a c t c c g c a a t g t a g c a g a a a c c a a a g c a t g a g a c t t c g t g g a c c
 t a a t t c t g g t g g t a t t t a a a c t t a a g c a a t c t a g t g t a a t a a t c a a t c a a
 a c a g t a g a c g a a t a g t g t a g g t a c a c c a a g t c a t c t c o a t a c a c a c t t t t
 t t t t c t a c a a a g a a t g c a a g c a t c a t c a c a a g t a a a t c a g g a g a c a a a
 a g a a t g a a t g t c a c t a c g g c a a a g t t t g a c t a c c t c c c a t c a c a a a a t c c
 t a c t g a t c a t c a t c a c t g c t t a t g t c a g a a t c a t c g t t t t c t t c c c c a t a
 g t g a a t t t c g t t t a c a t c t t c a t c g t c a t c t t c g c t t t c a t c c a t t t c g t
 g c t c g t g t g a a a t t g a g t t t t t g c g t g g g t t g c c t t g t c a a t c t a t
 t c a t c c a g a g c a a a a c a a a t a t t a a a t a t c a a a a g a c t a a a a c a c t t g g
 t c a a c t a a a t t a c t g a a t t c a a g t t a g a a t a t g a g g t t g t a t a t a c c t g t
 g c c a t c a a a g t t g c a a a c g a c c t g a c t t g t a a t a a a g g c t g a a t a g c
 t g c c c c t c t g g a t t t g g t a a c c t g c a a g g a g a a c a c a t c a a t t a a g t g c
 t t c a t c a a a t a t a t t a t a c c t a a g a t c a a a t c t g g t t c t t t a g g a g g t
 a t t g a g a g t t t t g g a t g a t g t c c c a c t a a a g g c a g a t a a t a a a t c c a a a
 a a c t g t t g a t a t t t g g g t g g a a t t t g t g t g a a a a t t c a t t t t g t t a g a g g
 t a a c a c t a t a a g a c a t t g a c g g t a g t c t a a c c a a g a a t c a t c t a a g c a t
 t g g t t t g t c t a t c a a g g g t a a t g g g t a a t g g t a t a c c a t a t g t a t a g g
 t c t a a a g c a g g c a c c t a g c a g c a c t g t g g g a a t a g t a c g a a a c t g c a a g
 t c a a a c a c a a a t c t t c a t a g g c t a c c g t a t t c c t t c a c g g a c a a t a t t c a
 118001 a a a g a a g a t c t t c a t c a a c c a c c a a t c t t a a t g a a a c t g t a a g t c g a g
 t t t t c c a t a t t g t t t t t c a g a a t a a t t g g a t t a c a a c t t a t a c a a a g t g a
 a a g a a g a c a g a g g g t a a a a c a t t g c t t a c c a g c a a c a t g t c a t t t t t t
 t t a g a c a a a c a a t a a a a t t t a t t a a a c a a t a a c a t g a a t t a c a a a a t a

MC-1217P5SSR316For

MC-1217P5SSR316Rev

Hypothetical protein with
Glycosyl hydrolases domain
and WD40 domain

(ORF 32) ↑

gtggataagaaccatccagcaaaactagctaagtcctctcaagtagtct
atccaagtaactaaaactgaacccaaatcattttacagctgctaacac
atcccacagtatgtgagaagagaataatccctgaaataagggagcaaat
118501 agtgacctggtataaaactagtactaattgggcacaaaattttaccttca
acaaggagctaagcatctgggttcagacttttgggacatgatcatga
ccgtgattaccaatgctgtaaatccatggaagaacttattcccact
gctagacctgttaatacataatcaaagattaaaactcaagtcatactaac
ttacactgtaatgctttccaaatggataaaaaaacaaaagcataaagt
tcaataacagctctttataaaaaccaaactcacggaccagagtcagatacac
attagcacaccagatactatggaattcagaattctacatcctctcaaaa
acacacacggaggaaaaacataatctcatgatacatgcataagttactg
gtagatttaaggtacaaattgtctgtataaaaccataactaacaattcat
tgttctaaaaagtgcatatataaagaaatctacacctcaaaaggacata
tataactctacattcatgcatgcatgaaatgaaaaccggaagaagaccaa
119001 aacagcaagaacaaacctgtttgcccagtgccaccagaacccccata
atgtgcatgctgactaggttctaataatgatgacacagcttccctcatctgg
cagatgatacgtatagatcagaacaaaatcaaatcaggaagtcacataa
caatataataagtaacaacatcatgccactgttcatacataagtaaaact
aaacaaaccaaccttcttatgcccataatctctgaagatcaataacct
taaatgttgagaattgagtgctgagtggaactgagtgatggtcatctactta
ctaagtatccctaatgctttcagctctgctcctccatgcaaattgctgctgt
gtcagcttcaagttctacatgcatctgcacataaccagcataacgtatg
aaacaagtaaaaccagcatcaggataccagaggtcaaataacataatggta
agccagtttgacactcaaaaataagatattgcatggtggaacagatctac
119501 agtcatcactaataaagcctaagtagttatataaggattacttatgttaca
tgcagaaaagttcaaatgcattcctgagtgactataacgccacctttg
ctcttcacaaggccagcacgacacctgctgcatccaaatcagccatttc
ttcatcttgccaaaactcaatttatcatgtcttttcttttctctgctgtg
aatcaaatacctttggcattggatgcaatgcctcttctgcatttgctcga
tctaatacaataactgttgaatcattgcacaaaataatcagctgaccgtca
aaaatacaagaaccggaataaaaatgtagggttatcacgatacgaattta
120001 taagaaacaataaatgttaatgcaagtgcatcatcaaaaaaagtaaaa
taaaattttggttttgattcctctcagaagggtccctaatgcatacgtat
ctcatgctatatagccagggaatcaacaacaaatgtgactaaatgtgg
tgagggggggaaacttgtgtgagaaatcttacctctattttgatcat
tgctccttcttagatttgacaagagactgactcattggtagaagaacccc
gtcatgggagctactcaactttatgtctgtccagaatgcacccaaaagt
tctggaatgaagtttccactggtgaccataagcaacaataacctgcca
gaagcagcttgcgaattccttgaataaccgctgcaaatattgtgggtaa
tgcaccttgtatgtttagacaaaatgtcttcagaagatagtgaaattt
ttgtaggcttgcactccgtaattcttcaatattctgtccaaacaaaag
120501 taacacacaccaacttctgaaatcgctaatacgtatagacctacatctcc
aacttcccactgtctatgcacctgctatctacgaagatggcagggtgct
ccattttgagaacaacacttgtgactgctttttccaccatctatactc
cactactgcaatatacctttaccaactgcagatgagaggatatacttccc
atcttcagtgaaatcatacagcgaacagcaccctaatgaaatgggtgaga
aaatattaacttaaaaaaatcacgggttaaaaaaatgttaaaaaataaaat
gaaggttagccctcctggtagctggaacttttggacacaatggtaaacca
aaatattgtccaggcatcaaacatcttaaatcaaatcgtacaaatgc
agctctcaaatatgttgaattacatgctggcctctcaatgggtgaaag
agggccacatagaggaggaagtgcaaatataggaaaatagttgccta
121001 acaaaggacattctaacagtggttaaccaacaagattatttgaagataa
gcagactgtaaacatgtacacatgtttaagcaccacaataagcaaaaagt
gtaaaagaaaaggagaatgtagatgcataaacactcacaggatggccaga
aaacttctgtatcttttgggtcagacaaattgaagatcttcatttgg
cagctgcagttgctaataatattccatctgggtgaggataaaaggccatcc
ccaatgtaactacagcaaaaataactaaatggaacaatctacaatacagtt
ttatttggatgcacataaaagctggctcacaataattgcaaattttaacc
tttttttaagcattagatcttcaacaatctaactgaaacctaatcatcc
gagctctctacctatacacaggaactgcaactagaattatcactaattt
caccatcgaattcctttcttctttttatgacttatgacaaatcagttct
tggacagcattagctaccaacttgccattcatcatcatataagttatca
tttctcaacttaattagttattaagacaacataaaggcatagctaataat
acaataaagacaggtttcagtgatgatgacaaattacaaaagcgatccc
121501 atatacttaactttcacatctctctaaaactactgtcctggtaaaatct
tggtttatgcaggagaatgctcattgacttggaaataattttgcagccaa
tcacgaaaacttgaatccagatagtatattcgtgatgtaattttcacat
agcagtatcaaacacaaatttaagcagaatgatgcataatgctgtaagcaa

aagtagaacaaaagatgtagggtgatggacatatgcatatgtaacatact
tgatgaaacagacatagaggagattgcctttgtggaagctttgaaactcc
ccgacaggtttcctgtcaacgaatcaatgtgacaaatcatcccacagca
ccggctgtgtaaagacatgatgtatcttctgcaaatgatatggagctaac
accgctgcagcgcagaaacatgcaacagagattaattagcgcccaaga
agtcaaaaacatacattaacacaaggcagtataaacatcaacataatatt
ttgacaagaagaaagttaagcagggagctaacccagcatggcaatcgt
122001 taactctcoacttcaaatgaccagcagccacatcaagcgccaaaacatca
ccgccaccagttcctaaaactaaaaatgaggaaccaagcttcctcttctt
ctgcaacaatgaacagtatagcataaaaaataaatttaatttaaaaaataa
aaattgctacagcatactgttaagctaaaaacaaagcaaatcgcaacc
tttcgttcggagagaaccatttcctgcaagtgaatcaactgagaggtg
ccctcttctcgttttagcatatatgcttgtatcttcggatgctacaatgt
ctgtaaacctcagctcggacctgccccttcacagtatcccaaatctaacac
atcccagaatgaaaattagggtttcataagaaatagagaaataaatc
aaccaatttcataaaattacattaccttgattcgaccgtcgccggagct
gatggcgaaacaagtccagagaaggactgaacgcccgtcaggagatctctga
122501 tghttgaaagagccattccggaagttagggtttggtaggagaggacaaag
gatgaggaggtgcagagctaaaaccctaaaccctagaacgaggagaaag
ggttggggcagctgaagtgatggcggcaacagctctgcttttagtttagg
gcgttattgatttttaattgtgggctgtttcaaatgttgggctttatttg
tctaggccaatattttctttttgtcaatttgatccaatgttactc
ttttaattttttttttataaagctatccacatattcattttacttct
gcaacacacttctcttatctgtccgattgaatgagttgaaagaagatcaat
gaaaaaaaacaaggtgtagagatgaaaatgcgtgtgtgaatagcgttat
cgattttgttttaggaaaagcatattttagagaataaaaataacacaatgct
gactcccaatcgcatctcatgtgattacgaggcaacaattgaatggaaaat
123001 tttctctaagattgaatgtcgtccaaaaatgagaggtaatgcttttcaa
cggttaaaataacaataatttactatcacaatacactagtatgtagt
caattagcataccctaattcacactccctaaaaagagaccaaggtctttt
gtgtttgaaacttgaattattatgatttcggactaaatcatattgctaa
aacatttaccgttctttgtaattaattttgttggtaggtaggattatg
gttattacttgtgacttaaggcaacaatcataacttaaaaacttgtttt
tttttccaagagtgaaaacgagtatggactaattttctgtgacgtcttca
tatgttatgaactcatgtaacatggaaagcgtgtaatgataaagtttaa
cgtcgagatttagtgattaaaatctaggggttacgaaaaattagatcat
gaaacataactaagagattttattaggaacaagcacaactagacttagc
123501 gcttatgaaattatgaatattcaaatatttttttaaatgttaaaaaata
aaaatttaaaacttttttcttccgaacaaacgatattatccacattaaa
ggatgggatgagcttaacctcacaatgagctagcaataatgtgattcaaa
ttcccctttgtataggaagaggaacatcactagaccatagtactaagtaa
caaatttaaaccttttaaacggtgaaaaaagaagtatcacaagagataat
atattccacgcaacgttatgtttggacaatcgaactgcatgatcgaaa
ccctaaaccatgaagctccatagaaattaattgactaacaaggacacac
accataggcattcccttctgtcttaaccttagcgcacaaataccctgtg
tataatttaaatccgctcttccccttcttaccgtaccatgactcaccgga
catgtttcccacaacagcaataacaatttcacaataacaggattttctt
124001 gttcttcggtgtatcctatacacttggattttaaagccattgaatcccga
tgtaccggaggatagtagaatttttgcaacaaggacactacacacaccgga
attttcaggtaaaacggaataaaaattgaaacagaaaatacgaacaataac
ccaagttcgtctacatcatcgtaccaaaaatttgcttacacatgatga
agaaatgaagctgtacaaataatctgcgggatgtggaagaagagatgcg
ggactgcaaatctacaggatgaagaagaagaagaagaacaaataatctg
aacatatgtgtcttattgaaaagaatgtagtctaatttttaattggcaatg
gtggaataatcttagtcttccctcaccgtagattttggctgaaaagta
taactggattttggaggaaattggtattataagaccgcgcttttccattc
aaggtgagtgcgaaaagcgggaggaggaaggatttgaatacagggtatt
124501 gaggtcggaactggctctgggaagacagactcagaagagccggaaacat
atgaacaattccaggtggtgggagatgcagaattcaatgacaatgagatg
ttggatagaagaagaagatgaatggcgattctttgatcccagtgaaagca
tccggcaattgtgatgtttgtgccaaccacaccttgcaatgtgatgat
ccaaaactgggagagcattaggatcgtatttgtcgtcaggatgggaacca
aactggccagaggcaccaaatgccatgtagatgttttccatttcgacatc
tgatgatgacgctctttgatgtaaccgctctaccctttgttgttcgaa
actgaatgccgctaaatgagttgttgaggcggaccttctcaacaagtacg
ttagaaatgccaccggacatctcactaccaaaggcaagagaagaaccagt
agctgatgtgaaggtgacacgcccgatgtgacgtttgcagtcggctctc
125001 cataggcaatcccgtactcgtccaaccactcttgaggccaatggcatca

```
taccccatgccaatgctgcaatcttctatgcacacaatatcagaagaatc
tgaaagcacagagaaaaattagtagcaatctcgccaaaaaggcgaaaagt
tctctcattcaattatcttcaaaacaagaagggaataatcaggatgcacaga
gaaaataactttgtaaaagtgggcaggaagctcaaatttagccaacatt
tctacttccattcattcgggataagaacaataacttcaacgtaaatcaca
gcaggaagtcaagagatgtcacaacacctgggactatgccgacagtatgag
gggagtccggaggagcagatacagagatgttgtgaacatgtacattactg
cattgtatcgattataaataaataaaaaatcttctttaaccgagaat
gaatgcaatgaaagaacaagcaccgctcatcagtgcccttagctaaca
cacctgcaatagactggatggatggttgaagcgggagcattcacgaatgt
gaggttcgaaaccaccagctatcttagacataacgaaaaggagctcg
```

125046

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 attccatggctagttgtatccgaaaagtagcaaaagaggtattaggagag
tccaagggctttgcccacacccaaaaggaatcttgggtggtggaatgagga
ggtataaacaaggtgaaggctaagaaggaatggtgtaaaagccttataca
aggataggaccgatgaaaatggtgaaaggtatagaaaagcgaagcaagag
cggaagaaagctgtgagagaagctaagttagcagcttatgacgatatgta
taagcgactagataccaaaagaaggagagttggatatctataaactagcta
gagcaagggaaaagaagacaagggacctaaaccaagtgaggtgcatcaag
gatgaggatggaaaggttcttgcctacagagAACCGGTTAAAGATAAATG
gagaggttatcttcataatcttttcaatgaaggacatgaaaggagtgctt
ctttaggggagttgagtaactcagaagagtgtagaaactactctttttat
501 cgtcgaatccggaaggaagaagtggttgtagctttgagaagatgaagca
tagaaaagcagtaggcccagacgatataccaatcgaagtgtggaaagttt
tgggagagacaggtataaacatggctcattgacctttcaataggattttg
aaaacaaagaagatgccaaatgagtggcgaacgagccctttgggtgcctat
ctacaagaataagggcgacgtacaaaattgcatgaactataggggtatta
agctaatgagtcatacaatgaagctctgggagagagtcattgagcataga
ttgaggcaagagacacgggttccggacaaccaattcgggttcatgccagg
gcgctcaaccatggaggcaatctatctcttacgaagattgatggaaagat
atagagatgggaaaaggtttacacatggctctttatagatttggaaaaa
gcgtatgatagggctcaagagacattatggaggattttagagaagaa
1001 aggagtacgagtagcatatccaagctataaaggatagtatgaaggag
caaagactgccgtgaaacttatgaaggacaaaccgaaagctttcccata
actgtaggattacatcaaggctcatccttaagtcttaccttttgcgtt
ggtaatggatgagttaacaggacatattcaagatgatattccttgggtgta
tgcttttcgcatacgatatagtgttgatagatgaaactcaggaaggggta
aatgcaaagcttaacctttggagagaagtgttggaaatctaaaggtcttg
cctaagccgatcaaagacagaatatatggagtgcaagttcagtgcaaatg
gaagccaaaacgagttaggggtgaggatcggagatcaagaaataccaaaag
agcgatcgttttcggttacctaggatctatcttgcaaaagaacggagaatt
agatgaagatctcaacctatagaatacaagctggatggatgaagtggaaga
1501 gtgcatcggcgtgttgtgtgaccgccgtatgccactgaaactcaagggga
aaattttataggcggcaataaggcggcgatgctgtatggcacataatg
ttgggcggtgaagcatcaacacatacacaaaatgggtgtagcggagatga
ggatgcttcggtggatgtgtgggcacacaagaaaggataaaaattaggaat
gaggatataccggggtaaagtaggagtacccgaaattgaaggaaagatgag
agaaaattggttacgggtggtttggacatgtgcaaaagaagcctactgacg
ctccgattagaagatgagctatgggacaaaggttcagggtcgaaaggggt
agaggaagacctagggaaactttggaagagactctaaagaaagacttaga
gtacttggatctaaggagacatgacacaggaccgagcacaatggcgtt
caaagattcatgtagccgatcccactcagtgacttggattttccaagctct
2001 tcaaccgagaagttttcctcactcggaaaattaagggaatactacctcaa
cctacatgctccactcacaagcttcaacatacaagcttcaacaaaagaa
```

RNA-directed DNA polymerase
(ORF33) ↓

aattcagagaacttagcgaagaaggctttggtgtatttaacacaatacgt
tgaatgaaggaaagctatttattgatatccccgataagtcataaatatg
tacatatacatgagtcAAAATAACAACAAGAGGGAGCCTTCAAAAAG
TTGCTTAGGAGAAGTCTCAGCAGTCGGTAGAGCCCCAGAAAGAGCAGGCA
CTGGAGGGGATCATTGGAGCCTCAGTACTGGACAGAACCCTAGAAGGA
GGAGGCAGCAGAGTTGATCATTGGAGCTTCATTACGCGGTACAGCCC
AGAAGACGAAAGCAATAAATGCCTTTGGAACAACCCACAATCTCTGAT
2501 gatcaagtaaaacctgaccatcggattccttcatctggtcaagcttcctc
ttcatgtttgtagtatagtcattgtgtgagctggtgcaactgtttattctc
atgcttgaaccctctaactcctgtttgagactcatcacttcagccgtca
atgattcaacttggcgggtcgagcaaataggcgttgggcatgttagac
acagaacctgcacactgaacactgagagccagagaatccttaacagccaa
ctcatcagaccgtttgaaagttagtctgttatctttgggagtgagaaggt
tcctggccactaccgcagcggctcatatcattcatcacggaatcccc
acggttaagaggaccagtaggggagacgaaggatgggcgcatatggtgtc
tggagaaggcgtggctgcctttcaacaagggtcaagtcaaaacaacggt
cgaaggaggtcggacaaatcaagatcttagaagtgaagaatggagcttc
3001 tactggtgagattcaagtgtgctttggaacttaatgccagcctctataa
aaatctgcactcgacggagcttcagaaatcgaagaggcatttgctttctc
aaaagctgggctactcagagaccccgagacaattgctttctcaaaagttg
ggctgctcaaaagaccacgaaggccgatctcagaaatcgaagaggcgtta
ctttctcaaaagctgggctgctcagagaccacgagagccgatctcagaaa
tcgacgaggcactactttccagccttgtcagcacctgtcagctttgcg
gaaattatgggcattatgtcgaagatttctggtgaagttagaaagcactg
aatcttactgttcaaatcaccacttcccacacgcaacattagctcatggg
taccacagataactttgcaaaagttttctgacaaagttgagacacgtgaa
gcttgcagctcccactacaccgctctgaccaagaagggtaaaagaatagc
aaagaaacagcactaacaagtttagacacataaattttgaaggtctagc
3501 taccatattattacctacaagggtaaaggacagtagcactgctggataa
ttggaaagtccctgtgtgtcaacctctgtgcttcatggcaagatagacta
gcaaacatgcccaacctttctcacattcgagaaaactcccacaaga
ttgctttctcaaaaatcgaagagacaccgctttccgaatctcgagagcca
gaccctagcaggattggttttcaaaaatcgaaaaggcatcgttctccg
aatctcgagagccagatctctgacaggattgcttgttcgaaaaccgaaga
ggcaccacttcccactcgaagagccgatctccttggataaagcttgt
ctgtaacttttacacgcaacatcagctttccagataccacagaccacttt
tcaaaagtctctgacagagttaaaacatgtgaagctggcagctcccact
accgtgttatgaccaagcaggttaaaggaaatagcattactacttgttgtt
4001 agggagactcctatatatgttgacctccatcccacaacacagggcagacc
tgcaaaaatgctcaacccttctcatatctgagagggcactcccacaaga
gccttttgaaatattcagctttctttcccccgataatacctctgcaaac
aagctatattagagcaagaatatctcatatcatcagggttaaaagcaaga
gtatcccataatcatgtttttccctgtctttcctttggccttgttctta
cctgacaagcaaggagaaagagagcaatcaatcaacacttggaaatcaagt
ttccagtcagaaactgactacctggaacccttacttgattacttacctg
gcattgctctcgagtactcatcttcaacatcttatgcttccaggggaagat
accgcatctacctgaggaacatatagggcaagtgagaaagatacaaggaa
gcatgtggagacaagcgtaacagcacacgtgccgatacatccactactct
4501 gtcaaaagcaaaagtatcccatatcagcagggcgaacgtactatagatt
tgatggacttgtttgacctcaaatcttctcagtcggccttatactctgg
aggaaccagaaaaccctccagccagttcaagaataagcctatggaaag
ttacttctcaaaagtaaaagtatctcatatcatctcttctcatttttct
tctctttatccttcatgctgctgcaagatagggagaatatgaacaatca
gcoggaagtcaaatcaagttctgatctgggactgattgcttggagctc
tgattgcttaccttgtctgtcacctcttcagcagatcccttagctcggc
gacttggaaagactcctactacatggtttgtatcgcgcttgaccaagcctg
aaactacaagtaagcttcaagtgaattgatacattaccttgtgcatctc
caccagtaaaagataaccaccctggatagaggaagagtagcttccagaaaa
5001 gatgccacatctacctacgagacagataaggcaagtcaagacgataccac
actccggaacttagaagttcgtgattacgagatcattcgcccacaatat
ttcctaattgctatttgtatcaaatcattcacttgtactcactaaaagaga
gcttgaatctatgtacttgtgtaaaccttcacaattaatgagaactcct
ctatttctgtggagctagccaatatgggtgaaccacgtacatctggtgttt
gctttcctatctctatccatttatatacttatccacactaatgaccggag
taatctagcgaagatcacaaaagcgaccgttttcgctacctaggatctat
cttgcaagagaacggagaattagatggagatctcaaccatagaatacaag
ctggatggaatgaagtgaagaatgcacccgctgttgtgtgaccgtcgt
aggctcactgaagctcaaggaaaaattttataggacggcaataaaggccagc

Putative reverse transcriptase

5501 gatggtgatgacatagaatgttggcggtgaagcatcaacacgtacaca (ORF35) ↓
 aatgggtgtagcggagataaggatgcttcgtgggatgtgtaggcacacg
 agaaaggataagattgtgaatgaggatattcgaggtaaagtaggagtagc
 cgaagtgaaaggaagatgagagaaaatcggttacgggtggttggacatg
 tgcaaaagatgcctactgacgctccggttcgaagatgtaactacgaaaca
 gaggttcggggccgaagggtagaggaagacctaggaaaactttggaaga
 gaccctaagaaaagacttagagtacttggatctaacggaggacatgacac
 aaaaccgagc[gc]aatggcgttctaggattc[at]atagccgaccccacttag
 taggaaaaagctttggttgttgggttgttgggttgttgggttgttgggtt
 gtatctcgtttatTTTTTTTTTTTTTatgtatcatctcaaaagttctcca
 6001 attcatttcacaattggatgaacttgt[agt]caatccatcctttgcaaa[tt
 aggtacaccatttctatatgtataccaccatgaaattaaaaaaataaaaa
 agcacacaattacttaaatatataatataatataatataatgattctcaagt
 aaaaaataataaaacattatccaataattatataaatatataacata
 tatataatgattctatcaagtaaaaaataataaaacattatccaata
 atttatataaatatataacatataatataatataatataatgtatgtat
 attccaagttgttgaatgataaaaacccacgtagtgctggtttatt
 aaaggttaaaaaacaaataaatccctgtatttctgaccctccaccca
 cctctaacacgcgcccaataatttccacagctcatatcattatcacactt
 tgaccogtttttacagtattctaattagtagtgaaaaaataatcaaggg
 6501 tatattcaatataattgccacaagatcctctctagattatTTTggagag
 aatcttaggaatttgtaaatcgtgttcgtttatcgtacatagtgcggtca
 atTTTTgttagttattgttgtatttaattttaataaaaaatattcaaaa
 taatttctgactgtacgataatcgatgaatgaatataattcgcagatctt
 tgagatcctcaaaaaaggatctaacgaggatccgaatccatataattgt
 atgcagacaagtgacttagaataatccactagaaatttcaatataatata
 tggagttgacacaccacacttaattatTTTccaatcaagttgtgaggct
 gcttcgtgatttctttgactaaaattaatgtaagaaattggattgtcct
 attagcgtgggcaacatcataggaaacctaatTTTccaaactcccctcat
 gaattctctttctccgcatctccttctaactttgataatTTTTTTTTT
 7001 acaaacgatatttttacgttaaagagaggcggtagcctcaca
 ttgactagcgataatgtggttcaaatcgcTTTTggcgagaatcgaacc
 taagaccttctactataaagtgaagatgaataaccactagaccctagtact
 aagtgacaactttgataatTTTTtaagctcttaagaaatctcaatggtat
 aaggaaaattttattgaataaaaaaaagttaacacctaatgtatgagtgga
 atataaaccttacttctttgaaacaaaatgaaatatacatgaaaccgata
 agagacattaatctgtcccacttaaggcaacaagaaaaatacaataaaa
 acagaggagacataaaactggatccttcttgaagcaaaacaagatataca
 agaaaaataggttaggaataaaaaataaaagttataaagctttgaaacacc
 gaaagaaatagttaggaataaaaaataaaataaacacattaggaagcaag
 7501 tctaacatgtatgccaggatggacactaatTTTaaaaatcgaaaaccagac
 aagctaagaaaattccattgcaagaagtaagaactgccaacgatatgagt
 catgctacaccaaagatgaggaaaacaagatcttaaatctcatcatagg
 tgtgtgagttcataggcgtatttatggtatgatttgcaaaatggtctcca
 aggaaaactaaatcagatttttcatTTTcatcgtcactttaatttatgcta
 ttatgtgtagagatccaaatgtattcaattattatggattctcatctga
 caagcaaatctcatgtagctataaaaaatggagcttaacattggtgagtgca
 aaacaaaaacttgcaagaatcatcagatccataaaatttcatTTTatttat
 tattgcaagttaaaggaaactgtaagagtatgagtataattcttacaag
 ttttgggataattgatattataaagaagtctaaatcatgttgacaagagt
 8001 taatccaatcatattggtatattgaggactcttgtggttggcttgttata
 ggactcctactaatacaaggaatgcagcttgtattccttgtaggattgt
 tataggccaatccatgtctttagtataaaaaccacaagccccttctctca
 caaaaaacgctcttattgttccaattacttataaactggagatcattga
 gttttgtagaagaggagaaggttgtgtgtagattcgtccatggcttcttca
 tccatgaacatgtctgtgttaacttagaaacacagcaacttttcaaatc
 aatttcataacttaggccttacgggaaggctttgttatatacactacaac
 atacaaaactttacaatttaaaatcacaggtgacagcacgccaacatgca
 cagcaccatctgctgtcttctgacagcacacatggaaaagacttctgca
 acagctcgcagatattgttctcagatattactcagatattgcatatgtag
 8501 aacatatttgatgctgttgggtgtgatttagttggtaaatcttgtattct
 tggctaagtatattataaagactgttcttattcttacctgtgggtatcaga
 gccattgaatttaacaactaaattcgatccaaaatcgttctcctaaaatct
 gcaacatataaagtttgtctatttgttttctaagaacaaggaatacaac
 gtogttataacatacactttgggttaaaaaatttatgaccgttggctgtt
 gaaaaacaaagaaatagggttttgttttttcaaaaatttgaccggttac
 cagctcgaacgagaagagaactggcgattctcgatagggttggagattt

MC-20SSR96C1aFor

MC-20SSR96C1aRev

tctttccacgcctgcataattgaagatgatgtatcaagttactaaagta
 tatctctatttaaacatcacaattttgatgtttgaaatgtattgctttcgct
 9001 gcagttggttatgatggtttgtagattaagataattttggtttattaaccc
 tcttgttttattagtgtaatgaataaacttgattctgatatgctaattg
 tgttattaagaaccattgaataaagttcatcttgtttctcttgttttatt
 gaaatctgaatgagaattgtcttttctgggaattggcgccgacgggatt
 ctgaccgataaactttggagatgattttgggtcgccatgtctgtaatggta
 tgaggtccataggaatttggagtgatggcaatccgcatggatgggtctcg
 agcattgatggataatttgttctgttgatttggagccgaggaattgctat
 tgggttgacgaatatgggtctagtgatgcattcgggtaactcagttcatt
 gtttccatccatgatttattgttctcctatatgtgcatcactagcattt
 tgatttcttctgttactgtagcgtgcttgaaggacctcttgtagctac
 9501 ctttttgcctgcttctcaagcttggctgatttcttctgttgtg
 tgactttgtcatcatcatcttcaacattcaaggttctgctcgcatgt
 cttttgctttaaattaagcaatttttttcttgcgccccactcatatgc
 ttatggtgccttgttgtttgttgcagccatggatattttcaagtgctt
 taagagcaatccatcaccattcttgtgaaggaaggatcgcaagtta
 ggggaacgatgttgaagcttgtgagcagtgactaacccaaaagcgtt
 gttcttctatgaaatataaacttgatgcatatcaagtcgtggtcttctga
 agtattttgggaggtgacagatgttggctgactaaatcgaagaagaacg
 catgtacatcgaggattcttcttgaatccttgcaccatggtcttgat
 aactcaccgcttctgttgaagcttcttggatcacatccttagcagagt
 10001 tgtgacttggaaatagcacttctgactactggagaagccatcttgtt
 agttttattgggaatggcttgaagatgtcttgagtctatcaaagatgtg
 cttaactaacgtggccttctcatgcccgtatgcatcttgtttagtta
 tgatcgataatccttctgtccttctgcatcattgagcattggtgttcag
 cgaccaacacatttcacacatcacaaggggagatgtcgatttcacttag
 gatctccacaatataggagtcttgcgcatccagggtaaatttacgacga
 ggttgttcttaacgtgaagaaaatctcatcgcaatagtcgaggattgc
 ttgaggttaccgctatttgttttgggcatatcacaagcttctttagag
 actcagggtaaaccagatgtgaatatttctcatggatccgattttggt
 ttaggatgccatgaagtacaaaagcttctgaagaaaagtggtcgtaaaa
 10501 agaccactaagccagaggagactcagaccctgaggtgccattggtt
 gctaagtgttggccctgctgagttgaaagtgtttagaatcttggcgt
 agcattagagcacatggaagaatcttaccctagccgttcttagcttatt
 ggcttgttaagttgttttcccaaagccgatgtaactttatccgccct
 agaatctcaaaatggctagcaagatgactacaggtgaatcttccagcca
 tgctattcagctcttagccaatacttatgatggcttaagtgttgtttcga
 actctgcaagcactgaagatcgtgttgcggtacttccctaccactatgtg
 tatggttgggtggctgagtaacttggcattcatttttcttctgcaagtt
 agacaagtcggggcttctctcatcatctcaatcaagctggggccttga
 tgacaaggtactctggcgtgtttctacaaagagttttagatgatttgcag
 11001 gcacaagcgttgtttagaagctgtgacggtttgaggatggaccatcttgc
 gagagtggtcggcagggcagggcctcatagatgattcacaccttgtt
 tctcagacttgtcttatttcacaagcttctgctcaaggtatggttctt
 cggcaatgagactgatgtattgttcagtcgtatagtcctcatcgttcag
 caggcagtttggcttctgtgataggagcatattttagcacttaattagct
 tgttttcttgcattaatgttgttaggtcctagttattttaaattttaaag
 ctatttctgtgtttaggtctaaaggttaaagtagcaagaagtgca
 aatttgaagcatttggagcagtttgggcttagaatggatagcacatgc
 ttggagcaaggtggatggacgaatgtgaagatcaaagaaggctaaagaacg
 agcaaaagaataaaggaaaataaattcaagacaaaagaatacaagagccc
 11501 agcaataaagaaggaatgttagccaaataccttatttctgctaatcct
 ttctaatcctacaccacctaatttccaactgcaaggaggatctctaat
 attttaggacactaaataaatgattctagaagacctaattcattcctga
 tatgtgcccaccttcttacccttttcttgccttgtatgactttccc
 cttttctctagtcttctgctgcatatcttcttaccatttcttccaaaatc
 tgccacatattcttcttcccttttctcaaatgtgtgagcagccacatc
 tcccttcttcatggattctaacatttaattacccttttctttaggat
 ctggagttctaatcttctccaaaatttaattgtgatgtgattcccttct
 cttaaataaaaggttctgcccaccacttgagaattcttgaacacaa
 cacacaccattccaaattcacaacccatccgctcatcaccattcaccat
 12001 acaccaatcatccaattcacaacccattcaccatccttgtgcccag
 caaagaagaagaatgaagaccctggacgtgcttgcattcaagcttggat
 tggtagaacattcttaggtgtaatccatcttggtttctgattttaa
 taagttatcttgttttgcgaacatgaggagctaaactcgttttagctaa
 aggagaattcaagccatgaacatatttgcataataaattgattacttcc
 agttatgatttcataaatcatgaatctaatttacttaactggttgattca

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

gaccttattattggttcggtgattaaggatgcataacttagttgcatgcat
 gaatttgatgctagaatataagggagttcatctaattgtatgaactta
 tattcgtaagtagtggaggttgctagtcacaatcgtgtaagtaaatcc
 ttgtcaatgcttatgattttcacaagctaatgatctttgattgtatct
 12501 ctattatggttctcatgtagggactattgaagaataattttggtgctga
 tgtggtatccatccaattcaatgacttaaggaaaatctgagggttaatta
 gtgctgttcacgattaatctgggctgtgaggtttatggtttattggaaa
 agcaactggaatcgatttgtatgcaagtgtgtcatgttagagaatgac
 cctctagctagccaatcacccatattttcccccaatttcgtgcaaat
 gtttaagtctttaatctacttgccttactttaattttgtcaaaaacca
 atccccttactttgctggtcaaatagttagaatataatccaacttgt
 gtttttaagtgttttgagtcaggttaaaatcaattttcgtccaatcacc
 ctttattgtctagttggagtctatttgattattttgtgctattttgagtc
 tctttagtttgtttcgagttctttgagtcagtttaagtgtttcaaacct
 13001 agttttgtgttttgagtcagttttaagttagcaatccctcataat
 cccggcctagaacgatccatacttaccttactacaattgtcaataa
 gagggctcaatttgtgctgctagtttatcacatcacttctgctagcata
 tgtcaggcaagctaaaggaaaacgcacaatcagaatcctccaagccgtg
 tatatgcattaggagtcctgcaactcgtgcatgtacaaatgtttctatcac
 cctgttggccaagacaagttcacgagtaatccagtgcctgtgccttg
 taggctggtggtcaaagtatatacataaagacttgggggctgcaagtgac
 actaattcttctcattcacgtaatgatgcccaggtgcaagctcatttgca
 agatagacctatggctttagctcctcaacaccggtgcttctctcaac
 atttgagctcttaagaggtgggtgatgaagttgattcacacgaggt
 13501 aaacttgtttgtgacagcgtccacaagttttgaacaagcagctgtatga
 aggtgctctaaatgacagtgacattaatttccatcataagaagaagaaag
 tgtttaggcctccttagtttgatgaacgtgtgtcacataagaaaaata
 tgtatttcttctttcttattgatttcttctgctttcattggtttagcttt
 tgtttctaacaccttcttcttcttcttggctgggtcttcttcttctt
 ttgatttggcgaacgcgggcattcattcctttacgaagatattatttggg
 ggaatctactcctctagacgagagatttaaggattctcctggtgcaagct
 tgttccaaatccaccaagttgccctagtttgcctatataagtgtaggta
 tgcaatagcctatcatgctctaacacctttgccagctagctccgagatc
 tcaacttagaggccgaaccttagtggcactcagcagcttatccattgcat
 14001 taaccttccgcccggccatcaatatcaaaagccatattgagaagaagatt
 caaagtgccatttagaggaccttgcattgctcaaggaggacttggcgaag
 cttgtttctacgattgacaaccttaacgttgattcctcgctcttgagagt
 caagattgctgaacttatggtgcctcaattgagttatttctttgcatg
 ttgtttctttgaagaagttgaatccaaggttaagctcagcatcttgca
 ataataaacttatcaattgcccaagtttgggtcttctcagcaagctactc
 ggaaggttatcaagccatagagacttctttagcttccatccaagcgcgtc
 tataagtgttggcgcgagagcaagagcagttggaaatcgaagcatcacga
 cttcaaagtgttctcctcgagcaggaagctacctttctcataatcaaga
 tgagattttatgctagagcaagagaaaggtgcccagctggaagtgccca
 14501 ccttatatcctaccgatgtagagactttgaagacttttgaaggcttgcct
 gaagatcgtcacctgagtttagggacgtagctttcaagtgtgtcttcc
 atttttgtactttaagccttctttgttattgctctttttgttaataagg
 gctagtttggattgctgtgctctaaacaaaaaaaaaacgcttttgttg
 tgctgtgaaaaataacaatttgaataaaagcagtagagttttggtaaac
 tttttttgtaaaagtgttttagaaaagaaaaaaaagcaatgttaagt
 gtttggtaaatttttatgtaaaaccgttgcaactgtgtaaatgactaaa
 aaatttatgatgttaatttacataataaattttttatcatttctttat
 tatgacatttataagatttatataaagtattataaa[cagttgtgagctt]
 [cgaagaatg]tacttttaccttttttattaagatatttataattttttta
 15001 ttctattattttctctaagagtcggtttattctatacctttccttatttta
 tttttgtagaacctttgtatattctatcttccgcccctcactccctctct
 ccctttctctctctctctctctctctctacctaactcctaaggaag
 gctaggtcgtcggggaaggaaggctagatcgcat[caaccgattcaactcc]
 [gatt]tcaggacttaccgtcgatgttagcgggtgaaggcatcgaggcttcc
 cattgagattttgggtgtcaagcaaaaggtgcaaatctagcattcttta
 tgcaaccagattttgacaaagcccagaaatctgacgaagagccctcaaa
 tttcgacaaaaaccagaaatagatgaacctttcgccgctcttgtcctag
 aatccgaaaaagctcagagtcgaagatggcgaagagagacgtggcaag
 15501 gttgacaacgaagtcgacgattataatgtataaatttatttaattgctaaa
 ggccattttgggttttggaaaatttcattaaaagcctcagcgcctcttgc
 gaaatcggcttgcgcaaaagctaggcaagcaaaaaagctagttttcaa
 agatgggttttgcacttttgtgtttttcattttttttcaccocaaacctgt

MC-20SSR96C1For

MC-20SSR96C1Rev

gaaaaaaaaagctaagtgtgatgtttaccaaacataaaaaagctgggag
 cttttttttattcctacttttttcagaagcacctcagtaacaaaccaag
 ctaaggcctttgtagccaaattcttcctttaaagaaataaagtattcacat
 tttgaatttgccttttattcttcaaagtgtttttgtttttgttgatgatg
 tgactatacttgaagttttgaaatccaagttgcctacgtctccttggtt
 aagagggatcaagtcagcgtggttcaaagagtgatggattgtttttt
 16001 acagttttggtggagattttgatgatgattttgattttgattttgtcgt
 tcacagtttatgctttttgtaggcctggagcattgggtgttgcctttttag
 ctcgtgcaacaaggagcattgtatgtcaccttttaagctcatgcaacc
 aggagcgttgtgtgtcgccttttaagcttgtgcaataaagaagcgttgtg
 ccgtgtgcattttaggctcatgcaagtgaggagcattgtgccgtgtgcag
 tttaggctcatgcaggcgaaggagcattgttgcctatgtgcagtataggtc
 gtgcaaatgaggagcattgagttgggttagggtaataatcgcttataaa
 acctaccgttgatgggggccaatttttaagcgtcttcggccatgattaa
 gtaggcacgttgggtgtagacctattgtatcacgtaaggttcatcctact
 ttgatgtgaacttgcattttgttttgtgagttgtgatgataggctttcgt
 16501 aatgccaaagacaagatctccagtttgaaagaccttgggccaactttttg
 ttgaatgtctttggatagccgtgcttggtaacagtcacaagtgttgttgagc
 ttcgagccttcttctgttgaagtcttcaactcttgaagtcgtggccttg
 tttttctcttccagtcacagcctttttatataatcatctttagtgaagg
 atttgactttcgagcggcataacagcttccacgccatagcaagagaata
 aggcataagcctgggttaggagtcagtggtgtcgtcctatagccaaagtc
 cttcgcctattcttttgtgccagcctttcttgttcggccaattaccttc
 tccaagaggttgcacaacgttttggttgaatgcttctacaagacggttggc
 tggagctgacacataaaaagacttgtgctgcttaaacctgtatttcttgc
 agagctcgtccatgagtcggttggagaactgtttccattgtcagtgatg
 17001 atgtagtgaggcacaccatagcagtagatgatagctctttgatgaaacg
 aacaacagtttcttcttaacttcttgggggtacggctttagccact
 tggagaagtagcatgttgcagctagtgcaacttacgaacataggttttggc
 tcctcgtttgtgcccaagttgagctctttgaggtcattgactacgacttg
 cccccctctcgagttgtggtggtgcagcagtgatgtctcctcgaggat
 tttgtctctttgtcttcttggattgtgataggaagacgtattggacct
 cctcttcagtggttgcctcttgggcatgttggcgtgaagattagccagtg
 tagataatggtgocctctttaccttcagttatctttttatgccaaacttc
 caaggttgcctggcctttcatcctagaaggaattgggcttcgaacgtcgt
 ccttctgatagacgtcaagcttttcttcttggcgttgtcttgttcttc
 17501 ctagaagcctatttgggtttttcatgtctttccatagccgacccgtcatgg
 aggagagctagctgtgttgcctttgttttttgggttttaacaaccttttga
 aaacagagccttaggggtgttggcgtgttaagccttttgaagatggagatt
 ggtcttgaccaccaatgagatcaagtgccgagattctgggttttgaacaa
 ttcagcctatcgaagactgacgttcgaggggoggggttacgctcatcttg
 attttgttcaacgcttgaatcttcacaggtgtattcgggtgaaagccaa
 gtccagccttgttgttgtcaactccgtaaacatgctcattcaacttctt
 tttagtctcgatgaggtcatgttctttgttgttattcttcttccaagat
 tttgaagagaggtgaaatcgtaccagcctttgacatgactttgtaggtg
 tttgggtcgaaacattcttctgatcattttggttcgaagagagccttggttc
 18001 cgtcccctttagcaggggttttaagaaaccttgggtggttttgaaactt
 tagtgtcgcctagctgtgttagaggtaaaattgcatttgccttatgcagc
 tttacatcgctttgtgcagttgtgtttagcttttcttattccagtttc
 gaacggggattggccattcttcttcttaataactgggatgtattgaaaga
 cgagtgtgttggctccttttgaaggcgtcacactctcttggttggttaagg
 gctttacaggctcatcagctctttgttgaataatggtacagtttccctt
 cttgtttcttgggcacggcttggcactcttatttcttaggtgaggctttg
 cctgtggatttgatttcttttggaaaagcttctgtgcaccatgtctttatc
 catgtagaacttggcgtccgtgaagtgtgatttggattcgggtgaaaggct
 tagtgtcgccttgcattcttcttcttactttaaagcattgatgaa
 18501 gaggtagcggcagggcaccacttcatctcgtagatccaaggccttctta
 atagcaagctgtaggaagtcttgcacatcaatcacgtggaatagcgtgctt
 tatttaagttcgcaatggtcatctccactcgaatcatgcccatcgctct
 ttatcttcttgggttaaaaccttggattatgaggcagctttgagatagtt
 catccaccttgatgttgattatggctattgttgactttgcatgatgttt
 atggcgcctcattatccacaagcatgcggttgaactttgtgctcccttac
 gtacccaagacaaagaaagaacagttgtaaggctttgatcccagcaaca
 agtctttatcagtaagttgattgcatcatgggcagcacacaacagctggca
 tattcatgtggccgaagattcaagccttcttcttgggttgcgtcgggact
 caccaagacggttactaatgcttttgcacatctcattgggaaattgcagca
 19001 cctcctcgatgctaaagtgtgtcggcaggccttctttaggtgtgaggatc
 tcccttctcgtaggcaatagctttaccttttaagggtccttctatctc

Hypothetical protein (ORF38) ↑

tatttcaaccatgtgacaagcaacagtaatgcattggttgaagaagtc
tcgagaagtaactcgtgtaaggaaatgggaatgtgtggtcttggttccata
ggttcccagcgtacgttggttaacatctctgacgtctttttgggctt
cctattgttgcgcagcagatgctttctccctgttccacatttgccttg
tggttgggtccttagcttcttatattttgtaggtcaccaatgtccat
cattcttcatcgtcagtaagtgcattttggtcacttgccccagcagtg
ttatgcaaaaacgtgcagtacggcttgagcattgacgggagtggtcatgcc
19501 ttatttagagaggtacgagatcgaatgatccaaacatgatcgtagtagtg
tacattgcggtcgtgtcttccaggcaggttgatttggccttgggtgc
cagctttatgatgagctccttgagaacgaagcacttgccaataggatgac
tcacgatatgatggtatttgcagtacctaagatagttaacgcgattcatc
tcttcaggcgtcttgcattcggatagctcaatcaccttctttccaacaa
gtcgtctaacattgcagccacgtcaaagtccaggaaagggtacgtcttct
gctccagctccctcaaggtgtctttgtacctgtcttgggtgcgagaatgc
tcccctctctttatctcattcatcttgttcttagaggatatttgacggg
cgtagatgaggtcttgatagggtagtattgacggtaaaagcttcatgg
cgagcttctcccagtttgtctaccttgaggcaaacacttatccttt
20001 ttaagccgggtgatcagctccttctcccatgattggcaatactcagctc
catgtcatgggagcaagttgccatctcctcaaagtctcgattttatgc
cttggaggatgtaatgtaacccccagtgcatgccttaaacacacatctca
atggtagaggtttcaaaaagccgatcttgcacccaatataatgaacg
ccactattgatgtagtcaatgaaacaggttcatcctccactgtttat
actcatcaactccagcatgttacgggtgtggcaggtgtgtagaatggat
tgaggaattccattcaagctggtctcagctattgatagactcgagctca
aggcgggtgaccagtaaaagcattaccttccagcaagtgcataattgc
ttgaagaggtagtctccctacgtcccagtggttgcaggttcaatgaa
atgggcaatatgtgttgggattgccttcccatcgaattgcataatt
20501 ttagcggctgataacctcttagcatccttaagggtgtaaatcttcttagag
tacggcttgaatacagcacgaagtcatgtgagcttcttctgactacgc
attgattgtgctcatgatcatctttgcaacttggatagagagaaatcc
catgagcatcgtagcttgggtcacctctagcttatcctcagcttctctca
ccagagcttctcttctcgtcagtttctcatttggcagatcaacctt
gggtcagcttctcatcgtgctgcttcttggacagttgacaagtgtaac
aatctgcaagtcattttcttccactgtcttgtcagcttgcattgctt
cattcatttgggccaactgctccttaatggacgtcgaccaaatagtcata
acttgcagtggtacaggatacaagctgatgcttgagtcagcatcagaagt
caacaactcgaataacctcatcgggcttcttcccttggcgtcttagcg
agggagggtgatcacgggctcgtgactcaggtgctcttgccttggc
21001 aaagttgatgtaggggtggaaggcacggcagagagagctcttgccttgtt
tcgattgtaaccccaagtgcctccactacggtaatgatgttcttgtt
cttcacattgattgctgggaaacagcttgatccttcttgatgccatgtgtg
tttcttgtggatttaaaaacatagatgataagcaaaagaaatcccaatagg
cgtgccaaatttgtaaacacgaaaattctgtaatgaatgaaacgagaaca
cgtgatacaaaagtagttgtattgatgaattgtgaaggtacaatctctg
tttacaatttctcctctgatacagcttcaaatgtgatgtagatgcataga
ttgtaattcaagggcagagatgacttgatctcaaacgaaacgataaaaaca
attgcttcaaatttgagcttgaacgatgcgtagatagcttccacctta
ggtttagggcagcggcaaaagggtggtgtgatgtaggatttctgtgattca
21501 aggtattggagacttgatcttgaacaaatgttgttacaagtttgagc
tttcaacaaagtcttgaaggaatcagcacaggtgcttgttattcttcaa
gggaatgcttgggttggcgaagaaagcttggccttgggtgtgcat
tcttctaagagaagtttggcagcatattagtcttcaaaagatttggcagag
gttttcttcttgtgagaattttagcccttcaatgtagaaattgtcgacct
tatgcttcttgggaccttacttatagacttccaaagtcatgcctt
gggtggatttggcttggatgtgtcttgcctcctcgggagaattta
ggcatgtttgtatcttaacttccggccatttcccttgccttggccgaatt
atagggcttcttgcctatttgtgagcaatcttcaattcttgccttggtt
tatgaagcagatttagggcctgttggatctaaccttgagaacgaaactc
22001 aaatttttatttttttctaaaaacacgttttgatttaaaaatttaaaat
tcaaaacctgttggatgcattttctcaaaactaatccaaacagaaaa
agtcaaacgccaacccaagcaattacctgacctcgcctccctctcatc
cccagcggcaaaagaccaatttccaaccaccgcatccctcactcccta
gtggtgactccttggctatttaccactccttactccaaactcgcgctg
acaaaggctaatcatgttggcggctcaaacctgagatccaatgaccata
gggaagatgtgggtatgggtgagcaagtgcagagagataagattga
agaggcaataaagaagagagactgaagagggcagagagagattgattgtc
ttggccgaagggagaagatagaggtggtggctatgggtgataaagtga
gaagggttggaaaggtaacagagaaaagaggattgaaaagggaatagg

22501 aaggagtagatcaaatgggatagaggaaagaaaggagaaagacggg
agtgataatccataaaataaaagggtctaaatacacactttttgtgtttc
atccaatagcccttgttttcaaataatttttcgttatggttttgagaatt
gtattcgaatccactaccaaaacaagattttaaatttgtgactcaaaact
tgtttttgagttaaaaacattggatccaaatacaaatcaaacaggccct
tagctttctttccggttttgggagctatttgggcccttgcgattttaa
gaagatttcttcttttgcacacttttgggaaggttttgacttctttg
cttgatttggccatagtctttgatgacttatccatttggataggtca
tatgtcacgttgagctttgtgatgcacctttgtatacaatgaaatttac
atgtctacaaatgccccacttcaaacattaaaataacaaaaacaatt
23001 aggcgataaaaaataaaataaaataaaaccaaacaattagccgataaa
aagaaaaataaaaaataacaatagctaataattatgaaagaacggacaat
aaaagacaattaaaaacaaaaacaaaaacaaaaaggaaagaaataaa
acaaaacagaaataatacaaaaagaaggaaatgaatgcacctggggagg
attcgaagtgcagcatggtgcaagaaacagaggccccggaaaaacaatg
aaaaagaaagcactgaagtgtagcgcgcttgtatcgaatccgtaacctgc
tggaaacaacataaagtgtctgagcctctcacacggggccacacatgctgat
gggaaccctaaggccatcttcaaccgaagggtccagaggccagatgatc
gaaaaatagtacgaaaacatctccaactgagggccaggccagaggctgg
ccacttccagccagccagccagccaacccccgggctagccagaatttttt
23501 aatgtgtcggttataatcgacactaatacatttatagtaataaatatga
atgtattcctgtcggttaaaactgataggaatgctttgaaatttttttt
ttacagttttattatttttttttttttttttttttttttttttttttt
attttttaataattttttttttttttttttttttttttttttttttttt
attt
tttcatt
ttatacaacattaaattaatataatataatataatataatataatata
taataataaccaacataaaaaattctacaacataaaaaaaaaacatttaaca
cataaaacttaacaacatttaagttgtagcttttaataataaattatg
tttggccctatggccctttggccctcggttggagacgggggcaaatatgg
24001 tcatgtactgttcattaaaaatattaatcttagaggatcttggaggacc
agagggctaaaaacgagccctctggctagccatcggttggagatggccta
cgagcttcaatataaatggagccgcctttcttatttctcactgctgcc
attcagatgtgctcattgcccgaataaaaaataaattcaaacgtaa
ataaaaagttagtacaacataatataactaagaatataataaacacaaa
ttatttaaaaaaatccccgcaacgggtgccaattttcttgatataaattt
tactacctacaaaagtaaggataaaaaatacttaaaaaacttgaagagt
acaaggttgtcgtagtatagcagcagcttaagcaaggctgttttctaca
ggattgaatgactaattttatgtaatacacaatcttaatttaattttga
aaacaaagattggaaaatgttgattttgaatttaaaaaataataaaaaacga
24501 atttaattaaaaagggttaagaaattagcaacgtaagaaataaagaaa
aacgttttgaaaaccaaattgtaaaagcctagggttccaccatcccttta
acaatcatatgcaattttcaattacttatgaatttctatatacatgct
ttgaaggttaggttttcttaataacatatttcccttggatggtcaagcga
aaacgtatctcaacatgcaaccatattgtgatggtcagattaaatata
acatgcaagactcattaagctttgtgaaaaccccttgaacacatgagac
ccttaagagcgtgatggtcgccttaagtgaacttacaattactaatcaca
agaagccctctcaatttcaggttgatgtccaacagaaattgcatcaaa
ttacctgtcaaaaaccccaatggtgcgcaatcactaagacaattagata
gtttaatcacagtgattaaagaattcaagcaagcatgcatccgttcataa
25001 gcaaatataaaatcacatattcatgctaaggcgaggcttaccctagc
aaacggaaattagttacaacaactataaaacaaaggagttttattgaaat
aaaaaaaaaaagatagaaaaaccccttgaatacaaatcgtcaaagcctag
ctaagttctctaaattgatgctcttcccctcaatggctaatagcctt
atttatactactacaaaataaaaccctaaggggggtgtttgtttgaccgga
ttaagggggactagactagactagactgtagtccttgtttggctctggac
aaggatgaggtttaatgggactaaaggggactcgtcccactgaacgttt
cgtaaatgttcttaaggagaccctcaatttcaaggggactgctagtct
cgttctctcccctgcccgtgatccctcgtcgtctcctttctccttgcct
tcttcttttgattcgcattctcaccactcccttgggccttcaactatcc
25501 cacctaaaagcctaaccagacctcaacaaaaactgaaaaacaattca
ctgattaccttcccttctctctogtttaattcactctcttctcagcaaa
atcatcatggttgcaaccaccagaccacatcagacgaccatcaccagaa
gctcgccatcgactcgcattcgtgcgccagaccagatgcaagtcaggac
cgaaagcgtaccaaagtcgaagaccagttgggggttttctcgggtgtg
gggaagaagctgagaaacttggaaatttagtagaaagttgtgaaatttg
gtatttgaattttgtgatgtgagaaacttgtaaattttgggccttatg
ttttgttatcaagagaagcggaggtggggtttgacgggaaaaagaggtag

agggagaaggaagagaggtttgacgggatggagaggtggggtgaaaaatg
 gttacagagattgtataagtaaaaagaagaggattctataaaaataaaaaa
 26001 aagggaaaaagggaaaaagataaaagtaaatagtaaaatattaatataa
 aataataaatattgcagtcctgcttcttagtccgacgctgcaccaaacc
 cttcactaaatcaatccagcttagtctaatacagccagtcacacatagt
 ccctgaagctagtccagtcagagatagtcgattcaacaaacgcactcta
 aaaggttttagaataaaaactaggaaacataataaaaataaaaacagaaa
 ataaaaggtttcctatgtgactaaaattcggacttctcagaaaattagt
 cttttgaccgaccaaatacaactccgatttggccaaaaatgtcgctttg
 aaagcttgagacatccccatacaaatcctcagaggaaatcgtccttaaaat
 atgcaactctgacctcaaaaaactcaaaacgctcagaaaacgctcaatctg
 ggaaagctgcgcgctgggtctttatctttcgccacggtcaaacggctg
 26501 gtgaaaaaatctgaaactttgacataatcatgttaaaagaggcatgaaca
 tccttcaattgaaatcacttcaaaattcatccgcttgatcactttttgct
 ccagagaaaagtcacatgtcctacattgaaaagatataacgaaagtttcaa
 attgtacaaaaataaccaataaaactataactaagaatagggtaaaatata
 tggatataatggactcatatgtagcctaatttttaagcatcaagcaca
 cttgccacaagcagccaactgcgccaaggaagcattgccatccttgaa
 ctacggttgaaggtgaggtcaatgcagaagtcattaacttattgttta
 ggtcctatgtcgttgtttcatagtatataaaactgaattatagagattt
 tgaacccaaaaagtacactttgtagaatagaagacgataatttgggtgtt
 aaagaaaaaaaaaattgaaaggagggtcagtcacatgctattgactcct
 27001 ccgataaaaatcagtagtatagtagccaaatctcctcactccaatttt
 tttttttttgggtgttatttttttctctagctacattattttaatttaa
 ttaagtttatttgggtttttgttcattgttgagtttaagtgtatgtttga
 ttatgattaataagggaaacaacaacacaagaatacaagatatatagggt
 aggagtaatacaacacactgatcacctttgaatcattcctacacggcagct
 ccataaaaataaggagcagaaagtaagtcagggaagaaagctaaaaagct
 taaccaccctataagtatgggaaagaagatcatgatataactagaagta
 aggtgacaactacaccctaataatcaacctaaccaatagctccaaatacc
 cactcctgactttcattactccccctcaagctgggttcagaatgttgatg
 gaatcgagcttggatagaagacacttaaaactgatttggaggtaatgggtt
 27501 ggtgaaaatgtcagctaaactgatcaacactccttgtgtagagagtttcaa
 tcacttgagcttt**tcacctgagcacagacatagtgccatcaacctcaat**
gttttctcgttcgtggaacactggattagatgctatgtgcattgcaact
tgattatcacagtagattgacataggttgccttgtttggaacctcaaatc
actcaaaagacccttaagccatattaattcacatggttggatgccatca
atctatactcagcttctgcactagattttgcaatgacattttgcttcta
ctcttccatgttactagatttctcccacaaacgtgtaataaacttgggt
agactttcgatttatttgggttctgcccagtcagcatcacaataaccaa
taacttgagtgttatcatttttattcataagaattcctctccaatcgat
cctttgagataccaaaagtattctcttaacaagattgaagtgttctatggt
 28001 aggtgcatgcatgaattgactagctaagctcattgcaaatgtaatatcta
 gtcttgtgatagtaataaataaacttccaacaagccggtgatagcaa
 ccaatattagggagaggttaccctccaagctaagccttgagcttgcctgc
 aaaaggagtttgaactggcttggcatccttcatgtttgattccgtgagat
 gatcgaggatataattcctctggttaagaaataacccttacttgatgag
 gtcatcttattcccagaaaatattttaaccttccaagatctctgatgga
 gaacttctgtttagggagaccttgaggttgtgatttcatcagcattgt
 ccctgtgattataattcatcaacatagatcaacactacaagttttcaa
 ccactcatgttccgaacaacatggatgagttttcatgacttctttgaaa
 ccagtttttcaagaaccgaattgagtttggcataccaggctccttgggga
 28501 ttgttttaatccataaacggatttgtgcaacctacccccatgccaggct
 cattgtattgatgatgtcctgggtgtagcttcatgtagacatctccaagg
 tcaacatgtaagaatgcatcttcacatccatttggatagagaccatca
 tttattcacaacaacagacaataaacacccttacagtggtcatcttggcaa
 ctagagtgaatgcttcttataagtcacacagcatatggttgagtgaaatcct
 ctagccaccaagctagcttgtatcttcaatcgatccatctgagtgaa
 tttgatcttgtatatacacttactaccaactgcattttttttttttgtga
 aagtctaactacactccagttgttgtgttgggtgagagcagagctctt
 cctccatggcatttctccacacttcttagcagcttgcctcttgaaaggtt
 tgaggctcacagtgattggtgatggtgttcaaataggcagcataagattg
 29001 tgaactttttggcggacatgaaaccactaattagatgtgaacatagca
 tgtgcacctaaagactttaagggtgtgacagattaacatcccttctttta
 acatgccattagagatttaaaactccaagacttgactaggttaactctgtt
 atgatatatgcagctatcatgacccttgggacaaaatctttttggtac
 tttcatgtgcaacattagtctcttgtttttcaagtagatcacgggtct
 ttatctcggctatgtcgttttgttggagtaccaacacaactattttga

Hypothetical protein with a
 domain for Reverse
 transcriptase (RNA-
 dependent DNA polymerase)
 (ORF40) ↑

tgcaaaattccattattgcttgaatattgtttcatgatgtgtgacatata
ttccgagccattattagatctcaaagttggatttgggaattgaaatgatt
tttgacaagattatgaaagtctatgaatgctcaacaacttcaacttaga
29501 tttcaaaaggtatagccaagtaactccaagaatcatctatgaaagtta
caaaacttataccatcaaagaatcaaaaatggccccatacatct
gaatgtacaagttcaaaaactgtttagtctatttgaggatgaaccaa
aggtaacctattagatttacaaaagtataagtaacacacaggatcactt
atttacaagggttggaaacaaaacttgacaaaatgggctcggatggatga
gctaactcttggatgccacaggttttggcttctgacttaagttatacttgac
ttgaaatcccttgagaaataaagtctcattggataggtaatagagaccat
ctaagaaaaaacctcaccaatcttttcttggatgacagctcctgaaa
atgacagtgtaaggggagaaagtaattgagcagtttaaggcatgtgtgat
ttttccaactgagagaatttgaaaaagaatgaaggaacgtacaaggctt
tagactttatattgtcaaacactagatgtaatttttcttccctaacatt
30001 ttagctccctccccattagcaactaaaacatgtgaagatgtaggaagttt
ttaaactcatggagcttgtaaaattggtttgtcataagctctgtggcacc
aaaactatgcccacaaatcatgcaattgatttatatttagagtagttt
tgaaagttgtcaatataccttgcattgttctcctgagtgagcttttctgag
tcggccagaaaacctgcaaaccttccagcaagcaattgagttccatt
tccaaaggtagcagcttcatcactcccagagacttcttctcatttgaa
ggatgacagcaaacctcattatgagtggttaggattagaggtaaaattc
aaaagatcatctgagcctcaattgataggataaaggcatgattagctt
gtaactggagaactgaggcttctttgaataaaccttcttcttcatga
agtctagtttaagctcagggtgaagaatataccaatgatccctcacatgt
30501 ccgtaggtcaccaatgttgcatttgagatgaggatttcttcccttga
tttcttattcattggagaagtatttcttactggaaaagtaagccctag
cttccgaggcaatcgtcgtggtaccagcgttcataacatttttctcatt
tcttcccttgaattgtgcacaaacattggcaaacgtgggaagctctgg
tttcatgagaatgtggctacgtaaatcctcatattagaccccagactag
agagaagctggaaaatttatcttcccaacccttttaagaagcaggttg
agtcggtgtagaggacaatatacctccaactcattcaacataccttt
agacttccaagaagctgaacaaaaggttttccatcttgttgcagctcgc
gatatccattttaaagttgaaacacacaggttgcattattctaatttccat
acatctcttgcagtgacttgcataactagaaagatgactcagaatagctg
31001 aatatctcagctatcttctcctccatggagttgaggagctatgacatgac
caactgatcttgttgagccaagattcataagtgagagaggaagcatcag
aagcttgcatacttccatttatgaaccctagctttgatcttcttgtata
gcaagtgaaaactgccccttaaccatggtaagtagttaaactcatttgaag
cacagaacttaacccttggatttggattacatcgatcttggttgaattg
gtgaccggctccaagtgagctatcagcctctgactcgaagatagggttt
cgtcagccatggaagaatacttgaagaacagagcaacaagatcacagag
acaggttattcttactctgataccatgttgagttttaaagttgttgtt
attatgattcataaggaaacaacaacacagatacaagatatatagggc
aaggatgattaacacattgatcacccttgaatcactcctacacgacaac
31501 ttcataaaaacaaggagcaaaaagtaaggaaggcaaaaagttaaaggct
ataaccacctataagtatgggaaaaagatacacaatataactagaagt
aagatgacaactacaccctaataatcaacctaaccaatagctccaaatac
ccactcttgactttcaatattcatttctccttttaattacctaatgtga
tatgatgatcataagaaggtatagttcgtgaggtatgtgatctttaa
tgttcgtttctggattttgcttaataatttgaaatatttcttgaaa
atgattgttaaaaaaagtgactttagtgtagcctatttaactaaagcat
tataagtaagaaatttagtggtatgattagtggtattcctcttaactgt
aaatgaaaggcatcttggttcgattctcagcaaggcgaagttgaacca
cattataggaagccattgtaggcttagccactccacactcccttagt
32001 gtacatactatcgtttgttcaaaaaaaatctaagaattatagaaagag
gtacggcaacacagagaggtgaattatgaggtgtgtattatctcacc
ttgtgcctttgtttataatagtagggaaggtaaattccctacccttagg
tattacattacaactcaggaagctataaaaaaatggtagaatcccataaa
gatttactaggatttacaacacacattcaatttctaactaaattaaaat
cgcaacaccctcccctgagtggtgtaaaaactcaagtaaatggtgcatcat
gtcttcatcgatgaagtatagttgatgaaattgtcggcacatgatcgaat
gcgagtctcaaatcaacgaaagaatgcatatagtaaaactcacaagac
ctcactatggtaaaacctaggatgggaggataaaaaagagaatttgag
aagtcaaaaatatacgtcttcagaacacaagtagaacgcacacaagggt
32501 tgactaacctaggatggatgacctaaaaccttgcaggtaacaaaaacca
gtggggaaaatgctcctaagtgtaggaaaaaaagtaacattagggttaag
tgagtatacttctagatactccctgagttttacaaaacttccaaatgag
aattacaatcattgcaaaataaatagttacgcataactaatcctcgaataa

gttttgttgaaggctgatgttagtagtgactttatgaagaggtcgacaag
 gttgtatggggcagattgcttgacttcaatctcttgatgctttgctgat
 gtgatgtagacgtaatgtgcttagcattgactttgttgatgtatcgtatc
 ttggcttagtagatcatgcagcatagtcttcatggactttgttgggtgga
 tgaaaaactgcaagttttcgaatatgctcaacaagagctcctaaccatgt
 ctttcatgtgtggcatagtgaggtgcggtgggccttgaacgatccgaaa
 33001 atttcgcaactaaggctctatattttcgcccttaagatattgcggtatccc
 taaaggtaaagacataaccatttaggaagatgtcttgtgcaggcttgata
 gttaacctgcggttagcataacaacaagttaaacacatcccagggaaccaag
 ggggttgaatctgcttgagatcggtgggctaaagctcaaatccatagtac
 cttcagtgtaatggagaacaacatctttatgccaacttaatggctatagtg
 aggccagtgcttcatctcaccaatagatcaacaatgaagaagatgtca
 ttttaatgcaatgaattaagtacaacaagtagaaaattaaacttatat
 atgcaactcaggttacataacctcttcaagggtctcatttgcataatagc
 gtatgaatgatccaaggtatactcaaaaggtaacacctccgggtgtagt
 tcgactggtggactagaataacctcggacaatgcttgaactgcaggctca
 33501 agacataacaggttttcccaagatccttcatctcaaatttcgtcacatgc
 ttggtagttttggaagctcttccagagtcatagtgagatcagtgcatcg
 atataaactgcaactctagcaattcggaaatagacttcttaattgaacag
 acaagggcatttaatcatalccctgactgatcaataatcacttagatgg
 gtataccacatctacccaaaattgcttaatctaagtgaaacgctcatttg
 aattgagaggggtgttctatggtctcactactatttgaattagtcocatgtaa
 ttctttggcaacttacatgtaaatctctgtatcaatatcccaatgaaaa
 aaattaatcattttgatatagctgtaatcaagaagccttgcgcataatg
 cgtgcattacatgctcattgttccattcatttccattacgctcatttccca
 cttgtaacacaacaaggctacctcggacagtgtaggaacaataggtccaa
 34001 atatacttcagtaaggaattaaacgtagattgtaatttcagttgttcaat
 tagttttaaattatttttatataataatgaggcattttttcttaattac
 taacatttgaaaaggggttaaaaaataatcatattttcattcaaatctc
 tgttgtttaaatttctacagtcctacgcttccgcaacgctgtgctcatt
 aagctggttggtcccaatgtaacttcaagaaataatctttaatcaaat
 aaactttggagttttctatccaaactgttgaactccagaactttgctgt
 tctactcttgaatccttttagcctgcccactatttatgaataaaaa
 tcaaaacaagttgcaggaaagctatgaagcaattgatcatgaattttcct
 acctccaactactaaataacgaaagtagctctctctgcacattgtaa
 cattagtaattggatgtaatatatttcttcaacagaatcgaagaagcaaaa
 34501 gaaaaagaaaaaaaagaaccttaacatatttgaatataacagttgata
 attctaagcattaaagaaaaataatcaagggtatattcgatataaattgta
 ctggcaaccagaattgactcactatgaatgcggaacaagtgattagaat
 attccactagaaatttcatgatataatggacttagcacaccacgcttga
 ttttttatttttttagatcaagttgtagaggcggcttgccttctct
 ggccaaaattaatgtaagatattggtatgtcctattagcatgggcaacat
 gtcacagcccgctccgagattttcttattaaggacgtgaaatgtcggatt
 tgccttagcgggggctatggtacgtgtgtgtgacattatttggattaat
 tctatataagttttggatttaatttttagttttataaaaatttgggttagg
 gaattagatgatggagggtgaggattggatgttggaccatttaaaccccc
 35001 tgcctctctcgcccttcccgtagcactctctctctcatccctcacggctt
 ctctctctctcactctcgaagtgtacggacgagaccctcaaacatccaaa
 actcagtggtcgaagcaacaaggtgacacctatcccaggacgagcgc
 atccagggacgtcagcaggggtacgacctatcagtgagtgagg
 caattttgttttctgattacctataactactgttttccagaaaatgcg
 tttatatagaaattatatttctaaaatgcatgcatgcatattatgagtta
 tgaattgataattgatgcatatataatgtgtgaattggtgccgtggacgca
 caggtgagatcaggtgagtttatatggtttatataatgaatgaatgatgatg
 tgattgcgttgtgagctcataaactgcaccttgggtattagtgcttatag
 tattcaccacaccgacgctcgccttggatccaagtaggtggatgtcgt
 35501 cagaccactagaggtggttccgacatgccagtcgtacagaccatagaagg
 gttccgactggtgggtgactttagattatgtgcacagatgattgatgaaa
 gaagcactagagcgtattatttaccatcttagtcatacagactactata
 ggtagttccgacttatgtgcagtgtaatgccgtacatgtcacatttgggtg
 actccggcacggcgtacaggtcacagttggtgactccggcttgatggga
 tattgagctatagaatcagccgtacaggaccactacagggctccagttg
 atttatttttccacttagttatattgatgcattcatattatattttggc
 atggcatggcatattctttctgagatggtatggtgatagatggtaagctg
 agttttgatgatataatataatataatataatggttatataactattttctg
 ggaaagtatacaggttttacgggtgaggggttagaaacgttttaataaat
 36001 gtttttttggaaaatctttggccgcctcgtaaaggaacatagggaatccc
 ccgacttctgaacctaatttccctgatattactcagctaggggaagcaa

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

tggcccaggctttacagaatgcaattcgtcctcctccccctcctcaaatg
acaccctggagaccatgtacaacttgaattagatcattttatgagtga
taagggcagcagggggcagagaaatggctagatcatattgaaaagcgt
ttcaggtgatgaaagtcaaggggaacttctgctgataggtgggttgag
accactacctggtttttggctgtgaactagcagcttgggtgggaaatca
gactaggtacatgtcacctgagcggctgctgactggaaagtattcaaag
agaactttatgaagagatttgcctccagagtatattgatcgcaagaaa
cagaaatttactcgattgaaacagaggaatgtcagcacatgagtacta
36501 caaaaagtactgatttgcacggtatgacctgatacagctggaatc
aggtagagacgcttcgacgcttcaagttgggaactaagagaaagtggcgg
acttttgcagtgacttccctgtgccgattatcatgagtttttcgagat
tttggtaggtgaggactccgataatgttcccagtgagagtgaggatg
acgaggacaagaatgatggtcagaagaaggatgataaaggtaaaggtatc
tttattctgggacctcgcaagactcagaactttaagaagagtggggcgag
ttcgagttcttctagtggatagatattactggcccaggagaggtg
gtggaagattttctgggtggaaccagatttcagaggcagagggattctggt
agagctggtggttctggcctccgtggtgccgtcgttgttattttcgtca
tcatggtgagtgaggaaggtggtggtgcttgttatacttgggacaaa
37001 cgggacatcgggcttctcactgtccccagggtcagcagagacccagcag
accactatgccacctccaacgcccattcagcagagcttggagcaggcaa
ttatggccagtcgggtcgtggtggtgcttaccactaccagagagatgctg
ctccttatgcttctggaccttatcagtatccctaggagccttattctcag
acggggattcccaggactttgggggttattcttcttattcttccatgcc
agctggtggtatcagtggttagggaggtcagccccgttaggggaaag
ttgctgctggtggtgaggtatcatctaggcagcctagtcaagtggccat
ggacgtactcccagggcgaggtaatcaaggcatcagaggtcgtggtgg
acgacagcaagctcaggggctgttatacacatctcgcttcaagatgctc
agaaccatcccgatattgatcatgggtacgttaaattgttcttgggtcattt
37501 gctagagttttgattgattgtggtgctatgcattctgtgatttctcatac
gtttgcctcaaatgactcaacctcaccctcgctccttaggatttgatttag
agtttgcctgctataggagataagtggtatggtgatagtggtatctt
aggtgccaagtgatggttagaggcagtagttatgctagctgatcttattcc
attagatattgtggtttttgatgtgattctcatagcagattggttgcatt
ataatcgtgccatattgatgttatgggaaatcagttactttttatcgt
cctggacaacccgaggttacttttgggtgaaagaagtggggtagaca
tggtgttatttctgccataagagcaaggaagttattatcgaaggggtgcc
aaggaatttagcacatgtggtgttaaattgatgttgttcttagcagtgta
gaagaagttggtgtggtcaggcattatcctgatgtatttctgatgatt
38001 gccgggggttggcgcagacagagatgtggaattctctattgatttgcctc
caggtactgatcctatctctaactcctatagaatgtctcatgctgag
ttgagagaaataaaaattcagttacaagaattacttgataaaggtttcat
tcaacctagttcgtcaccttggggagctccagttattttgtgaggaaga
aagatggaactctaagattgtgcatgattataggcaattgaaccgggta
acgataaaaaaccgttaccattgcctcgatcgatgatttgtttgacca
gctgaaaggtgagtggtatctctaagattgacttgagatctgggtatt
atcagttgaagattaaagatgaagatgtacctaaagacagcttccgaacc
cgttacgaacattatgaatttctgggtatgccatttaggttaaccaatgc
acctgcagcttctcatgaggtgatgaacgaggtattccagcaatatctg
38501 ataaattgttattgtttatattgatgatattctggtatactctaaatct
aaagcagatcatatccgacatcttaacttggattaaagaaattgagggga
acatcagttgatgccaagttcagtaaatgtcagtttgggtgactgaat
tggcatttttggagcatgtgtatcggctcaaggaattcaagtagatcct
caaaagatagcagcagtgagaaactgggagcaacctcgaaccgctcactga
ggtagcaagttttcttgggttggcaggttattaccgacggtttgttcagg
atcttctatgattgcttgcggttaacggagttgaccaggaaggatgtt
aaatttgagtggtgagaaattgtgagcggagtttccagcaattgaaata
ttgcctcactcatgctccggtattggtagcttctgatgatagcggtaatt
ttgagattacagtgatgcttcttgaatggttgggatgtgtttgatg
39001 cagcataatagagtgattgcatatgcttctcgatagttgaagaatcatga
aaggaattatcctactcacgatcttgaattggcagccattgtcttgcct
tgaagatttggaggcattttatattggtgagaagtgtaggattctcact
gatcacaagagttttagtatctatttactcaacatgatcttaactctcg
tcaacgaaggtggatggaattgctgagtgattatgactgcactattgagt
atcattcgggtcgtgctaatgtggtagctgaagcactgagtaggaacct
caagtgactcaatgcttataatgcttgcgtggttccccttcttgcctga
actggggcaactggagtaaggtggagttggaagatcgaagcgaagcttt
tcttgcctagttttcaagtcaggccagtttgggtgatcgtatactcgaag
ctcagatggtggatgaagaattcaagaaatggttcaattaagaaatgaa

39501 gggaaaaaagaagacctcaggattcagagaatcagatggcatgcttatgca
 ggagaacagaatgtatgtaccgaataatgaggaattaaagaatgaaattt
 tggatgaagcgcattgttcagcttatgcaatgcaccaggaggaactaaa
 atgtaccataccattcgaccattttattattggccgggtatgaagagga
 aattcctgagtagtgtagtagatgtattattggccagcaagtcaaagcaa
 aaaggaaaaagccttttgggcgaatgcagccacttcccgttcccagtg
 aatgggaaaaataactatggatttcgtgtataagcttctcgtacacg
 aatggatttgatggatttgggtgatgtggatcgactaccaagttag
 cgcatttcattccagtgagggaaaagtatcctctgaataaattggctaag
 ttggtcattacgaagattgtgaagtatcattaagttccagtgaaatattat
 40001 ttctgatcgagaccacaagatttacttttaattttgggtagcttttcagg
 aagctcctggtagcaatttgccttatagcactgcttatcatcctcaact
 gatgggcaatctgagaggacaattcagacgttggaggatagctgagatc
 ttcagtggtacaatttgggtgattcctggcattatcgtctggacttgatgg
 aatttgctacaataatagtttccactcgagcattggtagtaccattt
 gaggcactttatggtagggcgtgtcgtacaccattgtgtggctgaggt
 tgggtgaaagagtgtagaaggccctgagattgtggatgagactacccaaa
 atattcagtgatgaagctaatctgaaagcagccaggatcgacagaag
 agtttagctgacgacatgccactgacgagtgataatgtgggcgattg
 ggtattttgaaattgtcaccwgggagaggtgtggtagatttgga
 40501 gaggaaagctaagtcgagatacattggaccttatgagatcactgaaaga
 attggtagaagtgccactgggtggagtgccctccggagttatctaaggt
 ccataatgtgtccacgtctctatgcttcgacattatattatgatcctt
 tgcagtgatcctcctcaaccgctagagattaatcggatttgacat
 aatgaggaatcagtcactatcctggattgaaagataaggttctaatagaa
 caagacggtgagtttggtagaaggtgtgtggaggaaccattctgtagaag
 aagctacctgggagacaaaagatcaaagagagatgtaccaaggttg
 ttctatgggtattaagcggtttatttgattgtgggaatttcggggacgaa
 attctataaaggaggggagattgtcatagcccgctcccgagatttcttatt
 aaggacgtgaaatgtcggatttgccttagcgggggctatgttacgtgtg
 41001 tgtgacattatttggattaattctatataagtttggatttaaattttag
 tttataaaatttggtagggaattagatgatggagggtgaggattgga
 tgttgaccactaaacccctgctctctcgccttctcgtagcactctc
 tctctcctcctcacggtttctctctctctcactctcgaagtgtacg
 gacgagaccctaaacccatccaaactcagtggtcgaagcaacaaggtg
 acacctatcccagaaacgagcgcacccagggcagtcacaagggttacgac
 ccatcgacttatcagtgagtgaggcagtttgttttcgtattacctatata
 ttactgttttcccagaaaatgctttatatagaaattatatttctaaaatg
 ccatgcatgcatatgatgattgaattgataattgatgcatatataatg
 tgtgaattggtgcccgtggagcacaagtgagtagcaggtgagtttatatg
 41501 gtttatatgaatgatgatgtgattgtgtgtgagctcataaactgc
 accttgggtattagtgcttatagtagtaccacaccgcacgctcgccttg
 gatccaagtaggtggatgtcgtacagaccactagaggtggttccgaaatg
 ccagtcgtacagaccattagaggggttccgactggtgggtgactttagat
 tatgtgcacagatgattgatgagagaagcactagagcgtattatttacc
 atcttagtcgacagactactataggtagttctgacttatgtgcagtgta
 atgcccgtacaggtcacagttggtgactccggcagggccgtacaggtcaca
 gttgggtgactccggcttgatgggatattgagctatagaatcagccgtaca
 tgaccactgcaggggtctccggttgattttattttcacctgagttatatt
 42001 gatgcatcattatattatatttggcatggcatggcatattctttctgagat
 gttatggtgatagatggttagctgagtttgatgatataatataatgt
 ttatatactattttctgggaaagtatacaggttttacggtgaggggtta
 gaaacgttttaaataaatgttttttggaaaatcttgggttttactgacc
 cactcacctttgttttgcgcccctccaggttctagtttagcagttggggc
 tcacgaggtctttttaggcattctgacagacgttctgcatgtaggactca
 cctgcccgtgttatactttaattatggctcctacttgactacacttatacc
 tacgctctgaattcgtgtgttttactttataactctctcttgtagctag
 taggttatgctgctagtggttgggtttaaattccttcatatttctattata
 tcttgctttcgtgtcgcacttttggctacgtcacactcacgtaacggcca
 gcacgccttgattctaggatcgggggtgtgcacaacattcatgggaaat
 42501 taattttccaaactcccctcatggattctcttcccttcatatttattcta
 acttttagataattttctatgctcataaaaaacttcaatggtacaagaaa
 aattattgaaataaaaaaattacacctaatgagccgaatataaacctcat
 acctctaaaacaaaatgaaaaatatacaagaaaaaagagggagacataaa
 tctaaccctctttaaggcaacaagaaaaatacaataaaaagggggagac
 ataaaagcgatccttcttgaagcaaaaatgagatatacaagaaaaatagg
 ggatgcatgagcctaaccctctaaaaattataaagctttgaaacacca

MC-20SSR96C3For

MC-20SSR96C3Rev

aatattttcggaaataaaaataataaacacattaggaagtgatctaac
 ctgtatgccatgatgtgactaatctgaaaaatcagaagccagaccaact
 43001 aagaaaaattagaatgcaagaattaagaaccgccaacagatgtgagtc
 gctacaccaaagatgaggaacaaactcttaatcatcatcgtgggtgtg
 ttagctcatggacgtatttatggatgatttgcaaatgggttacaaagga
 aatgaaatcagattttcatctcactttaatttatgctattacgtgg
 tagagatccaaatgtatccaatttttatggatttatgattagttggtagg
 gtgcatttttgcaccacatgaactttgggtgatgctaccatacact
 ttattagttgtcatgtgtctcatcttctaaccctagtaaaacttgtaatt
 aatatacttttaaaattttaaaacaagccatccctcttattccactgt
 aactctgattttataaaagaaaacttaacaaatgtcccgcgaagaa
 aatataccccaaatcgaaaaaaattcaaaatctgtatgtgcacagaat
 43501 gtcctcgccaagagagaaagggctgggtctttaggtgaggccgctatcga
 attttatggctgcaatggcgcagccgtcgaagttgtctccggctattgtt
 gtatggcctcgattaccgagtcgacgatcatatagtcatagtggaagaa
 aaggtaaaaaaatgttctcatcgtcatcatccaatcttggcgtgggt
 ggggtgggtgttcatggggtgctgggtggagaagttgaaacactggaggg
 tcgattcttatgtgactagcaaaactcatgtagctataaaattggagctta
 acattatgtcccgataatattgataaaaagcttgcaagaatcatcagat
 ccataaatttcatttttatttatttattgagatgaaaggaatacagcag
 cattcactctttaatcaaaccttcaatattacggtagaagccagttttcg
 aagggaaaaaaatataaaggggattaaaaccacaccaatatttcatgga
 44001 aaagatgacacgagattgaatcacacatgaatgcagcagaaaaaatgtt
 tagttgtaggcatcctgggtgctccaagaggtagttctcaatcaacttaa
 gagggtggagccctctcttggcagccttcaactgttctccttgatct
 caacatcaccttgggtgtagtagtactggctcttgataatggaacca
 ctgtcggaagctaccaacttagtctcataagagatcttctcaattgtctc
 agagatggcatctccctcaatcacactatacttatacacaagttgtcct
 tgtcaagccatcgattctgtgcttcacatagctgtatgtgctaccttca
 ccaaagttgatcttcttaatgggtccaactcctccatctccttcaaggat
 ctcagtgctcttcaactgcttgggtgcaatcttggggatgaggttgtcag
 catcaaggacaaagcattgtacaacctagcaggggggatgacggaagt
 44501 aactcggattcgtatgtgaaaacaccatgattctgagatagtttagga
 agtgggaaaaataaataaccgatgaagaggaagactaaaaaagagatgatgt
 attgggagattttgagctgtgaagaagtgaggaggattatgatatttat
 aggctgaggctgaaggtgtgatcaaaggtttacgcaagctcctaactg
 ttaaatgggtctattcatcgtagacgtgggaaccattacttttgtacaa
 ggaattttgaatgtaatacaaggaaatgaccatctggaatgaataaag
 aaatgcccaattgctccgagggggactgacctaatcttcttttaattt
 ctcagtcatttctattttataagaaaaagagtaattgtattcactat
 atttttatactacatttctatactacttttaggttgcactctgatgtgaacg
 45001 gacacatcattgaaaattttgctaaaccaaggaaaagaaagagaaaga
 ctctttgtataccacaatcatcatttaattaactaatttttcttaattat
 tagttttataaataatgaaactaaatttaaaaatctgattaattcaaatga
 tgtgggtgtccacatcaaatgacacctaaagtgatagaaaatgtgat
 aaaaactgatttgaatagcattcctctaaaaaaaatgatcatatatt
 agccccaaataaaaacaatattgtctgtgggtcttgaaccgcgttggcg
 gcatgcccatagactggctgatatgaatcattaagcaaagctgagaga
 aaaataattttagagaacttaacgaaaagcttctagtagtctcaattt
 aacgaaaaatcacatttttactaaaaagctcaatccttgcactattcac
 tttaccctttattttgccttatcattaaaaactcaaagtttcaaacaat
 tttcattagttttccttttaattttatcctaatacgtacaataattttagt
 45501 ttttttgcgcatatacaatgcaatgaggatgagggtgtgattgagct
 gttatacacagccgacttttttaataattcaatttttgttgcgactgt
 ttagcatttgttatcgcacttttgatcaacgtttcccgtagcggttgcca
 cctactgactcgtcaaacaggatttaaagttgattaggttttttatga
 caagcactaaatgtattgtgagagatcaattgcatattagacaaat
 agaactctgagctgttatatgtatatttttatattgtcaattgatttt
 ataatagaatctcaactttcttcatggatcaaagctgactattcattcg
 acgtgccctggctcaattttctgggatcaaatcacctccttaagttggg
 ctttcgaaactcctgcttaaaaatattgggcctcgaagagatgattggc
 tctagtcctcggatttggatcctctcctgagctaaggagaggattcaaat
 46001 ccttgtcctcaatgatgactcggctttcgaaaaactagacacgcgtac
 tgaaggcatttttttaaatcacgatgtgctacgcgtgacttacacgtttta
 aatgtatttatatgagcaaatgacaaaaatgaaagtcaaatattgaagta
 aatgaataatcttagatcgtgctaaaaaaactcctcacaacaaatcaaa
 gcagctaaatttacatgatataaaatcgttttttcaattttcatctacattc

Mal d 1.03G ↑

MGVFTYESEFTSVIPPARLYNAFVLDADNLI
 PKIAPQAVKSTEILEGDDGGVGTIKKINFG
 EGSTYSYVKHRIDGLDKDNFVYKYSVIEGDA
 ISETIEKISYETKLVASDSGSIKSTSHY
 YTKGDVEIKKEQVKAGKEKASHLFKLIENYL
 LEHQDAYN

tattgaatttacattaagaatagagaaaataatattttaataaacaactga
ttgagtcacattattatgcacgtgcaaaagaccttttttagataaatattg
ttttcagaaaaaaaatcatttgacaattaatttaattacattattatgc
46501 acgtgcaaaagacgtttttataaccgacattacacgcctcttaggatgt
tttgaacgtgtttaaaaatggagaaaataatatttaataaacaactgatt
gagtcacattattatgtatgtgcaaaagaccttttttagataaatattgtt
tttagaaaaaacaatcatttaacaactaatttaacacattatgatgcac
gtgcaaaagacctttttataaccggcattacacgccttttaggatgttt
tgaacgtgttcaaaaatggagaaaataatatttaataaacaactgattga
atcacattattgttagcctattgtgaggtactaatttaaaaaatgtacaa
aaagaattaatttaaaacaacgctgttagaatgcaaaaaatcccctg
cattatttgggtattattttgggttgccttgaatttttgttttgggt
catttttgtccaaaaatttttgggtgaagctgtgaccccaaaagaggtg
47001 ttggctttatatttaaaaatattgttagctccaagatggtgacttgatc
aattcccaatgtgggaattagtttcccttgtgacctcgtgacaggccact
aatcagcaggtgtgtagagaatagaccttaactttcttcatagattta
gagcggactattcttgcacacgtctcgaacccatttttgggtcaa
ttacctccttaatttgggttgcgaaaagtggcctagtaccaacttgag
gatattgactgattgggttgcgaaaagttccgttgacactagtccctgaat
gatgattgggtccttaaaaaatttgcattagctttggaatgttggcttga
ttagtttcccaatgtggcatttgggttccctttgtgacccatgttaggca
actcgtgaggtgcatattggagaataggattgtatattagttatattgac
ataatatacaattactatagagagtttccaccagtaataattatcgagac
47501 attttgtgcaaaaactcaataccaataatattgtgattaagttgggacaa
tatcaacaatattggaggtagaccatgatggcctaaaaagtttcttcat
gatattagagcatgtggtcatgctgtgatgtccaatcgtcacacgtgat
gtacatcgtttattttgtatgtacatgtgttagacttaaaaattcgcca
cacctaagaggttatgttgagagtaaatctcacaacggagaaagaaatga
cattgcatgtgcttatgagtaattgtgctactccccatattgccaattgg
ttttgtagttgaacctcaactttcttcatatagtagaatacataaaatga
acaaagaagcatgcatgataaaaacaaaactaaagtcacacacatagagtc
gataaataatgtacaaaatattagctcgaactattattgtctcattactat
tcgagctcaatcaaagaagtcatttcaaatatcatctcaaagtttcaaa
aacactaatgtaacgatattttaaatcaatcatccgaactgtttgacta
48001 aaacacttggcaacgcataacattgcaatataatgagagtccttgtgttg
atcatgtcaaatatccaagacaccgtctaaatgctcacttccaacaaca
ccgatattgtctcctaacttattaattcccttgcacacttcaacacgtgtg
aggttttattacaaaggcctogataatattgttgggtgaggtttggatatt
taaactgttattttctctgtctctacctaattgtgggatattcaacacgac
cctccacttgggaccaactagtgaatcacacatgagagaatcacatc
gacaaccacgaggtactcaccacacacgtggaccaagggaacctaacatca
tcacactgggcaaaagcaaccacgcacacatggtagtagatcctgct
ttctagctccgataactaagtcataatgggtacaatcatctcagaaaaaac
aacccacttactcccaacatagtagtaaacaggggaaagcctcacat
48501 cacatgcatgatcgaagatcatgcaaatggcataatgaaatttcaaag
atataaacttgcacatagaatattcctctccgcaccacatcacatgtcat
gatcgaagatcatgcatatggcataatgaaatttcaaagatattgaactt
ccacatagaatattcctctctgtctctccctaattgtgggatattcaacac
gacccttcaacttgggaccaactaataaatcacacgtgagagaatcacat
atcgacaaccacgaggtactcaccacacacgtgggcaagggaacctagca
tcatcacacaagaccaaggcaaccacacatcacatggtaatatgatcct
gctttctatagctctgataactaagtcataatgggtgcaatcatctacaaaa
ataacaaccacttactcccaacatagtagtaaacaggggaaggcctcac
atatacatgtcatgatcgaagatcatgcaaatggcataatgaaatttca
49001 aatgatataaacttgcacatagaatattcctctccgcaccacatcacatg
tcatgatcgaagatcatgcatatggcataatgaaatttcaaagatattga
acttccaaatagaatatttctctatgtactcgagggcacattagccgacc
cacttaattgggaaaaacttgttgttgttgttgcactctcaaagggtg
catgtttggagataaaaaagaagcataagttgggagtaaaattaagtacaa
aaaaataaatgactagcaccataaacttgggggagtcgggatataaattaa
gtaacctatacccaccaccacgaaacatggataaagctaaagaatttcat
gcataaacactaatttttagcatcggagtagtttctttcaagcccttctcc
gaccttctacacgtgggttaactgtttatgattcataaagaatgttaagg
ttgtactgtaaaaccaattggcatttaatttttgggatgtagaattcactga
49501 acaactaattattggctttaggggtgagtgccaaggtctcatatacatat
catatattgcaaaacagttatcagacatctaccaatgtgcaatcactttta
tacaactgtaggaaaataaaggattgtaaagaatcttaaaaggattttca
gggtttatacagcatcttggcagtggtatagcattatcccattatttggg

agcccgtttaacgtcggtaacaaatgaattttcaagcaaaaaacgcata
gtacatatgatcaaccacagaaaaatatccatgcacgtcggcactaaga
tcgaaataagtttggagacaatcgattttaatgattgaatgtacttcaat
gtaagcctgattttcaacatcaaacactcctcccacaagccaccaactgc
gcagagacagggcattggccacccttggaaatcacgattggaggtgaatctag
tgtagaagtattatcttgccttttagagatttttaaacttgaattataaga
50001 gattttgcatcgaaaagtatactttgtagcatgcaaagtacacgccccg
actatgatatctccaaatatcagaataggcagctgtagctgacacccaa
gggtgacgaagccatttttaataacatgcaagtggtattggtaaacaaa
gtaaaataaaatagcaaacataaagaaatagcaaatgcggaaacgtgtt
cagagtatacaactaatcaagaataaagtgaagaattactataaaaaatt
acgaacaaaaggaatgagtcctacactaagaagactcgaagatatctatgt
ggaagtgcctaacgtcgggatcgtatgcctctgttctaaatcctaaatg
gggcacaaaacagaggggtgagttgaccaaagttataataataactaa
taataagaaaaccatttctcttctgaacataactaacctctgttttgaaa
actcatatagtactacataatagggttttctgaaaattctagcatgccatg
50501 aaaatgttcgtaaaagatgtatgtataaaacagtgctcagtaatgatgac
atagtagcccaaggcaaatccataaaaccttcgtaaaactcaaaataata
aagtaactcaactaggggtacgtatcatagttgcccaaggcataaacatc
attcgaaggtgaagttgtacgacactgggttacaccagtgaaaattatgg
taggtcccatcaccgaaggtgaagctatacaactctgagttacatcaa
gaagaaatatacataacacgctgacactagccgtagctagtgaagctgtc
tgacactaattttgctagtgaagctggggtaggtcataaatatcctcaa
ctatgttctatggctatagatgtatacactcaggttgtgtgtatgtgt
tgtatgattacccactagataagcacataaaatgtatctcaaaaaaaac
caaaacctcatcacctgaaatcataagaaatttcataaacgtgatttcta
51001 taaacataccattttataaaaccgtaaatagataaaccataaatatcttgt
aaaagcaatttaaataaatcataaattctccataaacataaatatagaa
atccgtaaaatataatataaaacataactcaatcattaaatagaaatgt
atggaagcctaattgttctataaaaaaacatgcaatttgagaaggggtcca
ctcacaacaaactctgtcgtcgaagagccgtgtaatttaagaaaacaggg
ttgcattaaacatacacacctaagaacataaaaaggaataatataaaaacc
ctacttaaacgattgaatttgggaaatggagtgaggatttagaatcaagg
cgtcgaattaccctaagaggggtattggaaaaaatttgtaaaaaatcaa
cgagaagtcaaccgagcatcgagccaatcaactacgttaaaactttggaa
ttccactaaatgcatgtcaaaaacaaacttggagcatcgaatttagatta
51501 ggagagttcctgagaaaaatttcgtaaaaagtcgacataaggaaaaatattc
ctagaattttctgggtcaccgagcgcgtcggcgccgaggaaccaccg
gaattgcgaaaaattggccagaaatagggcacttttcaaatggctatct
tctctcatttctcaaccaaatcgacccataatatacaaaaacaaagata
agaacaagattatacctgtttggagctaaaaagagggccagagatggcca
taaaacgcctgcaaaaggtgtggaggtcaccgaaaactgggaaaaattcaa
atggttataacttctttaactcaatgaaatgaaatgattcaaaaatga
aaatcatacctcgcagcagacgaagagaattgtacttttcttgatggct
agcttctcatggtttggccgaaaaataaccaaaaatgtcggcgcctgatt
tcattatgatcccaacgatctggaaagaagtttgaggttaggttgggtgt
52001 tctagatccgtgagaatgagacaaagcttgggtgttctttgtttgagtg
gtggtggttgggaaagttgtgcacagtgacaaagggagagatgggagagc
gggagagttccagagaaaacacacgagggagaaaggtggaagtggatgtg
ggccccatggcaacaataattatataaaataaagataaatataaaata
tctaaaatagtaattttacaaggtatctttggatttgtagtcccgtaa
agtctccgtcgtagtttgcgaaatcgagctacttacaccacatgctcgt
aacatcgagcactacaagaatataactaacataatgagtcatacgtcacac
aatacgaggggtcaacaaaaaatgatttctccttgaaaggcaatttcatac
ccacactttaaattataaaaatcgtaatattcgggacgggcccgttacag
aaagacgataaatttgggtgtattaaaaaagattgaagaaggaggggatcg
52501 gcaaaatgctacgacccccagacaagagtcgacatagtacagactcctcc
gacctaaggtcaataaacgtattaccgatcctcaccttttcaacattttt
tgtggttttttttcttgcactatattttatagttgagtttattttg
gttttcttcttttttcaactattaattatctaatttgatttaattttt
tttctttacaaaattataagaagatgtagtttctgggggtattaaatgct
ttaaattgttcttttcttttcttttcttttcttttcttttcttttctttt
ttggattttgcaactcattttcaaccattattctttactcattttcaat
tatttaaagatggatttgatgtatctttcactcattttcatggatgtgt
ctacaggtagacgatgtctctattaggcatggatagaatagtatattgt
ggatttgtagagaatattagacttagagagaagtcacgaaagagataga
53001 tgtcccattgataaataactcaggctagagatattatagggatcttaagag
tctttgtagggaaagtaattctttatcaaatcgacagactgtctaggagga

aatctaataaagataactcaaccctcctataatcttgattatggtgccta
atcctcagaatatcctaatttgaccttaaataggggttccctacttggga
ggcaagtaatatcttcaataggcctatgcctagcttgcggactgtgttt
catgggctatgggcattccgatccttctgtcacatgtcatgggctaga
aaatcatggtcccatacacgacttttccaaacagttcttctgaatgaaga
ataaaaaagatatggtgccatcagtggttttagttagggttaagatggtgt
tcaagccatgaatcaagattgaagtaaccagttttggaatcattgaagat
gttggtcaaaaaccaatggtgaaggtgtaggattttgggaaattatgagtt
53501 ttgctaaaaacttggttagaagaagatggtgaacgctcgaagtttaaagca
gaaagtcttgtaagcaaaagacaagataagagaaaagaaaagaatccag
gggtttatggatgaatcagtaagaagccttacgggttctgtctgagaat
ctaggaggaatgtcggggtatgctatgtaatgattgcaagaatcgaacag
ctcgggttggtcttttgcgttgcccaacgaaatttttaaagtgactgc
ttaatcttaaggcagtcataataagcaaaagaccaaaccttttggtgta
gctactttgtctagaaaatcgtggcacgtgagtagctgagttcgaacagc
cgtccctaagaatctggcacgtagtaggagcaaaagagaagtgtgttagtc
gatgcaaggtgtgcacgtaaatgggattgaggcaattaatgcaagatag
taggtcatttaaagaaggccaattggattatcctcccgtagctcatag
54001 gaaacttgcatgatgatccctggtgaaaaattagaaaataaacctt
gtggtcaaaaatgtagcaaattagtttaaagtgaagttttggttaa
tgacagttaaaagtataggtaaaactgtatcttcagttccaattgacatt
taaaataaataattttttataacaattaacaaaaaaaataatgctta
tttttaataaaaaagccgcaattcaatgtcaattggaactgaaaatac
agttttacccaacttttaaccgctccttaacaaaaatctacttttgga
ttaatcttgctaacattctagaccacaagggttaattcctagtttttct
atcacggatgtcatcatgcaagtgatgaccacagggatgaacaatc
caattagcctttaaagaatattaatcttaaaaaccattatgccaaagt
gtcaatgggcaagatctcagaagtgaggggcaatgttaggcctaaaat
54501 accaacttaggctgagatgggttgacctgggttgaaaaggccacgatact
tggttagtaagtatccctaggccagtggtcttttagccaagacaggg
ttggttcgactgagtcctcataaaagtaagattatgtaagattaagtgc
atatgtgctcagttgagtcctatgcaaaactgaaattgccaaggatcata
taagtcttgataaatacaaaactatcacagggtcataatgcctacacgat
tagataaagatgaaaaagttgcatgatgagatcaaatcagagacctagt
tagagtcaaaagatgatgagctggatgaagttttggctaaaaagggaa
aatgagataaaaagaaatattcaagtaaatgtcgagcaagcatagtcct
cgaactctgcctcgatagacttggtatcataaaagccttgccataaat
55001 aaaggaggtctgcacaactgggtgctagggccgctcataaatttttagg
cccggcgttgccaaaaaatttccctaactcatataataaaattattt
atcttcacatgtaagacatcgataaaattatacttagaaaagcagttcaa
actcaataatttaaaaacacagactaatttattttgtatcaagagaagga
aoccaaaaaactcttggttacaagtagatagacgataccacctcgatca
tcatttgccctgcagttataccttggtgttctactcaatagggtggtcca
tagaattttgaggtcggcgcgacacaaaaaatgtgccttaacttcata
taagaaattatttattttcacacataagacatcgataaaattatacttag
aaaagcagctcaaaactcaaaaatttagaaaacggagagactaatttatt
ttgtatccagagaaggaaccaaaaaactctagctcacatgtagataga
tgatgagttttgtttccatctgaacctttaaataatttttgtattaatc
55501 gtgaggtgttttttctcatttactaaccttgcaattgagactaaaaaa
ataatgatgtttctgagtccttaattatatagtataagtaaatgaatgggca
ctaaattaaacattttgataaacacgtcatatgattcgcaaatgtgctcta
aaatgtggaatacgtattactctttatagaagattagactgaaattcaa
acgaatatttaaaaatagcaaccacataatctactagctagtagctaaaa
ataaattacaacaaaaaaatattcgtaaaaattgaaaagaataaacac
gcatatagcatttcgaacttagaaaatctcgtaattttaaacaacaaaa
accattgcttcaattaaacttattgatcatttaaccaaatttaataaat
ttatactcttatgaaaagaattttgtaaaaattggaaagttaatgagcat
atggtggtgtttgaaacccaagacctctgcatttttcaaatagacaaaac
56001 cactactctagttaaatttctttgtcatcaatccaaattttataaattt
atactcttatgaaaacataataaataataccaccttaataattttggtgc
ccgaattttttgggcccacggacaatcgctaccacactgggctggg
acggccctaatgggtgccttttcaactcaacaacaactcaaacactct
atgtataaccattctcacaccttaagcaaaacactaactatgatatccc
tccatcaatacatcttcaatttgaataagtgacaacactagttcagga
attcccaacttcccactaatctgggctctcattttcttgaacccaagg
cctatatgtgttccatctgaggccctaatgtctcaagtgcagaagtcaac
agtgcatcgcacaccacaaccacatatactataactcgccaatcgag
cttggtcgcgagtttagcacgcccgcataatagaaggaggacgtaagatcat

56501 aattacttgacatgCGTgCCacacaagcattgtaatttctagactaaaca
 atatcaaaactacgatgtagaggacgatgtttgagtcAAaggTgattagT
 TgaaaatgCGctttacctaataaaattgTtaaaagTtGttatgtgatgagT
 ccaataataccataataatttaccataatcttagttctagtttattggTt
 attttggtagaatttgatactttgTtatataatttcaatgtaggacatt
 cgacttcatctggagcaaaaagTgatcaaacggatgaattaaGGTaaaca
 tcatgctcttaagggTtgcattggTTTTTcaaagggTTTTTcaaaagctta
 atgagtccttgcattgTtatatttgatctgaacatcacagaagggTtgcatt
 gTtagatatacgtTTTTgcttgaacatcacaggaaatattgattagaa
 aacctaaccttcaaagcattgatgtagaatttataagtaattggTaaaa

57001 ttgcaattgttaagaggaagCGaaccttaggcttttacaattggTttt
 caaaacttttccctttgTtttctttacttTgctaatttcttTaaacctttt
 taattaaattcgTTTTtattatttttaaattcaaaatcaaccttttccaa
 tctttgTtttcaaataaattaagattggTatttacataaattattc
 attcaatctctgtggaaaacaaccttactgagttgTtatactacaaca
 gattgcactcttacaagTgttttaagTgtttttatccttacttttatag
 Tgggtcattatgtaacaaaattggattataatgtagatggaaatta
 gatcataaacactcacattaacctttcaataggttttctgTTTTggT
 gtaccagagggcgTtgaggggagcgaaggTgcagTtgcataaggc
 ctcaaatTgaagggacccccaaaatttttGCCatctaattgTtatatt

57501 tagaaaattgcttttggtagtactttaaactctcattatacattcctcat
 aagagTaaattttgTtcccttctaataataaaaggTaaactagagata
 ctaaattTtgaactaaatgatgagtcaccaattggaaatgatcacgTtt
 atcaatgcttaagtaataattcaatcatcaacttatatgctatttaatt
 acaaaatttTgctacaaaatttagtatccctggcattactcaataaaaa
 aatttttagtacctgacataattctgattgTcaataaacacctaaaaa
 caatagTtttccaattttattatataataaggttatataattcaagtgaac
 tttaaagTtagagTtttgaagTttttttcttcatgTttgattgTttaag
 attgcactgcttttgattatctagtgaaggttatgtttatagctttttta
 attaatgacattTgtaaactTgcaatcagTgtgatttaagcgattgTtt

58001 tataacatcaatacattgTtttacttttttaataattatcttaacgagg
 gTcctttttataaatttgcacacagggcccccaaatctcagagacggccct
 aggtgtacctttcataaatgaggaccagctgaccaagggaagcaacctg
 agaacaatgaaattcatcacttctcttagattttctaattttcagTtc
 aatggTtttGttatataacacgggaagTgtgtctatactgctgTttg
 Tgggaaataaaaataaggaaaccataaattaaacctattgtgcatggaac
 atgtgCGtaagTgcaattTgatttaaaatacactgctggcgatcgggtcg
 atcgcggttTgctccctcactactgaataatgctcatgCCgatcgattgtg
 aatgTacaataaagtagaccacaaacactaattattaaaagtCGcatttaa
 taaaaaaaaaatgataaacagctaattattagTgtttaaccagataaaaaca

58501 attgcttatatcatatagTttgTtagTgcatcccagTtggTttcatatt
 TgaatacacaaccgaaaaaaaaaacagagTtGttgaggagaatatCGca
 cattggagagatgaagTttccgtcaaagatgtagaaaacgtctccattat
 TgcagtCCGCCattttttttcacagcttctcatatccttctcctcaaat
 ctctcttttGctttttggacttctcctccagattcaagataacagTatt
 ttttttttttttaaatattggaacaaaatgtTgaatttatgttaagct
 tattatattaacacttcacaaattgTtgaataaacccccacatacttaata
 cccccatatttgcttttTcacttaaaaattatcttaatacaccacagT
 actttctcataataacctcacacccacattctcagTatacaccttatat
 tttgtacataataccccacattttctatacatcccatatttctcaaatctt

59001 ataacactcatttttaattttcaaagactgaatcaaccatttgaacgTt
 tagTttgCGaaagatttcaaatcatcagcaatggcttcatTgtgaaggt
 cccatgagagctgggtTgatgaaaactccctcaatagaaatccccgTaa
 attattactcttagaaatttaagTgtTgaaatgacctTgtgaagggTca
 tctgacctTgtgaagccaacctggctcatTTTTaaagacataattgaag
 cctgtatcatatagTatgTttgTgactttgTttttgcaactaaacgatat
 atgtcaggtacatctaattgctTgtgatgcttaggataggcctgtagaca
 TCGaaatttcggaaaatttCGTgttcacaagggcactTgCCatacagat
 cataactTgTcaatgagctTgagcaggtTgtaagaacaaccgagctacta
 tttgacatttcttcaatgtcatctttgTtGCCaagatgataattaatatt

59501 aaattttaatgtgagagTgcattcaacattttatggggTgctaataaaa
 aagcctttatgtattttggccataaaaaatagaattttgTtagggTttca
 aaagTcattttataaaaggtTaaatattgTcactaaaatttccaataaaag
 TgggtgggaaaatagggtTgagaggTgggtttatcatcactctaaataag
 gTTTTatcacaattggTctctaaaatttaccatcccatcaaggTgtc
 tctgaaattgaaaatcaatcaatgtagTcctggaaataggTgtcgaaa
 tcaatataatccttccgttacaattctgttaaaaatttcatTaaagTgctg
 atgtggcacataaatgggtcccacaagatttacgggtttttatcacaat

gatccctaaaattgactcacaccatcaagatggccttgaattgaccta
 caccatcaagatggcctcctaaaattgaaaatcgatgaatgtagtccttaa
 60001 aaatattgtgtcgtaaatcaatatgatcattccgctacaattctgtaaaaa
 aaattggtatatgctaattgtgggtttaataattaatttaaaaagttgttt
 aaatacatttttgacgataaaaactttttcttatagtaggacctactcc
 gaaaagatatctctccataaaattcttaaggcacaagcttaaccaatg
 ccaataggggatttagaaaagaaaatacatttataagaaatgcataata
 agatttttggaaatgtcaatgacacgtccctaaaattttactaaaattgta
 ggatctgtgttcttagtccttgggtgaggttctcccaaaaatggtacaatg
 gaatagatagcctctaacttcaagttatattgtaggtaaatgatttttcat
 ttggaagggattagaagagactattgatacacgaaacttctccattctaa
 ccataccatttttgaaaagaattgcaatggataaattttctcttaagta
 60501 atccatctcctacctctatgtttgccacaaaatttttcattttctctta
 acataccttgaagtaaatggtttatccattcaaatccaacctttttat
 caaatgagacccttgataatcctattaccttcaaatgatagtagttt
 agcattgtaagctttcacacgcatgtaattagttggttaatgcatatttct
 tttaatagtagaaaatgatcttgcaagattaactctttgaaaatagtac
 atgcataaaaggaatgacagtgacaagaagaaaactgattgcaaatcttt
 gaattttattctatagctagacacacaagaagaaactacacatttac
 aagaatactaccctaattacaaatagtagcagttccagaagttcaagtt
 aggtggccttgacttgcatttcaagatgggctagctgcaattgggtcaa
 attaattgcccatcgtttgatcaaccttcagcctcaccctataaatacca
 61001 tacctcctcctcacttcttcaaaactcaaaatctcccattacaccatct
 tcttttccataaatttcttttctcactttccaaacacttcgcagaatca
 ttgggtggtttcacatacgaattcgagttcacttccgtcatccccctgct
 aggttgtaacaatgcctttgttcttgatgctgacaacctcatccctaaaat
 tgcaccacaggcagtgagagcactgagatccttgaaggagatggtggag
 ttggaaccattaagaagatcaacttcggtgaaggtagcacatacagctat
 gtgaagcacagaatcgatggggttgacaaggacaactttgtgtaccagta
 cagtgctattgagggagatgcaatctctgagacaattgagaaaatctcgt
 atgagactaaagtgggtggcttccggcagcgggtccggttatcaagagcatc
 agccactaccataccaagggtgatgttgagatcaaggaagagcatgtcaa
 61501 ggctggcaagagaaggcctcacatctcttcaagttgatggagaactacc
 tcttggagaccatgatgcctacaactaaattcttctgcccctagctaca
 atgattttctaagtattgataaaaaaaaaactttgtgtcctctttggatc
 agttaatgaattgttcggtgtggttttcattttcttttcttttttctttt
 tttttacttctaaaacactattctggcttgaattggagctttgattga
 agactcaataattgtatcttaccttatacaatgaataaaaggaaaatc
 atggatttgataaatatttgaatggttcttttaacaataggtttgatg
 agttggtctccttaagtgtggccatgtcgaattcaggaagggtggcagc
 tgaagccttgagagccgacctgttaggcatogaagatgcctccttaag
 62001 gattcttcttggaaaccacatcacataagagatgacctttaacaaacc
 aacgtatctaacaatcaattagggccaatttaagcatgcatatcctaaa
 tacggaatcatatttctaaatgaaatgctcatgttaaaaaaaaaaagac
 atgctcattaaaagtaatttatttcaaatctgtaccgaagccgggagtc
 gctgtttagagtgatgtgaccagcgggtcagagctcctttgaaatatt
 taaccggagcagagcaatgggtcatctccgatccgcgatcatagttgat
 actccgagtgagggtggctcaggcttgaaggctgaacttaagctactgg
 ggattgctcagagtggtattggatctgtgatccgcgctcgtcactccgac
 ccgagggctaccaaccagcagctccaataaattttagggtgtgatcca
 cacaccctttttatttttctcacaccttttaatttttaaccgtcggat
 cggatgaattgaagaagatcaacggataaaaaactaactaaagggtgtggga
 62501 gaattaaaaatgaatgtgtgaatagcacatcctaaaatttatggagtga
 attagttagtggggacttgacttgcatttcaacgtgggccacctgcgagc
 agctccagcgggaccaaactccccttgcgagtaggcaatccaatccattca
 tccccctcaaaaacctccagcccaatccagggatgggctccaggggga
 gcccaattctccatcagcggataatctattggccagcgcctggcctac
 ataacctaaagactgatgtcatcgtaacgttaaccacaactaaattatta
 aaaattcgttaaaaaataaaaaattaattaataaaaaataaaaaaat
 ttacataaatcatgacctggttctttttttttttttttttttataacca
 tgttctttcactttatcctacatttcaataattttcaataactttcaacca
 agcttttattatttctgtaataaaaaataatatttttaatatcaatagcc
 63001 tgaactattcagtttagagatgcaaatgacttgaacaattactgttcac
 tacagaaattactgtttacataagggaaactgagttgcttataaccatga
 ggcattttagagtggaatgctctgtcaaacattgattgctcattggaca
 aacagttgttatgaacaataattgaggttcgtcgaccttgggtaagcgg
 ccattttcttattccaatatccagatggcctccctcccttgacaaa

Mal d 1.07 ↓

MGVFTYEFEFSTVIPPARLYNAFVLDADNLI
 PKIAPQAVKSTEILEGDDGGVGTIKKINFGEG
 STYSYVKHRIDGVKDNFVYQYSVIEGDAIS
 ETIEKISYETKLVASGSGSVIKSISHYHTKG
 DVEIKEEHVKAGKEKASHLFLKIENYLLEHH
 DAYN

aatacaaaattacatgtacataaattaatggttcccacgccaacgaatag
 aaccaagcaggaactaccataaataaaccttttgatcaaccctcagcctc
 actctgaaaatacatatagctcctccctcacttcttcacagctcaaaaatctc
 ccggtgcatcatcttcccttttcatcaaattttttctcactttccaaca
 63501 tttcgtagaatc[atggggtgtttcacatacgaatccgagtcgacctccgt
 catccccctgctaggtgttcaatgcgactgctcttgatggtgacaaac
 tcatcgccaagcttgcaccgcaagctgtgaagagcgttgagatccttgaa
 ggagatggcggagctggaaccattatgaagatcagcttccggtgaaagtag
 cacatatggctacgtgaaaaaagaattgatgagatgacaaggagaact
 ttgtgtacaagtacagtatgattgagggagatgctatctctgagacaatt
 gagaagatctctatgagactatggtggtggttccagcaacgggttcgat
 catcaagcgccttaccactaccacacaaaggtgatgttgaatcaagg
 aagagcatctcaaggctggcaagagaagcctccacctcttgaagttg
 gttgagaactacctcttgagcaccaggatgcctacaac[taa]attttctt
 64001 ctacctcttactgattttctatgtgttgatcaaaactttgtgtcctctt
 tagcatcaatgaattatgtggttttcattcctttattttaacttctgct
 tctggaatactattaggcttgtaattcgagattttgagtgaccagtgaa
 tattgtatctttaaatttacaatgaataaaaagaaaaatcatggaggt
 ttgataatatattgtatttttttgtgacaccttggtaatgaatttaag
 cttcattttataagtatatgtgacccaaacttacatatgtaattcatatt
 aatcaccaactgatcttttaagagcaaatttcaactaaagattataata
 aacttatataaatccaagtggttagagtgttgaattttcatttgtgtcc
 cgagttcgaaactcctcatttcttaaatatttgtaatatttttaaagat
 aggcattccctcaacttaataataacccaaacctgaactaatccctccc
 64501 aaccataccccaccatccaatccagacctgaaccagctccttcccaaatc
 tcaaatgaatgggattgcatagagttgcatgatcgcaaccatttcttc
 gagctctacataaaacctctcccgcaaaacgagcagtggtgatcggttccg
 ttogagaatttctcaaaaaaaagcagttatcatagccacatttggga
 agcggagctttttgaacttttgattattctcctgatggcggcgatgacta
 cgccgctgatgagtcaggtcgatggtggtaattggtgactgcgaaa
 ggccgggatctaccggcgatagactaagggcctagaccagagagtgcaagg
 atgagaagatagagaggagtaggatgggcaatggttatcgggcgggta
 acccggttaatttaccataaaccgtttatggtcataaccgataaacat
 ttaccggttgggtaattgctaaacggttatactcatactcataaccggt
 65001 tataaacggttaaccatactcatagccgcataaccatttaaccgtaaccg
 tttaataccggttaccatattatttttaaccggttatcttctttt
 tttttccattaccctttttcaccgctcatatgtttttaacaacttga
 aaattaaaaaaaattgtcataattttttttaacaattaaacaccggt
 aaaaaaaattcatcatacatttccatattttaatttttttaagtcctta
 taccattcaataaattgaaataatagtttaacgaacaatatttttaattgt
 taccatgaaggtaaatataatatttgctaaagtattattgggttacaaaa
 attggttagaaggcatttctgtcaatttcaacggttaccacaatgaattt
 aaattaatttggccaattccttgatgatgagaatcatcatcctgtagta
 gtgttattgagagaatgaccgatgaattccttgctaatttattgggtgct
 65501 aagatatttggtcgagaaacatggtttgtgtagttttacatatataaa
 ataaatgggtaaacggttaccggttataaccgtagttaataaccataac
 cgccattttaaatttctggtggttaaacggttataaccataaccatttattt
 atctaaacggttacctataaacgtaaccggtttaaatttaaatgggcggtta
 accgcggttaccataaccaatgagatattgcccacccctagagaggaga
 cttttccgcaacgaggacaacggttgacgtggcggaggttaggtgggg
 tgggtgaagggttaaaccagaaaaataaaaacaaacaacaaaaatggtg
 atatctaggaaaaaacatttcaatataaaaaggaacacggtggtatattac
 gtgtttttatataaatggtgagaattttattttttaagttattaaattt
 ttaacacacaaaattccatcatttgtattgtgacacatagtgatcacctc
 66001 gtgtactggtcacactgaaaaatctctcaatcttagggacgcattcgtt
 tttttttataaaaaagaaattaataatttttatcagttggaatggaac
 aaattagcaactacgtcaatccataaatatggaatccattgttacaacat
 cattaacttaacaaattctctaataagatgtataactttgcaacggatttgt
 gagtttatgtatgacattgaaatgtgcaaaagatgaacatgcatttgtta
 tataacctaaagttgcagtggtaaaatgtaattttccaataattttcat
 tttgacaagaatagaatagggttagatgaaagaaatcgggtaaattacac
 aaaattacctcaactatgggtcgaactacaattaggagtggtgatgatgc
 ggttcagtgcaattccacccttaaatcgaaacgggaaccggaaaaatatca
 acagttcgggtttgatgagattccacttaaatattactagaaccttacg
 66501 gttcgggtctacaccggttccgattcgggttctgcccgttaccggttttga
 atactaaattatttgaacaccaaattcattatgcattcaaatatattctt
 aaaaatgattatataaagttgcatccaagacatcttttctaatgtaaatg

Mal d 1.09 ↓

MGVFTYESESTSVIPPARLNFATALDGDKLI
 AKLAPQAVKSVLEILEGDDGAGTIMKISFG
 ESSTYGVVKKRIDAIDKENFVYKYSMIEGDA
 ISETIEKISYETMLVASSNGSIKRTCHY
 HTKGDVEIKEEHLKAGKEKASHLLKLVENYL
 LEHQDAYN

t c o t t g g t a g t a g c a c c g a g t c a a c a a a a a g a t a a c t a a a g a a t a t a
c a t t g a t c a g c a a a c t t g t c a t a t g t g t c a g t t g t c a t g a g t c t c t t g g c
a c a t a a t t g a t t g g a c a a t g c a a c a a a a g t c g a t a g a c a c a t t a c c t c t g
t g g t a c a a a a t g c t t g t c c a a t a c c a a a a t t t t t t g c t t t t t c a t t t g
c c g t a t a a c a a c a a t t g a t t c t t g a c c c a c t a t a t t g g t c a g t c a g t c a a
c a a a t c a a a t t a a c c a a a t a a a a c c a a a a t c a t t c t c c a a t a c t c a t
67001 t t t a t t t c a c t a g a a a a c c a t g g a a a g a g a g a t a a c a c c t a t t t g t a a
c c a c t t a t a a a g t g a a a a t t a t g a c a t t t t c a t a g c c t t g a a t t t a a c a
a g a c a t t t c a a t a a t a t a g a t c a t g c a a g t a t t a a g c t a c a g t c a c c a a g
c a g t a t a c t g t a c a g a c t a a t c t a a g c a a t t c a a c a a g c a t a a c c t t a t t
t g t a g g c t t c a c a a t c a t c c a g t t a a a a a t t a a a a t g c t a g t a a t c t a g
c c t a c a c a c c t t a a t t a a g a t t t g c a a a a c a a a a a c a c a a a a c t a
c t a c a a a g a a a g g a g a a a g c c a c a t a a a g t a t t g t c a t g t a t t a t g a
t g c t g a a t c c c a g t c t g a g a a t a a g g a a a a g a c c a a a a t t t t c t a a t t a
a a a a t a t t a a a a t t t c t a a g a c c a a a a t t t g t a a c a t t a a g t g g t t c c c
a a c a c a t a a a c c a t g c c c a c a c t t c c t t a t a t c a g a c t c t t g t a c c a t t
67501 g a a t t a g c t t a a a c t g c a t a c t t c c t t t t a t c t a t g t c a a a g a t t t t a a
c a a t g c t t t t a g a t c a t a t a g c a a t g g a a g t t t g a t c a g c t t c a c a c t t c
c a a t g g a a c t c a a c c a t a t t c c t a a t c t c a a a t a t g a g a a t a g a t a g a g a
a a a t a c c a a g g a a c t g t c t a a g c t g t c a c g g a g g a c g g a t c g a t g g a g g a
c g a a a t g t c t a a g c t g t c g c g a g g a g g a t g a t g a g g g c c g t t g c a g t t g
a g g t t g t g g g c t g a a t c g c t g a t a t t t t g g a t g a g a g t c t g a g g t c g t g
a g t g g g t g g g t g g t c g c c g g t c g g c t c t c t g t g t g a a a t g g g a t g g g g g a
a t c a g a a a t g g a g t g t g a g t g t g t g a g g c c g t g a g g t c a g c g a c t a g t
g c g a a t g a a c c c a a t t g g g a t a a a a t c c a a t g c t t t g c a a t t t g g t g a a
a c a a a t t t t c a c c a a t c a t a c c a t g a c a t g t g t g t a t a t a t a t a t a t a a
68001 t t a a a t t a t t t a t t a t t a a t a a a a c g a t t c g g t t c g g t t a t g g c c g g t t t
a t g a t c a t t c g a a a t c g g a c c a a a a c c g g t t t g a a c g g t c c g g t t c g a t
t t t t g g c t c a a a a t t g a c t t t t c c g a t c a a c a c g g t t t t t t c g a t t t t t
t t t t c a t a c a g g t t c g a t g c g g t t t t g c a g t t t g g c a g t t t t a t g c c c a c
c c c t a a c t a c a a t c t c a t a c c t c a t g c t t t a a a a t t a c a a t g t c a t a c c
t c a t c t t t c g a a t t t g t t g c a a t g t t a g a c c t c g t c a g t t t t t c t t t a a
a t t t c t c t g t t a a a t g c t a a c g t a g c t t g a g g c g g g c c a c t t c c t a g t c
t a a c t a g a t t a a a a a t t a a t t a a a t a t t a a a a t a a a a t c a t t t a a a t a
t t t t t a t a a a a a a a a t a t g g a a c c c a c c c t a c a a c c a t c a c c c t c a c
c c t t c t t c t c c a g t t c a c c c g a t c c t g c c g c g c c a t c t t c c t c c t c a a
68501 a c c c c a g t a g a a c c a c c a c c a c t a c t a t c c a a c g a c g t c g t c a t a g t c
c t c c t c c t c a a c t t a a c a t a a g g a a g c a a c a a a a a c a a a a a t c c c a
t a c g a a a c a a t t t a t c a a a a t a t g c a g g c a a t c t a c t t t g t t t a t c c a a a
t c c a t c a a a c a a a g c a a a a g c a a c a c c a a t t t c a c a t c t t t a c a a a a a a
a t t c a g a a a c c c a c a a g c t g a a a t c c c t a a t t t a t c a a a t t t c a c c a t
t a t t g a t c a c t a c c c a c a a a g c t t c a a a c a g a t c g t t g c a a t c t g t a
g t c g a c g t a c t g g c t t t t t g g a g g c g g t g a c t a g a t c g g g a a t g g c a g t
c g t t t a g t a g t g a g g c c t g t g a g c t c c a a t c a a t t a g g a g c a a t g g g t c g
g a g g t a t g a t c g g a t c g a g a a g a g g a g g t t t c a a g a g a t g c a g c c t t g g
g t t c a c c g a t t c a a a g t t t t c a a c a g c g g c a a c a a t g g t a g c a t t t t g c
69001 a t g g g c g a g c t t c c a a t g c a c g a g g t c g g a g a g c t c g a c t c g a g t t c a t g
g t g g t g g c t a a g a a g a g a g t t t t a t t t t t c t t t t c a t t a a t t c t t t t t t a
a a a a a t c t t a t t t t a t t a a t t t a a t t a g t t t t t t a a t c c g g t t g g g t
t a g g g a g t g g g t c a t c c g t c t a c c a t g t c a g c a t t t a a c a g a a a t t t g a
c a a a a a g t t g a t g g a g g t t g a c a t t a c a c g a t a a g t t g a t g t a a a t
a t t g c a a t g t t t a a a a g a t a a g g c a t g a g a t t t a t a t g a g a c c a a a g t t
g a g g t g g t a a a a t g t a a t t t a c c a a a g a a a t t t c t c t t t t t g t a a t c c
t a c g a t t t t a a g g a g g a t t g g a a t a g a t a a g t t t t c t t c t t a a g t a a t t c
a t c t c t t a c t c c a t g a t t g t c a c a a a t t t t c a t t t c t t t c t t c a a c a t a
t c t t g a c a a t a g t t t t t t a c c t a t t c a a a t t c a a t c t c c t a t a t c a a a t g
69501 a g g c t a a g t a t c a t t c t t t a c c t t c a a c a t t a a a g t a a t t g c a g c a t
t g t a a g c t t c a c a t g c a t g t a a t t t g t t g t g a t g t a g t t t t t c t c a a
t a g t g g a a a t a g t g g g g c a a g a t t a g c t c t t c a a c a a t a g t c g c a t g c a
t t a c a t a a c g a a c t g a c a a t g c a a t a g g a a a t t a g t a t a c a a a t g a a a a
g a c a a a t a a a t t c t t a t c c t a t t c t a g a a a c c c t a c a c a a c c a t a t
a a a a c a c t c t a a c a t a g a a c t g a a c c a a g t t g t t a g c t t t t c c a a t a a a
t t t t a c a g a g t t t a a t t a g t t a g t a c g t a c c t t t c a a a t a g t a t g c a c c a
g c a a a a a g c a t g a g c g a t g c c g c g a g t t t a c a a c t c t a a t t c c g c t t a a
t t a c a a c a a t g t t t a t a a t t t a t t c c a a t t g t a c t c g t g t g t a c c a t a t
c c c t c t a c a a t a c c c c a c c a a a g g g g t t c a t g a t t c a t t g t g a c g c
70001 a t g c t t t a a a t t c a g g a a c t t t t a t t c c t t t t t t g t t c t t t a a t t a t t a a
a t t t a a a t c c t t g a g c g a t a c a t a t c t t a t a t t a a g a a g a c a t c a t a c a

ataacaatgaatgttttacactgtgcaagtggagcagctttcatttgcca
atctctcgtgcgataatgcaagctccattggtgaattaaaaaaaaagtaga
taaaattgaacccaaaaggcaaaagccttctctaacaatagatacatat
aaacagtcggttgggatatagatatgggtgtacctagaagaatagagatg
cactgagtcgagagagcaagaatgatagtggaatgttgacgaatact
ggacgatgacgaactgcacggctagagctgttgggtggcgtgattctgca
gcaaaagatagttagaaaaagattcaataaatcttacagaactgtaat
aattcagtcgatgagcttcagatcattggcatttcaataataggattcat
70501 aatttcaaattttaatatagaattaacagtttcaaaaattagaggtagct
gatggcatagcgtaccggcttcaagtatgctaaagttagacataaatcttc
agtatgtagaggtgtttctgtttctgaatttccctgcaacgggttcacag
ctggtttctcaaaatgttgaaaacgggagctgttagtatggatgtgtgt
aaatctctcgtaacaataggacgtataagaaaacgaagtgaaggtgca
tatataattcatgtttccctctctttttcatctactcaatcagatatatg
cgagtactaaattagtgcatttaaattaaaaagagttgacagaggcataa
ttggagaaaagaaaggtctttatatcaataaaaatcaaaagatgctttat
aattggacatagaaaatgaggaaaatgataggacgaaacaaagaagggc
cattaaaattttgacagagggcatttctcagcaagtaaaaaacagagggc
71001 tcccaaacgacaatacaaatattggaaaataaactgaggaggaaaaat
aaaaacaaaaaaataaaaaatgttaaactcggtagcaacatagaatg
cataagaattaagcatttgttttgagggttgagttgttttctaattt
gaagcaaaataacttgaatgaaatgcaaaataaaaaccccaaaacgatg
attaaaatgatattgagaaacataatacaaaaattatctgatattt
aaattgatgtgacctatctgatgtttattcagttcaagtatcaattcaaat
atcaaatagttgcttcgaatattagaaagattatgcaacctgcaattcaaa
tatctgaaattggcctgcttttgcaaaagtgcccaaaattcaggggttac
tcaatctcatgtttgaaagccatttctgtaattgaaaatccaaatcaga
ttcctttccaaagtctcaacaacactataactgaccacaaatccaaaaaa
71501 ccagttcaacccccactttgaccattttcactttcaattcgcatatca
tcactaaataaaaaccatccaagtcaaaacatattaagaacatagggtggag
agctgaatctaaggttctggaatcacatatgtattcccaaatcagtgga
aaacaatgaagctatgattacaaaaggctatgctaatttttgataaaca
cattgtatcaaatgttagatttattaaccaaatgtaaaaggagatggtag
agatcaccgacatctgtatcaaaaggttagttttattaacgaaattatta
gtacaattgttgcatttaatatgtatctaaatatccaccttgattta
gtgaaagtgatgttgggtggcgttggagaggggacgaagaacaaagaa
aaatcaaaacagaaaaatacacaaactaattcatcaatcaataccatctcc
ctcatgcgaaaaaacccagagtcacaaaaactacagaccaataacagttt
72001 gatctgattacaacaattagatggccaaatggaagaccacaaatctctct
accgcaaacgctcatatgcaaccacactccttacactccctgtctctttc
cctggccctgcttcaccaaagaccctgtttccatataccttgcataat
ctttcttggaaagcattacattttctttttgaaagccttctaccttgaggc
ctaattggaaaacgggtcgacctcgacgctgttcaatcgggggctgttc
aagcactctctgcacatcttcatcacaatcatgagcagcagatgatcg
gtgtagagtgtagatggacgttgttcatgaaaactctgagcagacagac
acaattttgaatttaggacaatctttgataatattccaacattcaaacca
ggatgaatgtttgcttttgccttttgccttttggcaccataccaagcttg
72501 ctgaaagtctacacaatcggggaggataaataattataatgagaatag
gtaacaaataaataatacaaaacaaataattataatgaaagttagtaacaa
ataaataatacaaaacaaataattataatgaaagttagtaacaaataatta
taattgaaagttagtaacaaataataatacaaaacaaataaataatatt
gttatctgatccgctaattttcccaacttcaaatattaccactagctgg
tgccaaggtgtctctccacgtactaaacacttggctaagtaatttccaac
gactggacatcgattctttagttctttgccaccattttctcaagataa
ttggtatgaataagactccacatttctcgcaactgcatctcattaccgct
aagcgaatcatgagtaacttgaaccagctggtatacaacgcaacatctt
caaaaagtgtccaattcgtacctgcttgagtagtcattttgttggaaaaa
aaaattggattcaaaccttgagagaaataggaatgtggttgaagat
73001 ttgagaaaaatgaaattgtggtgtaaggtagaagataatgagaaagtat
ttatagaaaagtaaaatcaaatttttttataatttttcataattttttt
ttcggattttaagaattttttacatttttttaattttttattcaccta
attaatctctaccatttcagaattttttaactgtttattccttttttttt
aatttcaaaagcctaataacgtggaccgttgatcttagatcaaacagatg
aaattaaatgaaaattttaaatgtttttatcgttggaaatccaacggc
ccataaaataggccattgcaatcaaacagacatggggaagccacgtggc
ttcccaacggtaacttttcaaaacagggctcgggggccacgggctgtt
ttttgttggaaagcgtgcgctggaagcacacggccttgacgcagccctc
gggctctcgggctggcaatccacgcgggctgcttctcgggctcgct

73501 gt cactcgagctt cctaacgctggagcctatcccccaacaatttgctcc
 tgtttgaaagcctggccccgagcaacctcctcgcgctggagttgctttaa
 acggtataggtatattctcacctctaccttttagggacagctccagagcg
 atcaaatccgtaggagtggaatctaatgcatcattcattctccctca
 aaaaccctcagccagtcagggcatggctccaagggagcccaattct
 tcaatcaggcgataaattcaatagcctaacgcctggcctatggtacccaag
 gttaatgtcatcgtgacgtcagccacaatttaaatattaaaaattataa
 aaaaaatataatcaaaaataaaaataaaaattatataaatatctaatac
 atggtcgttcaactttacaacatttcaatatttttaatactttcaacca
 tcttttatttatccgtaattaaaaataaagttatctttaattattttat
 74001 aaaaaatatttaataccaatagtctgagctattcattttaaaagtaaaa
 atacacttagataaactggttatctcaataaattgagttgctctatagc
 ctgagggcatttttaggtggaaatgctcagtcagtcaaatgctcagtcacatt
 gattggctcatggaacaaacagttatgataaataattgaggttcgctgacc
 atgtggtaagcggggcatttcttaatcgatatccagatccagatggtcc
 ctcccctctccagaaatacaaaaattatgtagaaaaattaatgctcca
 cgttgatgagtacacttttcatcaatcctcagcctcactctataaatac
 atacctctccttcaacttctcactatctcaaatctcccattacatcttc
 ctctcatcaatttttttttttctattttccaaacactttataaaatc
 atgggtggtttcacatacgaatccgagaccacctccgtcatccccctgc
 74501 taggttggtcaatgagactgctcttgatggatgaactcatcgccaagc
 ttgcaccacaggcagtgagagcattgagatccttgaaggagatgggga
 gttggaactggtcagaagatcatctcggtagaggtagcacaatggcta
 cgtgaaaaaagaattgattgattgacaaggacaactttgtgtacaagt
 acagtatgtaggggagatgctatctcagagacaattgagaagatctct
 tatgagactacggtgggtgctctcggcagcgggtccatcatcaagcgcac
 atgccactatcacactaagggtgatgtgaaattaatgaagagcatctca
 aggctagcaaaagagaagtctcccacctcttgaagttggttgagaactac
 ctcttgagaccaggatgcctacaactaaattctctctcgccttttcaa
 tgattttctgaagtgttgatcacaactttgtgtcctcttaggcatcagtg
 75001 aattgtgtggttttgccttgatattccttacatttcaaatgaata
 aaaaataaatcatgggggttgatgataatattgcatgttcttttttagg
 gaaaacttgatataagacaatttttttagtcaatggtgcaatttattaaa
 ataagtccaattccttaatttaattatttaattatcaataattatctat
 tcaatgataagatttattataaaaatcaaatatttcttaattatctctc
 tgattatttctttttattagttttttgttatttctgttcttctcttg
 ctttgaacccatttatagatttaattttttatttttataatcaacctat
 atcacaggttaacaggactcactcataaaaatcatccatgttttataac
 aaggagtataaaagaaaataaaaacatgtaattagataactaaactcact
 ataaaaatcatccatgttttgcactttgatctaaaagccccttctttt
 75501 atgcaccttggaatgaacttagagcttcattttataactatagtgacc
 caacttataatgtaattcatattaatcacctattatttttttaagagca
 aatatcaatcacctaaatatttataaaacttgataatctcagtgga
 taggctgttgattttcccatgctcgaatttaaaatttatcttctt
 taaattatgtaatggtttgaacgatgtgcatgccataatagtaaat
 taaaaataaataaataaactcatattgtcaccatgcatccctattca
 aaacttctcccttttaccattatgtaataattttgaacgcttctttattt
 ttaatagtaaatatttttttaagtaattcatataataaaaataattatac
 atctactctatgataccgatcaattttaaagcattacattttctagttg
 tatatgagttaccaatataaacacgtcaattagtaaatgattgtaattaa
 76001 tattatattgataacgtgacaagatgccaatctacctaactccttcata
 tttagccggcactaattaaaaatgtgacaattatcatttcatttttttat
 tttttaaaatgagacaactggtggcaaacacagtttaatttttttttta
 tccaaaataagctatgaagaatttaccgctgctcatgaccacaatcaagca
 gtcaaatatcatttcatttggcctagtacttcatattagccggccactaa
 ttaattgttaaagtctttaaattgataatgtgacaaatcatatatac
 attggacctacttcatattagccggccacaaagcatacgcgggttac
 cacttcaaagataatagtaaggatataatcagcttataattatattgga
 tgataaatttaagatacaaaagaaattttccagttacataatttgaaataa
 ttttaatttaatttctagaatttgtgaattttttttgttccgtaataaaa
 76501 gaacaataaaaatttaatacagaatttaatgagtataaaagtcattaat
 agataaatgagataaagttattagtaaatgtgttagcagtaatggcaacg
 acgtgggtgcaatagtggtagcagtagtgatgttgtgcaagtggtggctgt
 ggtggtgacgtacgctggtatgatgatggtgggtggtggtggcaacaata
 gtgatgatgataaatggtggtggtgacggcaagaacagtgacattggtat
 gtggtagtgacatggtggtggtggcgcctttgggttaagttatatttttt
 gttttatgtactatcaaaattgaaattttcaaccatttacgtgaaatt

Mal d 1.08 ↓

MGVFTYESETTSVIPPARLFNATALDGDLEI
 AKLAPQAVKSIEILEGDDGGVGTQKIFG
 EGSTNGYVKRIDVIDKDNFVYKYSMIEGDA
 ISETIEKISYETTLVASGSGSIKRTCHY
 HTKGDVEINEEHLKASKEKSSHLKLVENYL
 LEHQDAYN

taaacttgagaattgaaggtctaagatttaaaaaatttcttcttcgtgtg
 77001 aaaaatctaatttctatgtatgtaaattcaaaataagagaatttatgcat
 gtaatatataaaattctaatttagtcgaaacttcaaaattatttccct
 ttattcaaacgtaattgtattgaccagagctaactaaatgagttgaacca
 actgctataagacaaaaggctgcaattccaacagtgattccctatctcac
 acaaaatgaaacgactattttatagcatcatatctacaattcacaagtga
 ttctatctcacacaagacaaaacgatgattttaagcatcacacttt
 ccaactagatgacctgacctacacggcgaggcccatatcgtaaccta
 aggatcataacacaaatcaatctcatctaacatgcaattggagaatcca
 tttcgttttcttgtaaacacttagtaattcacaccataaatacactca
 tgagtgaaagaattttgaaacctcaactgatgctcattgacaagtaa
 ctcaagcttgcgatcactaaagctcaatttgggttttgagattgacat
 77501 aattcttcattttgatctctgagcgtcaaaccaatataaatgatccatg
 agcttgccaccgtcaatcgttgtagtcattctatgaaaacttcagtaa
 atcattttgtctcattgtttatataattgagagtttttatcattttat
 gaagaacaaaatgattgaacgatgtacaaaatcaaggactacttgatc
 gatttcaatttcaaacaccaaagtaaggagttgtatcaatctcacggac
 tattttagctaaaatgcgtttataaaaatcaaggagcaatcgtccgtcaat
 gggccggtcaaggcccacataaattggactggtcatgccaccctatgtg
 cccttcaaacgcatcccagggtagattcgcacaagtcagttcgtttgag
 tttacacttgaccggggcaggcacgtcaatttcaaagaggaaaggacag
 tgtagtaatttcacagacatcgtatatacagagttctctagaacgaacc
 78001 ccacaagaacacgcagaagctgttctcgactcttcgtttcttcggtt
 cccaaaaacaaaatcccttcagctttccgtctgctctgaaacagttcca
 ccttgctactgaatccggttttctccaatgtcaggcgtttcgattgctc
 ggattcggggagagaatcgttctacaacccaccagctgtacgacaacgg
 cagcagcagcagcagcaaaaagctacaggaagagaagcatcagaggcagtc
 ggtggattcagacgactgtcgcacttcatcacagttctggttcgggtcggg
 aagtggtagcgtatcagaatttggatcgggttttggagtacaccgca
 cctgtggttggtggccagatatttccgaaggtattgcaaatcaatcttaa
 aaatttctttggatttgaaaactttatgagagattgaaattttgaagt
 ttgggcatatgctatcgctgataaattttaaatttttgtatcttttaaac
 78501 ttggatggaagttttcaactttaaatttcttgtttgggtccgggtgcg
 ttgttttgtaatgtgtaattgtgtagaagtggttcttatgtgttcca
 tttgggtgatggatagacaagtgtgaagggtggaggactcgtgagtcaga
 attacatccatactttgtgcttggagatttatgggaatccttcaaggagt
 ggagcgttacggtgcggtgttctcttttgttgaatgggagtgactct
 gttactcagtaactatgttcttacttgtccgctattcagctctatgtaga
 cccatcaaaagccatcaacgagaaacaggtgagcaatatacaagattg
 acactttcaatcaattcaattatcttttaattgttctgctcatttctga
 ctatgggttagacatgaatgtcattacgtgggaaactgtcggagcaaa
 atttgaatttagtgaggtagattctgctgttcaattttatgtgatcaatt
 79001 tatttgaatcattagttacctgataataagacagatctgtaaagtatg
 atggtaagttaatctcatccactattgcagatacaacttggttgattggg
 ttacaattttatctactatagttatggcttagaggtcgcacttgggtgc
 gatggcaagtgccttcgcccagagcggtaggtctcgggttcgagacttg
 ggagcagcctcttcataaatggggtaaggctagccgacattcacctctc
 ccagaccagcgtaaagcgggagccttgtgcaactgggtacgacctttta
 tagttatggcttagagatatacagggataggaaatgtcgtcttatcaatct
 atgttcagaaaatggctgtggtagttagacagtgaggaaagatatattgga
 tgataggtgtgaaagacaaaataggtagtggagtgctgttggtttgaga
 cagcagtgagacatctataaacaatgggtgacaagatctttgatataat
 79501 tcatggtagcgtacttggataggtgttacttgggtgttgaggtgagacgt
 ctgatggtttatggcttaactcaactgaattggcaaaacattgcacag
 aattacatgctaattggctagtttgaactgtatctcccttcttttagtgc
 aacatgctagtactggagccaaagaattacctatctttatcatttgtgag
 ataaattgtgcagagcctccgtgcacactaatagctagttttgaatgtac
 tttcctttagtgaacacaaaatcaattttagttagttgtgagttaaat
 tgcgcttagaggtacacaaaatcaatgttcaactcagatgaaatcaatctg
 ttggcctagtaagtgttcttgtcacataaatttccacatgtgtgcatgtt
 gttacaggaggcctgggtgatgagagtgatgcagagtcacgagggagaca
 agtagcaacggcagcagtgattatggaacagaaaggatttaacaatgt
 80001 tccttatagtgctttaaagttggcagaatgctgcagatgtaaatggccatg
 gttggaataaaaacctcccagaaaataaacctttaaattgggtcctcaagt
 gatgaaagtggtaggcttgaaccctagtcagcttatatttcattat
 ggagcatgatcaaccatttggctgtgaaccttggctgataaggcaagtc
 acgcatcatctcagaacacttaatacgttgagtttccattttcatttat
 aagtttgatggtaacatgcactcattgctacctccaacagatttcagtt

Hypothetical protein with a
 conserved domain with unknown
 function: DUF789
 (ORF47) ↓

cttgcacatcaattccagagctgaggacatacaggagctgtgattatc
accttctagttggcttctgtagcctgttatgctttatcttactcctcgt
aaacacaggattgatgcatatgttatcataaatctaaaattgtttctgtg
80501 tatattcttgttaaggtatccgatataataggatacctacagggtccaaca
ttgcagagcctagatgcatgcttcttgacctttcattccctatcgactcc
caciaagggtacatctctatctctatgtacatagtattatattttgtg
tacgttcatgacttgcctgttttgcaaatcacgttataatataatcatgt
tgtcgggagcagggtgcaagcgtgatgaaactcagtggtggtgacgca
gctagagacaacaaaatgccgatttcaactatcattgccaatctttgg
gcttgcctcgtacaagttcaaggttccctttggaatctaaggagttta
caatgcccagaaggtgatctctattgagcagctgacaactggctca
ggcgattgcaagttcaatcatcctgatttcaatttctttgtgctgcagc
attccaaagggtgagaagggaatttatctatgatcagctcgagttatc
81001 atcccagtgtaaaagggtataaacgtactgatgacaaaactctattcac
tggccagtttttcatgttatgggtggatcagttcgactcctgtcgt
tatattacatagacttggccttgatcttatattatcatcaaagcaga
atgaagagtgattttggctctgtgagggtttggataggctgaaagaaga
atagtgatgacattatcgttaataaaggccttttaatttggggagtgct
ttcagaatggcgctcatgtattgcatcaggcattatgattacgactctc
ttagtctatttgcgtatcgagtttggcagtttaggtggaaggaacata
tgtcagtagctgtttgacaatcttctgttactgtttagtgtgatgacgatg
aaaatataataacttataatgtttatataatctccgtatagttgaggtg
ccttttgttagtatgtttgatcacaccgtgtgtatgaaactaacgtaaa
81501 gttcaacgcaaacatgttctgtttaatgtatattctgatcctacagtaaa
ttatcagtgagattatgatgaacaagggtgggtgtgctcctgcactcatgct
cactcttctgatggcaaatgaactttaggatgtgcccggcggaggactttg
atgcaattgcatgctgttgcaacatttcaagccaatcatgcattatcat
gctgtggttttcttggaaatgtgcccacagtgacatcctagtcgctggaag
ttcctcctgtgctgacctacatctggacaggtatagggtcttgatgggt
gcacatgggagtgatttccagggtactgttatagaccagcattcctaact
ccaattgaaatcctgggtgagtttcttacctccaatctgatcaaggcta
agcccgaatcggaaagcagcttgccttcccatttcaagcatgttatt
agtgtgcaatttatgcaaatatgtgttatttagtgaaaacaacgactcac
82001 gatagtagtacttccatggcagagtagctatcagtgctctacgttga
ggttgtgcagactgctgaatttaccgcaacatttattcttttcttttga
aaatttgaaagatttgattacaatacttactgtattcaatataatcatc
agttcaacctcgaatgaagattttatataacataattcaaaagaatt
acataaattactagatgggattatataaatgcatatataatattgtag
ccaatcctaaaattgaaagctttaatcaacctgaaagcaaaattgctat
atctgggttggattgatagaatcatagactaaaagcttacttatgttta
tagccaaacgcccagccttgatgcttccaaaacaataaatatggcaacct
accataaacgcttgaataatggtactttatgtaccacaaacactgagtt
caagcttccaacccacaactaccacttaagttaaatccttgataaacca
82501 atggaggaaatcaatcgcaacctcagatgacatattgctgctgcaacgta
gttttgcaaatcacttgaatgtcgatgagatttacttgcagcaaaataa
tttctagcaagtggtcgttttacattctcttttgcctcgcatgcacttat
ctttctttttttatattatgagaattaaataaatcaaaagaattaat
gacaaaaaaccttccggtaagattttcacaattttgttactctatttg
cacattgactcagatccccctcccctcacaccaacacaaaaaatctac
gaaactctattatgtatgaaaaaatgtgcaactatccattgtgacat
cacaccaacgcaaaaaatctacgaaaccttctgtatgaaaaaatatgc
aaactatcccttgtgacatcaaggcaccgcttccaaaggcgtcccttgg
tactagaattcgattgaaatatacgttacgttcaaacatgatcaatgaa
83001 gaaatgaaaaaatcttgcctcaattgaaaataaaaaataaattactcat
taaaaaacttattcgctctaatccttaaaatgaaaagatcagaaggaaa
atthttatcatgtcaaactccttccaaacgaaaaatcactcaccctcc
ccactcctcagtgactttccaactcaaatccattcttagacaccaat
ttataagctatcaactcgaatctattcttagacacaaacgagcctgaag
tgcccctatgcatatctaacctaaagtgtcactattccatttgattttg
tatcccgacttgggatcagtgcaaatatccctgacaaccttgggcaa
gttatcagaacatacagaagatagcaaacacagaagtgataaattta
ttccatattttcattatgacataaagggttgagtagattacaaatagat
agatgttctgatcatctaactattaccgcttgttacataaatagcaccac
83501 gttcttccctctcgtattacttgggtctgtgctctggtcctattaactta
tcatttccacatctcccactacactattaactcataaatcaaaaaata
gtcgacatcagcagtccttgaacaaactcacacctccgcttgcctcctc
ttggctccttgcgagataacgggtggctgttgaacgcttctgttgaacat
agaccagtgtagacctgaagaaatcatagtaattcaccgacacaaac

ggaaacacgtaaataagcacgcatgatttctgtgtgacatccagacct
caacacaatcccttaatactatccctacttactgtgtgtctatattccg
aaacaaaaaacaagatagaggactctcatttattactcaggcacaaaaac
ataaaaaaaaagataaaagataaaagtgtaaagtggatgatgttcat
accttccgagacctcttcccttgtgctctccttcacaaaagatggttca
actaaaggatgggtcttctgtactgacatcctcatcaaccgtgggagggt
84001 atataaaaggaagacaacaaggatcagcttcagtgttgaggaaagga
acgaagaccctcctagtagagattgtaaaactccttcccgcagttcttcc
cttaacatagaaaatgtgacccagatccctgcgaagttcctcttgcg
ctgaacgtatgtatggtcaaggagcacatgacccttacttaggacttaa
aaatattgacaaaactcaaaataaacatgtaaagcaaggtatcaaagttt
ttatgcaagaaaaataccatgccagcagattagagtgactgatgaaata
caggaagtgtcaacaaaaggacaaggatcaggtgacaatcgatcacatt
taataactagtctgcttgcggccagacacattaagatatgtgaaagtga
tttagatacttgaacataattttggtaaatgaggaaattttaaggat
aaaggttttatgattcctgtcacgccgttattagatggctttaaccaac
84501 tactaaacgataaaaaactcctccttgcattaccaactaattttatggca
caatggaatgcatgtaaatcataaccaaaccaaacatagaagtcacaac
aaaaggatctttaaacaactcccatgttgcgtagtaatatggctcaca
aggtttgcctcccaccctttcgacaaaactggacagacctgccaatgga
aaggaatgaacatttagttccataaaaaatccacgagagaaggacatcca
attgataacaatgggcaagatcataaacaatccaagaggaacattactc
cattctattaataaaaacaacaattagacctgaaagattatatactgta
tatcaaaaagaaaataactacaagcttggctcaaaccatcataaagac
cttcaaccaacacacacaaaactagcaagaagaataacaatattgtcggt
ctattaaggaacattgtcggtcacatcatatatactacatgaaccac
85001 ataagcagtcagacaaaactcactaaaatgaccccccaagatggccact
aatgtgcaactcaaccctactcaacaagcccttgcacaagcagagcca
aaccccaagagaaaaccaagttaaactgctagcatccttttgcagactcg
acaatgatacatatctctcccagaagttcactcactctctttgtaaatt
acaaaactagtgaacaatcacaacaaaaactgctgcttctaaagcaca
aacgaaaactagactccaacaaaacgcaatcattaaaatttataagggga
aaaacaacggaaagctcaacagttcgtatctccacacatttagaacaacc
actttccgaattaagctagaactaattaattacactgtaaaacttcaaa
ttaacaaaaaaaataagtgaaatagaacaaaataccccgttcttagcctc
cagcggatgctcctcatcgatgtggcagcaaaagcccaaggacatcaaat
85501 cctcggcgcaaaaaggacacaaaaactcggcctttaaatacatcgtccccg
tcagtttctcgtgcgaaaacagatctgaccaaaccaaccgcaatcaaca
cccccaaaaaccaaccataaaaaaaatcaagaaaatcaacgacaaaaca
tgaacactgccaggaacagttcggcaaacctaccggatcgcgacagatac
cgcctcgacgaggggggtcgagaagagactgctccacgaattatcagagtc
catttcttatgcagtgcaaaatcaaatcacggcgacggcgaattcggagc
tccaatcgaaatttgggggtccgaatttgtgaaatttgaaccctagaaaa
aaatggagacggtagataaaagcaggtgtggtgtgagagtgtgaggaagc
gaccaagaaattgcgagtgaggagagagagcagcagcaggaactctcta
ccaaacttccggagcgggttaaaccgcctttatagtagccacgtgcccg
86001 aatgcccctcgtacattccgggtgaatctcgggggatgccactatctacct
gagccgtgcgatttttctacgcatgtggatttcattggacggatgacgtg
tggtagacttataggttttatctcttttgggggttttctcaattgagatt
ggacagattttagttcttctaataaaattacaagatattaatgaaaattt
taataattttccaaataaatgtgtatttagtcattattttaagataaat
ttattaccatcttcaaaaaacaataataactattottactgtgtaattgtac
cccaattatataccatcttaggtggcatalgatgtagttgtccacgtcat
ttcaattaatcagaacttttattcaattaaaatcaatttaaaaaaaaaaaa
aaaaaaaaaaagctgcttaattatgactacgaggagtttggttccacga
ctcctctgggttttacgaggagctgcttcaagtttgcctccttccctcc
86501 ccttccctctcatgccaattcttcaccaccacaatctaatacaacttgggt
tttttaaatgtaggaaattttcatgtgcaattttccagctgctgacacc
cttgaattcaacaaaaacttccagaattgatccgaatgttgaacttac
tacaactttaaatacaaaaattccacttggtttaaaattttatattac
aacagacccaattattttttgaatctcgggaacaatttgaccgaatgt
aagcatgggtactcagattttagaacttgggtagtgattgcataagctaa
ttaaagttatagctttaaagtgaaagattcatgatttttgtgcacaata
caatttcagtagttgttagtgcaataacaaaattctcgacattgcaatg
gaggcgtagaaatataatgaaatagcttaattcttctagctaaacaa
cactaagcaatataatgaacaaaaatagagagagagagagagagact
87001 gatggagtttctcgggggttctcaggtggatcacaattgaacggaggag
gacgagaggggtcaagatcagaggaggagcacaaccagatagcccatcc

gtcgttgaggctccgaggagtttgctttcaagtcttctccttccctccc
ttctcttcatgacttcccttctccttctcttggattgttttttaattaa
tcatttaattgttagttttttaatttaattgaataaacgatta
atthaatttagtagatgttcacattagatgccacctaagatgatata
attgcagtacaattacatggttaagaatagcattactcttttaaaaatctg
gttgatthaaacatgattgtaataagtttattacattttaggatatt
taatgaacttttatttttaataatttttaaaaagttgaggaatttgagggt
87501 attaaattttttttatattgtatttcaagcattcacaaaagctctcatct
ttctgagatttaaaaagaattgaatcaaagttttatataaaaatgtctaca
agacagtttagctctgttaaaattgaattttataaaagtattatataaaa
tccatcaatattcaagaatttatatgactctattacatgaaagctctaca
agccacttggttgattttgtgacacgacgcggttttagctttccatgac
tcgccttcgctcacacgttatggaattttgatacaagtaatctatcagag
aaaattggaactaaaatgtcgacatccgtttgttagcagacaggacaaa
taatgaattcaaatattagtcctagttgtcatttaatatatagctagt
aatgtaattgttactcatttgaagttagatgttttaatttgactct
catcaaaagcacatttcaaccacatcatttttaactcattgtgagactta
88001 actagcattggttaattcctgatacaccgatacgcatacaaacgaaatt
aacaggtgttcggatcgacacgataacgaaatcggttcattatcgggta
ccaataagcatctattaaaataatgtgttggttcaggatatacgtgggt
aacacgatacagtttaagcaaaatattaattttataattttataacccta
aaaaactataataatttatataatataatataaacttttatacccta
aaaactataataatttatataatataatataatataatataatataat
aaaaattggtgaccattggatcggtttaaataacatatacattcaatttc
ttaagtgtatacatacaaaaataaataagtttaaatctacttaataata
aatatataatataatcagtgaaacttgatgggatacgaattatgtgaaa
ctaatttttaacgatccaaccgtcaaactgtttgtatagcttcgagatc
88501 gcatcagcaaaaaatcgcaaaaaacaaacattcagatatacaagtaacgag
acaaaacttttcgacggttatcaacgaaaaatcatgatttaattggttatt
ttaactccgattttgatgattttttacagctacactccttgaccctatat
gaatacaatgaataaaactcgatcttcaatttaaaatattaacactagttt
atacttaattgaaagtagaataaattccttaggtgtagtgaatctattgt
tttgatgggatacgcattctatgaaactagtttcaacgatccaattgtca
aacatgtttatataacttcgagatcgcataggcaaaaaattgcaaaaa
caaacattcagagatcagtaaacaggacaaaacttttcgacggttataatt
gaaaaatcatgatttaacggttatttttaactctgattttgatgattttt
ttttacagctacactccttgaccctatatgaatacaatgaattaattcaa
89001 tcttcaatttaaaatattacactagtgataccacaaaactttatgtta
tacttaattgaaagtataaataaaactctaagtgtagtgaatctatcgttt
tgatgggatacgcattctacgaaactaatttcaatgatccaaccgtcaaa
tttattttgcatatgcttcaagatcgcatacgcaaaaaatcacaaaaata
aacatttagtgatcaagtaaacaggacaaaacttttcaacggttatcaacg
aaaaatcacgatttaacggttatttttaactctgattttgatgctttttta
cagctacactccttgaccctatatgaatacaatgaatgaatttgatcttc
aatttaaaatatttactactagttgataccacaaaactttatggtataact
aataaaagtataaataaattcctaagtgtagtgaatctatttttgat
aggatcagcattttacgaaactagtttcaacgatccaatcgtcaaacatg
89501 tttgtatatacttcgagattgtatgcgcaaaaaattgcaaaaaacaaaca
ttcagagatcaagtaacgtgacaaaacttttcgacagttatcaacgaaaa
atcacgatttaacggttatttttaactccgattttgatgatttttttacag
ctacactccttgaccctatatgaatacaataaacgaattcgatcttcaat
ttaaaatatttactactagtagataccacaaaactttatggtataactta
gaaagtataaataaaactttaagtgtagtgaatctatcgttttgatggga
tacgcattctacaaaactaatttcaacgatccaatcgtcaaaccttggttg
tatatgcttcaagattgcataactcaaaaaatcgcaaaaaacaaacattca
aagatcaagtaacgggacaaaacttttcgacggttatcaacgaaaaatca
cgattttaacgattatttttaactccgattttgatgatttttttacagctaca
90001 ctcttgaccctatatgaatacaatgaatgaattcgatcttcaatttaaa
atatttacacttggtgataccacaaaactttatgcatgatgatcgatga
aaattgaagattttgtaaaatgaataacatgcaagctttgagttccattt
gcaatgtttaaacttttgagaaaggcttcacaaccttgaaaacaaaaact
ctttcattttgcataatccttgacggttataatattttgtaatagactagta
gtgtatggacggtttgtttatagctattattttcgagtgtagatgtagt
acttaaacgacaaatgtgttttaattgtttgaagtaataattatgtggta
atgtgtggtatgtgcaatttaagaaaataaaaatttcttaacgggtcat
aatgggtcgggtcactttaccgtaaggaccctttaagataacaagtgtag
acacgacatgaccggttttaataacaaggttacgaaaaatgacaaacac
aaccggtttgcccaggcctagacttagccattctctcactctgtttggc

90501 aatggtaattgtatTTTTGGATGAGATATTAactTTtaggcttaattgtaag
aatagTTTTCGAACCTTATTCCTAGTTTAAATTTATATCACGAACTCTTA
TATTAGTCTCTTAGTCTTCAAactCAACAATGTATTTACAATTAGTcat
TTTCGCTAACGTCATTTAGAGGTATATAGAACTTCATACCAACATTA
AAATCGCATTGCCAATATTAACAGTTGCTAAAGTCTAACCATTTTCATT
TTGCATATTTAAAGTTCAAATTTCAATTTGTTTTAAGTTTTAGACATTA
TTATTACTTAATACTTTGGTCTAGTGGTATCTTTTTAACTTATAAGTGG
GAGGTCTTAGATTCGGTCTCACAAAAACGAATTTGAATCACATTATTA
TGGCTGACCATTTGTGAGCCTTAGCTTAGTCCCCACCCTTATTGAAA
TAATATCGTTTGTAAAAAGAAAATGTTTTAGATTAATTTATAAAAT

91001 TTTATTTGTTGAAAATCGATGGTGGGTGCAGTGGGAATAAATAACAAGA
CACAAAATTAACAAGTTCTCTACAGTcattgggactgaagtacgtctt
cggggcagcagtggtgctttcattataatttagaataataggggtacata
AATAACTCTCTATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
TCTTCCGtGAGTCTCTCACTTCTTCTCTCTCTTCTCACATGCTTTTATAG
GAAAAATTAactGTAGCAACACTATTCactTTCAAAAATTTCCACAAATG
AATAGTGAAAATACTGTGGATAAATAGTAACAAATTTATCATGTGAGCCA
TTACATATTAATTCACAACACTCCCCTGGATGGCCACAAGTCTTCGATT
GACTCGTAAAAATCTTGATCTGTTTTGGATGCTCCCTCTGATGAGTATCT
TCCCCTGATCTAAATCATTtagTTTgatcattaatcattctgcttcagt
CTCTTAGCAAAAAGAGTTACCGAAAAGTGGTTACTGTTGTTAACTT
TCCACAGTTTGTCTTGAactGGGCTAGAGGGCTTCTGCTAGACTTGC
ATCAAAGTCTGAATTTACTTCTTTTATCTGGAGCGGAATTTGATTCAAG
GGTGGTGGCACTTGATCTTGAATCGGACTTGTGATTTCTCAAGGACTT
CGAGCTTGATCTTGAATGAATTAGACAATGATGAGCTTCAAGCGGCAA
GTGATCTTTGGCTTGTCAAGGATGAATCAGCACGTGTTGTTGCGCTTG
TCTCCAAATGCTTCAATGTATATTTTCACTTGCTTACTGTTCCCTT
GCATAAGTGGTATTTTCTTAGTTTCCCCCATGATGTGTGGCATCTTCT
TCTTCTGGGATTTGTCCACTGATGCAAGAAATGGAATGTAACCCTTGCAAT
TGGATGGTTCATCATTCTTAGTGACTGGAGACCAGCTCCAACAACAGT
TGAAGTAGTACTCGAGAGCAATACTAGGTAAGCAATCAAGAAAGATTC
CAGGAGTCGGTCTCTTACCAGAGTTTGAAGTGGAGTTTCGACATTTG
CTTCTTTATCATTGTCTTTGCAAAATAAGAATAAGGATAAAGGAAATGAC
AGGGAGATTGCATGATATGGGACTCTTGCTCTCGTCCCTGATGATAG
AGATACTCTTGCTCTTATGTGACTTGTTTGAAGAGGTGTTCTCGGAGAA
GAAGAAAACGAGTATTTGAGATACTTTGTTGAAGATGCATCTCGGAG
ATGAGGAAGAGTAAGGTATCTTGATGCTACTCTGTTATAGAGAATAA
AGGTCGACATATAAAGATTTCCCAATAGTAGGTAAGTGGTGGATGATG
GGCTTTGTCAATGCTTGTGGCAACAGTGTGTAGTAGGAATAGTAA
GATTTACACGTTTTCTACTTTATCAGAGATCTTTGACAAAAGTTGCACGTG
ATATCTAAAAGCTGAGATTGCGTCTGAGAAATGTTGAAGAactTTCTCTT
GAAAATCTAGCTTTTTGAAATTTGAGAGCCGTGCCTCTCAATCTCTGAA
CAAGTGGCTCTGTTGCCTCTCTTTTATAGAGACATCAATGTATTCAAG
AGTATGCTTAGAAAGTTGCAGCCTGTAGAATTTTCTCTTTTGCactT
CTGAGATTTATTTGACCTCATTTTTCTTCACTTTCTGAAAAATGGC
TTGCCATCTAACATTTGCTTAGATTTGAGTCTTAGGACTAATACAGGCA
AGCCGCATCAGGATAATATAGGTGCCCTCTTCTTATCTTCAAATGGT
CCTCTTACGGTTGGGGACTCTCTGATGAAGAAATGACATAATCGCTACAAT
GGTAGCTAGGAATCTTCTCACTCTCAAGGATAACAGAATCTCTTCAAACC
GGTCTGATGAGGCACCCGTTCAAGACTCATTGGCTCTCAGTATTCAGTGT
93001 GCGAactCTGTTCTAATATGGGTTAACGCCTACTTGCAAAAactCATCA
AGTTGAATCATTAAATGGTTGAAGTGGCAAGTCTAAACAAGATATTAGAG
GGCTCAAGCACGAGAATATAGTGTACACATGCTCGAAAATAATTAactCA
ACAAGCATGAAAAGAAAGCTTGACTAGCTACTAGAATCTAAAGGTGGGAC
TCAAAGTGACAATCAGGGGTCAAGTCTTATTCCAAGACATTCATTGT
CTCAACTGTTTAGAGTTTCACCAAGTACTGAAGCTTCACATGAGCAACCT
TTTTGAAGGCTCCCTCATGTTTTCATATGTATGTACATGTCTGTAATTT
TCAACGATATCAATAGATAAGCTTTATTTATCAATGTATTGTGTCAA
TACAATAAGCATATACTTCACTAAAATGGTACTTCTGGACAAAATCA
ATTCATCTTCTCCCTCACGAGGATGAATGATTTTTCTTAGTGTCTAATA
93501 TTAATAACCactTGAGTGTCAactCATGATCTTCAATCCATATGATCT
TTAATATCATAAACATCATGAATAAGATTCGTGTTGGTAGATTATTCGAT
CTGCAGCTCGAAAACAATTTTTCTCAACACCTTATAATTTCTTGACT
CTTTAGGAGTCTCAATTTAATATCATCAACTCTTGCAAGAGATTTTAGTA
TGTGCAACAATACCATAATGATTGTACTCATTAAAGGCAATTTATACCAT
CCATTGGTATTAAGTACATAATTTATGTTTTACCCTTCGGGGCTTACTT
CAAGCAATTTGAATCTTCAAGGATTTTTTACACATCATGACACATTTCT
CTTTGGAATTTATACAGAGCAATTTGCTCCCATGTACTTGAGGGATTT

acttatggagatgatgatgtgtgactcacacccttctgatataattct
ccattcatatgaactcgtttacaacctgtatcatatcatgtgagtgat
94001 ttcactgtattacaaatgccaatgtaacctccttggttcaacacctttt
ggctatttgtattatattaaatataatagatccattttccacatttcta
taaaattgtaaatggaattgcctttcttcattttggccaattattttggtg
acgaaaaagaacgagactcaatatcaacacagctcattgatgaatttagt
aactacttgtaaaaacaaaattaccattggttatcatcaatctatgatc
ccgtatttctcatgtgcatgcatgataaaagctttttatttctttgta
tatccattgctcttcaagggcattacattatctcttccatgatagtcgt
at ttcagagaatgcagatcataacctcttcttaagcgcgttatagagctt
ggatttttaactctacttctgggagtttaagtcat tggaaacctatagtcct
atcatgtttcatgcatgagacatcaatattcccatgtccgtggattgtcc
94501 tttatgggacgtgtatccatgcggtgactgtcatatttctggcatgcaac
agttatgcttctggcattcatatgaactcgtttacaacctgtgttatat
catatgagtgat ttcactgtattacaaat acccaatgtaacctccttgg
ttcaacaccttttggctatttgtattatattcaaatataggtccattt
ttccacgttcttacaaaattgtaatggaattgcctttctccattttgacc
aattatttggtaacgaaaaagaacgagactcaatatcaacacagcttat
tgatgaatttagtagctacttgtaaaaacaaaattaccattggttatta
tcaatctatgatcccgatttctcatgtgcatgcatgataaaagtttt
ttatttctttgtatatccattgctcttcaagggcattacactatctctt
ccatgatagtcgtattttagagaatgcgatcataacctcttcttgagcg
95001 ttatagagcttggatttttaactctacttctgggagttgagtcattggaa
cctatagtcctatcatgtttaatgcatgagacatcaatattcctatgtct
gtggatttcttcaatctatgatcccgatttctcatgtgcatgcatattt
tggcatgcaacagttatgcttctagcattcatatgaactcgtttacaacc
ttgtgtcatatcatgtgagtgat ttcactgtattacaaatgccaatgt
aacctccttggttcaacaccttttggctatttgtattatattcaaatata
taggtccatttttccacgttcttacaaaattgtaatggaattgcctttct
ccattttgaccaattatttttttaacgaaaaagaacgggactcaatatca
acacagcttattgatgaatttagtagctacttgtaaaaacaaaattacc
attggttattatcaatctatgatcccgatttctcatgtgcatgcatg
95501 ataaaagttttttatttctttggatattccattgctcttcaagggcatta
cactatctcttccatgatagtcgtattttagagaatgcgatcataacct
cttcttgagcgttatagagcttggatttttaactctacttctgggagttt
agtcattggaacctatagtcctatcatgtttcatgcataagacatcaata
ttccatgtctgtggattgtcctttatggaacgtgtatccatgcggtgac
tgtcatgctctagcatgcaacagttatgcttccgggaatgcaatagttt
ggagtgccatattcttctcttcaaatgactaaaccttttggagtcata
ggctgcccacgcttccaggtatgcaacagagaaatcatttactgcctcact
atattgtccaacagggacatttaattcttgtaggcacatttgaacaagta
tatgtgatttcatcacttcttgcacatttaaatgcatctggcatttgat
96001 tgatataatcgttacatcattaaattattttgtcttctgatttcatggtgat
tgcttcatgaaataaaaataagacaaattctcttctgcttcttctggaacagt
cgtttctcatcattcttctggaatgatatttctcatcatcttctctgga
cgatgttattttctccccctaaacggcaggaaaattgtcttataaaa
tgatttttctcatcatatcactaaaaggtatggatgtttgatctaaaa
aattcaaaaactgagggcctatatacaaaaatccctaagctaaatcgaat
tcaagggttaaacgagactcatttagttggagaactaagactcagtatat
aagctactcagtaactatggtccaatagtat tcatcttacttgtaaatga
gaggtcttatgttcgaatctcgtggatggtgaattcgataccaaattaga
ttgcccattgatgacttaaccaacccttagtggtaaaatataatcgttga
96501 ctaaaaataaaactcagtatataaatttgggagtaaggagttgttataa
gctctctagaacatttatagaaatataagaagacaaattgagtatgacgac
tttttaaatgattttggtatttattttgagtcattcctatagattatta
acatagcttgcagattaaagttgttcatgcttcttcttctttaaagtaa
cactgcgagacagaagacttgaaactcctctgcaaaaaggtcgacgacgcca
aggttgaggtcagaagtcagtttggagatcgacgacggcaaaagttga
catccgcccattggcctgtgttttggaccctctctagggaaactgtacaac
tcaactaagtcacatcaggagagatctctggaggaacctttggcctgtcaat
agcattatctgaaccaccgcttctgtttttcttctgtttgagtgctatttt
aattaatctaatctaaactaagggaggggtttcgaacttaggtgtagag
97001 gagagcgcactgctctatccaacctgtctacacaatatgtgatcttaccg
tataatttctcaactcatgccagagccagggtgtccactgtccagatac
agccagtcagggaatttgaacgagcaggagctgaattccttgatcaa
gaagtgcaaatagaagccccatacattgtgagctgcaataaaaactctct
cctggtagttgcccgttctgatgtattaaaaagaatcggaaatcaactcc
caccatcctacacaataaattctaactcagttattgctgccaattcttctt

gatgcagaatgtaaaactgccgcaagtaacctaacgtataatgagtggt
 ccactactaattgtgtcgcacgccttttgataaaaccgctattagccgca
 gattaaaattctcaaggggaaactagaaggagggtgaaagaataacaaacc
 tccggacatctgaatgtgtcaccatgtcagttgttataccgtggatgcta
 97501 gaatttggaaacatgttatttgagggttcatcagggtttggaaagcaactgtt
 agcgtgatttcagtaacatgtcctatctttctgcaatccagtggtttcaa
 gatccagaacaaatattttcacaagaacatccaaaactccgtagtccaaa
 atcttgggtgagcatcggagtattgtatgcttctgaagcatttcaaaggt
 gatttgccttagagtagctagtgttgatcttcccaaagatgctaatacgaga
 gaaggtgattggactccaattaggtacttgaagttggaatagcacatga
 gcatgccctcatttctcctcagcagacttggaaatagcacatggagcatgcc
 ctcaatttctcctcagcagactggaaaataaatttaacaaaccataagatt
 tgtagccatttgcgaagggacttggcagttaacagagggtgaaataattcaa
 ctttctaaacaagaaccgaaataccagtggtcatttttgagttacttaaatc
 98001 aagcaaatcaatcctgatgtcttttggctaccaacaatgccacagt
 gttcatgctcaaaatgaatgtgggttaggtcagtttgacagtcatccagt
 cccttgcacccaaatcgaaatcctctgtcccagggccgtccctttgcg
 ctcgagttaaaatcccctatcccgtccctaggtcaagcttcttcttca
 agtttagtgtttggactccagcatatgaccaattttatgtattgtttt
 tagtttaatttttagaaaacttgattttaatccttggcaaagatgacctt
 ttgattaaaaccatgagtaagatcaaaatctaatttttagtcccttgatac
 attaataacctcatttataattatattttgattttttaattattatct
 gtttcttctaatttttaatgtattaattttatttcattataatttttat
 tttcatttcattcctgtaaatatattttgttggtaacatttagataaact
 98501 aattttgttatacaaatatattttgatgtactttatcttctcatttagg
 tatattaattcattataatacattttgggtgcatacgtttgggtacacac
 atttcgggtacataatctcgatacaaatattaggtacattaattcaata
 taatacatttttgggtactaacatttcggtacacacatttaggtacattaa
 ttaataatataacatttcggtgcataatatttaggtacatacatttcggt
 actaacatttaggtacattaattgagtccttcaaatttgaagaatgttaa
 ataaaatataaaatataaaatccttttgggtcaaaaaataaattaaa
 atttgtgcattataaaatataaaatatttaaggggtgacataacattaat
 tttttgaattaaaggacacattagggactaaaatcaatatttaactaac
 accaggttttttttaatttttaactcttctgaaggattaaagttcaaac
 99001 ccctttaatttttctgttataaaaaataagcacataccattacactaaa
 gaggggtccatgccatttttaggaatctgcaatgccagcctgatcttcttt
 tattcatcaagcaaaaacaaaacccttaacaacaaaaatgaacacct
 ttcacgtgtaattatttttggcttagcttcttctcgaaaagaagtatgaac
 cctagcctcatcatcattcatttgaattctccttctctcacttctct
 tggctcttcttcttttcaaattctaggtctctctatactagtgaaggggc
 gaatcaaaatataatgattagattttaaagattgcaataaaaaactatgag
 ttaaatctttttaatttggtaaacctgaatttatattcattgttgtag
 atataatgattttagtgttactttttgttctggtaaaagatgttcata
 aatgataatgataaatataataagaattaattataaatgagtttttat
 99501 cacaaaaggtccttggaaatcacattcaccatcaaaatggtccttcaaat
 tgaacttaataatgtagtccctgaaaatagatgccgcaaatcaatgtg
 gtcttccgtcataatattgtttaaatttccggttaagtgtgatgtggca
 tataaatgagcccacaagatttgaggggttttttatcacaaatggtccct
 gaaattgaatcacaccaacaagatggtccttgaattgacccacaccatc
 aagatggtccctaaaattgaaaattcaatgaatctactccttgaaaaatag
 ttatcgtaaatcaatagatcattctgccacaattttgtaattttttta
 ttatgtgcttatgtgggttaataattaatctaaaaagttgtctaaatac
 ctttttgcacaatgaaattgtcaaaggcacaaggcttaaccaaggcctaa
 gatgggatttggaaatagttttcaaccaataatagacacacctcgggtat
 100001 aacctcattatctcaagatagacctcgccgttctttgagtcactagacca
 ccaaggaagcgtcgaacaagctctccaagattaagggtgtgaggatgttg
 atcgcctaaatgtaacggtgatgtcccagaagtcgttgccaatcagcca
 agcctcaccaaaaggtgatctcgaccaggttcaattcggtagagggtt
 tgaagtctgtaaaggtgggtgtattgagatttttttaagagaatagac
 agggaatgaggtaaaggaatgtggactgagattggagggtgattgcaggat
 tgtgttttggggttgacaactttttattaactattaatattaaacttg
 ttgatgcacaaatccgaaggggttttggaaacataaatccgaccgtga
 atctgcaagatataaataaacacaagatgtatcgtgggtcaccccaagg
 tttgggctatataccacactgatattgtattctgagggagagagagctct
 100501 gaatatgagagtgagagctttgagaggatgagagggcttgtgaggctctc
 cctaattaagagggtaatgggtcattttatagaataagggtcttctactt
 attacataatttgtccatttctttattacataattacatttaagttcccc
 gagtatttatcagaggtctaaatacggaggcccgaaagatggtataaaca

HEAT repeat-containing
 protein (ORF49) ↑

gtagtccccaaatcttcagtcagagagctctttggctggagagttgaa
attcagtcocatgtgtggccgaagtaactagatgtcgtctagaactgata
ctcgatagaagtagcatatgttgcgaggctgctctgcttgggttatgt
tgccttggctggctcggcttgtggcgttgaaggtgagggagtcctttt
atagaataagggatcgctcctcaatacataaataatgggctagagttgatg
cttgcggcagggcgggttgcctcagctggcagcgatgctctctaatgaaggt
101001 gaggagtccctttataaaataagggctcgctcctcagtcacatgaataat
gggttaggagtgacgctcgcgacgaggcgggttgcctcagctagcggcgatg
ctctctaatgaaggtgagagagtcactttataaaaataagggctcgctcc
tcagtacatgaataatagatgctctctaatgtaagtaagggagtccttt
tatagaataagggatcgctcccagtcacataaataatgggctaagtcctc
caagtatcttcataaagccagttgaggcccaatataatggtacataatg
taatccctcaagtcttcggccaatagagtcgttggctggagacttcaaa
ttgaatccatgatatgggccaagtgccggttgttcggatgaggatattgt
ataccctgactgtagctttagtagatgaagcttgcagtggaagcttga
agctagagctttagtaaatgaagctttagaagctggagctttagtaaatga
101501 agctttagaagctagagttctgtaaatgaagctttagaagctagaactct
gtaaatgaagctttagaagctagattgacataaataatgatgctcatgaatg
ttatgttgatgacatgagtgatgctcatgaatgtaacatgagtgatgc
ttatgaatgtgacatgaatgatgctcatgaatgtttatgatgattgac
ataagtgatgctcatgaatgtttatgatgattgacatgagtaatgctca
tgtataatttgaagtaactggacttactttgatcacctgggtgggtgata
atagcagcaggggtgctcgaataatttggagtaactgagcgtactttgatc
acctgggtgggtgataatagcggcaggggttgaataaattggagtaactg
ggcgtactttaaactcacctgggtgggtgataaataatggcagggctgta
aattttgaagtaactggcgtacttttgatcacctgggtgggtgtaaatg
102001 ggcaggctgctcgaataattttttagtactggacgtactttgatcacc
tgggttagtggaataagagggcctgactcttttaggcataatgggcttccg
ccttcacataaaccttccagcctattatttgggcttgcggttgggtt
tgtttttttttcattattaccctctgatggggttatacagatgctccg
aaagataagaaaaaataacatcattctgggtggggtgtttattcctgt
cttttgactgtccccatgtgcctccctttagcttctcttatttttct
ttcttttggcagatgacagacaaggaataatgtcatttctttttttt
tcttttgcctttgctttcacttttctttttgcttttgtttcttttctg
tagctttactagtggtctcccgcctcattgcttttgccttctgtattcct
tttctctaaagtgacaacacttttttttcttgattgatgttctctgcag
102501 attttgagctttgagatccacctgactagagcccacgtcagcagctctc
tgcctctcgagtcactagcttctactccttggtagaatgacttccgagcca
gcttccataatattcaagatcttcttcaatttggaaagtggcaacaaa
atctctgcatcttgtgcttgaccagaagaagcagttgaagtttctctag
acgtcttaacttttctctatcttctctataaaaaaaaaactttatctt
cttgtcccgcataaataatttgcggtccacagctagacgggaatgaatct
cctgctgtttcttgaacccttccaacttcttttctgagcctccctaact
tcagctggtaaaattcttacggagcatgcgctcgatggagcgaaggtggg
gatgcgaatgcccgttgtgcctgaggtgggtcaaaagcgcacatgggtcgg
gccaggtcgggttgggaagtgaaggagtgatgaggacagagagaaagac
103001 ggattcgagaggtaggttgcgagtagggcgggtgatggcaggggtgggtta
ccgtgtcgaggtgggttcaataggcagagggcgggaggaagttgatc
tttgtgagatcttcttgggttagacctcaagttggggttgtttgggtgg
atttgggtcagtgattgactcagaaaccttttctccatgactgaaatgg
aatctgagctagactggagatctcagctcctcggggacaagctgaaagt
ttgaagcttggaggagaaggagagagttcttattgactcactttcaca
gatactgggcatgtcctgcccgtcggaaatgacgttgttgcgatcaaacc
tgggtgggtgggtgagagccagcgggttggcccaggttgcctgaact
ctcaattcattttaaggtgattctgcttctctgcaactcgacctcag
tttcgaaatgtcctcttgggctaagtgaagctcatattcactaatgcta
103501 atttttttcttctactgtttcttctctctagcaaacattgagggctc
gcaactcggagagagttgaagcgatagatctgatcgggaggggaattgaa
ggagtgcagaggaagtctttagtgcgaatggcctggctctctgccat
catcgagcagcttgtggagtgcacgacccgtgagcaggtagttctca
tcgagcagaaaattcacgacgcaattgcacaacgacggcatttccacgctc
catcggagctcacaacctctgaaccttctcgcgacgccattgaaagaacca
aagaaccaaagcaggtctctctgtgtgtctctctatctctctctgtcttt
ctcttctgttacgtattttcttcaggagctcctccgatccagtcggtg
agttgtgagagagagagagagagagtttggctttttctgaagaggtg
aaagagatcagaagcaaggagaggttagagagagtttggggtccgcaa
104001 ggttttctgaaaaacccaaggagggaggttctaggttcttgggttgtgc
agattcacgatggtgagatgaaaaaatgaaagagaaccgacatagcttt

t c g t g t c g a t t c c c a c a g a t g g c g c c a a a t g t t g a t g c a c a a a a c c g g a g
g g g t t t t g g a a c a c a t a a a t c c g a c c g t g a a t c t g c a a g a t a t g t a a a t
a a c a c a a g a t g t a t c g t g g t t c a c c c c a a g g t t t g g g c t a c g t c c a c a c t
g a t a t t g t a t t t c t g a g a g a g a g g g a g c t c t g a a t a t g a g a g t g a g a g c t
t t g a g a g g a t g a g a g g g c t t g t g a g g c t c t c c t a a t t g t g a g g g t g a a a
g g t c t t t t a t a g a a t a a g g g c t c t c a c t t a t t a c a t a t t t g c t t t t c
c t t t a t t a c a t a a t t a c a t t t a a a t c c c c c g a g t a t t t a t a c g a g g t c t a
a a t a c g g a c c c t a a g t a t g g t a t a a c a a a a c c t a t t t g t a t t t t t t t
104501 t t a a t t t a a t t a g a c t t g t g t g a t c c a t t t a t g t g t c a c a t c a g c a c a t a
a c a a a t t t t t g a c a g a a t t g t g a c g g a a t a a c t a c a t t g a t t t g g a a c a
c t t a c t t g t g g g c t a c a t t g a t t g a t t t t c a a t t t t a g g g a c c a t t t t g
a t g c t g t g g t t a a t t t c a a g g a c c a a t t t g a t g t c a c a g t c a a t t t c a a
t g a t a a t t t g t g a g a a a a a t g a c t t a t g a g a c c t a t t t a t g t g c c g c a t
t a g c a t t t a g c g a a a a t t t a c a a a a t t g t g a t g g a a t g a c c a c a t t g a
t t t g c a a c a c g t t t t t c a a g g a c t a c a t t g a t t g a t t t c a a t t t c a a g
g a t c a t c t t g a t g g t g a a a g t t a a t t t c a a a g a t c g t t t t g t g a t a a a a c
t c a t t a t a a a t a a a a t a c a a a t t a t a t g a t c a c t a t c c a t t g a a a g a c c
105001 t a t g a a t t g a a t g a c c t t c t a a c a a g g a c t c a c c t a g g c c c c c t a a a t
c t t a g g g t t a g c c t t g c c g t t t c g t g a a c t a g a a t a g t t t a a a a t a g a a
a a c c g c c t g g t a a a g a a a c a t c a a c t t g a a g t a t t t c a t t t g c c t g a t g
t g t t c g a a a c g g g a a a g a a a c t a c a t t t c a t t t g c a a a t t t g c t t g a a g c
t g g a a a a a a t a g t t a t t g t g a t t a t a t a g g a a g g g g a g t g g g g g t g a g
a g t a g c a c a t g c g g a a c c c g c g t c a a g g c a a g g c c c t c c t c c c g c a t t
c t g a a c t a c c a g a g t c c t c g a g c t g c g c g g a g a a a t t t t c a t t g t g a
t c g a a a t t g a a t c g t a c a c c a c t t g t t t t a t a t a a a t t g t g a g a a a t t
t t a t t t t t a a t a c a t a t a t t c c a c c a t t t a t a t a a a c a c g t a g t a t a
c c a c t c t g t t t c g a t c t c t t a t g t g c a c g c g a c t t g c a g t c t t t a a g t a a
t t g a g t t t c a g a a t t t g g c g c t t c a c t a t t t c c t c a t g c t c t c c a g t t
105501 t t g g g c c t t t t g g a t a t g t a g c a t t t g a t g t g a t a t g t t t t g c a t g t t a g
g c t t t g a a t a a t t t a g g c c c t t t t c a t t t t c c c t t t g t t g g g t c a t t g t
a t a g g a a a c c t c a t t c c t c a c c t c g c t t t t t t c t t c t t c a c g t t g a g t g
a t a g g t a a a g a g c a g g t a c t t a g g t g c g c a a t c t t t g g c a c t c c a g t
a c g a t t a c c g a a t c t t c a t t a t a t g c t a g t c a c g a c c g c a a g a g t t t g t t
t t t t a c a a a c a a t a t a a a g a t t a t t t c a c a a a g c a t t c a t g t t t a g g t
a a t t c a t g t g t c g a c a a a t t g a c a g t t t a a t a c c a a t t a a g a t t t t t a
t g a a g a g a a t a a t a c t t t t a g a c t a t a g t a c t a c t t g c t a t t t a a a a t c t
t a a t t a a g t a a g g t a a c a c g c c a a c g t a c a g t a t a t g a t g c a t t a t c g t
106001 t a t t t a a c t c a t a t t a t g a a t t a t c a a c c a a a t a t c a a c a c t a t c t a g a a
c a t g t g a c a a a g c a c t a c t c t t t t a a a t c c c t t t t a a t t a g t a a c t a a c a
a g t a g g a a a a t t g t a a g t t a a t a t a g g g a t t a a t t g g c a a c g t g c g a g t t
a g c t a c a t g t g c a g t g g g a g g c a a a g a t g a t c a a a t c c t t a a a c c g a t t t
t g t g a a t a a a t a t g t g a t a a t c c a t c g t g g a a a g t g g g a c g c a g g a t c t c
g c a g g t g c t g c t t c t g t t a t g t t a t g t t a a c t t g c a c c a t g t c c a t g t g c
a t t c a t c c t a g g a a a a t a a a t a a t a a a t a a a t a a t t t a g a c c a t g t g c c
t g c a c c t a a c t a t t g t t t a t a g g g g a a a c a a c t g t g t t g t g g t g g t t g t
t t t a t g g g g c c a c c g g c c g c c g c c a c a c c g t t g c a a t a t a g c c a t a t a g
g a a g a t g a c t t t c a a t t a t g c a a t c a t c a t t t a t a t g a a t g c a t t g t g c a
106501 a t t t g t g t a g a t t c c t t c a c t a g t t t t a t g g g a a a a t t a a t a t a c c t a
g g t t a t g a t c c g a t t a g t c a t a a a c t a t a g g c t a g t t t t a g a a c a a t a t t
t t t a g c t c c a a t g g a a a t c c c t a a a a a t a t g t a t a c a a c t g a c c a t t g g a
a g a t c g a a a a c g g t a t a a c a t g a g a g a t t t t t t t t t g c g a c c a t t t c a a t
g t a t g a c a a g t c t a t t a c a t g a g t g a a t c a c t t t t t c a a c a c a t a g a a a a
t a a t a a t t g g c t a a g a g a a t c t c c a a a g g a g a t g t c a a a a t g t c t a c a t a
a g c t a t a a c a g t t a a a a a a a a a g a c c a c a c t a a t g t t c t c c a a c c g a a a
t g t c a a a t c t t a t g t g a c g a t g a c a t g g a t t g g t a g t a a g a g a g t t g c t g
t c a a a t t t g a c a g c a a c t t c a a a t t a a t t a t g a a t g c t t t t t a t t a a t t
t t t t a t t g g t t t a a t c a t c a t t g c a a t t a a t a t t g t t g t t g t c c c g t t t c
107001 c c t t t t g t t c a a t c g a g g a a c g a t c c c t c t g t t a a a a a g a a a t g g t g a a
t g a t c c g g a t t t c a t a a t t a t t a a a t a c c t t t t a t t a a g t t t t t a t t a a t
t t a a t c a t t t a t t t t a t a a t a t a a t a a a a t a a t a a t t t a a t t t g a c a t a t
c g g a t g g a c a a t a a a a c t t t a a a a g a t g t c a a a a t g c a a t a t a g a t t g t c
a a a t t t a a a t t t t a t t c a a a a t t t a a c a t t t c t t t a a a a g a t g c t c t a
a g t g a t g t c a a a a t g a c t a t g a a a t a g t t g t t a a g a g a a t c t c t c g t t a
a g a g t g g a g a g a g a a t c a c a a a c c a g a g g c c a a a t t t a t g a c a t g c a t g c
a g g t a t a t t g t a a c a a g a a g a g a c a g t c a c a a a a t a g a t a t c a a g t c t t
g t g a t a a a a c t t t a a c c a t g t a g a g c t g c a t g t g c a t g t g a t t a a g t t g t
g a c a t g a t c a t t g g t g a t a a a a c c a g a c a c a a g t g g g a a g g t t t t t a a c t
107501 t t g t g a t c a t c a t c g g a t t t c a t t c t t g g c a t a g t a g c t g a t a t a t c t t

ccaatgaatcaaaccctatataataaacagtcaagttgggggttgggggtaa
atcgaccatggacctacttagctaactctactacaatgcgcgcaccttgc
ttgcgagcctggcaaaagatgatacagagcctgtgagaggttgatttattt
gcttggatgacaggtgggtttcgcaagcttcaactcaataattcttcttc
tgttgaattcataaacattttctagatatctgcatgtggacgtgaacattg
cacaagaagtggatgtcgggaggaagcagaaactagctgccacgtgctc
gtcatcgtcatcatccatacttcttctttttgtataattattgaaagt
cctaattgttaccaaaaatggttttaaataagatgggacaacaaaccacc
gggacccttggatcctttacgaccaccacctaaataaaaaatttaaactg
108001 gtttataatttcaaagaactttaagaaaagctttcaattctcagtaactgt
ttattttaacaaaaaatcatattttataactaaaaagtttaattataata
ttatttactttactatttatttgtccttatagtttaactcaagttttcaa
gtcattttcattagttttacttataattttatctgtacagtggtgactca
gatcacaaccttaagtttttcaatcaccatataatgtctagttgtaca
ataccgataacatatacggatcgatctaccttaaaatcaagtttattttt
tcactttttatataatgcatgcaccactcggcagcctatctttctcagac
ttcatcagtttatttaccctacttggttggttctggaagatgatcgagaac
attgacaagtatgggtccaatttttacgtttgtgaaaactcaatttactgg
aaatctataatgggttgagagttgatcagataaactacattgagaactgtg
108501 ttgcaactgaagaagttataaggacatataatgcatgcatgcttaacacc
aaaacttcatcccatcttcatatactacgtaccaagcaaatatacccatg
aattgattctttaccaaaaaattactcatttatttagtgaacagtttgag
aagtataataactcagaaacagtaacgaattcaaagttctaaccattaca
agtacattcaaagttctaaaaagcaataaacgtaatacacagccagtggt
cggatccagaattttaaactcggggagttccaataataaatgtcaaaaa
tttatagacaaaaacaacattgaaaatttaaataataacatttcattaa
aaaatcatatgcaaaacaatattacaattataaaatcataattcaagcg
tccattacgaggtttcataagttactagggagtagcctaagttattgatta
tctaataatgaaccactaaattttaaaccgggtcaaataaaactctccc
109001 aaactcactctatctttcaaaccacataaccatttttctcttttactag
tttccaaacatccactttctctctcactaagaaaacagttgtcacaacctc
caccacccgctcctcccattctcccgaagccatggcagctccacccc
caaatttcaccacccatgacatgacataatcatatccatcttactaacct
tatgtaggaaatttcaccataagaaaaatctaaggtggttgaataaa
tttggggaaatttgggcttatgtactgcccgcagacgacagagaaagg
ttggaggagaagtgggtggcaccgatggatgagaccatggcagaagcaaa
gaggagagaggttagaaggatccttattttcttcttctcctggaacggt
catcttctcctggggcgtcatcttctcggagaagaagacaaccactgca
cctgcaacctcaaacctgacagaccttatgctcgagtacgaggggtcct
109501 cgaaaaatttatagcagcgtattaggaccactgacggatcctgggacctc
ccaggtccctatgtggatccgtacttgatcacagctggtataaactgatg
aattttccctcgcagcaacaataatactcgaacctcctctattttgt
gtccctagtgcagcgcgatagggacgcgtaagagggggcgacgtacg
ttacaactgaggggccccacaattctttctaaaaacccgatgaggggtgc
aacccatgctcttcccgcgttggcccagctctcccgaacgcacccg
cccttcttccatttgtacagtaagccggggcccgccggttggcgttaagg
acgccaccacgtgtccgtccgatttccctggtttagcaccgctcagcacg
tcacgtctgcgttgccttcgtcgctgatgtcgtggacgagcccagagtg
gcctagtgggtcccgaaccgagtcattcaatacttggcttaagtccttg
tgggtgggtcttgaactcgtgttcatccatcattgctggtacttgggtg
110001 agcacaactagctggatcttctgtgtaaaatttggattaaaatattgcag
tatccaacgaaaaaaaagggtaaaagtaaaacctcaaggaatctgtaaa
aaaaatttatggataatcggagtgaggacatgcatggatgcatgcttt
gttttttactttttgttaacaatcaactaaaagtggtaaaagagagca
acaaaatggtgtgaatgatgatgggaaggcatgccacaagtgattttg
acagtgattagatgcaagtgatgtaaatgtagcaacaagagaacatgtga
gagaatgggttaagtgagatgcaagccttttagtcaaggcaaaacaaaaat
ggacaatgattgttaacaacccaagaatgtgcaacaatggcacccaaaa
catgtgaaagggaagtgggtgaaaaacaaaaagagaaataaacagttgtg
gacacgcgtaagtgattatgatctattgatgcatgatgagagttggttag
110501 gttttctgtacttgggtttagagagcttacatggaccatgacgtcat
aatatagtagtgaaatgtttcatgtaagttattctcgtggatttaggtc
tttttatgcaattgctttaattaaccttgtatcttgtttctttatgatag
aattatgaaattcttgaatttatcttgaagcttttattttcaataaaac
ttgtacacacatctttttctttaaatacaagcatctactaaaatgatg
accatgaaacaatgacagccaaaacatatacaagaatgtacgttgaaaaa
ccaaaaacaagagaatgcaacaattctatgggocgaagcagatgatgagtc
ggtcatgctcgtggacttactcgcaaacatgaacgactaacgatatagc

atcacggttcaaatataaattacgctggtttcttccatgtaggttttcacat
ccaaaagcaaaaaagcaattaaagatgaaaactttgtcaagagatcacag
111001 tatatattctcacaaggcgcgatgcagaacaagactgaggctactatcata
tcatgtggttagtggttgggttgaaaagtaagtttggagaggatgtgtgctt
ttgcttgattgggacttgggagatgacactcatgatttaagtggcattt
gccaaaggagatgggggaaagttagcaaaagttagctttggactttttctc
caacttccacggtatgccacgaggcaattcaacagtagcagcagctccaa
tcacctgtaacctacccaaggagccacaagcatcgcaattttgcattgc
atitgcatttcogatatagtttaattggacgagcgtgaaatcaagctcgggt
ttacacacaaaatattatccatttttttgcgttaaaagcaaacctaaat
tactaaacctcaaaaagctagaacaagaacgataaataagtaattataa
aattaatcagtaactttatggatttaaatctctgctgggaaagcctagaa
111501 aagccagagcttaagcataaaagggaaccacactcgcagacgcagaggt
ttcattttgctggtttttcttgggcctggagaggagagttagtttaaggt
cacaaatttacagagttttcttttctgtgacggtgagagatgaaaagc
aaagggagcttggcggtttgttcatcgtaataagaaaaatgctagactaac
agggaaatatacgggtttcagaataggttagttgcttttagcatggctgtca
cagcttttcttgataaggaatcggaaagcaataactttttatatttaaat
tatagtttacaggattaatataattgggttagtttgctgattcagttccgc
tgtcttggattctatggtataaattgtgattacttggtaaaacttcaagt
ttctattcgaggtcttagattcgattttttctacgatctagctgataaa
agaagaaaaactgtcactattaatttgagttaatgtctatgcggaataaa
112001 gttaatgttccacctcaaaaacaaatgcggaataaagttgggatatgcaa
agcccccttagttttcaatgtggggcaaaattctgcatctttacaagcac
ctttacatataagataactcaaacaaaacagtatacatcacatgtcaaa
taaacaggtgtgatcgttatacttttacgcataagccaaatagtttgat
ttacaaatgataatacaattaacacataaatctatcatcaaacataccat
attcagagcaatgaagattaagaaattaagaaacgaataaagtaactagg
ttttaaaagaaaacatgcaaacacaccacaccacaccacaccacacc
aaaactgaaaagcaatgcatctgcaagtgcagggaggatgagtcagact
agggaaatgagagagcaaacataataggccaagtaagccttatcctgataa
atgtacccccaaaggacacggtgtcgagagggggaaacgggttgagaatgt
112501 ggtaactggtgggaaagtaaaagctggacacagacagcaacagcgcagcg
cagcagagaattacaccggttaactcaacggaggtaaaaaaattaataac
tcaacggagcagcttttgggaaaagaaacaaaagctcattgaacggactt
ccgcagccccctgtctgttccgttctgtcctgcatcacggggaatcatgt
ttgctcttggactaattggatttgtacctaattttcttttggttttctttc
tttttggctctcttctgtccttaaaattgaaaattttctccacagagttgtt
caatcacatctatttcttgagtataattttcataatgtatagaatttcatgt
aatttaccataatctgactccattgctgaacggttagtgtaaaaaattgacc
tacttgacttgctaatacgaattttttcataaattaaaattattttttgg
agttggtattataaaatttgcaggtttatgtatgaccactattgggtgaa
113001 ttgaaagtgggaaacctaaaccaccccaacaatgtatggaagtgggtgtt
tactttggcaacatgggtttccaagttttgacaaacgaatattatgttgg
tgtaaaacataatcgatcaaaatagcaaacattatcctctacatcgtca
tgaaaatggttttttctgttttgtacgtatataatgtaaggttgccttagg
tgtctcgatgcttattcatgatgcacttttgcataatttatctagtcaaa
cgtgtgttttatattttaacggtttagttattttcataatcttaaaact
ttcccgtatagtcacatttttgaaaaaacaagaagtttgcctatatttt
aacttttggtttcttttctaatgatagagtgcctagtgacccaaattgaat
ctcctacaagtaaatgaaactcaaaataaaatcaaaggttaaatcaaat
ggacgactttattattttctaaccttaataatttttctttaactaggtat
113501 cattggctgcatgattaaattagtaaaaggagtaggattctattcctttca
aatttcttctctctagtttttctatttgaacgatcacgattaaacc
atgtaaacatttatgggtgacttctttataaaaaataaaaacaaaagaaaa
gtaagaaagcatggaagaaaaggaaaatgatttggaggaagagatcta
actctattaaattcaaatacaattttttcagttttttagctcaacat
gttcaaaatttataaaataaaaatttggacaaaataaattaaatttttagat
aaactgatttgactttgattcaaacattggataaatttgaaggccaaacc
acaggaaacagataaaaacaggaaggtaaatagcagagttgaaaaggaaaa
gattatacaaaattacaatggtaagcagtaaatgaaggaaggaacttgt
tagttaatcaagtttaattatttttatttttatgtatagagctgga
114001 tccactctccaggagtgacccacctctgaataaaatattcaattattaa
aaaaaggtaacaaataaataaattaaattcataaaataaacccaaaacaca
gacaattagagcttgggaagcagagacagacaagtcagctctttcaactt
caatccgaagaatataacacacacacacatacatacatacatacataca
taataaaaaatttctcagctgaaaataaattttttcaattctgttggct
tttgctcaactgggttctttaggttttgaattgggtatgaaacaaaccg

	gagcttaagcatgtaattcgggttatctggtaggggattttatatataa ttatatatagatagagagatatcttgagtttattcgggttccgagattt aggttttttgaagctatgcatctgagcaacggcatgggctcgatgtcaac ggtagttcgagtgagcttgaaatcgggggatgtggctactgaccagt 114501 ttctgctggactaagagtctctgtggctcagacgatgatccta	MC-20SSR96C2For
	atgatcttgagaggatgctcagaacttgctctatgaaggtaactcttt ctgagtatc tagagactttgcttcatgggttttcttcaatccttctgtagatgtcaatt a	MC-20SSR96C2Rev
	cctgtctcccaacatggatttatctttctgtgggcgatttactgcaag ttttaccacttgggttataatcacttgtgtttttcatgctgttgtttaag ggatttggagcttgtggtagaagttaagatttggataatttgatatcttt tactgcttaagttgtgatttgtttggatttgaatcatgggttgtgttaa atctcgttacttgcctacatgggttgatttaatttttgtttctaaaaa ggtaaatcgtcttgttccagaagtaagatttggatttggtttgtgtctt 115001 tgtagattcatagccctgttttattaatttggttctatagttgtttatt tactaaatttgtggattgatgagttggttggttgggatttttttagtga caaatgtaatcgtgaggatcgacattcgttgctcgagagaacaag aacggattcgatattgttatcagtgatgtgcacatgccagacatggatgg attcaaacctcctagagcaagttggactggagatggactgccggttatca gtgagttcctcgtttgagtaatttggtagatttaactgttggtaaccg gaatcgtatgtgttaatttggagctgacacctgatgtgtaattccaca gtgatgtctgctgatgatggacaaagtgtgttatgaagggtgtgactca tgggtgctgtgattacctaatacaaccggttcgcattgaggcgtgaaga acatatggcagcatgtagtgcggaagaaaagaatgatggaaggatgtg 115501 gagcaatctggaagtgtggaagaaggggatcggcagcagaaaccatccga agatgcagattactcgtcctcggcaaatgaagggacttgaagaactcaa agagggaaggatgaggaagaagagacggatgagaggatgattcatcc acattaaagaagccgcggtcgtttggctcggttgaactcatcaacagtt tgtgggcgctgtcaatcaactaggcattgatagatgtcttctccctat actcttcaattatacttctgattgactgtagcctaataatgataaagt cttttgatcctttgagttttccacctgttcttcaactaggagctcagtc agttgaaattacgtacacactcctcacgcgaacggaatgggcatgcatg actcataattaggagctcaagtctcattttgtattgcaattattttaa taaaaattcaaaccttgattatttgatacagaggctgttctaagaaaatt 116001 ttggagttgatgaatgttctcgggcttactagagaaaatgttgccagcca tcttcaggtatggtaattatgcatattataataatttacgaggaggaatt ctcttacaatatgaatttgaagaaagtctgttcttatgtgacatggtta aaatagtttttggttgagattaactcatcctgccaatttatgaagcag aaatacagactacacttagaaggttaagtgggtgtctcagcaccagaa taatttggcaactcttcatgagtcctcaagaagcgtcttcggagcaa tgtcatcgtcagtgggctggatcttcagacgctagctgtcacaggtcaa ctccctgctcaaagccttgccacactccaagcagccgacttgggaggtc aacgataaaatcaggcgtacctatgcccttgttgatcaaaggaacttat tcagcttggaaaatccaaagttgaggttgcggaggggcaacatcatctg 116501 agcagtagcaaacggatgaacttacttcatgggattccaacaactatgga ggcaaaacagcttgccaacttgaccagtctgcacaatctctggggctg tcaatatgcaagtcaatgctcatggagtccagaacagttctctacttatg ccggtagctcaatcgagccaagagggcagatcttgaacgaaacttctgg aaatcatgttccacggctctcatcatcaatggggcatccatcatgtcta atgggtatgagtagtgagtttccgggaagaaatggaattgttgatagcggc agaggcggaggatataatccaatgcaacaaaactcgtcaatgttgaatta tcccctgaacaacagttcacaactaccatgcaacagtttactcttgga gtacaccgtaatgtccagcctcagctcaaaaggagcatttccagaagat gtaaaattctgatgtaaaaatatcttctggatttatgccaagttatgat 117001 gtttaacgagttacagcagcacaatcaaatgattgggacttgcagaatg tgggattgacttttgaccctctcaacatacaaaactccatgcacaatagc cttgatcgtcagttatagttcatcaaggattttcttctagccaaagagg tggacaaagcgggagtggtggaagagctctgttctcaatgggcaaggca gtcatcttgggaatgtgcaaaataccagtcaacgtcttaacggcttgtt gttgacaatacagttaggattaaagctgaaagtgcggccgatgcagcttc tcaaacctagcttcttcatgaacaatatggtcaggatgactaatgagtg cacttctcaaacaggtcagtttttgcctcattccattttacggatttt gaatttttagatagctgttctgagctatcaagggcatgattactcatga aatttcttttgttttccagcagcaagatggacttggaccocgctgaaatg 117501 agttgactttgatgggtattccttggataaatctctgtgtagaatgcg ctcctacagtattttgatgattgcagatctgggtatataagttgcggca atggaagtcatttcttctgcatcagatgggtgattcaagttgtttc	
		Multi-domain protein (a sensor histidine kinase and a response regulator) (ORF50) ↓

tacaccaagaaaaccagagtgcgtaatttcacaaatcatatctttcgtt
 gtgaatatcatgatgtgctgttttttaggagataaaagatgaggatgatgg
 ataaccgaggtggatcgaatatgtctaaagatatttgttgtgcacaatac
 tgotgatgtaggcaaagcatatgcaaaatttagaggggctagaaaag
 aagcatcgtttctggggggtgcccgtgaaacttgcctcggttttaggctt
 catcttgatattctagggtagggttttgatacttcttctaactctaataa
 118001 cacatgaagcttgccgttagtgaatttctttgccccttctcggctgcc
 tttccttctcgttccaggaattgagtttgatgtacatatcttctatttgg
 ttactcaagacataagtctggggatattctttctggttgttttcttcta
 cttaaaatttctaggaattcttgtatgtcccctattcttaataatcttct
 cctactcttagtctgaaggctttgggtgatttaacacaatcgttgaaat
 gaaggaaagcttatttattgatatccccgataagctacaaatattgacat
 atacatgagtcaaaataaacacacaagagggagccttcacaaaggttgc
 taggagaagtctcagcagtcggtagagcccagaaagagaaggcaccgga
 gggggatcatctggagcctcagtactggacagaaccctggaaggaggag
 catcagagggtgatcatttggagcttcattacgcgggtacagccccagaag
 118501 acgaaggcaataaatgcctttggaacaaactcacaatctctgatgatca
 agtaaaacctgacctcagtttcttcatctggcgaagcttctcttctcat
 gttttagatagtcagtcgagcgggtgcaactgtttattctcatgct
 tgagccctctaactctctgttggagactcatcactcagccgccaatgat
 tcaacttggcgggttcgagcaaataggcgttgggccaatattagacacaga
 acctgcacactgaacactgagagccagcgaatccttaacagctaactcat
 cagaccgtttgaaagtagtctgttatctttgggagtgagaaggttctctg
 gccaccaccgagcgggtcatatcattcttcatcacggaatccccacgggt
 aagagaccagtaggggagacgaaggatggcgccatattgttctggag
 aaggcggggtgcctctcaacaagggtcaagtcaaaacgacgggtcggag
 119001 gggccagacattttcaaagggttgaagagagaagaggtcggacaaatca
 agatcttagaagtgcaagaatgaagcttctactggtagattcaagtggt
 gctttagaacttaatgccagcccctataaaaatctgcaactcgacggagct
 tcagaaatcgaagaggcctgctcagaaatcggagagggcgcctgctcag
 aaatcgaagaggcgttcttctcaaaagctgagctgttagagatcac
 gagggtgatctcagaaatcgaagaggcttttgccttctcaatagttggg
 ctgctcaaaagaccacgaaggccgatctcagaaatcgaagaggcgcctcgt
 tctcaaaagctgggctcccagagaccagagggccgatctcagaaatc
 gaagaggcactacttttccagccttgtcagcacctgtcacacgcacact
 cagctttgcggaattatgggcattctgtcgaagacttctggggaagtag
 119501 aaaacacatgaattcttactgttcaatcactcacttcccacacgcaacaat
 agtcatgggtaccacagataactttgccaagttctctgccaagttga
 gcacgtgaaagcttgcagctcccactacatcgtctgaccaagaagggtaa
 aagaatagcaaaagaaacagcactaacaagtttagaccataaaattttga
 aggtctagctaccatattattaccacaagggtaaaggaaacagtaccact
 gctggataattggaaagtcctgtgtgtcaacctctgtgctcgtggcaa
 ggtagactagcaaacatgcccactcttactcacattcagaaaaaacctc
 ccaataagattgcttgcctcaaaatcgaagaggcaccgtcctccgaatct
 cgagagccagactcccacatgactactttatataaaatcgaagagaggg
 taaaggaaacagtaccattgctggataattggaaagtcctgtgtgtcaac
 120001 ctctgtgctcgtggcaaggtagactagcaaacatgcccacctttactc
 acattcgagaaaacactcccataagattgcttgcctcaaaatcgaagag
 gcaccgccttccgaatcttgagagccagactcccacatgattactttct
 caaaaatcgaagagacactgctcccgaatcttcgagagccagaccctta
 gcatgatgtcttctcaaaaatcgatgaggcatcgttctccgaatcaatc
 gaagaggcgtcgtttctcaaaagctgggctgccagagaccacgaggg
 ccgatctcagaaatcgaagaggcacctacttttctagccttgtcagcacc
 tgtcacacgcacactcagctttgcagaaatattgggcattctgtcgaaga
 tttctggtgaagtagaaagcacatgaatcttactgttcaattaccactt
 tccacacgcaacaatagctcatgggtaccacatataactttgccaagtt
 120501 ctctgccaagttgagcagctgaagcttgcagctcccactacatcgtct
 gaccaaaaaggttaaaagaatagcaaaagaacagcactaacaagtttag
 acacataaaattttgaaggtctagctaccatattattaccacaagggtaa
 aagaacagtaccactgctggataattggaaagtcctgtgtgtcaacctc
 tgtgcttctggtggaaggttagactagcaaacatgcccacctttactcaca
 ttogagaaaacactcccacaagattgcttgcctcaaaaatcgaagaggca
 ccgtcctccgaatctcgagagccagactctcaacatgactactttctcaa
 aatcgaagagagggtaaaaggaacagtaccattgctggataattggaaag
 cctgtgtgtcaacctttgcttctggtgcaaggttagactagcaaacatg
 cccaacctttactcaaatcgaagacaacactcccacaagattgcttgc
 121001 ccaaaatcgaagaggcaccgcccctccgaatctcgagagccagactccaa
 catgattacttctcaaaaatcgaagagacactgctctccgaatctcgag

Protein COBRA precursor,
putative
 (ORF52) ↑

agccagaccccagcatgattgctttctcaaaaatcgaagaggcatcgttc
tccgaatctcgagagccagataccacagaccactttttcaaagtgcctg
acagagttaaaacatgtgaagctggcagctcccactaccgtgctatgacc
aagcaggtaaaggaatagcattactactgttggtagggagacgcctat
atatgtcgacctccatcccccaacggacaggcagacctgcaaaaatgctca
accctccatcatatctgagagggcactcccaacgaagcctttcgaaat
tcagctttctttccccccgataatacctctgcaacaagctatactagag
caagaatatctcatatcatcaggggtaaaagcaagagatcccatatcat
121501 gctttttccctgccttttctttggccttgtttttacctgcaagacaagg
agaaagagagcaatcagtcagcacttggaaatcaagcttccagccaggaac
tgactgcctggaaccccttaoctgattacttaoctggcatgtctctcgag
tactcatctcaacatcttatgtttccaggaagattccgcatctgcttg
aggaacagatagggcaagtgcgaaggatacaaggaagcatgtggagacaa
gcgtaacagcacacgtgccgatacatccattactctgtcaaaagcaaaag
tatcccatatcagcaggggtggaacgtactctagatttgatggacttgttt
tgaccctcaaatcttccagtcggccttatactctggaggaaaccagaaaa
ccctccagctcagttcaagaataagcctgtggaaagttacttcttcaaaa
122001 gcaaaagtatctcatatcatctcttctcatttttcttctctttatccttc
atgctgttgcgaagatgggagaaggatgaaacatcagtcggagctctgatt
gcttaccttgtctgtcacctctttcagcagaccccctagctcggcgactt
ggggaactcctactacatggtttgtatcgcgctcgaccaagcctgaaact
acaagtaagcttcaagtgaattgatacatcacttctgtcatctccacca
gttaaagataaccaccctggatggaggaaaagtacttccacagaagatgc
cacatctactatgagacagataaggcaagtcaagacgacaccacactcc
gatacttagaagtttcgtgattacgagatcattctcccacaatatttct
aatgtcatttgtactaaatcattcacttgtactcactaaaggagagcttg
aacctatgacttgtgtaaaccttcacaattaatgagaactcttctatt
122501 ccgtggacgtagccaatctgggtgaaccacgtacatcttgtgtttgcttt
cctatctctatccatttataacttaccacactaatgacccggagcaatc
tagcgaagatcacaaaaagcgaaccgttttcgctacctaggatctatctg
caagagaacggagaattagatggaaatctcaaccatagaatcagagctgg
atggatgaaagtgcgaagcgcacccggcgtgtgtgtgaccgtcgtaggc
cactgaagctcaagggaaaattttataggacggcaataagccagcagatg
ttgtatggcacagaatgttggcggtgaagcatcaacacgtacacaaaat
gggtgtagcggagatgaggatgcttcgtgggatgtgtggcaccacgaaaa
aggataagattgggaatgaggatataccgaggtaaagtaggagttagccgaa
attgtaggaagatgagagaaaaatcggctccgggtgat ttggacatgtgca
aagaagggcagactgacgctccggctcgaagatgtgactacgggacagaggy
123001 ttcaggtgacgaagggtagaggaagacctaggaaaactttggaagagact
ctaagaaaagacttagagtagtggatctaaccggatgacatgacacaaaa
ccgaatgcaatggcgttctaggattcatatagccgaccccacttagtggg
aaaaggctttgttgttgttgttgttgttatttttcaagtaggga
gagaaactccgttgttgttgttgttactagtgaagctctcacaccgctg
gagtcagtagagtaggtgagattgaaatcatgaagtgtgtgtgcttt
tgcgagaggtgctgacgaggtggtcctaagttctatgctaaagttacaaa
tgagtaatgctagtgggtgatgtttcactatttttaggcactaattggaaa
agggtgat ttggcttgtgccggagggagagtaaatgggcaaacgggaacaa
tgcaacatgaagtggcggataagataataagaagtggccacgttaatt
123501 aagcatgagctttggaatcattccaaccgcccacaaaaagaagcgaaaaa
gaaaaatggaagagcctgcagaatagaatagaatattaacaaggaataa
tttcgggtgaacaagtcgcaaaatcgtttcactttatcctgtgaactct
tctccctaaaactcaaaagttacaattagaaaacttatgtgggaccatca
cttaataagtctatcaccggccaatgaatttgattataaaaccttactat
gogttaacatcaataataaaaaattaaaaataaaaataaaatcgagt
aaggataaaatgacgaatttaacataaaatattataaatctccattatgc
gttaacatctcacataacgaaat taacaaaaat taagcaatgtgaaacaa
at ttttcaggttaataagaaatgacgatttttgagt ttttaaggagcaaaatg
124001 agatccaaagtcaagttccaaggggcaagttgagcgaattttgagttc
agagagtaaaagttagcacttttgagttacaggaaataaaagtgaaatcaa
aatcgagccatagggagtgcaacaaaaataaagccttagtgcaattacca
cttgaacttaaaaataggataggcctattcacacgttcat tttcatctct
cacacccttgttaatttttgtcattgat tttcttcaattcattcgattc
atcgaccgaaaat taagagtggtgtgagaagtaaaaatgcacatggataa
cactaggggagaaaaaggcggcataagcactagtcttggaaacattagggt
gagctagaacgagattctaatggacactgctatcgcttccacaaccaac
aaagtgtatacaaaatacaaaaaagaatcaagaacaagcatctagttcaa
tatatatcgtttatacaaaaatagagaacattagctacatatagttttg
ctggtactttgcgggtgtgtcgatcctccctcagtat tttgcagatgaaa

124501 gcaaaagaataccagccacaaccaaataccgctaaaaacctctttaagt
accggcaatggcatgcacgccttctggtagttgggatttcgcaaaacctc
cccgttgagcatgcatcaatcaaaacctgaaagaatataaacatcagc
agttagtacaagatgtagcattgctcatccctcagctactgctgcatca
ccataaacgctacaaccattacgagaactgagaaaattagagatgtgcgt
tttctgtaaccactgctaggatagctgttaggaagagcgactcctccc
gttaaagtatactttcgtcgaaacccatctccaccaaccaaattgattc
cgggagtaatcttctttgtaaatgagagcactgactgctgtttcccaggc
accgtaggatccttctttgggttggtccatctgtttccgccacaagata
gttcaatcctttgaggccttgcataaatacggatattatgacaccattaa

125001 tttccaagtacttgcgttgaagagtagccttttcgaaaccaggccct
gctttatccagttgtactgcagcgaaccaatcaacaaaacggctttcatc
ccaattgaagactgtgaccttgactccatccacgagagtagtctgtat
ataaatgccagttgatgctgatcccacagttatcaccacaagggtgctcgt
tttgggactggtaggtgttaatatcagcccaagattttgcttgacagt
tcggttgtcaaagggacaagaagtgactctggtgggagaagcatagctg
gtgcagttgtgctacaagttcgaactgtattactagggcagccacaagcg
caagtgtgcatgggacgacagactcattgtagaagcagaaaatgaaac
acagcaactagggattgctcctctgggttgggtaaatattgcacacaacct
gccagctggctactgcagatatactcactggcaagccagatcggtcaggg

125501 aactgactgggactcaccggacaggaggccacatttataatcagggtt
gagcgtgcctttgattccccagttttgcggcggtagagctgggaccggt
tgagatttgaggcattttaaaaacttgcatctggaatgatgaaactgac
ttgcttgcatccattgttggggcaagatagtaccatttcggcaacaata
agggaccagaccaagaagtgtgtcattggcttttggtaggaaggtcaa
taattgttggcgtctttcacaattcaatacagttgagaagtctagattc
tcgtagaatgtacctgtttaccgaatatacaatcagaagaatcaacaac
agatggataagcccccttcggtgaaaatatagaactcatctccatccagt
cccagctcagtttccaattatcaagacgaccaagggagttatggttggcg
attgtcacttgcgctgtaatctgttgcggagtagtactgatcgcataca

126001 catcattgtaagatctccttttgacggggagggaactcatcgtccacag
tgatgtcgtttttaaatttgcgtctacgggtgcaacagacttgcatttca
tgggtccctgtgatgaacattatcagagcagtgatgaaaatgctcataa
acatggacggtcaaagtcgaaaatgaaaacccaacttagcgtatgaagta
aaacaaaagacataccctaacgatatagcattttttctctgtagatcaa
acgaacaaatttatcttgagatataagagagcgttgtcttctgaaaacga
aaaactaatgttttgggttcaagacattctctgtcactatcaaacttt
cctagtactgaaggtgaagtcagcagctttagggcagtttgataacaga
tgcatatcctaagagaccataaccgaaataaaatgttcaataaatcact
aaacaaaatccaaataccaagcattgtctcatcaaattcaactcattta

126501 atacaggtaattgcaataaacattcaaaagagctatcaaaagtattcaat
tcacaacagctaaatctaactgcatactccaaatcaaagtcataaaaa
accatcatccaaaaatgtaccgaaagaacacaaaaatgaagtttacgggc
tcaccttgcatggcagccgctgggcagacgaacccatcattggccagagt
aatattgaaagcagcagcagcagcagcagcagcagcagcagcagcagc
cgatcaatttgacctgaacctgcactctgggccaagtccccgcgctcta
accgccgtcttcagatcggcatcggataaccagcaaaaacgggtgccgtt
tccgacacccccagggatactagtaccgtcggccaaaacagcattggagg
cagagaccaagtactcgttggtaaacccccacaaacacctccacgac
ttgagatcgtcgagccggttgttgagcactgtgagaaccgactcgaactt

127001 gtaaggctgctgtttcgggtctttgacctcgggcgggagctggggctcgg
cagtatagcgtaggacaggaagacgccgttgacagagtcggaggccggc
gccggagctgcggcgggttggctgggcgagtgagaagggcatgatcgtgag
gataatgagaagattgagggccatgatcggaggccagttcctggagcccg
agtccatggtcttttgggttcttgggggtccccaaagtttggatttttga
atattggagaacacagatagggtttatataatgtgagcactgttctgtgtgcg
cttttttcaaacgcgtagagagagagagagagagagagagagagtaacgtt
ggattgtaaatatgtaataactaataactgcatagtggaagctgtaatttg
acaatatttttgggtttaaagttttttaaataatagttttatgtatttgtt
gcccgtctttgaaatgattcaattccaaaagtctttgaaatgtttgaaacgaa

127501 atggcaagaagtgcgtgggtgaaaaatactagtgtcaagactgctgctt
gtgtttgttttgagcgtgaaaaggtccttttgtctttgtctcaacggaac
ccttcccctctcttcttatttgattttgaaatctgtgttcgttaatacgggt
gagggagctgatcgtttcggatagaaatgttctcttatacaacttgagtt
aatagacaattaaattatgtatttatcagtcgtacgtttaggacaatgggt
tacaagtgaacgctcaacaaaagtcattgcgatcaaatctcgcatacca
ctgtctgtcaataaagcattctatggtctgataatttgaagagaat

MC-20SSR96C4For

MC-20SSR96C4Rev

gataacgctttgaaagaattctgtttcatgtaaataattgggtaaccagc
128001 aaagtcaatgcatgtcaatatgttgccgggcacgctccttttgtctgaagct
catagataccaaaacgtggctcactcgtttgcatattggacctatagtta
tgtaattggactctgcctggttttccacaatctgaccttgccggccctc
acatgggcccattttgaaaaggcctctacatgcagattatgttgagaggt
acaatttgggttggatcaaaatcttaactcgtttatgtttagtttgattt
taaaccgctttcgggcccgttagaattaagtaaacctcgtttcaacccc
accgatctcaattgggttatgtttgggtgagttgggctgagttgggtt
gatgcttctgtccatgtaactcaaatctaaccgaagtatattctga
tttctttcaacataactgacttcgcttcatattatacaagctacgta
taaatttgggtagaattacttgaaatttgggtgggtgattttcatttggg
caatgattgatgctattttagtttgacttttgtgtgaatgttaattatga
128501 aatgaattctaactagatttgacattatctagttcaatttttgaaaaat
ctttcagaataaataagatattaaaagattgaacttgataaagttgagca
accggaataaaccttagacctaatgaaacttgatcccgaagagcctgca
tgaatccaaacccaacccaagtccaaattgtgatagttgaatattagaga
tcaattgacattcaatccgcccgaccgagttgcacctctaaatcttatt
gtaacatataaataatcttgaaaaggatttttattgttacattaacccatg
aagatgacctatacagagagagctatcccaatagcttagattaataaa
ttatttggattaaacaataaacatccaattatgagattttgtttggaccct
tggatgtaagaggtttgttatgaccatcaactgaaagaaaattttgttgt
aaaattataactaatctcaagttatatacctttttattccccccccaaaa
129001 aaaggttgttattattattatttttagggggagggagaggttcgaaacc
agtgtcaaagttagaatttacaataatgaggcatgaataagaagcttac
aaaaactaaataacttttatttctaattggcaacgacatattctcatat
tttaaacatcgtatcacaacatcatttacaatagaatcaaatcacttt
tcattacatataaccaagttattattcgaccattgactttcattcatttt
tcattgtatacaaccaaattatcattcaatcattgattttccacacgatg
tgaatccttcacaatggtgatcactaaaaagctatccctattgtggata
ggcagtagttttggatgttttaattaaaaaataataaaaatgggtctct
ctattttgcttttcttagctattacagggctatataaaaaattcaaacggg
agatattcgaagtgggtgggtcagtgaggcaatgtctccatgttggc
129501 tctgcaactgttcaaaactactacaataattaagaagagagcaagtttta
aatccgagacacatagtagatagccaacatcctatatctaccattcgacc
acatgcatctcaagttgtacacctaaaaaactattgattatgctacagtg
ttagtactaggttagaagatgaacagacagaagagttagacattacatca
catgagacctctataagacactcaacggccgtcaactcacaaggggtca
acctttaccagtttaacactaaaaacggacatcaactattgaacttgaag
gaggacaatgacactttttagatttgggacaaatctaccttagtggaatg
cattaactcgttaaaattgacctctcattccaccagcaaataaatgggtg
attaggacataatagcaaacaggaaatctattacgtcatcattactata
tatccgctcacgtgagtttttgcatacacacaaacggatattacaattg
130001 taaagaaaaagttctctgtctattagtcogttgatcaaacacaatcttt
ataaaacaaaagatttacttacacacacaaactaacatcaatttctcaatc
gtcaaccgtttgtatgcatgtgaacgagtaactcttatgtatcaatcaag
gatggaaccaagagttttattacaagatgggctagcctgaatgtggttttg
tttagtggtatcaatcaaccactgccaatccttaattgtgcaacattagc
aaggttataaaagacgctaagcttttagtcaggatgactaggcgggggct
atgcaggtctaggcgccttcttaattttcaaacgcctaatgattaatcg
ggacagtggttaactgcctagcatttaggcgagggcctaggtggagctaga
cagagatttttaggacagtgaaacttagtaacatctatcaatggccgcca
accactcacatataaattgattaaaaataatgctatcatgttactatacta
130501 ggcatgtggtttgagaataaagcgtgaatgaagttctcttcaaagaaatc
aaaagggggccaagaatattatagcaagtgagactgcctaccacacatgca
actacctgtccgataagtatatgcatgatcaaatcgatcgagttcccct
tttacttttgtttgaaagcctatgagagatgcttgccttaataatgtca
aggtagagggggccaattagtaactcacttttggatcggaatatttaattt
tcatgcataatcctatgatttcatgaagtgatgtacgtattacaacgcta
atggcaatttagcaaggtgtataatcacaatttgccttatttagtttgc
atccgtgctcttgagaggtatttgcacgtatttaaagtatatttttatat
gagtataatatttattactatttactatttacttgaagtgtgatgatca
tgcttttttgggtttgcaagtgaactgaaccaacctatataatgaagca
131001 agaattaccgggttttggatttggatttggcatcaaactaggctatata
ttcacacttgaagaacttaagaaggttaattagttattttatcatgata
ttcacactctgaggcgaagtacgtagaagctacaggccattaaagaaat
aacaaaagtcaaaaggtgatcacgatcagtgaaagccaagattctttcaa
ttttatatttgaagctttttgtaattaaacattaattttcattttctaat
tagaatttaatatgcacatcatgatctttttacctttaattagaaagtac

cggccgggttgggattcttttgaaacctttgtgtaataaccagcaaatctg
aactagcaatatattccactgcatgccttttttttaaatggtgaggg
aactctttcaaaagtagaataatgtttgtatgaatgtttgtctatctcac
131501 agttaaagttaactcttgcgtaacattgtgcaaaaaatgaagggtg
gagagtttatggagagtcctcatcagcatttttttttaacaaatgatagt
attatatttaaatttttaaatcatggggacgggaattcgaacttttata
caaacagtaaaatcgcctactatctaactaacttgactaaacttaaat
acgcgcattataacaagtaattaaagaatgtaaacatggtttttcttt
tgttggaatgcaaacatggttagtattccacttaatacttatacatcga
gatgaaactaaagagacaaaataaaaattaaagagacaaaaggccgacat
ctttcaatgatctttgtatagactagtgtccaccggagttaatttggg
aaaggagaaggcattatttttaatgtttgtgaccactaattgtccac
tcaaacagccagctggaaagtccatgtggagattgtgatcttataatat
132001 ttccaaaatacatattacatgcatggcacaaaatataatgtgatttgatt
aaattaaaaattaaaaattaaaaacctattgagaataattattaccagt
aacatttatgtaaattcataggaatttatcttagcttaagaagttggtg
ttcgtcggttggcaccaatggcaagagaggctcattcatttccctcctag
tcatgccaaactcattgcacagaaatcaaacacaagatttcggataaaaa
ataatgattcttaatcatttgagttataaatttcgtatattctcggat
tttgattaacatcaagagagaaattccgtgacggggaagtattttattat
atztatgttatacagaagaatcgggttctgtttgtttattctgtgcac
ttatagttagccgcgtgaaaaatgcatcactgcactcaccatattgctg
gtggccatgtgatcgacctgcactgcactgcattattattattatt
132501 attacttccaacatttaactactaaatggagactctgaaattactcgac
ctcttaaccctaattactcagtggtcgggggaaaaaaaaaatggtataag
agctaattagctgagtaataaccaacttaaaataataacaatctaattta
ttaattttcgaacaaaaagagccgaatctctttattttctgtgatgcca
gctggcccatatctttgtgctctttaacaactcattccttatcacttatt
tcttgtcaagaagtattttgtattctgacgaatatgaacatgtatataga
tataatacaatgttgttccctacgtgtataacttagacaaaattaatcttt
ctcctccattttataaccagattaagaatgtatgtgatgttccattgacca
tgcagcttttttaattatatacatgtgtatataacgaacaaacattct
aagatcatctccaatcattgggttaaaacttaaaatgtttaacctataaa
atthaggttttaaccataaacaattttctcctccaacccttacgagtt
aaatttttagcccaaggttattaaagaatgaatttaggataatttttct
taaagtaaccttttaaaaaaattatataatgtacactatcataatttaat
tttatgaatattttaacctaaaaatatttagattccaataaatattgaaa
attcactaaaattgacatcataaaaactcgtagaacactatgaaagattatg
aaacacagaaaaataaatttttttaattacttttagacgttggatttaatt
tggaccgttagatcttttttttactgttgaatttgatcatattagattt
tagccgttgaattcgagctcggtagccgggatcctctagagtcga

132896

Appendix 2

Mal d 1 sequences added in the alignment of Chapter 2

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

53		Mal d 1.0703	FL	Not yet
54	Mal d 1.08	Mal d 1.0801.01	PM	AY789259
55		Mal d 1.0801.02	PM	AY789261
56		Mal d 1.0801.01	FL	Not yet
57	Mal d 1.09	Mal d 1.0901	PM	AY789262
58		Mal d 1.0902	GD	AY822720
59		Mal d 1.0903	GD	AY822721
60		Mal d 1.0903	FL	Not yet
61	Mal d 1.10	Mal d 1.1001	FL	Not yet
62	Mal d 1.11	Mal d 1.1101	FL	Not yet
63	Mal d 1.12	Mal d 1.1201	FL	Not yet

*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 Sthatis 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