

Genomic organisation of the *Mal d 1* gene cluster on linkage group 16 in apple

Giulia Pagliarani · Roberta Paris · Anna Rosa Iorio ·
Stefano Tartarini · Stefano Del Duca · Paul Arens ·
Sander Peters · Eric van de Weg

Received: 6 September 2010 / Accepted: 14 May 2011 / Published online: 24 June 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract European populations exhibit progressive sensitisation to food allergens, and apples are one of the foods for which sensitisation is observed most frequently. Apple cultivars vary greatly in their allergenic characteristics, and a better understanding of the genetic basis of low allergenicity may therefore allow allergic individuals to increase their fruit intake. *Mal d 1* is considered to be a major apple allergen, and this protein is encoded by the most

complex allergen gene family. Not all *Mal d 1* members are likely to be involved in allergenicity. Therefore, additional knowledge about the existence and characteristics of the different *Mal d 1* genes is required. In the present study, we investigated the genomic organisation of the *Mal d 1* gene cluster in linkage group 16 of apple through the sequencing of two bacterial artificial chromosome clones. The results provided new information on the composition of this family with respect to the number and orientation of functional and pseudogenes and their physical distances. The results were compared with the apple and peach genome sequences that have recently been made available. A broad analysis of the whole apple genome revealed the presence of new genes in this family, and a complete list of the observed *Mal d 1* genes is supplied. Thus, this study provides an important contribution towards a better understanding of the genetics of the *Mal d 1* family and establishes the basis for further research on allelic diversity among cultivars in relation to variation in allergenicity.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-011-9588-4) contains supplementary material, which is available to authorized users.

G. Pagliarani (✉) · R. Paris · S. Tartarini
Department of Fruit Tree and Woody Plant Sciences,
University of Bologna, Viale Fanin 46, 40127 Bologna,
Italy
e-mail: gpagliarani@agrsci.unibo.it

A. R. Iorio · S. Del Duca
Department of Biology es, University of Bologna,
Via Imerio 42, 40126 Bologna, Italy

G. Pagliarani · P. Arens · E. van de Weg
Plant Breeding, Plant Research International,
Wageningen University and Research Center,
Droevendaalsesteeg 1, 6708 PB Wageningen,
The Netherlands

S. Peters
Greenomics, Plant Research International,
Wageningen University and Research Center,
Droevendaalsesteeg 1, 6708 PB Wageningen,
The Netherlands

Keywords Allergen · PR-10 · *Malus × domestica* ·
Pru p 1 · *Prunus* · BAC clone

Introduction

The apple (*Malus × domestica*, Borkh., family Rosaceae, tribe Pyreae) is one of the most important fruit

species worldwide from both nutritional and economic points of view.

Consumption of apples is widely considered to have a positive effect on human health. However, European populations show progressive sensitisation to food allergens, and apples are one of the foods for which sensitisation is observed most frequently, ranking forth in this respect out of 24 foods examined in an extensive pan-European survey (Burney et al. 2010). Apple cultivars vary greatly in their allergenic characteristics, with some eliciting no or only slight responses in allergic patients following oral provocations (Bolhaar et al. 2005; Kootstra et al. 2007; Van der Maas and Schenk 2009; Vlieg-Boerstra et al. 2010). Understanding the genetic basis of low allergenicity may support the provision of low allergenicity apples and thereby allow allergic individuals to increase their fruit intake. Developing an allergy to apple often follows sensitisation to birch pollen due to cross-reactivity between the major birch pollen allergen, Bet v 1, and the apple allergen Mal d 1 (Fritsch et al. 1998). The cross-reactivity of these allergens is due to their high amino acid sequence and structural similarity (Vanek-Krebitz et al. 1995). *Bet v 1*-like genes have also been found in other Rosaceae fruit species, thus explaining the cross-reactivity that frequently occurs among these fruits (Fernandez-Rivas et al. 2006).

Mal d 1 is considered to be a major allergen in apple, and this protein is encoded by a highly complex gene family consisting of at least 21 members, of which 17 are organised in two clusters in the homoeologous linkage groups (LGs) 13 and 16, while *Mal d 1.05* is located on LG6 (Gao et al. 2005). The remaining three of these genes have been identified (Beuning et al. 2004; Gao et al. 2005) but have not yet been mapped. *Mal d 1* genes are classified into four subfamilies based on DNA sequence similarity, which coincide with differences in the length of the intron in these genes. *Mal d 1* coding sequences exhibit high levels of similarity: 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).

The expression of genes within the *Mal d 1* family has been found to exhibit tissue specificity (Puehringer et al. 2003; Beuning et al. 2004; Botton et al. 2008), similar to the *Bet v 1*-like gene families in other species (Lebel et al. 2010). Only a limited number of

Mal d 1 proteins and mRNAs have been traced back in apple fruit to date, despite the use of different methods such as mass spectrometry (Helsper et al. 2002), expressed sequence tag (EST) sequencing (Beuning et al. 2004) and targeted reverse-transcription PCR (Puehringer et al. 2003; Botton et al. 2008; Pagliarani et al. 2009). This indicates functional specialisation of the different gene family members, despite the high level of sequence and structural similarity among Mal d 1-like proteins. 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 immunoglobulin E (IgE) (Ma et al. 2006), suggesting that even the variants of these genes may differ in allergenicity.

Evidence for the functional specialisation of Bet v 1-like proteins has also been found in other crops, as different catalytic properties and ligand-binding specificities were found within this family in birch and peach (Markovic-Housley et al. 2003; Zubini et al. 2009), suggesting different functions associated with slight changes in three-dimensional structure. Differentiation into tissue-specific forms has resulted in Mal d 1 isoallergens that are not expressed in fruit and, thus, are not involved in allergic reactions. The observed functional differentiation among isoallergens and variants suggests the need for in-depth knowledge about the existence and performance of individual proteins and, thus, of individual genes and alleles. Performing such in-depth genetic studies has become feasible due to recent advances related to the genetics and genomics of apple, such as the many recently developed genetic tools, including molecular markers, bacterial artificial chromosome (BAC) libraries and collections of ESTs, as well as the draft genome sequence of the diploid Golden Delicious (GD) apple cultivar that has recently become available (Velasco et al. 2010). This sequence showed that apple has a relatively small genome size of approximately 750 Mb per haploid genome (Gasic et al. 2009; Höfer and Meister 2010). It also clarified the duplication patterns of the apple genome, supporting the hypothesis of an autopolyploid origin. Additionally, a strong collinearity was demonstrated between large segments of homoeologous chromosomes in the apple genome, i.e. between chromosomes 3 and 11, 5 and 10 or 13 and 16, only part of which was previously documented (Maliepaard et al. 1998; Liebhard et al. 2002; Chen et al. 2008). Moreover,

analysis of the markers mapped in more than one Rosaceae species revealed patterns of macro- and micro-synteny among *Malus* and *Prunus* (Dirlewanger et al. 2004; Chen et al. 2008; Sargent et al. 2009), and, because of this, genomic advances in peach are also of use in apple genetics (Shulaev et al. 2008). For instance, eight *Bet v 1*-like genes have been identified in peach. These genes, designated as *Pru p 1* genes, were mapped on chromosome G1 (Chen et al. 2008) in a region known to be syntenic to LG13 and 16 of apple (Dirlewanger et al. 2004).

The Mal d 1 and Bet v 1 proteins are not only known as major allergens (Breiteneder and Ebner 2000), but also as pathogenesis-related (PR-10) proteins that are thought to play a key role in selective biotic and abiotic stress responses (van Loon et al. 2006). Because of this, knowledge of these proteins is becoming available from different scientific disciplines. At the structural level, PR-10 proteins of different species share a highly conserved P-loop, a glycine-rich region (GXGGXGXXXK) that is involved in the formation of a large hydrophobic cavity in the protein, which opens toward the exterior and likely acts as a ligand-binding site. In fact, PR-10 proteins are capable to bind several different types of ligands, such as phytohormones, fatty acids and flavonoids (Gajhede et al. 1996; Spangfort et al. 1997; Koistinen et al. 2005) and to play a role in the storage and transport of biologically important molecules (Markovic-Housley et al. 2003; Mogensen et al. 2007). Moreover, the Mal d 1-orthologous Fra a 1 proteins in strawberry were found to be involved in the storage and/or transport of intermediates of the flavonoid pathway, suggesting that PR-10 proteins also play a role in fruit pigmentation (Muñoz et al. 2010). For PR-10 proteins with putative RNase activity, the P-loop is thought to act as a binding site for nucleotides (Radauer et al. 2008).

In this study, we aimed to identify the *Mal d 1* genes present in the cluster on LG16 and to elucidate their genomic organisation. Although we were aware of the imminent publication of the GD genome sequence, we decided to follow the approach based on the Sanger sequencing of BAC clones, anticipating difficulties in the assembly of sequences from a region containing an extensive family of highly similar genes by a next-generation whole-genome sequencing approach. We screened a BAC library successfully used for gene mapping and cloning

(Vinatzer et al. 1998, 2001, Cova et al. 2010; Galli et al. 2010), which was derived from the Vf scab-resistant cultivar Florina. Florina has recently been reported to have an intermediate level of allergenicity after a skin prick test analysis (Ricci et al. 2010). Two *Mal d 1*-containing BAC clones were selected and fully sequenced. This cluster was chosen for its greater diversity of *Mal d 1* genes compared to the homoeologous LG13 (Gao et al. 2005) and because some of its members seem to cause variation in allergenicity among apple cultivars (Gao et al. 2008). The results were then compared with the apple and peach genome sequences.

Materials and methods

PCR-based screening of the BAC library

A BAC library from the cultivar Florina (Vinatzer et al. 1998) already available at the Department of Fruit Tree and Woody Plant Sciences (University of Bologna) was used. The library consists of 36,864 BAC clones with an average insert size of 120 kb, representing approximately 5 × apple haploid genome equivalents. A bi-dimensional pooling method was performed following that of Cova (2008), and plasmids from the BAC clone pools were extracted using the alkaline extraction procedure (Birnboim and Doly 1979).

PCR-based screening of the library was carried out with four primer pairs specifically designed for the four *Mal d 1* subfamilies and a general primer pair for *Mal d 1* designed based on consensus regions (table provided as Online Resource 1) using Primer3 software (<http://frodo.wi.mit.edu/primer3/>). Positive BAC clones were picked from the library, singularised and tested by colony-PCR with the same primers used for the screening. All PCR amplifications were performed in a 17.5- μ l volume containing 200 ng of DNA from the BAC library pools or bacterial cells from singularised colonies and 0.1 μ M primers, 1.5 mM MgCl₂, 100 μ M dNTPs, 0.5 Units DNA Polymerase (Fisher Molecular Biology, Hampton, NH, USA) and 1 × 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, NY, USA) after electrophoresis on 1.5% (w/v) agarose gels and ethidium bromide staining.

BAC preparation and analysis

DNA from the positive BAC clones was prepared using a Maxi Prep protocol adapted as described by Untergasser (2006). Purified plasmid DNA (approximately 20 µg) was digested with 2 U *EcoRI* overnight at 37°C. Digested DNA fragments were loaded onto 1% Ultra Pure agarose gels (Lonza, Basel, Switzerland) and electrophoresed at 35 V overnight. Images of the *EcoRI*-digested DNA fragments of BAC clones were used to identify overlapping BAC clones.

To identify BAC clones that contained particular sets of *Mal d 1* genes, specific primer pairs were designed for each known isoallergen gene (table provided as Online Resource 2). In some cases, the primer specificity was increased by adding deliberate mismatches at the 3' end of a primer, as described by Gao et al. (2005). Because of the high sequence similarity, only one primer pair was designed for all the *Mal d 1.03* genes. All PCR amplifications were performed in a 20-µl volume containing 50 ng of plasmid DNA, 0.1 µM gene-specific primers, 1.5 mM MgCl₂, 100 µM dNTPs, 0.5 U of AmpliTaq Gold[®] DNA Polymerase (Applied Biosystems, Foster City, CA, USA) and 1 × 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) and a final extension of 7 min at 72°C.

BAC sequencing, sequence annotation and phylogenetic analysis

Two BAC clones carrying different *Mal d 1* genes were subjected to Sanger sequencing at Macrogen Inc. (Korea) after the creation of 2-kb and 6-kb (6×) insert libraries. The sequences were assembled by Greenomics (The Netherlands). Gaps between contigs were filled by direct sequencing performed with specific primer pairs (table provided as Online Resource 3) designed using PrimerSelect software (Lasergene[®] v8.0) for the ends of the contigs. Single-

run sequencing was performed by Bio-Fab Research srl (Pomezia, Italy). Final assembly was carried out manually with SeqMan software (Lasergene[®] v8.0). BAC sequences were annotated using the gene prediction programs GENESCAN (Burge and Karlin 1997) and GeneMark[™] (<http://opal.biology.gatech.edu/GeneMark/>). Predicted open reading frames (ORFs) were searched for similarity to known proteins using BLASTP software (Altschul et al. 1990) against the non-redundant protein database of the National Center for Biotechnology Information (NCBI) with a cut-off lower than E^{-15} . BLASTX was used to further verify the correspondence between the predicted ORFs and proteins in the databases. Phylogenetic analysis of the *Mal d 1* sequences was performed by searching for similarity to known nucleotide sequences using BLASTN with a cut-off lower than E^{-25} and by alignments of the coding sequences and predicted amino acid sequences of the *Mal d 1* members performed using ClustalW by Megalign (Lasergene[®] v8.0) with a gap open penalty of 10 and an extension penalty of 0.1. A phylogenetic tree corresponding to the amino acid sequence alignment was generated through Megalign (Lasergene[®] v8.0).

Anchoring BAC clones on the genetic map

The two assembled BAC sequences were analysed with Tandem Repeats Finder software (<http://tandem.bu.edu/trf/trf.html>), and 10 simple sequence repeat (SSR) markers were developed (table provided as Online Resource 4) using Primer3 software (<http://frodo.wi.mit.edu/primer3/>). PCR amplification and gel electrophoresis were performed following the method of Gianfranceschi et al. (1998). These SSRs were mapped in a Durello di Forlì × Fiesta population (population size $n = 174$) using JoinMap 3.0[®] (Van Ooijen and Voorrips 2001) with the Kosambi mapping function. The LOD value chosen for grouping the LG16 markers was equal to 7. The final image of LG16 was generated with MapChart (Voorrips 2001).

In-silico analysis of *Mal d 1* genes in the Golden Delicious draft genome sequence

The Golden Delicious (GD) draft genome sequence (Apple Genome v0.1 contigs) was searched for *Mal d 1* genes via the BLASTN server of the Genome

Database for Rosaceae (<http://www.rosaceae.org>) using all known *Mal d 1* sequences as queries and an expected value of 0.01. In order to distinguish loci and alleles and to check the intron/exon structure prediction of the genes downloaded from the Gbrowser, all the sequences with a cut-off lower than E^{-25} were carefully checked against the NCBI database using BLASTN software (Altschul et al. 1990) and aligned with known *Mal d 1* nucleotide sequences using ClustalW by Megalign (Lasergene[®] v8.0) with a gap open penalty of 15 and an extension penalty of 6.66. Moreover, to be sure of identifying all the *Mal d 1* sequences in the cluster on LG16, all the predicted ORFs in the chr16: 10733976–11466541 region (contigs MDC012403.580 to MDC003279.119) were manually annotated by searching for similarity with known sequences in the database using BLASTP software with a cut-off lower than E^{-15} . *Mal d 1* loci were classified as new when their predicted protein sequence showed less than 95% similarity to previously identified sequences. A phylogenetic tree corresponding to the coding sequences (CDS) alignment of the new genes was generated through Megalign (Lasergene[®] v8.0). They were named according to official allergen nomenclature (King et al. 1995), such that loci with over 95% DNA sequence similarity were denoted by adding a capital letter to their isoallergen name, thus following Gao et al. (2005). Next, their physical positions were compared with those of the BAC sequences of Florina.

In-silico analysis of *Pru p 1* gene cluster sequence on linkage group G1 of peach

The peach genome sequence (Peach Genome V0.1 scaffolds) was searched for *Pru p 1* genes via the BLASTN server of the Genome Database for Rosaceae (<http://www.rosaceae.org>) using the *Pru p 1.01* sequence EU424239 as a query. The region on the top of G1 was further examined for its collinearity with LG16 of apple. In particular, the region from position 9,450,000 to position 9,650,000 of scaffold 1 was downloaded from Gbrowser (http://www.rosaceae.org/gb/gbrowse/prunus_persica/) and analysed with the gene prediction programs GENESCAN and GeneMark[™]. The predicted ORFs were manually annotated using BLASTX and BLASTN software with a cut-off lower than E^{-25} . To identify any

putative coding sequences that the gene prediction methods failed to detect, fragments of 1,000 bp were also used as inputs for further BLASTX searches against the NCBI non-redundant database.

Results and discussion

BAC library screening and analysis of positive clones

The screening of the Florina BAC library resulted in 20 positive clones for *Mal d 1*, designated MC-1 to MC-20. The putative location of these BAC clones in the apple genome was determined by amplifying them with locus-specific *Mal d 1* primers (table provided as Online Resource 2) for which the position on linkage maps has been published (figure provided as Online Resource 5). Five clones yielded amplicons with primers specific for *Mal d 1* genes from LG16 (MC-1, -12, -14, -16, -20) and, hence, should be derived from chromosome 16. Similarly, eleven and four clones exhibited amplification of *Mal d 1* genes in LG13 and LG6, respectively. The *Mal d 1.04*-specific primers did not produce an amplicon in any BAC clone, which could be due to a failure during the screening step or to a lack of representation within the BAC library. The overlap among the BAC clones was assessed according to their digestion profiles. Next, clones were divided into groups that did not show any overlap: groups I and II for BACs from LG16, groups III and IV for LG13 and groups V and VI for LG6 (Online Resource 5). One BAC for LG13 remained ungrouped as its digestion pattern did not indicate an overlap with any of the other clones (data not shown). Some clones produced amplicons with primers from different linkage groups, suggesting the presence of as-yet-unidentified but highly similar isoallergen genes.

BAC clone sequencing and sequence annotation

Two non-overlapping but representative BAC clones belonging to LG16 were chosen for sequencing: MC-12 from group I because it was the clone that produced the greatest number of amplicons and MC-20 from group II because estimates based on *NotI* digestion showed that it was the longest of the group (data not shown). The first assembly of the

single reads gave four contigs for each clone. Further direct BAC sequencing steps were performed by primer walking (primers given in Online Resource 4), thereby enabling the assembly of unique full-length sequences for both clones, which were deposited in GenBank as FN823234 (MC-12, 125,046 nt) and FN823235 (MC-20, 132,896 nt).

The gene prediction software revealed an average coding percentage of 49.6% (~128 kb) between the two clones, with the presence of 56 putative ORFs, from ORF1 to ORF34 in clone MC-12 and from ORF35 to ORF56 in MC-20. Their putative annotations were recorded based on BLASTP scores (Table 1): 17 ORFs showed similarity with known proteins, 11 with retrotransposon elements, 8 were similar to hypothetical proteins of *Vitis vinifera*, and 20 had no significant match. Most of the ORFs putatively coding for known proteins (13/17) were similar to *Mal d 1* sequences, of which nine were in MC-12 and four in MC-20. Further analysis of these 13 sequences with BLASTN (Table 2) revealed five *Mal d 1*-like sequences that were not previously known to be located in LG16. For two of these new sequences, ESTs have previously been identified as *Mal d 1 m* (AY428588) and *Mal d 1 n* (AY428589), and their full-length genomic sequence is reported herein. *Mal d 1 m* (ORF10) includes an intron of 475 nt, which is longer than any of the other known *Mal d 1* introns. According to the official allergen nomenclature (King et al. 1995), we proposed to name this gene *Mal d 1.10* (FN823234). The sequence for *Mal d 1 n* (ORF20) has an intron of 208 nt. We proposed to name it *Mal d 1.11A* (FN823234). The ESTs for *Mal d 1.10* came from young fruit tissue, while the ESTs for *Mal d 1.11A* from mature fruit pulp tissue and partially senescent leaves. As they are expressed in fruit, they may both be relevant to allergies. The third new gene (ORF17) showed no perfect matches with other known *Mal d 1* sequences. It has 79% protein similarity with the Pru p 1.06A allergen of *Prunus dulcis* × *Prunus persica* and only 74% protein similarity with *Mal d 1.03G*, which was the highest similarity to any available *Malus* sequence (Tables 1, 2). This ORF was therefore classified as a putative new apple allergen gene belonging to the *Mal d 1* family, which we named *Mal d 1.12* (FN823234), and which has an intron of 375 nt. It must be further studied from the transcriptional point of view to confirm that it is functional and not a pseudogene.

The fourth new gene (ORF25) has the highest similarity at the nucleotide level (94%) to *Mal d Ips2.02* (AY827730), but is truncated after 283 nt. As this similarity is less than 95%, it is likely to be a new gene and we here classify it as a new pseudogene named *Mal d Ips3*. The fifth new sequence (ORF45) shows a high similarity with *Mal d 1.03G* (AY822733), with five non-synonymous single nucleotide polymorphisms (SNPs) corresponding to N110S and H132Q substitutions, thereby representing a new *Mal d 1.03G* allele for Florina at this locus (allele 02). Interestingly, this is the first *Mal d 1.03*-like gene found in LG16, as all the others have been mapped to LG13.

Apart from these five sequences, eight other genes previously mapped to LG16 were found in the BAC sequences (Fig. 1). In MC-12, we found one new allele for the pseudogene *Mal d Ips2*: *Mal d Ips2.04*. ORF29 showed 99% similarity with *Mal d Ips2.02* (AY827730), but with five SNPs being observed among these two sequences. Like the *Mal d Ips2.02* database sequence, ORF29 also contains a stop codon. ORF6 was similar to allele 02 of *Mal d 1.06A* (AY827697) but with one synonymous SNP in the coding region. Because this allele of *Mal d 1.06A* has been associated with low allergenicity (Gao et al. 2008), the intermediate level of allergenicity for Florina reported by Ricci et al. (2010) led us to hypothesise that this cultivar is heterozygous for this locus. Additionally, ORF13 and ORF28 showed only one synonymous SNP in the coding region compared to the isoallergen genes *Mal d 1.0201* (AY827654) and *Mal d 1.06C02* (AY827725), respectively; ORF23 was identical to *Mal d 1.06B02* (AY827712). Only intronless *Mal d 1*-like sequences with the conserved full length of 480 nt were found in BAC clone MC-20. ORF46 showed 99% similarity to *Mal d 1.0701* (AY822717) and is thus a new allele for this gene from Florina: *Mal d 1.0703*. It exhibits four synonymous and one non-synonymous SNP that caused the amino acid substitution K71R. ORF47 proved to be identical to *Mal d 1.0903* (AY822721), and ORF49 was identical to *Mal d 1.0801* (AY822719).

Examining the allelic composition of the *Mal d 1* genes on MC-12 from Florina suggests that this haplotype comes from the chromosome of Florina's mother, Jonathan (Gao et al. 2008). When we examined MC-20, however, no hypothesis could be made because this clone contains only intronless *Mal*

Table 1 Description of the predicted ORFs in the MC-12 and MC-20 sequences (FN823234 and FN823235, respectively) analysed with BLASTP software

ORF	Gene features		BLASTP results		E value	Description ^b	Source	Conserved domains	Mat d 1 isoform ^c
	Protein ID	Dir ^a	Start-end position (from T7-end)	Most similar sequence ID					
<i>On BAC clone MC-12</i>									
ORF1	CBL94129	↓	520–1891	AAT85304	4.00E-92	Reverse transcriptase (RNA-dependent DNA polymerase) domain containing protein	<i>Oryza sativa (japonica)</i> cultivar-group	RVE superfamily (Integrase domain)	
ORF2	CBL94130	↓	2837–3941			Hypothetical protein with no significant similarities			
ORF3	CBL94131	↓	5458–5673			Hypothetical protein with no significant similarities			
ORF4	CBL94132	↓	12339–13767	AAO45751	3.00E-11	Gag-protease polyprotein (retrotransposon protein)	<i>Cucumis melo</i>	Gag proteins from LTR retrotransposons	
ORF5	CBL94133	↓	14727–16116	ABA98191		Retrotransposon protein, putative, Ty3-gypsy subclass	<i>Oryza sativa (japonica)</i> cultivar-group	RT_LTR (reverse transcriptases from LTR retrotransposons); RNase H superfamily	
ORF6	CBL94134	↑	19701–20302	AAAX18298	1.00E-84	Major allergen Mal d 1.06A02	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.06A
ORF7	CBL94135	↓	22203–22903			Hypothetical protein with no significant similarities			
ORF8	CBL94136	↑	24478–25500			Hypothetical protein with no significant similarities			
ORF9	CBL94137	↓	29603–30001			Hypothetical protein with no significant similarities			
ORF10	CBL94138	↓	30450–31410	AA800052	1.00E-86	Mal d 1-like	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.10*
ORF11	CBL94139	↑	33218–37315	AAP52115	8.00E-57	Retrotransposon protein, putative, Ty1-copia subclass	<i>Oryza sativa (japonica)</i> cultivar-group	RVE superfamily (integrase domain); RVT_2 superfamily (reverse transcriptases domain)	
ORF12	CBL94140	↑	39819–42609			Hypothetical protein with no significant similarities			
ORF13	CBL94141	↓	45547–46197	AAD26548	2.00E-85	Major allergen Mal d 1	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.02
ORF14	CBL94142	↑	50864–51162			Hypothetical protein with no significant similarities			
ORF15	CBL94143	↑	52015–52470	CAN81839	9.00E-42	Hypothetical protein	<i>Vitis vinifera</i>		
ORF16	CBL94144	↑	53617–54123	AAG51247	2.00E-39	copa-type polyprotein, putative (retrotransposon protein)	<i>Arabidopsis thaliana</i>		
ORF17	CBL94145	↓	59754–60614	ACE80952	4.00E-46	Putative allergen Pru p 1.06A	<i>Prunus dulcis</i> x <i>Prunus persica</i>	Polyketide_cyc 2 superfamily	1.12*
ORF18	CBL94146	↓	63792–65322			Hypothetical protein with no significant similarities			
ORF19	CBL94147	↑	66153–66580			Hypothetical protein with no significant similarities			
ORF20	CBL94148	↓	68018–68719	AA800053	6.00E-92	Mal d 1-like	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.11A*

Table 1 continued

ORF	Protein ID	Gene features		BLASTP results		Description ^b	Source	Conserved domains	Mat d 1 isoform ^c
		Dir ^a	Start-end position (from T7-end)	Most similar sequence ID	E value				
ORF21	CBL94149	↑	70217–70631			Hypothetical protein with no significant similarities			
ORF22	CBL94150	↑	71128–72667			Hypothetical protein with no significant similarities			
ORF23	CBL94151	↓	73771–74403	AAAX18303	2.00E-84	Major allergen Mal d 1.06B	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.06B
ORF24	CBL94152	↑	75956–76767			Hypothetical protein with no significant similarities			
ORF25		↓	78571–78853			Mal d 1 pseudogene			
ORF26	CBL94154	↑	79009–83451	XP_002280852	0.0	Hypothetical protein	<i>Vitis vinifera</i>	RVE superfamily (integrase domain); RVT_2 superfamily (reverse transcriptases domain)	
ORF27	CBL94155	↓	87386–89959	ABA95230	1.00E-51	Retrotransposon protein, putative	<i>Oryza sativa (japonica cultivar-group)</i>	RVT_2 superfamily (reverse transcriptases domain)	
ORF28	CBL94156	↓	92219–92826	AAAX18306	5.00E-85	Major allergen Mal d 1.06C	<i>Malus domestica</i>	Polyketide_cyc 2 superfamily	1.06C
ORF29		↓	94875–95346			Mal d 1 pseudogene			
ORF30	CBL94158	↓	95952–96348			Hypothetical protein with no significant similarities			
ORF31	CBL94159	↓	96909–106394	ABA94905	2.00E-17	Retrotransposon protein	<i>Oryza sativa (japonica cultivar-group)</i>	Transposase domain	
ORF32	CBL94160	↓	107170–109602	CAB39638	4.00E-33	RNA-directed DNA polymerase-like protein	<i>Arabidopsis thaliana</i>	RT_like superfamily (reverse transcriptases domain)	
ORF33	CBL94161	↓	114358–115592	CAN79553	4.00E-71	Hypothetical protein	<i>Vitis vinifera</i>	HMA, heavy-metal-associated domain	
ORF34	CBL94162	↑	116800–121966	NP_196685	0.0	Transducin family protein/WD-40 repeat family protein	<i>Arabidopsis thaliana</i>	WD40 domain	
<i>On BAC clone MC-20</i>									
ORF35	CBL94163	↓	06–1982	ABD28426	4E-179	RNA-directed DNA polymerase (Reverse transcriptase)	<i>Medicago truncatula</i>	RT_nLTR (reverse transcriptase from non-LTR retrotransposons)	
ORF36	CBL94164	↑	2242–5241			Hypothetical protein with no significant similarities			
ORF37	CBL94165	↓	5409–5900	ABB00038	4E-47	Reverse transcriptase family member	<i>Glycine max</i>		
ORF38	CBL94166	↓	9631–11289	CAN74865	1E-35	Hypothetical protein	<i>Vitis vinifera</i>	PMD, plant mobile domain	

Table 1 continued

ORF	Protein ID	Gene features		BLASTP results		Description ^b	Source	Conserved domains	Mat d 1 isoform ^c
		Dj ^a	Start-end position (from T7-end)	Most similar sequence ID	E value				
ORF39	CBL94167	↓	13865–15433			Hypothetical protein with no significant similarities	<i>Vitis vinifera</i>	RVE superfamily (integrase domain)	
ORF40	CBL94168	↑	16118–17227	CAN69026	7E-09	Hypothetical protein	<i>Vitis vinifera</i>		
ORF41	CBL94169	↑	17718–23267	CAN62188	7E-38	Hypothetical protein	<i>Vitis vinifera</i>		
ORF42	CBL94170	↓	24714–25663			Hypothetical protein with no significant similarities			
ORF43	CBL94171	↑	27463–30549	CAN67317	3E-79	Hypothetical protein	<i>Vitis vinifera</i>	RVT_2 (reverse transcriptase) superfamily	
ORF44	CBL94172	↓	36000–40765	ABA98193	0.0	Retrotransposon protein, putative, Ty3-gypsy subclass	<i>Oryza sativa</i> (<i>japonica</i> cultivar-group)	RT_JTR: reverse transcriptases (RTs) from retrotransposons; RVP_2: single domain aspartyl proteases from retrotransposons; Retrotrans_gag: Gag or capsid-like proteins from LTR retrotransposons; Chromo (CHR)romatin Organisation Modifier domain; RVT_1: reverse transcriptase domain	I.03G
ORF45	CBL94173	↑	43900–44379	AAAX18324	7E-85	Major allergen Mal d 1.03G	<i>Malus domestica</i>	PolyKetide_cyc2 superfamily	I.07
ORF46	CBL94174	↓	60900–61379	AAAX18307	5E-86	Major allergen Mal d 1.07	<i>Malus domestica</i>	PolyKetide_cyc2 superfamily	I.09
ORF47	CBL94175	↓	63263–63742	AAAX20996	5E-87	Mal d 1.0903	<i>Malus domestica</i>	PolyKetide_cyc2 superfamily	I.08
ORF48	CBL94176	↑	66110–72637			Hypothetical protein with no significant similarities			
ORF49	CBL94177	↓	74151–74630	AAAX18309	6E-87	Major allergen Mal d 1.08	<i>Malus domestica</i>	PolyKetide_cyc2 superfamily	
ORF50	CBL94178	↑	75936–77515			Hypothetical protein with no significant similarities			
ORF51	CBL94179	↓	77779–79873	XP_002271465	5E-73	Hypothetical protein	<i>Vitis vinifera</i>	DUF789: conserved domain found in several plant proteins of unknown function	
ORF52	CBL94180	↑	83182–85403	BAB10341	3E-41	Drought-induced protein Di19-like protein	<i>Arabidopsis thaliana</i>	Drought-induced 19 protein (Di19) domain	

Table 1 continued

ORF	Protein ID	Gene features		BLASTP results		Description ^b	Source	Conserved domains	Mal d 1 isoform ^c
		Dir ^a	Start-end position (from T7-end)	Most similar sequence ID	E value				
ORF53	CBL94181	↑	87649–91977			Hypothetical protein with no significant similarities			
ORF54	CBL94182	↓	102592–103322			Hypothetical protein with no significant similarities			
ORF55	CBL94183	↓	114066–117194	XP_002519174	0.0	Sensor histidine kinase, putative	<i>Ricinus communis</i>	REC Response_reg, Response regulator receiver domain, myb-like DNA-binding domain, SHAQKYF class	
ORF56	CBL94184	↑	117937–126807	XP_002519175	0.0	Protein COBRA precursor, putative	<i>Ricinus communis</i>	COBRA superfamily	

^a Direction of the gene. The arrows are directed downward when the gene is directed from T7-end to Sp6-end and vice versa

^b ORF similar to *Mal d 1* genes are indicated in *bold* and pseudo-genes are highlighted in *gray*

^c Mal d 1 isoform names proposed following the official allergen nomenclature (King et al. 1995) and Gao et al. (2005)

* *Mal d 1* genes for which the genomic sequence is reported in this work for the first time

d 1 genes, and for these genes, no information is available about the alleles in Jonathan.

The phylogenetic tree of the total of 20 *Mal d 1* coding sequences (Online Resource 6) shows that the new gene *Mal d 1.10* fits in the clade of *Mal d 1.04* and *1.05* (subfamily II), while two of the other new genes (*Mal d 1.11A* and *1.12*) remain ungrouped. Based on examination of the gene structures, it must be noted that the length of the first exon is conserved in all the isoallergen genes, including *Mal d 1.10*, *1.11A* and *1.12*, despite their different intron lengths (Online Resource 6). Regarding the predicted amino acid sequences, most genes of the Bet v 1-like proteins encode predicted polypeptides of 151–162 amino acids (Liu and Ekramoddoullah, 2006), and all *Mal d 1* genes encode proteins of 159 amino acids, except the three newly identified functional genes, with *Mal d 1.10* and *Mal d 1.12* coding for 161 and *Mal d 1.11A* for 163 amino acids.

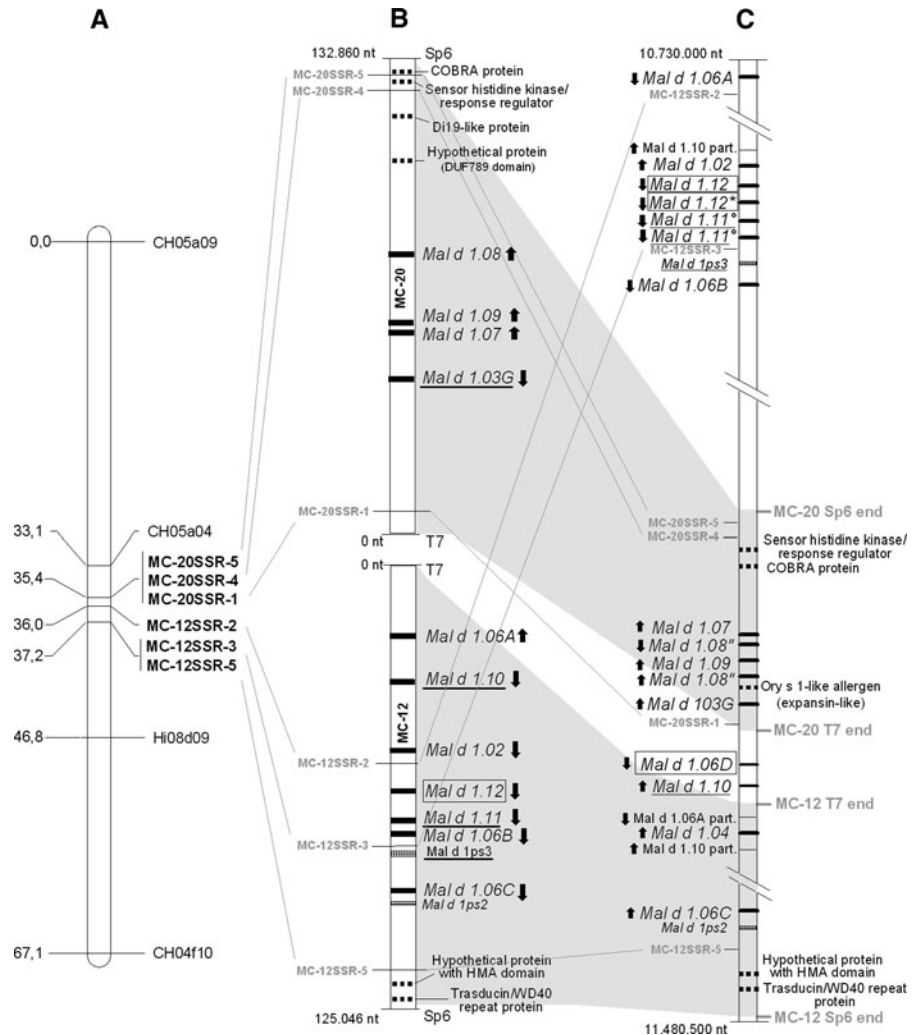
Genomic organisation of *Mal d 1* genes on LG16

The full sequences of the two BAC clones allowed us to study the genomic organisation of the *Mal d 1* gene cluster on LG16. All four intronless isoallergens were found in MC-20 over a region of approximately 30 kb. The seven intron-containing isoallergen genes in MC-12 were located in a region of approximately 75 kb. The distance between the isoallergen genes ranged from 10 to 15 kb, with the two exceptions being the ~5 kb between *Mal d 1.11A* and *Mal d 1.06B* and the ~2 kb between *Mal d 1.07* and *Mal d 1.09*, which are the two closest isoallergens in this cluster. All the isoallergen genes inside each clone are oriented in the same direction (from the T7 to the Sp6 end), except for *Mal d 1.06A* in MC-12 and *Mal d 1.03G* in MC-20 (Fig. 1). Their opposite orientation may have been caused by genomic re-arrangements during evolution, which is an assumption supported by the facts that *Mal d 1.06A* is positioned far from the other *Mal d 1.06* genes and that *Mal d 1.03G* is the only gene of the *Mal d 1.03* group in this region. Knowledge of cluster organisation within multigene families can be used to better understand gene expression data because intergenic regions and gene orientation are reported to affect gene expression levels (Bondino and Valle, 2009).

Table 2 Description of the *Mal d I*-like sequences in the MC-12 and MC-20 sequences (FN823234 and FN823235, respectively) analysed with BLASTN software

ORF	<i>Mal d I</i> gene/allele	<i>Mal d I</i> genes features				BLASTN results				
		Dir ^a	Full length (nt)	CDS (nt)	Intron/Exon	Most similar sequence ID	Description ^b	SNPs in CDS	SNPs in intron	E value
<i>On BAC clone MC-12</i>										
ORF6	<i>Mal d I.06A02</i>	↑	602	480	Exon I:184; Intron:122; Exon II: 296	AY827697	<i>Malus x domestica</i> clone 220903B10 <i>Mal d I.06A02 (Mal d I.06A) gene</i>	1	0	0.0
ORF10	<i>Mal d I.1001</i>	↓	961	486	Exon I:184; Intron:475; Exon II: 302	AY428588	<i>Malus x domestica</i> clone 1 m Mal d I-like mRNA	0	Not available	1.00E-156
ORF13	<i>Mal d I.0201</i>	↓	651	480	Exon I:184; Intron:171; Exon II: 296	AY827654	<i>Malus x domestica</i> clone 250803B10 <i>Mal d I.0201 (Mal d I.02) gene</i>	1	0	0.0
ORF17	<i>Mal d I.1201</i>	↓	861	486	Exon I:184; Intron:375; Exon II: 302	AY822733	<i>Malus x domestica</i> clone 231103G3 <i>Mal d I.03G01 (Mal d I.03G) gene</i>	–	–	3.00E-25
ORF20	<i>Mal d I.11A01</i>	↓	702	495	Exon I:184; Intron:208; Exon II: 310	AY428589	<i>Malus x domestica</i> clone In Mal d I-like mRNA	0	Not available	4.00E-160
ORF23	<i>Mal d I.06B02</i>	↓	633	480	Exon I:184; Intron:153; Exon II:296	AY827712	<i>Malus x domestica</i> clone 220903F10 <i>Mal d I.06B02 (Mal d I.06B) gene</i>	0	0	0.0
ORF25	<i>Mal d Ips301</i>	↓	283	–	With stop codons and truncated	AY827730	<i>Malus x domestica</i> clone 220903A7 Mal d Ips2 pseudogene	–	–	2.00E-113
ORF28	<i>Mal d I.06C02</i>	↓	608	480	Exon I:184; Intron:128; Exon II:296	AY827725	<i>Malus x domestica</i> clone 231103F11 <i>Mal d I.06C02 (Mal d I.06C) gene</i>	1	0	0.0
ORF29	<i>Mal d Ips204</i>	↓	470	–	With stop codons	AY827730	<i>Malus x domestica</i> clone 220903A7 Mal d Ips2 pseudogene	–	–	0.0
<i>On BAC clone MC-20</i>										
ORF45	<i>Mal d I.03G02</i>	↑	480	480	Intronless	AY822733	<i>Malus x domestica</i> clone 231103G3 <i>Mal d I.03G01 (Mal d I.03G) gene</i>	5	–	0.0
ORF46	<i>Mal d I.0703</i>	↓	480	480	Intronless	AY822717	<i>Malus x domestica</i> clone 231103B3 <i>Mal d I.0701 (Mal d I.07) gene</i>	5	–	0.0
ORF47	<i>Mal d I.0903</i>	↓	480	480	Intronless	AY822721	<i>Malus x domestica</i> clone 231103C1 <i>Mal d I.0903 (Mal d I.09) gene</i>	0	–	0.0
ORF49	<i>Mal d I.0801</i>	↓	480	480	Intronless	AY822719	<i>Malus x domestica</i> clone 231103F3 <i>Mal d I.0801 (Mal d I.08) gene</i>	0	–	0.0

Fig. 1 Genomic organisation of *Mal d 1* gene cluster on LG16. **a** Genetic map of Durello di Forlì LG16. SSRs developed based on the sequences of the two BACs are indicated in **bold**. **b** Physical map of the two BAC clones from cv Florina, MC-12 and MC-20. **c** Physical map of *Mal d 1* cluster on LG16 from GD draft genome sequence. In (b) and (c), the *Mal d 1* gene positions are indicated as **black bars**, the pseudogene positions as **striped bars** and the other genes in the cluster as **dotted bars**. The isoallergen genes previously known but located for the first time or designated by a new name are **underlined**, and the new isoallergen genes are indicated in **boxes**. The **arrows** indicate gene orientation; ° and " are identical sequences; **Mal d 1.12* sequence, but with a gap of 45 bp



Retrotransposons and other genes in the cluster

Four ORFs similar to other genes with known functions were found in the *Mal d 1* cluster, as listed in Table 1. ORF34 in MC-12 encoded for a protein of the transducin family containing a WD-40 domain (Stacey et al. 1999). In MC-20, ORF52 was found to be highly similar to the drought-induced Di19-like protein in *Arabidopsis* (Rodriguez Milla et al. 2006); ORF55 was similar to a conserved, multi-domain protein consisting of a sensor histidine kinase/response regulator (Yamada and Shiro 2008); and ORF56 was similar to a COBRA protein (Roudier et al. 2005). Conserved domains were also found in ORFs classified as hypothetical proteins (Table 1): ORF33 on MC-12 was similar to a hypothetical

protein of *Vitis vinifera* with a heavy metal-associated (HMA) domain (Zhou et al. 2009), and ORF51 on MC-20 was classified as a hypothetical protein of *V. vinifera* with DUF789 conserved domain.

A considerable accumulation of retrotransposon elements was observed in both MC-12 and MC-20 sequences, including a reverse transcriptase (RNA-dependent DNA polymerase), nucleotide-binding site, RNase H, RNA/DNA hybrid-binding site, integrase core domain, CHROMatin Organisation MODifier domain (CHROMO domain), plant mobile domain and retrotransposon gag proteins (Table 1). Retrotransposons represent the most abundant type of transposable element in the apple genome (Velasco et al. 2010). They play an important role in the plasticity of eukaryotic genomes because they can

drive gene duplications by inadvertently carrying copies of genes during transposition events and/or by facilitating unequal crossovers (Hancock 2005; Madlung et al. 2005). Therefore, it is possible to assume that these elements are involved in the increase and evolution of the *Mal d 1* family. The retention of many highly similar genes in plant genomes, such as *Mal d 1* genes in apple, has long been thought to supply the raw genetic material needed for plant adaptation and evolution (Lynch and Conery 2000). The retention of original function, the loss of function by the creation of pseudogenes, the acquisition of novel functions via neo-functionalisation and the partitioning of the ancestral gene function by sub-functionalisation might be included among the evolutionary strategies of plants (Force et al. 1999; Moore and Purugganan 2005). Because the *Mal d 1* proteins are not only known as major allergens but also as PR-10 proteins thought to be involved in biotic and abiotic stress responses and in fruit pigmentation (van Loon et al. 2006; Muñoz et al. 2010), the evolution of these genes may be driven by a stress-mediated mode of selection (Wagner et al. 2008), as was proposed for the PR-10 proteins of *Vitis vinifera* by Lebel et al. (2010).

Anchoring the physical map to the genetic map of the *Mal d 1* cluster on LG16

Six of the ten SSRs developed based on BAC clone sequences (Online Resource 4) were polymorphic and mapped on LG16 of the Durello di Forlì × Fiesta map (Fig. 1). This anchored the physical and the genetic maps, further confirming the location of these two BAC clones on LG16. Moreover, due to the presence of two progenies with DNA recombination between MC12SSR-2 and MC12SSR-3, the orientation of MC-12 within the LG could be determined. The approximately 30-kb physical distance between these two SSRs corresponded to a genetic distance of ± 1.2 cM (Fig. 1a, b). Because no recombinant individuals were found for SSRs developed in MC-20, no information that was useful for determining its orientation was obtained. Combining the entire physical map constructed with the BAC sequences and the genetic map obtained with SSR markers, the complete genetic distance of the region was approximately 2 cM, corresponding to a physical region of at least 260 kb. The difference between physical and genetic distance in this short region thus ranged from

25 to more than 130 kb/cM. Because the estimated overall genome size is approximately 750 Mb (Gasic et al. 2009; Höfer and Meister 2010), and the estimated length of the genetic map is approximately 1,400 cM, the average physical-to-genetic distance ratio for the apple genome was estimated to be 530 kb/cM (Patocchi et al. 1999; Cevik and King 2002). The physical/genetic distance ratio in the *Mal d 1* region proved to be below the estimated average, suggesting a higher recombination frequency in this region. As it is known that stress increases recombination frequency (Lucht et al. 2002), this finding is a further corroboration of the involvement of stress-mediated selection in the evolution of the *Mal d 1* gene family.

An overall agreement was found between the order of *Mal d 1* genes in the genetic map of LG16 (Gao et al. 2005) and in the physical map (Fig. 1), with only few discrepancies. First, *Mal d 1.09* was previously mapped proximal to SSR marker CH05a04, *Mal d 1.07* and *Mal d 1.08* but showed up between these last two genes on the physical map. Second, marker CH05a04 was not found in the two BAC sequences. This absence would be inconsistent with the order of the genetic map of Gao et al. (2005), but is feasible based on the order in the physical map, allowing the CH05a04 marker to be located proximal to the Sp6 end of the MC-20 BAC sequence. Third, *Mal d 1.02* was mapped 0.4 cM from the group of *Mal d 1.06A*, *1.06B* and *1.06C*, but it was located between *Mal d 1.06A* and *Mal d 1.06B* on the physical map. These discrepancies might be due to a few scoring errors of the marker in the mapping population and/or the presence of missing values, which negatively affect the accuracy of mapping. Finally, *Mal d 1.04* co-localised with *Mal d 1.02*, *Mal d 1.07* and *Mal d 1.08* on the genetic map but could not be localised to the physical map. One explanation for this difference might be that the two BAC clones do not overlap each other, and as these three genes are divided over the two clones, it is possible that *Mal d 1.04* maps in the gap between MC-12 and MC-20.

Genome-wide analysis of the Golden Delicious genome for *Mal d 1* genes and comparison with Florina BAC sequences

The recently released draft genome sequence of domesticated apple (Velasco et al. 2010) provided a

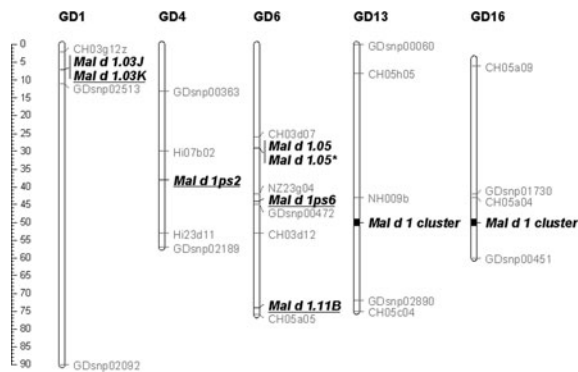


Fig. 2 Schematic overview of *Mal d 1* allergen gene positions in the apple genetic map. Genetic positions of *Mal d 1* loci are estimated through retrieval of their physical location in the GD whole genome relative to reference marker sequences. Genetic positions of reference markers are indicated according to Supplementary Figure 9 in Velasco et al. (2010). *Mal d 1* loci in new genomic regions are *underlined*; **Mal d 1.05* in tandem duplication

further opportunity to study the genomic organisation of the *Mal d 1* gene family in apple and to compare the Florina BAC sequences with the corresponding region of Golden Delicious (GD). A total of 49 *Mal d 1*-like sequences were retrieved from the whole GD genome and located on five different chromosomes, including the homeologous LG16 (region from position 10,737,669 to 11,436,747) and LG13 (from position 9,828,373 to 14,597,852), LG 6, LG 4 and LG 1 (Fig. 2). Ten of these sequences were additional to those recently retrieved by Yang et al. (2011) (Online Resources 7 and 8). Note that Yang et al. (2011) used different gene denotations for *Mal d 1.10* and following genes, probably because they were not aware that prior to the acceptance of their paper genes *Mal d 1.10* to *Mal d 1.12* had already been assigned through a NCBI deposited sequence (FN823234).

A cluster of 21 *Mal d 1*-like sequences was found on LG16 (Online Resource 7), including the 11 complete coding sequences and the two pseudogenes retrieved in the Florina BAC sequences. In addition to these, the *Mal d 1.04* sequence was found in the cluster (GD-ORF 40, Online Resource 7), as was expected based on the genetic map reported by Gao et al. (2005). The remaining seven ORFs coded for new sequences. Of these, GD-ORF35 had 97% amino acid and 98% nucleotide similarity with *Mal d 1.06A*. However, as *Mal d 1.06A* was already covered by GD-ORF1, this sequence has to be classified as a new

locus on LG16, which we propose to designate *Mal d 1.06D*. This locus was not found in the Florina BAC sequences, and, again, this can be explained based on the presence of the gap between the two BAC clones, with GD-ORF35 being located between the T7 ends of MC-12 and MC-20 (Fig. 1 and the table in Online Resource 7). The remaining six *Mal d 1*-like sequences on LG16 were complete or truncated duplications of previously assigned *Mal d 1* genes. These duplications may be explained by a further expansion of the cluster in the GD genotype with respect to Florina or by errors during the assembly of this region. The sequencing of the GD genome has also shown the presence of a relatively large number of repeated sequences, which hampered its assembly and were difficult to anchor unambiguously (Velasco et al. 2010). Moreover, the heterozygosity of the GD genome may have artificially increased the number of duplications in the assembled sequence through the inclusion of homologous and thereby highly similar sequence fragments during the assembly process, as happened in the construction of a physical map of the heterozygous grapevine Cabernet Sauvignon (Moroldo et al. 2008). The case of *Mal d 1.08* sequences (GD-ORF24 and 26) is a representative example of this because the two sequences are identical except for one Y (C/T) degenerate base in the first sequence, representing the SNP between the two alleles of GD. Finally, the annotation of GD-ORF28 as an Ory s 1-like sequence was particularly interesting. Ory s 1 is an expansin-like protein classified as the major allergen of *Orzya sativa* (<http://www.allergome.org>), which means that it may be a new apple allergen in the cluster of *Mal d 1* allergens.

Some discrepancies also emerged from the comparison of the entire *Mal d 1* gene cluster in LG16 of GD with the Florina BAC sequences. The MC-12 and MC-20 BAC ends were traced back in the GD genome, along with the SSR markers developed based on the Florina sequences (Fig. 1 and Online Resource 7), but while the distance between the Sp6 and T7 ends of BAC MC-20 is almost conserved in the GD sequence, the region corresponding to the MC-12 BAC is shorter in GD than in Florina. In particular, the size of the Florina BAC is 125 kb, while the same GD region spanning the Sp6-T7 ends is 8 kb shorter (117 kb). This difference is mainly due to the different localisation of the *Mal d 1.06A-Mal d 1.06B* block, which is located within the

MC-12 BAC clone in the Florina sequence and in a completely different region in the GD sequence (chr16:10737669..10806399) (Fig. 1). The deviation of the position of the markers MC12SSR-2 and MC12SSR-3 in the GD sequence further substantiated the erroneous alignment of that region of the GD genome sequence (Fig. 1). Other differences in the order and orientation of the *Mal d 1* genes were found; for instance, the position of *Mal d 1.10* in the GD genome is located within the MC-12 and MC-20 T7 ends (Fig. 1). Such discrepancies are expected to occur for genomic regions of complex gene families following a whole-genome shotgun approach of hundreds of Mb. The sequence length of individual BAC clones of approximately 120 kb reduces the complexity by decreasing the number of highly similar sequences and by confining their occurrence to a small defined genomic window. The complexity of the assembly is further reduced by the use of longer individual reads and the lower frequency of sequencing errors provided by Sanger sequencing compared to next-generation sequencing platforms. BAC clone Sanger sequencing can thus be useful for validating and improving the quality of genome sequences harbouring complex gene families.

Twenty-one of the 28 *Mal d 1*-like sequences in other apple chromosomes were found in the homologous cluster in LG13 (Fig. 2 and Online Resource 8). On this LG, beside the seven genes previously reported by Gao et al. (2005) and the seven additional genes reported by Yang et al. (2011), two new genes were identified and were named as *Mal d 1.03H* and *Mal d 1.11B* following King et al. (1995) and Gao et al. (2005). Among the five remaining sequences, three were complete duplications of the previously identified genes *Mal d 1.01*, *Mal d 1.03B* and *Mal d 1.03D*, and two were truncated duplications of *Mal d 1.01* and *Mal d 1.03F*. On chromosome 6, *Mal d 1.05* was found to be fully duplicated within a region of about 8 kb. Moreover, for the first time, two additional *Mal d 1*-like sequences were located on the bottom of chromosome 6, far from the *Mal d 1.05* sequences: a new pseudogene at approximately 5 Mb and a duplication of *Mal d 1.11B* (GD-ORF52) identical to GD-ORF73 on LG13 at approximately 15 Mb. Unlike Yang et al. (2011), we decided to delay the assignment of different names to all these duplicated genes until their existence has been validated and

assembly errors have been ruled out. On chromosome 1, two clustered *Mal d 1* loci were found, thus confirming Yang et al. (2011). Their genomic sequences showed highest similarities to the *Mal d 1.03* gene family, in particular to *Mal d 1.03D* (Online Resource 9), because of which we named them *Mal d 1.03J* and *Mal d 1.03K*, whereas Yang et al. (2011) related them to the genetically more distinct *Mal d 1.07* gene. Finally, a pseudogene was found alone on chromosome 4 (Fig. 2).

Comparison of the *Mal d 1* cluster on LG16 and the *Pru p 1* cluster on G1

The availability of the peach genome sequence (www.peachgenome.org) enabled us to perform a computational search of the orthologous *Mal d 1* genes in peach. In particular, the *Pru p 1* region on scaffold 1 (G1) was analysed because of its collinearity with the *Mal d 1* cluster on LG16 (Dirlewanger et al. 2004; Chen et al. 2008). A total of 36 predicted ORFs were retrieved in a 200-kb segment and designated as PpORF1-36, among which 20 *Pru p 1* isoallergen genes were identified. Six out of these 20 genes were new (Online Resources 10 and 11). The discovery of these new isoallergen genes suggests that a further increase in the number of *Pru p 1* genes in the peach genome may be expected.

The comparison of the apple and peach *Bet v 1*-like gene clusters indicated a certain level of microsynteny between *Malus* and *Prunus* species. In fact, in addition to the synteny at the whole-genome level previously assessed through markers, a striking level of gene content and of order conservation was retrieved for large blocks of sequence (Table 1, Online Resources 10 and 11). For example, the apple gene coding for a COBRA protein (ORF56) corresponded to peach PpORF1 and the apple gene for the transducin/WD40 repeat protein (ORF34) to PpORF30 (Table 1 and Online Resource 10). This order of the external genes in the peach cluster validated the orientation of the BAC clone MC-20, as shown in Fig. 1 and Online Resource 11. An exact correspondence between *Mal d 1* and *Pru p 1* genes was difficult to establish due to some inter-genera variability of these sequences, but the number of complete *Mal d 1* genes was less than the number of complete *Pru p 1* for this cluster (13 in *Malus* and 18 in *Prunus*). Moreover, a clear difference in the

dimensions of the two clusters became apparent, with at least 168 kb between the first and the last *Mal d 1* genes of the cluster and approximately 80 kb between the first and the last *Pru p 1* genes. These differences thus reflect the different evolution of the two gene families. As reported in the literature for genomes such as those of *Arabidopsis* or *Carica*, duplication and progressive gene loss following each polyploidisation can contribute to the evolution of genomes. Moreover, the gene retention rate appears to differ substantially in different lineages (Paterson et al. 2010). A lower retention of genes in the apple LG16 than in the peach cluster G1 would make sense given the presence of the homoeologous cluster in LG13. In fact, the number of *Bet v 1*-like complete coding sequences becomes much higher in apple than in peach if all the *Mal d 1* genes in LG16 and LG13 are taken into account (18 in *Prunus* and 31 in *Malus*).

Deduced amino acid sequences of *Mal d 1* genes

The alignment of all the different isoform sequences shown in Fig. 3 indicated that the P-loop region (glycine-rich loop, GXGXGXKK) is highly conserved among *Mal d 1* proteins, being reported for all the representative members of *Bet v 1*-like proteins in different species (Spangfort et al. 1997). Only a few substitutions appeared in this domain among the predicted *Mal d 1* proteins: a lysine replaced by a glutamine in *Mal d 1.08* and by a methionine in *Mal d 1.09*; the third glycine is replaced by glutamic acid in *Mal d 1.11A* and *Mal d 1.11B* and by arginine in *Mal d 1.12*. Unspecified amino acids are present in the P-loop of two new isoforms, *Mal d 1.03H* and *Mal d 1.03J*. It would be interesting to know whether any of these differences in amino acid composition affects the functionality of these isoforms and the allergenicity of apples.

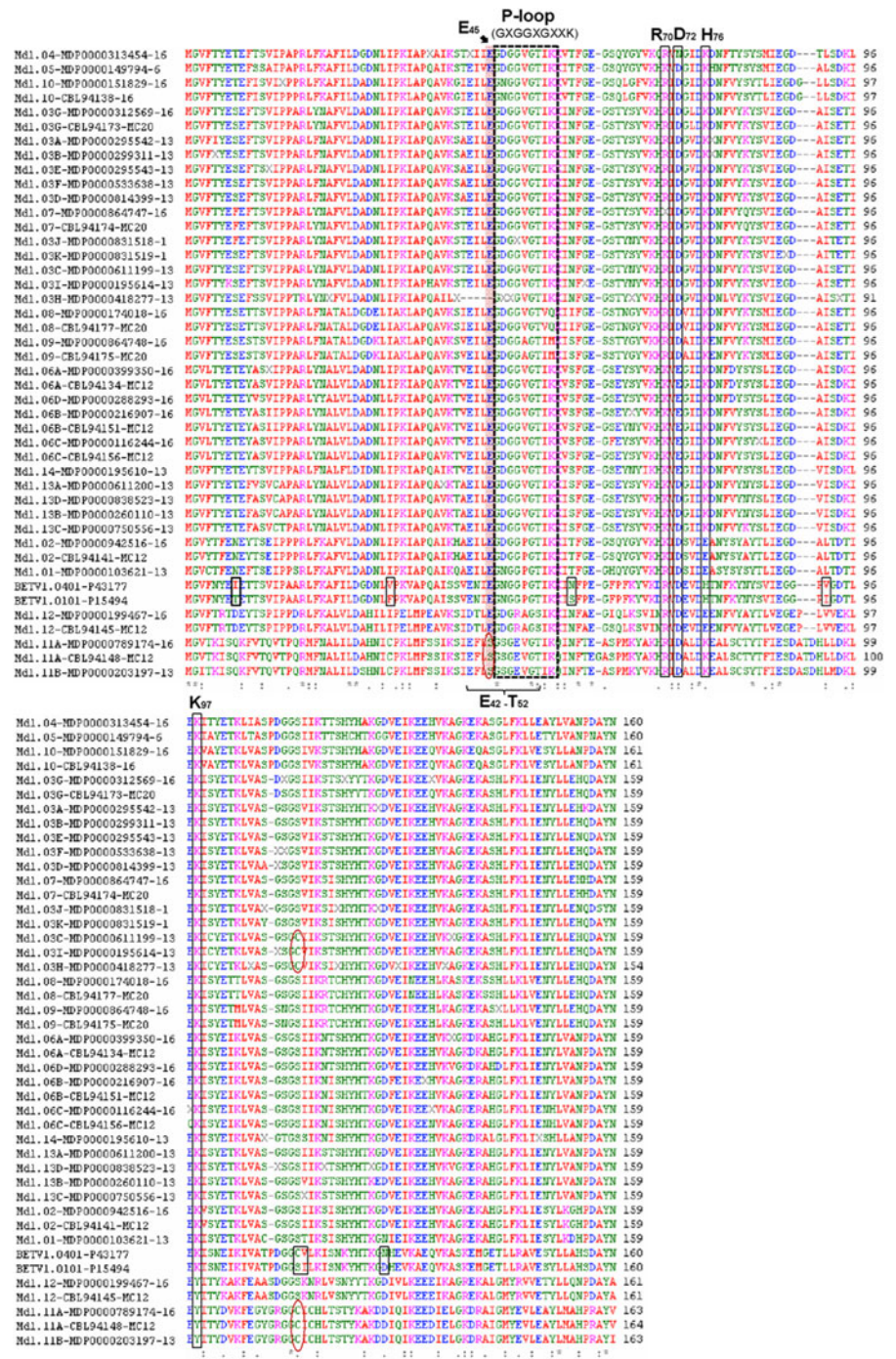
Given that little is known about the immunological properties of the various isoforms of both the *Bet v 1* and *Mal d 1* families, the ability of these different proteins to induce allergic responses is largely unknown. A high sequence similarity between proteins increases the chance of shared epitopes and thus of recognising the same antibody, whereas a single amino acid change may drastically influence the extent of allergenicity by increasing or decreasing the ability to bind that antibody. Wagner et al. (2008) showed that a few amino acid changes (from four to

nine changes) on the surface of three *Bet v 1* isoforms caused a difference in IgE induction. While *Bet v 1.0401* and *Bet v 1.1001* do not induce IgE synthesis, *Bet v 1.0101* can stimulate this process, and its associated IgE only partially cross-reacts with the two previous isoforms. The inclusion of *Bet v 1.01* and *Bet 1.04* in the *Mal d 1* alignment (Fig. 3) allowed us to examine the differences between the two birch and the apple isoforms. The S113C change of *Bet v 1.04* with respect to *Bet v 1.01* has previously been identified as being important for the ability of *Bet v 1.04* to form aggregates and to create a type of protection against IgE binding (Zaborsky et al. 2010). Given the similarity between *Mal d 1* and *Bet v 1* at the nucleotide and amino acid levels, an oligomerisation is also likely for some *Mal d 1* proteins. In particular, *Mal d 1.11A/B* and *Mal d 1.03C/I/H* are the best candidates for this because of the presence of C113. Moreover, a previous crystallographic study of the *Bet v 1*-antibody complex classified several residues as important for the antigenic surface, including the positions E42, E45, T52, R70, D72, H76 and K97. In particular, position 45 seems crucial because it is located in the core of the binding pocket of *Bet v 1.01* and fits well into the groove on the antibody surface (Ghosh and Bhattacharya 2007). The residue at position 45 is conserved both in apple and birch, but not in *Mal d 1.11A* and *B* (Fig. 3), further supporting its putative hypoallergenicity. Evidently, it will be of considerable interest to gain deeper insight into the specific immunological and biochemical features of these highly similar proteins, especially for the putative hypoallergenic isoforms *Mal d 1.11A* and *B*.

Conclusions

The BAC sequencing and genome-wide analysis conducted in this study increased our knowledge of the genomic organisation of the *Mal d 1* gene family in apple and led to the identification of 31 complete *Mal d 1* genes able to encode for different allergen isoforms in the apple genome. Our BAC-based sequence of Florina showed substantial agreement to the GD whole genome, although several differences were found in the assembly of the sequence, such as the occurrence of duplicated and truncated *Mal d 1* genes. The appropriateness of our assembly

Fig. 3 Alignment of predicted amino acid sequences of *Mal d 1* isoforms. *Mal d 1* sequences were retrieved from the BAC clones sequences and from Gbrowser of the apple genome sequence. Two *Bet v 1* isoforms, *Bet v 1.01* and *Bet v 1.04*, are also included. The sequences are indicated with the isoform name followed by the ID number and the LG in which they are located. The P-loop region is indicated by the dashed box; substitutions between *Bet v 1* sequences are indicated as small boxes; the position 45 is highlighted in red and the other amino acids putatively important for IgE recognition are within brackets or large boxes. Important amino acid substitutions are shown as circles



was validated by its consistency with genetic linkage maps, which means that other *Mal d 1* loci in the GD sequence that have been newly discovered *in silico* have to be validated. In fact, for in-depth genetic studies on complex gene families, it may be wise to

first accurately validate genome sequence data using other tools.

The insights gained here related to the constitution of the *Mal d 1* gene family and the fine physical positioning of its members is critical for further

clarifying the genetic basis of allergenicity in apple through expression and association studies, and may ultimately contribute to increasing the availability of low allergenicity apple fruit that can be consumed by individuals who are otherwise allergic to apple.

Acknowledgments This research was carried out as part of 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). Disclaimer: the views and opinions expressed in this publication are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission. The authors would like to thank Amy Claire Learmonth for providing some of the primer pairs in the table provided as Online Resource 2 and Professor Lars Ove Dragsted for useful comments on the draft of the manuscript. Finally, authors would like to thank the International Peach Genome Initiative (IPGI) for providing the peach genome sequence (available at www.peachgenome.org) and the Italian and the US apple genome projects for providing the draft of the apple genome sequence (http://www.rosaceae.org/gb/gbrowse/malus_x_domestica/).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- 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 HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513–1523
- Bolhaar ST, van de Weg WE, van Ree R, Gonzalez-Mancebo E, Zuidmeer L, Bruijnzeel-Koomen CAFM et al (2005) In vivo assessment with prick-to-prick testing and double-blind, placebo-controlled food challenge of allergenicity of apple cultivars. *J Allergy Clin Immunol* 116:1080–1086
- Bondino HG, Valle EM (2009) A small intergenic region drives exclusive tissue-specific expression of the adjacent genes in *Arabidopsis thaliana*. *BMC Mol Biol* 10:95
- Botton A, Lezzer P, Dorigoni A, Barcaccia G, Rupert B, Ramina A (2008) Genetic and environmental factors affecting allergen-related gene expression in apple fruit (*Malus domestica* L. Borkh). *J Agr Food Chem* 56:6707–6716
- Breiteneder H, Ebner C (2000) Molecular and biochemical classification of plant-derived food allergens. *J Allergy Clin Immunol* 106:27–36
- Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78–94
- Burney P, Summers C, Chinn S, Hooper R, van Ree R, Lidholm J (2010) Prevalence and distribution of sensitization to foods in the European community respiratory health survey: a EuroPrevall analysis. *Allergy* 65:1182–1188
- Cevik V, King GJ (2002) Resolving the aphid resistance locus Sd-1 on a BAC contig within a sub-telomeric region of *Malus* linkage group 7. *Genome* 45:939–945
- 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* 9:543
- Cova V (2008) Isolamento di marcatori microsatelliti strettamente associati al gene di resistenza a ticchiolatura Vm in melo. PhD Dissertation, Università di Bologna
- Cova V, Paris R, Passerotti S, Zini E, Gessler C, Pertot I, Loi N, Musetti R, Komjanc M (2010) Mapping and functional analysis of four apple receptor-like protein kinases related to LRPKml in HcrVf2-transgenic and wild-type apple plants. *Tree Genet Genomes* 6:389–403
- Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc Natl Acad Sci USA* 101:9891–9896
- 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
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Fritsch R, Bohle B, Vollmann U, Wiedermann U, Jahn-Schmid B, Krebitz M, Breiteneder H, Kraft D, Ebner C (1998) Bet v 1, the major birch pollen allergen, and Mal d 1, the major apple allergen, cross-react at the level of allergen-specific T helper cells. *J Allergy Clin Immunol* 102:679–686
- Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, Joost van Neerven RJ, Schou C, Løwenstein H, Spangfort MD (1996) X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat Struct Biol* 3:1040–1045
- Galli P, Broggin GAL, Kellerhals M, Gessler C, Patocchi A (2010) High-resolution genetic map of the Rvi15 (Vr2) apple scab resistance locus. *Mol Breed* 26:561–572
- Gao ZS, Van de Weg WEW, 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* 111:171–183
- Gao ZS, 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
- Gasic K, Gonzalez DO, Thimmapuram J, Liu L, Malnoy M, Gong G, Han Y, Vodkin LO, Aldwinckle HS, Carroll NJ, Orvis KS, Goldsbrough P, Clifton S, Pape D, Fulton L,

- Martin J, Theising B, Wisniewski ME, Fazio G, Feltus FA, Korban SS (2009) Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. *Plant Genome* 2:23–38
- Ghosh D, 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* 45:456–462
- 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
- Hancock JM (2005) Gene factories, microfunctionalization and the evolution of gene families. *Trends Genet* 21:591–595
- Helsper JPF, 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
- Höfer M, Meister A (2010) Genome size variation in *Malus* species. *J Bot doi:10.1155/2010/480873*
- 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, Venäläinen TA, Häyrinen J, Latikainen R, Peräkylä M, Tervahauta AI, Kärenlampi SO (2005) Birch PR-10c interacts with several biologically important ligands. *Phytochemistry* 66:2524–2533
- Kootstra HS, Vlieg-Boerstra BJ, Dubois AEJ (2007) Assessment of the reduced properties of the Santana apple. *Ann Allergy Asthma Immunol* 99:522–525
- Lebel S, Schellenbaum P, Walter B, Maillot P (2010) Characterisation of the *Vitis vinifera* PR10 multigene family. *BMC Plant Biol* 10:184
- 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 × domestica* Borkh.). *Mol Breed* 10:217–241
- Liu JJ, Ekramoddoullah AK (2006) The family 10 of plant pathogenesis-related proteins: their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol Mol Plant Pathol* 68:3–13
- Lucht JM, Mauch-Mani B, Steiner HY, Metraux JP, Ryals J, Hohn B (2002) Pathogen stress increases somatic recombination frequency in Arabidopsis. *Nat Genet* 30:311–314
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155
- 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
- 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* 41:221–230
- 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
- Markovic-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* 325:123–133
- Mogensen JE, Ferreras M, Wimmer R, Petersen SV, Enghild JJ, Otzen DE (2007) The major allergen from birch tree pollen, Bet v 1, binds and permeabilizes membranes. *Biochemistry* 46:3356–3365
- Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* 8:122–128
- Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di Gaspero G, Morgante M, Adam-Blondon AF (2008) A physical map of the heterozygous grapevine ‘Cabernet Sauvignon’ allows mapping candidate genes for disease resistance. *BMC Plant Biol* 8:66
- Muñoz C, Hoffmann T, Escobar NM, Ludemann F, Botella MA, Valpuesta V, Schwab W (2010) The strawberry fruit Fra a allergen functions in flavonoid biosynthesis. *Mol Plant* 3:113–124
- Pagliarani G, Paris R, Tartarini S, Sansavini S (2009) Cloning and expression of the major allergen genes in apple fruits. *J Horticult Sci Biotechnol ISAFRUIT special issue: 176–181*
- Paterson AH, Freeling M, Tang H, Wang X (2010) Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol* 61:349–372
- Patocchi A, Gianfranceschi L, Gessler C (1999) Towards the map-based cloning of Vf: fine and physical mapping of the Vf region. *Theor Appl Genet* 99:1012–1017
- 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. *BMC Evol Biol* 8:286
- Ricci G, Dondi A, Belotti T, Baldi E, Tartarini S, Paris R, Pagliarani G, Serafini-Fracassini D, Casadio R, Giannetti A, Masi M (2010) Allergenicity of different apple cultivars assessed by means of skin prick test and sensitisation to recombinant allergens Mal d 1 and Mal d 3 in a group of Italian apple-allergic patients. *Int J Food Sci Technol* 45:1517–1523
- Rodríguez Milla MA, Townsend J, Chang IF, Cushman JC (2006) The Arabidopsis AtDi19 gene family encodes a novel type of Cys2/His2 zinc-finger protein implicated in ABA-independent dehydration, high-salinity stress and light signaling pathways. *Plant Mol Biol* 61:13–30
- Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, Benfey PN (2005) COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* 17:1749–1763
- Sargent DJ, Marchese A, Simpson DW, Howad W, Fernández-Fernández F, Monfort A, Arús P, Evans KM, Tobutt KR (2009) Development of “universal” gene-specific markers from *Malus* spp. cDNA sequences, their mapping and use

- in synteny studies within Rosaceae. *Tree Genet Genomes* 5:133–145
- 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* 147:985–1003
- Spangfort MD, Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, van Neerven RJ, Schou C, Løwenstein H (1997) Three-dimensional structure and epitopes of Bet v 1. *Int Arch Allergy Immunol* 113:243–245
- Stacey MG, Hicks SN, von Arnim AG (1999) Discrete domains mediate the light-responsive nuclear and cytoplasmic localization of Arabidopsis COP1. *Plant Cell* 11:349–364
- Untergasser A (2006) Maxiprep - Alkaline Lysis Untergasser's Lab. Summer 2006. http://www.untergasser.de/lab/protocols/maxiprep_alkaline_lysis_v1_0.htm
- Van der Maas MP, Schenk MF (2009) Development of a protocol that allows safe consumption of the hypoallergenic apple cultivar Santana. *Acta Hort* 841:240–252
- Van Loon LC, Rep M, Pieterse CM (2006) Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* 44:135–162
- Van Ooijen J, Voorrips R (2001) Joinmap[®] Version 3.0: Powerful genetic linkage mapping made easy. Plant Research International, Wageningen, The Netherlands
- 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* 214:538–551
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troglio M et al (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 42:833–839
- Vinatzer BA, Zhang HB, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. *Theor Appl Genet* 97:1183–1190
- Vinatzer BA, Patocchi A, Gianfranceschi L, Tartarini S, Zhang HB, Gessler C, Sansavini S (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
- Vlieg-Boerstra BJ, van de Weg WE, van der Heide S, Kerkhof M, Arens P, Heijerman Peppelman G, Debois AEJ (2010) Identification of low allergenic apple cultivars using skin prick tests and oral food challenge tests. *Allergy*. doi:10.1111/j.1398-9995.2010.02499.x
- Voorrips R (2001) MapChart version 2.0: Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen
- 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* 121:246–252
- Yamada S, Shiro Y (2008) Structural basis of the signal transduction in the two-component system. *Adv Exp Med Biol* 631:22–39
- Yang Z, Ma Z, Chen L, Xie R, Zhang X, Zhang B, Lu M, Wu S, Gilissen LJWJ, van Ree R, Gao Z (2011) Differential transcript abundance and genotypic variation of four putative allergen-encoding gene families in melting peach. *Tree Genet Genomes* doi:10.1007/s11295-011-0383-5
- Zaborsky N, Brunner M, Wallner M, Himly M, Karl T, Schwarzenbacher R, Ferreira F, Achatz G (2010) Antigen aggregation decides the fate of the allergic immune response. *J Immunol* 184:725–735
- Zhou S, Sauv e R, Thannhauser TW (2009) Proteome changes induced by aluminium stress in tomato roots. *J Exp Bot* 60:1849–1857
- 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* 150:1235–1247