Theor Appl Genet (2005) 111: 1087–1097
DOI 10.1007/s00122-005-0034-z

ORIGINAL PAPER

Z. S. Gao · W. E. van de Weg · J. G. Schaart
G. van Arkel · H. Breiteneder
K. Hoffmann-Sommergruber · L. J. W. J. Gilissen

Genomic characterization and linkage mapping of the apple allergen genes *Mal d 2* (thaumatin-like protein) and *Mal d 4* (profilin)

Received: 25 April 2005 / Accepted: 28 June 2005 / Published online: 6 September 2005

Abstract Four classes of apple allergens (Mal d 1, −2, −3 and −4) have been reported. By using PCR cloning and sequencing approaches, we obtained genomic sequences of *Mal d 2* (thaumatin-like protein) and *Mal d 4* (profilin) from the cvs Prima and Fiesta, the two parents of a European reference mapping population. Two copies of the *Mal d 2* gene (*Mal d 2.01A* and *Mal d 2.01B*) were identified, which primarily differed in the length of a single intron (378 or 380 nt) and in one amino acid in the signal peptide. Both *Mal d 2.01A* and *Mal d 2.01B* were mapped at identical position on linkage group 9. Genomic characterization of four *Mal d 4* genes (*Mal d 4.01A* and *B, Mal d 4.02A* and *Mal d 4.03A*) revealed their complete gDNA sequences which varied among genes in length from 862 to 2017 nt. They all contained three exons of conserved length: 123, 138, and 135 nt. *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.

Communicated by H. Nybom

Z. S. Gao · W. E. van de Weg (✉) · L. J. W. J. Gilissen
Allergy Consortium Wageningen, Wageningen University and Research Centre, P.O. Box 16, 6700 Wageningen, The Netherlands
E-mail: eric.vandeweg@wur.nl
Tel.: +31-317-477281
Fax: +31-317-418094

Z. S. Gao
Experimental Nursery of Shanxi Forestry Department, Beiying Nan Lu 52, Taiyuan, Shanxi, 030031 People's Republic of China

J. G. Schaart · G. van Arkel · Z. S. Gao · W. E. van de Weg
L. J. W. J. Gilissen
Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700 Wageningen, The Netherlands

H. Breiteneder · K. Hoffmann-Sommergruber
Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, 1090 Vienna, Austria

Introduction

In apple, four classes of allergens (Mal d 1, −2, −3 and −4) have been reported with different clinical relevance. Mal d 1-based allergy is predominant in Central and Northern Europe. In contrast, allergies caused by Mal d 3 and Mal d 4 are more relevant in the Mediterranean area of Southern Europe (Van Ree et al. 1995; Hoffmann-Sommergruber and Radauer 2004). For Mal d 2, the geographical relevance is still unclear. Of the four classes of apple allergen genes, *Mal d 1* and *Mal d 3* have been characterized for their genomic DNA sequences and positions on apple molecular marker linkage maps (Gao et al. 2005a, b). This paper focuses on the remaining two allergen genes *Mal d 2* and *Mal d 4* encoding thaumatin-like protein (TLP) and profilin, respectively.

Thaumatin-like protein is one of the major protein constituents of mature apple fruit (Oh et al. 2000). It is homologous to an intensely sweet tasting protein, isolated from the fruit of *Thaumatococcus daniellii*. TLPs can be categorized into three subclasses based on their isoelectric point (pI): acidic, neutral and basic (Koiwa et al. 1994). Most TLPs have 16 conserved cysteines that form eight disulfide bonds contributing to the protein's resistance to proteolysis and heat (Breiteneder 2004). TLPs belong to the PR-5 family of pathogenesis-related proteins (Van Loon and Van Strien 1999). Several researchers provided evidence that TLPs play a role in plant defence against pathogens (Ibeas et al. 2000; Venisse et al. 2002; Velazhahan and Muthurishnan 2003; Han et al. 2004). However, TLPs are now recognized as a new class of panallergens in food as well as in pollen (Breiteneder 2004). Hsieh et al. (1995) were the first to identify a 31-kDa apple TLP as a major allergen. Its cDNA (AF090143) has been isolated and characterized from apple fruit (Oh et al. 2000). More recently, a 23.2-kDa TLP, deduced from a full-length cDNA clone (AJ243427) isolated from ripe apple fruits, was classified as Mal d 2 allergen and characterized as an antifungal

protein (Krebitz et al. 2003). The purified recombinant Mal d 2 appeared to have a double band at 31–32 kDa on SDS-PAGE, while MALDI-TOF mass spectrometry revealed a dominant peak close to the calculated 23 kDa (Krebitz et al. 2003).

Profilins are small (12–15 kDa) cytosolic proteins that are found in all eukaryotic cells. In pollen, their abundance increases 10-fold to 100-fold during maturation (Radauer and Hoffmann-Sommergruber 2004). In 1991, birch pollen profilin was identified as a relevant allergen (Valenta et al. 1991). In the meantime, 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 AA), domains and structure (Radauer and Hoffmann-Sommergruber 2004). Large profilin multigene families can be grouped into two classes: those constitutively expressed in vegetative tissues and those expressed mainly in reproductive tissue. Multiple profilin isoforms can be simultaneously expressed in individual tissues and cells (Huang et al. 1996). 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). Apart from pollen, many fruits contain profilins and their allergenic potency has been frequently reported (Scheurer et al. 2000, 2001; Wensing et al. 2002; Asero et al. 2003; Rodriguez-Perez et al. 2003; Westphal et al. 2004). When we started this research, three distinct mRNA sequences (AF129426, AF129427 and AF129428) encoding three isoallergens of apple profilin were available in the GenBank/EMBL nucleotide database. They all have an open reading frame (ORF) of 396 nt.

By applying similar PCR cloning and sequencing approaches as used for mapping of *Mal d 1* and *Mal d 3* genes (Gao et al. 2005a, b), we obtained the genomic sequences of *Mal d 2* and *Mal d 4* from cvs Prima and Fiesta, the two parents of a European reference mapping population (Maliepaard et al. 1998). Sequence-specific markers were created to map these genes on the apple linkage map. This resulted in the identification and mapping of two *Mal d 2* and four *Mal d 4* loci.

Materials and methods

Gene annotation

Mal d 2 was used as the basic symbol for apple TLP genes and *Mal d 4* for apple profilins. For further dividing each class into distinct genes and alleles, we adopted the same nomenclature as described previously for *Mal d 1* and *Mal d 3* (Gao et al. 2005a, b) while taking into account of published sequences of similar genes in the DNA databases.

PCR genomic cloning and sequencing

Prima and Fiesta, two diploid parental apple cultivars of a reference mapping population, were chosen for genomic PCR cloning and sequencing. They are the same cultivars that had been chosen for mapping *Mal d 1* (Gao et al. 2005a) and *Mal d 3* (Gao et al. 2005b). Two primer pairs located in the untranslated region (UTR) were used to amplify the targeted *Mal d 2* genomic sequences (Table 1). We designed one common forward primer and three different reverse primers to clone the three profilin genes, because these reference sequences (AF129426-AF129428) were conserved at their 5′ ends but divergent at the 3′ end. The PCR cloning and sequencing procedures were the same as described previously (Gao et al. 2005b).

Sequence analysis

DNA sequences and single nucleotide polymorphisms (SNPs) were analyzed using the Seqman program (DNAstar, Madison, WI, USA). Intron sizes were deduced by comparing the genome sequences with known cDNA sequences. Sequence identity percentages were calculated using the Clustal W of the Megalign program (DNAstar). Multiple DNA and amino acid sequence alignments were performed with the GeneDoc program (http://www.psc.edu/biomed/genedoc).

Designing and testing of sequence-specific markers

According to the DNA polymorphisms among the genomic sequences, three types of molecular markers were designed to distinguish a specific sequence or allele. These are (1) single nucleotide amplification polymorphism (SNAP) (Drenkard et al. 2000; Gao et al. 2005b), (2) cleaved amplification polymorphisms (CAPs) and (3) simple sequence repeat (SSR) markers. The methods for designing and testing of SNAP and SSR markers were according to Gao et al. (2005b).

Mapping genes on molecular linkage groups

Three molecular marker linkage maps derived from Prima × Fiesta (PM × FS, $n = 141$), Fiesta × Discovery (FS × DS, $n = 70$) and Jonathan × Prima (JO × PM, $n = 175$) were used to map *Mal d 2* and *Mal d 4* genes. Grouping and mapping of sequence-specific markers were performed with JoinMap 3.0 (Van Ooijen and Voorrips 2001), using the Kosambi mapping function. The LOD and recombination threshold was 4 and 0.45, respectively. Final drawings of the marker maps were generated with MapChart (Voorrips 2001).

Table 1 Primer pairs used for the genomic cloning of *Mal d 2* and *Mal d 4* genes

Allergen gene	Reference sequence	Primer sequence (5′–3′)	Position[a]	PCR Tm/cycles		Product (nt)	GenBank accession numbers	
				Pfu	*Taq*		Prima	Fiesta
Mal d 2	AJ243427	Mal d2-For1: GCAGCAAACAGGCAATTAAG	22–41	60/30	64/2	1,296–1,298	AY792598	AY792599
		Mald2-Rev1: TGTTCCTTTATTTGATCGTGT	919–939				AY792600	AY792601
		Mald2-For2: TCAATGTCGATGATGAAGAGCCA	48–70	58/25	60/2	1,159–1,161	AY792602	AY792603
		Mald2-Rev2: CATCATCTGCATATATAATCCCAT	805–827				AY792604	
Mal d 4.01	AF129426	Mald4-For: TGTCGTGGCAGGCGTACGTC	2–21	52/30	54/2	2,016	AY792606	AY792605
		Mald4.01-Rev: TTATAGGCCTTGATCAATCAGGTAGTCT	369–396				AY792608	AY792607
								AY792609
Mal d 4.02	AF129427	Mald4-For: see above	2–21	60/30	64/2	861, 862	AY792610	AY792611
		Mald4.02-Rev: TTAGAGACCCTGCTCAATGAGATAATCC	369–396				AY792612	AY792613
Mal d 4.03	AF129428	Mald4-For: see above	2–21	60/30	62/2	1,029–1,032	AY792614	AY792616
		Mald4.03-Rev: TTAGAGACCCTGCTCGATAAG	376–396				AY792615	AY792617

[a]Refer to reference sequence

Results

Genomic cloning and mapping of *Mal d 2* genes

Two primer pairs at different positions in the reference sequence (AJ243427) were used to clone *Mal d 2* genes from gDNA of Prima (PM) and Fiesta (FS) (Table 1). The sequencing of 29 true clones resulted in four different sequences from PM (GenBank accession nos: AY792598, AY792600, AY792602, AY792604) and three from FS (AY792599, AY792601, AY792603). These seven sequences had a nucleotide identity of 98.5–100%; two sequences from PM were identical to two sequences from FS (but were given different accession numbers). When compared with the original reference cDNA sequence, our *Mal d 2* genomic sequences apparently had one intron of either 378 or 380 nt (Table 2). The presence of four sequences from PM requires the presence of two *Mal d 2* loci, which we named *Mal d 2.01A* and *Mal d 2.01B*. The four sequences could be divided into two sets according to their intron length: the first set of two sequences (AY792598 and AY792600) with an intron of 378 nt, and the second set (AY792602 and AY792604) with an intron of 380 nt. We assumed that the intron size is locus specific, and assigned the first set of sequences to *Mal d 2.01A* and the second to *Mal d 2.01B*.

For mapping, three sequence-specific SNAP markers were created for *Mal d 2.01A* and two for *Mal d 2.01B* and tested on the PM × FS population (Table 3). The three markers for *Mal d 2.01A* showed a segregation pattern typical to a so-called *ef* × eg cross (Van Ooijen and Voorrips 2001), where *e* represents the common allele of two parents (Mal d 2.01A01.01), and *f* (Mal d 2.01A02.01) and *g* (Mal d 2.01A02.02) represent a unique allele of each parent. Mal d 2.01B01.02, one of the two markers for *Mal d 2.01B*, was present only in PM and had exactly the same segregation pattern as the marker Mal d 2.01A02.02. Both markers were always simultaneously present or absent, indicating that they are in coupling phase and located at identical map position. The other marker, Mal d 2.01B01.01, was amplified in both parental cultivars and did not segregate in their progeny. The absence of segregation indicates that this marker is present in a homozygous condition in (at least) one of the parents. We assume that FS is homozygous for this marker and PM heterozygous. In this way, markers Mal d 2.01B01.01 and -.02 are allelic for locus *Mal d 2.01B*.

During the cloning process, we also observed that a few clones did not have correct sequences due to PCR errors (reading mistakes, strand switching). One artifact sequence recombining *Mal d 2.01B* and *Mal d 2.01A01* sequences was obtained from FS. Sequence-specific primers for this artifact failed to amplify the expected band in FS, but surprisingly amplified a band of the expected size from gDNA of cv Discovery (DS). Moreover, this marker (Mal d 2.01B-DS) was mapped

Table 2 Exon and intron sizes (nt) of two *Mal d 2* and four *Mal d 4* genes

Gene	Total length	Coding	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3
Mal d 2.01A	1,119	741	61	378	680		
Mal d 2.01B	1,121	741	61	380	680		
Mal d 4.01A	2,017	396	123	343	138	1,278	135
Mal d 4.01B	2,017	396	123	343	138	1,278	135
Mal d 4.02A	862/863	396	123	248	138	218/219	135
Mal d 4.03A	1,030/1,032/1,033	396	123	386/388/389	138	248	135

on LG 9-DS at a similar position as *Mal d 2.01A* and *Mal d 2.01B* on PM and FS (Fig. 1). Sequencing of this fragment may reveal whether it is allelic to the two just-identified genes, or a new paralogue.

Genomic cloning and mapping of *Mal d 4* genes

At the onset of our research, the DNA database contained three different profilin sequences (AF129426, AF129427 and AF129428) with 75–80% identity, which have been named as Mal d 4.01, Mal d 4.02 and Mal d 4.03 isoallergens (http://www.allergen.org), respectively. On the basis of these three sequences, three different primer pairs (Table 1) were designed for PCR cloning in cvs PM and FS. From each primer pair, two genomic sequences from each cultivar were obtained, except for three sequences of *Mal d 4.01* from FS. These thirteen new sequences were deposited in GenBank with accession numbers AY792605 to AY792617 (Table 1). They all had the same deduced ORF length of 396 nt and two introns of very different length (Table 2). The deduced coding sequences were over 98.7% identical to the corresponding reference sequences.

The first primer pair (Mal d 4.01) amplified two sequences from PM and three from FS (AY792605–AY792609, Table 1). Comparison of these five sequences resulted in three different sequences with 99.6–99.8% identity. PM and FS shared two common sequences, while FS had a unique one (AY792605). Three SNAP markers were generated and tested (Table 3). Mal d 4.01A01 segregated in PM × FS and could be mapped on LG 9 of FS (Fig. 1), 15.5 cM from *Mal d 2* and 15.9 cM below the SSR marker CH05c07. We denoted the corresponding gene as *Mal d 4.01A*. The second marker (Mal d 4.01B) was present in all four parents, segregated only in JO × PM at a ratio of 3:1 (126:39), and mapped on LG 9-PM at similar position as *Mal d 4.01A* (15.4 cM below CH05c07). Because the marker Mal d 4.01B was heterozygous in PM and did not segregate in PM×FS, it must be homozygous in FS. Therefore *Mal d 4.01B* represents another locus. Considering the high sequence identity to *Mal d 4.01A*, we denoted this locus as *Mal d 4.01B*. These two loci are located near to each other, considering their distance to a linked reference marker (CH05c07, Fig. 1). The third marker (Mal d 4.01A02) was present in all parental cultivars and did not segregate in any population. It is most likely a *Mal d 4*.01 *A* allele, which is homozygous in PM and heterozygous in FS.

For *Mal d 4.02*, two different sequences were obtained from each of PM and FS (AY792610-AY792613, Table 1). AY792612 from PM and AY792613 from FS were identical. Markers for each of the three unique sequences were developed and tested on PM × FS (Table 3). These markers showed to be allelic, matching perfectly the segregation pattern of the model cross *ef* × eg (Van Ooijen and Voorrips 2001). These markers were mapped on LG 2 of PM and FS at similar distance from the common RFLP marker MC003 (Fig. 1).

With regard to *Mal d 4.03*, PM and FS each had two different sequences of variable lengths (1,030–1,033 nt) due to a CT-dinucleotide that repeated seven–nine times, and a deletion of a single nucleotide (AY792614–AY792617). Three kinds of markers were created to identify three of the four sequences (Table 3). A cleaved amplified polymorphism (CAP) marker was developed for sequence AY792614 (allele Mal d 4.03A01.01): the amplified product of 1,029 nt could be cleaved with *Eco* RI into two fragments of 713 and 316 nt. A SNAP marker was created for the sequence (AY792616) from FS. In addition, an SSR marker was developed (Mal d 4.03A-SSR), which revealed two bands in PM (160, 162 nt) and FS (162,164 nt) (Table 3). All these markers confirmed one another in their segregation. The mapping process resulted in locating the *Mal d 4.03A* locus on LG 8 (Fig. 1).

Genomic structure of *Mal d 2* and *Mal d 4*

By comparing the genomic sequences with the published reference mRNA sequences, the exons and introns could be identified (Table 2). The *Mal d 2* genes have two exons, one consisting of 61 nt and the other of 680 nt. These two exons are separated by an intron of 378 or 380 nt depending on the number of AT repeats within a mini SSR region. The four nucleotides at the 3′ and 5′ splicing sites of each exon and intron junction were AG:GT, the most common pattern in plants (Fig. 2).

All four profilin genes (*Mal d 4.01A/B, Mal d 4.02A, Mal d 4.03A*) have an ORF of identical length (396 nt) which is comprised of three exons of 123, 138 and 135 nt. The four exon–intron junction patterns were also identical for all genes except for the first site of *Mal d 4.03A* (Fig. 2). The two introns varied in size per gene (Table 2): 343 and 1,278 nt respectively for *Mal d 4.01*; 248 and 218/219 nt for *Mal d 4.02A*; and a variable intron of 386/388/389 nt and 248 nt for *Mal d 4.03A*. The four nucleotides at the 3′ and 5′ splicing sites of exon–intron junction were A/C,G:G/T,T/G (Fig. 2). The reported efficient branch-

Table 3 Sequence-specific molecular markers for mapping of *Mal d 2* and *Mal d 4*

Gene	Marker name[a] (detected sequences)	Primer sequence (5′–3′)[b]	Position[c]	Tm[d]	Cycles[e]	Product (nt)	Presence and segregation[f]	LG[g]
Mal d 2.01	Mal d 2.01A01.01	Fw-TCGATGATGAAGAGCCATGT	33–52	60	35	885	PM/FS	9
	AY792598	Rv-AAGTGGAGCCGGGCACA**T**CT	898–917				PMxFS (101:39)	
	AY792599							
	Mal d 2.01A02.01	Fw-TCAACTTGCCTATGTCTCTG	819–839	56	35	199	PM,	9
	AY792600	Rv-CCGGCGTATCATTCAGC**G**AG	988–1,017				PMxFS (72:68)	
	Mal d 2.01A02.02	Fw- TAAAAGTTCTCAATCCTG**C**C	338–357	57	35	844	FS,	9
	AY792601	Rv-ATCTGCATATATAATCCCATTTG**T**	1,158–1,181				PMxFS (72:68)	
	Mal d 2.01B01.01	Fw-CCTCACCTTGGCCATCCTCAT**T**	68–89	60	35	952	PM, **FS, JO**, DS[h]	
	AY792602	Rv-CCGGCGTATCATTCAGCG**A**G	1,000–1,019					
	AY792603							
	Mal d 2.01B01.02	Fw-CCTCACCTTGGCCATCCTCATT	68–89	51	35	301	PM	9
	AY792604	Rv-CACTTATATGTATCCCGATT**G**T	347–368				PMxFS (72:68)	
	Mal d 2.01B-DS	Fw-CCTCACCTTGGCCAT**C**CTCATT	68–89	62	35	852	DS	9
		Rv-AAGTGGAGCCGGGCACAC**C**T	900–919				FS×DS (33:37)	
Mal d 4.01 A/B	Mal d 4.01A01	Fw-GCAAAGTTTTTCTGGTCT**C**TA	1,828–1,848	59	45	113	FS	9
	AY792605	Rv-CATATATACCGAAAACCA**G**AG	1,920–1,940				PM × FS (64:70)	
	Mal d 4.01A02	Fw-GACGGCCACCATCTTACA	46–63	58	35	1,896	PM,FS,JO,DS	
	AY792606, AY792607	Rv-TCATATATACCGAA**A**ACCAAAG	1,920–1,941					
	Mal d 4.01B01	Fw-GGCAGGCGTACGTCGTT	7–24	56	35	597	PM,FS, JO,DS	9
	AY792608, AY792609	Rv-CTTCTTGCCACGAATC**A**CA	586–604				JO × PM (126:39)	
Mal d 4.02	Mal d 4.02A01.01	Fw-TCGTTTTTTTCACTTTCTTG**T**CT	131–153	56	35	674	PM,	2
	AY792610	Rv-CCAGGAGTCATCGGCACA**T**	788–804				PM × FS (70:71)	
	Mal d 4.02A01.02	Fw-CTGAGGAGGTGACTG**C**CA	379–396	58/56	10/30	428	FS	2
	AY792611	Rv-GTCAGGAGTCATCGG**C**ACG	788–806				PM × FS (60:81)	
	Mal d 4.02A02	Fw-C**T**GAGGAGGTGACTGT**C**G	379–396	60	35	428	PM, FS;	2
	AY792612, AY792613	Rv-GTCAGGAGTCATCGG**C**ACG	788–806				PM × FS (105:36)	
Mal d 4.03	Mal d 4.03A01.01	Fw-T**G**TCGTGGCAGGCGTA**C**GTC	CAPs- *Eco* RI	60	35	1,029	PM,	8
	AY792614	Rv-TTAGAGACCCTGCTCGATAAG	713/316				PM × FS (69:72)	
	Mal d 4.03A01.02b	Fw-GATCAACCCGGAACTCTT	551–568	60	35	244	FS,	8
	AY792616	Rv-TCTTCGACGCAAAC**T**CTCAA	775–794				PM × FS (79:62)	
	Mal d 4.03A-SSR-160	SSR-Fw-TTGGTCTCAGAGCG**C**CTCTT	96–115	62	35	160	PM,	
	AY792614	SSR-Rv-***GTTT***AACCGATCGACCGAACCAGT	228–247(+4)				PM × FS (60:66)	
	Mal d 4.03A-SSR-162	SSR-Fw-TTGGTCTCAGAGCGCCTCTT	96–115	62	35	162	PM, FS	
	AY792615, AY792616	SSR-Rv-***GTTT***AACCGATCGACCGAACCAGT	230–251(+4)				PM × FS (104:22)	
	Mal d 4.03A-SSR-164	SSR-Fw-TTGGTCTCAGAGCGCCTCTT	96–115	62	35	164	FS,	8
	AY792617	SSR-Rv-***GTTT***AACCGATCGACCGAACCAGT	232–253(+4)				PM × FS (54:72)	

[a]Marker name and its detected sequence or sequences. Marker *Mal d 2.01B-DS* was based on an artifact sequence from Fiesta but showed to be valid in Discovery (DS)
[b]Designed mismatching nucleotides are in bold and underlined
[c]Positions are counted from the start ATG codon in genomic sequence
[d, e]Annealing temperature (Tm) and number of PCR cycles. In the case of two values, a touch-down PCR was performed in two steps: the first number then refers to the Tm and number of cycles, respectively, of the first step
[f]Marker presence in parents, PM, Prima; FS, Fiesta; JO, Jonathan; DS, Discovery. Present:absent values are shown in brackets
[g]Linkage group (LG)
[h]**FS and JO** are homozygous for Mal d 2.01B01.01

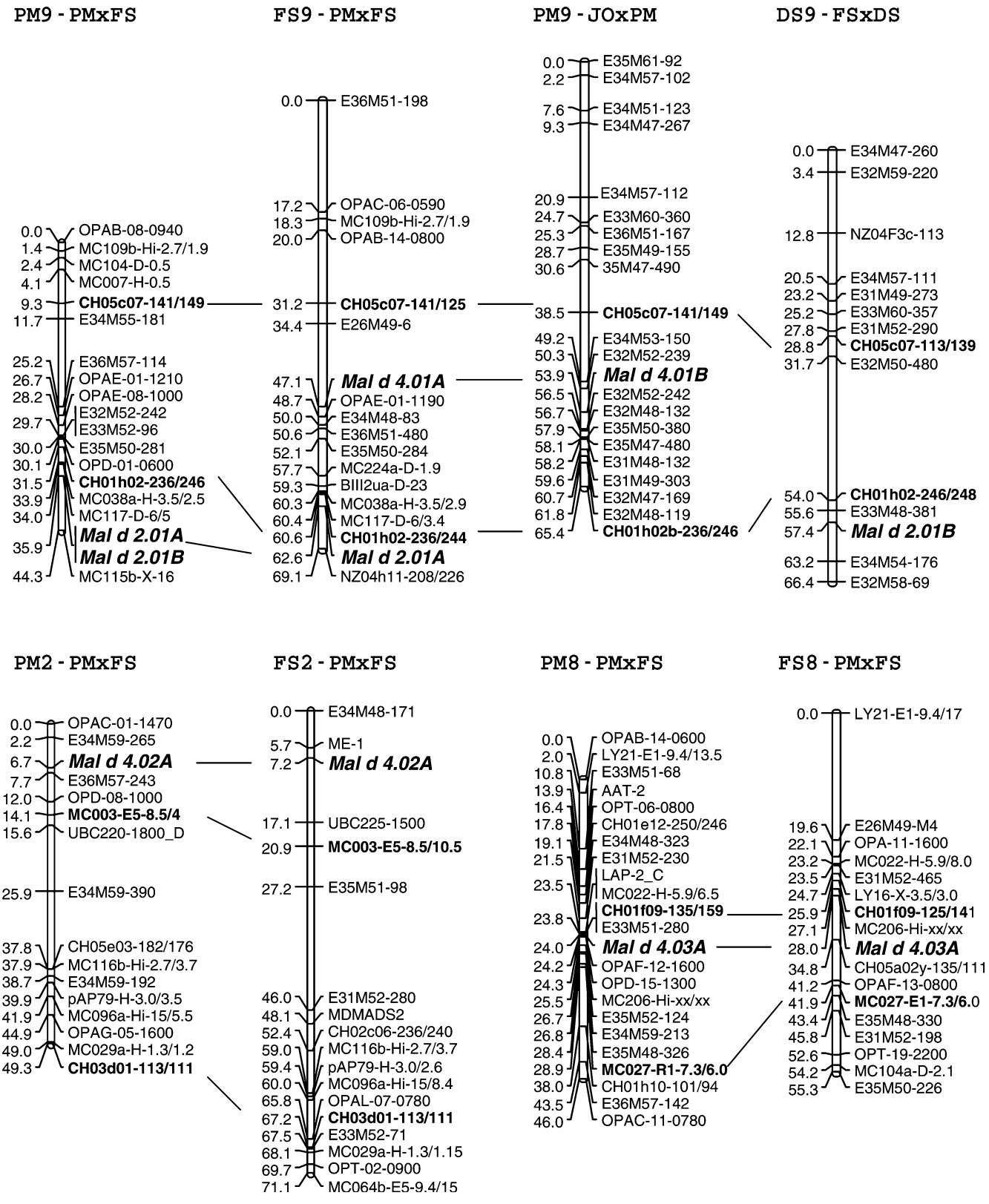

Fig. 1 Positions of two *Mal d 2* and four *Mal d 4* genes on molecular marker linkage maps from three mapping populations. Parental cultivars: *PM* Prima, *FS* Fiesta, *JO* Jonathan, *DS* Discovery

point consensus sequence 'CTRAY' (Simpson et al. 2002) was found in most introns of *Mal d 2* and *Mal d 4* genes near to the 5′ splicing sites (Fig. 2).

Sequence comparison of DNA and deduced amino acids of *Mal d 2* and *Mal d 4*

Twenty-three SNPs were present among five different sequences of two *Mal d 2* genes, of which 13 occurred in the exons and 10 in the introns. They resulted in three different amino acid sequences (Fig. 3). *Mal d 2.01A* differed consistently from *Mal d 2.01B* sequences in two and six SNPs in the first exon and intron, respectively, only resulting in one amino acid difference (6A or 6V) in the signal peptide (Fig. 3). The second exon showed 11 SNPs of which 10 consistently differentiated *Mal d 2.01A01* from the other four sequences, resulting in five amino acid changes (Fig. 3). The 11th SNP had a unique, silent polymorphism for *Mal d 2.01B01.02* and which was the only polymorphism with *Mal d 2.01B01.01*. Mature Mal d 2.01A01 had a calculated molecular weight of 23,210 Da and pI 4.57, and Mal d 2.01A02 of 23,190 Da and pI 4.26. The mature protein of Mal d 2.01B01 is the same as that of Mal d 2.01A02 (Fig. 3).

The three types of *Mal d 4* genes had 78% to 80% identity for both their coding DNA and amino acid sequences. These genes differed also in the regions where they showed allelic polymorphisms. For *Mal d 4.01*, nine out of ten SNPs were located in the first and the third exon, while the middle part was highly conserved despite its relatively long length (1,759 nt). The variation in the first exon gave rise to three variants for Mal d 4.01 (Fig. 4). *Mal d 4.02A* had only one SNP in the second and the third exon, which resulted in two variants, whereas the two introns had 17 SNPs. *Mal d 4.03A* was even more conserved in the coding region with only one silent nucleotide mutation. Its two introns had 13 SNPs and a SSR.

Discussion

In this study, we cloned and sequenced gDNA of two classes of apple allergen genes, TLP (Mal d 2) and profilin (Mal d 4). Subsequently, we identified and mapped two *Mal d 2.01* genes (A and B) and four *Mal d 4.01* genes by applying sequence-specific molecular markers on segregating populations. *Mal d 2.01A/B* and *Mal d 4.01A/B* mapped on linkage group 9; *Mal d 4.02A* on LG 2 and *Mal d 4.03A* on LG 8. With the new data presented here and with those of previous studies on *Mal d 1* and *Mal d 3* (Gao et al. 2005a, b), four classes of apple allergen genes have been mapped. The implications of these data for genetic analysis and breeding of low allergenic apple cultivars are discussed.

Mal d 2 and *Mal d 4* gene duplications within and across linkage groups of apple

Our data revealed two copies of *Mal d 2.01* which are slightly different in the signal peptide and intron size but not in the 5′ and 3′ UTR region and both copies were mapped at the same position on LG 9. Previous genomic southern blot analysis using two DNA probes containing an ORF or the 3′ UTR of a *Mal d 2* gene, which we now denoted as *Mal d 2.01A*, indicated the presence of a single copy of *Mal d 2.01* and also the presence of multiple similar TLP genes in the genome (Oh et al. 2000). Probably, these two *Mal d 2* genes are located so close to each other that no polymorphic fragment could be generated between these genes by the restriction enzymes used (Oh et al. 2000). Venisse et al. (2002) identified two different PR-5 EST sequences (AF494393 and

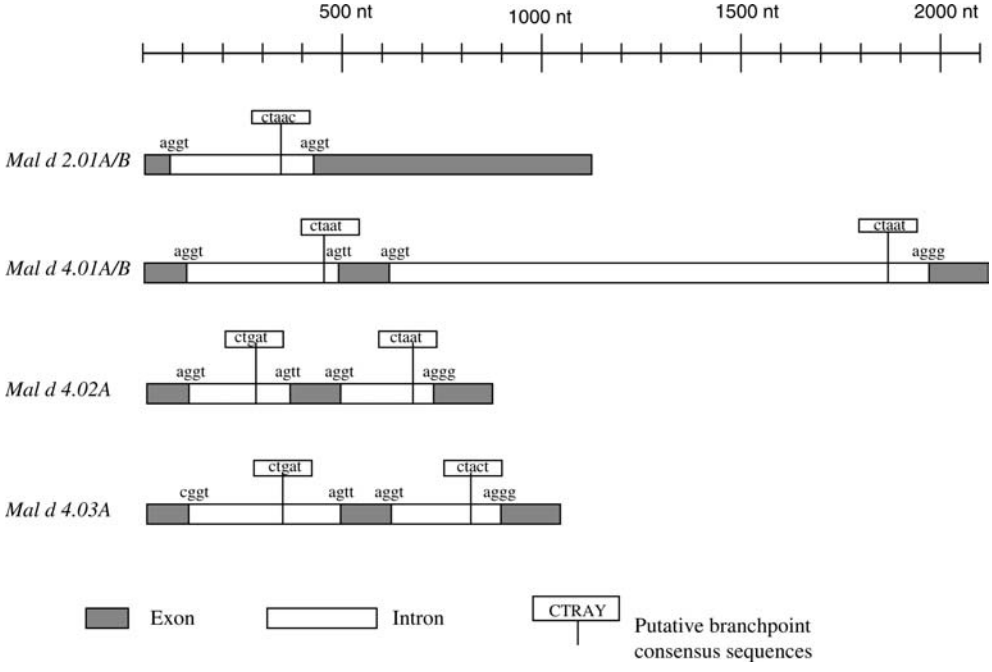

Fig. 2 Schematic illustration of the gene structure of *Mal d 2* and *Mal d* 4 genes

AF494394) in a study on defence responses of apple during compatible and incompatible interactions with *Erwinia amylovora* (the causal agent of fire blight). AF494393 is 99.2% identical to the allele *Mal d 2.01A01*, while AF494394 is much more similar to the second allele of *Mal d 2.01A02* and both alleles of *Mal d 2.01B*. Because these two sequences lack a part of the 5′ end of the coding sequences, we could not match them with certainty to *Mal d 2.01A* or *B* or both.

The profilin gene *Mal d 4.01* appeared to be duplicated and also located on LG 9. We did not find duplications for the other two profilin genes (*Mal d 4.02A* and *Mal d 4.03A*) which were mapped on LG 2 and LG 8, respectively.

The cultivated apple is an amphidiploid ($x = 17$) which evolved from a hybrid between two different ancestral species (Chevreau and Laurens 1987). The duplicated nature of the apple genome became apparent mainly through molecular markers on the reference map: PM × FS (Maliepaard et al. 1998) and its updated map (Van de Weg, unpublished). Information about the genomic sequences and mapping positions of *Mal d 1* and *Mal d 3* (Gao et al. 2005a, 2005b) supported the hybrid origin hypothesis. Most *Mal d 1* genes were mapped on the two homoeologous linkage groups (LG) 13 and 16, and two *Mal d 3* genes on the two homoeologous segments of LG 4 and 12. With regard to the three LG where *Mal d 2* and *Mal d 4* are located (Fig. 1), LG 9 is homoeologous to LG 17; LG 2 to LG 7; and LG 8 to a part of LG15 (Van de Weg, unpublished). We therefore expect that additional *Mal d 2* and *Mal d 4* loci are present on LG 17, LG 7, and LG 15 at positions corresponding to the current *Mal d 2* and *Mal d 4* locations. This expectation on the presence of additional genes was rewarded by the following database search. Sequence alignment of recently published TLP and profilin, which expressed sequence tags from apple in GenBank (some of which are listed in Table 4), definitely indicated the presence of two additional *Mal d 2* genes and two *Mal d 4* genes. One Mal d 2-like consensus encodes nearly the entire mature peptide and showed 12 amino acid differences (equivalent to 95% sequence identity) to the two genes mapped by us. Another sequence has a coding region of 738 nt and shows 67.5% amino acid identity to Mal d 2.01A. Similarly, at least two new apple profilin genes (*Mal d 4.02B* and *Mal d 4.03B*), which are present in the nucleotide databases, have the same length of 396 nt for the coding sequence (Table 4). *Mal d 4.02B* is closely related to *Mal d 4.02A* with 95.7% identity for the coding sequences, while the amino acid identity is 98.5% (only with a two amino acids difference). *Mal d 4.03B* has a coding sequence of 396 nt with 95% DNA sequence identity and 96.2% identity in amino acids compared to *Mal d 4.03A* (Fig. 4). Both the upstream and downstream of the coding sequences for these two new *Mal d 4* genes differ from their closely related genes. All these apple profilins have the same conserved G-actin-binding region (119–126 amino acids VERLGDYL) as pollen profilins from birch, celery, maize and rice (Ye et al. 2001; Radauer and Hoffmann-Sommergruber 2004). Still, the actual locations for these new *Mal d 2* and *Mal d 4* genes have to be determined.

Genomic structure of the TLP gene family in plants

In plants, TLPs are a big gene family with proteins of diverse pI values of either low or high molecular weight. Their exon number and length is also variable. While searching the *Arabidopsis* DNA database, 19 putative *TLP* genes have been identified, which are scattered on the chromosomes 1, 2, 4 and 5. Their ORF varied from 574 nt to 1998 nt, and mostly contained two or three

```
                          *        20         *        40         *        60
Mal d 2.01A01 : MMKSQVASLLGLTLAILFFSGAHAAKITFTNNCPNTVWPGTLTGDQKPQLSLTGFELASK :  60
Mal d 2.01A02 : ............................................................ :  60
Mal d 2.01B01 : .....A...................................................... :  60

                          *        80         *       100         *       120
Mal d 2.01A01 : ASRSVDAPSPWSGRFWGRTRCSTDAAGKFTCETADCGSGQVACNGAGAVPPATLVEITIA : 120
Mal d 2.01A02 : ..Q......................................................... : 120
Mal d 2.01B01 : ..Q......................................................... : 120

                          *       140         *       160         *       180
Mal d 2.01A01 : ANGGQDYYDVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNKVCPAPLQVKAADGSVISC : 180
Mal d 2.01A02 : ........................................MA...Q.............. : 180
Mal d 2.01B01 : ........................................MA...Q.............. : 180

                          *       200         *       220         *       240
Mal d 2.01A01 : KSACLAFGDSKYCCTPPNNTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDY : 240
Mal d 2.01A02 : ................D........................................... : 240
Mal d 2.01B01 : ................D........................................... : 240


Mal d 2.01A01 : VITFCP : 246
Mal d 2.01A02 : ...... : 246
Mal d 2.01B01 : ...... : 246
```

Fig. 3 Alignment of protein sequences of three variants of Mal d 2.01
Accession numbers for the deduced variants: Mal d 2.01A01: AY792598, AY792599 Mal d 2.01A02: AY792600, AY792601 Mal d 2.01B01: AY792602–AY792604

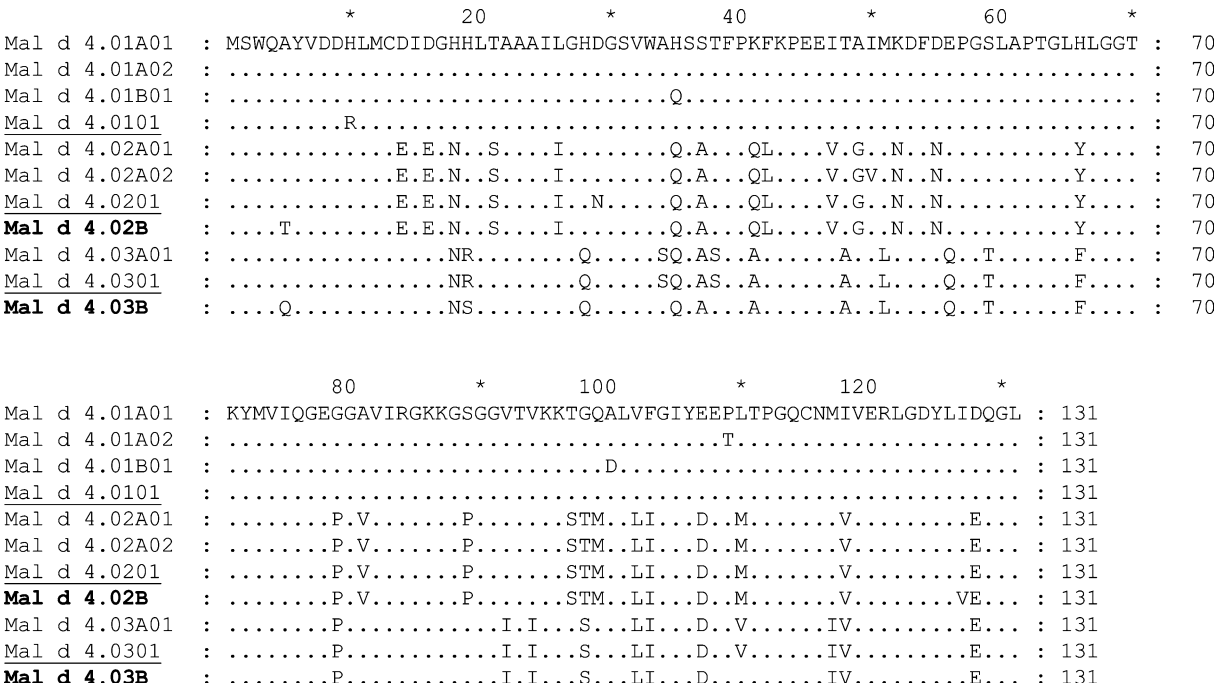

Fig. 4 Alignment of protein sequences of apple profilins

Mal d 4.01A01:AY792605 Mal d 4.01A02: AY792606, AY792607 Mal d 4.01B01: AY792608, AY792609 Mal d 4.0101: reference AF129426 Mal d 4.02A01: AY792610, AY792611 Mal d 4.02A02: AY792612, AY792613 Mal d 4.0201: reference AF129427 Mal d 4.03A01: AY792614-AY792617 Mal d 4.0301: reference AF129428 Mal d 4.02B, Mal d 4.03B in bold were deduced from EST sequences in the database listed in Table 4

exons, and one ORF had four exons. The length of these exons is also variable. TLPs in Rosaceous fruit crops showed a close relationship in amino acid identity and gene structure. In apple, two kinds of TLPs encoded by 741 nt or 738 nt were identified (for accession numbers, see Table 4). We have shown that the 741nt ORF sequence contains two exons of 61 and 680 nt (Table 2). The exon number and lengths of the 738 nt *TLP* are not yet known. A peach TLP (AF362988) has the same 741 nt coding sequence as *Mal d 2.01*, while the ORF of a TLP from sweet cherry (U32440) has 738 nt, like *Mal d 2*.03. TLP in Japanese pear (*Pyrus serotina*) has an ORF of 735 nt containing two exons (46 nt and 689 nt) separated by one intron of 351nt (accession AB036069). All these variations do not affect the presence of 16 cysteines that form eight disulfide bonds and thus maintain the high stability of these proteins.

Genomic structure of the profilin gene family in plants

All apple profilins have a coding sequence of 396 nt and two introns at conserved positions. The three exons have nearly equal sizes (123, 138, and 135 nt). These features appeared in most plant profilin genes, such as in maize, pear, cherry, peach and celery. In *Arabidopsis thaliana*, five profilin genes have been found, each of which has three exons. For two of these genes (profilin 1 and 2), exon sizes and splicing patterns are identical to those in apple, while for two other genes (profilin 3 and 4), a 9 nt longer first exon was found.

According to the conservation of splicing sites, we predict that the 402 nt birch pollen profilin encoding sequence (Bet v 2, M65179) consists of three exons of 129 nt, 138 nt and 135 nt, which are identical to those in apple except for a 6 nt longer first exon. From these conserved physical structures, it can be deduced that profilins have basic biological functions and may have been evolved from one ancestral gene similar to the Bet v 2 homologous genes. The similarity of plant, fungal, protist, insect and nematode profilins and their extreme divergence from the vertebrate profilins has striking implications for the evolution of fungal-spore- and plant-pollen-profilins as allergens (Huang et al. 1996).

Genetic analysis of apple allergenicity and breeding perspectives

Apple cultivars differ in the degree in which they elicit mild or severe allergic reactions in atopic patients (Vieths et al. 1994). To date, genes from all four classes of apple allergens (Mal d 1, Mal d 2, Mal d 3 and Mal d 4) have been mapped (Gao et al. 2005a, b and this paper). Elucidation of the genetic basis for the differences in allergenicity among cultivars will be complicated because of two aspects. Firstly, patients differ in their reactions to these four classes of allergens, and clinical assessment of allergenicity is not easy to perform on large numbers of patients. Secondly, the existence of gene families for at least two of these allergens (Mal d 1 and Mal d 4) may hamper clarification of the role of individual genes, especially when these genes are closely

Table 4 *Mal d 2* and *Mal d 4* gene sequences in GenBank/EMBL

Gene name proposed	Complete mRNA (Length of ORF[a])	Partial EST[b]	Tissue	Genomic sequence
Mal d 2.01A	AT000359,AF090143, AJ243427 (741 nt)	CO754260, CN996077, CN490732, CO052447, AF494393, AF494394, CV150401, AY548366, CN994747	Fruit, bud, leaf	AY792599–AY792601
Mal d 2.01B	(741 nt)	Not yet	Not known	AY792602–AY792604
Mal d 2.02		CO904477, CV082311, CO723595, CN445021	Fruit, bud	Not yet
Mal d 2.03	(738 nt)	CO866703, CO901275, CN495042, CV084040, CO866711, CN444138, CO866347, CN491711	Fruit, flower, bud	Not yet
Mal d 4.01	AF129426, AJ507457, CN497211, CO868479, CV129232, CV130122, CN495799, CN996348, CO865926 (396 nt)	CN488488, CO068373, CN993251, CO755896	Fruit, bud, flower	AY792606–AY792609
Mal d 4.02A	AF129427, AJ507458, CV086259, CV085867, CV084673 (396 nt)	CV084691, CN492781, CO066892, CO756641, CO756641, CO755013	Fruit, bud	AY792610–AY792613
Mal d 4.02B	CN578860, CN992900, CV082623 (396 nt)	CO415431, CN581130, CN493956, CN490579, CN491268, CO899800	Bud, flower	Not yet
Mal d 4.03A	AF129428, AJ507459, CN993521 (396 nt)	CO755279, CO867336, CO867114	Fruit, flower	AY792614–AY792617
Mal d 4.03B	CO722645, CO052497, CO418275 (396 nt)	CO051739, CV086318, CO415968, CN494198, CO067937	Fruit, flower	Not yet

[a] *ORF* open reading frame. Original ORF length ofAF090143 was reported as 738 nt, but we assign it to the 741 group based on the overall sequence identity
[b] *EST* expressed sequence tag, only some of the longer EST sequences were included

linked. Proteomic analyses will be helpful to identify those genes expressed in apple fruit.

For conventional breeding, it is essential to investigate the allelic diversity of the apple allergen genes among the current apple cultivars and to pinpoint individual isoforms relevant for low and high allergenicity. Once this has been accomplished, breeding for low allergenic cultivars becomes feasible with the aid of molecular markers. Another alternative to obtain hypoallergenic cultivars is the use of genetic modification. Currently, the silencing of the *Mal d 1* genes by RNA interference is under investigation, with promising result in leaves of in vitro growing apple shoots (Gilissen et al. 2005).

Acknowledgements This research was performed with financial support from the Commission of the European Communities for the projects SAFE QLK1-CT-2000-01394 and exploited unpublished results from FAIR5 CT97-3898 (DARE: Durable Apple Resistance in Europe). We acknowledge Dr. Bernard Koller and Professor Cesare Gessler for part of SSR data on the linkage maps of Prima×Fiesta and Fiesta × Discovery. We thank professors Evert Jacobsen and Richard Visser for critical reading of this manuscript, Dr. Yan Ma for checking the profilin sequences. Z. S. Gao is grateful to the International Agricultural Centre, Wageningen, and the ICSC-World Laboratory, Lausanne for providing fellowships to conduct this research.

References

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

Breiteneder H (2004) Thaumatin-like proteins—a new family of pollen and fruit allergens. Allergy 59:479–481

Chevreau E, Laurens F (1987) The pattern of inheritance in apple (*Malus domestica* Borkh.): further results from leaf isozyme analysis. Theor Appl Genet 75:90–95

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. Plant Physiol 124:1483–1492

Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AHM, Kodde J, Breiteneder H, Hoffmann-Sommergruber H, Bosch D, Gilissen LJWJ (2005a) 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 WE, Schaart JG, Van der Meer IM, Kodde L, Laimer M, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJWJ (2005b). 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

Gilissen LJWJ, Bolhaar STH, Matos C, Boone MJ, Krens FA, Zuidmeer L, Van Leeuwen A, Hoffmann-Sommergruber K, Knulst AC, Bosch D, Van de Weg WE, Van Ree R (2005) Silencing the major apple allergen Mal d 1 using the RNA interference approach. J Allergy Clin Immunol 115:364–369

Han CU, Lee CH, Jang KS, Choi GJ, Lim HK, Kim JC, Ahn SN, Choi JE, Cha JS, Kim HT, Cho KY, Lee SW (2004) Identification of rice genes induced in a rice blast-resistant mutant. Mol Cells 17:462–468

Hoffmann-Sommergruber K, Radauer C (2004) Bet v 1-homologous allergens. In: Mills ENC, Shewry PR (eds) Plant food allergens. Blackwell, Oxford, pp 125–140

Hsieh LS, Moos M, Lin Y (1995) Characterization of apple 18 and 31 kd allergens by microsequencing and evaluation of their content during storage and ripening. J Allergy Clin Immunol 96:960–970

Huang S, McDowell JM, Weise MJ, Meagher RB (1996) The *Arabidopsis* profilin gene family. Evidence for an ancient split between constitutive and pollen-specific profilin genes. Plant Physiol 111:115–126

Ibeas JI, Lee H, Damsz B, Prasad DT, Pardo JM, Hasegawa PM, Bressan RA and Narasimhan L (2000) Fungal cell wall phosphomannans facilitate the toxic activity of a plant PR-5 protein. Plant J 23:375–383

Koiwa H, Sato F, Yamada Y (1994) Characterization of accumulation of tobacco PR-5 proteins by the IEF-immunoblot analysis. Plant Cell Physiol 12:245–264

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

Maliepaard C, Alston FH, Van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, Van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, Den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-Van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

McKenna ST, Vidali L, Hepler PK (2004) Profilin inhibits pollen tube growth through actin-binding, but not poly-L-proline-binding. Planta 218:906–915

McKinney EC, Kandasmy MK, Meagher RB (2001) Small changes in the regulation of one *Arabidopsis* profilin isovariant, PRF1, alter seedling development. Plant Cell 13:1179–1191

Oh DH, Song KJ, Shin YK, Chung W (2000) 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 64:355–362

Radauer C, Hoffmann-Sommergruber K (2004) Profilins. In: Mills ENC, Shewry PR (eds) Plant food allergens. Blackwell, Oxford, pp 105–124

Ramachandran S, Christensen HEM, Ishimaru Y, Dong CH, Chao-Ming W, Cleary A, Chua NH (2000) Profilin play a role in cell elongation, cell shape maintenance, and flowering in *Arabidopsis*. Plant Physiol 124:1637–1647

Rodriguez-Perez R, Fernandez-Rivas M, Gonzalez-Mancebo E, Sanchez-Monge R, Diaz-Perales A, Salcedo G (2003) Peach profilin: cloning, heterologous expression and cross-reactivity with Bet v 2. Allergy 58:635–640

Scheurer S, Wangorsch A, Haustein D, Vieths S (2000) Cloning of the minor allergen Api g 4 profilin from celery (*Apium graveolens*) and its cross-reactivity with birch pollen profilin Bet v 2. Clin Exp Allergy 30:962–971

Scheurer S, Wangorsch A, Nerkamp J, Skov PS, Ballmer-Weber B, Wuthrich B, Haustein D, Vieths S (2001) Cross-reactivity 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 Bet v 2. J Chromatogra B 756:315–325

Simpson CG, Thow G, Clark GP, Jennings SN, Watters JA, Brown JWS (2002) Mutational analysis of a plant branchpoint and polypyrimidine tract required for constitutive splicing of a mini-exon. RNA 8:47–56

Valenta R, Duchene M, Pettenburger K, Breitenbach M, Sillaber C, Valenta P, Bettelheim P, Rumpold H, Kraft D, Scheiner O (1991) Identification of profilin as a novel pollen allergen. IgE auto-reactivity in sensitized individuals. Science 253:557–560

Van Loon LG, Van Strien EA (1999) The family of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type protein. Physiol Mol Plant Pathol 55:85–97

Van Ooijen J, Voorrips R (2001) Joinmap^{TM} version 3.0: powerful genetic linkage mapping made easy. Plant Research International, Wageningen

Van Ree R, Fernandez-Rivas M, Cuevas M, Van Wijngaarden M, Aalberse RC (1995) Pollen-related allergy to peach and apple: an important role for profilin. J Allergy Clin Immunol 95:726–734

Velazhahan V, Muthukrishnan S (2003) Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata*. Biol Plantarum 47:347–354

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

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 49:262–271

Voorrips R (2001) MapChart version 2.0: Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen

Wensing M, Akkerdaas JH, van Leeuwen WA, Stapel SO, Bruijnzeel CAFM, Aalberse RC, Bast BJEG, Knulst AC, van Ree R (2002) IgE to Bet v 1 and profilin: cross-reactivity patterns and clinical relevance. J Allergy Clin Immunol 110:435–442

Westphal S, Kempt W, Foeisch K, Retzek M, Vieths S, Scheurer S (2004) Tomato profilin Lyc e 1: IgE cross-reactivity and allergenic potency. Allergy 59:526–532

Ye Q, Xu Y, Yan F, Tang L, Chen F (2001) Molecular cloning and characterization of rice pollen profilin. Acta Biochim Biophys Sin 33:452–456