Localization of candidate allergen genes on the apple (*Malus domestica*) genome and their putative allergenicity

Gao Zhongshan

Promotoren:

Prof. dr. ir. E. Jacobsen

Hoogleraar in de Plantenveredeling, Wageningen Universiteit

Prof. dr. R.G. F. Visser

Hoogleraar in de Plantenveredeling, Wageningen Universiteit

Co-promotoren:

Dr. ir. W. E. van de Weg

Senior onderzoeker, Plant Research International, Wageningen Universiteit en Researchcentrum

Dr. L. J. W. J. Gilissen

Senior onderzoeker, Plant Research International, Wageningen Universiteit en Researchcentrum

Promotiecommissie:

Prof. dr. ir. H. F. J. Savelkoul, Wageningen Universiteit

Prof. dr. ir. M. A. J. S. van Boekel, Wageningen Universiteit

Prof. dr. R. F. Hoekstra, Wageningen Universiteit

Prof. dr. ir. L. J. L. D. van Griensven, Radboud Universiteit Nijmegen

Localization of candidate allergen genes on the apple (*Malus domestica*) genome and their putative allergenicity

Gao Zhongshan

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman
in het openbaar te verdedigen
op donderdag 30 juni 2005
des namiddags te vier uur in de Aula

CIP-DATA Koninklijke biobliotheek, Den Haag

Gao, Zhongshan
Localization of candidate allergen genes on the apple (*Malus domestica*) genome and their putative allergenicity

Ph.D. Thesis Wageningen University (2005), The Netherlands – with references – with summary in English, Dutch and Chinese
ISBN:90-8504-242-9

Contents

Chapter 1	Introduction	1
Chapter 2	Linkage map positions and allelic diversity of two *Mal d 3* (non-specific lipid transfer protein) genes in the cultivated apple (*Malus domestica*)	11
Chapter 3	Genomic cloning and linkage mapping of the *Mal d 1* (PR-10) gene family in apple (*Malus domestica*)	33
Chapter 4	Genomic characterisation and linkage mapping of the apple allergen genes *Mal d 2* (thaumatin-like protein) and *Mal d 4* (profilin)	55
Chapter 5	Allelic constitution of *Mal d 1* genes on linkage group 16 is related to the differences in allergenicity among apple cultivars	75
Chapter 6	General discussion	99
References		113
Summary		127
Samenvatting (summary in Dutch)		131
Summary in Chinese		137
Acknowledgements		141
Curriculum vitae		143
List of publications		144

Chapter 1

Introduction

Apple

Apple is the most important fruit produced in temperate climate regions of the world. Fresh apple is highly recommended as a healthy food. The world apple production in the marketing year 2003/04 was estimated 43 million tons (USDA/FAS). Traditional apple cultivars are Golden Delicious, Red Delicious, Cox, McIntosch, and several important new cultivars are Elstar, Granny Smith, Gala, Fuji, Jonagold and Braeburn. According to production efficiency, infrastructure and marketing, the leading world apple suppliers are New Zealand, Chile, The Netherlands, Austria and France (Belrose, Inc. 2001). The Netherlands has a very efficient production system and a long history of apple breeding. Elstar and Santana are the two most important cultivars released in The Netherlands. China started to introduce western apple cultivars from the USA and Europe since 1871 (Xin et al. 1996). After massive planting of apple trees in the 1980's and 1990's, China is now the main producer of apples, accounting for about 33% of the world's output and the top apple exporter.

Apple is a woody crop, a member of the Rosaceae family and the subfamily Maloideae. The origin of the cultivated apple is likely the result of interspecific hybridization, so the binominal *Malus x domestica* Borkh. has generally been accepted as an appropriate scientific name, replacing the previously used *M. pumila* (Korban and Skirvin 1984). The majority of apple cultivars is diploid (2n=34). Sax (1933) suggested that all Maloideae are allopolyploids, derived from doubled hybrids of two remote ancestral members of the subfamily Spiraeoideae (x=9) and Prunoideae (x=8). *Malus sieversii* Lebed., a wild apple species native to Central Asia, has been recognized as a major progenitor of the domesticated apple (Watkins, 1995). Molecular genetic markers indicated that the cultivated apple is most closely related to a series of *Malus* species (Harris et al. 2002). Till now, there is no conclusive answer to the origin of the domesticated apple. More phylogenetically informative markers are needed to support the hypotheses about its hybridization origin. Apple has a relatively small genome of about 1.6 pg/2C per nuleus for diploid apple cultivars, which is about three times the genome size of peach (Dickson et al. 1992) and seven times of *Arabidopsis thaliana*.

Apple cultivars are vegetatively propagated to maintain their genetic constitution. Some good cultivars have been used in production for several decades. Apple is not an easy crop for genetic studies and breeding. The long juvenile period, large field space requirement for planting and growing of seedlings, and the costs to maintain these plantings limit large-scale establishment and evaluation of apple breeding material (Maliepaard, 2000). Moreover, apple is highly heterozygous because of its gametophytic self-incompatibility system. It usually takes at least 20 years from crossing to the release of a new cultivar. To overcome these difficulties, molecular markers linked to important traits are the most promising modern tool to date. With these markers, the promising genotypes could be selected at young seedling stage long before they start bearing fruits.

Apple linkage maps and their application

Hemmat et al. (1994) published the first molecular linkage map on apple, covering only 15 of the 17 linkage groups with over 400 markers (most are randomly amplified polymorphic DNA (RAPD) markers). Because RAPD markers and other types of dominant markers, like AFLP are generally not transferable to other populations, the inclusion of additional codominant RFLP and SSR markers became essential. Meanwhile, an EU project 'The development of the European apple crop' (AIR-3: CT920473) was initiated in 1992 to construct a molecular marker linkage map. In 1998, the first apple reference map from the cross Prima x Fiesta (Maliepaard et al. 1998) was published, which aligned 17 linkage groups from male and female parents that were putatively corresponding to the seventeen haploid chromosomes. Since then, this map has been continuously updated, and now it covers around 1100 cM with 616 markers (RAPD, AFLP, RFLP, SSR, EST,) and ten monogenetic traits (Van de Weg, unpublished). This extended map serves as a world wide reference for the denotation and orientation of linkage groups of all recent molecular marker maps in apple and pear (Liebhard et al. 2002, 2003a; Yamamoto et al. 2004)

A comprehensive set of mapped SSR markers (Liebhard et al. 2002) simplified the alignment of different molecular marker maps and accelerated research on marker-trait associations. This allowed a directed genome scanning in BSA approaches (Patocchi et al. 2004) or by the development of an innovative procedure to assess marker trait associations and functional allelic diversity of identified genes by using multiple crosses and cultivars that are related by their pedigrees (Van de Weg et al. 2004). Many of these SSR markers could be

transferred to pear (Yamamoto et al. 2001, 2004; Dondini et al. 2004; Pierantoni et al. 2004). All these linkage maps are indispensable to study genome structure, and to localize genes of interest. They permitted to identify a few quantitative trait loci (QTLs) and major genes for fruit quality (Cheng et al. 1996; King et al, 2001; Maliepaard et al 2001; Liebhard et al. 2003b); disease resistance genes (Roche et al. 1997, Cheng et al. 1998; King et al. 1998; Durel et al.2000; Hemmat and Brown 2002; Evans and James 2003; Durell et al. 2003, Patocchi et al. 2004, Bus et al. 2005, James et al. 2005); and resistance gene analogues (Baldi et al. 2004; Calenge et al. 2005). Multiple loci of the same SSR and RFLP markers indicated several homoeologous linkage groups and segments across the linkage groups, such as LG 2 to LG 7; LG 3 to LG 11; LG 4 to LG 12; LG 5 to LG 10; LG 8 to part of LG 15; LG 9 to LG 17; and LG 13 to LG 16 (Maliepaard et al. 1998, Van de Weg, unpublished). These duplicated markers directly support the hybrid origin of the cultivated apple.

Food allergy in general

Food allergy is a hypersensitive reaction to normally harmless substances (allergens, mostly proteins) and involves humoral immune responses, mediated by immunoglobulin (Ig) E (with E derived from erythema) synthesized by B-lymphocytes (Bohle 2004). During sensitization, the immune system produces specific IgE antibodies against food allergens. The IgE antibodies bind to high affinity receptors on the surface of mast cells and basophiles. Upon re-exposure to the ingested food, cross-linking of the IgE antibodies with the allergen triggers the release of mediators, e.g., histamine, tryptase, and tumour nercrosis factor, causing the acute phase of the allergic immediate reaction. Clinically, the IgE-mediated hypersensitive reaction occurs within minutes to 2 h after ingestion (Fernandez-Rives and Miles 2004). Atopic patients have a genetic predisposition to develop allergy symptoms. There are two classes of food allergy. A class-I food allergy, in which the immune reaction takes place in the gastrointestinal tract, is especially found in children. The allergens involved in this class are very stable, and resistant to heat and digestion processes. A class-II food allergy is initiated by proteins that come into contact with the immune system through inhalation; e.g. the inhalation of pollen from several tree species and grasses. Class-II food allergies are generally found in adults.

Food allergy symptoms are usually mild local reactions of the oral cavity, the so-called oral allergy syndrome (OAS). Especially, fresh fruits and vegetables are causing such allergy

problems. Other symptoms are also observed in organs like skin (local or general urticaria, atopic eczema), gastro-intestinal tract (cramps, diarrhoea, vomiting), nose and lungs (rhinitis and asthma) and the cardiovascular system (anaphylactic shock) (Fernandez-Rivas and Miles, 2004).

The prevalence of food allergy is especially high (at least in the reported numbers) in the westernized countries. About 1 in 100 adults and nearly 1 in 10 children suffers from food allergy (Fernandez-Rivas and Miles 2004). Food allergy is an unsolved problem of great economic, social, and personal significance. To improve the quality of life for allergy patients, multidisciplinary approaches integrating medical science, food science, agriculture, environment and social science should be implemented to tackle this problem in Europe in the coming years. A number of projects, such as EU-PROTALL (FAIR CT-98-4356) and EU-SAFE (QLK1-CT-2000-01394) and the Allergy Consortium Wageningen have obtained substantial data and valuable insight into this complex problem. Development of strategies for allergy prevention is in progress.

Food allergens belong to a limited number of protein families (Breiteneder and Mills 2005) and are characterized by their biochemical and physiological properties. Pathogenesis related proteins such as Bet v 1 homologues (PR-10 proteins), thaumatin-like protein (TLP) (PR-5 proteins), and non-specific lipid transfer proteins (nsLTPs) (PR-14 proteins), represent several well-known allergens (Breiteneder and Ebner, 2000). TLP and nsLTP have similar intramolecular disulfide bonds attributing to thermal stability and resistance to proteolysis. Since PR proteins are induced by biotic and abiotic stresses, it is not a surprise that most allergens are present in the skin of fruits (Fernandez-Rivas and Cuevas 1999). Often, these proteins are highly conserved through the evolutionary trees of plants and animals, such as profilins and lipid transfer proteins. According to the clinical relevance and prevalence, major and minor allergens are classified for different populations and patients. Allergens are named according to the scientific name of the source material, the first three letters of the genus name, first letter of the species name (King et al. 1995), for example, ‘Mal d’ represents apple (*Malus domestica*) allergen. Different allergens are designated by Arabic numbers according to the time of their identification. The plain text refer to proteins, italics refer to genes. For example, Mal d 1 refers to first apple allergen and *Mal d 1* refers to this gene class.

Allergic cross-reactivity

Allergic cross-reactivity can occur in a patient reacting to similar allergens from different origins through the same IgE antibody type. IgE antibodies are not necessarily specific to a unique allergen. They will bind to the allergen that initiated IgE production by the immune system (primary sensitization) with the highest affinity, but can also bind to structurally similar allergens. The binding of IgE with the allergen precisely occurs at special recognition sites, called epitopes. In general, allergens that cross-react belong to the same protein family with a significant amino acid similarity and more importantly a high similarity in their three-dimensional structures (Neudecker et al. 2001; Jenkins et al. 2005). PR-10 (Bet v 1), profilin and nsLTP (non-specific lipid transfer proteins) are the most common allergen families in plants (Weber 2003). Bet v 1 cross-reactive allergens have been found in many fruits and vegetables and represent a large category, including apple, pear, cherry, strawberry, carrot, celery, etc. (Hoffmann-Sommergruber and Radauer, 2004). This list is growing along with new clinical reports and biological identifications. Recently, Bet v 1 homologues have been found in jackfruit (*Artocarpus integrifolia*, a tropical fruit native to India, East Asia, East Africa, South America), and sharon fruit (persimmon, belonging to the *Ebenaceae* family, native to China and Japan; Bolhaar et al. 2004a, 2005). These homologues are ubiquitously present in plants and are not limited to taxonomically specific species.

Apple allergy and allergen genes

Although apple is a healthy diet to the vast majority of the population, some individuals can not eat this fruit because it provokes allergic reactions. Apple allergy is common in Europe and affects about 2% of the population (Bolhaar, 2004b). The symptom of apple allergy is generally limited to oral allergy syndrome. Apple allergy was reported for the first time in Sweden in 1948, where patients with spring hay fever often complained of itching, tingling and swelling of the lips, tongue and throat, as well as sneezing and irritating of the eyes, after eating apples (Juhlin-Dannfelt, 1948). Mature apple fruits have more allergenic potency than young and freshly picked fruits (Vieths et al. 1993). Eriksson et al. (1982) demonstrated that birch-pollen allergic patients reported frequently (70%) to have hypersensitivity to various nuts, fruits and vegetables. Another type of apple allergy that is not related to birch-pollen allergy was reported mainly in Southern Europe countries, especially in Spain and Italy. In apple, four classes of allergens (Mal d 1, -2, -3 and -4) have been identified through

immunoblotting, protein and DNA sequencing. Below, we described these allergen classes in more detail.

Mal d 1

About 50–75% birch pollen-sensitized patients suffer from an oral allergy syndrome (OAS) after eating apples (Ebner et al. 1991; Pauli et al. 1996). This allergy results from the cross-reactivity between Bet v 1, the sensitizing birch allergen, and Mal d 1, the major apple allergen (Calkhoven et al 1987, Ebner et al. 1991, Fritsch et al. 1998). Both Bet v 1 and Mal d 1 belong to a group of pathogenesis-related (PR)10 proteins (Breiteneder et al. 2000). PR-10 gene families appear ubiquitous in plants (Wen et al. 1997; Hoffmann-Sommergruber et al. 2000; Liu and Ekramoddoullah 2004). Mal d 1 has been identified as a 17–18 kD protein of 158–159 amino acids encoded by 480–483 nucleotides (Vieths et al. 1995; Schoning et al. 1996; Hoffmann-Sommergruber et al. 1997, Puehringer et al. 2002). The molecular relationship between Mal d 1 and Bet v 1 has been demonstrated at the amino acid level (Vieths et al. 1994; Vieths et al. 1995; Vieths 1997; Helsper et al. 2002) and at the DNA level (Vanek-Krebitz et al. 1995; Holm et al. 2001). Recently, several new Mal d 1 genes were identified in fruit by cDNA sequencing (Beuning et al. 2004). Over 30 Mal d 1 sequences in seven groups (Puehringer et al. 2003) were available in the GenBank and EBML databases at the beginning of this research.

Mal d 1 allergy is predominant in Central and Northern Europe. But different apple cultivars can elicit mild or severe allergic reactions in atopic patients. The allergenicity of Golden Delicious apples increased significantly, but to a lesser degree of Boskoop apples during maturation (Vieths et al. 1993). Golden Delicious is highly allergenic as experienced by many apple-allergy patients, whereas Gloster causes only mild reactions (Vieths et al. 1994; Hsieh et al. 1995). Since both high allergenic cvs Golden Delicious and Granny Smith have a green skin, some clinical reports refer to “green apple allergy”. Recently, the allergenicity of a panel of apple cultivars were ranked according to their reaction by skin prick test. The significant difference in allergenicity between two cultivars was confirmed by double-blind placebo-controlled food challenges (DBPCFC) (Bolhaar 2004b). Golden Delicious proved to be high allergenic, whereas the new cultivar Santana (Plant Research International, Wageningen) proved to be low allergenic, with 30-fold reduced response in the DBPCFC (Bolhaar 2004b). Previously, differences in allergenecity among cultivars have been assigned to differences in Mal d 1 content, since the low-allergenic Boskoop and Gloster

fruits contained very low amount of this protein (Vieths et al 1993, Vieths et al. 1994). Isoforms and mutations of Bet v 1 and Mal d 1 with amino acid sequence variation displayed different binding capacities for specific IgE antibodies from allergic patients (Ferreira et al. 1996; Son et al. 1999). But no specific Mal d 1 isoform has been identified for low allergenicity. Apple allergy appeared to increase significantly during the birch pollen season (Hansen et al. 2001). Allergic reactions to apple could not be related to pesticide residues, which contrasts to the patient's believe that they react particularly to "chemically treated fruits" (Beil et al. 2001).

Mal d 2

Mal d 2, a homologue of the thaumatin-like protein (TLP), is one of the major protein constituents of a mature apple fruit. Most plant TLPs have 16 conserved cysteines that form eight disulfide bonds contributing to the protein's character of resistance to protease and heat (Breiteneder 2004). TLPs belong to the PR-5 family of pathogenesis-related proteins (Van Loon and Van Strien 1999). Several researches provided evidence that TLPs play a role in the plant defence against pathogens (Ibeas et al. 2000; Venisse et al. 2002; Velazhahan and Muthurishnan 2003; Han et al. 2004). Hsieh et al. (1995) were the first who identified a 31 kDa apple TLP as a major allergen. Later on, the encoding cDNA has been isolated from apple fruits (Oh, et al. 2000). Recently, a 23.2 kDa apple TLP deduced from a full-length cDNA was named Mal d 2 allergen and characterized as an antifungal protein (Krebitz et al. 2003). The major geographical area for the occurrence of Mal d 2 allergy is still unclear.

Mal d 3

Mal d 3 is a small protein of 9-kDa, a member of the non-specific lipid transfer proteins (nsLTPs) family. nsLTPs have been identified as major allergens in several Rosaceae fruits, including apple (Pastorello et al. 1999a; Diaz-Perales et al. 2002), peach (Pastorello et al. 1999b; Ballmer-Weber 2002), apricot (Pastorello et al. 2000), plum (Pastorello et al. 2001) and cherry (Scheurer et al. 2001). The prevalence of allergy to nsLTP-containing fruits is especially high in Mediterranean communities (Ballmer-Weber 2002). Peach nsLTP (Pru p3) was recognized as the primary sensitizing agent that was cross-reacting with other nsLTPs (Asero et al. 2003). Two Mal d 3 cDNA sequences obtained from apple fruit were published in the GenBank at the beginning of this study.

Mal d 4

Mal d 4 belong to the profilin protein family. Profilins are small (12-15 kDa) cytosolic proteins present in all eukaryotic cells. These profilins display striking features with regard to the conserved amino acid sequence, amino acid domains and structure (Radauer and Hoffmann-Sommergruber 2004). In 1991, birch pollen profilin was identified as the relevant allergen Bet v 2 (Valenta et al. 1991). Profilins from many fruit have been reported as allergens in recent years (Scheurer et al. 2000 and 2001; Asero et al. 2003; Wensing et al. 2002; Westphal et al 2004). It has assumed that profilins in foods cross-react to birch profilin Bet v 2 (Van Ree et al. 1995). Profilin allergy is more relevant in the Mediterranean area of Southern Europe (Van Ree et al. 1995; Hoffmann-Sommergruber and Radauer 2004). Three diverse mRNA sequences encoding three isoallergens of apple profilin were available in the GenBank/EMBL nucleotide database in 2001.

Scope of the thesis

The research described in this thesis was a part of the recently finished EU-SAFE project (Plant food allergies: field to table strategies for reducing their incidence in Europe, QLK-CT-2000-01394). This project was initiated to characterize the relevant allergens and their relationships to severe versus mild allergic symptomatology using apple as a model. One of the objectives was to map the allergen genes on established apple molecular marker linkage maps and subsequently to identify the allergen genes contributing to allergenicity.

Apple cultivars differ considerably in allergenicity. To design an apple-breeding programme aiming at hypoallergenic cultivars, it is essential to unravel the genetic basis of these differences at the qualitative and quantitative level of isoallergens and their variants. This thesis aims to provide the genome sequences and linkage map position as the basis for the elucidation of difference in allergenicity among different apple cultivars. The employed strategy is to clone and sequence all apple allergen genes from genomic DNA of the two apple cultivars Prima and Fiesta, which are the parents of a European reference mapping population. DNA sequencing allows the detection of single nucleotide polymorphisms (SNPs) between alleles of one locus. Molecular markers based on these sequences were developed and the segregating markers were mapped on established maps. Chapter 2 describes the linkage map position of two *Mal d 3* (non-specific lipid transfer protein) genes and their allelic diversity among ten cultivars. Chapters 3 and 4 present genomic sequences and linkage

map positions of three gene families, *Mal d 1*, *Mal d 2* and *Mal d 4*. Chapter 5 elucidates that the *Mal d 1* genes on linkage group 16 are closely associated with difference in allergenicity among various apple cultivars. This conclusion was reached by analysis of the changes in both the allelic constitution of a series of *Mal d 1* genes' and the allergenicty of a series of cultivars that were genetically related by their pedigrees. The best association was obtained with the allelic constitution of *Mal d 1.06A*. The developed molecular markers allow the genotyping of other cultivars, and may thus support future research on the generalization of our results.

Finally, several aspects related to gene cloning, linkage mapping, gene families and consequences for genetic analysis and breeding perspectives for hypoallergenic apple cultivars are summarized in the general discussion.

Chapter 2

Linkage map positions and allelic diversity of two *Mal d 3* (non-specific lipid transfer protein) genes in the cultivated apple (*Malus domestica*)

Z. S. Gao[1,2], W. E. van de Weg [1,3], J. G. Schaart[3], I. M. van der Meer[3], L. Kodde[3], M. Laimer[4], H. Breiteneder[5], K. Hoffmann-Sommergruber[5], L. J. W. J. Gilissen[1,3]

1: Allergy Consortium Wageningen, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands.
2: Experimental Nursery of Shanxi Forestry Department, Beiying Nan Lu 52, Taiyuan, Shanxi, 030031, People's Republic of China.
3: Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands.
4: Plant Biotechnology Unit, IAM, University of Natural Resources and Applied Life Sciences, Nussdorfer Laende 11, A-1190 Vienna, Austria
5: Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, A-1090 Vienna, Austria

Published in *Theoretical and Applied Genetics* (2005)110:479–491

Abstract

Non-specific lipid transfer proteins (nsLTPs) of Rosaceae fruits, such as peach, apricot, cherry, plum, and apple, represent major allergens for Mediterranean atopic populations. As a first step in elucidating the genetics of non-specific LTP, we directed the research reported here towards identifying the number and location of nsLTP (Mal d 3) genes in the apple genome and determining their allelic diversity. PCR cloning was initially performed on two cultivars, Prima and Fiesta, parents of a core apple mapping population in Europe, based on two *Mal d 3* sequences (AF221502 and AJ277164) in the GenBank. This resulted in the identification of two distinct sequences (representing two genes) encoding the mature nsLTP proteins. One is identical to accession AF221502 and has been named *Mal d 3.01*, and the other is new and has been named *Mal d 3.02*. Subsequent genome walking in the upstream direction and DNA polymorphism analysis revealed that these two genes are intronless and that they could be mapped on two homoeologous segments of linkage groups 12 and 4, respectively. Further cloning and sequencing of the coding and upstream region of both *Mal d 3* genes in eight cultivars was performed to identify allelic variation. Assessment of the deduced nsLTP amino acid sequences gave two variants at protein level for *Mal d 3.01* and three for *Mal d 3.02*. Consequences for allergen nomenclature and the breeding of low allergenic apple cultivars are discussed.

Introduction

Fruits are generally considered as an important component of a healthy diet for humans. However, in some individuals, several kinds of fruit may also cause severe allergic reactions. A food allergy results when the immune system reacts improperly to specific proteins, once sensitized, individuals can become allergic to homologous proteins of other food (fruits) via cross-reactivity. There are two classes of food allergy. A class-I food allergy, in which the immune reaction takes place in the gastrointestinal tract, is especially found in children. The allergens involved in this class are very stable. A class-II food allergy is initiated by proteins that contact with the immune system through inhalation; e.g. the inhalation of pollen from several tree species and grasses. Class-II food allergies are generally found in adults.

In apple, two major allergens have been identified: (1) a non-specific lipid transfer protein (nsLTP) called Mal d 3, which causes a class-I food allergy (Pastorello et al. 1999a); (2) Mal d 1, another major allergen, is a homologue of the sensitizing Bet v 1 allergen from birch pollen (Breiteneder and Ebner 2000) and causes a class-II food allergy. Many plant food allergens can be classified as members of the groups of pathogenesis-related (PR) proteins. Mal d 1 and Mal d 3 belong to the PR-10 and the PR-14 family, respectively (Breiteneder and Ebner 2000; Mills et al. 2003).

The nsLTPs were discovered about 30 years ago (for review, see Kader 1996). They form a multigene family encoding 9-kDa proteins (90–95 amino acids) that are distributed throughout the plant kingdom. In *Arabidopsis*, at least 6 individual nsLTP genes have been found on three chromosomes (Arondel et al. 2000). Kader (1996) suggested that nsLTP are pathogenesis-related since they can participate in plant defense reactions through anti-fungal and antibacterial activities. It has recently been confirmed that nsLTP lacks any specificity for fatty acids, phospholipids or the cutin monomers (Douliez et al. 2001). With respect to food allergies, most of the nsLTP allergens identified to date belong to the nsLTP1 subfamily, whose members have been identified as major allergens in several Rosaceae fruits, including peach (Pastorello et al. 1999b; Ballmer-Weber 2002), apple (Pastorello et al. 1999a; Diaz-Perales et al. 2002), apricot (Pastorello et al. 2000), plum (Pastorello et al. 2001) and cherry (Scheurer et al. 2001). The prevalence of allergy to nsLTP-containing fruits is especially high in Mediterranean communities (Ballmer-Weber 2002). With respect to Mal d 3, two cDNA sequences obtained from apple fruit were published in the GenBank at the beginning of this study. The first, AF 221502, comprises the complete coding sequence for the Mal d 3 precursor (including a signal peptide), while the second, AJ277164 (Diaz-Perales et al. 2002), encodes for the mature Mal d 3 protein. These two sequences differ in only one nucleotide in

the coding region for the mature protein. This sequence information was used as the start for genomic cloning.

The aim of this study was to contribute to the successful breeding of apple cultivars with reduced Mal d 3 allergenicity. At present, nothing is known about the genomic localization, the extension of the Mal d 3 gene family and the degree of its allelic variation. To answer these questions, we carried out genomic cloning on ten apple cultivars, including Prima, Fiesta, Jonathan and Discovery, four parental cultivars of three mapping populations. Two distinct Mal d 3 genes were localized on molecular marker linkage maps. Allelic diversity studies in ten cultivars showed the existence of two and three protein variants of both genes, respectively. This is the first report on the linkage map position of apple allergen genes.

Materials and Methods

Cultivars used for genomic cloning and DNA isolation

Apple (*Malus domestica*) cultivars Prima (PM) and Fiesta (FS) were used initially for genomic PCR cloning and sequencing because of the availability of a molecular marker linkage map from the population of PM x FS (Maliepaard et al. 1998). Subsequently, Jonathan (JO) and Discovery (DS) were added because maps from the populations of JO x PM and FS x DS became accessible (Van de Weg et al., unpublished). Ultimately six more cultivars, Golden Delicious (GD), Ingrid Marie (IM), Priscilla (PS), Cox (CO), Red Delicious (RD) and Fuji (FJ), were included in our investigation of allelic variation. These specific six cultivars were chosen because they are either common apple cultivars and/or frequently used in our apple-breeding programme. Golden Delicious, Ingrid Marie and Priscilla are the founders of the newly released cultivar Santana [(GDxIM)xPS] by Plant Research International, Wageningen.

Genomic DNA was extracted from young leaves of the cultivars and the population plants by the CTAB-based large-scale nuclei-isolation method (Roche et al. 1997).

Genomic PCR cloning and sequencing

Four rounds of PCR cloning were performed to obtain sequences of *Mal d 3* genes. The primer sequences, PCR conditions and product sizes are shown in Table 1. In the first round of cloning, *Mal d 3* genomic fragments of Prima and Fiesta were amplified by PCR using primers (Mald3-For and Mald3-Rev1) designed on the basis of two cDNA sequences AF 221502 and AJ277164 (Diaz-Perales et al. 2002). In the second round, genomic walking (see below) was carried out to reveal the upstream sequences of those obtained in the first round.

The third round of PCR cloning was based on the sequences obtained by genome walking for the purpose of searching polymorphisms in four parent cultivars of the three mapping populations, while the fourth round was aimed at investigating allelic diversity of the identified *Mal d 3* genes in eight more cultivars.

Table 1 Primer pairs used for four successive cloning rounds of two different *Mal d 3* genes

Round	Primer sequence (5′- 3′)	Position[b]	PCR[c]	Super-*Taq*	Product[d]	Reference
First	Mald3-For: ATAACATGTGGCCAAGTGA	73–91	60/30	64/2	274	AF221502
	Mald3-Rev1: ACTTCACGGTGGCGCAGTTG					
Second[a]	Mald3.01GSP1: AGGAAGCCCTGCTGCATTGTTA					First round
	Mald3.01GSP2: AGCCAATGCATGGCGCAAGGCTGCTGG					sequences
	Mald3.02GSP1: TGACTCCACACTTTCCAGGAAGCGATT					
	Mald3.02GSP2: GTAGTCAAAGCATGGTACAAGGTTGGA					
Third	Mald3.01UP-For: CATAGTTGTTGAATCATTGACCA	35–57	58/30	60/2	385–390	AY572500
	Mald3.01SSR-Rev: GTTTCTATTCCGTCTTCCTTC	399–419				
Fourth	Mald3.01-For: TATGTTTCTCACCATACTAGCCGA	1,515–1,538	60/30	62/2	901	AY572500
	Mald3-Rev2: TTACTTCACGGTGGCGCAGTTG	2,394–2,415				
	Mald3.02-For: GCTGCTGCTGTTCCTCCAGAT	7–27	60/30	62/2	985–986	AY572517
	Mald3-Rev2: TTACTTCACGGTGGCGCAGTTG	970–991				

[a] The second round was genomic walking, PCR procedures followed the manufacture's instruction (see text)

[b] Refer to reference sequence

[c] Temperature (^{0}C)/number of cycles (*Pfu*)

[d] Nuleotides (number)

In all cloning rounds with the exception of genome walking, PCR was performed in two steps using *Pfu* and *Taq* polymerase. In the first round, the PCR analysis was carried out using *Pfu* polymerase (Stratagene, La Jolla, Calif.) in a total reaction volume of 50 µl, containing 40 ng of genomic DNA (gDNA) of PM or FS, 5 µl of 10x *Pfu* buffer, 20 pmol forward and reverse primers, 0.2 µl 10 m *M* dNTPs, and 2.5 unit of *Pfu* polymerase enzyme. PCR was performed using the following program: an initial 2 min denaturation at 95°C; 25–30 cycles of 2 min at 95°C, 30 s at the optimized annealing temperature (see below), 3 min at 74°C (elongation); a final 10 min at 74°C. If this PCR resulted in the expected fragment, an additional first-round PCR was performed by adding 50 µl reaction mixture containing 5 µl 10x super-*Taq* buffer, 20 pmol of forward and reverse primers, 0.2 µl of 10 m *M* dNTPs, and 0.25 U of super-*Taq* polymerase to obtain A-tailing of the PCR product, which facilitates

subsequent cloning. Amplification was performed for 2 min at 94°C, followed by 2 cycles of 94°C for 1 min, 30 s at the optimized temperature, 2 min at 72°C and completed by 10 min at 72°C. The PCR product was separated on a 1.2 % agarose gel, and the target fragment was excised and purified by Qiaquick Gel Extraction Kit (Hilden, Germany) following the protocol supplied by the manufacturer. PCR products were then ligated into pGEM®-T easy vector (Promega, Madison, Wis.) and used to transform XL1 Blue competent cells (Stratagene) according to the protocol recommended by the supplier. For each fragment, 8–16 white colonies were subcultured for plasmid DNA isolation using the Qiaprep®Turbo BioRot®Kit (Hilden, Germany) and DNA sequencing (ABI 3700 sequencer, Applied Biosystems, Foster City, Calif.).

Genome walking (the second round) was applied to two cultivars, Prima and Fiesta, using the Universal Genome Walker kit (Clontech, Palo Alto, Calif.) to explore polymorphisms in the upstream region. For each cultivar, four libraries were constructed using *Dra*I, *EcoR*V, *Pvu*II and *Stu*I enzymes to digest 2.5 μg of the gDNA. Adaptors were ligated to the digested DNA fragments. Nested PCR was performed using two gene-specific reverse primers GSP1 and GSP2, which were based on new sequences from the first round cloning (Table 1), together with two adaptor primers (AP1 and AP2). The products obtained from two to three of the four libraries were excised from the gel and subsequently purified, ligated, transformed and sequenced as already described.

Single nucleotide polymorphisms (SNPs) were analysed using the SEQMAN programme (DNAstar, Madison, Wis.). The sequence identity plot was created using the CLUSTAR W method by MEGALIGN program (DNAstar). Multiple DNA and amino acid sequence alignments were performed with the GENEDOC programme.

Allele-specific (AS) primers

To discriminate *Mal d 3* allelic sequences, we designed primer pairs using the software programme PRIMER DESIGNER ver.2.0 (Scientific & Educational software, Cary, N.C.). The 3′ end in one or both of forward and reverse primers were exactly located at polymorphic nucleotides found anywhere in both the coding and the upstream sequences. To reduce the chance of primers annealing to the untargeted template, we employed the addition of a designed mismatch at the second or the third position at the 3′ end of the primer. This kind of marker has been called single-nucleotide amplified polymorphisms (SNAP) (Drenkard et al. 2000). In most cases, the designed mismatches were just complementary to the matching nucleotide, which were supposed to have a stronger effect. If this change did not meet the

primer setting criteria, two other nucleotides were examined. In case of ambiguity with respect to the allele specificity of primer pairs at a single annealing temperature, a touch down PCR was performed. Two steps of nested PCR (Bonants et al. 1997) were also applied to detect an allele that lacked a unique SNP.

Test for authenticity of polymorphisms revealed by cloning

The validity of some polymorphisms in the sequences was checked using a two-step SNAP marker test. The candidate allele-specific primer pair was tested first with gradient PCR on the cloned plasmid DNA of an allelic sequence and the results were compared with a negative plasmid DNA control with another allelic sequence. The primer pair was then applied to gDNA of the representative cultivar from which the sequence was obtained and two negative cultivars. In a final step, primer pairs were tested on all ten cultivars examined in this investigation and for which specificity could be validated due to obtained sequence information.

Optimisation of annealing temperature of PCR

This step was followed for each new primer pair. PCR was performed in 20-µl reaction volume containing 20 ng of gDNA or 0.1 ng of cloned plasmid DNA, 2 µl of 10x super-*Taq* buffer (HT Biotechnology, England), 0.2 m *M* of each dNTP, 5 pmol of forward and reverse primer and 0.25 U super-*Taq* polymerase. PCR was carried out using the PTC-200 machine (MJ Research. Waltham, Mass.) and following PCR conditions: an initial denaturation at 94°C for 2.5 min, then 35 cycles at 94°C for 30 s, 50.5°C–66.6 °C gradient for 30 s and 72°C for 1.5 minutes. PCR products were analysed by agarose gel electrophoresis to determine the optimum annealing temperature.

Mapping genes on molecular marker linkage maps

Molecular marker linkage maps from the crosses PM x FS and JO x PM were used to map *Mal d 3*. The PM x FS map (Maliepaard et al. 1998) has been extended within the European DARE project (FAIR5-CT97-3898) (Lespinasse et al. 2000) by Plant Research International and the Swiss Federal Institute of Technology (Van de Weg & Liebhard, unpublished). The JO x PM map was generated at Plant Research International (Van de Weg et al., unpublished). These two maps cover about 2,140 cM each and consist of 700 and 620 markers, respectively. Of the three parental cultivars, two are genetically related: Jonathan is a grandparent of Fiesta [Cox's Orange Pippin x Idared (=Jonathan x Wagner Apfel)]. This relationship is useful

because when one allele is present in homozygous condition in Fiesta, and can not be mapped in PM x FS, it may be mapped in JO x PM.

Grouping and mapping of allele-specific markers were performed with JoinMap ver. 3.0 (Van Ooijen et al. 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).

Nomenclature of the *Mal d 3* sequences

Current allergen nomenclature (King et al. 1995) has been followed and extended to name *Mal d 3* gene sequences by suffixes consisting of six Arabic numerals. The first two numerals refer to different genes of the same allergen, such as *Mal d 3.01* and *Mal d 3.02*. Alleles are differentiated at two levels: firstly, at the level of deduced amino acid sequence of mature nsLTPs (variant), which was designated by the middle two numerals; secondly, at the level of silent mutation in the coding sequence–indicated by a dot and final two numerals. For example, Mal d 3.0101.01 refers to a silent mutation of the first allele (variant) from the *Mal d 3.01* gene. In addition, a final lower case letter was used to indicate DNA polymorphisms in the flanking region of an allele.

Results

Genomic cloning and sequencing of *Mal d 3* genes from Prima and Fiesta

In the first round of PCR cloning using the gDNA of Prima and Fiesta as templates (Table 1), we obtained two distinct sequences, both of which were present in Prima and Fiesta. Both have a length of 274 nt, 2 nt shorter than the published cDNA due to the reverse primer lacking the final 2 nt of the gene, and they share 93% and 84% identity at DNA and amino acid level, respectively. One of the sequences is identical to AF221502 in the GenBank. We next develop two sequence-specific markers and tested these in the three mapping populations, PM x FS, JO x PM and DS x FS; the two markers were present in all samples (data not shown), indicating that these two sequences are not allelic variations at the same locus but are located at different loci, which we denoted as *Mal d 3.01* and *Mal d 3.02*, respectively.

Genome walking (the second round) was carried out on the parental cultivars Prima and Fiesta to identify DNA polymorphism in the upstream region of the two *Mal d 3* genes. When the sequence identified in the first PCR cloning is included in the calculation, the total length of sequences was 2,415–2,418 nt and 990–991 nt sequences for *Mal d 3.01* (accession:

AY572500–AY572502) and *Mal d 3.02* (accession: AY572503–AY572505), respectively, for both cultivars. These are presented as two consensus sequences in Figs. 1 and 2. Both consensus sequences contain a single exon of 348 nt, of which the first 72 nt code for a putative signal peptide that directs the molecule to the proper location and being cleaved off from the mature protein (Kader 1996). These two *Mal d 3* consensus sequences share 89% identity in the last 600-nt region including the coding sequences. Upstream of these 600-nt regions, the similarity is very low. *Mal d 3.01* is highly conserved, not only in the coding region for the mature protein but also up to at least 2 kb upstream. Polymorphisms were only found in a poly-T region about 1,730 nt upstream of the *Mal d 3* reading frame (Fig. 1). The number of T repeats varied from 14 to 16 in Prima and from 15 to 17 in Fiesta with steps of 1 nt. Each of these poly-T sequences was observed in at least two clones. To see whether any of these poly-T alleles segregated in the Prima x Fiesta population, two primers (forward: 5-TTAGCTAGAGACCCATCGAGAG-3; Reverse: 5-GTTTCTATTCCGTCTTCCTTC-3) flanking this region were designed and used to amplify the polymorphic fragments which were visualized by polyacrylamide gel electrophoresis with [^{33}P]-labeling. Prima and Fiesta were showed to have three major bands from 136 to 138 nt and from 137 to 139 nt, respectively, this corresponds to the T repeat numbers and were uniformly present in all descendents. Therefore, it was impossible to map *Mal d 3.01* with the simple sequence repeat (SSR) marker.

To further examine the poly-T region, a third round of PCR cloning (Table 1) was conducted on Prima and Fiesta to confirm the presence of poly-T alleles, as well as on Jonathan and Discovery to find additional polymorphisms. The amplified fragments were about 390 nt. Following sequencing from both sides, all six poly-T alleles reappeared in Prima and Fiesta accordingly, while Discovery possessed four major poly-T repeats of 15, 17, 18 and 19 nt (accession: CL449239, CL449240, CL449241 and CL449242). Surprisingly, Jonathan had 8 identical sequences with 5T and 1A followed by 10T (accession CL449238) and 3 sequences with 15T (CL449236) and 4 sequences of 16T (CL449237). All other sequences flanking this poly-T region were identical. Based on these data, we concluded that multiple number of poly-T sequences within a diploid apple cultivar indicates that *Mal d 3.01* is not a single copy gene.

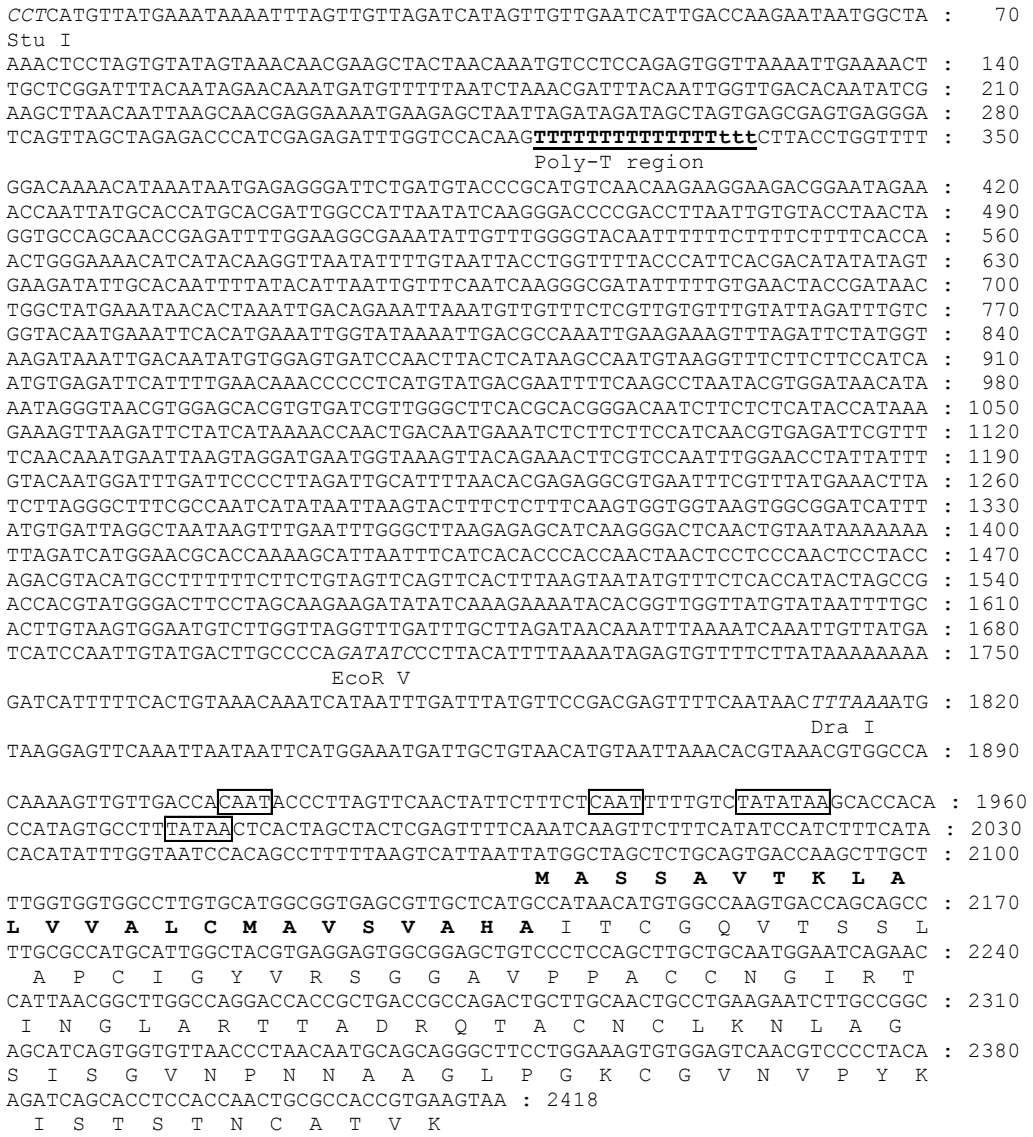

Fig. 1 Consensus sequence of gDNA and deduced amino acids of *Mal d 3.01* derived from apple cultivars Prima (PM) and Fiesta (FS). Nucleotides shown in *italics* indicate the three restriction sites used for genome walking. Putative CAAT and TATA *boxes* for promoter near the start code are indicated. Nucleotides in *bold* and *underlined* indicate poly-T region where T repeats varied by 14, 15, 16 for PM (accession numbers: AY572500, AY572501, AY572502) and 15, 16, 17 for FS (AY572503, AY572504, AY572505). Amino acids in *bold* comprise the signal peptide

Conversely, *Mal d 3.02* proved to be more variable than *Mal d 3.01*. When we aligned the two sequences from Prima and one sequence from Fiesta of an even shorter length of 991 nt, we were able to identify 14 SNPs. Most of them were located in the upstream region, only one was located in the part encoding for the signal peptide (Fig. 2).

Fig. 2 Consensus sequence of gDNA and deduced amino acids of *Mal d3.02* derived from cvs. Prima and Fiesta. Nucleotides in *italic* indicate the restriction sites used for the genome walking. Bold and underlined letters denote the SNPs within and between alleles identified from PM and FS: Y= C or T; W=A or T; R=G or A; M=A or C; S=C or G. Putative CAAT and TATA *boxes* for promoter near the start code were indicated. In the poly-T region, one allele from PM is one T shorter. Amino acids in bold indicate the signal peptide. Accession nos: AY572517, AY572518, AY572519.

Development of allele-specific markers to map *Mal d 3.01* and *Mal d 3.02*

Polymorphisms in the upstream of Mal d 3 coding sequences were used to create markers to map *Mal d 3.01* and *Mal d 3.02*. For *Mal d 3.01*, a forward primer (Mald 3.01Jo-f1, Table 3) was designed based on a single SNP (A/T) in the poly-T region of Jonathan with an introduced mismatch base G in the third position. Together with a common reverse primer (Mald 3.01SSR-r, Table 3), we first performed gradient PCR on DNA templates from Jonathan and Prima. An expected clear band of 116 nt occurred at a range of annealing temperatures (Tm, 51.4–60.6°C; Fig. 3a) for Jonathan but not for Prima. Secondly, a touch-down PCR program consisting of annealing temperatures of 58°C for 5 cycles and 56 °C for 35 cycles was applied to Jonathan, Prima and 12 of their descendants. This PCR test proved that Mal d 3.0101a-JO to be allele-specific (with regard to the other parental cultivar Prima) and segregating (Fig. 3b). Thirdly, 144 descendants were tested and showed a 1:1 segregation ratio (72:72).

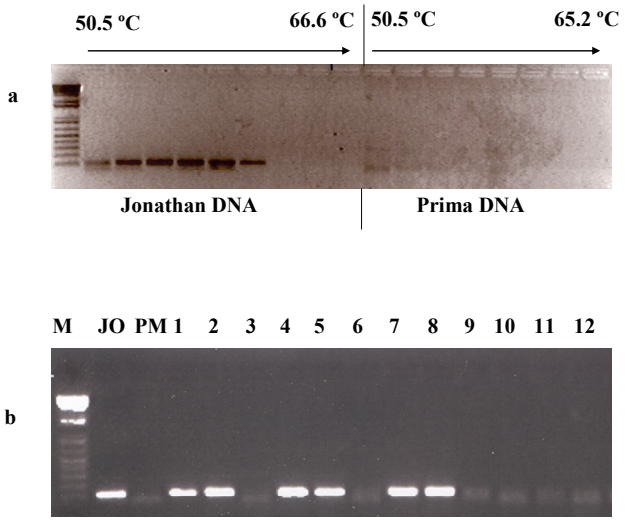

Fig. 3 Test of the allele specific marker Mal d 3.0101a-JO in the population of JO x PM **a** Test on specificity by gradient PCR. Only Jonathan shows the expected 116-nt band in a wide range of Tm (*lanes* 2 to lane 6; 51.4-60.6 ºC). **b** Test for segregation on 12 descendants of the cross JO x PM. *Lanes*: *M* marker (1 kb plus DNA ladder), *JO* Jonathan, *PM* Prima

With respect to the *Mal d 3.02* locus, Prima possesses two sequences that are different from that of Fiesta. Similarly, an allele-specific marker Mal d 3.0201c-PM (Table 3) segregated in a 71:68 ratio in PM x FS and in a 69:75 ratio in the JO x PM population. The new *Mal d 3* marker data allowed the mapping of both genes in the first or second round with Chi-square contributions of 0.53, 0.33 and 0.38 in the PM 4-PM x FS, PM 4-JO x PM and JO 12-JO x PM maps, respectively (Fig. 4).

Allelic diversity

To obtain information on the allelic diversity for both *Mal d 3* genes, we carried out additional cloning (the fourth round) on eight cultivars using gene-specific forward primers (Mald3.01-For, Mald3.02-For) with one common reverse primer (Mald3-Rev2) (Table 1). Their PCR products were expected to be around 901 and 986 nt in length, respectively (Table 1).

Comparison of all sequences of the same length (901–986 nt) from 10 cultivars and subsequent verification tests revealed only three different sequences in the group of *Mal d 3.01* and seven in that of *Mal d 3.02*. These sequences were further classified into variants based on the deduced mature nsLTP proteins: two variants for Mal d 3.01 and three variants

for Mal d 3.02 (Table 2). Figures 5 and 6 show the alignment of the complete coding DNA sequences and deduced amino acid sequences of the identified *Mal d 3.01* and *Mal 3.02* alleles together with reference sequences.

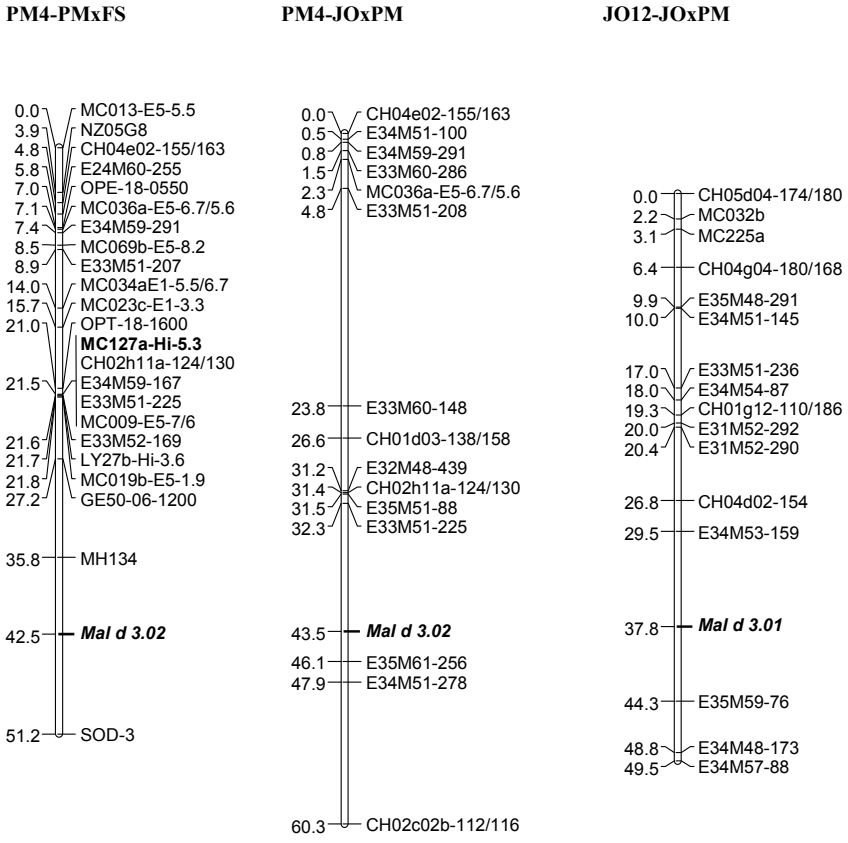

Fig. 4 Map position of *Mal d 3.01* and *Mal d 3.02* on the two homoeologous linkage groups 12 and 4 (JO 12 and PM 4, respectively) based on segregation patterns of the Mal d 3.0101a-JO and Mal d 3.0201c-PM marker

Only two variants were found in *Mal d 3.01*. One of these is *Mal d 3.0101*, which was present in all ten cultivars tested and is fully identical to AF221502 from Golden Delicious. Compared to AJ277164, *Mal d 3.0101* shows one nucleotide difference at position 105, where we found the consensus G instead of A (Fig. 5). However, this change does not lead to amino acid change. Another *Mal d 3.01* allele, *Mal d 3.0102*, is present only in Priscilla. This allele has two SNPs at position +128 and +178 which cause two amino acid changes—from S to N and from G to S, respectively—and an additional 15 SNPs in the upstream sequences of this allele. Through breeding pedigree we were able to trace the special allele (Mal d 3.0102) from cv. Priscilla back to its original source, *Malus floribunda* 821, and forward to a new cv. Santana. The finding of just one prevalent allele for *Mal d 3.01* coincided with little variation in the upstream region of the gene. The *Mal d 3.02* sequences were more diverse than those of

Mal d 3.01. Four different *Mal d 3.02* amino acid sequences of precursor nsLTP were obtained from ten culitvars, but they resulted in three variants of the mature protein after removal of the signal peptides. A single amino acid substitution, 20D/G or 66V/A, was found in two variants. More polymorphisms occurred in the upstream region.

The signal peptides are identical for both variants of *Mal d 3.01,* while those of *Mal d 3.02* differs in one or two amino acids from *Mal d 3.01*. Eight typical conserved cysteine residues (Kader 1996) have been found in all variants of Mal d 3.01 and Mal d 3.02. The Mal d 3.02 allelic constitution of cultivars reflect their pedigree relations: Ingrid Marie inherited Mal d 3.0202a allele from Cox, while Mal d 3.0201a in Fuji and Priscilla came from Red Delicious.

Table 2 Systematic classification of allelic variations of *Mal d 3.01* and *Mal d 3.02* among 10 apple cultivars (**PM** Prima; **FS** Fiesta; **GD** Golden Delicious; **PS** Priscilla; **IM** Ingrid Marie; **JO** Jonathan; **CO** Cox; **RD** Red Delicious; **FJ** Fuji; **DS** Discovery

Allergen	Iso-allergen[a]	Variant[a]	Additional polymorphism[a,b]	PM	FS	GD	IM	PS	JO	CO	RD	FJ	DS
Mal d 3	**01**	01	a	++ [c]	++	++	+	+	++	++	+	++	++
			b				+				+		
		02	a					+					
	02	01	a		++			+	++	+	++	+	
			b	+		++							+
			c	+				+					
			d									+	
			e										+
		02	a				+			+			
		03	a				+						

[a] NCBI GenBank accession numbers:

Mal d 3.0101a: AY572500, AY572501, AY572502, AY572503, AY572504, AY572505, AY572506, AY572507, AY572508, AY572509, AY572510, AY572511, AY572512, AY572513.

Mal d 3.0101b: AY572514, AY572515. **Mal d 3.0102a**: AY572516.

Mal d 3.0201a: AY572517, AY572520, AY572521, AY572522, AY572523, AY572524.

Mal d 3.0201b: AY572518, AY572525, AY572526. **Mal d 3.0201c**: AY572519, AY572527.

Mal d 3.0201d: AY572528. **Mal d 3.0201e**: AY572529.

Mal d 3.0202a: AY572530, AY572531. **Mal d 3.0203a**: AY572532.

[b] Polymorphisms in the region of about 600 nt upstream of the coding DNA sequences for the mature Mal d 3 proteins

[c] +, indicates heterozygous presence of polymorphic allele, + +, indicates homozygous presence of polymorphic allele

Table 3 Description of allele-specific markers for *Mal d 3* genes

Marker name	Primer pairs	Primer sequence [a] (5'-3')	Product (nt)	Annealing temperature ^{0}C [b]	Cycles [c]
Mal d 3.0101a-JO[d]	Mald 3.01Jo-f1	GATTTGGTCCACAAGTTTG**T**A	116	58/56	5/35
	Mald 3.01SSR-r	GTTTCTATTCCGTCTTCCTTC			
Mal d 3.0201c-PM[d]	Mald 3.02-Pm-f1	GTTGGTGCCGACATC**G**ATC**C**	571	63	35
	Mald 3.02-Pm-r1	AGGCCACCACCAAAGCAA**T**G			
Mal d 3.0101a	M3.0101a-f	ATGTTCCGACGAGTTTTC**T**A	259	58/57	5/35
	M3.0101ab-r	GATTACCAAAT**A**TGTGTATGAA**T**G			
Mal d 3.0101b	Mald 3.01-f1	ATGTTCCGACGAGTTTTC**T**G	565	60	35
	Mald 3.01-r1	AGGAAGCCCTGCTGCATTGTTA			
Mal d 3.0102a	Mald 3.01-f2	CCTACATTTTAA**C**ATAGAGTGT**CG**	503	53	35
	Mald 3.01-r2	GAGGGACAGCTCCGC**G**A**T**			
Mal d 3.0201a	Mald 3.02-f1	ACGAAGGTCTTAGCAATC**A****T**	665	53	40
	Mald 3.02-r1	ACTCTTCAGGCAGTTGCAAGC			
Mal d 3.0201b	Mald 3.02-f2	TGCCGACTGTCTAGCCA**G**AA	825	65/63	5/30
	Mald 3.02-r1	See above			
Mal d 3.0201c	Mald 3.01-f3	CAGTTAACGAGGTCAAACA**T**C	516	60	35
	Mald 3.02-r1	See above			
Mal d 3.0201d[e]	Mald 3.02-f4	GAAGAAAGAAGTTAGAAGCC**G**A	316	53/52	5/30
	Mald 3.02-r2	TTTCCAGGAAGCGATT**A**G**G**			
	Mald 3.02-r3	CTTATATTTTAAGGGG**G**C**A****T**			
Mal d 3.0201e	Mald 3.02-f5	TAACGAGGTCAAACAAAC**A****T**	512	60/58	5/30
	Mald 3.02-r1	See above			
Mal d 3.0202a	Mald 3.02-f6	GAAGAAAGAAGTTAGAAGCC**G****T**A	712	63/61	5/35
	Mald 3.02-r4	AGCTGGAGGGACAGGTC**G**G**T**			
Mal d 3.0203a	Mald 3.02-f7	ATTCATCGCCATGATTG**T****T**	458	55	35
	Mald 3.02-r1	See above			

[a] In bold, Allele-specific SNPs; underlined, designed mismatch nucleotide

[b,c] PCR annealing temperatures (Tm) and number of cycles, respectively. 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.

[d] Markers used for mapping.

[e] This marker is obtained by nested PCR: 1 µl of the PCR product of primer pair Mal d 3.02-f4 and –r2 was used as the template in a second-round PCR with primer pair Mal d 3.02-f4 and –r3. The same touch down procedure was followed for both PCR rounds

Development of allele-specific markers

Wherever the SNPs located, they were used to develop single nucleotide amplified ploymorphism (SNAP) markers for tagging the different *Mal d 3* genomic sequences. In all,

18 SNPs from *Mal d 3.01* sequences and 21 SNPs of *Mal d 3.02* were found. On the basis of these SNPs, we developed 10 allele-specific SNAP markers and tested these on the cultivars with the *Mal d 3* sequences information and other cultivars of interest in order to approve their authenticity. Their primer sequences, PCR conditions and product sizes are given in Table 3.

```
                          *       20        *       40        *       60        *
Mal d 3.0101.01a/b: ATGGCTAGCTCTGCAGTGACCAAGCTTGCTTTGGTGGTGGCCTTGTGCATGGCGGTGAGCGTTGCTCATG :  70
Mal d 3.0101.02   : ------------------------------------------------------------------
Mal d 3.0102      : ..................................................................... :  70
Mal d 3.0201a/b   : ..................T..................................................... :  70
Mal d 3.0201c/d/e : ..................T...C................................................ :  70
Mal d 3.0202      : ..................T...C................................................ :  70
Mal d 3.0203      : ..................T...C................................................ :  70

                         80        *      100        *      120        *      140
Mal d 3.0101.01a/b: CCATAACATGTGGCCAAGTGACCAGCAGCCTTGCGCCATGCATTGGCTACGTGAGGAGTGGCGGAGCTGT : 140
Mal d 3.0101.02   : --..............................A..................................... : 140
Mal d 3.0102      : ..................................................A.................... : 140
Mal d 3.0201a/b   : .............G....G.TC..A.....TA......T...A...................C.... : 140
Mal d 3.0201c/d/e : .............G....G.TC..A.....TA......T...A...................C.... : 140
Mal d 3.0202      : .............G....G.TC..A.....TA......T...A..............A....C.... : 140
Mal d 3.0203      : .............G....G.TC..A.....TA......T...A...................C.... : 140

                          *      160        *      180        *      200        *
Mal d 3.0101.01a/b: CCCTCCAGCTTGCTGCAATGGAATCAGAACCATTAACGGCTTGGCCAGGACCACCGCTGACCGCCAGACT : 210
Mal d 3.0101.02   : ..................................................................... : 210
Mal d 3.0102      : ....................................A............................... : 210
Mal d 3.0201a/b   : ..............................................A.......C...........G.. : 210
Mal d 3.0201c/d/e : ..............................................A.......C...........G.. : 210
Mal d 3.0202      : ..............................................A.......C...........G.. : 210
Mal d 3.0203      : ..............................................A.......C...........G.. : 210

                        220        *      240        *      260        *      280
Mal d 3.0101.01a/b: GCTTGCAACTGCCTGAAGAATCTTGCCGGCAGCATCAGTGGTGTTAACCCTAACAATGCAGCAGGGCTTC : 280
Mal d 3.0101.02   : ..................................................................... : 280
Mal d 3.0102      : ..................................................................... : 280
Mal d 3.0201a/b   : .................G.............G.................GG......C.A.TC..... : 280
Mal d 3.0201c/d/e : .................G.............G.................GG......C.A.TC..... : 280
Mal d 3.0202      : .................G.............G.................GG......C.A.TC..... : 280
Mal d 3.0203      : .................G.............G.................GG.....TC.A.TC..... : 280

                          *      300        *      320        *      340
Mal d 3.0101.01a/b: CTGGAAAGTGTGGAGTCAACGTCCCCTACAAGATCAGCACCTCCACCAACTGCGCCACCGTGAAGTAA : 348
Mal d 3.0101.02   : ..................................................................... : 348
Mal d 3.0102      : ..................................................................... : 348
Mal d 3.0201a/b   : ..................................................................... : 348
Mal d 3.0201c/d/e : ..................................................................... : 348
Mal d 3.0202      : ..................................................................... : 348
Mal d 3.0203      : ..................................................................... : 348
```

Fig. 5 Alignment of coding sequences of *Mal d 3.01* and *Mal d 3.02* among the ten cultivars together with two Mal d 3 reference sequences. Reference Mal d 3 sequence AF221502 is identical to Mal d 3.0101.01, AJ277164 has been proposed to be named as Mal d 3.0101.02

```
                                        1       *       20        *    36
Mal d 3.0101 : MASSAVTKLALVVALCMAVSVAHA ITCGQVTSSLAPCIGYVRSGGAVPPACCNGIRTING
Mal d 3.0102 : ........................ ..................N.............S
Pru p 3      : ------------------------ ......S.......P...G.............NV.N

Mal d 3.0201a: ......I.................. ......S.N.V..FD......P..............
Mal d 3.0201b: ......IN................. ......S.N.V..FD......P..............
Mal d 3.0202 : ......IN................. ......S.N.V..FD....D.P..............
Mal d 3.0203 : ......IN................. ......S.N.V..FD......P..............
Pyrus c 3    : ......I.................. ...S..SAN.....N.....................


                40        *        60        *        80        *
Mal d 3.0101 : LARRTTADRQTACNCLKNLAGSISGVNPNNAAGLPGKCGVNVPYKISTSTNCATVK : 91
Mal d 3.0102 : ........................................................ : 91
Pru p 3      : .....P...A......G.SA.VP........A.......SI.....A.....E.. : 91

Mal d 3.0201a: ..K..P...A......S....V.....G..ES....................... : 91
Mal d 3.0201b: ..K..P...A......S....V.....G..ES....................... : 91
Mal d 3.0202 : ..K..P...A......S....V.....G..ES....................... : 91
Mal d 3.0203 : ..K..P...A......S....V.....G.VES....................... : 91
Pyr c 3      : ..K..P...A...........V.....G..E........................ : 91
```

Fig. 6 Alignment of deduced amino acid sequences of the variants of Mal d 3.01 and Mal d 3.02 observed in the ten apple cultivars together with reference sequences from apple, pear and peach. The *first* 24 *underlined* amino acids are signal peptide, amino acids in *bold* are the same as the predicted antibody recognition sites of Pru p 3 (AJ277163). Amino acids (in bold and underlined) indicate amino acid substitutions. *Pyr c 3* (AF221503)

Discussion

Mal d 3 is one of the four allergens in apple identified to date. We report here, for the time, linkage mapping of apple allergen genes. By means of genomic PCR cloning and subsequent genome walking, we characterised genomic sequences of two Mal d 3 (nsLTP) isoallergen genes, which we denoted as *Mal d 3.01* and *Mal d 3.02*, mapped these two genes on linkage group 12 and 4, respectively. Comparison of the deduced amino acid sequences from ten apple cultivars showed that *Mal d 3.01* is highly conserved, since only two variants were observed, while *Mal d 3.02* is more variable, with three variants.

The amphidiploid nature of the apple genome

It is proposed that the cultivated apple is an amphidiploid ($\chi = 17$) evolved from a hybrid of two closely related ancestors (Chevreau and Laurens, 1987). The large number of duplicated restriction fragment length polymorphism (RFLP) markers have been found in the PM x FS linkage map (Maliepaard et al. 1998) support this hypothesis. In our updated map, the lower

part of linkage group 4 and 12 share two common RFLP markers, MC127 and MC105 (Van de Weg, unpublished). Alignment of the updated linkage groups 4 and 12 derived from Prima, Fiesta and Jonathan showed that *Mal d 3.01* and *Mal d 3.02* are located just between these two RFLP markers. The fact that these two genes are located on homoeologous chromosomes is clear evidence that *Mal d 3.01* and *Mal d 3.02* have originated from different ancestral genomes. Because the genome constitution of apple and pear is highly similar (Yamamoto et al 2002, 2004; Hemmat et al. 2003), we expect to locate: (1) the known pear nsLTP gene (Pyr c 3, accession AF221503), which is 93% identical to Mal d 3.02 (Fig. 6), on linkage group 4; and (2) another yet unidentified pear nsLTP gene on linkage group 12.

Multigene family of nsLTPs

Plant nsLTPs (PR-14) are encoded by a small gene family which is present throughout the plant kingdom (Mills et al. 2003). nsLTPs are generally secreted (Sterk et al. 1991) and externally associated with the cell wall (Clark and Bohnert 1999). Whlie their precise biological function remains uncertain, it has been postulated that they are involved in the formation of protective hydrophobic cutin and/or suberin layer and play a role in the defence of plants against microbial pathogens (Mills et al. 2003). One subfamily, nsLTP1, is characterised by a molecular weight of 9 kDa, a strictly conserved 8 cysteine motif and resistance to pepsin hydrolysis and thermal denaturation. Most nsLTP allergens identified to date belong to the nsLTP1 subfamily (Pastorello et al. 2000a/b, Chardin et al. 2003), including both nsLTPs in apple described here.

Each nsLTP member may have a different expression profile. In peach, Botton et al. (2002) identified two LTP genes in reproductive organs that had different expression levels, while Jung et al. (2003) identified three pathogen-inducible LTPs from pepper that were differentially activated by pathogens, abiotic and environmental stress (Jung et al. 2003). It is quite clear that in apple *Mal d 3.01* is being definitely expressed in the fruit since both cDNA (Diaz-Perales et al. 2002) and its protein (Pastorello et al. 1999a) have been identified. However, expression of *Mal d 3.02* in the fruit has not yet been shown. Typical TATA and CAAT boxes in the promoter are present in the proximity of the start codon at different positions for both *Mal d 3* genes (Figs. 1,2). There is also large sequence variation between the upstream region of *Mal d 3.01* and *Mal d 3.02* that points towards a potential individual expression profile. Furthermore, one or two amino acid differences in the signal peptide for *Mal d 3.01* and *Mal d 3.02* (Fig. 6) may lead to different trafficking to the target compartment of these two genes (quality of transport, in terms of time and possibly also efficacy in cleaving off the signal peptide but remaining the function of the mature protein).

Potential cross-reactivity with peach nsLTP (Pru p 3)

Peach fruit is one of the most frequent causes of food allergy in Mediterranean populations, and is considered as the primary sensitizing agent in this type of food allergy (Asero et al. 2000). Garcia-Casado et al. (2003) used the peach allergen, Pru p 3, as a model to find IgE epitopes. They identified three amino acid sequence regions (11–25, 31– 45, and 71–80) as major IgE epitopes and predicted 5 amino acid residues of R39, T40, R44, K80 and K91 as antibody recognition sites using IgE immuno-detection and synthetic peptides. Mal d 3.01 appears to be more similar to Pru p 3 than does Mal d 3.02 with regard to amino-acid sequence of the total protein as well as to the supposed epitope regions and recognition sites (Fig. 6). Compared with Pru p 3, the Mal d 3.01 and Mal d 3.02 protein sequences share 80% and 77% identity, respectively. Considering the recognition sites, Mal d 3.01 and Pru p 3 are fully identical, while Mal d 3.02 differs at a single position (R/K39). This may imply that Mal d 3.01 is more relevant to apple LTP allergenicity. Structure modelling and immunological testing of the recombinant proteins of all identified apple Mal d 3 isoallergens and variants will provide further information about their individual allergenicity.

Allelic diversity and association with phenotype

Neither of Mal d 3 genes was observed to be very diverse in the ten cultivars investigated; in particular, the *Mal d 3.01*, which we assume to be more important of the two considering its expression in fruit and potential allergenicity, showed very little diversity. Therefore, it is necessary to obtain more detailed information on allelic diversity for more cultivars and selections. Consequently, additional allele-specific markers can be developed and used in screening *Mal d 3* genotypes. Once the *Mal d 3* allelic constitution and the Mal d 3 related allergenicity for a wider range of apple cultivars have been assessed, it is possible to associate individual genes and alleles to low Mal d 3 allergenicity. On that basis, allelic-specific markers can be applied in breeding for low nsLTP allergenic cultivars.

Nomenclature of allergen genes

The occurrence of different loci and multi-allelism has its consequences for the nomenclature of these allergen genes. We followed the current allergen nomenclature (King et al. 1995) when naming the genomic sequences of *Mal d 3* as much as possible. Firstly, the current allergen nomenclature provides some guidelines to discriminate isoallergens and variants on the basis of amino acid and DNA sequence similarity. The distinction between isoallergens and variants is arbitrary since there is no a single threshold of amino acid identity can be applied in all situations. In the case of this investigation, two similar *Mal d 3* genes share 85%

identity at the amino acid level and have been mapped on two linkage groups, so we denoted them as two isoallergen genes: *Mal d 3.01* and *Mal d 3.02*. The Allergen Nomenclature Sub-Committee has accepted these names. Secondly, the current nomenclature recognise that each isoallergen may have multiple forms (designated as variants by two numerals) of cDNA sequences with nucleotide mutations which are either silent or can lead to single or multiple amino acid substitutions. The silent mutations will be designated by new variant numerals. This means that different variant names may represent the same protein, which may cause confusion about just how many different proteins are actually present when more allelic sequences from different sources become available. To avoid this confusion, we distinguished variants solely on the basis of amino acid sequence, and indicated silent mutations with two additional numerals preceded by a dot. For example, Accessions AF221502 and AJ277164, which have a silent mutation at position 105 (Fig. 5), will be named Mal d 3.0101.01 and Mal d 3.0101.02, respectively. This is a recommendation for updating the current allergen nomenclature. Thirdly, any polymorphism in an intron (if present) or flanking region is not covered in the current nomenclature. However, recognition of their existence by different names is necessary in some cases. Therefore, we extended the allele names by adding a letter in low case. For example, Mal d 3.0101a and Mal d 3.0101b refer to gDNA sequences for which the polymorphisms are located beyond the coding region of variant Mal d 3.0101.

Mapping genes of interest, genetic analysis and breeding aspects

With the advancement of sequencing facilities and the availability of complete genome sequence data of several model plant species, it can be expected that many gene sequences will be determined in the very near future for other plants too. With respect to genetics and breeding, knowledge of the location of these genes on the chromosome will be of great help in analysis of allelic diversity related to phenotypes. As illustrated in this paper, two almost identical *Mal d 3* cDNA sequences in the GenBank were explored to generate genomic sequences and to assess linkage map positions and allelic diversity of two *Mal d 3* genes. In addition to Mal d 3, three more apple allergens have been identified by gene cloning and sequencing: they are Mal d 1 (Vanek- Krebitz et al. 1995), Mal d 2 (Krebitz et al. 2003) and Mal d 4 (Van Ree et al. 1995). Similarly, the number of loci and their location on the linkage groups will also be determined in the same way.

For a long time fruit genetics and breeding aimed at parameters such as high yield, resistance and quality. At the present time there is a growing awareness about the paradox that healthy fruits can cause allergic reactions in some sensitive individuals. Based on knowledge

acquired from genetics and genomic research, future fruit breeding may be directed towards a new quality trait: reduction of allergenicity.

Acknowledgements

This research was funded in part by the EU (EU-SAFE, QLK1-CT-2000-01394). We took the advantage of using the updated molecular maker linkage map of Prima and Fiesta from the European DARE project (FAIR5-CT97-3898). The authors thank professors Evert Jacobsen and Richard Visser for their critical reading of this manuscript and their stimulating interest. The first author is grateful to ICSC-World Laboratory, Lausanne for providing a one-year scholarship in 2001.

Chapter 3

Genomic cloning and linkage mapping of the *Mal d 1* (*PR-10*) gene family in apple (*Malus domestica*)

Z. S. Gao[1,2,3], W. E. van de Weg [1,3], J. G. Schaart[3], H. J. Schouten[3], D. H. Tran[3,5], L. P. Kodde[3], I. M. van der Meer[3], A. H. M. van der Geest[3], J. Kodde[3], H. Breiteneder[4], K. Hoffmann-Sommergruber[4], D. Bosch[3], L. J. W. J. Gilissen[1,3]

1: Allergy Consortium Wageningen, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

2: Experimental Nursery of Shanxi Forestry Department, Beiying Nan Lu 52, Taiyuan, Shanxi, 030031. People's Republic of China.

3: Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands.

4: Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, A-1090 Vienna, Austria

5 Present address: Department of Biology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

This chapter has been accepted for publication in *Theoretical and Applied Genetics*

Abstract

Fresh apples can cause birch pollen-related food allergy in northern and central European populations primarily due to the presence of Mal d 1, the major apple allergen that is cross-reactive to the homologous and sensitizing allergen Bet v 1 from birch. Apple cultivars differ significantly in their allergenicity. Knowledge of the genetic basis of these differences would direct breeding for hypo-allergenic cultivars. PCR genomic cloning and sequencing were performed on two cultivars, Prima and Fiesta, which resulted in 37 different Mal d 1 gDNA sequences. Based on the mapping of sequence-specific molecular markers, these sequences appeared to represent 18 *Mal d 1* genes. Sixteen genes were located in two clusters, one cluster with seven genes on linkage group (LG) 13, and the other cluster with nine genes on the homoeologous LG16. One gene was mapped on LG 6, and one remained unmapped. According to sequence identity, these 18 genes could be subdivided into four subfamilies. Subfamilies I–III had an intron of different size that was subfamily and gene specific. Subfamily IV consisted of 11 intron-less genes. The deduced amino acid sequence identity varied from 65% to 81% among subfamilies, from 82% to 100% among genes within a subfamily, and from 97.5% to100% among alleles of one gene. This study provides a better understanding of the genetics of *Mal d 1* and the basis for further research on the occurrence of allelic diversity among cultivars in relation to allergenicity and their biological functions.

Introduction

Fresh apples can cause birch pollen-related food allergy in northern and central European populations. About 50–75% of birch pollen-sensitized patients suffer from an oral allergy syndrome (OAS) after eating apples (Ebner et al. 1991; Pauli et al. 1996). This allergy results from IgE-mediated cross-reactivity between Bet v 1, the major birch allergen, and Mal d 1, the major apple allergen. Both Bet v 1 and Mal d 1 belong to a group of pathogenesis-related (PR)10 proteins (Breiteneder et al. 2000). *PR-10* genes appear abundantly in plants (Wen et al 1997; Hoffmann-Sommergruber and Radauer 2004; Liu and Ekramoddoullah 2004). Expression of these genes is induced by stress, such as attack by plant pathogens, and occurs during ripening (Atkinson et al. 1996; Puehringer et al. 2000).

Mal d 1 has been identified as a 17–18 kDa protein of 158–159 amino acids encoded by 480–483 nucleotides (Vieths et al. 1995; Schoning et al. 1996; Hoffmann-Sommergruber et al. 1997). Over 30 *Mal d 1* DNA sequences obtained from cDNA and gDNA of leaves and fruits of various apple cultivars (Vanek-Krebitz et al. 1995; Hoffmann-Sommergruber et al. 1997; Son et al. 1999; Holm et al. 2001; Ziadi et al. 2001) were published in the database GenBank at the beginning of this study. Their phylogenetic tree indicated four distinct groups with seven members (Puehringer et al. 2003). Recently, Beuning et al. (2004) sequenced cDNA libraries of cultivar Royal Gala and identified 12 Mal d 1-related genes, five of which were new. The *Mal d 1* family seemed to consist of at least 15 members (Atkinson et al. 1996), which means that at least three *Mal d 1* homologous genes have not been cloned and sequenced. To date, neither is there a conclusive answer to the number of members in the *Mal d 1* gene family in the apple genome, nor about their positions on the linkage map.

Apple cultivars differ considerably in allergenicity, e.g., Golden Delicious is highly allergenic as experienced by many apple-allergy patients, whereas Gloster causes only mild reactions (Vieths et al. 1994; Hsieh et al. 1995). It is likely that genetic factors are involved. For breeding hypo-allergenic apple cultivars, it is essential to understand these differences at the qualitative and quantitative level of Mal d 1 isoallergens and their variants. As a first step, we examined the number of *Mal d 1* genes present in the genome and their location on molecular marker linkage maps of apple. For this purpose, we first cloned and sequenced *Mal d 1* genes from genomic DNA of two parental cultivars, Prima and Fiesta. Then we developed

sequence-specific markers for mapping. This strategy enabled us to characterize genomic sequences and to determine 18 *Mal d 1* genes and their map positions.

Materials and Methods

Plant materials

The PCR cloning and sequencing was conducted on two cultivars, Prima (PM) and Fiesta (FS). These two cultivars were chosen because they are used as parent in three available mapping populations: Prima x Fiesta (Maliepaard et al. 1998), Jonathan x Prima (JO x PM) and Fiesta x Discovery (FS x DS). With regard to the allergenicity of four parental cultivars, skin prick tests on Dutch apple allergenic patients revealed Fiesta and Discovery to be high and Prima and Jonathan to be intermediate allergenic (Van de Weg et al., unpublished data).

PCR primers

Primer pairs were designed using the software programme PRIMER DESIGNER ver. 2.0 (Scientific and Educational Software, Cary, N.C.) based on all available sequence information of *Mal d 1* genes. Preferably, the forward and reverse primers covered the whole length of the gene coding sequences, but primers for getting the middle part of a gene were also tried if the first primers failed to amplify the expected product or sequences. Both conserved and specific primers were used to obtain all possible *Mal d 1* sequences. Some of the cloning primers were adjusted after the first round of cloning and sequencing. The *Mal d 1* cloning primers were firstly derived from the sequences in the GenBank. Then new cloning primers were added when new *Mal d 1*-like sequences were obtained from unpublished research on gene regulation in apple stock root induced by wounding and auxin at Plant Research International, Wageningen. Primers for genome walking were designed according to new sequences obtained in this research. (All these primer sequences are listed in Table 1).

DNA isolation, PCR cloning and sequencing

Genomic DNA was extracted using the CTAB-based large-scale nuclei-isolation method (Roche et al. 1997). PCR cloning and sequencing procedures have been described previously (Gao et al. 2005). Here we mention only the key points and some changes. PCR was performed in two steps with *Pfu* polymerase (Stratagene, La Jolla, Calif.) and super*Taq* (HT

Table 1 PCR primer pairs used for cloning of the different *Mal d 1* genes

Primer pairs[a]	Reference sequence[a]	Primer name[b]	Primer sequence (5′– 3′)	Tm °C/cycles[c]		Product	Intron	Isoallergen gene member
				Pfu	*Taq*	(nt)	(nt)	
1	AY026910	M1a-For	ATCTCCAACACAATACTCTCAAC	58/25	60/2	800	168	Mal d 1.01
		M1a/b/c-Rev	AAAGCCACACAACCTTCGAC					
2	L42952 AF020542	Mald1b-For	CATCCTTGGTAGTTGCTTTC	52/30	54/2	723	171	Mal d 1.02
		M1b-Rev	ACCATAGAAACATATTAATTTAGT					
3	Z48969	M1a/b/c-For	TCATGGGTGTCTACACATTTG	50/35	50/2	439; 483	0	Mal d 1.03B; Mal d 1ps1
		M1c-Rev	ATTAATTTAGTTGTAGGCATCC					
4	Z72425–Z72428	HB20-For[b]	ATGGGTGTTTTCAATTACGAAACTGAG	54/30	56/2	480	0	Mal d 1.03A; Mal d 1.08; Mal d 1.09
		HB22-Rev[b]	TTAGTTGTAGGCATCGGAGTGTGCCAAGAGG					
5	AY026908 AY026909	M1.03-For1	ATGGGTGTTTTCACATACGAA	55/25	57/2	486	0	Mal d 1.07; Mal d 1.08; Mal d 1.09; Mal d 1.03C,D, E, G
		M1.03-Rev1	AAGAATTTAGTTGTAGGCATC					
6	Z72426–Z72428	M1.04-s27	GWYCACCTCTGTCATCCCTG	56/30	58/2	470	111	Mal d 1.04
		M1.04-as496	TCTCAACATCACCCTTAGCA					
7	Z72426-Z72428	M1.04-s62	AGGCYTTTATCCTTGATGGC	56/30	58/2	492; 500	111; 119	Mal d 1.04; Mal d 1.05
		M1.04-as553	GAAGCTTGAAGAGACCGGAA					
8	Based on 6	M1.04UP-GSP1	CTCCTTCAATCATAGAGTAACTGTATGT			747		Mal d 1.04
		M1.04UP-GSP2	CCAAAGGTAACTTTCTTGATGGTTCC					
9	Based on 6	M1.04DW-GSP1	ACATACAGTTACTCTATGATTGAAGGA			830		Mal d 1.04
		M1.04DW-GSP2	CATCAAAACCACTAGCCACTACCATGC					
10	Based on 7	M1.04UP-GSP1	See above			747		Mal d 1.05
		M1.05UP-GSP2	GAGAGAATATTGAGGGTTGTCTAGCTAA					
11	Based on 7	M1.04DW-GSP1	See above			477		Mal d 1.05
		M1.05DW-GSP2	ATTGACCGCATCTCCTGATGG					
12	3262-PRI 3425-PRI	M1.06-For	CATGGGTGTCCTCACATACGAAAC	55/25	57/2	609–634	128–142	Mal d 1.06A–C
		M1.06-Rev	TTAGTTGTAGGCATCAGGATTG					
13	Based on 12	M1.07-fw	ATGGGTGTCCTCACATACGAAACT	62/30	64/2	608	128	Mal d 1.06C
		M1.07-rev	TTAGTTGTAGGCATCAGGATTGGCCACAAGGTG					
14	3076-PRI	M1.03-For2	ATCATGGGTGTTTTCACATAT	55/25	57/2	491	0	Mal d 1.03A; Mal d 1.03B; Mal d 1.03F
		M1.03-Rev2	CATATAATTTAGTTGTAGGCGTCT					
15	Based on 14	M1.03-For2	See above	55/25	57/2	491	0	Mal d 1.03A; Mal d 1.03B; Mal d 1.03F
		M1.03-Rev3	CATATAATTTAGTTGTAGGCGT					

[a] primer pairs 1–7 were based on *Mal d 1* sequences in the GenBank; 8-11 are gene specific for genome walking to upstream (UP) and downstream (DW); 12–15 were derived from unpublished *Mal d 1* sequences at Plant Research International (PRI), Wageningen.

[b] HB20-for and HB22-Rev are original primers of Hoffmann-Sommergruber et al. 1997;

[c] Annealing temperature (Tm, ^{0}C)/number of cycles

Biotechnology, England) using a PTC-200 machine (MJ Research, Waltham, Mass.). The amplified fragments were purified by Qiaquick Gel Extraction Kit (Qiagen, Germany) and ligated into the pGEM-T easy vector (Promega, Madison, Wis.) and used to transform XL1

Blue competent cells (Stratagene) according to the protocols of the manufacturer. For each fragment, 8–10 white colonies were selected for the isolation of plasmid DNA using the Qiaprep®Turbo BioRot®Kit (Qiagen, Germany) by a Bio Robot 9600 (Qiagen, USA). DNA sequencing was performed on a 96-capillary system (ABI 3700; Applied Biosystems, Foster City, Calif.). If one primer pair produced more than two different sequences, then additional clones were sequenced or PCR cloning was performed again in order to obtain enough replicated sequences.

Genome walking approach

Genome walking was applied to gDNA of cvs. PM and FS using the Universal Genome Walker kit (Clontech, Palo Alto Calif.) to get precise sequences at the two ends of a gene and its flanking region. For each cultivar, four libraries were constructed using *Dra*I, *EcoR*V, *Pvu*II and *Stu*I enzymes to digest 2.5 µg of the gDNA. Adaptors were ligated to the digested DNA fragments. Four groups of gene-specific primers (Table 1) were designed and used for nested PCR together with two adapter primers (AP1 and AP2). The product obtained from one of the four libraries was excised from the gel and subsequently purified, ligated, transformed and sequenced as already described.

Sequence analysis

DNA sequences and single nucleotide polymorphisms (SNPs) were analyzed using the SEQMAN programme (DNAstar, Madison, Wis.). Intron sizes were deduced by comparing the genome sequences with known cDNA sequences or by putative splicing patterns. The phylogenetic tree was created and sequence identity percentages were calculated using the Clustal W by Megalign programme (DNAstar). Multiple DNA and amino acid sequence alignments were performed with the GenDoc programme (www.psc.edu/biomed /genedoc).

Designing and testing of sequence-specific markers

Two types of molecular markers were used to distinguish a specific sequence or allele in the context of the PM x FS or the JO x PM population: single nucleotide amplification polymorphism (SNAP) markers (Drenkard et al. 2000; Gao et al. 2005) and simple sequence repeat (SSR) markers. The SNAP markers were tested first on gDNA of Prima, Fiesta and eight individuals of their population to confirm PCR conditions, expected product size and

segregation pattern. Then the well-working markers were tested on the entire population. Some markers were also applied to JO x PM to map several Mal d 1 genes that did not segregate in PM x FS or to confirm the results of PM x FS.

SSR primers were designed for regions flanking the repetitive stretch. The reverse primers had so called pig-tails, i.e. GTTT at the 5'end according to Brownstein et al. (1996). Primer labelling, PCR amplification and gel electrophoresis were performed as described previously (Gianfranceschi et al. 1998).

Mapping genes on molecular linkage groups

Two molecular marker linkage maps, PM x FS (population size *n*=144) and JO x PM (population size *n*=196), were used to map sequence specific molecular markers. . Grouping and mapping were performed with JoinMap 3.0 (Van Ooijen et al. 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).

Nomenclature of the *Mal d 1* sequences

After consulting the Allergen Nomenclature Committee, we differentiated isoallergen genes when their protein sequences displayed less than 95% identity and denoted these genes according to current allergen nomenclature (King et al. 1995). Genes with more than 95% DNA sequence identity which are also clustered on the same linkage group were denoted by adding a capital letter to their isoallergen name, such as Mal d 1.03A, Mal d 1.03B, etc. Previous denotations of Mal d 1 isoallergens and variants (Mal d 1.01 to Mal d 1.04, www.allergen.org//isoall) were maintained as much as possible. Finally, we denoted silent mutations according to Gao et al. (2005).

Results

Generating and grouping *Mal d 1* sequences

Fifteen primer pairs (including four for genome walking) were used for PCR amplification using gDNA of two cultivars, PM and FS (Table 1), which resulted in about 300 raw *Mal d 1* sequences. These sequences could be aligned into 43 different sequences by the

Table 2 Classification of all *Mal d 1* sequences according to the phylogenetic tree and map positions

Subfamily	Gene	Previous names[a]	New sequences (alleles)		Reference sequences [b] (isoallergen and variants names)
			Prima	Fiesta	
I	Mal d 1.01	Mal d 1a/c/d; PR-10c, Mal d 1.01	AY789236 (Mal d 1.0105.01) AY789237 (Mal d 1.0105.03)	AY789238 (Mal d 1.0105.02)	X83672 (Mal d 1.0101), Z48969 (Mal d 1.0102), AF124823 (Mal d 1.0103), AF124829 (Mal d 1.0104), AF124830 (Mal d 1.0105), AF124831 (Mal d 1.0106), AF124832 (Mal d 1.0107), AF126402 (Mal d 1.0108), AY026910 (Mal d 1.0109), AJ417551, **AY428579**
	Mal d 1.02	Mal d 1b ; PR-10a; Mal d 1.02	AY789239 (Mal d 1.0201.09) AY789240 (Mal d 1.0201.01)	AY789241(Mal d 1.0201.01)	L42952 (Mal d 1.0201), AF020542 (Mal d 1.0206), AF074721, AF124822 (Mal d 1.0202), AF124824 (Mal d 1.0203), AF125825 (Mal d 1.0204), AF124826, AF124827, AF124828, AF124833, AF124834, AF124835 (Mal d 1.0205), AF124836, AY026911 (Mal d 1.0207), AJ488060 (Mal d 1.0208), **AY428578**
II	Mal d 1.04	Mal d 1.04	AY789242 (Mal d 1.0404.01) AY789243 (Mal d 1.0404.02a)	AY789244 (Mal d 1.0404.02b)	Z72426 (Mal d 1.0401), Z72427 (Mal d 1.0402), Z72428 (Mal d 1.0403)
	Mal d 1.05		AY789245 (Mal d 1.0501)	AY789246 (Mal d 1.0501) AY789247 (Mal d 1.0502)	
III	Mal d 1.06A	Mal d 1e; Mal d 1f	AY789248 (Mal d 1.06A02.01) AY789250 (Mal d 1.06A01.01b)	AY789249 (Mal d 1.06A01.01a)	**AY428580, AY428581**
	Mal d 1.06B	Mal d 1g; Mal d 1h	AY789252 (Mal d 1.06B02) AY789253 (Mal d 1.06B03)	AY789251 (Mal d 1.06B01)	**AY428582, AY428583**
	Mal d 1.06C		AY789255 (Mal d 1.06C02) AY789256 (Mal d 1.06C03)	AY789254 (Mal d 1.06C01)	

Table 2 continued

Subfamily	Gene	Previous names[a]	New sequences (alleles)		Reference sequences [b] (isoallergen and variants names)
			Prima	Fiesta	
IV	Mal d 1.07	PR-10d; Mal d 1k; Mal d 1.03	AY789258 (Mal d 1.0702)	AY789257 (Mal d 1.0701)	AY026909(Mal d 1.0303), **AY428586**
	Mal d 1.08		AY789259 (Mal d 1.0801.01) AY789261 (Mal d 1.0801.02)	AY789260 (Mal d 1.0801.01)	
	Mal d 1.09		AY789262 (Mal d 1.0901)		
	Mal d 1.03A	Mal d 1.03		AY789263 (Mal d 1.03A01)	Z72425 (Mal d 1.0301)
	Mal d 1.03B	Mal d 1.03	AY789264 (Mal d 1.03B01)	AY789265 (Mal d 1.03B02)	AY186248 (Mal d 1.0304)
	Mal d 1.03C	Mal d 1 l	AY789266 Mal d 1.03C01)		**AY428587**
	Mal d 1.03D	PR-10b, Mal d 1.03	AY789267 (Mal d 1.03D02)	AY789268 (Mal d 1.03D01)	AY026908 (Mal d 1.0302)
	Mal d 1.03E	Mal d 1i	AY789270 (Mal d 1.03E01)	AY789269 (Mal d 1.03E02)	**AY428584**
	Mal d 1.03F	Mal d 1j	AY789271 (Mal d 1.03F01) AY789272 (Mal d 1.03F01)	AY789273 (Mal d 1.03F02)	**AY428585**
	Mal d 1.03G			AY789274 (Mal d 1.03G01)	
	Mal d 1ps1		AY789275		

[a] In the reference papers and Mal d 1 isoallergen list

[b] Accession numbers in bold type were released in the GenBank shortly before submission of this manuscript. Isoallergen and variant names in bracket just behind a sequence were assigned by the Allergen Nomenclature Committee (see http://www.allergen.org)

SEQMAN programme. Most of our sequences have been identified from at least two clones. Those derived from a single clone had crosschecks to the GenBank or were confirmed by SNAP marker tests. Six sequences with PCR errors were excluded. Ultimately 37 correct sequences could be deduced. One out of these 37 sequences represented a pseudo gene with a deletion of 44 nt in the middle. Of the 36 protein coding sequences, 23 were present in Prima and 16 in Fiesta, three sequences were common to both cultivars (GenBank accessions AY789236–AY789275, Table 2). Apparently, Prima is more heterozygous for *Mal d 1* than Fiesta, and is therefore more suitable to create markers for mapping. By comparison with the known reference cDNA sequences in the GenBank, the coding sequences and intron sizes (if present) could be deduced for the newly obtained *Mal d 1* genomic sequences. According to a phylogenetic tree of the 39 Mal d 1 protein-coding sequences, we classified the *Mal d 1* gene family into four subfamilies which coincided with variation in intron size (Fig. 1). Subfamilies I–III contained members with a single intron, and the intron sizes are specific for each subfamily and gene, while subfamily IV included only intronless gene members.

By Clustal W alignment, all *Mal d 1* coding sequences in the database could be classified to these four subfamilies and members (Table 2). Subfamily I includes reference sequences that were formerly classified as Mal d 1a, Mal d 1b, Mal d 1c (Son et al. 1999) and Mal d 1 d (Beuning et al. 2004) isoforms or as PR-10a and PR-10c proteins (Ziadi et al. 2001). The new sequences of this subfamily were derived from the PCR primer pairs 1 and 2 (Table 1). Subfamily II includes representatives of the previously denoted isoallergen group Mal d 1.04 (Puehringer et al. 2003). The new sequences were derived from primer pairs 6 and 7 and subsequent genome walking in upstream and downstream directions (primer pairs 8–11) (Table 1). Subfamily III contains sequences similar to Mal d 1e, -f, -g, and -h (AY42580–AY42583, Beuning et al. 2004). Our sequences were obtained from primer pairs 12 and 13 based on two unpublished cDNA sequences (Tables 1 and 2). Subfamily IV includes four reference sequences that were formerly classified as Mal d 1.03 (Puehringer et al. 2003) and four recently released sequences Mal d 1i, -g, -k, -l (AY428584–AY428587, Beuning et al. 2004). The new sequences were derived from primer pairs 3–5, 14 and 15 (Table 1), which not only represent all the reference sequences, but also include some new members.

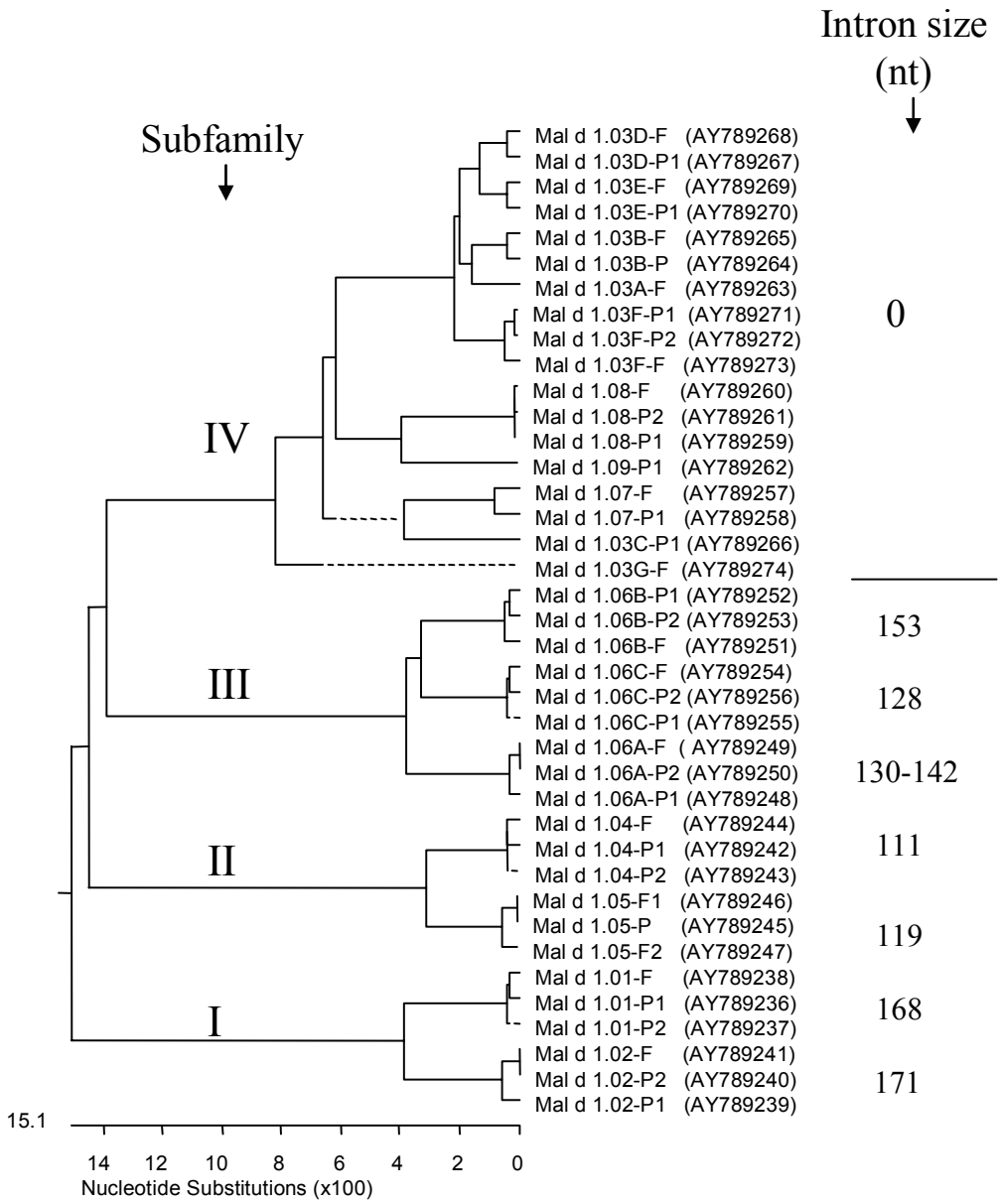

Fig. 1 Phylogenetic tree of 39 coding DNA sequences of the *Mal d 1* gene family of the cultivars Prima and Fiesta and their intron size. P and F behind the gene name indicate the source cultivar, *P* Prima, *F* Fiesta; '1' and '2' behind the cultivar symbols P and F refer to different alleles of the same gene. Mal d 1.02-P2 sequence is the same as Mal d 1.02-F, Mal d 1.05-P=Mal d 1.05-F1; Mal d 1.06A-P2=Mal d 1.06A-F.

Sequence-specific markers and mapping

To distinguish all 36 putatively functional sequences from PM and FS and a pseudogene sequence from Prima, we created 34 single nucleotide amplified polymorphism (SNAP) markers and three simple-sequence- repeat (SSR) markers (Table 3). In addition, one SNAP

marker (Mal d 1.03C02-GD) was developed according to a sequence from cv Golden Delicious (GenBank, AY822725). These markers were tested on the parental cultivars and on the mapping populations. As an example, Fig. 2 illustrates the specificity and segregation of the SNAP marker for AY78936 (Mal d 1.0105.01a). Most markers from PM were mapped in the population PM x FS, seven markers were homozygous in FS but heterozygous in JO, which enabled their mapping in JO x PM (Table 3). The segregating markers allowed the mapping of 17 loci on three linkage groups (LG): LG 6, LG13, and LG16 (Fig. 3). Seven genes were clustered around 0.6 cM on LG 13, including *Mal d 1.01* and six *Mal d 1.03* genes (*Mal d 1.03A–F*). *Mal d 1.01* and *Mal d 1.03F* were the anchor genes for the consensus *Mal d 1* gene cluster on LG 13 from JO and PM (Fig. 3). Nine *Mal d 1* genes were mapped in a cluster of 2 cM on LG 16, whereas one gene (*Mal d 1.05*) was mapped on LG 6. Within the *Mal d 1* gene cluster on LG13, one recombination event was found between two subclusters, each of them having three or four genes located at identical map position (no recombination events were observed) (Fig. 3, LG13-cons). Similarly, two subclusters were found on LG16, one including *Mal d 1.02, -04, -07, -08* and *Mal d 1ps1*, the other including three similar *Mal d 1.06* genes (Fig. 3, LG16-PM). The marker for *Mal d 1.03G* (AY789274) could not be mapped because it did not segregate in any of the three mapping populations. So this unmapped sequence represents a different gene rather than a different allele of the mapped genes. We fully clarified the allelic constitution of all seven intron-containing genes of PM and FS, as of three of the eleven intronless genes.

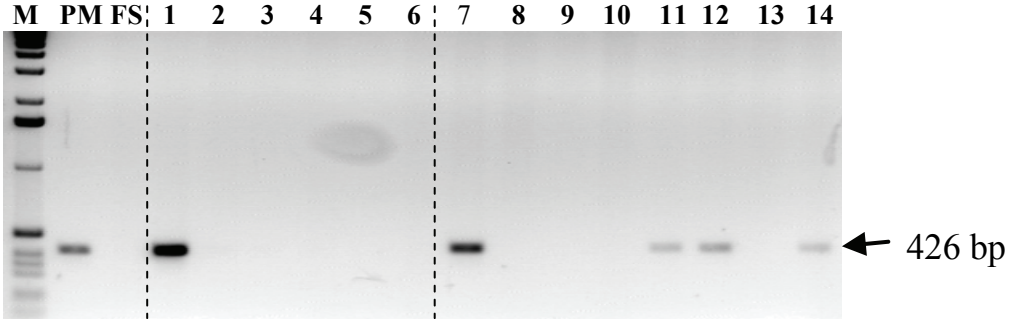

Fig. 2 Specificity and segregation of a Mal d 1.01 marker for sequence AY789236 (from Prima). Lanes (from left to right): M-molecular 1 kb ladder; PM, Prima; FS, Fiesta; 1, positive cloned DNA of sequence AY789236; 2 and 3 are cloned *Mal d 1.01* DNA of sequences AY789237 and AY789238, respectively; 4–6: three *Mal d 1.02* cloned DNA; 7–14: genomic DNA of 8 descendants of the cross PM x FS.

Table 3 Description of sequence-specific primer pairs for *Mal d 1* isoallergen genes

Sequence and marker name	Primer sequence (5'-3') [a]	Position[b] (gDNA)	PCR (nt)	Tm [c]	cycles [d]	Presence & segregation[e]	LG [f]
AY789236 Mal d 1.0105.01a	Fw-AGCTGAAATCCTTGAA**C**GAA	123–142	426	55/53	10/30	PM (PMxFS)	13
	Rv-CAATGTTTCCCTTGGTG**A**GA	528–547					
AY789237 Mal d 1.0105.03	Fw-TGAAGCACAGGATTGAC**A**CG	371–390	177	57/55	10/30	PM (PMxFS)	13
	Rv-CAATGTTTCCCTTGGTGT**C**G	528–547					
AY789238 Mal d 1.0105.02	Fw-TGAAGCACAGGATTGAC**G**CA	371–390	346	56	35	**FS** (JOxPM)	13
	Rv-CCACACAACCTTCGACT**C**A	698–716					
AY789239 Mal d1.0201.09	Fw-TACACCTCTGAGATTCCAC**G**AC	28–49	617	57/55	10/30	PM (PMxFS)	16
	Rv-TATGCGTCGGGGTGTCCCT**A**G	624–644					
AY789240–41 Mal d 1.0201.01	Fw-TCACTTTTGGTGAAGGTC**T**G	170–189	252	57	35	PM/**FS** (JOxPM)	16
	Rv-AGGCGTATGAGTAGTTT**A**CC	402–421					
AY789242 Mal d 1.0404.01	Fw-TCCTTGAGGCGATAACC**C**T	71–90	237	60/58	10/30	PM (PMxFS)	16
	CGTATTGACTGCCTACAAAC	288–307					
AY789243 Mal d 1.0404.02a	Fw-GGCATAAGTTCAAGTTTCTT**C**A	1240–1261	390	57	35	PM (PMxFS)	16
	Rv-TTAGGAATCAGAATGGGTG**T**G	1609–1629					
AY789244 Mal d 1.0404.02b	Fw-TTGCCAGGTCTACTTTCA**T**C	49–68	699	61	35	**FS**	
	Rv-AACACCTCCATCGCCTTCG**T**T	727–747					
AY789245–46 Mal d 1.0501.01	Fw-ATGGCGATAACCTCATCC**G**A	77–96	229	57	35	**PM**/FS	
	Rv-GCCTGCAAATGAATTAAAAC**TA**	284–305					
AY789247 Mal d 1.0502.01	Fw-TCAAAAGCACTGAAATCGTT	116–135	450	57	35	FS (PMxFS)	6
	Rv-TCAAGAAGCTTGAAGAGA**G**T	546–565					
AY789248 Mal d 1.06A02.01	Fw-GAAGGTTAGTTTAATTTCCACA	181–202	164	56	35	PM (PMxFS)	16
	Rv-***GTTT***GCTGTATTCACTCCCTGTCA	321–340+4					
AY789249 Mal d 1.06A01.01a	Fw-GAAGGTTAGTTTAATTTCCACA	181–202	152	56	35	**FS**	
	Rv-***GTTT***GCTGTATTCACTCCCTGTCA	309–328+4					
AY789250 Mal d 1.06A01.01b	Fw-GAAGGTTAGTTTAATTTCCACA	181–202	154	56	35	PM (PMxFS)	16
	Rv-***GTTT***GCTGTATTCACTCCCTGTCA	312–330+4					
AY789251 Mal d 1.06B01.01	Fw-GTA**G**AATGCCCTTGTCCTTG**T**C	57–78	511	57	35	**FS**	
	Rv-TCTTTGCCAGCCTTAACATGC**C**T	544–567					
AY789252 Mal d 1.06B02.01	Fw-CCACCATTTTCTCCATTAACTT**C**A	198–221	360	62/60	5/35	PM (PMxFS)	16
	Rv-GCCTTAACATGCTCTTCCTTGATT	534–557					
AY789253 Mal d 1.06B03.01	Fw-GAAATTCACGAGGGAGATGG**T**A	122–148	436	62/60	5/35	PM (PMxFS)	16
	Rv-GCCTTAACATGCTCTTCCTTGATT	534–557					
AY789254 Mal d 1.06C01.01	Fw-GAAACTGAATACGCCT**T**CG	19–37	383	57	57	**FS**	
	Rv-GGCATCTCCTTCAATCAA**G**T	382–401					
AY789255 Mal d 1.06C03.01	Fw-GAAACTGAATACGCCT**T**CG	19–37	513	55/53	5/35	PM (PMxFS)	16
	Rv-CCTTGACATGTTCTTCCT**A**A	513–531					
AY789256 Mal d 1.06C02.01	Fw-GAAACTGAATACGCCT**T**CG	19–37	412	53/51	5/35	PM (PMxFS)	16
	Rv-CTTAAT**G**TCATAAGAGATCTT**G**TG	417–440					
AY789257 Mal d 1.0701.01	Fw-GGTGTTTTCACATACGAA**C**T	3–23	266	60/58	5/35	**FS** (JOXPM)	16
	Rv-CTCCCTCAATGACACTGTAT**A**G	247–268					
AY789258 Mal d 1.0702.01	Fw-CAAGGAAAACTTTGTGTACC**T**G	228–249	215	56	35	PM (PMxFS)	16
	Rv-CAATCAACTTGAAGAGATGTG**T**A	420–442					
AY789259–60 Mal d 1.0801.01	Fw-GACTGCTCTTGATGGTG**T**TG	66–85	315	57/55	5/35	PM/**FS**	
	Rv-CAACATCACCCTTAGTGT**C**G	361–380					
AY789261 Mal d 1.0801.02	Fw-GACTGCTCTTGATGGTG**T**TG	66–85	315	58/56	5/35	PM (PMxFS)	16
	Rv-TCAACATCACCCTTAGTGT**C**A	360–380					
AY789262 Mal d 1.0901.01	Fw-TGCTAGGTTGTTCAATGCGAC	46–68	299	65/64	5/35	PM (PMxFS)	16
	Rv-GCTTGATGATCGAACCGTTGTT	325–346					

Table 3 continued

Sequence and marker name	Primer sequence (5'-3') [a]	Position[b] (gDNA)	PCR (nt)	Tm [c]	cycles [d]	Presence & segregation[e]	LG [f]
AY789263 Mal d 1.03A01	Fw-TCTGAGTTCACCTCCGTCATT Rv-TTCCTTGATCTCAACGTCA**G**T	22–42 370–390	369	60	35	**FS** (JOxPM)	13
AY789264 Mal d 1.03B01	Fw-GTCTTCACATACGAATCCG**C**A Rv-AACCGCTGCCGGAA**T**CCACT	7–27 315–335	329	65/63	5/30	**PM**	
AY789265 Mal d 1.03B02	Fw-TCCTTGAAGGAGATGGCG**A**C Rv-TCAATCACACTGTACTTGT**T**A	131–150 243–263	133	57	35	**FS** (JOxPM)	13
AY789266 Mal d 1.03C01	Fw-CCTTATCCCCAAGATTGCAC**A**T Rv-AACATCTCCCTTGGTG**C**GA	87–108 360–378	292	60	35	**PM**	
AY822725 Mal d 1.03C02-GD	Fw-CACCGAGATCCTTGGAG**T**AA Rv-GCTGGTGCTCTTGATG**T**TGC	123–142 335–354	232	57	40	JO (JOxPM)	13
AY789267 Mal d 1.03D01.01	Fw-AAGCAGTGAAGAGCGC**A**GAA Rv-TTGATGACGGAACCGCT**T**CT	110–129 325–344	235	57	35	PM (PMxFS)	13
AY789268 Mal d 1.03D02.01	Fw-GAGACTAAGTTGGTTGCTG**A**CG Rv-TTAGTTGTAGGCATCCTGA**A**G	304–325 460–480	177	58	35	PM/FS	
AY789269 Mal d 1.03E01.01	Fw-AGATCCTTGAAGGAGAT**A**G**C** Rv-TTAGTTGTAGGCATCCTGG**C**T	126–147 460–480	355	54	35	PM/FS	
AY789270 Mal d 1.03E02.01	Fw-CGAATCCGAGTTCACCT**G**TA Rv-TTAGTTGTAGGCATCCTGA**A**T	18–37 460–480	463	57	35	PM (PMxFS)	13
AY789271 Mal d 1.03F01.01	Fw-CTGCTAGGTTGTTCAAT**C**CA Rv-TGATGACGGAACCGCTGC**T**G	47–66 324–343	298	57/55	10/30	PM (PMxFS)	13
AY789272 Mal d 1.03F01.02	Fw-TGCTAGGTTGTTCAAT**C**CA Rv-CTTGATGACAGAACCGTT**C**CT	47–66 325–345	299	57/55	10/30	PM (PMxFS)	13
AY789273 Mal d 1.03F02.01	Fw-CTGCTAGGTTGTTCAAT**C**CA Rv-CTTGATGACAGAACCGTT**C**CT	47–66 325–345	399	55	35	**FS** (JOxPM)	13
AY789274 Mal d 1.03G01	Fw-AAGGACAACTTTGTGTACAAG**A**AT Rv-TCACCCTTGGTGTAGTAGT**C**A	229–252 354–374	146	57	35	**PM/FS**	?
AY789275 Mal d 1ps1	Fw-GTACAATGCCTTTATTCTTGA**C**T Rv-GAATATAGCATCTCCTCCT	57–79 261–279	223	57	30	PM (PMxFS)	16

[a] Nucleotides in bold type and underlined indicate introduced mismatches

[b] Positions count from the ATG start codon of the gDNA sequence.

[c, d] PCR annealing temperatures (Tm) and number of cycles, respectively. 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.

[e] PM, Prima; FS, Fiesta; bold type refer to homozygous presence of the sequence ; segregating mapping population are shown in brackets.

[f] Linkage Group (LG)

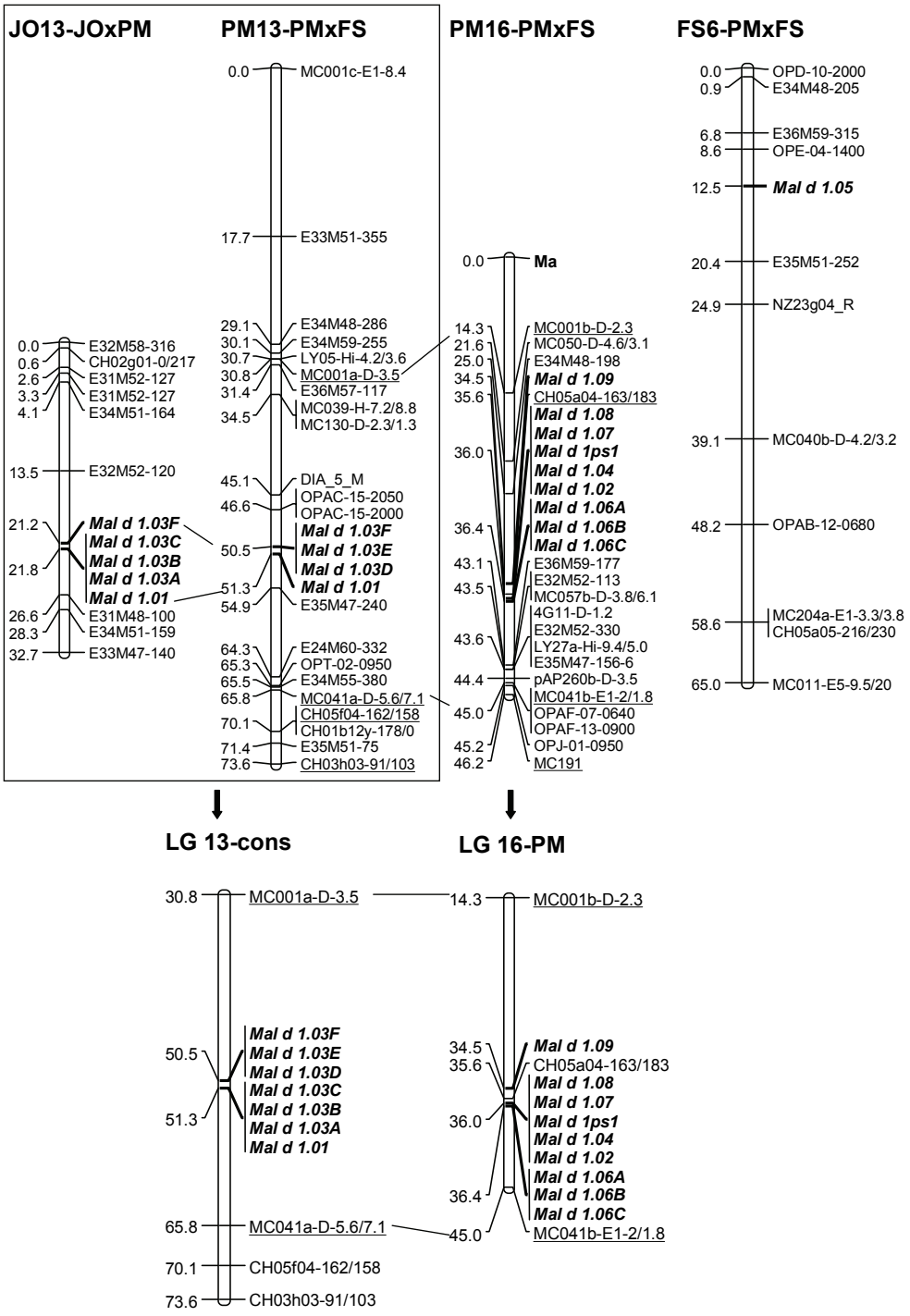

Fig. 3 Map positions of 17 loci of *Mal d 1* genes on linkage group (LG) 6, 13 and 16 of the cultivars Jonathan, Prima, and Fiesta. The linkage groups 13 and 16 are homoeologous, which is reflected by the presence of common RFLP markers (MC001 and MC041) and *Mal d 1* loci at similar mutual distances. The order of *Mal d 1* genes that are located at identical map position, is arbitrary. Reference RFLP and SSR markers flanking the *Mal d 1* clusters were included in the simplified consensus maps.

Genome characteristics of *Mal d 1* isoallergen genes

Sequence differences and similarities among *Mal d 1* genes can be examined for the coding, intron and upstream region. All 17 identified *Mal d 1* genes have a coding region of 480 nt, except for *Mal d 1.04* and *Mal d 1.05* in subfamily II with 483 nt as in *Bet v 1*. Comparison of our new *Mal d 1* coding sequences revealed different levels of identity: 71% to 83% among the four subfamilies, 86% to 98.1% among genes within a subfamily, 98.3% to100% among alleles of a single gene. Sequence identity for the intronless genes on LG 13 was over 95%. Some characteristic polymorphic nucleotides for each subfamily or each gene could be found by alignment of all *Mal d 1* genomic sequences. For instance, counted from the ATG start codon, 174T, 239A, 240C, 241T, and 248C nucleotides were exclusively present in *Mal d 1.01* and *Mal d 1.02* (data not shown). Gene specific polymorphic sites combined with allele specific polymorphisms in any position were frequently used to design sequence-specific primer pairs in this study.

As shown in Figure 1, the *Mal d 1* gene family can be split into two categories with or without intron. For the intron-containing genes, the intron size was specific for each subfamily and gene. *Mal d 1.01* has an intron of 168 nt and *Mal d 1.02* of 171 nt, which are the same as previously reported (Ziadi et al. 2001). *Mal d 1.04* has an intron of 111 nt (Hoffmann-Sommergruber et al 1997), and its new closely related *Mal d 1.05* gene has an intron of 119 nt. Although three genes in subfamily III are highly similar in their coding regions, they have three different intron sizes—*Mal d 1.06A* has an intron of variable size (130–142nt) due to variation in a simple sequence (CA) repeat, *Mal d 1.06B* of 153 nt and *Mal d 1.06C* of 128 nt. Their introns started always at position 184 and have the same 5'splicing site of AG:GT, whereas their 3' splicing sites were different. *Mal d 1.*02 had two 3'splicing patterns, AG:GC and AG:GT. *Mal d 1.01* of subfamily I and all genes of subfamily II (*Mal d 1.04* and *Mal d 1.05*) had AG:GC. All genes of subfamily III (*Mal d 1.06A, -B,* and *-C,*) had AG:GG. Another feature in the intron was the putative branchpoint sequence for the processing of pre-mRNA within 40 nt before the 3' splicing site: CTAAC was present in *Mal d 1.02* and *Mal d 1.06A*, CTAAT in *Mal d 1.01*, *Mal d 1.04*, *Mal d 1.05* and *Mal d 1.06B*, and CTAGT in *Mal d 1.06C*. Only the first two putative branchpoint sequences fit to the consensus mRNA sequence CURAY for efficient intron splicing (Simpson et al. 2002).

With regard to variation in the upstream region, we examined the sequence of 60 nt before the start codon ATG of eight genes. Intron containing genes (*Mal d 1.01*; *02*; *04* and *06A*) showed a conserved TCATC sequence directly preceding to ATG, while the intronless genes

(*Mal d 1.03A, -03D, -03E* and *-03F)* had the GAGAATC sequence. The five *Bet v 1* sequences (accession X77599, X77601, X15877, X81972, X82028) that have upstream parts also showed a conserved CATC (Swoboda et al. 1995). By scanning our sequences with Plantcare database online (Lescot et al. 2002) for putative binding sites, 594 nt upstream region of our three sequences of *Mal d 1.04* contained the same transcription factors as those found previously in four genomic sequences corresponding to *Mal d 1.01, -02, -03D,* and *-07* (AY026910, AY026911, AY026908, AY026909), such as the Box-W motif (the type member of the PR-10 family), ERE-element, TCA-element, AuxRR-core and ERRE-motif (involved in fungal elicitor, ethylene, salicylic acid production , auxin and biotic elicitor responsive elements, respectively) (Ziadi et al. 2001). *Mal d 1.04* had a second TATA box at 44 nt upstream to the start codon apart from a common TATAAAT box at around -100 nt found in four *PR-10* genes (Ziadi et al. 2001).

Deduced amino acid sequences

The 39 genomic sequences from PM and FS represented 28 different potential proteins (Fig. 4). Apparently, various silent mutations occurred due to which allelic variation at the genomic level did not result into variation at the protein level. For instance, three allelic gDNAs (AY789236–AY789238) of *Mal d 1.01* encoded the same amino acid sequence (Mal d 1.0105) (Table 2). Comparison between protein sequences showed that their identity varied from 65–81% among subfamilies, 82%–100% among genes within a subfamily, and 97.5%–100% among alleles within one gene. Mal d 1 proteins in subfamily III were 95.6–96.9% identical, so we named them as one isoallergen group Mal d 1.06. Within subfamily IV, *Mal d 1* genes located on LG 13 were more than 95% identical in amino acid sequence and were therefore classified as isoallergen Mal d 1.03, while three genes on LG 16 shared less than 95% identity to Mal d 1.03 and thus were assigned as different isoallergens (Mal d 1.07–09). As a special case, Mal d 1.03F01 (encoded by AY789271 and AY789272 from Prima) and Mal d 1.03B02 (encoded by AY789265 from Fiesta) had an identical amino acid sequence although these DNA sequences showed 3.5% dissimilarity. The unmapped gene was intronless and coded for a protein that is more closely related to Mal d 1.03 isoallergens than to Mal d 1.07, -08 and -09, and was subsequently denoted as *Mal d 1.03G*. The predicted molecular weights of our Mal d 1 isoallergens were in a narrow range of 17.3–17.6 kDa, whereas their calculated isoelectric point (pI) varied from 5.1 to 6.2.

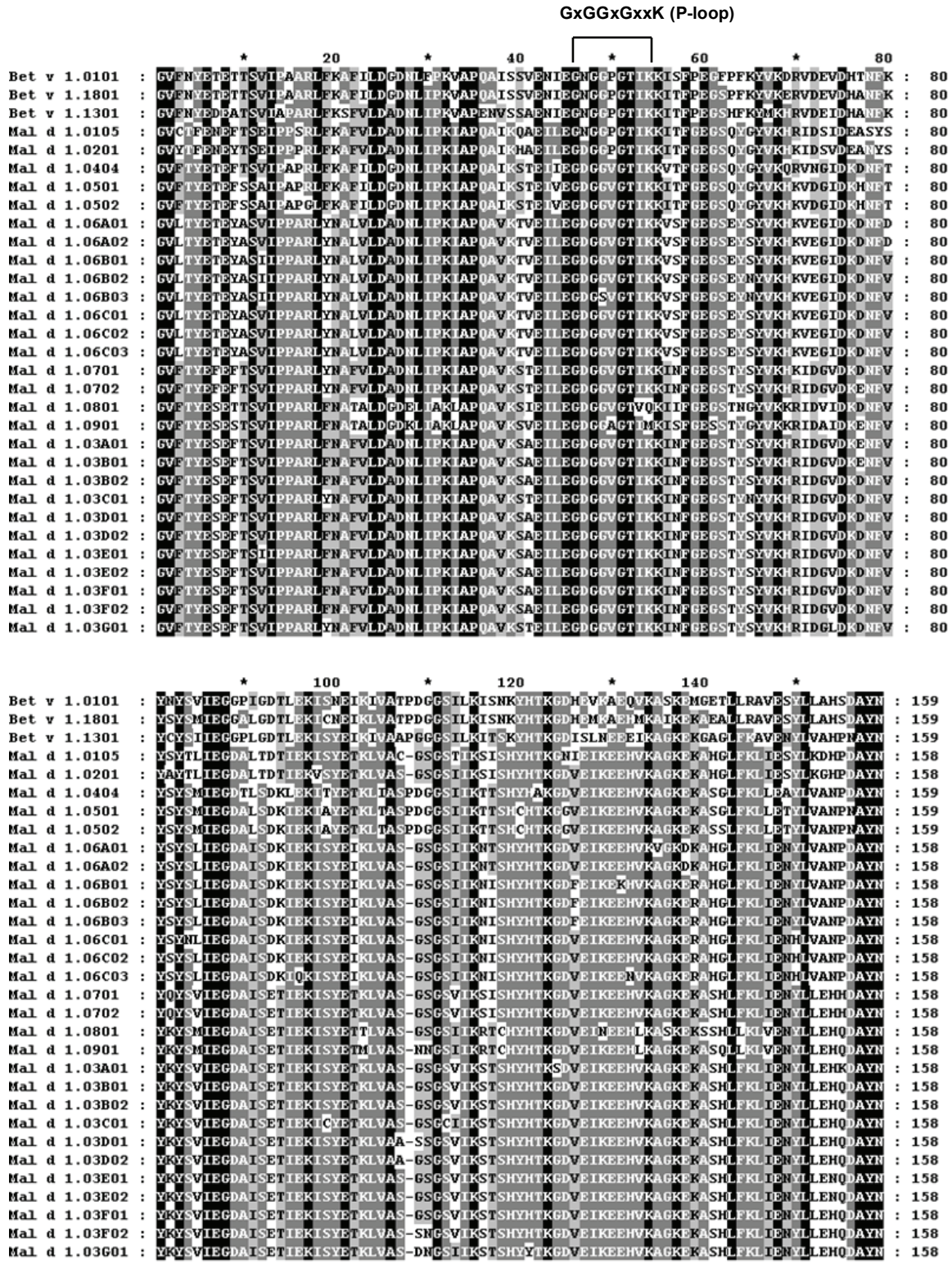

Fig. 4 Amino acid sequences of the Mal d 1 isoallergens and variants from cultivars Prima and Fiesta together with those of Bet v 1.0101, Bet v 1.1301 and Bet v 1.1801. Dashes at position 108 indicate gaps. Four levels of identity shown by shade: (1) black, 100%; (2) grey, 80%; (3) light grey, 60%; (4) white, differences

Discussion

Our results on genomic cloning and linkage mapping revealed that the *Mal d 1* gene family consists of at least 18 members. Except for *Mal d 1.05* on LG 6 and an unmapped gene, all these genes were located in two clusters on the two homoeologous linkage groups 13 and 16. This study forms the basis for a better understanding of the genetics of Mal d 1 and enables further research on the occurrence of allelic diversity among cultivars in relation to allergenicity and on biological functions.

Mal d 1 gene family organization and evolutionary origin

We identified 18 *Mal d 1* genes of which 16 are organized in a duplicated cluster located between two common RFLP markers (MC001 and MC041, Fig. 3) of LG 13 and 16 at similar mutual distances. We thus confirmed that LG 13 and LG 16 are homoeologous linkage groups, which fits with the known duplicate nature of the apple genome (Maliepaard et al. 1998). The position of the *Mal d 1* genes both in the phylogenetic tree (Fig. 1) and on the linkage maps (Fig. 3) gives some clues to the relationships among members in view of gene duplication. Being present on both the homoeologous LG 13 and LG 16, genes of subfamilies I and IV reflect the amphidiploid origin of the apple genome. Moreover, the intronless genes of subfamily IV are located closely above the intron containing gene of subfamily I on both linkage groups. Genes of subfamily II and III are only present on LG 16, but not on LG 13, indicating that these two linkage groups evolved differently. The presence of *Mal d 1.05* on linkage group 6 is unexpected, but its mapping is based on two reliable SNAP markers for the specific allele of FS. Two additional Mal d 1-related genes, Mal d 1m (AY428588) and Mal d 1n (AY428589) described by Beuning et al. (2004), were not covered because these sequences were not available at the time of our study. A DNA phylogenetic tree analysis indicated that Mal d 1m belongs to subfamily II, but is distinct from both *Mal d 1.04* and *Mal d 1.05* in sequence identity (data not shown) and in the size of its coding region (486 nt instead of 483 nt). Mal d 1n does not meet the threshold level of 67% sequence identify for assignment of different proteins to the same allergen (King et al. 1995), make its Mal d 1 membership questionable.

Comparisons between Mal d 1 and Bet v 1 amino acid sequences

Amino acid sequence comparison of 28 Mal d 1 isoallergens and variants from cultivars Prima and Fiesta with all known isoallergens of Bet v 1 revealed the highest identity with Bet v 1.13 (acc:X77601) and Bet v 1.18 (acc:Z724231) in a range of 58–64% and 59–62%, respectively. Figure 4 shows the alignment of these 28 amino acid sequences together with Bet v 1 isoallergen 01, 13 and 18. Subfamily II (Mal d 1.04 and Mal d 1.05) is the only one that has the same protein size as Bet v 1 (159 amino acids) and shows the highest identity (59%–67.7%) with Bet v 1, while all other Mal d 1 proteins consist of 158 amino acids. All these Mal d 1 and Bet v 1 sequences shared 44 conserved amino acids. The Mal d 1 amino acids in segment 108–113 are identical to Bet v 1.01 and Bet v 1.18 including the residue S112 which is essential for IgE binding and cross-reactivity (Son et al. 1999). The conserved sequence motif GXGGXGXXK or P-loop (Spangfort et al. 1997) at amino acid residues 47–52 was observed in all the Mal d 1 isoallergens except for Mal d 1.06B03, Mal d 1.08 and Mal d 1.09 (Fig. 4). Ferreira et al. (2000) suggested that amino acid positions 10, 30, 57, 112, 113 and 125 are important for IgE-binding of Bet v 1 and Mal d 1. Sequence alignment of our deduced Mal d 1 amino acids showed some substitutions in five of these six positions (Fig. 4). Recombinants of these isoallergens and variants identified in this study can be used to assess their allergenicity.

Common features of *Bet v 1* homologues

Our results are in line with the four common features of *Bet v 1* (*PR-10*) homologue genes. First, *Mal d 1* genes have either a 480- or 483-nt open reading frame (ORF) with one or two exons, which is similar to *Bet v 1* and its homologues (ORF of 465–489 nt) (Hoffmann-Sommergruber et al. 1997; Liu and Ekramoddoullah 2004). Second, all seven *Mal d 1* intron-containing genes have the same 5' splicing site at DNA position 184 (amino acid position 62) as most other *Bet v 1* homologues (Hoffmann-Sommergruber et al. 1997). Third, *Mal d 1* comprise a muligene family and most members are clustered in the genome. At least three *Bet v 1* genes (clone Sc1-3, also called Ypr-10 a–c) are within a genome segment of 14 kb (Hoffmann-Sommergruber et al. 2004). In soybean, the Bet v 1 homologue SAM22 allergen consists of ten genes present in a tandem array (Crowell et al. 1992; Kleine-Tebbe et al. 2002). Fourth, *Bet v 1*-related genes are expressed in different tissues and under biotic or

abiotic stresses (Hoffmann-Sommergruber and Radauer 2004). Several expression studies on apple *Mal d* 1 genes also showed some difference for individual genes. *Mal d 1.02* (GenBank accn. L42952, AF020542) is fruit ripening-related (Atkinson et al. 1996) and stress and pathogen inducible (Puehringer et al. 2000). PR-10c (*Mal d 1.01*) and PR-10a (*Mal d 1.02*) will express much more strongly than PR-10d (*Mal d 1.03C*) and PR-10b (*Mal d 1.03D*) upon a salicylic acid analogue induction (Ziadi et al. 2001), whereas *Mal d 1.06A, Mal d 1.03E* and *Mal d 1.03F* are auxin responsive (Kodde and Van der Geest, unpublished). A recent comprehensive expression study on *Mal d 1*-related genes in apple demonstrated that eight out of 12 genes were expressed in tree-ripened fruit, and most of these were also expressed in leaves in response to a challenge with *Venturia inaequalis* — a fungus causing apple scab (Beuning et al. 2004).

Application in defining different *Mal d 1* genes and allelic variations

Our mapping results clearly demonstrated that alleles of the same locus have over 98.3% sequence identity at the DNA level. A diploid cultivar allows at the most two different alleles for a single locus. These facts can be used to judge whether different *Mal d 1* sequences reflect allelic variation of a single gene or originated from different genes. As shown in Table 2, four very recent sequences identified by Beuning et al. (2004) in subfamily III are classified to two genes with two allelic sequences for each independent gene. If more than two almost identical sequences are identified from a single diploid cultivar, then some of these may contain PCR errors or belong to duplicated genes.

Breeding aspects

Once the alleles for high and low allergenicity have been identified and their linkage phase in the parental cultivars is known, genotyping within a breeding programme becomes useful. In this respect, it is worthwhile to investigate the allelic diversity of each *Mal d 1* gene for the founders of a breeding programme. We have indications that especially LG 16 is involved in allergenicity (Gao et al. unpublished results). LG 16 also contains genes for taste of apple fruit, including the *Ma* gene for acidity (Maliepaard et al. 1998; King et al. 2000) and a major QTL for juiciness and crispness (King et al. 2000). However, these genes are at 35 cM distance from the *Mal d 1* cluster (Fig. 3). This distance is large enough to allow frequent

recombination, which opens ways to breeding of new cultivars for good taste and low allergenicity.

In conclusion, this research provides fundamental knowledge of the genomic sequences and linkage map position of the *Mal d 1* gene family. Further investigation on allelic diversity of different *Mal d 1* genes and their protein expression in various cultivars would be one more step to determine the genetic causes for the difference in allergenicity among apple cultivars.

Acknowledgements

This research was carried out with financial support from the Commission of the European Communities for the projects SAFE QLK1-CT-2000-01394 and exploited the unpublished results from FAIR5 CT97-3898 (DARE: Durable Apple Resistance in Europe). We thank Bernard Koller and Cesare Gessler for their SSR data to update the linkage maps of the cross PM x FS. We also thank professors Evert Jacobsen and Richard Visser for critical reading of this manuscript. Z. S. Gao is grateful to the International Agricultural Centre, Wageningen, and the ICSC-World Laboratory, Lausanne for providing fellowships to conduct this research.

Chapter 4

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

Z. S. Gao[1,2,3], W. E. van de Weg [1,3], J. G. Schaart[3], G. van Arkel[3], H. Breiteneder[4], K. Hoffmann-Sommergruber[4], L. J. W. J. Gilissen[1,3]

1: Allergy Consortium Wageningen, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands

2: Experimental Nursery of Shanxi Forestry Department, Beiying Nan Lu 52, Taiyuan, Shanxi, 030031. People's Republic of China

3: Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands

4: Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, A-1090 Vienna, Austria

Submitted to *Theoretical and Applied Genetics*

Abstract

Four classes of apple allergens (Mal d 1, -2, -3 and -4) have been reported. 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, 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 differed primarily 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. Three different Mal d 4 isoallergens (Mal d 4.01, -02 and -03) were known from cDNA sequences (AF129426–AF129428). We obtained genomic sequences for each of these three isoallergens, which led to the identification of four *Mal d 4* genes (*Mal d 4.01A and B*, *Mal d 4.02A* and *Mal d 4.03A*). Their complete genomic DNA sequences varied among the genes in length from 862 to 2017 nt. They all contained two introns separating the protein coding sequences into three conserved exons of 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.

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.

TLP is one of the major protein constituents of a 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 researches 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; Cortegano et al. 2004). Hsieh et al. (1995) were the first who identified 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 thaumatin-like protein (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 acids are typically 70% to 80% similar. They

display striking protein features of conserved 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 expressed in individual tissues and cells (Huang et al. 1996). The basic biological functions of profilins have been found in 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 profilin allergy, many fruits contain profilins and their allergenic potency has been reported frequently (Scheurer et al. 2000, 2001; Wensing et al. 2002; Asero et al. 2003; Rodriguez-Perez et al. 2003; Westphal et al. 2004). Three distinct mRNA sequences (AF129426, AF129427 and AF129428) encoding three isoallergens of apple profilin were available in the GenBank/EMBL nucleotide database when we started this research. 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, 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 thaumatin-like protein genes and *Mal d 4* for apple profilins. For further dividing of 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) and taking account of the 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 as 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 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'ends. The PCR cloning and sequencing procedures were the same as described previously (Gao et al. 2005b).

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	1296–1298	AY792598	AY792599
		Mald2-Rev1: TGTTCCTTTATTTGATCGTGT	919–939				AY792600	AY792601
		Mald2-For2: TCAATGTCGATGATGAAGAGCCA	48–70	58/25	60/2	1159–1161	AY792602	AY792603
		Mald2-Rev2: CATCATCTGCATATATAATCCCAT	805–827				AY792604	
Mal d 4.01	AF129426	Mald4-For: TGTCGTGGCAGGCGTACGTC	2–21	52/30	54/2	2016	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	1029–1032	AY792614	AY792616
		Mald4.03-Rev: TTAGAGACCCTGCTCGATAAG	376–396				AY792615	AY792617

[a] Refer to reference sequence

Sequence analysis

DNA sequences and single nucleotide polymorphisms (SNPs) were analyzed using the Seqman program (DNAstar, Madison, Wis.). 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, Madison WI). Multiple DNA and amino acid sequence alignments were performed with the GeneDoc program (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 repeats

(SSR). 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 linkage maps derived from Prima x Fiesta (PM x FS, n=141), Fiesta x Discovery (FS x DS, n=70) and Jonathan x Prima (JO x 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).

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 and Fiesta (Table 1). The sequencing of 29 true clones resulted in four different sequences from Prima (GenBank accession nos: AY792598, AY792600, AY792602, AY792604) and three from Fiesta (AY792599, AY792601, AY792603). These seven sequences had a nucleotide identity of 98.5–100%; two sequences from Prima were identical to two sequences from Fiesta (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 nt or 380 nt (Table 2), and their coding sequences showed one SNP to 14 SNPs (Data not shown). Four *Mal d 2* genomic sequences from Prima could be divided into two sets, 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. These four sequences strongly suggested the presence of two loci which were named *Mal d 2.01A* and *Mal d 2.01B*.

For mapping, five sequence-specific SNAP markers were created and tested on the PM x FS population (Table 3). Four markers segregated and could be mapped on LG 9 of PM and FS at the same position (Fig. 1). Three markers showed a segregation pattern typical to a so called *ef* x *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. These three markers belong to a single locus named *Mal d 2.01A*. The fourth marker Mal d 2.01B01.01 was not segregating because it is homozygous in FS and heterozygous in PM, indicating a second locus named as *Mal d 2.01B*. The fifth marker Mal d 2.01B01.02 was present only in PM and had exactly the same segregation pattern as the marker Mal d 2.01A02.02, which imply that these two alleles are in coupling phase and located at identical map position. During the cloning process, we also observed a few clones that are not true due to PCR errors or recombination of different alleles. One artefact sequence recombining *Mal d 2.01B* and *Mal d 2.01A01* was obtained from FS. Sequence-specific primers for this artefact was failed to amplify the expected band in FS, but surprisingly valid on cv Discovery (DS). Moreover, this marker (Mal d 2.01B-DS) was mapped on LG 9-DS at the similar position as those on LG 9-PM and LG 9-FS (Fig. 1).

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	Exon2	Intron 2	Exon 3
Mal d 2.01A	1119	741	61	378	680		
Mal d 2.01B	1121	741	61	380	680		
Mal d 4.01A	2017	396	123	343	138	1278	135
Mal d 4.01B	2017	396	123	343	138	1278	135
Mal d 4.02A	862/863	396	123	248	138	218/219	135
Mal d 4.03A	1030/1032/1033	396	123	386/388/389	138	248	135

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 & segregation [f]	LG [g]
Mal d 2.01	Mal d 2.01A01.01	Fw-TCGATGATGAAGAGCCA**T**GT	33–52	60	35	885	PM/FS	9
	AY792598, AY792599	Rv-AAGTGGAGCCGGGCACA**T**CT	898–917				PMxFS (101:39)	
	Mal d 2.01A02.01	Fw-TCAACTTGCCTATGTCT**C**TG	819–839	56	35	199	PM,	9
	AY792600	Rv-CCGGCGTATCATTCAGCG**A**G	988–1017				PMxFS (72:68)	
	Mal d 2.01A02.02	Fw- TAAAAGTTCTCAATCCTGCC	338–357	57	35	844	FS,	9
	AY792601	Rv-ATCTGCATATATAATCCCATTT**G**T	1158–1181				PMxFS (72:68)	
	Mal d 2.01B01.01	Fw-CCTCACCTTGGCCATCCTC**A**TT	68–89	60	35	952	PM, **FS, JO**, DS	
	AY792602; AY792603	Rv-CCGGCGTATCATTCAGCG**A**G	1000–1019					
	Mal d 2.01B01.02	Fw-CCTCACCTTGGCCATCCTC**A**TT	68–89	51	35	301	PM	9
	AY792604	Rv-CACTTATATGTATCC**C**GATTGT	347–368				PMxFS (72:68)	
	Mal d 2.01B-DS	Fw-CCTCACCTTGGCCATCCTC**A**TT	68–89	62	35	852	DS	9
		Rv-AAGTGGAGCCGGGCACA**C**CT	900–919				FS x DS (33:37)	
Mal d 4.01A/B	Mal d 4.01A01	Fw-GCAAAGTTTTTCTGGTCT**C**TA	1828–1848	59	45	113	FS	9
	AY792605	Rv-CATATATACCGAAAACCAGAG	1920–1940				PMxFS (64:70)	
	Mal d 4.01A02	Fw-GACGGCCACCATCT**T**ACA	46–63	58	35	1896	PM,FS,JO,DS	
	AY792606, AY792607	Rv-TCATATATACCGAAAACCAAAG	1920–1941					
	Mal d 4.01B01	Fw-GGCAGGCGTACGTCG**T**T	7–24	56	35	597	PM,FS, JO;	9
	AY792608, AY792609	Rv-CTTCTTGCCACGAATCACA	586–604				JO x PM (126:39)	
Mal d 4.02A	Mal d 4.02A01.01	Fw-TCGTTTTTTTCACTTTCTT**G**TCT	131–153	56	35	674	PM,	2
	AY792610	Rv-CCAGGAGTCATCGGC**A**CA	788–804				PMxFS (70:71)	
	Mal d 4.02A01.02	Fw-CTGAGGAGGTGACTG**C**CA	379–396	58/56	10/30	428	FS	2
	AY792611	Rv-G**T**CAGGAGTCATCGGC**A**CG	788–806				PMx FS (60:81)	

Table 3 continued

Gene	Marker name [a] (detected sequences)	Primer sequence (5'–3') [b]	Position [c]	Tm [d]	Cycles [e]	Product (nt)	Presence & segregation [f]	LG [g]
Mal d 4.02A	Mal d 4.02A02	Fw-CTGAGGAGGTGACTG**T**CG	379–396	60	35	428	PM, FS;	2
	AY792612, AY792613	Rv-G**T**CAGGAGTCATCGGC**A**CG	788–806				PMxFS (105:36)	
Mal d 4.03A	Mal d 4.03A01.01	Fw-TGTCGTGGCAGGCGTACGTC	CAPs-*EcoR* I	60	35	1029	PM,	8
	AY792614	Rv-TTAGAGACCCTGCTCGATAAG	713/316				PMxFS (69:72)	
	Mal d 4.03A01.02b	Fw-GATCAACCCGGAAC**T**CTT	551–568	60	35	244	FS,	8
	AY792616	Rv-TCTTCGACGCAAACTCT**C**AA	775–794				PMxFS (79:62)	
	Mal d 4.03A-SSR-160	SSR-Fw-TTGGTCTCAGAGCGCCTCTT	96–115	62	35	160	PM,	
	AY792614	SSR-Rv-***GTTT***AACCGATCGACCGAACCAGT	228–247(+4)				PMxFS (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)				PMxFS (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)				PMxFS (54:72)	

[a] Marker name and its detected sequence or sequences, ***Mal d 2.01B-DS*** was based on an artefact 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 in the bracket. **FS and JO** are homozygous for Mal d 2.01B01.01

[g] Linkage group (LG)

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 (www.allergen.org), respectively. Based on 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 Fiesta. These thirteen new sequences were deposited in GenBank with accession numbers AY792605 to AY792617 (Table 1). They all had the same deduced open reading frame length of 396 nt, two introns of very different lengths (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 x 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 x PM at a ratio of 3:1 (105:36), 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 x 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 very close 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 likely a *Mal d 4.01A* 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 three unique sequences were developed and tested on PM x FS (Table 2). These markers showed to be allelic fitting perfectly to the segregation pattern of the model cross *ef* x *eg* (Van Ooijen and Voorrips, 2001) as we already mentioned in the mapping of *Mal d 2*. These markers were mapped on LG 2 of PM and FS at similar distance from a common RFLP marker MC003 (Fig. 1).

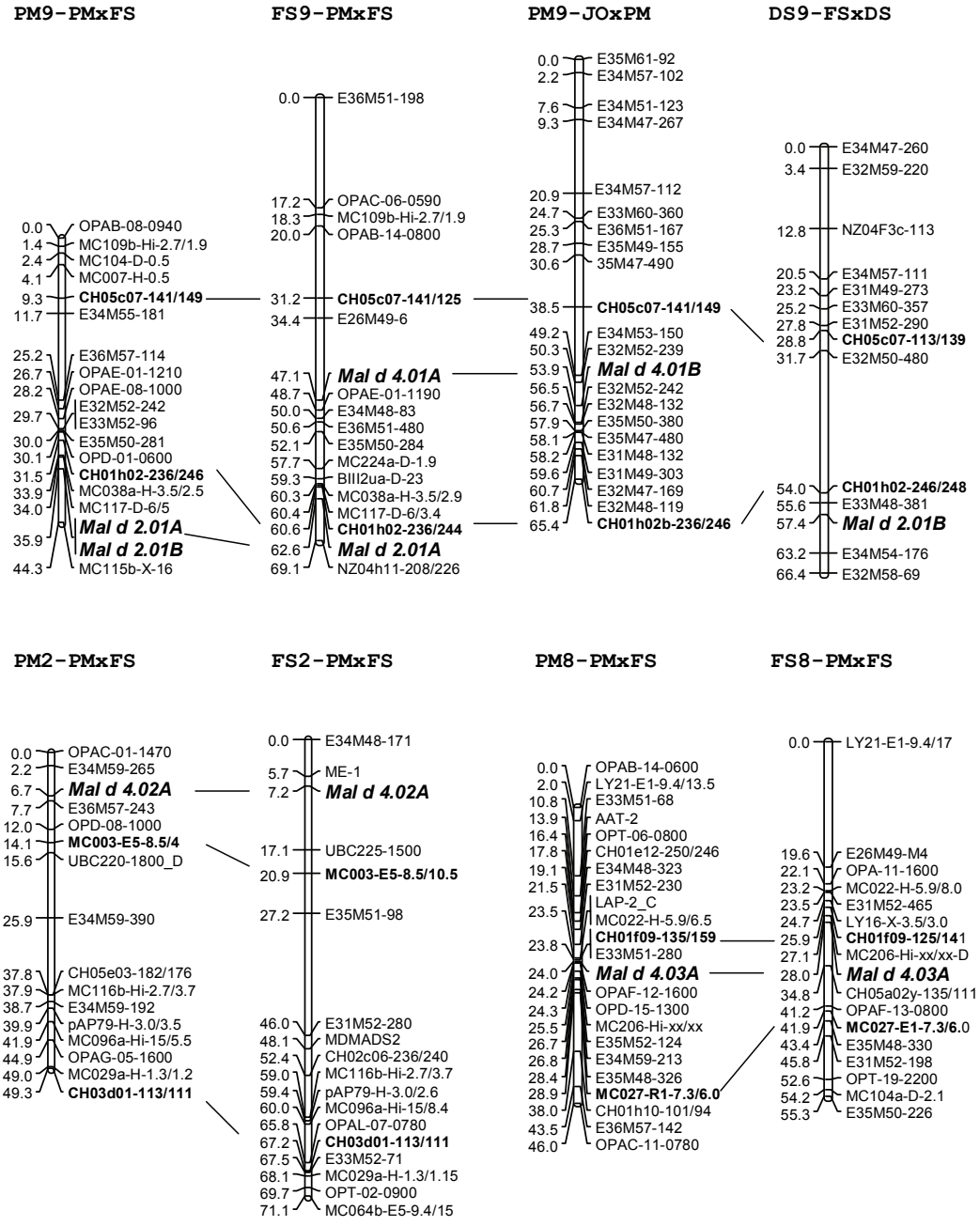

Fig. 1 Positions of two *Mal d 2* and four *Mal d 4* genes on molecular marker linkage maps from three mapping populations. Parental cultivars: PM-Prima, FS-Fiesta, JO-Jonathan, DS-Discovery

With regard to *Mal d 4.03*, PM and FS each had two different sequences of variable lengths (1030–1033 nt) due to a CT-dinucleotide that repeated 7 to 9 times, and a deletion of a single nucleotide (AY792614–AY792417). 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 1290 nt could be cleaved with *EcoR* I into two fragments of 713 nt and 316 nt. A SNAP marker was created for the sequence (AY792616) from FS. In addition, Mal d 4.03A-SSR primers revealed three bands of 160, 162 and 164 nt (Table 2). All these marker data confirmed each other. The mapping process resulted in 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 open reading frame 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 1278 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-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-one SNPs were present among five different sequences of two *Mal d 2* genes. *Mal d 2.01A* differed consistently from *Mal d 2.01B* sequences in nine SNPs in the first exon and intron, only resulting in one amino acid difference (6A or 6V) in the signal peptide (Fig. 3). The second exon from both alleles of *Mal d 2.01B* differed in one SNP from allele *Mal d 2.01A02*. Within *Mal d 2.01A*, two alleles of *Mal d 2.01A01* and *Mal d 2.01A02* differed in ten SNPs in the second exon, resulting in five amino acid changes (Fig. 3). 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.

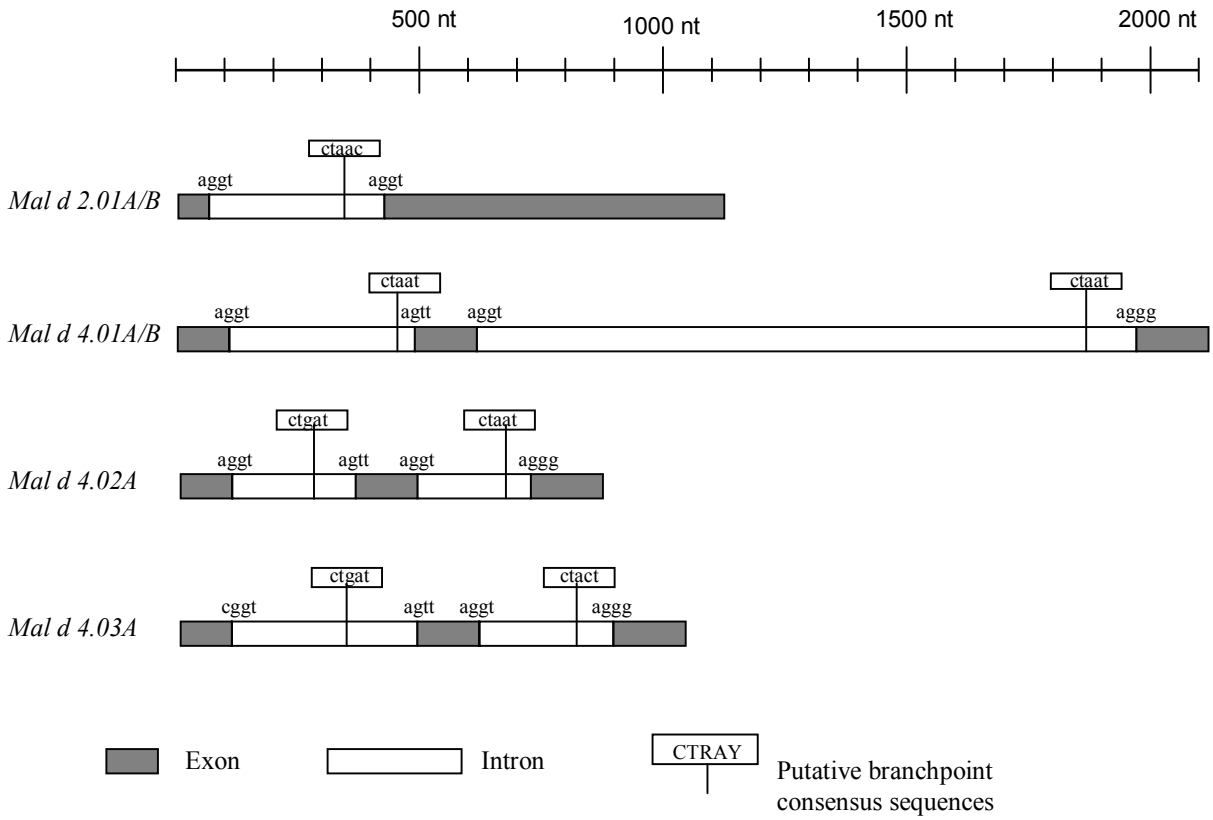

Fig. 2 Schematic illustration of the gene structure of *Mal d 2* and *Mal d* 4 genes

The sequence identity is 78% to 80% for both the coding DNA and amino acids among three *Mal d 4* isoallergen genes. 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 relative long length (1759 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 that 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 simple sequence repeat.

```
                        *        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 2.01A01: AY792600, AY792601; Mal d 2.01B01: AY792602–AY792504

```
                        *        20         *        40         *        60         *
Mal d 4.01A01  : MSWQAYVDDHLMCDIDGHHLTAAAILGHDGSVWAHSSTFPKFKPEEITAIMKDFDEPGSLAPTGLHLGGT :  70
Mal d 4.01A02  : ............................................................................ :  70
Mal d 4.01B01  : .................................Q.......................................... :  70
Mal d 4.0101   : .........R.................................................................. :  70
Mal d 4.02A01  : ............E.E.N..S....I........Q.A...QL....V.G..N..N..........Y.... :  70
Mal d 4.02A02  : ............E.E.N..S....I........Q.A...QL....V.GV.N..N..........Y.... :  70
Mal d 4.0201   : ............E.E.N..S....I..N.....Q.A...QL....V.G..N..N..........Y.... :  70
Mal d 4.02B    : ....T........E.E.N..S....I........Q.A...QL....V.G..N..N..........Y.... :  70
Mal d 4.03A01  : ..............NR........Q.....SQ.AS..A......A..L....Q..T......F.... :  70
Mal d 4.0301   : ..............NR........Q.....SQ.AS..A......A..L....Q..T......F.... :  70
Mal d 4.03B    : ....Q...........NS........Q......Q.A...A......A..L....Q..T......F.... :  70

                       80         *       100         *       120         *
Mal d 4.01A01  : KYMVIQGEGGAVIRGKKGSGGVTVKKTGQALVFGIYEEPLTPGQCNMIVERLGDYLIDQGL : 131
Mal d 4.01A02  : ..................................T....................... : 131
Mal d 4.01B01  : ............................D.............................. : 131
Mal d 4.0101   : ............................................................ : 131
Mal d 4.02A01  : ........P.V.......P.......STM..LI...D..M.......V.........E... : 131
Mal d 4.02A02  : ........P.V.......P.......STM..LI...D..M.......V.........E... : 131
Mal d 4.0201   : ........P.V.......P.......STM..LI...D..M.......V.........E... : 131
Mal d 4.02B    : ........P.V.......P.......STM..LI...D..M.......V........VE... : 131
Mal d 4.03A01  : ........P............I.I...S...LI...D..V......IV.........E... : 131
Mal d 4.0301   : ........P............I.I...S...LI...D..V......IV.........E... : 131
Mal d 4.03B    : ........P............I.I...S...LI...D.........IV.........E... : 131
```

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-AY792606

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.

Discussion

In this study, we cloned and sequenced the genomic DNA sequences of two classes of apple allergen genes, thaumatin-like protein (TLP, Mal d 2), and profilin (Mal d 4). Subsequently we have mapped two similar *Mal d 2.01* genes (A and B) and *Mal d 4.01A/B* on linkage group 9, *Mal d 4.02A* on LG 8 and *Mal d 4.03A* in LG 2 by applying molecular markers in segregating populations. With these new data presented here and previous studies on *Mal d 1* and *Mal d 3* (Gao et al. 2005a, b), four classes of apple allergen genes have been mapped

now. The implications of these data for genetic analysis and breeding of low allergenic apple cultivars are discussed.

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 (741nt)	CO754260, CN996077, CN490732, CO052447, AF494393, AF494394, CV150401, AY548366, CN994747	Fruit, bud, leaf	AY792599–AY792601
Mal d 2.01B	(741 nt)	Not yet	Not known	AY792502–AY792504
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 of AF090143 was reported as 738 nt, but we assign it in the 741 group based on the overall sequence identity.

[b] EST: expressed sequence tag, only some of the longer EST sequences were included

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 open reading frame (ORF) or the 3'untranslated region (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*, but 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 AF494394) in a study on defence responses of apple during compatible and incompatible interactions with *Erwinia amylovora* (the casual 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 or 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 certainly to *Mal d 2.01A* or *B* or both. Profilin gene *Mal d 4.01* appeared to be duplicated and located on LG 9, too. We did not find duplications for the other two Mal d 4 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 x 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 isoallergen genes on the two homoeologous segments of LG 4 and 12. With regard to the three linkage groups 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 LG15 at positions corresponding to the current *Mal d 2* and *Mal d 4* locations. This expectation was rewarded by the following database search. Sequence alignment of recently published thaumatin-like protein and profilin 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 is coding for nearly the entire mature peptide and showed 12 amino acid differences to the two genes mapped by us. Another consensus 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 in the nucleotide databases: *Mal d 4.02B* and *Mal d 4.03B*, both have the same length of 396 nt 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 two amino acids different). *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 are distinct 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). The actual locations for these new *Mal d 2* and *Mal d 4* genes can be determined using the similar approach.

Genomic structure of the TLP gene family in plants

In plants, TLPs are a big gene family with proteins of diverse pI values and two groups of molecular weight: high and low. The overall genomic structure was not conserved in view of exon numbers and lengths. Searching the model plant *Arabidopsis* DNA database for putative TLPs, 19 putative genes have been identified which are scattered on four chromosomes 1, 2, 4 and 5. The open reading frames (ORF) of these TLPs varied from 574 nt to 1998 nt, most of them consisted of two or three exons, and one ORF with four exons. The length of these exons is not conserved either. TLPs in Rosaceous fruit crops showed a closer relationship in amino acid identity and gene structure. In apple, two kinds of TLPs encoded by 741 nt or 738 nt were identified (accession numbers see Table 4). It is known from this research that the 741 nt sequence contains two exons, the first is 61 nt long and the second one is 680 nt. The exon constitute of the 738 nt is not known yet. TLP in Japanese pear (*Pyrus serotina*) has been revealed as an ORF of 735 nt comprises two introns (46 nt and 689 nt) and separated by one intron of 351nt (accession AB036069). A peach TLP (AF362988) also has the same 741 nt coding sequence as *Mal d 2.01*, another TLP from sweet cherry (U32440) has a 738 nt ORF as *Mal d 2.03*. But all these variations do not change the conserved 16 cysteines to form eight disulfide bonds. So the potential allergenicity remains.

Genomic structure of the profilin gene family in plants

The characterization of four apple profilin genomic sequences revealed two introns located in conserved positions that divide the coding sequences into three exons of nearly equal conserved sizes (123, 138, and 135 nt) with exactly the same splicing sites. These features appeared in most plant profilin genes. In *Arabidopsis thaliana* genome sequences, five profilin genes were present on three chrosomes (AC005169, AL161574, NC_003076). They all have three exons, of which profilin 1 and profilin 2 have the same length of three exons

and splicing pattern as those found in apple. Profilin 3, 4 and 5 have a different length of the first exon. Although the first exon of profilin 3 and 4 is 9 nt longer, the splicing pattern is the same as we found in apple. Two rice profilin loci 11 kb apart also had the same pattern in the gene structure. cDNA sequences from maize, pear, cherry, peach and celery in the databases all have 396 nt coding sequences and fit putatively to the conserved three exons in length and splicing sites. According to these common splicing sites, we predicted that the 402 nt birch pollen profilin encoding sequence (Bet v 2, M65179) is consisted of three exons: 129 nt, 138 nt and 135 nt. 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. Furthermore, it supports the "intron first" hypothesis of gene evolution (Hoffmann-Sommergruber et al. 1997) or "the Exon Theory of Genes" (Roy 2003). If the intron had been arisen late, then it would be unlikely that both introns are inserted into the conserved positions. Instead, introns are extremely ancient characteristics of 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

Different apple cultivars can elicit mild or severe allergic reactions in atopic patients (Vieths et al. 1994). The recently finished EU-SAFE project (Plant food allergies: field to table strategies for reducing their incidence in Europe) was initiated to characterise relevant allergens and their relationships to severe versus mild symptomatology using apple as a model. One of the objectives was to map allergen genes on established apple molecular marker linkage maps. Till now, representatives from all four classes of apple allergen genes (*Mal d 1*, *Mal d 2*, *Mal d 3* and *Mal d 4*) have been mapped on eight linkage groups (Gao et al. 2005a, b and this paper). To identify the genetic basis for difference in allergenicity among cultivars will be complicated because of two aspects. One aspect is that different patient populations react differently to these four classes of allergens, or even variation occurred within the same patients population. Clinical assessment of allergenicity is not easy for large numbers of patients. Another aspect is the existence of multigene gene families for these allergens and different linkage group locations. Proteomic analyses to identify those genes expressed in apple fruit are crucial

For conventional breeding, it is essential to investigate allelic diversity of the apple allergen genes among the current apple cultivars to pinpoint individual isoforms relevant for low allergenicity. Then molecular marker assisted selection can be applied. Another alternative is to use RNA interference to silence the responsible allergen genes. Such approach has been applied to Mal d 1 with a preliminary positive result in leaves of the transformed apple (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 x Fiesta and Fiesta x 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.

Chapter 5

Allelic constitution of *Mal d 1* genes on linkage group 16 is related to the differences in allergenicity among apple cultivars

Z. S. Gao [1,2,3], W. E. van de Weg [1,3], C. Matos[3], I. M. van der Meer[3], Y. H. Li [4], S. T. H. P. Bolhaar[5], A. C. Knulst[5], L. Zuidmeer[6], R. van Ree[6], K. Hoffmann-Sommergruber[7], L. J. W. J. Gilissen [1,3]

1: Allergy Consortium Wageningen, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands

2: Experimental Nursery of Shanxi Forestry Department, Beiying Nan Lu 52, Taiyuan, Shanxi, 030031, People's Republic of China

3: Plant Research International, Wageningen University and Research Centre, P.O. Box 16, 6700AA, Wageningen, The Netherlands

4: Institute of Crop Science, China Academy of Agriculture Science, Beijing, 100081, People's Republic of China

5: Department of Dermatology/Allergology, University Medical Center Utrecht, P.O. Box 85500, 3508GA Utrecht, The Netherlands

6: Department of Immunopathology, Sanquin Research, P.O. Box 9190, 1006AD Amsterdam, The Netherlands

7: Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, A-1090 Vienna, Austria

Abstract

Apple cultivars differ in their allergenicity. Mal d 1 is a major apple allergen. Previous genomic cloning and linkage mapping showed that the *Mal d 1* gene family consists of at least 18 members, of which 17 have been mapped on three linkage groups (LG). The aim of this study was to identify those Mal d 1 isoallergen genes or alleles that are closely related to the difference in allergenicity. To achieve this goal, we assessed the allelic diversity of seven representative *Mal d 1* genes among ten cultivars and tested the allergenicity of 14 cultivars by in vivo skin prick test (SPT). Of seven investigated *Mal d 1*genes, *Mal d 1.01* and *Mal d 1.02* were conserved in amino acid sequence: nine out of ten cultivars had one variant of protein, only one cultivar had a second variant. *Mal d 1.04*, *Mal d 1.05* and *Mal d 1.06* are more variable in amino acid sequences with three to six protein variants. Six to eight genomic alleles were identified for each gene. SPT on Dutch patients showed different allergenicity for 14 cultivars. Santana proved to be low-allergenic, similar to one of its parent (Priscilla), but is in contrast to its high-allergenic grandparent, Golden Delicious. Two homologous *Mal d 1* haplotypes on LG 16 of Golden Delicious were completely replaced in Santana, but the genes on LG 6 and LG 13 were kept the same. Based on this observation, we concluded that *Mal d 1* genes on LG 16 are relevant to apple allergenicity. Further genetic analysis combining the *Mal d 1* haplotypes on LG 16 and allergenicity of these 14 cultivars revealed that the two low-allergenic cultivars have a distinct allelic constitution from that of the high-allergenic ones. The favourable alleles for low allergenicity can be detected by the presence of the homozygous Mal d 1.06A-ssr-154 marker.

Introduction

Most birch pollen sensitized patients in central and northern Europe have oral allergy symptoms after eating fresh apples. This is caused by cross-reacting IgE antibodies directed against the major birch pollen allergen Bet v 1, that recognize the Mal d 1 from apple. The molecular relationship between Mal d 1 and Bet v 1 has been demonstrated by the radio allegro sorbent test (RAST) inhibition (Halmepuro et al. 1984), by comparing amino acid sequences (Vieths et al. 1994; Vieths et al. 1995; Vieths 1997) and complete DNA sequences (Vanek-Krebitz et al. 1995; Holm et al. 2001). Isoforms and mutants of Bet v 1 and Mal d 1 displayed different binding capacities to specific IgE antibodies from allergic patients (Ferreira et al. 1996; Son et al. 1999). It has been known for some years that Mal d 1-related allergenicity is apple cultivar dependent. Golden Delicious is highly allergenic as experienced by many apple-allergic patients, whereas Gloster causes mild reactions (Vieths et al. 1994; Hsieh et al. 1995). These observations raised the question whether these differences are due to different expression of Mal d 1 with regard to quality (isoforms) or quantity, or both. Son et al. (1999) did not find strain specific isoforms between the high and low allergenic cultivars after sequencing cDNAs of two Mal d 1 isoallergen genes (*Mal d 1.01* and *Mal d 1.02*) from seven cultivars. But they did observe a tenfold higher content of Mal d 1 protein in Golden Delicious compared to Gloster. Therefore, these authors suggested that genomic research might be helpful to identify the alleles that are expressed differently.

Recently, significant progress has been made in protein identification (Helsper et al. 2002), mRNA expression profiling (Beuning et al. 2004) and in determining the number and map position of *Mal d 1* genes in apple (Gao et al. 2005b). To date, at least 18 members have been identified, of which 17 have been mapped on three linkage groups. Most *Mal d 1* genes are clustered in two homoeologous linkage groups (LG 13 and LG 16), and a single gene on LG 6 (Gao et al. 2005b). Five *Mal d 1* genes (*Mal d 1.01*, *-02*, *03E*, *-06A*, *06B*) were expressed at the mRNA level in ripe Royal Gala apple fruits (Beuning et al. 2004; and gene denotations according to Gao et al. 2005b), two of which are located on LG 13 and three on LG 16. These expressed genes all have an intron except for *Mal d 1.03E*. Q-TOF-MS/MS analysis of natural Mal d 1 protein in ripe Granny Smith apple showed one abundant isoallergen (Mal d 1.02) and a small proportion of an unidentified isoallergen (Helsper et al 2002) which matched to Mal d 1.06A (Gao et al. 2005b). Both *Mal d 1.02* and *Mal d 1.06A* are located on LG 16. In addition, recent studies demonstrated a significant difference in allergenicity (by the skin

prick test, SPT) among apple cultivars (Bolhaar et al. 2004b). In particular, skin prick test as well as double-blind placebo-controlled food challenges (DBPCFC) showed that a new cultivar Santana (Plant Research International, Wageningen) is low-allergenic, while one of its grandparents, Golden Delicious, is high-allergenic (Bolhaar et al. 2004b). Therefore we wanted to find out whether *Mal d 1* genes have been changed in Santana compared to Golden Delicious.

The goal of this study was to get an impression of the allelic diversity of seven intron-containing *Mal d 1* genes and to assess which of these genes are associated with the low-allergenicity. To achieve this goal, we undertook three steps: (1) assessment of the allelic diversity of seven intron-containing *Mal d 1* genes (some with known expression in the fruit) among ten cultivars by sequencing; (2) development of sequence specific markers to genotype additional cultivars; (3) association of the *Mal d 1* genotypes (haplotype constitution) with allergenicity phenotypes.

Materials and Methods

Cultivars for cloning and sequencing

Eight cultivars were used for cloning and sequencing of *Mal d 1* isoallergen genes, they are Golden Delicious (GD) Priscilla (PS), Ingrid Marie (IM), Cox (CO), Jonathan (JO), Red Delicious (RD), Fuji (FJ) and Discovery (DS). They were chosen for several reasons. (1) GD, IM and PS are members in a pedigree from GD to Santana (ST) =[PS x Elstar (GD x IM)]. (2) GD, JO, CO and RD are founders used in many breeding programmes world wide, and RD, GD and FJ are the top five cultivars in apple production (world apple review). (3) JO and DS are parents of two mapping populations. In total, ten cultivars were under allelic diversity investigation including the two parental cultivars Prima (PM) and Fiesta (FS) used in our previous mapping (Gao et al. 2005b). These ten cultivars were also used for *Mal d 3* allelic diversity (Gao et al. 2005a).

Genomic cloning and sequencing of seven *Mal d 1* genes

Seven intron-containing *Mal d 1* genes on the three linkage groups were chosen to clone and sequence: *Mal d 1.05* on LG 6; *Mal d 1.01* located on LG 13; and *Mal d 1.02*, *Mal d 1.04*, *Mal d 1.06A*, *Mal d 1.06B*, *Mal d 1.06C* all located on LG 16. There are three reasons to chose

these seven intron-containing genes: (1) they are very well defined genes at genomic level and cover three linkage groups and have a longer genomic sequences, while the intronless genes cover only two linkage groups; (2) *Mal d 1.01* and *Mal d 1.02* have been investigated for diversity in cDNA among seven cultivars (Son et al. 1999); (3) *Mal d 1.02* and *Mal d 1.06A* are surely expressed with protein evidence and both are located on LG 16 as described above. Six primer pairs were used for cloning, of which those for *Mal d 1.04* and *Mal d 1.05* were newly designed (Table 1), while the others were reported by Gao et al. (2005b). The PCR amplification and sequencing procedures were described previously (Gao et al. 2005 a, b).

Table 1 Cloning primers and PCR conditions

Gene	Primers[a]	Pfu Tm/cycles[b]	Taq Tm/cycles[b]	Reference sequences
Mal d 1.01	For: ATCTCCAACACAATACTCTCAAC Rev: AAAGCCACACAACCTTCGAC	58/25	60/2	AY789236
Mal d 1.02	For: CATCCTTGGTAGTTGCTTTC Rev: ACCATAGAAACATATTAATTTAGT	52/25	54/2	AY789239
Mal d 1.04	For: CGTAGTTGGACAAGTGTCTTAGT Rev: AGGGTAACACACAAATTACATG	58/30	60/2	AY789242
Mal d 1.05	For: AGTTCATCATGGGTGTTTTC Rev: GGTAACACACAAATTACAAATATGC	53/30	55/2	AY789245
Mal d 1.06A-C	For: CATGGGTGTCCTCACATACGAAAC Rev: TTAGTTGTAGGCATCAGGATTG	55/25	57/2	AY789248
Mal d 1.06C	For: ATGGGTGTCCTCACATACGAAACT Rev: TTAGTTGTAGGCATCAGGATTGGCCACAAGGTG	62/30	64/2	AY789255

[a] Primers for Mal d 1.04 and Mal d 1.05 are new, others have been adopted from a previous study of Gao et al. 2005b

[b] Annealing temperature(Tm, ^{o}C) and cycle numbers

Sequence analysis

All the genomic DNA sequences from ten cultivars were analyzed to identify the consensus allele and single nucleotide polymorphisms (SNPs) using the Seqman program (DNAstar, Madison, WI). The coding sequences were deduced and translated into amino acid sequences. New Mal d 1 sequences were named according to Gao et al. (2005a, b). Multiple amino acid sequence alignments were performed with the GeneDoc program (www.psc.edu/biomed/genedoc).

Sequence-specific markers to genotype other cultivars

Sequence-specific SNAP markers were developed based on *Mal d 1.01, -02, -04, -05, -06A* and *-06B* sequences from the cultivars GD, IM and PS, and then apply to the unsequenced cultivars, Elstar and Santana in the pedigree. The primer designing and testing procedures

were described previously (Gao et al. 2005a, b). A Mal d 1.06A fluorescent labelling SSR marker was redesigned according to the new sequences. PCR reaction mixture for this SSR marker consisted of 2 µl 10x buffer, 1.2 µl $MgCl_2$ (25 mM), 0.4 µl dNTPs (10 mM), 1µl of each primer (2 µM) ,0.06 µl Taq (5 U/µl) and 1µl gDNA 10 (ng/µl) in a total volume of 20 µl. After an initial denaturation at 94°C for 2.5 min, the amplification was carried out for 34 cycles at 94°C for 30 s, 60 °C for 30 s and 72°C for 1 min, and a final extension at 72°C for 5 min. PCR products were analysed on an ABI 377 (Applied Biosystem, Foster City, Calif.).

Two Mal d 1.02 SNAP markers and one Mal d 1.06A-ssr marker were used to screen 47 cultivars and selections (including the sequenced ten cultivars). For some of these cultivars, skin prick test responses were available; others were included to follow the alleles over generations or for the interest of diversity.

Assessment of Mal d 1 allergenicity of apple cultivars by skin prick test (SPT)

Four skin prick test experiments were conducted at the University Medical Center Utrecht (UMCU) to assess the allergenicity of different cultivars. Some data were already presented in the thesis of Bolhaar (2004b). Six adult patients were recruited from the outpatient clinic of the department of Dermatology/Allergology of the UMCU. They all had birch pollinosis manifesting with rhinoconjunctivitis during the birch pollen season (April and May), as well as a positive SPT to fresh apple of at least half the diameter of the positive histamine control. All patients had a typical history of apple allergy, with oral allergy syndrome (OAS) symptoms like itching and mild swelling of the mouth, throat and sometimes rhinoconjuctivitis after eating an apple. These experiments were reviewed and approved by the Ethics Committee of the UMCU. All patients gave written informed consent before enrolment in the study.

Fourteen apple cultivars were used to screen, which were grown at the research orchard of Plant Research International, Wageningen, the Netherlands. Fruits were harvested at their usual degree of ripeness for consumption. SPT were performed on the flexor surface of the forearm using the prick-to-prick-technique according to Dreborg (1983; 1993). Histamine dihydrochloride (10mg/ml) was used as a negative control, and the glycerol diluent of the SPT-extracts were used as negative controls (ALK-ABELLO, Nieuwegein, The Netherlands). The wheal reaction was marked and transferred with transparent adhesive tape to a record sheet. The skin wheal areas were measured by a computer scanning (Poulsen et al. 1993). SPT

responses for each cultivar were standardized by dividing the original wheal area of the prick by that obtained for the reference cultivar for high allergenicity Golden Delicious histamine control and multiplied by 100. A single apple from each cultivar was used in the first two experiments, while the third experiment using two apples and two pricks per apple, and in the fourth using five apples and four pricks per apple (see Table 5). We used the responses of the low allergic patients because preliminary experiments indicate that other allergens played a role for the highly allergic patients (Van de Weg, unpublished). The final ranking results were obtained by averaging the responses from four experiments.

Results

Sequence analysis of the diversity of *Mal d 1* genes at DNA and deduced protein level

We obtained 117 gDNA sequences for seven *Mal d 1* genes from ten cultivars. The numbers of single nucleotide polymorphisms (SNPs) present in both the coding and intron regions are summarized in Table 2. The majority of SNPs were located in the coding region. These sequences were further classified according to polymorphisms in the deduced amino acid sequences (variants); coding sequence (silent mutation); and anywhere in the gDNA (Table 3). For the identification and denotation of variants, our amino acid sequences were compared with those of published isoallergens and variants.

Table 2 Number of SNPs in the genomic sequences of seven *Mal d 1* genes from ten cultivars

Gene [a]	Coding	Intron	Total
Mal d 1.01	4	3	7
Mal d 1.02	10	4	14
Mal d 1.04	13	4	17
Mal d 1.05	11	2	13
Mal d 1.06A	6	3	9
Mal d 1.06B	10	2	12
Mal d 1.06C	11	1	12

[a] Genomic gene lengths ranged from 608 to 651nt

Mal d 1.01 (on LG 13) showed to be highly conserved having only seven SNPs at the gDNA level (Table 2) and just two variants at the protein level matching the previously described variants Mal d 1.0105 and Mal d 1.0109 (Table 3, Fig. 1). Mal d 1.0105 was encoded by four different coding sequences and present in all 10 cultivars. Mal d 1.0109 was unique to Red Delicious. These two variants differed in only a single amino acid at position

135 (valine (V) or alanine (A) (Fig. 1). It is remarkable that both sequences of Priscilla were different from those of Red Delicious. Since Starking Delicious (almost identical to Red Delicious) is a parent of Priscilla, at least one common sequence was expected. It is therefore likely, that a single nucleotide mutation at position 404 occurred in this gene.

Mal d 1.02 (on LG 16) showed a similar high level of conservation at the protein level as *Mal d 1.01*. The first variant Mal d 1.0201 occurred in all 10 cultivars and was encoded by seven different coding sequences. The second variant, Mal d 1.0209, was found in just one cultivar (Discovery). These two variants differ in a single amino acid change at position 56 to asparagine (N) or lysine (K) (Fig. 1).

Mal d 1.04 (on LG 16) was more variable, having four functional variants, which differed in one or two amino acids at positions 94D/V, 136A/V and 156P/L (Fig. 1), and two different pseudo alleles (*Mal d 1.04ps1* and *Mal d 1.04ps2*). Variant Mal d.10404 and the two pseudo-alleles were predominantly present in most cultivars. The presence of two pseudo-alleles appeared not to be a rare phenomenon in apple cultivars, because seven cultivars had such alleles, and particularly, both *Mal d 1.04* alleles of Priscilla and Fuji were not functional. Compared to previously reported Mal d 1.04 variants (Z72426–Z72428), there are two and three different amino acids close to the N terminus and the C terminus, respectively (Fig. 1). The deviations in former sequences could be attributed to the use of degenerate primers (Hoffmann-Sommergruber, et al. 1997), which was determined by genome walking (Gao et al. 2005b).

In *Mal d 1.05* (on LG 6), only four genomic alleles were found among ten cultivars, which encoded three variants that differed in two amino acids (Fig. 1). Variant Mal d 1.0501 was predominantly present in all cultivars but Fuji (Table 3). The other variants occurred less frequently.

With regard to three slightly different *Mal d 1.06* genes (all on LG 16), we found three variants for Mal d 1.06A, five for Mal d 1.06B and six for Mal d 1.06C. Mal d 1.06A was less variable at the protein level, with three variants that differed in one or two amino acids at position 13 and 135 (Fig. 1). This gene specifically has a simple CA repeat of 6 to 16 times in the intron. No direct relationship exists between repeat number and variant identity (Table 3). Consequently, the parental origin of marker alleles should be taken into account when marker scores are linked to variants of Mal d 1.06 or to haplotypes (see below). Mal d 1.06B was quite variable with five functional variants. The fifth variant was unique to Discovery. Mal d 1.06C was the most variable isoallergen in our set of ten cultivars with six variants, although

its allelic variation was not fully disclosed because only a limited number of clones were obtained.

Table 3 Classification of the sequences of seven *Mal d 1* allergen genes from 10 apple cultivars

Isoallergen (linkage)	Variant[a]	Coding DNA[b]	gDNA[c]	Cultivar[d]										GenBank Accns.
				GD	PS	IM	CO	JO	RD	FJ	DS	PM	FS	
Mal d 1.05	01	1	1	+	+	+ +	+ +	+	+		+	+ +	+	AY789245, AY789246, AY827676–AY827682
(LG 6)	02	1	2								+		+	AY789247, AY827683
	02	2	3	+	+									AY827684, AY827685
	03	1	4					+	+	+ +				AY827686~AY827688
Mal d 1.01	05	1	1									+		AY789236
(LG 13)	05	1	2	+				+		+				AY827639~AY827641
	05	2	3	+	+	+ +	+ +	+			+ +		+ +	AY789238, AY827633–AY827638
	05	3	4		+							+		AY789237, AY827642
	05	4	5						+	+				AY827643, AY827644
	09	1	6						+					AY827645
Mal d 1.02	01	1	1	#			#	#				#	♦♦	AY789240, AY789241, AY827646–AY827648
(LG 16)	01	2	2									*		AY789239
	01	3	3		♦♦	*	*	*	*	*				AY827649–AY827654
	01	4	4	*						#				AY827655, AY827656
	01	5	5			#								AY827657
	01	6	6						#					AY827658
	01	7	7								*			AY827659
	09	1	8								*			AY827660
Mal d 1.04	04	1	1									*		AY789242
(LG 16)	04	2	2	#			#	#	#			#	♦♦	AY789243, AY789244, AY827661–AY827664
	05	1	3								*			AY827665
	06	1	4			#								AY827666
	07	1	5								*			AY827667
	ps1	1	6		♦♦			*	*	*				AY827668–AY827671
	ps2	1	7	*		*	*			#				AY827672–AY827675
Mal d1.06A	01	1	1-$(CA)_{10}$	#			#	#					♦♦	AY789249, AY827689–AY827691
(LG 16)	01	1	2-$(CA)_{11}$									#		AY789248
	02	1	3-$(CA)_{16}$									*		AY789250
	02	2	4-$(CA)_{6}$		♦♦	*	*	*	*	*				AY827692–AY827697
	02	3	5-$(CA)_{10}$			#			#					AY827698, AY827699
	02	4	6-$(CA)_{7}$								*			AY827700
	03	1	7-$(CA)_{7}$	*						#				AY827701–AY827702
	03	2	8-$(CA)_{7}$								*			AY827703
Mal d 1.06B	01	1	1	#	#		#	#					♦♦	AY789251, AY827704–AY827707
(LG 16)	02	1	2	*	*			*	*	♦♦		#		AY789252, AY827708–AY82712
	03	1	3						#			*		AY789253, AY827713
	03	2	4			#								AY827714
	04	1	5			*	*							AY827715, AY827716
	05	1	6								*			AY827717
	05	2	7								*			AY827718

Table 3 continued

Isoallergen (linkage)	Variant[a]	Coding DNA[b]	gDNA[c]	Cultivar[d]										GenBank Accns.
				GD	PS	IM	CO	JO	RD	FJ	DS	PM	FS	
Mal d1.06C	01	1	1	#	●		●	●					♦♦	AY789254, AY827719–AY827722
(LG 16)	02	1	2	*					*	●		#		AY789255, AY827723–AY827725
	03	1	3									*		AY789256
	04	1	4								●			AY827726
	05	1	5						#					AY827727
	06	1	6			●								AY827728

[a] Variants refer to polymorphisms at the amino acid level. Variants numbers of Mal d 1.01, Mal d 1.02 and Mal d 1.04 are following the existing names in the official allergen list (www.allergen.org), those for Mal d 1.05 and Mal d 1.06 are adopted from the nomenclature used in Gao et al. (2005b). *ps* refers to pseudogene allele

[b] Number refers to different coding sequences identified for the same variant

[c] Refers to different gDNA sequences. In Mal d 1.06A, the number after CA indicates the repeat times in the intron

[d] Cultivars in short: **PM,** Prima; **FS,** Fiesta; **GD,** Golden Delicious; **PS,** Priscilla; **IM,** Ingrid Marie; **JO,** Jonathan; **CO,** Cox; **RD,** Red Delicious; **FJ,** Fuji; **DS,** Discovery.

+ is used in the Mal d 1.01 and Mal d 1.05 genes to indicate the presence of an allele in heterozygous (+) or homozygous (+ +). For those genes on LG 16, ♦♦ refers to homozygous; # and * to heterozygous presence, and the same symbols in the same column refer to the alleles in coupling phase; ● indicate presence without knowing their heterozygous or homozygous condition in Mal d 1.06C section.

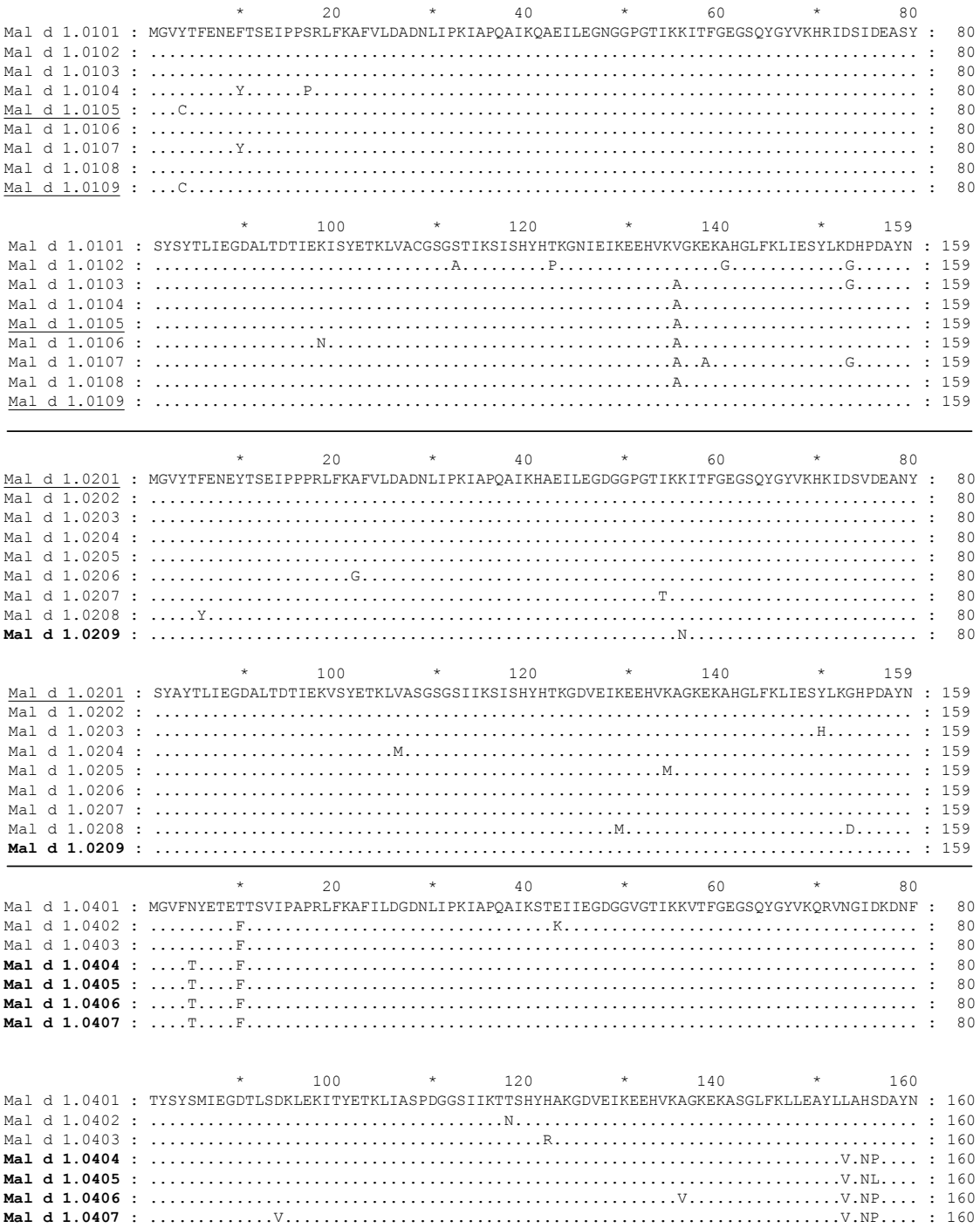

Fig. 1 Amino acid polymorphisms of seven *Mal d 1* genes (first part)

For each gene, the first variant is the reference to compare. The underlined variants were identified in the literature and confirmed by us, the variants in bold type are new.

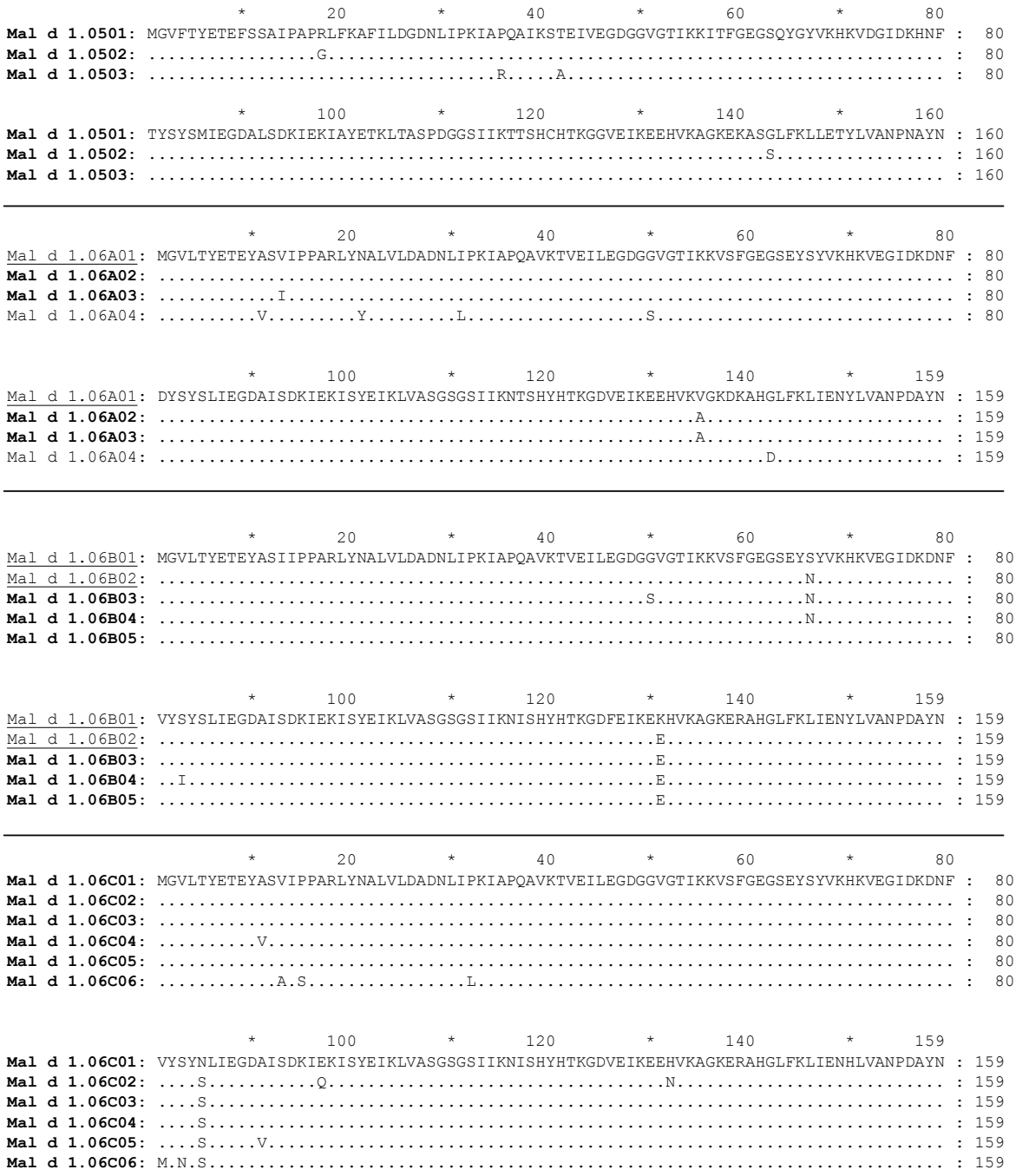

Fig. 1 Amino acid polymorphisms of seven *Mal d 1* genes (second part)

For each gene, the first variant is the reference to compare. The underlined variants were identified in the literature and confirmed by us, the variants in bold type are new. Reference sequences: Mal d 1.06A01, AY428580; Mal d 1.06A04, AY428581; Mal d 1.06B01, AY428582; Mal d 1.06B02, AY428583

Linkage phase and haplotypes

Linkage phases of the alleles of the examined genes of LG 16 could be deduced for Prima, Fiesta and Jonathan by the co-segregation patterns of sequence specific markers in the Prima x Fiesta and Jonathan x Prima progenies, in which the *Mal d 1* genes were mapped (Gao et al. 2005b). The linkage phase in Cox, Ingrid Marie, Fuji, Priscilla and Delicious could be deduced from direct pedigree relationships among these cultivars. Fiesta is a descendant of Cox x Idared (= Jonathan x Wagner Apfel); Ingrid Marie derived from an open pollination of Cox; Red Delicious is a parent of Priscilla (= Starking Delicious x PRI610-2) and Fuji (= Ralls Janet x Delicious), with Red Delicious and Starking Delicious being mutants of Delicious. The linkage phase for Golden Delicious was determined by marker test in a pedigree (see below). Table 4 presents nine haplotypes derived from these linkage phase analyses. *Mal d 1.06C* gene and Discovery were not included in these haplotypes, due to incomplete data on sequences and linkage phase.

Table 4 Haplotypes of four *Mal d 1* genes on Linkage Group 16 deduced from 9 cultivars

Haplotype	Mal d 1.02	Mal d 1.04	Mal d 1.06A	Mal d 1.06B	Presence	Mal d 1.06A-SSR marker
1	01 01	04 02	01 01	01 01	FS, GD, JO, CO	162
2	01 04	ps2	03 01	02 01	GD, FJ	156
3	01 03	ps2	02 02	04 01	IM, CO	154
4	01 03	ps1	02 02	02 01	PS, JO, RD, FJ	154
5	01 03	ps1	02 02	01 01	PS	154
6	01 05	06 01	02 03	03 02	IM	162
7	01 06	04 02	02 03	03 01	RD	162
8	01 01	04 02	01 01	02 01	PM	164
9	01 02	04 01	02 01	03 01	PM	174

Note: the first two numerals refer to the variant, and the last two numerals refer to silent mutation in the coding sequence (see Table 3)

Allergenicity scores of apple cultivars by SPT

SPT responses were available from 14 cultivars combining four independent experiments and were standardized by expressing the response of a single cultivar relative to the highly allergenic cultivar Golden Delicious in each experiment. The ranking of these cultivars is given in Table 5. Fiesta, Delblush, Pinova and Golden Delicious were within the high allergenic group (83–100%). Priscilla and Santana showed low SPT responses of 30–35% of that of Golden Delicious. Nine cultivars are intermediate (48–72%). Santana was already

identified as low-allergenic in comparison to Golden Delicious (Bolhaar et al., 2004). Of the two parents of the low allergenic cultivar Santana, the allergenicity of Elstar is intermediate, and Priscilla is as low as Santana

Table 5 Skin Prick Test (SPT) responses of 14 cultivars from four independent experiments

Cultivar	**SPT responses** [a]				
	Experiment [b]				
	I	**II**	**III**	**IV**	**Average**
Priscilla (PS)		30			**30**
Santana (ST)	33	30	41		**35**
Jonathan (JO)		48			**48**
Ecolette	39			62	**51**
Prima (PM)		61			**61**
Elstar (ES)	67		55		**61**
Fuji (FJ)	70		48	64	**61**
Gala	65		62		**64**
Elise	67				**67**
Bellida	72				**72**
Fiesta (FS)	67	99			**83**
Delblush	87				**87**
Pinova	89				**89**
Golden Delicious (GD)	100	100	100	100	**100**

[a] Responses are expressed in percentage relative to the response of Golden Delicious

[b] Data of the four experiments are based on 3, 5, 2 and 6 patients respectively with 1, 1, 2, and 5 apples per patient tested with 1, 1, 2, and 4 pricks per apple, respectively

Allelic diversity related to allergenicity

Fiesta and Golden Delicious are known as high-allergenic compared to the lowest one, Priscilla. These differences will not be related to the Mal d 1.05 alleles because Golden Delicious and Priscilla are both heterozygous for the same *Mal d 1.05* alleles (Table 3). The allelic presence pattern of *Mal d 1.01* was also not evident considering the small differences in amino acid and DNA sequence. *Mal d 1.01* and *Mal d 1.05* are located on LG 6 and LG 13, respectively. Inspecting the allelic variation of the *Mal d 1* genes located on LG 16, we found obvious differences between Fiesta/Golden Delicious and Priscilla. Fiesta and Golden Delicious both share one allele for each *Mal d 1* gene on LG 16. This preliminary allelic analysis suggests the *Mal d 1* genes on LG 16 to be correlated to allergenicity. This hypothesis will be elaborated below.

Pedigree analysis of the *Mal d 1* genes on three linkage groups for Santana

To confirm the hypothesis above, we carried out a pedigree analysis. Based on the *Mal d 1* genomic sequences of Golden Delicious, Ingrid Marie and Priscilla, 9 SNAP markers representing *Mal d 1.05*, *Mal d 1.01*, *Mal d 1.02*, *Mal d 1.04*, *Mal d 1.06B* and one Mal d 1.06A-SSR marker (Table 6) were developed for tracking the *Mal d 1* allelic constitution on

Table 6 Sequence-specific markers used for the pedigree analysis from Golden Delicious to Santana

Primer name	Primers 5'–3'	SNPs detected [a]	Product (nt)	Tm[b]	Cycles[c]
Mal d 1.01-3-GD/IM/PS	For-TGAAGCACAGGATTGAC**G**C**A**	390A	346	56	35
	Rev-CCACACAACCTTCGACT**C**A				
Mal d 1.01-2-GD	For-AGCTGAAATCCTTGAA**C**GAA	528T	426	55–53	10–30
	Rev-CAATGTTTCCCTTGGTG**A**G**A**				
Mal d 1.01-4-PS	For-TGAAGCACAGGATTGAC**A**C**G**	390G	177	57–55	10–30
	Rev-CAATGTTTCCCTTGGTGT**C**G	528C			
Mal d 1.0201-1-GD1/ IM2	For-TCACTTTTGGTGAAGGTC**T**G	402G	252	57	35
	Rev-AGGCGTATGAGTAGTTT**A**C**C**				
Mal d 1.0201-4-GD	For-ATTATTTAGATGGTTTCGCT**A**T	227T	439	55	35
	Rev-TATGCGTCGGGGTGTCCCT**A**G	624C			
Mal d 1.0201-3-PS/IM	For-GCCCTGGAACCATCAA**G**T**A**G	165G, 168G	213	61	35
	Rev-GGCTGCCTGTGAGACAA**C**T**G**	342C			
Mal d 1.04-6/7-GD2/IM	For-CATCCCGAAGATTGCTCC**G**T**T**	109T	395	57	35
	Rev-CCTTAGCATGGTAGTGGC**G**A**G**	464C			
Mal d 1.04-4-IM	For-CAAGGAAGAGCATGTTAAA**G**T	516A, 518T	136	60	35
	Rev-AGGGTAACACACAAATTACATG				
Mal d 1.06A-SSR	For-GGTGAAGGTTAGTTTAATTTCCACA	(CA) SSR		60	30
	Rev-***GT***TTCACATAGCTGTATTCACTCCCT				
Mal d 1.06B-2	For-CCACCATTTTCTCCATTAACTT**C**A	221A	360	62–60	5–35
	Rev-GCCTTAACATGCTCTTCCTTGATT				

[a] Position refers to the genomic sequence counted from the ATG starting codon

[b,c] PCR annealing temperature (Tm) and number of cycles, respectively. In case of two values, a touch-down PCR was performed in two steps: the first number refers to the Tm and the number of cycles, respectively, of the first step.

three linkage groups (LG 6, 13 and 16) of two unsequenced cultivars, Elstar and Santana. According to these marker tests (only *Mal d 1.06A* data are shown as representatives in Table 7), we illustrated schematically the *Mal d 1* gene flow on each linkage group in Fig. 2. There is just a single gene on LG 6, so it is easy to see that *Mal d 1.05* alleles in Santana are not changed in comparison to Golden Delicious. Similarly, *Mal d 1* segments on LG 13 are also

not changed at least with regard to *Mal d 1.01*. However, multiple marker tests of four *Mal d 1* genes (*Mal d 1.02, -04, -06A, 06B*) on LG 16 on Elstar and Santana revealed that Santana contains completely different *Mal d 1* haplotypes than Golden Delicious: with one from Ingrid Marie and the other from Priscilla (Fig. 2). Because Santana has a much lower allergenicity than Golden Delicious, this pedigree analysis furthermore suggests that the *Mal d 1* cluster on LG 16 is involved in allergenicity. Compared to Golden Delicious, Santana had the same deduced Mal d 1.0201 variant but encoded by the homozygous allele of Mal d 1.0201.03. Both Mal d 1.04 alleles in Santana were pseudo rather than the one in Golden Delicious. In addition, Santana had a new deduced Mal d 1.06A02 variant instead of Mal d 1.06A01 and -03 of Golden Delicious, while one of the two Mal d 1.06B variants has been changed in Santana (see the variants of different genes in Table 4).

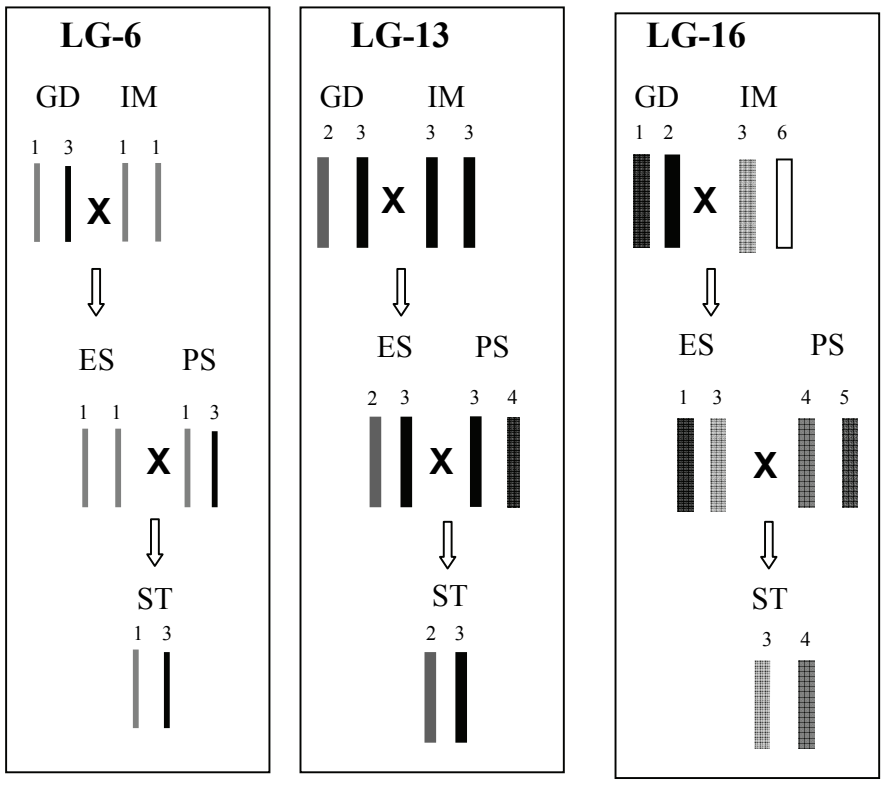

Fig. 2 Schematic illustration of Mal d 1 segment changes by representative genes on three linkage groups in a breeding pedigree. Only the *Mal d 1* genes on two LG 16 homologous chromosomes of Golden Delicious were completely replaced by one of the IM and PS in the new cultivar Santana (ST) Apple cultivars: GD, Golden Delicious; IM, Ingrid Marie; ES, Elstar; PS, Priscilla; ST, Santana. The number above the two homologous LG 6 and 13 refer to the gDNA allele number in Table 3; those in the LG16 section refer to the haplotypes in Table 4

Genotype-allergenicity (SPT) associations for LG 16 of additional cultivars

The validity of the conclusion above was further tested in a set of 14 cultivars. The haplotypes for five of these cultivars were already known (Table 4), nine cultivars without sequence information were genotyped as haplotypes by the SSR marker *Mal d 1.06A* using their parental origin and tests of two *Mal d 1.02* markers. The resulting haplotype constitution and their average SPT-responses are presented in Table 7. A clear association was observed between the high SPT responses and combination with haplotypes 1 and 2 or double 1. For cultivars with the lowest responses, Santana possesses haplotypes 3 and 4, and Priscilla has haplotypes 4 and 5 (Table 7). Six of the seven cultivars with intermediate responses had

Table 7 Haplotype constitutes of Mal d 1 genes on linkage group 16 and their association to allergenicity

Cultivar	**Pedigree**	***Mal d 1.06A* SSR marker (nt) and representing haplotypes** [a]					**SPT response** [b]
		154	156	162	164	174	
Priscilla	Starking Delicious x PRI610-2	**4-RD 5-PS**					**30**
Santana	Elstar x Priscilla	**3-IM 4-PS**					**35**
Jonathan		4-JO		1-JO			48
Ecolette	Elstar x Prima	3-IM			8-PM		51
Prima	NJ123249 x 14-510 (26829-2-2 x Golden Delicious)				8-PM	9-PM	61
Elstar	Golden Delicious x Ingrid Marie	3-IM		1-GD			61
Fuji	Ralls Janet x Delicious	4-RD	2-RJ				61
Gala	Kidd's Orange Red x Golden Delicious	3-CO		1-GD			64
Elise	Septer x Cox	3-CO		1-JO/GD			67
Bellida	Idared x Elstar			1-JO 1-GD			72
Fiesta	Cox x Idared			**1-CO 1-JO**			**83**
Delblush	Golden Delicious x Blushing Golden		**2-GD**	**1-GD**			**87**
Pinova	Clivia x Golden Delicious		**2-GD**	**1-CO**			**89**
Golden Delicious			**2-GD**	**1-GD**			**100**
Cox		3-CO		1-CO			
Ingrid Marie	Cox x ?	3-CO		6-IM			
Red Delicious		4-RD		7-RD			
Kidd's Orange Red	Delicious x Cox	3-CO		7-RD			

[a] The numbers refer to the haplotypes defined in Table 4

[b] SPT, skin prick test averaged percentage as shown in Table 5

haplotype combinations from those present in the lowest group and in the high group. Three haplotypes responsible for the low allergenic trait have a common Mal d 1.06A-SSR-154 marker and a SNAP marker Mal d 1.02-3PS/IM. When referring to the individual gene's alleles in Table 4, it is evident that various genomic Mal d 1.02 alleles all encoded the same protein variant (01); both Mal d 1.04 alleles are pseudo in the low haplotypes (3–5), while the high ones have at least one functional allele; in contrast, an obvious variant change occurred in Mal d 1.06A, variant Mal d 1.06A02 was associated with low responses, while variants -01 and -03 are associated with high responses. From this result, a useful genetic marker for high- or low-allergenicity became apparent.

Discussion

Skin prick tests on selected patients demonstrated the difference in Mal d 1-related allergenicity among cultivars. Cloning and sequencing of seven Mal d 1 genes (on three linkage groups) from ten cultivars revealed different allelic variation with regard to genomic alleles and deduced proteins. *Mal d 1.01* and *Mal d 1.02* were diverse at the gDNA level but conserved at the protein level: each had a second variant that differed in one amino acid in a single cultivar. *Mal d 1.04*, *Mal d 1.05* and three *Mal d 1.06* genes were variable at the protein level. Association between the allelic constitution (haplotype) and SPT values indicated that *Mal d 1* genes on LG 16 are related to allergenicity, and that Mal d 1.06A-ssr-154 marker correlated to low allergenicity.

Validity of sequences and its consequences for the existing true variants

Over 30 DNA sequences of *Mal d 1.01* and *Mal d 1.02* from other researchers are known in public databases with dozens of SNPs (Tables 8 and 9). Some sequences are from the same cultivars as we tested. Four SNPs in *Mal d 1.01* were confirmed by our sequences (Table 8). There were five *Mal d 1.01* sequences from Golden Delicious (one of them was actually from its seedling); only AF124830 was identical to one of our two sequences from Golden Delicious. AF126402 had a SNP at position 11A that differed at position of 11G in AF124830, which was due to its cloning primer used that introduced this base. For the same reason, other sequences had 11A, too. Our cloning primers were designed on sequences in the untranslated region and so avoided this error. AY026910 was from the Golden Delicious seedling (Ziadi et al. 2001) which is identical to one of the alleles of Red Delicious. It is

perhaps the result of pollination from one of the Delicious serial cultivars. We were not able to obtain the same sequence of Z48969 (Vanek-Krebitze et al. 1995) that was classified as Mal d 1c (Son et al. 1999), which is quite similar to *Mal d 1.01*, while the polymorphic SNPs resembled three other *Mal d 1* genes. This may be due to an artefact from strand switching during PCR. That is why also other researchers could not get completely identical sequences (Son et al. 1999; Beuning, 2004). The first 84 nt of AF124829 and AF124832 match to *Mal d 1.02* (typical SNPs are shown at the bottom of Table 8). Therefore, the actual number of *Mal d1.01* variants among apple cultivars is not as high as nine, for at least five of them can be considered as false. Similarly, we identified several suspect SNPs by analysis of all the published *Mal d 1.02* sequences (Table 9). This implies that there are only a few true variants in this gene. Mal d 1.06 is a new isoallergen group consisting of three members, four published sequences from Gala (GA) belong to two of these members (Fig. 1), three of these are perfectly matching to ours, one (Mal d 1.06A04) is different.

Mal d 1 genes on different linkage groups and their relevance

Linkage mapping of the *Mal d 1* gene family has shown that its members are located on three linkage groups (Gao et al. 2005b). Pedigree analysis tracing the genes in the low allergenic cultivar Santana led to a conclusion that at least some genes on LG 16 showed to be related to the Mal d 1-based allergenicity. Until now, mRNA expression evidence for five genes was observed in mature fruit (Beuning et al. 2004), representing two genes (*Mal d 1.01* and *-03E*) on LG 13 and three genes (*Mal d 1.02*, *-06A* and *-06B*) on LG 16. From the number of the Mal d 1 ESTs isolated from the cDNA library in ripe Gala fruit, *Mal d 1.01* and *Mal d 1.0*2 were expressed as mRNA in both skin and cortex, while *Mal d 1.06A*, *-06B*, *-03E* were present only in the skin of apple fruit. Although *Mal d 1.01* was highly expressed as mRNA in ripe fruit (Puehringer et al. 2003; Beuning et al. 2004), actual presence of protein has not been shown definitely. We assume that Mal d 1.01 will account for a very small proportion of total Mal d 1 if it is really expressed. Nine intronless genes were identified through sequencing and linkage mapping (Gao et al. 2005b), but only *Mal d 1.03E* (Mal d 1i) was found in the fruit skin (Beuning et al. 2004). Puehringer et al. (2003) also detected small fractions of *Mal d 1.03* genes by RT-PCR, but did not find *Mal d 1.04* mRNA in the fruit. At this moment, two

Table 8 SNPs present in the *Mal d 1.01* sequences found in the databases

Sequence(Allele) [a]	Cultivars [b]	Nucleotide position in coding sequences [c]																											
		9	11	15	18	21	29	49	75	84	222	294	333	334	360	364	404	405	408	413	419	420	435	453	456	458	465	468	471
Consensus		C	G	A	T	G	T	T	T	C	G	A	T	T	C	A	C	T	A	A	C	C	G	T	G	A	C	C	A
Mal d 1.0101(X83672)	GS		*A*														T												
Mal d 1.0102 (Z48969)	GD		*A*			*A*								*G*		*C*	T	*C*			*G*	*T*		*C*		*G*	*G*	*T*	*C*
Mal d 1.0103 (AF124823)	JB		*A*								A						T						A		*A*	*G*			
Mal d 1.0104 (AF124829)	JG		*A*				*A*	*C*	*C*	*T*					T														
Mal d 1.0105.01 (AF124830, AY428579)	GD, GA, **PM, GD, JO, FJ**														**T**														
Mal d 1.0105.02	**PM, FS, GD, PS, IM, JO CO, DS**										**A**												**A**						
Mal d 1.0105.03	**PM, PS**																												
Mal d 1.0105.04	**RD, FJ**										**A**																		
Mal d 1.0106 (AF124831)	GL											C																	
Mal d 1.0107 (AF124832)	GA		*A*				*A*		*C*	*T*	A								*C*	*C*		*T*		*C*					
Mal d 1.0108 (AF126402)	GD		*A*												T														
Mal d 1.0109 (AY026910)	GD-seedling, **RD**																**T**												
Mal d 1.02-CONS [d]			A				A	C	C	T		G				A			C			T				G			

[a] Aleles in bold are confirmed by our own sequences. The same reference sequences are given in brackets

[b] Additional cultivars in short apart from those in Table 2: GA-Gala, JB-Jamba, GL-Gloster, GD-seedling: plant material is from the seed of a GD apple, so it is not a cloned cultivar. Cultivar names in bold are from our source of sequences shown in Table 3

[c] Position refers to the coding sequence. SNP nucleotides in bold are our observations. The SNPs in italic and underlined were sure errors after cross checking.

[d] *Mal d 1.02* consensus SNPs compared with *Mal d 1.01*

Table 9 SNPs in the *Mal d 1.02* sequences found in the data bases

Sequence (allele) [a]	Cultivars [b]	Nucleotide position in coding sequences [c]																												
		9	18	19	23	65	161	165	168	180	186	222	231	273	306	313	316	369	386	387	401	414	420	448	453	458	459
Mal d 1.02 consensus		C	T	T	C	C	T	G	G	T	C	G	G	T	G	T	G	G	A	G	A	G	T	T	T	G	C
Mal d 1.0201.01 (L42952, AF074721, AY428578)	GD, GA, IR, **PM, FS, GD, JO, CO**							**A**			**T**																
Mal d 1.0201.02	**PM**											**A**	**A**												**C**		
Mal d 1.0201.03	**PS, IM, JO CO, RD, FJ**												**A**												**C**		
Mal d 1.0201.04 (AF124826)	IR, **GD, FJ**																								**C**		
Mal d 1.0201.05	**IM**															**C**											
Mal d 1.0201.06	**RD**															**C**		**A**									
Mal d 1.0201.07	**DS**																				**A**	**C**					
AF124827	GA				*T*			A			T			C													
AF124828	GS														C	C											
AF124833	GD																		*A*						C		
AF124834	GL												A														
AF124836	GA							A					A														*G*
Mal d 1.0202 (AF124822)	GS							A		G	T																
Mal d 1.0203 (AF124824)	GD							A			T													*C*			
Mal d 1.0204 (AF124825)	GL												A				A								C		
Mal d 1.0205 (AF124835)	JB												A							T					C		
Mal d 1.0206 (AF020542)	MT					G							A		A										C		
Mal d 1.0207 (AY026911)	GD-seedling						C																		C		
Mal d 1.0208 (AJ488060)	CO	*T*	*A*	*C*				A			T								*T*							*A*	
Mal d 1.0209	**DS**								**T**																**C**		

[a] Variant-alleles in bold type are from our own sequences (accession numbers in Table3). The same reference sequences are given in brackets

[b] Additional cultivars in short are additional to those in Table 2: GA-Gala, JB-Jamba, GL-Gloster, MT-McIntosh, IR-Idared, GS-Grany Smith, GD-seedling: plant material was development from the seed of a GD apple, so it is not a cloned cultivar. The bold type cultivars represent our sequences.

[c] Position refers to the coding sequence. SNP nucleotides in rectangular boundary were true, in italic and underlined were errors after cross-checking. Other SNPs are not certain.

isoallergens have been proven to be present in the apple fruit (Helsper et al. 2002; Beuning et al. 2004), the majority of Mal d 1 protein is Mal d 1.02 (Mal d 1b) and a minor part is Mal d 1.06A. Interestingly, both genes are located on linkage group 16. This is consistent with our result from genetic analysis.

Is allergenicity related to qualitative or quantitative differences in Mal d 1

Whether the apple Mal d 1 allergenicity is determined by the quality or quantity of Mal d 1 isoallergens is still a crucial question to answer. Previous studies demonstrated that Mal d 1 protein levels were tenfold different which correlated well with the allergenicity of four apple cultivars (Vieths et al. 1994 and 1995). Recent Mal d 1 content surveys among dozens of

cultivars using ELISA, RIA and inhibition methods also showed considerable differences (Zuidmeer et al. unpublished result), but the lower Mal d 1 content appeared not to be consistently associated with a lower allergenicity as revealed by SPT which seems more reliable and quantitatively reflecting the cultivar's allergenicity (Bolhaar et al. 2004b). It is, therefore, likely that Mal d 1-related allergenicity is affected by means of both qualitative and quantitative factors. Mal d 1.02 has been identified as the dominant protein (Helsper et al. 2002). Most cultivars have the same variant, so its difference in content may be related to the genomic differences at allelic level affecting their expression. In the case of Santana and Priscilla, its homozygous Mal d 1.0201.03 allele has a specific 231A (Table 9). The same 231A was found in Gloster (AF124825, AF124834), Jamba (AF124835), McIntosch (AF020542), and Gala (AF124836) (Table 9). Gloster and Jamba were reported to be lower allergenic than Golden Delicious (Son et al. 1999). It was noted that the variant (Mal d 1.0209) in Discovery, in which amino acid N replaced K at position 57, might affect its binding property to antibodies. This variant was also present in Topaz by a marker test. If the allergenicity is a matter of the variant of a single gene, then *Mal d 1.06A* will be the most outstanding candidate. Three variants were found in Mal d 1.06A. High-allergenic cultivars have two putative variants, Mal d 1.06A01 alone or -01 and -03 together, whereas low-allergenic cultivars have only the -02 variant. In the intermediate-alergenic cultivars, a combination of -01 and -02 was common. Amino acid 13 V/I and 135V/A occurred in these three variants (Fig. 1). Looking to this polymorphic position in the structure model of Mal d 1 (Neudecker et al. 2001), the first mutation is located in the first loop between the β1-strand and the α1-helix, the second is located in the α3-helix structure motif. These three amino acids all are hydrophobic but have different side chains. Recombinants of these three variants might reveal further evidence for the involvement of three amino acids.

Conventional breeding or gene silencing for low-allergenic cultivars

This study strongly suggests the involvement of a few *Mal d 1* genes on LG 16 in Mal d 1 allergenicity. Additional searching of *Mal d 1* candidate genes and genetic analyses will confirm the existence of specific alleles, such as the alleles represented by the Mal d 1.06A-ssr-154 marker, to be associated with low-allergenicity. For example, the cultivar Wijcik McIntosh is also homozygous for Mal d 1.06A-ssr-154 (data not shown), so we expect this cultivar will be low in its Mal d 1 allergenicity. If this holds true in other cultivars and

selections, then this result will facilitate future breeding programmes targeted at low allergenic apples enabling deliberated choice of parental cultivars and genetic markers for selection in young seedlings. At this moment, Santana is a new low-allergenic cultivar with good levels of resistance to scab, mildew and *Nectria* canker. Large-scale production of this cultivar has begun in the Netherlands. In addition, this cultivar may also has lower Mal d 3 content since its two parents Elstar and Priscilla are low (Ana Sancho, personal communication). Santana will, therefore, be a very good parent for further breeding of low-allergenic cultivars. Molecular markers from this research could be used to select favourable alleles in the progeny. Along with conventional breeding, the RNA interference (RNAi) technique to block the expression of the relevant Mal d 1 genes is another alternative to achieve this goal. A very recent research using such an RNAi approach to silence the *Mal d 1* gene has reduced the Mal d 1 allergenicity dramatically in the transformed apple leaves of in-vitro growing plants (Gilissen et al. 2005). It will take some more years to see the effect of this on the fruit and on the whole tree.

Acknowledgement

This research was performed with financial support from the Commission of the European Communities for the projects SAFE QLK1-CT-2000-01394. We acknowledge professors Evert Jacobsen and Richard Visser for critical reading of this manuscript.

Chapter 6

General Discussion

Apple was chosen as a model food for the EU-SAFE project for several reasons. Fruit allergy in general has a high prevalence and apple is one of the most frequently reported allergenic fruits. Apple is an important healthy food product, and therefore avoidance will have a negative impact on the nutritional value of this diet. Different apple cultivars have a variable degree of allergenicity (Hoffmann-Sommergruber et al. 2005). The SAFE project was carried out by a multidisciplinary consortium. It included laboratories with expertise in biomedical research, food chemistry, plant genetics, molecular and structural biology; in addition, clinicians, fruit growers, social scientists, representatives from patient organizations and a major fruit juice producer participated. As plant geneticists, our task in the project was to genetically characterize and map the currently known apple allergens: Mal d 1 (PR-10 protein), Mal d 2 (thaumatin-like protein, TLP, PR-5 protein), Mal d 3 (nsLTP, PR-14 protein) and Mal d 4 (profilin). We started our research with known information in two areas: (1) DNA sequences of four classes of apple allergens in the databases, and (2) Three molecular linkage maps were available to map the allergen genes. By PCR genomic cloning, over 200 new sequences have been produced from the four allergen genes and deposited in GenBank. Most of these sequences have been mapped on the apple molecular marker linkage maps. This thesis provides the basic knowledge of the four apple allergen genes regarding their genomic sequences, gene structures and linkage map positions. From a joint analysis of *Mal d 1* allelic diversity with the allergenicity phenotypes of skin prick tests on Dutch apple allergic patients, a conclusion was drawn that only the *Mal d 1* genes on LG 16 are responsible for birch-pollen related apple allergy. Furthermore, if the allergenicity is a matter of variant change, then the Mal d 1.06A is the most important protein to further investigate. Favourable alleles for low Mal d 1 allergenicity have been identified which can be used in the future approval tests.

Genomic PCR cloning and molecular markers for mapping

Most of the previously known Mal d sequences were obtained from mRNA. None of these were from the cvs Prima and Fiesta, the two parents of a core European reference linkage map population (Maliepaard et al. 1998). The intron number and size was known for three Mal d 1 genes. To map all these genes on the reference map from Prima x Fiesta and two alternative maps from Fiesta x Discovery and Jonathan x Prima (van de Weg, unpublished), proper representative markers were needed. The basic strategies employed here included four consecutive processes: (1) Search for the apple allergen sequences in the DNA databases (GenBank/EMBL) or other unpublished data, analyse these sequences and design PCR primers; (2) Clone and sequence these target genes on gDNA templates of Prima and Fiesta by PCR. (3) Analyse the polymorphisms of allelic sequences from the two parents and design primers to create markers. (4) Test these markers on the mapping population to obtain segregating data which enable to map these new markers onto the already known linkage maps. Fig.1 demonstrates this process.

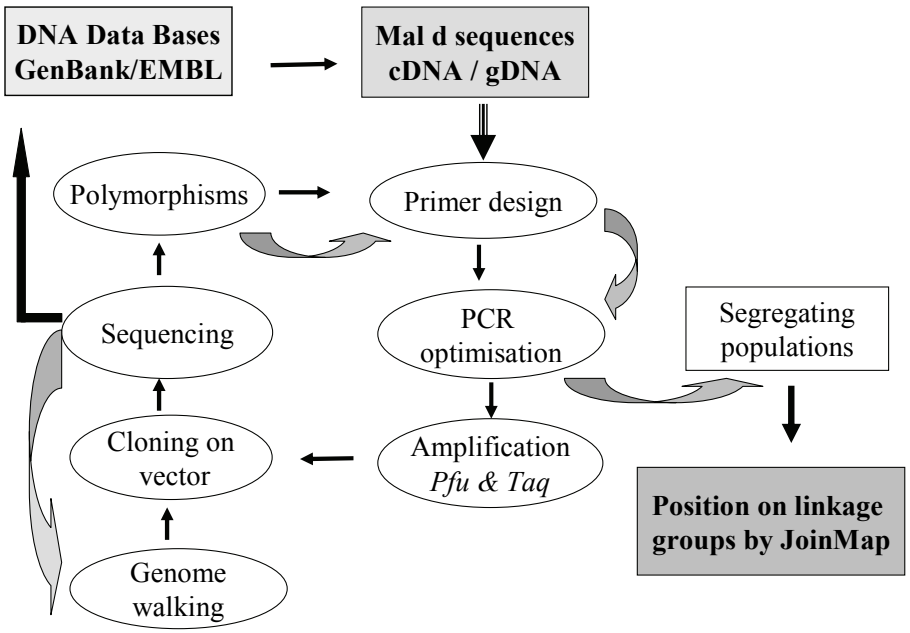

Fig. 1 The process of genomic PCR cloning and linkage mapping of apple allergen genes

PCR cloning techniques

In this process, we encountered three major problems. The first problem was the lack of allelic polymorphisms in the first round of PCR for *Mal d 3,* of which two distinct genes were identified (*Mal d 3.01* and *Mal d 3.02*, Chapter 2). To solve this problem, we applied the genome walking approach to sequence the upstream region for both genes. This led to a single SNP in a poly-T track in cv. Jonathan that enabled mapping of *Mal d 3.01*. The second problem was related to gene duplication. This problem was specific for the *Mal d 1* gene family with 18 members identified in this thesis. The DNA sequence identity for these members was over 70%. In several cases, one initial primer pair resulted in two or more individual genes. The intron size of some of these genes was a clear indicator of different genes, apart from the multiple sequence polymorphisms. Of the other three Mal d genes, multiple *Mal d 3* did exist; and at least two almost identical *Mal d 2* and *Mal d 4.01* genes became evident from the sequencing and mapping results. The existence of duplicated genes increases the difficulty to assign alleles for a single locus and to develop allele-specific markers. Therefore, a considerable number of clone replicates and highly specific markers were needed to tackle this problem. The primers on conserved regions could not guarantee that all the shared alleles/genes were amplified equally, which was observed in the first trial to clone two *Mal d 1* genes simultaneously. The third problem related to artificial PCR recombination. This happened more frequently when two gene copies existed and the cloning fragment was longer than 1Kb. This was the case in *Mal d 2* and *Mal d 4.01*. PCR recombinants were found (1 in 8 clones) with one common reverse primer to amplify two or more alleles of 900 nt length using *Pfu* polymerase. In the case of short fragments of less than 300 nt, PCR recombination was not observed even when both forward and reverse primers were in common. So, to avoid PCR recombination, we used gene-specific primers to amplify the target genes separately. Strategies to reduce and judge the artefact included the use of less PCR cycles to 25 (Zylstra et al. 1998) even for the long fragments, two independent PCR, UTR region primers, and sufficient replicates.

SNAP marker technique

The most frequent allelic variations at a certain locus are single nucleotide polymorphisms (SNPs). The average level of nucleotide diversity in *Arabidopsis thaliana* is one in every 3 kb (Drenkard et al. 2000). Apple allergen genes showed a more frequent SNP diversity of 1–19 SNPs per 1 kb. These polymorphisms within and between two parental cultivars of a mapping

population are vital for creating markers. The PCR-based agarose gel method to detect SNPs has evolved recently to facilitate tracing alleles or clones. The basis of this technique is that primers with a specific mismatch at the 3' end (SNPs) with addition of an extra mismatch within the last three bases of the primer will produce a significant reduction in the PCR product of the mismatch allele, but has relatively little effect on the amplification of the correct allele (Kwok et al. 1990). A modified allele-specific PCR procedure, called SNAP for single-nucleotide amplified polymorphisms, was used to generate *Arabidopsis* mapping makers with the aid of the SNAPER computer program to design possible primers (Drenkard et al. 2000). We simply designed primers using the 'Primer Designer' (version 2.0, Scientific & Educational Software, 1990) and empirically replaced one of the second or third bases. A mismatch at other positions was also applied to avoid dimer or hairpin formation in the primer itself or in the primer pair. Gradient PCR is powerful to find the best annealing temperature to distinguish a specific allele. Some primer pairs required a two-part cycling programme: 5–10 cycles with a high annealing temperature (Tm), followed by 30–35 cycles at lower annealing temperature. The initial high Tm is necessary to allow specificity of primer binding, whereas the late low Tm was needed to produce sufficient PCR product. Because of the introduction of mismatches in one of both primers, the PCR amplification plateau will be reached about 5 cycles later. SNAP is a dominant marker, but if the two alleles are specific, then two SNAP markers will serve as codominant markers. If there is any doubt about the score, a second marker can be checked. When the same samples are tested by three kinds of markers in *Mal d 4.03A* (as in Chapter 5, Table 3), the SNAP marker accuracy appeared to be over 99% in a cross-checking with SSR and CAPs markers. The SNAP marker can also be used to confirm whether an allele is truly or falsely identified by PCR. If it is an artefact, then this putative SNAP marker will only be amplified in the artefact clone but not in the original source.

Linkage map positions of apple allergen genes and relevance to other Rosaceae fruits

An intensive gene cloning and mapping process resulted in a comprehensive overview of the apple allergen genes on eight linkage groups (LG) (Fig. 2). The *Mal d 1* family consists of 18 gene members, of which 16 have been mapped as multiple gene clusters on the two homoeologous linkage groups (chromosomes) 13 and 16. One singe Mal d 1 locus was

identified on LG group 6 (Chapter 3). In a similar way, the genes, their loci and allelic diversity has also been analyzed for the other allergen genes in apple. Of *Mal d 2*, two gene copies were identified at the same position on LG 9. We still expect the presence of other *Mal d 2* genes on the homoeologous LG 17 (Chapter 4). *Mal d 3* genes were located on LG 4 and 12 (Chapter 2). Genomic characterization of *Mal d 4* revealed the existence of four genes of which two gene copies were found on LG 9 and two other single genes on LG 2 and LG 8. Also here, in view of linkage group duplication within the apple genome (van de Weg, unpublished) more genes on the homoeologous LG 7 and LG 17 are expected. These results have relevance for breeding. If the genomic map position of the expressed allergen gene is identified, breeding strategies for hypo-allergenic cultivars can be designed to replace that gene by a low-allergenic allele or by a gene with reduced expression. Gene sequences and map positions of these Mal d genes can be further analysed from the point of view of gene evolution and the origin of the apple genome. For example, the Mal d 1 subfamilies I and IV must have been present already in the common ancestral genome of *Malus domestica*, while the subfamily II and III developed later.

The major temperate fruit crops are apple, pear, peach, cherry, plum, apricot, quince and strawberry, which all belong to the Rosaceae family. Apple and pear belong to the Maloideae subfamily and have a similar genome structure with comparable linkage groups (Yamamoto et al. 2002, 2004; Hemmat et al. 2003). So the position of apple allergen genes can be used as references for pear. Also, preliminary results of the comparison between apple and *Prunus* maps suggest a high level of synteny between these two genera (Dirlewanger et al. 2004). *Prunus* LG 1 contains several common RFLP markers that aligned with apple LG 13 and LG 16 on which the majority of *Mal d 1* genes mapped; also with LG 4 and LG 8 where *Mal d 3.02* and *Mal d 4.03A* are located respectively. In addition, *Prunus* LG 3 aligns with apple LG 9 where apple *Mal d 2* and *Mal d 4.01* are located.

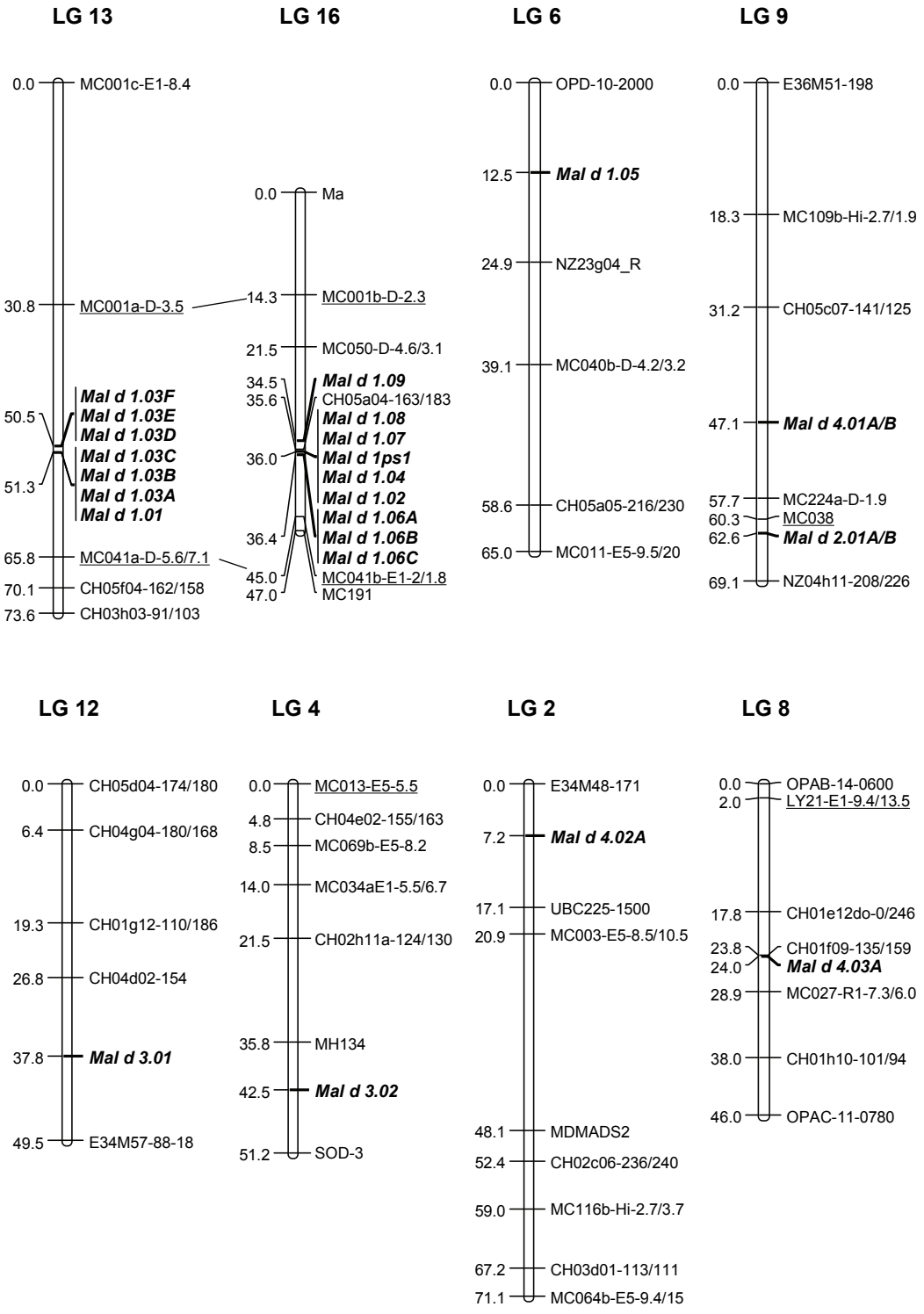

Fig. 2 Overview of the map positions of the four classes of apple allergen genes *Mal d 1* to *Mal d 4*

Gene families and relevance of individual genes in allergenicity

Food allergens belong to a limited number of protein families (Breiteneder and Mills 2005). Each protein member for a single species is encoded by multiple genes (gene family). All four apple allergen genes in this study are gene families, small or large. The number of gene member for some families at the genomic level is much larger than we previously thought. Study of allergen genes at the genomic level in plant becomes a huge task.

Genome sequences of *Arabidopsis thaliana* provide the foundation for detailed characterization of plant genes concerning genomic organization, function and evolution. 65% of the genes are duplicated (The Arabidopsis Genome Initiative, 2000). With regard to the four gene families of apple allergens, *Arabidopsis thaliana* genome sequences contain variable numbers of computational gene members for each gene family (Table 1).

Table 1 The computational number of members of four gene families on five *Arabidopsis thaliana* chromosomes

Gene family	Chrom-1	Chrom-2	Chrom-3	Chrom-4	Chrom-5	Total
Bet v 1 (PR-10)	17	3	2	3	4	29
LTP (PR-14)	19	14	16	25	24	98
TLP (PR-5)	8	2		6	3	19
Profilin		2		2	1	5

There are 29 putative *Bet v 1*-homologous genes scattered over the five chromosomes. More than half of these (17/29) are located on chromosome 1 in four clusters and each consisting of 3 to 7 genes. LTP genes in Arabidopsis form a super large gene family with 98 members, more or less distributed equally over all five chromosomes. Nineteen TLP genes are present on four chromosomes. Profilin genes form only a small gene family of five members, four of them have been characterized (Huang et al. 1996) located on chromosomes 2 and 4, the fifth is a conceptual one on chromosome 5.

PR-10 proteins are small, acidic, intracellular proteins with molecular masses ranging from 15 to 18 kDa (Van Loon and Van Strien 1999). Phylogenetic analyses performed with 67 *Bet v 1* homologous sequences derived from 22 species out of 7 plant families revealed that the

sequences from the same species were generally more similar to each other than to those from other species (Wen et al. 1997). As phylogenetic relationships have become better clarified, it becomes apparent that cross-reactivity does reflect taxonomy in the vast majority of cases (Weber 2003). Bet v 1 homologous sequences from the Rosaceae fruits were closer related to Bet v 1 from birch than to homologues from other plant families (Wen et al. 1997). This may result in the more frequent and higher degree of allergenicity of apple and other Rosaceae fruits.

PR-10 proteins are induced upon microbioal attack, wounding, or other physical and chemical stress. In addition, the constitutive accumulation of PR-10 proteins has been detected in various tissues during growth and development. The complicated expression pattern of these PR-10 proteins suggests that they have diverse biological functions. But how these genes are regulated is still a question to be answered. As to the allergenic features of Bet v 1 homologues, more than 50% of birch pollen-allergic patients show allergic symptoms towards various plant foods (Hoffmann-Sommergruber and Radauer 2004). Most Bet v 1-related allergens are heat sensitive and protease labile. So they provoke allergic reactions only in oral mucosa area (oral allergy syndrome).

TLPs and nsLTPs are also PR proteins ubiquitously present in plants. They are stable proteins because of intra-molecular disulfide bonds. This molecular property attributes to their allergenicity in a broader spectrum. Our two *Mal d 3* gene sequences do not have introns, which contrast to most nsLTPs that have a mini intron of 9–12 nt before the stop codon (Arondel et al. 2000). Perhaps other nsLTPs with an intron also exist in apple. Profilins are indispensable for any plant cell. They are very conserved at the protein level and are expressed differently in various tissues and at different development stages (Radauer and Hoffmann-Sommergruber 2004). From the information of ESTs for TLPs and profilin in apple fruit, multiple isoforms will be present in the fruit (Chapter 4). This indicates that more genes will be identified.

Because there exist so many gene members in a given plant species, they individually can be expressed differently in different tissues and at different times. To make a clear understanding which of these genes is responsible to the allergenicity is crucial for diagnosis. Molecular cloning can result in many and many PR-10 sequences. It is sure that not all of these proteins will encounter people and are allergenic. On the other hand, even for a specific tissue at a certain time, several proteins may simultaneously be present. Their individual quantity and allergenicity can be different. This gives rise to a great challenge for further

investigation. The amount of mRNA present in the tissue may not linearly relate to protein quantity. For example, Mal d 1.01 mRNA is abundantly present in the mature apple fruit, but its protein maybe far less than Mal d 1.02 with a higher mRNA expression than Mal d 1.01 (Helsper et al. 2002).

Epidemics of apple allergy and accurate diagnosis

Allergenicity refers to the reaction degree provoked by food or air-borne proteins after contact with sensitized patients. It is related to population and location. Mal d 1 related allergy is prevalent in Northern and Central Europe and North America where birch trees are present. Allergy to nsLTP mainly occurs in Mediterranean regions. Mal d 2 is at least of the same importance as Mal d 1 (Hsieh et al. 1995; Breiteneder et al. 2004). But there is still less clinical data available. Profilins as allergens in apple have been questioned recently (Wensing et al. 2002; Ebo et al. 2004). However, genetic analysis of the skin prick test scores from some patients on a progeny population (Fiesta x Discovery) of 39 individuals revealed a close association with linkage group 8 where *Mal d 4.03A* is located (van de Weg, unpublished).

Diagnostic tests for food allergy frequently resulted in poor sensitivity and specificity (van Ree et al. 2002). Therefore, the double blind placebo controlled food challenge (DBPCFC) is generally regarded as the golden standard. The major apple allergen Mal d 1 is labile upon disruption of fruit tissue and heat treatment. The prick to prick skin prick test (SPT) method with fresh fruit is, therefore, more reliable in comparison to the application of fruit extracts (Bolhaar et al. 2004). To assess the allergenicity of different apple cultivars, a large number of patients and repeated experiments are necessary. When the allergenicity of individual apple cultivars has been assessed, then the genetic and proteomic analyses of different genotypes will reveal the genes and alleles involved in low allergenic cultivars. For a given apple cultivar, not all the patients have the same SPT response. Separating the patients according to their SPT reaction is necessary for genetic analysis. The Mal d 1 allergenicity assessment of the 14 cultivars was from a selected group of patients with relatively low SPT reaction (Chapter 5). The allergic reaction is a complex matter with many uncertainties, such as the patient genetic make-up, his physical condition, apple source and state of maturity, and times of the year, etc. There is still much to do to improve the diagnostic methods.

Genetic and proteomic analysis of allergen genes

Genetic analysis to find the relevant *Mal d 1* genes causing the existing difference in allegenicity was carried out in Chapter 5. The first conclusion is that *Mal d 1* gene members on LG 16 have a major effect on allergenicity. Although we do not know the exact mechanism, at least we have information about the allelic constitution, and molecular markers are available to be used for research and application regarding this issue.

Furthermore, it appears that Mal d 3.01 is the dominant nsLTP in apple fruit with identical amino acid sequence for most current cultivars. However, differences in the concentration of nsLTP in the fruit of different cultivars have been identified (Ana I. Sancho, personal communication). This might be due to differences in the expression profile of individual isoallergens present in different cultivars. In this study, we identified four alleles of *Mal d 3.01*, including a special *Mal d 3.0102* allele with two amino acid changes in the cultivar Priscilla. This cultivar was classified in the group with the lowest amount of LTP measured by enzyme-linked immunosorbent assay (ELISA). The cv. Jonathan is in the group with high LTP content (Ana I. Sancho, personal communication). We found a special SNP in the poly-T region as well, which may be linked to some unknown regulatory elements. This allele has been passed on to cv. Idared as was shown by marker tests. The cultivars Red Delicious and Ingrid Marie possess a unique allele with an SNP of G on position 290 (Accn: AY572514; AY572515) resulting in CAGT rather than CAAT which is a putative enhancer element in the promoter region. From recent skin prick test results in the Spanish patients, cv. Fuji was more allergenic than cv. Golden Delicious, cv. Ecolette was the least allergenic (Bolhaar et al. 2004b). Because of the narrow genetic basis of the cultivated apple and the identified allele specific markers in this research present in the most frequently used parents of breeding programs, we can apply these markers to further analyse the relationship to LTP allergenicity in pedigrees and cross progenies.

Especially in the case of extended gene families, like pathogenesis related (PR) proteins which often have allergenic representatives, knowledge of the genomics of the allergen genes, their number in the genome and the sequence of the individual gene members is useful to identify the individual gene members that have come to expression. Proteomics approaches like QTOF and HPLC might reveal novel peptide sequences that can be traced back to the original gene sequences (Helsper et al., 2002; Reuter et al. 2005). In addition, genomics data

are useful to predict biochemical and physicochemical characteristics of the protein regarding its molecular weight, pI value, secondary and tertiary structure, thermal stability and resistances to proteolysis (Breiteneder and Mills, 2005).

Breeding perspectives

Selection for low allergenicity apples

In the EU-SAFE project, differences in allergenicity among apple cultivars were tested by prick-to-prick skin prick test (SPT) and double-blind placebo-controlled food challenges (DBPCFC) in well-documented birch pollen related apple allergic patients. For selection, a broad diversity of apple cultivars and genotypes was available at Plant Research International, Wageningen, from which over twenty apple cultivars have been analyzed (Bolhaar et al. 2004b). The differences in allergenicity among the tested cultivars were reproducible in fruits from two harvest years (Chapter 5)

The identification of Santana as a low-allergenic cultivar permits consumption of this cultivar by apple allergic patients. Confirmation of these results in a larger patient population is under way. This research shows the usefulness of the prick-to-prick SPT (combined with DBPCFC for confirmation) as a rapid and quantitative test for allergenicity in cultivar screening (Bolhaar et al. 2004). This selection strategy for the production of hypoallergenic cultivars is not restricted to apple but can be applied to other crops, like soybean (Codina et al. 2003) in which a diversity of phenotypes is available. Genetic analysis with the high and low allergenic Mal d 1-related apple cultivars pinpoints the allelic constitution for the low ones. At the same time, molecular markers representing these alleles were developed. These makers can be used to find low allergenic genotypes among the selections from the existing breeding programme and can be confirmed by a skin prick test or DBPCFC later when these plants are in fruit-bearing stage. For a longer term, proven low allergenic apple cultivars such as Santana can be used as parent.

Genetic modification for hypoallergen apple

The knockout strategy for the introduction of hypoallergenicity has been applied in rice (Tada et al. 1996) and soybean (Herman et al. 2003), and is expected to become a common procedure towards the production of hypoallergenic raw materials (Gilissen et al. 2005). The

production of an apple plant with a significant reduction of the overall expression of Mal d 1 from an existing economically successful cultivar using RNAi seems an attractive time-saving and simpler alternative than crossing strategies where each new genotype has to be tested for its market value, including tests for taste and texture, production and storage, consumer acceptability, and economic viability. Such tests will take at least 15 to 20 years. Reduction of expression of Mal d 1 appeared to be feasible using RNAi approach to reduce the expression of Mal d 1 (Gilissen et al. 2005).

With modern technologies for marker-assisted breeding and gene-silencing, there will be good hope in this new century for sufferers of apple allergy to eat an apple daily without the fear of an allergic reaction.

Future directions

Recombinant food allergens have been frequently produced for research and diagnosis. Most are comparable to those of their natural counterparts. In addition, the generation of hypoallergenic isoforms and mutants is useful for therapeutic purposes and for the determination of epitopes and cross-reactive structures (Lorenz, et al. 2001). Identification and quantification of Mal d 1 isoallergen and variants in apple fruits of Santana and Golden Delicious will provide further solid evidence on which *Mal d 1* genes and alleles are expressed and related to allergenicity. A recombinantly modified mutation of a Mal d 1 isoform (AJ41771) has shown reduced allergenicity (Hoffmann-Sommergruber et al. 2005).

MALDI-TOF or Q-TOF and peptide sequencing together with the known genomic information can identify different proteins in the natural allergen mixtures. But the accurate isoallergen and variant quantification will be difficult. One approach to solve this problem may be the development of monoclonal recombinant variants of all known Mal d 1 allergens by genome cloning and their application for Mal d 1 isoallergen separation and allergenicity test. Furthermore, molecular modelling of deduced variants and 3-D structural determination will elucidate clearly the molecular mechanism of allergenicity. Reference 3D structures for several allergens, such as Bet v 1 (Spangfort et al. 1997), Pru av 1 (Neudecker et al. 2001), Bet v 1l (Markovic-Housley et al. 2003), thaumatin-like protein in tobacco (Koiwa et al. 1999), rice and maize nsLTPs (Lee et al. 1998; Han et al. 2001) provide templates for such modelling.

To assess their different levels of allergenicity, allergen isoforms and their expression levels in individual apple cultivars should be analysed in future experiments. Because different apple cultivars give different reactions in the same patient group, genetic and proteomic analysis will determine the true allergens involved. Detailed investigation on Bet v 1 proteins in pollen using more sensitive methods is still needed to reveal the actual heterogeneity. Furthermore, Mal d 1 and Bet v 1 cross-reactivity will be better understood. Now that we have a relatively complete knowledge of the allergens, genomic sequences, linkage map positions and some allelic diversity, we are still left with deeper questions of the allergenicity difference among cultivars and populations. Allelic diversity on *Mal d 2* and *Mal d 4* on the same set of ten cultivars will provide a more complete picture of allergen genes for each cultivar.

The increase in the prevalence of allergy has been widespread recognized throughout the world. Much allergy research has been conducted in the developed countries, but it does not mean that there is no or less allergy problem in the developing countries. Food allergy in China was reported to affect 3.4–5.0% of residents in Beijing, Guangdong cities (Hill et al. 1997). In China, the most frequent allergenic foods are egg, milk, peanut, soybean and wheat. But there are only few data about the incidence of fruit allergy. With the prevailing modern life style and the current globalization of food production and consumption chains, the allergy matter will go beyond the regional level. For example, in case apple allergy is not a problem for Chinese consumers, the apple growers in China should still consider the consumers abroad, because China is now the top (in volume) apple exporter in the world. As it was put in an editorial of the Allergy journal: ‘allergy as a global problem: think globally, act globally’. Allergy is not limited to a specific population, not organ by organ, not allergen by allergen. The allergic disease should be considered globally from gene to molecular biology, cell biology, histopathology, symptomology, social and environmental sciences, and the quality of life (Bousquet 2002). Allergy is a multifactor-determined disease. Its prevention and treatment requires an interdisciplinary and integrated approach.

References

Arondel V, Vergnolle C, Cantrel C, Kader JC (2000) Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. Plant Sci 157:1–12

Asero R, Mistrello G, Roncarolo D, de Vries Sacco C, Gautier Marie F, Ciurana Caroline LF, Verbeek E, Mohammadi T, Knul Brettlova V, Akkerdaas Jaap H, Bulder I, Aalberse Rob C, van Ree R (2000) Lipid transfer protein: A pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. Int Arc Allergy Immunol 122:20–32

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

Atkinson RG, Perry J, Matsui T, Ross GS, Macrae EA (1996) A stress-, pathogenesis-, and allergen-related cDNA in apple fruit is also ripening-related. New Zealand J Crop Horti Sci 24:103–107

Baldi P, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologues in apple. Theor Appl Genet 109:231–239

Ballmer-Weber BK (2002) Lipid transfer protein as a potential panallergen? Allergy, 2002; 57:873–875

Beil D, Ring J, Vieter D, Kinder H, Behrendt H, Paschke A, Steinhart H (2001) Modulation of allergic reaction to apple by pesticide residues. Food Agri Immunol 13:93–102

Beuning LL, Bowen JH, Persson HA, Barraclough D, Bulley S, MacRae EA (2004) Characterisation of Mal d 1-related genes in *Malus*. Plant Mol Biol 55:369–388

Bohle B (2004) T lymphocytes and food allergy. Mol Nutri Food Res 48:424–433

Bolhaar STHP, Van Ree R, Bruijnzeel-Koomen CAFM, Knulst AC, Zuidmeer L (2004a) Allergy to Jackfruit: a noval example of Bet v 1-related food allergy. Allergy 59:1187–1192

Bolhaar STHP, Van de Weg WE, Van Ree R, Gonzalez-Mancebo E, Bruijnzeel-Koomen CAFM, Fernandez-Rivas M, Jansen J, Hoffmann-Sommergruber K, Knulst AC, Gilissen LJWJ (2004b) In vivo assessment by prick-to-prick testing and double-blinded placebo-controlled food chanllenge of allergenicity of apple cultivars. In PhD thesis of Bolhaar STHP, Food allergy: from sensitization and symptoms to prevention and treatment. University Medical Center Utrecht, p63–75

Bolhaar STHP, Van Ree R, Ma Y, Bruijnzeel-Koomen CAFM, Vietths S, Hoffmann-Sommergruber, Knulst AC, Zuidmeer L (2005) Severe allergy to Sharon fruit caused by birch pollen. Intern ArcAllergy Immunol 136:45–52

Bonants P, De Weerdt MH, Van Gent PM, Lacourt I, Cooke D, Duncan J (1997) Detection and identification of *Phytophthora fragariae* Hickman by the polymerase chain reaction. Eur J Plant Pathol 103:345–355

Botton A, Begheldo M, Rasori A, Bonghi C, Tonutti P (2002) Differential expression of two lipid transfer protein genes in reproductive organs of peach (*Prunus persica* L. Batsch). Plant Sci 163:993–1000

Bousquet J (2002) Allergy as a global problem: 'Think globally, act globally'. Allergy 57:661–662

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

Breiteneder H, Ebner C (2000) Molecular and biochemical classification of plant-derived food allergens. J Allergy Clin Immunol 106:27–36

Breiteneder H, Mills ENC (2005) Molecular properties of food allergens. J Allergy Clin Immunol 115:14–23

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. Biotechniques 20:1004–1010

Calenge F, Faure A, Goerre M, Gebhardt C, Van de Weg WE, Parisi L, Durel C-E (2004) A QTL analysis reveals both broadspectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of *Venturia inaequalis*. Phytopathology 94:370–379

Calenge F, Van der Linden CG, Van de Weg WE, Schoulten HJ, Van Arkel G, Denance C, Durel CE (2005a) Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. Thero Appl Genet 110:660–668

Calenge F, Drouet D, Denancé C, Van de Weg WE, Brisset MN, Paulin JP, Durel CE (2005b) Identification of a major QTL together with several minor additive or epistatic QTLs for resistance to fire blight in apple in two related progenies. Theor Appl Genet. In press

Calkoven PG, Aalbers M, Koshte VL, Pos O, Oei HD, Aalberse RC (1987) Cross-reactivity among birch pollen, vegetables and fruits as detected by IgE antibodies is due to at least three distinct cross-reactive structures. Allergy 42:382–390

Chardin H, Mayer C, Senechal H, Wal JM, Poncet P, Desvaux FX, Peltre G (2003) Lipid transfer protein 1 is a possible allergen in *Arabidopsis thaliana*. Int Arc Allergy Immunol 131:85–90

Cheng FS, Weeden NF, Brown SK (1996) Identification of co-dominant RAPD markers tightly linked to fruit skin color in apple. Theor Appl Genet 93:222–227

Cheng FS, Weeden NF, Brown SK, Aldwinckle HS, Gardiner SE, Bus VG (1998) Development of a DNA marker for *Vm*, a gene conferring resistance to apple scab. Genome 41: 208-214

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

Clark AM, Bohnert HJ (1999) Cell-specific expression of genes of the lipid transfer protein family from *Arabidopsis thaliana*. Plant Cell Physiol 40:69–76

Codina R, Ardusso L, Lockey RF, Crisci C, Medina I (2003) Allergenicity of varieties of soybean. Allergy 58:1293–1298

Conner PJ, Brown SK, Weeden NF (1997) Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. J Amer Soc Hort Sci 122:350–359

Cortegano I, Civantos E, Aceituno E, Del Moral A, Lopez E, Lombardero M, del Pozo V, Lahoz C (2004) cloning and expression of a maor allergen from Cupressus arizonica pollen, Cup a 3, a PR-5 protein expressed under polluted environment. Allergy 59:485–490

Crowell DN, John ME, Russell D, Amasino RM (1992) Characterization of a stress-induced, developmentally regulated gene family from soybean. Plant Mol Biol 18:459–466

Diaz-Perales A, Garcia-Casado G, Sanchez-Monge R, Garcia-Selles FJ, Barber D, Salcedo G (2002) cDNA cloning and heterologous expression of the major allergens from peach and apple belonging to the lipid-transfer protein family. Clin Exp Allergy 32:87–92

Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. Am J Bot 19:1081–1086

Dirlewanger E, Graziano E, Joobeur T, Gerriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci 101:9891–9896

Dondini L, Pierantoni L , Gaiotti F, Chiodini R, Tartarini S, Bazzi C, Sansavini S(2004) Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. Mol Breed 14:407-418

Douliez JP, Jegou S, Pato C, Molle D, Tran V, Marion D (2001) Binding of two mono-acylated lipid monomers by the barley lipid transfer protein, LTP1, as viewed by fluorescence, isothermal titration calorimetry and molecular modelling. Eur J Biochem 268:384–388

Dreborg S, Foucard T (1983) Allergy to apple, carrot and potato in children with birch pollen allergy. Allergy 38:167–172

Dreborg S (1993) Allergen standardisation and skin tests. EAACI position paper. Allergy 48 (suppl 14):49–82

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

Durel CE, Van de Weg WE, Venisse JS, Parisi L (2000) Localisation of a major gene for apple scab resistance on the European genetic map of the Prima x Fiesta cross. In: Integrated control of pome fruit diseases. IOBC-WPRS Bull 23:245–246

Durel CE, Parisi L, Laurens F, Van de Weg E, Liebhard R, Koller, B, Jourjon MF (2003) Genetic dissection of partial resistance against two monoconidial strains of the new race 6 of *Venturia inaequalis* in apple. Genome 46:224–234

Ebner C, Birkner T, Valenta R, Rumpold H, Breitenbach M, Scheiner O, Kraft D, Sakamoto T, Hayashi Y, Yamada M, Torii S, Urisu A, Vieths S, Schoning B, Petersen A, Janek K, Aulepp H, Rudeschko O, Fahlbusch B, Henzgen M, Schlenvoigt G, Herrmann D, Jager L (1991) Common epitopes of birch pollen and apples: Studies by Western and Northern blot. J Allergy Clin Immunol 88:588–594

Ebo DG, hangendorens MM, Bridts CH, De Clerk LS, Stevens WJ (2004) Sensitization to cross-reactive carbohydrate determinants and the ubiquitous protein profilin: mimickers of allergy. Clin Exp Allergy 34:137–144

Eriksson NE, Formgren H, Svenonius E (1982) Food hypersensitivity in patients with pollern allergy. Allery 33:437–443

Fernandez-Rivas M, Cuevas M (1999) Peels of Rosaceae fruits have higher allergenicity than pulps. Clin Exp Allergy 29:1239–1247

Fernandez-Rivas M, Miles S (2004) Food allergies: clinical and psychosocial perspectives. In: Mills ENC, Shewry PR (eds) Plant Food Allergens. Blackwell, pp1–23

Ferreira F, Hirtenlehner K, Jilek A, Godnik Cvar J, Breiteneder H, Grimm R, Hoffmann Sommergruber K, Scheiner O, Kraft D, Breitenbach M, Rheinberger Hans J, Ebner C (1996) Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen bet v 1: Potential use of hypoallergenic isoforms for immunotherapy. J Exp Med 183:599–609

Ferreira F, Hebenstreit D, Kramer B, Himly M, Breiteneder H, Scheiner O, Britza P, Ebner C, Breitenbach M (2000) Amino acid positions involved in the formation of IgE-binding epitopes of Api g 1 and Mal d 1 allergens. J Allergy Clin Immunol 105:S137

Fritsch R, Bole 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 Immonol 102:679–686

Gao ZS, Van den Weg WE, Schaart JG, Van der Meer IM, Kodde L, Laimer M, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJWJ (2005a). 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

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 (2005b) Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). Theor Appl Genet in press

Garcia-Casado G, Pacios LF, Diaz- Perales A, Sanchez-Monge R, Lombardero M, Garcia-Selles Francisco J, Polo F, Barber D, Salcedo G (2003) Identification of IgE-binding epitopes of the major peach allergen Pru p 3. J Allergy Clin Immunol 112:599–605

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

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

Halmepuro L, Vuontela K, Kalimo K, Björksten (1984) Cross-reactivity of IgE antibodies with allergens in birch pollen, fruit and vegetables. Int Archs Allergy Immunol 74:235–240

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(3):462–468

Han GW, Lee JY, Song HK, Chang C, Min K, Moon J, Shin DH, Kopka ML, Sawaya MR, Yuan HS, Kim TD, Choe J, Lim D, Moon HJ, Suh SW (2001) Structure basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography. J Mol Biol 308:263–278

Hansen KS, Vieths S, Vestergaard H, Skov PS, Bindslev-Jansen C, Poulsen LK (2001) Seasonal variation in food allergy to apple. J Chromatogr B 756:19–32

Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. Thrends Genet 18(8):426–430

Helsper JPFG, Gilissen Luud JWJ, 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

Hemmat M, Brown SK (2002) Tagging and mapping scab resistance genes from R12740-7A apple. J Am Soc Hortic Sci 127:365–370

Hemmat M, Weeden NF, Brown SK (2003) Mapping and evaluation of *Malus x domestica* microsatelites in apple and pear J Am Soc Hortic Sci 128:515–520

Hemmat M, Weeden NF, Maganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Heredity 85:4–11

Herman EM, Helm RM, Jung R, Kinney AJ (2003). Genetic modification removes an immunodominant allergen from soybean. Plant Physiol 132:36–43

Hill DJ, Hosking CS, Zhie CY, Leung R, Baratwidjaja K, Iikura N, Gonzalez-Andaya A, Wah LB, Hsieh KH (1997) The frequency of food allergy in Australia and Asia. Environ Toxicol and Pharmacol 4:101-110

Hoffmann-Sommergruber K, Vanek-Krebitz M, Radauer C, Wen J, Ferreira F, Scheiner O, Breiteneder H (1997) Genomic characterization of members of the *Bet v 1* family: Genes coding for allergens and pathogenesis-related proteins share intron positions. Gene 197:91–100

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

Hoffmann-Sommergruber K, the SAFE consortium (2005) The SAFE project: 'plant food allergies: field to table strategies for reducing their incidence in Europe' an EC-funded study. Allergy 60:436–442

Holm J, Baerentzen G, Gajhede M, Ipsen H, Larsen JN, Lowenstein H, Wissenbach M, Spangfort MD (2001) Molecular basis of allergic cross-reactivity between group 1 major allergens from birch and apple. J Chromatogr B 756:307–313

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 spite 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(3):375–383

James CM, Clare JB, Evens KM (2005) Identification of moleualr markers linked to the mildew resistance gene Pl-d in apple. Thero Appl Genet 110:175–181

Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteender H, Milss ENC (2005) Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: An *in silico* analysis. J Allergy Clin Immunol 115:103–170

Juhlin-Dannfelt C (1948) About the occurrence of various forms of pollen allergy in Sweden. Acta Med. Scand. 26:563–577

Jung HW, Kim W, Hwang BK (2003) Three pathogen-inducible genes encoding lipid transfer protein from pepper are differentially activated by pathogens, abiotic, and environmental stresses. Plant Cell Environ 26:915–928

Kader JC (1996) Lipid-transfer proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 47: 627–654

Kandasamy MK, McKinney EC Meagher RB (2002) Plant profilin isovariants are distinctly regulated in vegetative and reproductive tissues. Cell Motil Cytoskeleton 52(1):22-32

King GJ, Maliepaard C, Lynn JR, Alston FH, Durel CE, Evans KM, Griffon B, Laurens F, Manganaris AG, Schrevens E, Tartarini S, Verhaegh J (2000) Quantitative genetic

analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). Theor Appl Genet 100:1074–1084

King TP, Hoffman D, Lowenstein H, Marsh DG, Platts MT AE, Thomas W (1995) Allergen nomenclature. J Allergy Clin Immunol 96:5-14

Kleine-Tebbe J, Wangorsch A, Vogel L, Crowell DN, Haustein UF, Vieths S (2002) Severe oral allergy syndrome and anaphylactic reactions caused by a Bet v 1-related PR-10 protein in soybean, SAM22. J Allergy Clin Immunol 110:797–804

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

Koiwa H, Kato H, Nakatsu T, Oda J, Yamada Y, Sato F (1999) Crystal structure of tobacco PR-5 protein at 1.8 A resolution reveals a conserved acidic cleft structure in antifungal thaumatin-like proteins. J Mol Biol 286:1137–1145

Korban SS, Skirvin RM (1984) Nomenclature of the cultivated apple. HortScience 19:177–180

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

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsy JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Research 18: 999-1006

Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW (1998) Rice non-specific lipid transfer protein: the 1.6A crystal structure in the unliganded state reveals a small hydrophobic cavity. J Mol Biol 276:437–448

Lescot M, Dehais P, Moreau Y, De Moor B, Rouze P, Rombauts S (2002) PlantCARE: a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327

Liebhard 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

Liebhard R, Koller B, Giafranceschi L, Gessler C (2003a) Creating a saturated reference map for the apple (*Malus x domestica* Borkh.) genome. Theor Appl Genet 106:1497–1508

Liebhard B, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003b) Mapping quantitative physiological traits in apple (Malus x domestica Borkh.) Plant Mol Biol 52:511–526

Liu JJ, Ekramoddoullah AKM (2004) Characterization, expression and evolution of two novel subfamilies of *Pinus monticola* cDNAs encoding pathogenesis-related (PR)-10 proteins. Tree Physiol 24:1377–1385

Lorenz AR, Scheurer S, Haustein D, Vieths S (2001) Recombinant food allergens. J Chromatog B 756:255–279

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

Maliepaard C (2000) Genetic mapping in a full-sib family of apple. Doctoral thesis, Wageningen University. P7

Maliepaard C, Silanpää MJ , Van Ooijen JW, Jansen RC, Arjas E (2001) Bayesian versus frequentitst analysis of multiple quantitative trait loci with an application to an outbred apple cross. Theor Appl Genet 103:1243–1253

Markovis-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

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. The Plant Cell 13:1179–1191

Mills ENC, Madsen C, Shewry PR, Wichers HJ (2003) Food allergens of plant origin-their molecular and evolutionary relationships. Trends in Food Science & Technology 14:145–156

Neutecker P, Schweimer K, Nerkamp J, Scheurer S, Vieths S, Sticht H, Rösch P (2001) Allergic cross-reactivity made visible: solution structure of the major cherry allergen Pru av 1. J Biol Chem 276(25):22756–22763

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(2):355–362

Pastorello EA, Pravettoni V, Farioli L, Ispano M, Fortunato D, Monza M, Giuffrida MG, Rivolta F, Scibola E, Ansaloni R, Incorvaia C, Conti A, Ortolani C (1999a) Clinical role of a lipid transfer protein that acts as a new apple-specific allergen. J Allergy Clin Immunol 104:1099–1106

Pastorello EA, Farioli L, Pravettoni V, Ortolani C, Ispano M, Monza M, Baroglio C, Scibola E, Ansaloni R, Incorvaia C, Conti A (1999b) The major allergen of peach (*Prunus persica*) is a lipid transfer protein. J Allergy Clin Immunol 103:520-526

Pastorello EA, Farioli L, Pravettoni V, Ispano M, Scibola E, Trambaioli C, Giuffrida Maria G, Ansaloni R, Godovac Zimmermann J, Conti A, Fortunato D, Ortolani C (2000a) The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. J Allergy Clin Immunol 106:744-751

Pastorello EA, Purello D' Ambrosio F, Pravettoni V, Farioli L, Giuffrida G, Monza M, Ansaloni R, Fortunato D, Scibola E, Rivolta F, Incorvaia C, Bengtsson A, Conti A, Ortolani C (2000b) Evidence for a lipid transfer protein as the major allergen of apricot. J Allergy Clin Immunol 105:371–377

Pastorello EA, Farioli L, Pravettoni V, Giuffrida MG, Ortolani C, Fortunato D, Trambaioli C, Scibola E, Calamari AM, Robino AM, Conti A (2001) Characterization of the major allergen of plum as a lipid transfer protein. J Chromatogr B. 756:95-103

Patocchi A, Bigler B, Koller B, Kellerhals M, Gessler C. 2004. *Vr2*: a new apple scab resistance gene. Theor Appl Genet 109:1087–1092.

Pauli G, Oster JP, Deviller P, Heiss S, Bessot JC, Susani M, Ferreira F, Kraft D, Valenta R (1996) Allergens, IgE, mediators, inflammatory mechanisms: Skin testing with recombinant allergens rBet v 1 and birch profilin, rBet v 2: Diagnostic value for birch pollen and associated allergies. J Allergy Clin Immunol 97:1100–1109

Pierantoni L, Cho KH, Shin IS, Chiodini R, Tartarini S, Dondini L, Kang SJ, Sansavini S (2004) Characterization anf transferability of apple SSRs to two European pear F1 populations. Theor Appl Genet 109:1519–1524

Poulsen LK, Liisberg C, Bindslev-Jensen C, Malling HJ (1993) Precise determination of skin-prick tests: validation of a scaning device and software for a personal computer. Clin Exp Allergy 23:61–68

Puehringer H, Moll D, Hoffmann Sommergruber K, Watillon B, Katinger H, Laimer da Camara Machado M (2000) The promoter of an apple Ypr10 gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. Plant Sci 152:35–50

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 and Hoffmann-Sommergruber K (2004) Profilins. In: Mills ENC, Shewry PR (eds) Plant Food Allergens. Blackwell, pp105–124

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

Reuter A, Fortunato, D, Caroffo LP, Napolitana L, Scheurer S, Giuffrida MG, Vieths S, Conti A (2005) Novel isoforms of Pru av 1 with diverging immunoglobulin E binding properties identified by a synergistic combination of molecular biology and proteomics. Proteomics 5:282–289

Roche P, Alston FH, Maliepaard C, Evans KM, Vrielink R, Dunemann F, Markussen T, Tartarini S, Brown LM, Ryder C, King GJ (1997) RFLP and RAPD markers linked to the rosy leaf curling aphid resistance gene (*Sd1*) in apple. Theor Appl Genet 94:3–4

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

Rohac M, Birkner T, Reimitzer I, et al (1991) The immunological relationship of epitopes on major tree allergens. Mol immunol 28:897-906

Roy SW (2003) Recent evidence for the Exon Theory of Genes. Genetica 118;251–266

Sax K (1933) The origin of *Pomoideae*. J Am Soc Hortic Sci 30:147–150

Scheurer S, Pastorello EA, Wangorsch A, Kaestner M, Haustein D, Vieths S (2001) Recombinant allergens Pru av 1 and Pru av 4 and a newly identified lipid transfer protein in the in vitro diagnosis of cherry allergy. J Allergy Clin Immunol 107: 724–731

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

Schoning B, Ziegler WH, Vieths S, Baltes W (1996) Apple allergy: the cDNA sequence of the major allergen of apple, determined by performing PCR with a primer based on the N-terminal amino acid sequence, is highly homologous to the sequence of the major birch pollen allergen. J Sci Food Agri 71:475–482

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

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

Spangfort MD, Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, Joost Van Neerven RJ, Schou C, Lowenstein H (1997) Three-dimensional structure and epitopes of Bet v 1. Int Arc Allergy Immunol 113:243–245

Stebbins G (1950) Polyploidy and the origin of higher categories. In: Variation and evolution in plants. Columbia Univerity Press, New York, pp359–362

Sterk P, Booij H, Schellekens GA, Van Kammen A, De Vries SC (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. Plant Cell 3:907–922

Swoboda I, Jilek A, Ferreira F, Engel E, Hoffmann-Sommergruber K, Scheiner O, Kraft D, Breiteneder H, Pittenauer E, Schmid E, Vicente O, Heberle BE, Ahorn H, Breitenbach M (1995) Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry, and cDNA cloning. J Biol Chemistry 270:2607–2613

Tada Y, Nakase M, Adachi T, Nakamura R, Shimada H, takahashi M, fujimura T, Matsuda T (1996). Reduction of 14-16 kDa allergenic proteins in transgenic rice plants by antisense gene. FEBS Letters 91:341-345

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequences of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Velazhahan V and Muthukrishnan S (2003) Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata.* Biologia Plantarum 47 (3):347–354

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 de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM. (2004) Pedigree genotyping: a new pedigree-based approach of QTL identification and allele mining. Acta Hortic 663:45–50

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 ver.3.0: Powerful genetic linkage mapping made easy. Plant Research International, Wageningen, The Netherlands

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

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

Venisse JP, Malnoy M, Faize M, Paulin JP and Brisset MN (2002) Modulation of defense responses of *Malus* spp. During compatible and incompatible interactions with *Erwinia amylovora*. Mol Plant-Microbe Interact 15 (12):1204–1212

Vieths S, Schoning B, Jankiewicz A (1993) Occurrence of IgE binding allergens during ripening of apple fruits. Food Agri Immunol 5:93–105

Vieths S, Jankiewicz A, Schoning B, Aulepp H (1994) Apple allergy: the IgE-binding potency of apple strains is related to the occurrence of the 18-kDa allergen. Allergy 49:262–271

Vieths S, Janek K, Aulepp H, Petersen A (1995) Isolation and characterization of the 18-kDa major apple allergen and comparison with the major birch pollen allergen (Bet v I). Allergy 50:421–430

Vieths S (1997) Allergenic cross-reactivity, food allergy and pollen. Environ Toxicol Chem 4:61–70

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

Watkin R (1995) Apples and pears. In Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman, London, pp418–422

Weber RW (2003) Pattern of pollen cross-allergenicity. J Allergy Clin Immunol 112:229–239

Wen J, Vanek-Krebitz M, Hoffmann-Sommergruber K, Scheiner O, Breiteneder H (1997) The potential of *Betv1* homologues, a nuclear multigene family, as phylogenetic markers in flowering plants. Mol Phylogenet Evol 8:317–333

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 e1: IgE cross-reactivity and allergenic potency. Allergy 59:526–532

Xin PG, Xiang FG, Chen XS, Shen X, Wu Z (1996) Jounal of Fruit Science (Chinese) 13 (suppl) 93–110

Yamamoto TT, Kimura T, Sawamura Y, Kotobuki K, Ban Y, Hayashi T, Matsuta N (2001) SSRs isolated from apple can identify polymorphism and genetic diversity in pear. Theor Appl Genet 102: 865-870

Yamamoto T, Kimura T, Shoda M, Saito T, Sawamura Y (2002) Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. Theor Appl Genet 106:9–18

Yamamoto T, Kimura T, Saito T, Kotobuki K, Matsuta N, Liebhard R, Gessler C, van de Weg, WE and Hayashi T (2004) genetic linkage maps of Japanese and European pears aligned to the apple consensus map. Acta Hortic 663:51–56

Ye Q, Xu Y, Yan F, Tang L, Chen F (2001) Molecualr cloning and chracterization of rice pollen profilin. Acta Biochimica Biophysica Sinica 33:452–456

Zaidi S, Poupard P, Brisset MN, Paulin JP, Simoneau P (2001) Characterization in apple leaves of two subclasses of PR-10 transcripts inducible by acibenzolar-S-methyl, a functional analogue of salicylic acid. Physiol Mol Plant Pathol 59:33–43

Zylstra P, Rothenfluh H, Eller G, Blanden R, Steele EJ (1998) PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. Immunol Cell Biol 76:395–405

Summary

Apple is the most important fruit produced in temperate climate regions of the world. Nowadays, genetic analysis of apple is speeding up due to the development of molecular marker linkage maps and gene sequencing. Some molecular markers linked to disease resistance are currently in use in breeding programmes. Although apple is a healthy food product to the vast majority of the population, some individuals can not eat this fruit because of the occurrence of allergic reactions. Food allergy is a hypersensitive reaction to normally harmless substances (allergens, mostly proteins) and involves humoral immune responses, mediated by immunoglobulin E (IgE). Apple allergy is common in Europe and affects about 2% of the population. Since 1948, apple allergy has been investigated intensively with regard to its prevalence and the allergens involved. Up to now, four classes of allergens (Mal d 1, -2, -3 and -4) have been identified; their reference DNA sequences are available.

● Mal d 1, a member of the pathogenesis-related (PR) protein family 10, is the major apple allergen that is cross-reactive to IgE in patients sensitized by the allergen Bet v 1 from birch. Birch pollen-related apple allergy occurs in Northern and Central European populations. The symptoms are generally limited to the oral region (oral allergy syndrome), such as itching, tingling and swelling of the lips, tongue and throat, after eating apples. Apple cultivars differ considerably in their Mal d 1 allergenicity.

● Mal d 2 is a homologue of the thaumatin-like protein (TLP) which belongs to the PR-5 protein family. It represents one of the major proteins in the mature apple fruit. The major geographical area for the occurrence of Mal d 2 allergy is still unclear.

● Mal d 3 is the non-specific lipid transfer protein (nsLTP), a member of the PR-14 family. nsLTPs have been identified as major allergens in several Rosaceae fruits, including apple, peach, apricot, plum and cherry, and can provoke serious allergic reactions in the gastro-intestinal tract up to an anaphylactic shock. The prevalence of allergy to nsLTP-containing fruits is especially high in Mediterranean communities, such as Spain and Italy. Peach nsLTP was assumed the primary sensitizing allergen.

● Mal d 4 are apple profilins which have been assumed to be cross-reactive to birch profilin Bet v 2. Its clinical relevance is not clear.

This thesis is an outcome of the recently finished EU-SAFE project (Plant food allergies: field to table strategies for reducing their incidence in Europe, QLK-CT-2000-01394). This project was initiated to characterize the relevant allergens and their relationships to severe versus mild allergenic symptomatology using apple as a model system. To design an apple-breeding programme aiming at hypoallergenic cultivars, it is essential to unravel the genetic basis of the observed differences in allergenicity among apple cultivars at the qualitative and quantitative level of (iso)allergen genes and their variants. This thesis aims to provide the genome sequences and linkage map positions of major apple allergens. It also aims at the development of genetic markers for the identification of low-allergenicity.

The employed strategy is to clone and sequence all the identified apple allergen genes from genomic DNA of two cultivars, Prima and Fiesta, parents of a reference linkage mapping population. DNA sequencing allows the detection of single nucleotide polymorphisms (SNPs) between alleles of one locus. Molecular markers based on these sequences were created, with the majority of PCR agarose-gel based Single Nucleotide Amplified Polymorphism (SNAP) marker. Then the segregating markers were mapped on established maps. This strategy has been applied to the four allergen genes mentioned above.

In **Chapter 2,** two Mal d 3 isoallergen genes (*Mal d 3.01* and *Mal d 3.02*) were identified. *Mal d 3.01* is identical to accession AF221502, the reference sequence used for designing the primers to clone the genes. *Mal d 3.02* is new. Both genes were found to be without introns and mapped on two homoeologous segments of linkage groups (LG) 12 and 4, respectively. Neither of the *Mal d 3* genes was observed to be very diverse in the ten cultivars investigated; in particular, *Mal d 3.01*, which we assume to be the most important of the two, considering its expression in the fruit and its protein similarity to peach nsLTP, showed very little diversity. This suggests that the differences in Mal d 3-allergenicity among apple cultivars will mainly depend on the content of Mal d 3.01.

In **Chapter 3,** 18 *Mal d 1* genes were identified. Sixteen *Mal d 1* genes were located in two clusters, one cluster with seven genes on linkage group 13, and the other cluster with nine genes on the homoeologous LG 16. One gene was mapped on LG 6, and one remained unmapped. According to sequence identity, these 18 genes could be subdivided into four subfamilies. Subfamilies I, II and III had an intron of different size that was subfamily and gene specific. Subfamily IV consisted of 11 intron-less genes. The deduced amino acid

sequence identity varied from 65–81% among subfamilies, from 82%–100% among genes within a subfamily, and from 97.5%–100% among alleles of one gene. This study provided a better understanding of *Mal d 1* and formed the basis for further research on the occurrence of allelic diversity among cultivars in relation to allergenicity, which has been explored in Chapter 5.

Chapter 4 deals with the other two classes of apple allergen genes, *Mal d 2* and *Mal d 4*. Two copies of the *Mal d 2* gene (*Mal d 2.01A* and *Mal d 2.01B*) were identified, which differed primarily 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. Three different Mal d 4 isoallergens (Mal d 4.01, -02 and -03) were known from cDNA sequences (AF129426–AF129428). We obtained genomic sequences for each of these three isoallergens, which led to the identification of four *Mal d 4* genes (*Mal d 4.01A and B*, *Mal d 4.02A* and *Mal d 4.03A*). Their complete genomic DNA sequences varied among the genes in length from 862 to 2017 nt. They all contained two introns separating the protein coding sequences into three conserved exons of 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. According to the map position of *Mal d 2* and *Mal d 4.02A*, it is expected that similar genes may also be located on their homoeologous linkage group 17 and 7, respectively. New Mal d 2 (TLP) and Mal d 4 (profilin) EST sequences in the GenBank are positive signals to this expectation.

Chapter 5 identifies isoallergen genes that are potentially related to the differences in Mal d 1 allergenicity. The allelic diversity has been investigated the for seven intron- containing *Mal d 1* genes among ten cultivars. Next, the allelic constitution of 14 cultivars has been associated to the allergenicity of 14 cultivars as assessed by in vivo skin prick tests (SPT). *Mal d 1.01* and *Mal d 1.02* were conserved in amino acid sequence: nine out of ten cultivars had one variant of the protein, and only one cultivar had a second variant. *Mal d 1.04*, *Mal d 1.05* and *Mal d 1.06* were more variable in their amino acid sequence with three to six protein variants. Six to eight genomic alleles were identified for each gene. Skin prick tests on Dutch patients showed the difference in allergenicity of 14 cultivars. Santana had previously been proved to be low allergenic in contrast to its grandparent Golden Delicious, which is high-allergenic. The two homologous *Mal d 1* haplotypes on LG 16 of Golden Delicious were

found be replaced completely in Santana, but the gene on LG 6 was not changed and those on LG 13 were at the most slightly changed. This suggests that *Mal d 1* genes on LG 16 are relevant to apple allergenicity. Further analysis on the 14 cultivars revealed that allergenicity is indeed associated with the genetic constitution of LG 16, especially with the allelic composition of *Mal d 1.06A*. The favourable alleles for low allergenicity can be detected by the presence of the Mal d 1.06A-ssr-154 marker.

Chapter 6 discusses a number of issues regarding the strategies for cloning and mapping of allergen genes, the occurrence of allergen genes in multigene families, the relevance of linkage map position of apple allergen genes to other Rosaceae fruit crops, and breeding perspectives towards the production of hypoallergenic apple fruits. Techniques like PCR cloning to identify the true alleles and correct SNPs of the four allergen gene classes are summarized. An overview of the map positions of these classes of genes on eight linkage groups is shown. These maps are useful as a reference for other Rosaceae fruit crops, such as pear and peach where similar allergen genes exist. To identify exactly which of the genes in a gene family are responsible for the allergenicity is a great challenge. It requires protein identification, genetic analysis and linking the different gene members to their respective proteins. In addition, accurate diagnosis of allergenicity of a given food is not easy. Combining our genetic and SPT data enabled to assess the involvement of specific gene members in allergenicity. As a model for reducing the incidence of apple allergy in Europe, identification and production of low-allergenic apple cultivars is a direct benefit to the patients. In the future, by using modern technologies such as marker-assisted breeding and gene-knockout approaches, new apple varieties should enter the market which will allow apple allergic patients to eat apples daily without any problem.

Samenvatting

Hoewel appel bekend staat als een gezonde vrucht, kunnen bij het eten van een appel bepaalde eiwitten bij sommige consumenten allergische reacties oprepen. Er zijn vier groepen van dergelijke eiwitten in appel bekend. Dit zijn de allergene eiwitten Mal d 1, Mal d 2, Mal d 3 en Mal d 4. In dit proefschrift worden genen beschreven die voor deze eiwitten coderen, en worden deze genen gelokaliseerd op de genetische kaart van appel. Er zijn in totaal 26 genen gevonden, waarvan 18 voor Mal d 1, twee voor zowel Mal d 2 als Mal d 3, en vier voor Mal d 4. Voor patiënten in Noord West Europa wordt Mal d 1 als het belangrijkste allergeen van appel beschouwd. Uit dit onderzoek blijkt dat hierbij chromosoom 16 een belangrijke rol speelt. Op dit chromosoom liggen negen *Mal d 1* genen, waarvan het gen *Mal d 1.06A* een sterke associatie heeft met allergeniciteit. Blijkbaar is de aanwezigheid van bepaalde allergeenvarianten belangrijker dan de totale hoeveelheid Mal d 1 eiwit waar medisch onderzoek tot nu toe op gericht was.

Appel is de meest geteelde fruitsoort in de gematigde streken. De genetische analyse van appel is in een versnelling geraakt door de ontwikkeling van genetische kaarten en technieken waarmee de basenvolgorde van genen snel en goedkoop bepaald kan worden. Dit heeft onder andere merkers opgeleverd die gekoppeld zijn aan resistenties tegen diverse plantenziekten. Een deel van deze merkers worden op dit moment al in de appelveredeling toegepast voor het vroeg opsporen van potentieel resistente zaailingen. Hoewel appel bekend staat als een gezonde vrucht, kunnen sommige mensen toch geen appels eten omdat ze er allergisch voor zijn. Voedselallergie is een overgevoeligheidsreactie tegen stoffen (allergenen, meestal eiwitten) die normaalgesproken onschadelijk zijn. De reactie betreft een humorale immuunresponse, waarbij immunoglobuline E (IgE) een sleutelrol speelt. Appelallergie komt relatief veel voor in Europa. Ongeveer 2% van de bevolking heeft er last van. Sinds 1948 wordt er intensief onderzoek gedaan aan appelallergie. Dit onderzicht richt zich op het vóórkomen van deze vorm van allergie én op de allergenen die erbij betrokken zijn. Tot op heden zijn er vier appelallergenen geïdentificeerd. Dit zijn Mal d 1, -2, -3 en -4. Deze benaming heeft betrekking op de Latijnse naam van de appel, *Malus domestica*. Volgens de gangbare nomenclatuur wordt een allergeen genoemd naar de eerste drie letters van het geslacht van het organisme waarin het allergeen is aangetoond, gevolgd door de eerste letter van de soortnaam en een nummer dat aangeeft in welke volgorde deze allergenen beschreven

zijn. Bij het begin van dit promotieonderzoek waren van elk van de vier allergenen enkele DNA sequenties beschikbaar. Op basis hiervan is het onderzoek verder uitgebouwd. Hieronder volgt een korte beschrijving van de vier allergenen.

• **Mal d 1**, is het belangrijkste allergeen van appel. Dit allergeen geeft een kruisreactie met het IgE van patiënten die gesensibiliseerd zijn door het Bet v 1 allergeen van de berk. Berkenstuifmeel-gerelateerde appelallergie komt veel voor in Noord en Centraal Europa. De symptomen zijn overwegend mild en beperkt tot het orale gebied (vandaar de benaming 'orale allergie syndroom', OAS) en leidt tot jeuk, tintelingen en zwellingen van lippen, tong en keel na het eten van een appel. Het is gebleken dat de allergeniciteit van appels van ras tot ras sterk kan verschillen. Mal d 1 is ook een eiwit dat vaak aangemaakt wordt wanneer appels met ziekteverwekkers in contact komen. Het behoort daarom ook tot de zogenaamde pathogenese-gerelateerde (PR) eiwitten, in dit geval tot de PR10 familie.

• **Mal d 2** vertegenwoordigt een van de belangrijkste eiwitten in de rijpe appelvrucht. De geografische regio waar Mal d 2 allergie bij voorkeur voorkomt is nog niet duidelijk beschreven. Het is een homoloog van het thaumatine-achtige eiwit (TLP) dat tot de pathogenese-gerelateerde eiwitfamilie 5 (PR5) behoort.

• **Mal d 3** is het niet-specifieke lipide-overdragende eiwit (nsLTP). nsLTP's zijn als allergeen geïdentificeerd in diverse vruchten van de familie der roosachtigen, zoals appel, perzik, abrikoos, pruim en kers. Het kan ernstige allergische reacties veroorzaken in het maagdarmkanaal en kan zelfs leiden tot een anafylactische shock. Deze vorm van allergie komt vooral in mediterrane landen zoals Spanje en Italië voor. Verondersteld wordt dat patiënten gesensibiliseerd worden door perzik LTP. LTP behoort tot de pathogenese-gerelateerde eiwitfamilie14 (PR14).

• **Mal d 4** is appelprofiline, waarvan men vermoedt dat dit kruisreageert met het profiline in berkenstuifmeel, Bet v 2. De klinische relevantie van profiline als allergeen is niet duidelijk.

Dit proefschrift is een van de resultaten van het onlangs beëindigde EU-project EU-SAFE: "Plant food allergies - field to table strategies for reducung their incidence in Europe" (QLK-CT-2000-01394). Dit project is opgezet om de allergenen te karakteriseren die van belang zijn voor het optreden van ernstige en milde klachten. Appel is in dit project gebruikt als model gewas. Appelrassen verschillen onderling sterk in de mate waarin ze allergische reacties veroorzaken.

Dit biedt perspectieven voor het maken van nieuwe, laag allergene rassen via klassieke veredeling. Om efficiënt te veredelen, is het van belang om de genen te kennen die bepalend zijn voor de allergeniciteit van een ras. Dit kunnen aspecten zijn die betrekking hebben op kwalitatieve eigenschappen van de allergene eiwitten (isoallergen-varianten), of de hoeveelheid eiwit die geproduceerd wordt.
Dit proefschrift is gericht op het opsporen en karakteriseren van de genen die bepalend zijn voor de eiwitsamenstelling van de vier belangrijkste appelallergenen. Karakterisering van genen vindt plaats op het niveau van hun DNA sequenties, hun positie op de genetische kaart, en de aminozuursequentie van het eiwit dat ze coderen. Een andere doelstelling van het project is de ontwikkeling van genetische merkers waarmee de aanwezigheid van laag allergene gen varianten in een jong stadium van de plant aangetoond kan worden.

De gevolgde onderzoeksstrategie gaat uit van het sequencen van allergeen genen vanuit het genomisch DNA van de twee appelrassen Prima en Fiesta. Dit maakt het mogelijk om kleine DNA verschillen op te sporen. Vervolgens worden er moleculaire merkers ontwikkeld die deze verschillen zichtbaar maken. Voor Prima en Fiesta zijn genetische kaarten beschikbaar. Dit maakt het mogelijk om de nieuwe merkers snel op de genetische kaart van appel te lokaliseren. Deze strategie is succesvol gebleken.

Hoofdstuk 1 geeft en algemene inleiding in het onderzoeksveld.

Hoofdstuk 2 beschrijft de identificatie van twee genen voor Mal d 3 (*Mal d 3.01* en *Mal d 3.02*), waarvan het tweede gen nieuw is. Beide genen bleken geen intron te bevatten. Ze liggen op de chromosomen 12 en 4. De diversiteit in de DNA sequentie van deze twee genen is onderzocht in acht andere rassen. Hierbij werden echter nauwelijks verschillen gevonden.

Hoofdstuk 3 beschrijft de identificatie van 18 genen voor Mal d 1. Deze genen blijken meestal in groepen voor te komen. Één groep van 7 genen ligt op chromosoom 13. Een andere groep van 9 genen ligt op chromosoom 16. Een enkelvoudig gen ligt op chromosoom 6, terwijl de positie van het laatste gen niet vastgesteld kon worden. Op basis van hun sequentie konden deze 18 genen in vier subfamilies onderverdeeld worden. De genen van subfamilie I, II en III bleken een intron te bevatten waarvan de lengte specifiek was voor de subfamilie en het gen. Subfamilie IV omvatte 11 intronloze genen. De afgeleide aminozuurvolgorde vertoonde een overeenkomst van 65 tot 81% tussen subfamilies, van 82 tot 100% tussen genen binnen een subfamilie, en van 97.5 tot 100% tussen allelen van een enkel gen. Dit

onderzoek heeft tot een beter begrip geleid van de genetica van Mal d 1. In verder onderzoek is gekeken welke van deze genen allergische reacties kunnen veroorzaken (zie Hoofdstuk 5).

Hoofdstuk 4 beschrijft de identificatie van twee genen voor Mal d 2 en vier genen voor Mal d 4. De *Mal d* 2 genen liggen vlak naast elkaar op chromosoom 9. Op dit chromosoom liggen ook twee *Mal d 4* genen (*Mal d 4.01A* en *Mal d 401.B*). De twee andere Mal d 4 genen liggen op chromosoom 2 (*Mal d 4.02A*) en op chromosoom 8 (*Mal d 4.03A*). Op basis van de opbouw van het genoom wordt verwacht dat er nog andere *Mal d 2* genen op chromosoom 17 aanwezig kunnen zijn, en andere Mal d 4 genen op chromosoom 7, 15 en 17. De twee Mal d 2 genen lijken zeer sterk op elkaar. Hun eiwitten verschillen slechts in één aminozuur en hun intron verschilt slechts in twee basenparen. Het is daardoor niet waarschijnlijk dat de verschillen in allergeniciteit tussen appelrassen door dit allergeen veroorzaakt worden. De *Mal d 4* genen vertoonden wel grote verschillen voor hun coderende sequenties en in de lengte van hun introns. De lengte van het coderende gedeelte was wel sterk geconserveerd. Deze bestond steeds uit 3 exons van respectievelijk 123, 138 en 135 nucleotiden.

Hoofdstuk 5 laat zien dat chromosoom 16 een rol speelt bij de verschillen in allergeniciteit van appelrassen. Met name het gen *Mal d 1.06A* op chromosoom 16 gaf een goede associatie tussen allel sequentie (isoform-variant) en allergeniciteit. Het haplotype van chromosoom 16 voor lage allergeniciteit bleek gekarakteriseerd te worden door de aanwezigheid van de merker Mal d 1.06A-ssr-154. Hiermee is een essentiële stap gemaakt in de richting van merkergestuurde veredeling op lage allergeniciteit. Dit resultaat werd verkregen door de allelische constitutie van zeven *Mal d 1* genen van 14 onderling verwante appelrassen te bepalen, en deze in verband te brengen met de mate van allergeniciteit van deze appelrassen. In dit onderzoek zijn alleen de *Mal d 1* genen betrokken die een intron hebben. Nader onderzoek moet uitmaken of ook een of enkele van de intronloze genen van chromosoom 16 een rol kunnen spelen. Blijkbaar is de hoeveelheid van bepaalde allergeenvarianten belangrijker dan de totale hoeveelheid Mal d 1 eiwit waar medisch onderzoek tot nu toe op gebaseerd is.

In dit hoofdstuk is verder de variatie in gen sequenties tussen 10 appelrassen beschreven voor de zeven hierboven genoemde *Mal d 1* genen. Hieruit blijkt dat *Mal d 1.01* en *Mal d 1.02* eiwitten coderen met een geconserveerde aminozuurvolgorde: in negen van de tien cultivars kwam één eiwitvariant voor; slecht één cultivar bevatte een tweede variant. De aminozuur-

sequenties van Mal d 1.04, Mal d 1.05 en Mal d 1.06 vertoonden met drie tot zes eiwitvarianten een grotere verscheidenheid.

Hoofdstuk 6 bevat een algemene discussie over het belang van dit onderzoek voor andere fruit gewassen, en over technische aspecten van dit onderzoek en de perspectieven voor de veredeling van hypoallergene appels.

Kennis over de genlokatie van de allergie genen van appel is bruikbaar voor genetisch onderzoek in andere fruitsoorten zoals peer en perzik bruikbaar omdat deze overeenkomstige allergenen bevatten, en omdat bekend is dat hun genomen op vergelijkbare wijze georganiseerd zijn.

Enkele technische aspecten die besproken worden hebben betrekking op het kloneren van allergeen genen, het plaatsen van deze genen op de genetische kaart, het vóórkomen van allergeen genen in genfamilies, het voorkomen van foutieve sequenties ten gevolge van PCR-artefacten.

Uit dit proefschrift blijkt dat de exacte identificatie van de genen in een genfamilie die bij de allergeniciteit betrokken zijn, een grote uitdaging is, enerzijds omdat er meerdere allergenen een rol kunnen spelen en er voor elk allergeen meerdere genen (iso-allergenen) en allelen (varianten) aanwezig kunnen zijn, en anderzijds omdat patiënten onderling verschillen in hun gevoeligheid voor deze allergenen, iso-allergenen en varianten. Dit alles maakt een nauwkeurige diagnose van de allergeniciteit van een voedingsproduct niet eenvoudig. Onze combinatie van genetische data met de resultaten van de huidpriktesten maakte het evenwel mogelijk om de betrokkenheid van specifieke genen bij allergeniciteit aan te tonen. Als een model voor de verlaging van de incidentie van appelallergie in Europa levert de identificatie en de productie van laag allergene appelcultivars een rechtstreeks voordeel voor de patiënten. In de toekomst zullen, mede door de resultaten in dit proefschrift, en gebruikmakend van moderne technologieën zoals markergestuurde veredeling en uitschakeling van genexpressie (gene silencing) nieuwe hypoallergene appelcultivars op de markt gebracht kunnen worden die appelallergische patiënten in staat stellen een appel te eten zonder daar enige last van te krijgen.

题目：候选过敏原基因在苹果基因组上的定位及其过敏性推测

高 中 山

摘　　要

栽培苹果（*Malus domestica*，2x=34）是世界温带区最重要的水果之一。由于分子标记连锁图谱的建立和基因序列的测定，目前苹果遗传分析和育种进程在加速发展。一些抗病分子标记已在育种中应用。

尽管对于大多数欧洲人来讲，苹果是一种推荐的健康食品，但部分人（大约 2%）却因过敏反应而不能吃苹果。食物过敏是以免疫球蛋白 E（IgE）为中介的对无毒物质（过敏原，绝大多数为蛋白质）的超敏反应。从 1948 年开始，许多研究者对苹果过敏的临床观察，过敏原的鉴定做了大量调查研究。目前，有四类过敏原（分别称作 Mal d 1～Mal d 4）存在于苹果果实内，相关的 DNA 序列发表在 GenBank 和 EMBL 数据库中。

Mal d 1 为苹果主要过敏原，在蛋白质分类中属于植物发病相关蛋白（Pathogenesis related, PR）第 10 家族。它与桦树花粉中的过敏蛋白 Bet v 1 致敏产生的 IgE 发生交叉反应。多数对桦树花粉过敏的北欧和中欧居民吃苹果会发生口服过敏综合症（OAS）。一般症状表现为口腔，嘴唇，舌头，喉咙发痒，刺激和发肿。根据患者反应，不同苹果品种引发的过敏反应程度不同，如金冠 (Golden Delicious) 较强烈，而格洛斯特（Gloster）较弱。

Mal d 2 是一个类似甜蛋白（TLP）的过敏原， 属于植物发病相关蛋白第 5 家族（PR-5）。它是成熟苹果主要蛋白质组分，其引发过敏的地理区域不太清楚，有人认为其重要性与 Mal d 1 相当。

Mal d 3 是一种非特异脂肪转移蛋白（nsLTP），隶属于植物发病相关蛋白第 14 家族。许多蔷薇科水果（苹果，梨，桃，李子，杏和樱桃）中的 nsLTP 会引发胃肠道过敏反应，主要发生在欧洲南部地中海国家，特别是西班牙和意大利。

Mal d 4 是苹果抑制蛋白（profilin），许多研究者认为它会与桦树类似的 Bet v 2 发生交叉反应。临床观察证据和影响人群目前尚不清楚。

论文中研究结果来自于欧盟第五框架 EU-SAFE 项目（植物食品过敏：用来降低欧洲发生频率的田间到餐桌策略，QLK-CT-2000-01394）。此研究项目选择苹果为模式作物，研究导致在不同品种间存在过敏性差异的原因。其中的一个目标是把过敏原基因定位在连锁群图谱上，为将来低过敏苹果育种奠定基础。

主要讲究方法：根据已知的 DNA 序列设计 PCR 引物。选择一个参照作图群体的两个亲本品种——普瑞马（Prima）和菲斯达（Fiesta）为研究试材，通过克隆和测序发现两个品种间的单核苷酸多态性（SNPs），并依此设计产生基于 PCR 琼脂糖胶的分子标记（SNAP）。然后将在后代群体中分离的过敏原基因定位于参照图谱中，通过遗传多样性和系谱分析发掘重要候选基因。

论文的第二章对 2 个 *Mal d 3* 基因进行定位和多样性分析。*Mal d 3.01* 与已发表的 cDNA 序列（AF221502）100％同源，*Mal d 3.02* 为新发现的苹果 nsLTP 基因。这两个基因均没有内含子，分别定位于第 12 和第 4 连锁群的两个部分同源区段。分析了这两个基因在 10 个品种的测序多样性，结果表明二者都比较保守，没有太多的氨基酸变化，尤其是 *Mal d 3.01*, 但它确实在果实中表达而且与桃的 nsLTP 更相似。这说明不同品种的 Mal d 3 过敏性差别将要取决于过敏原 Mal d 3.01 含量的多少，而不是蛋白质种类的变化。

第三章主要论述了 *Mal d 1* 基因家族 18 个成员的基因序列和图谱位置。其中 7 个基因集中（集束）位于 13 连锁群， 9 个基因集束位于对应的部分同源的 16 连锁群。有一个例外的基因位于连锁群 6，还有一个基因没有定位。根据 DNA 序列的同一性，这 18 个基因可再分为四个亚家族。亚家族Ⅰ、Ⅱ、Ⅲ的基因都有一个内含子，但长度依亚家族和基因而别。亚家族Ⅳ包含 11 个没有内含子的基因。比较根据这些基因推导的氨基酸序列，亚家族之间拥有 65～81%的相同性，同一亚家族的成员间拥有 82～100%相同性，等位基因间拥有 97.5～100%的相同性。

第四章定位了其余两类过敏原基因，*Mal d 2* 和 *Mal d 4*。基因测序得到两个拷贝的 *Mal d 2*（*Mal d 2.01A* 和 *Mal d 2.01B*），两者主要差别在于内含子长度（378 和 380 个碱基）和信号多肽中一个氨基酸的不同。这两个基因均位于连锁群 9，且在同一位置，其中 *Mal d 2.01A* 与已经报道的 *Mal d 2* 序列相同。我们定位了 4 个 *Mal d 4* 基因(*Mal d 4.01A* 和 *B, Mal d 4.02A* 和 *Mal d 4.03A*)，它们分别来自 3 个已知 *Mal d 4* 同功过敏原（isoallergen）。这 4 个基因均属于抑制蛋白（Profilin），都有同样长的 396 核苷酸（nt）

编码区。基因组测序长度为 862～2017nt, 都有两个内含子，且三个外显子均为 123，138，135nt，Mal d 4.01 是一个重复基因（Mal d 4.01A 和 Mal d 4.01B），位于连锁群 (LG) 9，*Mal d 4.02A* 和 *Mal d 4.03A* 均为单拷贝，分别位于 LG 2 和 LG 8。

第五章结合 14 个品种的过敏性皮试结果，着重分析了 7 个有内含子的 Mal d1 基因在 10 个品种上的多样性。分析表明 *Mal d 1.01* 和 *Mal d 1.02* 基因都比较保守，在 10 个品种的 9 个推导出同一个氨基酸序列。*Mal d 1.04*，*Mal d 1.05*，*Mal d 1.06A-C* 变化较多，分别发现 3～6 种蛋白序列，而且每个基因都有 6～7 个同义突变。一个新品种桑塔那（Santana，荷兰国际植物所培育）的过敏性很低，但它的祖母亲本金冠为最高。分子标记谱系分析表明，位于第 16 连锁群上的 *Mal d 1* 基因对与 Mal d 1 相关的过敏性起决定作用，而非那些位于 LG 6 或 LG 13 的 Mal d 1 基因。LG 16 上的 Mal d 1 单配子基因组合和过敏性皮试结果间的关联分析表明，低过敏苹果品种拥有特殊的单倍性基因组合，并且可以用 Mal d 1.06A-SSR 标记鉴别。

第六章对本论文采用的实验方法、候选过敏基因家族、连锁图谱应用以及育种前景进行深入讨论。苹果过敏原基因连锁图对于其它蔷薇科果树上的类似基因研究有很大的参考价值。进一步重组表达已鉴定出不同过敏原对于过敏性研究和诊断治疗有潜在利用价值。可以预见，利用现代生物技术和传统育种方法，新一代低过敏性苹果品种在不久的将来问世并投入生产，这无疑是苹果过敏患者的一个福音。

Acknowledgements

It is eight years since I first came to Wageningen in January 1997. I have worked at Plant Research International (PRI) for more than five years over three periods. My PhD research began in 2001 during my second visit. After a relatively long time pursuing a PhD degree at Wageningen University, my dream has now finally been realized. This work and training experience will be of immense importance to my future career. It is my great pleasure to express my gratitude to those who have given me support, help and friendship throughout these years.

First, I owe a great deal to my copromoters Dr. Eric van de Weg and Dr. Luud Gilissen. Both have given me excellent scientific supervision, kind heartedness and sincere friendship. The fruit of my PhD degree is a result of their initiative and continuous cultivation.

Eric, you guided me to learn various molecular marker techniques and applications in genetic analyses throughout these years. More importantly, you led me to the right track in the apple allergy project, which enabled me to produce this thesis. I still remember that evening walk on the streets of Vienna in 2001 when we discussed seriously my intent to obtain a degree from WU. Your insight on genetic analyses impressed me. Your in-depth and critical comments on the experimental chapters in my thesis have been a great contribution to making them logical and concise. I have learned a lot from you. I have also been moved by your tireless effort to improve my thesis. Apart from the research, you cared about my personal life in Wageningen every time I came or left. Our friendship extends to our families.

Luud, I was lucky to join your part of the EU-SAFE project. Your ability to explore new research areas that combine various disciplines is a phenomenon to me. You communicate easily with a broad audience about your research. Your help with my English in this thesis was instrumental to my finishing it. I would not have finished it without your help. In addition, your managing of my studentship for the last three years was essential to me being able to be at PRI.

Second, I wish to thank Prof. Evert Jacobsen and Prof. Richard Visser, my two promoters. I am both grateful and honored to be one of your PhD students. In the early stage of outlining the thesis, both of you provided thoughtful comments and suggestions. Your critical reading and valuable comments on the thesis improved its readability. Your stimulating interests on apple allergy encouraged me to accomplish more and to demonstrate my will and ability, especially in the last months of finalizing the thesis when Evert pressed me to continue without delays. It has been proven that was a right a decision.

I am grateful to PRI colleagues who contributed directly to this thesis; they are Jan Schaart, Ingrid van der Meer, Catarina Matos, Linda Kodde, Henk Schouten, Hong Tran, Lonneke

van der Geest, Jan Kodde, Gert van Arkel, Dick Bosch and Yinghui Li. I also thank colleagues involved in the EU-SAFE project for collaboration and communication, especially to Karin Hoffmann-Sommergruber, Heimo Breiteneder, Suzane Bolhaar, Andre Knulst, Laurian Zuidmeer, Ronald van Ree and Margit Laimer for your inputs.

For experimental techniques, Gert van Arkel and Jan Schaart taught me a lot; you are the most important source of my lab skills. I am also indebted to Sjaak van Heusden for his supervision in molecular markers when I first came to Wageningen in 1997. Other colleagues in both business units of Genetics and Bioscience gave me much help. Ria Vrielink-van Ginkel, Linda Kodde, Catarina Matos, Tetty Chaidamsari, Wagner Augusto Benedito and Koen Pelgrom provided assistance in the lab. Chris Maliepaard helped me with the JoinMap software. I acknowledge Ton den Nijs and Roul Bino for inviting me to be a visiting scholar; Penelope Nestel for helping me with my English; Leo van Griensven, Ludo Luckerhoff, Ralph Litjens, Bert Schipper and Sun Zhongkui, who worked in the same office, for both the good atmosphere and their assistance. Jan Cordewener and Twan America provide valuable advice on analyzing MALDI-TOF data.

Many thanks go to my Chinese colleagues at PRI and friends that I got to know during my five years in Wageningen, in particular Zheng Sijun, Ge Yaxin, Li Yaling, Zhang Liru, Bai Yuling, Yan Zifu, Ji Qing, Liu Chunming, Yin Xinyou, Jing Haichun, Zhang Lei, Sun Zhongkui, Zhou Miaoping, Shi Zhiqi; Ren Xinying, Li Shipeng, Wang Xuekui, Wang Caiyun, Jing Qi, Zhong Jiayou, Xing Meiqing, Li Juming, Liang Yehong, Qi Yongyan, Yang Limei, Huang Sanwen, Zhou Shujun, Zhao Junliang, Hu Fuyi, Yu Yanfang, Lu Hongmei, Wang Miqia, Yan Liying, Tang Jifeng, Li Yinghui, Wang Chenzhu and Zhang Qiang.

I would also like to thank the Shanxi Government, China; IAC, the Netherlands; and the World Laboratory, Switzerland, for providing me fellowships to conduct my research at PRI. My sincere thanks go to Mr. Cui Jiguang, director of Experimental Nursery of Shanxi Forestry Department, who has supported me in many ways.

I dedicate this thesis to my beloved wife, Bi Shujun, and daughter, Gao Biyuan. My absence has been long, and that time has been tough for you. I express my gratitude and deep regret. Hopefully, the hard times are over, we shall be together soon. I should do something to compensate and repay for that suffering. My daughter drew an apple for the cover of my thesis. I am proud of her drawing ability. This thesis is also dedicated to my whole family, especially in memory of my father who encouraged me to be a determined scholar.

Gao, Zhongshan

Wageningen, the Netherlands, 18 May, 2005

Curriculum Vitae

Gao Zhongshan was born on the December 6th of 1964 in Shuozhou, Shanxi Province, China. From 1981 to 1988 he studied at the Horticultural Department, Shanxi Agricultural University, China, where he obtained bachelor and master degrees in agriculture. After graduating, he worked at the Shanxi Forestry Research Institute for nine years (1988–1996 and 1999) on walnut variety selection, propagation and cultivation. The major research achievement was the releasing of two new walnut varieties. In year 1997 and 1998, he worked at the Department of Vegetables and Fruit Crops, CPRO-DLO (now incorporated into Plant Research International, Wageningen, The Netherlands) as a visiting scientist. He participated in projects on apple molecular marker linkage mapping and genetic analysis of apple scab resistance. In March 2000, he moved to Experimental Nursery of Shanxi Forestry Department, Taiyuan, Shanxi, where he played a major role in application for two big projects (financed by the central government) on improving production and research facilities for forest nursery. In March 2001 he returned to Plant Research International (PRI) and participated in the EU-SAFE project to map four classes of apple allergen genes on apple molecular marker maps. In 2002, he was back to China to implement the two projects at Experimental Nursery of Shanxi Forestry Department. From March 2003 to June 2005, he continued the research on the apple allergy at PRI, which led to the completion of his PhD thesis. After the graduation, he will return to China.

List of Publications

Gao ZS, Van den Weg WE, Schaart JG, Van der Meer IM, Kodde L, Laimer M, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJWJ (2005). 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

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 (2005) Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). Theor Appl Genet (in press)

Gao ZS, van de Weg WE, Schaart JG, van Arkel G, Breiteneder H, Hoffmann-Sommergruber H, Gilissen LJWJ (2005) Genomic characterization and linkage mapping of the apple allergen genes Mal d 2 (Thaumatin-like protein) and Mal d 4 (Profilin) Submitted to Theor Appl Genet

Gilissen LJWJ, Bolhaar STHP, Knulst AC, Zuidmeer L, Van Ree R, **Gao ZS,** Van de Weg WE (2005) Production of hypoallergenic plant foods by selection, breeding and genetic modification. In Gilissen LJWJ, Wichers HJ, Savelkoul HFJ, Rogers RJ (eds) Allergy Matters: new approaches to allergy prevention and management, Wageningen UR Frontis Series Vol.10, Springer Science+Business Media, p95–104

Gao ZS and WE van de Weg (2005) The most widely explored Vf gene for scab resistance in apple is linked to sub-lethal genes. Submitted to Euphytica

Li KZ and **Gao ZS** (1990) Preliminary study on the mechanism of fruit cracking in Chinese jujube, J. of Fruit Sci. 1990(4)221-226

Wang G, **Gao ZS** (1992) A new walnut variety -Jinlong No.1, Acta Horticulture Sinica, 1992,19 (4) 287-288

G. Wang and **Gao ZS** (1993) Present extending situation and further developing direction of improved walnut varieties in Shanxi province, Economic Forest Researches, 1993 (1) 18-22

Lü ZS. Wang G and **Gao ZS**, New walnut cultivars and integrated cultural techniques for high yield and quality. Shanxi Higher Education Association Press, Taiyuan, 1993.

Abbreviations

AFLP	Amplified fragment length polymorphism
Bet v	*Betula verrucosa*
CAP	Cleavage amplified polymorphism
DBPCFC	Double-blind placebo controlled food challenge
ELISA	Enzyme-linked immunosorbent-assay
IgE	Immunoglobulin E
LG	Linkage group
LTP	Lipid transfer protein (Mal d 3)
Mal d	*Malus domestica*
MALDI-TOF	Matrix-assisted laser desorption/ionization-time-of -flight
OAS	Oral allergy syndrome
PR	Pathogenesis related
QTOF	Quadrupole time-of flight spectrometry
RAPD	Randomly amplified polymorphism DNA
RAST	Radio allergosorbent test
RFLP	Restriction fragment length polymorphism
RIA	RAST inhibition assay
RNAi	RNA interference
SNAP	Single nucleotide amplified polymorphism
SNP	Single nucleotide polymorphism
SPT	Skin prick test
SSR	Simple sequence repeat
TLP	Thaumatin-like protein (Mal d 2)

Cover illustration: Gao Biyuan
Reproduction: Ponsen & Looijen B. V., Wageningen